Cavity Field Control for Linear Particle Accelerators

Troeng, Olof

2019

Document Version:
Publisher's PDF, also known as Version of record

Cavity Field Control for Linear Particle Accelerators

Olof Troeng

LUND UNIVERSITY

Department of Automatic Control

PhD Thesis
ISRN LUTFD2/TFRT--1127--SE
ISBN 978-91-7895-276-2 (print)
ISBN 978-91-7895-277-9 (web)
ISSN 0280–5316

Department of Automatic Control
Lund University
Box 118
SE-221 00 LUND
Sweden

Cover image: Roger Eriksson/ESS.
Printed in Sweden by MediaTryck.
Lund 2019

Abstract

High-energy linear particle accelerators enable exploration of the microscopic structure of pharmaceuticals, solar cells, fuel cells, high-temperature superconductors, and the universe itself. These accelerators accelerate charged particles using oscillating magnetic fields that are confined in metal cavities. The amplitudes and phases of the electromagnetic fields need to be accurately controlled by fast feedback loops for proper accelerator operation.

This thesis is based on the author's work on performance analysis and control design for the field control loops of the linear accelerator at the European Spallation Source (ESS), a neutron microscope that is under construction in Lund, Sweden. The main contribution of the thesis is a comprehensive treatment of the field control problem during flat-top, which gives more insight into the control aspects than previous work. The thesis demonstrates that a key to understand the dynamics of the field control loop is to represent it as a single-input single-output system with complex coefficients. This representation is not new itself but has seen limited use for field control analysis.

The thesis starts by developing practical and theoretical tools for analysis and control design for complex-coefficients systems. This is followed by two main parts on cavity field control. The first part introduces parametrizations that enable a better understanding of the cavity dynamics and discusses the most essential aspects of cavity field control. The second part builds on the first one and treats a selection of more advanced topics that all benefit from the complex-coefficient representation: analysis of a polar controller structure, field control design in the presence of parasitic cavity resonances, digital downconversion for low-latency feedback, energy-optimal excitation of accelerating cavities, and an intuitive design method for narrowband disturbance rejection. The results of the investigations in this thesis provide a better understanding of the field control problem and have influenced the design of the field controllers at ESS.

Acknowledgments

Thank you, Bo Bernhardsson I could not have asked for a better supervisor. Your deep understanding of control in both theory and practice, your helpfulness, your optimism, and your sense of humor have made my PhD studies worthwhile. Your sharp eye for detail has greatly improved this thesis as well as my other manuscripts. I am also thankful to my co-supervisors Anders J Johansson and Rolf Johansson. Anders has led the low-level RF project at Lund University with great enthusiasm, organized visits to accelerator facilities, been supportive of my work and carefully reviewed this thesis. Rolf has carefully reviewed this thesis and skillfully negotiated contractual matters for my involvement with the ESS project.

The helpfulness and openness of the low-level RF experts around the world have been invaluable in my work on this thesis. I have spent week-long research visits with Mark Crofford and his colleagues at the Spallation Neutron Source, with Sven Pfeiffer and his colleagues at the Deutsches Elektronen-Synchrotron, and with Larry Doolittle and Carlos Serrano at Berkeley Labs. I have also made shorter visits to Philippe Baudrenghien at CERN, Thomas Schilcher at the Paul Scherrer Institute, and Chang V at the Institute of Modern Physics. I greatly thank you all for sharing your time and expertise with me. I am thankful to Claudio Rivetta for stimulating collaboration on complex-coefficient system.

I would like to express my sincere gratitude to the members of the beam physics and RF groups at ESS for taking their time to discuss with me and answer my questions. I would in particular like to thank Anders Svensson, Fredrik Kristensen, Christian Amstutz and Anirban Krishna Bhattacharyya for their work with the firmware and hardware of the ESS low-level RF system, the many interesting technical discussions that we have had, and all the fun travels that we have done.

Thank you Daniel Sjöberg, Anders Karlsson, and Gabriele Constanza at the Department of Electrical and Information Technology, Lund University, for sharing your expertise on the dynamics of electromagnetic cavity fields.

Thank you Åke Andersson and your colleagues at MAX IV for giving me a perspective on the ESS proton linac by sharing your expertise on electron synchrotrons and letting me visit your synchrotron on multiple occasions.

I am grateful to all my past and present colleagues at the Department of Automatic Control for excellent research discussions and all the fun that we have done together. Special thanks go to the administrative staff for making things run smoothly, to Leif Anderson for sharing his expertise on typography, and to my office mates throughout the years. Mattias Fält and Fredrik Bagge Carlson have been excellent travel companions and have provided great support on Julia programming. Mattias was also a great collaboration partner on an interesting optimization problem. Melika Shahriari from TRIUMF spent six months at the department and during this time we had many intriguing discussion on iterative learning control. Olle Kjellqvist did a great job of investigating numerical issues in Q design. Finally, Anton Cervin did an incredible job of reviewing this thesis, which led to significant improvements. Parts of the thesis were reviewed by Marcus Thelander Andrén, Martin Heyden, and Richard Pates.

Last but not least, I would like to thank those that are most important to me: my parents, my sister, my friends and Sara.

Financial Support

The author's work for the European Spallation Source was done under a collaboration agreement between ESS and Lund University. The author is a member of the ELLIIT Excellence Center (supported by the Swedish Government) and was a member of the LCCC Linnaeus Center (supported by the Swedish Research Council). His attendance at the American Control Conference 2017 was supported by the Ericsson Research Foundation.

Contents

Abbreviations and Notation

Abbreviations

Abbreviation	Meaning
ADC	Analog-to-Digital Converter
DAC	Digital-to-Analog Converter
DDC	Digital Downconversion
DTL	Drift-Tube Linac (a type of accelerating cavity)
ESS	European Spallation Source (Neutron source in Lund, Sweden)
Eu-XFEL	European X-ray FEL (FEL in Hamburg, Germany)
FEL	Free-Electron Laser
FPGA	Field-Programmable Gate Array
IF	Intermediate Frequency
ILC	Iterative Learning Control
IQ	In-phase and Quadrature (real and imaginary parts of a complex number)
LLRF	Low-Level Radio-Frequency
LCLS-II	Linac Coherent Light Source (FEL in Stanford, CA, USA)
LO	Local Oscillator
LTI	Linear Time-Invariant
MIMO	Multiple-Input Multiple-Output
MO	Master Oscillator
PI(D)	Proportional-Integral-(Derivative)
RF	Radio Frequency
RFQ	Radio-Frequency Quadrupole (a type of accelerating cavity)
rms	root-mean square
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SNS	Spallation Neutron Source (Neutron source in Oak Ridge, TN, USA)
TITO	Two-Input Two-Output

Notation

The list below has been limited to the most important symbols.

Symbol	Meaning
$G^*(s)$	The complex-conjugate transfer function ($\neq [G(s)]^*$ and $\neq G^{\mathsf{H}}(s)$, see p. 48)
$(\cdot)^*$	The complex conjugate of something
$(\cdot)^{\mathsf{T}}$	The transpose of something
$(\cdot)^{\mathsf{H}}$	The conjugate transpose of something
$(\cdot)^\star$	Indicates that something is optimal
$[-]$	Indicates that a quantity is dimensionless
ϕ_*	The angle of a phasor
θ_*	A phase shift in the control loop
γ_*	A decay rate
τ_*	A time delay

Complex signals and constants will be denoted by bold letters. In block diagrams with both real and complex signals the latter will be indicated by double arrows (⇉).

Nomenclature

Control will refer to automatic control and not industrial control (system). Cavity field errors will refer to what often is called RF jitter in the accelerator literature. Bandwidth will always refer to the bandwidth of a baseband system (equal to the half-bandwidth of the corresponding bandpass system).

1

Introduction

High-energy linear particle accelerators have many important applications: they drive spallation sources and free-electron lasers which reveal the structure of proteins, pharmaceuticals, and materials; they are integral components in most high-energy physics experiments; and they are researched as a means to cleaner nuclear energy.

For proper operation of a linear accelerator, it is crucial that the electromagnetic fields that accelerate the charged particles have the correct amplitude and phase. The work in this thesis has been motivated by the challenges to control the electromagnetic fields of the linear proton accelerator at the European Spallation Source.

1.1 The European Spallation Source

Neutron imaging reveals features that are invisible to X-rays (Figure 1.1) and is a crucial technology for understanding the atomic structure of materials and organic molecules. The European Spallation Source will, once completed, be the world's brightest neutron source. It has been said that its long, high-intensity pulses will bring a paradigm shift to neutron imaging [ESSa, 2017] that will enable breakthroughs in pharmacology, chemistry, solid-state physics, engineering sciences, material sciences, and archeology [Peggs et al., 2013].

The ESS is under construction outside of Lund, Sweden since 2014, as a collaboration between 13 European member countries. ESS will be a sustainable research facility and will be driven by renewable energy. The heat in its cooling water will be recycled to Lund's district-heating network.

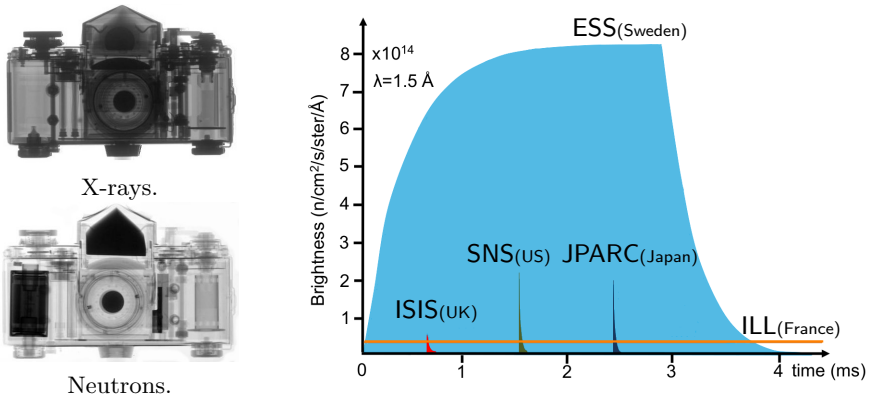


Figure 1.1 *Left*: Neutrons and X-rays interact differently with atoms which allow them to reveal complementary information. Note that certain parts of the camera are invisible to X-rays. Photo credit: Paul Scherrer Institute. *Right*: Comparison of the neutron pulse of ESS and the neutron pulses of the currently brightest neutron sources.

Spallation

The neutrons at ESS will be obtained through a nuclear reaction called spallation, by bombarding a tungsten target with protons (Figure 1.2). The emitted neutrons will be guided to the 15 different state-of-the-art neutron instruments by neutron guides. The main components of ESS are illustrated in Figure 1.3, for details see the technical design report [Peggs et al., 2013].

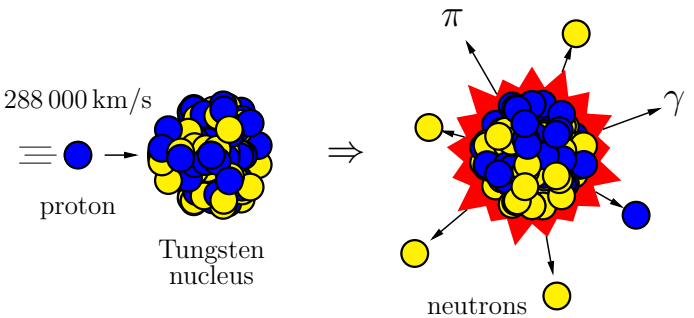


Figure 1.2 Smashing high-velocity protons into tungsten nuclei releases free neutrons through a nuclear reaction called spallation [Russell, 1990].

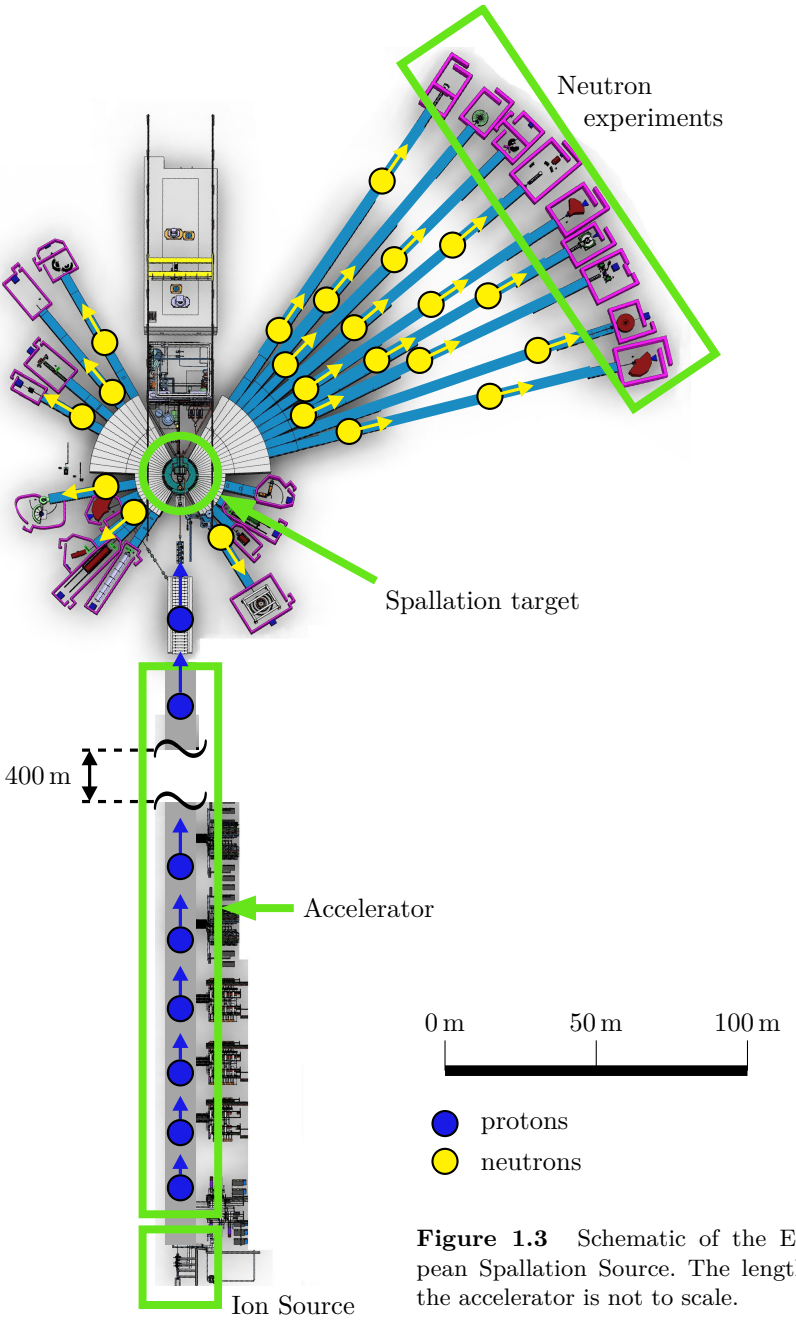


Figure 1.3 Schematic of the European Spallation Source. The length of the accelerator is not to scale.

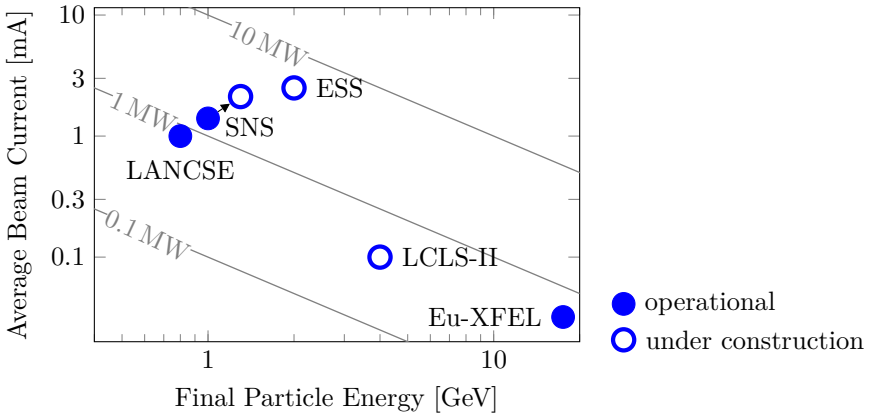


Figure 1.4 How the 5 MW average power of the ESS accelerator compares to the currently most powerful linear accelerators in the world.

1.2 Proton acceleration and the need for cavity field control

The protons at ESS will be accelerated by a 400-meter-long linear accelerator (linac). The accelerating force will be provided by oscillating electromagnetic fields confined in hollow metal structures called accelerating cavities. In total there will be 155 cavities of six different types along the accelerator.

The protons will be accelerated in bunches of roughly one billion protons each. The acceleration of the bunches and how well they are kept together (longitudinal focusing) depend on the amplitudes and the phases (relative to the bunches) of the accelerating electromagnetic fields. Unless the amplitudes and phases are accurately controlled there will be a focusing mismatch that leads to that protons are lost from the bunches. These protons crash into the cavity walls and induce radioactivity. Too much beam loss and consequent radioactivation delays hand-on maintenance of the accelerator, thereby reducing its availability.

The ESS linac will be the world's most powerful linac (Figure 1.4) due to its high beam current (62.5 mA) and high duty-factor (4 %). This means that only small field errors can be accepted before the radioactivation becomes too severe. The field-error requirements are unprecedentedly hard for a high-intensity proton linac. It is the challenges of meeting these requirements that have motivated this thesis.

The field control problem

A typical pulse of a pulsed linac is shown in Figure 1.5. Ripple from the RF amplifier's power supply and the beam current act as disturbances on the cavity field, giving rise to field errors. To keep the cavity field at its setpoint requires fast feedback loops that adjust the output of the RF amplifier based on measurements of the cavity field. Such a field control loop is illustrated in Figure 1.6.

Particular field control challenges for the ESS linac

There are a number of particular field control challenges for the ESS linac.

- The high beam current means that the cavities are more sensitive to disturbances such as amplifier ripple and beam-current ripple than in linacs with lower beam currents.
- The relative magnitude of the disturbances themselves are larger. For example, it is difficult to manufacture low-ripple pulsed power supplies for the RF amplifiers due to the massive output power that is required.
- The requirement that the field errors must be less than 0.2 % (rms) in amplitude and 0.2° (rms) in phase for the nine normal conducting cavities is unprecedentedly hard for a high-intensity proton linac.
- There will be a total of 155 cavities of six different types, driven by different types RF amplifiers. Many of the cavities have parasitic resonance modes that can cause feedback instability. The heterogeneity of the field control loops makes it necessary with a simple and understandable controller structure where parameter tuning is easy to automate.

1.3 Background literature and previous work

Previous work on cavity field control

There is no text book on cavity field control but the PhD thesis [Schilcher, 1998] is often suggested as an introduction. Introductory material is also provided by the CERN accelerator school [Baudrenghien, 2000; Schilcher, 2007] and in [Garoby, 1996]. In the author's experience the best way to learn about cavity field control is through personal communication with experts in the field. Recent work on cavity field control and field-control electronics is presented at the biannual Low-Level RF Workshop [LLRF].

Field control for the electron linacs of free-electron lasers (FELs) is discussed in several PhD theses [Brandt, 2007; Hoffmann, 2008; Schmidt, 2010; Pfeiffer, 2014; Rezaeizadeh, 2016]. However, it should be remembered that the field-control challenges for FEL linacs and high-intensity proton linacs are different. Two high-intensity proton linacs with similar characteristics

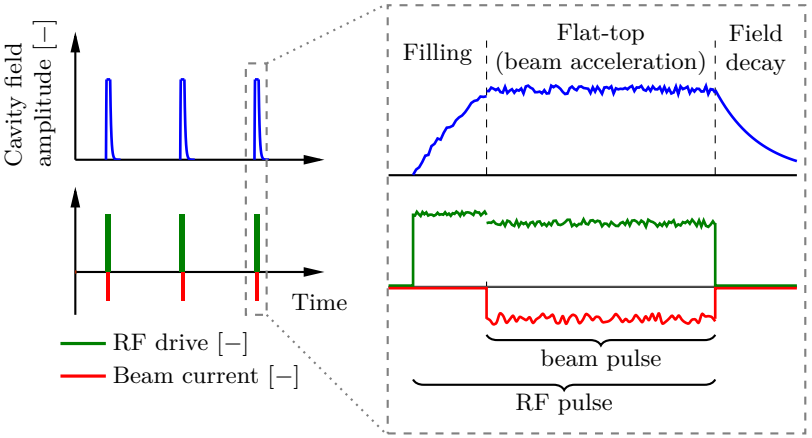


Figure 1.5 *Pulse structure of a pulsed linac.* The RF amplifier is turned on ahead of the beam pulse to build up the cavity field to its nominal level; this is called *filling*. During the beam pulse, or *flat-top* of the RF pulse, the beam gains energy from the cavity field which is maintained by the RF amplifiers. Note that ripple in the RF drive and the beam current act as disturbances that induce cavity field errors. To reduce the errors to acceptable levels it is necessary with feedback from a field controller.

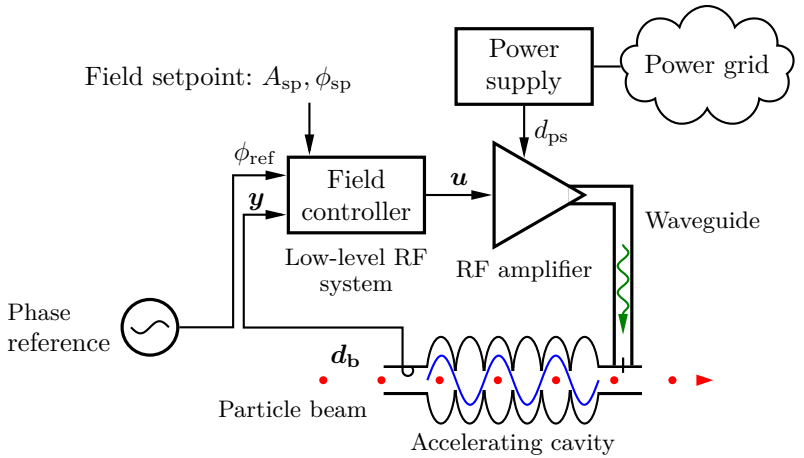


Figure 1.6 Typical field control loop for controlling the amplitude and phase of the electromagnetic field in an accelerator cavity. The amplitude and the phase (relative ϕ_{ref}) of the field measurement $\boldsymbol{y}$ should be kept at the setpoints A_{sp} and ϕ_{sp}. Variations d_{ps} of the amplifier supply voltage and beam-current variations $\boldsymbol{d_b}$ act as disturbances on the system.

and challenges as for the one at ESS are the linac at the Spallation Neutron Source [UT-Battelle, 2006] and Linac4 at CERN [Arnaudon, 2006]. Field control for these linacs is discussed in [Ma et al., 2006] and [Baudrenghien et al., 2014].

For extremely narrow-bandwidth cavities (e.g., at LCLS-II with 16 Hz) it is generally considered that one should use a field controller based on the self-excited-loop architecture introduced in [Delayen, 1978] and more recently discussed in [Doolittle, 2014]. This topic will not be discussed in this thesis.

Complex-coefficient models for cavity field control

This thesis demonstrates the advantages of analyzing the field-control loop as a single-input single-output LTI system with complex coefficients. Complex-coefficient representations of the cavity dynamics are common in the accelerator literature but have seen little use for field control analysis. Instead, equivalent real-coefficient models with two inputs and two outputs are used; these models give less intuition and are more cumbersome. The following was stated in a paper on cavity modeling for field control "The complex cavity representation simulates the circuit approximation but has rather academic importance" [Czarski et al., 2003].

For circular accelerators, the circulating beam gives rise to dynamics that cannot be captured by a complex-coefficient LTI system. The real-signal Pedersen model [Pedersen, 1975, Fig. 3] is traditionally used instead. A complex, *widely*-linear representation could still simplify the *modeling*, but the analysis would be the same as with a real representation.

The situation seems to be similar in the literature on feedback linearization of RF amplifiers (Section 4.6). A complex-coefficient representation is well known, but control analysis is done with equivalent, real-coefficient models [Briffa, 1996; Dawson, 2003].

Control theory for complex-coefficient systems

The complex-baseband representation of bandpass signals and systems has been well-known and extensively used in both engineering and physics [Gabor, 1946]. Particularly in the communications literature where it has been essential since the early days [Stein and Jones, 1967].

However, complex-coefficient systems have received little attention in the control literature, which is natural since the applications are few and little known. One interesting application is vibration suppression of rotating machinery [Byun and Lee, 1988; Ren et al., 2013]. Another is regulation of electric motors and generators [Novotny and Wouterse, 1976; Gataric and Garrigan, 1999; Harnefors, 2007; Dòria-Cerezo and Bodson, 2016]. In the context of the latter application a number of standard tools in automatic control have been generalized to the complex setting: the Nyquist stability criterion

[Gataric and Garrigan, 1999], amplitude and phase margins [Harnefors, 2007], and the root-locus method [Dòria-Cerezo and Bodson, 2016]. There have also been some extensions of theoretic results (for the sake of generality), i.e., LQR [Lancaster and Rodman, 1995], the Routh-Hurwitz criterion [Frank, 1946], and Kharitonov's theorem [Bose and Shi, 1987].

General background material

A good introduction to linear accelerators is [Wangler, 2008] and [Lapostolle and Septier, 1970] is a classic. Good introductions to automatic control are found in [Skogestad and Postlethwaite, 2007; Åström and Hägglund, 2006]. The low-level RF systems where field controllers are implemented have many similarities to the radio systems for wireless communications; a helpful introduction to such systems is [Ellingson, 2016].

Control of circular accelerators

The field control challenges for circular accelerators are different from those for linear accelerators. Most notably in that the dynamics of the circulating beam couple to the dynamics of the cavity field. Another challenge lies in the implementation of so-called RF gymnastics [Garoby, 2010]. These interesting problems are outside the scope of this thesis; for details see [Brennan, 1994; Baudrenghien, 2000; Mastoridis, 2005; Lonza and Schmickler, 2014]. A classic result on the stability of circular accelerators is *Robinson's stability theorem* [Robinson, 1964]. The Nobel prize-winning work on stochastic cooling [Van Der Meer, 1985] is another interesting read.

1.4 Contributions

The main contributions of this thesis are

1. An investigation of how the theory and tools of automatic control extend to the analysis of complex-coefficient systems.
2. A unified treatment of cavity field control for linear accelerators. The treatment is based on complex-coefficient LTI systems which simplifies understanding, analysis, and control design. Specific contributions are
   - A fundamental, energy-based parametrization of the cavity-field dynamics.
   - A suitable normalization of the cavity-field dynamics.
   - A derivation of the transfer function of an elliptical cavity with parasitic same-order modes.
   - A discussion of how to deal with parasitic cavity modes in the controller design.

- The observation that parasitic cavity modes together with beam loading may give a systematic control error of the accelerating mode.
- An in-depth discussion of the field-control problem from an automatic-control perspective and a demonstration of that simple PI and PID controllers often achieve close to optimal performance.
- A comparison of Cartesian and polar feedback for cavity field control.
- A discussion of digital downconversion for cavity field control.
- A discussion of directionality in the field-control problem, illustrated by "instantaneous" phasor diagrams.

3. An intuitive approach to tune disturbance rejecting peak-filters, based on considering complex-coefficient, first-order systems.

A minor contribution is the derivation and discussion of energy-optimal excitation of accelerating cavities.

The author worked closely with the physicists and engineers at ESS on the design and analysis of the field control loops for the ESS linac during 2014–2019. The topics considered in this thesis have been motivated by this work.

1.5 Publications

This section lists the publications on which this thesis is based.

The thesis is a further development of the licentiate[1] thesis

Troeng, O. (2017). *Cavity Field Control for High-Intensity Linear Proton Accelerators*. Licentiate thesis TFRT-3237. Dept. of Automatic Control, Lund University, Lund, Sweden.

Control for complex-coefficient LTI systems (Chapter 4) is a key concept used throughout this thesis. A collection of applications and theoretical results for such systems were presented in the following paper.

Troeng, O., B. Bernhardsson, and C. Rivetta (2017). "Complex-coefficient systems in control". In: *Proceedings of the 2017 American Control Conference*. (Seattle, WA, May 24–26, 2017).

Bo Bernhardsson and Claudio Rivetta independently observed that the field control loop can be analyzed as a complex-coefficient LTI system. Troeng investigated the practical implications of this observation, collected theory and applications of complex-coefficient LTI systems, and realized that also Cartesian feedback linearization can be analyzed as such a system. Troeng wrote

[1] A Swedish degree between MSc and PhD.

the paper, with Bernhardsson and Rivetta contributing helpful suggestions on the manuscript.

Modeling of accelerating cavities (Chapter 6, Appendices C and D) is the foundation for cavity field control.

The first of the following papers introduces an energy-based parameterization of the cavity dynamics, which avoids several issues of previous, equivalent-circuit-based parameterizations. The paper also proposes a normalization of the cavity dynamics that is suitable for field control analysis.

The second paper derives the complex-coefficient transfer function of an elliptical cavity with parasitic same-order modes. It also points out that beam loading variations may induce a systematic field error due to the interaction between the RF system and the parasitic modes, if the complex envelope of the accelerating mode is not directly measured.

Troeng, O. (2019a). "Modeling the accelerating mode of accelerator cavities". Submitted to Physical Review Accelerators and Beams.

Troeng, O. (2019b). "Modeling multicell elliptical cavities with parasitic same-order modes from a field-control perspective". Submitted to Physical Review Accelerators and Beams.

The field control challenges for different linac types (Chapter 3) are different and this point was discussed in the following paper.

Troeng, O., M. Eshraqi, A. Johansson, and S. Pfeiffer (2019). "Field control challenges for different linac types". In: *Proceedings of the 10th International Particle Accelerator Conference*. (Melbourne, Australia, May 19–24, 2019). Not subject to peer review.

Troeng wrote most of the manuscript and made the numerical computations. Mamad Eshraqi contributed to the section on drivers of field stability and Sven Pfeiffer contributed to the sections on cavity field control, in particular those related to free-electron-laser linacs. All co-authors contributed general comments and suggestions on the manuscript.

Digital downconversion (Chapter 12) is necessary for implementing field control on digital LLRF system. The publication below discusses how the requirements on digital downconversion differ between communications applications and control applications. Discussions and some numerical comparisons of different filters for low-latency digital downconversion are provided. Another contribution is that "two-sample reconstruction" is interpreted as standard digital downconversion with a two-tap FIR filter.

Troeng, O. and L. Doolittle (2019). "Low-latency digital downconversion for control applications". In Preparation.

The paper came about after discussions between the two authors. Troeng wrote the manuscript and Doolittle provided helpful comments and suggestions.

Energy-optimal cavity filling (Chapter 13) allows the energy for building up the electromagnetic fields in the accelerating cavities to be reduced.

Troeng, O. and B. Bernhardsson (2017). "Energy-optimal excitation of radio-frequency cavities". In: *Proceedings of the 20th World Congress of the International Federation of Automatic Control*. (Toulouse, France, July 9–14, 2017).

Troeng contributed the main ideas and wrote the paper. Bernhardsson suggested improvements and contributed the self-contained optimality proof.

Disturbance-rejecting peak filters (Chapter 14) are relatively well-known throughout the control community. However, the existing methods for designing these filters provide little intuition into the design trade-offs. The following paper proposed a simple, intuitive, and widely applicable approach to tune disturbance-rejecting peak filters.

The paper was motivated by the need to reduce the impact of power-supply ripple on the cavities at ESS, but the method has general applicability.

Troeng, O. and B. Bernhardsson (2018). "An intuitive design method for disturbance-rejecting peak filters". In: *Proceedings of the 14th IEEE International Conference on Control and Automation (ICCA)*. (Anchorage, AK, June 12–15, 2018).

Bernhardsson, with assistance from Troeng, performed initial experiments that showed that the impact of switching ripple could be reduced with a second-order filter. Troeng came up with the idea to decompose the second-order filter into two first-order complex-coefficient filters that could be interpreted as "bubbles" in the Nyquist diagram, thus enabling intuitive parameter tuning. Troeng wrote the paper and performed the numerical simulations, with Bernhardsson contributing helpful suggestions.

LLRF-workshop presentations

Most of the content in this thesis has been presented to the field control community at the bi-annual Low-Level RF workshop (no peer review).

Troeng, O., B. Bernhardsson, A. J. Johansson, and R. Johansson (2015). *Cavity field control for the European Spallation Source*. Low-Level Radio-Frequency Workshop 2015, Shanghai, Nov. 3–6, 2015.

Troeng, O., B. Bernhardsson, and A. J. Johansson (2017). *Towards a better understanding of cavity field control*. Low-Level Radio-Frequency Workshop 2017, Barcelona, Spain, Nov. 3–6, 2017.

Troeng, O., B. Bernhardsson, and A. J. Johansson (2019). *Perspectives on cavity field control*. Low-Level Radio-Frequency Workshop 2019, Chicago, IL, Sep. 29–Oct. 3, 2019.

Additional publications

Publications by the author, outside the scope of this thesis, are listed below.

Troeng, O. and M. Fält (2018). "A seemingly polynomial-time algorithm for optimal curve fitting by segmented straight lines". In: *Proceedings of the 57th IEEE Conference on Decision and Control*. (Miami Beach, FL, Dec. 17–19, 2018).

Troeng, O. and M. Fält (2019). "Sparsity-constrained optimization of inputs to second-order systems". In: *Proceedings of the 17th European Control Conference*. (Naples, Italy, June 25–28, 2019). IEEE.

Kjellqvist, O. and O. Troeng (2019). *Numerical pitfalls in Q-design*. Submitted to the 21st World Congress of the International Federation of Automatic Control.

Shahriari, Z., B. Bernhardsson, O. Troeng, and G. A. Dumont (2019). *Convergence analysis of iterative learning control using pseudospectra*. Submitted to International Journal of Control.

Gonzalez-Cava, J. M., F. B. Carlson, O. Troeng, A. Cervin, K. van Heusden, G. A. Dumont, and K. Soltesz (2019). *PID versus Youla parameter control of propofol anesthesia*. Submitted to IEEE Transactions on Control Systems Technology.

1.6 Thesis outline

Chapter 2, Background on Automatic Control introduces tools from automatic control that will be used throughout the thesis.

Chapter 3, Linear Accelerators and Cavity Field Control gives examples of high-energy linacs and discusses how the field control challenges differ between a high-intensity ion linac and a free-electron-laser linac.

Chapter 4, Complex-Coefficient Systems discusses how bandpass systems, such as the field control loop, can be modeled in the baseband (by the complex-envelope approximation). Baseband systems have complex-coefficient dynamics, so control theory for such systems is discussed.

Chapter 5, Modeling: The RF System provides baseband models of the components in the field control loop (except the accelerating cavity).

Chapter 6, Modeling: The Accelerating Cavity presents a suitable parametrization of the baseband dynamics of the electromagnetic cavity modes, and discusses various aspects of the cavity dynamics.

Chapter 7, Modeling: Summary collects the results from the previous two chapters and presents three baseband models of the field control loop with different levels of complexity. Disturbances are also discussed.

Chapter 8, Essentials of Cavity Field Control presents the most basic aspects of cavity field control, comparing the field control loop with regulation of a water tank. Control objectives for field control are discussed and the important topic of loop phase adjustment is treated.

Chapter 9, Limits of Field Control Performance compares the field control performance of PI(D)-controllers and the optimal LTI controller.

Chapter 10, Cartesian vs. Polar Feedback compares the standard, linear field controller architecture with a nonlinear polar architecture.

Chapter 11, Field Control of Cavities with Parasitic Modes presents different control strategies to avoid instability from parasitic cavity modes.

Chapter 12, Digital Downconversion for Cavity Field Control discusses the requirements on digital downconversion for cavity field control and analyzes two suitable downconversion filters.

Chapter 13, Energy-Optimal Cavity Filling presents the energy-optimal strategy to build up the electromagnetic cavity fields at the beginning of an RF pulse.

Chapter 14, Ripple-Rejecting Peak Filters presents how to augment a nominal controller with a peak filter for improved rejection of narrowband disturbances and an intuitive tuning method for the filter parameters.

Chapter 15, Conclusions and Future Work sums up the thesis.

2

Background on Automatic Control

This chapter gives short (and somewhat simplified) introductions to concepts in automatic control that will be used extensively throughout this thesis. The following topics will be discussed: transfer functions, feedback, Nyquist's stability criterion, and norms of signals and systems. For a proper introduction to these topics see Chapter 2 in [Åström and Hägglund, 2006]. An introduction to control written for physicists is [Bechhoefer, 2005].

2.1 Linear time-invariant systems

The dynamics of a linear, time-invariant (LTI) systems can be represented in the time domain by its impulse response $g(t)$, or in the frequency domain by its transfer function $G(s)$ [Åström and Hägglund, 2006]. The two representations are related via the Laplace transform. In this thesis we will mostly use the frequency-domain representation since it gives much insight into the field control problem.

An important property of (stable) linear time-invariant systems is that a sinusoidal input signal gives a sinusoidal output signal with the same frequency (after transients). The amplitudes and phases of the two sinusoids are related by the system's transfer function $G(s)$ as in Figure 2.1.

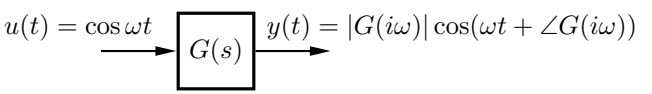


Figure 2.1 The transfer function $G(s)$ of an LTI system relates the Laplace transforms of the input and output signals via $Y(s) = G(s)U(s)$.

Example, transfer function of a first-order system

Assume that a first-order LTI system is defined by the following relationship between the input signal $u(t)$ and output signal $y(t)$,

$$\dot{y}(t) + 10y(t) = 10u(t). \tag{2.1}$$

By taking the one-sided Laplace transform $\mathcal{L}$ of both sides (for zero initial conditions) we obtain

$$sY(s) + 10Y(s) = 10U(s), \tag{2.2}$$

where $U(s) = \mathcal{L}u$ and $Y(s) = \mathcal{L}y$. By introducing the transfer function

$$G(s) = \frac{10}{s+10}$$

we can re-write (2.2) as

$$Y(s) = G(s)U(s).$$

Example, transfer function of a time delay

The system given by a time delay τ,

$$y(t) = u(t-\tau),$$

is after Laplace transformation seen to have the transfer function

$$G(s) = \mathrm{e}^{-s\tau}.$$

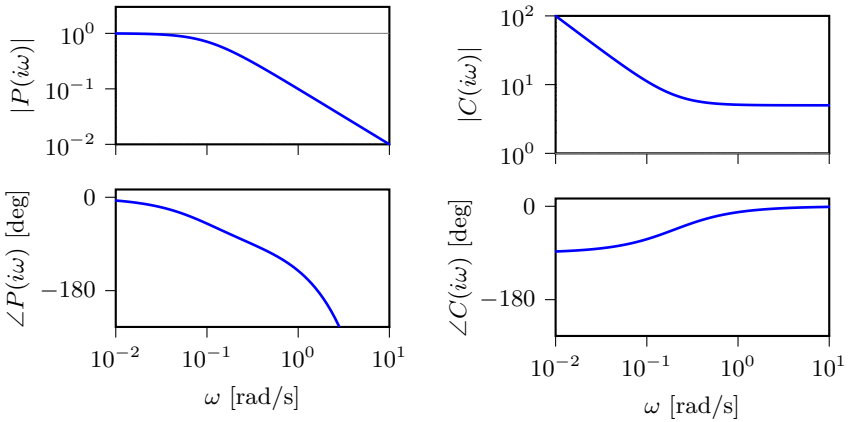
Bode diagrams

The Bode diagram is a convenient tool for visualizing transfer functions and obtain insight into the dynamics of the underlying LTI systems. The Bode diagram shows a transfer function's magnitude and phase as a function of frequency, with the magnitude and frequency on logarithmic scales.

Example: The Bode diagram for the process model

$$P(s) = \frac{10}{s+10}\mathrm{e}^{-s} \tag{2.3}$$

(the system (2.1) with a one-second delay) is shown in Figure 2.2a. The magnitude curve in Figure 2.2a shows that low-frequency variations ($\omega < 0.1\,\mathrm{rad/s}$) in the input signal give rise to output variations of the same amplitude, but that variations of higher frequencies are attenuated.



(a) Process $P(s) = \mathrm{e}^{-s}/(10s+1)$.

(b) Controller $C(s) = 5 + 1/s$.

Figure 2.2 Bode diagram for the process $P(s)$ in (2.3) and the PI controller (2.4) with $K = 5$ and $T_i = 5$. The upper diagram shows the amplitude of the transfer function $P(i\omega)$ as a function of frequency, and the lower diagram shows the phase of the transfer function.

2.2 Feedback

Given a stable LTI system we could in principle make its output $y(t)$ equal to a constant setpoint y_{sp} by an appropriate choice of the input signal $u(t)$. There are several problems with this approach: (1) it requires perfect knowledge of the process dynamics $P(s)$; (2) if there are disturbances acting on the process, see Figure 2.3a, there will be errors on y; and (3) if the system dynamics are slow it will take a long time for $y(t)$ to converge to y_{sp}. All these problems can be mitigated by introducing a feedback controller $C(s)$ that adjusts the input signal $u(t)$ based on the control error $e(t) = y_{\text{sp}} - y(t)$; see Figure 2.3b.

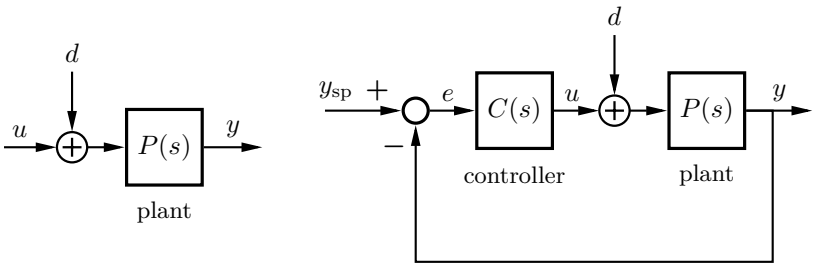
To illustrate the advantages of feedback, we consider a proportional-integral (PI) controller, which gives the control signal

$$u(t) = K\left(e(t) + \frac{1}{T_i}\int_{-\infty}^{t} e(t)\,dt\right).$$

The first term is proportional to the control error, and the second term is proportional to the integral of previous control errors.

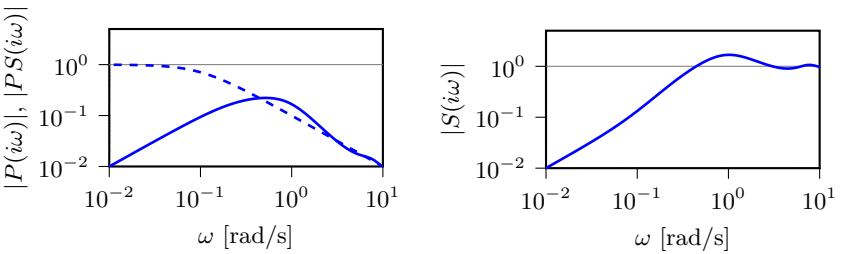
In the Laplace domain the PI controller takes the form

$$C(s) = K\left(1 + \frac{1}{sT_i}\right) \tag{2.4}$$



(a) No feedback. (b) With feedback.

Figure 2.3 Illustration of control without feedback (open-loop control), and with feedback (closed-loop control).



(a) Transfer function from disturbances d to the measured signal y; *dashed line:* P (without feedback), *solid line:* PS (with feedback).

(b) Sensitivity function $S = 1/(1 + PC)$.

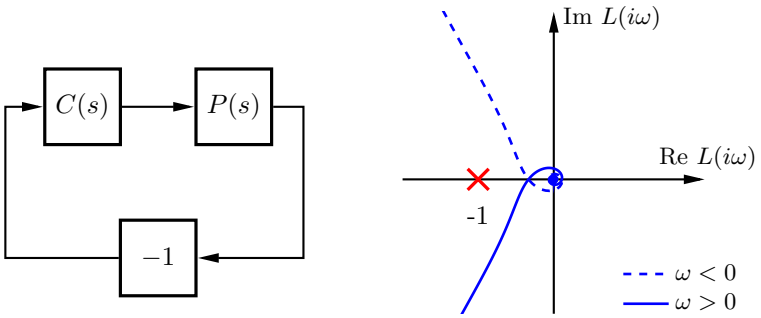
Figure 2.4 Bode magnitude diagrams for two closed-loop transfer functions when the process P and controller C are given by the transfer functions in Figure 2.2; the phases of the transfer functions are not shown since it is rarely of interest for closed-loop transfer functions.

see Figure 2.2b for the Bode diagram of the controller (2.4) with $K = 5$ and $T_i = 5$.

The sensitivity function

Without feedback (Figure 2.3a) the transfer function from disturbances d to control errors e is given by $P(s)$; by introducing a feedback controller $C(s)$ (Figure 2.3b) the transfer function from disturbances to the control errors becomes instead $P/(1+PC)$; Figure 2.4a shows the improved attenuation of low-frequency disturbances that is provided by the feedback controller.

The difference in disturbance attenuation from introducing feedback is



(a) Considered feedback interconnection.

(b) The Nyquist curve of the open-loop system $L(s) = C(s)P(s)$.

Figure 2.5 Illustration of the Nyquist criterion for the process P and controller C in Figure 2.2. From (b) it is seen that the Nyquist curve stays to the right of the point -1, thus, according to the Nyquist stability criterion, the closed-loop system in (a) is stable.

quantified by the so-called sensitivity function

$$S = \frac{1}{1 + PC}, \tag{2.5}$$

see Figure 2.4b.

Feedback stability

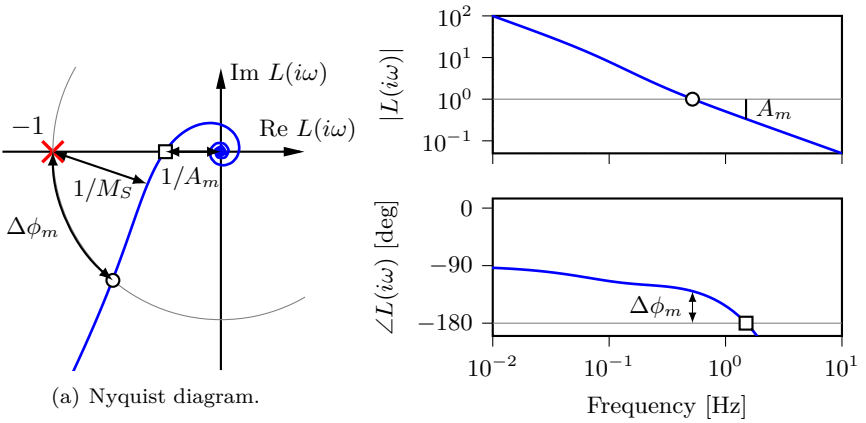
Stability is a fundamental concern for systems involving feedback. Instability results if the controller compensates control errors too aggressively. A typical manifestation of instability is that the system starts to oscillate uncontrollably. For the applications in this thesis, closed-loop stability is conveniently verified by the Nyquist stability criterion.

The Nyquist Stability Criterion (Simplified Version)
Let $L(s) = C(s)P(s)$ be a system's open-loop transfer function and assume that $L(s)$ is stable (has no poles in the right half-plane). Then the closed-loop system is stable if and only if the Nyquist curve $L(i\omega), \omega \in (-\infty, \infty)$ does not encircle and stays to the right of the critical point -1. □

See Figure 2.5 for an illustration of the Nyquist stability criterion.

Robustness

The Nyquist stability criterion shows that a feedback system is stable when the open-loop Nyquist curve stays to the right of the point -1. Hence it is not surprising that the shortest distance between the Nyquist curve and the



(a) Nyquist diagram.

(b) Bode diagram.

Figure 2.6 Illustration of different robustness measures in the Nyquist and open-loop Bode diagrams.

point -1 is a good indication of how robust the system is to modeling errors and process variations. Since the distance in the complex plane between the Nyquist curve and the point -1 is given by $|L(i\omega) - (-1)| = 1/|S(i\omega)|$, we see that the maximum value of the sensitivity function (2.5),

$$M_S := \sup_{\omega} |S(i\omega)| \tag{2.6}$$

indicates the robustness of a feedback interconnection. The smaller the value of M_S, the more robust is the feedback interconnection.

Amplitude and phase margins are two alternative robustness measures. The *amplitude margin* A_m quantifies the maximum permissible gain increase of the open-loop system without instability and the *phase margin* $\Delta\phi_m$ quantifies the maximum permissible phase drop of the open open-loop system without instability, see Figure 2.6.

Even if the amplitude and phase margins are large, the closed-loop system could have poor robustness [Åström and Murray, 2010, Figure 9.11]. On the other hand, a small value of M_S always guarantees reasonable amplitude and phase margins [Skogestad and Postlethwaite, 2007, Eq. (2.39)],

$$A_m \geq \frac{M_S}{M_S - 1}$$
$$\Delta\phi_m \geq 2 \arcsin \frac{1}{2M_S}.$$

For example, $M_S = 1.6$ guarantees an amplitude margin of *at least* 2.6, and a phase margin of *at least* 36°. Thus, a robustness specification in terms of

M_S is more general than amplitude and phase margins, and typically it is also more convenient to work with.

Fundamental limitations

It is important to remember that there are *fundamental* limitations on the achievable control performance [Stein, 2003]. If both the process and the controller are linear and time invariant, then one such limitation is given by Bode's integral formula (if both the open- and closed-loop systems are stable, and the gain of open-loop system rolls off sufficiently fast at high frequencies),

$$\int_{-\infty}^{\infty} \log |S(i\omega)| \, d\omega = 0, \tag{2.7}$$

where S is the sensitivity function (2.5).

Bode's integral formula can be viewed as a *conservation law of sensitivity* [Stein, 2003]. If the feedback pushes the sensitivity function below 1 at some frequencies it will unavoidably pop up above 1 at other frequencies; fittingly, this is known as the waterbed effect. The waterbed effect is seen in Figure 2.4a, where the control system reduces disturbances at low frequencies, but amplifies them around 1 rad/s.

2.3 Norms of signal and systems

In what follows it will be convenient to use the 2-norm of signals to quantify their energy/rms-value, and the $\mathcal{H}_\infty$-norm of closed-loop transfer functions to quantify the robustness of feedback interconnections. Below we provide the definitions of these norms for the single-input single-output case; for the multi-input multi-output case, as well as for more details, see [Skogestad and Postlethwaite, 2007, 2.8.1, 4.10, A.5].

2-norm of signal

The 2-norm of a scalar signal y with finite support $[0, T]$, and with one-sided Laplace transform Y, is defined as

$$\|y\|_2 = \|Y\|_2 := \sqrt{\int_0^T |y(t)|^2 \, dt} \tag{2.8a}$$

$$= \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(i\omega)|^2 \, d\omega}. \tag{2.8b}$$

From (2.8a) we see that the rms-value of y equals $\|y\|_2 / \sqrt{T}$.

If y is a stochastic signal from a stationary process with spectrum Y, the norm (2.8b) instead corresponds to the *expected* rms value of the signal as $T \to \infty$.

The 2-norm enable a compact representation of control performance—for example, if y is the amplitude error of the cavity field then $||Y||_2$ is the rms value of y, which is what needs to be kept within specified limits.

$\mathcal{H}_\infty$-norm of a system

The $\mathcal{H}_\infty$-norm (H-infinity norm) of a linear, time-invariant, single-input single-output system $G(s)$ is given by

$$||G||_\infty = \begin{cases} \sup_\omega |G(i\omega)| & \text{if } G \text{ is stable} \\ \infty & \text{if } G \text{ is unstable} \end{cases}. \tag{2.9}$$

Using the $\mathcal{H}_\infty$-norm, the robustness measure in (2.6) can be written

$$M_S := ||S||_\infty .$$

3

Linear Accelerators and Cavity Field Control

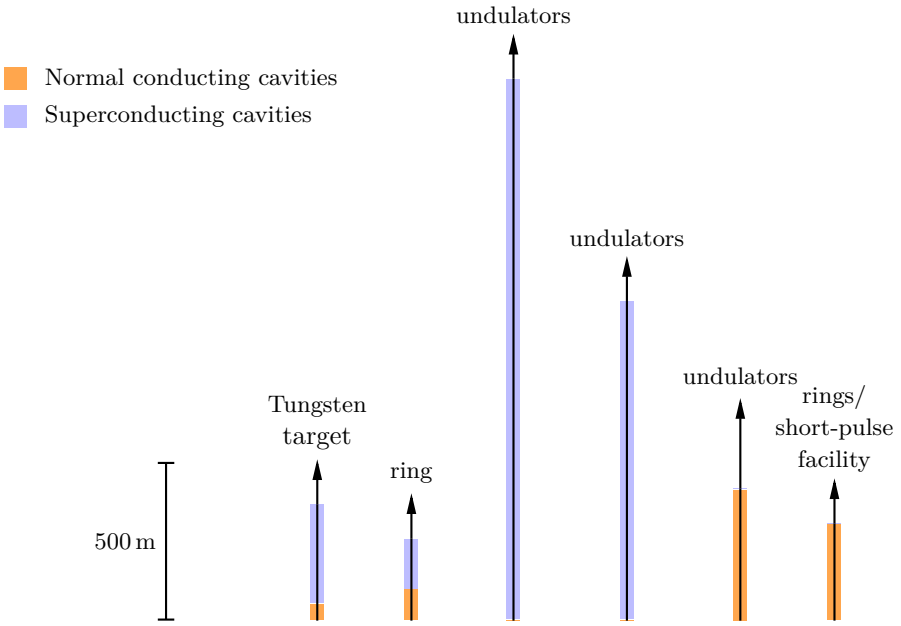
3.1 Linear Accelerators

High-energy linear particle accelerators have many important applications.

1. As the driver of free-electron lasers
2. As the driver of spallation sources
3. For high-energy physics experiments
    - Injecting into circular accelerators that reach even higher energies
    - Rare isotope production [Bollen, 2010]
    - Neutrino production [Baussan, 2014]
    - Linear colliders [LCC 2019]
4. For nuclear-energy applications
    - Transmutation of long-lived nuclear waste [Abderrahim et al., 2012]
    - Accelerator-driven fission reactors [Carminati et al., 1993]
    - Materials testing for fusion reactors [Garin and Sugimoto, 2009].

A large number of lower-energy linacs are used for cancer treatment and semiconductor manufacturing. However, the requirements on field control for these linacs are quite relaxed since they only have one or a few accelerating cavities.

Linear accelerators are generally classified as either ion linacs or electron linacs [Turner, 1994]. Parameters for a selection of high-energy linacs are shown in Figure 3.1.



	ESS	SNS	Eu-XFEL	LCLS-II SCL	Swiss-FEL	MAX IV
Accelerated particle	H^+	H^-	e^-	e^-	e^-	e^-
Avg. Beam Power [MW]	5	1.4 (2.8)	0.6	0.4	†	‡
Final Energy [GeV]	2	1 (1.3)	17.5	4	5.8	3.6
Beam Current* [mA]	62.5	25 (38)	5	0.1	†	‡
Pulse Length [ms]	2.86	1.0	0.65	CW	†	‡
Pulse Frequency [Hz]	14	60	10	CW	100	10/100
Bunch frequency (during pulse) [MHz]	352.21	402.5	4.5	0.9	†	‡
No. cavities	155	96	≈800	≈300	121	39
No. field control loops	155	96	27	≈300	36	0

* Averaged over the flat-top of the RF pulse.
† One or two 200 pC electron bunches per 3 µs RF pulse.
‡ One or three 100 pC electron bunches per 4.5 µs RF pulse.

Figure 3.1 Comparison between the parameters of a selection of high-energy linacs at different facilities: the European Spallation Source, Lund, Sweden [Peggs et al., 2013; Garoby, 2017]; Spallation Neutron Source, Oak Ridge, TN [UT-Battelle, 2006; Champion et al., 2017]; the European XFEL, Hamburg, Germany [Altarelli, 2007]; Linac Coherent Light Source II, Stanford, CA [Galayda, 2014]; and MAX IV, Lund, Sweden [MAX IV, 2010].

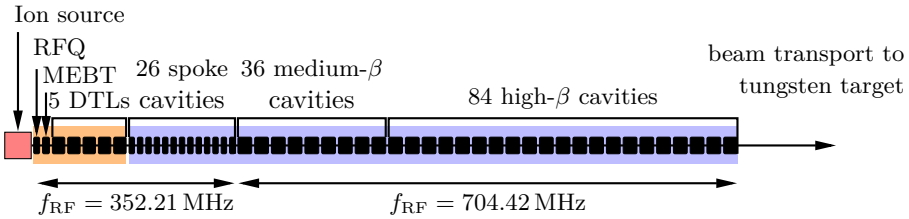


Figure 3.2 Block diagram of the proton linac at ESS. Depending on the proton velocity different accelerating cavities are used. The cavities in the orange section are made of copper (or copper plated) and are operated at room temperature. The downstream 146 cavities (blue color) are made of niobium and kept at 2 kelvin. This makes them superconducting, which reduces resistive losses and allows stronger electric fields.

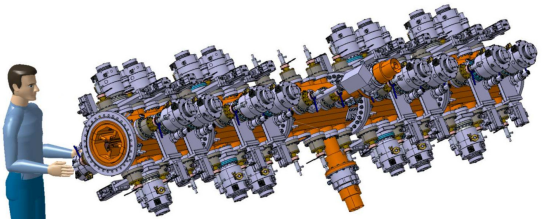
3.2 Example of an ion linac: The ESS linac

A schematic of the ESS linac is shown in Figure 3.2. Protons (H^+) are generated in the ion source by ionizing hydrogen gas with microwave radiation. The protons drift through the low-energy beam transport (not shown) to the radio-frequency quadrupole (RFQ). The RFQ is a special type of accelerating cavity whose electromagnetic field both bunches, focuses and accelerates the beam; the bunching is what allows the beam to be accelerated by the oscillating electromagnetic fields in the downstream cavities.

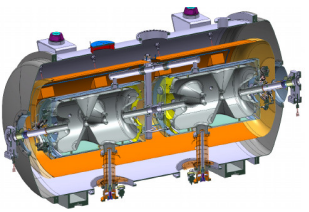
Downstream the RFQ, three buncher cavities in the medium-energy beam transport (MEBT) keep the beam bunched before five drift tube linacs (DTLs) accelerate the bunches from 9 % to 41 % of the speed of light.

These first 9 cavities (RFQ, bunchers, DTLs) are made of copper and are operated at room temperature. The following 146 cavities are made of niobium and are kept at a temperature of 2 K. This makes them superconducting which gives very low electric losses and allows strong electric fields. Three different superconducting cavities are used depending on the velocity of the proton bunches: spoke cavities, medium-β elliptical cavities, and high-β elliptical cavities. The RF frequency of the RFQ, bunchers, DTLs and spoke cavities is 352.21 MHz and for the elliptical cavities it is doubled to 704.42 MHz. Three of the cavity types for the ESS linac are shown in Figure 3.3.

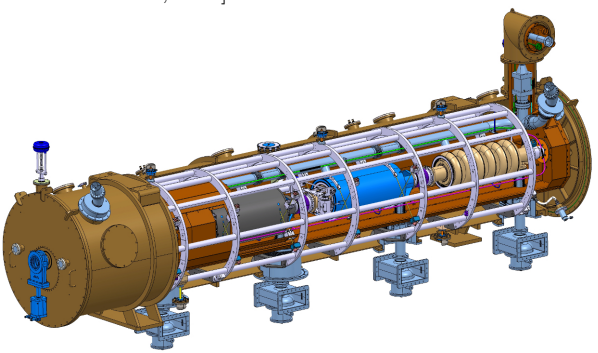
The motivation for the pulsed operation of the ESS linac is to allow higher peak intensities and to give a time separation of the neutron energies that arrive at the experiment stations. The specific pulse rate of 14 Hz is related to the time it takes for the neutrons to travel from the neutron target to the instrument stations through the neutron guides [Peggs et al., 2013].



(a) Radio-frequency quadrupole with vacuum pumps and tuners [Alberi and Lacroix, 2015].



(b) Cryomodule with two spoke cavities [Bousson et al., 2014].



(c) Cryomodule with four elliptical medium-β cavities [ESSb, 2017].

Figure 3.3 Three of the six types of accelerating cavities that will be used for the ESS linac. The spoke cavities and the elliptical cavities are kept in cryomodules to allow them to be cooled to 2 K, which makes them superconducting. Image credit: ESS.

3.3 Drivers of cavity field stability in linacs

Field-stability drivers of ion linacs

The beam in an ion linac needs to be accelerated over hundreds of meters before its velocity starts to approach c (the speed of light). Therefore the synchronous phase (the nominal phase between the particle bunches and the cavity field) needs to be selected so that the cavity fields provide longitudinal focusing (in addition to acceleration). If the electromagnetic field is longitudinally focusing then it must be transversally *de*focusing, which follows from Earnshaw's theorem [Wangler, 2008, Sec. 7.1]. Unless the amplitudes and phases of the fields are accurately controlled the beam is focused differently in the longitudinal and transverse planes which causes a mismatch that leads to halo production and loss of halo particles [Plum, n.d.] Proton losses on the structure cause radioactivation which delays hands-on maintenance and therefore reduces the availability of the linac. To allow hands-on maintenance

(within a reasonable time) the beam losses should not exceed 1 W/m [Mokhov and Chou, 1999]. It is often challenging to achieve this level of beam loss for multi-megawatt proton linacs. Many high-intensity proton linacs are *beam-loss limited*, i.e., many subsystem requirements, as those on field stability, are driven by the need to limit beam losses.

Experiments performed at LANSCE linac indicated that beam losses had an exponential dependence on field errors [Rybarcyk and McCrady, 2016]. Note that the beam losses depend on the beam current which means that cavity field control will become more important for future linacs with higher beam currents.

Example: Derivation of field-error requirements for the ESS linac[1]. For the 5 MW ESS linac the beam loss requirement of 1 W/m translates to that only 1–2 out of 10 000 particles can be lost. This was challenging to achieve in the linac design [Levinsen et al., 2016]. The field-error requirements for the ESS linac were derived by simulating the beam loss for different levels of field errors using the simulation software Tracewin [Uriot and Pichoff, 2015]. Thousands of bunches had to be simulated for each level of field error to obtain sufficient statistics. To accurately account for halo formation, which is a main reason for beam loss, required simulations with 10^5–10^6 macro particles (each representing some 1000 protons). The simulations were numerically demanding due to the large number of particles and the nonlinear Coulomb forces between them. To simulate the propagation of a single proton bunch along the linac took about one hour on a high-performance computer and completing the error studies took many months.

Finally, the beam physics group at ESS set the requirements on cavity field errors to 0.2 % rms in amplitude and 0.2° rms in phase for the 9 normal conducting cavities and 0.1 % and 0.1° rms for the 146 superconducting cavities [Eshraqi and Levinsen, 2016; Levinsen et al., 2016]. The difference in error levels was motivated by that it is easier to obtain small field errors for superconducting cavities, which we will see in Chapter 9. The error requirements on the amplitudes and phases were selected so that there would be equal contributions to beam loss from each type of error.

Certain disturbances give rise to field errors that are correlated between multiple cavities. For example, beam current ripple gives rise to errors that are correlated between all the cavities and ripple in the voltage supplied by the power supplies to a group of RF amplifiers gives rise to correlated field errors in the corresponding cavities. Error correlations were not accounted for in the design of the ESS linac due to a lack of disturbance models.

[1] Personal communication with Mamad Eshraqi, ESS.

Field-stability drivers of free-electron-laser linacs

For FEL linacs it is mainly beam parameters such as the bunch-to-bunch energy spread, the bunch compression in the injector, and the bunch arrival time at the undulator that dictate the required field stability. The European XFEL [Altarelli, 2007], LCLS-II [Doolittle, 2015], and SwissFEL [Ganter, 2012] require field errors smaller than 0.01 % (rms) in amplitude and 0.01° (rms) in phase.

Beam losses are less of an issue for FEL linacs than for high-intensity ion linacs due to[1]: (1) lower *average* beam current; (2) less Coulomb repulsion since the electrons are relativistic; (3) less transversal defocusing from the electric field; and (4) electrons induce less less radioactivity [Schmidt, 2016].

Derivation of field-error requirements for FEL linacs. The effect of cavity field errors on the energy spread and the arrival time jitter can be computed using the formulas in [Altarelli, 2007, Sec. 4.8.1.1] and [Merminga and Kraft, 1996], which is quite straight-forward relative to the beam-loss computations for high-intensity ion linacs. Examples of field error requirements for FEL linacs can be found in [Ganter, 2012, Sec. 2.4.4], [Stohr, 2011, Table 6.11] and [Altarelli, 2007]. These are roughly on the order of 0.01%/0.01°.

3.4 Field control challenges

It is interesting to note the diversity of field-control challenges that may arise from the high-level design decisions of a linac. The challenges for the high-intensity proton linac at ESS linac was discussed in Section 1.2. The challenges for the SNS linac are quite similar but larger field errors are tolerated. Below we briefly mention some field control challenges for the other linacs in Figure 3.1.

The European XFEL in Hamburg, Germany, was designed for high brilliance at X-ray-wavelengths [Altarelli, 2007]. For cost reasons, every RF amplifier drives *thirty-two* superconducting TESLA cavities [Schmidt, 2010; Pfeiffer, 2014]. The cavity bandwidth of 140 Hz means that calibration and cavity frequency control need careful consideration. There are additional feedback challenges in the control of the electron gun.

Linac Coherent Light Source II in Stanford, California, will have a superconducting linac operating in CW mode with a bunch repetition rate of almost 1 MHz [Galayda, 2014]. The cavities will have bandwidths of 16 Hz and will be driven by dedicated solid-state amplifiers. The extremely low bandwidth calls for field control by self-excited loops and makes cavity frequency control a challenge [Doolittle, 2015].

SwissFEL in Villigen, Switzerland, was designed to be an affordable and compact X-ray FEL capable of 1 angstrom wavelengths [Ganter, 2012; Milne, 2017]. Its operation principle is quite different from the two previous FELs—the RF pulses are extremely short and at most two electron bunches are accelerated during each RF pulse. The RF pulses are too short for intra-pulse feedback so it is necessary with pulse-to-pulse corrections using iterative learning control to deal with temperature transients that act as disturbances [Rezaeizadeh et al., 2017].

The MAX IV injector linac in Lund, Sweden, has klystrons with a small phase-pushing factor of $4\,°/\%$ and klystron power supplies with less than $\pm 0.01\,\%$ voltage ripple. This ensures that the pulse-to-pulse variations are small. Accurate control of the cooling and the air conditioning keeps long-term drifts small. The careful disturbances mitigation enables sufficient cavity field stability to be achieved with a simple and robust open-loop approach [MAX IV, 2010][2].

[2] Personal communication with Aleksandar Mitrovic, MAX IV.

4

Complex-Coefficient Systems

The analysis of a system with high-frequency input and output signals (e.g., a field control loop) is often simplified by considering how the system relates the complex envelopes of the input and output signals (Sections 4.1–4.2). If the original system is linear and time-invariant then the complex envelopes of the input and outputs signals are related via a *complex-coefficient* LTI system. Analysis and control design for complex-coefficient systems is discussed in Sections 4.3–4.4. The disturbances acting on the field control loop will have certain directionality, so we discuss such signals in Section 4.5. Lastly, some enjoyable examples of control applications whose analyses are simplified by a complex-coefficient representation are given in Section 4.6.

Certain effects in the field control loop give rise to dynamics that cannot quite be represented by a complex-coefficient LTI system since linearity only holds for multiplication by real numbers. This class of systems will be referred to as *widely linear* systems. We will only refer to this concept occasionally so the details have been placed in Appendix A.

4.1 Baseband transformation of bandpass signals

A high-frequency sinusoid

$$y_c(t) = A(t)\cos(\omega_c t + \phi(t)) \tag{4.1}$$

whose amplitude $A(t)$ and phase $\phi(t)$ vary slowly relative to the oscillation frequency ω_c has a narrowband spectrum around the oscillation frequency ω_c. This type signal is often called a *bandpass* signal.

The signal (4.1) can be written

$$y_c(t) = \mathrm{Re}\{\boldsymbol{y}(t)\mathrm{e}^{i\omega_c t}\} \tag{4.2}$$

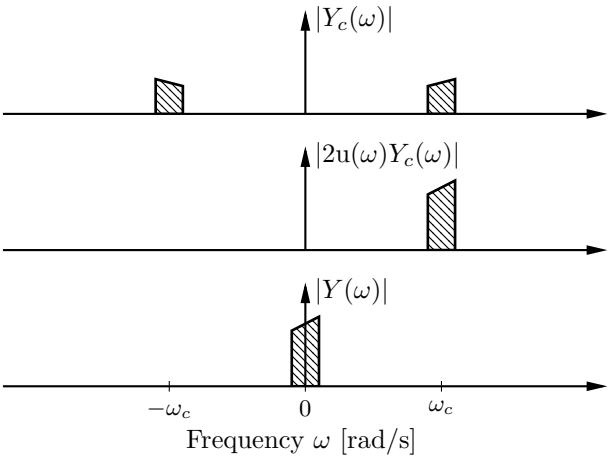


Figure 4.1 Frequency-domain illustration of how a bandpass signal $y_c(t)$ (with Fourier transform Y_c) is transformed to a baseband signal $\boldsymbol{y}(t)$ (with Fourier transform Y). The middlemost figure shows the signal after multiplication by 2 and truncation of negative frequency components ($\mathrm{u}(\cdot)$ denotes the Heaviside step function). The asymmetry of $Y(\omega)$ indicates that the corresponding time-domain signal $\boldsymbol{y}(t)$ is complex valued.

where $\boldsymbol{y}(t) = A(t)\mathrm{e}^{i\phi(t)}$ is a slowly-varying complex signal. For analysis it is often more convenient to work with $\boldsymbol{y}$ rather than y_c.

Note that $\boldsymbol{y}$ is not uniquely defined by the relation (4.2), [Schreier and Scharf, 2010, Example 1.4]. To avoid ambiguity we follow the communications literature [Proakis and Salehi, 2002] with the following definition.

Definition 4.1 The *complex envelope*, or the *equivalent baseband signal* of a finite-energy signal y_c is the signal $\boldsymbol{y}$ with Fourier transform

$$Y(s) = \begin{cases} 2Y_c(s + i\omega_c) & \text{if } \omega \geq -\omega_c \\ 0 & \text{otherwise} \end{cases},$$

where $Y_c(\omega)$ is the Fourier transform of y_c.

This definition corresponds to zeroing out Y_c at negative frequencies, multiplying by 2, and then translating along the frequency axis by $-\omega_c$, See Figure 4.1 for a visualization of these steps. The corresponding approach in the time domain is to first compute the *pre-envelope* $z(t) = y_c(t) + i(\mathcal{H}y_c)(t)$, where $\mathcal{H}y_c$ denotes the Hilbert transform of y_c, and then recover the complex envelope as $\boldsymbol{y}(t) = z(t)\mathrm{e}^{-i\omega_c t}$.

Real-time recovery of the complex envelope

In implementations that recover the envelope of a narrowband signal y_c, the truncation and the shift operations are applied in the opposite order of Definition 4.1. First the spectrum of $y_c(s)$ is frequency-shifted and scaled through multiplication by $2\mathrm{e}^{-i\omega_c t}$ (follows from standard rules for Fourier transforms). Then the resulting signal is low-pass filtered. To strictly adhere to the definition, only the frequency content at $\leq -\omega_c$ should be removed, but for practical purposes, any sensibly chosen low-pass filter does the job.

To illustrate the procedure, let $y_c(t) = \mathrm{Re}\left\{\boldsymbol{y}(t)\mathrm{e}^{-i\omega_c t}\right\}$. After frequency-shifting and scaling, we get

$$\begin{aligned} y_c(t) \cdot 2\mathrm{e}^{-i\omega_c t} &= \frac{1}{2}\left[\boldsymbol{y}(t)\mathrm{e}^{i\omega_c t} + \boldsymbol{y}^*(t)\mathrm{e}^{-i\omega_c t}\right] \cdot 2\mathrm{e}^{-i\omega_c t} \\ &= \boldsymbol{y}(t) + \boldsymbol{y}^*(t)\mathrm{e}^{-2i\omega_c t}. \end{aligned} \tag{4.3}$$

Low-pass filtering the signal (4.3) removes the double-frequency component $\boldsymbol{y}^*(t)\mathrm{e}^{-2i\omega_c t}$. With a non-ideal low-pass filter this component will not be fully removed and $\boldsymbol{y}$ will be slightly distorted, but these imperfections tend to be negligible in practice. This method of recovering the complex-envelope $\boldsymbol{y}$ is known as quadrature downconversion, or downconversion to baseband. See [Ellingson, 2016, p. 5.3] for a further discussion. For cavity field control, quadrature downconversion is typically implemented digitally. We discuss this further in Section 5.1.1 and Chapter 12.

Remark 4.1 (Telecommunications terminology) In the telecommunications literature, the real and imaginary parts of a baseband signal $\boldsymbol{y}$ are referred to as the in-phase and quadrature components. The reference frequency ω_c is referred to as the carrier frequency.

Examples of equivalent baseband signals

Example: Illustration of baseband signal. The impulse response of the system $G_c(s) = -30/(s^2 + 2s + 30^2)$ is approximately given by

$$y_c(t) = \mathrm{e}^{-t}\cos(30t + \pi/2), \tag{4.4}$$

The equivalent baseband signal of y_c with respect to $\omega_c = 29$ is given by

$$\boldsymbol{y}(t) = \mathrm{e}^{-t}\mathrm{e}^{i(t+\pi/2)}. \tag{4.5}$$

Both the original bandpass signal y_c and the equivalent baseband signal $\boldsymbol{y}$ are illustrated in Figure 4.2.

Example: A pure sinusoid with frequency ω_c,

$$y_c(t) = A_0\cos(\omega_c t + \phi_0),$$

has a constant complex envelope

$$\boldsymbol{y}(t) = A_0\mathrm{e}^{i\phi_0}.$$

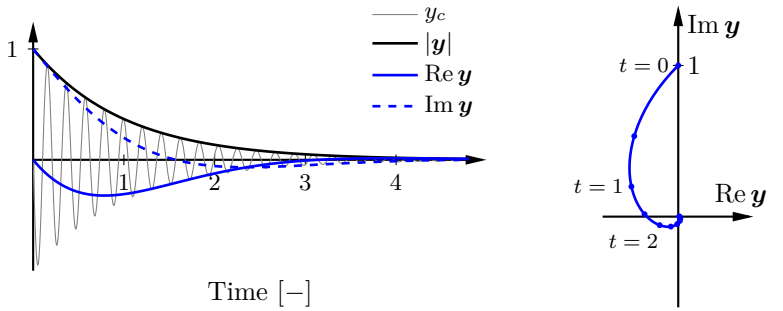


Figure 4.2 Illustration of the baseband signal in (4.5) as a function of time (left) and in the complex plane (right).

Example: A sinusoidal bandpass signal with frequency $\omega_c + \nu$,

$$y_c(t) = \cos\left((\omega_c + \nu)t\right),$$

has the complex envelope

$$\boldsymbol{y}(t) = \mathrm{e}^{i\nu t}.$$

This signal is often referred to as a *complex sinusoid* with frequency ν. Depending on the sign of ν, this frequency is either positive or negative.

This example shows that there is nothing strange with negative frequencies. A (baseband) frequency $\nu > 0$ corresponds to physical signals with a higher frequency than the carrier frequency, and a frequency $\nu < 0$ corresponds to a signal with a lower frequency than ω_c.

Example: A sinusoid with slowly varying amplitude,

$$y_c(t) = A(t)\cos\left(\omega_c t\right),$$

has the equivalent baseband representation

$$\boldsymbol{y}(t) = A(t).$$

Example: A sinusoid with slowly varying phase,

$$y_c(t) = \cos\left(\omega_c t + \phi(t)\right),$$

has the equivalent baseband representation

$$\boldsymbol{y}(t) = \mathrm{e}^{i\phi(t)}.$$

For small values of $\phi(t)$, we have that $\boldsymbol{y}(t) \approx 1 + i\phi(t)$.

4.2 Baseband transformation of bandpass systems

Assume that the LTI system $G_c(s)$ has narrowband dynamics around a frequency ω_c. The relationship between an input signal $U_c(s)$ and an output signal $Y_c(s)$ is, as for any LTI system, given by

$$Y_c(s) = G_c(s)U_c(s).$$

We would like the complex envelopes U and Y of U_c and Y_c to be related in the same way. That is, the *equivalent baseband system* $G(s)$ should satisfy the relation[1]

$$Y(s) = G(s)U(s). \tag{4.6}$$

Note that there is some freedom in the choice of $G(s)$ since baseband signals $U(s)$ are zero for frequencies $\leq -\omega_c$. One option is to take

$$G(s) = G_c(s + i\omega_c).$$

In control applications, where low-order system models are desirable, it is not practical to insist on that the frequency response is exactly zero for frequencies $\omega \leq -\omega_c$ since this would require an infinite-dimensional model.

4.2.1 Baseband transformation of rational LTI systems

Consider a bandpass LTI system with narrowband dynamics around frequencies $\pm\omega_c$ and assume that it has the state-space realization (A_c, B_c, C_c, D_c), which correspond to a transfer function $G_c(s) = C_c(sI - A_c)^{-1}B_c + D_c$. It is clear that the equivalent baseband system has a state-space realization $(A_c+i\omega_c I, B_c, C_c, D_c)$ since this gives a transfer function $G(s) = G_c(s+i\omega_c)$.

However, the system $G(s)$ has fast dynamics around $-2\omega_c$, which makes it numerically challenging to simulate and analyze the system. Thus, some kind of model reduction should be applied to $G(s)$ to get a low-order approximation without high-frequency dynamics.

A more explicit construction for obtaining a low-order model is to do a partial fraction decomposition of the transfer function $G_c(s)$ and then baseband transform each term individually as in the next subsection. This gives a good approximation when the damping ratio of G_c's poles are small, which typically is the case.

[1] Note that a different definition of the equivalent baseband system is common in the telecommunications literature [Proakis and Salehi, 2002], where it is defined in analogy with the equivalent baseband signals: by including a factor 2 and truncating the system dynamics at frequencies $\leq -\omega_c$. This makes the impulse response of a baseband system equal to the baseband transformed impulse response of the corresponding bandpass system. However, this gives an inconvenient factor 1/2 in the relation (4.6).

4.2.2 Baseband transformation of second-order systems

The second-order bandpass system

$$G_c(s) = \frac{2(b_1 s + b_0)}{s^2 + 2\zeta_0\omega_0 s + (1+\zeta_0^2)\omega_0^2} \tag{4.7}$$

has the partial fraction decomposition

$$G_c(s) = \frac{(1 - i\zeta_0)b_1 + (i/\omega_0)b_0}{s + \zeta_0\omega_0 + i\omega_0} + \frac{(1 + i\zeta_0)b_1 - (i/\omega_0)b_0}{s + \zeta_0\omega_0 - i\omega_0}.$$

Discarding the first term, which is small relative to the second one for frequencies close to ω_0, we get the following approximation to the equivalent baseband system of $G_c(s)$,

$$G(s) = G_c(s + i\omega_c) \approx \frac{(1 + i\zeta_0)b_1 - (i/\omega_0)b_0}{s + \zeta_0\omega_0 + i(\omega_c - \omega_0)}.$$

4.2.3 Baseband transformation of a time delay

Let

$$G_c(s) = \mathrm{e}^{-s\tau}.$$

The equivalent baseband system is given by

$$G(s) = G_c(s + i\omega_c) = \mathrm{e}^{-i\omega_c\tau}\mathrm{e}^{-s\tau}. \tag{4.8}$$

We see that the baseband model contains the same time delay as $G_c(s)$, but also an additional complex factor $\mathrm{e}^{-i\omega_c\tau}$. Since ω_c is typically quite large, a small change $\Delta\tau$ in τ gives a significant phase change $\omega_c\Delta\tau$ of $G(s)$, in addition to a small change $\Delta\tau$ of the delay.

4.2.4 Frequency response of equivalent baseband systems

The frequency response of an equivalent baseband system is typically not conjugate symmetric (i.e., $G(i\omega) \neq [G(-i\omega)]^*$) as seen in Figure 4.1. This means that both positive and negative frequencies need to be considered in control analyses. Consequently, the Bode diagram should show the response at negative frequencies and the Nyquist curve cannot be assumed to be symmetric with respect to the real axis; see Chapter 11 for examples of asymmetric Bode and Nyquist diagrams.

4.3 Control theory for complex-coefficient systems

In the previous section we saw that the baseband transformation of a bandpass system typically has complex coefficients. For this reason we will discuss

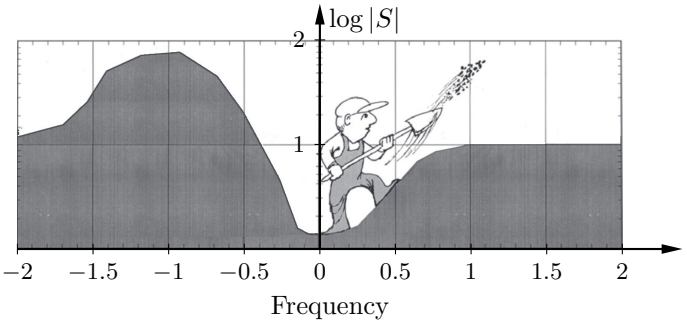


Figure 4.3 Illustration of that Bode's integral formula (2.7) must be taken over both positive and negative frequencies to hold for complex-coefficient LTI systems, since these may have $\int_0^\infty \log|S(i\omega)|\,d\omega < 0$. The figure is a modified version of the classic illustration in [Stein, 2003].

how standard tools of automatic control generalize to the complex setting. Overall, there are no significant difficulties but it should be remembered to consider negative frequencies and ensure that *conjugate* transposition is used instead of transposition. Previous work on control of complex-coefficient systems was mentioned in Section 1.3. To simplify the exposition we will only consider complex-coefficient single-input single-output (SISO) systems in this and the following sections.

4.3.1 Frequency-domain analysis

For frequency-domain analysis it is necessary to consider both positive and negative frequencies, as was mentioned in Section 4.2.4. For example, a factor $e^{i\epsilon}$ may give the impression of an improved phase margin if only positive frequencies are considered.

The Nyquist stability criterion and its proof require no change since the argument principle is valid for any meromorphic function, this was mentioned in [Gataric and Garrigan, 1999].

Bode's sensitivity integral (Section 2.2) is often only taken over positive frequencies as in [Freudenberg and Looze, 1985; Stein, 2003]. For complex-coefficient systems this formulation is not valid, and the two-sided version in (2.7) has to be used (the proof is the same in [Freudenberg and Looze, 1985]). That both positive and negative frequencies need to be considered is illustrated in Figure 4.3.

Bode's gain-phase relationship relates the amplitude and phase of a minimum-phase transfer function in an intuitive way [Skogestad and Postlethwaite, 2007]. The derivation relies on that the transfer function

is conjugate symmetric and there is no intuitive relationship for complex-coefficient transfer functions.

4.3.2 State-space analysis

Essentially all standard tools and concepts of state-space-based analysis, such as controllability, observability, stability, Gramians, computations of 2 norms and $\mathcal{H}_\infty$ norms, LQG design, and $\mathcal{H}_\infty$ design, directly generalize to the complex setting, as long as *conjugate* transposition is used. One exception to the use of complex conjugation is the standard algorithm for computing zeros of (possibly complex-coefficient) multivariable systems [Emami-Naeini and Van Dooren, 1982], where it in one step of the algorithm is imperative to use transposition *without* conjugation to avoid conjugating the zeros.

4.3.3 Computer tools for complex-coefficient control analysis

The popular computing environment MATLAB has reasonable support for complex-coefficient system, however, some caution is warranted. For example, the function `minreal` that computes minimal realizations does not support complex-coefficient systems and the command `nyquist` produces an incorrect Nyquist plot unless special measures are taken[2]. SIMULINK has limited support for complex signals in continuous-time models, so the real and imaginary parts often have to be treated individually.

4.3.4 Decomposition of a complex-coefficient transfer function

Given a complex-coefficient system $G(s)$ we can split its impulse response into real and imaginary parts

$$g(t) = g_{\mathrm{Re}}(t) + ig_{\mathrm{Im}}(t),$$

where $g_{\mathrm{Re}}(t)$ and $g_{\mathrm{Im}}(t)$ are real-valued. Denoting the Laplace transforms of $g_{\mathrm{Re}}(t)$ and $g_{\mathrm{Im}}(t)$ by $G_{\mathrm{Re}}(s)$ and $G_{\mathrm{Im}}(s)$ respectively we can write

$$G(s) = G_{\mathrm{Re}}(s) + iG_{\mathrm{Im}}(s). \tag{4.9}$$

Note that both $G_{\mathrm{Re}}(s)$ and $G_{\mathrm{Im}}(s)$ are real-coefficient (complex-valued) functions of the complex variable s.

It will be convenient to define the complex-conjugate G^* of a transfer function G as

$$G^*(s) \coloneqq G_{\mathrm{Re}}(s) - iG_{\mathrm{Im}}(s). \tag{4.10}$$

[2]Confirmed in version R2019a. Both issues were reported to MathWorks in Oct. 2017. The first issue was apparently hard to fix and the second issue was supposed to be resolved in 2018 (but was still present in version R2019a). However, we are happy to note that a third reported issue with `hinfsyn` was promptly resolved. A workaround for `nyquist` is to set `ShowFullContour` to `off` and supply a frequency vector.

This implies that $G^*(s)$ is the Laplace transform of $g^*(t)$ and that $G^*(s) = [G(s^*)]^*$. However, the most important observation is that

$$\boxed{G^*(i\omega) = [G(-i\omega)]^*.} \tag{4.11}$$

This relation says that $G^*(i\omega)$ gives the frequency-response of G at negative frequencies (up to conjugation).

With the definition (4.10), the decomposition (4.9) can be recovered from $G(s)$ as

$$G_{\mathrm{Re}}(s) = \frac{G(s) + G^*(s)}{2} \qquad \text{and} \qquad G_{\mathrm{Im}}(s) = \frac{G(s) - G^*(s)}{2i}. \tag{4.12}$$

4.3.5 Rotational invariance

A complex-coefficient LTI system that maps an input signal $\boldsymbol{u}$ to an output signal $\boldsymbol{y}$ will due to linearity also map the input signal $\mathrm{e}^{i\theta}\boldsymbol{u}$ to the output signal $\mathrm{e}^{i\theta}\boldsymbol{y}$. We will say that the system is *rotationally invariant* since multiplication by $\mathrm{e}^{i\theta}$ corresponds to a rotation in the complex plane.

We will more generally call a mapping $f : V_1 \to V_2$ between two complex vector spaces rotationally invariant if $f(\mathrm{e}^{i\theta}\boldsymbol{u}) = \mathrm{e}^{i\theta} f(\boldsymbol{u})$ for all $\boldsymbol{u}$ in V_1. Note that time invariance of a bandpass system corresponds to that the baseband system is rotationally invariant.

A real-valued function f on a complex vector space V (for example, the norm of a complex signal or the norm of a system) will be called rotationally invariant if $f(\mathrm{e}^{i\theta}\boldsymbol{u}) = f(\boldsymbol{u})$ for all $\boldsymbol{u}$ in V. The standard 2 norms and sup norms on complex signals and systems are rotationally invariant (even if the system itself is not).

4.4 Relation to equivalent real-coefficient representation

Equation (4.9) in the previous section showed that a complex-coefficient SISO transfer function $G(s)$ can be decomposed into

$$G(s) = G_{\mathrm{Re}}(s) + iG_{\mathrm{Im}}(s), \tag{4.13}$$

were $G_{\mathrm{Re}}(s)$ and G_{Im} have real coefficients. The action of $G(s)$ on a signal $\boldsymbol{u}(t) = u_{\mathrm{Re}}(t) + iu_{\mathrm{Im}}(t)$ can then be expressed as the action of the two-input two-output (TITO) system

$$\mathcal{G}_{\mathrm{equiv}}(s) = \begin{bmatrix} G_{\mathrm{Re}}(s) & -G_{\mathrm{Im}}(s) \\ G_{\mathrm{Im}}(s) & G_{\mathrm{Re}}(s) \end{bmatrix} \tag{4.14}$$

on real-valued signals $\underline{u} = \begin{bmatrix} u_{\mathrm{Re}} & u_{\mathrm{Im}} \end{bmatrix}^{\mathsf{T}}$. Conversely, it is seen that a real-coefficient system of the form (4.14) can be written as an equivalent complex-coefficient system (4.13).

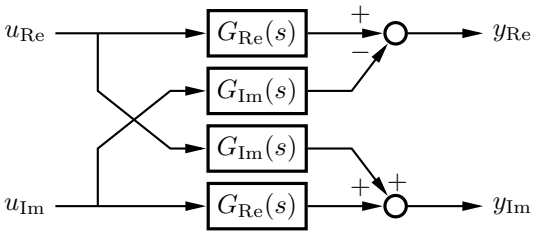


Figure 4.4 Illustration of how a complex transfer function $G(s) = G_{\mathrm{Re}}(s) + iG_{\mathrm{Im}}(s)$ acts on a signal $\boldsymbol{u} = u_{\mathrm{Re}} + iu_{\mathrm{Im}}$ to produce a signal $y(t) = y_{\mathrm{Re}} + iy_{\mathrm{Im}}$.

Much of the previous work on cavity field control dynamics has been based on the real TITO representation (4.13) rather than the complex SISO representation (4.12). Also, practical implementation of a complex-coefficient transfer function (4.13) must be done as in Figure 4.4, which corresponds to the representation (4.14). Conversely, certain real-world processes (see Section 4.6) which have dynamics of the form (4.14) are more intuitively analyzed using the representation (4.13). So it is meaningful to study the relationship between these two representations in some greater detail.

4.4.1 Diagonalization of equivalent real-valued TITO dynamics

The key to understand the relationship between the representations (4.13) and (4.14) is to note that (4.14) has the eigendecomposition

$$\mathcal{G}_{\mathrm{equiv}}(s) = T \begin{bmatrix} G(s) & 0 \\ 0 & G^*(s) \end{bmatrix} T^{\mathsf{H}}, \quad T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}. \tag{4.15}$$

Recalling (4.11), we see from (4.15) that the positive- and negative-frequency dynamics of $G(s)$ are intertwined in $\mathcal{G}_{\mathrm{equiv}}(s)$, which indicates that analysis using this representation is unnecessarily complicated.

Remark 4.2 The eigendecomposition (4.15) can be understood from noting that

$$\sqrt{2}T^{\mathsf{H}} = \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \tag{4.16}$$

maps $\begin{bmatrix} u_{\mathrm{Re}} & u_{\mathrm{Im}} \end{bmatrix}^{\mathsf{T}}$ to $\begin{bmatrix} \boldsymbol{u} & \boldsymbol{u}^* \end{bmatrix}^{\mathsf{T}}$, and that

$$\frac{1}{\sqrt{2}}T = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \tag{4.17}$$

maps $\begin{bmatrix} \boldsymbol{y} & \boldsymbol{y}^* \end{bmatrix}^{\mathsf{T}}$ to $\begin{bmatrix} y_{\mathrm{Re}} & y_{\mathrm{Im}} \end{bmatrix}^{\mathsf{T}}$. Consequently, $T^{\mathsf{H}}\mathcal{G}(s)T$ can be seen as mapping from $\begin{bmatrix} \boldsymbol{u} & \boldsymbol{u}^* \end{bmatrix}^{\mathsf{T}}$ to $\begin{bmatrix} \boldsymbol{y} & \boldsymbol{y}^* \end{bmatrix}^{\mathsf{T}}$.

Remark 4.3 The eigenvalue loci, i.e., the eigenvalues $\lambda_i(L(i\omega))$, where $L(s)$ is the open-loop transfer function of a MIMO system, is a generalization of the classic SISO Nyquist curve. From (4.15) it follows that the eigenvalue loci of a transfer function of the form (4.14) are given by the frequency response of (4.13) and its conjugate. In general, the eigenvalue loci are not too useful for robustness analysis [Skogestad and Postlethwaite, 2007]. In the special case of TITO systems of the form (4.14) they actually give information about the system's sensitivity function, but it is of course easier to work with the representation (4.13).

4.4.2 Norms

It is clear that the 2-norm of a complex signal, and its corresponding real representation are equal, $||\boldsymbol{u}||_2 = ||\begin{bmatrix} u_{\mathrm{Re}} & u_{\mathrm{Im}} \end{bmatrix}^{\mathsf{T}}||_2$.

The eigenvectors in the eigendecomposition (4.15) are orthogonal so the magnitudes of the eigenvalues are equal to the singular values of $\mathcal{G}(i\omega)$. From this observation we get the following theorem [Troeng et al., 2017].

THEOREM 1
If $G = G_{\mathrm{Re}} + iG_{\mathrm{Im}}$, and $\mathcal{G}_{\mathrm{equiv}}$ is given by (4.14), then

$$||\mathcal{G}_{\mathrm{equiv}}||_\infty = ||G||_\infty \tag{4.18}$$

$$||\mathcal{G}_{\mathrm{equiv}}||_2 = \sqrt{2}\,||G||_2\,. \tag{4.19}$$

Remark 4.4 This theorem has an important consequence. It shows that if a real-coefficient system $\mathcal{G}_{\mathrm{equiv}}$ of the form (4.14) is analyzed using the complex representation G in (4.13) and the sensitivity function is found to have a maximum magnitude M_S then the original system $\mathcal{G}_{\mathrm{equiv}}$ has the same level of robustness.

4.4.3 State-space realizations

If the complex-coefficient system (4.13) has a state-space realization

$$\left[\begin{array}{c|c} A_{\mathrm{Re}} + iA_{\mathrm{Im}} & B_{\mathrm{Re}} + iB_{\mathrm{Im}} \\ \hline C_{\mathrm{Re}} + iC_{\mathrm{Im}} & D_{\mathrm{Re}} + iD_{\mathrm{Im}} \end{array}\right], \tag{4.20}$$

then it is clear that the real-coefficient system (4.14) has a state-space representation given by

$$\left[\begin{array}{cc|cc} A_{\mathrm{Re}} & -A_{\mathrm{Im}} & B_{\mathrm{Re}} & -B_{\mathrm{Im}} \\ A_{\mathrm{Im}} & A_{\mathrm{Re}} & B_{\mathrm{Im}} & B_{\mathrm{Re}} \\ \hline C_{\mathrm{Re}} & -C_{\mathrm{Im}} & D_{\mathrm{Re}} & -D_{\mathrm{Im}} \\ C_{\mathrm{Im}} & C_{\mathrm{Re}} & D_{\mathrm{Im}} & D_{\mathrm{Re}} \end{array}\right], \tag{4.21}$$

and vice versa.

4.4.4 Advantages of complex-coefficient representation

The real-coefficient representation (4.14) is in many respects similar to the complex-coefficient representations (4.13), however, the complex-coefficient representation has some advantages.

Simplicity. The insight and simplicity of working with SISO systems is maintained. This enables the use of Nyquist diagrams and double-sided Bode diagrams to gain an intuitive understanding of control performance. With the real-coefficient representation one has to use the Bode diagrams of the 2×2 transfer-function matrix (4.14), which gives less insight.

Less computations. The amount of computations for analysis and control design is roughly halved if the complex representation (4.13) is used instead of the real representation (4.14) (multiplying two complex numbers requires 4 real multiplications but multiplying two 2×2 real matrices requires 8 real multiplications).

Implicit structure. There is no need to explicitly impose the rotationally-invariant structure of (4.14) when doing systems identification and control design. The structure is implicit in the complex-coefficient representation.

4.4.5 On the structure of the optimal controller

The following two theorems are not surprising, but it is nice to know that the optimal control does not have a more complicated structure than the plant itself. The results are proved in Appendix G, using the same technique as in [Bamieh et al., 2002]. Recall from Section 4.3.5 what is meant by rotational invariance.

THEOREM 2
The optimal control performance for a complex-coefficient plant with respect to a convex, rotationally-invariant control objective, subject to convex, rotationally-invariant constraints, can be attained by a complex-coefficient controller.

THEOREM 3
The optimal control performance for a real-coefficient plant with respect to a rotationally-invariant control objective, subject to rotationally-invariant constraints, can be attained by a real-coefficient controller.

4.5 Signals with directionality

To describe the second-order properties of stochastic complex signals it is necessary to use both the covariance function

$$r_{uu}(t_1, t_2) = \mathbb{E}[\boldsymbol{u}(t_1)\boldsymbol{u}(t_2)^{\mathsf{H}}]$$

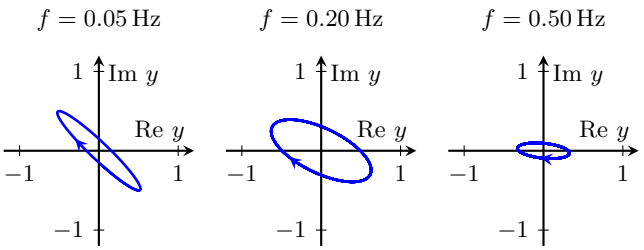


Figure 4.5 Lissajous ovals in the complex plane generated by excitation of $1/(s+1+i)$ by $u = \sin(2\pi f t)$.

and the *complementary covariance function* [Picinbono and Bondon, 1997; Schreier and Scharf, 2010]

$$\breve{r}_{uu}(t_1, t_2) = \mathbb{E}[\boldsymbol{u}(t_1)\boldsymbol{u}(t_2)^{\mathsf{T}}].$$

The following definitions are from [Schreier and Scharf, 2010].

Definition 4.2 A complex stochastic process is called wide-sense stationary if its covariance and complementary covariance only depend on $t_2 - t_1$, i.e., if $r_{uu}(t_1, t_2) = r_{uu}(t_2 - t_1)$, and $\breve{r}_{uu}(t_1, t_2) = \breve{r}_{uu}(t_2 - t_1)$.

Definition 4.3 A complex wide-sense stationary signal $\boldsymbol{u}$ is called *proper* if $\breve{r}(t) = 0$ for all t.

For control design, we will assume all signals to be proper. However the actual disturbances on a field control loop tend to have certain directionality, this will be briefly discussed in Chapter 15.2.

Response to a signal with a specific direction

The time-response of a stable complex-coefficient system $G(s)$ to a signal with a specific direction, say $\boldsymbol{u}(t) = \cos(\omega_0 t)\boldsymbol{u_0}$, is (after transients) given by

$$\begin{aligned}\boldsymbol{y}(t) &= \left[G(i\omega_0)\frac{1}{2}\mathrm{e}^{i\omega_0 t} + G(-i\omega_0)\frac{1}{2}\mathrm{e}^{-i\omega_0 t}\right]\boldsymbol{u_0} \\ &= \big[\,|G_{\mathrm{Re}}(i\omega_0)|\cos\left(\omega_0 t + \angle G_{\mathrm{Re}}(i\omega_0)\right) \\ &\qquad + i\,|G_{\mathrm{Im}}(i\omega_0)|\cos\left(\omega_0 t + \angle G_{\mathrm{Im}}(i\omega_0)\right)\big]\boldsymbol{u_0}. \quad (4.22)\end{aligned}$$

The signal (4.22) traces out a so-called Lissajous oval in the complex plane, see Figure 4.5. In the special case of a transfer function with real coefficients ($G_{\mathrm{Im}}(s) = 0$), the output signal has the same direction as the input signal.

4.6 Other applications of complex-coefficient systems

There are a few other real-world control applications, apart from cavity field control, where it is helpful to model the system dynamics as a complex-coefficient LTI system. The complex-coefficient dynamics in these applications either arise from transforming a bandpass system to baseband (Sections 4.1 and 4.2) or from representing a system with physically rotationally-invariant dynamics as a complex-coefficient system.

Control applications where complex-coefficient dynamics arise from consideration of baseband dynamics are MEMS gyroscopes [Saggin et al., 2019] and Cartesian feedback linearization of power amplifiers (discussed below).

An example of a physical system with rotationally-invariant dynamics is the motion of the Foucault pendulum[3] in the xy-plane. The dynamics of the Foucault pendulum can, subject to small-angle approximation, be described by the following complex-coefficient differential equation [Arnold, 1978]

$$\ddot{z} + i2\Omega \sin\lambda_0 \dot{z} + \omega_0^2 z = 0,$$

where $z = x + iy$, ω_0 is the natural frequency of the pendulum, Ω is the rotational frequency of the Earth, and λ_0 is the latitude where the pendulum is located. See [Pippard, 2007] for similar examples.

Control applications where complex-coefficient dynamics arise from rotational invariance are three-phase electric machines (after dq transformation) [Harnefors, 2007; Dòria-Cerezo and Bodson, 2016] and active vibration damping of rotary machinery [Byun and Lee, 1988]. Another rotationally-invariant system that can be analyzed as a complex-coefficient system is the "spinning body" from the robust control literature (discussed below).

4.6.1 "The spinning body"

"The spinning body" is a classic example from the robust control literature [Zhou et al., 1996; Skogestad and Postlethwaite, 2007] that is used for demonstrating that loop-by-loop analysis with, e.g., the Nyquist criterion, is insufficient to guarantee closed-loop robustness of a MIMO system.

The "spinning body" is assumed to spin around the z axis with constant velocity; the controlled states are the rotational velocities around the x and y axes; the control signals are torque inputs around the same axes; and the states are measured in scaled and rotated coordinates. The system dynamics are given by

$$\mathcal{P}(s) = \frac{1}{s^2 + a^2} \begin{bmatrix} s - a^2 & a(s+1) \\ -a(s+1) & s - a^2 \end{bmatrix} \tag{4.23}$$

and control by unity feedback, $\mathcal{K}(s) = I$, is assumed.

[3] An experiment conceived by Léon Foucault in 1851 to demonstrate the Earth's rotation.

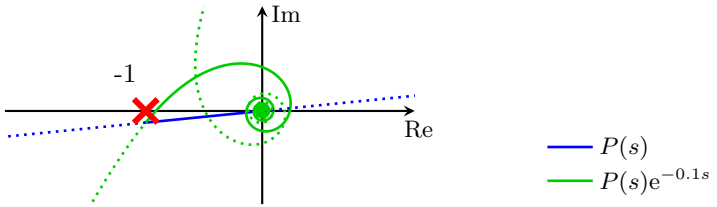


Figure 4.6 Nyquist curve for the classic spinning-body example when considered as a single-input single-output, complex-coefficient system

The system (4.23) has the structure in (4.14) and writing it in the equivalent complex form (4.13) we get

$$P(s) = \frac{1 - ia}{s + ia}.$$

The controller corresponds to $K(s) = 1$. The robustness of this system is straightforward to evaluate using its Nyquist diagram which is shown in Figure 4.6. It clear that the robustness is miserable. Due to Remark 4.4 this conclusion also holds for the original system (4.23).

4.6.2 Cartesian feedback linearization of RF amplifiers

Cartesian feedback linearization of power amplifiers has been studied as a means of reducing power consumption and adjacent channel interference in telecommunications applications [Johansson, 1991; Briffa and Faulkner, 1996; Dawson, 2003].

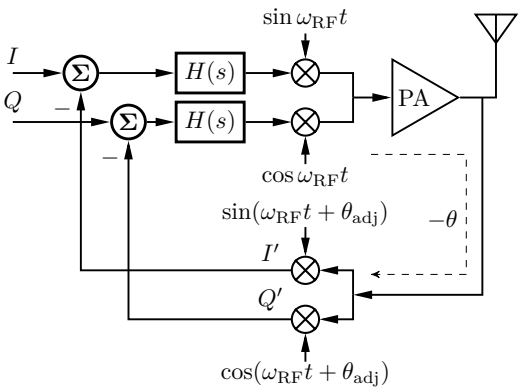


Figure 4.7 Schematic of Cartesian feedback linearization of an RF amplifier [Dawson, 2003]. It is crucial to select the phase adjustment angle θ_{adj} to cancel the phase shift θ.

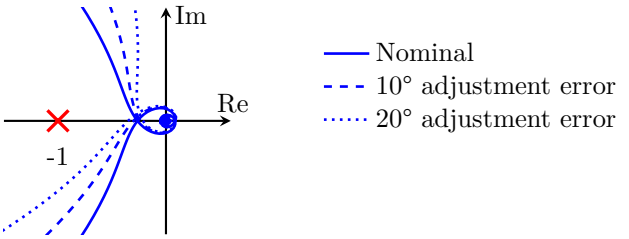


Figure 4.8 Nyquist curves of a Cartesian feedback loop with different phase adjustment errors δ. The data for the nominal curve is from [Johansson, 1991, Sec. 4.2].

If the amplifier is operating in an almost linear region then the open-loop system is well approximated by

$$G(s) = H(s)P(s)\mathrm{e}^{-s\tau}\mathrm{e}^{i\delta}, \tag{4.24}$$

where $H(s)$ is the loop filter, $P(s)$ is a baseband model of the mixer and amplifier dynamics, τ is the loop delay, and $\delta \coloneqq \theta_{\text{adj}} - \theta$ is the phase adjustment error. To avoid instability and performance degradation the phase adjustment error δ must be small.

The system (4.24) was simulated as a complex-coefficient system in [Briffa, 1996] but the equivalent TITO representation (4.14) was used for algebraic stability analysis; after several calculations it was shown that a phase adjustment error δ reduces the phase margin by the same amount.

With the complex SISO representation (4.24) this conclusion follows trivially [Troeng et al., 2017] from the Nyquist stability criterion and from noting that the factor $\mathrm{e}^{i\delta}$ corresponds to a rotation of the nominal Nyquist curve $H(i\omega)P(i\omega)\mathrm{e}^{-i\omega\tau}$ by δ radians, see Figure 4.8 for an illustration.

Part I

Modeling and Essentials of Cavity Field Control

5

Modeling: The RF System

This chapter provides baseband models of the components of the *RF system*. The main responsibility of the RF system is to establish, maintain, and control the electromagnetic field in the accelerating cavity(ies). Cavity modeling is discussed in the next chapter. An RF station for the medium-β section of the ESS linac is shown in Figure 5.1.

Before going into components of the RF system it is a good time to explain the operation of a typical field control loop (see Figure 5.2). The RF amplifier generates an electromagnetic wave that propagates through the waveguide and maintains a standing electromagnetic field in the cavity. The cavity field accelerates the particle bunches and, as discussed in Section 3.3, it must have the correct amplitude and phase (relative to the bunches). For this reason the cavity field is sensed by a pickup probe (i.e., antenna) and the complex envelope of the signal is detected by the *receiver* in the *low-level RF* (LLRF) system. The phase of the envelope is adjusted with respect to the phase reference. The field controller compares the adjusted complex envelope to the given setpoint and adjusts the drive signal to the RF amplifier accordingly.

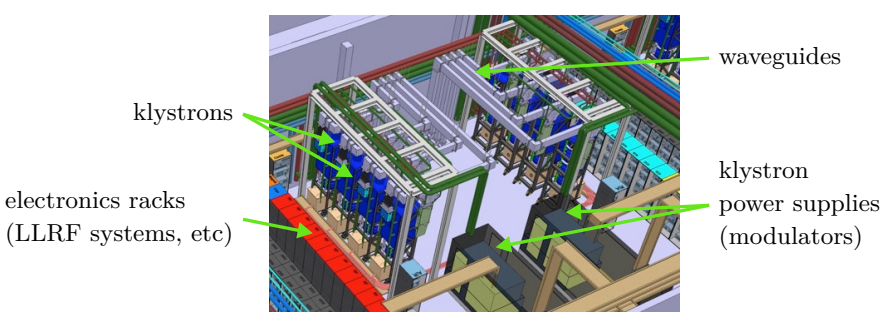


Figure 5.1 RF station for the medium-β section of the ESS linac. It drives eight cavities with one klystron per cavity. Image credit: ESS.

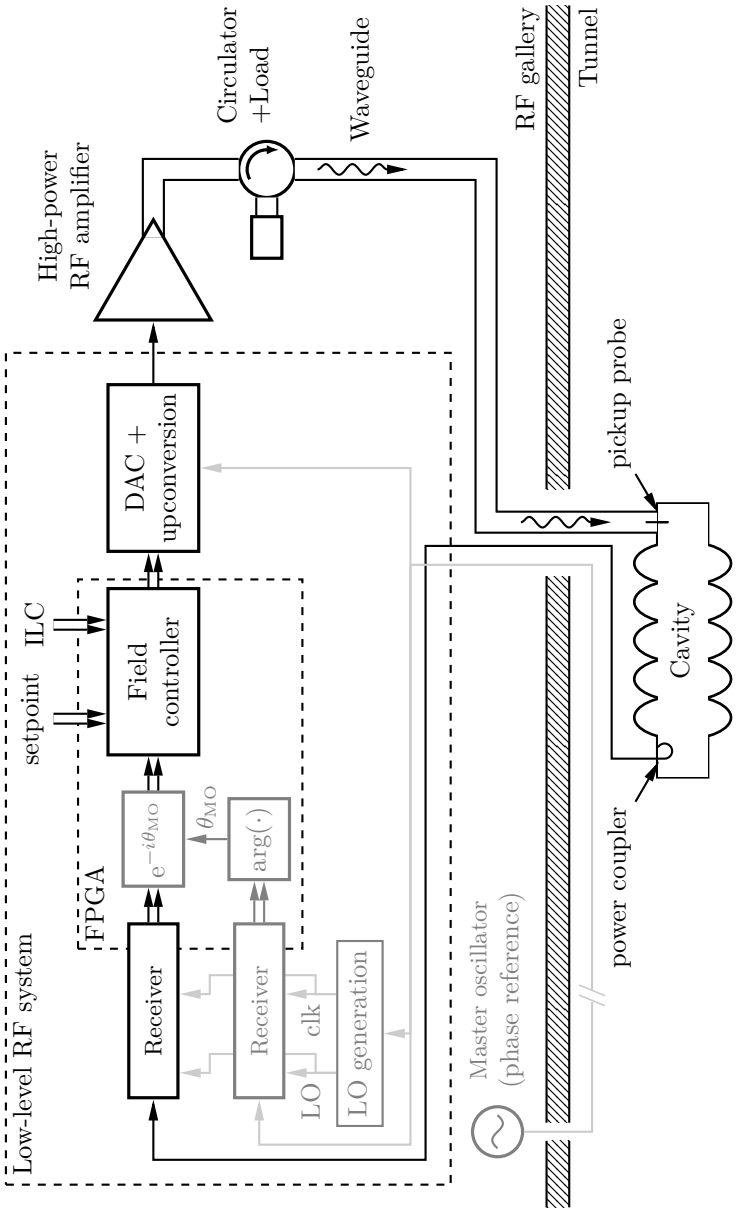


Figure 5.2 Schematic of a typical field-control loop. Double arrows (⟹) indicate complex signals (implemented by two real signal in the FPGA). The grayed-out parts provide a stable phase reference for the field detection and the upconversion. To avoid radiation damage to the RF-system electronics, the accelerating cavities are often located in a separate tunnel.

5.1 The low-level RF system

The low-level RF (LLRF) system is the platform where the field controller (the topic of this thesis) is implemented. The LLRF system could also contain feedback controllers for the cavity's resonance frequency, or an outer loop for beam-energy feedback [Pfeiffer, 2014].

A schematic of a typical, digital LLRF system is shown in Figure 5.3. Additional measured signals of, e.g., the amplifier output, the voltage of the amplifier's power supply, or the beam current, could also be used by the field controller to improve the control performance. Nowadays, almost all field controllers are implemented digitally on FPGAs, but in the old days, analog electronics were used.

Examples of LLRF hardware are shown in Figure 5.4. For an introduction to the electronic components used in LLRF systems, see [Gallo, 2010; Ellingson, 2016]. For details on the LLRF system for ESS, see Appendix B.

5.1.1 Receiver

Hardware for recovering the complex envelope of an RF signal is called a *receiver*. A common receiver architecture for digital LLRF systems is the heterodyne digital-quadrature architecture (Figure 5.5). First, the RF signal is downconverted to an intermediate-frequency (IF) signal to reduce the sensitivity to ADC sampling jitter, then it is sampled by an ADC, after which the sampled signal is digitally downconverted in the FPGA. Alternative receiver architectures are discussed in [Ellingson, 2016, Sec. 15.6]. For example, previous-generation analog LLRF systems typically used amplitude and phase detectors [Pedersen, 1975].

For field control, the filter $H_{\mathrm{DDC}}(z)$ in Figure 5.5 is often chosen as a moving average filter $H_{\mathrm{DDC}}(z) = (1 + z^{-1} + \ldots + z^{-(N-1)})/N$ [Schilcher, 2007; Hoffmann, 2008], see Chapter 12 for a further discussion.

To include the receiver dynamics in a baseband model of the RF system, we need a model of the dynamics from the complex envelope $\boldsymbol{y}$ of y_{RF}, to its estimate $\hat{\boldsymbol{y}}$. Such a model is shown in Figure 5.6, where we have neglected the fast dynamics of the mixer and the bandpass filter.

5.1.2 Controller implementation on FPGAs

Almost all modern field controllers are implemented on field-programmable gate arrays (FPGAs), which are reconfigurable, integrated circuits. FPGAs support massive parallelization of the controller computations and enable controller rates of 10–100 MHz. Compared to previous-generation field controllers based on analog electronics, the use of FPGAs allows easier reconfiguration and more elaborate controller architectures. Programming FPGAs requires specialized knowledge of hardware description languages.

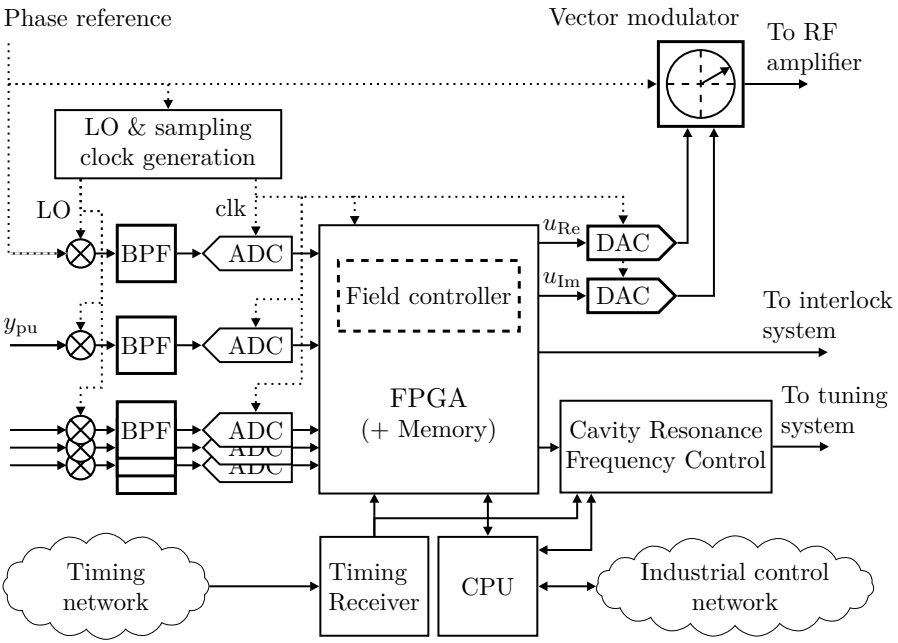


Figure 5.3 Schematic of a typical LLRF system. Fast feedback loops, such as for field control are implemented in the FPGA. A number of additional signals from the RF system are also sampled for interlocks and monitoring. The CPU handles local, general-purpose computations, and connects the LLRF system with the accelerator's industrial control system [EPICS; TANGO]. The timing receiver receives pulses from the timing network so that the RF pulse can be synchronized to the beam pulse.

Figure 5.4 Components of an LLRF system based on the MicroTCA standard. *Left*: Digitizer SIS8300-KU, containing ADCs and an FPGA (Struck). *Middle*: Signal conditioning board DWC8VM1, containing analog down- and upconversion (Struck). *Right*: 12-slot MicroTCA4 crate (Schroff). Photo Credit: Struck Innovative Systeme GmbH, and nVent Schroff.

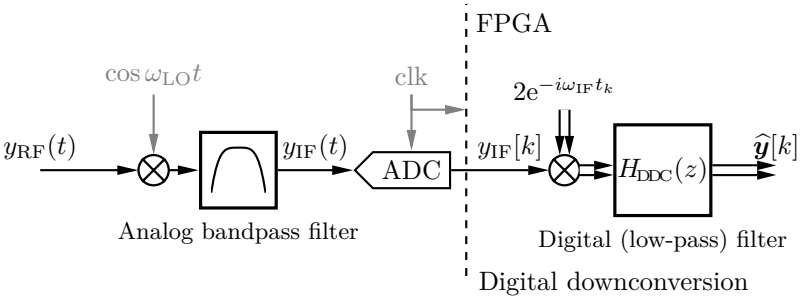


Figure 5.5 Typical LLRF receiver architecture for recovering the complex envelope $\boldsymbol{y}$ of an RF signal $y_{\mathrm{RF}}(t) = \mathrm{Re}\left\{\boldsymbol{y}(t)\mathrm{e}^{i\omega_{\mathrm{RF}} t}\right\}$.

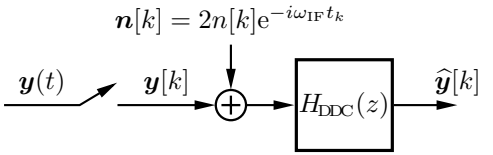


Figure 5.6 Baseband model of the receiver in Figure 5.5. Note that all signals are complex, so single lines can be used without confusion.

Every clock cycle of processing from the measurement signal to the control signal adds to the delay of the field control loop so careful implementation of the field controller is crucial. Current FPGAs used for field control operate with clock frequencies on the order of 100 MHz [Doolittle et al., 2016].

Other automatic control applications that rely on FPGAs are those with fast dynamics, such as power electronics [Monmasson et al., 2011] and lasers [Leibrandt and Heidecker, 2015]; or where the control strategies are computationally demanding, for example, model predictive control [Ling et al., 2006].

5.1.3 Digital-to-analog conversion and upconversion

There are different architectures for upconverting a digital baseband signal [Ellingson, 2016, Figure 17.1d–f]. Both direct quadrature upconversion [Schilcher, 2007, Section 4] (Figure 5.7) and heterodyne upconversion of a digitally synthesized intermediate-frequency signal are used.

DAC quantization errors. DACs give rise to quantization errors which act as a load disturbance in a similar way as the amplifier ripple and beam current ripple. However, the errors have relatively small magnitude and high frequency so it has a small impact on the control performance.

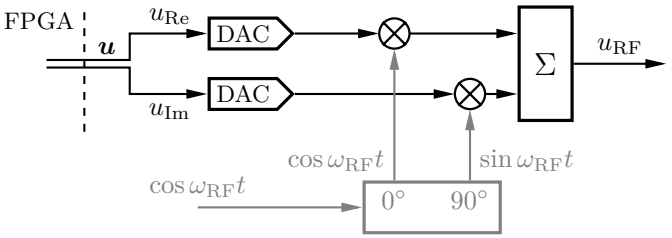


Figure 5.7 Schematic for direct quadrature upconversion. Often, the hybrid, mixers and the combiner are collectively referred to as a *vector modulator*, or a *quadrature modulator*.

5.1.4 Upconverter IQ imbalance

The baseband signal $\boldsymbol{u}$ should ideally be upconverted to the bandpass signal

$$u_{\mathrm{RF}}(t) = \mathrm{Re}\left\{\boldsymbol{u}(t)\mathrm{e}^{i\omega_{\mathrm{RF}}t}\right\} = u_{\mathrm{Re}}(t)\cos(\omega_{\mathrm{RF}}t) - u_{\mathrm{Im}}(t)\sin(\omega_{\mathrm{RF}}t). \tag{5.1}$$

However, in the case of direct quadrature upconversion, this is typically not the case. In the next subsection we discuss how this can be modeled; not because it tends to be a severe problem, but because it is an example of how *widely linear* dynamics arise in the field control loop (see Appendix A).

Analog quadrature upconversion suffers from different hardware imperfections [Schilcher, 2007]:

1. The phase between the RF signals to the two mixers may differ from $\pi/2$ by an angle θ_{skew} known as the *quadrature skew* or the *phase imbalance*.
2. The gains of the Re (I) and Im (Q) branches may differ by a factor ρ_{gain} known as the *gain imbalance*.
3. The DACs may have DC offsets.

The first two effects are collectively referred to as IQ imbalance and they amount to that the output of the upconverter is not given by (5.1) but instead

$$\breve{u}_{\mathrm{RF}}(t) = u_{\mathrm{Re}}\cos(\omega_{\mathrm{RF}}t) - u_{\mathrm{Im}}\rho_{\mathrm{gain}}\sin(\omega_{\mathrm{RF}}t + \theta_{\mathrm{skew}}).$$

Denoting the complex envelope of $\breve{u}_{\mathrm{RF}}$ by $\breve{\boldsymbol{u}} = \breve{u}_{\mathrm{Re}} + i\breve{u}_{\mathrm{Im}}$, we have

$$\begin{bmatrix}\breve{u}_{\mathrm{Re}}\\ \breve{u}_{\mathrm{Im}}\end{bmatrix} = \begin{bmatrix}1 & -\rho_{\mathrm{gain}}\sin\theta_{\mathrm{skew}}\\ 0 & \rho_{\mathrm{gain}}\cos\theta_{\mathrm{skew}}\end{bmatrix}\begin{bmatrix}u_{\mathrm{Re}}\\ u_{\mathrm{Im}}\end{bmatrix}. \tag{5.2}$$

This relation can also be written [Zou et al., 2008]

$$\breve{\boldsymbol{u}} = K_1\boldsymbol{u} + K_2\boldsymbol{u}^* \tag{5.3}$$

where $K_1 = (1 + \rho_{\mathrm{gain}}\mathrm{e}^{i\theta_{\mathrm{skew}}})/2$, and $K_2 = (1 - \rho_{\mathrm{gain}}\mathrm{e}^{i\theta_{\mathrm{skew}}})/2$.

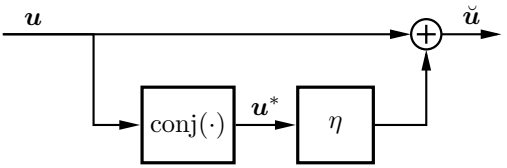


Figure 5.8 Baseband model of how IQ imbalance in the upconverter affects the complex envelope of the RF output. Note that the conjugation operation breaks the rotational invariance in the relation between $\boldsymbol{u}$ and $\breve{\boldsymbol{u}}$.

By redefining $\breve{\boldsymbol{u}}$ to include the factor $1/K_1$, we can write

$$\breve{\boldsymbol{u}} = \boldsymbol{u} + \eta \boldsymbol{u}^*, \tag{5.5}$$

where

$$\eta := K_2/K_1 \approx K_2 \approx (1 - \rho_{\text{gain}})/2 - i\theta_{\text{skew}}/2. \tag{5.6}$$

The approximations hold when ρ_{gain} and θ_{skew} are small.

Equation (5.5) which describes the impact of IQ imbalance can be visualized as in Figure 5.8. Due to the conjugation operation, this effect cannot be modeled as a linear, complex-coefficient system (the system is still *widely linear*, see Appendix A.

To reduce IQ imbalance to acceptable levels one may either do analog calibration[1], or estimate the matrix in (5.2) and apply its inverse at the controller output. The DAC offsets can similarly be compensated digitally. One approach to estimate the IQ imbalance and the DC offset is to use Fitzgibbon's elegant ellipse-fitting method [Fitzgibbon et al., 1999] as discussed in [Rojas et al., 2011]. Typically, the imperfections can be well compensated. Furthermore, the remaining imperfections occur at the output of the controller, and is hence much less problematic than if they would have occurred at the input, see Chapter 8. For these reasons we will not consider IQ imbalance in what follows.

Remark 5.1 The problem with IQ imbalance is not present in the receiver since *digital* quadrature downconversion is used for exactly this reason.

5.2 Phase-reference system

The absolute phase between the particle bunches and the cavity fields cannot be measured or estimated with sufficient speed and accuracy for feedback

[1] This approach was used at SNS, but their newest LLRF systems use heterodyne upconversion from a digitally synthesized intermediate-frequency signal to avoid the problem altogether [Personal communication with Mark Crofford, SNS].

control. Instead, the phases of all cavities are controlled relative to an extremely stable phase reference (*master oscillator*) that is distributed along the linac. The bunch phase in a high-intensity ion linac is determined by the phase of the radio-frequency quadrupole that bunches the beam. Since the RFQ is controlled with respect to the phase reference, it is possible to use the phase references as a substitute to the true beam phase.

A common approach to phase-reference distribution along a linac is to use a rigid coaxial line [Piller et al., 2005]. Temperature stabilization is important to avoid phase drifts [Olofsson et al., 2018]. Reference phase distribution by a laser through optical fibers is used when high precision is necessary in long linacs [Sikora et al., 2018].

One approach to measure the cavity signal relative to the phase reference is to sample both the cavity signal and the phase reference and subtract the reference phase from the cavity signal. This approach is sometimes called *reference tracking* [Ludwig, 2011]. If both the LO frequency, sampling clock and the upconversion are locked to the master oscillator it suffices to only sample the cavity signal, however, this approach is prone to drifts. Drift calibration for field control was discussed in [Ludwig, 2017].

Good phase stability of the master oscillator and the phase distribution system is crucial for keeping the phase between the beam and cavity field constant. Drifts in the phase distribution system leads to systematic phase errors unless calibration is done. The impact of the master oscillator's phase noise is discussed in [Ludwig et al., 2006].

The stability issues of the master oscillator and the phase-distribution system are orthogonal to the field control problem. In what follows we will assume that all signals are measured with respect to perfect phase reference.

5.3 RF cables

RF cables are used for connecting the cavity pickup probe to the LLRF system, and the LLRF system to the RF amplifier. Typical group velocities in RF cables range from $0.66c$ up towards c [Kaiser, 2005].

Example: The ESS linac will use a 3/8 inch, dielectric, coaxial cable (SCF38-50JFN from Radio Frequency Systems) between the pickup probes and the LLRF systems. The propagation velocity in this cable is $0.82c$, which gives a delay of 160 ns for a typical cable length of 40 m.

5.4 RF distribution system

The RF distribution system transfers the electromagnetic waves generated by the RF amplifier to the cavity. This is most commonly done with rectangular

waveguides. Alternatives are coaxial waveguides, and coaxial cables (for low-power test setups). In many circumstances there is a reverse wave traveling backward from the cavity through the waveguide [Wangler, 2008]. The reverse wave could damage the RF amplifier so *circulators* are used for directing the reverse wave to an RF load where it is absorbed [Wangler, 2008].

The dynamics of a waveguide is given by its propagation delay. The electromagnetic wave is also attenuated as it passes through the waveguide. From a field control perspective this corresponds to a constant gain which disappears after normalization.

Rectangular Waveguides

The group velocity at which an electromagnetic wave propagates through a rectangular waveguide is given by $v_{\mathrm{wg}} = \sqrt{1-(c/(2af))^2}\cdot c$, where f is the frequency of the wave, a is the longest side of the waveguide's cross section, and c is the speed of light. Staying within recommended waveguide dimensions gives group velocities in the range $0.60c$–$0.85c$ [EIA RS-216-B].

Example: In the first part of the ESS linac the RF frequency is $f_{\mathrm{RF}} = 352.21\,\mathrm{MHz}$ and the waveguide dimension is $a = 0.5842\,\mathrm{m}$ which gives $v_{\mathrm{wg}} = 0.68c$. In the second part of the linac the frequency is doubled and the waveguide dimensions are halved which also gives $v_{\mathrm{wg}} = 0.68c$. The propagation delay through a 40 m long waveguide is given by $40\,\mathrm{m}/0.68c = 200\,\mathrm{ns}$.

5.5 High-power RF amplifiers

Many types of RF amplifiers are used for particle accelerators, depending on the operation frequency and required output power. Klystrons, solid-state amplifiers, tetrodes, and inductive output tubes are the most common ones. For an introduction to RF amplifiers, see [Carter, 2010].

In reality, there are no ideal amplifiers. Below we discuss the nonidealities that are most relevant for field control performance, namely:

- the limited bandwidth,
- the nonlinear dependence on input signal's amplitude,
- the gain and phase dependence on the supply voltage (klystrons), and
- noise and spurious components (solid-state amplifiers).

5.5.1 Limited bandwidth

We will model the limited bandwidth of the RF amplifier by a first-order low-pass filter as in [Schilcher, 1998; Serrano et al., 2017],

$$P_{\mathrm{amp}}(s) = \frac{\omega_{\mathrm{amp}}}{s+\omega_{\mathrm{amp}}}, \tag{5.7}$$

where ω_{amp} is the 3dB bandwidth in rad/s.

The bandwidth of typical RF amplifiers is at least 1 MHz which is more than ten times the cross-over frequency of a typical field control loop. Hence the amplifier dynamics has a rather small impact on the field control loop, but it is prudent to account for it in field control analysis.

Bode magnitude curves for two klystrons that will be used at ESS are shown in Figure 5.9. The model (5.7) is obviously not a good model for the Thales klystron but well-designed field controllers should have sufficient roll off to avoid instability from the resonance peak[2], see Figure 8.9.

A more detailed model should account for that klystrons contain several high-bandwidth cavities that are coupled via an electron beam. Each cavity has baseband dynamics of the form $\gamma_k/(s+\gamma_k-i\Delta\omega_k)$ where γ_k is the cavity's decay rate and $\Delta\omega_k$ is the difference between the cavity resonance frequency and the nominal RF frequency.

5.5.2 Nonlinearity

The gain and the phase shift of an RF amplifier has a nonlinear dependence on the magnitude of the input signal [Schenk, 2008]. Assuming that the nonlinearity is static (memoryless) it can be represented as

$$\mathbf{f}_{\text{amp}}(\boldsymbol{u}) = \mathrm{f}_{\text{amp}}(|\boldsymbol{u}|)\mathrm{e}^{i(\angle \boldsymbol{u}+\theta_{\text{amp}}(|\boldsymbol{u}|))}, \tag{5.8}$$

where $\mathbf{f}_{\text{amp}}$ is the baseband output of the amplifier (in normalized units), $\boldsymbol{u}$ is the input signal, and θ_{amp} and f_{amp} are nonlinear functions describing the nonlinearity. Typical nonlinear amplifier characteristics are shown in Figure 5.10. Note that the output saturates at a level $\mathrm{f}_{\text{amp sat}}$.

The magnitude curve of Figure 5.10 shows that the nonlinearity becomes more severe closer to saturation. For this reason RF amplifiers for field control are often operated with some overhead (back-off) to saturation. Overheads in the range of 25–40 % are typical[3].

The linearization of equation (5.8) around the operating point corresponding to a real input signal $\boldsymbol{u}_0 = u_0 > 0$ is (up to the phase shift from $\theta_{\text{amp}}(u_0)$) given by

$$\begin{bmatrix}\tilde{\mathrm{f}}_{\text{Re}}\\ \tilde{\mathrm{f}}_{\text{Im}}\end{bmatrix} = \begin{bmatrix}\mathrm{f}'_{\text{amp}}(u_0) & 0\\ \mathrm{f}_{\text{amp}}(u_0)\theta'_{\text{amp}}(u_0) & \mathrm{f}_{\text{amp}}(u_0)/u_0\end{bmatrix}\begin{bmatrix}\tilde{u}_{\text{Re}}\\ \tilde{u}_{\text{Im}}\end{bmatrix}. \tag{5.9}$$

The matrix in (5.9) is not rotationally invariant (i.e., of the form (4.14)) and shows that the linearization of a complex-valued nonlinearity may give rise to IQ imbalance (or widely linear dynamics, see Section A).

[2] At Linac4 at CERN, a notch filter was added to the (LQG) controller to avoid instability from the interaction between a similar resonance in their Thales klystrons and a parasitic mode in their pi-mode-structure cavities [Personal communication with Robert Bonner and Philippe Baudrenghien, CERN].

[3] The target value for ESS is 25 % [Personal communication with Morten Jensen, ESS]. In [Omet, 2014] it is stated that 40 % is common.

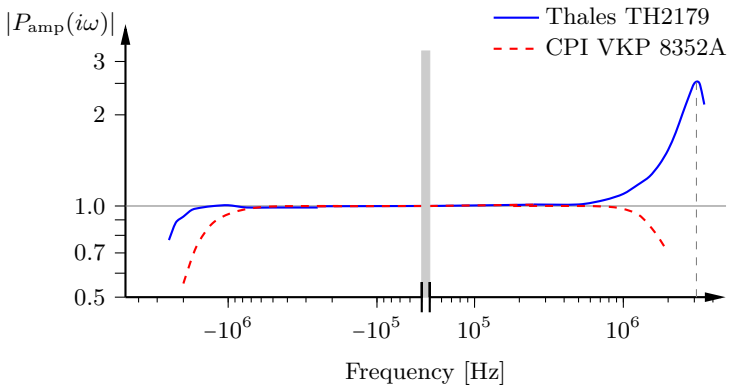


Figure 5.9 Bode magnitude curves (in baseband) for two 352 MHz klystrons that will be used at ESS. Phase curves are typically not provided by klystron manufacturers. The CPI klystron will power the RFQ and two DTLs and a Thales klystron with similar characteristics will power the remaining three DTLs. Note that the Thales klystron has a resonance peak at around 3 MHz. Peaks like these are typical for klystron from the manufacturers Thales and Canon. Credit: Data from Thales and CPI, provided by Chiara Marrelli, ESS.

One obvious approach to deal with the nonlinearity is to include an inverse model of it in the field controller (called pre-distortion in the telecommunications literature) [Omet, 2014; Ellingson, 2016]. Another approach is to close an inner loop around the klystron [Baudrenghien et al., 2014], but the loop delay of the inner loop is often too long for this to be efficient.

Remark 5.2 Higher-order harmonics from RF amplifiers (typically quantified by the third-order intercept point (IP_3)) is a major concern in telecommunications due to the strict requirements on adjacent channel interference. However, such higher-order harmonics are of little concern for field control due to the relatively narrow bandwidth of the cavities.

5.5.3 Gain and phase-shift variations

Small variations $\tilde{g}_{\text{amp}}$ and $\tilde{\theta}_{\text{amp}}$ of the RF amplifier's gain and phase-shift can be seen as a multiplicative disturbance $(1 + \tilde{g}_{\text{amp}})\text{e}^{i\tilde{\theta}_{\text{amp}}}$ acting at the amplifier output.

Pushing. Variations in an RF amplifier's supply voltage changes its gain and phase shift. This effect is called *pushing* and can be modeled as

$$\tilde{g}_{\text{amp}}(t) = \xi_g d_{\text{ps}}(t) \qquad \text{and} \qquad \tilde{\theta}_{\text{amp}}(t) = \xi_\theta d_{\text{ps}}(t), \tag{5.10}$$

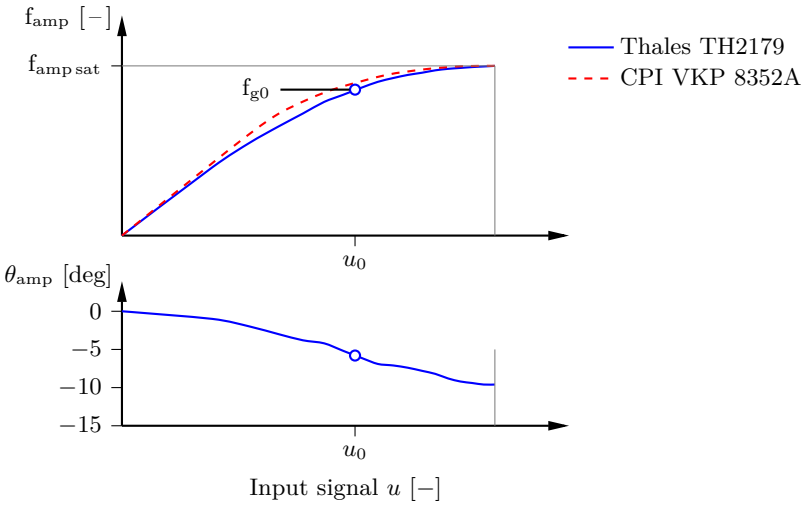


Figure 5.10 Nonlinear input-output characteristics (transfer curves) of the same klystrons as in Figure 5.9; the phase curve was not available for the CPI klystron. The circles indicate an operating point at 25% below saturation (in terms of power) corresponding to $f_{g0} = 0.87 \cdot f_{amp\,sat}$. Credit: Data from Thales and CPI, provided by Chiara Marrelli, ESS.

where $d_{ps}(t)$ is the relative variation of the supply voltage, ξ_g is the *amplitude-pushing factor* and ξ_θ is the *phase-pushing factor*. The pushing factors depend on the amplifier's operating point.

Pushing factors for klystrons. For klystrons it holds that $\xi_g \approx 5/4$ (to first order) [Gilmour, 2011]. The phase-pushing factor ξ_θ depends on the operating point and the design parameters of the klystron [Hara et al., 1998]. Typical values are around 10 °/%. For the klystrons at ESS values of ξ_θ in the range 6–8°/% ≈ 10–13 rad have been measured[4]. Note that these values are significantly larger than $\xi_g = 5/4 = 1.25$, i.e., the phase variations from supply-voltage ripple are large compared to the amplitude variations.

Noise and spurious components. Certain power amplifiers, in particular solid-state amplifiers, contribute both *flicker* noise, with a $1/f$-dependence in the baseband, spurious peaks originating from combinations of switching and nonlinear phenomena, and to a minor extent also white, thermal noise. These disturbances are most conveniently described by a spectral model.

[4] Personal communication with Morten Jensen, ESS.

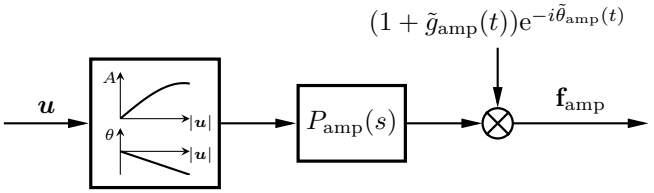


Figure 5.11 RF amplifier model which accounts for nonlinearity, limited bandwidth, and gain and phase variations from, e.g., DC-supply voltage variations. Note that the nonlinearity is complex valued, with the output amplitude and phase *shift* depending on the amplitude of the input. Of course, this separated RF amplifier model is only an approximation and the order of the three blocks was chosen to simplify analysis.

Table 5.1 The RF amplifiers that will be used for the ESS accelerator.

Cavity Type (No. Units)	$f_{\mathbf{RF}}$ [MHz]	**Output Power** (peak) [kW]	**Amplifier Type**
RFQ (1)	352.21	2900	Klystron*
Buncher (3)	–"–	30	Solid-State Amplifier
DTL (5)	–"–	2900	Klystron*
Spoke (26)	–"–	2×200	Tetrode
Medium-β (36)	704.42	1500	Klystron†
High-β (84)	–"–	1500	Klystron†

*Will have 500 W preamplifiers.
†Will have 200 W preamplifiers.

5.5.4 Example: High-power RF amplifiers for ESS[4]

High-intensity ion linacs, as the one at ESS, have several different cavity types, which calls for different RF amplifier technologies. At ESS it is planned to use three different amplifier types[5], as shown in Table 5.1.

In-house designed DC power supplies will be used to power the klystrons, since a pulsed draw of 200 MW at 14 Hz would induce excessive flicker on the regional power gridThese power supplies, which are called modulators, will have capacitors banks which are charged continuously and then discharged during the pulses.

[5]The initial plan was to use multi-beam inductive output tubes, which are more energy-efficient than klystrons, for the high-β section [Jensen et al., 2018]. This would have made ESS the first particle accelerator with such amplifiers. Unfortunately, these plans were abandoned due to a lack of funding and that the manufacturers had problems to meet the reliability requirements.

5.6 Typical delays in field control loops

Field control loops with digital LLRF systems tend to have time delays in the range of 1–2 µs. Approximate values for a few different linacs are:

- 1 µs for ESS,
- 1 µs for LCLS-II [Huang et al., 2016][6], and
- 2 µs for the European XFEL[7].

A break-down of the estimated delay of the field control loops at ESS is shown in Table 5.2. Note that 400 ns of the delay are from propagation of electromagnetic waves.

Table 5.2 Estimated time delays in the field control loops at ESS. The dynamics of the RF amplifier is modeled by a first-order system $P_{\text{amp}}(s)$ instead of as a group delay, see Section 5.5.

Source of delay	**Delay [ns]**
Propagation of electromagnetic waves	
Amplifier to cavity (waveguide), $40\,\text{m} \div 0.68c$	200
LLRF to amplifier (cable), $10\,\text{m} \div 0.82c$	40
Cavity probe to LLRF (cable), $40\,\text{m} \div 0.82c$	160
LLRF system	
ADC	100
FPGA, 48 clock cycles @117 MHz	400
DAC	80
Total	980

[6] Not counting time constants of filters in the controller.

[7] About 1 µs is due to master–slave communication. Personal communication with Sven Pfeiffer, DESY.

6

Modeling: The Accelerating Cavity

6.1 Introduction

For field control analysis it is essential with a model of how the cavity field interacts with the beam and the RF system. To a very good approximation, the cavity field can be described as a linear combination of modes that evolve independently. For most analyses it is sufficient to consider just one of these modes, namely the *accelerating mode*, that is intended for beam acceleration. In some cases, as for multi-cell elliptical cavities, it is necessary to consider *parasitic modes* to ensure feedback stability.

In this chapter we discuss how to model and understand the dynamics of the electromagnetic modes of an accelerating cavity. We begin with a short introduction to electromagnetic fields in accelerating cavities (Section 6.2), and then present a convenient, energy-based parameterization of the accelerating mode (Section 6.3) [Troeng, 2019]. The model is an extension (with beam loading) of Haus' approach to model optical cavities [Haus, 1983]. As we discuss in Section 6.4, this model avoids the problems of previous parametrizations that were inspired by equivalent electric circuits. We then introduce a type of phasor diagram in Section 6.5, which is helpful in illustrating the dynamics of the accelerating mode, and which will be used throughout the thesis. In Section 6.6 we discuss how the detuning and waveguide coupling of the accelerating mode should be selected to minimize the steady-state energy consumption. After that, we provide a suitable normalization of the accelerating mode in Section 6.7. Then, in Section 6.8 we introduce parasitic modes, and discuss how these can be normalized. Finally, in Section 6.9 we state some special relationships that hold between the properties of parasitic same-order modes in elliptical multicell cavities.

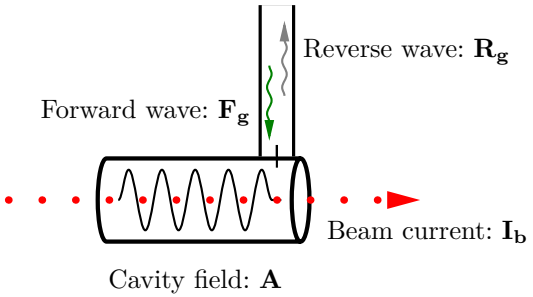


Figure 6.1 Illustration of the quantities involved in the derivation of the cavity dynamics. Measurements of the reverse wave can be used in various calibration schemes and for estimating the detuning, but we will not use it in this thesis.

6.2 Electromagnetic fields in accelerating cavities

Mode expansion of the cavity field

The dynamics of the electromagnetic cavity field is given by Maxwell's equations. If there is no beam present, then the electric field $\boldsymbol{\mathcal{E}}$ can be shown to satisfy

$$\nabla^2 \boldsymbol{\mathcal{E}} - \epsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \boldsymbol{\mathcal{E}} = 0 \tag{6.1}$$

in the interior of the cavity, where ϵ_0 denotes the vacuum permittivity and μ_0 denotes the vacuum permeability. For a closed cavity with perfectly conducting walls, the electric field can be expanded as a sum of orthogonal eigenmodes $\boldsymbol{E}_k$

$$\boldsymbol{\mathcal{E}}(\boldsymbol{r}, t) = \sum_{k=a,1,2,\ldots}^{\infty} e_k(t) \boldsymbol{E}_k(\boldsymbol{r}), \tag{6.2a}$$

where the mode amplitudes $e_k(t)$ evolve independently as

$$\frac{d^2}{dt^2} e_k(t) = -\omega_k^2 e_k(t). \tag{6.2b}$$

The subscript $k = a$ in (6.2a) indicates the accelerating mode. We see from (6.2b) that the amplitude of each mode varies sinusoidally with frequency ω_k. We will assume $\boldsymbol{E}_k$ and e_k to be scaled so that the oscillation amplitude of e_k equals the square root of the energy stored in mode k.

Cavity–beam interaction

The acceleration provided by the cavity field to a particle with charge q traversing the cavity along straight trajectory $z(t)$, is commonly modeled

under the approximation that the trajectory does not depend on the cavity field [Wangler, 2008]. Due to the linearity of Maxwell's equations, the effect of the cavity modes on the charge can then be considered individually. In what follows we will assume that only the *accelerating* mode is excited; parasitic modes are treated similarly. The energy gained by the particle from crossing the cavity is given by

$$\Delta W = q \int_0^L e_a(t(z)) E_{az}(z)\, dz, \tag{6.3}$$

where we have expressed time as a function of the particle position.

Recall that $e_a(t)$ evolves sinusoidally, with an amplitude that equals the square root of the mode energy $\sqrt{U_a}$. Thus we can write (6.3) as

$$\Delta W = q \int_0^L E_{az}(z) \cos(\omega_a t(z) + \phi)\, dz \sqrt{U_a}, \tag{6.4}$$

for some ϕ. The maximum value of the integral with respect to ϕ equals [Wangler, 2008, Exercise 2.12]

$$\alpha_a \coloneqq \left| \int_0^L E_{az}(z) \mathrm{e}^{i\omega_a t(z)}\, dz \right|. \tag{6.5}$$

We will refer to α_a as the cavity–beam-coupling parameter. It can be seen, perhaps most easily from (6.4), that α_a has units $\mathrm{V}/\sqrt{\mathrm{J}}$.

With this definition we can re-write (6.4) as

$$\Delta W = q\alpha_a \sqrt{U_a} \cos\phi_b, \tag{6.6}$$

where the angle ϕ_b is defined with respect to the bunch phase that would give maximum energy gain.

Remark 6.1 The existing accelerator literature uses the normalized shunt impedance $(r/Q)_a = \alpha_a^2/\omega_a$ (also called r-over-Q) [Wangler, 2008], for relating the accelerating voltage seen by a particle traversing the cavity to the energy stored in a cavity mode. Note that both $(r/Q)_a$ and α_a depend on the particle velocity which is implicitly given by cavity position along the linac. Apart from the particle velocity, the two quantities only depend on the cavity geometry.

Remark 6.2 The particle bunches should arrive to the cavity slightly before the instant when they would obtain maximal acceleration (quantified by the phase ϕ_b), to ensure that they are kept together by *longitudinal focusing* [Wangler, 2008, Sec. 1.3]. Longitudinal focusing is particularly important in ion linacs, where ϕ_b typically is chosen in the range $-30°$ to $-15°$. Typical values of ϕ_b for electron linacs are in the range $-3°$ to $0°$.

6.3 Baseband model of the accelerating cavity mode

Let the nominal operating RF frequency of the cavity be given by ω_{RF}, and denote the complex envelope (Chapter 4) of e_a by $\mathbf{A}$, i.e.,

$$e_a(t) = \mathrm{Re}\{\mathbf{A}_a(t)\mathrm{e}^{i\omega_{\mathrm{RF}}t}\}.$$

From the definition of e_a it follows that also $\mathbf{A}_a$ has units $\sqrt{\mathrm{J}}$, and that $|\mathbf{A}_a|^2$ equals the energy stored in the accelerating mode.

Up until the sections on parasitic modes we will only consider the accelerating mode and we therefore drop the mode index (when possible).

Transforming the differential equation (6.2b), which assumed a closed cavity with no losses, to baseband gives

$$\frac{d}{dt}\mathbf{A} = i\Delta\omega\mathbf{A} \tag{6.7}$$

where $\Delta\omega := \omega_a - \omega_{\mathrm{RF}}$ is the so-called *detuning*. By also including resistive losses, beam loading, coupling to a waveguide, and an incident wave in the waveguide (from the RF amplifier) we get (see Appendix C) the following equation for the accelerating mode

$$\boxed{\frac{d\mathbf{A}}{dt} = (-\gamma + i\Delta\omega)\mathbf{A} + \sqrt{2\gamma_{\mathrm{ext}}}\mathbf{F_g} + \frac{\alpha}{2}\mathbf{I_b}.} \tag{6.8}$$

Here, $\gamma = \gamma_0 + \gamma_{\mathrm{ext}}$ is the total decay rate of the cavity field (γ_0 corresponds to resistive losses and γ_{ext} to decay through the power coupler); $\mathbf{F_g}$ models the incident wave from the RF amplifier with $|\mathbf{F_g}|^2$ equal to the wave power ($\mathbf{F_g}$ has units $\sqrt{\mathrm{W}}$); $\mathbf{I_b}$ models the beam of charged particles with $|\mathbf{I_b}|$ equal to the corresponding DC current.

We will refer to $\mathbf{F_g}$ as the RF drive (phasor) and $\mathbf{I_b}$ as the beam phasor. These quantities and reverse wave $\mathbf{R_g}$ (see Remark 6.3 below) are illustrated in Figure 6.1. Typical cavity parameters are given in Table 6.1.

Remark 6.3 (Reverse wave) As shown in Figure 6.1, there is typically a reverse wave in the waveguide that propagates away from the cavity. The complex envelope of the reverse wave, which also has units $\sqrt{\mathrm{W}}$, takes the form

$$\mathbf{R_g} = -\mathbf{F_g} + \sqrt{2\gamma_{\mathrm{ext}}}\mathbf{A}, \tag{6.9}$$

i.e., it is a superposition of the reflection of the incident wave, and a term from field decay through the power coupler. See Appendix C.5 for details.

Table 6.1 Parameters for some different real-world cavities in high-energy linacs [Altarelli, 2007; Doolittle et al., 2016]. The first group of parameters are inherent to the cavity (sometimes γ_{ext} can be tuned), the second group of parameters gives the nominal operating point, and the third group are derived quantities of interest. The cavities of the ESS linac were illustrated in Figure 3.3.

Cavity	$\gamma_0/2\pi$ kHz	$\gamma_{\text{ext}}/2\pi$ kHz	α MV/$\sqrt{\text{J}}$	A_0 $\sqrt{\text{J}}$	I_{DC} mA	ϕ_b °	ΔW MeV	$\lvert\mathbf{F_{go}}\rvert^2$ kW
RFQ (ESS)	24	36	3.1	1.6	62.5	-45^*	3.5	1000
DTL (ESS)	3.2	8.8	3.7	5.2	62.5	-25	18	2200
Medium-β (ESS)	0	0.5	1.3	11.2	62.5	-15	14	900
TESLA (Eu-XFEL)	0	0.14	2.9	8.1	5.0	-3	24	120
TESLA (LCLS II)	0	0.016	2.9	5.5	0.1	-3	16	2.5

* The beam loading in an RFQ is always relative to the phase of the accelerating mode, i.e., $\mathbf{I_b}(t) = I_b(t)\mathrm{e}^{i(\pi-\phi_b)} \cdot \mathrm{e}^{i \arg \mathbf{A}(t)}$. The value ϕ_b for the whole RFQ is never given in RFQ specifications; the value $-45°$ is an estimate by the author based on the parameters of the individual RFQ cells.

Remark 6.4 The effective accelerating voltage $\mathbf{V}$, which is commonly used in existing cavity models, is related to $\mathbf{A}$ through $\mathbf{V} = \alpha\mathbf{A}$. If an equation in $\mathbf{V}$ is desired, multiplying (6.8) by α gives

$$\frac{d\mathbf{V}}{dt} = (-\gamma + i\Delta\omega)\mathbf{V} + \alpha\sqrt{2\gamma_{\text{ext}}}\mathbf{F_g} + \frac{\alpha^2}{2}\mathbf{I_b}. \tag{6.10}$$

Most of the advantages mentioned in Section 6.4 apply to this parametrization as well.

Remark 6.5 In equation (6.8) there are four parameters: $\gamma = \gamma_0 + \gamma_{\text{ext}}$, γ_{ext}, $\Delta\omega$ and α. The beam-coupling coefficient α is intrinsic to the geometry of the cavity, but depends on the particle velocity. The resistive decay rate γ_0 depends on the material of the cavity and its geometry (to some extent they also depend on the field amplitude). The waveguide coupling γ_{ext} is sometimes tunable (depending on the coupler design). The detuning $\Delta\omega$, corresponding to the resonance frequency of the cavity, varies due to perturbations of the cavity geometry. For normal conducting cavities, these are typically due to temperature variations which have time-scales of minutes. For superconducting cavities, the variations have time-scales of milliseconds, and are due to the Lorentz force (the electromagnetic force of the cavity field on the cavity walls), microphonics (vibrations and pressure variations in the cryogenic systems), and intentional actuation by tuning mechanisms to counter-act previous two effects.

Remark 6.6 (Decay rates) In the proposed parametrization (6.8) we considered the factor γ as a decay rate, rather than a bandwidth as in (6.11b) in the next section. This is consistent with the laser literature [Siegman, 1986], and makes it natural to write the total decay rate as $\gamma = \gamma_0 + \gamma_{\text{ext}}$; a relation that would be less intuitive in terms of bandwidths. For frequency-domain considerations, one should, of course, think of γ as a bandwidth.

6.4 On equivalent-circuit-based cavity parametrizations

Since the accelerating mode provides a voltage to the charged particles and the particle beam corresponds to a current, it is natural to model the accelerating mode as an equivalent (electric) circuit. This is the standard approach to model accelerating cavities [Schilcher, 1998; Tückmantel, 2011][1] and is also a common technique in RF engineering [Montgomery et al., 1948].

However, equivalent-circuit-based parameterizations have a number of issues. These mainly stem from that the amplifier drive needs to be considered as a fictitious (generator) current to fit it into the equivalent-circuit framework [Tückmantel, 2011].

In this section we compare the energy-based parameterization (6.8) to two of the most popular equivalent-circuit-based parametrizations. It should be emphasized that all three parametrizations describe exactly the same dynamics.

Equivalent-circuit-based parametrizations in the existing literature

Two popular, equivalent-circuit-based parametrizations in the existing literature are: Tückmantel's [Tückmantel, 2011][2],

$$\frac{d\mathbf{V}}{dt} = -\left[\frac{\omega_a}{2Q_{\text{ext}}} + \frac{\omega_a}{2Q_0} - i\Delta\omega\right]\mathbf{V} + \omega_a \frac{r_\circ}{Q}\mathbf{I_g} + \omega_a \frac{1}{2}\frac{r_\circ}{Q}\mathbf{I_{b,RF}}, \tag{6.11a}$$

and Schilcher's [Schilcher, 1998][3],

$$\frac{d\mathbf{V}}{dt} = (-\omega_{1/2} + i\Delta\omega)\mathbf{V} + R_L\omega_{1/2}\left(2\mathbf{I_g} + \mathbf{I_{b,RF}}\right). \tag{6.11b}$$

In the above equations, $\mathbf{V}$ denotes the complex envelope of the effective accelerating voltage of the cavity field; $\mathbf{I_g}$ denotes the "generator current"

[1] Often the emphasis is on steady-state relations [Padamsee et al., 2008; Wangler, 2008] which is not enough when transients and field control are considered. See also [Wangler, 2008, Sec. 5.7] for an interesting explanation of the dynamics of a waveguide-coupled cavity, and the classic references [Montgomery et al., 1948; Slater, 1950; Pedersen, 1975].

[2] Eq. (53), with minor modifications for consistency.

[3] Eq. (3.49) interpreted for complex signals, with $\mathbf{I} = 2\mathbf{I_g} + \mathbf{I_{b,RF}}$.

which models the RF amplifier drive; and $\mathbf{I}_{\mathbf{b},\mathrm{RF}}$ denotes the RF component of the beam current ($|\mathbf{I}_{\mathbf{b},\mathrm{RF}}| = 2I_{\mathrm{DC}}$ [Schilcher, 1998, A4]), see Tables 6.2a and 6.2b for a complete list of the quantities involved. The normalized shunt impedance $r_\circ/Q$ in (6.11a) is defined with respect to the *equivalent-circuit convention*, and so is the loaded resistance R_L in (6.11b). There is also the *linac convention* for which $r/Q = 2(r_\circ/Q)$, [Tückmantel, 2011].

Note that the generator current $\mathbf{I_g}$ is a fictitious quantity that is introduced to make the RF drive term fit into the equivalent-circuit framework. An additional relation is needed for how $\mathbf{I_g}$ relates to the drive power P_g, which is the physical quantity of interest,

$$P_\mathrm{g} = \frac{1}{2}\frac{r_\circ}{Q} Q_\mathrm{ext} \left|\mathbf{I_g}\right|^2 . \tag{6.11c}$$

With Table 6.2 it is easy to verify that the three parametrizations (6.11a), (6.11b) and (6.8) are equivalent. Next we compare the advantages of the proposed energy-based parametrization (6.8) and the equivalent-circuit-based parametrizations (6.11a) and (6.11b). Note that field error requirements of the form $x\%$ amplitude error and y° phase error are identical for $\mathbf{V}$ and $\mathbf{A}$.

Advantages of the energy-based parametrization (6.8)

1. The dynamic equation is cleaner. The expression for the drive power (6.11c) is simplified to $P_\mathrm{g} = \left|\mathbf{F_g}\right|^2$ and there is no need to consider the "RF component" of the beam current.

   Remark: Quantifying signal amplitudes in terms of square root of power, as for $\mathbf{F_g}$, is common in RF engineering. It is with respect to such power waves that scattering matrices (S matrices) are defined [Pozar, 2009].

   Example: The RF drive power necessary to maintain an accelerating voltage $\mathbf{V_0} = \alpha\mathbf{A}_0$, while accelerating a beam modeled by $\mathbf{I_{b0}}$, is easily found from (6.8) as

   $$P_\mathrm{g} = \frac{1}{2\gamma_\mathrm{ext}} \left|(-\gamma + i\Delta\omega)\mathbf{V_0}/\alpha + \frac{\alpha}{2}\mathbf{I_{b0}}\right|^2 .$$

   This expression is more convenient, and easier to remember, than (6.11a/6.11b) together with (6.11c).

2. The effect of variations in the cavity–waveguide coupling γ_ext and cavity–beam coupling α are transparent. The same cannot be said for the parametrizations (6.11a)/(6.11b).

   Example: The quantity $\mathbf{F_g}$ that represents the RF drive is independent of γ_ext. Contrast this to (6.11c) which includes both Q_ext and

Table 6.2 Physical quantities in the equivalent-circuit-based parametrizations (6.11a)/(6.11b), and in the proposed parametrization (6.8). The rightmost column contains the quantities expressed in the parameters of the other parameterization.

a) Quantities common to (6.11a)/(6.11b) and (6.8)			
ω_a	rad/s	Resonance frequency of the accelerating mode	
$\Delta\omega$	rad/s	Detuning of the accelerating mode, $= \omega_a - \omega_{\mathrm{RF}}$	
b) Quantities in (6.11a)/(6.11b)			
$\mathbf{V}$	V	Accelerating voltage	$\mathbf{V} = \alpha\mathbf{A}$
$\mathbf{I_g}$	A	Generator current	$2\sqrt{2\gamma_{\mathrm{ext}}}/\alpha\mathbf{F_g}$
$\mathbf{I}_{\mathbf{b},\mathrm{RF}}$	A	Beam current (RF component)	$2\mathbf{I_b}$
Q_0	–	Unloaded quality factor	$\omega_a/(2\gamma_0)$
Q_{ext}	–	External quality factor	$\omega_a/(2\gamma_{\mathrm{ext}})$
β	–	Coupling factor $= Q_0/Q_{\mathrm{ext}}$	$\gamma_{\mathrm{ext}}/\gamma_0$
Q_L	–	Loaded quality factor $= Q_0/(\beta+1)$	$\omega_a/(2\gamma)$
$\omega_{1/2}$	rad/s	Cavity bandwidth $= \omega_a/(2Q_L)$	γ
r/Q	Ω	Normalized shunt impedance, linac convention	α^2/ω_a
$r_\circ/Q$	Ω	Normalized shunt impedance, equivalent-circuit convention $= (r/Q)/2$	$\alpha^2/(2\omega_a)$
R_L	Ω	Loaded shunt impedance $= (r_\circ/Q)\cdot Q_0/(1+\beta)$	$\alpha^2/(4\gamma)$
c) Quantities in (6.8)			
$\mathbf{A}$	$\sqrt{\mathrm{J}}$	Mode amplitude	$\mathbf{V}/\sqrt{\omega_a(r/Q)}$
$\mathbf{F_g}$	$\sqrt{\mathrm{W}}$	RF drive	$\sqrt{(r/Q)Q_{\mathrm{ext}}/4}\mathbf{I_g}$
$\mathbf{I_b}$	A	Beam current	$\mathbf{I}_{\mathbf{b},\mathrm{RF}}/2$
α	$\mathrm{V}/\sqrt{\mathrm{J}}$	Field–beam coupling parameter	$\sqrt{\omega_a(r/Q)}$
γ_0	1/s	Resistive decay rate	$\omega_a/(2Q_0)$
γ_{ext}	1/s	External decay rate	$\omega_a/(2Q_{\mathrm{ext}})$
γ	1/s	Total decay rate $= \gamma_0 + \gamma_{\mathrm{ext}}$	$\omega_{1/2}$

r/Q. Thus, the parameterization (6.8) avoids the issue mentioned in [Tückmantel, 2011][4].

Example: From (6.11b) we see that the transfer function from beam-current variations to cavity field errors is

$$G_{\mathbf{I_b}\to\boldsymbol{e}}(s) = \frac{\omega_{1/2}}{s+\omega_{1/2}} R_L.$$

Recognizing the first factor as a low-pass filter with bandwidth $\omega_{1/2}$, one is led to believe that reducing $\omega_{1/2}$ reduces the field errors from beam-current ripple. However, this is incorrect, since R_L depends inversely on $\omega_{1/2}$. This confusion does not arise from (6.8).

3. The mode amplitude $\mathbf{A}$ is intrinsic to the cavity field, while the effective cavity voltage $\mathbf{V}$ depends on the velocity of the accelerated particles [Wangler, 2008]. Without a given particle velocity $\mathbf{V}$ is not well-defined.

   For electron linacs, one could assume that the velocity equals c and hence uniquely quantify the amplitude of the accelerating mode by its effective voltage. But the amplitudes of parasitic same-order modes of multi-cell cavities can obviously not be quantified this way. For example, the same-order modes of the TESLA cavity all have a negligible coupling to the beam, i.e., zero effective voltage, but they are crucial to consider in field-control analysis.

4. The equation (6.8) can be derived using basic properties of Maxwell's equations (see Appendix C). This arguably allows for a better understanding of how the model parameters relate to physical cavity properties.

Disadvantages of the energy-based parametrization (6.8)

A natural objection to the parameterization (6.8) is that it does not explicitly contain the accelerating voltage, which is the most important quantity from a beam perspective. In that case (6.10) could be used which retains the understandable structure of (6.8). Overall, there is less reason to use the energy-based parametrization (6.8) when the focus is on how the beam is affected by the cavity field. It is when the focus is on the cavity field, as from an RF system perspective, that (6.8) brings helpful intuition.

Another objection would be that the parameters in (6.8) are not the ones commonly used to for specifying cavities. However, the parameters in (6.8) arguably have more physical significance than those in (6.11a) and (6.11b).

[4] "*One word of caution is required here:* /.../ for considerations where Q_{ext} varies—as for a variable coupler e.g. in context with RF vector feedback loop gain—or where (R/Q) varies—as when particles of different speed $v = \beta c$ pass the same cavity /.../—the model currents cannot be considered constant; they have to be re-adapted each time Q_{ext} or (R/Q) change."

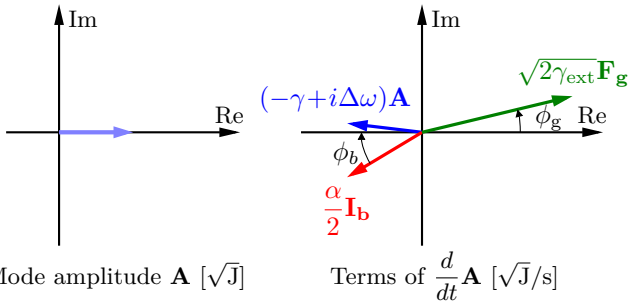


Figure 6.2 Phasor diagrams for visualizing the terms in the differential equation for the accelerating cavity mode (6.8). *Left:* Phasor for the mode amplitude. *Right:* Terms that affect the time derivative of the mode amplitude. The phasors sum to zero which indicates steady-state operation.

6.5 Phasor diagrams

To better understand the dynamics of the accelerating mode (6.8), which we restate for convenience,

$$\frac{d\mathbf{A}}{dt} = (-\gamma + i\Delta\omega)\mathbf{A} + \sqrt{2\gamma_{\text{ext}}}\mathbf{F_g} + \frac{\alpha}{2}\mathbf{I_b},$$

it will be helpful to use phasor diagrams as in Figure 6.2. The coloring of the phasors coincides with that in the above equation. To avoid clutter, the phasor for the cavity field is drawn separately from the phasors that affect its derivative. We will refer to them as the field-decay phasor (blue), the RF-drive phasor (green) and the beam-loading phasor (red).

Remark 6.7 (Phasor diagrams in existing literature) Phasor diagrams are common in existing accelerator [Wangler, 2008; Padamsee et al., 2008; Wiedemann, 2015]. In these diagrams the phasors for the RF drive and the beam loading are typically shown together with their so-called induced voltages, i.e., their steady-state effect on the cavity field. For cavity field control, one needs to understand how *variations* of the phasors $\mathbf{F_g}$ and $\mathbf{I_b}$ affect the cavity field. In this regard, the induced voltages are of little interest and leaving them out reduces clutter.

Remark 6.8 (The choice of reference phase) In previous literature, the reference phase is often chosen so that the beam phasor $\mathbf{I_b}$ is oriented along the negative real axis [Wangler, 2008, p. 348]. This is reasonable, since after all, the beam phase is the phase reference relative to which

the cavity field should be controlled. However, in a field-control context, where the objective is to keep the cavity field close to a setpoint and beam variations act as disturbances, it is arguably more natural to choose the reference phase so that the cavity-field phasor lies on the positive real axis.

With this convention we get a nice symmetry in the phasor diagrams for optimally tuned cavities (Figure 6.3), with the cavity-field phasor and the RF-drive phasor lying on the real axis. Also, amplitude variations and (small) phase variations of these two phasors correspond to variations of their real and imaginary parts, respectively.

6.6 Power-optimal coupling and detuning

In this section we compute the detuning $\Delta\omega$ and coupling γ_{ext} that minimize the RF-drive power for steady-state operation. These are standard calculations [Schilcher, 1998; Wangler, 2008], but we go through them to show what they look with the parametrization (6.8) and because we need the results in the next section.

Assume that the nominal cavity field is given by[5] $\mathbf{A_0} = A_0 > 0$ and the nominal beam phasor is given by

$$\mathbf{I_{b0}} = I_b e^{i(\pi-\phi_b)} = -I_b\cos(\phi_b) + iI_b\sin(\phi_b), \tag{6.12}$$

where[6] $-\pi/2 \le \phi_b \le 0$. Note that $\phi_b \neq \angle\mathbf{I_{b0}}$.

The corresponding stationary RF drive $\mathbf{F_{g0}}$ satisfies

$$0 = (-\gamma + i\Delta\omega)A_0 + \sqrt{2\gamma_{\text{ext}}}\mathbf{F_{g0}} + \frac{\alpha}{2}\mathbf{I_{b0}}, \tag{6.13}$$

which gives an RF drive power of

$$\begin{aligned} P_{\text{g}} = |\mathbf{F_{g0}}|^2 &= \frac{1}{2\gamma_{\text{ext}}}\left|(-\gamma + i\Delta\omega)A_0 + \frac{\alpha}{2}\mathbf{I_{b0}}\right|^2 \\ &= \frac{1}{2\gamma_{\text{ext}}}\left[\left((\gamma_0+\gamma_{\text{ext}})A_0 + \frac{\alpha}{2}I_b\cos\phi_b\right)^2 + \left(\Delta\omega A_0 + \frac{\alpha}{2}I_b\sin\phi_b\right)^2\right]. \end{aligned}$$

From this expression we see that the detuning $\Delta\omega$ that minimizes the power consumption is given by

$$\Delta\omega^\star = -\frac{\frac{\alpha}{2}I_b\sin\phi_b}{A_0}. \tag{6.14}$$

[5] Recall Remark 6.8.

[6] This condition is necessary in order for the beam to be both accelerated and longitudinally focused [Wangler, 2008, p. 179].

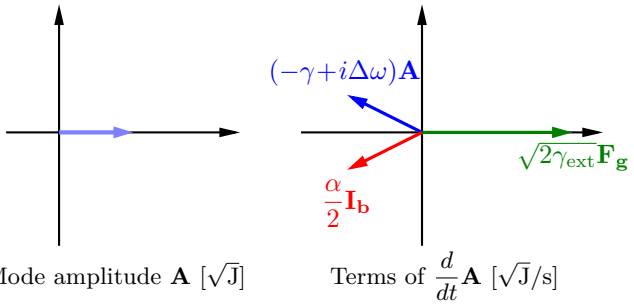


Figure 6.3 Phasor diagram for a superconducting cavity that is optimally tuned and optimally coupled.

If the cavity is optimally tuned, then (at steady-state) the RF-drive phasor lies on the positive real axis, and the imaginary parts of the decay phasor and the beam-loading phasor have equal magnitudes, but opposite signs. It can be seen that Figures 6.3 and 6.4 correspond to optimally tuned cavities, but that Figure 6.2 does not.

With power-optimal detuning we have

$$P_{\rm g}\Big|_{\Delta\omega=\Delta\omega^\star} = \frac{1}{2\gamma_{\rm ext}}\left((\gamma_0+\gamma_{\rm ext})A_0 + \frac{\alpha}{2}I_b\cos\phi_b\right)^2.$$

Minimizing this expression with respect to $\gamma_{\rm ext}$, gives the power-optimal coupling coefficient[7]

$$\gamma_{\rm ext}^\star = \gamma_0 + \frac{\frac{\alpha}{2}I_b\cos\phi_b}{A_0}. \tag{6.15}$$

Thus, given A_0 and $\mathbf{I_{b0}}$, the minimal power consumption equals

$$P_{\rm g}^\star = 2\gamma_0 A_0^2 + \alpha A_0 I_b\cos\phi_b.$$

That is, all energy in the forward wave is either dissipated in the cavity walls, or is transferred to the particle beam—no power is wasted in the reverse wave.

The total decay rate, assuming optimal coupling, is given by

$$\gamma = 2\gamma_0 + \frac{\frac{\alpha}{2}I_b\cos\phi_b}{A_0}. \tag{6.16}$$

[7] We wish to minimize $2f(x) = (ux+v)^2/x = u^2x + 2uv + v^2/x$ wrt $x > 0$. Differentiating gives $2f'(x) = u^2 - v^2/x^2$, from which we find the optimal point $x^\star = v/u$, at which $f(x^\star) = 2vu$.

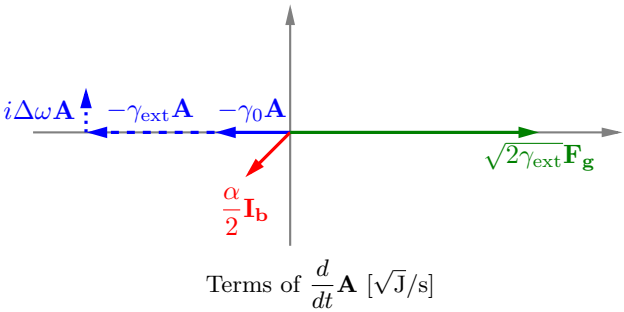


Figure 6.4 Phasor diagram with the terms of the time derivative of the cavity field. The considered cavity is optimally tuned ($\Delta\omega A_0 = -(\alpha/2)\text{Im}\,\mathbf{I_{b0}}$), optimally coupled ($\gamma_{\text{ext}}A_0 = \gamma_0 A_0 - (\alpha/2)\text{Re}\,\mathbf{I_{b0}}$), and normal conducting ($\gamma_0 > 0$); compare (6.14) and (6.15).

We have found that it sometimes gives insight to think about the second term in (6.15) as the *decay rate of the cavity field due to beam loading* (at the nominal operating point). This motivates the definition

$$\gamma_{\text{beam}} = \gamma_{\text{beam}}(A_0, \mathbf{I_{b0}}) \coloneqq \frac{\frac{\alpha}{2} I_b \cos\phi_b}{A_0}. \tag{6.17}$$

Using (6.17), we can write (6.15) more intuitively as

$$\gamma^\star_{\text{ext}} = \gamma_0 + \gamma_{\text{beam}}.$$

Remark 6.9 For pulsed accelerators, the value of γ_{ext} that minimizes the overall power consumption is somewhat larger than $\gamma^\star_{\text{ext}}$ since this gives a shorter filling time [Ayvazyan et al., 2010]. This is particularly important when the pulses are short compared to the filling time. It is also better to choose γ_{ext} larger than $\gamma^\star_{\text{ext}}$ if there are significant detuning variations during the flat-top (e.g., from microphonics) [Merminga and Delayen, 1996].

6.7 Normalized cavity dynamics

Normalization

Requirements on cavity-field errors, and specifications on amplifier ripple and beam-current ripple are typically given in relative terms, i.e., on the form $x\%$ and $y°$. By normalizing the dynamics from disturbances to field errors (of the accelerating mode), it is easy to compute the relative field errors that result from relative disturbances.

For control design, it is convenient if the static gain from control action to mode amplitude equals 1.

Let us therefore introduce normalized phasors for the cavity field, RF drive, and beam loading,

$$\mathbf{a} := \frac{1}{A_0}\mathbf{A} \tag{6.18a}$$

$$\mathbf{f_g} := \frac{1}{\gamma A_0}\sqrt{2\gamma_{\text{ext}}}\mathbf{F_g} \tag{6.18b}$$

$$\mathbf{i_b} := \frac{1}{\gamma A_0}\frac{\alpha \mathbf{I_b}}{2}. \tag{6.18c}$$

Scaling equation (6.8) by $1/A_0$, now gives

$$\boxed{\dot{\mathbf{a}} = (-\gamma + i\Delta\omega)\mathbf{a} + \gamma(\mathbf{f_g} + \mathbf{i_b}).} \tag{6.19}$$

The transfer function from $\mathbf{f_g}$ and $\mathbf{i_b}$ to $\mathbf{a}$ is given by

$$P_a(s) := \frac{\gamma}{s + \gamma - i\Delta\omega}, \tag{6.20}$$

where the subscript a indicates that we are considering the accelerating mode.

Relations at nominal operating point

Consider steady-state operation at some nominal operating point ($\mathbf{a_0} = 1, \mathbf{f_{g0}}, \mathbf{i_{b0}}$). For an *optimally tuned* cavity it follows from (6.14) that

$$\gamma \text{Im}\, \mathbf{i_{b0}} + \Delta\omega = 0. \tag{6.21}$$

For an *optimally coupled* cavity, it follows from (6.15) that

$$-1 \leq \text{Re}\, \mathbf{i_{b0}} \leq 0. \tag{6.22}$$

For an *optimally tuned and optimally coupled* cavity we have that $\mathbf{f_{g0}} = 1 - i\Delta\omega/\gamma - \mathbf{i_{b0}}$ is real, and that

$$1 \leq \mathbf{f_{g0}} \leq 2. \tag{6.23}$$

For a superconducting cavity ($\gamma_0 = 0$) that is optimally tuned and coupled, we have that $\text{Re}\, \mathbf{i_b} = -1$ and $\mathbf{f_g} = 2$.

Dynamics around nominal operating point

Cavity field stability is typically evaluated around some nominal operating point. For this reason it is meaningful to introduce the normalized field error $\boldsymbol{z}$ through

$$\boldsymbol{z} = \mathbf{a} - 1.$$

Table 6.3 Normalized parameters for the cavities in Table 6.1.

Cavity	$\gamma/2\pi$ kHz	$\Delta\omega/2\pi$ kHz	$\mathbf{f_{g0}}$ –	$\lvert\mathbf{i_{b0}}\rvert$ –	ϕ_b °
RFQ (ESS)	60	5.1	1.1	0.2	−45
DTL (ESS)	12	1.5	1.3	0.3	−25
Medium-β (ESS)	0.5	0.1	2.1	1.1	−15
TESLA (Eu-XFEL)	0.14	≈ 0	2.0	1.0	≈ 0
TESLA (LCLS II)	0.016	≈ 0	1.3	0.3	≈ 0

It is clear that a small field error $\boldsymbol{z} = z_{\mathrm{Re}} + iz_{\mathrm{Im}}$ approximately corresponds to an amplitude error of $z_{\mathrm{Re}} \cdot 100\,\%$ and a phase error of z_{Im} rad.

In the case of an ideal amplifier, we would have $\mathbf{f_g} = \mathbf{f_{g0}} + \tilde{\mathbf{f}}_\mathbf{g}$ where $\tilde{\mathbf{f}}_\mathbf{g}$ corresponds to control action from the field controller. However, due to variations of the amplifier's gain and phase shift, denoted by $\tilde{g}_{\mathrm{amp}}$ and $\tilde{\theta}_{\mathrm{amp}}$ (see Section 5.5), we have

$$\mathbf{f_g} = (1 + \tilde{g}_{\mathrm{amp}})\mathrm{e}^{-i\tilde{\theta}_{\mathrm{amp}}}(\mathbf{f_{g0}} + \tilde{\mathbf{f}}_\mathbf{g}). \tag{6.24}$$

We will assume that the variations $\tilde{g}_{\mathrm{amp}}$ and $\tilde{\theta}_{\mathrm{amp}}$ are small, and introduce, $\boldsymbol{d_g} \coloneqq \tilde{g}_{\mathrm{amp}} - i\tilde{\theta}_{\mathrm{amp}}$, which allows us to write

$$(6.24) \approx (1 + \boldsymbol{d_g})(\mathbf{f_{g0}} + \tilde{\mathbf{f}}_\mathbf{g}) \approx \mathbf{f_{g0}} + \tilde{\mathbf{f}}_\mathbf{g} + \mathbf{f_{g0}}\boldsymbol{d_g}. \tag{6.25}$$

Similarly, relative beam loading variations $\boldsymbol{d_b}$ affect $\mathbf{i_b}$ according to

$$\mathbf{i_b} = (1 + \boldsymbol{d_b})\mathbf{i_{b0}}.$$

Plugging these expressions into (6.19) and ignoring second-order terms give

$$\dot{\boldsymbol{z}} = (-\gamma + i\Delta\omega)\boldsymbol{z} + \gamma\left(\tilde{\mathbf{f}}_\mathbf{g} + \mathbf{f_{g0}}\boldsymbol{d_g} + \mathbf{i_{b0}}\boldsymbol{d_b}\right). \tag{6.26}$$

We see that the transfer function from relative disturbances $\boldsymbol{d_g}$ and $\boldsymbol{d_b}$ to relative field errors $\boldsymbol{z}$ are given by

$$P_{d_\mathrm{g}\to z}(s) = \mathbf{f_{g0}}P_a(s), \tag{6.27a}$$
$$P_{d_\mathrm{b}\to z}(s) = \mathbf{i_{b0}}P_a(s), \tag{6.27b}$$

where $P_a(s)$ is defined in (6.20). In these equations, the nominal phasors $\mathbf{f_{g0}}$ and $\mathbf{i_{b0}}$ act as complex-valued coefficients that quantify the impact of relative disturbances.

Example: The corresponding normalized cavity parameters γ, $\mathbf{f_{g0}}$, and $\mathbf{i_{b0}}$ for the cavities in Table 6.1 are shown in Table 6.3.

Remark 6.10 (Time-varying detuning) In superconducting cavities, the detuning $\Delta\omega$ varies over time due to microphonics and (in pulsed accelerators) Lorenz-force detuning. Denote the detuning variations by $\widetilde{\Delta\omega}$ and the define the normalized detuning variations $d_{\Delta\omega} \coloneqq \widetilde{\Delta\omega}/\gamma$. These normalized detuning variations correspond to a term $\gamma \cdot i d_{\Delta\omega}$ on the right-hand side of (6.26). The transfer function from normalized detuning variations to field errors is given by $P_{d_{\Delta\omega}\to z}(s) = P_a(s)$.

The effect of Lorenz force detuning can be mitigated by iterative learning control and fast cavity-resonance-frequency control. The microphonics are typically infeasible to predict, but they are relatively slow (20–200 Hz) and have a magnitude of tens of hertz. Hence, it is only for the most narrow-bandwidth cavities and strictest requirements on control performance that this effect is non-negligible. For this reason, it was not included in (6.26).

An example[8] of where the impact of detuning on field errors is significant is for the superconducting cavities of LCLS-II which have bandwidths $\gamma = 16\,\text{Hz}$ and need a phase-stability better than $0.01°$.

Directionality of disturbances

The cavity dynamics around an operating point (6.19) is rotationally invariant, in the sense that its complex-valued dynamics is linear (corresponds to time-invariance of the physical bandpass system). However, the disturbances $\boldsymbol{d_g}$ and $\boldsymbol{d_b}$ have far from circular distributions, and rather tend to have certain directionality (Section 4.5).

For example, RF amplifiers such as klystrons, typically have more phase variations than amplitude variations, corresponding to that the $\boldsymbol{d_g}$ is dominantly imaginary. Beam-current ripple on the other hand affects the magnitude of $\mathbf{i_b}$ (corresponding to a real $\boldsymbol{d_b}$)[9]. The directionality of these disturbances are readily visualized as in Figure 6.5, using the phasor diagrams of Section 6.5. For circular accelerators, synchrotron oscillations would correspond to an imaginary $\boldsymbol{d_b}$.

6.8 Parasitic modes

When parasitic modes need to be considered in addition to the accelerating mode, the subscript a will indicate the accelerating mode, and numeric subscripts $1, 2, \ldots$ will indicate parasitic modes.

[8] Personal communication with Larry Doolittle, LBNL.

[9] There will also be variations in the beam phase (arrival time of the bunches) due to field-control errors in the upstream cavities; but with satisfactory field control these variations should be small.

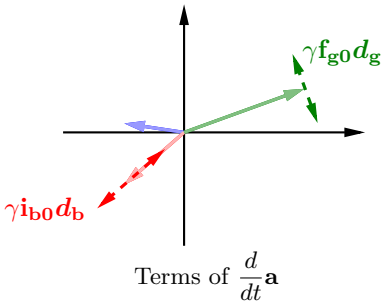


Figure 6.5 Visualization of the directions in which phase variations of the RF drive ($\boldsymbol{d_g}$ purely imaginary) and amplitude variations of the beam current ($\boldsymbol{d_b}$ purely real) affect the cavity field. The effect on the accelerating cavity mode obtained by filtering the illustrated variations through the transfer function $P_a(s)$ in (6.20).

Model in physical units

The complex envelopes of the parasitic modes in (6.2) satisfy essentially the same dynamics as the accelerating mode, namely

$$\frac{d\mathbf{A}_k}{dt} = (-\gamma_k + i\Delta\omega_k)\mathbf{A}_k + \sqrt{2\gamma_{\mathrm{ext}\,k}}\mathbf{F_g} + \frac{\boldsymbol{\alpha}_k}{2}\mathbf{I_b}.$$

The phase of complex envelope $\mathbf{A}_k$ of the kth mode has been defined so that the coefficient in front of $\mathbf{F_g}$ is real. Note that the cavity–beam-coupling parameters $\boldsymbol{\alpha}_k$ of the parasitic modes in general are complex. The coupling between the cavity modes and the pickup probe is quantified by complex coefficients $\boldsymbol{c}_k$, i.e., the voltage sensed by the pickup probe is given by

$$\mathbf{V}_{\mathrm{pu}} = \sum_{k=a,1,2,\ldots}^{N} \boldsymbol{c}_k \mathbf{A}_k.$$

The model is conveniently visualized with the block diagram in Figure 6.6.

Example: Figure 6.7 shows the variation of the cavity–beam-coupling parameter of the accelerating mode and a few parasitic modes of the superconducting cavities along the ESS proton linac.

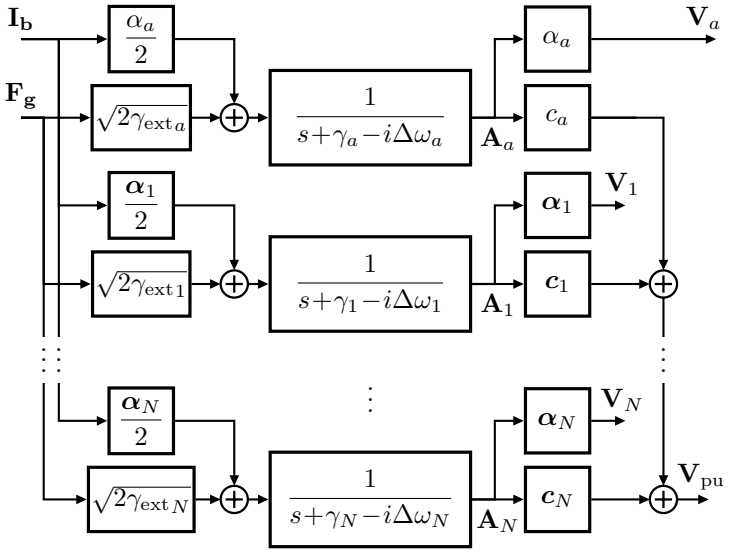


Figure 6.6 Block diagram of how the accelerating mode and the parasitic modes of a cavity couple to the RF system (through the power coupler and the pickup probe) and the beam. The coefficients $\boldsymbol{c}_k$ quantify the coupling between the modes to the pickup probe.

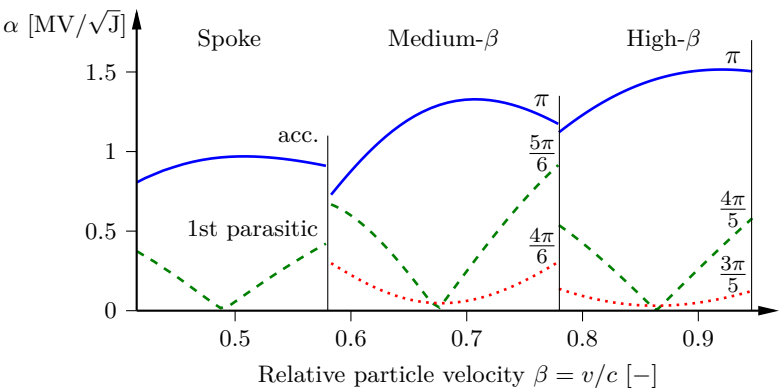


Figure 6.7 The cavity–beam-coupling parameter as a function of the particle velocity for the accelerating mode and the closest parasitic modes of the superconducting cavities along the ESS linac. The relation between the cavity–beam-coupling parameter and the more commonly used parameter r/Q is given in Remark 6.1. Data provided by Aaron Farricker.

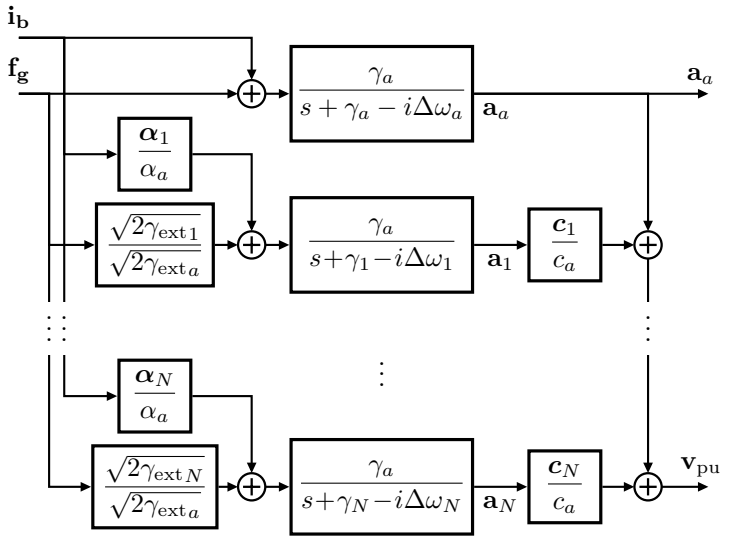


Figure 6.8 Normalized model of a cavity with parasitic modes which is derived from the model in Figure 6.6 by normalizing according to (6.28)

Normalized model with parasitic modes

For field control analysis, it is useful to normalize the model in Figure 6.6 as in the previous section, by introducing the dimension-free variables

$$\mathbf{a}_k := \frac{1}{A_{a0}}\,\mathbf{A}_k \tag{6.28a}$$

$$\mathbf{f_g} := \frac{1}{\gamma_a A_{a0}}\sqrt{2\gamma_{\mathrm{ext}\,a}}\mathbf{F_g} \tag{6.28b}$$

$$\mathbf{i_b} := \frac{1}{\gamma_a A_{a0}}\frac{\alpha_a \mathbf{I_b}}{2} \tag{6.28c}$$

$$\mathbf{v}_{\mathrm{pu}} := \mathbf{V}_{\mathrm{pu}}/(c_a A_{a0}). \tag{6.28d}$$

This gives the model in Figure 6.8.

Lumped model

The model in Figure 6.8 can be lumped into the one in Figure 6.9, where

$$P_a(s) = \frac{\gamma_a}{s+\gamma_a-i\Delta\omega_a},$$

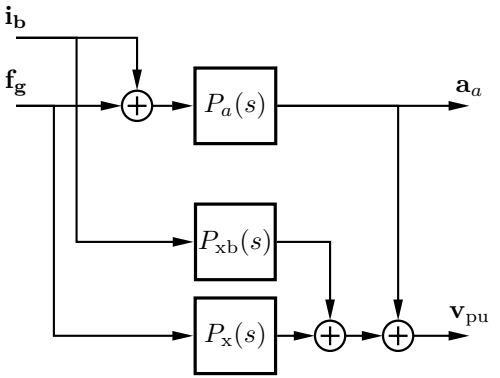


Figure 6.9 Lumped normalized cavity model which is equivalent to the one in Figure 6.8, but with the dynamics of the parasitic modes lumped into $P_{\mathrm{x}}(s)$ and $P_{\mathrm{xb}}(s)$, which are defined in (6.29) and (6.30). Note that $P_{\mathrm{xb}}(s) = 0$ for elliptical cavities operating at their design velocity (typically holds for electron cavities), see Remark 6.12.

defined in (6.20), is the dynamics of the accelerating mode,

$$P_{\mathrm{x}}(s) \coloneqq \gamma_a \sum_{k=1}^{N} \frac{\sqrt{2\gamma_{\mathrm{ext}\,k}}}{\sqrt{2\gamma_{\mathrm{ext}\,a}}} \frac{\boldsymbol{c}_k}{\boldsymbol{c}_a} \frac{1}{s + \gamma_k - i\Delta\omega_k} \tag{6.29}$$

is the transfer function from $\mathbf{f_g}$, through the parasitic modes, to $\mathbf{v}_{\mathrm{pu}}$, and

$$P_{\mathrm{xb}}(s) \coloneqq \gamma_a \sum_{k=1}^{N} \frac{\boldsymbol{\alpha}_k}{\alpha_a} \frac{\boldsymbol{c}_k}{c_a} \frac{1}{s + \gamma_k - i\Delta\omega_k} \tag{6.30}$$

is the transfer function from $\mathbf{i_b}$, through the parasitic modes, to $\mathbf{v}_{\mathrm{pu}}$.

Simplified, normalized model

In many cases it is acceptable to approximate the model in Figure 6.9, with the one in Figure 6.10, where

$$P_{\mathrm{cav}}(s) \coloneqq P_a(s) + P_{\mathrm{x}}(s). \tag{6.31}$$

The model in Figure 6.10 has only one input, which simplifies analysis.

Comparing the simplified model to the full model in Figure 6.8, we see that the simplification amounts to putting $\boldsymbol{\alpha}_n = \alpha_a \sqrt{2\gamma_{\mathrm{ext}\,k}}/\sqrt{2\gamma_{\mathrm{ext}\,a}}$. This approximation is far from true, but since $P_a(s)$ is narrowband, these approximation errors have little impact on $\mathbf{a}_a$, which is the quantity of interest. The fact that the spectrum of $\mathbf{i_b}$ tends to be narrowband around zero frequency

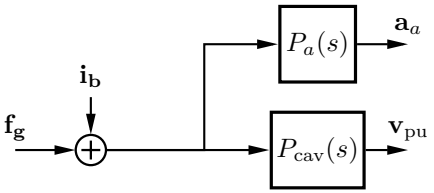


Figure 6.10 Simplified normalized cavity model, obtained from the model in Figure 6.9 by replacing the coupling $P_{\mathrm{xb}}(s)$ between the beam and the pickup probe through the parasitic modes by $P_{\mathrm{x}}(s)$. The error due to this approximation is typically small, since the spectrum of the beam-variations tend to narrowband around the nominal RF frequency. Note that this model does not capture the issue with beam-loading induced control errors that is discussed in Section 11.6. Recall from (6.31) that $P_{\mathrm{cav}}(s) = P_a(s) + P_{\mathrm{x}}(s)$.

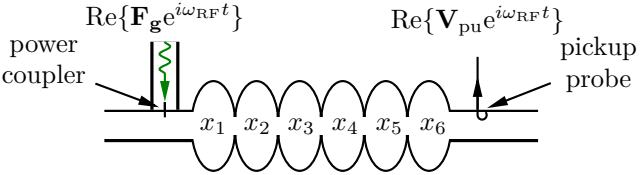


Figure 6.11 Illustration of a 6-cell elliptical cavity. The following quantities of the bandpass model in Appendix D are indicated: the RF drive, the probe voltage, and the electromagnetic-field amplitudes in each cell. The fields in the cells are weakly coupled and give rise to the same type of mode shapes as for a chain of weakly-coupled oscillators.

and that notch filters often are introduced at the parasitic-mode frequencies, further reduce the errors resulting from this approximation.

However, the issue with beam-loading-induced field-control errors that is discussed in Section 11.6 is not captured by this simplified model.

6.9 Parasitic same-order modes of elliptical multicell cavities

Elliptical multicell cavities (Figure 6.11) are suitable for accelerating particle beams with velocities greater than $0.5c$ (half the light speed). However, these cavities intrinsically have parasitic electromagnetic modes that are close in frequency to the accelerating mode. These modes are known as same-order modes, or fundamental-passband modes, and arise from the coupling between the fundamental electromagnetic modes of each cell. They correspond to the eigenmodes of a chain of weakly coupled oscillators.

The dynamics of the parasitic same-order modes of elliptical cavities are captured by the general mode structure in Figure 6.8. However, due to the special cavity geometry the parameters $\gamma_{\mathrm{ext}\,k}$, $\boldsymbol{c}_k$ and $\Delta\omega_k$ are related and depend only on the number of cells N, the cell-to-cell coupling factor k_{cc}, the coupling γ_{pc} between the waveguide and the first cell through the power coupler, and the cell resonance frequency ω_{cell}.

Remark 6.11 There are two different definitions of the cell-to-cell coupling factor k_{cc} in the accelerator literature. There is the *full-passband-width* definition, for which k_{cc} is the relative frequency difference between the same-order modes with lowest and highest frequencies [Wangler, 2008; Sekutowicz, 2010]. For this definition, k_{cc} corresponds to the per-cycle decay rate of a cell's electromagnetic energy due to coupling to one neighboring cell. There is also the *half-passband-width* definition [Ferrario et al., 1996; Padamsee et al., 2008], for which k_{cc} is half as large. In this thesis we will use the second definition since this simplifies the computations in Appendix D.

Notation. Instead of labeling the modes by a, 1, 2, ... as in the previous chapter, we will use the standard notation for same-order modes. In an N-cell cavity, the modes are referred to as the π/N mode, the $2\pi/N$ mode, up to the π mode. This naming indicates the cell-to-cell phase of the mode. Typically, it is the π mode that is used as the accelerating. We will denote the π mode with a subscript π, and quite opposite to the previous convention, denote the $n\pi/N$ mode by a subscript n.

Parameters for the dynamics parasitic same-order modes

The coefficients of the parasitic same-order modes ($n < N$) are shown Appendix D to be related to the coefficients of the π mode according to

$$\Delta\omega_n \approx (R_n^2 - 2)k_{cc}\omega_{\mathrm{cell}} \tag{6.32}$$

$$\gamma_{\mathrm{ext}\,n} = R_n^2\gamma_{\mathrm{ext}\,\pi} \tag{6.33}$$

$$\boldsymbol{c}_n = (-1)^{N-n}R_n, \tag{6.34}$$

where $R_n \coloneqq \sqrt{2}\sin(n\pi/(2N))$. Values of R_n^2 for some different N and n are shown in Table 6.4. The resistive decay rate equals γ_0 for all the modes. The parameters for the same-order modes of the ESS medium-β cavities are given in Table 6.5.

With the relationships (6.32)–(6.34) the general model in Figure 6.8 simplifies to the special form in Figure 6.12.

Remark 6.12 When the particle velocity equals the design velocity of an elliptical cavity it holds that $\boldsymbol{\alpha}_k = 0$ for the same-order modes. This relation typically always holds for elliptical cavities in electron linacs, since electrons quickly approach the speed of light, due to their small rest mass.

Table 6.4 The ratio $\gamma_{\mathrm{ext}_n}/\gamma_{\mathrm{ext}_\pi} = R_n^2 = 2\sin^2(n\pi/(2N))$ evaluated for different values of n and N.

N	n							
	8	7	6	5	4	3	2	1
5					1.81	1.31	0.69	0.19
6				1.87	1.50	1.00	0.50	0.13
9	1.94	1.77	1.50	1.17	0.83	0.50	0.23	0.06

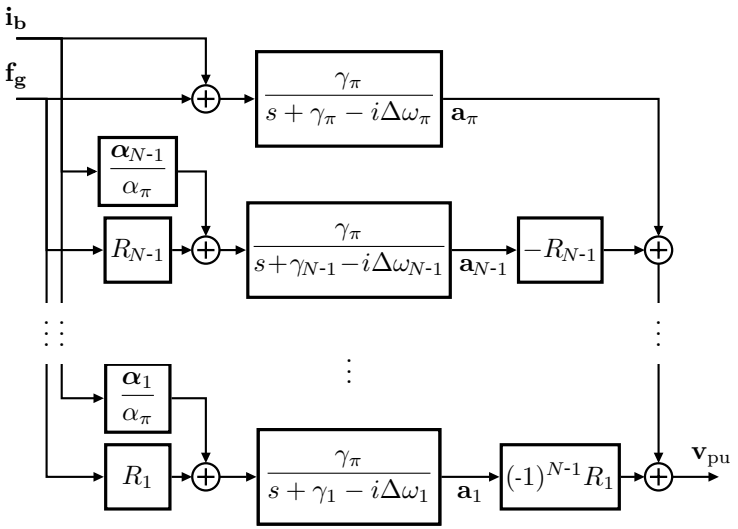


Figure 6.12 Normalized cavity model with parasitic same-order modes. A subscript n indicates the $n\pi/N$ mode.

Table 6.5 Parameters of the same-order modes of an ESS medium beta cavity (6 cells, $k_{cc} = 0.0077$). The table shows the offset frequencies $\Delta\omega_n$ relative to the accelerating π mode and the external decay rates γ_{ext_n} which correspond to the (half) bandwidths of the modes. Note that it is only the $5\pi/6$ mode that is within 1 MHz of the accelerating mode.

	π	$5\pi/6$	$4\pi/6$	$3\pi/6$	$2\pi/6$	$\pi/6$
$\Delta\omega_n$ [MHz]	0	-0.7	-2.7	-5.3	-8.0	-10
γ_{ext_n} [Hz]	500	940	750	500	250	70

Transfer function from RF drive to pickup probe

For field control analysis, it is important to be familiar with the transfer function from $\mathbf{f_g}$ to $\mathbf{v}_{\text{pu}}$ in Figure 6.12; it is given by

$$\begin{aligned} P_{\text{cav}}(s) &= \gamma_\pi \sum_{n=1}^{N} (-1)^{N-n} \frac{R_n^2}{s + \gamma_n - i\Delta\omega_n} \\ &= \frac{\gamma_\pi}{\gamma_{\text{ext}_\pi}} \sum_{n=1}^{N} (-1)^{N-n} \frac{\gamma_{\text{ext}_n}}{s + \gamma_n - i\Delta\omega_n}, \end{aligned} \tag{6.35}$$

where $\gamma_n \coloneqq \gamma_0 + \gamma_{\text{ext}_n}$ is the total decay rate of the $n\pi/N$ mode. For superconducting cavities with $\gamma_0 \ll \gamma_{\text{ext}_n}$, i.e., $\gamma_{\text{ext}_n} \approx \gamma_n$, (6.35) simplifies to

$$P_{\text{cav}}(s) = \sum_{n=1}^{N} (-1)^{N-n} \frac{\gamma_n}{s + \gamma_n - i\Delta\omega_n}. \tag{6.36}$$

Transfer function characteristics of elliptical multicell cavities

Let us observe some characteristics of the transfer function (6.35). First, since the numbers γ_n are small relative to the differences between the numbers $\Delta\omega_n$, the transfer function $P_{\text{cav}}(s)$ has sharp resonance peaks at (baseband) frequencies $\Delta\omega_n$. Note that $\Delta\omega_n < 0$ for $n < N$, so all parasitic same-order modes have negative (baseband) frequencies.

For superconducting cavities ($\gamma_0 \ll \gamma_{\text{ext}_n}$) we see from (6.36) that all peaks approximately have magnitude one. For the opposite extreme of $\gamma_0 \gg \gamma_{\text{ext}_n}$ which could correspond to S21-measurements of a room-temperature cavity with a network-analyzer, we have instead that the peak magnitude of the $n\pi/N$ mode equals $\gamma_{\text{ext}_n}/\gamma_{\text{ext}_\pi} = R_n^2$. See Table 6.4 for values of R_n^2 for some different number of cells. Other cases fall in between these extremes. See Figures 6.13 and 6.14 for Bode magnitude plots of the extreme cases. Figure 6.14 additionally shows the good agreement of the model to measurement data.

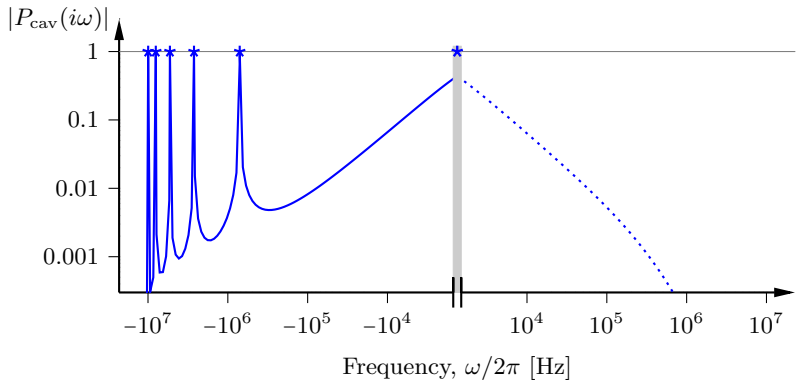


Figure 6.13 Bode magnitude plot of the transfer function (6.35) with the parameters of a superconducting 6-cell ESS medium-β cavity which has the nominal parameters $\omega_\pi/(2\pi) = 704.42\,\mathrm{MHz}$, $k_{cc} = 0.0077$ (half-passband-width definition), $\gamma_{\mathrm{ext}\,\pi}/(2\pi) = 500\,\mathrm{Hz}$, and $\gamma_0 = 0$.

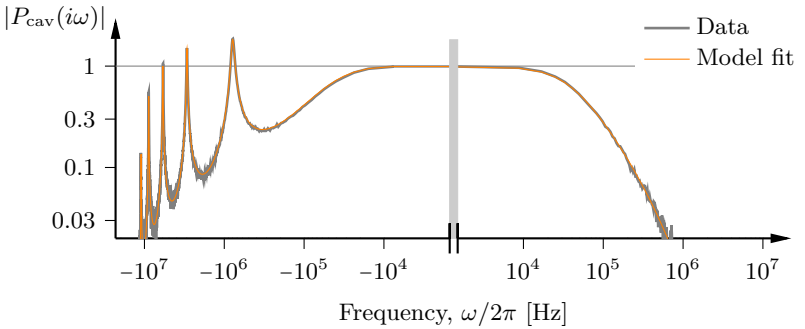


Figure 6.14 Parameter fit of the model (6.35) to the measured transfer function of a *room-temperature* ESS medium-β cavity (6 cells). The data was frequency-shifted so that the peak of the π mode was at zero frequency. The resistive decay rate was estimated to $\gamma_0/(2\pi) = 35\,\mathrm{kHz}$. Note that for superconducting cavities where the external decay rate $\gamma_{\mathrm{ext}\,\pi}$ is significantly larger than γ_0 the resonance peaks have equal magnitudes (Figure 6.13), but for a normal conducting cavity as in this figure, the peak magnitudes are related by R_n^2. Note also that the models in this and the preceding figure are normalized with respect to the accelerating mode; the peaks of the normal conducting cavity would all be lower than the peaks of the superconducting cavity in absolute terms. The measurement was done by P. Pierini, ESS.

7

Modeling: Summary

7.1 Modeling for cavity field control

7.1.1 Full nonlinear model

Collecting the results from the two previous chapters we get the baseband model of the field control loop in Figure 7.1, where τ_{LLRF} is the delay from signal processing in the LLRF system, τ_1 is the propagation delay of the electric waves from the RF amplifier to the cavity, and τ_2 is the propagation delay from the cavity pickup probe to the LLRF system. The factors $\mathrm{e}^{-i\theta_1}$ and $\mathrm{e}^{-i\theta_2}$ are phase shifts from considering the propagation delays in the baseband (see Section 4.2). We have left out the phase shift between the LLRF system and the amplifier since we will not consider inner loops around the amplifier.

The control objective is to keep the mode amplitude of the accelerating mode, $\mathbf{a}_a$, at some nominal value $\mathbf{a}_a^\star$, which due to normalization is given by $\mathbf{a}_a^\star = 1$. However, the mode amplitude $\mathbf{a}_a$ is not directly measured by the LLRF system, since the measurement $\mathbf{v}_{\text{pu}}$ from the pickup probe is a linear combination of all the mode amplitudes, including those of parasitic modes. The problem of deciding the setpoint $\boldsymbol{y}_{\text{sp}}$ so that $\mathbf{a}_a$ is controlled to its desired value is an interesting calibration problem. Another problem is to decide on a desirable value for $\mathbf{a}_a$. These questions are outside the scope of cavity field control. We will simply assume that $\mathbf{a}_a$ should be controlled to its nominal value of 1.

7.1.2 Linear large-signal model

A full model as the one in Figure 7.1 is straightforward to implement and simulate using a modeling language such as Simulink. However, we will make some minor simplifications that allow for linear control analysis and a better understanding of the field control dynamics. All these assumptions are rather standard in the field control literature.

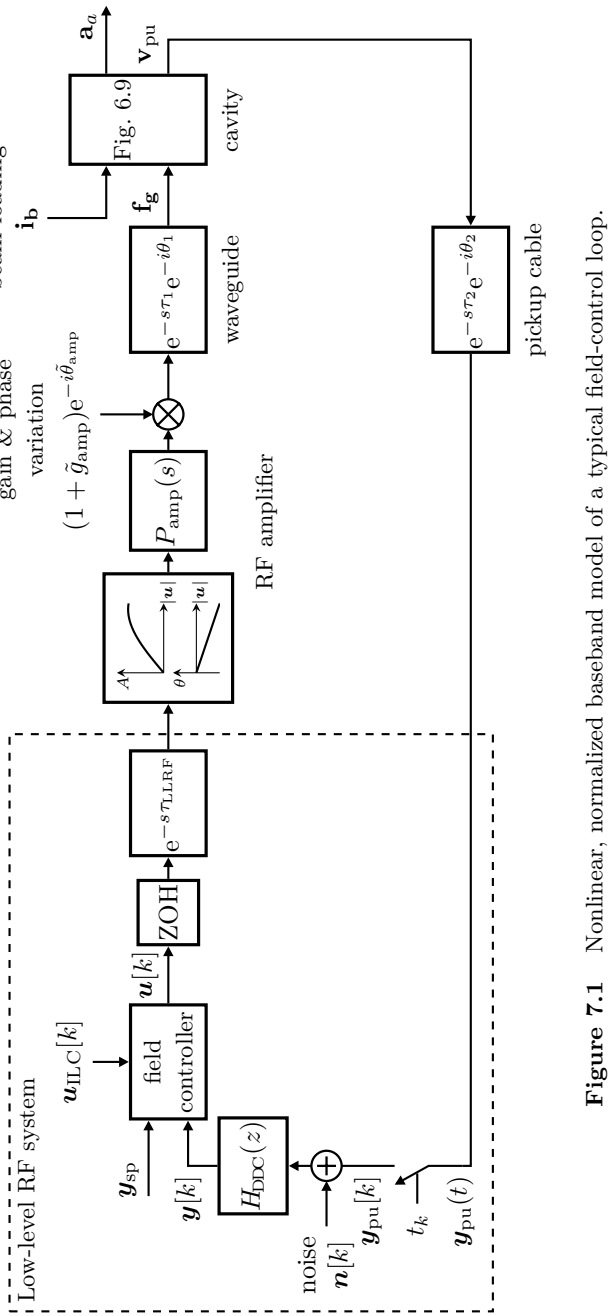


Figure 7.1 Nonlinear, normalized baseband model of a typical field-control loop.

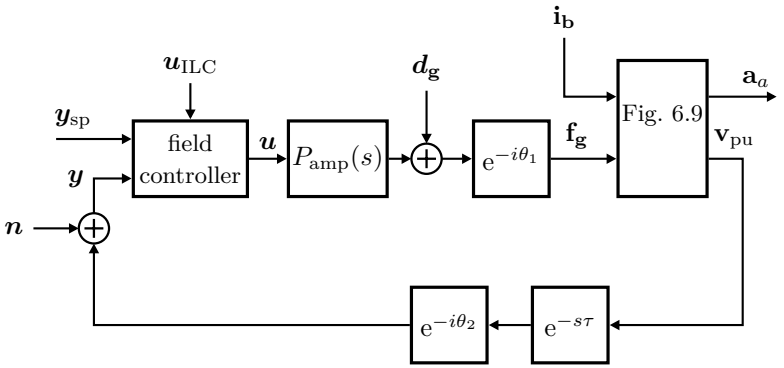


Figure 7.2 Linear, large-signal, baseband model of the field control loop, which has been derived from the nonlinear model in Figure 7.1 through some simplifying assumption.

1. The amplifier will be assumed to be linear. Often the amplifier is operated with some margin from saturation (back-off) which makes this a reasonable approximation. If amplifier nonlinearity is an issue, it could be considered to invert the nonlinearity in the controller (i.e., predistortion) [Omet, 2014], in which case the amplifier would behave very linearly.

2. We will model the impact of the RF amplifier's gain and phase variations (which are assumed to be small) by an additive disturbance $\mathbf{f_{g0}}\boldsymbol{d_g}$ as in (6.25). This approximation assumes that gain and phase-shift variations are small, and that the amplifier drive is close to its nominal value.

   For klystrons we have that $\boldsymbol{d_g} = \tilde{g}_{\mathrm{amp}} - i\tilde{\theta}_{\mathrm{amp}} = (5/4 - i\xi_\theta)\, d_{\mathrm{ps}}$, where ξ_θ is the phase-pushing factor and d_{ps} is the variation of power-supply voltage (see Section 5.5).

3. It will be assumed that the controller operates in continuous time. The typical controller frequencies for field controllers are on the order of tens of MHz, while typical closed-loop bandwidths of field controllers are on the order of 100 kHz, making this an acceptable approximation [Åström and Wittenmark, 1997].

4. The discrete digital-downconversion filter will be neglected. Its dynamics could be approximated by increasing the loop delay, see Chapter 12.

We will also lump the time delays into $\tau := \tau_1 + \tau_2 + \tau_{\mathrm{LLRF}}$.

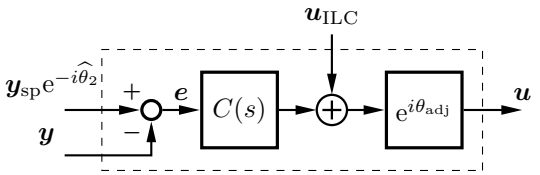


Figure 7.3 Typical structure of a field controller.

Remark 7.1 We could of course have kept the discrete-time controller, modeled the field-control loop by its zero-order hold equivalent, and then analyzed the resulting discrete-time system. This would have been somewhat more correct but the continuous-time formulation simplifies the exposition and most textbooks in process control take this approach.

7.1.3 Linear small-signal model

To study field control performance around a nominal operating point (flat-top operation), it is helpful to consider the relative error $\boldsymbol{z} = \mathbf{a} - 1$ of the accelerating mode, as in Section 6.7.

We will assume that the controller has the form in Figure 7.3, and since we are considering the dynamics around a nominal operating point, we can neglect $\boldsymbol{u}_{\mathrm{ILC}}$. We will also assume that the loop-phase adjustment angle is chosen so that $\theta_{\mathrm{adj}} = \theta_1 + \theta_2$. For the analysis we will do, we may assume that $\theta_2 = 0$ (this corresponds to re-defining the phase at which the inputs and outputs of the controller are considered).

We will use the simplified cavity model in Figure 6.10, and lump the two disturbances into one[1],

$$\boldsymbol{d} = \mathbf{f_{g0}}\boldsymbol{d_g} + \mathbf{i_{b0}}\boldsymbol{d_b}. \tag{7.1}$$

with these simplifications we get the model in Figure 7.4. It may seem that we have made many approximations in this and the previous subsection, but all of them are very reasonable since the involved components (except the amplifier), have linear, well-behaved dynamics. A simple model as the one in Figure 7.4 thus captures the essential aspects of field control around a nominal operating point.

7.2 Types of disturbances

For pulsed linacs it is meaningful to separate the disturbances acting on the field control loop into those that are repetitive from pulse to pulse, and those that are not.

[1] If the disturbance $\boldsymbol{d_g}$ is not stationary, then it should be delayed by τ_1 in (7.1).

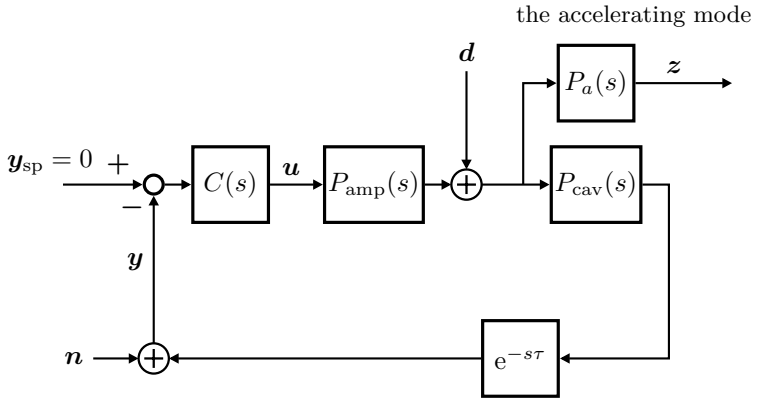


Figure 7.4 Small-signal version of the model in Figure 7.2. The loop phase is assumed to be perfectly calibrated. We have reused $\boldsymbol{y}$ and $\boldsymbol{u}$ for denoting the small-signal variations of the measured and controlled signals.

For continuously operating linacs one may use repetitive control to reject repetitive disturbances of a certain frequency [Inoue et al., 1981; Hara et al., 1988; Wang et al., 2009], but we will not consider this topic further[2].

Repetitive disturbances

Disturbances that are similar from pulse to pulse include

- The nominal beam pulse $\mathbf{i_b}$.
- Variations of $\Delta\omega_a$, from mechanical deformation of the accelerator cavity caused by the Lorentz force[3].

Since these disturbances can be anticipated, it is in principle possible to cancel them completely by adding a suitable signal $\boldsymbol{u}_{\mathrm{ILC}}[k]$, which is synchronized to the start of the beam pulse, to the control signal. A common technique for updating the compensation signal $\boldsymbol{u}_{\mathrm{ILC}}[k]$, based on the control error $\boldsymbol{e}[k]$ in the previous pulse, is iterative learning control (ILC)[4]. We will not consider this topic in this thesis and refer to the ILC literature [Norrlöf, 2000; Bristow et al., 2006; Rogers et al., 2010].

Non-repetitive disturbances

Disturbances which are not repeatable from pulse to pulse include

[2] *Historic remark:* Repetitive control was developed to improve the control performance of a synchrotron magnet power supply [Inoue et al., 1981].

[3] The electromagnetic field acting on the cavity walls.

[4] Frequently referred to as adaptive feedforward in the accelerator literature.

- variations of the beam current ($\boldsymbol{d_b}$),
- variations of the detuning $\Delta\omega_a$ due to microphonics[5],
- spurious noise from the RF amplifier ($\boldsymbol{d_g}$),
- switching ripple from the amplifier's power supply ($\boldsymbol{d_g}$) (unless it is synchronized to the pulse[6]).

In some cases, it is possible to measure these disturbances and reduce their impact through feedforward. However, this is often challenging due to the fast time scales, and the quality of typically available measurements. Except for feedforward, there is no way to avoid that these disturbances give rise to field errors. The only way to reduce the errors is by feedback.

Remark 7.2 With feedforward from beam-current monitors (BCMs) it could be possible to reduce the impact of non-repetitive beam-current variations. The effectiveness depends on the time-delay of the feedforward signal relative to the beam. Even if the feedforward correction arrives later than the disturbance it may still improve the performance[7]. The scheme is more effective for ion linacs where the initial beam velocity is relatively low. Analog feedforward of beam current variations was discussed in [Jameson and Wallace, 1971]. See also [Butkowski et al., 2018].

Example: Beam current variations at ESS

The ESS will have a pulsed microwave discharge ion source [Neri et al., 2014; Celona et al., 2016] that generates free protons (H^+) by ionizing hydrogen gas (H_2) to a plasma. From the plasma, which is confined by magnetic fields, a proton current of up to 70 mA is extracted by a 75 keV potential.

From a field control perspective, the most important characteristic of an ion source is the beam current variations. For overall accelerator performance it is typically more important with a low beam emittance.

A normalized beam pulse from the ESS ion source together with its repetitive and non-deterministic components are shown in Figure 7.5. The repetitive component can be canceled by iterative learning control while feedback is necessary to cancel the non-deterministic component. An estimate of the power spectrum of the non-deterministic components of the ESS beam pulse is shown in Figure 7.6.

[5] Unwanted mechanical vibrations, for example, from nearby traffic or the cryo plant.

[6] Good rejection of repeatable switching ripple has been demonstrated in practice by synchronizing the switching pattern to the RF pulse. [Personal correspondence with Thomas Schilcher, Paul Scherrer Institute].

[7] Investigations by B. Bernhardsson for the ESS linac, based on the beam spectrum of the LINAC2 ion source, indicated that with a delay of 0.5 µs, the errors could be reduced by 40 %, under ideal conditions. A delay of 1 µs gave an error reduction by 25 %.

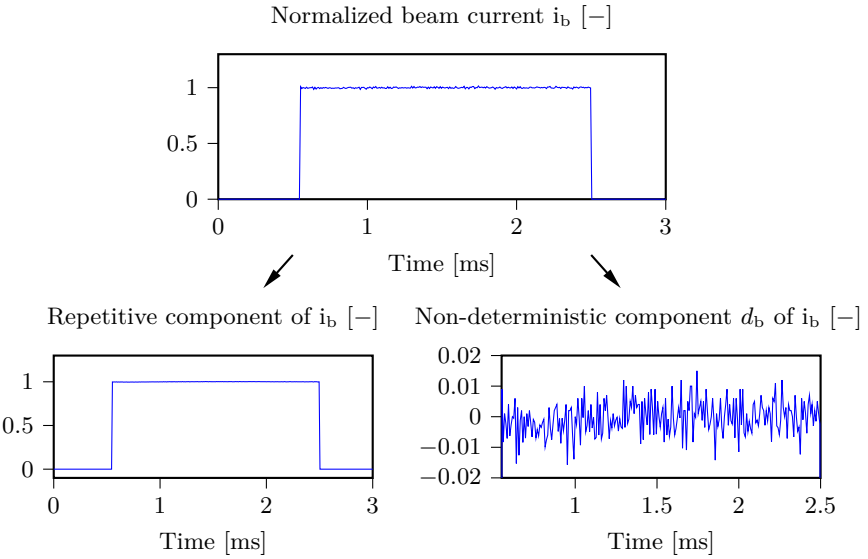


Figure 7.5 *Top:* Normalized beam pulse from the ESS ion source (magnitude 62.5 mA before normalization). Note that the beam pulse is shorter than the nominal length of 2.86 ms. *Bottom row:* The beam pulse can be divided into a repetitive component and a non-deterministic component. *Additional info:* The data (49 pulses) was collected on 2019-07-02 by Ryoichi Miyamoto, ESS. The measurement is from downstream the LEBT chopper and has been truncated according to the 10 ns rise time of the MEBT chopper (which is not installed). The ripple characteristics depend on several tuning parameters which have not yet been decided.

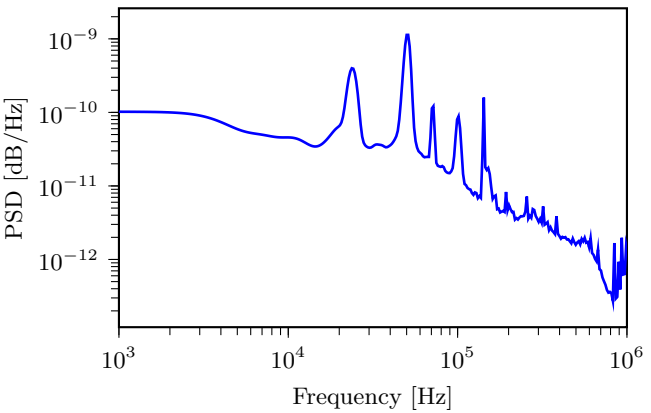


Figure 7.6 Estimated power spectral density of the non-deterministic components in the beam pulse from of the ESS ion source.

7.3 Parameter variations

The disturbances in the previous section enter additively in the linearized model in Figure 7.2, but from the original nonlinear model in Figure 7.1, we see that some of them correspond to parameter variations in the linearized model. In this section we briefly mention some of these parameter variations. Neither of them tends to be severe enough to warrant the use of an adaptive controller, but it is prudent to give them some attention when considering controller robustness.

Droop. In pulsed linacs the supply voltage to the RF amplifiers commonly decreases (droops) during the pulse. As discussed in Section 5.5.3, this changes the gain and phase-shift of the RF amplifier through gain and phase pushing. Assuming a phase-pushing factor of $\xi_\theta = 10\,°/\%$ then $1\,\%$ droop gives a phase shift of $\tilde{\theta}_{\text{amp}} = 10°$. As will be discussed in Section 8.4, this may give a phase-margin reduction of up to 10° which is significant.

Temperature variations. Temperature variations change the electrical length of waveguides and RF cables, which correspond to variations of the phase shifts θ_1 and θ_2 in Figure 7.2. Variations of up to 10° per day (correlated with the building temperature) were reported in [Brandt, 2007]. The temperature variations are slow, and the variations in θ_1 and θ_2 can be handled by re-calibrating the controller's phase-adjustment angle θ_{adj} (see Figure 7.3). Loop-phase adjustment will be further discussed in Section 8.4.

Detuning variations. The detuning $\Delta\omega_a$ may vary due to Lorenz force detuning and microphonics. However, the variations are small (hundreds of hertz) compared to the typical bandwidths of field-control loops (tens of kilohertz), the variation in the dynamics should not be significant. The interaction between detuning variations and mechanical resonance modes of the cavity may (for very narrowband cavities) give rise to so-called ponderomotive instabilities [Delayen, 2006], however this topic is outside the scope of this thesis.

Variations of parasitic-mode frequencies. The resonance frequencies of the parasitic modes $\Delta\omega_1, \ldots, \Delta\omega_N$ vary for the same reasons as the detuning $\Delta\omega_a$. Chapter 11 will discuss how notch filters can be used for avoiding feedback instability from parasitic modes. It is important that these notch filters are sufficiently wide to handle variations of the $\Delta\omega_k$'s.

8

Essentials of Cavity Field Control

To explain the essential principles of cavity field control in the clearest way possible we will in this chapter assume that: the detuning is zero ($\Delta\omega = 0$), the loop phase is perfectly adjusted, there are no parasitic modes, and the RF amplifier is ideal; these assumptions give the block diagram in Figure 8.1.

8.1 A water-tank analogy

The process in Figure 8.1 can be recognized as a first-order system with a time-delay. The dynamics are the same as for a water tank linearized around a nominal operating point, together with a measurement delay. A difference is that the signals in Figure 8.1 are complex valued, but for the discussions in this chapter (except Section 8.4), this is of no importance.

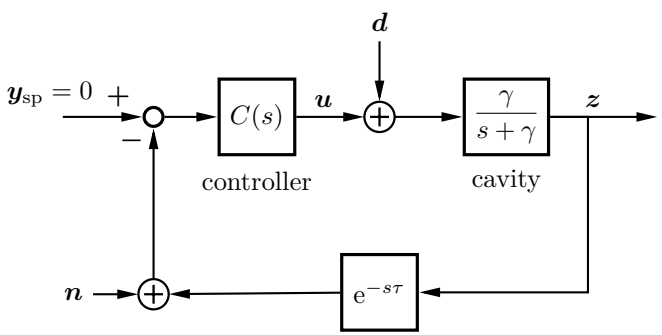


Figure 8.1 A simplified model of a field-control loop that captures many of the important control aspects.

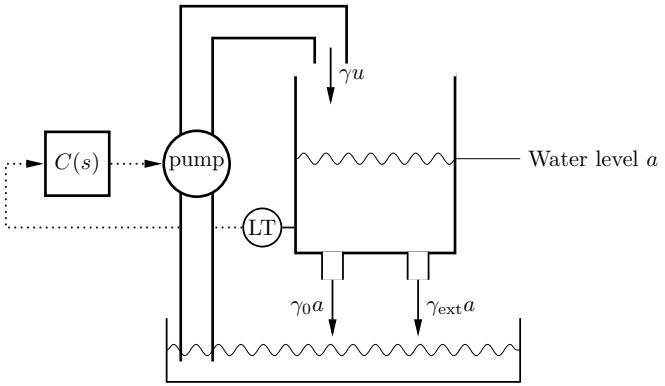


Figure 8.2 From a control perspective, the cavity field control loop can be thought of as a water tank (with a complex-valued water level). The water level corresponds to the amplitude of the electromagnetic field. (LT = level transmitter)

A water tank with a time-delay is a common first lab process in undergraduate control education [Åström and Östberg, 1986], see Figure 8.2. Practically all students are able to achieve acceptable performance with a PI controller that they tune by trial and error [Tank Lab]. This indicates that it should be possible to achieve reasonable field control performance by rather simple means.

Process models consisting of a first-order system together with a time-delay are studied extensively in the process control literature, where they are referred to as FOTD models. A good book on the control of such systems is [Åström and Hägglund, 2006].

Differences

The field control problem has two very important features which are not captured by the water-tank analogy, namely

- the need for loop-phase adjustment and
- parasitic modes.

Either one of these features is likely to cause instability if not properly mitigated. However, if the loop-phase is properly adjusted (Section 8.4) and notch filters are introduced at the frequencies of problematic parasitic modes (Chapter 11) then the resulting system is well described by a water tank with a delay.

Some other, less crucial, differences are that the disturbances and the objective function have certain directionality in the complex plane (Chapter 15.2), the measurement process, i.e., the digital downconversion is a bit

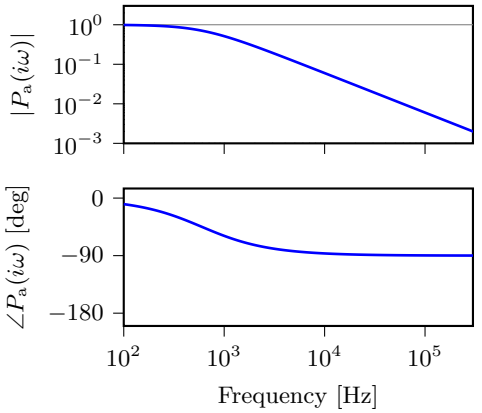


Figure 8.3 Normalized transfer function for the accelerating mode of a superconducting cavity with 500 Hz bandwidth.

more involved than measuring the water level (Chapter 12), and the detuning makes the dynamics for positive and negative frequencies slightly different.

While the water-tank analogy captures many of the essential *control aspects* of the field-control problem, it should be emphasized that the physical behavior of the water tank and cavity system differ significantly[1].

8.2 Basic aspects of cavity field control

8.2.1 The need for feedback

Without feedback, the impact of relative amplifier disturbances $\boldsymbol{d_g}$ and beam variations $\boldsymbol{d_b}$ on the accelerating cavity mode is given by the signal $\boldsymbol{d} = \mathbf{f_{g0}}\boldsymbol{d_g} + \mathbf{i_{b0}}\boldsymbol{d_b}$ in (7.1) filtered through $P_a(s)$. The frequency response of $P_a(s)$ is shown in Figure 8.3. It is seen that the cavity acts as a low-pass filter; only high-frequency disturbances are attenuated. With a feedback controller, it is possible to reduce the impact of low-frequency disturbances.

8.2.2 Closed-loop transfer functions

The change in disturbance rejection due to a feedback controller C, is, as discussed in Section 2.2, quantified by the so-called sensitivity function

$$S = 1/(1 + PC). \tag{8.1}$$

The examples of transfer functions given in this chapter correspond to a superconducting cavity with 500 Hz bandwidth controlled by a PI controller.

[1] The physics of electrical *DC* circuits can be described by hydraulic models but the cavity model (6.8) is a baseband model of an RF system.

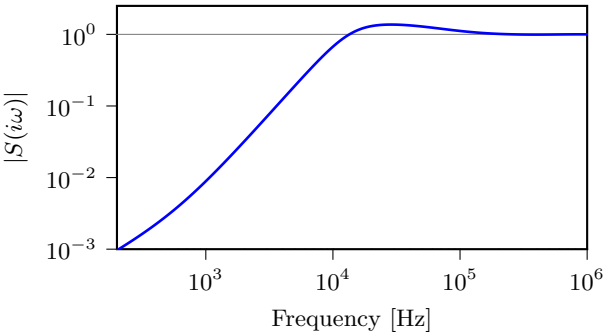


Figure 8.4 Sensitivity function $S = 1/(1 + PC)$ for a superconducting cavity controlled with a PI controller.

The PI controller has high gain at low frequencies, which reduces the sensitivity to low-frequency disturbance, see Figure 8.4. While the figure shows an improved disturbance rejection for low frequencies, it also shows that disturbances around 10–100 kHz are amplified. This is unavoidable and known as the waterbed effect (recall Section 2.2).

8.2.3 Impact of load disturbances on field errors

With feedback, the transfer function from disturbances to field errors becomes

$$G_{zd} = P_a/(1 + PC) = P_a S, \tag{8.2}$$

A typical transfer function $G_{zd}(s)$, together with the open-loop cavity dynamics $P_a(s)$, are shown in Figure 8.5. Recall that, similarly to (6.27), we have that the closed-loop transfer functions from relative amplifier variations and relative beam variations, to field errors, are given by $\mathbf{f_{g0}} G_{zd}$ and $\mathbf{i_{b0}} G_{zd}$.

8.2.4 Impact of measurement noise on field errors

The transfer function

$$G_{zn} = P_a C/(1 + PC) \tag{8.3}$$

quantifies the impact of measurement noise on the controlled signal z and the response to setpoint changes. A typical transfer function G_{zn} is shown in Figure 8.6.

For pulsed linacs, the step response characteristics can typically be improved to acceptable levels using iterative learning control, so this is not a crucial aspect in the design of $C(s)$.

Typical field-control loops have bandwidths of about 30–100 kHz, so the impact of measurement noise on field errors is, unlike the impact of load

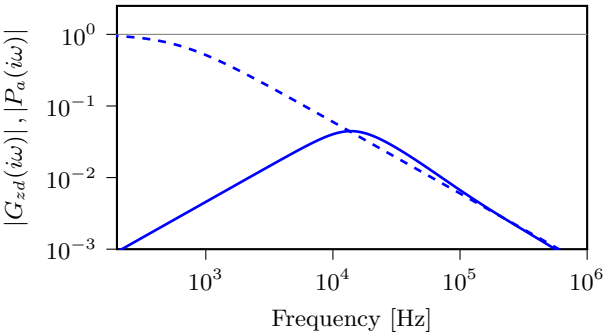


Figure 8.5 Transfer functions from disturbances d to field errors z, with feedback ($G_{zd} = P_a(1+PC)$, solid line), and without feedback (P_a, dashed line).

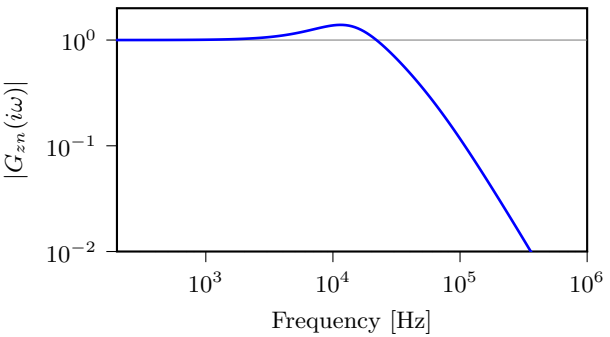


Figure 8.6 The transfer function $G_{zn} = P_aC/(1+PC)$ from measurement noise n to field errors z.

disturbances, relatively similar for different cavities and linacs. For FEL linacs, which require field errors below 0.01 %, the impact of measurement noise through $G_{zn}(s)$ is a concern. Nevertheless, field errors below 0.01 % are achieved for these linacs. Hence, for field-error requirements of 0.1–1 % the direct impact of measurement noise through $G_{zn}(s)$ should be quite negligible (assuming a decent LLRF system).

8.2.5 Impact of measurement noise on control signal activity

The transfer function

$$G_{un} = C/(1+PC) = CS \tag{8.4}$$

quantifies how measurement noise n is amplified to the control signal u.

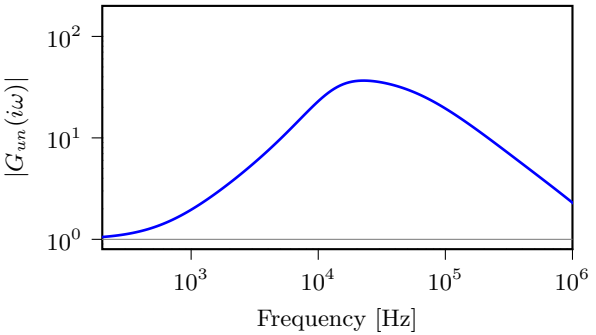


Figure 8.7 The transfer function $G_{un} = C/(1 + PC)$ from measurement noise n to the control signal u.

High control signal activity tends to give poor control performance due to excitation of unmodeled dynamics and an increased impact of amplifier nonlinearities. Another good reason to limit the control signal activity is to avoid excessive transversal kick to the beam [Hellert and Dohlus, 2018]. This effect cannot be seen or understood from the models that we have presented.

The effects of excessive control signal activity are hard to quantify, but it is prudent to put conservative limits on control signal activity during preliminary design and analysis. A high controller gain is typically used for superconducting cavities, so a low-pass filter in the controller is needed to limit the amplification of high-frequency measurement noise to the control signal.

The transfer functions (8.1)–(8.4) are collectively known as the "gang of four" [Åström and Murray, 2010] and capture the essentials of the closed-loop control performance. Typical gang-of-fours for a superconducting cavity and a normal conducting cavity are shown in Figure 8.8.

8.2.6 Stability

It is not sufficient to look at the magnitude of the closed-loop transfer function, but as discussed in Section 2.2, it is also important to consider stability when feedback is involved. For the applications in this thesis, closed-loop stability is conveniently verified by the Nyquist criterion (Section 2.2). The Nyquist and Bode diagrams for the normal conducting and superconducting cavities from our previous examples are shown in Figure 8.9.

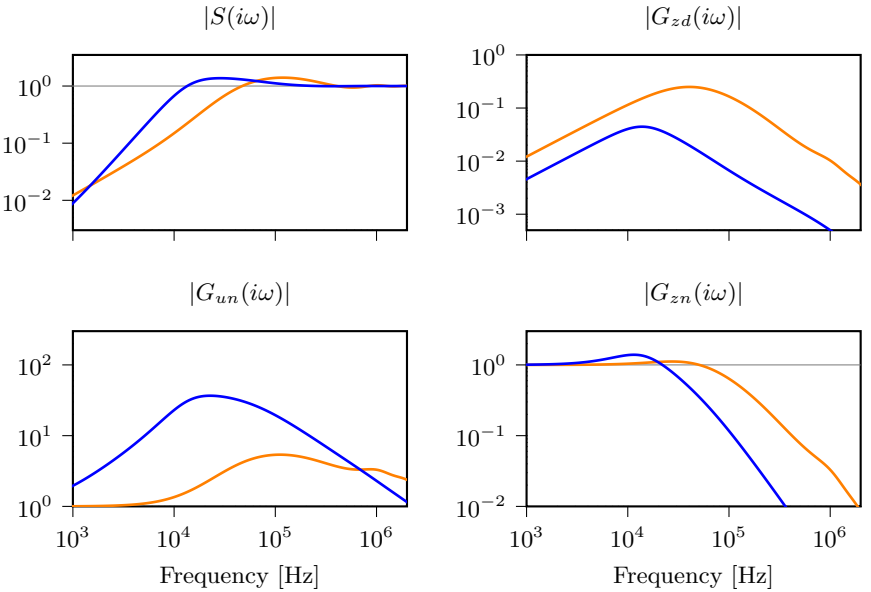


Figure 8.8 The transfer functions known as the "gang of four" for a normal conducting cavity (12 kHz bandwidth, orange lines) and a superconducting cavity with (0.5 kHz bandwidth, blue lines). Both cavities are controlled with PI controllers tuned to have a maximum sensitivity $M_S \leq 1.6$.

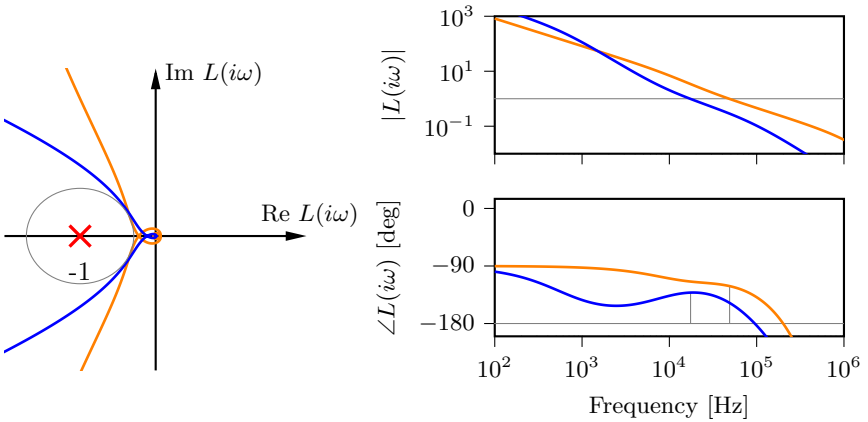


Figure 8.9 Nyquist and Bode diagrams for a normal conducting and a superconducting cavity controlled by a PI controller. The gray circle in the Nyquist diagram corresponds to a sensitivity constraint of $M_S \leq 1.6$, and the gray lines in the phase curve of the Bode diagram illustrate phase margins of about 45°; both measures indicate good closed-loop robustness.

8.3 Disturbance sensitivity vs. external coupling

The block diagram in Figure 8.1 suggests that the bandwidth $\gamma = \gamma_0 + \gamma_{\text{ext}}$ determines a field-control loop's relative disturbance sensitivity and one is led to believe that by decreasing γ_{ext} (and thus γ) the disturbance sensitivity could be made arbitrarily small (if $\gamma_0 = 0$). The analysis in [Ludwig et al., 2006] points in the same direction. However, this conclusion only valid if the beam loading is negligible, an assumption that rarely holds.

In this section we clarify how the disturbance sensitivity depends on γ_0, γ_{ext}, and γ_{beam} (γ_{beam} is the decay rate due to beam loading, see (6.17)).

To simplify the discussion, we will assume that $\Delta\omega = \phi_b = 0$, in which case $\gamma \mathbf{i_{b0}} = -\gamma_{\text{beam}}$ and $\gamma \mathbf{f_{g0}} = \gamma_0 + \gamma_{\text{ext}} + \gamma_{\text{beam}}$. This lets us re-write (6.27) as

$$P_{d_g \to z}(s) = \frac{\gamma_0 + \gamma_{\text{ext}} + \gamma_{\text{beam}}}{s + \gamma_0 + \gamma_{\text{ext}}}, \tag{8.5a}$$

$$P_{d_b \to z}(s) = \frac{\gamma_{\text{beam}}}{s + \gamma_0 + \gamma_{\text{ext}}}. \tag{8.5b}$$

From these expressions we observe the following effects of decreasing γ_{ext}.

- Decreasing γ_{ext} reduces the impact of high-frequency RF-drive variations. However, the impact cannot be made smaller than

$$\frac{\gamma_0 + \gamma_{\text{beam}}}{s + \gamma_0}.$$

  Relative to an optimally tuned cavity ($\gamma_{\text{ext}} = \gamma_0 + \gamma_{\text{beam}}$), this corresponds to a reduction by factor smaller than 2. At low frequencies, there are competing effects from the γ_{ext} in the numerator and the denominator of (8.5a).

- Decreasing γ_{ext} hardly affects the impact of high-frequency beam variations. The impact of low-frequency beam variations is increased.

Increasing γ_{ext} gives the opposite effects. In this analysis we did not account for the effect of feedback. This effect is captured by the sensitivity function $S(s) = 1/(1 + C(s)P(s))$ which was introduced earlier in this chapter. Increasing γ_{ext} allows a slightly faster controller and slightly better rejection of low-frequency disturbances.

It should be remembered that deviations of γ_{ext} from $\gamma^\star_{\text{ext}} = \gamma_0 + \gamma_{\text{beam}}$ increase the steady-state power consumption (Section 6.6). Hence there is limited freedom in selecting γ_{ext} in practice. However, small changes of γ_{ext} from $\gamma^\star_{\text{ext}}$ are tolerable. For example, $0.82\gamma^\star_{\text{ext}} < \gamma_{\text{ext}} < 1.22\gamma^\star_{\text{ext}}$ gives less than one percent increase in power consumption.

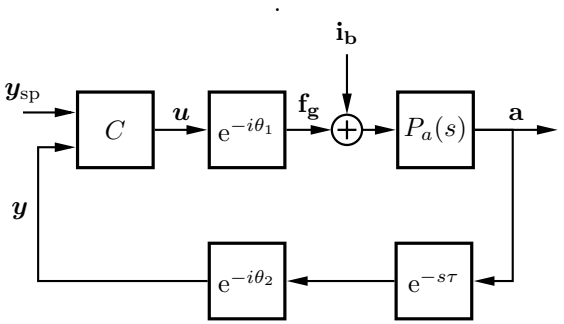


Figure 8.10 Linear, large-signal model of a field control loop, which is a simplified version of the model in Figure 7.2. This model will be used to illustrate the importance of correct loop phase adjustment. In this section we will consider a linear controller C, but a nonlinear controller will be considered in Chapter 10. The transfer function for the accelerating mode is given by $P_a(s) = \gamma/(s + \gamma - i\Delta\omega)$.

8.4 Loop-phase adjustment

We will now discuss the importance of the loop-phase adjustment angle $\theta_{\rm adj}$. Consider the model in Figure 8.10 which is a simplification (only considers the accelerating mode) of the model in Figure 7.2. The transfer function of the plant is given by

$$P(s) := \mathrm{e}^{-i\theta} P_0(s), \tag{8.6}$$

where $P_0(s) := P_a(s)\mathrm{e}^{-s\tau}$ is the nominal transfer and $\theta := \theta_1 + \theta_2$.

We will assume that the field controller has the structure in Figure 7.3 and thus has the transfer function $\mathrm{e}^{i\theta_{\rm adj}} C(s)$. The open-loop transfer function then takes the form

$$L(s) = \mathrm{e}^{i\theta_{\rm adj}} C(s) \mathrm{e}^{-i\theta} P_0(s) = \mathrm{e}^{i\delta} L_0(s) \tag{8.7}$$

where $L_0(s) = C(s)P_0(s)$, and $\delta = \theta_{\rm adj} - \theta$. It follows that the Nyquist curve of $L(s)$ equals the Nyquist curve of $L_0(s)$ rotated by δ radians around the origin, see Figure 8.11.

If $L_0(s)$ has real coefficients (i.e., detuning and parasitic modes are neglected) then it is clear that choosing $\delta = 0$ gives the best robustness (Theorem 3 on p. 52). It is also clear that a non-zero value of δ gives a phase-margin reduction by δ radians (if the nominal feedback loop is stable).

To illustrate the importance of δ, note that the transfer function from the control error $\boldsymbol{e} = -\boldsymbol{z}$ to the corrective control action is given by

$$\mathrm{e}^{-s\tau} C_0(s) \mathrm{e}^{i\delta}.$$

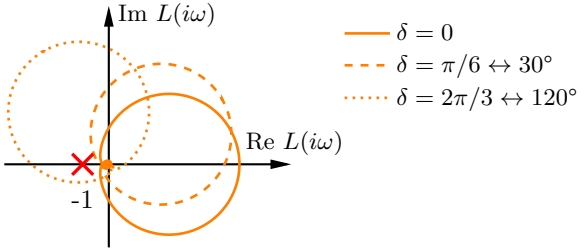


Figure 8.11 Nyquist curves of a normal conducting cavity controlled with a proportional controller. It is seen that a loop-phase misadjustment δ gives a corresponding phase-margin reduction which may cause instability.

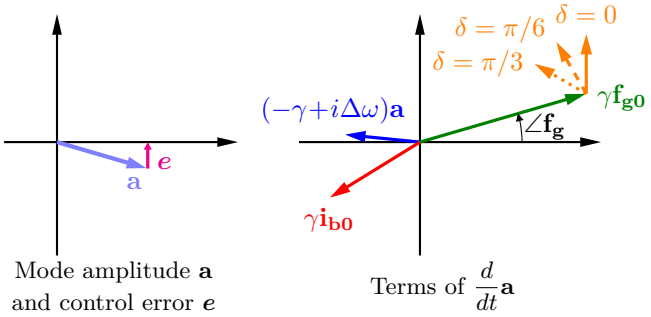


Figure 8.12 Phasor diagram illustrating the effect of a loop-phase misadjustment δ. The control action should ideally be applied in the direction of the control error $\boldsymbol{e} = -\boldsymbol{z}$ which corresponds to the solid orange arrow. With a loop-phase misadjustment $\delta \neq 0$, the corrective action is applied in a less helpful direction. The extreme case of $\delta = \pi$ corresponds to positive feedback which is likely to cause instability.

From this we see that the direction in which the control action is applied depends on δ. A poorly chosen value of δ may give amplified control errors and instability, see Figure 8.12.

Remark 8.1 In the field control literature the phase shift $(\theta_1 + \theta_2)$ is typically referred to as the *loop phase* [Brandt, 2007]. This terminology is somewhat unfortunate, since in a control context, the term loop phase refers to the quantity $\angle L(i\omega)$ as a function of the frequency ω. Other texts refer to $\theta_1 + \theta_2$ as the *fractional time delay*. In this thesis we will avoid referring explicitly to $\theta_1 + \theta_2$, but will refer the process of choosing a suitable value of θ_{adj} as *loop-phase adjustment*.

Remark 8.2 Loop-phase adjustment is also important for Cartesian feedback linearization of RF amplifiers, as was discussed in Section 4.6.

Calibration of the loop-phase adjustment angle

To calibrate the loop-phase adjustment angle θ_{adj}, the most straight-forward approach would be to do systems identification of $L(s)$ and then adjust θ_{adj} for optimal robustness. Realizing that adequate phase margins are most critical around the open-loop system's cross-over frequency, the identification should focus on this frequency range. Automation of this calibration task was discussed in [Brandt, 2007]. Causes for variations in the parameters θ_1 and θ_2 were discussed in Section 7.3.

Remark 8.3 (Mathematics-free loop-phase adjustment) The following calibration approach[2] has the advantage of being very simple:

1. find a value of θ_{adj} such that the field control loop is stable and denote this value by θ_{init};
2. increase θ_{adj} from θ_{init} until the feedback loop starts to oscillate and denote the corresponding value of θ_{adj} by θ_{max};
3. decrease θ_{adj} from θ_{init} until the feedback loop starts to oscillate and denote the corresponding value of θ_{adj} by θ_{min}; and finally
4. select $\theta_{\text{adj}} = (\theta_{\text{max}} + \theta_{\text{min}})/2$.

Remark 8.4 It might be tempting to estimate the loop phase by simply taking the argument of the steady-state response $P(0)$. However, this is not a good idea since this estimate would also include the "detuning angle" $\tan^{-1}(\Delta\omega/\gamma)$, since $P(0) = \mathrm{e}^{-i\theta}\gamma/(\gamma - i\Delta\omega)$. It is not at low-frequencies that the phases at positive and negative frequencies should be balanced, but rather for frequencies around the open-loop cross-over frequency ω_{co}. Namely, one would like to have $\angle L(i\omega_{co}) = -\angle L(-i\omega_{co})$. A better approach to loop-phase calibration could be to excite the process with a real-valued (baseband) signal $2\cos\omega_{co}t = \mathrm{e}^{i\omega_{co}t} + \mathrm{e}^{-i\omega_{co}t}$ and then compare the frequency shifts of the positive and negative frequency components.

[2] Oral tradition of the field control community. The author learned about it from Mark Crofford, SNS.

9

Limits of Field Control Performance

9.1 Introduction

An important aspect of control design is to decide on the controller structure—it should be flexible enough to allow good control performance but also as simple as possible for ease of understanding and implementation. For field control it has been proposed to use: PI(D) control, linear–quadratic–Gaussian (LQG) control [Baudrenghien et al., 2014], and Smith predictor-based control [Simrock, 2005]. All these controllers types are linear and time invariant (LTI). Regardless of the controller structure there are *fundamental* limitations on the achievable control performance. One such fundamental limitation is given by Bode's integral formula (2.7).

For LTI systems it is often possible to compute the best control performance achievable by *any* LTI controller [Boyd and Barratt, 1991]. In this chapter we will do this for the field control problem and compare the performance of PI(D) controllers and the optimal LTI controller. The work in this chapter was inspired by [Garpinger, 2009], where similar[1] investigations were made for low-order process models.

We will compare the performance of three classes of controllers:

- all linear time-invariant controllers C(s);
- PI controllers of the form

$$C_{\mathrm{PI}}(s) = K\left(1 + \frac{1}{sT_i}\right) \cdot \frac{1}{sT_f + 1}, \tag{9.1}$$

  where T_f is the time-constant of the low-pass filter; and

[1] The performance objective considered in [Garpinger, 2009] was the integrated absolute value (L_1 norm) of the control error and the load disturbance was assumed to be a step. In this chapter we will minimize the rms control error and consider a slightly higher-frequency disturbance spectrum.

- PID controllers of the form[2]

$$C_{\mathrm{PID}}(s) = K\left(1 + \frac{1}{sT_i} + \frac{sT_d}{sT_d/N + 1}\right) \cdot \frac{1}{sT_f + 1}, \tag{9.2}$$

  where T_d is the derivative time and N is a constant that limits the derivative action at high frequencies.

To make meaningful comparisons, the field control requirements from Section 8.2 are formulated mathematically in Section 9.2. In Section 9.3 we provide the disturbance and plant model for the comparison. In Section 9.4 we compare the performance of the optimal LTI controller to the performance of optimal PI and PID controllers for different loop delays and different levels of control signal activity. In Appendix E.1 we give some details on how the optimal linear controllers were designed and in Appendix E.2 we briefly discuss how the PI and PID controllers were tuned.

9.2 Mathematical formulation of control specifications.

In Section 8.2 we discussed that a good field controller should keep the field errors small in the presence of load disturbances, avoid excessive control signal activity, and be sufficiently robust. Next, we formulate these requirements mathematically using the signal and system norms from Section 2.3.

Cavity field errors. Denoting the spectrum of the normalized load disturbances $\boldsymbol{d}$ by $D(s)$, the resulting normalized error (rms) of the accelerating mode's amplitude is given by

$$\text{normalized field errors (rms)} = \left|\left|G_{zd}D\right|\right|_2 .$$

The impact of measurement noise on field errors has not been included since it should only be a concern for reaching the field errors necessary for FEL linacs (see Section 8.2.4). Leaving this term out makes it possible to scale the results to different levels of noise and disturbances.

Control signal activity. The control signal activity resulting from measurement noise with spectrum $N(s)$ is given by

$$\text{control signal activity (rms)} = \left|\left|G_{un}N\right|\right|_2 .$$

This control signal activity should be compared to the power overhead of the RF amplifier. In the preliminary field control analysis for LCLS-II, the control signal activity was constrained to 4 % rms [Doolittle et al., 2016].

[2] Another PID-controller parametrization [Åström and Hägglund, 2006] which only use four parameters is $C_{\mathrm{PID}}(s) = K(1 + 1/(sT_i) + sT_d)/(1 + sT_f + (sT_f)^2/2)$. However, this parameterization can sometimes give *slightly* worse than a PI controller [Garpinger, 2009]. This may seem confusing and therefore we consider the form (9.2).

Robustness. The robustness of the feedback loop is conveniently quantified by the maximum value of the sensitivity function (8.1),

$$M_S = \sup_{\omega} |S(i\omega)| = ||S||_\infty .$$

This robustness measure is, as we discussed in Section 2.2, more general and convenient than the traditionally used gain and phase margins. A typical upper bound on M_S is between 1.4 and 2 with better robustness for smaller values.

Mathematical formulation of the field control problem. We are now in a position to formulate the field control problem mathematically.

Control Problem
From a given class $\mathcal{C}$ of LTI controllers, find the controller C that solves the optimization problem

$$\begin{aligned} \underset{C \in \mathcal{C}}{\text{minimize}} \quad & \left|\left| \frac{P_a}{1+PC} D \right|\right|_2 && \text{(9.3a)} \\ \text{subject to} \quad & \left|\left| \frac{1}{1+PC} \right|\right|_\infty \leq S_{\max} && \text{(9.3b)} \\ & \left|\left| \frac{C}{1+PC} N \right|\right|_2 \leq B_{un}. && \text{(9.3c)} \end{aligned}$$

where D is the spectrum of load disturbances, N is the spectrum of measurement noise, $S_{\max}$ is the limit on maximum sensitivity and B_{un} is the limit on control signal activity due to measurement noise.

9.3 Problem formulation

Below we introduce the assumed process models $P(s)$ and $P_a(s)$, the disturbance spectrum $D(s)$, and the measurement noise spectrum $N(s)$.

Plant dynamics. We will consider the plant model in Figure 8.1 which has the transfer function

$$P(s) = \frac{\gamma_a}{s+\gamma_a} P_{\text{amp}}(s) \mathrm{e}^{-s\tau}.$$

For this model it was assumed that the detuning is zero to get a conjugate symmetric transfer function. The assumption that the detuning is zero has little impact on field control performance since the closed-loop bandwidth is much larger than typical values of detuning.

Table 9.1 Parameters used in the numerical experiments of Section 9.4.

Parameter	**Unit**	**Value**	
Amplifier bandwidth, $\omega_{\text{amp}}/(2\pi)$	MHz	1.5	
Loop-delay, τ	µs	1.0	
		n.c. cavity	**s.c. cavity**
Bandwidth of the accel. mode, $\gamma_a/(2\pi)$	kHz	12	0.5

Two bandwidths γ_a of the accelerating mode will be considered: 12 kHz which corresponds to the (normal conducting) DTLs at ESS and 0.5 kHz which is typical for the superconducting cavities at ESS.

We also assume that $\theta_{\text{adj}} = \theta = 0$. The amplifier dynamics $P_{\text{amp}}(s)$ is given by (5.7). The parameter values that will be considered in the next section are given in Table 9.1. The loop delay of 1 µs is typical for field control loops and matches the expected loop delay at ESS (see Table 5.2).

Disturbance spectrum It will be assumed that the disturbances are rotationally invariant and have a power spectral density

$$\Phi(\omega) = |D(i\omega)|^2 \quad \text{with} \quad D(s) = 30 \cdot \frac{(s+a)}{s(s+b)}, \tag{9.4}$$

where $a = 2\pi \cdot 3 \times 10^3$, and $b = 2\pi \cdot 3 \times 10^4$, see Figure 9.1. The large magnitude of $|D(i\omega)|$ at low frequencies corresponds to amplifier droop and pulse-to-pulse variations of the beam current, while the higher-frequency spectrum corresponds to beam current ripple and switching ripple from the RF amplifier's power supply.

Measurement noise spectrum Since we only consider how the measurement noise drives control signal activity we only model the broadband white noise that mostly originates from the ADC. We will assume that the white measurement noise is normalized to unity with respect to the Nyquist frequency of the controller. This corresponds to

$$N(i\omega) = 1/\sqrt{f_{\text{reg}}}$$

where f_{reg} is a controller frequency. We will assume that $f_{\text{reg}} = 100$ ns which is typical for current-generation LLRF systems. The normalization makes it possible to compare the results for different noise levels and different constraints on control signal activity. The assumption of white noise also enables a numerically better problem formulation to design optimal LTI controller.

White noise was also assumed in the field control analysis in [Serrano et al., 2017]. For an example of receiver noise in an LLRF system see [Ludwig et al., 2019, Fig. 1].

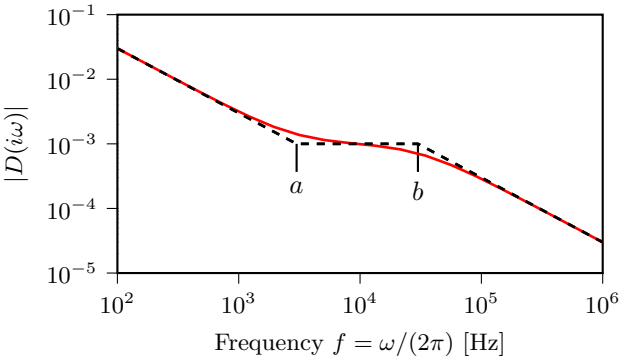


Figure 9.1 Assumed load disturbance spectrum for the numerical examples considered in this section. Since we will consider the performance of different control approaches relative to each other, the absolute scale of the disturbance spectrum is not important.

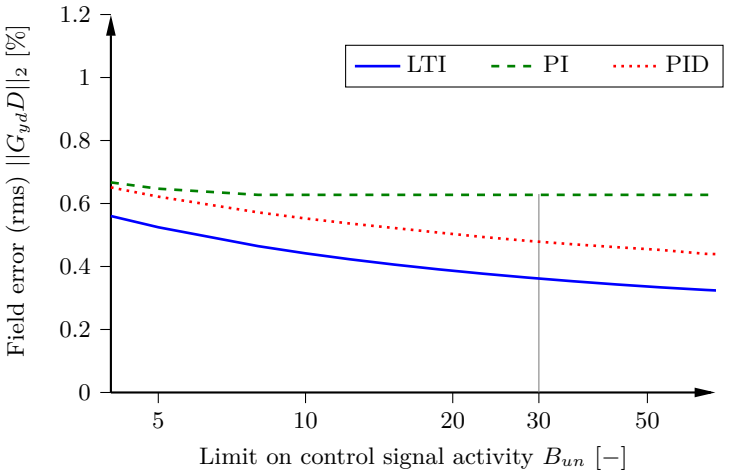
9.4 Results

We consider the following nominal case: time delay $L = 1\,\mu s$, maximum sensitivity $S_{\max} = 1.6$, and maximum control signal activity $B_{un} = 30$. For this nominal a case, a comparison of the open-loop frequency responses and the gang-of-fours for the considered controllers are shown in Figure 9.5 and Figure 9.4. How the cavity field errors depend on the allowed control signal activity B_{un} is shown in Figure 9.2 and how cavity field errors depend on the time delay τ is shown in Figure 9.3.

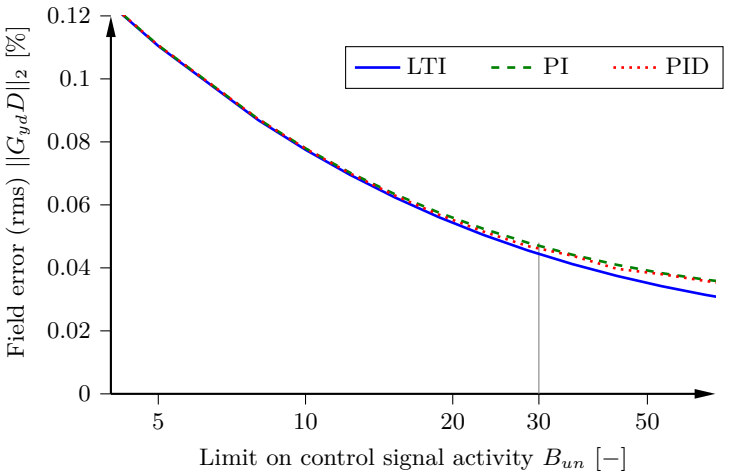
Remark 9.1 The sensitivity constraint of $M_S \leq S_{\max} = 1.6$ is motivated by that $M_S \leq 1.4$ is commonly used in process control [Åström and Hägglund, 2006]. In process control, the large number of controllers makes it infeasible to obtain accurate process models and robustness is often more important than pushing the performance to its limits.

The typical field control loop has linear dynamics (apart from the amplifier) and for a billion-euro accelerator it is economically justifiable to spend more time on system identification. This makes it reasonable to allow slightly larger values of $S_{\max}$ which enables better control performance.

The nominal constraint $B_{un} = 30$ corresponds to that $-60\,\text{dB}$ measurement noise gives to control signal variations of less than $3\,\%$ rms.

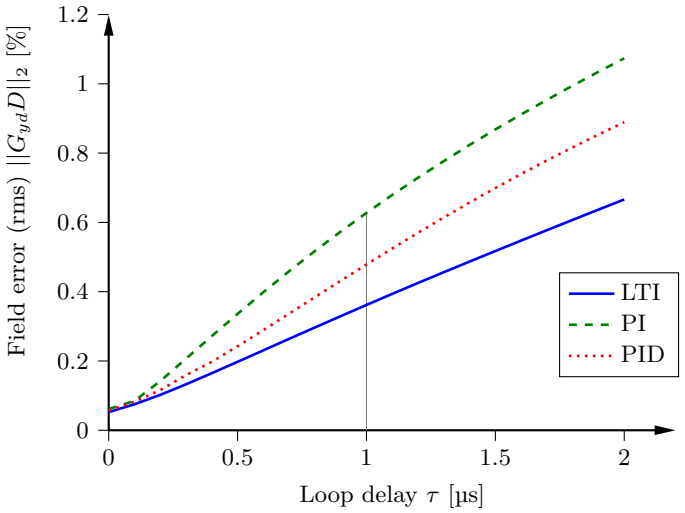


(a) Normal conducting cavity.

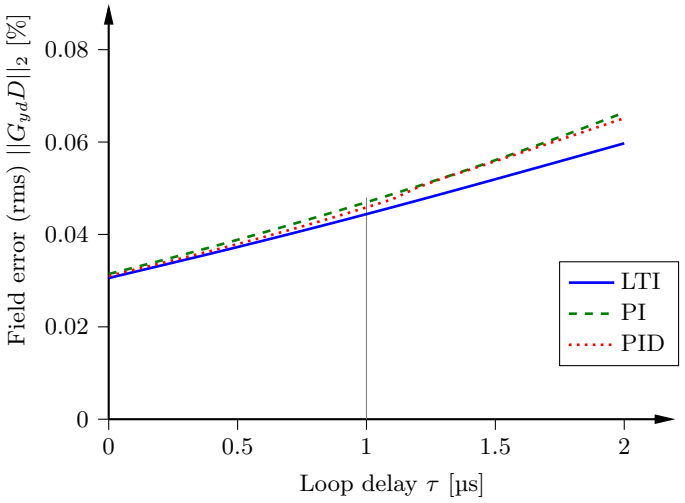


(b) Superconducting cavity.

Figure 9.2 Optimal control performance vs. allowed control signal activity B_{un}, for different controller types. For a small value of B_{un} it is not possible for the general LTI controller to use its flexibility to suitably shape the open-loop dynamics, since this requires a large control signal. With greater values of B_{un}, the performance advantage of the general LTI controller over the PI and PID controllers increases. For the superconducting cavity there is little advantage of using anything other than a PI controller.

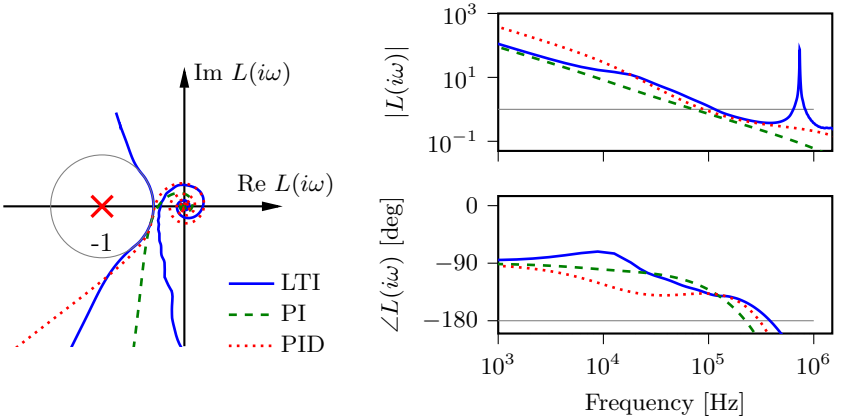


(a) Normal conducting cavity.

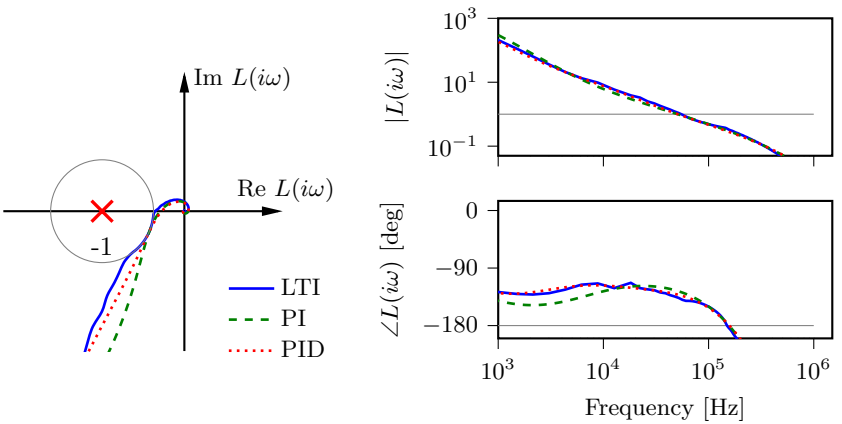


(b) Superconducting cavity.

Figure 9.3 Field errors vs. loop delay τ (with $S_{\max} = 1.6$ and $B_{un} = 30$).n It is seen that the achievable control performance, both for the normal conducting and the superconducting cavity depends strongly on the time-delay τ. This is most noticeable for the normal conducting cavity where the control performance is almost proportional to the time delay.

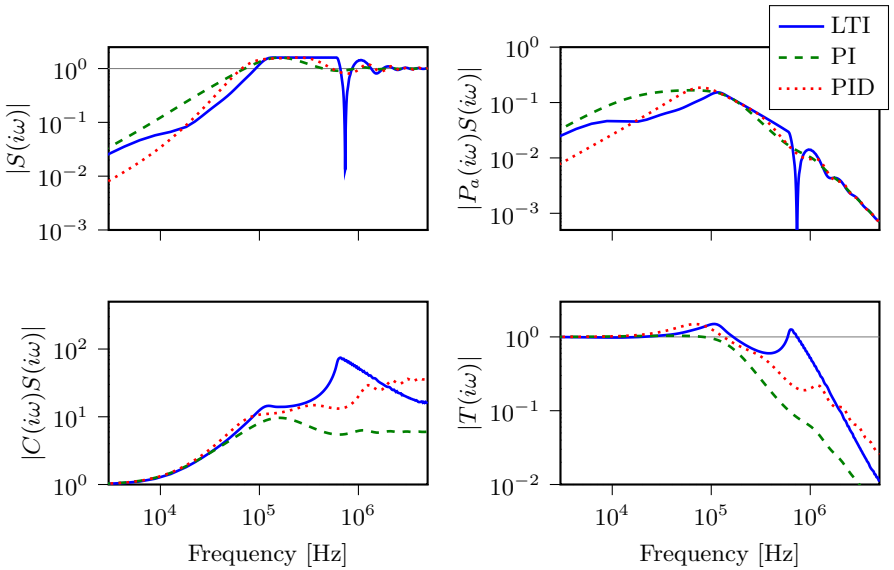


(a) Normal conducting cavity.

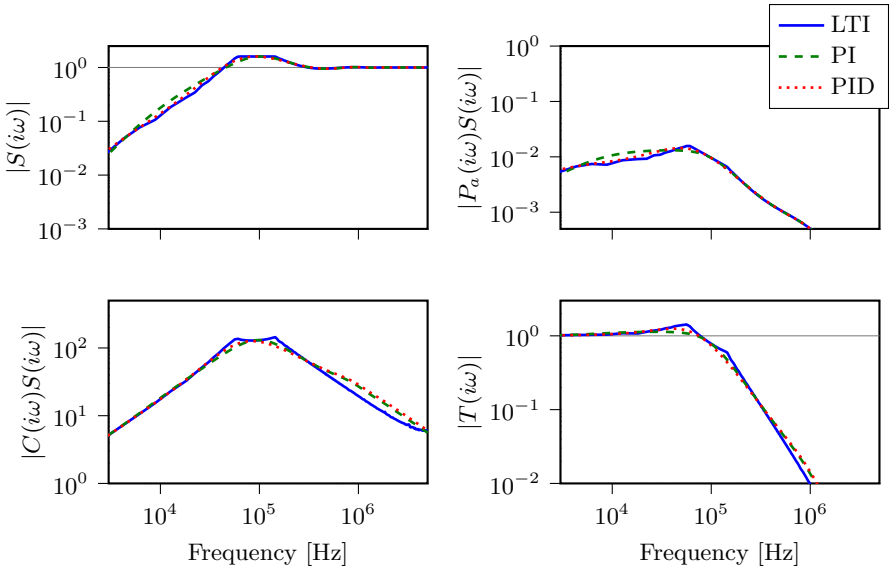


(b) Cold cavity.

Figure 9.4 Comparison between the open-loop frequency responses for three different controller types. Note that the Nyquist curve of the optimal LTI controller in subfigure (a) makes a huge turn into the upper half plane before going to zero for high frequencies.



(a) Warm cavity.



(b) Cold cavity.

Figure 9.5 Gang of four for different controllers that have been optimized for minimal field errors subject to the constraints $M_S \leq 1.6$ and $||G_{un}||_2 \leq 30$.

9.5 Conclusions

From the Bode diagrams in Figure 9.4 we see that the open-loop transfer functions for the different controllers are similar, with high gain at low frequencies and roll-off for higher-frequencies. However, the optimal LTI controllers do a better job of pushing up the loop gain at around 30 kHz, giving better disturbance rejection around that frequency (see also the sensitivity function in Figure 9.5). This is enabled by the high flexibility of the LTI controllers which allows them to push the sensitivity function all the way to the sensitivity constraint $M_S \leq 1.6$. This is also seen in Figure 9.4 where the Nyquist curves for the LTI controllers tightly follow the sensitivity circles.

In order to both push up the loop gain at around 30 kHz, while still respecting the sensitivity constraint, the optimal LTI controller for the normal conducting cavity makes a big turn into the upper half-plane before finally going to zero for high frequencies. To implement such a controller in practice is very challenging, and it would likely be sensitive to implementation errors and process variations.

Field errors vs. control signal activity

From Figure 9.2 we see that control performance is strongly dependent on the allowed control signal activity B_{un}, in particular for superconducting cavities. Superconducting cavities have low bandwidths, so it is necessary to increase the system bandwidth with a high-gain controller. The high gain makes the controller sensitive to measurement noise and the constraint on control signal activity leaves little freedom for the general LTI controller to achieve better performance than the PI(D)-controllers.

With the given constraints on control signal activity there is little benefit from using a more complicated controller than a PI controller for controlling superconducting cavities. For normal conducting cavities there are some benefits of using a PID controller or a more complicated controller over a PI controller. Note that decreased levels of measurement noise directly leads to a corresponding allowed increase of B_{un}, and thus quite significant performance improvements.

Field errors vs. loop delay

From Figure 9.3 it is seen that the field control performance is strongly dependent on loop delay. In particular, for normal conducting cavities for which the field error is almost proportional to the loop delay. This is because the limiting factor on the control performance of the normal conducting cavity is the robustness constraint $M_S \leq 1.6$, which becomes easier to satisfy for shorter delays. For the superconducting cavity, the performance is also limited by the constraint on the control signal activity.

Part II

Selected Topics in Cavity Field Control

10

Cartesian vs. Polar Feedback

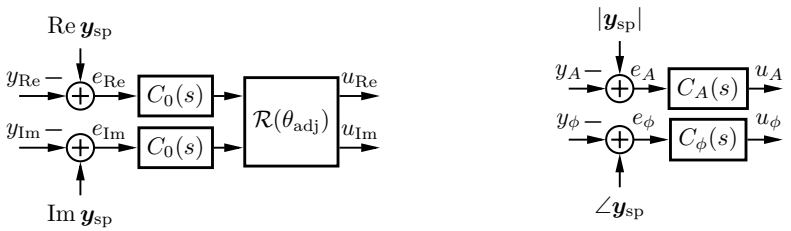
10.1 Introduction

In the Section 8.4 we considered controllers of the form $C(s) = C_0(s)\mathrm{e}^{i\theta_{\mathrm{adj}}}$. If $C_0(s)$ has real coefficients it can be seen as a decoupled controller between the real and imaginary parts of the measured signal and the control signal with the decoupling matrix [Skogestad and Postlethwaite, 2007] given by the rotation matrix

$$\mathcal{R}(\theta_{\mathrm{adj}}) := \begin{bmatrix} \cos\theta_{\mathrm{adj}} & -\sin\theta_{\mathrm{adj}} \\ \sin\theta_{\mathrm{adj}} & \cos\theta_{\mathrm{adj}} \end{bmatrix}, \tag{10.1}$$

see Figure 10.1a. We will refer to this controller structure as a Cartesian controller. Modern LLRF systems rely on quadrature up- and downconversion for interfacing the baseband field controller to the bandpass dynamics of the RF system. For these platforms it is natural to use Cartesian feedback.

Another controller structure for (nonlinear) decoupled field control is the polar controller where separate loops are used between the amplitudes and



(a) Cartesian controller.

(b) Polar controller.

Figure 10.1 Two alternatives to implement decoupled field control. The rotation matrix $\mathcal{R}(\theta)$ is defined in (10.1).

phases of the measured and controlled signals (Figure 10.1b). This controller structure is typical for previous-generation analog LLRF systems that used amplitude and phase detectors for detecting the complex envelope of the cavity signal. However, even for digital LLRF systems, there are certain advantages of polar feedback, in particular for circular accelerators [Salom et al., 2017].

A number of publications have briefly touched upon various aspects of this design decision between Cartesian and polar feedback [Pirkl, 1994; Garoby, 1996; Brandt, 2007; Plawski, 2014; Salom et al., 2017], often in the context of analog LLRF systems. We are not aware of that the performance and robustness around a nominal operating point has been analyzed, which is the main topic of this chapter. We also discuss a number of additional aspects for the case of linacs and digital LLRF systems.

For feedback linearization of RF amplifiers (Section 4.6 and [Ellingson, 2016]) there is similar design decision between Cartesian or polar feedback, however the considerations are quite different from the field-control setting.

Remark 10.1 The terminology Cartesian and polar feedback was chosen for consistency with the literature on telecommunications [Ellingson, 2016]. Existing field control literature often refer to the two approaches as IQ (in-phase-and-quadrature) control and amplitude-and-phase control.

10.2 Small-signal robustness of polar feedback

We will consider the model in Figure 8.10 with the polar controller in Figure 10.1b. The transfer function of the plant is given, as in (8.6), by

$$P(s) = \mathrm{e}^{-i\theta} P_0(s),$$

where $P_0(s)$ is the nominal open-loop transfer function and θ is a phase shift.

Polar feedback

The two controllers $C_A(s)$ and $C_\phi(s)$ in Figure 10.1b can be seen as interfaced to the model in Figure 8.10 through the coordinate transformations

$$\begin{aligned} f_{\mathbb{C}\to A\phi}(\cdot): \quad x+iy &\mapsto \left(\sqrt{x^2+y^2},\, \arctan(y,x)\right) \\ &= (A,\, \phi), \end{aligned} \tag{10.2a}$$

$$f_{A\phi\to\mathbb{C}}(\cdot): \quad (A,\, \phi) \mapsto (A\cos\phi,\, A\sin\phi) = x+iy, \tag{10.2b}$$

where $\arctan(\cdot,\cdot)$ refers to four-quadrant inverse tangent. This is illustrated in Figure 10.2. Note that the second output of $f_{\mathbb{C}\to A\phi}$ is not defined at $(0,0)$, this would have to be handled in practical implementations. In an analog

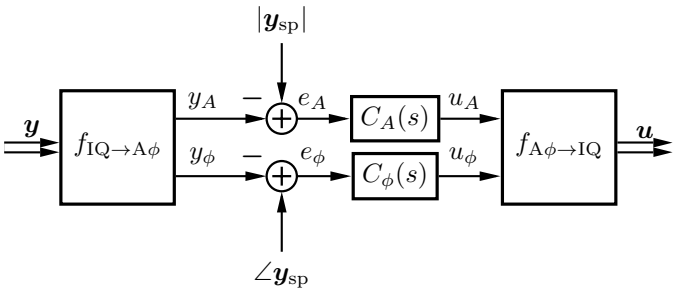


Figure 10.2 Model for how a polar controller is connected to the complex-baseband model in Figure 8.10.

LLRF system the polar transformations (10.2a)–(10.2b) would be implicitly implemented with amplitude and phase detectors together with variable-gain amplifiers and phase shifters. Digital implementations, using so-called CORDICs, add a delay of about 15 clock cycles[1] for each of the operations in (10.2a) and (10.2b).

The controller in Figure 10.2 is nonlinear and it is thus challenging to investigate global stability properties. We will not pursue this question here, but one may note that polar feedback has worked sufficiently well to have seen extensive practical use for field control. Our objective is instead to consider robustness and local stability for operation around a nominal operating point $(\mathbf{a_0}=1, \mathbf{f_{g0}}, \boldsymbol{y_0}, \boldsymbol{u_0})$.

The qualitative behavior of polar feedback around the given operating point is illustrated in Figure 10.3. It is seen that control errors are compensated at an angle equal to the angle of the RF-drive phasor relative to the cavity field. This angle equals 0 for a perfectly tuned cavity ($\Delta\omega = -\mathrm{Im}\,\gamma\mathbf{i_{b0}}$).

A quantitative analysis of the robustness and performance around the nominal operating point is possible by linearizing the static nonlinearities (10.2a) and (10.2b). Around a point $\boldsymbol{x}_0 = x_{0\rm Re} + ix_{0\rm Im} = A_0 \mathrm{e}^{i\phi_0}$, we have

$$\left.\frac{\partial f_{\mathbb{C}\to A\phi}}{\partial(x_{\rm Re}, x_{\rm Im})}\right|_{(A_0,\phi_0)} = \begin{bmatrix} \cos\phi_0 & \sin\phi_0 \\ -\dfrac{1}{A_0}\sin\phi_0 & \dfrac{1}{A_0}\cos\phi_0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1/A_0 \end{bmatrix} \mathcal{R}(-\phi_0) \tag{10.3a}$$

$$\left.\frac{\partial f_{A\phi\to\mathbb{C}}}{\partial(A,\phi)}\right|_{(A_0,\phi_0)} = \begin{bmatrix} \cos\phi_0 & -A_0\sin\phi_0 \\ \sin\phi_0 & A_0\cos\phi_0 \end{bmatrix} = \mathcal{R}(\phi_0) \begin{bmatrix} 1 & 0 \\ 0 & A_0 \end{bmatrix}, \tag{10.3b}$$

[1] A total delay of 32 clock cycles for the two transformations was reported in [Salom et al., 2017]. Delays of 10–15 clock cycles per transformation were reported by Christian Amstutz, ESS (the original implementation is due to Fredrik Kristensen).

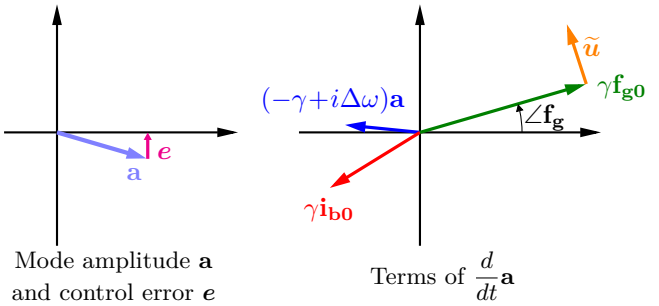


Figure 10.3 Illustration of the direction in which a polar controller applies the control action in response to a phase error. If the cavity is poorly tuned (i.e., $\angle\mathbf{f_{g0}} \neq 0$) then, as illustrated in the figure, the control action is not applied in the correct direction. Qualitatively this has the same effect as a loop-phase misadjustment $\delta = \angle\mathbf{f_{g0}}$.

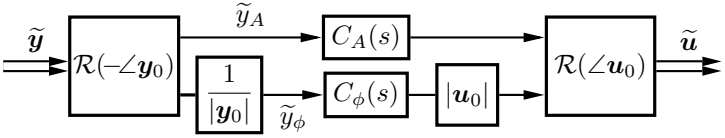


Figure 10.4 Linearization of the controller in Figure 10.2 around its nominal operating point. Small deviations of the control signal and measured signal are denoted by $\widetilde{\boldsymbol{u}}$ and $\widetilde{\boldsymbol{y}}$, respectively. By assuming $C_\phi(s) = (|\boldsymbol{y_0}| \,/\, |\boldsymbol{u_0}|)C_A(s)$ the controller becomes rotationally invariant, and can be represented as a complex-coefficient LTI system.

where $\mathcal{R}(\cdot)$ denote rotation matrices as in (10.1). Using (10.3) it follows that the controller in Figure 10.2 can be linearized[2] as in Figure 10.4.

If the disturbances and control objective are assumed to be rotationally invariant then it follows from Theorem 2 on page 52 that it is optimal to choose the amplitude and phase controllers so that

$$C_A(s) = C_0(s)$$
$$C_\phi(s) = (|\boldsymbol{y_0}| \,/\, |\boldsymbol{u_0}|)C_0(s).$$

With these choices, the polar controller in Figure 10.4 can be represented by

[2]Some justification of the linearizations (10.3) are in order. Typical field controllers keep the field errors well within 1 %/1°, while the control signal activity is kept within 15 % due to the RF amplifier's limited power-overhead. A similarly tuned phase loop gives 0.15 rad $\approx$ 9° phase variations of the control signal. This implies that our small-angle approximations are reasonable.

the complex-baseband system

$$C_{A\phi}(s) = \mathrm{e}^{i(-\angle \boldsymbol{y_0} + \angle \boldsymbol{u_0})} C_0(s). \tag{10.5}$$

Assuming that the closed-loop system converges to the given operating point $(\mathbf{a_0} = 1, \mathbf{f_{g0}}, \boldsymbol{y_0}, \boldsymbol{u_0})$, we see from Figure 8.10 that $\boldsymbol{u_0} = \mathrm{e}^{i\theta_1} \cdot \mathbf{f_{g0}}$, and $\boldsymbol{y_0} = \mathrm{e}^{-i\theta_2} \cdot 1$. Using these relations, we can re-write (10.5) as $C_{A\phi}(s) = \mathrm{e}^{i(\theta_2 + \theta_1 + \angle \mathbf{f_{g0}})} C_0(s)$.

Together with the system dynamics (10.2) we get the open-loop system

$$L_{A\phi}(s) = \mathrm{e}^{i\angle \mathbf{f_{g0}}} P_0(s) C_0(s). \tag{10.6}$$

From (10.6) we see that $\angle \mathbf{f_{g0}}$ plays the role of δ in (8.7).

Recall from Section 6.6 that $\angle \mathbf{f_{g0}} = 0$ corresponds to that the cavity is optimally tuned. Consequently, a poorly tuned cavity degrades the robustness and performance of polar feedback in the same way as loop-phase misadjustment does in the case of Cartesian feedback.

10.3 Aspects of Cartesian vs. polar feedback

We first summarize the implications from the previous section and then discuss other aspects of the choice between Cartesian or polar feedback.

Performance around a nominal operating point. With Cartesian feedback it is crucial to adjust the loop phase with θ_{adj} to ensure stability and acceptable phase margins. Although it is rather easy to calibrate θ_{adj}, this entails additional system complexity. With polar feedback there is no need to worry about the loop phase. It is instead the tuning of the cavity that affects the performance.

Linearity and transients during pulsed operation. The Cartesian controller is linear so transients from the pulsed operation and beam loading are easily handled. With polar feedback the effective gain in the ϕ direction depends on the magnitude of the input $\boldsymbol{u}$ and the output $\boldsymbol{y}$. It is certainly possible to patch this up by changing the gain based on $|\boldsymbol{u}|$ and $|\boldsymbol{y}|$ (instead of using $|\boldsymbol{u_0}|$ and $|\boldsymbol{y_0}|$), but this adds complexity.

For pulsed operation, one should probably also include a good initial guess of the integral state in the ϕ-controller to minimize the required settling time at every pulse. At $\boldsymbol{y_0} = 0$ the phase is undefined, this is another issue that needs consideration if polar feedback is used.

Implementation of iterative learning control for compensation of repetitive disturbances is less straight-forward if done in polar coordinates. One may of course convert to Cartesian coordinates for the iterative learning control but this would add complexity.

Control signal limitation and anti-windup. Limiting the output signal from the LLRF system is necessary to avoid over-driving the klystron. The limitation is trivial to implement in polar coordinates, but is more complicated in Cartesian coordinates if the number of clock cycles are to be kept low. Similarly, the implementation of anti-windup is simpler with polar feedback since it suffices to augment the amplitude loop with standard anti-windup. Anti-windup for a Cartesian controller is somewhat more complicated, see, Appendix F.

Delay. Implementing polar feedback on a digital LLRF system requires two polar transformation to be implemented in hardware, which as discussed above, adds a delay of about 30 clock cycles. This delay is significant compared typical delays of the field-control loop and it would hence give noticeably reduced control performance.

10.4 Summary

With Cartesian feedback it is *crucial* to adjust the loop-phase-adjustment angle θ_{adj}, either manually or with high-level software routines. However, for digital LLRF systems it is rather straightforward to implement such a calibration scheme. The loop phase is not an issue with polar feedback but it is necessary to ensure that the cavity is properly tuned. All aspects of transient operation are also more involved with polar feedback. Both controllers can provide similar levels of performance but Cartesian feedback seems to be the most convenient choice for typical field control loops (in linacs). The different aspects that have been discussed are summarized in Table 10.1.

Table 10.1 Pros and cons of Cartesian and polar feedback for digital field control (of linacs). Aspects for analog LLRF systems are mentioned in [Simrock and Geng, 2013] and [Garoby, 1996].

Cartesian feedback	Polar feedback
+ Simple and linear	+ No need to adjust loop phase
+ Not sensitive to cavity tuning	+ Easy implementation of saturation and anti-windup
− Control signal limitation and anti-windup are slightly more complicated to implement	− Inherently nonlinear, transient operation needs extra consideration
− Loop-phase adjustment is crucial	− Reduced robustness for poorly tuned cavities
	− Increased loop delay

11

Field Control of Cavities with Parasitic Modes

11.1 Introduction

Accelerating cavities have infinitely many parasitic modes in addition to the accelerating mode (Section 6.8). The interaction of these parasitic modes with the beam and the field control loop is detrimental to beam quality and control performance. The beam interacts with the modes that are close in frequency to the *machine lines* (integral multiples of the bunch frequency). These interactions need consideration to ensure sufficient beam quality [Ainsworth and Molloy, 2012]. For field control it is the modes that are close to the RF frequency that need consideration. It is often sufficient to consider just the one or two parasitic modes closest to the RF frequency, see Figure 11.1.

An early work on field control of cavities with parasitic modes is [Schilcher, 1998]. In [Pfeiffer et al., 2012; Schmidt et al., 2012], parasitic modes were handled by designing an $\mathcal{H}_\infty$ controller for a plant model on the real TITO form (4.14). Another work is [Vogel, 2007] which only considered the element $G_{\mathrm{Re}}(s)$ of the real-coefficient TITO representation (4.14). This approximation is valid if only the accelerating mode is considered but is not adequate for analyzing parasitic modes, since in this case $G_{\mathrm{Im}}(s)$ cannot be neglected (cf. Section 9.6 in [Zhou et al., 1996] and Section 4.6).

In this chapter we present and compare different approaches for dealing with parasitic modes in the control design, demonstrating how the complex-coefficient SISO representation (4.13) brings insight to the analysis. The investigations in this chapter are qualitative in nature and are mainly intended to demonstrate the utility of double-sided Bode diagrams and to illustrate approaches for dealing with parasitic modes in the field control design.

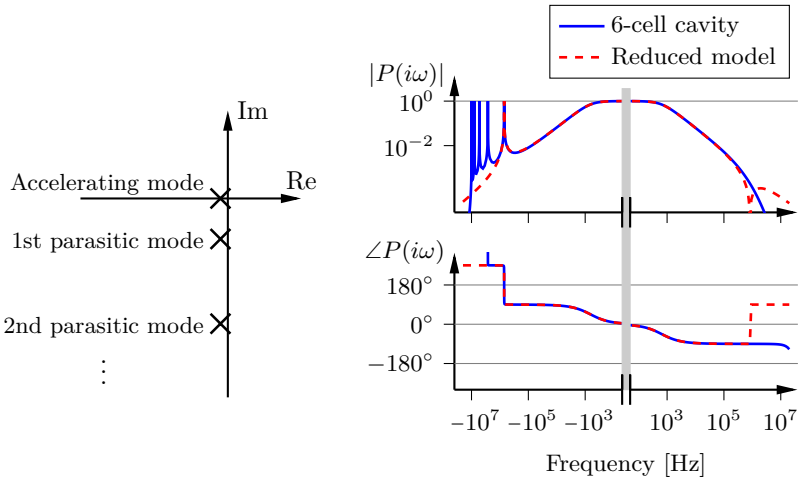


Figure 11.1 Pole–zero diagram and Bode diagram of the transfer function of a 6-cell elliptical cavity.

The danger of parasitic modes

Consider the plant model $P(s)$ from Section 9.3 where the cavity model *does not* include parasitic modes and assume that a nominal PI controller $C_0(s)$ is used for this plant. The resulting Nyquist curve is shown to the left in Figure 11.2.

If the same controller $C_0(s)$ is used when a parasitic mode is accounted for in the cavity model we get the Nyquist curve to the right in Figure 11.2. It is seen that there is a huge "bubble" from the parasitic mode encircling the point -1 which implies that the closed-loop system is unstable. This simple example illustrates that parasitic modes need to be considered in the design of field controllers.

11.2 Process model

Throughout this chapter we will consider the model in Figure 7.4, where the plant has the transfer function

$$P(s) = P_{\text{cav}}(s)P_{\text{amp}}(s)\mathrm{e}^{-s\tau}.$$

The time delay and amplifier dynamics will be assumed to be the same as in Section 9.3 ($\tau = 1\,\mu\text{s}$, $\omega_{\text{amp}} = 1.5\,\text{MHz}$). The cavity will be assumed to have a single parasitic mode

$$P_{\text{cav}}(s) = \frac{\gamma_a}{s+\gamma_a} - \frac{\gamma_1}{s+\gamma_1 - i\Delta\omega_1}. \tag{11.1}$$

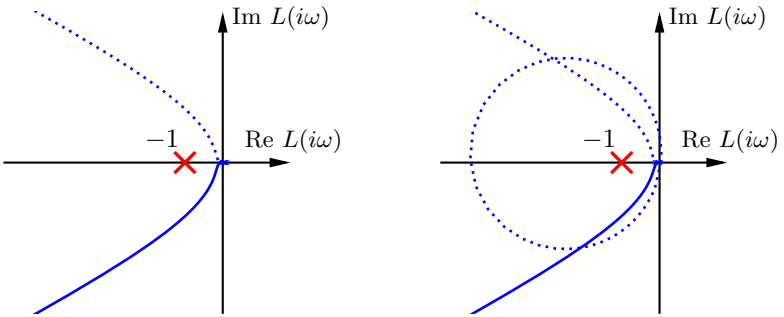


Figure 11.2 *Left:* Nyquist diagram for a superconducting cavity without parasitic modes controlled by a nominal PI controller. The dotted line indicates negative frequencies. *Right:* Nyquist curve for when the same nominal controller is used for a cavity *with* a parasitic modes. It is seen that the "bubble" from the parasitic mode encircles the point -1 which implies closed-loop instability. This example illustrates that parasitic modes must be considered in the controller design.

The cavity parameters will be taken as those of the ESS medium-β cavities, namely, $\gamma_a/(2\pi) = 500\,\text{Hz}$, $\Delta\omega_1/(2\pi) = 700\,\text{kHz}$, and $\gamma_1/(2\pi) = 930\,\text{Hz}$, see Table 6.5. The red dashed line of Figure 11.1 shows the transfer function (11.1) for these parameter values. In practice it could be necessary to consider more than one parasitic mode, but in this chapter we consider only one mode to simplify the exposition.

11.3 Control strategies for parasitic modes

To avoid instability from the parasitic mode, the corresponding bubble in the Nyquist curve $L(i\omega) = C(i\omega)P(i\omega)$ should not encircle or come close to the point -1. This can be achieved in two ways:

1. By sufficient *attenuation* of the parasitic mode so that the open-loop transfer function $L(s)$ is well below 1 around this frequency $\Delta\omega_1$.

2. By using the controller $C(s)$ for *phase adjustment* of the open-loop transfer function $L(s)$ around $\Delta\omega_1$ so that the bubble from the parasitic mode is directed away from the critical point -1.

For the first option, the phase of $L(i\omega)$ around $\Delta\omega_1$ is not important for stability and the closed loop is robust to variations of the loop delay. However, for the second option, where the bubble from the parasitic mode is allowed to remain be large, the design is sensitive to phase variations at the parasitic-mode frequency $\Delta\omega_1$.

Attenuation

To attenuate the parasitic mode one may use a controller with a high-order low-pass filter or a notch filter.

High-order low-pass filter. Augmenting the controller with low-pass filters of the form

$$F_1(s) = \frac{1}{sT_f + 1} \tag{11.2}$$

and

$$F_2(s) = \frac{1}{s^2T_f^2 + sT_f + 1}, \tag{11.3}$$

attenuates not only the closest parasitic mode with frequency $\Delta\omega_1$ but also higher-frequency parasitic modes. This strategy is not sensitive to the exact frequency of the parasitic mode.

Notch filter. To specifically reduce the controller gain around $\Delta\omega_1$ one may use a single-sided notch filter of the form

$$F_n(s) = \frac{s - i\omega_n}{s + \zeta_n\omega_n - i\omega_n} \tag{11.4}$$

and choose $\omega_n = \Delta\omega_1$.

Remark 11.1 (Alternative notch filter implementation) A discrete-time notch filter with very low implementation complexity is

$$\widetilde{F}_n(z) = \frac{1}{2}\left(1 + z^{-N_n}\right), \tag{11.5}$$

which was suggested in [Vogel, 2007]. Denoting the sampling frequency by h, the filter (11.5) gives a notch at the frequency $1/(2N_nh)$. A more sophisticated version of the filter that allows the notch frequency to be adjusted was also presented in [Vogel, 2007]. Disadvantages of the filter (11.5) are that the notch width cannot be adjusted and that it typically gives an unnecessarily large phase drop in the control loop.

11.4 Numerical comparison

We have considered the same objective and constraints as in Chapter 9. The disturbance spectrum and measurement noise spectrum was assumed to be the same, as well as the constraints on robustness and sensitivity ($S_{\max} = 1.6$ and $B_{un} = 30$). The approach to find the optimal controller parameters was the same as for the PI(D) controllers in the Chapter 9 (see Appendix E.2).

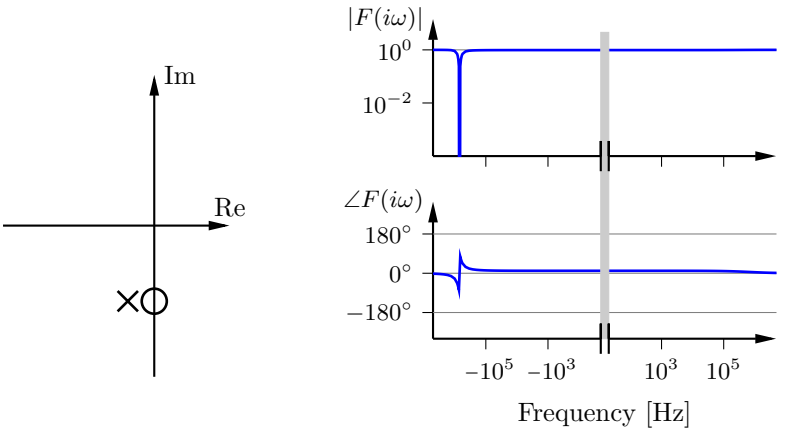


Figure 11.3 Pole–zero diagram and Bode diagram for the single-sided notch filter $F_n(s)$ in (11.4).

To enforce attenuation of the parasitic modes, the following constraint was imposed in the controller design

$$|P(i\omega)C(i\omega)| \leq 1 - \frac{1}{S_{\max}} \text{ for } |\omega| \geq \omega_{\lim}. \tag{11.6}$$

The frequency $\omega_{\lim}$ should be greater than the assumed cross-over frequency and lower than the frequencies of the parasitic modes.

Controller structures

We considered PI controllers of the form

$$C(s) = K\left(1 + \frac{1}{sT_i}\right) F(s)\mathrm{e}^{i\theta_{\mathrm{adj}}} \tag{11.7}$$

where the filter $F(s)$ was chosen as either:

1. The first-order low-pass filter $F_1(s)$ in (11.3).
2. The second-order low-pass filter $F_2(s)$ in (11.3).
3. A third-order low-pass filter $F_3(s) := F_1(s)F_2(s)$.
4. A fourth-order low-pass filter $F_4(s) := F_2(s)F_2(s)$.
5. A first-order low-pass filter plus a notch filter, $F_1(s)F_n(s)$.
6. A second-order low-pass filter plus a notch filter, $F_2(s)F_n(s)$.

Table 11.1 The performance for different approaches of dealing with parasitic modes in the control design. The table is normalized with respect to the performance for a cavity model without parasitic modes controlled by a PI controller with a first-order low-pass filter.

Type	Controller Structure	Objective
Baseline, no parasitic mode	PI+1st order filter	1.0
Attenuation of the parasitic mode	PI+1st order filter (Figure 11.4)	4.4
	PI+2nd order filter (Figure 11.5)	2.4
	PI+3rd order filter	2.1
	PI+4th order filter	2.0
	PI + notch filter + 1st order filter	1.1
	PI + notch filter + 2nd order filter (Figure 11.6)	1.1
Phase adjustment of the parasitic mode	PI+1st order filter	4.3
	PI+2nd order filter	2.1
	PI+3rd order filter (Figure 11.7)	1.6
	PI+4th order filter	1.4

The reason for including the factor $e^{i\theta_{adj}}$ in (11.7) is that the transfer function $P(s)$ has complex coefficients from the parasitic mode and hence it is typically suboptimal to choose $\theta_{adj} = 0$ as was done in Chapter 9.

Results

For the case of attenuation of the parasitic mode the filters 1–6 were considered and for the case of phase adjustment the filters 1–3 were considered.

The field errors relative to a baseline field control loop without parasitic modes (i.e., the model in Chapter 9), are given in Table 11.1. For some controller designs we have provided illustrative frequency domain plots, see Figures 11.4 to 11.7. A frequency-domain comparison between four different control approaches is provided in Figure 11.8.

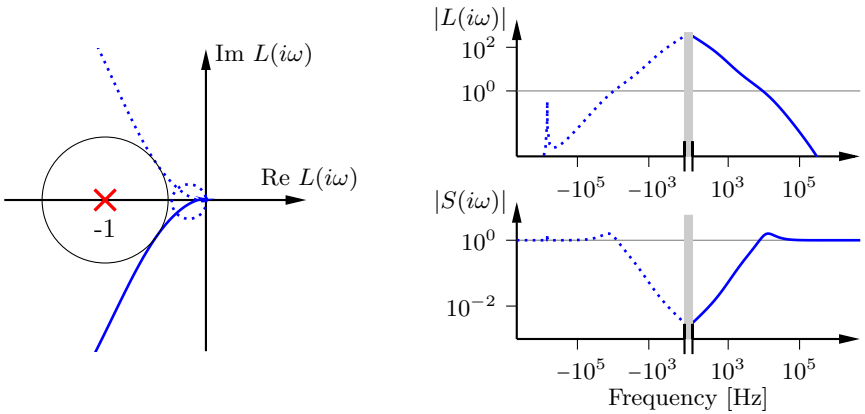


Figure 11.4 Attenuation of the parasitic mode with a PI controller with a first-order filter. The constraint on sufficiently low attenuation of the parasitic mode has resulted in a low cross-over frequency. This implies poor rejection of low-frequency disturbances.

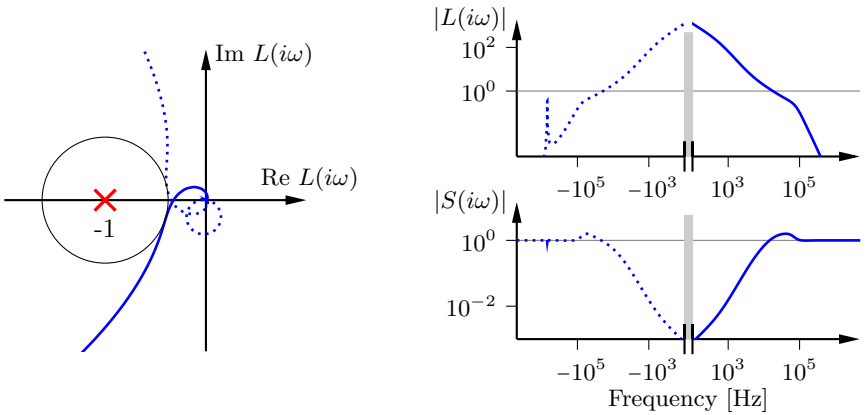


Figure 11.5 Attenuation of the parasitic mode with a PI controller with a second-order low-pass filter.

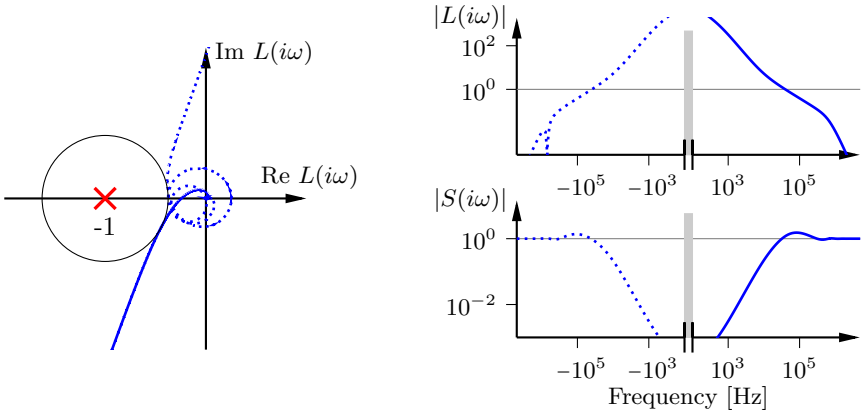


Figure 11.6 Attenuation of the parasitic mode using a PI controller with a notch filter and a second-order low-pass filter. Note that the parasitic mode is attenuated by the notch filter so the second-order low-pass filter can have a relatively fast time constant. This enables a higher cross-over frequency and better rejection of low-frequency disturbances.

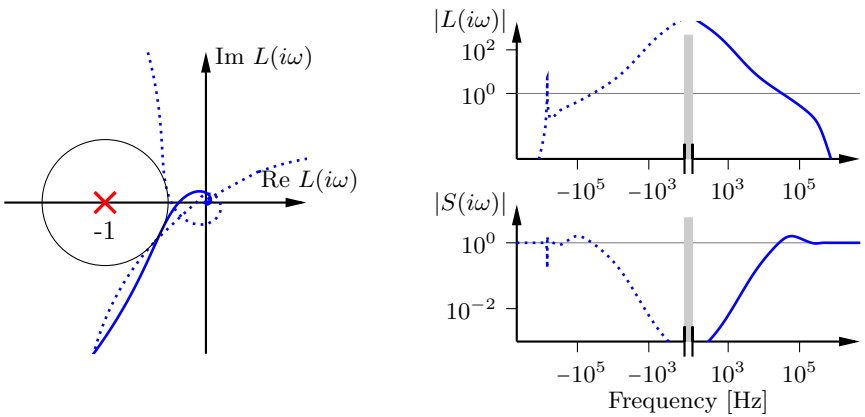


Figure 11.7 Phase adjustment of the parasitic mode with a PI controller and a third-order low-pass filter. Note the big bubble in the Nyquist curve that results from that the parasitic mode is not attenuated. Even a small changes in the loop phase might lead to that the bubble encircles the critical point -1 which would give instability.

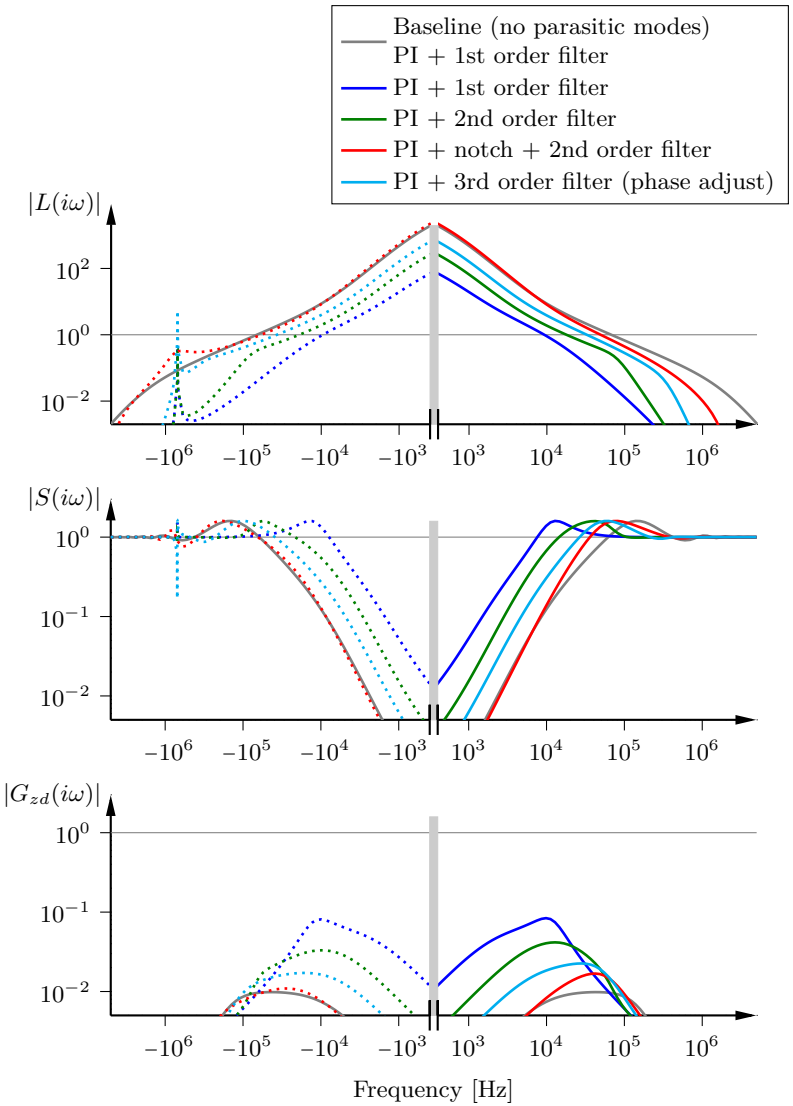


Figure 11.8 Comparison between different control strategies to avoid instability from parasitic cavity modes. It is seen that the achievable control bandwidth depends on the controller structure. The controller with a notch filter achieves the highest bandwidth.

11.5 Conclusions

To avoid instability from parasitic cavity modes that are close in frequency to the accelerating mode, the gain of the field controller must be reduced around the problematic modes (unless the dubious approach of phase adjustment is used). The performance degradation from parasitic modes depends strongly on the strategy chosen for dealing with them, see Table 11.1.

One simple approach to deal with parasitic modes is to design robust controller for a cavity model without parasitic mode and then augment it with notch filters at the frequencies of problematic parasitic modes. A perhaps better approach would be to include notch filters with desired characteristics in the process model and then design the controller based on this model. The main consideration with a notch-filter approach is that the zero frequencies of the notch filters need to be matched sufficiently well with the parasitic-mode frequencies.

If notch filters are not used, the bandwidth of the control loop must be reduced to achieve acceptable robustness, see Figure 11.8. This leads to reduced attenuation of low-frequency disturbances.

This investigations in this chapter have been far from exhaustive. The relative performance of the different control approaches depend on the assumed spectrum of the load disturbances. If the disturbance spectrum would have been assumed to have more low-frequency character than the one in Figure 9.1 then it would have been more important to push up the controller bandwidth.

11.6 Beam-loading-induced field-control error

In this section we discuss another problem with parasitic modes that arises when a linac is calibrated with a lower beam current than nominal.

Typical field controllers measure and regulate the voltage $\mathbf{v}_{\mathrm{pu}}$ measured by the cavity pickup probe to some given setpoint $\mathbf{v}^{\star}_{\mathrm{pu}}$. The actual objective, however, is to keep the voltage of the accelerating mode $\mathbf{a}_a$ at some desired value $\mathbf{a}^{\star}_a$. If parasitic modes couple to the pickup then beam loading may induce a systematic field-control error $\boldsymbol{\delta} = \mathbf{a}_a - \mathbf{a}^{\star}_a$. We demonstrate this by using the lumped model in Figure 6.9 and by considering two situations, tuning and operation.

A. Tuning

Let $\mathbf{a}^{\star}_a$ denote the accelerating-mode amplitude that is optimal from a beam-physics perspective. During accelerator tuning, beam-physicists search for a setpoint $\mathbf{v}^{\star}_{\mathrm{pu}}$, that makes field amplitude $\mathbf{a}_a = \mathbf{a}^{\star}_a$. This is typically done using short, or low-current pulses. In these situations the beam will not significantly affect the cavity fields that it sees, and we may assume that $\mathbf{i_b} = 0$. At the desired conditions (in stationarity), we have, for some $\mathbf{f}^{A}_{\mathbf{g0}}$,

$$\mathbf{v}^{\star}_{\mathrm{pu}} = [P_a(0) + P_{\mathrm{x}}(0)]\,\mathbf{f}^{A}_{\mathbf{g0}} \tag{11.8a}$$

$$\mathbf{a}^{\star}_a = P_a(0)\mathbf{f}^{A}_{\mathbf{g0}}. \tag{11.8b}$$

B. Operation

During nominal operation with beam loading $\mathbf{i_{b0}}$ the field-controller keeps the pickup voltage at the desired setpoint $\mathbf{v}^{\star}_{\mathrm{pu}}$. This gives

$$\mathbf{v}^{\star}_{\mathrm{pu}} = [P_a(0) + P_{\mathrm{x}}(0)]\,\mathbf{f}^{B}_{\mathbf{g0}} + [P_a(0) + P_{\mathrm{xb}}(0)]\,\mathbf{i_{b0}} \tag{11.9a}$$

$$\mathbf{a}^{B}_a = P_a(0)\mathbf{f}^{B}_{\mathbf{g0}} + P_a(0)\mathbf{i_{b0}}. \tag{11.9b}$$

It can be seen that there will be a systematic error $\boldsymbol{\delta} = \mathbf{a}^{B}_a - \mathbf{a}^{\star}_a$ during operation. To see this, note that from (11.8a) and (11.8b) we have $\mathbf{v}^{\star}_{\mathrm{pu}} = (P_a(0) + P_{\mathrm{x}}(0))/P_a(0)\mathbf{a}^{\star}_a$. Plugging this expression into (11.9a) gives

$$\mathbf{f}^{B}_{\mathbf{g0}} = \frac{1}{P_a(0)}\mathbf{a}^{\star}_a - \frac{P_a(0) + P_{\mathrm{xb}}(0)}{P_a(0) + P_{\mathrm{x}}(0)}\mathbf{i_{b0}}.$$

Substituting this expression for $\mathbf{f}^{B}_{\mathbf{g0}}$ in (11.9b) we see that

$$\begin{aligned}\boldsymbol{\delta} = \mathbf{a}^{B}_a - \mathbf{a}^{\star}_a &= -P_a(0)\frac{P_a(0) + P_{\mathrm{xb}}(0)}{P_a(0) + P_{\mathrm{x}}(0)}\mathbf{i_{b0}} + P_a(0)\mathbf{i_{b0}} \\ &= \frac{P_{\mathrm{x}}(0) - P_{\mathrm{xb}}(0)}{P_a(0) + P_{\mathrm{x}}(0)}P_a(0)\mathbf{i_{b0}} \approx (P_{\mathrm{x}}(0) - P_{\mathrm{xb}}(0))\mathbf{i_{b0}}.\end{aligned} \tag{11.10}$$

Note that $\boldsymbol{\delta}$ is (approximately) proportional to $\mathbf{i_{b0}}$. From (6.29) and (6.30) we see that $\boldsymbol{\delta}$ is also (approximately) proportional to γ_π.

Remark 11.2 There will be a non-zero error $\boldsymbol{\delta}$ even if the beam does not couple to the parasitic modes.

Remark 11.3 This issue cannot be observed from the simplified model in Figure 6.10, since there $P_{\mathrm{x}}(s) = P_{\mathrm{xb}}(s)$.

Example: Let us compute the beam-loading-induced error for a medium-β cavity at the European Spallation Source. These are superconducting, 6-cell elliptical cavities with $\gamma_\pi/(2\pi) = 500\,\mathrm{Hz}$ and $\Delta\omega_{5\pi/6}/(2\pi) = -700\,\mathrm{kHz}$. For simplicity, include only the $5\pi/6$ mode in $P_{\mathrm{x}}(s)$, assume that the particle velocity equals the cavity's design velocity ($P_{\mathrm{xb}}(s) = 0$), and that $\mathbf{i_{b0}} = 1$. This gives

$$\boldsymbol{\delta} \;=\; P_{\mathrm{x}}(0)\mathbf{i_{b0}} \;\approx\; \frac{\gamma_{5\pi/6}}{\gamma_{5\pi/6} + i\Delta\omega_{5\pi/6}} \;\approx\; \frac{\gamma_{5\pi/6}}{i\Delta\omega_{5\pi/6}} \;=\; \frac{R_5^2\gamma_\pi}{i\Delta\omega_{5\pi/6}} \;\approx\; 0.0013i.$$

This corresponds to a phase error of $0.0013\,\mathrm{rad} \approx 0.076°$. This value is significant relative to field error requirement of $0.1°$ for the ESS medium-β section.

Possible remedies

The issue with beam-loading-induced biasing could be remedied by using a Kalman filter to estimate the true value of $\mathbf{a}_a$ and use this value for feedback and calibration. Another possibility is to recalibrate the setpoint $\mathbf{v}^\star_{\mathrm{pu}}$ during operation with nominal beam current. If the error $\boldsymbol{\delta}$ is much smaller than the allowed phase errors one could just ignore it.

12

Digital Downconversion for Cavity Field Control

Digital LLRF systems typically use digital (quadrature) downconversion to recover the complex envelopes of the cavity signal and other RF signals of interest. Digital downconversion is illustrated in Figure 12.1 and was also discussed in Section 5.1.1. The focus of this chapter is the filter $H(z)$ in Figure 12.1 which should attenuate the double-frequency component of the mixer output and reduce aliasing. The filter $H(z)$ must have a low latency to minimize the impact on field control performance.

There is much literature on how to design the filter $H(z)$ for communications applications [Crochiere and Rabiner, 1983; Creaney and Kostarnov, 2008; Ellingson, 2016], but in this literature the main design objectives for $H(z)$ are passband flatness, stopband attenuation, and resource usage. The latency is usually not a major concern.

This chapter presents and compares two approaches to digital downconversion that have successfully been used for cavity field control: moving averaging [Simrock et al., 2006; Hoffmann, 2008; Schilcher, 2007] and two-sample reconstruction [Doolittle et al., 2006; Doolittle, 2008].

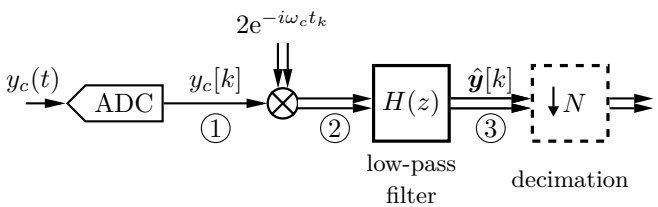


Figure 12.1 Typical implementation of digital downconversion. The output $\hat{\boldsymbol{y}}$ is an estimate of the complex envelope of y_c. The low-pass filter should reject the double-frequency components of the mixer output and reduce aliasing (if the signal is decimated).

Remark 12.1 Control of MEMS gyroscopes [Saggin et al., 2019] and single-phase phasor estimation by phasor-measurement units in three-phase power systems [Phadke et al., 1983; Phadke and Thorp, 2017] are two other feedback applications where low-latency digital downconversion is used.

12.1 Background on digital downconversion

Recall the following representation of a bandpass signal from (4.2),

$$y_c(t) = \mathrm{Re}\{\boldsymbol{y}(t)\mathrm{e}^{i\omega_c t}\},$$

where the complex envelope $\boldsymbol{y}$ varies slowly. In Section 4.1 we briefly discussed how the complex envelope of $\boldsymbol{y}$ can be recovered from y_c. In many applications this recovery is implemented digitally. Assume that $y_c(t)$ is sampled with period h at time instances $t_k = kh$. A sampled version of the complex envelope $\boldsymbol{y}$ can then be recovered by digital downconversion as shown in Figure 12.1.

To better understand Figure 12.1, note that the signals at the indicated points are given by

$$\textcircled{1} \qquad \begin{aligned} y_c[k] &= \mathrm{Re}\big\{\boldsymbol{y}[k]\mathrm{e}^{i\omega_c t_k}\big\} \\ &= (\boldsymbol{y}[k]\mathrm{e}^{i\omega_c t_k} + \boldsymbol{y}[k]^*\mathrm{e}^{-i\omega_c t_k})/2, \end{aligned}$$

$$\textcircled{2} \qquad \boldsymbol{y}[k] + \boldsymbol{y}[k]^*\mathrm{e}^{-2i\omega_c t_k}.$$

The low-pass filter removes the double-frequency component around $-2\omega_c$, giving an estimate $\hat{\boldsymbol{y}}[k] \approx \boldsymbol{y}[k]$ at ③.

As illustrated in Figure 4.1, one may think of downconversion in the frequency-domain as a translation of the frequency spectrum by $-\omega_c$ followed by truncation of low frequencies. For an illustration more specific to *digital* downconversion that shows the aliased spectral components present in sampled signals, see Figure 18.14 in [Ellingson, 2016].

Baseband model of digital downconversion

For detailed analysis of control performance it is convenient with a complex-baseband model of Figure 12.1. Such a model of the dynamics from the complex envelope $\boldsymbol{y}$ of the measured signal to the downconverted signal $\hat{\boldsymbol{y}}$ is shown Figure 12.2. It is seen that the baseband behavior of digital downconversion amounts to addition of complex-valued, cyclostationary noise $\boldsymbol{n}[k]$ and filtering through a linear, time-invariant (LTI) filter $H(z)$. For cavity field control the filter $H(z)$ is selected to have low-latency and hence little impact on closed-loop performance. It is nevertheless prudent to include the filter dynamics $H(z)$ and the measurement noise $\boldsymbol{n}[k]$ in performance analysis.

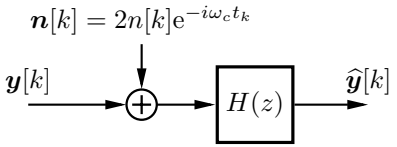


Figure 12.2 Baseband model of digital downconversion that includes measurement noise. The double-frequency component $\boldsymbol{y}^*[k]\mathrm{e}^{-2i\omega_c t_k}$ is not included since it should be rejected by properly designed filters $H(z)$.

Digital downconversion for communications applications

In communications applications, the signal of interest has a relatively flat spectrum across the channel bandwidth. It is desirable to extract this signal without distortion, while avoiding unwanted spectral content which is typically present close to the channel. This calls for a filter $H(z)$ that has flat amplitude characteristics over the channel bandwidth and then a fast transition to a well-attenuated stop band.

The desired features $H(z)$ could be summarized as

1. flat amplitude response in the passband, and
2. excellent stopband suppression.

In particular the second item requires significant engineering efforts since it is often difficult to achieve sufficient attenuation of adjacent channels and spurious interference[1].

Typically, $H(z)$ is taken as an FIR filter $H(z) = h_0 + h_1 z^{-1} + \cdots + h_{N-1} z^{-(N-1)}$ with symmetric coefficients ($h_k = h_{N-1-k}$). The symmetry ensures that $H(z)$ is a *linear phase*[2] filter which has the same phase response as a time delay of $(N-1)/2$ samples. This time delay is often referred to the *group delay*[3] of the system. For communications applications the delay requirement is rarely a major concern[1].

The impulse and frequency responses of a typical filter $H(z)$ for communications applications are shown in Figures 12.3 and 12.4.

In addition to the requirements 1–2 above, the design of H should allow an implementation with

3. low power consumption, and
4. few hardware resources.

[1] Personal communication with Bo Bernhardsson, former filter expert at Ericsson.

[2] The phase is a linear function of frequency.

[3] The group delay $\tau_g(\omega) := -d\angle H(i\omega)/d\omega$ of a linear system $H(s)$ is a common concept in physics and filter design. Group delay is rarely considered in control since it is rather the phase $\angle H(i\omega)$ of the linear systems that is of interest.

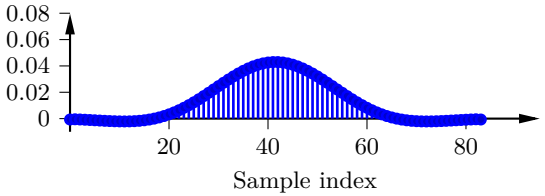


Figure 12.3 Impulse response of a typical low-pass filter for digital downconversion in communications applications.

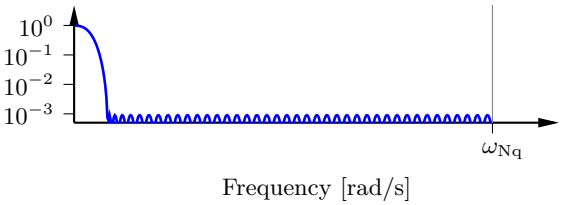


Figure 12.4 Magnitude response of the filter in Figure 12.3. The phase response equals that of a 42-sample time delay.

For these reasons, the filter $H(z)$ is typically implemented as a cascade of filters at different rates. The first filter in the cascade is often a cascaded-integrator–comb filter [Hogenauer, 1981]. For more details, see [Creaney and Kostarnov, 2008; Ellingson, 2016].

To reduce the computations in the downstream baseband processing, the signal downstream $H(z)$ is typically decimated as much as possible. This can be done without loss of information due to that the spectrum of the baseband signal is centered around zero frequency.

12.2 Low-latency digital downconversion for field control

12.2.1 Desired features for low-latency downconversion

The requirements on digital downconversion for control applications are quite different from those in communications. The signal of interest has (due to the feedback) very narrow spectrum and hence it gives rise to ADC and mixer[4] harmonics, as well as a strong double-frequency component. It is important to avoid that decimation folds these products onto the signal of interest. However, unlike in communications applications, there is no need to worry about folding of spectral content from adjacent channels.

[4] If a heterodyne architecture is used.

A main concern is instead to attenuate measurement noise (mostly ADC quantization noise) since it gives rise to control errors and increased control signal activity. However, most important is to maintain sufficient phase margins. The desired features of $H(z)$ for low-latency feedback applications could be summarized as

1. small phase drop around the cross-over frequency of the open-loop system,
2. rejection of (ADC) measurement noise,
3. high attenuation of the double-frequency component,
4. high attenuation of the DC-offset spur, and
5. attenuation of mixer and ADC harmonics.

The trade-off between 1 and 2 is studied in the control literature [Åström and Wittenmark, 1997, Sec. 7.4]. The next three subsections introduce filters that address points 3–5, however it may still be necessary with additional low-pass filtering to meet requirement 2.

The filters $H(z)$ that we consider in this section are rather simple, so resource usage should not be a concern when using modern FPGAs. Additionally, the hardware tends to be less penny pinched than for communications applications, so resources are less of an issue in the first place.

Since neither power consumption nor hardware resources tend to be an issue for field control it is feasible to use little, or no, decimation after the filter $H(z)$. Decimation by a factor N gives a controller period $h_{\text{reg}} = Nh$. The trade-offs in the selection of h_{reg} are discussed extensively in [Åström and Wittenmark, 1997]. A controller period h_{reg} effectively gives a latency of $h_{\text{reg}}/2$ compared to an ideal, continuous-time controller [Åström and Wittenmark, 1997]. For this reason h_{reg} should be kept small. On the other hand, this increases the resolution necessary for coefficients and computations in the controller.

Remark 12.2 Requirement 2 in the list above is typically understood in a 2-norm sense, while the requirement on stop-band suppression in Section 12.1 is typically understood in a sup-norm sense.

12.2.2 IQ sampling

Define the sampling frequency $f_s := 1/h$ and let $f_c := \omega_c/(2\pi)$. A simple, computationally efficient, and low-latency approach to digital downconversion is to take $f_c = f_s/4$ and $H(z) = 1 + z^{-1}$. This is known as *IQ sampling* [Doolittle et al., 2006] or $f_s/4$ *sampling* [Ellingson, 2016]. The zero of $H(z)$ in -1 eliminates the double-frequency component at $-f_s/2$. This scheme is efficient to implement since it requires no multiplications [Ellingson, 2016].

The narrow bandwidth of the cavity signal gives rise to mixer and ADC harmonics. With IQ sampling, all odd harmonics alias to the zero baseband frequency [Doolittle et al., 2006].

12.2.3 Near-IQ sampling

To avoid the drawbacks of IQ sampling one can use *near-IQ (non-IQ) sampling* [Doolittle et al., 2006; Schilcher, 2007] where N samples are taken over M periods of the carrier signal, i.e., $M/f_c = N/f_s$.

For near-IQ sampling, a slightly more complex filter $H(z)$ is necessary to eliminate the double-frequency component. Two alternatives for the filter $H(z)$ are the moving average filter and two-sample reconstruction; these are discussed in the following subsections.

It will be convenient to introduce the phase advance between consecutive samples as

$$\Delta \coloneqq \omega_c h = 2\pi M/N. \tag{12.1}$$

Note that Δ corresponds to ω_c in normalized angular frequency.

Example: The LLRF systems for ESS use moving averaging (with decimation). The sampling frequency is $f_s = 704.42\,\text{MHz}/6 = 117.403\,333\,\text{MHz}$ and the intermediate frequency is 25.1579 MHz which gives $M/N = 3/14$.

Example: The LLRF systems for LCLS-II use two-sample reconstruction. The sampling frequency is $f_s = 94.3\,\text{MHz}$ and the intermediate frequency is $f_c = 20\,\text{MHz}$ which gives $M/N = 7/33$ [Doolittle et al., 2016].

12.2.4 Near-IQ filter: Moving averaging

With near-IQ sampling, the N-sample moving average filter

$$H_{\text{MA-}N}(z) = \frac{1}{N}\left(1 + z^{-1} + \ldots + z^{-(N-1)}\right) = \frac{1 - z^{-N}}{1 - z^{-1}} \tag{12.2}$$

is a common choice [Schilcher, 2007; Phadke and Thorp, 2017].

The impulse and frequency responses of an 11-sample moving average filter are shown in Figures 12.5 and 12.6. An attractive feature of the moving average filter (12.2) is that it has zeros at the frequencies of the double-frequency component, the DC-offset spur, and all ADC and mixer harmonics (except multiples of N).

If decimation by a factor N is used after the moving average filter (12.2) is a special case of a single-stage cascaded-integrator–comb (CIC) filter [Hogenauer, 1981]. In [Hoffmann, 2008] this is referred to as *step window detection* and the case of no decimation is referred to as *sliding window detection*.

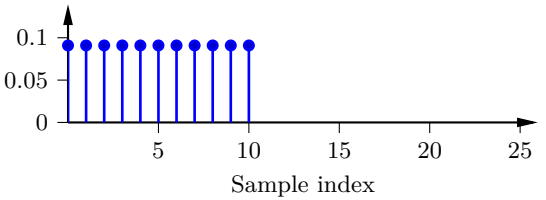


Figure 12.5 Impulse response of an 11-sample moving average filter.

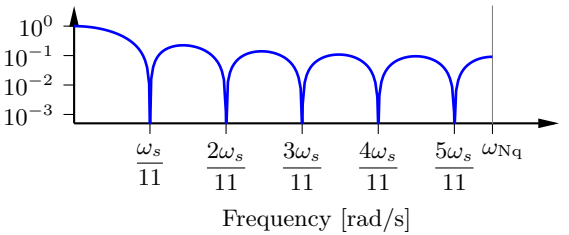


Figure 12.6 Magniude response of an 11-sample moving average filter; $\omega_s = 2\pi/h$. The phase response is that of a time delay of $11h/2$.

Digital downconversion with a moving average filter has two somewhat illuminating interpretations: (1) as a short-time, discrete-time Fourier transform with a single bin (at frequency ω_c) [Hoffmann, 2008]; and (2) as the least-squares estimate of $\boldsymbol{y}$ given N observations $y[k], \ldots, y[k+N-1]$ [Schilcher, 2007].

Note that the moving average filter is linear phase with relatively little amplitude drop up for low frequencies. Thus, it can be modeled as a time-delay of $(N-1)h/2$ for stability and disturbance-rejection analysis. The power spectrum of white noise filtered through a moving average filter is given by [Hogenauer, 1981] $|N(i\omega)|^2 = \left[\sin(\omega/2)/\sin(\omega/(2N))\right]^2$.

Moving average filters have relatively high side lobes which indicates that there would be significant aliasing if the filter output is decimated. However, if the signal is decimated by the filter length N these side lobes fold away from the zero frequency, allowing them to be rejected by a low-pass filter operating at the lower rate [Hogenauer, 1981].

12.2.5 Near-IQ filter: Two-sample reconstruction

Two-sample reconstruction estimates the complex envelope $\boldsymbol{y}$ of a signal y_c based on only two samples. In this chapter we take a slightly different approach than in [Doolittle, 2008] and represent the algorithm on the form in Figure 12.2. This representation is good for analysis but the formulation in [Doolittle, 2008] gives a more efficient implementation.

To motivate the algorithm, assume that two samples of y_c are taken at times $t_{k-1} = (k-1)h$ and $t_k = kh$, and that $\boldsymbol{y}$ does not change between these two samples. The two consecutive samples can be written

$$y_c[k-1] = \mathrm{Re}\{\boldsymbol{y}\mathrm{e}^{i\omega_c t_{k-1}}\} = \frac{1}{2}\left(\boldsymbol{y}\mathrm{e}^{i\omega_c t_{k-1}} + \boldsymbol{y}^*\mathrm{e}^{-i\omega_c t_{k-1}}\right) \tag{12.3a}$$

$$y_c[k] = \mathrm{Re}\{\boldsymbol{y}\mathrm{e}^{i\omega_c t_k}\} = \frac{1}{2}\left(\boldsymbol{y}\mathrm{e}^{i\omega_c t_k} + \boldsymbol{y}^*\mathrm{e}^{-i\omega_c t_k}\right) \tag{12.3b}$$

Now, taking (12.3b) times $\mathrm{e}^{-i\omega_c t_k}$ minus (12.3a) times $\mathrm{e}^{-2i\Delta}\mathrm{e}^{-i\omega_c t_{k-1}}$ gives

$$\mathrm{e}^{-i\omega_c t_k}y_c[k] - \mathrm{e}^{-2i\Delta}\mathrm{e}^{-i\omega_c t_{k-1}}y_c[k-1] = \frac{1}{2}\left(\boldsymbol{y} - \mathrm{e}^{-2i\Delta}\boldsymbol{y}\right).$$

Thus, $\boldsymbol{y}$ can be recovered as

$$\boldsymbol{y} = \frac{2}{1-\mathrm{e}^{-2i\Delta}}\left(\mathrm{e}^{-i\omega_c t_k}y_c[k] - \mathrm{e}^{-2i\Delta}\mathrm{e}^{-i\omega_c t_{k-1}}y_c[k-1]\right). \tag{12.4}$$

The process of recovering $\boldsymbol{y}$ according to the equation above can be seen as filtering the signal $2\mathrm{e}^{-i\omega_c t_k}y[k]$ through the filter

$$\begin{aligned} H_{2\mathrm{S}}(z) &= \frac{1}{1-\mathrm{e}^{-2i\Delta}}\left[1 - \mathrm{e}^{-2i\Delta}z^{-1}\right] \\ &= \frac{\mathrm{e}^{i\Delta}}{2i\sin\Delta}\left[1 - \mathrm{e}^{-2i\Delta}z^{-1}\right] \end{aligned} \tag{12.5}$$

according to Figure 12.1. The complex factor $\mathrm{e}^{i\Delta}/i$ may be discarded if only relative phase is of interest, but for our analysis it is convenient to have $\angle H_{2\mathrm{S}}(1) = 0$. From the magnitude plot in Figure 12.5 we see that $H_{2\mathrm{S}}(z)$ is not low-pass filter but rather a notch filter with notch frequency $-2\omega_c$ (from the zero in $\mathrm{e}^{-2i\Delta}$). The notch eliminates the double-frequency component of the mixer output.

Note from (12.5) that $H_{2\mathrm{S}}(z)$ has complex coefficients which implies that: (1) its frequency-response is not conjugate symmetric (see Fig 12.7) and (2) the complex-signal representation in Figure 12.1 must be used to describe two-sample reconstruction, the version with two real-valued branches in [Ellingson, 2016, Sec. 18.4] is not sufficiently general.

From (12.5) it is seen that Δ should be close to $\pi/2$ so that $|\sin\Delta| \approx 1$ which avoids amplification of measurement noise [Doolittle, 2008].

For implementation details and Verilog code for two-sample reconstruction see [Doolittle, 2008].

DC-spur rejection. The communications-grade ADCs used in LLRF systems can have significant DC offset errors[5]. From Figure 12.2 it is seen that

[5]The offset error can be on the order of 1 % of the maximum amplitude of the sine wave y_c [Personal communication with Larry Doolittle, LBNL]. For example, LTC2175 from Linear Technologies may have an offset error of up to 1.2% relative to the maximum sine wave amplitude.

such an offset n_0 at the ADC output generates a spur $2n_0 e^{-i\omega_c t_k}$. This spur can be removed by introducing a DC-rejection filter $(1 - z^{-2})/2$ before the digital mixer in Figure 12.1 [Doolittle, 2008]. The DC-rejection filter is conveniently analyzed by transforming it to the baseband ($z \mapsto e^{i\omega_c h} z = e^{i\Delta} z$) where it corresponds to a filter $(1 - e^{-2i\Delta} z^{-2})/2$. Normalizing for unity gain at zero baseband frequency gives

$$H_{\mathrm{DCR}}(z) = \frac{1 - e^{-2i\Delta} z^{-2}}{1 - e^{-2i\Delta}} = \frac{e^{i\Delta}}{2i \sin(\Delta)} \left[1 - e^{-2i\Delta} z^{-2}\right]. \tag{12.6}$$

This filter should only be used if $\Delta \approx \pi/2$ to avoid amplifying measurement noise. An alternative is to reject the DC component at the mixer input by a high-pass filter $(z-1)/(z-p)$ where p is a real number slightly smaller than one[6]. This filter is used for the LLRF systems at LCLS-II where nominally $p = 15/16$. The filter corresponds to an IIR notch filter in the baseband.

Magnitude responses of the filters $H_{\mathrm{DCR}}(z)$ and $H_{2\mathrm{S}}(z)$ are shown in Figure 12.7.

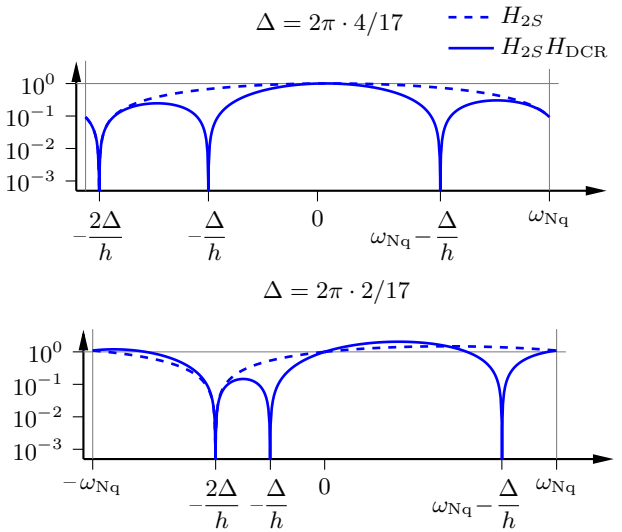


Figure 12.7 Magnitude responses for the two-sample reconstruction filter (12.5) for near-IQ ratios of 4/17 and 2/17; with DC rejection filters (12.6) (solid lines) and without (dashed lines). Note that $\sin 2\pi/17 \approx 0.36$, $\sin 4\pi/17 \approx 0.67$, $\sin 8\pi/17 \approx 1.00$.

[6] Personal communication with Larry Doolittle, LBNL.

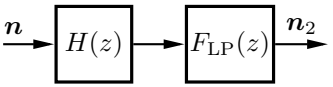


Figure 12.8 An additional low-pass filter $F_{\mathrm{LP}}(z)$ can be used to achieve acceptable levels of filtered noise $\boldsymbol{n}_2$ if the noise rejection of the downconversion filter $H(z)$ is insufficient.

12.3 Additional low-pass filtering

Typically, the noise rejection from the filters $H_*(z)$ in the preceding subsections is not sufficient and additional low-pass filtering by a filter $F_{\mathrm{LP}}(z)$ is necessary. Such a filter $F_{\mathrm{LP}}(z)$ is often considered and designed as part of the controller. In this section we will analyze two aspects related to additional low-pass filtering. To simplify the analysis we will study the interconnection in Figure 12.8 rather than a full feedback loop. We will assume the noise $\boldsymbol{n}$ to be white and that $F_{\mathrm{LP}}(z)$ is a first-order low-pass filter with bandwidth ω_{LP},

$$F_{\mathrm{LP}}(z) = \frac{1-a}{1-az^{-1}}, \qquad a = \mathrm{e}^{-\omega_{\mathrm{LP}}h}.$$

The quantity that we will be interested in is the rms level of the filtered noise $\boldsymbol{n}_2$. We can compute this quantity as the 2-norm of the impulse response of $H(z)F_{\mathrm{LP}}(z)$. We will assume a sampling period of $h = 10\,\mathrm{ns}$ which is typical for current-generation LLRF systems.

Noise rejection vs. phase-margin reduction

The moving average filter (12.2) has a longer latency than two-sample reconstruction (12.5) but it rejects more high-frequency noise which is helpful in reducing the control signal activity. A comparison of the noise rejection relative to the phase drop at typical cross-over frequencies for two moving average filters of different lengths and two-sample reconstruction (for a specific ratio M/N) is shown in Figure 12.9. It is seen that two-sample reconstruction comes out on top but that the performance difference between the two approaches is small.

Order of low-pass filtering and decimation

If decimation is used, a natural question is whether to implement the filter $F_{\mathrm{LP}}(z)$ before or after the decimation operation, see Figure 12.10. Implementing the filter at the lower rate facilitates the implementation but gives more aliasing. However, due to the nice folding properties of the filters (12.2) and (12.5) (if $\sin\Delta \approx 1$) the increase in aliasing is not as severe as one would expect.

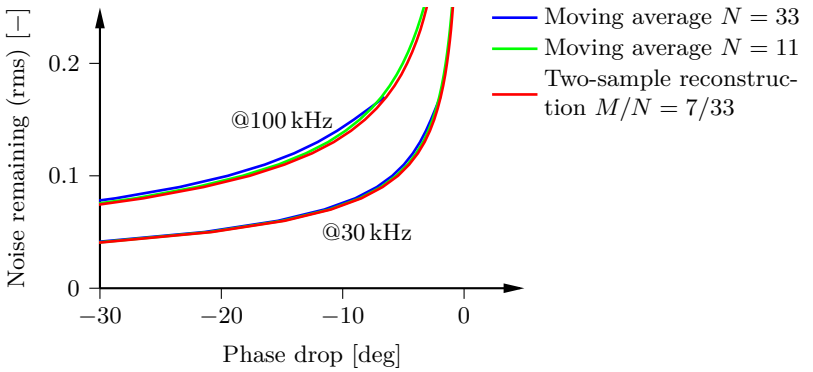


Figure 12.9 Comparison of the noise rejection as a function of phase drop for three different DDC filters in series with a first-order low-pass filter $F_{\mathrm{LP}}(z)$ (Figure 12.8). The bandwidth of the low-pass filter was swept over a wide range of frequencies to generate the plot. The phase drop at two different frequencies are considered. This figure shows that even if two-sample reconstruction has a shorter delay than a moving average filter the difference becomes small if noise rejection is desired. Nevertheless there is a slight performance advantage to using two-sample reconstruction.

(a) Decimation after $F_{\mathrm{LP}}(z)$.

(b) Decimation before $F_{\mathrm{LP}}(\mathrm{z})$.

Figure 12.10 Two alternatives for introducing additional low-pass filtering with a low-pass filter $F_{\mathrm{LP}}(z)$ (note that the filter coefficients differ between (a) and (b) due to the different sampling rates). The approach in (a) is advantageous in that it reduces aliasing from the decimation but it requires the filter to run at a higher rate than in (b).

Figures 12.11 and 12.12 show comparisons of how the order of filtering and decimation affects the resulting noise level as a function of the bandwidth ω_{LP} of the low-pass filter.

In the case of moving average filters, we see from Figure 12.11 that the order of low-pass filtering and decimation has little impact on the noise rejection if the bandwidth ω_{LP} is lower than 100 kHz. For higher bandwidths ω_{LP} there is a significant difference for the longer moving average filters.

In the case of two-sample reconstruction, we see from Figure 12.12 that it is crucial to have the ratio $M/N \approx 4$ if decimation is used before low-pass filtering.

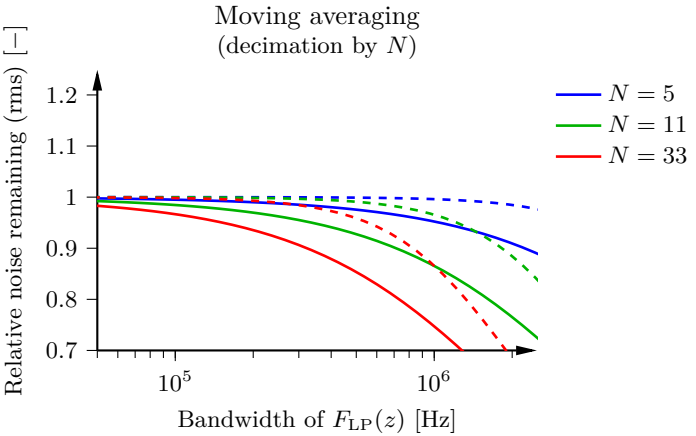


Figure 12.11 A comparison of the remaining noise for decimation *before* low-pass filtering (dashed, Figure 12.10b) and decimation *after* low-pass filtering (solid, Figure 12.10a), when using an N-sample moving average filter in series with a first order low-pass filter $F_{\mathrm{LP}}(z)$. The results are shown relative to only using $F_{\mathrm{LP}}(z)$. Three different lengths of the moving average filter and a range of bandwidths of the filter $F_{\mathrm{LP}}(z)$ were considered. The difference in noise rejection becomes apparent when the bandwidth of the filter $F_{\mathrm{LP}}(z)$ is high and the moving average filter is long.

12.4 Summary

This chapter has discussed low-latency digital downconversion for cavity field control. The requirements on digital downconversion for this application were seen to be quite different from those for communications applications. We considered two approaches for low-latency rejection of the double-frequency component at the output of the complex mixer: moving averaging and two-sample reconstruction.

With moving averaging there is no need for additional filters to reject the DC-offset spur. If decimation is used there is little need to worry about aliasing due to its nice folding properties. For two sample reconstruction it is crucial that the ratio of the intermediate frequency and the sampling frequency is approximately 1/4 to avoid excessive noise amplification. Their performance was similar in terms of noise rejection relative to phase drop around typical cross-over frequencies, with two-sample reconstruction performing slightly better. Taking these observations into account the moving average filter is the simple and care-free option. Advanced users may go with two-sample reconstruction. In particular if low-latency is more important than noise rejection.

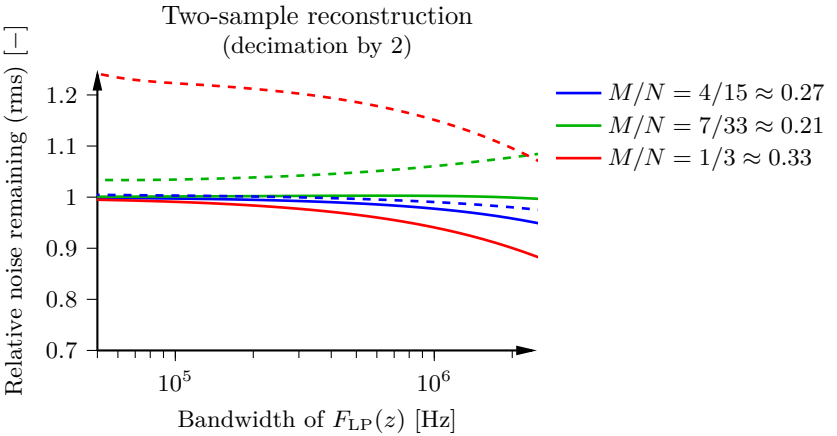


Figure 12.12 The same comparison as in Figure 12.11 but for the case of two-sample reconstruction, dashed lines correspond to decimation *before* low-pass filtering (Figure 12.10b) and solid lines corresponds to decimation *after* low-pass filtering (Figure 12.10a). If M/N is not close to 1/4 there is a significant advantage in low-pass filtering before decimation.

An alternative to these two filters could be to use a first-order IIR notch filter to reject the double frequency component. This avoids both the delay from a long moving average filter and the constraint $\Delta \approx \pi/2$ for two-sample reconstruction. However, implementing an IIR filter at the sampling rate of the ADC could be technically challenging.

13

Energy-Optimal Cavity Filling

13.1 Introduction

For pulsed linacs, the energy required to *fill* the cavities (build up the electromagnetic fields) is significant but does not contribute to particle acceleration. As remarked in [Bhattacharyya et al., 2015], little work has been done on energy-optimal filling of accelerating cavities. In that article it was also analytically derived how to minimize the energy *reflected* from a cavity during filling, which corresponds to minimization of the wall-plug energy consumption for *ideal amplifiers*. This chapter, which is based on [Troeng and Bernhardsson, 2017], shows how to minimize the wall-plug energy consumption for *arbitrary* power-consumption characteristics. We will also discuss energy-optimal filling subject to non-zero time-varying detuning $\Delta\omega(t)$.

In Section 13.2 we formulate an optimal-control problem for the RF drive that minimization the fill energy. In Section 13.3 we solve the optimal-control problem by decoupling it into two problems: one for the phase (corresponding to [Ayvazyan et al., 2010]) and one for the amplitude. In Section 13.4 we compare the energy-optimal filling strategies for different power-consumption characteristics and different amplifier saturation levels. We conclude with a remark on cryogenic losses and a discussion of the results.

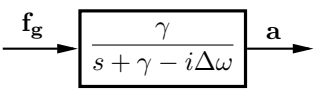


Figure 13.1 This chapter considers energy-optimal cavity filling, i.e., finding the RF-drive profile $\mathbf{f_g}(t)$ that drives the mode amplitude $\mathbf{a}(t)$ of the accelerating mode to its nominal value $\mathbf{a_0} = 1$ at final time t_f, while minimizing the energy consumption $\int_0^{t_f} p_{\text{amp}}(|\mathbf{f_g}(t)|)dt$.

Remark 13.1 The optimization of the RF-drive amplitude is only of interest for amplifiers whose power consumption depends on the output power, such as solid-state amplifiers and inductive output tubes (IOTs). Klystrons, which is the most common RF amplifier for high-energy accelerators, have constant power consumption and the energy-optimal filling strategy is simply to minimize the fill time by driving them at saturation.

The investigations of this chapter were motivated by that the 84 high-β cavities of the ESS linac would be powered by multi-beam inductive output tubes. However, as mentioned in Footnote 5 on page 71, the plan has changed to use klystrons instead.

13.2 Problem formulation

We will compare the wall-plug energy consumption for different filling approaches *relative* to each other. Hence, we may use the normalized dynamics of the accelerating mode (6.8), where we put $\mathbf{i_b} = 0$ since the beam pulse starts after the filling. To further simplify the exposition, we also scale time by the cavity time constant γ which gives

$$\frac{d\mathbf{a}}{dt} = (-1 + i\Delta\omega(t))\,\mathbf{a} + \mathbf{f_g}, \tag{13.1}$$

where the time-scaled detuning $\Delta\omega$ is redefined to include a factor $1/\gamma$.

We want to determine the amplifier drive $\mathbf{f_g}(t)$ that fills the cavity, i.e., achieves $\mathbf{a}(t_f) = 1$ at some final time t_f, while minimizing the normalized energy consumption

$$\mathrm{w} = \int_0^{t_f} p_{\mathrm{amp}}(|\mathbf{f_g}|)\,dt, \tag{13.2}$$

where $p_{\mathrm{amp}}(\mathrm{f_g})$ is the normalized power-consumption as a function of RF-drive amplitude. The final time t_f will be considered as a free parameter and there is an upper limit $\mathrm{f_{amp\,sat}}$ on $|\mathbf{f_g}|$, see Figure 13.2 for an illustration.

We may now state the optimal-control problem as

$$\underset{\mathbf{f_g},\,t_f}{\text{minimize}} \quad \int_0^{t_f} p_{\mathrm{amp}}(|\mathbf{f_g}|)\,dt \tag{13.3a}$$

$$\text{subject to} \quad \dot{\mathbf{a}} = (-1 + i\Delta\omega(t))\mathbf{a} + \mathbf{f_g} \tag{13.3b}$$

$$|\mathbf{f_g}| \le \mathrm{f_{amp\,sat}} \tag{13.3c}$$

$$\mathbf{a}(0) = 0 \tag{13.3d}$$

$$\mathbf{a}(t_f) = 1. \tag{13.3e}$$

In the problem formulation (13.3), we have assumed the detuning $\Delta\omega(t)$ for $0 \le t \le t_f$ to be known in advance, that the amplifier has no dynamics,

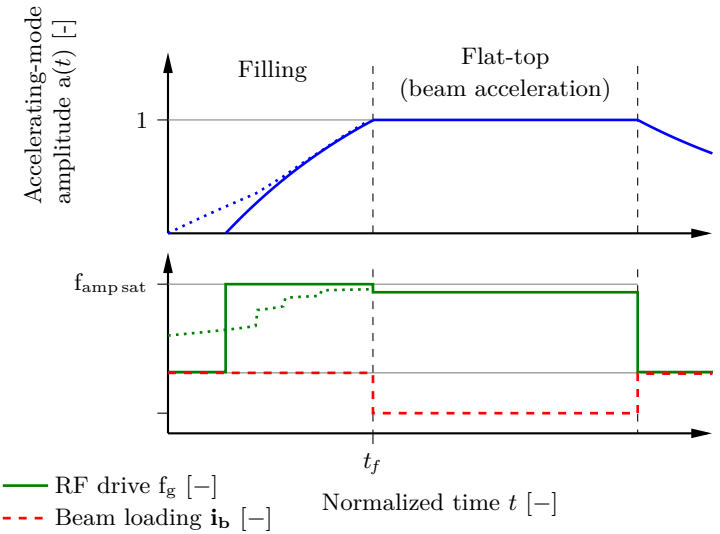


Figure 13.2 Normalized amplitudes of the cavity field and the RF drive during an RF pulse, for minimum-time filling, i.e., $|\mathbf{f_g}| = \mathrm{f_{amp\,sat}}$ (solid lines), and energy-optimal filling for an inductive output tube (dotted lines).

that there are no disturbances, and that the cavity parameters are perfectly known. These assumptions are approximations, and the computed results should be seen as an indication of what can be achieved in practice.

13.3 RF drive for energy-optimal filling

The cavity equation in polar coordinates

In polar coordinates the cavity equation (13.3b) takes the form [Brandt, 2007, (3.22)],

$$\mathrm{a}\dot{\phi}_\mathrm{a} - \Delta\omega \mathrm{a} = \mathrm{f_g}\sin(\phi_\mathrm{g} - \phi_\mathrm{a}) \tag{13.4a}$$

$$\dot{\mathrm{a}} + \mathrm{a} = \mathrm{f_g}\cos(\phi_\mathrm{g} - \phi_\mathrm{a}), \tag{13.4b}$$

where $\mathrm{a} \geq 0$, ϕ_a, $\mathrm{f_g} \geq 0$, and ϕ_g are defined via

$$\mathbf{a}(t) = \mathrm{a}(t)\mathrm{e}^{i\phi_\mathrm{a}(t)} \tag{13.5}$$

$$\mathbf{f_g}(t) = \mathrm{f_g}(t)\mathrm{e}^{i\phi_\mathrm{g}(t)}. \tag{13.6}$$

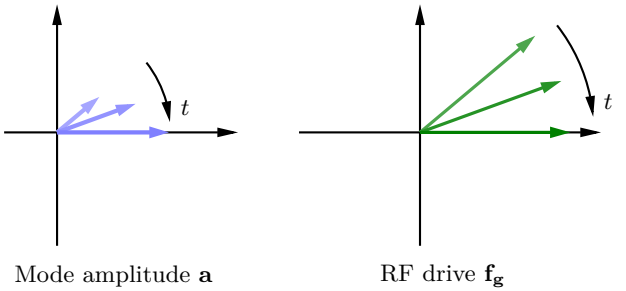


Figure 13.3 For energy-optimal cavity filling, the RF drive should have the same phase as the cavity field, see equations (13.7)–(13.8).

Optimal phase of the RF drive

By considering (13.4b), we see that choosing ϕ_{g} as

$$\phi_{\mathrm{g}}^{\star}(t) = \phi_{\mathrm{a}}(t), \tag{13.7}$$

maximizes $\dot{\mathrm{a}}$ for all $\mathrm{f_g} \geq 0$. Since the cost (13.3a) is independent of ϕ_{a} and we wish to minimize the cost for reaching $\mathrm{a}(t_f) = 1$, it is clear that (13.7) is optimal. With this choice of $\phi_{\mathrm{g}}(t)$, equation (13.4a) reduces to $\dot{\phi}_{\mathrm{a}} = \Delta\omega$; since we want $\phi_{\mathrm{a}}(t_f) = 0$, we must have

$$\phi_{\mathrm{a}}(t) = -\int_{t}^{t_f} \Delta\omega(t')\, dt'. \tag{13.8}$$

From (13.7) it follows that the optimal phase profile $\phi_{\mathrm{g}}^{\star}(t)$ of the RF drive equals the right-hand side of (13.8). This is the same conclusion that was reached in [Ayvazyan et al., 2010]. See Figure 13.3 for an illustration.

Remark 13.2 For superconducting cavities, the detuning $\Delta\omega(t)$ depends on the cavity field amplitude $\mathrm{a}(t)$ via the Lorentz force detuning. However, since the optimization of ϕ_{g} and $\mathrm{f_g}$ is decoupled, the optimal $\mathrm{f_g}(t)$ and corresponding $\mathrm{a}(t)$ can be found first, before finding the optimal phase via (13.8).

Optimal amplitude of the RF drive

The optimal phase of the RF drive is given by (13.7) so we only need to consider the amplitude dynamics given by (13.4b). This gives the following

optimization problem for the RF-drive amplitude $\mathrm{f_g}(t)$

$$\underset{\mathrm{f_g},\, t_f}{\text{minimize}} \qquad \int_0^{t_f} p_{\mathrm{amp}}(\mathrm{f}_g(t))\, dt \tag{13.9a}$$

$$\text{subject to} \qquad \dot{\mathrm{a}}(t) = -\mathrm{a}(t) + \mathrm{f_g}(t) \tag{13.9b}$$

$$|\mathrm{f_g}(t)| \leq \mathrm{f_{amp\,sat}} \tag{13.9c}$$

$$\mathrm{a}(0) = 0 \tag{13.9d}$$

$$\mathrm{a}(t_f) = 1. \tag{13.9e}$$

From (13.9), it seems reasonable that the optimal $\mathrm{f_g}(t)$ should, at each time instant, maximize the ratio between the increase of the cavity field and the power consumption, i.e.,

$$\mathrm{f_g^\star}(t) = \underset{\mathrm{f_g}}{\operatorname{argmax}} \frac{-\mathrm{a}(t) + \mathrm{f_g}}{p_{\mathrm{amp}}(\mathrm{f_g})}. \tag{13.10}$$

That (13.10) indeed is optimal follows from the Hamilton–Jacobi–Bellman equation. See Appendix G.3 for a self-contained proof and some comments.

13.4 Comparison of filling strategies

In this section we compare the energy consumption for three filling strategies,

- *Minimum time*, i.e., $\mathrm{f_g}(t) = \mathrm{f_{amp\,sat}}$
- *Minimum reflection*, i.e., $\mathrm{f_g}(t) = \min(\mathrm{e}^t / \sinh t_f,\, \mathrm{f_{amp\,sat}})$ with $t_f = 2$ [Bhattacharyya et al., 2015]
- *Energy optimal*, according to (13.10),

considering typical efficiency characteristics $\eta(\mathrm{f_g}) \coloneqq \mathrm{f_g^2}/p_{\mathrm{amp}}(\mathrm{f_g})$ of four different RF amplifier types [Bhattacharyya et al., 2015], see Figure 13.4,

- Tetrode,
- Solid-state amplifier (SSA),
- Inductive output tube (IOT),
- Constant efficiency (ideal) amplifier,

and two saturation levels,

- $\mathrm{f_{amp\,sat}} = 1.5$,
- $\mathrm{f_{amp\,sat}} = 2.25$.

The normalized saturation levels correspond to a normal conducting cavity and an optimally coupled superconducting cavity, respectively.

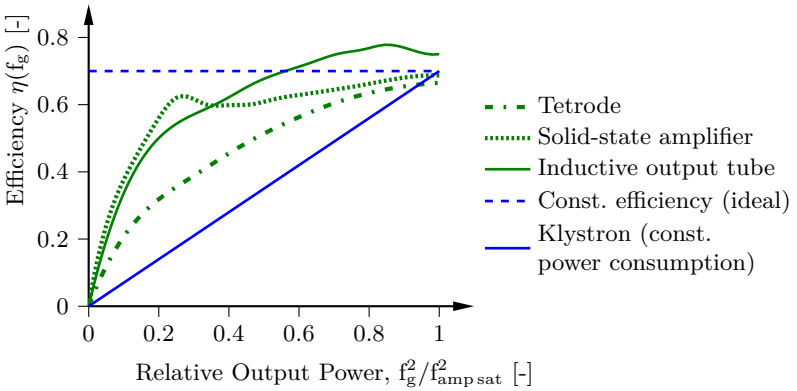


Figure 13.4 Efficiency $\eta(f_g) = f_g^2/p_{amp}(f_g)$ as a function of the relative output power $(f_g/f_{amp\,sat})^2$, for different amplifiers types. The data is from [Bhattacharyya et al., 2015, Fig. 9], and has been slightly smoothed.

The energy consumption for the different filling strategies and parameter combinations are shown in Figure 13.5. The relative energy consumption of energy-optimal filling compared to minimum-time and minimum-reflection filling is given in Table 13.1. The corresponding profiles for the mode amplitude a and RF drive f_g are shown in Figure 13.6.

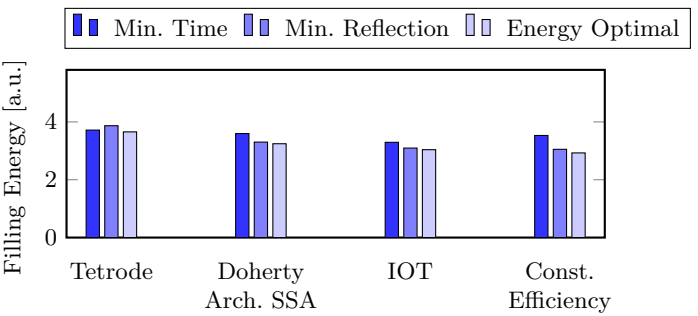
Remark 13.3 If multi-beam IOTs would have been used for the high-β section ($f_{amp\,sat} = 2.25$) of the ESS linac the energy reduction from using energy-optimal filling, relative to minimum-time filling, would be 13% which corresponds[1] to 17 k€/year.

Remark 13.4 (Cryogenic losses) From Figure 13.6 it is seen that the energy-optimal filling profiles take longer time than those for minimum-time filling. This implies increased RF heating of the cavities and a higher load on the cryogenic system. Cryogenic losses can be incorporated into the optimization problem (13.9) by adding a term $k_{cryo}a(t)^2$ to the integrand in (13.9a). Experiments of including cryogenic losses gave similar, but slightly faster, filling profiles.

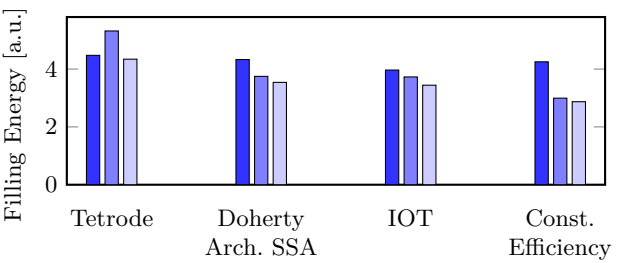
[1] The filling constitutes 0.2 ms/(2.86 + 0.2) ms = 6.5 % of the RF pulse and the average power consumption of the RF amplifiers of the high β section would have been 6.5 MW. This gives the cost estimate of the filling energy: 6 % × 6 MW × 5000 h/year × 0.07 €/kWh = 130 k€/year.

Table 13.1 Energy consumption of minimum-time (MT) and minimum-reflection (MR) filling relative energy-optimal (EO) filling.

Amplifier Type	$f_{amp\,sat} = 1.5$		$f_{amp\,sat} = 2.25$	
	$\frac{w_{EO}}{w_{MT}}$	$\frac{w_{EO}}{w_{MR}}$	$\frac{w_{EO}}{w_{MT}}$	$\frac{w_{EO}}{w_{MR}}$
Tetrode	98 %	94 %	97 %	82 %
Doherty Arch. SSA	90 %	98 %	82 %	94 %
IOT	92 %	98 %	87 %	92 %
Constant Efficiency	83 %	96 %	68 %	96 %

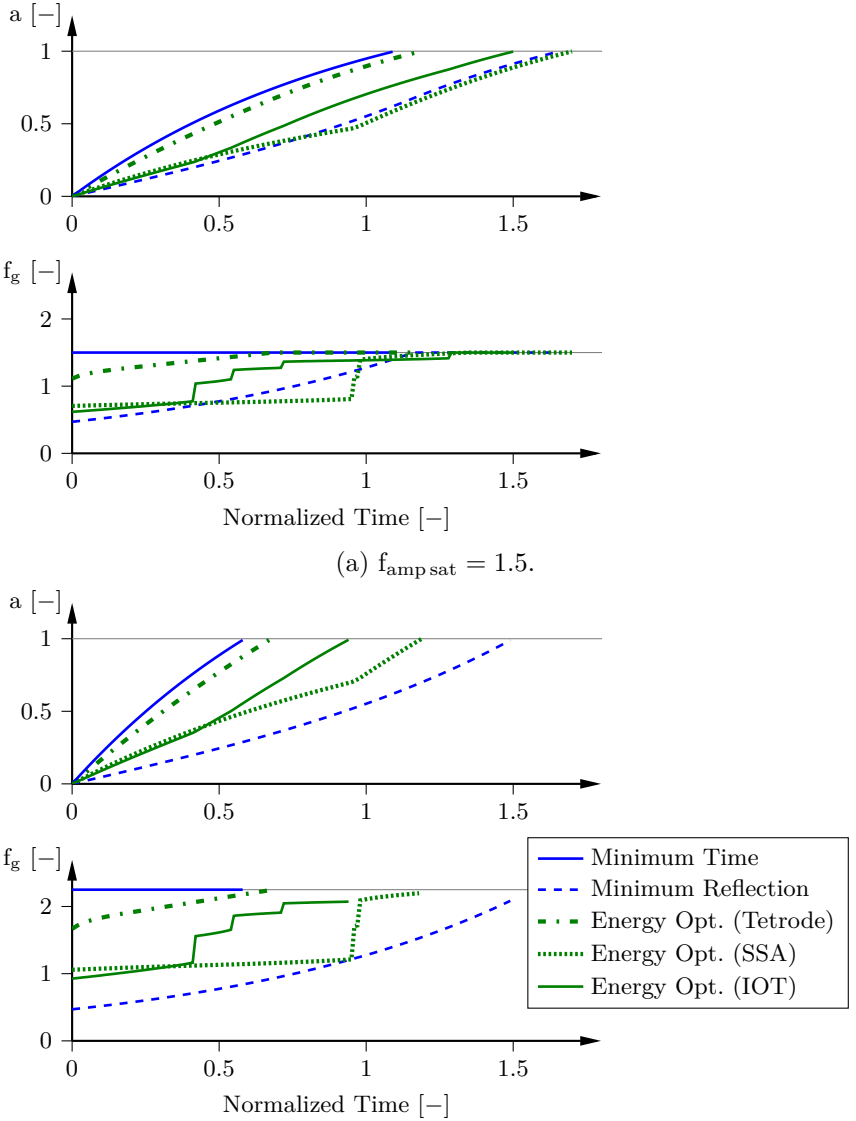


(a) $f_{amp\,sat} = 1.5$.



(b) $f_{amp\,sat} = 2.25$.

Figure 13.5 Comparison of the energy consumption for minimum-time filling (standard approach), minimum-reflection filling according to [Bhattacharyya et al., 2015], and energy-optimal filling according to (13.10). Four different amplifier types and two different saturation levels are considered.



(a) $f_{amp\,sat} = 1.5$.

(b) $f_{amp\,sat} = 2.25$.

Figure 13.6 Accelerating-mode amplitude and RF drive for different filling strategies: minimum-time, minimum-reflection (↔ constant efficiency amplifier) and energy-optimal for the efficiency characteristics in Figure 13.4. Note that minimum-time and minimum-reflection filling are independent of the amplifier efficiency characteristics. Two saturation levels are considered.

13.5 Conclusions

We have shown how to minimize the energy consumption for establishing the electromagnetic field in accelerating cavities. We proved that the amplitude and phase for the optimal RF-drive profile (normalized with respect to the cavity time constant) are given by

$$\mathrm{f}_\mathrm{g}^\star(t) = \operatorname*{argmax}_{\mathrm{f_g}} \frac{-\mathrm{a}(t) + \mathrm{f_g}}{p_\mathrm{amp}(\mathrm{f_g})}$$

$$\phi_\mathrm{g}^\star(t) = -\int_t^{t_f} \Delta\omega(t)\, dt.$$

We compared the energy savings for different amplifier characteristics and found that the energy consumption could be reduced by up to 30%. If inductive output tubes would have been used for the high-β section of the European Spallation Source we estimated the yearly savings could be 17 k€. Since the energy-optimal filling profiles, or at least approximate versions thereof, are relatively easy to implement, they provide a straight-forward way to reduce the operating costs and environmental footprint of pulsed particle accelerators.

14

Ripple-Rejecting Peak Filters

14.1 Introduction

The PI(D) controller is an attractive choice when the plant has simple dynamics (e.g., accelerating cavities) and the load disturbances have low-frequency character (see Chapter 9). However, the PI(D) controller is outperformed by more advanced controllers in the case of narrowband load disturbances. This shortcoming of the PID controller can be understood from Francis' internal model principle [Francis and Wonham, 1976], which states that the controller needs to contain a model of any disturbance that it should reject; the PID controller is *too* simplistic in this regard.

By augmenting a nominal (e.g., PID) controller with a second-order peak filter it is possible to selectively increase the controller gain and reject disturbances with a specific frequency (see Figure 14.1). This approach gives only a small increase in controller complexity. The filter can be seen as a model of the narrowband disturbance; see [Åström and Hägglund, 2006] for

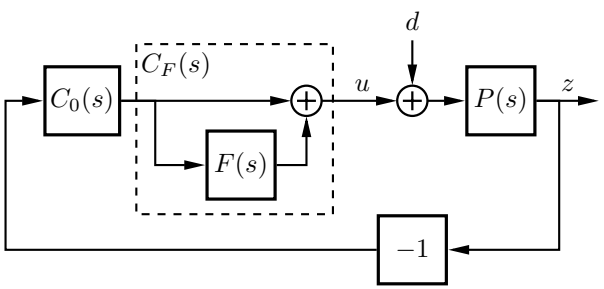


Figure 14.1 A nominal controller $C_0(s)$ can be augmented with a peak filter $C_F(s)=1+F(s)$ for improved rejection of narrowband disturbances d.

an insightful interpretation. Peak filters have been used for improved disturbance rejection in hard-disk-drive track following [Kim et al., 2005; Zheng et al., 2006; Atsumi et al., 2007; Xu et al., 2012] and several other real-world applications [Sievers and Flotow, 1992; Han et al., 2008; Castilla et al., 2009].

Although the basic idea is straightforward it is crucial that the filter has the proper phase at the peak frequency to avoid robustness degradation or closed-loop instability. This critical detail has received surprisingly little attention in the control literature. We are only aware of [Sievers and Flotow, 1992] and [Kim et al., 2005] which provide qualitative discussions based on root-locus arguments, and [Zheng et al., 2006] that derives suitable filter coefficients by considering a transformed system.

This chapter discusses the approach to peak-filter design that was introduced in [Troeng and Bernhardsson, 2018]. The approach is based on consideration of the open-loop Nyquist curve, which makes the design trade-offs easier to understand than in previous approaches. The most natural filter design, based on Nyquist diagram consideration, recovers the design in [Zheng et al., 2006]. However, the flexibility of our approach makes it easy to consider other performance metrics as well, and it also readily extends to complex-coefficient systems. We will also discuss the convergence time of the filter, and provide practical implementation details.

The work in this chapter was motivated by early simulations of the klystron power supplies for ESS[1] that indicated significant ripple at 90 kHz. This ripple frequency falls within the range where field control loops are the most sensitive to disturbances (see Chapter 9). Simulations indicated that it would not be possible to achieve the field-error requirement for the normal conducting cavities at ESS (0.2 %/0.2°) with only PI(D) control.

14.2 Intuitive design of disturbance-rejecting peak filters

Assume that a single-input single-output plant $P(s)$ and a controller $C(s)$ are given. We will assume that they both have real coefficients and operate on real signals; this is done in order to show that this approach has applications outside of cavity field control. The extension to complex-coefficient systems, such as field control loops, are straight-forward and are discussed at the end of this section.

Recall from Section 2.2 that the magnitude of the sensitivity function

$$S(i\omega) = \frac{1}{1 + P(i\omega)C(i\omega)}$$

[1] A single time series from a Simulink simulation provided by the power-converter group at ESS in November 2015. It was also stated that the ripple would not be synchronized with the beam pulse (i.e., it would not be rejected by iterative learning control).

quantifies the closed-loop robustness and how the feedback from $C(s)$ affects the disturbance rejection.

We will consider how the disturbance rejection around a specific frequency ω_0 can be improved by augmenting a nominal controller (e.g., a PID controller) C_0 with a second-order filter

$$C_F(s) = \frac{s^2 + 2\zeta_z\omega_z s + \omega_z^2}{s^2 + 2\zeta_0\omega_0 s + \omega_0^2}, \tag{14.1}$$

where $\omega_z \approx \omega_0$, and $\zeta_z > \zeta_0$. We will refer to $C_F(s)$ as a *peak filter*, since this captures the feature that is relevant to us; the terms resonant filter and bandpass filter are also common.

If ω_z is chosen equal to ω_0, we have $C_F(0) = 1$, $C_F(\pm\infty) = 1$, and $C_F(i\omega_0) = \zeta_z/\zeta_0$, enabling the gain of the augmented controller $C = C_0C_F$ to be selectively increased at the disturbance frequency ω_0. For maintained robustness, and closed-loop stability, also the phase of the open-loop system

$$L(s) = P(s)C_0(s)C_F(s)$$

around ω_0 needs to be considered.

Choosing ω_z different from ω_0 gives two degrees of freedom, ω_z and ζ_z, making it possible to place $L(i\omega_0)$ arbitrarily in the complex plane. We will see that this allows robustness and stability to be maintained.

To clarify the behavior of the filter C_F, we introduce the parametrization

$$C_F(s) = 1 + F(s) \tag{14.2}$$

with

$$F(s) \coloneqq K\frac{2\Delta(s\cos\alpha - \omega_0\sin\alpha)}{s^2 + 2\Delta s + \omega_0^2}, \tag{14.3}$$

where K and α are new parameters and $\Delta = \zeta_0\omega_0$ is the width of the peak. Note that $F(\pm i\omega_0) = K\mathrm{e}^{\pm i\alpha}$. The parametrization (14.2)–(14.3) is inspired by the one in [Zheng et al., 2006], but we have modified it slightly to simplify the exposition[2]. See Figure 14.1 for an illustration of the controller structure and Figure 14.2 for a typical pole–zero diagram of $C_F(s)$.

[2]The definition $F_{\text{zheng}}(s) \coloneqq Ks(\omega_0\cos\alpha - s\sin\alpha)/(s^2 + 2\zeta_0\omega_0 s + \omega_0^2)$ in [Zheng et al., 2006] gives that $F_{\text{zheng}}(i\omega_0) = K\mathrm{e}^{-i\alpha}/(2\zeta_0)$; the ζ_0 dependence of peak magnitude and the minus sign in front of α makes this parameterization slightly less intuitive.

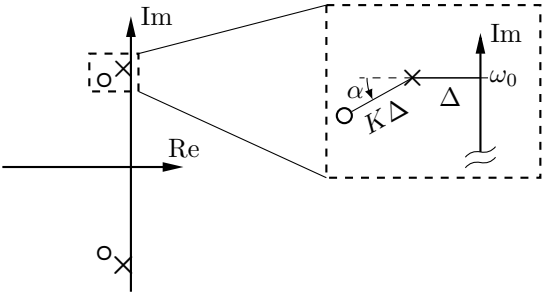


Figure 14.2 Typical pole–zero diagram for the peak filter $C_F(s) = 1 + F(s)$ with $F(s)$ given by (14.2). Having different imaginary parts for the poles and zeros gives freedom to select the phase of $C_F(s)$ at ω_0. In the limit of $\Delta \to 0$ the poles are given by $p_{1,2} = -\Delta \pm i\omega_0$ and the zeros by $z_{1,2} = p_{1,2} - K\Delta e^{\pm i\alpha}$.

14.2.1 Nyquist bubbles

Instead of considering the second-order filter (14.3) we simplify the situation even further and introduce the first-order (complex-coefficient) filter

$$\begin{aligned} B(s) &= K\mathrm{e}^{i\alpha}\left(\frac{1}{2} + \frac{1}{2}\frac{1-(s-i\omega_0)/\Delta}{1+(s-i\omega_0)/\Delta}\right) \\ &= K\mathrm{e}^{i\alpha}\frac{\Delta}{s-i\omega_0+\Delta}. \end{aligned} \tag{14.4}$$

For small Δ we now have

$$F(s) \approx B(s) + B^*(s), \tag{14.5}$$

where $B^*(s)$ is the complex-conjugate transfer function defined in (4.10). Equality holds in the limit[3] $\Delta \to 0$. Note that the right-hand side in (14.5) has real coefficients even if $B(s)$ does not.

The Nyquist curve of $B(s)$ in (14.4) is shown in Figure 14.3. It is seen to be a circle/bubble with diameter K that passes through the origin and is directed at an angle α. The frequency ω_0 is mapped to the point on the circle that is furthest away from the origin and the frequencies $\omega_0 \pm \Delta$ are mapped to the points on the circle that are exactly between $B(i\omega_0)$ and the origin[4].

[3]The exact expression is given by

$$B(s) + B^*(s) = K\frac{2\Delta\,(s\cos\alpha - \omega_0\sin\alpha + \Delta\cos\alpha)}{s^2 + 2\Delta s + \omega_0^2 + \Delta^2}.$$

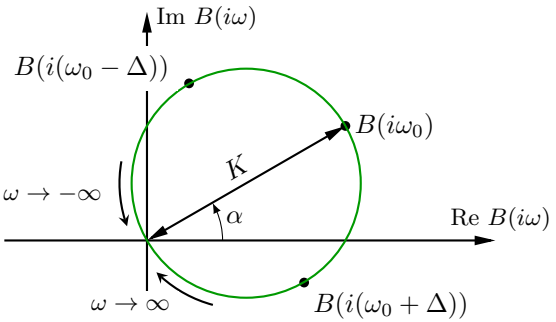


Figure 14.3 The Nyquist curve of the first-order system (14.4).

Remark 14.1 From (14.4) it is seen that the frequency response of the Nyquist bubble $B(s)$ goes to zero at high frequencies and that $B(0) \neq 0$. If the peak is inserted at a frequency above or close to the open-loop cross-over frequency it might be better to use a peak filter of the form

$$\widetilde{B}(s) = B(s)\frac{s}{i\omega_0}, \tag{14.6}$$

which distorts the nominal system less in the critical cross-over region.

14.2.2 Nyquist diagram interpretation

For studying control performance and robustness it often gives more insight to consider the Nyquist curve of the open-loop system $L = PC$ than the sensitivity function $S = 1/(1 + L)$. The distance between a point $L(i\omega)$ on the open-loop Nyquist curve and the point -1 equals $1/|S(i\omega)|$, but the Nyquist curve also shows useful phase information.

To understand the effect of the peak filter, let $L_0 = PC_0$ be the nominal open-loop system and let $L = L_0C_F = L_0(1 + F)$ be the augmented system. If the peak filter is narrow, we have that

$$L(s) \approx L_0(s) + L_0(i\omega_0)F(s). \tag{14.7}$$

From the discussion in the previous section it is clear that the Nyquist curve of the augmented system is similar to the nominal one but with two circular bubbles at frequencies $\pm\omega_0$ due to the term $L_0(i\omega_0)F(s)$. The bubble at

[4]Note that the linear fractional transformation $B_0(s) = \frac{1}{2} + \frac{1}{2}\frac{1-s}{1+s}$ maps the imaginary axis to a circle with radius $1/2$ centered around the point $1/2$. It is also seen that $B_0(0) = 1$, $|B_0(\pm 1)| = 1/\sqrt{2}$, and $|B_0(\pm\infty)| = 0$. Scaling $B_0(s)$ gives the middlemost expression of (14.4).

frequency ω_0 is oriented in the direction $\alpha' = \alpha + \angle L_0(i\omega_0)$ and have diameter $K' = K \cdot |L_0(i\omega_0)|$, see Figure 14.4 for an illustration.

14.2.3 Filter design for maximum sensitivity reduction

For a given value of K, it is clear that maximum sensitivity reduction at ω_0 is achieved by directing the bubbles away from the point -1. This corresponds to $\alpha' = \angle(L_0(i\omega_0) - (-1))$ which gives

$$\alpha = \angle\left(L_0(i\omega_0) - (-1)\right) - \angle L_0(i\omega_0) = -\angle \frac{L_0(i\omega_0)}{L_0(i\omega_0) + 1} = -\angle T_0(i\omega), \quad (14.8)$$

where $T_0 = L_0/(1 + L_0)$ is the complementary sensitivity function of the nominal system.

The choice (14.8) is the same as in [Zheng et al., 2006] but the Nyquist-bubble-based derivation in this section arguably provides more intuition.

It will be convenient to denote the nominal sensitivity function by $S_0(s) := 1/(1 + L_0(s))$. Recall that $1/|S_0(i\omega)|$ is the distance between the nominal open-loop Nyquist curve $L_0(i\omega)$ and the point -1. Choosing α according to (14.8) and K according to

$$K = \frac{N}{|S_0(i\omega_0)|} \frac{1}{|L_0(i\omega_0)|} = \frac{N}{|T_0(i\omega_0)|}, \quad (14.9)$$

increases the distance between Nyquist curve $L(i\omega)$ at ω_0 by a factor $(N+1)$, which gives a sensitivity reduction by the same factor at that frequency. The choices (14.8) and (14.9) are illustrated in Figure 14.4.

14.2.4 Closed-loop behavior

With the choices of α and N in (14.8)–(14.9), we can derive an approximate expression for how the filter affects the closed-loop sensitivity around ω_0. First, re-write the sensitivity function of the augmented system as

$$\begin{aligned} S(s) &= \frac{1}{1 + L_0(s)(1 + F(s))} \\ &= \frac{1}{(1 + L_0(s)) + L_0(s)F(s)} = S_0(s) \cdot \frac{1}{1 + T_0(s)F(s)}. \end{aligned} \quad (14.10)$$

Then, using (14.8)–(14.9) and that $T_0(s)F(s) \approx T_0(i\omega_0)F(s)$,

$$S(s) \approx S_0(s) \frac{s^2 + 2\zeta_0\omega_0 s + \omega_0^2}{s^2 + 2(N+1)\zeta_0\omega_0 s + \omega_0^2}. \quad (14.11)$$

For $N > 0$ the second factor (14.11) is a notch filter with a convergence time on the order of

$$T_F = 1/((N+1)\omega_0\zeta_0). \quad (14.12)$$

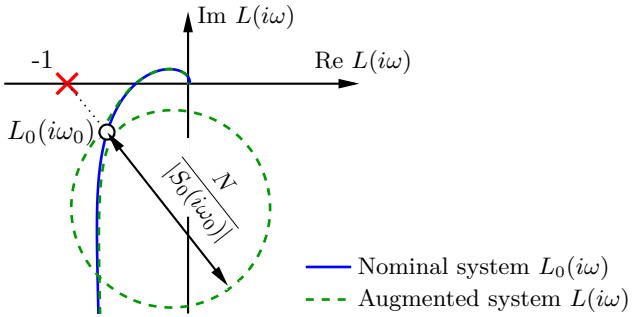


Figure 14.4 By augmenting a nominal controller with a peak filter it is possible to introduce a bubble in the open-loop Nyquist curve which allows the Nyquist curve at ω_0 to be moved to an arbitrary point. The figure illustrates the bubble when the filter parameters are chosen according to (14.8) and (14.9). With these choices the sensitivity is reduced by a factor $(N+1)$ at ω_0.

The convergence time T_F indicates how long it takes the peak filter to reject a narrowband disturbance.

Since $(N+1)\omega_0\zeta_0$ is the 3 dB bandwidth of the notch around ω_0, the same number also determines the degradation of the closed-loop performance at frequencies away from ω_0, due to the waterbed effect.

14.2.5 Implementation aspects

When the filter C_F is used for augmenting a PI(D) controller, it will probably be advantageous to place it before the PI(D) controller in order to reduce the impact on the anti-windup implementation.

A discrete-time counterpart to the filter (14.1), obtained through pole–zero matching, is given by

$$C_F(z) = \frac{(z - r_z e^{i\omega_z h})(z - r_z e^{-i\omega_z h})}{(z - r_0 e^{i\omega_0 h})(z - r_0 e^{-i\omega_0 h})} = \frac{z^2 - 2r_z \cos(\omega_z h) z + r_z^2}{z^2 - 2r_0 \cos(\omega_0 h) z + r_0^2},$$

where h is sample time, $r_0 = e^{-\zeta_0 \omega_0 h}$ and $r_z = e^{-\zeta_z \omega_z h}$.

14.2.6 Design guidelines

The filter $F(s)$ in (14.3) has four parameters which gives freedom to select

- the frequency of the peak,
- the phase of the peak,
- the magnitude of the peak, and

- the width of the peak (which affects settling time, robustness, and numerical stability).

One approach to select the parameters of the peak filter (14.3) is:

1. Select ω_0 as the disturbance frequency.
2. Select the desired sensitivity reduction N at ω_0, and put $K = N/|T_0(i\omega_0)|$.
3. Select Δ based on the spectral width of the disturbance, the required convergence time, and numerical implementation considerations.
4. Select α to achieve the best possible robustness, the typical choice would be according to (14.8).

Some comments on point 3) are in order. If it is considered more important to maintain a large amplitude or phase margin, it would make sense to choose α different than in (14.8). Another case where a different value of α should be considered, is when the bubbles are far away from their ideal circular shape. This is seen in Figure 14.10a, where a small increase of α would improve the closed-loop robustness. Note that the intuition from Section 14.2.1 makes these modifications obvious

14.2.7 Peak-filter design for complex-coefficient systems

In the case of complex-coefficient systems, the frequency response and disturbance spectrum might not be symmetric about the zero frequency. The filter $F(s)$ defined in (14.3) is not sufficiently flexible in this situation since the positive- and negative-frequency bubbles $B(s)$ and $B^*(s)$ have the same parameters (up to conjugation). In the complex-coefficient case one should instead take $F(s) := B_+(s) + B_-(s)$ which allows the parameters of the positive- and negative-frequency bubbles $B_+(s)$ and $B_-(s)$ to be chosen independently. If the disturbance only occurs at one side of the frequency spectrum then a single bubble is sufficient.

14.3 Example: Rejection of RF-drive ripple

The motivation for this chapter was that preliminary simulations[5] indicated that the power supplies for the klystrons at ESS would have significant 90 kHz voltage ripple, see Figures 14.5 and 14.6. These supply-voltage variations would change the output phase of the klystron (Section 5.5.3) and consequently act as a disturbance on the cavity fields. The design method presented in this chapter makes it easy to design disturbance-rejecting peak filters that reduce the impact of this disturbance.

[5]Simulink simulations performed by the Power Converter group at ESS in Nov. 2015.

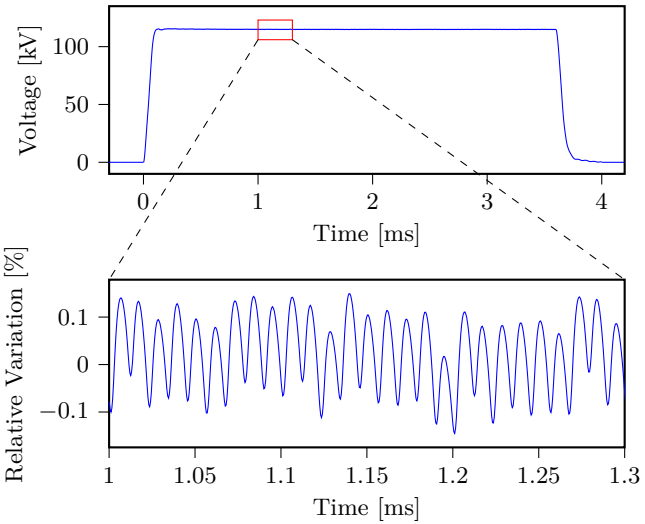


Figure 14.5 Simulated voltage pulse from the DC power supply that will be used for the ESS klystrons (see footnote on p. 176). The zoom reveals the presence of 90 kHz switching ripple.

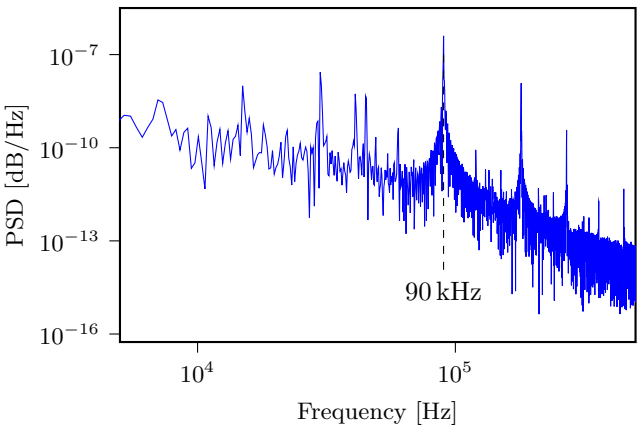


Figure 14.6 Estimated power spectral density of the DC supply ripple. The dominant harmonic at 90 kHz lies in the frequency range where the disturbance rejection of the cavity feedback loop is the worst.

Plant dynamics. We will model the field control with the following normalized baseband model from Chapter 9,

$$P(s) = \frac{\gamma_a}{s+\gamma_a} P_{\text{amp}}(s) \mathrm{e}^{-s\tau}, \tag{14.13}$$

where γ_a is the bandwidth of the accelerating cavity mode, P_{amp} is the amplifier dynamics, and $\tau = 1\,\mu\text{s}$ is the loop delay of the field control loop. In our numerical example we will consider three different values of the bandwidth γ_a which correspond to the ESS cavities in Table 6.3.

Impact of ripple. The voltage ripple d_{ps} of the DC power supply acts as a load disturbance given in (5.5.3) by

$$d(t) = \mathrm{f}_{\text{g0}} \xi_\theta d_{\text{ps}}(t),$$

where f_{g0} is the nominal RF drive (assuming optimal tuning real so that $\mathbf{f_{g0}}$ is real) and $\xi_\theta = 8^\circ/\%$ is a typical phase-pushing factor of the RF amplifier at ESS.

Nominal controller design. For each cavity type, a nominal PI controller C_0 was designed for rejection of low-frequency disturbances subject to the same constraints on the control signal activity ($B_{un} = 30$) as in Chapter 9. A slightly tighter constraints on sensitivity than in Chapter 9 was used ($S_{\max} = 1.5$) due to that there is a small sensitivity degradation from introducing the peak filter.

A typical transfer function from load disturbances d to control errors, when the cavity is controlled by a PI controller is shown in Figure 14.7. It is seen that the disturbance rejection is worst at around 50 kHz to 100 kHz, which is due to the fundamental limitations from the 1 µs loop delay.

Peak-filter design. The nominal controllers were augmented with peak filters designed according to (14.8) and (14.9). The same peak filter parameters $\omega_0/(2\pi) = 90\,\text{kHz}$, $\Delta/(2\pi) = 1\,\text{kHz}$, and $N = 5$ were used for all three cavity types.

Simulation results. The improvement in ripple rejection provided by the peak filters is shown in Table 14.1. The Nyquist plots for the different designs are shown in Figure 14.10. Note that the bubbles from the peak filters are oriented quite different, showing the advantage with an automatic design procedure.

The resulting closed-loop transfer function from disturbances to control errors for a DTL and different controllers are shown in Figure 14.7, and the corresponding power spectra are shown in Figure 14.8. As can be seen in Figure 14.7, there is significant improvement in rejection of the disturbance at 90 kHz, with small performance degradation for other frequencies. The

time response of a DTL controlled with and without a peak filter is shown in Figure 14.9. The Nyquist plots in Figure 14.10 show that the filters only give a small impact on robustness.

Table 14.1 Comparison of the remaining phase errors for PI controllers with and without peak filters. The numbers *only* include the impact of power-supply ripple—during operation there will be additional disturbances. Recall that the ripple data is from a preliminary simulation. Reliable measurements of the ripple of the production series power supply that will be used at ESS are not available at the time of writing.

	RFQ	**DTL**	**Medium-β**
Bandwidth $\gamma_a/(2\pi)$ [kHz]	60	12	0.5
f_{g0}	1.1	1.3	2
Phase error (rms)			
Requirement	0.20°	0.20°	0.10°
PI	0.35°	0.11°	0.01°
PI + peak filter	0.14°	0.05°	0.006°

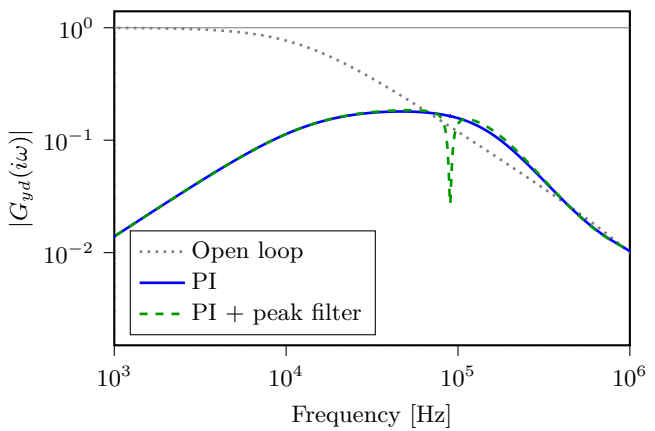


Figure 14.7 Transfer functions from load disturbances d to cavity field errors z for a DTL with different types of field control.

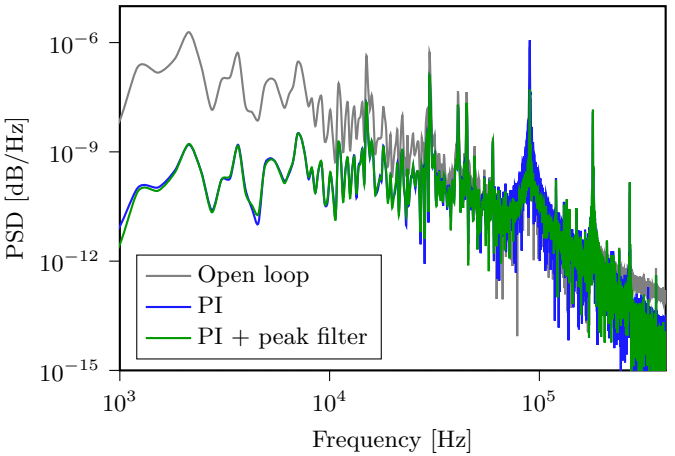


Figure 14.8 Simulated power spectral density of the field control errors of a DTL subjected to the DC supply ripple in Figure 14.6. It is seen that the ripple filter significantly reduces the impact of 90 kHz ripple.

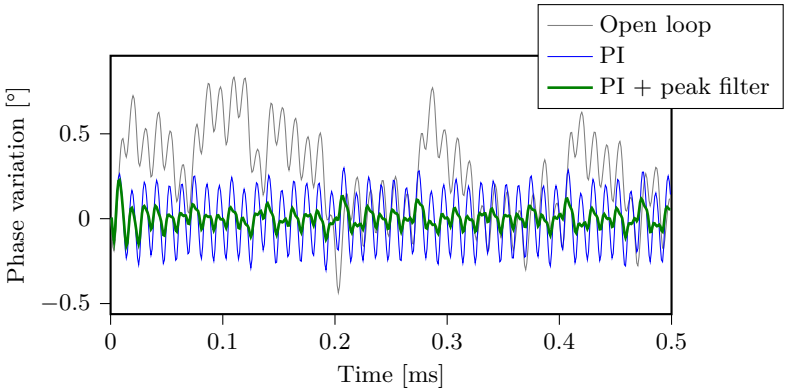
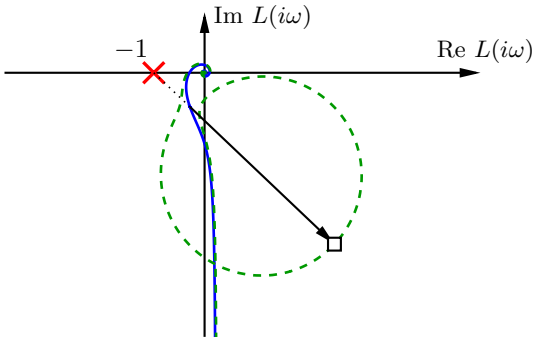
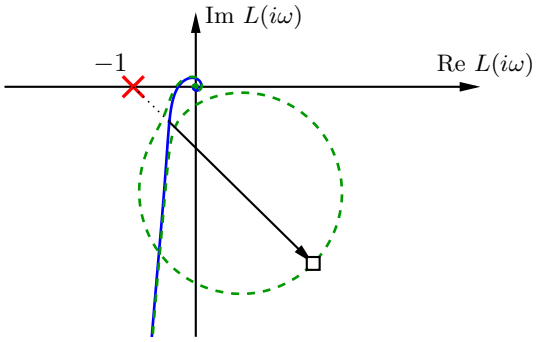


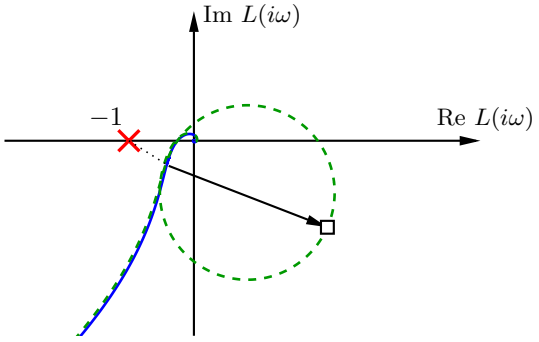
Figure 14.9 Field control errors for a DTL due to the voltage ripple in Figure 14.5 for three different types of control. Note the fast convergence rate of the controller with the peak filter which is consistent with the estimate $T_F \approx 20\,\mu s$ from (14.12). The convergence will mostly occur during the filling of the RF cavity, i.e., before the beam is turned on.



(a) Radio-frequency quadrupole.



(b) Drift-tube linac.



(c) Elliptical medium-β cavity.

Figure 14.10 Nyquist curves for three different field control loops. The blue solid lines correspond to nominal open-loop systems, and the green dashed lines to systems augmented with peak filters.

14.4 Summary and discussion

This chapter has discussed how to improve a nominal controller's ability to reject narrowband disturbances of a specific frequency. The idea is to augment the controller with a peak filter that selectively increases the controller gain at the disturbance frequency. The peak filter can be seen to add two bubbles to the nominal system's Nyquist curve. These bubbles need to have the correct phase to achieve disturbance rejection and avoid instability. This chapter proposed a parameterization of the peak filter that captures the behavior of the bubbles that it produces. This interpretation made it easy to understand how the peak filter parameters should be chosen for good disturbance rejection and maintained closed-loop robustness.

(End of Part II)

15

Conclusions and Future Work

15.1 Summary

This thesis has addressed amplitude and phase control of the accelerating electromagnetic cavity fields in linear particle accelerators. The topics covered were selected based on their relevance to the high-intensity proton linac at ESS, but the material is also relevant for other linacs.

Chapter 3 discussed the drivers of field stability for different linac types. Field-control challenges for some specific linacs were also mentioned.

Chapter 4 introduced how systems with narrowband dynamics, like accelerating cavities, can be modeled by complex-coefficient baseband systems. This representation is common in physics and telecommunications where it greatly simplifies understanding and analysis. However, this representation has seen little use for field control analysis and is relatively unheard of in the control community. Therefore we discussed how standard tools of automatic control generalize to complex-coefficient LTI systems. The extensions were mostly straight-forward and the main difference is in that negative frequencies must be considered.

Part I. The first part of the thesis addressed baseband modeling of the field control loop and essential aspects of cavity field control.

In Chapter 5 we provided baseband models of the components in the RF system. In Chapter 6 we discussed the dynamics of accelerating cavities which is a key to understand cavity field control. We proposed an energy-based parameterization of the cavity dynamics that gives more intuition and avoids certain issues with previous parameterizations. We also proposed a suitable normalization of the cavity dynamics. In Chapter 7 we collected the results from Chapters 5 and 6, proposing models of the field control loop with varying levels of complexity. We also discussed the different types of

disturbances that are relevant to field control. The modeling in all these chapters was greatly simplified by the complex coefficient representation.

Chapter 8 considered basic aspects of cavity field control and discussed the features of a good field controller. In this chapter it was also demonstrated how the complex-coefficient representation makes it possible to interpret the crucial aspect of loop-phase adjustment as a rotation of the Nyquist curve around the origin.

If only the accelerating cavity mode is considered the field control loop is essentially a first-order system with a time-delay. Assuming that the disturbance spectrum has low-pass character it was demonstrated in Chapter 9 that PI(D)-controllers achieve close to optimal control performance. Since PI(D)-controllers are intuitive and easy to implement we argued that they are a suitable choice for cavity field control.

Part II At the end of Part I it was shown that PI(D) control is suitable for cavity field control if parasitic modes can be neglected and the disturbances have low-pass character. If parasitic modes are present or if the disturbances have narrow spectra, one can retain the simple PID structure if it is augmented with notch filters (Chapter 11) and peak filters. A simple and intuitive tuning method for such disturbance-rejecting peak filters was introduced in Chapter 14.

Three additional topics were also discussed. Chapter 10 compared Cartesian and polar feedback for cavity field control in the context of a digital LLRF system and a linear accelerator. Low-latency digital downconversion for cavity field control was discussed and analyzed in Chapter 12. The problem of minimizing the energy consumption for filling the accelerating cavities at the start of each RF pulse was solved in Chapter 13.

15.2 Future work: Directionality in the field control problem

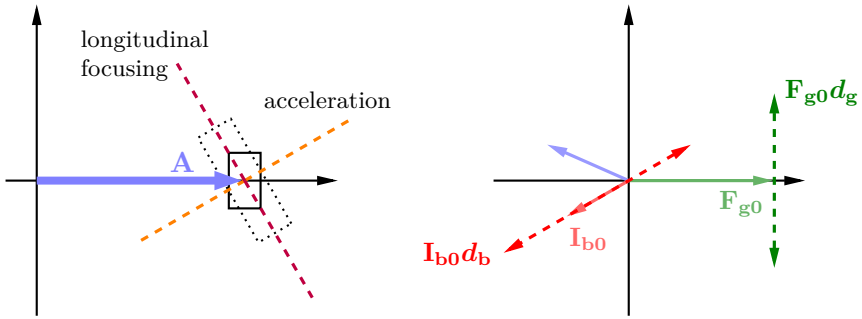
Throughout this thesis it has been assumed that the field control loop can be modeled as a single-input single-output complex-coefficient system $P(s)$ (if IQ imbalance and amplifier nonlinearity are compensated). This representation implies that the dynamics of the field control loop is rotationally invariant (see Section 4.3.5). However, as have been discussed in this thesis, the objective function, the disturbances distributions, and the control signal limitation (around a nominal operating point) are typically not rotationally invariant.

Directionality of disturbances and their impact. The disturbances acting on the field control loop have certain directionality, as was mentioned in Section 6.7. For example, the RF amplifier predominantly gives rise to phase variations of the RF drive. A robust field control loop should have $G(i\omega) \approx [G(-i\omega)]^*$ around the open-loop cross-over frequency where the disturbance sensitivity is the greatest (see Chapter 8.4). This implies that $G_{\mathrm{Im}}(i\omega) = (G(i\omega) - G^*(i\omega))/(2i)$ is small compared $G_{\mathrm{Re}}(i\omega)$ in this frequency range. From (4.22) it then follows that a directional disturbance in this frequency range gives rise to field errors in its own direction.

Directionality of the objective function. Field error requirements are typically stated in terms of the allowed rms errors of the amplitude and the phase of the accelerating mode. Often the maximum level of amplitude errors in percent is equal to the maximum level of phase errors in degrees. This requirement is somewhat asymmetric if considered in the complex plane since $1^\circ = 0.017\,\mathrm{rad}$, see the solid rectangle in Figure 15.1a. For some FEL linacs the phase requirements of certain cavities are even further relaxed relative to the amplitude requirements. For example, the X-band cavities at SwissFEL have the error requirement $0.018\%/0.072^\circ$ [Ganter, 2012].

The beam in an ion linac experiences an acceleration proportional to $\mathrm{Re}\{\mathbf{A}\cdot\mathbf{I_b^*}\}$ where $\mathbf{A}$ is the phasor of the accelerating mode and $\mathbf{I_b}$ is the beam-current phasor. The longitudinal focusing of the beam depends on $\mathrm{Im}\{\mathbf{A}\cdot\mathbf{I_b^*}\}$. The directions in which field errors affect these two quantities are indicated in Figure 15.2a. In ion linacs, where the beam is accelerated with a synchronous phase ϕ_b significantly lower than zero, these two quantities that the beam cares about are not quite captured by requirements on amplitude and phase.

The optimal controller. Given a rotationally-invariant plant $P(s)$, a rotationally-invariant control objective, and rotationally-invariant disturbances it follows from Theorem 2 on p. 52 that optimal control performance can be achieved with a rotationally-invariant controller $C(s)$. However, if the control objective, the signal constraints, and the disturbances are not rotationally invariant then it is unlikely that the optimal controller is rotationally



(a) Directionality of the control objective. (b) Directionality of disturbances.

Figure 15.1 *Left:* The directions in which field control errors affect acceleration and focusing of the particle bunches. The small solid rectangle corresponds to a specification in terms of amplitude and phase errors. The dotted rectangle shows a specification in terms of acceleration and longitudinal focusing. Such a specification could potentially lead to better accelerator performance at a lower cost. *Right:* Illustration of the directions that modulator ripple (green dashed line) and the beam current ripple (red dashed line) affect the cavity field in.

invariant. The optimal controller must hence be represented as a widely linear system (see Appendix A) which makes analysis and controller optimization more involved. So even if a complex-coefficient LTI controller is not optimal in a mathematical sense it is likely quite optimal in a practical sense if also understandability and maintainability are considered.

Open questions. We are not aware of that the directionality of the field control problem has been discussed in the existing literature and it seems like an interesting topic for further studies. For example, it would be interesting to know how much performance that is sacrificed by restricting the controller structure to complex-coefficient LTI systems rather than widely linear systems. To compute the optimal control performance achievable by a widely linear controller one could use Q design as in Chapter 9.

Another question is how to specify the field error requirements for an ion linac. It would be interesting to know if the beam losses are better quantified in terms of field errors in the direction indicated in Figure 15.1a than using uncorrelated amplitude and phase errors. Such requirements could be captured by a cost function of the field errors $\boldsymbol{z} = \mathbf{a} - 1$ of the form

$$J(\mathbf{z}) = \begin{bmatrix} \operatorname{Re} \boldsymbol{z} & \operatorname{Im} \boldsymbol{z} \end{bmatrix} Q_J \begin{bmatrix} \operatorname{Re} \boldsymbol{z} \\ \operatorname{Im} \boldsymbol{z} \end{bmatrix} \tag{15.1}$$

where Q_J is a symmetric positive-definite matrix. A bound on the cost function in (15.1) would correspond to an ellipsoidal constraint in Figure 15.1a.

15.3 Future work: Transient operation and iterative learning control

This thesis has focused on closed-loop performance around a nominal operating point (flat-top operation). For pulsed linacs it is also important to consider transient operation. Iterative learning control (ILC) is a good tool for reducing the errors from repetitive set-point changes and disturbances, as was briefly discussed in Section 7.2. The update law of the filter-based ILC formulation in [Norrlöf, 2000] is conveniently modeled using complex-coefficient systems as

$$\boldsymbol{u}_{\mathrm{ILC}k+1} = Q(z)\left(\boldsymbol{u}_{\mathrm{ILC}k} - L(z)\mathbf{e}_k\right),$$

and where $\boldsymbol{u}_{\mathrm{ILC}k}$ is the (complex) feedforward signal, $\mathbf{e}_k$ is the error vector from the previous pulse, $L(z)$ is a suitably chosen (possibly non-causal) filter, and $Q(z)$ is a zero-phase low-pass filter.

It is planned to use this ILC filter-based approach for the ESS linac. The author implemented and experimentally demonstrated it on a bench-top test cavity in October 2015, but due to FPGA hardware upgrades it has not been feasible to repeat the experiments.

15.4 Conclusions

The thesis has provided a solid foundation for cavity field control in terms of normalized complex-baseband models. Based on this foundation, several aspects of field control were analyzed from novel directions and novel results were provided. It is hoped that this thesis has convinced the reader of the merits of complex-coefficient SISO models in field control analysis for linear accelerators.

Bibliography

LCC (2019). The Linear Collider Collaboration. URL: http://www.linearcollider.org/.

Abderrahim, H. A., P. Baeten, D. De Bruyn, and R. Fernandez (2012). "MYRRHA—A multi-purpose fast spectrum research reactor". *Energy conversion and management* **63**, pp. 4–10.

Ainsworth, R. and S. Molloy (2012). "Studies of parasitic cavity modes for proposed ESS linac lattices". In: *Proceedings of LINAC2012*. (Tel-Aviv, Israel, Sept. 9–14, 2012).

Alberi, A. and M. Lacroix (2015). *RFQ mechanical design*. Presentation at CDR2 for the ESS RFQ, CEA Saclay, Paris.

Altarelli, M. et al. (2007). *The European X-Ray Free-Electron Laser, Technical design report*. Tech. rep. DESY 2006-097. Deutsches Elektronen-Synchrotron.

Arnaudon, L. et al. (2006). *Linac4 Technical Design Report*. Tech. rep. CERN-AB-2006-084. CERN, Geneva. URL: http://cds.cern.ch/record/1004186.

Arnold, V. I. (1978). "Mathematical methods of classical mechanics".

Åström, K. J. and B. Wittenmark (1997). *Computer-Controlled Systems: Theory and Design*. 3rd ed. Prentice Hall, Upper Saddle River, NJ.

Åström, K. J. and T. Hägglund (2006). *Advanced PID Control*. The Instrumentation, Systems, and Automation Society, Research Triangle Park, NC. ISBN: 978-1-55617-942-6.

Åström, K. J. and R. M. Murray (2010). *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, NJ.

Åström, K. J. and A.-B. Östberg (1986). "A teaching laboratory for process control". *IEEE Control Systems Magazine* **6**:5, pp. 37–42.

Atsumi, T., A. Okuyama, and M. Kobayashi (2007). "Track-following control using resonant filter in hard disk drives". *IEEE/ASME Transactions on Mechatronics* **12**:4, pp. 472–479.

Automatic Control: Basic Course, Laboratory Exercise 1, PID Control (2019). Laboratory Manual. Dept. of Automatic Control, Lund University, Lund, Sweden.

Ayvazyan, V., S. Choroba, Z. Geng, G. Petrosyan, S. Simrock, and V. Vogel (2010). "Optimization of filling procedure for TESLA-type cavities for klystron RF power minimization for European XFEL". In: *Proceedings of 1st International Particle Accelerator Conference*. Vol. 10.

Bamieh, B., F. Paganini, and M. A. Dahleh (2002). "Distributed control of spatially invariant systems". *IEEE Transactions on Automatic Control* **47**:7, pp. 1091–1107.

Baudrenghien, P. (2000). "Low-level RF systems for synchrotrons". In: *Proceedings of CERN Accelerator School 2000 — Radio Frequency Engineering*. CERN, Geneva, Switzerland.

Baudrenghien, P., D. Valuch, D. Stellfeld, J. Galindo, J. Noirjean, and G. Hagmann (2014). "Commissioning of the Linac4 low level RF and future plans". In: *Proceedings of LINAC2014*. (Geneva, Switzerland, Aug. 31–Sept. 5, 2014).

Baussan, E. et al. (2014). "A very intense neutrino super beam experiment for leptonic CP violation discovery based on the European spallation source linac". *Nuclear Physics B* **885**, pp. 127–149.

Bechhoefer, J. (2005). "Feedback for physicists: A tutorial essay on control". *Reviews of Modern Physics* **77**:3, p. 783.

Bhattacharyya, A. K., V. Ziemann, R. Ruber, and V. Goryashko (2015). "Minimization of power consumption during charging of superconducting accelerating cavities". *Nuclear Instruments and Methods in Physics Research Section A* **801**, pp. 78–85.

Bollen, G. (2010). "FRIB—Facility for rare isotope beams". In: *AIP Conference Proceedings*. Vol. 1224. 1. AIP, pp. 432–441.

Bose, N. and Y. Shi (1987). "A simple general proof of Kharitonov's generalized stability criterion". *IEEE Transactions on Circuits and Systems I* **34**:10, pp. 1233–1237.

Bousson, S. et al. (2014). "The ESS spoke cavity cryomodules". In: *AIP Conference Proceedings*. Vol. 1573. 1. American Institute of Physics, pp. 665–672.

Boyd, S. and C. Barratt (1991). *Linear Controller Design: Limits of Performance*. Prentice-Hall, Englewood Cliffs, NJ. ISBN: 0-13-538687-X. URL: `http://stanford.edu/~boyd/lcdbook/`.

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

Brennan, J. M. (1994). *RF Beam Control for the AGS Booster*. Tech. rep. BNL-52438. Brookhaven National Lab.

Briffa, M. A. (1996). *Linearization of RF Power Amplifiers*. PhD thesis. Dept. of Electrical and Electronic Engineering, Victoria University of Technology, Melbourne, Australia.

Briffa, M. A. and M. Faulkner (1996). "Stability analysis of Cartesian feedback linearisation for amplifiers with weak nonlinearities". *IEE Proceedings - Communications* **143**:4, pp. 212–218.

Bristow, D. A., M. Tharayil, and A. G. Alleyne (2006). "A survey of iterative learning control". *IEEE Control Systems* **26**:3, pp. 96–114.

Butkowski, L., J. Branlard, M. Omet, R. Rybaniec, H. Schlarb, and C. Schmidt (2018). "Implementation of the beam loading compensation algorithm in the LLRF system of the European XFEL". In: *Proccedings of the 29th Linear Accelerator Conference*. (Beijing, China, Sept. 16–21, 2018), pp. 594–597.

Byun, S.-W. and C.-W. Lee (1988). "Pole assignment in rotating disk vibration control using complex modal state feedback". *Mechanical Systems and Signal Processing* **2**:3, pp. 225–241.

Carminati, F., R. Klapisch, J. P. C. Revol, C. Roche, J. A. Rubio, and C. Rubbia (1993). *An energy amplifier for cleaner and inexhaustible nuclear energy production driven by a particle beam accelerator*. Tech. rep. CERN/AT/93-47 (ET). CERN.

Carter, R. G. (2010). "RF power generation". In: *Proceedings of CERN Accelerator School 2010 — RF for Accelerators*. CERN, Geneva, Switzerland.

Castilla, M., J. Miret, J. Matas, L. G. de Vicuna, and J. M. Guerrero (2009). "Control design guidelines for single-phase grid-connected photovoltaic inverters with damped resonant harmonic compensators". *IEEE Transactions on Industrial Electronics* **56**:11, pp. 4492–4501.

Celona, L. et al. (2016). "Preliminary commissioning results of the proton source for ESS at INFN-LNS". In: *Proceedings of the 7th International Particle Accelerator Conference*. (Busan, Korea, May 8–13, 2016), pp. 2628–2631.

Champion, M., M. Plum, M. Howell, J. Galambos, R. Dean, and B. Riemer (2017). "Progress on the proton power upgrade of the spallation neutron source". In: *Proceedings of the 8th International Particle Accelerator Conference*. (Copenhagen, Denmark, May 14–19, 2019).

Creaney, S. and I. Kostarnov (2008). *Designing efficient digital up and down converters for narrowband systems*. XAPP1113. Application Note. Xilinx.

Crochiere, R. E. and L. R. Rabiner (1983). *Multirate Digital Signal Pocessing*. Prentice Hall, Englewood Cliffs, NJ.

CVX Research Inc (2012). *CVX: Matlab software for disciplined convex programming, version 2.0*. `http://cvxr.com/cvx`.

Czarski, T., K. T. Pozniak, R. S. Romaniuk, and S. Simrock (2003). "Cavity control system essential modeling for the TESLA linear accelerator". In: *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments*. Vol. 5125. International Society for Optics and Photonics, pp. 205–213.

Dawson, J. L. (2003). *Feedback Linearization of RF Power Amplifiers*. PhD thesis. Dept. of Electrical Engineering, Stanford University, Stanford, CA.

Delayen, J. R. (1978). *Phase and Amplitude Stabilization of Superconducting Resonators*. PhD thesis. California Institute of Technology, Pasadena, CA.

Delayen, J. (2006). "Ponderomotive instabilities and microphonics—a tutorial". *Physica C: Superconductivity* **441**:1-2, pp. 1–6.

Doolittle, L. (1989). *Understanding 5-cell mode structures*. Tech. rep. CEBAF-TN-0120. Thomas Jefferson National Accelerator Facility, Newport News, VA.

Doolittle, L. (2008). "Digital IQ reconstruction from a non-IQ sampled waveform". URL: `http://recycle.lbl.gov/~ldoolitt/down/` (visited on 01/03/2019).

Doolittle, L. (2014). *The Hows and Whys of SEL feedback for LLRF*. Tech. rep. Lawrence Berkeley National Lab, Berkeley, CA.

Doolittle, L. et al. (2015). "The LCLS-II LLRF system". In: *Proceedings of the 6th International Particle Accelerator Conference*. (Richmond, VA, May 3–8, 2015).

Doolittle, L., H. Ma, and M. S. Champion (2006). "Digital low-level RF control using non-IQ sampling". In: *Proceedings of LINAC 2006*. (Knoxville, TN, Aug. 20–25, 2006).

Doolittle, L. et al. (2016). "LLRF control of high QL cavities for the LCLS-II". In: *Proceedings of the 7th International Particle Accelerator Conference*. (Busan, Korea, May 8–13, 2016).

Dòria-Cerezo, A. and M. Bodson (2016). "Design of controllers for electrical power systems using a complex root locus method". *IEEE Transactions on Industrial Electronics* **63**:6, pp. 3706–3716.

Ellingson, S. W. (2016). *Radio Systems Engineering*. Cambridge University Press, Cambridge, UK. ISBN: 978-1-107-06828-5.

Emami-Naeini, A. and P. Van Dooren (1982). "Computation of zeros of linear multivariable systems". *Automatica* **18**:4, pp. 415–430.

EPICS — Experimental Physics and Industrial Control System. `https://epics-controls.org/`. Accessed: 2019-09-01.

Eshraqi, M. and Y. Levinsen (2016). *LLRF needs for Beam Physics*. Tech. rep. ESS-0049330. ESS internal document. ESS, Lund, Sweden.

ESSa (2017). *The unique capabilities of ESS*. Accessed 2017-04-12. URL: `https://europeanspallationsource.se/unique-capabilities-ess`.

ESSb (2017). *Elliptical SRF*. Accessed 2017-05-21. URL: `https://europeanspallationsource.se/elliptical-srf`.

Ferrario, M., A. Mosnier, L. Serafini, F. Tazzioli, and J. Tessier (1996). "Multi-bunch energy spread induced by beam loading in a standing wave structure". *Particle Accelerators* **52**:1, pp. 1–30.

Fitzgibbon, A., M. Pilu, and R. B. Fisher (1999). "Direct least square fitting of ellipses". *IEEE Transactions on pattern analysis and machine intelligence* **21**:5, pp. 476–480.

Francis, B. A. and W. M. Wonham (1976). "The internal model principle of control theory". *Automatica* **12**:5, pp. 457–465.

Frank, E. (1946). "On the zeros of polynomials with complex coefficients". *Bulletin of the American Mathematical Society* **52**:2, pp. 144–157.

Freudenberg, J. S. and D. P. Looze (1985). "Right half plane poles and zeros and design trade-offs in feedback systems". *IEEE Transactions on Automatic Control* **30**:6, pp. 555–565.

Gabor, D. (1946). "Theory of communication. part 1: the analysis of information". *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering* **93**:26, pp. 429–441.

Galayda, J. (2014). "The new LCLS-II project: Status and challenges". In: *Proceedings of the 5th International Particle Accelerator Conference*. (Dresden, Germany, June 15–20, 2014).

Gallo, A. (2010). "Basics of RF electronics". In: *Proceedings of CERN Accelerator School 2010 — RF for Accelerators*. CERN, Geneva, Switzerland.

Ganter, R. (2012). *Swissfel-conceptual design report*. Tech. rep. PSI 10-04. Paul Scherrer Institute, Villigen, Switzerland.

Garin, P., M. Sugimoto, et al. (2009). "Main baseline of IFMIF/EVEDA project". *Fusion Engineering and Design* **84**:2-6, pp. 259–264.

Garoby, R. (1996). "Low level RF and feedback". In: *Joint US-CERN-Japan Accelerator School on Frontiers of Accelerator Technology*. URL: `https://cds.cern.ch/record/328252`.

Garoby, R. (2010). "RF gymnastics in synchrotrons". In: *Proceedings of CERN Accelerator School 2010 — RF for Accelerators*. CERN, Geneva, Switzerland.

Garoby, R. et al. (2017). "The European spallation source design". *Physica Scripta* **93**:1, p. 014001.

Garpinger, O. (2009). *Design of Robust PID Controllers with Constrained Control Signal Activity*. Licentiate thesis TFRT-3245. Dept. of Automatic Control, Lund University, Lund, Sweden.

Gataric, S. and N. R. Garrigan (1999). "Modeling and design of three-phase systems using complex transfer functions". In: *30th Annual IEEE Power Electronics Specialists Conference*. Vol. 2, pp. 691–697.

Gilmour, A. (2011). *Klystrons, Traveling Wave Tubes, Magnetrons, Crossed-Field Amplifiers, and Gyrotrons*. Artech House, Norwood, MA. ISBN: 978-1608071845.

Gonzalez-Cava, J. M., F. B. Carlson, O. Troeng, A. Cervin, K. van Heusden, G. A. Dumont, and K. Soltesz (2019). *PID versus Youla parameter control of propofol anesthesia*. Submitted to IEEE Transactions on Control Systems Technology.

Han, C.-H., C.-C. Wang, and M. Tomizuka (2008). "Suppression of vibration due to transmission error of harmonic drives using peak filter with acceleration feedback". In: *Proceedings of the 10th IEEE International Workshop on Advanced Motion Control*. (Trento, Italy, Mar. 26–28, 2008), pp. 182–187.

Hara, M., T. Nakamura, and T. Ohshima (1998). "A ripple effect of a klystron power supply on synchrotron oscillation". *Particle Accelerators* **59**, pp. 143–156.

Hara, S., Y. Yamamoto, T. Omata, and M. Nakano (1988). "Repetitive control system: a new type servo system for periodic exogenous signals". *IEEE Transactions on Automatic Control* **33**:7, pp. 659–668.

Harnefors, L. (2007). "Modeling of three-phase dynamic systems using complex transfer functions and transfer matrices". *IEEE Transactions on Industrial Electronics* **54**:4, pp. 2239–2248.

Haus, H. A. (1983). *Waves and Fields in Optoelectronics*. Prentice-Hall, Englewood Cliffs, NJ. ISBN: 0-13-946053-5.

Hellert, T. and M. Dohlus (2018). "Detuning related coupler kick variation of a superconducting nine-cell 1.3 GHz cavity". *Physical Review Accelerators and Beams* **21** (4), p. 042001.

Henke, H. and M. Filtz (1993). "Envelope equations for transients in linear chains of resonators". In: *Proceedings of the 1993 Particle Accelerator Conference*. (Washington D.C. May 17–20, 1993). Vol. 2, pp. 901–903.

Hespanha, J. P. (2009). *Linear Systems Theory*. Princeton University Press, Princeton, NJ. ISBN: 978-0-691-14021-6.

Hoffmann, M. (2008). *Development of a multichannel RF field detector for the low-level RF control of the free-electron laser at Hamburg*. PhD thesis. TU Hamburg-Harburg, Hamburg, Germany.

Hogenauer, E. (1981). "An economical class of digital filters for decimation and interpolation". *IEEE Transactions on Acoustics, Speech, and Signal Processing* **29**:2, pp. 155–162.

Huang, G. et al. (2016). "High precision RF control for the LCLS-II". In: *Proceedings of the 2nd North American Particle Accelerator Conference*. (Chicago, IL, Oct. 9–14, 2016).

Inoue, T., M. Nakano, T. Kubo, S. Matsumoto, and H. Baba (1981). "High accuracy control of a proton synchrotron magnet power supply". In: *Proceedings of the 8th IFAC World Congress*. (Kyoto, Japan, Aug. 24–28, 1981).

Jameson, R. and J. Wallace (1971). "Feedforward control of accelerator RF fields". *IEEE Transactions on Nuclear Science* **18**:3, pp. 598–600.

Jamieson, S. (2006). *Micro Telecommunications Computing Architecture Short Form Specification*. Tech. rep. PICMG MTCA.0. PICMG. URL: `https://www.picmg.org/wp-content/uploads/MicroTCA_Short_Form_Sept_2006.pdf`.

Jensen, M. et al. (2018). "Testing of the ESS MB-IOT prototypes". In: *Proceedings of the 9th International Particle Accelerator Conference*. (Vancouver, Canada, Apr. 29–May 4, 2018). JACOW Publishing, Geneva, Switzerland, pp. 1759–1764.

Johansson, M. (1991). *Linearization of RF power amplifiers using Cartesian feedback*. Licentiate thesis. Dept. of Applied Electronics, Lund University, Lund, Sweden.

Kaiser, K. L. (2005). *Transmission lines, matching, and crosstalk*. CRC Press, Boca Raton, FL.

Kim, Y., C. Kang, and M. Tomizuka (2005). "Adaptive and optimal rejection of non-repeatable disturbance in hard disk drives". In: *Proc. IEEE/ASME Int. Conf. Adv. Intelligent Mechatronics*. Vol. 1, pp. 1–6.

Kjellqvist, O. and O. Troeng (2019). *Numerical pitfalls in Q-design*. Submitted to the 21st World Congress of the International Federation of Automatic Control.

Lancaster, P. and L. Rodman (1995). *Algebraic Riccati Equations*. Oxford University Press, Oxford, UK.

Lapostolle, P. M. and A. L. Septier, (Eds.) (1970). *Linear Accelerators*. North-Holland Publishing Company, Amsterdam, The Netherlands.

Leibrandt, D. R. and J. Heidecker (2015). "An open source digital servo for atomic, molecular, and optical physics experiments". *Review of Scientific Instruments* **86**:12, p. 123115.

Levinsen, Y., M. Eshraqi, R. Miyamoto, M. Munoz, A. Ponton, and R. D. Prisco (2016). "Beam dynamics challenges in the ESS linac". In: *Proceedings of 57th ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams*. (Malmö, Sweden, July 3–8, 2016).

Liberzon, D. (2012). *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton University Press, Princeton, NJ. ISBN: 978-0-691-15187-8.

Liepe, M. U. (2001). *Superconducting Multicell Cavities for Linear Colliders*. PhD thesis. University of Hamburg, Hamburg, Germany.

Ling, K., S. Yue, and J. Maciejowski (2006). "A FPGA implementation of model predictive control". In: *Proceedings of the 2006 American Control Conference*. (Minneapolis, MN, June 14–16, 2006). IEEE, pp. 1930–1935.

Lonza, M. and H. Schmickler (2014). "Digital signal processing in RF applications". In: *Proceedings of CERN Accelerator School 2013 — Advanced Accelerator Physics Course*.

LLRF (2019). *Low-level RF workshop*. URL: `llrf.net`.

Ludwig, F. (2011). "Tutorial on RF (receiver fundamentals)". In: *Presented at the LLRF Workshop*. (Hamburg, Germany, Oct. 17–20, 2011).

Ludwig, F. (2017). *Drift calibration for the European XFEL*. Presented at the Low-Level Radio-Frequency Workshop 2017, Barcelona, Spain, Nov. 3–6, 2017.

Ludwig, F., M. Hoffmann, H. Schlarb, and S. Simrock (2006). "Phase stability of the next generation RF field control for VUV-and X-ray free electron laser". In: *Proceedings of the 10th European Particle Accelerator Conference*. (Edinburgh, Scotland, June 26–30, 2006), pp. 1453–1455.

Ludwig, F., J. Branlard, L. Butkowski, M. Czwalinna, et al. (2019). "RF controls towards femtosecond and attosecond precision". In: *Proceedings of the 10th International Particle Accelerator Conference*. (Melbourne, Australia, May 19–24, 2019).

Ma, H. et al. (2006). "Low-level RF control of Spallation Neutron Source: System and characterization". *Physical Review Special Topics - Accelerators and Beams* **9** (3), p. 032001. URL: `https://link.aps.org/doi/10.1103/PhysRevSTAB.9.032001`.

Mandic, D. P. and V. S. L. Goh (2009). *Complex valued nonlinear adaptive filters: noncircularity, widely linear and neural models*. Vol. 59. John Wiley & Sons, West Sussex, UK. ISBN: 978-0-470-06635-5.

Mastoridis, T. (2005). *Radio Frequency Station—Beam Dynamics Interaction in Circular Accelerators*. PhD thesis. Dept. of Electrical Engineering, Stanford University, Stanford, CA.

MAX IV (2010). *Detailed Design Report on the MAX IV facility*. Tech. rep.

Merminga, L. and G. Kraft (1996). *Energy spread from RF amplitude and phase errors*. Tech. rep. CEBAF-PR-96-011. Thomas Jefferson National Accelerator Facility, Newport News, VA.

Merminga, L. and J. R. Delayen (1996). *On the optimization of Qext under heavy beam loading and in the presence of microphonics*. Tech. rep. TN-96-022. CEBAF.

MTCA.4 (2017). *MicroTCA.4 for industry and research*. DESY. URL: `http://mtca.desy.de/`.

Milne, C. et al. (2017). "SwissFEL: the Swiss X-ray free electron laser". *Applied Sciences* **7**:7, p. 720.

Mokhov, N. V. and W. Chou, (Eds.) (1999). *Beam Halo and Scraping — Proceedings of the 7th ICFA Mini-Workshop on High-Intensity and High-Brightness Hadron Beams*. (Lake Como, Wisconsin, Sept. 13–15, 1999). Fermi National Accelerator Laboratory, Batavia, IL.

Monmasson, E., L. Idkhajine, and M. W. Naouar (2011). "FPGA-based controllers". *IEEE Industrial Electronics Magazine* **5**:1, pp. 14–26.

Montgomery, C. G., R. H. Dicke, and E. M. Purcell (1948). *Principles of Microwave Circuits*. McGraw-Hill, New York, NY.

Neri, L. et al. (2014). "Improved design of proton source and low energy beam transport line for European Spallation Source". *Review of Scientific Instruments* **85**:2, 02A723.

Norrlöf, M. (2000). *Iterative Learning Control: Analysis, Design, and Experiments*. PhD thesis. Dept. of Electrical Engineering, Linköping University, Linköping, Sweden.

Novotny, D. W. and J. H. Wouterse (1976). "Induction machine transfer functions and dynamic response by means of complex time variables". *IEEE Transactions on Power Apparatus and Systems* **95**:4, pp. 1325–1335.

Olofsson, B., B. Bernhardsson, R. Zeng, P. Andersson, and R. Johansson (2018). "Temperature stabilization of the phase-reference line at the european spallation source". In: *Proceedings of the 2nd IEEE Conference on Control Technology and Applications*. (Copenhagen, Denmark, Aug. 21–24, 2018), pp. 1369–1376.

Omet, M. (2014). *Digital Low Level RF Control Techniques and Procedures Towards the International Linear Collider*. PhD thesis. Dept. of Accelerator Science, The Graduate University for Advanced Studies, Kanagawa, Japan.

Padamsee, H., J. Knobloch, and T. Hays (2008). *RF Superconductivity for Accelerators*. 2nd ed. Wiley-VCH, Weinheim, Germany.

Pedersen, F. (1975). "Beam loading effects in the CERN PS booster". *IEEE Transactions on Nuclear Science* **22**:3, pp. 1906–1909.

Peggs, S. et al. (2013). *ESS Technical Design Report*. Tech. rep. ESS-doc-274. European Spallation Source, Lund, Sweden.

Pfeiffer, S. (2014). *Symmetric Grey Box Identification and Distributed Beam-Based Controller Design for Free-Electron Lasers*. PhD thesis. TU Hamburg-Harburg, Hamburg, Germany.

Pfeiffer, S., G. Lichtenberg, C. Schmidt, H. Schlarb, and H. Werner (2012). "Design of an optimal and robust controller for a free-electron laser exploiting symmetries of the RF-system". In: *Proceedings of the 51st IEEE Conference on Decision and Control*. (Maui, HI, Dec. 10–13, 2012). IEEE, pp. 4253–4258.

Phadke, A. G., J. S. Thorp, and M. G. Adamiak (1983). "A new measurement technique for tracking voltage phasors, local system frequency, and rate of change of frequency". *IEEE Transactions on Power Apparatus and Systems* 5, pp. 1025–1038.

Phadke, A. G. and J. S. Thorp (2017). *Synchronized phasor measurements and their applications*. 2nd ed. Springer, Cham, Switzerland. ISBN: 978-3-319-50584-8.

Picinbono, B. and P. Bondon (1997). "Second-order statistics of complex signals". *IEEE Transactions on Signal Processing* **45**:2, pp. 411–420.

Picinbono, B. and P. Chevalier (1995). "Widely linear estimation with complex data". *IEEE Transactions on Signal Processing* **43**:8, pp. 2030–2033.

Piller, M., M. Champion, M. Crofford, H. Ma, and L. Doolittle (2005). "The Spallation Neutron Source RF reference system". In: *Proceedings of the 2005 Particle Accelerator Conference*. (Knoxville, TN, May 16–20, 2005). IEEE, pp. 3573–3575.

Pippard, A. B. (2007). *The Physics of Vibration*. Cambridge University Press, New York, NY. ISBN: 9780521033336.

Pirkl, W. (1994). "RF systems for linacs". In: *Proceedings of CERN Accelerator School 1994 — Cyclotrons, Linacs and their Applications*. CERN, Geneva, Switzerland.

Plawski, T. (2014). *Digital RF Control System for Superconducting Cavity with Lorenz Force Detuning Coefficient*. PhD thesis. Inst. of Electronic Systems, Warsaw University of Technology, Warsaw, Poland.

Plum, M. A. "Beam loss in linacs", pp. 39–62.

Pozar, D. M. (2009). *Microwave Engineering*. 4th ed. John Wiley & Sons, New York.

Proakis, J. G. and M. Salehi (2002). *Communication Systems Engineering*. 2nd ed. Prentice Hall, Upper Saddle River, NJ.

Ren, Y., D. Su, and J. Fang (2013). "Whirling modes stability criterion for a magnetically suspended flywheel rotor with significant gyroscopic effects and bending modes". *IEEE Transactions on Power Electronics* **28**:12, pp. 5890–5901.

Rezaeizadeh, A., T. Schilcher, and R. S. Smith (2017). "Control of the Swiss free electron laser: Methods for precision control of pulsed-mode accelerator beams". *IEEE Control Systems Magazine* **37**:6, pp. 30–51.

Rezaeizadeh, A. (2016). *Automatic Control Strategies for the Swiss Free Electron Laser*. PhD thesis. Swiss Federal Institute of Technology, Zürich, Switzerland.

Robinson, K. W. (1964). *Stability of Beam in Radiofrequency System*. Tech. rep. CEAL-1010. Cambridge Electron Accelerator, Cambridge, MA.

Rogers, E. et al. (2010). "Norm-optimal iterative learning control with application to problems in accelerator-based free electron lasers and rehabilitation robotics". *European Journal of Control* **16**:5, pp. 497–522.

Rojas, C. R., P. Zetterberg, and P. Handel (2011). "Transceiver in-phase/quadrature imbalance, ellipse fitting, and the universal software radio peripheral". *IEEE Transactions on Instrumentation and Measurement* **60**:11, pp. 3629–3639.

Russell, G. J. (1990). "Spallation physics — an overview". In: *Proceedings of International Collaboration on Advanced Neutron Sources XI*. (Tsukuba, Japan, Oct. 22–26, 1990).

Rybarcyk, L. and R. McCrady (2016). "The effect of DTL cavity field errors on beam spill at LANSCE". In: *Proceedings of LINAC2016*. (East Lansing, MI, Sept. 25–30, 2016).

Saggin, F., A. Korniienko, and G. Scorletti (2019). *On Phasor Control for Linear Time Invariant systems*. Research Report. Ecole Centrale Lyon. URL: `https://hal.archives-ouvertes.fr/hal-02068370`.

Salom, A., F. Pérez, and E. Morales (2017). "Development of a DLLRF using commercial uTCA platform".

Schenk, T. (2008). *RF imperfections in high-rate wireless systems: Impact and digital compensation*. Springer Science & Business Media, Dordrecht, The Netherlands.

Schilcher, T. (1998). *Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconducting Cavities*. PhD thesis. University of Hamburg, Hamburg, Germany.

Schilcher, T. (2007). "Digital signal processing in RF applications". In: *Proceedings of CERN Accelerator School 2007 — Digital Signal Processing*. CERN, Geneva, Switzerland.

Schmidt, C., S. Pfeiffer, J. Branlard, H. Schlarb, and W. Jalmuzna (2012). "Precision regulation of RF fields with MIMO controllers and cavity-based notch filters". In: *Proceedings of LINAC2012*. (Tel-Aviv, Israel, Sept. 9–14, 2012).

Schmidt, C. (2010). *RF System Modeling and Controller Design for the European XFEL*. PhD thesis. TU Hamburg-Harburg, Hamburg, Germany.

Schmidt, R. (2016). "Introduction to machine protection", pp. 1–20.

Schreier, P. J. and L. L. Scharf (2010). *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, Cambridge, UK. ISBN: 978-0-511-67772-4.

Sekutowicz, J. (2010). "Superconducting elliptical cavities". In: *Proceedings of CERN Accelerator School 2010 — RF for Accelerators*. CERN, Geneva, Switzerland.

Serrano, C., L. R. Doolittle, G. Huang, and P. J. Emma (2017). *LCLS-II Longitudinal Beam Dynamics: Physics & Simulations*. Tech. rep. LCLSII-TN-17-06. Lawrence Berkeley National Lab.

Siegman, A. E. (1986). *Lasers*. University Science Books, Mill Valley, CA. ISBN: 978-0935702118.

Sievers, L. A. and A. H. von Flotow (1992). "Comparison and extensions of control methods for narrow-band disturbance rejection". *IEEE Transactions on Signal Processing* **40**:10, pp. 2377–2391.

Sikora, D. et al. (2018). *Phase drift compensating RF link for femtosecond synchronization of E-XFEL*. eprint: `arXiv:1806.09197`.

Simrock, S. (2005). "Digital low-level RF conctrols for future superconducting linear colliders". In: *Proceedings of the 2005 Particle Accelerator Conference*. (Knoxville, TN, May 16–20, 2005), pp. 515–519.

Simrock, S. and Z. Geng (2013). "Cavity field control, LLRF lecture part 3". In: *Proceedings of the 8th International Accelerator School for Linear Colliders*.

Simrock, S., M. Hoffmann, F. Ludwig, M. Grecki, and T. Jezynski (2006). "Considerations for the choice of the intermediate frequency and sampling rate for digital RF control". In: *Proceedings of the 10th European Particle Accelerator Conference*. (Edinburgh, Scotland, June 26–30, 2006), pp. 1462–1464.

Skogestad, S. and I. Postlethwaite (2007). *Multivariable Feedback Control: Analysis and Design*. 2nd ed. John Wiley & Sons, Chichester, UK.

Slater, J. C. (1950). *Microwave Electronics*. Van Nostrand, Princeton, NJ.

Stein, G. (2003). "Respect the unstable". *IEEE Control Systems* **23**:4, pp. 12–25.

Stein, S. and J. J. Jones (1967). *Modern communication principles: With application to digital signaling*. McGraw-Hill, New York.

Stohr, J. et al. (2011). *Linac coherent light source II (LCLS-II) conceptual design report*. Tech. rep. SLAC-I-060-003-000-00. SLAC National Accelerator Lab, Menlo Park, CA.

TANGO Controls. `https://www.tango-controls.org/`. Accessed: 2019-09-01.

Troeng, O. and B. Bernhardsson (2018). "An intuitive design method for disturbance-rejecting peak filters". In: *Proceedings of the 14th IEEE International Conference on Control and Automation (ICCA)*. (Anchorage, AK, June 12–15, 2018), pp. 969–974.

Troeng, O. (2017). *Cavity Field Control for High-Intensity Linear Proton Accelerators*. Licentiate thesis TFRT-3237. Dept. of Automatic Control, Lund University, Lund, Sweden.

Troeng, O. (2019). "Modeling the accelerating mode of accelerator cavities". Submitted to Physical Review Accelerators and Beams.

Troeng, O. and B. Bernhardsson (2017). "Energy-optimal excitation of radio-frequency cavities". In: *Proceedings of the 20th World Congress of the International Federation of Automatic Control*. (Toulouse, France, July 9–14, 2017), pp. 10644–10648.

Troeng, O., B. Bernhardsson, and C. Rivetta (2017). "Complex-coefficient systems in control". In: *Proceedings of the 2017 American Control Conference*. (Seattle, WA, May 24–26, 2017), pp. 1721–1727.

Tückmantel, J. (2011). *Cavity-Beam-Transmitter Interaction Formula Collection with Derivation*. Tech. rep. CERN-ATS-Note-2011-002 TECH. CERN, Geneva, Switzerland.

Turner, S., (Ed.) (1994). *Proceedings of CERN Accelerator School 1994 — Cyclotrons, Linacs and Their Application*. CERN, Geneva, Switzerland. ISBN: 92-9083-083-2.

Uriot, D. and N. Pichoff (2015). "Status of the Tracewin code". In: *Proceedings of the 6th International Particle Accelerator Conference*. (Richmond, VA, May 3–8, 2015).

UT-Battelle, L. (2006). *Spallation Neutron Source Project Completion Report*. Tech. rep. SNS 100000000-BL0005-R00. Oak Ridge National Laboratory, Oak Ridge, TN.

Van Der Meer, S. (1985). "Stochastic cooling and the accumulation of antiprotons". *Reviews of Modern Physics* **57**:3, p. 689.

Van Tartwijk, G. and D. Lenstra (1995). "Semiconductor lasers with optical injection and feedback". *Quantum and Semiclassical Optics: Journal of the European Optical Society Part B* **7**:2, p. 87.

Vogel, E. (2007). "High gain proportional RF control stability at TESLA cavities". *Physical Review Special Topics - Accelerators and Beams* **10**:5, p. 052001.

Wang, Y., F. Gao, and F. J. Doyle III (2009). "Survey on iterative learning control, repetitive control, and run-to-run control". *Journal of Process Control* **19**:10, pp. 1589–1600.

Wangler, T. P. (2008). *RF Linear Accelerators*. 2nd ed. Wiley-VCH, Weinheim, Germany. ISBN: 978-3-527-40680-7.

Wiedemann, H. (2015). *Particle Accelerator Physics*. 3rd ed. Springer, Berlin.

Xu, J. X., D. Huang, V. Venkataramanan, and T. C. T. Huynh (2012). "Adaptive compensation of contact-induced vibration in high density hdd servo systems using peak filter method". In: *IEEE Int. Symp. on Ind. Electronics*, pp. 797–802.

Youla, D., H. Jabr, and J. Bongiorno (1976). "Modern Wiener-Hopf design of optimal controllers—Part II: the multivariable case". *IEEE Transactions on Automatic Control* **21**:3, pp. 319–338.

Zheng, J. et al. (2006). "Phase lead peak filter method to high TPI servo track writer with microactuators". In: *Proceedings of the 2006 American Control Conference*. (Minneapolis, MN, June 14–16, 2006). IEEE, pp. 1309–1314.

Zhou, K., J. C. Doyle, and K. Glover (1996). *Robust and Optimal Control*. Prentice Hall, Englewood Cliffs, NJ.

Zou, Y., M. Valkama, and M. Renfors (2008). "Analysis and compensation of transmitter and receiver I/Q imbalances in space-time coded multi-antenna OFDM systems". *EURASIP Journal on Wireless Communications and Networking* **2008**, pp. 1–16.

Appendices

A

Widely Linear Systems

Certain systems which are conveniently modeled with complex signals do not have a strictly complex-linear relationship between the input signal $\boldsymbol{u}$ and the output signal $\boldsymbol{y}$, but rather depend on both $\boldsymbol{u}$ through a transfer function $G(s)$ and the conjugate $\boldsymbol{u}^*$ through a transfer function $\breve{G}(s)$, see Figure A.1. We will refer to the system in Figure A.1 as a *widely linear*[1] system and denote it by the two-tuple $[G, \breve{G}]$.

The widely linear system $[G, \breve{G}]$ can, similarly to (4.14), be written as an equivalent real-coefficient TITO system acting on signals $\begin{bmatrix} u_{\mathrm{Re}} & u_{\mathrm{Im}} \end{bmatrix}^{\mathsf{T}}$,

$$\mathcal{G}_{\mathrm{equiv}}(s) = \begin{bmatrix} G_{\mathrm{Re}}(s) & -G_{\mathrm{Im}}(s) \\ G_{\mathrm{Im}}(s) & G_{\mathrm{Re}}(s) \end{bmatrix} + \begin{bmatrix} \breve{G}_{\mathrm{Re}}(s) & -\breve{G}_{\mathrm{Im}}(s) \\ \breve{G}_{\mathrm{Im}}(s) & \breve{G}_{\mathrm{Re}}(s) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad \text{(A.1)}$$

Another, possibly more insightful representation of (A.1), is obtained by applying the same unitary transformation T as in (4.15), to obtain the so-called *augmented system* [Schreier and Scharf, 2010]

$$\mathcal{G}_{\mathrm{aug}}(s) = T^{\mathsf{H}} \mathcal{G}_{\mathrm{equiv}}(s) T = \begin{bmatrix} G(s) & \breve{G}(s) \\ \breve{G}^*(s) & G^*(s) \end{bmatrix}. \quad \text{(A.2)}$$

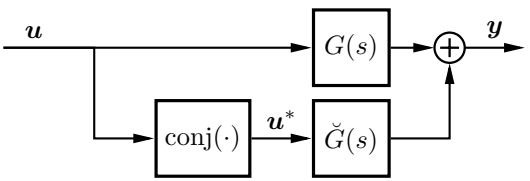


Figure A.1 The action of the widely linear system $[G, \breve{G}]$ on a complex signal $\boldsymbol{u}$. Both $G(s)$ and $\breve{G}(s)$ have complex coefficients in general.

[1] Widely linear is the term used in the telecommunications literature [Picinbono and Chevalier, 1995; Schreier and Scharf, 2010]. The power-electronics literature use the terms unbalanced [Dòria-Cerezo and Bodson, 2016] and asymmetric [Harnefors, 2007].

This system can (recall Remark 4.2) be thought of as mapping signals $\begin{bmatrix} \boldsymbol{u} & \boldsymbol{u}^* \end{bmatrix}^\mathsf{T}$ to signals $\begin{bmatrix} \boldsymbol{y} & \boldsymbol{y}^* \end{bmatrix}^\mathsf{T}$. The augmented representation (A.2) is the standard representation of widely linear systems in the communications literature [Schreier and Scharf, 2010]. Note that the second row of $\mathcal{G}_{\text{aug}}$ is redundant in that it can be obtain from the first row through conjugation.

For cavity field control, widely linear behavior arises from IQ imbalance in analog quadrature up- and downconversion and from the amplifier nonlinearity (when linearizing around a nominal operating point). In the first case, the widely linear behavior is due to a lack of time-invariance and in the second case it is due to that the linearization of a rotationally-invariant system around some operating point other than 0 is not guaranteed to be linear.

Much of the dynamics of electrical machines is rotationally invariant, but the rotor dynamics give rise to widely linear dynamics [Harnefors, 2007]. Similarly, for optical feedback in diode lasers, the optical cavity and the feedback path have linear dynamics. However, the lasing dynamics (linearized around an operating point) is widely linear [Van Tartwijk and Lenstra, 1995]. In signal processing it is not uncommon with widely linear models, see for example [Schreier and Scharf, 2010; Mandic and Goh, 2009].

Remark A.1 Any real-coefficient system with an even number of inputs and outputs can be identified with a widely linear system. However, there is only reason to expect that this representation is helpful if the system is almost rotationally invariant in the sense that $\breve{G}$ is small relative to G.

Interconnections of widely linear systems

The set of widely linear systems $[G, \breve{G}]$ constitute an associative algebra over the complex numbers if addition is defined by

$$[G, \breve{G}] + [H, \breve{H}] \coloneqq [G + H, \breve{G} + \breve{H}], \tag{A.3}$$

multiplication by a complex number w is defined by

$$w[G, \breve{G}] \coloneqq [wG, w\breve{G}], \tag{A.4}$$

and the bilinear product is defined by

$$[G, \breve{G}] \cdot [H, \breve{H}] \coloneqq [GH + \breve{G}\breve{H}^*, G\breve{H} + \breve{G}H^*]. \tag{A.5}$$

This is equivalent to consider widely linear systems to be represented by their augmented representation (A.2), however, the notation $[G, \breve{G}]$ avoids redundancy.

Worst-case robustness degradation due to unbalanced dynamics

Definition A.1 The $\mathcal{H}_\infty$ norm of a widely linear system is defined by

$$\left|\left|[G, \breve{G}]\right|\right|_\infty := \left|\left|\begin{bmatrix} G & \breve{G} \\ \breve{G}^* & G^* \end{bmatrix}\right|\right|_\infty .$$

THEOREM 4
Assume that the complex-coefficient system L_0 subject to unity negative feedback is stable and has a maximum sensitivity of $M_{S0} := ||1/(1+L_0)||_\infty$. Then the widely linear system $[L_0, \breve{L}]$ where $\breve{L}$ is a small unbalance contribution with $||\breve{L}||_\infty < 1/M_{S0}$ has, subject to unity feedback, its maximum sensitivity $\breve{M}_S$ bounded as

$$\breve{M}_S \leq \frac{M_{S0}}{1 - M_{S0}||\breve{L}||_\infty} \tag{A.6}$$

Remark A.2 If the unbalance is small, a first-order expansion gives that $\breve{M}_S \lessapprox M_{S0} + (M_{S0})^2||\breve{L}||_\infty$.

Proof. The proof follows the lines of the proof of Theorem 9.7 in [Zhou et al., 1996]. From the small-gain theorem it follows that the feedback interconnection is stable. The equivalent real-coefficient TITO representations of G_0 and $\breve{G}$ are

$$\mathcal{G}_0 = \begin{bmatrix} G_{0\mathrm{Re}} & -G_{0\mathrm{Im}} \\ G_{0\mathrm{Im}} & G_{0\mathrm{Re}} \end{bmatrix}, \quad \breve{\mathcal{G}} = \begin{bmatrix} \breve{G}_{\mathrm{Re}} & -\breve{G}_{\mathrm{Im}} \\ \breve{G}_{\mathrm{Im}} & \breve{G}_{\mathrm{Re}} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & -1 \end{bmatrix}$$

Note that for any two complex matrices A and B it holds that $\underline{\sigma}(A+B) \geq \underline{\sigma}(A) - \overline{\sigma}(B)$, where $\overline{\sigma}$ denotes the largest singular value and $\underline{\sigma}$ denotes the smallest singular value. Also note that the assumptions of the theorem ensure that $\underline{\sigma}(I + \mathcal{G}_0 + \breve{\mathcal{G}}) > 0$. These two observations show that the frequency response of the sensitivity function $\breve{\mathcal{S}} = (I + \mathcal{G}_0 + \breve{\mathcal{G}})^{-1}$ at each frequency ω satisfies the following inequalities (where the frequency argument is suppressed)

$$\begin{aligned} \overline{\sigma}(\breve{\mathcal{S}}) &= \overline{\sigma}\left((I + \mathcal{G}_0 + \breve{\mathcal{G}})^{-1}\right) \\ &= \frac{1}{\underline{\sigma}(I + \mathcal{G}_0 + \breve{\mathcal{G}})} \leq \frac{1}{\underline{\sigma}(I + \mathcal{G}_0) - \overline{\sigma}(\breve{\mathcal{G}})} \leq \frac{M_{S0}}{1 - M_{S0}\overline{\sigma}(\breve{\mathcal{G}})}. \end{aligned}$$

This inequality together with $||\breve{\mathcal{G}}||_\infty = ||\breve{G}||_\infty$ show (A.6). □

B

The LLRF System for ESS

The LLRF system for the ESS linac will be based on components that are compliant with the MicroTCA.4 standard [MTCA.4 2017], which was developed as a subspecification of the MicroTCA standard [Jamieson, 2006], to allow a flexible, carrier-grade platform for big science projects.

Table B.1 Components of the ESS LLRF system.

Component	**Model/Comments**
12-slot chassis	nVent Schroff 11850-026
MicroTCA Carrier Hub	NAT-MCH-PHYS
Power Supply Unit	Wiener MTCA.4 1000 W PS
Digitizer (AMC*)	Struck SIS8300-KU Kintex Ultrascale FPGA 10 ADCs, 16-bit, 125 MS/s 2 high speed DACs 2 GByte DDR4 Memory
Downconversion & vector modulator (RTM†)	Struck DWC8VM1(-LF) 8 down-conversion channels 1 up-conversion channel (vector modulator)
Timing receiver (AMC*)	Micro-Research Finland mTCA-EVR-300U
CPU (AMC*)	Concurrent Technologies AM G64/471 99
LO-generation 352.21 MHz (stand-alone unit)	In-kind contribution form ESS Bilbao
LO-generation 704.42 MHz (RTM†)	In-kind contribution from the Polish Electronic Group‡

*Advanced Mezzanine Card, inserted from the front of the chassis.
†Rear Transition Module, inserted from the back of the chassis.
‡Collaboration between National Centre for Nuclear Research, TU Warsaw, and TU Łódź.

C

Derivation: Dynamics of a Cavity Mode

The following derivation of (6.8) is based on [Haus, 1983, Section 7.2], but additionally includes beam loading. We will consider the kth mode but will drop the mode index except in Section C.3 where we also mention the accelerating mode. The starting point of this derivation will be the baseband dynamics (6.7) (relative to the nominal RF frequency ω_{RF}) of a closed and lossless cavity. We restate this equation for convenience;

$$\frac{d}{dt}\mathbf{A} = i\Delta\omega\mathbf{A},$$

where $\Delta\omega := \omega_a - \omega_{\mathrm{RF}}$.

C.1 Waveguide coupling

Next, let us introduce a coupling port to a waveguide. This allows an incident, forward wave in the waveguide to excite the cavity mode, but energy will also escape the cavity through the port, and propagate away in a reverse wave (see Figure 6.1). We will denote the complex-envelopes (with respect to ω_{RF}) of the forward and reverse waves by $\mathbf{F_g}$ and $\mathbf{R_g}$. We also assume them normalized so that $|\mathbf{F_g}|^2$ is the power carried by the forward wave ($\mathbf{F_g}$ has units $\sqrt{\mathrm{W}}$), and similarly for $\mathbf{R_g}$.

Due to the linearity of Maxwell's equations, we have

$$\frac{d}{dt}\mathbf{A} = i\Delta\omega\mathbf{A} - \gamma_{\mathrm{ext}}\mathbf{A} + \kappa_g\mathbf{F_g} \tag{C.1}$$

where γ_{ext} is the decay rate of the cavity field through the coupling port, and κ_g is a, possibly complex-valued, proportionality constant quantifying the effect of the forward wave on the cavity field. It turns out that γ_{ext} and κ_g are in fact related and we next derive how.

C.2 Relation between γ_{ext} and κ_g

Throughout this section we will consider a particular solution to (C.1): from the initial condition $\mathbf{A}(0) = 1$, with $\Delta\omega = 0$ and $\mathbf{F_g}(t) \equiv 0$. It is clear that the solution is given by

$$\mathbf{A}(t) = \mathrm{e}^{-\gamma_{\text{ext}} t}, \qquad t \geq 0.$$

Recall that the stored energy in the cavity equals $|\mathbf{A}|^2$. The stored energy changes by $\frac{d}{dt}|\mathbf{A}|^2 = -2\gamma_{\text{ext}}\mathrm{e}^{-2\gamma_{\text{ext}}t}$, and due to conservation of energy, this power is carried away by the reverse wave, so

$$|\mathbf{R_g}|^2 = 2\gamma_{\text{ext}}\mathrm{e}^{-2\gamma_{\text{ext}}t}. \tag{C.2}$$

If $\boldsymbol{\mathcal{E}}(\boldsymbol{r}, t)$ is a solution to (6.1)—valid in the cavity and the waveguide—then so is the time-reversed solution $\boldsymbol{\mathcal{E}}_r(\boldsymbol{r}, t) \coloneqq \boldsymbol{\mathcal{E}}(\boldsymbol{r}, -t)$.

We next consider the time-reversed version of the particular solution of this section. The mode amplitude for the time-reversed solution is given by

$$\mathbf{A}_r(t) = \mathbf{A}(-t) = \mathrm{e}^{\gamma_{\text{ext}} t}, \quad t \leq 0, \tag{C.3}$$

and the forward wave turns into reverse wave, and vice versa. Thus, for $\mathbf{F}_{\mathbf{g}r}$ (the forward wave in the time-reversed solution) we have that $|\mathbf{F}_{\mathbf{g}r}| = |\mathbf{R_g}|$, which gives

$$\mathbf{F}_{\mathbf{g}r} = \mathrm{e}^{i\phi_0} \cdot \sqrt{2\gamma_{\text{ext}}}\mathrm{e}^{\gamma_{\text{ext}} t} \tag{C.4}$$

for some phase ϕ_0.

Recalling that $\Delta\omega = 0$, and plugging (C.3) and (C.4) into (C.1) gives

$$\gamma_{\text{ext}}\mathrm{e}^{\gamma_{\text{ext}} t} = -\gamma_{\text{ext}}\mathrm{e}^{\gamma_{\text{ext}} t} + \kappa_g \mathrm{e}^{i\phi_0} \cdot \sqrt{2\gamma_{\text{ext}}}\mathrm{e}^{\gamma_{\text{ext}} t},$$

from which it follows that

$$\kappa_g = \mathrm{e}^{-i\phi_0} \cdot \sqrt{2\gamma_{\text{ext}}}.$$

Since we are free to define the reference phase for the forward wave, we may assume $\phi_0 = 0$, and thus get

$$\kappa_g = \sqrt{2\gamma_{\text{ext}}}.$$

$$\frac{d}{dt}\mathbf{A} = i\Delta\omega\mathbf{A} - \gamma_{\text{ext}}\mathbf{A} + \kappa_g \mathbf{F_g}. \tag{C.5}$$

We have now recovered the same cavity equation as in [Haus, 1983]. To model accelerating cavities we also need to include beam loading.

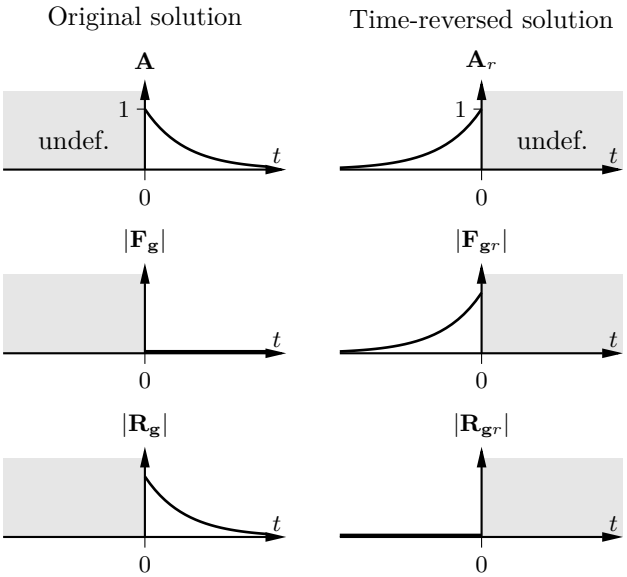


Figure C.1 Illustration of the specific solution considered in Section C.2.

C.3 Beam loading

A sufficiently closely spaced bunch train traversing the cavity can be modeled by a complex envelope $\mathbf{I_b}$ whose magnitude $|\mathbf{I_b}|$ equals the corresponding DC current of the bunch train. The phase of $\mathbf{I_b}$ (i.e., the bunch phase) is defined so that $\angle\mathbf{I_b} = -\pi$ gives maximum acceleration from the nominal field of the accelerating mode (corresponding to $\angle\mathbf{A}_a = 0$).

Introduce the cavity–beam-coupling parameter $\boldsymbol{\alpha}_k$ of the mode[1] according to (cf. [Wangler, 2008])

$$\text{power to beam from mode } k = -\text{Re}\{\boldsymbol{\alpha}_k^* \mathbf{I}_\mathbf{b}^* \mathbf{A}_k\}. \tag{C.6}$$

Note that the cavity–beam-coupling parameter of the accelerating mode is real and non-negative, $\boldsymbol{\alpha}_a = \alpha_a \geq 0$, due to the definition of the $\mathbf{I_b}$.

The bunch train induces an electromagnetic field in the cavity, corresponding to a term $-\mu_0\, d\boldsymbol{J}/dt$, where $\boldsymbol{J}$ is current density, on the right-hand side of (6.1). This effect is linear and assuming for a moment that $\Delta\omega = 0$ and there is no waveguide coupling ($\gamma_{\text{ext}} = 0$, this effect is linear and can be

[1] The cavity–beam-coupling parameter depends on the velocity of the particle bunches and is related to the traditionally used parameter $(r/Q)_k$ (linac convention) as $|\boldsymbol{\alpha}_k| = \sqrt{\omega_k (r/Q)_k}$.

added by the superposition principle) we have that the impact on the kth mode is given by

$$\frac{d\mathbf{A}_k}{dt} = \boldsymbol{\kappa}_{\mathbf{b}k}\mathbf{I}_\mathbf{b}$$

where $\boldsymbol{\kappa}_{\mathbf{b}k}$ is a complex coefficient. Using this expression we have that the energy of mode k changes due to beam loading as

$$\frac{d}{dt}\left|\mathbf{A}_k\right|^2 = 2\mathrm{Re}\{\boldsymbol{\kappa}^*_{\mathbf{b}k}\mathbf{I}^*_\mathbf{b}\mathbf{A}_k\}. \tag{C.7}$$

Conservation of energy yields that (C.6) and (C.7) sum to zero. Since this holds for all $\mathbf{I}_\mathbf{b}$ it follows that $\boldsymbol{\kappa}_{\mathbf{b}k} = \boldsymbol{\alpha}_k/2$.

C.4 Putting it together

By combing the results from the two preceding sections and adding a term $-\gamma_0\mathbf{A}$ for resistive losses (assuming that these do not significantly change the mode distribution) we arrive at

$$\boxed{\frac{d\mathbf{A}}{dt} = (-\gamma + i\Delta\omega)\mathbf{A} + \sqrt{2\gamma_{\mathrm{ext}}}\mathbf{F}_\mathbf{g} + \frac{\boldsymbol{\alpha}}{2}\mathbf{I}_\mathbf{b}}$$

where $\gamma = \gamma_0 + \gamma_{\mathrm{ext}}$. This is exactly (6.8).

C.5 Reverse wave

As in [Haus, 1983, (7.36)] we may derive an expression for the envelope of the reverse wave. From the linearity of Maxwell's equations we know that the reverse wave $\mathbf{R}_\mathbf{g}$ depends linearly on the forward wave and the mode amplitude:

$$\mathbf{R}_\mathbf{g} = c_F\mathbf{F}_\mathbf{g} + c_A\mathbf{A},$$

where c_F and c_A are complex constants. We already know from (C.2) that $|c_A| = \sqrt{2\gamma_{\mathrm{ext}}}$, and by selecting which phase $\mathbf{R}_\mathbf{g}$ is defined with respect to, we can make $c_A = \sqrt{2\gamma_{\mathrm{ext}}}$.

Next, we derive an expression for c_F. Conservation of energy gives that

$$\left|\mathbf{F}_\mathbf{g}\right|^2 - \left|\mathbf{R}_\mathbf{g}\right|^2 = \frac{d}{dt}\left|\mathbf{A}\right|^2, \tag{C.8}$$

and from (C.5) it follows that

$$\frac{d}{dt}\left|\mathbf{A}\right|^2 = -2\gamma_{\mathrm{ext}}\left|\mathbf{A}\right|^2 + \sqrt{2\gamma_{\mathrm{ext}}}\Big(\mathbf{A}^*\mathbf{F}_\mathbf{g} + \mathbf{F}^*_\mathbf{g}\mathbf{A}\Big). \tag{C.9}$$

Putting (C.8) equal to (C.9), and then substituting $\mathbf{A} = (\mathbf{R_g} - c_F\mathbf{F_g})/\sqrt{2\gamma_{\text{ext}}}$ gives

$$\begin{aligned}
|\mathbf{F_g}|^2 - |\mathbf{R_g}|^2 &= -2\gamma_{\text{ext}}\,|\mathbf{A}|^2 + \sqrt{2\gamma_{\text{ext}}}\Big(\mathbf{A}^*\mathbf{F_g} + \mathbf{F_g^*}\mathbf{A}\Big) \\
&= -\Big(|\mathbf{R_g}|^2 - c_F\mathbf{R_g^*}\mathbf{F_g} - c_F^*\mathbf{F_g^*}\mathbf{R_g} + |c_F|^2\,|\mathbf{F_g}|^2\Big) \\
&\quad + \Big(\mathbf{R_g^*}\mathbf{F_g} - c_F^*\,|\mathbf{F_g}|^2 + \mathbf{F_g^*}\mathbf{R_g} - c_F\,|\mathbf{F_g}|^2\Big).
\end{aligned}$$

From this equation it follows that

$$|(1 + c_F)\mathbf{F_g} - \mathbf{R_g}|^2 = |\mathbf{R_g}|^2\,.$$

For this equality to hold for all $\mathbf{F_g}$ and $\mathbf{R_g}$, we must have that

$$c_F = -1.$$

Thus the reverse wave is given, as in [Haus, 1983, (7.36)], by

$$\mathbf{R_g} = -\mathbf{F_g} + \sqrt{2\gamma_{\text{ext}}}\mathbf{A}.$$

D

Derivation: Dynamics of Elliptical Multicell Cavities

Dynamic models of same-order modes have been used for studying how beating of same-order modes affects the voltage seen by the beam [Henke and Filtz, 1993; Ferrario et al., 1996]. Models that capture the shapes of the modes and their resonance frequencies (but not their dynamics and decay rates) have been considered for field-flatness tuning Padamsee et al., 2008, Sec. 7.2; Wangler, 2008; Sekutowicz, 2010; Doolittle, 1989.

For field-control design in the frequency domain it is useful to know the (S21) transfer function from the incident RF drive to the pickup-probe signal. A real-coefficient, two-input two-output transfer function from RF drive to pickup probe (valid for superconducting) cavities was presented without motivation in [Schilcher, 1998]. An incorrect attempt to derive the same transfer function was made in [Vogel, 2007] (see Remark D.1).

In this appendix, we will start from the bandpass state-space model of an N-cell cavity in [Ferrario et al., 1996], perform modal decomposition (diagonalization), and then transform the diagonal model to baseband.

D.1 Bandpass model of an N-cell cavity

Consider the elliptical N-cell cavity in Figure 6.11 which has $N-2$ identical inner cells and two end-cells that are joined with the beam pipe. Cell 1 can be excited through a power coupler connected to the beam pipe. Adjacent cells are connected by irises which allow the electromagnetic field to disperse throughout the cavity. Cell N contains a pickup probe for measuring the electric field.

We will only consider the lowest energy mode in each cell, and denote the electric field amplitudes of these by $x = \begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix}^\mathsf{T}$. We will assume that x_j is normalized so that the squared magnitude of its complex envelope equals the energy stored in cell j. Let the coupling between cells be quantified

by the cell-to-cell coupling factor k_{cc}; the resonance frequencies of the inner cells be given by ω_{cell}; the resonance frequencies of the end cells be given by $\omega_{\text{cell}}\sqrt{1+2k_{cc}}$; the RF drive (i.e., the forward wave entering the power coupler) be modeled by its complex envelope $\mathbf{F_g}$, with $|\mathbf{F_g}|^2$ equaling the power in the forward wave; the coupling between the waveguide and the field in cell 1 be quantified by the decay rate γ_{pc} of the field in this through the power coupler; and assume that field decay through the pickup probe is negligible. See Figure 6.11 for an illustration. This is a rather standard description of a multicell cavity [Padamsee et al., 2008; Sekutowicz, 2010], although the notation is slightly different.

According to [Ferrario et al., 1996, (B-1)] the cell amplitudes x evolve as a chain of weakly coupled oscillators, with the dynamics

$$\ddot{x} + 2\gamma_0\dot{x} + 2\gamma_{\text{pc}}E_1\dot{x} + \omega_{\text{cell}}^2 x + \omega_{\text{cell}}^2 k_{cc}Kx = 2\sqrt{2\gamma_{\text{pc}}}e_1\frac{d}{dt}\text{Re}\{\mathbf{F_g}e^{i\omega_{\text{RF}}t}\}, \quad \text{(D.1)}$$

where

$$K = \begin{bmatrix} 3 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \vdots \\ 0 & 0 & \ddots & & 0 \\ \vdots & & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 3 \end{bmatrix}, \quad \text{(D.2)}$$

$$E_1 = \text{diag}(1, 0, \ldots, 0),$$

$$e_1 = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^{\mathsf{T}}.$$

D.2 Eigenvectors and eigenvalues of the matrix K

By recalling standard trigonometric identities and doing some algebra—alternatively looking up [Padamsee et al., 2008, Sec. 7.2]—it can be verified that the matrix K in (D.2) has the eigenfactorization $Q\Lambda Q^{\mathsf{T}} = K$, where

$$Q \coloneqq \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_N \\ | & | & & | \end{bmatrix}, \quad \text{(D.3)}$$

$$\Lambda \coloneqq \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_N), \quad \text{(D.4)}$$

and the eigenvalues λ_n are given by

$$\lambda_n = 2\left(1 - \cos\frac{n\pi}{N}\right), \quad \text{(D.5)}$$

and the orthonormal eigenvectors q_n are given by

$$q_n = \sqrt{\frac{2}{N}} \begin{bmatrix} \sin\left[(1-\frac{1}{2})\frac{n\pi}{N}\right] \\ \sin\left[(2-\frac{1}{2})\frac{n\pi}{N}\right] \\ \vdots \\ \sin\left[(N-\frac{1}{2})\frac{n\pi}{N}\right] \end{bmatrix}, \quad q_N = \sqrt{\frac{1}{N}} \begin{bmatrix} 1 \\ -1 \\ \vdots \\ (-1)^{N-1} \end{bmatrix}. \tag{D.6}$$
$$(n < N)$$

The entries of the Nth mode have equal magnitude and opposite signs. For this reason, it is almost always used as the accelerating mode.

We will, as in existing literature, refer to the nth mode as the $n\pi/N$ mode, due to the cell-to-cell phase advance. However, to keep the notation manageable, we keep the subscripts n, except for when we wish to emphasize its special nature of the π mode (the Nth mode).

D.3 Diagonalizing the dynamics

With the change of variables $x = Q\xi$ (ξ corresponds to the mode amplitudes in the case of no losses), we can diagonalize all of (D.1) except the third term,

$$\ddot{\xi} + 2\gamma_0\dot{\xi} + 2Q^{\mathsf{T}}E_1Q\gamma_{\mathrm{pc}}\dot{\xi} + \omega_{\mathrm{cell}}^2(I + k_{cc}\Lambda)\xi = 2\sqrt{2\gamma_{\mathrm{pc}}}Qe_1\frac{d}{dt}\mathrm{Re}\{\mathbf{F_g}\mathrm{e}^{i\omega_{\mathrm{RF}}t}\}, \tag{D.7}$$

where I is the identify matrix. For convenience, denote Q's first row times $\sqrt{N}$ by

$$\begin{aligned} R &\coloneqq \begin{bmatrix} R_1 & \cdots & R_n & \cdots & R_N \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{2}\sin\dfrac{\pi}{2N} & \cdots & \sqrt{2}\sin\dfrac{n\pi}{2N} & \cdots & 1 \end{bmatrix}. \end{aligned}$$

Then we can write the third term of (D.7) as $2R^{\mathsf{T}}R/N\gamma_{\mathrm{pc}}\dot{\xi}$. This term, which originates from field decay through the power coupler, dynamically couples the modes. We may however assume that this interaction averages out to 0, since the beating period between the different modes is significantly shorter than the timescales at which the (complex) mode amplitudes change. Thus it suffices to keep the diagonal entries of $R^{\mathsf{T}}R$, which correspond to the external decay rates of the modes. Denoting them by

$$\gamma_{\mathrm{ext}\,n} \coloneqq R_n^2\,\gamma_{\mathrm{pc}}/N, \tag{D.8}$$

we get N uncoupled differential equations from (D.7), one for each mode,

$$\ddot{\xi}_n + 2\gamma_0\dot{\xi}_n + 2\gamma_{\mathrm{ext}\,n}\dot{\xi}_n + \omega_{\mathrm{cell}}^2(1 + k_{cc}\lambda_n)\xi_n \\ = 2\sqrt{2\gamma_{\mathrm{ext}\,n}}\frac{d}{dt}\mathrm{Re}\{\mathbf{F_g}\mathrm{e}^{i\omega_{\mathrm{RF}}t}\}. \quad \text{(D.9)}$$

It can be seen from (D.8) that the external decay rate for the π mode is given by $\gamma_{\mathrm{ext}\,\pi} = \gamma_{\mathrm{pc}}/N$, and that the external decay rates of the parasitic same-order modes ($n < N$) are related to $\gamma_{\mathrm{ext}\,\pi}$ as

$$\gamma_{\mathrm{ext}\,n} = R_n^2\gamma_{\mathrm{ext}\,\pi} = 2\sin^2\frac{n\pi}{2N}\cdot\gamma_{\mathrm{ext}\,\pi} \quad \text{(D.10)}$$

D.4 Baseband state-space model

The eigenfrequencies of the modes in (D.9) are given by

$$\omega_n = \omega_{\mathrm{cell}}\sqrt{1 + k_{cc}\lambda_n}.$$

Denote their offsets from the nominal RF frequency ω_{RF} by $\Delta\omega_n \coloneqq \omega_n - \omega_{\mathrm{RF}}$.

Note that the eigenvalues in (D.5), for $n < N$, can be written $\lambda_n = 2R_n^2$. Assuming that the nominal RF frequency approximately equals the resonance frequency of the π mode, $\omega_{\mathrm{RF}} \approx \omega_\pi = \omega_{\mathrm{cell}}\sqrt{1 + 2\cdot 2k_{cc}}$, then for $n < N$

$$\Delta\omega_n \approx \omega_{\mathrm{cell}}\sqrt{1 + R_n^2\cdot 2k_{cc}} - \omega_{\mathrm{cell}}\sqrt{1 + 2\cdot 2k_{cc}} \approx (R_n^2 - 2)k_{cc}\omega_{\mathrm{cell}}. \quad \text{(D.11)}$$

Introducing $\mathbf{A}_n$ as the complex envelope of the $n\pi/N$ mode we see that the baseband transformation of (D.9) with respect to ω_{RF} is given by[1]

$$\dot{\mathbf{A}}_n = \left[-(\gamma_0 + \gamma_{\mathrm{ext}\,n}) + i\Delta\omega_n\right]\mathbf{A}_n + \sqrt{2\gamma_{\mathrm{ext}\,n}}\mathbf{F_g}. \quad \text{(D.12)}$$

By introducing $\underline{\mathbf{A}} \coloneqq \begin{bmatrix}\mathbf{A}_1 & \cdots & \mathbf{A}_N\end{bmatrix}^{\mathsf{T}}$, $\Delta\Omega \coloneqq \mathrm{diag}(\Delta\omega_1, \ldots, \Delta\omega_n)$, and $\Gamma_{\mathrm{ext}} \coloneqq \mathrm{diag}(\gamma_{\mathrm{ext}\,1}, \ldots, \gamma_{\mathrm{ext}\,N})$, we can write the equations (D.12) as

$$\dot{\underline{\mathbf{A}}} = \left[-(\gamma_0 I + \Gamma_{\mathrm{ext}}) + i\Delta\Omega\right]\underline{\mathbf{A}} + \sqrt{2\gamma_{\mathrm{ext}\,\pi}}R^{\mathsf{T}}\mathbf{F_g}. \quad \text{(D.13)}$$

The voltage signal $\mathbf{V}_{\mathrm{pu}}$ from the pickup probe is proportional to the field amplitude in cell N, $\mathbf{V}_{\mathrm{pu}} \propto Q_{N:}\underline{\mathbf{A}}$ where $Q_{N:}$ is the Nth row of Q. Noting that[2] $\sqrt{N}Q_{Nn} = (-1)^{n-1}R_n$, we can write

$$\mathbf{V}_{\mathrm{pu}} = C\underline{\mathbf{A}},$$

[1] A reasonable baseband approximation of $2s/(s^2 + 2\zeta_0\omega_0 s + \omega_0^2)$ with respect to the frequency ω_{RF} is $1/(s + \zeta_0\omega_0 - i(\omega_0 - \omega_{\mathrm{RF}}))$.

[2] $Q_{Nn} = \sin\left[(N - \frac{1}{2})\frac{n\pi}{N}\right] = \sin\left[n\pi - \frac{n\pi}{2N}\right] = (-1)^{n-1}\sin\frac{n\pi}{2N} = (-1)^{n-1}Q_{1n} = (-1)^{n-1}R_n/\sqrt{N}$.

where

$$C = \begin{bmatrix} c_1 & \cdots & c_{N\text{-}1} & c_\pi \end{bmatrix} \coloneqq \kappa_{\mathrm{pu}} R \begin{bmatrix} (-1)^{N-1} & & & \\ & \ddots & & \\ & & -1 & \\ & & & 1 \end{bmatrix}, \tag{D.14}$$

and κ_{pu} is a proportionality constant (which can be assumed to be real since the phase that $\mathbf{V}_{\mathrm{pu}}$ is defined with respect to can be chosen freely).

Combining (D.13) with (D.14) and also including cavity–beam interaction, quantified by parameters $\boldsymbol{\alpha}_n$ as in Section 6.8, we have the following state-space realization for the dynamics of the same-order modes

$$\dot{\underline{\mathbf{A}}} = \boldsymbol{\mathcal{A}}\underline{\mathbf{A}} + B_{\mathrm{g}}\mathbf{F_g} + \boldsymbol{B}_{\mathrm{b}}\mathbf{I_b} \tag{D.15}$$

$$\mathbf{V}_{\mathrm{pu}} = C\underline{\mathbf{A}} \tag{D.16}$$

where

$$\begin{aligned} \boldsymbol{\mathcal{A}} &= -(\gamma_0 I + \Gamma_{\mathrm{ext}}) + i\Delta\Omega \\ B_{\mathrm{g}} &= \sqrt{2\gamma_{\mathrm{ext}\,\pi}}R^{\mathsf{T}} \\ \boldsymbol{B}_{\mathrm{b}} &= \frac{1}{2}\begin{bmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_{N\text{-}1} & \alpha_\pi \end{bmatrix}^{\mathsf{T}} \\ C &= \text{given by (D.14)}\,. \end{aligned}$$

Remark D.1 (Previous transfer-function models) In [Schilcher, 1998] it was correctly stated (but without motivation) that the same-order modes of a superconducting cavity have the dynamics (6.36). This is also consistent with Figure 5.9 in [Liepe, 2001], which indicates equal peak magnitudes of $|P_{\mathrm{cav}}(i\omega)|$. However, in [Vogel, 2007] the mode terms in (6.36) are additionally multiplied by their coupling strengths to the RF system (essentially R_n^2) although these factors are already included in the numerators of middlemost expression of (6.35).

E

Controller Optimization

E.1 Design of optimal linear controllers

Approximately optimal LTI controllers can often be found by utilizing the Youla parametrization [Youla et al., 1976], whereby all internally stable closed-loop transfer functions are parametrized by a stable transfer function $Q(s)$. If the objective and constraints are convex functions of the closed-loop transfer functions, which they often are, it is possible to find an approximate solution with standard convex solvers [Boyd and Barratt, 1991]. This method is called Q design; see [Hespanha, 2009] for a practical implementation in the modeling system `cvx` [CVX Research Inc, 2012].

The basic ideas of Q design are straight-forward, but it is common to encounter various numerical issues for all but the smallest problems [Kjellqvist and Troeng, 2019]. Some key ideas for simple and successful implementation of Q design were presented in [Troeng, 2017]. For the problem that we consider in Chapter 9 it is sufficient with the following version of the Youla parametrization.

Youla Parametrization (Simplified Version)
Assume that the SISO system $P(s)$ is stable, then

$$C(s) = \frac{Q(s)}{1 - P(s)Q(s)}$$

is an internally stabilizing controller for $P(s)$, if and only if $Q(s)$ is stable.□

With this parametrization, all closed-loop transfer functions become affine in $Q(s)$, for example, the sensitivity function $S = (1 - PQ)$, and we can re-write the control problem (9.3) as

$$\begin{aligned} \underset{Q}{\text{minimize}} \quad & ||P_a(1 - PQ)D||_2 && \text{(E.1a)} \\ \text{subject to} \quad & ||1 - PQ||_\infty \leq S_{\max}. && \text{(E.1b)} \\ & ||QN||_2 \leq B_{un}. && \text{(E.1c)} \end{aligned}$$

Note that the objective and the constraints are convex in Q.

Obtaining a finite-dimensional problem

To solve the problem (E.1) numerically, Q is approximated using a basis $\{Q_k\}_{k=1}^N$ of stable transfer functions,

$$Q = \sum_{k=1}^{N} \beta_k Q_k. \tag{E.2}$$

The norms in (E.1) are furthermore evaluated over a discrete frequency grid $\Omega = \{\omega_m\}_{\ell=1}^M$. To allow for arbitrary disturbance spectra $D(i\omega)$, perhaps *empirically* measured, the 2-norms in (E.1) can be expressed as frequency-domain integrals

$$||G||_2^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |G(i\omega)|^2 \, d\omega \approx \frac{1}{2\pi} \sum_{m=1}^{M-1} |G(i\omega_m)|^2 \, \Delta\omega_m, \tag{E.3}$$

where $\Delta\omega_m = \omega_{m+1} - \omega_m$. In principle we could have selected the frequency grids differently for each of the three norms in (E.1). For example, the sensitivity constraint typically only needs to be enforced around the bandwidth of the closed-loop system.

Practical considerations

Basis functions. The basis functions were selected according to

$$Q_k(t) = \begin{cases} 1 & \text{if } t \in [t_k, \, t_{k+1}), \\ 0 & \text{otherwise} \end{cases} \tag{E.4}$$

see Figure E.1. This is similar to the choice in [Garpinger, 2009] where a discrete-time formulation was used (corresponding to $t_{k+1} - t_k = h$). However, allowing the interval length to be adjusted makes it possible to increase the resolution of the impulse response of Q for small times t, where fast variations typically occur. Also note that the basis functions Q_k are orthogonal in $\mathcal{H}_2$.

Constraint on control signal activity. If the measurement noise is white, i.e., $N(i\omega) = N(0)$, and if the basis functions Q_k are orthogonal, the constraint on control signal activity (E.1c) can be expressed as

$$||QN||_2^2 = N(0)^2 \, ||Q||_2^2 = N(0)^2 \sum_{n=1}^{N} |\beta_n|^2 \, ||Q_n||_2^2 \leq B_{un}^2. \tag{E.5}$$

This expression can be seen as a regularization constraint on the coefficient vector $\{\beta_n\}$. Numerical experiments indicates that a constraints of this form alleviates some of the numerical issues in Q design [Troeng, 2017]. A constraint of the form (E.5) was also likely contributing to the successful solution of the Q design problem in [Gonzalez-Cava et al., 2019].

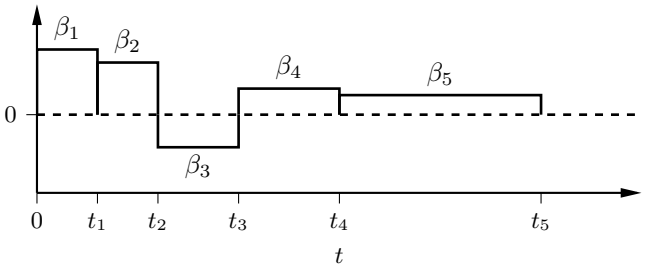


Figure E.1 Illustration of the impulse response of $Q = \sum \beta_k Q_k$ when the basis functions are of the form (E.4).

E.2 Design of low-order controllers

It is challenging to hand-tune controller parameters well enough to make a fair comparison between different controller structures. Hence, locally optimal controller parameters for the different controller structures considered in Chapters 9 and 11 were found by local, gradient-based search with the MATLAB function `fmincon`. Since the problem is non-convex there are no guarantees that optimal parameters are found. The command `Multistart` was used for running multiple local searches from different starting points to increase the likelihood of finding good parameter values.

F

Anti-Windup for Complex-Coefficient Systems

For feedback controllers with integral action, a notorious effect called *integrator windup* can arise when the actuator is driven into saturation [Åström and Hägglund, 2006]. It is caused by integration of a persistent non-zero control error during a period of actuator saturation. When the control error eventually changes sign, it takes significant time before the controller begins to counter-act the error, due to the large value of the integral part.

For most field control loops, integrator windup should not be an issue since all but the smallest control errors are unacceptable and care is taken to have sufficient overhead to avoid RF amplifier saturation. However, it is good practice to always implement some means of anti-windup for controllers with integral action. Therefore, we briefly discuss anti-windup of field control.

In the standard (real-valued) single-input single-output setting, there are several anti-windup strategies to choose from [Åström and Hägglund, 2006]. One simple approach called clamping, is to stop increasing the I part when the control signal saturates. Another approach is back-calculation [Åström and Hägglund, 2006] (Figure F.1), which directly generalizes to the complex setting if saturation of a complex signal is defined by saturating the magnitude and leaving the phase unchanged,

$$\mathrm{sat}(\boldsymbol{u}(t)) = \begin{cases} \boldsymbol{u}(t) & \text{if } |\boldsymbol{u}(t)| \leq 1 \\ \boldsymbol{u}(t)/\,|\boldsymbol{u}(t)| & \text{if } |\boldsymbol{u}(t)| > 1. \end{cases} \tag{F.1}$$

A nice feature of back-calculation in the complex setting, is that the phase of the resulting control signal $\boldsymbol{u}_{\mathrm{sat}}(t)$ equals that of the nominal control signal $\boldsymbol{u}(t)$. This seems tricky to achieve with clamping-based approaches to anti-windup.

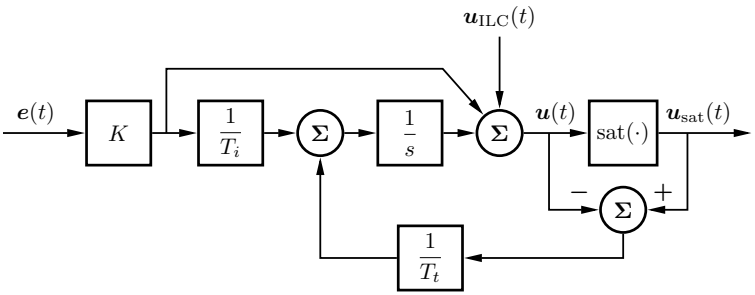


Figure F.1 Illustration of complex-valued anti-windup based on back-calculation for a nominal PI controller given by $C(s) = K\,(1 + 1/(sT_i))$. The saturation function is defined in (F.1). The tracking time constant T_t should be smaller than the integral time T_i [Åström and Hägglund, 2006]. The signal $\boldsymbol{u}_{\mathrm{ILC}}(t)$ is the feedforward term that is updated by the ILC algorithm.

G

Proofs

G.1 Proof of Theorem 2

The Youla parametrization [Youla et al., 1976] gives that all internally stable closed-loop transfer functions with a widely linear controller (Appendix A) from a set of external inputs w to a set of controlled variables z can be parameterized as

$$G_{zw} = H_1 + H_2 \left[Q, \breve{Q} \right] H_3$$

where $H_1, H_2, H_3 \in \mathcal{H}_\infty$ and $[Q, \breve{Q}]$ is a stable widely linear system.

Let $[Q^\star, \breve{Q}^\star]$ be the Q parameter corresponding to a controller that achieves optimal performance (and satisfies the constraints). We will now show, reasoning as in [Bamieh et al., 2002], that just the strictly linear part $Q^\star$, achieves the same performance (and also satisfies all constraints).

The transfer function from a subset w' of the external signals, to a subset z' of the controlled signals can be written

$$G_{z'w'} = H_1' + H_2' \left[Q^\star, \breve{Q}^\star \right] H_3'.$$

where H_k' is a submatrix of H_k.

For any widely linear system $[G, \breve{G}]$ and rotationally-invariant systems norm $||\cdot||$ we have

$$\left|\left|\left[G, \breve{G} \right]\right|\right| = \left|\left| -iI \left[G, \breve{G} \right] iI \right|\right| = \left|\left|\left[G, -\breve{G} \right]\right|\right|, \tag{G.1}$$

where I denotes the identity system. The first equality follows from the rotational invariance of the norm and the second equality follows from (A.4).

Using (G.1) we have that

$$\begin{aligned}
&\left\|H_1' + H_2'\left[Q^\star, \breve{Q}^\star\right] H_3'\right\| \\
&\quad = \frac{1}{2}\left(\left\|H_1' + H_2'\left[Q^\star, \breve{Q}^\star\right] H_3'\right\| + \left\|H_1' + H_2'\left[Q^\star, -\breve{Q}^\star\right] H_3'\right\|\right) \\
&\quad \geq \left\|H_1' + H_2'\left[Q^\star, \frac{\breve{Q}^\star - \breve{Q}^\star}{2}\right] H_3'\right\| = \|H_1' + H_2' Q^\star H_3'\| \qquad \text{(G.2)}
\end{aligned}$$

This shows that better performance cannot be achieved with the widely linear Q-parameter $[Q^\star, \breve{Q}^\star]$, than with the strictly linear $Q^\star$. □

G.2 Proof of Theorem 3

From Theorem 2 on p. 52 it follows that optimal performance is attained by a strictly linear $Q^\star$. Similarly to the preceding proof of Theorem 2, averaging this $Q^\star = Q^\star_{\mathrm{Re}} + iQ^\star_{\mathrm{Im}}$ with its complex conjugate $(Q^\star)^* = Q^\star_{\mathrm{Re}} - iQ^\star_{\mathrm{Im}}$ (defined in (4.10)) shows that also $Q^\star_{\mathrm{Re}}$ attains optimal performance. □

G.3 Theorem on optimal control of first-order systems

THEOREM 5
Consider the optimal control problem

$$\begin{aligned}
\underset{u,\, t_f}{\text{minimize}} \quad & \int_0^{t_f} r(u(t))\, dt && \text{(G.3a)}\\
\text{subject to} \quad & \dot{x}(t) = f(x(t), u(t)) && \text{(G.3b)}\\
& x(0) = 0 && \text{(G.3c)}\\
& x(t_f) = 1 && \text{(G.3d)}\\
& u(t) \in \mathcal{U}, && \text{(G.3e)}
\end{aligned}$$

where $\mathcal{U}$ is a compact set, $r(u) > 0$ for all $u \in \mathcal{U}$, $f(x,u)$ and $r(u)$ are continuous functions of u. We will assume that there is a constant $c > 0$ such that

$$\forall x \in [0,1] \;\; \exists u \in \mathcal{U} \text{ so that } f(x,u) \geq c. \qquad \text{(G.4)}$$

Define

$$u^\star(x) := \underset{u \in \mathcal{U}}{\operatorname{argmax}} \; \frac{f(x,u)}{r(u)}. \qquad \text{(G.5)}$$

and assume that $u^\star$ is sufficiently well-behaved for $\dot{x} = f(x, u^\star(t))$ to have a unique solution $x^\star(t)$. Then the optimal control signal is given by $u(t) = u^\star(x^\star(t))$.

Proof The following is a self-contained proof based on the Hamilton–Jacobi–Bellman principle, cf. Example 5.1 in [Liberzon, 2012].

Define the optimal cost-to-go function

$$V(x) := \int_x^{t_f} \frac{r(u^\star(x'))}{f(x', u^\star(x'))}\, dx'.$$

Let u be an arbitrary control signal with $u(t) \in \mathcal{U}$, such that the corresponding state trajectory x satisfies (G.3b)–(G.3d). It then holds that

$$\begin{aligned} r(u) + \frac{d}{dt}V(x(t)) &= r(u) + \frac{dV}{dx} f(x,u) \\ &= r(u) - \frac{r(u^\star(x))}{f(x,u^\star(x))} f(x,u) \\ &= \frac{r(u) r(u^\star(x))}{f(x,u^\star(x))} \left(\frac{f(x,u^\star(x))}{r(u^\star(x))} - \frac{f(x,u)}{r(u)} \right) \geq 0, \end{aligned} \quad \text{(G.6)}$$

where the inequality follows from $f(x, u^\star(x)) > 0$, $r(u) > 0$ and the definition (G.5) of $u^\star$. Equality holds for $u = u^\star$.

Integration of (G.6) gives, since $x(0) = 0$ and $x(t_f) = 1$,

$$\int_0^{t_f} r(u(t))\, dt \geq V(x(0)) - V(x(t_f)) = V(x(0)),$$

with equality for $u = u^\star$. This proves optimality of $u^\star$. □

Remark G.1 The assumption (G.4) guarantees finite-time feasibility.

Remark G.2 The maximum in (G.5) exists since a continuous function is optimized over a compact set. If several u maximize the expression, any can be chosen.

Remark G.3 It is clear that the functions $r(u)$ and $f(x,u)$ considered in the optimal control problem (13.9) give a well-behaved $u^\star$.

Remark G.4 A constraint $t_f \leq t_{\max}$ on the final time can be handled by adding a constant term to $r(u)$, and doing a binary search over that constant.

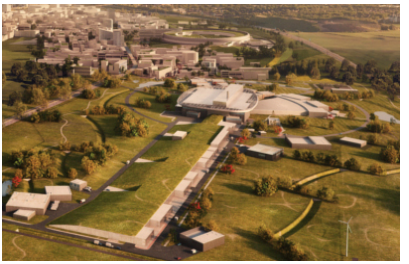
Popular science summary (in Swedish)

Reglering av elektriska fält i linjära partikelacceleratorer

Olof Troeng
Institutionen för Reglerteknik

Populärvetenskaplig sammanfattning av doktorsavhandlingen *Cavity Field Control for Linear Particle Accelerators*, November 2019.
Avhandlingen kan laddas ner från: `https://portal.research.lu.se`

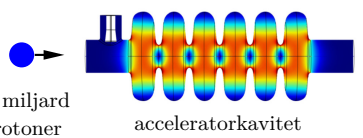
Världens starkaste neutronmikroskop, ESS (European Spallation Source), håller just nu på att byggas utanför Lund i ett samarbete mellan tretton europeiska länder. ESS kommer att bidra till förståelsen och utvecklingen av framtidens läkemedel, solceller, bränsleceller och tekniska material.



ESS fungerar ungefär som en röntgenmaskin, men neutronerna avslöjar detaljer som är osynliga för röntgenstrålar. Bildrättigheter: ESS.

Neutronerna kommer att frigöras genom att wolframatomer krossas med protoner som accelererats till nära ljusets hastighet. För att få skarpa bilder behövs många neutroner, och alltså även många protoner. Små moln med cirka en miljard protoner kommer att accelereras av växlande elektriska fält inneslutna i acceleratorkaviteter. För att protonerna ska accelereras på rätt sätt måste styrkan på de elektriska fälten vara rätt inom 0.1 %. Det är också nödvändigt att protonmolnen och de elektriska fälten är synkroniserade så när som på en biljondels sekund. För att klara dessa krav kommer styrkan och fasen på de elektriska fälten att mätas och justeras miljontals gånger i sekunden av snabba regulatorer.

Avhandlingen belyser olika aspekter av hur man kan förbättra och lättare designa dessa regulatorer. Ett genomgående tema är att förståelsen underlättas om man ser det som att reglersystemet arbetar med komplexvärda signaler, dvs. signaler med både realdel och imaginärdel. Med detta synsätt får systemet *en* mätsignal och *en* styrsignal, vilket underlättar förståelsen jämfört med det traditionella synsättet med två reella mätsignaler och två reella styrsignaler.



Protonerna på ESS kommer att accelereras av växlande elektriska fält som skjutsar på protonerna när de passerar. De elektriska fälten kommer att vara inneslutna i acceleratorkaviteter som den på bilden. Bildrättigheter: G. Constanza.

Idén att representera två signaler med *en* komplexvärd signal är vanlig inom telekommunikation, men det finns mycket få exempel på reglersystem där denna representation är meningsfull. På grund av detta har praktiska och teoretiska aspekter av reglerdesign för komplexvärda system varit relativt outforskade. I avhandlingen utvecklas därför viss ny och grundläggande teori för sådana system.

Nyttan av en komplexvärd systembeskrivning exemplifieras av att den förenklat härledningen av ett flertal av avhandlingens nya resultat. Bland dessa kan nämnas en intuitiv designmetod för att minska verkan av periodiska störningar från t.ex. förstärkarelektronik. Ett annat exempel är en förenkling av de digitala filter som förhindrar att oönskade elektriska svängningsmoder ger instabilitet i regleringen.

De nya resultat och intuitiva synsätt som presenteras i avhandlingen kommer att underlätta förståelsen för regleringen av de elektriska fälten på ESS och i världens många andra partikelacceleratorer.