View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by CERN Document Server

Real-time control of beam parameters

M. Dehler
Paul Scherrer Institut, Villigen, Switzerland

Abstract

This article gives an overview of the theory and application of real-time control of accelerator beams. The design and structure of orbit feedbacks are described, going from basic local feedbacks to modern state-of-the art global systems. The time domain behaviour is analysed for the building blocks of the systems as well as from the spectrum of random sources driving the orbit perturbations. The use of predictive filtering is shown for the design of the control algorithm. A second important class is the control of tunes and chromaticities. Advanced tune measurements are performed using a digital phase-locked loop. The feedback systems are typically hybrid, simultaneously working on tune and coupling and chromaticity. Adaptive feed-forward algorithms are shown to be a suitable approach for use in energy ramping. For application in a high-speed bunch-by-bunch feedback system, efficient low-noise data processing is presented for a digital filter. Also here, predictive filtering is shown to give well-adapted high-order filters.

1 Introduction

One of the countless corollaries to Murphy's Law states that everything able to go out of adjustment will do so with the utmost enthusiasm. As a consequence, today's accelerator facilities employ a myriad of controllers, from minor ones stabilizing individual components and subsystems, to the ones described here and directly affect the beam behaviour.

These systems are best classified according to the parameter in the phase-space distribution which they are supposed to control. This phase space describes the particle motion in the accelerator in six dimensions: the three spatial coordinates and the corresponding momenta. The most important are mean values (position, energy and flight angle) and standard deviations (size and energy/momentum spread) as a function of time and position inside the machine.

The longitudinal position and momentum (as well as the frequency of the longitudinal synchrotron oscillation in a ring machine) are mainly influenced by the amplitude and phase of the main RF. These are controlled by the regulation loop of the RF system, which is not covered here. Exceptions are fast instabilities caused, for example, by higher order modes in the system, which are typically controlled by bunch-by-bunch feedbacks.

Transversally, we observe drifts in the beam orbit due to ground motion and mechanical drifts. Other than the longitudinal case, sources of perturbations as well as required corrections are distributed over the whole machine, so that feedbacks and feed-forwards need multiple sensors and multiple corrections. As in the longitudinal case, there are fast transverse oscillations due to wake fields etc., which are controlled via dedicated bunch-by-bunch feedback systems.

Then there are the higher order properties of the beam such as the transverse tunes relating to the focusing strength of the magnet optics in the respective planes and the chromaticity of the optics, which describes the dependency between focusing strength and particle energy. The tunes, the coupling between the horizontal and vertical tunes, and the chromaticity are intrinsically related, so that typically hybrid stabilization schemes are employed.

1.1 Feedback or feed-forward

Implicitly, all controllers make use of pre-knowledge about the system and the excitations. The way this is done defines two general classes, feedbacks and feed-forward systems.

The main task of a feedback is to control random, non-predictable fluctuations in the accelerator parameters, which may be due to internal or external noise. (Notable exceptions are bunch-by-bunch feedbacks. Here the system is inherently unstable and the goal of the feedback is to change an unstable behaviour into a stable one.)

But there are also systematic changes like ramping the energy of an accelerator. While ramping the beam energy, the magnets controlling orbit and focusing have to be adjusted correspondingly, something which could in principle be done by a feedback system. But, since the parameters for the ramping curve are perfectly known in advance, a feed-forward system is the more suitable approach, where the corresponding set of magnet currents are precomputed and set directly.

A pure feed-forward system blindly sets values derived from a machine model and will not look at the quality of the correction. So the usual approach is to use an adaptive feed-forward, which combines the classical feed-forward with a feedback. The correction values from the feedback are used to update and optimize those used by the feed-forward system.

2 Controlling the orbit

One of the fundamental beam properties to control is deviations from the reference orbit in the accelerator. The measurement is done using beam-position monitors (BPMs), in which the passing beam induces electromagnetic fields into sets of electromagnetic electrodes in the shape of buttons or striplines. The beam position is proportional to the difference in signal strength coming from the electrodes. As an alternative, the synchrotron light radiated by the beam as it is deflected by the magnets can also serve for the position measurement. As the light strikes the electrodes of X-ray BPMs, it stimulates the emission of photoelectrons. Comparing the photocurrent from different electrodes again results in a position measurement. Corrector magnets—typically dipole magnets of low strength—give a transverse kick to the beam as it passes the corrector and thus adjust the orbit.

2.1 Local control

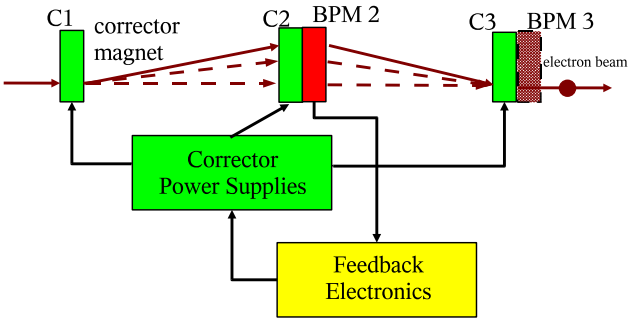

Fig. 1: Simple local orbit feedback

A model of the most basic local orbit feedback is shown in Fig. 1. It contains three corrector magnets C1 to C3 spaced at a certain distance and a BPM positioned directly at corrector C2. If we need to change the position at the BPM, we apply a kick to the beam at the corrector C1. Corrector C2 will be driven so that the overall change in beam offset at C2 is zero and corrector C3 will compensate for any change in the flight angle. The magnet triplet C1 to C3 is called a magnet bump; it will generate a local excursion in the beam trajectory, which (in first order) is transparent outside the bump.

A second triplet centred around BPM3 using correctors C2, C3 and a following corrector would help control the position there. Interlacing this bump with the one centred around BPM2 gives full control of beam position and angle between BPMs 2 and 3. Whereas a local feedback using a triplet is a classical setup around a collision experiment, a four-magnet bump is fairly common in synchrotron light sources, since the beam angle defines the direction of the synchrotron light generated, an important parameter for the machine.

It should be clear though that even with perfect magnets, local orbit bumps and feedbacks are local only in a first-order sense. Setting up a bump, we change the path length of the trajectory. As the beam has a different time of flight along the orbit, the synchronization with respect to the accelerating RF changes results in a different beam energy, which via dispersion leads to global orbit changes.

2.2 Going global

In principle, a global orbit feedback can be set up by interlacing multiple local feedbacks, such that we cover the whole accelerator. In practice, we do not have total freedom in placing the components of the feedback. The BPMs may be in different locations than the correctors, and in addition, the number of BPMs may be higher or lower than that of the correctors.

To characterize the relationship between correctors and BPMs, a response measurement is used. The machine is set to the reference orbit, we drive one of the correctors c_i and measure the resultant orbit excursion at all BPMs x_j inside the ring. If we restrict ourselves to small amplitudes, magnet non-linearities and hysteresis effects will be negligible, so that we end up with a linear relationship. Measuring offsets at all k BPMs using all n correctors, gives us the $n \times k$ size response matrix M:

$$\vec{x} = M\vec{c}\,. \tag{1}$$

To correct the orbit, we need the inverse mapping, giving us corrector settings $\vec{c}$ for a given orbit distortion $\vec{x}$. Trying to invert M will fail in most cases. The matrix may not be quadratic and, even if it is quadratic and invertible, it may be badly conditioned, meaning that even tiny errors in measuring beam offsets lead to huge errors in the computed corrector settings. Effectively, we need a suitable approximation of M that we can invert without encountering the problems discussed above.

The appropriate way out of this dilemma is given by the singular value decomposition (SVD) algorithm [1]. The $n \times p$ matrix M is decomposed into a product of three matrices

$$M = USV^T\,, \tag{2}$$

where U and V are unitary $n \times n$ and $p \times p$ matrices:

$$U^TU = I\,, \tag{3}$$
$$V^TV = I\,. \tag{4}$$

The vectors u_i and v_i constituting U and V are called the left and right singular vectors and, together with the singular values σ_i, define a relationship similar to eigenvalues and eigenvectors:

$$\begin{aligned} Mv_i &= \sigma_i u_i \\ M^Tu_i &= \sigma_i v_i\,. \end{aligned}$$

As can be shown from these relations, the singular values define the diagonal matrix S:

$$s_{ii} = \sigma_i, i = 1 \ldots n \tag{5}$$
$$s_{ij} = 0, \text{else}\,. \tag{6}$$

The most simple approach computes the decomposition as follows. The left singular vectors v_i are given by the eigenvectors with non-zero eigenvalue of the matrix product MM^T. In the same way, the right

singular u_i are eigenvectors with non-zero eigenvalue of the matrix product $M^T M$. The singular values themselves are given by the square roots of the non-zero eigenvalues of either MM^T or $M^T M$—they are equal. (For a tutorial on SVD see Ref. [2].)

Given the components of the decomposition, we can define a pseudoinverse mapping as

$$\tilde{M}^{-1} = \sum v_i \sigma_i^{-1} u_i, i = 1 \ldots n \,, \tag{7}$$

which will be a least-square approximation of the inverse of M. The smallest singular values σ_i correspond to states which are very sensitive to noise and errors in the input. Therefore one typically restricts the calculation of the pseudoinverse even further by using only the first $r < n$ largest singular values, so that one ends up with a well-conditioned mapping.

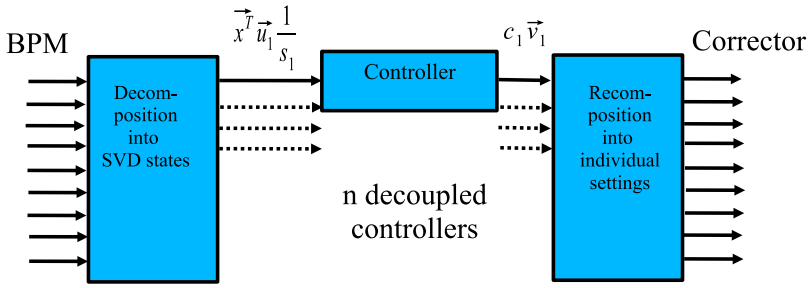

Fig. 2: Conversion of a MIMO feedback system into multiple SISO feedbacks using the singular value decomposition

Apart from computing a pseudoinverse, the approach has a second extremely useful feature. Let us assume we change the corrector magnet settings to values corresponding to one of the right singular vectors v_i. By definition, we see a change in BPM readings, which is proportional to the corresponding left singular vector u_i and nothing else. Thus we obtain a set of decoupled states of the machine, for each of which we can define a single input single output (SISO) controller, as shown in Fig. 2. The data coming from the BPMs is decomposed into decoupled states using the matrix U; we multiply the state variables with the scalar SVD inverse σ_i^{-1} and feed the result into the scalar controller to compute correction factors. Multiplying these with the right singular vectors gives us the actual machine settings for the correctors. Instead of having a large-dimension multiple-input multiple-output (MIMO) system, we deal with a (large) number of scalar SISO systems, which are far easier to manage in terms of analysis and design.

In doing the decomposition, we implicitly assumed purely linear components in the system. For the corrector magnets, this assumption may not hold. For large amplitudes, a magnet with a yoke may show hysteresis effects corresponding to a variable gain. Superconducting magnets exhibit decay and snap back effects due to persistent currents in the magnet. If these effects are encountered, additional local control loops measuring and stabilizing the field strength should be used to compensate, so that the global orbit feedback sees more or less linear devices. For a discussion of the problem for the Large Hadron Collider superconducting correctors, see, for example, Ref. [3].

Finally, there is one important thing to keep in mind in applying this decomposition. Especially for a MIMO system, it is extremely important to avoid saturations in selected parts of the system such as ADCs, DACs, amplifiers or corrector power supplies. In a system with a single input and output, a saturating element will simply change the effective gain of the whole system. In a MIMO system, however, only parts of the response matrix saturate, so the singular value decomposition used by the controller is no longer valid and system behaviour changes in an unpredictable manner.

2.3 System topology

Given the large number of input and output parameters, data flux and processing seem to be rather complex. All the BPM data in the accelerator has to be collected to do the decomposition into the various machine states. The same thing looks true for the corrections. At first glance, this seems to favour a centralized processing architecture where one big central node is responsible for receiving, processing, and sending out data.

But looking at the local feedback using a three-corrector bump as described before, we can observe something interesting. If we forget about higher order coupling due to path length changes, this feedback already represents one of the decoupled states of the machine. Further, as has been mentioned already, we could set up a global feedback by interleaving bumps using the local feedbacks on the bump. The local feedbacks would correspond to the SISO controllers working on the decoupled SVD states of the machine, the corresponding hardware architecture would be that of a distributed system requiring only a localized data exchange.

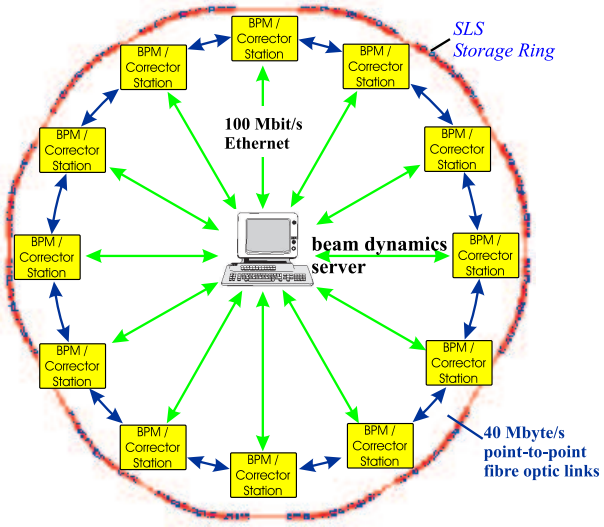

Fig. 3: Topology of orbit feedback system (SLS)

In general, the distribution of BPMs and correctors will differ from the idealized case discussed here, but we will still have a certain localization of right and left singular vectors relating to the fact that to correct an orbit deviation in a selected location of the ring, we have to adjust only correctors in the vicinity of that location and not over the whole machine. An appropriate architecture for a global system will most likely look like that in Fig. 3, showing the layout of the SLS feedback [4]. We have twelve BPM/corrector stations distributed along the ring, each being connected to six BPMs and correctors. Fast fibre optic links connect adjacent stations allowing the real-time exchange of data. For the computation of each corrector, the data from up to 18 BPMs in the vicinity can be processed. A central server connected to the stations via Ethernet is responsible for the distribution of the precomputed SVD coefficients and manages the data transfer to and from the control system.

2.4 Dynamics

What is the time and frequency domain behaviour of the overall system? The first and most important component, the beam, sees changes in the corrector fields instantaneously and propagates the information about corrector kicks as it passes through the accelerator.

The next component to consider is the position measurement and the controller itself. The hardware layout of a BPM/corrector station is shown in Fig. 4. An analog RF front-end converts the 500 MHz

raw signals coming from button BPMs down to an intermediate frequency, which is sampled digitally at a high sample rate of 33 MS/s. The digital down converter filters and down-samples the signal to a 2 kHz, 4 kS/s data stream, from which the first DSP calculates the position data. The second DSP performs the SVD calculations and contains the controller algorithms. The computed kicks are sent via fibre optic links to the digital power supply controllers steering the corrector magnets.

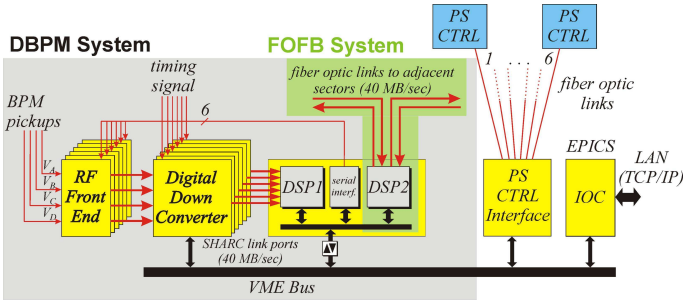

Fig. 4: Hardware layout of a SLS BPM/corrector station

The influence of these components on the frequency domain behaviour of the control loop is relatively simple. The principal effects are latencies due to data transfer, which in the case of the SLS amount to 1.5 milliseconds. Both digital down converters and corrector power supplies each have a 2 kHz bandwidth, which is large compared to the low pass behaviour due to eddy currents in the magnet core and the vacuum chamber wall, described in the following.

The magnetic flux in the corrector magnet builds up in two steps. First the field induced by the coils propagates into the spaces between the metal sheets of the lamination, a process which happens within picoseconds. In the second step the magnetic field enters the sheet material itself. Here the bulk of attenuation and delay takes place due to eddy currents.

The effect can be calculated analytically with good accuracy. The frequency dependency of the flux comes out to be

$$\frac{\psi(\omega)}{\psi_{DC}} = \frac{2}{kd}\tan(kd/2) \tag{8}$$

$$k^2 = \omega^2\epsilon\mu - j\omega\mu\kappa\,. \tag{9}$$

An example [5] of the resultant frequency behaviour is shown in Fig. 5.

A second, similar effect are eddy current losses in the wall of the vacuum chamber. Other than the effects in the lamination of the corrector, the eddy currents also have an effect on the distribution of the magnetic field, so this needs to be computed numerically. As an example, Fig. 6 shows the loss density in the wall together with the resulting frequency dependency. These two effects, eddy current losses in the corrector magnets and the chamber wall, are the main determinants of the feedback performance.

2.5 Sources of orbit drift

Before talking about suitable controllers, we first need to know what effects the system is supposed to follow or correct.

There are plenty of sources of orbit drift and jitter such as, for example, machinery or human traffic, but the fundamental and unavoidable contribution is due to random drift of the ground. Tectonic drifts, settling effects, and other natural effects lead to a stochastic ground motion equivalent to a random walk or Brownian motion. If we look at two given points on the ground and try to predict the evolution of their distance $L(t)$ over time, we can make only statistical statements. There is no preferred direction, so the expected distance $E(L(t))$ will stay constant. But as time goes by, we can say less and less what

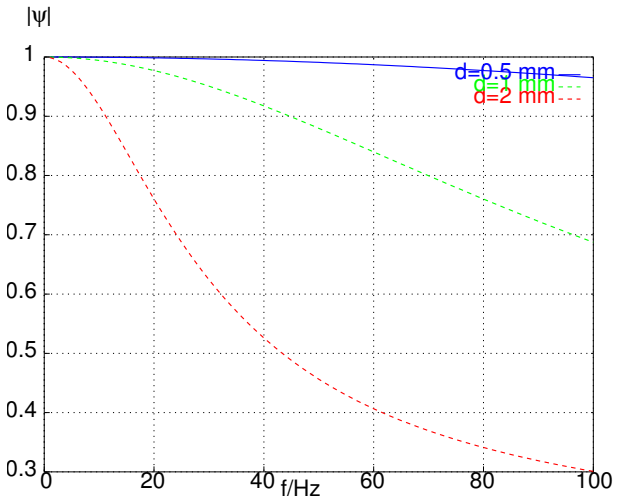

Fig. 5: Frequency dependency of the magnetic flux generated by a corrector magnet with laminated yoke. Thickness of lamination d = 0.5, 1, 2 mm. Steel type EBG 1200-100A.

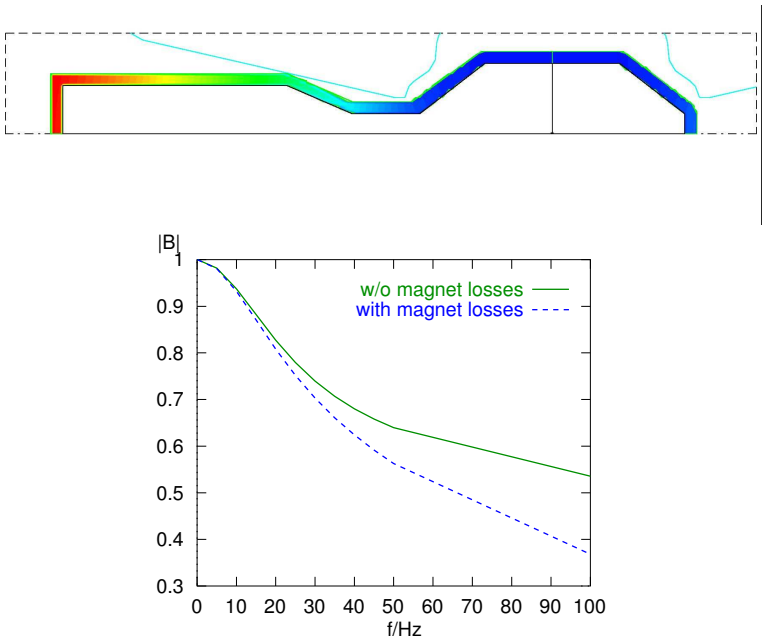

Fig. 6: Upper part: Eddy current loss density on the vacuum chamber walls at 50 Hz. Lower: Resulting frequency dependency with and without losses in the corrector magnets.

exactly it is going to be—the variance $E((L(t) - L(0))^2)$ will increase, in this case linear with time. This is described by the ATL law [6]:

$$E(L(t)) = L(0)\,. \tag{10}$$

$$E\big((L(t) - L(0))^2\big) = AtL\,. \tag{11}$$

The drift constant A has been measured in various parts of the Earth; a relatively constant value of approximately

$$A = 10^{-5}\frac{\mu m^2}{sm} \tag{12}$$

has been obtained. The corresponding noise power density comes out to be

$$S(j\omega) = \frac{A}{j\omega} \,. \tag{13}$$

The beam itself sees these shifts only indirectly via changes in the magnetic fields, as the magnets drift. As the field of quadrupole magnets varies linearly with position, it is the quadrupole motion which has the strongest effect on the beam. Thus the beam trajectory will fluctuate proportionally to the ground noise modulated by the mechanical transmission from ground to the magnets. The corresponding transmission function may add frequency peaks due to mechanical resonances in the girder and will change the spectrum considerably, especially for the higher frequencies. A measurement of the ground spectrum together with the corresponding quadrupole spectrum is shown in Fig. 7.

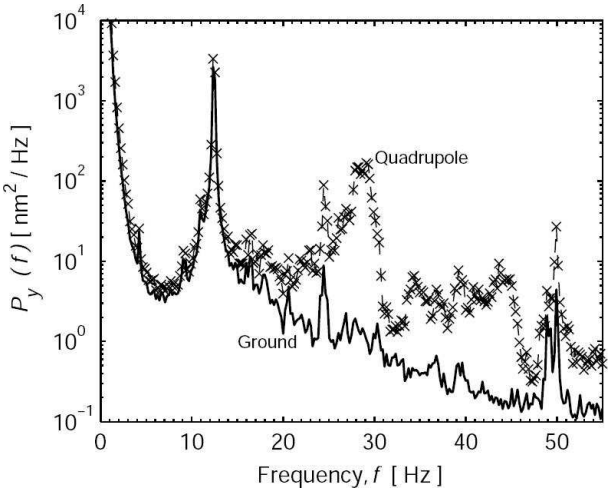

Fig. 7: Power spectrum of mechanical motion at the SLS measured on tunnel floor and quadrupole magnets [7]

A typical power spectrum of the beam motion is shown in Fig. 8, along with the corresponding curve for a running feedback using a PID controller. These have been computed from two-second samples of the beam motion, so the DC peak due to the random walk drift is not very visible. Most prominent is the line at 50 Hz, which shows the beam motion not from mechanical drifts but due to stray fields of the 50 Hz power mains. Next we see a few peaks between 10 and 30 Hz, which correspond to mechanical resonances of the girder structures. At 3 Hz there is still a tiny peak in the spectrum, which is also due to stray fields. In this case, these originate from magnets in the SLS booster, which is ramping in a 3 Hz cycle.

2.6 Proceeding further

The classic textbook example of a controller is a PID controller containing a proportional part, an integrator, and a differentiator. We determine only three parameters and, as the results in Fig. 8 show, good performance can already be obtained.

Nonetheless, it is interesting to consider what a truly optimum controller design could look like for a given machine, a given noise spectrum exciting the beam and—not to be forgotten—the internal noise sources present in the measurement, controller and corrector hardware.

This looks like a nice task for a specialized filter. The problem is that the filter is part of a feedback system, so its characteristic acts back on the feedback loop. Taking this into account leads to an optimization process involving the solution of a differential equation of the Riccati type. For purely Gaussian, white, and mutually uncorrelated noise sources, the solution is given by a linear quadratic regulator LQR [8].

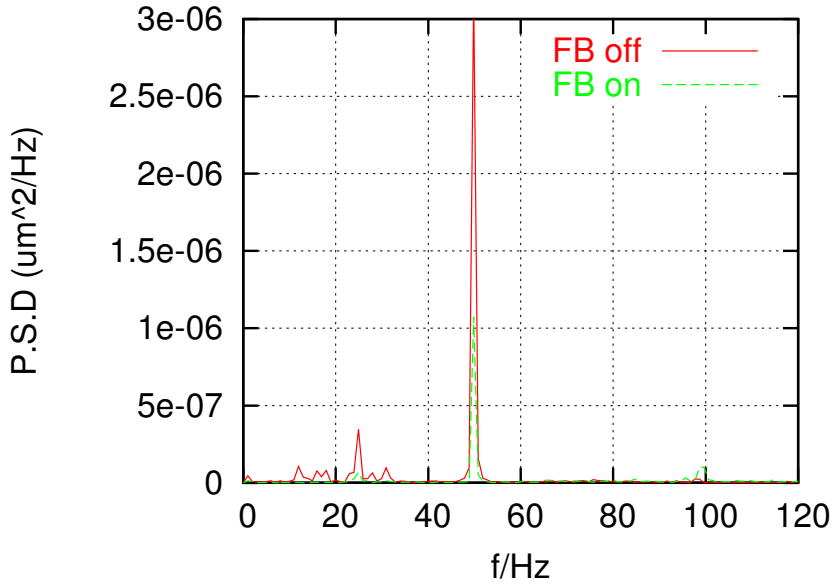

Fig. 8: Power spectrum of beam motion at the SLS with and without feedback using a PID controller

In the case of orbit correction schemes, these assumptions do not hold and we would have to find an adapted solution to the equivalent Riccati equation, and synthesize an equivalent FIR and/or IIR filter. Here, we restrict ourselves to a more approximative approach. We look for an appropriate stochastic filter fitting conditions which are explained in the following, neglecting the effects due to the feedback loop. Closing the loop and optimizing performance is then done using a classical (e.g. PID) controller. (For an in-depth overview on the topic see, for example, Ref. [9].)

2.6.1 A short introduction to predictive filtering

To keep things simple, we will assume linear, weak, stationary systems having time invariant means and variances.

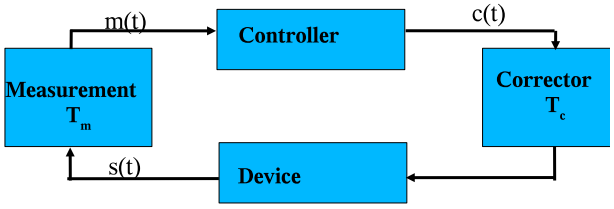

Fig. 9: Basic feedback loop

Given the basic feedback loop in Fig. 9, let us discuss what a perfect controller should look like. Any correction applied by the controller will affect the machine state only after a certain delay T_c. Furthermore, measuring the state of the device takes an additional delay (and causes bandwidth limitations and noise); the controller acts on old and only partially reliable information and needs to correct a future state of the machine. So we would like to have the best guess or estimate of the state of the machine at a future time $s(t + T_c)$, as computed from the currently available measurement values $m(t), t < 0$:

$$\tilde{s}(t + T_c) = E\{s(t)|m(\tau), \tau < t\} \; . \tag{14}$$

This describes an estimator filter

$$\tilde{s}(t) = \int m(\tau)h(t-\tau)d\tau$$

with a yet to be determined pulse response $h(t)$. Requiring the estimation to minimize the mean square error

$$P = E\big((s(t) - \tilde{s}(t))^2\big)$$

leads to the defining equation of a Wiener filter [10]

$$R_{sm}(t,\zeta) = \int_{T_1}^{T_2} h(\tau)R_{mm}(\tau,\zeta)d\tau \ , \tag{15}$$

where R_{mm} and R_{sm} are the auto and cross correlation functions of $m(t)$ and $s(t)$. The integration limits $T_1 < T_2 < t$ choose the time span of data we are extrapolating from[1].

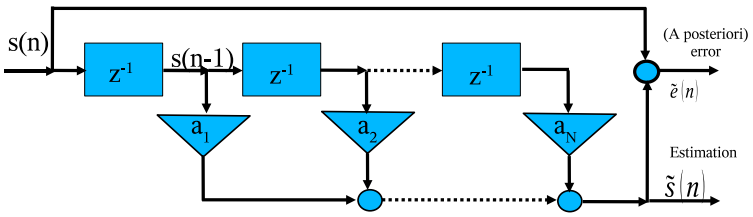

Fig. 10: A FIR predictor filter

In the discrete world, the estimator corresponds to a FIR filter as shown in Fig. 10. Minimizing the estimation error in terms of the unknown filter coefficients a_k leads to N equations of the form

$$E\{\big(s(n) - \sum_{k=1}^{N} a_k s(n-k)\big)s(n-m)\} = 0, 1 \le m \le N \tag{16}$$

or in terms of the auto correlation $R(n)$:

$$R(m) - \sum_{k=1}^{N} a_k R(m-k) = 0, 1 \le m \le N \ . \tag{17}$$

These are called the Yule–Walker equations. The corresponding matrix is a Toeplitz matrix, the system can be solved by Levinson's recursion [11].

Given the coefficients a_k, we create the filter of Fig. 10 and, subtracting estimate and measurement, we can compute the a posteriori error $\tilde{e}(n)$. An ideal filter will extract only the predictable parts of the signal, so the error will not be zero but white noise with autocorrelation

$$R_{\tilde{e}\tilde{e}}(k) = E\delta(k)$$

and power E. This property is also valid in reverse. If the signal to be estimated is white noise, then the corresponding Wiener filter will be zero in all coefficients. Signal levels for white noise can not be predicted by definition, so this certainly makes sense.

This argument can be extended a little bit further. If the signal can be modelled as an IIR filter as in Fig. 11, driven by white noise

$$s(m) = \sum_{k=1}^{N} a_k s(m-k) + n(m) \ ,$$

[1] Using, for example, $T_1 < t < T_2$ will interpolate and $t < T_1 < T_2$ will give a best estimate of past values in terms of future ones.

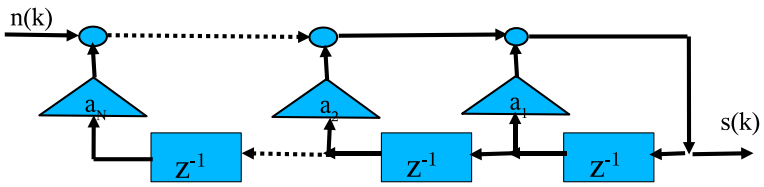

Fig. 11: IIR filter

the inverse filter, which is an FIR type, will convert $s(m)$ back into white noise and so the corresponding Wiener filter gets the coefficients

$$h_k = a_k \ .$$

As we have seen, the autocorrelation of the *a posteriori* error $\tilde{e}(n)$ contains information about the quality of the estimator. From the estimation process itself, we can also compute an a priori error along with its autocorrelation. If we have a non-stationary process, which means the stochastic properties of the sources and devices involved vary with time, we can monitor both error signals and continuously update the filter coefficients. The resultant adaptive system is called a Kalman filter [12].

2.6.2 Application to typical spectra

The most simple model of orbit motion is given by the ATL law describing a random walk. In this case, the orbit deviation is simply an IIR integrator being driven by white noise:

$$s(k) = s(k-1) + An(k) \ .$$

Assuming a perfect measurement without additional noise, we can apply the result from the last section and directly write down the estimator as

$$\tilde{s}(k) = s(k-1) \ .$$

Effectively, we take the value of the last measurement as the best guess of the current state.

In the next step, we take the same model for the orbit drift, but now we are more realistic in adding some uncorrelated white noise to the measurement $m(k)$.

$$\begin{aligned} s(k) &= s(k-1) + An_1(k) \ . \\ m(k) &= s(k) + Bn_2(k) \ . \end{aligned}$$

In this case, measurement and signal differ. Minimizing the least-squares error between signal and estimates leads us to the following defining equations:

$$\begin{aligned} \sum_{l=1}^{N} a_l R_{mm}(l,k) &= R_s m(k); k = 1, N \\ R_{mm}(k) &= R_{ss}(k) + B^2\delta(k) \\ R_{sm}(k) &= R_{ss}(k) \end{aligned}$$

Since the noise sources n_1, n_2 are uncorrelated, the correlation matrices R_{mm} and R_{sm} can be derived easily from those of $s(k)$. The result is shown in Fig. 12. Before, we had an all pass filter with a single tap pulse response, now we obtain a low pass filter giving us a high response at frequencies where the

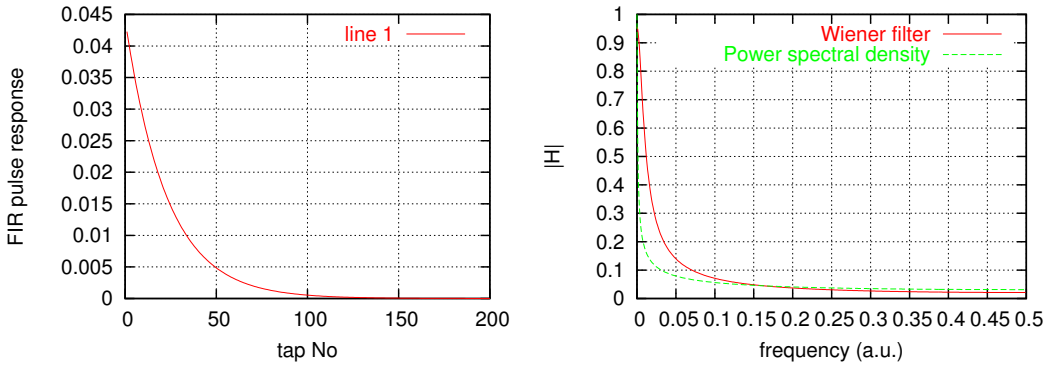

Fig. 12: Predictor filter assuming basic random walk orbit drift and white noise in measurement. Left: pulse response; right: frequency domain behaviour.

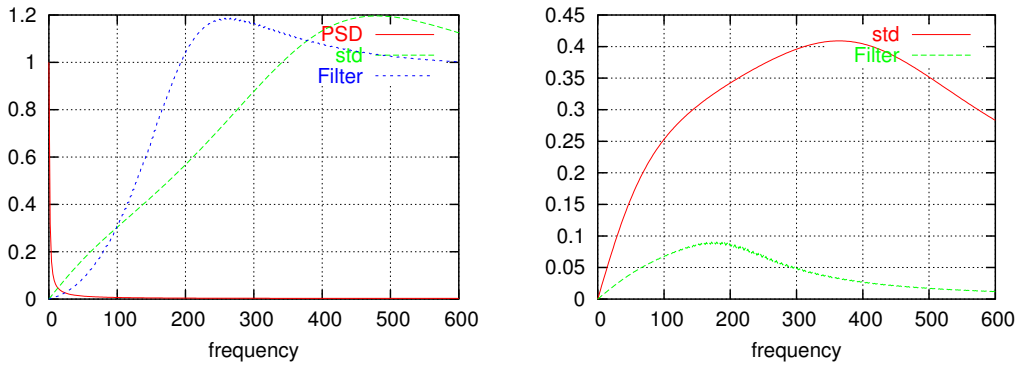

Fig. 13: Closed loop performance with and without basic predictor filters, as shown in Fig. 12. Left: suppression of orbit jitters caused by ground motion, right: additional orbit jitter excited by measurement noise.

signal is high compared to the noise. This does not come for free, as one sees in the spread-out pulse response and increased group delay.

What is the closed loop performance of a feedback using this predictor compared to just taking the measurement value? Figure 13 shows the relevant graphs assuming a PI controller. The prolonged latency of the filter will worsen the effectiveness of the controller in suppressing the effect of ground motion as can be seen in the left graph. The big advantage of the estimator is visible on the right graph which shows how noise in the measurement leads to additional orbit fluctuations in both cases. Orbit noise contains contributions from both sources; the estimator will provide a good compromise in limiting both.

As a second generic example, Fig. 14 shows a predictive filter assuming a random walk, where the ground spectrum is amplified by a mechanical resonance, for example in the girder. Varying the strength of the uncorrelated noise in the measurement, we see a similar effect as before. As we increase the measurement noise, the filter converges to a narrow band behaviour, limiting the influence of the noise, but at a high cost in terms of latency.

One observation should be made here with respect to a real world optimization of the closed loop performance. The objective is to minimize jitters in the estimation $\tilde{s}(k)$ and not in the measurement signal $m(k)$. This approach has some inherent danger. Depending on whether the resonance in Fig. 14 is really a girder resonance or, for example, a measurement artefact (and consequently, how we determine the filter coefficients), the predictor will enhance or suppress it, so that the result may be well off the

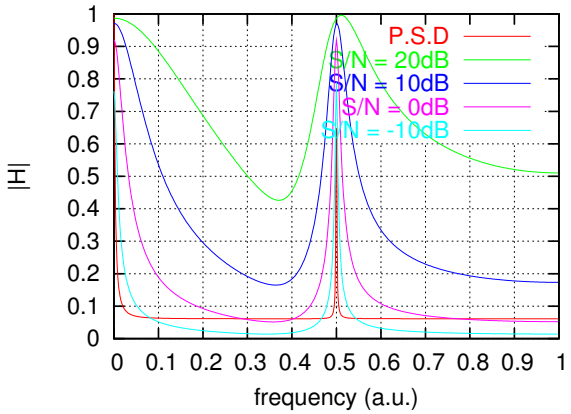

Fig. 14: Predictor filters assuming random walk and girder resonance for various noise levels in the measurement

optimum. A good and clean assessment of the system and its excitations are a prerequisite for the successful use of this technique.

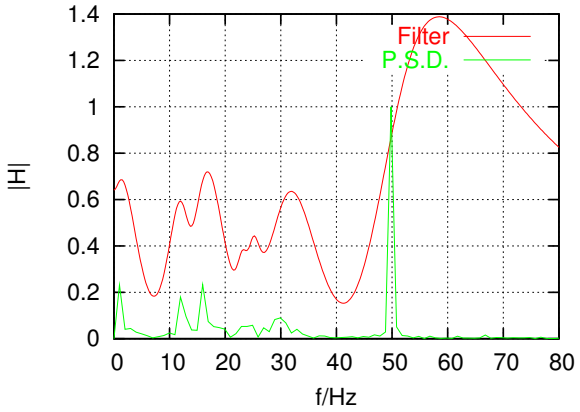

Fig. 15: Predictor generated from power spectral density of beam at SLS

As the last example, we look at a filter generated for real world data. Figure 15 show the power spectrum and the filter characteristics for beam jitter data seen at the Swiss Light Source. As seen before, the filter removes the peaks in the power spectral density; nonetheless the characteristic shown is not the optimum for the task. We have non-random components in the signal such as the 50 Hz line (which is due to magnetic stray fields of the power lines) or the 3 Hz line, which corresponds to the SLS booster cycle. These are completely predictable and should be handled by feed-forwards. Not visible in the spectrum are artificial lines in the spectrum showing up from other sources via cross-talk and interference, which also have to be handled separately. More on this topic is shown in the section on bunch-by-bunch feedbacks later.

3 Tune and chromaticity feedbacks

Quadrupole magnets show focusing effects only in one of the transverse planes; beam offsets in the other will be magnified. By setting a horizontally defocusing quadrupole into the focal point of a horizontally focusing one of equal strength, one obtains a net focusing action of the whole assembly. Applying this principle simultaneously for both planes, we get an arrangement of magnets called a FODO lattice. Other magnet types such as sextupoles or octupoles may be included in the design to correct for aberrations in

the optics. A detailed introduction to beam optics can be found in Ref. [13].

The focusing strength not only controls the transverse dimensions of the circulating bunches, but also affects the trajectories of bunches, which are off-axis. A bunch at an offset will oscillate transversally until damping effects due to synchrotron radiation cause it to settle down to the reference trajectory. The number of oscillations per machine turn is called the betatron tune Q and is proportional to the focusing strength of the machine.

The deflection of a particle due to magnetic fields is inversely proportional to its energy. Correspondingly the tune will vary with beam energy. Analogous to optics, this parameter is called the chromaticity Q'. Each ring optic has a natural chromaticity, which can be influenced by additional sextupole magnets. As with the tune values, we may have different chromaticities for each plane.

For a storage ring running at a constant energy and having a working orbit stabilization, we can expect only minor and very slow variations in tune and chromaticity. The situation is different if the ring is ramped in energy. Non-linearities of the magnets lead, even for a gentle ascent of the magnet currents, to fluctuations in the beam optics.

The control of the tunes and their associated parameters involves several aspects. The first is doing a clean measurement of the tunes, the second setting up a feedback. Third comes accounting for higher order effects due to tune coupling in the horizontal and vertical planes and the interaction with chromaticity measurement and feedback, which typically leads to a hybrid system design.

3.1 Tune measurement

In principle, all techniques of measuring the tune look at the oscillation frequency of particles around the design trajectory. The most direct way consists in kicking the bunches and doing a Fast Fourier Transform on the BPM data. This has several drawbacks. The beam quality (emittance) gets worse. The kick needs to be strong enough to give a visible signal, so the resultant orbit excursions cause particles in the beam halo to get lost and produce radiation—an effect which is of special concern for high-energy hadron colliders like RHIC or LHC. The latency of the measurement is mainly given by the rate by which the beam is kicked. A short latency, required for good feedback operation, means a high kick rate, but conflicts with the need for low emittance and particle loss rates.

An alternative relies on the fact that the charge distribution in a bunch is never 100% homogeneous– it consists of discrete, randomly distributed particles, so that we always have residual inhomogeneities exhibiting so-called Schottky or shot noise. These inhomogeneities oscillate with the tune frequency and can be measured directly in base band with Schottky monitors [14]. There is no external excitation to the beam, so this is the ideal measurement in terms of emittance dilution and the effect on beam loss rates. The headache is that we derive our measurements from noise spectra, so the signal-to-noise ratio may vary strongly depending on the initial condition of the injected beam and the machine settings. For low levels of Schottky noise, a relatively long measurement and measurement latency may be needed to obtain the required resolution, causing problems for the feedback.

The third and last option, which is nowadays used for feedback applications [15, 16] and which is described in more detail here, is to run an active phase-locked loop on the beam using fast kickers and pickups.

The typical structure is shown in Fig. 16. A numerically controlled oscillator (NCO) feeds a sinusoidal signal into a kicker deflecting the beam. The transverse beam oscillation has the maximum amplitude at the tune frequency. At this frequency, the motion is in phase with the excitation driving the kicker, whereas we have a phase difference approaching ± 90 degrees off crest. A phase detector compares the beam signal coming from a pickup with the reference from the NCO. The resultant phase difference signal is fed into a controller steering the frequency of the NCO.

The most basic phase detector would be a simple mixer followed by a low pass filter. With the

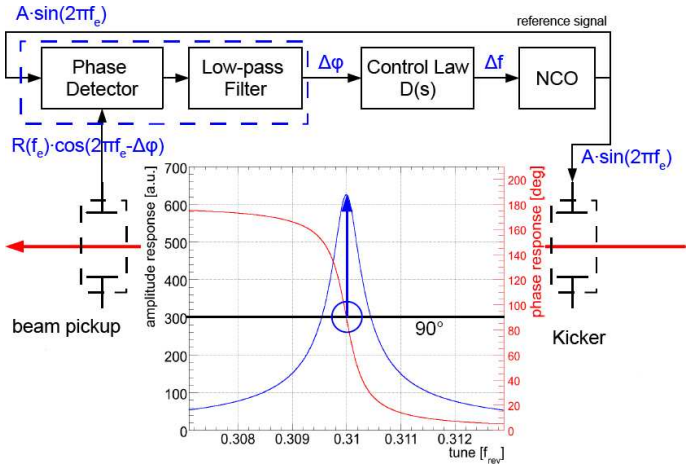

Fig. 16: Classical setup for tune measurement using a PLL loop

beam response

$$H(\omega) = A(\omega) * e^{-j\phi(\omega)} \ ,$$

we get

$$\begin{aligned} z_r &= A(\omega) \underbrace{\cos(\omega t)}_{\text{NCO signal}} \cdot \underbrace{A(\omega)\cos(\omega t + \phi(\omega))}_{\text{beam response}} \\ &= \frac{A(\omega)}{2}\Big(\underbrace{\sin(\phi(\omega))}_{\approx\phi(\omega)} + \underbrace{\sin(2\omega t + \phi(\omega))}_{\text{removed by low pass}} \Big) \ . \end{aligned}$$

This detector can be implemented in a relatively simple and robust way using analog circuitry. The disadvantage are non-linearities arising from the linearization of the sine argument and the amplitude dependency of the beam response, which can lead to reduced performance and even chaotic behaviour of the PLL loop [17].

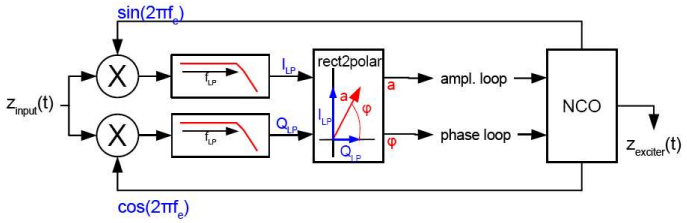

Fig. 17: Advanced PLL detector decoupling phase and amplitude as used at the LHC

An advanced digital detector design used in the LHC [18] which avoids these problems is shown in Fig. 17. The beam signal is split into two parts, which are mixed in phase and out of phase with the signal from the NCO followed by a low pass filter. A rectangular-to-polar converter generates pure amplitude and phase signals. In addition to the phase-locked loop controlling the frequency and obtaining the tune, a second controller keeps the amplitude constant. In the case of the LHC design, the controller used is a PI design, which was optimized with Youla's affine parametrization method [19], which is also described in the Appendix. The bandwidth of the measurement is of the order of 8 Hz.

If we compare the PLL measurement to the kicked excitation, a much smaller amplitude of the beam oscillation is required to keep the PLL locked. At prototype tests at the SPS ring at CERN, amplitudes well below 1 μm were required, while the emittance blow up and particle losses due to the measurement were negligible.

3.2 Full system and extensions

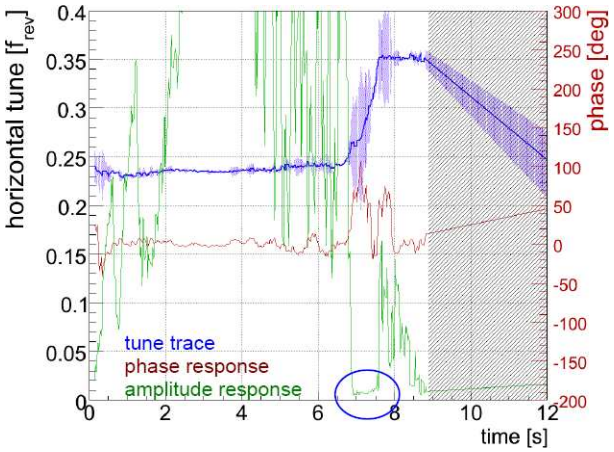

Fig. 18: Example of tune stabilization using PLL at the SPS/CERN

The closed loop performance of a tune feedback using the PLL is demonstrated in Fig. 18. Here a prototype was tested in the energy ramping cycle at the SPS, where within roughly seven seconds protons get accelerated from 26 GeV to 450 GeV. Apart from the tune, the output from the phase and amplitude control loop is also shown. Monitoring these value gives a good indication of whether the PLL is still locked. The region where the lock is lost (no more beam due to extraction) is marked with the blue circle.

The coupling of the beam motion between the horizontal and vertical planes can obscure the measurement, especially if, at certain times, both tunes are nearly equal, as happens for example during the ramp in RHIC [20]. In that case, exciting the bunches at the tune will result in a mixed vertical and horizontal motion. Also, the residual coupling leads to a perturbation of the tune values, decreasing the accuracy of the measurement. So one has to go one step further and use a hybrid feedback which simultaneously controls both tunes and coupling.

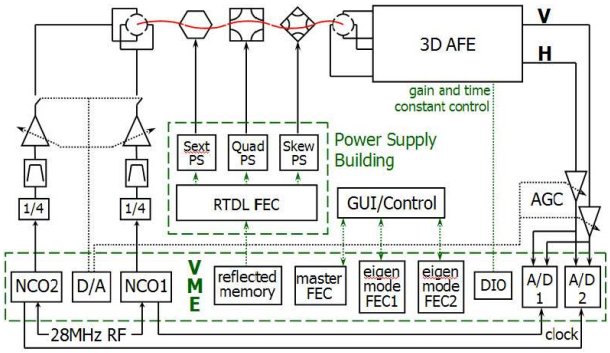

Fig. 19: Combined tune and coupling feedback at RHIC

The layout is shown in Fig. 19. We have phase-locked loops (analog front end AFE, A/D converter, numerically controlled oscillators NCO1 and 2) similar to that shown locking to the beam resonances,

which correspond to the perturbed tunes. Data processing for the individual planes is done by dedicated front end computers FEC1 and 2; a master FEC computes the required corrections, which are applied by a fourth computer to the sextupoles, quadrupoles, and skew quadrupole families.

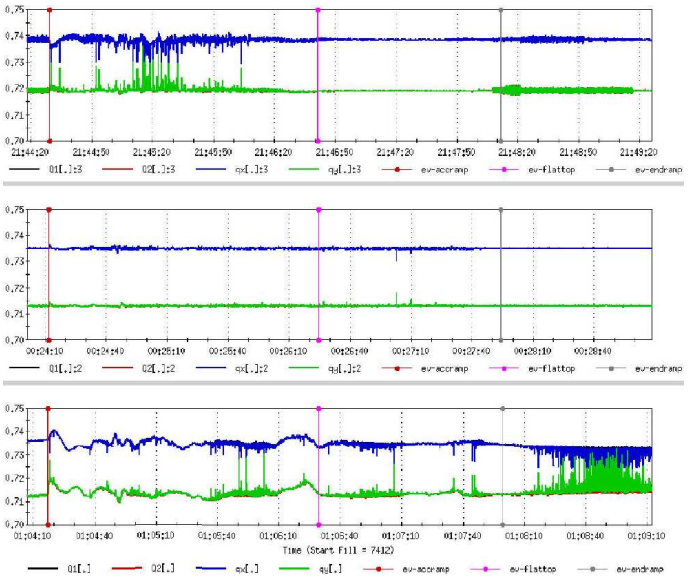

Fig. 20: Set and measured tunes of combined tune and coupling feedback at RHIC; uppermost trace: feedback on; middle trace: feedback on plus feed-forward on magnets; lower trace: feedback off

Figure 20 demonstrates the performance of the system for a 100 GeV development ramp done over five minutes. The lowest set of traces shows the tunes with the feedback switched on. In the upper-most set, the feedback is switched on showing a marked improvement.

As presented in the introduction, feedback should be used only to correct for non-predictable perturbations. Fluctuations which are predictable either due to their physics or because they are periodic like in an accelerator cycle are best handled by a feed-forward system. The middle set of traces are a nice demonstration of this. Magnet correction values from previous ramps were used to update an adaptive feed-forward system. The feedback itself has to perform only minor corrections, the result is an excellent stability of the system.

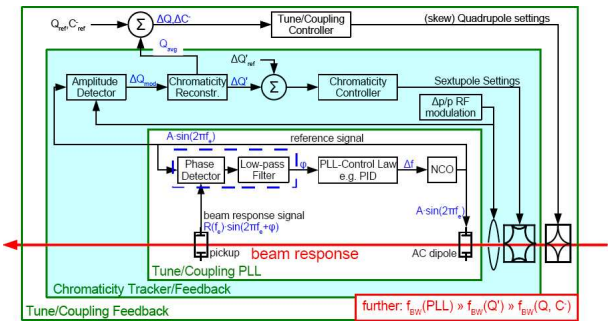

Fig. 21: Combined tune and chromaticity feedback at the LHC

For the LHC, a hybrid system is planned stabilizing all tune, coupling, and chromaticity. As shown in Fig. 21, it consists of a cascade of three different feedbacks. The innermost part is the PLL loop used to measure the tune. By modulating the RF, and measuring tune versus longitudinal momentum, the average

tune and the chromaticity are obtained and used for controlling the sextupole magnets. The measured tunes are used to compute the coupling and the unperturbed tunes and fed into a third controller steering the normal and skew quadrupole magnets.

It is clear that coupling between the nested feedbacks can be a problem, especially since the tune feedback would minimize the momentum driven modulation of the tune and render the chromaticity measurement useless. So the highest bandwidth of 8 Hz is assigned to the tune PLL, followed by one of circa 1 Hz for the chromaticity feedback. The tune feedback is the slowest one.

4 Fast data-processing algorithms for a longitudinal bunch-by-bunch feedback

Bunch-by-bunch feedback algorithms are covered in depth in Ref. [21], the idea for this section is to cover certain data processing features of these high-speed systems and keep the general description to a minimum.

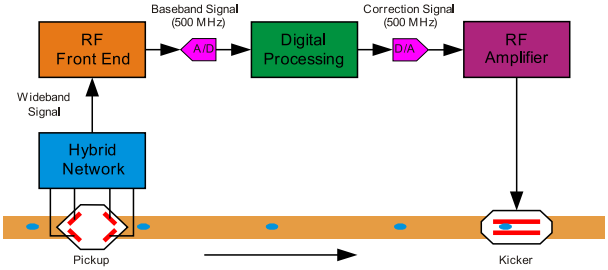

Fig. 22: Layout of bunch-by-bunch feedback system

The basic layout is shown in Fig. 22 (Ref. [22]). The effects to be corrected are unstable longitudinal oscillation driven by trapped resonances in the vacuum chamber, for example. For this application, the beam can be seen as resonant at the synchronous sidebands of the revolution harmonics of the accelerator.

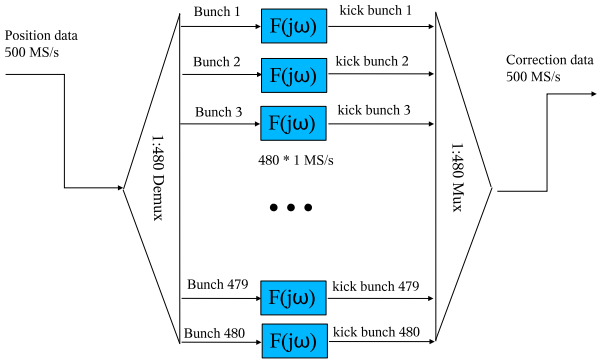

Fig. 23: Multiplexing and demultiplexing of data streams

The feedback is a single-input single-output (SISO) system with the special challenge that there are a large number of potentially instable resonances (e.g., 480 for the Swiss Light Source, 3564 for the LHC). By multiplexing the incoming signal into multiple channels (Fig. 23), each containing the data for one individual bunch, the problem can be simplified considerably. Owing to the inherent down sampling, all the resonant sidebands of the full signal get mapped to the same frequency for the single-bunch data stream, something which facilitates the controller design considerably. It may be tempting to see this approach as a decomposition of a decoupled state similar to using the singular-value decomposition

method for orbit stabilization. But the instabilities lead to a coupling of bunch motion, so this is not the case.

4.1 Signal and noise

Using the example of the SLS parameters, let us take a look at signal and noise levels. The detector works at a centre frequency of 1.5 GHz running at a band width of $B = 500$ MHz. Omitting for the moment cross-talk, beam harmonics, and other external noise sources, we are left with the thermal noise floor, which at a temperature of $T = 300$ K comes out as

$$N = k_B T B = -77 \text{ dBm} .$$

The RF front end may add another 6 dB, so that we end up with a noise power of -71 dBm.

As for the signal, we are not interested in static phase offsets, but in the instable oscillatory part. This is equivalent to a phase modulation of the beam signal and seen in the frequency domain as side bands of the revolution harmonics. The signal picked up by the RF front end is

$$V(t) = Ae^{j(\omega t + \phi \cos \omega_s t)} , \tag{18}$$

where ω is one of the revolution harmonics and ω_s is the synchronous frequency. For small oscillations, we can approximate

$$Ae^{j(\omega t + \phi \cos \omega_s t)} = Ae^{j\omega t} e^{j\phi \cos \omega_s t} \approx A\big(1 + j\phi \cos \omega_s t\big) e^{j\omega t} .$$

Applying standard trigonometric identities, we obtain three frequencies, the beam harmonic at ω with amplitude A and the two side bands $\omega \pm \omega_s$ with amplitude $A\phi$. A beam of 400 mA creates a main peak of -1 dBm, the noise floor of approximately -71 dBm corresponds to a phase noise of

$$\Delta\phi = 10^{-70/20 \text{ dB}} \text{ rad} = 0.3 \text{ mrad} \sim 2 \text{ ps at } 1.5 \text{ GHz} .$$

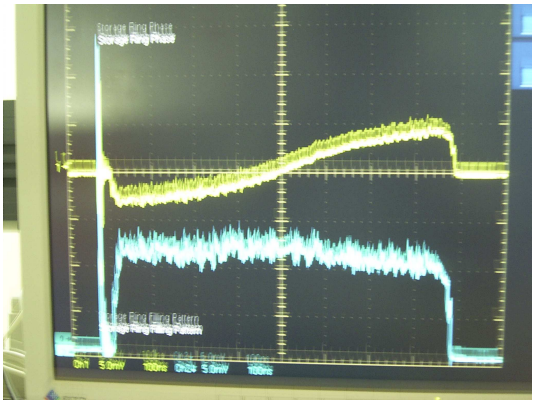

Fig. 24: Synchronous phase (yellow line) for fill pattern with gap (blue line), data from the SLS

On the analog side, everything looks fine for a successful conversion to digital and, for a homogeneous fill pattern with bunches in all RF buckets, this would hold true. But typically an inhomogeneous filling as in Fig. 24 is used, employing a gap to avoid accumulation of ions near the beam trajectory. As this inhomogeneous pattern circulates around the ring, the voltage in the cavity gets modulated by the fill pattern, so that the synchronous phase of the individual bunch will also vary from the reference phase. The result is that the ADC range has to be adjusted not only to cover the amplitudes of any dynamic

oscillations coming from instabilities, but also the static phase offsets. We end up with a pronounced decrease in resolution and a corresponding increase in quantization noise.

To make the picture complete, there are still other noise sources like truncation noise (due to the finite precision of the computations done in the controller) or white noise contributions (coming from the digital-to-analog converter and the power amplifiers). Nonetheless, the essential contribution is due to ADC quantization noise.

Is this critical for the system performance? In terms of a broadband excitation of beam jitter, the answer is essentially no, since the dynamic of the beam exhibits a narrow band-pass behaviour, so that it will react only to components in the noise spectrum close to the synchronous frequency. The big challenge lies in a possible saturation of digital controller, DAC, and the power amplifiers due to noise which can severely compromise the function and efficiency of the feedback. An example of how to set up the filter in terms of algorithm, data flux, and format in a manner that avoids these problems is shown in the next section.

4.2 Low-level processing

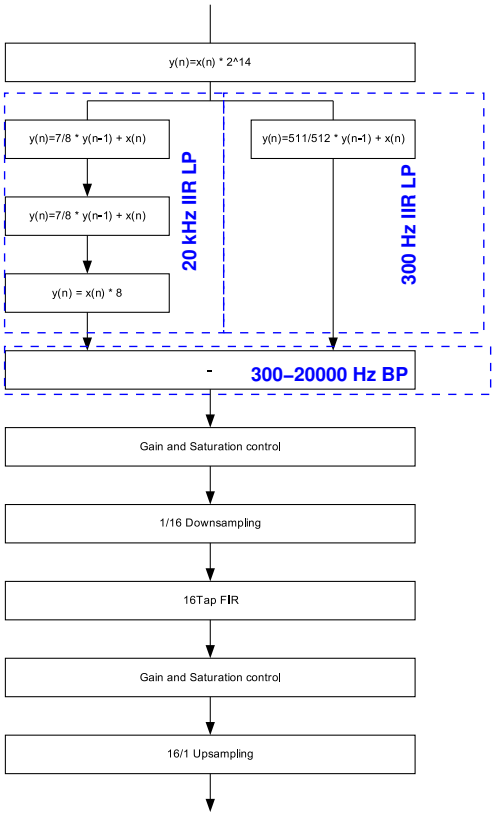

Fig. 25: Basic blocks of the total filter

After the analog-to-digital conversion, we are left with predominantly white noise, spread over the full spectrum. The only way of reducing it is to cut the bandwidth while leaving the signal intact. Given that the signal occupies frequencies in the near vicinity of the synchronous resonance[2], a band-pass design centred around the synchronous frequency is the most suitable approach. This implementation uses two filter stages. The first is a fixed IIR band-pass implemented so as to minimize internally generated noise,

[2]For the SLS, for example, we have a few hundred hertz around a synchronous frequency of 2–5 kHz compared with a full bandwidth of 500 kHz.

which cuts the noise and limits the bandwidth from DC –500 kHz to 0.3–20 kHz. Following that, the data rate is reduced by a 1:16 down sampling stage. Following that is the second, user configurable 16 tap filter. Figure 25 shows the layout.

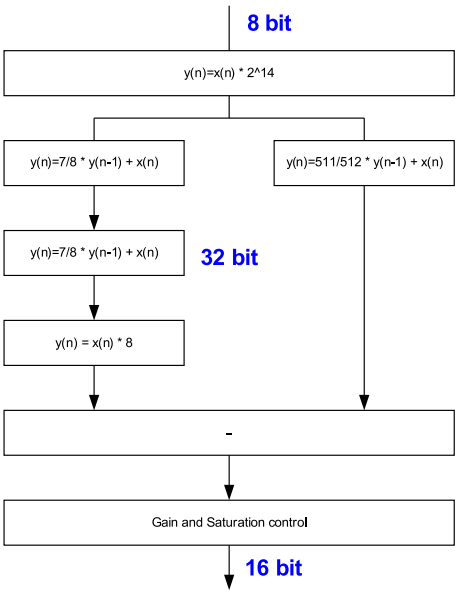

Fig. 26: Introductory bandpass IIR

Figure 26 gives a close look at the introductory IIR filter. The band pass is constituted as the difference of two IIR low-pass systems resulting in a bandwidth of 300–20 000 Hz. The internal computation is done in 32-bit fixed point. Finally a gain and saturation control stage reduces the data format to 16 bit.

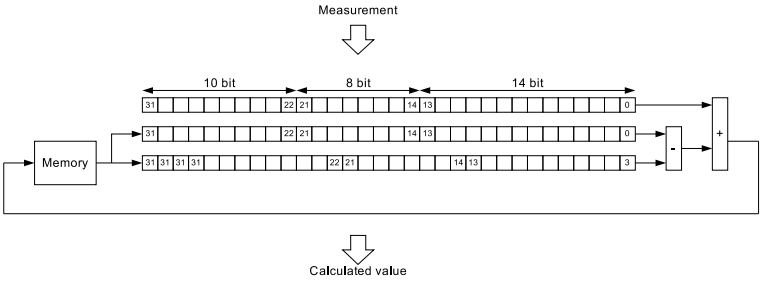

Fig. 27: Implementing the IIR low pass $y_n = \frac{7}{8}y_{n-1} + x_n$ with minimum noise

The low level implementation of one of the stages (shown in Fig. 27) computes the recursion

$$y_n = \frac{7}{8}y_{n-1} + x_n \ .$$

The input data is inserted into bits 14 to 21 of a 32-bit double word. The multiplication by $7/8$ is done by writing the data y_{n-1} into two registers, into the first as is and into the second using a shift by three bits (or divided by eight). Subtracting both gives the required prefactor 7/8. Only the shift operation results in a truncation error in the least significant bit. The maximum gain of this iteration is eight, so the result has significant values in bits 0 to 24.

Putting the result into the second IIR stage fills its output 32-bit word from bits 0 to 27, the following multiplication by eight (realized again as a bit shift) makes use of bits from 0 to 30 without

incurring any noise effects. The IIR low pass in the second branch is done in an analog way. The gain and saturation stage at the end does a user controlled truncation to a 16-bit data format, choosing the right gain will give the optimum balance between saturation and noise.

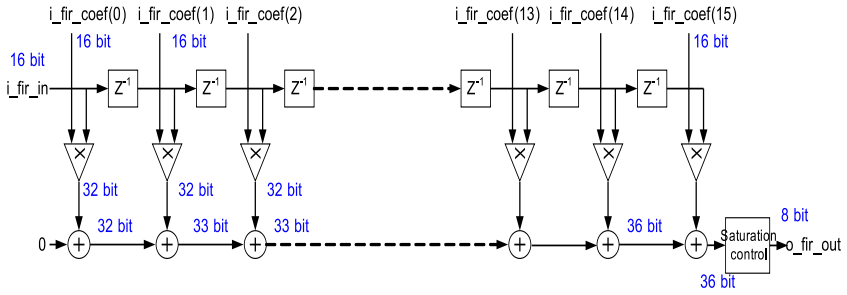

Fig. 28: Second stage: data flux and format of the 16-tap FIR filter

After the down sampling stage, the data rate is low enough to implement the user configurable FIR filter using 16-bit for coefficients and input (Fig. 28). Multiplying coefficients and input data without truncating leads to a 32-bit number. Adding two 32-bit numbers results in a 33-bit size, four 32-bit numbers give a 34-bit and so on until we end up with an untruncated data flow of 36-bits width. A second user configurable gain and saturation stage helps the user find a good balance between saturation and noise.

What remains is how to choose the sixteen coefficients. A quick way, suitable for a manual trial and error optimization, is shown in the next section.

4.3 Quick and easy controllers

A standard controller typically found in textbooks and used for universal application is a PID controller:

$$F(s) = P + \frac{b}{s} + cs \; .$$

Only three parameters have to be chosen to optimize the control loop—sometimes even the differentiator is omitted reducing these to two. For a bunch-by-bunch feedback, this approach cannot work, the main reason being that the PID controller is designed for operation around DC, but the perturbation is centred at a non-zero frequency. So it is interesting to ask what a corresponding filter looks like for a non-zero frequency ω_0.

For the proportional part, things are easy. A constant will stay that way also for non-zero frequency, so there are no changes here:

$$F_{DC} = F_\omega = P \; . \tag{19}$$

Next we have the integrator. In DC, we have a pole at $s = 0$, so the shifted integrator needs to have two at $s = j\omega_0$ and $s = -j\omega_0$, consequently, it should look like

$$F(s) = \frac{1}{s^2 + \omega_0^2} \; .$$

To be realizable, the corresponding time domain function should be real. If we look in a table of Laplace transforms, we find two candidates

$$\begin{aligned} \frac{s}{s^2 + \omega_0^2} &\iff \cos \omega_0 t \\ \frac{\omega_0}{s^2 + \omega_0^2} &\iff \sin \omega_0 t \end{aligned}$$

which correspond to in-phase and out-of-phase time domain responses and which we can identify with a real and imaginary prefactor. The integrator comes out to be

$$F_\omega = \left(b_r + b_i \frac{\omega_0}{s}\right) \frac{s}{s^2 + \omega_0^2} \,. \tag{20}$$

The design of the differentiator essentially follows the same pattern. Instead of a DC zero $s = 0$, we need two zeros at $s = \pm j\omega_0$. Again, there is an in-phase and an out-of-phase candidate, so that we have a real and imaginary prefactor giving us

$$F_\omega(s) = \left(c_r + c_i \frac{s}{\omega_0}\right) \frac{s^2 + \omega_0^2}{s} \,. \tag{21}$$

The corresponding time domain response is a pair of opposing pulses combined with a step function.

Table 1: Comparison of filter functions for PID controller centre on DC and offset at ω_0

Type	**PID at DC**	**PID at ω_0**
P	$F(s) = A$	$F(s) = a$
I	$F(s) = B\frac{1}{s}$	$F(s) = \left(b_r + b_i \frac{\omega_0}{s}\right) \frac{s}{s^2+\omega_0^2}$
D	$F(s) = Cs$	$F(s) = \left(c_r + c_i \frac{s}{\omega_0}\right) \frac{s^2+\omega_0^2}{s}$

An overview of the resulting functions is shown in Table 1. Where a standard PID controller has three parameters, we now specify five. To do the digital implementation, the values of the pulse response at discrete times are used to define the coefficients of an equivalent FIR filter. As an alternative, a combined FIR/IIR design is possible. A description for a similar controller used for low-level RF control can be found in Ref. [23].

4.4 Applying predictive filters

The purpose of the feedback is not to suppress wide band noise, but to shift the instable resonances back into a region with positive damping. We are only interested in the response of the predictor in the vicinity of the resonance—multiplied with an appropriate scalar factor, the predictor can be used directly as a controller.

What is the measurement and what is the signal or state we are looking for? The power spectral density of the longitudinal momentum is described by a resonance at the synchronous frequency with positive or negative damping coefficient. Since the bunch oscillates in the longitudinal phase space, measurement (the bunch phase) and state (the bunch momentum) are at 90 degree phase offset with respect to each other. In addition, the measurement m contains white noise, giving us the following power spectra (while neglecting the damping terms):

$$\begin{aligned} S_m(j\omega) &= \frac{A\omega^2}{\left(\omega^2 - \omega_s^2\right)^2} + N \\ S_s(j\omega) &= \frac{A}{\left(\omega^2 - \omega_s^2\right)^2} \,. \end{aligned}$$

The resulting 16 tap filters in Fig. 29 show the typical behaviour. As the measurement becomes noisy, the controller converges more and more to a band pass.

In a real machine, we are not only going to observe the synchrotron tune obscured by white noise. The signal also contains static offsets combined with low-frequency noise due to the regulation loops of

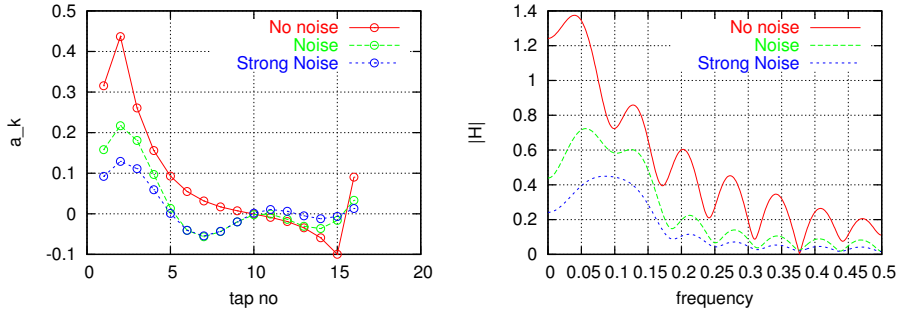

Fig. 29: Pulse and frequency response assuming only synchrotron resonance $f_s = 0.1$ for varying levels of measurement noise

the main RF, for example, something which we can approximate as a stochastic drift process. The naive approach would be to include this as a true signal in the description of the power spectrum of the state. The resulting filter would show a band-pass behaviour near the synchronous frequency as well as a peak near DC to bring out the drift signal. But the drift signal is precisely the one that the feedback should not react to! So, what we do is to define it as being a part of the measurement noise and let the predictor suppress it:

$$S_m(j\omega) = \underbrace{A\frac{1\omega^2}{\left(\omega^2-\omega_s^2\right)^2}}_{\text{synchrotron oscillation}} + \underbrace{B\frac{1}{\omega^2}+N}_{\text{drift/measurement noise}}$$

Figure 30 shows the result. We get a minimum suppressing the drift, as one would expect.

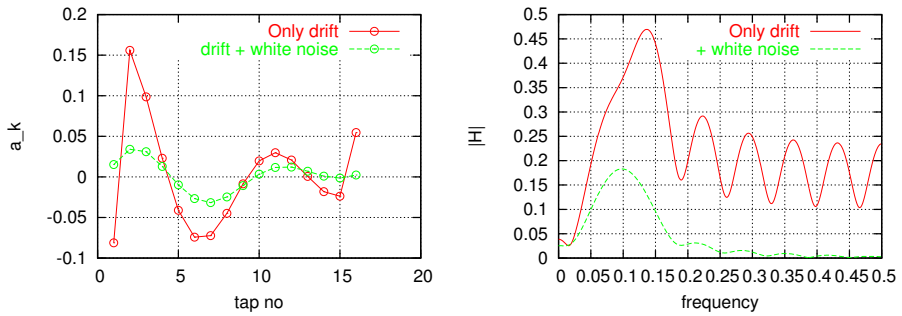

Fig. 30: Using predictor to suppress unwanted drift signals in the spectrum

4.4.1 Accommodating estimation and model errors

In the following discussion, let us take a look at how to handle tolerances in setting up the physical model (e.g., the value of the synchronous frequency) or non-stationary effects. The normal approach would be to use a Kalman filter, but incorporating an adaptive, time-varying filter into a feedback loop could easily lead to unstable closed loop behaviour. Also, if changes are rapid, the adaptive filter may not be able to follow within the required time.

In the context of bunch-by-bunch feedbacks, there is a good example for that. Often, synchrotron light sources have higher harmonic cavities (HHC) built into the ring. These are passive devices excited by the beam. By changing the slope of the RF voltage versus phase seen by the beam, they produce longer bunches with higher stability thresholds with respect to intra bunch oscillations. Since the synchronous

frequency depends on the slope, it also varies with the induced voltage in the HHC. In the case of the SLS, it varies from a standard value of approximately $\nu_s = 5 \cdot 10^{-3}$ down to values of $2 \cdot 10^{-3}$.

If we optimize the feedback only for the current operating value, the following vicious circle can appear: In the beginning, we may have the onset of a longitudinal instability. As the beam is oscillating, the lines in the beam spectrum widen up and drop in height, so the beam excites less voltage in the HHC. As a result, the synchronous frequency changes from the design value and so the feedback becomes less efficient in stabilizing the situation. The beam oscillation grows, the voltage in the HHC drops, the synchronous frequency deviates even more. We end up in a full blown instability. The process happens really quickly, too short for an adaptive filter algorithm to react.

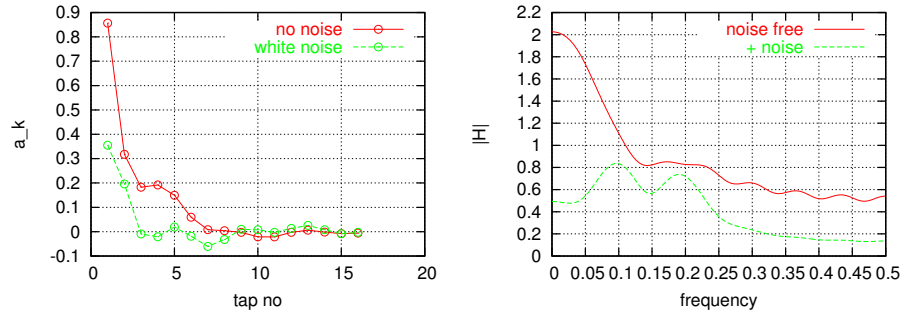

Fig. 31: Including variations in the synchrotron frequency into the predictor ($0.1 < f_s < 0.2$)

So what we could do is to include the expected range of the synchronous frequency (as well as other variables) as a stochastic density in the design process. Staying with our generic examples, we assume a variation of f_s between 0.1 and 0.2 and take a constant density distribution of

$$D_{fs}(f_s) = \begin{cases} C; 0.1 < f_s < 0.2 \\ 0; \text{ else} \end{cases}$$

and create an adapted power density to give

$$S_s(j\omega) = \int S(\omega, f_s) D_{fs}(f_s) df_s = \int_{0.1}^{0.2} \frac{CA\omega^2}{\left(\omega^2 - \omega_s^2\right)^2} df_s \; .$$

The resultant filter shown in Fig. 31 takes account of this spread of frequencies. A similar filter to this is actually in use at the SLS bunch-by-bunch feedbacks.

5 Outlook

Accelerator-based feedbacks have come a long way from early analog systems used for basic local stabilization. Several trends are visible. Digital systems allowing for more flexibility have replaced more and more components in the analog signal path. The system architecture went from local systems to ones taking into account cross coupling effects, be it the global orbit or the relationships between tunes, tune coupling, and chromaticity. Digital systems always have the disadvantage of higher latencies compared to analog solutions, so lots of work is devoted to speeding up processing, data flux, and data rates.

Given that we will reach a point of vanishing returns, the next step is to replace the currently used standard controllers by optimized feedback controllers. The successful use of predictor filters or LQR controllers needs a profound knowledge of the accelerator and of the stochastic properties of the processes determining the perturbations to be expected.

Acknowledgements

I would like to thank many people for the information and data provided, especially Ralph Steinhagen, who provided lots of material on LHC feedbacks, Peter Cameron, Thomas Schilcher and Goran Marinkovic. Special thanks go to Glenn Behrmann for valuable help in hunting down (I hope all) clumsy formulations in the text.

Appendix

A Youla–Kucera parametrization

Finding optimum controllers (with respect to internal noise, transient behaviour and so on) by performing an unrestricted optimization over a whole function space often runs into severe problems. The resultant system may be unstable or the controller not physically realizable. So it is quite attractive to first generate the complete set of controller functions with stable properties, before doing a restricted optimization on this set. One method, which will be shown in the following, is the Youla–Kucera parametrization.

In order to start, we first introduce some terms coming from pure mathematics. A ring R is a set of elements with the following properties:

- On all elements of R, there exists a commutative addition
$$A + B = B + A \, .$$
- On all elements of R, there exists a commutative multiplication operation
$$A \cdot B = B \cdot A \, .$$
- An element C having a multiplicative inverse D is called a unit element. Not every element of R needs to be a unit element.

Examples for rings are the following:

- The set of integer numbers $\mathbf{Z}$.
- The set of proper Hurwitz stable functions, where both nominator and denominator of the fraction are Hurwitz polynomials.
$$R_H(s) = \frac{\sum a_i s^i}{\sum b_i s^i} \, .$$
- The set of discrete FIR functions
$$R_f(z) = \sum a_k z^k \, .$$

If all elements in a ring are unit elements, that is, we can define a division for all elements, it is called a field. Examples for fields are the sets of real numbers or the sets of rational functions.

Equations on a ring are the so-called Diophantine equations [24]. An example for this is the search for solutions of
$$A^n + B^n = C^n \text{ for } A, B, C \in \mathbf{Z}, n = 2, 3, ...$$
which for $n > 2$ is the subject of the famous last theorem of Fermat [25]. For our purposes, we are interested in solutions of the linear equation
$$AX + BY = C \, , \tag{A.1}$$
where $A, B, C, X, Y \in \mathbf{R}$, A, B, C are given and we are looking for solutions (X, Y). This equation has the following interesting property. Assume, we have already found one pair X' and Y' solving Eq. A.1. Then we can create other pairs by setting
$$X \;\; = \;\; X' + BW \tag{A.2}$$

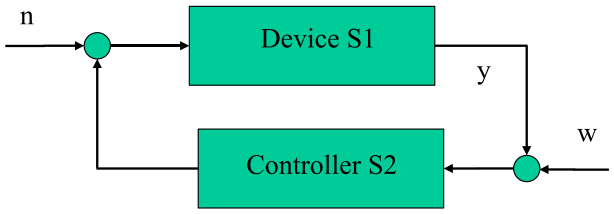

Fig. A.1: Simple control loop

$$Y \quad = \quad Y' + AW,\ W \in \mathbf{R} \tag{A.3}$$

which will solve the equation. What is more, this method will create all solutions in **R**.

How do we use that rather abstract result in the context of designing controllers? For that, let us take a look at the simple feedback schematic in Fig. A.1. We assume that we can write both device and controller transfer functions as fractions of Hurwitz stable functions $R_H(s)$:

$$\begin{aligned} S_1(s) &= \frac{B(s)}{A(s)}\,, \\ S_2(s) &= -\frac{Y(s)}{X(s)}\,. \end{aligned}$$

With this representation, the closed loop behaviour of the parameter $y(s)$ is given as

$$y(s) = \frac{BX}{AX+BY}n(s) + \frac{-BY}{AX+BY}w(s) \tag{A.4}$$

and a condition for $y(s)$ to be stable is that the denominator $AX + BY$ be a Hurwitz stable function.

Now let us assume that we already know one stabilizing controller, which we again decompose into a fraction:

$$S_2'(s) = -\frac{Y'(s)}{X'(s)}\,. \tag{A.5}$$

Using this particular solution and arbitrary functions $W(s) \in \mathbf{R_H(s)}$, we can generate a whole set of solutions

$$\begin{aligned} X(s) &= X'(s) + B(s)W(s) \\ Y(s) &= Y'(s) - A(s)W(s)\,, \end{aligned}$$

since, according to our Diophantine equation above

$$AX + BY = AX' + BY' \in \mathbf{R_H(s)}\,.$$

So the parametrization of all stabilizing controllers will take the form

$$S_2(s) = -\frac{Y'(s) - A(s)W(s)}{X'(s) + B(s)W(s)}, \forall W \in \mathbf{R_H(s)}\,. \tag{A.6}$$

A.1 Example

As a simple example, we assume the following characteristics for device and controller. The device is proportional with unity gain, so that we can write it simply as

$$S_1(s) = 1 = \frac{1}{1} = \frac{A(s)}{B(s)}\,.$$

As a particular case of a stabilizing controller, we assume an integrator

$$S_2(s) = -\frac{1}{s} .$$

When decomposing the transfer function into a fraction, we need both nominator and denominator to be Hurwitz stable, writing for example:

$$S_2(s) = -\frac{1}{s} = -\frac{\frac{1}{s+1}}{\frac{s}{s+1}} = -\frac{Y'(s)}{X'(s)} .$$

Now the parametrized controller can be written as

$$S_2(s, W(s)) = -\frac{\frac{1}{s+1} - W(s)}{\frac{s}{s+1} + W(s)}, \forall W(s) \in \mathbf{R_H(s)} ,$$

using for example

$$W(s) = \frac{1}{s+a}; a > 0 .$$

A possible disadvantage of this approach is, that the parametrization is in terms of a ring of functions and not functional space like a Hilbert space, for which we have lots of techniques for optimization and synthesis. Nonetheless, one is typically looking for controller designs of a low order and complexity, so that the effort stays manageable.

As an alternative, we can use the same approach with controller and device transfer functions exchanged. As a result, we obtain a parametrization of all devices, which can be stabilized by a given controller. Comparing these with the expected variations in the device behavior due to tolerances and drifts gives a valuable criterion for the robustness of the chosen controller.

References

[1] G.H. Golub and C.F. Van Loan, *Matrix Computations* (Johns Hopkins University Baltimore Press, 1989).

[2] `http://web.mit.edu/be.400/www/SVD/Singular_Value_Decomposition.htm`

[3] T. Wijnands *et al.*, Requirements for real-time correction of decay and snapback in the LHC superconducting magnets, Proc. of the EPAC 2000, pp. 367–369, Vienna, Austria, 2000.

[4] M. Boege *et al.*, Fast closed orbit control in the SLS storage ring, Proc. of the 1999 Particle Accelerator Conference, pp. 1129–1131, New York, USA, 1999.

[5] M. Dehler, Eddy current calculations for the SLS corrector magnets, SLS note SLS-TME-TA-1998–0004, Villigen-PSI, Switzerland (1998).

[6] Vladimir Shiltsef, Space-time ground diffusion: The ATL law for accelerators, Proc. of the 4th IWAA95, Nov. 14–17, 1995, KEK, Japan. IV 352.

[7] S. Redaelli *et al.*, Vibration measurements at the Swiss Light Source (SLS), Proc. of the EPAC 2004, Lucerne, Switzerland, pp. 2278–2280, 2004.

[8] B.D.O. Anderson and J.B. Moore, *Optimal Control: Linear Quadratic Methods* (Prentice Hall, Englewood Cliffs, NJ, 1990).

[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. (McGraw-Hill, New York, 1991).

[10] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (MIT Press, Cambridge, MA, 1950).

[11] N. Levinson, The Wiener RMS error criterion in filter design and prediction, *J. Math. Phys.* 25 (1947) 261–278.

[12] R.E. Kalman, A new approach to linear filtering and prediction problems, *ASME Trans.* vol. 82D (1960) 35–45.

[13] A. Hofmann, Single particle dynamics 1 – Basic phase space, in *Beam Measurement*, Proc. Joint US – CERN – Japan – Russia School on Particle Accelerators, ISBN 981-02-3881-9, 1999.

[14] F. Nolden, Instrumentation and diagnostics using Schottky signals, Proc. 5th European Workshop on Beam Diagnostics and Instrumentation for Particle Accelerators DIPAC 2001, pp. 6–10, Grenoble, France, 2001.

[15] P. Cameron *et al.*, PLL tune measurement during RHIC 2001, Proc. European Particle Accelerator Conference EPAC 2002, pp. 1855–1857, Paris, France, 2002.

[16] R. Steinhagen, Feedbacks on tune and chromaticity, Proc. 8th European Workshop on Beam Diagnostics and Instrumentation for Particle Accelerators DIPAC 2007, Venice, Italy (2007).

[17] B. Harb *et al.*, Chaos and bifurcation in a third order phase locked loop, Chaos, Solitons and Fractals 19 (2004) 667–672, doi:10.1016/S0960-0779(03)00197-8.

[18] R. Steinhagen, Real time feedback on beam parameters, Proc. Fourth Asian Particle Accelerator Conference APAC 2007, Indore, India (2007).

[19] D.C. Youla *et al.*, Modern Wiener–Hopf Design of Optimal Controllers, *IEEE Trans. Automat. Control*, 1976, Vol. 21, No. 1, pp. 3–13 (part I) and pp. 319–338.

[20] P. Cameron *et al., PRST-AB* 9 (2006) 122801.

[21] M. Lonza, Multibunch feedback systems, these proceedings.

[22] M. Dehler *et al.*, State of the SLS multi bunch feedbacks, Proc. Fourth Asian Particle Accelerator Conference APAC 2007, Indore, India, 2007.

[23] L. Doolittle *et al.*, Digital low-level RF control using non-IQ sampling, Proc. LINAC 2006, pp. 568–570, Knoxville TN, USA, 2006.

[24] V. Kucera, Diophantine equations in control - a survey, *Automatica* 29 (1993) 1361–1375, UTIA AV CR. research report No. 1778, Academy of Sciences of the Czech Republic, 1993. Also to be found at `ftp://ftp.utia.cas.cz/pub/reports/utia1778.ps.Z`.

[25] S. Singh, *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem* (Anchor Books, New York, 1998).