LOW-LEVEL RF SYSTEMS FOR SYNCHROTRONS
Part II: High intensity. Compensation of beam-induced effects

P. Baudrenghien
CERN, Geneva, Switzerland

Abstract

The high-intensity regime is reached when the voltage induced by the beam in the RF cavities is of an amplitude comparable to the desired accelerating voltage. In steady state this *beam loading* can be compensated by providing extra RF power. Transient beam loading occurs at injection or in the presence of a beam intensity that is not uniform around the ring. The transients are periodic at the revolution frequency. Without correction transient beam loading can be very harmful: the stable phase and bucket area will not be equal for all bunches. Strong beam loading often occurs with longitudinal instabilities because the RF cavities are a large contributor to the total ring impedance. The low-level systems that reduce the effect of transient beam loading will also increase the threshold intensity of longitudinal instability caused by the cavity impedance at the fundamental RF frequency. Four classic methods are presented here: feedforward, RF feedback, long delay feedback, and bunch-by-bunch feedback. The first three fight against both transient beam loading and longitudinal instability, if caused by the cavity impedance (at the fundamental). The last cures longitudinal instability (dipole mode) caused by any impedance in the machine but has no effect on beam loading. These techniques have been made possible by the recent advent of fast digital circuitry and an emphasis will be put on implementation.

1 THE PROBLEM OF BEAM-INDUCED VOLTAGE

1.1 Transient beam loading

This Section is a brief presentation of the basics of beam loading. Refer to the literature for more details [1]–[5].

The beam crossing the RF cavity induces an electromagnetic field within it, thereby creating a decelerating voltage V_b acting on the beam in return (Fig. 1). The accelerating voltage seen by the beam is thus the vector sum of the voltage produced by the generator V_g and the beam-induced voltage V_b

$$V_t = V_g + V_b = Z_g I_g + Z_b I_b \tag{1}$$

where I_g is the RF drive and I_b is the beam current.

In the case of a *standing wave cavity*[1], the voltages V_g and V_b are simply the accelerating voltages at the cavity gap created by the generator and by the beam, respectively [6]. Around one resonance ω_0 the standing wave cavity can be modelled as a lumped Resistor–Inductor–Capacitor (RLC) parallel circuit. The two impedances Z_g and Z_b are proportional, with a ratio that is a function of the main coupler transformation ratio. To simplify the notations, we will make them equal

$$Z_b(\omega) = Z_g(\omega) \tag{2}$$

$$Z_g(\omega) = \frac{R}{1 + j2Q\frac{\Delta\omega}{\omega_0}} \tag{3}$$

[1] This applies to ferrite-loaded cavities as well.

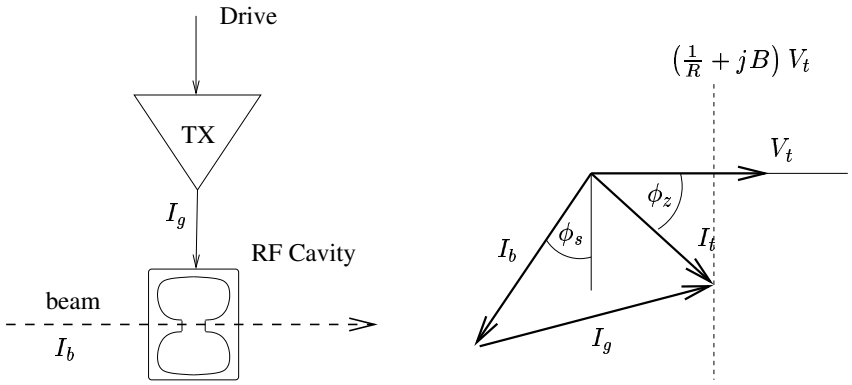

Fig. 1: Left: principle of beam loading. The beam current I_b induces a voltage in the cavity that modifies the total accelerating voltage V_t. Right: vector diagram relevant to stationary beam loading in a standing-wave cavity.

where Q is the quality factor and $\Delta\omega = \omega - \omega_0$.

For a *travelling-wave cavity*, which looks like a matched transmission line for the power generator, $Z_g(\omega)$ is proportional to $\sin\tau/\tau$, with τ being the transit time factor. $Z_g(\omega)$ is purely real. First it decreases for increasing $|\Delta\omega|$; it crosses zero at a frequency offset equal to the inverse filling time of the cavity; and it changes sign thereafter (deceleration) [7]. The impedance $Z_b(\omega)$ is also a function of τ but it is not simply proportional to $Z_g(\omega)$. It has both real and imaginary parts.

If the bunch intensity is uniform around the ring and if we analyse the situation well after injection, we are in the *stationary* situation. In the frequency range of the RF system, the beam current I_b is a single spectral line at the RF frequency f_{rf}. The currents and voltages of Eq. (1) can be represented as vectors in the complex plane (Fig. 1). Consider a standing-wave cavity. Let $Y(\omega)$ be the cavity admittance

$$Y(\omega) = 1/Z_g(\omega) = 1/R + jB(\omega) \tag{4}$$

with

$$B(\omega) = 2\frac{Q}{R}\frac{\Delta\omega}{\omega_0} \ . \tag{5}$$

Making $Z_b = Z_g$, Eq. (1) can be rewritten

$$I_g + I_b = YV_t \ ; \tag{6}$$

I_g must compensate I_b to keep the modulus of the total accelerating voltage V_t at the desired value (desired bucket area). This is done by adjusting the amplitude of the generator drive and by detuning the cavity (angle ϕ_z in Fig. 1). This latter changes the value of the cavity susceptance B, while the conductance $1/R$ remains constant. In the vector diagram of Fig. 1 the vector $I_t = I_g + I_b$ follows the dashed line perpendicular to V_t as the cavity tune is changed. The amplitude $|I_g|$ is minimum when V_t and I_g are in phase. This is the desired working point for the power generator. If the stable phase angle ϕ_s is not zero (acceleration), extra RF power is needed to compensate stationary beam loading. For a standing-wave cavity, the low-level system needed includes a so-called cavity field amplitude loop, which adjusts the amplitude of the generator drive $|I_g|$ to get the desired $|V_t|$, and a slow cavity tuning loop, which maintains V_t and I_g in phase (matched conditions at the cavity input minimizing the generator drive $|I_g|$) [1],[3],[5]. No tuning loop is needed with a *travelling-wave cavity*. It remains a matched load in the presence of beam loading [7]. Stationary beam loading is compensated by the cavity field amplitude loop (identical to the one used with standing-wave cavities).

A *transient* situation is encountered at injection. In the time domain, the beam current suddenly jumps from zero to the ring distribution and we cannot, in the frequency domain, reduce our analysis to

the single spectral component at f_{rf}. Equation (1) is still valid, but the two sides are Fourier transforms with a broad frequency spectrum. Because the RF system is only reacting to a narrow band around f_{rf} we can consider these signals as a band-limited modulation of the carrier frequency f_{rf}. When injecting a bunched beam, the voltage V_b varies from zero to the full beam loading. The low-level system must react in a short time compared with the synchrotron period, to restore the proper bucket area and the correct phase of the total voltage V_t with respect to the beam current I_b. Two loops come into action. The fast beam phase loop, present in most hadron machines [8], tries to restore the stable phase between $I_b(\omega_{rf})$ and $V_t(\omega_{rf})$, while the cavity field amplitude loop adjusts the drive amplitude $|I_g(\omega_{rf})|$ to restore the desired $|V_t(\omega_{rf})|$. For the beam phase loop, V_t is the vector sum of the total voltages in all the cavities. In the case of a standing-wave cavity a third loop (cavity tuning loop) tries to keep the generator current $I_g(\omega_{rf})$ in phase with the total voltage $V_t(\omega_{rf})$ (matched conditions at the cavity input). The argument $\omega_{rf} = 2\pi f_{rf}$ emphasizes that these loops work on the carrier component only, i.e. the component at the exact RF frequency. This works fine for small beam current, that is, when the accelerating voltage is predominantly determined by the generator current. For higher beam currents, a variation of the amplitude of $I_g(\omega_{rf})$ not only results in a variation of the amplitude of $V_t(\omega_{rf})$ but also of its phase. The loops that were independent at low beam currents become coupled, and unstable behaviour of the system results above a certain beam current threshold. Let I_0 be the generator current required to produce the accelerating voltage $|V_t|$ without beam loading and with the cavity tuned to resonance: $I_0 = |V_t|/R$. The beam loading is characterized by the ratio $|I_b|/I_0$, which is called the relative beam loading. It can be shown that the loops become unstable when this ratio approaches two [9]. The result is that the loops do not lock at injection: the beam is not captured.[2]

Transient beam loading also occurs if the ring is not filled uniformly (such as in the presence of a hole reserved for the possible triggering of a beam dump) or if the filling time of the cavities is not very large compared to the revolution period T_{rev}. The value of V_t will vary in amplitude and phase along the batch because of the modulation of the beam current I_b. The bucket area and the phase of the bunch with respect to V_g will also be modulated along the batch. They will not be correct for some bunches in the batch since the beam phase loop and the cavity field amplitude loop only adjust their average values (values at the carrier frequency f_{rf}). Modulation of the bucket area may result in a loss of particles from some bunches in the batch because the bucket area is too small (Fig. 2).

In the case of a collider the modulation of the longitudinal position of the bunches (phase of the bunch with respect to V_g) will displace the collision point for some bunches. In the case of an injector it will reduce capture efficiency when the bunched beam is injected into the receiving machine because some bunches will not see the correct RF phase (if the transfer is of the bunch-into-bucket type). For high-intensity proton machines, the beam loading can be greater than the RF voltage. The nominal LHC beam in the CERN SPS, for example, (1.05×10^{11} protons per bunch, DC current 0.67 A) induces 4.5 MV total in the RF cavities, compared with a matched capture voltage of 0.65 MV (bunch emittance 0.35 eVs, bunch length 4 ns). Fortunately transient beam loading compensation is only needed at a few discrete frequencies: as explained in Section 4.1 (Eq. (69)), the voltage induced by the beam in the cavity consists of a line at the carrier frequency f_{rf} (stationary beam loading) plus sidebands at multiples of the revolution frequency on each side of the carrier (transient beam loading). The strength of the sidebands decreases as we move away from f_{rf} with an envelope that is a function of the ring pattern. Therefore, transient beam loading compensation is only needed at the frequencies

$$f = f_{rf} \pm n f_{rev} \quad . \tag{7}$$

In the above equation the index n goes from 0 to ∞. In practice, however, the spectrum of the beam-induced voltage is limited to the bandwidth of the RF cavities. Ideally the compensation should cover a significantly larger bandwidth to get a good correction of the fast components present when the head of the batch enters the cavity, or when the beam is injected into the machine.

[2] If we avoid the strong transient beam loading at injection, by slow accumulation scheme for example, the loops would still become unstable above the same beam current threshold.

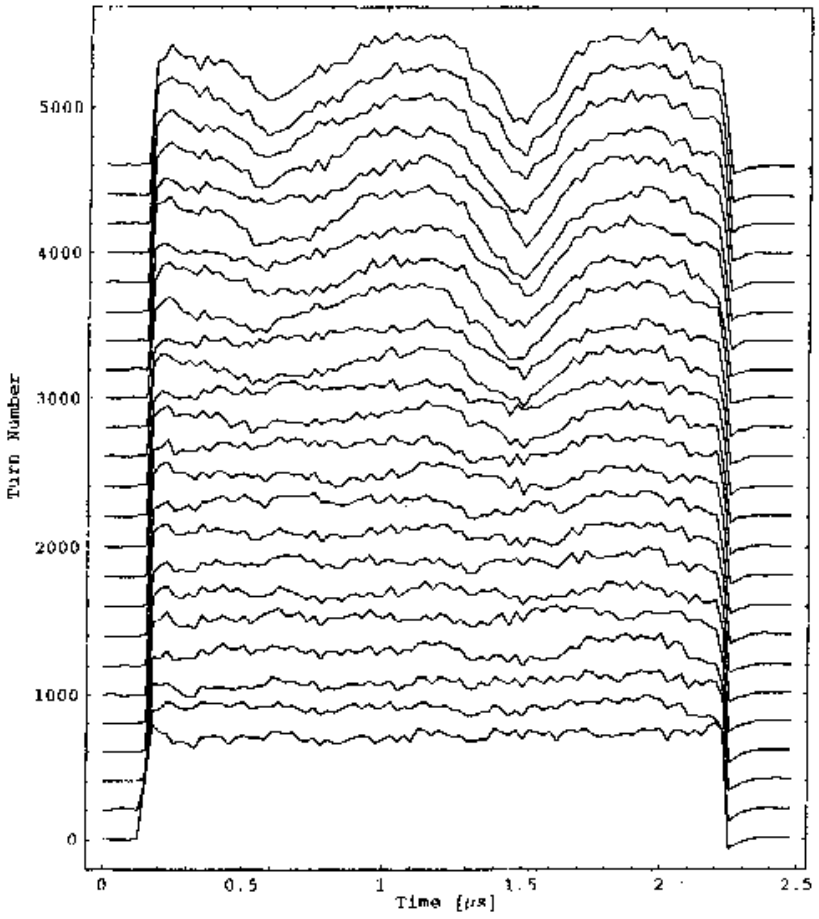

Fig. 2: Beam loss due to transient beam loading in the CERN SPS (proton beam for LHC, 0.5×10^{11} protons per bunch in one batch of 81 bunches). The beam consists of one batch filling 1/11 th ring ($\approx 2\mu$s). The cavity filling time is 800 ns (much shorter than the revolution period 23 μs). Each trace shows the envelope of the bunch intensity along the batch. The bottom trace shows the first turn (26 GeV/c). Traces are separated vertically by 200 turns. The capture voltage is 550 kV. No acceleration. A modulation of the bunch intensity develops along the batch after about 2000 turns.

1.2 Longitudinal instabilities

This Section is a brief qualitative presentation of the longitudinal instabilities caused by cavity impedance at its fundamental resonance. A complete theory can be found in Ref. [10]. At high beam current, the interaction between a bunched beam and an RF cavity can lead to longitudinal instabilities. This process can be broken down into three steps:

- Step 1: when crossing the cavity, the beam induces an electro-magnetic wave called wakefield.
- Step 2: this wakefield modifies the accelerating voltage seen by the beam (phase of V_t when the beam crosses the cavity and bucket height) and acts in return on the current profile along the beam.
- Step 3: this modifies the wakefield created on the next crossing.

If the gain and phase shift of the above natural beam/cavity feedback is unfavourable, instability will grow. The bunches start a longitudinal oscillation at the synchrotron frequency f_s (Fig. 3).

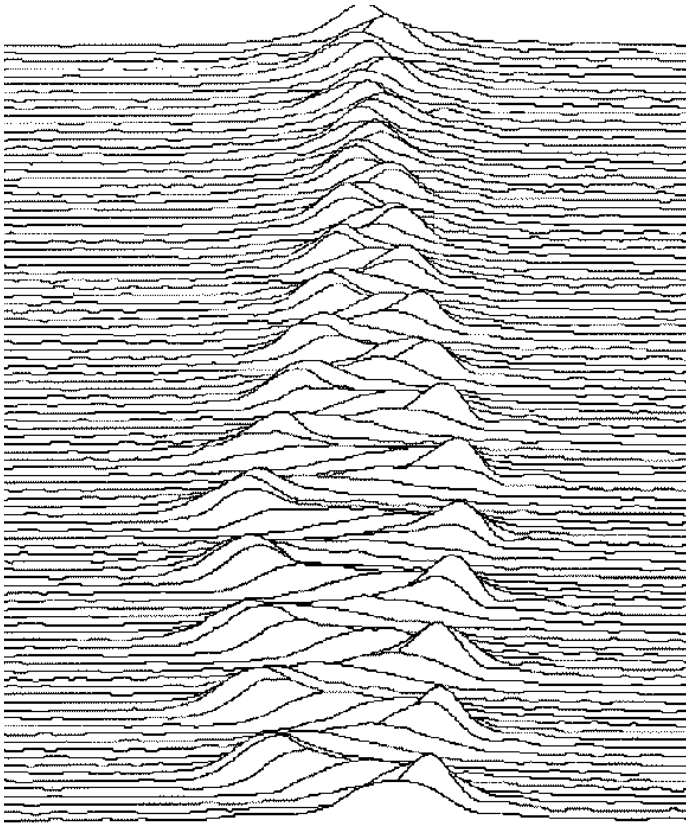

Fig. 3: Mountain range display of a positron bunch in the CERN SPS, from 100 ms after injection (3.8 GeV/c, bottom trace) to 122 ms after injection (4.2 GeV/c, top trace). Time goes from bottom to top. Ten turns (0.23 ms) between traces. Horizontal window = 10 ns. $f_s = 430$ Hz (1 synchrotron period = 100 turns = 10 traces). The amplitude of the dipole oscillation rises to 3 ns maximum. The RF consists of a fixed 0.750 MV at 100 MHz, plus a second harmonic at 200 MHz whose amplitude is zero on the bottom trace, and rises to 0.7 MV on the last trace. The oscillation is here excited by an external cause. Its amplitude (in ns) decreases as the acceleration proceeds thanks to the addition of the higher frequency RF and natural damping (synchrotron radiation). A strong quadrupole oscillation is also present.

In the presence of dipolar motion the bunches create a beam current that is periodic at f_{rev} with a phase modulation at the synchrotron frequency f_s. In Section 4.1 (Eq. (78)) we show that its spectral power is concentrated at the frequencies

$$f = f_{rf} \pm n f_{rev} \pm m f_s \tag{8}$$

with dominant sidebands at $m = 1$. Pure dipole oscillation goes without change in the bunch shape. But the cavity impedance can excite higher order shape oscillations of the bunch. Figure 4 shows a quadrupole oscillation: there is (almost) no motion of the centre of the charge distribution but the second order moment (related to the bunch length) oscillates at twice the synchrotron frequency. Higher order

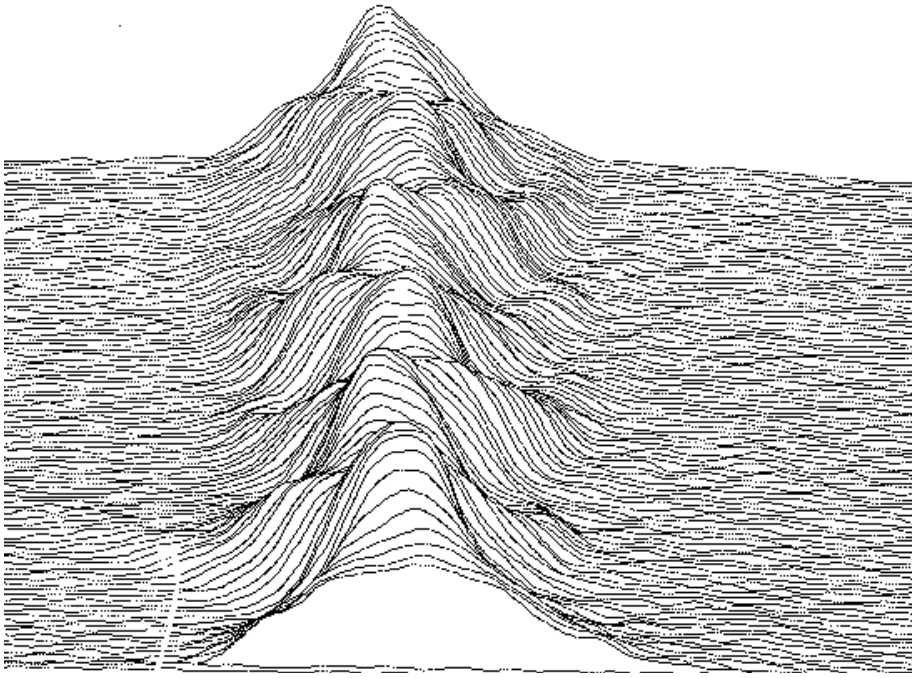

Fig. 4: Quadrupole oscillation (plus a small dipole oscillation) of a proton bunch at injection from the CPS into the CERN SPS (26 GeV/c, 200 MHz RF). The first non-zero trace at the bottom shows the first turn of the injected beam. Horizontal window = 10 ns. Proton beam for the LHC. Note that this oscillation is not due to an instability but to an RF mismatch.

modes can also be excited: sextupole, octopole, etc. [10]. Several modes can occur at the same time. The oscillation shown in Fig. 3 is the superpositions of a dipole and a quadrupole mode. These oscillations will sample the cavity impedance at the frequencies (Eq. (79))

$$f = f_{rf} \pm n f_{rev} \pm m f_s \tag{9}$$

and the low-level system must thus reduce the apparent cavity impedance there.[3] Comparing the above equation to Eq. (7) we notice that it is more difficult to prevent the longitudinal instabilities than to correct the beam loading: the synchrotron frequency f_s is typically much smaller than the revolution frequency f_{rev} and a complete prevention of the instabilities calls for a reduction of the cavity impedance at closely spaced frequencies around each revolution frequency line. Fortunately, higher order modes of oscillation are not as easily excited by cavity impedance (at the fundamental frequency) due to the form factor and are usually damped by the ever-present synchrotron frequency spread, so impedance reduction is hopefully only required on the first few synchrotron sidebands [10].

[3] The fundamental resonance of the RF cavity is most efficient at driving mode m, when the bunch width is comparable to m times $T_{rf}/2$ [10]. The quadrupole mode (or even higher modes) can thus be excited by the cavity fundamental with a full bucket, as is often the case at injection into a hadron machine [11]. In lepton machines the bunches are typically much shorter than the bucket width and the cavity will mainly excite the dipole mode. Higher order modes can always be excited by the undesired high-frequency resonances in the cavity or elsewhere in the beam pipe. The low-level systems presented here have no action on these external causes.

2 CURES

2.1 Feedforward

Method: Figure 5 shows the feedforward technique: the beam current I_b measured by a pick-up is filtered by a Bandpass Filter (BPF) centred at the centre frequency of the cavity response f_0 and fed into the feedforward filter. The feedforward transfer function H_{opt} corrects the generator drive such that the resulting generator current I_g^{comp} gives a cavity voltage V_g equal but opposite to V_b.

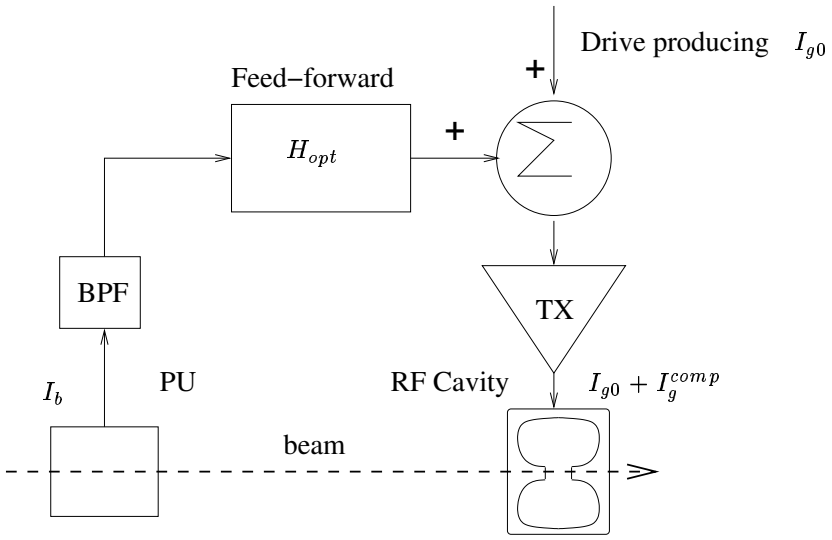

Fig. 5: Block diagram of the feedforward method

From Eq. (1) we get

$$Z_g I_g^{comp} = -Z_b I_b \ . \tag{10}$$

For a standing-wave cavity $Z_g = Z_b$ so H_{opt} is a constant gain. In the case of a travelling-wave cavity perfect compensation is impossible because Z_g is zero at frequencies where Z_b is non-zero [7]. But partial correction is possible with the help of a more complex filtering function H_{opt} [12] (see Section 3.4). In practice the delays in the cables and in the electronics do not permit measurement of the beam current and correction of the cavity voltage in the same turn. An intentional delay of one full turn is thus implemented in the chain.[4] **Advantage:** The feedforward loop is closed through the beam, but from the electronics point of view it is an open-loop system. As such it is not limited by the time constant inherent in a closed-loop system such as the long delay feedback (Section 2.3). At injection the beam current is measured on the first turn and the full compensation can be applied on the second turn. This makes locking of the beam phase loop possible at injection, even with high beam currents.
Limitation: The drawback is that this open-loop system is very sensitive to gain and phase drifts of the RF generator. This limits the long-term performance of the method. It is also difficult to set up for a varying RF frequency (acceleration ramp) because the pick-up to cavity delay must be varied continuously. A phase error of 5° at the RF frequency will degrade a perfect beam loading compensation to only 91% compensation ($2\sin 2.5^\circ = 0.09$).
Example: Figure 6 shows results obtained with the feedforward system installed on the 200 MHz travelling-wave cavities in the CERN SPS [12],[13].

[4] The synchrotron frequency f_s is typically much smaller than the revolution frequency f_{rev}. The current profile along the beam therefore varies very little in one turn.

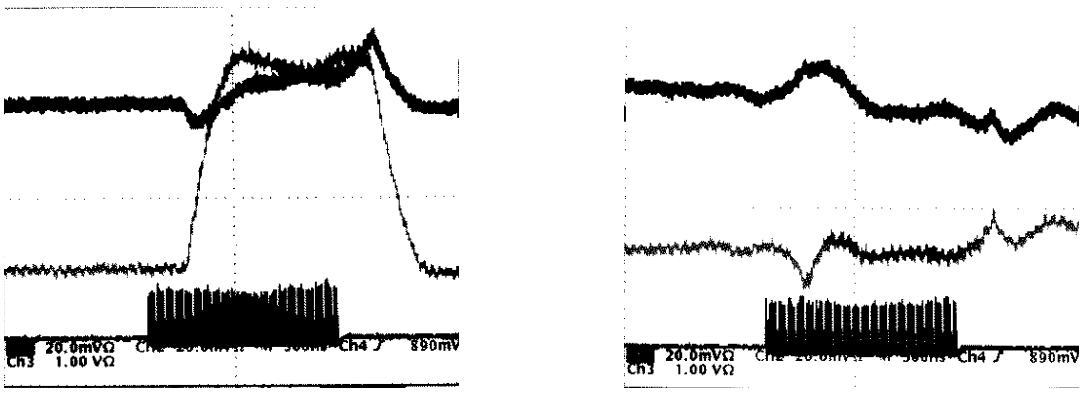

Fig. 6: Transient beam loading voltage measured in the CERN SPS 200 MHz TWC cavities with the LHC beam (1 batch = 81 bunches spaced by 25 ns, 4×10^{12} protons total, 500 ns/div). Left: feedforward OFF. Right: feedforward ON. The bottom trace is the beam current I_b measured with a wide-band pick-up, the upper two traces are the I and Q components (AC coupled) obtained by demodulating the total cavity voltage V_t with a Local Oscillator (LO) at the RF frequency (see Section 3.1 for details on I/Q demodulation). $I^2 + Q^2$ measures the power of the cavity voltage error.

2.2 RF feedback

Method: RF feedback around an amplifier is generally used to reduce the effects of the drifts in gain and phase encountered with power amplifiers. It also reduces distortion by flattening their response: this closed loop makes the overall response substantially independent of the response of the amplifier itself [14]. We can use it here to reduce the influence of the beam-induced voltage by effectively reducing cavity impedance. The principle is shown in Fig. 7. A probe measures the total accelerating voltage in the cavity V_t. It is compared to the desired voltage V_{ref} and the error is used to regulate the drive of the power amplifier. The RF feedback is a closed-loop system. As such it is relatively insensitive to the

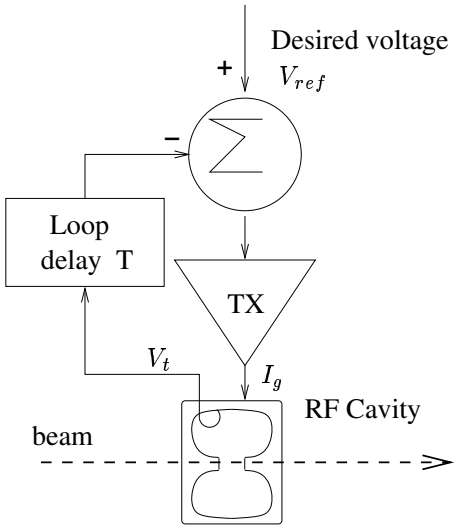

Fig. 7: Block diagram of the RF feedback method

small drifts of the gain and phase of the power generator. It is very easy to implement on a single-cell cavity. RF feedback on a multi-cell cavity is much more complex: a cavity consisting of N identical cells coupled together will show N closely spaced resonances corresponding to its N modes (phase shift of $2\pi n/N$ between cells). Since the coupling between two adjacent cells is slightly different for the

different modes, the phase shift from I_g to V_t will be different. A filter must be placed in the probe signal to adjust the open loop gain and phase independently for each resonance within the bandwidth of the system [15]. Otherwise the undesired resonances will make the loop unstable even though the corresponding field in the cavity may be harmless to the beam.

Limitation: The unavoidable loop delay T clearly appears in Fig. 7 because it limits the achievable impedance reduction. Figure 8 shows a block diagram of the RF feedback in the Laplace domain: G is the gain of the return path, A is the amplifier gain and e^{-Ts} is the delay operator.

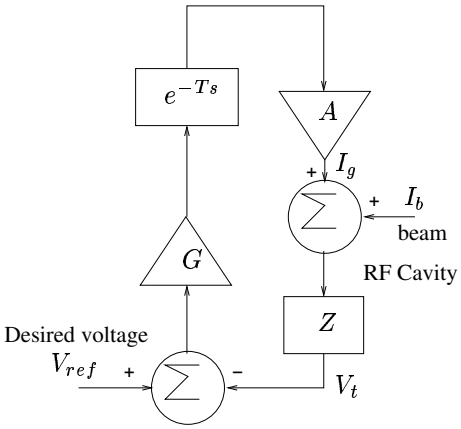

Fig. 8: Block diagram of the RF feedback in the Laplace domain

The beam current I_b is an added perturbation at the input of the standing-wave cavity of impedance $Z_g(\omega) = Z_b(\omega) = Z(\omega)$. The overall delay in the loop is T. It is the sum of the cable delay and the delay in the electronics (inversely proportional to the bandwidth of the amplifier).

Near its resonant frequency ω_0 a standing-wave cavity can be represented as an RLC circuit

$$Z(\omega) = \frac{R}{1 + j2Q\frac{\Delta\omega}{\omega_0}} \tag{11}$$

where Q is the quality factor and $\Delta\omega = \omega - \omega_0$. With the feedback loop closed, the beam loading voltage is

$$V_b(\omega) = \frac{Z(\omega)}{1 + GAe^{-jT\Delta\omega}Z(\omega)} I_b(\omega) \ . \tag{12}$$

In the above we have assumed that the phase of the return path has been adjusted so that the phase shift of the total loop is zero degrees at the frequency ω_0. Therefore a large open loop gain GAR leads to a good impedance reduction. The stability of the loop imposes a limit: outside the ω_0/Q band the cavity is purely reactive

$$Z(\omega) \approx \frac{R}{j2Q\frac{\Delta\omega}{\omega_0}} \ . \tag{13}$$

The phase shift is thus $-\frac{\pi}{2}$. A classic indicator of stability for a feedback loop is the phase margin, defined as the amount by which the phase of the open loop response exceeds $-\pi$ when the modulus of its gain is one [16]. To keep a phase margin of $\frac{\pi}{4}$, the open loop gain must have decreased to 1 when the delay T has added an extra $-\frac{\pi}{4}$ phase shift, that is at $\Delta\omega = \frac{\pi}{4T}$

$$GA|Z(\frac{\pi}{4T})| \leq 1 \tag{14}$$

$$GAR \leq \frac{\pi}{2}\frac{Q}{\omega_0}\frac{1}{T} = G_{max}AR \ . \tag{15}$$

Let $Z_{fbk}(\omega)$ be the apparent cavity impedance with the feedback loop. From Eq. (12) we get

$$Z_{fbk}(\omega) = V_b(\omega)/I_b(\omega) = \frac{Z(\omega)}{1 + GAe^{-jT\Delta\omega}Z(\omega)} \ . \tag{16}$$

Using the dimensionless variable $x = \Delta\omega T$ and the value of $G_{max}A$ given by Eq. (15) we can rewrite $Z_{fbk}(\omega)$ as

$$Z_{fbk}(\omega) = \frac{\frac{1}{GA}}{\frac{1}{GAR} + e^{-jx} + j\frac{4}{\pi}\frac{G_{max}A}{GA}x} \ . \tag{17}$$

At resonance, we have

$$Z_{fbk}(\omega_0) = \frac{R}{1 + GAR} \ . \tag{18}$$

A reduction of the apparent cavity impedance is thus possible only if GAR is larger than 1. Assuming that $GAR >> 1$, the first term in the denominator of Eq. (17) can be neglected and Z_{fbk} can be plotted as a function of $x = \Delta\omega T$ for different values of G (Fig. 9). For $G = G_{max}$ ($k = 1$), the frequency response presents a 3 dB overshoot on the edges of the passband. This is avoided by reducing the gain to $0.7G_{max}$ ($k = 0.7$). This flattens the frequency response in the passband, but the achieved value of the impedance at resonance ($x = 0$) is 3 dB larger. The figure also shows the effect of increasing the feedback gain above G_{max} ($k = 1.3$): the strong overshoot in the frequency response indicates that we are approaching the instability limit.

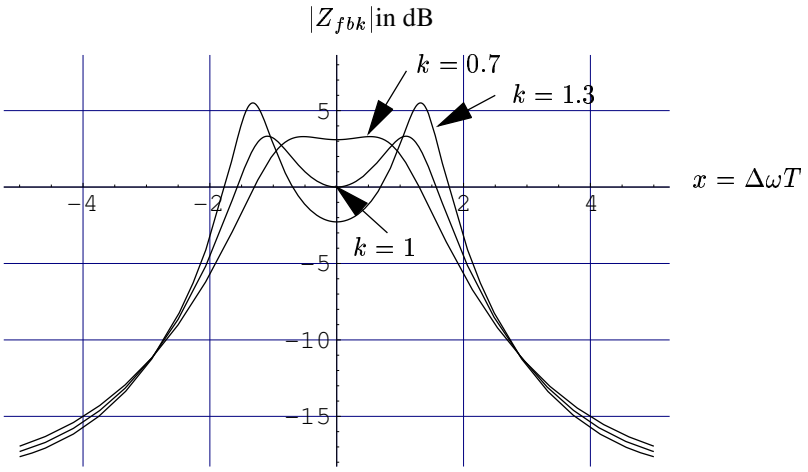

Fig. 9: Modulus of the apparent cavity impedance for three values of the normalized feedback gain $k = G/G_{max}$ as a function of $x = \Delta\omega T$

The minimum achievable value for the apparent cavity impedance at resonance ($\Delta\omega = 0$) is [5]

$$R_{min} = \frac{R}{1 + G_{max}AR} \approx \frac{2}{\pi}\frac{R}{Q}\omega_0 T \ . \tag{19}$$

The ultimate performance depends only on the cavity geometry R/Q and the loop delay T. The lesson is that the delay must be kept short, i.e. a broadband amplifier located as close to the cavity as possible. The achievable bandwidth is a function of the loop delay only. For the flat frequency response ($G = 0.7G_{max}$) we get

$$\Delta\omega_{-3dB} \approx \frac{1.3}{T} \ . \tag{20}$$

[5] If the flat frequency response ($G = 0.7G_{max}$) is preferred, the impedance will be 3 dB higher.

Examples: The present RF system of the storage ring at the Synchrotron Radiation Research Center (SRRC) in Taiwan consists of two independent and identical chains [17]: a 60 kW klystron is connected to a warm pill-box cavity via a circulator (Doris cavity, 500 MHz centre frequency, $Q_L = 14500$). It has been decided to replace this system by a single superconducting cavity [18]. As a first step in this upgrade project, an RF feedback has been tested on the warm cavities. Figure 10 shows the apparent impedance Z_{fbk} with a feedback gain $GAR = 11.2$ (or 21 dB). This measurement is obtained by sweeping the frequency of the klystron drive and recording the amplitude of the cavity voltage: it is the transfer function from V_{ref} to V_t in Fig. 8 and it is proportional to $|Z_{fbk}|$, with a proportionality factor equal to GAe^{-Ts}. The most important parameter is the loop delay T, here equal to 450 ns, including the 150 ns group delay of the klystron. From Eq. (15) we derive the maximal loop gain $G_{max}AR = \frac{\pi}{2}\frac{Q}{\omega_0}\frac{1}{T} = 16.111$. The flat response of Fig. 9 should thus be obtained with a 30% lower gain, that is $GAR = 11.28$, in good agreement with the experiment ($GAR = 11.2$). Eq. (20) predicts a bandwidth of 920 kHz (two sided), that is a Q equal to 543. This again is in good agreement with the measured Q of 562.29. On this system, the RF feedback reduces the apparent cavity impedance by 11.2.

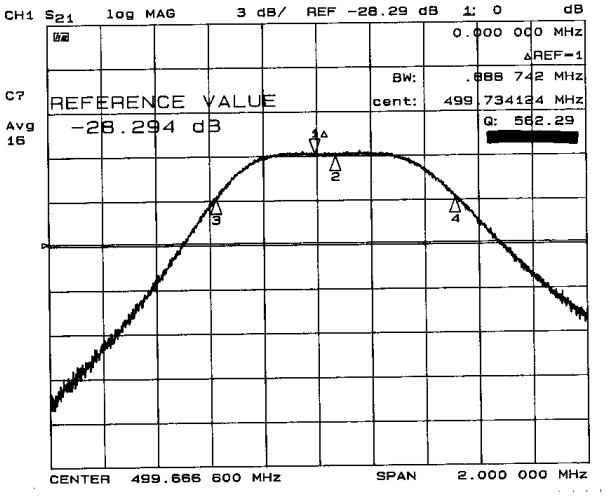

Fig. 10: Apparent cavity impedance of the SRRC RF system with an open-loop gain $GAR = 11.2$

Another example is the RF feedback installed in the CERN PS on the 40 MHz cavities [19]. A tube amplifier is connected to the warm re-entrant cavity ($Q_L = 10000$) via a short (60 cm) coaxial line. The total loop delay, including the amplifier group delay, is only 220 ns. The maximal loop gain is derived from Eq. (15) and we get $G_{max}AR = \frac{\pi}{2}\frac{Q}{\omega_0}\frac{1}{T} = 284$ or 49 dB. The system is operated with a gain of 43 dB, i.e. half the maximal value. This feedback reduces the cavity impedance at the resonance by a factor of 140.

2.3 Long delay feedback

Method: In the previous section it was concluded that the amplifier was best located close to the cavity for a good reduction of beam-induced effects with RF feedback. The amplifier is, however, often placed outside the accelerator tunnel in order to ease maintenance and reduce down-time in case of failure. In that case, impedance reduction is still possible if the cavity is not narrow band. It is first noted that beam loading needs only to be compensated at the frequencies (Eq. (69))

$$f = f_{rf} \pm n f_{rev} \tag{21}$$

while the prevention of instability calls for a reduction of the cavity impedance at the frequencies (Eq. (79))

$$f = f_{rf} \pm n f_{rev} \pm m f_s \ . \tag{22}$$

The synchrotron frequency f_s is typically much smaller than the revolution frequency f_{rev}. We thus conclude that our feedback needs gain around the revolution frequency lines only, and in a bandwidth sufficient to include the first synchrotron sidebands. The long delay in the loop will not affect the phase at these frequencies *if it is an exact multiple of the revolution period* T_{rev}. Figure 11 shows the block diagram of the long delay feedback [20]: the cable and electronics delay T is extended to one full turn in the feedback loop. The phase of the correction is thus $0°$ at multiples of f_{rev}. It is wrong by $180°$ at the centre, between two revolution frequency lines. The open-loop gain must have dropped to a value smaller than 1 in order to maintain the stability of the loop there. The frequency response of the filter (including the delay of one turn) is shown in Fig. 12. It is similar to a comb with high gain plus zero phase shift on the revolution frequency lines, and low gain plus $180°$ phase shift in between.

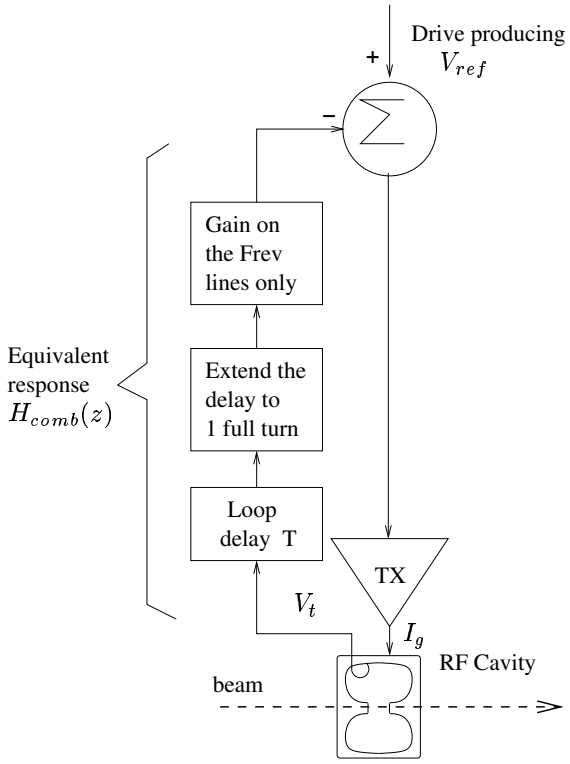

Fig. 11: Block diagram of the long delay feedback method

The comb filter is easily implemented with digital technology.[6] The clock f_c, used to sample the input signal, is obtained by division of the RF frequency f_{rf} so that

$$f_c = M f_{rev} \ . \tag{23}$$

Inside the digital filter a delay of M clock periods therefore implements an exact revolution period, i.e. a delay of one turn. And this remains true with the RF frequency varying during acceleration. The transfer function of the filter[7], including the one-turn delay, is

$$H_{comb}(z) = G \frac{1-a}{1-az^{-M}} z^{-M} \tag{24}$$

[6] Section 4.2 presents the basics of digital filters.
[7] Transfer functions and z-transforms are explained in Section 4.2.3.

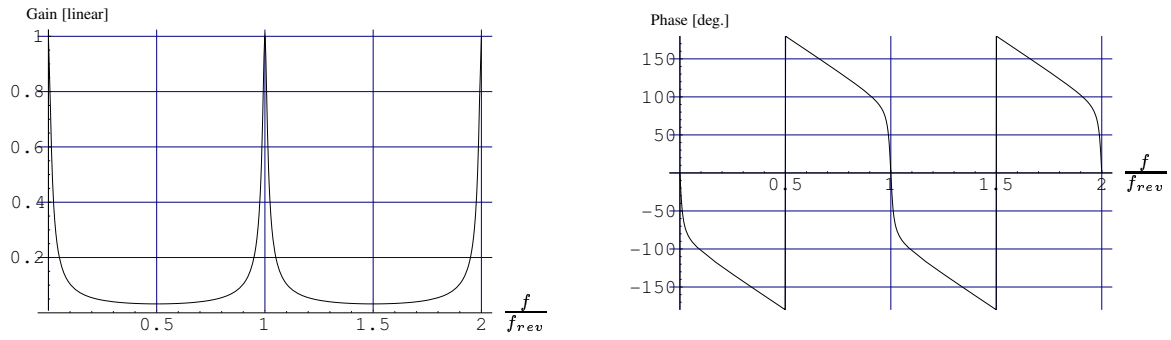

Fig. 12: Frequency response of the revolution frequency comb $H_{comb}(z)$ ($a = 15/16$). Although the phase slips by π halfway between two successive revolution frequency lines, the gain has decreased enough to maintain stability.

Limitation: The parameter a fixes the bandwidth of the filter around each revolution frequency line

$$\Delta f_{-3db} \approx \frac{1}{2\pi M}(1-a)f_c = \frac{1}{2\pi}(1-a)f_{rev} \tag{25}$$

where the approximation holds for a value of a close to 1 (small passband). It also fixes the reduction of the transfer function at half distance between passbands (where $z^M = -1$)

$$H_{min} = G\frac{1-a}{1+a} \approx G(1-a)/2 \ . \tag{26}$$

To maintain reasonable stability, the open-loop gain must be significantly below 1 when the phase is π [16]. It is usual to impose a gain margin of 10 dB ($\approx 1/3$ linear). We get

$$G(1-a)/2 \leq 1/3 \ . \tag{27}$$

The impedance reduction in the passbands is equal to G. Equation (27) shows that a large reduction is possible with a value of the parameter a close to 1. However, Eq. (25) shows that the passband then vanishes. This may not be a problem for the compensation of the transient beam loading (except at injection, see below) but a bandwidth covering at least the first synchrotron sidebands is necessary if the long delay feedback is to act against longitudinal instabilities. The optimal value of a therefore depends on the synchrotron tune $Q_s = f_s/f_{rev}$. The smaller the tune, the more efficient the long delay feedback can be.

Another solution is to use a dedicated system to compensate the beam loading with a transfer function $H_{comb}(z)$ given by Eq. (24) and a parameter a very close to 1. A second system with a different transfer function having gain only on the first synchrotron sidebands (double peaked comb filter) will fight against the dipole mode longitudinal instabilities [21]. An implementation of this double-peaked comb filter is presented in Section 3.3. The transient response at injection may also limit the value of a even though we are only concerned by beam loading. The time needed at injection to compensate beam loading is inversely proportional to the bandwidth Δf_{-3db}. If this time is too long compared to the synchrotron frequency, the mismatch between the bunch emittance and the beam loaded bucket will last sufficiently to create loss of particles. It may even prevent the locking of the beam phase loop. This is unlike a feedforward compensation, which corrects the transient beam loading on the second turn (Section 2.1).

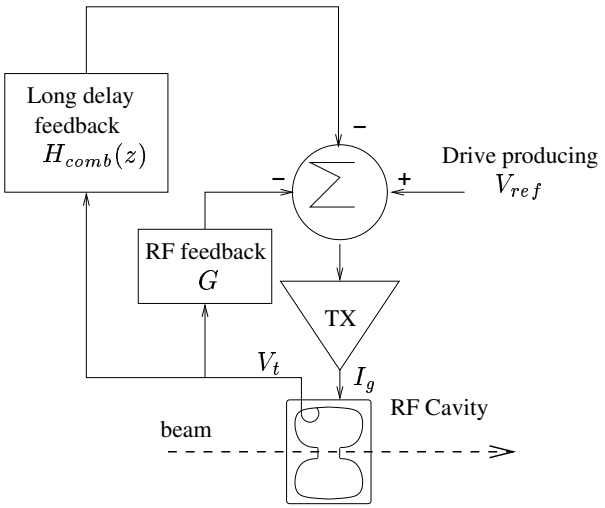

Fig. 13: Long delay feedback and RF feedback on a high Q cavity

Remark: In this section we have not included the cavity impedance Z_g in the open-loop response. If the long delay feedback is used alone on a narrow band cavity, the open-loop response will be modified by the cavity impedance Z_g and the gain and phase shift will not be equal on all frequency lines. Good and stable overall performances will not be easily achieved. In the case of a high Q cavity the long delay feedback is best used in conjunction with an RF feedback (Section 2.2) that effectively transforms the narrowband impedance of the cavity into a broadband response. The long delay feedback is then further reducing the impedance on the revolution frequency lines, in the band where the RF feedback has flattened the cavity response. The whole system is shown in Fig. 13. The drive producing V_{ref} will be generated by the slow cavity amplitude loop and the beam phase loop of the low-level system, which, respectively, try to keep the modulus and phase of V_t at the desired values. They only act on the line at the exact RF frequency f_{rf}. In order to avoid the interference of the long delay feedback with these loops the passband, centred on f_{rf}, is best cancelled in the comb response. An elegant way of rejecting this band is presented in Section 3.1. This combination of RF feedback and long delay feedback is used on the 352 MHz superconducting cavities in the CERN SPS. In the CERN PS it is used on the ferrite cavities [22]. Travelling-wave cavities are different [7]: their bandwidth is large enough for a long delay feedback alone, without RF feedback. As seen from the generator they present a matched load even in the presence of beam loading, so no tuning is needed. The two impedances Z_g and Z_b are, however, different and vary in a complex fashion with frequency. When designing the long delay feedback it is therefore necessary to consider the Z_g impedance. The 200 MHz travelling-wave cavities of the CERN SPS are equipped with a long delay feedback whose response includes a Post Filter $H_{pf}(z)$ in series with the classic comb $H_{comb}(z)$. The Post Filter compensates the effects of the cavity response Z_g [12],[13].

2.4 Bunch-by-bunch feedback

Method: The last method does not involve the RF cavity, neither as a probe nor acting on the beam. It will have no effect on beam loading: its goal is to prevent dipole mode instabilities. It is broadly used in high-intensity lepton machines (synchrotron light sources). The principle is shown in Fig. 14. The longitudinal oscillation of each bunch (dipole mode, shown in Fig. 3) is measured independently. The corresponding signals are processed in parallel in order to generate a dedicated longitudinal kick on each bunch. The kicks are adjusted so that they reduce the amplitude of the oscillations [23].

Let $\hat{\phi}_k$ be the phase of the RF when the centre of charge of the bunch of index k crosses the pick-up. We saw in the first part of these two lectures that, in the presence of a modulation of the RF

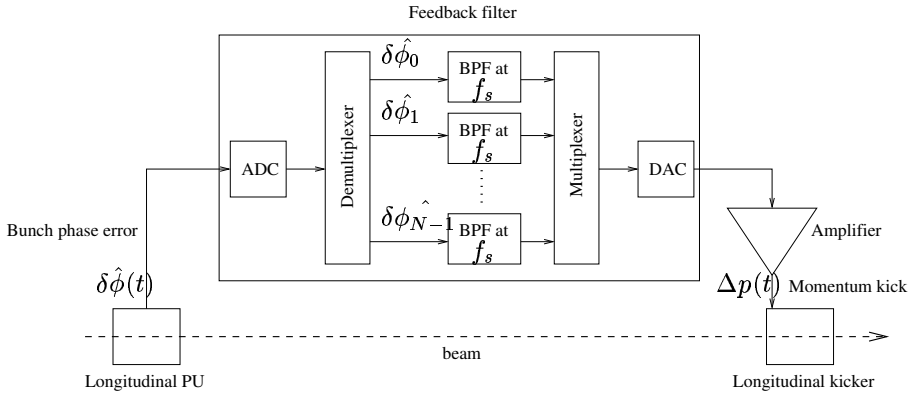

Fig. 14: Block diagram of the bunch-by-bunch feedback method (N bunches)

frequency $\delta\omega_{rf}$, the longitudinal motion of the centre of charge of each bunch obeys the equation

$$\frac{d^2\delta\hat{\phi}_k}{dt^2} + \Omega_s^2\,\delta\hat{\phi}_k = \frac{d\delta\omega_{rf}}{dt} \ , \tag{28}$$

where $\delta\hat{\phi}_k = \hat{\phi}_k - \phi_s$. ϕ_s is the stable phase and $\Omega_s = 2\pi f_s$ [8]. A similar equation can be written if the driving term is a small momentum kick Δp

$$\frac{d^2\delta\hat{\phi}_k}{dt^2} + \Omega_s^2\,\delta\hat{\phi}_k = \frac{\eta h\,(\beta c)^2}{2\pi R_0^2 p}\Delta p \ , \tag{29}$$

where η is the slippage factor related to the energy

$$\eta = \frac{1}{\gamma_t^2} - \frac{1}{\gamma^2} \tag{30}$$

with $\gamma = E/E_0$ (E_0 being the rest energy), γ_t is the value of γ at the transition energy, h is the harmonic number, β is the normalized velocity ($\beta = v/c$), c being the speed of light, $2\pi R_0$ is the machine circumference, and p is the average momentum of the bunch. To obtain damping, we must introduce a term proportional to the first derivative of $\delta\hat{\phi}_k$. This is achieved if we make the momentum kick Δp_k proportional to the *derivative* of the phase error $\delta\hat{\phi}_k$

$$\Delta p_k = a\frac{d\delta\hat{\phi}_k}{dt} \ . \tag{31}$$

The equation of motion then becomes

$$\frac{d^2\delta\hat{\phi}_k}{dt^2} + \Omega_s^2\,\delta\hat{\phi}_k = -\alpha_f\frac{d\delta\hat{\phi}_k}{dt} \tag{32}$$

or

$$\frac{d^2\delta\hat{\phi}_k}{dt^2} + \alpha_f\frac{d\delta\hat{\phi}_k}{dt} + \Omega_s^2\,\delta\hat{\phi}_k = 0 \tag{33}$$

with

$$\alpha_f = -a\frac{\eta h\,(\beta c)^2}{2\pi R_0^2 p} \ . \tag{34}$$

Taking as initial conditions for bunch k a non-zero phase error and a zero frequency error, the feedback would make the phase error go to zero with a time constant $\tau_f = 2/\alpha_f$ (if $\alpha_f \leq 2\Omega_s$) in the

absence of the instability mechanism. The dipole mode instability can be modelled as an additional term, proportional to the first derivative of $\delta\hat{\phi}_k$, with a negative factor

$$\frac{d^2\delta\hat{\phi}_k}{dt^2} + (\alpha_f - \alpha_i)\frac{d\delta\hat{\phi}_k}{dt} + \Omega_s^2\,\delta\hat{\phi}_k = 0 \quad . \tag{35}$$

If the damping rate $\alpha_f/2$ is larger than the instability growth rate $\alpha_i/2$ the feedback will keep the centre of charge of the bunches stable: no dipole mode instability will occur.
Advantage: Bunch-by-bunch feedback fights against *all* sources of dipole mode instabilities no matter where they are located in the machine.
Limitation: Beam loading is not compensated. There is no effect on the instabilities of modes higher than one either (quadrupole mode, sextupole mode, etc.).
Remark: If the feedback filter has the characteristic of a pure differentiator, the high-frequency noise in the pick-up signal will severely degrade the performance (emittance blow-up and reduction of life-time). Since the bunches oscillate at the synchrotron frequency, we can use a bandpass centred at the synchrotron frequency f_s and implement a phase shift of $\pi/2$. For the synchrotron oscillation the band-pass will act as a differentiator and the noise present outside the band will be rejected. Rejection of the DC component is also desirable. We actually measure $\hat{\phi}_k$ but must provide a kick proportional to the derivative of $\delta\hat{\phi}_k = \hat{\phi}_k - \phi_s$. The stable phase offset ϕ_s appears as a DC component in the signal $\hat{\phi}_k$ and the filter will automatically remove it. Other offsets in the acquisition electronics will also have no consequence for the kick. A large bandwidth is required for the acquisition of the phase of each bunch and for the generation of the kicks. Assuming N evenly spaced bunches, the minimum bandwidth is half the bunch frequency $Nf_{rev}/2$ (250 MHz for PEPII and for the Japanese b-Factory KEKB [24]). At first sight the requirement for processing power seems discouragingly high. The synchrotron frequency is, however, much smaller than the revolution frequency (small Q_s). For a given bunch, the momentum kick need not be recomputed at each turn. We can thus reduce the sampling rate (downsampling), i.e. acquire the phase of a bunch only every D turns, compute the corresponding kick, and apply the same kick on the bunch for the coming D turns (interpolation). It may be sufficient to process the signal ten times per synchrotron period instead of every turn. This would result in a large saving if Q_s is small. The downsampling/interpolation method is treated in detail in Section 3.6.

3 IMPLEMENTATION

3.1 Heterodyning

Motivation: The feedforward and long delay feedback imply complex signal processing in a relatively narrow band around the fundamental RF frequency f_{rf}. The bandwidth is rarely more than a few tens of MHz. It is limited by the bandwidth of the RF power chain. On the other hand, the RF frequency can be as high as 500 MHz. When using a digital filter, the processing bandwidth is typically limited to one third of the sampling frequency (see Section 4.2.1). If the RF frequency is low, the cavity signal can be sampled directly. This is the case in the long delay feedback of the CERN PS. The RF frequency remains below 10 MHz. The cavity voltage is sampled at $80f_{rev}$ (from 33.2 MHz at lower energy to 38.4 MHz at higher energy) [22]. Treatment is thus possible up to 13 MHz. The RF frequencies used in the CERN SPS are 200 MHz (travelling-wave cavities) and 352 MHz (superconducting standing-wave cavities). They are both equipped with a long delay feedback whose processing bandwidth covers about 13 MHz on each side of f_{rf}. The sampling frequency should thus be around 1 GHz (for the 352 MHz cavities) if the cavity signal was processed directly, while a more reasonable sampling rate of 40 MHz will be sufficient for the heterodyne system.
Method: The input signal $x(t)$ at the RF frequency is either the cavity voltage V_t (the long delay feedback of Fig. 11) or the beam current I_b (feedforward of Fig. 5). It is fed into the RF input of the I/Q demodulator (Fig. 15).

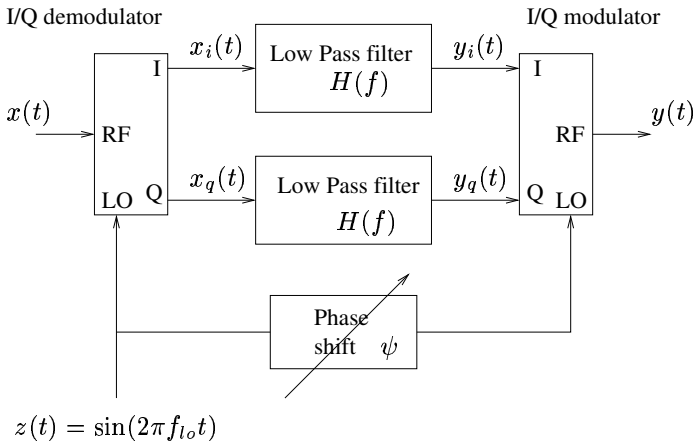

Fig. 15: Heterodyning: bandpass filtering at the RF frequency using two identical low pass filters

A pure sine wave $z(t)$ at frequency f_{lo} is fed into the LO input. The demodulator decomposes the signal $x(t)$ into two components that are in phase and in quadrature with $z(t)$, respectively

$$x(t) = x_i(t)\sin(2\pi f_{lo}t) + x_q(t)\cos(2\pi f_{lo}t) \tag{36}$$

with

$$x_i(t) = x(t)\sin(2\pi f_{lo}t) \tag{37}$$

$$x_q(t) = x(t)\cos(2\pi f_{lo}t) \ . \tag{38}$$

Let $X(f)$ be the Fourier transform of $x(t)$; the Fourier transforms of $x_i(t)$ and $x_q(t)$ are

$$X_i(f) = \frac{X(f - f_{lo}) - X(f + f_{lo})}{2j} \tag{39}$$

$$X_q(f) = \frac{X(f - f_{lo}) + X(f + f_{lo})}{2} \ . \tag{40}$$

The signals $x_i(t)$ and $x_q(t)$ are fed into two identical Low Pass Filters (LPF) whose outputs are $y_i(t)$ and $y_q(t)$

$$Y_i(f) = H(f)X_i(f) \tag{41}$$

$$Y_q(f) = H(f)X_q(f) \ . \tag{42}$$

The I/Q modulator receives on its LO input the sine wave at f_{lo} shifted in phase by ψ. It produces an RF output signal $y(t)$

$$y(t) = y_i(t)\sin(2\pi f_{lo}t + \psi) + y_q(t)\cos(2\pi f_{lo}t + \psi) \ . \tag{43}$$

The Fourier transform of $y(t)$ is

$$Y(f) = \frac{Y_i(f - f_{lo})e^{j\psi} - Y_i(f + f_{lo})e^{-j\psi}}{2j} + \frac{Y_q(f - f_{lo})e^{j\psi} + Y_q(f + f_{lo})e^{-j\psi}}{2} \ . \tag{44}$$

Now using Eqs. (39)–(42) in (44) we get, after simplifications

$$Y(f) = \frac{H(f - f_{lo})e^{j\psi} + H(f + f_{lo})e^{-j\psi}}{2}X(f) \ . \tag{45}$$

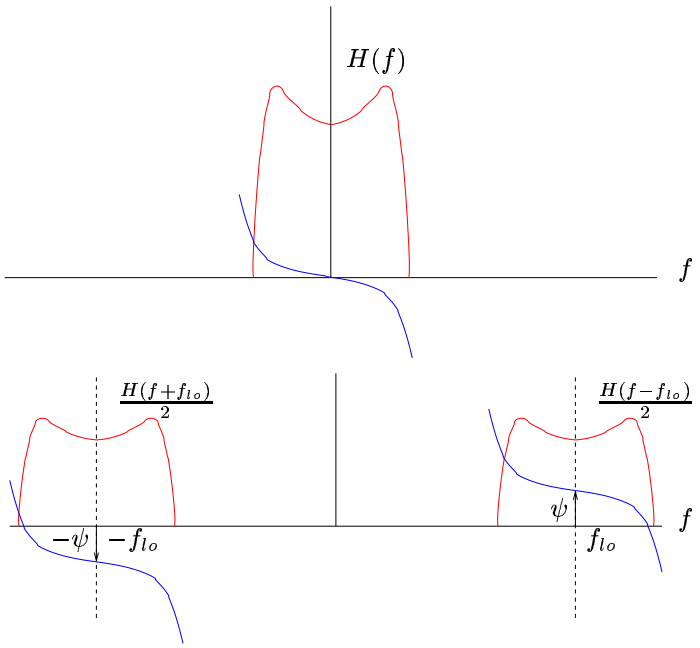

Fig. 16: Top: low pass model. Bottom: BPF implemented by the heterodyne system.

Figure 16 illustrates the above equation: the response at the top is the modulus and phase of the frequency response of the LPF $H(f)$. The response at the bottom is the filtering described by Eq. (45). The response $H(f - f_{lo})$ is identical to $H(f)$, but shifted by f_{lo}, while $H(f + f_{lo})$ is identical to $H(f)$ shifted by $-f_{lo}$. The result is a BPF centred at f_{lo} with an amplitude response, in the band, identical to the low pass model while the phase response is shifted by a constant value ψ. In a heterodyne implementation of the long delay feedback, $x(t)$ is the total cavity voltage V_t. The signal at the LO input of the modulator and demodulator is the RF frequency ($f_{lo} = f_{rf}$). The signals $x_i(t)$ and $x_q(t)$ are AC-coupled to the digital LPF $H_{comb}(z)$ (z-transform given by Eq. (24)). The AC coupling introduces a zero in the overall filtering at the exact frequency f_{rf} so that the long delay feedback does not interfere with the other low-level loops (cavity field amplitude loop, beam phase loop, and the possible cavity tuning loop). The clock frequency for the digital filters is a multiple of the revolution frequency. It is obtained by dividing the RF frequency. The leakage of the LO at the output of the modulator is not important. It introduces a small error in the cavity voltage at f_{rf}, but this is corrected by the other low-level loops. The phase shifter between the LO references fed into the demodulator and modulator is an easy way to finely adjust the phase of the output at the RF frequency. This implementation is used in the CERN SPS for the long delay feedback on the 200 MHz travelling-wave cavities. The digital filters are clocked at 40 MHz. The processing bandwidth is 13 MHz on each side of the RF frequency [12]. In a heterodyne implementation of the feedforward, the input $x(t)$ is the beam current after bandpass filtering to isolate the interesting band around the cavity centre frequency f_0. The signal at the LO inputs is a sine wave at f_0. The digital filters are DC coupled and implement the delay of one turn. Their clock frequency f_c must be a multiple of the revolution frequency so that the overall delay remains one full turn when the RF varies. In the case of a travelling-wave cavity the response H_{opt} is more complex (see Section 3.4). The leakage of the LO at the output of the modulator must be minimized because it is not at the RF frequency f_{rf} and will thus not be corrected by the low-level loops. The CERN SPS 200 MHz cavities are equipped with a feedforward working in tandem with a long delay feedback for the acceleration of the proton beam for the LHC. The RF frequency ramps from 200.264 MHz to 200.395 MHz. The cavity centre frequency (f_{lo}) is constant at 200.222 MHz. The digital filters are clocked at $1/10$ the RF frequency (20 MHz) [12].

In the range of frequencies used in particle accelerators, I/Q modulators and demodulators are readily available, for example:

- MIQ family from Mini-Circuits: RF/LO 9 MHz to 1.9 GHz, I/Q DC to 5 MHz, 40 dB sideband rejection.
- IM/ID family from Pulsar: RF/LO 10 MHz to 1.9 GHz, I/Q DC to 50 MHz, 30 dB sideband rejection.
- QM/SM family from Synergy: RF/LO 20 MHz to 1.9 GHz, I/Q DC to 50 MHz, 30 dB sideband rejection.

The heterodyne system of Fig. 15 implements an RF bandpass filter using two identical digital low pass filters. An alternative is shown in Fig. 17: instead of mixing the carrier frequency down to DC, we mix it down to an intermediate frequency (IF). In order to implement the IF bandpass filter $H_{if}(f)$ with digital technology, the IF output of the mixer must be sampled.

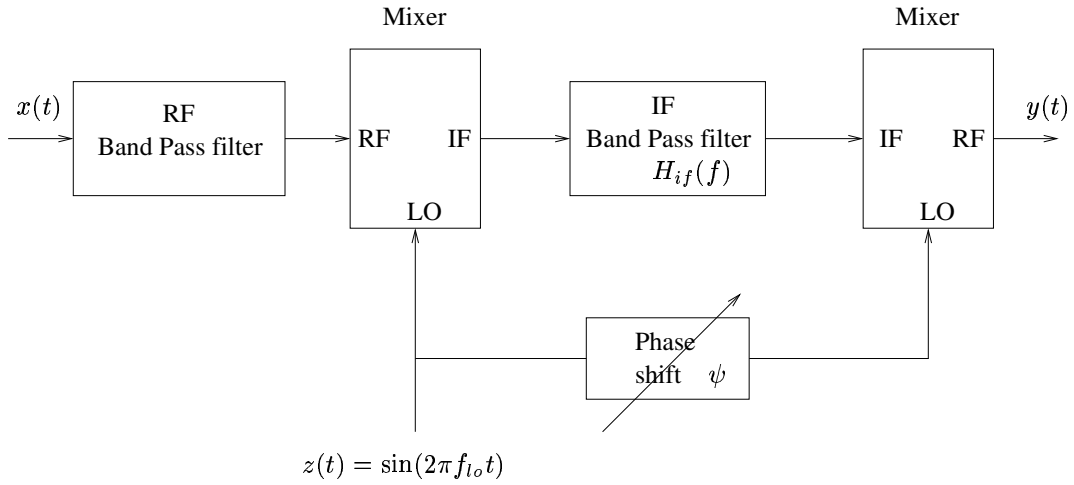

Fig. 17: Heterodyning: Bandpass filtering at the RF frequency using a BPF at the IF frequency

Let us take, as an example, the long delay feedback presented above. The desired bandwidth is 13 MHz on each side of the RF frequency at 200 MHz. The use of an IF frequency at 13 MHz, that is $f_{lo} = 187$ MHz, is not recommended because the noise in the band from 161 MHz to 187 MHz will create, at the mixer output, image signals corrupting the interesting signal band (from DC to 26 MHz). The RF bandpass filter centred at 200 MHz first selects the desired signal band with minimal distortion (from 187 MHz to 213 MHz), and ideally rejects the rest of the spectrum so no image signals are present at the mixer output. In practice, accepting 1 dB attenuation at the extreme of the signal band ($f_{rf} \pm$ 13 MHz), we will get 20 dB attenuation at $f_{rf} \pm 25$ MHz using an RF bandpass filter available from industry (of 4 to 5 sections).[8] The image at the mixer output will thus be rejected by 20 dB if we use an intermediate frequency at 25 MHz ($f_{lo} = 175$ MHz). At the input of the digital filter the signal will cover the band from 12 MHz to 38 MHz, implying a minimum sampling frequency of around 100 MHz, compared to the 40 MHz sufficient in the system of Fig. 15. This example shows that the system using the I/Q demodulator and modulator is less demanding on digital bandwidth. But two identical digital filters are needed while one is sufficient in the system of Fig. 17.

[8] Theoretical attenuation curves give, respectively, 22 dB and 29 dB attenuation at ± 25 MHz for a five section Butterworth and a four section 0.3 dB passband ripple Chebishev, with 1 dB attenuation at ± 13 MHz. In practice, given these passband specifications, companies designing custom RF filters (RLC Electronics for example) will propose a $4-5$ section filter providing 20 dB attenuation at ± 25 MHz.

In the following sections, we study heterodyne implementations of the feedforward and long delay feedback systems. One could also implement an RF feedback using an I/Q demodulator, followed by two digital low pass filters and an I/Q modulator, as shown in Fig. 15. The I and Q components of the desired accelerating voltage should be subtracted from the I and Q output of the demodulator (cavity voltage) so that the loop gives precise control of the accelerating voltage [25]. This architecture is used to linearize power amplifiers in mobile communication systems, where it is called Cartesian feedback [26]. For its application to an accelerator cavity, and if the main concern is beam loading, one must make sure that the digital processing does not contribute significantly to loop delay, thereby limiting the achievable impedance reduction. A variant of the I/Q feedback is proposed for the RF cavities of the synchrotron light source Soleil [27].

3.2 Analog-to-digital conversion and back

The performance (speed and resolution) of the Analog to Digital and Digital to Analog Converters (ADC and DAC) available on the market are sufficient for our applications. State-of-the-art examples are available from Analog Devices: 12-bit ADC at 105 Msamples/s (AD9432), 14-bit DAC at 125 Msamples/s (AD9754). Note that, with a 12-bit ADC covering the analog range of 1 V peak-to-peak, one Least Significant Bit (LSB) corresponds to 250 μV, i.e. -65 dBm. The noise in the RF signal plus eventual glitches on the power supply lines and noise because of poor grounding must stay below that level to take full advantage of the digital resolution. An efficient ground plane is essential to prevent the pollution of the sensitive analog signals by the noise due to the fast commutations of the digital circuitry. The designers of the ADC and DAC provide valuable information (see for example Ref. [28]).

3.3 Implementation using discrete numerical operators

Principle: The filtering realized by a digital filter can be represented as a difference equation relating the output sequence y_n to the input sequence x_n. In our applications n will be the time index: $x_n = x(nT_c)$, where T_c is the period of the clock ($T_c = 1/f_c$). For example, the output y_n of the comb filter H_{comb} (z-transform given in Eq. (24)) is related to the input x_n by the difference equation

$$y_n = a\, y_{n-M} + G(1-a)x_{n-M} \quad , \tag{46}$$

where the sequence x_{n-M} is obtained by delaying the sequence x_n by M samples. (Section 4.2.4 shows how to derive a transfer function from the difference equations.) A linear, constant-coefficients, difference equation can be implemented with only three basic operations: addition, delay and multiplication by a constant. The digital filter can thus be realized using three discrete components:

- Addition/Subtraction using 16-bit Arithmetic/Logical Units (ALU) such as IDT7381 (Integrated Device Technology) or L4C381 (Logic Devices). Operation in 15 ns.
- Delay of any depth using First-In First-Out memories (FIFO) such as CY7C42X5 (Cypress) or IDT722X5 (Integrated Device Technology). An 18-bit FIFO of variable depth up to 4 k words. Operation in 10 ns.
- Multiplication using 16×16-bit multipliers such as IDT7216 (Integrated Device Technology). Operation in 16 ns.

Example: As an example we present the double-peaked comb filter installed on the 352 MHz superconducting cavities of the CERN SPS. Four cavities providing a total accelerating voltage of 32 MV were installed to accelerate leptons to 22 GeV for LEP. They are equipped with a strong RF feedback (Section 2.2) that flattens the cavity response in a band of 1 MHz around 352 MHz [15]. The tetrode amplifiers are located next to the cavities and the overall loop delay is 500 ns. The SPS is also accelerating a high-intensity proton beam and the impedance of these cavities triggers longitudinal instability for this beam. Beam loading is not a problem. The ring is almost completely filled with bunches spaced by 5 ns, so that the beam current I_b has very little power around 352 MHz. An additional long delay

feedback was designed to further reduce the cavity impedance at the synchrotron sidebands of the revolution frequency lines. The two systems work together as shown in Fig. 13. The long delay feedback uses the heterodyne method shown in Fig. 17. The intermediate frequency (IF) is 4 MHz. The LO is thus at a 4 MHz offset from the centre frequency 352 MHz. The RF bandpass filter selects a band of ± 1.2 MHz around 352 MHz (1 dB bandwidth) and gives 13 dB attenuation at ± 4 MHz. Note that the interesting band is rather small (2.4 MHz). The digital filter is clocked at 20 MHz. It is a variant of the double-peaked comb filter proposed in Ref. [21]. The synchrotron tune is so small ($f_s \leq 1$ kHz, $f_{rev} \approx 43$ kHz) that the peaks on the sidebands can be realized by placing zeros and double poles on the revolution frequency lines. The z-transform is[9]

$$H_{sbd}(z) = \frac{1 - z^{-M}}{(1 - az^{-M})(1 - az^{-M})} z^{-M} \tag{47}$$

with $M = 462$, $f_c = 20$ MHz. The parameter a can be varied. As it gets closer to 1, the achievable impedance reduction increases but the peaks move closer to the revolution frequency lines. We use $a = 31/32$ or $a = 63/64$ providing an impedance reduction of 20 dB and 26 dB respectively on the synchrotron sidebands with the classic 10 dB gain margin. Figure 18 shows the modulus of the frequency response.

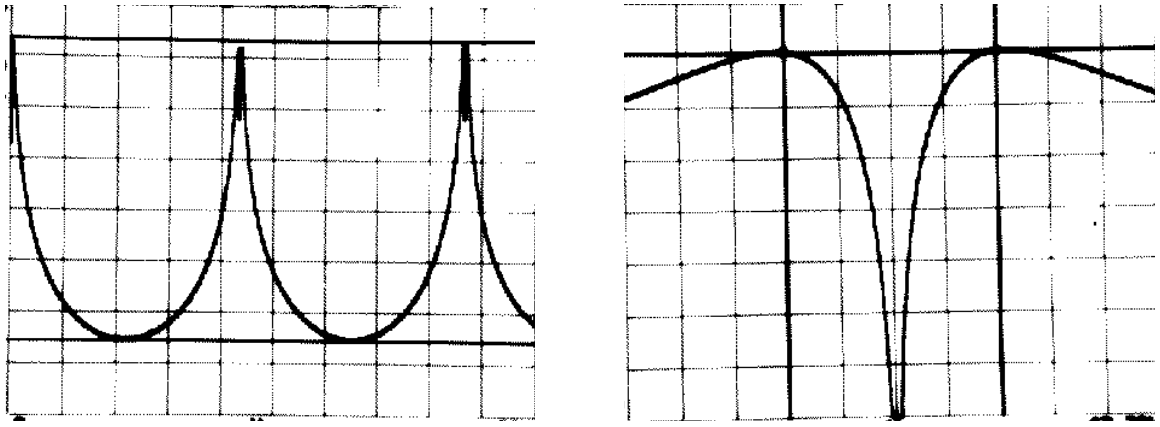

Fig. 18: Frequency response of the double-peaked comb filter, showing peaks on the synchrotron sidebands of the revolution frequency lines. Left: 100 kHz span, 5 dB/div. Right: enlargement (1 kHz span, 3 dB/div) around one f_{rev} line, showing the two peaks on the sidebands ($a = 31/32$, $f_{rev} = 43$ kHz).

The corresponding impedance reduction is shown in Fig. 19. The smooth trace shows the cavity impedance with the RF feedback. The comb shows the additional reduction achieved around the revolution frequency lines with the double-peaked long delay feedback.

In the time domain, the filtering realized by $H_{sbd}(z)$ can be implemented as the following set of difference equations (see also Section 4.2.4). The term x_n is the input and y_n the output:

$$v_n = a\, v_{n-M} + x_n \tag{48}$$

$$w_n = a\, w_{n-M} + v_n - v_{n-M} \tag{49}$$

$$y_n = w_{n-M} \ . \tag{50}$$

Figure 20 shows a direct implementation of the above equations. The delays are implemented using FIFOs. The multiplications by $31/32$ are replaced by subtraction of a shifted version of the operand

[9] This z-transform is analysed in Section 4.2.3.

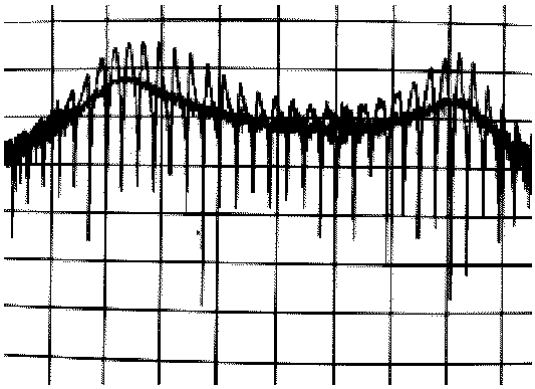

Fig. 19: Apparent impedance of the CERN SPS 352 MHz superconducting cavity. The smooth trace shows the impedance with the RF feedback alone. The second trace shows the additional reduction around the revolution frequency lines with the double-peaked long delay feedback: $a = 31/32$, 150 kHz/div, 3 dB/div. The resolution of the measurement is not sufficient to show the full 20 dB reduction or to separate the sidebands.

from itself. (In two's complement arithmetic a scaling by 32 is easily implemented by shifting the binary word five positions towards the LSB, with extension of the most significant bit.) Additions and subtractions are realised in the ALUs.

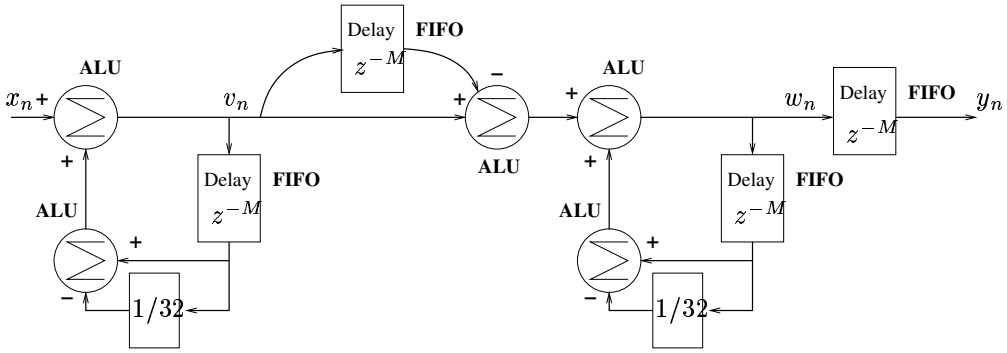

Fig. 20: Implementation of the double-peaked comb filter using ALUs and FIFOs

The advantage of such an implementation using discrete numerical operators is its speed. A new output signal is available on each clock pulse. A 50 MHz clock rate is easily achievable with 16-bit words. It is a straightforward implementation for digital filters containing loops, i.e. with poles in their z-transform (also called Infinite Impulse Response filters or IIR).
The main limitation is related to the use of fixed-point arithmetic:

- Rounding noise: after the shift by five positions (multiplication by $1/32$) in the above loops, the binary word must be truncated (or rounded) to remain a 16-bit word. This rounding creates an error (called rounding noise) that will propagate to the output of the filter.
- Rounding of filter coefficients: the coefficients of the filter must also be quantized. In the above design we chose $a = 31/32$, which can be implemented exactly with 16 bits but, in general, the coefficients must be rounded to fit the binary word and this will modify the filter response.
- Scaling: the input must be scaled properly to guarantee that the outputs of the ALUs do not overflow.

- Choice of structure: given a z-transform, many sets of difference equations can realize it. But the choice of the realization is important because some structures are much more sensitive to the negative effects of fixed-point arithmetic than others. In general, if the filter has a high order (many poles and zeros), a cascade of low-order sections is preferred (bi-quad sections having a maximum of two zeros and two poles) instead of direct form realization.

This type of implementation does not have much future. There is no new development from chip designers regarding discrete ALUs and multipliers. New FIFOs are, however, periodically being introduced.

3.4 Implementation with video products

Principle: Recently (end 1998) a new family of products was introduced for digital image processing in High Definition TV (HDTV) applications. Their speed and resolution make them ideal for our applications. The LF3320 (Horizontal Digital Image Filter) from Logic Devices implements a 32-tap Finite Impulse Response filter (FIR, see Section 4.2.2) at data rate of 83 MHz with a resolution of 12 bits for data and coefficients $h_0, h_1, ..., h_{L-1}$

$$y_n = h_0 x_n + h_1 x_{n-1} + h_2 x_{n-2} + \cdots + h_{L-1} x_{n-L+1} \quad . \tag{51}$$

The advantages of such an implementation are

- Accuracy: if the input data x_n and the coefficients $h_0, h_1, ..., h_{L-1}$ are 12-bit words in the above equation, the result y_n must be extended to 29 bits to avoid overflow and rounding (for $L = 32$). Internally, the chip keeps all these bits, and by programming the output scaling one chooses which 16 bits come out.
- Flexibility: by programming (with an on-board ROM for example or external host), one changes the filter coefficients, the output scaling and the output limiting. This facility to bound the output is very attractive in set-ups where the filter output is the drive to a power amplifier. An overflow in two's complement arithmetic results in the output of the DAC dropping from the maximum positive value to the minimum negative value. This fast transient is likely to trip the power amplifier.
- It is cascadable for larger filters (more than 32 coefficients).
- It supports decimation up to 16:1 with a correspondingly increased number of filter coefficients.
- It is targeted at a promising new market and new developments can be expected.

The disadvantages are

- It can implement FIR filters only. These have only zeros in their transfer function (no pole).
- The range is limited by the use of fixed-point arithmetic (12 bits). It suffers from the need to quantize the coefficients. Rounding noise is not a big problem. It is here limited to one LSB at the output since the FIR keeps all bits in the computation.

Example: The performance of the feedforward on the CERN SPS 200 MHz travelling-wave cavities was presented in Section 2.1 (Fig. 6). The implementation uses the heterodyne method shown in Fig. 15. Each of the two identical digital filters is realized using two FIRs (LF3320) clocked at $f_c = 20$ MHz. The optimal transfer function H_{opt} is complex for a travelling-wave cavity because Z_g is not equal to Z_b. We have [29]

$$Z_g(\Delta\omega) = l\sqrt{\frac{Z_0 r_2}{2}} \left(\frac{\sin \tau/2}{\tau/2}\right) \tag{52}$$

where the phase slip $\tau(\Delta\omega)$ is

$$\tau(\Delta\omega) = \frac{l}{v_g}\left(1 - \frac{v_g}{v}\right)\Delta\omega \tag{53}$$

with $\Delta\omega = \omega - \omega_0$. The term ω_0 is the angular frequency of the cavity fundamental resonance, l is the interaction length of the cavity, v is the particle velocity, v_g is the group velocity in the cavity, r_2 is the series impedance (Ω/m^2), and Z_0 is the characteristic impedance of the RF chain (50 Ω). Notice that Z_g is purely real but its sign changes periodically as a function of frequency. The impedance Z_b is

$$Z_b(\Delta\omega) = -\frac{l^2 r_2}{8}\left[\left(\frac{\sin\tau/2}{\tau/2}\right)^2 - 2j\left(\frac{\tau-\sin\tau}{\tau^2}\right)\right] \tag{54}$$

It has both a real and an imaginary part. The naive solution $H_{opt} = -Z_b/Z_g$ will not work for a travelling-wave cavity because Z_g vanishes at frequencies where the imaginary part of Z_b is non-zero. Let us decompose H_{opt} into a real and an imaginary part:

$$H_{opt} = H_{opt}^{re} + jH_{opt}^{im} \ . \tag{55}$$

Since Z_g is real valued, our design goal is

$$H_{opt}^{re} Z_g + jH_{opt}^{im} Z_g \approx -Re[Z_b] - jIm[Z_b] \ . \tag{56}$$

By inspection of Z_g, (Eq. (52)), and Z_b, (Eq. (54)), it follows that the real parts on the two sides can be made identical:

$$H_{opt}^{re} = \frac{l}{4}\sqrt{\frac{r_2}{2Z_0}}\left(\frac{\sin\tau/2}{\tau/2}\right) \tag{57}$$

and the compensation of the resistive part of beam loading is exact. The impulse response of H_{opt}^{re} is rectangular (inverse Fourier transform of H_{opt}^{re}), lasting for a time equal to $\frac{l}{v_g}\left(1-\frac{v_g}{v}\right)$. In the case of the SPS cavities this corresponds to 12 samples at 20 MHz. This is implemented with the first FIR filter (12 coefficients are equal to 1, the other coefficients being 0). The value jH_{opt}^{im} is implemented with a second FIR filter. Its impulse response h_n^{im} is limited to 31 samples ($-15 \le n \le 15$) and must be odd-symmetric ($h_{-n}^{im} = -h_n^{im}$) so that its frequency response is purely imaginary. Perfect compensation is not possible, however, because Z_g is zero at frequencies where the imaginary part of Z_b is non-zero. We must therefore choose a criterion for computing the optimal coefficients $h_1^{im}, h_2^{im}, ..., h_{15}^{im}$. The details are presented in Ref. [12]. The outputs of the two FIR filters are finally added together using an ALU (IDT7381 or L4C381).

3.5 Implementation with Programmable Gate Array (PGA)

The PGA (or Programmable Logic Device, PLD) is a high-density gate array (up to 200 000 gates on one chip) that the user can configure to implement the desired function. Modern devices also include a memory and some specialized functions on the chip (embedded programmable logic arrays).

The advantages of the PGA are:

- It helps ensure very compact hardware by minimizing the external connections between chips.
- It can be re-configured on the board.
- Powerful development and debugging tools are available. It is possible to simulate both the logic function and the delays to evaluate the possible overall processing rate.
- PGAs are widely used and we can therefore expect new developments in the field.

The drawback is that the resulting design will implement fixed-point arithmetic only.

The bunch-by-bunch feedback at the Japanese b-Factory KEKB is an example of an implementation with PGA [24].

Figure 21 shows a block diagram of the feedback filter. The longitudinal position of each bunch is measured with a wide band phase detection system (not shown here) capable of distinguishing individual bunches spaced by less than 2 ns (maximum bunch frequency 509 MHz) [31]. The result is an analog

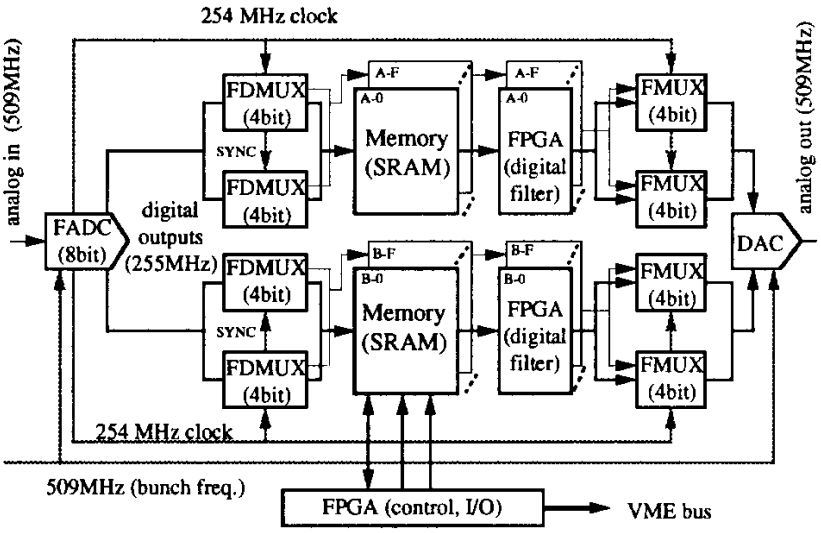

Fig. 21: Digital filter of the KEKB bunch-by-bunch feedback (reproduced from Ref. [30]). The Field Programmable Gate Array (FPGA) implements the 2-tap filtering of Eq. (58).

signal whose average value, in each 2 ns window, is a measurement of the phase $\hat{\phi}_k$ of the corresponding bunch. This signal is fed, on the left, into the digital filter shown on the figure. The ADC output is first demultiplexed in 2 channels, each being further demultiplexed in 16 channels. The processing rate is thus finally reduced to 16 MHz. The Field Programmable Gate Array (FPGA) QL16 $\times$ 24B made by Quick Logic [32] implements a subtraction between its two 8-bit operands. These are actually the same time sequence x_n, but one is delayed by L samples with respect to the other, thereby implementing a two-tap FIR filter with fixed coefficients 1 and -1. The output of the subtractor y_n is related to its input x_n by the difference equation

$$y_n = x_n - x_{n-L} \ . \tag{58}$$

The delay L is adjusted to get the required 90° phase shift at the synchrotron frequency. With these coefficients (1 and -1) the FIR filter also eliminates the DC component (stable phase).

The outputs of the 32 channels are finally multiplexed and converted into analog (the output signal is shown on the right of the figure). This analog signal is fed into the IF port of a mixer, receiving twice the bunch frequency on its LO (not shown here). The RF output drives the amplifiers feeding the wide band kicker. The design of the digital filter dates from the mid-1990s. The FPGA is only used to implement the subtraction between two 8-bit words, which are properly aligned in time by the memories (Static Random Access Memories, marked SRAM on Fig. 21). These two functions could be integrated in a single Embedded PGA chip nowadays, which would make the implementation more attractive. Since the speed of the components has improved, using the latest technology would also reduce the number of channels. As an example of the power of modern PGAs, Altera proposes software called MegaCore to implement FIR filters using its FLEX10KE family. With this tool, it quotes the realization of a 19-tap FIR filter (8-bit input) at 101 MHz throughput rate [33]. In the CERN SPS we have used this technology to upgrade the electronics of the transverse damper for the LHC beam: a digital filter for closed-orbit rejection[10] has been designed with a single chip FLEX 10K100E from Altera. It works with 12-bit inputs and has a throughput of 80 MHz [30].

[10] The minimal digital filter of a *transverse* bunch-by-bunch feedback produces notches at the multiple of the revolution frequency and introduces a delay of one full turn [34]: $H(z) = (1 - z^{-M})z^{-M/2}$. The required 90° phase shift can be obtained from the betatron phase advance between pick-up and kicker.

3.6 Implementation with digital signal processors

Digital Signal Processors (DSPs) are microprocessors that specialize in performing repetitive mathematical operations. Most DSPs are based on the Harvard architecture. The core processor is connected to two separate memories by two separate buses so that two memory accesses can be made in one cycle. For example a new instruction can be fetched from one memory while data is fetched from the other. In the implementation of an FIR (Eq. (81)), one item of data x_{n-i} can be fetched from one memory and one coefficient h_i fetched from the other memory (if the instruction is already in a third smaller memory called the cache). All DSPs incorporate a hardware multiplier/accumulator unit, which, combined with the Harvard architecture, makes it possible to perform a complete floating-point multiply-accumulate operation in a single clock cycle, including the fetch of the two operands. Older DSPs implement only fixed-point arithmetic. The ADSP-2106x from Analog Devices and the TMS320C6x from Texas Instruments are in use in particle accelerators. In the first part of these two lectures [8] the beam control system of the RHIC at Brookhaven National Laboratory was presented. It uses two DSPs (TMS320C40) clocked at 50 MHz [35]. Some of the systems presented here can also be implemented with DSPs.

The advantages of the DSPs are:

- Modern DSPs implement floating-point operations in the 32-bit IEEE standard. It was shown above that the limited dynamic range of fixed-point arithmetic results in the need for careful scaling of inputs and intermediate signal levels to avoid overflow. Here, the need for such scaling is essentially eliminated by using floating-point arithmetic. Rounding noise is still introduced at each operation but the signal-to-noise ratio at the output of the filter is significantly better for floating point arithmetic as compared with fixed-point arithmetic.[11] The filter coefficients still need to be quantized but this is done at constant relative precision.
- The industry provides complete DSP systems. For example a printed circuit board containing several DSPs linked together, external memories, input/output ports and a VME bus for connection to an external host. Very complete software tools are also provided, such as extensive signal processing libraries to implement FIR and IIR filters, Fast Fourier Transform (FFT). The prototype can be tested in an emulator. Debugging tools are also available. The algorithm implemented by the DSP can be re-programmed at will.
- The filtering realized by the DSP can be very complex. For example one could imagine adjusting the coefficients of the bunch-by-bunch feedback BPF continuously (Fig. 14) so that it remains centred on the synchrotron frequency f_s during the acceleration. The low-level system developed at RHIC and presented in the first part of these two lectures [8] is a good example of the sophistication made possible by DSPs: the feedback gains (radial loop, beam phase loop and synchronization loop) are varied during the acceleration to keep the poles of the closed-loop response at the optimal location while the synchrotron frequency varies. Notice, however, that these gains are adjusted at a very slow rate: only 200 times during the 74 s long acceleration ramp [35].
- Control and diagnostic functions can easily be integrated. Additional software code can be added to change the filtering function or to read interesting signals (see the example of ALS shown below).
- The market for DSPs is growing fast. We can expect continuing improvements in their performance over coming years.

The drawback is that the DSPs are still very slow for RF applications. They may have a very fast clock frequency, they still have only one multiplier/accumulator unit and will perform only one multiplication at a time.[12] It therefore takes many clock cycles to produce a single output sample when

[11] This is obvious if we compare the 32-bit floating point IEEE standard (23-bit in the mantissa) to the classic 8- or 16-bit fixed point format. But even with a fair comparison taking the same number of bits in the mantissa of the floating format as in the fixed-point format one concludes that the output signal-to-noise ratio is better with the floating point implementation [36].

[12] The Single-Instruction-Multiple-Data (SIMD) DSPs from Analog Devices have two arithmetic units executing the same instruction. But performance increases only if two independent filtering channels have to be implemented.

the DSP must implement a filtering operation. For example, suppose that we wish to use a DSP to implement a L taps FIR filter for the feedforward presented in Section 3.4 (feedforward on a travelling-wave cavity). Generating the code using the optimized assembly-language library of the ADSP-21000 family, it takes $5 + L$ cycles to compute a single output value y_n [37]. Thirty-six cycles are needed to generate one output of the 31-tap FIR. If the DSP is clocked at 100 MHz,[13] we get an output signal every 360 ns, compared with the 12 ns needed by the video filter LF3320 clocked at 83 MHz. In this application, the sampling rate of the DSP implementation would be 2.8 MHz and the useful bandwidth less than 1 MHz (one-third the sampling rate). In our RF applications DSPs are only a good candidate if the processing rate can be much reduced by downsampling and interpolation. If parallel processing is possible, one could also compensate the low rate by using several DSPs in parallel, as shown in the following example.

The bunch-by-bunch feedback system of the ALS (synchrotron light source) at LBNL is implemented using DSPs (Fig. 22) [23].

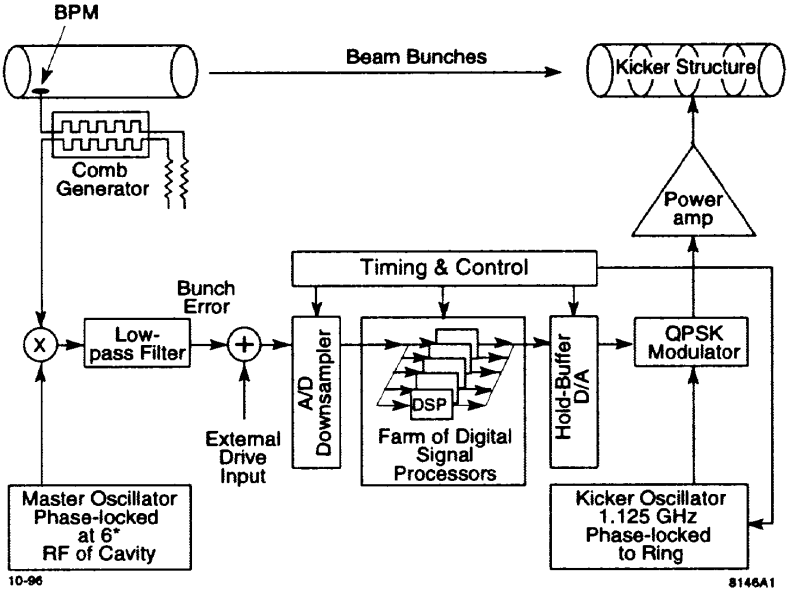

Fig. 22: Block diagram of the longitudinal bunch-by-bunch feedback used at ALS (reproduced from Ref. [23])

The machine parameters are 328 bunches, 2 ns spacing, revolution frequency $f_{rev} = 1.5$ MHz, and synchrotron frequency $f_s = 11$ kHz. The acquisition part is similar to the one implemented at KEKB and we shall not detail it. The bunch phase signal $\hat{\phi}$ is fed into the ADC. Recall that, for each bunch k, the bunch-by-bunch feedback must implement a separate BPF processing the average bunch phase $\hat{\phi}_k$. Since the processing of each bunch is independent of the others, we can treat them in parallel, using a bank of DSPs, each branch implementing one BPF with an L-tap FIR. The throughput rate could be reduced to a comfortable 1.5 MHz (f_{rev}) with 328 DSPs. Realization becomes possible once we notice that the synchrotron tune is small ($Q_s = f_s/f_{rev} = 7.3 \times 10^{-3}$). The phase of each bunch does not change much between two turns. Let us therefore reduce the processing rate by a factor D: for each bunch, we treat its signal every D turns only. This process is called downsampling. A reasonable value for the downsampling ratio D is a fraction of $1/Q_s$. For N bunches the total number of multiply–accumulate

[13] Texas Instruments advertises a new 600 MHz DSP including four multiplier/accumulator units [38] (TMS320C64x). We have not evaluated it yet.

cycles per second (MAC/s) using an L-tap FIR is[14]

$$\frac{NLf_{rev}}{D} \quad . \tag{59}$$

In the case of the ALS, taking 10 samples per synchrotron period ($D = 1/(10Q_s) \approx 14$) and with a 20-tap FIR, the processing power required is 7.03×10^8 MAC/s. Considering a modern DSP clocked at 100 MHz, realizing one multiply–accumulate per cycle, we see that the above bunch-by-bunch feedback can be implemented using only eight such DSPs, each one processing the signal from 41 bunches and producing one output per bunch every 14 turns. The kick value for each bunch will thus be kept constant over 14 turns (Hold–Buffer DAC).[15] This derivation is, however, overly optimistic: the ALS design actually uses 40 DSPs in parallel. In addition the DSPs provide a powerful diagnostic tool. The phase signal from each bunch can be observed individually. The growth rate of each mode of instability can be measured by switching the feedback off, letting the oscillation develop, and switching the feedback back on. This allows measurements in the small-signal regime [23]. Acquisition of these signals is easily achieved through the existing interface between the DSP and its host computer.

4 ANNEXE

4.1 Spectrum of the beam-induced voltage

First consider a uniform ring distribution with all buckets filled and suppose that the bunch length is infinitely short. The beam current measured in a pick-up then consists of a series of infinitely narrow pulses spaced in time by the RF period.

$$i_\delta(t) = I_0 T_{rf} \sum_{l=-\infty}^{+\infty} \delta(t - lT_{rf}) \tag{60}$$

where I_0 is the DC component of the beam current. Using the identity

$$\sum_{l=-\infty}^{+\infty} \delta(t - lT) = \frac{1}{T} \sum_{m=-\infty}^{+\infty} e^{j2\pi mt/T} \tag{61}$$

the beam current can be rewritten

$$i_\delta(t) = I_0 \sum_{m=-\infty}^{+\infty} e^{j2\pi m f_{rf} t} \quad . \tag{62}$$

The corresponding spectrum is an infinite set of discrete lines at the RF frequency and its harmonics: $f = \pm m f_{rf}$. In practice the bunches are not infinitely short. Let $\lambda(t)$ be the normalized longitudinal charge density, the beam current is

$$i_\lambda(t) = I_0 T_{rf} \sum_{l=-\infty}^{+\infty} \lambda(t - lT_{rf}) = i_\delta(t) * \lambda(t) \quad . \tag{63}$$

The current i_λ is the convolution of the two time domain signals $i_\delta(t)$ and $\lambda(t)$. The spectrum of the result is thus obtained by multiplying the spectrum of i_δ by the Fourier transform of the bunch charge density λ. This spectrum still consists of a set of discrete lines at $f = \pm m.f_{rf}$ but the amplitude of the spectral lines in the frequency domain will decrease with an envelope equal to the Fourier transform of the charge density $\lambda(t)$. The width of the envelope is inversely proportional to the bunch length. The

[14] We have neglected the small overhead of 5 cycles in the total number of cycles $5 + L$.

[15] A proper downsampling/interpolation filter (also called multirate filter) is somewhat more complex. See Section 4.2.5 for more details.

bandwidth of the accelerating system around f_{rf} being much narrower than the RF frequency we can neglect all the harmonics. The non-zero bunch length then only introduces a scaling factor in the spectral component at f_{rf}. For our analysis the beam current can thus be simplified to

$$i_{uniform}(t) \propto \sin(2\pi f_{rf} t + \phi) \ . \tag{64}$$

that is, a single line at the RF frequency. This remains valid if the bunch spacing is a multiple of the RF period (uniform ring but not all buckets filled), assuming that the bunch frequency (inverse of the bunch spacing) is much larger than the bandwidth of the RF system so only the spectral component at f_{rf} need to be considered.

We now consider a non-uniform ring distribution. This is the case if the ring pattern consists of one or several batch(es) of bunches and one (or more) hole(s) without bunches. As seen in a pick-up (or in a cavity), the beam current is then modulated by a function $a(t)$ representing the batch envelope

$$i_{batch}(t) \propto a(t)\sin(2\pi f_{rf} t + \phi) \ . \tag{65}$$

The same bunch pattern periodically crosses the pick-up. The modulation function $a(t)$ is thus periodic in time with a period equal to the revolution period T_{rev}. We can expand it as a Fourier series and the beam current becomes

$$\begin{aligned} i_{batch}(t) \quad \propto \quad & [a_0 + a_1\cos(2\pi f_{rev} t) + a_2\cos(4\pi f_{rev} t) + a_3\cos(6\pi f_{rev} t) + \cdots \\ & + b_1\sin(2\pi f_{rev} t) + b_2\sin(4\pi f_{rev} t) + b_3\sin(6\pi f_{rev} t) + \cdots]\sin(2\pi f_{rf} t + \phi) \end{aligned} \tag{66}$$

where $a_0, a_1, a_2, \ldots, b_1, b_2, \ldots$ are the coefficients of the Fourier series. The amplitude of the coefficients will typically decrease with increasing index in a manner that is a function of the shape of the batch. Consider the classic situation where only a fraction α of the ring is evenly filled. By choosing the time origin so that $a(t)$ is even-symmetric, we get

$$a_n = \frac{\sin\pi\alpha n}{\pi\alpha n} \tag{67}$$

$$b_n = 0 \ . \tag{68}$$

Figure 23 shows the spectrum of the beam current $i_b(t)$ in that case. It consists of the carrier frequency f_{rf} plus a set of sidebands at $f_{rf} \pm n.f_{rev}$. The voltage V_b induced by the beam in the cavity will thus consist of a discrete set of lines at the frequencies

$$f = f_{rf} \pm n f_{rev} \ . \tag{69}$$

We now return to the uniform ring distribution and consider a beam undergoing longitudinal dipole oscillations, as shown in Fig. 3. Consider a single infinitely narrow bunch undergoing a pure dipole oscillation, the measured current is

$$i_{bunch,dipole}(t) = I_0 T_{rev} \sum_{l=-\infty}^{+\infty} \delta(t - lT_{rev} - \hat{\tau}\sin(2\pi f_s t + \psi)) \ , \tag{70}$$

where $\hat{\tau}$ is the amplitude of the synchrotron oscillation (in time) and f_s is the synchrotron frequency. Using the identity from Eq. (61) we rewrite the bunch current

$$i_{bunch,dipole}(t) = I_0 \sum_{n=-\infty}^{+\infty} e^{j2\pi n f_{rev}[t - \hat{\tau}\sin(2\pi f_s t + \psi)]} \ . \tag{71}$$

Now using the identity

$$e^{-jx\sin\phi} = \sum_{m=-\infty}^{+\infty} (-1)^m J_m(x) e^{jm\phi} \tag{72}$$

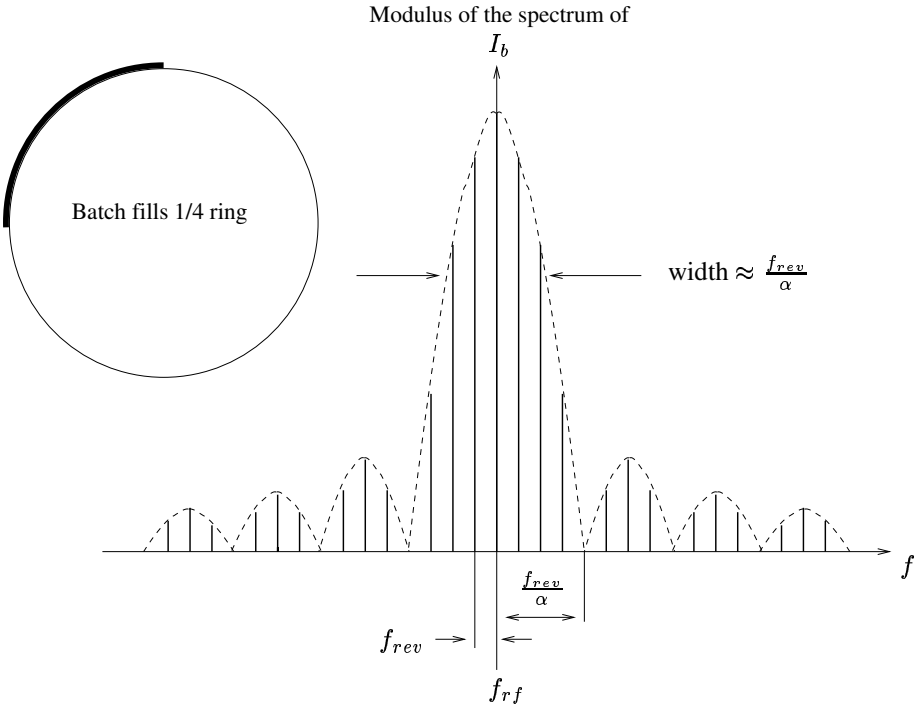

Fig. 23: Frequencies where beam loading must be compensated: spectrum of the beam current I_b in the case where only a fraction α of the ring is filled with bunches ($\alpha = 1/4$)

we get

$$i_{bunch,dipole}(t) = I_0 \sum_{n=-\infty}^{+\infty} e^{j2\pi n f_{rev} t} \sum_{m=-\infty}^{+\infty} (-1)^m J_m(2\pi n f_{rev}\hat{\tau}) e^{j[2\pi m f_s t + m\psi]} \tag{73}$$

$$i_{bunch,dipole}(t) = I_0 \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} (-1)^m J_m(2\pi n f_{rev}\hat{\tau}) e^{j[2\pi (n f_{rev} + m f_s) t + m\psi]} \tag{74}$$

Around each harmonic of the revolution frequency (at nf_{rev}) there is an infinite number of synchrotron sidebands (at $nf_{rev} + mf_s$). The spectral amplitude of the m th sideband of the n th revolution line is given by $J_m(2\pi n f_{rev}\hat{\tau})$ (Bessel function of order m). The spectral lines falling in the RF cavity impedance (near the fundamental) have index $n \approx h$ (where h is the harmonic number). The amplitudes of their sidebands are proportional to

$$J_m(2\pi n f_{rev}\hat{\tau}) \approx J_m(2\pi h f_{rev}\hat{\tau}) = J_m(2\pi \frac{\hat{\tau}}{T_{rf}}) \ . \tag{75}$$

When the instability starts growing, the amplitude of the longitudinal oscillation is much smaller than the RF period. The dominant sidebands are thus the first ones (at $nf_{rev} \pm f_s$) since, for a small value of their argument δ, the Bessel function of higher order are close to zero ($J_m(\delta) \approx (\delta/2)^m / m!$). Taking, for example, an oscillation of $\pm 10^\circ$ in the RF bucket ($\tau/T_{rf} = 10/360$), the strength of the sidebands relative to the revolution frequency line will be 0.09 for $m = 1$, 0.004 for $m = 2$, 0.0001 for $m = 3$, etc. Figure 3 shows an oscillation of larger amplitude: ± 1.5 ns in the 100 MHz bucket ($\tau/T_{rf} = 1.5/10$), and the sidebands of large index are not negligible. Relative to the revolution frequency line we get an amplitude of 0.53 for $m = 1$, 0.13 for $m = 2$, 0.02 for $m = 3$, etc. We show only one bunch in Fig. 3. The other bunches execute similar dipole oscillations at the same frequency f_s but each bunch k has its own phase ψ_k. For N evenly spaced bunches of equal intensity, each bunch executing a dipolar

longitudinal oscillation of the same amplitude but with different phases, the beam current is

$$i_{beam,dipole}(t) = I_0 T_{rev}/N \sum_{k=0}^{N-1} \sum_{l=-\infty}^{+\infty} \delta(t - lT_{rev} - \frac{k}{N}T_{rev} - \hat{\tau}\sin(2\pi f_s t + \psi_k)) \ . \quad (76)$$

Following the derivation used for a single bunch, we can rewrite the beam current as

$$i_{beam,dipole}(t) = I_0 \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} (-1)^m J_m(2\pi n f_{rev}\hat{\tau}) e^{j[2\pi(nf_{rev}+mf_s)t]} \sum_{k=0}^{N-1} e^{j[m\psi_k - 2\pi kn/N]} \ . \quad (77)$$

The spectrum is similar to the one created by a single bunch. The only difference is the last factor, which combines the phases ψ_k of the dipolar oscillations of the individual bunches to enhance or reduce the spectral lines. We conclude that the spectrum of the voltage V_b induced in the cavity by a beam undergoing a small longitudinal dipole oscillation contains power only at the frequencies

$$f = f_{rf} \pm nf_{rev} \pm mf_s \quad (78)$$

with the amplitude of the sidebands decreasing quickly with increasing index m.

The quadrupole oscillation shown in Fig. 4 does not modulate the phase of the beam current but only its amplitude. Since the frequency of the quadrupole mode is twice the synchrotron frequency we can expect that this mode of oscillation will create dominant sidebands at the frequencies $f = f_{rf} \pm nf_{rev} \pm 2f_s$. Generalizing to the higher order modes, we conclude that the voltage V_b induced in the cavity (around its fundamental resonance) by a beam undergoing a longitudinal oscillation has power only at the discrete frequencies

$$f = f_{rf} \pm nf_{rev} \pm mf_s \ . \quad (79)$$

4.2 Digital filters

4.2.1 Nyquist rate

Nyquist theorem states that a continuous-time signal can be sampled without loss of information if the sampling rate is at least twice the highest frequency present in the signal spectrum. In a mixed signal set-up, one first filters the analog signal with an analog anti-aliasing filter (low pass filter). This filter has a stop band starting at a frequency lower or equal to half the sampling frequency. In order not to distort the signal band, some frequency margin must be left for the transition band (between passband and stop band), so that the maximum interesting frequency in the continuous-time signal does not exceed one-third of the sampling frequency.

4.2.2 Finite Impulse Response and Infinite Impulse Response filters

A digital filter is an operator that maps an input sequence x_n to an output sequence y_n. If the filter is linear and invariant in time, the output y_n is related to the input x_n via the convolution

$$y_n = h_n * x_n \ , \quad (80)$$

h_n is the filter impulse response, defined as the output when the filter is excited by a Dirac pulse δ_n at its input ($\delta_n = 0$ for $n \neq 0$, and $\delta_0 = 1$). If the impulse response is of finite duration the filter is called a Finite Impulse Response (FIR) filter. Otherwise it is called an Infinite Impulse Response (IIR) filter. An FIR filter, of impulse duration L samples, realizes the following filtering operation

$$y_n = h_0 x_n + h_1 x_{n-1} + h_2 x_{n-2} + \cdots + h_{L-1} x_{n-L+1} \quad (81)$$

It is easily implemented with a tap delay line of length L, fed at its input by the signal x_n, and with an adder that computes the weighted sum of the tap outputs, the weights being the values of the impulse

response. IIR filters have feedback loops relating the value of the output at time n with the past values of the output. For example, the filter implementing the recursion

$$y_n = a\, y_{n-1} + x_n \tag{82}$$

has the impulse response $h_n = a^n$ for $n \geq 0$ and $h_n = 0$ for $n < 0$. The impulse response lasts forever.

4.2.3 z-transform

The Laplace transform is a very powerful tool for analysing continuous-time linear time-invariant systems. It provides answers to important questions of stability and frequency response. The equivalent for the analysis of discrete-time linear time-invariant systems is the z-transform, defined as

$$X(z) = \sum_{n=-\infty}^{+\infty} x_n z^{-n} \ . \tag{83}$$

The transfer function of the filter is the z-transform of its impulse response

$$H(z) = \sum_{n=-\infty}^{+\infty} h_n z^{-n} \ . \tag{84}$$

Because the z-transform of the convolution of two sequences is the product of the z-transforms of the two sequences, it follows from Eq. (80) that the z-transform $Y(z)$ of the filter output is the product of the z-transform of its input and the transfer function $H(z)$,

$$Y(z) = H(z)\, X(z) \ . \tag{85}$$

The variable z is the equivalent of the variable s for the Laplace transform. It is complex valued. The poles (zeros) of the transfer function are defined as the values of z for which the denominator (numerator) of $H(z)$ equals zero. Let us consider the double-peaked comb filter of Eq. (47). We have M zeros at

$$z_k = e^{j2\pi \frac{k}{M}} \tag{86}$$

for $k = 0, 1, 2, \cdots, M-1$. And M poles at

$$z_p = a^{\frac{1}{M}} e^{j2\pi \frac{p}{M}} \tag{87}$$

for $p = 0, 1, 2, \cdots, M-1$. Each pole has a multiplicity of two: it appears twice in the denominator. Recall that, for continuous-time filters, the frequency response is obtained by evaluating the Laplace transform on the imaginary axis. This gives the Fourier transform. Similarly, the frequency response of the digital filter is obtained from the evaluation of its transfer function on the *unit circle*. Let $\Omega = \frac{f}{f_c}$ be the normalized frequency of the input sine wave. The filter output will be a sine wave at the same frequency, but its amplitude and phase are given by the modulus and phase of $H(e^{j2\pi\Omega})$. Let us consider the double-peaked comb filter again. Because the zeros are located on the unit circle (Eq. (86)), the filter will show zeros of transmission at the exact multiples of the revolution frequency as shown in Fig. 18. The parameter a is very close to one, so that the pairs of poles (Eq. (87)) will be located very close to the corresponding zero but slightly inside the unit circle. This produces the desired pair of resonances on the synchrotron sidebands (Fig. 18).

4.2.4 From difference equations to transfer function

From a set of difference equations, we can easily get the filter transfer function by applying the z-transform to both sides of the equations. Two properties of the z-transform must be recalled:

Linearity: the z-transform of a sum of two sequences x_n and y_n is the sum of the z-transform of each sequence, $X(z) + Y(z)$.

Delay property: the z-transform of a sequence delayed by one sample is z^{-1} times the z-transform of the original sequence.

Using these properties, we derive from Eqs. (48),(49) and (50)

$$V(z) = az^{-M} V(z) + X(z) \tag{88}$$

$$W(z) = az^{-M} W(z) + \left(1 - z^{-M}\right) V(z) \tag{89}$$

$$Y(z) = z^{-M} W(z) \ , \tag{90}$$

where $X(z)$, $Y(z)$, $V(z)$ and $W(z)$ are the z-transforms of the sequences x_n, y_n, v_n and w_n respectively. We can now eliminate $V(z)$ and $W(z)$ from the above three equations and we get

$$Y(z) = \frac{1 - z^{-M}}{(1 - az^{-M})(1 - az^{-M})} z^{-M} X(z) \ . \tag{91}$$

This gives the transfer function $H_{sbd}(z)$ of Eq. (47). Notice, however, that several different sets of difference equations will lead to the same transfer function and they therefore realize the same filter. For example H_{sbd} is also realized by the following difference equation

$$y_n = 2a\, y_{n-M} - a^2\, y_{n-2M} + x_{n-M} - x_{n-2M} \ . \tag{92}$$

However, some realizations will be much more sensitive to quantization effects and the cascade form of Fig. 20 is often preferred.

4.2.5 Multirate filters

Throwing away $D - 1$ data out of every D data as proposed in Section 3.6 will unavoidably worsen the signal-to-noise ratio. Also, simply holding the output constant for D turns will introduce distortion. It is, however, possible to reduce the necessary processing without loss of precision [39]. The input sequence x_n is first processed by a so-called decimation filter (impulse response h_n^d). It is a low pass filter with a passband extending to the maximal synchrotron frequency $f_{s,max}$ (so that it covers the band of the BPF) and a stopband starting at $f_{rev}/D - f_{s,max}$. This puts a first limit on the downsampling ratio $D < f_{rev}/2f_s = 1/2Q_s$. All input data is used as input to the decimation filter but its output is computed every D turns only (downsampling). This sequence is fed into the original BPF, called the kernel filter in multirate filtering, and now operated at the reduced rate f_{rev}/D (rejection of the DC component and of the noise outside the synchrotron frequency band, $\pi/2$ phase shift at the synchrotron frequency). The saving in processing time in the BPF only is proportional to D^2 because the number of coefficients required to implement a given passband characteristic (in Hz) scales linearly with $1/D$ if the transition bandwidth is kept constant (from passband to stopband in Hz). We here assume that all filters are implemented as FIR. The original sampling rate is recovered at the BPF output using an interpolation filter h_n^u (upsampling) that produces an output, at every turn, from its input sequence at the rate f_{rev}/D. The interpolation filter must reject the image spectra created by the upsampling process. This requirement is satisfied if we choose the same filter for interpolation and decimation ($h_n^u = h_n^d$). Notice that, out of every D samples at the input of h_n^u (rate f_{rev}), $D - 1$ samples are equal to zero. The unnecessary multiplications by zero are not performed, thereby saving on the overall processing time. With this multirate scheme there is no degradation in performance compared to the implementation of the BPF at the rate f_{rev}, but a significant saving in processing time is possible if $f_{rev} >> 2f_{s,max}$. While the processing time required by the BPF decreases as $1/D^2$, the time spent computing the outputs of the decimation and interpolation filters increases with D and the overall processing time is therefore minimal for some optimal value of D.

REFERENCES

[1] A. Gamp, Servo control of RF cavities under beam loading, these proceedings.

[2] D. Boussard, Beam loading, Proc. CERN Accelerator School, Oxford, 1985, CERN 97-03 (1997).

[3] D. Boussard, Control of cavities with high beam loading, *IEEE Trans. Nuc. Sci.* **NS-32** (1985) No. 5.

[4] M.M. Karliner, Beam–cavity interaction, beam loading, in *Frontiers of Accelerator Technology*, Proc. Joint US–CERN–Japan International School, Tsukuba, 1996 (World Scientific, Singapore, 1999).

[5] R. Garoby, Beam loading in RF cavities, *Beam Intensity Limitations*, Joint US–CERN School, Hilton Head Island, South Carolina, 1990 (Lecture Notes in Physics No. 400, Springer, Berlin).

[6] J. Le Duff, High frequencies non-ferrite cavities, these proceedings.

[7] D. Boussard, Travelling-Wave Structures, in *Frontiers of Accelerator Technology*, Proc. Joint US–CERN–Japan International School, Tsukuba, 1996 (World Scientific, Singapore, 1999).

[8] P. Baudrenghien, Low-level RF systems for synchrotrons, Part I: Low intensity case, CAS course on RF Engineering, Seeheim, May 8-16 2000, these proceedings. (Also published as SL-Note-2001-008 HRF, 2001).

[9] F. Pedersen, Beam loading effects in the CERN PS booster, *IEEE Trans. Nucl. Sci.* **NS-22** (1975) No. 3.

[10] F.J. Sacherer, A longitudinal stability criterion for bunched beams, *IEEE Trans. Nucl. Sci.* **NS-20** (1973) No. 3.

[11] D. Boussard, G. Dôme, T.P.R. Linnecar, Acceleration in the CERN SPS, present status and future developments, *IEEE Trans. Nucl. Sci.* **NS-26** (1979) No. 3.

[12] P. Baudrenghien, Reducing the impedance of the travelling wave cavities: feed-forward and one turn delay feed-back, Proc. Workshop on LEP-SPS Performance, Chamonix, 2000, CERN-SL-2000-007 DI.

[13] P. Baudrenghien, Control of strong beam loading: results with beam, Proc. Workshop on LEP-SPS Performance, Chamonix, 2001, CERN-SL-2001-003 DI.

[14] F.E. Terman, *Radio Engineering* (McGraw-Hill, London, 1951).

[15] D. Boussard, H.P. Kindermann, V. Rossi, RF feedback applied to a multicell superconducting cavity, Proc. EPAC, Rome, Italy, 1988.

[16] Gene F. Franklin, J. David Powell, Abbas Emami-Naeini, *Feedback Control of Dynamic Systems* (Addison-Wesley, Reading, MA, 1994).

[17] R.C. Sah, J.R. Chen, C.C. Kuo, and G.H. Luo, Status report on the Synchrotron Radiation Research Center, Proc. APAC, Tsukuba, 1998.

[18] G.H. Luo, L.H. Chang, C.C. Kuo, M.C. Lin, R. Sah, T.T. Yang, Ch. Wang, The superconducting RF cavity and 500 mA beam current: upgrade project at Taiwan Light Source, Proc. EPAC, Vienna, 2000.

[19] R. Garoby, D. Grier, E. Jensen, A. Mitra, R.L. Poirier, The PS 40 MHz bunching cavity, Proc. PAC, Vancouver, 1997, CERN/PS/97-03.

[20] D. Boussard, G. Lambert, Reduction of the apparent impedance of wide band accelerating cavities by RF feedback, *IEEE Trans. Nucl. Sci.* **NS-30** (1983) No. 4.

[21] F. Pedersen, RF cavity feedback, CERN PS 92-59 RF (1992).

[22] F. Blas, R. Garoby, Design and operational results of a 'one-turn-delay feedback' for beam loading compensation of the CERN PS ferrite cavities, Proc. PAC, San Francisco, CA, 1991.

[23] J.D. Fox, Bunch feedback systems and signal processing, Proc. Joint US–CERN–Japan–Russia School on Particle Accelerators, Montreux, 1998.

[24] E. Kikutani, M. Tobiyama, S. Kurokawa, Development of a high-speed digital signal-process board for the KEKB bunch feedback systems, Proc. EPAC, Sitges, 1996.

[25] R. Garoby, Low level RF and feedback, in *Frontiers of Accelerator Technology*, Proc. Joint US–CERN–Japan International School, Tsukuba, 1996 (World Scientific, Singapore, 1999).

[26] P.B. Kenington, *High Linearity RF Amplifier Design* (Artech House, 2000).

[27] A. Mosnier, F. Orsini, B. Phung, Analysis of the heavily beam-loaded SOLEIL RF system, Proc. EPAC, Stockholm, 1998.

[28] *Mixed-Signal and DSP Design Techniques* (Analog Devices, 2003).

[29] G. Dôme, The SPS accelerating system, travelling wave drift-tube structure for the CERN SPS, CERN-SPS/ARF/77-11 (1977).

[30] E. Kikutani, J. Flanagan, M. Tobiyama, Limitations of multibunch feedback systems and extrapolation, Proc. EPAC, Vienna, 2000.

[31] M. Tobiyama, E. Kikutani, T. Obina, Y. Minagawa, T. Kasuga, Initial test of a bunch feedback system with a two-tap FIR filter board, Proc. 7th Beam Instrumentation Workshop (BIW96), Argonne National Laboratory, IL, 1996 (KEK Preprint 96-22, 1996).

[32] Y. Minagawa, E. Kikutani, S. Kurokawa, M. Tobiyama, Study of a transverse bunch-by-bunch feedback system based on the two tap FIR filter, *Nucl. Instrum. Methods Phys. Res.* **416** (1998).

[33] FIR Compiler MegaCore Function, version 1.12, ALTERA, http://www.altera.com .

[34] V. Rossi, CERN SL/HRF, private communication.

[35] J.M. Brennan, A. Campbell, J. DeLong, T. Hayes, E. Onillon, J. Rose, K. Vetter, RF beam control system for the Brookhaven Relativistic Heavy Ion Collider RHIC, presented at EPAC, Stockholm, 1998.

[36] *ADSP-21000 Family, Applications Handbook Vol. 1* (Analog Devices, 1995).

[37] A. Oppenheim, R. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

[38] Texas Instruments, Technology Innovations Bulletin **8** March 2001, http://www.ti.com/sc/techinnovations8 .

[39] N.J. Fliege, *Multirate Digital Signal Processing* (Wiley, Chichester, 1994).