Position sensitive detector technology, data reduction and count rates

Introduction

Over the years I have found that there exist quite some misunderstandings and misconceptions about position sensitive detectors and data collecting issues. This text is intended to shed some light on different aspects of detectors. It should be clear from the start that I'm convinced of the fact that the ideal detector, suitable for all applications, does not exist[1]. Also be aware that this is an introduction to detectors and not an 'every aspect and final word' document. I leave the writing of that document to the real experts. Also one should keep in mind that technology is changing in time. My aim is to show some issues and let the individual researcher determine for him/her/itself what to him/her/it are the important ones. I'm aware of the fact that some people will make a career out of remapping detector grids and also I'm also aware of the tendency of people to ignore issues that they think are too complicated or not consistent with their prejudices. As long as one gets nice reliable results without over- or under-interpreting the data I don't mind. The thing I'm really allergic to is to publish nonsense and afterwards use the comment 'but no one ever told me…..'
My background in X-ray research is time-resolved Small Angle X-ray Scattering using synchrotron radiation sources, where I developed my own particular set of prejudices. This document is intended for students who use this technique but it is not forbidden for users of other techniques to read and think about it (and comment). I also would like to stress that I have detector developers, second hand car dealers and even some pachyderms, but very few smooth talking politicians, among my friends. My mate Chris Hall, who is a detector developer, has shed a critical eye on this text[2].

[1] To use the phrase of the great and wise, Amsterdam born, football professor and philosopher Johan Cruyff: 'Ellek foordeel hep se nadeel'. (For the Anglosaxons: 'Effery afuntage hass hiss dissadfuntage')
[2] But should not share the blame for errors and misconceptions on my part.

Detector technology with respect to count rates

Position sensitive detectors exist in several varieties but they can be subdivided into two broad categories. The first are photon counting devices and the second are the integrating detectors.

Photon counters

This type of detector counts incoming photons as individual events. The detector itself generates a signal when a photon arrives and the contents of a counting register are incremented with one. This might be done in a slightly convoluted way like for instance in gas filled wire chambers equipped with a delay line in which the electronics generate an address in a multiscaler register. The gas counting technology can also be used with small detecting elements each equipped with it's own individual scaler. The strong point of these detectors is that after the time frame is finished one only has to read out the digital numbers stored in the memories. The read-out process doesn't influence the contents of the registers. The weak point of these detectors is that the detection of each individual photon is a rather time consuming process during which the detector is not registering other incoming photons. This intrinsically limits the count rate capacity of these detectors. For instance when one uses a 100 nsec long delay line one can only detect 10 x 10^6 counts/sec (in detector jargon 10 MHz). This is an average rate for randomly arriving events. To be on the safe side we usually say 1 Mhz.
In practice even this figure is not obtained with delay line detectors. When individual pixels are equipped with their own counters the limiting factor becomes the detection process in the gas. This improves this situation dramatically so that a detector with a similar active length can handle 400 MHz. The great advantage of these detectors is that it is possible to detect a very high dynamic range in the pattern. This is the difference between the highest and lowest scattered intensity in a scattering pattern. This difference can become 1:10^{6-7} with this type of detector. Another disadvantage is that the detecting element can not be made too small without risking that the physical process inside the detector will not trigger of a single pixel but also it's adjacent ones. The limit for this is roughly 200 – 300 micron for a gas detector. Smaller for solid state detectors. Yet another disadvantage is that gas detectors need some gas depth in order to be able to detect photons efficiently. When a photon hits the detector at an angle this leads to a parallax problem in which a photon is detected at a different position than where it should be. This problem is practically not very well suitable for mathematical correction.

Integrating detectors

This type of detector uses a solid state physics process in which a charge is generated when a piece of semiconducting material is being hit by a visible light photon (CCD detector). The visible light is generated by the interaction of the X-ray with a phosphor material. The light is then channeled to the detecting chip by either a lens or by large bundles of fiber optics. Unavoidably some photons will be lost in this way which means that the efficiency of the detector is somewhat lower than a gas detector[3]. The charges are accumulated in a small capacitor which at the end of the time frame has to be read out. The charge is then converted

[3] However, this is very much dependent on the energy of the incoming photon. A gas detector will have difficulties to efficiently detect photons above 15 keV, unless the gas is pressurised.

to the approximate number of photons that has hit the detecting pixel. The big advantage of this type of detector is that the count rate is nearly unlimited[4].

There are some disadvantages. The first of these is that during the read-out process of the pixel some noise is added due to the read-out process. This is on top of the noise that is accumulated on top of the signal during data collection. The second is that not every pixel is equipped with it's own physical connections so that the read-out takes place by reading out the first pixel in the row, erase this one, shift the contents of pixel number two to number one, read this etc. etc. This makes the read-out process for the total detector cumbersome and slow. One can speed this process up but this is generally at the expense of an increase in read-out noise. The third problem is that the capacitor used for the charge accumulation only has a limited capacitance. This means that when it is full it has to be read out. This means that all other pixels will have to be read out as well even the ones that are not completely full. Since every read-out cycle adds read-out noise the weaker signals will suffer strongly thus limiting the dynamic range. Electronic crosstalk between the pixels and horizontal drift in the phosphor will reduce the spatial resolution of this type of detector somewhat but in general it is better than the spatial resolution of a gas detector.

Summary of detector properties

In the table below one can find a coarse summary of the properties of the two types of detectors.

	Counting	Integrating
Countrate	Relatively low	High
Spatial resolution	Low (200 – 300 micron)	High (25 micron)
Dynamic range	High 1:10^{6-7}	Low 1:10^{4-5}
Read-out noise	None	Weak signals will be influenced
Read-out speed detector	Fast (microseconds)	Slow (100 microseconds)
Maintenance	Not easy	easy
Parallax problems	At wider angles	None
Efficiency	High (5 – 12 keV) Low > 12 keV	Low but much better than gas detector > 12 keV

Obviously this is very much a generalization. It is possible to build detectors for specific purposes that will satisfy the needs for a single application. For instance it is possible to construct gas detectors with a very high spatial resolution. However, this will not be suitable for high count rate applications. The same as it is possible to photon count with solid state devices if one reads out very carefully (i.e. very slow).

The factor of maintenance is rather subjective but for a beamline scientist on a multi user synchrotron radiation beamline it is definitely not a parameter that can be neglected which means that for a beamline a detector system can be chosen which is sub-optimal for many applications but which operates reliably. The purchase and maintenance cost of a single system is in general so high that it will not be possible to equip a beamline with more than one system. (This leaves the station scientist often in the position that he has to defend his

[4] There is a rate limit for phosphors/scintillators, where the light output goes non-linear with x-ray flux. However, it is large. Anecdotally we have reached it at 10 Mhz from a 100 micron square beam = 10^9/mm^2/sec (Chris Hall)

choice for a certain detector system and therefore will slack of other types of detectors. This is a sin which normally fades away after the age of 35 or 10 years of experience on a beamline, whatever comes first).

Basic statistics of counting[5]

In the following text no attention is paid to the effects that electronic crosstalk between channels will have and every pixel is treated individually. Also Gaussian statistics is being used while the more correct procedure for the integrating case would be to use Poisson statistics. The main reason for this is that I think it more important to get a feel for what is possible and which parameters are important rather than writing the ultimate text on detectors. I'm not an expert in error propagation through calculations. For more expert views take a look at the book of Rabinovich.

The basic equation for primary data reduction reads:

$$I_s(q) = \frac{\frac{I_{sc}(q)}{A_{sc}(q)} - \frac{I_c(q)}{A_c(q)}}{F(q)}$$

$I_s(q)$ scattered intensity due to sample
$I_{sc}(q)$ scattered intensity due to sample and cell/beamline windows/air etc.
$I_c(q)$ scattered intensity due to cell/beamline windows/air etc.
$A_{sc}(q)$ intensity of transmitted beam with filled sample cell
$A_c(q)$ intensity of transmitted beam with empty sample cell
$F(q)$ pattern produced by even illumination of detector (flood field/detector response)

In this q actually should be expressed in the Cartesian coordinates $q(x,y)$ but for simplicity sake I will refrain from doing that. Now make some simplifying assumptions.
1- the detector response curve is much better than the actual data. This is a valid assumption in the case of dealing with proper experimenters.
2 - the scaling factors for the intensity have a zero error. This again is not without basis since this is a parameter integrated over time.

This reduces the equation to be considered for the error propagation to:

$$I_s(q) = I_{sc}(q) - I_c(q)$$

<u>Photon counting detector</u>

Defining error margins and dropping the q to reduce the writing:

$$I_c \pm \Delta I_c$$

$$I_{sc} \pm \Delta I_{sc}$$

With a photon counting detector one can basically assume Gaussian statistics. This gives for the error margin:

$$\Delta I_s = \sqrt{(\Delta I_{sc})^2 + (\Delta I_c)^2}$$

[5] There are lies, bigger lies and statistics. (inscription found on the wall of a toilet in Herculaneum, next to the inscription 'Kilroy was here')

Integrating detectors

With an integrating detector one finds two noise terms on top of the statistical noise intrinsic to any measurement. These are the dark-current and the read-out noise. The first is a parameter that is linear with time. The read-out noise is a function of the read-out speed of the detector. The faster the read-out is the higher this parameter is.

$$I_c \pm \Delta I_c \pm D$$

$$I_{sc} \pm \Delta I_{sc} \pm D$$

The dark current and read-out noise are independent of the countrate.

Now we can consider two regimes. The first where the countrate is sufficiently high that the intrinsic experimental error is much larger than the dark current and read-out noise:

$$\Delta I_{sc} >> D$$

In this case the integrating detector performs equally well as a photon counting detector when considering equal countrates. However, integrating detectors have in general a much higher countrate compared with photon counters so it is likely that an integrator will outperform a counter here.

The second regime is the opposite case where the detector noise is higher than the experiment noise:

$$\Delta I_{sc} << D$$

In this case the statistical error can be approximated by:

$$\Delta I_s = \sqrt{2}D$$

It will be clear that a counter outperforms an integrator here.

One could argue that with enough flux this is not too much of a problem since one will never enter into this regime. However, the limitation on integrating detectors is the full well depth of the integrating element. If this is over-filled the detector either saturates or has to be read-out again. Now, keeping in mind that a lot of SAXS patterns fall of in intensity as $I(q) \sim 1/q^4$ at wider angles it is not unlikely that at some point in the data range one will encounter this problem. In my opinion this is therefore closely related to what q-range one wants to cover in a single experiment.

As an example take the case where at q = 0.01 $Å^{-1}$ the intensity reaches the full well depth of a 16 bit detector. This is 2^{16} = 65536. Where do we reach the point, $q = z$, then where the counts, N, fall off to a real low level?

$$I(q = z) = \frac{I(q = 0)}{(q - 0.01)^4} = N$$

$$N = 1 \Rightarrow z = 4$$

$$N = 10 \Rightarrow z = 1.3$$

$$N = 100 \Rightarrow z = 0.4$$

For $N = 100$ the statistical error margin due to counting statistics is 10%. Now if the detector reads out fast, one can sum several frames to artificially increase the dynamic range. At the high counting end this is no problem but as soon as $\Delta I_{sc} \approx D$ one does not make any large gains in the low counting end of the spectrum. One should also keep in mind that *D* increases when one is reading out fast.

For a photon counting detector no such problems exist but it might take a longer time to reach the 2^{16} level due to the fact that the maximum countrate is much lower than with an integrator. So it is a matter of choosing between Scylla and Charybdis (or the devil and the deep blue sea).

The message in this is that if one just needs a pattern to study orientation effects, phase transformations etc. an integrating detector will be a good choice. This is the case in experiments where one is mainly interested in phase recognition like for instance high pressure experiments on polymers. If one needs a large dynamic range then one is much better of with a photon counter. This is the case in for instance solution scattering experiments. Some people will claim that photon counting detectors are 'best' for time resolved experiments as well. I prefer to leave this issue to be determined by the individual experimenter.

Small changes on a large background

When considering small changes in a time-resolved experiment there are some tricks that can be used. Consider the time-dependent version of the background subtraction at time t_1:

$$I_s(q,t_1) = I_{sc}(q,t_1) - I_c(q,t_1)$$

$$\Delta I_s(q,t_1) = \sqrt{(\Delta I_{sc}(q,t_1))^2 + (\Delta I_c(q,t_1))^2}$$

A similar formula holds for t_2 .
Now one is interested in the difference:

$$I_d(q,t_2 - t_1) = I_s(q,t_2) - I_s(q,t_1)$$

Conventionally this is obtained by subtracting two curves like in the equation above. Now the error margin in this difference is:

$$\Delta I_d(q,t_2 - t_1) = \sqrt{(\Delta I_{sc}(q,t_2))^2 + (\Delta I_c(q,t_2))^2 + (\Delta I_{sc}(q,t_1))^2 + (\Delta I_c(q,t_1))^2} \approx$$
$$\approx \sqrt{2}\Delta I_s(q,t_1)$$

However, if one assumes that the background doesn’t change, which is not completely unrealistic since we’re only dealing with small changes, one skips two mathematical operations and the error in the data becomes:

$$\Delta I_d(q,t_2 - t_1) \approx \sqrt{2}\Delta I_{sc}(q,t_1)$$

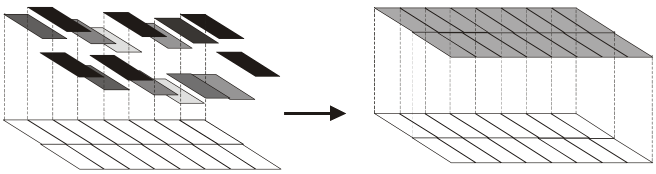
This is by definition smaller than the previous error. One should keep in mind that the reduction in the background due to the increase in the signal is not taken into account so that the total change in the actual signal is probably slightly smaller but this is offset by a much smaller error margin. This method looks particular favorable in the case of for instance the growth of a crystalline peak above an amorphous background. It should be noted that as long as the count rate on the detector is high enough the read-out noise of an integrating detector will not play a role.

Data reduction issues

Detector pictures are not always what they appear to be. Before one can start with data analysis several artifacts introduced by the detector should be taken care of. These problems are not the same for every type of detector. Even with supposedly identical detectors variations can occur.
The order in which this happens is sometimes important. Also one should be careful to keep the amount of mathematical operations, in which experimental scattering patterns are used, to a minimum since errors on the experimental data will propagate. For instance if one divides one pattern, with a statistical error margin of 5%, by another pattern with the same error margin, the resulting pattern will have an error margin of about 7%.

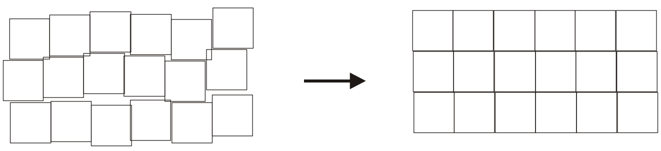
Pixel to pixel sensitivity

It is very unlikely that every pixel, or detecting element, in a detector has the same sensitivity. Some will count 9 out of 10 photons another will only detect 7. One has to compensate for this effect. See figure . The experimental procedure for this is to illuminate the detector with a uniform scattering pattern and divide the original pattern by this uniform pattern.



Remapping in the detector plane

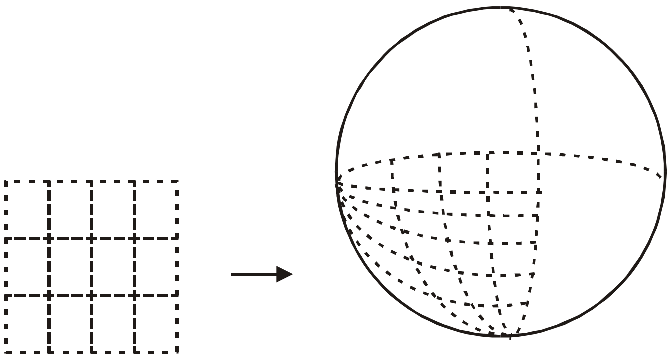
Ideally a detector is divided up in a grid with square pixels. This is often not the case in real life. In a CCD the detecting phosphor is coupled by fiber optics to the detecting chip. The individual fibers cannot be bonded with mathematical precision. In a wire chamber, equipped with a delay line, the delay line can introduce errors since the thickness of the line influences the speed of the electrons which means that sometimes 100 pixels are mapped on 1 cm line and at another position maybe 95. See figure.



One way of correcting this is to use a metal grid, with regular arrays of holes, which is placed before the detector. After this the detector is illuminated with a fairly uniform source. The dark stripes in the pattern should be visible as straight lines. If this is not the case the pixels have to be shifted and readjusted in dimensions. Sometimes there is software available to do this. However, one should try to asses how important this issue is in the total data reduction.

Remapping on Ewald sphere

Generally position sensitive detectors are flat. However, diffraction theory tells us that in principle one shouldn’t measure in a flat plane but everywhere the distance from the sample to the detector should be the same. This means that one has to use, in principle, a detector with a spherical surface. For linear detectors this is possible but for area detectors this is not practical. This means that one has to remap the square detector grid onto a spherical surface. In fact this is the inverse problem that producers of atlases face. This requires specialized software. See for instance the CCP13 software suite.



In small angle scattering this problem is much less important than in wide angle scattering. A cynical remark is that researchers working with biological fibers apparently have discovered this problem many years ago but that synthetic polymer researchers hardly see this as a problem. Another cynical remark is that people looking at biological fibers apparently have so little data that they can afford to spend years on the analysis of a single pattern.

Data noise simulations

In combined experiments, like for instance SAXS combined with WAXS or Raman spectroscopy, the issue sometimes crops up what the time-correlation is between parameters derived from the different techniques.

In the following the results of a simulation of noisy data are shown. The first curve (blue) represents the function:

$y = 40 \pm 2$ $\quad t \leq 40$

$y = 40 + 2t \pm 2$ $\quad t > 40$

The second curve (red) is:

$y = 10 \pm 6$ $\quad t \leq 40$

$y = 10 + 2t \pm 6$ $\quad t > 40$

The error margins represent the intrinsic statistical noise levels associated with experiments. The offsets of 40 and 10 respectively are purely for graphical reasons in this case.

The green curve is

$y = 10 \pm 6 + (3 \pm 1)$ $\quad t \leq 40$

$y = 10 + 2t \pm 6 + (3 \pm 1)$ $\quad t > 40$

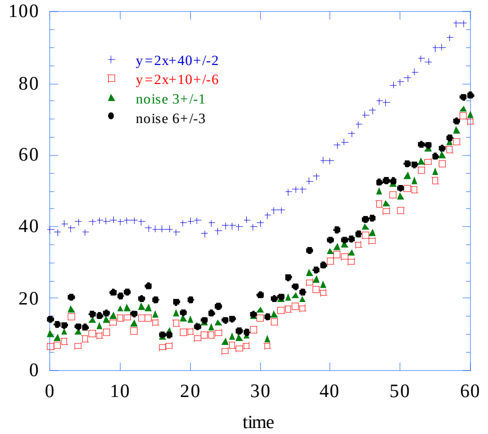
This is the same data as above but now with a background and noise level added. This could for instant represent the read-out noise of an integrating detector.

The black curve is again with extra background and noise added:

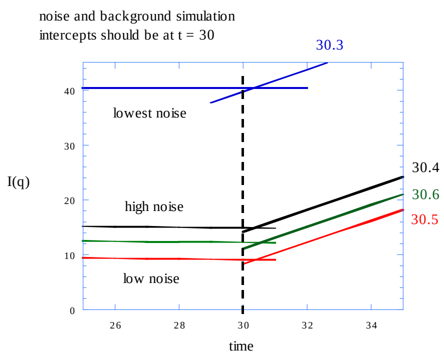
$y = 10 \pm 6 + (6 \pm 3)$ $\quad t \leq 40$

$y = 10 + 2t \pm 6 + (6 \pm 3)$ $\quad t > 40$

The results are shown below.



Now we want to determine from this data the moment at which the data points start to increase. This can be done by fitting the straight parts to linear functions and determining the intercept. The result is shown in the following figure:



From this figure we find values of 30.5, 30.6. 30.4 and 30.3 seconds while the underlying mathematics tells us that it should be 30.0 seconds. The absence of any systematic tendency is somewhat counterintuitive. In this simulation we see that noisy data doesn't always have to inhibit the finding of the intercept. It should be noted that only point to point noise is taken into account here. To be clearer the differences between I(t) and I(t+1) and I(t-1). In a real experiment one has to take into account the intrinsic error margin as well. This is the margin that one can determine by repeating the experiment several times and then calculating the average. This is the variation between $I_n(t)$ and $I_m(t)$ with n and m indicating different repeats of the same experiment.

The reason that one still finds the proper intercept is mathematical explainable. The horizontal part can be represented by:

$f(t) = c$

the other part by:
$g(t) = c + at$

For the intercept $f(t) = g(t) \Rightarrow t = 0$

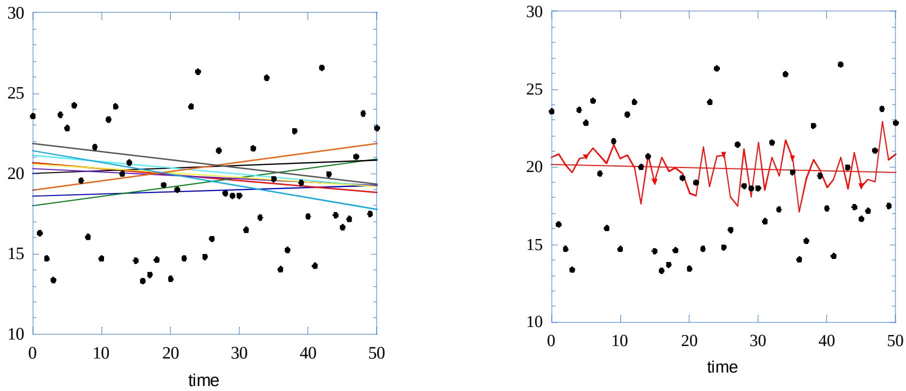
If now a background or noise level Δc is added it reads like:

$$f(t) = c \pm \Delta c_f;\ g(t) = c + a\,t \pm \Delta c_g;$$
$$f(t) = g(t);\ c \pm \Delta c_f = c + a\,t \pm \Delta c_g;$$
$$\Rightarrow t = \frac{\sqrt{2}\Delta c}{a}$$

Assuming the noise levels in both curves to be the same at the moment that the increase in intensity starts we can apply conventional error analysis to this and estimate that the expected error margin is √2 times a single error margin. So variations in the position of the intercept are only due to the higher noise levels which influences that accuracy of the linear fits and thus the position of the intercept. This ultimately leads to a larger error margin but not to a systematic tendency. This is easily verified by numerical simulations.

So variations in the position of the intercept are only due to the higher noise levels which influences that accuracy of the linear fits and thus the position of the intercept. This ultimately leads to a larger error margin but not to a systematic tendency.

This is illustrated in the following graph. In this ten fits are shown to different (repeat-) data sets and from one of the data sets all data points. The data-sets are created by taking a constant value of f(x) = 20 and add to that random noise with as maximum 1.5 x √20



In the right hand panel the same data set is shown. The connected line is the average of the ten data sets and the line the fit to this data set. It will be clear that the variation in the position of an intercept with a sloped line with similar error margins due to a high noise levels can vary considerably. These problems are quite important at low detector countrates. (This is off course something that we all know and generally prefer to forget since it means that for taking accurate data we need either to repeat the experiments several times and take an average or that the data quality should be high).

The method of determination of the exact point where the intensity starts to rise also can have an influence. In the case that one uses the method of two linear intercepts we mathematically get out the correct answer but it might not be compatible with the underlying physics.

It should be noted that when the changes in scattered intensity are less than the experimental statistical noise level any attempt to detect a change is useless.

Detector technology with respect to spatial resolution

The conventional attitude towards the spatial resolution of detectors is that 'the higher the better'. A slightly modified attitude could save quite a bit of work and improve results and ease of data analysis.

In the case that one wants to drill a small hole in a wall to hang a frame for a photograph one uses a small electric drill. In the case that one wants to drill a hole in the ground to look for oil hidden 5 kilometers (3 miles for the Anglosaxons) below the surface of the earth, one uses a drill head 1 meter wide (1.1 yard) equipped with special drill bits. Most people understand this logic but forget it as soon as position sensitive X-ray detectors are being discussed.

Assuming that one is working on a beamline where the size of the X-ray beam in the focal plane of the beamline is approximately 300 microns there is very little to be gained by using a position sensitive detector with a spatial resolution much better than 300 microns. The mathematical theory of this was developed some scores of years ago by a guy called Nyquist (1928) and formalized mathematically by Shannon in 1949 (see reference). Mind you, it does not hurt to oversample your data but your data files will be larger and might give you the impression that your spatial resolution is better than it actually is.

Time resolution

When performing time-resolved experiments there are many experimental aspects that one has to consider. These range from the suitability of the X-ray equipment, like detectors, to the sample environment that one is using. The latter especially deserves attention since in a time resolved experiment one generally varies a physical or chemical parameter in order to study the response of the sample to this perturbation. The crucial aspect here is that this parameter should be varied in a controlled and uniform way.

Before starting a time-resolved experiment it is important to check if it is possible to obtain the information that one needs with the experimental equipment that is available. Generally the real limits can only be determined empirically but by carefully assessing the factors that influence the time resolution it might be possible to see beforehand if there is not a parameter that inhibits the experimenter to perform the experiment at the time-resolution needed in order to render relevant results. These parameters are:

1 - X-ray flux

With synchrotron radiation sources the flux is dependent on the photon energy. The higher the flux the faster the experiment can be performed.

2 - detector efficiency

Different types of detectors have different efficiencies which are also dependent on the used energy. Also one should keep in mind that some types of detectors, like CCD's, add dark current and read-out noise to the signal which influences the dynamic range. A higher efficiency and low noise level will have a positive effect on the achievable time resolution.

3 - maximum countrate a detector can handle

This is dependent on the physical process by which the photons are detected and on the associated electronics. If this value is high more photons can be detected in a single frame and will therefore have a positive effect on the time resolution.

4 - read-out time of the detector

This depends on the number of 'pixels' to be read out and on the physical process by which the detector is read-out. The shorter the better (noise being equal)

5 - attenuation of X-rays by windows around sample and extra background scatter

In some cases this factor can be eliminated by placing both sample and detector in a vacuum chamber. The smaller this number is the faster the experiment can be performed. However, even when X-ray flux is abundant extra background scatter will have a negative impact on the quality of the data.

6 - X-ray contrast

In an X-ray scattering experiment this is the electron density contrast giving rise to the scattering. If the contrast is high one needs less time to obtain a statistical significant result.

7 - statistical accuracy needed

This parameter depends heavily on which parameters one wants to derive from the experiment but in general this is also a human factor. Some people have 'the eye of faith' and others don't. The higher this value is the better the data quality but the slower the time resolution.

8 - available memory to store the different time frames

This is a crucial consideration and might be a limiting factor in applications where it is crucial to use two-dimensional detectors. (Although computer memory is cheap nowadays it is still not trivial to store large data volumes rapidly)

9 – dynamic range

The maximum intensity a detector can detect in comparison with the minimum is called the dynamic range. In scattering experiments where the intensity generally falls of at wider angles as the fourth power of the scattering vector this is not a trivial issue.

All these parameters can be summarized in a symbolic formula as follows:

$$\text{Time resolution} = \frac{\text{read - out} \times \text{statistics} \times \text{attenuation}}{\text{flux} \times \text{detector efficiency} \times \text{maximum countrate} \times \text{contrast}} \times \left|\text{memory}\right| \times \left|\text{dynamic range}\right|$$

The higher the 'value' of the parameter time resolution the longer the length of a time frame and consequently the slower the experiment. This 'value' should, as mentioned in the previous section, be convoluted with the time scale on which it is possible to perturb the sample uniformly. The parameters 'memory' and 'dynamic range' act as a type of delta functions. If these conditions are satisfied the experiment, regarding this aspect, is possible. If this is not the case the experiment is not possible with the equipment under scrutiny. The fastest experiments, using conventional methods and non-cyclic data collection, are at the moment limited to roughly 0.1 sec/frame. When dealing with a process that can be accurately cycled one can play different tricks and the time resolution can be pushed towards the millisecond/frame resolution at the expense of a very complicated experimental protocol.

My personal opinion is that for most X-ray scattering experiments at the moment the achievable time-resolution is not necessarily limited by the X-ray beamlines and detectors but by the speed with which a sample can be perturbed in a controlled and uniform fashion.

Literature

Semyon G. Rabinovich, *Measurement errors and uncertainties*, Springer 1995, ISBN 0-387-98835-1.

C. E. Shannon, Proceedings IRE, **37**, 1949, 10-21.