LS-273

Introduction to the Measurement of Noise with Application to Particle Accelerator Beam Stabilization

Glenn Decker

Advanced Photon Source
Accelerator Systems Division

December 1998

1.0 Introduction

One of the most important figures of merit for a synchrotron radiation source, once specified beam intensity and energy have been achieved, is charged particle beam stability. While a significant effort has been expended at the Advanced Photon Source (APS) to reduce or eliminate undesirable sources of beam motion, it will be necessary to employ active feedback to stabilize the user photon beams to the very stringent levels required. This becomes especially important when one considers that transverse beam stability is generally quoted as a fraction of beam dimensions. Since source brightness tends to be inversely proportional to these transverse dimensions, it should be evident that x-ray beamline users in general will support any and all efforts to reduce the transverse charged particle beam dimensions. The obvious corollary to this is that coincident with emittance reduction efforts must come improvements in our ability to both measure and correct the particle beam trajectory.

Presently, there are at least two active proposals at the APS for reducing both horizontal and vertical emittance. A simple change in lattice functions gives a factor of two reduction in horizontal and vertical beam emittance, while a machine studies program focusing on the correction of horizontal-vertical coupling will allow a reduction of vertical emittance by a factor of 100 or more.

The Advanced Photon Source presently operates with the design natural emittance of 8.2 nm-rad, and without coupling correction, the vertical emittance is approximately 3% of this value. The APS design value for coupling is 10%. Given the design lattice functions, the effect of horizontal-vertical coupling on insertion device source beam size is shown in Table 1.

Table 1. Effect of coupling variation on insertion device source size. The natural emittance assumed is 8.2 nm-rad.

Coupling	10%	3%	1%	0.3%	0.1%
Horizontal rms beam size σ_x (μm)	325	336	340	341	341
Vertical rms beam size σ_y (μm)	86	49	28	15.7	9.1

Presently, the most stringent requirement on rf beam position monitors (RFBPMs) is derivable from the vertical orbit stability specification, namely, that the vertical charged particle beam trajectory must be stable to within 5% of the rms vertical beam size.

At the time the APS beam position monitor system was designed, it was not known that such a specification was even achievable, and even with completely noise-free electronics, other sources of apparent beam motion come into play (e.g., thermo-mechanical effects impacting both the accelerator and the beamlines).

Fortunately, due to a careful and extensive design specification process, the presently installed RFBPM system has been demonstrated to have an AC sensitivity that can resolve vertical orbit motions much smaller than 5% of the "design" vertical beam size, i.e., assuming 10% coupling [1]. The reason for qualifying the above statement as applying only to AC beam motions should be clear to those familiar with the theory of low frequency electronic noise. Long-term drift of electronic circuit characteristics is a very difficult problem. Ultimately, what the x-ray user is concerned about is the uniformity of x-ray flux striking the sample under study; this is the ultimate aim of our orbit correction and feedback efforts.

2.0 Noise Fundamentals

In all of the preceding discussion, emphasis has been placed on measurable quantities having the dimensions of length, namely beam size and orbit motions. References to temporal structure have been qualitative, making use of language such as "AC position sensitivity," and "long-term drift." We need to do a bit better than that if we are to understand what properties of the BPM system are critical for real-time orbit correction with the objective of reducing or eliminating unwanted noise in the beam position.

As an example, Figure 1 shows the ground motion that occurred in APS storage ring sector 19 when a forklift drove into the storage ring infield through the vehicle access tunnel and passed directly underneath the storage ring where a ground motion transducer (a seismometer) was located. There are several things about this data that are brutally obvious to the most casual of observers, but, when we attempt to write down something quantitative, it becomes clear that much care is needed in order to include all of the essential subtleties

.

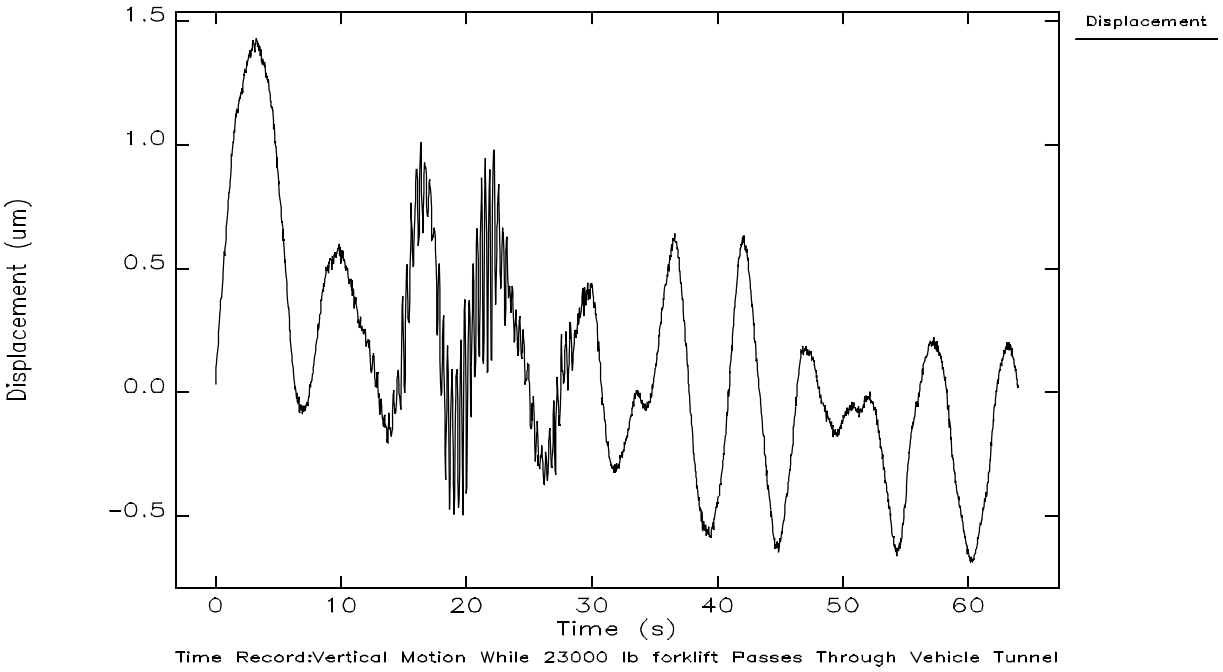

Figure 1. Time series ground motion data collected for the experiment hall floor, sector 19.

For one thing, it is apparent that the forklift spent about 20 out of the 64-second time sample actually traversing the tunnel. During the time it was in the tunnel, a strong and fast vibration occurred. In addition to this, an even longer, slower motion is observed, with about a 1 micron peak-to-peak amplitude, and with a period of about 7 seconds. This longer period motion in fact is not periodic at all, with changes in both amplitude and phase clearly observable in the time sample shown. Finally, and this usually goes without saying but is extremely important in our analysis, the data shown in Figure 1 is incomplete. In fact, it is infinitely incomplete. We have no idea what happened before time t = 0 or after time t = 64 seconds. For all we know, the ground may have been dead quiet at all other times and it is only due to some cosmic conspiracy that the data shows the displayed form. Antithetically, at t = 65 seconds, it may have been that a magnitude 9 earthquake occurred. There is simply no way from looking at the data to disprove either conjecture.

Common sense tells us, however, that the slow, 7-second period motion probably was present before time t = 0 and after t = 64 seconds, and that it probably had about the same average amplitude that it did while we were looking, namely a peak-to-peak amplitude of 1 micron. On

the other hand, it is a good bet that the fast ground motion (actually, pretty close to 3 Hz) probably was not present prior to time t = 0 and probably will occur one more time for about 20 seconds duration sometime after time t = 64 seconds, unless the forklift driver plans to take a lunch break parked somewhere within the storage ring infield. Incidentally, geophysicists are quite familiar with our observed 7-second period motion, which is generally detectable everywhere on earth. It is almost certainly associated with ocean waves lapping languidly up on beaches, up and down the continental coast lines.

In addition to the long time scale information, it should also be apparent that very rapid motions occurring faster than the sampling rate are not represented accurately in the data of Figure 1. Here we have 2048 numbers, measured once every 31.25 milliseconds for a total time of 64 seconds. Clearly, events occurring faster than this, i.e., at high frequency, are not accurately characterized. What exactly is meant by high frequency will soon be made clear.

How then are we to cast the remarks made in the preceding paragraphs into a quantitative framework for the purpose of characterizing our success in achieving the goal of orbit stability? This was alluded to via remarks such as "1 micron peak-to-peak," "period of about 7 seconds," and "pretty close to 3 Hz." In other words, definitions used to answer the questions "How big is it?" and "How fast is it?" are needed.

With regard to a quantitative measure of signal amplitude, one standard measure is that of mean-square amplitude, defined as:

$$\langle x^2 \rangle \;=\; \frac{1}{N} \sum_{n\,=\,0}^{N-1} x_n^2 \,, \tag{1}$$

where x_n represents the sequence of N numbers, for example those plotted in Figure 1. The rms amplitude is just the square root of $\langle x^2 \rangle$, by definition. Implicit in the definition of Eq. (1) might be the assumption that our data is truly AC in nature, specifically, that it has no DC component (please excuse the EE vernacular).

Quantitatively, we are saying that:

$$\langle x \rangle \equiv \frac{1}{N} \sum_{n\,=\,0}^{N\,-\,1} x_n \;=\; 0 \,. \tag{2}$$

While for many very large data sets this might be a very good approximation, it's not good enough in the context of micron-scale beam position measurement work. The quantity $\langle x \rangle$, sometimes written $\bar{x}$, is called the mean value of the signal for a particular data set, in this case the mean displacement. In order to accommodate correctly data having a non-zero mean value $\langle x \rangle$, it is necessary to first introduce a baseline offset in the data prior to performing the sum indicated in Eq. (1). In this way, one arrives at the standard deviation σ_x for a data set, defined according to:

$$\sigma_x^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x_n - \langle x \rangle)^2 \quad . \tag{3}$$

The mean $\langle x \rangle$ and standard deviation σ_x for a data set $\{x_n\}$ are extremely important statistical properties characterizing a data set. It is evident, however, that a large amount of information has been discarded in performing the sums indicated in the definitions for $\langle x \rangle$ and σ_x. Instead of the N values contained in the data set $\{x_n\}$, we have boiled them down to the two numbers $\langle x \rangle$ and σ_x. We have gone some distance in answering the question "How big is it?" at the expense of throwing away any information that could be used to answer the second question, "How fast is it?" To address this second question, we must introduce the concepts of Fourier transforms and power spectral densities [2].

3.0 Analog Fourier Transforms and Power Spectral Density

Making use of the conventions used in Ref. 2, we define the following Fourier transform pair:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{2\pi i f t} dt \tag{4}$$

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{-2\pi i f t} df \quad . \tag{5}$$

Of particular interest here is an object referred to in Ref. 2 as 'total power,' with representations in both time and frequency domains:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \ , \tag{6}$$

which is a statement of what is commonly known as Parseval's theorem. The resemblance between Eq. (6) and Eq. (1) is the basis for defining a quantity known as the power spectral density (PSD), which is to be interpreted as the quantity of mean-squared signal amplitude contained within a frequency range from f to $f + df$. When integrated over all frequencies, the PSD yields the integrated mean-square signal amplitude. The motivation behind using the term 'power' comes from electrical signal spectrum analysis, where the power is proportional to the square of voltage. In the present context of ground or beam motion measurements, the PSD has dimensions of (length)^2 per unit frequency.

For an analog, real function of time $x(t)$, it is the convention in Ref. 2 to take advantage of the symmetry properties of the Fourier transform and make use of only positive frequencies in the definition of the power spectral density $P_h(f)$:

$$P_h(f) \equiv |X(f)|^2 + |X(-f)|^2 = 2|X(f)|^2; 0 \le f < \infty \quad , \tag{7}$$

such that

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{0}^{\infty} P_h(f) df \ . \tag{8}$$

4.0 Discrete Fourier Transforms

If instead of the continuous function *x(t)*, we have in our possession only samples of it at regularly spaced intervals Δ:

$$x_n = x(t_n), \qquad t_n = n\Delta, \qquad n = 0, 1, 2, 3, \ \ldots \ \text{N-1} \ , \tag{9}$$

we can in a completely analogous manner to Eqs. (4) and (5) define a discrete Fourier transform pair (departing slightly from the convention of Ref. 2):

$$X_k = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{2\pi i n k/N}, \tag{10}$$

$$x_n = \sum_{k=0}^{N-1} X_k e^{-2\pi i n k/N}. \tag{11}$$

The correspondence between analog and discrete formulations can be approximated as

$$X(f_k) = X\left(\frac{k}{N\Delta}\right) \approx N\Delta X_k . \tag{12}$$

The discrete form of Parseval's theorem becomes:

$$\frac{1}{N}\sum_{n=0}^{N-1} |x_n|^2 = \sum_{k=0}^{N-1} |X_k|^2, \tag{13}$$

i.e., the sum of the squares of the X_k yields the mean-square value of the sampled data set $\{x_n\}$. Given the symmetry properties of X_k for x_n real, namely that $X_{-k} = X_{N-k}$, and $X_{-k} = X^*_k$, it is customary to display the sequence $\{|X_k|\}$ for positive k only, extending from $k = 0$ to $k = N/2$, corresponding to frequencies extending from 0 up to $f_c = 1/2\,\Delta$, also known as the Nyquist critical frequency.

At this point, to the great misfortune of anyone actually trying to accomplish anything quantitative, the number of methods for displaying discretely sampled data in the frequency domain is vast, and a majority of the literature explaining the conventions, for example the users manual for a spectrum analyzer, tend to be vague and relatively impenetrable. While we were able with some confidence to write down the definitions represented by Eqs. (7) and (8), these formulae in practice are never used except in an abstract manner by theoreticians. Virtually all data these days is collected and analyzed in the form of discrete samples. We are essentially guaranteed that a geophysicist, an accelerator physicist, and an rf engineer all mean something different when they refer to a graph showing "the spectrum" corresponding to a specific data set.

We can be fairly confident, however, that at least for k unequal to zero or $N/2$, when a physicist or engineer shows a plot of amplitude vs. frequency, they will be showing a graph of

something proportional to either the sequence of numbers $\{|X|_k\}$, or the 'power' spectrum $\{|X_k|^2\}$. Once a normalization is found such that the rms value described in Eq. (13) comes out right, we are assured of being able to answer noise/bandwidth questions quantitatively.

Inspection of the definition in Eq. (10) setting $k = 0$ tells us immediately why the case $k = 0$ is generally handled separately; X_0 is nothing other than the mean value of the data set $\{x_n\}$ defined in Eq. (2). Removing the $k = 0$ term from the right-hand sum in Eq. (13) yields an alternate formula for the standard deviation of Eq. (3):

$$\sigma_x^2 = \frac{1}{N}\sum_{n=0}^{N-1}(x_n - \langle x\rangle)^2 = \langle x^2\rangle - \langle x\rangle^2 = \sum_{k=1}^{N-1}|X_k|^2 , \qquad (14)$$

as can be easily verified from the definitions in Eqs. (1), (2), (3), together with Parseval's theorem written as in Eq. (13).

One additional statement that all will agree on is that sample number k of the sequence $\{X_k\}$ corresponds to a frequency:

$$f_k = \frac{k}{N\Delta} = 2f_c\frac{k}{N}, \qquad k = 0, 1, \ldots \frac{N}{2} , \qquad (15)$$

where N is the total number of samples collected in the time series, and Δ is the time in seconds between samples, assuming a uniform sampling rate.

As a specific example of the application of the formulae presented in the preceding sections, Figure 2 shows FFT data computed directly from the data of Figure 1, using the routine sddsfft authored by M. Borland [3]. While 2048 data points ($n = 0$ through $n = N = 2047$) are displayed in Figure 1, there are only 1025 data points $\left(\frac{N}{2}+1\right)$ shown in Figure 2. Roughly half of the information has been lost, because only the magnitude of the complex numbers X_k are displayed. Phase information is not shown. The $k = 0$ element $|X_0|$ indeed turns out to be exactly the mean value of the data set $\langle x\rangle$, 0.125 mm in this case. The normalization of the $k\neq 0$ data in Figure 2 is set such that the sum of squares from $k = 1$ to $k = \frac{N}{2} = 1024$ (i.e., not including the DC offset) gives the standard deviation according to Eq. (12)*. Apparently, for $1 \le k \le \frac{N}{2}$, what

* It actually gives a value for σ_x^2 equal to N/(N-1) times Eq. (12), which is a very minor effect, relevant only for statistics of small data sets.

has been displayed in Figure 2, is the quantity

$$\sqrt{|X_k|^2 + |X_{-k}|^2} \;=\; \sqrt{2}|X_k| \;. \tag{16}$$

For $k = \frac{N}{2}$, the value shown is simply $|X_{N/2}|$ and not $\sqrt{2}$ times this value, as for the other k values; a consequence of avoiding the double-counting of the $k = N/2$ term when evaluating the sum in Eq. (13). We are now in a position, finally, to quantitatively answer the question posed earlier, namely, for the data set of Figure 1, or for any set of time series data, "How fast is it already?"

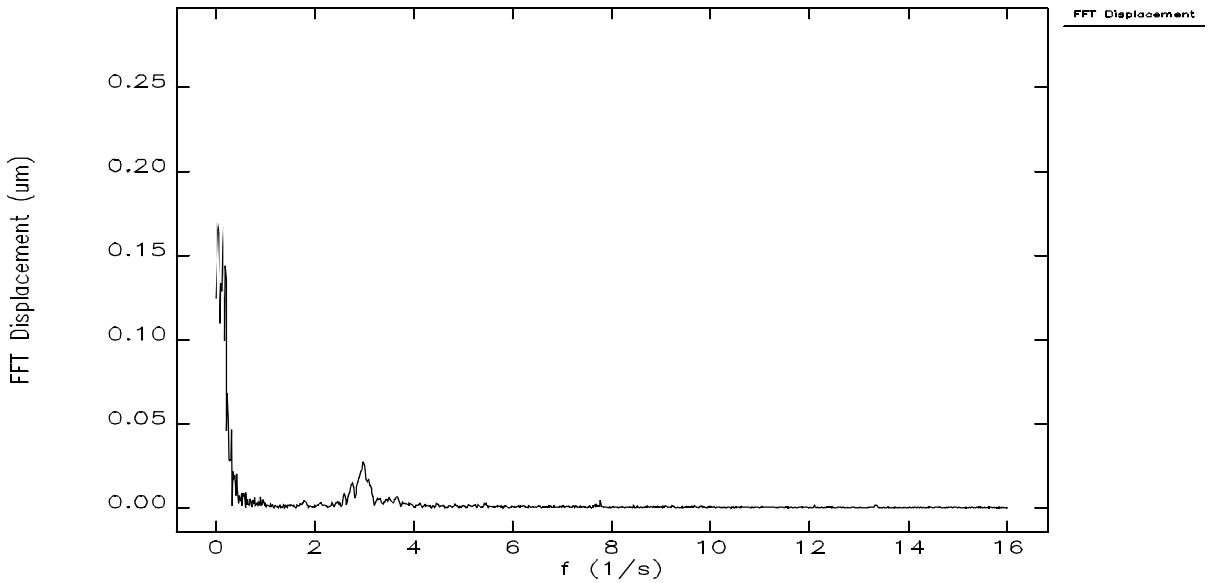

Figure 2. FFT of the data from Fig. 1 using program sddsfft.

The sum of the squares of the values shown in Figure 2, over a span of k values (corresponding to a span of frequency values according to Eq. (15)), yields a number that can be interpreted as being the mean square amount of signal noise contained within that frequency span. Thus, when an rf engineer, for example, refers to the "noise power in a 10 MHz frequency band," what he is referring to, whether he knows it or not, is the partial sum of k values in Eq. (13) corresponding to the frequency range of interest. He'll probably take the log of the answer he gets, expressed relative to 1 mW of power dissipated in a 50-Ohm load, and quote the value in dBm. So long as the meaning behind Eq. (13) is remembered, there should never be a problem with physicists and engineers communicating.

With regard to the information lost during the sampling process in Figure 1, we can say that it lies either between frequency 0 and $\delta f = \frac{1}{N\Delta}$ (i.e., long-term drift), or above the Nyquist frequency $f_c = \frac{1}{2\Delta}$, since there is no way to quantify changes occurring between time samples spaced Δ apart. Long-term drift can be addressed by plotting the DC term X_0 for successive (or overlapping) data sets as a function of time. For the high frequency data, the only practical approach, if this information is of real interest, is to sample faster, making Δ smaller.

As one final illustrative example, Figure 3 shows the same data as for Figure 2, now squared and plotted on a logarithmic scale. This format is quite common on rf spectrum analyzers, while low frequency (< 10's of kHz) signal analyzers tend to display results as rms or peak amplitude vs. frequency, or as power spectral densities, with built in 'band power' calculations involving cursors, etc.

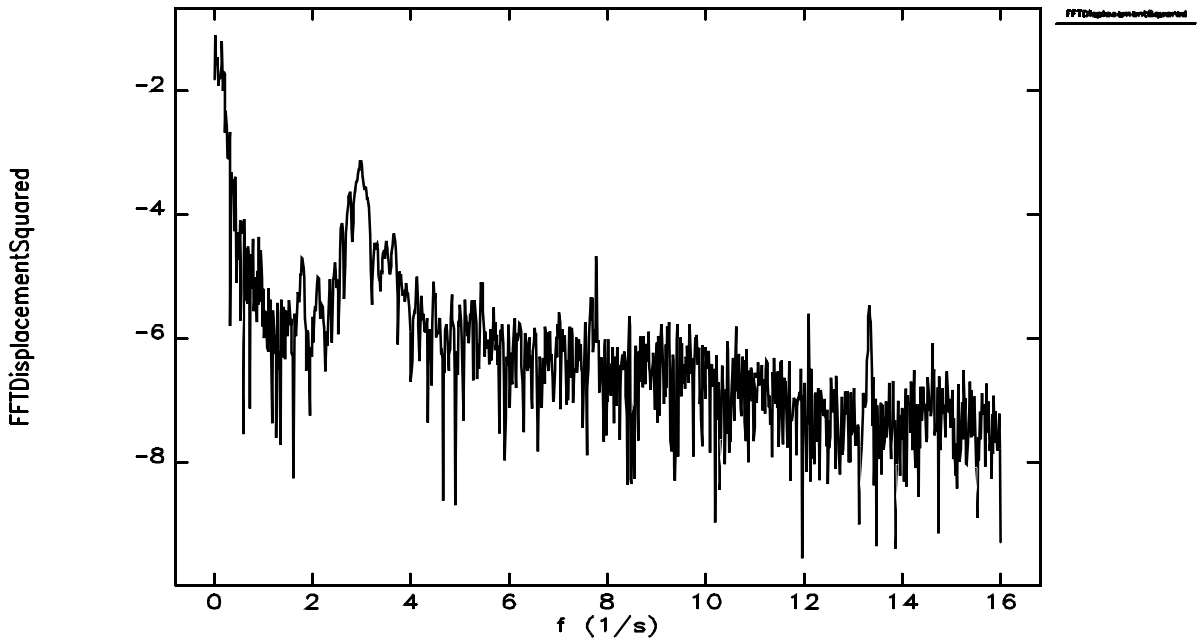

Figure 3. The square of the data of Fig. 2, plotted on a $\log_{10}$ scale.

One final comment about frequency domain data is that if one were to plot the square of the data in Figure 2 as a function of frequency on a linear scale, the area under the 'curve' will be equal to the total mean-square signal amplitude multiplied by the Nyquist frequency $\frac{1}{2\Delta}$. Thus, unless you know what the sampling rate was for the time series data used to generate a plot of

'amplitude' vs. frequency, the overall scale factor to get back quantitative rms or noise power values will be indeterminate. The best solution is if the FFT box you are using has built-in noise power calculations; however, we are often in the situation of having a computer file of signal vs. time or vs. frequency. It is hoped that the discussion in this section will aid in the analysis of signals which have a significant random (noise) component.

References

[1] Y. Chung, E. Kahana, "Performance at the Beam Position Monitor for the Advanced Photon Source," Rev. Sci. Instrum. 66 (9), September 1995.

[2] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, "Numerical Recipies in C," New York, Cambridge University Press, pp. 496-504 and pp. 549-553, 1992.

[3] M. Borland, L. Emery, "The Self-Describing Data Sets File Protocol and Program Toolkit," Proceedings of the International Conference on Accelerator and Large Experimental Physics Control Systems ICALEPCS '95, pp. 653-662, January 1997. Also online at URL http://www.aps.anl.gov/icalepcs95/icaleps95.html