RF applications in digital signal processing

T. Schilcher
Paul Scherrer Institut, Villigen, Switzerland

Abstract
Ever higher demands for stability, accuracy, reproducibility, and monitoring capability are being placed on Low-Level Radio Frequency (LLRF) systems of particle accelerators. Meanwhile, continuing rapid advances in digital signal processing technology are being exploited to meet these demands, thus leading to development of digital LLRF systems. The first part of this course will begin by focusing on some of the important building-blocks of RF signal processing including mixer theory and down-conversion, I/Q (amplitude and phase) detection, digital down-conversion (DDC) and decimation, concluding with a survey of I/Q modulators. The second part of the course will introduce basic concepts of feedback systems, including examples of digital cavity field and phase control, followed by radial loop architectures. Adaptive feed-forward systems used for the suppression of repetitive beam disturbances will be examined. Finally, applications and principles of system identification approaches will be summarized.

1 Introduction

In this lecture we shall examine various aspects of Low-Level Radio Frequency (LLRF) systems for particle accelerators. These LLRF systems typically collect measurements of amplitude, phase, and frequency, perform various signal processing and computational processes on that data, and then use the outcome to monitor, control, and regulate RF fields in particle accelerators. The LLRF applications can be classified into monitoring, feedback, and feed-forward systems. The range of typical RF frequencies in particle accelerators varies widely starting from a few MHz (e.g., in ion accelerators) and going up to tens of GHz (e.g., high-gradient acceleration structures operating at 30 GHz). The variety of accelerators demands LLRF systems with different requirements. The boundary conditions for the LLRF systems depend on their field of application, e.g., linear or circular machines, normal or super-conducting RF systems, pulsed or CW operation, or even electron or hadron accelerators. A short incomplete list of typical LLRF applications is given in Table 1. The rapid advances in digital technology over the last decade have prepared the ground for applications of digital systems in the field of LLRF. The question arises why implementations of digital systems have found their way into LLRF applications to such a large extent. There are some disadvantages with respect to their analog counterparts but on the other hand, they offer many advantages. A comparison of digital and analog LLRF systems is given in Table 2 [1]. In feedback systems, the biggest advantage of analog systems is their short loop latency. Whenever high loop bandwidths are required, analog systems still continue to be the only realistic alternative.

Table 1: Typical low-level RF applications

Application	Task
Cavity field loops	Control amplitude and phase of accelerating RF fields in cavities
Tuner loops	Control resonant frequencies of RF cavities
Radial and phase loops	Control radial beam positions and phases in circular machines
Klystron loops	Control amplitude and phase of klystrons
RF gymnastics	Control bunch splitting and merging in circular machines

Table 2: Comparison of digital versus analog RF applications

	Digital	Analog
implementation	learning curve + software effort	easier / known
latency	longer	short
data acquisition/control	(I/Q) sampling (also direct) or Digital Down Conversion (DDC)	amplitude/phase, IF down conversion
algorithms	sophisticated state machines, exception handling	simple linear, time-invariant (example: PID control)
multi-user	full	limited
remote control + diagnostics	easy, often no additional hardware necessary	difficult, often extra HW necessary
flexibility/reconfigurability	high, easy upgrades	limited
drift/tolerance	no drifts, repeatability	drift (temperature, etc.), component tolerance
signal transport distance without distortions	longer	short
radiation sensitivity	high	small

However, the increasing processing speed of digital systems now often allows them to cover the loop bandwidth demands and therefore supersede analog systems.

To characterize digital LLRF systems, we can break them down into four main building blocks which are shown in Fig. 1. RF signals from the accelerator, e.g., from probe antennas or directional couplers, are down-converted (if necessary) and conditioned in the first block, which typically includes filtering, amplification and/or attenuation. The second block comprises the digitization process, while

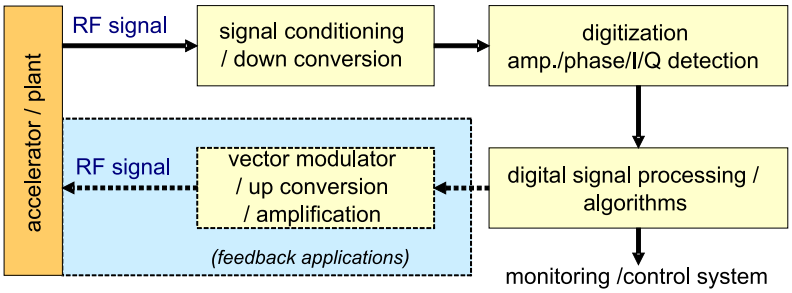


Fig. 1: Main blocks of RF applications. Feedback and feed-forward systems close the loop (blue box) while monitoring applications provide the processed RF field information to other sub-systems or to the control system.

the third block deals with the digital signal processing of the sampled RF fields. Depending on the hardware and algorithms, the extracted information is supplied to the control system for monitoring purposes or to any other sub-system requiring this information. In case of feedback applications, which close the signal path, the correction signals resulting from the control algorithms usually need to be converted back to analog RF signals of precise amplitudes, phases, and frequencies. This can be achieved by digital-to-analog converters and various up-conversion schemes—very often so-called vector modulators are used. Basic principles of these techniques will be described in the section covering the fourth building block. These four basic building blocks which form LLRF systems also form the basis of many other accelerator applications, e.g., diagnostic applications like beam-position monitoring measurements, or-

bit feedbacks, and bunch-by-bunch feedback systems. Sharing of digital hardware platforms provides grounds for common development. As in every feedback system, the ultimate remaining error is dominated by the measurement process which includes systematic errors, accuracy, linearity, repeatability, stability, resolution, and noise. In the following sections, we shall first discuss these four building blocks before proceeding on to an introduction to feedback systems, adaptive feed-forward algorithms, and system identification.

2 Signal conditioning and down conversion

Digital systems require the transformation of analog signals into the digital domain. This is achieved by analog-to-digital converters (ADC, A/D converters). In most cases, analog preprocessing has to be applied prior to the A/D conversion because the sampling frequencies, analog input bandwidths, and the bit resolution of ADCs are limited. This preprocessing typically consists of amplitude scaling (amplification, attenuation) along with filtering and frequency translation to an intermediate frequency or to baseband. The digitization of high-frequency carrier signals is very often not possible or reasonable. Although today's ADCs already reach the giga-samples per second domain with even higher analog input bandwidths, the requirements for ADC clock and aperture jitter become more and more stringent. This is even more critical with under-sampling schemes in which signals are sampled outside the first Nyquist zone, which means any clock and/or aperture jitter has a large impact on the resulting error. The system designer always has to evaluate the trade-off between sampling speed and dynamic range of an ADC. Each additional effective bit of an ADC provides 6 dB more in terms of dynamic range. Very often it is better to use analog circuits in conjunction with ADCs to implement automated gain control (AGC) functions to ensure that the signal to be sampled falls within the ideal dynamic range of the chosen ADC. Taking into account these technical limits, frequency translations to lower intermediate frequencies are essential in many RF applications. This is usually achieved by RF mixers, whose basic functionality will presently be introduced. Since frequency conversions are non-linear operations, RF mixers cannot be realized by linear-time-invariant (LTI) components or circuits. Instead, the translation is achieved by either time varying or non-linear circuits, e.g., diodes. An ideal mixer is usually represented by a multiplier symbol (see Fig. 2) with two input ports and one output port. The signal at the output port is the vector

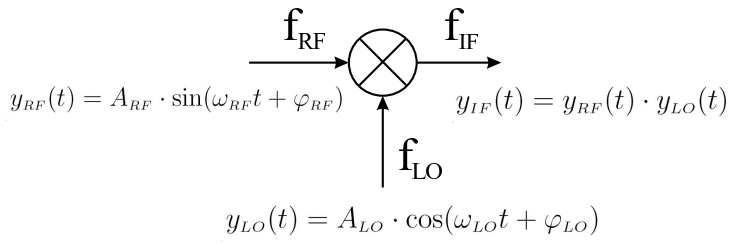


Fig. 2: Ideal RF mixer with input signals $y_{RF}(t)$ and $y_{LO}(t)$

multiplication of the signals at the two input ports. One of the input signals is the reference signal, the local oscillator (LO). Using some trigonometric product-to-sum identities and assuming two sinusoidal input signals with amplitudes A_{RF}, A_{LO} and frequencies f_{RF}, f_{LO}, the multiplied output signal y_{IF} can be represented as

$$\begin{aligned} y_{IF}(t) &= y_{RF}(t) \cdot y_{LO}(t) \\ &= \frac{1}{2} A_{LO} A_{RF} \cdot \Big(\underbrace{\sin\big[(\omega_{RF} - \omega_{LO})\, t + (\varphi_{RF} - \varphi_{LO})\big]}_{\text{lower sideband}} \\ &\qquad + \underbrace{\sin\big[(\omega_{RF} + \omega_{LO})\, t + (\varphi_{RF} + \varphi_{LO})\big]}_{\text{upper sideband}} \Big) \ . \end{aligned} \tag{1}$$

The input signal is translated to two output frequencies, the upper ($f_{RF}+f_{LO}$) and the lower ($f_{RF}-f_{LO}$) sideband signals. Mixers are used for down- and up-conversion. In the case of down-conversion, the RF and LO signals are high-frequency inputs while the resulting output signal is the intermediate frequency signal [Fig. 3(a)]. Likewise, an intermediate frequency can be used as input to perform an up-conversion [Fig. 3(b)]. Usually, either the upper or the lower sideband of the mixing process is selected by filtering

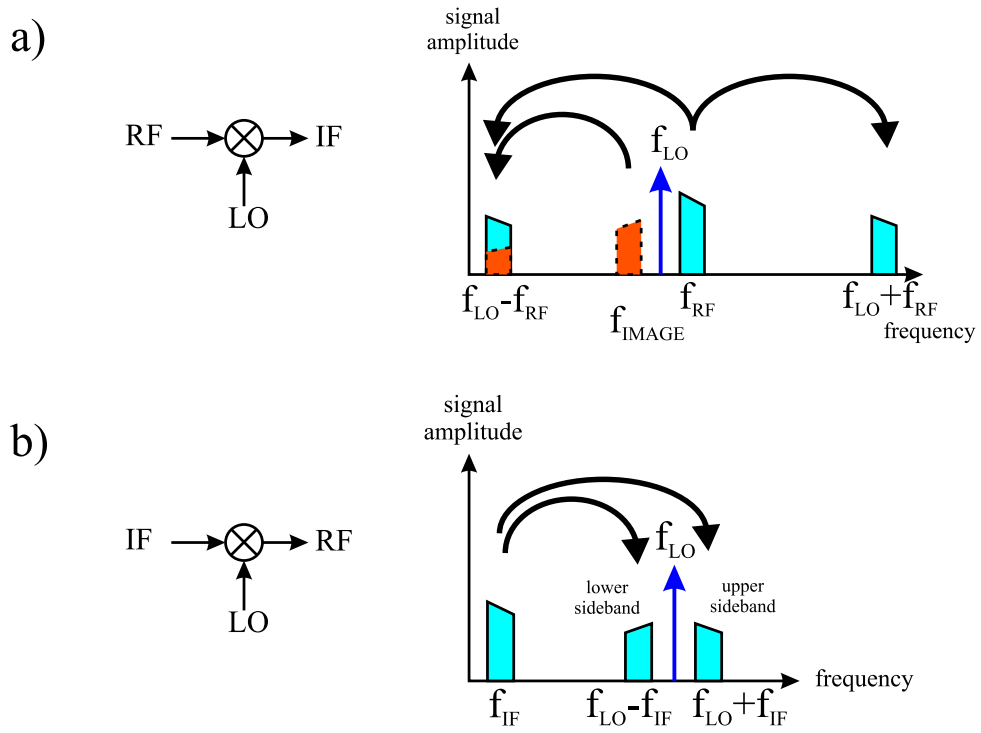


Fig. 3: (a) down conversion and (b) up conversion schemes with RF mixers

the output signal. Note that for a given LO frequency, signals at frequencies LO±IF are converted to the same IF frequency. Therefore care has to be taken by applying proper filtering in order to avoid the conversion of any noise or signal at the image frequency to the IF frequency [Fig. 3(a)], otherwise it can degrade system performance, often severely. For receiver applications, to which particular attention will now be given, the lower sideband is extracted by low-pass filtering. With Eq. (1) this yields for the IF signal

$$y_{IF}(t) = A_{IF} \cdot \sin\Big(\,\omega_{IF}t + \varphi_{IF}\Big)$$

with the frequency, amplitude and phase definition

$$\begin{aligned}
\omega_{IF} &= \omega_{RF} - \omega_{LO}, \\
A_{IF} &= \tfrac{1}{2}A_{LO}A_{RF} \quad \sim \quad A_{RF} \qquad \text{with constant } A_{LO}\ , \\
\varphi_{IF} &= \varphi_{RF} - \varphi_{LO} \quad \sim \quad \varphi_{RF} \qquad \text{with constant } \varphi_{LO}\ .
\end{aligned}$$

With a constant LO signal, i.e., with a constant A_{LO} and φ_{LO}, the amplitude and phase of the down-converted signal are directly proportional to those of the input signal, in other words, all of the basic properties of an RF signal are conserved in the frequency conversion process. In this context, two important facts should be highlighted: first of all, a phase change or phase jitter at f_{RF} is exactly the same phase change at the IF frequency f_{IF}. As an example, a 1° phase change at 3 GHz corresponds also to a 1° phase change at any arbitrary IF frequency. Secondly, timing jitter of a given magnitude at a higher frequency leads to a bigger phase variation than at a lower frequency. This has to be taken into account

when deciding whether an RF signal should be directly sampled or down-converted first. ADC clock jitters are therefore much more critical in direct sampling applications and tougher requirements have to be applied. As an illustration, a 10 ps clock jitter at 500 MHz corresponds to 1.8° phase variation while the same clock jitter results in only 0.18° phase variation at 50 MHz.

So far, ideal mixers have been considered. As pointed out earlier, real mixers are non-linear devices. In addition to the desired ideal mixing product, their output spectra contain many undesired spurious signals which are located at $mf_{RF} \pm nf_{LO}$. Figure 4 depicts the output spectrum of a down-conversion mixer. The image frequency is denoted by IM. Usually, the signal at the mixer output is

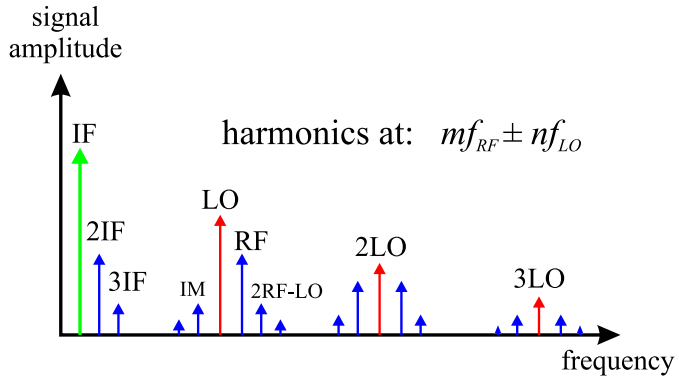


Fig. 4: Output spectrum of a real mixer

filtered by a low-pass filter in order to suppress the undesired spurs sufficiently. Introducing low-pass filters with a steep roll-off, however, will mean a trade-off between good suppression and increased group delay which is very important for low latency and feedback applications. Other common ways to reduce spurious outputs are the use of double-balanced mixers and image rejection mixers to minimize image mixing.

3 Detection of amplitude and phase in digital systems

Several techniques can be applied to detect amplitude and phase of RF signals in digital systems, an overview of which is shown in Fig. 5. Common amplitude detectors like Schottky diodes and phase detectors [2] provide amplitude and phase information as a continuous analog signal which can be digitized by ADCs for further processing [Fig. 5(a)]. In contrast to this, the polar representation of an RF signal can also be decomposed into its Cartesian representation (I/Q, see Section 3.1) by analog IQ demodulators. The individual I/Q components can then be digitized [Fig. 5(b)]. A third alternative is the so-called digital IQ sampling and Digital Down Conversion (DDC) where an RF signal is first down-converted to an intermediate frequency and then directly digitized. Depending on the frequency, the down-conversion process can be omitted. The I/Q information is extracted from these samples in a purely digital way. These techniques will be further discussed in Section 3.1. Since the focus of this lecture is on digital signal processing, the first two detection schemes will not be dealt with any further.

3.1 *IQ* sampling

The terminology 'IQ' originates from the representation of an RF signal which can be in either polar (amplitude/phase representation) or in Cartesian coordinates. Any sinusoidal RF signal

$$y(t) = A \cdot \sin(\omega t + \varphi_0) \tag{2}$$

can be modelled as a phasor which is a rotating vector with amplitude A, frequency ω, and an initial phase φ_0. A positive frequency corresponds to an anticlockwise rotating phasor, which can be decomposed into

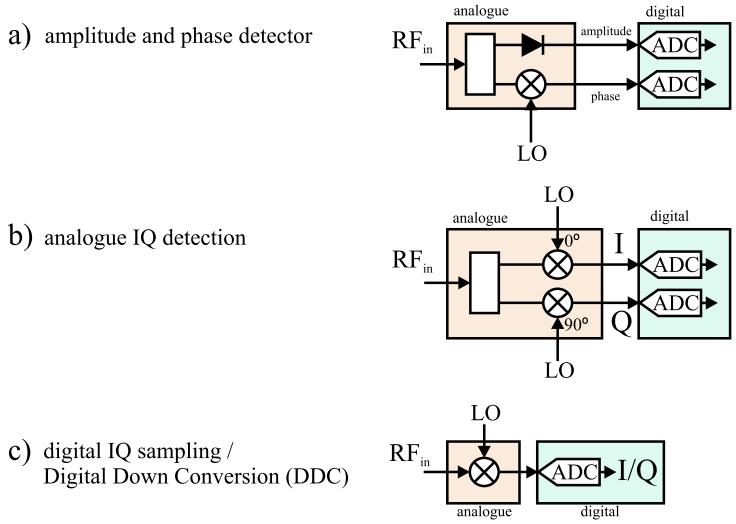
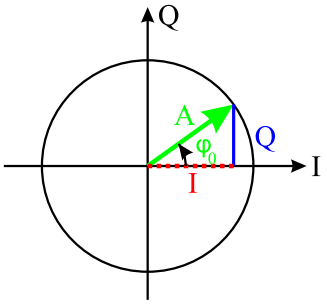


Fig. 5: different schemes of amplitude and phase detection in digital systems

its sin and cos components using basic trigonometric functions.

$$\begin{aligned} y(t) &= \underbrace{A\cos\varphi_0}_{=:I}\sin\omega t + \underbrace{A\sin\varphi_0}_{=:Q}\cos\omega t \\ y(t) &= I\cdot\sin\omega t + Q\cdot\cos\omega t\,. \end{aligned} \tag{3}$$

The amplitude of the sine component can be defined as the *in-phase* component (I), while the amplitude of the cosine component is called the *quadrature-phase* component (Q). This definition is somewhat arbitrary and can therefore be found inversed in the literature. The Cartesian representation of an RF signal (I/Q) is widely used in numerical applications. The well-known relationship between Cartesian and polar coordinates [Eq. (4)] makes switching between both representations possible—if necessary, and if computing resources allow the time-consuming calculation of square roots and trigonometric functions. Sampling of a sinusoidal RF signal with an ADC delivers only one component of the rotating phasor



$$\begin{aligned} I &= A\cdot\cos\varphi_0 & A &= \sqrt{I^2+Q^2} \\ Q &= A\cdot\sin\varphi_0 & \varphi_0 &= \operatorname{atan}\left(\frac{Q}{I}\right) \end{aligned} \tag{4}$$

Fig. 6: Phasor representation of RF signal

at a time. Nevertheless it is both easy and possible to extract I/Q information based on the sampled data stream, as will now be shown. We assume that the vertical component is measured by the ADC (this assumption is also arbitrary but it has no consequence on the following results). The goal of sampling the RF or IF signal is to extract its amplitude/phase or I/Q information, respectively. If I/Q of the rotating phasor is known initially and measured again later at a well-defined time, an algorithm can rotate the phasor back to an initial reference phasor. A comparison of these two measurements will then reveal whether the incoming RF signal has changed its amplitude and phase. The usual IQ sampling is

achieved if the sampling frequency f_s and the intermediate frequency f_{IF} (or in general, the incoming RF signal) are related by

$$f_s = 4 \cdot f_{IF} \, . \tag{5}$$

In this case, the phase advance between two samples amounts to 90°. With Eq. (3) the digitization at four consecutive sampling times results in

$$\begin{array}{ll} \omega t_0 = 0 : & y(t_0) = \quad Q \\ \omega t_1 = \pi/2 : & y(t_1) = \quad I \\ \omega t_2 = \pi : & y(t_2) = -Q \\ \omega t_3 = 3\pi/2 : & y(t_3) = -I \, . \end{array} \tag{6}$$

This is illustrated in Fig. 7. At each sampling time, the I and Q components of the phasor can be determined by two consecutive samples, $y(t_i)$ and $y(t_{i+1})$. These vectors have to be rotated backwards

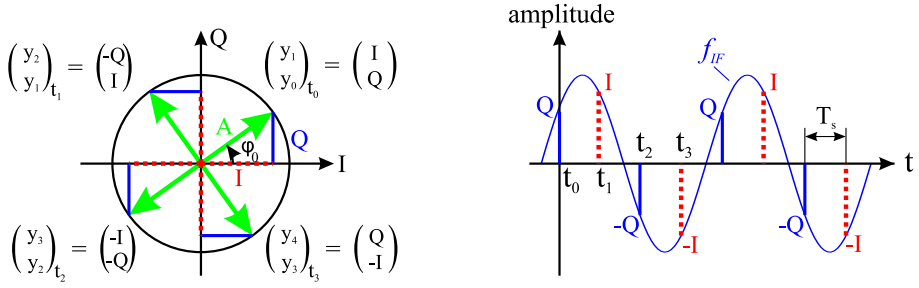


Fig. 7: IQ sampling with 90° phase advance between consecutive samples

by $\Delta\varphi_i = 0°$, –90°, –180° and –270° in order to compare them with the initial vector. The underlying rotation algorithm is

$$\begin{pmatrix} I \\ Q \end{pmatrix}_{t_i} = \begin{pmatrix} \cos\Delta\varphi_i & -\sin\Delta\varphi_i \\ \sin\Delta\varphi_i & \cos\Delta\varphi_i \end{pmatrix} \cdot \begin{pmatrix} y_{i+1} \\ y_i \end{pmatrix} . \tag{7}$$

The rotation algorithm is carried out at the sample rate f_s. It is based on the assumption that the incoming RF signal does not change its amplitude and phase substantially between two samples since the components of the (I/Q) phasor are based on two successive samples. If we were to choose to measure the horizontal component of the phasor with an ADC, the result would be exactly the same, which can easily be seen in Fig. 7. Instead of the sequence Q, I, –Q, –I we would measure the sequence I, –Q, –I, Q which only corresponds to a phase shift of 90°.

The restriction to choose a sampling frequency four times the carrier frequency simplifies the calculation since the elements of the rotation matrix only consist of 1, 0 and –1. Despite this fact, it is possible in principle to choose the sampling frequency such that it is a multiple of the IF frequency.

$$f_s/f_{IF} = m, \qquad m : \text{integer} \, .$$

In this case, the phase difference between two samples amounts to

$$\Delta\varphi = \frac{2\pi}{m} \, . \tag{8}$$

As long as the input signal does not change substantially, each IF period is sampled at the same location on the circle in the IQ plane. With two consecutive samples ($y_n = Q_n$, $y_{n+1} = Q_{n+1}$) and the rotation of the I/Q vector at time t_{n+1} to the vector at time t_n by the angle $-\Delta\varphi$, a set of linear equations is obtained.

$$\begin{pmatrix} I_n \\ Q_n \end{pmatrix} = \begin{pmatrix} \cos\Delta\varphi & \sin\Delta\varphi \\ -\sin\Delta\varphi & \cos\Delta\varphi \end{pmatrix} \cdot \begin{pmatrix} I_{n+1} \\ Q_{n+1} \end{pmatrix} \, .$$

Solving this set of equations to eliminate I_{n+1}, Q_{n+1} yields

$$\begin{pmatrix} I_n \\ Q_n \end{pmatrix} = \frac{1}{\sin\Delta\varphi} \cdot \begin{pmatrix} 1 & -\cos\Delta\varphi \\ 0 & \sin\Delta\varphi \end{pmatrix} \cdot \begin{pmatrix} y_{n+1} \\ y_n \end{pmatrix} .$$

Similar to the case where the IF frequency is sampled four times per period, the phasor (I_n, Q_n) has to be rotated backwards to the initial phasor (I_0, Q_0) by a rotation angle $-n\Delta\varphi$ where the phase advance $\Delta\varphi$ is defined in Eq. (8).

$$\begin{pmatrix} I_0 \\ Q_0 \end{pmatrix} = \frac{1}{\sin\Delta\varphi} \cdot \begin{pmatrix} \cos n\Delta\varphi & -\cos(n+1)\Delta\varphi \\ -\sin n\Delta\varphi & \sin(n+1)\Delta\varphi \end{pmatrix} \cdot \begin{pmatrix} y_{n+1} \\ y_n \end{pmatrix} .$$

In the most general case, the initial phasor can also be rotated by a user defined angle $-\varphi$ in order to match initial conditions. Including this last rotation the (I,Q) vector can be obtained by

$$\begin{pmatrix} I \\ Q \end{pmatrix} = \frac{1}{\sin\Delta\varphi} \cdot \begin{pmatrix} \cos(\varphi + n\Delta\varphi) & -\cos(\varphi + (n+1)\Delta\varphi) \\ -\sin(\varphi + n\Delta\varphi) & \sin(\varphi + (n+1)\Delta\varphi) \end{pmatrix} \cdot \begin{pmatrix} y_{n+1} \\ y_n \end{pmatrix} . \tag{9}$$

The calculation of I and Q belongs to the class of FIR filters of first order since the values are based on two consecutive samples. This sequence of processing is illustrated in Fig. 8. Although the phase

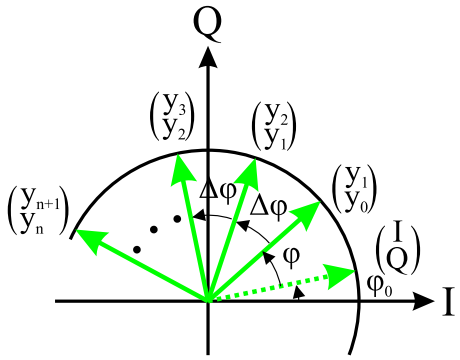


Fig. 8: IQ sampling with a phase advance of $\Delta\varphi = 2\pi/m$ between consecutive samples

advance can be chosen as illustrated in Eq. (8), the calculation of I and Q becomes extremely sensitive to noise and sampling errors if the phase advance is far away from 90° or 270° [see Eq. (9)].

There are a number of potential problems associated with this typical mode of IQ sampling ($\Delta\varphi$ = 90°). DC offsets at the input of the ADC or phase advances between samples (which differ from the expected 90°) systematically result in a ripple with frequency f_{IF}. These errors are easily detectable since their signatures reveal their origin. Other error sources are much more difficult to identify at first glance. Non-linearities of mixers and differential non-linearities of ADCs generate higher harmonics of the input signal frequency f_{IF}. Sampling this signal four times per period maps the second harmonic on the Nyquist frequency, while the third harmonic aliases directly on the IF frequency again. In general, all odd harmonics alias to the carrier frequency while even harmonics alias to DC or to the Nyquist frequency, respectively (see Fig. 9). Any harmonics which map onto the IF frequency are indistinguishable from the carrier by standard IQ sampling. As the carrier frequency and phase changes, the distortion changes, resulting in false measurement data.

3.2 Non-IQ sampling

A solution to the problems described is to change the sampling frequency in such a way that the following relation holds:

$$f_s = \frac{N}{M} \cdot f_{IF} \quad \Longleftrightarrow \quad N \cdot T_s = M \cdot T_{IF} \qquad \text{with } M, N\text{: integer} .$$

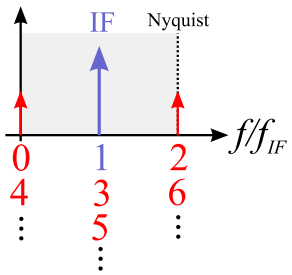


Fig. 9: Spectrum of IQ sampling with a phase advance of 90° between consecutive samples

The above is equivalent to taking N samples in M successive IF periods. Depending on the values chosen for M, N the circle in the IQ plane is now sampled at different locations where the samples only repeat after M IF periods. The phase advance between two consecutive ADC readings is

$$\Delta\varphi = \omega_{IF} T_s = 2\pi \frac{T_s}{T_{IF}} = 2\pi \frac{M}{N} \; .$$

Expressing the N successive samples with Eq. (3), the result is a set of linear equations [3], [4].

$$\begin{array}{lcll} y_0 & = & I \cdot \sin\varphi_0 & + \; Q \cdot \cos\varphi_0 \\ y_1 & = & I \cdot \sin\varphi_1 & + \; Q \cdot \cos\varphi_1 \\ y_2 & = & I \cdot \sin\varphi_2 & + \; Q \cdot \cos\varphi_2 \qquad \text{with } \varphi_i = i \cdot \Delta\varphi = i \cdot 2\pi \frac{M}{N}, \; i\text{: integer} \; . \\ \ldots & & & \\ y_{(N-1)} & = & I \cdot \sin\varphi_{(N-1)} & + \; Q \cdot \cos\varphi_{(N-1)} \end{array} \tag{10}$$

In the ideal case of no measurement errors, any two equations of this set would give a solution common to both for I, Q. Under real conditions with noise, quantization errors, and clock jitter, the over-constrained system of equations can be solved by applying the Least Mean Square algorithm (LMS). The basis for this algorithm is to minimize the function

$$f(I,Q) = \sum_{i=0}^{N-1} \left(I \cdot \sin\varphi_i + Q \cdot \cos\varphi_i - y_i \right)^2 \; . \tag{11}$$

The conditions for a minimum are vanishing derivatives with respect to I and Q.

$$\frac{\partial f}{\partial I} = 0, \qquad \frac{\partial f}{\partial Q} = 0 \; .$$

This yields

$$\begin{array}{lcl} \dfrac{\partial f}{\partial I} & = & 2I \displaystyle\sum_{i=0}^{N-1} \sin^2\varphi_i + 2Q \sum_{i=0}^{N-1} \sin\varphi_i \cos\varphi_i - 2 \sum_{i=0}^{N-1} y_i \sin\varphi_i = 0 \\ \dfrac{\partial f}{\partial Q} & = & 2Q \displaystyle\sum_{i=0}^{N-1} \cos^2\varphi_i + 2I \sum_{i=0}^{N-1} \sin\varphi_i \cos\varphi_i - 2 \sum_{i=0}^{N-1} y_i \cos\varphi_i = 0 \; . \end{array} \tag{12}$$

With the definitions

$$p_{11} = \sum_{i=0}^{N-1} \sin^2\varphi_i, \qquad p_{12} = p_{21} = \sum_{i=0}^{N-1} \sin\varphi_i \cos\varphi_i, \qquad p_{22} = \sum_{i=0}^{N-1} \cos^2\varphi_i$$

$$s_1 = \sum_{i=0}^{N-1} y_i \sin\varphi_i, \qquad s_2 = \sum_{i=0}^{N-1} y_i \cos\varphi_i \ ,$$

Eq. (12) can be written as

$$\begin{aligned} p_{11}\cdot I \ &+ \ p_{12}\cdot Q \ = \ s_1 \\ p_{21}\cdot I \ &+ \ p_{22}\cdot Q \ = \ s_2 \ . \end{aligned}$$

The coefficients p_{12} and p_{21} can be simplified using the definition for φ_i of Eq. (10).

$$p_{12} = p_{21} = \frac{1}{2}\sum_{i=0}^{N-1} \sin 2\varphi_i = \frac{1}{2}\underbrace{\sum_{i=0}^{N-1} \sin 4\pi M \frac{i}{N}}_{=0} = 0 \ .$$

Likewise, the coefficients p_{11} and p_{22} can be calculated to

$$p_{11} = \frac{1}{2}\sum_{i=0}^{N-1}(1-\cos 2\varphi_i) = \frac{N}{2} \ , \qquad p_{22} = \frac{1}{2}\sum_{i=0}^{N-1}(1+\cos 2\varphi_i) = \frac{N}{2} \ .$$

Therefore the I and Q values can be simply calculated by

$$I \ = \ \frac{2}{N}\cdot\sum_{i=0}^{N-1} y_i \cdot \sin(i\cdot\Delta\varphi) \tag{13}$$

$$Q \ = \ \frac{2}{N}\cdot\sum_{i=0}^{N-1} y_i \cdot \cos(i\cdot\Delta\varphi) \ . \tag{14}$$

It should be noted that the sine and cosine terms can be precalculated and stored in lookup tables since the phase advance is constant and defined in Eq. (10). Owing to the non-IQ sampling, most of the harmonics no longer line up with the carrier frequency, which can be seen in Fig. 10. Digital filtering can now be

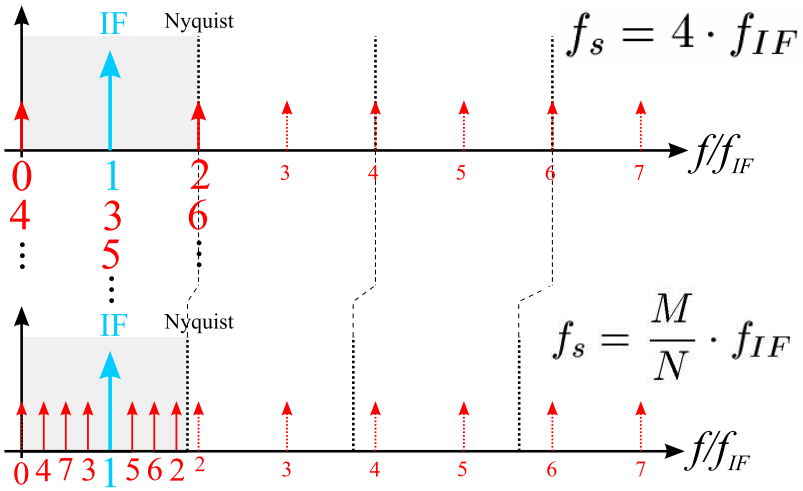


Fig. 10: Spectrum of non-IQ sampling for the case $M = 15$, $N = 4$

applied to separate the carrier from the harmonics. Because of the sampling of the carrier frequency at several locations on the circle in the IQ plane, errors from ADC non-linearities, quantization, DC offsets, and clock jitter are reduced. In contrast to IQ sampling, the calculation of I, Q by Eq. (13) and Eq. (14) results in more latency. Special care has to be taken if a digital filter is introduced to distinguish the carrier from the harmonics because of the resulting additional group delay. A trade-off between improved linearity and low latency has to be made.

3.3 Digital Down Conversion—DDC

Another possible way of extracting amplitude and phase information from a sampled carrier signal is by means of Digital Down Conversion (DDC), sometimes also referred to as Digital Drop Receiver (DDR). In many RF applications, the band of interest of an RF signal lies around a carrier or IF frequency. The frequency band of interest is very often narrow compared to the carrier frequency itself. According to the Nyquist theorem, the sampling rate of a band-limited signal must be at least twice the highest analog frequency component in order to be able to reconstruct the signal fully. This leads to high data rates and imposes high demands on subsequent digital processing stages. However, Nyquist–Shannon's theorem states that the sampling frequency only has to be at least twice the maximum information bandwidth. Digital down conversion is a technique which takes a band-limited high-sample-rate digitized signal, shifts the band of interest to a lower frequency and reduces the sample rate while retaining all the information. The basic function of DDC is shown in Fig. 11. To illustrate the DDC principle, we assume

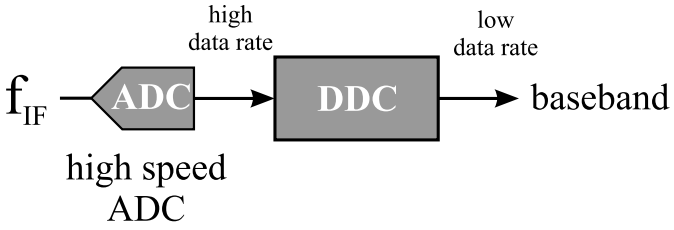


Fig. 11: Basic principle of Digital Down Conversion (DDC)

a 40 MHz carrier (IF signal). A band-limited signal with a bandwidth of 1 MHz is located around the carrier frequency. If oversampling is applied, sampling has to be carried out with a frequency of at least 82 MHz. However, an output sample rate of just 2 MHz would be sufficient to meet the requirement of Shannon's theorem if the band of interest is first shifted to baseband.

Although DDC can be implemented on Digital Signal Processors (DSPs), the main application area is on Application Specific Integrated Circuits (ASICs) or Field Programmable Gate Arrays (FPGAs). DDC can be divided into two classes, wideband and narrowband, based on their decimation ratio. The sampling rate of wideband signals is typically reduced by a smaller amount (decimation ratio < 32). This can be achieved by FIR or so-called multi-rate FIR filters. Narrowband DDC provides large decimation ratios of more than 32, consequently their implementation on FPGA with FIR filters would take too many gates. A better approach is the use of Cascaded Integrator Comb (CIC) filters (see later in this lecture). A digital down converter is composed of three major building blocks (see Fig. 12) which will be addressed in the following sub-sections. The digital mixers convert the incoming digitized RF signal

DDC building blocks:

- mixer (digital mixer)
- local oscillator (numerically controlled oscillator, NCO)
- decimating low-pass filter (LPF)

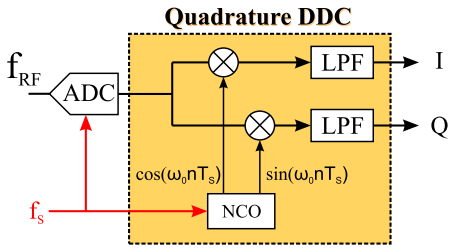


Fig. 12: Main building blocks of Digital Down Conversion

down to baseband and are realized as ideal multipliers, just like their ideal analog counterpart. The local oscillator inputs are supplied by numerically controlled oscillators (NCOs). As shown in Section 2, the resulting output consists of the sum and the difference frequency with respect to the input frequency. The purpose of the subsequent low-pass filter is to suppress the sum frequency component and to provide

anti-aliasing filtering. This is necessary in order to limit the signal spectrum prior to decimation. The decimating anti-aliasing filter is often implemented as CIC filter and/or FIR filter.

3.3.1 Numerically Controlled Oscillator—NCO

The numerically controlled oscillator in a DDC is implemented as a quadrature digital oscillator which generates a stream of sine and cosine samples. If the frequency of the NCO is chosen to match the carrier frequency, the difference signals at the outputs of the digital mixers represent the I and Q components of the input signal. One of the advantages of the NCO is that the two sine and cosine output signals are perfectly synchronized, since they are derived from a common clock. Moreover, the signals have a very precise 90° phase shift. There are several ways to implement a NCO. Its basic functionality is that of a phase accumulator and a phase-to-amplitude conversion block (see Fig. 13). A programmable

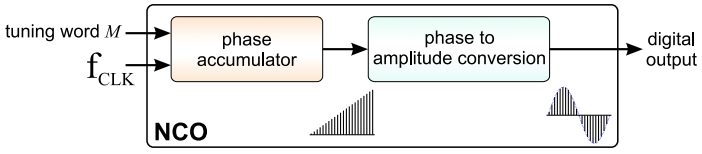


Fig. 13: Basic principle of a Numerically Controlled Oscillator (NCO)

phase increment is added within the phase accumulator at each clock cycle. This phase increment is determined by the tuning word M. The resulting total phase is then converted to the corresponding amplitude value taken from a memory-based lookup table in which a sine wave is stored. A phase roll-over is achieved easily if the table size is chosen to be a binary multiple. In this case, the roll-over takes place automatically. Numerically controlled oscillators have several advantages compared to their analog counterparts. First of all, the tuning word is programmable, which means that almost every frequency up to nearly half the clock frequency (Nyquist frequency) can be chosen by means of the software. NCOs have extremely fast hopping speeds in tuning the output frequency, which means they are extremely frequency agile. The frequency change occurs immediately after loading a new phase increment into the tuning word M. In addition, this technique provides a phase-continuous frequency hop without over- or undershoot and without the loop settling time anomalies encountered in analog circuit variable oscillators. The sine lookup table can contain a full sine wave or just one quarter of one; this is sufficient for reconstruction of the sine wave and reduces the amount of memory space required (however, the additional calculations necessary to end up with the correct amplitude have to be traded off against the saving of the memory space). The total size of the memory is dependent on the desired resolution (output amplitude bit-width) of the entries and on the number of entries stored in the table. The achievable output frequency f_{out} is a function of the clock frequency f_{CLK}, the tuning word M, the bit width of the phase accumulator N, and is given by

$$f_{out} = M \cdot \frac{f_{CLK}}{2^N} \, .$$

NCOs can reach very high frequency resolutions if the word length of the phase accumulator is chosen properly. As an example, if a clock frequency of 50 MHz is assumed and the phase accumulator width is set to 32-bits, the resulting frequency tuning resolution will be Δf= 12 mHz. On the other hand, the question that needs to be addressed is whether the same number of entries in the lookup table is needed as phase outputs are possible. Sticking to our example of 32-bit phase accumulation, and assuming an amplitude bit width of 8-bits, we would require a 4 GByte lookup table. One way out of such an impractical implementation is the method of *phase truncation*. Before the resulting phase of the accumulator is looked up in the table, it is truncated to a defined number of upper bits while the phase in the accumulator is preserved. This reduces the need for large memories tremendously. However, one consequence of this approach is the introduction of errors in the phase-to-amplitude conversion process. Regardless of the

chosen tuning word, these truncation errors are periodic and appear as line spectra in the frequency domain (although certain tuning words result in no phase truncation errors at all, which should be obvious). These lines in the spectra are known as phase truncation spurs which reduce the Spurious Free Dynamic Range (SFDR) of the output spectrum.

Although NCOs for digital down conversion do not require more elements than are shown in the block diagram shown in Fig. 13, we shall briefly extend this chapter to *Direct Digital Synthesis* for the sake of completeness. Direct Digital Synthesis or DDS is based on the same building blocks as used for NCO but is extended by a DAC to generate freely programmable, high-frequency, analog waveforms (see Fig. 14). The staircase-like output of the DAC (zero-order-hold function) implies that the output

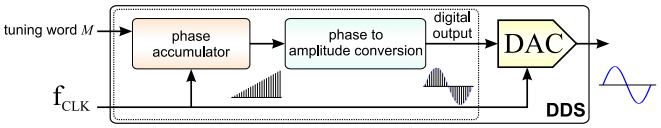


Fig. 14: basic principle of Direct Digital Synthesis (DDS)

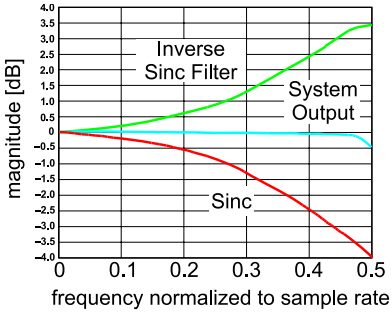
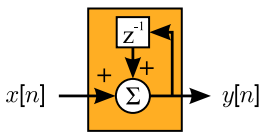


Fig. 15: inverse sin(x)/x correction at the output of DDS

amplitude spectrum follows a $\sin(x)/x$ function. This roll-off can be quite significant if higher output frequencies are desired. A solution to this problem is to pre-compensate the roll-off by digital inverse $\sin(x)/x$ filters (also called inverse sinc filters) before the data is sent to the DAC (Fig. 15). In this way, flat output amplitudes within ± 0.1 dB over a bandwidth of 80% of the Nyquist frequency can be obtained. In modern DDS chips, the $\sin(x)/x$ correction is already built in.

3.3.2 Cascaded Integrator Comb filter—CIC

As the final blocks in the DDC application, Cascaded Integrator Comb (CIC) filters are well suited as decimating low-pass filters. They belong to the class of multi-rate FIR filter which performs decimation or interpolation. This type of filter was introduced by Eugene Hogenauer in 1981 [5]. A CIC filter is a computationally efficient implementation of a narrowband low-pass filter; they require no multiplication, as we shall now see. The two basic building blocks of a CIC filter are an integrator and a comb (see Fig. 16). The integrator is a one-pole IIR filter with a unity feedback coefficient. The difference equation

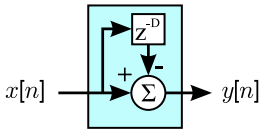


basic integrator

$$y[n] = y[n-1] + x[n]$$

transfer function:

$$H_I(z) = \frac{1}{1 - z^{-1}}$$



basic comb

$$y[n] = x[n] - x[n-D]$$

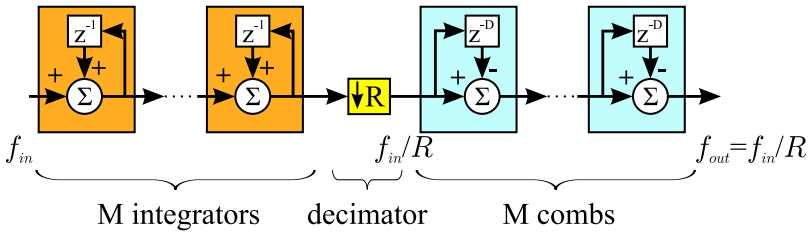
transfer function:

$$H_C(z) = 1 - z^{-D} \tag{15}$$

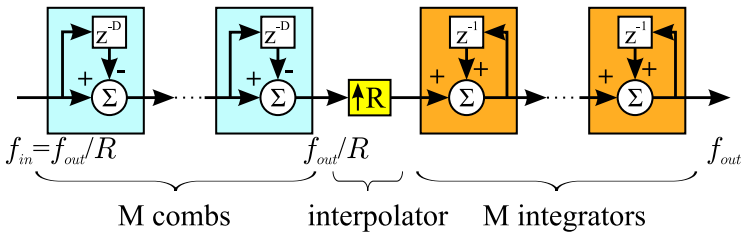
Fig. 16: Basic elements of a CIC filter

of the integrator and the corresponding transfer function are given in Fig. 16. As usual in digital signal

processing theory, the *delay by one clock* operator (z^{-1}) in the z-domain has been introduced. In a comb filter, a delayed version of the input is subtracted from itself, which causes constructive and destructive interference at the output. The resulting frequency response consists of a series of equally spaced spikes, which resemble a comb. The difference equation of the comb filter and its transfer function are also given in Fig. 16. The parameter D is often referred to as *differential delay*. In general, a cascade of M integrators and the same number of comb filters make up an M stage CIC filter. Additionally, a rate changer is inserted between integrator and comb, which changes the data rate by a factor of R [see Fig. 17(a)]. The decimation operation ↓R means to discard all but every R$^{\text{th}}$ sample. The integrators



a) decimating CIC filter



b) interpolating CIC filter

Fig. 17: Basic structure of a decimating (a) and interpolating (b) CIC filter

push date at a rate f_{in} while the comb filter stages operate at a reduced rate $f_{out} = f_{in}/R$, which minimizes power consumption in the high-speed digital hardware. Likewise, an interpolating CIC filter is composed of the same building blocks as a decimating CIC filter, only the sequence of integrators and combs are exchanged and a rate expander is inserted in between [Fig. 17(b)]. Thus this structure of CIC filters clearly reveals that only additions and subtractions are required for their implementation. The total transfer function of the decimating CIC filter is obtained by multiplying the individual transfer functions of the building blocks. However, care has to be taken with the comb transfer function since the comb stages subsequently operate at a reduced clock rate. If we refer the comb transfer function to the high input rate f_{in}, we have to replace the differential delay D in Eq. (15) by $R \cdot D$, since a delay of D at the reduced clock rate f_{out} corresponds to $R \cdot D$ clock cycles with respect to f_{in}. The transfer function of the basic comb filter after the rate changer is then given by

$$H_C(z) = 1 - z^{-RD} \; .$$

The total transfer function of an M stage CIC filter is therefore expressed by

$$H(z) = (H_I)^M (H_C)^M = \frac{(1 - z^{-RD})^M}{(1 - z^{-1})^M} = \left(\sum_{k=0}^{RD-1} z^{-k} \right)^M \; , \tag{16}$$

where the geometric sum is used to express $H(z)$ as a cascade of M FIR filters. Although initially it may have seemed that a CIC filter could be unstable—since it contains poles from the integrator—it is always stable because it is a pure FIR filter.

Up to now, the building blocks of CIC filters have been introduced without regard to any other particular motivation. An alternative approach to understanding the structure of CIC filters can be found by looking at recursive running-sum filters, which are well-known. In fact, CIC filters are derived from the concept of those filters [6]. In the following, we shall begin with a moving average filter (or box-car filter) of length N and demonstrate how this structure very closely resembles a CIC filter of first order. The difference equation of a moving average filter is given by

$$y[n] = \frac{1}{N}\Big(x[n] + x[n-1] + \cdots + x[n-N+1]\Big) = \sum_{k=n-N+1}^{n} \Big(\frac{1}{N} \cdot x[k]\Big) \,. \tag{17}$$

This is a standard FIR filter with equal coefficients. The corresponding transfer function can be easily deduced from this equation and can be expressed as

$$H(z) = \frac{1}{N}\Big(1 + z^{-1} + \cdots + z^{-(N-1)}\Big) = \frac{1}{N}\sum_{k=0}^{N-1} z^{-k} = \frac{1}{N}\frac{1-z^{-N}}{1-z^{-1}} \,. \tag{18}$$

Here the formula for a geometric series has been used to represent the sum in a more convenient way. An alternative implementation of this box-car filter is the recursive running-sum. Instead of summing up all past N elements, it is sufficient to add only the latest scaled new sample $x[n]$ to the previous average value $y[n-1]$ and to subtract the oldest scaled sample $x[n-N]$.

$$\text{recursive running-sum:} \qquad y[n] = y[n-1] + \frac{1}{N}\Big(x[n] - x[n-N]\Big)$$

The recursive running-sum can also be calculated in a slightly different, cascaded way. Rewriting Eq. (17) to

$$y[n] = \frac{1}{N}\Big(\sum_{k=-\infty}^{n} x[k] - \sum_{k=-\infty}^{n-N} x[k]\Big) \,.$$

and defining

$$w[n] := \sum_{k=-\infty}^{n} x[k]$$

yields the set of equations

$$w[n] = w[n-1] + x[n] \tag{19}$$

$$y[n] = \frac{1}{N}\Big(w[n] - w[n-N]\Big) \,. \tag{20}$$

In many applications, the box-car filter is followed by decimation of factor $R = N$ which means that only after N samples is the average value passed on for further processing. The cascaded box-car implementation given by Eq. (19) and Eq. (20) with decimation is visualized in Fig. 18. Since all operations

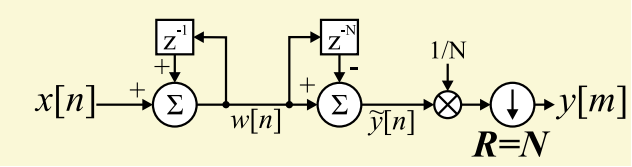


Fig. 18: Alternative representation of a recursive running-sum filter with decimation by factor N

are linear, it is possible to move the decimation from the end to just after the first integrator, i.e., decimate

$w[n]$. Only the *delay by N samples* operator (z^{-N}) has to be changed to z^{-D} with $D = N/R$ because of the rate change. If we compare the resulting transfer function

$$H_{box-car}(z) = \frac{1}{N}\frac{1-z^{-RD}}{1-z^{-1}}$$

with the one which we obtained for a first order CIC ($M = 1$) in Eq. (16), it becomes clear that box-car or recursive running-sum filters have the same transfer function as a first-order CIC—apart from the $1/N$ scaling factor and a more general differential delay D in CIC filters.

Cascaded integrator comb filters are typically employed in applications which have a large excess sample rate, i.e., in systems where the system sample rate is much larger than the bandwidth occupied by the signal. The purpose for which CIC filters are used in these applications are anti-aliasing filtering and decimation. To see the low-pass characteristic, we need to examine the frequency response. This is obtained by evaluating the transfer function along the unity circle in the z-plane by inserting

$$z = e^{i\omega T_S} = e^{i2\pi\frac{f}{f_S}}\ , \qquad f_S = f_{in}\text{: sampling or input frequency}$$

into Eq. (16). The variable f_S is the sampling frequency or the input frequency f_{in} to the CIC filter. As a result, the magnitude response of a CIC filter can be calculated to

$$|H(e^{i\omega T_S})| = \left|\frac{\sin\left(\pi RD\frac{f}{f_S}\right)}{\sin\left(\pi\frac{f}{f_S}\right)}\right|^M .$$

Since the decimation factor R is freely programmable, it is more convenient to express the magnitude response with respect to the output frequency

$$f_{out} = \frac{f_S}{R} = \frac{f_{in}}{R}$$

which yields

$$|H(f)| = \left|\frac{\sin\left(\pi D\frac{f}{f_{out}}\right)}{\sin\left(\frac{\pi}{R}\frac{f}{f_{out}}\right)}\right|^M . \tag{21}$$

The differential delay D, the decimation factor R, and the number of CIC stages M are the filter design parameters, chosen in such a way so as to fulfil the desired passband characteristics. The output spectrum has zeros, if the frequency is a multiple integer of f_{out}/D.

$$\text{filter zeros at} \qquad f = k\cdot\frac{f_{out}}{D} \qquad k\text{ : integer .}$$

Another property of CIC filters is their gain. When we compared CIC filters with box-car filters, it already became clear that they are almost identical apart from a scaling factor. The DC gain for CIC decimators can be calculated as

$$\lim_{f\to 0}|H(f)| = |H(0)| = (RD)^M . \tag{22}$$

Because of this fact, the magnitude response of a CIC filter is usually plotted relative do the DC magnitude. In other words, the magnitude is very often given as $|H(f)|/|H(0)|$. As a consequence of the DC gain, an internal register growth in the CIC path occurs. Each additional integrator must add the required number of bit width to account for the register growth which is equal to $\log_2(RD)$. As an example, if a differential delay $D = 1$ and a decimation of $R = 8$ is chosen, a register bit width of three additional bits is required at each integrator. In addition, overflows occur at each integrator but they are of

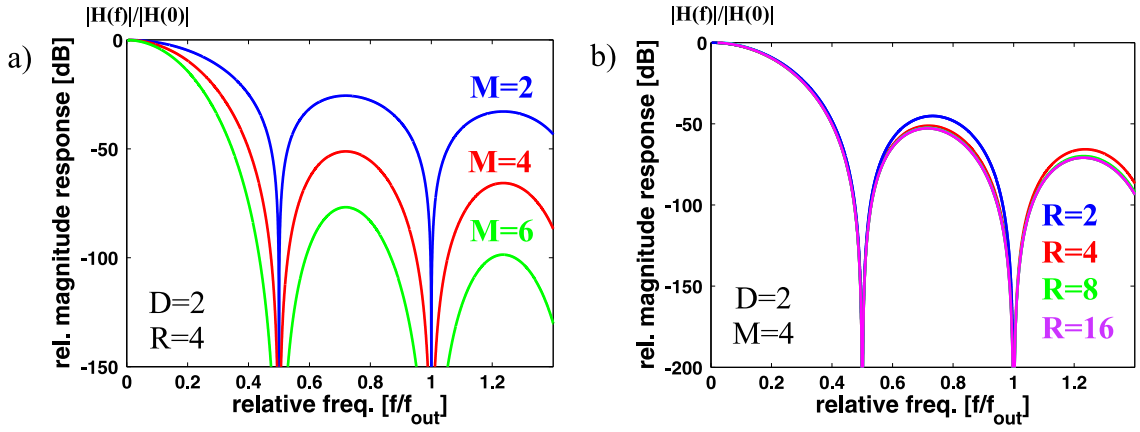


Fig. 19: Frequency response of CIC filter as a function of (a) the integrator and comb stages M; (b) the decimation ratio R

no consequence so long as the filter is implemented with two's complement (non-saturating) arithmetic. The frequency response of a CIC filter as a function of the integrator and comb stages M is shown in Fig. 19(a), and as a function of the decimation ratio R in Fig. 19(b). In both cases, the differential delay is held constant at $D = 2$. There are several properties to note. First of all, the differential delay D sets the number of zeros in the frequency band up to f_{out}. Secondly, the shape of the filter is not influenced very much by the decimation ratio R. For values of R above approximately 16, changes in the filter shape are negligible. Thirdly, the passband attenuation is a function of the integrator and comb stages M. As can be seen in Fig. 19(a), the first side lobe is attenuated by 26 dB for $M = 2$ and increases with the increasing number of CIC stages M; this acts to improve alias rejection. On the other hand, increasing M also increases the passband droop, which might not be acceptable in some applications. A solution to this problem could be to compensate the droop by using an additional non-CIC-based filter. For decimating CIC filters, the compensation filter is placed after the CIC, thus operating at a reduced clock speed (left plot in Fig. 20). In contrast to this, a pre-compensation filter is placed before the CIC

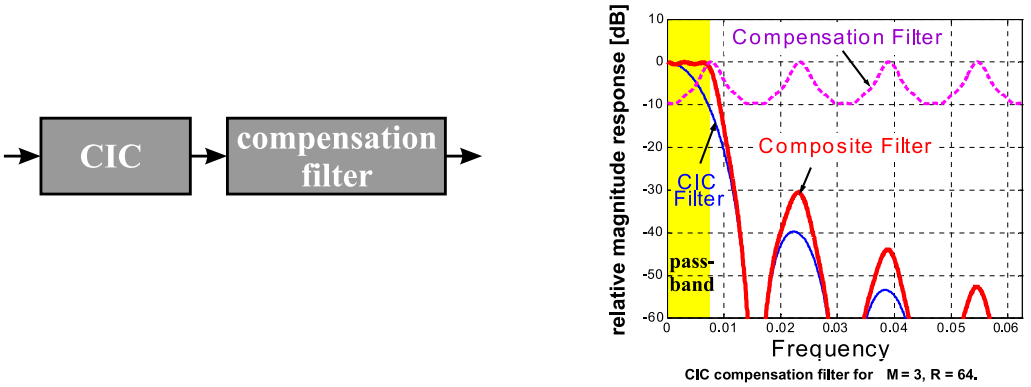


Fig. 20: Compensation filter for CIC to compensate the passband droop [7]

stage for interpolating CIC filters in order to work at the lower clock speed again. On account of the compensation filter, the composite filter can provide a sufficiently flat response within the passband. As an example, Fig. 20 shows a three-stage CIC filter (blue line) with a decimation ratio of $R = 64$ combined with a $x/\sin(x)$-shaped compensation filter (pink dotted line); the frequency response of the composite filter (thick red line) is sufficiently flattened within the desired passband (yellow shaded box).

3.3.3 Comparison of IQ sampling and DDC

After having introduced digital down conversion and IQ and non-IQ sampling, we shall now briefly compare the different approaches. In fact, both schemes have lots in common. The data stream of IQ sampling [see Eq. (6)], $Q, I, -Q, -I$, can also be regarded as the result of a mixing process with $f_{LO} = f_S = 4 \cdot f_{IF}$ which corresponds to the NCO frequency. The output of the mixing process contains the difference and the sum frequencies, $f_S - f_{IF}$ and $f_S + f_{IF}$. The two frequencies represent the 3rd and the 5th harmonic which are mapped on the IF frequency due to aliasing. In contrast to DDC, in which the NCO is usually set to the IF frequency and the sum frequency has to be filtered by the successive low-pass filter, with IQ sampling this is not required, owing to the aliasing of both components to the IF frequency. Some basic properties of DDC and IQ sampling are provided in Table 3. DDC has advantages

Table 3: Basic properties of DDC and IQ demodulation

DDC	**IQ demodulation**
– long group delay (depending on clock speed and number of taps in the CIC/FIR filters) – very flexible (NCO can follow f_{IF} over a broad range) – data reduction and good S/N ratio	– low latency, simple implementation – f_S is locked to $4 \cdot f_{IF}$ – sensitive to clock jitter and non-linearities – non-IQ sampling provides better S/N ratio on cost of latency

in applications with large, varying, IF frequencies and where a good signal-to-noise ratio is required, but at the cost of reasonable latency times. IQ demodulation, on the other hand, is mainly used in feedback applications where very short latency is important and where the sampling frequency can be locked to the IF frequency.

4 Up-conversion

In applications where the control of RF signals is required (like in feedback or in feed-forward systems), it is necessary to convert a digital baseband signal to a real passband signal. Ideally, a single DAC could generate the desired RF signal directly (Direct Digital Synthesis, DDS, see Fig. 14). However, today's DACs do not yet meet the precision and speed requirements for RF signals in the range of several hundreds of MHz. A series of DACs and mixers have to be used to provide the conversion into the RF domain. A brief overview about different up-conversion schemes will now be given. Basically, up-conversion techniques can be split into two approaches, homodyne (direct) and heterodyne.

4.1 Homodyne up-conversion

In homodyne up-conversion (also called direct or baseband up-conversion), the baseband I and Q signals are converted to analog signals and are then mixed with the in-phase and quadrature-phase components of a local oscillator (LO). The LO signal is split with a 90° hybrid in order to generate the required phase shift between the two paths. The translated baseband signals are then summed to generate the final RF signal (see Fig. 21 with the corresponding frequency spectrum). This device is generally known as *vector modulator*. The LO frequency is located at the desired RF carrier frequency and no intermediate frequency is required. The direct up-conversion is widely used because of its simplicity and cost effectiveness. The mixers in the vector modulator are operated as amplitude control elements since the I and Q inputs (baseband) directly control the amplitude of the two signal paths. We remember that in Eq. (2) and Eq. (3) we have seen that an arbitrary RF signal can be decomposed into its sine and cosine

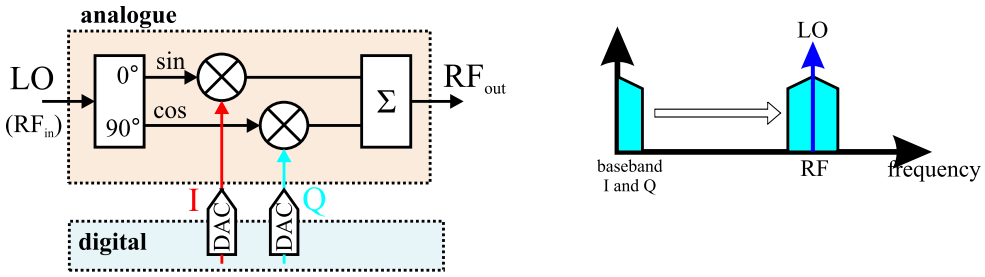


Fig. 21: Homodyne up-conversion approach (vector modulator)

components.

$$\begin{aligned} RF_{out}(t) &= I \cdot A_{LO} \cdot \sin \omega t + Q \cdot A_{LO} \cdot \cos \omega t \\ &= A_{out} \cdot \sin(\omega t + \varphi_0) \; . \end{aligned}$$

And vice versa, if the I and Q amplitudes are controlled, the amplitude and phase of the outgoing carrier signal can be changed. The corresponding output amplitude and phase are given by

$$A_{out} = A_{RF} \sqrt{I^2 + Q^2} \qquad \varphi_0 = \text{atan}\left(\frac{Q}{I}\right) . \tag{23}$$

With this approach, a pure amplitude or a pure phase modulation can be implemented very easily. For amplitude modulation, the I and Q inputs are modulated with a common time-varying amplitude function $A_0(t)$.

amplitude modulation:

$$\begin{aligned} I(t) &= A_0(t) \cdot \cos \varphi_0 \\ Q(t) &= A_0(t) \cdot \sin \varphi_0 \; . \end{aligned}$$

For pure phase modulation, the I and Q inputs follow a common time-varying phase function $\varphi_0(t)$ while the amplitude is kept constant.

phase modulation:

$$\begin{aligned} I(t) &= A_0 \cdot \cos \varphi_0(t) \\ Q(t) &= A_0 \cdot \sin \varphi_0(t) \; . \end{aligned}$$

Although direct up-conversion is often preferred due to its simplicity, the approach is prone to several sources of errors. The use of analog components with their part-to-part variation and non-linearities introduces many errors like DC offsets, gain imbalance, LO phase noise, and quadrature skew. Any DC offsets in either I or Q will lead to carrier leakage, while a gain imbalance between the two signal paths results in phase-to-amplitude modulation. Ideally, the split LO signal is exactly 90° out of phase. In reality, the quadrature skew—the deviation from the 90° phase shift—introduces coupling between I and Q. Note that digital signal processing can be employed to pre-distort the input of the vector modulator in order to minimize the quadrature skew:

$$\begin{pmatrix} I_{out} \\ Q_{out} \end{pmatrix} = \frac{1}{\cos \varphi_S} \cdot \begin{pmatrix} \cos \varphi_S & -\sin \varphi_S \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} I_{in} \\ Q_{in} \end{pmatrix} .$$

The skew phase is denoted by φ_S. The ideal inputs I_{in}, Q_{in} are pre-distorted to I_{out}, Q_{out} in order to counteract the skew caused by the analog vector modulator components.

4.2 Heterodyne up-conversion

Heterodyne up-conversion (also called IF up-conversion) applies the concept of digital up-conversion of I and Q signals to an IF frequency which is converted to an analog signal by a single DAC. The analog IF

signal is then mixed by means of at least one additional local oscillator (single-stage mixing) or several mixers (multistage mixing) to translate the IF signal to the desired RF frequency. The LO frequency has to be chosen such that the desired RF frequency corresponds to the sum of the LO and IF frequencies. In addition to the sum frequency, the mixing process also generates an image frequency at $f_{LO} - f_{IF}$. A schematic of single-stage heterodyne up-conversion along with the corresponding frequency spectrum is shown in Fig. 22. This approach is also called double-sideband modulation because of the generation

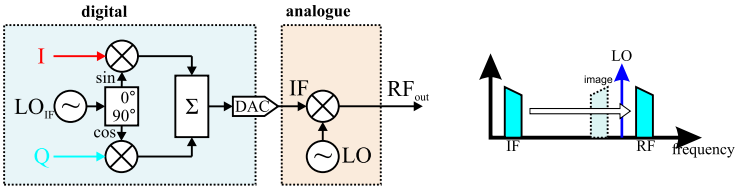


Fig. 22: Heterodyne up-conversion approach (IF up-conversion) as double-sideband modulator

of both frequency bands, i.e., the desired RF along with its image frequency band. Using this method requires filtering to ensure that the image frequency band is sufficiently suppressed and does not interfere with the RF frequency band. This concept is also known as the *filtering method*. In contrast to the homodyne up-conversion approach, the translation of I and Q to an IF frequency takes place digitally, thus it is not susceptible to gain imbalance and quadrature skews. However, the required DACs have to provide a higher bandwidth and are thus more subject to errors such as harmonic distortion and passband ripple.

An alternative for avoiding the image band is the single-sideband modulation scheme which is shown in Fig. 23. The digitally up-converted I and Q signals are converted to analog sine and cosine

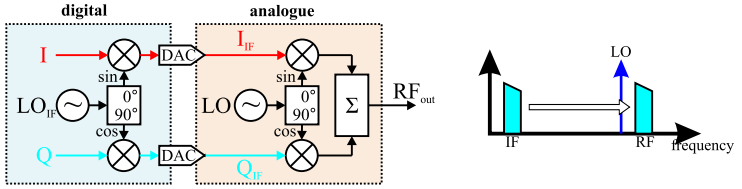


Fig. 23: Single-sideband modulator scheme

waveforms which are then applied to a vector modulator (direct quadrature modulator). Mixing the I_{IF} and Q_{IF} signals with the frequency f_{LO} generates frequencies at $f_{LO} \pm f_{IF}$. The translated I frequency $f_{LO} - f_{IF}$ is 180° out of phase with respect to the translated Q frequency $f_{LO} - f_{IF}$. These two frequencies will cancel each other out (at least in the ideal case), thus leaving a single sideband at the output of the vector modulator. A graphical representation of the image cancellation is shown in Fig. 24. In reality, gain imbalance and quadrature skew prevent complete cancellation; this results in

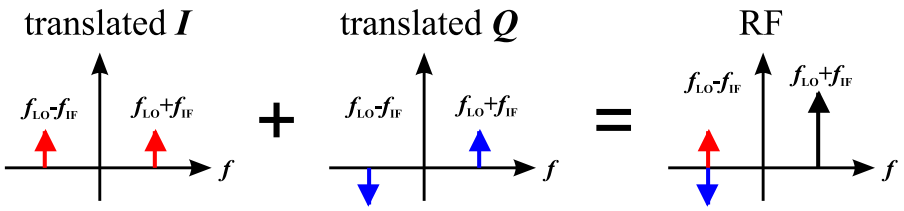


Fig. 24: Graphical representation of the image cancellation in the single-sideband modulator scheme

the unsuppressed sideband at $f_{LO} - f_{IF}$. The amplitude ratio between the desired single sideband at $f_{LO} + f_{IF}$ and the unwanted image sideband is called sideband suppression and is measured in [dBc].

5 Algorithms in RF applications

Following this discussion of signal conditioning/down conversion, digitization, IQ detection, as well as the up-conversion process (see Fig. 1), we shall now turn to common algorithms used in RF applications. In this respect the algorithms will be divided into feedback and adaptive feed-forward algorithms. Finally, a brief introduction to system identification will be given at the end.

5.1 Feedback algorithms

Among the many algorithms applied in the field of RF, we shall concentrate on only two particular cases for the purposes of this lecture: amplitude and phase feedback algorithms for RF cavity control, and phase and radial loops as applied in circular accelerators.

5.1.1 Amplitude and phase feedback for RF cavities

The task of these amplitude and phase feedback systems, generally known as Low-Level RF systems (LLRF), is to precisely maintain the phase and amplitude of the accelerating fields used to accelerate charged particle beams. There are many different types of RF cavities in use today (travelling wave structures, standing wave structures, spoke cavities, RFQ, ferrite loaded cavities, etc.), and their operating frequencies cover the range of some several MHz up to tens of GHz. Amplitude and phase feedback systems range from the control of normal and superconducting cavities (which are operated in pulsed or CW mode) up to the control of single and multi-cell cavities, with one or more cavities per RF amplifier (vector sum control). During the design process, one has to decide whether an analog, a digital, or a combined feedback system will be used. Likewise, the pros and cons of IQ versus amplitude/phase control have to be considered. The typical requirements for relative amplitude stability are in the range of $10^{-2} - 10^{-4}$ and for phase stability in the range of $10^{-2} - 10^{-4}$ rad, whereas the toughest requirements originate from FEL linac facilities. Apart from the pure amplitude and phase control, the systems very often have to fulfil additional tasks such as exception handling along with providing automated calibration routines and extensive diagnostics capabilities. If large bandwidths are required, analog systems—with their usually shorter loop latencies—have a big advantage, but more and more digital systems have been developed and are now in operation at accelerator facilities.

In the following section, we shall use standard control theory to analyse accelerator amplitude and phase feedback systems in more detail. A basic feedback loop of a system (e.g., an RF plant) with transfer matrix $G(s)$ and a controller with transfer matrix $C(s)$ is depicted in Fig. 25. The output of

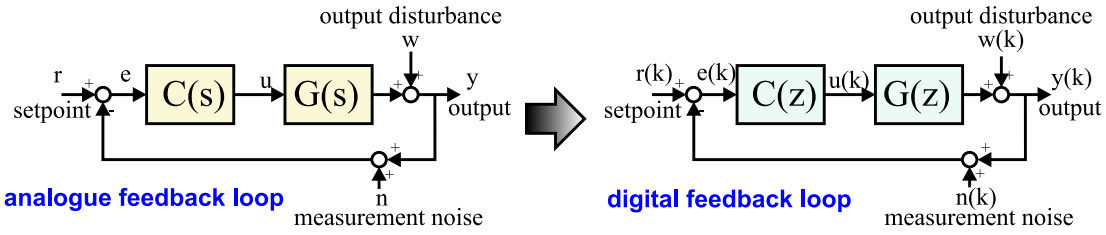


Fig. 25: Translation of a basic analog to a digital feedback loop

the system is described by y, the set-point by r, and the controller output by u. In addition, the output disturbance w and the measurement noise n are included. In order to describe a digital feedback loop, all continuous variables need to be transferred to the discrete domain. Mathematically speaking, the transfer matrices have to be transformed from the Laplace domain into the z-domain. While the controller can be directly designed in digital form, the transformation of the continuous plant into the z-domain requires a zero-order-hold block to be inserted. This since a digital controller only outputs discrete values (the latter are kept constant between two samples by a DAC, in order to deliver continuous inputs to the plant). The cascade of zero-order-hold block and continuous transfer matrix has to be transformed from the Laplace

into the time domain by the inverse Laplace operator $\mathcal{L}^{-1}$ and then to the z-domain by $\mathcal{Z}$. The resulting rule is

$$\begin{aligned} G(z) &= \mathcal{Z}\Big\{\mathcal{L}^{-1}\Big\{G(s)\cdot H_{ZOH}(s)\Big\}\Big|_{t=kT_S}\Big\} \qquad T_s:\ \text{sampling period},\ k:\ \text{integer} \\ G(z) &= \frac{z-1}{z}\mathcal{Z}\Big\{\mathcal{L}^{-1}\Big\{\frac{G(s)}{s}\Big\}\Big|_{t=kT_S}\Big\}\,. \end{aligned} \tag{24}$$

The continuous transfer matrix $G(s)$ includes all sub-systems such as modulators, pre-amplifiers, cavities, klystrons, etc. In order to design the controller for the feedback loop, it is necessary to model the plant sufficiently well. For the sake of simplicity, we reduce the feedback system to a Single Input–Single Output (SISO) system. In this case, the transfer matrices reduce to transfer functions. Based on the system shown in Fig. 25, the output y and the tracking error $e_T = r - y$ can be expressed by

$$\begin{aligned} y &= \frac{GC}{1+GC}[r-n] + \frac{1}{1+GC}w \\ e_T &= \frac{GC}{1+GC}n + \frac{1}{1+GC}[r-w]\,. \end{aligned}$$

The variable (z) has been omitted for clarity. The transfer function GC denotes the open loop transfer function. With the commonly defined functions

$$\begin{aligned} S &:= \frac{1}{1+GC} && \text{sensitivity function} \\ T &:= \frac{GC}{1+GC} && \text{complementary sensitivity function (closed loop transfer function)}\,, \end{aligned}$$

it becomes obvious that the sum of sensitivity and complementary sensitivity function is equal to 1.

$$S(z) + T(z) = 1\,. \tag{25}$$

There are several things to note here. Any measurement error n (e.g., IQ detection error) behaves like a change in the set-point r. In order to keep the influence of n on the output y and on the tracking error e_T small, $T(z)$ should be kept small; this improves robustness. On the other hand, the output y should be insensitive to frequency output disturbances w over the range of interest. Very often it is the low-frequency range where $S(z)$ should be minimized for performance reasons. Since $S(z)$ and $T(z)$ are complementary [see Eq. (25)], a trade-off between robustness and performance has to be made. Moreover, it is not possible to minimize $S(z)$ over the entire frequency range. This fact is expressed by the Bode integral formula [8, 9]. If we suppose that the open loop transfer function $G(s)C(s) = n(s)/d(s)$ is a rational function with two more poles than zeros (i.e., the degree($n(s)$) – degree($d(s)$) ≥ 2) and $G(s)C(s)$ is a linear time-invariant system which has no unstable poles (no poles in the right hand plane), then the Bode integral theorem states

$$\int_{-\infty}^{\infty} \ln|S(i\omega)|d\omega = 0\,. \tag{26}$$

Likewise, the Bode integral formula can also be written for linear discrete time-invariant systems. If $G(z)C(z) = n(z)/d(z)$ is a rational function with one more pole than the number of zeros, where no poles lie outside the unity circle (unstable poles), then the Bode integral theorem is expressed by

$$\int_{-\pi}^{\pi} \ln|S(e^{i\omega})|d\omega = 0\,. \tag{27}$$

This means the integral of the sensitivity is conserved. Low sensitivity (i.e., good disturbance rejection of the feedback system) over a certain frequency range must be compensated for by a sensitivity with

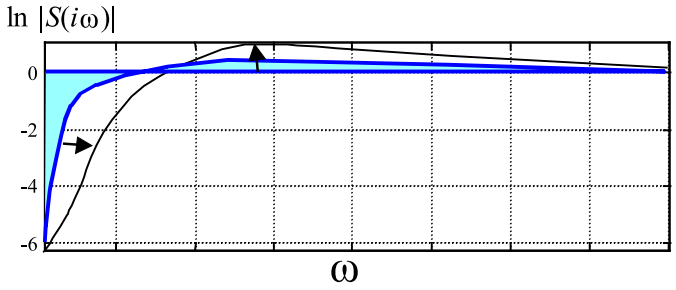


Fig. 26: Principle of Bode's integral theorem (waterbed effect); lowering the sensitivity in one frequency range increases it somewhere else

$|S| > 1$ across another range. This effect is sketched in Fig. 26. If the gain of a feedback loop is increased in order to increase the suppression bandwidth in the low-frequency range, the potential to amplify disturbances in the high-frequency range is also increased. This effect is frequently referred to as the *waterbed effect*. Care has to be taken to make sure that no dominant disturbance lines are located in the frequency range where the absolute value of the sensitivity function is greater than one.

In order to apply control theory to cavity amplitude and phase feedback systems, a model of the RF plant is required. Once all devices have been modelled (i.e., transfer matrices have been obtained), the total RF plant transfer matrix is the result of the transfer matrix products of the individual components. First of all, we shall deduce the transfer matrix of an RF cavity. Resonant modes in cavities can be described by means of resonant LCR circuits [10]. In the following we restrict ourselves to the simple case of a single-passband mode of a cavity. A simplified model of a LCR circuit fed by a generator current $\tilde{I}_g$ and a beam current I_b is shown in Fig. 27. The true generator current I_g which is supplied by

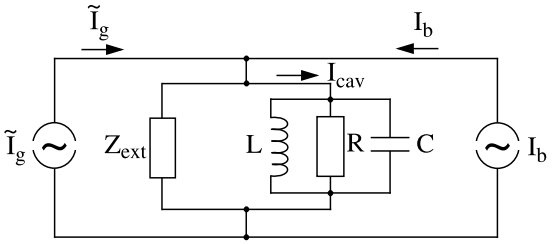


Fig. 27: Simplified model of resonant RF cavity with generator current $\tilde{I}_g$ and beam current I_b

an RF source and transported to the cavity via a transmission line is transformed into the cavity system by means of the cavity input coupler. This transformation is taken into account in Fig. 27 by representing the generator current as $\tilde{I}_g$ [11]. Based on the model depicted in Fig. 27, the differential equation of a driven LCR circuit can be obtained. The current $\mathbf{I}(t)$ is the sum of the generator and beam currents. For RF

$$\ddot{\mathbf{V}}(t) + \frac{\omega_0}{Q_L}\dot{\mathbf{V}}(t) + \omega_0^2\mathbf{V}(t) = \frac{\omega_0 R_L}{Q_L}\dot{\mathbf{I}}(t)$$

$\mathbf{V}(t)$:	cavity voltage
$\mathbf{I}(t)$:	driving current (from generator and beam)
ω_0:	resonance frequency of undamped cavity
Q_L:	loaded quality factor of cavity
R_L:	cavity resistance incl. external load

applications, the driving current and the cavity voltage are harmonic oscillations with time dependence $e^{i\omega t}$, where ω is the frequency of the driving current. Therefore, we separate fast RF oscillations from the slowly changing amplitude and phase variations.

$$\mathbf{V}(t) = \begin{pmatrix} V_r(t) \\ V_i(t) \end{pmatrix} \cdot e^{i\omega t}\,, \qquad \mathbf{I}(t) = \begin{pmatrix} I_r(t) \\ I_i(t) \end{pmatrix} \cdot e^{i\omega t}\,. \tag{28}$$

In this formula, the notation of real and imaginary parts of the current and voltage have been chosen instead of the I and Q representation. Inserting Eq. (28) into the cavity differential equation results in the description of the cavity field envelope as a matrix equation.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} V_r \\ V_i \end{pmatrix} = \begin{pmatrix} -\omega_{1/2} & -\Delta\omega \\ \Delta\omega & -\omega_{1/2} \end{pmatrix} \cdot \begin{pmatrix} V_r \\ V_i \end{pmatrix} + \begin{pmatrix} R_L\omega_{1/2} & 0 \\ 0 & R_L\omega_{1/2} \end{pmatrix} \cdot \begin{pmatrix} I_r \\ I_i \end{pmatrix} . \tag{29}$$

The parameter $\omega_{1/2}$ is the cavity bandwidth of the loaded cavity, and $\Delta\omega$ is the difference frequency between the resonance frequency and the frequency of the driving source. Both parameters are defined by

$$\omega_{1/2} = \frac{\omega_0}{2Q_L} , \qquad \Delta\omega = \omega_0 - \omega . \tag{30}$$

In order to transform the differential Equation (29) into an algebraic equation, Laplace transformation can be applied which yields

$$\underbrace{\begin{pmatrix} V_r(s) \\ V_i(s) \end{pmatrix}}_{V(s)} = \underbrace{\frac{\omega_{1/2}}{\Delta\omega^2 + (s+\omega_{1/2})^2} \begin{pmatrix} s+\omega_{1/2} & -\Delta\omega \\ \Delta\omega & s+\omega_{1/2} \end{pmatrix}}_{H_{cav}(s) = \begin{pmatrix} H_{11}(s) & H_{12}(s) \\ H_{21}(s) & H_{22}(s) \end{pmatrix}} \cdot \underbrace{\begin{pmatrix} R_L \cdot I_r(s) \\ R_L \cdot I_i(s) \end{pmatrix}}_{U(s)} . \tag{31}$$

Having done so, we have obtained the continuous 2×2 transfer matrix of a resonant cavity. A few properties of this matrix will now be highlighted. If the cavity is operated on resonance ($\Delta\omega = 0$), then the off-diagonal elements of the transfer matrix vanish. The cavity acts like a low-pass filter of first order with a roll-off frequency of $\omega_{1/2}$.

$$\begin{aligned} \Delta\omega = 0 \text{ (cavity on resonance) :} \qquad H_{11}(s) = H_{21}(s) &= \frac{\omega_{1/2}}{s+\omega_{1/2}} \\ H_{12}(s) = H_{21}(s) &= 0 . \end{aligned} \tag{32}$$

In this case, the real and imaginary parts of the cavity voltage (or the I and Q components), and thus amplitude and phase, are completely decoupled. If the cavity is detuned ($\Delta\omega \neq 0$), coupling between I and Q occurs. Especially for superconducting cavities, which are subject to Lorentz force detuning at high gradients, the cavity resonance frequency and therefore $\Delta\omega$ becomes a function of the accelerating field in the resonator. If the cavity is operated in a pulsed mode, then the cavity filling causes a transition of the resonance frequency. This leads to $\Delta\omega = \Delta\omega(t)$. Thus the transfer function $H_{cav}(s)$ becomes time variant, therefore complicating the control task. The stability of the feedback loop has to be guaranteed under these parameter changes, which may require modern control theory approaches to synthesize an *optimal controller*.

As a next step, the transformation from the continuous into the discrete cavity model has to be carried out according to Eq. (24). Although an analytical expression of $H_{cav}(z)$ could be given, the calculation is tedious. Most of the time it is sufficient to use computer programs (*octave, Scilab, matlab, etc.*) and to analyse the feedback loop with its discrete transfer matrices. An alternative to using the transfer matrix approach is the analysis of amplitude and phase feedback systems in the state space formalism [12] where the differential equation (29) can be written as

$$\dot{\mathbf{x}}(t) = \mathbf{A}\cdot\mathbf{x}(t) + \mathbf{B}\cdot\mathbf{u}(t) \tag{33}$$

$$\mathbf{y}(t) = \mathbf{C}\cdot\mathbf{x}(t) \tag{34}$$

with the definition

$$\mathbf{x}(t) = \begin{pmatrix} V_r(t) \\ V_i(t) \end{pmatrix} , \qquad \mathbf{u}(t) = \begin{pmatrix} I_r(t) \\ I_i(t) \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} -\omega_{1/2} & -\Delta\omega \\ \Delta\omega & -\omega_{1/2} \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} R_L\omega_{1/2} & 0 \\ 0 & R_L\omega_{1/2} \end{pmatrix}, \qquad \mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

In this lecture we shall not explore the approach of state space formalism in more detail. The transfer matrix analysis of the closed amplitude and phase feedback loop in the digital domain requires the loop delay (which is inherent due to digital signal processing) to be taken into account. Also all cable delays must be considered. This delay of N sample periods is expressed by the z^{-N} transfer matrix in Fig. 28. In most cases, the digital controller for amplitude and phase feedback for RF cavities is typically imple-

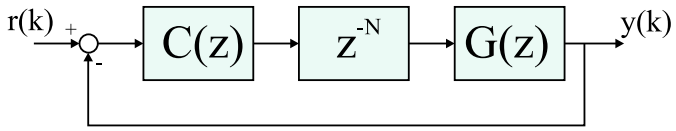


Fig. 28: Block diagram of a digital feedback loop with loop delay of N sample periods

mented using a PID controller. Nevertheless, studies are currently underway in many laboratories that aim at improving the feedback performance by synthesizing an optimal controller. Although PID controllers have been analysed thoroughly in many publications, we shall nevertheless investigate the loop performance with this type of controller in more detail, since PID controllers play such an important role in RF cavity field control. A standard PID control algorithm is expressed by

$$u(k) = u(k-1) + (K_p + K_i + K_d) \cdot e(k) - (K_p + 2K_d) \cdot e(k-1) + K_d \cdot e(k-2) \qquad (35)$$

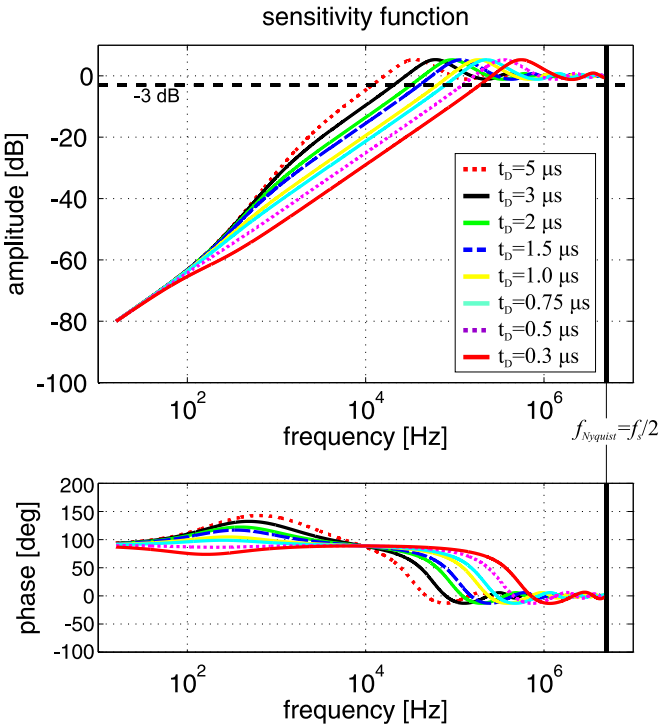
with the error value $e(k)$ at sample time t_k, and the proportional, integral, and derivative gains K_p, K_i, K_d respectively. To simplify matters, we regard the RF plant as consisting only of the cavity, neglecting all other components such as vector modulators, pre-amplifiers, klystrons, pickups, etc. Even with this simplified example, some important properties of the feedback design will be revealed. The plant $G(z)$ in Fig. 28 is described by the transfer function given in Eq. (31). As a standard rule of thumb, control theory recommends a gain margin of at least between 6–8 dB and a phase margin of between 40° and 60°. With a constant integral gain ($K_i = 0.1$), the proportional gain is adjusted in the above-mentioned example such that a gain margin of 8 dB is achieved. Depending on the total loop delay t_D, different maximum proportional gains can be set which yield different loop bandwidths. The sensitivity function as a result of this simulation is shown in Fig. 29. The achievable maximum bandwidth strongly depends on the total loop delay and is very sensitive if the total delay is on the order of the sampling period. Thus the loop delay is an important parameter which has to be minimized. High-speed digital technology is required to keep the delay in the digital signal processing part as short as possible. However, despite ever more powerful digital technology, the larger delays typical of digital feedback systems (compared to their analog counterparts) are still the limiting factor in their application. This becomes most evident if superconducting and normal conducting cavity field feedback systems are compared (see Table 4). In

Table 4: Comparison of basic feedback parameters of normal and superconducting cavities

Superconducting cavity	Normal conducting cavity
Q_L: ≈ few 10^5–10^7	Q_L: ≈10–10^5
$f_{1/2}$: ≈ few 100 Hz	$f_{1/2}$: ≈100 kHz
τ_{cav}: ≈ few 100 μs	τ_{cav}: ≈ few μs
feedback loop delay **small** compared to τ_{cav}	feedback loop delay **in the order** of τ_{cav}

this table, the cavity bandwidth $f_{1/2}$ is given by Eq. (30) and the cavity time constant (standing wave cavity) is the time constant of a resonator which is defined by

$$\tau_{cav} = \frac{1}{\omega_{1/2}} = \frac{Q_L}{\pi f_{RF}}. \qquad (36)$$



parameters:
superconducting cavity with
$f_0 = 1.3$ GHz,
$Q_L = 3 \cdot 10^6$,
$\omega_{1/2} = 216$ Hz,
$f_s = 10$ MHz;
controller: PI;
integral gain: constant ($K_i = 0.1$);
proportional gain: adjusted to keep gain margin constant at 8 dB.

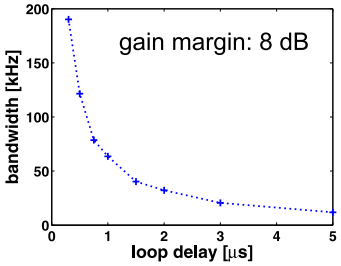
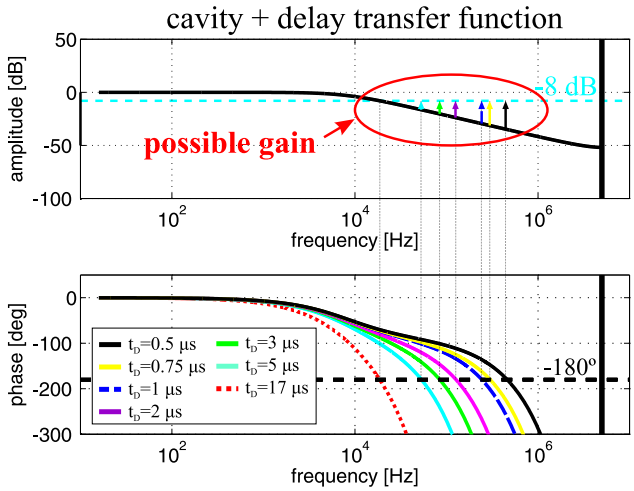


Fig. 29: Bode plot of the sensitivity function of a cavity-field feedback loop for different loop delays

Since the loaded quality factors Q_L of normal conducting cavities are some order of magnitudes smaller than those of superconducting cavities, the bandwidths of the normal conducting cavities are usually in the range of some hundred kHz. As a result, the cavity time constants are comparable with the loop delays which are typical of digital systems nowadays. The consequence is that loop gains are limited because unity gain is already reached with moderate gains at frequencies where the loop phase (dominated by the phase advance of the loop delay) approaches 180°. The achievable gains are highlighted in the open loop Bode plot (simplified model which consists of a series of cavity and loop delay transfer function) in Fig. 30. If the bandwidth provided by a digital feedback system is insufficient, analog or analog/digital



normal conducting cavity
Q_L=2·10^4
f_{RF}=324 MHz

Fig. 30: Open loop Bode plot of a normal conducting RF cavity for different loop delays

hybrid systems might offer an alternative. As a final example, a digital cavity amplitude and phase feedback will now be shown. A schematic layout of the J-PARC LLRF system for the normal conducting

proton linac (Drift Tube Linac, DTL, and Separate-type Drift Tube Linac, SDTL) is depicted in Fig. 31 [13, 14]. The linac is operated in pulsed mode with a repetition rate of 12.5 Hz or 25 Hz, respectively,

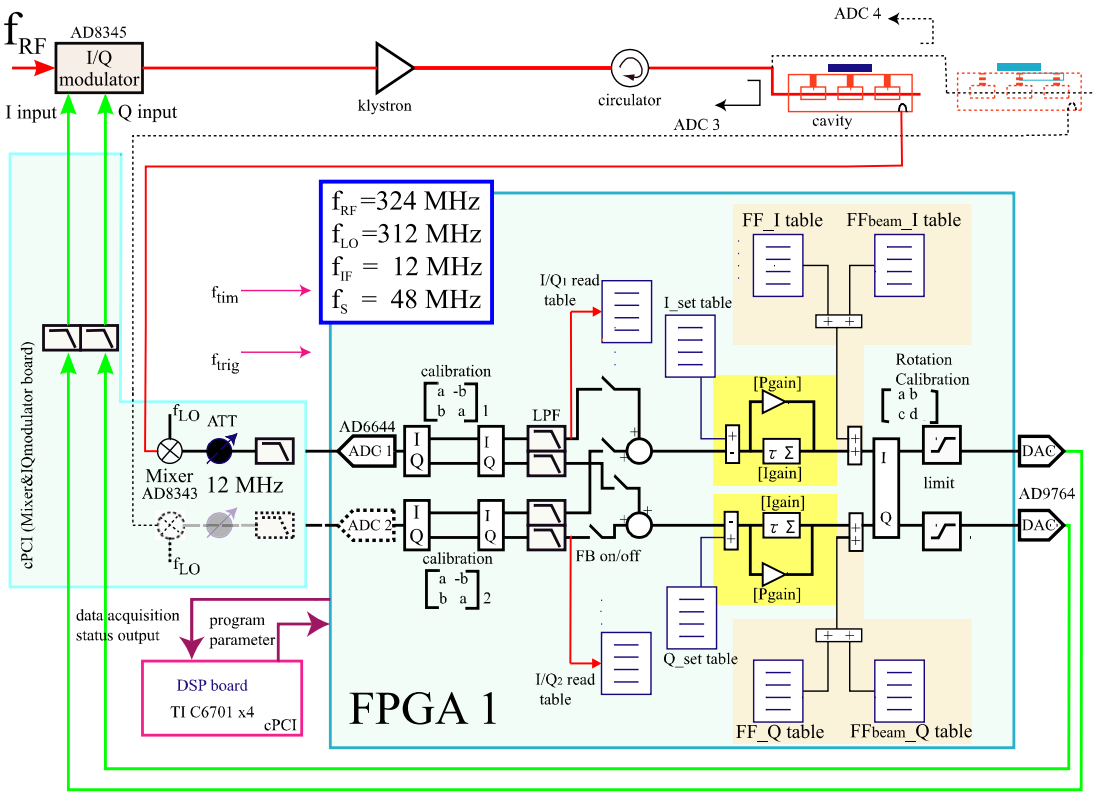


Fig. 31: Example of digital cavity amplitude and phase feedback system (J-PARC linac)

and a pulse length between 500 μs to 650 μs. One klystron feeds two drift tube structures. Sampling the down-converted IF frequencies of 12 MHz with 48 MHz sampling frequency ADCs deliver a stream of I and Q data. The successive rotation matrices perform the 90° phase rotation, including an absolute phase shift, which accounts for the individual cable delays from the cavity pickups to the ADCs. The digital LLRF system calculates and controls the vector sum using a PI control algorithm. In addition, feed-forward is applied by means of lookup tables which are adapted from pulse to pulse. The cavity amplitude and phase control in the form of I/Q control is implemented in a single FPGA, while Digital Signal Processors (DSPs) are used for tuner control, communication and configuration of the FPGA. The normal conducting cavities along the linac have loaded quality factors Q_L between 8 000 and $3 \cdot 10^5$ and typical cavity time constants of around 100 μs due to the low operating frequency of 324 MHz. It has been demonstrated that the required amplitude and phase stabilities of $<\pm 1\%$ and $<\pm 1°$, respectively, have been superseded by a factor of six. The total loop delay amounts to about 500 ns and the loop bandwidth reaches 100 kHz with moderate proportional gains of 10 and integral gains of 0.01.

5.1.2 Radial and phase loops

Booster synchrotrons for hadron or ion acceleration are often designed to accelerate a variety of particles with different charge-to-mass ratios. These machines must be able to capture and very often adiabatically rebunch the injected beam, and then accelerate it to the desired extraction energy. The different species of particles require very different acceleration cycles. The relation between the necessary magnetic field change dB/B in a synchrotron, the change of the revolution frequency df/f, the change of the mean

radial position dR/R and the relative particle energy γ is given by [15]

$$\frac{dB}{B} = \gamma^2 \frac{df}{f} + (\gamma^2 - \gamma_{tr}^2) \frac{dR}{R} \tag{37}$$

where γ_{tr} is the transition energy. The LLRF drive system must be very flexible and fast to support the dynamic changes throughout the ramp cycle. The beam control system's task is to calculate the revolution frequency as a function of dipole field, control the cavity voltages, keep the mean radial position of the beam on the desired value, and lastly, to maintain the synchronous phase Φ_s between beam and cavity field. In addition, it must synchronize the RF phase to a master RF clock in order to provide synchronization of the beam transport to other accelerator rings. The RF frequency can change up to a factor of ten during the ramp. The values typically vary from several hundreds of kHz to several tens of MHz. In order to support these large frequency changes, low Q cavities are used which are often ferrite-loaded cavities or magnetic alloy based cavities. In contrast to electron machines, no damping of synchrotron oscillations occurs in hadron machines. As a result, any particles deviating from the synchronous phase Φ_s will lead to oscillations. Moreover, all errors due to phase noise, imperfections in the magnetic field B, power supply ripple, etc., will cause phase and radial errors of the accelerated beam with respect to their design values. Hence radial and phase loop feedback systems are required. To summarize, the beam control system consists of the following main components.

- *Frequency program*
  It calculates the frequency based on the B field and based on desired radial position, it optimizes the frequency ramp to improve injection efficiency, and it generates dual harmonic RF drive signals for cavities in order to provide bunch shaping capabilities.
- *Beam phase loop*
  It forces the cavity RF phase to follow the beam phase, i.e., it maintains the synchronous phase. Thus it damps coherent synchrotron oscillations from injection errors (energy, phase), bending magnet noise, and low-frequency synthesizer phase noise.
- *Radial loop*
  It keeps the beam to its design radial position during acceleration and therefore forces the beam to have the synchronous energy.
- *Cavity amplitude loop*
  It compensates for imperfections in the cavity amplifier chain and provides the capability to control the amplitude according to a ramping function.
- *Synchronization loop*
  It locks the RF phase of the synchrotron to a master RF clock frequency in order to synchronize different accelerator rings which in turn ensures proper beam transfer.

The required frequency signals were commonly generated by analog methods, mainly by mixers and voltage controlled oscillators (VCOs), which were used to accommodate the large frequency changes during the acceleration cycle. The disadvantage of VCOs is their lack of absolute accuracy. Their stability limitations become evident if frequency tuning is necessary over a broad range. With the recent large advances in digital technology, VCOs were replaced more and more by Direct Digital Synthesis (DDS) in the late 1980s. The agility, reproducibility, and precision achieved with DDS supersedes the application of analog technology in most cases. Since DDS devices are completely digital except for the output DAC, their stability is completely defined by an external, highly stable fixed frequency oscillator. However, the spectral purity depends on the clock frequency and the operating frequency. After the digital implementation of the frequency program, the phase and radial loops were also more and more often implemented as digital feedback systems. In recent years, fully digital beam control systems have been implemented at machines like LEIR, AGS, RHIC, etc. A basic system layout is shown in Fig. 32. All RF signals such as those from cavity pickups, phase pickups (e.g., Wall Current Monitor, WCM), and

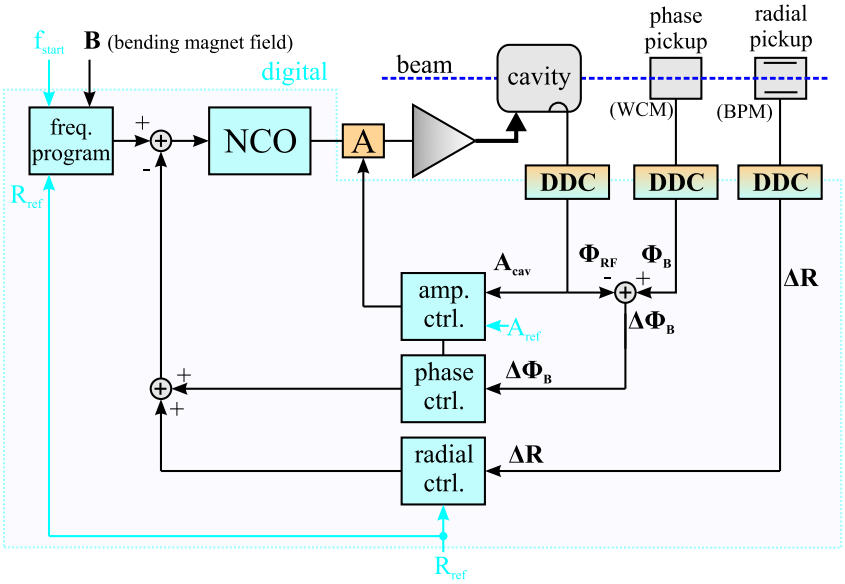


Fig. 32: Basic structure of digital radial and phase loops for hadron/ion synchrotrons

from radial pickups (Beam-Position Monitor, BPM), are digitized by means of I/Q sampling or Digital Down Conversion (DDC). The latter method is usually used since it provides the better possibility to track the changing frequencies as has been shown in Section 3.3. While the amplitude loop acts directly on the amplitude of the drive signal, the radial and phase loop apply a correction to the calculated frequency from the frequency program. The feedback gains of the purely digital control loops can be a function of the beam parameters in order to maintain the same loop performance throughout the acceleration cycle. To demonstrate this statement in more detail, the transfer functions of the radial and phase feedback loops will now be examined. The open loop model of both feedback systems is depicted in Fig. 33 [16]. The output of the DDS is amplified in a klystron and can be modelled in the simple case as a pure

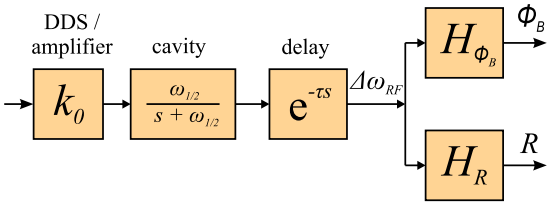


Fig. 33: Open loop model of radial and phase loops in hadron/ion synchrotrons

amplification with factor k_0. The transfer function of the broadband cavity is given in Eq. (32) (cavity on resonance) and is followed by the loop delay τ. It can be shown by beam dynamics calculation that the transfer function from an RF frequency change $\Delta\omega_{RF}$ to beam phase difference $\Delta\phi_B = \phi_B(s) - \phi_{RF}(s)$ is given by [16, 17]

$$H_{\phi_B}(s) = \frac{\Delta\phi_B}{\Delta\omega_{RF}} = \frac{s}{s^2 + \omega_S^2}\,, \tag{38}$$

which is based on the model of a lossless oscillation around the synchronous phase Φ_S (linearized synchrotron oscillations). The frequency ω_S is the synchrotron frequency and can be calculated to

$$\omega_S = f_{rev} \cdot \sqrt{\frac{2\pi e V_{RF} \cos\Phi_S h|\eta|}{E}}\,,$$

where f_{rev} is the asymptotic revolution frequency, V_{RF} the amplitude of the cavity voltage, h the harmonic number, E the particle energy, and $|\eta| = |1/\gamma^2 - 1/\gamma_{tr}^2|$ with γ_{tr} the transition energy. Likewise,

the transfer function from an RF frequency change $\Delta\omega_{RF}$ to a radial position change (or to the radial position itself) can be modelled to

$$H_R(s) = \frac{dR}{\Delta\omega_{RF}} = \frac{b}{s^2 + \omega_S^2} \tag{39}$$

with b as a function of energy and synchronous phase.

$$b = \frac{ceV_{RF}\cos\Phi_S}{2\pi\beta\gamma_{tr}E} \,.$$

We shall not go further into the details of beam dynamic calculations here, but just emphasize that both transfer functions, $H_{\phi_B}(s)$ and $H_R(s)$, are functions of energy which increases during the acceleration cycle. Hence the underlying model is time varying. These models are usually difficult to analyse and are treated in the frame of Linear Parameter Varying (LPV) models. Controllers for LPV models are often scheduled, i.e., the controller parameters are adjusted over time to meet the changing plant dynamics and thus guarantee constant loop performance and stability. As an example, the implementation of a radial loop is shown in Fig. 34. It is a test system implementation for LEIR (Low Energy Ion Ring) which has

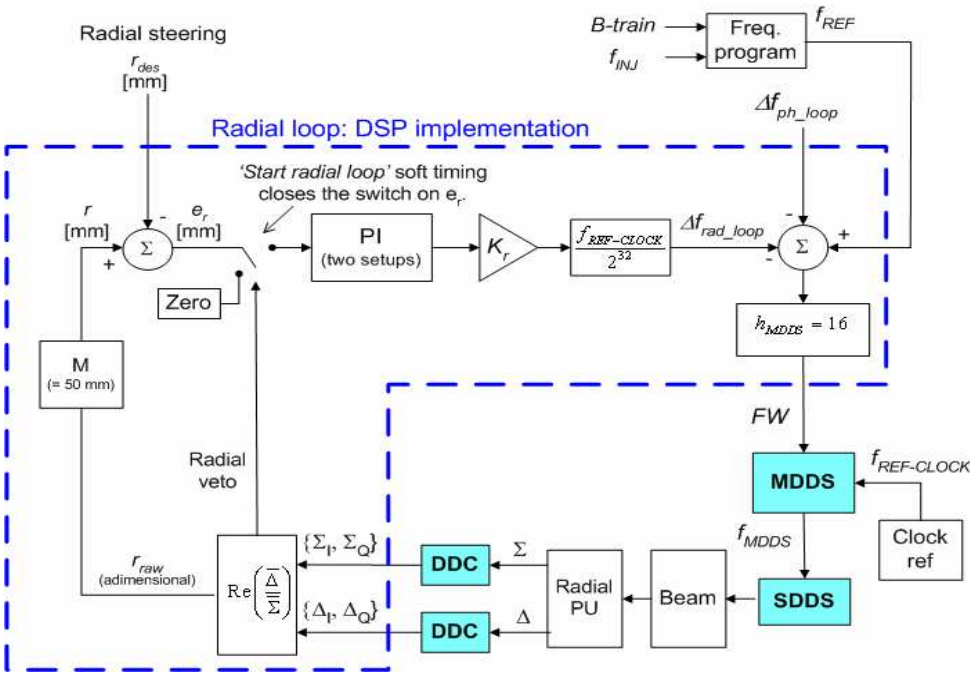


Fig. 34: Example of a radial loop implementation: LEIR system test at the PS Booster at CERN

been successfully tested at the PS Booster at CERN. Parts of the loop run on digital signal processors (feedback loop) while processing-intensive tasks like DDC have been implemented on FPGAs. Apart from the PI controller—which contains two different sets of control parameters (gain scheduling) to match the different parameters at injection and extraction—many of the concepts which have been presented in this course have been applied. The frequency is calculated by the *frequency program* based on the 'B-train' input, derived from a field measurement in a reference magnet which is connected in series with the bending magnets in the ring. Corrections from the radial and phase loop are added to the frequency output and applied to a Master DDS (MDDS). The distributed frequency is received by so-called Slave DDS cards (SDDSs) which locally generate the RF drive frequency for their respective RF cavities. The signals from the beam-position monitor pickups (sum and difference signal) are down-converted by DDCs in which a 1st order CIC with decimation factor $R = 64$ has been implemented. Based on the digitized sum and difference values (I,Q), the position is calculated with a geometric scaling factor M. Finally, a NCO stage follows the PI controller.

5.2 Adaptive feed-forward

The performance of RF feedback systems with repetitive or well-known errors can be improved by the application of adaptive feed-forward algorithms; the goal of these algorithms is to unburden the feedback and/or to suppress known disturbances in advance. The use of feed-forward schemes to compensate beam loading in pulsed accelerating structures is a particularly common practice. The beam arrival time and duration is known in advance, but the induced voltage of the beam can only be compensated by feedback after a certain time, owing to the loop delay. Feed-forward systems have proven effective for suppressing the transients which normally occur if only feedback is used. Since the repetitive disturbances can slowly vary over time, the feed-forward signal has to be adapted by means of adaptive algorithms. The standard feedback loop shown in Fig. 25 is extended by the adaptive feed-forward scheme (see Fig. 35). The

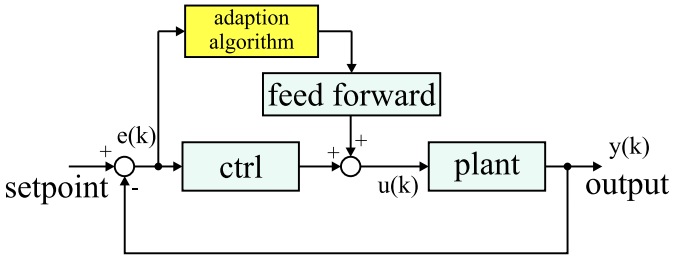


Fig. 35: Basic structure of an adaptive feed-forward in addition to a standard feedback loop

question remains as to how the feed-forward correction should be obtained. Recording the error signal $e(k)$, shifting it in time to compensate for the loop delay, then applying the opposite correction $-e(k)$ will not work, since this approach does not take into account the plant dynamics. What is necessary is to generate the proper additional input to the plant which then results in the required output $-e(k)$. In other words, an inverse model of the plant is necessary to extract the input which leads to the output $-e(k)$. However, this is not always easy in reality. A model to describe the dynamic behaviour of the system is either not known or not precise enough. In addition, stochastic errors are superimposed on the measured total error signal. Averaging methods are necessary in order to separate repetitive from stochastic errors. There are various approaches to attack this problem. One possibility is to obtain a system model by system identification (see Section 5.3), another way is to measure the system response with test input signals (step inputs), record the output, and calculate the system response matrix [18]. Assuming a linear time-invariant Single Input–Single Output (SISO) system, an input $u(k) = u(t_k)$ will cause an output $y(k) = y(\tau_k)$. Note that the definition of the sampling time is made such that the output occurs not earlier than

$$\tau_k = t_k + t_D \tag{40}$$

due to the loop delay t_D. If we change the input by small increments $\Delta u(t_k)$ around the chosen working point with input $u(t_k)$, the output changes by $\Delta y(\tau_k)$ for linear systems (or for sufficient small changes in a first order approximation). The first change can be written as

$$\Delta y(\tau_1) = R_{11} \cdot \Delta u(t_1) \; ,$$

where R_{11} describes the linear system response to the input $\Delta u(t_1)$. The second sampled output change is a linear combination of the previous two test input signals, expressed by

$$\Delta y(\tau_2) = R_{21} \cdot \Delta u(t_1) + R_{22} \cdot \Delta u(t_2) \; .$$

Applying successive test input signals $\Delta u(t_k)$ to the system results in a sequence of output changes which can be written in a matrix equation.

$$\begin{pmatrix} \Delta y(\tau_1) \\ \vdots \\ \Delta y(\tau_p) \end{pmatrix} = \begin{pmatrix} R_{11} & & 0 \\ \vdots & \ddots & \\ R_{p1} & \dots & R_{pp} \end{pmatrix} \cdot \begin{pmatrix} \Delta u(t_1) \\ \vdots \\ \Delta u(t_p) \end{pmatrix} \quad \Longleftrightarrow \quad \Delta \vec{\mathbf{y}} = \mathbf{R} \cdot \Delta \vec{\mathbf{u}} \; . \tag{41}$$

For reasons of causality, the upper half of the matrix is zero, since an input at time t_k cannot be detected earlier than $\tau_k = t_k + t_D$. Owing to the time definition in Eq. (40), the matrix $\mathbf{R}$ does not contain zero rows or columns. Hence, the response matrix $\mathbf{R}$ can be inverted in the general case and Eq. (41) solved for the required input which produces a measured output change.

$$\Delta\vec{\mathbf{u}} = \mathbf{T} \cdot \Delta\vec{\mathbf{y}} \; .$$

Once the inverted response matrix $\mathbf{T} = \mathbf{R}^{-1}$ has been measured, it is possible to calculate the necessary additional input $\Delta\vec{\mathbf{ff}}$ (feed-forward) to cancel the recorded repetitive error vector $\vec{\mathbf{e}}$.

$$\Delta\vec{\mathbf{ff}} = \mathbf{T} \cdot (-\vec{\mathbf{e}}) = \mathbf{T} \cdot (\vec{\mathbf{y}} - \vec{\mathbf{r}}) \; . \tag{42}$$

The vector $\vec{\mathbf{r}}$ denotes the set-point. As mentioned earlier, the repetitive error has to be separated from stochastic errors before being applied to the feed-forward adaption algorithm. A successful example of adaptive feed-forward in a pulsed cavity field feedback system (for compensation of repetitive errors due to beam loading and Lorentz force detuning) has been demonstrated at the TESLA Test Facility (now called FLASH) [19, 20]. The basic principle of the algorithm is shown in Fig. 36. The indicated feed-

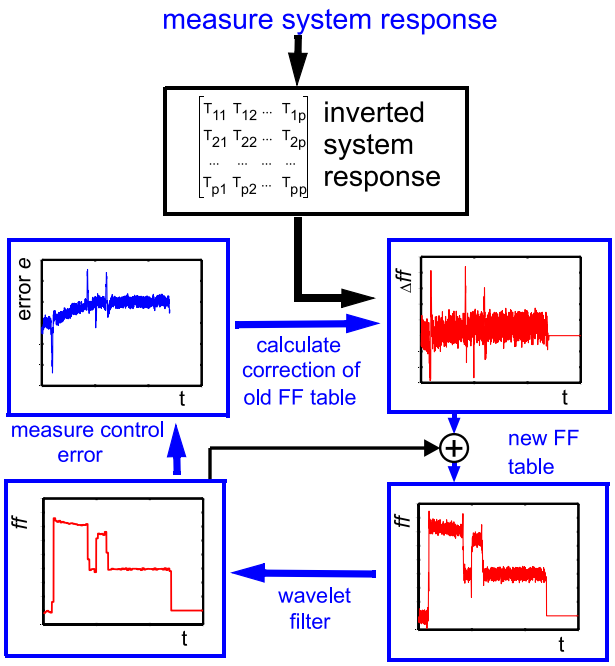


Fig. 36: Structure of an adaptive feed-forward algorithm at the TESLA Test Facility to compensate beam loading and Lorentz force detuning

forward table output forms one of the DAC output values to the vector modulator. Before the calculated feed-forward table is applied, it is wavelet filtered in order to remove stochastic noise introduced by the measurement process. Although this algorithm works fine, it turned out that any change of the working point (gradient and/or phase in the cavities) required the response matrix to be re-measured, since the whole system is highly non-linear. The response matrix measurement and the following adaption process tended to be too slow for the dynamic changes during machine operation. The need for a fast and robust adaptive feed-forward algorithm has led to the development of the so-called *time-reversed filtering* to generate the feed-forward tables; this is described in more detail in Ref. [21]. This approach of generating the feed-forward tables has been successfully demonstrated at FLASH and at the Spallation Neutron Source (SNS) [22] but is only applicable for pulsed systems for which the entire set of measurement data recorded during the pulse is available for the adaption process.

5.3 System identification

The field of system modelling and identification has advanced at a rapid pace during the past decades. System identification refers to the general process of extracting information about a system from measured input–output data. However, it is not possible to treat this subject to anything near its full extent in this lecture. We shall concentrate only on basic concepts and show an example of an application to a real RF plant. More detail about system identification is given in the lectures published in Ref. [12]. The interest in applying system identification to the area of RF control is to aid in the design or synthesis of high-performance controllers. The synthesis of a controller is based upon a model of the system. If a model exists, one can use it to develop a model-based controller. In case a model is not available but a sufficient amount of input–output data is accessible, the data can be used to develop a model of the system first, and then use that identification model to design the controller. Also, the implementation of efficient adaptive feed-forward algorithms requires a model of the system in order to cancel repetitive errors as shown in Section 5.2. The starting point for system identification is the choice of the underlying model type which describes the plant and its dynamics. The general model structure is based on differential equations and transfer functions. In practice, several model structures are used and can be classified by their attributes (linear/non-linear, time variant/time invariant, non-parametric/parametric like linear parameter varying (LPV) models etc.). Common model structures are AR (**A**uto-**R**egressive model), ARX (**A**uto-**R**egressive model with e**X**ogenous inputs), ARMAX (**A**uto-**R**egressive **M**oving **A**verage with E**X**ogenous Input), Box-Jenkins (combination of moving average and autoregressive approaches), **O**utput **E**rror (OE), or FIR. System identification is based on the following main steps:

1. record output data with proper input signal
   The input signal has to excite all relevant frequencies of the system in order to provide significant output data.
2. choose model structure
   - grey box model
     based on a physics model; preserves known structures with a number of unknown free parameters
   - black box model
     no underlying physics model, parameters have no direct physical meaning
3. estimate model parameter
   The error $e(k)$ has to be minimized by finding an appropriate parameter set θ for the model.
4. validate the model using a set of data different from that employed in the identification process

Attempts to identify the dynamics of superconducting cavities with Lorentz force detuning have been made [23]. As an example, the Output Error model applied to an RF cavity is illustrated in Fig. 37. In this model, no parameters are used to describe the disturbance characteristics $d(t)$ and are therefore

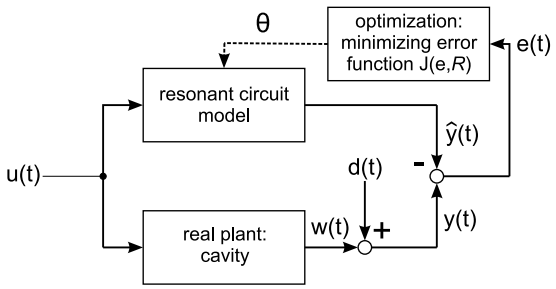


Fig. 37: Structure of the Output Error (OE) model for system identification

simply added to the system output. The model describes the system dynamics separately. To identify the

dynamics, the same inputs $u(k)$ are applied to the model as to the real plant. The resulting differences in the output $e(t) = y(t) - \hat{y}(t)$ are used to optimize the parameter set θ which describes the model. This is achieved by defining an error function with weighting matrix R. Apart from the goal of synthesizing an optimal controller and obtaining a model for feed-forward generation, system identification can also give insights about the system dynamics and provide information about not directly measurable states. Examples for these states are Lorentz force detuning $\Delta\omega(t)$ during an RF pulse [see Eq. (29)], the beam phase with respect to the RF field, or the cavity bandwidth, which is mainly determined by the external coupling [24]. The principle of this approach is illustrated in Fig. 38. As is indicated, either the *grey*

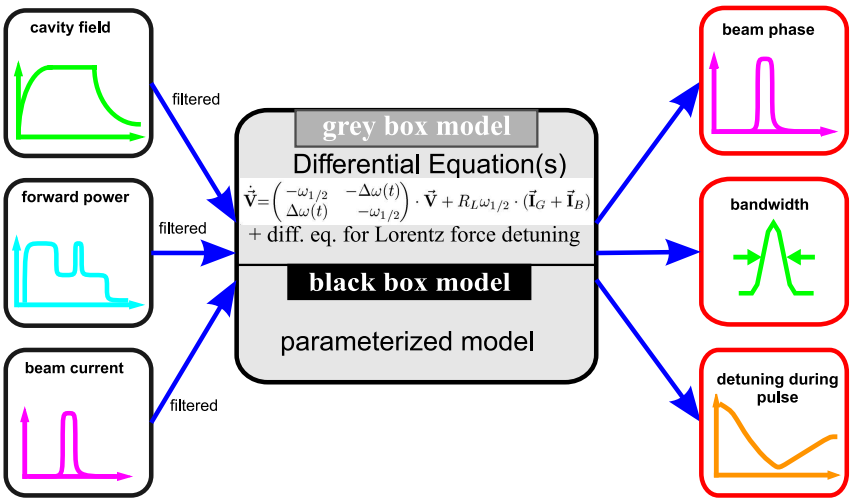


Fig. 38: Example: system identification in pulsed high-gradient superconducting cavities with Lorentz force detuning

box model (based on the cavity model of a LCR circuit in conjunction with differential equations for the Lorentz force detuning [25]) can be chosen, or a *black box model* is applied, in which case the determination of the appropriate model structure and order is typically a trial-and-error process.

6 Conclusion

The development of digital systems in the field of RF applications for accelerators has expanded significantly in the past decade due to the availability of greatly improved high-speed digital electronic hardware. Fast feedback systems, which were usually the domain of analog systems, have been replaced more and more by digital systems if the required regulation bandwidths are not too high. Digital low-level RF systems have much in common with many other accelerator RF sub-systems such as beam diagnostics. Common platforms for both areas are possible, but with dedicated analog front-end electronics. The extensive diagnostics capabilities in digital RF systems allow the implementation of fully automated procedures and calibration schemes for those complex systems. Because of high flexibility and ease of reconfiguration of software-based digital systems, more sophisticated algorithms than currently used are expected to emerge in coming years.

References

[1] M.E. Angoletta, *Digital Low Level RF*, EPAC'06, Edinburgh, Scotland, 2006, p. 1847.

[2] R. Garoby, *Low Level R.F. Building Blocks*, CAS CERN Accelerator School, RF Engineering for Particle Accelerators, Oxford, 1991, CERN–92–03, pp. 428–457.

[3] M. Grecki, T. Jezynski and A. Brandt, *Estimation of IQ Vector Components of RF Field - Theory and Implementation*, MIXDES 2005, Cracow, Poland, 2005, pp. 783–788.

[4] L. Doolittle, H. Ma and M.S. Champion, *Digital Low-Level RF Control Using Non-IQ Sampling*, LINAC'06, Knoxville, Tennessee, USA, 2006, p. 568.

[5] E. B. Hogenauer, *An Economical Class of Digital Filters for Decimation and Interpolation*, IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-29, no. 2, April 1981, pp. 155–162.

[6] M.E. Frerking, *Digital Signal Processing in Communication Systems* (Van Nostrand Reinhold, New York, USA, 1994).

[7] Xilinx Intellectual Property Center, datasheet, *Cascaded Integrator-Comb (CIC) Filter*, California USA, March 2002.

[8] B.F. Wu and E.A Jonckheere, *A Simplified Approach to Bode's Theorem for Continuous-Time and Discrete-Time Systems*, IEEE Trans. Autom. Control, vol. 37, Issue 11, Nov. 1992, pp. 1797–1802.

[9] B.F. Wu and E.A Jonckheere, *Chaotic Disturbance Rejection and Bode Limitation*, American Control Conference, Chicago, Illinois USA, June 24–26, 1992, TP1, pp. 2227–2231.

[10] C.G Montgomery, R.H. Dicke and E.M. Purcell, *Principles of Microwave Circuits*, Radiation Laboratory Series Vol. 8 (McGraw-Hill, New York, 1942), Ch. 7.

[11] T. Schilcher, *Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities*, DESY Print, TESLA 1998-20, 1998.

[12] S. Simrock, *Lecture on Control Theory*, these proceedings, 2007.

[13] S. Michizono *et al., Digital Feedback System for J-Parc Linac RF Source*, LINAC'04, Luebeck, Germany, 2004, p. 742.

[14] S. Michizono *et al., Performance of a Digital LLRF Field Control System for the J-PARC Linac*, LINAC'06, Knoxville, Tennessee, USA, 2006, p. 574.

[15] J. Belleman, V. Chohan, J.L. Gonzalez, S. Johnston, E. Schulte and E. Thivent, *Measurement of the Mean Radial Position of a Lead Ion Beam in the CERN PS*, EPAC'96, Sitges, Spain, 1996.

[16] E. Onillon and J.M. Brennan, *The New BNL AGS Phase, Radial and Synchronization Loops*, EPAC'96, Sitges, Spain, 1996.

[17] D. Boussard, *Design of a Ring RF System*, CAS CERN Accelerator School, RF Engineering for Particle Accelerators, Oxford, 1991, CERN–92–03, pp. 474–500.

[18] R. Zhang, I. Ben-Zvi and J. Xie, *A Self Adaptive Feedforward RF Control System for Linacs*, Nucl. Instrum. Methods Phys. Res. **A324** (1993), pp. 421–428.

[19] M. Liepe and S. Simrock, *Adaptive Feed Forward for the Digital RF Control System at the TESLA Test Facility*, EPAC'98, Stockholm, Sweden, 1998, pp. 1735–1737.

[20] M. Liepe, *Regelung supraleitender Resonatoren mit Strahlbelastung am TESLA-Test-Linearbeschleuniger*, diploma thesis, University of Hamburg, 1998.

[21] A. Brandt, *Development of a Finite State Machine for the Automated Operation of the LLRF Control at FLASH*, PhD thesis, University of Hamburg, 2007.

[22] H. Ma, private communication, ORNL, Tennessee USA, 2007.

[23] G. Koch, *Modelling of an Accelerator Based X-ray Free Electron Laser System for Controller Design*, diploma thesis, Technical University of Hamburg-Harburg, 2005.

[24] S. Simrock, *Digital Low-Level RF Controls for Future Superconducting Linear Colliders*, PAC'05, Knoxville, Tennessee USA, 2006, pp. 515–519.

[25] M. Hüning, S. Simrock, *System Identification for the Digital RF Control System at the TESLA Test Facility*, EPAC'98, Stockholm, Sweden, 1998, pp. 1732–1734.