Particle Acceleration and Detection

Harald Klingbeil
Ulrich Laier
Dieter Lens

Theoretical Foundations of Synchrotron and Storage Ring RF Systems

Springer

Theoretical Foundations of Synchrotron and Storage Ring RF Systems

Particle Acceleration and Detection

springer.com

The series *Particle Acceleration and Detection* is devoted to monograph texts dealing with all aspects of particle acceleration and detection research and advanced teaching. The scope also includes topics such as beam physics and instrumentation as well as applications. Presentations should strongly emphasize the underlying physical and engineering sciences. Of particular interest are

- contributions which relate fundamental research to new applications beyond the immediate realm of the original field of research
- contributions which connect fundamental research in the aforementioned fields to fundamental research in related physical or engineering sciences
- concise accounts of newly emerging important topics that are embedded in a broader framework in order to provide quick but readable access of very new material to a larger audience

The books forming this collection will be of importance for graduate students and active researchers alike.

For further volumes:
http://www.springer.com/series/5267

Harald Klingbeil • Ulrich Laier • Dieter Lens

Theoretical Foundations of Synchrotron and Storage Ring RF Systems

Springer

Harald Klingbeil
Dreieich-Offenthal
Germany

Ulrich Laier
Messel
Germany

Dieter Lens
Darmstadt
Germany

Open Access funded by SCOAP[3]

ISSN 1611-1052
ISBN 978-3-319-07187-9 ISBN 978-3-319-07188-6 (eBook)
DOI 10.1007/978-3-319-07188-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014948728

Preface

This book is based on a course that one of the authors (Klingbeil) has been offering at the Technische Universität Darmstadt, Germany, since 2007. It is addressed to graduate students who intend to study accelerator physics and engineering or who would like to deepen their knowledge in this area, especially in the field of synchrotron and storage ring RF (radio frequency) systems. Furthermore, this book should provide a basis for understanding more advanced accelerator physics literature. In this context, it is especially suitable for engineering students who are interested in a deeper background in physics (e.g., nonlinear dynamics) and for physics students who want to become familiar with engineering aspects (e.g., closed-loop control).

There are many excellent books on accelerator physics. Most of them have a strong focus on transverse beam dynamics, while this book takes a completely different approach. The main topics are RF systems and longitudinal beam behavior, and both theoretical background and practical aspects are discussed. However, the main objective is to provide a solid theoretical basis for further studies.

We did not try to present a complete description of the field. Instead, we selected several specific topics that in our opinion will lead to a deep understanding of the basic facts.

Much effort was expended to maintain a consistent engineering notation throughout the book. In order to achieve this aim, it was sometimes necessary to use symbols that differ from those frequently used in some physics disciplines.

This book was strongly influenced by our work in the Ring RF department of the GSI Helmholtzzentrum für Schwerionenforschung GmbH and at the Technische Universität Darmstadt. Therefore, writing it would not have been possible without many fruitful discussions that we had with our colleagues at GSI, at the university, and at other institutions. Specifically, we would like to thank Priv.-Doz. Dr. Peter Hülsmann, Dr. Hans Günter König, Dr. Gerald Schreiber (all with GSI), Prof. Dr. Jürgen Adamy, and Prof. Dr. Thomas Weiland (both at TU Darmstadt). We are also grateful to the former staff of the GSI ring RF group and the accelerator division, especially Dr. Norbert Angert, Dr. Klaus Blasche, Dipl.-Phys. Martin Emmerling, Dr. Bernhard Franzke, and Dr. Klaus Kaspar. Last but

not least, we thank our managing editor Dr. Christian Caron, the external referees (especially Dr. Frank Zimmermann), and all those involved at Springer for their friendly and efficient collaboration and helpful comments.

Any errors that may be present in the text are, of course, our own responsibility.

Dreieich-Offenthal, Germany — Harald Klingbeil
Messel, Germany — Ulrich Laier
Darmstadt, Germany — Dieter Lens
March 2014

Contents

Formula Symbols

Latin symbols: symbol, description, example for chapter (no.)

A

a	Real number, constant	(2, 5, 6)
a	Principal axis	(5)
a_n	Real Fourier components	(2)
a_0, a_1, a_2, a_m	Constants	(7)
$a_k(t)$	Functions, ODE coefficients	(2, 5)
a_{ik}	Elements of matrix A	(2)
a_x, a_y	Moments of particle ensemble	(5)
A	Constant, parameter, substitution variable	(2, 3, 6)
A, ∂A	Two-dimensional domain and its boundary	(2, 4)
$A(t)$	Time-dependent area	(2)
A	Mass number (integer)	(2)
A	Matrix	(2, 7)
A_c	Matrix (dynamic output feedback)	(7)
A_cl	Matrix (closed-loop dynamics)	(7)
$\vec{A}$	Vector potential	(2, 4)
A_c	Cathode surface	(6)
A_r	Relative mass factor $= \frac{m_0}{m_\mathrm{u}}$	(2, 3)
A_0	Initial area enclosed by a trajectory	(5)
A_1, A_2	Areas enclosed by trajectories	(5)
A'	Deformed area $A(t + \Delta t)$	(2)
$A_{\mathrm{B,stat}}^{\Delta\varphi_\mathrm{RF},\Delta W}$	Bucket area in phase space $(\Delta\varphi_\mathrm{RF}, \Delta W)$, stationary case	(3)

$A_{\mathrm{B,stat}}^{\Delta t,\Delta W}$	Bucket area in phase space $(\Delta t, \Delta W)$, stationary case	(3, 5)
$A_{\mathrm{b,DC}}$	Bunch area of coasting beam, long. emittance	(3, 5)
$A_{\mathrm{B}}^{\Delta t,\Delta W}$	Bucket area in phase space $(\Delta t, \Delta W)$	(3, 5)
$A_{\mathrm{B}}^{\Delta \varphi,\delta}$	Bucket area in phase space $(\Delta \varphi, \delta)$	(5)
A_{R}	Richardson constant	(6)
A_z^E	Long. component of vector potential (TM mode)	(4)
A_z^H	Long. component of vector potential (TE mode)	(4)
AM	Amplitude margin	(7)
ΔA	Surface area of volume element	(2)
$\mathrm{d}A$	Infinitesimal area	(2, 4)
$\mathrm{d}\vec{A}$	Normal vector of $\mathrm{d}A$	(2, 4)

B

$b(t)$	Function, ODE coefficient	(2)
b	Constant	(5)
b	Principal axis	(5)
b_0, b_1, b_2, b_m	Constants	(7)
b_n	Real Fourier components	(2)
$\vec{b}_1, \vec{b}_2$	Input vectors	(5)
B	Matrix	(2, 7)
B	Constant	(2, 6)
B	Induction/magnetic field	(4, 5)
$B(t)$	Magnetic field ramp	(3, 5)
$\vec{B}$	Magnetic induction vector, magnetic field vector	(1, 2, 4, 5)
B_{c}	Matrix (dynamic output feedback)	(7)
B_{extr}	Magnetic field at extraction energy	(5)
B_{f}	Bunching factor	(2, 5)
B_{inj}	Magnetic field at injection energy	(5)
B_{max}	Maximum magnetic field	(2, 5)

B_r	Residual induction	(4)
B_n	Magnetic field of dipole magnets during turn n	(3)

C

c	Speed of light	(2)
c	Constant	(5)
$\vec{c}$	Vector of constants	(2)
c_n	Complex Fourier components	(2, 5)
c_0	Speed of light in vacuum	(2, 3, 5)
$c_1, \ldots, c_n$	Components of $\vec{c}$	(2)
c_1, c_2	Constants	(7)
C	Curve	(2)
C	Constant	(2)
C	Capacitance	(2, 4)
C	Matrix	(7)
C_c	Matrix (dynamic output feedback)	(7)
C_k	Constants	(2)
C^k	Class of k-times continuously differentiable functions	(2)

D

d	Length, distance	(3, 6)
d_core	Core thickness	(4)
d_n	Fourier amplitudes	(2)
$D(\Delta u)$	Probability of interval $\pm\Delta u$	(2)
$D, \partial D$	Domain and its boundary	(2)
D_c	Matrix (dynamic output feedback)	(7)
$\vec{D}$	(Electric displacement) field vector	(2, 5)
$D_A^\lambda(t)$	Diagonal matrix	(2)
$D_B^\lambda(t)$	Diagonal matrix	(2)
D_n	Normal component of $\vec{D}$	(5)
D_n^F	Determinant with system order n	(2)
$D\vec{v}$	Jacobian matrix	(2)

$D\vec{v}(\vec{r}_\mathrm{F})$	Jacobian matrix at $\vec{r} = \vec{r}_\mathrm{F}$	(2)
D_ρ, D_φ, D_z	Components of $\vec{D}$ in cylindrical coordinates	(5)

E

e	Elementary charge	(2, 6)
$\vec{e}_x$, $\vec{e}_y$, $\vec{e}_z$	Cartesian unit vectors	(2)
$\vec{e}_r$	Radial unit vector	(3)
$\mathrm{E}(\gamma, k)$	Elliptic integral of the second kind	(2, 3)
$\mathrm{E}(k)$	Complete elliptic integral of the second kind	(3)
$\hat{E}$	Amplitude of electric field	(4)
$\vec{E}$	(Electric) field vector	(1, 2, 4, 5)
$E(g(X))$	Expected value of $g(X)$	(2)
$E(X^k)$	kth moment	(2)
E_surf	Electric field on the cathode surface	(6)
E_ρ, E_φ, E_z	Components of electric field vector $\vec{E}$ in cylindrical coordinates	(4)

F

f	Frequency $= \frac{\omega}{2\pi}$	
$f(\cdot)$	Function, probability density function	(2, 3, 5, 7)
f_res	Resonant frequency	(4)
f_k	Samples of function f at $k\Delta t$	(2)
f_max	Maximum frequency	(2)
f_sampl	Sampling frequency $\frac{1}{\Delta t}$	(2)
f_DC	DC component of $f(t)$	(2)
f_R	Revolution frequency of reference particle	(1, 3, 4, 5)
f_RF	RF frequency $= hf_\mathrm{R}$	(1, 3, 4, 5, 7)
$f_\mathrm{S,0}$	Synchrotron frequency for small oscillation amplitudes	(3, 5)
$f_\mathrm{S,0,stat}$	Synchrotron frequency for small oscillations, stationary case	(3)

$f_{S,1}$	Synchrotron frequency before voltage increase	(5)
$f_{S,2}$	Synchrotron frequency after voltage increase	(5)
f_T	Transit time factor	(4)
f_1, f_2	Constants	(5)
Δf_{3dB}	3 dB bandwidth	(4)
$\vec{f}(\cdot)$	Vector function	(5)
$\vec{F}$	Force	(1, 3)
$F(s)$	Laplace transform of $f(t)$	(2)
$F(\cdot)$	Function	(2)
$\mathrm{F}(\gamma, k)$	Elliptic integral of the first kind	(2, 3)
$F(W)$	Distribution function	(6)
$\vec{F}(\cdot)$	Map, vector function	(2, 3)
F_R	Accelerating force	(3)
$F_1(q, Q, t)$	Generating function	(5)
$F_2(q, P, t)$	Generating function	(5)
$F_3(p, Q, t)$	Generating function	(5)
$F_4(p, P, t)$	Generating function	(5)

G

$g(\cdot)$	Function	(2, 3, 5)
g	Standard acceleration of gravity	(2, 5)
g_0	Geometry factor	(5)
g_1, g_2	Constants	(5)
$G(s)$	Laplace transform of $g(t)$	(2)
G_{Vgain}	Gain of driver and tetrode amplifiers	(7)

H

h	Harmonic number	(1, 3, 4, 5)
$h(t)$	Function, idealized beam signal	(2)
$h(t)$	Function, impulse response	(7)
$h(\cdot)$	Homeomorphism	(2)
h	Height	(2, 5)
h	Planck's constant	(6)

h_b	Number of bunches	(3)
Δh	Height of volume element	(2)
$\vec{H}$	Magnetic field vector	(1, 2, 4)
$H(s)$	Laplace transform of $h(t)$	(2, 7)
$H(q, p, t)$	Hamiltonian	(2)
H	Value of Hamiltonian	(3)
$H(q, p, \lambda)$	Hamiltonian with parameter	(5)
H	Magnetizing field	(4)
H_bias	DC bias field/average ring core bias field	(4)
H_c	Coercive magnetizing field	(4)
$H_\mathrm{dy}(s)$	Disturbance to output transfer function	(7)
$H_i(j\omega)$	Frequency responses ($i = 1, 2, \ldots$)	(7)
$H_\mathrm{open}(s)$	Open-loop transfer function	(7)
$H_\mathrm{p}(s)$	Process/plant transfer function	(7)
$H_\mathrm{open,delay}$	Open-loop transfer function with delay	(7)
$H_\mathrm{m}(s)$	Measurement transfer function	(7)
$H_\mathrm{ry}(s)$	Reference to output transfer function	(7)
H_s	Value of Hamiltonian at saddle point	(3)
$H_\mathrm{s}(s)$	Sensitivity function	(7)
H_t	Tangential magnetic field	(4)
H_ν	Hurwitz determinants	(7)
H_0	Magnetic field for current I_0	(2)

I

i	Index, integer	(2, 7)
I	Current	(2)
I	Unit matrix	(2, 7)
$I(q_k, p_k)$	Constant of motion	(2)
I_a	Anode current	(7)
$I_\mathrm{beam}(t)$	Beam signal, beam current	(2, 3, 4, 5)
I_max	Maximum current	(2)
I_bias	Bias current	(4)

$\bar{I}_{\text{beam}}$	Average (DC) beam current	(2, 5)
$I_{\text{beam,max}}$	Maximum beam current	(2, 5)
I_L	Inductive current	(4)
I_{tot}	Total current	(4)
I_C	Capacitive current	(4)
I_{cav}	Current from generator to cavity	(4)
I_{gen}	Generator current	(4, 5)
$\underline{\hat{I}}_{\text{beam}}(\omega)$	Phasor of $I_{\text{beam}}(t)$	(5)

J

j	Imaginary number $j = \sqrt{-1}$	(2)
j_{mn}	Zeros of Bessel function $\mathrm{J}_m(x)$ of first kind	(4)
j'_{mn}	Zeros of $\mathrm{J}'_m(x)$	(4)
J	Action variable	(2, 3, 5)
$\vec{J}$	Current density vector	(2, 4, 5, 6)
J_{S}	Saturation current density	(6)
J_x, J_y, J_z	Components of current density $\vec{J}$ in Cartesian coordinates	(2, 5)

K

k	Index, integer	(2, 5)
k	Constant, parameter	(5, 6)
$k \in [0, 1]$	Modulus, argument of elliptic integrals	(2, 3)
k'	Complementary modulus $= \sqrt{1 - k^2}$	(2)
k_{B}	Boltzmann constant	(6)
k_z	Propagation constant	(4, 5)
K	Constant	(2, 4, 6)
K	Spring constant	(2)
K	Hamiltonian	(5)
$\mathrm{K}(k)$	Complete elliptic integral of the first kind	(2, 3)
K	Gain	(7)
K_{c}	Controller gain	(7)
K_{cd}	Gain of capacitive divider	(7)
K_{ff}	Feedforward gain	(7)

K_{mod}	Modulator gain	(7)
K_{P}, K_{I}, K_{D}	Gains of PID controller	(7)
K_{Vgain}	Gains of driver and tetrode amplifiers	(7)
$K_{0,\nu}$, $K_{\mathrm{r},\nu}$, $K_{\mathrm{c1},\nu}$, $K_{\mathrm{c2},\nu}$	Constants	(7)
K_{ω}	Constant	(7)

L

l	Index, integer	(2)
Δl_{gap}	Gap length	(4)
l_{pillbox}	Length of pillbox cavity	(4)
l_{R}	Orbit length of reference particle, synchrotron circumference	(3, 4, 5)
$\bar{l}_{\mathrm{R}}$	Lorentz-transformed synchrotron circumference	(5)
l_n	Orbit length $l_{\mathrm{R}} + \Delta l_n$ for nonreference particle	(3)
L	Positive real number	(2)
L	Inductance	(2)
$L(\vec{r})$	Lyapunov function	(2)
L_{p}	Inductance of parallel equivalent circuit	(4)
L_{s}	Inductance of series equivalent circuit	(4)

M

m	Index, integer	(2, 5, 7)
m	Mass	(2, 3, 5)
m	Order of Bessel functions J_m, Y_m, I_m, and K_m	(4, 5)
m	Mode number (within-bunch oscillations)	(4, 5)
m	Polynomial order of numerator	(7)
m_0	Rest mass	(2, 5)
m_1, m_2	Polynomial orders, positive integers	(7)
m_{e}	Electron mass	(6)

m_u	Unified atomic mass unit	(2)
M	Matrix	(2)
M	Constant	(5)

N

n	Index, integer	(2, 5)
n	Number of coordinates q_i, degrees of freedom, dimension of space	(2)
n	Mode number (coupled-bunch oscillations)	(4, 5)
n	Polynomial order of denominator	(7)
n_critical	Number of open-loop poles on imaginary axis	(7)
n_h	Number of cavities for harmonic number h	(3)
n_unstable	Number of unstable open-loop poles	(7)
n_1, n_2	Polynomial orders, positive integers	(7)
Δn	Number of particles in volume element ΔV	(2)
$\mathrm{d}n$	Density of occupied states	(6)
$\mathrm{d}n_\mathrm{stat}$	Density of states inside the metal	(6)
N	Integer, number	(2, 4, 5, 7)
N_b	Number of particles per bunch	(5)
N_beam	Total number of particles in the synchrotron	(5)
N_bias	Number of bias current loops/windings	(4)

P

p	Index, integer	(2, 3, 5)
p	Momentum	(2, 5, 6)
p	Generalized momentum	(2, 5)
p	Mode number	(4)
$p,\ p^*$	Poles and complex conjugates	(7)
$\vec{p}$	Momentum vector	(2, 3)

Δp	Momentum deviation $p - p_{\mathrm{R}}$, continuous counterpart of Δp_n	(2, 3, 5)
$\frac{\Delta p}{p_{\mathrm{R}}}$	Momentum spread	(5)
δp	Change of reference momentum during one turn $= p_{\mathrm{R},n+1} - p_{\mathrm{R},n}$	(3)
$p_{\mathrm{c},\nu}$	Nonzero complex poles	(7)
p_i	Generalized momentum variables	(2)
$p_i(0)$	Open-loop poles	(7)
p_n	Momentum $= p_{\mathrm{R},n} + \Delta p_n$ of nonreference particle (turn n)	(3)
Δp_n	Momentum deviation of nonreference particle (turn n)	(3)
$p_{\mathrm{r},\nu}$	Nonzero real poles	(7)
$p_{\mathrm{R},n}$	Momentum of reference particle during turn n	(3)
p_{R}	Reference momentum, continuous version of $p_{\mathrm{R},n}$	(3, 5)
p_{vac}	Minimum momentum of electrons to leave the metal	(6)
p_x, p_y	Cartesian components of $\vec{p}$	(6)
p_1, p_ν	Poles	(7)
P	Generalized momentum, transformed	(2, 5)
P_i	Generalized momenta, transformed	(2)
$\overline{P}_{\mathrm{loss}}$	(Time-averaged) power loss	(4)
PM	Phase margin	(7)

Q

q	Generalized coordinate	(2, 5)
q_i	Generalized coordinates	(2)
Q	Charge	(1, 2, 3, 5)
Q	Generalized coordinate, transformed	(2, 5)
Q	Quality factor, Q factor of ring core material	(4)

Q_i	Generalized coordinates, transformed	(2)
Q_p	Q factor of cavity	(5)
$Q_{\mathrm{p},0}$	Unloaded Q factor of cavity	(4)

R

r	Radius, coordinate	(4)
$\bar{r}$	Mean ring core radius	(4)
$\bar{r}$	Lorentz-transformed r	(5)
$\vec{r}, \vec{r}(t)$	Position vector $= (x, y)$	(1, 2, 7)
$\vec{r}(t_0) = \vec{r}_0$	Initial value	(2, 7)
$\vec{r}\,'$	Vector $= (x', y')$	(2)
$\mathrm{d}\vec{r}$	Infinitesimal position vector	(4)
$\vec{r}_\mathrm{F}$	Vector, fixed point	(2)
$\vec{r}_k$	Discrete samples of position vector	(2)
r_a	Inner radius (anode)	(6)
r_beam	Beam radius	(5)
r_bp	Radius of beam pipe	(5)
r_c	Outer radius (cathode)	(6)
$r_\mathrm{i}, r_\mathrm{o}$	Inner, outer core radius	(4)
r_pillbox	Radius of pillbox cavity	(4)
r_R	Bending radius of the reference particle	(3, 5)
r_n	Bending radius $= r_\mathrm{R} + \Delta r_n$ of nonreference particle (turn n)	(3)
$\vec{r}_1, \vec{r}_2$	Vectors	(2)
R	Pendulum length	(2, 5)
$R_{\mathrm{p},0}$	Shunt impedance, resistance of parallel equivalent circuit of cavity	(4, 5)
R_p	Total resistance of generator and cavity	(4)
R_s	Resistance of series equivalent circuit	(4)
R_S	Surface resistivity	(4)

S

s	Complex Laplace variable $\sigma + j\omega$	(2, 7)
s	Coordinate, longitudinal path length	(3)
s^2	Sample variance, estimator of variance	(2)
$s_{h,k}$	Positions of cavities	(3)
S	Synchrotron rest frame (laboratory frame)	(5)
$\bar{S}$	Reference frame/rest frame of beam	(5)

T

t	Time	(1, 2, 4, 5, 6)
Δt	Time difference, time deviation	(2, 3, 5)
$\Delta\hat{t}$	Amplitude of Δt	(3)
t_k	Discrete times	(2)
$t_{h,k}$	Arrival time of bunch at cavity k	(3)
t_n	Gap arrival time $t_{\mathrm{R},n} + \Delta t_n$ of asynchronous particle (turn n)	(3)
t_R	Reference time	(3)
$t_{\mathrm{R},n}$	Time of reference particle reaching the gap during turn n, $t_{\mathrm{R},0} = 0$	(3)
t_0	Time shift, initial time	(7)
t_1, t_2, t_3, t_4	Time values	(5)
Δt_0	Initial time deviation	(5)
Δt_n	Time deviation in turn n	(5)
$\Delta t_{\mathrm{max,stat}}$	Max. Δt of separatrix, stationary case	(3)
T	Oscillation period $= \frac{2\pi}{\omega}$	(2)
T	Time shift	(2)
T	Absolute temperature	(6)
ΔT	Continuous counterpart of $T_n - T_{\mathrm{R},n}$	(3)

Δt_{gap}	Time of flight through gap	(4)
T_{cav}	Cavity time constant	(7)
$T_{\text{c1}}, T_{\text{c2}}$	Time constants	(7)
T_{d}	Time delay	(7)
T_{D}	Time constant of derivative part	(7)
T_{det}	Detector time constant	(7)
T_k	Time shifts of Dirac pulses	(2)
T_n	Revolution period of nonreference particle in turn n	(3)
T_{R}	Revolution time of reference particle	(1, 3, 5)
T_{RF}	RF period	(1)
$T_{\text{R},n}$	Revolution time of reference particle in turn n	(3)
T_{ry}	Closed-loop time constant	(7)
T_0	Period $= \frac{2\pi}{\Omega_0}$	(5)
T_1	Period of single-sine pulse $= \frac{2\pi}{\omega_1}$	(5)
$T_1, \ldots, T_4$	Temperatures	(6)

U

u	Velocity (absolute value)	(2)
$u(t)$	Input signal, function	(5, 7)
u_n	Velocity (nonreference particle, turn n)	(3)
u_{R}	Reference velocity, continuous counterpart of $u_{\text{R},n}$	(3)
$u_{\text{R},n}$	Reference particle velocity (turn n)	(3)
Δu	Velocity deviation	(2)
$\vec{u} = \frac{\text{d}\vec{r}}{\text{d}t}$	Instantaneous velocity (in S)	(1, 2, 3)
$\bar{\vec{u}}$	Particle velocity measured in $\bar{S}$	(2)
$\vec{u}$	Input vector	(7)
$\vec{u}_{\text{c}}$	Output of dynamic output feedback	(7)
$\vec{u}_{\text{F}}$	Equilibrium vector	(7)
$\Delta\vec{u}$	Deviation from $\vec{u}$	(7)
U	Neighborhood set	(2)
$U_{(\vec{r}_0, t_0)}$	Neighborhood of $(\vec{r}_0, t_0)$	(2)

V

v	Velocity	(2, 3, 4, 5)
$\vec{v}$	Relative velocity of $\bar{S}$ w.r.t. S	(2)
$\vec{v}$	Velocity field, vector field	(2, 6, 7)
$\vec{v}_{\mathrm{ho}}$	Vector function with higher-order terms	(7)
v_n	Normal component of velocity w.r.t. surface of V	(2)
v_x, v_y, v_z	Components of velocity $\vec{v}$	(2, 6)
v_x, v_y	Moments of particle ensemble	(5)
v_0	Constant, steady-state value of v_x	(5)
$v_1, \ldots, v_n$	Components of $\vec{v}$	(7)
$\vec{v}_1, \vec{v}_2$	Vector fields, vector functions	(2, 7)
V	Voltage	(1, 2, 3, 4)
$V, \partial V$	Three-dimensional domain and its boundary	(2)
ΔV	Volume element	(2)
V	Neighborhood set	(2)
$V(t)$	Total accelerating voltage	(3)
$\hat{V}$	Voltage amplitude	(1, 3, 5)
V_{a}	Anode voltage	(6)
$V_{\mathrm{a,DC}}$	DC component of anode voltage	(6)
V_{h}	Heating voltage	(6)
$\hat{V}_{h,k}$	Gap voltage amplitude of cavity k	(3)
$\hat{V}_{\mathrm{c,off}}$	Offset voltage	(7)
$\hat{V}_{\mathrm{c}}$	Control effort (low-level)	(7)
$\hat{V}_{\mathrm{dr}}$	Driver input amplitude	(7)
$\hat{V}_{\mathrm{e}}$	Amplitude error (low-level)	(7)
$\hat{V}_{\mathrm{gap,det}}$	Detected amplitude (low-level)	(7)
$\hat{V}_{\mathrm{inj}}$	Voltage amplitude at injection	(5)
$V_{h,k}(t)$	Gap voltage of cavity k	(3)
V_{beam}	Voltage seen by the beam	(4)
$V_{\mathrm{gap}}(t)$	Gap voltage	(4, 5)
$\hat{V}_{\mathrm{gap}}$	Gap voltage amplitude	(5, 7)
$\Delta\hat{V}_{\mathrm{gap}}$	Amplitude error	(7)
$\hat{V}_{\mathrm{gap,ref}}$	Reference amplitude	(7)
V_{gen}	Generator voltage	(4)
V_{g1}	Tetrode control grid voltage	(4, 6)
V_{g2}	Tetrode screen grid voltage	(6)
V_{max}	Maximum voltage	(2)

V_n	Voltage $= V_{\mathrm{R},n} + \Delta V_n$ in turn n	(3)
ΔV_n	Gap voltage difference in turn n	(5)
V_R	Reference voltage	(1)
$\hat{V}_\mathrm{ref}$	Reference amplitude (low-level)	(7)
$\Delta\hat{V}_\mathrm{ref}$, $\Delta\hat{V}_\mathrm{c}$, $\Delta\hat{V}_\mathrm{e}$, $\Delta\hat{V}_\mathrm{gap,det}$	Deviations w.r.t. setpoint	(7)
$V_{\mathrm{R},n}$	Reference voltage during turn n	(3)
$V'(t)$, $V''(t)$, $V'''(t)$	Derivatives of $V(t)$ w.r.t. time	(5)
$\underline{\hat{V}}$	Phasor of $V(t)$	(5)

W

$\vec{w}$	Vector	(2)
$w_{ki}, \ldots$	Components of $\vec{w}_k$	(2)
$\vec{w}_k$	Eigenvector of A to λ_k	(2)
W	Energy	(2, 6)
ΔW	Energy gain, energy deviation, energy difference	(1, 3, 5, 6)
$\Delta\hat{W}$	Amplitude of ΔW	(3)
W_bind	Binding energy	(6)
W_el	(Stored) electric energy	(2, 4)
W_magn	(Stored) magnetic energy	(2, 4)
$\overline{W}_\mathrm{el}$	(Stored) electric energy, time-average	(2, 4)
$\overline{W}_\mathrm{magn}$	(Stored) magnetic energy, time-average	(2, 4)
W_F	Fermi energy	(6)
W_kin	Kinetic energy	(2, 3)
$W_\mathrm{kin,u}$	Kinetic energy per unified atomic mass unit	(3)
ΔW_max	Maximum value of separatrix in ΔW	(3)
$\Delta W_\mathrm{max,stat}$	Separatrix maximum, stationary case	(3, 5)
W_pot	Potential energy	(2, 5)
W_rest	Rest energy	(2)
$W_\mathrm{rest,u}$	Rest energy per atomic mass unit	(3)
W_R	Total energy of reference particle	(3, 5)
W_tot	Total energy	(2, 4)
ΔW_tot	Energy deviation	(2)
W_vac	Potential energy of electron in vacuum just outside the metal	(6)

X

x	Space coordinate in S	(2)
$x(t)$	Real function	(2)
Δx	Deviation $x - \mu$	(2)
$\bar{x}$	Space coordinate in $\bar{S}$, Lorentz-transformed x	(2)
$\bar{x}$	Mean value, sample mean, estimator of μ	(2, 5)
$\vec{x}(t)$	State vector	(5, 7)
$\vec{x}(0)$	Initial value for $t = 0$	(7)
$\Delta\vec{x}$	Deviation from $\vec{x}$	(7)
$\vec{x}_\mathrm{F}$	Equilibrium vector	(7)
x_k	Samples	(2)
x_k	Particle position in phase space	(5)
x_k	Variables of tracking equations	(3)
$\vec{x}_\mathrm{c}$	State vector of dynamic output feedback	(7)
$\vec{x}_\mathrm{e}$	Vector with control errors	(7)
$\vec{x}_\mathrm{op}$	Operating point	(5)
Δx_rms	Root mean square	(2)
$x_\mathrm{step}(t)$	Piecewise constant function	(7)
$x_1, \ldots, x_n$	State variables	(2, 7)
$x_1(t), x_2(t)$	Functions	(7)
$x_1, \ldots, x_5$	States	(5)
$\Delta x_1, \Delta x_3$	Deviations of x_1, x_3 from the operating point	(5)
X	Random variable	(2)
$X(s)$	Laplace transform of $x(t)$	(2, 7)
$X_A(0)$	Matrix (w_{ki}) of eigenvectors	(2)
$X_A(t)$	Matrix	(2)
$X_B(t)$	Matrix	(2)
$X_\mathrm{d1}, X_\mathrm{d2}, X_\mathrm{d3}$	Disturbance signals	(7)
$X_\mathrm{e}(s)$	Control error	(7)
X_k	Random variable	(2)
X_n	Spectral components, discrete Fourier transform of samples x_k	(2)
X_n^*	Complex conjugate of X_n	(2)
$\vec{X}(s)$	Laplace transform of state vector $\vec{x}(t)$	(7)

Y

y	Space coordinate in S	(2)
$y(t)$	Function, beam signal	(2)
$y(t)$	Function, ODE solution	(2, 5)
$y(t)$	Output signal	(7)
$\bar{y}$	Space coordinate in $\bar{S}$, Lorentz-transformed y	(2)
$\bar{y}$	Mean value	(5)
$\vec{y}$	Output vector	(7)
$\vec{y}_\mathrm{r}$	Vector with reference values	(7)
$\vec{y}_\mathrm{m}$	Vector with measurement values	(7)
$y_\mathrm{delay}(t)$	Delayed output signal	(7)
y_k	Variables of tracking equations	(3)
y_k	Particle position in phase space	(5)
$y_\mathrm{step}(t)$	Output response to $x(t) = x_\mathrm{step}(t)$	(7)
$y_\Theta(t)$	Response to step function	(7)
$y_\mathrm{trans}(t)$	Transient response	(7)
$y_1(t), y_2(t)$	Functions	(7)
Y	Random variable	(2)
$Y(s)$	Laplace transform of $y(t)$	(2, 7)
$Y(\varphi_\mathrm{R})$	Bucket height reduction factor $= \frac{\Delta W_\mathrm{max}}{\Delta W_\mathrm{max,stat}}$	(3)
$Y_\mathrm{m}(s)$	Measured output signal	(7)
$Y_\mathrm{r}(s)$	Setpoint/reference	(7)
Y_s	Admittance $= 1/Z_\mathrm{s}$	(4)
Y_tot	Total cavity admittance	(4, 5)
$Y_\mathrm{trans}(s)$	Laplace transform of $y_\mathrm{trans}(t)$	(7)

Z

z	Space coordinate in S	(2)
$z(t)$	Function, ODE solution	(2)
z	Longitudinal coordinate	(4)
z	Cylindrical coordinate	(4, 5)
z, z^*	Zeros and complex conjugates	(7)
$\bar{z}$	Space coordinate in $\bar{S}$, Lorentz-transformed z	(2, 5)
$z_{\mathrm{r},\nu}$	Nonzero real zeros	(7)
$z_{\mathrm{c},\nu}$	Nonzero complex zeros	(7)

z_1, z_2, z_m, z_v	Zeros	(7)
z_q	Charge number	(2)
Z	Atomic number	(2)
Z_s	Impedance $= j\omega L_\mathrm{s} + R_\mathrm{s}$	(4)
Z_sc	Longitudinal space charge impedance	(5)
Z_tot	Total cavity impedance	(4, 5)
$Z_\mathrm{tot}(s)$	Cavity impedance as Laplace transform	(4)
Z_0	Impedance of free space	(4, 5)

Greek Symbols: Symbol, description, example for chapter (no.)

α	Constant	(2, 4)
α	Angle of mathematical pendulum	(2, 5)
$\alpha, \alpha_\mathrm{min}, \alpha_\mathrm{max}$	Parameter and its limits	(2)
$\alpha(\varphi_\mathrm{R})$	Bucket area reduction factor $= \frac{A_\mathrm{B}}{A_\mathrm{B,stat}}$	(3, 5)
$\hat{\alpha}$	Angle amplitude	(5)
α_adiab	Adiabaticity parameter	(5)
α_c	Momentum compaction factor	(3)
$\beta, \beta_\mathrm{min}, \beta_\mathrm{max}$	Parameter and its limits	(2)
β	Lorentz factor	(3, 5)
$\Delta\beta$	Continuous counterpart of $\Delta\beta_n$	(3)
$\Delta\beta_n$	Beta deviation in turn n	(3)
β_n	β of non-synchronous particle $= \beta_{\mathrm{R},n} + \Delta\beta_n$ during turn n	(3, 5)
β_u	Lorentz factor (for u)	(2)
β_v	Lorentz factor (for v)	(2)
β_R	Lorentz beta for reference particle	(3, 5)
$\beta_{\mathrm{R},n}$	Lorentz beta for reference particle (turn n)	(5)
γ	Argument of elliptic integral	(2, 3)
γ	Lorentz factor	(3, 5)
$\Delta\gamma$	Continuous counterpart of $\Delta\gamma_n$	(3, 5)
γ_n	γ of non-synchronous particle (turn n)	(3)
$\Delta\gamma_n$	Gamma deviation in turn n	(3, 5)
γ_u	Lorentz factor (for u)	(2)
γ_v	Lorentz factor (for v)	(2)

γ_R	Lorentz gamma for reference particle	(3, 5)
$\gamma_{\mathrm{R},n}$	Lorentz gamma for reference particle (turn n)	(3, 5)
γ_T	Transition gamma	(3)
$\delta(x)$, $\delta(t)$	Dirac's delta distribution	(1, 2, 7)
δ	Skin depth	(4)
δ	Relative momentum deviation	(5)
δ	Damping factor	(7)
$\delta_{\mathrm{max,stat}}$	$1/2$ of the bucket height, stationary case	(5)
δ_max	$1/2$ of the bucket height, general case	(5)
δ_n	Relative momentum deviation (turn n)	(5)
δ_μ	Angle of complex $\underline{\mu}$	(4)
ϵ	Small quantity	(2)
ϵ	Permittivity	(2)
ϵ_r	Relative permittivity	(5)
$\epsilon_{\mathrm{r,eff}}$	Effective relative permittivity	(4)
ϵ_0	Vacuum permittivity	(2, 6)
$\zeta(\rho)$	Function	(5)
η_fill	Bucket filling factor	(3)
η_R	Phase slip factor $= \alpha_\mathrm{c} - \frac{1}{\gamma_\mathrm{R}^2}$	(3, 5)
$\eta_{\mathrm{R},n}$	Phase slip factor (turn n)	(3, 5)
θ	Angle variable	(2)
θ	Angle along the synchrotron ring	(5)
$\Delta\theta$	Angle deviation	(5)
$\theta_{h,k}$	Angle of cavity position along the ring	(3)
$\Theta(t)$	Heaviside step function	(2, 5, 7)
κ	Conductivity	(2, 4)
κ	Constant	(5)
λ, λ_k	Eigenvalue	(2)
$\lambda_q(t)$, $\lambda_q(z)$	Line charge density	(3, 5)
$\lambda(t)$	Function, control parameter	(5)
λ	Wavelength	(4)
$\bar{\lambda}_q$	Lorentz-transformed line charge density	(5)
λ_H	Eigenvalues of Hessian matrix	(2)
λ_norm	Normalized density function	(5)

$\bar{\lambda}_{q,0,\mathrm{DC}}$	Lorentz-transformed line charge density (DC beam)	(5)
$\lambda_{q,0,\mathrm{DC}}$	Line charge density of DC beam	(5)
μ	Mean of Gaussian distribution	(2)
μ	Permeability	(2, 4)
$\underline{\mu}$	Complex permeability $= \mu'_\mathrm{s} - j\mu'_\mathrm{s}$	(4)
μ'_p, μ''_p	Permeability parameters of parallel equivalent circuit	(4)
$\mu'_\mathrm{p,r}$, $\mu''_\mathrm{p,r}$, $\mu'_\mathrm{s,r}$, $\mu''_\mathrm{s,r}$	Relative permeabilities	(4)
μ_r	Relative permeability	(4)
μ'_s, μ''_s	Permeability parameters of series equivalent circuit	(4)
μ_Δ	Differential or incremental permeability	(4)
μ_0	Vacuum permeability	(4)
μ_0, μ	Positive constants	(5)
ν	Index	(7)
ξ, ξ'	Abbreviations for $\frac{\partial(x,y)}{\partial(\alpha,\beta)}$ and $\frac{\partial(x',y')}{\partial(\alpha,\beta)}$	(2)
ξ	Determinant $\frac{\partial(Q,P)}{\partial(q,p)}$	(2)
ξ	Factor, Jacobian	(5)
ξ	Moment of particle ensemble	(5)
ρ_q	(Beam) Charge density	(2, 5, 6)
ρ	Cylindrical coordinate	(4, 5)
ρ	Particle density	(2)
$\bar{\rho}_q$	Lorentz-transformed charge density	(5)
$\rho_{q,0}(z)$	Charge density function	(5)
$\rho_{q,0,\mathrm{DC}}$	Constant charge density (DC beam)	(5)
σ	Standard deviation of Gaussian distribution	(2)
σ	Damping parameter of Z_tot	(4)
σ_q	Surface charge density	(5)
τ	Pulse width	(2)
τ	Cavity time constant	(4)
$\Delta\tau$	Sample time	(7)
τ_k	Time shift of a bunch	(5)
τ_ν	Sampling times	(7)
φ	Phase	(2, 5, 7)

φ	Cylindrical coordinate	(4, 5)
$\Delta\varphi$	Phase deviation	(5)
$\varphi_{h,k}$	RF phase shift at cavity k	(3)
$\Delta\varphi_{\mathrm{gap}}(t)$	Phase modulation	(5)
$\Delta\varphi_{\mathrm{l}}$	Left border of bucket	(3)
$\Delta\varphi_{\mathrm{saddle}}$	Right border of bucket, saddle point	(3)
φ_n	Phase of Fourier coefficients	(2)
φ_n	RF phase w.r.t. zero crossing of gap voltage	(5)
$\Delta\varphi_{\mathrm{Nyquist}}$	Argument change of Nyquist plot	(7)
φ_{R}	Synchronous (RF) phase	(3, 5)
φ_{RF}	RF phase	(1, 3)
$\Delta\varphi_{\mathrm{RF}}$	Deviation $\varphi_{\mathrm{RF}} - \varphi_{\mathrm{R}}$ of RF phase w.r.t. reference	(3, 5)
$\Delta\hat{\varphi}_{\mathrm{RF}}$	RF phase amplitude $= \omega_{\mathrm{RF}}\Delta\hat{t}$	(3, 5)
$\varphi_{\mathrm{RF,B,len}}$	Bucket length in $\Delta\varphi_{\mathrm{RF}}$	(3)
$\Phi(\Delta u)$	Area below curve $f(x)$, probability	(2)
Φ	Scalar potential	(2, 4, 6)
Φ_{a}	Anode potential	(6)
Φ_{m}	Magnetic flux	(2, 4)
Φ_t	Flow, continuous map	(2)
$\Phi_{\mathrm{m,tot}}$	Total flux of N ring cores	(4)
$\Phi_{\mathrm{m,1}}$	Flux through one single core	(4)
$\omega = \frac{2\pi}{T}$	Angular frequency, frequency of the first harmonic	(2, 3, 5, 7)
ω_{c}	Angular cutoff frequency	(4, 7)
ω_{R}	Angular revolution frequency of reference particle	(1, 3)
ω_{res}	Resonant (angular) frequency	(2, 7)
ω_{RF}	RF frequency $= 2\pi f_{\mathrm{RF}}$	(1, 3, 5)
ω_{ry}	Closed-loop bandwidth	(7)
$\omega_{\mathrm{S,0}}$	Synchrotron frequency for small oscillation amplitudes	(3)
$\omega_{\mathrm{S,0,stat}}$	Synchrotron frequency for small oscillations, stationary case	(3, 5)
$\omega_{\mathrm{S,stat}}$	Synchrotron frequency for arbitrary oscillation amplitudes, stationary case	(3)

ω_S	Synchrotron frequency $= \frac{2\pi}{T_S} = 2\pi f_S$	(5)
$\omega_0 = \frac{2\pi}{T_0}$	Fundamental angular frequency and period	(2)
Ω	Special angular frequency	(2)
$\Omega(t), \Omega_0(t)$	Special time-dependent angular frequency	(5)

Operators, General Notations: Symbol, description (A, B are placeholders for arbitrary quantities)

$A(t) * B(t)$	Convolution
$[A, B]$	Closed interval
$]A, B[$	Open interval
$\hat{A}$	Amplitude of periodic function $A(t)$
$\bar{A}$	Lorentz transform of A or mean value of A
$\vec{A}$	Vector quantity
$\underline{\hat{A}}$	Phasor
ΔA	Deviation/difference or Laplace operator
A^*	Complex conjugate of A
A^{T}	Transpose of matrix A
$\dot{A}, \ddot{A}, \dddot{A}$	Time derivatives of A
$\measuredangle A$	Phase/angle of A

Chapter 1
Introduction

The motion of a charged point particle is influenced by electromagnetic fields according to the **Lorentz force**[1]

$$\boxed{\vec{F} = Q\left(\vec{E} + \vec{u} \times \vec{B}\right).} \tag{1.1}$$

Here Q denotes the charge of the particle, and $\vec{E}(\vec{r}, t)$ and $\vec{B}(\vec{r}, t)$ are the electric field and the magnetic induction vectors, respectively, at a certain time t and at the position $\vec{r}(t)$ of the charge. In this book, we will alternatively call $\vec{B}$ the magnetic field, since the distinction between $\vec{B}$ and the magnetic field vector $\vec{H}$ is obvious. The vector $\vec{u} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}$ is the instantaneous velocity of the point charge.

For the energy gain due to the electromagnetic field, we obtain

$$\Delta W = \int_C \vec{F} \cdot \mathrm{d}\vec{r} = Q \int_C \vec{E} \cdot \mathrm{d}\vec{r} + Q \int_C \left(\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} \times \vec{B}\right) \cdot \mathrm{d}\vec{r},$$

where C is the flight path of the particle. For the second integral, we obtain

$$\int_C \left(\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} \times \vec{B}\right) \cdot \mathrm{d}\vec{r} = \int_{t_1}^{t_2} \left(\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} \times \vec{B}\right) \cdot \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\, \mathrm{d}t = 0,$$

since two factors of the scalar triple product are equal. Therefore, it is impossible to use magnetic fields to change the energy of the charged particle.[2] Magnetic fields may be used to deflect particles, but an acceleration in the sense of changing their energy by

[1] The contribution of the electric field is the Coulomb force, whereas the term "Lorentz force" is used to specify the magnetic contribution in a more specific sense.

[2] Time-dependent magnetic fields, however, may be used to induce a voltage that allows acceleration. Also in this case, the accelerating field is an electric field.

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6_1

$$\boxed{\Delta W = Q \int_C \vec{E} \cdot \mathrm{d}\vec{r} = QV} \tag{1.2}$$

requires electric fields that lead to a voltage

$$V = \int_C \vec{E}(\vec{r}, t) \cdot \mathrm{d}\vec{r} = \int_{t_1}^{t_2} \vec{E}(\vec{r}(t), t) \cdot \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\, \mathrm{d}t.$$

Please note that we will always define the voltage V to be oriented in the same way as the direction of flight. For acceleration, we need $V > 0$ in case of $Q > 0$ (e.g., protons or other positive ions) and $V < 0$ in case of $Q < 0$ (e.g., electrons).

The simplest choice of an electric field is a time-independent field, i.e., a DC field. The total energy that may be reached by DC fields, however, is limited by high-voltage sparkovers.

The next logical step to increase the overall voltage would be to use not only one accelerating section but several adjacent ones. This does not solve the problem, however, since either the DC voltages will add up to a voltage that again leads to sparkovers, or—depending on the grounding concept of the sections—DC voltages in the reverse direction will be present, leading to sections with undesired deceleration.

These limits may be exceeded if AC fields are used, especially in the radio frequency (RF) range, because the particles may then gain energy several times. In a linear accelerator, the beam passes different cavities, which may consist of different accelerating cells. In a ring accelerator such as a **synchrotron** or a **storage ring**, the particles repeatedly gain energy in the same cavity or in the same number of cavities, since they arrive at the same place after one revolution.

When using RF fields, one still has to make sure that the acceleration that is realized during one-half of the RF period does not lead to a deceleration during the other half of the period. In linear accelerators (**LINAC**s), this may be accomplished by so-called **drift tubes**, which shield the particles against electric fields with the wrong polarity (cf. [1]). In synchrotrons and storage rings, the particles will be located inside the conducting beam pipe during those time intervals in which the field has the wrong polarity. The electric field may be generated in a so-called **ceramic gap**. Such a ceramic gap is a short ceramic tube that interrupts the metallic beam pipe. Since its material is nonconducting, a voltage can be induced even though the beam pipe is still evacuated.

Figure 1.1 shows the main elements (cf. [2] for further reading) of a synchrotron or storage ring in a schematic way:

- A metallic **beam pipe** is evacuated so that flying particles will not hit gas molecules. For storage rings with long storage times (e.g., on the order of several hours), a better vacuum quality is usually required than is to be found in synchrotrons that are used only for comparatively short acceleration phases.
- An **injection system** is used to deflect the beam (which comes from a linear accelerator or a **booster synchrotron**) onto its target trajectory. Following the

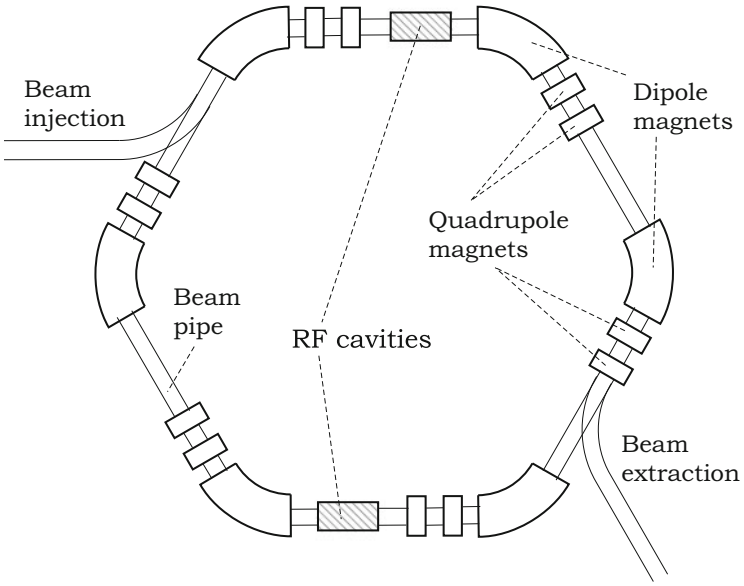

Fig. 1.1 Schematic drawing of a synchrotron

acceleration, which may need thousands of revolutions, the beam is extracted by a similar system. Injection and **extraction** are shown only schematically in Fig. 1.1.

- **Dipole magnets** are used to deflect particles in such a way that a closed orbit is realized. Inside a dipole magnet, a constant or slowly varying B field is oriented in the vertical direction so that the beam is bent horizontally. Dipole magnets are therefore the arcs of a synchrotron. They are also called bending magnets.
- **Quadrupole magnets** are used as focusing elements. A quadrupole magnet leads to transverse focusing in one direction (e.g., in the radial x direction) and to defocusing in the other direction (e.g., in the vertical y direction). Fortunately, the net effect of two quadrupoles the first of which produces focusing in the x direction (and defocusing in the y direction) while the second produces focusing in the y direction (and defocusing in the x direction) is to focus in both directions. Therefore, two (a so-called **quadrupole doublet**) or three quadrupole magnets (a **quadrupole triplet**) are typically combined. One also speaks of magnetic quadrupole lenses, since the effect in ion or electron optics is comparable with the effect in light optics.
- In the straight sections of the synchrotron, a set of **RF cavities** is used to produce the electric field, mentioned above, that is required for the desired energy gain ΔW.

In a synchrotron, there is a specific orbit that is the desired one. Of course, not all particles will follow this orbit precisely, because the ideal situation that all particles

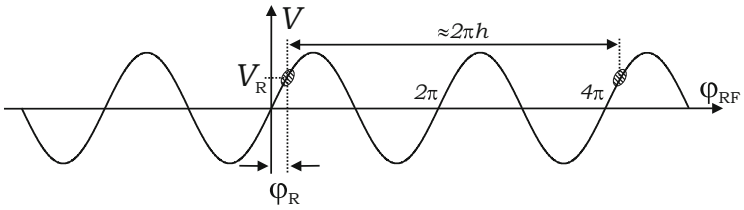

Fig. 1.2 Phase-focusing principle

have no transverse offset from this **reference orbit** cannot be realized. However, we may assume that a so-called **reference particle** will follow the reference orbit.

When the reference particle gains energy in an RF cavity, it is clear that the dipole field B has to be increased to keep the reference particle on the reference orbit (path length l_R for one revolution). One immediately sees that these conditions can be fulfilled only if all parameters—the magnetic dipole field B, the RF voltage amplitude, and the RF frequency—fit together. These parameters have to be varied synchronously, whence the name "synchrotron."

The reference particle (with reference energy W_R) is also called a **synchronous particle**. Typically, the energy gain ΔW in each revolution is small in comparison with the total energy W_R. The desired total energy gain is reached only because of the very large number of revolutions.

We now discuss the effect of the RF cavity on positive charges $Q > 0$. A sinusoidal RF voltage $V(t)$ is sketched in Fig. 1.2. This voltage is specified by

$$V = \hat{V} \sin(\varphi_\mathrm{RF}),$$

where the RF phase is given by

$$\varphi_\mathrm{RF} = \int_0^t \omega_\mathrm{RF}(\tilde{t})\, \mathrm{d}\tilde{t}. \tag{1.3}$$

In general, the amplitude $\hat{V}$ and the RF frequency[3] $f_\mathrm{RF} = \frac{\omega_\mathrm{RF}}{2\pi}$ are time-dependent quantities.

Let us assume that a certain level V_R of the electric voltage leads to the "correct" energy gain, i.e., after passing the RF cavity, the particle has an energy that allows it to travel on the desired path in the beam pipe and to "see" the correct voltage V_R the next time it arrives again at the RF cavity. In other words, in each revolution,

[3]Throughout this book, we always use the notation

$$\omega = 2\pi f, \qquad f = 1/T.$$

Here f denotes the frequency, T the period, and ω the angular frequency. This notation is used for every index that may be present.

the reference particle will experience an energy gain that allows it to stay on the reference path. The energy gain ΔW of the particle is small in comparison with its energy W_R. Also, the relative change in its velocity is small. Therefore, one may regard the RF frequency as constant during several periods, so that

$$\varphi_\mathrm{RF} \approx \omega_\mathrm{RF} t$$

may be written instead of Eq. (1.3).

The voltage V_R that is required after one revolution usually does not differ much from the voltage V_R in the previous revolution. This is why Fig. 1.2 shows almost the same voltage after the **revolution time** T_R. The voltage V_R is determined by

$$V_\mathrm{R} = \hat{V} \sin \varphi_\mathrm{R},$$

where φ_R is the **reference phase** or the **synchronous phase**. Please note that in LINACs, the synchronous phase is usually defined in a different way, namely with respect to the crest instead of the zero crossing of the RF voltage.

As the figure shows, it is not necessary that the **RF frequency** $f_\mathrm{RF} = 1/T_\mathrm{RF}$ equal the **revolution frequency** $f_\mathrm{R} = 1/T_\mathrm{R}$ for the same voltage V_R as in the previous revolution affecting the particle. If the number of RF periods that have passed after the revolution time T_R has elapsed is a positive integer, the particle will still be influenced by the same voltage V_R, and the slope of $V(t)$ also will look identical. Therefore, it is sufficient if

$$\boxed{f_\mathrm{RF} = h \cdot f_\mathrm{R}} \tag{1.4}$$

holds, where the **harmonic number** h is a positive integer (in Fig. 1.2, we have $h = 2$). Finally, a particle will reappear at the cavity after h RF periods.

Now we consider an asynchronous particle that arrives at the RF cavity a bit later than the reference particle. It is obvious that this particle will experience a higher voltage than the reference particle, leading to a higher energy gain. Therefore, one would expect it to arrive earlier at the cavity the next time (later in this book, we will point out that this is true only below the so-called **transition energy**). It will therefore move toward the reference particle.[4] Analogously, a particle that arrives earlier than the reference particle "sees" a lower voltage and therefore gains less energy than the reference particle. Hence, it will arrive later the next time. Both cases show that there is some stable region around the positive slope of the RF voltage where particles may be "focused." This principle is therefore called **phase focusing** or **phase stability** (cf. [3–7]).

[4] Of course, the particle does not move toward the reference particle on the same slope of the voltage $V(t)$. After each revolution, it will be located on a different slope. Similar to the triggering of an oscilloscope, however, we may project all these slopes onto each other so that a virtual movement of the particles becomes visible.

Analogous reasoning shows that the area around the negative slope of the RF voltage is an unstable region.

For negative charges, we would arrive at the conclusion that particles will be focused around the negative zero slope, whereas the positive slope is an unstable region in this case. As mentioned above, this is again true only below the so-called transition energy.

Later in this book, it will be shown that the asynchronous particles will not approach the synchronous one asymptotically. Instead, they will oscillate around it. This is the so-called **synchrotron oscillation**, which will be analyzed in detail in the main chapters of this book.

The phase focusing principle shows that a charged particle beam has to be "bunched" (i.e., it has to consist of **bunches**) if it is to be accelerated. In contrast to the acceleration case, a DC beam, which means a homogeneous distribution of particles in the longitudinal direction (also called a **coasting beam**) may exist at a constant reference energy W_R. Therefore, bunched beams are always possible, whereas a coasting beam may exist for a longer time only if the reference energy is constant.[5]

Various **beam diagnostic** instruments exist that allow one to evaluate the quality of the beam and to track problems [8–12]. Here we mention only some nondestructive methods. A **beam position monitor** (BPM) can be used to determine the transverse position of the beam (cf. [8, Sect. 5.4]—various BPM applications are discussed in [9]). This is done by evaluating the difference between the measurement signals of two opposite plates or buttons (in the horizontal and/or vertical direction, BPM Δ signal). If instead of the difference between the two signals, the sum of the signals is used (BPM Σ signal), one obtains a signal that is not primarily dependent on the transverse position of the beam but which represents the beam signal. Of course, sufficient bandwidth and sufficient dynamic range are required if the signal form is actually to represent the longitudinal bunch shape.

As an example, Fig. 1.3 shows an oscilloscope measurement of the BPM Σ signal in the synchrotron SIS18 at GSI dated 21 August 2008. The beam consisted of $^{40}\mathrm{Ar}^{18+}$ ions[6] at an energy of 11.4 MeV/u. An RF voltage of $\hat{V} = 6\,\mathrm{kV}$ at $h = 4$ was applied. This means that each pulse shown in the diagram corresponds to one of four bunches. If the oscillogram had been recorded for a longer time, the next visible pulse would correspond to the same bunch as the first, since this bunch reappears at the BPM after one revolution. In general, a maximum[7] of h bunches may circulate in a synchrotron if it is operated at the harmonic number h.

[5] Here we neglect synchrotron radiation, which may be significant in electron synchrotrons but which is usually negligible in ion synchrotrons.

[6] This notation will be explained in Sect. 2.7.

[7] As we will see in the main parts of this book, the h stable regions where bunches may exist are called **buckets**. Not all buckets have to be occupied by bunches; one speaks of **empty buckets** in this case.

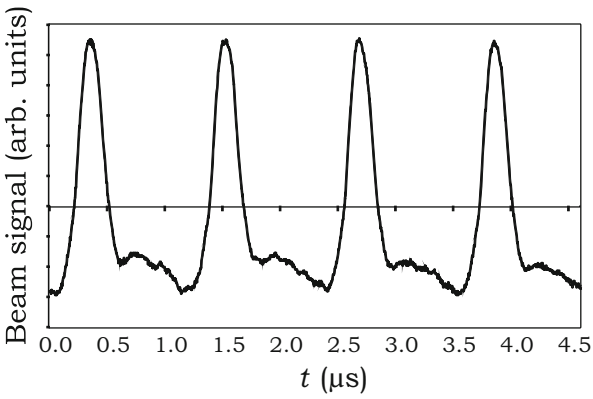

Fig. 1.3 Beam current measurement in a synchrotron

Usually, the BPM Σ signal is not calibrated with respect to the total beam current. As a nondestructive way to determine the beam current, one may, e.g., use a **beam current transformer** (BCT) [8, 11]. A DC beam current transformer (DCCT) does not provide the bunch shape but only the average current.[8]

Another important beam diagnostic procedure is the **Schottky measurement** [12–14]. Here we mention only the longitudinal Schottky measurement for the unbunched, i.e., coasting, beam. For this purpose, one analyzes the beam signal delivered by a suitable pickup, usually a broadband device, with a spectrum analyzer. For an ideal coasting beam with a continuous charge distribution corresponding to a constant beam current, one would not expect any spectral components other than the DC component. In reality, however, the coasting beam consists of a finite number of particles (discrete charges), and its noise therefore contains spectral components around the revolution frequency and its harmonics, which can be observed in the frequency domain. The spectrum is usually evaluated in a frequency range centered at several times the revolution frequency. As a result of this longitudinal Schottky measurement of the coasting beam, one may determine the revolution frequency f_R and also its distribution. This is one possible way of determining the required RF frequency $f_{RF} = hf_R$ with sufficient accuracy.

We do not want to finish our brief introduction without mentioning that the operation of a synchrotron or a storage ring requires several further technical systems. A **centralized control system** is needed that controls the different devices (magnets, RF cavities, beam diagnostics systems, etc.) in real time. **Cooling media** (at least water and air) and electrical power distribution and conversion systems are needed as well. The **vacuum system** mentioned above is another complex subsystem of an accelerator.

Before we proceed with our main topics, "particle acceleration" and "RF systems," we shall summarize several basic points in the next chapter:

[8] However, there exist **fast beam current transformers** that provide the bunch shape.

- Fourier analysis
- mathematical statistics
- electromagnetic fields
- special relativity
- nonlinear dynamics

These sections will include only the most important fundamental results that are needed in the rest of the book. It is, of course, impossible to aim at completeness, since each of these topics could fill several books and be the subject of its own university course.

References

1. T.P. Wangler, *RF Linear Accelerators* (Wiley-VCH Verlag GmbH&Co.KGaA, Weinheim, 2008)
2. A.W. Chao, M. Tigner, *Handbook of Accelerator Physics and Engineering*, 3rd edn. (World Scientific/New Jersey/London/Singapore/Beijing/Shanghai/Hong Kong/Taipei/Chennai, 2006)
3. D.A. Edwards, M.J. Syphers, *An Introduction to the Physics of High Energy Accelerators* (Wiley-VCH Verlag GmbH&Co.KGaA, Weinheim, 2004)
4. S.Y. Lee, *Accelerator Physics* (World Scientific, Singapore/New Jersey/London/Hong Kong, 1999)
5. K. Wille, Physik der Teilchenbeschleuniger und Synchrotronstrahlungsquellen (B. G. Teubner, Stuttgart, 2. Auflage, 1996)
6. P.J. Bryant, K. Johnsen, *The Principles of Circular Accelerators and Storage Rings* (Cambridge University Press, Cambridge/New York/Melbourne/Madrid/Cape Town, 2005)
7. H. Wiedemann, *Particle Accelerator Physics I & II*, 2nd edn. (Springer, Berlin/Heidelberg/New York, 2003)
8. P. Strehl, *Beam Instrumentation and Diagnostics* (Springer, Berlin/Heidelberg/New York, 2006)
9. M.G. Minty, F. Zimmermann, *Measurement and Control of Charged Particle Beams* (Springer, Berlin/Heidelberg/New York, 2003)
10. H. Koziol, Beam diagnostics for accelerators, in *CAS - CERN Accelerator School: Introduction to Accelerator Physics*, Loutraki, Greece, 2–13 Oct 2000 (2000), pp. 154–197
11. V. Smaluk, *Particle Beam Diagnostics for Accelerators. Instruments and Methods* (VDM Verlag Dr. Müller, Saarbrücken, 2009)
12. P. Forck, Lecture Notes on Beam Instrumentation and Diagnostics. Joint University Accelerator School, Archamps, France, Jan–March 2011 (2011). http://www-bd.gsi.de/conf/juas/juas_script.pdf
13. D. Boussard, Schottky noise and beam transfer function diagnostics, in *CAS - CERN Accelerator School: 5th Advanced Accelerator Physics Course*, Rhodes, 20 Sep–1 Oct 1993 (1995), pp. 749–782
14. F. Caspers, Schottky signals for longitudinal and transverse bunched-beam diagnostics, in *CAS - CERN Accelerator School: Course on Beam Diagnostics*, Dourdan, 28 May–6 Jun 2008 (2008), pp. 407–425

Chapter 2
Theoretical Fundamentals

In this chapter, we summarize some theoretical fundamentals. We assume that the reader is already familiar with these basic facts. The main purpose of this chapter is to introduce the notation that is used in this book and to provide a reference. Therefore, the explanations are brief, and no proofs are given.

2.1 Fourier Analysis and Application to Beam Signals

In this section, several formulas for Fourier series and the Fourier transform are summarized. However, we do not discuss the properties of a function that are necessary for the existence of the transformation. For those foundations, the reader should consult the references cited here.

2.1.1 Fourier Series

A real-valued periodic function $f(t)$ with period T may be decomposed into Fourier components according to the **Fourier series**

$$\boxed{f(t) = \sum_{n=-\infty}^{\infty} c_n \, e^{jn\omega t} \qquad \text{with } \omega = \frac{2\pi}{T},} \tag{2.1}$$

where the complex coefficients c_n are determined by

$$\boxed{c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\varphi) \, e^{-jn\varphi} \mathrm{d}\varphi}$$

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6_2

or by

$$\boxed{c_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t)\, e^{-jn\omega t}\, \mathrm{d}t,} \tag{2.2}$$

where we made the substitution $\varphi = \omega t$.

With the substitution $x = t + T$, we obtain

$$\int_{-T/2}^{0} f(t)\, e^{-jn\omega t}\, \mathrm{d}t = \int_{T/2}^{T} f(x - T)\, e^{-jn\omega x} e^{jn\omega T}\, \mathrm{d}x.$$

Due to $\omega = \frac{2\pi}{T}$, the last exponential function equals 1. Furthermore, we have $f(x - T) = f(x)$, so that

$$\int_{-T/2}^{0} f(t)\, e^{-jn\omega t}\, \mathrm{d}t = \int_{T/2}^{T} f(x)\, e^{-jn\omega x}\, \mathrm{d}x = \int_{T/2}^{T} f(t)\, e^{-jn\omega t}\, \mathrm{d}t$$

holds. Therefore, we may use

$$c_n = \frac{1}{T} \int_{0}^{T} f(t)\, e^{-jn\omega t}\, \mathrm{d}t \tag{2.3}$$

instead of Eq. (2.2).

2.1.2 *Spectrum of a Dirac Comb*

In this book, the Dirac delta distribution is used in a heuristic way without the foundations of distribution theory. Therefore, the reader should be aware that the results presented still have to be proven mathematically. For example, we use the formula

$$\int_{-\infty}^{+\infty} f(x)\, \delta(x - x_0)\, \mathrm{d}x = f(x_0),$$

even though it does not have any meaning in the scope of classical analysis.

A strongly bunched beam may be approximated by a sum of Dirac delta pulses

$$f(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT),$$

which is called **Dirac comb**. For this special sum of Dirac pulses, one obtains the following Fourier coefficients (only the Dirac pulse with $k = 0$ is located inside the interval $-T/2 \leq t \leq +T/2$):

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} \delta(t)\, e^{-jn\omega t} \mathrm{d}t = \frac{1}{T}. \tag{2.4}$$

Hence, all coefficients are equal. According to Eq. (2.1), we get

$$\sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{jn\omega t}.$$

This can also be written as

$$\sum_{k=-\infty}^{\infty} \omega\, \delta(\omega t - \omega k T) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{jn\omega t}$$

$$\Rightarrow 2\pi \sum_{k=-\infty}^{\infty} \delta(\varphi - 2\pi k) = \sum_{n=-\infty}^{\infty} e^{jn\varphi}.$$

2.1.3 *Different Representations of the Fourier Series*

The general definition of the Fourier series shows that the c_n are defined in such a way that both positive and negative frequencies occur. If only positive frequencies are to be allowed, one may write Eq. (2.1) as follows:

$$f(t) = c_0 + \sum_{n=1}^{\infty} \left(c_n e^{jn\omega t} + c_{-n} e^{-jn\omega t} \right) = \tag{2.5}$$

$$= c_0 + \sum_{n=1}^{\infty} \left(c_n \left[\cos(n\omega t) + j\ \sin(n\omega t) \right] \right.$$

$$\left. + c_{-n} \left[\cos(n\omega t) - j\ \sin(n\omega t) \right] \right). \tag{2.6}$$

We obtain the result

$$f(t) = c_0 + \sum_{n=1}^{\infty} \left[(c_n + c_{-n}) \cos(n\omega t) + j(c_n - c_{-n}) \sin(n\omega t) \right].$$

By means of the definition

$$a_n = c_n + c_{-n}$$

and

$$b_n = j(c_n - c_{-n}),$$

one obtains

$$\boxed{f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(n\omega t) + b_n \sin(n\omega t)\right].} \tag{2.7}$$

Taking $a_0 = 2c_0$ and $b_0 = 0$ into account, one may calculate the coefficients c_n if a_n and b_n are known:

$$c_n = \frac{a_n - jb_n}{2}. \tag{2.8}$$

For the special case that $c_n = 1/T$ holds for all n (Dirac comb; see Sect. 2.1.2), one obtains $a_n = 2/T$ and $b_n = 0$. According to Eq. (2.7), this means that the average, i.e., the DC component, of a strongly bunched beam is exactly one-half the fundamental harmonic:

$$\sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T} + \frac{2}{T} \sum_{n=1}^{\infty} \cos(n\omega t). \tag{2.9}$$

Now we return to the general case. Instead of using a_n and b_n, one may also use amplitudes and phases:

$$\boxed{f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} d_n \ \cos(n\omega t + \varphi_n).} \tag{2.10}$$

A comparison with Eq. (2.7) shows that

$$a_n \cos(n\omega t) + b_n \sin(n\omega t) = d_n \ \cos(n\omega t + \varphi_n)$$
$$\Rightarrow a_n \cos(n\omega t) + b_n \sin(n\omega t) = d_n \ \cos(n\omega t) \ \cos\varphi_n - d_n \ \sin(n\omega t) \ \sin\varphi_n.$$

This leads to the following conditions:

$$a_n = d_n \ \cos\varphi_n, \tag{2.11}$$

$$b_n = -d_n \ \sin\varphi_n. \tag{2.12}$$

According to Eq. (2.8), we therefore have

$$\varphi_n = \measuredangle\, c_n \tag{2.13}$$

and

$$d_n = \sqrt{a_n^2 + b_n^2}. \tag{2.14}$$

Due to

$$|c_n| = \frac{1}{2}\sqrt{a_n^2 + b_n^2}, \tag{2.15}$$

one obtains

$$d_n = 2|c_n| \tag{2.16}$$

as the physical amplitudes (peak values). By inserting Eqs. (2.11) and (2.12) into Eq. (2.8), one gets

$$c_n = \frac{d_n}{2} e^{j\varphi_n}.$$

The same result is obtained by combining Eqs. (2.13)–(2.15).

2.1.4 Discrete Fourier Transform

The discrete Fourier transform is a powerful tool for spectral analysis of signals that are given in digital form, e.g., on a computer. Therefore, we briefly discuss some important features here.

2.1.4.1 Motivation of the Transformation Formula

Let us now assume that a real-valued periodic function $f(t)$ with period $T = \frac{2\pi}{\omega}$ is discretized according to

$$f_k = f(k\Delta t),$$

where k is an integer. The period T is divided into $N \in \mathbb{N}$ time intervals

$$\Delta t = \frac{T}{N}$$

such that $f_0 = f_N$ holds. Therefore, the N samples $f_0, f_1, \ldots, f_{N-1}$ are sufficient to describe the function $f(t)$, provided that N is large enough. We now replace the integral in Eq. (2.3) by the Riemann sum

$$c_n \approx \frac{1}{T} \sum_{k=0}^{N-1} f(k\Delta t) e^{-jn\omega k\Delta t} \Delta t = \frac{1}{T} \sum_{k=0}^{N-1} f_k e^{-j2\pi nk/N} \Delta t = \frac{1}{N} \sum_{k=0}^{N-1} f_k e^{-j2\pi nk/N}.$$

This formula is used to define the **discrete Fourier transform (DFT)**

$$\boxed{X_n = \frac{1}{N} \sum_{k=0}^{N-1} x_k \; e^{-j2\pi nk/N}.} \tag{2.17}$$

This obviously yields an approximation of the Fourier coefficients c_n of the periodic function $f(t)$, provided that the number N of samples $x_k = f(k\Delta t)$ is large enough.

2.1.4.2 Symmetry Relations

Based on Eq. (2.17), we find that

$$X_{n+N} = \frac{1}{N} \sum_{k=0}^{N-1} x_k \; e^{-j2\pi nk/N} \; e^{-j2\pi Nk/N} = \frac{1}{N} \sum_{k=0}^{N-1} x_k \; e^{-j2\pi nk/N} \cdot 1 = X_n.$$

Therefore, all X_n are known if those for $0 \le n \le N-1$ are specified. One sees that for a sample $(x_0, x_1, \ldots, x_{N-1})$, one obtains a sample $(X_0, X_1, \ldots, X_{N-1})$ as the spectrum.

Since we have assumed that the signal $f(t)$ is real-valued and periodic, the same is true for the samples x_k. Based on Eq. (2.17), it is then obvious that the symmetry relation

$$X_{-n} = X_n^*$$

holds. We may also combine these two symmetry relations to obtain

$$X_{N-n} = X_{-(n-N)} = X_{n-N}^* = X_n^*.$$

Therefore, only about one-half of the coefficients X_n with $0 \le n \le N-1$ have to be calculated.

Table 2.1 Overview of DFT components of real-valued signals

Time	Sample	Frequency	Spectral component	Comment
0	x_0	0	X_0 (real)	DC component of the signal
Δt	x_1	$\frac{1}{T} = \frac{1}{N\Delta t} = \frac{1}{N} f_{\text{sampl}}$	X_1	Peak value: $2\lvert X_1\rvert$
$2\Delta t$	x_2	$\frac{2}{T} = \frac{2}{N\Delta t} = \frac{2}{N} f_{\text{sampl}}$	X_2	Peak value: $2\lvert X_2\rvert$
$3\Delta t$	x_3	$\frac{3}{T} = \frac{3}{N\Delta t} = \frac{3}{N} f_{\text{sampl}}$	X_3	Peak value: $2\lvert X_3\rvert$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(N-2)\Delta t$	x_{N-2}	$\frac{N-2}{T} = \frac{N-2}{N\Delta t} = \frac{N-2}{N} f_{\text{sampl}}$	$X_{N-2} = X_2^*$	Peak value: $2\lvert X_2\rvert$
$(N-1)\Delta t$	x_{N-1}	$\frac{N-1}{T} = \frac{N-1}{N\Delta t} = \frac{N-1}{N} f_{\text{sampl}}$	$X_{N-1} = X_1^*$	Peak value: $2\lvert X_1\rvert$
$N\Delta t$	x_N	$\frac{N}{T} = \frac{1}{\Delta t} = f_{\text{sampl}}$	$X_N = X_0$	Repetition
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

2.1.4.3 Interpretation of the Spectral Components

According to Eq. (2.1), the sample X_0 belongs to the DC component of the signal. The sample X_1 obviously belongs to the angular frequency

$$1 \cdot \omega = \frac{2\pi}{T}.$$

Therefore, the spectrum $(X_0, X_1, \ldots, X_{N-1})$ has a resolution of $f = 1/T$, where T is the total time that passes between the samples x_0 and x_N. It is obvious that X_{N-1} belongs to the frequency

$$f_{\max} = \frac{N-1}{T} = \frac{N-1}{N}\frac{1}{\Delta t} \approx \frac{1}{\Delta t} = f_{\text{sampl}}.$$

This approximation is, of course, valid only for large samples with $N \gg 1$. Hence we conclude that the frequency resolution is given by the inverse of the total time T, whereas the maximum frequency is determined by the sampling frequency $f_{\text{sampl}} = 1/\Delta t$. However, due to $X_{N-n} = X_n^*$, only one-half of this frequency range between 0 and $f_{\max}$ actually contains information. In other words, and in compliance with the **Nyquist–Shannon sampling theorem**, sampling has to take place with at least twice the signal bandwidth.

These properties are visualized in Table 2.1.

If one makes sure that the N equidistant samples x_n of the periodic function represent an integer number of periods (so that duplicating $(x_0, x_1, \ldots, x_{N-1})$ does not introduce any severe discontinuities), one may obtain good results even without sophisticated windowing techniques.

For the interpretation of the spectrum, please note that the DC component is equal to

$$f_{\text{DC}} = \frac{a_0}{2} = c_0 = X_0,$$

i.e., to the first value of the DFT.

According to Eq. (2.16), the amplitude (peak value) at the frequency p/T is given by

$$d_p = 2|c_p| = 2|X_p|.$$

The discussion above shows that the sample $(X_0, X_1, \ldots, X_{N-1})$ contains all the information about the spectrum, but that the DFT spectrum is infinite. It does not even decrease with increasing frequencies. At first glance, this looks strange, but in our introduction to the DFT, we assumed only that the integral over Δt may approximately be replaced by a product with Δt. We made no assumption as to how the function $f(t)$ varies in the interval Δt. This explains the occurrence of the high-frequency components.

It should be clear from the Nyquist–Shannon sampling theorem that the spectrum for frequencies larger than $f_{\max}/2$ cannot contain any relevant information, since the sampling frequency is fixed at $\Delta t \approx 1/f_{\max}$.

Therefore, in the next section, we filter out those frequencies to obtain the inverse transform.

2.1.4.4 Inverse DFT

As mentioned above, the Nyquist–Shannon sampling theorem tells us that we should consider only frequencies f_x with

$$-\frac{f_{\max}}{2} \le f_x \le +\frac{f_{\max}}{2}.$$

This corresponds to

$$-\frac{N-1}{2T} \le f_x \le +\frac{N-1}{2T},$$

or

$$-\frac{N-1}{2}\omega \le \omega_x \le +\frac{N-1}{2}\omega.$$

For the sake of simplicity, we assume that $N \ge 3$ is an odd number. If we have a look at Eq. (2.1),

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \, e^{jn\omega t},$$

it becomes clear that only those n with

$$-\frac{N-1}{2} \le n \le +\frac{N-1}{2}$$

lead to the aforementioned frequencies $\omega_x = 2\pi f_x = n\omega$. Therefore, we expect to be able to reconstruct the signal based on

$$f(t) = \sum_{n=-(N-1)/2}^{+(N-1)/2} c_n \, e^{jn\omega t}.$$

We now apply the discretization

$$f_k = f(k\Delta t) = \sum_{n=-(N-1)/2}^{+(N-1)/2} c_n \, e^{jn\omega k\Delta t} = \sum_{n=-(N-1)/2}^{+(N-1)/2} c_n \, e^{j2\pi nk/N} \tag{2.18}$$

and obtain

$$\sum_{n=(N+1)/2}^{N-1} c_n \, e^{j2\pi nk/N} = \sum_{l=-N/2+1/2}^{-1} c_{l+N} \, e^{j2\pi k\frac{l+N}{N}}.$$

Here we introduced the new summation index $l = n - N$. The last formula leads to

$$\sum_{n=(N+1)/2}^{N-1} c_n \, e^{j2\pi nk/N} = \sum_{l=-(N-1)/2}^{-1} c_l \, e^{j2\pi kl/N}.$$

On the right-hand side, we may now rename l as n again. This shows that the sum from $-(N-1)/2$ to -1 included in Eq. (2.18) may be replaced by the sum from $(N+1)/2$ to $N-1$:

$$f_k = \sum_{n=0}^{N-1} c_n \, e^{j2\pi nk/N}.$$

This defines the formula for the inverse DFT (not only for odd N):

$$\boxed{x_k = \sum_{n=0}^{N-1} X_n \, e^{j2\pi nk/N}.}$$

Please note that in the literature, the factor $1/N$ is sometimes not included in the definition of the DFT, but it appears in that of the inverse DFT. Our choice was determined by the close relationship to the Fourier series coefficients discussed above. Apart from the factor $1/N$, the DFT and the inverse DFT differ only by the sign in the argument of the exponential function.

2.1.4.5 Conclusion

We have summarized only a few basic facts that will help the reader to interpret the DFT correctly. There are many other properties that cannot be mentioned here.

For large sample sizes equal to a power of 2, the so-called **fast Fourier transform (FFT)** algorithm may be used, which is a dramatically less time-consuming implementation of the DFT.

2.1.5 Fourier Transform

The **Fourier transform** $X(\omega)$ of a real-valued function $x(t)$ depending on the time variable t is given by

$$\boxed{X(\omega) = \int_{-\infty}^{+\infty} x(t)\, e^{-j\omega t}\, \mathrm{d}t,} \tag{2.19}$$

the inverse transform by

$$\boxed{x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega)\, e^{j\omega t}\, \mathrm{d}\omega.} \tag{2.20}$$

This relation is visualized by the correspondence symbol

$$x(t) \circ\!\!-\!\!\bullet\, X(\omega).$$

The Fourier transform is a linear transformation. It is used to determine the frequency spectrum of signals, i.e., it transforms the signal $x(t)$ from the time domain into the frequency domain. It is possible to generalize the definition of the Fourier transform to generalized functions (i.e., distributions), which also include the Dirac function [1, 2].

Please note that various definitions for the Fourier transform and for its inverse transform exist in the literature. The factor $\frac{1}{2\pi}$ may be distributed among the original transformation and the inverse transformation in a different way, and even the sign of the argument of the exponential function may be defined in the opposite way.

Some common Fourier transforms are summarized in Table A.3 on p. 417. Further relations can also be found using symmetry properties of the Fourier transform. Consider the Fourier transform

$$x(t) \quad \circ\!\!-\!\!\bullet \quad X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, \mathrm{d}t.$$

If the time t in $x(t)$ is replaced by ω, and $x(\omega)$ is regarded as a Fourier transform, its inverse transform is given by

$$x(\omega) \quad \bullet\!\!-\!\!\circ \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} x(\omega)\, e^{j\omega t}\, \mathrm{d}\omega = \frac{1}{2\pi} X(-t).$$

In other words, the inverse transform of $x(\omega)$ is obtained by replacing ω in the function $X(\omega)$ by $-t$.

2.1.5.1 Fourier Transform of a Single Cosine Pulse

Let

$$x(t) = \begin{cases} 1 + \cos(\Omega t) & \text{for } -\pi < \Omega t < \pi, \\ 0 & \text{otherwise,} \end{cases} \tag{2.21}$$

define a single cosine pulse. This leads to

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{+\infty} x(t)\, e^{-j\omega t}\mathrm{d}t = \int_{-\pi/\Omega}^{+\pi/\Omega} [1 + \cos(\Omega t)]\; e^{-j\omega t}\mathrm{d}t = \\ &= \int_{-\pi/\Omega}^{+\pi/\Omega} \left[e^{-j\omega t} + \frac{1}{2}\, e^{j(\Omega-\omega)t} + \frac{1}{2}\, e^{-j(\Omega+\omega)t} \right] \mathrm{d}t = \\ &= \left[\frac{e^{-j\omega t}}{-j\omega} + \frac{1}{2}\, \frac{e^{j(\Omega-\omega)t}}{j(\Omega-\omega)} + \frac{1}{2}\, \frac{e^{-j(\Omega+\omega)t}}{-j(\Omega+\omega)} \right]_{-\pi/\Omega}^{+\pi/\Omega} = \\ &= \sin\left(\pi \frac{\omega}{\Omega}\right) \left[\frac{2}{\omega} + \frac{1}{\Omega-\omega} - \frac{1}{\Omega+\omega} \right] = \sin\left(\pi \frac{\omega}{\Omega}\right) \frac{2\Omega^2}{\omega(\Omega^2-\omega^2)} \end{aligned}$$

$$\Rightarrow X(\omega) = \frac{2\pi}{\Omega}\, \frac{\mathrm{si}\left(\pi \frac{\omega}{\Omega}\right)}{1 - \left(\frac{\omega}{\Omega}\right)^2}. \tag{2.22}$$

In the last equation, we used the definition

$$\mathrm{si}(x) = \begin{cases} \frac{\sin x}{x} & \text{for } x \neq 0, \\ 1 & \text{for } x = 0. \end{cases}$$

For the sake of uniqueness, we call this function $\mathrm{si}(x)$ instead of $\mathrm{sinc}(x)$.

2.1.5.2 Convolution

The **convolution** is given by

$$\boxed{h(t) * x(t) = \int_{-\infty}^{+\infty} h(\tau)\, x(t-\tau)\, \mathrm{d}\tau,}$$

and one obtains

$$x(t) * h(t) = h(t) * x(t) \circ\!\!-\!\!\bullet\, H(\omega)X(\omega).$$

We consider the special case that

$$h(t) = \sum_k \delta(t - T_k)$$

is a sequence of Dirac pulses. This leads to

$$h(t) * x(t) = \sum_k \int_{-\infty}^{+\infty} \delta(\tau - T_k)\; x(t - \tau)\; \mathrm{d}\tau = \sum_k x(t - T_k).$$

Hence, by convolution with a sequence of Dirac pulses, we may produce a repetition of the function $x(t)$ at the locations of the delta pulses.

2.1.5.3 Relation to the Fourier Series

We consider the special case

$$X(\omega) = \sum_{k=-\infty}^{+\infty} p_k\; \delta(\omega - k\omega_0).$$

According to Eq. (2.20), this leads to

$$x(t) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_k\; \delta(\omega - k\omega_0)\; e^{j\omega t}\; \mathrm{d}\omega = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} p_k\; e^{jk\omega_0 t}.$$

If we set

$$p_k = 2\pi c_k,$$

we obtain the correspondence

$$x(t) = \sum_{k=-\infty}^{+\infty} c_k\; e^{jk\omega_0 t} \qquad \circ\!\!-\!\!\bullet \qquad X(\omega) = 2\pi \sum_{k=-\infty}^{+\infty} c_k\; \delta(\omega - k\omega_0),$$

which is an ordinary Fourier series, as Eq. (2.1) shows.

Hence, if we calculate the Fourier transform of a periodic function with period $T_0 = \frac{2\pi}{\omega_0}$, we get a sum of Dirac pulses that are multiplied by 2π and the Fourier coefficients. The factor 2π is obvious because of the correspondence

$$1 \circ\!\!-\!\!\bullet\, 2\pi\delta(\omega).$$

2.1.6 Consequences for the Spectrum of the Beam Signal

We first model an idealized beam signal $h(t)$ as a periodic sequence of Dirac pulses. Even if the bunches oscillate in the longitudinal direction, periodicity may be satisfied if the beam signal repeats itself after one synchrotron oscillation period. The sequence of delta pulses will be defined by

$$h(t) = \sum_k \delta(t - T_k)$$

as above. Thus, we get a realistic beam signal by convolution with the time function $x(t)$, which represents a single bunch:

$$y(t) = h(t) * x(t).$$

Since $h(t)$ is to be periodic, it may be represented by a Fourier series. As shown in the previous section, this leads to the Fourier transform

$$H(\omega) = 2\pi \sum_{k=-\infty}^{+\infty} c_k^{h(t)}\ \delta(\omega - k\omega_0).$$

The function $x(t)$ describes a single pulse and is therefore equal to zero outside a finite interval. Therefore, the spectrum $X(\omega)$ will be continuous. This shows that

$$\begin{aligned} Y(\omega) &= H(\omega)\ X(\omega) \\ &= 2\pi \sum_{k=-\infty}^{+\infty} c_k^{h(t)}\ X(\omega)\ \delta(\omega - k\omega_0) \\ &= 2\pi \sum_{k=-\infty}^{+\infty} c_k^{h(t)}\ X(k\omega_0)\ \delta(\omega - k\omega_0) \end{aligned}$$

is a Fourier series whose Fourier coefficients are

$$c_k^{y(t)} = c_k^{h(t)}\ X(k\omega_0). \tag{2.23}$$

As an example and as a test of the results obtained so far, we analyze the convolution of a Dirac comb

$$h(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_0)$$

with a single cosine pulse. According to Eq. (2.4), the Fourier coefficients of the Dirac comb are

$$c_k^{h(t)} = \frac{1}{T_0}.$$

Here T_0 denotes the time span between the pulses. For the single cosine pulse with time span $T = \frac{2\pi}{\Omega}$ that was defined in Eq. (2.21), one obtains—based on Eq. (2.22)—the Fourier transform

$$X(\omega) = \frac{2\pi}{\Omega} \frac{\mathrm{si}\left(\pi \frac{\omega}{\Omega}\right)}{1 - \left(\frac{\omega}{\Omega}\right)^2}.$$

According to Eq. (2.23), the Fourier coefficients of the convolution function $y(t) = h(t) * x(t)$ are therefore

$$c_k^{y(t)} = \frac{1}{T_0} \frac{2\pi}{\Omega} \frac{\mathrm{si}\left(\pi k \frac{\omega_0}{\Omega}\right)}{1 - \left(k \frac{\omega_0}{\Omega}\right)^2} = \frac{\omega_0}{\Omega} \frac{\mathrm{si}\left(\pi k \frac{\omega_0}{\Omega}\right)}{1 - \left(k \frac{\omega_0}{\Omega}\right)^2}. \tag{2.24}$$

We will now analyze this result for several special cases.

- **Constant beam current:** In this first case, we assume that the different single-cosine pulses overlap according to $T = 2T_0$, which is equivalent to $\omega_0 = 2\Omega$. In this case, we obtain $c_k^{y(t)} = 0$ for $k \neq 0$. For $c_0^{y(t)}$, which corresponds to the DC component, one obtains

  $$c_0^{y(t)} = \frac{1}{T_0} \frac{2\pi}{\Omega} = 2,$$

  which is the expected result for a constant function that equals 2.
- **Continuous sine wave:** In this case, we make use of the simplification $\Omega = \omega_0$, so that $y(t)$ corresponds to a simple cosine function that is shifted upward:

  $$c_k^{y(t)} = \frac{\mathrm{si}(\pi k)}{1 - k^2}.$$

  We obviously have

  $$c_0^{y(t)} = 1.$$

  For $k = \pm 1$, we may use l'Hôpital's rule:

  $$c_{\pm 1}^{y(t)} = \lim_{k \to \pm 1} \frac{\mathrm{si}(\pi k)}{1 - k^2} = \lim_{k \to \pm 1} \frac{\sin(\pi k)}{\pi\ (k - k^3)} = \lim_{k \to \pm 1} \frac{\pi\ \cos(\pi k)}{\pi\ (1 - 3k^2)} = \frac{1}{2}.$$

All other coefficients are zero. Thus we obtain

$$y(t) = \sum_{k=-\infty}^{+\infty} c_k^{y(t)} e^{jk\omega_0 t} = 1 + \frac{1}{2}\, e^{j\omega_0 t} + \frac{1}{2}\, e^{-j\omega_0 t} = 1 + \cos(\omega_0 t),$$

which is in accordance with our expectation.

- **Dirac comb:** For this last case, we first observe that the area under each single-cosine pulse defined in Eq. (2.21) is T. If we want to have an area of 1 instead, we have to divide the function $y(t)$ by T:

$$\tilde{y}(t) = \frac{y(t)}{T}.$$

Hence, the Fourier coefficients in Eq. (2.24) also have to be divided by T:

$$c_k^{\tilde{y}(t)} = \frac{1}{T_0} \frac{\mathrm{si}\left(\pi k\, \frac{\omega_0}{\Omega}\right)}{1 - \left(k\, \frac{\omega_0}{\Omega}\right)^2}.$$

We now consider the case $T \to 0$ while assuming a fixed value of T_0. Hence $\omega_0/\Omega \to 0$, and we obtain

$$c_k^{\tilde{y}(t)} = \frac{1}{T_0},$$

which is the expected result for a Dirac comb.

Finally, our simple beam signal model that was constructed by a combination of single-cosine pulses is able to describe all states between unbunched beams and strongly bunched beams. In the case of long bunches (continuous sine wave), the DC current equals the RF current amplitude. As the bunches become shorter ($\omega_0 < \Omega$), Eq. (2.24) can be used to determine the ratio between RF current amplitude and DC current.

2.2 Laplace Transform

The **Laplace transform** is one of the standard tools used to analyze closed-loop control systems. In the scope of the book at hand, we deal only with the one-sided Laplace transform [3, 4], which is useful because processes can be described whereby signals are switched on at $t = 0$. Hence, the name "Laplace transform" will be used as a synonym for "one-sided Laplace transform." Such a one-sided Laplace transform of a function $f(t)$ with $f(t) = 0$ for $t < 0$ is given by

$$\boxed{F(s) = \int_0^\infty f(t)\, e^{-st}\, \mathrm{d}t.} \tag{2.25}$$

Here $s = \sigma + j\omega$ is a complex parameter. It is obvious that the Laplace transform has a close relationship to the Fourier transform that is obtained for $\sigma = 0$ if only functions with $f(t) = 0$ for $t < 0$ are allowed. The real part of s is usually introduced to obtain convergence for a larger class of functions (please note that the Fourier transform of a sine or cosine function already leads to nonclassical Dirac pulses, as we saw in Sect. 2.1.5.3).

The Laplace transform $F(s)$ of a function $f(t)$ is an analytic function, and there is a unique correspondence between $f(t)$ and $F(s)$ if the classes of functions/distributions that are considered in the time domain and the Laplace domain are chosen accordingly [1, 4]. Since the integral in Eq. (2.25) exists only in some region of the complex plane, the Laplace transform is initially defined in only this region as well. If, however, a closed-form expression is obtained for the Laplace transform, e.g., a rational function, it is possible to extend the domain of definition by means of analytic continuation (cf. [5, Sect. 2.1]; [6, Sect. 10-9]; [7, Sect. 5.5.4]). Therefore, the Laplace transform $F(s)$ should be defined as the analytic continuation of the function defined by Eq. (2.25). Apart from poles, a Laplace transform $F(s)$ may thus be defined in the whole complex plane.

Like the Fourier transform, the Laplace transform is a linear transformation. If according to

$$f(t) \circ\!\!-\!\!\bullet\, F(s), \qquad g(t) \circ\!\!-\!\!\bullet\, G(s),$$

we use the correspondence symbol again, the Laplace transform has the following properties (n is a positive integer, and a is a real number):

- Laplace transform of a derivative[1]:

$$\frac{\mathrm{d}f}{\mathrm{d}t} \circ\!\!-\!\!\bullet\, s\ F(s) - f(0+),$$

$$\frac{\mathrm{d}^n f}{\mathrm{d}t^n} \circ\!\!-\!\!\bullet\, s^n F(s) - s^{n-1} f(0+) - s^{n-2}\frac{\mathrm{d}f}{\mathrm{d}t}(0+) - \cdots - s^0 \frac{\mathrm{d}^{n-1} f}{\mathrm{d}t^{n-1}}(0+).$$

- Derivative of a Laplace transform:

$$-t\ f(t) \circ\!\!-\!\!\bullet\, \frac{\mathrm{d}F(s)}{\mathrm{d}s}, \qquad t^n\ f(t) \circ\!\!-\!\!\bullet\, (-1)^n \frac{\mathrm{d}^n F(s)}{\mathrm{d}s^n}.$$

- Laplace transform of an integral:

$$\int_0^t f(\tau)\ \mathrm{d}\tau \circ\!\!-\!\!\bullet\, \frac{F(s)}{s}.$$

[1] We use the notation

$$f(0+) := \lim_{\epsilon \to 0} f(\epsilon) \qquad \text{with} \quad \epsilon > 0.$$

- Shift theorems:

$$e^{-at} f(t) \circ\!\!-\!\!\bullet\, F(s+a),$$

$$f(t-a) \circ\!\!-\!\!\bullet\, e^{-as} F(s) \qquad \text{for} \quad a > 0. \tag{2.26}$$

- Convolution:

$$f * g = g * f \circ\!\!-\!\!\bullet\, F(s)\, G(s). \tag{2.27}$$

- Scaling ($a > 0$):

$$f(at) \circ\!\!-\!\!\bullet\, \frac{1}{a} F\left(\frac{s}{a}\right),$$

$$\frac{1}{a} f\left(\frac{t}{a}\right) \circ\!\!-\!\!\bullet\, F(as).$$

- Limits:

$$f(0+) = \lim_{s\to\infty} (s\, F(s)),$$

$$f(\infty) := \lim_{t\to\infty} f(t) = \lim_{s\to 0} (s\, F(s)). \tag{2.28}$$

  Here f and its derivative must satisfy further requirements [4]. Before using the final-value theorem (2.28), for example, one should verify that the function actually converges for $t \to \infty$.

Like the Fourier transform, the Laplace transform may also be generalized in order to cover distributions (i.e., generalized functions) [1]. Some common Laplace transforms are summarized in Table A.4 on p. 418.

2.3 Transfer Functions

Some dynamical systems[2] may be described by the equation

$$Y(s) = H(s)\, X(s).$$

[2] This will be discussed in Sect. 7.1.1.

In this case, $X(s)$ and $Y(s)$ are the Laplace transforms of the input signal $x(t)$ and the output signal $y(t)$, respectively. The Laplace transform $H(s)$ is called the **transfer function** of the system. We discuss two specific input signals:

- Let us assume that the input function $x(t)$ is a **Heaviside step function**

$$\Theta(t) = \begin{cases} 0 & \text{for } t < 0, \\ 1 & \text{for } t \geq 0, \end{cases} \quad \circ\!\!-\!\!\bullet \quad \frac{1}{s}.$$

  In this case, the output is

$$Y(s) = \frac{H(s)}{s}.$$

  If we now apply Eq. (2.28), we obtain

$$y(\infty) = \lim_{s \to 0} H(s)$$

  as the long-term **(unit-)step response** of the system.
- If generalized functions are allowed, we may use $x(t) = \delta(t)$ as an input signal. In this case, the correspondence

$$\delta(t) \circ\!\!-\!\!\bullet \; 1$$

  leads to

$$Y(s) = H(s),$$

  which means that the transfer function $H(s)$ corresponds to the **impulse response** $h(t)$ of the system. The final value of the response $y(t) = h(t)$ is then given by

$$y(\infty) = \lim_{s \to 0} \left(s \; H(s) \right).$$

Let us assume that a system component is specified by the transfer function $H(s)$. If we calculate the **phase response**[3] of this component according to $\varphi(\omega) = \measuredangle H(j\omega)$, the **group delay** can be defined by

$$\tau_{\mathrm{g}} = -\frac{\mathrm{d}\varphi}{\mathrm{d}\omega}.$$

[3]The function $A(\omega) = |H(j\omega)|$ is the **amplitude response**. Phase response and amplitude response define the **frequency response** $H(j\omega)$.

Taking a **dead-time element** with $H(s) = e^{-sT_{\text{dead}}}$ (see shift theorem (2.26)) as an example, one obtains the frequency-independent, i.e., constant group delay

$$\tau_{\text{g}} = T_{\text{dead}}.$$

Hence, the dead-time element is an example of a device with linear phase response.

2.4 Mathematical Statistics

The results summarized in this chapter can be read in more detail in [8].

2.4.1 Gaussian Distribution

The **Gaussian distribution** (also called the normal distribution) is given by the **probability density function**

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},} \tag{2.29}$$

where $\mu, \sigma \in \mathbb{R}$ with $\sigma > 0$ are specified. In order to ensure that $f(x)$ is in fact a valid probability distribution, the equation

$$\boxed{\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1}$$

must hold. We show this by substituting

$$u = \frac{x-\mu}{\sigma}, \qquad \frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{\sigma}.$$

This leads to

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}\,\mathrm{d}u.$$

By means of standard methods of mathematical analysis, one may show that

$$\int_0^{\infty} e^{-a^2u^2}\,\mathrm{d}u = \frac{\sqrt{\pi}}{2a},$$

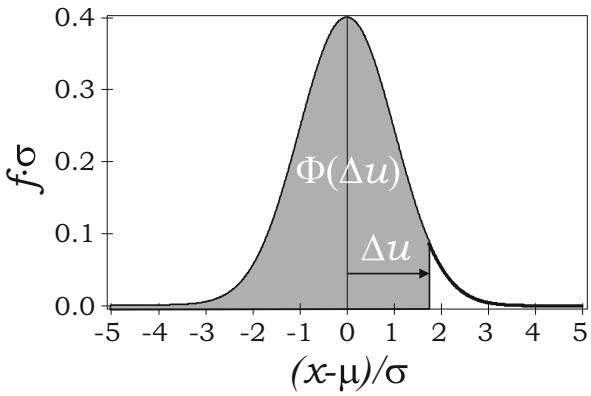

Fig. 2.1 Gaussian distribution

which actually leads to the result

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2}\,\mathrm{d}u = 1. \tag{2.30}$$

For a given measurement curve that has the shape of a Gaussian distribution, one may use curve-fitting techniques to determine the parameters μ and σ. A simpler method is to determine the **FWHM (full width at half maximum) value**. According to Eq. (2.29), one-half of the maximum value is obtained for

$$e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \stackrel{!}{=} \frac{1}{2} \qquad \Rightarrow -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 = -\ln\,2 \qquad \Rightarrow |x-\mu| = \sigma\sqrt{2\,\ln\,2}.$$

The FWHM value equals twice this distance (one to the left of the maximum and one to the right of the maximum):

$$\mathrm{FWHM} = \sigma\;2\;\sqrt{2\;\ln\;2} \approx 2.35482\;\sigma.$$

This formula may, of course, lead to less-accurate results than those obtained by the curve-fitting concept if zero line or the maximum cannot be clearly identified in the measurement data.

2.4.2 Probabilities

We now consider the area below the curve $f(x)$ that is located to the left of $x = \mu + \Delta x$, where $\Delta x > 0$ holds. This area will be denoted by Φ:

$$\Phi = \int_{-\infty}^{\mu+\Delta x} f(x)\,\mathrm{d}x.$$

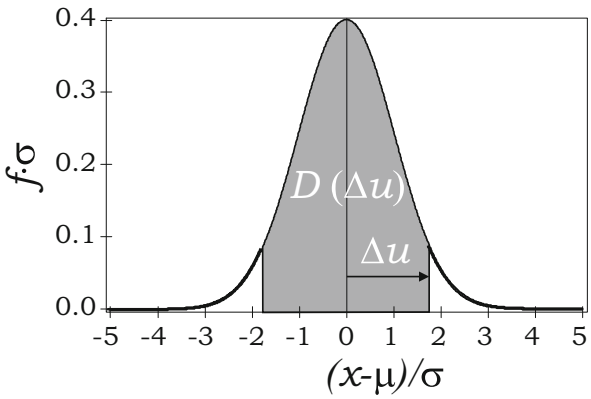

Fig. 2.2 Gaussian distribution

It obviously specifies the probability that the **random variable** X is less than $\mu + \Delta x$. By applying the same substitution as that mentioned above, one obtains

$$\Phi = \int_{-\infty}^{\Delta x/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \, \mathrm{d}u.$$

According to Fig. 2.1 we set

$$\Delta u = \frac{\Delta x}{\sigma}$$

and get

$$\Phi(\Delta u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\Delta u} e^{-\frac{1}{2}u^2} \, \mathrm{d}u.$$

The area D that is enclosed between $\mu - \Delta x$ and $\mu + \Delta x$ (see Fig. 2.2) can be calculated as follows:

$$D(\Delta u) = \Phi(\Delta u) - \Phi(-\Delta u).$$

Due to symmetry, we have

$$\Phi(-\Delta u) = 1 - \Phi(\Delta u),$$

which leads to

$$D(\Delta u) = 2\,\Phi(\Delta u) - 1.$$

Table 2.2 Integrals of the Gaussian probability density function

Δu	$\Phi_0(\Delta u)$	Δx	$D(\Delta u) = 2\ \Phi_0(\Delta u)$
0	0	0	0
0.5	0.1915	$0.5\ \sigma$	0.3829
1	0.3413	σ	0.6827
1.5	0.4332	$1.5\ \sigma$	0.8664
1.6449	0.45	$1.6449\ \sigma$	0.9000
1.9600	0.475	$1.9600\ \sigma$	0.9500
2	0.4772	$2\ \sigma$	0.9545
2.5	0.4938	$2.5\ \sigma$	0.9876
2.5758	0.4950	$2.5758\ \sigma$	0.9900
3	0.4987	$3\ \sigma$	0.9973
3.2906	0.4995	$3.2906\ \sigma$	0.9990
3.5	0.4998	$3.5\ \sigma$	0.9995

Often, the area Φ_0 is considered, which is located between μ and $\mu + \Delta x$:

$$\Phi_0(\Delta u) = \Phi(\Delta u) - \frac{1}{2}, \qquad \Phi_0(\Delta u) = \frac{1}{\sqrt{2\pi}} \int_0^{\Delta u} e^{-\frac{1}{2}u^2}\ \mathrm{d}u.$$

This shows that D may also be written in the form

$$D(\Delta u) = 2\ \Phi_0(\Delta u).$$

Some examples for these quantities are summarized in Table 2.2.

As an example, the table shows that the random variable is located in the **confidence interval** between $\mu - 2\sigma$ and $\mu + 2\sigma$ with a probability of 95.45%.

2.4.3 Expected Value

Let X be a random variable with probability density function $f(x)$. Then the **expected value** of the function $g(X)$ is given by

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x)\ f(x)\ \mathrm{d}x.$$

It is obvious that the expected value is linear:

$$E(a\ g_1(X) + b\ g_2(X)) = a\ E(g_1(X)) + b\ E(g_2(X)).$$

For $g(X) = X^k$, one obtains the kth **moment**:

$$E(X^k) = \int_{-\infty}^{+\infty} x^k\ f(x)\ \mathrm{d}x.$$

By definition, the first moment is the **mean** of the random variable X. For the Gaussian distribution, we obtain

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma u + \mu)\, e^{-\frac{1}{2}u^2} \,\mathrm{d}u.$$

The term σu in the parentheses leads to an odd integrand, so that this part of the integral vanishes.

Using Eq. (2.30), one obtains the mean

$$E(X) = \mu,$$

which is geometrically obvious.

If we always (not only for the Gaussian distribution) denote the mean by μ, then the kth **central moment** is given by

$$E((X-\mu)^k) = \int_{-\infty}^{+\infty} (x-\mu)^k \; f(x)\,\mathrm{d}x.$$

The second central moment is called the **variance**. For the Gaussian distribution, we obtain

$$E((X-\mu)^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma u)^2\, e^{-\frac{1}{2}u^2} \,\mathrm{d}u.$$

With

$$a = u\, e^{-u^2/4}, \quad a' = e^{-u^2/4}\left(1 - \frac{u^2}{2}\right),$$
$$b' = u\, e^{-u^2/4}, \quad b = -2\, e^{-u^2/4},$$

an integration by parts yields

$$\int_{-\infty}^{\infty} u^2\, e^{-u^2/2} \,\mathrm{d}u = -2u\, e^{-u^2/2}\Big|_{-\infty}^{\infty} + 2\int_{-\infty}^{+\infty} e^{-u^2/2}\left(1 - \frac{u^2}{2}\right) \mathrm{d}u.$$

The first term on the right-hand side vanishes, and we get

$$2\int_{-\infty}^{\infty} u^2\, e^{-u^2/2} \,\mathrm{d}u = 2\int_{-\infty}^{+\infty} e^{-u^2/2} \,\mathrm{d}u.$$

The remaining integral is known from Eq. (2.30):

$$\int_{-\infty}^{\infty} u^2 \, e^{-u^2/2} \, \mathrm{d}u = \sqrt{2\pi}.$$

Hence we obtain

$$E((X-\mu)^2) = \sigma^2.$$

The variance is generally denoted by σ^2 (not only for the Gaussian distribution), and its square root, the value σ, is called the **standard deviation**.

For a random sample with m values $x_1, x_2, \ldots, x_m$, one defines the **sample mean**

$$\bar{x} = \frac{1}{m} \sum_{k=1}^{m} x_k$$

and the **sample variance**

$$s^2 = \frac{1}{m-1} \sum_{k=1}^{m} (x_k - \bar{x})^2.$$

For large samples, this value does not deviate much from $\Delta x_{\mathrm{rms}}^2$, where the **root mean square (rms)** is defined[4] as

$$\Delta x_{\mathrm{rms}} = \sqrt{\frac{1}{m} \sum_{k=1}^{m} (x_k - \bar{x})^2}.$$

[4]The **root mean square (rms)** of a continuous-time signal $f(t)$ in the interval $[T_1, T_2]$ is defined by

$$\boxed{f_{\mathrm{rms}} = \sqrt{\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} f^2(t) \, \mathrm{d}t}.} \tag{2.31}$$

If the time interval is divided into m equal subintervals, one obtains approximately

$$f_{\mathrm{rms}} = \sqrt{\frac{1}{T_2 - T_1} \sum_{k=1}^{m} f_k^2 \frac{T_2 - T_1}{m}}$$

$$\Rightarrow \boxed{f_{\mathrm{rms}} = \sqrt{\frac{1}{m} \sum_{k=1}^{m} f_k^2}} \tag{2.32}$$

if f_k is regarded as a (time-discrete) sample of $f(t)$ in the kth subinterval. This equation is used in general to define the rms value of a set of values f_k ($k \in \{1, 2, \ldots, m\}$).

2.4.4 Unbiasedness

The individual values x_k of a sample are the observed realizations of the random variables X_k that belong to the same distribution. Also,

$$\bar{X} = \frac{1}{m} \sum_{k=1}^{m} X_k$$

is a random variable for which one may calculate the expected value. From $E(X_k) = \mu$ we obtain

$$E(\bar{X}) = \frac{1}{m} \sum_{k=1}^{m} E(X_k) = \mu,$$

which means that $\bar{X}$ is an **unbiased estimator** of the mean value μ of the population. We now check whether the sample variance

$$S^2 = \frac{1}{m-1} \sum_{k=1}^{m} (X_k - \bar{X})^2$$

is unbiased as well. We have

$$E(S^2) = \frac{1}{m-1} \sum_{k=1}^{m} \left[E(X_k^2) - 2E(X_k \bar{X}) + E(\bar{X}^2) \right]. \tag{2.33}$$

First of all, we need an expression for $E(X_k^2)$. For this purpose, we point out that all the random variables X_k belong to the same distribution, so that

$$\sigma^2 = E((X_k - \mu)^2) = E(X_k^2) - 2\mu E(X_k) + \mu^2$$

holds. From $E(X_k) = \mu$, we obtain

$$\sigma^2 = E(X_k^2) - \mu^2$$

and

$$E(X_k^2) = \sigma^2 + \mu^2. \tag{2.34}$$

Now we analyze the second expression in Eq. (2.33), i.e., the expected value of

$$X_k \bar{X} = \frac{1}{m} \sum_{l=1}^{m} X_k X_l.$$

For **independent random variables** X and Y, we have the equation

$$E(XY) = E(X)\ E(Y).$$

In our case, this is satisfied only for $k \neq l$, which means for $m - 1$ terms. The term with $k = l$ leads to the expected value $E(X_k^2)$ derived above. Therefore, we have

$$E(X_k \bar{X}) = \frac{1}{m} \sum_{l=1}^{m} E(X_k X_l) = \frac{1}{m} \left[(m-1)\mu^2 + (\sigma^2 + \mu^2)\right] = \mu^2 + \frac{\sigma^2}{m}. \tag{2.35}$$

Finally, we calculate the expected value of

$$\bar{X}^2 = \frac{1}{m^2} \sum_{k=1}^{m} \sum_{l=1}^{m} X_k X_l$$

in an analogous way, obtaining

$$E(\bar{X}^2) = \frac{1}{m^2} \left((m^2 - m)\mu^2 + m(\sigma^2 + \mu^2)\right) = \mu^2 + \frac{\sigma^2}{m}. \tag{2.36}$$

The results (2.34)–(2.36) may now be used in Eq. (2.33):

$$\begin{aligned} E(S^2) &= \frac{m}{m-1} \left[(\sigma^2 + \mu^2) - 2\left(\mu^2 + \frac{\sigma^2}{m}\right) + \left(\mu^2 + \frac{\sigma^2}{m}\right)\right] \\ &= \frac{m}{m-1} \left(\sigma^2 - \frac{\sigma^2}{m}\right) = \sigma^2. \end{aligned}$$

This shows that the sample variance is an unbiased estimator of the population variance. This is obviously not true for rms values. For large samples, however, this difference is no longer important.

We now calculate the variance of the sample mean $\bar{X}$:

$$E((\bar{X} - \mu)^2) = E(\bar{X}^2) - 2\mu E(\bar{X}) + \mu^2 = \left(\mu^2 + \frac{\sigma^2}{m}\right) - \mu^2 = \frac{\sigma^2}{m}.$$

This shows that an estimate of the population mean from the sample mean becomes better as the sample size becomes larger.

2.4.5 *Uniform Distribution*

According to

$$f(x) = \begin{cases} \frac{1}{2\Delta x} & \text{for } |x-\mu| \le \Delta x, \\ 0 & \text{elsewhere,} \end{cases} \qquad \text{with the constant } \Delta x > 0,$$

we now calculate the variance of a uniform distribution:

$$\sigma^2 = E((X-\mu)^2) = \int_{-\infty}^{+\infty} (x-\mu)^2 \, f(x) \, \mathrm{d}x = \int_{-\infty}^{+\infty} u^2 \, f(u+\mu) \, \mathrm{d}u.$$

In the last step, we substituted $u = x - \mu$ to obtain

$$\sigma^2 = \int_{-\Delta x}^{+\Delta x} u^2 \, \frac{1}{2\Delta x} \, \mathrm{d}u = 2 \int_0^{\Delta x} u^2 \, \frac{1}{2\Delta x} \, \mathrm{d}u = \frac{1}{\Delta x} \, \frac{u^3}{3} \Big|_0^{\Delta x} = \frac{\Delta x^2}{3}$$

$$\Rightarrow \sigma = \frac{1}{\sqrt{3}} \, \Delta x.$$

For large samples, we get

$$\Delta x_{\mathrm{rms}} \approx \frac{1}{\sqrt{3}} \, \Delta x.$$

2.5 Bunching Factor

Let us consider a beam signal $I_{\mathrm{beam}}(t)$ of a bunched beam as shown, for example, in Fig. 1.3 on p. 7. The **bunching factor** is defined as

$$\boxed{B_{\mathrm{f}} = \frac{\bar{I}_{\mathrm{beam}}}{I_{\mathrm{beam,max}}},} \tag{2.37}$$

i.e., it is the ratio of the average beam current to the maximum beam current (cf. Chao [9, Sect. 2.5.3.2, p. 131] or Reiser [10, Sect. 4.5.1, p. 263]). Obviously, the equation

$$\bar{I}_{\mathrm{beam}} = \frac{1}{T_{\mathrm{RF}}} \int_{-T_{\mathrm{RF}}/2}^{T_{\mathrm{RF}}/2} I_{\mathrm{beam}}(t) \, \mathrm{d}t$$

holds, where T_{RF} denotes the period.

Now one may replace the true shape of the beam current pulse by a rectangular one with the same maximum value. For $-T_{\mathrm{RF}}/2 < t < T_{\mathrm{RF}}/2$, we then have

$$I_{\mathrm{beam}}(t) = \begin{cases} I_{\mathrm{beam,max}} & \text{for } |t| \le \tau/2, \\ 0 & \text{elsewhere,} \end{cases}$$

where we have assumed that the bunch is centered at $t = 0$. In this case, one has to choose a pulse width τ in such a way that the same average beam current is obtained:

$$\bar{I}_{\mathrm{beam}} = \frac{\tau}{T_{\mathrm{RF}}}\, I_{\mathrm{beam,max}}.$$

Under these conditions, we obtain the expression

$$B_{\mathrm{f}} = \frac{\tau}{T_{\mathrm{RF}}}$$

for the bunching factor.

We now assume that the beam current pulse has the shape of a Gaussian distribution. This is, of course, possible only if the pulses are significantly shorter than the period time T_{RF}. Under this condition, the beam current will be close to zero before the next pulse starts.

Making use of Eq. (2.29), one may write $I_{\mathrm{beam}}(t)$ in the form

$$I_{\mathrm{beam}}(t) = K\, \frac{1}{\sigma\, \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2} \qquad \text{for} \qquad -T_{\mathrm{RF}}/2 < t < T_{\mathrm{RF}}/2.$$

We have

$$\int_{-\infty}^{+\infty} I_{\mathrm{beam}}(t)\, \mathrm{d}t = K.$$

The average beam current is obtained using the above-mentioned approximation:

$$\bar{I}_{\mathrm{beam}} = \frac{1}{T_{\mathrm{RF}}} \int_{-T_{\mathrm{RF}}/2}^{T_{\mathrm{RF}}/2} I_{\mathrm{beam}}(t)\, \mathrm{d}t \approx \frac{K}{T_{\mathrm{RF}}}.$$

For the maximum current, we obtain

$$I_{\mathrm{beam,max}} = \frac{K}{\sigma\, \sqrt{2\pi}},$$

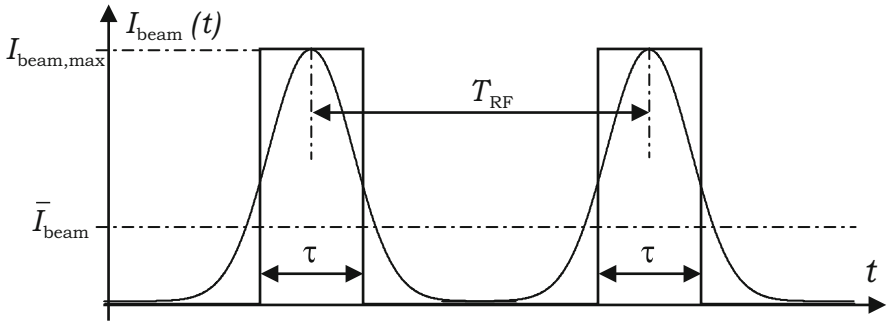

Fig. 2.3 Gaussian beam signal

so that the bunching factor

$$B_{\text{f}} \approx \frac{\sigma\ \sqrt{2\pi}}{T_{\text{RF}}}$$

is obtained. The equivalent length τ of a rectangular pulse is therefore

$$\tau = \sigma\ \sqrt{2\pi} \approx 2.5\ \sigma.$$

The two slopes of the rectangular pulse are therefore located at about $\pm 1.25\ \sigma$. This leads to the conversion between the Gaussian bunch and the rectangular signal that is visualized in Fig. 2.3.

2.6 Electromagnetic Fields

We summarize in this section a few basic formulas that may be found in standard textbooks (cf. [11–17]). We begin with **Maxwell's equations** in their integral form.

In the following, A denotes a two-dimensional domain, and V a three-dimensional domain. For a domain D (two- or threedimensional), ∂D denotes its boundary (with mathematically positive orientation, if applicable).

Maxwell's first equation (**Ampère's law**) in the time domain is

$$\boxed{\oint_{\partial A} \vec{H} \cdot \mathrm{d}\vec{r} = \int_A \left(\vec{J} + \dot{\vec{D}}\right) \cdot \mathrm{d}\vec{A},} \tag{2.38}$$

where $\vec{H}$ is the magnetizing field, $\vec{J}$ the current density, and $\vec{D}$ the electric displacement field.

Maxwell's second equation in the time domain (**Faraday's law**) reads

$$\boxed{\oint_{\partial A} \vec{E} \cdot \mathrm{d}\vec{r} = -\int_A \dot{\vec{B}} \cdot \mathrm{d}\vec{A}.} \tag{2.39}$$

Here $\vec{E}$ is the electric field, and $\vec{B}$ is the magnetic field.

Maxwell's third equation states that no magnetic charge exists:

$$\boxed{\oint_{\partial V} \vec{B} \cdot \mathrm{d}\vec{A} = 0.} \tag{2.40}$$

The electric charge Q inside a three-dimensional domain V is determined by Maxwell's fourth equation (**Gauss's law**):

$$\boxed{\oint_{\partial V} \vec{D} \cdot \mathrm{d}\vec{A} = \int_V \rho_q \, \mathrm{d}V = Q.} \tag{2.41}$$

Here, ρ_q denotes the charge density.

The current through a certain region A is given by

$$\boxed{I = \int_A \vec{J} \cdot \mathrm{d}\vec{A}},$$

and the voltage along a curve C is defined by

$$\boxed{V = \int_C \vec{E} \cdot \mathrm{d}\vec{r}.}$$

Please note that we use the same symbol for voltage and for threedimensional domains, but according to the context this should not lead to confusion.

In material bodies, the simplest relationships (linear isotropic media with relaxation times that are much smaller than the minimum time intervals of interest) between the field vectors are

$$\boxed{\vec{D} = \epsilon \vec{E}, \qquad \vec{B} = \mu \vec{H}, \qquad \vec{J} = \kappa \vec{E}.}$$

The material parameters are the permittivity ϵ, the permeability μ, and the conductivity κ. In vacuum, and approximately also in air, we have

$$\epsilon = \epsilon_0, \qquad \mu = \mu_0.$$

At least for fixed nonmoving domains A, we can write Eq. (2.39) in the form

$$\oint_{\partial A} \vec{E} \cdot \mathrm{d}\vec{r} = -\frac{\mathrm{d}\Phi_\mathrm{m}}{\mathrm{d}t},$$

where

$$\Phi_\mathrm{m} = \int_A \vec{B} \cdot \mathrm{d}\vec{A}$$

is the magnetic flux through the domain A. This form is suitable for induction problems.

Based on the integral form of Maxwell's equations presented above, one may derive their differential form if integral theorems are used:

$$\boxed{\begin{aligned} \mathrm{curl}\ \vec{H} &= \vec{J} + \dot{\vec{D}}, && (2.42)\\ \mathrm{curl}\ \vec{E} &= -\dot{\vec{B}}, && (2.43)\\ \mathrm{div}\ \vec{D} &= \rho_q, && (2.44)\\ \mathrm{div}\ \vec{B} &= 0. && (2.45) \end{aligned}}$$

Taking Eq. (2.44) into account, the divergence of Eq. (2.42) leads to the **continuity equation**

$$\boxed{\mathrm{div}\vec{J} + \dot{\rho_q} = 0.} \qquad (2.46)$$

We will discuss the physical meaning of this equation in Sect. 2.9.

In certain cases (here we assume that domains are filled homogeneously with linear isotropic material), Maxwell's equations may be solved by means of the **vector potential**[5] $\vec{A}$, defined by

$$\vec{B} = \mathrm{curl}\ \vec{A}, \qquad (2.47)$$

and the **scalar potential** Φ, defined by

$$\vec{E} = -\dot{\vec{A}} - \mathrm{grad}\ \Phi, \qquad (2.48)$$

both connected by the **Lorenz gauge condition**

$$\mathrm{div}\ \vec{A} = -\mu\epsilon\dot{\Phi}. \qquad (2.49)$$

[5]Please note that we have used the symbol $\mathrm{d}\vec{A}$ for the area element and A for two-dimensional domains. The absolute value of the vector potential would be $A = |\vec{A}|$ as well, but the meaning should be clear from the context.

Using these definitions, one obtains the **wave equations**

$$\boxed{\Delta\vec{A} - \frac{1}{c^2}\ddot{\vec{A}} = -\mu\vec{J},} \tag{2.50}$$

$$\boxed{\Delta\Phi - \frac{1}{c^2}\ddot{\Phi} = -\frac{\rho_q}{\epsilon}.} \tag{2.51}$$

Here

$$\boxed{c = \frac{1}{\sqrt{\mu\epsilon}}} \tag{2.52}$$

denotes the speed of light in the material under consideration. The speed of light in vacuum is

$$\boxed{c_0 = \frac{1}{\sqrt{\mu_0\epsilon_0}}.} \tag{2.53}$$

For static problems, there is no time dependence of the fields, and according to Maxwell's equations, electric and magnetic fields are therefore decoupled. In this case, the vector potential and the scalar potential also do not depend on time.

Equations (2.48) and (2.51) for homogeneous media thus reduce to

$$\vec{E} = -\text{grad}\,\Phi$$

and the **Poisson equation**

$$\boxed{\Delta\Phi = -\frac{\rho_q}{\epsilon},} \tag{2.54}$$

respectively. This equation has to be solved for electrostatic problems.

2.7 Special Relativity

The primary objective of this section is to introduce the nomenclature that is used in this book. This nomenclature is close to that of the introductory text [17] (in German). In any case, the reader should consult standard textbooks on special (and general) relativity (cf. [11, 13, 18–20] in English or [14, 21–28] in German) for an extensive introduction. However, the remainder of the book can also be understood if the formulas presented in this section are regarded as given.

The speed of light c_0 in vacuum has the same value in every **inertial frame**. Therefore, the equation of the wave front

$$x^2 + y^2 + z^2 = c_0^2 t^2 \tag{2.55}$$

in one inertial frame S (e.g., light flash at $t = 0$ at the origin of S) is transformed into a wave front equation

$$\bar{x}^2 + \bar{y}^2 + \bar{z}^2 = c_0^2 \bar{t}^2$$

that has the same form in a different inertial frame $\bar{S}$. Such transformation behavior is satisfied by the general **Lorentz transformation**. If one restricts generality in such a way that at $t = 0$, the origins of the two inertial frames are at the same position and that one frame $\bar{S}$ moves with constant velocity v in the z-direction relative to the other frame S, then one obtains the special Lorentz transformation

$$\bar{x} = x, \tag{2.56}$$

$$\bar{y} = y, \tag{2.57}$$

$$\bar{z} = \frac{z - vt}{\sqrt{1 - \frac{v^2}{c_0^2}}}, \tag{2.58}$$

$$\bar{t} = \frac{t - \frac{v}{c_0^2} z}{\sqrt{1 - \frac{v^2}{c_0^2}}}. \tag{2.59}$$

The inverse transformation can be generated if the quantities with a bar (e.g., $\bar{y}$) are replaced by the same quantities without the bar (e.g., y) and vice versa. In that case, $\bar{v} = -v$ has to be used (if $\bar{S}$ moves with respect to S with velocity v in the positive z direction, S will move with respect to $\bar{S}$ in the negative z direction), and c_0 remains the same. This concept for generating inverse transformation formulas may also be applied to electromagnetic field quantities, whose transformation behavior is discussed below.

The square root in the denominator of Eqs. (2.58) and (2.59) is typical of expressions in special relativity. Therefore, the so-called **Lorentz factors** are defined:

$$\boxed{\beta_v = \frac{v}{c_0},}$$

$$\boxed{\gamma_v = \frac{1}{\sqrt{1 - \beta_v^2}}.}$$

Special relativity may be built up by defining so-called **four-vectors** and **four-tensors**. For example, the space coordinates are combined with the time "coordinate" in order to define the components of a four-vector that specifies the position in **space-time**:

$$(\theta^i) = (x, y, z, c_0 t)^{\mathrm{T}} \quad \text{with} \quad i \in \{1, 2, 3, 4\}.$$

Specific values of this four-vector can be interpreted as **events**. In combination with the special choice (**signature**)

$$(g_{ik}) = (g^{ik}) = (\bar{g}_{ik}) = (\bar{g}^{ik}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

for the **metric tensor** ($i, k \in \{1, 2, 3, 4\}$), one obtains the desired transformation behavior of the wave front equation, because

$$\theta^i \theta_i = g_{ik} \theta^i \theta^k = 0,$$

which reproduces Eq. (2.55), is a tensor equation with a tensor of rank 0 (scalar) on the right-hand side. Here we use the **Ricci calculus** and **Einstein's summation convention**. The special Lorentz transformation given above can now be reproduced by

$$\bar{\theta}^i = \bar{a}^i_k \theta^k,$$

which corresponds to the matrix equation

$$(\bar{\theta}^i) = (\bar{a}^i_k) \cdot (\theta^i)$$

if the transformation coefficients

$$(\bar{a}^i_k) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \gamma_v & -\beta_v \gamma_v \\ 0 & 0 & -\beta_v \gamma_v & \gamma_v \end{pmatrix}$$

are chosen ($i =$ row, $k =$ column).

Similarly to the construction of the position four-vector, the vector potential and the scalar potential in electromagnetic field theory may be combined to form the electromagnetic **four-potential** $\mathcal{A}$ according to

$$(\mathcal{A}^i) = (A_x, A_y, A_z, \Phi/c_0)^{\mathrm{T}}.$$

This, for example, allows one to write the Lorenz gauge condition (2.49) for free space in the form

$$\mathcal{A}^i|_i = 0$$

of a tensor equation, where the vertical line indicates a **covariant derivative**, which—in special relativity—corresponds to the partial derivative because the metric coefficients are constant.

The **four-current density** $\mathcal{J}$ is defined by

$$(\mathcal{J}^i) = (J_x, J_y, J_z, \rho_q c_0)^{\mathrm{T}},$$

so that the tensor equation

$$\mathcal{J}^i|_i = 0$$

represents the continuity equation (2.46). The transformation law obviously yields

$$\bar{J}_x = J_x, \tag{2.60}$$

$$\bar{J}_y = J_y, \tag{2.61}$$

$$\bar{J}_z = \gamma_v \left(J_z - v\rho_q\right), \tag{2.62}$$

$$\bar{\rho}_q = \gamma_v \left(\rho_q - \frac{v}{c_0^2} J_z\right). \tag{2.63}$$

With $\vec{v} = v\vec{e}_z$ defining the parallel direction $\parallel$, this may be written in the generalized form

$$\vec{\bar{J}}_\perp = \vec{J}_\perp, \tag{2.64}$$

$$\vec{\bar{J}}_\parallel = \gamma_v \left(\vec{J}_\parallel - \vec{v}\rho_q\right), \tag{2.65}$$

$$\bar{\rho}_q = \gamma_v \left(\rho_q - \frac{\vec{v} \cdot \vec{J}}{c_0^2}\right). \tag{2.66}$$

The **electromagnetic field tensor** may be defined as

$$(\mathcal{B}^{ik}) = \begin{pmatrix} 0 & B_z & -B_y & -E_x/c_0 \\ -B_z & 0 & B_x & -E_y/c_0 \\ B_y & -B_x & 0 & -E_z/c_0 \\ E_x/c_0 & E_y/c_0 & E_z/c_0 & 0 \end{pmatrix},$$

while its counterpart for the other field components in Maxwell's equations may be defined as

$$(\mathcal{H}^{ik}) = \begin{pmatrix} 0 & H_z & -H_y & -c_0 D_x \\ -H_z & 0 & H_x & -c_0 D_y \\ H_y & -H_x & 0 & -c_0 D_z \\ c_0 D_x & c_0 D_y & c_0 D_z & 0 \end{pmatrix},$$

where i specifies the row, and k the column. The introduction of these four-vectors and four-tensors allows one to write Maxwell's equations as[6]

$$\mathcal{H}^{ik}|_i = -\mathcal{J}^k, \tag{2.67}$$

$$\mathcal{B}^{*ik}|_i = 0, \tag{2.68}$$

so that their form remains the same if a Lorentz transformation from one inertial frame to a different one is performed. This form invariance of physical laws is called **covariance**. The covariance of Maxwell's equations implies the constancy of c_0 in different inertial frames, since c_0 is a scalar quantity, a tensor of rank 0. Because $\mathcal{B}^{ik}$ and $\mathcal{H}^{ik}$ are tensors of rank 2, they are transformed according to the transformation rule

$$\bar{\mathcal{B}}^{ik} = \bar{a}^i_l \bar{a}^k_m \mathcal{B}^{lm}, \qquad \bar{\mathcal{H}}^{ik} = \bar{a}^i_l \bar{a}^k_m \mathcal{H}^{lm}.$$

Taking the second transformation rule as an example, this may be translated into the matrix equation

$$(\bar{\mathcal{H}}^{ik}) = (\bar{a}^i_k) \cdot (\mathcal{H}^{ik}) \cdot (\bar{a}^i_k)^{\mathrm{T}}.$$

A long but straightforward calculation then leads to the transformation laws for the corresponding field components:

$$\bar{H}_x = \gamma_v (H_x + v D_y), \tag{2.69}$$

$$\bar{H}_y = \gamma_v (H_y - v D_x), \tag{2.70}$$

$$\bar{H}_z = H_z, \tag{2.71}$$

$$\bar{D}_x = \gamma_v \left(D_x - \frac{\beta_v}{c_0} H_y \right), \tag{2.72}$$

$$\bar{D}_y = \gamma_v \left(D_y + \frac{\beta_v}{c_0} H_x \right), \tag{2.73}$$

$$\bar{D}_z = D_z. \tag{2.74}$$

[6]The asterisk is an operator that generates the dual tensor $\mathcal{B}^*$,

$$\mathcal{B}^{*ik} = \frac{1}{2} e^{iklm} \mathcal{B}_{lm},$$

of the tensor $\mathcal{B}$. Here e^{iklm} denotes the complete asymmetric tensor of rank 4.

The generalized form is

$$\vec{\bar{H}}_\perp = \gamma_v \left(\vec{H}_\perp - \vec{v} \times \vec{D}_\perp \right), \tag{2.75}$$

$$\vec{\bar{H}}_\parallel = \vec{H}_\parallel, \tag{2.76}$$

$$\vec{\bar{D}}_\perp = \gamma_v \left(\vec{D}_\perp + \frac{\vec{v} \times \vec{H}_\perp}{c_0^2} \right), \tag{2.77}$$

$$\vec{\bar{D}}_\parallel = \vec{D}_\parallel. \tag{2.78}$$

The remaining transformation laws are obtained analogously:

$$\bar{B}_x = \gamma_v \left(B_x + \frac{v}{c_0^2} E_y \right), \tag{2.79}$$

$$\bar{B}_y = \gamma_v \left(B_y - \frac{v}{c_0^2} E_x \right), \tag{2.80}$$

$$\bar{B}_z = B_z, \tag{2.81}$$

$$\bar{E}_x = \gamma_v (E_x - \beta_v c_0 B_y), \tag{2.82}$$

$$\bar{E}_y = \gamma_v (E_y + \beta_v c_0 B_x), \tag{2.83}$$

$$\bar{E}_z = E_z, \tag{2.84}$$

$$\vec{\bar{B}}_\perp = \gamma_v \left(\vec{B}_\perp - \frac{\vec{v} \times \vec{E}_\perp}{c_0^2} \right), \tag{2.85}$$

$$\vec{\bar{B}}_\parallel = \vec{B}_\parallel, \tag{2.86}$$

$$\vec{\bar{E}}_\perp = \gamma_v \left(\vec{E}_\perp + \vec{v} \times \vec{B}_\perp \right), \tag{2.87}$$

$$\vec{\bar{E}}_\parallel = \vec{E}_\parallel. \tag{2.88}$$

In the scope of this book, there is no need to develop the theory further. Nor do we discuss such standard effects as time dilation, Lorentz contraction, and the transformation of velocities. However, we need some relativistic formulas for mechanics.

The definition of the Lorentz force (1.1) is valid in special relativity—it corresponds to a covariant equation (the charge Q is invariant; it is a scalar quantity).

Also, the equation

$$\boxed{\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t}}$$

with the momentum definition

$$\boxed{\vec{p} = m\,\vec{u}}$$

based on the velocity

$$\boxed{\vec{u} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}}$$

still holds. However, the mass m is not invariant. Only the **rest mass** m_0 is a tensor of rank zero, i.e., a scalar:

$$\boxed{m = \frac{m_0}{\sqrt{1 - \frac{u^2}{c_0^2}}} = m_0 \gamma_u.}$$

Please note that we strictly distinguish between the velocities v and u and also between the related Lorentz factors. The velocity $\vec{v} = v\,\vec{e}_z$ is defined as the relative velocity of the inertial frame $\bar{S}$ with respect to the inertial frame S, i.e., the velocity between these two reference frames. The velocity $\vec{u}$ is the velocity of a particle measured in the first inertial frame S. Consequently, $\bar{\vec{u}}$ is the velocity of the same particle in the inertial frame $\bar{S}$. If it is clear what is meant by a certain Lorentz factor, one may, of course, omit the subscript.

The **total energy** of a particle with velocity $\vec{u}$ is given by

$$\boxed{W_{\mathrm{tot}} = m\,c_0^2 = \gamma_u m_0 c_0^2.} \tag{2.89}$$

Consequently, the **rest energy** is obtained for $\vec{u} = 0$, which leads to $\gamma_u = 1$:

$$\boxed{W_{\mathrm{rest}} = m_0 c_0^2.}$$

Therefore, the **kinetic energy** is

$$\boxed{W_{\mathrm{kin}} = W_{\mathrm{tot}} - W_{\mathrm{rest}} = m_0 c_0^2 (\gamma_u - 1).}$$

Using the Lorentz factors, one may write the momentum in the form

$$\vec{p} = m\vec{u} = m_0 c_0 \vec{\beta}_u \gamma_u,$$

leading to the absolute value

$$p = mu = m_0 c_0 \beta_u \gamma_u. \tag{2.90}$$

Here we used the definition

$$\vec{\beta}_u = \frac{\vec{u}}{c_0}. \tag{2.91}$$

If we have a look at Eqs. (2.89)–(2.91), we observe that γ is related to the energy, the product $\beta\gamma$ to the momentum, and β corresponds to the velocity. It is often helpful to keep this correspondence in mind when complicated expressions containing a large number of Lorentz factors are evaluated. One should also keep in mind that when one of the expressions β, γ, $\beta\gamma$ is known, the others are automatically fixed as well.

This is why we can also convert expressions for relative deviations into each other. For example, we may calculate the time derivative of

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \tag{2.92}$$

as follows:

$$\dot{\gamma} = -\frac{-2\beta\dot{\beta}}{2\,(1-\beta^2)^{3/2}} = \beta\gamma^3\dot{\beta}$$

$$\Rightarrow \frac{\dot{\gamma}}{\gamma} = \beta^2\gamma^2\frac{\dot{\beta}}{\beta}.$$

Here we can use the relation

$$\gamma^2 - \beta^2\gamma^2 = 1, \tag{2.93}$$

which follows directly from Eq. (2.92):

$$\frac{\dot{\gamma}}{\gamma} = (\gamma^2 - 1)\frac{\dot{\beta}}{\beta}.$$

Expressions of this type are very helpful, because they can be translated as follows:

$$\frac{\Delta W_{\text{tot}}}{W_{\text{tot}}} = (\gamma^2 - 1)\frac{\Delta u}{u}.$$

This conversion is possible if the relative change in the quantities is sufficiently small. In the example presented here, one can see directly that a velocity deviation of 1% is transformed into an energy deviation of 3% if $\gamma = 2$ holds.

As a second example, we can calculate the time derivative of Eq. (2.93):

$$2\gamma\dot{\gamma} - 2(\beta\gamma)\frac{\mathrm{d}(\beta\gamma)}{\mathrm{d}t} = 0$$

$$\Rightarrow \frac{\dot{\gamma}}{\gamma} = \beta^2\frac{1}{(\beta\gamma)}\frac{\mathrm{d}(\beta\gamma)}{\mathrm{d}t}.$$

Table 2.3 Conversion of relative deviations

$\frac{\Delta\beta}{\beta}$	$= \frac{\Delta\beta}{\beta}$	$= \frac{1}{\gamma^2-1}\frac{\Delta\gamma}{\gamma} = \frac{1}{(\beta\gamma)^2}\frac{\Delta\gamma}{\gamma}$	$= \gamma^{-2}\frac{\Delta(\beta\gamma)}{\beta\gamma}$	$= \frac{1}{\gamma(\gamma+1)}\frac{\Delta\gamma}{\gamma-1}$
$\frac{\Delta\gamma}{\gamma}$	$= (\gamma^2-1)\frac{\Delta\beta}{\beta} = (\beta\gamma)^2\frac{\Delta\beta}{\beta}$	$= \frac{\Delta\gamma}{\gamma}$	$= \beta^2\frac{\Delta(\beta\gamma)}{\beta\gamma}$	$\left(1-\gamma^{-1}\right)\frac{\Delta\gamma}{\gamma-1}$
$\frac{\Delta(\beta\gamma)}{\beta\gamma}$	$= \gamma^2\frac{\Delta\beta}{\beta}$	$= \beta^{-2}\frac{\Delta\gamma}{\gamma}$	$= \frac{\Delta(\beta\gamma)}{\beta\gamma}$	$= \frac{\gamma}{\gamma+1}\frac{\Delta\gamma}{\gamma-1}$
$\frac{\Delta\gamma}{\gamma-1}$	$= \gamma(\gamma+1)\frac{\Delta\beta}{\beta}$	$= \frac{\gamma}{\gamma-1}\frac{\Delta\gamma}{\gamma}$	$= \left(1+\gamma^{-1}\right)\frac{\Delta(\beta\gamma)}{\beta\gamma}$	$= \frac{\Delta\gamma}{\gamma-1}$

Note: $\frac{\Delta W_{\text{tot}}}{W_{\text{tot}}} = \frac{\Delta\gamma}{\gamma}$, $\frac{\Delta p}{p} = \frac{\Delta(\beta\gamma)}{\beta\gamma}$, $\frac{\Delta W_{\text{kin}}}{W_{\text{kin}}} = \frac{\Delta\gamma}{\gamma-1}$

This can be translated into

$$\frac{\Delta W_{\text{tot}}}{W_{\text{tot}}} = \beta^2\frac{\Delta p}{p}.$$

Relations like these are summarized in Table 2.3.

In accelerator physics and engineering, specific units that contain the elementary charge e are often used to specify the energy of the beam. This is due to the fact that the energy that is gained by a charge[7] $Q = z_q e$ is given by formula (1.2),

$$\Delta W = QV = z_q eV.$$

An electron that passes a voltage of $V = 1\,\text{kV}$ will therefore lose or gain an energy of 1 keV, depending on the orientation of the voltage. We have only to insert the quantities into the formula without converting e into SI units. In order to convert an energy that is given in eV into SI units, one simply has to insert $e = 1.6022{\cdot}10^{-19}\,\text{C}$, so that $1\,\text{eV} = 1.6022\cdot10^{-19}\,\text{J}$ holds.

Also, the rest energy of particles is often specified in eV. For example, the electron rest mass $m_\text{e} = 9.1094{\cdot}10^{-31}\,\text{kg}$ corresponds to an energy of 510.999 keV.

As we saw above, the energy directly determines the Lorentz factors and the velocity. Therefore, it is desirable to specify the energy in a unit that directly corresponds to a certain velocity. Due to

$$W_{\text{kin}} = mc_0^2 - m_0c_0^2 = m_0c_0^2(\gamma-1),$$

a kinetic energy of 1 MeV leads to different values for γ if different particle rest masses m_0 are considered. This is why one introduces another energy unit for ions. An ion with mass number A has rest mass

$$m_0 = A_\text{r}m_\text{u},$$

[7] For ions, z_q is the **charge number**. For electrons, one has to set $z_q = -1$, while for protons and positrons, $z_q = 1$ has to be defined.

where $m_\mathrm{u} = 1.66054 \cdot 10^{-27}\,\mathrm{kg}$ denotes the **unified atomic mass unit** (as mentioned below, A_r differs slightly from A). Therefore, one obtains

$$W_\mathrm{kin,u} = \frac{W_\mathrm{kin}}{A_\mathrm{r}} = m_\mathrm{u} c_0^2 (\gamma - 1).$$

If the value on the right-hand side is specified now, γ is determined in a unique way, since m_u and c_0 are global constants. As an example, an ion beam with a kinetic energy of[8] 11.4 MeV/u corresponds to $\gamma = 1.0122386$ and $\beta = 0.15503$. We do not need to specify the ion species.

Ions are usually specified by the notation

$$^{\mathrm{A}}_{\mathrm{Z}}\mathrm{Element}^{z_\mathrm{q}+}.$$

Here A is the (integer) **mass number**, i.e., the number of **nucleons** (protons plus neutrons); Z is the **atomic number**, which equals the number of protons and identifies the element. For example,

$$^{238}_{92}\mathrm{U}^{28+}$$

indicates a uranium ion that has $A - Z = 146$ neutrons. Different uranium **isotopes**[9] exist with a different number of neutrons. The number of protons however, is the same for all these isotopes. Therefore, Z is redundant information that is already included in the element name. In the last example, the uranium atom has obviously lost 28 of its 92 electrons, leading to the charge number $z_q = 28$.

The unified atomic mass unit m_u is defined as $1/12$ of the mass of the atomic nucleus $^{12}_{6}\mathrm{C}$. For different ion species and isotopes, the mass is not exactly an integer multiple of m_u (reasons: different mass of protons and neutrons, relativistic mass defect due to binding energy). For $^{238}\mathrm{U}$, for example, one has $A_\mathrm{r} = 238.050786$, which approximately equals $A = 238$.

2.8 Nonlinear Dynamics

A continuous **dynamical system** of first order may be described by the following first-order **ordinary differential equation (ODE)**:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v(x, t).$$

[8] read: MeV per nucleon.

[9] A **nuclide** is specified by the number of protons Z and the number of neutrons $A - Z$. The chemical element is determined by the number of protons only. Different nuclides belonging to the same chemical element are called isotopes (of that element).

The state of a dynamical system of order n is represented by the values of n variables $x_1, x_2, \ldots, x_n$, which may be combined into a vector $\vec{r} = (x_1, x_2, \ldots, x_n)$. Hence, a dynamical system of order n is described by the system of ordinary differential equations

$$\boxed{\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}, t).} \tag{2.94}$$

One should note that the system of ODEs is still of order 1, but of dimension n. Such a system is called **autonomous** when $\vec{v}(\vec{r}, t)$ does not depend on the time[10] t, i.e., when

$$\vec{v}(\vec{r}, t) = \vec{v}(\vec{r})$$

holds. The next sections will show that Eq. (2.94), which may look very simple at first sight, includes a huge variety of problems.

2.8.1 *Equivalence of Differential Equations and Systems of Differential Equations*

Let us consider the nth-order linear ordinary differential equation

$$\frac{\mathrm{d}^n y}{\mathrm{d}t^n} + a_{n-1}(t)\frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}} + \cdots + a_1(t)\frac{\mathrm{d}y}{\mathrm{d}t} + a_0(t)y(t) = b(t)$$

with dimension 1. One sees that by means of the definitions

$$\begin{aligned} x_1 &= y, \\ x_2 &= \frac{\mathrm{d}y}{\mathrm{d}t}, \\ &\cdots \\ x_n &= \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}}, \end{aligned}$$

it may be converted into the form

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= x_3, \\ &\cdots \\ \dot{x}_n &= b(t) - a_0(t)x_1 - a_1(t)x_2 - \cdots - a_{n-1}(t)x_n, \end{aligned}$$

[10] The variable t is not necessarily the time, but we will often call it the time in order to make the interpretation easier.

which is equivalent to the standard form

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}, t).$$

If b and all the a_k do not depend on time (ODE of order n with constant coefficients), then $\vec{v}$ will also not depend on time explicitly, so that an autonomous system is present.

Although the vector field $\vec{v}$ is called a velocity function, it does not always correspond to a physical velocity. As already mentioned, the variable t is not necessarily the physical time. However, we will use this notation because the reader may always interpret these variables in terms of the mechanical analogy, which may help to understand the physical background.

The above-mentioned equivalence is also valid for nonlinear ODEs of the form

$$\frac{\mathrm{d}^n y}{\mathrm{d}t^n} = F\left(t, y, \frac{\mathrm{d}y}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}}\right),$$

where[11] $F \in C^1$. Also here, we may use

$$\begin{aligned} x_1 &= y, \\ x_2 &= \frac{\mathrm{d}y}{\mathrm{d}t}, \\ &\ldots \\ x_n &= \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}}, \end{aligned}$$

to obtain the standard form

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= x_3, \\ &\ldots \\ \dot{x}_n &= F(t, x_1, x_2, \ldots, x_n) \\ &\Leftrightarrow \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}, t). \end{aligned}$$

---

[11] In this book, C^k denotes the class of functions that are k-times **continuously differentiable**: C^0 is the class of **continuous** functions, C^1 the class of continuously differentiable functions.

An autonomous system results when F and $\vec{v}$ do not explicitly depend on the time variable t.

2.8.2 Autonomous Systems

Hereinafter, we will consider only autonomous systems if a time dependence is not stated explicitly.

2.8.2.1 Time Shift

An advantage of autonomous systems is the fact that if a solution $y(t)$ of

$$\frac{\mathrm{d}^n y}{\mathrm{d}t^n} = F\left(y, \frac{\mathrm{d}y}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}}\right)$$

is known, then $z(t) = y(t - T)$ will also be a solution if T is a constant time shift. This can be shown as follows:

The solution $y(t)$ is the first component of the vector $\vec{r}(t)$ that satisfies the differential equation[12]

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}).$$

Therefore, $z(t)$ is the first component of the vector

$$\vec{r}_{\text{shift}}(t) = \vec{r}(t - T).$$

We obtain

$$\frac{\mathrm{d}\vec{r}_{\text{shift}}}{\mathrm{d}t} = \left.\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right|_{t-T} = \vec{v}(\vec{r})\big|_{t-T} = \vec{v}(\vec{r}_{\text{shift}}).$$

One sees that $\vec{r}_{\text{shift}}(t)$ satisfies the system of ODEs in the same way as $\vec{r}(t)$ does. Due to the equivalence with the differential equation of order n, $z(t)$ will be a solution as well.

This explains, for instance, why $\sin(\omega t)$ must be a solution of the homogeneous differential equation

$$\ddot{y} + \omega^2 y = 0$$

---

[12] In the text, we will not explicitly distinguish between systems of ordinary differential equations and ordinary differential equations, since this difference should be obvious based on the notation.

if one knows that $\cos(\omega t)$ is a solution. This ODE is autonomous, and these two solutions differ only by a time shift.

2.8.2.2 Phase Space

The **phase space** may be defined as the continuous space of all possible states of a dynamical system. In our case, the dynamical system is described by an autonomous system of ordinary differential equations.

The graphs of the solutions $\vec{r}(t)$ of the differential equation are the **integral curves** or **solution curves** in the n-dimensional phase space. Such an integral curve contains the dependence on the parameter t (which is usually but not necessarily the time). A different parameterization therefore leads to a different integral curve.

The set of all image points of the map $t \mapsto \vec{r}(t)$ is called the **orbit**. An orbit does not contain dependence on the parameter t. A different parameterization therefore leads to the same orbit, since the same image points are obtained simply by a different value of the parameter t.

Different orbits of an autonomous system are often drawn in a **phase portrait**, which may be defined as the set of all orbits.

2.8.3 Existence and Uniqueness of the Solution of Initial Value Problems

The standard form

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r})$$

has the advantage that it can be solved numerically according to the (explicit) **Euler method**:

$$\vec{r}_{k+1} = \vec{r}_k + \Delta t \cdot \vec{v}(\vec{r}_k)$$
$$t_k = t_0 + k\ \Delta t.$$

It is obvious that by defining the initial condition

$$\vec{r}_0 = \vec{r}(t_0),$$

the states

$$\vec{r}_k \approx \vec{r}(t_k) = \vec{r}(t_0 + k\Delta t)$$

of the system at different times can be derived iteratively for $k > 0$ ($k \in \mathbb{N}$). The states of the system may be calculated for both future times $t > t_0$ and past times $t < t_0$ in a unique way by selecting the sign of Δt. However, this is possible only in a certain neighborhood around t_0, as we will see in the next sections.

It is obvious that by defining $\vec{r}_0$, n scalar initial conditions are required to make the solution unique.

2.8.3.1 Existence of a Local Solution

The existence of a solution is ensured by the following theorem:

Theorem 2.1 (Peano). *Consider an initial value problem*

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}, t) \qquad \vec{r}(t_0) = \vec{r}_0$$

with a continuous $\vec{v} : D \to \mathbb{R}^n$ *on an open set* $D \subset \mathbb{R}^{n+1}$. *Then there exists* $\alpha(\vec{r}_0, t_0) > 0$ *such that the initial value problem has at least one solution in the interval* $[t_0 - \alpha, t_0 + \alpha]$.

(See Aulbach [29, Theorem 2.2.3].)

Remark. We may easily see that $\vec{v}$ must be continuous. If we choose $v = \Theta(t)$ (Heaviside step function) in the one-dimensional case, we immediately see that the derivative $\frac{\mathrm{d}r}{\mathrm{d}t}$ is not defined at $t = 0$. Therefore, in the scope of classical analysis, we have to exclude functions that are not continuous. In the scope of distribution theory, the solution $r = t\,\Theta(t)$ is obvious.

2.8.3.2 Uniqueness of a Local Solution

Uniqueness can be ensured if the vector field $\vec{v}$ satisfies a **Lipschitz condition** or if it is continuously differentiable.

Definition 2.2. The vector function $\vec{v}(\vec{r}, t) : D \to \mathbb{R}^n$ ($D \subset \mathbb{R}^{n+1}$ open) is said to satisfy a global Lipschitz condition on D with respect to $\vec{r}$ if there is a constant $K > 0$ such that for all $(\vec{r}_1, t), (\vec{r}_2, t) \in D$, the condition

$$\left\|\vec{v}(\vec{r}_1, t) - \vec{v}(\vec{r}_2, t)\right\| \leq K \left\|\vec{r}_1 - \vec{r}_2\right\|$$

holds. Instead of saying that a function satisfies a global Lipschitz condition, one also speaks of a function that is Lipschitz continuous.

(Cf. Aulbach [29, Definition 2.3.5] and Perko [30, p. 71, Definition 2].)

Definition 2.3. The vector function $\vec{v}(\vec{r}, t) : D \to \mathbb{R}^n$ ($D \subset \mathbb{R}^{n+1}$ open) is said to satisfy a local Lipschitz condition on D with respect to $\vec{r}$ if for each $(\vec{r}_0, t_0) \in D$,

there exist a neighborhood $U_{(\vec{r}_0,t_0)} \subset D$ of $(\vec{r}_0, t_0)$ and a constant $K > 0$ such that for all $(\vec{r}_1, t), (\vec{r}_2, t) \in U_{(\vec{r}_0,t_0)}$, the condition

$$\left\|\vec{v}(\vec{r}_1, t) - \vec{v}(\vec{r}_2, t)\right\| \leq K \left\|\vec{r}_1 - \vec{r}_2\right\|$$

holds. Instead of saying that a function satisfies a local Lipschitz condition, one also speaks of a function that is locally Lipschitz continuous.

(Cf. Aulbach [29, Definition 2.3.5], Wirsching [31, Definition 3.4], and Perko [30, p. 71, Definition 2].)

In other words, the function satisfies a local Lipschitz condition if for every point, we can find a neighborhood such that a "global" Lipschitz condition holds in that neighborhood.

Example. The function $f(x) = x^2$ is locally Lipschitz continuous, but it is not Lipschitz continuous.

Theorem 2.4 (Picard–Lindelöf). *Consider the initial value problem*

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}, t) \qquad \vec{r}(t_0) = \vec{r}_0$$

with continuous $\vec{v} : D \to \mathbb{R}^n$ *(*$D \subset \mathbb{R}^{n+1}$ *open). Suppose that the vector function* $\vec{v}(\vec{r}, t)$ *is locally Lipschitz continuous with respect to* $\vec{r}$*. Then there exists* $\alpha(\vec{r}_0, t_0) > 0$ *such that the initial value problem has a unique solution in the interval* $[t_0 - \alpha, t_0 + \alpha]$.

(See Aulbach [29, Theorem 2.3.7].)

Every locally Lipschitz continuous function is also continuous.

Every continuously differentiable function satisfies a local Lipschitz condition, i.e., is locally Lipschitz continuous (Aulbach [29, p. 77], Arnold [32, p. 279], Perko [30, lemma on p. 71]).

Therefore, the Picard–Lindelöf theorem may simply be rewritten for continuously differentiable functions instead of locally Lipschitz continuous functions (Perko [30, p. 74]: "The Fundamental Existence-Uniqueness Theorem," Guckenheimer/Holmes [33, Theorem 1.0.1]).

2.8.3.3 Maximal Interval of Existence

One may try to make the solution interval larger by using the endpoint of the solution interval as a new initial condition. If this strategy is executed iteratively, one obtains the **maximal interval of existence**. It is an open interval (cf. [30, p. 89, Theorem 1]). The maximal interval of existence does not necessarily correspond to the full real time axis. Further requirements are necessary to ensure this.

2.8.3.4 Global Solution

A continuously differentiable vector field $\vec{v}$ is called **complete** if it induces a **global flow**,[13] i.e., if its integral curves are defined for all $t \in \mathbb{R}$.

Every differentiable vector field with compact support is complete.

The following theorem shows that certain restrictions on the "velocity" $\vec{v}(\vec{r})$ are sufficient for completeness:

Theorem 2.5. *Let the vector function $\vec{v}(\vec{r})$ with $\vec{v} : D \to \mathbb{R}^n$ ($D \subset \mathbb{R}^n$ open) be continuously differentiable and linearly bounded with $K, L \geq 0$:*

$$\left\|\vec{v}(\vec{r})\right\| \leq K \left\|\vec{r}\right\| + L.$$

Then the initial value problem

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}) \qquad \vec{r}(t_0) = \vec{r}_0$$

has a global flow.

(Cf. Zehnder [34, Proposition IV.3, p. 130], special form of Theorem 2.5.6, Aulbach [29].)

According to Amann [35, Theorem 7.8], the solution will then be bounded for finite time intervals.

Like many other authors, Perko [30, p. 188, Theorem 3] requires that $\vec{v}(\vec{r})$ satisfy a global Lipschitz condition

$$\left\|\vec{v}(\vec{r}_1) - \vec{v}(\vec{r}_2)\right\| \leq K \left\|\vec{r}_1 - \vec{r}_2\right\|$$

for arbitrary $\vec{r}_1, \vec{r}_2 \in \mathbb{R}^n$. For $\vec{r}_2 = 0$, this leads to linear boundedness, as one may show by means of the reverse triangle inequality, but it is a stronger condition.

Example. The ODE

$$\dot{y} = 1 + y^2$$

is obviously satisfied for

$$y = \tan\, t = \frac{\sin t}{\cos t} \qquad \dot{y} = \frac{\cos^2 t + \sin^2 t}{\cos^2 t} = 1 + \tan^2 t.$$

This solution may be found by separation of variables. An arbitrary initial condition $y(0) = y_0$ may be satisfied if the shifted solution

$$y = \tan(t + \tau)$$

[13] The exact definition of a flow will be given in Sect. 2.8.6 on p. 61.

is considered. In any case, however, the solution curve reaches infinity while t is still finite. The "vector" field $v(y) = 1 + y^2$ is not complete, and it is obviously not linearly bounded.

If we simplify the results of this section, we may summarize them as follows:

- The existence of a local solution is ensured by continuity of $\vec{v}$.
- Local Lipschitz continuity ensures uniqueness of the solution. If $\vec{v}$ is continuously differentiable, uniqueness is also guaranteed.
- If linear boundedness of $\vec{v}$ is required in addition, a global solution/global flow exists.

For the sake of simplicity, we will consider only complete vector fields in the following.

2.8.3.5 Linear Systems of Ordinary Differential Equations

For linear systems of differential equations with

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = A \cdot \vec{r}, \qquad \vec{r}(t_0) = \vec{r}_0,$$

where A is a quadratic matrix with real constant elements, we may use the **matrix norm**:

$$\left\|\vec{v}(\vec{r})\right\| = \left\|A \cdot \vec{r}\right\| \leq \|A\| \left\|\vec{r}\right\| = K \left\|\vec{r}\right\|.$$

Therefore, the conditions of Theorem 2.5 are satisfied, and a unique solution with a global flow exists. One may specifically use the **Frobenius norm**

$$\|A\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{n}\sum_{k=1}^{n} |a_{ik}|^2},$$

which is compatible with the **Euclidean norm**

$$\left\|\vec{r}\right\| = \sqrt{\sum_{i=1}^{n} |r_i|^2}$$

of a vector, so that

$$\left\|A \cdot \vec{r}\right\| \leq \|A\|_{\mathrm{F}} \left\|\vec{r}\right\|$$

holds.

2.8.4 Orbits

Two distinct orbits of an autonomous system do not intersect. In order to prove this, we assume the contrary. Suppose that two distinct orbits defined by $\vec{r}_1(t)$ and $\vec{r}_2(t)$ intersect according to

$$\vec{r}_1(t_1) = \vec{r}_2(t_2).$$

Please note that the intersection point may be reached for different values t_1 and t_2 of the parameter t, since we require only that the orbits (i.e., the images of the solution curves) intersect. As shown in Sect. 2.8.2.1,

$$\vec{r}_{\text{shift}}(t) = \vec{r}_1(t + t_1 - t_2)$$

is also a solution of the differential equation. Therefore, we have

$$\vec{r}_{\text{shift}}(t_2) = \vec{r}_1(t_1) = \vec{r}_2(t_2).$$

Hence $\vec{r}_{\text{shift}}(t)$ and $\vec{r}_2(t)$ satisfy the same initial conditions at the time t_2. This means that the solution curves $\vec{r}_{\text{shift}}(t)$ and $\vec{r}_2(t)$ are identical.

Since $\vec{r}_1(t)$ is simply time-shifted with respect to $\vec{r}_{\text{shift}}(t) = \vec{r}_2(t)$, the images, i.e., the orbits, will be identical. This means that two orbits are completely equal if they have one point in common.

In other words, each point of phase space is crossed by only one orbit.

2.8.5 Fixed Points and Stability

Vectors $\vec{r} = \vec{r}_{\text{F}}$ for which

$$\vec{v}(\vec{r}) = 0$$

holds are called **fixed points** (or **equilibrium points** or **stationary points** or **critical points**) of the dynamical system given by

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}).$$

This nomenclature is obvious, since a particle that is initially located at

$$\vec{r}(t_0) = \vec{r}_{\text{F}}$$

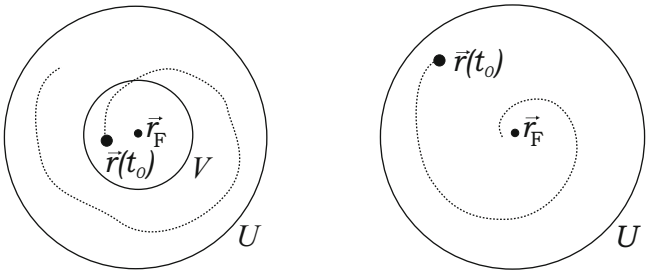

Fig. 2.4 Stability (*left*) and asymptotic stability (*right*) of a fixed point

will stay there forever:

$$\vec{r}(t) = \vec{r}_F \qquad \text{for } t > t_0.$$

Definition 2.6. A fixed point of an autonomous dynamical system is called an **isolated fixed point** or a **nondegenerate fixed point** if an environment of the fixed point exists that does not contain any other fixed points.

(Cf. Sastry [36, Definition 1.4, p. 13], Perko [30, Definition 2, p. 173].)

We now define the stability of fixed points according to Lyapunov.

Definition 2.7. A fixed point is called **stable** if for every neighborhood U of $\vec{r}_F$, another neighborhood $V \subset U$ of $\vec{r}_F$ exists such that a trajectory starting in V at $t = t_0$ will remain in U for all $t \geq t_0$ (see Fig. 2.4). Otherwise, the fixed point is called **unstable**.

Please note that it is usually necessary to choose V smaller than U, because the shape of the orbit may cause the trajectory to leave U for some starting points in U even if $\vec{r}_F$ is stable.

Definition 2.8. A stable fixed point $\vec{r}_F$ is called **asymptotically stable** if a neighborhood U of $\vec{r}_F$ exists such that for every trajectory that starts at $t = t_0$ in U, the following equation holds:

$$\lim_{t \to \infty} \vec{r}(t) = \vec{r}_F.$$

(See, e.g., Perko [30, Definition 1, p. 129].)

Definition 2.9. A function $L(\vec{r})$ with $L \in C^1$ and $L : U \to \mathbb{R}$ ($U \subset \mathbb{R}^n$ open) is called a **Lyapunov function** for the fixed point $\vec{r}_F$ of the autonomous system

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}) \qquad \vec{v} \in C^1(D) \qquad D \subset \mathbb{R}^n \text{open}$$

if

$$L(\vec{r}_\mathrm{F}) = 0$$

and

$$L(\vec{r}) > 0 \qquad \text{for} \quad \vec{r} \in U \backslash \{\vec{r}_\mathrm{F}\},$$
$$\dot{L} = \vec{v} \cdot \operatorname{grad} L \leq 0 \qquad \text{for} \quad \vec{r} \in U \backslash \{\vec{r}_\mathrm{F}, \}$$

hold in a neighborhood $U \subset D$ of $\vec{r}_\mathrm{F}$.

A Lyapunov function is called a **strict Lyapunov function** if

$$\dot{L} = \vec{v} \cdot \operatorname{grad} L < 0 \qquad \text{for} \quad \vec{r} \in U \backslash \{\vec{r}_\mathrm{F}\}$$

holds.

(Cf. Perko [30, p. 131, Theorem 3], La Salle/Lefschetz [37, Sect. 8], Guckenheimer/Homes [33, Theorem 1.0.2].)

Theorem 2.10. *If a Lyapunov function for a fixed point $\vec{r}_\mathrm{F}$ of an autonomous system exists, then this fixed point $\vec{r}_\mathrm{F}$ is stable. If a strict Lyapunov function exists, then this fixed point $\vec{r}_\mathrm{F}$ is asymptotically stable.*

(Cf. Perko [30, p. 131, Theorem 3].)

It is easy to see that this theorem is valid. For two-dimensional systems with the particle trajectory $\vec{r}(t) = x(t)\, \vec{e}_x + y(t)\, \vec{e}_y$, we obtain, for example,[14]

$$\dot{L} = \frac{\mathrm{d}L}{\mathrm{d}t} = \frac{\partial L}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial L}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} = \vec{v} \cdot \operatorname{grad} L.$$

If this expression is negative, the strict Lyapunov function will decrease while the particle continues on its path. Since the minimum of the Lyapunov function is obtained for $\vec{r}_\mathrm{F}$, it is clear that the particle will move toward the fixed point.

Similar reasoning applies for a Lyapunov function that is not strict. In this case, the particle cannot move away from the fixed point, because the Lyapunov function does not increase. However, it will not necessarily get closer to the fixed point.

[14] According to the definition, the Lyapunov function depends on the location (x, y). The chain rule that is applied here implies that we are following the trajectory $\vec{r}(t)$, so that $\dot{L}$ depends only on the time t.

The gradient of L points in the direction in which L increases. Therefore, the scalar product $\vec{v} \cdot \operatorname{grad} L$ becomes negative if $\vec{v}$ has a component in the opposite direction (decreasing L).

2.8.6 *Flows of Linear Autonomous Systems*

Having shown above that a linear autonomous system possesses a global flow, we shall now compute this flow. If an autonomous system of order n is linear, we may describe it by

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r})$$

with

$$\vec{v}(\vec{r}) = A \cdot \vec{r},$$

where A is a quadratic $n \times n$ matrix with real constant elements. The ansatz

$$\vec{r} = \vec{w}\, e^{\lambda t}$$

leads to

$$\lambda \vec{w} = A \cdot \vec{w},$$

or

$$(A - \lambda I) \cdot \vec{w} = 0.$$

For nontrivial solutions $\vec{w} \neq 0$, the condition

$$\boxed{\det(A - \lambda I) = 0}$$

is necessary, which determines the **eigenvalues** λ. For the sake of simplicity, we now assume that all n eigenvalues are distinct and that there is one **eigenvector** belonging to each eigenvalue (A is diagonalizable in this case). The overall solution of the homogeneous system of ODEs may then be written in the form

$$\vec{r}(t) = \sum_{k=1}^{n} C_k\ \vec{w}_k e^{\lambda_k t}, \tag{2.95}$$

where $\vec{w}_k$ denotes the eigenvector that belongs to the eigenvalue $\lambda = \lambda_k$ and where the C_k are constants. For the initial condition at $t = 0$, we therefore have

$$\vec{r}_0 = \vec{r}(0) = \sum_{k=1}^{n} C_k\ \vec{w}_k.$$

According to Eq. (2.95), the solution is obviously asymptotically stable if and only if

$$\boxed{\mathrm{Re}\{\lambda_k\} < 0} \tag{2.96}$$

holds for all $k \in \{1, 2, \ldots, n\}$, since only then does

$$\lim_{t \to \infty} \vec{r}(t) = 0$$

hold for arbitrary initial conditions. In this case, $\vec{r} = 0$ is an asymptotically stable fixed point.

Now we raise the question whether further fixed points exist. This is the case for

$$\vec{v}(\vec{r}) = A \cdot \vec{r} = 0$$

with

$$\vec{r} \neq 0,$$

i.e., only for

$$\det\ A = 0.$$

In Sect. 2.8.8, we will see that this is the condition for a degenerate, i.e., nonisolated, fixed point (see Definition 2.6, p. 59).

Let us now determine a map that transforms an initial value $\vec{r}_0$ into a vector $\vec{r}(t)$ that satisfies the linear autonomous system of ODEs

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}) = A \cdot \vec{r}.$$

The overall solution

$$\vec{r}(t) = \sum_{k=1}^{n} C_k\ \vec{w}_k e^{\lambda_k t}$$

with eigenvectors

$$\vec{w}_k = \begin{pmatrix} w_{k1} \\ w_{k2} \\ \cdots \\ w_{kn} \end{pmatrix}$$

may, due to

$$x_i(t) = \sum_{k=1}^{n} C_k \; w_{ki} e^{\lambda_k t},$$

be written as the matrix equation

$$\vec{r}(t) = \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{n1} \\ w_{12} & w_{22} & \cdots & w_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ w_{1n} & w_{2n} & \cdots & w_{nn} \end{pmatrix} \cdot \begin{pmatrix} C_1 e^{\lambda_1 t} \\ C_2 e^{\lambda_2 t} \\ \cdots \\ C_n e^{\lambda_n t} \end{pmatrix}.$$

We define

$$X_A(0) = \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{n1} \\ w_{12} & w_{22} & \cdots & w_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ w_{1n} & w_{2n} & \cdots & w_{nn} \end{pmatrix}$$

(matrix of the eigenvectors),

$$\vec{c} = \begin{pmatrix} C_1 \\ C_2 \\ \cdots \\ C_n \end{pmatrix},$$

and

$$D_A^{\lambda}(t) = \begin{pmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{pmatrix}.$$

Hence we have

$$\vec{r}(t) = X_A(0) \cdot D_A^{\lambda}(t) \cdot \vec{c}.$$

For $t = 0$, we obtain

$$\vec{r}(0) = X_A(0) \cdot D_A^{\lambda}(0) \cdot \vec{c} = X_A(0) \cdot I \cdot \vec{c} = X_A(0) \cdot \vec{c}.$$

Due to

$$D_A^{\lambda}(0) = I,$$

the definition

$$X_A(t) = X_A(0) \cdot D_A^{\lambda}(t) \tag{2.97}$$

makes sense. Finally, we obtain

$$\vec{r}(t) = X_A(t) X_A(0)^{-1} \vec{r}(0).$$

Using the **matrix exponential function**

$$e^{tA} = X_A(t) X_A(0)^{-1}, \tag{2.98}$$

one also writes

$$\vec{r}(t) = e^{tA}\, \vec{r}(0). \tag{2.99}$$

This equation obviously determines the global flow (Guckenheimer [33, Eqn. (1.1.9), p. 9]) if the following definition is used:

Definition 2.11. A (global) **flow** is a continuous map $\Phi : \mathbb{R} \times D \to D$, which transforms each initial value $\vec{r}(0) = \vec{r}_0 \in D$ ($D \subset \mathbb{R}^n$ open) into a vector $\vec{r}(t)$ ($t \in \mathbb{R}$) satisfying the following conditions:

$$\Phi_0 = \mathrm{id}, \quad \text{i.e.,} \quad \Phi_0(\vec{r}_0) = \vec{r}_0 \quad \text{for all } \vec{r}_0 \in D,$$

$$\Phi_{t_1+t_2} = \Phi_{t_1} \circ \Phi_{t_2}, \quad \text{i.e.} \quad \Phi_{t_1+t_2}(\vec{r}_0) = \Phi_{t_1}(\Phi_{t_2}(\vec{r}_0)) \quad \text{for all } \vec{r}_0 \in D, \quad t_1, t_2 \in \mathbb{R}.$$

Here we have defined $\Phi_t(\vec{r}) := \Phi(t, \vec{r})$.

(Cf. Wiggins [38, Proposition 7.4.3, p. 93], Wirsching [31, Definition 8.6], Amann [39, p. 123/124].)

The interpretation of this definition is simple: If no time passes, one remains at the same point. Instead of moving from a first point to a second one in the time span

t_2 and then from this second one to a third one in a time span t_1, one may go directly from the first to the third in the time span $t_1 + t_2$.

Remark.

- A flow (also called a **phase flow**) is called a **global flow** if it is defined for all $t \in \mathbb{R}$ (as in Definition 2.11), a **semiflow** if it is defined for all $t \in \mathbb{R}^+$, and a **local flow** if it is defined for $t \in I$ (open interval I with $0 \in I$).
- For semiflows and local flows, Definition 2.11 has to be modified.
- In the modern mathematical literature, a dynamical system is *defined* as a flow. In our introduction, however, the dynamical system was initially described by an ODE, and the corresponding velocity vector field induced the flow.

Final remark: If the matrix A does not possess n linearly independent eigenvectors, then no diagonalization is possible, in which this case **generalized eigenvectors** may be used (cf. Guckenheimer [33, p. 9]). These are defined by

$$(A - \lambda I)^p \cdot \vec{w} = 0,$$
$$(A - \lambda I)^{p-1} \cdot \vec{w} \neq 0,$$

and may be used to transform any quadratic matrix A into **Jordan canonical form** (cf. Burg et al. [40, vol. II, p. 293] or Perko [30, Sect. 1.8]). As the formula shows, eigenvectors are also generalized eigenvectors (for $p = 1$).

2.8.7 Topological Orbit Equivalence

In this section, we shall define what it means to say that two vector fields are topologically orbit equivalent. Of course, the term **topological orbit equivalence** will include the case that the two vector fields can be transformed into each other by a simple rotation.

In order to simplify the situation even further, we assume that the two vector fields are given by

$$\vec{v}_1(\vec{r}_1) = A \cdot \vec{r}_1 \tag{2.100}$$

and

$$\vec{v}_2(\vec{r}_2) = B \cdot \vec{r}_2, \tag{2.101}$$

where A and B are quadratic matrices.

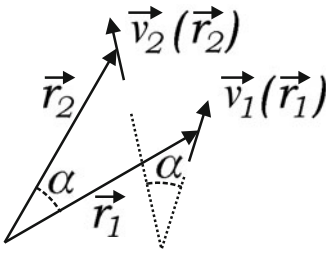

Fig. 2.5 Rotation of a vector field

If one vector field can be obtained as a result of rotating the other one, there must be a rotation matrix M such that

$$\vec{v}_2 = M \cdot \vec{v}_1 = M \cdot A \cdot \vec{r}_1 \tag{2.102}$$

holds. Now $\vec{v}_2$ depends on $\vec{r}_1$. In order to make $\vec{v}_2$ dependent on $\vec{r}_2$, we must rotate the coordinates in the same way as the vector field (see Fig. 2.5):

$$\vec{r}_2 = M \cdot \vec{r}_1. \tag{2.103}$$

Hence, we obtain

$$\vec{v}_2 = M \cdot A \cdot M^{-1} \cdot \vec{r}_2.$$

Since M is invertible as a rotation matrix, the matrix

$$B = M \cdot A \cdot M^{-1}$$

describes the well-known similarity transformation that may also be written in the form

$$B \cdot M = M \cdot A.$$

A **similarity transformation**, however, is usually written in the form

$$B = \tilde{M}^{-1} \cdot A \cdot \tilde{M},$$

so that we have to define $\tilde{M} = M^{-1}$ here.

Now we observe that two orbits can be identical even though the corresponding solution curves are parameterized in a different way.

A flow for Eq. (2.100) will be denoted by

$$\vec{r}_1(t_1, \vec{r}_{10}),$$

and a flow for Eq. (2.101) by

$$\vec{r}_2(t_2, \vec{r}_{20}).$$

In order to transform the orbits of these flows into each other, the starting points must be mapped first:

$$\vec{r}_{20} = M \cdot \vec{r}_{10}.$$

Our requirement that different parameterizations be allowed for both solution curves may be translated as follows: For every time t_1, there is a time t_2 such that

$$\vec{r}_2(t_2, \vec{r}_{20}) = M \cdot \vec{r}_1(t_1, \vec{r}_{10}),$$

and therefore

$$\vec{r}_2(t_2, M \cdot \vec{r}_{10}) = M \cdot \vec{r}_1(t_1, \vec{r}_{10})$$

holds. This formula must be included in the general definition of topological orbit equivalence if rotations are to be allowed as topologically equivalent transformations.

The previous considerations make the following definition transparent:

Definition 2.12. Two C^1 vector fields $\vec{v}_1(\vec{r}_1)$ and $\vec{v}_2(\vec{r}_2)$ are called topologically orbit equivalent[15] if a homeomorphism h exists such that for every pair $\vec{r}_{10}, t_1$, there exists t_2 such that

$$\Phi_{t_2}^{\vec{v}_2}(h(\vec{r}_{10})) = h(\Phi_{t_1}^{\vec{v}_1}(\vec{r}_{10})) \tag{2.104}$$

holds. Here, the orientation of the orbits must be preserved. If in addition, the parameterization by time is preserved, the vector fields are called **topologically conjugate**. In this definition, $\Phi_t^{\vec{v}}$ denotes the flow that is induced by the vector field $\vec{v}$.

(Cf. Sastry [36, Definition 7.18, p. 303], Wiggins [38, Definition 19.12.1, p. 346], Guckenheimer [33, p. 38, Definition 1.7.3].)

[15] Sometimes, the abbreviation "TOE" is used for the property "topological orbit equivalence" (cf. Jackson [41]).

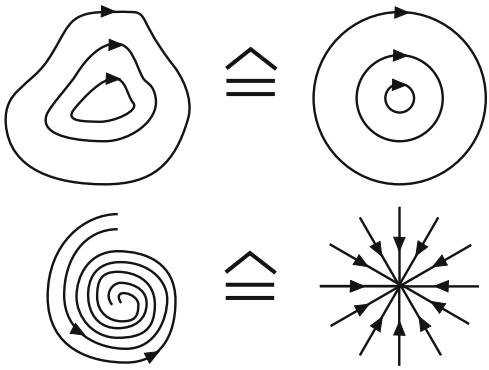

Fig. 2.6 Examples of topological orbit equivalence

Remark. A **homeomorphism** is a continuous map whose inverse map exists and is also continuous. The fact that a homeomorphism is used as a generalization of the rotation matrix, used as an example above, leads to the following features:

- Not only linear maps are allowed, but also nonlinear ones.
- The requirement that the map be continuous guarantees that neighborhoods of a point are mapped to neighborhoods of its image point. Therefore, the orbits are deformed but not torn apart. Two examples for topological orbit equivalence are shown in Fig. 2.6.

The fact that the validity of Eq. (2.104) is required for each initial point ensures that all orbits are transformed into each other. Hence, the entire phase portraits will be equivalent.

The preservation of the orientation may be checked by means of a continuously differentiable function $t_2(\vec{r}_{10}, t_1)$ with $\frac{\partial t_2}{\partial t_1} > 0$ (cf. Perko [30, Sect. 3.1, Remark 2, p. 183/184]).

Please note that different authors use slightly different definitions. In cases of doubt, one should therefore check the relevant definitions thoroughly.

Let us now consider the case that a vector field $\vec{v}(\vec{r}) = A \cdot \vec{r}$ is given by a real $n \times n$ matrix A and that we want to check whether this vector field is topologically orbit equivalent to a simpler vector field. Often, diagonalization is possible. This case will be discussed in the following.

Remark.

- In case diagonalization is not possible, it is *always* possible to transform the matrix into Jordan canonical form.
- Diagonalization of an $n \times n$ matrix is possible if and only if for each eigenvalue, the **algebraic multiplicity** (multiplicity of the zeros of the characteristic polynomial) equals the **geometric multiplicity** (number of linearly independent eigenvectors).

- Diagonalization of an $n \times n$ matrix is possible if and only if it possesses n linearly independent eigenvectors.
- Diagonalization is possible for every symmetric matrix with real elements.

Consider a matrix A for which diagonalization is possible. We will show now that the diagonal matrix[16]

$$B = X_A(0)^{-1} \cdot A \cdot X_A(0) \tag{2.105}$$

does in fact lead to a topologically orbit equivalent vector field. Here, $\tilde{M} = M^{-1} = X_A(0)$ denotes the matrix of the n eigenvectors of A. These are linearly independent, since diagonalization of A is possible (cf. Burg/Haf/Wille [40, vol. II, p. 280, Theorem 3.52]).

According to Eqs. (2.98) and (2.99), the flows are given by

$$\Phi_{t_1}^{\vec{v}_1}(\vec{r}_{10}) = X_A(t_1) X_A(0)^{-1} \vec{r}_{10}$$

for A and by

$$\Phi_{t_2}^{\vec{v}_2}(\vec{r}_{20}) = X_B(t_2) X_B(0)^{-1} \vec{r}_{20}$$

for B. In our case, the homeomorphism h is given by the matrix M. We therefore obtain

$$\Phi_{t_2}^{\vec{v}_2}(h(\vec{r}_{10})) = \Phi_{t_2}^{\vec{v}_2}(M \cdot \vec{r}_{10}) = X_B(t_2) X_B(0)^{-1} X_A(0)^{-1} \cdot \vec{r}_{10}.$$

On the other hand,

$$h(\Phi_{t_1}^{\vec{v}_1}(\vec{r}_{10})) = M \cdot \Phi_{t_1}^{\vec{v}_1}(\vec{r}_{10}) = X_A(0)^{-1} X_A(t_1) X_A(0)^{-1} \cdot \vec{r}_{10}$$

holds. We see that these expressions are equal for every initial vector $\vec{r}_{10}$ if

$$X_B(t_2) X_B(0)^{-1} X_A(0)^{-1} = X_A(0)^{-1} X_A(t_1) X_A(0)^{-1},$$

or

$$X_B(t_2) X_B(0)^{-1} = X_A(0)^{-1} X_A(t_1),$$

---

[16] Please note that $X_A(0)$ is the matrix of eigenvectors of A such that Eq. (2.105) actually represents the diagonalization step.

is valid. Due to Eq. (2.97), we know that

$$X_A(t_1) = X_A(0) \cdot D_A^{\lambda}(t_1)$$

and

$$X_B(t_2) = X_B(0) \cdot D_B^{\lambda}(t_2)$$

hold, so that the equation

$$X_B(0) \cdot D_B^{\lambda}(t_2) X_B(0)^{-1} = X_A(0)^{-1} X_A(0) \cdot D_A^{\lambda}(t_1) = D_A^{\lambda}(t_1)$$

has to be verified. Since B is a diagonal matrix, the Cartesian unit vectors are eigenvectors, so that

$$X_B(0) = I$$

is valid. Therefore, we have only to check whether

$$D_B^{\lambda}(t_2) = D_A^{\lambda}(t_1)$$

is true. Since the eigenvalues of A and B are equal due to diagonalization, we obtain

$$D_A^{\lambda}(t_1) = D_B^{\lambda}(t_1).$$

Here we had only to set $t_2 = t_1$. We have shown that diagonalization leads to topologically orbit equivalent vector fields.

2.8.8 *Classification of Fixed Points of an Autonomous Linear System of Second Order*

The considerations presented above indicate that a similarity transformation

$$B = \tilde{M}^{-1} \cdot A \cdot \tilde{M}$$

always leads to topologically orbit equivalent vector fields. Every similarity transformation leaves the eigenvalues unchanged (cf. Burg/Haf/Wille [40, vol. II, p. 272, 3.17]). This leads us to the assumption that the eigenvalues of a matrix at least influence the topological properties of the related vector field. Therefore, the eigenvalues are now used to characterize the fixed points.

We calculate the eigenvalues of a two-dimensional matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

with constant real elements. This leads to

$$\begin{aligned} &(a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = 0 \\ \Rightarrow\ &\lambda^2 - \lambda(a_{11} + a_{22}) + a_{11}a_{22} - a_{12}a_{21} = 0 \\ \Rightarrow\ &\lambda^2 - \lambda\ \mathrm{tr}A + \det A = 0. \end{aligned} \tag{2.106}$$

Hence, we obtain

$$\lambda = \frac{\mathrm{tr}A}{2} \pm \sqrt{\frac{(\mathrm{tr}A)^2}{4} - \det\ A} = B \pm \sqrt{C} \tag{2.107}$$

with

$$B = \frac{\mathrm{tr}A}{2}, \qquad C = \frac{(\mathrm{tr}A)^2}{4} - \det A. \tag{2.108}$$

We now try to distinguish as many cases as possible:

1. Both eigenvalues are real ($C \geq 0$).
   (a) Both are positive
      (i) $\lambda_1 > \lambda_2 > 0$
      (ii) $\lambda_1 = \lambda_2 > 0$
   (b) Both are negative
      (i) $\lambda_1 < \lambda_2 < 0$
      (ii) $\lambda_1 = \lambda_2 < 0$
   (c) One is positive, one is negative: $\lambda_1\lambda_2 < 0$
   (d) One equals 0 ($\lambda_1 = 0$):
      (i) $\lambda_2 > 0$
      (ii) $\lambda_2 < 0$
      (iii) $\lambda_2 = 0$
2. Imaginary eigenvalues: $C < 0$, $B = 0$, $\lambda_2 = -\lambda_1$
3. Complex eigenvalues: $C < 0$, $\lambda_1 = \lambda_2^*$
   (a) $B = \mathrm{Re}\{\lambda\} < 0$
   (b) $B = \mathrm{Re}\{\lambda\} > 0$

Hence we have found 11 distinct cases. If one eigenvalue is zero, then Eq. (2.106) leads to

$$\det\ A = 0.$$

As we will show now, this means in general that one row is a multiple of the other row, which contains the special case that one or both rows are zero.

If we now assume that the second row is a multiple of the first one,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ ka_{11} & ka_{12} \end{pmatrix},$$

we obtain the following condition for the eigenvalues:

$$\begin{aligned} &(a_{11} - \lambda)(ka_{12} - \lambda) - ka_{11}a_{12} = 0 \\ &\Rightarrow \lambda^2 - \lambda(a_{11} + ka_{12}) = 0. \end{aligned}$$

This shows that at least one eigenvalue is zero. The same can be shown for the case that the first equation is a multiple of the second one. Since one eigenvalue is 0, the relation

$$\det(A - \lambda I) = 0$$

leads to

$$\det\ A = 0.$$

In conclusion, the following statements for our two-dimensional case are equivalent:

- $\det\ A = 0$.
- One row of A is a multiple of the other row.
- At least one eigenvalue is 0.

The following general theorem holds:

Theorem 2.13. *The following statements for a quadratic matrix A are equivalent:*

- *The quadratic matrix A is regular.*
- *All row vectors (or column vectors) of A are linearly independent.*
- $\det\ A \neq 0$.
- *All eigenvalues of A are nonzero.*
- *A is invertible.*

In our two-dimensional case, the statement that for $\det A = 0$, one row is a multiple of the other one means that the equation $\vec{v}(\vec{r}) = A \cdot \vec{r} = 0$ is satisfied along a line through the origin or even everywhere. Hence, we have an infinite number of fixed points that are not separated from each other. In this case, we speak of **degenerate fixed points** (see Definition 2.6 on p. 59). This case will henceforth be excluded[17] (case 1.d), so that the number of relevant cases is reduced from 11 to 8. According to the behavior of the vector field in the vicinity of the fixed point $\vec{r}_F = 0$, the fixed points are named as follows:

1. Both eigenvalues are real ($C \geq 0$).
   (a) Both positive
      (i) $\lambda_1 > \lambda_2 > 0$: **unstable node**
      (ii) $\lambda_1 = \lambda_2 > 0$: **unstable improper node or unstable star**
   (b) Both negative
      (i) $\lambda_1 < \lambda_2 < 0$: **stable node**
      (ii) $\lambda_1 = \lambda_2 < 0$: **stable improper node or stable star**
   (c) One positive, one negative: $\lambda_1 \lambda_2 < 0$: **saddle point**
2. Imaginary eigenvalues: $C < 0$, $B = 0$, $\lambda_2 = -\lambda_1$: **center** or **elliptic fixed point**
3. Complex eigenvalues: $C < 0$, $\lambda_1 = \lambda_2^*$
   (a) $\mathrm{Re}\{\lambda\} < 0$: **stable spiral point** or **stable focus**
   (b) $\mathrm{Re}\{\lambda\} > 0$: **unstable spiral point** or **unstable focus**

The classification of fixed points is summarized in Table 2.4 on p. 74 (cf. [29, 42, 43]). As the table shows, not only the eigenvalues can be used to characterize the fixed points. In the column "Topology", the numbers in parentheses denote the quantity of eigenvalues with positive real part. The column "Topology" also contains the so-called **index** in brackets. In order to calculate the index, one considers a closed path around the fixed point with an orientation that is mathematically positive. Now one checks how many revolutions the vectors of the vector field perform while "walking" on the path. If, e.g., the vectors of the vector field also perform one revolution in mathematically positive orientation, the index is $+1$. If the vector field rotates in opposite direction, the index is -1.

Figures 2.7, 2.8, 2.9, 2.10, 2.11, and 2.12 on p. 75 and 76 show how orbits in the vicinity of the fixed point look in principle for each type of fixed point. In case the fixed point is a stable node, star, or spiral point, the orientation of the solution curves will be towards the fixed point in the middle; in case of an unstable node, star, or spiral point, all solution curves will be directed outwards. Each picture is just an example; in the specific case under consideration the orbits may of course be deformed significantly.

[17] Please note that small changes of the entries of the matrix A are likely to convert degenerate fixed points into nondegenerate (isolated) fixed points (cf. Sastry [36, p. 2]).

Table 2.4 Classification of isolated fixed points in the plane (regular Jacobian matrix)

	Stability	**Topology**	**Eigenvalues**	**Fixed Point**
$\mathrm{Re}\{\lambda\} = 0$	stable	center [+1] (-)	imaginary $\lambda_1 = -\lambda_2$	center =elliptic fixed point
hyperbolic $\mathrm{Re}\{\lambda\} \neq 0$	asymptotically stable	sink [+1] (0) $\mathrm{Re}\{\lambda\} < 0$	real $\lambda_1 < \lambda_2 < 0$	stable improper node (two[a] tangents)
			real $\lambda_1 = \lambda_2 < 0$	stable improper node (one tangent, one eigenvector)
				stable proper node =stable star (many tangents, two eigenvectors)
			complex $\mathrm{Re}\{\lambda\} < 0$	stable spiral point =stable focus
	unstable	source [+1] (2) $\mathrm{Re}\{\lambda\} > 0$	real $\lambda_1 > \lambda_2 > 0$	unstable improper node (two[a] tangents)
			real $\lambda_1 = \lambda_2 > 0$	unstable improper node (one tangent, one eigenvector)
				unstable proper node =unstable star (many tangents, two eigenvectors)
			complex $\mathrm{Re}\{\lambda\} > 0$	unstable spiral point =unstable focus
		saddle point [-1] (1)	real $\lambda_1 < 0 < \lambda_2$	saddle point

[a]one tangent for two orbits, one tangent for all other orbits

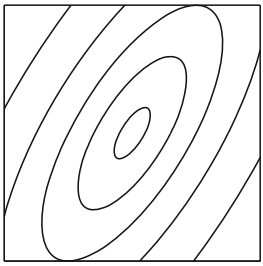

Fig. 2.7 Center

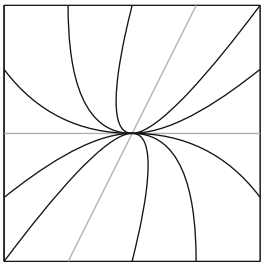

Fig. 2.8 Node with two tangents

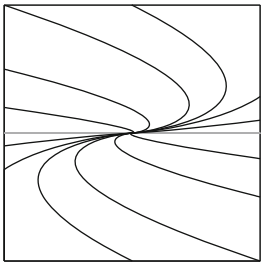

Fig. 2.9 Node with one tangent

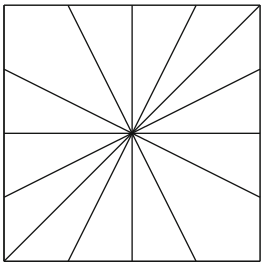

Fig. 2.10 Star

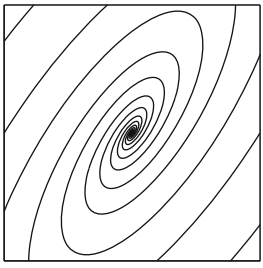

Fig. 2.11 Spiral point

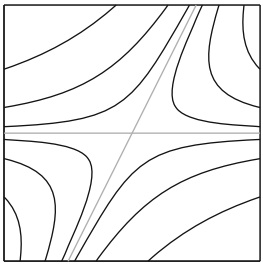

Fig. 2.12 Saddle point

2.8.9 *Nonlinear Systems*

Consider the nonlinear autonomous system

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}) \tag{2.109}$$

with the initial condition

$$\vec{r}(0) = \vec{r}_0,$$

where $\vec{v}(\vec{r}) \in C^2$ has a fixed point $\vec{r}_\mathrm{F}$ with

$$\vec{v}(\vec{r}_\mathrm{F}) = 0.$$

As the results summarized in this section show, the **linearization**

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = A \cdot \vec{r} \tag{2.110}$$

of the system (2.109), where $A = D\vec{v}(\vec{r}_\mathrm{F})$ is the **Jacobian matrix**[18] at $\vec{r} = \vec{r}_\mathrm{F}$, is a powerful tool for analyzing a nonlinear system in the vicinity of its fixed points.

[18] The Jacobian matrix is defined as

$$D\vec{v}(\vec{r}) := \frac{\partial \vec{v}}{\partial \vec{r}} := \begin{pmatrix} \frac{\partial v_1}{\partial x_1} & \frac{\partial v_1}{\partial x_2} & \cdots & \frac{\partial v_1}{\partial x_n} \\ \frac{\partial v_2}{\partial x_1} & \frac{\partial v_2}{\partial x_2} & \cdots & \frac{\partial v_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_n}{\partial x_1} & \frac{\partial v_n}{\partial x_2} & \cdots & \frac{\partial v_n}{\partial x_n} \end{pmatrix}$$

where $\vec{r} = x_1\,\vec{e}_1 + x_2\,\vec{e}_2 + \cdots + x_n\,\vec{e}_n$.

The determinant of the Jacobian matrix is called **Jacobian determinant** or simply **Jacobian**, and one writes

$$\det\ D\vec{v}(\vec{r}) = \frac{\partial(v_1, v_2, \ldots, v_n)}{\partial(x_1, x_2, \ldots, x_n)}.$$

Please note that $\vec{r} = 0$ in Eq. (2.110) corresponds to $\vec{r} = \vec{r}_\mathrm{F}$ in Eq. (2.109), i.e., the fixed point of the nonlinear system was shifted to the origin of the linearized system.

Theorem 2.14. *Consider the nonlinear system (2.109) with the linearization (2.110) at a fixed point $\vec{r}_\mathrm{F}$. If A is nonsingular, then the fixed point $\vec{r}_\mathrm{F}$ is isolated (i.e., nondegenerate).*

(See Sastry [36, Proposition 1.5, p. 13], Perko [30, Definition 2, p. 173].)

If one or more eigenvalues of the Jacobian matrix are zero, the fixed point is a degenerate fixed point. This is the generalization of the linear case.

Definition 2.15. A fixed point is called a hyperbolic fixed point if no eigenvalue of the Jacobian matrix has zero real part.

Theorem 2.16. *If the fixed point $\vec{r}_\mathrm{F}$ is a hyperbolic fixed point, then there exist two neighborhoods U of $\vec{r}_\mathrm{F}$ and V of $\vec{r} = 0$ and a homeomorphism $h : U \to V$, such that h transforms the orbits of Eq. (2.109) into orbits of*

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = A \cdot \vec{r} \qquad \textit{with} \qquad A = D\vec{v}(\vec{r}_\mathrm{F}).$$

Orientation and parameterization by time are preserved.

(Cf. Guckenheimer [33, Theorem 1.3.1 (Hartman-Grobman), p. 13], Perko [30, Theorem Sect. 2.8, p. 120], and Bronstein [44, Sect. 11.3.2].)

In other words, we may state the following theorem.

Theorem 2.17 (Hartman–Grobman). *Let $\vec{r}_\mathrm{F}$ be a hyperbolic fixed point. Then the nonlinear problem*

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}), \qquad \vec{v} \in C^1(D), \qquad D \subset \mathbb{R}^n \textit{ open},$$

and the linearized problem

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = A \cdot (\vec{r} - \vec{r}_\mathrm{F})$$

with

$$A = D\vec{v}(\vec{r}_\mathrm{F})$$

are topologically conjugate in a neighborhood of $\vec{r}_\mathrm{F}$.

(See Wiggins [38, Theorem 19.12.6, p. 350].)

If the fixed point is not hyperbolic, i.e., a center (elliptic fixed point), then the smallest nonlinearities are sufficient to create a stable or an unstable spiral point. This is why the theorem refers to hyperbolic fixed points only.

If the real parts of all eigenvalues of $D\vec{v}(\vec{r}_\mathrm{F})$ are negative, then $\vec{r}_\mathrm{F}$ is asymptotically stable. If the real part of at least one eigenvalue is positive, then $\vec{r}_\mathrm{F}$ is unstable:

Theorem 2.18. *Let $D \subset \mathbb{R}^n$ be an open set, $\vec{v}(\vec{r})$ continuously differentiable on D, and $\vec{r}_\mathrm{F}$ a fixed point of*

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}).$$

If the real parts of all eigenvalues of $D\vec{v}(\vec{r}_\mathrm{F})$ are negative, then $\vec{r}_\mathrm{F}$ is asymptotically stable. If $\vec{r}_\mathrm{F}$ is stable, then no eigenvalue has positive real part.

(Cf. Bronstein [44, Sect. 11.3.1], Perko [30, Theorem 2, p. 130].)

A saddle point has the special property that two trajectories exist that approach the saddle point for $t \to \infty$, whereas two different trajectories exist that approach the saddle point for $t \to -\infty$ (cf. [30, Sect. 2.10, Definition 5]). These four trajectories define a **separatrix**. Loosely speaking, a separatrix is a trajectory that "meets" the saddle point.

2.8.10 Characteristic Equation

Consider the autonomous linear homogeneous nth-order ordinary differential equation

$$a_n \frac{\mathrm{d}^n y}{\mathrm{d}t^n} + a_{n-1} \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}} + \cdots + a_1 \frac{\mathrm{d}y}{\mathrm{d}t} + a_0 y(t) = 0. \tag{2.111}$$

As usual (see Sect. 2.8.1), we define the vector $\vec{r} = (x_1, x_2, \ldots, x_n)^\mathrm{T}$ by

$$\begin{aligned} x_1 &= y, \\ x_2 &= \frac{\mathrm{d}y}{\mathrm{d}t}, \\ &\cdots \\ x_n &= \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}}, \end{aligned}$$

which leads to

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= x_3, \\ &\cdots \\ \dot{x}_n &= -\frac{a_0}{a_n} x_1 - \frac{a_1}{a_n} x_2 - \cdots - \frac{a_{n-1}}{a_n} x_n, \end{aligned}$$

in order to obtain the standard form

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}).$$

This may be written as

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = A \cdot \vec{r}$$

if the following $n \times n$ matrix is defined:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & -\frac{a_2}{a_n} & -\frac{a_3}{a_n} & \dots & -\frac{a_{n-2}}{a_n} & -\frac{a_{n-1}}{a_n} \end{pmatrix}. \tag{2.112}$$

According to Sect. 2.8.6, Eq. (2.96), we know that asymptotic stability is reached if all eigenvalues of this system matrix have negative real part. Therefore, we now describe how to find the eigenvalues based on the requirement that the determinant

$$D_n^F = \det(A - \lambda I) \tag{2.113}$$

equal zero. Let us begin with $n = 2$ as an example:

$$A = \begin{pmatrix} 0 & 1 \\ -\frac{a_0}{a_2} & -\frac{a_1}{a_2} \end{pmatrix}$$

$$\Rightarrow D_2^F = \det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ -\frac{a_0}{a_2} & -\frac{a_1}{a_2} - \lambda \end{vmatrix} = \lambda^2 + \frac{a_1}{a_2}\lambda + \frac{a_0}{a_2} \stackrel{!}{=} 0.$$

For $n = 3$, we obtain

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{a_0}{a_3} & -\frac{a_1}{a_3} & -\frac{a_2}{a_3} \end{pmatrix}$$

$$\Rightarrow D_3^F = \det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ -\frac{a_0}{a_3} & -\frac{a_1}{a_3} & -\frac{a_2}{a_3} - \lambda \end{vmatrix} = -\lambda^3 - \frac{a_2}{a_3}\lambda^2 - \frac{a_1}{a_3}\lambda - \frac{a_0}{a_3} \stackrel{!}{=} 0.$$

These two results lead us to the assumption that

$$D_n^F = (-1)^n \left(\lambda^n + \frac{a_{n-1}}{a_n}\lambda^{n-1} + \frac{a_{n-2}}{a_n}\lambda^{n-2} + \cdots + \frac{a_2}{a_n}\lambda^2 + \frac{a_1}{a_n}\lambda + \frac{a_0}{a_n} \right)$$
$$= (-1)^n \sum_{k=0}^{n} \frac{a_k}{a_n}\lambda^k \tag{2.114}$$

holds in general. In Appendix A.6, it is shown that this is indeed true. The requirement that the polynomial in Eq. (2.114) equal zero is called the **characteristic equation** of the ODE (2.111). One easily sees that the characteristic equation is also obtained if the Laplace transform is applied to the original ODE (2.111):

$$\left(a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0\right) Y(s) = 0$$
$$\Rightarrow a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0 \qquad \text{for} \quad Y(s) \neq 0.$$

The matrix A is called the **Frobenius companion matrix** of the polynomial. Please note that instead of finding the zeros (roots) of the polynomial, one may also determine the eigenvalues of the companion matrix A and vice versa.

Hence, asymptotic stability of the dynamical system defined by the ODE (2.111) is equivalently shown

- if all zeros of the characteristic equation have negative real part.
- if all eigenvalues of the system matrix have negative real part.

In case of asymptotic stability, one also calls the system matrix a **strictly** or **negative stable matrix** (cf. [45, Definition 2.4.2]) or a **Hurwitz matrix**.

2.9 Continuity Equation

Consider a particle density ρ in space and a velocity field $\vec{v}$ that moves the particles. We will now calculate how the particle density in a fixed volume V changes due to the velocity field.

For this purpose, we consider a small volume element ΔV at the surface of the three-dimensional domain V. As shown in Fig. 2.13, this contains in total

$$\Delta n = \Delta V\ \rho = \Delta h\ \Delta A\ \rho$$

particles. During the time interval Δt, this quantity of

$$\Delta n = v_n \Delta t\ \Delta A\ \rho = \rho\ \Delta t\ \vec{v} \cdot \Delta \vec{A}$$

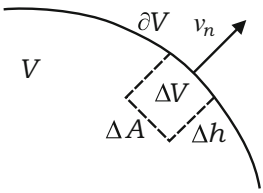

Fig. 2.13 Volume element at the surface of a region

particles will leave the domain V, where $v_n = \Delta h/\Delta t$ denotes the normal component of the velocity vector $\vec{v}$ with respect to the surface of the domain V. Hence, we have

$$\int_V \rho(t + \Delta t)\, \mathrm{d}V = \int_V \rho(t)\, \mathrm{d}V - \oint_{\partial V} \rho \Delta t\, \vec{v} \cdot \mathrm{d}\vec{A}.$$

As a limit for $\Delta t \to 0$, one therefore obtains

$$\int_V \dot{\rho}\, \mathrm{d}V = -\oint_{\partial V} \rho\, \vec{v} \cdot \mathrm{d}\vec{A}.$$

According to Gauss's theorem,

$$\oint_{\partial V} \vec{V} \cdot \mathrm{d}\vec{A} = \int_V \mathrm{div}\, \vec{V}\, \mathrm{d}V,$$

one concludes by setting $\vec{V} = \rho\, \vec{v}$:

$$\int_V \dot{\rho}\, \mathrm{d}V = -\int_V \mathrm{div}(\rho\, \vec{v})\, \mathrm{d}V.$$

Since this equation must be valid for arbitrary choices of the domain V, one obtains

$$\boxed{-\dot{\rho} = \mathrm{div}(\rho\, \vec{v}).} \tag{2.115}$$

This is the **continuity equation**, for which we only assumed that no particles disappear and no particles are generated. Instead of the particle density, one could have considered different densities, such as the mass density, assuming mass conservation in that case. If we take the charge density as an example, charge conservation leads to

$$\boxed{-\dot{\rho}_q = \mathrm{div}(\rho_q\, \vec{v}) = \mathrm{div}\, \vec{J},}$$

which we already know as Eq. (2.46) and where $\vec{J} = \rho_q \vec{v}$ is the **convection current density**.

Remark. If $\dot{\rho}_q = 0$ holds, then the density will remain constant at every location; one obtains a stationary flow with

$$\operatorname{div}(\rho_q \, \vec{v}) = 0 \qquad \text{or} \qquad \operatorname{div} \vec{J} = 0.$$

This equation is known in electromagnetism for steady currents.

2.10 Area Preservation in Phase Space

In this section, we discuss how an area or a volume that is defined by the contained particles is modified when the particles are moving.

2.10.1 Velocity Vector Fields

Consider an arbitrary domain A in $\mathbb{R}^2$ at time t. Particles located inside the domain and on its boundary at time t will move a bit farther during the time span Δt. This movement is determined by the velocity field $\vec{v}(x, y)$.

Let us define a parameterization of the domain such that $x(\alpha, \beta)$ and $y(\alpha, \beta)$ are given depending on the parameters α and β. This leads to the area

$$A(t) = \int_A \mathrm{d}A = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} \left| \frac{\partial(x, y)}{\partial(\alpha, \beta)} \right| \, \mathrm{d}\alpha \, \mathrm{d}\beta.$$

The coordinates $\vec{r} = (x, y)$ denote each point of A. Such a point $\vec{r}$ will move to the new point

$$\vec{r}\,' = \vec{r} + \vec{v}(\vec{r}) \Delta t$$

after the time span Δt. Since $\vec{r}$ depends on α and β, it follows that $\vec{r}\,'$ will also depend on these parameters. For the area of the deformed domain at the time $t + \Delta t$, we therefore get

$$A(t + \Delta t) = A' = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} \left| \frac{\partial(x', y')}{\partial(\alpha, \beta)} \right| \, \mathrm{d}\alpha \, \mathrm{d}\beta$$

with

$$x' = x + v_x \Delta t,$$
$$y' = y + v_y \Delta t.$$

Using the abbreviation

$$\xi = \frac{\partial(x,y)}{\partial(\alpha,\beta)} = \begin{vmatrix} \frac{\partial x}{\partial \alpha} & \frac{\partial x}{\partial \beta} \\ \frac{\partial y}{\partial \alpha} & \frac{\partial y}{\partial \beta} \end{vmatrix} = \det(\vec{r}_\alpha, \vec{r}_\beta)$$

or

$$\xi' = \det(\vec{r}\,'_\alpha, \vec{r}\,'_\beta)$$

leads to

$$\begin{aligned}\xi' &= \det(\vec{r}_\alpha + \vec{v}_\alpha \Delta t, \vec{r}_\beta + \vec{v}_\beta \Delta t) = \det(\vec{r}_\alpha, \vec{r}_\beta + \vec{v}_\beta \Delta t) + \det(\vec{v}_\alpha \Delta t, \vec{r}_\beta + \vec{v}_\beta \Delta t) = \\ &= \det(\vec{r}_\alpha, \vec{r}_\beta) + \det(\vec{r}_\alpha, \vec{v}_\beta \Delta t) + \det(\vec{v}_\alpha \Delta t, \vec{r}_\beta) + \det(\vec{v}_\alpha \Delta t, \vec{v}_\beta \Delta t) = \\ &= \xi + \Delta t \left(\det(\vec{r}_\alpha, \vec{v}_\beta) + \det(\vec{v}_\alpha, \vec{r}_\beta)\right) + \Delta t^2 \det(\vec{v}_\alpha, \vec{v}_\beta).\end{aligned}$$

One obtains

$$\begin{aligned}\frac{\partial \xi}{\partial t} &= \lim_{\Delta t \to 0} \frac{\xi' - \xi}{\Delta t} = \det(\vec{r}_\alpha, \vec{v}_\beta) + \det(\vec{v}_\alpha, \vec{r}_\beta) = \\ &= \frac{\partial x}{\partial \alpha}\frac{\partial v_y}{\partial \beta} - \frac{\partial y}{\partial \alpha}\frac{\partial v_x}{\partial \beta} + \frac{\partial v_x}{\partial \alpha}\frac{\partial y}{\partial \beta} - \frac{\partial v_y}{\partial \alpha}\frac{\partial x}{\partial \beta} = \\ &= \frac{\partial x}{\partial \alpha}\left(\frac{\partial v_y}{\partial x}\frac{\partial x}{\partial \beta} + \frac{\partial v_y}{\partial y}\frac{\partial y}{\partial \beta}\right) - \frac{\partial y}{\partial \alpha}\left(\frac{\partial v_x}{\partial x}\frac{\partial x}{\partial \beta} + \frac{\partial v_x}{\partial y}\frac{\partial y}{\partial \beta}\right) + \\ &+ \frac{\partial y}{\partial \beta}\left(\frac{\partial v_x}{\partial x}\frac{\partial x}{\partial \alpha} + \frac{\partial v_x}{\partial y}\frac{\partial y}{\partial \alpha}\right) - \frac{\partial x}{\partial \beta}\left(\frac{\partial v_y}{\partial x}\frac{\partial x}{\partial \alpha} + \frac{\partial v_y}{\partial y}\frac{\partial y}{\partial \alpha}\right) = \\ &= \frac{\partial v_y}{\partial y}\left(\frac{\partial x}{\partial \alpha}\frac{\partial y}{\partial \beta} - \frac{\partial x}{\partial \beta}\frac{\partial y}{\partial \alpha}\right) + \frac{\partial v_x}{\partial x}\left(\frac{\partial x}{\partial \alpha}\frac{\partial y}{\partial \beta} - \frac{\partial x}{\partial \beta}\frac{\partial y}{\partial \alpha}\right)\end{aligned}$$

$$\Rightarrow \frac{\partial \xi}{\partial t} = \xi \operatorname{div} \vec{v}.$$

Since $\xi \neq 0$ is valid ($\xi = |\xi| \operatorname{sgn} \xi$, $\operatorname{sgn} \xi$ constant), one gets

$$\frac{\partial |\xi|}{\partial t} = |\xi| \operatorname{div} \vec{v}.$$

Due to

$$A(t) = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} |\xi| \, \mathrm{d}\alpha \, \mathrm{d}\beta,$$

one obtains

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} \frac{\partial|\xi|}{\partial t}\, \mathrm{d}\alpha\, \mathrm{d}\beta = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} |\xi|\, \mathrm{div}\, \vec{v}\, \mathrm{d}\alpha\, \mathrm{d}\beta.$$

Now it is obvious that the area remains constant for div $\vec{v} = 0$. If we were talking here about a fluid, such a fluid would obviously be **incompressible**; were one to try to compress it, the shape would be modified, but the total area (or volume) occupied by the particles would remain the same.

2.10.2 Maps

Now we analyze in a more general way how an area

$$A = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} \left| \frac{\partial(x, y)}{\partial(\alpha, \beta)} \right| \mathrm{d}\alpha\, \mathrm{d}\beta$$

is modified by a map

$$\vec{r}\,' = \vec{F}(\vec{r}),$$

which transforms each vector $\vec{r} = (x, y)$ into a vector $\vec{r}\,' = (x', y')$. The parameterization will remain the same. Each point of the domain moves to a new point, so that the shape of the domain will change in general. Hence, we have to calculate

$$A' = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} \left| \frac{\partial(x', y')}{\partial(\alpha, \beta)} \right| \mathrm{d}\alpha\, \mathrm{d}\beta.$$

According to Appendix A.5, we have

$$A' = A$$

if

$$\left| \frac{\partial(x', y')}{\partial(x, y)} \right| = 1$$

is satisfied; the Jacobian of area-preserving maps is obviously $+1$ or -1.

We now check this general formula for the situation discussed in the previous section, where a special map

$$\vec{r}\,' = \begin{pmatrix} x' \\ y' \end{pmatrix} = \vec{r} + \vec{v}\, \Delta t = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} v_x \\ v_y \end{pmatrix} \Delta t$$

was given. The Jacobian is then

$$\begin{vmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{vmatrix} = \begin{vmatrix} 1+\frac{\partial v_x}{\partial x}\,\Delta t & \frac{\partial v_x}{\partial y}\,\Delta t \\ \frac{\partial v_y}{\partial x}\,\Delta t & 1+\frac{\partial v_y}{\partial y}\,\Delta t \end{vmatrix} = 1+\Delta t\left(\frac{\partial v_x}{\partial x}+\frac{\partial v_y}{\partial y}\right)$$
$$+\Delta t^2\left(\frac{\partial v_x}{\partial x}\,\frac{\partial v_y}{\partial y}-\frac{\partial v_x}{\partial y}\,\frac{\partial v_y}{\partial x}\right).$$

If one now wants to calculate

$$\frac{\partial \xi}{\partial t} = \lim_{\Delta t\to 0}\frac{\xi'-\xi}{\Delta t},$$

one obtains, due to

$$\xi' = \frac{\partial(x',y')}{\partial(x,y)}\,\xi$$

(see Appendix A.5), the relation

$$\frac{\partial \xi}{\partial t} = \xi\,\lim_{\Delta t\to 0}\frac{\frac{\partial(x',y')}{\partial(x,y)}-1}{\Delta t} = \xi\,\left(\frac{\partial v_x}{\partial x}+\frac{\partial v_y}{\partial y}\right) = \xi\ \mathrm{div}\,\vec{v},$$

as above.

2.10.3 Liouville's Theorem

The statement derived above that the condition

$$\boxed{\mathrm{div}\,\vec{v}=0}$$

leads to **area preservation** or—depending on the dimension—to **volume preservation** in phase space is called **Liouville's theorem**. This equation is also given as the condition for **incompressible flows**. Please note that one can speak of area preservation only if the area is defined in a unique way. This is, for example possible, if a continuous particle density ρ with clear boundaries in phase space is assumed, but not for a discrete distribution of individual particles (or only approximately if large numbers of particles are present). We will return to this problem later.

Liouville's theorem (and therefore also area/volume preservation) is also valid if

$$\vec{v}(\vec{r},t)$$

depends explicitly on time (cf. Szebehely [46, p. 55], Fetter [47, p. 296], or Budó [48, p. 446]).

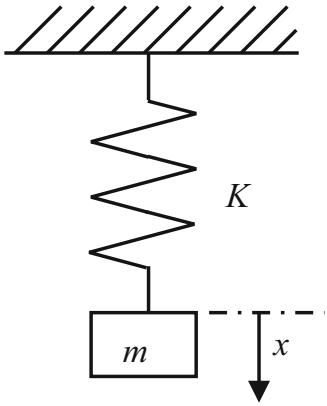

Fig. 2.14 Spring–mass system

2.11 Hamiltonian Systems

Hamiltonian theory is usually developed in the scope of classical mechanics after introduction of the Lagrangian formulation (cf. [19]). Here we choose a different approach by introducing Hamiltonian functions directly. This can, of course, be no replacement for intense studies of Hamiltonian mechanics, but it is sufficient to understand some basics that are relevant in the following chapters of this book.

2.11.1 Example for Motivation

Consider the system sketched in Fig. 2.14. The spring constant K and the mass m are known. The force balance leads to

$$\begin{aligned} m\ddot{x} &= -Kx \\ \Leftrightarrow \ddot{x} + \frac{K}{m}\, x &= 0. \end{aligned} \tag{2.116}$$

The general solution is obtained using the following ansatz:

$$\begin{aligned} x &= A\ \cos(\omega t) + B\ \sin(\omega t), \\ \dot{x} &= -A\omega\ \sin(\omega t) + B\omega\ \cos(\omega t), \\ \ddot{x} &= -A\omega^2\ \cos(\omega t) - B\omega^2\ \sin(\omega t). \end{aligned}$$

We obviously obtain

$$\omega^2 = \frac{K}{m}.$$

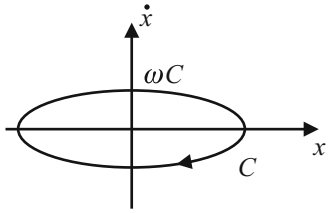

Fig. 2.15 Trajectory of the spring–mass system

One can alternatively write x in the form

$$x = C\ \cos(\omega t - \varphi) = C\ \cos(\omega t)\ \cos\varphi + C\ \sin(\omega t)\ \sin\varphi. \tag{2.117}$$

This leads to:

$$A = C\ \cos\varphi,$$
$$B = C\ \sin\varphi.$$

For $\dot{x}$ and $\ddot{x}$ one obtains

$$\dot{x} = -C\omega\ \sin(\omega t - \varphi), \tag{2.118}$$

$$\ddot{x} = -C\omega^2\ \cos(\omega t - \varphi) = -\omega^2 x. \tag{2.119}$$

The result may be drawn as shown in Fig. 2.15.

The quantity φ obviously determines only the initial conditions, whereas C is the oscillation amplitude and thus characterizes the energy of the system.

The quantity C (regarded as a system property), as well as the energy W, remains constant on the trajectory. If, in general, we have an invariant H that depends on two variables q and p, then the trajectory $(q(t), p(t))$ will have the property that

$$\frac{\mathrm{d}H}{\mathrm{d}t} = 0$$

holds. One concludes that

$$\frac{\partial H}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}t} + \frac{\partial H}{\partial p}\frac{\mathrm{d}p}{\mathrm{d}t} = 0$$
$$\Leftrightarrow \frac{\partial H}{\partial q}\dot{q} + \frac{\partial H}{\partial p}\dot{p} = 0.$$

This equation is obviously satisfied if the following system of equations is valid:

$$\frac{\partial H}{\partial p} = \dot{q}, \tag{2.120}$$

$$\frac{\partial H}{\partial q} = -\dot{p}. \tag{2.121}$$

These equations are called **Hamilton's equations**. The function $H(q, p)$ is called the **Hamiltonian**. We will now check whether this system of equations is actually satisfied in our example.

It is clear that the total energy of the system remains constant:

$$W(x, v) = \frac{K}{2}\, x^2 + \frac{1}{2} m v^2.$$

This can also be seen formally if the differential equation

$$m\, \ddot{x} + K\, x = 0$$

is multiplied by $\dot{x}$:

$$\begin{aligned} & m\, \dot{x}\, \ddot{x} + K\, \dot{x}\, x = 0 \\ \Leftrightarrow\ & \frac{m}{2} \frac{\mathrm{d}}{\mathrm{d}t}(\dot{x}^2) + \frac{K}{2} \frac{\mathrm{d}}{\mathrm{d}t}(x^2) = 0 \\ \Leftrightarrow\ & \frac{\mathrm{d}W}{\mathrm{d}t} = 0. \end{aligned}$$

Here we obviously have

$$\begin{aligned} \frac{\partial W}{\partial x} &= K x, \\ \frac{\partial W}{\partial v} &= m v, \\ \dot{x} &= v, \\ \dot{v} = \ddot{x} &= -\frac{K}{m} x. \end{aligned}$$

As a result, one obtains

$$\begin{aligned} \frac{\partial W}{\partial x} &= -m\dot{v}, \\ \frac{\partial W}{\partial v} &= m\dot{x}. \end{aligned}$$

These equations are not yet equivalent to the above-mentioned Hamilton equations. However, it is not a big step to work with $p = mv$ instead of v:

$$\frac{\partial W}{\partial x} = -\dot{p},$$
$$\frac{\partial W}{\partial p} = \frac{1}{m}\frac{\partial W}{\partial v} = \dot{x}.$$

Now the equations actually have the desired form; Hamilton's equations are satisfied. The function $W(x, p)$ is called a Hamiltonian, since it satisfies Hamilton's equations.

As shown above, C is also constant along the trajectory. We can obviously determine C as follows, based on Eqs. (2.117) and (2.118):

$$(\omega C)^2 = (\omega x)^2 + \dot{x}^2.$$

It seems to be useful to define the following quantities in order to get $C = \sqrt{\bar{q}^2 + \bar{p}^2}$:

$$\bar{q} = x, \qquad \bar{p} = \frac{\dot{x}}{\omega}.$$

Calculating the partial derivatives leads to

$$\frac{\partial C}{\partial \bar{p}} = \frac{1}{2C}\, 2\bar{p},$$
$$\frac{\partial C}{\partial \bar{q}} = \frac{1}{2C}\, 2\bar{q},$$
$$\dot{\bar{q}} = \dot{x} = \omega \bar{p}.$$

With the help of Eq. (2.116), one obtains

$$\dot{\bar{p}} = \frac{\ddot{x}}{\omega} = -\omega x = -\omega \bar{q}.$$

We therefore get two coupled differential equations:

$$\frac{\partial C}{\partial \bar{p}} = \frac{1}{\omega C}\,\dot{\bar{q}} \quad \Rightarrow \omega C\, \frac{\partial C}{\partial \bar{p}} = \dot{\bar{q}},$$
$$\frac{\partial C}{\partial \bar{q}} = -\frac{1}{\omega C}\,\dot{\bar{p}} \quad \Rightarrow \omega C\, \frac{\partial C}{\partial \bar{q}} = -\dot{\bar{p}}.$$

This is reminiscent of the product rule

$$\frac{\partial (C^2)}{\partial \bar{p}} = 2C\frac{\partial C}{\partial \bar{p}}, \qquad \text{or} \qquad \frac{\partial (C^2)}{\partial \bar{q}} = 2C\frac{\partial C}{\partial \bar{q}}.$$

If we therefore set

$$H = C^2 \frac{\omega}{2},$$

we again obtain Hamilton's equations:

$$\frac{\partial H}{\partial \bar{p}} = \dot{\bar{q}},$$
$$\frac{\partial H}{\partial \bar{q}} = -\dot{\bar{p}}.$$

We conclude that on the trajectory, the Hamiltonian

$$H(\bar{q}, \bar{p}) = \frac{\omega}{2}\bar{q}^2 + \frac{\omega}{2}\bar{p}^2$$

is constant if in our example, the generalized coordinate

$$\bar{q} = x$$

and the generalized momentum

$$\bar{p} = \dot{x}/\omega$$

are used. In our special case, $\bar{q}$ is a physical coordinate, but $\bar{p}$ is not the physical momentum. In general, $\bar{q}$ also does not need to be a physical coordinate. This explains the terms "generalized coordinate" and "generalized momentum." Here they formally play a similar mathematical role. We summarize:

- $H(q, p)$ is called a Hamiltonian if Hamilton's equations (2.120) and (2.121) are satisfied.
- The Hamiltonian describes a dynamical system.
- The quantities q and p are called a **generalized coordinate** and **generalized momentum**, respectively. They do not necessarily have to be identical to the physical coordinates and momenta.
- Different Hamiltonians may exist for the same dynamical system (in our example, $\frac{\omega}{2}\ C^2$ and W), and also different definitions of q and p are possible.

2.11.2 Arbitrary Number of Variables

Our introductory example contained only one coordinate and one momentum variable. For an arbitrary number of coordinate variables, Hamilton's equations are

$$\boxed{\frac{\partial H}{\partial p_i} = \dot{q}_i\,,} \tag{2.122}$$

$$\boxed{\frac{\partial H}{\partial q_i} = -\dot{p}_i\,.} \tag{2.123}$$

In this case, the Hamiltonian

$$H(q_i\,, p_i\,, t)$$

depends on n generalized coordinates q_i $(1 \leq i \leq n)$, on n generalized momentum variables p_i, and in general, explicitly on the time t. Its total derivative with respect to time is

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{i=1}^{n} \left[\frac{\partial H}{\partial q_i}\dot{q}_i + \frac{\partial H}{\partial p_i}\dot{p}_i \right] + \frac{\partial H}{\partial t}.$$

By means of Eqs. (2.122) and (2.123), one obtains

$$\boxed{\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial t}.}$$

This shows that if the Hamiltonian does not explicitly depend on time (as in our introductory example), it is constant along the trajectory. In contrast to this case, an explicit time dependence directly determines the time dependence along the trajectory.

Autonomous Hamiltonian systems $H(q_i\,, p_i)$ with no explicit time dependence are conservative systems, because H does not change with time (i.e., along the trajectory).

2.11.3 Flow in Phase Space

In general, we consider a system with n degrees of freedom. In this case, we have n generalized coordinates q_i and n generalized momentum variables p_i $(i \in \{1, 2, \ldots, n\})$.

The $2n$-dimensional space that is generated by these variables is called the phase space. If the $2n$ variables q_i and p_i are given at a time t_0, the system state is determined completely, and $q_i(t)$, $p_i(t)$ can be calculated for arbitrary times t (in the maximal interval of existence; see Sects. 2.8.3.3 and 2.8.3.4).

In order to show this, we combine coordinate and momentum variables as follows:

$$\vec{r} = \begin{pmatrix} q_1 \\ q_2 \\ \dots \\ q_n \\ p_1 \\ p_2 \\ \dots \\ p_n \end{pmatrix}, \qquad \vec{v} = \dot{\vec{r}} = \begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dots \\ \dot{q}_n \\ \dot{p}_1 \\ \dot{p}_2 \\ \dots \\ \dot{p}_n \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial p_1} \\ \frac{\partial H}{\partial p_2} \\ \dots \\ \frac{\partial H}{\partial p_n} \\ -\frac{\partial H}{\partial q_1} \\ -\frac{\partial H}{\partial q_2} \\ \dots \\ -\frac{\partial H}{\partial q_n} \end{pmatrix}.$$

In the last step, we used Hamilton's equations (2.122) and (2.123). Based on this definition, the problem has the standard form (2.94) of a dynamical system (see p. 50). We obtain

$$\boxed{\operatorname{div} \vec{v} = \sum_{k=1}^{n} \left(\frac{\partial^2 H}{\partial p_k \partial q_k} - \frac{\partial^2 H}{\partial q_k \partial p_k} \right) = 0.}$$

Therefore, the flow in phase space corresponds to an incompressible fluid. Thus, Liouville's theorem is valid automatically, stating that the area/volume in phase space remains constant. We have assumed only the preservation of the number of particles and the validity of Hamilton's equations.

Liouville's theorem (and area/volume preservation) is also valid if the Hamiltonian

$$H(q, p, t)$$

explicitly depends on time (cf. Szebehely [46, p. 55], Lichtenberg [49, p. 13]).

2.11.4 *Fixed Points of a Hamiltonian System in the Plane*

For the fixed points of an autonomous Hamiltonian system with one degree of freedom, we have

$$\vec{r} = \begin{pmatrix} q \\ p \end{pmatrix}, \qquad \vec{v} = \begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial p} \\ -\frac{\partial H}{\partial q} \end{pmatrix} = 0.$$

The Jacobian matrix is

$$D\vec{v} = \begin{pmatrix} \frac{\partial^2 H}{\partial p \partial q} & \frac{\partial^2 H}{\partial p^2} \\ -\frac{\partial^2 H}{\partial q^2} & -\frac{\partial^2 H}{\partial q \partial p} \end{pmatrix}.$$

If we calculate the eigenvalues of the matrix

$$A = D\vec{v}(\vec{r}_{\mathrm{F}}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{2.124}$$

according to Eqs. (2.107) and (2.108), we obtain

$$B = \frac{a_{11} + a_{22}}{2} = 0, \qquad C = -\det\ A,$$

and therefore

$$\lambda = \pm\sqrt{C} = \pm\sqrt{-\det\ A}.$$

Hence, the two eigenvalues are either real with opposite sign or imaginary with opposite sign.

All fixed points of the linearized system are therefore either centers or saddle points. The linearized system cannot have any sources or sinks. This is consistent with the name "conservative system."

We now regard the Hamiltonian $H(q, p)$ as a function that describes a two-dimensional surface in three-dimensional space.

The fixed-point condition

$$\begin{pmatrix} \frac{\partial H}{\partial p} \\ -\frac{\partial H}{\partial q} \end{pmatrix} = 0$$

is necessary for the existence of a **relative extremum** (also called a local extremum) of $H(\vec{r})$ at $\vec{r} = \vec{r}_{\mathrm{F}}$, because the gradient of H must be zero. A sufficient condition for a **relative minimum** is that the **Hessian matrix**

$$\begin{pmatrix} \frac{\partial^2 H}{\partial q^2} & \frac{\partial^2 H}{\partial q \partial p} \\ \frac{\partial^2 H}{\partial p \partial q} & \frac{\partial^2 H}{\partial p^2} \end{pmatrix}$$

of H be positive definite at $\vec{r} = \vec{r}_{\mathrm{F}}$ (all eigenvalues positive). If the Hessian matrix is negative definite (all eigenvalues negative), then a **relative maximum** is present. If the Hessian matrix is indefinite (both positive and negative eigenvalues), a **saddle point** is present.

Obviously, we find for the Hessian matrix

$$\begin{pmatrix} \frac{\partial^2 H}{\partial q^2} & \frac{\partial^2 H}{\partial q \partial p} \\ \frac{\partial^2 H}{\partial p \partial q} & \frac{\partial^2 H}{\partial p^2} \end{pmatrix} = \begin{pmatrix} -a_{21} & a_{11} \\ a_{11} & a_{12} \end{pmatrix}.$$

The eigenvalues λ_H of the Hessian matrix can be determined as follows:

$$\begin{aligned}
&(-a_{21} - \lambda_H)(a_{12} - \lambda_H) - a_{11}^2 = 0 \\
&\Rightarrow \lambda_H^2 + \lambda_H(a_{21} - a_{12}) - (a_{11}^2 + a_{12}a_{21}) = 0 \\
&\Rightarrow \lambda_H = \frac{a_{12} - a_{21}}{2} \pm \sqrt{\frac{(a_{12} - a_{21})^2}{4} + a_{11}^2 + a_{12}a_{21}} \qquad (2.125) \\
&\Rightarrow \lambda_H = \frac{a_{12} - a_{21}}{2} \pm \sqrt{\frac{(a_{12} + a_{21})^2}{4} + a_{11}^2}.
\end{aligned}$$

The argument of the square root is not negative, so that only real eigenvalues exist (symmetry of the Hessian matrix).

Hence, we have three possibilities for the value of the square root:

- It is greater than the absolute value of the first fraction. In this case, it determines the sign of the eigenvalues. Therefore, a positive eigenvalue and a negative eigenvalue exist, and the Hessian matrix is indefinite. Hence, we have a (geometric) saddle point. In this case, due to Eq. (2.125), we have

$$a_{11}^2 + a_{12}a_{21} > 0,$$

  or with $a_{11} = -a_{22}$ (see Eq. (2.124)),

$$\begin{aligned}
&a_{11}a_{22} - a_{12}a_{21} < 0 \\
&\Leftrightarrow \det\ A < 0.
\end{aligned}$$

  Due to the restriction

$$\lambda = \pm\sqrt{C} = \pm\sqrt{-\det\ A}$$

  for the eigenvalues of the Jacobian matrix, the fixed point is also a saddle point.
- It is less than the absolute value of the first fraction in Eq. (2.125), so that $\det\ A > 0$ holds. Due to

$$\lambda = \pm\sqrt{-\det\ A},$$

  the eigenvalues are imaginary, and the fixed point is a center. The first fraction in Eq. (2.125) decides which sign the eigenvalues of the Hessian matrix have. For $a_{12} > a_{21}$, we have a relative minimum of the Hamiltonian, and for $a_{12} < a_{21}$, one obtains a relative maximum.
- It equals the first fraction in Eq. (2.125). Then, one eigenvalue is zero, and $\det\ A = 0$ holds, which we have excluded (degenerate fixed point).

In conclusion, the Hamiltonian has a geometric saddle point if the corresponding fixed point is a saddle point. It has a relative minimum or maximum if the corresponding fixed point is a center.

2.11.5 Hamiltonian as Lyapunov Function

As in the previous sections, let us consider an autonomous Hamiltonian system with only one degree of freedom.

In Sect. 2.11.4, we saw that the linearization of such a system may have only centers and saddle points as fixed points. Let us assume that the system is linearized at a specific fixed point and that the fixed point of the linearized system is a saddle point. According to Theorem 2.17 (p. 77), the fixed point of the original system must be a saddle point as well. Theorem 2.17 applies, because the saddle point is a hyperbolic fixed point.

These arguments cannot be adopted for a center as a fixed point, because centers are not hyperbolic fixed points. If we want to show that a center of the linearized system corresponds to a center of the original nonlinear system, we need a different approach, which is presented in the following.

In many cases, the Hamiltonian of an autonomous system is defined in such a way that $H \geq 0$ holds and that for the fixed points, $H(\vec{r}_\mathrm{F}) = 0$ is valid. If under these conditions, $H(\vec{r})$ has a minimum at $\vec{r} = \vec{r}_\mathrm{F}$, then $L := H$ is a Lyapunov function, since one has

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \vec{v} \cdot \operatorname{grad} L = \begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} \cdot \operatorname{grad} H = \begin{pmatrix} \frac{\partial H}{\partial p} \\ -\frac{\partial H}{\partial q} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial H}{\partial q} \\ \frac{\partial H}{\partial p} \end{pmatrix} = 0 \leq 0.$$

Under these conditions, one is therefore able to show that the autonomous Hamiltonian system has a center.[19]

Theorem 2.19. *Let $H \in C^2(D)$ be a Hamiltonian ($D \subset \mathbb{R}^{2n}$ open). If $\vec{r}$ is an isolated minimum (strict minimum) of the Hamiltonian, then $\vec{r}$ is a stable fixed point.*

(See Amann [35, Sect. 18.11 b]; Walter [50, Sect. 30, Chap. XII d].)

Since for Hamiltonians, the question whether a minimum of maximum exists is just a matter of the sign,[20] one concludes in general the following result:

[19] This is not astonishing, since we obtained $\frac{\mathrm{d}L}{\mathrm{d}t} = 0$, so that L and H are constant on the trajectory.

[20] The reader may verify that if $H(q, p)$ is a Hamiltonian, then $\tilde{H} = -H$ with $\tilde{q} = p$ and $\tilde{p} = q$ is also a Hamiltonian. Furthermore, one can easily see that adding a constant to a Hamiltonian does not modify Hamilton's equations. Therefore, every relative maximum or minimum of a Hamiltonian at $\vec{r} = \vec{r}_\mathrm{F}$ may be transformed into a relative minimum with $H(\vec{r}_\mathrm{F}) = 0$.

Theorem 2.20. *Every nondegenerate fixed point $\vec{r}_F$ of a Hamiltonian system is a saddle point or a center. It is a saddle point if and only if the Hamiltonian has a saddle point with*

$$\det\left[D\vec{v}(\vec{r}_F)\right] < 0.$$

It is a center if and only if the Hamiltonian has a strict minimum or strict maximum with

$$\det\left[D\vec{v}(\vec{r}_F)\right] > 0.$$

(See Perko [30, Sect. 2.14, Theorem 2].)

2.11.6 Canonical Transformations

We consider **canonical transformations** as transformations that preserve the phase space area and that transform one set of Hamilton's equations (depending on q, p) into another set of Hamilton's equations (depending on Q, P).

According to Appendix A.5, preservation of the phase space area means

$$\xi = \frac{\partial(Q,P)}{\partial(q,p)} = \begin{vmatrix} \frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p} \\ \frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \end{vmatrix} = \frac{\partial Q}{\partial q}\frac{\partial P}{\partial p} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial q} = 1. \tag{2.126}$$

We consider only a very specific[21] subset of canonical transformations for which the value of the Hamiltonian remains unchanged. In this case,

$$\dot{q} = \frac{\partial H}{\partial p}, \qquad \dot{p} = -\frac{\partial H}{\partial q},$$

[21] A sophisticated theory of canonical transformations exists and is described in many textbooks on theoretical physics. The transformed generalized coordinates Q_i and momenta P_i ($i \in \{1, 2, \ldots, n\}$) may depend on the original generalized coordinates q_i and momenta p_i and on the time t:

$$\begin{aligned} Q_i &= Q_i(q_1, q_2, \ldots, q_n, p_1, p_2, \ldots, p_n, t), \\ P_i &= P_i(q_1, q_2, \ldots, q_n, p_1, p_2, \ldots, p_n, t). \end{aligned}$$

The Hamiltonian $H(q_1, q_2, \ldots, q_n, p_1, p_2, \ldots, p_n, t)$ is transformed into the Hamiltonian $K(Q_1, Q_2, \ldots, Q_n, P_1, P_2, \ldots, P_n, t)$. Canonical transformations can be constructed using four basic types of **generating function** $F_1(q, Q, t)$, $F_2(q, P, t)$, $F_3(p, Q, t)$, $F_4(p, P, t)$. This theory is outside the scope of this book. An introduction may be found, for example, in [19].

must be transformed into

$$\dot{Q} = \frac{\partial H}{\partial P}, \qquad \dot{P} = -\frac{\partial H}{\partial Q}.$$

For all points in phase space we have

$$\begin{aligned} \frac{\partial H}{\partial p} &= \frac{\partial H}{\partial Q}\frac{\partial Q}{\partial p} + \frac{\partial H}{\partial P}\frac{\partial P}{\partial p} \qquad &\Rightarrow \dot{q} = -\dot{P}\,\frac{\partial Q}{\partial p} + \dot{Q}\,\frac{\partial P}{\partial p}, \\ \frac{\partial H}{\partial q} &= \frac{\partial H}{\partial Q}\frac{\partial Q}{\partial q} + \frac{\partial H}{\partial P}\frac{\partial P}{\partial q} \qquad &\Rightarrow -\dot{p} = -\dot{P}\,\frac{\partial Q}{\partial q} + \dot{Q}\,\frac{\partial P}{\partial q}. \end{aligned}$$

Now we have to check whether these restricted transformations are actually canonical ones, i.e., whether $\xi = 1$ holds.

For this purpose, we eliminate all derivatives of P in Eq. (2.126) by means of the last two results:

$$\xi = \frac{\partial Q}{\partial q}\left(\frac{\dot{q}}{\dot{Q}} + \frac{\dot{P}}{\dot{Q}}\frac{\partial Q}{\partial p}\right) - \frac{\partial Q}{\partial p}\left(-\frac{\dot{p}}{\dot{Q}} + \frac{\dot{P}}{\dot{Q}}\frac{\partial Q}{\partial q}\right) = \frac{1}{\dot{Q}}\left(\frac{\partial Q}{\partial q}\dot{q} + \frac{\partial Q}{\partial p}\dot{p}\right).$$

The last expression in parentheses is equal to $\dot{Q}$, so that $\xi = 1$ indeed holds.

2.11.7 Action-Angle Variables

In this section, we will briefly discuss special coordinates for oscillatory Hamiltonian systems, the so-called **action-angle variables**. Again, the general theory is outside the scope of this book, but we will use some results of this theory to determine the oscillation frequency of nonlinear systems.

2.11.7.1 Introductory Example

As an introductory example, we consider a parallel LC circuit as shown in Fig. 2.16 for which

$$V = L\dot{I} \qquad \Rightarrow \dot{I} = \frac{V}{L}$$

and

$$-I = C\dot{V} \qquad \Rightarrow \dot{V} = -\frac{I}{C}$$

hold. For

$$q = I, \qquad p = V,$$

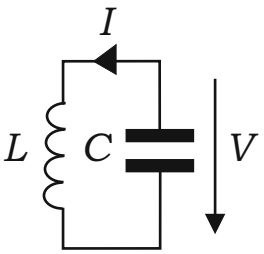

Fig. 2.16 Parallel LC circuit

this may be transformed into Hamilton's equations:

$$\dot{q} = \frac{p}{L} = \frac{\partial H}{\partial p}, \qquad \dot{p} = -\frac{q}{C} = -\frac{\partial H}{\partial q}.$$

By means of an integration, we obtain the Hamiltonian

$$H = \frac{p^2}{2L} + \frac{q^2}{2C} = \frac{V^2}{2L} + \frac{I^2}{2C}.$$

For the initial conditions

$$V = V_{\max}, \qquad I = 0,$$

one obtains

$$H = \frac{V_{\max}^2}{2L}.$$

This value of the Hamiltonian is preserved, so that

$$I^2 = 2CH - \frac{C}{L}V^2 = \frac{C}{L}(V_{\max}^2 - V^2)$$

is valid. Hence, the orbit in phase space is an ellipse with semiaxes $V_{\max}$ and

$$I_{\max} = \sqrt{\frac{C}{L}}V_{\max}.$$

For the area enclosed by this orbit, one obtains

$$A = \pi V_{\max} I_{\max} = \pi\sqrt{\frac{C}{L}}V_{\max}^2 = 2\pi\sqrt{LC}\,H.$$

Since we know that the resonant angular frequency of a parallel LC circuit is

$$\omega = \omega_{\text{res}} := \frac{1}{\sqrt{LC}},$$

we see at once that

$$A = \frac{2\pi}{\omega} H = TH,$$

where T is the period of the oscillation. One may therefore guess that the resonant frequency or period may be derived from the area enclosed by the orbit even if less-trivial examples are considered. If that works (and it does, as we will see soon), it will obviously not be necessary to actually solve the differential equation.

2.11.7.2 Basic Principle

Let us consider an autonomous Hamiltonian system with one degree of freedom. We assume that in the (q, p) phase space, a center exists such that closed orbits are present. We are now looking for a specific canonical transformation that introduces the new generalized coordinate/momentum pair (Q, P).

The idea of action-angle variables is to require that one of the transformed coordinates not depend on time:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \dot{P} = 0.$$

Hamilton's equations

$$\dot{q} = \frac{\partial H}{\partial p}, \qquad \dot{p} = -\frac{\partial H}{\partial q},$$
$$\dot{Q} = \frac{\partial H}{\partial P}, \qquad \dot{P} = -\frac{\partial H}{\partial Q},$$

then show that

$$\frac{\partial H}{\partial Q} = 0$$

is valid, so that H cannot depend on Q but only on P:

$$H = H(P).$$

Therefore, $\frac{\partial H}{\partial P}$ also depends only on P. Furthermore, P is constant with respect to time, so that

$$\frac{\partial H}{\partial P} = \frac{\mathrm{d}H}{\mathrm{d}P}$$

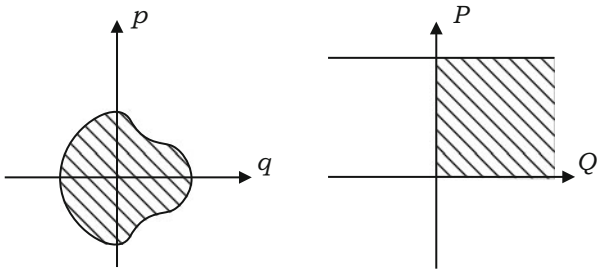

Fig. 2.17 Transition to action-angle variables

cannot depend on time either. Due to Hamilton's equation, one then obtains

$$\dot{Q} = \frac{\mathrm{d}H}{\mathrm{d}P} = K(P),$$

and therefore

$$Q(t) = Q(0) + K(P)\, t.$$

In the original phase space (q, p), one revolution lasted for time T. Hence, it is clear that in the transformed phase space (Q, P), the variable Q will increase by the amount

$$\Delta Q = K(P)\, T, \tag{2.127}$$

while P remains constant. This is visualized in Fig. 2.17.

Since the area in phase space is kept constant by a canonical transformation, the (Q, P) phase space is a surface of a cylinder (cf. Percival/Richards [51, p. 105]). This indicates why generalized coordinates are called **cyclic** if the Hamiltonian does not depend on them. In our case, Q is a cyclic coordinate.

Heretofore, we required only that P not depend on time. This is satisfied, for example, if to every point (q, p), we assign the area A that is enclosed by the orbit that goes through the point (q, p):

$$P = A = \iint_A \mathrm{d}q\, \mathrm{d}p.$$

If this definition for P is used, then shaded area in the (q, p) phase space in Fig. 2.17 is equal to P. The shaded area in the (Q, P) phase space is equal to $P\ \Delta Q$ (see the right-hand diagram in Fig. 2.17). Since both areas must be equal, we obtain

$$\Delta Q = 1,$$

and therefore, based on Eq. (2.127),

$$K(P)\ T = 1$$
$$\Leftrightarrow K(P) = \frac{\mathrm{d}H}{\mathrm{d}P} = \frac{1}{T}.$$

In conclusion, we may use H to calculate the period of the oscillation directly without solving the differential equation explicitly.

Instead of taking the area $A = P$ directly as a generalized coordinate, one defines the **action variable**

$$J = \frac{P}{2\pi}$$
$$\Leftrightarrow \boxed{J = \frac{1}{2\pi} \iint_A \mathrm{d}q\ \mathrm{d}p}$$

and the **angle variable**[22]

$$\theta = 2\pi\ Q.$$

As the name implies, the angle variable obviously increases by 2π during every period of the oscillation. Hence, one obtains

$$\frac{\mathrm{d}H}{\mathrm{d}J} = \frac{\mathrm{d}H}{\mathrm{d}P}\ 2\pi = \frac{2\pi}{T}$$
$$\Rightarrow \boxed{\frac{\mathrm{d}H}{\mathrm{d}J} = \frac{2\pi}{T} = \omega}$$

for Hamilton's equations

$$\boxed{\dot{\theta} = \frac{\partial H}{\partial J}, \qquad \dot{J} = -\frac{\partial H}{\partial \theta} = 0.}$$

Please note that for these considerations, we assumed that the Hamiltonian does not depend on time and that the orbits are closed. Therefore, the action variable is

[22] It is easy to show that multiplying the generalized coordinate by a constant and dividing the generalized momentum by that same constant is a canonical transformation (cf. Sect. 5.1.3).

defined in a unique way, and by Liouville's theorem, it is obvious that the phase space area, and thus also the action variable, remains constant.

2.11.8 LC Circuit with Nonlinear Inductance

The characteristic curve $B(H)$ of a magnetic material can be approximated by

$$B = B_{\max} \frac{2}{\pi} \arctan \frac{H}{H_0} = B_{\max} \frac{2}{\pi} \arctan \frac{I}{I_0}.$$

The magnetic material will be used to build an inductor with N windings. With the magnetic flux $\Phi_{\mathrm{m}} = BA$, it follows that

$$V = N \frac{\mathrm{d}\Phi_{\mathrm{m}}}{\mathrm{d}t} = N\, A B_{\max} \frac{2}{\pi} \frac{1}{1 + \left(\frac{I}{I_0}\right)^2} \frac{1}{I_0} \frac{\mathrm{d}I}{\mathrm{d}t}.$$

Therefore, from

$$V = L \frac{\mathrm{d}I}{\mathrm{d}t},$$

one obtains

$$L(I) = \frac{L_0}{1 + \left(\frac{I}{I_0}\right)^2}.$$

The corresponding inductor in parallel with a capacitor can now be used to form an LC oscillator as shown in Fig. 2.16. For the capacitance of the oscillating circuit,

$$-I = C \frac{\mathrm{d}V}{\mathrm{d}t}.$$

is valid. The magnetic energy is

$$W_{\mathrm{magn}} = \int V\, I\, \mathrm{d}t = \int L(I)\, I\, \frac{\mathrm{d}I}{\mathrm{d}t}\, \mathrm{d}t = L_0 \int \frac{I}{1 + \left(\frac{I}{I_0}\right)^2} \frac{\mathrm{d}I}{\mathrm{d}t}\, \mathrm{d}t$$

$$= L_0 I_0^2 \int \frac{x}{1 + x^2} \frac{\mathrm{d}x}{\mathrm{d}t}\, \mathrm{d}t,$$

where we have used

$$x = \frac{I}{I_0}, \qquad \text{or} \qquad I = x\, I_0.$$

Because of

$$\int \frac{x}{1+x^2}\,\mathrm{d}x = \frac{1}{2}\,\ln|1+x^2| + \text{const},$$

one obtains

$$W_{\text{magn}} = \frac{L_0 I_0^2}{2}\,\ln\left(1+\left(\frac{I}{I_0}\right)^2\right).$$

Together with

$$W_{\text{el}} = \frac{1}{2}CV^2,$$

this leads to

$$W(V,I) = W_{\text{el}} + W_{\text{magn}} = \frac{1}{2}CV^2 + \frac{L_0 I_0^2}{2}\,\ln\left(1+\left(\frac{I}{I_0}\right)^2\right).$$

If we define $p = V$, we obtain

$$\frac{\partial W}{\partial p} = Cp = CV = CL(I)\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{CL_0}{1+\left(\frac{I}{I_0}\right)^2}\frac{\mathrm{d}I}{\mathrm{d}t}.$$

If this is one of the two Hamilton's equations, the right-hand side must be equal to $\dot{q}$, and one obtains

$$q = CL_0 I_0\,\arctan\frac{I}{I_0}.$$

Therefore, the Hamiltonian is

$$H(q,p) = \frac{1}{2}Cp^2 + \frac{L_0 I_0^2}{2}\,\ln\left(1+\tan^2\frac{q}{CL_0 I_0}\right)$$

$$\Rightarrow H(q,p) = \frac{1}{2}Cp^2 - L_0 I_0^2\,\ln\,\cos\frac{q}{CL_0 I_0}.$$

We still have to check the second of Hamilton's equations. One obtains

$$\frac{\partial H}{\partial q} = -L_0 I_0^2\,\frac{1}{\cos\frac{q}{CL_0 I_0}}\left(-\sin\frac{q}{CL_0 I_0}\right)\frac{1}{CL_0 I_0} = \frac{I_0}{C}\,\tan\frac{q}{CL_0 I_0} = \frac{I}{C}.$$

In fact, the right-hand side equals $-\frac{\mathrm{d}V}{\mathrm{d}t}$, which is equal to $-\dot{p}$, and both Hamilton's equations are satisfied.

In order to calculate the oscillation frequency, we compute the action:

$$\begin{aligned} J(H) &= \frac{1}{2\pi} \iint \mathrm{d}q \, \mathrm{d}p \\ &= \frac{1}{\pi} \int_{q_1}^{q_2} p \, \mathrm{d}q \\ &= \frac{1}{\pi} \sqrt{\frac{2}{C}} \int_{q_1}^{q_2} \sqrt{H + L_0 I_0^2 \, \ln \, \cos \frac{q}{C L_0 I_0}} \, \mathrm{d}q. \end{aligned}$$

The limits q_1 and q_2 are determined by the zeros of p where the trajectory crosses the q-axis. The substitution

$$x = \frac{q}{C L_0 I_0}, \qquad \frac{\mathrm{d}x}{\mathrm{d}q} = \frac{1}{C L_0 I_0},$$

leads to

$$J(H) = \frac{1}{\pi} \sqrt{\frac{2}{C}} C L_0 I_0 \int_{x_1}^{x_2} \sqrt{H + L_0 I_0^2 \, \ln \, \cos x} \, \mathrm{d}x. \tag{2.128}$$

Due to

$$\ln \, \cos x \approx -\frac{x^2}{2} - \frac{x^4}{12} - \cdots$$

the simplest approximation for $I \ll I_0$ is

$$\begin{aligned} J(H) &= \frac{1}{\pi} \sqrt{2C} L_0 I_0 \sqrt{\frac{L_0}{2}} I_0 \int_{x_1}^{x_2} \sqrt{H \frac{2}{L_0 I_0^2} - x^2} \, \mathrm{d}x \\ &= \frac{1}{\pi} \sqrt{L_0 C} L_0 I_0^2 \int_{x_1}^{x_2} \sqrt{H \frac{2}{L_0 I_0^2} - x^2} \, \mathrm{d}x. \end{aligned}$$

The integral describes the area of a semicircle with radius

$$\sqrt{\frac{2H}{L_0 I_0^2}},$$

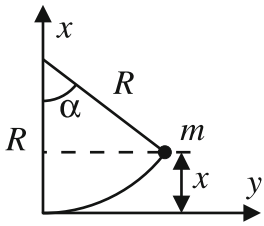

Fig. 2.18 Mathematical pendulum

so that

$$J(H) = \frac{1}{\pi}\sqrt{L_0 C}\; L_0 I_0^2 \frac{2H}{L_0 I_0^2}\frac{\pi}{2} = \sqrt{L_0 C}\; H$$

is obtained. As expected, one obtains

$$\omega = \frac{\mathrm{d}H}{\mathrm{d}J} = \frac{1}{\sqrt{L_0 C}}.$$

If the approximation is undesirable, one may directly calculate the derivative of Eq. (2.128). Then the integral may be evaluated numerically in order to calculate the amplitude-dependent oscillation frequency $\omega(\hat{I})$. As mentioned above, no direct solution of the differential equation is required.

2.11.9 Mathematical Pendulum

Consider a mathematical pendulum with mass m that is suspended by means of a massless cord of length R (see Fig. 2.18). Suppose that initially, the mass m is at height $x = h$ (corresponding to the angle $\alpha = \hat{\alpha}$) with zero velocity.

2.11.9.1 Energy Balance

The sum of the potential energy and kinetic energy must remain constant:

$$W_{\mathrm{pot}} + W_{\mathrm{kin}} = \mathrm{const}, \qquad \text{with} \quad W_{\mathrm{pot}} = mgx \qquad \text{and} \quad W_{\mathrm{kin}} = \frac{1}{2}mu^2$$

$$\Leftrightarrow\; mgx + \frac{1}{2}mu^2 = \mathrm{const}$$

$$\Leftrightarrow\; g(R - R\,\cos\alpha) + \frac{1}{2}R^2\dot{\alpha}^2 = \mathrm{const}.$$

We now calculate the time derivative of this equation:

$$gR\ \sin\alpha\ \dot{\alpha} + R^2\dot{\alpha}\ddot{\alpha} = 0.$$

As a result, we obtain

$$\ddot{\alpha} + \frac{g}{R}\ \sin\alpha = 0. \tag{2.129}$$

2.11.9.2 Hamilton's Equations

We now try to convert Eq. (2.129) into a pair of Hamilton's equations using our standard approach

$$q = \alpha, \qquad p = \dot{\alpha},$$

which leads to

$$\dot{q} = p,$$
$$\dot{p} = -\frac{g}{R}\ \sin q.$$

If this is to be in accord with Hamilton's equations,

$$\frac{\partial H}{\partial p} = \dot{q},$$
$$\frac{\partial H}{\partial q} = -\dot{p},$$

we obtain by integration

$$H = \frac{p^2}{2} + f(q),$$
$$H = -\frac{g}{R}\ \cos q + g(p).$$

Putting both results together, one obtains

$$H(q, p) = \frac{p^2}{2} - \frac{g}{R}\ \cos q. \tag{2.130}$$

If we additionally require $H(0,0) = 0$, we may add a constant accordingly:

$$H(q,p) = \frac{p^2}{2} + \frac{g}{R}\,(1 - \cos q). \tag{2.131}$$

2.11.9.3 Oscillation Period

In order to calculate the oscillation period, we first determine the action variable:

$$J(H) = \frac{1}{2\pi} \iint \mathrm{d}q\,\mathrm{d}p = \frac{1}{\pi} \int_{q_1}^{q_2} p\,\mathrm{d}q.$$

Equation (2.130) leads to

$$p = \sqrt{2\left[H + \frac{g}{R}\,\cos q\right]}$$

for the upper part of the curve in phase space. By means of

$$a = \frac{2}{\pi^2} H,$$

$$b = \frac{2}{\pi^2} \frac{g}{R}, \tag{2.132}$$

$$\frac{a}{b} = \frac{HR}{g} \tag{2.133}$$

one obtains the following integral:

$$J(H) = \int_{q_1}^{q_2} \sqrt{a + b\,\cos q}\,\mathrm{d}q. \tag{2.134}$$

The limits q_1 and q_2 of integration are determined by the zeros of p. We obviously have $p = 0$ for

$$q_{1,2} = \mp \arccos \frac{-a}{b}.$$

Therefore, $q_1 = -q_2$ holds, so that we can use the symmetry of the integrand:

$$J(H) = 2 \int_0^{q_2} \sqrt{a + b\,\cos q}\,\mathrm{d}q.$$

According to the first formula 2.576 in [3], for $|a| \le b$ and $0 \le q < \arccos(-a/b)$, the integral has the following value:[23]

$$J(H) = 2\sqrt{\frac{2}{b}}\left[(a-b)\,\mathrm{F}\left(\gamma, \frac{1}{r}\right) + 2b\,\mathrm{E}\left(\gamma, \frac{1}{r}\right)\right]_0^{q_2} \tag{2.135}$$

with

$$r = \sqrt{\frac{2b}{a+b}} \qquad \text{and} \qquad \gamma = \arcsin\sqrt{\frac{b(1-\cos q)}{a+b}}.$$

For $q = 0$, one obviously has $\gamma = 0$. For $q = q_2$,

$$\cos q = \frac{-a}{b} \qquad \Rightarrow \gamma = \arcsin\sqrt{\frac{b\left(1+\frac{a}{b}\right)}{a+b}} = \arcsin\ 1 = \frac{\pi}{2}$$

is valid. The expression in square brackets in Eq. (2.135) is equal to zero at the lower integration limit $\gamma = 0$, since $\mathrm{F}(0,k) = 0$ and $\mathrm{E}(0,k) = 0$:

$$(a-b)\mathrm{F}\,(0,k) + 2b\ \mathrm{E}\,(0,k) = 0.$$

Here we set

$$k = \frac{1}{r} = \sqrt{\frac{1}{2} + \frac{a}{2b}} \qquad \text{and} \qquad k' = \sqrt{1-k^2} = \sqrt{\frac{1}{2} - \frac{a}{2b}}. \tag{2.136}$$

From $\mathrm{F}(\pi/2,k) = \mathrm{K}(k)$ and $\mathrm{E}(\pi/2,k) = \mathrm{E}(k)$, one concludes, based on Eq. (2.135), that

$$J(H) = 2\sqrt{\frac{2}{b}}\left[(a-b)\ \mathrm{K}(k) + 2b\ \mathrm{E}(k)\right] = 4\sqrt{2b}\left[\mathrm{E}(k) - k'^2\mathrm{K}(k)\right]. \tag{2.137}$$

---

[23] Here $\mathrm{F}(\gamma,k)$ denotes the **elliptic integral of the first kind**, whereas $\mathrm{E}(\gamma,k)$ is an **elliptic integral of the second kind** [52]:

$$\mathrm{F}(\gamma,k) = \int_0^\gamma \frac{\mathrm{d}\theta}{\sqrt{1-k^2\ \sin^2\theta}} \qquad \mathrm{E}(\gamma,k) = \int_0^\gamma \sqrt{1-k^2\ \sin^2\theta}\ \mathrm{d}\theta.$$

The **complete elliptic integral of the first kind** is defined by $\mathrm{K}(k) = \mathrm{F}(\pi/2,k)$, while the **complete elliptic integral of the second kind** is given by $\mathrm{E}(k) = \mathrm{E}(\pi/2,k)$. In this book, we make use of only the **modulus** k ($0 \le k \le 1$). Alternatively, one can also use the **parameter** m or the **modular angle** α:

$$k = \sin\ \alpha, \qquad m = k^2.$$

The **complementary modulus** k' is given by $k^2 + k'^2 = 1$, and the **complementary parameter** $m_1 = k'^2$ is therefore defined by $m + m_1 = 1$.

The angular frequency of the oscillation may be calculated according to

$$\omega = \frac{\partial H}{\partial J}.$$

Since by definition, H depends only on the action variable J but not on the angle variable θ, and since H does not depend on time in our case, the partial derivative is in fact a total derivative:

$$\omega = \frac{\partial H}{\partial J} = \frac{\mathrm{d}H}{\mathrm{d}J} = \left(\frac{\mathrm{d}J}{\mathrm{d}H}\right)^{-1}.$$

We therefore need $\frac{\mathrm{d}J}{\mathrm{d}H}$. From Eqs. (2.133), (2.136), and (2.137), we get

$$\frac{\mathrm{d}J}{\mathrm{d}H} = \frac{\mathrm{d}J}{\mathrm{d}k}\frac{\mathrm{d}k}{\mathrm{d}H} = 4\sqrt{2b}\left[\frac{\mathrm{dE}(k)}{\mathrm{d}k} - 2k'\frac{-2k}{2k'}\,\mathrm{K}(k) - (1-k^2)\frac{\mathrm{dK}(k)}{\mathrm{d}k}\right]\frac{1}{2k}\frac{R}{2g} =$$

$$= \frac{2}{\pi k}\sqrt{\frac{R}{g}}\left[\frac{\mathrm{dE}(k)}{\mathrm{d}k} + 2k\,\mathrm{K}(k) - (1-k^2)\frac{\mathrm{dK}(k)}{\mathrm{d}k}\right].$$

In the last step, we made use of Eq. (2.132), which led to

$$\sqrt{2b} = \frac{2}{\pi}\sqrt{\frac{g}{R}}.$$

With

$$\boxed{\frac{\mathrm{dK}(k)}{\mathrm{d}k} = \frac{\mathrm{E}(k)}{kk'^2} - \frac{\mathrm{K}(k)}{k}}$$

and

$$\boxed{\frac{\mathrm{dE}(k)}{\mathrm{d}k} = \frac{\mathrm{E}(k) - \mathrm{K}(k)}{k},}$$

we obtain

$$\frac{\mathrm{d}J}{\mathrm{d}H} = \frac{2}{\pi k}\sqrt{\frac{R}{g}}\,\frac{\mathrm{E}(k) - \mathrm{K}(k) + 2k^2\,\mathrm{K}(k) - \mathrm{E}(k) + k'^2\mathrm{K}(k)}{k} = \frac{2}{\pi}\sqrt{\frac{R}{g}}\,\mathrm{K}(k).$$

This finally leads to

$$\omega = \frac{\mathrm{d}H}{\mathrm{d}J} = \frac{\pi}{2\mathrm{K}(k)}\sqrt{\frac{g}{R}}. \tag{2.138}$$

The calculation presented here can be simplified significantly if the derivative with respect to H is determined before the integral is evaluated. This is done in Sect. 3.16 for an analogous problem.

Now our considerations are complete in principle. Only the geometric meaning of the modulus k remains to be clarified.

From Eq. (2.136), one obtains

$$k = \sqrt{\frac{1}{2} + \frac{HR}{2g}}.$$

Initially, the mass is momentarily at rest, so that we have $p = \dot{\alpha} = 0$. Therefore, according to Eq. (2.130),

$$H = -\frac{g}{R} \cos\hat{\alpha}$$

is the value of the Hamiltonian (which remains constant). This leads to

$$k = \sqrt{\frac{1 - \cos\hat{\alpha}}{2}}.$$

Since

$$\cos\hat{\alpha} = \cos^2\frac{\hat{\alpha}}{2} - \sin^2\frac{\hat{\alpha}}{2} = 1 - 2\,\sin^2\frac{\hat{\alpha}}{2},$$

this may be written in the form

$$k = \sin\frac{\hat{\alpha}}{2}.$$

2.11.10 Vlasov Equation

From the formula

$$\operatorname{div}(\rho\,\vec{v}) = \vec{v} \cdot \operatorname{grad}\rho + \rho\,\operatorname{div}\vec{v},$$

which is known from vector analysis, the continuity equation (2.115) for incompressible flows leads to the differential equation

$$-\dot{\rho} = \vec{v} \cdot \operatorname{grad}\rho.$$

If we consider a Hamiltonian system with one degree of freedom that describes the incompressible flow, we have

$$\vec{r} = \begin{pmatrix} q \\ p \end{pmatrix}, \qquad \vec{v} = \dot{\vec{r}} = \begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix}.$$

Therefore, one obtains

$$\frac{\partial \rho}{\partial t} + \dot{q}\,\frac{\partial \rho}{\partial q} + \dot{p}\,\frac{\partial \rho}{\partial p} = 0$$

$$\Rightarrow \frac{\partial \rho}{\partial t} + \frac{\partial H}{\partial p}\,\frac{\partial \rho}{\partial q} - \frac{\partial H}{\partial q}\,\frac{\partial \rho}{\partial p} = 0.$$

This is the **Vlasov equation**. It describes how the particle density ρ at different locations changes with time.

2.11.11 Outlook

A dynamical system is called **conservative** if the total energy (or the area in phase space) remains constant.

Every autonomous Hamiltonian system is conservative. However, there exist non-Hamiltonian systems that are conservative.

A function $I(q_k, p_k)$ that does not depend on t and that does not change its value on the trajectory is called a **constant of the motion**. Such a constant of the motion allows one to reduce the order of the problem by 1 (cf. Tabor [53, p. 2]), since one may express one variable in terms of the other variables by means of this function.

For a non-Hamiltonian system of order n, one therefore needs $n - 1$ constants of the motion in order to completely solve the differential equation by means of quadratures (cf. Tabor [53, p. 39]).

A Hamiltonian system is called **integrable** if the solution can be determined by quadratures (cf. Rebhan [22, vol. I, p. 287]). This is the case if the problem can be written in action-angle variables.

In contrast to non-Hamiltonian systems, one needs only n constants of the motion if a Hamiltonian system of order $2n$ with n degrees of freedom is considered (instead of $2n - 1$, as in the general case).

Conservative Hamiltonian systems with one degree of freedom (order 2) are integrable (cf. Rebhan [22, vol. I, p. 359]). This is obvious, because the Hamiltonian itself is a constant of the motion.

Chaotic behavior is possible only in nonintegrable systems (cf. Rebhan [22, vol. I, pp. 335 and 359]). Therefore, chaos is not possible in autonomous Hamiltonian systems with one degree of freedom. However, if more degrees of freedom are present, chaotic behavior may also occur in autonomous Hamiltonian systems.

References

1. A.H. Zemanian, *Distribution Theory and Transform Analysis* (Dover, New York, 1987)
2. K.B. Howell, *Principles of Fourier Analysis* (Chapman & Hall/CRC, Boca Raton/London/New York/Washington, 2001)
3. I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products*, 6th edn. (Academic, San Diego/San Francisco/New York/Boston/London/Sydney/Tokyo, 2000)
4. D. Müller-Wichards, *Transformationen und Signale* (B. G. Teubner, Stuttgart/Leipzig, 1999)
5. B. Davies, *Integral Transforms and Their Applications*, 3rd edn. (Springer, Berlin/Heidelberg/New York, 2002)
6. W.R. LePage, *Complex Variables and the Laplace Transform for Engineers* (McGraw-Hill, New York/Toronto/London, 1961)
7. B. Girod, R. Rabenstein, A. Stenger, *Einführung in die Systemtheorie*, 2. Auflage (B. G. Teubner, Stuttgart/Leipzig/Wiesbaden, 2003)
8. E. Kreyszig, *Statistische Methoden und ihre Anwendungen* (Vandenhoeck & Ruprecht, Göttingen, 1967)
9. A.W. Chao, M. Tigner, *Handbook of Accelerator Physics and Engineering*, 3rd edn. (World Scientific/New Jersey/London/Singapore/Beijing/Shanghai/Hong Kong/Taipei/Chennai, 2006)
10. M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New York/Chichester/Brisbane/Toronto/Singapore, 1994)
11. J.A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York/London, 1941)
12. R.E. Collin, *Field Theory of Guided Waves* (IEEE Press, New York, 1990)
13. J.D. Jackson, *Classical Electrodynamics*, 3rd edn. (Wiley, New York/Chichester/Weinheim/Brisbane/Singapore/Toronto, 1998)
14. J.D. Jackson, *Klassische Elektrodynamik*, 3. Auflage (Walter de Gruyter, Berlin/New York, 2002)
15. G. Lehner, *Electromagnetic Field Theory for Engineers and Physicists* (Springer, Heidelberg/Dordrecht/London/New York, 2010)
16. G. Lehner, *Elektromagnetische Feldtheorie für Ingenieure und Physiker*, 3. Auflage (Springer, Berlin/Heidelberg/New York/Barcelona/Budapest/Hongkong/London/Mailand/Paris/Santa Clara/Singapur/Tokio, 1996)
17. H. Klingbeil, *Elektromagnetische Feldtheorie. Ein Lehr- und Übungsbuch*, 2. Auflage (B. G. Teubner, Stuttgart/Leipzig/Wiesbaden, 2010)
18. W. Rindler, *Introduction to Special Relativity* (Clarendon Press, Oxford, 1982)
19. H. Goldstein, C. Poole, J. Safko, *Classical Mechanics*, 3rd edn. (Pearson Education, Addison-Wesley, June 25, 2001). ISBN-10: 0201657023, ISBN-13: 978-0201657029
20. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (W. H. Freeman and Company, San Francisco, 1973)
21. E. Schmutzer, *Grundlagen der Theoretischen Physik* (Deutscher Verlag der Wissenschaften, Berlin, 1989)
22. E. Rebhan, *Theoretische Physik*, Band 1, 1. Auflage (Spektrum Akademischer Verlag, Heidelberg/Berlin, 1999)
23. H. Goenner, *Einführung in die spezielle und allgemeine Relativitätstheorie*, 1. Auflage (Spektrum Akademischer Verlag, Heidelberg/Berlin/Oxford, 1996)
24. T. Fließbach, *Elektrodynamik* (Bibliographisches Institut & F. A. Brockhaus AG, Mannheim/Leipzig/Wien/Zürich, 1994)
25. T. Fließbach, *Allgemeine Relativitätstheorie*, 4. Auflage (Spektrum Akademischer Verlag, Heidelberg/Berlin, 2003)
26. W. Greiner, J. Rafelski, *Theoretische Physik, Band 3A: Spezielle Relativitätstheorie*, 2. Auflage (Verlag Harri Deutsch, Frankfurt am Main, 1989)
27. H. Mitter, *Mechanik* (Bibliographisches Institut & F. A. Brockhaus AG, Mannheim/Wien/Zürich, 1989)

28. H. Mitter, *Elektrodynamik*, 2. Auflage (Bibliographisches Institut & F. A. Brockhaus AG, Mannheim/Wien/Zürich, 1990)
29. B. Aulbach, *Gewöhnliche Differenzialgleichungen*, 2. Auflage (Spektrum, München, 2004)
30. L. Perko, *Differential Equations and Dynamical Systems* (Springer, New York/Berlin/Heidelberg, 2002)
31. G.J. Wirsching, *Gewöhnliche Differentialgleichungen* (B. G. Teubner Verlag/GWV Fachverlage GmbH, Wiesbaden, 2006)
32. V.I. Arnold, *Gewöhnliche Differentialgleichungen*, 2. Auflage (Springer, Berlin/Heidelberg/New York, 2001)
33. J. Guckenheimer, P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York/Berlin/Heidelberg/Tokyo, 1986)
34. E. Zehnder, *Lectures on Dynamical Systems. Hamiltonian Vector Fields and Symplectic Capacities* (European Mathematical Society, Zürich, 2010)
35. H. Amann, *Gewöhnliche Differentialgleichungen*, 2. Auflage (de Gruyter, Berlin/New York, 1995)
36. S. Sastry, *Nonlinear Systems: Analysis, Stability and Control*, 1st edn. (Springer, New York /Berlin/Heidelberg/Barcelona/Hong Kong/London/Milan/Paris/Singapore/Tokyo, 1999)
37. J. La Salle, S. Lefschetz, *Die Stabilitätstheorie von Ljapunov* (Bibliographisches Institut AG, Mannheim, 1967)
38. S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, 2nd edn. (Springer, New York/Berlin/Heidelberg/Hong Kong/London/Milan/Paris/Tokyo, 2003)
39. H. Amann, *Ordinary Differential Equations. An Introduction to Nonlinear Analysis* (Walter de Gruyter & Co., Berlin/New York, 1990)
40. K. Burg, H. Haf, F. Wille, *Höhere Mathematik für Ingenieure* (B. G. Teubner, Stuttgart, 2002) (Bd. I: 5. Aufl. 2001, Bd. II: 4. Aufl. 2002, Bd. III: 4. Aufl. 2002, Bd. IV: 2. Aufl. 1994, Bd. V: 2. Aufl. 1993)
41. E.A. Jackson, *Perspectives of Nonlinear Dynamics* (Cambridge University Press, Cambridge/New York/New Rochelle/Melbourne/Sydney, 1991)
42. W.E. Boyce, R.C. DiPrima, *Gewöhnliche Differentialgleichungen*, 1. Auflage (Spektrum Akademischer Verlag, Heidelberg/Berlin/Oxford, 1995)
43. W.E. Boyce, R.C. DiPrima, *Elementary Differential Equations*, 5th edn. (Wiley, New York, 1991)
44. I.N. Bronstein, K.A. Semendjajew, *Taschenbuch der Mathematik*, 24. Auflage (Verlag Harri Deutsch, Thun und Frankfurt am Main, 1989)
45. U. van Rienen, *Numerical Methods in Computational Electrodynamics. Linear Systems in Practical Applications* (Springer, Berlin/Heidelberg/New York/Barcelona/Hong Kong/London/Milan/Paris/Singapore/Tokyo, 2001)
46. V. Szebehely, *Theory of Orbits* (Academic, New York/San Francisco/London, 1967)
47. A.L. Fetter, J.D. Walecka, *Theoretical Mechanics of Particles and Continua* (McGraw-Hill, New York/St. Louis/San Francisco/Auckland/Bogotá/Hamburg/Johannesburg/London/Madrid/Mexico/Montreal/New Delhi/Panama/Paris/São Paulo/Singapore/Sydney/Tokyo/Toronto, 1980)
48. A. Budó, *Theoretische Mechanik*, 9. Auflage (Deutscher Verlag der Wissenschaften, Berlin, 1978)
49. A.J. Lichtenberg, M.A. Liebermann, *Regular and Stochastic Motion* (Springer, New York/Heidelberg/Berlin, 1983)
50. W. Walter, *Gewöhnliche Differentialgleichungen*, 7. Auflage (Springer, Berlin/Heidelberg/New York, 2000)
51. I. Percival, D. Richards, *Introduction to Dynamics* (Cambridge University Press, Cambridge/New York/Melbourne, 1982)
52. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965)
53. M. Tabor, *Chaos and Integrability in Nonlinear Dynamics* (Wiley, New York/Chichester/Brisbane/Toronto/Singapore, 1989)

Chapter 3
RF Acceleration

This chapter is devoted to the longitudinal motion of charged particles in a synchrotron.

3.1 Centripetal Force

For the derivation of the equations of motion in a synchrotron, we need the centripetal force. Therefore, we briefly show that the expression for the centripetal force is the same in special relativity as in classical mechanics.

For the momentum vector, we have

$$\vec{p} = m\,\vec{u} = m(\dot{x}\,\vec{e}_x + \dot{y}\,\vec{e}_y + \dot{z}\,\vec{e}_z),$$

where

$$m = m_0\gamma$$

is the velocity-dependent mass. We assume that wherever the centripetal force is active, the absolute value of the velocity will not be changed. Therefore, one obtains

$$\dot{u} = 0, \qquad \dot{\beta} = 0, \qquad \dot{\gamma} = 0, \qquad \dot{m} = 0.$$

Hence, only

$$\vec{F} = \dot{\vec{p}} = \dot{m}\,\vec{u} + m\,\dot{\vec{u}} = m(\ddot{x}\,\vec{e}_x + \ddot{y}\,\vec{e}_y + \ddot{z}\,\vec{e}_z)$$

has to be evaluated. On a circular orbit, we obtain

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6_3

$$
\begin{aligned}
x &= r\ \cos(\omega t), & u_x &= \dot{x} = -\omega r\ \sin(\omega t), & \dot{u}_x &= \ddot{x} = -\omega^2 r\ \cos(\omega t),\\
y &= r\ \sin(\omega t), & u_y &= \dot{y} = +\omega r\ \cos(\omega t), & \dot{u}_y &= \ddot{y} = -\omega^2 r\ \sin(\omega t),\\
z &= \text{const}, & u_z &= \dot{z} = 0, & \dot{u}_z &= \ddot{z} = 0.
\end{aligned}
$$

If we define the unit vector

$$\vec{e}_r = \vec{e}_x\ \cos(\omega t) + \vec{e}_y\ \sin(\omega t)$$

pointing radially outward, we can see directly the following relations:

$$\dot{\vec{u}} = -\omega^2 r\ \vec{e}_r,$$

$$u^2 = \vec{u} \cdot \vec{u} = \omega^2 r^2.$$

Hence we obtain

$$\vec{F} = -m\ \omega^2 r\ \vec{e}_r = -\frac{mu^2}{r}\ \vec{e}_r.$$

This is the well-known formula for the centripetal force, which is now verified in the scope of special relativity, provided that the energy of the particle remains constant (as satisfied in pure magnetic fields). In Appendix A.2.3, it is shown that this result remains true for arbitrary plane curves (e.g., if the magnetic field is no longer constant).

3.2 Simplified Model Synchrotron

In the scope of this book, we are interested only in longitudinal particle motion. Therefore, we significantly decrease the complexity of the problem by employing a model synchrotron that comprises only two straight sections and two dipole magnets. We have to emphasize that such a synchrotron will not work, because quadrupole magnets are essential for transverse focusing. If we keep this in mind, however, we may use the model nevertheless to study longitudinal motion in principle.

For the dipole magnets, we obtain

$$F = QuB = m\frac{u^2}{r_\mathrm{R}}$$

$$\Rightarrow \boxed{p = Qr_\mathrm{R}B.} \tag{3.1}$$

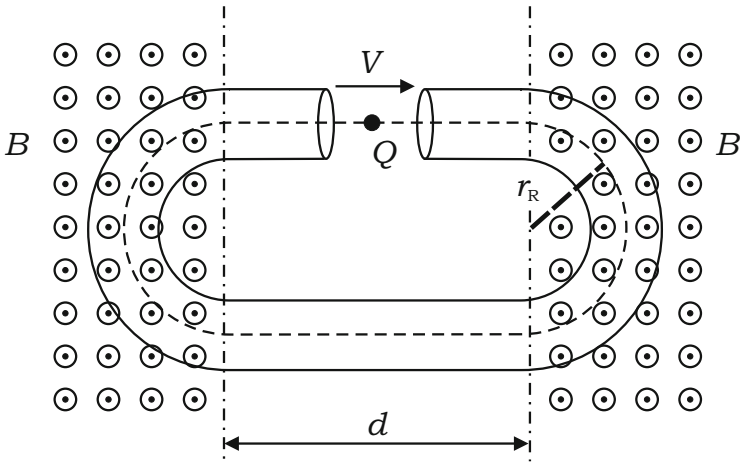

Fig. 3.1 Strongly simplified synchrotron

The product $r_R B$ is the **magnetic rigidity**. In one of the straight sections (the upper one in Fig. 3.1), we place a short ceramic gap. Let us assume that at $t = 0$, the reference particle is located at the gap. By means of the gap voltage, the energy[1] of the particle will increase from $\gamma_{R,0}$ to $\gamma_{R,1}$. Therefore, in the nth revolution, the particle has energy $\gamma_{R,n}$. Hereinafter, we will in general use the index R for quantities that are related to the reference particle.

By definition, the magnetic field is increased in such a way that the reference particle always remains in the same orbit ($r_R = \text{const}$).

For the reference particle, we therefore obtain

$$p_{R,n} = Q r_R B_n. \tag{3.2}$$

Here B_n is the magnetic dipole field, and $p_{R,n}$ is the momentum of the reference particle in the nth revolution. The reference particle will reach the beginning of the gap at time $t_{R,n}$ after the nth revolution ($t_{R,0} = 0$).

For the orbit length, one obtains

$$l_R = 2\pi r_R + 2d.$$

From

$$u_{R,n} = \frac{l_R}{T_{R,n}},$$

[1] Of course, the physical energy is obtained only after multiplying the Lorentz gammas by $m_0 c_0^2$, but we will also briefly refer to the gammas as energies.

it follows that

$$t_{\mathrm{R},n} = t_{\mathrm{R},n-1} + T_{\mathrm{R},n} = t_{\mathrm{R},n-1} + \frac{l_{\mathrm{R}}}{\beta_{\mathrm{R},n} c_0}. \tag{3.3}$$

Before the nth revolution takes place, the energy of the particle is increased from $\gamma_{\mathrm{R},n-1}$ to $\gamma_{\mathrm{R},n}$. This happens due to the voltage $V_{\mathrm{R},n-1}$ at the gap:

$$(\gamma_{\mathrm{R},n} - \gamma_{\mathrm{R},n-1}) m_0 c_0^2 = Q V_{\mathrm{R},n-1}$$

$$\Rightarrow \gamma_{\mathrm{R},n} = \gamma_{\mathrm{R},n-1} + \frac{Q}{m_0 c_0^2} V_{\mathrm{R},n-1}. \tag{3.4}$$

This completes the analysis of the reference particle.

After the nth revolution, an asynchronous particle, i.e., an off-momentum particle, will be located at the gap at time t_n. For the nth revolution, this asynchronous particle will need time T_n:

$$u_n = \frac{l_{\mathrm{R}} + \Delta l_n}{T_n},$$

$$t_n = t_{n-1} + \frac{l_{\mathrm{R}} + \Delta l_n}{\beta_n c_0}. \tag{3.5}$$

Before the nth revolution, the energy of the asynchronous particle increases from γ_{n-1} to γ_n. This energy step is caused by the voltage V_{n-1}:

$$\gamma_n = \gamma_{n-1} + \frac{Q}{m_0 c_0^2} V_{n-1}. \tag{3.6}$$

From now on, we shall use Δ quantities to specify differences between the asynchronous particle and the reference particle (i.e., the synchronous particle). If we set $\Delta t_n = t_n - t_{\mathrm{R},n}$ and $\Delta \gamma_n = \gamma_n - \gamma_{\mathrm{R},n}$, then Eqs. (3.3) and (3.5) lead to

$$\Delta t_n = \Delta t_{n-1} + \frac{l_{\mathrm{R}} + \Delta l_n}{\beta_n c_0} - \frac{l_{\mathrm{R}}}{\beta_{\mathrm{R},n} c_0} =$$

$$= \Delta t_{n-1} + \frac{\beta_{\mathrm{R},n} l_{\mathrm{R}} + \beta_{\mathrm{R},n} \Delta l_n - \beta_n l_{\mathrm{R}}}{\beta_n \beta_{\mathrm{R},n} c_0}$$

$$= \Delta t_{n-1} + l_{\mathrm{R}} \frac{\frac{\Delta l_n}{l_{\mathrm{R}}} \beta_{\mathrm{R},n} - \Delta\beta_n}{\beta_n \beta_{\mathrm{R},n} c_0}$$

$$\Rightarrow \Delta t_n = \Delta t_{n-1} + \frac{l_{\mathrm{R}}}{\beta_n c_0} \left(\frac{\Delta l_n}{l_{\mathrm{R}}} - \frac{\Delta\beta_n}{\beta_{\mathrm{R},n}} \right). \tag{3.7}$$

Based on Eqs. (3.4) and (3.6), one obtains, by means of $\Delta V_n = V_n - V_{\mathrm{R},n}$,

$$\Delta\gamma_n = \Delta\gamma_{n-1} + \frac{Q}{m_0 c_0^2} \Delta V_{n-1}. \tag{3.8}$$

We are now interested in the change of the orbit length that is caused by the change of momentum.

This is relevant, because the orbit of the asynchronous particle will have a different radius r_n from that of the reference particle, since its momentum is different:

$$p_n = Q r_n B_n.$$

The magnetic field B_n, however, is the same as for the reference particle, since we assume that the asynchronous particle is still inside the straight section when the reference particle is located at the gap.

We obviously have

$$\frac{r_n}{r_{\mathrm{R}}} = \frac{p_n}{p_{\mathrm{R},n}}, \qquad r_n = r_{\mathrm{R}} \frac{p_n}{p_{\mathrm{R},n}} = r_{\mathrm{R}} + \left(\frac{p_n}{p_{\mathrm{R},n}} - 1 \right) r_{\mathrm{R}}.$$

It follows that

$$l_n = 2\pi r_n + 2d = l_{\mathrm{R}} + 2\pi \left(\frac{p_n}{p_{\mathrm{R},n}} - 1 \right) r_{\mathrm{R}}.$$

Using $\Delta l_n = l_n - l_{\mathrm{R}}$ and $\Delta p_n = p_n - p_{\mathrm{R},n}$, one obtains

$$\Delta l_n = 2\pi r_{\mathrm{R}} \frac{\Delta p_n}{p_{\mathrm{R},n}}.$$

Since l_{R} and r_{R} are constant, the quantity

$$\alpha_{\mathrm{c}} = \frac{2\pi r_{\mathrm{R}}}{l_{\mathrm{R}}} \tag{3.9}$$

can be defined, leading to

$$\boxed{\frac{\Delta l_n}{l_\mathrm{R}} = \alpha_\mathrm{c} \frac{\Delta p_n}{p_{\mathrm{R},n}}.} \tag{3.10}$$

Due to this equation, α_c is called the **momentum compaction factor**. The simple expression in Eq. (3.9) is due to our very special simplified model synchrotron. In general, it depends on the **synchrotron lattice**, i.e., on the special combination of all magnets. In Appendix A.3, a more general equation, namely (A.26), is discussed. In most synchrotrons, the momentum compaction factor is positive (cf. [1, Sect. 2.1.1]):

$$\alpha_\mathrm{c} > 0.$$

In this book, we will restrict ourselves to this case.

By means of equation (3.10), we may convert Eq. (3.7) as follows:

$$\Delta t_n = \Delta t_{n-1} + \frac{l_\mathrm{R}}{\beta_n c_0} \left(\alpha_\mathrm{c} \frac{\Delta p_n}{p_{\mathrm{R},n}} - \frac{\Delta \beta_n}{\beta_{\mathrm{R},n}} \right).$$

With

$$\begin{aligned} \beta_n \gamma_n - \beta_{\mathrm{R},n} \gamma_{\mathrm{R},n} &= (\beta_{\mathrm{R},n} + \Delta\beta_n)(\gamma_{\mathrm{R},n} + \Delta\gamma_n) - \beta_{\mathrm{R},n}\gamma_{\mathrm{R},n} = \\ &= \beta_{\mathrm{R},n}\Delta\gamma_n + \Delta\beta_n \gamma_{\mathrm{R},n} + \Delta\beta_n \Delta\gamma_n, \end{aligned}$$

the equation

$$\frac{\Delta p_n}{p_{\mathrm{R},n}} = \frac{\beta_n \gamma_n - \beta_{\mathrm{R},n}\gamma_{\mathrm{R},n}}{\beta_{\mathrm{R},n}\gamma_{\mathrm{R},n}}$$

leads to

$$\frac{\Delta p_n}{p_{\mathrm{R},n}} = \frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}} + \frac{\Delta\beta_n}{\beta_{\mathrm{R},n}} + \frac{\Delta\beta_n}{\beta_{\mathrm{R},n}} \frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}}.$$

We conclude that

$$\Delta t_n = \Delta t_{n-1} + \frac{l_\mathrm{R}}{\beta_n c_0} \left(\alpha_\mathrm{c} \left(\frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}} + \frac{\Delta\beta_n}{\beta_{\mathrm{R},n}} + \frac{\Delta\beta_n}{\beta_{\mathrm{R},n}} \frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}} \right) - \frac{\Delta\beta_n}{\beta_{\mathrm{R},n}} \right). \tag{3.11}$$

Please note that up to this point, all derivations have been exact in the scope of our simplified synchrotron model. We now make first approximations, assuming

$$\frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}} \ll 1 \quad \text{and} \quad \frac{\Delta\beta_n}{\beta_{\mathrm{R},n}} \ll 1.$$

In Table 2.3 (p. 48), we find the conversion

$$\frac{\Delta\beta_n}{\beta_{\mathrm{R},n}} \approx \frac{1}{\gamma_{\mathrm{R},n}^2\beta_{\mathrm{R},n}^2}\frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}}. \tag{3.12}$$

This leads to

$$\Delta t_n = \Delta t_{n-1} + \frac{l_\mathrm{R}}{\beta_n c_0}\frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}}\left(\alpha_\mathrm{c}\left(1+\frac{1}{\gamma_{\mathrm{R},n}^2\beta_{\mathrm{R},n}^2}\right)-\frac{1}{\gamma_{\mathrm{R},n}^2\beta_{\mathrm{R},n}^2}\right),$$

and due to

$$1+\beta^2\gamma^2 = \gamma^2,$$

one obtains

$$\Delta t_n = \Delta t_{n-1} + \frac{l_\mathrm{R}}{\beta_n c_0}\frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}}\left(\alpha_\mathrm{c}\frac{1}{\beta_{\mathrm{R},n}^2}-\frac{1}{\gamma_{\mathrm{R},n}^2\beta_{\mathrm{R},n}^2}\right).$$

By means of

$$\eta_{\mathrm{R},n} = \alpha_\mathrm{c} - \frac{1}{\gamma_{\mathrm{R},n}^2}, \tag{3.13}$$

it follows that

$$\Delta t_n = \Delta t_{n-1} + \frac{l_\mathrm{R}\eta_{\mathrm{R},n}}{\beta_n\beta_{\mathrm{R},n}^2 c_0}\frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}}. \tag{3.14}$$

If we assume that the time $T_{\mathrm{R},n}$ needed by the particle for one revolution is short in comparison with the times in which Δt_n changes significantly, we may calculate the following limit:

$$\frac{\Delta t_n - \Delta t_{n-1}}{T_{\mathrm{R},n}} \to \frac{\mathrm{d}\Delta t}{\mathrm{d}t}.$$

The quantities without the index R that describe the asynchronous particle and the quantities with the index R that describe the synchronous particle differ only by a small amount. Therefore, we obtain, by the limiting process,

$$\frac{\mathrm{d}\Delta t}{\mathrm{d}t} = \frac{l_\mathrm{R}\eta_\mathrm{R}}{T_\mathrm{R}\beta_\mathrm{R}^3\gamma_\mathrm{R}c_0}\,\Delta\gamma$$

$$\Leftrightarrow \frac{\mathrm{d}\Delta t}{\mathrm{d}t} = \frac{\eta_\mathrm{R}}{\beta_\mathrm{R}^2\gamma_\mathrm{R}}\,\Delta\gamma. \tag{3.15}$$

In an analogous way, we may perform a limiting process for Eq. (3.8):

$$\frac{\mathrm{d}\Delta\gamma}{\mathrm{d}t} = \frac{Q}{T_\mathrm{R}m_0c_0^2}\,\Delta V.$$

If we assume that the voltage

$$V(t) = \hat{V}(t)\ \sin(\varphi_\mathrm{RF}(t)) \qquad \text{with} \qquad \varphi_\mathrm{RF}(t) = \int_0^t \omega_\mathrm{RF}(\tilde{t})\ \mathrm{d}\tilde{t}$$

is harmonic (see Chap. 1) and that the reference particle experiences the **reference phase** (also called **synchronous phase**) $\varphi_\mathrm{RF} = \varphi_\mathrm{R}$ when it passes the gap, we obtain

$$V_\mathrm{R} = V(t_\mathrm{R}) = \hat{V}(t_\mathrm{R})\ \sin\varphi_\mathrm{R} \tag{3.16}$$

for the synchronous particle and

$$V(t = t_\mathrm{R} + \Delta t) \approx \hat{V}(t_\mathrm{R})\ \sin(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R})$$

for an off-momentum particle. Here we have assumed that neither the amplitude $\hat{V}$ nor the frequency ω_RF changes significantly during the time Δt. It follows that

$$\frac{\mathrm{d}\Delta\gamma}{\mathrm{d}t} = \frac{Q\hat{V}}{T_\mathrm{R}m_0c_0^2}\left(\sin(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}) - \sin\varphi_\mathrm{R}\right). \tag{3.17}$$

Equations (3.15) and (3.17) may be written in the form

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{v}(\vec{r}),$$

which is compatible with Sect. 2.8 if

$$\vec{r} = \begin{pmatrix} \Delta t \\ \Delta\gamma \end{pmatrix} \qquad \text{and} \qquad \vec{v} = \begin{pmatrix} v_{\Delta t} \\ v_{\Delta\gamma} \end{pmatrix}$$

are defined. For the divergence of $\vec{v}$, one obtains

$$\mathrm{div}\ \vec{v} = \frac{\partial v_{\Delta t}}{\partial\Delta t} + \frac{\partial v_{\Delta\gamma}}{\partial\Delta\gamma} = 0.$$

Thus we have shown that the flow in phase space preserves the phase space area (see Sect. 2.10.3); Liouville's theorem is satisfied.[2] Therefore, one usually uses the unit eVs for the phase space area, i.e., a product of time and energy. This unit is obtained if we calculate with ΔW instead of $\Delta\gamma$. This does not change the invariance of the phase space area, since the required factor $m_0c_0^2$ is constant:

$$\frac{\mathrm{d}\Delta t}{\mathrm{d}t} = \frac{\eta_\mathrm{R}}{m_0c_0^2\beta_\mathrm{R}^2\gamma_\mathrm{R}}\,\Delta W$$

$$\Leftrightarrow \boxed{\frac{\mathrm{d}\Delta t}{\mathrm{d}t} = \frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2}\,\Delta W,} \tag{3.18}$$

$$\boxed{\frac{\mathrm{d}\Delta W}{\mathrm{d}t} = \frac{Q\hat{V}}{T_\mathrm{R}}\left(\sin(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}) - \sin\varphi_\mathrm{R}\right).} \tag{3.19}$$

Here we used the total energy

$$\boxed{W_\mathrm{R} = mc_0^2 = m_0c_0^2\gamma_\mathrm{R}} \tag{3.20}$$

of the synchronous particle.

3.3 Tracking Equations

Equations (3.8) and (3.14) may be used as recurrence steps for a so-called **particle tracking** program. It is remarkable that the result of Eq. (3.8) has to be inserted on the right-hand side of Eq. (3.14). Therefore, one cannot evaluate both equations at the same time. If one replaced $\Delta\gamma_n$ in Eq. (3.14) by $\Delta\gamma_{n-1}$, which does not seem to change things significantly at first sight, then unstable orbits instead of stable orbits would be the result. This may be verified as follows:

Equation (3.14) obviously has the form

$$x_k = x_{k-1} + f(y_k).$$

Equation (3.8), however, has the form

$$y_k = y_{k-1} + g(x_{k-1}).$$

[2]The validity of Liouville's theorem is of course dependent on the assumptions that we made to derive the equations of motion in longitudinal phase space.

If we insert the second equation into the first in order to have only quantities on the right-hand side that belong to step $k-1$, we obtain

$$\begin{aligned} x_k &= x_{k-1} + f(y_{k-1} + g(x_{k-1})), \\ y_k &= y_{k-1} + g(x_{k-1}). \end{aligned}$$

A map

$$\vec{r}_k = \vec{F}(\vec{r}_{k-1})$$

is area-preserving if the absolute value of the Jacobian equals 1, as we discussed in Sect. 2.10.2. In the case under consideration here, the Jacobian is

$$\frac{\partial(x_k, y_k)}{\partial(x_{k-1}, y_{k-1})} = \begin{vmatrix} 1 + f'(y_{k-1} + g(x_{k-1}))\, g'(x_{k-1}) & f'(y_{k-1} + g(x_{k-1})) \\ g'(x_{k-1}) & 1 \end{vmatrix} = 1.$$

Hence, the system that is described by the **tracking equations** (3.8) and (3.14) is actually area-preserving. The discrete tracking equations therefore are a reasonable discretization of the continuous system; the corresponding iteration algorithm is called a leapfrog scheme (cf. [2, Appendix E]).

If, however, one starts with the symmetric pair of equations

$$\begin{aligned} x_k &= x_{k-1} + f(y_{k-1}), \\ y_k &= y_{k-1} + g(x_{k-1}), \end{aligned}$$

one obtains the Jacobian

$$\frac{\partial(x_k, y_k)}{\partial(x_{k-1}, y_{k-1})} = \begin{vmatrix} 1 & f'(y_{k-1}) \\ g'(x_{k-1}) & 1 \end{vmatrix} \neq 1.$$

In contrast to the continuous system, these discrete equations do not describe a conservative system. We may therefore conclude that Eqs. (3.8) and (3.14) are correct, even though they are not symmetric.

By means of the tracking equations, we are now able to simulate the behavior of particle clouds in **longitudinal phase space**. An example is shown in Fig. 3.2. At the start of the simulation, several particles are randomly distributed in phase space inside an ellipse. A harmonic voltage

$$V(t) = \hat{V}\ \sin(\omega_{\mathrm{RF}} t)$$

is assumed. Each particle moves in phase space according to the tracking equations.

The simulation shows some effects that can also be seen in the snapshots displayed in Fig. 3.2:

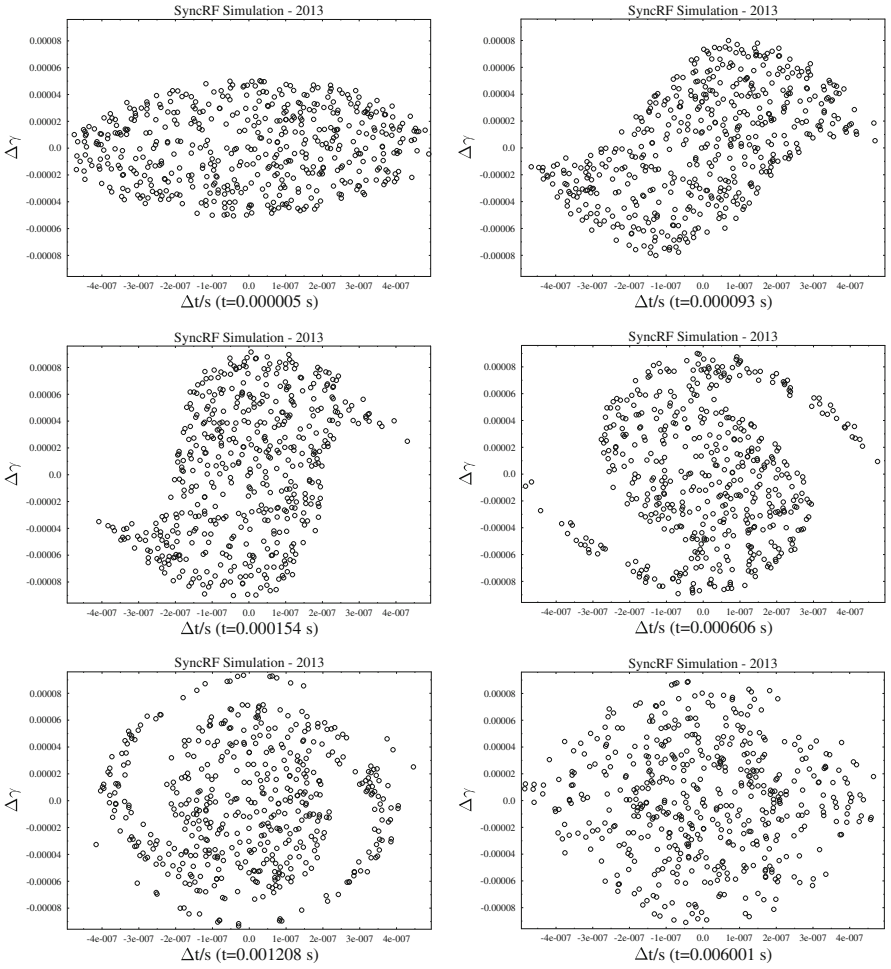

Fig. 3.2 Example of a tracking simulation of particles in phase space

- The particles perform rotations in longitudinal phase space (counterclockwise). This is the so-called **synchrotron oscillation** of individual particles that was already mentioned in the introduction. The conclusion that the particles oscillate around the synchronous phase, i.e., that they perform oscillations on the axis Δt, is now visualized in a different way: since the energy deviation is added in phase space, this oscillation on the time axis is converted into a rotation in phase space. For the sake of clarity, it should again be emphasized (cf. Footnote 4 on p. 5) that the oscillation does not occur around a single slope of the voltage $V(t)$ but that an overlay of several revolutions is necessary to make this oscillation (and also the rotation in phase space) visible. After all, the particles fly in the longitudinal direction with very high speed.

- The combination of all the particles (e.g., in the first diagram of Fig. 3.2) is called a **bunch**.
- The third picture shows approximately the initial bunch rotated by 90° in phase space (and stretched due to the arbitrary scaling of the axes). However, the bunch shape is not identical to the original elliptical bunch, because the outer particles rotate with a lower oscillation frequency; they lag behind. In the beginning, this leads to an S-shaped bunch, and later, spiral galaxies are formed in phase space.
- After several revolutions, the different spiral arms can no longer be distinguished from each other.
- Based on the first four (or even five) pictures, one may verify that the area occupied by the particles in phase space remains constant, as proven above. In the last picture, the occupied area seems to be larger, but this is due to the effect mentioned earlier that the spiral arms can no longer be identified clearly. One speaks of **filamentation** and **phase space dilution**. The reason is the limited number of particles used instead of a continuous particle distribution. Winding up the spiral arms will involve some empty phase space area that is also wound up in between. Strictly speaking, the definition of the area that is occupied by particles is possible only for a continuous particle distribution with clear boundaries, i.e., for an infinite number of particles. For a finite number of particles, the definition may be used only approximately. In the beginning, the number of particles is large enough to identify the boundaries of the bunch. Later, the same number of particles is no longer sufficient to identify the more complex bunch shape with the wound up spiral arms. Therefore, as a result of the lack of a clear definition of area for a finite number of particles, the visible area occupied by the particles in the last diagram is larger than that in the first diagrams. We will return to this problem in Sects. 3.24, 5.4, and 5.6.4.
- The phase space area that is occupied by particles is called the **longitudinal emittance**. Due to Liouville's theorem, the longitudinal emittance is constant.[3] For the longitudinal emittance, one usually uses eVs as the physical unit. In this book, we are not dealing with transverse beam dynamics, and we will therefore omit the word "longitudinal" from time to time.

[3] As mentioned above, this statement depends, of course, on the assumptions that were made to derive the equations of motion in longitudinal phase space. Examples of effects that do not keep the phase space area occupied by the particles constant are:

– Active beam cooling (which reduces the area occupied by particles in phase space).
– Synchrotron radiation (especially relevant for electron beams but negligible for ion beams if they are not accelerated to extremely high energies).

As discussed before, it is also possible that an increase in the phase space area occupied by particles is observed based on the finite number of particles. This may, for example, be caused by undesired disturbances. In this case, one should speak rather of an increase in the RMS emittance that will be introduced in Sect. 5.6.4. Calling this effect purely "emittance increase" may lead to misunderstandings caused by the validity of Liouville's theorem in combination with the difficulty in defining an area for a finite number of particles.

- If the simulation runs even longer than displayed here, the macroscopic shape of the bunch will no longer change significantly. The individual particles will, of course, continue to move, but the distribution will no longer change. As we will see later, this is the case of a **matched bunch**, for which

$$\frac{\partial \rho}{\partial t} = 0$$

holds everywhere in phase space.
- All particles that are located inside a small time interval (i.e., inside a narrow vertical strip) will contribute to the beam current during this time interval. In other words, the specific instantaneous energy of these particles does not matter. Therefore, the **beam signal** is obtained as the **projection**

$$I_{\mathrm{beam}}(t) = I_{\mathrm{beam}}(t_{\mathrm{R}} + \Delta t) = \int \rho_q(\Delta t, \Delta W)\, \mathrm{d}(\Delta W)$$

of the bunch onto the time axis. This is why both the particle distribution in phase space and the pulses of the beam signal are called bunches.

Of course, the example shown in Fig. 3.2 is a very special one. If, for example, a larger bunch had been assumed as an initial distribution, one would have seen that the outer particles may also move on unstable trajectories. Therefore, we will discuss the phase space behavior in more detail in the following sections. Nevertheless, the individual particles will still show a very transparent behavior that justifies our phase space analysis.

The phenomena discussed here were based on a computer simulation of several particles. These tracking simulations are inevitable for complicated situations. Nevertheless, it is possible to study the basic effects analytically. This will be done in the following sections.

3.4 Phase Slip Factor and Transition Energy

The quantity[4]

$$\boxed{\eta_{\mathrm{R}} = \alpha_{\mathrm{c}} - \frac{1}{\gamma_{\mathrm{R}}^2},} \tag{3.21}$$

[4]The index n that specifies the revolution is omitted in this section, since the quantities will be regarded as quasicontinuous here.

which was introduced in Eq. (3.13), is called[5] a slip factor (cf. [4]), **phase slip factor** (cf. [5]), frequency slip factor (cf. [6]), or slippage factor [7]. This name will be explained in the following.

Based on

$$u_\mathrm{R} = \frac{l_\mathrm{R}}{T_\mathrm{R}} \qquad \Leftrightarrow T_\mathrm{R} = \frac{l_\mathrm{R}}{\beta_\mathrm{R} c_0},$$

we may obtain the following relation by calculating the derivative with respect to time:

$$\dot{T}_\mathrm{R} = \frac{\dot{l}_\mathrm{R}\beta_\mathrm{R}c_0 - \dot{\beta}_\mathrm{R}c_0 l_\mathrm{R}}{\beta_\mathrm{R}^2 c_0^2}$$

$$\Rightarrow \frac{\dot{T}_\mathrm{R}}{T_\mathrm{R}} = \frac{\dot{l}_\mathrm{R}}{l_\mathrm{R}} - \frac{\dot{\beta}_\mathrm{R}}{\beta_\mathrm{R}}.$$

Therefore, for a deviation ΔT of the revolution time, one obtains approximately

$$\frac{\Delta T}{T_\mathrm{R}} = \frac{\Delta l}{l_\mathrm{R}} - \frac{\Delta \beta}{\beta_\mathrm{R}}.$$

On the one hand, we may make use of Eq. (3.10),

$$\frac{\Delta l}{l_\mathrm{R}} = \alpha_\mathrm{c} \frac{\Delta p}{p_\mathrm{R}},$$

to calculate the deviation Δl of the orbit length. On the other hand, one obtains

$$\gamma^2 = \frac{1}{1-\beta^2} \qquad \Rightarrow \gamma^2 - \beta^2\gamma^2 = 1 \qquad \Rightarrow \beta\gamma = \sqrt{\gamma^2 - 1}$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}(\beta\gamma) = \frac{2\gamma\dot{\gamma}}{2\sqrt{\gamma^2-1}} = \frac{\dot{\gamma}}{\beta}$$

$$\Rightarrow \frac{1}{\beta\gamma}\frac{\mathrm{d}}{\mathrm{d}t}(\beta\gamma) = \frac{1}{\beta^2}\frac{\dot{\gamma}}{\gamma} \qquad \Rightarrow \frac{\Delta p}{p_\mathrm{R}} \approx \frac{1}{\beta_\mathrm{R}^2}\frac{\Delta\gamma}{\gamma_\mathrm{R}}.$$

Together with Eq. (3.12),

$$\frac{\Delta\beta}{\beta_\mathrm{R}} = \frac{1}{\gamma_\mathrm{R}^2\beta_\mathrm{R}^2}\frac{\Delta\gamma}{\gamma_\mathrm{R}},$$

[5] Please note that the slip factor is sometimes defined with a negative sign (cf. [3]).

one obtains

$$\frac{\Delta\beta}{\beta_\mathrm{R}} = \frac{1}{\gamma_\mathrm{R}^2}\frac{\Delta p}{p_\mathrm{R}},$$

which may be inserted above:

$$\frac{\Delta T}{T_\mathrm{R}} = \left(\alpha_\mathrm{c} - \frac{1}{\gamma_\mathrm{R}^2}\right)\frac{\Delta p}{p_\mathrm{R}}.$$

Therefore,

$$\boxed{\frac{\Delta T}{T_\mathrm{R}} = \eta_\mathrm{R}\,\frac{\Delta p}{p_\mathrm{R}}}$$

is valid. In conclusion, the factor η_R describes by what percentage the revolution time changes if the momentum changes by a certain percentage. Every time deviation Δt of a particle with respect to the reference particle may be converted into a phase deviation:

$$\Delta\varphi_\mathrm{RF} = \omega_\mathrm{RF}\Delta t = 2\pi h\frac{\Delta t}{T_\mathrm{R}}$$

$$\Leftrightarrow \frac{\Delta\varphi_\mathrm{RF}}{2\pi} = h\,\frac{\Delta t}{T_\mathrm{R}}.$$

This explains the term "phase slip factor."

For sufficiently low energies, $\eta_\mathrm{R} < 0$ is valid, so that a larger momentum causes the particles to arrive again earlier at the gap. For

$$\alpha_\mathrm{c} > \frac{1}{\gamma_\mathrm{R}^2},$$

this is no longer valid. Then the path that has to be taken by the particles increases so much that the momentum increase cannot compensate for it; the particles will arrive later at the gap. One defines the **transition gamma** γ_T by

$$\boxed{\alpha_\mathrm{c} = \frac{1}{\gamma_\mathrm{T}^2} \qquad \gamma_\mathrm{T} = \frac{1}{\sqrt{\alpha_\mathrm{c}}},}$$

so that this transition point is reached for $\gamma_\mathrm{R} = \gamma_\mathrm{T}$. The corresponding energy is called the **transition energy**.

3.5 Accelerating Voltage

The magnetic field B_n will be given at each revolution n. This leads to the question how large the voltage $V_{\mathrm{R},n}$ must be to ensure that the synchronous particle always takes the same reference path of length l_R.

Due to Eq. (3.2), we first of all note that from a knowledge of B_n, the momentum $p_{\mathrm{R},n}$ of the reference particle is known for each revolution n. Therefore, $\beta_{\mathrm{R},n}$ and $\gamma_{\mathrm{R},n}$ are also known.

From $\gamma = \sqrt{1+(\beta\gamma)^2}$, Eq. (3.4) leads to

$$V_{\mathrm{R},n} = \frac{m_0c_0^2}{Q}(\gamma_{\mathrm{R},n+1} - \gamma_{\mathrm{R},n}) = \frac{m_0c_0^2}{Q}\left(\sqrt{1+\left(\frac{p_{\mathrm{R},n+1}}{m_0c_0}\right)^2} - \sqrt{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2}\right).$$

If we insert

$$p_{\mathrm{R},n+1} = p_{\mathrm{R},n} + \delta p,$$

we obtain

$$\sqrt{1+\left(\frac{p_{\mathrm{R},n+1}}{m_0c_0}\right)^2} = \sqrt{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2 + 2\frac{p_{\mathrm{R},n}\delta p}{(m_0c_0)^2} + \frac{\delta p^2}{(m_0c_0)^2}} =$$

$$= \sqrt{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2}\sqrt{1+\frac{2\frac{p_{\mathrm{R},n}\delta p}{(m_0c_0)^2} + \frac{\delta p^2}{(m_0c_0)^2}}{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2}}.$$

Using $\sqrt{1+x} \approx 1 + \frac{x}{2}$ for $x \ll 1$ in the second square root and neglecting the quadratic term leads to

$$\sqrt{1+\left(\frac{p_{\mathrm{R},n+1}}{m_0c_0}\right)^2} \approx \sqrt{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2}\left(1+\frac{\frac{p_{\mathrm{R},n}\delta p}{(m_0c_0)^2}}{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2}\right).$$

This yields the approximation

$$V_{\mathrm{R},n} \approx \frac{m_0c_0^2}{Q}\sqrt{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2}\,\frac{\frac{p_{\mathrm{R},n}\delta p}{(m_0c_0)^2}}{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2} = \frac{1}{m_0Q}\frac{p_{\mathrm{R},n}\delta p}{\sqrt{1+\left(\frac{p_{\mathrm{R},n}}{m_0c_0}\right)^2}}.$$

Due to $\frac{p_{\mathrm{R},n}}{m_0c_0} = \gamma_{\mathrm{R},n}\beta_{\mathrm{R},n}$ and $\sqrt{1+(\gamma_{\mathrm{R},n}\beta_{\mathrm{R},n})^2} = \gamma_{\mathrm{R},n}$, it follows that

$$V_{\mathrm{R},n} \approx \frac{p_{\mathrm{R},n}\delta p}{m_0 Q \gamma_{\mathrm{R},n}} = \frac{c_0 \beta_{\mathrm{R},n}\delta p}{Q} = \frac{u_{\mathrm{R},n}\delta p}{Q} = \frac{l_\mathrm{R}\delta p}{T_{\mathrm{R},n} Q}.$$

With the help of Eq. (3.2), one obtains

$$\delta p = p_{\mathrm{R},n+1} - p_{\mathrm{R},n} = Q r_\mathrm{R}(B_{n+1} - B_n),$$

and one finally obtains

$$V_{\mathrm{R},n} \approx l_\mathrm{R} r_\mathrm{R} \frac{B_{n+1} - B_n}{T_{\mathrm{R},n}},$$

which leads to the continuous counterpart

$$\boxed{V_\mathrm{R} \approx l_\mathrm{R} r_\mathrm{R} \dot{B}.} \tag{3.22}$$

This is the voltage that the reference particle actually experiences, not the required amplitude. The relation between these two quantities is given by Eq. (3.16):

$$\boxed{V_\mathrm{R} = \hat{V} \sin\varphi_\mathrm{R}.} \tag{3.23}$$

Without loss of generality, the reference phase φ_R will always be in the range $\varphi_\mathrm{R} \in [-\pi, +\pi]$. We discuss two cases:

- $\dot{B} = 0$: This means that the magnetic field remains constant, i.e., that the reference particle does not change its momentum or energy, as Eq. (3.1) shows. In this case, $V_\mathrm{R} = 0$ must hold. This can be realized either by $\hat{V} = 0$ or by $\hat{V} > 0$ and $\varphi_\mathrm{R} \in \{0, \pm\pi\}$. The first case means that none of the particles experiences a voltage, so that a coasting beam will be present. The second case means that—in contrast to the reference particle—the off-momentum particles will experience a voltage. The phase-focusing principle (see Chap. 1) will lead to a bunched beam in this case. The larger the amplitude $\hat{V}$ is, the stronger will be the bunching effect, i.e., the shorter the bunches (lower bunching factor B_f according to Eq. (2.37)). The condition $\dot{B} = 0$ is usually satisfied before acceleration and after acceleration, i.e., at injection energy and at extraction energy. One also speaks of the **injection plateau** and the **extraction plateau**. The latter is also called the **flat top energy**.
- $\dot{B} > 0$: This means that the magnetic field, and hence the momentum and energy of the reference particle, increases. This is the acceleration phase, which is possible only if $\hat{V} > 0$ and $\varphi_\mathrm{R} > 0$ for $Q > 0$ (or $\varphi_\mathrm{R} < 0$ for $Q < 0$) are chosen. Only bunched beams can be accelerated.

We will return to these facts and discuss them in more detail in Sect. 3.22.

3.6 Synchrotron Oscillation

For small time deviations Δt with $\omega_{\mathrm{RF}}\Delta t \ll 1$, the approximation

$$\begin{aligned}\sin(\omega_{\mathrm{RF}}\Delta t + \varphi_{\mathrm{R}}) &= \sin(\omega_{\mathrm{RF}}\Delta t)\ \cos\varphi_{\mathrm{R}} + \cos(\omega_{\mathrm{RF}}\Delta t)\ \sin\varphi_{\mathrm{R}} \\ &\approx \omega_{\mathrm{RF}}\Delta t\ \cos\varphi_{\mathrm{R}} + \sin\varphi_{\mathrm{R}}\end{aligned}$$

is valid, so that a linear approximation of Eq. (3.19) leads to the relation

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}t} = \frac{Q\hat{V}\ \omega_{\mathrm{RF}}\ \cos\varphi_{\mathrm{R}}}{T_{\mathrm{R}}}\ \Delta t. \tag{3.24}$$

Together with Eq. (3.18), it follows that

$$\frac{\mathrm{d}^2\Delta W}{\mathrm{d}t^2} = \frac{\eta_{\mathrm{R}}\ Q\hat{V}\ \omega_{\mathrm{RF}}\ \cos\varphi_{\mathrm{R}}}{W_{\mathrm{R}}\beta_{\mathrm{R}}^2 T_{\mathrm{R}}}\ \Delta W,$$

if one takes into account that the quantities in the fraction will change slowly. This is an oscillator equation of the form

$$\frac{\mathrm{d}^2\Delta W}{\mathrm{d}t^2} + \omega_{\mathrm{S},0}^2\Delta W = 0$$

with

$$\omega_{\mathrm{S},0} = \sqrt{-\frac{\eta_{\mathrm{R}}\ Q\hat{V}\ \omega_{\mathrm{RF}}\ \cos\varphi_{\mathrm{R}}}{W_{\mathrm{R}}\beta_{\mathrm{R}}^2 T_{\mathrm{R}}}}$$

$$\Leftrightarrow\ \omega_{\mathrm{S},0} = \sqrt{-\frac{\eta_{\mathrm{R}}\ Q\hat{V}\ h2\pi\ \cos\varphi_{\mathrm{R}}}{W_{\mathrm{R}}\beta_{\mathrm{R}}^2 T_{\mathrm{R}}^2}}$$

$$\Leftrightarrow\ \omega_{\mathrm{S},0} = \frac{2\pi}{T_{\mathrm{R}}}\sqrt{-\frac{\eta_{\mathrm{R}}\ Q\hat{V}\ h\ \cos\varphi_{\mathrm{R}}}{2\pi W_{\mathrm{R}}\beta_{\mathrm{R}}^2}}$$

$$\Leftrightarrow\ \boxed{f_{\mathrm{S},0} = f_{\mathrm{R}}\sqrt{-\frac{\eta_{\mathrm{R}}\ Q\hat{V}\ h\ \cos\varphi_{\mathrm{R}}}{2\pi W_{\mathrm{R}}\beta_{\mathrm{R}}^2}}.} \tag{3.25}$$

The oscillation frequency $f_{\mathrm{S},0}$ is called the **synchrotron frequency**. The index 0 indicates that it is valid for particles with small oscillation amplitudes with respect to the reference particle. Please note that for positive charges $Q > 0$ below the transition $\eta_{\mathrm{R}} < 0$, one chooses $0 \le \varphi_{\mathrm{R}} < \pi/2$, so that the argument of the square

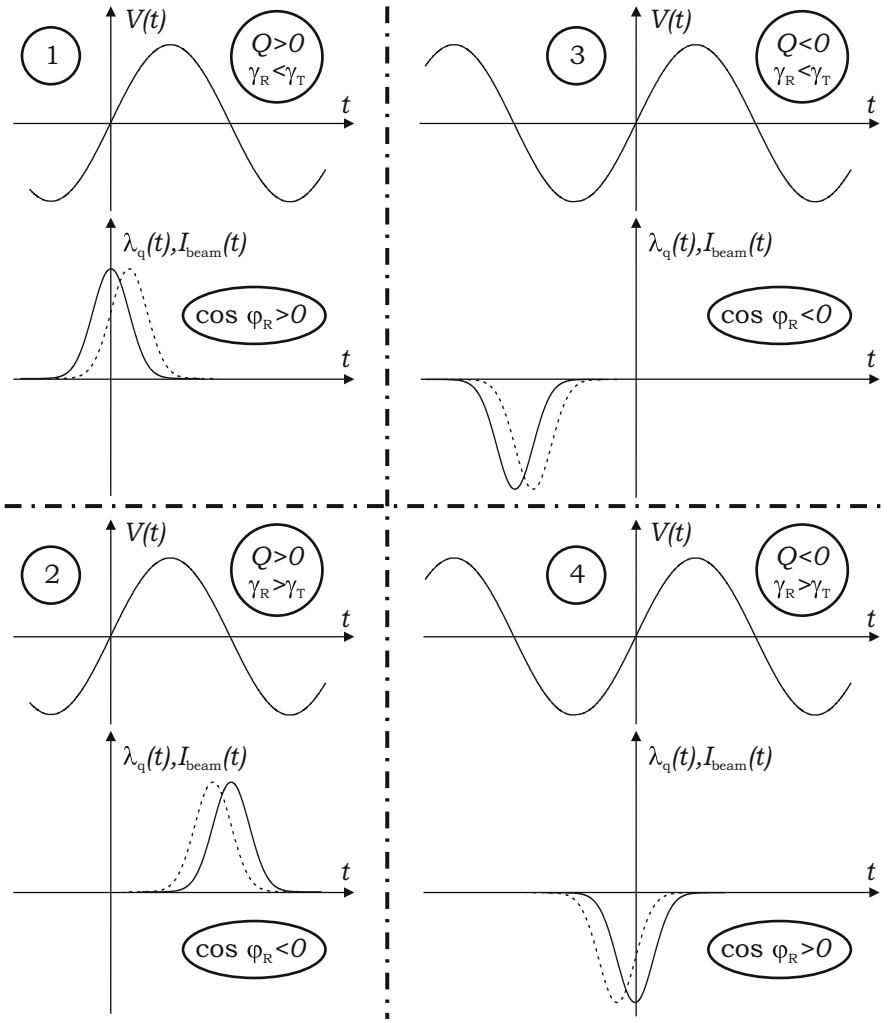

Fig. 3.3 Phase relation between gap voltage and beam signal (*solid line*: stationary case with $\varphi_R = 0$ for $\eta_R Q < 0$ or $\varphi_R = \pm\pi$ for $\eta_R Q > 0$; *dotted line*: acceleration case)

root is positive. The various cases have to be treated accordingly (see Fig. 3.3), so that

$$\eta_R \; Q \; \cos\varphi_R < 0 \tag{3.26}$$

holds.

3.7 Principal Axes

Due to the simple oscillator equation for small oscillation amplitudes, we get elliptic trajectories in the $(\Delta t, \Delta W)$ phase space. This is true for all particles inside a bunch that is sufficiently short. Now we try to find the principal axes of the bunch ellipse in phase space for small oscillation amplitudes. For this purpose, we begin with Eq. (3.24),

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}t} = \frac{Q\,\hat{V}\;\omega_{\mathrm{RF}}\;\cos\varphi_{\mathrm{R}}}{T_{\mathrm{R}}}\;\Delta t,$$

and insert the ansatz

$$\Delta W = \Delta\hat{W}\;\sin\omega_{\mathrm{S},0}t.$$

This leads to

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}t} = \Delta\hat{W}\;\omega_{\mathrm{S},0}\;\cos\omega_{\mathrm{S},0}t = \frac{Q\,\hat{V}\;\omega_{\mathrm{RF}}\;\cos\varphi_{\mathrm{R}}}{T_{\mathrm{R}}}\;\Delta t.$$

Hence, one obtains

$$\Delta t = \pm\Delta\hat{t}\;\cos\omega_{\mathrm{S},0}t$$

with

$$\Delta\hat{t} = \pm\Delta\hat{W}\;\frac{\omega_{\mathrm{S},0}T_{\mathrm{R}}}{Q\,\hat{V}\;\omega_{\mathrm{RF}}\;\cos\varphi_{\mathrm{R}}}.$$

Thus, we obtain the ratio of the principal axes

$$\frac{\Delta\hat{W}}{\Delta\hat{t}} = \pm\frac{Q\,\hat{V}\;\omega_{\mathrm{RF}}\;\cos\varphi_{\mathrm{R}}}{\omega_{\mathrm{S},0}T_{\mathrm{R}}} = \pm\frac{Q\,\hat{V}\;h\;f_{\mathrm{R}}^2\;\cos\varphi_{\mathrm{R}}}{f_{\mathrm{S},0}}.$$

From Eq. (3.25), it follows that

$$Q\,\hat{V}\;h\;\cos\varphi_{\mathrm{R}} = \frac{f_{\mathrm{S},0}^2}{f_{\mathrm{R}}^2}\frac{2\pi W_{\mathrm{R}}\beta_{\mathrm{R}}^2}{-\eta_{\mathrm{R}}}, \tag{3.27}$$

leading to

$$\boxed{\frac{\Delta\hat{W}}{\Delta\hat{t}} = f_{\mathrm{S},0}\;\frac{2\pi W_{\mathrm{R}}\beta_{\mathrm{R}}^2}{|\eta_{\mathrm{R}}|}.} \tag{3.28}$$

3.8 Hamiltonian

Now we want to analyze the longitudinal phase space dynamics without restricting ourselves to small oscillation amplitudes. For this purpose, we will determine a Hamiltonian. Equations (3.18) and (3.19) are converted into Hamilton's equations if we set

$$\frac{\partial H}{\partial \Delta t} = \Delta \dot{W},$$

$$\frac{\partial H}{\partial \Delta W} = -\Delta \dot{t}.$$

By integration of the first equation with respect to Δt, one obtains

$$H = \frac{Q\hat{V}}{T_\mathrm{R}} \left(-\frac{1}{\omega_\mathrm{RF}} \cos(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}) - \Delta t \ \sin\varphi_\mathrm{R} \right) + C_1(\Delta W).$$

Here we assumed that $\hat{V}$, T_R, ω_RF, φ_R are constant or change sufficiently slowly. An integration of the second equation with respect to ΔW leads to

$$H = -\frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2} \frac{\Delta W^2}{2} + C_2(\Delta t).$$

We combine both results to obtain

$$H = -\frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2} \frac{\Delta W^2}{2} - \frac{Q\hat{V}}{T_\mathrm{R}} \left(\frac{1}{\omega_\mathrm{RF}} \cos(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}) + \Delta t \ \sin\varphi_\mathrm{R} \right) + \text{const.}$$

The constant is now fixed so that $H(0,0) = 0$ holds:

$$\boxed{H = -\frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2} \frac{\Delta W^2}{2} - \frac{Q\hat{V}}{T_\mathrm{R}} \left(\frac{1}{\omega_\mathrm{RF}} [\cos(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}) - \cos\varphi_\mathrm{R}] + \Delta t \ \sin\varphi_\mathrm{R} \right).} \tag{3.29}$$

3.9 Separatrix

We will now analyze the properties of the Hamiltonian $H(\Delta W, \Delta t)$ in Eq. (3.29). First of all, we determine the fixed points of the system. For these fixed points, $\frac{\partial H}{\partial \Delta W} = 0$ and $\frac{\partial H}{\partial \Delta t} = 0$ must be valid (see Sect. 2.11.4). The first of these requirements leads directly to $\Delta W = 0$. The second equation leads to

$$-\sin(\omega_{\mathrm{RF}}\Delta t + \varphi_{\mathrm{R}}) + \sin\varphi_{\mathrm{R}} = 0.$$

One sees that this equation is satisfied for $\omega_{\mathrm{RF}}\Delta t + \varphi_{\mathrm{R}} = \varphi_{\mathrm{R}}$ and for[6] $\omega_{\mathrm{RF}}\Delta t + \varphi_{\mathrm{R}} = \pi\ \mathrm{sign}\,\varphi_{\mathrm{R}} - \varphi_{\mathrm{R}}$. Hence, we obtain the two fixed points $(\Delta W, \Delta t) = (0, 0)$ and $(\Delta W, \Delta t) = \left(0, \frac{\pi\ \mathrm{sign}\,\varphi_{\mathrm{R}} - 2\varphi_{\mathrm{R}}}{\omega_{\mathrm{RF}}}\right)$, for which, of course, further fixed points exist periodically.

According to Appendix A.4, the first one is the stable fixed point, for which H has a minimum or a maximum, and the second one is a saddle point. For the saddle point,

$$\Delta\varphi_{\mathrm{saddle}} = \pi\ \mathrm{sign}\,\varphi_{\mathrm{R}} - 2\varphi_{\mathrm{R}},$$

H has the value

$$\begin{aligned} H_s &= -\frac{Q\hat{V}}{T_{\mathrm{R}}}\left(\frac{1}{\omega_{\mathrm{RF}}}\left[\cos(\pi\ \mathrm{sign}\,\varphi_{\mathrm{R}} - \varphi_{\mathrm{R}}) - \cos\varphi_{\mathrm{R}}\right] + \frac{\pi\ \mathrm{sign}\,\varphi_{\mathrm{R}} - 2\varphi_{\mathrm{R}}}{\omega_{\mathrm{RF}}}\ \sin\varphi_{\mathrm{R}}\right) = \\ &= -\frac{Q\hat{V}}{T_{\mathrm{R}}\omega_{\mathrm{RF}}}\left(\cos(\pi\ \mathrm{sign}\,\varphi_{\mathrm{R}} - \varphi_{\mathrm{R}}) - \cos\varphi_{\mathrm{R}} + (\pi\ \mathrm{sign}\,\varphi_{\mathrm{R}} - 2\varphi_{\mathrm{R}})\ \sin\varphi_{\mathrm{R}}\right) = \\ &= \frac{Q\hat{V}}{T_{\mathrm{R}}\omega_{\mathrm{RF}}}\left(2\ \cos\varphi_{\mathrm{R}} + (2\varphi_{\mathrm{R}} - \pi\ \mathrm{sign}\,\varphi_{\mathrm{R}})\ \sin\varphi_{\mathrm{R}}\right). \end{aligned} \tag{3.30}$$

The Hamiltonian keeps this value H_s on the whole separatrix (see Sect. 2.8.9). Therefore, using Eq. (3.29), one obtains for the separatrix

$$\begin{aligned} -\frac{\eta_{\mathrm{R}}}{W_{\mathrm{R}}\beta_{\mathrm{R}}^2}\ \frac{\Delta W^2}{2} = \frac{Q\hat{V}}{T_{\mathrm{R}}\omega_{\mathrm{RF}}}\big[&\cos(\omega_{\mathrm{RF}}\Delta t + \varphi_{\mathrm{R}}) - \cos\varphi_{\mathrm{R}} + \omega_{\mathrm{RF}}\Delta t\ \sin\varphi_{\mathrm{R}} + \\ &+\ 2\ \cos\varphi_{\mathrm{R}} + (2\varphi_{\mathrm{R}} - \pi\ \mathrm{sign}\,\varphi_{\mathrm{R}})\ \sin\varphi_{\mathrm{R}}\big], \end{aligned}$$

$$\begin{aligned} \Delta W^2 = \frac{2W_{\mathrm{R}}\beta_{\mathrm{R}}^2 Q\hat{V}}{-\eta_{\mathrm{R}}T_{\mathrm{R}}\omega_{\mathrm{RF}}}\big[&\cos(\omega_{\mathrm{RF}}\Delta t + \varphi_{\mathrm{R}}) + \omega_{\mathrm{RF}}\Delta t\ \sin\varphi_{\mathrm{R}} + \cos\varphi_{\mathrm{R}} \\ &+ (2\varphi_{\mathrm{R}} - \pi\ \mathrm{sign}\,\varphi_{\mathrm{R}})\ \sin\varphi_{\mathrm{R}}\big]. \end{aligned} \tag{3.31}$$

[6] The **sign function** is introduced here in order to have the fixed points in the range $-\pi \le \varphi_{\mathrm{RF}} \le +\pi$ while the reference phase is in the range $-\pi \le \varphi_{\mathrm{R}} \le +\pi$. In this book, we use the definitions

$$\mathrm{sgn}(x) = \begin{cases} -1 & \text{for } x < 0, \\ 0 & \text{for } x = 0, \\ +1 & \text{for } x > 0, \end{cases} \qquad \mathrm{sign}(x) = \begin{cases} -1 & \text{for } x < 0, \\ +1 & \text{for } x \ge 0, \end{cases}$$

for $x \in \mathbb{R}$.

The maximum of this function is characterized by the fact that the derivative with respect to Δt vanishes, so that

$$-\omega_{\mathrm{RF}} \sin(\omega_{\mathrm{RF}} \Delta t + \varphi_{\mathrm{R}}) + \omega_{\mathrm{RF}} \ \sin \varphi_{\mathrm{R}} = 0$$

holds. This is satisfied for $\Delta t = 0$ and $\omega_{\mathrm{RF}} \Delta t + \varphi_{\mathrm{R}} = \pi \ \mathrm{sign}\, \varphi_{\mathrm{R}} - \varphi_{\mathrm{R}}$ (further zeros occur periodically). For the latter of these two values, $\Delta W = 0$, so that it does not correspond to the desired maximum. Therefore, the maximum is obtained for $\Delta t = 0$:

$$\Delta W_{\mathrm{max}} = \left(\frac{2 W_{\mathrm{R}} \beta_{\mathrm{R}}^2 Q \hat{V}}{-\eta_{\mathrm{R}} T_{\mathrm{R}} \omega_{\mathrm{RF}}} (2 \ \cos \varphi_{\mathrm{R}} + (2\varphi_{\mathrm{R}} - \pi \ \mathrm{sign}\, \varphi_{\mathrm{R}}) \ \sin \varphi_{\mathrm{R}}) \right)^{1/2}$$

$$\Leftrightarrow \ \Delta W_{\mathrm{max}} = \left(\frac{W_{\mathrm{R}} \beta_{\mathrm{R}}^2 Q \hat{V}}{-h \pi \eta_{\mathrm{R}}} (2 \ \cos \varphi_{\mathrm{R}} + (2\varphi_{\mathrm{R}} - \pi \ \mathrm{sign}\, \varphi_{\mathrm{R}}) \ \sin \varphi_{\mathrm{R}}) \right)^{1/2}.$$

For the stationary case, i.e., for $\varphi_{\mathrm{R}} = 0$ if Q and η_{R} have opposite signs (this is our standard case if nothing else is stated) or for $\varphi_{\mathrm{R}} = \pm\pi$ if Q and η_{R} have identical signs, one obtains

$$\boxed{\Delta W_{\mathrm{max,stat}} = \sqrt{\frac{2 \ W_{\mathrm{R}} \beta_{\mathrm{R}}^2 \ |Q| \ \hat{V}}{\pi h \ |\eta_{\mathrm{R}}|}}.} \tag{3.32}$$

It follows that

$$\boxed{\Delta W_{\mathrm{max}} = \Delta W_{\mathrm{max,stat}} \, Y(\varphi_{\mathrm{R}})} \tag{3.33}$$

with

$$\boxed{Y(\varphi_{\mathrm{R}}) = \sqrt{\left| \cos \varphi_{\mathrm{R}} + \left(\varphi_{\mathrm{R}} - \frac{\pi}{2} \ \mathrm{sign}\, \varphi_{\mathrm{R}} \right) \ \sin \varphi_{\mathrm{R}} \right|}.} \tag{3.34}$$

Let us now briefly analyze the function $Y(\varphi_{\mathrm{R}})$. An extremum is expected for

$$-\sin \varphi_{\mathrm{R}} + \sin \varphi_{\mathrm{R}} + \left(\varphi_{\mathrm{R}} - \frac{\pi}{2} \ \mathrm{sign}\, \varphi_{\mathrm{R}} \right) \ \cos \varphi_{\mathrm{R}} = 0,$$

i.e., for $\varphi_{\mathrm{R}} = \pm\pi/2$. As shown in Fig. 3.4, this is a minimum. In the interval $0 \leq \varphi_{\mathrm{R}} \leq \pi/2$, the function $Y(\varphi_{\mathrm{R}})$ decreases monotonically from 1 to 0. Please note that Eq. (3.34) is valid for all four cases discussed in the next section, i.e., for $-\pi \leq \varphi_{\mathrm{R}} \leq +\pi$; $Y(\varphi_{\mathrm{R}})$ is an even function.

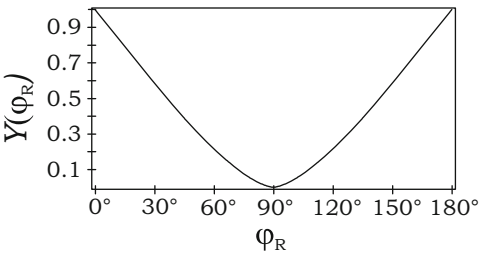

Fig. 3.4 Bucket height reduction factor $Y(\varphi_R)$

3.10 Symmetry with Respect to Transition Energy and Sign of Charge

Depending on whether positive or negative charges are accelerated, and depending on whether operation below or above transition energy is chosen, different cases have to be distinguished that are shown in Fig. 3.3 on p. 133 and Table 3.1 on p. 139.

One usually tries to avoid a crossing of the transition energy during acceleration.[7] However, in principle, **transition crossing** is also possible (cf. [8, Sect. 4.7] and [4, Sect. 2.2.3]). In this case, a γ_T jump (cf. [9, Sect. 7.7] and [8, Sect. 4.7]) may, for example, be applied in order to cross the transition energy quickly. According to Table 3.1, this includes a shift of the RF phase.

We will now analyze different substitutions in Eq. (3.29):

- If $\varphi_R' = -\varphi_R$ and $\Delta t' = -\Delta t$ are substituted, the expression in parentheses will still have the same form. If, in addition, the signs of Q and of η_R are reversed, the Hamiltonian will change its sign, which does not modify the trajectories. These steps lead us from case 1 to case 4 or from case 2 to case 3. In total, the separatrix and the trajectories will experience a reflection across the ordinate.
- If $\varphi_R' = \varphi_R - \pi \operatorname{sign} \varphi_R$ is substituted (this corresponds to $\varphi_R = \varphi_R' - \pi \operatorname{sign} \varphi_R'$ if the values 0, π, and $-\pi$ are excluded), the expression in parentheses will still have the same form, but its sign will change. If, in addition, the sign of Q is inverted, the Hamiltonian will not change at all. These steps lead us from case 1 to case 3 or from case 2 to case 4. In total, the separatrix and the trajectories will not change at all.

In sum, all four cases lead to the same trajectories and separatrices. Only a reflection across the ordinate may be necessary.

This allows us to restrict the following analysis of the orbits and the bucket geometry to case 1 without suffering any loss of generality.

[7] According to Eq. (3.25), the synchrotron frequency becomes zero at transition, which makes the bunches prone to unstable behavior.

Table 3.1 Symmetry of the buckets (acceleration case only)

Case	Charge	Operation below/above transition energy	Phase slip factor	Synchronous phase	Saddle point
1.	$Q > 0$	$\gamma_R < \gamma_T$	$\eta_R < 0$	$0 < \varphi_R < \frac{\pi}{2}$	$\pi > \Delta\varphi_{saddle} > 0$
2.	$Q > 0$	$\gamma_R > \gamma_T$	$\eta_R > 0$	$\frac{\pi}{2} < \varphi_R < \pi$	$0 > \Delta\varphi_{saddle} > -\pi$
3.	$Q < 0$	$\gamma_R < \gamma_T$	$\eta_R < 0$	$-\pi < \varphi_R < -\frac{\pi}{2}$	$\pi > \Delta\varphi_{saddle} > 0$
4.	$Q < 0$	$\gamma_R > \gamma_T$	$\eta_R > 0$	$-\frac{\pi}{2} < \varphi_R < 0$	$0 > \Delta\varphi_{saddle} > -\pi$

3.11 Orbits

From Eq. (3.29), one concludes for $Q > 0$, $\eta_R < 0$, and $0 < \varphi_R < \pi/2$, if only the upper half of the orbit is considered, that

$$\Delta W = \sqrt{\frac{2W_R\beta_R^2}{-\eta_R}\, H + \frac{Q\hat{V}}{T_R}\frac{2W_R\beta_R^2}{-\eta_R}\left(\frac{1}{\omega_{RF}}\left[\cos(\omega_{RF}\Delta t + \varphi_R) - \cos\varphi_R\right] + \Delta t\ \sin\varphi_R\right)}.$$

Using Eq. (3.32) leads to

$$\Delta W = \Delta W_{max,stat}\sqrt{\frac{\pi h}{Q\hat{V}}\, H + \frac{1}{2}\left(\cos(\omega_{RF}\Delta t + \varphi_R) - \cos\varphi_R + \omega_{RF}\Delta t\ \sin\varphi_R\right)}. \tag{3.35}$$

For $H = H_s$, the first term under the square root sign equals $\frac{1}{2}(2\cos\varphi_R + (2\varphi_R - \pi)\ \sin\varphi_R) = [Y(\varphi_R)]^2$, as Eqs. (3.30) and (3.34) show. For $\Delta t = 0$, it follows that $\Delta W = \Delta W_{max,stat} Y(\varphi_R) = \Delta W_{max}$, as expected.

If one wants to draw an orbit inside the separatrix, one inserts a value $\tilde{H} < [Y(\varphi_R)]^2$ for $\tilde{H} = \frac{\pi h}{Q\hat{V}}\, H$. For trajectories outside the separatrix, however, one sets $\tilde{H} > [Y(\varphi_R)]^2$. The orbits themselves are obtained if ΔW as a function of $\Delta\varphi_{RF} = \omega_{RF}\Delta t$ is plotted according to Eq. (3.35). Of course, one obtains only the upper half of the orbits in this case.

Figure 3.5 shows a typical sketch of the longitudinal phase space. It clearly shows the characteristics that were discussed in Sect. 2.11.4. The stable fixed point in the middle is a center. The smaller the orbits around this center, the more closely they resemble an ellipse. The unstable fixed point is a saddle point. The orbits that approach this saddle point define the separatrix. Inside the separatrix, the orbits are closed. Outside the separatrix, the orbits are unstable. The region inside the separatrix is also called a **bucket**, since the particles that are inside will remain inside (provided that parameters such as the amplitude $\hat{V}$ are not changed rapidly). Only parts of the bucket are usually filled with particles. The region that is filled with particles is called a **bunch**.

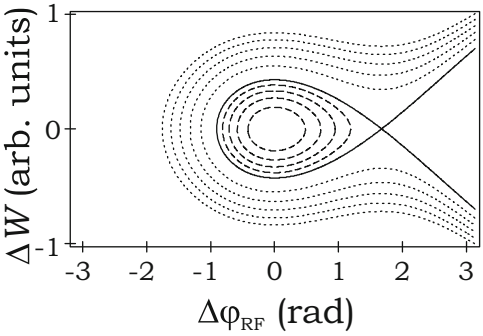

Fig. 3.5 Orbits in longitudinal phase space, acceleration with $\varphi_R = 42°$

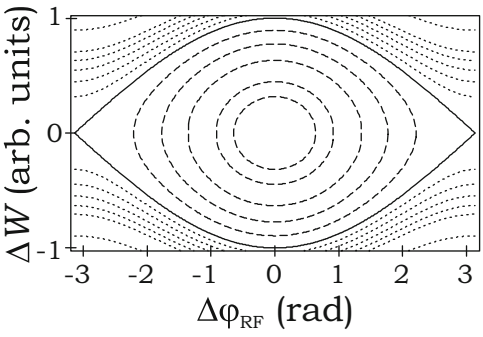

Fig. 3.6 Orbits in longitudinal phase space, stationary case ($\varphi_R = 0°$)

The bucket shown in Fig. 3.5 corresponds to a positive reference phase $\varphi_R = 42°$, i.e., to the acceleration case; one sometimes calls this a fishlike bucket. In the stationary case, defined by $\varphi_R = 0°$ for $Q > 0$ and $\eta_R < 0$, no overall acceleration takes place. The stationary bucket leads to an eye-shaped separatrix (see Fig. 3.6) that is symmetric with respect to both the energy and time axes. The bucket height of the stationary bucket is given by $2\Delta W_{\mathrm{max,stat}}$, according to Eq. (3.32), whereas the bucket height for the acceleration case is $2\Delta W_{\mathrm{max}}$, according to Eq. (3.33).

Figure 3.7 shows the same two cases for a larger range of $\Delta\varphi_{RF}$ values. It is obvious that the fixed points occur periodically. The buckets for the stationary case are larger than those for the acceleration case if all parameters except φ_R are kept constant. This effect will be analyzed in the Sect. 3.12 below. In the acceleration case, the energy deviation ΔW of a particle on an unstable orbit (outside the separatrix) will grow without bound, as Fig. 3.7 shows. The same is therefore valid for its deviation from the reference orbit; it will sooner or later hit the beam pipe walls, which is, of course, undesirable. Therefore, one always has to minimize the number of particles that are not captured inside a bucket.

Strictly speaking, the trajectories shown in Figs. 3.5, 3.6, and 3.7 are obtained only if an autonomous system is assumed, i.e., if no parameters (such as voltage $\hat{V}$) change with time. We will see later, however (cf. Sect. 5.3), that slow changes of control parameters are acceptable.

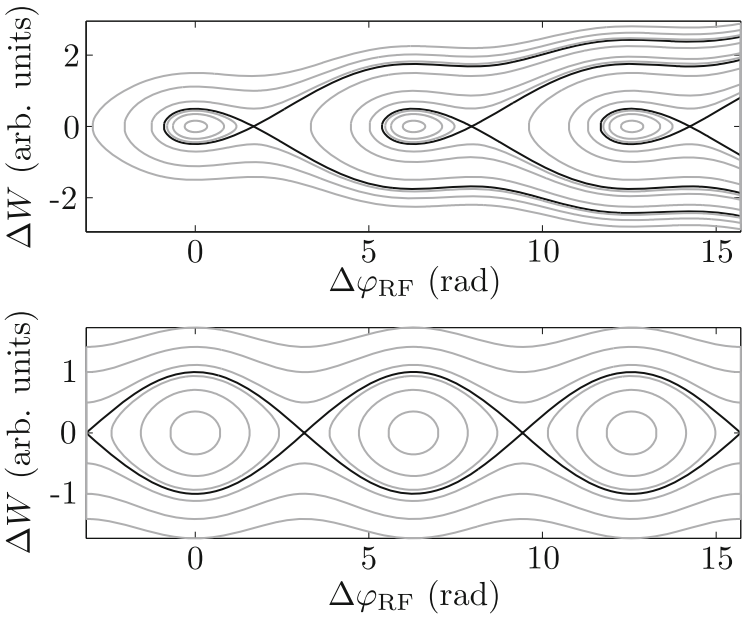

Fig. 3.7 Orbits in longitudinal phase space (*top*: acceleration with $\varphi_R = 42°$, *bottom*: stationary case $\varphi_R = 0°$)

3.12 Bucket Area

Now we want to determine the area inside the separatrix. First of all, we consider the stationary case with $\eta_R Q < 0$ and $\varphi_R = 0$ (or $\eta_R Q > 0$ and $\varphi_R = \pm\pi$, respectively). For the separatrix, we then obtain, using Eq. (3.31),

$$\Delta W = \left[\frac{2W_R \beta_R^2 |Q| \hat{V}}{|\eta_R| T_R \omega_{RF}} \left(\cos(\omega_{RF} \Delta t) + 1\right) \right]^{1/2} . \tag{3.36}$$

The value $\Delta W = 0$ is achieved for $\Delta\varphi_{RF} = -\pi$ and $\Delta\varphi_{RF} = +\pi$, so these values are the integration limits. Due to

$$\sqrt{1 + \cos \Delta\varphi_{RF}} = \sqrt{1 + \cos^2 \frac{\Delta\varphi_{RF}}{2} - \sin^2 \frac{\Delta\varphi_{RF}}{2}} = \sqrt{2} \cos \frac{\Delta\varphi_{RF}}{2},$$

one obtains

$$\begin{aligned} \int_{-\pi}^{+\pi} \sqrt{1 + \cos \Delta\varphi_{RF}} \, \mathrm{d}\Delta\varphi_{RF} &= 2 \int_0^{+\pi} \sqrt{2} \cos \frac{\Delta\varphi_{RF}}{2} \, \mathrm{d}\Delta\varphi_{RF} \\ &= 4\sqrt{2} \left[\sin \frac{\Delta\varphi_{RF}}{2} \right]_0^{\pi} = 4\sqrt{2}, \end{aligned} \tag{3.37}$$

and it follows that

$$A_{\mathrm{B,stat}}^{\Delta\varphi_{\mathrm{RF}},\Delta W}/2 = \int_{-\pi}^{+\pi} \Delta W \, \mathrm{d}\Delta\varphi_{\mathrm{RF}} = \sqrt{\frac{2W_{\mathrm{R}}\beta_{\mathrm{R}}^2|Q|\hat{V}}{|\eta_{\mathrm{R}}|T_{\mathrm{R}}\omega_{\mathrm{RF}}}}\, 4\sqrt{2}$$

$$\Rightarrow A_{\mathrm{B,stat}}^{\Delta\varphi_{\mathrm{RF}},\Delta W} = 8\sqrt{2}\,\sqrt{\frac{2W_{\mathrm{R}}\beta_{\mathrm{R}}^2|Q|\hat{V}}{|\eta_{\mathrm{R}}|T_{\mathrm{R}}\omega_{\mathrm{RF}}}}.$$

Since we calculated the integration with respect to $\Delta\varphi_{\mathrm{RF}} = \omega_{\mathrm{RF}}\Delta t$, this is the bucket area in $(\Delta W,\Delta\varphi_{\mathrm{RF}})$ coordinates. In order to obtain the bucket area in $(\Delta W,\Delta t)$ coordinates, we have to divide by ω_{RF}:

$$\boxed{A_{\mathrm{B,stat}}^{\Delta t,\Delta W} = \frac{4\sqrt{2}}{\pi h}T_{\mathrm{R}}\,\sqrt{\frac{W_{\mathrm{R}}\beta_{\mathrm{R}}^2|Q|\hat{V}}{\pi h|\eta_{\mathrm{R}}|}} = \frac{8\sqrt{2}}{\omega_{\mathrm{RF}}}\,\sqrt{\frac{W_{\mathrm{R}}\beta_{\mathrm{R}}^2|Q|\hat{V}}{\pi h|\eta_{\mathrm{R}}|}}.} \tag{3.38}$$

This equation corresponds to formula (2.75) in Edwards/Syphers [4].

If φ_{R} is not equal to zero, we have to integrate ΔW in Eq. (3.31) instead of ΔW in Eq. (3.36). This integration can be performed for the case $Q > 0$, $\eta_{\mathrm{R}} < 0$, $0 < \varphi_{\mathrm{R}} < \pi/2$ in the following, because the result will not depend on this choice, due to the symmetry as discussed in Sect. 3.10. In order to find the integration limits, we first have to determine the zeros of ΔW as a function of $\Delta\varphi_{\mathrm{RF}}$. The coefficients of $\sin\varphi_{\mathrm{R}}$ in parentheses in Eq. (3.31) become zero for

$$\Delta\varphi_{\mathrm{RF}} = \Delta\varphi_{\mathrm{saddle}} = \pi - 2\varphi_{\mathrm{R}}.$$

For this value, the remaining terms in parentheses also vanish; this obviously corresponds to the upper integration limit. In order to find the lower integration limit $\Delta\varphi_{\mathrm{RF}} = \Delta\varphi_{\mathrm{l}}$, the equation

$$\sin\varphi_{\mathrm{R}}\,(\sin\Delta\varphi_{\mathrm{l}} + \pi - 2\varphi_{\mathrm{R}} - \Delta\varphi_{\mathrm{l}}) = \cos\varphi_{\mathrm{R}}\,(\cos\Delta\varphi_{\mathrm{l}} + 1) \tag{3.39}$$

(which is obtained by setting the expression in parentheses equal to zero) has to be solved numerically with $\Delta\varphi_{\mathrm{l}}$ in the range between $-\pi$ and 0. For this purpose, one may determine the zero of the function

$$f(\Delta\varphi_{\mathrm{l}}) = \sin\varphi_{\mathrm{R}}\,(\sin\Delta\varphi_{\mathrm{l}} + \pi - 2\varphi_{\mathrm{R}} - \Delta\varphi_{\mathrm{l}}) - \cos\varphi_{\mathrm{R}}\,(\cos\Delta\varphi_{\mathrm{l}} + 1)$$

by means of an interval division method, since for $0 < \varphi_{\mathrm{R}} < \pi/2$, the function is always positive at $\Delta\varphi_{\mathrm{l}} = -\pi$, whereas it is always negative at $\Delta\varphi_{\mathrm{l}} = +\pi$.

The resulting function $\Delta\varphi_\mathrm{l}(\varphi_\mathrm{R})$ is shown in Fig. 3.9. It may be extended easily to the range $-\pi \le \varphi_\mathrm{R} \le +\pi$ (see Sect. 3.10), because $\Delta\varphi_\mathrm{l}(\varphi_\mathrm{R})$ is an odd function.

If one now constructs the integral

$$A_\mathrm{B}^{\Delta\varphi_\mathrm{RF},\Delta W}/2 = \int_{\Delta\varphi_\mathrm{l}}^{\Delta\varphi_\mathrm{saddle}} \Delta W \, \mathrm{d}\Delta\varphi_\mathrm{RF} = \int_{\Delta\varphi_\mathrm{l}}^{\pi-2\varphi_\mathrm{R}} \Delta W \, \mathrm{d}\Delta\varphi_\mathrm{RF},$$

one sees by comparison with Eqs. (3.31), (3.36), and (3.37) that it is smaller than that for the stationary case by a factor of

$$\alpha(\varphi_\mathrm{R}) = \frac{1}{4\sqrt{2}} \cdot \cdot \int_{\Delta\varphi_\mathrm{l}}^{\pi-2\varphi_\mathrm{R}} \sqrt{(2\varphi_\mathrm{R}+\Delta\varphi_\mathrm{RF}-\pi)\sin\varphi_\mathrm{R}+\cos\varphi_\mathrm{R}+\cos(\varphi_\mathrm{R}+\Delta\varphi_\mathrm{RF})} \, \mathrm{d}\Delta\varphi_\mathrm{RF}. \tag{3.40}$$

The denominator of the first fraction equals the integral for the stationary case with $\varphi_\mathrm{R} = 0$, as shown above. We obtain

$$A_\mathrm{B}^{\Delta\varphi_\mathrm{RF},\Delta W} = A_\mathrm{B,stat}^{\Delta\varphi_\mathrm{RF},\Delta W} \, \alpha(\varphi_\mathrm{R}).$$

The conversion to $(\Delta W, \Delta t)$ coordinates leads to the same factor on both sides, and we obtain

$$\boxed{A_\mathrm{B}^{\Delta t,\Delta W} = A_\mathrm{B,stat}^{\Delta t,\Delta W} \, \alpha(\varphi_\mathrm{R}) = \frac{4\sqrt{2}}{\pi h} T_\mathrm{R} \sqrt{\frac{W_\mathrm{R}\beta_\mathrm{R}^2|Q|\hat{V}}{\pi h|\eta_\mathrm{R}|}} \, \alpha(\varphi_\mathrm{R}).} \tag{3.41}$$

By means of a numerical calculation of the functions $\Delta\varphi_\mathrm{l}(\varphi_\mathrm{R})$ and $\alpha(\varphi_\mathrm{R})$ and subsequent fitting by suitable functions, the following approximations may be found:

$$\Delta\varphi_\mathrm{l}(\varphi_\mathrm{R}) \approx \pi \left[\left(\frac{2}{\pi}\varphi_\mathrm{R} \right)^{0.4373} - 1 \right], \tag{3.42}$$

$$\alpha(\varphi_\mathrm{R}) \approx 1 - \sin\varphi_\mathrm{R} - a\left(\frac{2}{\pi}\varphi_\mathrm{R}\right) + 2a\left(\frac{2}{\pi}\varphi_\mathrm{R}\right)^2 - a\left(\frac{2}{\pi}\varphi_\mathrm{R}\right)^3 \qquad \text{with } a = 1.0879. \tag{3.43}$$

The accuracy of the approximation for $\Delta\varphi_\mathrm{l}$ is better than $\pm 2°$. The accuracy of the approximation for α in the range $0 \le \varphi_\mathrm{R} \le 85°$ relevant in practical applications is better than $+5\%$ or -16%, respectively. Figure 3.8 shows the dependence of α on the synchronous phase φ_R. Please note that the function $\alpha(\varphi_\mathrm{R})$ may easily be extended to the range $-\pi \le \varphi_\mathrm{R} \le +\pi$, because it is an even function. The RF phase that corresponds to one end of the bucket, $\Delta\varphi_\mathrm{RF} = \Delta\varphi_\mathrm{l}$, is shown in

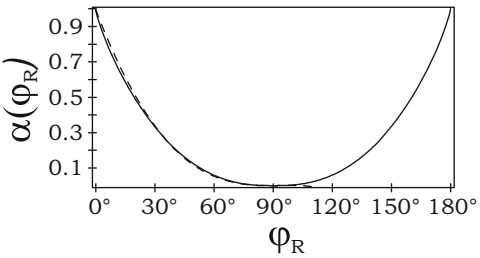

Fig. 3.8 Bucket area reduction factor $\alpha(\varphi_R)$; *solid line*: Eq. (3.40), *dashed line*: approximation (3.43)

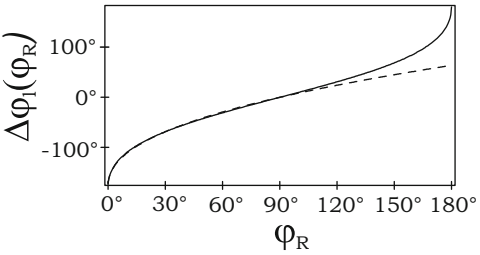

Fig. 3.9 RF phase $\Delta\varphi_l(\varphi_R)$; *solid line*: solution of Eq. (3.39), *dashed line*: approximation (3.42)

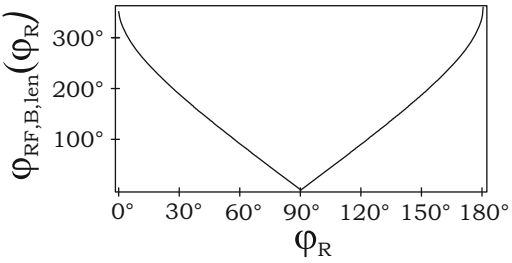

Fig. 3.10 Bucket length $\varphi_{RF,B,len}(\varphi_R)$

Fig. 3.9; it is an odd function $\Delta\varphi_l(\varphi_R)$. Since the other end of the bucket is given by $\Delta\varphi_{RF} = \Delta\varphi_{saddle} = \pi \text{ sign}\,\varphi_R - 2\varphi_R$, the bucket length is

$$\varphi_{RF,B,len} = |\Delta\varphi_{saddle} - \Delta\varphi_l| = |\pi \text{ sign}\,\varphi_R - 2\varphi_R - \Delta\varphi_l|.$$

This function is shown in Fig. 3.10; it is an even function (see Sect. 3.10). In all these diagrams, the left half of the curves ($0 \le \varphi_R \le 90°$) corresponds to case 1 (stationary case for $\varphi_R = 0$), whereas the right half ($90° \le \varphi_R \le 180°$) corresponds to case 2 (stationary case for $\varphi_R = \pi$) discussed in Sect. 3.10.

3.13 Approximation of Bucket Area

If we simply approximate the shape of the stationary bucket by an ellipse, we obtain the formula

$$A_{B,stat}^{\Delta t,\Delta W} \approx \pi \Delta t_{max,stat} \Delta W_{max,stat}.$$

By means of

$$\Delta t_{\mathrm{max,stat}} = \frac{T_\mathrm{R}}{2h}$$

and the bucket height (3.32),

$$\Delta W_{\mathrm{max,stat}} = \sqrt{\frac{2\ W_\mathrm{R}\beta_\mathrm{R}^2|Q|\hat{V}}{\pi h|\eta_\mathrm{R}|}},$$

it follows that

$$A_{\mathrm{B,stat}}^{\Delta t,\Delta W} \approx \frac{\pi}{\sqrt{2}}\ \frac{T_\mathrm{R}}{h}\ \sqrt{\frac{W_\mathrm{R}\beta_\mathrm{R}^2|Q|\hat{V}}{\pi h|\eta_\mathrm{R}|}}.$$

The coefficient $\frac{\pi}{\sqrt{2}} \approx 2.22$ does not differ much from that of the exact solution ($\frac{4\sqrt{2}}{\pi} \approx 1.80$; see Eq. (3.38)). The relative deviation is about 23%.

3.14 Ratio of Bucket Height to Bucket Length

The bucket length $2\Delta t_{\mathrm{max,stat}}$ of a stationary bucket is obviously given by the condition

$$\omega_\mathrm{RF}\Delta t_{\mathrm{max,stat}} = \pi.$$

Together with Eq. (3.32), this leads to

$$\frac{\Delta W_{\mathrm{max,stat}}}{\Delta t_{\mathrm{max,stat}}} = \sqrt{\frac{2\ W_\mathrm{R}\beta_\mathrm{R}^2\ |Q|\ \hat{V}}{\pi h\ |\eta_\mathrm{R}|}}\ \frac{\omega_\mathrm{RF}}{\pi}.$$

We may now use Eq. (3.25) for the stationary case,

$$f_{\mathrm{S,0,stat}} = f_\mathrm{R}\sqrt{\frac{|\eta_\mathrm{R}\ Q|\ \hat{V}\ h}{2\pi W_\mathrm{R}\beta_\mathrm{R}^2}} \qquad \Rightarrow |Q|\hat{V} = \frac{2\pi W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|\ h}\left(\frac{f_{\mathrm{S,0,stat}}}{f_\mathrm{R}}\right)^2,$$

and obtain

$$\frac{\Delta W_{\mathrm{max,stat}}}{\Delta t_{\mathrm{max,stat}}} = \frac{2W_{\mathrm{R}}\beta_{\mathrm{R}}^2}{|\eta_{\mathrm{R}}|\,h}\,\frac{f_{\mathrm{S,0,stat}}}{f_{\mathrm{R}}}\,\frac{\omega_{\mathrm{RF}}}{\pi}$$

$$\Rightarrow \boxed{\frac{\Delta W_{\mathrm{max,stat}}}{\Delta t_{\mathrm{max,stat}}} = f_{\mathrm{S,0,stat}}\,\frac{4W_{\mathrm{R}}\beta_{\mathrm{R}}^2}{|\eta_{\mathrm{R}}|}.} \tag{3.44}$$

If we compare this with the ratio of the principal axes for small oscillation amplitudes

$$\frac{\Delta\hat{W}}{\Delta\hat{t}} = f_{\mathrm{S,0,stat}}\,\frac{2\pi W_{\mathrm{R}}\beta_{\mathrm{R}}^2}{|\eta_{\mathrm{R}}|}$$

according to Eq. (3.28), we see that the expressions differ only by the factor 4 vs. 2π. In comparison with small-amplitude orbits, the bucket shape is flattened. This is also visible in Fig. 3.6.

Equation (3.25) for the synchrotron frequency

$$f_{\mathrm{S,0,stat}} = f_{\mathrm{R}}\sqrt{\frac{|\eta_{\mathrm{R}}\;Q|\hat{V}\;h}{2\pi W_{\mathrm{R}}\beta_{\mathrm{R}}^2}}$$

in a stationary bucket ($\varphi_{\mathrm{R}} = 0$ or $\varphi_{\mathrm{R}} = \pm\pi$, respectively) and Eq. (3.38) for the bucket area

$$A_{\mathrm{B,stat}}^{\Delta t,\Delta W} = \frac{4\sqrt{2}}{\pi h}T_{\mathrm{R}}\sqrt{\frac{W_{\mathrm{R}}\beta_{\mathrm{R}}^2\;|Q|\;\hat{V}}{\pi h\;|\eta_{\mathrm{R}}|}}$$

have similar expressions. Therefore, we get

$$\frac{f_{\mathrm{S,0,stat}}}{A_{\mathrm{B,stat}}^{\Delta t,\Delta W}} = \frac{\pi h f_{\mathrm{R}}^2}{4\sqrt{2}}\,\frac{|\eta_{\mathrm{R}}|\;h}{\sqrt{2}\;W_{\mathrm{R}}\beta_{\mathrm{R}}^2} = \frac{\pi f_{\mathrm{RF}}^2\;|\eta_{\mathrm{R}}|}{8W_{\mathrm{R}}\beta_{\mathrm{R}}^2}. \tag{3.45}$$

This allows us to write the principal axes ratio given by Eq. (3.28),

$$\frac{\Delta\hat{W}}{\Delta\hat{t}} = f_{\mathrm{S,0,stat}}\,\frac{2\pi W_{\mathrm{R}}\beta_{\mathrm{R}}^2}{|\eta_{\mathrm{R}}|},$$

in the form

$$\boxed{\frac{\Delta\hat{W}}{\Delta\hat{t}} = A_{\mathrm{B,stat}}^{\Delta t,\Delta W}\,\frac{\pi^2}{4}\,f_{\mathrm{RF}}^2 = A_{\mathrm{B,stat}}^{\Delta t,\Delta W}\,\frac{\omega_{\mathrm{RF}}^2}{16}.} \tag{3.46}$$

This leads us to the following expression for the bunch area if only small bunches are considered:

$$A_{\mathrm{b,stat}}^{\Delta t,\Delta W} \approx \pi \Delta\hat{t}\,\Delta\hat{W} = \pi \Delta\hat{t}^2 A_{\mathrm{B,stat}}^{\Delta t,\Delta W} \frac{\omega_{\mathrm{RF}}^2}{16} = \frac{\pi}{16} \Delta\hat{\varphi}_{\mathrm{RF}}^2 A_{\mathrm{B,stat}}^{\Delta t,\Delta W}.$$

Here $\Delta\hat{t}$ and $\Delta\hat{W}$ are interpreted as the semiaxes of the outermost particle in the bunch. If we define the bucket filling factor η_{fill} by

$$\eta_{\mathrm{fill}} = \frac{A_{\mathrm{b,stat}}^{\Delta t,\Delta W}}{A_{\mathrm{B,stat}}^{\Delta t,\Delta W}},$$

we obtain

$$\eta_{\mathrm{fill}} \approx \frac{\pi}{16} \Delta\hat{\varphi}_{\mathrm{RF}}^2 \tag{3.47}$$

for sufficiently small bunches. In Sect. 3.17, an exact expression for arbitrary bunch sizes will be derived.

For later use, we write Eq. (3.45) in the form

$$\boxed{\frac{\omega_{\mathrm{S,0,stat}}}{A_{\mathrm{B,stat}}^{\Delta t,\Delta W}} = \frac{\omega_{\mathrm{RF}}^2 \,|\eta_{\mathrm{R}}|}{16\, W_{\mathrm{R}} \beta_{\mathrm{R}}^2}.} \tag{3.48}$$

From now on, we will often discard the superscript specification of the phase space coordinates. If nothing else is stated, we assume

$$A_{\mathrm{B}} := A_{\mathrm{B}}^{\Delta t,\Delta W}, \qquad A_{\mathrm{B,stat}} := A_{\mathrm{B,stat}}^{\Delta t,\Delta W}, \qquad A_{\mathrm{b}} := A_{\mathrm{b}}^{\Delta t,\Delta W}, \qquad A_{\mathrm{b,stat}} := A_{\mathrm{b,stat}}^{\Delta t,\Delta W}.$$

3.15 Choice of the Harmonic Number

According to Eq. (3.38), the total area of h buckets,

$$h \cdot A_{\mathrm{B,stat}} = \frac{4\sqrt{2}}{\pi} T_{\mathrm{R}} \sqrt{\frac{W_{\mathrm{R}} \beta_{\mathrm{R}}^2 |Q| \hat{V}}{\pi h |\eta_{\mathrm{R}}|}},$$

remains constant if both the amplitude $\hat{V}$ and the harmonic number h are increased by the same factor.

If we assume, however, that the **ramp rate** $\dot{B}$ is the same, the required accelerating voltage V_{R} will also remain the same, as Eq. (3.22),

$$V_{\mathrm{R}} \approx l_{\mathrm{R}} r_{\mathrm{R}} \dot{B}, \tag{3.49}$$

shows. Therefore, an increase in $\hat{V}$ and h by the same factor will lead to a reduced reference phase φ_{R}, since

$$V_{\mathrm{R}} = \hat{V} \sin \varphi_{\mathrm{R}} \tag{3.50}$$

will remain constant. This leads to a larger value of the bucket reduction factor $\alpha(\varphi_{\mathrm{R}})$, which results in a larger area

$$h \cdot A_{\mathrm{B}} = h \cdot A_{\mathrm{B,stat}} \alpha(\varphi_{\mathrm{R}})$$

of the accelerated buckets.

In conclusion, for the acceleration case, the simultaneous increase of RF amplitude and harmonic number by the same factor leads to a larger area in phase space, since the reference phase is reduced.

3.16 Revolution Time in the Stationary Bucket

Let us consider the Hamiltonian in Eq. (3.29) for the stationary case with $Q > 0$, $\eta_{\mathrm{R}} < 0$, $\varphi_{\mathrm{R}} = 0$:

$$H = -\frac{\eta_{\mathrm{R}}}{W_{\mathrm{R}} \beta_{\mathrm{R}}^2} \frac{\Delta W^2}{2} - \frac{Q \hat{V}}{T_{\mathrm{R}}} \frac{1}{\omega_{\mathrm{RF}}} \left[\cos(\omega_{\mathrm{RF}} \Delta t) - 1\right].$$

Setting

$$A = -\frac{\eta_{\mathrm{R}}}{2\, W_{\mathrm{R}} \beta_{\mathrm{R}}^2} \tag{3.51}$$

leads to

$$\Delta W^2 = \frac{1}{A} \left(H + \frac{Q \hat{V}}{T_{\mathrm{R}}} \frac{1}{\omega_{\mathrm{RF}}} \left[\cos(\omega_{\mathrm{RF}} \Delta t) - 1\right] \right).$$

One obtains the action variable

$$\begin{aligned} J &= \frac{1}{2\pi} \int \int \mathrm{d}(\Delta t)\, \mathrm{d}(\Delta W) \\ &= \frac{1}{\pi \sqrt{A}} \int_{-\Delta \hat{t}}^{\Delta \hat{t}} \sqrt{H + \frac{Q \hat{V}}{T_{\mathrm{R}}} \frac{1}{\omega_{\mathrm{RF}}} \left[\cos(\omega_{\mathrm{RF}} \Delta t) - 1\right]}\, \mathrm{d}(\Delta t). \end{aligned} \tag{3.52}$$

By means of the substitution

$$\Delta\varphi_{\mathrm{RF}} = \omega_{\mathrm{RF}}\Delta t, \qquad \frac{\mathrm{d}\Delta\varphi_{\mathrm{RF}}}{\mathrm{d}\Delta t} = \omega_{\mathrm{RF}},$$

we obtain, since the integrand is an even function,

$$J = \frac{2}{\pi\omega_{\mathrm{RF}}\sqrt{A}} \int_0^{\Delta\hat{\varphi}_{\mathrm{RF}}} \sqrt{H + \frac{Q\hat{V}}{T_{\mathrm{R}}} \frac{1}{\omega_{\mathrm{RF}}} [\cos\ \Delta\varphi_{\mathrm{RF}} - 1]}\ \mathrm{d}(\Delta\varphi_{\mathrm{RF}}).$$

In order to convert this into a well-known elliptic integral, we make use of the trigonometric identity

$$\cos\ \Delta\varphi_{\mathrm{RF}} = \cos^2 \frac{\Delta\varphi_{\mathrm{RF}}}{2} - \sin^2 \frac{\Delta\varphi_{\mathrm{RF}}}{2} = 1 - 2\ \sin^2 \frac{\Delta\varphi_{\mathrm{RF}}}{2}$$

and get

$$J = \frac{2}{\pi\omega_{\mathrm{RF}}\sqrt{A}} \int_0^{\Delta\hat{\varphi}_{\mathrm{RF}}} \sqrt{H - 2\ \frac{Q\hat{V}}{T_{\mathrm{R}}\ \omega_{\mathrm{RF}}} \sin^2 \frac{\Delta\varphi_{\mathrm{RF}}}{2}}\ \mathrm{d}(\Delta\varphi_{\mathrm{RF}}).$$

If we now set

$$B = 2\ \frac{Q\hat{V}}{T_{\mathrm{R}}\ \omega_{\mathrm{RF}}} = \frac{Q\hat{V}}{h\ \pi} \tag{3.53}$$

and

$$x = \frac{\Delta\varphi_{\mathrm{RF}}}{2}, \qquad \frac{\mathrm{d}x}{\mathrm{d}\Delta\varphi_{\mathrm{RF}}} = \frac{1}{2},$$

we obtain

$$J = \frac{4}{\pi\omega_{\mathrm{RF}}\sqrt{A}} \int_0^{x_{\max}} \sqrt{H - B\ \sin^2 x}\ \mathrm{d}x.$$

Obviously,

$$x_{\max} = \arcsin\sqrt{\frac{H}{B}}$$

is valid, because the integrand must vanish where the orbits cross the x-axis. One obtains

$$\frac{\mathrm{d}J}{\mathrm{d}H} = \frac{4}{\pi\omega_{\mathrm{RF}}\sqrt{A}}\left[\frac{\mathrm{d}x_{\max}}{\mathrm{d}H}\sqrt{H - B\ \sin^2 x_{\max}} + \int_0^{x_{\max}} \frac{\mathrm{d}x}{2\sqrt{H - B\ \sin^2 x}}\right].$$

The first term equals zero:

$$\frac{\mathrm{d}J}{\mathrm{d}H} = \frac{2}{\pi\omega_{\mathrm{RF}}\sqrt{HA}}\int_0^{x_{\max}} \frac{\mathrm{d}x}{\sqrt{1 - \frac{B}{H}\ \sin^2 x}}.$$

For

$$k^2 = \frac{B}{H},$$

one obtains an elliptic integral of the first kind:

$$\frac{\mathrm{d}J}{\mathrm{d}H} = \frac{2}{\pi\omega_{\mathrm{RF}}\sqrt{HA}}\,\mathrm{F}(x_{\max}, k). \tag{3.54}$$

Obviously,

$$x_{\max} = \arcsin\frac{1}{k}$$

is valid. Now we make use of formula 17.4.15 in Abramowitz/Stegun [10]:

$$\mathrm{F}(\varphi, k) = \frac{1}{k}\,\mathrm{F}\left(\arcsin(k\sin\varphi), \frac{1}{k}\right)$$

$$\Rightarrow \mathrm{F}(x_{\max}, k) = \frac{1}{k}\,\mathrm{F}\left(\arcsin(1), \frac{1}{k}\right).$$

Because of

$$\arcsin(1) = \frac{\pi}{2},$$

this corresponds to a complete elliptic integral of the first kind:

$$\mathrm{F}(x_{\max}, k) = \frac{1}{k}\,\mathrm{F}\left(\frac{\pi}{2}, \frac{1}{k}\right) = \frac{1}{k}\,\mathrm{K}\left(\frac{1}{k}\right).$$

Equation (3.54) therefore yields

$$\frac{\mathrm{d}J}{\mathrm{d}H} = \frac{2}{\pi\omega_{\mathrm{RF}}\sqrt{BA}}\,\mathrm{K}\left(\sqrt{\frac{H}{B}}\right).$$

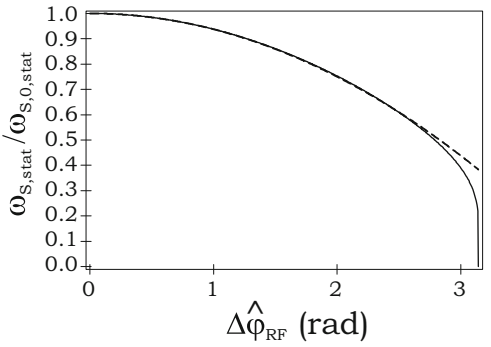

Fig. 3.11 Decrease of synchrotron frequency vs. oscillation amplitude (*solid line*: exact formula (3.55), *dashed line*: approximation (3.56))

We see from Eqs. (3.51), (3.53), and (3.25) that

$$AB = -\frac{\eta_R}{2\,W_R\beta_R^2}\frac{Q\hat{V}}{h\,\pi} = \left(\frac{\omega_{S,0,stat}}{\omega_R}\right)^2\frac{1}{h^2} = \left(\frac{\omega_{S,0,stat}}{\omega_{RF}}\right)^2$$

is valid, so that

$$\frac{dJ}{dH} = \frac{2}{\pi\omega_{S,0,stat}}\,K\left(\sqrt{\frac{H}{B}}\right)$$

$$\Rightarrow \omega_{S,stat} = \frac{dH}{dJ} = \omega_{S,0,stat}\frac{\pi}{2\,K\left(\sqrt{\frac{H}{B}}\right)}$$

results. Instead of the dependence on the value of the Hamiltonian, the dependence on the maximum phase deviation of the particles is of course more transparent. Due to

$$x_{max} = \arcsin\sqrt{\frac{H}{B}} = \frac{\Delta\hat{\varphi}_{RF}}{2},$$

one finally obtains

$$\boxed{\omega_{S,stat} = \omega_{S,0,stat}\frac{\pi}{2\,K\left(\sin\frac{\Delta\hat{\varphi}_{RF}}{2}\right)}.} \tag{3.55}$$

This formula for the synchrotron frequency of particles with arbitrary oscillation amplitude may also be found in Lee [5, p. 240] or Ng [1, Sect. 2.1.2.1]. The decrease in the oscillation frequency is shown in Fig. 3.11. For zero argument, we have

$$\mathrm{K}(0) = \frac{\pi}{2},$$

and

$$\omega_{\mathrm{S,stat}} = \omega_{\mathrm{S,0,stat}}$$

results for small phase amplitudes $\Delta\hat{\varphi}_{\mathrm{RF}}$, as expected. By means of a Taylor series expansion of Eq. (3.55), one obtains

$$\frac{\omega_{\mathrm{S,stat}}}{\omega_{\mathrm{S,0,stat}}} \approx 1 - \frac{\Delta\hat{\varphi}_{\mathrm{RF}}^2}{16} \tag{3.56}$$

for sufficiently small amplitudes $\Delta\hat{\varphi}_{\mathrm{RF}}$. This, however, does not restrict the validity significantly, because the error of this formula is below 5% if $\Delta\hat{\varphi}_{\mathrm{RF}} < 164°$ holds.

3.17 Bunch Area

In this section, we calculate the bunch area, i.e., the area that is enclosed by a particle trajectory in phase space. We restrict ourselves to the stationary case with $Q > 0$, $\eta_{\mathrm{R}} < 0$, $\varphi_{\mathrm{R}} = 0$. The bunch area already appeared in Eq. (3.52):

$$A_{\mathrm{b,stat}} = 2\pi J = \frac{2}{\sqrt{A}} \int_{-\Delta\hat{t}}^{\Delta\hat{t}} \sqrt{H + \frac{Q\hat{V}}{2\pi h}\left[\cos(\omega_{\mathrm{RF}}\Delta t) - 1\right]}\, \mathrm{d}(\Delta t).$$

According to Eq. (3.51), we have

$$A = -\frac{\eta_{\mathrm{R}}}{2\, W_{\mathrm{R}} \beta_{\mathrm{R}}^2}.$$

By means of the substitution

$$\Delta\varphi_{\mathrm{RF}} = \omega_{\mathrm{RF}}\Delta t \qquad \Rightarrow \frac{\mathrm{d}\Delta\varphi_{\mathrm{RF}}}{\mathrm{d}\Delta t} = \omega_{\mathrm{RF}},$$

one gets

$$\begin{aligned} A_{\mathrm{b,stat}} = 2\pi J &= \frac{2}{\omega_{\mathrm{RF}}\sqrt{A}} \int_{-\Delta\hat{\varphi}_{\mathrm{RF}}}^{\Delta\hat{\varphi}_{\mathrm{RF}}} \sqrt{H + \frac{Q\hat{V}}{2\pi h}\left[\cos(\Delta\varphi_{\mathrm{RF}}) - 1\right]}\, \mathrm{d}(\Delta\varphi_{\mathrm{RF}}) = \\ &= \frac{4}{\omega_{\mathrm{RF}}\sqrt{A}} \int_{0}^{\Delta\hat{\varphi}_{\mathrm{RF}}} \sqrt{a + b\, \cos\Delta\varphi_{\mathrm{RF}}}\, \mathrm{d}(\Delta\varphi_{\mathrm{RF}}) \end{aligned} \tag{3.57}$$

with

$$a = H - \frac{Q\hat{V}}{2\pi h}$$

and

$$b = \frac{Q\hat{V}}{2\pi h}. \tag{3.58}$$

The integration limit is defined by the integrand being zero:

$$H + \frac{Q\hat{V}}{2\pi h}\left[\cos(\Delta\hat{\varphi}_{\mathrm{RF}}) - 1\right] = 0$$

$$\Rightarrow a = -\frac{Q\hat{V}}{2\pi h}\ \cos(\Delta\hat{\varphi}_{\mathrm{RF}}) = -b\ \cos(\Delta\hat{\varphi}_{\mathrm{RF}}),$$

$$H = b + a = b\left(1 - \cos(\Delta\hat{\varphi}_{\mathrm{RF}})\right). \tag{3.59}$$

The integral in Eq. (3.57) is already well known from Sect. 2.11.9, Eq. (2.134). If we take the different definitions of the parameters a and b into account, we may adopt the solution (2.135). We get

$$A_{\mathrm{b,stat}} = \frac{4}{\omega_{\mathrm{RF}}\sqrt{A}}\sqrt{\frac{2}{b}}\left[(a-b)\mathrm{F}\left(\gamma, \frac{1}{r}\right) + 2b\mathrm{E}\left(\gamma, \frac{1}{r}\right)\right]_0^{\Delta\hat{\varphi}_{\mathrm{RF}}}$$

with

$$r = \sqrt{\frac{2b}{a+b}} \qquad \text{and} \qquad \gamma = \arcsin\sqrt{\frac{b(1-\cos\Delta\varphi_{\mathrm{RF}})}{a+b}}.$$

Here $\mathrm{E}(\gamma, k)$ is an elliptic integral of the second kind (see Footnote 23 on p. 108). For $\Delta\varphi_{\mathrm{RF}} = 0$, we get $\gamma = 0$, and the expression in the square brackets disappears. For $\Delta\varphi_{\mathrm{RF}} = \Delta\hat{\varphi}_{\mathrm{RF}}$, we get

$$\gamma = \arcsin\sqrt{\frac{b(1-\cos\Delta\hat{\varphi}_{\mathrm{RF}})}{a+b}}.$$

Because of Eq. (3.59), we therefore obtain

$$\gamma = \arcsin\ 1 = \frac{\pi}{2}.$$

Due to

$$\mathrm{F}\left(\frac{\pi}{2},k\right)=\mathrm{K}(k),\qquad \mathrm{E}\left(\frac{\pi}{2},k\right)=\mathrm{E}(k),$$

we obtain

$$A_{\mathrm{b,stat}}=\frac{4\sqrt{2}}{\omega_{\mathrm{RF}}\sqrt{Ab}}\left[(a-b)\mathrm{K}(k)+2b\mathrm{E}(k)\right]$$

if we set

$$k=\frac{1}{r}=\sqrt{\frac{a+b}{2b}}=\sqrt{\frac{H}{2b}}=\sqrt{\frac{1-\cos\Delta\hat{\varphi}_{\mathrm{RF}}}{2}}$$
$$\Rightarrow k=\sin\frac{\Delta\hat{\varphi}_{\mathrm{RF}}}{2}.$$

Since

$$a-b=H-2b=-b(1+\cos\Delta\hat{\varphi}_{\mathrm{RF}})$$

is valid, we obtain

$$A_{\mathrm{b,stat}}=\frac{4\sqrt{2}}{\omega_{\mathrm{RF}}\sqrt{Ab}}\,2b\left[\mathrm{E}(k)-\frac{1+\cos\Delta\hat{\varphi}_{\mathrm{RF}}}{2}\,\mathrm{K}(k)\right]$$
$$\Rightarrow A_{\mathrm{b,stat}}=\frac{8\sqrt{2}\,\sqrt{b}}{\omega_{\mathrm{RF}}\sqrt{A}}\left[\mathrm{E}(k)-\cos^2\frac{\Delta\hat{\varphi}_{\mathrm{RF}}}{2}\,\mathrm{K}(k)\right]$$
$$\Rightarrow A_{\mathrm{b,stat}}=\frac{8\sqrt{2}}{\omega_{\mathrm{RF}}}\sqrt{-\frac{Q\hat{V}\,W_{\mathrm{R}}\beta_{\mathrm{R}}^2}{\pi h\eta_{\mathrm{R}}}}\left[\mathrm{E}(k)-\mathrm{K}(k)\,\cos^2\frac{\Delta\hat{\varphi}_{\mathrm{RF}}}{2}\right].$$

In the last step, we used Eqs. (3.51) and (3.58). A comparison with Eq. (3.38) leads to

$$A_{\mathrm{b,stat}}=A_{\mathrm{B,stat}}\left[\mathrm{E}(k)-\mathrm{K}(k)\,\cos^2\frac{\Delta\hat{\varphi}_{\mathrm{RF}}}{2}\right]$$
$$\Leftrightarrow\ \boxed{A_{\mathrm{b,stat}}=A_{\mathrm{B,stat}}\,\eta_{\mathrm{fill}}}\tag{3.60}$$

with the bucket filling factor

$$\boxed{\eta_{\mathrm{fill}}(\Delta\hat{\varphi}_{\mathrm{RF}})=\mathrm{E}\left(\sin\frac{\Delta\hat{\varphi}_{\mathrm{RF}}}{2}\right)-\mathrm{K}\left(\sin\frac{\Delta\hat{\varphi}_{\mathrm{RF}}}{2}\right)\cos^2\frac{\Delta\hat{\varphi}_{\mathrm{RF}}}{2}.}\tag{3.61}$$

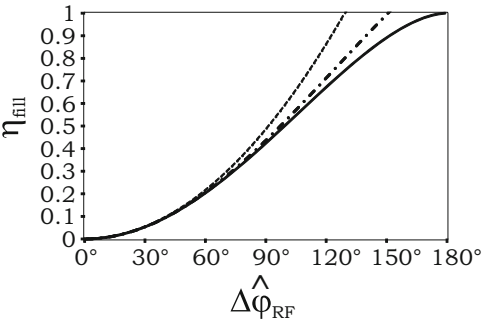

Fig. 3.12 Bucket filling factor (*solid line*: exact solution according to Eq. (3.61), *dashed line* (*uppermost curve*): approximation (3.62), *dash-dotted line*: improved approximation (3.65))

Please note that $\Delta\hat{\varphi}_{\text{RF}}$ denotes the RF phase amplitude at the border of the bunch, not the bucket limit.

Now we check our result for small values of $\Delta\hat{\varphi}_{\text{RF}}$. By means of the approximations

$$\mathrm{E}(k) \approx \frac{\pi}{2}\left(1-\frac{k^2}{4}\right) = \frac{\pi}{2}\left(1-\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{16}\right),$$

$$\mathrm{K}(k) \approx \frac{\pi}{2}\left(1+\frac{k^2}{4}\right) = \frac{\pi}{2}\left(1+\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{16}\right),$$

$$\cos^2\frac{\Delta\hat{\varphi}_{\text{RF}}}{2} = 1-\sin^2\frac{\Delta\hat{\varphi}_{\text{RF}}}{2} \approx 1-\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{4},$$

we obtain

$$\eta_{\text{fill}}(\Delta\hat{\varphi}_{\text{RF}}) \approx \frac{\pi}{2}\left[1-\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{16}-\left(1+\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{16}\right)\left(1-\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{4}\right)\right] \approx$$

$$\approx \frac{\pi}{2}\left[-\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{16}-\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{16}+\frac{\Delta\hat{\varphi}_{\text{RF}}^2}{4}\right]$$

$$\eta_{\text{fill}}(\Delta\hat{\varphi}_{\text{RF}}) \approx \frac{\pi}{16}\,\Delta\hat{\varphi}_{\text{RF}}^2. \tag{3.62}$$

As expected, this is equivalent to Eq. (3.47). Figure 3.12 shows the difference between the exact solution and this approximation. For $\Delta\hat{\varphi}_{\text{RF}} \leq 90°$, the error is below 15%. An improved approximation will be derived in Sect. 3.18.

3.18 Ratio of Bunch Height to Bunch Length

We now determine the ratio of bunch height to bunch length for the stationary case ($\varphi_R = 0$ or $\varphi_R = \pm\pi$). In this case, Eq. (3.29) reads

$$H = -\frac{\eta_R}{W_R \beta_R^2} \frac{\Delta W^2}{2} - \frac{Q\hat{V}}{T_R} \left(\frac{1}{\omega_{RF}} [\pm \cos(\omega_{RF} \Delta t) \mp 1] \right).$$

If we assume that the value of H is given for a specific closed trajectory that marks the outer limits of the bunch, we may calculate the bunch height $2\Delta W = 2\Delta\hat{W}$ by setting $\Delta t = 0$:

$$\Delta\hat{W} = \sqrt{\frac{2W_R\beta_R^2}{-\eta_R} H}. \tag{3.63}$$

The bunch length $2\Delta t = 2\Delta\hat{t}$ is obtained if we set $\Delta W = 0$ for the same value of H:

$$H = -\frac{Q\hat{V}}{T_R} \frac{1}{\omega_{RF}} \left[\pm \cos(\omega_{RF} \Delta\hat{t}) \mp 1 \right].$$

Due to

$$\cos x - 1 = -2 \sin^2 \frac{x}{2},$$

we obtain

$$H = \pm \frac{Q\hat{V}}{2\pi h} 2 \sin^2 \frac{\Delta\hat{\varphi}_{RF}}{2}$$

with

$$\Delta\hat{\varphi}_{RF} = \omega_{RF} \Delta\hat{t}.$$

Inserting this into Eq. (3.63) leads to

$$\Delta\hat{W} = \sqrt{\pm \frac{2W_R\beta_R^2 Q\hat{V}}{-\pi h \eta_R}} \sin \frac{\Delta\hat{\varphi}_{RF}}{2} = \sqrt{\frac{W_R\beta_R^2 |Q| \hat{V}}{2\pi h |\eta_R|}} 2 \sin \frac{\Delta\hat{\varphi}_{RF}}{2}.$$

Due to Eq. (3.27), we have

$$|Q\eta_R|\hat{V} h = 2\pi W_R \beta_R^2 \frac{f_{S,0,stat}^2}{f_R^2},$$

which leads to

$$\Delta\hat{W} = \frac{W_\mathrm{R}\beta_\mathrm{R}^2}{h\,|\eta_\mathrm{R}|}\frac{f_\mathrm{S,0,stat}}{f_\mathrm{R}}\,2\,\sin\frac{\Delta\hat{\varphi}_\mathrm{RF}}{2} = \frac{W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|}\frac{\omega_\mathrm{S,0,stat}}{\omega_\mathrm{RF}}\,2\,\sin\frac{\Delta\hat{\varphi}_\mathrm{RF}}{2}$$

$$\Rightarrow \boxed{\frac{\Delta\hat{W}}{\Delta\hat{t}} = \frac{W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|}\,\omega_\mathrm{S,0,stat}\,\mathrm{si}\frac{\Delta\hat{\varphi}_\mathrm{RF}}{2}.} \tag{3.64}$$

For small values of $\Delta\hat{\varphi}_\mathrm{RF}$, i.e., for small bunches, this is equivalent to Eq. (3.28), as expected. If the bunch fills the whole bucket, we have $\Delta\hat{\varphi}_\mathrm{RF} = \pi$, which leads to

$$\mathrm{si}\frac{\Delta\hat{\varphi}_\mathrm{RF}}{2} = \mathrm{si}\frac{\pi}{2} = \frac{2}{\pi},$$

so that

$$\frac{\Delta\hat{W}}{\Delta\hat{t}} = \frac{2W_\mathrm{R}\beta_\mathrm{R}^2}{\pi|\eta_\mathrm{R}|}\,\omega_\mathrm{S,0,stat} = \frac{4W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|}\,f_\mathrm{S,0,stat}$$

is obtained. As expected, this is equivalent to Eq. (3.44).

One may sometimes approximate the bunch area by the formula for the area of an ellipse:

$$A_\mathrm{b,stat} \approx \pi\Delta\hat{t}\,\Delta\hat{W} = \pi\Delta\hat{t}^2\frac{W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|}\,\omega_\mathrm{S,0,stat}\,\mathrm{si}\frac{\Delta\hat{\varphi}_\mathrm{RF}}{2}.$$

If we make use of Eq. (3.48),

$$\frac{\omega_\mathrm{S,0,stat}}{A_\mathrm{B,stat}} = \frac{\omega_\mathrm{RF}^2\,|\eta_\mathrm{R}|}{16\,W_\mathrm{R}\beta_\mathrm{R}^2},$$

we get

$$\eta_\mathrm{fill} = \frac{A_\mathrm{b,stat}}{A_\mathrm{B,stat}} \approx \frac{\pi}{16}\Delta\hat{\varphi}_\mathrm{RF}^2\,\mathrm{si}\frac{\Delta\hat{\varphi}_\mathrm{RF}}{2} \tag{3.65}$$

for the bucket filling factor. In comparison with the exact solution (3.61), the maximum error is 23.4% for $\Delta\hat{\varphi}_\mathrm{RF} = \pi$ (180°). For $\Delta\hat{\varphi}_\mathrm{RF} < 110°$, the error is smaller than 5%, while for $\Delta\hat{\varphi}_\mathrm{RF} < 140°$, the error is smaller than 10%. This shows that it is often justified to simplify the shape of the trajectory by replacing it with an ellipse.

3.19 Frequency and RF Amplitude

Inside the deflection magnets, the magnetic field acts as centripetal force:

$$m\frac{u_\mathrm{R}^2}{r_\mathrm{R}} = Q\ u_\mathrm{R}\ B$$

$$\Rightarrow\ p_\mathrm{R} = Q\ r_\mathrm{R}\ B. \tag{3.66}$$

By means of this formula, we may calculate the revolution frequency of the particles. One obtains

$$u_\mathrm{R} = \beta_\mathrm{R} c_0 = l_\mathrm{R} f_\mathrm{R}$$

$$\Rightarrow\ f_\mathrm{R} = \frac{c_0}{l_\mathrm{R}}\ \beta_\mathrm{R}.$$

Since

$$p_\mathrm{R} = m_0 \gamma_\mathrm{R} \beta_\mathrm{R} c_0 \qquad \Rightarrow \gamma_\mathrm{R}\beta_\mathrm{R} = \frac{Q r_\mathrm{R} B}{m_0 c_0}$$

is valid, we obtain, with $\beta_\mathrm{R} = \frac{\gamma_\mathrm{R}\beta_\mathrm{R}}{\sqrt{1+\gamma_\mathrm{R}^2\beta_\mathrm{R}^2}}$,

$$\boxed{f_\mathrm{R} = \frac{c_0}{l_\mathrm{R}}\ \frac{\frac{Q r_\mathrm{R} B}{m_0 c_0}}{\sqrt{1+\left(\frac{Q r_\mathrm{R} B}{m_0 c_0}\right)^2}}.} \tag{3.67}$$

For a given accelerator, l_R and r_R are constants. Therefore, we see that the revolution frequency is given by the ratio

$$\frac{Q r_\mathrm{R} B}{m_0 c_0} = \beta_\mathrm{R}\gamma_\mathrm{R} \tag{3.68}$$

in a unique way. Due to

$$\gamma_\mathrm{R}^2 = \frac{1}{1-\beta_\mathrm{R}^2} \qquad \Leftrightarrow \gamma_\mathrm{R}^2 - \beta_\mathrm{R}^2\gamma_\mathrm{R}^2 = 1 \qquad \Leftrightarrow \beta_\mathrm{R}\gamma_\mathrm{R} = \sqrt{\gamma_\mathrm{R}^2 - 1},$$

the frequency is alternatively determined by a given γ_R. This shows that a normalization of the kinetic energy with respect to the mass according to

$$\frac{W_\mathrm{kin}}{m_0} = \frac{m_0 c_0^2(\gamma_\mathrm{R}-1)}{m_0} = c_0^2(\gamma_\mathrm{R}-1)$$

leads to values that can be uniquely assigned to a certain revolution frequency. This is why energies are often specified using the unit MeV/u. Independent of the ion species, one obtains the same revolution frequency and the same γ_R for the same value in MeV/u.

According to Eq. (3.68), however, this does not determine the magnetic field in a unique way, since the mass-to-charge ratio is still relevant.

Now we present a simplified derivation of Eq. (3.22). Based on Eq. (3.66), we determine the force that accelerates the particles in the longitudinal direction:

$$\Rightarrow F_R = \frac{\mathrm{d}p_R}{\mathrm{d}t} = Q\ r_R\ \dot{B}.$$

If $\dot{B}$ is constant, which is approximately true during one revolution, this force F_R has to be applied continuously. We therefore imagine that the force F_R is continuously distributed[8] along the ring accelerator. Then the kinetic energy that the particle gains during one revolution is

$$W = \int F_R\ \mathrm{d}s = Q\ r_R\ \dot{B}\ l_R.$$

This energy is delivered by the accelerating voltage in the accelerating gap:

$$W = \int QE\ \mathrm{d}s = Q\ V_R$$
$$\Rightarrow V_R = r_R\ \dot{B}\ l_R. \tag{3.69}$$

If $\dot{B}$ is given also, the required accelerating voltage is known. The acceleration, however, does not take place at the maximum of the accelerating voltage:

$$V_R = \hat{V}\ \sin\varphi_R$$
$$\Rightarrow \hat{V} = \frac{r_R\ \dot{B}\ l_R}{\sin\varphi_R}.$$

The frequency of the accelerating voltage is larger than the revolution frequency by a factor of h:

$$V(t) = \hat{V}\ \sin(\omega_{RF}t + \varphi_R) = \hat{V}\ \sin(2\pi h f_R t + \varphi_R).$$

---

[8] This is, of course, not true in reality, because the force is present in the accelerating gaps. Therefore, the derivation presented here should be regarded with skepticism. It is given here only because of its simplicity.

3.20 Voltage Versus Bunch Length

As an application of Eq. (3.28) for the ratio of the principal axes in phase space, we now determine the voltage that is necessary to produce a bunch of a given length.

For this purpose, we assume that the bunch area

$$A_\mathrm{b} = \pi \Delta\hat{t}\, \Delta\hat{W}$$

in phase space, i.e., the longitudinal emittance, is given. This formula includes the assumption that the target bunch in phase space is an ellipse. Therefore, the bunch must be sufficiently short in comparison with the bucket. Furthermore, only the stationary case is considered. Together with Eq. (3.28),

$$\frac{\Delta\hat{W}}{\Delta\hat{t}} = f_{\mathrm{S},0,\mathrm{stat}}\, \frac{2\pi W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|},$$

one obtains

$$A_{\mathrm{b,stat}} = \pi \Delta\hat{t}^2\; f_{\mathrm{S},0,\mathrm{stat}}\; \frac{2\pi W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|}.$$

If we now insert the synchrotron frequency from Eq. (3.25), it follows that

$$A_{\mathrm{b,stat}}^2 = \pi^2 \Delta\hat{t}^4\; f_\mathrm{R}^2\; \frac{|\eta_\mathrm{R}|\,|Q|\,\hat{V}\,h}{2\pi W_\mathrm{R}\beta_\mathrm{R}^2}\; \frac{(2\pi W_\mathrm{R}\beta_\mathrm{R}^2)^2}{\eta_\mathrm{R}^2} = \pi^2 \Delta\hat{t}^4\; f_\mathrm{R}^2\; \frac{|Q|\,\hat{V}\,h\,2\pi W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|}.$$

We expand the fraction with $2\pi h$ in order to get ω_RF:

$$A_{\mathrm{b,stat}}^2 = \pi^2 \Delta\hat{t}^4\; \omega_\mathrm{RF}^2\; \frac{|Q|\,\hat{V}\; W_\mathrm{R}\beta_\mathrm{R}^2}{2\pi h|\eta_\mathrm{R}|}.$$

Finally, we take

$$\Delta\hat{\varphi}_\mathrm{RF} = \omega_\mathrm{RF}\; \Delta\hat{t}$$

into account and obtain

$$A_{\mathrm{b,stat}}^2 = \pi^2 \Delta\hat{\varphi}_\mathrm{RF}^4\; \frac{|Q|\,\hat{V}\; W_\mathrm{R}\beta_\mathrm{R}^2}{2\pi h|\eta_\mathrm{R}|\omega_\mathrm{RF}^2}$$

$$\Rightarrow \hat{V} = \frac{2h|\eta_\mathrm{R}|\;\omega_\mathrm{RF}^2\; A_{\mathrm{b,stat}}^2}{\pi\;|Q|\;W_\mathrm{R}\;\beta_\mathrm{R}^2\;\Delta\hat{\varphi}_\mathrm{RF}^4}. \tag{3.70}$$

Therefore, the required voltage decreases as the fourth power of the bunch length; for reducing the bunch length to one-half its value, the voltage has to be increased by a factor of 16. Equation (3.70) can also be found in Edwards/Syphers [4] as formula (2.76).

3.21 Coasting Beam

Often, the (relative) **momentum spread**

$$\frac{\Delta p}{p_\mathrm{R}}$$

of the coasting beam, i.e., of the nonbunched beam, is given. Due to

$$\frac{\mathrm{d}\gamma}{\gamma} = \beta^2 \frac{\mathrm{d}(\gamma\beta)}{\gamma\beta},$$

one obtains

$$\frac{\Delta W}{W_\mathrm{R}} = \beta_\mathrm{R}^2 \frac{\Delta p}{p_\mathrm{R}},$$

so that the energy deviation can be calculated with the help of

$$W_\mathrm{R} = mc_0^2 = \gamma_\mathrm{R}\, m_0\, c_0^2.$$

It is important to note that W_R is the total energy. If, for example, $^{238}\mathrm{U}^{28+}$ with a kinetic energy of 11.4 MeV/u is given, one obtains

$$\begin{aligned} W_\mathrm{R} &= A_\mathrm{r}\,(W_\mathrm{kin,u} + W_\mathrm{rest,u}) = A_\mathrm{r}\left(W_\mathrm{kin,u} + m_\mathrm{u}c_0^2\right) = \\ &= 238.05 \cdot (11.4\,\mathrm{MeV} + 931.49\,\mathrm{MeV}) = 224.455\,\mathrm{GeV} \end{aligned}$$

for the total energy. By means of

$$A_\mathrm{b,DC} = 2\,\Delta W\,\Delta t,$$

one may now calculate the longitudinal emittance of the coasting beam. The factor 2 is due to the fact that the energy is in the range from $W_\mathrm{R} - \Delta W$ to $W_\mathrm{R} + \Delta W$. The time deviation corresponds to the revolution time of the particles:

$$A_\mathrm{b,DC} = 2\,\Delta W\,T_\mathrm{R} = \frac{2\,\Delta W}{f_\mathrm{R}} = \frac{2\,W_\mathrm{R}\beta_\mathrm{R}^2}{f_\mathrm{R}} \frac{\Delta p}{p_\mathrm{R}}.$$

We know that the phase space area is preserved (Liouville's theorem). Therefore, it should be possible to convert the coasting beam with emittance $A_{\mathrm{b,DC}}$ into a bunched beam produced by a cavity operating with harmonic number h.

In this case, the phase space area will be split among the h bunches according to

$$A_{\mathrm{b}} = \frac{A_{\mathrm{b,DC}}}{h}.$$

In Sect. 5.3, we will see that this conversion from a coasting beam to a bunched beam may be accomplished by an **adiabatic capture process**.

3.22 Ramps

We have seen that according to Eq. (3.67), the revolution frequency f_{R} is determined in a unique way as soon as the magnetic field B is given. With respect to Eq. (1.4), it is clear that the RF frequency f_{RF} is then also determined if a certain harmonic number h is chosen. Finally, Eq. (3.22) shows that the voltage $\hat{V}$ is also influenced by the choice of the magnetic field.

In other words, all these quantities have to be in synchrony, whence the name "synchrotron." A synchrotron is usually operated by transmitting time functions such as $B(t)$, $f_{\mathrm{RF}}(t)$, $\hat{V}(t)$ to the synchrotron devices. These time functions are often called **ramps**.

We will now discuss how ramps may be generated for a simple accelerating cycle. Such a **machine cycle** has to satisfy certain requirements:

- The momentum of the particles, and hence also the energy, is directly related to the magnetic rigidity. Therefore, the magnetic field of the bending magnets is low when the particles are injected into the synchrotron, and it must be high after the particles are accelerated to the final energy until they are extracted. Therefore, the magnetic field B has to be increased.
- Some time before the acceleration starts at injection energy and some time after the acceleration is completed at extraction energy, the beam must be bunched. On the one hand, we have $\dot{B} = 0$ during these phases, and on the other hand, we need some voltage amplitude $\hat{V} > 0$. According to Eq. (3.49), we have $V_{\mathrm{R}} = 0$, and due to Eq. (3.50), this leads to $\varphi_{\mathrm{R}} = 0$ or $\varphi_{\mathrm{R}} = \pm\pi$, depending on whether particles with positive or negative charge are considered and on whether operation takes place below or above the transition energy.
- During the accelerating phase, i.e., when $\dot{B} > 0$ holds, we need a larger voltage according to Eq. (3.49). Furthermore, we must have $\varphi_{\mathrm{R}} \neq 0, \pm\pi$, because the reference particle has to experience an accelerating voltage $V_{\mathrm{R}} \neq 0$ (Eq. (3.50)).
- In other words, we have a stationary bucket at injection energy, an acceleration bucket during acceleration, and a stationary bucket at extraction energy.

- During the whole process, we would, of course, like to have **matched bunches**,[9] because in that case, the phase space area that is occupied by the bunches will remain constant. A larger phase space area caused by the filamentation of **unmatched bunches** is undesired, because more voltage would then be needed. The strategy for having matched bunches permanently is to control the bucket area by means of the voltage $\hat{V}$.

One typically begins with the magnetic field $B(t)$ of the dipole magnets. As a first approximation, one would start with a piecewise linear curve (especially for this function, the name "ramp" is justified because the magnetic dipole field is ramped up):

$$B(t) = \begin{cases} B_{\mathrm{inj}} & 0 < t < t_1, \\ \frac{B_{\mathrm{extr}}(t-t_1)+B_{\mathrm{inj}}(t_2-t)}{t_2-t_1} & t_1 < t < t_2, \\ B_{\mathrm{extr}} & t_2 < t < t_3, \\ \frac{B_{\mathrm{extr}}(t_4-t)+B_{\mathrm{inj}}(t-t_3)}{t_4-t_3} & t_3 < t < t_4. \end{cases} \tag{3.71}$$

Such a choice, however, leads to a ramp $\dot{B}(t)$ that is not continuous. One result is that the time function for the reference phase $\varphi_{\mathrm{R}}(t)$ would contain jumps. The bunches would be unable to follow those jumps (unmatched bunch), and undesired longitudinal beam oscillations and filamentation would be the consequence.

Therefore, one may insert transitions at the times t_1, t_2, and t_3 in the ramp $B(t)$ specified above. These transitions may be defined in such a way that the ramp rate $\dot{B}$ increases or decreases to the desired value of the next segment. If this is done in a linear way, the ramp $\dot{B}(t)$ will be piecewise linear. The ramp $B(t)$ will then look similar to the one specified by Eq. (3.71), but its edges will be rounded off.

In conclusion, we have to keep in mind that transitions are needed in Eq. (3.71). Nevertheless, we will now discuss the meaning of the different phases.

- During the phase $0 < t < t_1$, the beam is injected into the synchrotron while the magnetic field is constant (phase A in Fig. 3.13). The magnetic field, of course, has to fit to the energy of the particles (injection energy/plateau). The RF voltage may already be switched on (so-called injection into stationary buckets), or it is slowly[10] ramped up[11] (**adiabatic capture**; see Sect. 5.3.2). The latter case is shown as phase B in Fig. 3.13. In phase C, the beam is kept bunched at constant energy for a while.

[9] We introduced matched bunches earlier as bunches that do not change their distribution. From our discussion of the trajectories inside the bucket, we now know that one of the requirements to obtain a matched bunch is that the boundary of the bunch correspond to an orbit with a constant value of the Hamiltonian.

[10] Here it is sufficient to regard the word "adiabatic" as a synonym for "slow." The mathematical concept of adiabaticity will be introduced in Sect. 5.3.

[11] The RF frequency has to be adapted to h times the revolution frequency, which may be determined by means of a Schottky measurement, as described in Chap. 1.

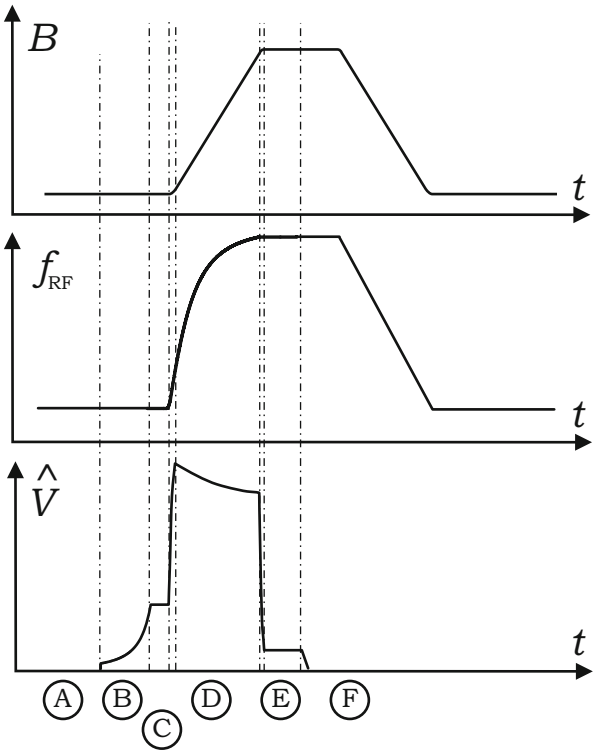

Fig. 3.13 An example of ramps in a machine cycle

- The phase $t_1 < t < t_2$ is the accelerating phase (phase D in Fig. 3.13).
- The phase $t_2 < t < t_3$ is called the "flat top" (phase E in Fig. 3.13). Here, at maximum energy, the beam is extracted from the synchrotron (extraction energy/plateau). Both **fast extraction** of bunches and **slow "spill" extraction** in order to get a steady beam current are possible (cf. [11, Sect. 7.4.1]).
- In the last phase, $t_3 < t < t_4$, the magnetic field is ramped down again in order to have the same initial conditions for the next machine cycle (phase F in Fig. 3.13). The actual shape of the ramps in this phase is unimportant, because no beam is present.
- Afterward, the next machine cycle begins. Therefore, t_4 is the cycle time (if we neglect the above-mentioned transitions), which should be kept as low as possible if a large intensity, i.e., a large number of bunches per time, is to be offered.

The magnetic field in the bending magnets is restricted for technical reasons. This is valid for both the minimum of the magnetic field B_{inj} at injection and the maximum of the magnetic field B_{extr} at extraction. In general, the dipole fields have to be controlled to very high accuracy (e.g., 10^{-5} or 10^{-6}).

We now derive all relevant ramps based on the definition of the magnetic field ramp $B(t)$. First of all, we may directly calculate the revolution frequency with the help of Eq. (3.67):

$$f_{\mathrm{R}} = \frac{c_0}{l_{\mathrm{R}}} \frac{\frac{Qr_{\mathrm{R}}B}{m_0c_0}}{\sqrt{1+\left(\frac{Qr_{\mathrm{R}}B}{m_0c_0}\right)^2}}.$$

The RF frequency is then given by

$$f_{\mathrm{RF}} = h\ f_{\mathrm{R}}.$$

It is easy to determine other quantities such as the velocity, the relativistic Lorentz factors β_{R} and γ_{R}, and kinetic and total energy.

Let us now assume that an adiabatic capture is performed, in which case, one would calculate the longitudinal emittance $A_{\mathrm{b,DC}}$ of the coasting beam based on the known momentum spread $\Delta p/p$ of the coasting beam (see Sect. 3.21). At the end of the capture process, the sum $h \cdot A_{\mathrm{B}}$ of the phase space areas of the h buckets must, of course, be larger than this emittance in order to allow all particles to move into the buckets.[12] In reality, however, some losses must be accepted. One may increase the voltage even further if bunching is to be enhanced, i.e., if shorter bunches are to be generated that fill a smaller fraction of the final bucket area A_{B}.

In the simplest case, one may increase the voltage slowly (more slowly than the period of the synchrotron oscillation, i.e., adiabatically) in a linear fashion. However, one may also use the isoadiabatic ramps derived in Sect. 5.3.2 in order to save time.

For the sake of simplicity, we assume a linear increase from 0 to $\hat{V}_{\mathrm{inj}}$ in the time interval $t_A < t < t_B$, where $0 < t_A < t_B < t_1$ holds. The time t_A is needed to inject the beam into the synchrotron and to give it enough time to form a real coasting beam with respect to the original momentum spread. For $t_B < t < t_1$, the voltage is kept constant at $\hat{V} = \hat{V}_{\mathrm{inj}}$.

Now we have to calculate how the voltage $\hat{V}$ varies in the time interval $t_1 < t < t_2$. For this purpose, we assume that a constant area in phase space is available for the beam:[13]

$$A_{\mathrm{B}} = \mathrm{const} \qquad \text{and} \qquad A_{\mathrm{B}} = A_{\mathrm{B,stat}}.$$

We use Eq. (3.41),

[12] Particles that are outside the bucket move on unstable trajectories. When the acceleration phase starts, the energy of these particles will deviate more and more from the reference value. This energy deviation corresponds to large transverse deviations from the reference orbit, which will lead finally to a loss of the particle on the beam pipe wall. Beam loss is, of course, undesirable, since it leads to less beam current, radioactive activation, and bad vacuum conditions.

[13] In heavy-ion synchrotrons, the bucket filling factor is often rather high, so that the requirement $A_{\mathrm{B}} = \mathrm{const}$ in combination with adiabaticity roughly implies $A_{\mathrm{b}} = \mathrm{const}$.

$$A_\mathrm{B} = A_{\mathrm{B,stat}}\,\alpha(\varphi_\mathrm{R}) = \frac{4\sqrt{2}}{\pi h} T_\mathrm{R} \sqrt{\frac{W_\mathrm{R}\beta_\mathrm{R}^2 |Q| \hat{V}}{\pi h |\eta_\mathrm{R}|}}\,\alpha(\varphi_\mathrm{R}), \tag{3.72}$$

which allows us to calculate A_B for the stationary case with $\varphi_\mathrm{R} = 0$ (or $\varphi_\mathrm{R} = \pm\pi$) and $\alpha(\varphi_\mathrm{R}) = 1$ and $\hat{V} = \hat{V}_\mathrm{inj}$. For the time interval $t_1 < t < t_2$, this value A_B remains constant, according to the requirements, whereas φ_R and $\hat{V}$ may change. The devolution of the other quantities is already determined, as mentioned above.

We now return to Eqs. (3.22) and (3.23), which lead to

$$\hat{V}\ \sin\varphi_\mathrm{R} = l_\mathrm{R} r_\mathrm{R} \dot{B} \tag{3.73}$$

for harmonic gap voltages. With the exception of φ_R and $\hat{V}$, all other quantities are already determined. Therefore, the last two equations allow us to determine both φ_R and $\hat{V}$ numerically.

At the flat top, the voltage $\hat{V}$ will be reduced, due to $\varphi_\mathrm{R} = 0$. One may even decrease it to zero in order to create a coasting beam again—this time, however, with a significantly higher (extraction) energy.

We now show how a numerical calculation of φ_R may be performed. For this purpose, we insert Eq. (3.73) into Eq. (3.72) in order to eliminate $\hat{V}$:

$$A_\mathrm{B}\sqrt{\sin\varphi_\mathrm{R}} = \frac{4\sqrt{2}}{\pi h} T_\mathrm{R} \sqrt{\frac{W_\mathrm{R}\beta_\mathrm{R}^2 |Q|\ l_\mathrm{R} r_\mathrm{R} \dot{B}}{\pi h |\eta_\mathrm{R}|}}\,\alpha(\varphi_\mathrm{R}).$$

$$\Leftrightarrow \sqrt{\sin\varphi_\mathrm{R}} - q\ \alpha(\varphi_\mathrm{R}) = 0 \tag{3.74}$$

Here we defined

$$q = \frac{4\sqrt{2}\ T_\mathrm{R}}{\pi h\ A_\mathrm{B}} \sqrt{\frac{W_\mathrm{R}\beta_\mathrm{R}^2 |Q|\ l_\mathrm{R} r_\mathrm{R} \dot{B}}{\pi h |\eta_\mathrm{R}|}}.$$

Now we see that the left-hand side of Eq. (3.74) equals $-q$ for $\varphi_\mathrm{R} = 0$, i.e., it is negative. For $\varphi_\mathrm{R} = \pi/2$, it equals 1, which is positive. Therefore, the root may easily be determined by means of a bisection method (i.e., by repeatedly bisecting the interval $\varphi_\mathrm{R} \in [0, \pi/2]$).

3.23 Multicavity Operation

Let us assume that several cavities are distributed along the synchrotron. A cavity is said to work at harmonic number h if its operating frequency equals h times the revolution frequency of the reference particle. The circumference of the synchrotron is denoted by l_R, and s will be a coordinate that describes the path length. For a specific harmonic number h, we have a number n_h of cavities that work at this harmonic number h and that are installed at the positions $s_{h,k}$ $(k = 1, \ldots, n_h)$. The number of bunches circulating in the synchrotron is denoted by h_b.

Each of the n_h cavities produces an RF signal

$$V_{h,k}(t) = \hat{V}_{h,k} \ \sin(h\omega_\mathrm{R} t - \varphi_{h,k}).$$

One usually wants to ensure that a particle bunch passes all these cavities in such a way that it reaches them at the same RF phase. The particle bunch will reach the kth cavity at the time

$$t_{h,k} = \frac{s_{h,k}}{u_\mathrm{R}},$$

where u_R is determined by

$$\omega_\mathrm{R} = \frac{2\pi}{T_\mathrm{R}} = \frac{2\pi u_\mathrm{R}}{l_\mathrm{R}}.$$

Therefore, we have

$$t_{h,k} = 2\pi \ \frac{s_{h,k}}{\omega_\mathrm{R} l_\mathrm{R}}.$$

The requirement that the bunch see the same phase at the different cavities means that

$$2\pi h \ \frac{s_{h,k}}{l_\mathrm{R}} - \varphi_{h,k} = 2\pi p,$$

where the left-hand side is obtained by inserting $t_{h,k}$ into the argument of the sine function. The right-hand side with an integer p results from the fact that all periods of the sine function are equivalent.

We now define

$$\theta_{h,k} = 2\pi \ \frac{s_{h,k}}{l_\mathrm{R}}$$

as the angle that describes the position of the cavity in the synchrotron. This leads to

$$\varphi_{h,k} = h\theta_{h,k} - 2\pi p.$$

As an example, we now consider the synchrotron SIS18 at GSI. It has two cavities that are exactly on opposite positions in the ring, so that we have $\theta_{h,1} = 0$ and $\theta_{h,2} = \pi$. For the standard operation at $h = 4$ or any other even harmonic number h, one gets

$$\varphi_{h,1} = 0 \qquad \text{and } \varphi_{h,2} = 0,$$

where p is chosen in such a way that $\varphi_{h,k}$ is located in the interval $]-\pi, \pi]$. For odd values of h, one immediately sees that one must choose

$$\varphi_{h,1} = 0 \qquad \text{and } \varphi_{h,2} = \pi.$$

In addition, we have to ensure that all bunches are treated equally. The distance between the h_b buckets is $l_\mathrm{R}/h_\mathrm{b}$. Assuming that two adjacent buckets are filled, their time difference equals

$$\Delta t = \frac{l_\mathrm{R}}{h_\mathrm{b}\, u_\mathrm{R}} = \frac{2\pi}{h_\mathrm{b}\, \omega_\mathrm{R}}.$$

If we insert this into the argument of the sine function, we see that

$$2\pi \frac{h}{h_\mathrm{b}}$$

must be an integer multiple of 2π if the same RF phase is to be obtained for all bunches. Hence, only integer multiples of h_b are allowed for h.

Furthermore, at least one group of cavities with $h = h_\mathrm{b}$ must exist in order to create the h_b bunches.

3.24 Bunch Shape

Several topics that we have presented thus far were based on the simple model that the bunch occupies a well-defined area in phase space. This allowed us, for example, to easily compare the longitudinal emittance (i.e., the bunch area in phase space) with the bucket area. Arguments of this type are implicitly based on the assumption that the particles have a distribution in phase space with clear margins. In reality, this

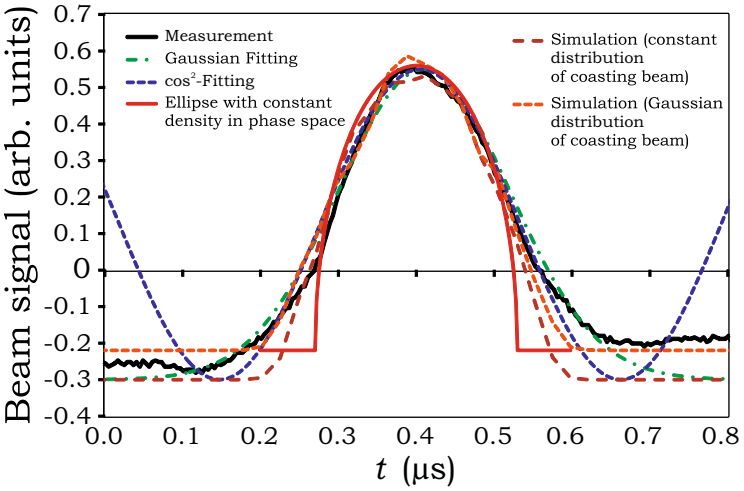

Fig. 3.14 Comparison of beam current measurement with different curves

is, of course, not true. For realistic bunches, one has to use rms values to describe the dimensions of the bunches in phase space. For example, one may define the longitudinal emittance[14] by

$$A_\mathrm{b} \sim \pi \Delta t_\mathrm{rms} \Delta W_\mathrm{rms}.$$

It is obvious that this formula is obtained if the area of an ellipse is calculated by means of the two semiaxes. In practice, different constants are used in the literature to make the definition complete (cf. [6, Sect. 5.4.4]).

Now we return to our beam current example, which was shown in Fig. 1.3 (Measurement No. 43 dated August, 21, 2008, $^{40}\mathrm{Ar}^{18+}$, 11.4 MeV/u, 6 kV, $h = 4$, $N_\mathrm{beam} = 1.5 \cdot 10^9$ particles). According to

$$\boxed{\bar{I}_\mathrm{beam} = \frac{N_\mathrm{beam} z_q e}{T_\mathrm{R}} = N_\mathrm{beam} z_q e f_\mathrm{R},} \tag{3.75}$$

this corresponds to an average beam current of about 0.9 mA. Since the quality of the beam was very good during this experiment, the measured pulse shape may be used for an analysis of the distribution.

Figure 3.14 shows different curves in comparison with the measurement.

If one discards the problem that the zero line of the beam current is hard to identify, the Gaussian distribution (second curve) is an excellent approximation

[14] See also Sect. 5.6.4.

of the measured pulse shape (with $\sigma = 78\,\text{ns}$). Please note that the Gaussian distribution cannot, in spite of the good results, represent the "absolute truth," since a bunch that is captured inside a bucket must always have a finite size. The Gaussian distribution, however, represents the typical shape of the bunches of a low-intensity ion beam when space-charge effects are low (cf. [6, Sect. 5.4.7]).

Also, the $\cos^2$ distribution (third curve) provides good results—except that at the margins of the bunch, larger deviations are visible (the $\cos^2$ function must, of course, be clipped beyond the minima on both sides—this was not done in the figure in order to show the periodicity of this function).

The fourth curve shows the results of a simulation in which a coasting beam with a constant distribution of 30,000 macroparticles in phase space with $\Delta p/p|_{\max} = \pm 3.5 \cdot 10^{-4}$ was assumed. In the simulation, the complete capture process was simulated. The projection that is needed to calculate the bunch shape was based on 200 equidistant bins. One sees that the constant distribution does not describe the reality very well, but for qualitative analyses it is often sufficient.

The fifth curve shows the simulation results in which a coasting beam with a Gaussian distribution ($\Delta p/p|_{\sigma} = 2 \cdot 10^{-4}$) was assumed at the beginning. Again, 30,000 macroparticles with 200 bins were used, and the complete capture process was simulated. The good agreement with the measurement is obvious.

Finally, the theoretical curve for a constant particle density inside an ellipse in phase space is shown. In this case, the projection is an ellipse as well.

Of course, there exist other models for the distribution of the particles (e.g., the parabolic bunch, which is an important distribution for space-charge dominated beams [6, 12]).

References

1. K.Y. Ng, *Physics of Intensity Dependent Beam Instabilities* (World Scientific, New Jersey/London/Singapore/Beijing/Shanghai/Hong Kong/Taipei/Chennai, 2006)
2. M. Conte, W.W. MacKay, *An Introduction to the Physics of Particle Accelerators* (World Scientific, New Jersey/London/Singapore/Beijing/Shanghai/Hong Kong/Taipei, Chennai, 2008)
3. H. Wiedemann, *Particle Accelerator Physics I & II*, 2nd edn. (Springer, Berlin/Heidelberg/New York, 2003)
4. D.A. Edwards, M.J. Syphers, *An Introduction to the Physics of High Energy Accelerators* (Wiley-VCH Verlag GmbH&Co.KGaA, Weinheim, 2004)
5. S.Y. Lee, *Accelerator Physics* (World Scientific, Singapore/New Jersey/London/Hong Kong, 1999)
6. M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New York/Chichester/Brisbane/Toronto/Singapore, 1994)
7. A.W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (Wiley, New York/Chichester/Brisbane/Toronto/Singapore, 1993)
8. A.W. Chao, M. Tigner, *Handbook of Accelerator Physics and Engineering*, 3rd edn. (World Scientific, New Jersey/London/Singapore/Beijing/Shanghai/Hong Kong/Taipei/Chennai, 2006)
9. P.J. Bryant, K. Johnsen, *The Principles of Circular Accelerators and Storage Rings* (Cambridge University Press, Cambridge/New York/Melbourne/Madrid/Cape Town, 2005)

10. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965)
11. E.J.N. Wilson, *An Introduction to Particle Accelerators* (Oxford University Press, Oxford/New York, 2001)
12. B.W. Zotter, S.A. Kheifets, *Impedances and Wakes in High-Energy Particle Accelerators* (World Scientific, Singapore/New Jersey/London/Hong Kong, 1998)

Chapter 4
RF Cavities

This introduction and Sect. 4.1 are based on the article "Ferrite cavities" [2] published in CERN Report CERN-2011-007 under the CC BY 3.0 Attribution License (http://creativecommons.org/licenses/by/3.0/). The original content (http://cds.cern.ch/record/1411778/files/p299.pdf) has been modified slightly.

The revolution frequency of charged particles in synchrotrons or storage rings is usually lower than 10 MHz. Even if we consider comparatively small synchrotrons (e.g., like HIT/HICAT in Heidelberg, Germany, or CNAO in Pavia, Italy, with about 20–25 m diameter, both used for tumor therapy), the revolution time will be greater than 200 ns, since the particles cannot reach the speed of light. Since according to Eq. (1.4),

$$f_{\mathrm{RF}} = h \cdot f_{\mathrm{R}},$$

the RF frequency is an integer multiple of the revolution frequency, the RF frequency will typically be lower than 10 MHz if only small harmonic numbers h are desired. For such an operating frequency, the spatial dimensions of a conventional RF resonator would be far too large to be used in a synchrotron. One possibility for solving this problem is to reduce the wavelength by filling the cavity with magnetic material. This is the basic idea of **ferrite-loaded cavities** (cf. [1]). Furthermore, this type of cavity offers a simple means to modify the resonant frequency in a wide range (typically up to a factor of 10) and in a comparatively short time (typically at least 10 ms cycle time). Therefore, ferrite cavities are suitable for ramped operation in a synchrotron. The possibility of adjusting the resonant frequency of a cavity to the desired operating frequency is called **tuning**.

Due to the low operating frequencies, the transit-time factor (cf. Sect. 4.3) of traditional ferrite-loaded cavities is almost 1 and is therefore not of interest.

If a synchrotron is operated at comparatively high harmonic numbers, the RF frequency will reach values that can be realized as resonant frequencies of classical RF resonators (typically 300 MHz or higher). Furthermore, if the particles are

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6_4

already relativistic (i.e., $\beta_R \approx 1$) when they are injected into the synchrotron,[1] there is no need to make the cavity tunable to different resonant frequencies. In this case, the use of classical resonators is possible.

Many accelerating cavities of this classical type may be regarded as modifications and/or combinations of the so-called **pillbox cavity** [3, 4]. The pillbox cavity is the simplest type of resonator that can be used for particle acceleration. It will be discussed in Sect. 4.4.

4.1 Ferrite-Loaded Cavities

The main purpose of this section is to derive the general lumped element circuit for cavities and to discuss the main properties of RF cavities in synchrotrons and storage rings. Furthermore, some specifics that are typical for ferrite-loaded cavities are discussed.

We will see that a ferrite-loaded cavity may be regarded as roughly a transformer whose primary coil consists of only one winding fed by an RF power source in which the beam acts as the secondary coil. Consequently, some conclusions that are valid for transformers are also valid here. For example, the cavity will not work properly if the frequency is too low, because the reactance (product of inductance and angular frequency) will be too small in comparison with the ohmic parts, thereby decreasing the transformation ratio. If the frequency becomes too large, flux leakage and distributed effects will become important, so that a simple magnetoquasistatic analysis is no longer possible. Hence, an optimum operating frequency range can be specified for a ferrite-loaded cavity, similar to that of a transformer, taking the material properties and the geometry into account. For our analysis, which begins in Sect. 4.1.2, it is assumed without further notice that the considered frequency belongs to this optimum frequency range.

4.1.1 Permeability of Magnetic Materials

In this section, all calculations are based on permeability quantities μ for which

$$\mu = \mu_r \mu_0$$

holds. In material specifications, the relative permeability μ_r is given, which means that we have to multiply by μ_0 to obtain μ. This comment is also valid for the incremental/differential permeability introduced below.

[1] Please note that according to Sect. 2.7, relativistic particles will still gain energy even though the increase in speed is negligible.

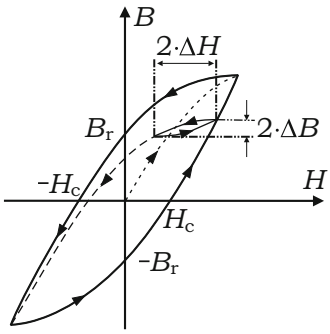

Fig. 4.1 Hysteresis loop

In RF cavities, only so-called **soft magnetic materials** that have a narrow hysteresis loop are of interest, since their losses are comparatively low (in contrast to **hard magnetic materials**, which are used for permanent magnets.[2])

Figure 4.1 shows the **hysteresis loop** of a ferromagnetic material. It is well known that the hysteresis loop leads to a **residual induction** B_r if no magnetizing field H is present and that some **coercive magnetizing field** H_c is needed to set the induction B to zero.

Let us now assume that some cycles of the large hysteresis loop have already passed and that H is currently increasing. We now stop to increase the magnetizing field H in the upper right-hand part of the diagram. Then H is decreased by a much smaller amount $2 \cdot \Delta H$, then increased again by that amount $2 \cdot \Delta H$, and so forth.[3] As the diagram shows, this procedure will lead to a much smaller hysteresis loop whereby B changes by $2 \cdot \Delta B$. We may therefore define a **differential** or **incremental permeability**[4]

$$\boxed{\mu_\Delta = \frac{\Delta B}{\Delta H},}$$

which describes the slope of the local hysteresis loop. It is this quantity μ_Δ that is relevant for RF applications. One can see that μ_Δ can be decreased by increasing the DC component of H. Since H is generated by currents, one speaks of a **bias**

[2]No strict separation exists between hard and soft magnetic materials.

[3]The factor 2 was assumed in order to have the same total change of $2 \cdot \Delta H$ as in the equation

$$H_\mathrm{AC}(t) = \Delta H \cos \omega t,$$

which is usually used for harmonic oscillations.

[4]In a strict sense, the differential permeability is the limit

$$\mu_\Delta = \frac{\mathrm{d}B}{\mathrm{d}H}$$

for $\Delta H, \Delta B \to 0$.

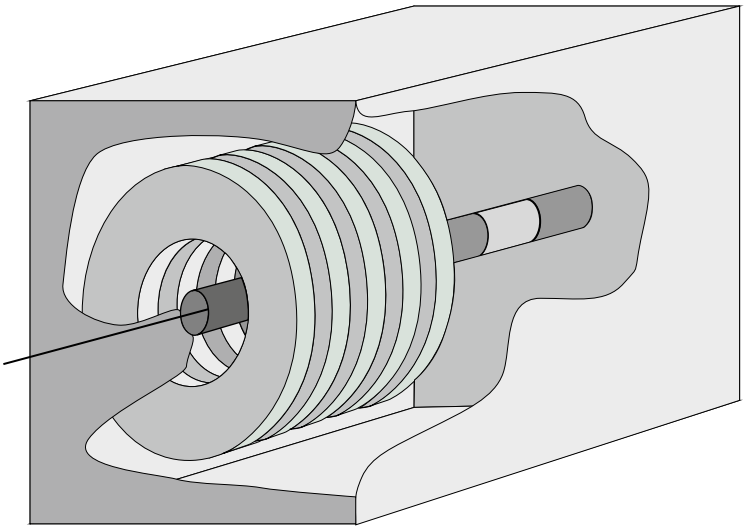

Fig. 4.2 Simplified 3D sketch of a ferrite-loaded cavity

current that is applied in order to shift the **operating point** to higher inductions B, leading to a lower differential permeability μ_Δ.

If no biasing is applied, the maximum μ_Δ is obtained, which is typically of order a few hundred or a few thousand times μ_0.

The hysteresis loop and the AC permeability of ferromagnetic materials can be described in a phenomenological way by the so-called **Preisach model**, which is explained in the literature (cf. [5]). Unfortunately, the material properties are even more complicated, since they are also frequency-dependent. One usually uses the **complex permeability**

$$\boxed{\underline{\mu} = \mu_s' - j\mu_s''} \tag{4.1}$$

in order to describe losses (hysteresis loss, eddy current loss, and residual loss). The parameters μ_s' and μ_s'' are frequency-dependent. In the following, we will assume that the complex permeability $\underline{\mu}$ describes the material behavior in rapidly alternating fields as the above-mentioned real quantity μ_Δ does when a biasing field H_{bias} is present. However, we will omit the index Δ for the sake of simplicity.

4.1.2 Magnetoquasistatic Analysis of a Ferrite Cavity

Figure 4.2 shows the main elements of a ferrite-loaded cavity. The beam pipe is interrupted by a **ceramic gap**. This gap ensures that the beam pipe may still be

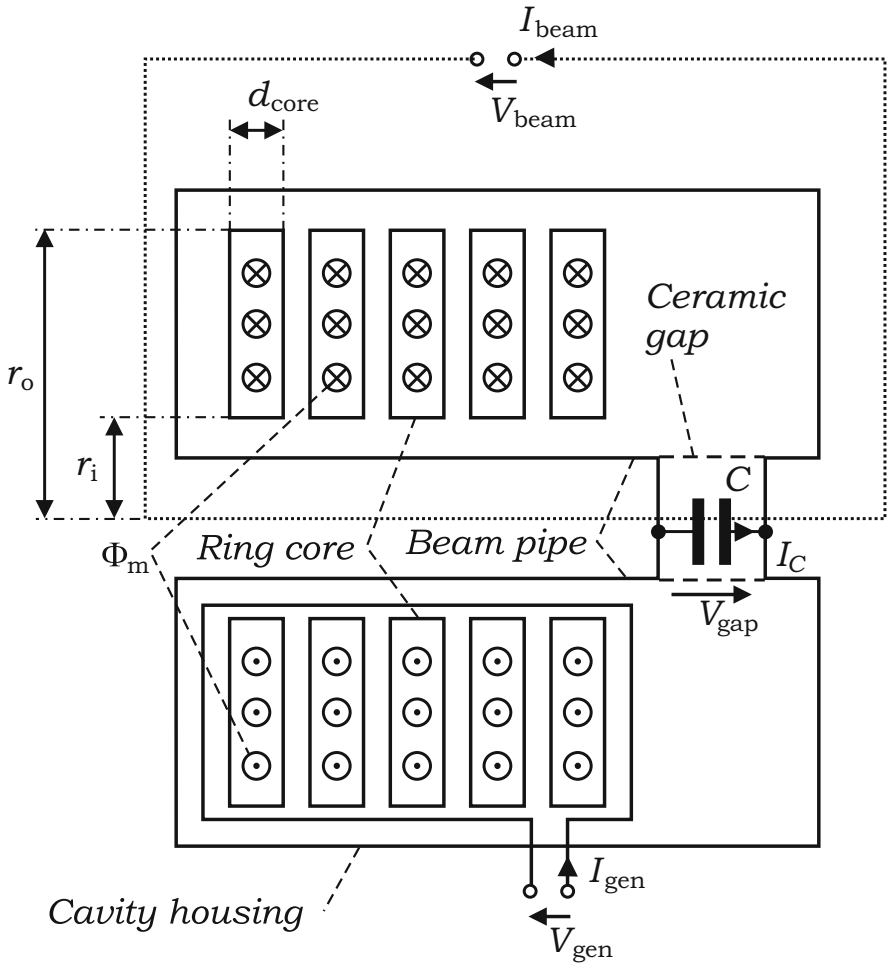

Fig. 4.3 Simplified model of a ferrite cavity

evacuated, but it allows a voltage V_{gap} to be induced in the longitudinal direction. Several magnetic **ring cores** are mounted in a concentric way around the beam and beam pipe (five ring cores are drawn here as an example). The whole cavity is surrounded by a metallic housing, which is connected to the beam pipe.

Figure 4.3 shows a cross section through the cavity. The dotted line represents the beam, which is located in the middle of a metallic beam pipe (for analyzing the influence of the beam current, this dotted line is regarded as a part of a circuit that closes outside the cavity, but this is not relevant for understanding the basic operational principle). The ceramic gap has a parasitic capacitance, but additional lumped-element capacitors are usually connected in parallel, leading to the overall capacitance C. Starting at the generator port located at the bottom of the figure, an

inductive coupling loop surrounds the ring core stack. This loop was not shown in Fig. 4.2.

Please note that due to the cross-section approach, we obtain a wire model of the cavity with two wires representing the cavity housing. This is sufficient for practical analysis, but one should keep in mind that in reality, the currents are distributed.

In the following, we will represent voltages, currents, field, and flux quantities as phasors, i.e., complex amplitudes/peak values for a given frequency $f = \omega/2\pi$. In this case, for a quantity X in the time domain, we write $\underline{\hat{X}}$ for the phasor in the frequency domain. The function $X(t)$ can be reconstructed by means of the complex function

$$\underline{X} = \underline{\hat{X}} e^{j\omega t}$$

according to

$$X(t) = \mathrm{Re}\{\underline{X}\} = \mathrm{Re}\{\underline{\hat{X}} e^{j\omega t}\}.$$

Let us consider a contour that surrounds the lower left ring core stack. Based on Maxwell's second equation in the time domain (Faraday's law),

$$\oint_{\partial A} \vec{E} \cdot \mathrm{d}\vec{r} = -\int_A \dot{\vec{B}} \cdot \mathrm{d}\vec{A},$$

we obtain

$$\underline{\hat{V}}_{\mathrm{gen}} = +j\omega \underline{\hat{\Phi}}_{\mathrm{m,tot}} \tag{4.2}$$

in the frequency domain. If we now use the complete lower cavity half as the integration path, we obtain

$$\underline{\hat{V}}_{\mathrm{gap}} = +j\omega \underline{\hat{\Phi}}_{\mathrm{m,tot}}.$$

Hence we obtain

$$\underline{\hat{V}}_{\mathrm{gap}} = \underline{\hat{V}}_{\mathrm{gen}}. \tag{4.3}$$

Here we assumed that the stray field B in the air region is negligible in comparison with the field inside the ring cores (due to their high permeability). Finally, we consider the beam current contour:

$$\underline{\hat{V}}_{\mathrm{beam}} = +j\omega \underline{\hat{\Phi}}_{\mathrm{m,tot}} = \underline{\hat{V}}_{\mathrm{gap}}.$$

For negligible displacement current, we have Maxwell's first equation (Ampère's law)

$$\oint_{\partial A} \vec{H} \cdot \mathrm{d}\vec{r} = \int_A \vec{J} \cdot \mathrm{d}\vec{A}.$$

We use a concentric circle with radius ρ around the beam as integration path:

$$H\ 2\pi\rho = I_{\mathrm{tot}}. \tag{4.4}$$

In the frequency domain, this leads to

$$\underline{\hat{B}} = \underline{\mu} \frac{\underline{\hat{I}}_{\mathrm{tot}}}{2\pi\rho} \tag{4.5}$$

with

$$\underline{\hat{I}}_{\mathrm{tot}} = \underline{\hat{I}}_{\mathrm{gen}} - \underline{\hat{I}}_C - \underline{\hat{I}}_{\mathrm{beam}}. \tag{4.6}$$

For the flux through a single ring core, we get

$$\underline{\hat{\Phi}}_{\mathrm{m},1} = \int \underline{\hat{\vec{B}}} \cdot \mathrm{d}\vec{A} = d_{\mathrm{core}} \int_{r_{\mathrm{i}}}^{r_{\mathrm{o}}} \underline{\hat{B}}\, \mathrm{d}\rho = \frac{d_{\mathrm{core}} \underline{\mu} \underline{\hat{I}}_{\mathrm{tot}}}{2\pi}\ \ln \frac{r_{\mathrm{o}}}{r_{\mathrm{i}}}.$$

With the complex permeability

$$\underline{\mu} = \mu_{\mathrm{s}}' - j\mu_{\mathrm{s}}''$$

and assuming that N ring cores are present, we obtain

$$\underline{\hat{V}}_{\mathrm{gap}} = j\omega \underline{\hat{\Phi}}_{\mathrm{m,tot}} = j\omega N\ \underline{\hat{\Phi}}_{\mathrm{m},1} = j\omega \frac{N d_{\mathrm{core}} (\mu_{\mathrm{s}}' - j\mu_{\mathrm{s}}'') \underline{\hat{I}}_{\mathrm{tot}}}{2\pi}\ \ln \frac{r_{\mathrm{o}}}{r_{\mathrm{i}}}.$$

Therefore, we obtain

$$\underline{\hat{V}}_{\mathrm{gap}} = \underline{\hat{I}}_{\mathrm{tot}} (j\omega L_{\mathrm{s}} + R_{\mathrm{s}}) = \underline{\hat{I}}_{\mathrm{tot}} Z_{\mathrm{s}} \tag{4.7}$$

if

$$Z_{\mathrm{s}} = \frac{1}{Y_{\mathrm{s}}} = j\omega L_{\mathrm{s}} + R_{\mathrm{s}}, \tag{4.8}$$

$$\boxed{L_\mathrm{s} = \frac{N d_\mathrm{core} \mu_\mathrm{s}'}{2\pi} \ln \frac{r_\mathrm{o}}{r_\mathrm{i}},}$$

$$\boxed{R_\mathrm{s} = \omega \frac{N d_\mathrm{core} \mu_\mathrm{s}''}{2\pi} \ln \frac{r_\mathrm{o}}{r_\mathrm{i}} = \omega \frac{\mu_\mathrm{s}''}{\mu_\mathrm{s}'} L_\mathrm{s} = \frac{\omega L_\mathrm{s}}{Q},} \tag{4.9}$$

are defined. Here,

$$Q = \frac{\mu_\mathrm{s}'}{\mu_\mathrm{s}''} = \frac{1}{\tan \delta_\mu} \tag{4.10}$$

is the **quality factor** (or **Q factor**) of the ring core material. Using Eq. (4.6), we obtain

$$\begin{aligned} \underline{\hat{V}}_\mathrm{gap} Y_\mathrm{s} &= \underline{\hat{I}}_\mathrm{tot} = \underline{\hat{I}}_\mathrm{gen} - \underline{\hat{I}}_\mathrm{beam} - \underline{\hat{V}}_\mathrm{gap}\ j\omega C \\ \Rightarrow \boxed{\underline{\hat{V}}_\mathrm{gap}} &= \frac{\underline{\hat{I}}_\mathrm{gen} - \underline{\hat{I}}_\mathrm{beam}}{Y_\mathrm{s} + j\omega C} = Z_\mathrm{tot}(\underline{\hat{I}}_\mathrm{gen} - \underline{\hat{I}}_\mathrm{beam}). \end{aligned} \tag{4.11}$$

This equation corresponds to the equivalent circuit shown in Fig. 4.4. In the last step, we defined

$$Y_\mathrm{tot} = \frac{1}{Z_\mathrm{tot}} = Y_\mathrm{s} + j\omega C.$$

In the literature, one often finds a different version of Eq. (4.11), in which $\underline{\hat{I}}_\mathrm{beam}$ has the same sign as $\underline{\hat{I}}_\mathrm{gen}$. This corresponds to both currents having the same direction (flowing into the circuits in Figs. 4.4 and 4.5). In any case, one has to make sure that the correct phase between beam current and gap voltage is established.

In Chap. 3, we studied the stationary case, whereby the gap voltage is given by

$$V_\mathrm{gap}(t) = \hat{V}_\mathrm{gap} \sin(\omega_\mathrm{RF} t)$$

and the bunches are located at $t = 0, \pm T_\mathrm{RF}, \pm 2T_\mathrm{RF}, \ldots$ (operation with positively charged particles below transition energy). Therefore, the fundamental harmonic of the beam current will be proportional to $\cos(\omega_\mathrm{RF} t)$, which corresponds to a 90° phase shift between V_gap and I_beam. For low beam currents and for a cavity that is tuned to resonance, the phase of the gap voltage is equal to the phase of the generator current. For higher beam currents, however, not only the generator current, but also the beam current will have an influence on the gap voltage due to the **beam impedance** Z_tot, as can be seen in Eq. (4.11) and in Fig. 4.4. This phenomenon is called **beam loading**.

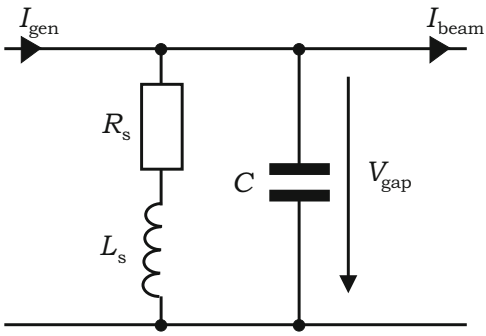

Fig. 4.4 Series equivalent circuit

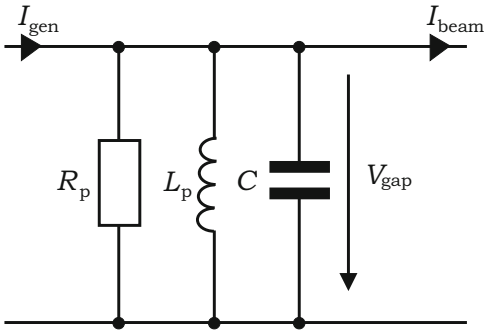

Fig. 4.5 Parallel equivalent circuit

4.1.3 Parallel and Series Lumped Element Circuit

In the vicinity of the resonant frequency, it is possible to convert the lumped element circuit shown in Fig. 4.4 into a parallel circuit as shown in Fig. 4.5. The admittances of both circuits will be assumed equal:

$$Y_{\text{tot}} = j\omega C + \frac{1}{R_s + j\omega L_s} = j\omega C + \frac{1}{R_p} + \frac{1}{j\omega L_p}$$

$$\Rightarrow \frac{R_s - j\omega L_s}{R_s^2 + (\omega L_s)^2} = \frac{1}{R_p} + \frac{1}{j\omega L_p}.$$

A comparison of the real and imaginary part yields

$$R_p = \frac{R_s^2 + (\omega L_s)^2}{R_s}, \tag{4.12}$$

$$\omega L_p = \frac{R_s^2 + (\omega L_s)^2}{\omega L_s}. \tag{4.13}$$

For the inverse relation, we modify the first equation according to

$$(\omega L_\mathrm{s})^2 = R_\mathrm{s}(R_\mathrm{p} - R_\mathrm{s})$$

and use this result in the second equation:

$$\begin{aligned}
&\omega L_\mathrm{p}\sqrt{R_\mathrm{s}(R_\mathrm{p} - R_\mathrm{s})} = R_\mathrm{s}R_\mathrm{p} \\
\Rightarrow\ &(\omega L_\mathrm{p})^2(R_\mathrm{p} - R_\mathrm{s}) = R_\mathrm{s}R_\mathrm{p}^2 \\
\Rightarrow\ &\boxed{R_\mathrm{s} = \frac{(\omega L_\mathrm{p})^2}{R_\mathrm{p}^2 + (\omega L_\mathrm{p})^2}R_\mathrm{p}.}
\end{aligned}$$

Equations (4.12) and (4.13) directly provide

$$R_\mathrm{p}R_\mathrm{s} = (\omega L_\mathrm{p})(\omega L_\mathrm{s}), \tag{4.14}$$

which leads to

$$\boxed{\omega L_\mathrm{s} = \frac{R_\mathrm{p}}{\omega L_\mathrm{p}}R_\mathrm{s} = \frac{R_\mathrm{p}^2}{R_\mathrm{p}^2 + (\omega L_\mathrm{p})^2}\omega L_\mathrm{p}.}$$

Since it is suitable to use both types of lumped element circuit, it is also convenient to define the complex permeability $\underline{\mu}$ in a parallel form:

$$\boxed{\frac{1}{\underline{\mu}} = \frac{1}{\mu'_\mathrm{p}} + j\frac{1}{\mu''_\mathrm{p}}.} \tag{4.15}$$

This is an alternative representation of the series form shown in Eq. (4.1), which leads to

$$\frac{1}{\underline{\mu}} = \frac{\mu'_\mathrm{s} + j\mu''_\mathrm{s}}{\mu'^2_\mathrm{s} + \mu''^2_\mathrm{s}}.$$

Comparing the real and imaginary parts of the last two equations, we obtain

$$\mu'_\mathrm{p} = \frac{\mu'^2_\mathrm{s} + \mu''^2_\mathrm{s}}{\mu'_\mathrm{s}}, \tag{4.16}$$

$$\mu''_\mathrm{p} = \frac{\mu'^2_\mathrm{s} + \mu''^2_\mathrm{s}}{\mu''_\mathrm{s}}. \tag{4.17}$$

These two equations lead to

$$\mu'_{\mathrm{p}}\mu'_{\mathrm{s}} = \mu''_{\mathrm{p}}\mu''_{\mathrm{s}}.$$

Putting this together with Eqs. (4.9), (4.10), and (4.14), we conclude that

$$\boxed{Q = \frac{\mu'_{\mathrm{s}}}{\mu''_{\mathrm{s}}} = \frac{\omega L_{\mathrm{s}}}{R_{\mathrm{s}}} = \frac{R_{\mathrm{p}}}{\omega L_{\mathrm{p}}} = \frac{\mu''_{\mathrm{p}}}{\mu'_{\mathrm{p}}}.} \tag{4.18}$$

With these expressions, we may write Eqs. (4.16) and (4.17) in the form

$$\boxed{\mu'_{\mathrm{p}} = \mu'_{\mathrm{s}}\left(1 + \frac{1}{Q^2}\right),} \tag{4.19}$$

$$\boxed{\mu''_{\mathrm{p}} = \mu''_{\mathrm{s}}\left(1 + Q^2\right).} \tag{4.20}$$

If we use Eq. (4.18),

$$Q = \frac{\omega L_{\mathrm{s}}}{R_{\mathrm{s}}},$$

we may rewrite Eqs. (4.12) and (4.13) in the form

$$\boxed{\begin{aligned} R_{\mathrm{p}} &= R_{\mathrm{s}}(1 + Q^2), \qquad (4.21)\\ L_{\mathrm{p}} &= L_{\mathrm{s}}\left(1 + \frac{1}{Q^2}\right). \qquad (4.22) \end{aligned}}$$

By combining Eqs. (4.21) and (4.9), we obtain

$$R_{\mathrm{p}} = (1 + Q^2)\omega\frac{Nd_{\mathrm{core}}\mu''_{\mathrm{s}}}{2\pi}\ \ln\frac{r_{\mathrm{o}}}{r_{\mathrm{i}}}.$$

With the help of Eqs. (4.18) and (4.19), we obtain

$$\mu''_{\mathrm{s}} = \frac{\mu'_{\mathrm{s}}}{Q} = \frac{\mu'_{\mathrm{p}}}{Q + \frac{1}{Q}} = \frac{\mu'_{\mathrm{p}}Q}{1 + Q^2}.$$

The last two equations lead to

$$R_{\mathrm{p}} = \omega\frac{Nd_{\mathrm{core}}\mu'_{\mathrm{p}}Q}{2\pi}\ \ln\frac{r_{\mathrm{o}}}{r_{\mathrm{i}}} = Nd_{\mathrm{core}}\mu'_{\mathrm{p}}Qf\ \ln\frac{r_{\mathrm{o}}}{r_{\mathrm{i}}}.$$

This shows that R_p is proportional to the product $\mu'_p Qf$, which is a material property. The other parameters refer to the geometry. Therefore, the manufacturers of ferrite cores sometimes specify the $\boldsymbol{\mu_r}$**Qf product** (for the sake of simplicity, we define $\mu_r := \mu'_{p,r}$).[5]

For $Q \geq 5$, we may use the approximations

$$R_p \approx R_s\, Q^2, \qquad L_p \approx L_s, \qquad \mu'_p \approx \mu'_s, \qquad \mu''_p \approx \mu''_s\, Q^2, \tag{4.23}$$

which then have an error of less than 4%.

4.1.4 Frequency Dependence of Material Properties

As an example, the frequency dependence of the permeability is shown in Figs. 4.6 and 4.7 for the special ferrite material Ferroxcube 4, assuming small magnetic RF fields without biasing. All the data presented for this material are taken from [6]. It is obvious that the behavior depends significantly on the choice of the material. Without biasing, a constant $\mu'_s \approx \mu'_p$ may be assumed only up to a certain frequency (see Fig. 4.6). With increasing frequency from 0, the Q factor will decrease (compare Figs. 4.6 and 4.7). Figure 4.8 shows the resulting frequency dependence of the $\mu_r Qf$ product.

If the magnetic RF field is increased, both Q and $\mu_r Qf$ will decrease in comparison with the diagrams in Figs. 4.6, 4.7, and 4.8. The effective incremental permeability μ_r will increase for rising magnetic RF fields, as one can see by interpreting Fig. 4.1. Therefore, it is important to consider the material properties under realistic operating conditions (the maximum RF B-field is usually of order 10 . . . 20 mT).

If biasing is applied, the $\mu_r Qf$ curve shown in Fig. 4.8 will be shifted to the lower right-hand side; this effect may approximately compensate the increase in $\mu_r Qf$ with frequency [6]. Therefore, the $\mu_r Qf$ product may sometimes be regarded as approximately a constant if biasing is used to keep the cavity at resonance for all frequencies under consideration.

[5] Here, the index r again denotes the relative permeability, i.e.,

$$\mu'_{p,r} = \frac{\mu'_p}{\mu_0}.$$

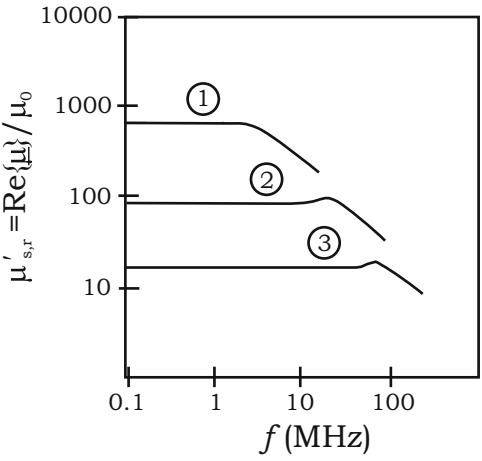

Fig. 4.6 $\mu'_{s,r}$ versus frequency for three different types of ferrite material (1: Ferroxcube 4A, 2: Ferroxcube 4C, 3: Ferroxcube 4E). Data taken from [6]

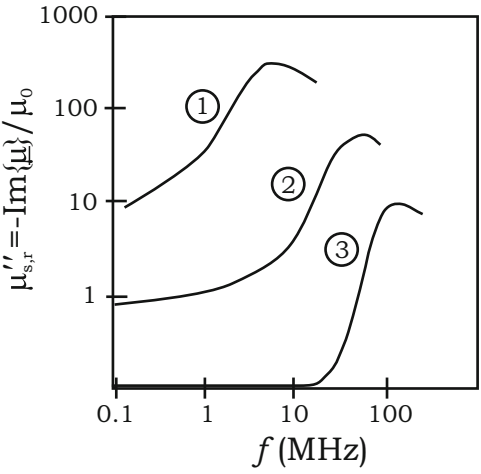

Fig. 4.7 $\mu''_{s,r}$ versus frequency for three different types of ferrite material (1: Ferroxcube 4A, 2: Ferroxcube 4C, 3: Ferroxcube 4E). Data taken from [6]

4.1.5 Quality Factor of the Cavity

The quality factor of the equivalent circuit shown in Fig. 4.5 is obtained if the **resonant (angular) frequency**

$$\boxed{\omega_{\mathrm{res}} = 2\pi f_{\mathrm{res}} = \frac{1}{\sqrt{L_{\mathrm{p}} C}}}$$

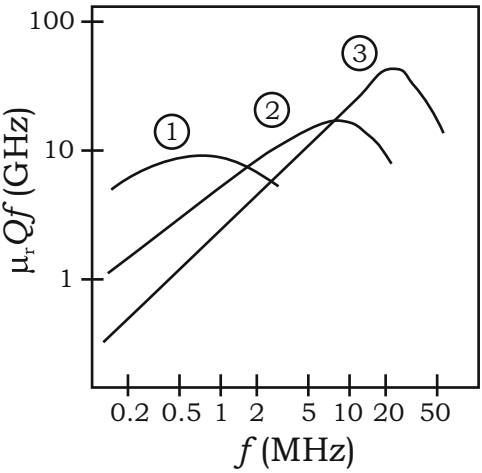

Fig. 4.8 $\mu'_{s,r}Qf$ product versus frequency for three different types of ferrite material (1: Ferroxcube 4A, 2: Ferroxcube 4C, 3: Ferroxcube 4E). Please note that $\mu_r := \mu'_{p,r} \approx \mu'_{s,r}$ holds, due to Eq. (4.23). No bias field is present, and small magnetic RF field amplitudes are assumed. Data taken from [6]

is inserted into Eq. (4.18):

$$\boxed{Q_\mathrm{p} = R_\mathrm{p}\sqrt{\frac{C}{L_\mathrm{p}}}.}$$

In general, all parameters μ'_s, μ''_s, μ'_p, μ''_p, R_s, L_s, R_p, L_p, Q, and Q_p are frequency-dependent. It depends on the material whether the parallel or the series lumped element circuit is the better representation in the sense that its parameters may be regarded as approximately constant in the relevant operating range. In the following, we will use the parallel representation.

We briefly show that Q_p is in fact the quality factor defined by

$$Q_\mathrm{p} = \omega\frac{\overline{W}_\mathrm{tot}}{\overline{P}_\mathrm{loss}},$$

where $\overline{W}_\mathrm{tot}$ is the stored energy and $\overline{P}_\mathrm{loss}$ is the power loss (both time-averaged):

$$\boxed{\overline{P}_\mathrm{loss} = \frac{|\hat{V}_\mathrm{gap}|^2}{2R_\mathrm{p}},} \tag{4.24}$$

$$\overline{W}_\mathrm{el} = \frac{1}{4}\,C\,|\hat{V}_\mathrm{gap}|^2,$$

$$\overline{W}_\mathrm{magn} = \frac{1}{4}\,L_\mathrm{p}\,|\hat{I}_L|^2 = \frac{1}{4}\,L_\mathrm{p}\,\frac{|\hat{V}_\mathrm{gap}|^2}{\omega^2 L_\mathrm{p}^2} = \frac{|\hat{V}_\mathrm{gap}|^2}{4\omega^2 L_\mathrm{p}}.$$

At resonance, we have $\overline{W}_\mathrm{el} = \overline{W}_\mathrm{magn}$, which leads to

$$Q_\mathrm{p} = 2\omega\frac{\overline{W}_\mathrm{el}}{\overline{P}_\mathrm{loss}} = 2\omega\frac{R_\mathrm{p}C}{2} = R_\mathrm{p}\sqrt{\frac{C}{L_\mathrm{p}}},$$

as expected. The parallel resistor R_p defined by Eq. (4.24) is sometimes called a **shunt impedance**. Please note that different definitions for shunt impedance exist in the literature. Sometimes, especially in the LINAC community, the shunt impedance is defined as twice R_p (cf. [7]).

4.1.6 Impedance of the Cavity

The impedance of the cavity

$$Z_\mathrm{tot} = \frac{1}{\frac{1}{R_\mathrm{p}} + j\left(\omega C - \frac{1}{\omega L_\mathrm{p}}\right)} = \frac{\sqrt{\frac{L_\mathrm{p}}{C}}}{\frac{1}{R_\mathrm{p}}\sqrt{\frac{L_\mathrm{p}}{C}} + j\left(\omega\sqrt{L_\mathrm{p}C} - \frac{1}{\omega\sqrt{L_\mathrm{p}C}}\right)}$$

may be written as

$$Z_\mathrm{tot} = \frac{\frac{R_\mathrm{p}}{Q_\mathrm{p}}}{\frac{1}{Q_\mathrm{p}} + j\left(\frac{\omega}{\omega_\mathrm{res}} - \frac{\omega_\mathrm{res}}{\omega}\right)}$$

$$\Rightarrow \boxed{Z_\mathrm{tot} = \frac{R_\mathrm{p}}{1 + j\ Q_\mathrm{p}\left(\frac{\omega}{\omega_\mathrm{res}} - \frac{\omega_\mathrm{res}}{\omega}\right)}}. \tag{4.25}$$

The Laplace transformation yields

$$Z_\mathrm{tot}(s) = \frac{R_\mathrm{p}}{1 + s\frac{Q_\mathrm{p}}{\omega_\mathrm{res}} + \frac{Q_\mathrm{p}\omega_\mathrm{res}}{s}} = \frac{R_\mathrm{p}\frac{\omega_\mathrm{res}}{Q_\mathrm{p}}s}{s\frac{\omega_\mathrm{res}}{Q_\mathrm{p}} + s^2 + \omega_\mathrm{res}^2}, \tag{4.26}$$

which may be found in the literature (cf. [8, 9]) in the form

$$Z_\mathrm{tot}(s) = \frac{2R_\mathrm{p}\sigma\ s}{s^2 + 2\sigma s + \omega_\mathrm{res}^2}$$

if

$$\sigma = \frac{\omega_\mathrm{res}}{2Q_\mathrm{p}}$$

is defined.

We now determine the 3-dB bandwidth[6] of the cavity. The corresponding corner frequencies are reached if the absolute value of the impedance is $1/\sqrt{2}$ of the maximum value R_p. Equation (4.25) shows that this is fulfilled for

$$\frac{\omega}{\omega_\mathrm{res}} - \frac{\omega_\mathrm{res}}{\omega} = \pm\frac{1}{Q_\mathrm{p}} \qquad \Rightarrow \omega^2 \mp \omega\frac{\omega_\mathrm{res}}{Q_\mathrm{p}} - \omega_\mathrm{res}^2 = 0$$

$$\Rightarrow \omega = \pm\frac{\omega_\mathrm{res}}{2Q_\mathrm{p}} \pm \sqrt{\frac{\omega_\mathrm{res}^2}{4Q_\mathrm{p}^2} + \omega_\mathrm{res}^2} = \pm\frac{\omega_\mathrm{res}}{2Q_\mathrm{p}} \pm \omega_\mathrm{res}\sqrt{1 + \frac{1}{4Q_\mathrm{p}^2}}.$$

The absolute value of the second expression is always larger than that of the first expression. Since ω must be positive, one obtains

$$\omega = \omega_\mathrm{res}\sqrt{1 + \frac{1}{4Q_\mathrm{p}^2}} \pm \frac{\omega_\mathrm{res}}{2Q_\mathrm{p}}.$$

This obviously leads to the 3-dB bandwidth

$$\Delta\omega_\mathrm{3dB} = \frac{\omega_\mathrm{res}}{Q_\mathrm{p}} \qquad \Rightarrow \boxed{Q_\mathrm{p} = \frac{\omega_\mathrm{res}}{\Delta\omega_\mathrm{3dB}} = \frac{f_\mathrm{res}}{\Delta f_\mathrm{3dB}}.}$$

4.1.7 Length of the Cavity

In the previous sections, we assumed that the ferrite ring cores can be regarded as lumped-element inductors and resistors. This is, of course, true only if the cavity is short in comparison with the wavelength.

As an alternative to the transformer model introduced above, one may therefore use a coaxial transmission line model. For example, the section of the cavity that is located on the left side of the ceramic gap in Fig. 4.3 may be interpreted as a coaxial line that is homogeneous in the longitudinal direction and that has a short circuit at the left end. The cross section consists of the magnetic material of the ring cores, air between the ring cores and the beam pipe, and air between the ring cores and the cavity housing. This is, of course, an idealization, since cooling disks, conductors, and other air regions are neglected. Taking the SIS18 cavity at GSI as an example, the ring cores have $\mu_\mathrm{r} = 28$ at an operating frequency of 2.5 MHz. The ring cores have a relative dielectric constant of 10–15, but this is reduced to an effective value

[6] Please note that a voltage level of −3 dB below the maximum corresponds to 70.7946% of the peak voltage. This is a good approximation for $1/\sqrt{2}$ (70.7107%).

of $\epsilon_{\mathrm{r,eff}} = 1.8$, since the ring cores do not fill the full cavity cross section. These values lead to a wavelength of $\lambda = 16.9\,\mathrm{m}$. Since 64 ring cores with a thickness of 25 mm are used, the effective length of the magnetic material is $1.6\,\mathrm{m} = 0.095\ \lambda$ (which corresponds to a phase of 34°). In this case, the transmission line model leads to deviations of less than 10% with respect to the lumped-element model. The transmission line model also shows that the above-mentioned estimation for the wavelength is too pessimistic; it leads to $\lambda = 24\,\mathrm{m}$, which corresponds to a cavity length of only 24°.

This type of model makes it understandable why the ferrite cavity is sometimes referred to as a **shortened quarter-wavelength resonator**.

Of course, one may also use more detailed models in which subsections of the cavity are modeled as lumped elements. In that case, computer simulations can be performed to calculate the overall impedance. If one is interested in resonances that may occur at higher operating frequencies, one should perform full electromagnetic simulations.

In any case, one should always remember that some parameters are difficult to determine, especially the permeability of the ring core material under different operating conditions. This uncertainty may lead to larger errors than simplifications of the model. Measurements of full-size ring cores in the requested operating range are inevitable when a new cavity is developed. Also, parameter tolerances due to the manufacturing process have to be taken into account.

In general, one should note that the total length and the dimensions of the cross section of the ferrite cavity are not determined by the wavelength as for a conventional RF cavity. For example, the SIS18 ferrite cavity has a length of 3 m flange to flange, although only 1.6 m is filled with magnetic material. This provides space for the ceramic gap, the cooling disks, and further devices such as the bias current bars. In order to avoid resonances at higher frequencies, one should not waste too much space, but there is no exact size of the cavity housing that results from the electromagnetic analysis.

4.1.8 Differential Equation and Cavity Filling Time

The equivalent circuit shown in Fig. 4.5 was derived in the frequency domain. As long as no parasitic resonances occur, this equivalent circuit may be generalized. Therefore, we may also analyze it in the time domain (allowing slow changes of L_p with time):

$$I_C = C \cdot \frac{\mathrm{d}V_\mathrm{gap}}{\mathrm{d}t}, \qquad V_\mathrm{gap} = L_\mathrm{p} \cdot \frac{\mathrm{d}I_L}{\mathrm{d}t}, \qquad V_\mathrm{gap} = (I_\mathrm{gen} - I_L - I_C - I_\mathrm{beam})\ R_\mathrm{p}$$

$$\Rightarrow I_L = -\frac{V_\mathrm{gap}}{R_\mathrm{p}} + I_\mathrm{gen} - I_C - I_\mathrm{beam} \tag{4.27}$$

$$\Rightarrow V_{\text{gap}} = L_{\text{p}} \left(-\frac{1}{R_{\text{p}}} \frac{\mathrm{d}V_{\text{gap}}}{\mathrm{d}t} + \frac{\mathrm{d}}{\mathrm{d}t}(I_{\text{gen}} - I_{\text{beam}}) - C \frac{\mathrm{d}^2 V_{\text{gap}}}{\mathrm{d}t^2} \right)$$

$$\Rightarrow L_{\text{p}} C \, \ddot{V}_{\text{gap}} + \frac{L_{\text{p}}}{R_{\text{p}}} \, \dot{V}_{\text{gap}} + V_{\text{gap}} = L_{\text{p}} \frac{\mathrm{d}}{\mathrm{d}t}(I_{\text{gen}} - I_{\text{beam}})$$

$$\Rightarrow \boxed{\ddot{V}_{\text{gap}} + \frac{2}{\tau} \, \dot{V}_{\text{gap}} + \omega_{\text{res}}^2 V_{\text{gap}} = \frac{1}{C} \frac{\mathrm{d}}{\mathrm{d}t}(I_{\text{gen}} - I_{\text{beam}}).} \tag{4.28}$$

Here we used the definition

$$\boxed{\tau = 2 R_{\text{p}} C.}$$

The product $R_{\text{p}} C$ is also present in the expression for the quality factor:

$$Q_{\text{p}} = R_{\text{p}} \sqrt{\frac{C}{L_{\text{p}}}} = \frac{R_{\text{p}} C}{\sqrt{L_{\text{p}} C}} = \frac{1}{2} \tau \omega_{\text{res}}$$

$$\Rightarrow \boxed{\tau = \frac{2 Q_{\text{p}}}{\omega_{\text{res}}} = \frac{Q_{\text{p}}}{\pi f_{\text{res}}}.}$$

Under the assumption $\omega_{\text{res}} > \frac{1}{\tau}$ ($Q_{\text{p}} > \frac{1}{2}$), the approach $V_{\text{gap}} = V_0 e^{\alpha t}$ (with a complex constant α) for the homogeneous solution of Eq. (4.28) actually leads to

$$\alpha = -\frac{1}{\tau} \pm j \omega_{\text{d}}$$

with exponential decay time τ and oscillation frequency

$$\omega_{\text{d}} = \omega_{\text{res}} \sqrt{1 - \frac{1}{(\tau \omega_{\text{res}})^2}} = \omega_{\text{res}} \sqrt{1 - \frac{1}{4 Q_{\text{p}}^2}}.$$

This leads to $\omega_{\text{d}} \approx \omega_{\text{res}}$ even for moderately high $Q_{\text{p}} > 2$ (error less than 4%). Sometimes, the resonant frequency ω_{res} is called the **undamped natural frequency**, whereas ω_{d} is called the **damped natural frequency**.

The time τ is the time constant for the cavity, which also determines the **cavity filling time**. Furthermore, the time constant τ is relevant for amplitude and phase jumps of the cavity (see, e.g., [10] and Appendices A.7.1 and A.7.2). We will visualize this fact in Sect. 4.2.

Sometimes, especially in the LINAC community, the cavity filling time is defined as $Q_{\text{p}}/\omega_{\text{res}}$ (one-half of our definition; cf. [7]) in order to specify the energy decay instead of the field strength or voltage decay.

4.1.9 Power Amplifier

Up to now, the Q factor of the cavity has been called Q_p. What we have not mentioned is that the Q factor of the cavity itself is the so-called **unloaded Q factor**. From now on, this unloaded Q factor will be denoted by $Q_{\mathrm{p},0}$. In accordance with this, the parallel resistor will be denoted by $R_{\mathrm{p},0}$. The reason for this is the following: An RF power amplifier that feeds the cavity may often be represented by a voltage-controlled current source (e.g., in the case of a tetrode amplifier as discussed in Chap. 6). The impedance of this current source will be connected in parallel to the equivalent circuit, thereby reducing the ohmic part R_p according to $R_\mathrm{p} = R_{\mathrm{p},0} || R_\mathrm{gen}$ (see Sect. 4.1.12). Therefore, the **loaded Q factor** Q_p will usually be reduced in comparison with the unloaded Q factor $Q_{\mathrm{p},0}$. Also, the cavity filling time will be reduced due to the impedance of the power amplifier.

The formulas that were derived for the parallel equivalent circuit are valid for both cases, the cavity alone and the combination of cavity and amplifier. This is why they were based on R_p.

It must be emphasized that for ferrite cavities, $50\,\Omega$ impedance matching is not necessarily used in general. The cavity impedance is usually on the order of a few hundred ohms or a few kilohms. Therefore, it is often more suitable to connect the tetrode amplifier directly to the cavity. Impedance matching is not mandatory if the amplifier is located close to the cavity. Short cables have to be used, since they contribute to the overall impedance/capacitance. Cavity and RF power amplifier must be considered as one unit; they cannot be developed individually, since that the impedance curves of the cavity and the power amplifier influence each other.

4.1.10 Cooling

Both the power amplifier and the ferrite ring cores need active cooling. Of course, the Curie temperature of the ferrite material (typically $> 100\,^\circ\mathrm{C}$) must never be reached. Depending on the operating conditions (e.g., CW or pulsed operation), **forced air cooling** may be sufficient or **water cooling** may be required. **Cooling disks** between the ferrite cores may be used. In this case, one has to ensure that the thermal contact between cooling disks and ferrite cores is good.

4.1.11 Cavity Tuning

We already mentioned in Sect. 4.1.1 that a DC bias current may be used to decrease μ_Δ, which results in a higher resonant frequency. This is one possible way to realize **cavity tuning**. Strictly speaking, one deals with a quasi-DC bias current, since the resonant frequency must be modified during a synchrotron machine cycle if it is to

equal the variable RF frequency. Such a tuning of the resonant frequency f_{res} to the RF frequency f_{RF} is usually desirable, since the large Z_{tot} makes it possible to generate large voltages with moderate RF power consumption.

Sometimes, the operating frequency range is small enough in comparison with the bandwidth of the cavity that no tuning is required.

If tuning is required, one has at least two possibilities to realize it:

1. Bias current tuning
2. Capacitive tuning

The latter may be realized by a variable capacitor (see, e.g., [11, 12]) whose capacitance may be varied by a stepping motor. This mechanical adjustment, however, is possible only if the resonant frequency is not changed from machine cycle to machine cycle or even within one machine cycle.

In the case of bias current tuning, one has two different choices, namely **perpendicular biasing** (also called **transverse biasing**) and **parallel biasing** (also called **longitudinal biasing**). Also, a mixture of both is possible [13]. The terms parallel and perpendicular refer to the orientation of the DC field H_{bias} in comparison with the RF field H.

Parallel biasing is simple to realize. One adds bias current loops, which may in principle be located in the same way as the inductive coupling loop shown in Fig. 4.3. If only a few loops are present, current bars with large cross sections are needed to withstand the bias current of several hundred amps. The required DC current may, of course, be reduced if the number N_{bias} of loops is increased accordingly (keeping the ampere-turns constant). This increase in the number of bias current windings may be limited by resonances. On the other hand, a minimum number of current loops is usually applied to guarantee a certain amount of symmetry, which leads to a more homogeneous flux in the ring cores.

Perpendicular biasing is more complicated to realize; it requires more space between the ring cores, and the permeability range is smaller than for parallel biasing. The main reason for using perpendicular biasing is that lower losses can be reached (see, e.g., [14]). One can also avoid the so-called **Q-loss effect** or **high loss effect**. The Q-loss effect often occurs when parallel biasing is applied and if the bias current is constant or varies only slowly. After a few milliseconds, one observes that the induced voltage breaks down by a certain amount even though the same amount of RF power is still applied (see, e.g., [15, 16]). For perpendicular biasing, the Q-loss effect was not observed. The Q-loss effect is not fully understood. However, there are strong indications that it may be caused by mechanical resonances of the ring cores induced by magnetostriction effects [17]. It is possible to suppress the Q-loss effect by mechanical damping. For example, in some types of ferrite cavities, the ring cores are embedded in a sealing compound [18], which should damp mechanical oscillations. Not only the Q-loss effect but also further anomalous loss effects have been observed [15].

When the influence of biasing is described, one usually defines an average bias field H_{bias} for the ring cores. For this purpose, one may use the magnetic field

$$H_{\text{bias}} = \frac{N_{\text{bias}} I_{\text{bias}}}{2\pi \bar{r}}$$

located at the mean radius

$$\bar{r} = \sqrt{r_{\text{i}} r_{\text{o}}}.$$

Of course, this choice is somewhat arbitrary from a theoretical point of view, but it is based on practical experience.

A combination of bias current tuning and capacitive tuning has also been applied to extend the frequency range [19].

4.1.12 Resonant Frequency Control

A method that has traditionally been applied to decide whether a ferrite-loaded cavity is at resonance is the measurement of the phase of the gap voltage and of the phase of the control grid voltage (in case of a tetrode power amplifier, cf. Chap. 6).

At first glance, it seems to be clear that the cavity is at resonance if and only if the cavity impedance is purely resistive, i.e., if the inductance of the lumped element (parallel) circuit exactly compensates the capacitance (see Fig. 4.5). Therefore, it is obvious that the generator current and the gap voltage must be in phase for the cavity operating at resonance.

In order to analyze this fact in detail, however, one should be aware that a tube amplifier also contributes to the overall cavity impedance. It is not only the ohmic output impedance that will contribute to the overall impedance of the cavity, but also the capacitance of the tetrode and its circuitry (and also some inductances). As shown in Fig. 4.9, we may use a model [20] in which the tetrode power amplifier is represented as a voltage-controlled current source with an internal resistor R_{gen} in parallel. The capacitance of the power amplifier is represented by a capacitor C_{gen} (if necessary, C_{gen} may be frequency-dependent to include the effect of parasitic inductances). The cavity without the power amplifier is shown on the right side of the circuit. It consists of $R_{\text{p,0}}$, L_{p}, and C_0. Now it becomes obvious that the resonant frequency of the overall system is

$$\omega_{\text{res}} = \frac{1}{\sqrt{L_{\text{p}}(C_0 + C_{\text{gen}})}}.$$

If the cavity is tuned to resonance, it is not the current I_{cav} that is in phase with the gap voltage V_{gap}, but the current I_{gen}. This current I_{gen} cannot be measured, however, since it may be regarded as an internal current of the tetrode. Fortunately, the control grid voltage V_{g1} of the tetrode is usually in phase with the internal current I_{gen} (cf. Chap. 6). Therefore, a **resonant frequency control** loop may compare the phase of V_{g1} with the phase of V_{gap}. If both are in phase (or 180° out of phase, depending

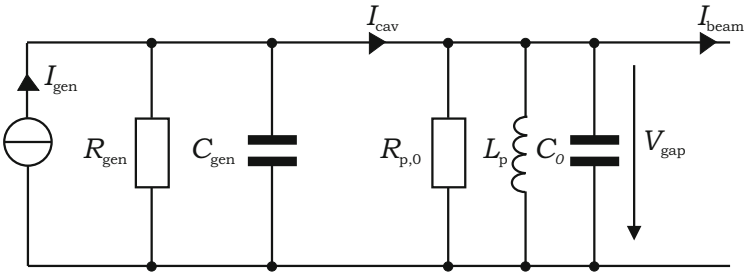

Fig. 4.9 Equivalent circuit of RF generator and cavity

on the chosen orientation of the voltages), the whole cavity system consisting of generator and cavity will be at resonance. One typically uses a **phase detector**, which provides the phase difference between V_{g1} and V_{gap}. This output signal is then used by a closed-loop controller (cf. Chap. 7) to modify the bias current of the ferrite cavity so that I_{gen} and V_{g1} are in phase with V_{gap}, as desired.

It must be emphasized that the equivalent circuit in Fig. 4.5 and the related formulas are still valid. We just have to interpret the circuit in a slightly different way. If we compare it with Fig. 4.9, it becomes obvious that $C = C_{gen} + C_0$ is the total capacitance of generator and cavity, and $R_p = R_{p,0} || R_{gen} = \frac{R_{p,0} R_{gen}}{R_{p,0} + R_{gen}}$ also includes both contributions.

Sometimes, it is desired to operate the cavity not at resonance but slightly off-resonance. In this case, one may choose a target value that differs from zero (or 180°, respectively) for the phase difference.

Completely detuning the cavity may be a choice to deactivate the cavity without having a high beam impedance. Then, of course, no gap voltage is produced, so that the closed-loop control system will not work. However, one may modify the bias current in an open-loop mode in this case.

4.1.13 Further Complications

We already mentioned that the effective differential permeability depends on the hysteresis behavior of the material, i.e., on the history of bias and RF currents. It was also mentioned that due to the spatial dimensions of the cavity, we have to deal with ranges between lumped-element circuits and distributed elements. The anomalous loss effects are a third complication. There are further points that make the situation even more complicated in practice:

- If no biasing is applied, the maximum of the magnetic field is present at the inner radius r_i. One has to ensure that the maximum ratings of the material are not exceeded.

- Bias currents lead to an ρ^{-1} dependency of the induced magnetic field H_{bias}. Therefore, biasing is more effective in the inner parts of the ring cores than in the outer parts, resulting in a μ_Δ that increases with ρ. According to Eq. (4.5), this will modify the ρ^{-1} dependency of the magnetic RF field. As a result, the dependence on ρ may be much weaker than without a bias field.
- The permeability depends not only on the frequency, on the magnetic RF field, and on the biasing. It is also temperature-dependent.
- Depending on the thickness of the ferrite cores, on the conductivity of the ferrite, on the material losses, and on the operating frequency, the magnetic field may decay from the surface to the inner regions, reducing the effective volume.
- At higher operating frequencies with strong bias currents, the differential permeability will be rather low. This means that the magnetic flux will no longer be guided perfectly by the ring cores. The fringe fields in the air regions will be more important, and resonances may occur.

4.1.14 Cavity Configurations

A comparison of different types of ferrite cavities can be found in [21–23]. We summarize a few aspects here that lead to different solutions.

- Instead of using only one stack of ferrite ring cores and only one ceramic gap, as shown in Fig. 4.3, one may also use more sections with ferrites (e.g., one gap with half the ring cores on the left side and the other half on the right side of the gap, for reasons of symmetry) or more gaps. Sometimes, the ceramic gaps belong to different independent cavity cells, which may be coupled by copper bars (e.g., by connecting them in parallel). Connections of this type must be short to allow operation at high frequencies.
- One configuration that is often used is a cavity consisting of only one ceramic gap and two ferrite stacks on each side. Figure-eight windings surround these two ferrite stacks (see, e.g., [24]). With respect to the magnetic RF field, this leads to the same magnetic flux in both stacks. In this way, an RF power amplifier that feeds only one of the two cavity halves will indirectly supply the other cavity half as well. This corresponds to a 1:2 transformation ratio. Hence, the beam will see four times the impedance compared with the amplifier load. Therefore, the same RF input power will lead to higher gap voltages (but also to a higher beam impedance). The transformation law may be derived by an analysis that is similar to that in Sect. 4.1.2.
- Instead of the inductive coupling shown in Fig. 4.3, one may also use capacitive coupling if the power amplifier is connected to the gap via capacitors. If a tetrode power amplifier is used, one still has to provide it with a high anode voltage. Therefore, an external inductor (choke coil) is necessary, which allows the DC anode current but blocks the RF current from the DC power supply. Often

a combination of capacitive and inductive coupling is used (e.g., to influence parasitic resonances). The coupling elements will contribute to the equivalent circuit.
- Another possibility is inductive coupling of individual ring cores. This leads to lower impedances, which ideally allow a 50-Ω impedance matching to a standard solid-state RF power amplifier (see, e.g., [25]).
- If a small relative tuning range is required, it is not necessary to use biasing for the ferrite ring cores inside the cavity. One may use external tuners (see, e.g., [26, 27]), which can be connected to the gap. For external tuners, both parallel and perpendicular biasing may be applied [28].

No general strategy can be defined for how a new cavity is to be designed. Many compromises have to be found. A certain minimum capacitance is given by the gap capacitance and the parasitic capacitances. In order to reach the upper limit of the frequency range, a certain minimum inductance has to be realized. If biasing is used, this minimum inductance must be reached using the maximum bias current, but the effective permeability should still be high enough to reduce stray fields. Also, the lower frequency limit should be reachable with a minimum but nonzero bias current. There is a maximum RF field $B_{\mathrm{RF,max}}$ (about 15 mT), which should not be exceeded for the ring cores. This imposes a lower limit on the number of ring cores. The required tuning range in combination with the overall capacitance will also restrict the number of ring cores. As mentioned above, the amplifier design should be taken into account from the very beginning, especially with respect to the impedance. The maximum **beam impedance** that is tolerable is defined by beam dynamics considerations. This impedance budget also defines the power that is required. If more ring cores can be used, the impedance of the cavity will increase, and the power loss will decrease for a given gap voltage.

4.1.15 The GSI Ferrite Cavities in SIS18

As an example for a ferrite cavity, we summarize the main facts about GSI's SIS18 ferrite cavities (see Figs. 4.10 and 4.11). Two identical ferrite cavities are located in the synchrotron SIS18.

The material Ferroxcube FXC 8C12m is used for the ferrite ring cores. In total, $N = 64$ ring cores are used per cavity. Each core has the following dimensions:

$$d_{\mathrm{o}} = 2\, r_{\mathrm{o}} = 498\,\mathrm{mm}, \quad d_{\mathrm{i}} = 2\, r_{\mathrm{i}} = 270\,\mathrm{mm}, \quad d_{\mathrm{core}} = 25\,\mathrm{mm},$$

$$\bar{r} = \sqrt{r_{\mathrm{i}} r_{\mathrm{o}}} = 183\,\mathrm{mm}.$$

For biasing,

$$N_{\mathrm{bias}} = 6$$

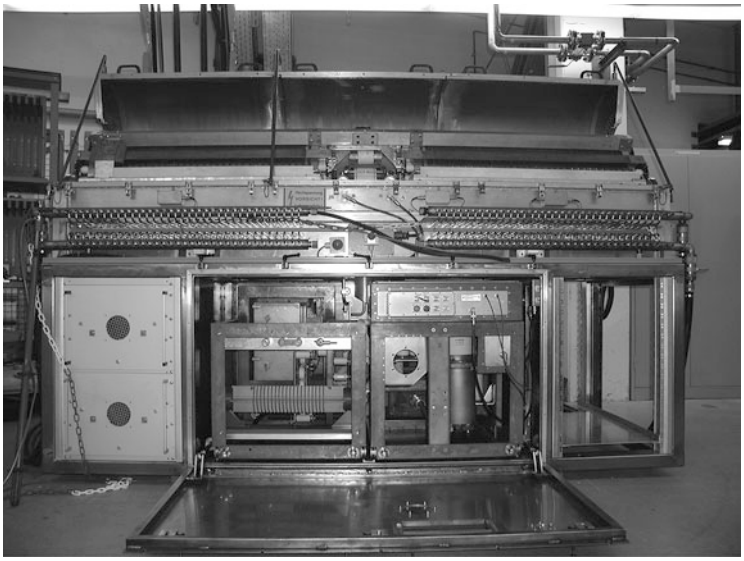

Fig. 4.10 SIS18 ferrite cavity. Photography: GSI Helmholtzzentrum für Schwerionenforschung GmbH, T. Winnefeld

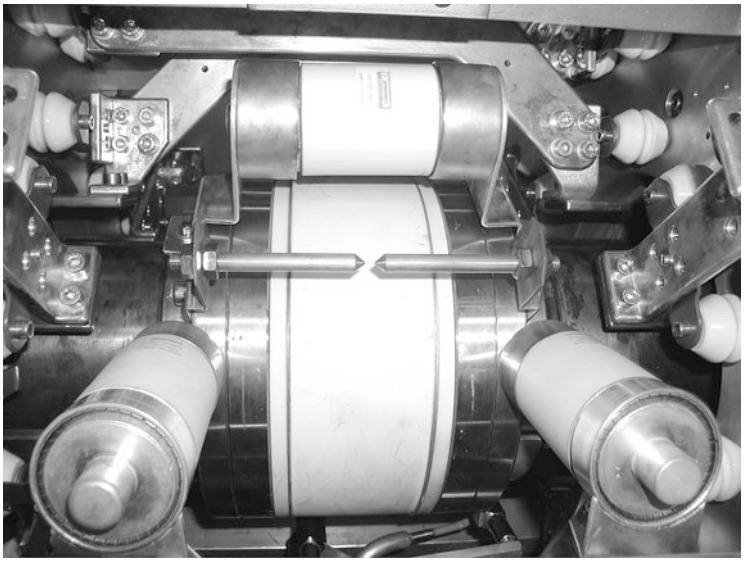

Fig. 4.11 Gap area of the SIS18 ferrite cavity. Photography: GSI Helmholtzzentrum für Schwerionenforschung GmbH, T. Winnefeld

Table 4.1 Equivalent circuit parameters for SIS18 ferrite cavities (without influence of tetrode amplifiers)

Resonant frequency f_{res}	620 kHz	2.5 MHz	5 MHz
Relative permeability $\mu'_{\mathrm{p,r}}$	450	28	7
Magnetic bias field H_{bias} at mean radius	25 A/m	700 A/m	2750 A/m
Bias current I_{bias}	4.8 A	135 A	528 A
$\mu'_{\mathrm{p,r}}Qf$ product	$4.2 \cdot 10^9\ \mathrm{s}^{-1}$	$3.7 \cdot 10^9\ \mathrm{s}^{-1}$	$3.3 \cdot 10^9\ \mathrm{s}^{-1}$
Q-factor $Q_{\mathrm{p,0}}$	15	53	94
L_{s}	88.2 µH	5.49 µH	1.37 µH
L_{p}	88.5 µH	5.49 µH	1.37 µH
$R_{\mathrm{s,0}}$	22.8 Ω	1.63 Ω	0.46 Ω
$R_{\mathrm{p,0}}$	5,200 Ω	4,600 Ω	4,100 Ω
Cavity time constant τ	7.7 µs	6.7 µs	6.0 µs

figure-eight copper windings are present. The total capacitance amounts to

$$C = 740\,\mathrm{pF},$$

including the gap, the gap capacitors, the cooling disks, and other parasitic capacitances. The maximum voltage that is reached under normal operating conditions is $\hat{V}_{\mathrm{gap}} = 16\,\mathrm{kV}$.

Table 4.1 shows the main parameters for three different frequencies. All these values are consistent with the formulas presented in this book. It is obvious that both $\mu'_{\mathrm{p,r}}Qf$ and R_{p} do not vary strongly with frequency, justifying the parallel equivalent circuit. This compensation effect was mentioned at the end of Sect. 4.1.4.

All the parameters mentioned here refer to the beam side of the cavity. The cavity is driven by an RF amplifier coupled to only one of two ferrite core stacks (consisting of 32 ring cores each). The two ring core stacks are coupled by the bias windings. Therefore, a transformation ratio of $1 : 2$ is present from amplifier to beam. This means that the amplifier has to drive a load of about $R_{\mathrm{p,0}}/4 \approx 1.1\,\mathrm{k\Omega}$. For a full amplitude of $\hat{V}_{\mathrm{gap}} = 16\,\mathrm{kV}$ at $f = 5\,\mathrm{MHz}$, the power loss in the cavity amounts to 31 kW.

The SIS18 cavity is supplied by a single-ended tetrode power amplifier using a combination of inductive and capacitive coupling.

It has to be emphasized that the values in Table 4.1 do not contain the amplifier influence. Depending on the operating point of the tetrode, R_{p} will be reduced significantly in comparison with $R_{\mathrm{p,0}}$, and all related parameters vary accordingly.

4.1.16 Further Practical Considerations

For measuring the gap voltage, one needs a **gap voltage divider** in order to decrease the high-voltage RF to a safer level. This can be done by capacitive voltage dividers.

Gap relays are used to short-circuit the gap if the cavity is temporarily unused. This reduces the beam impedance, which may be harmful for beam stability. If cycle-by-cycle switching is needed, semiconductor switches may be used as gap switches instead of vacuum relays. Another possibility to temporarily reduce the beam impedance is to detune the cavity.

The capacitance/impedance of the gap periphery devices must be considered when the overall capacitance C and the other elements in the equivalent circuit are calculated. Also, further parasitic elements may be present.

On the one hand, the cavity dimensions should be as small as possible, since space in synchrotrons and storage rings is valuable and since undesired resonances may be avoided. On the other hand, certain minimum distances have to be kept in order to prevent high-voltage sparkovers. For reasons of **EMC** (**electromagnetic compatibility**), RF seals are often used between conducting metal parts of the cavity housing to reduce electromagnetic emission.

In order to satisfy high vacuum requirements, it may be necessary to allow a **bakeout** of the vacuum chamber. This can be realized by integrating a **heating jacket** that surrounds the beam pipe. It has to be guaranteed that the ring cores are not damaged by heating and that safety distances (for RF purposes and high-voltage requirements) are kept.

If the cavity is used in a radiation environment, the radiation hardness of all materials is an important topic.

4.1.17 Magnetic Materials

A large variety of magnetic materials is available. Nickel-zinc (NiZn) ferrites may be regarded as the traditional standard material for ferrite-loaded cavities. At least the following material properties are of interest for material selection, and they may differ significantly for different types of material:

- permeability
- magnetic losses
- saturation induction (typically 200–300 mT for NiZn ferrites)
- maximum RF inductions (typically 10–20 mT for NiZn ferrites)
- relative dielectric constant (on the order of 10–15 for NiZn ferrites but, e.g., very high for MnZn ferrites) and dielectric losses (usually negligible for typical NiZn applications)
- maximum operating temperature, thermal conductivity, and temperature dependence in general
- magnetostriction
- specific resistance (very high for NiZn ferrites, very low for MnZn ferrites)

In order to determine the RF properties under realistic operating conditions (large magnetic flux, biasing), thorough reproducible measurements in a fixed test setup are inevitable.

Amorphous and nanocrystalline magnetic alloy (MA) materials have been used to build very compact cavities that are based on similar principles as the classical ferrite cavities (see, e.g., [12, 23, 29–31]). These materials allow a higher induction and have a very high permeability. This means that a smaller number of ring cores is needed for the same inductance. MA materials typically have lower Q factors than those of ferrite materials. Low Q factors have the advantage that frequency tuning is often not necessary and that it is possible to generate signal forms including higher harmonics instead of pure sine signals (cf. Sect. 5.5.1). MA cavities are especially of interest for pulsed operation at high field gradients. If a low Q-factor is not desired, it is also possible to increase it by cutting the MA ring cores.

Microwave garnet ferrites have been used at frequencies in the range 40–60 MHz in connection with perpendicular biasing, since they provide comparatively low losses (see, e.g., [32–34]).

4.2 Cavity Excitation

In Sect. 4.1.8, we derived the differential equation (4.28) that is obtained for the standard lumped element circuit of the cavity shown in Fig. 4.5. This is valid for several types of cavities.

We already mentioned that the cavity time constant τ determines how the cavity reacts to excitations, i.e., changes in the generator current and the beam current.

In Appendix A.7.1, a solution (A.37) of the ODE (4.28) is derived for a special excitation, namely that the sinusoidal generator current is switched on.

We now evaluate this solution for a specific case. Consider the following parameters as an example:

$$R_\mathrm{p} = 2\,\mathrm{k\Omega}, \qquad C = 500\,\mathrm{pF}, \qquad L_\mathrm{p} = 50.66\,\mu\mathrm{H}.$$

This leads to

$$Q_\mathrm{p} \approx 6.28, \qquad \tau = 2\,\mu\mathrm{s}$$

and a resonant frequency of 1 MHz. The impedance is shown in Fig. 4.12.

Figures 4.13, 4.14, 4.15, 4.16, 4.17, and 4.18 show what happens if the cavity is excited with a current

$$I_\mathrm{gen}(t) = \hat{I}_\mathrm{gen}\ \sin(\omega t)$$

that is switched on at $t = 0$ with an amplitude of $\hat{I}_\mathrm{gen} = 5\,\mathrm{A}$.

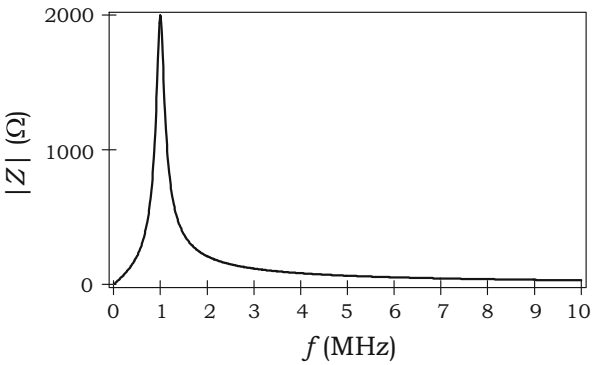

Fig. 4.12 Impedance of the cavity versus frequency

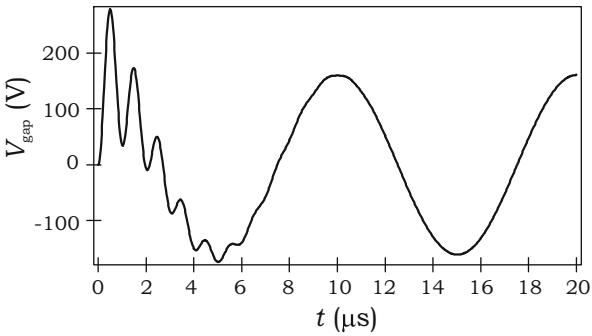

Fig. 4.13 Excitation with $f = 0.1$ MHz

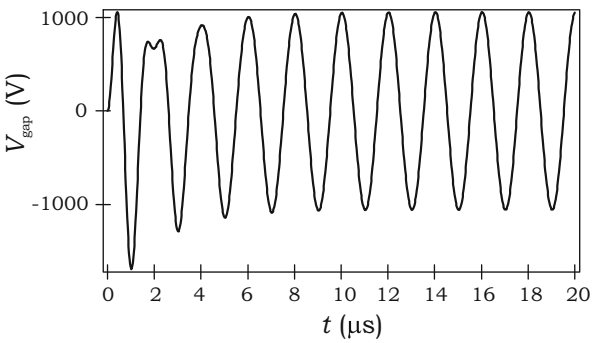

Fig. 4.14 Excitation with $f = 0.5$ MHz

If the excitation frequency differs much from the resonant frequency of the cavity, the cavity first tries to oscillate with the resonant frequency. After a while, this transient behavior ends, and the cavity oscillates with the excitation frequency. However, no significant voltage is obtained.

If the excitation frequency is close to the resonant frequency, one sees some overshoot before the stationary conditions are reached.

If the cavity is excited with its resonant frequency, the maximum voltage is achieved for a given current. The time constant τ is clearly visible.

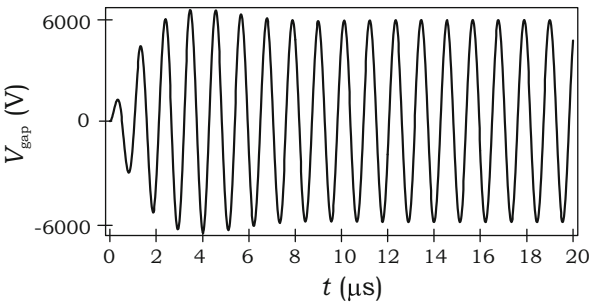

Fig. 4.15 Excitation with $f = 0.9\,\text{MHz}$

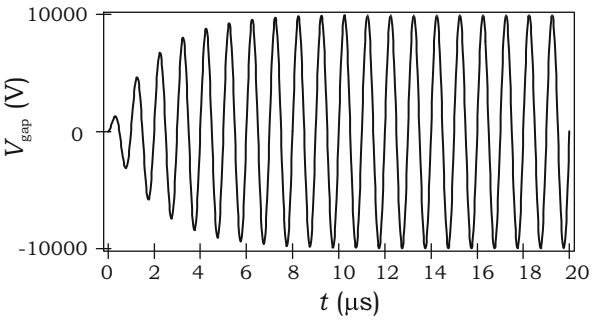

Fig. 4.16 Excitation with $f = 1\,\text{MHz}$

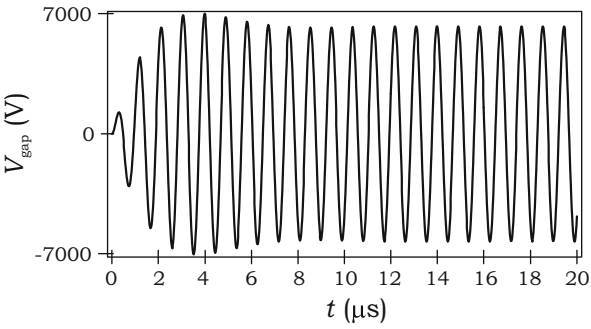

Fig. 4.17 Excitation with $f = 1.1\,\text{MHz}$

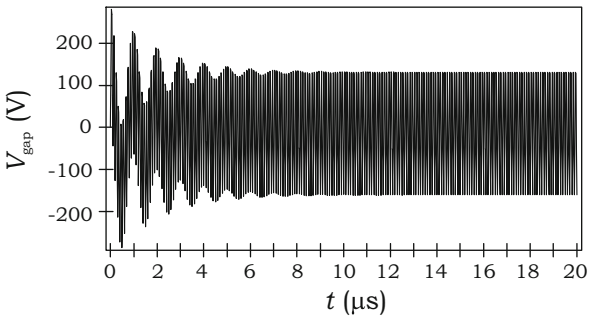

Fig. 4.18 Excitation with $f = 10\,\mathrm{MHz}$

4.3 Transit Time Factor

In the simplest case, the longitudinal component of the electric field may be regarded as constant in the ceramic gap, and it has a harmonic shape:

$$E(t) = \hat{E}\ \cos(\omega_{\mathrm{RF}} t).$$

We now consider a particle that passes the center of the gap at $z = 0$ exactly at the time $t = 0$, when the electric field is at its crest.

The particle then experiences the voltage

$$V = \int_{-\Delta l_{\mathrm{gap}}/2}^{+\Delta l_{\mathrm{gap}}/2} E\ \mathrm{d}z = \int_{-\Delta t_{\mathrm{gap}}/2}^{+\Delta t_{\mathrm{gap}}/2} E(t)\ u_{\mathrm{R}}\ \mathrm{d}t.$$

Here we assume that the percentage of change in the particle velocity is small enough that one may simply write

$$u_{\mathrm{R}} = \frac{z}{t} = \frac{\Delta l_{\mathrm{gap}}}{\Delta t_{\mathrm{gap}}}.$$

The quantity Δt_{gap} denotes the time of flight through the gap. In this way, we obtain

$$V = u_{\mathrm{R}}\hat{E} \int_{-\Delta t_{\mathrm{gap}}/2}^{+\Delta t_{\mathrm{gap}}/2} \cos(\omega_{\mathrm{RF}} t)\ \mathrm{d}t = 2u_{\mathrm{R}}\hat{E}\left[\frac{\sin(\omega_{\mathrm{RF}} t)}{\omega_{\mathrm{RF}}}\right]_0^{\Delta t_{\mathrm{gap}}/2}$$

$$= 2u_{\mathrm{R}}\hat{E}\,\frac{\sin(\omega_{\mathrm{RF}}\Delta t_{\mathrm{gap}}/2)}{\omega_{\mathrm{RF}}}$$

$$\Rightarrow V = \Delta l_{\mathrm{gap}}\hat{E}\ \frac{\sin(\omega_{\mathrm{RF}}\Delta t_{\mathrm{gap}}/2)}{\omega_{\mathrm{RF}}\Delta t_{\mathrm{gap}}/2}.$$

If the time of flight is, according to

$$\Delta t_{\mathrm{gap}} \ll T_{\mathrm{RF}} \Leftrightarrow \omega_{\mathrm{RF}}\Delta t_{\mathrm{gap}} \ll 2\pi,$$

small in comparison with the RF period, it follows that

$$V \to V_0 = \Delta l_{\mathrm{gap}} \hat{E}.$$

If this is not the case, one obtains

$$V = V_0 \cdot f_{\mathrm{T}},$$

where f_{T} denotes the **transit time factor**

$$f_{\mathrm{T}} = \mathrm{si}\left(\omega_{\mathrm{RF}} \Delta t_{\mathrm{gap}}/2\right) = \mathrm{si}\left(\frac{\omega_{\mathrm{RF}} \Delta l_{\mathrm{gap}}}{2u_{\mathrm{R}}}\right).$$

In a synchrotron,

$$\omega_{\mathrm{RF}} = h\frac{2\pi}{T_{\mathrm{R}}}$$

and

$$u_{\mathrm{R}} = \frac{l_{\mathrm{R}}}{T_{\mathrm{R}}}$$

are valid, so that

$$f_{\mathrm{T}} = \mathrm{si}\left(\frac{\pi h \Delta l_{\mathrm{gap}}}{l_{\mathrm{R}}}\right)$$

is obtained.

In the synchrotron SIS18 at GSI, the gap length Δl_{gap} of the ferrite cavities amounts to about 0.1 m, the circumference of the synchrotron is $l_{\mathrm{R}} = 216\,\mathrm{m}$, and the maximum harmonic number is 4. This leads to

$$f_{\mathrm{T}} \approx 0.999994,$$

so that the transit time effect is not at all relevant.

The situation is different in a linear accelerator, in which a gap length of 10 cm may have a significant effect. In the linear accelerator UNILAC at GSI, $\beta = 0.15$ is reached at an RF frequency of 108 MHz, so that

$$f_{\mathrm{T}} \approx 0.9079$$

holds.

The transit time factor that was derived here should be regarded as only a rule of thumb to check whether the transit time effect is relevant. The formula was based on the following drastic simplifications:

- The arrival time was chosen in such a way that the particle experiences the maximum field strength.
- The field was assumed to be homogeneous in the longitudinal direction.
- The relative change in the particle velocity inside the gap was assumed to be small.

4.4 Pillbox Cavity

The pillbox cavity may be regarded as the fundamental cavity type, especially in linear accelerators. Other accelerating structures may often be considered as a modification or combination of pillbox cavities. Therefore, the pillbox cavity is analyzed in this section. The notation used here is close to that used in [35, 36].

Firstly, we discuss the TE modes and the TM modes in circular waveguides (perfectly conducting hollow cylinder). In both cases, a field solution for the pillbox cavity will afterward be generated by introducing ideally conducting end plates.

4.4.1 TM Modes

We begin with the equations for the vector potential (2.50) and the scalar potential (2.51). Since we are interested in only time-harmonic solutions, phasors will be used—even though we do not mark[7] them specifically as phasors in this section:

$$\Delta\vec{A} + \frac{\omega^2}{c^2}\,\vec{A} = -\mu\vec{J},$$

$$\Delta\Phi + \frac{\omega^2}{c^2}\,\Phi = -\frac{\rho_q}{\epsilon}.$$

The inside of the cavity is evacuated and is therefore free of charges and free of currents. Hence, we are looking for solutions of the **vector Helmholtz equation**

$$\Delta\vec{A} + k^2\,\vec{A} = 0$$

and the **scalar Helmholtz equation**

[7] In other chapters of this book, the phasor of a time-domain function $X(t)$ is denoted by $\underline{\hat{X}}$ in the frequency domain.

$$\Delta\Phi + k^2\,\Phi = 0$$

with

$$k = \frac{\omega}{c_0},$$

which are coupled by the Lorenz gauge condition (2.49)

$$\text{div}\ \vec{A} = -j\omega\mu_0\epsilon_0\Phi. \tag{4.29}$$

As we will see, nontrivial solutions are still possible if

$$\vec{A} = A_z^E\vec{e}_z \tag{4.30}$$

is assumed. This leads to the following equation:

$$\Delta A_z^E + k^2\,A_z^E = 0. \tag{4.31}$$

According to Eqs. (2.47) and (2.48), the fields can be derived from the vector potential by means of

$$\vec{B} = \text{curl}\ \vec{A} = \vec{e}_\rho\frac{1}{\rho}\frac{\partial A_z^E}{\partial\varphi} - \vec{e}_\varphi\frac{\partial A_z^E}{\partial\rho} \tag{4.32}$$

and

$$\vec{E} = -j\omega\vec{A} - \text{grad}\ \Phi = -j\omega\vec{A} - \frac{1}{-j\omega\mu_0\epsilon_0}\ \text{grad div}\ \vec{A} = -j\omega\vec{A} + \frac{c_0^2}{j\omega}\ \text{grad div}\ \vec{A}.$$

Here, Eq. (4.29) was used. For $\vec{A} = A_z^E\vec{e}_z$, we have

$$\text{div}\ \vec{A} = \frac{\partial A_z^E}{\partial z}$$

and

$$\text{grad div}\ \vec{A} = \vec{e}_\rho\,\frac{\partial^2 A_z^E}{\partial\rho\partial z} + \vec{e}_\varphi\,\frac{1}{\rho}\,\frac{\partial^2 A_z^E}{\partial\varphi\partial z} + \vec{e}_z\,\frac{\partial^2 A_z^E}{\partial z^2},$$

so that

$$\vec{E} = \frac{c_0^2}{j\omega}\left[\vec{e}_\rho\,\frac{\partial^2 A_z^E}{\partial\rho\partial z} + \vec{e}_\varphi\,\frac{1}{\rho}\,\frac{\partial^2 A_z^E}{\partial\varphi\partial z} + \vec{e}_z\left(\frac{\partial^2 A_z^E}{\partial z^2} + \frac{\omega^2}{c_0^2}\,A_z^E\right)\right] \tag{4.33}$$

is the result. Equations (4.32) and (4.33) were obtained by expressing the curl, div, and grad operators in cylindrical coordinates. The first result shows that our assumption (4.30) for the vector potential implies that there are no longitudinal components of the magnetic field. We will therefore get **transverse magnetic (TM) waves**, which are also called **E waves**. This explains why we called the longitudinal component of the vector potential A_z^E. In cylindrical coordinates, Eq. (4.31) can be written as

$$\frac{\partial^2 A_z^E}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial A_z^E}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 A_z^E}{\partial \varphi^2} + \frac{\partial^2 A_z^E}{\partial z^2} + k^2 A_z^E = 0.$$

We solve this equation by inserting the separation ansatz (cf. [36])

$$A_z^E = A_0(\rho, \varphi) A_3(z).$$

After division by A_z^E, one gets

$$\left(\frac{1}{A_0}\frac{\partial^2 A_0}{\partial \rho^2} + \frac{1}{\rho}\frac{1}{A_0}\frac{\partial A_0}{\partial \rho} + \frac{1}{\rho^2}\frac{1}{A_0}\frac{\partial^2 A_0}{\partial \varphi^2}\right) + \frac{1}{A_3}\frac{\partial^2 A_3}{\partial z^2} + k^2 = 0.$$

It is obvious that the first term in parentheses may depend only on ρ and φ, and the second term only on z. Since the last term, k^2, however, does not depend on any of these coordinates, the two terms must be constant:

$$\begin{aligned}
\frac{1}{A_0}\frac{\partial^2 A_0}{\partial \rho^2} + \frac{1}{\rho}\frac{1}{A_0}\frac{\partial A_0}{\partial \rho} + \frac{1}{\rho^2}\frac{1}{A_0}\frac{\partial^2 A_0}{\partial \varphi^2} &= -C_0, \\
\frac{1}{A_3}\frac{\partial^2 A_3}{\partial z^2} &= -C_3, \\
C_0 + C_3 &= k^2.
\end{aligned}$$

If we multiply the first equation by ρ^2, we get

$$\left(\rho^2\frac{1}{A_0}\frac{\partial^2 A_0}{\partial \rho^2} + \rho\frac{1}{A_0}\frac{\partial A_0}{\partial \rho} + C_0\rho^2\right) + \frac{1}{A_0}\frac{\partial^2 A_0}{\partial \varphi^2} = 0.$$

We insert the separation ansatz

$$A_0 = A_1(\rho)\, A_2(\varphi)$$

and obtain

$$\left(\rho^2\frac{1}{A_1}\frac{\partial^2 A_1}{\partial \rho^2} + \rho\frac{1}{A_1}\frac{\partial A_1}{\partial \rho} + C_0\rho^2\right) + \frac{1}{A_2}\frac{\partial^2 A_2}{\partial \varphi^2} = 0.$$

The first term in parentheses may depend only on ρ, whereas the second term may depend only on φ. Therefore, they must both be constant:

$$\rho^2 \frac{1}{A_1} \frac{\partial^2 A_1}{\partial \rho^2} + \rho \frac{1}{A_1} \frac{\partial A_1}{\partial \rho} + C_0 \rho^2 = C_1,$$

$$\frac{1}{A_2} \frac{\partial^2 A_2}{\partial \varphi^2} = -C_1.$$

The three ordinary differential equations may be rewritten in the following form:

$$\rho^2 \frac{\mathrm{d}^2 A_1}{\mathrm{d}\rho^2} + \rho \frac{\mathrm{d}A_1}{\mathrm{d}\rho} + (C_0 \rho^2 - C_1) A_1 = 0, \tag{4.34}$$

$$\frac{\mathrm{d}^2 A_2}{\mathrm{d}\varphi^2} + C_1 A_2 = 0, \tag{4.35}$$

$$\frac{\mathrm{d}^2 A_3}{\mathrm{d}z^2} + C_3 A_3 = 0, \tag{4.36}$$

$$C_0 + C_3 = k^2. \tag{4.37}$$

For real constants $C_1, C_3 > 0$, the last two differential equations describe a harmonic oscillator, so that the solutions

$$A_3 = A_{31} \cos(\sqrt{C_3} z) + A_{32} \sin(\sqrt{C_3} z) \tag{4.38}$$

and

$$A_2 = A_{21} \cos(\sqrt{C_1} \varphi) + A_{22} \sin(\sqrt{C_1} \varphi) \tag{4.39}$$

are obvious.[8] In the first equation (4.34), the substitution $\tilde{\rho} = \sqrt{C_0}\rho$ leads to

$$\tilde{\rho}^2 \frac{\mathrm{d}^2 A_1}{\mathrm{d}\tilde{\rho}^2} + \tilde{\rho} \frac{\mathrm{d}A_1}{\mathrm{d}\tilde{\rho}} + (\tilde{\rho}^2 - C_1) A_1 = 0.$$

This is equivalent to **Bessel's differential equation**

$$\boxed{x^2 \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + x \frac{\mathrm{d}y}{\mathrm{d}x} + (x^2 - m^2) y = 0,} \tag{4.40}$$

[8] The assumption $C_1 < 0$ would lead to solutions that are not 2π-periodic with respect to φ, which cannot provide unique field values, because an angle advance of 2π corresponds to the same point inside the circular cross section. For the same reason, the constant C_1 must be real; it cannot be complex. The assumption $C_3 < 0$ would lead to evanescent waves. Since the circular waveguide is free of losses, complex values of C_3 are impossible; the transmitted power must be constant in the z-direction. Due to Eq. (4.37) a real C_3 also leads to a real C_0.

where we define

$$m = \sqrt{C_1}.$$

A set of solutions is given by the **Bessel functions of the first kind** $\mathrm{J}_m(x)$ and **Bessel functions of the second kind** $\mathrm{Y}_m(x)$ (also called **Neumann functions** $\mathrm{N}_m(x)$)—see Table A.2 on p. 415 and Figs. A.5 and A.6. In our case, we get

$$A_1 = A_{11}\mathrm{J}_m(\sqrt{C_0}\rho) + A_{12}\mathrm{Y}_m(\sqrt{C_0}\rho).$$

Since the field components and A_z^E must not change if integer multiples of 2π are added to φ, the restriction $m \in \{0, 1, 2, \dots\}$ must be valid, as Eq. (4.39) shows. Negative m do not lead to new degrees of freedom ($\mathrm{J}_{-m} = (-1)^m\mathrm{J}_m$, $\mathrm{Y}_{-m} = (-1)^m\mathrm{Y}_m$) and can therefore be excluded. Because $\mathrm{Y}_m(x)$ has a pole at $x = 0$, which would lead to singularities of the field components at $\rho = 0$, it cannot correspond to a physical solution. Hence, we omit the last term. The total solution is therefore given by

$$A_z^E = \mathrm{J}_m(\sqrt{C_0}\rho)\,[A_{21}\ \cos(m\varphi) + A_{22}\ \sin(m\varphi)]\left[A_{31}\ \cos(\sqrt{C_3}z) + A_{32}\ \sin(\sqrt{C_3}z)\right].$$

We finally define $K = \sqrt{C_0}$ and $k_z = \sqrt{C_3}$, so that

$$A_z^E = \mathrm{J}_m(K\rho)\,[A_{21}\ \cos(m\varphi) + A_{22}\ \sin(m\varphi)]\,[A_{31}\ \cos(k_z z) + A_{32}\ \sin(k_z z)]$$

with

$$K^2 + k_z^2 = k^2$$

is valid. According to Eqs. (4.33) and (4.32), this leads to the following field components:

$$\begin{aligned} E_\rho &= \frac{c_0^2}{j\omega}\,\frac{\partial^2 A_z^E}{\partial\rho\partial z} = \\ &= \frac{c_0^2}{j\omega}\,Kk_z\,\mathrm{J}'_m(K\rho)\,[A_{21}\ \cos(m\varphi) + A_{22}\ \sin(m\varphi)] \\ &\quad \times [-A_{31}\ \sin(k_z z) + A_{32}\ \cos(k_z z)]\,, \end{aligned} \tag{4.41}$$

$$\begin{aligned} E_\varphi &= \frac{c_0^2}{j\omega}\,\frac{1}{\rho}\,\frac{\partial^2 A_z^E}{\partial\varphi\partial z} = \\ &= \frac{c_0^2}{j\omega}\,\frac{m}{\rho}\,k_z\mathrm{J}_m(K\rho)\,[-A_{21}\ \sin(m\varphi) + A_{22}\ \cos(m\varphi)] \\ &\quad \times [-A_{31}\ \sin(k_z z) + A_{32}\ \cos(k_z z)]\,, \end{aligned} \tag{4.42}$$

$$
\begin{aligned}
E_z &= \frac{c_0^2}{j\omega}(k^2 - k_z^2)A_z^E = \\
&= \frac{c_0^2}{j\omega}\, K^2\, \mathrm{J}_m(K\rho)\,[A_{21}\ \cos(m\varphi) + A_{22}\ \sin(m\varphi)] \\
&\quad \times [A_{31}\ \cos(k_z z) + A_{32}\ \sin(k_z z)]\,, \qquad (4.43) \\
B_\rho &= \frac{1}{\rho}\frac{\partial A_z^E}{\partial \varphi} = \\
&= \frac{m}{\rho}\, \mathrm{J}_m(K\rho)\,[-A_{21}\ \sin(m\varphi) + A_{22}\ \cos(m\varphi)] \\
&\quad \times [A_{31}\ \cos(k_z z) + A_{32}\ \sin(k_z z)]\,, \qquad (4.44) \\
B_\varphi &= -\frac{\partial A_z^E}{\partial \rho} = \\
&= -K\, \mathrm{J}'_m(K\rho)\,[A_{21}\ \cos(m\varphi) + A_{22}\ \sin(m\varphi)] \\
&\quad \times [A_{31}\ \cos(k_z z) + A_{32}\ \sin(k_z z)]\,, \qquad (4.45) \\
B_z &= 0. \qquad (4.46)
\end{aligned}
$$

At the conducting surface at $\rho = r_{\text{pillbox}}$, the tangential components of the electric field (i.e., E_φ, E_z) and the normal component of the magnetic field (i.e., B_ρ) must vanish.[9] This leads to the condition

$$\mathrm{J}_m(Kr_{\text{pillbox}}) = 0.$$

The zeros of $\mathrm{J}_m(x)$ with $x > 0$ will be denoted by j_{mn}, where $n \in \{1, 2, 3, \dots\}$. Selecting n therefore determines K according to

$$K = \frac{j_{mn}}{r_{\text{pillbox}}}.$$

In conclusion, the field pattern of a TM mode is determined by the two numbers m and n. Therefore, one writes TM_{mn} or E_{mn} to specify the modes in a circular waveguide. The propagation constant for such a mode is given by

$$k_z = \sqrt{k^2 - K^2}.$$

[9] If we had assumed $C_0 < 0$, Bessel's modified differential equation would have been obtained. This ODE will be discussed later in this book, and its solutions without poles, the modified Bessel functions $\mathrm{I}_m(x)$, do not have zeros for $x > 0$. Therefore, it would not have been possible to satisfy this boundary condition at $\rho = r_{\text{pillbox}}$.

It is obvious that a minimum (angular) frequency $\omega = \omega_c$ is needed if k_z is to be real, which means that the wave is propagating. This **cutoff frequency** of the TM mode under consideration is given by $k = K$, which is equivalent to

$$\omega_c = c_0 \frac{j_{mn}}{r_{\text{pillbox}}}.$$

A pillbox cavity corresponds to a circular waveguide with additional conducting walls at $z = 0$ and $z = l_{\text{pillbox}}$. Since E_ρ and E_φ must be zero at $z = 0$, one then selects $A_{32} = 0$. Because one has to satisfy $E_\rho = E_\varphi = 0$ at $z = l_{\text{pillbox}}$ in addition, the condition

$$k_z l_{\text{pillbox}} = p\pi$$

must hold. For characterizing the modes in a resonator, we therefore have to introduce a third number $p \in \{0, 1, 2, \dots\}$. If the three numbers m, n, and p are fixed, then K and k_z are fixed as well, and the equation

$$\frac{\omega}{c_0} = k = \sqrt{k_z^2 + K^2}$$

determines the resonant frequency $f_{\text{res}} = \frac{\omega_{\text{res}}}{2\pi} = \frac{\omega}{2\pi}$. In a cavity with perfectly conducting walls, electromagnetic fields may exist only at these discrete resonant frequencies (eigenvalue problem). The corresponding modes are denoted by TM_{mnp} or E_{mnp}.

4.4.2 TE Modes

The solutions derived in the previous section are based on the assumption that the inside of the waveguide or the cavity is evacuated. Therefore, the Maxwell's equations that had to be solved reduce to

$$\text{curl } \vec{H} = j\omega\epsilon_0 \vec{E}, \tag{4.47}$$

$$\text{curl } \vec{E} = -j\omega\mu_0 \vec{H}, \tag{4.48}$$

$$\text{div } \vec{B} = 0, \tag{4.49}$$

$$\text{div } \vec{D} = 0. \tag{4.50}$$

The general approach to deriving the vector potential is based on the equation

$$\text{div } \vec{B} = 0,$$

so that

$$\vec{B} = \text{curl } \vec{A}$$

was required above. Now we see that in the special case that the waveguide is evacuated, we can alternatively satisfy

$$\text{div } \vec{D} = 0$$

by defining

$$\vec{D} = \text{curl } \vec{A}. \tag{4.51}$$

If we insert this into Eq. (4.47), we get

$$\text{curl} \left(\vec{H} - j\omega \, \vec{A} \right) = 0,$$

which may be satisfied if the scalar potential Φ is defined according to

$$\vec{H} - j\omega \, \vec{A} = \text{grad } \Phi \qquad \Rightarrow \vec{H} = j\omega \, \vec{A} + \text{grad } \Phi. \tag{4.52}$$

The expressions for $\vec{D}$ and $\vec{H}$ are now inserted into Eq. (4.48):

$$\begin{aligned} &\text{curl curl } \vec{A} = -j\omega\mu_0\epsilon_0 \left(j\omega \, \vec{A} + \text{grad } \Phi \right) \\ \Rightarrow &\text{grad div } \vec{A} - \Delta\vec{A} = \frac{\omega^2}{c_0^2} \, \vec{A} - j \frac{\omega}{c_0^2} \text{ grad } \Phi \\ \Rightarrow &\Delta\vec{A} + k^2\vec{A} = \text{grad} \left(\text{div } \vec{A} + j \frac{\omega}{c_0^2} \, \Phi \right). \end{aligned}$$

By means of the gauge condition

$$\text{div } \vec{A} = -j \frac{\omega}{c_0^2} \, \Phi, \tag{4.53}$$

this leads to the vector Helmholtz equation

$$\Delta\vec{A} + k^2\vec{A} = 0.$$

Our new ansatz for the vector potential obviously leads to the same equations to be solved as in the previous section. However, in this case, the curl of the vector potential now determines the electric field instead of the magnetic one. One therefore speaks of **TE waves** instead of TM waves.

We now determine the field components of the TE waves (also called **H waves**) in the same way as those for the TM waves (also known as E waves). Therefore,

$$\vec{A} = A_z^H \vec{e}_z \tag{4.54}$$

is assumed. This leads to the following equation:

$$\Delta A_z^H + k^2 \, A_z^H = 0. \tag{4.55}$$

According to Eqs. (4.51) and (4.52), the fields can be derived from the vector potential by means of

$$\vec{D} = \text{curl } \vec{A} = \vec{e}_\rho \frac{1}{\rho} \frac{\partial A_z^H}{\partial \varphi} - \vec{e}_\varphi \frac{\partial A_z^H}{\partial \rho} \tag{4.56}$$

and

$$\vec{H} = j\omega \vec{A} + \text{grad } \Phi = j\omega \vec{A} - \frac{c_0^2}{j\omega} \text{ grad div } \vec{A}.$$

Here, Eq. (4.53) was used. For $\vec{A} = A_z^H \vec{e}_z$, we have

$$\text{div } \vec{A} = \frac{\partial A_z^H}{\partial z}$$

and

$$\text{grad div } \vec{A} = \vec{e}_\rho \, \frac{\partial^2 A_z^H}{\partial \rho \partial z} + \vec{e}_\varphi \, \frac{1}{\rho} \, \frac{\partial^2 A_z^H}{\partial \varphi \partial z} + \vec{e}_z \, \frac{\partial^2 A_z^H}{\partial z^2},$$

so that

$$\vec{H} = -\frac{c_0^2}{j\omega} \left[\vec{e}_\rho \, \frac{\partial^2 A_z^H}{\partial \rho \partial z} + \vec{e}_\varphi \, \frac{1}{\rho} \, \frac{\partial^2 A_z^H}{\partial \varphi \partial z} + \vec{e}_z \left(\frac{\partial^2 A_z^H}{\partial z^2} + \frac{\omega^2}{c_0^2} \, A_z^H \right) \right] \tag{4.57}$$

is the result. Equations (4.56) and (4.57) were obtained by expressing the curl, div, and grad operators in cylindrical coordinates. The first result shows that our assumption (4.54) for the vector potential implies that there are no longitudinal components of the electric field. We will therefore get transverse electric (TE) waves. This explains why we called the longitudinal component of the vector potential A_z^H.

Since the Helmholtz equation (4.55) is still identical to that of the E mode derivation (4.31), the same separation ansatz may be used as in the previous section.

The total solution is therefore given by

$$A_z^H = \text{J}_m(K\rho) \, [A_{21} \, \cos(m\varphi) + A_{22} \, \sin(m\varphi)] \, [A_{31} \, \cos(k_z z) + A_{32} \, \sin(k_z z)]$$

with

$$K^2 + k_z^2 = k^2.$$

According to Eqs. (4.56) and (4.57), this leads to the following field components:

$$
\begin{aligned}
D_\rho &= \frac{1}{\rho}\frac{\partial A_z^H}{\partial \varphi} = \frac{m}{\rho}\,\mathrm{J}_m(K\rho)\,[-A_{21}\ \sin(m\varphi) + A_{22}\ \cos(m\varphi)] \\
&\quad \times [A_{31}\ \cos(k_z z) + A_{32}\ \sin(k_z z)]\,, \\
D_\varphi &= -\frac{\partial A_z^H}{\partial \rho} = -K\,\mathrm{J}'_m(K\rho)\,[A_{21}\ \cos(m\varphi) + A_{22}\ \sin(m\varphi)] \\
&\quad \times [A_{31}\ \cos(k_z z) + A_{32}\ \sin(k_z z)]\,, \\
D_z &= 0, \\
H_\rho &= -\frac{c_0^2}{j\omega}\,\frac{\partial^2 A_z^H}{\partial \rho \partial z} = \\
&= -\frac{c_0^2}{j\omega}\,K k_z\,\mathrm{J}'_m(K\rho)\,[A_{21}\ \cos(m\varphi) + A_{22}\ \sin(m\varphi)] \\
&\quad \times [-A_{31}\ \sin(k_z z) + A_{32}\ \cos(k_z z)]\,, \\
H_\varphi &= -\frac{c_0^2}{j\omega}\,\frac{1}{\rho}\,\frac{\partial^2 A_z^H}{\partial \varphi \partial z} = \\
&= -\frac{c_0^2}{j\omega}\,\frac{m}{\rho}\,k_z \mathrm{J}_m(K\rho)\,[-A_{21}\ \sin(m\varphi) + A_{22}\ \cos(m\varphi)] \\
&\quad \times [-A_{31}\ \sin(k_z z) + A_{32}\ \cos(k_z z)]\,, \\
H_z &= -\frac{c_0^2}{j\omega}(k^2 - k_z^2) A_z^H = \\
&= -\frac{c_0^2}{j\omega}\,K^2\,\mathrm{J}_m(K\rho)\,[A_{21}\ \cos(m\varphi) + A_{22}\ \sin(m\varphi)] \\
&\quad \times [A_{31}\ \cos(k_z z) + A_{32}\ \sin(k_z z)]\,.
\end{aligned}
$$

At the conducting surface at $\rho = r_{\text{pillbox}}$, the tangential component of the electric field (i.e., E_φ and hence also D_φ) and the normal component of the magnetic field (i.e., B_ρ and hence also H_ρ) must vanish. This leads to the condition

$$\mathrm{J}'_m(K r_{\text{pillbox}}) = 0.$$

The zeros of $\mathrm{J}'_m(x)$ with $x > 0$ will be denoted by j'_{mn}, where $n \in \{1, 2, 3, \dots\}$. Selecting n therefore determines K according to

$$K = \frac{j'_{mn}}{r_{\text{pillbox}}}.$$

In conclusion, the field pattern of a TE mode is determined by the two numbers m and n. Therefore, one writes $\mathrm{TE}_{\mathrm{mn}}$ or H_{mn} to specify the modes in a circular waveguide. The propagation constant for such a mode is given by

$$k_z = \sqrt{k^2 - K^2}.$$

It is obvious that a minimum (angular) frequency $\omega = \omega_{\mathrm{c}}$ is needed if k_z is to be real, which means that the wave is propagating. This frequency is given by $k = K$, which is equivalent to

$$\omega_{\mathrm{c}} = c_0 \frac{j'_{mn}}{r_{\mathrm{pillbox}}}.$$

We again introduce end plates at $z = 0$ and $z = l_{\mathrm{pillbox}}$ in order to convert the waveguide into a pillbox cavity. Since D_ρ and D_φ must be zero at $z = 0$, one then selects $A_{31} = 0$. Because one has to satisfy $D_\rho = D_\varphi = 0$ at $z = l_{\mathrm{pillbox}}$ in addition, the condition

$$k_z l_{\mathrm{pillbox}} = p\pi$$

must hold ($p \in \{1, 2, \ldots\}$). Please note that $p = 0$ is not an option for TE modes, since all field components disappear in that case. If the three numbers m, n, and p are fixed, K and k_z are fixed as well, and the equation

$$\frac{\omega}{c_0} = k = \sqrt{k_z^2 + K^2}$$

determines the resonant frequency. The corresponding modes are denoted by $\mathrm{TE}_{\mathrm{mnp}}$ and $\mathrm{H}_{\mathrm{mnp}}$.

According to Table 4.2 on p. 216, j'_{11} is smaller than j_{01}. Hence, the dominant mode in the circular waveguide, i.e., the mode with the lowest cutoff frequency, is the TE_{11} mode.

Since the TE modes do not have any longitudinal component of the electric field, they cannot be used for acceleration in a pillbox cavity. Therefore, in a pillbox cavity, the mode TM_{010} is usually used for acceleration. In this case, the TM_{010} mode should also be the dominant mode. Therefore, one has to exclude the situation in which the resonant frequency of the TE_{111} mode is lower. For the TE_{111} mode, we have

$$k^2 = \frac{\omega_{\mathrm{res}}^2}{c_0^2} = k_z^2 + K^2$$

with

$$k_z = \frac{\pi}{l_{\mathrm{pillbox}}}, \qquad K = \frac{j'_{11}}{r_{\mathrm{pillbox}}}.$$

Table 4.2 Zeros j_{mn} of the Bessel function $\mathrm{J}_m(x)$ and zeros j'_{mn} of its derivative $\mathrm{J}'_m(x)$; values taken from [37]

j_{mn} / j'_{mn}	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=1$	**2.40483**	3.83171	5.13562	6.38016	7.58834
	3.83171	**1.84118**	3.05424	4.20119	5.31755
$n=2$	5.52008	7.01559	8.41724	9.76102	11.06471
	7.01559	5.33144	6.70613	8.01524	9.28240
$n=3$	8.65373	10.17347	11.61984	13.01520	14.37254
	10.17347	8.53632	9.96947	11.34592	12.68191
$n=4$	11.79153	13.32369	14.79595	16.22347	17.61597
	13.32369	11.70600	13.17037	14.58585	15.96411

Please note that $\mathrm{J}'_0(x)$ also has a zero at $x=0$, which explains the modified order in [37]. The smallest values for j_{mn} and for j'_{mn}, respectively, are printed in bold type.

For the TM_{010} mode,

$$k = K = \frac{j_{01}}{r_{\mathrm{pillbox}}}$$

is valid. Therefore we require

$$\left(\frac{\pi}{l_{\mathrm{pillbox}}}\right)^2 + \left(\frac{j'_{11}}{r_{\mathrm{pillbox}}}\right)^2 > \left(\frac{j_{01}}{r_{\mathrm{pillbox}}}\right)^2,$$

$$\frac{r_{\mathrm{pillbox}}}{l_{\mathrm{pillbox}}} > \frac{1}{\pi}\sqrt{j_{01}^2 - j'^2_{11}} \approx 0.49.$$

Roughly speaking, the length of the cavity must therefore be smaller than the diameter.

4.4.3 *Energy Considerations for the* TM_{010} *Mode*

As mentioned above, the mode that is used for acceleration is the TM_{010} mode. For this mode,

$$m=0, \qquad n=1, \qquad p=0 \quad \Rightarrow k_z = 0 \quad \Rightarrow K = k = \frac{\omega}{c_0}$$

is valid.

Equations (4.41)–(4.46) then reduce to

$$E_z = \frac{c_0^2}{j\omega}\, K^2\, \mathrm{J}_0(K\rho)\, A_{21}\, A_{31} = -j\omega\, A_0\, \mathrm{J}_0(K\rho), \tag{4.58}$$

$$B_\varphi = -K\ \mathrm{J}_0'(K\rho)\ A_{21}\ A_{31} = -A_0\ K\ \mathrm{J}_0'(K\rho). \tag{4.59}$$

Here we defined $A_0 = A_{21}\ A_{31}$ as a new constant. The voltage that is generated along the z-axis is

$$\hat{V} = \left| \int_0^{l_{\text{pillbox}}} E_z(\rho = 0)\ \mathrm{d}z \right| = \omega\ |A_0|\ \mathrm{J}_0(0)\ l_{\text{pillbox}} = \omega\ |A_0|\ l_{\text{pillbox}}, \tag{4.60}$$

where $\mathrm{J}_0(0) = 1$ was used.

4.4.3.1 Electric Energy

We now calculate the total time-averaged **electric energy** in the cavity:

$$\begin{aligned}
\overline{W}_{\text{el}} &= \frac{1}{4}\mathrm{Re}\left\{ \int \vec{D}^* \cdot \vec{E}\ \mathrm{d}V \right\} = \\
&= \frac{1}{4}\epsilon_0 \int_0^{l_{\text{pillbox}}} \int_0^{r_{\text{pillbox}}} \int_0^{2\pi} |E_z|^2 \rho\ \mathrm{d}\varphi\ \mathrm{d}\rho\ \mathrm{d}z = \frac{2\pi}{4} l_{\text{pillbox}} \epsilon_0 \int_0^{r_{\text{pillbox}}} |E_z|^2 \rho\ \mathrm{d}\rho = \\
&= \frac{\pi}{2} l_{\text{pillbox}} \epsilon_0 \omega^2 |A_0|^2 \int_0^{r_{\text{pillbox}}} \mathrm{J}_0^2(K\rho)\ \rho\ \mathrm{d}\rho.
\end{aligned}$$

According to Gradshteyn [38, Sect. 5.5, formula 5.54-2],

$$\boxed{\int \mathrm{J}_m^2(K\rho)\ \rho\ \mathrm{d}\rho = \frac{\rho^2}{2}\left\{ \mathrm{J}_m^2(K\rho) - \mathrm{J}_{m-1}(K\rho)\ \mathrm{J}_{m+1}(K\rho) \right\}}$$

$$\Rightarrow \int \mathrm{J}_0^2(K\rho)\ \rho\ \mathrm{d}\rho = \frac{\rho^2}{2}\left\{ \mathrm{J}_0^2(K\rho) - \mathrm{J}_{-1}(K\rho)\ \mathrm{J}_1(K\rho) \right\}$$

is valid. Due to Eq. (A.72), Table A.2 on p. 415, we have

$$\mathrm{J}_{-m}(z) = (-1)^m \mathrm{J}_m(z) \qquad \Rightarrow \mathrm{J}_{-1}(z) = -\mathrm{J}_1(z),$$

so that

$$\int \mathrm{J}_0^2(K\rho)\ \rho\ \mathrm{d}\rho = \frac{\rho^2}{2}\left\{ \mathrm{J}_0^2(K\rho) + \mathrm{J}_1^2(K\rho) \right\}$$

holds. In our case, the upper integration limit is r_{pillbox}, which leads to the argument $K\rho = K r_{\text{pillbox}} = j_{01}$. Therefore, we obtain

$$\int_0^{r_{\text{pillbox}}} \mathrm{J}_0^2(K\rho)\ \rho\ \mathrm{d}\rho = \frac{r_{\text{pillbox}}^2}{2}\ \mathrm{J}_1^2(j_{01}).$$

This leads to

$$\overline{W}_{\text{el}} = \frac{\pi}{4} l_{\text{pillbox}} r_{\text{pillbox}}^2 \epsilon_0 \omega^2 |A_0|^2\ \mathrm{J}_1^2(j_{01}) = \frac{\pi}{4} \frac{r_{\text{pillbox}}^2}{l_{\text{pillbox}}} \epsilon_0 \hat{V}^2\ \mathrm{J}_1^2(j_{01}). \tag{4.61}$$

In the last step, Eq. (4.60) was used.

4.4.3.2 Magnetic Energy

For the sake of completeness, we now check that the stored (time-averaged) **magnetic energy** equals the electric energy stored in the cavity:

$$\begin{aligned}
\overline{W}_{\text{magn}} &= \frac{1}{4}\mathrm{Re}\left\{\int \vec{H}^* \cdot \vec{B}\ \mathrm{d}V\right\} = \\
&= \frac{1}{4\mu_0}\int_0^{l_{\text{pillbox}}}\int_0^{r_{\text{pillbox}}}\int_0^{2\pi} |B_\varphi|^2 \rho\ \mathrm{d}\varphi\ \mathrm{d}\rho\ \mathrm{d}z = \frac{2\pi l_{\text{pillbox}}}{4\mu_0}\int_0^{r_{\text{pillbox}}} |B_\varphi|^2 \rho\ \mathrm{d}\rho = \\
&= \frac{\pi}{2}\frac{l_{\text{pillbox}}}{\mu_0} K^2 |A_0|^2 \int_0^{r_{\text{pillbox}}} J'^2_0(K\rho)\ \rho\ \mathrm{d}\rho.
\end{aligned}$$

According to Table A.2 on p. 415, Eq. (A.77),

$$\mathrm{J}'_0(z) = -\mathrm{J}_1(z),$$

is valid, which leads to

$$\overline{W}_{\text{magn}} = \frac{\pi}{2}\frac{l_{\text{pillbox}}}{\mu_0} K^2 |A_0|^2 \int_0^{r_{\text{pillbox}}} \mathrm{J}_1^2(K\rho)\ \rho\ \mathrm{d}\rho.$$

We may again apply formula 5.54-2 in [38]:

$$\int \mathrm{J}_1^2(K\rho)\ \rho\ \mathrm{d}\rho = \frac{\rho^2}{2}\left\{\mathrm{J}_1^2(K\rho) - \mathrm{J}_0(K\rho)\ \mathrm{J}_2(K\rho)\right\}.$$

One gets

$$\int_0^{r_{\text{pillbox}}} \mathrm{J}_1^2(K\rho)\ \rho\ \mathrm{d}\rho = \frac{r_{\text{pillbox}}^2}{2}\ \mathrm{J}_1^2(j_{01}). \tag{4.62}$$

This leads to

$$\overline{W}_{\text{magn}} = \frac{\pi}{4}\frac{l_{\text{pillbox}} r_{\text{pillbox}}^2}{\mu_0} K^2 |A_0|^2 \, \mathrm{J}_1^2(j_{01})$$
$$= \frac{\pi}{4}\frac{l_{\text{pillbox}} r_{\text{pillbox}}^2}{\mu_0}\frac{\omega^2}{c_0^2}|A_0|^2 \, \mathrm{J}_1^2(j_{01}) = \frac{\pi}{4}\frac{r_{\text{pillbox}}^2}{l_{\text{pillbox}}}\epsilon_0 \hat{V}^2 \, \mathrm{J}_1^2(j_{01}),$$

which verifies $\overline{W}_{\text{el}} = \overline{W}_{\text{magn}}$ after comparison with Eq. (4.61). In the last step, Eq. (4.60) was used.

4.4.3.3 Power Loss

According to the power loss method, the (time-averaged) losses in the conductor (conductivity κ) can be calculated approximately by

$$\boxed{\overline{P}_{\text{loss}} = \frac{R_{\text{surf}}}{2}\int |H_t|^2 \, \mathrm{d}A.}$$

Here

$$\boxed{R_{\text{surf}} = \frac{1}{\kappa\delta}}$$

denotes the **surface resistivity** with **skin depth**

$$\boxed{\delta = \sqrt{\frac{2}{\omega\mu\kappa}}.}$$

In the scope of the power loss method, one assumes that the fields outside the conductor (i.e., inside the cavity) do not change significantly if the ideal conductors are replaced by real ones with sufficiently high κ. Therefore, the tangential field H_t according to our solution (4.59) in the previous section can be used. For the cylindrical surface at $\rho = r_{\text{pillbox}}$, we therefore have

$$H_t = H_\varphi(\rho = r_{\text{pillbox}}) = -A_0 \frac{K}{\mu_0} \mathrm{J}_0'(K r_{\text{pillbox}}) = A_0 \frac{K}{\mu_0} \mathrm{J}_1(j_{01}),$$

which leads to

$$\overline{P}_{\text{loss},1} = \frac{R_{\text{surf}}}{2} 2\pi r_{\text{pillbox}} l_{\text{pillbox}} \, |A_0|^2 \, \frac{K^2}{\mu_0^2} \, \mathrm{J}_1^2(j_{01}) =$$
$$= \pi R_{\text{surf}} \frac{r_{\text{pillbox}}}{l_{\text{pillbox}}} \, \hat{V}^2 \, \frac{\epsilon_0}{\mu_0} \, \mathrm{J}_1^2(j_{01}) = \pi \frac{R_{\text{surf}}}{Z_0^2} \frac{r_{\text{pillbox}}}{l_{\text{pillbox}}} \, \hat{V}^2 \, \mathrm{J}_1^2(j_{01}).$$

Here

$$\boxed{Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 376.73\ \Omega}$$

denotes the **impedance of free space**.

For each of the two end plates of the pillbox cavity, one obtains

$$\begin{aligned}\overline{P}_{\text{loss},2} &= \frac{R_{\text{surf}}}{2}\int_0^{r_{\text{pillbox}}}\int_0^{2\pi} |H_\varphi|^2 \rho \, \mathrm{d}\varphi \, \mathrm{d}\rho = \\ &= \frac{R_{\text{surf}}}{2\mu_0^2} 2\pi |A_0|^2 K^2 \int_0^{r_{\text{pillbox}}} \mathrm{J}_0'^2(K\rho)\ \rho \, \mathrm{d}\rho = \\ &= \pi R_{\text{surf}} \frac{\epsilon_0}{\mu_0} |A_0|^2 \omega^2 \int_0^{r_{\text{pillbox}}} \mathrm{J}_1^2(K\rho)\ \rho \, \mathrm{d}\rho.\end{aligned}$$

Here we may use the result (4.62) and Eq. (4.60) again:

$$\overline{P}_{\text{loss},2} = \frac{\pi}{2} \frac{R_{\text{surf}}}{Z_0^2} \hat{V}^2 \frac{r_{\text{pillbox}}^2}{l_{\text{pillbox}}^2} \mathrm{J}_1^2(j_{01}).$$

Since two end plates are present, the total **power loss** is

$$\overline{P}_{\text{loss}} = \overline{P}_{\text{loss},1} + 2\,\overline{P}_{\text{loss},2} = \pi \frac{R_{\text{surf}}}{Z_0^2}\ \hat{V}^2\, \mathrm{J}_1^2(j_{01}) \left(\frac{r_{\text{pillbox}}}{l_{\text{pillbox}}} + \frac{r_{\text{pillbox}}^2}{l_{\text{pillbox}}^2} \right). \tag{4.63}$$

This leads to the Q factor

$$\begin{aligned}Q_{\text{p},0} &= \omega_{\text{res}} \frac{\overline{W}_{\text{total}}}{\overline{P}_{\text{loss}}} = \omega_{\text{res}} \frac{2\overline{W}_{\text{el}}}{\overline{P}_{\text{loss}}} \\ &= \omega_{\text{res}} \frac{\frac{\pi}{2} \frac{r_{\text{pillbox}}^2}{l_{\text{pillbox}}} \epsilon_0 \hat{V}^2\, \mathrm{J}_1^2(j_{01})}{\pi \frac{R_{\text{surf}}}{Z_0^2}\ \hat{V}^2\, \mathrm{J}_1^2(j_{01}) \left(\frac{r_{\text{pillbox}}}{l_{\text{pillbox}}} + \frac{r_{\text{pillbox}}^2}{l_{\text{pillbox}}^2} \right)} = \frac{\omega_{\text{res}} r_{\text{pillbox}} \mu_0}{2R_{\text{surf}} \left(1 + \frac{r_{\text{pillbox}}}{l_{\text{pillbox}}} \right)}.\end{aligned}$$

Due to

$$\boxed{\omega_{\text{res}} = c_0 \frac{j_{01}}{r_{\text{pillbox}}},} \tag{4.64}$$

this may also be written as

$$\boxed{Q_{\mathrm{p},0} = \frac{c_0 \mu_0 j_{01}}{2R_{\mathrm{surf}}\left(1 + \frac{r_{\mathrm{pillbox}}}{l_{\mathrm{pillbox}}}\right)} = \frac{Z_0 j_{01}}{2R_{\mathrm{surf}}\left(1 + \frac{r_{\mathrm{pillbox}}}{l_{\mathrm{pillbox}}}\right)} = \frac{r_{\mathrm{pillbox}}}{\delta} \frac{1}{1 + \frac{r_{\mathrm{pillbox}}}{l_{\mathrm{pillbox}}}}.} \tag{4.65}$$

Equation (4.63) leads directly to the shunt impedance

$$\boxed{R_{\mathrm{p},0} = \frac{\hat{V}^2}{2\,\overline{P}_{\mathrm{loss}}} = \frac{Z_0^2}{2\pi R_{\mathrm{surf}}\,\mathrm{J}_1^2(j_{01})\frac{r_{\mathrm{pillbox}}}{l_{\mathrm{pillbox}}}\left(1 + \frac{r_{\mathrm{pillbox}}}{l_{\mathrm{pillbox}}}\right)}} \tag{4.66}$$

with

$$j_{01} \approx 2.40483 \qquad \text{and} \qquad \mathrm{J}_1(j_{01}) \approx 0.51915.$$

Again we have to emphasize that the shunt impedance is often defined with an additional factor of 2. Furthermore, the transit time factor may be included in the definition of the shunt impedance (cf. [39, 40] or [41, vol. II, Sect. 6.1.4]).

4.4.4 *Practical Considerations*

In order to calculate the losses of the pillbox cavity, we introduced a transition from ideal conductors to realistic conductors with finite conductivity. The cavity with ideal conductors does not have any losses at all. Therefore, $Q \to \infty$ and $R_{\mathrm{p},0} \to \infty$ are valid. According to Eqs. (4.65) and (4.66), this result is also obtained for $\kappa \to \infty$, $\delta = 0$, $R_{\mathrm{surf}} = 0$.

In the case of a lossless cavity, fields are present only for specific resonant frequencies. If losses are present, this discrete spectrum of resonant frequencies is transformed into a continuous one. The resonant frequencies correspond to local maxima of the absolute value of the impedance; each resonance has a finite bandwidth in the frequency domain (according to the Q factor of that specific mode). At the fundamental resonance, one therefore usually describes the pillbox cavity with the same lumped element circuit (parameters ω_{res}, $Q_{\mathrm{p},0}$, $R_{\mathrm{p},0}$, Eq. (4.25), Fig. 4.5) that was obtained previously for the ferrite cavity.

In the analytical calculations above, we completely neglected the beam pipe, the beam itself, and coupling elements. The beam pipe of course has a smaller diameter than the cavity, but it still allows undesired **higher-order modes (HOM)** to propagate. Of course, the presence of the beam pipe also modifies the solutions obtained above. The beam itself is sometimes modeled as an RF current; it induces fields inside the cavity (beam loading) that may act back on the beam.

Coupling elements are needed both to excite the fields inside the cavity and to measure them. These coupling elements (e.g., small coupling loops) are usually

designed in such a way that they can be connected to waveguides or transmission lines with a specific impedance (e.g. 50 Ω).

Instead of the unmodified pillbox cavity, one often uses rounded geometries. Several individual cavities may be combined into a larger structure in order to supply it with only one RF power source.

A cooling system is required to keep the lossy parts at constant temperature. This is especially important for high Q factors, since temperature changes may otherwise lead to significant drifts of the resonant frequency.

Since an unmodified pillbox cavity cannot be tuned in real time, i.e., since its resonant frequency cannot be changed during operation, it is suitable only for synchrotrons that work with a fixed RF frequency. This is, for example, possible if ultrarelativistic electrons are accelerated where $\beta \approx 1$ is valid. Of course, tuning is at least necessary during the commissioning phase of a cavity. Due to tolerances, the desired resonant frequency will usually not be hit after manufacture. Therefore, possibilities to slightly modify the geometry must be offered. For this purpose, plungers may, for example, be moved during normal operation in LINAC structures.

4.4.5 Example

As an example, we consider an 805-MHz pillbox cavity [42, 43]. The geometric dimensions are

$$r_{\text{pillbox}} = 0.1562\,\text{m}, \qquad l_{\text{pillbox}} = 0.0519\,\text{m}, \qquad r_{\text{pillbox}}/l_{\text{pillbox}} = 3.0096.$$

This leads to a theoretical resonant frequency of

$$f_{\text{res}} = \frac{c_0 j_{01}}{2\pi r_{\text{pillbox}}} = 734.6\,\text{MHz}.$$

In reality, the modifications of the ideal pillbox cavity (e.g., rounded edges, beam pipe, coupling elements, etc.) lead to the above-mentioned resonant frequency of 805 MHz. At this frequency, one gets

$$\delta = 2.3292\,\mu\text{m}, \qquad R_{\text{surf}} = 7.40225\,\text{m}\Omega$$

if one assumes a conductivity of $5.8 \cdot 10^7$ S/m (copper). With Eq. (4.65), this leads to

$$Q_{\text{p},0} = 15262,$$

which is close to the measured value of $Q_{\text{p},0} = 15080$ [43]. According to Eq. (4.66), one obtains

$$R_{\text{p},0} = 943\,\text{k}\Omega, \qquad \Rightarrow R_{\text{p},0}/l_{\text{pillbox}} = 18.18\,\text{M}\Omega/\text{m}.$$

The design value in [42, 43] corresponds to 19 MΩ/m, since the definition of the shunt impedance differs by a factor of 2 from our circuit definition and since the transit time factor is included in [43] (but not in [42]).

References

1. J.M. Brennan, Ferrite loaded cavities (Chapter 7.3.8), in *Handbook of Accelerator Physics and Engineering*, ed. by A.W. Chao, M. Tigner (World Scientific, Singapore, 1999), pp. 570–572
2. H. Klingbeil, Ferrite cavities, in *CAS - CERN Accelerator School: RF for Accelerators*, Ebeltoft, 8–17 Jun 2010, pp. 299–317
3. E. Jensen, Cavity basics, in *CAS - CERN Accelerator School: RF for Accelerators*, Ebeltoft, 8–17 Jun 2010, pp. 259–275
4. F. Gerigk, Cavity types, in *CAS - CERN Accelerator School: RF for Accelerators*, Ebeltoft, 8–17 Jun 2010, pp. 277–298
5. I.D. Mayergoyz, *Mathematical Models of Hysteresis and Their Applications* (Elsevier Science, New York, 2003)
6. F.G. Brockman, H. van der Heide, M.W. Louwerse, Ferroxcube für Protonensynchrotrons. *Philips Technische Rundschau* **30**(11/12), 323–342 (1969/1970)
7. F. Gerigk, RF basics I and II, in *CAS - CERN Accelerator School: High Power Hadron Machines*, Bilbao, 24 May–02 Jun 2011, pp. 71–116
8. F. Pedersen, Beam loading effects in the CERN PS booster. IEEE Trans. Nucl. Sci. **22**(3), 1906–1909 (1975)
9. D. Boussard, Stability considerations, in *7th LEP Performance Workshop*, Chamonix, Jan 1997, pp. 108–113
10. S. Papureanu, Ch. Hamm, A. Schnase, H. Meuth, Performance test of a ferrite-loaded cavity under operation conditions, in *15th IEEE Particle Accelerator Conference*, Washington, May 1993
11. M. Morvillo, R. Garoby, D. Grier, M. Haase, A. Krusche, P. Maesen, M. Paoluzzi, C. Rossi, The PS 13.3–20 MHz RF system for LHC, in *20th IEEE Particle Accelerator Conference*, Portland, May 2003
12. P. Hülsmann, G. Hutter, W. Vinzenz, The bunch compressor system for SIS18 at GSI, in *EPAC* (1994), pp. 1165–1167
13. C. Völlinger, F. Caspers, E. Jensen, The effect of 2-directional magnetic biasing used for tuning of a ferrite-loaded re-entrant cavity. IEEE Trans. Nucl. Sci. **NS-60**(3), 2170–2174 (2013)
14. W.R. Smythe, T.G. Brophy, RF cavities with transversely biased ferrite tuning. IEEE Trans. Nucl. Sci. **NS-32**(5), 2951–2953 (1985)
15. J.E. Griffin, G. Nicholls, A review of some dynamic loss properties of Ni-Zn accelerator RF system ferrite. IEEE Trans. Nucl. Sci. **26**(3), 3965–3967 (1979)
16. K. Kaspar, H.G. König, T. Winnefeld, Studies on maximum RF voltages in ferrite-tuned accelerating cavities, in *EPAC* (2004), pp. 985–987
17. H.G. König, S. Schäfer, Reduction of Q-loss-effects in ferrite-loaded cavities, in *EPAC* (2008), pp. 985–987
18. V.S. Arbuzov et al., Accelerating RF station for HIRFL-CSR, Lanzhou, China, in *RuPAC XIX*, Dubna (2004), pp. 332–334
19. X. Pei, S. Anderson, D. Jenner, D. McCammon, T. Sloan, A wide tuning range Rf cavity with external ferrite biasing, in *15th IEEE Particle Accelerator Conference*, Washington, May 1993
20. U. Hartel, Modellierung des Regelungs- und Steuerungssystems einer Beschleunigungseinheit für Synchrotrons. Diplomarbeit, Technische Universität Darmstadt, Darmstadt, 2011
21. A. Susini, Low frequency ferrite cavities, in *EPAC* (1988), pp. 1416–1417

22. I.S.K. Gardner, Ferrite dominated cavities, in *CAS - CERN Accelerator School: RF Engineering for Particle Accelerators*, April 1991, pp. 349–374
23. A. Schnase, Cavities with a swing, in *CAS - CERN Accelerator School: Radio Frequency Engineering*, May 2000, pp. 236–272
24. A. Krusche, M. Paoluzzi, The new low frequency accelerating systems for the CERN PS booster, in *EPAC* (1998)
25. J. Dey, I. Kourbanis, D. Wildman, A new RF system for bunch coalescing in the Fermilab main ring, in *16th IEEE Particle Accelerator Conference*, Dallas, May 1995
26. R.M. Hutcheon, A perpendicular-biased ferrite tuner for the 52 MHz PETRA II cavities, in *12th IEEE Particle Accelerator Conference*, Washington, March 1987
27. C.C. Friedrichs, R.D. Carlini, G. Spalek, W.R. Smythe, Test results of the Los Alamos ferrite-tuned cavity, in *12th IEEE Particle Accelerator Conference*, Washington, March 1987
28. R.L. Poirier, Perpendicular biased ferrite-tuned cavities, in *15th IEEE Particle Accelerator Conference*, Washington, May 1993
29. C. Fougeron, P. Ausset, D. de Menezes, J. Peyromaure, G. Charruau, Very wide range and short accelerating cavity for MIMAS, in *15th IEEE Particle Accelerator Conference*, Washington, May 1993
30. K. Saito, K. Matsuda, H. Nishiuchi, M. Umezawa, K. Hiramoto, R. Shinagawa, RF accelerating system for a compact ion synchrotron, in *19th IEEE Particle Accelerator Conference*, Chicago, June 2001
31. R. Stassen, K. Bongardt, F.J. Etzkorn, H. Stockhorst, S. Papureanu, A. Schnase, The HESR RF-system and tests in COSY, in *EPAC* (2008)
32. L.M. Earley, H.A. Thiessen, R. Carlini, J. Potter, A high-Q ferrite-tuned cavity. IEEE Trans. Nucl. Sci. **NS-30**(4), 3460–3462 (1983)
33. K. Kaspar, Design of ferrite-tuned accelerator cavities using perpendicular-biased high-Q ferrites. Technical report, Los Alamos National Laboratory, NM, November 1984. LA-10277-MS
34. G. Schaffer, Improved ferrite biasing scheme for booster RF cavities, in *EPAC* (1992), pp. 1234–1236
35. G. Piefke, *Feldtheorie I, II, III* (Bibliographisches Institut & F. A. Brockhaus AG, Mannheim/Wien/Zürich, Bd. I: 1977, Bd. II: 1973, Bd. III: 1977)
36. T. Weiland et al., *Skriptum zu Grundlagen der Elektrodynamik und Technische Elektrodynamik* (TU Darmstadt, Darmstadt, 2011)
37. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965)
38. I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products*, 6th edn. (Academic, San Diego/San Francisco/New York/Boston/London/Sydney/Tokyo/, 2000)
39. H. Padamsee, J. Knobloch, T. Hays, *RF Superconductivity for Accelerators* (Wiley, New York/Chichester/Weinheim/Brisbane/Singapore/Toronto, 1998)
40. P. Schmüser, Basic principles of RF superconductivity and superconducting cavities, in *11th Workshop on RF Superconductivity*, Lübeck/Travemünde, 8–12 Sep 2003, pp. 180–198
41. H. Wiedemann, *Particle Accelerator Physics I & II*, 2nd edn. (Springer, Berlin/Heidelberg/New York, 2003)
42. D. Li et al., Design and fabrication of an 805 MHz RF cavity with Be windows for a high RF power testing for a Muon cooling experiment, in *PAC* (2001)
43. D. Li et al., RF tests of an 805 MHz Pillbox cavity at Lab G of Fermilab, in *PAC* (2003)

Chapter 5
Advanced Topics

In this chapter, some more-advanced aspects of longitudinal beam manipulations in synchrotrons are discussed.

5.1 Different Phase Space Descriptions

In this section, we discuss alternatives to the longitudinal phase space $(\Delta t, \Delta W)$.

5.1.1 Phase Space (φ, δ)

In Chap. 3, we carefully derived the tracking equations (3.8),

$$\Delta\gamma_n = \Delta\gamma_{n-1} + \frac{Q}{m_0 c_0^2}\Delta V_{n-1},$$

and (3.14),

$$\Delta t_n = \Delta t_{n-1} + \frac{l_\mathrm{R}\eta_{\mathrm{R},n}}{\beta_n \beta_{\mathrm{R},n}^2 c_0}\frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}}.$$

This allowed us to show that the phase space area measured in eVs is invariant. These tracking equations are, of course, not the only possible ones. For example, we may use the approximation

$$\delta := \frac{\Delta p}{p_\mathrm{R}} \approx \frac{1}{\beta_\mathrm{R}^2}\frac{\Delta\gamma}{\gamma_\mathrm{R}} \tag{5.1}$$

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6_5

to convert the first equation into

$$\delta_n = \delta_{n-1} + \frac{Q}{\beta_\mathrm{R}^2 W_\mathrm{R}} \Delta V_{n-1}.$$

For harmonic gap voltages, we get

$$\delta_n = \delta_{n-1} + \frac{Q\hat{V}}{\beta_\mathrm{R}^2 W_\mathrm{R}} \left(\sin\varphi_{\mathrm{RF},n-1} - \sin\varphi_\mathrm{R}\right). \tag{5.2}$$

Here we assumed that $\beta_\mathrm{R}^2\gamma_\mathrm{R}$ does not change significantly from revolution $n-1$ to revolution n. We may also multiply the second equation by

$$\omega_\mathrm{RF} = 2\pi h f_R = 2\pi h \frac{c_0\beta_\mathrm{R}}{l_\mathrm{R}}$$

to get

$$\Delta\varphi_{\mathrm{RF},n} = \Delta\varphi_{\mathrm{RF},n-1} + 2\pi h \frac{\eta_{\mathrm{R},n}}{\beta_n\beta_{\mathrm{R},n}} \frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}}.$$

Here we assumed that the RF frequency does not change significantly from revolution $n-1$ to revolution n. If one also assumes that $\varphi_{\mathrm{R},n}$ does not differ significantly from $\varphi_{\mathrm{R},n-1}$, one gets

$$\varphi_{\mathrm{RF},n} = \varphi_{\mathrm{RF},n-1} + 2\pi h \frac{\eta_{\mathrm{R},n}}{\beta_n\beta_{\mathrm{R},n}} \frac{\Delta\gamma_n}{\gamma_{\mathrm{R},n}},$$

since

$$\varphi_{\mathrm{RF},n} = \varphi_{\mathrm{R},n} + \Delta\varphi_{\mathrm{RF},n}$$

holds. If we furthermore make use of the approximation (5.1), we get

$$\varphi_{\mathrm{RF},n} = \varphi_{\mathrm{RF},n-1} + 2\pi h \eta_\mathrm{R} \delta_n. \tag{5.3}$$

Equations (5.2) and (5.3) may also be found, for example, as Eq. (3.28) in the textbook by Lee [1]. They obviously use the phase space $(\varphi_\mathrm{RF}, \delta)$ instead of our original phase space $(\Delta t, \Delta W)$. If one calculates the Jacobian

$$\frac{\partial(\delta_n, \varphi_{\mathrm{RF},n})}{\partial(\delta_{n-1}, \varphi_{\mathrm{RF},n-1})},$$

one finds that it equals 1. Therefore, the modified tracking equations preserve the phase space area. It should be clear, however, that this is an artifact caused by our

sloppy derivation of the modified tracking equations. We know that the phase space $(\Delta t, \Delta W)$ leads to area invariance on a long-term basis. Due to the approximations we used to derive the tracking equations for the phase space $(\varphi_{\mathrm{RF}}, \delta)$, we cannot be sure that these equations are still exact. Although the approximations are valid with a certain precision in each tracking step, they may lead to large deviations on a long-term basis. This should not astonish the reader, since we already pointed out previously that making the tracking equations more symmetric also destroys area preservation, although this modification is negligible in each step. There is no canonical transformation that converts the phase space coordinates $(\Delta t, \Delta W)$ into the phase space coordinates $(\varphi_{\mathrm{RF}}, \delta)$ or vice versa.

Of course, one may analyze the phase space $(\varphi_{\mathrm{RF}}, \delta)$ in the same way as we did for the phase space $(\Delta t, \Delta W)$. A Hamiltonian may be derived, and the bucket height or the bucket area can be calculated.

As an example, we consider the ratio of the principal axes. According to Eq. (3.28), we have

$$\frac{\Delta \hat{W}}{\Delta \hat{t}} = f_{\mathrm{S},0} \, \frac{2\pi W_{\mathrm{R}} \beta_{\mathrm{R}}^2}{|\eta_{\mathrm{R}}|}.$$

If we use the momentum spread

$$\delta = \frac{\Delta p}{p_{\mathrm{R}}} \approx \frac{1}{\beta_{\mathrm{R}}^2} \frac{\Delta W}{W_{\mathrm{R}}}$$

instead of ΔW and the RF phase deviation

$$\Delta \varphi_{\mathrm{RF}} = \omega_{\mathrm{RF}} \Delta t$$

instead of Δt, we obtain

$$\frac{\hat{\delta}}{\Delta \hat{\varphi}_{\mathrm{RF}}} = f_{\mathrm{S},0} \, \frac{2\pi}{|\eta_{\mathrm{R}}| \, \omega_{\mathrm{RF}}} = \frac{1}{|\eta_{\mathrm{R}}| h} \frac{f_{\mathrm{S},0}}{f_{\mathrm{R}}}.$$

This corresponds to Eq. (3.55) in Lee [1].

Since the phase space coordinates $(\varphi_{\mathrm{RF}}, \delta)$ lead to an area preservation that is not justified from a physical point of view, we will not use this phase space in most parts of this book.

5.1.2 Relation to Phase Space $(\Delta t, \Delta W)$

Now we determine the bucket area for the phase space $(\Delta \varphi_{\mathrm{RF}}, \delta)$. For this purpose, we consider the variables

$$Q = \Delta t, \qquad P = \Delta W,$$
$$q = \Delta\varphi_{\mathrm{RF}}, \qquad p = \delta = \frac{\Delta p}{p_{\mathrm{R}}}.$$

This leads to

$$\frac{q}{Q} = \frac{\Delta\varphi_{\mathrm{RF}}}{\Delta t} = \omega_{\mathrm{RF}} = \frac{2\pi h}{T_{\mathrm{R}}} \tag{5.4}$$

and

$$\frac{p}{P} = \frac{\delta}{\Delta W} = \frac{1}{\beta_{\mathrm{R}}^2 m_0 c_0^2 \gamma_{\mathrm{R}}}. \tag{5.5}$$

For the first of these two equations, one has to take into account that $\Delta\varphi_{\mathrm{RF}}$ is the RF phase. If this varies from $-\pi$ to π, the variable Δt will move through only one bucket with the time span T_{R}/h, not through the whole circumference equivalent to T_{R}.

For deriving the second equation, we made use of the relation

$$\frac{\Delta W}{W_{\mathrm{R}}} = \frac{\Delta\gamma}{\gamma_{\mathrm{R}}} \approx \beta_{\mathrm{R}}^2 \frac{\Delta p}{p_{\mathrm{R}}}.$$

For the bucket area, we have

$$A_{\mathrm{B}}^{\Delta\varphi_{\mathrm{RF}},\delta} = \int\int \mathrm{d}q\ \mathrm{d}p$$

and

$$A_{\mathrm{B}}^{\Delta t,\Delta W} = \int\int \mathrm{d}Q\ \mathrm{d}P,$$

respectively. The transformation law is

$$A_{\mathrm{B}}^{\Delta\varphi_{\mathrm{RF}},\delta} = \int\int \left|\begin{matrix} \frac{\partial q}{\partial Q} & \frac{\partial q}{\partial P} \\ \frac{\partial p}{\partial Q} & \frac{\partial p}{\partial P} \end{matrix}\right| \mathrm{d}Q\ \mathrm{d}P,$$

the Jacobian equals

$$\xi = \left|\begin{matrix} \frac{\partial q}{\partial Q} & \frac{\partial q}{\partial P} \\ \frac{\partial p}{\partial Q} & \frac{\partial p}{\partial P} \end{matrix}\right| = \frac{2\pi h}{T_{\mathrm{R}}\beta_{\mathrm{R}}^2 m_0 c_0^2 \gamma_{\mathrm{R}}} = \frac{2\pi h}{T_{\mathrm{R}}\beta_{\mathrm{R}}^2 W_{\mathrm{R}}},$$

as Eqs. (5.4) and (5.5) show. Please note that this factor ξ changes if an acceleration takes place. The transformation from q, p to Q, P does not correspond to a

canonical transformation. The factor ξ does not depend on the integration variables, so that we may take it out of the integral:

$$A_{\mathrm{B}}^{\Delta\varphi_{\mathrm{RF}},\delta} = \xi \; A_{\mathrm{B}}^{\Delta t,\Delta W}.$$

Therefore, Eq. (3.41),

$$A_{\mathrm{B}}^{\Delta t,\Delta W} = \frac{4\sqrt{2}}{\pi h} T_{\mathrm{R}} \sqrt{\frac{W_{\mathrm{R}}\beta_{\mathrm{R}}^2 |Q| \hat{V}}{\pi h |\eta_{\mathrm{R}}|}} \, \alpha(\varphi_{\mathrm{R}}),$$

leads to

$$A_{\mathrm{B}}^{\Delta\varphi_{\mathrm{RF}},\delta} = 8\sqrt{2} \sqrt{\frac{|Q| \hat{V}}{\pi h |\eta_{\mathrm{R}}| W_{\mathrm{R}}\beta_{\mathrm{R}}^2}} \, \alpha(\varphi_{\mathrm{R}})$$

for the bucket area of a single bucket. By means of Eq. (5.5), we may convert the bucket height in Eq. (3.32),

$$\Delta W_{\mathrm{max,stat}} = \sqrt{\frac{2 \; W_{\mathrm{R}}\beta_{\mathrm{R}}^2 \; |Q| \; \hat{V}}{\pi h \; |\eta_{\mathrm{R}}|}},$$

into the bucket height

$$\delta_{\mathrm{max,stat}} = \Delta W_{\mathrm{max,stat}} \, \frac{1}{W_{\mathrm{R}}\beta_{\mathrm{R}}^2} = \sqrt{\frac{2 \, |Q| \hat{V}}{\pi h |\eta_{\mathrm{R}}| W_{\mathrm{R}}\beta_{\mathrm{R}}^2}}$$

for the momentum spread. For an accelerated bucket, we have

$$\delta_{\mathrm{max}} = \delta_{\mathrm{max,stat}} Y(\varphi_{\mathrm{R}}).$$

5.1.3 *Scale Transformation with Invariant Bucket Area*

If we transform the original coordinate/momentum pair (q, p) into a new one (Q, P), and if Q and P depend on q and p only via constant factors, then the bucket area changes by the factor

$$\xi = \left| \begin{array}{cc} \frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p} \\ \frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \end{array} \right|,$$

as shown above. If we furthermore assume that Q depends only on q, and P only on p, then

$$\xi = \frac{\partial Q}{\partial q}\frac{\partial P}{\partial p}$$

holds. If the bucket area is to remain unchanged under the transformation, the condition $\xi = 1$ is required. Therefore, if one variable p is multiplied by a factor f ($P = fp$), the other variable q must be divided by that factor f ($Q = q/f$) in order to preserve the bucket area.

Let us, for example, begin with the phase space $(\Delta t, \Delta W)$ and multiply the first variable Δt by ω_{RF} in order to obtain $\Delta\varphi_{\mathrm{RF}}$. Then ΔW must be divided by ω_{RF}. Hence, the coordinates $(\Delta\varphi_{\mathrm{RF}}, \Delta W/\omega_{\mathrm{RF}})$ lead to the same bucket area as the coordinates $(\Delta t, \Delta W)$. If instead of the RF phase $\Delta\varphi_{\mathrm{RF}}$, one considers the angle θ of the whole accelerator ring, i.e.,

$$\Delta\theta = \Delta\varphi_{\mathrm{RF}}/h,$$

then the pair $(\Delta\theta, \Delta W/\omega_{\mathrm{R}})$ is obtained as another alternative.

The transition to the new coordinates may be regarded as a canonical transformation. Even though we did not introduce generating functions in the scope of this book, we now use $F_3(p, Q)$. The reader may consult, for example, the book of Goldstein [2] for the definition and the properties of generating functions. We begin with

$$q = \Delta t, \qquad p = \Delta W, \qquad Q = \Delta\varphi_{\mathrm{RF}},$$

and due to

$$q = -\frac{\partial F_3(p, Q)}{\partial p},$$

we obtain

$$\Delta t = -\frac{\partial F_3(\Delta W, \Delta\varphi_{\mathrm{RF}})}{\partial \Delta W}.$$

This relation is obviously satisfied for

$$F_3 = -\Delta W\ \Delta\varphi_{\mathrm{RF}}\ \frac{1}{\omega_{\mathrm{RF}}},$$

since $\Delta\varphi_{\mathrm{RF}} = \omega_{\mathrm{RF}}\Delta t$ holds. This allows us to determine the new generalized momentum variable P:

$$P = -\frac{\partial F_3}{\partial Q} = -\frac{\partial F_3}{\partial \Delta\varphi_{\mathrm{RF}}} = \frac{\Delta W}{\omega_{\mathrm{RF}}}.$$

As a result, we obtain the same pair of variables $(\Delta\varphi_{\mathrm{RF}}, \Delta W/\omega_{\mathrm{RF}})$ as above.

Now we consider the case that the factor used for the scale transformation (ω_{RF}) may vary with time. Just in order to use a different generating function this time, we choose $F_2(q, P, t)$. We set

$$F_2 = q\ P\ \lambda(t).$$

Therefore, we obtain

$$p = \frac{\partial F_2}{\partial q} = P \cdot \lambda(t) \qquad \Rightarrow P = \frac{p}{\lambda(t)},$$

$$Q = \frac{\partial F_2}{\partial P} = q \cdot \lambda(t) \qquad \Rightarrow Q = q \cdot \lambda(t).$$

The new Hamiltonian is

$$K = H + \frac{\partial F_2}{\partial t} = H + q\ P\ \frac{\mathrm{d}\lambda}{\mathrm{d}t} = H + QP\ \frac{1}{\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t}.$$

We see that only for constant λ is the same Hamiltonian $K = H$ obtained.

In Schmutzer [3, volume I, p. 417], it is shown that for all canonical transformations, i.e., time-dependent ones, the Jacobian equals 1 in general. Therefore, the phase space area remains the same even though the Hamiltonian becomes time-dependent.

We finally check that the new function K actually is a Hamiltonian:

$$\begin{aligned}
\frac{\partial K}{\partial Q} &= \frac{\partial H}{\partial q}\ \frac{\partial q}{\partial Q} + \frac{\partial H}{\partial p}\ \frac{\partial p}{\partial Q} + P\frac{\dot\lambda}{\lambda} = \\
&= -\dot p\ \frac{1}{\lambda} + \dot q \cdot 0 + P\frac{\dot\lambda}{\lambda} = \\
&= -(\dot\lambda P + \lambda \dot P)\frac{1}{\lambda} + P\frac{\dot\lambda}{\lambda} = \\
&= -\dot P,
\end{aligned}$$

$$\begin{aligned}\frac{\partial K}{\partial P} &= \frac{\partial H}{\partial q}\,\frac{\partial q}{\partial P} + \frac{\partial H}{\partial p}\,\frac{\partial p}{\partial P} + Q\frac{\dot{\lambda}}{\lambda} = \\ &= -\dot{p}\cdot 0 + \dot{q}\,\lambda + Q\frac{\dot{\lambda}}{\lambda} = \\ &= \frac{\dot{Q}\lambda - Q\dot{\lambda}}{\lambda^2}\,\lambda + Q\frac{\dot{\lambda}}{\lambda} = \\ &= \dot{Q}.\end{aligned}$$

5.2 Special Remarks on Linear ODEs of Second Order

Linear ODEs of second order with variable coefficients occur very often in mathematical physics. In the following, we will therefore discuss some of their properties.

5.2.1 Removing the Attenuation Term

As mentioned in Kamke [4, volume 1, Sect. 16.3], the second term of a linear ODE may be removed by means of a suitable transformation. For the special case

$$a_2(t)\ddot{y} + a_1(t)\,\dot{y} + a_0(t)\,y = 0, \tag{5.6}$$

the ansatz

$$\begin{aligned}y &= uv,\\ \dot{y} &= \dot{u}v + u\dot{v},\\ \ddot{y} &= \ddot{u}v + 2\dot{u}\dot{v} + u\ddot{v},\end{aligned}$$

leads to the ODE

$$\ddot{u}(a_2 v) + \dot{u}(2a_2\dot{v} + a_1 v) + u(a_2\ddot{v} + a_1\dot{v} + a_0 v) = 0$$

$$\Rightarrow \ddot{u} + \dot{u}\left(2\frac{\dot{v}}{v} + \frac{a_1}{a_2}\right) + u\left(\frac{\ddot{v}}{v} + \frac{a_1}{a_2}\frac{\dot{v}}{v} + \frac{a_0}{a_2}\right) = 0.$$

Here we make the second term vanish by setting

$$\frac{\dot{v}}{v} = -\frac{a_1}{2a_2},$$

which leads to

$$\int \frac{\mathrm{d}v}{v} = -\frac{1}{2}\int \frac{a_1}{a_2}\mathrm{d}t$$

$$\Rightarrow \ln|v| = -\frac{1}{2}\int \frac{a_1}{a_2}\mathrm{d}t + \text{const}$$

$$\Rightarrow v = v_0 \exp\left(-\frac{1}{2}\int \frac{a_1}{a_2}\mathrm{d}t\right).$$

This choice leads to the ODE

$$\ddot{u} + u\left(\frac{\ddot{v}}{v} + \frac{a_1}{a_2}\frac{\dot{v}}{v} + \frac{a_0}{a_2}\right) = 0,$$

without attenuation term. Due to

$$\dot{v} = -\frac{1}{2}\frac{a_1}{a_2}v,$$

$$\ddot{v} = -\frac{1}{2}\frac{\dot{a}_1 a_2 - \dot{a}_2 a_1}{a_2^2}v - \frac{1}{2}\frac{a_1}{a_2}\dot{v} = \left(-\frac{1}{2}\frac{\dot{a}_1}{a_2} + \frac{1}{2}\frac{\dot{a}_2}{a_2}\frac{a_1}{a_2} + \frac{1}{4}\frac{a_1^2}{a_2^2}\right)v,$$

one obtains

$$\ddot{u} + u\left(\frac{a_0}{a_2} - \frac{1}{2}\frac{\dot{a}_1}{a_2} + \frac{1}{2}\frac{\dot{a}_2}{a_2}\frac{a_1}{a_2} - \frac{1}{4}\frac{a_1^2}{a_2^2}\right) = 0. \tag{5.7}$$

In principle, the trick that was presented here that made the first-order derivative vanish cannot change the physical behavior of the system under consideration. The attenuation that was obviously present in the original ODE was just (partly) shifted into the function $v(t)$. Furthermore, it is clear that the transformation will usually lead to a very complicated ODE of the type (5.7), which is no easier to solve analytically than the original ODE (5.6).

However, the mathematical trick showed us that it is very useful to analyze ODEs of the type

$$\ddot{u} + K(t)\, u = 0,$$

because it also allows statements about the original ODE (5.6).

5.2.2 Solution by Integration of the Phase

We begin with the homogeneous linear ODE

$$\ddot{u} + \Omega^2(t)u = 0 \tag{5.8}$$

of second order.[1] If Ω did not depend on time, we could use sine, cosine, or exponential functions as an ansatz for the solution. For a time-dependent Ω, it is therefore straightforward to test the ansatz

$$u(t) = u_0(t) \exp\left(j \int \Omega_0(t)\, \mathrm{d}t\right),$$

which leads to

$$\begin{aligned}
\dot{u} &= (\dot{u}_0 + j\,\Omega_0 u_0)e^{j\varphi} \\
\ddot{u} &= (\ddot{u}_0 + j\,\dot{\Omega}_0 u_0 + j\,\Omega_0 \dot{u}_0 + j\,\Omega_0(\dot{u}_0 + j\,\Omega_0 u_0))e^{j\varphi} = \\
&= (\ddot{u}_0 + j\,\dot{\Omega}_0 u_0 + 2j\,\Omega_0 \dot{u}_0 - \Omega_0^2 u_0)e^{j\varphi}.
\end{aligned}$$

Here we used

$$\varphi(t) = \int \Omega_0(t)\, \mathrm{d}t.$$

If we insert this into the ODE, we obtain

$$\ddot{u}_0 + (\Omega^2 - \Omega_0^2)u_0 = 0$$

for the real part and

$$\dot{\Omega}_0 u_0 + 2\Omega_0 \dot{u}_0 = 0$$

for the imaginary part. According to

$$\int \frac{\mathrm{d}u_0}{u_0} = -\frac{1}{2} \int \frac{\mathrm{d}\Omega_0}{\Omega_0},$$

the latter may be separated, so that

$$\ln|u_0| = -\frac{1}{2} \ln|\Omega_0| + \text{const}$$

[1] If instead of $\Omega^2(t)$, there were a periodic function $K(t)$, which does not necessarily have to be positive, one would be faced with the **Hill's differential equation**.

$$\Rightarrow u_0 = \frac{M}{\sqrt{\Omega_0}}$$

is obtained. If $\Omega_0(t)$ is known, then $u_0(t)$ is determined as well, and a solution $u(t)$ of the ODE is found.

We obviously have

$$\dot{u}_0 = -\frac{M}{2}\Omega_0^{-3/2}\dot{\Omega}_0,$$
$$\ddot{u}_0 = \frac{3M}{4}\Omega_0^{-5/2}\dot{\Omega}_0^2 - \frac{M}{2}\Omega_0^{-3/2}\ddot{\Omega}_0,$$

which may be inserted into the ODE for $u_0(t)$:

$$\frac{3}{4}\Omega_0^{-5/2}\dot{\Omega}_0^2 - \frac{1}{2}\Omega_0^{-3/2}\ddot{\Omega}_0 + (\Omega^2 - \Omega_0^2)\Omega_0^{-1/2} = 0$$

$$\Rightarrow \Omega^2 = \Omega_0^2 + \frac{1}{2}\frac{\ddot{\Omega}_0}{\Omega_0} - \frac{3}{4}\frac{\dot{\Omega}_0^2}{\Omega_0^2}. \tag{5.9}$$

In general, it is, of course, very difficult to solve this nonlinear ODE if $\Omega(t)$ is given. In Appendix A.8, however, we use this result to construct a test scenario that can be solved analytically.

5.2.3 *Discussion of a Sample Solution*

Now we consider some concrete numbers for the test scenario defined in Appendix A.8. In the synchrotron SIS18 at GSI, U^{73+} is stored at a kinetic energy of 11.4 MeV/u. According to Eq. (A.67), this leads to

$$a = \frac{-\eta_R}{\beta_R^2\gamma_R} = 38.732.$$

At the time $t = 0$, the total gap voltage is chosen in such a way that a synchrotron frequency of 1640 Hz is obtained:

$$\hat{\Omega}_0 = 10{,}304\,\mathrm{s}^{-1}.$$

We fix

$$k = 10\,\mathrm{s}^{-1}$$

and consider the time span from $t = 0$ to $t = 100\,\text{ms}$ (see Eq. (A.62)). Finally, we analyze an asynchronous particle with an initial time deviation

$$\Delta t_0 = 300\,\text{ns}.$$

For this example, Fig. 5.1 shows a phase space plot. In order to keep the plot simple, only three time intervals with a length of 10 ms each are displayed. It is remarkable that the revolutions are almost closed. Therefore, one may calculate the area approximately that is enclosed by the trajectory. According to the figure, this area obviously does not change significantly.

Figures 5.2 and 5.3 show how the situation changes if k has a larger value. In all cases, $k\,t$ sweeps from 0 to 1, so that Ω_0 decreases to about 30% of its initial value.

For values of k that are too large, the curves are strongly deformed, and we may therefore no longer regard them as closed. This makes it more difficult to define the area that is enclosed in a unique way.

Figures 5.4, 5.5, and 5.6 show that for sufficiently small k, the values of Ω and Ω_0 are almost equal.

If we consider Eq. (5.9),

$$\Omega^2 = \Omega_0^2 + \frac{1}{2}\frac{\ddot{\Omega}_0}{\Omega_0} - \frac{3}{4}\frac{\dot{\Omega}_0^2}{\Omega_0^2},$$

we immediately see that the two requirements

$$\left|\frac{\dot{\Omega}_0}{\Omega_0}\right| \ll \frac{1}{T_0}, \tag{5.10}$$

$$\left|\frac{\ddot{\Omega}_0}{\Omega_0}\right| \ll \frac{1}{T_0^2}, \tag{5.11}$$

lead to

$$\Omega \approx \Omega_0 = \frac{2\pi}{T_0}.$$

Therefore, the behavior that is observed in our example is valid in general.

Now we check under what conditions the curves in phase space are almost closed. Equations (A.69) and (A.70) are abbreviated according to

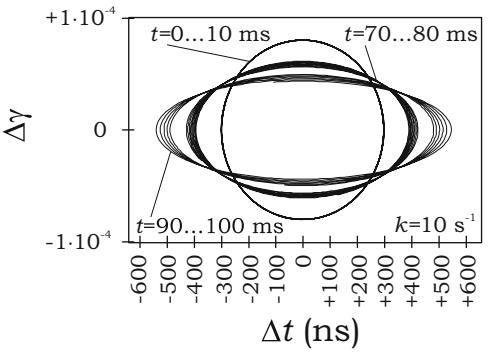

Fig. 5.1 Phase space plot for $k = 10\,\mathrm{s}^{-1}$

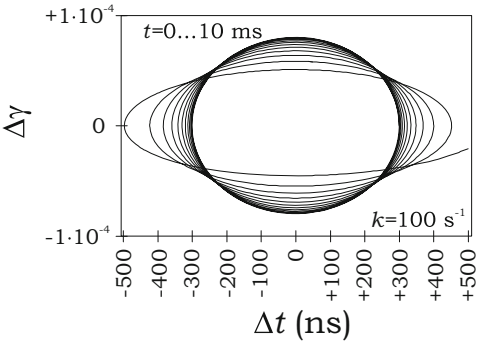

Fig. 5.2 Phase space plot for $k = 100\,\mathrm{s}^{-1}$

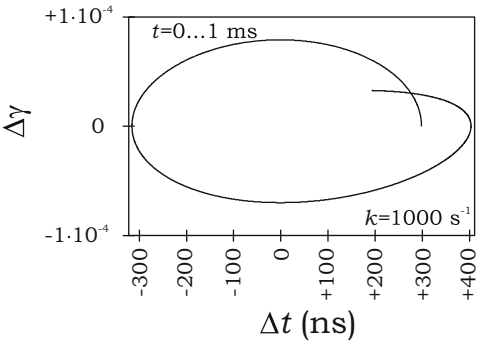

Fig. 5.3 Phase space plot for $k = 1000\,\mathrm{s}^{-1}$

$$\Delta t = \Delta t_0 c^{-1} \cos\varphi$$

and

$$\Delta\gamma = \frac{\Delta t_0}{a}\left[-k\ c^{-2}\ s\ \cos\varphi + \hat{\Omega}_0\ c\ \sin\varphi\right].$$

For the sake of simplicity, we consider one revolution from $\varphi = 2\pi m + \pi/2$ to $\varphi = 2\pi(m+1) + \pi/2$. In this case, at the starting point and at the end point, we

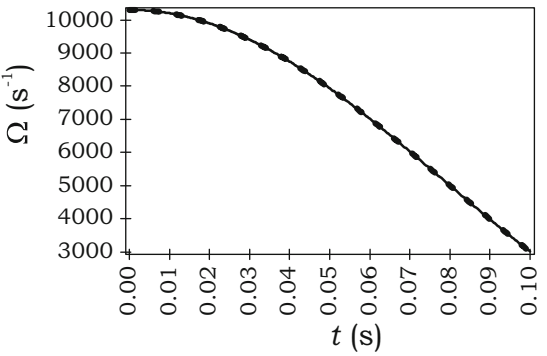

Fig. 5.4 Ω_0 (*solid line*) and Ω (*dotted line*) for $k = 10\,\mathrm{s}^{-1}$

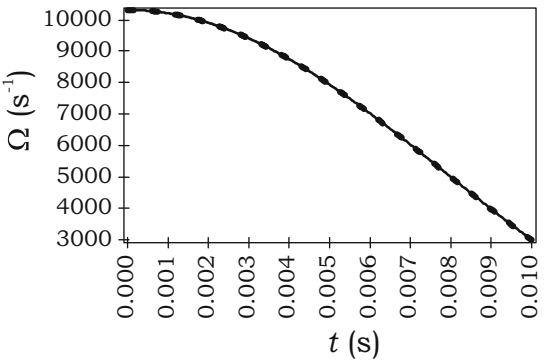

Fig. 5.5 Ω_0 (*solid line*) and Ω (*dotted line*) for $k = 100\,\mathrm{s}^{-1}$

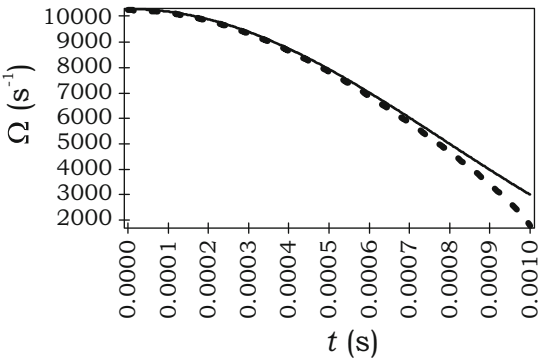

Fig. 5.6 Ω_0 (*solid line*) and Ω (*dotted line*) for $k = 1000\,\mathrm{s}^{-1}$

have

$$\Delta t = 0,$$

and $\Delta\gamma$ changes only according to the cosine function

$$c = \cos(kt).$$

Hence, it is obvious that during one revolution, $\Delta\gamma$ changes by the factor

$$\frac{\cos(k(t+T_0))-\cos(kt)}{\cos(kt)} = \cos(kT_0) - \tan(kt)\ \sin(kT_0) - 1.$$

For our choice

$$\Omega_0(t) = \hat{\Omega}_0 \cos^2(kt),$$

the requirements above lead to

$$(kT_0)^2 \ll 1,$$

so that

$$\frac{\cos(k(t+T_0))-\cos(kt)}{\cos(kt)} \approx -\frac{(kT_0)^2}{2} - kT_0\ \tan(kt)$$

is obtained. Due to

$$kT_0\ \tan(kt) \ll \frac{1}{2},$$

the expression on the right-hand side is much less than 1. Hence, with the required inequalities (5.10) and (5.11), the revolutions are almost closed.

Since the curves are almost closed, we may answer the question how large the enclosed area is. For this purpose, we may use Leibniz's sector formula

$$\boxed{A_0 = \int_0^{T_0} \frac{\dot{y}\,x - \dot{x}\,y}{2}\,\mathrm{d}t,}$$

which gives the area inside a closed curve $(x(t), y(t))$ parameterized by the parameter $t \in [0, T_0]$. In order to define this area in a unique way, we demand that for a given t, the expression kt will be kept constant during the following revolution. Due to $kT_0 \ll 1$, the value of kt will not change significantly. The expression $\varphi \approx \Omega_0 t$, however, increases by 2π during one revolution. For a given revolution, φ will then be the only variable quantity on the right-hand side of the following expressions:

$$\Delta t = \Delta t_0\, c^{-1}\ \cos\varphi,$$
$$\Delta\gamma = \frac{\Delta t_0}{a}\left[-k\ c^{-2}\ s\ \cos\varphi + \hat{\Omega}_0\ c\ \sin\varphi\right].$$

Due to

$$\Delta\dot{t} = -a\ \Delta\gamma,$$
$$\Delta\dot{\gamma} = b\ \Delta t,$$

and

$$b = \frac{\Omega^2}{a},$$

we obtain

$$A_0 = \int_0^{T_0} \frac{\Delta\dot{\gamma}\ \Delta t - \Delta\dot{t}\ \Delta\gamma}{2}\ \mathrm{d}t = \frac{1}{2}\int_0^{T_0} \left(b\ \Delta t^2 + a\Delta\gamma^2\right)\ \mathrm{d}t =$$
$$= \frac{1}{2}\int_0^{T_0} \left[\frac{\Omega^2}{a}\Delta t_0^2\ c^{-2}\ \cos^2\varphi + \frac{\Delta t_0^2}{a}\left(-k\ c^{-2}\ s\ \cos\varphi + \hat{\Omega}_0\ c\ \sin\varphi\right)^2\right]\ \mathrm{d}t =$$
$$= \frac{\Delta t_0^2}{2a}\int_0^{T_0} \Big[\left(\Omega^2\ c^{-2} + k^2\ c^{-4}\ s^2\right)\ \cos^2\varphi - 2k\ c^{-1}\ s\ \hat{\Omega}_0\ \sin\varphi\ \cos\varphi$$
$$+\ \hat{\Omega}_0^2\ c^2\ \sin^2\varphi\Big]\ \mathrm{d}t.$$

With

$$\varphi = \int \Omega_0\ \mathrm{d}t, \qquad \frac{\mathrm{d}\varphi}{\mathrm{d}t} = \Omega_0 = \hat{\Omega}_0\ c^2,$$

it follows that

$$A_0 = \frac{\Delta t_0^2}{2a}\int_0^{2\pi} \left[\left(\frac{\Omega^2}{\hat{\Omega}_0}\ c^{-4} + \frac{k^2}{\hat{\Omega}_0}\ c^{-6}\ s^2\right)\ \cos^2\varphi - 2k\ c^{-3}\ s\ \sin\varphi\ \cos\varphi\right.$$
$$\left.+\ \hat{\Omega}_0\ \sin^2\varphi\right]\ \mathrm{d}\varphi.$$

Now we have

$$\int_0^{2\pi} \sin^2\varphi\ \mathrm{d}\varphi = \int_0^{2\pi} \cos^2\varphi\ \mathrm{d}\varphi = \pi \qquad \text{and} \int_0^{2\pi} \sin\varphi\ \cos\varphi\ \mathrm{d}\varphi = 0,$$

so that

$$A_0 = \frac{\pi\Delta t_0^2}{2a\hat{\Omega}_0}\left(\Omega^2\ c^{-4} + k^2\ c^{-6}\ s^2 + \hat{\Omega}_0^2\right)$$

is the result. We may now substitute Δt_0 and a if we define the oscillation amplitudes

$$\Delta\hat{\gamma} = \frac{\Delta t_0 \hat{\Omega}_0 c}{a}, \qquad \Delta\hat{t} = \Delta t_0 c^{-1} \qquad \Rightarrow \Delta\hat{t}\ \Delta\hat{\gamma} = \frac{\Delta t_0^2 \hat{\Omega}_0}{a}.$$

One should note that the product $\Delta\hat{t}\ \Delta\hat{\gamma}$ is constant, even though the factors are changing with time:

$$A_0 = \frac{\pi}{2}\ \Delta\hat{t}\ \Delta\hat{\gamma}\left(\frac{\Omega^2}{\hat{\Omega}_0^2}\ c^{-4} + \frac{k^2}{\hat{\Omega}_0^2}\ c^{-6}\ s^2 + 1\right).$$

Due to

$$\Omega \approx \Omega_0 = \hat{\Omega}_0 c^2 \qquad \text{and} \qquad T_0 = \frac{2\pi}{\Omega_0} = \frac{2\pi c^{-2}}{\hat{\Omega}_0},$$

this may be written approximately as

$$A_0 \approx \frac{\pi}{2}\ \Delta\hat{t}\ \Delta\hat{\gamma}\left(2 + \left[\frac{kT_0}{2\pi}\tan(kt)\right]^2\right). \tag{5.12}$$

Finally, with the help of

$$kT_0\ \tan(kt) \ll \frac{1}{2},$$

we find that

$$A_0 \approx \pi\ \Delta\hat{t}\ \Delta\hat{\gamma}$$

is almost constant.

We analyze this phenomenon more thoroughly. According to Eq. (A.62),

$$\Omega_0(t) = \hat{\Omega}_0\ \cos^2(kt),$$

the angular frequency $\Omega_0 = \hat{\Omega}_0$ is obtained for $t = 0$. This value will be reduced to a specific final value Ω_{final}. This leads to a certain requirement for kt. By choosing k, we may now select whether this final value is reached in a shorter or a longer time $t = t_{\text{final}}$. Under these conditions, we have

$$A_1 = A_0(t = 0) = \pi\ \Delta\hat{t}\ \Delta\hat{\gamma}$$

and

$$A_2 = A_0(t = t_{\text{final}}) \approx \pi\ \Delta\hat{t}\ \Delta\hat{\gamma}\left(1 + \frac{1}{8}\left[\frac{kT_0}{\pi}\tan(kt_{\text{final}})\right]^2\right).$$

For the same final value Ω_{final}, the tangent function always has the same value, since $k\,t_{\text{final}}$ is thereby determined in a unique way. However, we see that the difference between A_1 and A_2 can be made as small as desired by choosing k. If k is very small, i.e., if the frequency change is slow and the final frequency Ω_{final} is reached for large t_{final}, this difference will be very small. In fact, it can be made as small as one likes by increasing t_{final} accordingly.

5.3 Adiabaticity

The example that was presented in the previous section was a very specific one. Nevertheless, we can use it to make some assumptions about phenomena of a general nature.

We summarize the observations we made based in the example:

- The frequency was chosen in such a way that it does not change significantly during one revolution in phase space.
- Therefore, the curves in phase space were almost closed.
- From a mathematical point of view, we required that the inequalities (5.10),

$$\left|\frac{\dot{\Omega}_0}{\Omega_0}\right| \ll \frac{1}{T_0},$$

and (5.11)

$$\left|\frac{\ddot{\Omega}_0}{\Omega_0}\right| \ll \frac{1}{T_0^2}$$

be valid.
- The area that is enclosed by the orbits remains almost constant (its shape changes, however)—even during long time intervals. The slower the frequency is changed, the smaller is the change in the area.

The observation described in the second bullet point motivates the following definition:

Definition 5.1. Consider a Hamiltonian system with the Hamiltonian

$$H(q, p, \lambda)$$

that depends on a time-dependent control parameter $\lambda(t)$. For constant λ, suppose that the orbits in phase space are closed. Under these conditions, the function

$$J(t) = \frac{1}{2\pi} \int_{A(t)} \mathrm{d}q\ \mathrm{d}p$$

is called an action variable of the system. The domain $A(t)$ is defined in such a way that the boundary curve $\partial A(t)$ is a closed orbit $(q(\tilde{t}), p(\tilde{t}))$ of the system with $\lambda = \lambda(t)$ kept fixed while the time parameter $\tilde{t}$ changes. Furthermore, the point $(q(t), p(t))$ must belong to $\partial A(t)$.

This means that for each time t, the actual position (q, p) in phase space is determined. Based on this position, a specific orbit is considered that would be obtained if λ were constant (even though λ varies a bit in reality). This trick leads to a closed orbit that is determined in a unique way for each time t. Please note that this closed orbit is not an orbit that the system actually traverses, since the real system orbit is not closed. Therefore, we cannot use area preservation in phase space for time-dependent Hamiltonians as an argument for the case described in this definition. The action $J(t)$ may in principle be time-dependent.

Please note that Definition 5.1 is a generalization of the definition of J in Sect. 2.11.7.2, which was applicable only to closed orbits. With Definition 5.1, one may now also consider orbits that are not closed or almost closed.

A quantity for which the phenomenon applies that was described in the last bullet point above is called an (Ehrenfest) adiabatic invariant.

According to Vladimir Arnold (see reprint of the article "Small denominators and problems of stability of motion in classical and celestial mechanics," Russ. Math. Surveys 18:6 85–191(1963) in [5]), one makes the following definition.

Definition 5.2. Consider a Hamiltonian system with one degree of freedom specified by a Hamiltonian

$$H(q, p, \lambda)$$

that depends on a time-dependent control parameter $\lambda(t)$. Suppose that for constant λ, the orbits in phase space are closed and that the control parameter varies according to $\lambda(t) = f(\mu t)$, where $f(x)$ is a smooth function.

A function J is called an adiabatic invariant of this system $H(q, p, \lambda)$ if for every $\kappa > 0$, there exists $\mu_0 > 0$ such that for all μ satisfying $0 < \mu < \mu_0$, the relation

$$|J(t) - J(0)| < \kappa \qquad \text{for all } t \text{ with } 0 < \mu t < 1$$

holds.

This definition becomes transparent if we have a look at our example. Let us assume that kt sweeps from 0 to 1.5. We may then set

$$\mu = \frac{k}{1.5},$$

so that

$$0 < kt < 1.5$$

and

$$0 < \mu t < 1$$

are equivalent. Therefore, the normalization that is performed by the transition from k to μ ensures that the final value is obtained for $\mu t = 1$. Our example showed that for every given deviation κ, a maximum value for k can be specified that corresponds to the maximum value μ_0 for μ in the definition.

If we define $t_{\text{final}} = 1/\mu$ and $t_{\text{final,min}} = 1/\mu_0$, we may rewrite Definition 5.2 in the following form:

Definition 5.3. Consider a Hamiltonian system with one degree of freedom specified by the Hamiltonian

$$H(q, p, \lambda)$$

that depends on a time-dependent control parameter $\lambda(t)$. Suppose that for constant λ, the orbits in phase space are closed and that the control parameter varies according to $\lambda(t) = f(t/t_{\text{final}})$, where $f(x)$ is a smooth function.

A function J is called an adiabatic invariant of this system $H(q, p, \lambda)$ if for every $\kappa > 0$, there exists $t_{\text{final,min}} > 0$ such that for all t_{final} satisfying $t_{\text{final}} > t_{\text{final,min}}$, the relation

$$|J(t) - J(0)| < \kappa \qquad \text{for all } t \text{ with } 0 < t < t_{\text{final}}$$

holds.

This means that the tolerated increase in the adiabatic invariant can be fixed a priori as small as desired. By fixing the function $f(x)$, it is clear how the control parameter will change in principle, but since t_{final} is not fixed, the speed of changing the control parameter—and hence also the time span for this process—is open. If J is actually an adiabatic invariant, it is possible to specify a minimum time that is needed for the process in order to satisfy the requirement concerning the tolerated increase of J.

The ODE in our example is a simple case for a Hamiltonian system with one degree of freedom given by the Hamiltonian

$$H(q, p, \lambda),$$

which depends on a time-dependent parameter $\lambda(t)$. For this type of system, the following theorem holds (cf. [6, 7]):

Theorem 5.4. *The action variable J of a Hamiltonian system with one degree of freedom is an adiabatic invariant, provided that the revolution frequency in phase space is not equal to zero.*

This shows that it is theoretically possible to limit the increase in the longitudinal emittance to an arbitrarily small value by just making the beam manipulation process slow enough. Considering longitudinal beam manipulations at constant reference energy, it becomes clear that the change in the voltage amplitude $\hat{V}$ has to be adiabatic.

Theorem 5.4 does not make any statement about how the slowness of the changes of the parameter may be determined in order for the area deviation to remain below a certain limit.

For this practical purpose, in the case of longitudinal beam manipulations, one usually defines the adiabaticity parameter

$$\epsilon = \frac{1}{\omega_S^2} \left| \frac{\mathrm{d}\omega_S}{\mathrm{d}t} \right| = \frac{T_S}{2\pi} \left| \frac{\dot{\omega}_S}{\omega_S} \right| \tag{5.13}$$

(cf. [8, p. 316]). This definition becomes transparent if one looks at requirement (5.10), which states that $2\pi\epsilon \ll 1$.

5.3.1 Pendulum with Variable Length

A famous example using an adiabatic invariant is a pendulum similar to the one shown in Fig. 2.18 on p. 105. However, it does not have a fixed suspension point where the massless thread is fixed. Instead, the thread passes through a hole in a metal plate at the top. Hence, the length R of the pendulum may be varied by moving the thread up and down through the hole. This type of pendulum is called a Rayleigh pendulum or Lorentz–Einstein pendulum.

If the length R is shortened very slowly by pulling the thread upward, the action variable will be an adiabatic invariant.

Before we consider a variable length R, we keep R fixed, and we analyze the case of small oscillation amplitudes with $\alpha \ll 1$. Due to

$$\cos\ \alpha = \cos^2 \frac{\alpha}{2} - \sin^2 \frac{\alpha}{2} = 1 - 2\ \sin^2 \frac{\alpha}{2},$$

we obtain

$$1 - \cos\ \alpha = 2\ \sin^2 \frac{\alpha}{2} \approx \frac{\alpha^2}{2},$$

so that the potential energy is

$$W_{\mathrm{pot}} = mgx = mgR(1 - \cos\ \alpha) \approx mgR\frac{\alpha^2}{2}.$$

Due to $v = R\dot{\alpha}$, the kinetic energy is

$$W_{\text{kin}} = \frac{1}{2}\, mv^2 = \frac{1}{2}\, mR^2\dot{\alpha}^2.$$

The sum $W_{\text{pot}} + W_{\text{kin}}$ remains constant. Hence, the time derivative is zero:

$$mgR\alpha\dot{\alpha} + mR^2\dot{\alpha}\ddot{\alpha} = 0.$$

Therefore, we obtain the well-known ODE

$$\ddot{\alpha} + \frac{g}{R}\alpha = 0$$

with angular frequency

$$\omega = \sqrt{\frac{g}{R}}.$$

We now use the length coordinate

$$q = R\alpha$$

and the momentum

$$p = mR\dot{\alpha}$$

as the pair of coordinates to define the Hamiltonian

$$H(q, p) = \frac{p^2}{2m} + \frac{m}{2}\frac{g}{R}q^2$$

as the sum $W_{\text{pot}} + W_{\text{kin}}$. We verify that H is actually a Hamiltonian:

$$\frac{\partial H}{\partial p} = \frac{p}{m} = \dot{q},$$

$$\frac{\partial H}{\partial q} = m\frac{g}{R}q = mg\alpha.$$

Due to

$$\dot{p} = mR\ddot{\alpha} = -mR\frac{g}{R}\alpha = -mg\alpha,$$

we actually obtain

$$\frac{\partial H}{\partial q} = -\dot{p}.$$

For small oscillation amplitudes, a solution of the ODE is

$$\begin{aligned}
\alpha &= \hat{\alpha}\ \cos(\omega t),\\
\dot{\alpha} &= -\omega\ \hat{\alpha}\ \sin(\omega t),\\
q &= \hat{q}\ \cos(\omega t),\\
p &= -\hat{p}\ \sin(\omega t),
\end{aligned}$$

with

$$\hat{q} = R\hat{\alpha}, \qquad \hat{p} = mR\omega\ \hat{\alpha},$$

leading to the following value of the Hamiltonian:

$$H = \frac{1}{2}\, mR^2\omega^2\ \hat{\alpha}^2\ \sin^2(\omega t) + \frac{1}{2}\, m\frac{g}{R}R^2\hat{\alpha}^2\ \cos^2(\omega t) = \frac{1}{2}\, mR^2\omega^2\ \hat{\alpha}^2 = \frac{1}{2}\, m\omega^2\ \hat{q}^2.$$

The action may easily be determined as the area of the ellipse divided by 2π:

$$J = \frac{1}{2\pi}\int\int \mathrm{d}q\ \mathrm{d}p = \frac{1}{2\pi}\pi\hat{q}\,\hat{p} = \frac{1}{2}mR^2\omega\hat{\alpha}^2 = \frac{1}{2}\, m\omega\ \hat{q}^2. \tag{5.14}$$

The last two equations show that

$$\omega = \frac{H}{J} = \sqrt{\frac{g}{R}} \tag{5.15}$$

is valid.

Now we allow slow changes of the control parameter R with time so that the Hamiltonian is modified[2] according to

$$H(q, p, R(t)) = \frac{p^2}{2m} + \frac{m}{2}\frac{g}{R(t)}q^2.$$

[2]This generalization should not be taken for granted. If, for example, the Hamiltonian in Eq. (2.130) with the pair of variables $(q, p) = (\alpha, \dot{\alpha})$ is taken instead, one obtains different results. Therefore, we took the physical coordinates and momenta and the physical energy conservation as a basis to derive the Hamiltonian. Of course, one may also keep R as time-variable from the very beginning in order to derive the equations of motion and the Hamiltonian.

We know that J is an adiabatic invariant, i.e., its value will remain almost constant if $R(t)$ is modified slowly. Therefore, according to Eq. (5.15), we obtain

$$H \sim \omega \sim R^{-1/2}.$$

Due to Eq. (5.14),

$$\hat{q} \sim \omega^{-1/2} \sim R^{1/4}$$

holds. This leads to

$$\hat{\alpha} = \frac{\hat{q}}{R} \sim R^{-3/4} \sim \omega^{3/2}.$$

Please note that H is not constant, since energy is exchanged with the pendulum if the thread is moved up or down through the hole. When R is reduced, ω and the amplitude $\hat{\alpha}$ of the angle increase. However, the amplitude $\hat{q} = R\hat{\alpha}$ decreases.

A more detailed analysis of the Rayleigh pendulum with the same results can be found in [9, 10].

5.3.2 *Iso-Adiabatic Ramps*

Consider a coasting beam that is stored in a synchrotron at constant energy. This beam will be captured, i.e., bunched, by means of an adiabatic increase of the voltage amplitude $\hat{V}$. This procedure is called an adiabatic capture process. The voltage amplitude $\hat{V}$ plays the role of the control parameter λ discussed earlier. As we will see, the voltage begins at a small but positive value, so that the synchrotron frequency does not vanish.

We assume that the adiabaticity parameter

$$\alpha_{\text{adiab}} = T_{\text{S}} \frac{\dot{\omega}_{\text{S}}}{\omega_{\text{S}}} = 2\pi \frac{\dot{\omega}_{\text{S}}}{\omega_{\text{S}}^2}$$

remains constant (so-called isoadiabatic ramp). Please note that in the literature, the adiabaticity parameter is sometimes defined without the additional factor 2π, and sometimes the absolute value is used as well, so that the coefficient is always nonnegative (cf. Garoby in [8, Sect. 4.8.1]); see Eq. (5.13).

Due to Eq. (3.25), we know that in the stationary case,

$$\omega_{\text{S},0,\text{stat}} = k\sqrt{\hat{V}} \qquad \text{with} \qquad k = f_{\text{R}} \sqrt{\frac{2\pi |\eta_{\text{R}}\, Q|\, h}{W_{\text{R}} \beta_{\text{R}}^2}} > 0$$

holds, so that one obtains

$$\alpha_{\text{adiab}} = \frac{2\pi}{k^2 \hat{V}} \frac{k}{2\sqrt{\hat{V}}} \dot{\hat{V}} = \frac{\pi}{k} \frac{\dot{\hat{V}}}{\hat{V}^{3/2}}.$$

Separation of this ODE yields

$$\int \alpha_{\text{adiab}} \, \mathrm{d}t = \frac{\pi}{k} \int \hat{V}^{-3/2} \mathrm{d}\hat{V}$$

$$\Rightarrow -2\hat{V}^{-1/2} = \beta_{\text{adiab}} t + c_{\text{adiab}} \qquad \text{with} \quad \beta_{\text{adiab}} = \frac{k\alpha_{\text{adiab}}}{\pi}$$

$$\Rightarrow \hat{V} = \frac{4}{\left(\beta_{\text{adiab}} t + c_{\text{adiab}}\right)^2}, \qquad \beta_{\text{adiab}} t = -\frac{2}{\sqrt{\hat{V}}} - c_{\text{adiab}}. \tag{5.16}$$

This solution is now written depending on the initial value $\hat{V}(t = 0) = \hat{V}_1$ and the final value $\hat{V}(t = T) = \hat{V}_2$:

$$\hat{V}_1 = \frac{4}{c_{\text{adiab}}^2}, \qquad c_{\text{adiab}} = -\frac{2}{\sqrt{\hat{V}_1}},$$

$$\hat{V}_2 = \frac{4}{\left(\beta_{\text{adiab}} T + c_{\text{adiab}}\right)^2}, \qquad \beta_{\text{adiab}} = \frac{1}{T}\left(-\frac{2}{\sqrt{\hat{V}_2}} - c_{\text{adiab}}\right) = \frac{2}{T}\left(\frac{1}{\sqrt{\hat{V}_1}} - \frac{1}{\sqrt{\hat{V}_2}}\right).$$

For increasing synchrotron frequencies ω_{S}, the amplitude $\hat{V}$ will increase, and α_{adiab} is positive, so that β_{adiab} is positive as well. For decreasing synchrotron frequencies ω_{S}, the amplitude $\hat{V}$ will decrease, and α_{adiab} is negative, so that β_{adiab} is negative as well. It follows that

$$\hat{V} = \frac{4}{\left(\frac{t}{T}\left(\frac{2}{\sqrt{\hat{V}_1}} - \frac{2}{\sqrt{\hat{V}_2}}\right) - \frac{2}{\sqrt{\hat{V}_1}}\right)^2}$$

$$\Rightarrow \boxed{\hat{V} = \frac{\hat{V}_1}{\left(\frac{t}{T}\left(\sqrt{\frac{\hat{V}_1}{\hat{V}_2}} - 1\right) + 1\right)^2}.}$$

The total time T must, of course, be chosen sufficiently large that

$$|\alpha_{\text{adiab}}| \ll 1$$

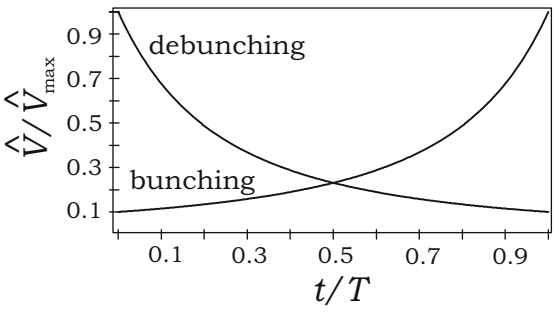

Fig. 5.7 Example of isoadiabatic ramps

holds, where α_{adiab} can be calculated as follows:

$$\alpha_{\text{adiab}} = \frac{\pi}{k}\beta_{\text{adiab}} = \frac{2\pi}{kT}\left(\frac{1}{\sqrt{\hat{V}_1}} - \frac{1}{\sqrt{\hat{V}_2}}\right) = \frac{1}{f_{\text{R}}T}\sqrt{\frac{2\pi\ W_{\text{R}}\beta_{\text{R}}^2}{|\eta_{\text{R}}\ Q|\ h}}\left(\frac{1}{\sqrt{\hat{V}_1}} - \frac{1}{\sqrt{\hat{V}_2}}\right).$$

An example of the amplitude ramp is shown in Fig. 5.7. Let us briefly discuss the bunching case with $\hat{V}_2 > \hat{V}_1$. Since the voltage $\hat{V}_1$ is switched on instantaneously at the very beginning, a certain increase in the longitudinal emittance cannot be avoided. One therefore tries to keep $\hat{V}_1$ as small as technically achievable. For a required value of α_{adiab}, however, the time T will then be large. Therefore, a compromise has to be found.

5.4 Bunch Compression and Unmatched Bunches

In the following, we will discuss some beam manipulations such as bunch compression and barrier bucket operation. These schemes belong to advanced scenarios that are sometimes called "RF gymnastics" [11].

If only the innermost part of the bucket is filled with particles, one may describe the motion in phase space by small values for the momentum spread and the time deviation. In this case, we may linearize the equations of motion. Both the time deviation and the energy deviation will then be harmonic functions. If the first one is proportional to $\cos(\omega_{\text{S}}t)$, the latter one will be proportional to $\sin(\omega_{\text{S}}t)$. Assuming suitable scaling, the orbits in phase space will be circles. If the phase space occupation looks like a horizontal bar at the beginning, this bar will be oriented in an upright way after one quarter of the synchrotron period (similar to the first three pictures in Fig. 3.2 on p. 125). Hence, a long bunch with a comparatively small momentum spread will be transformed into a short bunch with a larger momentum spread. Therefore, this process, which corresponds to a 90° rotation in phase space, is called a **fast bunch compression**.

It is easy to see what happens in the time domain if one analyzes how many particles are located inside a certain time interval. At the beginning (horizontal bar),

one will find a comparatively small number of particles in a certain time interval. This corresponds to a long bunch with small beam current. After one-quarter revolution, we will find many particles in the same time interval, provided that this time interval is close enough to the bunch center (or we will find zero particles outside). This corresponds to a short bunch with large beam current.

If we observe this process depending on time, we will identify an amplitude and length modulation of the bunch. In reality, the bucket will also be filled in the vicinity of the separatrix, so that nonlinearities will be present. The angular velocity of the particles in phase space will no longer be constant, but it will decrease by an amount that increases with the distance of the particles from the bucket center. A bar-shaped bunch that is oriented horizontally at the beginning will therefore be deformed in an S-shaped way after one-quarter of a synchrotron period.

We now analyze the simple case that the bunch is located in the linear region of the bucket and that it fills an elliptical area. The ellipse will be matched to the bucket so that the ratio of the principal axes is given by Eq. (3.28),

$$\frac{\Delta\hat{W}_1}{\Delta\hat{t}_1} = f_{\mathrm{S},1}\,\frac{2\pi W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|}.$$

The index 1 denotes the situation before the voltage is increased at $t = 0$. Therefore, the total length is $2\Delta\hat{t}_1$. We now consider a particle at $(\Delta t, \Delta W) = (0, -\Delta\hat{W}_1)$. Due to the voltage increase, it will move to the point $(\Delta t, \Delta W) = (\Delta\hat{t}_2, 0)$ after one-quarter of the synchrotron period. The new ratio of the principal axes is given by

$$\frac{\Delta\hat{W}_1}{\Delta\hat{t}_2} = f_{\mathrm{S},2}\,\frac{2\pi W_\mathrm{R}\beta_\mathrm{R}^2}{|\eta_\mathrm{R}|},$$

since the synchrotron frequency has changed (in the numerator, we replaced $\Delta\hat{W}_2$ by $\Delta\hat{W}_1$, since this corresponds to the starting point of the trajectory). We have to take into account that the energy W_R of the reference particle does not change when the voltage increases.

We calculate the quotient of the last two equations:

$$\frac{\Delta\hat{t}_2}{\Delta\hat{t}_1} = \frac{f_{\mathrm{S},1}}{f_{\mathrm{S},2}}.$$

According to Eq. (3.25), the synchrotron frequency is proportional to the square root of the voltage amplitude, so that

$$\frac{\Delta\hat{t}_2}{\Delta\hat{t}_1} = \sqrt{\frac{\hat{V}_1}{\hat{V}_2}}$$

is obtained. Hence, we have found an approximation for the bunch length reduction as a result of the voltage increase. The above-mentioned S-shaped deformation is, of course, not taken into account.

As described before, the bunch compression is intentionally based on the phenomenon that the bunch contour does not fit to the trajectories. If the beam is extracted in time, the short bunch that is generated may be used for an experiment.

The same phenomenon is also observed in a weaker form if the bunch is unmatched by accident (**unmatched bunch**). In this case, the particles will smear over the whole phase space area that is determined by the particles that correspond to the outermost trajectory. As a result, this outermost trajectory will be filled with particles in a homogeneous way after several revolutions in phase space. As mentioned in Sect. 3.3, this effect is called filamentation and phase space dilution. The particles finally cover more phase space area than in the beginning.

This observation seems to contradict Liouville's theorem (here we are even confronted with the case that the Hamiltonian does not explicitly depend on time, so that Liouville's theorem is valid in its simplest form). Especially during the first revolution, we may easily see, however, that the S-shaped area filled by the particles has, in fact, the same area as the initial ellipse. Even though this observation is more complicated to verify in the following revolutions, one will always come to the conclusion that area preservation is still valid. Hence, Liouville's theorem is still satisfied.

After some revolutions, however, a lot of "air" will be enclosed between the particles, so that the impression of a completely filled phase space area is induced. Since the number of particles is large but still finite, it is difficult to determine the border of the area that is filled with particles. Therefore, the contradiction is resolved: due to the finite number of particles, the "air" that is curled up leads to an effective increase in the phase space area that cannot be withdrawn. The filamentation process thus leads to an emittance increase.

Please note that Liouville's theorem was derived for a continuous distribution of particles in phase space. When the spiral arms have become longer and longer so that they contain only a few individual particles, this assumption of a continuous distribution is no longer justified, and Liouville's theorem is no longer applicable. In other words, the seeming contradiction may be resolved by the fact that on the one hand, the bunch is described as a continuum, and on the other hand, as a cloud of discrete particles.

The increase in the longitudinal emittance that is caused by unmatched bunches is of course undesirable. The larger emittance usually leads to a larger bucket area that is required to keep all particles bunched. According to Eq. (3.38), this also leads to larger RF voltage (and hence RF power installation) requirements. This is clearly a negative effect of the filamentation process. There are, however, also positive aspects of the momentum spread of a bunch that may also be explained based on our unmatched bunch example. The unmatched bunch as a whole obviously performs coherent oscillations, which may primarily be measured as amplitude oscillations of the beam current. After filamentation has taken effect, these oscillations will have stopped. Therefore, the filamentation process may also be regarded as a damping effect. The momentum spread of the bunch suppresses a collective motion of the particles. This effect is called **Landau damping**, and it may prevent beam

instabilities from growing, provided that Landau damping is stronger than the growth of the specific instability (cf. [12, 13, Sect. 9.5]).

5.5 Dual-Harmonic Operation and Barrier Buckets

According to Sect. 2.1.3 (see Eq. (2.9)) and Sect. 2.1.6, for a strongly bunched beam, the beam current amplitude at the fundamental harmonic is twice as large as the DC component, i.e., the average beam current.

In general, the interaction of the beam with the environment or with itself may lead to instabilities. Some of these instabilities become stronger as the ratio of maximum beam current to average beam current increases. Hence, the bunching factor

$$B_\mathrm{f} = \frac{\bar{I}_\mathrm{beam}}{I_\mathrm{beam,max}}$$

defined in Eq. (2.37) should be as large as possible. Please note that the wording "bunching factor" is misleading because this factor decreases when the bunching gets stronger.

Increasing the bunching factor means to fill a larger fraction of the synchrotron circumference l_R with bunches and/or to distribute the particles more homogeneously in the longitudinal direction. One method to achieve this is to add a second RF voltage with a higher frequency to the normal accelerating voltage (**dual-harmonic operation**). For ions with positive charge below transition, the bunches are located on the rising slope of the RF voltage. If the higher-harmonic RF that is added has a negative slope at this point, the phase focusing effect will be reduced locally. Hence, the bunches will show a flat profile.

A different method that also increases the bunching factor is **barrier bucket** operation. Let us assume that h bunches are present in the synchrotron ring and that they are captured as usual by a sinusoidal RF voltage. If some of these sine waves are now omitted, the adjacent bunches will now merge across the missing sine slope. Hence, a longer bunch with a sausage-like shape is created. In other words, two single-sine pulses may be used to keep the beam bunched between them. It is even possible to shift the phase of one of these sine pulses with respect to the revolution frequency. This allows one to adiabatically squeeze the bunch together, thereby increasing its momentum spread. One speaks of **moving barriers** in this case.

5.5.1 Barrier Bucket Signal Generation

Let us assume that a standard cavity system with quality factor Q_p (loaded Q) is used to generate single-sine pulses for barrier bucket operation. The cavity system

can be described by the standard equivalent circuit shown in Fig. 4.5 on p. 181, so that the transfer function is given by Eq. (4.26)

$$Z_{\text{tot}}(s) = \frac{R_\text{p} \frac{\omega_\text{res}}{Q_\text{p}} s}{s \frac{\omega_\text{res}}{Q_\text{p}} + s^2 + \omega_\text{res}^2}$$

$$\Rightarrow Y_{\text{tot}}(s) = \frac{1}{Z_{\text{tot}}(s)} = s \frac{Q_\text{p}}{R_\text{p} \omega_\text{res}} + \frac{1}{R_\text{p}} + \frac{\omega_\text{res} Q_\text{p}}{R_\text{p}} \frac{1}{s}.$$

We want to generate a single-sine pulse

$$\begin{aligned} V_{\text{gap}}(t) &= \hat{V}_{\text{gap}} \ \sin(\omega_1 t) \left[\Theta(t) - \Theta(t - T_1)\right] = \\ &= \hat{V}_{\text{gap}} \ \sin(\omega_1 t) \ \Theta(t) - \hat{V}_{\text{gap}} \ \sin(\omega_1 (t - T_1)) \ \Theta(t - T_1) \end{aligned}$$

with $\omega_1 = 2\pi / T_1$. According to Table A.4 (p. 418) and to the time shifting property of the Laplace transform (Sect. 2.2), this corresponds to

$$V_{\text{gap}}(t) \quad \circ\!\!-\!\!\bullet \quad V_{\text{gap}}(s) = \hat{V}_{\text{gap}} \frac{\omega_1}{s^2 + \omega_1^2} \left(1 - e^{-T_1 s}\right).$$

Now we want to determine the required generator current for negligible beam current. In the Laplace domain, we get

$$I_{\text{gen}}(s) = Y_{\text{tot}}(s) \ V_{\text{gap}}(s) = \left(1 - e^{-T_1 s}\right) \ I_x(s) \tag{5.17}$$

with

$$I_x(s) = \frac{\hat{V}_{\text{gap}}}{R_\text{p}} \left[\omega_1 \frac{Q_\text{p}}{\omega_\text{res}} \frac{s}{s^2 + \omega_1^2} + \frac{\omega_1}{s^2 + \omega_1^2} + \omega_\text{res} Q_\text{p} \omega_1 \frac{1}{s(s^2 + \omega_1^2)} \right].$$

Now we need a transformation back to the time domain. The first two expressions in the square brackets are directly available from Table A.4. For the last term, a partial fraction decomposition leads to

$$\frac{1}{s(s^2 + \omega_1^2)} = \frac{1}{\omega_1^2} \left(\frac{1}{s} - \frac{s}{s^2 + \omega_1^2} \right).$$

Now the transformation back to the time domain can be accomplished with the help of Table A.4:

$$I_x(t) = \Theta(t) \ \frac{\hat{V}_{\text{gap}}}{R_\text{p}} \left[\left(\omega_1 \frac{Q_\text{p}}{\omega_\text{res}} - \omega_\text{res} \frac{Q_\text{p}}{\omega_1} \right) \ \cos(\omega_1 t) + \sin(\omega_1 t) + \omega_\text{res} \frac{Q_\text{p}}{\omega_1} \right].$$

The required generator current is obtained using the time shifting property based on Eq. (5.17):

$$I_{\mathrm{gen}}(t) = I_x(t) - I_x(t - T_1).$$

It is obvious that the solution can be simplified significantly if the cavity is tuned to the angular frequency ω_1 that corresponds to the duration T_1 of the single-sine pulse, i.e., if $\omega_{\mathrm{res}} = \omega_1$:

$$I_x(t) = \Theta(t)\,\frac{\hat{V}_{\mathrm{gap}}}{R_{\mathrm{p}}}\,\left[Q_{\mathrm{p}} + \sin(\omega_1 t)\right]$$

$$\Rightarrow I_{\mathrm{gen}}(t) = \Theta(t)\,\frac{\hat{V}_{\mathrm{gap}}}{R_{\mathrm{p}}}\,\left[Q_{\mathrm{p}} + \sin(\omega_1 t)\right] - \Theta(t-T_1)\,\frac{\hat{V}_{\mathrm{gap}}}{R_{\mathrm{p}}}\,\left[Q_{\mathrm{p}} + \sin(\omega_1 (t - T_1))\right]$$

$$\Rightarrow I_{\mathrm{gen}}(t) = \frac{\hat{V}_{\mathrm{gap}}}{R_{\mathrm{p}}}\,\left[Q_{\mathrm{p}} + \sin(\omega_1 t)\right]\left[\Theta(t) - \Theta(t - T_1)\right]. \tag{5.18}$$

We see that even though the cavity is tuned to the frequency of the single-sine pulse, a DC offset pulse is needed that is Q_{p} times higher than the peak value of the sinusoidal current component. This is not very efficient if a **narrowband cavity** is used. Hence, the quality factor (loaded Q) of the barrier bucket cavity system should be kept as low as possible (**broadband cavity** system). Nevertheless, it is possible to produce single-sine voltage pulses by means of the superposition of a DC current pulse and a sinusoidal current pulse even if Q_{p} is on the order of ten [14].

Of course, in reality, there are always deviations from the ideal behavior. Therefore, it may be necessary to correct the theoretical generator current given by Eq. (5.18) slightly in order to avoid microbunching effects. One way to optimize the shape of the gap voltage is to perform a Fourier analysis of the measured gap voltage, which makes it possible to calculate a predistorted control signal [15].

5.5.2 *Phase and Amplitude Relations for Dual-Harmonic Operation*

As in most cases, we assume ions with positive charge below transition. Let us consider the case that a higher-harmonic component $k\omega_{\mathrm{RF}}$ is added to the fundamental harmonic ω_{RF}. This leads to the total voltage

$$V(t) = \hat{V}_1\ \sin(\omega_{\mathrm{RF}} t + \varphi_1) + \hat{V}_2\ \sin(k\omega_{\mathrm{RF}} t + \varphi_2).$$

We now require that a saddle point be created at the location of the bunch and that the slope be positive to the left and to the right of this saddle point (as it was before

the second harmonic component was added). For the bunch center at $t = 0$, we therefore require

$$V(t=0) = V_0, \qquad V'(t=0) = 0, \qquad V''(t=0) = 0, \qquad \text{and} \qquad V'''(t=0) > 0.$$

Due to

$$\begin{aligned} V'(t) &= \omega_{\mathrm{RF}}\ \hat{V}_1\ \cos(\omega_{\mathrm{RF}} t + \varphi_1) + k\omega_{\mathrm{RF}}\ \hat{V}_2\ \cos(k\omega_{\mathrm{RF}} t + \varphi_2), \\ V''(t) &= -\omega_{\mathrm{RF}}^2\ \hat{V}_1\ \sin(\omega_{\mathrm{RF}} t + \varphi_1) - k^2\omega_{\mathrm{RF}}^2\ \hat{V}_2\ \sin(k\omega_{\mathrm{RF}} t + \varphi_2), \\ V'''(t) &= -\omega_{\mathrm{RF}}^3\ \hat{V}_1\ \cos(\omega_{\mathrm{RF}} t + \varphi_1) - k^3\omega_{\mathrm{RF}}^3\ \hat{V}_2\ \cos(k\omega_{\mathrm{RF}} t + \varphi_2), \end{aligned}$$

this leads to:

$$V_0 = \hat{V}_1\ \sin\varphi_1 + \hat{V}_2\ \sin\varphi_2, \tag{5.19}$$

$$\hat{V}_1\ \cos\varphi_1 = -k\ \hat{V}_2\ \cos\varphi_2, \tag{5.20}$$

$$\hat{V}_1\ \sin\varphi_1 = -k^2\ \hat{V}_2\ \sin\varphi_2, \tag{5.21}$$

$$-\hat{V}_1\ \cos\varphi_1 - k^3\ \hat{V}_2\ \cos\varphi_2 > 0. \tag{5.22}$$

The last equation shows that the substitution

$$\varphi_2' = \varphi_2 - \pi$$

is suitable, since we have

$$\cos\varphi_2 = -\cos\varphi_2'$$

and

$$\sin\varphi_2 = -\sin\varphi_2'.$$

Equations (5.20) and (5.21) then lead to

$$\tan\varphi_1 = k\ \tan\varphi_2'. \tag{5.23}$$

Equation (5.20) provides

$$\frac{\hat{V}_2}{\hat{V}_1} = \frac{\cos\varphi_1}{k\ \cos\varphi_2'}. \tag{5.24}$$

If we insert Eq. (5.21) into Eq. (5.19), we obtain

$$V_0 = \hat{V}_1\ \sin\varphi_1 \left(1 - \frac{1}{k^2}\right),$$

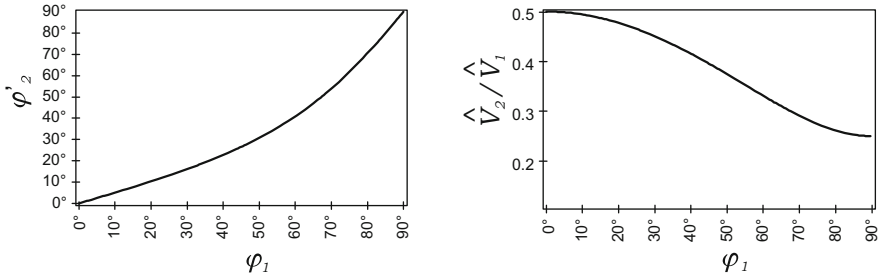

Fig. 5.8 Phase and amplitude relations between first and second harmonics for $k = 2$

or

$$V_0 = \hat{V}_2 \ \sin\varphi_2' \left(k^2 - 1\right).$$

From $k > 1$ and $V_0 \geq 0$, it follows that

$$0 \leq \varphi_1, \varphi_2' \leq \pi.$$

Phases outside the interval $]-\pi, +\pi]$ do not lead to signal forms that differ from those within the interval and are therefore excluded.

Inserting Eq. (5.20) into the inequality (5.22) leads to

$$\hat{V}_1 \ \cos\varphi_1 \left(k^2 - 1\right) > 0$$

and

$$\hat{V}_2 \ \cos\varphi_2' \left(k^3 - k\right) > 0.$$

This reduces the selection to

$$0 \leq \varphi_1, \varphi_2' < \frac{\pi}{2}.$$

The phase relation given by Eq. (5.23) and the amplitude relation given by Eq. (5.24) are both displayed in Fig. 5.8 for the simplest case $k = 2$.

The signal forms $V(t)$ that are obtained for different values of φ_1 are shown in Fig. 5.9. It is obvious that phases $\varphi_1 > 45°$ are usually not of interest, because the corresponding bucket[3] becomes too small.

[3]The shape of a dual-harmonic bucket differs from that of a single-harmonic bucket discussed in Sect. 3.11.

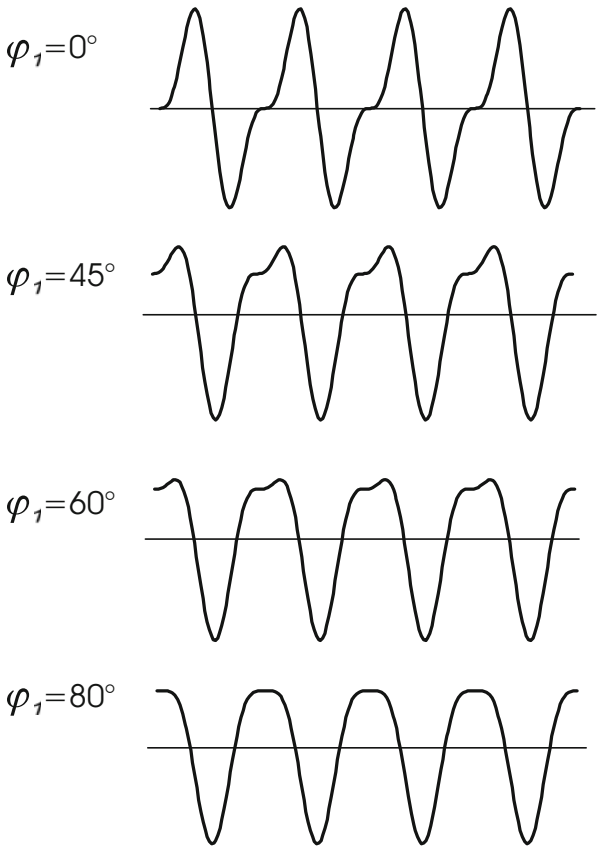

Fig. 5.9 Dual-harmonic signal forms $V(t)$ for $k = 2$

5.6 Bunch Description by Means of Moments

We will now analyze the influence of gap voltage modulations. The results presented in this section and in the next one are based on the article [16] and were extended in [17].

Based on the unmodulated harmonic gap voltage

$$V(t) = \hat{V}_0(t)\ \sin(\varphi_{\mathrm{RF}}(t))$$

with

$$\varphi_{\mathrm{RF}}(t) = \int \omega_{\mathrm{RF}}(t)\ \mathrm{d}t,$$

we now introduce a phase modulation $\Delta\varphi_{\mathrm{gap}}(t)$ and an amplitude modulation $\epsilon(t)$ such that

$$V(t) = \hat{V}_0(t)(1+\epsilon(t))\ \sin(\varphi_{\mathrm{RF}}(t) - \Delta\varphi_{\mathrm{gap}}(t)) \tag{5.25}$$

is obtained.

The reference particle is defined in such a way that it arrives at the accelerating gap when $\varphi_{\mathrm{RF}}(t) = \varphi_{\mathrm{R}}(t) + 2\pi k$ is satisfied. Here, the integer k is the bunch repetition number. For an asynchronous particle, the arrival time is defined by $\varphi_{\mathrm{RF}}(t) = \varphi_{\mathrm{R}}(t) + 2\pi k + \Delta\varphi_{\mathrm{RF}}(t)$. The magnetic field B in the bending dipoles and the quantities $\hat{V}_0$, ω_{RF}, and φ_{R} are chosen in such a way that the reference particle follows the reference path. All these quantities vary slowly with time in comparison with the synchrotron oscillation. The modulation functions $\epsilon(t)$ and $\Delta\varphi_{\mathrm{gap}}(t)$, however, may vary faster.

According to Eqs. (3.18) and (3.19), the nonlinear differential equations are

$$\Delta\dot{\varphi}_{\mathrm{RF}} = \frac{2\pi h\eta_{\mathrm{R}}}{T_{\mathrm{R}}\beta_{\mathrm{R}}^2 W_{\mathrm{R}}}\Delta W, \tag{5.26}$$

$$\Delta\dot{W} = \frac{Q\hat{V}_0}{T_{\mathrm{R}}}[(1+\epsilon)\cdot\sin(\varphi_{\mathrm{R}} + \Delta\varphi_{\mathrm{RF}} - \Delta\varphi_{\mathrm{gap}}) - \sin\varphi_{\mathrm{R}}]. \tag{5.27}$$

For small values of $|\Delta\varphi_{\mathrm{RF}} - \Delta\varphi_{\mathrm{gap}}| \ll 1$, we have

$$\Delta\dot{W} \approx \frac{Q\hat{V}_0\ \cos\varphi_{\mathrm{R}}}{T_{\mathrm{R}}}(1+\epsilon)\left(\Delta\varphi_{\mathrm{RF}} - \Delta\varphi_{\mathrm{gap}} + \frac{\epsilon\ \tan\varphi_{\mathrm{R}}}{1+\epsilon}\right).$$

Defining

$$\Delta\tilde{\varphi}_{\mathrm{gap}} = \Delta\varphi_{\mathrm{gap}} - \frac{\epsilon}{1+\epsilon}\ \tan\varphi_{\mathrm{R}} \tag{5.28}$$

leads to

$$\Delta\dot{W} \approx \frac{Q\hat{V}_0\ \cos\varphi_{\mathrm{R}}}{T_{\mathrm{R}}}(1+\epsilon)(\Delta\varphi_{\mathrm{RF}} - \Delta\tilde{\varphi}_{\mathrm{gap}}). \tag{5.29}$$

By a combination of Eqs. (5.26) and (5.29), we get

$$\Delta\ddot{\varphi}_{\mathrm{RF}} = \frac{2\pi h\eta_{\mathrm{R}} Q\hat{V}_0\ \cos\varphi_{\mathrm{R}}}{T_{\mathrm{R}}^2\beta_{\mathrm{R}}^2 W_{\mathrm{R}}}(1+\epsilon)(\Delta\varphi_{\mathrm{RF}} - \Delta\tilde{\varphi}_{\mathrm{gap}}).$$

According to Eq. (3.25), the synchrotron frequency is defined by

$$\omega_{\mathrm{S}} = \omega_{\mathrm{S},0} = 2\pi f_{\mathrm{S},0} = \sqrt{\frac{2\pi h \hat{V}_0(-\eta_{\mathrm{R}}\ Q\ \cos\varphi_{\mathrm{R}})}{T_{\mathrm{R}}^2 \beta_{\mathrm{R}}^2 W_{\mathrm{R}}}},$$

which yields

$$\Delta\ddot{\varphi}_{\mathrm{RF}} + \omega_{\mathrm{S}}^2(1+\epsilon)\Delta\varphi_{\mathrm{RF}} = \omega_{\mathrm{S}}^2(1+\epsilon)\Delta\tilde{\varphi}_{\mathrm{gap}}. \tag{5.30}$$

Using the new variables

$$x = \Delta\varphi_{\mathrm{RF}}, \qquad y = C\Delta\dot{\varphi}_{\mathrm{RF}},$$

we obtain

$$\dot{x} = \frac{1}{C}\ y,$$
$$\dot{y} = -C\omega_{\mathrm{S}}^2(1+\epsilon)(x - \Delta\tilde{\varphi}_{\mathrm{gap}}).$$

Here the factor C was introduced because we want the trajectories to be circles if no excitations are present ($\epsilon = 0$ and $\Delta\tilde{\varphi}_{\mathrm{gap}} = 0$). Based on this requirement, we can easily determine C:

$$x = \cos(\omega_{\mathrm{S}} t) \qquad \Rightarrow y = C\dot{x} = -C\omega_{\mathrm{S}}\ \sin(\omega_{\mathrm{S}} t) \qquad \Rightarrow C = -\frac{1}{\omega_{\mathrm{S}}}.$$

Thus, we obtain (note that C and ω_{S} vary slowly, and therefore we neglect the time derivative)

$$\dot{x} = -\omega_{\mathrm{S}}\ y, \tag{5.31}$$
$$\dot{y} = \omega_{\mathrm{S}}(1+\epsilon)(x - \Delta\tilde{\varphi}_{\mathrm{gap}}). \tag{5.32}$$

5.6.1 Phase Oscillations

Whereas Eqs. (5.31) and (5.32) are valid for individual particles, we now consider bunches with N_{b} particles. The mean values are defined as

$$\bar{x} = \frac{1}{N_{\mathrm{b}}}\sum_{k=1}^{N_{\mathrm{b}}} x_k, \qquad \bar{y} = \frac{1}{N_{\mathrm{b}}}\sum_{k=1}^{N_{\mathrm{b}}} y_k.$$

This leads to

$$\dot{\bar{x}} = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} \dot{x}_k = -\omega_\mathrm{S} \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} y_k = -\omega_\mathrm{S} \bar{y} \tag{5.33}$$

and

$$\dot{\bar{y}} = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} \dot{y}_k = \omega_\mathrm{S}(1+\epsilon) \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (x_k - \Delta\tilde{\varphi}_\mathrm{gap})$$

$$\Rightarrow \dot{\bar{y}} = \omega_\mathrm{S}(1+\epsilon)(\bar{x} - \Delta\tilde{\varphi}_\mathrm{gap}). \tag{5.34}$$

If we combine the results (5.33) and (5.34) for slowly varying ω_S, we obtain

$$\ddot{\bar{x}} = -\omega_\mathrm{S} \dot{\bar{y}} = -\omega_\mathrm{S}^2 (1+\epsilon)(\bar{x} - \Delta\tilde{\varphi}_\mathrm{gap}). \tag{5.35}$$

We may interpret $\bar{x}$ as the bunch center. Hence, we see that the equation for the bunch center has the same form as Eq. (5.30) for the individual particles. This means that the whole bunch may oscillate with the synchrotron frequency. One should note, however, that we allowed only small oscillation amplitudes to get to this result. Such an oscillation of a whole bunch is called a **coherent dipole oscillation**. It may, for instance, be generated by placing a matched bunch off-center into the bucket. Since all particles rotate in phase space with the synchrotron frequency, this will, in this case, also be true for the whole bunch. However, it is clear that for a realistic bunch size, such an oscillation will soon lead to filamentation. Hence, our linearization will describe only the coherent dipole oscillation in the very beginning and only for small oscillation amplitudes.

5.6.2 Amplitude Oscillations

We define the following quantities:

$$a_x = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} x_k^2, \qquad a_y = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} y_k^2, \qquad \xi = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} x_k y_k, \tag{5.36}$$

$$v_x = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (x_k - \bar{x})^2 = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (x_k^2 - 2x_k\bar{x} + \bar{x}^2) = a_x - \bar{x}^2,$$

$$v_y = a_y - \bar{y}^2.$$

Please note that v_x corresponds to the variance of the quantities x_k if a division by $N_\mathrm{b} - 1$ is used instead of division by N_b. Since we are interested only in large (particle) numbers N_b, this difference is negligible.

The quantity $\sqrt{v_x}$ represents the (rms) bunch length, whereas $\sqrt{v_y}$ represents the (rms) height of the bunch in the phase space (x, y). We get

$$\dot{a}_x = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} 2 x_k \dot{x}_k = -2\omega_\mathrm{S} \xi,$$

$$\dot{a}_y = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} 2 y_k \dot{y}_k = 2\omega_\mathrm{S}(1+\epsilon) \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} y_k (x_k - \Delta\tilde{\varphi}_\mathrm{gap})$$

$$\Rightarrow \dot{a}_y = 2\omega_\mathrm{S}(1+\epsilon)(\xi - \bar{y}\Delta\tilde{\varphi}_\mathrm{gap})$$

$$\dot{\xi} = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (\dot{x}_k y_k + x_k \dot{y}_k) =$$

$$= -\omega_\mathrm{S} a_y + \omega_\mathrm{S}(1+\epsilon)(a_x - \bar{x}\Delta\tilde{\varphi}_\mathrm{gap}),$$

$$\dot{v}_x = \dot{a}_x - 2\bar{x}\dot{\bar{x}} = -2\omega_\mathrm{S}\xi + 2\omega_\mathrm{S}\bar{x}\bar{y} = -2\omega_\mathrm{S}\alpha. \tag{5.37}$$

Here we defined $\alpha = \xi - \bar{x}\bar{y}$ in order to have the same form in the expressions for a_x and for v_x. We obtain

$$\dot{v}_y = \dot{a}_y - 2\bar{y}\dot{\bar{y}} = 2\omega_\mathrm{S}(1+\epsilon)(\xi - \bar{y}\Delta\tilde{\varphi}_\mathrm{gap}) -$$

$$- 2\omega_\mathrm{S}(1+\epsilon)\bar{y}(\bar{x} - \Delta\tilde{\varphi}_\mathrm{gap})$$

$$\Rightarrow \dot{v}_y = 2\omega_\mathrm{S}(1+\epsilon)\alpha, \tag{5.38}$$

$$\dot{\alpha} = \dot{\xi} - \dot{\bar{x}}\bar{y} - \bar{x}\dot{\bar{y}} =$$

$$= -\omega_\mathrm{S} a_y + \omega_\mathrm{S}(1+\epsilon)(a_x - \bar{x}\Delta\tilde{\varphi}_\mathrm{gap}) +$$

$$+ \omega_\mathrm{S}\bar{y}^2 - \omega_\mathrm{S}(1+\epsilon)\bar{x}(\bar{x} - \Delta\tilde{\varphi}_\mathrm{gap})$$

$$\Rightarrow \dot{\alpha} = -\omega_\mathrm{S} v_y + \omega_\mathrm{S}(1+\epsilon) v_x. \tag{5.39}$$

Now we are able to derive a differential equation for v_x, i.e., for the bunch length oscillation.

Combining Eqs. (5.37) and (5.38) yields

$$\dot{v}_y = -(1+\epsilon)\dot{v}_x. \tag{5.40}$$

We now combine Eq. (5.37) with Eq. (5.39):

$$\ddot{v}_x = 2\omega_S^2 v_y - 2\omega_S^2(1+\epsilon)v_x \tag{5.41}$$

$$\Rightarrow \dddot{v}_x = 2\omega_S^2 \dot{v}_y - 2\omega_S^2(1+\epsilon)\dot{v}_x - 2\omega_S^2 \dot{\epsilon} v_x.$$

Using Eq. (5.40), we finally get

$$\dddot{v}_x = -4\omega_S^2(1+\epsilon)\dot{v}_x - 2\omega_S^2 \dot{\epsilon} v_x. \tag{5.42}$$

Please note that for $\epsilon = 0$, the standard differential equation

$$\ddot{v}_x + (2\omega_S)^2 v_x = \text{const} \tag{5.43}$$

is obtained, which corresponds to an oscillation with frequency $2\omega_S$, a so-called **quadrupole oscillation**. Due to the linearization, an initial quadrupole oscillation will continue forever.

We saw above that a dipole oscillation is generated if a matched bunch is placed off-center in the bucket. In order to generate a quadrupole oscillation, we do not place the bunch off-center in the bucket, but we consider an unmatched bunch. Since a matched bunch has a circular shape in our phase space coordinates (x, y), this means that a slightly elliptical bunch has to be considered. Let us assume that the major axis of the ellipse in phase space is oriented in the x direction at the beginning. Since the individual particles rotate in phase space with the synchrotron frequency, it is then clear that after a quarter of a synchrotron period, the major axis will be directed in the y direction, i.e., the ellipse is then standing upright. After half a synchrotron period, the major of the ellipse will again be oriented horizontally. This explains the oscillation frequency $2\omega_S$ of the quadrupole oscillation. The following points have to be emphasized:

- The derivations presented in this section imply that only small deviations from the matched bunch are allowed.
- The larger the bunch is, the more time will the particles on the bunch contour need for one revolution in phase space. Therefore, the oscillation is more accurately described if an effective synchrotron frequency ω_S is considered instead of the synchrotron frequency $\omega_{S,0}$, which is valid for particles close to the bucket center. In this case, ω_S describes the coherent motion of the particles.
- For realistic bunch sizes, filamentation will occur after a few oscillation periods, so that that a pure quadrupole oscillation will be visible only at the very beginning.

Now we derive the differential equation for v_y, i.e., for the amplitude oscillation. Equation (5.38) yields

$$\frac{\dot{v}_y}{1+\epsilon} = 2\omega_S \alpha.$$

The time derivative is

$$\frac{\ddot{v}_y(1+\epsilon)-\dot{\epsilon}\dot{v}_y}{(1+\epsilon)^2} = -2\omega_S^2 v_y + 2\omega_S^2(1+\epsilon)v_x,$$

where we used Eq. (5.39) on the right-hand side. We divide by $(1+\epsilon)$:

$$\frac{\ddot{v}_y}{(1+\epsilon)^2} - \frac{\dot{\epsilon}\dot{v}_y}{(1+\epsilon)^3} = -2\omega_S^2 \frac{v_y}{1+\epsilon} + 2\omega_S^2 v_x.$$

Now another time derivative leads to $\dot{v}_x$ on the right-hand side, so that we can use Eq. (5.40) to eliminate v_x completely. After some steps, we obtain

$$\dddot{v}_y - \frac{3\ddot{v}_y\dot{\epsilon}}{1+\epsilon} + \dot{v}_y\left(4\omega_S^2(1+\epsilon) - \frac{\ddot{\epsilon}}{1+\epsilon} + \frac{3\dot{\epsilon}^2}{(1+\epsilon)^2}\right) = 2\omega_S^2\dot{\epsilon}v_y. \quad (5.44)$$

This differential equation for v_y differs from Eq. (5.42) for v_x only by terms that are of higher order with respect to ϵ. Furthermore, the sign of the excitation term $2\omega_S^2\dot{\epsilon}v_y$ is different for v_x and v_y, which matches our expectation, since the bunch is short when its amplitude is high, whereas the bunch is long when its amplitude is small.

Please note that we have not introduced any approximations to derive the differential equations (5.42) and (5.44) from Eqs. (5.31) and (5.32).

According to [4], these differential equations have the following solution:

$$v_x = C_{x1}w_{x1}^2 + C_{x2}w_{x1}w_{x2} + C_{x3}w_{x2}^2, \quad (5.45)$$

$$v_y = C_{y1}w_{y1}^2 + C_{y2}w_{y1}w_{y2} + C_{y3}w_{y2}^2. \quad (5.46)$$

The functions w_{x1} and w_{x2} are the linearly independent solutions of

$$\ddot{w}_x + \omega_S^2(1+\epsilon)w_x = 0,$$

whereas the functions w_{y1} and w_{y2} are the linearly independent solutions of

$$\ddot{w}_y - \frac{\dot{\epsilon}}{1+\epsilon}\dot{w}_y + \omega_S^2(1+\epsilon)w_y = 0.$$

In the trivial case $\epsilon = 0$, we may choose

$$w_{x1} = w_{y1} = \cos(\omega_S t), \qquad w_{x2} = w_{y2} = \sin(\omega_S t)$$

as a solution. Due to Eqs. (5.45) and (5.46), v_x and v_y will oscillate with twice the frequency, in accord with Eq. (5.43).

5.6.3 *Linearization*

The derived equations (5.33), (5.34), (5.41), and (5.40),

$$\begin{aligned}
\dot{\bar{x}} &= -\omega_S \bar{y}, \\
\dot{\bar{y}} &= \omega_S(1+\epsilon)(\bar{x} - \Delta\tilde{\varphi}_{gap}), \\
\ddot{v}_x &= 2\omega_S^2 v_y - 2\omega_S^2(1+\epsilon)v_x, \\
\dot{v}_y &= -(1+\epsilon)\dot{v}_x,
\end{aligned}$$

can be written as a **state-space model**[4]

$$\dot{\vec{x}} = \vec{f}(\vec{x}, \epsilon, \Delta\tilde{\varphi}_{gap})$$

with **state vector**

$$\vec{x} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{pmatrix}^T = \begin{pmatrix} \bar{x} & \bar{y} & v_x & \dot{v}_x & v_y \end{pmatrix}^T$$

and the nonlinear function

$$\vec{f}(\vec{x}, \epsilon, \Delta\tilde{\varphi}_{gap}) = \begin{pmatrix} -\omega_S x_2 \\ \omega_S(1+\epsilon)(x_1 - \Delta\tilde{\varphi}_{gap}) \\ x_4 \\ 2\omega_S^2 x_5 - 2\omega_S^2(1+\epsilon)x_3 \\ -(1+\epsilon)x_4 \end{pmatrix}.$$

In the following, a linearization with $\Delta\vec{x} = \vec{x} - \vec{x}_{op}$ around the **operating point**

$$\vec{x}_{op} = \begin{pmatrix} 0 & 0 & v_0 & 0 & v_0 \end{pmatrix}^T, \quad \epsilon_{op} = 0, \quad \Delta\tilde{\varphi}_{gap,op} = 0$$

is performed, which corresponds to the matched circle-shaped bunch. This linearization (see Sect. 7.1.3, cf. [18]) leads to the linear system

$$\Delta\dot{\vec{x}}(t) = A \cdot \Delta\vec{x}(t) + \vec{b}_1\epsilon(t) + \vec{b}_2\Delta\tilde{\varphi}_{gap}(t) \tag{5.47}$$

with the system matrix (Jacobian matrix)

$$A = \left.\frac{\partial \vec{f}}{\partial \vec{x}}\right|_{op} = \begin{pmatrix} 0 & -\omega_S & 0 & 0 & 0 \\ \omega_S & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -2\omega_S^2 & 0 & 2\omega_S^2 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}$$

[4] A general discussion about the state-space representation is presented in Sect. 7.1.2.

and the input matrices

$$\vec{b}_1 = \left.\frac{\partial \vec{f}}{\partial \epsilon}\right|_{\text{op}} = \begin{pmatrix} 0 & 0 & 0 & -2\omega_S^2 v_0 & 0 \end{pmatrix}^{\text{T}}$$

and

$$\vec{b}_2 = \left.\frac{\partial \vec{f}}{\partial \Delta\tilde{\varphi}_{\text{gap}}}\right|_{\text{op}} = \begin{pmatrix} 0 & -\omega_S & 0 & 0 & 0 \end{pmatrix}^{\text{T}}.$$

Please note that the matrix A has a block diagonal structure with one block corresponding to the dynamics of the bunch center $\bar{x}$ and one to the dynamics of the bunch variance v_x. In addition, the bunch center is influenced only by $\Delta\tilde{\varphi}_{\text{gap}}$, and the bunch variance only by ϵ.

Comparing the equations for Δx_3 and Δx_5 in (5.47) yields $\Delta\dot{x}_3(t) = -\Delta\dot{x}_5(t)$ and thus

$$\Delta x_3(t) + \Delta x_5(t) = v_x(t) + v_y(t) - 2v_0 = \text{const}, \tag{5.48}$$

which implies that the bunch variances are connected by an algebraic equation and cannot be controlled independently. It must be possible in principle that the solution $\vec{x}$ of the differential equation reaches the operating point $\vec{x}_{\text{op}}$ (e.g., as an initial condition). For the operating point,

$$\Delta x_3 + \Delta x_5 = 0 \tag{5.49}$$

is valid, which therefore holds in general for every t, due to Eq. (5.48).

With (5.47) and (5.49), linear differential equations of second order can be derived for the bunch center using $\Delta x_1 = \bar{x}$ and for the bunch variance using $\Delta x_3 = v_x - v_0$:

$$\ddot{\bar{x}} + \omega_S^2 \bar{x} = \omega_S^2 \Delta\tilde{\varphi}_{\text{gap}},$$

$$\ddot{v}_x + 4\omega_S^2 (v_x - v_0) = -2\omega_S^2 v_0 \epsilon.$$

A phase modulation mainly influences the dipole oscillation whereas an amplitude modulation primarily affects the quadrupole oscillation.

5.6.4 RMS Emittance

In the previous sections, we defined the longitudinal emittance of the beam as the area in phase space that is filled by the particles. This is a very transparent definition

from a geometric point of view. For a real particle distribution whose density differs from point to point in phase space, however, this definition is not satisfactory. In practice, one cannot decide easily where the boundaries of a given bunch are.

Therefore, one sometimes defines the emittance as the area of a contour that contains 95% of all particles. On the one hand, this percentage is chosen arbitrarily, and on the other hand, it is not easy to define this contour in a unique way. Therefore, an emittance definition is needed that can be determined for every cloud of particles. The motivation for this definition of the **RMS emittance** [19,20], which is presented in the following, is based on [21].

Let us assume that particle number i is located in a Cartesian coordinate system (which represents our phase space) at the position $\vec{r}_i = x_i\ \vec{e}_x + y_i\ \vec{e}_y$. The area of the triangle formed by particle number i, particle number k, and the origin is then given by the vector product

$$\frac{1}{2}\,\vec{r}_i \times \vec{r}_k = \frac{1}{2}\,(x_i\ \vec{e}_x + y_i\ \vec{e}_y) \times (x_k\ \vec{e}_x + y_k\ \vec{e}_y) = \frac{\vec{e}_z}{2}\,(x_i\ y_k - x_k y_i).$$

$$\Rightarrow A_{ik} = \frac{1}{2}\,|x_i\ y_k - x_k y_i\,|.$$

For the next step, we abandon the idea of calculating an exact area for the particle cloud. We just need an expression that has some similarity to an area. Therefore, we simply sum up the squares of these areas for all possible particle pairs:

$$A^2 = \frac{1}{N_\mathrm{b}^2 - N_\mathrm{b}} \sum_{i=1}^{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} A_{ik}^2.$$

The terms where $i = k$ holds do not contribute anything to the sum. Therefore, we divided by $N_\mathrm{b}(N_\mathrm{b} - 1)$ to get the average. We obtain

$$A^2 = \frac{1}{4(N_\mathrm{b}^2 - N_\mathrm{b})} \sum_{i=1}^{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (x_i\ y_k - x_k y_i)^2 = \frac{1}{4(N_\mathrm{b}^2 - N_\mathrm{b})} \sum_{i=1}^{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (x_i^2\ y_k^2 + x_k^2\ y_i^2 - 2x_i\ x_k\ y_i\ y_k).$$

The first two terms lead to the same sum, since only the roles of i and k are interchanged:

$$A^2 = \frac{1}{4(N_\mathrm{b}^2 - N_\mathrm{b})} \sum_{i=1}^{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (2x_i^2\ y_k^2 - 2x_i\ x_k\ y_i\ y_k) = \frac{1}{2(N_\mathrm{b}^2 - N_\mathrm{b})} \sum_{i=1}^{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (x_i^2\ y_k^2 - x_i\ x_k\ y_i\ y_k). \tag{5.50}$$

We now return to our definitions (5.36),

$$a_x := \overline{x^2} = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} x_k^2$$

and

$$a_y := \overline{y^2} = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} y_k^2,$$

which lead to

$$a_x a_y = \overline{x^2}\,\overline{y^2} = \frac{1}{N_\mathrm{b}^2} \left(\sum_{i=1}^{N_\mathrm{b}} x_i^2 \right) \left(\sum_{k=1}^{N_\mathrm{b}} y_k^2 \right) = \frac{1}{N_\mathrm{b}^2} \sum_{i=1}^{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} (x_i^2 y_k^2).$$

This obviously reproduces the first term in Eq. (5.50). In a similar way, the definition

$$\xi := \overline{xy} = \frac{1}{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} x_k y_k$$

leads to

$$\xi^2 = (\overline{xy})^2 = \frac{1}{N_\mathrm{b}^2} \sum_{i=1}^{N_\mathrm{b}} \sum_{k=1}^{N_\mathrm{b}} x_i x_k y_i y_k,$$

which reproduces the second term in Eq. (5.50). We therefore get

$$2A^2 \left(1 - \frac{1}{N_\mathrm{b}} \right) = a_x a_y - \xi^2 = \overline{x^2}\,\overline{y^2} - (\overline{xy})^2.$$

Please note that the term in parentheses is close to 1 for sufficiently large numbers N_b. Since we summed up the squares of the areas of the triangles, we have to apply the square root to get a quantity with the correct dimension. This defines the **RMS emittance**

$$\varepsilon = C_\varepsilon \sqrt{\overline{x^2}\,\overline{y^2} - (\overline{xy})^2},$$

for which different constant factors C_ε are used in the literature (e.g., $C_\varepsilon = 4$ in [19]).

Due to the construction of the emittance (based on triangles that have the origin as one vertex), it is obvious that the value ε is not invariant under translations of the origin. Points that are far from the origin contribute with large areas, whereas points close to the origin contribute with small areas. Therefore, the origin should correspond to the center of the particle cloud:

$$\bar{x} = 0, \qquad \bar{y} = 0.$$

Nevertheless, the RMS emittance has the disadvantage that particles at the boundaries of the bunch have more influence than they deserve and that it is not always a measure for the area (especially for deformed bunches). For bunches with small distortions, however, the RMS emittance may often be used successfully.

We may regard the RMS definition as a special case for a finite number of particles. For a continuous distribution, one would use expected values instead:

$$\left(\frac{\varepsilon}{C_\varepsilon}\right)^2 = E(X^2)\ E(Y^2) - E^2(XY), \qquad E(X) = 0, \qquad E(Y) = 0.$$

If one now assumes that the random variables X and Y are independent, one sees that

$$\left(\frac{\varepsilon}{C_\varepsilon}\right)^2 = E(X^2)\ E(Y^2)$$

holds. In this case,

$$\frac{\varepsilon}{C_\varepsilon} = \sigma_X\ \sigma_Y$$

is obtained. For a bunch that has an elliptical shape in phase space, this result is expected, because the area of an ellipse with the two semiaxes r_X, r_Y is $\pi r_X r_Y$. If the distribution of the particles is Gaussian, it is obvious that C_ε can be used to define an elliptical contour that contains a certain percentage of particles.

For purposes of illustration, we conclude this section with a simple example. Let us consider only $N_\mathrm{b} = 3$ particles with the following positions:

$$\vec{r}_1 = (+10, +10),$$
$$\vec{r}_2 = (0, +10),$$
$$\vec{r}_3 = (-10, +10).$$

This leads to $N_\mathrm{b}(N_\mathrm{b} - 1)/2 = 3$ triangles with areas 50, 50, and 100, respectively. The average of the squares is $15000/3 = 5000$, so that

$$A = \sqrt{5000} \approx 70.71$$

is obtained. This can also be calculated formally based on

$$a_x = \frac{200}{3}, \qquad a_y = \frac{300}{3} = 100, \qquad \xi = 0.$$

In this example, the origin does not correspond to the center of the particle cloud. If we shifted the origin to this center so that

$$\bar{x} = 0, \qquad \bar{y} = 0$$

holds, we would get an area of zero ($A = 0$). This is due to the fact that the three particles are located on a straight line.

5.7 Longitudinal Bunch Oscillations

We now analyze a few specific oscillations of bunched beams in longitudinal phase space. The undisturbed bunch, i.e., a bunch that is matched to the bucket, is the starting point. As in the previous section, a scaling of the phase space coordinates will be performed in such a way that the contour of the matched bunch is a circle. Instead of physical phase space variables, we again use simple coordinates (x, y) for which the undisturbed bunch is a unit circle:

$$x = \cos\varphi, \qquad y = \sin\varphi.$$

The physical phase space representation is obtained if both coordinates are multiplied by the corresponding factors.

5.7.1 Coherent Dipole Mode

As we already discussed in the previous section, the coherent dipole mode of oscillation is obtained if the bunch as a whole is shifted along one coordinate so that it is located off-center in the bucket afterward:

$$x = \epsilon + \cos\varphi, \qquad y = \sin\varphi.$$

For the radius r, we obtain

$$r^2 = x^2 + y^2 = 1 + \epsilon^2 + 2\epsilon\ \cos\varphi.$$

With the help of

$$\sqrt{1+a} \approx 1 + \frac{a}{2},$$

we find, for sufficiently small $\epsilon \ll 1$,

$$r \approx 1 + \epsilon\ \cos\varphi.$$

5.7.2 Quadrupole Mode

A quadrupole oscillation is obtained if the ratio of the principal axes is slightly modified. For our circular bunch, this means that it becomes elliptical:

$$x = a\ \cos\varphi, \qquad y = b\ \sin\varphi,$$

$$r^2 = a^2\cos^2\varphi + b^2\sin^2\varphi = a^2 + (b^2 - a^2)\sin^2\varphi.$$

Due to

$$\sin^2\varphi = \frac{1}{2} - \frac{1}{2}\cos(2\varphi),$$

we obtain

$$r^2 = \frac{a^2+b^2}{2} + \frac{a^2-b^2}{2}\ \cos(2\varphi) = \frac{a^2+b^2}{2}\left(1 + \frac{a^2-b^2}{a^2+b^2}\ \cos(2\varphi)\right).$$

For

$$a = 1 + \frac{\epsilon}{2}, \qquad b = 1 - \frac{\epsilon}{2},$$

we obtain

$$r \approx 1 + \epsilon\ \cos(2\varphi).$$

The quadrupole mode leads to the effect that the bunch is elongated in phase space. The rotation in phase space then leads to an oscillation between short bunches with large peak current and long bunches with small peak current. Hence, a bunch length and bunch amplitude modulation is present.

5.7.3 Generalization

We are now able to see that the two cases discussed before may be generalized to the formula

$$r \approx 1 + \epsilon\ \cos(m\varphi).$$

The dipole mode is obtained for $m = 1$, the quadrupole oscillation for $m = 2$. For $m = 3$, the **sextupole oscillation** is obtained; for $m = 4$, the **octupole mode**, etc.

The mode number m also specifies the eigenfrequency of the oscillation. As the diagrams in Fig. 5.10 show, the bunch in phase space is a polygon with m rounded

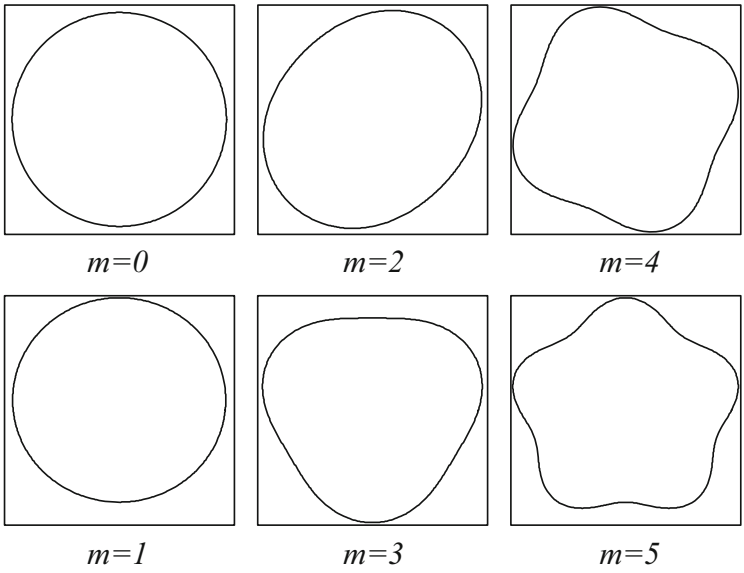

Fig. 5.10 Longitudinal modes of oscillation

corners. The rotation of the bunch with frequency f_S results in a projection onto the time axis whose frequency is m times higher. Together with the revolution period, one obtains spectral components at $kf_R \pm mf_S$.

In the case of coupled-bunch oscillations, the h bunches in the ring interact with each other. In general, not only do the bunches oscillate as a whole (as would be the case for $m = 1$), but the individual bunches may be deformed, as described by the **bunch shape mode number** m.

If we consider coupled oscillations for a specific m, we find that there are h modes altogether, which are characterized by the **coupled-bunch mode number** n with $0 \leq n \leq h-1$, since h bunches may oscillate in h ways with respect to one another. In the simplest case, $m = 1$ (dipole oscillation of individual bunches) and $h = 2$, the two bunches may oscillate either in phase ($n = 0$) or out of phase ($n = 1$). In general, the phase advance from bunch to bunch is

$$2\pi \frac{n}{h}.$$

Hence, coupled-bunch oscillations may be characterized using the two mode numbers m and n (cf. [22] and [23, Sect. 5.6]).

The modes of oscillation may alternatively be defined on the basis of the spectral lines observed in the beam signal. If, for example, strong spectral components are observed at $k f_R \pm 2 f_S$, this would indicate a quadrupole oscillation by definition.

The modes of oscillation that were introduced here may be excited if the initial conditions do not correspond to those of a matched bunch. If, for example, the bunch is not centered inside the bucket, coherent dipole oscillations will be the result.

A different cause of longitudinal oscillations is (longitudinal) impedances that act on the beam. If these impedances lead to an unstable situation, specific modes of oscillation will be excited; they have certain growth rates. With respect to the modes of oscillation defined above with mode numbers m and n, one then speaks of coupled-bunch instabilities. Instabilities may be damped using **feedback systems** (cf. [24]). So-called longitudinal feedback systems are used to reduce undesired longitudinal beam oscillations, whereas transverse feedback systems damp transverse beam oscillations.

5.7.4 Spectrum of the Dipole Oscillation

Consider the time function

$$f(t) = \sum_{k=-\infty}^{+\infty} \delta(t - kT_R - \tau_k), \tag{5.51}$$

where τ_k is periodic with period T_S of the synchrotron oscillation. Here T_R is the revolution time,[5] and T_S is an integer multiple of T_R, so that $f(t)$ is strictly periodic with period T_S. The function $f(t)$ obviously represents a strongly bunched beam that performs coherent dipole oscillations. The corresponding Fourier coefficients are

$$c_n = \frac{1}{T_S} \int_{-T_S/2}^{+T_S/2} f(t)\, e^{-jn\omega_S t}\, \mathrm{d}t.$$

Hence we have

$$c_n = \frac{1}{T_S} \sum_{k \in M} e^{-jn\omega_S [kT_R + \tau_k]}\, \mathrm{d}t.$$

---

[5] We assume $h = 1$ here.

Here M denotes the set of all indices for which $kT_\mathrm{R} + \tau_k$ runs from $-T_\mathrm{S}/2$ to $+T_\mathrm{S}/2$. In order to determine this set in a unique way, we want to satisfy

$$\tau_k \approx \epsilon \ \sin(\omega_\mathrm{S} t)$$

approximately. Since $\omega_\mathrm{S} \ll \omega_\mathrm{R}$ holds, we may insert

$$t = kT_\mathrm{R}$$

approximately, so that

$$\tau_k = \epsilon \ \sin(k \ \omega_\mathrm{S} T_\mathrm{R}) = \epsilon \ \sin(k \ 2\pi T_\mathrm{R}/T_\mathrm{S}) \tag{5.52}$$

is now considered as an exact definition. The limits of integration will be obtained for

$$kT_\mathrm{R} = \pm T_\mathrm{S}/2,$$

i.e., for

$$k = \pm \frac{T_\mathrm{S}}{2T_\mathrm{R}}.$$

At this point, we require that $T_\mathrm{S}/T_\mathrm{R} = \omega_\mathrm{R}/\omega_\mathrm{S}$ be even in order to have an integer k at the integration limits. For these values of k, the quantity τ_k vanishes. Therefore, the integration limits are located exactly on two Dirac pulses. Since we have to integrate only one period, only one of these two Dirac pulses must be taken into account in the summation. Therefore, we have

$$M = \left\{ -\left(\frac{T_\mathrm{S}}{2T_\mathrm{R}} - 1 \right), \ldots, \frac{T_\mathrm{S}}{2T_\mathrm{R}} \right\}.$$

Therefore, we obtain

$$c_n = \frac{1}{T_\mathrm{S}} + \frac{1}{T_\mathrm{S}} e^{-jn\omega_\mathrm{S}\left[\frac{T_\mathrm{S}}{2} + \epsilon \sin \pi\right]} + \frac{2}{T_\mathrm{S}} \sum_{k=1}^{\frac{T_\mathrm{S}}{2T_\mathrm{R}}-1} \cos\left(n\omega_\mathrm{S}\left[kT_\mathrm{R} + \epsilon \ \sin\left(2\pi k T_\mathrm{R}/T_\mathrm{S}\right)\right]\right).$$

The first term is obtained for $k = 0$, the second one for $k = \frac{T_\mathrm{S}}{2T_\mathrm{R}}$. It follows that

$$\boxed{c_n = \frac{1}{T_\mathrm{S}} \left[1 + (-1)^n + 2 \sum_{k=1}^{\frac{T_\mathrm{S}}{2T_\mathrm{R}}-1} \cos\left(2\pi n \frac{1}{T_\mathrm{S}} \left[kT_\mathrm{R} + \epsilon \ \sin\left(2\pi k T_\mathrm{R}/T_\mathrm{S}\right)\right]\right)\right].} \tag{5.53}$$

We are now looking for a specific expression for the coefficients c_n at the revolution harmonics.

The coefficients c_n correspond to the frequencies $n\omega_S$. Hence the revolution harmonics are located at

$$p\omega_R = p\frac{\omega_R}{\omega_S}\omega_S = p\frac{T_S}{T_R}\omega_S \qquad \Rightarrow n = p\frac{T_S}{T_R}.$$

Since T_S/T_R is an even number, n is also even, and it follows that

$$c_n = \frac{2}{T_S} + \frac{2}{T_S}\sum_{k=1}^{\frac{T_S}{2T_R}-1} \cos\left(2\pi p\frac{\epsilon}{T_R}\ \sin\left(2\pi k T_R/T_S\right)\right).$$

For $\epsilon = 0$ (no coherent dipole oscillation), we obtain

$$c_n = \frac{2}{T_S} + \frac{2}{T_S}\left(\frac{T_S}{2T_R} - 1\right) = \frac{1}{T_R},$$

which is what one expects for a simple Dirac comb without phase modulation. Now we would have to show that the coefficients c_n that do not correspond to revolution harmonics vanish for $\epsilon = 0$. At this point, however, we omit the calculation (for odd ratios T_S/T_R, the calculation is presented in the next section).

Now let T_S/T_R be an odd integer. Then the summation limits are defined by

$$k = \pm\frac{1}{2}\left(\frac{T_S}{T_R} - 1\right).$$

It follows that

$$\boxed{c_n = \frac{1}{T_S}\left[1 + 2\sum_{k=1}^{\frac{1}{2}\left(\frac{T_S}{T_R}-1\right)} \cos\left(2\pi n\frac{1}{T_S}\left[kT_R + \epsilon\ \sin\left(2\pi k T_R/T_S\right)\right]\right)\right].} \tag{5.54}$$

Also here we seek a specific expression for the coefficients c_n at the revolution harmonics. To this end, we make the substitution

$$n = p\frac{T_\mathrm{S}}{T_\mathrm{R}}$$

and obtain

$$c_n = \frac{1}{T_\mathrm{S}}\left[1 + 2\sum_{k=1}^{\frac{1}{2}\left(\frac{T_\mathrm{S}}{T_\mathrm{R}}-1\right)} \cos\left(2\pi p\frac{\epsilon}{T_\mathrm{R}}\ \sin\left(2\pi k T_\mathrm{R}/T_\mathrm{S}\right)\right)\right].$$

As expected, we get for $\epsilon = 0$,

$$c_n = \frac{1}{T_\mathrm{S}}\left[1 + \left(\frac{T_\mathrm{S}}{T_\mathrm{R}} - 1\right)\right] = \frac{1}{T_\mathrm{R}}.$$

Now we show that all other coefficients that do not belong to the revolution harmonics equal zero. As one finds in Gradshteyn [25] (formula 1.342,2),

$$\sum_{k=1}^{m}\cos(kx) = \cos\left(\frac{m+1}{2}x\right)\ \sin\frac{mx}{2}\ \mathrm{cosec}\frac{x}{2}$$

holds. From Eq. (5.54), we conclude for $\epsilon = 0$ with the help of this formula that

$$\begin{aligned} c_n = \frac{1}{T_\mathrm{S}}\Bigg(1 + 2\ \cos\left[\frac{1}{4}\left(\frac{T_\mathrm{S}}{T_\mathrm{R}} + 1\right)\ 2\pi n\frac{T_\mathrm{R}}{T_\mathrm{S}}\right]\ \cdot \\ \cdot \sin\left[\frac{1}{4}\left(\frac{T_\mathrm{S}}{T_\mathrm{R}} - 1\right)\ 2\pi n\frac{T_\mathrm{R}}{T_\mathrm{S}}\right]\ \mathrm{cosec}\left[\pi n\frac{T_\mathrm{R}}{T_\mathrm{S}}\right]\Bigg) \end{aligned}$$

$$\Rightarrow c_n = \frac{1}{T_\mathrm{S}}\left(1 + 2\ \cos\left[\frac{\pi}{2}n\left(1 + \frac{T_\mathrm{R}}{T_\mathrm{S}}\right)\right]\ \sin\left[\frac{\pi}{2}n\left(1 - \frac{T_\mathrm{R}}{T_\mathrm{S}}\right)\right]\ \mathrm{cosec}\left[\pi n\frac{T_\mathrm{R}}{T_\mathrm{S}}\right]\right).$$

From

$$a = \frac{\pi}{2}n \qquad \text{and} \qquad b = \frac{\pi}{2}n\frac{T_\mathrm{R}}{T_\mathrm{S}},$$

it follows that

$$c_n = \frac{1}{T_\mathrm{S}}\left[1 + 2\frac{\cos(a+b)\ \sin(a-b)}{\sin(2b)}\right].$$

With the help of trigonometric identities, one easily shows that

$$\frac{\cos(a+b)\ \sin(a-b)}{\sin(2b)} = \frac{1}{2}\left(\frac{\sin(2a)}{\sin(2b)} - 1\right)$$

is valid. This leads to

$$c_n = \frac{1}{T_\mathrm{S}} \frac{\sin(2a)}{\sin(2b)} = \frac{1}{T_\mathrm{S}} \frac{\sin(\pi n)}{\sin\left(\pi n \frac{T_\mathrm{R}}{T_\mathrm{S}}\right)}.$$

The numerator is always equal to zero. The coefficient c_n can equal $1/T_\mathrm{S}$ only if the denominator is zero as well, which is the case if $nT_\mathrm{R}/T_\mathrm{S}$ is an integer, i.e., if

$$n \frac{T_\mathrm{R}}{T_\mathrm{S}} = p$$

holds. This, however, is just the condition derived above for the beam harmonics. Hence, the spectrum for $\epsilon = 0$ corresponds to a Dirac comb, as expected for a Dirac comb in the time domain.

The Fourier coefficients derived here (Eq. (5.53) for even ratios $T_\mathrm{S}/T_\mathrm{R}$ and Eq. (5.54) for odd ratios $T_\mathrm{S}/T_\mathrm{R}$) are exact formulas without approximations that are valid for the coherent dipole oscillation defined by Eqs. (5.51) and (5.52). These formulas can also be proven in the scope of distribution theory [16]. It is easily possible to consider realistic bunches instead of the Dirac pulses if one performs a convolution as shown in Sect. 2.1.6.

5.8 A Simple Space Charge Model

We have heretofore assumed that each charged particle with charge Q that belongs to a bunch experiences the same voltage $V(t)$ (depending, of course, on its arrival time) produced by a cavity. If the density of particles becomes larger and larger, this is no longer the case. The charge distribution of the whole particle cloud will influence an individual particle. Such phenomena are called **space charge effects**.

In this section, a simple space charge model is analyzed. For this purpose, we consider a reference frame in which the particle bunches are at rest. Therefore, a pure electrostatic problem with $\vec{B} = 0$ has to be solved. In this case, Maxwell's equations reduce to

$$\mathrm{div}\ \vec{D} = \rho_q, \tag{5.55}$$

$$\mathrm{curl}\ \vec{E} = 0. \tag{5.56}$$

5.8.1 Field in the Rest Frame of the Bunch

We now assume that the beam pipe is perfectly conducting and that it has a cylindrical shape (radius r_bp). The longitudinal axis of the beam pipe defines the z-axis of a cylindrical coordinate system (coordinates ρ, φ, z), and the beam pipe is

assumed to be infinitely long. This means that the curvature of the synchrotron is neglected. The beam itself is modeled as a charge distribution with nonzero charge density inside a cylinder of radius $r_{\text{beam}} < r_{\text{bp}}$:

$$\rho_q(\rho, z) = \begin{cases} \rho_{q,0}(z) & \text{for} \quad 0 \le \rho < r_{\text{beam}} \quad \text{(region A)}, \\ 0 & \text{for} \quad r_{\text{beam}} \le \rho < r_{\text{bp}} \quad \text{(region B)}. \end{cases}$$

In each cross section at constant z, the space charge density is constant for $\rho < r_{\text{beam}}$. According to[6]

$$\lambda_q(z) = \pi r_{\text{beam}}^2 \rho_{q,0}(z),$$

this space charge density ρ_q may be converted into a line charge density λ_q.

Due to the special setup, we may require $D_\varphi = 0$ and no φ-dependence of the fields. In cylindrical coordinates, Eqs. (5.55) and (5.56) may be written as

$$\frac{\partial D_\rho}{\partial \rho} + \frac{1}{\rho} D_\rho + \frac{\partial D_z}{\partial z} = \rho_q(\rho, z), \tag{5.57}$$

$$\frac{\partial D_\rho}{\partial z} - \frac{\partial D_z}{\partial \rho} = 0. \tag{5.58}$$

We calculate the derivative of the first equation with respect to ρ and insert the second equation:

$$\frac{\partial^2 D_\rho}{\partial \rho^2} - \frac{1}{\rho^2} D_\rho + \frac{1}{\rho} \frac{\partial D_\rho}{\partial \rho} + \frac{\partial^2 D_\rho}{\partial z^2} = 0. \tag{5.59}$$

If one uses the derivative of Eq. (5.57) with respect to z instead, one obtains by inserting Eq. (5.58),

$$\frac{\partial^2 D_z}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial D_z}{\partial \rho} + \frac{\partial^2 D_z}{\partial z^2} = \frac{\partial \rho_q}{\partial z}. \tag{5.60}$$

We now attempt to solve Eq. (5.59) by means of a separation ansatz:

$$D_\rho = f(\rho)\, g(z)$$

[6] The total charge of each bunch is obtained by

$$N_b Q = \int \rho_q \, \mathrm{d}V = \int \lambda_q \, \mathrm{d}z.$$

$$\Rightarrow g\,\frac{\partial^2 f}{\partial\rho^2} - \frac{1}{\rho^2}\,f\,g + \frac{1}{\rho}\,g\,\frac{\partial f}{\partial\rho} + f\,\frac{\partial^2 g}{\partial z^2} = 0$$

$$\Rightarrow \frac{1}{f}\,\frac{\partial^2 f}{\partial\rho^2} - \frac{1}{\rho^2} + \frac{1}{\rho f}\,\frac{\partial f}{\partial\rho} + \frac{1}{g}\,\frac{\partial^2 g}{\partial z^2} = 0.$$

The last term on the left-hand side may depend only on z, whereas all the other terms may depend only on ρ. Therefore, these terms must be constant:

$$\frac{1}{f}\,\frac{\partial^2 f}{\partial\rho^2} - \frac{1}{\rho^2} + \frac{1}{\rho f}\,\frac{\partial f}{\partial\rho} = C_0,$$

$$\frac{1}{g}\,\frac{\partial^2 g}{\partial z^2} = -C_0.$$

We need solutions that are periodic in the z direction, because the fields must repeat themselves after one revolution in the synchrotron (we assume that this requirement in combination with the straight cylindrical beam pipe leads to solutions that are similar to a closed, bent beam pipe). Therefore, $C_0 > 0$ will be valid, and we set $C_0 = k_z^2$:

$$\rho^2\frac{\mathrm{d}^2 f}{\mathrm{d}\rho^2} + \rho\,\frac{\mathrm{d}f}{\mathrm{d}\rho} - (\rho^2 k_z^2 + 1)\,f = 0, \tag{5.61}$$

$$\frac{\mathrm{d}^2 g}{\mathrm{d}z^2} + k_z^2 g = 0.$$

The second equation obviously has the solution

$$g(z) = g_1\,\cos(k_z z) + g_2\,\sin(k_z z)$$

with constants g_1 and g_2. Due to the periodicity of the solutions after one synchrotron revolution,

$$k_z l = 2\pi p \qquad \Rightarrow k_z = p\frac{2\pi}{l} \qquad \text{with} \qquad p \in \{0, 1, 2, \ldots\}$$

must be valid.

By means of the substitution $u = k_z\rho$, the first equation (5.61) may be transformed into the **modified Bessel's differential equation**

$$\boxed{u^2\frac{\mathrm{d}^2 f}{\mathrm{d}u^2} + u\,\frac{\mathrm{d}f}{\mathrm{d}u} - (u^2 + m^2)\,f = 0}$$

with $m = 1$. Two independent solutions are the **modified Bessel functions** $\mathrm{I}_m(u)$ and $\mathrm{K}_m(u)$; see Table A.2 on p. 415 and Fig. A.7. Hence, we get

$$f(\rho) = f_1\ \mathrm{I}_1(k_z\rho) + f_2\ \mathrm{K}_1(k_z\rho)$$

$$\Rightarrow D_\rho = (f_1\ \mathrm{I}_1(k_z\rho) + f_2\ \mathrm{K}_1(k_z\rho))\,(g_1\ \cos(k_z z) + g_2\ \sin(k_z z))$$

with constants f_1 and f_2. For the derivative with respect to z, we obtain

$$\frac{\partial D_\rho}{\partial z} = k_z\,(f_1\ \mathrm{I}_1(k_z\rho) + f_2\ \mathrm{K}_1(k_z\rho))\,(-g_1\ \sin(k_z z) + g_2\ \cos(k_z z))\,.$$

According to Eq. (5.58), we get D_z by means of an integration with respect to ρ:

$$D_z = k_z\,(-g_1\ \sin(k_z z) + g_2\ \cos(k_z z)) \int (f_1\ \mathrm{I}_1(k_z\rho) + f_2\ \mathrm{K}_1(k_z\rho))\ \mathrm{d}\rho.$$

Since the functions $\mathrm{K}_m(u)$ have poles at $u = 0$ for $m \in \{0, 1, 2, \ldots\}$, the function $\mathrm{K}_1(u)$ cannot be used in region A ($0 \le \rho < r_{\mathrm{beam}}$), because there is no singular charge density at $\rho = 0$. Hence, the solution in region A is

$$D_\rho = \mathrm{I}_1(k_z\rho)\ \left(g_1^A\ \cos(k_z z) + g_2^A\ \sin(k_z z)\right), \tag{5.62}$$

$$D_z = \zeta(\rho)\ \left(-g_1^A\ \sin(k_z z) + g_2^A\ \cos(k_z z)\right), \tag{5.63}$$

where

$$\zeta(\rho) = \int_0^\rho k_z \mathrm{I}_1(k_z\rho)\ \mathrm{d}\rho + \zeta(0) = \mathrm{I}_0(k_z\rho) - 1 + \zeta(0). \tag{5.64}$$

In the last step, we used Table A.2 on p. 415, formula (A.78),

$$\boxed{\mathrm{I}_0'(u) = \mathrm{I}_1(u),} \tag{5.65}$$

which implies

$$\boxed{\int k_z \mathrm{I}_1(k_z\rho)\ \mathrm{d}\rho = \mathrm{I}_0(k_z\rho) + \mathrm{const}}$$

and the function value $\mathrm{I}_0(0) = 1$. In the following, we will also make use of the general formula (A.77),

$$\boxed{\mathrm{K}_0'(u) = -\mathrm{K}_1(u),} \tag{5.66}$$

which implies

$$\boxed{\int k_z \mathrm{K}_1(k_z\rho)\,\mathrm{d}\rho = -\mathrm{K}_0(k_z\rho) + \text{const.}}$$

In region B ($r_{\text{beam}} \le \rho < r_{\text{bp}}$), we have

$$D_\rho = \left(f_1^B\ \mathrm{I}_1(k_z\rho) + f_2^B\ \mathrm{K}_1(k_z\rho)\right)\left(g_1^B\ \cos(k_z z) + g_2^B\ \sin(k_z z)\right), \tag{5.67}$$

$$\begin{aligned} D_z &= \left(f_1^B \int_{r_{\text{beam}}}^{\rho} k_z \mathrm{I}_1(k_z\rho)\,\mathrm{d}\rho + f_2^B \int_{r_{\text{beam}}}^{\rho} k_z \mathrm{K}_1(k_z\rho)\,\mathrm{d}\rho + \zeta(r_{\text{beam}})\right) \cdot \\ &\quad \cdot \left(-g_1^B\ \sin(k_z z) + g_2^B\ \cos(k_z z)\right) = \\ &= \left(f_1^B\,[\mathrm{I}_0(k_z\rho) - \mathrm{I}_0(k_z r_{\text{beam}})] + f_2^B\,[\mathrm{K}_0(k_z r_{\text{beam}}) - \mathrm{K}_0(k_z\rho)] + \zeta(r_{\text{beam}})\right) \cdot \\ &\quad \cdot \left(-g_1^B\ \sin(k_z z) + g_2^B\ \cos(k_z z)\right). \end{aligned} \tag{5.68}$$

The field continuity between the two regions A and B at $\rho = r_{\text{beam}}$ leads to the integration constant $\zeta(r_{\text{beam}})$ in the last equation, and it also implies

$$g_1^A = g_1^B = g_1, \qquad g_2^A = g_2^B = g_2.$$

Furthermore, we get

$$\mathrm{I}_1(k_z r_{\text{beam}}) = f_1^B\ \mathrm{I}_1(k_z r_{\text{beam}}) + f_2^B\ \mathrm{K}_1(k_z r_{\text{beam}}). \tag{5.69}$$

The field continuity is also the reason why we did not use different symbols for k_z in the two regions.

At $\rho = r_{\text{bp}}$, the ideally conducting beam pipe leads to the condition $E_z = 0$ (the longitudinal component of the electric field remains unchanged by the Lorentz transformation, so that $E_z = 0$ in the laboratory frame corresponds to $E_z = 0$ in the rest frame of the beam). This leads to

$$f_1^B\left[\mathrm{I}_0(k_z r_{\text{bp}}) - \mathrm{I}_0(k_z r_{\text{beam}})\right] + f_2^B\left[\mathrm{K}_0(k_z r_{\text{beam}}) - \mathrm{K}_0(k_z r_{\text{bp}})\right] + \zeta(r_{\text{beam}}) = 0. \tag{5.70}$$

The last two equations can be used to determine the constants f_1^B and f_2^B. However, $\zeta(0)$ still has to be calculated in order to get $\zeta(r_{\text{beam}})$ by means of Eq. (5.64).

Now we determine the space charge density in region A. Due to

$$\frac{\partial D_\rho}{\partial \rho} = k_z \mathrm{I}_1'(k_z\rho)\ \left(g_1^A\ \cos(k_z z) + g_2^A\ \sin(k_z z)\right),$$

$$\frac{\partial D_z}{\partial z} = k_z\ \zeta(\rho)\ \left(-g_1^A\ \cos(k_z z) - g_2^A\ \sin(k_z z)\right),$$

we get

$$\rho_q(z) = \frac{\partial D_\rho}{\partial \rho} + \frac{1}{\rho}\, D_\rho + \frac{\partial D_z}{\partial z} =$$
$$= \left(g_1^A\, \cos(k_z z) + g_2^A\, \sin(k_z z)\right) \left(k_z \mathrm{I}_1'(k_z\rho) + \frac{1}{\rho}\, \mathrm{I}_1(k_z\rho) - k_z \zeta(\rho)\right).$$

According to [26, p. 376, formula (9.6.28)], we have

$$\frac{1}{u}\frac{\mathrm{d}}{\mathrm{d}u}\left(u\mathrm{I}_1(u)\right) = \mathrm{I}_0(u)$$
$$\Rightarrow \boxed{\frac{1}{u}\, \mathrm{I}_1(u) + \mathrm{I}_1'(u) = \mathrm{I}_0(u)}$$
$$\Rightarrow \frac{1}{k_z\rho}\, \mathrm{I}_1(k_z\rho) + \mathrm{I}_1'(k_z\rho) = \mathrm{I}_0(k_z\rho),$$

so that

$$\rho_q(z) = \left(g_1^A\, \cos(k_z z) + g_2^A\, \sin(k_z z)\right) \left(k_z \mathrm{I}_0(k_z\rho) - k_z \zeta(\rho)\right),$$

or

$$\rho_q(z) = \xi\, \left(g_1^A\, \cos(k_z z) + g_2^A\, \sin(k_z z)\right), \tag{5.71}$$

with

$$\xi = k_z \mathrm{I}_0(k_z\rho) - k_z \zeta(\rho) = k_z \left(\mathrm{I}_0(k_z\rho) - \zeta(\rho)\right)$$

is obtained. The derivative with respect to ρ is

$$\frac{\mathrm{d}\xi}{\mathrm{d}\rho} = k_z^2 \mathrm{I}_0'(k_z\rho) - k_z^2 \mathrm{I}_0'(k_z\rho) = 0.$$

Here we used Eq. (5.64). With Eq. (5.71),

$$\frac{\partial \rho_q}{\partial \rho} = 0$$

is also valid, which means that in each region, the space charge density depends only on z, as required. Please note that the quantity ξ introduced above does not depend on ρ, even though the individual terms do depend on ρ. One may therefore evaluate ξ for all values of ρ. For $\rho = 0$, we get

$$\xi = k_z \left(1 - \zeta(0)\right), \tag{5.72}$$

since $\mathrm{I}_0(0) = 1$ holds.

As a last step, we now have to satisfy the boundary condition for the normal component of the electric field at $\rho = r_{\mathrm{bp}}$:

$$D_n = \sigma_q .$$

In general, σ_q denotes the surface charge density defined by

$$\int\int\int \rho_q \, \mathrm{d}V = \int\int \sigma_q \, \rho \, \mathrm{d}\varphi \, \mathrm{d}z = \int \lambda_q \, \mathrm{d}z .$$

In our special case, the radius is $\rho = r_{\mathrm{bp}}$, $D_n = -D_\rho$ holds, and the charge on the beam pipe must be the same as the charge of the beam with negative sign (image charge in the rest frame of the beam):

$$\sigma_q \, 2\pi r_{\mathrm{bp}} = -\lambda_q = -\pi r_{\mathrm{beam}}^2 \rho_q$$

$$\Rightarrow D_\rho|_{\rho = r_{\mathrm{bp}}} = \frac{\lambda_q}{2\pi r_{\mathrm{bp}}} = \frac{r_{\mathrm{beam}}^2 \rho_q}{2 r_{\mathrm{bp}}} .$$

According to Eqs. (5.67) and (5.71) this leads to the condition

$$f_1^B \, \mathrm{I}_1(k_z r_{\mathrm{bp}}) + f_2^B \, \mathrm{K}_1(k_z r_{\mathrm{bp}}) = \frac{r_{\mathrm{beam}}^2 \xi}{2 r_{\mathrm{bp}}} . \tag{5.73}$$

In combination with Eqs. (5.69), (5.70), and (5.72), this defines how $\zeta(\rho)$ and ξ depend on each other.

The solution discussed above does not include the case that the charge density and the fields are constant in the longitudinal direction. For $p = 0$, all field components vanish.

Therefore, we now consider Eq. (5.59) for the case that D_ρ does not depend on z:

$$\frac{\mathrm{d}^2 D_\rho}{\mathrm{d}\rho^2} - \frac{1}{\rho^2} \, D_\rho + \frac{1}{\rho} \, \frac{\mathrm{d}D_\rho}{\mathrm{d}\rho} = 0$$

$$\Rightarrow \rho^2 \, \frac{\mathrm{d}^2 D_\rho}{\mathrm{d}\rho^2} + \rho \, \frac{\mathrm{d}D_\rho}{\mathrm{d}\rho} - D_\rho = 0 .$$

This is a homogeneous **Euler–Cauchy ODE**, which can be solved by the substitution

$$\rho = e^u ,$$

so that we have

$$\frac{\mathrm{d}D_\rho}{\mathrm{d}u} = \frac{\mathrm{d}D_\rho}{\mathrm{d}\rho}\frac{\mathrm{d}\rho}{\mathrm{d}u} = \rho\,\frac{\mathrm{d}D_\rho}{\mathrm{d}\rho}$$

and

$$\frac{\mathrm{d}^2 D_\rho}{\mathrm{d}u^2} = \frac{\mathrm{d}}{\mathrm{d}\rho}\left(\rho\,\frac{\mathrm{d}D_\rho}{\mathrm{d}\rho}\right)\frac{\mathrm{d}\rho}{\mathrm{d}u} = \left(\frac{\mathrm{d}D_\rho}{\mathrm{d}\rho} + \rho\frac{\mathrm{d}^2 D_\rho}{\mathrm{d}\rho^2}\right)\rho = \rho\,\frac{\mathrm{d}D_\rho}{\mathrm{d}\rho} + \rho^2\frac{\mathrm{d}^2 D_\rho}{\mathrm{d}\rho^2}.$$

We thereby obtain

$$\frac{\mathrm{d}^2 D_\rho}{\mathrm{d}u^2} - D_\rho = 0.$$

The ansatz

$$D_\rho \sim e^{ku}$$

leads to the characteristic equation

$$k^2 - 1 = 0 \qquad \Rightarrow k = \pm 1,$$

which yields the solutions

$$D_\rho \sim \rho \qquad \text{and} \qquad D_\rho \sim 1/\rho.$$

In region A ($0 \leq \rho < r_{\mathrm{beam}}$), the second solution would lead to a singularity at $\rho = 0$, although the charge distribution will not be singular. Therefore, we have

$$D_\rho^A = h_1^A \rho$$

and

$$D_\rho^B = h_1^B \rho + \frac{h_2^B}{\rho}.$$

We evaluate Eq. (5.57),

$$\frac{\partial D_\rho}{\partial \rho} + \frac{1}{\rho}\,D_\rho + \frac{\partial D_z}{\partial z} = \rho_q(z),$$

for the two regions A ($0 < \rho < r_{\mathrm{beam}}$) and B ($r_{\mathrm{beam}} < \rho < r_{\mathrm{bp}}$):

$$2h_1^A + \frac{\partial D_z^A}{\partial z} = \rho_{q,0}(z) = \rho_{q,0,\mathrm{DC}},$$

$$h_1^B - \frac{h_2^B}{\rho^2} + h_1^B + \frac{h_2^B}{\rho^2} + \frac{\partial D_z^B}{\partial z} = 0 \qquad \Rightarrow 2h_1^B + \frac{\partial D_z^B}{\partial z} = 0.$$

From a physical point of view, solutions for D_z that increase or decrease linearly with z may be excluded, since the fields must be periodic with respect to the synchrotron circumference. Therefore, and because of $\frac{\partial D_z}{\partial \rho} = \frac{\partial D_\rho}{\partial z} = 0$, only a constant D_z may be considered. However, D_z must vanish for $\rho = r_{\text{bp}}$. Due to the field continuity at $\rho = r_{\text{beam}}$, D_z must then be zero everywhere. Therefore, we have

$$h_1^A = \frac{\rho_{q,0,\text{DC}}}{2}, \qquad h_1^B = 0.$$

For $\rho = r_{\text{beam}}$, we obtain

$$\frac{h_2^B}{r_{\text{beam}}} = h_1^A r_{\text{beam}} = \frac{\rho_{q,0,\text{DC}}}{2} r_{\text{beam}}.$$

Therefore, we get the following solution:

$$D_\rho^A = \frac{\rho_{q,0,\text{DC}}}{2}\,\rho,$$

$$D_\rho^B = \frac{\rho_{q,0,\text{DC}} r_{\text{beam}}^2}{2\rho}$$

Due to $\lambda_{q,0,\text{DC}} = \pi r_{\text{beam}}^2\, \rho_{q,0,\text{DC}}$, we get

$$D_\rho^A = \frac{\lambda_{q,0,\text{DC}}}{2\pi r_{\text{beam}}^2}\,\rho, \tag{5.74}$$

$$D_\rho^B = \frac{\lambda_{q,0,\text{DC}}}{2\pi\rho}. \tag{5.75}$$

The general solution is the sum of the DC charge distribution result (Eqs. (5.74) and (5.75)) and the harmonic solutions (Eqs. (5.62), (5.63), (5.67), and (5.68)). Before we write down the general solution, we now assume that it belongs to a reference frame $\bar{S}$ that is the rest frame of the beam. Later, we will analyze a Lorentz transformation to the frame S that is the rest frame of the synchrotron, i.e., the laboratory frame. In the frame $\bar{S}$, we now have[7]

[7]Rewriting the equations that were derived above is accomplished by adding a bar to those quantities that belong to the frame $\bar{S}$ and by adding an index k for the different Fourier components.

$$\bar{D}_\rho^A = \frac{\bar{\lambda}_{q,0,\mathrm{DC}}}{2\pi r_{\mathrm{beam}}^2}\,\bar{\rho} + \sum_{k=1}^{\infty} \mathrm{I}_1(\bar{k}_{zk}\bar{\rho})\left(g_{1k}\ \cos(\bar{k}_{zk}\bar{z}) + g_{2k}\ \sin(\bar{k}_{zk}\bar{z})\right), \quad (5.76)$$

$$\bar{D}_z^A = \sum_{k=1}^{\infty} \zeta_k(\bar{\rho})\left(-g_{1k}\ \sin(\bar{k}_{zk}\bar{z}) + g_{2k}\ \cos(\bar{k}_{zk}\bar{z})\right), \quad (5.77)$$

$$\bar{D}_\rho^B = \frac{\bar{\lambda}_{q,0,\mathrm{DC}}}{2\pi\bar{\rho}} + \sum_{k=1}^{\infty} \left(f_{1k}\ \mathrm{I}_1(\bar{k}_{zk}\bar{\rho}) + f_{2k}\ \mathrm{K}_1(\bar{k}_{zk}\bar{\rho})\right) \cdot \cdot \left(g_{1k}\ \cos(\bar{k}_{zk}\bar{z}) + g_{2k}\ \sin(\bar{k}_{zk}\bar{z})\right), \quad (5.78)$$

$$\bar{D}_z^B = \sum_{k=1}^{\infty} \left(f_{1k}\left[\mathrm{I}_0(\bar{k}_{zk}\bar{\rho}) - \mathrm{I}_0(\bar{k}_{zk}r_{\mathrm{beam}})\right] + f_{2k}\left[\mathrm{K}_0(\bar{k}_{zk}r_{\mathrm{beam}}) - \mathrm{K}_0(\bar{k}_{zk}\bar{\rho})\right] + \zeta_k(r_{\mathrm{beam}})\right) \cdot \cdot \left(-g_{1k}\ \sin(\bar{k}_{zk}\bar{z}) + g_{2k}\ \cos(\bar{k}_{zk}\bar{z})\right). \quad (5.79)$$

Taking the Lorentz transformation into account, we see that the constant $\bar{k}_{zk}$ equals

$$\bar{k}_{zk} = k\frac{2\pi}{\bar{l}_{\mathrm{R}}} = k\frac{2\pi\gamma}{l_{\mathrm{R}}}.$$

It is assumed that the charge density distribution is given by Eq. (5.83) below, so that the constants g_{1k}, g_{2k} are known.

For each k, the three constants $f_{1k}, f_{2k}, \zeta_k(r_{\mathrm{beam}})$ that were introduced above and two further constants $\zeta_k(0), \xi_k$ can then be determined by solving the linear system of equations that consists of Eqs. (5.80), (5.81), (5.82), (5.84), and finally Eq. (5.85) for the specific value $\bar{\rho} = r_{\mathrm{beam}}$.

The first of these five equations is a rewritten form of Eq. (5.69),

$$\mathrm{I}_1(\bar{k}_{zk}r_{\mathrm{beam}}) = f_{1k}\ \mathrm{I}_1(\bar{k}_{zk}r_{\mathrm{beam}}) + f_{2k}\ \mathrm{K}_1(\bar{k}_{zk}r_{\mathrm{beam}}), \quad (5.80)$$

the second of Eq. (5.70),

$$f_{1k}\left[\mathrm{I}_0(\bar{k}_{zk}r_{\mathrm{bp}}) - \mathrm{I}_0(\bar{k}_{zk}r_{\mathrm{beam}})\right] + f_{2k}\left[\mathrm{K}_0(\bar{k}_{zk}r_{\mathrm{beam}}) - \mathrm{K}_0(\bar{k}_{zk}r_{\mathrm{bp}})\right] + \zeta_k(r_{\mathrm{beam}}) = 0, \quad (5.81)$$

Please note that the upper indices A and B for the constants f_{1k}, f_{2k}, g_{1k}, g_{2k} are no longer necessary, since they are unique.

and the third of Eq. (5.73),

$$f_{1k}\ \mathrm{I}_1(\bar{k}_{zk} r_{\mathrm{bp}}) + f_{2k}\ \mathrm{K}_1(\bar{k}_{zk} r_{\mathrm{bp}}) = \frac{r_{\mathrm{beam}}^2 \xi_k}{2 r_{\mathrm{bp}}}. \tag{5.82}$$

From Eq. (5.71), we obtain, for the charge density in region A,

$$\bar{\rho}_q = \bar{\rho}_{q,0,\mathrm{DC}} + \sum_{k=1}^{\infty} \xi_k \left(g_{1k}\ \cos(\bar{k}_{zk}\bar{z}) + g_{2k}\ \sin(\bar{k}_{zk}\bar{z}) \right), \tag{5.83}$$

with

$$\xi_k = \bar{k}_{zk} \left(\mathrm{I}_0(\bar{k}_{zk}\bar{\rho}) - \zeta_k(\bar{\rho}) \right)$$

constant, so that this expression may be evaluated for different values of $\bar{\rho}$, e.g., $\bar{\rho} = 0$:

$$\xi_k = \bar{k}_{zk} \left(1 - \zeta_k(0)\right). \tag{5.84}$$

Equation (5.64) now reads

$$\zeta_k(\bar{\rho}) = \mathrm{I}_0(\bar{k}_{zk}\bar{\rho}) - 1 + \zeta_k(0). \tag{5.85}$$

5.8.2 *Transformation to the Rest Frame of the Synchrotron*

For the Lorentz transformation, the following formulas are valid in our specific case (cf. Eqs. (2.56)–(2.59), (2.60)–(2.63), (2.69)–(2.74)):

$$\rho = \bar{\rho}, \qquad \bar{z} = \gamma(z - vt), \qquad \bar{k}_{zk}\bar{z} = \gamma \bar{k}_{zk} z - \gamma \bar{k}_{zk} vt =: \phi(z,t),$$

$$D_\rho^A = \gamma \bar{D}_\rho^A, \qquad D_\rho^B = \gamma \bar{D}_\rho^B,$$

$$D_z^A = \bar{D}_z^A, \qquad D_z^B = \bar{D}_z^B,$$

$$H_\varphi^A = \gamma v \bar{D}_\rho^A, \qquad H_\varphi^B = \gamma v \bar{D}_\rho^B,$$

$$J_z = \gamma v\ \bar{\rho}_q,$$

$$\rho_q = \gamma\ \bar{\rho}_q,$$

$$\int \lambda_q\ \mathrm{d}z = \int \bar{\lambda}_q\ \mathrm{d}\bar{z} \qquad \Rightarrow \bar{\lambda}_q = \frac{1}{\gamma}\ \lambda_q.$$

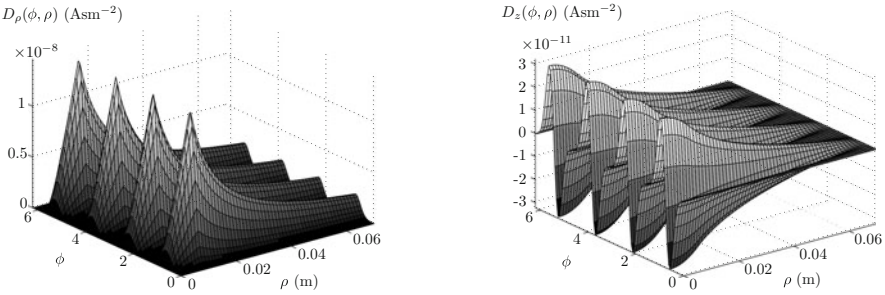

Fig. 5.11 Electric displacement field $\vec{D}$ for the line charge density given in Fig. 5.12 (parameters: $l_\mathrm{R} = 216\,\mathrm{m}$, $r_\mathrm{bp} = 7\,\mathrm{cm}$, $r_\mathrm{beam} = 1\,\mathrm{cm}$)

Therefore, we obtain

$$D_\rho^A = \frac{\lambda_{q,0,\mathrm{DC}}}{2\pi r_\mathrm{beam}^2}\,\rho + \gamma \sum_{k=1}^{\infty} \mathrm{I}_1(\bar{k}_{zk}\rho)\,(g_{1k}\ \cos\phi + g_{2k}\ \sin\phi)\,, \tag{5.86}$$

$$D_z^A = \sum_{k=1}^{\infty} \zeta_k(\rho)\,(-g_{1k}\ \sin\phi + g_{2k}\ \cos\phi)\,, \tag{5.87}$$

$$D_\rho^B = \frac{\lambda_{q,0,\mathrm{DC}}}{2\pi\rho} + \gamma \sum_{k=1}^{\infty} \left(f_{1k}\ \mathrm{I}_1(\bar{k}_{zk}\rho) + f_{2k}\ \mathrm{K}_1(\bar{k}_{zk}\rho)\right) \cdot \\ \cdot\,(g_{1k}\ \cos\phi + g_{2k}\ \sin\phi)\,, \tag{5.88}$$

$$D_z^B = \sum_{k=1}^{\infty} \Big(f_{1k}\left[\mathrm{I}_0(\bar{k}_{zk}\rho) - \mathrm{I}_0(\bar{k}_{zk}r_\mathrm{beam})\right] \\ + f_{2k}\left[\mathrm{K}_0(\bar{k}_{zk}r_\mathrm{beam}) - \mathrm{K}_0(\bar{k}_{zk}\rho)\right] + \zeta_k(r_\mathrm{beam})\Big) \cdot \\ \cdot\ (-g_{1k}\ \sin\phi + g_{2k}\ \cos\phi)\,, \tag{5.89}$$

$$\rho_q = \rho_{q,0,\mathrm{DC}} + \gamma \sum_{k=1}^{\infty} \xi_k\,(g_{1k}\ \cos\phi + g_{2k}\ \sin\phi)\,,$$

$$\dot{\rho}_q = -\gamma^2 v \sum_{k=1}^{\infty} \bar{k}_{zk}\xi_k\,(-g_{1k}\ \sin\phi + g_{2k}\ \cos\phi)\,.$$

Figure 5.11 presents an example for the field $\vec{D}$ that is obtained if a line charge density according to Fig. 5.12 is assumed.

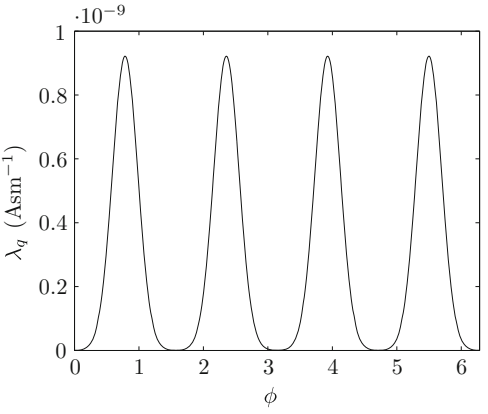

Fig. 5.12 Line charge density $\lambda_q = \pi r_{\text{beam}}^2 \rho_q$ with an average beam current of $\bar{I}_{\text{beam}} = 13.6\,\text{mA}$ for $\beta = 0.15583$, and $\hat{I}_{\text{beam}} = 20.0\,\text{mA}$ at $h = 4$

5.8.3 Longitudinal Electric Field

Now we introduce some approximations. We assume that only a few Fourier components, say 10, are needed to describe the beam current. Furthermore, we assume that the maximum energy is below $\gamma = 6$. Therefore, the maximum $\bar{k}_{zk}$ is equal to

$$\bar{k}_{zk} = k\frac{2\pi}{\bar{l}_{\text{R}}} = k\frac{2\pi\gamma}{l_{\text{R}}},$$

with $k = 10$ and $\gamma = 6$. Here, l_{R} is the circumference of the synchrotron, for which a length contraction

$$\bar{l}_{\text{R}} = \frac{l_{\text{R}}}{\gamma}$$

has to be taken into account in the rest frame $\bar{S}$ of the beam. If we now assume that the beam pipe radius r_{bp} is smaller than 10 cm whereas the synchrotron circumference l_{R} is larger than 200 m, we get

$$\bar{k}_{zk} r_{\text{bp}} < 0.1885.$$

Therefore, we will assume

$$\bar{k}_{zk} r_{\text{bp}} < 0.2$$

in the following. The specific numbers may, of course, be modified for different cases.

For small arguments z, one obtains[8]

$$\boxed{\mathrm{I}_0(z) \approx 1 + \frac{z^2}{4}, \qquad \mathrm{K}_0(z) \approx -\ln z, \qquad \mathrm{I}_1(z) \approx \frac{z}{2}, \qquad \mathrm{K}_1(z) \approx \frac{1}{z}.}$$

We use these approximations to simplify Eqs. (5.80)–(5.82):

$$\frac{\bar{k}_{zk} r_{\mathrm{beam}}}{2} = f_{1k} \, \frac{\bar{k}_{zk} r_{\mathrm{beam}}}{2} + f_{2k} \, \frac{1}{\bar{k}_{zk} r_{\mathrm{beam}}}, \tag{5.90}$$

$$f_{1k} \, \frac{\bar{k}_{zk}^2 (r_{\mathrm{bp}}^2 - r_{\mathrm{beam}}^2)}{4} + f_{2k} \, \ln \frac{r_{\mathrm{bp}}}{r_{\mathrm{beam}}} + \frac{\bar{k}_{zk}^2 r_{\mathrm{beam}}^2}{4} + \zeta_k(0) = 0, \tag{5.91}$$

$$f_{1k} \, \frac{\bar{k}_{zk} r_{\mathrm{bp}}}{2} + f_{2k} \, \frac{1}{\bar{k}_{zk} r_{\mathrm{bp}}} = \frac{r_{\mathrm{beam}}^2 \xi_k}{2 r_{\mathrm{bp}}} = \frac{r_{\mathrm{beam}}^2}{2 r_{\mathrm{bp}}} \, \bar{k}_{zk} \, (1 - \zeta_k(0)) \, . \tag{5.92}$$

In the last two equations, we used Eqs. (5.84) and (5.85). The first of these three equations leads to

$$f_{1k} = 1 - f_{2k} \, \frac{2}{(\bar{k}_{zk} r_{\mathrm{beam}})^2}. \tag{5.93}$$

If we insert this into Eq. (5.91), we get

$$\frac{\bar{k}_{zk}^2 r_{\mathrm{bp}}^2}{4} + f_{2k} \left(\ln \frac{r_{\mathrm{bp}}}{r_{\mathrm{beam}}} - \frac{1}{2} \frac{r_{\mathrm{bp}}^2}{r_{\mathrm{beam}}^2} + \frac{1}{2} \right) + \zeta_k(0) = 0. \tag{5.94}$$

If one inserts Eq. (5.93) into Eq. (5.92), one obtains

$$\frac{\bar{k}_{zk} r_{\mathrm{bp}}}{2} + \frac{f_{2k}}{\bar{k}_{zk}} \left(\frac{1}{r_{\mathrm{bp}}} - \frac{r_{\mathrm{bp}}}{r_{\mathrm{beam}}^2} \right) = \frac{r_{\mathrm{beam}}^2}{2 r_{\mathrm{bp}}} \, \bar{k}_{zk} \, (1 - \zeta_k(0))$$

$$\Rightarrow \; f_{2k} = -\frac{\bar{k}_{zk}^2 r_{\mathrm{beam}}^2 r_{\mathrm{bp}}^2}{2(r_{\mathrm{beam}}^2 - r_{\mathrm{bp}}^2)} + \frac{r_{\mathrm{beam}}^4}{2(r_{\mathrm{beam}}^2 - r_{\mathrm{bp}}^2)} \, \bar{k}_{zk}^2 \, (1 - \zeta_k(0)) \, .$$

This may be inserted into Eq. (5.94):

$$\frac{\bar{k}_{zk}^2 r_{\mathrm{bp}}^2}{4} + \left[\frac{\bar{k}_{zk}^2 r_{\mathrm{beam}}^2 r_{\mathrm{bp}}^2}{2(r_{\mathrm{bp}}^2 - r_{\mathrm{beam}}^2)} - \frac{\bar{k}_{zk}^2 r_{\mathrm{beam}}^4}{2(r_{\mathrm{bp}}^2 - r_{\mathrm{beam}}^2)} \, (1 - \zeta_k(0)) \right] \left(\ln \frac{r_{\mathrm{bp}}}{r_{\mathrm{beam}}} - \frac{1}{2} \frac{r_{\mathrm{bp}}^2}{r_{\mathrm{beam}}^2} + \frac{1}{2} \right) + \zeta_k(0) = 0$$

[8] For $0 \le z \le 0.2$, the error is below $3 \cdot 10^{-5}$ for $\mathrm{I}_0(z)$, below 8.2% for $\mathrm{K}_0(z)$, below 0.5% for $\mathrm{I}_1(z)$, and below 4.7% for $\mathrm{K}_1(z)$.

$$\Rightarrow \zeta_k(0)\left[1+\frac{\bar{k}_{zk}^2 r_{\text{beam}}^4}{2(r_{\text{bp}}^2-r_{\text{beam}}^2)}\left(\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}-\frac{1}{2}\frac{r_{\text{bp}}^2}{r_{\text{beam}}^2}+\frac{1}{2}\right)\right]=$$

$$=\bar{k}_{zk}^2\left[-\frac{r_{\text{bp}}^2}{4}+\frac{r_{\text{beam}}^4-r_{\text{beam}}^2 r_{\text{bp}}^2}{2(r_{\text{bp}}^2-r_{\text{beam}}^2)}\left(\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}-\frac{1}{2}\frac{r_{\text{bp}}^2}{r_{\text{beam}}^2}+\frac{1}{2}\right)\right]$$

$$=\bar{k}_{zk}^2\left[-\frac{r_{\text{bp}}^2}{4}-\frac{r_{\text{beam}}^2}{2}\left(\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}-\frac{1}{2}\frac{r_{\text{bp}}^2}{r_{\text{beam}}^2}+\frac{1}{2}\right)\right]$$

$$=\bar{k}_{zk}^2\left[-\frac{r_{\text{beam}}^2}{2}\left(\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}+\frac{1}{2}\right)\right]$$

$$\Rightarrow \zeta_k(0)=-\bar{k}_{zk}^2\,\frac{r_{\text{beam}}^2\left(\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}+\frac{1}{2}\right)}{2+(\bar{k}_{zk}r_{\text{beam}})^2\frac{\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}-\frac{1}{2}\frac{r_{\text{bp}}^2}{r_{\text{beam}}^2}+\frac{1}{2}}{\frac{r_{\text{bp}}^2}{r_{\text{beam}}^2}-1}}.$$

For $\frac{r_{\text{bp}}}{r_{\text{beam}}} > 1$, the fraction in the denominator is in the range $]-0.5, 0[$. If we also take into account that $\bar{k}_{zk}r_{\text{beam}} < \bar{k}_{zk}r_{\text{bp}} < 0.2$ holds, we see that this part in the denominator can be neglected in comparison with 2:

$$\zeta_k(0)=-\frac{\bar{k}_{zk}^2 r_{\text{beam}}^2}{2}\left(\frac{1}{2}+\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}\right)=-\frac{\bar{k}_{zk}^2 r_{\text{bp}}^2}{2}\left[\frac{r_{\text{beam}}^2}{r_{\text{bp}}^2}\left(\frac{1}{2}+\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}\right)\right]. \tag{5.95}$$

Therefore, according to Eq. (5.87), the longitudinal field on the beam axis is

$$D_z|_{\rho=0}=-\frac{r_{\text{beam}}^2}{2}\left(\frac{1}{2}+\ln\frac{r_{\text{bp}}}{r_{\text{beam}}}\right)\sum_{k=1}^{\infty}\bar{k}_{zk}^2\left(-g_{1k}\ \sin\phi+g_{2k}\ \cos\phi\right). \tag{5.96}$$

By means of Eq. (5.83), we obtain

$$\frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}}=\pi r_{\text{beam}}^2\frac{\mathrm{d}\bar{\rho}_q}{\mathrm{d}\bar{z}}=\pi r_{\text{beam}}^2\sum_{k=1}^{\infty}\bar{k}_{zk}\,\xi_k\left(-g_{1k}\ \sin(\bar{k}_{zk}\bar{z})+g_{2k}\ \cos(\bar{k}_{zk}\bar{z})\right). \tag{5.97}$$

Due to $\frac{r_{\text{bp}}}{r_{\text{beam}}} > 1$, the expression in Eq. (5.95) in square brackets is in the range $]0, 0.5[$. With $\bar{k}_{zk}r_{\text{bp}} < 0.2$, one sees that

$$|\zeta_k(0)| \ll 1,$$

so that

$$\xi_k \approx \bar{k}_{zk}$$

is valid according to Eq. (5.84). Therefore, the sums in the two equations (5.96) and (5.97) are identical. Hence, we obtain

$$D_z|_{\rho=0} = -\frac{1}{2\pi}\left(\frac{1}{2} + \ln\frac{r_{\mathrm{bp}}}{r_{\mathrm{beam}}}\right)\frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}}.$$

According to the Lorentz transformation,

$$\frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}} = \frac{1}{\gamma^2}\frac{\partial\lambda_q}{\partial z}$$

is valid, and we finally obtain

$$D_z|_{\rho=0} = -\frac{1}{4\pi\gamma^2}\left(1 + 2\,\ln\frac{r_{\mathrm{bp}}}{r_{\mathrm{beam}}}\right)\frac{\partial\lambda_q}{\partial z} = -\frac{g_0}{4\pi\gamma^2}\frac{\partial\lambda_q}{\partial z}$$

with the **geometry factor**

$$\boxed{g_0 = 1 + 2\,\ln\frac{r_{\mathrm{bp}}}{r_{\mathrm{beam}}}.}$$

The result

$$\boxed{E_z|_{\rho=0} = -\frac{g_0}{4\pi\epsilon_0\gamma^2}\frac{\partial\lambda_q}{\partial z}} \tag{5.98}$$

may also be found in Edwards/Syphers [27] as formula (6.33) and in Ng [28] as formula (2.42). In these references, a much simpler derivation is offered, which, however, requires some advance knowledge about the field. Please note that in our derivation and in [27], λ_q denotes the charge density. Sometimes, $\lambda_{q,\mathrm{norm}}$ is defined as a normalized density function for the bunch, so that the total bunch charge $N_b Q$ has to be added explicitly as a factor:

$$\lambda_q = N_b Q \lambda_{q,\mathrm{norm}}.$$

Finally, it has to be emphasized that the derivation presented here is based on the simplest model with transversally constant charge density. If different models are used, one also obtains different expressions for the geometry factor. This is discussed, for example, in Reiser [29, Sect. 6.3.2], and in Zotter [23, Sect. 12.1.1].

5.8.4 Space Charge Impedance

We determine the current at a constant position z:

$$I_{\text{beam}} = \int \vec{J} \cdot \mathrm{d}\vec{A} = \pi r_{\text{beam}}^2 J_z = \pi r_{\text{beam}}^2 \gamma v \bar{\rho}_q = \gamma v \bar{\lambda}_q .$$

In $\bar{S}$, the line charge density $\bar{\lambda}_q$ depends only on $\bar{z}$ and not on time, since it determines the static solution. By means of the Lorentz transformation, however, $\bar{z} = \gamma(z - vt)$ depends on z and on t in the laboratory frame S. Therefore, the time derivative is

$$\frac{\partial I_{\text{beam}}}{\partial t} = \gamma v \frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}} \frac{\partial \bar{z}}{\partial t} = -\gamma^2 v^2 \left. \frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}} \right|_{\bar{z}=\gamma(z-vt)} . \tag{5.99}$$

Now we calculate the voltage that is seen by a specific particle of the bunch. In the rest frame $\bar{S}$ of the bunch, we have

$$\bar{E}_z|_{\rho=0} = E_z|_{\rho=0} = -\frac{g_0}{4\pi\epsilon_0} \frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}} .$$

Since the particle is moving with the bunch, the derivative $\frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}}$ is always constant.[9] Therefore, the integration is simple. We have only to multiply this expression by l_{R} to get the voltage in S:

$$V = \int \vec{E} \cdot \mathrm{d}\vec{r} = \int E_z \, \mathrm{d}z = -\frac{g_0 l_{\text{R}}}{4\pi\epsilon_0} \frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}} . \tag{5.100}$$

The expression $\frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}}$ has to be evaluated at the position of the specific particle. In the frame S, it will be given by $z = z_0 + vt$, so that

$$\bar{z}_0 = \bar{z} = \gamma z_0$$

is obtained. If the beam current is also measured at $z = z_0 + vt$, we may combine Eqs. (5.99) and (5.100) and obtain

$$V = -\frac{g_0 l_{\text{R}}}{4\pi\epsilon_0} \frac{1}{-\gamma^2 v^2} \frac{\partial I_{\text{beam}}}{\partial t} = \frac{g_0 l_{\text{R}}}{4\pi\epsilon_0 \gamma^2 v^2} \frac{\partial I_{\text{beam}}}{\partial t} .$$

[9] Please note that the synchrotron oscillation period T_{S} is much larger than the revolution time T_{R}.

If this equation is transformed into the frequency domain, the time derivative will be converted into a multiplication by $j\omega$:

$$\underline{\hat{V}}(\omega) = j\omega \frac{g_0 l_\mathrm{R}}{4\pi\epsilon_0 \gamma^2 v^2}\, \underline{\hat{I}}_\mathrm{beam}(\omega).$$

According to our standard equivalent circuit in Fig. 4.5 on p. 181, we have to define the longitudinal space charge impedance as

$$Z_\mathrm{sc} = -\frac{\underline{\hat{V}}(\omega)}{\underline{\hat{I}}_\mathrm{beam}(\omega)}$$

if we want to treat it similarly to the impedance of a cavity. Hence we get

$$Z_\mathrm{sc} = -j\omega \frac{g_0 l_\mathrm{R}}{4\pi\epsilon_0 \gamma^2 v^2}. \tag{5.101}$$

If we evaluate this expression at the nth harmonic of the revolution frequency, i.e., at

$$\omega = n\omega_\mathrm{R} = \frac{2\pi n}{T_\mathrm{R}} = \frac{2\pi n v}{l_\mathrm{R}},$$

we obtain

$$\boxed{\frac{Z_\mathrm{sc}}{n} = -j\,\frac{g_0}{2\epsilon_0 \gamma^2 v} = -j\,\frac{g_0 Z_0}{2\gamma^2 \beta}.} \tag{5.102}$$

Here we used

$$c_0 = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \qquad \text{and} \qquad Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 376.73\ \Omega,$$

respectively the velocity of light and the impedance of free space. According to the sign in Eq. (5.102), the space charge impedance is capacitive, although the frequency dependence corresponds to an inductance. Formula (5.102) can be found in many publications and textbooks on accelerator physics (cf. Hofmann/Pedersen [30, eqn. (8)]; Edwards/Syphers [27, eqn. (6.50)]; Zotter [23, eqn. (12.1)]; Ng [28, eqn. (2.45)]; Chao/Tigner [8, Sect. 2.5.3.1, p. 128, formula (1)]; Reiser [29, Sect. 6.3.3, eqn. (6.114)]). In many cases, however, the definition of the sign is reversed. Instead of the impedance itself, one may also consider the impedance per length. In this sense, Eq. (5.101), e.g., corresponds to equation (6.94b) in Reiser [29].

According to Eq. (5.98), the electric field induced by the space charge is proportional to $-\frac{\partial \lambda_q}{\partial z}$. In terms of time, it is therefore proportional to $+\frac{\partial \lambda_q}{\partial t}$. This is due to the fact that we have

$$\frac{\partial \bar{\lambda}_q}{\partial z} = \frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}} \frac{\partial \bar{z}}{\partial z} = \gamma \frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}}$$

and

$$\frac{\partial \bar{\lambda}_q}{\partial t} = \frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}} \frac{\partial \bar{z}}{\partial t} = -\gamma v \frac{\mathrm{d}\bar{\lambda}_q}{\mathrm{d}\bar{z}}.$$

From a physical point of view, this change of sign is also obvious, since the head of the bunch is located at larger z than the tail of the bunch. With respect to time, the head of the bunch will reach a certain point earlier than the tail of the bunch. We now have a look at the four cases shown in Fig. 3.3 on p. 133. If we draw the electric field that is generated by the space charge for each of these four cases, we get rising slopes in the bunch center for the right-hand diagrams and falling slopes in the left-hand diagrams. This corresponds to a defocusing effect in the upper two diagrams, since the slope is opposite to that of $V(t)$, and to a focusing effect in the lower two diagrams, since the slope is the same as that of $V(t)$. Therefore, the longitudinal space charge leads to a defocusing effect below transition. Above transition, it may lead to a beam instability (cf. Edwards [27, Sect. 6.2]; Reiser [29, Sect. 6.3.3]).

We now take GSI's synchrotron SIS18 at injection energy (positive charges below transition) as an example. Let us consider the stationary case. Under these conditions, we have

$$\beta = 0.15503, \qquad \gamma = 1.012238,$$

and we are interested in the Fourier component of the space charge voltage at $n = h = 4$. We assume

$$\frac{r_{\mathrm{bp}}}{r_{\mathrm{beam}}} = 7 \qquad \Rightarrow g_0 = 4.89,$$

which leads to a space charge impedance of

$$Z_{\mathrm{sc}} = -j \cdot 23\, k\Omega$$

if Eq. (5.102) is evaluated. If we now assume a beam current amplitude of

$$\underline{\hat{I}}_{\text{beam}} = 20\,\text{mA}$$

at the RF frequency ($h = 4$), this results in a space charge voltage

$$\underline{\hat{V}} = -Z_{\text{sc}}\,\underline{\hat{I}}_{\text{beam}} = +j \cdot 460\,\text{V}.$$

A beam current

$$I_{\text{beam}}(t) = 20\,\text{mA} \cdot \cos(\omega_{\text{RF}} t)$$

therefore leads to a voltage

$$V(t) = 460\,\text{V} \cdot \cos(\omega_{\text{RF}} t + \pi/2) = -460\,\text{V} \cdot \sin(\omega_{\text{RF}} t).$$

This is actually defocusing, because the RF voltage must be proportional to $+\sin(\omega_{\text{RF}} t)$ to keep the particles bunched.

References

1. S.Y. Lee, *Accelerator Physics* (World Scientific, Singapore/New Jersey/London/Hong Kong, 1999)
2. H. Goldstein, C. Poole, J. Safko, *Classical Mechanics*, 3rd edn. (Pearson Education, Addison-Wesley, June 25, 2001). ISBN-10: 0201657023, ISBN-13: 978-0201657029
3. E. Schmutzer, *Grundlagen der Theoretischen Physik* (Deutscher Verlag der Wissenschaften, Berlin, 1989)
4. E. Kamke, *Differentialgleichungen, Lösungsmethoden und Lösungen* (Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1956)
5. R.S. MacKay, J.D. Meiss, *Hamiltonian Dynamical Systems* (Adam Hilger, Bristol/Philadelphia, 1987)
6. V.I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd edn. (Springer, New York/Berlin/Heidelberg/Hong Kong/London/Milan/Paris/Tokyo, 1989)
7. V.I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer, New York/Berlin/Heidelberg, 1988)
8. A.W. Chao, M. Tigner, *Handbook of Accelerator Physics and Engineering*, 3rd edn. (World Scientific, New Jersey/London/Singapore/Beijing/Shanghai/Hong Kong/Taipei/Chennai, 2006)
9. A. Sommerfeld, *Atomic Structure and Spectral Lines*, vol. 1, 3rd edn. (Methuen&Co., London, 1934)
10. A. Sommerfeld, *Atombau und Spektrallinien*, Band 1 (Verlag Harri Deutsch, Thun, 1978)
11. R. Garoby, RF gymnastics in synchrotrons, in *CAS - CERN Accelerator School: RF for Accelerators*, Ebeltoft, 8–17 Jun 2010, pp. 431–446
12. H.G. Hereward, Landau damping, in *CAS - CERN Accelerator School: Accelerator Physics*, Oxford, 16–27 Sep 1985, pp. 255–263
13. E.J.N. Wilson, *An Introduction to Particle Accelerators* (Oxford University Press, Oxford/New York, 2001)

14. G. Schreiber, Barrier-Buckets am Experimentierspeicherring der Gesellschaft für Schwerionenforschung. Dissertation, Universität Frankfurt, 2006
15. R. Stassen, K. Bongardt, F.J. Etzkorn, H. Stockhorst, S. Papureanu, A. Schnase, The HESR RF-system and tests in COSY, in *EPAC* (2008)
16. H. Klingbeil, D. Lens, M. Mehler, B. Zipfel, Modeling longitudinal oscillations of bunched beams in synchrotrons (November 2010). arXiv:1011.3957v1 [physics.acc-ph]
17. D. Lens, *Modeling and Control of Longitudinal Single-Bunch Oscillations in Heavy-Ion Synchrotrons*. Fortschrittberichte VDI, Reihe 8, Mess-, Steuerungs- und Regelungstechnik; Nr. 1209; Dissertation, Technische Universität Darmstadt, 2012
18. J.-J.E. Slotine, W. Li, *Applied Nonlinear Control* (Prentice Hall, Englewood Cliffs, 1991)
19. P.M. Lapostolle, Possible emittance increase through filamentation due to space charge in continuous beams. IEEE Trans. Nucl. Sci. **NS-18**(3), 1101–1104 (1971)
20. F.J. Sacherer, RMS envelope equations with space charge. IEEE Trans. Nucl. Sci. **NS-18**(3), 1105–1107 (1971)
21. J. Buon, Beam phase space and emittance, in *CAS - CERN Accelerator School: 4th General Accelerator Physics Course*, Jülich, 17–28 Sep 1990, pp. 30–52
22. F.J. Sacherer, A longitudinal stability criterion for bunched beams, in *5th IEEE Particle Accelerator Conference*, San Francisco, 5–7 Mar 1973, pp. 825–829
23. B.W. Zotter, S.A. Kheifets, *Impedances and Wakes in High-Energy Particle Accelerators* (World Scientific, Singapore/New Jersey/London/Hong Kong, 1998)
24. S. Khan, *Collective Phenomena in Synchrotron Radiation Sources* (Springer, Berlin/Heidelberg/New York, 2006)
25. I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products*, 6th edn. (Academic, San Diego/San Francisco/New York/Boston/London/Sydney/Tokyo, 2000)
26. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965)
27. D.A. Edwards, M.J. Syphers, *An Introduction to the Physics of High Energy Accelerators* (Wiley-VCH Verlag GmbH&Co.KGaA, Weinheim, 2004)
28. K.Y. Ng, *Physics of Intensity Dependent Beam Instabilities* (World Scientific, New Jersey/London/Singapore/Beijing/Shanghai/Hong Kong/Taipei/Chennai, 2006)
29. M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New York/Chichester/Brisbane/Toronto/Singapore, 1994)
30. A. Hofmann, F. Pedersen, Bunches with local elliptic energy distributions. IEEE Trans. Nucl. Sci. **NS-26**(3), 3526–3528 (1979)

Chapter 6
Power Amplifiers

In this chapter, power amplifiers for synchrotron and storage ring RF systems are discussed. The main task of the power amplifier is to provide RF power to the synchrotron cavity. This is done by conversion from DC to RF power. The DC power is drawn from the line power by means of power supplies that are responsible for providing all voltages and currents that are required to operate the power amplifier.

A variety of devices are used in power amplifiers for particle accelerators [1,2]:

- **Gridded vacuum tubes** (e.g., triodes, tetrodes) [3,4]
- **Inductive output tubes, IOTs** (klystrode) [5]
- **Klystrons** [6]
- **Solid state amplifiers**

The first three types (gridded tubes, inductive output tubes, and klystrons) are **vacuum tubes**. These devices thermionically liberate electrons from a cathode to create a high-current electron beam, which is accelerated by a high DC voltage toward the anode (collector). The beam always travels within a vacuum to reduce interaction with rest gas. After emission, the beam is bunched. If gridded tubes and inductive output tubes are used, this bunching is done by a fine grid in front of the cathode. The voltage of this grid is modulated with the desired frequency, thereby directly modulating the current through the tube. In a klystron, this bunching is done by modulating the voltage in one or typically several buncher cavities, thereby bunching the coasting beam emitted from the gun in flight.

For gridded tubes, the voltage of the anode is modulated by the impacting beam, and the RF current is directly drawn from there. If an inductive output tube and klystron are used, then the modulated electron beam travels through a catcher cavity in front of the collector, where it excites an electromagnetic field. The power from this catcher cavity is extracted by a coupler and can be delivered to the load.

Solid-state amplifiers consist of a large number of transistors operating in parallel. Due to the reduced mobility of the electrons within the transistor compared to vacuum, the output power of solid-state amplifiers at high frequencies is rather

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6_6

limited compared to vacuum tubes, usually restricting them to medium power applications such as driver amplifiers (even though their range of application is being extended continually).

The three different types of vacuum tubes are used in different applications with respect to particle accelerators. The factors that are most important in determining the choice are the operating frequency of the power amplifier and whether it is a narrowband or a broadband application. Generally, gridded tubes are favorable at frequencies below a few hundred megahertz. At high frequencies, gridded tubes are limited by time-of-flight effects. Tubes usually have a high bandwidth, since they do not exhibit any low-frequency limitations. Klystrons, on the other hand, are unsuited for low-frequency operation, since their dimensions become impractical in this case. In addition, the frequency band of klystrons is rather limited, due to the bandwidth of the cavities it uses. Inductive output tubes fall between the operating frequencies of gridded tubes and klystrons. Summarizing, gridded tubes are favorable in applications below 100 MHz, especially if broadband behavior is required. Klystrons are optimal for high-frequency narrowband applications. In general, hadron synchrotron and storage ring RF systems fall in the former category (low frequency, broadband), and therefore gridded vacuum tubes are usually the best choice for a power amplifier for this type of RF system. Tubes are also much less sensitive than semiconductors to a high radiation dose environment. Therefore, the following section will focus on power amplifiers using gridded vacuum tubes.

6.1 Gridded Vacuum Tubes

In this section, different types of gridded vacuum tubes are discussed. More detailed information may be found, for example, in [3], which also influenced the presentation in this book.

6.1.1 Diode

The simplest type of vacuum tube is a **diode**, as shown in Fig. 6.1. A vacuum diode features two **electrodes**: a **cathode** and an **anode** (also called **plate**) in a vacuum environment. A cylindrical arrangement with an outer anode cylinder enclosing an inner cathode cylinder is usually used. This assembly is enclosed in a compartment made of metal or glass with several feedthroughs to connect the electrodes. The cathode is heated by means of an electric current to initiate the **thermionic emission** of electrons, thereby introducing free charge carriers into the vacuum. When one now applies a positive voltage to the anode with respect to the cathode, these electrons may travel through the diode, creating a negative current flowing from the cathode to the anode.

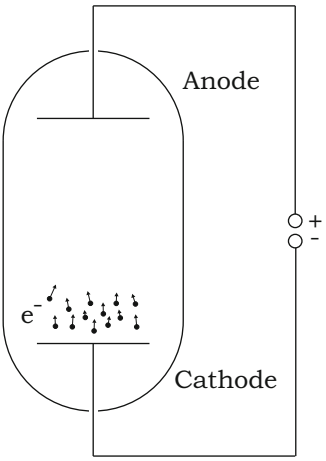

Fig. 6.1 Schematic view of a diode

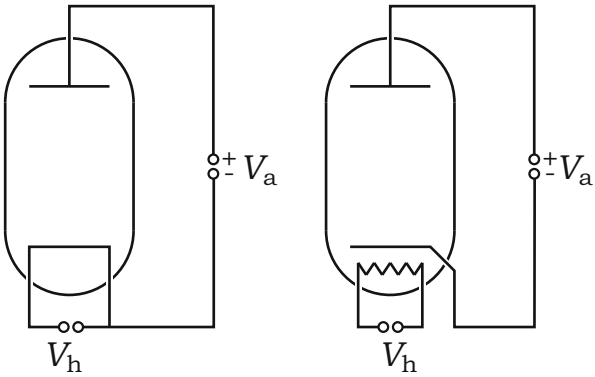

Fig. 6.2 Directly heated cathode (*left diagram*) and indirectly heated cathode (*right diagram*)

As we will see later, this current is dependent on the type of vacuum tube (geometry, material selection), the temperature of the cathode, the vacuum pressure, and the voltage between anode and cathode.

Reversing the polarity of both electrodes results in no current flowing through the device, since the anode is not heated. No electrons will be emitted from the anode that can travel through the tube. On the other hand, all electrons emitted from the cathode will also not be able to pass through the tube, due to the voltage barrier introduced by the anode. The transport of (negative) current through a vacuum tube is possible from only the cathode to the anode. Therefore, a vacuum diode may be used for rectification purposes, similar to semiconductor diodes.

The heating of the cathode can be done in two different ways. For **direct heating**, as shown in the left-hand diagram of Fig. 6.2, the current used to increase the temperature of the cathode travels directly through the cathode. For **indirect heating** (see the right-hand diagram of Fig. 6.2), a dedicated **filament** allows the heating of the cathode. In addition to this, the heating may be performed by DC or AC.

A directly heated cathode with one side of the cathode at ground potential and the other side at the heating potential exhibits a voltage drop over its length. This voltage

drop leads to different voltages between different parts of the cathode and the anode, thereby influencing the current traveling through the tube. To reduce this effect, as well as the self-heating of the cathode by the emitted electrons, the resistance of the cathode is usually chosen to be fairly low. Nevertheless, in case of AC heating, a modulation of the tube current with the frequency of the heating current (usually 50 Hz) is observed, and a correction may be necessary, depending on the application. In an indirect heating scheme, the whole cathode remains at an almost constant potential. In addition to this, the total area of an indirectly heated cathode is usually much larger than that of the wire configuration of a directly heated one. AC heating without any corrections is applied whenever possible. On the other hand, the required heating power is much larger than in the direct heating scheme, due to the larger volume that is kept at high temperature. Especially for high-power tubes, this may become a limiting factor. It will also take much longer for an indirectly heated cathode to reach its operating temperature. The temperature and therefore the current density that can be emitted from an indirectly heated cathode is smaller than that of a directly heated one.

In the following, the current density emitted by such a heated cathode will be discussed. Electrons within the cathode are bound to the metal. To release them into the vacuum, they have to overcome the binding energy of the metal (electron work function), which is usually on the order of a few electron volts. The electrons inside the metal exhibit different energies; they follow a band structure with a partly occupied conduction band.

The emitted current density is calculated assuming an infinite plane surface. The emission takes place perpendicular to this surface, along the z-axis.

The density of states n_{stat} inside the metal is given by

$$\mathrm{d}n_{\mathrm{stat}} = 4\pi \left(\frac{2m_{\mathrm{e}}}{h^2}\right)^{3/2} \sqrt{W}\,\mathrm{d}W,$$

with the electron mass m_{e}, **Planck's constant** h, and the energy W of the observed state. The density of occupied states $\mathrm{d}n$ is given as the product of the density of states $\mathrm{d}n_{\mathrm{stat}}$ and the distribution function $F(W)$ obeying the **Fermi–Dirac statistics**

$$F(W) = \frac{1}{1 + e^{\frac{W - W_{\mathrm{F}}}{k_{\mathrm{B}} T}}} \approx e^{\frac{W_{\mathrm{F}} - W}{k_{\mathrm{B}} T}}, \tag{6.1}$$

where W_{F} is the **Fermi energy**, $k_{\mathrm{B}} = 8.6173 \cdot 10^{-5}\,\mathrm{eV/K}$ **Boltzmann's constant**, and T the (absolute) temperature. One can see that at $T = 0\,\mathrm{K}$, the energy distribution of the electrons has a sharp edge (at the Fermi energy) with all lower states occupied and all states above empty. Once the temperature increases, higher-energy states inside the band structure of the metal and even unbound states begin to become populated. Here, we are interested in electrons featuring these unbound states, because they represent electrons that are thermionically emitted from the cathode. Taking tungsten as an example, the temperature must be below the melting point 3695 K. Therefore, $k_{\mathrm{B}} T$ is below 0.318 eV, which is much smaller than the

binding energy $W_{\text{bind}} = 4.5\,\text{eV}$. This explains the approximation in the last step, since we are interested in unbound states where $W - W_{\text{F}}$ is sufficiently large in comparison with $k_{\text{B}}T$.

The product of the last two equations yields

$$\mathrm{d}n = F(W)\,\mathrm{d}n_{\text{stat}} = 4\pi\left(\frac{2m_{\text{e}}}{h^2}\right)^{3/2} e^{\frac{W_{\text{F}}-W}{k_{\text{B}}T}}\sqrt{W}\mathrm{d}W.$$

The calculation is simpler when we are discussing the momentum p instead of the energy $W = \frac{p^2}{2m_{\text{e}}}$:

$$\mathrm{d}n = \frac{8\pi}{h^3}e^{\frac{W_{\text{F}}-\frac{p^2}{2m_{\text{e}}}}{k_{\text{B}}T}}p^2\mathrm{d}p.$$

We will now use the formula

$$4\pi p^2\mathrm{d}p = \mathrm{d}p_x\,\mathrm{d}p_y\,\mathrm{d}p_z,$$

where $\mathrm{d}p_x$, $\mathrm{d}p_y$, and $\mathrm{d}p_z$ are the differentials of the momentum components in Cartesian coordinates. Furthermore, due to

$$\vec{J} = \rho_q\vec{v},$$

the differential of the current density in the z-direction is given by

$$\mathrm{d}J_z = -ev_z\,\mathrm{d}n = -e\frac{p_z}{m_{\text{e}}}\,\mathrm{d}n.$$

This results in[1]

$$\mathrm{d}J_z = -\frac{2e}{h^3m_{\text{e}}}e^{\frac{W_{\text{F}}-\frac{p^2}{2m_{\text{e}}}}{k_{\text{B}}T}}p_z\,\mathrm{d}p_x\,\mathrm{d}p_y\,\mathrm{d}p_z.$$

Integrating this equation in the x- and y-directions from minus infinity to infinity and in the z-direction from p_{vac} to infinity accounts only for such electrons with sufficient energy to be able to leave the metal, thereby forming a current density J_z in the z-direction:

$$J_z = -\frac{2e}{h^3m_{\text{e}}}e^{\frac{W_{\text{F}}}{k_{\text{B}}T}}\int_{-\infty}^{\infty}e^{-\frac{p_x^2}{2m_{\text{e}}k_{\text{B}}T}}\mathrm{d}p_x\int_{-\infty}^{\infty}e^{-\frac{p_y^2}{2m_{\text{e}}k_{\text{B}}T}}\mathrm{d}p_y\int_{p_{\text{vac}}}^{\infty}e^{-\frac{p_z^2}{2m_{\text{e}}k_{\text{B}}T}}p_z\mathrm{d}p_z\,.$$

[1] Please note that in this equation and in the following, the same symbol e is used for the elementary charge and Euler's constant. Since the latter is the base of the exponential function, this should not lead to misunderstanding.

The first two integrals are of the type

$$\int_{-\infty}^{\infty} e^{-ax^2} \mathrm{d}x = \sqrt{\frac{\pi}{a}} \qquad (a > 0),$$

and the third integral is

$$\int x e^{-ax^2} \mathrm{d}x = -\frac{e^{-ax^2}}{2a} + \text{const.}$$

This leads to

$$J_z = -\frac{4\pi e k_\mathrm{B} T}{h^3} e^{\frac{W_\mathrm{F}}{k_\mathrm{B} T}} m_\mathrm{e} k_\mathrm{B} T e^{-\frac{p_\mathrm{vac}^2}{2 m_\mathrm{e} k_\mathrm{B} T}}.$$

Substituting the momentum again for the energy and observing[2] that

$$W_\mathrm{bind} = W_\mathrm{vac} - W_\mathrm{F}$$

yields the **Richardson or Richardson–Dushman equation**

$$J_\mathrm{S} = -J_z = A_\mathrm{R} T^2 e^{-\frac{W_\mathrm{bind}}{k_\mathrm{B} T}} \tag{6.2}$$

with the **Richardson constant**

$$A_\mathrm{R} = \frac{4\pi m_\mathrm{e} k_\mathrm{B}^2 e}{h^3} \approx 120 \, \frac{\mathrm{A}}{\mathrm{cm}^2 \, \mathrm{K}^2}.$$

For the operation of a vacuum tube, a high-saturation current density J_S is desirable. Therefore, the cathode is operated at high temperature, and the cathode material is chosen so as to exhibit a low binding energy and the ability to withstand operation at these high temperatures. For high-power applications, tungsten or thoriated tungsten is usually employed.

In applying an electric field at the surface of the metal, the binding energy is slightly reduced. This leads to an increase in emission from the cathode. This phenomenon is called the **Schottky effect**. The reduction in binding energy amounts to

[2] p_vac is the momentum corresponding to the energy W_vac, which denotes the potential energy of an electron in vacuum just outside of the metal; W_bind is the binding energy, which denotes the minimal energy (at $T = 0$) required to remove an electron from the metal. For metals, this quantity is equal to the work function, which at $T = 0$, denotes the energy required to transfer an electron from Fermi energy to the outside of the metal (vacuum energy). Usually, all energies are referenced relative to $W_\mathrm{vac} = 0$.

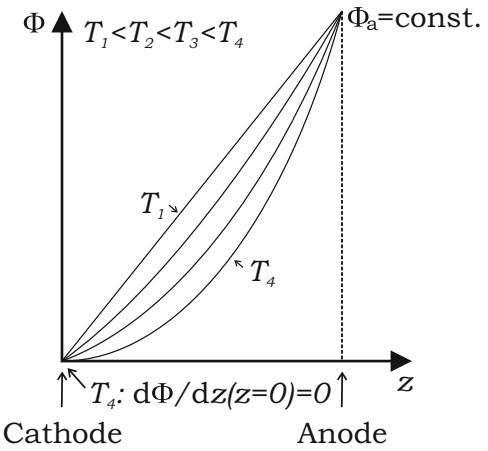

Fig. 6.3 Evolution of an idealized potential distribution between cathode and anode with increasing temperature (increasing charge density)

$$\Delta W = \sqrt{\frac{e^3 E_{\text{surf}}}{4\pi\epsilon_0}}, \tag{6.3}$$

with E_{surf} the electric field strength at the surface of the cathode.

Under some circumstances, not all of the electrons that are emitted by the cathode can travel to the anode. The potential distribution between the cathode and the anode is defined not only by the potential difference and the geometric arrangement between these electrodes. The electrons that are traveling from the cathode to the anode also influence the potential distribution due to the charge they are introducing. This effect is called **space charge**. Under some operational conditions, the presence of these electrons can constitute a limiting factor with respect to the total current that can travel through the tube. To illustrate this effect, consider two parallel electrodes. Without any space charge effects, the potential distribution between these electrodes is given by the straight line shown in Fig. 6.3 denoted by T_1 (no heating). If we now begin to heat the cathode, electrons are emitted, and all of them travel from the cathode to the anode, thereby introducing charge carriers in the space between the electrodes. These charge carriers will now begin to have an influence on the potential, as shown by the curve T_2 in Fig. 6.3. They will shield the potential seen from the anode. This effect is most pronounced at the cathode. When one increases the temperature of the cathode, the number of electrons that are emitted increases according to Eq. (6.2). All these electrons will travel to the anode, thereby further amplifying the space charge effects and flattening out the potential distribution. These mechanisms continue with rising temperature (T_2, T_3) until the gradient of the modified potential at the location of the cathode becomes zero (grad $\Phi = -\vec{E} = 0$) at temperature T_4. The electric field at the location of the cathode now vanishes. And with rising temperature, any additional electrons that are emitted by the cathode will be unable to travel from the cathode to the anode. An electron cloud is formed at the cathode, where electrons emitted from the cathode are in equilibrium with

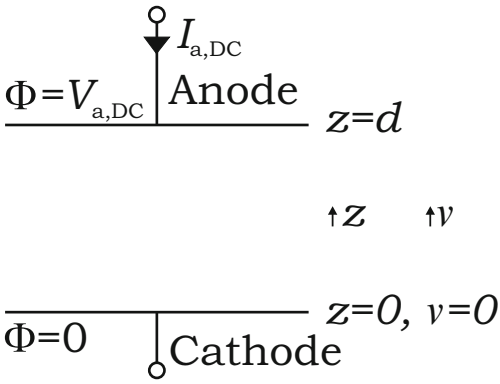

Fig. 6.4 Arrangement of two parallel electrodes used to derive the space charge limited current

electrons that move from the cloud back into the cathode. So from the temperature T_4 upward, the current transported through the tube is almost independent of the cathode temperature; the potential distribution also remains unchanged.

Let us now calculate this effect under the following conditions (see Fig. 6.4):

- Two infinitely extended plane electrodes that are separated in the z-direction by d with constant potentials:

$$\Phi|_{z=0} = 0, \qquad \Phi|_{z=d} = V_{a,DC} \qquad \text{(Dirichlet boundary condition).}$$

- The current density is limited by space charge effects:

$$\left.\frac{d\Phi}{dz}\right|_{z=0} = 0.$$

- The electrons are the only charge carriers between the electrodes.
- The electrons are emitted with zero velocity:

$$v|_{z=0} = 0.$$

- The electron energies are sufficiently small; relativistic effects are negligible.
- The change in anode voltage is slow compared to the transit time of the electrons.
- All magnetic fields in the diode caused by the moving electrons in the vacuum as well as in the conductors (e.g., cathode) are neglected.

The distribution of the potential is given by the Poisson equation (2.54),

$$\Delta\Phi = -\frac{\rho_q}{\epsilon_0}, \tag{6.4}$$

where Φ is the electric potential and ρ_q the space charge density.

Calculating the potential for two plate electrodes of infinite dimensions that are separated in the z-direction leads to

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}z^2} = -\frac{\rho_q}{\epsilon_0}. \tag{6.5}$$

The charge and the current density are connected by the relation

$$J = -J_z = -\rho_q v, \tag{6.6}$$

where v is the velocity of the electrons.

The kinetic energy of the electrons (nonrelativistic case)[3] is equal to the electric energy gain:

$$\Phi e = \frac{1}{2} m_\mathrm{e} v^2. \tag{6.7}$$

Here m_e denotes the mass of the electrons. Combining the last three equations yields

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}z^2} = kJ\Phi^{-1/2}, \tag{6.8}$$

with

$$k = \frac{1}{\epsilon_0}\sqrt{\frac{m_\mathrm{e}}{2e}}$$

constant. It follows that

$$2\frac{\mathrm{d}\Phi}{\mathrm{d}z}\frac{\mathrm{d}^2\Phi}{\mathrm{d}z^2} = 2kJ\Phi^{-1/2}\frac{\mathrm{d}\Phi}{\mathrm{d}z}$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}z}\left[\left(\frac{\mathrm{d}\Phi}{\mathrm{d}z}\right)^2\right] = 2kJ\Phi^{-1/2}\frac{\mathrm{d}\Phi}{\mathrm{d}z}.$$

The current density J remains constant with varying z. By means of an integration, we obtain

$$\left(\frac{\mathrm{d}\Phi}{\mathrm{d}z}\right)^2 = 2kJ\int \Phi^{-1/2}\frac{\mathrm{d}\Phi}{\mathrm{d}z}\,\mathrm{d}z$$

$$\Rightarrow \left(\frac{\mathrm{d}\Phi}{\mathrm{d}z}\right)^2 = 4kJ\Phi^{+1/2} + A.$$

[3] The following example illustrates this assumption: for electrons with a kinetic energy of 20 keV one gets

$$\beta = 0.2719, \qquad \gamma = 1.03914,$$

and the error of the nonrelativistic formula for the kinetic energy is about 5.9%.

At the surface of the cathode,

$$\Phi|_{z=0} = 0, \qquad \left.\frac{\mathrm{d}\Phi}{\mathrm{d}z}\right|_{z=0} = 0$$

is valid. This results in $A = 0$, and we obtain

$$\frac{\mathrm{d}\Phi}{\mathrm{d}z} = 2\sqrt{kJ}\Phi^{1/4}$$

$$\Rightarrow \int \Phi^{-1/4}\mathrm{d}\Phi = 2\sqrt{kJ}\int \mathrm{d}z$$

$$\Rightarrow \frac{4}{3}\Phi^{3/4} = 2\sqrt{kJ}z + B.$$

Using the boundary condition

$$\Phi|_{z=0} = 0$$

at the cathode and

$$\Phi|_{z=d} = V_{\mathrm{a,DC}}$$

results in $B = 0$ and

$$\frac{2}{3}V_{\mathrm{a,DC}}^{3/4} = \sqrt{kJ}d$$

$$\Rightarrow V_{\mathrm{a,DC}} = \left(\frac{3}{2}d\sqrt{k}\right)^{4/3} J^{2/3}.$$

This equation gives the dependency of the potential in the space-charge-limited case. The line depicting T_4 in Fig. 6.3 follows this $d^{4/3}$ dependency. Regarding the current density, we have

$$J = \frac{4}{9kd^2}V_{\mathrm{a,DC}}^{3/2}, \tag{6.9}$$

or in general, for different geometric conditions and integrating over the whole cathode surface A_{c}, we have

$$I_{\mathrm{a,DC}} = KA_{\mathrm{c}}V_{\mathrm{a,DC}}^{3/2},$$

with $I_{\mathrm{a,DC}}$ the total current traveling through the tube. If we assume that the voltage and the current will change slowly enough to justify our static field calculation, we may use this equation also in the non-DC case:

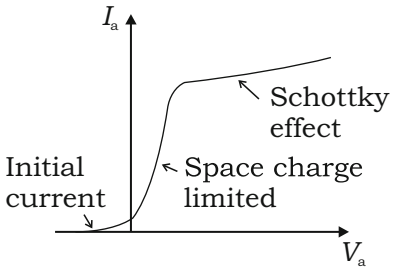

Fig. 6.5 Characteristic current curve of a vacuum diode; the axes are not true to scale

$$I_a = K A_c V_a^{3/2}. \tag{6.10}$$

This equation is called **Child's law or the Langmuir–Child law**, stating that the current is proportional to the power of $3/2$ of the voltage between both electrodes. The constant K depends on the geometry of the tube. For the planar electrodes considered so far, we get

$$K = \frac{4}{9kd^2} = \frac{4\epsilon_0}{9d^2}\sqrt{\frac{2e}{m_e}}.$$

A more detailed analysis of the behavior of electrons between two parallel electrodes under space charge conditions, which also takes different initial velocities into account, is given in [7].

For tubes, a cylindrical assembly of a cathode with an outer radius r_c and an anode with an inner radius of r_a is often realized ($r_a > r_c$). This configuration yields, for a long cylinder [8],

$$K = \frac{4\epsilon_0}{9r_c^2}\sqrt{\frac{2e}{m_e}}\left(\frac{r_a}{r_c} - 1\right)^{-1/2}\left(\ln\frac{r_a}{r_c}\right)^{-3/2}. \tag{6.11}$$

Figure 6.5 shows a typical characteristic curve of a vacuum diode. Plotted here is the current through the diode as a function of the voltage between cathode and anode at a constant cathode temperature.

One can distinguish three different sections. The first part, where the anode voltage is negative, is called area of initial current. Here the current rises exponentially with the voltage. This behavior can be described by Eq. (6.2) modified in that the electrons now have to overcome not only the binding energy of the metal but also the potential difference between both electrodes:

$$J = A_R T^2 e^{-\frac{W_{bind} - eV_{a,DC}}{k_B T}}. \tag{6.12}$$

This equation is valid as long as the current is not limited by space charge effects, which will become relevant even at negative anode voltages. Only the electrons that

leave the metal with sufficient kinetic energy are capable of doing so; therefore, the current is strongly dependent on the temperature of the cathode. In general, the magnitude of this initial current is much lower (several orders of magnitude) than that of the saturation current. With increasing anode voltage, the behavior of the anode current is dominated by space charge effects; it is almost independent of the cathode temperature; its devolution with the anode voltage is given by Eq. (6.10). This behavior remains intact with increasing anode voltage until the current reaches the saturation current given by Eq. (6.2). At that point, all electrons that are emitted from the cathode can travel through the diode; space charge effects are no longer relevant with respect to the current. The current will again become dominated by the temperature of the cathode. This transition is not sharp, since the temperature is not constant over the whole cathode. When one continues to raise the voltage, the saturation current increases slowly according to the Schottky effect; see Eq. (6.3).

In the scope of this chapter, vacuum tubes will be applied in amplifiers. Here, one wants the current of the tube to be strongly influenced by the anode or grid voltage. In addition, it is beneficial if the current does not depend on the temperature of the cathode, allowing much more relaxed requirements regarding temperature stability. Therefore, vacuum tubes in amplifier configuration are generally operated in the space-charge-dominated region.

6.1.2 Triode

The vacuum diode presented in the previous section is unsuitable for application in an amplifier design, because for a given cathode heating, the tube current is not controllable independently of the anode voltage. To overcome this limitation, an additional cold electrode is introduced into the tube, called the **grid**, resulting in a **triode**. The grid is located between the cathode and the anode; it consists of a metal wire helix or metal mesh with an electrical connection to the outside of the tube. Now, for a given heating current, the electron flux from cathode to anode is a function of the grid and the anode voltage. The following considerations illustrate these dependencies: For a diode, the current in the space-charge-dominated region is given by Eq. (6.10),

$$I_\mathrm{a} = \tilde{K}_\mathrm{diode} V_\mathrm{a}^{3/2},$$

where V_a is given by the capacitance between anode and cathode and the charge induced on the cathode:

$$V_\mathrm{a} = \frac{Q_\mathrm{c}}{C_\mathrm{ac}}.$$

In the case of a triode, the current traveling to the anode can be expressed by

$$I_\mathrm{a} = \tilde{K}_\mathrm{triode} \tilde{V}_\mathrm{ctrl}^{3/2}$$

as a function of an effective control voltage

$$\tilde{V}_\mathrm{ctrl} = \frac{Q_\mathrm{c}}{C_\mathrm{tot}}.$$

Here

$$C_\mathrm{tot} = C_\mathrm{ac} + C_\mathrm{gc}$$

represents the combined capacitances of the cathode with respect to the anode and the grid, respectively. The charge Q_c is given by

$$Q_\mathrm{c} = C_\mathrm{ac} V_\mathrm{a} + C_\mathrm{gc} V_\mathrm{g},$$

resulting in

$$I_\mathrm{a} = \tilde{K}_\mathrm{triode} \left(\frac{C_\mathrm{ac} V_\mathrm{a} + C_\mathrm{gc} V_\mathrm{g}}{C_\mathrm{ac} + C_\mathrm{gc}} \right)^{3/2},$$

or

$$I_\mathrm{a} = K_\mathrm{triode} \left(V_\mathrm{g} + \frac{V_\mathrm{a}}{\mu_\mathrm{a}} \right)^{3/2}. \tag{6.13}$$

Here the behavior of the triode in the space-charge-dominated region is described by means of the amplification factor

$$\mu_\mathrm{a} = \frac{C_\mathrm{gc}}{C_\mathrm{ac}} > 1$$

and

$$K_\mathrm{triode} = \tilde{K}_\mathrm{triode} \left(\frac{C_\mathrm{gc}}{C_\mathrm{ac} + C_\mathrm{gc}} \right)^{3/2}.$$

The amplification factor μ_a indicates how much stronger the influence of the grid voltage on the anode current turns out to be than that of the anode voltage. The reverse factor $1/\mu_\mathrm{a}$ describes the ratio between the influence of a change in the anode voltage on the tube current with respect to a change in the grid voltage.

One can see that a triode can be described as a diode with its anode at the position of the grid whose voltage with respect to the cathode is given by $\tilde{V}_\mathrm{ctrl}$. According to Eq. (6.13), three quantities of a triode (I_a, V_g, and V_a) are interconnected. To evaluate the performance of a triode, typically two of these quantities are plotted while one

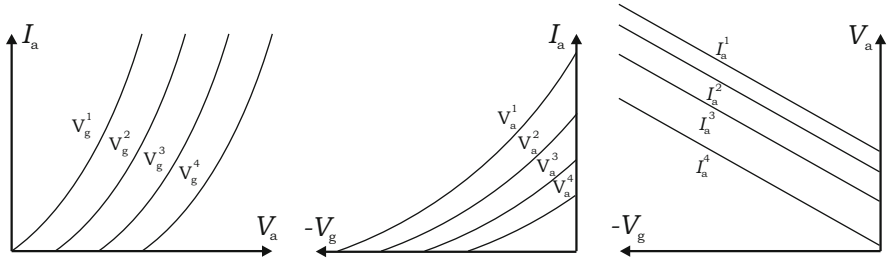

Fig. 6.6 Plate characteristics, mutual characteristics, and constant current curves of a triode

is held constant. If the anode current as a function of the anode voltage is plotted for several constant grid voltages, the plate characteristics are obtained as shown in the first diagram of Fig. 6.6. One can see that for different grid voltages, the curves are shifted with respect to each other along the anode voltage axis, which is due to the different impacts of the grid and anode voltages on the anode current.

The mutual characteristics as shown in the second diagram of Fig. 6.6 display the anode current as a function of the grid voltage for constant anode voltages. Lastly, the constant current curves (right-hand diagram in Fig. 6.6) are obtained if the grid voltage is plotted as a function of the anode voltage for constant anode currents.

In addition to these curves there are three useful quantities to describe the behavior of a triode. These are the amplification factor μ_a mentioned above, the mutual conductance g_m, and the internal resistance R_i. They are given by the following expressions:

$$\mu_a = -\left.\frac{\partial V_a}{\partial V_g}\right|_{I_a=\text{const}}, \tag{6.14}$$

$$g_m = \left.\frac{\partial I_a}{\partial V_g}\right|_{V_a=\text{const}}, \tag{6.15}$$

$$R_i = \left.\frac{\partial V_a}{\partial I_a}\right|_{V_g=\text{const}}. \tag{6.16}$$

The significance of the amplification factor has already been introduced. It can be deduced, for example, from the plate characteristics. The ratio between the shift in anode voltage and the differences in grid voltages between two neighboring constant anode current curves can be used to determine μ_a. The mutual conductance describes the sole impact of a change in grid voltage on the anode current. It is a key figure concerning the amount of amplification that can be realized by a given tube. The mutual conductance can be deduced, for example, by determining the slope of a tangent in the mutual characteristics. Finally, the internal resistance R_i is a measure

of the influence of the anode voltage on the anode current. It is given, for example, by the reciprocal value of the slope of a tangent in the plate characteristic curves.

Based on Eq. (6.13),

$$I_\mathrm{a} = K_\mathrm{triode}\left(V_\mathrm{g} + \frac{V_\mathrm{a}}{\mu_\mathrm{a}}\right)^{3/2}$$

$$\Leftrightarrow V_\mathrm{a} = \mu_\mathrm{a}\left[\left(\frac{I_\mathrm{a}}{K_\mathrm{triode}}\right)^{2/3} - V_\mathrm{g}\right],$$

we now calculate the partial derivatives mentioned above. The first result,

$$\frac{\partial V_\mathrm{a}}{\partial V_\mathrm{g}} = -\mu_\mathrm{a},$$

confirms Eq. (6.14). The other two derivatives are

$$g_\mathrm{m} = \frac{\partial I_\mathrm{a}}{\partial V_\mathrm{g}} = \frac{3}{2}K_\mathrm{triode}\left(V_\mathrm{g} + \frac{V_\mathrm{a}}{\mu_\mathrm{a}}\right)^{1/2},$$

$$R_\mathrm{i} = \frac{\partial V_\mathrm{a}}{\partial I_\mathrm{a}} = \frac{2}{3}\mu_\mathrm{a}\left(\frac{I_\mathrm{a}}{K_\mathrm{triode}}\right)^{-1/3}\frac{1}{K_\mathrm{triode}} = \frac{2}{3}\mu_\mathrm{a}\left(V_\mathrm{g} + \frac{V_\mathrm{a}}{\mu_\mathrm{a}}\right)^{-1/2}\frac{1}{K_\mathrm{triode}}.$$

The product of these two equations yields

$$g_\mathrm{m} R_\mathrm{i} = \mu_\mathrm{a}, \tag{6.17}$$

which is often called the **Barkhausen equation**.

6.1.3 Tetrode

In amplifier configurations, especially for RF cavities, **tetrodes** are often employed. Tetrodes exhibit two cold electrodes between the cathode and the anode, the **control grid** (grid 1) and the **screen grid** (grid 2). The main purpose of the control grid is to modulate the anode current, whereas the screen grid is introduced to reduce the reverse amplification factor of the anode. The anode current in the space-charge-limited region is given by

$$I_\mathrm{a} = K_\mathrm{tetrode}\left(V_\mathrm{g1} + \frac{V_\mathrm{g2}}{\mu_\mathrm{g2}} + \frac{V_\mathrm{a}}{\mu_\mathrm{a}}\right)^{3/2} \tag{6.18}$$

as a generalization of Eq. (6.13).

6.2 Tube Amplifiers

In this section, RF power amplifiers based on vacuum tubes are discussed. More detailed information may be found, for example, in [9].

In the context of tube amplifiers for synchrotron cavities, the main task of the amplifier is to provide the AC current to achieve the required gap voltages. As discussed in Sect. 4.1.15, the gap voltages used in hadron synchrotron cavities for low harmonic numbers are typically in the range of a few hundred up to a few thousand volts. Since the shunt impedance of these cavities is usually fairly low, at least compared to LINAC cavities or electron synchrotron cavities, the required amplitude of the AC currents is in the range of a few amperes up to dozens of amperes. So in general, the type of amplifier discussed here has to provide high currents at voltage levels of several kilovolts, resulting in power levels of up to several hundreds of kilowatts. In designing an amplifier for this type of application, usually two main goals have to be taken into account: Firstly, the signal quality is an important factor. Every distortion in the current delivered by the tube will lead to a distortion of the gap voltage seen by the beam, albeit reduced by the loaded quality factor of the cavity. These distortions will have an effect on the shape of the bucket. Secondly, the power level of this type of amplifier requires at least some considerations regarding the efficiency of the conversion from DC to AC power. These two requirements contradict each other. In general, an increase in efficiency of a tube amplifier is accompanied by an increase in distortions. So when designing an RF system consisting of amplifier and cavity, a trade-off has to be achieved based on the requirements of this particular system. In classical terms, this type of amplifier can be located somewhere between a broadcast amplifier and a high-power HIFI amplifier. In general, the amplification factor (the ratio between the input and output voltage) is of less concern here, since the input voltage can easily be preamplified to sufficient levels.

Another important topic in designing the power amplifier of a synchrotron RF system is impedance matching. Therefore, it is necessary in most applications to consider the design of the power amplifier and the synchrotron cavity as a combined task. There are many different approaches to achieving this matching, such as by fixing the impedance of the cavity or parts of the cavity that will be driven in parallel to a fixed impedance, usually 50 Ω. In that case, usually one or more semiconductor amplifiers are used. Due to the 50-Ω regime, it is possible to separate the cavity from the amplifier (e.g., housing the radiation-sensitive semiconductor amplifier in an area separated from the beam line). Another approach regarding the impedance matching is to locate the amplifier very close to the cavity, thereby reducing the length of the power transmission lines as much as possible. In doing so, it is not necessary to match the impedance of the transmission line. We will focus on this scheme, since it allows the transmission of very high power levels from the amplifier to the cavity. In this scheme, again, the impedance of the cavity is a major factor in designing the power amplifier, and it may also be beneficial in designing the cavity

to adapt the impedance seen by the power amplifier in order to allow better matching to certain tubes.

In dealing with this set of requirements, the natural choice of power source is the tetrode. Compared to a triode, the tetrode exhibits a higher amplification factor, due to a screen grid that can be controlled independently of the control grid and the anode voltage. A tetrode can also be operated at lower anode voltages, due to the fact that the control voltage can be kept sufficiently high by a proper choice of the screen grid voltage. This enables a tetrode to be operated in a given scenario at lower DC anode voltages compared to a triode, thereby increasing the efficiency. These benefits generally outweigh the fact that the current emitted by the cathode of a tetrode partly flows to the positive screen grid, reducing the anode current. This effect is fairly small, especially since the geometry of the tetrodes is optimized to minimize it.

Pentodes feature a suppressor grid that prevents secondary electrons emitted by the anode from traveling to the screen grid. In comparing pentodes to tetrodes in the high current operation scheme, this is usually not beneficial, since for the high voltage application discussed here, it is easily achievable to maintain a sufficiently high voltage difference between the anode and the screen grid to block this anode to screen grid current.[4]

Figure 6.7 shows a typical layout of a tetrode-based amplifier in grounded cathode configuration. Here, the cathode of the tube is kept at ground potential. Several voltage supplies are required to operate the tube. A high voltage power supply is responsible for obtaining the voltage difference between cathode and anode. This power supply has to provide fairly high DC currents. In addition, a voltage supply is required to provide the DC voltage of the control grid and another supply will ensure the DC voltage of the screen grid. A filament supply, not shown in Fig. 6.7, provides the necessary current to achieve sufficient heating of the cathode. The designer of power tube amplifiers will always separate the DC resistance seen by the anode voltage supply from the AC resistance. This can be done by the use of a transformer, as shown in Fig. 6.7 or, for example, by implementing a large choke coil in parallel with the load resistor. Each remaining DC resistance will reduce the anode voltage and thereby increase the power that has to be delivered by the voltage supply to reach a given AC voltage amplitude on the load resistor. Usually, the input AC voltage to the control grid is also provided via a transformer, separating the AC generator (e.g., driver amplifier) from the DC control grid voltage. The input impedance of this transformer is usually matched to 50 Ω to allow the transmission of this input signal over a larger distance. According to the definition, the mutual conductance is given by

$$g_{\mathrm{m}} = \left.\frac{\partial I_{\mathrm{a}}}{\partial V_{\mathrm{g1}}}\right|_{V_{\mathrm{a}}}$$

[4] In most cases, the potential barrier introduced by the space charge effect of the high anode current alone is sufficient to inhibit the secondary electrons from traveling to the screen grid.

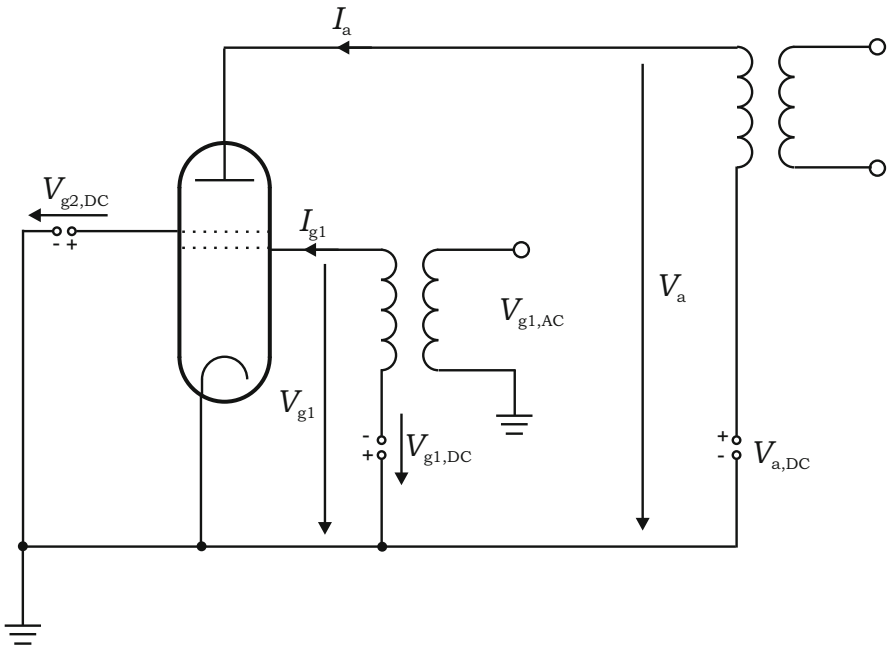

Fig. 6.7 Layout of a single-ended tube amplifier in grounded cathode configuration

for a sinusoidal control grid voltage

$$V_{g1,AC} = \hat{V}_{g1,AC} \cos(\omega t).$$

When using the orientations of voltages and currents as shown in Fig. 6.7, the anode AC current is given by

$$I_{a,AC} = g_m V_{g1,AC}.$$

This equation holds only in case of a short circuit operation. When R_a is larger than zero, the retroaction of the anode voltage on the anode current, moderated by the reverse amplification factor $1/\mu_a$, has to be taken into account (see Eq. (6.18)), resulting in

$$I_{a,AC} = g_m V_{ctrl,AC} \qquad \text{with} \qquad V_{ctrl,AC} = V_{g1,AC} + \frac{V_{a,AC}}{\mu_a}.$$

Using the Barkhausen equation (6.17),

$$\frac{g_m R_i}{\mu_a} = 1,$$

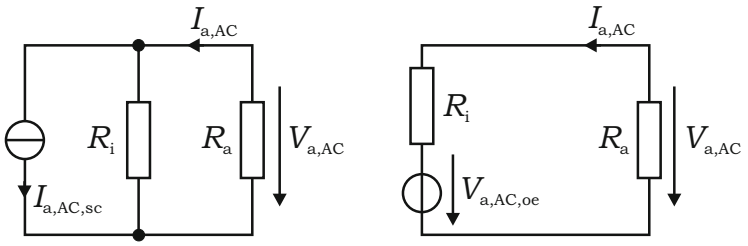

Fig. 6.8 Representation of a tube amplifier by a current source (*left*) or voltage source (*right*)

and $V_{a,AC} = -I_{a,AC} R_a$ leads to

$$I_{a,AC} = \frac{\mu_a V_{g1,AC}}{R_i + R_a},$$

$$V_{a,AC} = -\mu_a V_{g1,AC} \frac{R_a}{R_i + R_a}.$$

In case of a short circuit operation, this simplifies to

$$I_{a,AC,sc} = \frac{\mu_a V_{g1,AC}}{R_i},$$

$$V_{a,AC,sc} = 0,$$

and for open-ended operation, to

$$I_{a,AC,oe} = 0,$$

$$V_{a,AC,oe} = -\mu_a V_{g1,AC}.$$

The voltage amplification $\mu = -\hat{V}_{a,AC}/\hat{V}_{g1,AC}$ varies with the load resistance, remaining always below the open-ended amplification $\mu = \mu_a$.

The tube may be described as a current source or as a voltage source, as depicted in Fig. 6.8. These equivalent circuits describe the small-signal AC behavior of the tube as seen from the load. They cannot be used to describe the DC performance of the tube or the internal processes within the tube.

For each load resistance R_a, the relation between $V_{a,AC}$ and $I_{a,AC}$ is given by

$$V_{a,AC} = -R_a \, I_{a,AC}.$$

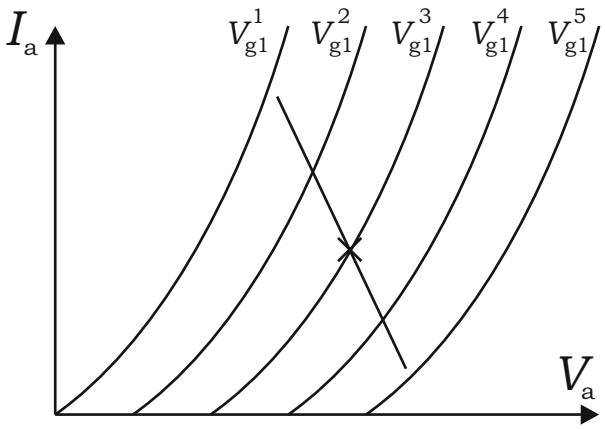

Fig. 6.9 Plate characteristics and example of a load line

The actual value of the anode current is given by

$$I_a = I_{a,DC} + I_{a,AC}$$

$$\Rightarrow I_a = I_{a,DC} - \frac{V_{a,AC}}{R_a}.$$

This represents the equation of a line with slope $-1/R_a$. The resulting line is called **load line** of the tube. An example of such a load line is shown in Fig. 6.9 together with exemplary plate characteristics of a tube. Such a diagram is a very efficient tool for visualizing the behavior of a tube amplifier under certain operational conditions. The line shows the actual values of the anode voltage, the anode current, and the control grid voltage for each moment in time. The point on the load line where $V_{g1} = V_{g1,DC}$ is called the **operating point** of the amplifier (cross in Fig. 6.9).

An amplifier can be used in different modes of operation, depending on the amount of time during which the amplifier provides current to the load. Here one distinguishes between class A, class B, and class C operation. The mode of operation is determined by the type of tube (the associated characteristic curves) and the choice of DC anode voltage, DC control grid voltage, AC control grid voltage, and screen grid voltage. Depending on these parameters, the tube may block the current flow for a certain period of one RF cycle.

If the parameters are chosen in such a way that the tube can provide current during the whole RF period, e.g., by keeping $V_{g1,DC}$ sufficiently high or $V_{g1,AC}$ sufficiently low for a given $V_{a,DC}$, the tube is used in **class A operation**. In class A operation, the anode current at the operating point differs only slightly from the DC anode current. The anode DC current is almost independent of the amplitude of the control grid AC voltage.

If the parameters ($V_{a,DC}$, $V_{g1,DC}$, $V_{g1,AC}$, $V_{g2,DC}$) are chosen in such a way that a given tube provides current for exactly half of each RF period, this is called

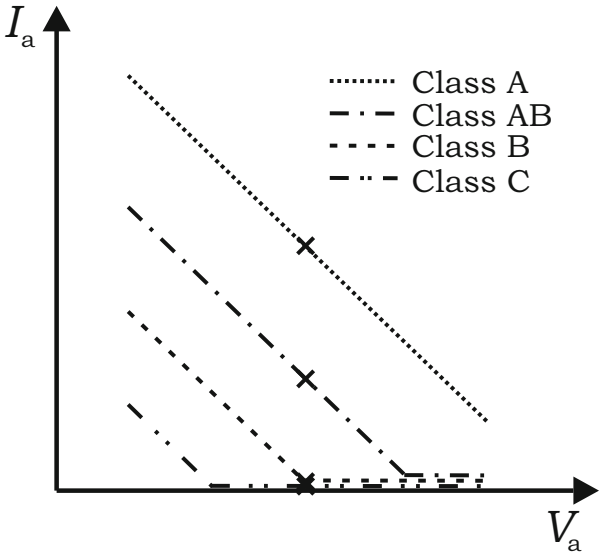

Fig. 6.10 Different classes of operation, typical operating lines of class A, AB, B, and C

class B operation. The anode current at the operating point is zero in an idealistic[5] approximation, and the amplitude of the AC control grid voltage will have a strong influence on the DC anode current. All modes of operation between class A and class B are called **class AB** (e.g., if the tube delivers current for three-fourths of an RF period).

If the tube delivers current for less than half of an RF period, this operation is called **class C**. Again, the anode current at the operating point will be zero, and the amplitude of the control grid voltage will have an even stronger influence on the DC anode current. Idealized operating lines for different classes of operation are shown in Fig. 6.10. As will be discussed later, the efficiency of the amplifier as well as the amplitudes of the higher harmonic current components will increase when the class of operation changes from A to B to C.

One aim of amplifier development is to ensure that the output signal of the amplifier follows the input signal in a linear way. In the case of the amplifier discussed here, the actual value of the anode AC current has to be proportional to the actual value of the AC voltage at the control grid that will be delivered by a driver amplifier. There are several mechanisms that reduce this linearity. One reason for nonlinearities arises from positive control grid values. If the control grid is positive, a fraction of the current will begin to switch from the anode to the control grid. In addition, the activation of the control grid now requires input power during certain times of the period; this may induce further distortions of the

[5] Due to the characteristics of the tetrode, the anode current will not drop to zero sharply. It will be a continuous process.

control grid voltage. These imperfections can easily be avoided if the control grid is always operated at negative voltages. A second type of nonlinearity arises from a nonconstant amplification factor, which will have an impact on the anode current. If the anode voltage becomes sufficiently low compared to the screen grid voltage, a significant portion of the current emitted from the cathode may switch from the anode to the screen grid. Under these operating conditions, a slight variation of the anode voltage has a huge impact on the anode current; hence the amplification factor rises sharply in that region, producing nonlinearities. The extent of this type of distortion can be greatly reduced when the operating line of the tetrode is chosen in such a way that there is always a sufficient gap between the anode and the screen grid voltage. As mentioned earlier, the mutual conductance is not a constant; it varies strongly with the operating parameters of the tetrode, as can be seen here:

$$g_\mathrm{m} = \left.\frac{\partial I_\mathrm{a}}{\partial V_\mathrm{g1}}\right|_{V_\mathrm{a}}.$$

Using Eq. (6.18) yields

$$g_\mathrm{m} = \frac{3}{2} K_\mathrm{tetrode} \sqrt{V_\mathrm{g1} + \frac{V_\mathrm{a}}{\mu_\mathrm{a}} + \frac{V_\mathrm{g2}}{\mu_\mathrm{g2}}}.$$

According to Eq. (6.18), this introduces a third type of nonlinearity that cannot be avoided. The amplitude of higher harmonic components of the anode current rises with the ratio between the AC anode and the DC anode current, as we will see now. Neglecting the influence of the anode voltage (short circuit operation), the anode current around the operating point can be expanded into a Taylor series

$$I_\mathrm{a} = I_\mathrm{a,DC,op} + A_1 V_\mathrm{g1,AC} + \frac{1}{2} A_2 V_\mathrm{g1,AC}^2 + \frac{1}{6} A_3 V_\mathrm{g1,AC}^3 + \frac{1}{24} A_4 V_\mathrm{g1,AC}^4 + \cdots$$

with

$$A_1 = \left.\frac{\mathrm{d}I_\mathrm{a}}{\mathrm{d}V_\mathrm{g1}}\right|_\mathrm{op}, \qquad A_2 = \left.\frac{\mathrm{d}^2 I_\mathrm{a}}{\mathrm{d}V_\mathrm{g1}^2}\right|_\mathrm{op}, \qquad A_3 = \left.\frac{\mathrm{d}^3 I_\mathrm{a}}{\mathrm{d}V_\mathrm{g1}^3}\right|_\mathrm{op}, \qquad A_4 = \left.\frac{\mathrm{d}^4 I_\mathrm{a}}{\mathrm{d}V_\mathrm{g1}^4}\right|_\mathrm{op}, \ldots.$$

Assuming a distortion-free signal

$$V_\mathrm{g1,AC} = \hat{V}_\mathrm{g1,AC} \cos(\omega t)$$

at the control grid results in

$$I_\mathrm{a} = I_\mathrm{a,DC,op} + A_1 \hat{V}_\mathrm{g1,AC} \cos(\omega t) + \frac{1}{2} A_2 \hat{V}_\mathrm{g1,AC}^2 \cos^2(\omega t) + \frac{1}{6} A_3 \hat{V}_\mathrm{g1,AC}^3 \cos^3(\omega t) + \\ + \frac{1}{24} A_4 \hat{V}_\mathrm{g1,AC}^4 \cos^4(\omega t) + \cdots.$$

Using

$$\cos^2(\omega t) = \frac{1}{2} + \frac{1}{2}\cos(2\omega t),$$
$$\cos^3(\omega t) = \frac{3}{4}\cos(\omega t) + \frac{1}{4}\cos(3\omega t),$$
$$\cos^4(\omega t) = \frac{3}{8} + \frac{1}{2}\cos(2\omega t) + \frac{1}{8}\cos(4\omega t),$$

yields

$$I_\mathrm{a} = I_\mathrm{a,DC,op} + \Delta I_\mathrm{a} + \hat{I}_\mathrm{a,AC,\omega}\cos(\omega t) + \hat{I}_\mathrm{a,AC,2\omega}\cos(2\omega t) + \hat{I}_\mathrm{a,AC,3\omega}\cos(3\omega t) + \cdots$$

with

$$\Delta I_\mathrm{a} = \frac{1}{4}A_2\hat{V}^2_\mathrm{g1,AC} + \frac{1}{64}A_4\hat{V}^4_\mathrm{g1,AC} + \cdots, \tag{6.19}$$
$$\hat{I}_\mathrm{a,AC,\omega} = A_1\hat{V}_\mathrm{g1,AC} + \frac{1}{8}A_3\hat{V}^3_\mathrm{g1,AC} + \cdots, \tag{6.20}$$
$$\hat{I}_\mathrm{a,AC,2\omega} = \frac{1}{4}A_2\hat{V}^2_\mathrm{g1,AC} + \frac{1}{48}A_4\hat{V}^4_\mathrm{g1,AC} + \cdots, \tag{6.21}$$
$$\hat{I}_\mathrm{a,AC,3\omega} = \frac{1}{24}A_3\hat{V}^3_\mathrm{g1,AC} + \cdots. \tag{6.22}$$

The quantity ΔI_a represents the shift of the DC anode current with respect to the anode current at the operating point due to the asymmetric characteristic current curves. As mentioned during the discussion of the different classes of amplifier operation, this shift increases with rising amplitude of the control grid voltage and especially in moving from class A to class B or class C. The quantity $\hat{I}_\mathrm{a,AC,n\omega}$ describes the amplitude of the nth fundamental of the anode current. Again, the higher harmonic content rises when the amplitude of the control grid voltage increases as well as in moving from class A to class B or class C operation.

In order to quantify the harmonic distortion, we will use the ratio

$$k_{\mathrm{hd},n} = \frac{\hat{I}_\mathrm{a,AC,n\omega}}{\hat{I}_\mathrm{a,AC,\omega}}.$$

for the individual harmonics.

We will now perform a first estimation regarding the amount of harmonic distortion for a particular mode of operation of a tube. Note that these approximations

assume that the tube is operated in the space-charge-limited region, which is strictly true only in class A operation. Furthermore, for the sake of simplicity, the effect of a variation of the anode voltage on the anode current is again neglected (short-circuit approximation).

In the space-charge-limited region, we have

$$I_\mathrm{a} = K_\mathrm{tetrode} V_\mathrm{ctrl}^\zeta = K_\mathrm{tetrode} \left(V_\mathrm{g1} + \frac{V_\mathrm{g2}}{\mu_\mathrm{g2}} + \frac{V_\mathrm{a}}{\mu_\mathrm{a}} \right)^\zeta .$$

This equation is based on Eq. (6.18), but we now assume a general exponent ζ instead of the specific value $3/2$, since this may be useful in describing specific tubes more accurately.

For the operating point,

$$I_\mathrm{a,DC,op} = K_\mathrm{tetrode} V_\mathrm{ctrl,DC,op}^\zeta$$

is valid, so that

$$A_1 = \left. \frac{\mathrm{d}I_\mathrm{a}}{\mathrm{d}V_\mathrm{g1}} \right|_\mathrm{op} = \zeta \; K_\mathrm{tetrode} V_\mathrm{ctrl,DC,op}^{\zeta-1} = \zeta \frac{I_\mathrm{a,DC,op}}{V_\mathrm{ctrl,DC,op}}$$

$$A_2 = \left. \frac{\mathrm{d}^2 I_\mathrm{a}}{\mathrm{d}V_\mathrm{g1}^2} \right|_\mathrm{op} = \zeta(\zeta-1) \; K_\mathrm{tetrode} V_\mathrm{ctrl,DC,op}^{\zeta-2} = \zeta(\zeta-1) \frac{I_\mathrm{a,DC,op}}{V_\mathrm{ctrl,DC,op}^2}$$

is obtained. Equation (6.20) leads to

$$\hat{I}_{\mathrm{a,AC},\omega} \approx A_1 \hat{V}_\mathrm{g1,AC} = \zeta I_\mathrm{a,DC,op} \frac{\hat{V}_\mathrm{g1,AC}}{V_\mathrm{ctrl,DC,op}} . \tag{6.23}$$

Equation (6.19) yields

$$\Delta I_\mathrm{a} \approx \frac{1}{4} A_2 \hat{V}_\mathrm{g1,AC}^2 = \frac{1}{4} \zeta(\zeta-1) I_\mathrm{a,DC,op} \frac{\hat{V}_\mathrm{g1,AC}^2}{V_\mathrm{ctrl,DC,op}^2} .$$

By means of Eq. (6.23), this may be written as

$$\Delta I_\mathrm{a} \approx \frac{1}{4} \frac{\zeta-1}{\zeta} \frac{\hat{I}_{\mathrm{a,AC},\omega}^2}{I_\mathrm{a,DC,op}} .$$

Equation (6.21) leads to

$$\hat{I}_{\mathrm{a,AC},2\omega} \approx \frac{1}{4} A_2 \hat{V}_\mathrm{g1,AC}^2 = \frac{1}{4} \zeta(\zeta-1) I_\mathrm{a,DC,op} \frac{\hat{V}_\mathrm{g1,AC}^2}{V_\mathrm{ctrl,DC,op}^2} . \tag{6.24}$$

By means of Eq. (6.23), this may be written as

$$\hat{I}_{\mathrm{a,AC},2\omega} \approx \frac{1}{4}\frac{\zeta-1}{\zeta}\frac{\hat{I}^2_{\mathrm{a,AC},\omega}}{I_{\mathrm{a,DC,op}}}.$$

This allows us to calculate

$$k_{\mathrm{hd},2} = \frac{\hat{I}_{\mathrm{a,AC},2\omega}}{\hat{I}_{\mathrm{a,AC},\omega}} \approx \frac{1}{4}\frac{\zeta-1}{\zeta}\frac{\hat{I}_{\mathrm{a,AC},\omega}}{I_{\mathrm{a,DC,op}}}. \tag{6.25}$$

It can be seen that in this approximation, the second harmonic distortion increases proportionally to the ratio of the amplitude of the current of the fundamental harmonic compared to the current of the operating point. So these distortions can be reduced either by reducing the AC current or by increasing the current at the operating point.

By means of Eq. (6.23), the result (6.25) may alternatively be written as

$$k_{\mathrm{hd},2} = \frac{\hat{I}_{\mathrm{a,AC},2\omega}}{\hat{I}_{\mathrm{a,AC},\omega}} \approx \frac{1}{4}(\zeta-1)\frac{\hat{V}_{\mathrm{g1,AC}}}{V_{\mathrm{ctrl,DC,op}}}.$$

The efficiency of the class A power amplifier may be defined by the ratio between the RF power at the fundamental harmonic (only this is useful for standard acceleration) delivered to the cavity and the DC power received by the amplifier:

$$\eta_{\mathrm{ampl}} = \frac{P_{\mathrm{a,AC},\omega}}{P_{\mathrm{a,DC,op}}}.$$

With

$$P_{\mathrm{a,AC},\omega} = \frac{1}{2}\hat{V}_{\mathrm{a,AC},\omega}\hat{I}_{\mathrm{a,AC},\omega}$$

and

$$P_{\mathrm{a,DC,op}} = V_{\mathrm{a,DC,op}} I_{\mathrm{a,DC,op}},$$

we obtain

$$\eta_{\mathrm{ampl}} = \frac{1}{2}\frac{\hat{V}_{\mathrm{a,AC},\omega}\hat{I}_{\mathrm{a,AC},\omega}}{V_{\mathrm{a,DC,op}} I_{\mathrm{a,DC,op}}}.$$

Here one can see that the efficiency is proportional to the ratio of the amplitude of the current of the fundamental harmonic to the DC current of the operating point. This emphasizes that system efficiency and minimal distortion are contradicting targets in the design of a synchrotron RF amplifier. In general, the ratio $\hat{I}_{\mathrm{a,AC},\omega}/I_{\mathrm{a,DC,op}}$ cannot

exceed 1 in class A operation, and the ratio $\hat{V}_{\mathrm{a,AC},\omega}/V_{\mathrm{a,DC,op}}$ always stays below 1, due to the fact that the anode voltage always has to remain sufficiently above the screen grid voltage to ensure that (electron) current can flow from the cathode to the anode. Therefore, the efficiency in class A operation is limited to

$$\eta_{\mathrm{ampl}} < \frac{1}{2}.$$

6.3 Tube Operation

The power delivered by the tetrodes used for synchrotron cavities is rather high, and these tubes are operated at moderate efficiencies, resulting in a significant heat load of the tetrodes. Most synchrotron cavity power amplifiers are designed in such a way that they can also be operated without RF input for some time. In that case, the efficiency is zero, and the whole power delivered by the power supplies is dissipated in the tube. This power has to be removed by active cooling systems to prevent overheating of the tetrode, which would result in melting and the destruction of the tube. Usually two different methods of cooling are applied: Most of the power is dissipated in the anode of the tetrode, which is cooled effectively by deionized water. The water requirements such as conductivity, flow rate, and pressure drop are usually listed in the data sheet of the tube. The remaining heat load (e.g., grid, cathode) is handled by forced air cooling of the tube casing and socket.

The tubes must be handled with care to maximize the lifetime. This includes certain precautions in transporting or handling the tube. It is, for example, very important to restrict the maximum acceleration force experienced by the tube. Certain operating parameters of the tetrode, e.g., anode or grid voltages and currents, must not exceed predefined values. These values have to be observed during operation, and appropriate action, e.g., a normal or a fast shutdown procedure, has to be taken when one of these values exceeds its threshold. In some cases, especially in case of an overcurrent on the grids, this shutdown procedure has to be performed very fast. Therefore, the crucial parameters have to be monitored by a fast electronic module, and the reacting device—usually located in the power supply unit—that is responsible for the grounding of tetrode voltages also has to be fast (e.g., ignitron or solid state switch). In case of a pulsed RF system that uses tetrodes above their CW rating, it is also necessary to monitor the total amount of power dissipated in the tetrode during each pulse. This can be done by integrating the DC power delivered to the tetrode and the RF power provided by it. When the difference between these two quantities exceeds a predefined threshold, the tube has to be shut down.

Special care has to be taken regarding the activation and deactivation of the tetrode. The sequence used to activate and deactivate the different power supplies is crucial. Regarding activation, first the cathode supply (filament) is turned on. It may take up to several minutes for the cathode to reach a stable temperature;

afterward, the negative DC voltage of the control grid is applied before the main anode voltage is turned on. Finally, the screen grid voltage is activated. Now the tube is ready to operate. This sequence ensures that there will be no excessive current from the cathode to one of the grids or toward the anode. This sequence is reversed in deactivating the tube.

References

1. K.R. Spangenberg, *Vacuum Tubes* (McGraw-Hill, New York/Toronto/London, 1948)
2. R.G. Carter, RF power generation, in *CAS - CERN Accelerator School: RF for Accelerators*, Ebeltoft, 8–17 Jun 2010, pp. 173–207
3. H. Barkhausen, E.-G. Woschni, *Lehrbuch der Elektronenröhren*, Band 1, 12. Auflage (S. Hirzel Verlag, Leipzig, 1969)
4. A.H.W. Beck, *Thermionic Valves: Their Theory and Design* (Cambridge University Press, Cambridge, 1953)
5. D. Preist, M. Shrader, The klystrode — an unusual transmitting tube with potential for UHF-TV. Proc. IEEE **70**(11), 1318–1325 (1982)
6. A.S. Gilmour, *Klystrons, Traveling Wave Tubes, Magnetrons, Crossed-Field Amplifiers, and Gyrotrons* (Artech House, Norwood, 2011)
7. I. Langmuir, The effect of space charge and initial velocities on the potential distribution and thermionic current between parallel plane electrodes. Phys. Rev. **21**, 419–435 (1923)
8. X. Chen, J. Dickens, L.L. Hatfield, E.-H. Choi, M. Kristiansen, Approximate analytical solutions for the space-charge-limited current in one-dimensional and two-dimensional cylindrical diodes. Phys. Plasmas **11** (2004) http://dx.doi.org/10.1063/1.1743309
9. H. Barkhausen, E.-G. Woschni, Lehrbuch der Elektronenröhren, Band 2, 10. Auflage (S. Hirzel Verlag, Leipzig, 1968)

Chapter 7
Closed-Loop Control

In this chapter, an introduction to the basics of continuous-time feedback systems is given. For more detailed treatments, the reader is referred to textbooks such as [1–5]. A simple amplitude control loop serves as an example in the following sections. The concepts presented here, however, may also be applied to more advanced control loops (cf. [6, 7]). The RF control loops are often called **low-level RF (LLRF) systems** to distinguish them from the high-power parts.

7.1 Basics of Continuous-Time Feedback Systems

Since many discrete feedback systems may be treated as quasicontinuous if the sampling time is small enough, discrete-time systems are not covered in the following. The analysis of discrete-time systems is, however, possible in an analogous way to continuous-time systems with the $\mathcal{Z}$-transform instead of the Laplace transform [8]. Most feedback analysis and design methods may then be used for discrete systems in a very similar way.

7.1.1 Linear Time-Invariant Systems

The systems under consideration are assumed to be linear and time-invariant (they are so-called **LTI systems**). Assume a general dynamic system

$$y(t) = \varphi\{x(t)\}$$

that maps the input signal $x(t)$ to the output signal $y(t)$. If the system is time-invariant, a time shift at the input will lead to the shifted output

$$\varphi\{x(t - t_0)\} = y(t - t_0). \tag{7.1}$$

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6_7

In case of a linear system, a linear combination of two input signals $x_1(t)$ and $x_2(t)$ will lead to the same linear combination of their corresponding outputs $y_1(t) = \varphi\{x_1(t)\}$ and $y_2(t) = \varphi\{x_2(t)\}$, i.e.,

$$\varphi\{a_1 x_1(t) + a_2 x_2(t)\} = a_1 y_1(t) + a_2 y_2(t) \tag{7.2}$$

holds for arbitrary constants a_1 and a_2.

A consequence of properties (7.1) and (7.2) is that the output of LTI systems can be calculated in the Laplace domain as

$$Y(s) = H(s)\, X(s), \tag{7.3}$$

where the transfer function $H(s)$ corresponds to the impulse response $h(t)$ of the system as defined in Sect. 2.3, and $X(s)$ is the Laplace transform of the system input $x(t)$. This fact is of particular importance, because it enables the analysis and design of feedback systems in the Laplace domain. For a demonstration of the fact that Eq. (7.3) holds for any LTI system, we follow [9] and approximate the input signal $x(t)$ by the step function

$$x_{\text{step}}(t) = \sum_{\nu=0}^{\infty} x(\tau_\nu) \Big(\Theta(t - \tau_\nu) - \Theta(t - \tau_{\nu+1}) \Big) \approx x(t),$$

where $\tau_\nu = \nu \Delta\tau$ are discrete sampling times with distance $\Delta\tau$ and $\Theta(t)$ is the Heaviside step function. It is assumed that $x(t)$ is zero for $t < 0$, as introduced in Sect. 2.2, for all functions for which the one-sided Laplace transform is used. The step response of the system, i.e., the output for $x(t) = \Theta(t)$, will be denoted by $y_\Theta(t)$ in the following. For the input $x_{\text{step}}(t)$, the LTI properties then lead to the output response

$$y_{\text{step}}(t) = \sum_{\nu=0}^{\infty} x(\tau_\nu) \Big(y_\Theta(t - \tau_\nu) - y_\Theta(t - \tau_{\nu+1}) \Big).$$

The continuous output response $y(t)$ is obtained for the limit $\Delta\tau \to 0$:[1]

$$\begin{aligned} y(t) &= \lim_{\Delta\tau \to 0} \sum_{\nu=0}^{\infty} x(\tau_\nu) \, \frac{y_\Theta(t - \tau_\nu) - y_\Theta(t - \tau_{\nu+1})}{\Delta\tau} \, \Delta\tau \\ &= \int_0^\infty x(\tau) \, \dot{y}_\Theta(t - \tau) \, \mathrm{d}\tau. \end{aligned}$$

[1] The assumption is made that the step response $y_\Theta(t)$ is continuous at $t = 0$, continuously differentiable for $t > 0$, and zero for $t < 0$. However, the proof is also possible if $y_\Theta(t)$ is piecewise analytic for $t > 0$ and zero for $t < 0$ [10].

If the derivative $\dot{y}_\Theta(t)$ is denoted by the function $h(t)$, this is a convolution integral, and Eq. (2.27),

$$y(t) = h(t) * x(t) \quad \circ\!\!-\!\!\bullet \quad Y(s) = H(s)\, X(s),$$

holds. The choice of $h(t)$ is indeed not coincidental, because a comparison with Sect. 2.3 shows that due to $\dot{\Theta}(t) = \delta(t)$, this is the already defined impulse response, and the relation

$$h(t) = \dot{y}_\Theta(t) \quad \circ\!\!-\!\!\bullet \quad H(s) = s\, Y_\Theta(s)$$

holds for $t > 0$, i.e., the impulse response $h(t)$ is the derivative of the step response with respect to time. Conversely, it can easily be shown that systems defined by Eq. (7.3) are linear, because in the Laplace domain, the output $Y(s)$ results from a simple multiplication of the input $X(s)$ and the transfer function [10]. In addition, they are time-invariant, because the shifted input

$$x(t - t_0) \quad \circ\!\!-\!\!\bullet \quad X(s)\, e^{-t_0 s}$$

leads to the output

$$Y(s)\, e^{-t_0 s} \quad \bullet\!\!-\!\!\circ \quad y(t - t_0).$$

In summary, we can conclude that the definition of LTI systems by the properties (7.1) and (7.2) is equivalent to Definition (7.3).

In many cases, the transfer function $H(s)$ has the form

$$H(s) = \frac{b_0 + b_1 s + \ldots + b_m s^m}{a_0 + a_1 s + \ldots + a_n s^n} \tag{7.4}$$

with real coefficients b_ν and a_ν and nonzero coefficients $b_m \neq 0$ and $a_n \neq 0$. This is a rational transfer function, and the system (7.3) is then represented in the time domain by the linear ODE

$$a_0 y(t) + a_1 \dot{y}(t) + \ldots + a_n \frac{\mathrm{d}^n}{\mathrm{d}t^n} y(t) = b_0 x(t) + b_1 \dot{x}(t) + \ldots + b_m \frac{\mathrm{d}^m}{\mathrm{d}t^m} x(t) \tag{7.5}$$

with constant coefficients. A transfer function (7.4) is called **proper** if $m \leq n$ and **strictly proper** if $m < n$. In the latter case, $H(s)$ tends to zero as $|s| \to \infty$.

It is sometimes more convenient to use the **zero-pole-gain representation**

$$H(s) = \frac{K(s - z_1)(s - z_2) \ldots (s - z_m)}{s^N (s - p_1) \ldots (s - p_{n-N})}. \tag{7.6}$$

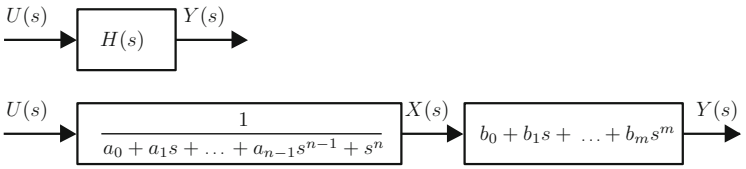

Fig. 7.1 Derivation of the state-space representation

The **zeros** z_ν are those values for which $H(s)$ becomes zero, whereas the **poles** $p_\nu \neq 0$ are singularities of $H(s)$. In case a pole and a zero are exactly equal, they cancel and do not influence the input–output behavior of the system. The gain can also be expressed as $K = b_m/a_n$.

As will be discussed in the following, the system represented by $H(s)$ is called **stable** if all poles have a negative real part, i.e., $\mathrm{Re}\{p_\nu\} < 0$ and $N = 0$. In this case, all poles lie in the open left half of the complex s-plane, which is referred to as OLHP. The abbreviations ORHP (open right half-plane), LHP (left half-plane), RHP (right half-plane) follow accordingly. If at least one pole has a positive real part, the system is **unstable**.

7.1.2 State-Space Representation

The higher-order ODE (7.5) can be rewritten as a system of ODEs of first order. Consider the transfer function $H(s)$ with input U and output Y, as shown in Fig. 7.1. The input variable $U(s)$ corresponds to $X(s)$ in the previous section. The notation is changed here to be consistent with the standard notation in the control system literature. Without loss of generality, it is assumed that $H(s)$ has the form (7.4) but with $a_n = 1$, i.e., the coefficients of $H(s)$ are normalized by $a_n \neq 0$. By splitting $H(s)$ in two blocks with its denominator and numerator, a new variable $X(s)$ may be defined as shown in Fig. 7.1.

In the time domain, the following ODEs can be derived from this block diagram:

$$u(t) = a_0 x(t) + a_1 \dot{x}(t) + \ldots + a_{n-1} \frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} x(t) + \frac{\mathrm{d}^n}{\mathrm{d}t^n} x(t),$$

$$y(t) = b_0 x(t) + b_1 \dot{x}(t) + \ldots + b_m \frac{\mathrm{d}^m}{\mathrm{d}t^m} x(t).$$

Defining the **states** (see also Sect. 2.8.1)

$$x_1 := x, \quad x_2 := \dot{x}, \quad \ldots, \quad x_n := \frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} x(t), \tag{7.7}$$

leads to the system of equations

$$
\begin{aligned}
\dot{x}_1(t) &= x_2(t),\\
\dot{x}_2(t) &= x_3(t),\\
&\vdots\\
\dot{x}_{n-1}(t) &= x_n(t),\\
\dot{x}_n(t) &= -a_0 x_1(t) - \ldots - a_{n-1} x_n(t) + u(t),\\
y(t) &= b_0 x_1(t) + \ldots + b_m x_{m+1}(t).
\end{aligned}
$$

With the definition of the **state vector**

$$\vec{x}(t) := \left[x_1(t)\ x_2(t)\ \ldots\ x_n(t)\right]^{\mathrm{T}},$$

the matrix representation

$$
\begin{aligned}
\frac{\mathrm{d}\vec{x}(t)}{\mathrm{d}t} &= \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & \ldots \\ & & & \ddots & \\ & & & & 1 \\ -a_0 & -a_1 & \ldots & & -a_{n-1} \end{bmatrix} \cdot \vec{x}(t) + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \cdot u(t),\\
y(t) &= \left[b_0 \ldots b_m\ 0 \ldots 0\right] \cdot \vec{x}(t),
\end{aligned}
$$

is obtained, which is called the **controllable canonical form** and is a special case of a **state space representation**. Different choices of the states (7.7) lead to different representations, but these have the general form

$$
\begin{aligned}
\frac{\mathrm{d}\vec{x}(t)}{\mathrm{d}t} &= A \cdot \vec{x}(t) + B \cdot \vec{u}(t),\\
\vec{y}(t) &= C \cdot \vec{x}(t),
\end{aligned}
$$

with the state vector $\vec{x}$ of dimension n, the input vector $\vec{u}$ of dimension p, the output vector $\vec{y}$ of dimension q, the $n \times n$ system matrix A, the $n \times p$ input matrix B, and the $q \times n$ output matrix C. All matrices are assumed to have constant and real elements. A feedthrough matrix for a direct influence of $\vec{u}$ on $\vec{y}$ can be avoided in most practical cases. The Laplace transform of these equations yields[2]

$$s\,\vec{X}(s) - \vec{x}(0) = A\,\vec{X}(s) + B\,\vec{U}(s),$$

[2]In the following, we write $\vec{x}(0)$ instead of $\vec{x}(0+)$ because we assume that the value at $t = 0$ is defined by the limit $t \to 0$ for positive values of t.

or

$$\vec{X}(s) = (s\,I - A)^{-1}\left(B\,\vec{U}(s) + \vec{x}(0)\right), \tag{7.8}$$

where I denotes the $n \times n$ identity matrix. The Laplace transform for the output $\vec{y}$ leads to

$$\vec{Y}(s) = C\,(s\,I - A)^{-1}\left(B\,\vec{U}(s) + \vec{x}(0)\right).$$

In case of a system with a single input and a single output (**SISO system**), the transfer function $H(s)$ is obtained as

$$H(s) = C(s\,I - A)^{-1}B,$$

where C is a row vector and B a column vector ($p = 1, q = 1$).

7.1.3 Linearization of Nonlinear Systems

Every practical system contains nonlinearities. Examples are nonlinear friction and constraints on the input that lead to saturation. Fortunately, in many cases, the considered nonlinear system behaves similarly to a linear system in the vicinity of its operating point. Consider a nonlinear system described by

$$\frac{\mathrm{d}\vec{x}(t)}{\mathrm{d}t} = \vec{v}(\vec{x}(t), \vec{u}(t))$$

with the analytic vector function $\vec{v}$. Suppose that $\vec{x} = \vec{x}_\mathrm{F}$ and $\vec{u} = \vec{u}_\mathrm{F}$ constitute a constant equilibrium point, i.e.,

$$\vec{v}(\vec{x}_\mathrm{F}, \vec{u}_\mathrm{F}) = 0.$$

With the use of the Jacobian matrix

$$\frac{\partial \vec{v}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial v_1}{\partial x_1} & \frac{\partial v_1}{\partial x_2} & \cdots & \frac{\partial v_1}{\partial x_n} \\ \frac{\partial v_2}{\partial x_1} & \frac{\partial v_2}{\partial x_2} & \cdots & \frac{\partial v_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial v_n}{\partial x_1} & \frac{\partial v_n}{\partial x_2} & \cdots & \frac{\partial v_n}{\partial x_n} \end{pmatrix},$$

the Taylor series expansion around the equilibrium can be written as

$$\frac{\mathrm{d}\vec{x}(t)}{\mathrm{d}t} = \vec{v}(\vec{x}_\mathrm{F}, \vec{u}_\mathrm{F}) + \frac{\partial \vec{v}}{\partial \vec{x}}\Big|_\mathrm{F} \cdot (\vec{x} - \vec{x}_\mathrm{F}) + \frac{\partial \vec{v}}{\partial \vec{u}}\Big|_\mathrm{F} \cdot (\vec{u} - \vec{u}_\mathrm{F}) + \vec{v}_\mathrm{ho}(\vec{x} - \vec{x}_\mathrm{F}, \vec{u} - \vec{u}_\mathrm{F}),$$

where $\big|_{\mathrm{F}}$ denotes the value at the equilibrium and $\vec{v}_{\mathrm{ho}}$ are higher-order terms. For small deviations

$$\Delta\vec{x}(t) = \vec{x}(t) - \vec{x}_{\mathrm{F}}, \quad \Delta\vec{u}(t) = \vec{u}(t) - \vec{u}_{\mathrm{F}}$$

from equilibrium, the higher-order terms may be neglected, and the linear system

$$\frac{\mathrm{d}\Delta\vec{x}(t)}{\mathrm{d}t} = A \cdot \Delta\vec{x}(t) + B \cdot \Delta\vec{u}(t)$$

with

$$A = \frac{\partial\vec{v}}{\partial\vec{x}}\Big|_{\mathrm{F}}, \quad B = \frac{\partial\vec{v}}{\partial\vec{u}}\Big|_{\mathrm{F}},$$

can be used as a linearization of the nonlinear system.

7.1.4 Dynamic Response of LTI Systems

The output of an LTI system depends on its transfer function $H(s)$ and on the input signal $u(t)$. In the following, the response of a general LTI system with respect to important test signals is discussed. This prepares the definition of stability. It is assumed that the poles and zeros of $H(s)$ are all distinct, apart from N poles at $s = 0$. In most cases, this is a valid assumption. The calculations for the case with poles or zeros of higher multiplicity are similar but more intricate. Because the coefficients in Eq. (7.4) are real, nonreal poles p or zeros z are always accompanied by their complex conjugate counterparts p^* and z^*. The complex conjugate operator commutes with every holomorphic function $f(x)$ on its domain of definition if $f(x)$ is real for real x. Thus in this case, $f^*(x)$ equals $f(x^*)$. In particular, this applies to every polynomial and rational function with real coefficients.

According to Eq. (7.6), the considered transfer function can be written as

$$H(s) = \frac{K\prod_{\nu=1}^{m_1}(s - z_{\mathrm{r},\nu})\prod_{\nu=1}^{m_2}(s - z_{\mathrm{c},\nu})(s - z_{\mathrm{c},\nu}^*)}{s^N\prod_{\nu=1}^{n_1}(s - p_{\mathrm{r},\nu})\prod_{\nu=1}^{n_2}(s - p_{\mathrm{c},\nu})(s - p_{\mathrm{c},\nu}^*)}, \tag{7.9}$$

where the $z_{\mathrm{r},\nu}$ and $p_{\mathrm{r},\nu}$ are the nonzero real zeros and poles, $z_{\mathrm{c},\nu}$ and $p_{\mathrm{c},\nu}$ are the nonzero complex zeros and poles, and K is the real gain. The total polynomial degree equals $m = m_1 + 2m_2$ for the numerator and $n = N + n_1 + 2n_2$ for the denominator. For a proper transfer function, $n \geq m$ holds.

7.1.4.1 Impulse Response

The impulse response is of practical interest for the study of pulse-shaped disturbances that may act on the feedback loop. In addition, this case is equivalent to the response of the state-space representation with zero input and certain nonzero initial conditions $\vec{x}(t=0) \neq 0$.

According to Eq. (7.3), the excitation of the system with the Dirac function

$$u(t) = \delta(t) \quad \circ\!\!-\!\!\bullet \quad U(s) = 1$$

yields

$$Y(s) = H(s) \cdot 1 = H(s)$$

in the Laplace domain. To calculate the response in the time domain, the partial fraction decomposition

$$Y(s) = H(s) \stackrel{!}{=} \sum_{\nu=0}^{N} \frac{K_{0,\nu}}{s^{\nu}} + \sum_{\nu=1}^{n_1} \frac{K_{\mathrm{r},\nu}}{s - p_{\mathrm{r},\nu}} + \sum_{\nu=1}^{n_2} \left(\frac{K_{\mathrm{c}1,\nu}}{s - p_{\mathrm{c},\nu}} + \frac{K_{\mathrm{c}2,\nu}}{s - p^*_{\mathrm{c},\nu}} \right) \tag{7.10}$$

is used. Here one assumes that the transfer function $H(s)$ is proper, i.e., $n \geq m$. The constants $K_{\mathrm{r},\nu}$ can be calculated as follows. Multiplying Eqs. (7.9) and (7.10) by $(s - p_{\mathrm{r},i})$ for a specific $i = 1, \ldots, n_1$ and setting $s = p_{\mathrm{r},i}$ leads to

$$\begin{aligned} K_{\mathrm{r},i} &= [H(s)\ (s - p_{\mathrm{r},i})]_{s=p_{\mathrm{r},i}} \\ &= \frac{K \prod_{\nu=1}^{m_1} (p_{\mathrm{r},i} - z_{\mathrm{r},\nu}) \prod_{\nu=1}^{m_2} (p_{\mathrm{r},i} - z_{\mathrm{c},\nu})(p_{\mathrm{r},i} - z^*_{\mathrm{c},\nu})}{p_{\mathrm{r},i}^N \prod_{\nu=1,\nu\neq i}^{n_1} (p_{\mathrm{r},i} - p_{\mathrm{r},\nu}) \prod_{\nu=1}^{n_2} (p_{\mathrm{r},i} - p_{\mathrm{c},\nu})(p_{\mathrm{r},i} - p^*_{\mathrm{c},\nu})}. \end{aligned}$$

The constants $K_{\mathrm{r},\nu}$ are always real, because in the denominator, the expression

$$(p_{\mathrm{r},i} - p_{\mathrm{c},\nu})(p_{\mathrm{r},i} - p^*_{\mathrm{c},\nu}) = (p_{\mathrm{r},i} - \mathrm{Re}\{p_{\mathrm{c},\nu}\})^2 + (\mathrm{Im}\{p_{\mathrm{c},\nu}\})^2$$

is real, and the same applies to the numerator. A similar calculation yields the constants

$$\begin{aligned} K_{\mathrm{c}1,i} &= [H(s)\ (s - p_{\mathrm{c},i})]_{s=p_{\mathrm{c},i}}\,, \\ K_{\mathrm{c}2,i} &= \left[H(s)\ (s - p^*_{\mathrm{c},i})\right]_{s=p^*_{\mathrm{c},i}}\,, \end{aligned}$$

and using the above-mentioned commutability property of the complex conjugate operator leads to

$$K_{\mathrm{c}2,i} = K^*_{\mathrm{c}1,i}.$$

The constant $K_{0,N}$ is obtained by multiplying by s^N; it reads

$$K_{0,N} = \left[H(s)\, s^N\right]_{s=0} = \frac{K \prod_{\nu=1}^{m_1}(-z_{\mathrm{r},\nu}) \prod_{\nu=1}^{m_2} |z_{\mathrm{c},\nu}|^2}{\prod_{\nu=1}^{n_1}(-p_{\mathrm{r},\nu}) \prod_{\nu=1}^{n_2} |p_{\mathrm{c},\nu}|^2}.$$

For the remaining constants $K_{0,i}$, a system of N linear equations is obtained by evaluating Eqs. (7.9) and (7.10) at N points $s = s_i$ that are different from the zeros and poles of the system. The constant $K_{0,0}$ is zero for strictly proper transfer functions $H(s)$, i.e., for $n > m$.

The transformation of Eq. (7.10) into the time domain

$$Y(s) = H(s) \quad \bullet\!\!-\!\!\circ \quad y(t) = h(t)$$

yields the impulse response

$$h(t) = K_{0,0}\,\delta(t) + \sum_{\nu=1}^{N} K_{0,\nu} \frac{t^{\nu-1}}{(\nu-1)!} + \sum_{\nu=1}^{n_1} K_{\mathrm{r},\nu}\, e^{p_{\mathrm{r},\nu} t} + \sum_{\nu=1}^{n_2} \left(K_{\mathrm{c1},\nu}\, e^{p_{\mathrm{c},\nu} t} + K^*_{\mathrm{c1},\nu}\, e^{p^*_{\mathrm{c},\nu} t} \right),$$

as Table A.4 shows ($\Theta(t)$ is omitted for the sake of simplicity). The elements of the last sum can be rewritten as

$$e^{\mathrm{Re}\{p_{\mathrm{c},\nu}\}t} \left[K_{\mathrm{c1},\nu}\, e^{j\,\mathrm{Im}\{p_{\mathrm{c},\nu}\}t} + K^*_{\mathrm{c1},\nu}\, e^{-j\,\mathrm{Im}\{p_{\mathrm{c},\nu}\}t} \right],$$

and the term of this expression in square brackets is equal to

$$|K_{\mathrm{c1},\nu}| \left(e^{j\left(\mathrm{Im}\{p_{\mathrm{c},\nu}\}t + \measuredangle K_{\mathrm{c1},\nu}\right)} + e^{-j\left(\mathrm{Im}\{p_{\mathrm{c},\nu}\}t + \measuredangle K_{\mathrm{c1},\nu}\right)} \right),$$

where $|K|$ and $\measuredangle K$ are the amplitude and phase of the complex number K, respectively. Altogether, the impulse response for $t \geq 0$ is

$$\begin{aligned} h(t) = K_{0,0}\,\delta(t) + \sum_{\nu=1}^{N} K_{0,\nu} \frac{t^{\nu-1}}{(\nu-1)!} + \sum_{\nu=1}^{n_1} K_{\mathrm{r},\nu}\, e^{p_{\mathrm{r},\nu} t} + \\ + 2 \sum_{\nu=1}^{n_2} |K_{\mathrm{c1},\nu}|\, e^{\mathrm{Re}\{p_{\mathrm{c},\nu}\}t} \cos\left(\mathrm{Im}\{p_{\mathrm{c},\nu}\}t + \measuredangle K_{\mathrm{c1},\nu}\right), \end{aligned} \tag{7.11}$$

and it tends to zero as $t \to \infty$ if all poles have negative real parts.

7.1.4.2 Step Response

The response $y(t) = y_\Theta(t)$ to a step command

$$u(t) = \Theta(t)$$

can be calculated in an analogous way with $U(s) = 1/s$. An alternative is the use of the convolution integral

$$y_\Theta(t) = \int_0^t h(\tau)\,\Theta(t-\tau)\,\mathrm{d}\tau = \int_0^t h(\tau)\,\mathrm{d}\tau. \tag{7.12}$$

With

$$\begin{aligned}\int_0^t e^{\delta\tau}\,\cos(\omega\tau+\varphi)\,\mathrm{d}\tau &= \frac{\delta}{\delta^2+\omega^2}\left(e^{\delta t}\,\cos(\omega t+\varphi)-\cos\varphi\right)\\ &\quad+\frac{\omega}{\delta^2+\omega^2}\left(e^{\delta t}\,\sin(\omega t+\varphi)-\sin\varphi\right)\\ &= \frac{1}{\sqrt{\delta^2+\omega^2}}\left(e^{\delta t}\,\cos\left(\omega t+\varphi-\arctan\frac{\omega}{\delta}\right)\right.\\ &\quad\left.-\cos\left(\varphi-\arctan\frac{\omega}{\delta}\right)\right),\end{aligned}$$

the integration of (7.11) for $t \geq 0$ yields

$$\begin{aligned}y_\Theta(t) &= K_{0,0}+\sum_{\nu=1}^{N}K_{0,\nu}\frac{t^\nu}{\nu!}+\sum_{\nu=1}^{n_1}\frac{K_{\mathrm{r},\nu}}{p_{\mathrm{r},\nu}}\left(e^{p_{\mathrm{r},\nu}t}-1\right)+\\ &\quad+\sum_{\nu=1}^{n_2}\frac{2|K_{\mathrm{c1},\nu}|}{|p_{\mathrm{c},\nu}|}\left(e^{\mathrm{Re}\{p_{\mathrm{c},\nu}\}t}\,\cos(\mathrm{Im}\{p_{\mathrm{c},\nu}\}t+\measuredangle K_{\mathrm{c1},\nu}-\measuredangle p_{\mathrm{c},\nu})\right.\\ &\quad\left.-\cos(\measuredangle K_{\mathrm{c1},\nu}-\measuredangle p_{\mathrm{c},\nu})\right).\end{aligned} \tag{7.13}$$

This calculation shows that the step response y_Θ will approach a finite value for large times t if and only if the conditions

$$N = 0, \quad p_{\mathrm{r},\nu} < 0, \quad \mathrm{Re}\{p_{\mathrm{c},\nu}\} < 0$$

are satisfied, i.e., all poles have negative real parts. Because the limit $\lim_{t\to\infty} y_\Theta(t)$ is then finite, the final value theorem can be applied:

$$\lim_{t\to\infty} y_\Theta(t) = \lim_{s\to 0} s\,Y_\Theta(s) = H(0).$$

The initial value theorem leads to

$$\lim_{t\to+0} y_\Theta(t) = K_{0,0} = \lim_{s\to\infty} s\,Y_\Theta(s) = \lim_{s\to\infty} H(s) = \begin{cases} 0 & \text{if } n > m, \\ K & \text{if } n = m. \end{cases}$$

For this reason, systems with $n = m$ are also said to have direct feedthrough. In contrast, strictly proper transfer functions with $n > m$ have a continuous output response at $t = 0$.

7.1.4.3 Frequency Response

An important test signal is the harmonic excitation

$$u(t) = \sin(\omega t) \quad \circ\!\!-\!\!\bullet \quad \frac{\omega}{s^2+\omega^2}.$$

The amplitude of the test signal may be chosen arbitrarily because of the linearity property (7.2). If it is assumed that none of the poles of $H(s)$ is equal to $\pm j\omega$, the decomposition of the output response in the Laplace domain can be written as

$$Y(s) = H(s)\,\frac{\omega}{s^2+\omega^2} = \frac{K_\omega}{s-j\omega} + \frac{K_\omega^*}{s+j\omega} + Y_{\text{trans}}(s),$$

where Y_{trans} has the same structure as the expression in Eq. (7.10), but with $K_{0,0} = 0$. A multiplication by $(s - j\omega)$ and the evaluation at $s = j\omega$ leads to

$$K_\omega = \left[H(s)\,\frac{\omega(s-j\omega)}{s^2+\omega^2}\right]_{s=j\omega} = \left[H(s)\,\frac{\omega}{s+j\omega}\right]_{s=j\omega} = \frac{1}{2j}H(j\omega).$$

In the time domain, the output response reads

$$\begin{aligned} y(t) &= K_\omega e^{j\omega t} + K_\omega^* e^{-j\omega t} + y_{\text{trans}}(t) \\ &= H(j\omega)\frac{e^{j\omega t}}{2j} - H^*(j\omega)\frac{e^{-j\omega t}}{2j} + y_{\text{trans}}(t) \\ &= |H(j\omega)|\,\sin(\omega t + \measuredangle H(j\omega)) + y_{\text{trans}}(t). \end{aligned}$$

If the transfer function $H(s)$ has only poles with negative real parts, the transient response y_{trans} will tend to zero, and $y(t)$ tends to a constant oscillation. The amplitude and phase of this oscillation with respect to the excitation $u(t)$ is determined by $H(j\omega)$, i.e., the value of the transfer function at $s = j\omega$. Because of the linearity property, this also applies to any shifted or scaled sinusoidal excitation. For this reason, the function $H(j\omega)$ depending on the frequency ω is called the

frequency response of the system $H(s)$ and is obtained by introducing $s = j\omega$ into $H(s)$.

There are two main reasons why the frequency response is important for feedback systems. First, $H(j\omega)$ can easily be measured by exciting the system with different frequencies ω, even if the transfer function $H(s)$ of the physical system is not known. Second, $H(j\omega)$ can be used for the stability analysis of the closed feedback loop with the Nyquist criterion (see Sect. 7.4.2).

So far, it has been assumed that $j\omega$ is not a pole of $H(s)$. Without further calculation, it can be reasoned that if $j\omega$ is a pole, $H(s)$ has a singularity at $H(j\omega)$ and the excitation with frequencies close to ω will lead to very large amplitudes. If the chosen frequency is exactly ω, this will result in a perfect resonance, and the oscillation at the output will grow without bound, although the input is a bounded signal.

7.1.4.4 General Input Function

In the previous sections, the Laplace transform was used to calculate specific output responses for SISO systems. In case of general input functions, **multiple-input and multiple-output (MIMO) systems**, or initial values, it is often more convenient to consider the state-space representation. In Sect. 2.8.6, it was shown that autonomous linear systems of differential equations

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = A \cdot \vec{r} \qquad \vec{r}(0) = \vec{r}_0$$

have the solution (2.99),

$$\vec{r}(t) = e^{tA}\,\vec{r}_0,$$

where e^{tA} is the matrix exponential function. In the presence of an input vector $\vec{u}(t)$, the system is no longer autonomous in general. The input may be a control effort or a disturbance such as a noise signal. In Sect. 7.1.2, the Laplace domain solution of a system with inputs was given by Eq. (7.8) as

$$\vec{X}(s) = (sI - A)^{-1}\vec{x}(0) + (sI - A)^{-1}B\vec{U}(s).$$

Comparing this with the solution $\vec{r}(t)$ of the autonomous system, it is apparent that

$$(sI - A)^{-1} \quad \bullet\!\!-\!\!\circ \quad e^{tA}$$

must hold, i.e., we have found the Laplace transform of the matrix exponential function. Transforming $\vec{X}(s)$ into the time domain thus leads to

$$\vec{x}(t) = e^{tA}\vec{x}(0) + \int_0^t e^{(t-\tau)A}B\vec{u}(\tau)\,\mathrm{d}\tau.$$

It can be shown that the matrix exponential function has the following properties, similar to those of an ordinary exponential function (cf. [11]):

- series representation: $e^{tA} = I + \sum_{\nu=1}^{\infty} A^{\nu} \frac{t^{\nu}}{\nu!}$
- inverse: $\left(e^{tA}\right)^{-1} = e^{-tA}$
- multiplication: $e^{t_2 A} e^{t_1 A} = e^{(t_2+t_1)A}$

7.1.5 Stability

In Sects. 2.8.6 and 2.8.10, it was shown that a linear autonomous system is asymptotically stable if and only if all eigenvalues of the system matrix A have negative real parts, i.e., are situated in the OLHP. Equivalently, the same holds for the roots of the characteristic equation. Asymptotic stability for autonomous systems implies that a trajectory that starts at some initial value will tend to a fixed point.

For a system with nonzero inputs $\vec{u}(t)$, this definition may not be sufficient. The input can be a persistent disturbance with a certain amplitude that prevents the system from approaching the fixed point. For a feedback system, it is, however, necessary that the states $\vec{x}(t)$ or the output $\vec{y}(t)$ remain bounded. This motivates the following definition:

Definition 7.1. A dynamical system

$$\frac{\mathrm{d}\vec{x}(t)}{\mathrm{d}t} = \vec{v}_1(\vec{x}(t), \vec{u}(t)), \quad \vec{y}(t) = \vec{v}_2(\vec{x}(t), \vec{u}(t))$$

with input $\vec{u}(t)$, states $\vec{x}(t)$, and output $\vec{y}(t)$ is assumed to be in equilibrium for $t = t_0$ with arbitrary real t_0, i.e., $\vec{x}(t_0) = \vec{x}_\mathrm{F}$, where $\vec{x}_\mathrm{F}$ is a fixed point. This fixed point is said to be **bounded-input bounded-output (BIBO) stable** if for every finite c_1 with $\|\vec{u}(t)\| < c_1$ for $t \geq t_0$, there exists a finite c_2 such that $\|\vec{y}(t)\| \leq c_2$ for $t \geq t_0$.

(See, e.g., Ludyk [11, Definition 3.37, p. 159].)

The step response (7.13) shows that $y_\Theta(t)$ is bounded if all poles of $H(s)$ have negative real parts. Because of Eq. (7.12), this is also true if

$$|y_\Theta(t)| = \left| \int_0^t h(\tau)\, \mathrm{d}\tau \right| \leq \int_0^t |h(\tau)|\, \mathrm{d}\tau \leq \int_0^\infty |h(\tau)|\, \mathrm{d}\tau \leq c_2 < \infty$$

holds, i.e., if the impulse response $h(t)$ is absolutely integrable.

In general, the following theorem holds.

Theorem 7.2. *An LTI SISO system is BIBO stable if and only if the following (equivalent) conditions are satisfied:*

- *The transfer function $H(s)$ has only poles with negative real parts.*
- *the impulse response $h(t)$ is absolutely integrable.*

(See, e.g., Ludyk [11, Theorems 3.39 and 3.40, p. 160].)

In addition, there is a close relationship between BIBO and asymptotic stability. The transfer function $H(s)$ can be written as

$$H(s) = C(sI - A)^{-1}B = \frac{C\,\mathrm{adj}(sI - A)\,B}{\det(sI - A)},$$

where $\mathrm{adj}(A)$ denotes the adjugate[3] matrix of A. Thus, the poles of $H(s)$ are obtained by calculating the roots of the characteristic equation

$$\det(sI - A) = 0,$$

and these are identical to the eigenvalues of A. However, due to pole–zero cancelations, the poles are, in general, a subset of the eigenvalues of A, i.e., not every eigenvalue is a pole of $H(s)$. If A has only eigenvalues with negative real parts, the system is asymptotically stable, and this always implies that the poles have negative real parts. This consideration leads to the following theorem:

Theorem 7.3. *An LTI system that is asymptotically stable is also BIBO stable, but a BIBO stable system is not always asymptotically stable.*

(See, e.g., Ludyk [Theorem 3.41, p. 160][11].)

7.2 Standard Closed Loop

The block diagram in Fig. 7.2 is called the standard feedback loop. It has one input and one output and is thus also called a **single-input single-output (SISO)** system.

The feedback system can be described by the following equations:

$$\begin{aligned} Y(s) &= X_{\mathrm{d2}}(s) + H_{\mathrm{p}}(s)\Big[X_{\mathrm{d1}}(s) + U(s)\Big], \\ U(s) &= H_{\mathrm{c}}(s)X_{\mathrm{e}}(s), \\ X_{\mathrm{e}}(s) &= Y_{\mathrm{r}}(s) - H_{\mathrm{m}}(s)\Big[X_{\mathrm{d3}}(s) + Y(s)\Big]. \end{aligned}$$

Solving these equations for the output $Y(s)$ leads to

[3]The cofactor matrix of A is a matrix that consists of the (i, k) minors of A multiplied by the factor $(-1)^{i+k}$. The adjugate matrix of A is the transpose of the cofactor matrix of A.

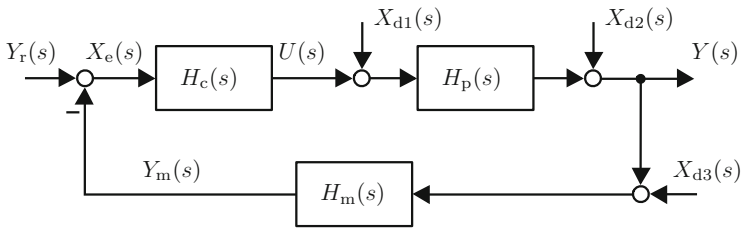

Fig. 7.2 Standard feedback loop: transfer functions of process $H_p(s)$, controller $H_c(s)$, and measurement $H_m(s)$. The signals are reference $Y_r(s)$, control error $X_e(s)$, input $U(s)$, output $Y(s)$, and disturbances $X_{d1}(s)$, $X_{d2}(s)$, and $X_{d3}(s)$

$$Y(s) = H_{ry}(s)\Big[Y_r(s) - H_m(s)X_{d3}(s)\Big] + H_{dy}(s)\Big[H_p(s)X_{d1}(s) + X_{d2}(s)\Big] \tag{7.14}$$

with the reference to output transfer function

$$H_{ry}(s) = \frac{H_p(s)H_c(s)}{1 + H_p(s)H_c(s)H_m(s)} \tag{7.15}$$

and the disturbance to output transfer function

$$H_{dy}(s) = \frac{1}{1 + H_p(s)H_c(s)H_m(s)}. \tag{7.16}$$

A **unity feedback system** has $H_m(s) = 1$, and in this case, the disturbance to output transfer function

$$H_{dy}(s) = \frac{1}{1 + H_p(s)H_c(s)}$$

is also called the **sensitivity function**, and the reference to output transfer function

$$H_{ry}(s) = \frac{H_p(s)H_c(s)}{1 + H_p(s)H_c(s)}$$

is the **complementary sensitivity function**. Note that $H_{dy}(s) + H_{ry}(s) = 1$.

Usually, the process transfer function $H_p(s)$ has to be determined in a separate modeling step before the analysis or the design of the feedback loop. The modeling can be based on analytical equations if the underlying physical principles are well known. If this is not the case, measurements may be used for a system identification. In both cases, modeling assumptions have to be made to limit the complexity of the system. Often, nonlinearities in the feedback loop are linearized, and high-frequency dynamics are omitted.

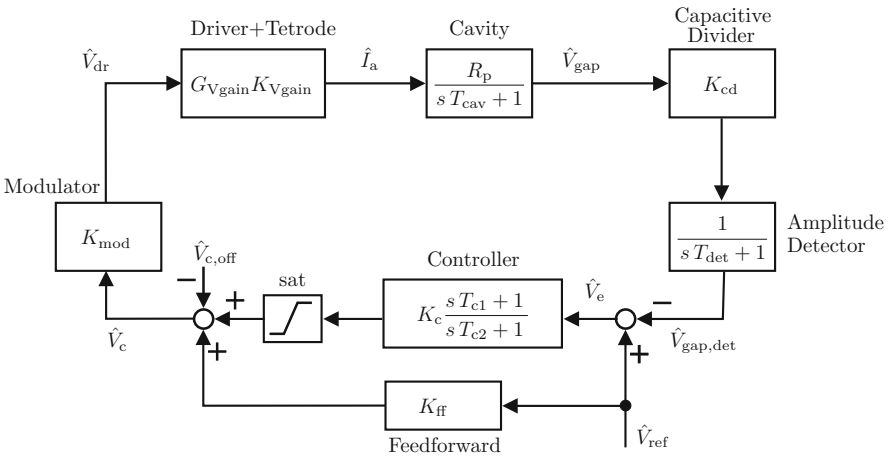

Fig. 7.3 Model of the amplitude feedback loop

7.3 Example: Amplitude Feedback

As a realistic example of a feedback loop, the amplitude feedback control of a ferrite-loaded cavity will be considered. The feedback is needed to hold the amplitude $\hat{V}_{\mathrm{gap}}$ of the RF voltage close to a given reference value $\hat{V}_{\mathrm{gap,ref}}$. In our example, the cavity feedback loop behaves highly nonlinearly with respect to the RF frequency f_{RF} and the reference amplitude $\hat{V}_{\mathrm{ref}}$. In the following, the operating point

$$f_{\mathrm{RF}} = 3\,\mathrm{MHz}, \quad \hat{V}_{\mathrm{gap,ref}} = 2\,\mathrm{kV},$$

will be considered. A model of the feedback loop was obtained in [12] based on measurements, and the corresponding block diagram is shown in Fig. 7.3. In the following, only amplitudes of RF signals are used, not the RF signals themselves.

The feedback loop consists of the following subcomponents:

- The **cavity** is driven by the anode current with the amplitude $\hat{I}_{\mathrm{a}}$. The amplitude of the resulting gap voltage $\hat{V}_{\mathrm{gap}}$ acts approximately as a first-order system (PT_1) with respect to $\hat{I}_{\mathrm{a}}$ (see also Appendix A.7.1). The "gain" is equal[4] to $R_{\mathrm{p}} \approx 2700\,\Omega$, and the time constant is $T_{\mathrm{cav}} \approx 4\,\mu\mathrm{s}$. The set points are $\hat{V}_{\mathrm{gap}} = 2\,\mathrm{kV}$ and $\hat{I}_{\mathrm{a}} = 0.75\,\mathrm{A}$.
- A **capacitive divider** is used to downscale the gap voltage of one-half the gap with a factor of 1000. This has no significant influence on the time constants in the loop. With respect to the total gap voltage, the scaling is $K_{\mathrm{cd}} = 1/2000$.

[4] Due to the output impedance of the tetrode, this value is about one-half the pure cavity impedance specified in Table 4.1 on p. 198.

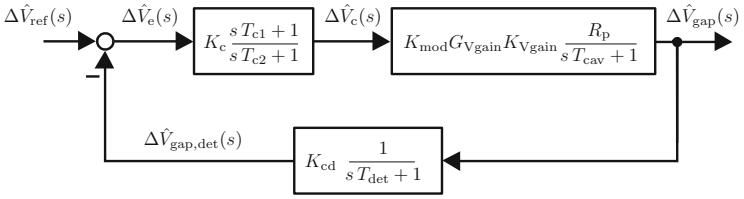

Fig. 7.4 Small-signal model of the amplitude feedback loop

- An **amplitude detector** with time constant $T_{\mathrm{det}} = 5\,\mu\mathrm{s}$ is used to obtain the amplitude $\hat{V}_{\mathrm{gap,det}}$. This amplitude is then compared to the reference $\hat{V}_{\mathrm{ref}}$. The set points are $\hat{V}_{\mathrm{gap,det}} = 1\,\mathrm{V}$ and $\hat{V}_{\mathrm{ref}} = 1.04\,\mathrm{V}$.
- The parameters of the **controller** are $K_{\mathrm{c}} = 14.9$, $T_{\mathrm{c1}} = 17.2\,\mu\mathrm{s}$, and $T_{\mathrm{c2}} = 487.2\,\mu\mathrm{s}$. A saturation limit **sat** follows that limits the control output to $\pm 7.23\,\mathrm{V}$. The offset voltage is $\hat{V}_{\mathrm{c,off}} = 0.2\,\mathrm{V}$. In the **feedforward** loop, the gain is $K_{\mathrm{ff}} = 0.6$. According to these values, the set point of the control effort is $\hat{V}_{\mathrm{c}} = 1.02\,\mathrm{V}$.
- The (amplitude) **modulator** produces a sinusoidal signal modulated with $\hat{V}_{\mathrm{c}}$. The sinusoidal signal with initial amplitude 0.316 V (0 dBm) is damped with a factor of −12.2 dB; this corresponds to a factor of 0.245 for the voltage amplitude. Altogether, the modulator can be modeled as a gain $K_{\mathrm{mod}} = 0.316{\cdot}0.245$. Hence, the set point of the driving voltage is $\hat{V}_{\mathrm{dr}} = 79\,\mathrm{mV}$.
- The gains of the **driver and tetrode amplifiers** depend on the RF frequency and the amplitude of the gap voltage. For the chosen setting, we have $G_{\mathrm{Vgain}} \approx 27\,\mathrm{S}$ and $K_{\mathrm{Vgain}} \approx 0.35$.

Signal time delays with a magnitude of about 1 μs are neglected in the following. However, they would be important for larger feedback gains.

The given set-point values were obtained by choosing $\hat{V}_{\mathrm{gap}} = 2\,\mathrm{kV}$. Because the stationary gain of the cavity transfer function is R_{p}, the necessary anode current amplitude equals $\hat{I}_{\mathrm{a}} = \hat{V}_{\mathrm{gap}}/R_{\mathrm{p}}$. All other set-point values in the feedback loop follow accordingly. This results in a reference $\hat{V}_{\mathrm{ref}}$ that is slightly higher than $\hat{V}_{\mathrm{gap,det}}$ and thus in a stationary control error $\hat{V}_{\mathrm{e}} = 40\,\mathrm{mV}$. This steady-state error could be avoided by introducing an integral controller in the loop. However, it is also possible to adjust the reference in such a way that the desired value $\hat{V}_{\mathrm{gap}}$ is reached, as has been done in this case.

The system is nonlinear due to the saturation function. This function and the offset values $\hat{V}_{\mathrm{ref}}$ and $\hat{V}_{\mathrm{c,off}}$ can be neglected if only small deviations with respect to the set point are considered. This leads to the linearized feedback loop in standard notation, as shown in Fig. 7.4 with amplitude error $\Delta\hat{V}_{\mathrm{gap}} = \hat{V}_{\mathrm{gap}} - \hat{V}_{\mathrm{gap,ref}}$ and reference $\Delta\hat{V}_{\mathrm{ref}} = 0$. Similarly, all other values are defined relative to their set-point values, e.g., the relative control effort is $\Delta\hat{V}_{\mathrm{c}} = \hat{V}_{\mathrm{c}} - 1.02\,\mathrm{V}$.

A calculation of the reference to output transfer function according to (7.15) yields

$$H_{\mathrm{ry}}(s) = \frac{K_{\mathrm{c}} K_{\mathrm{mod}} G_{\mathrm{Vgain}} K_{\mathrm{Vgain}} R_{\mathrm{p}} (sT_{\mathrm{c1}} + 1)(sT_{\mathrm{det}} + 1)}{(sT_{\mathrm{det}} + 1)(sT_{\mathrm{c2}} + 1)(sT_{\mathrm{cav}} + 1) + K_{\mathrm{c}} K_{\mathrm{mod}} G_{\mathrm{Vgain}} K_{\mathrm{Vgain}} R_{\mathrm{p}} (sT_{\mathrm{c1}} + 1) K_{\mathrm{cd}}}.$$

A zero-pole-gain representation of this transfer function can be obtained by a numerical calculation of the poles and zeros. The gain is equal to the ratio of the factors of the highest order in s in the numerator and denominator. For the amplitude loop, these orders are s^2 and s^3, respectively, and the gain is

$$K = \frac{K_{\mathrm{c}} K_{\mathrm{mod}} G_{\mathrm{Vgain}} K_{\mathrm{Vgain}} R_{\mathrm{p}} T_{\mathrm{c1}} T_{\mathrm{det}}}{T_{\mathrm{det}} T_{\mathrm{c2}} T_{\mathrm{cav}}} = 2.6 \cdot 10^8 \ s^{-1}.$$

The resulting zero-pole-gain representation is

$$H_{\mathrm{ry}}(s) = \frac{\Delta \hat{V}_{\mathrm{gap}}(s)}{\Delta \hat{V}_{\mathrm{ref}}(s)} = 2.6 \cdot 10^8 \, \mathrm{s}^{-1} \ \frac{(s - z_1)(s - z_2)}{(s - p_1)(s - p_2)(s - p_3)}$$

with zeros

$$z_1 = -5.81 \cdot 10^4 \, \mathrm{s}^{-1}, \quad z_2 = -2 \cdot 10^5 \, \mathrm{s}^{-1},$$

and poles

$$p_1 = -2.42 \cdot 10^4 \, \mathrm{s}^{-1}, \quad p_{2,3} = -(2.14 \pm j \ 1.44) \cdot 10^5 \, \mathrm{s}^{-1}.$$

Thus, the closed-loop system is BIBO stable. The pole p_1 is closest to the imaginary axis and dominates the dynamics of the feedback. The dominating pole corresponds to a closed-loop bandwidth and a time constant of

$$\omega_{\mathrm{ry}} = -p_1 = 2.42 \cdot 10^4 \, \mathrm{s}^{-1} \quad \Rightarrow \quad T_{\mathrm{ry}} = -\frac{1}{p_1} \approx 40 \, \mu\mathrm{s}.$$

The absolute values of the remaining poles are larger by an order of magnitude. They are thus negligible for a first rough evaluation of the closed-loop dynamics.

7.4 Analysis and Stability

The closed-loop transfer function

$$H_{\mathrm{ry}}(s) = \frac{b_0 + b_1 s + b_2 s^2}{a_0 + a_1 s + a_2 s^2 + s^3}$$

can be obtained from the given open-loop transfer function using only basic manipulations. The calculation of the poles p_i from the characteristic equation

$$0 = a_0 + a_1 s + a_2 s^2 + s^3$$

is a more complex task, and numerical computations are necessary for higher-order systems in general. For a stability analysis, one may, however, not be interested in the exact values of the poles, but only in the decision whether all poles have negative real parts. There are several **stability criteria** that can be applied without solving the characteristic equation directly. The Hurwitz and Nyquist criteria will be presented in the next sections.

7.4.1 Routh–Hurwitz Stability Criterion

The **Routh–Hurwitz criterion** is a necessary and sufficient condition for the roots of the polynomial

$$a_0 + a_1 s + \ldots + a_{n-1} s^{n-1} + s^n \tag{7.17}$$

to have only negative real parts, in which case the polynomial is then called a **Hurwitz polynomial**. The criterion is of particular interest if the coefficients a_i contain undetermined parameters. An example of such a parameter is the controller gain in the feedback loop. With the Routh–Hurwitz criterion, inequalities in these parameters can then be obtained for the closed loop to be stable.

A first necessary condition is given by the following theorem:

Theorem 7.4. *If the polynomial* (7.17) *is Hurwitz, then it has only positive coefficients* $a_i > 0$, $i = 0, 1, \ldots, n-1$.

(See, e.g., Ludyk [11, Theorem 3.43, p. 161].)

This enables a first simple test whether a polynomial can be Hurwitz. If any of the coefficients is missing, i.e., $a_i = 0$, or any a_i is negative, there will be roots with nonnegative real part, and the polynomial is not Hurwitz.

A necessary and sufficient condition is presented by the Hurwitz criterion. It uses the $\nu \times \nu$ Hurwitz determinants

$$H_\nu := \det \begin{pmatrix} a_{n-1} & a_{n-3} & a_{n-5} & \ldots & a_{n-2\nu+1} \\ 1 & a_{n-2} & a_{n-4} & \ldots & a_{n-2\nu+2} \\ 0 & a_{n-1} & a_{n-3} & \ldots & a_{n-2\nu+3} \\ 0 & 1 & a_{n-2} & \ldots & a_{n-2\nu+4} \\ 0 & 0 & a_{n-1} & \ldots & a_{n-2\nu+5} \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots & a_{n-\nu} \end{pmatrix}, \tag{7.18}$$

where the coefficients a_i in the matrix with an index $i < 0$ are set to zero. As an example, the first three determinants for a polynomial with degree $n \geq 5$ are

$$H_1 := a_{n-1},$$

$$H_2 := \det \begin{pmatrix} a_{n-1} & a_{n-3} \\ 1 & a_{n-2} \end{pmatrix},$$

$$H_3 := \det \begin{pmatrix} a_{n-1} & a_{n-3} & a_{n-5} \\ 1 & a_{n-2} & a_{n-4} \\ 0 & a_{n-1} & a_{n-3} \end{pmatrix}.$$

The Hurwitz criterion is given by the following theorem:

Theorem 7.5. *The polynomial* (7.17) *is Hurwitz if and only if the Hurwitz determinants* H_ν *defined by* (7.18) *are positive for* $\nu = 1, \ldots, n$.

(See, e.g., Gantmacher [13].)

A simplified version of this theorem needs only half the determinants:

Theorem 7.6. *Suppose that all the coefficients of the polynomial* (7.17) *are positive. For odd n, the polynomial is Hurwitz if and only if the Hurwitz determinants* $H_2, H_4, \ldots, H_{n-1}$ *are positive. For even n, the polynomial is Hurwitz if and only if the Hurwitz determinants* $H_3, H_5, \ldots, H_{n-1}$ *are positive.*

(See, e.g., Gantmacher [13].)

Consider as an example the amplitude feedback introduced in Sect. 7.3. The denominator of the closed-loop transfer function reads

$$a_0 + a_1 s + a_2 s^2 + s^3$$

with

$$a_0 = 1.6129 \cdot 10^{15}\,\mathrm{s}^{-3}, \quad a_1 = 7.6901 \cdot 10^{10}\,\mathrm{s}^{-2}, \quad a_2 = 4.5205 \cdot 10^{5}\,\mathrm{s}^{-1}.$$

In the following, the physical units of these coefficients will be ignored to avoid confusion with the Laplace variable s. Since all coefficients are positive, this polynomial with $n = 3$ is Hurwitz, because

$$H_2 = \det \begin{pmatrix} a_2 & a_0 \\ 1 & a_1 \end{pmatrix} = a_2 a_1 - a_0 = 3.3150 \cdot 10^{16} > 0.$$

Now assume that the feedback gain K_c in the loop of Fig. 7.4 is a free parameter. As a consequence, the coefficients a_0 and a_1 become parameter-dependent:

$$a_0 = 1.0136 \cdot 10^{14} K_\mathrm{c} + 1.0263 \cdot 10^{14}, \quad a_1 = 1.7435 \cdot 10^{9} K_\mathrm{c} + 5.0924 \cdot 10^{10}.$$

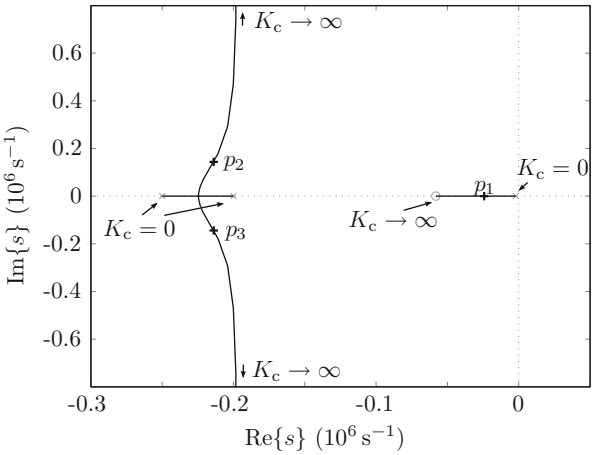

Fig. 7.5 Root locus of the amplitude feedback. The poles p_1, p_2, and p_3 are obtained for $K_c = 14.9$

The Hurwitz criterion now leads to the conditions

$$\begin{aligned} a_0 > 0 \quad &\Rightarrow \quad K_c > -1.01, \\ a_1 > 0 \quad &\Rightarrow \quad K_c > -29.21, \\ H_2 > 0 \quad &\Rightarrow \quad K_c > -33.37. \end{aligned}$$

Thus, the feedback loop is stable for $K_c > -1.01$. Due to the stability of the open-loop system, the closed-loop system obviously remains stable even if the feedback gain is slightly negative. A positive feedback gain K_c, however, is the typical case for the amplitude control. Figure 7.5 shows the closed-loop poles in the complex s-plane as a function of the positive gain $K_c > 0$. This type of diagram is also referred to as a the **root locus**. For $K_c = 0$, the closed-loop poles are equal to the open-loop poles

$$p_1(0) = -\frac{1}{T_{c2}}, \quad p_2(0) = -\frac{1}{T_{det}}, \quad p_3(0) = -\frac{1}{T_{cav}}$$

that are obtained from the open-loop transfer function (cf. Fig. 7.4)

$$H_{open}(s) = \frac{\Delta\hat{V}_{gap,det}(s)}{\Delta\hat{V}_e(s)}.$$

For increasing K_c, the closed-loop pole p_1 moves to the left toward the open-loop zero

$$z_1(0) = -\frac{1}{T_{c1}}$$

of $H_{open}(s)$, whereas the poles p_2 and p_3 approach each other and for a certain K_c between 0 and 14.9, a complex conjugate pole pair arises. The root locus indicates that the closed loop remains stable also for higher $K_c \to \infty$, because all three branches of the root locus remain in the OLHP. Since the branches of the root locus are the positions of the closed-loop poles,[5] the closed loop is stable. This is in agreement with the result of the Hurwitz criterion.

Please note that for a practical implementation, very large feedback gains K_c would not be recommendable for several reasons:

- For sufficiently large gains, the complex pair $p_{2,3}$ dominates the dynamics of the loop, resulting in an unacceptable oscillatory behavior.
- Large gains may increase disturbances, especially the measured noise.
- The feedback of the real system may become unstable for very large gains due to unmodeled high-frequency dynamics and delays.

7.4.2 Bode Plots and Nyquist Criterion

The Hurwitz stability criterion is based on the characteristic equation, i.e., on the denominator polynomial of the closed-loop transfer function. The **Bode plots** and the **Nyquist criterion** are approaches that are different in the sense that they rely on the open-loop transfer function

$$H_{open}(s) := H_c(s)\, H_p(s)\, H_m(s)$$

of the standard feedback loop; cf. Fig. 7.2. Consider as an example the system

$$H_{open}(s) = \frac{K\left(1 - \frac{s}{z_1}\right)}{s^N \left(1 - \frac{s}{p_1}\right)\left(1 - \frac{s}{p_2}\right)\left(1 - \frac{s}{p_2^*}\right)}. \tag{7.19}$$

This system is assumed to have a real zero $z_1 \neq 0$, a real pole $p_1 \neq 0$, a complex pole pair p_2 and p_2^*, and N poles at $s = 0$. The frequency response of $H_{open}(s)$ is given by

$$H_{open}(j\omega) = \frac{K\left(1 - \frac{j\omega}{z_1}\right)}{(j\omega)^N \left(1 - \frac{j\omega}{p_1}\right)\left(1 - \frac{j\omega}{p_2}\right)\left(1 - \frac{j\omega}{p_2^*}\right)}.$$

[5] The root locus is usually obtained by a numerical calculation of the closed-loop poles for different values of the gain.

The complex pole pair can also be written as

$$\left(1-\frac{j\omega}{p_2}\right)\left(1-\frac{j\omega}{p_2^*}\right)=1-\frac{\omega^2}{p_2 p_2^*}-j\omega\left(\frac{1}{p_2}+\frac{1}{p_2^*}\right)=1-\frac{\omega^2}{|p_2|^2}-j\omega\frac{2\mathrm{Re}\{p_2\}}{|p_2|^2}.$$

In a Bode diagram, the amplitude and phase of H_{open} are plotted versus the frequency $\omega > 0$. A logarithmic scale is used, which has the advantage that the multiplication of two transfer functions is equivalent to the sum of their Bode diagrams. The amplitude of H_{open} in decibels (dB) is calculated as

$$|H_{\text{open}}(j\omega)|_{\text{dB}} := 20\log_{10}|H_{\text{open}}(j\omega)|.$$

In our example, using the properties of the logarithmic function leads to

$$|H_{\text{open}}(j\omega)|_{\text{dB}} = 20\log_{10}|K| + 20\log_{10}\sqrt{1+\frac{\omega^2}{z_1^2}} - 20N\log_{10}\omega - \\ - 20\log_{10}\sqrt{1+\frac{\omega^2}{p_1^2}} - 20\log_{10}\sqrt{\left(1-\frac{\omega^2}{|p_2|^2}\right)^2+\left(\frac{2\omega\mathrm{Re}\{p_2\}}{|p_2|^2}\right)^2}.$$

This expression is the sum of five components. The first is the constant

$$|H_1(j\omega)|_{\text{dB}} := 20\log_{10}|K|.$$

The second function is due to the zero and can be approximated by two asymptotes:

$$|H_2(j\omega)|_{\text{dB}} := 20\log_{10}\sqrt{1+\frac{\omega^2}{z_1^2}} \approx \begin{cases} 0\,\text{dB} & \text{for } \omega \ll |z_1|, \\ 3\,\text{dB} & \text{for } \omega = |z_1|, \\ 20\log_{10}\omega - 20\log_{10}|z_1| & \text{for } \omega \gg |z_1|. \end{cases} \tag{7.20}$$

The N-fold integrator leads to

$$|H_3(j\omega)|_{\text{dB}} := -20N\log_{10}\omega.$$

For the pole p_1, the result is similar to the case of zero z_1, but with opposite signs:

$$|H_4(j\omega)|_{\text{dB}} := -20\log_{10}\sqrt{1+\frac{\omega^2}{p_1^2}} \approx \begin{cases} 0\,\text{dB} & \text{for } \omega \ll |p_1|, \\ -3\,\text{dB} & \text{for } \omega = |p_1|, \\ -20\log_{10}\omega + 20\log_{10}|p_1| & \text{for } \omega \gg |p_1|. \end{cases}$$

Finally, the pole pair has the following asymptotes:

$$|H_5(j\omega)|_{\mathrm{dB}} := -20\log_{10}\sqrt{\left(1-\frac{\omega^2}{|p_2|^2}\right)^2+\left(\frac{2\omega\mathrm{Re}\{p_2\}}{|p_2|^2}\right)^2}$$
$$\approx \begin{cases} 0\,\mathrm{dB} & \text{for } \omega \ll |p_2|, \\ -40\log_{10}\omega + 40\log_{10}|p_2| & \text{for } \omega \gg |p_2|. \end{cases}$$

The phase of H_{open} is given by

$$\measuredangle H_{\mathrm{open}}(j\omega) = \sum_{i=1}^{5} \measuredangle H_i(j\omega).$$

The phases $\measuredangle H_i(j\omega)$ can be approximated by asymptotes in a similar way as shown for the amplitudes. For example, the zero leads to the phase

$$\measuredangle H_2(j\omega) = \measuredangle\left(1-\frac{j\omega}{z_1}\right) = \begin{cases} \approx 0 & \text{for } \omega \to 0, \\ +\frac{\pi}{4} & \text{for } \omega = |z_1| \text{ and } z_1 < 0, \\ \approx +\frac{\pi}{2} & \text{for } \omega \to \infty \text{ and } z_1 < 0, \\ -\frac{\pi}{4} & \text{for } \omega = |z_1| \text{ and } z_1 > 0, \\ \approx -\frac{\pi}{2} & \text{for } \omega \to \infty \text{ and } z_1 > 0. \end{cases}$$

Figure 7.6 shows the Bode plots of the transfer functions $H_i(j\omega)$ with their asymptotes for a system with $N = 1$, positive gain K, and with the zero and poles in the OLHP, i.e., a stable system. The following observations can be made:

- The gain $H_1(j\omega) = K$ leads to an amplitude shift of the open-loop transfer function H_{open}.
- The zero $z_1 > 0$ raises the amplitude and phase; cf. $H_2(j\omega)$. At the frequency $\omega = |z_1|$, the amplitude is close to 3 dB, and the phase equals $\pi/4$. For large frequencies, the amplitude increases with 20 dB per (frequency) decade and the phase approaches $\pi/2$.
- The amplitude of the integrator $H_3(j\omega)$ tends to infinity for small frequencies. This fact enables steady-state accuracy for the closed loop with regard to stepwise disturbances. However, the phase of $-\pi/2$ may lead to stability problems in some cases. This can be shown with the Nyquist stability criterion, which will be presented below.
- The pole p_1 has the opposite effect to that of the zero z_1. For large frequencies, the amplitude slope is -20 dB per decade, and the phase approaches $-\pi/2$.
- For small or large frequencies, the complex pole pair acts as a double pole at $\omega = |p_2|$. However, for frequencies close to $|p_2|$, a resonance may occur. This means that $|H_5(j\omega)|$ may become considerably larger than 1. The frequency at

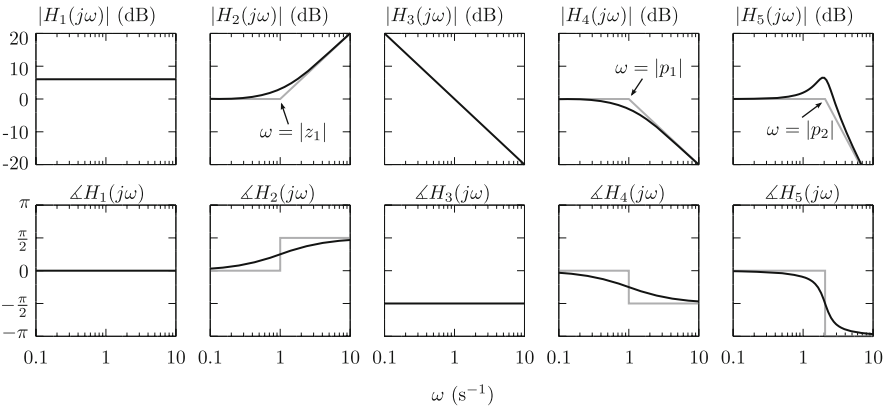

Fig. 7.6 Bode plots of the open-loop transfer function (7.19) for $N = 1$, $K = 2$, $z_1 = -1$, $p_1 = -1$, and $p_2 = -\frac{1}{2} + 2j$

which the maximum of $|H_5(j\omega)|$ occurs can be calculated analytically, and it reads

$$\omega_{\text{res}} = \sqrt{\text{Im}\{p_2\}^2 - \text{Re}\{p_2\}^2} \approx 1.94,$$

i.e., $\text{Im}\{p_2\} > \text{Re}\{p_2\}$ is a necessary condition for a resonance. Disturbances or input signals with frequencies close to ω_{res} will be amplified significantly in the open loop. A resonance in the open loop may be one reason why feedback is necessary. Feedback can provide additional damping, so that the resonance is not present in the closed-loop frequency response.

For the Bode plot of the system with the transfer function H_{open}, the Bode plots of the subsystems H_i have to be combined. As already shown, this simply corresponds to the sum of the amplitude and phase plots due to the use of a logarithmic scale. This also applies to the asymptotes. To sketch the asymptotes of the Bode plot of H_{open}, it is therefore possible to proceed as follows. First, the break points are calculated as the absolute value of the zeros and poles, i.e., $\omega = |z_i(0)|$ and $\omega = |p_i(0)|$. The argument 0 for both z_i and p_i emphasizes that the open-loop zeros and poles are used. Next, one begins with the asymptote of the N-fold integrator H_3. This asymptote is a line with slope $-20N$ dB per decade (of the frequency ω) that crosses the point with amplitude $20\log_{10}(K)$ at $\omega = 1\,\text{s}^{-1}$. For $N = 0$, the Bode plot begins with a horizontal asymptote. One then proceeds to higher frequencies, changing the slope of the asymptote at every break point. For a single pole, the slope changes by -20 dB per decade; for a single zero, by 20 dB per decade; and for multiple poles or zeros, accordingly with the multiple of these slopes. For the phase plot, one begins with a horizontal asymptote of $-N\frac{\pi}{2}$. At the break points, the asymptote is changed stepwise with $\frac{-\pi}{2}$ for a single pole, $\frac{\pi}{2}$ for a zero, and a multiple of $\frac{\pi}{2}$ for multiple poles or zeros. For the amplitude feedback, this procedure leads to

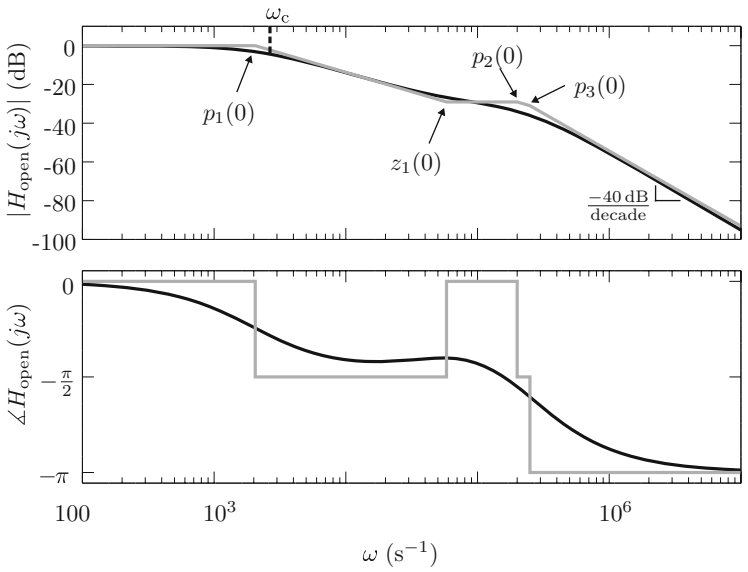

Fig. 7.7 Bode plot of the amplitude feedback example

the asymptotes as shown in Fig. 7.7 for $K_c = 1$. The exact Bode plot is shown as a solid black curve. The static open-loop gain equals

$$H_{open}(j\omega = 0) = K_{mod} G_{Vgain} K_{Vgain} R_p K_{cd} = 0.99$$

for $K_c = 1$. At $\omega = p_1(0)$, the first pole leads to a negative slope of $-20\,dB$ per decade. Next, the zero $z_1(0)$ raises the slope to zero, before the two remaining poles finally lead to a slope or **cutoff rate** of $-40\,dB$ per decade. The phase begins at zero and drops to

$$-\frac{\pi}{2} + \frac{\pi}{2} - \frac{\pi}{2} - \frac{\pi}{2} = -\pi$$

for large frequencies.

The frequency at which the amplitude drops by $-3\,dB$ is called the **cutoff frequency**. It is denoted by $\omega_c = 2004\,\frac{1}{s}$ in Fig. 7.7 and is also called the **bandwidth** of the open-loop transfer function [1].

Because the Bode plot contains all information about the open loop, there is a unique correspondence between this diagram and the transfer function $H_{open}(s)$. If the open loop is stable, the Bode plot can be obtained by measuring the frequency response $H_{open}(j\omega)$. An equivalent diagram that is very useful for determining the stability of the closed loop is the **Nyquist plot**. It is obtained by plotting the curve

$$H_{open}(j\omega) = \mathrm{Re}\{H_{open}(j\omega)\} + j\ \mathrm{Im}\{H_{open}(j\omega)\}$$

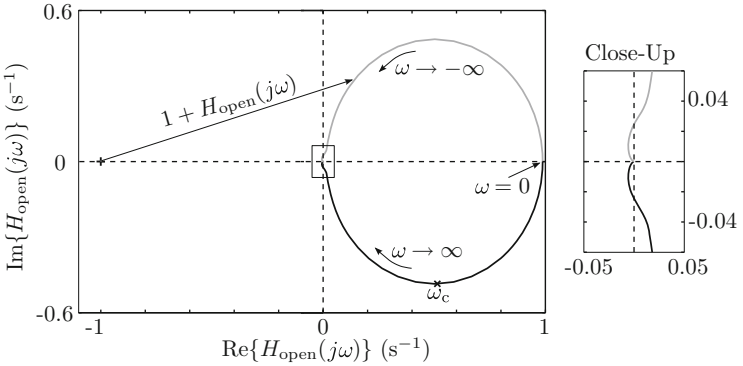

Fig. 7.8 Nyquist plot of the amplitude feedback example

in the complex plane for $\omega \in \mathbb{R}$. The Nyquist plot of the amplitude feedback example is shown in Fig. 7.8.

Due to

$$H_{\mathrm{open}}(-j\omega) = H^*_{\mathrm{open}}(j\omega),$$

the part of the Nyquist plot for negative frequencies ω is always axially symmetric to the part for positive frequencies. For this reason, the Nyquist plot is usually analyzed for only positive frequencies. From the discussion of the Bode plot, it is already known that the Nyquist plot begins at $H_{\mathrm{open}}(j\,0) = 0.99$ and approaches the origin for large ω. Also, the phase approaches $-\pi$, as can be observed from the closeup view in Fig. 7.8. The vector

$$1 + H_{\mathrm{open}}(j\omega)$$

points from $-1 + j\,0$ to the Nyquist plot, as shown in Fig. 7.8. Its behavior is essential for the stability of the closed loop. If we follow this vector from $\omega = 0$ to $\omega \to \infty$, we can define the change of its argument as

$$\Delta\varphi_{\mathrm{Nyquist}} := \lim_{\omega\to\infty} \measuredangle\left(1 + H_{\mathrm{open}}(j\omega)\right) - \measuredangle\left(1 + H_{\mathrm{open}}(j\,0)\right). \tag{7.21}$$

The **general Nyquist stability criterion** can now be used to determine the stability of the closed loop:

Theorem 7.7. *The closed loop is asymptotically stable if and only if the continuous change of the argument as defined in Eq. (7.21) is equal to*

$$\Delta\varphi_{\mathrm{Nyquist}} = n_{\mathrm{unstable}}\pi + n_{\mathrm{critical}}\frac{\pi}{2},$$

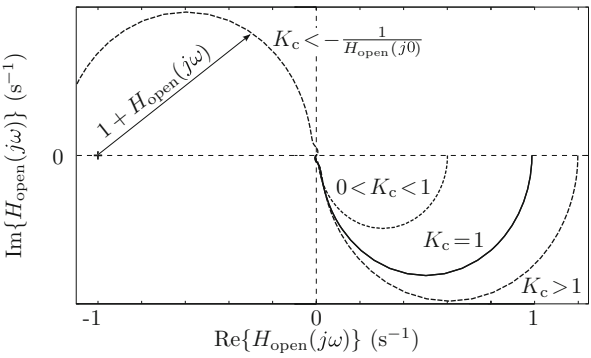

Fig. 7.9 Nyquist plot for different gains K_c

where n_{unstable} *is the number of (unstable) open-loop poles in the ORHP and* n_{critical} *is the number of open-loop poles on the imaginary axis.*

(See, e.g., Unbehauen [14, p. 156].)

Only the continuous change in the argument is considered. If, for example, the Nyquist plot consists of several branches due to open-loop poles on the imaginary axis, then $\Delta\varphi_{\text{Nyquist}}$ can be determined for each branch separately, and the total change is the sum of these results.

Since the amplitude feedback system in our example contains only stable open-loop poles, a necessary and sufficient condition for stability is

$$\Delta\varphi_{\text{Nyquist}} = 0,$$

as is the case for $K_c = 1$ in Fig. 7.8. Changing the gain K_c will only scale the Nyquist plot, as shown in Fig. 7.9. For positive gains $K_c > 0$, the closed loop will always be stable, because $\Delta\varphi_{\text{Nyquist}} = 0$. In the case of negative K_c, the Nyquist plot is also rotated by 180°, and the critical point $-1 + j0$ is crossed for

$$K_c = -\frac{1}{H_{\text{open}}(j0)} = -1.01$$

and the change in the argument is $\Delta\varphi_{\text{Nyquist}} = +\pi$. Thus, the closed loop is unstable for $K_c < -1.01$, a result already obtained with the Hurwitz criterion.

7.4.3 Time Delay

If the feedback loop contains a considerable time delay T_d, this can be taken into account in the Laplace transform of the open loop $H_{\text{open}}(s)$. If, for example, the measurement of the output $y(t)$ is delayed, this leads to

$$y_{\text{delay}}(t) = y(t - T_d).$$

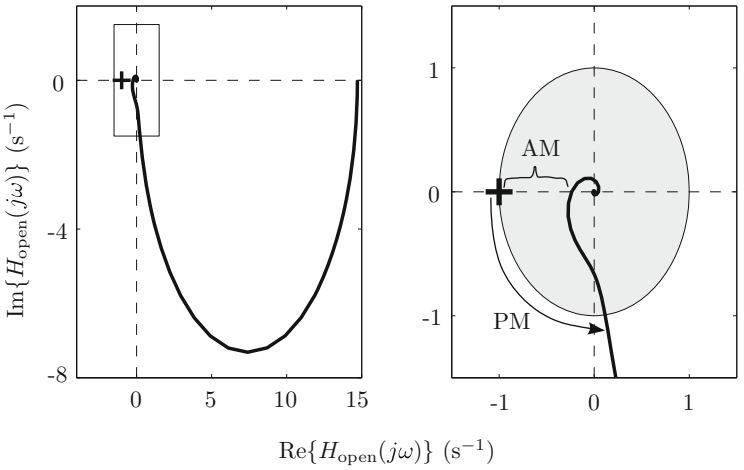

Fig. 7.10 Nyquist plot (*left*) and closeup (*right*) of the amplitude feedback with delay $T_\mathrm{d} = 5\,\mu\mathrm{s}$ and definition of the amplitude margin (*AM*) and phase margin (*PM*)

Due to the shift theorem of the Laplace transform, every open loop with a single delay can be expressed by

$$H_{\mathrm{open,delay}}(s) = H_{\mathrm{open}}(s)\, e^{-T_\mathrm{d} s}.$$

The consequence of the exponential function is that the characteristic equation of the closed loop is no longer an algebraic equation, but a transcendental one. The number of poles becomes infinite, and the stability analysis is thus more involved. Fortunately, the Nyquist criterion can still be applied [15]. For the frequency response,

$$\begin{aligned} |H_{\mathrm{open,delay}}(j\omega)| &= |H_{\mathrm{open}}(j\omega)| \\ \measuredangle H_{\mathrm{open,delay}} &= \measuredangle H_{\mathrm{open}} - T_\mathrm{d}\omega \end{aligned}$$

holds, i.e., the delay leads to a faster decrease of the phase, but does not affect the amplitude. Figure 7.10 shows the Nyquist plot of the amplitude feedback with the nominal feedback gain of $K_\mathrm{c} = 14.9$ and an additional time delay of $T_\mathrm{d} = 5\,\mu\mathrm{s}$. This time delay is a worst-case scenario for signal transit times due to a distance of about 100 m between the cavity and the LLRF unit [12]. The closeup shows that the closed loop is still stable, but not for arbitrary $K_\mathrm{c} > 0$. The Nyquist plot crosses the horizontal axis at -0.237. Increasing the gain K_c by a factor of

$$AM = 20\log\left(\frac{1}{0.237}\right) = 12.5\,\mathrm{dB}$$

will therefore lead to a crossing of the critical point $-1 + j0$ and to instability. This factor is called the **amplitude margin** and is a measure for variations in the amplitude of the process transfer function that can be tolerated. For larger amplitude margins, the feedback is more robust against such variations. In addition, Fig. 7.10 shows that the Nyquist plot crosses the unit circle at an angle of about $-83°$. The frequency of this crossing is $\omega = 34.2 \cdot 10^3\,\mathrm{s}^{-1}$. The **phase margin**

$$PM = 180° - 83° = 97°$$

is defined as the distance to the critical point in terms of the phase, i.e., the tolerable variation in the phase of the process transfer function. A simple estimate[6] shows that an additional time delay of $T_\mathrm{d} = 50\,\mu\mathrm{s}$ would lead to a phase decrease of

$$\omega T_\mathrm{d} \approx 34.2 \cdot 10^3\,\mathrm{s}^{-1} \cdot 50\,\mu\mathrm{s} \approx 98°,$$

i.e., the feedback will remain stable for time delays up to this order of magnitude.

7.4.4 Steady-State Accuracy

The standard closed loop in Fig. 7.2 on p. 341 is said to have no **steady-state error** if

$$x_\mathrm{e}(\infty) := \lim_{t\to\infty} x_\mathrm{e}(t) = \lim_{t\to\infty} (y_\mathrm{r}(t) - y_\mathrm{m}(t)) = 0$$

is guaranteed, i.e., if the measured value converges to the reference value. From Fig. 7.2, the following expression for the steady-state error can be obtained:

$$\begin{aligned} X_\mathrm{e}(s) = {} & \frac{1}{1 + H_\mathrm{p}(s)H_\mathrm{c}(s)H_\mathrm{m}(s)}\, Y_\mathrm{r}(s) - \frac{H_\mathrm{p}(s)H_\mathrm{m}(s)}{1 + H_\mathrm{p}(s)H_\mathrm{c}(s)H_\mathrm{m}(s)}\, X_\mathrm{d1}(s) - \\ & - \frac{H_\mathrm{m}(s)}{1 + H_\mathrm{p}(s)H_\mathrm{c}(s)H_\mathrm{m}(s)}\, (X_\mathrm{d2}(s) + X_\mathrm{d3}(s)). \end{aligned} \tag{7.22}$$

In the following, it is assumed that all transfer functions in this expression are stable, i.e., have only poles in the OLHP. In this case, we can use the final-value theorem for Laplace transforms (cf. Sect. 2.2). Without disturbances, this leads to

$$x_\mathrm{e}(\infty) = \lim_{s\to 0} \big(s\, X_\mathrm{e}(s)\big) = \lim_{s\to 0} \left(\frac{s\, Y_\mathrm{r}(s)}{1 + H_\mathrm{p}(s)H_\mathrm{c}(s)H_\mathrm{m}(s)} \right) = \lim_{s\to 0} \left(\frac{s\, Y_\mathrm{r}(s)}{1 + H_\mathrm{open}(s)} \right).$$

[6] Because the amplitude does not depend on the time delay, the crossing of the unit circle always occurs at the same frequency.

It is now particularly important which type of reference signal $y_r(t)$ is assumed. For a step function, we have $Y_r(s) = K/s$ and[7]

$$x_e(\infty) = \begin{cases} \frac{K}{1+H_{open}(0)} & \text{for } |H_{open}(0)| < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

This shows that an integrator $(1/s)$ in the feedback loop—in the controller, the process, or the measurement transfer function—is sufficient for a vanishing steady-state error. For other reference signals, this may not be sufficient. For example, a ramp signal $(1/s^2)$ requires at least two integrators in the transfer functions of the feedback loop. However, too many integrators may lead to stability problems, because each integrator lowers the phase of the open-loop transfer function by $-\pi/2$.

If significant disturbances are present, it is usually necessary that the integrator be contained in the controller, as can be seen from the other transfer functions in Eq. (7.22). Assuming that the process and measurement transfer functions have no integrator, $H_p(0)$ and $H_m(0)$ are finite, and an integral controller will lead to $x_e(\infty) = 0$ for stepwise disturbances.

7.5 Feedback Design

7.5.1 Tradeoff Between Performance and Robustness

The transfer function $H_p(s)$ in Fig. 7.2 on p. 341 usually describes the physical behavior of the real process only approximately. Reasons for model errors can be nonlinearities, dependence on time or operating conditions, and unmodeled high-frequency dynamics. In many cases, the model errors may be described by parameter variations in the numerator and denominator of the transfer function $H_p(s)$. These variations will lead to a change in performance of the closed-loop control. To estimate this effect, the **sensitivity function**

$$H_s := \frac{\partial H_{ry}}{\partial H_p} \frac{H_p}{H_{ry}}$$

is defined as the relative change of the closed-loop transfer function $H_{ry}(s)$ with respect to variations of the process transfer function $H_p(s)$. With Eq. (7.15), this leads to

[7] Note that $1 + H_{open}(0) = 0$ is impossible, since that would imply that $s = 0$ would be a pole, and this has been excluded by considering stable transfer functions.

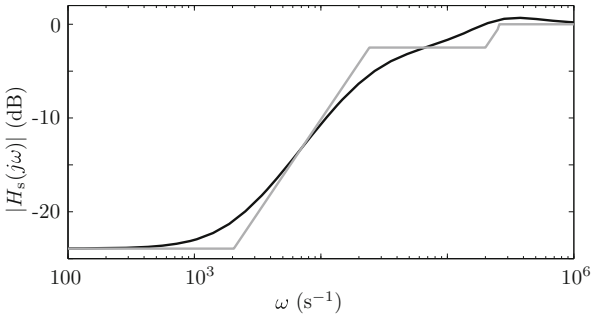

Fig. 7.11 Sensitivity function of the amplitude feedback ($T_d = 0$)

$$H_s = \frac{\partial\left(\frac{H_p H_c}{1+H_p H_c H_m}\right)}{\partial H_p}\,\frac{H_p}{H_{ry}} = \frac{H_c(1+H_p H_c H_m) - H_p H_c^2 H_m}{(1+H_p H_c H_m)^2}\,\frac{H_p(1+H_p H_c H_m)}{H_p H_c}$$

and finally to the sensitivity function

$$H_s(s) = \frac{1}{1+H_p(s)H_c(s)H_m(s)}.$$

This is exactly the disturbance-to-output transfer function $H_{dy}(s)$ (cf. Eq. (7.16)) that was derived from Fig. 7.2. It is apparent that a sufficiently large feedback gain $|H_c|$ will lead to both a small sensitivity $|H_s|$ and a good disturbance rejection. However, a large feedback decreases the amplitude margin AM in many cases and may lead to instability. This shows that a tradeoff between performance and robustness specifications is usually necessary. Please note that for the open-loop system, $H_c = 0$, and the sensitivity equals 1. For the closed-loop system, $|H_s|$ also approaches 1 for large frequencies, because for most practical cases, $|H_p H_c H_m|$ tends to zero.

For our amplitude feedback example, the sensitivity function is equal to

$$H_s(s) = \frac{(s-z_1)(s-z_2)(s-z_3)}{(s-p_1)(s-p_2)(s-p_3)}$$

with

$$z_1 = -2.5\cdot 10^5\,\mathrm{s}^{-1}, \quad z_2 = -2\cdot 10^5\,\mathrm{s}^{-1}, \quad z_3 = -2.05\cdot 10^3\,\mathrm{s}^{-1},$$
$$p_1 = -2.42\cdot 10^4\,\mathrm{s}^{-1}, \quad p_{2,3} = (-2.14 \pm j\ 1.44)\cdot 10^5\,\mathrm{s}^{-1}.$$

Its amplitude $|H_s(j\omega)|$ is shown in Fig. 7.11. In contrast to $|H_{\text{open,delay}}(j\omega)|$, the amplitude of the sensitivity function depends on the time delay.

The sensitivity shows that the amplitude feedback rejects disturbances or noise with frequency components up to about 10 kHz. The closed loop is also less sensitive with respect to model variations than the open loop in this frequency range.

However, the sensitivity is not zero for $\omega \to 0$. This implies that the closed loop does not reject DC offsets completely and may thus have a steady-state error. This can be shown as follows. From the standard feedback loop, the control error can be calculated as

$$\begin{aligned} X_\mathrm{e}(s) &= Y_\mathrm{r}(s) - H_\mathrm{m}(s) \cdot Y(s) \\ &= Y_\mathrm{r}(s) - H_\mathrm{m}(s) \cdot \frac{H_\mathrm{c}(s)\ H_\mathrm{p}(s)}{1 + H_\mathrm{m}(s)\ H_\mathrm{c}(s)\ H_\mathrm{p}(s)} \cdot Y_\mathrm{r}(s) \\ &= \frac{1}{1 + H_\mathrm{m}(s)\ H_\mathrm{c}(s)\ H_\mathrm{p}(s)} \cdot Y_\mathrm{r}(s). \end{aligned}$$

If we assume that the closed loop is stable and the reference signal is equal to a unit step, i.e., $Y_\mathrm{r} = 1/s$, then the final value of the control error is given by

$$\begin{aligned} \lim_{t\to\infty} x_\mathrm{e}(t) &= \lim_{s\to 0} \left(s \cdot \frac{1}{1 + H_\mathrm{m}(s)\ H_\mathrm{c}(s)\ H_\mathrm{p}(s)} \cdot \frac{1}{s} \right) \\ &= \frac{1}{1 + H_\mathrm{m}(0)\ H_\mathrm{c}(0)\ H_\mathrm{p}(0)}. \end{aligned}$$

Thus, the value of the sensitivity function for $\omega = 0$ is equal to the relative steady-state error of the closed-loop system. For the amplitude feedback loop, a value of 6.4%, or -23.9 dB, is obtained. This steady-state error will also be apparent in the simulation results in the next section.

7.5.2 Design Goals and Specifications

The main design goals of feedback are stability, a fast dynamic response, disturbance rejection, a small tracking error, and robustness against parameter variations. In addition, the control effort should comply with the physical limitations of the process. There exist several parameters to describe these specifications quantitatively. In the time domain, the response to a step disturbance or reference signal is often considered, and the following quantities are used to describe the dynamic response:

- **Rise time**: transit time from 10% to 90% of the final value, i.e., of the output step size.
- Percentage of **overshoot**.
- **Settling time**: time after which the output stays inside a $\pm 5\%$ or $\pm 2\%$ interval around the final value.
- **Steady-state error** between the reference signal and the output.

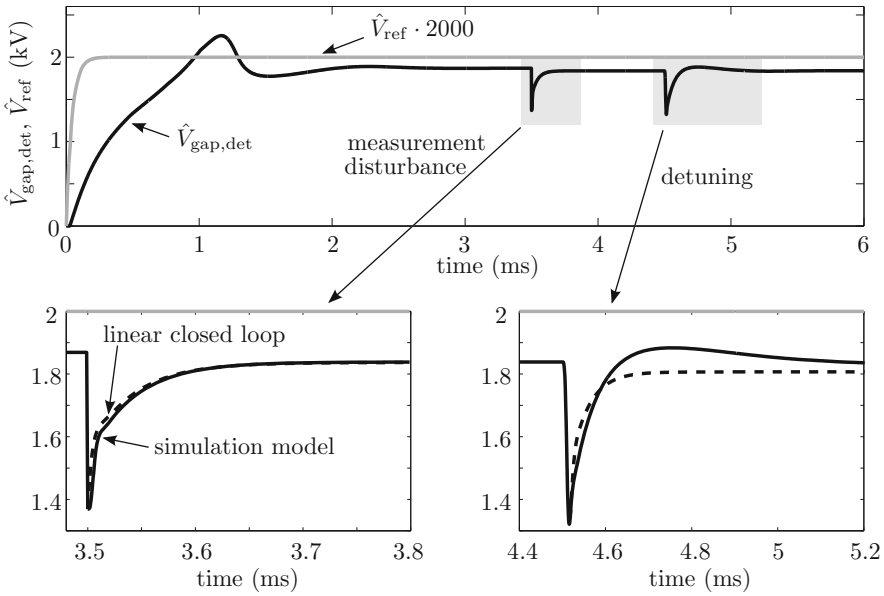

Fig. 7.12 Performance of the amplitude feedback

The performance of the amplitude feedback example is shown in Fig. 7.12. The curve $\hat{V}_{\mathrm{gap,det}}$ is obtained from a simulation model from [12], which is in good agreement with measurements. The reference signal $\hat{V}_{\mathrm{ref}}$ is initially raised from zero to 1 V. Due to a prefilter with a time constant of 43 µs, the reference signal is raised not stepwise, but smoothly. The simulation model includes not only the amplitude feedback, but also a resonance frequency feedback to ensure that the cavity is in resonance. At the beginning of the simulation, the resonance frequency feedback has to settle and has a strong coupling with $\hat{V}_{\mathrm{gap}}$. At $t \approx 3$ ms, both feedback loops have reached their equilibrium.

The amplitude feedback is excited at $t = 3.5$ ms with a stepwise disturbance of the measurement $\hat{V}_{\mathrm{gap,det}}$. The dynamic response of the simulation model is compared to the response of the linear closed loop $H_{\mathrm{ry}}(s)$ with $T_{\mathrm{d}} = 0$ (Fig. 7.12, bottom left). This shows that the transfer function $H_{\mathrm{ry}}(s)$ describes the behavior very well for small deviations from equilibrium. From the simulation results, a rise time of 73 µs, a 5% settling time of 103 µs, and a steady-state error of 6.4% are obtained.

At $t = 4.5$ ms, the cavity is detuned, so that the gap voltage drops by about 0.5 kV. This time, the simulation model shows a different behavior due to the interaction of the resonance frequency feedback with the amplitude feedback. This demonstrates that nested control loops are dynamically coupled in general. If the coupling is strong, it is necessary to take this into account during the analysis

and design of the feedback. Nested control loops can be described by MIMO or multivariable control systems [16].

In addition to the mentioned parameters, there also exist specifications in the frequency domain:

- **Resonant peak**: the maximum of the closed-loop frequency response $|H_{\mathrm{ry}}(j\omega)|$ indicates relative stability and is recommended to be between 1.1 and 1.5 [1].
- **Bandwidth**: the frequency at which $|H_{\mathrm{ry}}(j\omega)|$ has decreased by $-3\,\mathrm{dB}$ with respect to the zero-frequency value.
- **Cutoff rate**: the slope of $|H_{\mathrm{ry}}|$ at high frequencies.
- **Amplitude margin** and **phase margin** (cf. Sect. 7.4.3): an *AM* larger than 6 dB and a *PM* between 30° and 60° are regarded as a good tradeoff between robustness and performance [1].

In our example, the bandwidth of $H_{\mathrm{ry}}(j\omega)$ equals $30.3 \cdot 10^3\,\mathrm{s}^{-1}$ (which corresponds to $\Delta f = 4831$ Hz), and the cutoff-rate is $-20\,\mathrm{dB/decade}$.

7.5.3 PID Control

A general proper **PID control** algorithm is given by

$$H_{\mathrm{c}}(s) = \frac{U(s)}{X_{\mathrm{e}}(s)} = K_{\mathrm{P}} + K_{\mathrm{I}}\frac{1}{s} + K_{\mathrm{D}}\frac{s}{T_{\mathrm{D}}s + 1};$$

it is a combination of a proportional, an integral, and a derivative controller. The transfer function can also be written as

$$H_{\mathrm{c}}(s) = \frac{(K_{\mathrm{P}}T_{\mathrm{D}} + K_{\mathrm{D}})s^2 + (K_{\mathrm{P}} + K_{\mathrm{I}}T_{\mathrm{D}})s + K_{\mathrm{I}}}{s(T_{\mathrm{D}}s + 1)}; \tag{7.23}$$

it has two zeros and two poles. A pure derivative is obtained for $T_{\mathrm{D}} = 0$. However, this leads to an improper transfer function. In the time domain, the controller is described by the differential equation

$$T_{\mathrm{D}}\dot{u}(t) + u(t) = (K_{\mathrm{P}}T_{\mathrm{D}} + K_{\mathrm{D}})\dot{x}_e(t) + (K_{\mathrm{P}} + K_{\mathrm{I}}T_{\mathrm{D}})x_{\mathrm{e}}(t) + K_{\mathrm{I}}\int_0^t x_{\mathrm{e}}(\tau)\,\mathrm{d}\tau.$$

In steady state, the control error x_{e} must be zero due to the integration.

The controller of the amplitude feedback example is of PDT_1 type. This can be shown as follows. A general **PDT_1 controller** can be written as

$$H_{\mathrm{c}}(s) = K_{\mathrm{P}} + \frac{K_{\mathrm{D}}s}{T_{\mathrm{D}}s + 1} = K_{\mathrm{P}}\frac{s\left(T_{\mathrm{D}} + \frac{K_{\mathrm{D}}}{K_{\mathrm{P}}}\right) + 1}{s\,T_{\mathrm{D}} + 1}.$$

With

$$K_c = K_\mathrm{P}, \quad T_{c1} = T_\mathrm{D} + \frac{K_\mathrm{D}}{K_\mathrm{P}}, \quad T_{c2} = T_\mathrm{D},$$

we obtain the amplitude controller that is shown in Fig. 7.4.

To design a general PID controller, it is necessary to determine the four degrees of freedom K_P, K_D, K_I, and T_D so that the specifications are met. If the open-loop system is stable, the two zeros of $H_\mathrm{c}(s)$ may be used to compensate open-loop poles. The time constant T_D should not be chosen too small, because that would amplify high-frequency noise.

Several so-called **tuning rules** exist for the design of PI and PID controllers [5]. A simple tuning rule is described in [16] that is based on the approximation of the process transfer function with a first-order model

$$H_\mathrm{approx}(s) = \frac{K}{T \cdot s + 1}\, e^{-T_\mathrm{d} \cdot s},$$

with the gain K, the time constant T, and a time delay T_d. For a PI controller, the tuning rule is (cf. [16, p. 57])

$$K_\mathrm{P} = \frac{1}{K}\frac{T}{T_\mathrm{tune} + T_\mathrm{d}}, \qquad K_\mathrm{I} = \frac{K_\mathrm{P}}{\min\{T, 4(T_\mathrm{tune} + T_\mathrm{d})\}},$$

with a single tuning parameter T_tune. A small value of this parameter will lead to fast output performance, whereas a large value implies a high robustness and smaller values of the input. A typical tradeoff is the choice $T_\mathrm{tune} = T_\mathrm{d}$.

This tuning rule can be applied to the amplitude feedback loop example. From Fig. 7.4, the open-loop transfer function

$$\frac{\Delta\hat{V}_\mathrm{gap,det}(s)}{\Delta\hat{V}_\mathrm{c}(s)} = \frac{K_\mathrm{mod} G_\mathrm{Vgain} K_\mathrm{Vgain} R_\mathrm{p} K_\mathrm{cd}}{(s\, T_\mathrm{cav} + 1)(s\, T_\mathrm{det} + 1)}$$

is obtained. For this type of transfer function, the following first-order approximation may be used; cf. [16, p. 58]:

$$K = K_\mathrm{mod} G_\mathrm{Vgain} K_\mathrm{Vgain} R_\mathrm{p} K_\mathrm{cd} = 0.9877, \qquad T = T_\mathrm{det} + \frac{1}{2} T_\mathrm{cav} = 7\,\mu\mathrm{s},$$

$$T_\mathrm{d} = \frac{1}{2} T_\mathrm{cav} = 2\,\mu\mathrm{s}.$$

With $T_\mathrm{tune} = T_\mathrm{d}$ as the choice of the tuning parameter, the coefficients of the resulting PI controller are

$$K_\mathrm{P} = 1.7718, \qquad K_\mathrm{I} = 2.5312 \cdot 10^5\, \frac{1}{\mathrm{s}}.$$

The settling time of the linear amplitude feedback with this controller is 16.4 µs for a 5% interval around the set point. This is considerably faster than the PDT_1 controller. Furthermore, the PI controller leads to a zero steady-state error. Note, however, that for the design in this section, we have neglected any interaction of the amplitude loop with the resonance frequency feedback loop.

For the practical implementation of a PID controller, some issues should be taken into account. If the process is stable, it is often sufficient to use a PI controller. Derivative action, i.e., $K_D \neq 0$, will lead to an increased sensitivity with respect to measurement noise. If the reference signal $y_r(t)$ contains steps and a derivative action is needed, it is usually better to use the measured output $y_m(t)$ as input of the derivative part of the controller instead of the control error $x_e(t)$; cf. [16, p. 56] and [5, p. 317]. One challenge for the integral action is the so-called **integrator windup** [5], a nonlinear effect.

We can illustrate this effect by means of Fig. 7.3. We assume that the controller has integral action and generates a value that exceeds the constraints of the subsequent saturation function. In this case, the output of the feedback will be a constant value as long as the saturation function is active. This may be interpreted as a feedback loop that is no longer closed, because the output of the controller does not depend on the control error. The integral controller will, however, continue to integrate the control error, and this may result in a poor overall feedback performance. Measures that prevent windup are known as **antiwindup**.

7.5.4 Stability Issues for Nonlinear Systems

As described in Sect. 7.1.3, almost every practical feedback system is, in fact, a nonlinear system

$$\frac{\mathrm{d}\vec{x}(t)}{\mathrm{d}t} = \vec{v}_1(\vec{x}(t), \vec{u}(t)), \tag{7.24a}$$

$$\vec{y}_m(t) = \vec{v}_2(\vec{x}(t)), \tag{7.24b}$$

where $\vec{y}_m$ is the output vector with the measured quantities of the process. A common approach is to calculate the linearization

$$\frac{\mathrm{d}\Delta\vec{x}(t)}{\mathrm{d}t} = A \cdot \Delta\vec{x}(t) + B \cdot \Delta\vec{u}(t), \tag{7.25a}$$

$$\Delta\vec{y}_m(t) = C \cdot \Delta\vec{x}(t) \tag{7.25b}$$

of the system for a certain equilibrium and to use it for the analysis or design of a linear controller so that the closed-loop behavior is stable. This approach has also been chosen in the previous sections. An important question that now arises is whether the linear controller will also be able to stabilize the nonlinear system.

The stability theory of Lyapunov that was described in Sect. 2.8.5 is useful to obtain some conclusions concerning this question. In order to use the theory of Lyapunov, it is necessary to analyze the feedback loop in the time domain, because the frequency domain approach is in general not applicable to nonlinear systems.

Consider first a very general linear controller in state-space representation

$$\frac{\mathrm{d}\vec{x}_\mathrm{c}(t)}{\mathrm{d}t} = A_\mathrm{c} \cdot \vec{x}_\mathrm{c}(t) + B_\mathrm{c} \cdot \vec{x}_\mathrm{e}(t), \tag{7.26a}$$

$$\vec{u}_\mathrm{c}(t) = C_\mathrm{c} \cdot \vec{x}_\mathrm{c}(t) + D_\mathrm{c} \cdot \vec{x}_\mathrm{e}(t), \tag{7.26b}$$

where $\vec{x}_\mathrm{e} = \Delta\vec{y}_\mathrm{r} - \Delta\vec{y}_\mathrm{m}$ denotes the vector with measured control errors, $\vec{u}_\mathrm{c}$ is the actuator value that can be used as input to the process (i.e., $\Delta\vec{u} = \vec{u}_\mathrm{c}$), and $\vec{x}_\mathrm{c}$ contains the internal states of the controller. This type of controller is also known as a **dynamic output feedback**, because the controller has a dynamic structure and it uses the output vector $\vec{y}_\mathrm{m}$ as the only information about the process. This type of controller also contains the PID controller as a special case: rewriting the transfer function (7.23) as the sum of a constant and a remaining polynomial leads to

$$H_\mathrm{c}(s) = \frac{U_\mathrm{c}(s)}{X_\mathrm{e}(s)} = \left(K_\mathrm{P} + \frac{K_\mathrm{D}}{T_\mathrm{D}}\right) + \frac{\left(K_\mathrm{I} - \frac{K_\mathrm{D}}{T_\mathrm{D}^2}\right)s + \frac{K_\mathrm{I}}{T_\mathrm{D}}}{s^2 + \frac{1}{T_\mathrm{D}}s}.$$

Using the results of Sect. 7.1.2 and taking the additional direct feedthrough into account leads to the following state-space representation of the controller:

$$\frac{\mathrm{d}\vec{x}_\mathrm{c}(t)}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{1}{T_\mathrm{D}} \end{bmatrix} \cdot \vec{x}_\mathrm{c}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot x_\mathrm{e}(t),$$

$$u_\mathrm{c}(t) = \left[\frac{K_\mathrm{I}}{T_\mathrm{D}} \left(K_\mathrm{I} - \frac{K_\mathrm{D}}{T_\mathrm{D}^2}\right)\right] \cdot \vec{x}_\mathrm{c}(t) + \left(K_\mathrm{P} + \frac{K_\mathrm{D}}{T_\mathrm{D}}\right) \cdot x_\mathrm{e}(t).$$

This is a dynamic output feedback. Note that the case of a pure derivative controller ($T_\mathrm{D} = 0$) is not included in this representation.[8] Due to Eq. (7.26), the transfer function of the controller can be obtained by

$$H_\mathrm{c}(s) = C_\mathrm{c} \cdot (sI - A_\mathrm{c})^{-1} \cdot B_\mathrm{c} + D_\mathrm{c}.$$

Connecting the controller (7.26) with system (7.25) (i.e., by $\Delta\vec{u} = \vec{u}_\mathrm{c}$) leads directly to the following dynamics of the closed loop:

[8] This is, however, not a serious limitation, since a pure derivative would be both undesirable in the presence of noise and is not realizable on any physical hardware.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\Delta\vec{x}(t)\\ \vec{x}_\mathrm{c}(t)\end{bmatrix} = \underbrace{\begin{bmatrix}A - B\cdot D_\mathrm{c}\cdot C & B\cdot C_\mathrm{c}\\ -B_\mathrm{c}\cdot C & A_\mathrm{c}\end{bmatrix}}_{A_\mathrm{cl}}\cdot\begin{bmatrix}\Delta\vec{x}(t)\\ \vec{x}_\mathrm{c}(t)\end{bmatrix} + \begin{bmatrix}B\cdot D_\mathrm{c}\\ B_\mathrm{c}\end{bmatrix}\cdot\Delta\vec{y}_\mathrm{r}(t). \tag{7.27}$$

We assume that the controller is designed properly, so that the closed-loop dynamics are stable. According to the results of Sect. 7.1.5, this is the case if A_cl has only eigenvalues with negative real parts.

After the controller design, the controller will be connected to the real nonlinear process. One possible choice for the input of the nonlinear system (7.24) is then

$$\vec{u}(t) = \vec{u}_\mathrm{F} + \vec{u}_\mathrm{c}(t) = \vec{u}_\mathrm{F} + C_\mathrm{c}\cdot\vec{x}_\mathrm{c}(t) + D_\mathrm{c}\cdot\vec{x}_\mathrm{e}(t),$$

where $\vec{u}_\mathrm{F}$ is a feedforward value that equals the input value at the equilibrium point $\vec{x} = \vec{x}_\mathrm{F}$. In other words,

$$\vec{v}_1(\vec{x}_\mathrm{F}, \vec{u}_\mathrm{F}) = 0$$

is assumed, and the controller has only to correct deviations from the equilibrium. The control error is now given by

$$\vec{x}_\mathrm{e} = \vec{y}_\mathrm{r} - \vec{y}_\mathrm{m} = \vec{y}_\mathrm{r} - \vec{v}_2(\vec{x}).$$

These choices of the closed-loop connection lead to the following dynamics:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\vec{x}(t)\\ \vec{x}_\mathrm{c}(t)\end{bmatrix} = \begin{bmatrix}\vec{v}_1(\vec{x}(t), \vec{u}(t))\\ A_\mathrm{c}\cdot\vec{x}_\mathrm{c}(t) + B_\mathrm{c}\cdot(\vec{y}_\mathrm{r}(t) - \vec{v}_2(\vec{x}(t)))\end{bmatrix}. \tag{7.28}$$

A linearization around $\vec{x} = \vec{x}_\mathrm{F}$, $\vec{u} = \vec{u}_\mathrm{F}$, $\vec{x}_\mathrm{c} = 0$, and $\vec{y}_\mathrm{r} = \vec{v}_2(\vec{x}_\mathrm{F})$ leads to the same linear dynamics as Eq. (7.27). This is reasonable, because it means that the same result is obtained either by linearizing the nonlinear closed-loop dynamics or by using the linearization (7.25) of the open-loop system (7.24) to obtain the linear closed-loop model (7.27).

We already assumed that Eq. (7.27) is stable, and we can now use theorem 2.18. For $\Delta\vec{y}_\mathrm{r} = 0$ and the previous assumption of a strictly stable matrix A_cl (the real parts of all eigenvalues are negative), the theorem can be applied to Eq. (7.28), and the consequence is a stable equilibrium of the nonlinear setup. This is an important motivation for using linear control design in many cases, even for systems that are practically nonlinear.

Note, however, that the linear system (7.27) is asymptotically stable in the global sense, i.e., for arbitrary initial values, whereas in general, the asymptotic stability of the nonlinear system (7.28) is given only in a local neighborhood around the equilibrium. This neighborhood, also called a **region of attraction**, may be so small that from a practical point of view, the equilibrium is in fact unstable. The size of

the region of attraction can be estimated using Lyapunov functions as defined in Sect. 2.8.5.

A nonzero reference value $\Delta\vec{y}_\mathrm{r} \neq 0$ acts as an excitation. As long as it is not too large, the closed loop will be stable.

If further disturbances act on the system (7.24) or the model is inaccurate, this may lead to a steady-state error. In most cases, an integral controller will help to avoid such an error. A pure integral controller can be written as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x_\mathrm{c}(t) &= x_\mathrm{e}(t),\\ u_\mathrm{c}(t) &= K_\mathrm{I}x_\mathrm{c}(t).\end{aligned}$$

Therefore, $A_\mathrm{c} = 0$, $B_\mathrm{c} = 1$, $C_\mathrm{c} = K_\mathrm{I}$, and $D_\mathrm{c} = 0$. The closed-loop dynamics for a SISO system are then

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\vec{x}(t)\\ x_\mathrm{c}(t)\end{bmatrix} = \begin{bmatrix}\vec{v}_1(\vec{x}(t), u_\mathrm{F} + K_\mathrm{I}\,x_\mathrm{c}(t))\\ y_\mathrm{r} - v_2(\vec{x}(t))\end{bmatrix},$$

and from the bottom row, we have the equilibrium

$$\frac{\mathrm{d}}{\mathrm{d}t}x_\mathrm{c} = 0 \quad \Rightarrow \quad y_\mathrm{r} = v_2(\vec{x}) = y_\mathrm{m},$$

and the steady-state error will therefore tend to zero for stepwise reference signals.

References

1. W.S. Levine (ed.), *The Control Handbook: Control System Fundamentals*, 2nd edn. (CRC Press/Taylor & Francis Group, West Palm Beach/London, 2011)
2. J.W. Polderman, J.C. Willems, *Introduction to the Mathematical Theory of Systems and Control* (Springer, New York, 1998)
3. J. Doyle, B. Francis, A. Tannenbaum, *Feedback Control Theory* (Macmillan, New York, 1990)
4. E.D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Textbooks in Applied Mathematics, Number 6, 2nd edn. (Springer, New York, 1998)
5. K.J. Aström, R.M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers* (Princeton University Press, Princeton, 2012)
6. P. Baudrenghien, Low-level RF, in *CAS - CERN Accelerator School: RF for Accelerators*, Ebeltoft, 8–17 June 2010, pp. 341–367
7. D. Lens, Modeling and control of longitudinal single-bunch oscillations in heavy-ion synchrotrons. Fortschrittberichte VDI, Reihe 8, Mess-, Steuerungs- und Regelungstechnik; Nr. 1209; Dissertation, Technische Universität Darmstadt, 2012
8. K.J. Aström, B. Wittenmark, *Computer-Controlled Systems* (Prentice Hall, Englewood, 1997)
9. O. Föllinger, *Laplace- und Fourier-Transformation* (Hüthig, Heidelberg, 1990)
10. O. Föllinger, *Regelungstechnik: Einführung in die Methoden und ihre Anwendung* (Hüthig Buch, Heidelberg, 1990)
11. G. Ludyk, *Theoretische Regelungstechnik 1* (Springer, Berlin, 1995)

12. U. Hartel, Modellierung des Regelungs- und Steuerungssystems einer Beschleunigungseinheit für Synchrotrons. Diplomarbeit, Technische Universität Darmstadt, Darmstadt, 2011
13. F.R. Gantmacher, *Matrizentheorie* (Deutscher Verlag der Wissenschaften, Berlin, 1986)
14. H. Unbehauen, *Regelungstechnik I*, 15. Auflage (Vieweg+Teubner Verlag, Wiesbaden, 2008)
15. O. Föllinger, Zur Stabilität von Totzeitsystemen, Regelungstechnik, S. 145–149 (1967)
16. S. Skogestad, I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design* (Wiley, London, 2005)

Correction to: Theoretical Foundations of Synchrotron and Storage Ring RF Systems

Correction to:
Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*, https://doi.org/10.1007/978-3-319-07188-6

The updated version of this book can be found at https://doi.org/10.1007/978-3-319-07188-6

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6_8

Appendix

A.1 Description of an Ellipse in the Plane

If the **semiaxes** $a, b > 0$ of an **ellipse** are oriented in parallel to the $\tilde{x}$- and $\tilde{y}$-axes of a Cartesian coordinate system, one may describe the ellipse by

$$\boxed{\begin{aligned} \tilde{x} &= a \ \cos\varphi, \\ \tilde{y} &= b \ \sin\varphi, \end{aligned}} \qquad \begin{aligned} &(A.1) \\ &(A.2) \end{aligned}$$

where $0 \leq \varphi < 2\pi$ is the range of the parameter φ. This obviously leads to the equation

$$\frac{\tilde{x}^2}{a^2} + \frac{\tilde{y}^2}{b^2} = 1,$$

which does not include the parameterization.

In general, the ellipse may, of course, have a different orientation. Therefore, we will now use a **rotation matrix** according to

$$\boxed{\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\zeta & -\sin\zeta \\ \sin\zeta & \cos\zeta \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}}$$

in order to move the points of the ellipse counterclockwise (cf. Fig. A.1). We obtain

$$\begin{aligned} x &= a\cos\zeta\cos\varphi - b\sin\zeta\sin\varphi, \\ y &= a\sin\zeta\cos\varphi + b\cos\zeta\sin\varphi, \end{aligned}$$

as a parametric description of the rotated ellipse. This description includes a very special parameterization, because the angle φ is used as a parameter. One could alternatively use the path length or a parameter without any special meaning.

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6

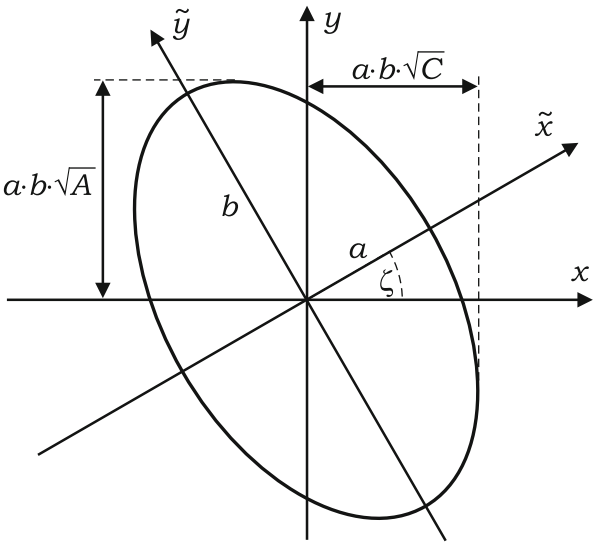

Fig. A.1 Ellipse with arbitrary orientation

Therefore, the objective of the following calculation is to find a description that is free of parameters. This may be achieved by eliminating the parameter φ. To this end, we first calculate the following expressions:

$$x^2 = a^2 \cos^2\zeta \cos^2\varphi + b^2 \sin^2\zeta \sin^2\varphi - 2ab \sin\zeta \cos\zeta \sin\varphi \cos\varphi,$$
$$y^2 = a^2 \sin^2\zeta \cos^2\varphi + b^2 \cos^2\zeta \sin^2\varphi + 2ab \sin\zeta \cos\zeta \sin\varphi \cos\varphi,$$
$$xy = a^2 \sin\zeta \cos\zeta \cos^2\varphi - b^2 \sin\zeta \cos\zeta \sin^2\varphi + ab \sin\varphi \cos\varphi(\cos^2\zeta - \sin^2\zeta).$$

We now use the trigonometric identities

$$\sin^2\varphi = \frac{1}{2} - \frac{1}{2}\cos(2\varphi), \tag{A.3}$$
$$\cos^2\varphi = \frac{1}{2} + \frac{1}{2}\cos(2\varphi), \tag{A.4}$$
$$\sin\varphi \cos\varphi = \frac{1}{2}\sin(2\varphi), \tag{A.5}$$

to obtain

$$x^2 = \frac{a^2\cos^2\zeta + b^2\sin^2\zeta}{2} + \frac{a^2\cos^2\zeta - b^2\sin^2\zeta}{2}\cos(2\varphi) - ab\sin\zeta\cos\zeta\sin(2\varphi),$$
$$y^2 = \frac{a^2\sin^2\zeta + b^2\cos^2\zeta}{2} + \frac{a^2\sin^2\zeta - b^2\cos^2\zeta}{2}\cos(2\varphi) + ab\sin\zeta\cos\zeta\sin(2\varphi),$$

$$xy = \frac{(a^2 - b^2)\sin\zeta\cos\zeta}{2} + \frac{(a^2 + b^2)\sin\zeta\cos\zeta}{2}\cos(2\varphi) + \frac{ab(\cos^2\zeta - \sin^2\zeta)}{2}\sin(2\varphi).$$

Using Eqs. (A.3)–(A.5) for ζ, this may also be written as

$$\begin{aligned}
x^2 &= \frac{a^2 + b^2}{4} + \frac{a^2 - b^2}{4}\cos(2\zeta) + \left(\frac{a^2 - b^2}{4} + \frac{a^2 + b^2}{4}\cos(2\zeta)\right)\cos(2\varphi) \\
&\quad - \frac{ab\sin(2\zeta)}{2}\sin(2\varphi), \\
y^2 &= \frac{a^2 + b^2}{4} + \frac{b^2 - a^2}{4}\cos(2\zeta) + \left(\frac{a^2 - b^2}{4} - \frac{a^2 + b^2}{4}\cos(2\zeta)\right)\cos(2\varphi) \\
&\quad + \frac{ab\sin(2\zeta)}{2}\sin(2\varphi), \\
xy &= \frac{(a^2 - b^2)\sin(2\zeta)}{4} + \frac{(a^2 + b^2)\sin(2\zeta)}{4}\cos(2\varphi) \\
&\quad + \frac{ab\cos(2\zeta)}{2}\sin(2\varphi).
\end{aligned}$$

We now combine the first and the last equations in order to eliminate $\sin(2\varphi)$:

$$\begin{aligned}
\xi_x &= x^2\cos(2\zeta) + xy\sin(2\zeta) = \\
&= \frac{a^2 + b^2}{4}\cos(2\zeta) + \frac{a^2 - b^2}{4} + \left(\frac{a^2 - b^2}{4}\cos(2\zeta) + \frac{a^2 + b^2}{4}\right)\cos(2\varphi).
\end{aligned}$$

Also, to eliminate $\sin(2\varphi)$, the combination of the second and the third equations yields

$$\begin{aligned}
\xi_y &= y^2\cos(2\zeta) - xy\sin(2\zeta) = \\
&= \frac{a^2 + b^2}{4}\cos(2\zeta) + \frac{b^2 - a^2}{4} + \left(\frac{a^2 - b^2}{4}\cos(2\zeta) - \frac{a^2 + b^2}{4}\right)\cos(2\varphi).
\end{aligned}$$

Now we see that $\cos(2\varphi)$ can be eliminated by generating the following expression:

$$\begin{aligned}
&\xi_x\left(\frac{a^2 - b^2}{4}\cos(2\zeta) - \frac{a^2 + b^2}{4}\right) - \xi_y\left(\frac{a^2 - b^2}{4}\cos(2\zeta) + \frac{a^2 + b^2}{4}\right) = \\
&= \frac{a^4 - b^4}{16}\cos^2(2\zeta) + \frac{(a^2 - b^2)^2}{16}\cos(2\zeta) - \frac{(a^2 + b^2)^2}{16}\cos(2\zeta) - \frac{a^4 - b^4}{16} +
\end{aligned}$$

$$+\frac{b^4-a^4}{16}\cos^2(2\zeta)+\frac{(b^2-a^2)^2}{16}\cos(2\zeta)-\frac{(a^2+b^2)^2}{16}\cos(2\zeta)+\frac{a^4-b^4}{16}=$$
$$=-\frac{a^2b^2}{8}\cos(2\zeta)\cdot 4=-\frac{a^2b^2}{2}\cos(2\zeta)$$

$$\Rightarrow \xi_x\left(\frac{a^2+b^2}{2a^2b^2\cos(2\zeta)}+\frac{b^2-a^2}{2a^2b^2}\right)+\xi_y\left(\frac{a^2+b^2}{2a^2b^2\cos(2\zeta)}+\frac{a^2-b^2}{2a^2b^2}\right)=1.$$

If we now insert the definitions of ξ_x and ξ_y, we obtain the implicit equation of the ellipse that we were looking for:

$$\boxed{Ax^2+2Bxy+Cy^2=1.} \tag{A.6}$$

The constants are obviously given by the following equations:

$$A=\frac{a^2+b^2}{2a^2b^2}+\frac{b^2-a^2}{2a^2b^2}\cos(2\zeta), \tag{A.7}$$
$$2B=\frac{b^2-a^2}{a^2b^2}\sin(2\zeta), \tag{A.8}$$
$$C=\frac{a^2+b^2}{2a^2b^2}+\frac{a^2-b^2}{2a^2b^2}\cos(2\zeta). \tag{A.9}$$

These constants obviously satisfy the inequalities

$$A\geq 0 \qquad \text{and} \qquad C\geq 0. \tag{A.10}$$

We now determine how to return from the implicit equation (A.6) to the original parameters a, b, ζ. We find that

$$A-C=\frac{b^2-a^2}{a^2b^2}\cos(2\zeta), \tag{A.11}$$

so that in combination with Eq. (A.8), we have

$$\boxed{\tan(2\zeta)=\frac{2B}{A-C}.}$$

Without loss of generality, we now assume that $b>a>0$ holds (the case $a=b>0$ is analyzed below). According to Eqs. (A.8) and (A.11), the signs of the expressions $2B$ and $A-C$ may be used to determine $2\zeta\in\,]-\pi,+\pi]$ in a unique way. Hence, the rotation angle is in the range

$$\zeta\in\left]-\frac{\pi}{2},+\frac{\pi}{2}\right].$$

This is sufficient to describe an arbitrary ellipse, because a rotation by $+90°$ leads to the same ellipse as a rotation by $-90°$. Finally, we reconstruct a and b. Due to

$$(A-C)^2+(2B)^2=\frac{(b^2-a^2)^2}{a^4b^4} \qquad \Rightarrow \sqrt{(A-C)^2+(2B)^2}=\frac{b^2-a^2}{a^2b^2}$$

and

$$A+C=\frac{a^2+b^2}{a^2b^2} \qquad \text{with} \quad A+C>0,$$

we obtain

$$A+C+\sqrt{(A-C)^2+(2B)^2}=2\frac{b^2}{a^2b^2}=\frac{2}{a^2}$$

and

$$A+C-\sqrt{(A-C)^2+(2B)^2}=2\frac{a^2}{a^2b^2}=\frac{2}{b^2},$$

which leads to

$$\boxed{a=\sqrt{\frac{2}{A+C+\sqrt{(A-C)^2+(2B)^2}}},} \tag{A.12}$$

$$\boxed{b=\sqrt{\frac{2}{A+C-\sqrt{(A-C)^2+(2B)^2}}}.} \tag{A.13}$$

The area of the ellipse is

$$\pi ab=\frac{2\pi}{\sqrt{(A+C)^2-[(A-C)^2+(2B)^2]}}=\frac{2\pi}{\sqrt{4AC-4B^2}}$$

$$\Rightarrow \boxed{\pi ab=\frac{\pi}{\sqrt{AC-B^2}}.} \tag{A.14}$$

This result may also be found, for example, in [1, volume 2, Sect. 339.6]. This implies that Eq. (A.6) describes an ellipse only if

$$\boxed{AC-B^2>0} \tag{A.15}$$

holds. As shown above and in order to make a and b real,

$$\boxed{A+C>0} \tag{A.16}$$

must be valid. The first inequality implies $AC > 0$, so that the constants A and C must either both be positive or both be negative. Together with the second inequality, this leads to

$$\boxed{A, C > 0.}$$

As a conclusion, the two conditions (A.15) and (A.16) imply all the conditions mentioned above. If the matrices

$$X = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad M = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$$

are defined, Eq. (A.6) may also be written in the form

$$\boxed{X^T \cdot M \cdot X = 1.}$$

The conditions (A.15) and (A.16) for an ellipse are obviously equivalent[1] to

$$\boxed{\det M > 0, \qquad \text{tr } M > 0.}$$

The transformation to the original ellipse may also be performed by means of a **principal axis transformation**. One may easily check that the eigenvalues of M are $\lambda_1 = 1/a^2$ and $\lambda_2 = 1/b^2$. The eigenvectors correspond to the direction of the **principal axes**.

We now analyze the case of a circle with radius $a = b > 0$. According to Eqs. (A.7)–(A.9), we obtain

$$A = C = \frac{1}{a^2}, \qquad B = 0.$$

The parameter ζ can no longer be determined, because the rotation matrix does not modify the circle. However, the equations and inequalities (A.12)–(A.16) are still valid. Therefore, the circle is included in the treatment as a special ellipse.

We have analyzed only ellipses that are centered at the origin of the coordinate system. In order to cover the most general case of an ellipse in the plane, one may add a translation in Eq. (A.6) if x and y are replaced by $x - x_0$ and $y - y_0$,

---

[1] For $\det M < 0$, a hyperbola is obtained.

respectively. This leads to the general equation

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0,$$

for which the decision whether it represents an ellipse becomes more complicated (cf. Burg et al. [2, volume II, Sect. 3.9.9]).

A.2 Path Length and Curvature

In this section, we determine the path length of a particle.

A.2.1 Path Length

As shown in Fig. A.2, the path is given by $\vec{r}(t)$, where t is an arbitrary parameter. According to Fig. A.2, we obtain

$$\Delta\vec{r} = \vec{r}(t + \Delta t) - \vec{r}(t).$$

If we divide this equation by Δt, the left-hand side will change its absolute value but not its direction. If one then considers the limit as $\Delta t \to 0$, it is clear that a tangent vector of arbitrary length is obtained:

$$\lim_{\Delta t \to 0} \frac{\vec{r}(t + \Delta t) - \vec{r}(t)}{\Delta t} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}.$$

According to

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \left|\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right| \vec{e}_t,$$

this tangent vector may be normalized in order to get the unit vector $\vec{e}_t$. The path length is obviously given by

$$\int \mathrm{d}s = \int \left|\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right| \mathrm{d}t,$$

where $\mathrm{d}s = |\mathrm{d}\vec{r}|$.

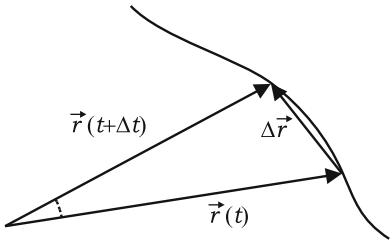

Fig. A.2 Calculation of path length

A.2.2 Curvature

It is well known that the **curvature** is the inverse of the **local radius** and that it is given by

$$\boxed{k_{\text{curv}} = \frac{1}{r_{\text{curv}}} = \left|\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}s^2}\right|}. \tag{A.17}$$

Here ds denotes the arc-length element of the curve. We will now check how this formula must be modified if the dependence of $\vec{r}$ on s is not known but the dependence on an arbitrary parameter t is given as above. For the derivative with respect to t, we use a dot from now on. For the first derivative, we obtain

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}s} = \dot{\vec{r}}\,\frac{\mathrm{d}t}{\mathrm{d}s}.$$

Since

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \left|\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right| = |\dot{\vec{r}}|$$

holds, we obtain

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}s} = \frac{\dot{\vec{r}}}{|\dot{\vec{r}}|}.$$

As the second derivative, we now obtain

$$\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}s^2} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{\vec{r}}}{|\dot{\vec{r}}|}\right)\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{1}{|\dot{\vec{r}}|}\,\frac{\ddot{\vec{r}}|\dot{\vec{r}}| - \frac{\ddot{\vec{r}}\cdot\dot{\vec{r}}}{|\dot{\vec{r}}|}\dot{\vec{r}}}{|\dot{\vec{r}}|^2} = \frac{\ddot{\vec{r}}}{|\dot{\vec{r}}|^2} - \frac{\ddot{\vec{r}}\cdot\dot{\vec{r}}}{|\dot{\vec{r}}|^4}\,\dot{\vec{r}}.$$

In the second step, we made use of

$$\frac{\mathrm{d}|\dot{\vec{r}}|}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\dot{\vec{r}}\cdot\dot{\vec{r}}} = \frac{2\,\ddot{\vec{r}}\cdot\dot{\vec{r}}}{2\sqrt{\dot{\vec{r}}\cdot\dot{\vec{r}}}} = \frac{\ddot{\vec{r}}\cdot\dot{\vec{r}}}{|\dot{\vec{r}}|}.$$

We obtain

$$\left|\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}s^2}\right|^2 = \frac{\mathrm{d}^2\vec{r}}{\mathrm{d}s^2}\cdot\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}s^2} = \frac{\ddot{\vec{r}}\cdot\ddot{\vec{r}}}{|\dot{\vec{r}}|^4} + \frac{(\ddot{\vec{r}}\cdot\dot{\vec{r}})^2}{|\dot{\vec{r}}|^8}\,|\dot{\vec{r}}|^2 - 2\frac{\ddot{\vec{r}}\cdot\dot{\vec{r}}}{|\dot{\vec{r}}|^6}\,\dddot{\vec{r}}\cdot\dot{\vec{r}}$$

$$\Rightarrow \left|\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}s^2}\right|^2 = \frac{|\ddot{\vec{r}}|^2|\dot{\vec{r}}|^2 - (\dot{\vec{r}}\cdot\ddot{\vec{r}})^2}{|\dot{\vec{r}}|^6}.$$

Finally, we obtain

$$\boxed{k_{\mathrm{curv}} = \left|\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}s^2}\right| = \frac{\sqrt{|\ddot{\vec{r}}|^2|\dot{\vec{r}}|^2 - (\dot{\vec{r}}\cdot\ddot{\vec{r}})^2}}{|\dot{\vec{r}}|^3}.} \tag{A.18}$$

For the sake of simplicity, we assume from now on that the whole path is located in a two-dimensional plane. In this case, we get

$$\begin{aligned}
|\ddot{\vec{r}}|^2 &= \ddot{x}^2 + \ddot{y}^2,\\
|\dot{\vec{r}}|^2 &= \dot{x}^2 + \dot{y}^2,\\
\dot{\vec{r}}\cdot\ddot{\vec{r}} &= \dot{x}\ddot{x} + \dot{y}\ddot{y}.
\end{aligned}$$

This leads to:

$$\begin{aligned}
|\ddot{\vec{r}}|^2|\dot{\vec{r}}|^2 - (\dot{\vec{r}}\cdot\ddot{\vec{r}})^2 &= \ddot{x}^2\dot{x}^2 + \ddot{x}^2\dot{y}^2 + \ddot{y}^2\dot{x}^2 + \ddot{y}^2\dot{y}^2 - \dot{x}^2\ddot{x}^2 - \dot{y}^2\ddot{y}^2 - 2\dot{x}\dot{y}\ddot{x}\ddot{y} =\\
&= \ddot{x}^2\dot{y}^2 + \ddot{y}^2\dot{x}^2 - 2\dot{x}\dot{y}\ddot{x}\ddot{y} =\\
&= (\ddot{x}\dot{y} - \ddot{y}\dot{x})^2.
\end{aligned}$$

As a special form of Eq. (A.18), we therefore get

$$\boxed{k_{\mathrm{curv,signed}} = \frac{\ddot{y}\dot{x} - \ddot{x}\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}.} \tag{A.19}$$

The sign was chosen in such a way that the curvature $k_{\mathrm{curv,signed}}$ is positive for a convex curve.

We finally analyze the unit vectors that are tangential or normal to the curve. With the results derived above, we obtain

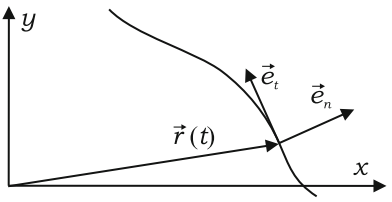

Fig. A.3 Normal and tangential unit vectors

$$\vec{e}_t = \frac{\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}}{\left|\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right|} = \frac{\dot{x}\,\vec{e}_x + \dot{y}\,\vec{e}_y}{\sqrt{\dot{x}^2+\dot{y}^2}}. \tag{A.20}$$

Figure A.3 shows that the corresponding normal unit vector $\vec{e}_n$ that satisfies $\vec{e}_n \cdot \vec{e}_t = 0$ may be obtained from $\vec{e}_t$ by exchanging the components in the following way:

$$\vec{e}_n = \frac{\dot{y}\,\vec{e}_x - \dot{x}\,\vec{e}_y}{\sqrt{\dot{x}^2+\dot{y}^2}}. \tag{A.21}$$

We now calculate the derivative of this normal unit vector:

$$\begin{aligned}
\frac{\mathrm{d}\vec{e}_n}{\mathrm{d}t} &= \frac{(\ddot{y}\,\vec{e}_x - \ddot{x}\,\vec{e}_y)\sqrt{\dot{x}^2+\dot{y}^2} - (\dot{y}\,\vec{e}_x - \dot{x}\,\vec{e}_y)\frac{2\dot{x}\ddot{x}+2\dot{y}\ddot{y}}{2\sqrt{\dot{x}^2+\dot{y}^2}}}{\dot{x}^2+\dot{y}^2} = \\
&= \frac{(\ddot{y}\,\vec{e}_x - \ddot{x}\,\vec{e}_y)(\dot{x}^2+\dot{y}^2) - (\dot{y}\,\vec{e}_x - \dot{x}\,\vec{e}_y)(\dot{x}\ddot{x}+\dot{y}\ddot{y})}{(\dot{x}^2+\dot{y}^2)^{3/2}} = \\
&= \frac{\vec{e}_x\,(\ddot{y}\dot{x}^2 + \ddot{y}\dot{y}^2 - \dot{y}\dot{x}\ddot{x} - \dot{y}^2\ddot{y}) + \vec{e}_y\,(-\ddot{x}\dot{x}^2 - \ddot{x}\dot{y}^2 + \dot{x}^2\ddot{x} + \dot{x}\dot{y}\ddot{y})}{(\dot{x}^2+\dot{y}^2)^{3/2}} = \\
&= \frac{\vec{e}_x\,\dot{x}(\ddot{y}\dot{x} - \dot{y}\ddot{x}) + \vec{e}_y\,\dot{y}(-\ddot{x}\dot{y} + \dot{x}\ddot{y})}{(\dot{x}^2+\dot{y}^2)^{3/2}}.
\end{aligned}$$

According to Eq. (A.19), this is equivalent to

$$\frac{\mathrm{d}\vec{e}_n}{\mathrm{d}t} = k_{\mathrm{curv,signed}}(\dot{x}\,\vec{e}_x + \dot{y}\,\vec{e}_y) = k_{\mathrm{curv,signed}}\,|\dot{\vec{r}}|\,\vec{e}_t. \tag{A.22}$$

The derivative of the tangential unit vector is

$$\begin{aligned}
\frac{\mathrm{d}\vec{e}_t}{\mathrm{d}t} &= \frac{(\ddot{x}\,\vec{e}_x + \ddot{y}\,\vec{e}_y)\sqrt{\dot{x}^2+\dot{y}^2} - (\dot{x}\,\vec{e}_x + \dot{y}\,\vec{e}_y)\frac{2\dot{x}\ddot{x}+2\dot{y}\ddot{y}}{2\sqrt{\dot{x}^2+\dot{y}^2}}}{\dot{x}^2+\dot{y}^2} = \\
&= \frac{(\ddot{x}\,\vec{e}_x + \ddot{y}\,\vec{e}_y)(\dot{x}^2+\dot{y}^2) - (\dot{x}\,\vec{e}_x + \dot{y}\,\vec{e}_y)(\dot{x}\ddot{x}+\dot{y}\ddot{y})}{(\dot{x}^2+\dot{y}^2)^{3/2}} =
\end{aligned}$$

$$= \frac{\vec{e}_x\,(\ddot{x}\dot{x}^2 + \dot{x}\dot{y}^2 - \ddot{x}\dot{x}^2 - \dot{x}\dot{y}\ddot{y}) + \vec{e}_y\,(\ddot{y}\dot{x}^2 + \ddot{y}\dot{y}^2 - \dot{y}\dot{x}\ddot{x} - \dot{y}^2\ddot{y})}{(\dot{x}^2 + \dot{y}^2)^{3/2}} =$$

$$= \frac{\vec{e}_x\;\dot{y}(\ddot{x}\dot{y} - \dot{x}\ddot{y}) + \vec{e}_y\;\dot{x}(\ddot{y}\dot{x} - \dot{y}\ddot{x})}{(\dot{x}^2 + \dot{y}^2)^{3/2}}$$

$$\Rightarrow \frac{\mathrm{d}\vec{e}_t}{\mathrm{d}t} = k_{\text{curv,signed}}(-\dot{y}\;\vec{e}_x + \dot{x}\;\vec{e}_y) = -k_{\text{curv,signed}}\;|\dot{\vec{r}}|\;\vec{e}_n. \tag{A.23}$$

Similar formulas exist for curves in three dimensions. They are known as **Frenet–Serret formulas**.

A.2.3 Centripetal Force

As a simple application of the formulas derived so far, we briefly calculate the centripetal force for a particle that follows an arbitrary curve in the xy-plane. We shall assume that the particle has a constant kinetic energy, so that $\dot{m} = 0$ holds. Due to

$$\vec{u} = \dot{x}\;\vec{e}_x + \dot{y}\;\vec{e}_y \qquad \Rightarrow u = \sqrt{\dot{x}^2 + \dot{y}^2},$$

we obtain, by means of Eq. (A.19),

$$\pm\frac{m\;u^2}{r_{\text{curv}}} = m\;u^2\;k_{\text{curv,signed}} = m\;\frac{\ddot{y}\dot{x} - \ddot{x}\dot{y}}{u}. \tag{A.24}$$

On the other hand, we get

$$\vec{F} = m\;\dot{\vec{u}} = m\;(\ddot{x}\vec{e}_x + \ddot{y}\vec{e}_y).$$

The center of the osculating circle is located in the direction normal to the tangent vector. Therefore, we calculate the normal component of $\vec{F}$. By means of Eq. (A.21), we see that

$$\vec{F}\cdot\vec{e}_n = m\;\frac{\ddot{x}\dot{y} - \ddot{y}\dot{x}}{u}$$

holds. If we compare this with Eq. (A.24), we obtain

$$F_n = -\frac{m\;u^2}{r_{\text{curv}}},$$

as expected.

A.3 Some Results Concerning Transverse Optics in Synchrotrons

In this section, some results concerning transverse beam dynamics (linear case) are summarized. They may be found, for instance, in [3–5]. As discussed before, a reference particle with reference momentum p_R travels on the reference path (**closed orbit**) of length l_R. In transverse optics, one often uses the coordinates x, y, s. Here x is the horizontal coordinate (positive for larger radii), y is the vertical coordinate (positive for a larger height), and $s \in [0, l_\mathrm{R}]$ measures the length on the reference path. For the reference path, we have $x = y = 0$. Instead of $s \in [0, l_\mathrm{R}]$, one may use an angle variable $\theta \in [0, 2\pi]$ describing the azimuthal position in the ring. A particle with momentum $p_\mathrm{R} + \Delta p$ will experience a displacement

$$\Delta x = D_x(s)\,\frac{\Delta p}{p_\mathrm{R}}$$

from the reference orbit (cf. Edwards/Syphers [3, Sect. 3.3.1]) if $\Delta p \ll p_\mathrm{R}$ holds. The function $D_x(s)$ is called the momentum dispersion function or simply **dispersion function**. From $D_x(s) = D_x(s+l_\mathrm{R})$, we see that it is a periodic function for typical synchrotron lattices.

Let us assume that the reference path is given by $\vec{r}_0(\theta)$. In this case, the length of the reference path is

$$l_\mathrm{R} = \int \mathrm{d}r_0 = \int_0^{2\pi} \left|\frac{\mathrm{d}\vec{r}_0}{\mathrm{d}\theta}\right| \mathrm{d}\theta. \tag{A.25}$$

The path for $\delta = \Delta p/p_\mathrm{R} \neq 0$ is given by

$$\vec{r}(\theta) = \vec{r}_0(\theta) + D_x(\theta)\delta\vec{e}_x(\theta).$$

The derivative with respect to θ is

$$\frac{\mathrm{d}\vec{r}}{\mathrm{d}\theta} = \frac{\mathrm{d}\vec{r}_0}{\mathrm{d}\theta} + D_x\delta\,\frac{\mathrm{d}\vec{e}_x}{\mathrm{d}\theta} + \frac{\mathrm{d}D_x}{\mathrm{d}\theta}\,\delta\vec{e}_x.$$

For the sake of simplicity, we write derivatives with respect to θ with a dot:

$$\dot{\vec{r}} = \dot{\vec{r}}_0 + D_x\,\delta\dot{\vec{e}}_x + \dot{D}_x\,\delta\vec{e}_x.$$

Due to Eq. (A.22) we have

$$\dot{\vec{e}}_x = \vec{e}_s\,k|\dot{\vec{r}}_0|,$$

where $k(\theta) = k_\mathrm{curv,signed} = 1/\rho(\theta)$ is the local curvature. This leads to

$$\dot{\vec{r}} = \dot{\vec{r}}_0 + D_x\ \delta\ k|\dot{\vec{r}}_0|\vec{e}_s + \dot{D}_x\ \delta\vec{e}_x.$$

Taking into account that

$$\dot{\vec{r}}_0 = |\dot{\vec{r}}_0|\ \vec{e}_s$$

is valid, one obtains

$$\dot{\vec{r}} \cdot \dot{\vec{r}} = \left(|\dot{\vec{r}}_0| + D_x\ \delta\ k|\dot{\vec{r}}_0|\right)^2 + \left(\dot{D}_x\ \delta\right)^2$$

$$\Rightarrow |\dot{\vec{r}}| = |\dot{\vec{r}}_0|\ \sqrt{(1 + D_x\ \delta\ k)^2 + \frac{\dot{D}_x^2\ \delta^2}{|\dot{\vec{r}}_0|^2}}.$$

The term with $\dot{D}_x$ is usually much smaller than that with D_x, so that

$$|\dot{\vec{r}}| \approx |\dot{\vec{r}}_0|\ (1 + D_x\ \delta\ k)$$

holds. Hence, as the path length of the off-momentum particle, we obtain

$$l = \int_0^{2\pi} \left|\frac{\mathrm{d}\vec{r}}{\mathrm{d}\theta}\right| \mathrm{d}\theta = \int_0^{2\pi} |\dot{\vec{r}}|\ \mathrm{d}\theta = \int_0^{2\pi} |\dot{\vec{r}}_0|\ \mathrm{d}\theta + \int_0^{2\pi} |\dot{\vec{r}}_0|\ D_x \delta k\ \mathrm{d}\theta.$$

According to Eq. (A.25), the first integral on the right-hand side equals l_R. Therefore, taking into account that $\mathrm{d}s = |\dot{\vec{r}}_0|\ \mathrm{d}\theta$ holds, we see that

$$\Delta l = l - l_\mathrm{R} = \int_0^{2\pi} |\dot{\vec{r}}_0|\ D_x \delta k\ \mathrm{d}\theta = \delta \int_0^{l_\mathrm{R}} D_x k\ \mathrm{d}s$$

is valid. The quantity Δl obviously specifies the length difference between the closed orbit of an off-momentum particle and that of a particle with the reference momentum.

This leads to the momentum compaction factor

$$\alpha_\mathrm{c} = \frac{\Delta l / l_\mathrm{R}}{\Delta p / p_\mathrm{R}} = \frac{\Delta l / l_\mathrm{R}}{\delta} = \frac{1}{l_\mathrm{R}} \int_0^{l_\mathrm{R}} D_x(s)\ k(s)\ \mathrm{d}s.$$

Since the curvature equals the inverse local radius ($k(s) = 1/\rho(s)$), we obtain

$$\alpha_\mathrm{c} = \frac{1}{l_\mathrm{R}} \int_0^{l_\mathrm{R}} \frac{D_x(s)}{\rho(s)}\ \mathrm{d}s. \tag{A.26}$$

Hence the momentum compaction factor is the average of the fraction $D_x(s)/\rho(s)$.

The particles perform **betatron oscillations** around their closed orbit in the transverse case in a manner similar to the synchrotron oscillation in the longitudinal case:

$$x(s) = A\sqrt{\beta_x(s)}\,\cos\left[\psi_x(s) + \psi_{x,0}\right]. \tag{A.27}$$

This formula satisfies the **Mathieu–Hill differential equation**, or **Hill's differential equation**

$$\frac{\mathrm{d}^2x}{\mathrm{d}s^2} + K_x(s)\,x = 0 \qquad \text{with} \qquad K_x(s + l_R) = K_x(s)$$

if

$$K_x(s)\beta_x^2(s) - \frac{1}{4}\left(\frac{\mathrm{d}\beta_x}{\mathrm{d}s}\right)^2 + \frac{1}{2}\beta_x(s)\frac{\mathrm{d}^2\beta_x}{\mathrm{d}s^2} = 1$$

and

$$\frac{\mathrm{d}\psi_x}{\mathrm{d}s} = \frac{1}{\beta_x(s)} \qquad \Leftrightarrow \qquad \psi_x(s) = \int \frac{\mathrm{d}s}{\beta_x(s)} \tag{A.28}$$

are valid. The periodic function $K_x(s)$ is determined by the lattice. The function $\psi_x(s)$ is the **phase function**, while the function $\beta_x(s)$ is the **amplitude function**, **betatron function**, or **beta function**. The beta function obviously determines the envelope of the beam. For standard synchrotron lattices, it is a periodic function with $\beta_x(s + l_\mathrm{R}) = \beta_x(s)$.

The number of betatron oscillations per revolution (which is not an integer for proper synchrotron operation)[2] is called the (horizontal or vertical) **tune**:

$$\nu_x = \frac{T_\mathrm{R}}{T_{\beta,x}} = \frac{f_{\beta,x}}{f_\mathrm{R}}, \qquad \nu_y = \frac{T_\mathrm{R}}{T_{\beta,y}} = \frac{f_{\beta,y}}{f_\mathrm{R}}.$$

Due to Eq. (A.28), it is obtained by

$$\nu_x = \frac{1}{2\pi}\int_0^{l_\mathrm{R}} \frac{\mathrm{d}s}{\beta_x(s)}.$$

If the momentum of the particle differs from the reference momentum p_R, the tune will change according to

[2]Moreover, fractions consisting of small integers have to be avoided, because the beam is then repeatedly subject to transverse kicks in a similar state of oscillation, so that it becomes unstable (**resonance**). The pair (ν_x, ν_y) is the **working point** of the machine, and it is usually visualized in a **working diagram** showing the resonance lines of different orders (cf. [5, Sect. 6.4.1]).

$$\Delta\nu_x = \xi_x(p)\,\frac{\Delta p}{p_\mathrm{R}}, \qquad \Delta\nu_y = \xi_y(p)\,\frac{\Delta p}{p_\mathrm{R}}.$$

Here ξ_x, ξ_y denotes the **chromaticity**.

If the derivative with respect to s is denoted by a prime, Eq. (A.27) leads to

$$x'(s) = \frac{\mathrm{d}x}{\mathrm{d}s} = A\frac{\beta_x'(s)}{2\sqrt{\beta_x(s)}}\,\cos\left[\psi_x(s)+\psi_{x,0}\right] - A\frac{1}{\sqrt{\beta_x(s)}}\,\sin\left[\psi_x(s)+\psi_{x,0}\right] \tag{A.29}$$

if Eq. (A.28) is taken into account. The pair (x, x') defines the phase space[3] in the horizontal plane. If one uses Eqs. (A.27) and (A.29) to eliminate the sine and cosine functions, one obtains

$$x'^2\frac{\beta_x}{A^2} - 2xx'\frac{\beta_x'}{2A^2} + x^2\left(\frac{1}{A^2\beta_x} + \frac{\beta_x'^{\,2}}{4A^2\beta_x}\right) = 1. \tag{A.30}$$

This is the equation

$$\tilde{\gamma}_x x^2 + 2\tilde{\alpha}_x xx' + \tilde{\beta}_x x'^2 = 1$$

of an ellipse in the (x, x')-plane ($\tilde{\gamma}_x\tilde{\beta}_x - \tilde{\alpha}_x^2 > 0$, $\tilde{\beta}_x > 0$), whose area is

$$\frac{\pi}{\sqrt{\tilde{\gamma}_x\tilde{\beta}_x - \tilde{\alpha}^2}} = \frac{\pi}{\sqrt{\frac{1}{A^4} + \frac{\beta_x'^{\,2}}{4A^4} - \frac{\beta_x'^{\,2}}{4A^4}}} = \pi A^2,$$

as we found in Eq. (A.14). If the outermost particle with maximum A is considered, the total phase space area enclosed by all particles can (similar to the longitudinal case) be used to define the **transverse emittance** ϵ_x,

$$\pi\epsilon_x = \pi A^2,$$

in the (x, x')-plane.[4] In this case, one may set $A = \sqrt{\epsilon_x}$ in Eqs. (A.27) and (A.29), and Eq. (A.30) can be written as the **Courant–Snyder invariant**

$$\gamma_x x^2 + 2\alpha_x xx' + \beta_x x'^2 = \epsilon_x$$

[3] The (x, x')-plane is sometimes called trace space (cf. [6]) in order to distinguish it from the overall phase space, which has six dimensions and whose volume is conserved due to the choice of canonical coordinates.

[4] Sometimes, the emittance is defined with different constant factors. Similarly to the longitudinal case, one may define a transverse RMS emittance to cope with the difficulties in defining an area for a finite number of particles.

with **Courant–Snyder parameters** or **Twiss parameters**

$$\alpha_x = -\frac{\beta_x'}{2}, \quad \beta_x, \quad \gamma_x = \frac{1}{\beta_x}\left(1 + \frac{{\beta_x'}^2}{4}\right).$$

In the main chapters of this book, we pointed out that the area in the phase space $(\Delta t, \Delta W)$ is preserved if no synchrotron radiation or other damping or blowup mechanisms are present. Due to

$$\begin{aligned} \Delta s &= v_\mathrm{R}\Delta t = \beta_\mathrm{R} c_0 \Delta t, \\ \frac{\Delta W}{W_\mathrm{R}} &= \beta_\mathrm{R}^2 \frac{\Delta p}{p_\mathrm{R}} \quad \Rightarrow \Delta p = \frac{1}{\beta_\mathrm{R} c_0}\Delta W, \end{aligned}$$

a simple scale transformation converts the $(\Delta t, \Delta W)$ phase space into the (s, p_s) phase space[5] (cf. Sect. 5.1.3: one coordinate is multiplied by $c_0\beta_\mathrm{R}$, and the other one is divided by the same factor). Therefore, the phase space (s, p_s) also leads to area preservation.

In the transverse case, area preservation is given in an analogous way if the phase space (x, p_x) is used. However, we have used the pair (x, x') so far. According to

$$p_x = m\frac{\mathrm{d}x}{\mathrm{d}t} = m_0\gamma_\mathrm{R}\frac{\mathrm{d}x}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = m_0 c_0 \gamma_\mathrm{R}\beta_\mathrm{R} x',$$

we also have to multiply the (x, x') area obtained above by $\gamma_\mathrm{R}\beta_\mathrm{R}$ in order to obtain an invariant quantity:

$$\pi\epsilon_x\gamma_\mathrm{R}\beta_\mathrm{R} = \text{const.}$$

Therefore, one defines the **normalized transverse emittance**

$$\epsilon_{\mathrm{n}x} = \gamma_\mathrm{R}\beta_\mathrm{R}\epsilon_x,$$

which is preserved (also during acceleration) if no synchrotron radiation or other damping or blowup effects are present. According to Eq. (A.27), the maximum amplitude of the betatron oscillation is

$$\hat{x}_\mathrm{max}(s) = A\sqrt{\beta_x(s)} = \sqrt{\epsilon_x\beta_x(s)} = \sqrt{\frac{\epsilon_{\mathrm{n}x}\beta_x(s)}{\gamma_\mathrm{R}\beta_\mathrm{R}}}.$$

Therefore, the betatron oscillations are damped during acceleration. This is called **adiabatic damping**.

[5] Here the coordinates from our longitudinal phase space considerations were renamed according to $p_s = \Delta p$, $s = \Delta s$.

A.4 Characterization of Fixed Points for the Longitudinal Beam Dynamics System

The starting point for the classification of fixed points is the Hamiltonian in Eq. (3.29),

$$H = -\frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2}\,\frac{\Delta W^2}{2} - \frac{Q\hat{V}}{T_\mathrm{R}}\left(\frac{1}{\omega_\mathrm{RF}}\left[\cos(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}) - \cos\varphi_\mathrm{R}\right] + \Delta t\ \sin\varphi_\mathrm{R}\right), \tag{A.31}$$

which depends on $q = \Delta W$ and $p = \Delta t$. We calculate the partial derivatives:

$$\frac{\partial H}{\partial \Delta W} = -\frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2}\,\Delta W,$$

$$\frac{\partial H}{\partial \Delta t} = -\frac{Q\hat{V}}{T_\mathrm{R}}\left(-\sin(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}) + \sin\varphi_\mathrm{R}\right).$$

This leads to the following second-order partial derivatives:

$$\frac{\partial^2 H}{\partial \Delta W\,\partial \Delta t} = 0,$$

$$\frac{\partial^2 H}{\partial \Delta W^2} = -\frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2},$$

$$\frac{\partial^2 H}{\partial \Delta t^2} = \frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\ \cos(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}).$$

By means of

$$\vec{v} = \begin{pmatrix}\dot{q}\\ \dot{p}\end{pmatrix} = \begin{pmatrix}\frac{\partial H}{\partial p}\\ -\frac{\partial H}{\partial q}\end{pmatrix},$$

one obtains the Jacobian matrix

$$D\vec{v} = \begin{pmatrix}\frac{\partial^2 H}{\partial q\,\partial p} & \frac{\partial^2 H}{\partial p^2}\\ -\frac{\partial^2 H}{\partial q^2} & -\frac{\partial^2 H}{\partial q\,\partial p}\end{pmatrix}.$$

First, we consider the fixed point $\vec{r}_{F1}$, for which $\Delta t = 0$ and $\Delta W = 0$ hold. The corresponding Jacobian matrix is

$$D\vec{v}(\vec{r}_{F1}) = \begin{pmatrix} 0 & \frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\,\cos\varphi_\mathrm{R} \\ \frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2} & 0 \end{pmatrix}.$$

For its eigenvalues, one obtains

$$\lambda^2 - \frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2}\frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\,\cos\varphi_\mathrm{R} = 0$$

$$\Rightarrow \lambda^2 = \frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2}\frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\,\cos\varphi_\mathrm{R}.$$

Due to $\eta_\mathrm{R} Q\,\cos\varphi_\mathrm{R} < 0$ (Eq. (3.26)), two imaginary eigenvalues exist. Hence, the linearized problem has a center at $\vec{r}_{F1}$, i.e., a stable fixed point.

As we discussed in Sect. 2.11.5, one cannot conclude that the nonlinear system also has a center at $\vec{r}_{F1}$. However, the alternative approach that was presented in Sect. 2.11.5 is successful, as the following discussion demonstrates.

We would like to show that H is a Lyapunov function. For the sake of simplicity, we shall consider only $Q > 0$, $\eta_\mathrm{R} < 0$. The term in Eq. (A.31) that depends on ΔW is then always greater than or equal to zero. The term that depends on Δt has a minimum at $\Delta t = 0$ for $-\pi/2 < \varphi_\mathrm{R} < \pi/2$, since we have

$$\left.\frac{\partial H}{\partial \Delta t}\right|_{\Delta t=0} = 0$$

and

$$\left.\frac{\partial^2 H}{\partial \Delta t^2}\right|_{\Delta t=0} > 0.$$

Therefore, this term is greater than zero in a neighborhood of $\Delta t = 0$. It is clear from the general statements presented in Sect. 2.11.5 that H is a Lyapunov function and that $\vec{r}_{F1}$ is a stable fixed point. Therefore, the nonlinear Hamiltonian system also has a center at $\vec{r}_{F1}$.

Now we consider the fixed point $\vec{r}_{F2}$ at $\Delta t = \frac{\pi\,\mathrm{sign}\,\varphi_\mathrm{R} - 2\varphi_\mathrm{R}}{\omega_\mathrm{RF}}$ and $\Delta W = 0$. Due to

$$\left.\frac{\partial^2 H}{\partial \Delta t^2}\right|_{\omega_\mathrm{RF}\Delta t = \pi\,\mathrm{sign}\,\varphi_\mathrm{R} - 2\varphi_\mathrm{R}} = \frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\,\cos(\pi\,\mathrm{sign}\,\varphi_\mathrm{R} - \varphi_\mathrm{R}) = -\frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\,\cos\,\varphi_\mathrm{R},$$

the Jacobian matrix is

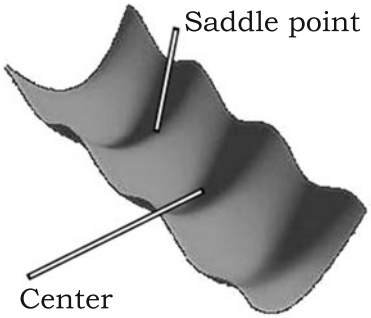

Fig. A.4 Hamiltonian $H(\Delta W, \Delta t)$ as a surface

$$D\vec{v}(\vec{r}_{F2}) = \begin{pmatrix} 0 & -\frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\,\cos\varphi_\mathrm{R} \\ \frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2} & 0 \end{pmatrix}.$$

For the eigenvalues, one obtains

$$\lambda^2 + \frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2}\frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\,\cos\varphi_\mathrm{R} = 0$$

$$\Rightarrow \lambda^2 = -\frac{\eta_\mathrm{R}}{W_\mathrm{R}\beta_\mathrm{R}^2}\frac{Q\hat{V}}{T_\mathrm{R}}\,\omega_\mathrm{RF}\,\cos\varphi_\mathrm{R}.$$

Due to $\eta_\mathrm{R} Q\,\cos\varphi_\mathrm{R} < 0$, there are two real eigenvalues with opposite signs; $\vec{r}_{F2}$ is a saddle point of the linearized system. Since the saddle point is a hyperbolic fixed point, the nonlinear system also has a saddle point (see Sect. 2.8.9).

In Fig. A.4, the Hamiltonian $H(\Delta W, \Delta t)$ is drawn as a surface in three-dimensional space for $\varphi_\mathrm{R} = 20°$. The particles move on the level curves of this mountain-like surface, because an autonomous system was assumed such that the Hamiltonian remains constant along the trajectory. If these level curves are drawn in the $(\Delta W, \Delta t)$ plane, a diagram like the upper one shown in Fig. 3.7 is obtained (except for the different choice of φ_R).

A.5 Change of Variables for Multiple Integrals

Consider a surface integral for which the surface is parameterized by α and β:

$$F = \int\int \Phi(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\beta_\mathrm{min}}^{\beta_\mathrm{max}} \int_{\alpha_\mathrm{min}}^{\alpha_\mathrm{max}} \Phi(\alpha, \beta)\,\left|\frac{\partial(x, y)}{\partial(\alpha, \beta)}\right|\,\mathrm{d}\alpha\,\mathrm{d}\beta.$$

Now we transform the coordinates x and y into new coordinates x' and y', and we consider an analogous surface integral:

$$F' = \int\int \Phi(x', y')\, \mathrm{d}x'\, \mathrm{d}y' = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} \Phi(\alpha, \beta) \left| \frac{\partial(x', y')}{\partial(\alpha, \beta)} \right| \mathrm{d}\alpha\, \mathrm{d}\beta.$$

We define

$$\xi = \frac{\partial(x, y)}{\partial(\alpha, \beta)} = \begin{vmatrix} \frac{\partial x}{\partial \alpha} & \frac{\partial x}{\partial \beta} \\ \frac{\partial y}{\partial \alpha} & \frac{\partial y}{\partial \beta} \end{vmatrix} = \frac{\partial x}{\partial \alpha}\, \frac{\partial y}{\partial \beta} - \frac{\partial x}{\partial \beta}\, \frac{\partial y}{\partial \alpha},$$

$$\xi' = \frac{\partial(x', y')}{\partial(\alpha, \beta)} = \begin{vmatrix} \frac{\partial x'}{\partial \alpha} & \frac{\partial x'}{\partial \beta} \\ \frac{\partial y'}{\partial \alpha} & \frac{\partial y'}{\partial \beta} \end{vmatrix} = \frac{\partial x'}{\partial \alpha}\, \frac{\partial y'}{\partial \beta} - \frac{\partial x'}{\partial \beta}\, \frac{\partial y'}{\partial \alpha}$$

$$\Rightarrow \xi' = \left(\frac{\partial x'}{\partial x}\, \frac{\partial x}{\partial \alpha} + \frac{\partial x'}{\partial y}\, \frac{\partial y}{\partial \alpha} \right) \left(\frac{\partial y'}{\partial x}\, \frac{\partial x}{\partial \beta} + \frac{\partial y'}{\partial y}\, \frac{\partial y}{\partial \beta} \right) -$$

$$- \left(\frac{\partial x'}{\partial x}\, \frac{\partial x}{\partial \beta} + \frac{\partial x'}{\partial y}\, \frac{\partial y}{\partial \beta} \right) \left(\frac{\partial y'}{\partial x}\, \frac{\partial x}{\partial \alpha} + \frac{\partial y'}{\partial y}\, \frac{\partial y}{\partial \alpha} \right) =$$

$$= \frac{\partial x'}{\partial x}\, \frac{\partial y'}{\partial y} \left(\frac{\partial x}{\partial \alpha}\, \frac{\partial y}{\partial \beta} - \frac{\partial x}{\partial \beta}\, \frac{\partial y}{\partial \alpha} \right) + \frac{\partial x'}{\partial y}\, \frac{\partial y'}{\partial x} \left(\frac{\partial x}{\partial \beta}\, \frac{\partial y}{\partial \alpha} - \frac{\partial x}{\partial \alpha}\, \frac{\partial y}{\partial \beta} \right)$$

$$\Rightarrow \xi' = \left(\frac{\partial x'}{\partial x}\, \frac{\partial y'}{\partial y} - \frac{\partial x'}{\partial y}\, \frac{\partial y'}{\partial x} \right) \xi = \xi \begin{vmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{vmatrix}.$$

If we write this in a different way, we obtain the **chain rule for Jacobian determinants**:

$$\boxed{\frac{\partial(x', y')}{\partial(\alpha, \beta)} = \frac{\partial(x', y')}{\partial(x, y)}\, \frac{\partial(x, y)}{\partial(\alpha, \beta)}.}$$

Hence we have shown that

$$F' = \int\int \Phi(x', y')\, \mathrm{d}x'\, \mathrm{d}y' = \int_{\beta_{\min}}^{\beta_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} \Phi(\alpha, \beta) \left| \frac{\partial(x', y')}{\partial(x, y)} \right| \left| \frac{\partial(x, y)}{\partial(\alpha, \beta)} \right| \mathrm{d}\alpha\, \mathrm{d}\beta =$$

$$= \int\int \Phi(x, y) \left| \frac{\partial(x', y')}{\partial(x, y)} \right| \mathrm{d}x\, \mathrm{d}y$$

holds. This is the **transformation rule for double integrals**, which may be generalized to multiple integrals.

The surface integral obviously keeps its value according to

$$F' = F$$

if the absolute value of the Jacobian equals 1:

$$\left| \frac{\partial(x', y')}{\partial(x, y)} \right| = 1.$$

This is the condition for area preservation in two dimensions. For higher dimensions, one speaks of volume preservation in n-dimensional space:

$$\boxed{\left| \frac{\partial(x_1', x_2', x_3', \ldots, x_n')}{\partial(x_1, x_2, x_3, \ldots, x_n)} \right| = 1.}$$

A.6 Characteristic Equation and Companion Matrix

In Sect. 2.8.10, the relationship between the characteristic equation and the companion matrix of an ODE was presented. Here, we will show that the roots of the characteristic equation are in fact equal to the eigenvalues of the companion matrix.

For this purpose, we will need the following $n \times n$ determinant:

$$D_n^\lambda := \begin{vmatrix} -\lambda & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\lambda & 1 & \ldots & 0 & 0 \\ 0 & 0 & -\lambda & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -\lambda & 1 \\ 0 & 0 & 0 & \ldots & 0 & -\lambda \end{vmatrix}.$$

One way to calculate this determinant is to consider a Laplace expansion along the first column. This leads directly to

$$D_n^\lambda = -\lambda \; D_{n-1}^\lambda.$$

Since $D_1^\lambda = -\lambda$ is obvious, we obtain

$$D_n^\lambda = (-\lambda)^n.$$

We begin with the proof that the eigenvalues of the companion matrix are identical to the solutions of the characteristic equation. We use mathematical induction. In Sect. 2.8.10, we showed that the statement (Eq. (2.114)) is true for

$n = 2$ and $n = 3$ (basis). For the inductive step, we now assume that the statement is true for $n-1$, so that we have to show that it holds for n.

According to Eqs. (2.112) and (2.113), we obtain

$$D_n^F = \det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & -\lambda & 1 \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & -\frac{a_2}{a_n} & -\frac{a_3}{a_n} & \dots & -\frac{a_{n-2}}{a_n} & -\frac{a_{n-1}}{a_n} - \lambda \end{vmatrix},$$

where the determinant is an $n \times n$ array. We expand the determinant along the last column:

$$\begin{aligned} D_n^F = -1 \cdot & \begin{vmatrix} -\lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & -\lambda & 1 \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & -\frac{a_2}{a_n} & -\frac{a_3}{a_n} & \dots & -\frac{a_{n-3}}{a_n} & -\frac{a_{n-2}}{a_n} \end{vmatrix} + \\ & + \left(-\frac{a_{n-1}}{a_n} - \lambda \right) \begin{vmatrix} -\lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\lambda & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & -\lambda \end{vmatrix}. \end{aligned} \tag{A.32}$$

These are $(n-1) \times (n-1)$ determinants. The second determinant obviously equals $D_{n-1}^{\lambda} = (-\lambda)^{n-1}$. The first determinant is similar—but not equal—to D_{n-1}^F. One difference is that in D_{n-1}^F, an additional term $-\lambda$ in the last element is present. The other difference is that in the last row of D_{n-1}^F, the denominator is a_{n-1} instead of a_n. We therefore now regard the last column of D_{n-1}^F as the sum of two column vectors, one with $-\lambda$ as the last element, the other with $-\frac{a_{n-2}}{a_{n-1}}$ as the last element. Therefore, we obtain

$$
D_{n-1}^F = \begin{vmatrix}
-\lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & -\lambda & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & -\lambda & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & 1 & 0 \\
0 & 0 & 0 & 0 & \dots & -\lambda & 1 \\
-\frac{a_0}{a_{n-1}} & -\frac{a_1}{a_{n-1}} & -\frac{a_2}{a_{n-1}} & -\frac{a_3}{a_{n-1}} & \dots & -\frac{a_{n-3}}{a_{n-1}} & -\frac{a_{n-2}}{a_{n-1}} - \lambda
\end{vmatrix} =
$$

$$
= \begin{vmatrix}
-\lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & -\lambda & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & -\lambda & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & 1 & 0 \\
0 & 0 & 0 & 0 & \dots & -\lambda & 1 \\
-\frac{a_0}{a_{n-1}} & -\frac{a_1}{a_{n-1}} & -\frac{a_2}{a_{n-1}} & -\frac{a_3}{a_{n-1}} & \dots & -\frac{a_{n-3}}{a_{n-1}} & -\frac{a_{n-2}}{a_{n-1}}
\end{vmatrix} +
$$

$$
+ \begin{vmatrix}
-\lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & -\lambda & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & -\lambda & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & 1 & 0 \\
0 & 0 & 0 & 0 & \dots & -\lambda & 0 \\
-\frac{a_0}{a_{n-1}} & -\frac{a_1}{a_{n-1}} & -\frac{a_2}{a_{n-1}} & -\frac{a_3}{a_{n-1}} & \dots & -\frac{a_{n-3}}{a_{n-1}} & -\lambda
\end{vmatrix}.
$$

The last determinant may be expanded along the last column, which leads directly to $-\lambda \cdot D_{n-2}^{\lambda} = (-\lambda)^{n-1}$. If we multiply the last row of the first of the two determinants on the right-hand side by $\frac{a_{n-1}}{a_n}$, we get the determinant that we need in Eq. (A.32). Hence, this determinant equals

$$
\frac{a_{n-1}}{a_n} \left(D_{n-1}^F - (-\lambda)^{n-1} \right),
$$

which may be used in Eq. (A.32):

$$
D_n^F = -\frac{a_{n-1}}{a_n} \left(D_{n-1}^F - (-\lambda)^{n-1} \right) + \left(-\frac{a_{n-1}}{a_n} - \lambda \right) (-\lambda)^{n-1} = -\frac{a_{n-1}}{a_n} D_{n-1}^F + (-\lambda)^n .
$$

According to the assumption of our induction step,

$$
D_{n-1}^F = (-1)^{n-1} \sum_{k=0}^{n-1} \frac{a_k}{a_{n-1}} \lambda^k
$$

is valid, so that

$$D_n^F = -(-1)^{n-1} \sum_{k=0}^{n-1} \frac{a_k}{a_n} \lambda^k + (-\lambda)^n = (-1)^n \sum_{k=0}^{n-1} \frac{a_k}{a_n} \lambda^k + (-1)^n \lambda^n$$
$$= (-1)^n \sum_{k=0}^{n} \frac{a_k}{a_n} \lambda^k$$

follows. This completes the induction step and the proof.

A.7 Cavity Response to Excitations

In this section, we analyze how an ideal cavity with the standard lumped element circuit shown in Fig. 4.5 on p. 181 reacts to different types of excitation.

A.7.1 Amplitude Jumps

First of all, we consider the case that the generator current is switched on and that the beam current is negligible. We are interested in the resulting change in the voltage.

In other words, we have to solve the ODE (4.28),

$$\ddot{V}_{\text{gap}} + \frac{2}{\tau} \dot{V}_{\text{gap}} + \omega_{\text{res}}^2 V_{\text{gap}} = \frac{1}{C} \frac{\mathrm{d}}{\mathrm{d}t} (I_{\text{gen}} - I_{\text{beam}}),$$

for the generator current

$$I_{\text{gen}}(t) = \begin{cases} 0 & t < 0, \\ \hat{I}_{\text{gen}} \ \sin(\omega t) & t > 0, \end{cases}$$

with

$$I_{\text{beam}}(t) = 0.$$

A.7.1.1 Solution of the Homogeneous Equation

We use the ansatz

$$V_{\text{gap}} \sim e^{\alpha t},$$
$$\dot{V}_{\text{gap}} \sim \alpha \ e^{\alpha t},$$
$$\ddot{V}_{\text{gap}} \sim \alpha^2 \ e^{\alpha t},$$

and insert it into the homogeneous ODE:

$$\alpha^2 + \frac{2}{\tau}\alpha + \omega_{\text{res}}^2 = 0$$

$$\Rightarrow \alpha = -\frac{1}{\tau} \pm \sqrt{\left(\frac{1}{\tau}\right)^2 - \omega_{\text{res}}^2}.$$

For the quality factor, we assume

$$Q_{\text{p}} = \frac{\omega_{\text{res}}\tau}{2} > \frac{1}{2},$$

so that

$$\omega_{\text{res}}^2 > \frac{1}{\tau^2}$$

holds. Therefore, the argument of the square root is always negative.

We obtain

$$\alpha = -\frac{1}{\tau} \pm \sqrt{\frac{1}{\tau^2} - \omega_{\text{res}}^2}$$

$$\Rightarrow \alpha = -\frac{1}{\tau} \pm j\omega_{\text{d}}$$

with

$$\boxed{\omega_{\text{d}} = \sqrt{\omega_{\text{res}}^2 - \frac{1}{\tau^2}} = \omega_{\text{res}}\sqrt{1 - \frac{1}{\tau^2\omega_{\text{res}}^2}} = \omega_{\text{res}}\sqrt{1 - \frac{1}{4Q_{\text{p}}^2}}.} \tag{A.33}$$

Thus, we may write the ansatz given above in the following form:

$$V_{\text{gap}} = e^{-t/\tau}\left(A\ \cos(\omega_{\text{d}}t) + B\ \sin(\omega_{\text{d}}t)\right), \tag{A.34}$$

$$\dot{V}_{\text{gap}} = e^{-t/\tau}\left(\cos(\omega_{\text{d}}t)\left[-\frac{A}{\tau} + B\omega_{\text{d}}\right] + \sin(\omega_{\text{d}}t)\left[-\frac{B}{\tau} - A\omega_{\text{d}}\right]\right),$$

$$\ddot{V}_{\text{gap}} = e^{-t/\tau}\left(\cos(\omega_{\text{d}}t)\left[\frac{A}{\tau^2} - \frac{B\omega_{\text{d}}}{\tau} - \frac{B\omega_{\text{d}}}{\tau} - A\omega_{\text{d}}^2\right] + \right. \tag{A.35}$$

$$\left. + \ \sin(\omega_{\text{d}}t)\left[\frac{B}{\tau^2} + \frac{A\omega_{\text{d}}}{\tau} + \frac{A\omega_{\text{d}}}{\tau} - B\omega_{\text{d}}^2\right]\right).$$

We may easily verify that this ansatz satisfies the homogeneous differential equation if we insert it and compare the cosine coefficients:

$$\frac{A}{\tau^2} - \frac{2B\omega_\mathrm{d}}{\tau} - A\omega_\mathrm{d}^2 - \frac{2A}{\tau^2} + \frac{2B\omega_\mathrm{d}}{\tau} + \omega_\mathrm{res}^2 A = 0.$$

The sine coefficients lead to

$$\frac{B}{\tau^2} + \frac{2A\omega_\mathrm{d}}{\tau} - B\omega_\mathrm{d}^2 - \frac{2B}{\tau^2} - \frac{2A\omega_\mathrm{d}}{\tau} + \omega_\mathrm{res}^2 B = 0.$$

A.7.1.2 Particular Solution

The particular solution $V'_\mathrm{gap}(t)$ is the one that is obtained if the standard AC circuit theory with phasors is used. We obtain

$$\underline{\hat{V}}'_\mathrm{gap} = \underline{\hat{I}}_\mathrm{gen} \cdot Z$$

with

$$Z = 1/Y,$$

$$Y = \frac{1}{R_\mathrm{p}} + j\omega C + \frac{1}{j\omega L_\mathrm{p}},$$

and

$$|Y| = \frac{1}{|Z|} = \sqrt{\frac{1}{R_\mathrm{p}^2} + \left(\omega C - \frac{1}{\omega L_\mathrm{p}}\right)^2}.$$

For the generator current

$$I_\mathrm{gen}(t) = \hat{I}_\mathrm{gen}\ \sin(\omega t) = \hat{I}_\mathrm{gen}\ \cos\left(\omega t - \frac{\pi}{2}\right),$$

we get the phasor

$$\underline{\hat{I}}_\mathrm{gen} = -j\ \hat{I}_\mathrm{gen},$$

and therefore

$$V'_\mathrm{gap}(t) = \hat{V}'_\mathrm{gap}\ \cos\left(\omega t + \varphi_Z - \frac{\pi}{2}\right) = \hat{V}'_\mathrm{gap}\ \sin(\omega t + \varphi_Z)$$

with

$$\hat{V}'_\mathrm{gap} = \hat{I}_\mathrm{gen}|Z|$$

and

$$\varphi_Z = \measuredangle Z = -\measuredangle Y = -\arctan\frac{\omega C - \frac{1}{\omega L_\mathrm{p}}}{1/R_\mathrm{p}} = \arctan\left(\frac{R_\mathrm{p}}{\omega L_\mathrm{p}} - \omega R_\mathrm{p} C\right).$$

A.7.1.3 Overall Solution and Initial Conditions

The total solution for $t > 0$ is obtained as the sum of the homogeneous and the particular solution:

$$V_\mathrm{gap}(t) = e^{-t/\tau}\left(A\ \cos(\omega_\mathrm{d} t) + B\ \sin(\omega_\mathrm{d} t)\right) + \hat{V}'_\mathrm{gap}\ \sin(\omega t + \varphi_Z). \tag{A.36}$$

Now we have to take the initial conditions into account. First of all, it is clear that the voltage at the capacitor, i.e., V_gap, may not jump:

$$V_\mathrm{gap}(0-) = V_\mathrm{gap}(0+) = 0$$

$$\Rightarrow A + \hat{V}'_\mathrm{gap}\ \sin(\varphi_Z) = 0$$

$$\Rightarrow A = -\hat{V}'_\mathrm{gap}\ \sin(\varphi_Z).$$

Due to

$$\sin(\arctan(x)) = \frac{x}{\sqrt{1+x^2}},$$

the equation

$$\sin(\varphi_Z) = \frac{\frac{R_\mathrm{p}}{\omega L_\mathrm{p}} - \omega R_\mathrm{p} C}{\sqrt{1 + \left(\frac{R_\mathrm{p}}{\omega L_\mathrm{p}} - \omega R_\mathrm{p} C\right)^2}} = \frac{\frac{1}{\omega L_\mathrm{p}} - \omega C}{\sqrt{\frac{1}{R_\mathrm{p}^2} + \left(\frac{1}{\omega L_\mathrm{p}} - \omega C\right)^2}} = -|Z|\ \mathrm{Im}\{Y\}$$

is valid, and we obtain

$$A = \hat{I}_\mathrm{gen}|Z|^2\ \mathrm{Im}\{Y\}.$$

Also, the current inside the inductor cannot jump. Therefore,

$$I_L = I_\mathrm{gen} - I_C - I_R$$

must be continuous. The reason that I_R must be continuous is that we have already shown that V_gap is continuous. Also, I_gen is continuous at $t = 0$. In conclusion, I_C must be continuous, which implies the continuity of $\dot{V}_\mathrm{gap}$. In general, we have

$$\dot{V}_{\mathrm{gap}}(t) = e^{-t/\tau}\left(\cos(\omega_{\mathrm{d}}t)\left[-\frac{A}{\tau} + B\omega_{\mathrm{d}}\right] + \sin(\omega_{\mathrm{d}}t)\left[-\frac{B}{\tau} - A\omega_{\mathrm{d}}\right]\right)$$
$$+\hat{V}'_{\mathrm{gap}}\omega\ \cos(\omega t + \varphi_Z).$$

For $t = 0$, we must therefore require

$$\left[-\frac{A}{\tau} + B\omega_{\mathrm{d}}\right] + \hat{V}'_{\mathrm{gap}}\omega\ \cos(\varphi_Z) = 0$$

$$\Rightarrow B = \frac{1}{\omega_{\mathrm{d}}}\left(\frac{A}{\tau} - \hat{V}'_{\mathrm{gap}}\omega\ \cos(\varphi_Z)\right).$$

Due to

$$\cos(\arctan(x)) = \frac{1}{\sqrt{1+x^2}},$$

we obtain

$$\cos(\varphi_Z) = \frac{1}{\sqrt{1 + \left(\frac{R_{\mathrm{p}}}{\omega L_{\mathrm{p}}} - \omega R_{\mathrm{p}} C\right)^2}} = \frac{1}{R_{\mathrm{p}}\sqrt{\frac{1}{R_{\mathrm{p}}^2} + \left(\frac{1}{\omega L_{\mathrm{p}}} - \omega C\right)^2}} = \frac{|Z|}{R_{\mathrm{p}}},$$

leading to

$$B = \frac{1}{\omega_{\mathrm{d}}}\left(\frac{\hat{I}_{\mathrm{gen}}|Z|^2\mathrm{Im}\{Y\}}{\tau} - \frac{\hat{I}_{\mathrm{gen}}|Z|^2\omega}{R_{\mathrm{p}}}\right)$$

$$\Rightarrow B = \hat{I}_{\mathrm{gen}}|Z|^2\left(\frac{\mathrm{Im}\{Y\}}{\omega_{\mathrm{d}}\tau} - \frac{\omega}{\omega_{\mathrm{d}}R_{\mathrm{p}}}\right).$$

This completely determines the solution:

$$\frac{V_{\mathrm{gap}}(t)}{\hat{I}_{\mathrm{gen}}|Z|} = \left(\sin(\omega t + \varphi_Z) + e^{-t/\tau}\left[|Z|\ \mathrm{Im}\{Y\}\ \cos(\omega_{\mathrm{d}}t)\right.\right. \tag{A.37}$$
$$\left.\left. + \left(\frac{|Z|\mathrm{Im}\{Y\}}{\omega_{\mathrm{d}}\tau} - \frac{\omega}{\omega_{\mathrm{d}}}\frac{|Z|}{R_{\mathrm{p}}}\right)\sin(\omega_{\mathrm{d}}t)\right]\right).$$

Finally, we consider the special case that the cavity is excited with the resonant frequency according to

$$\omega = \omega_{\mathrm{res}}.$$

Then we obtain

$$\mathrm{Im}\{Y\} = 0, \qquad \varphi_Z = 0, \qquad |Z| = R_{\mathrm{p}}, \qquad A = 0, \qquad B = -\hat{I}_{\mathrm{gen}}|Z|\frac{\omega}{\omega_{\mathrm{d}}},$$

and the total solution

$$V_{\mathrm{gap}}(t) = \hat{I}_{\mathrm{gen}} R_{\mathrm{p}} \left(\sin(\omega t) - e^{-t/\tau} \frac{\omega}{\omega_{\mathrm{d}}} \sin(\omega_{\mathrm{d}} t) \right).$$

For sufficiently high Q factors, one gets $\omega_{\mathrm{d}} \approx \omega$, and our excitation

$$I_{\mathrm{gen}}(t) = \hat{I}_{\mathrm{gen}} \sin(\omega t)$$

for $t > 0$ leads to the response

$$V_{\mathrm{gap}}(t) = I_{\mathrm{gen}}(t) R_{\mathrm{p}} \left(1 - e^{-t/\tau}\right).$$

Hence, under these simplifications, an amplitude step of the generator current

$$\hat{I}_{\mathrm{gen}}(t) = \hat{I}_{\mathrm{gen},0}\, \Theta(t)$$

leads to a gap voltage amplitude response

$$\hat{V}_{\mathrm{gap}}(t) = \hat{I}_{\mathrm{gen},0}\, \Theta(t)\; R_{\mathrm{p}} \left(1 - e^{-t/\tau}\right),$$

so that the transfer function

$$\frac{\hat{V}_{\mathrm{gap}}(s)}{\hat{I}_{\mathrm{gen}}(s)} = R_{\mathrm{p}} \frac{\frac{1}{s} - \frac{1}{s+\frac{1}{\tau}}}{\frac{1}{s}} = R_{\mathrm{p}} \left(1 - \frac{s}{s + \frac{1}{\tau}}\right) = \frac{R_{\mathrm{p}}}{1 + s\tau}$$

is obtained in the Laplace domain; this is a PT_1 behavior for the amplitudes.

A.7.2 Phase Jumps

In this section, we analyze how the phase of the gap voltage reacts to a jump in the phase of the beam current $\Delta\varphi_{\mathrm{beam}}$. The resulting change in the phase of the gap voltage is denoted by $\Delta\varphi$. We will determine this quantity in this section.

We consider only small phase jumps, since that is what is relevant for typical stability considerations.

Let us assume that the cavity is in a steady-state condition for $t < 0$, so that the following signals are present:

$$V_{\text{gap}}(t) = \hat{V}_{\text{gap}} \ \sin(\omega t - \varphi_0), \tag{A.38}$$

$$I_{\text{beam}}(t) = \hat{I}_{\text{beam}} \ \cos(\omega t - \varphi_{\text{beam}} - \varphi_0), \tag{A.39}$$

$$I_{\text{gen}}(t) = \hat{I}_{\text{gen}} \ \cos(\omega t - \varphi_{\text{gen}} - \varphi_0).$$

Please note that according to the choice of the sine and cosine functions, for $\varphi_{\text{beam}} = 0$ the maximum of the beam current is located at the positive zero crossing of the gap voltage (stationary conditions for $Q > 0$ and $\eta_{\text{R}} < 0$).

At $t = 0$, the phase of the beam current will jump from φ_{beam} to $\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}}$. Since this jump will in principle be able to hit any phase of the listed signals, the phase φ_0 was introduced.

A.7.2.1 Particular Solution

For $t > 0$, we have

$$I_{\text{beam}}(t) = \hat{I}_{\text{beam}} \ \cos(\omega t - \varphi_{\text{beam}} - \Delta\varphi_{\text{beam}} - \varphi_0), \tag{A.40}$$

$$I_{\text{gen}}(t) = \hat{I}_{\text{gen}} \ \cos(\omega t - \varphi_{\text{gen}} - \varphi_0).$$

The generator current remains unchanged, but the beam current performs a phase jump according to our formulation of the problem. In the steady state for $t \to \infty$, we obtain the following particular solution:

$$V_{\text{gap}}(t) = \hat{V}'_{\text{gap}} \ \sin(\omega t - \Delta\varphi_{\text{gap}} - \varphi_0) = \hat{V}'_{\text{gap}} \ \cos\left(\omega t - \Delta\varphi_{\text{gap}} - \varphi_0 - \frac{\pi}{2}\right). \tag{A.41}$$

Of course, not only a different steady-state phase has to be assumed here, but also an amplitude $\hat{V}'_{\text{gap}}$ that differs from the original amplitude $\hat{V}_{\text{gap}}$.

If we convert these time signals into complex amplitudes, we get

$$\underline{\hat{I}}_{\text{beam}} = \hat{I}_{\text{beam}} \ e^{-j(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} + \varphi_0)},$$

$$\underline{\hat{I}}_{\text{gen}} = \hat{I}_{\text{gen}} \ e^{-j(\varphi_{\text{gen}} + \varphi_0)},$$

$$\underline{\hat{V}}_{\text{gap}} = \hat{V}'_{\text{gap}} \ e^{-j\left(\frac{\pi}{2} + \Delta\varphi_{\text{gap}} + \varphi_0\right)}.$$

The equivalent circuit leads to

$$\underline{\hat{I}}_{\text{gen}} - \underline{\hat{I}}_{\text{beam}} = \underline{\hat{V}}_{\text{gap}} \ |Y| \ e^{-j\varphi_Z}.$$

For the real part of this equation, one obtains

$$\hat{I}_{\text{gen}}\ \cos\left(\varphi_{\text{gen}}+\varphi_0\right)\ -\ \hat{I}_{\text{beam}}\ \cos\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)=$$
$$=\hat{V}'_{\text{gap}}\ |Y|\ \cos\left(-\frac{\pi}{2}-\Delta\varphi_{\text{gap}}-\varphi_0-\varphi_Z\right).$$

The imaginary part is

$$-\hat{I}_{\text{gen}}\ \sin\left(\varphi_{\text{gen}}+\varphi_0\right)\ +\ \hat{I}_{\text{beam}}\ \sin\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)=$$
$$=\hat{V}'_{\text{gap}}\ |Y|\ \sin\left(-\frac{\pi}{2}-\Delta\varphi_{\text{gap}}-\varphi_0-\varphi_Z\right).$$

The last two equations may be written in the following form:

$$-\hat{I}_{\text{gen}}\ \cos\left(\varphi_{\text{gen}}+\varphi_0\right)\ +\ \hat{I}_{\text{beam}}\ \cos\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)=$$
$$=\hat{V}'_{\text{gap}}|Y|\ \sin\left(\Delta\varphi_{\text{gap}}+\varphi_0+\varphi_Z\right), \tag{A.42}$$
$$\hat{I}_{\text{gen}}\ \sin\left(\varphi_{\text{gen}}+\varphi_0\right)\ -\ \hat{I}_{\text{beam}}\ \sin\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)=$$
$$=\hat{V}'_{\text{gap}}|Y|\ \cos\left(\Delta\varphi_{\text{gap}}+\varphi_0+\varphi_Z\right). \tag{A.43}$$

The quotient of these two results is

$$\tan\left(\Delta\varphi_{\text{gap}}+\varphi_0+\varphi_Z\right)=\frac{-\hat{I}_{\text{gen}}\ \cos\left(\varphi_{\text{gen}}+\varphi_0\right)+\hat{I}_{\text{beam}}\ \cos\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)}{\hat{I}_{\text{gen}}\ \sin\left(\varphi_{\text{gen}}+\varphi_0\right)-\hat{I}_{\text{beam}}\ \sin\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)}$$

$$\Rightarrow\quad\left[\hat{I}_{\text{gen}}\ \sin\left(\varphi_{\text{gen}}+\varphi_0\right)-\hat{I}_{\text{beam}}\ \sin\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)\right]$$
$$\cdot\ \sin\left(\Delta\varphi_{\text{gap}}+\varphi_0+\varphi_Z\right)=$$
$$=\left[-\hat{I}_{\text{gen}}\ \cos\left(\varphi_{\text{gen}}+\varphi_0\right)+\hat{I}_{\text{beam}}\ \cos\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)\right]$$
$$\cdot\ \cos\left(\Delta\varphi_{\text{gap}}+\varphi_0+\varphi_Z\right).$$

The derivative of this equation with respect to $\Delta\varphi_{\text{beam}}$ is[6]

$$-\hat{I}_{\text{beam}}\ \cos\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)\ \sin\left(\Delta\varphi_{\text{gap}}+\varphi_0+\varphi_Z\right)+$$
$$+\left[\hat{I}_{\text{gen}}\ \sin\left(\varphi_{\text{gen}}+\varphi_0\right)-\hat{I}_{\text{beam}}\ \sin\left(\varphi_{\text{beam}}+\Delta\varphi_{\text{beam}}+\varphi_0\right)\right]\cdot$$
$$\cdot\ \cos\left(\Delta\varphi_{\text{gap}}+\varphi_0+\varphi_Z\right)\ \frac{\mathrm{d}\Delta\varphi_{\text{gap}}}{\mathrm{d}\Delta\varphi_{\text{beam}}}=$$

[6]We assume that φ_Z remains unchanged, because the operating frequency and the cavity parameters R_{p}, L_{p}, and C will not be modified by the phase jump.

$$= -\hat{I}_{\text{beam}}\ \sin\left(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} + \varphi_0\right)\ \cos\left(\Delta\varphi_{\text{gap}} + \varphi_0 + \varphi_Z\right) -$$
$$- \left[-\hat{I}_{\text{gen}}\ \cos\left(\varphi_{\text{gen}} + \varphi_0\right) + \hat{I}_{\text{beam}}\ \cos\left(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} + \varphi_0\right)\right] \cdot$$
$$\cdot\ \sin\left(\Delta\varphi_{\text{gap}} + \varphi_0 + \varphi_Z\right)\ \frac{\mathrm{d}\Delta\varphi_{\text{gap}}}{\mathrm{d}\Delta\varphi_{\text{beam}}}.$$

It follows that

$$K := \left.\frac{\mathrm{d}\Delta\varphi_{\text{gap}}}{\mathrm{d}\Delta\varphi_{\text{beam}}}\right|_{\Delta\varphi_{\text{beam}}=0} =$$
$$= \frac{\hat{I}_{\text{beam}}\ \cos\left(\varphi_{\text{beam}} + \varphi_0\right)\ \sin\left(\varphi_0 + \varphi_Z\right) - \hat{I}_{\text{beam}}\ \sin\left(\varphi_{\text{beam}} + \varphi_0\right)\ \cos\left(\varphi_0 + \varphi_Z\right)}{\hat{I}_{\text{gen}}\ \sin\left(\varphi_{\text{gen}} + \varphi_0\right)\ \cos\left(\varphi_0 + \varphi_Z\right) - \hat{I}_{\text{beam}}\ \sin\left(\varphi_{\text{beam}} + \varphi_0\right)\ \cos\left(\varphi_0 + \varphi_Z\right) - \cdots} \cdots$$

$$\cdots \frac{}{\cdots - \hat{I}_{\text{gen}}\ \cos\left(\varphi_{\text{gen}} + \varphi_0\right)\ \sin\left(\varphi_0 + \varphi_Z\right) + \hat{I}_{\text{beam}}\ \cos\left(\varphi_{\text{beam}} + \varphi_0\right)\ \sin\left(\varphi_0 + \varphi_Z\right)}.$$

Here we used the fact that for $\Delta\varphi_{\text{beam}} = 0$, we also have $\Delta\varphi_{\text{gap}} = 0$ by definition. Now we see that the trigonometric sum and difference identities may be applied, and we finally obtain

$$K = \frac{\hat{I}_{\text{beam}}\ \sin\left(\varphi_Z - \varphi_{\text{beam}}\right)}{\hat{I}_{\text{gen}}\ \sin\left(\varphi_{\text{gen}} - \varphi_Z\right) + \hat{I}_{\text{beam}}\ \sin\left(\varphi_Z - \varphi_{\text{beam}}\right)}.$$

This factor K obviously has to be multiplied by $\Delta\varphi_{\text{beam}}$ to determine the phase deviation $\Delta\varphi_{\text{gap}}$ when an equilibrium is reached after a small phase jump $\Delta\varphi_{\text{beam}}$ has occurred:

$$\boxed{\Delta\varphi_{\text{gap}} = K\ \Delta\varphi_{\text{beam}}.} \tag{A.44}$$

The expression for K may be converted into a form that does not depend on φ_{gen}. For this purpose, we multiply Eq. (A.42) by $\sin(\varphi_Z + \varphi_0)$ and Eq. (A.43) by $\cos(\varphi_Z + \varphi_0)$. The sum of the resulting equations leads to

$$\hat{I}_{\text{gen}}\ \sin\left(\varphi_{\text{gen}} - \varphi_Z\right) + \hat{I}_{\text{beam}}\ \sin\left(\varphi_Z - \varphi_{\text{beam}}\right) = \hat{V}_{\text{gap}}\ |Y|$$

if $\Delta\varphi_{\text{beam}} = \Delta\varphi_{\text{gap}} = 0$, $\hat{V}'_{\text{gap}} = \hat{V}_{\text{gap}}$ is considered the "operating point." We therefore obtain

$$K = \frac{\hat{I}_{\text{beam}}\ \sin\left(\varphi_Z - \varphi_{\text{beam}}\right)}{\hat{V}_{\text{gap}}\ |Y|}.$$

Sometimes, the relative **beam loading factor**[7]

$$\boxed{\xi = \frac{\hat{I}_{\text{beam}} R_{\text{p}}}{\hat{V}_{\text{gap}}} = \frac{\hat{I}_{\text{beam}}}{\hat{V}_{\text{gap}} \, |Y| \, \cos\varphi_Z}} \tag{A.45}$$

is defined, so that we obtain

$$K = \frac{\hat{I}_{\text{beam}} \, \sin\left(\varphi_Z - \varphi_{\text{beam}}\right)}{\hat{I}_{\text{beam}}} \, \xi \, \cos\varphi_Z$$

$$\Rightarrow \boxed{K = \xi \, \cos\varphi_Z \, \sin\left(\varphi_Z - \varphi_{\text{beam}}\right).} \tag{A.46}$$

Now we determine the change in the amplitude at steady state that is caused by the phase jump. For this purpose, we add the squares of Eqs. (A.42) and (A.43):

$$\hat{I}_{\text{gen}}^2 + \hat{I}_{\text{beam}}^2 - 2\hat{I}_{\text{gen}}\hat{I}_{\text{beam}}\left[\cos\left(\varphi_{\text{gen}} + \varphi_0\right) \, \cos\left(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} + \varphi_0\right) + \right.$$
$$\left. + \, \sin\left(\varphi_{\text{gen}} + \varphi_0\right) \, \sin\left(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} + \varphi_0\right)\right] = \left(\hat{V}'_{\text{gap}}|Y|\right)^2$$

$$\Rightarrow \hat{V}'_{\text{gap}} = |Z|\sqrt{\hat{I}_{\text{gen}}^2 + \hat{I}_{\text{beam}}^2 - 2 \, \hat{I}_{\text{gen}}\hat{I}_{\text{beam}} \, \cos(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} - \varphi_{\text{gen}})}.$$

If no phase jump occurs, we have

$$\Rightarrow \hat{V}_{\text{gap}} = |Z|\sqrt{\hat{I}_{\text{gen}}^2 + \hat{I}_{\text{beam}}^2 - 2 \, \hat{I}_{\text{gen}}\hat{I}_{\text{beam}} \, \cos(\varphi_{\text{beam}} - \varphi_{\text{gen}})}. \tag{A.47}$$

By means of

$$\cos(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} - \varphi_{\text{gen}}) \approx \cos(\varphi_{\text{beam}} - \varphi_{\text{gen}}) - \sin(\varphi_{\text{beam}} - \varphi_{\text{gen}}) \, \Delta\varphi_{\text{beam}},$$

we obtain

$$\hat{V}'_{\text{gap}} \approx \hat{V}_{\text{gap}} \sqrt{1 + \frac{2 \, \hat{I}_{\text{gen}}\hat{I}_{\text{beam}} \, \sin(\varphi_{\text{beam}} - \varphi_{\text{gen}})}{\hat{I}_{\text{gen}}^2 + \hat{I}_{\text{beam}}^2 - 2 \, \hat{I}_{\text{gen}}\hat{I}_{\text{beam}} \, \cos(\varphi_{\text{beam}} - \varphi_{\text{gen}})} \Delta\varphi_{\text{beam}}}.$$

Finally, we obtain

---

[7] In the literature, this factor is usually denoted by Y, which we have avoided, since Y is the admittance here.

$$\boxed{\hat{V}'_{\text{gap}} \approx \hat{V}_{\text{gap}}(1 + J\ \Delta\varphi_{\text{beam}})} \tag{A.48}$$

with

$$J = \frac{\sin(\varphi_{\text{beam}} - \varphi_{\text{gen}})}{\frac{\hat{I}_{\text{gen}}}{\hat{I}_{\text{beam}}} + \frac{\hat{I}_{\text{beam}}}{\hat{I}_{\text{gen}}} - 2\ \cos(\varphi_{\text{beam}} - \varphi_{\text{gen}})}. \tag{A.49}$$

Not only K, but also J may be written in a form that is independent of φ_{gen}. For this purpose, we multiply Eq. (A.42) by $\sin(\varphi_{\text{beam}} + \varphi_0)$ and Eq. (A.43) by $\cos(\varphi_{\text{beam}} + \varphi_0)$. The sum of the resulting equations is ($\Delta\varphi_{\text{beam}} = \Delta\varphi_{\text{gap}} = 0$, $\hat{V}'_{\text{gap}} = \hat{V}_{\text{gap}}$)

$$\hat{I}_{\text{gen}}\ \sin(\varphi_{\text{gen}} - \varphi_{\text{beam}}) = \hat{V}_{\text{gap}}|Y|\ \cos(\varphi_{\text{beam}} - \varphi_Z). \tag{A.50}$$

Now we multiply Eq. (A.42) by $-\cos(\varphi_{\text{beam}}+\varphi_0)$ and Eq. (A.43) by $\sin(\varphi_{\text{beam}}+\varphi_0)$. The sum of the resulting equations is ($\Delta\varphi_{\text{beam}} = \Delta\varphi_{\text{gap}} = 0$, $\hat{V}'_{\text{gap}} = \hat{V}_{\text{gap}}$):

$$\hat{I}_{\text{gen}}\ \cos(\varphi_{\text{gen}} - \varphi_{\text{beam}}) - \hat{I}_{\text{beam}} = \hat{V}_{\text{gap}}|Y|\ \sin(\varphi_{\text{beam}} - \varphi_Z). \tag{A.51}$$

The two equations (A.50) and (A.51) are now inserted into Eq. (A.49):

$$J = \frac{-\hat{V}_{\text{gap}}|Y|\ \cos(\varphi_{\text{beam}} - \varphi_Z)}{\frac{\hat{I}^2_{\text{gen}}}{\hat{I}_{\text{beam}}} + \hat{I}_{\text{beam}} - 2\ \left(\hat{I}_{\text{beam}} + \hat{V}_{\text{gap}}|Y|\ \sin(\varphi_{\text{beam}} - \varphi_Z)\right)}.$$

Due to Eq. (A.45), we obtain

$$\frac{\hat{V}_{\text{gap}}|Y|}{\hat{I}_{\text{beam}}} = \frac{1}{\xi\ \cos\varphi_Z}, \tag{A.52}$$

and it follows that

$$J = \frac{-\ \cos(\varphi_{\text{beam}} - \varphi_Z)}{\xi\ \cos\varphi_Z \left[\frac{\hat{I}^2_{\text{gen}}}{\hat{I}^2_{\text{beam}}} - 1 - 2\frac{\sin(\varphi_{\text{beam}} - \varphi_Z)}{\xi\ \cos\varphi_Z}\right]}. \tag{A.53}$$

In order to substitute the fraction $(\hat{I}_{\text{gen}}/\hat{I}_{\text{beam}})^2$, we make use of Eq. (A.47):

$$\frac{\hat{V}_{\mathrm{gap}}|Y|}{\hat{I}_{\mathrm{beam}}} = \sqrt{\frac{\hat{I}_{\mathrm{gen}}^2}{\hat{I}_{\mathrm{beam}}^2} + 1 - 2\,\frac{\hat{I}_{\mathrm{gen}}}{\hat{I}_{\mathrm{beam}}}\,\cos(\varphi_{\mathrm{beam}} - \varphi_{\mathrm{gen}})}.$$

Now we insert Eq. (A.51):

$$\frac{\hat{V}_{\mathrm{gap}}|Y|}{\hat{I}_{\mathrm{beam}}} = \sqrt{\frac{\hat{I}_{\mathrm{gen}}^2}{\hat{I}_{\mathrm{beam}}^2} + 1 - 2\left[1 + \frac{\hat{V}_{\mathrm{gap}}|Y|}{\hat{I}_{\mathrm{beam}}}\,\sin(\varphi_{\mathrm{beam}} - \varphi_Z)\right]}.$$

Equation (A.52) may now be applied twice:

$$\frac{1}{\xi\,\cos\varphi_Z} = \sqrt{\frac{\hat{I}_{\mathrm{gen}}^2}{\hat{I}_{\mathrm{beam}}^2} - 1 - 2\frac{\sin(\varphi_{\mathrm{beam}} - \varphi_Z)}{\xi\,\cos\varphi_Z}}.$$

The argument of the square root is exactly the expression that is needed in Eq. (A.53):

$$\boxed{J = -\xi\,\cos\varphi_Z\,\cos(\varphi_{\mathrm{beam}} - \varphi_Z).} \tag{A.54}$$

A.7.2.2 Overall Solution and Initial Conditions

The homogeneous solution is the same one that we derived in Sect. A.7.1.1, since only the excitation differs in the two cases (amplitude jumps vs. phase jumps).

The total solution for $t > 0$ is the sum of the homogeneous solution (A.34) and the particular solution (A.41):

$$V_{\mathrm{gap}}(t) = e^{-t/\tau}\,(A\,\cos(\omega_{\mathrm{d}}t) + B\,\sin(\omega_{\mathrm{d}}t)) + \hat{V}'_{\mathrm{gap}}\,\sin(\omega t - \Delta\varphi_{\mathrm{gap}} - \varphi_0) \tag{A.55}$$

Now we have to incorporate the initial conditions. First of all, it is clear that the voltage at the capacitor V_{gap} cannot have any steps (the voltage for $t < 0$ is given by Eq. (A.38)):

$$V_{\mathrm{gap}}(0-) = V_{\mathrm{gap}}(0+)$$

$$\Rightarrow -\hat{V}_{\mathrm{gap}}\,\sin\varphi_0 = A - \hat{V}'_{\mathrm{gap}}\,\sin(\Delta\varphi_{\mathrm{gap}} + \varphi_0)$$

$$\Rightarrow A = \hat{V}'_{\mathrm{gap}}\,\sin(\Delta\varphi_{\mathrm{gap}} + \varphi_0) - \hat{V}_{\mathrm{gap}}\,\sin\varphi_0. \tag{A.56}$$

Also, the current I_L of the inductance cannot have any steps. Since the gap voltage and the generator current are continuous as well, Eq. (4.27) shows that $I_C + I_{\mathrm{beam}}$ must be continuous, too. Based on Eqs. (A.38) and (A.39), we obtain for $t < 0$,

$$I_C = C\ \dot{V}_{\text{gap}} = \omega C\ \hat{V}_{\text{gap}}\ \cos(\omega t - \varphi_0),$$
$$I_{\text{beam}} = \hat{I}_{\text{beam}}\ \cos(\omega t - \varphi_{\text{beam}} - \varphi_0).$$

Due to Eqs. (A.55) and (A.40), we have for $t > 0$,

$$I_C = C\ \dot{V}_{\text{gap}} = C\ e^{-t/\tau}\left(\cos(\omega_{\text{d}} t)\left[-\frac{A}{\tau} + B\omega_{\text{d}}\right] + \sin(\omega_{\text{d}} t)\left[-\frac{B}{\tau} - A\omega_{\text{d}}\right]\right) +$$
$$+\ \omega C\ \hat{V}'_{\text{gap}}\ \cos(\omega t - \Delta\varphi_{\text{gap}} - \varphi_0),$$
$$I_{\text{beam}} = \hat{I}_{\text{beam}}\ \cos(\omega t - \varphi_{\text{beam}} - \Delta\varphi_{\text{beam}} - \varphi_0).$$

The continuity requirement

$$I_C(0-) + I_{\text{beam}}(0-) = I_C(0+) + I_{\text{beam}}(0+)$$

leads to the following condition:

$$\omega C\ \hat{V}_{\text{gap}}\ \cos\varphi_0 + \hat{I}_{\text{beam}}\ \cos(\varphi_{\text{beam}} + \varphi_0) =$$
$$= C\left[-\frac{A}{\tau} + B\omega_{\text{d}}\right] + \omega C\,\hat{V}'_{\text{gap}}\cos(\Delta\varphi_{\text{gap}} + \varphi_0) + \hat{I}_{\text{beam}}\cos(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} + \varphi_0)$$

$$\Rightarrow B = \frac{1}{\omega_{\text{d}}}\left[\frac{A}{\tau} + \frac{\hat{I}_{\text{beam}}}{C}\left(\cos(\varphi_{\text{beam}} + \varphi_0) - \cos(\varphi_{\text{beam}} + \Delta\varphi_{\text{beam}} + \varphi_0)\right) +\right.$$
$$\left. + \omega\left(\hat{V}_{\text{gap}}\ \cos\varphi_0 - \hat{V}'_{\text{gap}}\ \cos(\Delta\varphi_{\text{gap}} + \varphi_0)\right)\right]. \tag{A.57}$$

Now we determine approximations for the expressions A and B that are valid for small $\Delta\varphi_{\text{beam}}$ and small $\Delta\varphi_{\text{gap}}$. Together with Eq. (A.48), Eq. (A.56) leads to

$$A \approx \hat{V}_{\text{gap}}(1 + J\ \Delta\varphi_{\text{beam}})\left(\Delta\varphi_{\text{gap}}\ \cos\varphi_0 + \sin\varphi_0\right) - \hat{V}_{\text{gap}}\ \sin\varphi_0$$
$$\Rightarrow A \approx \hat{V}_{\text{gap}}\left(\Delta\varphi_{\text{gap}}\ \cos\varphi_0 + J\ \Delta\varphi_{\text{beam}}\ \sin\varphi_0\right)$$
$$\Rightarrow \boxed{A \approx \hat{V}_{\text{gap}}\ \Delta\varphi_{\text{beam}}\left(K\ \cos\varphi_0 + J\ \sin\varphi_0\right).} \tag{A.58}$$

In the last step, we used Eq. (A.44).

In an analogous way, Eq. (A.57) leads to

$$B \approx \frac{1}{\omega_\mathrm{d}}\left[\frac{A}{\tau}+\frac{\hat{I}_\mathrm{beam}}{C}\left(\cos(\varphi_\mathrm{beam}+\varphi_0)-\cos(\varphi_\mathrm{beam}+\varphi_0)+\Delta\varphi_\mathrm{beam}\ \sin(\varphi_\mathrm{beam}+\varphi_0)\right)+\right.$$
$$\left.+\ \omega\left(\hat{V}_\mathrm{gap}\ \cos\varphi_0-(1+J\Delta\varphi_\mathrm{beam})\hat{V}_\mathrm{gap}\left[\cos\varphi_0-\Delta\varphi_\mathrm{gap}\ \sin\varphi_0\right]\right)\right]$$

$$B \approx \hat{V}_\mathrm{gap}\ \Delta\varphi_\mathrm{beam}\left(\frac{K\ \cos\ \varphi_0+J\ \sin\ \varphi_0}{\tau\omega_\mathrm{d}}+\frac{\hat{I}_\mathrm{beam}\ \sin(\varphi_\mathrm{beam}+\varphi_0)}{\hat{V}_\mathrm{gap}\,\omega_\mathrm{d}C}\right.$$
$$\left.+\frac{\omega}{\omega_\mathrm{d}}(K\ \sin\ \varphi_0-J\ \cos\ \varphi_0)\right).$$

In the second term, we may expand the fraction by R_p, so that the definition (A.45) for ξ may be applied:

$$B \approx \hat{V}_\mathrm{gap}\ \Delta\varphi_\mathrm{beam}\left(\frac{K\ \cos\varphi_0+J\ \sin\varphi_0}{\tau\omega_\mathrm{d}}+\frac{2\xi\ \sin(\varphi_\mathrm{beam}+\varphi_0)}{\tau\omega_\mathrm{d}}\right.$$
$$\left.+\frac{\omega}{\omega_\mathrm{d}}(K\ \sin\varphi_0-J\ \cos\varphi_0)\right).$$

We finally use Eq. (A.46):

$$\boxed{B \approx \hat{V}_\mathrm{gap}\ \Delta\varphi_\mathrm{beam}\left(\frac{K\ \cos\varphi_0+J\ \sin\varphi_0}{\tau\omega_\mathrm{d}}+\frac{2K\ \sin(\varphi_\mathrm{beam}+\varphi_0)}{\tau\omega_\mathrm{d}\ \sin(\varphi_Z-\varphi_\mathrm{beam})\ \cos\varphi_Z}+\right.}$$
$$\boxed{\left.+\ \frac{\omega}{\omega_\mathrm{d}}(K\ \sin\varphi_0-J\ \cos\varphi_0)\right).} \tag{A.59}$$

A.7.2.3 Phase Jump of the Gap Voltage

In order to determine the transient behavior $\Delta\varphi(t)$ of the phase jump of the gap voltage, we now write $V_\mathrm{gap}(t)$ for $t > 0$ in the form

$$V_\mathrm{gap}(t)=\hat{V}''_\mathrm{gap}\ \sin(\omega t-\Delta\varphi-\varphi_0)=\hat{V}''_\mathrm{gap}\ \sin(\omega t)\ \cos(\Delta\varphi+\varphi_0)$$
$$-\hat{V}''_\mathrm{gap}\ \cos(\omega t)\ \sin(\Delta\varphi+\varphi_0).$$

We now define

$$\boxed{\omega_\mathrm{d} = \omega + \Delta\omega_\mathrm{d},}$$

and a comparison of the $\sin(\omega t)$ and $\cos(\omega t)$ terms with Eq. (A.55) yields

$$\hat{V}''_\mathrm{gap}\ \cos(\Delta\varphi + \varphi_0) = e^{-t/\tau}\ (-A\ \sin(\Delta\omega_\mathrm{d} t) + B\ \cos(\Delta\omega_\mathrm{d} t)) + \hat{V}'_\mathrm{gap}\ \cos(\Delta\varphi_\mathrm{gap} + \varphi_0)$$

$$-\hat{V}''_\mathrm{gap}\ \sin(\Delta\varphi + \varphi_0) = e^{-t/\tau}\ (A\ \cos(\Delta\omega_\mathrm{d} t) + B\ \sin(\Delta\omega_\mathrm{d} t)) - \hat{V}'_\mathrm{gap}\ \sin(\Delta\varphi_\mathrm{gap} + \varphi_0).$$

The negative quotient of the right-hand side of these equations will be denoted by f:

$$f = \frac{e^{-t/\tau}\ (A\ \sin(\Delta\omega_\mathrm{d} t) - B\ \cos(\Delta\omega_\mathrm{d} t)) - \hat{V}'_\mathrm{gap}\ \cos(\Delta\varphi_\mathrm{gap} + \varphi_0)}{e^{-t/\tau}\ (A\ \cos(\Delta\omega_\mathrm{d} t) + B\ \sin(\Delta\omega_\mathrm{d} t)) - \hat{V}'_\mathrm{gap}\ \sin(\Delta\varphi_\mathrm{gap} + \varphi_0)}.$$

An approximation of the quotient of the left-hand sides for $\Delta\varphi \ll 1$ leads to

$$\cos\varphi_0 - \Delta\varphi\ \sin\varphi_0 \approx f(\sin\varphi_0 + \Delta\varphi\ \cos\varphi_0)$$

$$\Rightarrow \Delta\varphi \approx \frac{\cos\varphi_0 - f\ \sin\varphi_0}{\sin\varphi_0 + f\ \cos\varphi_0}.$$

Here we insert the expression for f determined above:

$$\begin{aligned}\Delta\varphi \approx\ & \frac{e^{-t/\tau}\,[A\ \cos(\Delta\omega_\mathrm{d} t)\ \cos\varphi_0 + B\ \sin(\Delta\omega_\mathrm{d} t)\ \cos\varphi_0 -}{e^{-t/\tau}\,[A\ \cos(\Delta\omega_\mathrm{d} t)\ \sin\varphi_0 + B\ \sin(\Delta\omega_\mathrm{d} t)\ \sin\varphi_0 +} \cdots \\ \cdots\ & \frac{-A\ \sin(\Delta\omega_\mathrm{d} t)\ \sin\varphi_0 + B\ \cos(\Delta\omega_\mathrm{d} t)\ \sin\varphi_0] +}{+A\ \sin(\Delta\omega_\mathrm{d} t)\ \cos\varphi_0 - B\ \cos(\Delta\omega_\mathrm{d} t)\ \cos\varphi_0] +} \cdots \\ \cdots\ & \frac{+\hat{V}'_\mathrm{gap}\left[-\sin(\Delta\varphi_\mathrm{gap} + \varphi_0)\ \cos\varphi_0 + \cos(\Delta\varphi_\mathrm{gap} + \varphi_0)\ \sin\varphi_0\right]}{+\hat{V}'_\mathrm{gap}\left[-\sin(\Delta\varphi_\mathrm{gap} + \varphi_0)\ \sin\varphi_0 - \cos(\Delta\varphi_\mathrm{gap} + \varphi_0)\ \cos\varphi_0\right]}.\end{aligned}$$

The trigonometric addition and subtraction identities lead to a significant simplification:

$$\boxed{\Delta\varphi \approx \frac{e^{-t/\tau}\,[A\ \cos(\Delta\omega_\mathrm{d} t + \varphi_0) + B\ \sin(\Delta\omega_\mathrm{d} t + \varphi_0)] - \hat{V}'_\mathrm{gap}\sin\Delta\varphi_\mathrm{gap}}{e^{-t/\tau}\,[A\ \sin(\Delta\omega_\mathrm{d} t + \varphi_0) - B\ \cos(\Delta\omega_\mathrm{d} t + \varphi_0)] - \hat{V}'_\mathrm{gap}\cos\Delta\varphi_\mathrm{gap}}.}$$

For sufficiently large t, the exponential functions disappear, so that

$$\Delta\varphi \to \Delta\varphi_\mathrm{gap}$$

holds for $\Delta\varphi_{\mathrm{gap}} \ll 1$. This is a cross-check, since the phase jump of the overall solution must approach the phase jump of the particular solution asymptotically for large times t.

Again, we make some approximations for $\Delta\varphi_{\mathrm{gap}} \ll 1$:

$$\Delta\varphi \approx \frac{-e^{-t/\tau}\left[\frac{A}{\hat{V}'_{\mathrm{gap}}}\cos(\Delta\omega_{\mathrm{d}}t+\varphi_0)+\frac{B}{\hat{V}'_{\mathrm{gap}}}\sin(\Delta\omega_{\mathrm{d}}t+\varphi_0)\right]+\Delta\varphi_{\mathrm{gap}}}{-e^{-t/\tau}\left[\frac{A}{\hat{V}'_{\mathrm{gap}}}\sin(\Delta\omega_{\mathrm{d}}t+\varphi_0)-\frac{B}{\hat{V}'_{\mathrm{gap}}}\cos(\Delta\omega_{\mathrm{d}}t+\varphi_0)\right]+1}.$$

If we now take into account that both $A/\hat{V}'_{\mathrm{gap}}$ and $B/\hat{V}'_{\mathrm{gap}}$ are proportional to $\Delta\varphi_{\mathrm{beam}}$, i.e., very small, we may neglect the denominator:

$$\Delta\varphi \approx \Delta\varphi_{\mathrm{gap}} - e^{-t/\tau}\left[\frac{A}{\hat{V}'_{\mathrm{gap}}}\cos(\Delta\omega_{\mathrm{d}}t+\varphi_0)+\frac{B}{\hat{V}'_{\mathrm{gap}}}\sin(\Delta\omega_{\mathrm{d}}t+\varphi_0)\right].$$

Inserting the approximation (A.48),

$$\hat{V}'_{\mathrm{gap}} \approx \hat{V}_{\mathrm{gap}}(1+J\Delta\varphi_{\mathrm{beam}}),$$

would lead to only second-order terms, which can be neglected:

$$\boxed{\Delta\varphi \approx K\ \Delta\varphi_{\mathrm{beam}} - e^{-t/\tau}\left[\frac{A}{\hat{V}_{\mathrm{gap}}}\cos(\Delta\omega_{\mathrm{d}}t+\varphi_0)+\frac{B}{\hat{V}_{\mathrm{gap}}}\sin(\Delta\omega_{\mathrm{d}}t+\varphi_0)\right].} \tag{A.60}$$

Equation (A.60) determines the time response of the phase of the gap voltage that results from a jump in the beam current phase. It has to be emphasized that the calculations presented here have to be regarded as exact for sufficiently small phase jumps.

Finally, we show that the phase response corresponds to a PT_1 behavior under certain conditions. For this purpose, we assume that the cavity is excited at resonance and that the quality factor is high:

$$\omega = \omega_{\mathrm{res}}, \quad \tau\omega_{\mathrm{d}} \gg 1, \quad \tau\omega_{\mathrm{d}} \gg \xi, \quad \omega_{\mathrm{d}} \approx \omega, \quad \Delta\omega_{\mathrm{d}} \approx 0. \tag{A.61}$$

Equations (A.46),

$$K = \xi \ \cos\varphi_Z \ \sin\left(\varphi_Z - \varphi_{\text{beam}}\right),$$

and (A.54),

$$J = -\xi \ \cos\varphi_Z \ \cos\left(\varphi_Z - \varphi_{\text{beam}}\right),$$

lead to the following expressions:

$$\begin{aligned} K \ \cos\varphi_0 + J \ \sin\varphi_0 &= \xi \ \cos\varphi_Z \ \sin\left(\varphi_Z - \varphi_{\text{beam}} - \varphi_0\right), \\ K \ \sin\varphi_0 - J \ \cos\varphi_0 &= \xi \ \cos\varphi_Z \ \cos\left(\varphi_Z - \varphi_{\text{beam}} - \varphi_0\right). \end{aligned}$$

According to Eq. (A.58), we get

$$A \approx \hat{V}_{\text{gap}} \ \Delta\varphi_{\text{beam}} \ \xi \ \cos\varphi_Z \ \sin\left(\varphi_Z - \varphi_{\text{beam}} - \varphi_0\right).$$

Using the approximations (A.61), Eq. (A.59) leads to

$$B \approx \hat{V}_{\text{gap}} \ \Delta\varphi_{\text{beam}} \ \xi \ \cos\varphi_Z \ \cos\left(\varphi_Z - \varphi_{\text{beam}} - \varphi_0\right).$$

These results allow us to determine the expression in brackets in Eq. (A.60):

$$\begin{aligned} &\left[\frac{A}{\hat{V}_{\text{gap}}} \ \cos(\Delta\omega_{\text{d}} t + \varphi_0) + \frac{B}{\hat{V}_{\text{gap}}} \ \sin(\Delta\omega_{\text{d}} t + \varphi_0)\right] \\ &\quad \approx \Delta\varphi_{\text{beam}} \ \xi \ \cos\varphi_Z \ \sin\left(\varphi_Z - \varphi_{\text{beam}}\right) = K \ \Delta\varphi_{\text{beam}}. \end{aligned}$$

Hence, Eq. (A.60) takes the simplified form

$$\Delta\varphi \approx K \ \Delta\varphi_{\text{beam}} \left(1 - e^{-t/\tau}\right),$$

which actually is a PT_1 response.

A.8 Example for Adiabaticity

In this section, we construct an example of an exact solution of the ODE (5.8) in Sect. 5.2.2. For this purpose, we just define the solution

$$\Omega_0(t) = \hat{\Omega}_0 \ \cos^2(kt), \tag{A.62}$$

where $\hat{\Omega}_0$ and k are constants. We obtain[8]

$$\dot{\Omega}_0 = -2\hat{\Omega}_0 k \ \sin(kt) \ \cos(kt), \tag{A.63}$$

$$\ddot{\Omega}_0 = -2\hat{\Omega}_0 k^2 \left[\cos^2(kt) - \sin^2(kt)\right] = -2\hat{\Omega}_0 k^2 \left[2 \ \cos^2(kt) - 1\right], \tag{A.64}$$

$$\dddot{\Omega}_0 = -2\hat{\Omega}_0 k^2 \ 4k \ \cos(kt)(-\sin(kt)) = 8\hat{\Omega}_0 k^3 \ \sin(kt) \ \cos(kt) = -4k^2\dot{\Omega}_0. \tag{A.65}$$

Using Eq. (5.9), we obtain

$$\Omega^2(t) = \hat{\Omega}_0^2 \ \cos^4(kt) - k^2[1 - \tan^2(kt)] - 3k^2 \tan^2(kt)$$

$$\Rightarrow \Omega^2(t) = \hat{\Omega}_0^2 \ \cos^4(kt) - k^2[1 + 2 \ \tan^2(kt)] \tag{A.66}$$

as the "original" frequency in the ODE. Solutions of the ODE

$$\ddot{u} + \Omega^2(t)u = 0$$

with this choice of Ω are therefore (cf. Sect. 5.2.2)

$$u(t) = \frac{M}{\sqrt{\hat{\Omega}_0}} \ \cos^{-1}(kt) \ \cos\left(\int \Omega_0(t) \ \mathrm{d}t\right)$$

and

$$u(t) = \frac{M}{\sqrt{\hat{\Omega}_0}} \ \cos^{-1}(kt) \ \sin\left(\int \Omega_0(t) \ \mathrm{d}t\right).$$

Here

$$\varphi(t) = \int \Omega_0(t) \ \mathrm{d}t = \hat{\Omega}_0 \int \cos^2(kt) \ \mathrm{d}t = \frac{\hat{\Omega}_0}{2} \int (1 + \cos(2kt)) \ \mathrm{d}t = \frac{\hat{\Omega}_0}{2}\left(t + \frac{1}{2k}\sin(2kt)\right)$$

$$\Rightarrow \varphi(t) = \frac{\hat{\Omega}_0}{2} t \ (1 + \mathrm{si}(2kt))$$

holds, where the integration constant is omitted, since we require

[8] The higher-order derivatives are needed later.

$$\varphi(0) = 0.$$

For $kt \ll 1$, we have $\varphi \approx \hat{\Omega}_0 t$ as a zero-order approximation. This is not astonishing, since the time derivative of $\Omega_0(t)$ vanishes for $t = 0$ and $\Omega_0 \approx \hat{\Omega}_0$ is quasiconstant.

One has to make sure that t is never so large that $\Omega^2 < 0$ becomes true. In any case, one has to choose $kt < \pi/2$ in order to keep the tan function in Eq. (A.66) finite. Furthermore, $0 < k < \hat{\Omega}_0$ must hold in order to have $\Omega^2(0) > 0$.

Now we continue with an application of the test scenario. For this purpose, Eq. (3.15),

$$\frac{\mathrm{d}\Delta t}{\mathrm{d}t} = \frac{\eta_\mathrm{R}}{\beta_\mathrm{R}^2 \gamma_\mathrm{R}} \Delta\gamma,$$

is written in the form

$$\Delta\dot{t} = -a \ \Delta\gamma,$$

where

$$a = \frac{-\eta_\mathrm{R}}{\beta_\mathrm{R}^2 \gamma_\mathrm{R}} \tag{A.67}$$

is defined. Here we assume that the energy of the synchronous particle remains constant, so that a does not depend on time.

Equation (3.17),

$$\frac{\mathrm{d}\Delta\gamma}{\mathrm{d}t} = \frac{Q\hat{V}}{T_\mathrm{R} m_0 c_0^2} \left(\sin(\omega_\mathrm{RF}\Delta t + \varphi_\mathrm{R}) - \sin\varphi_\mathrm{R}\right),$$

is linearized, which leads to

$$\Delta\dot{\gamma} = b \ \Delta t. \tag{A.68}$$

The voltage amplitude $\hat{V}$ is assumed to be time-dependent, and therefore b also depends on time. Since a is constant, the ODE for Δt is simple:

$$\Delta\ddot{t} + ab \ \Delta t = 0.$$

Due to the time dependence of

$$\Omega^2(t) := a \ b(t),$$

the ODE

$$\Delta\ddot{t} + \Omega^2(t)\ \Delta t = 0$$

for Δt that was discussed above for $u(t)$ is obtained. For the choice of Ω and Ω_0 that served as an example above, the solution

$$\Delta t = \Delta t_0\ \cos^{-1}(kt)\ \cos\left(\int \Omega_0(t)\ \mathrm{d}t\right) \tag{A.69}$$

is obtained, where we have used the abbreviation

$$\Delta t_0 = \Delta t|_{t=0} = \frac{M}{\sqrt{\hat{\Omega}_0}}.$$

It follows that

$$\begin{aligned}\Delta\dot{t} = \Delta t_0 \Bigg[& k\ \cos^{-2}(kt)\ \sin(kt)\ \cos\left(\int \Omega_0(t)\ \mathrm{d}t\right) \\ & -\Omega_0\ \cos^{-1}(kt)\ \sin\left(\int \Omega_0(t)\ \mathrm{d}t\right)\Bigg] = \\ = \Delta t_0 \Bigg[& k\ \cos^{-2}(kt)\ \sin(kt)\ \cos\left(\int \Omega_0(t)\ \mathrm{d}t\right) \\ & -\hat{\Omega}_0\ \cos(kt)\ \sin\left(\int \Omega_0(t)\ \mathrm{d}t\right)\Bigg],\end{aligned}$$

so that

$$\begin{aligned}\Delta\gamma = -\frac{\Delta\dot{t}}{a} = \frac{\Delta t_0}{a}\Bigg[& -k\ \cos^{-2}(kt)\ \sin(kt)\ \cos\left(\int \Omega_0(t)\ \mathrm{d}t\right) \\ & +\hat{\Omega}_0\ \cos(kt)\ \sin\left(\int \Omega_0(t)\ \mathrm{d}t\right)\Bigg]\end{aligned} \tag{A.70}$$

is the result.

For verification purposes, we now check the validity of the ODE for $\Delta\gamma$ (the reader may skip this calculation, which is printed only for the sake of completeness). Equation (A.68) leads to

$$\Delta\ddot{\gamma} = \dot{b}\Delta t + b\Delta\dot{t} = \frac{\dot{b}}{b}\Delta\dot{\gamma} - ab\Delta\gamma$$

$$\Rightarrow \Delta\ddot{\gamma} - \frac{\dot{b}}{b}\Delta\dot{\gamma} + ab\Delta\gamma = 0.$$

Due to

$$\Omega^2 = ab \qquad \Rightarrow 2\Omega\dot{\Omega} = a\dot{b} \qquad \Rightarrow 2\frac{\dot{\Omega}}{\Omega} = \frac{\dot{b}}{b},$$

the differential equation is

$$\Delta\ddot{\gamma} - 2\frac{\dot{\Omega}}{\Omega}\,\Delta\dot{\gamma} + \Omega^2\Delta\gamma = 0,$$

or

$$\Omega^2\Delta\ddot{\gamma} - 2\dot{\Omega}\;\Omega\;\Delta\dot{\gamma} + \Omega^4\Delta\gamma = 0. \tag{A.71}$$

In our example, we have (cf. Eq. (A.66))

$$\Omega^2(t) = \hat{\Omega}_0^2\;\cos^4(kt) - k^2[1 + 2\;\tan^2(kt)]$$

$$\Rightarrow 2\Omega\dot{\Omega} = -4k\hat{\Omega}_0^2\;\cos^3(kt)\;\sin(kt) - 4k^3\;\cos^{-3}(kt)\;\sin(kt).$$

According to Eq. (A.70), the solution of the ODE is

$$\begin{aligned}
\Delta\gamma &= \frac{\Delta t_0}{a}\left[-k\;\cos^{-2}(kt)\;\sin(kt)\;\cos\varphi + \hat{\Omega}_0\;\cos(kt)\;\sin\varphi\right],\\
\Delta\dot{\gamma} &= \frac{\Delta t_0}{a}\left[\left(-2k^2\;\cos^{-3}(kt)\;\sin^2(kt) - k^2\;\cos^{-1}(kt) + \hat{\Omega}_0\;\Omega_0\;\cos(kt)\right)\cos\varphi +\right.\\
&\quad\left. + \left(-\hat{\Omega}_0\,k\;\sin(kt) + k\;\Omega_0\;\cos^{-2}(kt)\;\sin(kt)\right)\sin\varphi\right] =\\
&= \frac{\Delta t_0}{a}\left(-2k^2\;\cos^{-3}(kt) + k^2\;\cos^{-1}(kt) + \hat{\Omega}_0^2\;\cos^3(kt)\right)\cos\varphi,\\
\Delta\ddot{\gamma} &= \frac{\Delta t_0}{a}\left[\left(-6k^3\;\cos^{-4}(kt) + k^3\;\cos^{-2}(kt) - 3k\hat{\Omega}_0^2\;\cos^2(kt)\right)\sin(kt)\;\cos\varphi -\right.\\
&\quad\left. - \Omega_0\left(-2k^2\;\cos^{-3}(kt) + k^2\;\cos^{-1}(kt) + \hat{\Omega}_0^2\;\cos^3(kt)\right)\sin\varphi\right] =\\
&= \frac{\Delta t_0}{a}\left[\left(-6k^3\;\cos^{-4}(kt) + k^3\;\cos^{-2}(kt) - 3k\hat{\Omega}_0^2\;\cos^2(kt)\right)\sin(kt)\;\cos\varphi +\right.\\
&\quad\left. + \left(2k^2\hat{\Omega}_0\;\cos^{-1}(kt) - k^2\hat{\Omega}_0\;\cos(kt) - \hat{\Omega}_0^3\;\cos^5(kt)\right)\sin\varphi\right].
\end{aligned}$$

All these formulas are now inserted into the ODE (A.71). The abbreviations

$$c = \cos(kt), \qquad s = \sin(kt)$$

are helpful. First we write down the coefficients of $\cos\varphi$:

$$\begin{aligned}
&\left(\hat{\Omega}_0^2\, c^4 - k^2[1 + 2s^2c^{-2}]\right)\left(-6k^3\, c^{-4} + k^3\, c^{-2} - 3k\hat{\Omega}_0^2\, c^2\right) s + \\
&+ \left(4k\hat{\Omega}_0^2\, c^3\, s + 4k^3\, c^{-3}\, s\right)\left(-2k^2\, c^{-3} + k^2\, c^{-1} + \hat{\Omega}_0^2\, c^3\right) + \\
&+ \left(\hat{\Omega}_0^2\, c^4 - k^2[1 + 2s^2c^{-2}]\right)^2\left(-k\, c^{-2}\, s\right).
\end{aligned}$$

Here we analyze the coefficients of the powers of $\hat{\Omega}_0$:

$$\begin{aligned}
\hat{\Omega}_0^4 \;:\;& c^4(-3k\, c^2)s + 4k\, c^3\, s\, c^3 + c^8\left(-k\, c^{-2}\, s\right) = 0, \\
\hat{\Omega}_0^2 \;:\;& c^4\,(-6k^3\, c^{-4} + k^3\, c^{-2})s - 3k\, c^2 s(-k^2[1 + 2s^2c^{-2}]) + \\
&+ 4k\, c^3\, s(-2k^2\, c^{-3} + k^2\, c^{-1}) + c^3\, 4k^3\, c^{-3}\, s + \\
&+ k\, c^{-2}\, s\, 2c^4k^2[1 + 2s^2c^{-2}] = \\
&= -6k^3s + k^3c^2s + 3k^3c^2s + 6k^3s^3 - \\
&- 8k^3s + 4k^3c^2s + 4k^3s + \\
&+ 2k^3c^2s + 4k^3s^3 = \\
&= -10k^3s + 10k^3c^2s + 10k^3s^3 = \\
&= -10k^3s + 10k^3c^2s + 10k^3s - 10k^3sc^2 = 0, \\
\hat{\Omega}_0^0 \;:\;& -k^2[1 + 2s^2c^{-2}]\left(-6k^3\, c^{-4} + k^3\, c^{-2}\right) s + \\
&+ 4k^3\, c^{-3}\, s\left(-2k^2\, c^{-3} + k^2\, c^{-1}\right) + \\
&+ k^4[1 + 2s^2c^{-2}]^2\left(-k\, c^{-2}\, s\right) = \\
&= 6k^5\, c^{-4}\, s - k^5\, c^{-2}\, s + 12k^5\, c^{-6}\, s^3 - 2k^5\, c^{-4}\, s^3 - \\
&- 8k^5c^{-6}s + 4k^5c^{-4}s - \\
&- k^5c^{-2}\, s - 4k^5c^{-4}s^3 - 4k^5c^{-6}s^5 = \\
&= 6k^5\, c^{-4}\, s - k^5\, c^{-2}\, s + 12k^5\, c^{-6}\, s - 12k^5\, c^{-4}\, s - 2k^5\, c^{-4}\, s + 2k^5\, c^{-2}\, s - \\
&- 8k^5c^{-6}s + 4k^5c^{-4}s -
\end{aligned}$$

$$
\begin{aligned}
&- k^5 c^{-2}\, s - 4k^5 c^{-4} s + 4k^5 c^{-2} s - 4k^5 c^{-6} s^3 + 4k^5 c^{-4} s^3 = \\
&= -4k^5\, c^{-4}\, s + k^5\, c^{-2}\, s + 4k^5\, c^{-6}\, s - \\
&- k^5 c^{-2}\, s - 4k^5 c^{-4} s + 4k^5 c^{-2} s - 4k^5 c^{-6} s + 4k^5 c^{-4} s + 4k^5 c^{-4} s - \\
&- 4k^5 c^{-2} s = 0.
\end{aligned}
$$

We have shown that the coefficients of $\cos\varphi$ are equal to zero. Now we write down the coefficients of $\sin\varphi$ in Eq. (A.71):

$$
\begin{aligned}
&\left(\hat{\Omega}_0^2\, c^4 - k^2[1 + 2s^2 c^{-2}]\right)\left(2k^2 \hat{\Omega}_0\, c^{-1} - k^2 \hat{\Omega}_0\, c - \hat{\Omega}_0^3\, c^5\right) + \\
&+ \left(\hat{\Omega}_0^2\, c^4 - k^2[1 + 2s^2 c^{-2}]\right)^2 \hat{\Omega}_0\, c.
\end{aligned}
$$

The coefficients of the powers of $\hat{\Omega}_0$ are

$$
\begin{aligned}
&\hat{\Omega}_0^5 \;:\; -c^9 + c^9 = 0, \\
&\hat{\Omega}_0^3 \;:\; c^4(2k^2\, c^{-1} - k^2\, c) + c^5 k^2[1 + 2s^2 c^{-2}] - 2c^4 k^2[1 + 2s^2 c^{-2}]\, c = \\
&\quad = 2k^2 c^3 - k^2 c^5 + k^2 c^5 + 2k^2 c^3 s^2 - 2k^2 c^5 - 4k^2 c^3 s^2 = \\
&\quad = 2k^2 c^3 - k^2 c^5 + k^2 c^5 + 2k^2 c^3 - 2k^2 c^5 - 2k^2 c^5 - 4k^2 c^3 + 4k^2 c^5 = 0, \\
&\hat{\Omega}_0^1 \;:\; -k^2[1 + 2s^2 c^{-2}]\left(2k^2\, c^{-1} - k^2\, c\right) + k^4[1 + 2s^2 c^{-2}]^2\, c = \\
&\quad = -k^2[-1 + 2c^{-2}]\left(2k^2\, c^{-1} - k^2\, c\right) + k^4[-1 + 2c^{-2}]^2\, c = \\
&\quad = \left(2k^4\, c^{-1} - k^4\, c\right) + \left(-4k^4\, c^{-3} + 2k^4\, c^{-1}\right) + k^4 c - 4k^4 c^{-1} + 4k^4 c^{-3} = 0.
\end{aligned}
$$

In conclusion, we have shown that the coefficients of $\sin\varphi$ are also equal to zero—the example solution indeed satisfies the ODE (A.71) without any approximations.

A.9 Tables and Diagrams

Table A.1 Fundamental constants

Speed of light in vacuum	$c_0 = 2.99792458 \cdot 10^8 \, \frac{\mathrm{m}}{\mathrm{s}}$
Permittivity of vacuum	$\epsilon_0 = 8.8541878176 \cdot 10^{-12} \, \frac{\mathrm{As}}{\mathrm{Vm}}$
Permeability of vacuum	$\mu_0 = 4\pi \cdot 10^{-7} \, \frac{\mathrm{Vs}}{\mathrm{Am}}$
Impedance of free space	$Z_0 = 376.73031346\,\Omega$
Elementary charge	$e = 1.60218 \cdot 10^{-19}\,\mathrm{C}$
Rest mass of the electron	$m_\mathrm{e} = 9.1094 \cdot 10^{-31}\,\mathrm{kg}\ (= 510.999\,\mathrm{keV}/c_0^2)$
Rest mass of the protons	$m_\mathrm{p} = 1.67262 \cdot 10^{-27}\,\mathrm{kg}\ (= 938.27\,\mathrm{MeV}/c_0^2)$
Rest mass of the neutron	$m_\mathrm{n} = 1.67493 \cdot 10^{-27}\,\mathrm{kg}\ (= 939.565\,\mathrm{MeV}/c_0^2)$
Unified atomic mass unit	$m_\mathrm{u} = 1.66054 \cdot 10^{-27}\,\mathrm{kg}\ (= 931.49\,\mathrm{MeV}/c_0^2)$
Planck's constant	$h = 6.6261 \cdot 10^{-34}\,\mathrm{Js}$
Boltzmann's constant	$k_\mathrm{B} = 1.381 \cdot 10^{-23} \, \frac{\mathrm{J}}{\mathrm{K}}$
Avogadro's number	$N_\mathrm{A} = 6.02214 \cdot 10^{23}\,\mathrm{mol}^{-1}$
Molar volume of ideal gas	$V_0 = 2.2414 \cdot 10^{-2} \, \frac{\mathrm{m}^3}{\mathrm{mol}}$ (at 273.15 K, 1013.25 hPa)
Standard acceleration of gravity	$g = 9.81 \, \frac{\mathrm{m}}{\mathrm{s}^2}$
Gravitational constant	$G = 6.67 \cdot 10^{-11} \, \frac{\mathrm{Nm}^2}{\mathrm{kg}^2}$
Mass of the sun	$m_\mathrm{sun} = 1.99 \cdot 10^{30}\,\mathrm{kg}$
Mass of the earth	$m_\mathrm{earth} = 5.98 \cdot 10^{24}\,\mathrm{kg}$
Mass of the moon	$m_\mathrm{moon} = 7.36 \cdot 10^{22}\,\mathrm{kg}$
Average radius of the sun	$r_\mathrm{sun} = 6.96 \cdot 10^8\,\mathrm{m}$
Average radius of the earth	$r_\mathrm{earth} = 6.37 \cdot 10^6\,\mathrm{m}$
Average radius of the moon	$r_\mathrm{moon} = 1.74 \cdot 10^6\,\mathrm{m}$
Average distance sun–earth	$d_\mathrm{s-e} = 1.496 \cdot 10^{11}\,\mathrm{m}$
Average distance earth–moon	$d_\mathrm{e-m} = 3.844 \cdot 10^8\,\mathrm{m}$

Table A.2 Formulas for Bessel functions of the first kind $\mathrm{J}_m(x)$ and Bessel functions of the second kind $\mathrm{Y}_m(x)$, modified Bessel functions of the first kind $\mathrm{I}_m(x)$, and modified Bessel functions of the second kind $\mathrm{K}_m(x)$ for $x \in \mathbb{R}$ and integers k

Formula		Meaning of $\mathcal{Z}$
$\mathcal{Z}_{-k}(x) = (-1)^k \, \mathcal{Z}_k(x)$	(A.72)	J or Y
$\mathcal{Z}_{-k}(x) = \mathcal{Z}_k(x)$	(A.73)	I or K
$x \left[\mathcal{Z}_{k-1}(x) + \mathcal{Z}_{k+1}(x)\right] = 2k \ \mathcal{Z}_k(x)$	(A.74)	J or Y
$x \left[\mathrm{I}_{k-1}(x) - \mathrm{I}_{k+1}(x)\right] = 2k \ \mathrm{I}_k(x)$	(A.75)	
$x \left[\mathrm{K}_{k-1}(x) - \mathrm{K}_{k+1}(x)\right] = -2k \ \mathrm{K}_k(x)$	(A.76)	
$x \frac{\mathrm{d}\mathcal{Z}_k(x)}{\mathrm{d}x} = k \mathcal{Z}_k(x) - x \mathcal{Z}_{k+1}(x)$	(A.77)	J, Y or K
$x \frac{\mathrm{d}\mathrm{I}_k(x)}{\mathrm{d}x} = k \mathrm{I}_k(x) + x \mathrm{I}_{k+1}(x)$	(A.78)	

The equations in this table are based on [7, 8]

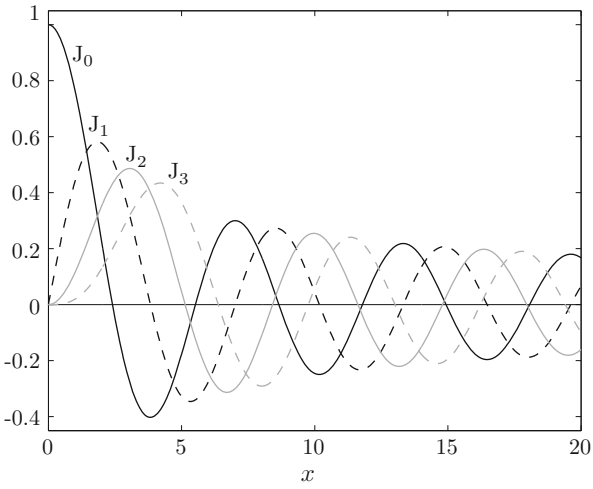

Fig. A.5 Bessel functions of the first kind $J_m(x)$ for $m \in \{0, 1, 2, 3\}$

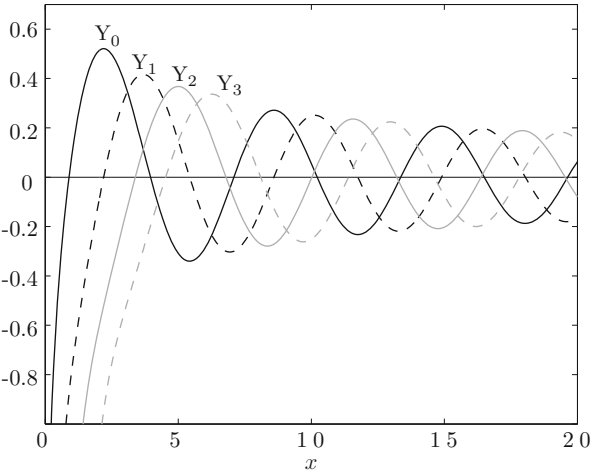

Fig. A.6 Bessel functions of the second kind $Y_m(x)$ for $m \in \{0, 1, 2, 3\}$

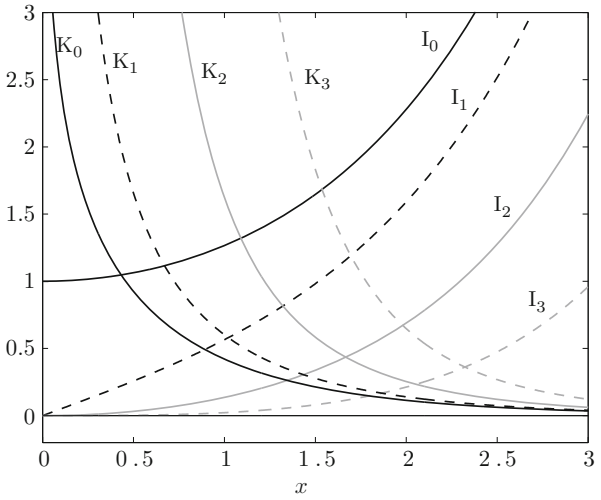

Fig. A.7 Modified Bessel functions of the first kind $\mathrm{I}_m(x)$ and modified Bessel functions of the second kind $\mathrm{K}_m(x)$ for $m \in \{0, 1, 2, 3\}$

Table A.3 Some Fourier transforms

$x(t)$	$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, dt$
$\delta(t)$	1
1	$2\pi\delta(\omega)$
$\Theta(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t \geq 0 \end{cases}$	$\pi\delta(\omega) + \dfrac{1}{j\omega}$
$\cos(\Omega t)$	$\pi[\delta(\omega - \Omega) + \delta(\omega + \Omega)]$
$\sin(\Omega t)$	$\dfrac{\pi}{j}[\delta(\omega - \Omega) - \delta(\omega + \Omega)]$
$e^{j\Omega t}$	$2\pi\delta(\omega - \Omega)$
$e^{-t^2/2a},\ a > 0$	$\sqrt{2\pi a}\, e^{-a\omega^2/2}$
$e^{-a\lvert t\rvert},\ a > 0$	$\dfrac{2a}{a^2 + \omega^2}$
$\displaystyle\sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T}\sum_{k=-\infty}^{\infty} e^{j2\pi kt/T}$	$\displaystyle\frac{2\pi}{T}\sum_{k=-\infty}^{\infty} \delta\left(\omega - k\frac{2\pi}{T}\right) = \sum_{k=-\infty}^{\infty} e^{jk\omega T}$
$x(t) = \begin{cases} 1 & \text{for } -T < t < T \\ 0 & \text{otherwise} \end{cases}$	$2T\,\mathrm{si}(T\omega)$
$x(t) = \begin{cases} 1 - \dfrac{\lvert t\rvert}{T} & \text{for } -T < t < T \\ 0 & \text{otherwise} \end{cases}$	$T\,\mathrm{si}^2\left(\dfrac{T\omega}{2}\right)$
$x(t) = \begin{cases} 1 + \cos(\Omega t) & \text{for } -\pi < \Omega t < \pi \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{2\pi}{\Omega}\,\dfrac{\mathrm{si}\left(\pi\frac{\omega}{\Omega}\right)}{1 - \left(\frac{\omega}{\Omega}\right)^2}$
$x(t) = \begin{cases} \cos^2\left(\frac{\Omega t}{2}\right) & \text{for } -\pi < \Omega t < \pi \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{\pi}{\Omega}\,\dfrac{\mathrm{si}\left(\pi\frac{\omega}{\Omega}\right)}{1 - \left(\frac{\omega}{\Omega}\right)^2}$
$x(t) = \begin{cases} 1 & \text{for } a < t < b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{j}{\omega}\left(e^{-jb\omega} - e^{-ja\omega}\right)$

Table A.4 Some one-sided Laplace transforms

$f(t)$	$F(s) = \int_0^\infty f(t)\, e^{-st}\, dt$
$\delta(t)$	1
$\Theta(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t \geq 0 \end{cases}$	$\dfrac{1}{s}$
$\dfrac{t^n}{n!}\,\Theta(t)$	$\dfrac{1}{s^{n+1}},\ n = 0, 1, \ldots$
$e^{-at}\,\Theta(t)$	$\dfrac{1}{s+a}$
$\dfrac{t^n}{n!}\,e^{-at}\,\Theta(t)$	$\dfrac{1}{(s+a)^{n+1}},\ n = 0, 1, \ldots$
$\dfrac{1}{a^{n+1}}\left[1 - \sum_{k=0}^{n} \dfrac{(at)^k}{k!}\, e^{-at}\right]\Theta(t)$	$\dfrac{1}{s(s+a)^{n+1}},\ a \neq 0,\ n = 0, 1, \ldots$
$\dfrac{1}{b-a}\left[e^{-at} - e^{-bt}\right]\Theta(t)$	$\dfrac{1}{(s+a)(s+b)},\ a \neq b$
$\sin(at)\,\Theta(t)$	$\dfrac{a}{s^2 + a^2}$
$\cos(at)\,\Theta(t)$	$\dfrac{s}{s^2 + a^2}$
$e^{-bt}\sin(at)\,\Theta(t)$	$\dfrac{a}{(s+b)^2 + a^2}$
$e^{-bt}\cos(at)\,\Theta(t)$	$\dfrac{s+b}{(s+b)^2 + a^2}$
$e^{-bt}\cos(at + c)\,\Theta(t)$	$\dfrac{(s+b)\cos c - a \sin c}{(s+b)^2 + a^2}$
$\sinh(at)\,\Theta(t)$	$\dfrac{a}{s^2 - a^2}$
$\cosh(at)\,\Theta(t)$	$\dfrac{s}{s^2 - a^2}$
$\cosh(at + b)\,\Theta(t)$	$\dfrac{s\cosh b + a \sinh b}{s^2 - a^2}$

Table A.5 Dependence of some bucket quantities on the reference phase φ_R

φ_R (°)	$\Delta\varphi_l$ (°)	$\varphi_{RF,B,len}$ (°)	Y	α	φ_R (°)	$\Delta\varphi_l$ (°)	$\varphi_{RF,B,len}$ (°)	Y	α
1	−154.95	332.95	0.98627	0.95410	2	−145.47	321.47	0.97252	0.91756
3	−138.49	312.49	0.95873	0.88451	4	−132.79	304.79	0.94491	0.85374
5	−127.89	297.89	0.93107	0.82468	6	−123.56	291.56	0.91721	0.79698
7	−119.64	285.64	0.90333	0.77044	8	−116.06	280.06	0.88943	0.74491
9	−112.74	274.74	0.87552	0.72026	10	−109.65	269.65	0.86160	0.69641
11	−106.74	264.74	0.84767	0.67330	12	−103.99	259.99	0.83373	0.65087
13	−101.38	255.38	0.81979	0.62908	14	−98.89	250.89	0.80585	0.60789
15	−96.51	246.51	0.79192	0.58726	16	−94.23	242.23	0.77799	0.56717
17	−92.03	238.03	0.76407	0.54760	18	−89.91	233.91	0.75016	0.52853
19	−87.86	229.86	0.73626	0.50993	20	−85.88	225.88	0.72238	0.49180
21	−83.96	221.96	0.70852	0.47411	22	−82.09	218.09	0.69469	0.45686
23	−80.27	214.27	0.68088	0.44004	24	−78.50	210.50	0.66710	0.42363
25	−76.77	206.77	0.65335	0.40762	26	−75.08	203.08	0.63963	0.39202
27	−73.43	199.43	0.62595	0.37679	28	−71.82	195.82	0.61232	0.36195
29	−70.24	192.24	0.59872	0.34749	30	−68.69	188.69	0.58517	0.33339
31	−67.17	185.17	0.57167	0.31966	32	−65.67	181.67	0.55823	0.30628
33	−64.21	178.21	0.54483	0.29325	34	−62.76	174.76	0.53150	0.28057
35	−61.34	171.34	0.51823	0.26823	36	−59.95	167.95	0.50502	0.25623
37	−58.57	164.57	0.49188	0.24456	38	−57.21	161.21	0.47880	0.23322
39	−55.88	157.88	0.46581	0.22220	40	−54.56	154.56	0.45289	0.21151
41	−53.25	151.25	0.44005	0.20113	42	−51.97	147.97	0.42729	0.19106
43	−50.70	144.70	0.41462	0.18130	44	−49.44	141.44	0.40204	0.17185
45	−48.20	138.20	0.38955	0.16270	46	−46.97	134.97	0.37715	0.15385
47	−45.75	131.75	0.36486	0.14529	48	−44.55	128.55	0.35267	0.13702
49	−43.36	125.36	0.34059	0.12904	50	−42.17	122.17	0.32861	0.12135
51	−41.00	119.00	0.31676	0.11394	52	−39.84	115.84	0.30501	0.10680
53	−38.69	112.69	0.29339	0.09994	54	−37.55	109.55	0.28190	0.09335
55	−36.42	106.42	0.27053	0.08702	56	−35.29	103.29	0.25929	0.08096
57	−34.18	100.18	0.24819	0.07516	58	−33.07	97.07	0.23723	0.06961
59	−31.97	93.97	0.22642	0.06431	60	−30.87	90.87	0.21575	0.05927
61	−29.79	87.79	0.20524	0.05446	62	−28.70	84.70	0.19489	0.04990
63	−27.63	81.63	0.18470	0.04558	64	−26.56	78.56	0.17468	0.04148
65	−25.50	75.50	0.16482	0.03762	66	−24.44	72.44	0.15515	0.03397
67	−23.38	69.38	0.14566	0.03055	68	−22.33	66.33	0.13636	0.02734
69	−21.29	63.29	0.12725	0.02435	70	−20.25	60.25	0.11835	0.02155
71	−19.21	57.21	0.10965	0.01896	72	−18.18	54.18	0.10116	0.01657
73	−17.15	51.15	0.09290	0.01437	74	−16.13	48.13	0.08487	0.01235
75	−15.10	45.10	0.07707	0.01051	76	−14.08	42.08	0.06953	0.00885
77	−13.07	39.07	0.06224	0.00735	78	−12.05	36.05	0.05522	0.00602
79	−11.04	33.04	0.04848	0.00484	80	−10.03	30.03	0.04203	0.00382
81	−9.02	27.02	0.03590	0.00293	82	−8.02	24.02	0.03009	0.00218
83	−7.01	21.01	0.02464	0.00156	84	−6.01	18.01	0.01955	0.00106
85	−5.00	15.00	0.01488	0.00067	86	−4.00	12.00	0.01065	0.00039
87	−3.00	9.00	0.00692	0.00019	88	−2.00	6.00	0.00377	0.00007
89	−1.00	3.00	0.00133	0.00001	90	0.00	0.00	0.00000	0.00000

Table A.6 Dependence of some trajectory parameters on $\Delta\hat{\varphi}_{\mathrm{RF}}$ for $\varphi_{\mathrm{R}} = 0$

$\Delta\hat{\varphi}_{\mathrm{RF}}$ (°)	$\frac{\omega_{\mathrm{S,stat}}}{\omega_{\mathrm{S,0,stat}}}$	η_{fill}	$\frac{\Delta\hat{W}/\Delta\hat{t}}{\left(\Delta\hat{W}/\Delta\hat{t}\right)\big\vert_{\mathrm{lin.}}}$	$\Delta\hat{\varphi}_{\mathrm{RF}}$ (°)	$\frac{\omega_{\mathrm{S,stat}}}{\omega_{\mathrm{S,0,stat}}}$	η_{fill}	$\frac{\Delta\hat{W}/\Delta\hat{t}}{\left(\Delta\hat{W}/\Delta\hat{t}\right)\big\vert_{\mathrm{lin.}}}$
2	0.99992	0.00024	0.99995	4	0.99970	0.00096	0.99980
6	0.99931	0.00215	0.99954	8	0.99878	0.00382	0.99919
10	0.99810	0.00597	0.99873	12	0.99726	0.00859	0.99817
14	0.99627	0.01169	0.99751	16	0.99513	0.01525	0.99675
18	0.99383	0.01928	0.99589	20	0.99239	0.02377	0.99493
22	0.99079	0.02873	0.99387	24	0.98904	0.03414	0.99271
26	0.98714	0.04000	0.99144	28	0.98509	0.04631	0.99008
30	0.98289	0.05306	0.98862	32	0.98054	0.06025	0.98705
34	0.97803	0.06788	0.98539	36	0.97537	0.07593	0.98363
38	0.97257	0.08440	0.98177	40	0.96961	0.09328	0.97982
42	0.96650	0.10257	0.97776	44	0.96325	0.11226	0.97561
46	0.95984	0.12234	0.97336	48	0.95628	0.13280	0.97101
50	0.95258	0.14364	0.96857	52	0.94872	0.15485	0.96603
54	0.94472	0.16641	0.96340	56	0.94056	0.17832	0.96067
58	0.93626	0.19057	0.95785	60	0.93181	0.20315	0.95493
62	0.92721	0.21605	0.95192	64	0.92246	0.22925	0.94882
66	0.91756	0.24276	0.94562	68	0.91252	0.25654	0.94234
70	0.90732	0.27061	0.93896	72	0.90198	0.28493	0.93549
74	0.89649	0.29951	0.93193	76	0.89085	0.31433	0.92828
78	0.88506	0.32937	0.92455	80	0.87913	0.34462	0.92073
82	0.87304	0.36007	0.91681	84	0.86681	0.37571	0.91282
86	0.86043	0.39152	0.90874	88	0.85389	0.40749	0.90457
90	0.84721	0.42361	0.90032	92	0.84038	0.43985	0.89598
94	0.83340	0.45620	0.89156	96	0.82626	0.47266	0.88706
98	0.81898	0.48920	0.88248	100	0.81154	0.50580	0.87782
102	0.80394	0.52246	0.87308	104	0.79619	0.53915	0.86826
106	0.78829	0.55587	0.86337	108	0.78022	0.57258	0.85839
110	0.77200	0.58928	0.85334	112	0.76361	0.60595	0.84822
114	0.75506	0.62256	0.84302	116	0.74634	0.63911	0.83775
118	0.73746	0.65557	0.83241	120	0.72840	0.67193	0.82699
122	0.71916	0.68816	0.82151	124	0.70974	0.70425	0.81595
126	0.70014	0.72018	0.81033	128	0.69035	0.73592	0.80464
130	0.68036	0.75146	0.79889	132	0.67016	0.76678	0.79307
134	0.65976	0.78186	0.78718	136	0.64913	0.79667	0.78123
138	0.63828	0.81119	0.77522	140	0.62718	0.82540	0.76915
142	0.61582	0.83928	0.76302	144	0.60419	0.85280	0.75683
146	0.59228	0.86594	0.75058	148	0.58004	0.87868	0.74427
150	0.56747	0.89098	0.73791	152	0.55453	0.90282	0.73150
154	0.54118	0.91418	0.72503	156	0.52737	0.92502	0.71851
158	0.51304	0.93531	0.71194	160	0.49813	0.94503	0.70532
162	0.48253	0.95413	0.69865	164	0.46613	0.96257	0.69193
166	0.44874	0.97032	0.68517	168	0.43014	0.97734	0.67836
170	0.40994	0.98356	0.67150	172	0.38759	0.98893	0.66461
174	0.36205	0.99337	0.65767	176	0.33120	0.99681	0.65069
178	0.28902	0.99910	0.64367	180	0.00000	1.00000	$2/\pi \approx 0.63662$

References

1. G.M. Fichtenholz, *Differential- und Integralrechnung* (Verlag Harri Deutsch, Frankfurt am Main, 1997) [Bd. 1: 14. Aufl. 1997, Bd. 2: 10. Aufl. 1990, Bd. 3: 12. Aufl. 1992]
2. K. Burg, H. Haf, F. Wille, *Höhere Mathematik für Ingenieure* (B. G. Teubner, Stuttgart, 2002) [Bd. I: 5. Aufl. 2001, Bd. II: 4. Aufl. 2002, Bd. III: 4. Aufl. 2002, Bd. IV: 2. Aufl. 1994, Bd. V: 2. Aufl. 1993]
3. D.A. Edwards, M.J. Syphers, *An Introduction to the Physics of High Energy Accelerators* (Wiley-VCH Verlag GmbH&Co.KGaA, Weinheim, 2004)
4. H. Wiedemann, *Particle Accelerator Physics I & II*, 2nd edn. (Springer, Berlin/Heidelberg/New York, 2003)
5. E.J.N. Wilson, *An Introduction to Particle Accelerators* (Oxford University Press, Oxford/New York, 2001)
6. M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New York/Chichester/Brisbane/Toronto/Singapore, 1994)
7. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965)
8. I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products*, 6th edn. (Academic, San Diego/San Francisco/New York/Boston/London/Sydney/Tokyo, 2000)

Index

Harald Klingbeil, *Theoretical Foundations of Synchrotron and Storage Ring RF Systems*,
https://doi.org/10.1007/978-3-319-07188-6