Development of a Finite State Machine for the Automated Operation of the LLRF Control at FLASH

Dissertation zur Erlangung des Doktorgrades
Department Physik der Universität Hamburg

vorgelegt von
Alexander Brandt
aus Hamburg

Hamburg 2007

Gutachter der Dissertation	:	Prof. Dr. J. Roßbach Prof. Dr. T. Weiland
Gutachter der Disputation	:	Prof. Dr. J. Roßbach Prof. Dr. S. Khan
Datum der Disputation		25. Mai 2007
Vorsitzende des Prüfungsausschusses	:	Prof. Dr. C. Hagner
Vorsitzender des Promotionsausschusses	:	Prof. Dr. G. Huber
Dekan der MIN-Fakultät	:	Prof. Dr. A. Frühwald
Departmentleiter Physik	:	Prof. Dr. R. Klanner

Abstract

The entry of digital signal processors in modern control systems not only allows for extended diagnostics compared to analog systems but also for sophisticated and tricky extensions of the control algorithms. With modern DSP- and FPGA-technology*, the processing speed of digital systems is no longer inferior to analog systems in many applications. A higher degree of digitalization leads to an increased complexity of the systems and hence to higher requirements on their operators. The focus of research and development in the field of high frequency control has changed in the last few years and moved towards the direction of software development and complexity management.

In the presented thesis, a frame for an automation concept of modern high frequency control systems is developed. The developed automation is based on the concept of finite state machines (FSM), which is established in industry for years. A flexible framework was developed, in which procedures communicate using standardized interfaces and can be exchanged easily. With that, the developer of high frequency control components as well as the operator on shift shall be empowered to improve and adapt the automation to changed conditions without special programming skills required.

Along the automation concept a number of algorithms addressing various problems were developed which satisfy the needs of modern high frequency control systems. Among the developed and successfully tested algorithms are the calibration of incident and reflected wave of resonators without antennas, the fast adaptive compensation of repetitive errors, the robust estimation of the phase advance in the control loop and the latency adjustment for the rejection of instabilities caused by passband modes.

During the development of the resonator theory, high value was set on the usability of the equation in algorithms for high frequency control. The usage of the common nomenclature of control theory emphasizes the underlying mathematical structures of the equations. Several physical limitations and requirements, for example the limits of the vector sum calibration, were newly and adequately calculated based on the developed theory.

The linear accelerator of the Free-Electron Laser in Hamburg (FLASH) served as the main platform for testing of the algorithms and concepts. The developed automation, in particular the flexible and transparent framework and methods for the reduction of the complexity of the various communication channels (quantization) is not only suited for high fequency control but also for other aspects of an accelerator and beyond.

*DSP=Digital Signal Processor, FPGA=Field Programmable Gate Array

Zusammenfassung

Der Einzug digitaler Signalprozessoren in moderne Regelsysteme erlaubt nicht nur die Implementation weit umfangreichere Diagnosemöglichkeiten als bei analogen Systemen sondern auch sensible und trickreiche Erweiterungen der Regelalgorithmen selber. Durch moderne DSP- und FPGA-Technologie[†] steht die Geschwindigkeit digitaler Systeme der von analogen in vielen Anwendungsbereichen nicht mehr nach. Mit der Digitalisierung ist die Komplexität der Systeme gestiegen und damit auch die Anforderungen an diejenigen, die sie bedienen sollen. Damit ist die Optimierung der Regelung in Software und die Beherrschung der Komplexität mehr ins Zentrum von Forschung und Entwicklung bei Hochfrequenzregelungen gerückt als es noch vor ein paar Jahren der Fall war.

In der vorliegenden Arbeit wird der Rahmen für ein Automatisierungskonzept moderner Hochfrequenzregelungen entwickelt. Die hier entwickelte Automatisierung baut auf dem Konzept des endlichen Automaten auf, welches für industrielle Anwendungen als etabliert gilt. Es wurde ein modulares Framework entwickelt, in dem Prozeduren über standardisierte Schnittstellen miteinander kommunizieren und problemlos ausgetauscht werden können. Damit soll sowohl dem Entwickler von Komponenten der Regelung als auch dem Operateur auf Schicht die Möglichkeit gegeben werden, einzelne Aspekte der Automatisierung zu verbessern beziehungsweise geänderten Bedingungen anzupassen, ohne dass dafür besondere Programmierkenntnisse nötig wären.

Entlang des Automatisierungskonzeptes wurden neue Algorithmen zu zahlreichen Problemstellungen entwickelt, welche den gestiegenen Anforderungen moderner Hochfrequenzregelungen gerecht werden. Unter den entwickelten und erfolgreich getesteten Algorithmen sind die Kalibration einfallender und reflektierter Welle bei Resonatoren ohne Antenne, die schnelle adaptive Vorauskompensation repititiver Fehler, die robuste Bestimmung des Phasenvorschubs im Regelkreis und der Latenzzeitabgleich zum Zwecke der Unterdrückung von Instabilitäten verursacht durch Passbandmoden.

Bei der Entwicklung der Resonatortheorie wird Wert auf die praktische Verwendbarkeit der Formeln in den Regelalgorithmen gelegt. Durch die Verwendung der in der Regelungstheorie üblichen Nomenklatur wird die Struktur der zugrundeliegenden Gesetzmäßigkeiten hervorgehoben. Auf Basis der entwickelten Theorie konnten an zahlreichen Stellen physikalische Limitierungen, zum Beispiel bei der Vektorsummenkalibration, neu und angemessen ermittelt werden.

Getestet wurden die entwickelten Verfahren zu größten Teilen am Linearbeschleuniger des Free-Electron Lasers in Hamburg (FLASH). Die entwickelte Automatisierung, insbesondere das flexible und transparente Framework und die Verfahren zur Reduktion der Komplexität der zahlreichen Kommunikationskanäle (Quantisierung) ist nicht nur für die Hochfrequenzregelung geeignet sondern auch für andere Bereiche eines Teilchenbeschleunigers und darüber hinaus.

[†]DSP=Digital Signal Processor, FPGA=Field Programmable Gate Array

Table of Contents

Light at the end of the tunnel.

1. Introduction

The Free-Electron Laser Hamburg (FLASH) is a unique pulsed light source for a variety of modern sciences including medicine, biology and material research. The free-electron laser (FEL) is to be distinguished from other light sources by its high intensity ($\sim$ 1 GW) in a spectral, spatial and temporal confined space, in summary by its high brilliance. The wavelengths of FLASH are tunable in the range between 13 and 45 nm, which makes FLASH suiteable for the exploration of very small structures.

The electrons participating at the FEL process are subject to strong requirements on their longitudinal and transversal dimensions (required peak current $\sim$ 2 kA, required emittance $\sim$ 3 mm-mrad). The properties of the electron beam are determined by the radio frequency (RF) system in combination with the beam optics, which puts strong requirements on the RF system. Typically, an amplitude respectively phase stability of 0.1 % and 0.1° is required. For next generation light sources, such as the european XFEL project, the requirements are even tighter by one order of magnitude.

FLASH uses superconducting nine cell cavities for particle acceleration, operated at a frequency of 1.3 GHz. One RF amplifier (klystron) feeds up to 16 cavities. In the next expansion stage, this will be extended to 24 cavities, in future linear accelerator projects it might even be 32 or 48. An peculiar challenge for the low level RF (LLRF) control at FLASH origins from the superconducting nature of the cavities as well as the concept of vector-sum control, where one LLRF system controls the collection of 16 cavities together with one klystron as actuator. Superconducting cavities have an extremelty high quality factor ($Q_0 \approx 10^9$), leading to time constants of the loaded cavities of the order of the pulse length itself (1−2 ms). Additionally, the bandwidth ($\omega_{1/2} \approx 200$ Hz) is so small that changes in geometry of either mechanical or electromagnetic origin affect the field quality significantly.

With the entering of digital signal processors in modern control systems the complexity of service has increased. Digital systems offer a broader variety of diagnostics as well as the possibility for more sophisticated control algorithms. With modern computing devices, such as digital signal processors (DSP) and field programable gate arrays (FPGA), the processing speed of digital systems is no longer inferior compared to analog systems for many applications. The optimization of control algorithms as well as the control of complexity of the whole system moved more into the foreground of todays research and development compared to a few years ago.

Industrial automation is often based on finite state machines. This concept, developed by the informational sciences, offers with the help of state charts powerful tools for modeling of discrete systems. Many software packages, including the control system of FLASH, support this concept. Contrary to industrial automations, it is required from the automation of the high frequency control of FLASH that it is still functioning even if the operator is bypassing the automation and modifies parameters directly at the high frequency control. Further, the automation is automat-

ing an environment which is still developing and changing. While the former requires a solid treatment of exceptions, the latter needs a flexible automation concept.

Chapter 2 gives a brief introduction into the theory of free-electron lasing with the example of FLASH.

Chapter 3 develops the formalism used for the algorithms developed in the framework of this automation. The formalism emphasizes the general nature of resonant modes and the different possible couplings to the modes. To this end it uses the language of control theory and introduces physical constants at a late stage. In section 3.9.1 and 3.9.2, the formalism is used for the estimation of the calibration requirements for vector-sums, where the variable under control is the sum of two dimensional vectors.

The next chapter, chapter 4, goes into details of special problems of LLRF control that are either resolved with the developed formalism and/or covered by the developed automation scheme. Among the developed and successfully tested algorithms are the calibration of incident and reflected wave of resonators without antennas, the fast adaptive compensation of repetitive errors, the robust estimation of the phase advance in the control loop and the latency adjustment for the rejection of instabilities caused by passband modes.

A new method for an automatic calibration of the measured incident and reflected waves at pulsed resonators is presented in section 4.3.1. The application of the calibration, which is performed fully automated within a few pulses, allowes the estimation of the fields inside the resonator as the sum of the measured and calibrated incident and reflected wave. An algorithm for the pre-compensation of repetitive errors was developed and successfully tested within a few tenthousend iterations. It is described in section 4.4.6. The FLASH high frequency control holds ready tables for the pre-compensation of repetitive errors which are applied in every pulse. The algorithm calculates these tables on the basis of the data of the preceding pulse using digital filters and is thus very well suited for the implementation in digital devices. A method for the estimation of the phase advance in the control loop was developed and proved to be suitable for the compensation of fluctuations in the so-called loop phase. As described in section 4.4.3, the phase advance in the loop can fluctuate by several ten degrees under the influence of variations of the temperature which would be critical for the stability of the control loop without compensation. A very simple algorithm which is capable of increasing the latency of the control loop by several nanoseconds was used for the demonstration of the destabilizing effects of modes beside the fundamental accelerating mode in the superconducting resonators. The same algorithm can be used in the normal operation of the control loop for adjusting the latency in a way that the modes beside the fundamental mode practically do not destabilize the control loop. The results are presented in section 4.4.2.

In chapter 5, the concept for an automation framework based on finite state machines is presented. A flexible framework was developed, in which procedures communicate using standardized interfaces and can be exchanged easily. With that, the developer of high frequency control components as well as the operator on shift shall be empowered to improve and adapt the automation to changed conditions without special programming skills required.

The last chapter 6 gives an overview over the available procedures. Each procedure is compatible with the developed framework and can be applied in a modular fashion.

2. Linac Based Free-Electron Laser

2.1. Linear Accelerator for Free-Electron Lasers

When building a free-electron laser based on a linear accelerator (linac), the largest effort in terms of manpower and finance goes into the linac itself. Its purpose is to deliver a bright, high-energy electron beam. A linear accelerator, compared to a circular one, does not suffer from energy-loss due to synchrotron-radiation and the longitudinal emittance is thus not limited by the synchrotron-radiation process. While in a synchrotron or storage ring particles reach their maximum energy after several revolutions, a linac transfers the energy with a single pass. The main components of a linear accelerator shall briefly be introduced at the example of FLASH (figure 2.1).

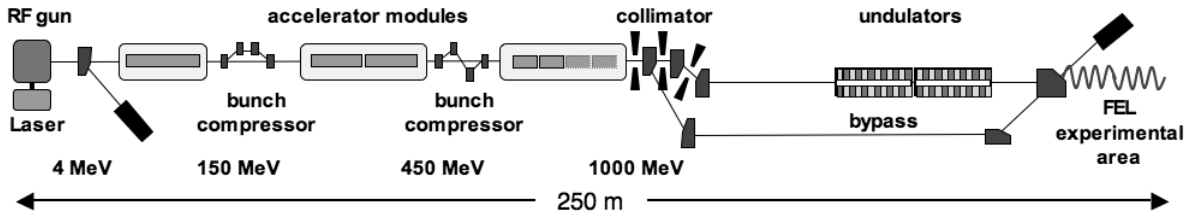

Figure 2.1.: Layout of FLASH showing the main components of the linac and the undulators.

Particle Source

The particle source is the origin of the electrons that are accelerated further downstream the linac. The particle source at FLASH is a photoinjector, where electrons are produced by means of the photoeffect. A Nd:YLF laser at 1047 nm is twice frequency doubled and focused on a cathode of cesium telluride where it liberates electrons. The pulses of the laser arrive with a periodicity of 1 μs and a width of about 4 ps rms and thereby define the bunch structure of the electrons [1]. The cathode is surrounded by a $1\frac{1}{2}$-cell copper resonator which provides an accelerating field directly at the origin of particle production of up to 45 MV/m. Within a short distance, the electrons are accelerated to 5 MeV, an energy at which the emittance of the beam is no longer broadened by space charge effects. The resonator of an RF photoinjector is often referred to as the RF gun [2].

Superconducting Linac

Particles are further accelerated downstream the RF gun to the required energy for FEL operation of about half a GeV. The acceleration is achieved by the alternating fields of 9-cell resonators build from pure niobium [3]. If properly matched to the generator frequency, RF resonators can build up high alternating electric fields with low power losses. The resonators are helium-cooled and operated at a temperature of 1.7 K, where niobium is superconducting. The quality factor of resonators made from superconducting niobium is of the order of 10^{10} which allows for very high field at low RF losses compared to normal-conducting resonators. However, superconducting resonators introduce additional subtleties like limits on duty-cycles due to heat disspiation at very low temperatures and a sensitivity to environmental noise due to thin walls and very narrow bandwidths. Geometric distortions lead to a significant shift in the center frequency of the accelerating mode, referred to as detuning. Detuning is caused by mechanical influences, microphonics [4], or time varying radiation pressure, Lorentz force [5].

At FLASH, eight nine-cell superconducting cavities are grouped into one cryomodule. One RF-amplfier (klystron) serves for one to four modules. Figure 2.3 shows the pulse-structure for a typical example-configuration. The total pulse length is of the order of the time-constant of the cavity itself. For each pulse, beam acceleration is only done during the flattop of 800 μs in total. Bunches traverse the cavity every 1 μs. The flat top starts after the filling time which lasts for about 500 μs.

Beam Optics

The trajectory of the particle beam through the vacuum of the beam pipe is determined by the beam optics which is based upon the deflection of charged particles under the influence of a magnetic field. The different types of magnets used in the accelerator correspond to the elements of a multipole-expansion of a general magnetic field distribution. Dipoles are used to steer the beam through the vacuum pipe by causing a deflection of all particles in a plane that is perpendicular to the magnetic field. Combinations of alternating focussing and defocussing quadrupoles control the diameter of the beam in the horizontal as well as in the vertical plane along the linac. Some dispersive sections of FLASH are equipped with sextupoles for correction of chromaticity. Octupolemagnets are not forseen at FLASH. An overview over current beam optics studies is found in [6].

Bunch Compression

A bunch compressors can be considered a special purpose optics device. Its purpose ist to compress each bunch in the longitudinal direction in order to achieve the peak currents of several kA that are required for the FEL process. In a bunch compressor, four dipoles are set up in a way that the beam undergoes a chicane, either D-shaped or S-shaped (refer to figure 2.1). The bending radius of a single particle trajectory under the influence of a homogeneous magnetic field is inversely proportional to the energy of the particles, therefore particles of higher energy pass the bunch-compressor on a straighter trajectory. Higher energy particles, if they arrive later in the bunch, catch up with the lower energy particles. Off-crest acceleration in the module before the

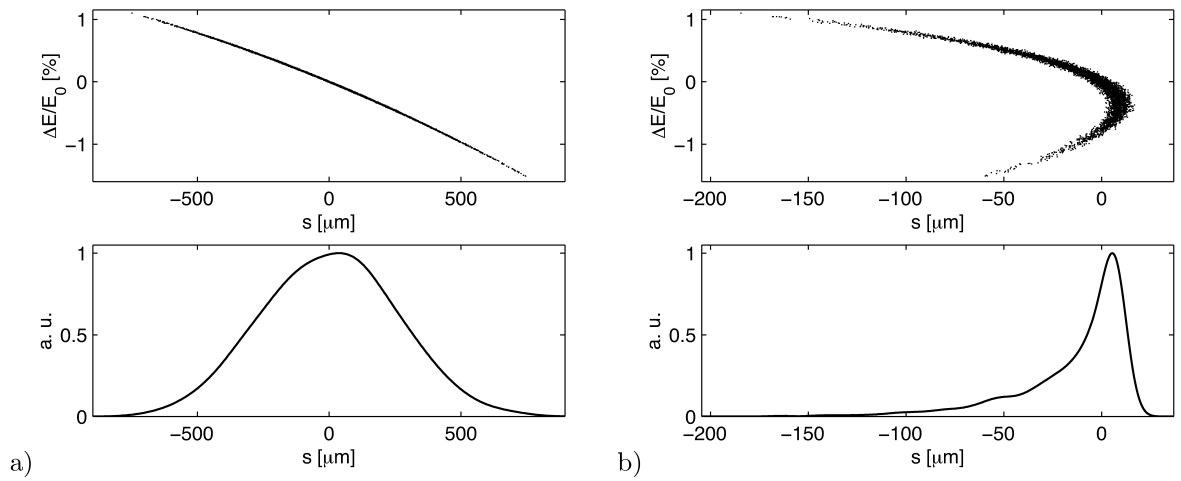

Figure 2.2.: The longitudinal phase space distribution and the charge profile are shown in front of the chicane (a) and behind it (b). Due to the small initial curvature, the phase space distribution folds over during compression and a huge amount of charge accumulates at the head of the bunch.

bunch compressor ensures that the later particles have a higher energy. Figure 2.2 illustrates the process inside a bunch compressor. The two columns show the electron energy distribution and the longitudinal electron density before and after the bunch compressor. Due to the change of the electron positions inside the bunch, a region with a very high peak current is produced. A detailed description of the processes inside a bunch compressor can be found in [7]. In order to remove nonlinear longitudinal energy profiles on the bunch an RF section with a higher harmonic frequency of the fundamental accelerating mode can be used.

Diagnostics

The beam diagnostic of an accelerator is a field that uses a great variety of techniques in order to determine energy, position, charge, shape and arrival-time of individual bunches or bunch trains. Transverse positions are typically determined from electromagnetic interaction with an additional impedance inserted into the vacuum pipe (beam position monitor). Cameras with a direct connection to the accelerator control system provide the operator with images of luminescent screens. The screens image properties of the particle beam by direct interaction with the electrons or with the emitted photons in dispersive regions. The charge of the electron beam is measured by toroids. The beam induces a voltage in the windings of a coil that is further processed and digitized. Information about the longitudinal shape of the bunch is obtained by a transverse deflecting structure that tilts the spacial orientation of a single bunch. The tilted bunch is projected on a luminescent screen and thus reveals information over its longitudinal profile (refer also to [8]). Further installations for the bunch shape diagnostic include electro-optical methods, like the timing electro-optical (TEO) experiments. Altogether, the beam diagnostic is

a complex system of monitoring devices that allowes to adjust the various number of parameters in a linac in order to have a good beam quality.

Control System

The control-system of a linear accelerator is a distributed network with the purpose to interconnect all active components. Physically, the Distributed Object Oriented Control System (DOOCS) of FLASH is built from distributed CPUs which are connected by the internet protocol (IP). Each CPU is a full-fledged computer with a hard drive, memory and a Unix operating system installed.

The majority of these units is embedded in industrial VME busses with direct connection to the hardware of the active components of the accelerator, as for example the beam diagnostics. The purpose of the embedded computers is to collect data from the hardware (e. g. beam diagnostics) and supply it to the network. On the other hand, the embedded computers receive commands from the network and translate them into signals of e. g. a stepmotor of a valve of the accelerator. The embedded computers form the lowest layer of the three layer model of DOOCS, their purpose is to provide data from the devices to the network and to communicate commands from the network to the decives.

The middle layer of the control system is again a set of CPUs with a direct connection to the control network. These CPUs are usually not embedded into a bus system and not necessarily have a direct cable connection to devices of the accelerator. The middle layer CPUs are high performance machines with large processing capabilities. The programs that run on the middle layer CPUs are called middle layer servers. Their purposes are manifold and range from signal processing to data archiving. For example, middle layer servers can collect data from several locations of the lower layer CPUs and mathematically derive new quantities. These derived quantities are available in the control network via the same protocols as the data from the lower layer servers. Another example of a middle layer server is the implementation of the automation presented in this thesis.

On the top level, the control system presents a graphical interface to the operator. The elements display information from the two lower layers of the control system, as textual or graphical representation to the operator. The interface also allows to modify quantities at the system remotely and with this trigger predefined action at the two layers below, as for example switching of a high power amplifier or the adjustment of the current in a magnetic coil.

2.2. Free-Electron Laser Process

The free-electron laser process (FEL process) is a collective instability of the electron beam where the energy is transferred into a photon beam of high brilliance. Its high power at small wavelengths, tuneability and coherence it yet unreached by conventional lasers and opens new possibilities for several fields of science. The electron motion in an alternating magnetic field is in the starting point for the FEL process. This section introduces the particle trajectory and spontaneous radiation in an undulator and the low and high gain FELs.

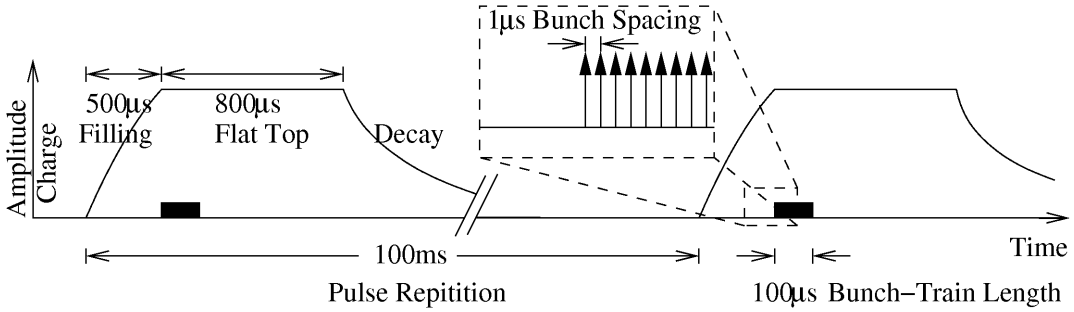

Figure 2.3.: Typical configuration of the pulsed operation at FLASH. The figure introduces names for the individual phases of a pulse (filling, flattop and decay) and gives examples for the duration of individual elements.

Electron Trajectory in an Undulator

At the end of the linac described in section 2.1, a so-called undulator magnet establishes an alternating magnetic field from permanent magnets which forces the electrons on a periodic trajectory and makes them emit spontaneous synchrotron radiation. With λ_u and B_u being the undulator period and the peak magnetic field, a condition for positive interference can be stated for the photons emitted at a wavelength λ:

$$\lambda = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2}\right), \quad \text{where} \quad \gamma = \frac{E}{m_e c^2} \quad \text{and} \quad K = \frac{eB_u\lambda_u}{2\pi m_e c} \tag{2.1}$$

The interference condition has its origin in the relativistic Doppler-effect (factor $1/2\gamma$), the relativistic length contraction ($1/\gamma$) and the slippage of the electron orbit with respect to a relativistic particle on a straight orbit ($K^2/2$). A photon emitted by an electron in the forward direction traverses the undulator on a straight path which is shorter than the periodic trajectory of the electron, see also figure 2.4. The resonance condition 2.1 is true if the retardation of the electron path between a period of the undulator with respect to a relativistic particle on a straight path is equal to the photon wavelength. A more useful form of the undulator parameter K is $K \approx 0.93\lambda_u\,[\text{cm}]B_u\,[\text{T}]$. For a given undulator and a given wavelength of the electromagnetic wave λ, a resonant γ_{res} can be found according to equation 2.1 and is defined as

$$\gamma_{res} = \sqrt{\frac{\lambda_u}{2\lambda}\left(1 + \frac{K^2}{2}\right)}. \tag{2.2}$$

Spontaneous Radiation

While a radiating dipole has a cylindric radiation characteristics in the far zone, the synchrotron radiation of a relativistic particle is pointing in the direction of the particle velocity with an opening angle that is in good approximation $1/\gamma$. In an undulator, the radiation characteristics is

additionally determined by the number of periods. An undulator is distinguished from a wiggler which is also an alternating magnetic structure, by the parameter K, which is larger than 1 for a wiggler and less or equal to 1 for an undulator. If $K < 1$, the angle between the symmetry axis of the undulator and velocity vector of the charged particles is less or equal to $1/\gamma$ which ensures overlapping of the emitted light cones. The opening angle of the total emitted radiation decreases by interference with the square-root of the number of periods, $\sqrt{N_\mathrm{u}}$. The relative amount of the power emitted in the forward direction is therefore proportional to N_u. On the other hand, the total emitted energy by an electron is proportional to the path length in the magnetic field, or equivalently to the number of periods, N_u. Altogether, one can state that the total amount of energy in an undulator that is emitted in the forward direction is proportional to N_u^2/γ.

The spectrum of undulator radiation is dominated by a line at the frequency that is given by the resonance condition in equation 2.1. Odd higher harmonics of the fundamental wavelength λ exist with an amplitude that is smaller by several orders of magnitude compared to the fundamental. The spectral width of the fundamental line is proportional to $1/N_\mathrm{u}$.

The radiation power emitted by a single point-like particle that carries out a harmonic oscillation at a circular frequency ω_0 with an amplitude x_0 scales with the square of the charge q^2,

$$P = \frac{q^2 x_0^2 \omega_0^4}{6\pi\epsilon_0 c^3}. \tag{2.3}$$

Synchrotron radiation from a bunch of electrons is a stochastic process and the photons emitted from different electrons add up incoherently. The total power of the synchrotron radiation therefore scales linearly with the number of electrons, N. Coherent emission of photons would scale with the square of the number of photons, N^2. However, coherent emission requires that the sources of the radiation are positioned with equal distance along the longitudinal axis of the radiation. This is the case FEL radiation.

Low Gain FEL

In a low gain FEL, the interaction between an electron beam on its oscillatory trajectory through the undulator and a coaxial electromagnetic wave,

$$E(z,t) = E_0 \cos(kz - \omega t + \psi_0) \tag{2.4}$$

is considered. In the low gain approximation, the amplitude E_0 is assumed to be constant since the energy transfer between the electron beam and the electromagnetic wave is small compared to the energy of the electromagnetic wave. The transverse dependency of the electromagnetic wave is neglected. In a low gain FEL, the particle beam in the undulator is an active medium that is capable of amplifying an electromagnetic wave by a few percent in a single passage. In optical resonators with mirrors of high reflectivity, low gain FELs are used as sources for intensive light of high coherence, compareable to conventional lasers. The achievable wavelengths is limited by the availability of mirrors of high reflectivity and is typically above 150 nm.

The trajectory of the electrons is determined by the alternating magnetic field of the undulator. Each position along the z-axis of the undulator can be assigned a phase $k_\mathrm{u} z$ of the electron motion

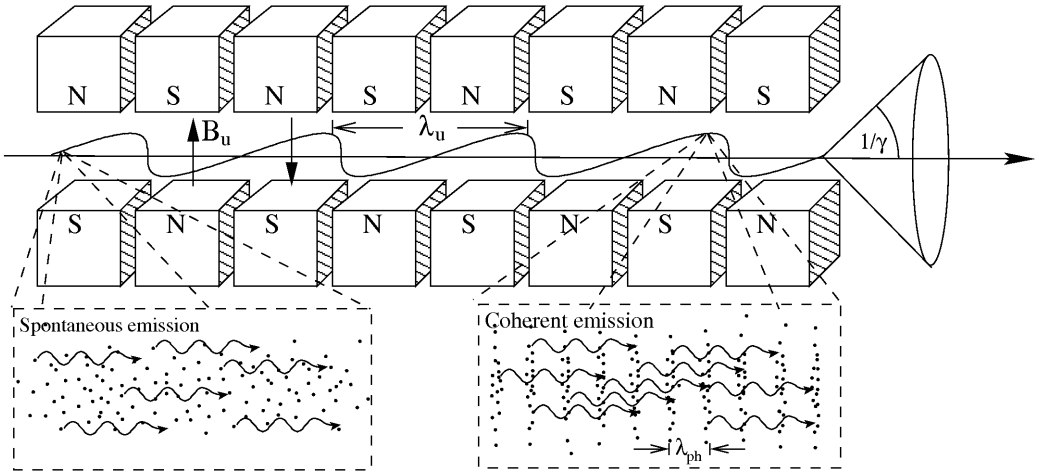

Figure 2.4.: Illustration of the SASE-process. Incorherent radiation dominates at the beginning of the undulator while at the end the electron beam shows a density modulation with a periodicity corresponding to the emitted photon-wavelength.

with $k_\mathrm{u} = 2\pi/\lambda_\mathrm{u}$. The sum of the phases of the electromagnetic wave in equation 2.4 and the phase of an electron trajectory along the undulator is defined as the ponderomotive phase ψ,

$$\psi(z,t) := (k + k_\mathrm{u})z - \omega t + \psi_0. \tag{2.5}$$

The ponderomotive phase ψ depends on the frequency of the electromagnetic wave and the geometry of the undulator. Equation 2.5 assigns to each point in space and time a ponderomotive phase. The periodicity in space of the ponderomotive phase ψ along the undulator axis for a fixed time is nearly the wavelength λ, since $k \gg k_\mathrm{u}$. Each electron is assigned a ponderomotive phase according to its position $z(t)$ at the time t. In an electron bunch with a length of greater than the wavelength λ, the electrons will cover all ponderomotive phases between 0 and 2π uniformly.

The pondermotive phase of an electron determines if spontaneous radiation of the electrons interfers constructively or destructively with the electromagnetic wave. If the ponderomotive phases of a bunch of electrons that obey the resonance condition 2.1 is distributed uniformly, the net energy transfer between the electromagnetic wave is zero.

In the low gain approximation it is shown that a uniform distribution of ponderomotive phases of a collection of electrons that are slightly above or below the resonance energy (equation 2.2) changes. The result is a non-uniform distribution of the ponderomotive phases and a net energy transfer. Deceleration and acceleration leads to a microbunching of the electrons on the scale of photon wavelength. If the initial electron energies are slightly above the resonance condition, the energy transfer will lead to an increased energy of the electromagnetic wave. In case the electron energies are below the resonance energy, the electrons will gain energy from the electromagnetic wave.

High Gain FEL

Creation of a high power beam of coherent photons at wavelengths below 150 nm requires a scheme without optical elements since. If an electromagnetic wave as in equation 2.4 is amplified by several orders of magnitude by the interaction with an electron beam in an undulator the process is no longer described by the low gain approximation. The electromagnetic wave in equation 2.4 is extended by a z-dependent envelope,

$$E(z,t) = E_0(z)\cos(kz - \omega t + \psi_0(z)). \tag{2.6}$$

The theoretical treatment of the high gain FEL calculates a solution for the envelope of the electromagnetic wave $E_0(z)$ and the charge density of the electromagnetic beam that is consistent with the Maxwell equations and the dynamics of the electrons that result from the low gain approximation. The charge density distribution is no longer negligible since the microbunching is in the order of the wavelength of the emitted radiation. The ansatz for charge density is

$$\rho(z) = \rho_0 + \rho_1(z)\cos\psi(z) \tag{2.7}$$

with a uniformly distributed part (ρ_0) and a part that is bunched with the periodicity of the ponderomotive phase (ρ_1). The treatment of the equations yields a third order differential equation that can be solved analytically ([9]). In a high gain FEL, the interaction between the electrons and the electromagnetic wave leads to a microbunching of the particles which, in turn, increases the power of the electromagnetic wave by coherent emission.

The result is an exponential amplification of the electromagnetic wave along the z-axis of the undulator with a characteristic gain length L_g,

$$L_\mathrm{g} = \frac{1}{\sqrt{3}}\sqrt[3]{\frac{4\gamma^3 m_\mathrm{e}}{\mu_0 K^2 q^2 k_\mathrm{u} N}}. \tag{2.8}$$

Due to the high amplification of a high gain FEL which can be 10^6 the amplitude of the initial electromagnetic wave can be rather small. At FLASH, the initial electromagnetic wave is obtained from the spontenaous synchrotron radiation. An FEL starting from noise is abbreviated SASE or self amplifying spontaneous emission. The first lasing of FLASH is summarized in [10].

2.3. Requirements on the Accelerating Fields

The requirements on the accelerating fields for SASE operation can be derived from the required properties of the electron beam at the entrance of the undulator.

From equation 2.1 one can calculate the required energies for certain wavelengths. The FLASH undulator is a fixed 12 mm gap permanent magnet device with a period length of 2.73 cm and a peak magnetic field of 0.47 T, which leads to $K = 1.198$ [11]. Table 2.1 gives the required beam energies for a number of wavelengths. The lowest wavelength reached by FLASH so far is

13 nm and is limited by the available energy of the electron beam.

λ	γ	E
1 μm	153	78.2 MeV
10 nm	1531	782.4 MeV
0.1 nm	15311	7.8 GeV

Table 2.1.: Resonant electron energies for wavelength between 1 μm and 1 Å of the FLASH undulator.

Besides the total energy of the beam the energy distribution is of importance. The required maxiumum width of the energy distribution is in principle given by the spectral amplification profile that follows from equation 2.1. In [9] it is shown for the high gain FEL theory that the amplification profile narrows compared to the low gain FEL theory for an increasing number of undulator periods. The relative width of the amplification curve is typically below 10^{-3}. This has consequences for the tolerable energy spread inside a bunch, from bunch to bunch and from pulse to pulse. Inside an electron bunch, the energy spread may not exceed the 10^{-3}. Electrons outside this energy range will not participate at the FEL process. The energy spread from bunch to bunch is of particular interest in case the undulator is used inside a resonator with a revolving photon beam. In this case, the center of the allowed electron energy distribution is given by the wavelength of the light. A pulse to pulse fluctuation of the center energy of the electron bunches of larger than the 10^{-3} leads to a fluctuation of the wavelength of the radiation in case the FEL is self-amplified. Fluctuations of the wavelength are usually not wanted by the users of the FEL radiation. If the FEL is starting from a seed laser, only the electrons with energies around the accepted energy will participate at the FEL amplification.

Another set of requirements is related to the bunch compression and applies to the RF sections before the bunch compressors. FLASH has two bunch compressors installed and the process in these devices is briefly introduced in section 2.1. The process and thus the compression ratio depend strongly on the acceleration phase of the electrons in the RF sections before the bunch compressor. The sensitivity on phases is even increased if an RF section with a higher order mode is used in order to compensate for non-linearities in the energy profile. For an optimal compression, several working-points of the RF system in amplitude and phase exist. In [12], an amplitude and phase stabiliy of a few tenth of a percent and a few hundreth of a degree is estimated as a requirement in a typical scenario with a third order RF system. In the absence of a higher order RF system, the requirements are relaxed by about an order of magnitude.

Additional constraints can be the arrival-time of the electron beam in order to be able to support experiments that need synchronization with a secondary (pump-probe) laser. A typical number is 100 fs which is of the order of the bunch length [13].

A start to end simulation [14] that considers the setup of FLASH and takes into account several effects including the bunch-compression effects and the required energy spread in front of the undulator results in amplitude and phase requirements of 0.02 % and 0.05° for all accelerating modules. In this simulation, the errors in different RF stations with a single klystron are considered uncorrelated, which reduces the required field stability in single stations.

3. Field Dynamics in (Superconducting) Cavities

3.1. Envelope Approximation for Cavity Modes

Our general understanding of the field dynamics in super-conducting cavities is that each resonant mode can be treated as a harmonic oscillator. An LCR-circuit with properties equivalent to a cavity mode is introduced in [15] and [16] and from this the fundamental equations that describe the cavity dynamics are derived. The introduction of equivalent LCR-cirquits is only a convenience, while the "important equations refer to the resonant frequencies, their widths, amplitudes, and phases, and to the strength of coupling, quantities which at these frequencies are directly or indirectly easily measurable" ([17]). Refraining from the LCR-cirquit model has advantages when it comes to the calibration of in- and outgoing signals (e. g. for the loop-phase measurement in 4.4.3). Within this and the following chapters, a nomenclature known from control theory is utilized, where the system (the cavity mode) is treated as a two-port with u as input and x as output. While the x is usually interpreted as the accelerating voltage, the u is the exciter of the oscillation and can be the beam current or a field amplitude of the forward traveling wave.

A harmonic oscillator describing a resonant mode of a cavity (superconducting or normal-conducting) is given by the equation

$$\ddot{x} + 2\omega_{1/2}\dot{x} + \omega_0^2 x = 2\omega_{1/2}\omega_0 u, \tag{3.1}$$

where $u(t)$ and $x(t)$ are the high-frequency in- and output signals of the system. Assuming the stimulation to be harmonic*, $u(t) = u_0 \cdot e^{i\omega t}$, the response of the system in steady-state will be of the same form, $x(t) = x_0 \cdot e^{i\omega t}$. Here, $u_0, x_0 \in \mathbb{C}$ and have amplitude and phase but are constant in steady-state. The exciting frequency ω does not necessarily coincide with the resonant frequency ω_0 of the mode. $\omega_{1/2}$ is called the half-bandwidth of the system.

For $u_0 = 1$ the steady-state amplitude and phase can be calculated to

$$|x_0| = \frac{2\omega_{1/2}\omega_0}{\sqrt{(\omega_0^2 - \omega^2)^2 + (2\omega_{1/2}\omega)^2}} \tag{3.2}$$

$$\angle x_0 = \arctan \frac{-2\omega_{1/2}\omega}{\omega_0^2 - \omega^2} \tag{3.3}$$

*Already here we use the complex notation. Later, this will simplify the notation significantly. For physical quantities usually only the real part is relevant.

For the modes of the cavities at FLASH, the half-bandwidth is small compared to the center frequency, $\omega_{1/2} \ll \omega_0$, therefore one can approximate $\omega_0^2 - \omega^2 = 2\omega_0\Delta\omega$ for small $\Delta\omega := \omega_0 - \omega$. $\Delta\omega$ is called the detuning of a mode[†]. In the case of accelerating modes, time variation of the detuning is caused by changes of the cavity mode itself (changes of ω_0) rather than changes of the exciting frequency ω. Equations 3.2 and 3.3 can now be written as[‡]

$$|x_0| = \frac{1}{\sqrt{1 + \left(\frac{\Delta\omega}{\omega_{1/2}}\right)^2}} \tag{3.4}$$

$$\angle x_0 = -\arctan\frac{\omega_{1/2}}{\Delta\omega} \tag{3.5}$$

It is still assumed that $u_0 = 1$. $\angle x_0$ therefore is $-90°$ for $\Delta\omega = 0$. Knowing that $-\arctan(1/x) = \arctan(x) + \pi$ one usually defines

$$\psi := \arctan\frac{\Delta\omega}{\omega_{1/2}} \tag{3.6}$$

and calls it the detuning-angle, $\psi = 0$ for $\Delta\omega = 0$. It should be noted here that $x_0 = 1$ for $\Delta\omega = 0$ thanks to the choice of the right-hand side of equation 3.1 which might have appeared arbitrary initially. Figure 3.1 illustrates the amplitude and phase of the system. With the definition of ψ, the equation for the amplitude, equation 3.4, can be expressed as $|x_0|^2 = 1/(1 + \tan^2\psi) = \cos^2\psi$ or $|x_0| = \cos\psi$. For a constant exciting amplitude u_0, the steady-state amplitudes of the fields therefore lie on a circle. Now that the steady-state of the system is described in terms of constant u_0 and x_0, the question arises if one can describe the transient behavior of u_0 and x_0 with a differential equation similar to equation 3.1. Therefore, x_0 and u_0 are no longer considered constant,

$$\begin{aligned} x &= x_0 e^{i\omega t}, \quad \dot{x} = (\dot{x}_0 + i\omega x_0)e^{i\omega t}, \quad \ddot{x} = (\ddot{x}_0 + 2i\omega\dot{x}_0 - \omega^2 x_0)e^{i\omega t}, \\ u &= u_0 e^{i\omega t}, \quad \dot{u} = (\dot{u}_0 + i\omega u_0)e^{i\omega t}, \quad \ddot{u} = (\ddot{u}_0 + 2i\omega\dot{u}_0 - \omega^2 u_0)e^{i\omega t}, \end{aligned} \tag{3.7}$$

and 3.1 reads

$$\ddot{x}_0 + 2(i\omega + \omega_{1/2})\dot{x}_0 + (\omega_0^2 - \omega^2 + 2i\omega\omega_{1/2})x_0 = 2\omega_{1/2}\omega_0 u_0. \tag{3.8}$$

Substituting the same approximation for $\omega_0^2 - \omega^2$ as above and dividing by $2i\omega$ yields

$$\frac{\ddot{x}_0}{2i\omega} + (1 + \frac{\omega_{1/2}}{i\omega})\dot{x}_0 + (-i\Delta\omega + \omega_{1/2})x_0 = -i\omega_{1/2}\frac{\omega_0}{\omega}u_0. \tag{3.9}$$

The fraction ω_0/ω can be set to one in good approximation. ω is considered large with respect

[†]With this definition, $\Delta\omega := \omega_0 - \omega$, a sign of the detuning is defined.

[‡]Because of the very good approximations $\omega_0 \approx \omega$ and $\omega_0 \gg \Delta\omega$, "$\approx$" is omitted. For the super conducting cavities at FLASH, $\omega_{1/2}/\omega_0 \sim 10^{-7}$ and for the normal conducting gun $\omega_{1/2}/\omega_0 \sim 10^{-5}$.

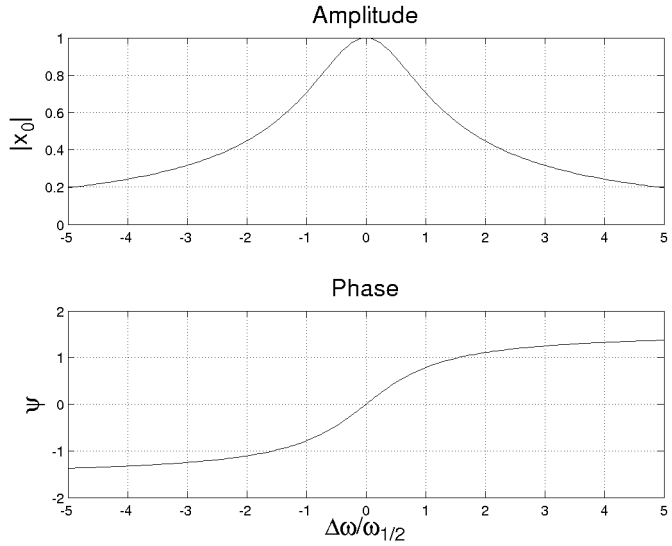

Figure 3.1.: Amplitude and phase of a bandpass in steady-state if the half-bandwidth is much smaller then the resonance frequency.

to $\dot{x}_0$ and $\ddot{x}_0/\omega_{1/2}$, therefore the terms with ω in the denominator are neglected,

$$\dot{x}_0 + (\omega_{1/2} - i\Delta\omega)x_0 = -i\omega_{1/2}u_0. \tag{3.10}$$

It is again desirable to have an equation with zero phase difference between x_0 and u_0 in the resonant case as earlier, when the detuning ψ was introduced. This is achieved by replacing $-iu_0$ with u_0. This leads to the following formula for the envelope of the signals that are transformed by a bandpass. Since the bandpass formula will no longer be used, the index 0 is omitted in the envelope approximation:

$$\dot{x} + (\omega_{1/2} - i\Delta\omega)x = \omega_{1/2}u \tag{3.11}$$

3.2. Transient Behavior of the Cavity Mode Equation

As depicted in figure 3.1, the steady-state solution of 3.11 for u = const. is given by

$$\begin{aligned} x_{\text{steady-state}} &= \frac{\omega_{1/2}}{\omega_{1/2} - i\Delta\omega}u \\ &= \frac{\omega_{1/2}}{\sqrt{\omega_{1/2}^2 + \Delta\omega^2}}e^{i\arctan\frac{\Delta\omega}{\omega_{1/2}}}u, \end{aligned} \tag{3.12}$$

which is at the same time a solution of the inhomogeneous equation. One homogeneous solution is

$$x_{\text{hom}} = e^{-(\omega_{1/2} - i\Delta\omega)t}. \tag{3.13}$$

From this, the step-response of the system ($u = 0$ for $t < 0$, $u = 1 = \text{const.}$ else) is easily found,

$$x_{\text{step}} = \frac{\omega_{1/2}}{\omega_{1/2} - i\Delta\omega}(1 - e^{-(\omega_{1/2} - i\Delta\omega)t})u. \tag{3.14}$$

Looking at the first derivative $\dot{x}_{\text{step}}(0) = \omega_{1/2}$ at $t = 0$, one sees that the direction of change of the mode points directly into the direction of the excitation (that is, $\dot{x}_{\text{step}}(0)$ is real). This is an important observation relevant for the control of a system where the excitation is updated with a clock-rate much smaller than the typical time-scale of the mode, $\tau = \omega_{1/2}^{-1}$. For such a controller, the mode never is in steady-state.

The figures 3.2, 3.3 and 3.4 show the transient behavior of the cavity at a timescale of $7\omega_{1/2}^{-1}$ for different detunings of up to $5\omega_{1/2}$. The parametric plot 3.4 gives an impression of the convergence of the system towards the circle that is formed by the steady-state solutions for different detunings. It should be noticed that the slopes of the amplitude and the phase in the beginning of the step-response are identical for all detunings in case of the amplitude but differ for the phases.

A lengthy calculation (refer to appendix A) reveals the slope of the transient amplitude and phase (still assuming that the drive is switched on at $t = 0$, i. e. $u = 1$ for $t > 0$):

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}|x_{\text{step}}|\right]_{t=0} = \omega_{1/2} \quad \text{and} \quad \left[\frac{\mathrm{d}}{\mathrm{d}t}\angle x_{\text{step}}\right]_{t=0} = \frac{\Delta\omega}{2} \tag{3.15}$$

These formulas are very helpful for examining a running system via “optical inspection”, since they provide fast information about bandwidth and detuning of a cavity.

While equation 3.15 provides information about bandwidth and detuning at the beginning of a pulse for constant drive, 3.13 (homogeneous solution) yields the same information at the end of a pulse (i. e. $u = 0$):

$$\frac{\mathrm{d}}{\mathrm{d}t}|x_{\text{hom}}| = \omega_{1/2} \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\angle x_{\text{hom}} = \Delta\omega \tag{3.16}$$

3.15 and 3.16 are similar except for a factor $1/2$ for the detuning-term. One more conversion of 3.16 reveals the beauty of the complex notation (obtained by differentiating the logarithm of 3.13, the homogeneous solution):

$$\operatorname{Re}\frac{\dot{x}_{\text{hom}}}{x_{\text{hom}}} = \omega_{1/2} \quad \text{and} \quad \operatorname{Im}\frac{\dot{x}_{\text{hom}}}{x_{\text{hom}}} = \Delta\omega \tag{3.17}$$

The last formula has a practical meaning, since it contibutes to a simplification of algorithms.

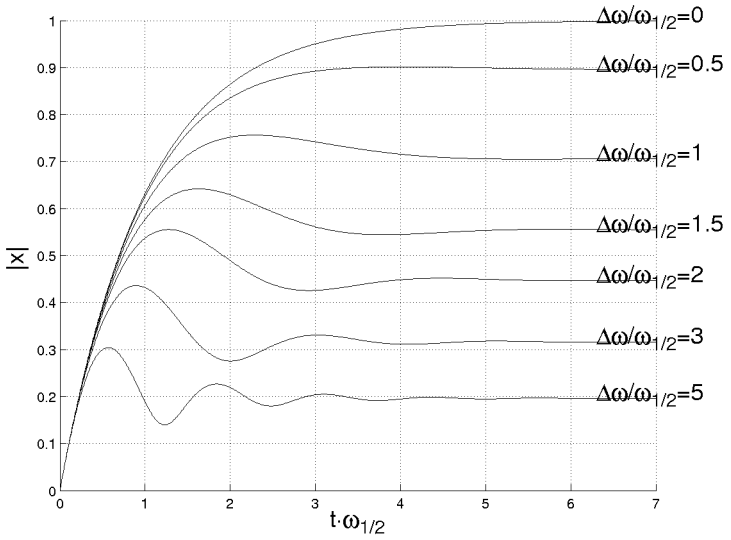

Figure 3.2.: Step response of the amplitude of a resonator for different detunings.

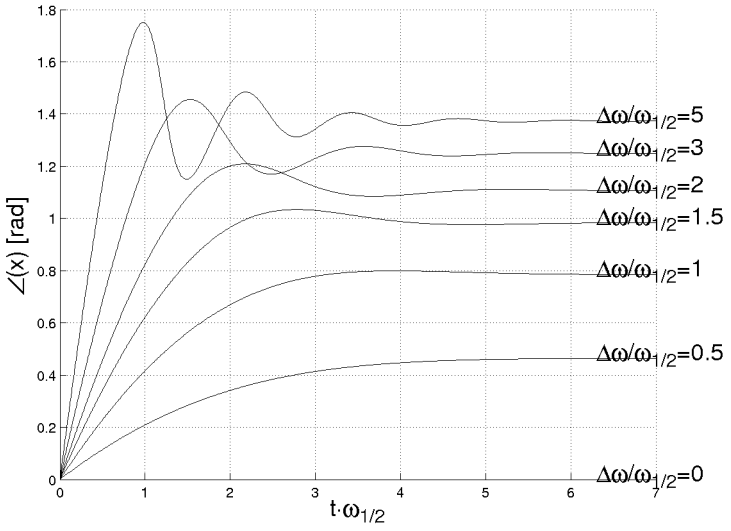

Figure 3.3.: Step response of the phase of a resonator for different detunings.

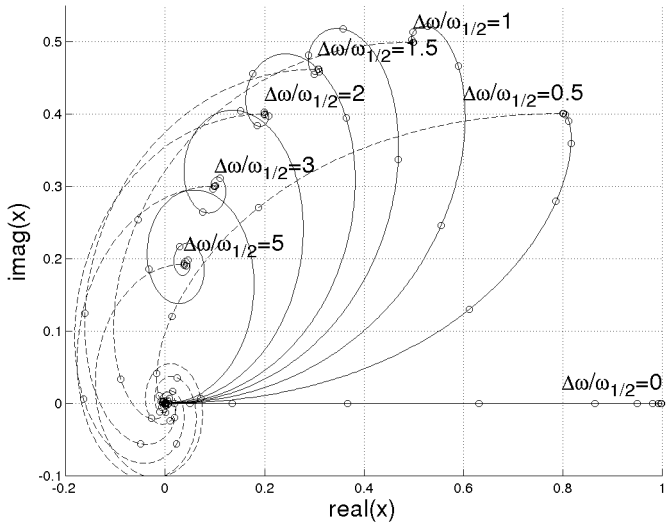

Figure 3.4.: Transient behavior in the complex plane for a constant excitation at different detunings. The step-response is shown for switch-on (solid) and switch-off (dashed) of the drive signal. The small circles mark the elapsing of a time-constant τ.

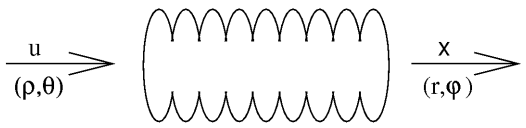

Figure 3.5.: Picture clarifying the nomenclature throughout this chapter: $u = \rho e^{i\theta}$ is transformed by the transfer function representing the cavity into $x = re^{i\varphi}$.

3.3. Cavity Mode Equation in Polar Coordinates

For a number of applications it is useful to have an equivalent for equation 3.11 in polar coordinates. The envelope of the signals is therefore expressed in polar coordinates.

$$u = \rho e^{i\theta},\ x = re^{i\varphi},\ \dot{x} = (\dot{r} + ir\dot{\varphi})e^{i\varphi} \tag{3.18}$$

Figure 3.5 illustrates this nomenclature. The cavity equation 3.11 can be written as

$$[\dot{r} + ir\dot{\varphi} + (\omega_{1/2} - i\Delta\omega)r]e^{i\varphi} \quad = \quad \omega_{1/2}\rho e^{i\theta} \tag{3.19}$$

This is separated into real and imaginary part by adding respectively substracting its complex

conjugate from each side of the equation. Two equations are the result.

$$\begin{aligned}(\dot{r} + r\omega_{1/2})\cos\varphi - (r\dot{\varphi} - r\Delta\omega)\sin\varphi &= \omega_{1/2}\rho\cos\theta \\ (\dot{r} + r\omega_{1/2})\sin\varphi + (r\dot{\varphi} - r\Delta\omega)\cos\varphi &= \omega_{1/2}\rho\sin\theta \end{aligned} \tag{3.20}$$

This is a linear system of two equations. It is simplified by eliminating the expressions in brackets,

$$\begin{aligned}(r\dot{\varphi} - r\Delta\omega)(\cos^2\varphi + \sin^2\varphi) &= \omega_{1/2}\rho(\sin\theta\cos\varphi - \cos\theta\sin\varphi) \\ (\dot{r} + r\omega_{1/2})(\cos^2\varphi + \sin^2\varphi) &= \omega_{1/2}\rho(\cos\theta\cos\varphi + \sin\theta\sin\varphi). \end{aligned} \tag{3.21}$$

Resolving some trigonometric identities yields the cavity equation in polar coordinates:

$$\begin{aligned} r\dot{\varphi} - r\Delta\omega &= \omega_{1/2}\rho\sin(\theta - \varphi) \\ \dot{r} + r\omega_{1/2} &= \omega_{1/2}\rho\cos(\theta - \varphi) \end{aligned} \tag{3.22}$$

It shall be emphasized that in the second equation of 3.22 the detuning $\Delta\omega$ does not appear. This will turn out to be useful for calibration procedures (e. g. the calibration of forward and reflected power of superconducting cavities in section 4.3.2) as well as for the calculation of the detuning of a cavity (section 4.4.4).

3.4. Coupling of the Beam to the Cavity Field

The dynamics of cavity modes has been deduced very generally. This and the following sections will introduce physical interpretations for these formulas. The introduction of physical quantities means to replace u and x by meaningful variables.

At FLASH, the electron beam traversing the cavities is bunched with a bunch spacing of $1\,\mu$s (see also the pulse structure in figure 2.3). Due to the finite length of each bunch, the electron beam can be considered as a sequence of Gaussian shaped charge clouds. Due to the limited bandwith of the cavity, only the fourier components of the electron beam around the resonance frequency of the cavity will excite fields. It has been shown in [15] that for a bunch length that is small with respect to the bunch distance, the fourier component I_{beam} at the resonance frequency is twice the average beam current $I_{\mathrm{beam},0}$.

In steady-state the beam induces the voltage of $U = R_{\mathrm{sh}} I_{\mathrm{beam}} l$ in a free-running cavity of length l. R_{sh} is called the shunt impedance of a cavity and depends on the material and on the geometry of the cavity. It is measured in MV/m. If the cavity has external losses, e. g. due to a connection to a waveguide system, the voltage induced in steady state by the beam is decreased to $U = R_{\mathrm{l}} I_{\mathrm{beam}} l$. R_{l} is the loaded impedance that is formed from the internal shunt impedance R_{sh} of the cavity together with an external impedance R_{ext} that accounts for external losses, $R_{\mathrm{l}}^{-1} = R_{\mathrm{sh}}^{-1} + R_{\mathrm{ext}}^{-1}$.

Figure 3.6 shows an LCR-cirquit that is connected to an external system with a transmission line of the characteristic impedance Z_0. The coupling strength is represented by a transformator with a winding ratio of $1 : n$. The impedance of the left hand side is transformed into the right

hand side into R_{ext}. The total resistance seen by the LCR cirquit is $R_{\mathrm{l}} = (1/R_{\mathrm{sh}} + 1/R_{\mathrm{ext}})^{-1}$. With $\omega_0 = 1/\sqrt{LC}$ and $\omega_{1/2} = 1/(2R_{\mathrm{l}}C)$ the results of the previous sections on bandpass dynamics can be applied with

$$x := U \quad \text{and} \quad u := R_{\mathrm{l}} I_{\mathrm{beam}}. \tag{3.23}$$

All results on steady-state values and transient behavior obtained in the previous sections can be applied on the beam induced field after replacing x and u.

At this point it is convenient to introduce a quantity that characterizes the losses in a LCR cirquit. The loaded quality factor Q_{l} is defined as the ratio of center frequency of a resonator and its full bandwidth.

$$Q_{\mathrm{l}} = \frac{\omega_0}{2\omega_{1/2}} \tag{3.24}$$

An alternative but equivalent definition for the quality factor is given in the context of power consumtion in section 3.7. In a general parallel LCR circuit, the quality factor can be calculated as $Q = R\sqrt{C/L}$, therefore the ratio R/Q is a constant independent of external losses. With this, a normalized shunt impedance that is independent of external coupling is defined and abbreviated as (r/Q). It is normalized per meter and is for TESLA cavities

$$\left(\frac{r}{Q}\right) = 1024\frac{\Omega}{\mathrm{m}}. \tag{3.25}$$

From

$$R_{\mathrm{l}} = \left(\frac{r}{Q}\right) Q_{\mathrm{l}} \tag{3.26}$$

for the loaded shunt impedance follows with the definition of Q_{l}

$$R_{\mathrm{l}} = \frac{1}{2}\left(\frac{r}{Q}\right)\frac{\omega_0}{\omega_{1/2}}. \tag{3.27}$$

In other words, the loaded shunt impedance is proportional to the inverse half-bandwidth $\omega_{1/2}^{-1}$. With that, also the amplitudes of the beam induced fields are proportional to $\omega_{1/2}^{-1}$. The beam induced field in steady state is larger for a narrow banded system than for a broad banded system.

The amplitude step-response of a cavity mode with normalized in- and outputs has been calculated to (refer to equation 3.15)

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}|x_{\mathrm{step}}|\right]_{t=0} = \omega_{1/2}. \tag{3.28}$$

In case of excitation by a beam current one has $u = R_l I_{\mathrm{beam}} = (r/Q)\omega_0/(2\omega_{1/2})I_{\mathrm{beam}}$. The chain rule of differentiation yields

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}|U_{\mathrm{step}}|\right]_{t=0} = \frac{1}{2}\left(\frac{r}{Q}\right)\omega_0 I_{\mathrm{beam}}. \tag{3.29}$$

This means that the initial step response does not depend on the bandwidth of the system.

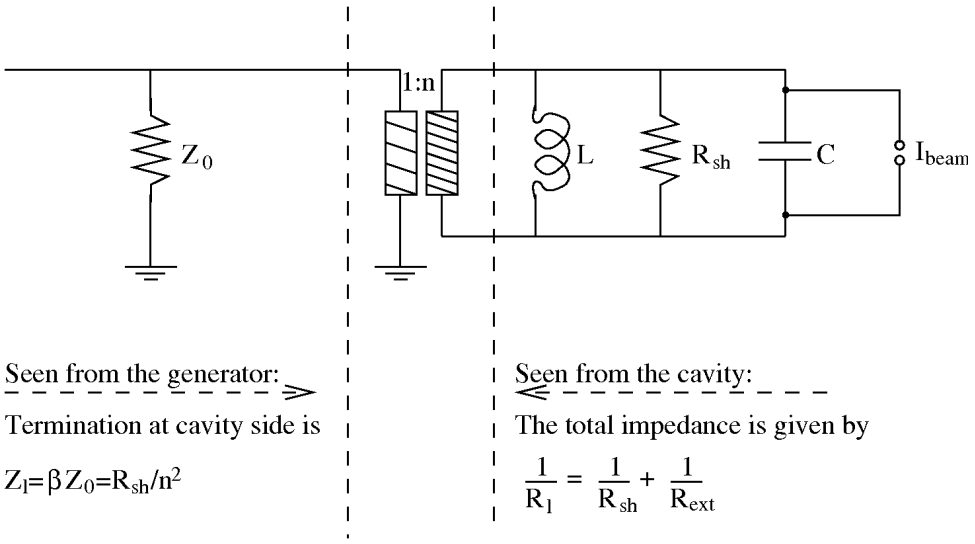

Figure 3.6.: Simple LCR-circuit with a current source and a connection to an external transmission line.

Figure 3.7 shows the transient response on current-like excitements. The amplitude is scaled such that in steady-state $|x| = 1$ for an arbitrary reference half-bandwidth $\omega_{1/2,\mathrm{r}}$. The transient behavior is identical for $t \ll \omega_{1/2,\mathrm{r}}^{-1}$ for all depicted situations. This will be important for the beam-based vectorsum calibration method discussed in section 4.3.3.

3.5. Forward and Reflected Waves

Before discussing the coupling of the RF amplifier to the cavity field, the forward and reflected waves are introduced. At FLASH, one RF amplifier (klystron) powers up to 16 and in the future even 24 cavities, for XFEL this will be even more. One klystron emits up to 10 MW that is transported by the waveguide system. A circulator protects the klystron from reflected power. This circulator has two consequences. First, the cavities see a properly terminated source and second, the klystron sees a properly terminated drain.

The reflection that occurs close to the high power couplers of individual cavities is depicted in figure 3.8. The top of the figure shows the schematic setup of a waveguide together with a high power coupler and an antenna that has a limited penetration depth. At the end of the antenna, the wave has to pass a vacuum pipe of limited diameter smaller than the wavelength and therefore shows exponential evanescence.

The thick line in the graph below the schematic setup in figure 3.8 shows the amplitude of the forward traveling wave. This thick line is at the same time the amplitude of the reflected wave due to the open end reflection in case the cavity is infinitely far away. An antinode is at the place of the open end and the forward traveling wave and the reverse traveling wave have thus the same phase at the end of the coupler. The result is a standing wave pattern on the waveguide and is

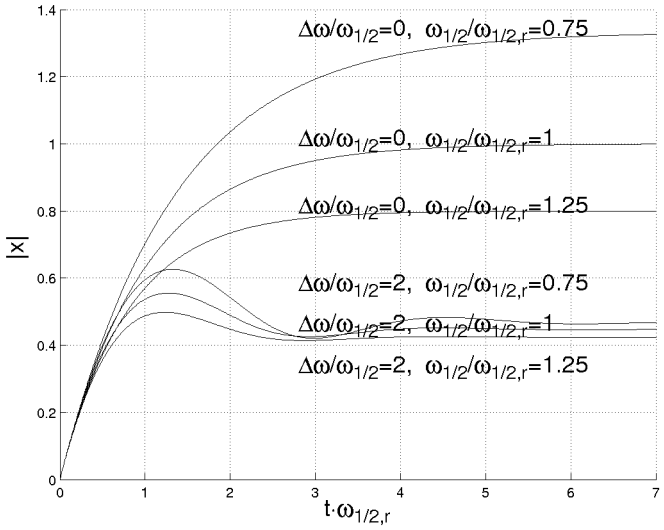

Figure 3.7.: Transient amplitude behavior for current-like excitation. Different loaded bandwidths $\omega_{1/2}$ are given as multiples of a reference bandwidth $\omega_{1/2,\mathrm{r}}$. The slope in the beginning is identical for all curves. (See also figure 3.2.)

depicted as a dashed line in the upper graph of figure 3.8. The reverse traveling wave does not lead to further reflections on the generator side due to the circulator.

The case that the cavity is in the reach of the evanescent field of the power coupler is depicted in the lower graph of figure 3.8. After the exponential decay only a small fraction of the original amplitude excites the field in the cavity. The very high quality factor of the cavity leads on resonance to a very high field in the resonator, indicated by the interrupted arrow. The phase difference between the field of the forward wave at the end of the coupler and the cavity field is the detuning angle ψ plus 90°. ψ was defined such that it is zero on resonance while in reality the system on resonance follows an excitation with 90° phase shift. While the field of the forward wave close to the coupler excites a field in the cavity, the field in the cavity excites a field on the waveguide. Identically to the forward wave that excites the cavity field, also the excitation of the cavity field on the waveguide is subject to exponential evanescence. The phase advance however is 90° (and not $\psi + 90°$), since the waveguide is a broadbanded system. The added phase advance therefore is $\psi + 180°$ from the forward traveling wave to the reflection at the cavity. The amplitude of the reflection at the cavity, denoted as thick dotted line in the lower graph of figure 3.8, is added to the open end reflection and is traveling away from the cavity as the reflected wave.

The two waves that travel away from the cavity are from now on referred to as the open-end reflection and the cavity reflection. The superposition of the open-end reflection and the cavity reflection form the (total) reflected wave.

The ratio of the impedance of the transmission line Z_0 and the impedance of the transformed

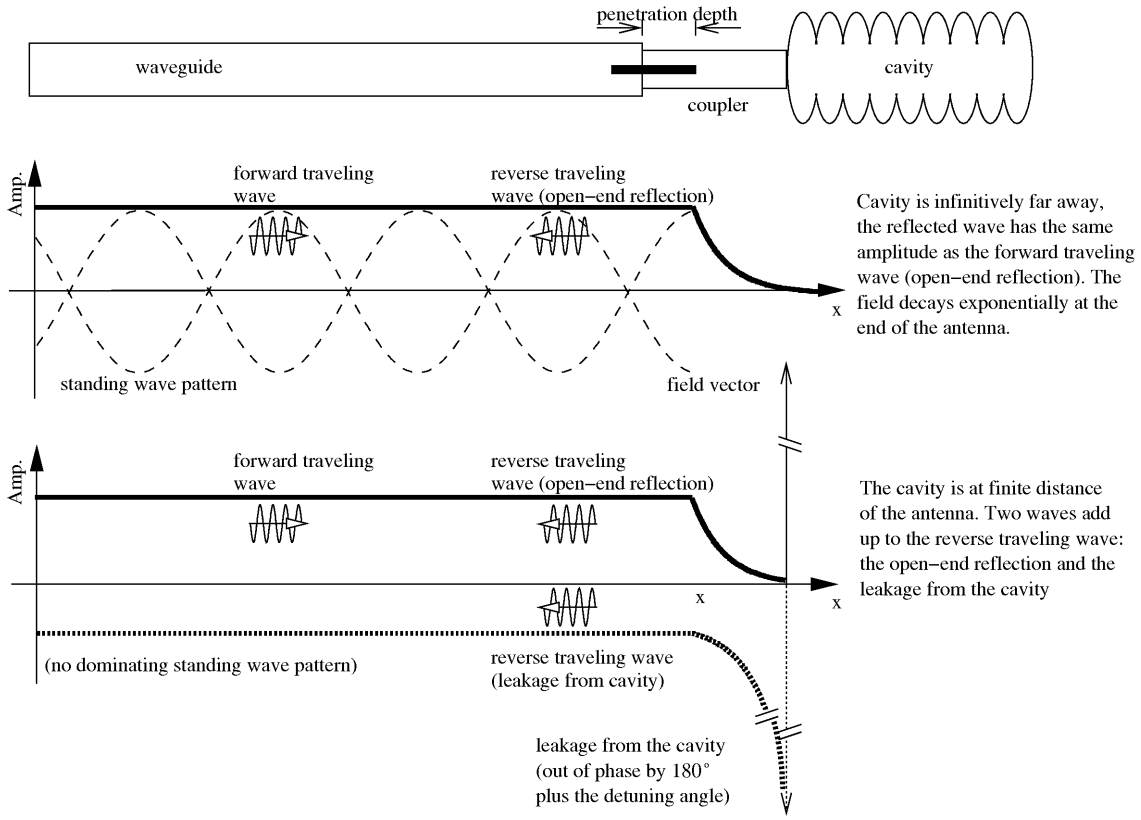

Figure 3.8.: Coupling of waves to the cavity transmitted via a waveguide system. The evanescent field is small from waveguide into the cavity and vice versa.

Z_l (see figure 3.6) cavity load defines the coupling β,

$$\beta = \frac{Z_\mathrm{l}}{Z_0}. \tag{3.30}$$

This ratio of impedances is identical to the ratio of the quality factors of the transmission line Q_ext and the (unloaded) cavity Q_0,[§]

$$\beta = \frac{Q_0}{Q_\mathrm{ext}} = Q_0\left(\frac{1}{Q_\mathrm{l}} - \frac{1}{Q_0}\right) = \frac{Q_0}{Q_\mathrm{l}} - 1 \tag{3.31}$$

The coupling can be understood figuratively. Three cases, $\beta < 1$, $\beta = 1$ and $\beta \gg 1$ are considered for a cavity that is in steady-state:

1. $\beta < 1$ (weak coupling). The cavity is far away from the antenna of the high power coupler. The total reflected wave is dominated by the open-end reflection at the cavity.
2. $\beta = 1$ (matched coupling). The distance of the cavity is such that the open-end reflection on resonance is interfering destructively with the cavity reflection. The total reflected wave is zero.
3. $\beta \gg 1$ (strong coupling). The cavity reflection is twice the amplitude of the open-end reflection but opposite in sign. The total reflected wave is of same amplitude as the amplitude of the forward traveling wave but opposite in sign.

A strongly coupled cavity that is not in steady-state but filled in a limited amount of time shall be considered. For this cavity with $\beta \gg 1$, the transient filling process will appear as if the three steady-state situations described above ($\beta < 1$, $\beta = 1$ and $\beta \gg 1$) are undergone one after the other. At the very beginning of the filling, the reflected wave will be the open end reflection since no field is present in the cavity — as in the case $\beta < 1$ above. At some point during the filling (precisley at the time $t = \ln(2)/\omega_{1/2}$), the total reflection will be zero due to destructive interference of the open-end reflection and the cavity reflection. This is similar to the case $\beta = 1$ described above. As the field is further increased towards its steady-state value, it will reach the situation described above as $\beta \gg 1$, where the total reflection is of same amplitude but opposite phase as the forward traveling wave.

3.6. RF-Amplifier Coupling

The field at the high power coupler is a superposition of a forward and a reverse traveling wave with the complex field strengths at the high power coupler U_for and U_ref. The total field strength on the transmission line at the high power coupler is $U_\mathrm{trans} = U_\mathrm{for} + U_\mathrm{ref}$. The plus sign is convention. For $\beta \gg 1$, $|U_\mathrm{trans}| = 2|U_\mathrm{for}|$ in steady-state. The vectors U_for, U_ref and U_trans are

[§]In order to avoid confusion caused by the identical index 0, it shall be emphasized that Z_0 is a property of the transmission line and Q_0 is a property of the resonator. Q_ext is the quality factor of the transmission line and Q_l is the *loaded* quality factor, thus a property of the combined system built from the transmission line and the resonator.

depicted in figure 3.9. The possible values for U_{trans} lie on a circle that is determined by the bandpass properties of the cavity. Figure 3.9 shows three circles for three possible values of β. Upper and lower half of the plane correspond to positive ($\Delta\omega > 0$) and negative detuning ($\Delta\omega < 0$). Since U_{trans} is directly proportional to the field in the cavity, stabilizing the accelerating field U in the cavity can be achieved by stabilizing U_{trans}.

U_{trans} can be expressed wit the help of the reflection coefficient $\Gamma = U_{\text{ref}}/U_{\text{for}}$. This leads to $U_{\text{trans}} = (1+\Gamma)U_{\text{for}}$. With the help of the reflection coefficient in steady-state and on resonance Γ_0, the dynamics between the complex amplitude of the forward traveling wave an the total complex amplitude at the high power coupler is given by the bandpass dynamics with the substitution

$$x := U_{\text{trans}} \quad \text{and} \quad u := (1 + \Gamma_0)U_{\text{for}}. \tag{3.32}$$

The value $U'_{\text{for}} = (1 + \Gamma_0)U_{\text{for}}$ can be imagined as the fraction of the forward traveling wave that is exciting field in the cavity. The field $U_{\text{for}} - U_{\text{for}'}$ can then be imagined as being reflected at the high power coupler directly without the influence of the field in the cavity.

The field inside the cavity is given by transforming the total field on the transmission line, U_{trans} into the cavity, U, via the winding ration $1 : n$, $U = nU_{\text{trans}}$. For strongly coupled cavities, The loaded resistance R_{l} on the cavity side is dominated by the external losses, R_{ext}. The external losses are the impedance of the transmission line transformed via the winding ratio, $R_{\text{l}} \approx R_{\text{ext}} = Z_0 n^2$. On the other hand, the loaded resistance is given by the normalized shunt impedance multiplied by the loaded quality factor, $R_{\text{l}} = (r/Q)Q_{\text{l}}$. With this, the n in $U = nU_{\text{trans}}$ can be replaced and one gets

$$U = \sqrt{\frac{1}{Z_0}\left(\frac{r}{Q}\right)Q_{\text{l}}}U_{\text{trans}}. \tag{3.33}$$

Since $Q_l \propto \omega_{1/2}^{-1}$, a difference to the beam induced current becomes visible. For a change in coupling of overcoupled cavities, the field scales with $\omega_{1/2}^{-1}$ for the beam induced fields and with $\sqrt{\omega_{1/2}}^{-1}$ for the RF amplifier induced fields.

The equation 3.33 is true in steady-state as well as in the transient case. Adding the measurements of forward and reflected waves allows to deduce U_{trans} that is proportional to the field in the cavity itself. This can be used for the control of the cavity. The problem of calibrating the correct phase and amplitude relations between the measured signals is addressed in section 3.9.1 (requirements on signal calibration) and 4.3.1 (calibration method). From now on, U_{trans} is often identified with U itself, leading to the formula

$$U = U_{\text{for}} + U_{\text{ref}}. \tag{3.34}$$

The coupling β relates to the loaded and the unloaded quality factor via

$$1 + \beta = \frac{Q_0}{Q_{\text{l}}}. \tag{3.35}$$

This is equivalent to

Equation 3.34 can be used to calculate the precise shape of the reflected power for an arbitrary

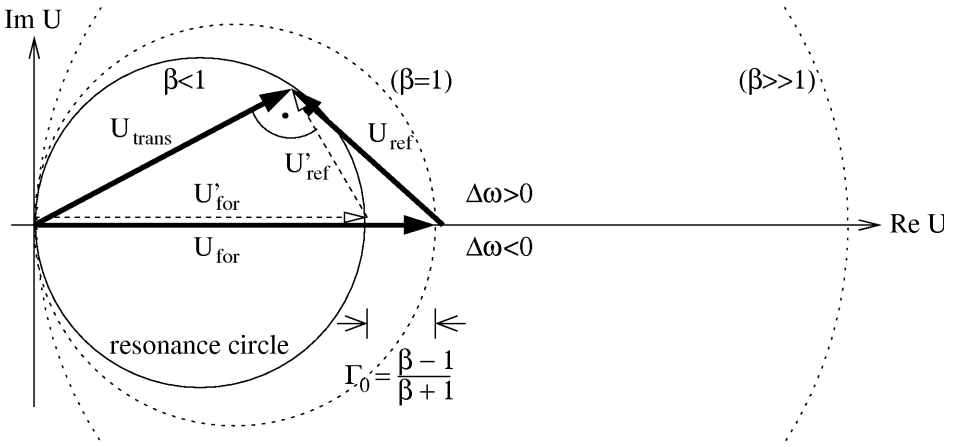

Figure 3.9.: Resonance circle. The dotted circles are the valid circles for critical coupling ($\beta = 1$) and a high overcoupling ($\beta >> 1$).

shape of the forward power. This has been done in figure 3.10 for the cases $\beta = 0.5$, $\beta = 1$ and $\beta \gg 1$. The time-axis are scaled in multiples of the characteristic time-constant $\omega_{1/2}^{-1}$. The top row shows the amplitude of the forward and reflected wave and the field amplitude for FLASH pulse shapes. The lower row shows a square excitation. The graph on the lower right shows the excitation of an overcoupled cavity. As discussed earlier, the reflected amplitude changes its sign during the filling process. During the transient process, at $t = \ln(2)/\omega_{1/2}$, the reflected amplitude is zero for an infinitively small moment.

3.7. Power Consumption

The quality factor Q is an important quantity to characterize a resonator. It is defined as

$$Q = 2\pi \frac{\text{energy stored in the resonator}}{\text{energy loss within one cycle}}. \tag{3.36}$$

In the absence of sources the cavity field and also the stored energy decay exponentially. The time-constant of the exponential decay of the energy is half the time constant of the field decay since the stored energy depends quadratically on the field. From $\mathrm{d}/\mathrm{d}tU = \omega_{1/2}U$ follows $\mathrm{d}/\mathrm{d}tU^2 = 2\omega_{1/2}U^2$. With this, the quality factor is

$$Q_\mathrm{l} = \frac{\omega_0}{2\omega_{1/2}}. \tag{3.37}$$

The index l has been introduced, since $\omega_{1/2}$ usually refers to the half-bandwidth of the loaded cavity. The quality factor of the loaded cavity is given by the coupling to the waveguide distribution system. With the quality-factor, the loaded resistance (equation 3.27) can be expressed

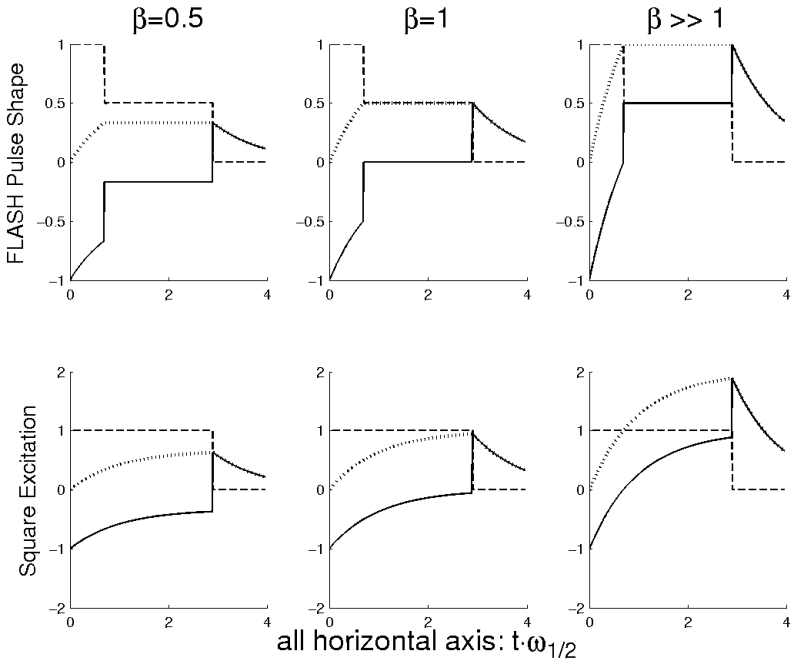

Figure 3.10.: Shape of forward (dashed line), reflected (solid line) and field amplitude (dotted line) for different forward pulse shapes and different βs. The curves of forward and reflected amplitudes are not power-waves, the same level of the forward amplitude for different couplings does not necessarily mean identical forward power (see text). The cavities are driven on resonance.

as

$$R_l = \left(\frac{r}{Q}\right) Q_l. \tag{3.38}$$

From the cavity envelope equation (3.11), the total power required to maintain a field U in the presence of a beam current I_{beam} that has a phase φ_{beam} relative to the cavity field can be calculated. x and u are then defined as

$$x := U \quad \text{and} \quad u := (1 + \Gamma_0)U_{\text{for}} + R_l I_{\text{beam}}. \tag{3.39}$$

I_{beam} is a complex number with amplitude and phase where the phase relative to the field U is called the beam-phase φ_{beam}. The absolute of U_{for} is then directly proportional to the square root of the power required from the RF amplifier. In this consideration, the bandwidth $\omega_{1/2}$ is assumed to be constant (i. e., no additional reference bandwidth $\omega_{1/2,\text{r}}$ appears in the equation). The field induced in the cavity in steady-state is the superposition of the fields induced by the RF amplifier (represented by U_{for}) and the beam current (represented by I_{beam}). Solving the steady-state solution (equation 3.12) for U_{for} yields

$$U_{\text{for}} = \frac{1}{1 + \Gamma_0}\left(\frac{\omega_{1/2} - i\Delta\omega}{\omega_{1/2}} U - R_l I_{\text{beam}}\right). \tag{3.40}$$

This expression in brackets is the difference of two complex numbers. The left summand has an angle of $-\psi$ with respect to the cavity field U and an amplitude $|U|/\cos\psi$. ψ and $\cos\psi$ are the phase rotation and amplitude reduction in steady-state that results from the resonance curve defined by the expression $\omega_{1/2}/(\omega_{1/2} + i\Delta\omega)$ in the presence of a detuning $\Delta\omega$, refer also to the equations 3.4 to 3.6. The relative phase between the two complex numbers is just the difference of their angles, $\varphi_{\text{beam}} + \psi$, and the absolute value of U_{for} is obtained by application of the law of cosines.

$$|U_{\text{for}}|^2 = \frac{1}{(1 + \Gamma_0)^2}\left(\frac{|U|^2}{\cos^2\psi} + R_l^2 I_{\text{beam}}^2 - 2\frac{|U|}{\cos\psi} R_l I_{\text{beam}} \cos(\psi + \varphi_{\text{beam}})\right) \tag{3.41}$$

This expression is further transformed by replacing

$$1/\cos^2\psi = 1 + \tan^2\psi \quad \text{and} \tag{3.42}$$

$$R_l^2 I_{\text{beam}}^2 = R_l^2 I_{\text{beam}}^2(\sin^2\varphi_{\text{beam}} + \cos^2\varphi_{\text{beam}}) \quad \text{and the identity} \tag{3.43}$$

$$\cos(\psi + \varphi_{\text{beam}}) = \cos\psi\cos\varphi_{\text{beam}} - \sin\psi\sin\varphi_{\text{beam}}, \tag{3.44}$$

which leads to

$$|U_{\text{for}}|^2 = \frac{1}{(1 + \Gamma_0)^2}\Big(|U|^2 + \tan^2\psi|U|^2 + R_l^2 I_{\text{beam}}^2 \sin^2\varphi_{\text{beam}} + R_l^2 I_{\text{beam}}^2 \cos^2\varphi_{\text{beam}} - 2|U|R_l I_{\text{beam}} \cos\varphi_{\text{beam}} + 2|U|R_l I_{\text{beam}} \sin\varphi_{\text{beam}} \tan\psi\Big). \tag{3.45}$$

And finally, by identifying binomial identities, one gets

$$|U_{\text{for}}|^2 = \frac{1}{(1+\Gamma_0)^2}\left[(|U| - R_l I_{\text{beam}}\cos\varphi_{\text{beam}})^2 + (|U|\tan\psi + R_l I_{\text{beam}}\sin\varphi_{\text{beam}})^2\right] \tag{3.46}$$

The power is obtained from the last expression by $P_{\text{for}} = |U_{\text{for}}|^2/(2R_l)$. When evaluating this expression, attention has to be paid to the beam phase: φ_{beam} was defined above to be the phase between the cavity field and the beam, which means $\varphi_{\text{beam}} = 180°$ for on-crest operation (maximum acceleration). In practice, it is common to define the beam phase such that it is zero for on-crest operation, $\varphi'_{\text{beam}} = \varphi_{\text{beam}} - 180°$.

For TESLA cavity ($(r/Q) = 1024\,\Omega/\text{m}$) driven at a field of 25 MV/m and half-bandwidth of $\omega_{1/2} = 216\,\text{Hz}$ (that is $Q_l = 3 \cdot 10^6$) without beam, the power-consumption can be approximated as

$$P_{\text{for}} \approx 25\,\text{kW}\left(1 + \left[\frac{\Delta\omega}{\omega_{1/2}}\right]^2\right). \tag{3.47}$$

The reflection coefficient in steady state for zero detuning Γ_0 was set to 1 due to the fact that the TESLA cavities are highly overcoupled ($\beta \gg 1$).

3.8. Discrete Cavity Model and Digital Filters

Z-Transformation

The concept of digital filters turns out to be very efficient for the implementation of several LLRF algorithms. Digital filters are closely connected to Z-transformations. If a function $u(t)$ is known only for discrete arguments $t_n = nT$, T being the time between subsequent samples of e. g. an analog to digital converter, one defines the Z-transform of that function as the infinite row

$$U(z) = \mathscr{Z}\{u(t)\} = \sum_{j=0}^{\infty} u_j \left(\frac{1}{z}\right)^j, \tag{3.48}$$

where $u_k = u(kT)$. Obviously, a translation by jT in time-domain is equivalent to a multiplication by z^{-j} in Z-domain.

If the sampled time-domain data of two signals $u(t)$ and $x(t)$ are connected via the equation

$$\sum_{j=0}^{n} a_j x_{n-j+k} = \sum_{j=0}^{m} b_j u_{m-j+k} \quad \text{for all } k \tag{3.49}$$

one can conclude for the corresponding Z-transforms $U(z)$ and $X(z)$

$$\sum_{j=0}^{n} a_j z^{-j} X(z) = \sum_{j=0}^{m} b_j z^{-j} U(z) \quad \text{or} \quad X(z) = H(z)U(z) \quad \text{with} \quad H(z) = \frac{\sum_{j=0}^{m} b_j z^{-j}}{\sum_{j=0}^{n} a_j z^{-j}}. \tag{3.50}$$

The identity between the time-domain equation which is valid "for all k" and the Z-domain equation which states equality between polynomials becomes clear if one considers each polynomial coefficient individually. The transfer-function $H(z)$ defined by a_k and b_k can be implemented in digital devices very efficiently. The only operations required are multiplications and additions.

For example, a discrete approximation of a transfer function that maps u to its derivative,

$$u(t) \mapsto x(t) = \dot{u}(t) \quad \text{(continuous)} \tag{3.51}$$

is given by

$$u_k \mapsto x_k = \frac{u_k - u_{k-1}}{T} \quad \text{(discrete)}, \tag{3.52}$$

which means $a = [0, T]$ and $b = [1, -1]$.

Digital Filter Design

Several patterns have been developed to design transfer functions $H(z)$. As an example, the step-invariant design shall be demonstrated for a low-pass of the corner frequency ω_{lp}. At the corner frequency ω_{lp}, the signal strength is reduced by -3 dB. For frequencies beyond the corner frequency, the signal strenght is attenuated by -10 dB per decade. A low-pass is described by the differential equation

$$\dot{x} + \omega_{\text{lp}} x = \omega_{\text{lp}} u. \tag{3.53}$$

The transfer function $H(z)$ of this low-pass the quotient of the Z-transformation of the step-response $(1 - e^{-t\omega_{\text{lp}}})$ and the Z-transformation of the step itself ($\Theta(t)$, where $\Theta(t) = 1$ for $t > 0$ otherwise $\Theta(t) = 0$). The Z-transformation of the step can be calculated from the fractional representation of an infinite geometric series.

$$U(z) = \mathscr{Z}\{\Theta(t)\} = \sum_{j=0}^{\infty} z^{-j} = \frac{1}{1 - z^{-1}} \tag{3.54}$$

The Z-transformation of the step-response $1 - e^{-t\omega_{\text{lp}}}$ is also obtained knowing the fractional representation of infinite geometric series.

$$X(z) = \mathscr{Z}\{1 - e^{-t\omega_{\text{lp}}}\} = \sum_{j=0}^{\infty} (1 - e^{jT\omega_{\text{lp}}}) z^{-j} = \frac{(1 - e^{T\omega_{\text{lp}}}) z^{-1}}{(1 - z^{-1})(1 - e^{T\omega_{\text{lp}}} z^{-1})} \tag{3.55}$$

Division of $X(z)$ and $U(z)$ yields the transfer function of the low-pass in the step-invariant design.

$$H(z) = \frac{X(z)}{U(z)} = \frac{(1 - e^{T\omega_{\text{lp}}}) z^{-1}}{1 - e^{T\omega_{\text{lp}}} z^{-1}} \tag{3.56}$$

	$H(z)$	b	a
Impuls invariant	$\frac{T\omega_{\mathrm{lp}}}{1-e^{-T\omega_{\mathrm{lp}}}z^{-1}}$	$[T\omega_{\mathrm{lp}}]$	$[1, -e^{-T\omega_{\mathrm{lp}}}]$
Step invariant	$\frac{(1-e^{-T\omega_{\mathrm{lp}}}z^{-1}}{1-e^{T\omega_{\mathrm{lp}}}z^{-1}}$	$[0, (1-e^{-T\omega_{\mathrm{lp}}})]$	$[1, -e^{-T\omega_{\mathrm{lp}}}]$
Trapezoid	$\frac{\omega_{\mathrm{lp}}+\omega_{\mathrm{lp}}z^{-1}}{2/T+\omega_{\mathrm{lp}}-(2/T-\omega_{\mathrm{lp}})z^{-1}}$	$[\omega_{\mathrm{lp}}, \omega_{\mathrm{lp}}]$	$[2/T+\omega_{\mathrm{lp}}, -2/T+\omega_{\mathrm{lp}}]$

Table 3.1.: Different digital approximations for low-pass filters with corner-frequency ω_{lp}.

The filter-coefficients therefore are $b = [0, 1 - e^{-T\omega_{\mathrm{lp}}}]$ and $a = [1, -e^{-T\omega_{\mathrm{lp}}}]$. The transformation rule for the time-domain sequence can be stated according to equation 3.49:

$$u_1..u_N \mapsto x_1..x_N : \; x_{k+1} = e^{-T\omega_{\mathrm{lp}}}x_k + (1 - e^{-T\omega_{\mathrm{lp}}})u_k \tag{3.57}$$

With this formula, the heuristic filter used currently in the DSP of the FLASH LLRF system for smoothing of the vector-sum signal (see [15], page 63),

$$u_1..u_N \mapsto x_1..x_N : \; x_{k+1} = (1 - \frac{1}{N})x_k + \frac{1}{N}u_k \tag{3.58}$$

can be identified as a low-pass with corner-frequency $\omega_{\mathrm{lp}} = -\ln(1 - 1/N)/T$ which is 20 kHz for $N = 4$.

Other design patterns are the impuls-invariant design and the trapezoid (Tustin) approximation. Since the simple low-pass is of high importance for the cavity dynamics, a summary of the different digital low-passes is given in table 3.1. The trapezoid-approximation has simple coefficients and is therefore interesting for application where real-time devices calculate the filter coefficients on-line.

Discrete State-Space Model for Cavity Modes

The presented low-pass approximations can now be used for modeling the cavity dynamics. The dynamics of the envelope approximation for resonant modes (equation 3.11) is very similar to a simple low-pass (equation 3.53). The step-invariant design applied on the envelope equation 3.11 yields the filter-coefficients

$$b = \frac{\omega_{1/2}}{\omega_{1/2} - i\Delta\omega}[0, 1 - e^{-T(\omega_{1/2}-i\Delta\omega)}] \;\text{ and }\; a = [1, -e^{T(\omega_{1/2}-i\Delta\omega)}]. \tag{3.59}$$

In order to transform any of the low-pass filter coefficients of table 3.1 into coefficients for the envelope equation, one can proceed according to this rule: replace every occurrence of ω_{lp} by $\omega_{1/2} - i\Delta\omega$, then multiply the b-coefficient by the factor $\omega_{1/2}/(\omega_{1/2} - i\Delta\omega)$.[¶]

[¶]This rule is identical to replacing all occurrences of ω_{lp} in equation 3.53 by $\omega_{1/2} - i\Delta\omega$ except the one on the right hand side of the equal sign in front of the u, which is just replaced by $\omega_{1/2}$. The resulting is then equation 3.11.

The usage of complex filter coefficients is unorthodox but turns out to be very useful.[‖] The transformation rule for the discrete model of the cavity envelope equation can now be stated as

$$u_1..u_N \mapsto x_1..x_N : \; x_{k+1} = e^{-T(\omega_{1/2}-i\Delta\omega)}x_k + \frac{\omega_{1/2}}{\omega_{1/2}-i\Delta\omega}(1 - e^{T(\omega_{1/2}-i\Delta\omega)})u_k \tag{3.60}$$

This state-space equation is of practical meaning. It is the discrete state-space model of the cavity envelope equation. Since it connects only u_ks and x_ks of subsequent time-steps, it allows to use different detunings $\Delta\omega_k$ for all steps k. The formula can be used for modeling a time varying detuning in case the change of the detuning is small compared to the half-bandwidth during a sample period T.

The formula for the discrete state-space model 3.60 was obtained using the methods of digital filter design. Since digital filter design is not unique, the state-space representation is not unique, too. An alternative way to obtain a state-space representation can be found in [18] page 386. Assuming that the input function $u(t)$ is constant between subsequent sample points, a state-space representation identical to formula 3.60 is obtained by stepwise integration over the sample period T.

[‖]It is even supported by numerical platforms like Matlab®.

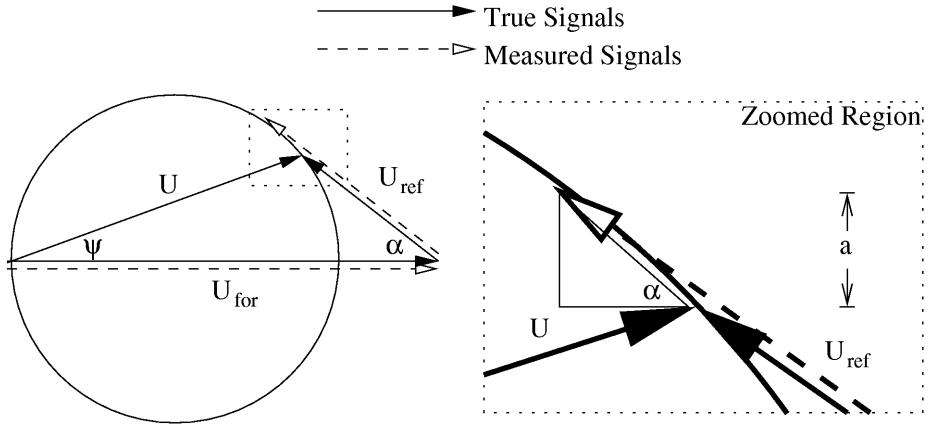

Figure 3.11.: Vector-constellation in a complex plane for approximating the errors on the constructed cavity field. The dashed arrows represent the measurements made by the controller, e. g. a DSP. The solid arrows are the true values. In this simplified example, the measured forward vector coincides with the true forward vector, while the measured reflected vector has an error in amplitude compared to the true reflected vector.

3.9. Requirements and Limits on Signal Calibration

3.9.1. Probeless Single Cavity Calibration

In the absence of an antenna, a quantity that is proportional to the field in the cavity can be calculated from the forward and the reflected power signals. Although the following consideration is valid for superconducting cavities as well as for normal conducting cavities, it is of special interest for the normal conducting electron gun at FLASH. Presently, the normal conducting, $1\frac{1}{2}$-cell resonator of the electron source of FLASH does not have a field probe. The field control has to derive the control variable as a sum of two vectors. The following consideration on the limits of the field estimation is also true for superconducting cavities, while the developed calibration-method in section 4.3.1 is exclusively applicable on cavities that are driven in pulsed mode with pulse length that are large compared to the typical time-constant of the resonator.

In section 3.6, this sum of the complex amplitudes of the forward and the reflected waves was identified with the field in the cavity itself,

$$U = U_{\text{for}} + U_{\text{ref}}. \tag{3.61}$$

U_{for} and U_{ref} are the complex amplitudes of the forward and the reflected wave measured in units of the cavity field, U. In practice, constant calibration errors are made when determining the observables, $(\Delta A/A)$ in amplitude and $\Delta\varphi$ in phase. Figure 3.11 demonstrates how a constant calibration error in amplitude of the reflected field can lead to a time-varying error of the estimated cavity field phase. It shows the complex plane with solid arrows representing the true

values as well as dashed arrows representing measurements by the controller, e. g. a digital signal processor (DSP). It is important to realize that for LLRF control, constant errors on the determined cavity field U are irrelevant. Constant errors on the observables U_{for} and U_{ref}, however, can lead to a time-varying error on the field U in the presence of a time-varying detuning.

In order to estimate the time-varying error of the cavity field phase, a zoomed region is presented in figure 3.11. The orthogonal projection of the error of the sum of the vectors is between the error of the reflected wave $(\Delta A/A) \cdot U_{\text{ref}}$ and a (see figure). The term orthogonal in "orthogonal projection" has to be understood as orthogonal with respect to the direction given by the cavity field U. For nearly matched coupling and small detuning, the orthogonal projection is approximated by a. The error in phase is, for small angles ψ, just the ratio between the length of a and the field amplitude, $|U|$. The law of sines yields $|U_{\text{ref}}| \sin\alpha = |U| \sin\psi$. Therefore, the phase error of the cavity-field is

$$\begin{aligned} \Delta\varphi'_{\text{cavity}} &= \frac{a}{|U|} \\ &= \frac{\left(\frac{\Delta A}{A}\right)|U_{\text{ref}}| \sin\alpha}{|U|} \\ &= \left(\frac{\Delta A}{A}\right)\sin\psi \end{aligned} \tag{3.62}$$

where $(\Delta A/A)$ is the amplitude calibration error of the reflected wave. In reality, both forward and reflected waves will have amplitude calibration errors. Provided the errors are of same size, distributed gaussian and uncorrelated, this can be included in the presented calculation by just multiplying with a factor $\sqrt{2}$. Its interpretation is that the measured fields are transformed into a coordinate system where the forward field has no calibration error. The calibration error of the forward wave is described by that transformation and increases the error of the reflected wave signal.

$$\Delta\varphi_{\text{cavity}} = \sqrt{2}\left(\frac{\Delta A}{A}\right)\sin\psi \tag{3.63}$$

Similarly, an approximation of the contribution of a constant phase calibration error $\Delta\varphi$ on the measured cavity field amplitude can be found,

$$\left(\frac{\Delta A}{A}\right)_{\text{cavity}} = \sqrt{2}\Delta\varphi \sin\psi. \tag{3.64}$$

A phase calibration error to a phase error and from amplitude calibration to amplitude error is of second order and is neglected here. Also, the change in cavity field amplitude due to detuning in connection with a change of the detuning angle ψ is of second order and therefore neglected.

A typical value for the detuning over a pulse is $\psi = 3°$. A calibration error of 1% in amplitude and 1° in phase will allow to measure the field up to 0.1 % in amplitude and 0.04° in phase. This calibration requirement is in reach of the methods presented in section 4.3.1.

The electron gun is subject to thermomechanical stress. More than one MW of power is dis-

sipated in the walls of the cavity during an RF pulse of a few ten microseconds. The change in geometry due to heating might influence the coupling β of the high power coupler. Beside the required calibration of the measured forward and reflected waves it is thus of interest which change in β due to for example thermomechanical stress is tolerable. This is answered by considering first the change in field and second the change in the reconstructed field caused by a change in β. The change in β is assumed to be small around the matched case, $\beta = 1 + \Delta\beta$ and $\Delta\beta \ll 1$. A change in β leads to a change in the Q_l, the loaded quality factor. A field inside the resonator induced by an RF amplifier scales with $\sqrt{Q_\mathrm{l}}$. $\sqrt{Q_\mathrm{l}}$ can be expressed with β using equation 3.31. For the relative change of $\sqrt{Q_\mathrm{l}}$ one yields

$$\begin{aligned}
\frac{1}{\sqrt{Q_\mathrm{l}}}\frac{\mathrm{d}}{\mathrm{d}\beta}\sqrt{Q_\mathrm{l}} &= \sqrt{\frac{1+\beta}{Q_0}}\frac{\mathrm{d}}{\mathrm{d}\beta}\sqrt{\frac{Q_0}{1+\beta}} \\
&= -\frac{\sqrt{1+\beta}}{\sqrt{Q_0}}\frac{\sqrt{Q_0}}{2\sqrt[3]{1+\beta}} \\
&= -\frac{1}{2(1+\beta)}
\end{aligned} \tag{3.65}$$

For $\beta = 1 + \Delta\beta$, the relative change of the field inside the resonator is $-\Delta\beta/4$. On the other hand, the reconstructed field U' changes, too. This can be traced back to the reflection coefficient in steady-state on resonance Γ_0 via $U' = U_\mathrm{for} + U_\mathrm{ref} = (1 + \Gamma_0)U_\mathrm{for}$.

$$\Gamma_0 = \frac{\beta - 1}{\beta + 1} \quad \Leftrightarrow \quad 1 + \Gamma_0 = \frac{2\beta}{\beta + 1} \tag{3.66}$$

With that, the relative change of the measurement can be calculated.

$$\begin{aligned}
\frac{1}{1+\Gamma_0}\frac{\mathrm{d}}{\mathrm{d}\beta}1 + \Gamma_0 &= \frac{\beta+1}{2\beta}\frac{\mathrm{d}}{\mathrm{d}\beta}\frac{2\beta}{\beta+1} \\
&= \frac{\beta+1}{2\beta}\frac{2(\beta+1) - 2\beta}{(\beta+1)^2} \\
&= \frac{1}{2\beta(\beta+1)}
\end{aligned} \tag{3.67}$$

For $\beta = 1$, the relative change of the reconstructed field is therefore $\Delta\beta/4$. This change in the reconstructed field is just opposite in sign compared to the change in the field. The total error made is therefore $\Delta\beta/2$! If one requires an error on the reconstructed field of less than 0.1 % in amplitude at a matched resonator ($\beta = 1$), the coupling β may not change by more than 0.05 due to thermal stress. In section 4.3.1 it is shown by measurement that there is no indication that the coupling exceeds this value.

Since the equations 3.65 and 3.67 are a little bit hard to digest, this section shall be closed with the following plausibility considerations.

- If β is increased, the coupling is stronger and the field in the cavity is decreased as in

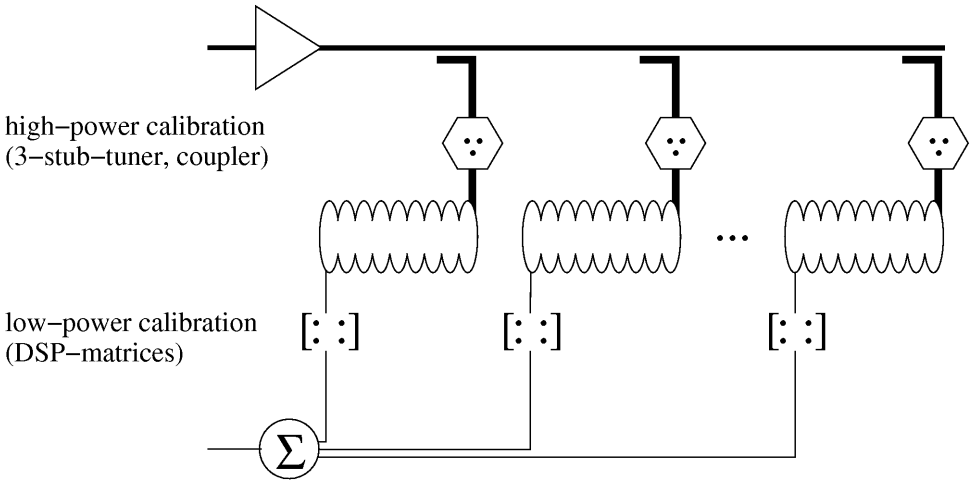

Figure 3.12.: The vector-sum calibration consists of two aspects: on the high power side, the distribution system has to be adjusted in a way that the incident amplitudes and phases of all cavities are identical. (However, there might be situations in which it is preferable that the amplitudes differ from each other.) On the low power side, the calibration matrices shall reflect all attenuations and phase rotation on the sensor side.

equation 3.65. More power is reflected if the coupling is increased and less power is available for the field.

- If β is increased, the reconstructed field is larger. This can be explained at figure 3.9, where the radius of the circles is increasing for stronger coupling. In steady-state and on resonance, the reconstructed field is just the diameter of the resonance circle. For $\beta > 1$, the vector of the reflected wave points in the same direction as the vector of the forward wave.
- For $\beta \gg 1$, the increase in the reconstructed field is much less than the decrease of the true cavity field. For high β, one already is very close to a total reflection at the power coupler. However, increasing β can still increase the field inside the cavity. The right hand side of equation 3.65 is of order -1 in β while the right hand side of 3.67 is of order -2 in β.

3.9.2. Vector-Sum Calibration

The analytical approach will lead to simple formulas that allow to calculate the requirements for the single cavity calibration from the aimed vector-sum precision. It is similar to the chapter about the probe-less cavity calibration and is followed by a subsection about a numerical approach covering more details that refine the requirements.

The reason for controlling vector-sums rather than single cavities is a cost argument. One very high-power klystron is significantly cheaper than several smaller ones. Also, vector-sum control requires only one LLRF control system. Vector-sum calibration has two sides which need to be

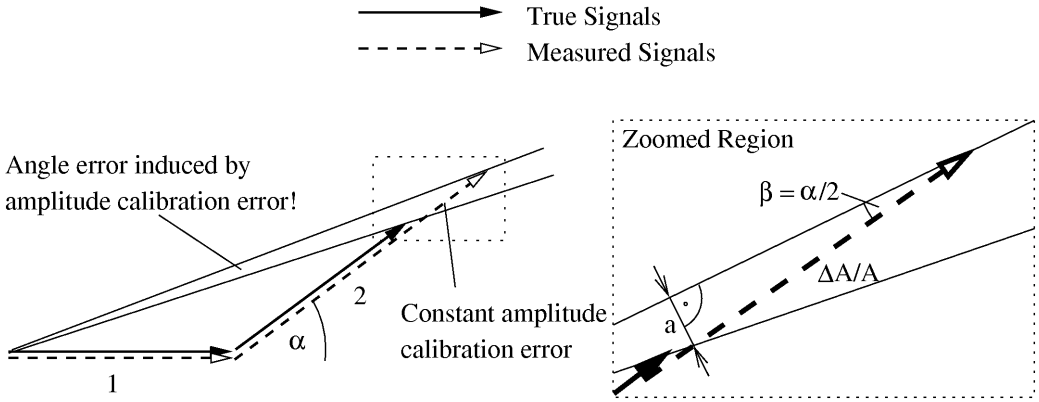

Figure 3.13.: Only two vectors contribute to a simple vectorsum (1 and 2). Only the second vector has a small calibration error in amplitude. The amplitude of the first vector as well as both phases have no calibration error. Still, a phase error is visible on the measured vector-sum. The variable a is the orthogonal projection of the error.

distinguished from each other, as illustrated in figure 3.12. On the high power side, the three-stub tuners and high-power couplers need to be adjusted such that the incident waves have identical amplitudes and phases for an optimal efficieny.** An instruction on three-stub tuning can be found in [19] and is not discussed here. The low-power calibration has to assure that the vector-sum calculated by the DSP is identical to the sum of accelerating voltages actually seen by the beam.

One can state the goals of the low-power and the high-power calibration as follows:

$$\text{high-power:} \qquad U_1 = U_2 = \ldots = U_N \tag{3.68}$$

$$\text{low-power:} \qquad \sum_{j=1}^{N} U_j = \sum_{j=1}^{N} c_j U_j^* \tag{3.69}$$

U_j denotes the true accelerating voltage experienced by the beam in the jth-cavity. U_j^* is the measured field as seen by the DSP and c_j is the corresponding complex calibration factor, thus a scaling in amplitude and phase. Ideally, $c_j U_j^*$ is identical to U_j for all j. The accelerating voltage U is considered as a complex quantity with absolute value $|U|$ and phase $\angle U$.††

Analytical Approach

Before going into a detailed numerical analysis, the impact of signal calibration errors of a vector-sum from two vectors shall be estimated analytically. A simple vector-sum of two cavities (as in figure 3.13) is considered with a time-varying detuning of maximum amplitude $\Delta\omega^*$. The varying detuning leads to a varying angle α between the two vectors (see figure). The variation in amplitude due to a varying detuning is of second order (as seen for example in figure 3.1, the resonance curve of a bandpass) and remains unattended in this analytical approach. The calibration error shall be constant for an individual cavity but is distributed gaussian from cavity to cavity with relative amplitude $(\Delta A/A) = 1 - |c_j U_j^*/U_j|$ and phase $\Delta\varphi = \angle(c_j U_j/U_j)$.

With the same argument made in section 3.9.1 for the calibration of forward and reflected signals, the error from the first vector can be transferred to the second one by increasing it by a factor $\sqrt{2}$. Identically, one can assume that the first vector has no varying detuning while the variation in detuning from pulse to pulse of the second vector is increased by an additional factor of $\sqrt{2}$. In other words, the situation is transformed into a system, where the first vector is constant and has no calibration error.

A constant amplitude calibration error $(\Delta A/A)$ of a cavity can be evaluated into a phase error of the vector-sum. Similar to the considerations in section 3.9.1, an orthogonal projection a of the error of the sum of the vectors is introduced (see figure 3.13). The term orthogonal in "orthogonal projection" has to be understood as orthogonal with respect to the direction given by the vector-sum. In order to calculate an angle, a is divided by the length of the vector-sum which is about twice the length of a single vector. This introduces an additional factor $1/2$.

$$\Delta\varphi_{\text{2-vector-sum}} = \frac{1}{2} \cdot \sqrt{2} \cdot \sqrt{2} \cdot \frac{\psi^*}{2} \cdot \left(\frac{\Delta A}{A}\right) = \frac{\psi^*}{2} \cdot \left(\frac{\Delta A}{A}\right). \tag{3.70}$$

The angle ψ^* is the maximum angle by which a vector is changing under the influence of microphonics. Since a coordinate system is considered in which the first vector is fixed, the angle α in figure 3.13 is approximated by $\sqrt{2}\psi^*$. The orthogonal projection a is in small angle approximation given by $(\Delta A/A) \cdot \alpha/2$. For a pulsed system with a pulse length of the order of the time-constant of the resonator, the worst case will be the steady-state detuning angle. So one can approximate here $\psi^* \approx \tan\psi^* = \Delta\omega^*/\omega_{1/2}$.

Figure 3.14 shows a simple example were the estimation of a sum of two vectors has a time-varying error in amplitude that depends on the relative angle between them. Identical to the previous consideration, a constant phase calibration error $\Delta\varphi$ can contribute to an amplitude error of the vector-sum. It can be approximated according to figure 3.14 as

$$\left(\frac{\Delta A}{A}\right)_{\text{2-vector-sum}} = \frac{\psi^*}{2} \cdot \Delta\varphi. \tag{3.71}$$

[**] In certain situations an amplitude spread is desirable, as for example in a module directly after the particle source. Systematic reduction of the gradient in the first few cavities is called adiabatic acceleration and helps minimizing the emittance of the beam.

[††] For this consideration, the absolute of x of the general envelope equation 3.11 is interpreted as the accelerating voltage per meter length, U.

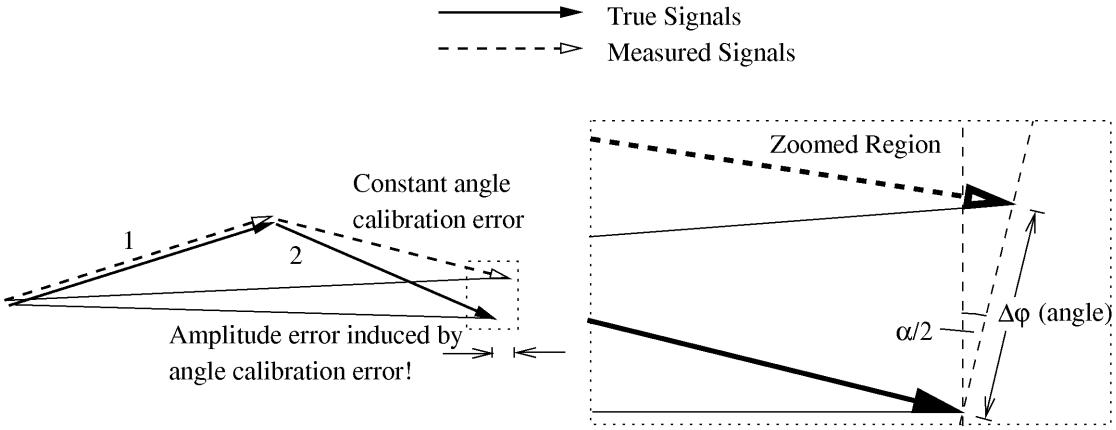

Figure 3.14.: Only two vectors contribute to a simple vectorsum (1 and 2). Only the second vector has a small calibration error in phase. The phase of the first vector as well as both amplitudes have no calibration error. Still, an amplitude error is visible on the measured vector-sum.

So far, only phase-calibration errors contributing to vector-sum amplitude errors and amplitude-calibration errors contributing to vector-sum phase errors have been considered. The other two possible combinations (amplitude-calibration errors contributing to amplitude errors on the vector-sum and phase-calibration errors contributing to phase errors on the vector-sum) are proportional to $1 - \cos\psi^*$ instead of $\sin\psi^*$, which is of second order and therefore remains unattended in the analytical approach. Figures 3.19 and 3.20 from the numerical approach justify this assumption.

As an example the error that is expected for a two-vector-sum with a maximum for the detuning induced by microphonics of $\Delta\omega^*/\omega_{1/2} = 1/10$ (about 20 Hz for a FLASH cavity) is considered. It is assumed that the calibration of individual cavities in amplitude and phase is 1 % and 0.5°. As a result from the equations 3.70 and 3.71 one gets

$$\Delta\varphi_{\text{2-vector-sum}} = 8.8 \cdot 10^{-5} = 0.028°, \quad \left(\frac{\Delta A}{A}\right)_{\text{2-vector-sum}} = 0.044\,\%.$$

A remark shall be made on the distribution of the microphonics. The previous consideration implicitly assumed all variations to be gaussian. While this is a good assumption for the spread of the calibration errors, it is not the case for microphonics. Microphonics result from mechanical resonances of the cavity. The detuning caused by microphonics is thus varied in time sinusoidally at frequencies that correspond to the mechanical resonances of the cavity. Gaussian error propagation is not adequate for a sinusoidally distributed variable. The simulations considers different kinds of distributions.

Figures 3.15 and 3.16 shall motivate the choice of $\psi^* = 1/10 \approx 5°$ since it is used in the following numerical analysis. Basis for 3.16 is a measurement that is established since June 2005 that records the measured amplitude and phase of each cavity at the time that the beam

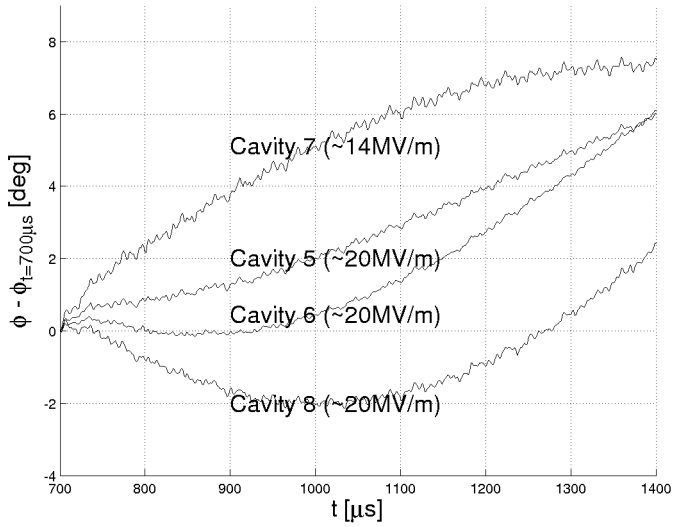

Figure 3.15.: Phases of four cavities of the first accelerating modules (ACC1) during one pulse during closed-loop operation. The phase at $t = 700\,\mu$s has been subtracted. Even if one neglects the cavity with the lowered gradient, relative phases between individual cavities of up to 5° are visible during the pulse.

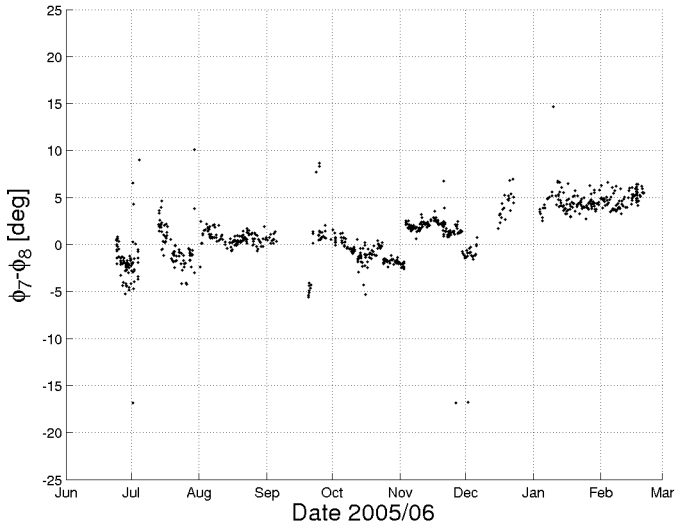

Figure 3.16.: Phase difference at beam arrival between the measured cavity phase of the seventh and the eigth cavity in the first module of the FLASH. The relative phases between these two cavities are subject to changes of the order of 3° (short-term, width of the band) respectively 5° (long-term).

passes the module (usually 700 μs after the RF pulse start).[‡‡] Drawing conclusions from this measurement is tricky since re-calibration of the waveguide distribution system happend during this time which leads to several discontinuities. It can be stated for long terms (e. g. one month) that $\psi^* \approx 1/10 \approx 5^\circ$ is a realistic assumption. For short-terms (width of the band), $\psi^* \approx 1/20 \approx 3^\circ$ seems to be a reasonable value. Figure 3.15 shows the phase of the last four cavities of the first accelerating module (ACC1) during the flattop of one pulse (i. e. between $t = 700\,\mu$s and $t = 1400\,\mu$s). The intra-pulse variation of the relative phases between individual cavities visible here is at least 5°.

Numerical Approach

A numerical analysis avoids the approximations made in the previous consideration. In addition, it in principle allows to accomodate for further effects that might influence the vector-sum measurement. The following steps are necessary to simulate the effects of constant single-cavity calibration errors on dynamic vector-sum errors. They were implemented using Matlab®.

1. Dice a set of calibration errors ($\Delta A/A$) and $\Delta\varphi$ of the individual cavities. Assume a normal distribution.
2. Dice a set of detunings, one $\Delta\omega$ for each cavity. Assume an sinusoidal distribution, since the detuning caused by microphonics follows a sine-like shape in time domain. Calculate ψ^* from the steady-state phase of the detuning of an individual cavity. Additionally, calculate the amplitude contraction due to detuning. (Amplitude distortion was not considered in the analytical approach.)
3. Calculate the vector-sum seen by the beam and seen by the DSP, compare. Record the error and go back to step 2 several times.
4. Calculate the RMS over the recorded vector-sum errors in amplitude and phase. This will be the error made for the calibration diced in 1. Note: the mean value is not of interest, since constant vector-sum errors do not need to be considered. Go back to step 1 several times.
5. Average over the RMS-deviations of the vector-sum errors in amplitude and phase and determine its RMS. The mean value of the RMS-derivations is the expected average vector-sum measurement error. Its RMS divided by the square-root of the number of diced calibration sets is the confidence interval of the result.

The input-parameters required for a simulation are: 1) The number of cavities N, 2) RMS calibration errors of amplitude and phase ($\Delta A/A$) and $\Delta\varphi$, 3) boundaries for the detuning-spread, $\Delta\omega^*$ and 4) the predetuning of the cavities.

Figures 3.17-3.22 show the results of the simulation. In all simulations, sets of input-parameters have been varied against each other. The figures have horizontal lines indicating the control requirements in amplitude and phase for the XFEL. The error bars that result from this stochastic approach are included in the diagrams. The size of the error bars is reduced by an increased set

[‡‡] Each time the main-parameters are printed into the electronic-logbook of FLASH, a snapshot from all phases and amplitudes at the time of beam injection is stored. Basis for this measurement is the "LLRF-Tool".

of simulated samples.

The figures 3.17 and 3.18 confirm the results from the analytical considerations for two-vector-sums. The analytical analysis implicitly assumed gaussian error propagation. The simulation agrees with the analytical estimations only for gaussian distributed errors.

Figures 3.19 and 3.20 confirm what has been stated in the previous section: amplitude calibration errors lead mainly to vector-sum phase errors while phase calibration errors cause mainly vector-sum amplitude errors. Additionally, figures 3.19 and 3.20 contain the analytical result scaled with $1/\sqrt{N}$, to accommodate for the different number of cavities and $2/3$, which is an empiric factor that accounts for the difference in gaussian and sinusoidal distributions.

The numerical approach justifies the assumption that errors are proportional to the maximum detuning $\Delta\omega$, the square-root of the number of cavities and the amplitude and phase calibration, $(\Delta A/A)$ and $\Delta\varphi$.

Summary

The numerical analysis suggests that a good requirement for the X-FEL target precision of

$$\left(\frac{\Delta A}{A}\right)_{N\text{-vector-sum}} = 0.01\% \quad \text{and} \quad \Delta\varphi_{N\text{-vector-sum}} = 0.01^\circ \tag{3.72}$$

requires a calibration of single cavities with an error less than

$$\left(\frac{\Delta A}{A}\right) = 1\,\% \quad \text{and} \quad \Delta\varphi = 0.5^\circ \tag{3.73}$$

These limits assume $N = 32$ cavities and a maximum microphonics of $\Delta\omega^*/\omega_{1/2} = 1/10$. By looking at the diagrams 3.19 and 3.20 a value that is less strong might appear appropriate. However, since the value depends strongly on the distribution chosen for the microphonics (see figures 3.18 and 3.17) it is indicated to chose rather conservative limits.

The analytical and the numerical method are in good agreement for $N = 2$. The numerical approach suggests a linear dependency of the calibration limits from the maximum detuning $\Delta\omega$, the square-root of the number of cavities $\sqrt{N}$ and the amplitude and phase calibration, $(\Delta A/A)$ and $\Delta\varphi$. A rough analytical estimate for the required precision is therefore given by

$$\left(\frac{\Delta A}{A}\right)_{N\text{-vector-sum}} = \frac{2}{3}\frac{1}{\sqrt{N}}\frac{\psi^*}{2}\cdot\Delta\varphi \quad \text{and} \quad \Delta\varphi_{N\text{-vector-sum}} = \frac{2}{3}\frac{1}{\sqrt{N}}\frac{\psi^*}{2}\cdot\left(\frac{\Delta A}{A}\right). \tag{3.74}$$

The factor $2/3$ is obtained empirically and is supposed to accommodate for the differences between gaussian and sinusoidally distributed microphonics. This approximation is included in figures 3.19 and 3.20 as a dashed line.

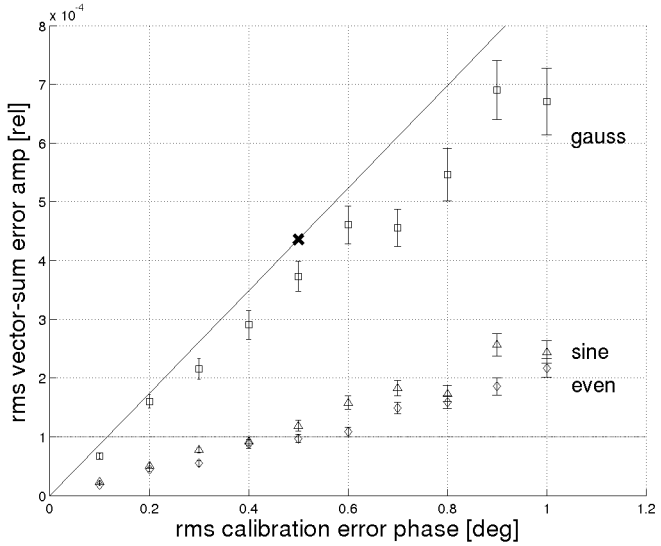

Figure 3.17.: Simulation of the relative vector-sum amplitude error: $N = 2$, $\Delta\omega^*/\omega_{1/2} = 1/10$, $(\Delta A/A)_{\mathrm{rms}} = 0.01$, $\Delta\varphi_{\mathrm{rms}} = 0..1°$ (x-axis). The types of the microphonics distributions are indicated. The solid horizontal line marks the 10^{-4}-target, the cross represents the analytical estimation made previously.

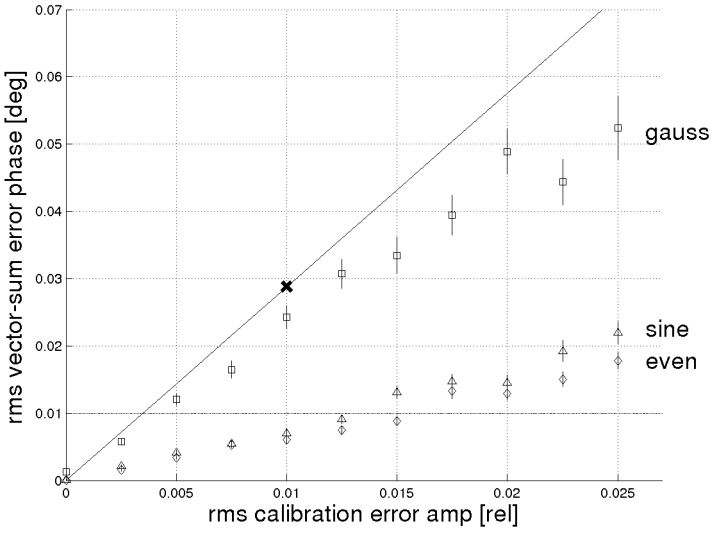

Figure 3.18.: Simulation of the vector-sum phase error: $N = 2$, $\Delta\omega^*/\omega_{1/2} = 1/10$, $\Delta\varphi_{\mathrm{rms}} = 0.5°$, $(\Delta A/A)_{\mathrm{rms}} = 0..0.025$ (x-axis). The types of the microphonics distributions are indicated. The solid horizontal line marks the $0.01°$-target, the cross represents the analytical estimation made previously.

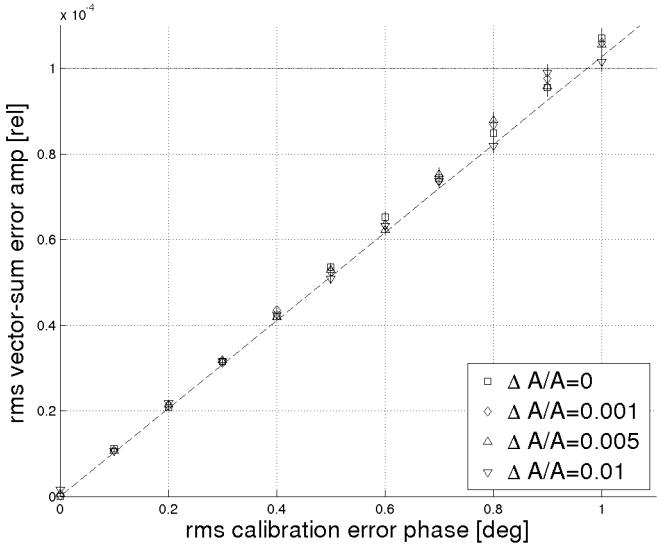

Figure 3.19.: Simulation of the relative vector-sum amplitude error: $N = 32$, $\Delta\omega^*/\omega_{1/2} = 1/10$, $\Delta\varphi_{\text{rms}} = 0..1°$ (x-axis), $(\Delta A/A)_{\text{rms}} = 0, 0.001, 0.005, 0.01$. Microphonics are distributed sinusoidally. The solid horizontal line marks the 10^{-4}-target.

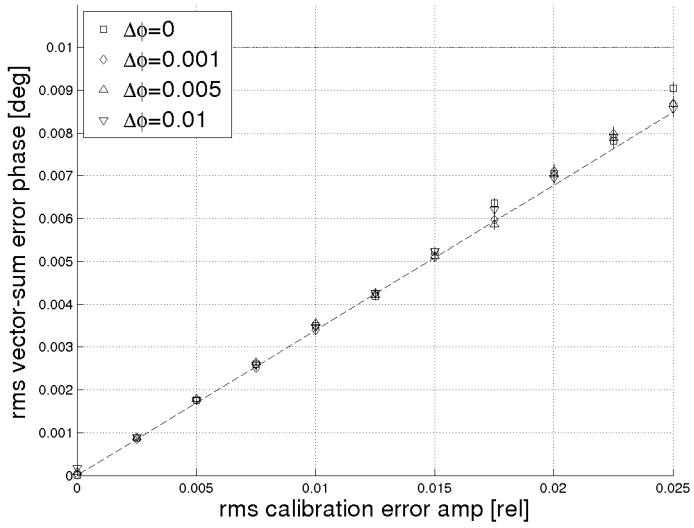

Figure 3.20.: Simulation of the vector-sum phase error: $N = 32$, $\Delta\omega^*/\omega_{1/2} = 1/10$, $(\Delta A/A)_{\text{rms}} = 0..0.025$ (x-axis), $\Delta\varphi_{\text{rms}} = 0, 0.1, 0.5, 1°$. Microphonics are distributed sinusoidally. The solid horizontal line marks the $0.01°$-target.

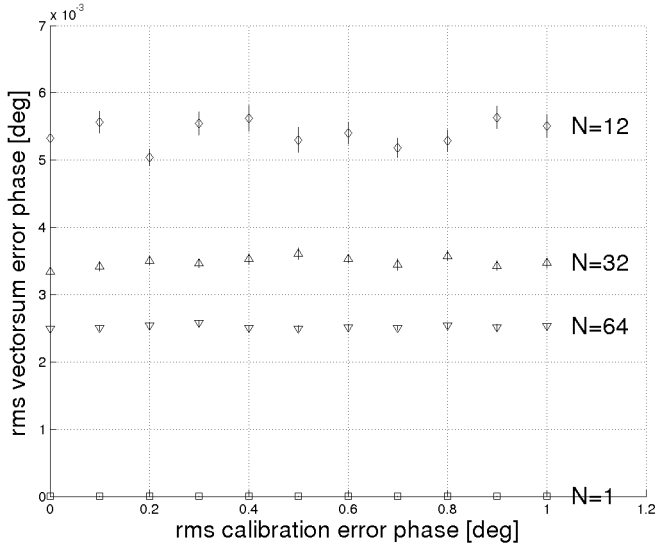

Figure 3.21.: Simulation of the vector-sum phase error: $N = 1, 12, 32, 64$, $\Delta\omega^*/\omega_{1/2} = 1/10$, $\Delta\varphi_{\text{rms}} = 0..1°$ (x-axis), $(\Delta A/A)_{\text{rms}} = 0.01$. Microphonics are distributed sinusoidally. The solid horizontal line marking the 0.01°-target is not visible.

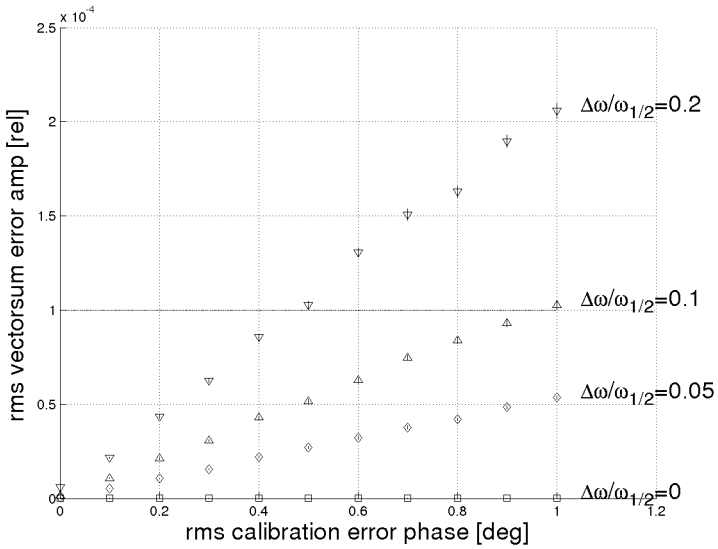

Figure 3.22.: Simulation of the relative vector-sum amplitude error: $N = 32$, $\Delta\omega^*/\omega_{1/2} = 0, 1/10, 1/20, 2/10$, $\Delta\varphi_{\text{rms}} = 0..1°$ (x-axis), $(\Delta A/A)_{\text{rms}} = 0.01$. Microphonics are distributed sinusoidally. The solid horizontal line marks the 0.01°-target.

4. Low Level Radio Frequency Control of Accelerating Cavities

4.1. FLASH LLRF Overview and Sources for Field Errors

A characterizing feature of the FLASH LLRF control scheme that also distinguishes it from other installations is the concept of vector-sum control. Figure 4.1 shows the main components of the FLASH control scheme. With the vector-sum concept, one klystron powers a set of eight to 16 resonators (for future accelerator projects, this number will be 32 or even more). This is cost efficient compared to a solution where each cavity has its own LLRF station consisting of one controller and one amplifier. At FLASH, the power is distributed from a high power klystron (typically 5 MW to 10 MW) to the cavities via a forked set of waveguides. The synchronism of the incident waves to the cavities is ensured by three-stub tuners in front of each cavity (refer to [19]) which allow an adjustment of amplitudes and phases. This calibration can be derived from the analysis of beam based transients as in section 4.3.3, where it is referred to as the high power calibration.

The center of figure 4.1 shows the reference master oscillator and the timing system which plays a central role in the LLRF system. The goal of LLRF control can be stated as locking the field inside the resonators to the reference.

Probes measure the field directly inside the cavities by coupling out a small fraction of the power. The measured fields are digitized and further processed by a digital signal processor in order to calculate the vector-sum and from that, by comparison with the setpoint, the drive that is applied to the klystron. In control theory, a system is decomposed into plant, sensor, controller and actuator.

In the following, the components of the control loop are explained. Along with that, the main sources of errors that have their origin in each component are illuminated.

Plant

The plant is the quantity under control. For a set of resonators driven by a single amplifier this is the integrated acceleration over all cavities of one RF station in amplitude and phase. Due to the factual un-measurability of the integrated acceleration, the control loop measures the individual fields and combines them to an equivalent of the integrated beam acceleration, the vector-sum. The plant of the control loop is therefore of virtual character. The physical plant, the set of superconducting cavities in the cryomodules is subject to intrinsic sources of errors, mediated by the beam current and detuning caused by microphonics or Lorentz force detuning. With that, the

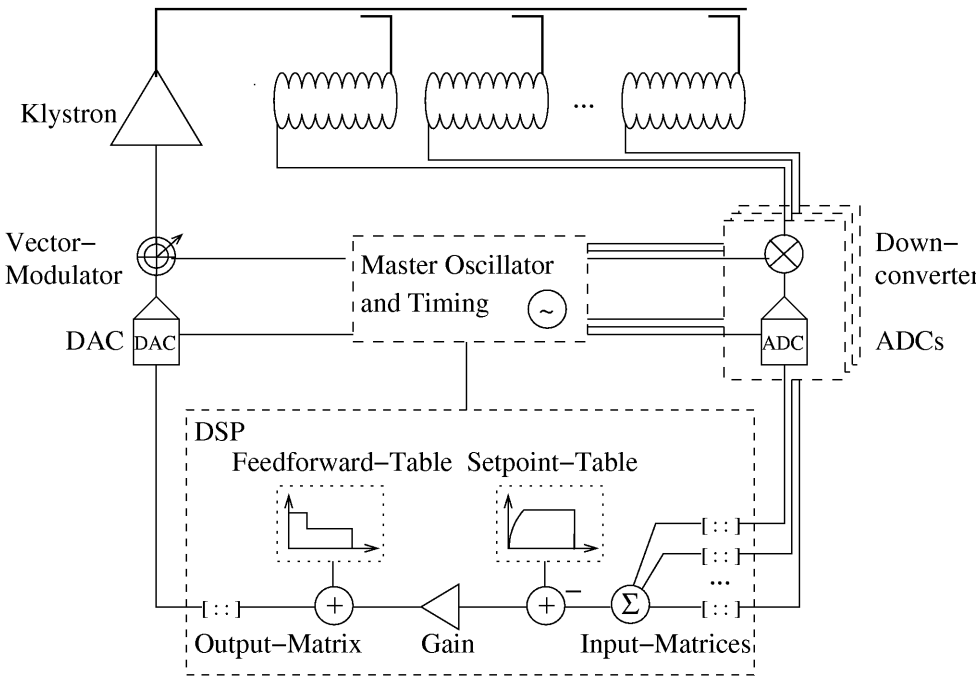

Figure 4.1.: Schematic of the LLRF system used at FLASH. The picture shows the plant consisting of klystron and cavities, a set of sensors and the simplified control algorithm itself.

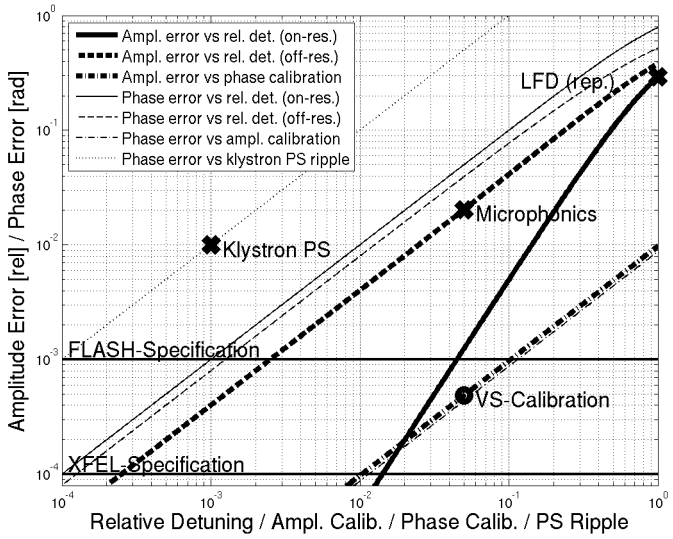

Figure 4.2.: Summary of the largest error-contributions in the LLRF control system. Each curve displays the dependency of the error-contribution and its actuating variable. The thick dashed line for example shows the dependency on the amplitude error in steady-state of the relative detuning. A few points indicate the current errors at FLASH. An ‘x’ indicates an error that can be reduced by the feedback system. An ‘o’, as for the vector-sum calibration, indicates that the feedback-system can not compensate for that error.

control theoretical plant is also subject to these effects.

In figure 4.2 the influence of detuning without feedback is shown. The thick solid line is interpreted as follows. If the x-axis is the detuning measured in multiples of the bandwidth, the thick solid line represents the error in amplitude caused by a certain detuning. The underlying physical law follows from the shape of the resonance curve in amplitude of a bandpass and is quadratic in first approximation. Therefore the slope of the thick line is two in the logarithmic representation of figure 4.2. If off resonance, the resonance curve follows a rather linear law. Therefore, the influence of microphonics on amplitude stability, represented by the thick dashed line, has slope one in the logarithmic representation.

Actuator

The actuator in the LLRF system in figure 4.1 is the path from the controller to the plant. With that definition, the border between the controller and the actuator is at the same time the border between a digital and an analog part. Part of the actuator chain is the vector modulator. This is a device that modulates an incoming RF wave in amplitude and phase according to a vector that is provided in cartesian coordinates. The vector modulator is the connecting element between the DC signals produced by the DAC and the RF signals that are amplified by the klystron. The incoming signals of the vector modulator are the reference oscillator and the vector that represents the envelope of the drive signal with respect to the reference oscillator. The envelope is generated from the DAC directly after the controller.

Amplitude errors originating from the actuator are compensated by the control loop in the same way as amplitude errors in the cavity. In small angle approximation, this is also true for errors on the phase. Large errors on the phase have a significant impact on the loop phase and can even lead to an instability. Slow drifts of the loop phase can be calculated and corrected as desribed in section 4.4.3. This section also approximates the influence of small changes in the loop phase. Figure 4.2 shows the influence of fast ripples on the power supply (PS) of a klystron on the phase of the plant.

An imperfection in the high power calibration (asynchronous phases, unequal amplitudes) is allocated at the actuator but can not be considered as a source of noise as long as the imperfections are constant.

Sensor

The path between the resonators and the controller in figure 4.1 is considered as sensor, including the antennas in the cavities and the analog to digital converter (ADC) close to the controller. Along the sensor chain, signals are transformed from the RF (1.3 GHz) to an intermediate frequency (IF) by mixers. Mathematically speaking, mixers are multiplicators in time-domain. The trigonometric identity

$$\sin(x)\sin(y) = \frac{1}{2}\cos(x-y) + \frac{1}{2}\cos(x+y) \tag{4.1}$$

reveals that a mixer transforms two sinusoidal signals at its inputs to a superposition of two sinusoidal signals at its output. The frequencies of the output signal are the sum and the difference

of the frequencies of the input signals. With a bandpass at the difference of the frequencies, the sum of the frequencies can be filtered out and the mixer becomes a downconverter.

The mixers of the downconverters at FLASH have as input the RF from the cavity and as second signal the local oscillator (LO), which (usually) is a harmonic signal. The difference in frequency between the LO and the RF is the frequency of the IF which is 250 kHz. The whole information in amplitude and phase of the RF with respect to the 1.3 GHz reference is contained in the IF with respect to a 250 kHz signal that is locked to the 1.3 GHz reference. For the RF part, the reference is the 1.3 GHz provided by the master oscillator. For the IF part, this is in principle a 250 kHz-signal that is locked to the 1.3 GHz of the master oscillator. In practice, it is the clock of the ADC that is synchronized to the 1.3 GHz. At the current FLASH LLRF installation, the ADCs are clocked with 1 MHz. Therefore, the ADC takes four samples of the IF per period. The sequence obtained in one period of the 250 kHz is $(I, Q, -I, -Q)$, which is the technical expression for the real part (I), the imaginary part (Q), the negative real part ($-I$) and the negative imaginary part ($-Q$) of the field vector.

At this point, it is worth noticing that the LO of the FLASH LLRF actually is not a continuous sine at 1.3 GHz + 250 kHz but a 1.3 GHz sine that is switched in phase by 90° every microsecond. The ADCs do not take notice of this since the switching of 90° takes place between subsequent sampling points.

In contrast to error contributions in the actuator chain, error contributions in the sensor can not be compensated by the feedback.

Controller

Figure 4.1 shows the controller together with the real-time algorithm executed on the DSP. Apart from signal calibration (application of a rotation and a scaling) it is the subtraction from the setpoint curve, multiplication with a loop gain (or application of a linear transfer-function that has proportional, differential and integral components) and addition of a feedforward-curve.

The calibration matrices can be determined by the method of beam based vector-sum calibration (introduced in section 4.3.3). Errors in the input calibration matrix can be considered as errors originating in the controller. Section 3.9.2 points out that even constant calibration errors can lead to time-varying errors over the pulse in the presence of microphonics or Lorentz force detuning. After the summing junction in figure 4.1, errors caused by calibration matrices appear as measurement errors. Therefore, the closed feedback loop can not correct for them. Figure 4.2 also contains curves that summarizes the results of section 3.9.2, where an analytic estimation for errors caused by calibration errors is given.

4.2. Parameter Optimization in Case of a Maximum-Gradient Spread

A few peculiarities occur if one has to control a number of cavities connected to one RF amplifier instead of just a single cavity. Properties of vector-sums as the plant under control shall be

discussed considering the following question as an example:

"Given that the cavities that are controlled by one RF unit differ in their maximum gradient (due to, e. g., different quench limits), what is the best total gradient that can be reached by varying only coupling, incident phase and detuning of individual cavities?"

Usually, the weakest cavity would determine the maximum allowed gradient in all cavites of one RF station. If one cavity reaches its performance limit already at 92% of the nominal gradient, the overall gradient would necessarily be reduced by 8%, leaving the remaining cavities far below their limits.

As an approach for solving this problem, two points inside an rf pulse as depicted in figure 2.3 shall be considered rather than the whole pulse. The end of the fill time is a point where the cavity is not in steady-state. At the end of the flat-top, the cavity is close to its steady-state value. The longer the flat-top, the closer the cavity is to steady-state at the end of the flat-top. At both points, the gradient distribution has to reflect the individual limits of the cavities.

If one leaves the incident phase and the detuning zero and reduces the coupling for the limited cavities, the end of the fill time would be at a reduced level and the end of the flat-top would be increased at the same time. Literally speaking, the reduced coupling narrows the bandwidth of the cavity which decelerates the filling-process but allows for higher steady-state field values.

The change in coupling can be denoted by a change in half-bandwidth $\omega_{1/2}$ with respect to a reference-bandwidth $\omega_{1/2,\mathrm{r}}$, where $\omega_{1/2,\mathrm{r}}$ determines the fill time via $t_{\mathrm{fill}} = \ln(2)/\omega_{1/2,\mathrm{r}}$. In addition to the change in bandwidth, a change in (constant) detuning can help to match the levels at the end of the fill time and the flat-top. The levels for a certain pair of $(\omega_{1/2}, \Delta\omega)$ are given by the absolute values of

$$x_{\text{fill-end}} = \frac{\omega_{1/2}}{\omega_{1/2} - i\Delta\omega} e^{-(\omega_{1/2} - i\Delta\omega)\cdot t_{\mathrm{fill}}} \cdot \sqrt{\frac{\omega_{1/2,\mathrm{r}}}{\omega_{1/2}}} \tag{4.2}$$

$$x_{\text{flat-end w/o beam}} = \frac{\omega_{1/2}}{\omega_{1/2} - i\Delta\omega} \left(e^{-(\omega_{1/2} - i\Delta\omega)\cdot (t_{\mathrm{fill}} + t_{\mathrm{flat}})} - \frac{1}{2} e^{-(\omega_{1/2} - i\Delta\omega)\cdot t_{\mathrm{flat}}} \right) \tag{4.3}$$

where t_{flat} is the duration of the flat-top. The calculations assume the feedforward shape of a perfect (undetuned) pulse: constant phase and a flat amplitude which is decreased by a factor $1/2$ after the fill time. The amplitude of the drive signal during filling is just twice the amplitude during the flat-top.

In the $(\omega_{1/2}, \Delta\omega)$-plane one can now calculate the absolute valus of 4.2 and 4.3. Two sets of equipotential lines reflecting the two equations 4.2 and 4.3 are shown in figure 4.3. In addition, the plot shows intersections of lines of equal potential, marked by circles. They mark the appropriate set $(\omega_{1/2}, \Delta\omega)$ for a cavity that is limited in its gradient.

However, even though the circles mark points of equal level (i. e. equality of the absolute values of equation 4.2 and 4.3), this does not include equality of phases. The phases at the beginning and the end of the flattop are not necessarily identical. The choice of the incident phase is therefore still open. A remaining transient behaviour therefore is expected due to phase mismatch. The remaining transient will be of importance if operated in closed loop and will

demand for a further tweaking, since the parameters were determined for open loop and optimal pulse-shape only.

An example shall therefore be considered. Four cavities have linearly distributed performance limits with values 90 %, 93.3 %, 96.7 % and 100 % of the nominal gradient. These four cavities form a vector-sum that is driven from one RF amplifier. The goal of an optimization is to come close to the theoretical average performance of 95 %. Figure 4.4 shows the pulse shape of the common drive together with the amplitudes and phases of the four cavities. The drive is optimized to ensure that the vector-sum of all cavities follows the desired setpoint-curve, i.e. exponential filling and a flat flat-top. In this figure, the effects of the transients caused by the ambiguity of the phases become visible. The levels at the end of the flat-top do not resemble the underlying linear gradient spread. The calculated gradient limit for this configuration is 87.7 %, which is even below the limit of 90 % given by the weakest cavity. A closer look at the end of the flattop reveals that the achieved spread of amplitudes seems to be larger than the expected 10 % after optimization. An obvious fix for that is to repeat the calculation above, but now requiring different spreads for the beginning and the end of the flat-top in the open loop mode. For the results of figure 4.5, the calculation above was repeated but now for a spread of 10 % at the beginning of the flat-top and a spread of 4 % at the end of the flat-top. In the presence of feedback, this spread seems to be widened close to the required 10 % at the beginning and the end of the feedback. In this configuration, a performance of 92.2 % is achieved which is between 90 % for the expected performance without parameter modification and 95 % as the average of the four performance limits. The parameters used in this example are

	I	II	III	IV
gradient limit	0.90	0.93	0.97	1.00
hald-bandwidth $\omega_{1/2}/\omega_{1/2,\mathrm{r}}$	0.68	0.78	0.88	1.00
detuning $\Delta\omega/\omega_{1/2,\mathrm{r}}$	-0.86	-0.74	-0.56	0.00
actual maximum gradient	0.88	0.91	0.95	1.00

With the presented method, models for eight or even 32 cavities can be optimized. These are the results for an eight-cavity optimization, assuming a linear spread of the maximum gradient:

peak-to-peak spread	0.10	0.20	0.40
theoretical limit	0.95	0.90	0.80
result w/o optimization	0.90	0.80	0.60
result w/ optimization	0.93	0.86	0.71

The presented method is limited to constant bandwidths and detunings, which is not always within reach.

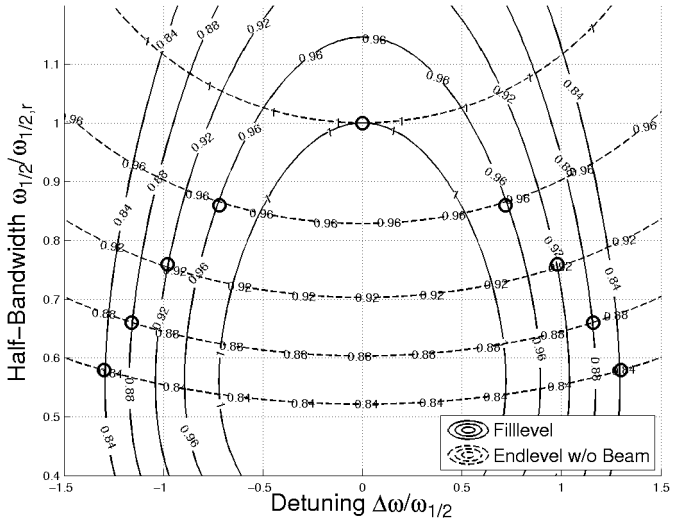

Figure 4.3.: Levels of equal potential in the $\omega_{1/2}$-$\Delta\omega$-plane for different points inside a pulse: directly after the filling (solid lines) and at the end of a 800 μs-flat-top. The circles mark intersections of lines of same potential for the filllevel and the endlevel without beam.

4.3. Signal Calibration

4.3.1. Forward and Reflected Power at the Electron Gun

Calibration Principle based on Resonance Circles

In the absence of a probe, a calibration can be done by an investigation of the resonance circle as depicted in figure 3.9. Presumption for this calibration is that the measured forward and reflected signal need to be corrected only for constant distortions,

$$\begin{aligned} U_{\text{for}} &= aU^*_{\text{for}} + bU^*_{\text{ref}} \\ U_{\text{ref}} &= cU^*_{\text{for}} + dU^*_{\text{ref}} \end{aligned} \tag{4.4}$$

with $a, b, c, d \in \mathbb{C}$ reflecting the amplitude and phase corrections that need to be applied on the measured signals (indicated by the *) in order to get the calibrated signals. In case $b, c \neq 0$, equation 4.4 would also compensate for linear crosstalk between the forward and the reflected channel.

If the non-diagonal elements b, c are neglected, the reflection coefficient for the measured signals $\Gamma^* = U^*_{\text{ref}}/U^*_{\text{for}}$ differs from the true reflection coefficient $\Gamma = U_{\text{ref}}/U_{\text{for}}$ by a complex factor d/a which is a rotation and a scaling. If measured for several different detunings in steady-state, the collection true reflection coefficients will form a circle. For any coupling β, this circle

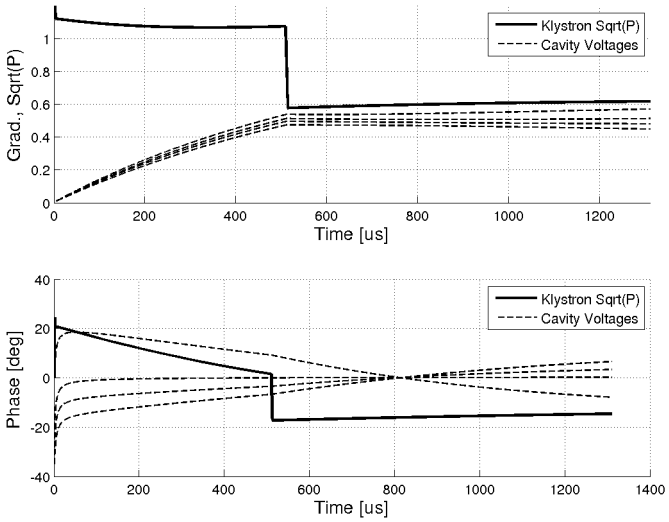

Figure 4.4.: Four vector-sums with a certain set of $(\omega_{1/2}, \Delta\omega)$. The set of constant parameters for each cavitiy is not optimal.

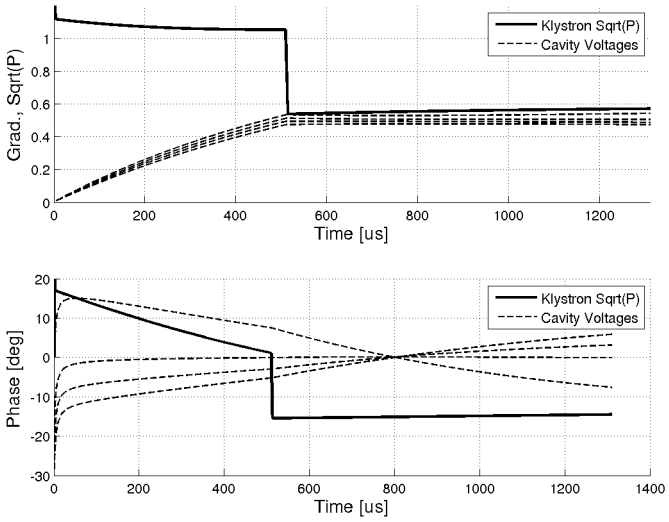

Figure 4.5.: Four vector-sum with a certain set of $(\omega_{1/2}, \Delta\omega)$. The set of constant parameters for the individual cavities is much better than what is achieved without optimization.

will include the point $(-1, 0)$ which represents the total reflection at maximum detuning (see also figure 3.1). A collection of reflection coefficients measured at different detunings will form a circle, too. However, due to constant calibration errors it will not necessarily include the point $(-1, 0)$.

The requirement that the full resoanance circle has to include the point $(-1, 0)$ can be used in order to develop a calibration procedure. Usually, it is not possible to completely detune a resonator in order to find this point. If just detuned by one half-bandwidth, the field inside the cavity already changes its phase by 45° (refer to figure 3.1). That means that already a quarter (90°) of a whole circle is covered by detuning by a half-bandwidth in each direction. This is good enough for fitting a circle with radius r and center c ($c \in \mathbb{C}$) through the measured points by minimizing

$$\chi^2 = \sum_i (|\Gamma_i^* - c| - r)^2 \tag{4.5}$$

for all measured Γ_i^*. With c and r, the uncalibrated reflection coefficient for maximum detuning can be identified as

$$\Gamma_\infty^* = (|c| + r)\frac{c}{|c|}. \tag{4.6}$$

This is the point on the fitted circle which has the largest distance from the origin. Since $\Gamma_\infty = -1$, the factor d/a is just $-1/\Gamma_\infty^*$. Once the calibration d/a is known, the reflection coefficient for zero detuning in steady-state Γ_0 can be calculated. This is the point on the circle with the least distance to the origin.

$$\begin{aligned} \Gamma_0 &= \frac{d}{a}(|c| - r)\frac{c}{|c|} \\ &= -\frac{(|c| - r)}{(|c| + r)} \end{aligned} \tag{4.7}$$

Once Γ_0 is known, the coupling β can be calculated as $\beta = (1 + \Gamma_0)/(1 - \Gamma_0)$.

Measuring the Resonance Circle for the Calibration

There are several ways to detune the cavity in order to cover a significant fraction of the resonance circle. An obvious way is the temperature-scan, where the temperature of the resonator is varied over time. The resonance frequency of the FLASH photoinjector changes with about 20 kHz/°C. 3 °C in temperature detune the electron gun by a half-bandwidth ($\sim$ 65 kHz). Figure 4.6 shows an example of such a measurement, where a point from the center of the flat-top was chosen of each pulse and plotted into the complex reflection-coefficient plane.

It is of special interest, if the amplitude- and phase-calibration or the coupling β show a dependency on variables like the average dissipated power. The measurement-series where the example of figure 4.6 originates from contains resonance circles for different average powers, where the average power was changed by changing the pulse length between 100 μs and 600 μs. The average power at the repitition-rate of 2 Hz is in the kW-regime. Figure 4.7 shows the result of this analysis for the coupling. The relative rms over different average powers is $3.3 \cdot 10^{-3}$ and

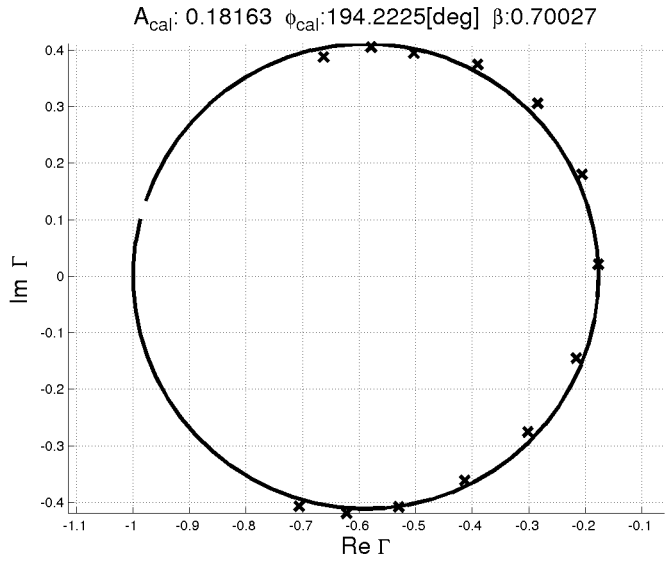

Figure 4.6.: Example of a partial resonance circle obtained by a temperature scan. The solid line is the best circle that fits through the points according to the χ^2-criteria 4.5. The temperature was varied by ±3°C around the assumed resonance temperature in steps of 0.5°C. The plot shown is already scaled by the factor $-1/\Gamma_\infty^*$ such that the point of maximal reflection lies at $(-1, 0)$. The amplitude- and phase-calibration are shown as well as the coupling β on top of the plot for this measurement. The calibration values are specific for the used setup.

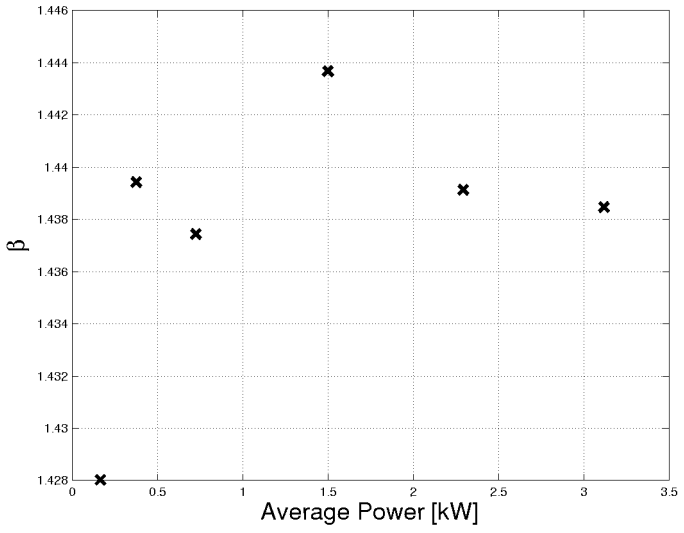

Figure 4.7.: Dependency of the coupling β on the average power. The variation of the coupling (relative rms) is $3.3 \cdot 10^{-3}$ and does not show a realiable trend. It is compatible with the requirement developed in chapter 3.9.1.

is compatible with the requirement developed in chapter 3.9.1.

The temperature-scan is a lengthy method for detuning the cavity, since temperature-control is a slow procedure by nature. Instead of changing the temperature, the resonance-frequency can be changed by changing the center frequency of the master-oscillator (figure 4.1). In a test-setup at the Photoinjector Test Stand in Zeuthen (PITZ), the master-oscillator was replaced by a frequency generator with the ability to change the center frequency periodically. By sweeping around the center-frequency by ±65 kHz with a repitition of 200 Hz, a fraction of the resonance circle within a single pulse.

Figure 4.8 shows the (time-resolved) reflection-coeficients of several pulses while the center frequency was changed periodically. The frequency generator was not locked with the timing-system, therefore it is completely arbitrary, at which frequency a pulse starts (which is not a disadvantage). In figure 4.8, lines are visible that are formed from a number of measured reflection coefficients. One line is the set of reflection coefficients measured during one pulse. Each line has a beginning, represented by a small circle in figure 4.8. There is no single circle that serves as a good fit for all reflection coefficients obtained. A systematic drift in phase seems to be visible during the pulse. Since the reflection-coefficients obtained during the pulse are now time-resolved, it is possible to fit resonance-circles in points from the tth-microsecond of a pulse. Figure 4.9 shows the evaluation of the phase-calibration inside the pulse. The drift in phase that already was suggested from figure 4.8 is confirmed, its magnitude is about 0.02° per μs. The source of the drift could not yet be determined.

A third method besides the temperature scan and the frequency variation of the reference oscillator exists that is even applicable during normal operation, [20]. It induces a detuning by digital frequency generation. It is done by applying a constant slope in phase on the RF generator and therefore requires a sufficient resolution of the DAC of the controller. A digitally generated slope on the phase requires that the generated frequency offset (which is identical to the slope on the phase) is well below the Nyquist frequency of the sampling system. At a sampling rate of 1 MHz, a frequency offset of 50 kHz appears feasible. At 1 MHz, the time distance between subsequent samples of the controller is 1 μs, therefore subsequent samples of the envelope of the drive signal need to have a difference in phase of 18° in order to produce a frequency offset of 50 kHz. The slope in phase should last at least a few times the typical time-constant of the resonator in order to be able to exclude transient effects and also provide enough data within one pulse for averaging. The typical time-constant for a resonator of half-bandwidth $\omega_{1/2} = 2\pi \cdot 65\,\text{kHz}$ is 2.4 μs, the slope should last for a few 10 μs.

Figure 4.10 shows the application of the introduced scheme at FLASH, where the controller of the photoinjector allows to have a secondary pulse after the primary pulse for the beam. The secondary pulse is independent of the gradient- or phase-setpoint and feedback-setting of the primary pulse. It has a significantly lower amplitude and does not have feedback applied. The slope of the phase can be chosen from pulse to pulse. Every slope setting represents a different detuning. For each setting of the slope of the phase, the average reflection coefficient is calculated from all samples of the secondary pulse, excluding a few samples in the beginning of the secondary pulse for avoiding transient behavior. The right side of figure 4.10 shows the measured reflection coefficients together with the fittet resonance circle and the derived parameters.

The three presented measurement methods lead to comparable results and are all suited for

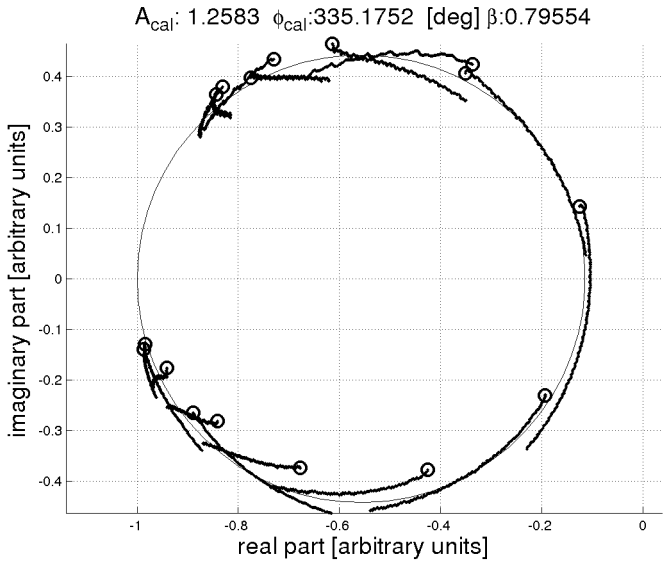

Figure 4.8.: Each connected line is a sequence of reflection coefficients recorded during one pulse of 250 μs length. The small circles mark the beginning of the pulses. The large circle is the best circle fitted with the χ^2-criteria 4.5. From theory one would expect that all points lie on a perfect circle. Instead, one can see a systematic phase-drift of the points during the pulse.

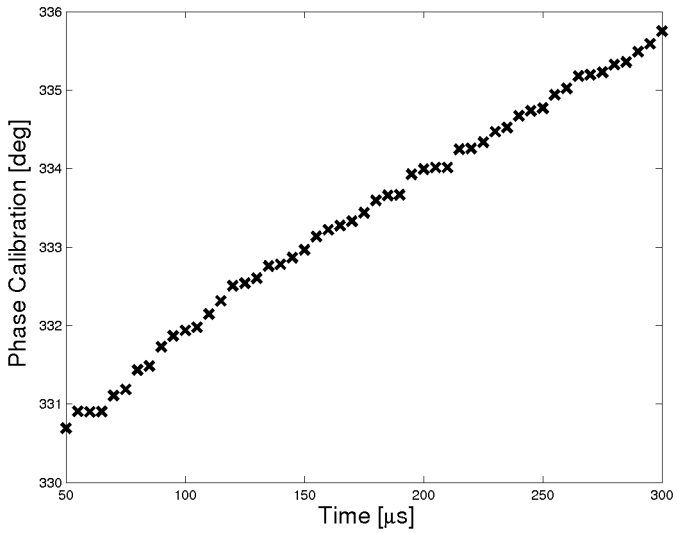

Figure 4.9.: Variation of the phase calibration over the pulse. Here, the data from many pulses was drawn into 250 resonance-circle diagrams, where each diagram represents the ocurrence in time of the data inside a pulse. For example, the small circles from 4.8 form the first microsecond of their pulse and therefore are taken for the calculation of the phase calibration of the first microsecond.

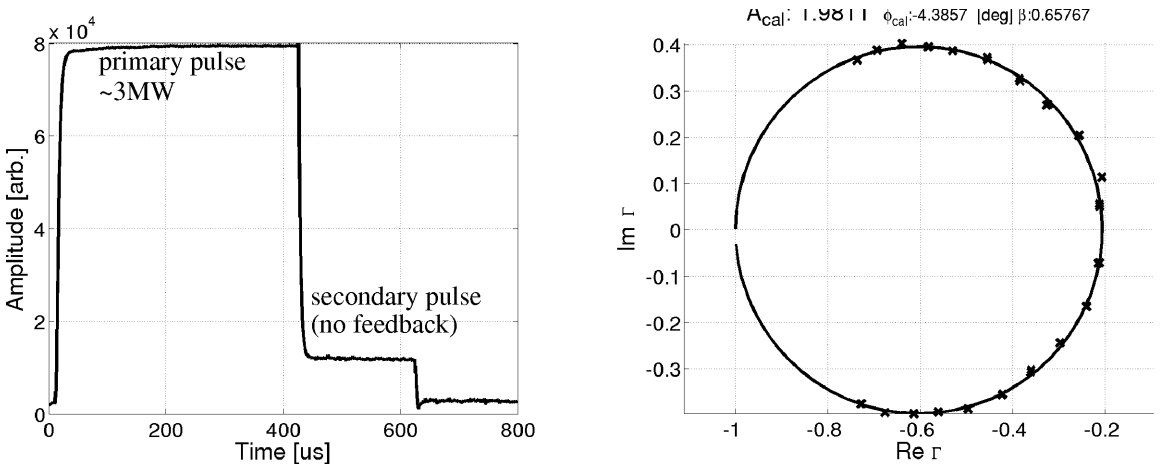

Figure 4.10.: Practical example for a calibration performed with the introduced algorithm at FLASH. The small step at the end of the primary pulse on the left picture is used for the calibration. The right picture shows the reflection coefficients in the complex plane after calibration together with all calibration coefficients.

the calibration of the forward and reflected signal at the electron gun. The results are compatible with the precision requirement stated in section 3.9.1 of about 1 % (amplitude) and 1° (phase). A comparison on the sub-percent or sub-degree was not done. The method based on the DAC induced detuning is suited best for FLASH since it does not disturb the operation.

4.3.2. Forward and Reflected Power at Superconducting Cavities

A calibration of the amplitudes of the forward and the reflected wave of the superconducting cavities is given by constant (or slowly varying) $a, b, c, d \in \mathbb{C}$ that fulfill

$$\begin{aligned} U_{\text{for}} &= aU^*_{\text{for}} + bU^*_{\text{ref}} \\ U_{\text{ref}} &= cU^*_{\text{for}} + dU^*_{\text{ref}} \\ \text{with } U &= U_{\text{for}} + U_{\text{ref}}. \end{aligned} \tag{4.8}$$

U^*_{for} and U^*_{ref} are the measured fields from the forward and the reflected wave. U_{for} and U_{ref} are the calibrated fields. U, without an index, is the field inside the cavity. a, b, c and d are constant calibration factors that correct for the various elements between the actual measurement and the digitized signals. b and c are non-diagonal elements that compensate for proportional coupling from the forward to the reflected channel and vice versa. This crosstalk can be quantified by looking at the inverted equation (in matrix formalism):

$$\begin{pmatrix} U^*_{\text{for}} \\ U^*_{\text{ref}} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} U_{\text{for}} \\ U_{\text{ref}} \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} U_{\text{for}} \\ U_{\text{ref}} \end{pmatrix}. \tag{4.9}$$

The measured forward signal U^*_{for} is thus a linear combination of the true forward signal U_{for} and the true reflected signal U_{ref}. The ratio of the coefficients of the contributions from U_{for} and U_{ref} to U^*_{for} is the crosstalk. The crosstalk can be calculated as

$$\begin{aligned} U_{\text{for}} \text{ to } U^*_{\text{ref}} &: -\frac{b}{d} \\ U_{\text{ref}} \text{ to } U^*_{\text{for}} &: -\frac{c}{a}. \end{aligned} \tag{4.10}$$

a, b, c and d are four complex (eight real) unknowns which means four complex degrees of freedom. Two can be eliminated by a simple linear regression. A third one is fixed by looking at the end of a pulse where the klystron is switched off and the fourth one requires deeper knowledge of the cavity envelope approximation in equation 3.11 and derived equations. The three steps for the determination of the calibration coefficients will be presented.

With $X := a + c$ and $Y := b + d$ one can write

$$U = XU^*_{\text{for}} + YU^*_{\text{ref}}. \tag{4.11}$$

From a single pulse with sampled data points $U(t_k), U^*_{\text{for}}(t_k), U^*_{\text{ref}}(t_k),\ k = 1..N$ this be can used to state a linear, overdetermined system of equations:

$$\begin{pmatrix} U(t_1) \\ U(t_2) \\ \vdots \\ U(t_N) \end{pmatrix} = \begin{pmatrix} U^*_{\text{for}}(t_1) & U^*_{\text{ref}}(t_1) \\ U^*_{\text{for}}(t_2) & U^*_{\text{ref}}(t_2) \\ \vdots & \vdots \\ U^*_{\text{for}}(t_N) & U^*_{\text{for}}(t_N) \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \tag{4.12}$$

Via multiple linear regression the best estimates for X and Y can be determined*. This reduces the number of complex degrees of freedom to two. The fact that a rather large region in the complex IQ-plane is covered during one pulse ensures that the matrix is not (close to being) singular and that numerical inversion is possible.

The next degree of freedom is determined from a pulse at the time where no forward power is present. For the FLASH timing this is the case at $t_{\text{off}} = 1.4\,\text{ms}$ after a pulse has started. From $U_{\text{for}} = aU^*_{\text{for}} + bU^*_{\text{ref}}$ one can simply conclude

$$-U^*_{\text{ref}}/U^*_{\text{for}} = a/b =: Z \quad \text{for } t > t_{\text{off}} \tag{4.13}$$

since U_{for} (the real forward amplitude) is zero at that time, i. e., no forward traveling wave is present at the directional coupler. In the absence of a forward traveling wave the term reflected wave may be misleading, therefore it is more appropriate to speak about the backward traveling wave instead.

The last complex degree of freedom finally requires knowledge of the cavity dynamics. Given

*The Matlab® language offers the "\"-operator for a multiple linear regression. The regression then has a similar syntax to a simple matrix division.

that X, Y, Z and e. g. a are known, b, c and d follow directly from

$$\begin{aligned} b &= \frac{a}{Z} \\ c &= X - a \\ d &= Y - \frac{a}{Z} \end{aligned} \tag{4.14}$$

The complex coefficient a can be found numerically by requiring that the bandwidth is constant over the pulse. Equation 3.22 on page 19 (cavity mode equation in polar coordinates) can be used to formulate a χ^2-criterion. The second equation of 3.22,

$$\dot{r} + r\omega_{1/2} = \omega_{1/2}\rho\cos(\theta - \varphi) \tag{4.15}$$

is very convenient since it does not contain the detuning $\Delta\omega$ explicitly. With $x = re^{i\varphi} = U$ being the cavity field and $u = \rho e^{i\theta} = 2U_{\text{for}}$ (the factor 2 is owed to the coupling), a χ^2-criterion can be stated:

$$\chi^2(a) = \sum_{t_k} \left[\frac{\mathrm{d}}{\mathrm{d}t}|U(t_k)| + (|U(t_k)| - 2|U_{\text{for}}(t_k)|\cos(\angle U_{\text{for}}(t_k) - \angle U(t_k)))\,\omega_{1/2,\text{end}} \right]^2 \tag{4.16}$$

$\omega_{1/2,\text{end}}$ is the half-bandwidth of the cavity determined at the end of the pulse and U_{for} is the amplitude of the calibrated forward power calculated with a certain a. A good starting point for the numerical search for the coefficient a is of course the X that was determined previously since the expected crosstalk is rather small. Figure 4.12 shows the change of χ^2 for an example pulse. For the left picture, the amplitude was varied while for the right picture, the phase was varied. On both plots, the units of the x-axis is relativ to the complex X. One can see that the difference between a and X is small but significant.

Figure 4.11 shows the cavity field and the forward and reflected amplitudes. The crosstalk amplitudes determined from this pulse are $|b/d| = 2.9\%$ (forward to reflected) and $|c/a| = 1.7\%$ (reflected to forward) which are typical values.

The determination of the calibration coefficients is fundamental for the RF-based de-tuning estimation (section 4.4.4) and the on-line beam phase measurement. The confirmation of the de-tuning measurement by the pulse-shortening-method presented in section 4.4.4 about the hardware detuning monitor can be considered as a confirmation of the presented calibration parameter estimation.

Although the calibration procedure described turns out to be robust and reproducible, a few things should be mentioned.

- The procedure assumes that only crosstalk between the measured forward and reflected wave exists. However, a non-negligible amount is coupled between channels of different cavities. In the present FLASH LLRF installation, RF signals are downconverted on a single board in groups of eight channels. It is known that crosstalk between the channels of one board exists. In the presence of a large inter-channel crosstalk, the first step in the presented algorithm (determination of X and Y) leads to suspicious results: the shapes of

U and $XU^*_{\text{for}} + YU^*_{\text{ref}}$ differ significantly by about a percent for some cavities.

- In this chapter, no U^* (measured cavity field) is distinguished from U (real cavity field). The presented calibration with the coefficients a, b, c, d is only applicable except for an unknown factor (with amplitude and phase) that scales the measured probe signal. For many applications like the detuning measurement this additional complex coefficient is not of interest. Therefore, the amplitudes of forward and reflected wave are only calibrated with respect to the measured cavity field. In order to determine the power from the measurement it would be sufficient to know the shunt impedance and have the field probe calibrated in MV/m.
- a, b, c and d can only compensate for linear crosstalk. There are indications that these coefficients are not constant but vary with the amplitude and the phase of the cavity field.

4.3.3. Beam Based Vector-Sum Calibration

Beam based vector-sum calibration is a well-known method for the high- and the low-power calibration depicted in figure 3.12, it is discussed for example in [15] and [16]. The high power calibration ensures that all cavities in a vector-sum experience excitation with synchronized phases and identical amplitudes from the waveguide distribution system. The low power calibration on the other hand compensates for amplitude- and phase-distortions done during the field-measurement occurring on the path from the sensor to the signal processor. Both can be extracted by looking at the transients induced by the beam.

The precondition for the beam based vector-sum calibration is that the beam, while passing through the module, excites fields in all cavities by the same strength at exactly the same time with a uniform bunch-profile (i.e., charge and phase are identical over the bunch-train). This assumption is not completely fulfilled since the shunt impedance varies between cavities. The resulting limits are discussed later in this section.

Ideally, the calibration procedure would be performed while no initial field is present inside the cavities. This would allow to clearly identify the contribution from the beam load. In practice, this is not possible for the first few modules of a linac where the energy is still comparatively low. The energy loss of the beam is so high that it would hardly reach the end of a module, unless it enters already at a very high energy. In order to be able to distinguish between the RF-amplifier induced cavity field and the beam induced field, one can assure that the drive applied on the RF amplifier is constant in amplitude and phase. The field in the cavities excited from the RF amplifier therefore changes only linear in time if at all and is relatively easy to distinguish from the beam induced field, since the total field is just a superposition of both.

Figure 4.13 shows beam based vector-sum calibration for a single cavity. Three linear fits are made in the region of interest, leading to three slopes, $a_{\text{pre-beam}}$, $a_{\text{during-beam}}$ and $a_{\text{after-beam}}$. The slopes are complex since the function $t \mapsto U^*(t)$ has complex values (the star * indicates the uncalibrated measurement). At the time the beam stops, $t = t_{\text{beam-stop}}$, the field in the cavity $U(t_{\text{beam-stop}})$ is a superposition of the generator induced voltage $U_{\text{gen}}(t_{\text{beam-stop}})$ and the beam induced voltage $U_{\text{beam}}(t_{\text{beam-stop}})$. With the slopes from the linear fits and the duration of the beam,

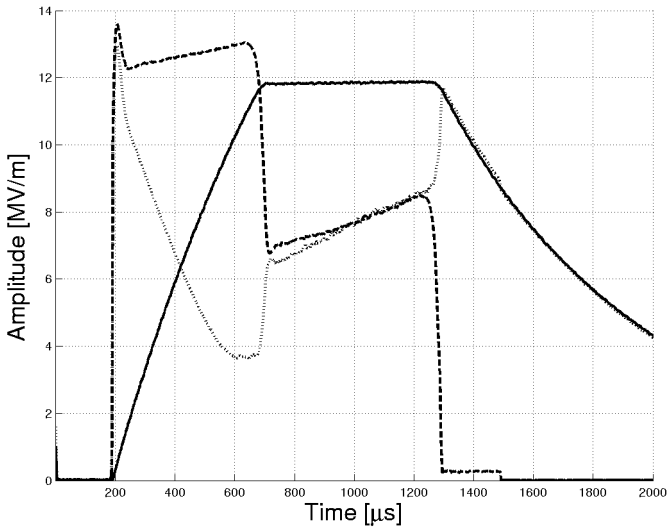

Figure 4.11.: Amplitudes of the calibrated signals of one cavity. The cavity field U is the solid line, forward (U_{for}) and reflected wave (U_{ref}) are dashed respectively dotted.

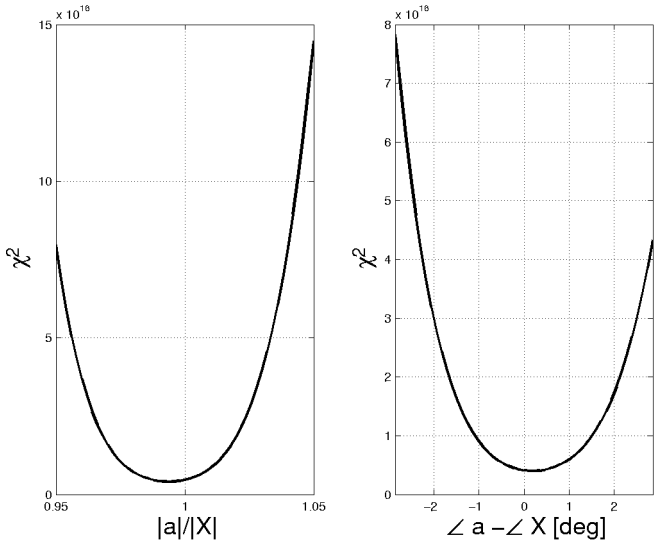

Figure 4.12.: Change of the χ^2 mentioned in the text. On the left picture, the amplitude of parameter a is changed relative to X, while on the right picture the phase of a is changed relative to the phase of X. The minima of the curves do not fall together with 0, which means that a slightly differs from X.

	I	II	III	IV	V	VI	VII	IIX
	low power amplitude (DSP calibration)							
A	1.00	0.79	1.00	1.03	1.28	0.94	0.82	1.11
$A_{\text{error}}/A \cdot 100$	0.00	0.91	1.06	1.43	0.96	0.78	0.66	1.82
	low power phase (DSP calibration)							
φ	0.00°	81.52°	−29.40°	27.51°	99.35°	−70.60°	46.03°	−83.55°
φ_{error}	0.00°	0.31°	0.51°	0.34°	0.44°	0.36°	0.37°	0.35°
	high power amplitude (waveguide distribution)							
A	1.00	1.02	0.99	1.02	1.66	1.62	1.15	1.61
$A_{\text{error}}/A \cdot 100$	0.00	1.17	1.06	1.41	1.23	1.35	0.95	2.63
	high power phase (waveguide distribution)							
φ	0.00°	4.21°	3.84°	−0.54°	3.55°	−8.15°	−2.27°	3.72°
φ_{error}	0.00°	0.29°	0.52°	0.27°	0.43°	0.42°	0.41°	0.33°

Table 4.1.: Coefficients for low- and high-power calibration obtained with the method of beam based vector-sum calibration for all cavities (I-IIX) of the first module ACC1 of FLASH.

Δt_{beam}, estimates can be made:

$$\begin{aligned} U^*_{\text{beam}}(t_{\text{beam-stop}}) &= \left(\frac{a_{\text{pre-beam}} + a_{\text{after-beam}}}{2} - a_{\text{during-beam}}\right)\Delta t_{\text{beam}} \\ U^*_{\text{gen}}(t_{\text{beam-stop}}) &= U(t_{\text{beam-stop}}) - U_{\text{beam}}(t_{\text{beam-stop}}) \end{aligned} \tag{4.17}$$

The high power and the low power calibration are directly calculated from $U^*_{\text{gen}}(t_{\text{beam-stop}})$ and $U^*_{\text{beam}}(t_{\text{beam-stop}})$. The argument $t_{\text{beam-stop}}$ is omitted from now on.

- The low power amplitude calibration factor is proportional to the inverse of the measured amplitude transient, $1/|U^*_{\text{beam}}|$.
- The high power amplitude calibration is the measured amplitude of the generator induced voltage corrected by the measured beam induced voltage, $|U^*_{\text{gen}}|/|U^*_{\text{beam}}|$.
- The high power phase calibration is the difference between the phase of the beam induced amplitude, $\angle U_{\text{beam}} - \angle U_{\text{gen}}$.
- The low power phase calibration is the difference between the acceleration angle $\angle U^*_{\text{beam}} - \angle U^*_{\text{gen}}$ and the measured generator angle $\angle U^*_{\text{gen}}$. Thus, it is $\angle U^*_{\text{beam}} - 2\angle U^*_{\text{gen}}$.

It should be noted that a single transient yields four calibration numbers, namely amplitude and phase for the low-power and the high-power calibration. For the low-power calibration it is common at FLASH to normalize all calibration in a way that the first cavity has amplitude calibration factor 1 and phase calibration 0°. The amplitude calibration factors carry no units after the normalization.

Phase and amplitudes of low and high power calibration have been calculated during a vector-sum calibration of the first module (ACC1) of the superconducting accelerator of FLASH and

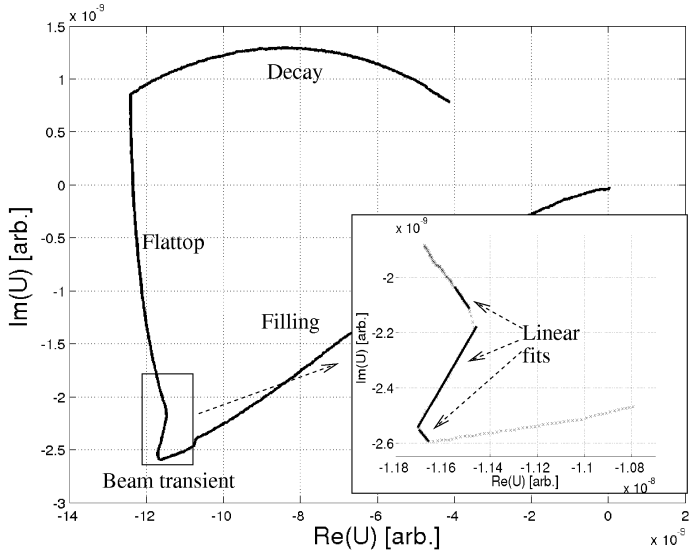

Figure 4.13.: Beam induced transients used for vector-sum calibration. The big picture shows the accelerating voltage of a single cavity in open-loop mode as a parameter plot in the complex plane. The small picture shows the region of interest together with linear fits.

lead to the results in table 4.1. The results for the amplitudes A and the phases ϕ are normalized to the first cavity. The errors A_{error}, ϕ_{error} are the statistical error of the mean which is the standard deviation divided by the squareroot of the number of measurements. The statistical errors are acceptable compared to the requirements developed in section 3.9.2. However, there exist systematical errors that are explained later in this section. The results for the high power amplitude calibration reflect that the first four cavities are reduced in gradient compared to the last four. In addition, it shows attenuation of cavity seven which is caused by an extra attenuator that has been mounted due to a reduced performance of that cavity.

The beam based vector-sum calibration has several weaknesses. The beam induced field is subject to phase rotations due to detuning. The acceleration angle is not related to the phase of the beam induced field but rather to the phase of the beam current. The error made in presence of a detuning can be estimated from the transient response in phase in equation 3.15, $[(\mathrm{d}/\mathrm{d}t)\angle x_{\text{step}}]_{t=0} = \Delta\omega/2$. At a detuning of 100 Hz and a beam duration of 30 μs, a detuning angle of 0.5° is expected for the beam induced field. This is already of the order of the critical precision that is determines in section 3.9.2. In principle it is possible to calculate the detuning and take it into account for the vector-sum calibration.

Another weakness is the assumption that the beam current induces identical fields in all cavities of a vector-sum. The normalized shunt impedance (r/Q) of the whole cavity is a geometrical constant. For coupled cells the coupling between the cells determines the overall shunt impedance. The loaded shunt impedance $R_{\text{l}} = (r/Q)Q_{\text{l}}$ relates the accelerating voltage experi-

enced by a particle to the power loss of the cavity. Given that $(r/Q)_{\text{cell}}$ is the normalized shunt impedance of a single cell, an estimate for the normalized shunt impedance of the coupled system of a 9-cell resonator can be made based on the different amplitudes in the cells. In the following consideration, the overall normalized shunt impedance is determined as the ratio of the squared sum of the accelerating voltages in the individual cells U_i divided by the sum of the power losses in the cavities. Since the normalized shunt impedance is usually given per meter, the length of a cavity l or of a single cell l/N is used in this consideration.

$$l\left(\frac{r}{Q}\right)_{\text{cavity}} = \frac{(\sum U_i)^2}{\sum U_i^2 / \left(\left(\frac{r}{Q}\right)_{\text{cell}} \cdot \frac{l}{N}\right)} \tag{4.18}$$

This equation can be simplified by inserting a well known formula for the root mean square of a variable, $U_{\text{rms}}^2 =< U_i^2 > - < U_i >^2$.

$$\begin{aligned}
\left(\frac{r}{Q}\right)_{\text{cavity}} &= \frac{1}{N}\left(\frac{r}{Q}\right)_{\text{cell}} \frac{N^2\left(\frac{1}{N}\sum U_i\right)^2}{N\frac{1}{N}\sum U_i^2} \\
&= \left(\frac{r}{Q}\right)_{\text{cell}} \frac{< U_i >^2}{< U_i^2 > - < U_i >^2 + < U_i >^2} \\
&= \left(\frac{r}{Q}\right)_{\text{cell}} \frac{< U_i >^2}{U_{\text{rms}}^2 + < U_i >^2} \\
&= \left(\frac{r}{Q}\right)_{\text{cell}} \frac{1}{\left(\frac{U_{\text{rms}}^2}{< U_i >^2}\right)^2 + 1}
\end{aligned} \tag{4.19}$$

Since the transient induced by a beam is directly proportional to the shunt impedance of a cavity, 4.19 can be used to verify the validity of the assumption that the beam induces identical voltages in all cavities of a cryomodule. For a linear distribution of amplitudes in the cells of a cavity with a relative peak to peak distance of 20%, the normalized RMS value $U_{\text{rms}}/ < U_i >$ is 7 %. From 4.19 follows a reduction of the overall shunt-impedance of 0.5 %, which is already half of the critical region determined for the XFEL in section 3.9.2.

LLRF control is sensitive to the calibration of the field vectors contributing to the vector-sum. The beam based calibration is a reliable and fast method for obtaining it. The application of the beam based vector-sum calibration interrupts the normal operation of the accelerator. The precision achieved with the beam based method is sufficient for accelerators like FLASH but might need improvement for future accelerators like the European XFEL project. A transient measurement technique based on an RF notch filter is currently under development and has first promising results [21]. The new technique allows to evaluate transients of small beam currents to a high precision.

4.4. Control Theory Issues

4.4.1. Introduction to Control Theory

This section will introduce some basic elements of control theory that will be used throughout this and the following sections. A few elements, namely discrete transfer functions, have already been introduced in section 3.8, where Z-transformations and digital filters were discussed. A more comprehensive introduction to control theory can be found in [22].

Transfer Functions

Transfer functions describe linear, time-invariant systems with one input and one output. The differential equation between an input $u(t)$ and an output $x(t)$ of such a system can be transformed into Laplace- or Fourier-space, where linear operators like $\mathrm{d}/\mathrm{d}t$ turn into algebraic functions of an independent variable s (Laplace) or ω (Fourier). Application of transfer functions on signals or associative combination of transfer functions is purely multiplicative in the Laplace- or Fourier-domain.

The Laplace transformation and its inverse of a function $f(t)$ (following the definition in [22]) are given by

$$F(s) = \mathscr{L}[f(t)] = \int_0^\infty f(t)e^{-st}\mathrm{d}t \tag{4.20}$$

$$f(t) = \mathscr{L}^{-1}[F(s)] = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} F(s)e^{st}\mathrm{d}s. \tag{4.21}$$

The Fourier transformation can be obtained from a Laplace transform via the substitution $s \mapsto i\omega$. Table 4.2 gives some examples for Laplace transformations. A longer list can be found in [22] on page 636.

Name	Time domain $f(t)$	Laplace domain $F(s)$
Unit impulse	$\delta(t)$	1
Unit step	$u(t)$	$\frac{1}{s}$
Exponential	e^{-at}	$\frac{1}{s+a}$
Derivative	$\frac{\mathrm{d}}{\mathrm{d}t}$	s
Integral	$\int \mathrm{d}t$	$\frac{1}{s}$
Multiplication	$g(t) \cdot h(t)$	$G(s) * H(s)$
Convolution	$g(t) * h(t)$	$G(s) \cdot H(s)$

Table 4.2.: A few examples for Laplace transformations on functions and operators.

Block Diagrams

Block diagrams as in figure 4.18 are typically used to visualize control algorithms. Central elements of a block diagram are boxes representing transfer functions. The input and the output of a two-port (box) are connected via the transfer-function. Additional elements in a block diagram are summing junctions and signal routing. A block diagram that has an input and an output can be interpreted as a two-port itself and can therefore be described by a transfer function. For simple topologies the combined transfer function for a whole block diagram is just a multiplication of its elements. For topologies with loops (as in figure 4.18), some rules can be applied in order to determine the overall transfer function. For a simple closed loop, the transfer function can be determined with Mason's gain formula [22],

$$\text{closed-loop transfer function} = \frac{\text{transfer function of the straight path}}{1 - \text{open-loop transfer function}}. \tag{4.22}$$

The open-loop transfer function is the transfer function that is obtained by concatenating all transfer function that are part of the loop in a diagram. For figure 4.18, the transfer function of the straight path is $C(s) \cdot A(s) \cdot P(s)$ which is at the same time the open-loop transfer function. The closed-loop transfer function in that example is therefore

$$H(s) = \frac{C(s)A(s)P(s)}{1 + C(s)A(s)P(s)}. \tag{4.23}$$

The plus-sign is owed to the fact that the summing-junction in figure 4.18 carries a minus sign indicating that signals should be subtracted.

Bode Diagrams and Pole-Zero Maps

Linear transfer functions with a sinusoidal signal at their input produce sinusoidal signals at their output in steady-state. The amplitude and phase difference of in- and outgoing signals can be visualized by means of a Bode diagram. The resonance curve that is shown in figure 3.1 is a typical Bode blot, where the amplitude and the phase-response are shown in different diagrams depending on the frequency. From a Bode plot, the basic effect of a transfer function on a signal can be estimated.

Another way to visualize transfer functions is to display their poles and zeros in a complex plane. For transfer functions that can be expressed as rational functions the zeros are the roots of the numerator and the poles the roots of the denominator. The position of the poles reveals the dynamics of a transfer function, therefore the pole-zero map does not only inform about the steady-state behavior of a system but also its transient response. The real part of a pole gives the damping time while the imaginary part gives the oscillation frequency.

Stability Criteria

Looking at Mason's gain formula it appears that the open-loop transfer function (e. g. a proportional gain for the controller C) should be as large as possible in order to reject disturbances. In

practice restriction apply that become clear when looking at the open-loop bode diagram. Due to the elements in the open-loop transfer function, the absolute phase advance can exceed 180° for frequencies higher than a critical frequency ω_{crit}. This is equivalent to a flip in sign and therefore equivalent to adding the plant error signal instead of subtracting it at the summing junction (a look at the example in figure 4.18 is useful, here). The stability of the system depends on the amplitude of the open-loop transfer function for frequencies larger than ω_{crit}. If it is larger than 1 (amplifcation), errors with that frequency will be amplified rather than suppressed. In practice, a safety phase margin from the unity gain or a safety phase margin from 180° is used in order to guarantee stability. [22] recommends as a rule of thumb to have a gain margin of 8 dB (factor 2.51) and a phase margin of 50°.

Another stability criterion can applied in order to test if a system is so-called bounded-input, bounded-output (BIBO)-stable, that is, if a system has always a limited output if the input is limited. It can be tested by looking at the pole-zero map of a transfer function. If poles exist on the positive half of the complex plane, BIBO-stability is not given.

Step-Response

A step-response is the response $x = x_{\text{step}}$ of a two-port on a step $u = \theta$, $\theta(t) = 0$ for $t < 0$, $\theta(t) = 1$ else, at its input. By looking at the transfer function H,

$$H(s) = \frac{b_M s^M + b_{M-1} s^{M-1} + \ldots + b_0}{a_N s^N + a_{N-1} s^{N-1} + \ldots + a_0} \tag{4.24}$$

or the equivalent differential equation in time domain

$$b_M \frac{\mathrm{d}^M}{\mathrm{d}t^M} u(t) + b_{M-1} \frac{\mathrm{d}^{N-1}}{\mathrm{d}t^{M-1}} u(t) + \ldots + b_0 u(t) = a_N \frac{\mathrm{d}^N}{\mathrm{d}t^N} u(t) + a_{N-1} \frac{\mathrm{d}^{N-1}}{\mathrm{d}t^{N-1}} u(t) + \ldots + a_0 u(t) \tag{4.25}$$

one can draw conclusions for the step response of the system: the steady-state value is determined by the ratio of the coefficients with the lowest power in s that are not both equal to zero. A division by zero would indicate that a step-response grows until infinity. This rule can be proven by applying equation 4.25 in the steady-state case, where all derivatives vanish. On the other hand, the immediate response of the system ($x_{\text{step}}(\epsilon)$ with $\epsilon \to 0$ and $\epsilon > 0$) is obtained from the ration of the coefficients with the highest power in s that are not both equal to zero. A division by zero indicates that the system is non-causal (i. e., the degree of the numerator is higher than the degree of the denominator). This is proven by integrating both sides of equation 4.25 several times. Integrals of $u(t)$ and $x(t)$ then vanish since both are zero for $t < 0$.

4.4.2. Passband Modes

Other modes exist in a cavity beside the accelerating π-mode. In the case of TESLA cavities, the closest mode is the $8/9\pi$-mode with a center-frequency that is only 800 kHz away. In the digital controller, signals are represented as an envelope with respect to a 1.3 GHz-reference. The π-mode is therefore represented by a DC-signal while e. g. the $8/9\pi$-mode is represented by a

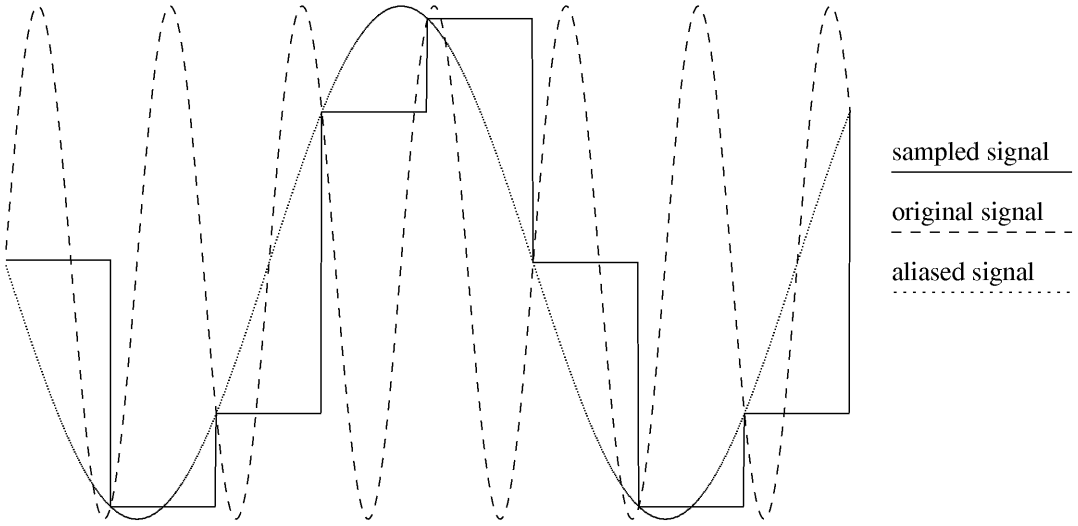

Figure 4.14.: A signal is sampled at a frequency that is beyond the Nyquist frequency. The sampled signal and the aliased signal are shown.

800 kHz-signal. 800 kHz however is beyond the Nyquist frequency of 500 kHz at a sampling frequency of 1 MHz. In the following, the attenuation of a signal beyond Nyquist will be deduced. It is followed by a measurement that evaluates the potential of the cavity modes to lead to an instability.

The Fourier component at frequency ω of a sinusoidal signal with the frequency ω that is sampled at a frequency ω_s is the integral over a whole period of the function

$$g(t) = e^{i\omega t}\sin(\omega D(t)), \quad D(t) = \frac{2\pi}{\omega_\mathrm{s}} \cdot \mathrm{floor}\left(\frac{\omega_\mathrm{s} t}{2\pi}\right) \tag{4.26}$$

$D(t)$ is the function that maps each t to the biggest discrete sampling point smaller than t ("floor"), whereas the sampling frequency is given by ω_s. The function $\sin(\omega D(t))$ is the stepwise constant function shown in figure 4.14. Since the e-function and the sine in equation 4.26 have both the same ω in their arguments, it is helpful to introduce the possible, discrete times t_k and discrete phases φ_k,

$$\varphi_k = \omega t_k = \omega k \frac{2\pi}{\omega_\mathrm{s}}. \tag{4.27}$$

For a rational ratio ω_s/ω, a numbers $N, M \in \mathbb{N}$ exist for which the series $\varphi_j,\ j = 1, 2, \ldots$ repeats after N points, $\varphi_j = \varphi_{j+N} - M \cdot 2\pi$. For example, for $\omega = 2\pi \cdot 800\,\mathrm{kHz}$ and $\omega_\mathrm{s} = 2\pi \cdot 1\,\mathrm{MHz}$ one has $\varphi_j = 2\pi/5, 4\pi/5, .., 2\pi$. The function g can be written

$$g(t) = e^{i\omega[t - D(t)]} e^{i\omega D(t)} \sin \omega D(t). \tag{4.28}$$

The argument $t - D(t)$ is in the interval $[0, 2\pi/\omega_\mathrm{s}]$. The fourier amplitude of the component is

given by

$$F(\omega) = \left| \frac{2}{NT} \int_0^{NT} g(t) \mathrm{d}t \right|, \tag{4.29}$$

where $T = 2\pi/\omega_\mathrm{s}$ is the time between subsequent samples and NT the time period after which the values of the sampled data points repeat. The factor 2 is owed to the fact that the frequency ω is also aliasing to $-\omega$. The last expression can be expressed with the stepwise constant phases φ_k and a sum over the total number N of different phases.

$$\begin{aligned} F(\omega) &= \frac{2}{N}\frac{\omega_\mathrm{s}}{2\pi} \left| \int_0^{2\pi/\omega_\mathrm{s}} e^{i\omega t} \mathrm{d}t \right| \left| \sum_{j=1}^{N} e^{i\varphi_j} \sin\varphi_j \right| \\ &= \frac{2}{N}\frac{\omega_\mathrm{s}}{2\pi} \left| \frac{1}{i\omega} \left(e^{2\pi i\omega/\omega_\mathrm{s}} - 1 \right) \right| \left| \sum_{j=1}^{N} \sin\varphi_j \cos\varphi_j + i \sum_{j=1}^{N} \sin^2\varphi_j \right| \end{aligned} \tag{4.30}$$

The φ_j are equally spaced and repeat after N samples and thus cover at least one turn of a whole circle. Thus the sinus-cosinus sum vanishes while the sinus-squared sum is $0.5N$. The absolute value of $e^{2\pi i\omega/\omega_\mathrm{s}} - 1$ is determined with the law of cosines, $|e^{2\pi i\omega/\omega_\mathrm{s}} - 1| = \sqrt{2 - \cos 2\pi\omega/\omega_\mathrm{s}}$,

$$F(\omega) = \frac{1}{2\pi}\frac{\omega_\mathrm{s}}{\omega}\sqrt{2 - 2\cos 2\pi\omega/\omega_\mathrm{s}}. \tag{4.31}$$

Figure 4.15 shows the shape of $F(\omega)$ for a sample frequency of $\omega_\mathrm{s} = 1\,\mathrm{MHz}$ (10 MHz). For ω equal to the Nyquist-frequency of 500 kHz, the value of $F(\omega)$ is just 64 % (or −4 dB). For $\omega = 2\pi \cdot 800\,\mathrm{kHz}$, $F(\omega)$ is 23 % (or −13 dB). (For 10 MHz, the numbers are 99.6 % (−0.04 dB) at 500 kHz and 99.0 % (−0.09 dB) at 800 kHz.

So far, formula 4.31 was used for calculating the strength of a certain frequency ω after the digitization process. It can also be used for calculating the strenght of a frequency ω' that is resulting from aliasing after a certain frequency ω was digitized. For example, if a frequency $\omega = 2\pi \cdot 800\,\mathrm{kHz}$ is sampled at $\omega_\mathrm{s} = 2\pi \cdot 1\,\mathrm{MHz}$ it will alias to 200 kHz, 1.2 MHz and so on. For the digitation process it does not make any difference whether the original frequency ω is 800 kHz or any of its aliases. Therefore, the strength of a certain alias frequency can be calculated directly with $F(\omega')$.

Passband modes can lead to instabilities in a closed loop feedback. Section 4.4.1 explained that instability can occur if the closed-loop transfer functions has a phase shift of 180° and an amplitude larger than one for certain frequency ranges. This can be extended, since frequency regions with an amplitude larger than one and a phase shift of 180° plus an integer multiple of 360° also lead to an instability. The phase shift can directly be influenced by the time delay in the loop. Figures 4.16 and 4.17 show the results of a measurement, where a single cavity is operated in closed loop and the delay in the loop is systematically increased. The proportional gain of 35 was chosen such that a significant instability is visible. Figure 4.16 shows time-domain pictures of a single pulses whith different extra delays. In the right part of figure 4.17, the rms over the

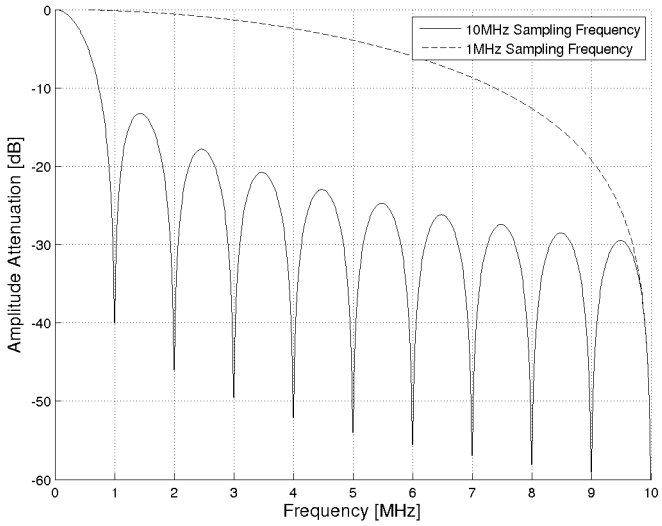

Figure 4.15.: Attenuation of digitally synthesized frequencies. The two curves show the relative amplitudes of frequencies generated with a sampling frequency of 1 MHz and 10 MHz.

flat-top is plotted versus the extra time delay. The periodicity is about 1.3 μs, which corresponds to a frequency of about 770 kHz. The left part of that figure shows the fourier transform of a pulse that is close to instability. It shows a major frequency component around 780 kHz. Since the $8/9\pi$-mode of a Tesla cavity is known to be about 800 kHz away from the accelerating π-mode, it is the best candidate for the origin of the instabilities presented here. The preceeding consideration of attenuation beyond the Nyquist frequency has shown that the attenuation of the $8/9\pi$-mode is not sufficient in order to avoid instabilities.

4.4.3. Determination of Loop Phase and System Gain

Introduction of Loop Phase and System Gain

In a linear single-cavity control-loop, loop phase and system gain can easily be identified with angle (loop phase) and magnitude (system gain) of the factor $A \in \mathbb{C}$ in

$$\dot{x} + (\omega_{1/2} - i\Delta\omega)x = A\omega_{1/2}u. \tag{4.32}$$

Except for the factor A, this equation is identical to the envelope approximation for cavity modes, equation 3.11. The meaning of A is a constant scaling in amplitude and a distortion in phase caused by the physical elements between the drive u and the detected field x, in particular by actuator- and sensor chain. If the control system produces the complex number u, the cavity will act as if it was driven by the complex number Au. In practice, A will rather be a function in time,

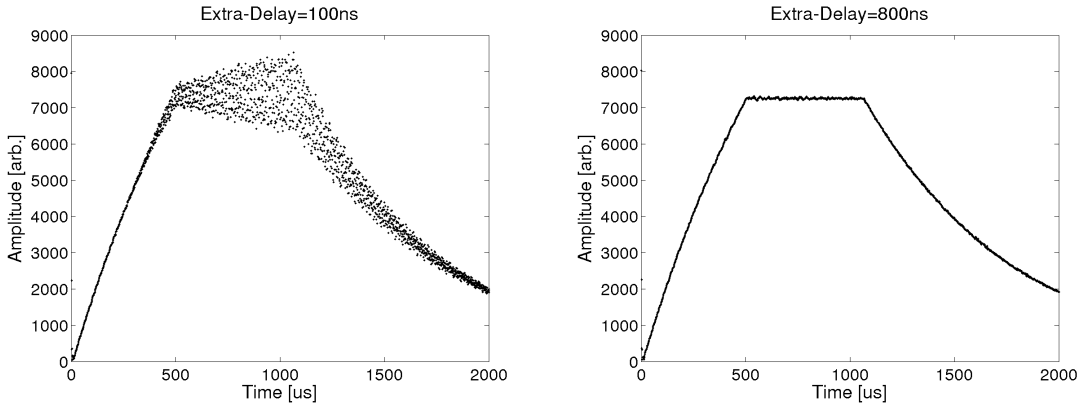

Figure 4.16.: Single-shot time-domain pictures of the amplitude of a cavity driven in closed loop with different extra delays. The instability on the left plot (extra delay is 100 ns) is clearly visible compared to the right plot (extra delay is 800 ns), even though the total loop delay is shorter.

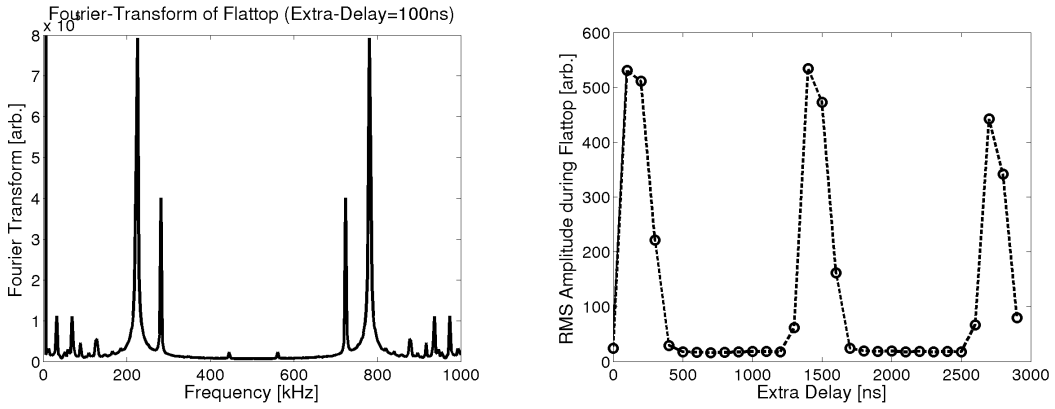

Figure 4.17.: The left plot is the fourier transform of a flat-top with instabilies at 100 ns extra-delay. The right plot shows the rms values of the flat-top amplitude versus a systematic scan of the extra delay.

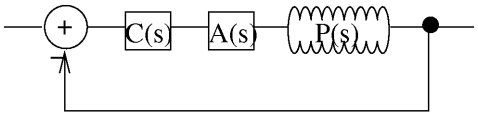

Figure 4.18.: Closed-loop diagram for a plant with linearly distorted inputs (represented by A). The distortion can be associated with the controller, C. If P is linear, it does not make a difference if A is an input- or an output-distortion of the plant P.

depending especially on the temperature (slow drifts).

Initially, before presenting a method that determines A, the meaning of loop phase and system gain shall be illuminated. In the presence of a distorting A, the closed-loop transfer function turns into

$$H = \frac{ACP}{1 + ACP} \tag{4.33}$$

with C being the controller and P the undistorted plant. A can be thought of as being associated to C rather than to P as depicted in figure 4.18. An interpretation of A would be that it just changes the controller C into AC. The effects of a $|A| = 1$ (but nonzero phase) and $\angle A = 0$ (but non-unity amplitude) will now be discussed separately and thereby lead to the notions system gain and loop phase.

For a moment it is assumed that A has a zero phase. For any (not too large) $|A|$, the loop can be closed without the risk of an instability. Given that the proportional feedback gain is 1 (i. e. $C = 1$ for the controller) and assuming that $P = 1$ in steady-state (which is the case for a cavity on resonance[†]), the transfer function from the setpoint to the cavity field can be written as

$$H = \frac{ACP}{1 + ACP} = \frac{A}{1 + A}. \tag{4.34}$$

Or, in other words: if and only if $A = 1$, the cavity will be at 50% of the setpoint value in steady-state and closed-loop operation. If A is larger, H in equation 4.34 will be larger (closer to unity) which is equivalent to a higher effective proportional gain. Therefore, knowledge of $|A|$ is important for the knowledge of the true loop gain. From here on, system gain will be defined as $|A|$, because it can be considered as the contribution from the system to the loop gain. In practice, compensation for A in the digital signal processor can be adjusted by the input matrices of figure 4.1. From equation 4.34, a simple way of determining the system gain is visible: For a feedback-only driven cavity, one has to alter the loop gain at the DSP until the cavity field is at 50% of its setpoint value.

Now the system gain $|A|$ shall be assumed to be equal to unity in order to illuminate the influence of a non-zero phase $\angle A$ on the closed loop. Figure 4.19 gives an example where a loop phase of $180\,^\circ$ causes an instablity in closed loop operation. The diagram shows the complex plane, which is in radio frequency engineering often referred to as the IQ-plane with I ("in phase") be-

[†]The coupling contribution $1 + \Gamma_0$ is by intention not taken into account for this definition of the system gain. Therefore, the field is not twice as large as the scaled drive in the near-to steady-state part of figure 4.21.

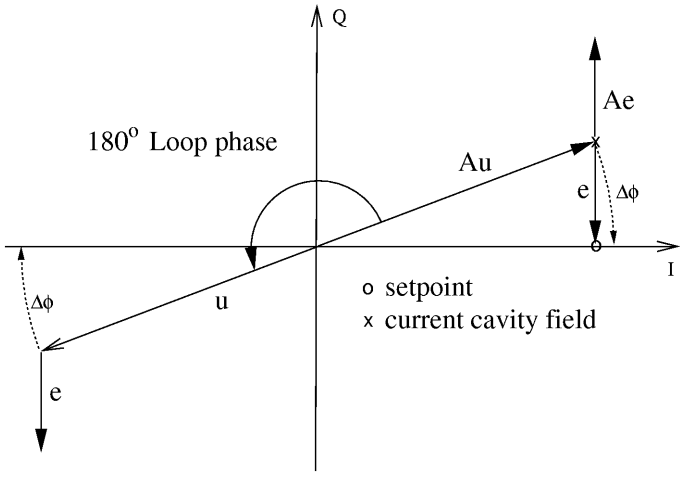

Figure 4.19.: Instability caused by an inadequate loop phase. The diagram shows the complex IQ-plane. The drive u produced by the controller is rotated by 180 ° compared to the drive Au arriving at the cavity. This leads to the instability described in the text. The dotted lines show the phase error and phase correction respectively in the case of Aϕ-control. Here, a loop phase can not cause instability.

ing the real part of the field and Q ("quadrature") the imaginary part. The current field vector x is indicated by an x and the setpoint x_0 by o. The drive Au (as seen from the undistorted cavity) falls together with the field itself, since steady-state operation on resonance is assumed. The controller output u produced by the control system is 180° out of phase compared to the plant input as seen by the cavity, Au. For a loop gain of 1, the feedback algorithm would directly add the error $e = x_0 - x$ to the controller output u. The cavity will then experience a drive that is corrected by Ae and thus move in the opposite direction of the setpoint! Due to the feedback, the distance in the IQ-plane between the field and its setpoint will grow exponentially.

As a side-remark it shall be mentioned that the illustrated instabilty does occur in IQ-control but not in Aϕ-control (amplitude-phase control). Figure 4.19 shows the phase-correction (but not the amplitude correction) that would be applied in Aϕ-control as a dotted segment. A rotation preserves the orientation even in the presence of a phase distortion and hence the feedback applied in this case does not lead to an instability.

Even in IQ-control, a small loop phase usually does not necessarily lead to an instability. A loop phase rather introduces a coupling between I and Q. From figure 4.19 it is visible that instabilty occurs if the loop phase is larger than 90°. In order to make a quantitative statement on how much a loop phase different from zero affects the field quality, equation 4.33 shall be investigated. Again, a cavity on resonance in steady-state is considered with a proportional controller. That is, ACP is replaced by a complex number z with the amplitude $|z|$ corresponding to the proportional gain AC and a phase equal to the loop phase $\angle A$. The phase $\Delta\varphi$ of the

expression $z/(1+z)$ is thus of interest:

$$\begin{aligned}\tan\Delta\varphi &= \frac{\mathrm{Im}(z/(1+z)))}{\mathrm{Re}(z/(1+z))} \\ &= \frac{z+z\bar{z}-\bar{z}-z\bar{z}}{z+z\bar{z}+\bar{z}+z\bar{z}} \\ &= \frac{\mathrm{Im}\,z}{\mathrm{Re}\,z+|z|^2} \\ &= \frac{\mathrm{Im}\,z/\,\mathrm{Re}\,z}{1+|z|^2/\,\mathrm{Re}\,z}\end{aligned} \tag{4.35}$$

For small angles $\angle z$ (small loop phase) and large $|z|$ (large proportional gain), one can replace $|z|^2/\,\mathrm{Re}\,z$ just with $|z|$ (the proportional gain) and yields the not quite unexpected result: the final error on the plant phase is just the loop phase divided by the proportional gain.

Measurement of Loop Phase and System Gain

Looking at equation 4.32, it seems obvious how to obtain A in a pulsed system (assuming that A does not change over the pulse): the knowledge of x, $\dot{x}$, $\omega_{1/2}$, $\Delta\omega$ and (non-zero) u for a single point inside a pulse allows to calculate A. $\omega_{1/2}$ and $\Delta\omega$ can be determined very convenient at the end of a pulse (refer to equation 3.17 on page 16). The bandwidth and the detuning at time $t = t_{\mathrm{end}}$, the end of the flat-top, shall be labeled $\omega_{1/2,\mathrm{end}}$ and $\Delta\omega_{1/2,\mathrm{end}}$. Now, taking x, $\dot{x}$ and u at the end of the pulse (or rather close-to the end of the flat-top, since after the flat-top the drive usually is zero) would mean to loose a lot of information since only a few samples can be involved. Especially the estimation of the derivative $\dot{x}$ suffers from taking only a few samples close to the end of the flat-top. Instead, one can take advantage from the fact that the detuning over a pulse follows a rather linear shape (as seen for example in figure 4.24) during the flat-top. An

$$A'(t) = \frac{\dot{x}(t) + (\omega_{1/2,\mathrm{end}} - i\Delta\omega_{\mathrm{end}})x(t)}{\omega_{1/2,\mathrm{end}}u(t)} \tag{4.36}$$

can now be defined. In this equation, the dependency on t is explicitly given in order to make clear that this $A'(t)$ is built from $x(t)$, $\dot{x}(t)$ and $u(t)$ from the whole pulse but with $\omega_{1/2,\mathrm{end}}$ and $\Delta\omega_{\mathrm{end}}$ taken from the end of the flat-top only. The derivative of x can be calculated inside an algorithm very quickly with a digital filter as in equation 3.52 on page 30. Since usually $\omega_{1/2}(t)$ and $\Delta\omega(t)$ change during the pulse, $A = A'(t)$ only for $t = t_{\mathrm{end}}$. Still, all of $A'(t)$ for t inside the flat-top of the pulse can be used for the estimation of $A = A'(t_{\mathrm{end}})$ by doing a linear extrapolation. This is illustrated in figure 4.20, where $A'(t)$ is plotted for the whole pulse and a linear fit has been performed in the flat-top region (with a safety margin towards the beginning and the end of the flat-top). The x marks the extrapolated $A'(t_{\mathrm{end}})$ at the end of the flat-top either for the system gain and the phase. Figure 4.21 shows the corresponding forward power and vector-sum signal of a single pulse. A vertical line separates the regions where RF is applied and where it is off. The forward power signal of figure 4.21 has already been scaled with the results of the system gain and loop phase determination.

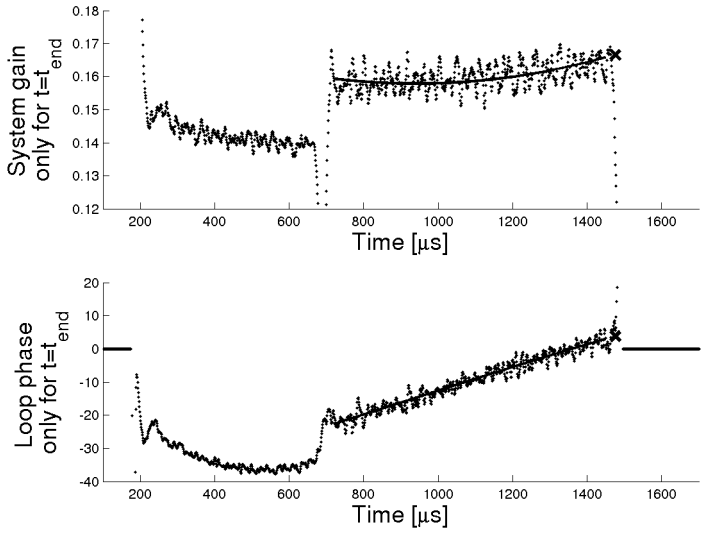

Figure 4.20.: Amplitude and phase of the time-dependent quantity $A'(t)$ for the pulse in figure 4.21 are plotted. A linear fit is performed for the region of the flat-top. The linear fit has a visible bend in the amplitude plot (still it is a linear fit).

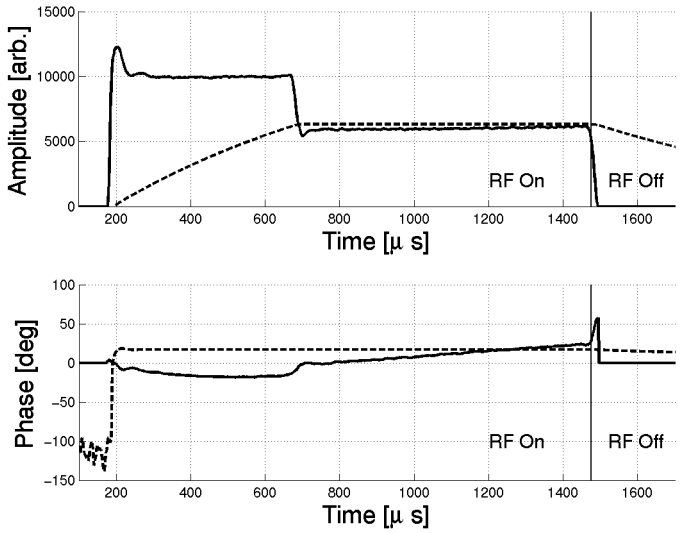

Figure 4.21.: Forward power signal (solid line) and vector-sum (dashed line) for a single pulse. The results of the loop phase and system gain determination have already been applied.

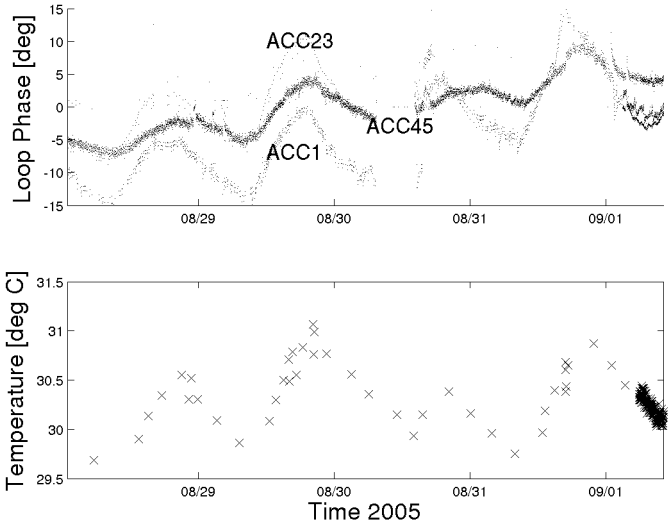

Figure 4.22.: Parallel measurement of the loop phase of the accelerating vector-sums and the temperature taken from a cable temperature sensor at the cryo hall annex or the FLASH accelerator.

Even though developed explicitly for single resonances (single cavity), the presented algorithm for the determination of loop phase and system gain has been applied successfully on vector-sums as plant. Figure 4.22 shows the measurement of the loop phase based on the presented method for a period of nearly five days. A clear day-night periodicity is visible. The peak-to-peak value of this drift is at about 30°. The drift of the loop phase can clearly be correlated with the temperature measured at a fixed location of the end of the linac thus indicating a potential candidate of the origin of the phase drift.

4.4.4. FPGA-Based Single Cavity Detuning Monitor

With the cavity equation in polar coordinates, equation 3.22, the detuning of an oscillator can be calculated as

$$\Delta\omega = \dot{\varphi} - \omega_{1/2}\frac{\rho}{r}\sin(\theta - \varphi). \tag{4.37}$$

Practically speaking, in order to derive the detuning of a cavity from RF signals, one needs to measure the amplitude and the phase of the incident wave (ρ and θ) and of the field probe (x and φ). The measurements need to be calibrated in a way that amplitudes and phases match each other. In addition, for precise measurement it is necessary to correct the signals for crosstalk. A procedure for the calibration of measured signals was developed in section 4.3.2.

Deriving the detuning from RF signals is very attractive in the face of real-time detuning compensation. The derived detuning can be used as a sensor to feedback loops that correct for the detuning during the pulse. As a first attempt, equation 4.37 was implemented as a real time

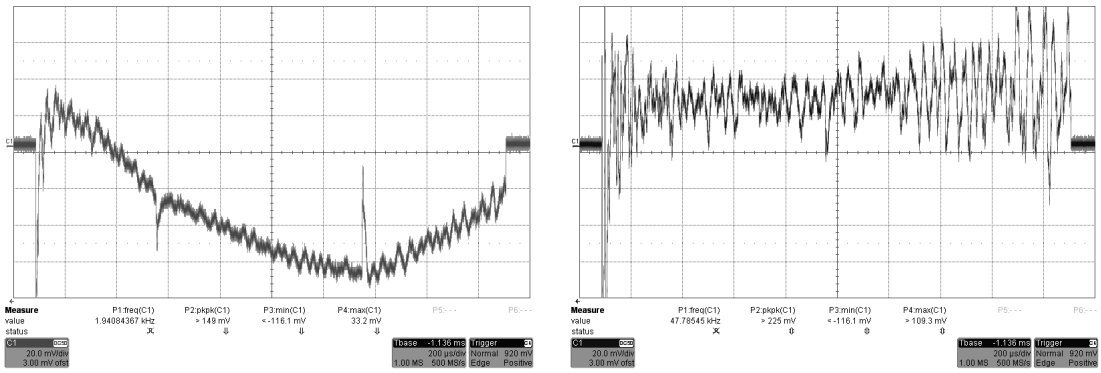

Figure 4.23.: Results of the hardware detuning-monitor obtained by a digital oscilloscope. The details on the bottom of the plots are not of importance. Both screenshots show the detuning over a full pulse. One the horizontal axis, one division corresponds to 200 μs. One vertical division is about 50 Hz detuning. The left picture was taken during operation at 19 MV/m while for the right picture, the gradient was 2 MV/m.

application in a Field Programmable Gate Array (FPGA), [23].

The calibration of the signals can be performed offline. In spite of the simple nature of equation 4.37, the implementation on an FPGA is not trivial. Besides the digital downconversion-process and the calculation of amplitude and phase, which is a challenge on its own, a number of things need to be considered. Typically in an FPGA only fixed-point calculations are available, therefore a carful choice of the number of bits in a word has to be made. The derivative of the phase is implemented by means of a digital filter as in 3.52, convoluted with an appropriate averaging filter. The lengths of the averaging filter determines the total delay of the algorithm. The division in the formula requires additional bits for precision. The implementation of the sine-function finally needs a dedicated algorithm, in this case the CORDIC-algorithm [24] has been chosen.

Figure 4.24 shows a comparison of the presented RF-method with the pulse-shortening method, where the pulse is subsequently shortened. The detuning of the cavity is determined as the slope of the phase at the end of the pulse, where no drive is present (refer to to equation 3.17). With the pulse-shortening method, detuning can be measured that is repeating in every pulse. It can not measure effects where the detuning is excited by the previous pulse. This can be the explaination for the obvious difference of the two methods in the beginning of the pulse. The agreement between the two measurements is increasing towards the end of the pulse.

Figure 4.23 shows examples of detunings obtained from the hardware detuning-monitor. It is important to notice that the measurements are made in real time with a delay of a few microseconds (varying with the length of the averaging filter). The signals are therefore available for further processing by feedback algorithms that compensate for the measured detuning. While the left part of 4.23 shows the cavity at its operating gradient of 19 MV/m, the gradient has been reduced to about 2 MV/m on the right picture. Clearly visible are the impacts of the discontinuities of the drive power on the signals (start and stop of the flat-top), where the derived detuning shows spikes. It is also visible that the noise increases strongly for low gradients, caused by the

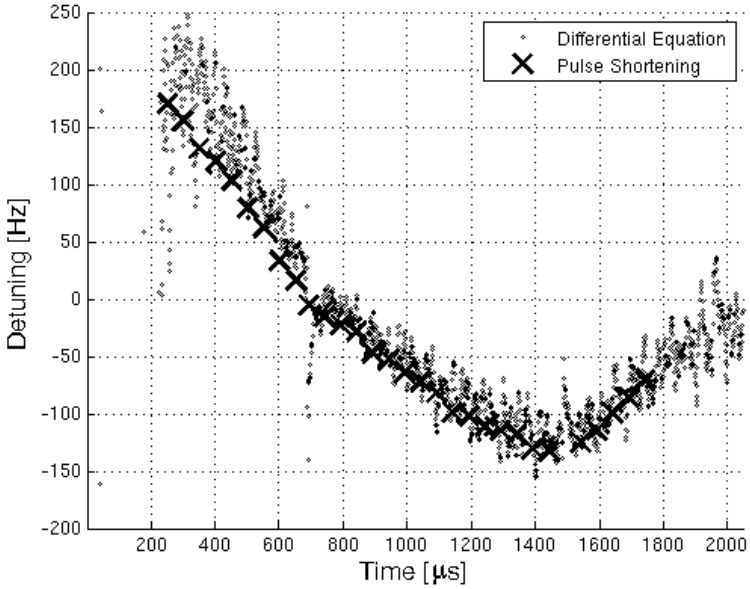

Figure 4.24.: Comparison of different methods for the detuning estimation. The crosses are determined by the pulse-shortening method while the circles are calculated from the RF signals. The comparison is only of qualitative nature as explained in the text.

derivative in equation 4.37.

4.4.5. Analytical Approach to Iterative Feedforward Generation

The concept of feedforward originates in the repetitive nature of the control problem. It is added to every pulse at the leftmost summing junction in figure 4.1. Its purpose is to compensate for errors that occur repetitively in every pulse. In that sense, the desired field shape (in time domain) is an error that demands for a certain shape of the drive signal that takes into account several, also nonlinear effects like saturation of sensors and actuators, Lorentz force detuning and the beam load. Some of these influences are undergoing changes depending on outer (e. g. temperature) or inner (e. g. desired gradient) parameters. It is desirable to have an automatic adaption of the feedforward tables.

The most obvious, self-adapting scheme is as follows (from here on it will be referred to as the naive scheme): take the drive (i. e. after the leftmost summing-junction in figure 4.1 where the current feedforward and the feedback are added) of the current pulse as feedforward of the next pulse. Practice has shown that such an approach is unstable already after a few iterations.

This section shall introduce a formalism that allows to analyze certain linear adaption schemes for pulsed systems. It explains, why the naive approach just introduced (and many of its derivatives) is instable. The next section then will introduce a method that solves the adaption problem satisfactorily using linear filters, but somehow is not covered by linear theories.

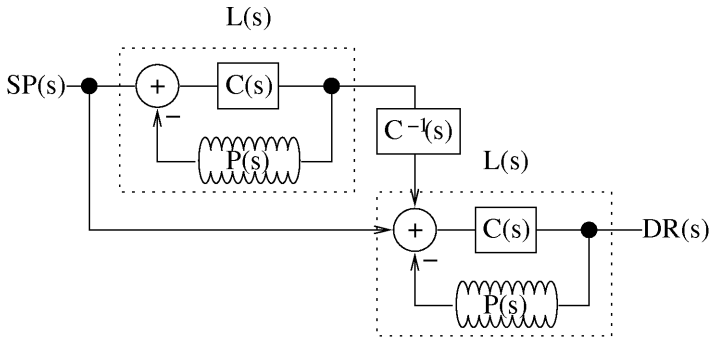

Figure 4.25.: First iteration of the naive adaptive feedforward, where the two involved pulses are depicted as disjoint systems. It is important to see here that the left system (representing the first pulse) can influence the right system (representing the second pulse) but not the other way round.

The basic idea of treating the pulsed structure in the frame of the linear control theory as for example discussed in [22] is to think of separate pulses as of separate systems. Instead of having one system exercising pulses one after the other, identical systems exercise exactly one pulse. Figure 4.25 illustrates the impact of the first iteration of the naive adaptive feedforward. First of all, two control loops are visible. They are described in the Laplace-space and contain a plant $P(s)$ (symbolized by a nine-cell cavity) and a controller $C(s)$. The transfer-function from the input of one system to its output, namely the function between the set-point table ($SP(s)$) and the total drive ($DR(s)$) is

$$L(s) = \frac{C(s)}{1 + P(s)C(s)}. \tag{4.38}$$

It is worth noting that the definition of $L(s)$ is not the most common way to define a transfer-function: usually, transfer-functions connect input (or setpoint in this example) with the output of the physical system (here the cavity field). $L(s)$, instead, communicates between the setpoint and the drive of the controlled system. The transfer-function between setpoint and cavity field would naturally be $P(s)/(1 + P(s)C(s))$.

The two systems in figure 4.25 shall be identified with two subsequent pulses of a single system. The left system is the first pulse, it does not have any feedforward applied at all. The right sytem has an extra contribution at its summing junction that originates from the left system. The path from the left to the right system has a $C^{-1}(s)$ which is ment to compensate for the controller $C(s)$ of the second system: by doing so, the introduction of a second summing junction in front of the plant $P(s)$ is avoided. The concatenation of the two system is feasible, since it is warranted that the earlier pulse influences the later one but not the other way round.

The next consequent step is to build a cascade of n such system as depicted in figure 4.26. This picture reflects the topology of figure 4.25. Figure 4.26 also indicates that parts of it shall be repeated $(n - 1)$-times. Following this picture, one can derive a linear transfer function $H(s)$ after n iterations between the setpoint table and the drive (the independend variable s is omitted

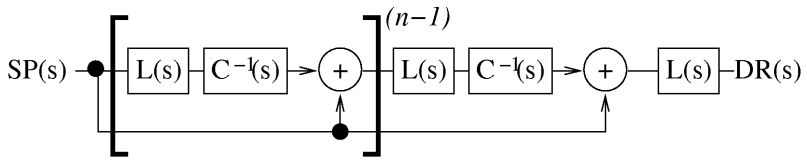

Figure 4.26.: Cascade of several iterations of the naive feedforward adaption. The bold bracket with the superior $n-1$ indicates that this part is repeated according to the number of iterations.

from here on),

$$\begin{aligned} H &= (((L \cdot C^{-1} + 1) \cdot L \cdot C^{-1} + 1) \ldots) \cdot L \\ &= [(L \cdot C^{-1})^{n-1} + (L \cdot C^{-1})^{n-2} + \ldots + 1] \cdot L. \end{aligned} \tag{4.39}$$

The last line can be identified as a geometric series,

$$\begin{aligned} H &= \frac{(L \cdot C^{-1})^n - 1}{L \cdot C^{-1} - 1} \cdot L \\ &= (1 - (L \cdot C^{-1})^n) \cdot P^{-1}, \end{aligned} \tag{4.40}$$

where the last line takes advantage of the equality $P = (1 - LC^{-1})/L$ which follows from 4.38. For reasons that will become clear in appendix B, this is reformulated to

$$H = (1 - [1 - LP]^n)P^{-1}. \tag{4.41}$$

With this, a surprisingly simple formula is presented that describes the evolution of the feedforward tables over n pulses. It has the following, natural properties. In case $n \to \infty$, the term in square brackets goes towards zero for $|1 - LP| < 1$, leading to $H \to P^{-1}$. For finite n, the difference $H - P^{-1}$ is $-[1 - LP]^n P^{-1}$, which involves the inverse of a stable transfer-function P, which has poles in the positve half-plane. This is already a severe reason, why the naive feedforward might cause problems. Figure 4.27 shows the result of five simulated iterations of the naive adaptive feedforward. The overshoot is clearly visible. Even though this scheme converges for $n \to \infty$, the steps in between show inacceptable behavior.

The amount of the overshoot can be calculated. As stated in section 4.4, it is the ratio of the coefficients of the highest degree in denominator and nominator in the transfer function that determines the height of the immediate response. This can be applied on 4.40, where $L =$

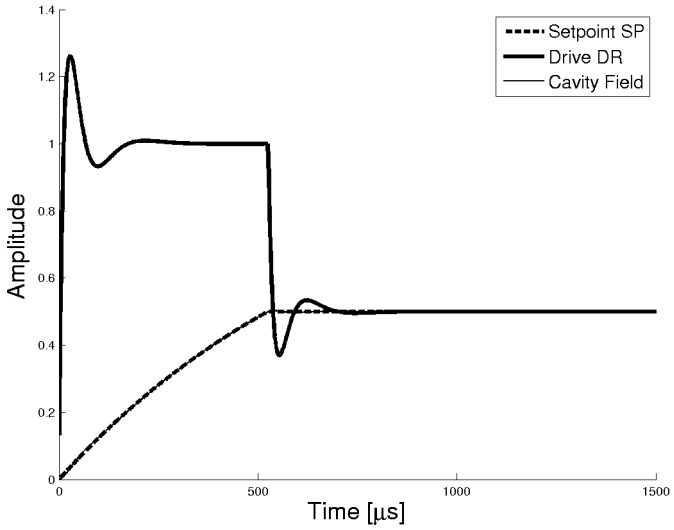

Figure 4.27.: Simulated results after five iterations of the naive adaptive feedforward. The parameters are $\omega_{1/2} = 2\pi \cdot 210\,\mathrm{Hz}$, $\Delta\omega = 0$, no beam, proportional controller with loop-gain 20. The overshoots are clearly visible and correspond to what is observed in reality.

$C/(1 + CP)$ is already replaced:

$$\begin{aligned} H &= \left(1 - \frac{1}{(1+CP)^n}\right)/P \\ &= \left(1 - \frac{1}{\left(1 + \frac{C\omega_{1/2}}{s+\omega_{1/2}}\right)^n}\right)\frac{s+\omega_{1/2}}{\omega_{1/2}} \\ &= \left(1 - \frac{(s+\omega_{1/2})^n}{[s+(C+1)\omega_{1/2}]^n}\right)\frac{s+\omega_{1/2}}{\omega_{1/2}} \\ &= \frac{[s+(C+1)\omega_{1/2}]^n - (s+\omega_{1/2})^n}{[s+(C+1)\omega_{1/2}]^n}\frac{s+\omega_{1/2}}{\omega_{1/2}} \end{aligned} \tag{4.42}$$

In the last expression, the coefficient of the highest power (n) in the denominator is clearly $\omega_{1/2}$. In the numerator, the coefficient of the power n is $nC\omega_{1/2}$ which means that the overshoot of the step-response grows linearly with the number of iterations.

A natural extension of the naive adaptive feedforward can be stated as follows. Take the drive of the current pulse and subtract the feedforward, the result will be the feedback contribution within this pulse. Apply a digital filter D on the feedback and add the filtered feedback on top of the existing feedforward table. Appendix B shows in a lengthy calculation that the transfer

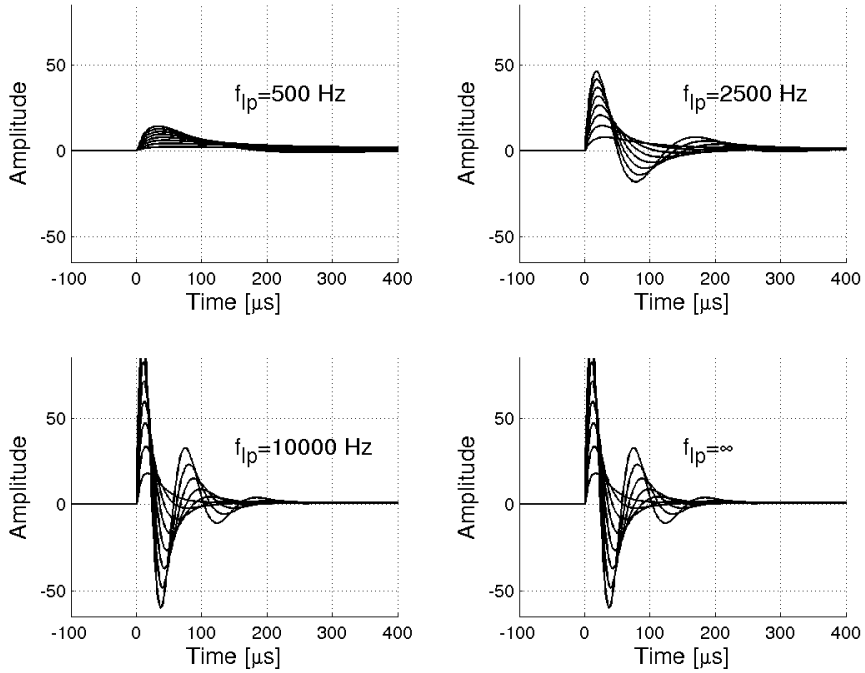

Figure 4.28.: Step-response of the transfer-function that desribes the adaptive feedforward-iterations, here up to eight iterations. The parameters are chosen for a system without beam, half-bandwidth $\omega_{1/2} = 2\pi \cdot 210\,\text{Hz}$, zero detuning, proportional loop gain 50 and different digital low-pass filters D with corner frequencies starting from 500 Hz. The step-responses with the highest amplitude are alway the ones after eight iterations.

function between the setpoint table and the drive after n iterations is given by

$$H = (1 - [1 - LDP]^n)\, P^{-1}, \tag{4.43}$$

which has a similar appearance as equation 4.41.

The transfer-function D in equation 4.43 does not suffer from a restriction that typically applies to applications of control theory: it does not have to be causal. At the time the feedforward table is calculated, the pulse is already over and the information of the whole pulse is available.

With 4.43, the step-responses for the adaptive feedforward can be calculated. This is plotted in figure 4.28 for the first eight iterations of a system without beam, half-bandwidth $\omega_{1/2} = 2\pi \cdot 210\,\text{Hz}$, zero de-tuning, proportional loop gain 50 and different digital low-pass filters D with corner frequencies starting from 500 Hz. From the step-responses it can be seen that the transfer-functions by themselves are stable in the bounded-input-bounded-output (BIBO) sense. But the amplitude is increasing linearly with the number of iterations, independed of the choice

of D. For this reason, the feedforward-algorithms built from this scheme are instable in the sense that they do converge after a large number of iterations but show high overshoots in intermediate iteration steps. The situation changes if the bandwidth of the system under control is higher: simulation show that the leading overshoot is very small.

4.4.6. Fast Adaptive Feedforward Generation with Time Reversed Filters

The section will present a very simple, efficient and stable method for the adaptive generation of feedforward tables. It is, however, not covered by linear theories. First the scheme will be introduced in a cookbook-fashion. Then, after a few examples, an attempt to a more analytic phrasing will be made.

The scheme is described best by the following steps, which are illustrated in figure 4.29. It is later on referred to as the reverse-lowpass feedforward.

1. Take the feedback-contribution of a pulse (shaded area in the figure).
2. Reverse that signal in time (solid line represents reversed signal).
3. Apply a digital lowpass-filter with the corner-frequency ω_{lp} on that signal.
4. Again reverse this filtered signal in time.
5. Time-shift this by $-\Delta t_{\mathrm{aff}}$.
6. Add the reverse-filtered, time-shifted signal on top of the current feedforward-tables. The newly added part is shaded in the figure.

This scheme can, of course, only be applied in a pulsed environment. It takes advantage of the fact that the whole pulse information is available for the feedforward calculation. It can very efficiently be calculated in digital devices, since only multiplications and additions are involved. The application of the filter is implemented by subsequent multiplications and additions, as described in section 3.8.

The reverse-lowpass feedforward scheme has two free parameters: the lowpass corner frequency ω_{lp} and the time-shift Δt_{aff}. By experience it can be stated that the best results in terms of convergence speed and stability are obtained for values where the corner frequency is of the order of the loop bandwidth and the time-shift of the order of the loop delay. A typical loop-bandwidth is 20 kHz for a cavity corner frequency of 200 Hz and a proportional gain of 100. A typical loop latency is of the order of 1 μs.

Figure 4.30 and 4.31 show examples for the performance of the algorithm for high beam-load simulation. The copper-cavity is driven with feedforward plus feedback consisting of a proportional and an integral gain, the corner frequency of the reversed-lowpass is 50 kHz. Beam is simulated by a waveform generator at an additional input in front of the amplifier that drives the cavity. The small pictures on the left show a zoom into the critical area where the simulated beam starts. The right small picture in 4.30 shows the development of the relative peak-to-peak deviation in the critical area while the number of iterations is increasing. Initially, the peak-to-peak error is 5 %. The error is only visible in the region around the beam-start. After a few ten microseconds, the beam load is fully compensated by the integrator for the rest of the pulse. The

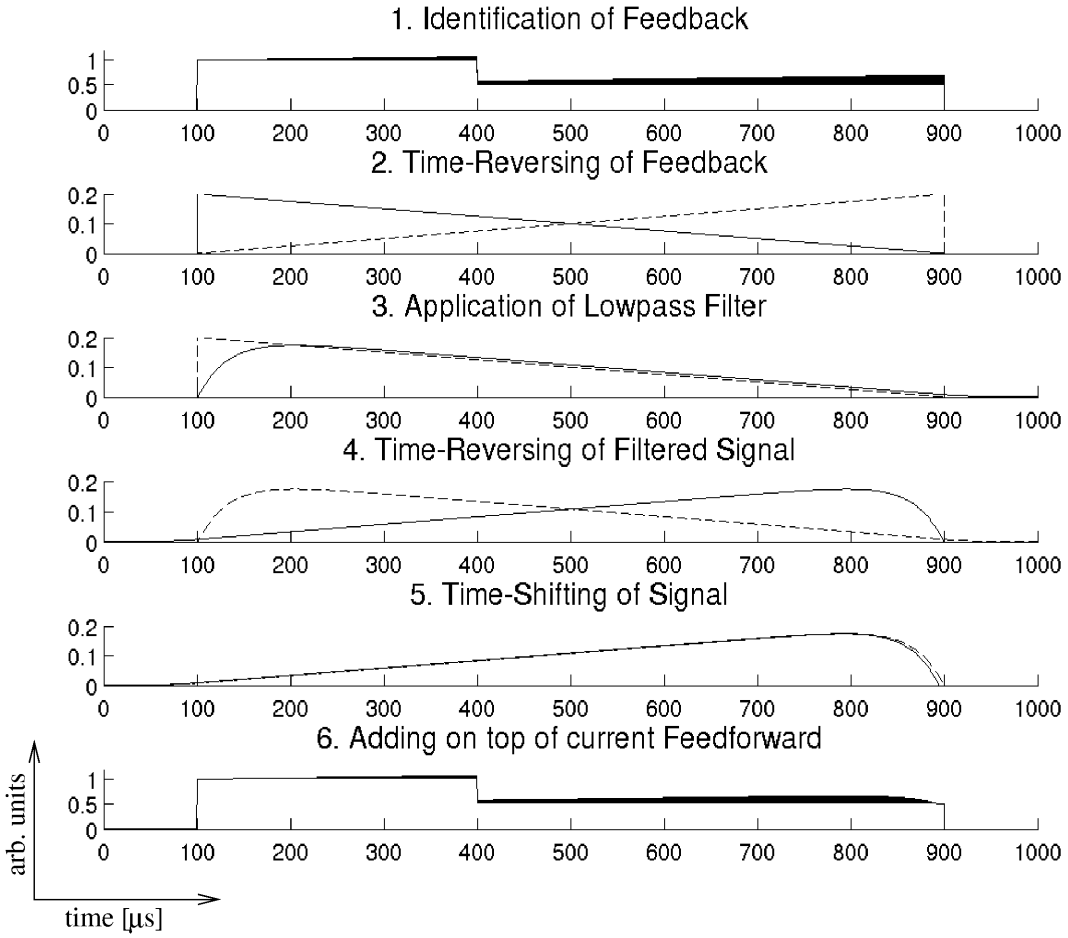

Figure 4.29.: Illustration of the reverse-lowpass feedforward scheme. The steps are explained in the text.

adaptive feedforward is able to reduce the peak-to-peak error in the critical region down to less than 1 % within 5 iterations. The feedback contribution to the drive is minimized over the whole pulse.

The pictures reveal a certain stair-step structure. The stairs are produced by the nature of the underlying controller[‡], which stores the feedforward-tables differentially. The size of the differential steps is constrained by a memory-saving scheme.

Although the method of reverse-filtered feedforward generation does not fit into linear theories, it is worth a closer look. A lowpass-filter is determined by the equation

$$\dot{x}(t) + \omega_{\mathrm{lp}} x(t) = \omega_{\mathrm{lp}} u(t) \quad \text{and} \quad sx(s) + \omega_{\mathrm{lp}} x(s) = u(s) \tag{4.44}$$

(time domain and Laplace representation), therefore the equation for the time-reversed lowpass can be obtained by substituting $t \mapsto -t$. Consequently, the transfer function for the time-reversed lowpass is just

$$H(s) = \frac{\omega_{\mathrm{lp}}}{\omega_{\mathrm{lp}} - s}. \tag{4.45}$$

The first interesting observation with this transfer function is that it is BIBO-unstable according to section 4.4.1 (the single pole lies in the positive half-plane), which would indicate that the results grow until infinity. However, the reverse lowpass filter is stable. This becomes clear from a closer look to boundary conditions of filters that are applied on waveforms with a finite length as it is the case for a pulsed LLRF system. A normal lowpass applied on a signal $u(t)$ that is defined on the finite interval $[t_{\mathrm{start}}, t_{\mathrm{stop}}]$ will yield a filtered signal $x(t)$ for $t \in [t_{\mathrm{start}}, t_{\mathrm{stop}}]$. In most implementations, the filtered signal will have zero as its initial value, $x(t_{\mathrm{start}}) = 0$ (unless other boundary conditions apply). The reverse lowpass, however, has the natural boundary condition $x(t_{\mathrm{stop}}) = 0$ which prevents exponential growth on the interval $[t_{\mathrm{start}}, t_{\mathrm{stop}}]$.

For the stability analysis of several feedforward algorithms in the previous chapter it was implicitly supposed that the filtered signals are zero for $t = t_{\mathrm{start}}$. Only for that type of initial condition it makes sense to talk about the initial height of a step response. The linear growth of the initial response with the number of iterations was the reason why all previously described algorithms failed.

4.5. Concept for a Self-Adapting Beam Based Feedforward

The previous section presented a fast and robust algorithm that shapes the feedforward table according to the demands of the beam current and other circumstances. Even though it is fast enough to adapt within 1-2 pulses, there are situations conceivable where this is not sufficient. Especially in a long linear accelerator with numerous RF stations along the linac and optics between these sections, beam transmission can only be achieved after the RF is adapted for all stations. On the other hand, adaption to a certain beam current can only be achieved if the beam is already present in a module. In a long linac, without any futher measures, adaption

[‡]The experiment was performed with the current controller of the Spallation Neutron Source (SNS) accelerator at the Oak Ridge National Laboratory (ORNL).

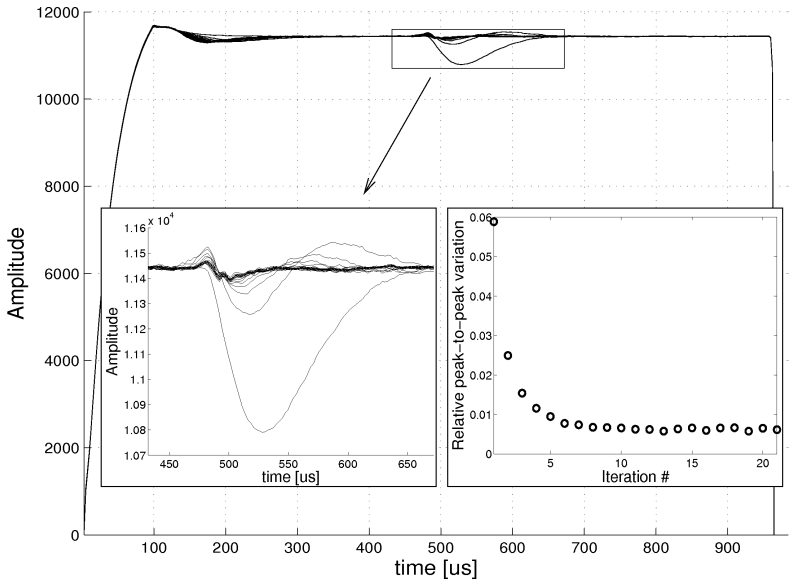

Figure 4.30.: Convergence of the cavity field with the reverse-lowpass feedforward scheme in the presence of a high beam load at a copper cavity. The large picture shows the field-amplitudes of several iterations of the adaptive feedforward. The left small picture shows a zoom into the critical region, the right small picture shows the relative peak-to-peak variation in the zoomed area after each iteration.

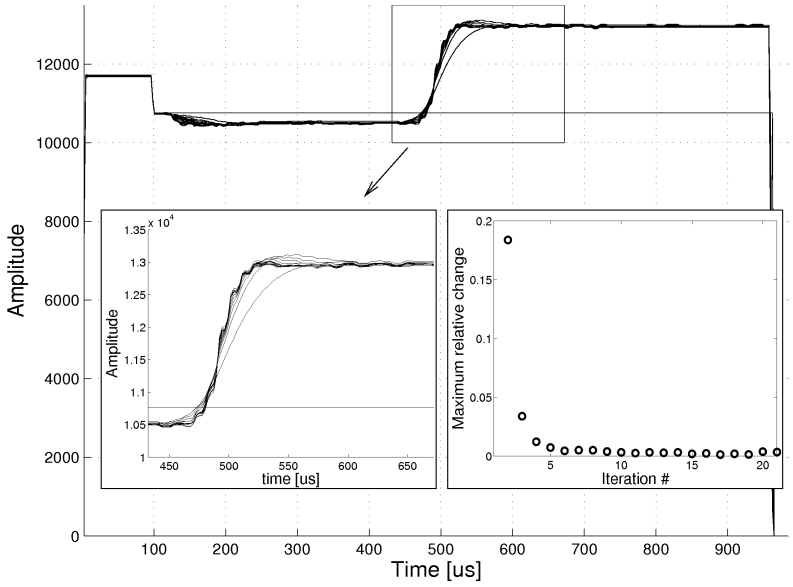

Figure 4.31.: Convergence of the feedforward with the reverse-lowpass feedforward scheme in the presence of a high beam load at a copper cavity. The large picture shows the feedforward-amplitudes of several iterations of the adaptive feedforward. The left small picture shows a zoom into the critical region, the right small picture shows the maximum relative variation in the zoomed area from one iteration to the next.

would successively be accomplished one module after the other and therefore might take long depending on the number of RF-stations. This section will introduce an extension of the adaptive feedforward scheme that takes this into account.

A first and basic concept refers to changes in the setpoint amplitude or the phase by the operator (automated or human). If any of these change, the feedforward table needs to change, too. Instead of having to adapt tables to a new setpoint, it is much faster to scale the existing table linear with the change of the setpoint. Therefore, it is helpful to have a reference feedforward table (i. e. a normalized tabel) in the memory of the signal processor that is scaled by the setpoint amplitude and phase before it is applied to the system. The algorithm that adapts the feedforward between pulses acts rather on the reference table than on the full-scaled table that is put on the DAC.

Adaption to different beam-currents and bunch-lengths can not be performed by linear scaling. Instead, the basic idea is to define a finite set of N scenarios. A scenario is uniquely defined by, for example,

- the targeted beam current,
- the number of bunches in a pulse and
- the desired pulse shape.

This is information that is known before the pulse happens. It should be communicated to the FPGA prior to the pulse. Ideally, this information is encoded in the signals from the timing distribution system. The signal processor holds as many reference feedforward tables as there are scenarios. The feedforward that is applied in a pulse is taken from the unique reference feedforward table that corresponds to that scenario. After a pulse, the feedforward adaption algorithm is applied and the result alters the feedforward table of the regarding scenario.

There are cases, in which adaption is unwanted, especially if interlocks were triggered. Therefore, the signal processor needs an additional input logic which decides whether adaption is wanted or not.

After a while, reference-feedforward tables will have built up for all scenarios. Once they are built up, beam transmission will be granted even in the case that the beam load or the pulse length changes.

5. A Framework Based on Finite State Machines for LLRF Automation

5.1. Finite State Machines

An automation-system has to act in a similar way as an experienced operator. At any point in time, it has to make decisions that are based on the information that was obtained until that moment. An automation system can thus be a decision table: available data is tested against a predefined set of conditions. Every possible combination of true and false conditions corresponds to an action performed by the automation. However, the possible combinations scale exponentially with the number of conditions which makes the development of decision tables for the automation of complex system impractical. Finite state machines (FSM) provide a way to solve the complexity without loosening the demands for a strict determinism. State-charts allow a simple visualization of systems that would otherwise be represented by complex decision tables.

FSMs are commonly used in industry and thus supported by various platforms and languages ranging from real-time FPGA systems to high level applications as in the simulation of a character in a computer game. Various tools and patterns exist that allow a formal analysis of FSMs, including tests if all states can be reached or an FSM is complete in the face of all possible inputs. Thus FSMs combine the advantages of a strictly formal system with a descriptive representation. This section gives a formal introduction to FSMs for automation and its visualization by state-charts. The next section introduces the concept of quantization that is needed in order to manage systems with a vast number of input channels and is closed with a formal definition that demonstrates the equivalence to the original FSM concept.

A very general, formal definition can be given for FSMs, [25]: a finite state machine $A = (Q, \Sigma, \Gamma, \delta, \lambda, q_0)$ is a six-tuple where Q is a finite nonempty set of states, Σ is a finite nonempty set of inputs (input alphabet), Γ is a finite nonempty set of outputs (output alphabet), $\delta : Q \times \Sigma \to Q$ is the transition function mapping, $\lambda : Q \to \Gamma$ in a Moor FSM [26] and $\lambda : Q \times \Sigma \to \Gamma$ in a Mealy FSM [27] is the output mapping and $q_0 \in Q$ is an initial state.* Practically, an FSM can be considered as a mapping between discrete functions, in the same way that a functional (e. g.

*A few variations of this definition exist in literature. Sometimes, the output alphabet Γ is omitted and replaced by a set of final states, F. These FSMs are called acceptor FSM, while the definition used in the text referres to a transducer FSM. The former is rather used in text processing than in automation applications.

$\mathrm{d}/\mathrm{d}t$) maps between continuous functions. The following to mappings shall serve as examples:

$$\begin{array}{lll} \frac{\mathrm{d}}{\mathrm{d}t} & : & \sin \mapsto \cos \qquad\qquad\qquad\qquad\qquad \text{(mapping of continuous functions)} \\ G & : & (u,u,d,u,u,\ldots) \mapsto (2,3,2,3,4,\ldots) \qquad \text{(mapping of discrete mappings)} \end{array}$$

While sin is a continuous function $\mathbb{R} \to \mathbb{R}$, the ordered set $(u,u,d,u,u,\ldots)$ can be considered as a discrete mapping $\mathbb{N} \to \{u,d\}$. Similarly, cos maps $\mathbb{R} \to \mathbb{R}$ and $(2,3,2,3,4,\ldots)$ maps $\mathbb{N} \to \{1,2,3,4,5\}$. The mapping G is therefore an FSM and can be visualized by a state-chart (figure 5.1). A state-chart [28] is the visualization of an FSM by means of a directed graph. The

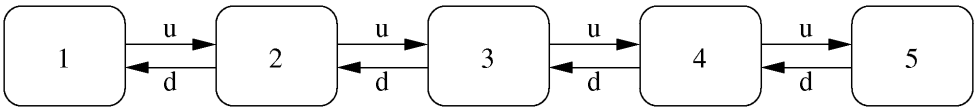

Figure 5.1.: Simple FSM that translates sequences built from a finite input alphabet $\{u,d\}$ into sequences of the same length built from a finite output alphabet $\{1,2,3,4,5\}$. The FSM can be considered as a model for a five-gear half-automatic gear-box that has two input signals (u: gear up and d: gear down).

visualized FSM G can be a model for the behavior of a gear-box, where the driver has two buttons (gear-up and gear-down) and can only chose one gear after the other. The input-sequence of the FSM G is therefore a sequence of us and ds, while the output-sequences (sequence of states) is the sequences of selected gears 1-5. The sequence of states is, for this simple model, identical with the output sequence.

Usually, the concept of FSMs is extended by a few features.

1. Conditions. In the FSM G, the appeareance of u in an input-sequence causes a transition if the FSM is in one of the states 1-4. The concept of conditions allows to connect a transition with a required value of a secondary input sequence (or time-dependend function).
2. Parallel flows. One can imagine to have an additional flow as the one depicted in figure 5.1 in parallel. This is equivalent to having a secondary output sequence.
3. Superstates. A superstate is a state that contains other states. The states inside a superstate form in fact an FSM by their own. The purpose of superstates is mainly to make state chart diagrams more readable by allowing the designer of FSMs to introduce different levels of abstraction. If and only if a superstate is active, the contained flows are active.
4. History states. The whole FSM has an initial state (equivalent to an initial condition of a differential equation). Additionally, sub-states of a superstate can be made so-called history states. That means, on re-entry of a superstate, the sub-state that was active last will be the initially active state.

For Moor FSMs, the output sequence is obtained from a unique mapping from the current state to the output alphabet (which is not necessarily identical to the set of states). For Mealy FSMs, the output sequence is obtained from a unique mapping from the last transition to the output alphabet. As there are typically much more transitions than states, a Mealy FSM is more

complex than a Moor FSM. However, it can be shown that for each Mealy FSM an equivalent Moor FSM can be constructed.

So far, an FSM is a language for modeling systems with discrete in- and outputs in the same way that transfer functions are a language for modeling linear systems with continuous in- and outputs. A state-chart (as in figure 5.1) is a graphical way to represent FSMs and their behavior. Another way to represent FSMs are transition tables.

The DOOCS control system [29], offers a basic implementation of FSMs [30]. A DOOCS FSM is event-driven, which means that the elements of the input-sequence (Σ in the formal definition) are not restricted by periodic time-patterns. However, the DOOCS FSM framework foresees a timing event which occurs synchronously to the accelerators timing system on every pulse of machine operation. The concept of output-sequences (Γ in the formal definition) is modified in the DOOCS implementation in the sense that functions are invoked in certain situations: states do have *enter()*-, *during()*- and *exit()*-functions that are invoked on entry into a state, while the state is active (with the periodicity of the pulsed structure of the accelerators timing) and on exit. Additionally, transitions offer functions that are invoked if this transition happens. The programmer has the freedom of choice to fill these function-templates. Obviously, if the programmer uses only the function templates that are connected to transitions, he followes the concept of a Mealy FSM. In a Moor FSM, the programmer would rather use the *enter()*-functions. It is possible to mix between these two concept with the DOOCS FSM implementation.

5.2. Discretization of Signals

The analogy between transfer-functions and FSMs is continued with their applications. While transfer-functions can be used to represent the controller in a continuous closed-loop system, FSMs can be used for modeling the controller in a system with discrete in- and output data. Precondition for the usage of an FSM as a controller is the presence of a discrete data flow of the system under control (plant). One needs to distinguish between value-discrete (value-continuous) and time-discrete (time-continuous) data. A channel is value-dicrete, if the possible set of values is finite. If it is infinite in the sense that there is an infinite number of possible values between[†] two values, it is value-continuous. It is time-discrete, if the value is constant for periodic intervals in time while it is time-continuous, if the value can change at any time. Transfer-functions are the appropriate choice for systems that are value-continuous and time-continuous (Laplace-transform) or value-continuous and time-discrete (Z-transform). The fact that digital systems operate at a limited digital resolution (e. g. 14 bit) is usually not considered as a value-discretization of the system unless it affects the precision of calculations.

Real systems as the LLRF system at FLASH usually have discrete values (e. g. interlock status, feedback status) as well as continuous values (e. g. setpoint, phase). Additionally, the number of channels is very high. In order to automate such a system with an FSM, the signal spaces need to be partitioned. An example for a partitioning of a continuous value-space is to measure the level of water in a bottle with discrete values like “empty”, “half-full” and “full” instead of a precise

[†]The usage of the term “between” implies that an order or “greater than” relation of the values exists.

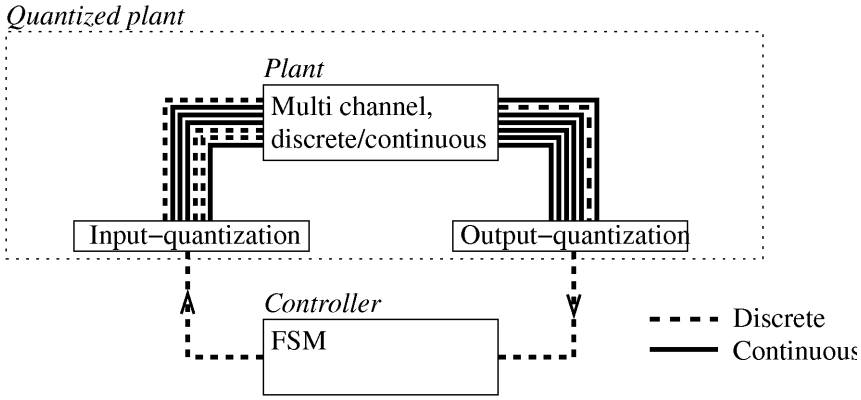

Figure 5.2.: Automation of a plant with continuous and discrete signal spaces with an FSM and a quantization. The system under control is complicated due to its large number of either discrete or continuous in- and output-channels. With the quantization in place, it appears to the outside world as an FSM itself (quantized plant, dotted line).

measure in milliliter.

The concept of an FSM-based automation of a plant with discrete and continuous signals involving a quantization is depicted in figure 5.2. The complexity of the system is masked by the in- and output quantization which are the only access-points to the system for the automation. The system in the dotted borders itself acts as a quantized sytem. Techniques exist that allow to construct an FSM that models its behavior [31].

FSMs are supposed to react on events. For the LLRF automation, an event can be defined as a significant change in a variable. Since the DOOCS control system provides access to its variables via a pull-mechanism, it is rather inefficient to react on a change of a variable since it involves frequent pulling and comparison of that variable. In particular reactions on variables that are derived from many raw variables of the DOOCS system produce a large network load. For automation purposes, it would be preferable to have the quantized signals available immediately in the FLASH environment via push-mechanisms. That is, that the producer of a value in the DOOCS system is capable of detecting changes and notifying registered clients. However, this would require changes of the DOOCS control system. Presently, a centralized data acquisition system is developed that will be able to solve this problem in the future. The central data acquisition system, [32], would be a good candidate for the preparation of the quantized data in the future of DOOCS. The quantization-scheme developed within this thesis considers the fact that the number of channels in the LLRF system is very high, that the data is not available centrally and that the computing and networking capabilities are limited.

The developed framework introduces procedure servers which hold a number of procedures. The procedure server reacts on a discrete input (comparable to a push of a button) with the invocation of a procedure. Only a few restrictions apply on these procedures:

- If invoked, the procedure is unaware of previous invocations of itself or other procedures. This simplifies the procedures. Additionally, it requires the procedure to check its require-

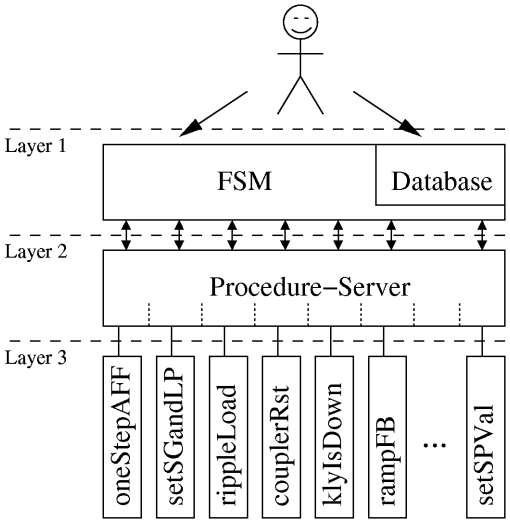

Figure 5.3.: Three-layer display of the automation framework presented in the text.

ments on each invocation with regard to machine protection and operability.

- While running, the procedure does not interact with the operator. This simplifies the interface of the procedure and allows a sequential execution of the tasks inside a procedure.
- The procedure holds all the knowledge necessary to solve its intended task. Additional knowledge may come from a database or a configuration file. With this, a separation between control strategy (implemended in the controller FSM) and problem solving (implemented in procedures) is achieved.
- Within a reasonable amount of time (1-100 pulses of the accelerator) the procedure has to terminate. Special purpose procedures are imaginable that might take longer for execution. The procedure server foresees a mechanism to cancel procedures that overexceed a defined time-limit.
- Before termination, a discrete return-value indicating the success of the procedure is handed over to the procedure server. In addition, the procedure is supposed to produce a human-readable result that can be stored in log-files.

A descriptive term for these procedures is “fire-and-forget”-procedure. Figure 5.3 shows the procedure and the procedure-server in the context of an automation with an FSM. In the developed framework, the discrete results of the procedures are the sole quantized data that represent the plant. The invocation of a procedure leads to a discrete result that is determined by the state of the plant. One can say that one retrieves partial information of the state of the plant by invoking appropriate procedures.

It is obvious that the procedure server along with its procedures fulfill the functionality of the input-quantization in figure 5.2. A push of a button of the procedure server is a discrete signal that is translated into a sequence of different continuous or discrete signals handled by the procedure. On the other hand, it is not quite obvious to see that the procedure server can act as the

output quantization. Obviously, the procedure server can provide the current discrete information only on request, namely after a button has been pushed. Additionally, the information about the machine is always incomplete. Only what is defined in a certain procedure is revealed after the procedure has finished.

In order to see that this is (nearly) equivalent with the situation where the quantized state is directly accessible from the output (that is, $\lambda : Q \to \Gamma$ is unique), the presented concept will be pointed out with a formal definition.

A quantized plant where the states are directly visible from the output shall be given by

$$\begin{aligned} A_{\mathrm{qp}} &= (Q, \Sigma, \Gamma, \delta, \lambda, q_0) \\ Q &= Q_1 \times Q_2 \times \ldots \times Q_N \\ \lambda &: Q \to \Gamma \text{ unique.} \end{aligned} \tag{5.1}$$

This is just a standard FSM following the formal definition with a multi-dimensional state-space. Now, an FSM $\bar{A}_{\mathrm{qp}}$ shall be constructed from the definition of A_{qp}, that has the properties of a plant where the states are only indirectly visible over the detour of the introduced concept of a procedure server and a finite number of procedures. The FSM $\bar{A}_{\mathrm{qp}} = (\overline{Q}, \overline{\Sigma}, \overline{\Gamma}, \overline{\delta}, \overline{\lambda}, \overline{q})$ shall be introduced step by step by the elements of its six-tuple.

First of all, the set of states of $\overline{A}_{\mathrm{qp}}$ is identical to the set of states of A_{qp}.

$$\overline{Q} = Q \tag{5.2}$$

The set of inputs of A_{qp} shall be extended by a dimension that corresponds to the finite number of N prodedures that are available through the procedure server. It is assumed, that N also is the number of dimensions of the state space Q.

$$\overline{\Sigma} = \Sigma \times \Sigma_Q, \quad \Sigma_Q = \{1, \ldots, N\} \tag{5.3}$$

The transition function $\overline{\delta}$ shall be identical to the transition function δ. It simply ignores the input elements from the additional dimension Σ_Q

$$\begin{aligned} \overline{\delta} &= \overline{\Sigma} \times \overline{Q} = \Sigma \times \Sigma_Q \times Q \to Q = \overline{Q} \\ &\quad (\sigma, j, q) \mapsto \overline{\delta}(\sigma, j, q) = \delta(\sigma, q) \end{aligned} \tag{5.4}$$

The output alphabet $\overline{\Gamma}$ differs from Γ. The output alphabet is no longer identical to the state space Q. Instead, it is identical to the state space but includes a $\emptyset$ to each dimension.

$$\overline{\Gamma} = \{\emptyset, Q_1\} \times \ldots \times \{\emptyset, Q_N\} \tag{5.5}$$

The output mapping now is constructed such that only the state of the jth dimension of the

state-space, that was previously addressed at the input, is visible at the output.

$$\begin{aligned}\overline{\lambda} \quad : \quad & \overline{\Sigma} \times \overline{Q} = \Sigma \times \Sigma_Q \times Q \to \overline{\Gamma}, \\ & (\sigma, j, q) \mapsto \overline{\lambda}(\sigma, j, q) = (\emptyset, \ldots, \emptyset, q_j, \emptyset, \ldots, \emptyset), \\ & \text{where } q = (q_1, \ldots, q_N)\end{aligned} \tag{5.6}$$

With that definition, the formal description of an output quantization that uses the procedure server concept has been given. While the FSM A_{qp} is a Moor-FSM, $\overline{A}_{\text{qp}}$ is a Mealy FSM.

Disadvantages of this concept arise if the states of the quantized plant change much faster than the elements of the input-sequence arrive. Even though a concept that needs certain inputs in order to get information about the state of the quantized plant may appear unnatural, there exists information in the LLRF plant that can only be accessed via invasive procedures. An example of those "hidden states" is the DC-offset of the vector-modulator. In order to measure this using the DSP one needs to open the loop and scan different power levels. This makes this property of the plant inaccessible during normal operation.

The procedures in the presented concept fulfill a two edged role. On the one hand, they are used in order to produce transitions in the quantized plant, on the other hand, they are supposed to deliver information about the state of the plant.

5.3. Requirements of an Automation System for LLRF Control

The analogy between transfer functions and FSMs ends when it comes to the design of an optimal FSM for automation. An optimal controller can be designed on the basis of a set of numbers describing its performance goals. Once in operation, the controller can be tested against its goals. In [22] numerous standard-methods are described that can be used for controller-design.

The design goals towards an automation-system are rather weak compared to the goals in the classic control problem and are for example:

- Ease the operators job or even protect the machine from damage caused by operator.
- Prevent from faults by early detection of problems. Many of the known faults can be foreseen already by a careful online-analysis of the LLRF data, as for example the quenching of the cavities.
- Increase up-time by fast fault recovery. Simple known faults that require only the reset of a subsytem can be recovered automatically much faster than a human operator could do.
- Increase field-stability by continuous optimization of parameters (slow-feedback).
- Offer interface to standard-tasks that need to be performed on request. Even though not invoked autonomously, many applications can be integrated into the automation system and presented to the operator though a common interface.
- Deal with several ways of bypassing the automation system. In this point, the automation for LLRF control differs from most industrial automation systems, where the machines are

accessible only via the automation system.

Besides the goals of automation, the location where the automation takes places is worth a discussion. Many features that are already in the DSP system of the FLASH accelerator can be considered automation, e.g. the quench detection. Also the interlock system (machine protection) is part of the automation. The automation discussed in this thesis is placed higher in hierarchie at a level on a par with the operator himself.

Further, it needs to be decided which parts of the automation shall be invoked completely autonomously and which on request. This leads to the question, which tasks shall be integrated in an automation system and which not. Many tasks need to be performed only during the commissioning phase. Others are only accomplished rarely by experts who use their own programs.

And finally, it has to be considered, that an accelerator is an evolving system where subsystems are changed on a regular basis. This is accompanied with the fact that expert operators tend to bypass the automation system by directly accessing the machine. In industrial automation systems it is common that every access to the machine goes through the automation interface.

5.4. Outline of a Finite State Machine for LLRF Automation

From these considerations, two main guidelines for the automation of the FLASH LLRF system where derived. The first is the fundamental set of states of the automation FSM itself, the second is the premise under which the DOOCS framework for FSM generation has been extended, namely a very high flexibility.

The structure of the FSM follows the behavior of an experienced operator. Figure 5.4 shows the top level of the FSM. One can see two parallel flows. The flow titled operation flow has as a major element a sequential chain that leads from a null state to the operate state. In the null state, the status of the LLRF system is completely unknown while in the operate super-state the machine is operated normally. In between these two super-states are super-states in which preparations for the operation are performed that are best done with the klystron being in a secure high-voltage mode (secure super-state) and those that can only be done when the klystron is at its full high-voltage setting (full super-state). The tweak super-state is a state, where the RF system is principally running but non-invasive parameter fine-tuning is done. Inside the super-states of this chain is usually a sequence that fulfills the goal of the super-state. In case a procedure reports an error, the error state is approached from any state. Depending on the severity of the error, it is foreseen to jump back to the sequential preparation chain. In case of severe errors, a jump back to secure is probable while in less severe cases the operate super-state is appropriate.

The next element in the operation flow is a number of applications that are only accessible if the system is in operate state. These are invasive parameter optimizations as the transient based DSP calibration or simply a routine for controlled access to the tunnel which requires a controlled ramp down of all RF systems. The state connected with the controlled acces is abbreviated as zz in figure 5.4.

The parallel flow titled observer flow holds a super-state that itself holds many flows as the one depicted in figure 5.5. Usually, the exception-flow oscillates periodically between two states: being idle and monitoring. While the idle state is nothing but a pause for a certain number of

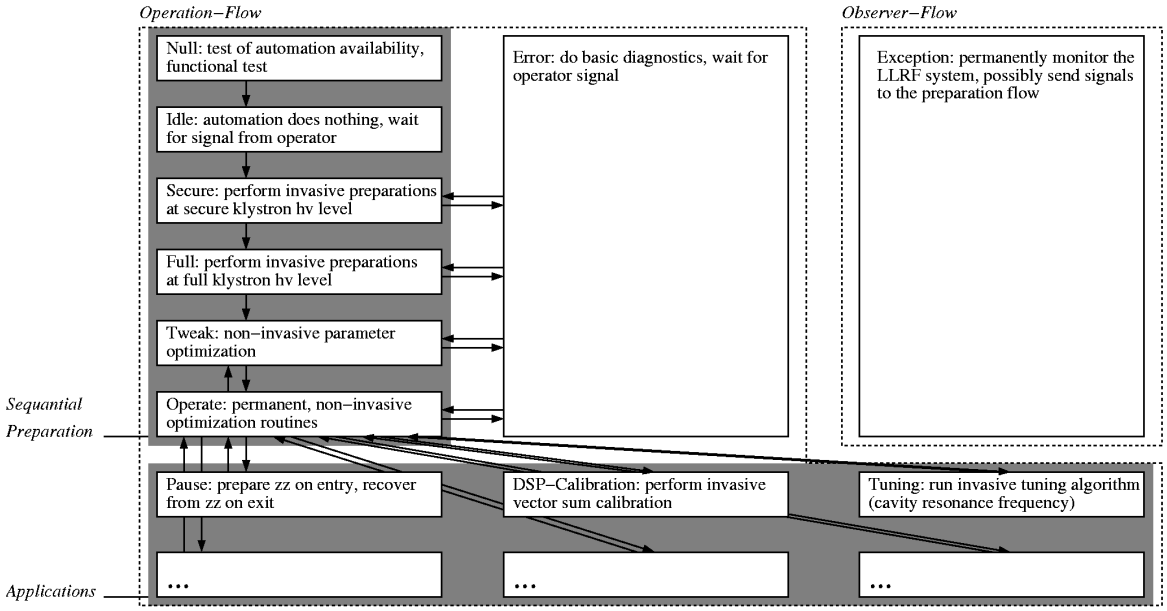

Figure 5.4.: Outline of the implemented FSM. The four main-components are the sequential operation flow, a number of special-purpose applications, an error-state and an exception and monitoring flow.

pulses, the monitor state is foreseen to execute a procedure that monitors a certain aspect of the accelerator that can be obtained parasitically. In case the monitor state encounters something suspicious, it jumps to the error flow which can be configured such that an event is sent to the operation flow, causing it to jump back in the preparation chain or go to an error. The exception flows can be used for many occasions. They can be configured for oscillating exclusively between the idle and the monitor state or they can perform branched logics by using also the remaining states in figure 5.5.

Apart from this design of the states, the flexibility of the design of the automation system seems to be very important. Therefore, the DOOCS FSM framework has been extended by a few features. They cover the on-line re-configurability of the system as well as the off-line re-structurability. Reconfiguring the FSM means to either change the procedure that is invoked on entry into a state or change the conditions under which transitions are initiated. Re-configurability could also mean a change in the structure of states. The techniques for re-configuration and re-structuring are described in detail in [33].

An important experience with the high degree of flexibility of the presented approach is summarized as: The more flexible an automation scheme is, the faster it will converge towards a stationary (no further changing) solution.

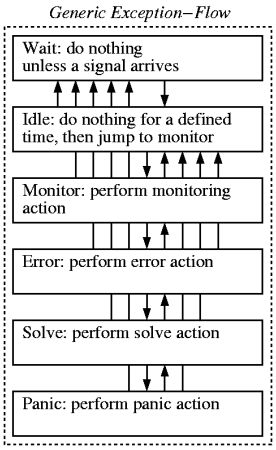

Figure 5.5.: Single exception flow. In the exception super-state in figure 5.4, numerous parallel exception flows as shown here are active. They can be freely configured by an FSM-expert.

6. Stateless Procedures for LLRF Automation

6.1. Overview

The collection of procedures which constitute the quantization of the LLRF system is not unique. A good choice requires intuition and permanent monitoring and revision. A few guidelines for a good design can be stated:

1. A procedure should check fulfillment of its requirements each time it is invoked before performing any action.
2. While running, a procedure should continue checking its requirements.
3. A procedure should monitor all relevant variables while running in order to detect interference by the operator or by other procedures.
4. Procedures working together on the same set of variables need to avoid conflicts by themselves. This can be ensured by a semaphore mechanism, which is not (yet) part of the procedure server.
5. A procedure combines a number of actions which are performed one after the other, therefore it is a linear sequence. Before each of its actions, the procedure should check the requirements of this particular step and quit and return with a unique error code in case the requirements are not fulfilled.

The procedures are a collection of ready made tools to be used by a proficient operator for solving standard tasks in operating and maintaining the LLRF. The standardized interface of the procedure server allows this collection to be accessed by a finite state machine.

Within this thesis, a number of procedures have been developed and are briefly summarized in table 6.1. The table gives a short summary of the procedures as well as the context in which the procedure is used within the FSM. The stated context of each procedure corresponds directly to either a state or to a flow of the FSM in figure 5.4. Selected algorithms are described in more depth in the following sections.

6.2. Loop Phase and System Gain Control

Description of the Procedure

The purpose of this algorithm is to measure the loop phase and the system gain of the control loop using the equations developed in section 4.4.3. A configuration flag determines whether the

Name	Description	Context
offCal	calibration of actuator offsets in I and Q	invasive low power preparation during start-up
setStepAFF	measure and/or correct system gain	tweaking, noticeable only if large deviations need to be corrected
oneStepAFF	performs one iteration of adaptive feedforward	tweaking
defaultFF	restores original feedforward table	tweaking
couplerRst	check and reset coupler interlocks	exception monitoring and handling
fieldQuality	check field quality, return discrete value (category)	exception monitoring
checkNetwork	monitor network availability	exception monitoring
rampFB	ramp feedback to desired target value	tweaking
operatorAction	monitor operator actions on target and setpoint values	exception monitoring
cathLaser	monitor operator actions on charge (cathode laser settings)	exception monitoring
getDSPCalData	retrieve DSP calibration data and store in file	application
evalDSPCalData	evaluate DSP calibration of forward and reflected power and store in DB	application
signalCalilb	calculate calibration of forward and reflected power and store in database	exception monitoring
klyIsDown	check if klystron is alive	exception monitoring
findFancyPulses	search for strange pulse shapes	exception monitoring
rippleLoad	apply ripple compensation table	application
gunCalib	apply RF-gun calibration algorithm	application

Table 6.1.: Available procedures with description and context in the automation

algorithm should just monitor loop phase and system gain or if it should also correct for it. As extra input-parameters that can be changed from FSM-experts, upper and lower bounds for the allowed system gain as well as a threshold for the operating gradient above which the algorithm is inactive can be adjusted. Once the algorithm is invoked it collects all necessary data and performs a number of consistency checks. If any of these checks fails, the procedure terminates with a negative return-code without changing anything in the LLRF-system. The consistency checks incorporate

- the expert flag that allows or forbids intervention on loop phase and system gain through the automation,
- the feedback and feedforward switches at the DSP,
- timing consistency be demanding for a flat set-point during the flat-top time that was calculated from the timing,
- a constant amplitude and phase setpoint for all pulses that are recorded and used for averaging,
- RMS requirements in flattop and filling to avoid fluctuations due to noisy data, which occurs especially at low gradient set-points and
- an acceptance range for the measured half-bandwidth of the vector-sum.

If all points are fulfilled, the algorithm starts calculating the loop phase and the system gain. In case the loop-phase differs from zero by more than one degree or the system gain differs from one by more than five percent, a correction is applied at the DSP.

Usage of the Procedure in the FSM-Context

The algorithm is currently used in the exception-block of the FSM. This block is foreseen for periodical execution of algorithms. One of the structures as in figure 5.5 is configured such that the loop phase and system gain are measured parasitically in the monitoring state. If the return code indicates that a correction is necessary, the correction algorithm is invoked in the error state of the exception-block.

6.3. Adaptive Feedforward Generation

Description of the Procedure

The procedure `oneStepAFF` follows directly the scheme that is developed in section 4.4.6. The invocation of `oneStepAFF` performs one iteration of the fast adaptive feedforward generation with time reversed filters.

Directly after the invocation, the procedure checks if feedback and feedforward are enabled at the DSP and if a dedicated user-flag is set that allows the algorithm to operate. An additional element of safety is that the user can define a lower limit of the setpoint amplitude below which the algorithm is inactive. After these checks are passed, data is collected that is used for the calculation of the new feedforward tables. After the data is collected, it is checked if the beam

inhibit system (BIS) was active during the collection of data. In case the BIS system reported an incident, the feedforward generation stopps.

The calculation of the feedforward tables is then based on the collected data. It is followed by a scaling of the resulting new feedforward table, for the following reason: the DSP does not store the feedforward table as it is written out to the DAC. Instead, it keeps a normalized table that is scaled with the setpoint value (amplitude and phase) before it is written to the DAC. This leads to a linear scaling of the total feedforward tables in the case that the setpoint changes. The linear scaling was described as a part of the self-adapting beam-based feedforward in section 4.5.

Usage of the Procedure in the FSM-Context

First, the procedure is used in the tweak superstate. It is embedded in a sequence of algorithms that work together. Second, it is referenced to in the operate state itself. The periodic invocation of the algorithm does not disturb operation. In case the feedforward table is already optimal, additional invocation will not be noticed by the system.

6.4. Tracking of Operator Interference

Description of the Procedure

An explicit design goal of the LLRF automation is the ability to detect and handle interventions of the operator that do not pass the automation-interface. As an example, the operator could change gradient and phase set-points while these properties are under the control of the automation system. For this reason, an algorithm `operatorAction` that is reading numerous variables of the control system and compares them a) to the desired target value and b) to an internal tracking variable has been developed. The internal tracking variable allows the FSM to distinguish between changes of the variables due to operator interference and changes caused by the FSM. Each change to the control system variables performed by the FSM is applied similarly to the tracking variables (this is a design rule and has to be followed by each algorithm that acts on the control system variables). Based on the configuration found in the control system, different actions have to be considered.

In addition to the tracking and the target variable, bool variables were introduced that determine the behaviour of the algorithm. “Auto-Mode” determines if a variable is under the control of the FSM. “Overwrite-Mode” indicates the behaviour of the algorithm in case an interference from the operator is detected. Table 6.2 summarizes all variables that are relevant to the algorithm and gives information on their type. Each cell corresponds to a variable. The first row indicates the variables that are held by the DSP, namely amplitude and phase setpoints and settings for feedback and feedforward. The next row is a set of variables with a one to one correspondence to these variables. These are the target variables. The row labeled tracking is one more set of variables that corresponds to the first row. Here, the automation tracks the changes applied by the state machine. Operator interference can be detected by comparing the DSP variables and the tracking variables. The row auto-mode shows the bool variables that were introduced in order to disable or enable automatic control of individual variables. The last row finally is a switch

	Gradient	Phase	FF-Mode	FB-Mode	FB-Gain
DSP	float	float	bool	bool	float
Target	float	float	bool	bool	float
Tracking	float	float	bool	bool	float
Auto-Mode	bool			bool	bool
Overwrite-Mode	bool				

Table 6.2.: This table indicates all variables inspected by `operatorAction`. An entry indicates that one variable exists. Connected cells, like in the last row, indicate that a variable is held that is authoritative for several aspects. Except for the DSP-values, all values are held in the FSM-database.

that allows a final veto from the operator. If this is set to false, the automation is not allowed to change the DSP system at all.

The procedure `operatorAction` cannot change the values inside the DSP itself. Instead it is terminating with a return-code that causes the FSM to trigger subsequent algorithms that correct these settings. Only a restricted number of algorithms can act on the DSP settings. Table 6.3 summarizes all possible behaviours of the algorithm. Apart from gradient and phase set-point, the algorithm is also applied on feedback-gain, feedforward-mode and feedback-mode.

Usage of the Procedure in the FSM-Context

The algorithm is invoked periodically in the exception-block of the FSM. If the return-code of the algorithm indicates the necessarity of an update of DSP variables, the exception-block sends a signal to the operation-block. Since the algorithm is reading a number of variables from the control system, periodical network traffic is produced. Instead of periodically *pulling* the concerned variables, *pushing* of the control system would be desirable due to the reduced network traffic. This is not yet supported by the DOOCS control system.

6.5. Ramping of the Feedback

Description of the Procedure

The goal of this algorithm is to compare the current gain of the proportional feedback with the target and if necessary change its value. The algorithm cooperates with the previously introduced algorithm that tracks operator interference in the sense that any change to the feedback caused by this algorithm is written into a tracking variable, making it possible to distinguish between changes in the feedback-gain originating from the operator and this algorithm.

Once the algorithm is invoked, it checks if it is allowed to touch DSP-settings at all. If not, it quits. Then it compares the current gain with the target gain and stops if both are identical. If the target gain is lower than the current gain, it simply decreases the gain. If the target gain is higher, it checks if the target gain is below a reasonable value and quits if not. Then, it calculates

ovw	sp=tar	trk=tar	trk=sp	action
true	*true*	*true*	*true*	Do nothing.
true	*true*	*true*	*false*	Do nothing (impossible combination).
true	*true*	*false*	*true*	Do nothing (impossible combination).
true	*true*	*false*	*false*	Update tracker (inconsistency).
true	*false*	*true*	*true*	Do nothing (impossible combination).
true	*false*	*true*	*false*	Operator changed sp. Disable auto-mode. Update target and tracker.
true	*false*	*false*	*true*	Operator changed target. Send signal if in auto-mode.
true	*false*	*false*	*false*	Inconsistency. Update target and tracker.
false	*true*	*true*	*true*	Do nothing.
false	*true*	*true*	*false*	Do nothing (impossible combination).
false	*true*	*false*	*true*	Do nothing (impossible combination).
false	*true*	*false*	*false*	Inconsistency. Update tracker.
false	*false*	*true*	*true*	Do nothing (impossible combination).
false	*false*	*true*	*false*	Operator changed sp. Send signal if in auto-mode.
false	*false*	*false*	*true*	Operator changed target. Send signal if in auto-mode.
false	*false*	*false*	*false*	Inconsistency. Update tracker. Send signal if in auto-mode.

Table 6.3.: Full decision table for tracking of operator-action. The decision table is valid for the observation of gradient, phase, feedback-gain, feedforward-mode and feedback-mode. “ovw”: true if overwrite mode is active (operator-interference overwrites the FSM). “sp=tar”: setpoint (amplitude, phase, feedback-gain, feedforward-mode, feedback-mode) is equal to the target value of the FSM. “trk=tar” tracking variable is equal to the target value. “trk=sp” tracking variable is equal to the setpoint. As a further switch, the “auto-mode” is considered here. If the setpoint should be updated and if auto-mode is true, a signal is send that causes the FSM to trigger action for updating the settings.

an increasing sequence of gain-values, starting from the current gain and ending with the target gain. Basically, the sequence is defined by the powers of 1.7 (which is a heuristical value) with a few corrections at the boundaries. In order to increase the gain from 5 to 40, the sequence would be $[5, 1.7^4, 1.7^5, 1.7^6, 40] = [5, 8.3, 14.2, 24.1, 40]$. Then, the sequence is applied to the loop-gain property of the DSP system. After setting a value from the sequence, data is taken from the current vector-sum. If it turns out that an overshoot of more than 20% occurred, the algorithm quits.

Usage of the Procedure in the FSM-Context

The algorithm is part of the tweak superstate of the FSM. Therefore it is included in a chain of algorithms that work together. The tweak superstate is active whenever a minor parameter optimization is necessary. Tweaking is usually triggered by the states in the exception block. An example can be that `operatorAction` recognizes a change in the setpoint variables.

7. Conclusions

Signal Calibration

Section 3.9.1 approximates the minimum calibration requirements for a resonator without a field probe, e. g. a normal-conducting electron gun, as

$$\Delta\varphi_{\text{cavity}} = \sqrt{2}\left(\frac{\Delta A}{A}\right)\sin\psi \tag{7.1}$$

for the amplitude calibration and

$$\left(\frac{\Delta A}{A}\right)_{\text{cavity}} = \sqrt{2}\Delta\varphi\sin\psi. \tag{7.2}$$

for the phase calibration. A calibration error of 1% in amplitude and 1° in phase will allow to measure the field up to 0.1% in amplitude and 0.04° in phase.

This requirement is in reach of the methods described in section 4.3.1. The new method is based on resonance circles that are obtained from DAC induced detuning. A simple chi-square fit of a circle yields the calibration parameters at an accuracy of better than a percent in amplitude and one degree in phase.

In section 3.9.2, the requirements for vector-sum calibration are approximated as

$$\Delta\varphi_{N\text{-vector-sum}} = \frac{2}{3}\frac{1}{\sqrt{N}}\frac{\psi^*}{2}\cdot\left(\frac{\Delta A}{A}\right). \tag{7.3}$$

in amplitude and

$$\left(\frac{\Delta A}{A}\right)_{N\text{-vector-sum}} = \frac{2}{3}\frac{1}{\sqrt{N}}\frac{\psi^*}{2}\cdot\Delta\varphi \tag{7.4}$$

in phase.

The control requirements of the European XFEL, 10^{-4} in amplitude and 0.01° in phase can be evaluated to a calibration requirement of

$$\left(\frac{\Delta A}{A}\right) = 1\,\% \quad\text{and}\quad \Delta\varphi = 0.5°. \tag{7.5}$$

This consideration assumes $N = 32$ cavities in a single vector-sum and a maximum microphonics of $\Delta\omega^*/\omega_{1/2} = 1/10$.

The method presented in 4.3.3 for beam-based vector-sum calibration allows to calibrate the

vector-sums up to a few percent in amplitude and a few degrees in phase at FLASH. While this is enough for FLASH requirements, it is not for XFEL, in praticular due to the discussed limitations derived from the presence of detuning and differences in shunt impedance from different cavities.

While section 4.3.1 discusses the calibration of the measurement of the forward wave and reflected wave for a normal-conducting resonator, section 4.3.2 discusses a method for superconducting cavities. The method for superconducting cavities differs significantly from the method for the normal-conducting cavities. First of all, it requires a field-probe. The method discussed is even capable of determining the cross talk from the forward to the reflected channel and vice versa. It is the basis for measurements that require a very good calibration, as for example the measurement of the detuning discussed in section 4.4.4, the FPGA-based detuning monitor. Even though the calibration method produces good results, it is very sensitive to crosstalk from channels other than the forward and reflected channel. The limitations are also discussed in section 4.3.2.

Not quite a signal calibration problem, but a signal adjustment problem is discussed in section 4.2. Usually, the weakest cavity limits the performance of a vector sum. Section 4.2 presents a method to calculate values for predetuning and coupling that need to be adjusted in order to increase this performance limit.

Algorithms and Procedures

A number of new algorithms have been developed and discussed within this thesis.

In section 4.4.2, the effects of passband modes beside the accelerating mode is discussed. The potential of instability is demonstrated together with a simple measure to avoid instablity. Stability is achieved by simply increasing the latency in the loop. This is a surprising result since an increased latency is usually rather a source of instability than of stability.

Section 4.4.3 introduces a new tool for the determination of the loop phase and the system gain. In particular the loop phase is a quantity that is subject to thermal drift. Since a wrong loop phase can lead to an instability in the loop, it is crucial to apply corrections to the loop phase compensation in the controller on a regular basis. This is achieved by the new tool.

In section 3.3 of the theory chapter the polar representation of the cavity equations is discussed. This leads to a simple formula for the detuning of a cavity, derived from equation 3.22. In section 4.4.4 it is presented how this formula was implemented into a real time device (FPGA). This allows to calculate the detuning inside a pulse. The algorithm is very sensitive to the calibration of the signals. The signal calibration is performed off-line and is discussed in section 4.3.2. The FPGA-based detuning monitor can serve as a basis for several developments of algorithm that need information about the detuning of a pulse.

Section 4.4.5 introduces a new method for the analytical evaluation of linear, iterative schemes for the calculation of feedforward tables. Central result of this calculation is that linear schemes have the intrinsic potential of becoming instable after a number of iterations. The analytic expressions of linear schemes contains the inverse of the transfer function of the cavity (plant) P^{-1}, which is already a strong indication for the instability of such a scheme. A solution for the feedforward problem is offered in section 4.4.6, where the linear scheme was modified by the introduction of time-reversed filters. This new scheme is not covered by the linear theory but has

proven to be successful in various tests.

Automation Framework

Chapter 5 introduces a new framework for LLRF automation. The framework is designed in a way that everybody, in particular hardware developers and operators, are capable of adapting the automation to a changing environment. The framework is based on the principles of finite state machines which is an established concept for automation in industry. At the heart of the automation scheme is a procedure server that acts as quantization from the automation theory point of view. Section 5.2 discusses the solution formally and shows equivalence to the FSM picture. A number of differences compared to industrial automation exist. First of all, the framework for LLRF automation is required to deal with operators bypassing it. Thus, a framework was developed that is flexible enough to allow changes of the automation without prior programming experience. Further, the procedures contain comprehensive exception handling to be able to detect interference from operators acting directly on the parameters of the control system while bypassing the automation.

Transparency was one of the guiding motives during the development of the framework. Transparency is needed in order to introduce an automation scheme without objection from the operators, who learned to use the machine without automation. The challenge at FLASH was to introduce an automation for a small number of only five RF stations. Due to the small number, the acceptance barrier of automation scheme and the fear of failures caused by a not fully matured scheme was high. Therefore, only small aspects of the presented automation scheme have gained the full acceptance from the operators, in particular the automated loop phase correction and, for long beam operation, the adaptive feedforward scheme.

The presented concept is due to its flexibility a prototype and a basis for the automation of future accelerators. Especially projects like the European XFEL face the need of a high degree of automation due to its large number of about 30 RF stations.

Appendices

A. Slopes of Amplitude and Phase of the Step Response

In this appendix the initial slopes of amplitude and phase for a step-response of a cavity mode shall be calculated. The results of this lengthy calculation were already presented in chapter 3.2. The expression of interest is the transient response of a cavity mode to a step at the drive (identical to 3.14),

$$x_{\text{step}} = \frac{\omega_{1/2}}{\omega_{1/2} - i\Delta\omega}(1 - e^{-(\omega_{1/2} - i\Delta\omega)t})u. \tag{A.1}$$

Generally, from the formulas of amplitude $|z| = \sqrt{z\bar{z}}$ and phase $\angle z = \arctan \operatorname{Im} z / \operatorname{Re} z$ one derives the first derivatives with respect to t as (for well-behaving complex functions z)

$$\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{z\bar{z}} = \operatorname{Re}\left(\dot{z}\sqrt{\frac{\bar{z}}{z}}\right) \tag{A.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\arctan\left(\frac{\operatorname{Im} z}{\operatorname{Re} z}\right) = \operatorname{Im}\frac{\dot{z}}{z}. \tag{A.3}$$

A.2 needs to be evaluated for $z = x_{\text{step}}$ for $t \to 0$ in order to retrieve the slope of the amplitude of the step-response. The expression under the root $\sqrt{}$ clearly becomes unity, while the first derivative of x_{step} is $\omega_{1/2}u$ for $t \to 0$, therefore,

$$\lim_{t\to 0}\frac{\mathrm{d}}{\mathrm{d}t}|x_{\text{step}}| = \omega_{1/2}|u|. \tag{A.4}$$

The expression for the phase, A.3, can be further simplyfied since

$$\frac{a + ib}{c + id} = \frac{ac + bd + i(bc - ad)}{c^2 + d^2} \text{ or} \tag{A.5}$$

$$\operatorname{Im}\frac{\dot{z}}{z} = \frac{\operatorname{Im}\dot{z}\operatorname{Re} z - \operatorname{Re}\dot{z}\operatorname{Im} z}{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2} \tag{A.6}$$

The only term from A.1 contributing to a derivative in phase is $z = 1 - e^{-(\omega_{1/2} - i\Delta\omega)t}$. z and $\dot{z}$ can be separated in their real and imaginary parts,

$$z = 1 - e^{-(\omega_{1/2} - i\Delta\omega)t} = 1 - e^{-\omega_{1/2}t}\cos\Delta\omega t - ie^{-\omega_{1/2}t}\sin\Delta\omega t \tag{A.7}$$

$$\begin{aligned}\dot{z} &= (\omega_{1/2} - i\Delta\omega)e^{-(\omega_{1/2} - i\Delta\omega)t} \\ &= (\omega_{1/2}\cos\Delta\omega t + \Delta\omega\sin\Delta\omega t)e^{-\omega_{1/2}t} + i(\omega_{1/2}\sin\Delta\omega t - \Delta\omega\cos\Delta\omega t)e^{-\omega_{1/2}t}.\end{aligned} \tag{A.8}$$

Now the derivative of the phase can be written as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\angle z &= \frac{\omega_{1/2}e^{-\omega_{1/2}t}\sin\Delta\omega t - \Delta\omega e^{-\omega_{1/2}}\cos\Delta\omega t + \Delta\omega e^{-\omega_{1/2}t}}{1 + e^{-2\omega_{1/2}t} - 2e^{-\omega_{1/2}t}\cos\Delta\omega t} \\ &= \frac{\omega_{1/2}\sin\Delta\omega t - \Delta\omega\cos\Delta\omega t + \Delta\omega e^{-\omega_{1/2}t}}{e^{-\omega_{1/2}t} + e^{\omega_{1/2}t} - 2\cos\Delta\omega t}\end{aligned} \tag{A.9}$$

The approximation of A.9 for $t \to 0$ can, for example, be done by expanding the trigonometric and exponential functions into their Taylor series. Doing so, it turns out that all zero- and first-order term in either numerator and denominator cancel each other out. Finally, one yields

$$\lim_{t\to 0}\frac{\mathrm{d}}{\mathrm{d}t}\angle z = \lim_{t\to 0}\frac{\frac{\Delta\omega}{2}t^2(\omega_{1/2}^2 + \Delta\omega^2)}{t^2(\omega_{1/2}^2 + \Delta\omega^2)} = \frac{\Delta\omega}{2}. \tag{A.10}$$

In summary, the slopes of amplitude and phase of the step-response can be written as

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}|x_{\mathrm{step}}|\right]_{t=0} = \omega_{1/2}|u| \quad \text{and} \quad \left[\frac{\mathrm{d}}{\mathrm{d}t}\angle x_{\mathrm{step}}\right]_{t=0} = \frac{\Delta\omega}{2}. \tag{A.11}$$

B. Derivation of the Extended Naive Adaptive Feedforward

Extended Naive Adaptive Feedforward

This appendix derives the analytical expression for the extended naive feedforward. The rules for this algorithm shall be stated in words first:

1. Subtract the current feedforward table from the total drive. The result is the feedback contribution to the drive.
2. Apply a digital filter D on the feedback.
3. Add the filtered feedback on top of the existing feedforward curve. That will be used as feedforward in the next pulse.

Similarly a diagram can be found that represents this algorithm. Figure B.1 shows such a diagram. Three loops are visible, where bold brackets around the center loop indicate that this part of the diagram has to be repeated $n-2$ times. In the figure, certain signals are labled $I_1..I_n$ (in the following considerations, the transfer function that leads from SP to I_i is identified with I_i itself). As visible from the diagram, I_1 represents the field of the first cavity filtered by D. The negative of this signal (indicated by the small minus sign) is added to the summing junction of the next cavity in the cascade, representing the next pulse. In other words: the setpoint of the second loop is reduced by the D-filtered field-achievements of the first loop and the setpoint for the second loop is just "what is leftover after the first loop did its job". I_2 plus I_1 then would be the field of the cavity after the first iteration of the algorithm. The setpoint of the third loop then would be reduced by the D-filtered achievements of the previous loops and so on. An expressions for the I_is shall now be found (keeping in mind that the sum of all I_is is the field inside the cavity after

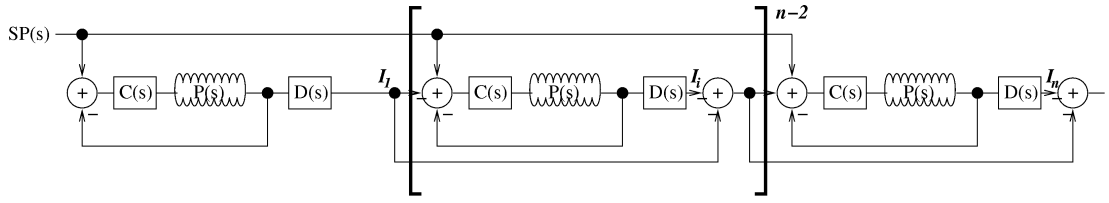

Figure B.1.: Diagram representing the extended naive adaptive feedforward. Three control loops are shown in this figure, each representing an iteration of the algorithm. The center control loop is surrounded by bold brackets, indicating that this part of the diagram shall be repeated $n-2$ times.

n iterations).

$$\begin{aligned} I_1 &= \frac{CP}{1+CP}D \\ I_2 &= (1-I_1)\frac{CP}{1+CP}D = (1-I_1)I_1 \\ I_3 &= (1-I_1-I_2)\frac{CP}{1+CP}D = (1-I_1-I_2)I_1 \\ &\vdots \\ I_n &= (1-\sum_{i=1}^{n-1} I_i)\frac{CP}{1+CP}D = (1-\sum_{i=1}^{n-1} I_i)I_1 \end{aligned} \tag{B.1}$$

These equations can directly be developed by looking at figure B.1. They can further be simplified.

$$\begin{aligned} I_1 &= I_1 \\ I_2 &= (1-I_1)I_1 \\ I_3 &= (1-I_1-I_1+I^2)I_1 = (1-I_1)^2 I_1 \\ &\vdots \\ I_n &= (1-I_1)^{n-1} I_1 \end{aligned} \tag{B.2}$$

Comparing the equations B.1 and B.2 is a tricky way getting an expression for the sum $\sum_i I_i$. This sum is at the same time the transfer function from the setpoint to the cavity field after n iterations and differs from the feedforward just by the inverse of P. This way, an expression for the extended naive adaptive feedforward is already found (H beeing the transfer function from the setpoint to the feedforward after n iterations):

$$\begin{aligned} HP &= \sum_{i=1}^{n} I_i \\ &= 1-(1-I_1)^n \\ &= 1-\left(1-\frac{CP}{1+CP}D\right)^n \end{aligned} \tag{B.3}$$

Again, the transfer function for the single loop from setpoint to drive, $L = C/(1+CP)$ can be substituted.

$$H = \left[1-(1-LDP)^n\right] P^{-1} \tag{B.4}$$

This equation is again a very simple expression and differs from the naive adaptive feedforward of equation 4.41 only by the insertion of a transfer function D.

Other Schemes

The presented way of deriving linear transfer functions for linear, iterative operations on pulsed systems is very powerful and can be used to develop analytical expressions for lots of very different algorithms.

Another topology, which is described in words as

1. take the error-signal (setpoint minus cavity field) of the feedforward-only driven system and apply a filter Q. Q can be a filter very close to the inverse of the real plant, P^{-1}.
2. Add the Q-filtered error signal on top of the existing feedforward. If $Q = P^{-1}$, the new feedforward-signal would lead to perfect agreement between the cavity field and the setpoint.

is useful for systems which do not have real-time feedback inside the pulse but which apply a correction between pulses. It is especially interesting for normal-conducting resonators where the bandwidth is large and the pulse-length short such that a real-time feedback can not be applied due to stability considerations.

Mathematically, this scheme can be formulated such that the drive J_{n+1} of the $(n+1)$th-iteration is constructed from the previous one by the formula

$$J_{n+1} = J_0 + J_n(1 - PQ) \tag{B.5}$$

Again, this can be reduced to a geometric series with the result (transfer-function between the setpoint-table and the feedforward after the nth iteration)

$$H = \left[1 - (1 - PQ)^n\right] P^{-1}. \tag{B.6}$$

However, instability of this scheme can be proven with the same arguments that were made in section 4.4.5.

Acknowledgments

I acknowledge the steady support of my Professor Jörg Roßbach. Special thanks goes to Stephan Simrock for uncounted fruitful discussions which not only taught me deep insides about LLRF but also lessons on how to discuss. I am very grateful to a number of people at DESY and elsewhere who shared their wisdom with me. Especially the engineers I worked with deserve deep appreciation for the patience they had with me.

Bibliography

[1] I. Will, G. Koss, and I. Templin. The upgraded photocathode laser of the TESLA Test Facility. *NIM A*, 541:467–477, 2005.

[2] B. Dwersteg, K. Floettmann, J. Sekutowicz, and Ch. Stolzenburg. RF gun gesign for the TESLA VUV Free Electron Laser. *NIM A*, 393:93–95, 1997.

[3] B. Aune et al. The Superconducting TESLA Cavities. *Phys. Rev. ST-AB*, 092001, 2001.

[4] M. Liepe and S. Belomestnykh. Microphonics Detuning in the 500 MHz Superconducting CESR Cavities. In *Proceedings of PAC 2003, Portland, USA*, 2003.

[5] Valeri Ayvazyan and Stephan Simrock. Dynamic Lorentz Force Detuning Studies in TESLA Cavities. In *Proceedings of EPAC 2004, Lucerne, Switzerland*, 2004.

[6] V. Balandin, P. Castro, and N. Golubeva. Beam Optics Studies for the TESLA Test Facility Linac. In *Proceedings of LINAC 2004, Lübeck, Germany*, 2004.

[7] Frank Stulle. *A Bunch Compressor for small Emittances and high Peak Currents at the VUV Free-Electron Laser*. Doktorarbeit, Universität Hamburg, 2004.

[8] Andy Bolzmann. *Investigation of the longitudinal charge distribution of electron bunches at the VUV-FEL using the transverse deflecting cavity LOLA*. Diplomarbeit, Julius-Maximilians-Universität Würzburg, 2005.

[9] E. L. Saldin, E. A. Schneidmiller, and M. V. Yurkov. *The Physics of Free-electron Lasers*. Springer, Berlin, Heidelberg, 2000.

[10] Valeri Ayvazyan et al. First operation of a free-electron laser generating GW power radiation at 32 nm wavelength. *The European Physical Journal D*, 37:297–303, 2006.

[11] The Tesla Test Facility Team. SASE FEL at the TESLA Facility, Phase 2. *DESY Print TESLA-FEL*, 01-2002, 2002.

[12] M. Dohlus and T. Limberg. Bunch Compression Stability Dependence on RF Parameters. In *Proceedings of FEL 2005, Stanford, California, USA*, 2005.

[13] Stefan Simrock, Frank Ludwig, and Holger Schlarb. Synchronization systems for ERLs. *NIM A*, 557:293–298, 2006.

[14] Yujong Kim et al. S2E simulations on jitter for european XFEL project. In *Proceedings of EPAC 2004, Lucerne, Switzerland*, 2004.

[15] Thomas Schilcher. *Vector-sum Control of Pulsed Accelerating Fields in Lorenz Force detuned Superconducting Cavities*. Doktorarbeit, Universität Hamburg, 1998.

[16] Matthias Liepe. *Regelung supraleitender Resonatoren mit Strahlbelastung am TESLA-Test-Linearbeschleuniger*. Diplomarbeit, Universität Hamburg, 1998.

[17] D.E. Nagel. Coupled Resonator Model of Linear Accelerator Tanks. In *Conference on Proton Linear Accelerators, MURA 714*, page 21ff., 1964.

[18] Jan Lunze. *Regelungstechnik 2, Mehrgrößensysteme und Digitale Regelung*. Springer, 2nd edition, 2002.

[19] Markus Hüning. *Selbstoptimierende Parametersteuerung der Hochfrequenz des supraleitenden Linearbeschleunigers TESLA Test Facility*. Diplomarbeit, Rheinisch-Westfälische Technische Hochschule Aachen, 1998.

[20] Alexander Brandt, Piotr Pucyk, and Stefan Simrock. Field Estimation and Signal Calibration of RF Guns without Field Probe. *DESY Print TESLA-FEL*, 01-2007, 2007.

[21] P. Pawlik, M. Grecki, and S. Simrock. System for High Resolution Detection of Beam Induced Transients in RF Signals. In *Proceedings of Mixdes 2005, Krakow, Poland*, 2005.

[22] Charles L. Phillips and Roye D. Harbor. *Feedback Control Systems*. Prentice Hall, 2000.

[23] Wojciech Jalmuzna. *Design and implementation of universal mathematical library supporting algorithm development for FPGA based systems in high energy physics experiments*. Master thesis, Warsaw University of Technology, 2006.

[24] Jack E. Volder. The CORDIC Trigonometric Computing Technique. *IRE Transactions on Electronic Computers*, 1959.

[25] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Einführung in die Automatentheorie, Formale Sprachen und Komplexitätstheorie*. Pearson Studium, München, 2002.

[26] Edward F. Moore. Gedanken-Experiments on Sequential Machines. *Automata Studies*, pages 129–153, 1956.

[27] George H. Mealy. Method for Synthesizing Sequential Circuits. *Bell System Technical Journal*, 34:1045–1079, 1955.

[28] D. Harel. A Visual Formalism for Complex Systems. *Science of Computer Programming*, 8(3):231–274, 1987.

[29] Kay Rehlich. Status of the TTF VUV-FEL Control System. In *Proceedings of PCaPAC, Hayama, Japan*, 2005.

[30] Artur Aghababyan, Vitali Kocharyan, and Kay Rehlich. An Integrated Finite State Machine to Automate TTF operation. In *Proceedings of PCaPAC, Hamburg, Germany*, 2000.

[31] M. Krabs. *Das ROSA-Verfahren zur Modellierung dynamischer Systeme durch Regeln mit statistischer Relevanzbewertung*, volume 404. VDI Verlag, Düsseldorf, 1994.

[32] Artur Aghababyan et. al. Data Acquisition System for a VUV-FEL Linac. In *Proceedings of PCaPAC, Hamburg, Germany*, 2005.

[33] Alexander Brandt. Implementation Details of FLASH LLRF Automation and Algorithms. *(LLRF Internal Note)*, 2006.