ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CAS CERN ACCELERATOR SCHOOL
ADVANCED ACCELERATOR PHYSICS

The Queen's College, Oxford, England
16–27 September 1985

PROCEEDINGS
Editor: S. Turner
Vol. II

GENEVA
1987
ABSTRACT

This advanced course on general accelerator physics is the second of the biennial series given by the CERN Accelerator School and follows on from the first basic course given at Gif-sur-Yvette, Paris, in 1984 (CERN Yellow Report 85-19). Stress is placed on the mathematical tools of Hamiltonian mechanics and the Vlasov and Fokker-Planck equations, which are widely used in accelerator theory. The main topics treated in this present work include: nonlinear resonances, chromaticity, motion in longitudinal phase space, growth and control of longitudinal and transverse beam emittance, space-charge effects and polarization. The seminar programme treats some specific accelerator techniques, devices, projects and future possibilities.
CONTENTS

Foreword vii

Opening address 1

Local coordinates for the beam and frequently used symbols 3

J.S. Bell

Hamiltonian mechanics 5
  Introduction 5
  Poisson brackets 8
  Stationary and varying action 10
  Poincaré invariant 11
  Lagrange invariant 13
  Symplectic 14
  Liouville's theorem 15
  Conserved quadratic form 16
  Characteristic exponents 19
  Canonical transformations 20
  Point transformation 22
  Change of independent variable 23
  Scaling 24
  Dynamical evolution as canonical transformation 25
  Poincaré invariant 25
  Lagrange invariant 27
  Liouville invariant 29
  One degree of freedom 30
  Action and angle variables 31
  Small deviations from closed orbit 31
  Adiabatic invariance of J 35
  Small canonical transformation 36
  Canonical perturbation theory 37

E.J.N. Wilson

Nonlinear resonances 41
  Introduction 41
  The general form of the Hamiltonian 43
  The magnetic vector potential for multipoles 45
  Linear dynamics in action angle variables 46
  Perturbation theory 52
  Effect of nonlinearities far from a resonance 55
  Resonances 56
  The third-integer resonance 57
  The trajectory of a third-integer resonance 59
  The effect of an octupole 62
  Phase-space topology with amplitude frequency variation 63
  Amplitude growth on crossing a resonance 65
  Synchrotron resonances 66
  Beam lifetime due to magnet imperfections 68
  The effect of two dimensions of transverse motion 69
  Three dimensions of magnetic field 72
  Conclusions 73
<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.N. Lashmore-Davies</td>
<td>Waves in plasmas</td>
<td>235-247</td>
</tr>
<tr>
<td></td>
<td>Waves in a field-free plasma</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>Waves in a magnetized plasma</td>
<td>236-240</td>
</tr>
<tr>
<td></td>
<td>Low-frequency waves in a magnetized plasma</td>
<td>240-245</td>
</tr>
<tr>
<td></td>
<td>Raman scattering</td>
<td>247</td>
</tr>
<tr>
<td>H.G. Hereward</td>
<td>Landau damping</td>
<td>255-261</td>
</tr>
<tr>
<td></td>
<td>Spectrum of linear oscillations</td>
<td>255-259</td>
</tr>
<tr>
<td></td>
<td>Longitudinal instability</td>
<td>259</td>
</tr>
<tr>
<td></td>
<td>Nonlinear oscillations</td>
<td>261</td>
</tr>
<tr>
<td>J.L. Laclare</td>
<td>Bunched beam coherent instabilities</td>
<td>264-325</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>264</td>
</tr>
<tr>
<td></td>
<td>Longitudinal instabilities</td>
<td>264</td>
</tr>
<tr>
<td></td>
<td>Transverse instabilities</td>
<td>306</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>325</td>
</tr>
<tr>
<td>I. Hofmann</td>
<td>Space charge dominated beam transport</td>
<td>327-332</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>327</td>
</tr>
<tr>
<td></td>
<td>Basic properties</td>
<td>327</td>
</tr>
<tr>
<td></td>
<td>Emittance and field energy</td>
<td>329</td>
</tr>
<tr>
<td></td>
<td>Application to emittance growth</td>
<td>332</td>
</tr>
<tr>
<td>L. Palumbo and V.G. Vasaró</td>
<td>Wake fields, impedances and Green's function</td>
<td>341-369</td>
</tr>
<tr>
<td></td>
<td>Longitudinal wake potential and impedance</td>
<td>341-344</td>
</tr>
<tr>
<td></td>
<td>Longitudinal impedance and wake potential for simple structures</td>
<td>344-357</td>
</tr>
<tr>
<td></td>
<td>General analysis</td>
<td>357</td>
</tr>
<tr>
<td></td>
<td>Circular accelerators</td>
<td>361</td>
</tr>
<tr>
<td></td>
<td>Transverse-wake potential and impedance</td>
<td>363</td>
</tr>
<tr>
<td></td>
<td>Remarks</td>
<td>368</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td>369</td>
</tr>
<tr>
<td>G. Döme</td>
<td>Diffusion due to RF noise</td>
<td>370-392</td>
</tr>
<tr>
<td></td>
<td>Statistical properties of random variables</td>
<td>370</td>
</tr>
<tr>
<td></td>
<td>Fokker-Planck equation</td>
<td>372</td>
</tr>
<tr>
<td></td>
<td>Differential equations for a stationary bucket with amplitude and phase noise</td>
<td>374-377</td>
</tr>
<tr>
<td></td>
<td>Computation of the coefficients $A_1$, $A_2$ in the Fokker-Planck equation</td>
<td>377-382</td>
</tr>
<tr>
<td></td>
<td>Case of a sinusoidal RF voltage</td>
<td>382</td>
</tr>
<tr>
<td></td>
<td>Finite difference equations</td>
<td>386</td>
</tr>
<tr>
<td></td>
<td>Diffusion equation</td>
<td>388</td>
</tr>
<tr>
<td></td>
<td>Contribution of RF noise to the finite beam lifetime in the SPS collider</td>
<td>392</td>
</tr>
<tr>
<td>A. Pivénekí</td>
<td>Intra-beam scattering</td>
<td>402-411</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>402</td>
</tr>
<tr>
<td></td>
<td>Calculation of rise times and damping times</td>
<td>403-411</td>
</tr>
<tr>
<td>Author</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>D. Boussard</td>
<td>Schottky noise and beam transfer function diagnostics</td>
<td>416</td>
</tr>
<tr>
<td></td>
<td>Schottky signals</td>
<td>417</td>
</tr>
<tr>
<td></td>
<td>Beam detectors</td>
<td>425</td>
</tr>
<tr>
<td></td>
<td>Observation of Schottky signals</td>
<td>440</td>
</tr>
<tr>
<td></td>
<td>Beam transfer functions</td>
<td>445</td>
</tr>
<tr>
<td>D. Möhl</td>
<td>Stochastic cooling</td>
<td>453</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>453</td>
</tr>
<tr>
<td></td>
<td>Simplified theory, time-domain picture</td>
<td>455</td>
</tr>
<tr>
<td></td>
<td>A more detailed presentation of betatron cooling, frequency domain</td>
<td>478</td>
</tr>
<tr>
<td></td>
<td>of electron cooling, momentum scaling</td>
<td>510</td>
</tr>
<tr>
<td>H. Poth</td>
<td>Electron cooling</td>
<td>534</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>534</td>
</tr>
<tr>
<td></td>
<td>What electron cooling is</td>
<td>534</td>
</tr>
<tr>
<td></td>
<td>Why electron cooling?</td>
<td>535</td>
</tr>
<tr>
<td></td>
<td>Ion-beam and storage-ring properties</td>
<td>535</td>
</tr>
<tr>
<td></td>
<td>How electron cooling works in principle</td>
<td>537</td>
</tr>
<tr>
<td></td>
<td>Introduction to electron cooling theory (for pedestrians)</td>
<td>538</td>
</tr>
<tr>
<td></td>
<td>Experimental realization of electron cooling</td>
<td>543</td>
</tr>
<tr>
<td></td>
<td>More on the theory of electron cooling</td>
<td>552</td>
</tr>
<tr>
<td></td>
<td>Recombination</td>
<td>555</td>
</tr>
<tr>
<td></td>
<td>Electron cooling experiments</td>
<td>557</td>
</tr>
<tr>
<td></td>
<td>Simulation of electron cooling in storage rings</td>
<td>561</td>
</tr>
<tr>
<td></td>
<td>Electron cooling diagnostics</td>
<td>562</td>
</tr>
<tr>
<td></td>
<td>Applications of electron cooling</td>
<td>564</td>
</tr>
<tr>
<td></td>
<td>Electron cooling projects</td>
<td>566</td>
</tr>
<tr>
<td>J. Jowett</td>
<td>Electron dynamics with radiation and nonlinear wigglers</td>
<td>570</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>570</td>
</tr>
<tr>
<td></td>
<td>The dynamics of electrons in a storage ring</td>
<td>571</td>
</tr>
<tr>
<td></td>
<td>Normal modes and optical functions</td>
<td>579</td>
</tr>
<tr>
<td></td>
<td>Radiation damping</td>
<td>594</td>
</tr>
<tr>
<td></td>
<td>Quantum fluctuations and Fokker-Planck equations</td>
<td>587</td>
</tr>
<tr>
<td></td>
<td>Nonlinear wigglers</td>
<td>591</td>
</tr>
<tr>
<td>J. Lé Duff</td>
<td>Beam break up</td>
<td>610</td>
</tr>
<tr>
<td></td>
<td>Experimental evidence</td>
<td>610</td>
</tr>
<tr>
<td></td>
<td>Deflecting modes in circular iris loaded waveguides</td>
<td>614</td>
</tr>
<tr>
<td></td>
<td>Regenerative beam break up</td>
<td>617</td>
</tr>
<tr>
<td></td>
<td>Cumulative beam break up</td>
<td>621</td>
</tr>
<tr>
<td>D. Boussard</td>
<td>Beam loading</td>
<td>626</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>626</td>
</tr>
<tr>
<td></td>
<td>Single bunch passage in a cavity</td>
<td>627</td>
</tr>
<tr>
<td></td>
<td>Multiple bunch passages</td>
<td>629</td>
</tr>
<tr>
<td></td>
<td>Limiting case $\delta_0 = 0$</td>
<td>630</td>
</tr>
<tr>
<td></td>
<td>The case of a travelling wave structure</td>
<td>633</td>
</tr>
<tr>
<td></td>
<td>Transient correction</td>
<td>635</td>
</tr>
<tr>
<td></td>
<td>RF drive generation</td>
<td>637</td>
</tr>
</tbody>
</table>
J. Buon

Polarization in electron and proton beams

Introduction 647
Generalities on polarization and spin motion 649
Acceleration of polarized protons in synchrotrons 661
Polarization of electrons in storage rings 671

H. Maia, G. Ripken, A. Hreulich and F. Schmidt

Particle tracking 690
Introduction 690
Hamiltonian description of the proton motion 690
Dynamic aperture 693
Particle tracking 694
Qualitative theory of dynamical systems 696
Studies of chaotic behaviour in HERA caused by transverse magnetic multipole fields 699
Summary 704

M. Puglisi

The radiofrequency quadrupole linear accelerator 706
Introduction 706
The accelerating structure 706
Outline of the T-K expansion 709
The wane tips shaping 712
Physical considerations 717
The structure of an RFQ 720
Design and technical considerations 728
Recent developments 733

H. Piel

Fundamental features of superconducting cavities for high energy accelerators 736
Introduction 736
Some cavity fundamentals 737
Superconducting cavities 741
Cavity design 746
Anomalous losses 751
Cavities covered with superconducting thin films 761
Current accelerator projects and achievements 765

J. Le Duff

High field electron linacs 772
Introduction 772
Extrapolation of present technologies 772
RF compression scheme 775
Ultimate accelerating gradients in conventional structures 783
A survey of accelerating structures 785
RF power source: the lasertron 787

G. Dattoli, A. Renieri and A. Torre

Free electron lasers: a short review of the theory and experiments 792
Introduction 792
FEL: theory and design criteria 795
FEL storage ring operation 803
Single passage FEL operation 809
Conclusions 814

D.A. Gray and G.R. Rees

ISIS, the accelerator based neutron source at RAL 817
Introduction 817
Linac and synchrotron 817
Target station 819
Experimental facilities 821
High intensity performance of the ISIS synchrotron 822
H. Gubrød

Heavy ions: the present and future*)

U. Amaldi

Linear colliders versus storage rings*)

List of participants

*) Contribution not received
SCHOTTKY NOISE AND BEAM TRANSFER FUNCTION DIAGNOSTICS

D. Boussard
CERN, Geneva, Switzerland

ABSTRACT
Following the analysis of Schottky signals for the unbunched and the bunched beam cases, a general study of electromagnetic detectors is presented. Here the image-current approach and the Lorentz reciprocity theorem will be used to evaluate the detector (or pick-up) performance for several typical examples. Then, signal-processing techniques, which play an important role in the Schottky signal analysis, will be reviewed. The beam transfer function which relates the beam response to an external excitation also provides very useful information about the accelerator behaviour. It requires an element to excite the beam (kicker) which will be shown to be equivalent to a detector working in reverse. With beam-transfer-function measurements an assessment of beam stability limits can be made, leading to the determination of the overall ring impedance.

The noise generated in an old fashioned electron tube is governed by the Schottky formula which simply reflects the fact that the anode current is composed of individual electrons randomly emitted by the cathode. Very similarly, the beam current in a circular particle accelerator, also exhibits a random component, called the Schottky noise, which results from the large, but finite, number of particles in the beam. In the absence of random quantum emissions (i.e. for hadron machines) the analysis of Schottky noise signals (or Schottky signals, for brevity) is a very powerful tool to study the accelerator behaviour. Historically, Schottky signals have been observed first on unbunched beam machines (CERN ISR),\(^1\,^2\) leading to the development of the very successful stochastic cooling technique. For bunched beams, the presence of strong “macroscopic” beam signals renders the observation of the tiny Schottky signals more difficult. However improved signal processing techniques have recently made their observation possible.

Following the analysis of Schottky signals for the unbunched and the bunched beam cases, a general study of electromagnetic detectors is presented. Here the image-current approach and the Lorentz reciprocity theorem will be used to evaluate the detector (or pick-up) performance for several typical examples. Then, signal-processing techniques, which play an important role in the Schottky signal analysis, will be reviewed.

The beam transfer function which relates the beam response to an external excitation also provides very useful information about the accelerator behaviour. It requires an element to excite the beam (kicker) which will be shown to be equivalent to a detector working in reverse. With beam-transfer-function measurements an assessment of beam stability limits can be made, leading to the determination of the overall ring impedance.
1 SCHOTTKY SIGNALS

1.1 Unbunched beam, longitudinal

For a single particle circulating in the machine (charge e, revolution period $T_1 = 1/f_1$), the beam current, at a given location in the ring, is composed of an infinite train of delta pulses (Fig. 1a) separated in time by $T_1$. In frequency domain, this periodic waveform is represented by a line spectrum (Fig. 1b), the distance between lines being $f_1 = \omega_1/2\pi$.

$$i_i(t) = ef_i \sum_{n=-\infty}^{+\infty} \exp jn\omega_1 t \quad \text{(1)}$$

Looking at positive frequencies only:

$$i_i(t) = ef_i + 2ef_0 \sum_{n=1}^{+\infty} \cos n\omega_1 t \quad \text{(2)}$$

The first term represents the DC component, the others are simply the successive harmonics of the revolution frequency.

![Graphs](image)

**Fig. 1** a) Time domain δ pulses   b) Frequency domain: line spectrum

For $N$ particles, randomly distributed in azimuth along the ring circumference (debunched beam case) and having slightly different $\omega_1$, each line at frequency $nf_1$ which is infinitely narrow in the case of a single particle, will be replaced by a band of frequencies (Schottky band) whose width is simply:

$$\Delta f = n \Delta f_1 = n f_0 \eta \frac{\Delta p}{p} \quad \text{(3)}$$

$\Delta f_1$ is the spread in particle's revolution frequencies resulting from the relative momentum spread $\Delta p/p$ and the machine parameter $\eta = (1/\gamma_0^2 - 1/\gamma^2)$. $f_0$ is the average revolution frequency.
When averaging equation (2) over \( N \) particles, only the DC terms remain (\( i_{\text{DC}} = N i_0 \)), the other components cancel due to the random azimuth phase factor. However, the r.m.s. current per band which is given by the sum:

\[
<i^2> = [2ef_o (\cos \theta_1 + \cos \theta_2 + \ldots + \cos \theta_{\frac{N}{2}})]^2
\]

(4)
does not vanish because of the \( \cos \theta^2 \) terms. One obtains:

\[
i_{\text{rms}} = \sqrt{<i^2>} = 2ef_o \sqrt{N \cos \theta_1} \frac{\sqrt{2}}{2}
\]

(5)

(6)

The r.m.s. current per band (Schottky current) is independent of \( n \) (harmonic number) and proportional to the square root of the number of particles \( N \).

As indicated on Fig. 2, the power spectral density, proportional to \( <i^2>/\Delta f \), decreases with \( n \) until overlap occurs \( (\Delta f > f_o) \). For a given band the local power density is obviously proportional to the number of particles per unit frequency. If the parameter \( n \) is known \( (n \text{ may be frequency dependent}) \), the measurement of the power spectral density, in one particular Schottky band gives directly the \( \Delta p/p \) distribution of the beam.

![Power spectral density of Schottky lines with increasing n.](image)

This forms the basis of \( \Delta p/p \) beam distribution measurements in DC coasting machines, (cooling and accumulation rings in particular).

Note that the noise signals pertaining to successive Schottky bands are not correlated because the random azimuthal phase factor is multiplied by \( n \) in Eq. (4).

1.2 Unbunched beam, transverse

For a single particle, the beam current \( i_1(t) \) must be replaced by the dipole moment: \( d_1(t) = a_1(t) \cdot i_1(t) \), where \( a_1(t) \) is the transverse displacement. The \( i^{th} \) particle executes a sinusoidal betatron oscillation, of amplitude \( a_i \), which can be written:
\[ a_i(t) = a_i \cos(q_i \omega_i t + \varphi_i) \]  

(7)

Here \( q_i \), \( f_i \) is the observed frequency, at a fixed location in the ring, \( q_i \) being the non-integer part of the betatron tune (Fig. 3a).

In frequency domain:

\[ d_i(t) = a_i \cos(q_i \omega_i t + \varphi_i) ef_o \lim_{n \to \infty} \sum_{n=-\infty}^{+\infty} \exp jn\omega_i t \]  

(8)

![Diagram](a)

![Diagram](b)

Fig. 3 Time (a) and Frequency (b) domain representations of a single-particle transverse oscillation.

\[ d_i(t) = a_i \cos(q_i \omega_i t + \varphi_i) ef_o \left[ \sum_{n=-\infty}^{+\infty} \exp \{j(n + q_i) \omega_i t + \varphi_i) \} \right] \]  

(9)

The spectrum is again a series of lines spaced by the revolution frequency of the \( i^{th} \) particle, but shifted in frequency by \( q_i f_i \). Looking at positive frequencies only (Fig. 3b) one obtains two betatron lines per revolution frequency band as in the case of an amplitude-modulated carrier which exhibits two symmetrical sidebands.

For \( N \) particles in the beam, again randomly distributed in azimuth and in betatron phases, averaging equation (9), for a given value of \( n + q_i \), gives:

\[ \langle d \rangle = 0 \]

\[ \langle d^2 \rangle = \langle a_i^2 \rangle e^2 f_o \frac{N}{2} \]  

(10)

\[ \text{d rms} = ef_o a_{\text{rms}} \sqrt{\frac{N}{2}} \]  

(11)

Again, the total power per Schottky band is independent of its location in the frequency spectrum; it is proportional to the number of particles in the beam and to the square of the r.m.s. oscillation amplitude.
Each Schottky band has now a finite width which results from the spread of revolution frequencies $\Delta f_1/f_o = \eta \Delta p/p$ and from the spread of betatron frequencies $\Delta q_1$. The latter usually comes from the machine chromaticity $\xi$: $\Delta q_1 = Q \xi \Delta p/p$, but may also result from space charge, beam-beam or nonlinear effects.

The line width of two adjacent Schottky bands $(n \pm q)$ is given by:

$$\Delta f = (n \pm q) \Delta f_1 \pm 2 \pi f_o \Delta q_1$$

(12)

$$\Delta f = f_o \frac{\Delta p}{p} \left[ (n \pm q) \eta \pm Q \xi \right]$$

(13)

if only chromaticity contributes to the betatron frequency spread.

Equation (13) shows that the width of the two Schottky bands is not the same, due to the machine chromaticity. However, by comparing the two bands $n \pm q$, one can determine the $\Delta q_1$ of the beam. Even more, if one can identify similar points on the distribution (resonances, for instance), their $q$ can be determined by the formula:

$$q = \frac{1}{2} \left( 1 + \frac{\Delta f_c}{f_o} \right)$$

(14)

$\Delta f_c$ being the measured frequency difference between them. This technique was extensively used in the ISR to monitor the working line of the machine distribution in transverse tunes.

Comparing equations (10) and (6) gives a direct measure of the r.m.s. betatron amplitude:

$$\frac{d_{\text{rms}}}{i_{\text{rms}}} = \frac{a_{\text{rms}}}{2}$$

(15)

Equation (15) can be used to measure directly the transverse beam emittance, if the beam distribution is known. This obviously requires well calibrated longitudinal and transverse detectors to measure accurately $d_{\text{rms}}$ and $i_{\text{rms}}$ unless only relative measurements are sought (evolution of $\alpha$A transverse emittance, for instance).

1.3 Bunched beam, longitudinal

In the bunched beam case, every individual particle executes synchrotron oscillations at the frequency $Q_s/2\pi$. The time of passage of the particle in front of the detector is modulated according to:
\[ \tau_i(t) = \tau_i \sin \left( \Omega_s t + \psi_i \right) \]  \hspace{1cm} (16)

\( \tau_i(t) \) is the time difference with respect to the synchronous particle (frequency \( f_0 \)) and \( \tau_i \) is the amplitude of the synchrotron oscillation, assumed to be linear. In the time domain, the beam current is represented in Fig. 4, as a series of delta pulses, with a modulated time of passage. It can be written:

\[ I_i(t) = ef_0 + 2 ef_0 R_e \left\{ \sum_{n=1}^{\infty} \exp jn \omega_0 (t + \tau_i \sin (\Omega_s t + \psi_i)) \right\} \]  \hspace{1cm} (17)

![Fig. 4 Time domain representation of a single particle current in a bunched beam.](image)

Using the relation:

\[ \exp (j(z \sin \theta)) = \sum_{p=-\infty}^{+\infty} J_p(z) e^{j p \theta} \]  \hspace{1cm} (18)

where \( J_p \) is the Bessel function of order \( p \), one can expand the nth harmonic in equation (17) and obtain:

\[ I_n = 2 ef_0 R_e \left\{ \sum_{p=-\infty}^{+\infty} J_p(n \omega_0 \tau_i) \exp j(n \omega_0 t + p \Omega_s t + p \psi_i) \right\} \]  \hspace{1cm} (19)

Each revolution frequency line (nf) now splits into an infinity of synchrotron satellites, spaced by \( \Omega_s / 2\pi \), the amplitudes of which being proportional to the Bessel functions of argument \( n \omega_0 \tau_i \) as shown in Fig. 5.

The amplitudes of the synchrotron satellites become negligible beyond a certain value of \( p \). This is because \( J_p(x) \approx 0 \) for \( p > x \) if \( x \) is large. Therefore, the synchrotron satellites are, in practice, confined into a limited bandwidth:

\[ 2p \Omega_s = 2n \omega_0 \tau_i \Omega_s \]  \hspace{1cm} (20)
The spread in the instantaneous revolution frequency of the $i$th particle due to the synchrotron oscillation is simply:

$$2\Delta\omega_i = 2\Omega_s n\omega_i t_i$$  \hspace{1cm} \text{(21)}$$

Consequently, for large values of $n$, the significant bandwidth around line $i$ is the same as that of a beam of many particles having the same $\Delta\omega_i$ and therefore the same $\Delta p/p$.

Consider now the case of many particles, with randomly distributed synchrotron phases $\psi_i$ and $\hat{t}_i$ ranging from 0 to $\hat{t}_m$ ($2\hat{t}_m$ being the total bunch length).

For a given $n$, the central line ($p = 0$) shows the same phase factor ($\exp jn\omega t$) for all particles: the current in the central line is therefore proportional to $N$ and not $\sqrt{N}$; this is simply the macroscopic RF current of the bunch. On the contrary, the synchrotron satellites ($p \neq 0$) add r.m.s. wise because of the random phase factor

$$\exp j(n\omega t + p\Omega_s + p\psi_i)$$

(Fig. 6).

Each line is infinitely narrow if the synchrotron oscillation is purely linear ($\Omega_s$ is the same for all particles) and if the machine has no imperfections. However, magnet and RF fluctuations broaden in practice each individual line. In addition a spread in synchrotron frequency within the bunch $\Delta\Omega_s$ transforms each satellite ($p \neq 0$) into a band of width $p\Delta\Omega_s$. For large values of $n$, overlap between successive synchrotron satellites ($p\Delta\Omega_s > \Omega_s$) can occur within the significant width of the Schottky band of order $n$ (Fig. 6b).

If we consider two Schottky bands with different values of $n$, their corresponding synchrotron satellites (of order $p$) are correlated. This results from Eq. (19), where the random phase factor $p\psi_i$ is the same, even for different values of $n$.

Another way to look at the coherence between successive Schottky bands is to examine the bunch signal in time domain (Fig. 7). It is composed of a steady component.
Fig. 6  Longitudinal Schottky spectrum of a bunched beam.

Fig. 7  Time domain of representation of bunched beam Schottky signal.

(macroscopic signal resulting from the terms: \( \sum \tilde{J}_p (n\omega_0 \hat{t}_m) \)) and a fluctuating Schottky signal (\( \sum \sum J_p (n\omega_0 \hat{t}_m) \)). The fluctuating signal extends in time over \( 2\tau_m \) and can be Fourier decomposed into components at multiples of the fundamental bunch frequency \( f_b = 1/2\tau_m \). All information concerning the Schottky signal is contained into those components (in the limit \( \Omega_s \ll \omega_0 \)). In other words significant information about the Schottky signal only appears every \( f_b \) frequency interval, the other spectral lines in (19) (every \( f_o \)) simply give redundant information, i.e., they are correlated.

As a consequence, sampling of Schottky signals at \( f_o \), which folds many \( n f_o \) bands on top of each other and only gives one Schottky signal, does not introduce any loss of information, if the bandwidth before sampling is limited to \( \pm f_b/2 \).

1.4  Bunched beam. Transverse

Here we have to combine the amplitude modulation (betatron oscillation) and the time modulation (synchrotron oscillation). One obtains:
\[ d_1(t) = a_1 \cos(q_1 \omega_0 t + \psi_1) \left( \sum_{n=-\infty}^{\infty} \exp jn\omega_0 (t + \frac{t}{\tau_1} \sin(\Omega_S t + \psi_1)) \right) \]  

(22)

If \(q_1\) is independent of \(\omega_1\), the \(n\)th sum becomes:

\[ d_n = ef_0 \alpha_1 \left( \sum_{p=-\infty}^{\infty} J_{p}((n+q)\omega_0 \tau_1) \exp j((n+q)\omega_0 + p\Omega_S) t + pu_1 + \psi_1) \right) \]  

(23)

Again, each betatron line splits into an infinite number of synchrotron satellites (Fig. 8). The significant bandwidth, as in the longitudinal case, approaches that of coasting beams with the same \(\Delta p/p\), for large values of \(n\). On the contrary, for small values of \(n\), most of the energy is concentrated in the \(p = 0\) line.

![Decomposition of each betatron line into synchrotron satellites.](image)

For a non-zero chromaticity, the argument of the Bessel function \((n + q)\omega_0 \tau_1\) should be replaced by \([n + q - Q \Omega/n] \omega_0 \tau_1\). In this case, the relative amplitudes of the synchrotron satellites also depend on the chromaticity. In particular, for the chromatic frequency:

\[ \omega_\xi = \frac{Q \Omega}{n} \omega_0 \]  

(24)

only the term \(J_0\) is significant: all the energy of the \(n\)th Schottky band is concentrated in the central line.

With many particles, we should average over the two random variables \(\psi_1\) and \(\psi_1\). Unlike the longitudinal case, the central lines (\(\sigma=0\)) add up \(r.m.s\). wise due to the random betatron phase factor \(\psi_1\), the consequence being that there is no transverse macroscopic signal. Successive bands are correlated as in the longitudinal case, again, because all the signal is concentrated in the time interval \(2\tau_m\) and not \(T_0 = 1/f_0\) as if the beam were unbunched.

The width of the central line is determined by RF and magnetic field fluctuations, but also by transverse nonlinearities (tune spread due to octupole fields, beam-beam or
space charge forces). In addition, the synchrotron satellites are broadened by the spread in synchrotron frequencies within the bunch (width $p\Delta \Omega$ as in Fig. 6).

The total power per band (for a given $n$) is given by:

$$\langle d_n^2 \rangle = e^2 f_0^2 \langle a^2 \rangle \frac{N}{2} \sum_p \int_{-\infty}^{+\infty} J_p^2(x) \omega_{nq}^\perp \, dx$$

(25)

With the identity:

$$\sum_{p=-\infty}^{+\infty} J_p^2(x) = 1$$

(26)

one obtains:

$$\langle d_n^2 \rangle = e^2 f_0^2 \langle a^2 \rangle \frac{N}{2}$$

(27)

The total power per band is the same as in the coasting beam case, for the same total number of particles and the same transverse oscillation amplitude (Fig. 9).

![Horizontal Schottky signals in the SPS.
Top: debunched beam
Bottom: bunched beam.](image)

Fig. 9 Horizontal Schottky signals in the SPS.
Top: debunched beam
Bottom: bunched beam.

2 BEAM DETECTORS

2.1 The image-current approach

Consider the very simple geometry of Fig. 10a, where a round beam circulates in the center of a cylindrical smooth vacuum chamber. This is a two-dimensional problem, and it is well known that the electromagnetic fields are purely transverse, as in a coaxial line,
in the limit \( v = c \). It follows that for all frequencies the beam and wall currents are opposite:

\[
i_b = -i_w
\]  

(28)

Equation (28) is only valid up to some upper frequency, depending on the particle relativistic factor \( \gamma \) and the transverse dimensions of the vacuum chamber. However, for most practical cases (high energy storage rings) this is not a limitation.

If now we cut a gap in the circular wall we introduce a coupling between the inside and the outside of the vacuum pipe. The latter is characterised by the impedance \( Z \) which we can measure between the two sides of the gap. As the energy lost by the beam when passing through the detector is much smaller than the particle's energy, the current \( i_b \), and hence \( i_w \), is independent of the gap voltage: it means that the wall current \( i_w \) which flows through \( Z \) can be represented by a pure current source (Fig. 10b).

The detector, which seen from the gap appears like an impedance \( Z \), delivers its output signal in the load \( R_o \) (Fig. 10a). The sensitivity of the detector (longitudinal in this case) is defined by:

\[
S = \frac{V_{out}}{i_b} \frac{\text{in } R_o}{i_b}
\]  

(29)

For a lossless network between gap and \( R_o \), one can easily obtain, from power considerations:

\[
S = \sqrt{\frac{R_o \cdot R_e}{Z}}
\]

The following examples will illustrate the image current approach for the evaluation of beam detectors (or beam pick-ups).
a) The resistive-gap pick-up

In this case the load resistor $R_o$ is simply connected to the vacuum chamber gap. However, to provide a low impedance DC return path for the wall current, a short-circuited coaxial line is built around the vacuum chamber, as shown on Fig. 11. The line is filled with lossy material (ferrites) such that, for the operating frequency of the pick-up, it appears as a terminated line. This introduces a low-pass characteristic in the detector response.

![Diagram of resistive-gap pick-up]

Fig. 11  Resistive-gap pick-up

The upper frequency limit is determined by the parasitic capacitance at the gap. Making $R_o$ small (several parallel resistors) will push the upper frequency limit, at the expense of sensitivity.

The SPS wide-band longitudinal detector\(^3\) uses eight parallel 50 $\Omega$ strip lines symmetrically connected to the gap, and a ferrite loaded coaxial line with 25 $\Omega$ characteristic impedance. This arrangement gives $Z = 5$ $\Omega$. The eight gap signals are combined in an eight port power combiner giving an overall sensitivity, in a 50 $\Omega$ load:

$$S = 5\sqrt{8} = 14 \Omega$$

instead of the maximum $S = \sqrt{6.25 \times 50} = 17.6 \Omega$ if no power would be lost in the ferrites (very high impedance coaxial line).

The bandwidth extends from 4 MHz to 4 GHz with almost no resonances. To improve the low-frequency response the inductance of the short circuited line can be increased by lossless ferrites, but high-frequency resonances may be difficult to suppress.

b) The directional-coupler pick-up

As shown on Fig. 12a, there are two gaps in this detector, joined together by a piece of coaxial line of characteristic impedance $R_o$, surrounding the vacuum chamber. With the two load resistors $R_o$ which are connected to each gap, one can draw the equivalent
circuit of Fig. 12b. The two beam current sources, at each gap, are in opposite
direction, and are shifted in phase by the beam transit time.

The current flowing in the load $R_o$ on the right is the sum of the contributions
from the two current sources:

$$\frac{i_b}{2} \exp(-j\omega l / v_\psi) \quad \text{left source}$$

$$- \frac{i_b}{2} \exp(-j\omega l / v_p) \quad \text{right source}$$

$v_\psi$ and $v_p$ being the wave and beam velocities and $l$ the distance between gaps.

The total current:

$$\frac{i_b}{2} \left( \exp(-j\omega l / v_\psi) - \exp(-j\omega l / v_p) \right) \quad (30)$$

vanishes if $v_p$ and $v_\psi$ are equal.

For the load $R_o$ on the left, one finds easily the current:

$$\frac{i_b}{2} \left( 1 - \exp(-\omega \left( \frac{1}{v_\psi} + \frac{1}{v_b} \right) l) \right) \quad (31)$$

and the corresponding sensitivity:
for \( \nu_p = \nu = \nu \).

If this synchronous condition is fulfilled, for instance if \( \nu_p = \nu \) and the coaxial line is in vacuum, this detector is directional: the signal only appears at the upstream port (with respect to beam velocity). With counterrotating beams (\( p \) and \( p' \) for instance) the directional pick-up can separate the signals from the two types of particles. In practice the directivity is of the order of 30 to 35 dB. Note that directivity can, in principle, be obtained also by combining the signals of several identical detectors.

The sensitivity of the detector, given by Eq. (32) is frequency dependent (Fig. 13). It shows a succession of zeros and maxima corresponding to:

\[
\begin{align*}
\omega = & \frac{\lambda}{2}, \frac{3\lambda}{2} : & \text{zeros} \\
\omega = & \frac{\lambda}{4}, \frac{5\lambda}{4} : & \text{maxima}
\end{align*}
\]

the sensitivity being simply \( R_0 \) at the maxima.

![Fig. 13 Transfer functions of the directional-coupler pick-up.](image)

The transient response of the detector can be obtained by making the inverse Fourier transform of Eq. (32), but it is obvious from the equivalent circuit of Fig. 12b that it is composed of two opposite delta pulses separated in time by twice the transit time (\( 2\lambda/c \)) (Fig. 14a).

Several identical pick-ups can be combined to increase the overall sensitivity. With power combiners, the output signals are added power wise giving an overall sensitivity

\[ S = R_0 \sqrt{N} \quad \text{for identical detectors, and the same frequency response.} \]

One can also
combine several directional coupler detectors in cascade and obtain, with the proper delays, a transient response as in Fig. 14b. There the maximum sensitivity is proportional to \( \eta \), but the frequency response now shows a \( \sin f/t \) curve peaked at \( t = \lambda/4 \). In other words, the higher sensitivity (proportional to \( \eta \)) results in a narrower bandwidth.

\[ t \]
\[ 2 \frac{1}{\lambda} \]

\[ f \]

**Fig. 14** Transient response of directional coupler
a): single
b): multiple (with the associated frequency response).

Directional coupler pick-ups are in fact mostly used as transverse detectors. With several strips symmetrically arranged in the vacuum chamber, as in Fig. 15a, the total wall current \( i_w \) should be replaced by \( i_w \theta \omega/2\pi \) for each strip, provided the beam is in the center. For a non-centered beam the problem is truly three dimensional near the gaps. By approximating the electromagnetic field by that of a pure TEM wave one can obtain the wall current distribution along the vacuum chamber azimuth which obviously depends on the beam position. For small beam displacement, \( \Delta x \), the difference of the signals of two opposite strip lines is proportional to \( \Delta x \):

\[ \Delta V = V_2 - V_1 = S_A i_b \Delta x \]  \( \text{(33)} \)

\( S_A \) being defined by equation (33) as the transverse sensitivity of the detector (in ohms/meter).

\[ \theta \]
\[ \Delta x \]

**Fig. 15** Cross section of transverse directional-coupler pick-up
a): circular
b): rectangular
In the case of a rectangular geometry, often used in wide-aperture cooling rings for instance, the sensitivity $S_\Delta$ is given by:

$$S_\Delta = \frac{R_0}{2h} \left( \tanh \frac{\pi h}{4h} \right).$$

(34)

The form factor $\tanh(hw/h)$ simply reflects the fact that some fraction of the wall current flows outside the strip line gaps.

This type of pick-up (sometimes called loop coupler) is widely used in cooling systems. It offers a good compromise between bandwidth (of the order of one octave) and sensitivity. The signals of many couplers are often added power wise on a combiner board, inside vacuum, to increase the overall sensitivity. If only one type of particle is present, the downstream resistor $R_0$, where no current flows, can be replaced by a short circuit (hence the name of loop coupler), but microwave resonances may be harmful in this case.

c) The electrostatic pick-up

If the coaxial line of Fig. 12a is much shorter than the wavelength ($\ll \lambda$), it can be represented by a simple capacitor $C = \varepsilon / R_0$ (Fig. 16b). For a very high load resistor, the equivalent circuit of Fig. 16b represents the electrostatic detector, with the two current sources phase shifted by $\omega t / \nu_b$.

\[ \text{Fig. 16 The electrostatic pick-up} \]

The voltage developed on the line (or the electrode) is simply:

$$V = \frac{1}{jC\omega} \left[ i_b - i_b \exp \left( -j\omega \frac{t}{\nu_b} \right) \right]$$

(35)

$$V = \frac{i_b}{jC\omega} - \frac{i_b}{\nu_b} e^{j\omega t}.$$  \hspace{1cm} (36)

The quantity $\varepsilon i_b / \nu_b$ is the beam charge $q$ contained in the detector length, (assuming a slowly varying charge distribution with respect to the electrode length). It follows:
\[ V = \frac{q}{C} \]  

(37)

as the electrostatic theory would have given immediately.

For \( v = v_b \) and in the approximation of a high load resistor, Eq. (36) combined with \( C = \frac{k}{R_o} \) leads to the very simple result:

\[ S = R_o. \]

The sensitivity is independent of the frequency and of the length of the detector. Of course this is only true at medium frequencies. The non-infinite load resistor (usually an amplifier with high input impedance) introduces a low frequency cut off whereas at high frequencies the approximation \( k \ll \lambda \) is no longer valid.

The transverse version of the electrostatic pick-up can be obtained by splitting the electrode cylinder in two halves along a linear cut. (Fig. 17). Electrostatic theory shows that the difference in voltage between the two plates is a linear function of the beam displacement. Many versions of the transverse electrostatic pick-up with various shapes could be found in the literature (circular, rectangular, elliptical)\(^4\). They are mostly used for closed orbit measurements (sometimes horizontal and vertical pick-ups are combined in a single unit).

![Fig. 17 Transverse electrostatic pick-up linearly cut](image)

If the linearity requirement is less important, the linear cut could be abandoned, for instance in the so called "buttons" to be used in LEP (Fig. 18). There, only the high-frequency response is important, and consequently the load resistor is a 50 \( \Omega \) cable. The linearity can be restored by a proper algorithm at the signal processing level.

The electrostatic detector can be made resonant, with a coil (or transformer) connected to the electrode. A transverse version is sketched in Fig. 19a, with the equivalent circuit of Fig. 19b.

At resonance the voltage across the plates \( V' \) is given by:

\[ V' = j \omega \frac{k}{v_b} R \frac{2\Delta x}{d} \]  

(38)
Fig. 18  Cross section of the LEP position detector

Fig. 19a): Resonant electrostatic pick-up
b): Equivalent circuit

which is transformed in the output load into:

\[ V_{out} = \sqrt{R_o R} \frac{\omega}{\nu_b} \frac{2\Delta x}{d} i_b \]  \( (39) \)

for a lossless transformer.
Taking into account the ohmic losses of the coil \( Q_0 = \text{quality factor of the resonant circuit} \), \( Q_L = \text{loaded quality factor} \), one obtains:

\[
S_A = \frac{2}{d} \sqrt{\frac{R\omega Q_L}{C}} \frac{\hat{k}}{v_b} \left( 1 - \frac{Q_L}{Q_0} \right).
\]  (40)

This technique has been used in the CERN SPS, for a dedicated, very sensitive Schottky detector\(^5\) (sensitivity: 75Ω/mm).

### 2.2 Pick-up evaluation using the reciprocity theorem

The reciprocity theorem, well known in antenna theory, results from Maxwell equations applied to a linear, isotropic system. If we have two sets of current sources in the system \( J' \) and \( J'' \) which produce the electric fields \( E' \) and \( E'' \) and the magnetic fields \( H' \) and \( H'' \), the following relation is valid:

\[
\iint_S (E'' \times H' - E' \times H'') \cdot d\mathbf{s} = \iint_V (E'.J'' - E'' . J') \, d\mathbf{v}
\]  (41)

where the volume \( V \) is enclosed by the surface \( S \) (\( \hat{n} \) is the unity vector on that surface).

For the application of the reciprocity theorem, (Fig. 20), we take \( \hat{J}' = \hat{i}_b \), \( (\hat{i}_b \ \text{is the beam current along the detector axis}) \), and \( J'' = I_1 \) (\( I_1 \) is a pure current source applied across the load resistor \( R_0 \)).

We consider an integration volume limited by the metallic enclosure of the pick-up, where the electric fields are normal to the surface, which makes the left side of Eq. (41) vanish and leads to:

\[
\iiint_V \mathbf{E}_b \cdot \mathbf{i}_1 \, d\mathbf{v} = \iiint_V \mathbf{E}'. \mathbf{i}_b \, d\mathbf{v}
\]  (42)

\[
I_1 \cdot V_{\text{out}} = \int_z E_z i_b \, dz
\]  (43)

where \( V_{\text{out}} \) is the output voltage of the detector when excited by \( i_b \), and \( E_z \) is the on axis component of the field in the pick-up structure when excited by \( I_1 \). For a given geometry and a given field configuration, \( E_z \) can be related to \( I_1 \), from power considerations. Then application of Eq. (43) directly gives the detector sensitivity \( S_{V_{\text{out}}/i_b} \) for cases where the image current approach would fail (e.g. microwave structures).
Fig. 20 Application of the reciprocity theorem to a beam detector

Note that the reciprocity theorem, transposed in circuit theory, simply states that, for a passive quadrupole, the determinant of its transfer matrix is unity.

Application of the reciprocity theorem will be illustrated in the following by two examples: the slow-wave and the slot-line pick-ups.

The slow-wave pick-up is essentially an electromagnetic wave guide in which the phase velocity has been slowed down to match the velocity of the particles. Dielectric slabs (Fig. 21a) or corrugations (Fig. 21b) have been considered for this purpose. A description of the field in the structure will be given by standard wave guide theory. With respect to the transverse dimension, the \( E_z \) field configuration is either symmetrical (even mode) or antisymmetrical (odd mode), leading to a longitudinal or a transverse detector respectively.

Fig. 21 Slow-wave pick-ups  a) dielectric slab  b) corrugated wall

In the case of a pure travelling-wave structure, terminated at both ends by resistors \( R_o \) via matched transitions, the power flow \( P_o \) in the waveguide is related to \( I_1 \) by:
Note that only $I_1/2$ flows towards the waveguide, the rest being dissipated in the load resistor $R_0$.

The sensitivity is given by:

$$\int_{z} E_{z} i_b \exp(jk_0 z/B_p) \, dz = I_1 V_{out} \tag{45}$$

It is found to be proportional to the transit time factor:

$$\sin k_0 \left( \frac{\frac{1}{B_p} - \frac{1}{B_p'}}{2} \right) \frac{k}{B_p}$$

$$= k_0 \left( \frac{\frac{1}{B_p} - \frac{1}{B_p'}}{2} \right) \frac{k}{B_p'} \tag{46}$$

$k_0$, $B_p$, $B_p'$: propagation constants in free space, waveguide and beam respectively.

The sensitivity is optimum for $B_p$ and $B_p'$ (synchronism condition) as expected. For a given frequency, optimum dimensions of the waveguides are given by the synchronism condition (as in Fig. 14b). Making the detector longer increases the sensitivity (proportional to $k$) but reduces its bandwidth according to (46).

The slot-line pick-up$^{7,8}$ offers another interesting example, in which the waves propagate in a direction perpendicular to that of the beam (Fig. 22). A thin slot in a metallic plane on a dielectric substrate can support quasi TEM waves in the upper region. The electric field, not too close to the slot, is purely tangential: its amplitude is given by:

$$E_\phi = V_0 \frac{k_c}{2j} H^{(1)}_1(k_c r) \tag{47}$$
with:

$$k_c = \frac{2\pi j}{\lambda} \sqrt{\left(\frac{\lambda^2}{\lambda'}\right)^2 - 1}$$  \hspace{1cm} (48)

\(V_o\) is the voltage across the slot, \(\lambda'\) the wavelength along the slot and \(H_{11}\) the Hankel function of first order.

The longitudinal field \(E_z\), along the beam (at a distance \(d\) from the metallic plane) is simply:

$$E_z = \frac{d}{r} V_o$$  \hspace{1cm} (49)

For \(\lambda' - \lambda\) (\(c_r\) not too large) one can replace \(k_c H_{11}(k_c r)\) by \(2j \pi r\), which gives:

$$E_z = \frac{V_o d}{\pi r^2}$$  \hspace{1cm} (50)

\(V\) is related to the power flow \(P_o\) along the gap by the slot-line impedance \(Z\):

$$P_o = \frac{V^2}{2Z}.$$  \hspace{1cm} (combined with (44) and (45), one obtains:

$$S = \sqrt{\frac{R_o Z}{2}} \int_z^\infty \frac{1}{\pi r^2} \cos \frac{2\pi z}{\lambda} dz$$  \hspace{1cm} (51)

$$S = \sqrt{\frac{R_o Z}{2}} \exp \left(- \frac{2\pi d}{\lambda}\right).$$  \hspace{1cm} (52)

Equation (52) can be shown to be valid also even if \(\lambda\) and \(\lambda'\) are not very close: \(\lambda\) should then be replaced by \(\lambda'\) in (52).

It is interesting to remark that in the limit \(d \ll \lambda\), Eq. (52) reduces to

$$S = \sqrt{\frac{R_o Z}{2}}$$  \hspace{1cm} (which is the result given by the image current approach. With the reciprocity theorem, transverse propagation which was previously neglected can be taken into account.

If the signals of two symmetrical plates with two slots are combined, a transverse detector can be built. Its sensitivity would be:

$$S_A = \sqrt{\frac{R_o Z}{h}} \frac{1}{\sinh\left(\theta h/\lambda'\right)}$$  \hspace{1cm} (53)

\(h\) being the distance between plates.
Slot-line pick-ups would be interesting, because they can be easily produced by standard printed-circuit techniques, even in the microwave region. Their bandwidth is only limited by that of the slot-line to strip-line transitions (the wave on the slot is coupled to outside via a strip line deposited on the opposite side of the dielectric). Because of the transverse propagation, the inherent delay of the detector depends on the transverse beam position. This could be useful for some stochastic cooling schemes.

2.3 Impulse response

Consider again a travelling wave detector like, for instance, the corrugated wall waveguide, where a number of cells (or individual rectangular boxes) are coupled together via the beam pipe. When excited by a short beam pulse, the response of the detector is, in first approximation, an RF burst (Fig. 23) of amplitude $V_{\text{out}}$ and duration $\tau$. After the time $\tau$, all the energy deposited in the detector has been transported with the group velocity $v_g$ to the end of the structure and then to the terminating resistor $R_o$.

![Fig. 23 Impulse response of a travelling-wave detector](image)

What is the relation between the detector sensitivity $S$ and its output voltage $V_{\text{out}}$ in this case? If we assume a periodic train of short beam pulses (charge $q$), separated in time by $\tau$, the RF component of the beam current $i_b$ at the central frequency of the pick-up is simply $i_b = 2q/\tau$. Obviously the output amplitude $V_{\text{out}}$ is constant, with that particular beam input, which gives:

$$V_{\text{out}} = S i_b = 2S q/\tau$$  \hspace{1cm} (54)

for the longitudinal case and:

$$V_{\text{out}} = 2S \Delta x/\tau$$  \hspace{1cm} (55)

for the transverse case.

The energy $W$ deposited by the charge $q$ in the detector is related to the geometry of the structure via its "loss parameter" defined by:

$$W = k q^2$$  \hspace{1cm} (56)

The $k$ factor is also the $R/Q$ of the structure ($k = \frac{Q}{Q} R/Q$).
Combining (54), (56) and the relation:

$$W = \frac{1}{2} \frac{V_{\text{out}}}{R_0} \tau$$

valid for a lossless detector one finds:

$$S = \sqrt{\frac{E_{\text{f}}}{2}}$$

(58)

and a similar equation for the transverse case.

The loss parameter $k$ depends essentially upon the cell geometry, and can be calculated analytically in some simple cases (neglecting the effect of the beam hole) or evaluated by computer codes like SUPERFISH for instance. On the other hand $\tau$ characterizes the cell-to-cell coupling via $v_E$.

From Eq. (58), the maximum sensitivity is again proportional to $l$ (detector length) as both $k$ and $\tau$ are themselves proportional to $l$. Of course the bandwidth decreases correspondingly as was shown in the example of the multiple directional coupler (Fig. 14b). Note that this multiple directional coupler can be considered as a backward-travelling wave structure with $v_E = c$.

In the following example, we shall evaluate the $k$ factor for the simple geometry of Fig. 24: a chain of coupled cylindrical cavities. We consider the node $E_{011}$ (transverse detector) where the electric field is only longitudinal:

$$E_x = E_\phi = 0; \quad E_z = E_0 \frac{J(2\pi x/\lambda)}{\cos \phi}$$

(59)

The energy lost by charge $q$ is given by:

$$W = \frac{1}{2} q \int_{\mathbf{E}_z} \frac{E_z \exp \frac{\text{imuE}}{c}}{c} \, dz$$

(60)

The factor 1/2 simply reflects the fact that the charge $q$ only sees one half of its own induced voltage (fundamental theorem of beam loading). $W$ is also obtained by integrating $E_z^2$ over the whole cavity volume:

$$W = \frac{1}{2} \iiint \varepsilon_0 E_z^2 \, dv$$

(61)

Eliminating $E_0$ between (59), (60) and (61) finally gives:
Fig. 24  A chain of coupled cylindrical cavities as transverse detector

\[ W = \frac{\mu_0}{4\pi} \frac{k\omega^2}{2.38} \left( \frac{2\pi}{\lambda} \right)^2 \left( \frac{\sin(\pi L/\lambda)}{\pi L/\lambda} \right)^2 q^2 x^2. \]  

(62)

Equation (62) shows the interest of high-frequency detectors as far as sensitivity is concerned (factors \( \omega^2 \) and \( 1/\lambda^2 \)). But the influence of the beam hole which has been neglected in this simplified analysis will become more and more important. An example of this type of detector, using the first transverse mode of the accelerating cavities in the CERN SPS is given in Ref. 9.

3  OBSERVATION OF SCHOTTKY SIGNALS

3.1  Spectral analysis

As already mentioned in section 1 the measurement of the power spectral density of the Schottky signals gives the particle distribution in either momentum or betatron tune (or a combination of both). Therefore, spectral analysis is the natural technique for observing Schottky signals.

The frequency span of interest is of the order of the revolution frequency, or even less, (in most cases below 100 kHz). Consequently, the Fast Fourier Transform (FFT) or, more precisely, the Digital Fourier Transform (DFT) techniques which operate at low frequencies, can be used to evaluate in real time the signal spectrum. The Schottky band to be analysed must be translated at low frequency prior to FFT analysis, as in a conventional spectrum analyser. This may require a careful prefiltering to reject the unwanted image frequencies.

In the DFT technique, the signal is sampled and digitized at frequency \( f_s \). Each digital word is stored in a memory with \( M \) locations (typically \( 2^{10} = 1024 \) locations): the duration of the signal sample to be analysed is then \( T = M/f_s \). The frequency content (frequency span) of the sampled signal extends only up to \( f_s/2 \) (Nyquist theorem), and the resolution of the frequency analysis is of the order of \( 1/T \) (Fig. 25).
Fig. 25 Spectral analysis (DFT) of Schottky signals

Depending on the choice of the signal processing "windowing", the resolution varies a little: 1/T for the rectangular window; 1.4/T for the "Hamming window", better optimized for noise signals.

For a given resolution of the beam distribution measurement (in Δp/p, or ΔQ/Q), T is minimum for the largest width of the Schottky band. For instance, in the longitudinal, debunched beam case, one would minimize T by looking at the highest frequency Schottky bands (width nΔf) limited by either f_s/2, the detector sensitivity, or the overlap condition. This is of particular interest for the observation of "pseudo" Schottky signals in pulsed machines to measure the beam momentum spread during debunching. (T is there strictly limited by the duration of the magnetic cycle flat top). The beam develops, during debunching at high intensity, a very complicated structure which is more or less equivalent to random noise, but of macroscopic nature ("pseudo" Schottky signal). Its spectrum analysis provides an estimate of the momentum spread of the beam during debunching.

Even if T can be made very long, the result of the DFT on a noise signal does not give a good estimate of its spectral density. This is because the variance of the power measurement is comparable to its mean value: it does not decrease when T is made longer.

A better "estimation" of the true power density is obtained by averaging several spectra taken at different time intervals. The "degree of confidence" of the measurement increases with the number of averaged spectra (Fig. 26), at the expense of the total analysis time (which may be distributed over several machine cycles in the the previous example).
3.2 Parasitic signals of the Schottky spectrum

Due to the very low level of the Schottky signals, many sources of disturbance can be harmful and should be eliminated whenever possible.

Parasitic signals may come from the beam itself; if there is a coherent excitation (i.e. transverse) it will appear as a betatron signal, but with an amplitude proportional to $N$ and not $\sqrt{N}$ as for the Schottky signal. The longitudinal line in a transverse Schottky scan can be suppressed by careful centering of the pick-up on the beam axis. In the bunched beam case, additional sharp filtering with a crystal filter is necessary.  

The line related components are reduced by a careful design of the amplifiers, power supplies and earth connections. If this is not sufficient, narrow band synchronous filtering (locked to the mains frequency) can also be employed.

Of more fundamental nature is the disturbance due to the thermal noise of the first preamplifier, after the detector. The amplifier is characterised by its noise factor $F$ (excess noise with respect to a simple resistor $R_0$). The available noise spectral density resulting from the amplifier and which is given by: $F k T R_0$ , ($k$ : Boltzmann constant, $T$ : temperature) must be considerably smaller than the Schottky noise spectral density. One possibility is to cool the preamplifier and the terminating resistors of the
pick-ups (ACOL). From Fig. 2, it is clear that the best signal to noise ratio is obtained at low frequencies (for the case of unbunched beams). Unfortunately, observation of Schottky signals at high frequency is more favourable as far as sensitivity and analysis time are concerned.

3.3 Bunched-beams signal processing

As all Schottky lines in a frequency interval $f_b$ are correlated (see section 1), it is interesting to sample the beam signal at the revolution frequency. All lines will be folded in the base band giving a much better signal to noise ratio as will be shown in the following.

Consider the RF burst amplitude $V_{out}$ of Fig. 23 delivered by a travelling wave transverse pick-up, when excited by a short bunch. For a transverse r.m.s. beam displacement $x$, the output voltage of the detector and the thermal noise voltage of the amplifier, referred to the input, are respectively:

$$V_{rms} = 2S_0^2 \frac{x}{f}$$

$$V_{th} = Fk_t \frac{BB}{oo}$$

(63)

(64)

$B$ being the amplifier bandwidth.

The power signal to noise ratio, during the time interval $\tau$ is therefore:

$$\frac{1}{U} = \frac{2Nf_o S_0^2 x^2}{F \tau^2 k_t B}$$

(65)

This is also the signal to noise ratio after sampling. We can select $B$ ($B = B_{opt}$) to optimize $1/U$. $B_{opt}$ is the minimum bandwidth for which the useful signal is not reduced significantly. This happens if the rise time of the band limited RF burst is of the order of its length: $1/B = \tau$, as illustrated in Fig. 27. More precisely $B_{opt}$ is that of the so called "optimum filter" (radar terminology) for which the impulse response is the time reversed image of the RF burst. With that condition (65) becomes:

$$\frac{1}{U} = \frac{1}{\tau f_o N} \frac{e^2 f_o S_0^2 x^2}{F k_t B}$$

(66)

which is the same as for the debunched beam case, except for the enhancement factor $1/\tau f_o$ which can be much larger than unity.
The overall signal processing system for a bunched beam transverse Schottky signal is displayed on Fig. 28. Frequency translation down to the base band frequency can be done by peak detection, as indicated, or with a synchronous detector driven by the sum signal of the pick-up. In this case, it is interesting to remark that the odd synchroton satellites are rejected for an in phase detection (like for a peak detection), whereas for a quadrature detection, it is the even synchrotron satellites which are rejected. This feature may be useful if one wants to isolate the central $J_0$ line of the Schottky band.

Fig. 27 Optimum filtering of an RF burst from a beam detector

Fig. 28 Bunched beam signal processing system

Although the thermal noise of the preamplifier is of less importance for bunched beam signal processing, the effect of spurious coherent excitation of the beam may be more of a problem. This is because, even a low frequency excitation, near the first betatron line, appears everywhere in the spectrum, contrary to the debunched beam case, and may spoil even a high frequency Schottky system. A solution to that problem is to reject that part of the detector signal which is coherent from one bunch to the next.\(^9\)}
4 BEAM TRANSFER FUNCTIONS

4.1 Principle of beam transfer functions

The name of beam transfer function almost speaks for itself: it relates the response
of the beam (amplitude and phase) to a known excitation. In the case of a transverse
excitation by a deflector (or kicker), the beam response is measured by a transverse
pick-up as indicated on Fig. 29a, whereas Fig. 29b shows the arrangement for the
measurement of a longitudinal transfer function.\(^1\)

\[
\text{Transverse} \quad \begin{array}{c}
\text{kicker} \\
\text{Pu} \\
\text{excitation} \\
\text{Network} \\
\text{Analysers}
\end{array}
\quad \begin{array}{c}
\text{Longitudinal} \\
\text{Cavity} \\
\text{Pu} \\
\text{excitation} \\
\text{Network} \\
\text{Analysers}
\end{array}
\]

\(a\) \quad \text{b)}

Fig. 29 Principle of beam transfer function measurement

To minimize the analysis time and the disturbance to the beam, it is interesting to
excite the beam with a white noise spectrum (all frequencies are present in the band of
interest). There the output will also be a noise signal, similar to the Schottky noise,
and for which similar processing techniques can be applied. To extract the phase
information spectral density measurements are not sufficient and a dual channel DFT
instrument is needed. Again averaging many transfer functions reduces the variance of the
estimate (Fig. 26). In Fig. 29a and b, a new element appears, namely the kicker (either
transverse or longitudinal) which will be examined more in detail in the following.

4.2 Kickers

A longitudinal kicker is a fairly straightforward device in which a longitudinal
electric field \(E_z\) is produced. The particle gains an energy \(\Delta W\), when crossing the
kicker (or cavity), which is simply given by:

\[
\Delta W = \int e E_z \, dz. \quad (67)
\]

The application of the reciprocity theorem to a longitudinal beam detector has led us to
Eq. (43), which combined with (67) results in:

\[
\Delta W = e S I_1, \quad (68)
\]
showing that the energy gain of the kicker and the sensitivity of the pick-up are simply proportional. In other words, a longitudinal kicker is nothing but a longitudinal detector working in reverse. This is almost obvious for cavity-like detectors, but is also true for a directional coupler type of pick-up, for instance, where a quasi TEM wave propagates. There, only the field at the ends of the coupler are useful for beam excitation.

Fig. 30 Application of the induction law to the evaluation of the transverse force

Consider now the case of a transverse deflection produced by the Lorentz force:

$$\Delta p = \int_{z} e (R + v \times B) \, dz$$  \hspace{1cm} (69)$$

which projected on the x-axis (Fig. 30) can be written:

$$\Delta p_x = \int_{z} e (E_x + v.B_y) \, dz$$  \hspace{1cm} (70)$$

To evaluate the quantity $E_x + v.B_y$, we apply the induction law to a small rectangle in the xo$\bar{z}$ plane:

$$\oint_C \mathbf{E} \cdot d\mathbf{r} = -\frac{\partial}{\partial t} \Phi_B$$  \hspace{1cm} (71)$$

$\Phi_B$ being the flux of the magnetic field $B$ on the contour $C$. One obtains

$$\frac{dE_x}{dx} \frac{dx}{dz} + \frac{dE_x}{dz} \frac{dz}{dx} = -j\omega B_y \frac{dx}{dz}$$  \hspace{1cm} (72)$$

$$\frac{dE_x}{dx} + \frac{dE_x}{dz} = -j\omega B_y$$  \hspace{1cm} (73)$$

With:
\[ \frac{dE_x}{dz} = \frac{dE_z}{dz} + j\omega E_x \frac{1}{\nu} \]  
\hspace{1cm} (74) 

it comes:

\[ E_x + v \cdot B_y = \frac{\nu}{j\omega} \frac{dE_z}{dx} \]  
\hspace{1cm} (75) 

Equation (75) shows that only the longitudinal field \( E_z \) (more precisely \( dE_z/dx \)) is important for transverse deflection. This is a well known result (linac theory for instance) which has a few interesting corollaries. For instance, one cannot deflect a beam neither with a pure TEM wave nor with a pure H mode in a cavity if the end effects are neglected. A transverse kicker must show a longitudinal electric field, in the same way as a transverse pick-up extracts energy from the longitudinal velocity of the particles. There is complete equivalence between pick-ups and kickers even in the transverse plane. This will be illustrated in the following example.

The "TEM" travelling wave kicker has the same geometry as the transverse directional coupler pick-up (Fig. 31a). The field is that of a TEM wave along the two lines, except at the two ends where a longitudinal component \( E_z \) exists (Fig. 31b). Assume, for simplicity \( \nu = \nu' \); the particles receive successively two opposite transverse kicks at either end of the kicker, the result being a zero deflection (another way of saying the same thing is that the electric and magnetic deflections along the line exactly cancel each other). On the contrary, for \( \nu' = -\nu \) (beam in the opposite direction) the two kicks add exactly if they are separated by half a period of the RF wave (\( t = \lambda/4 \)). This gives a variation of the type \( \sin 2\pi \frac{1}{\lambda} \). The kicker efficiency \( K_z \) is, from Eq. (75) proportional to:

\[ K_z \sim e^{\frac{\nu}{j\omega} \frac{dE_z}{dx} \sin 2\pi \frac{1}{\lambda}} \]  
\hspace{1cm} (76) 

Fig. 31 "TEM" travelling wave kicker
which can be written:

\[ K_1 \sim \frac{\alpha}{h} \frac{\sin \frac{2\pi t}{\lambda}}{2\pi t/\lambda} \]  

(77)

for \( v = c \).

The first term is proportional to the DC deflection (proportional to \( \alpha \)) and the term in brackets gives the form factor which is frequency dependent.

### 4.3 Debunched beam transfer function

The beam is composed of a collection of particles, each having its own oscillation frequency \( q_1 \omega_1 \), submitted to a common driving force \( F(\omega) \). The equation of motion, for each individual particle is, in linear approximation:

\[ x_1'' + (q_1 \omega_1)^2 x = F(\omega) \]  

(78)

with a forced solution of the form:

\[ x_1 = X_1 \exp j\omega t \]

\[ X_1 = \frac{F(\omega)}{(q_1 \omega_1)^2 - \omega^2} = \frac{1}{2\omega} \frac{F(\omega)}{q_1 \omega_1 - \omega} \]  

(79)

The average beam response \( \langle X_1 \rangle / F(\omega) \) is given by the integral:

\[ \frac{\langle X_1 \rangle}{F(\omega)} = \frac{1}{2\omega} \int_{q_1 \omega_1 - \omega}^{q_1 \omega_1 + \omega} \frac{\rho(q_1 \omega_1)}{q_1 \omega_1 - \omega} d(q_1 \omega_1) \]  

(80)

where \( \rho(q_1 \omega_1) \) is the normalized distribution of the betatron frequencies within the beam, \( (q_1 \omega_1 \text{ and } q_2 \omega_2 \text{ being the two extreme frequencies}).

This is a singular integral because of the pole at \( q_1 \omega_1 = \omega \). It can be decomposed into its Cauchy principal value, which is real, and its residue at the pole (imaginary):

\[ \frac{\langle X_1 \rangle}{F(\omega)} = \frac{1}{2\omega} \left[ \text{Princ. Value} \ - j\pi \rho(\omega) \right] \]  

(81)

We now replace \( \langle X_1 \rangle \) by \( j\omega \langle X_1 \rangle \) to obtain a real transfer function \( B(\omega) \) when energy is absorbed (force and displacement in quadrature) and obtain:
\[ B(\omega) = \frac{1}{2} (\pi \rho(\omega) + j \text{ Princ. Value}) \]  

The real part of the transfer function gives the particle distribution in tune like the spectral power density of the Schottky signal. Outside the frequency band \((q_1 \omega, q_2 \omega)\) the real part of \(B(\omega)\) vanishes (pure imaginary response). The fact that a collection of lossless oscillators responds like a damped resonator is the basis of Landau damping and is illustrated in Fig. 32.

![Diagram showing phase and amplitude vs frequency](image)

**Fig. 32** Response of a large number of lossless resonators

--- individual particles  

average

The evaluation of the stability of the beam certainly corresponds to the most interesting application of beam transfer function measurements. Collective effects (and in particular beam instabilities) result from the presence of parasitic impedances in the machine which generate a deflecting force (in the transverse case), when excited by a collective displacement of the beam. In other words the excitation \(F(\omega)\) in Eq. (78) should be combined with a term proportional to the beam response \(j\omega X_1\). This leads to the well known feedback loop of Fig. 33, where \(H(j\omega)\) is linked to machine parameters and is proportional to the impedance of the machine \(Z(\omega)\). For instance in the transverse case:

\[ H(j\omega) = \frac{e \omega l_b}{j 2\omega m c_y} Z(\omega) \]  

\(m_0\) is the rest mass of the particle.
Fig. 33 Feedback loop due to the machine impedance

From Fig. 33 the new transfer function becomes:

\[ B'(\omega) = \frac{j\omega \langle X_L \rangle}{F} = \frac{B(\omega)}{1 - B(\omega)H(\omega)} \quad (84) \]

\[ \frac{1}{B'(\omega)} = \frac{1}{B(\omega)} - H(\omega) \quad (85) \]

By plotting the curve \(1/B(\omega)\) for different beam intensities \(i_B\) one obtains a family of curves shifted in the complex plane by the quantity \(H(\omega)\) (Fig. 34). This shift being proportional to \(Z(\omega)\), the machine impedance can be directly measured at any frequency.\(^{11}\)

When the shifted \(1/B(\omega)\) curve reaches the complex plane origin, stability of the beam is lost (\(B(\omega) \to \infty\)), this means that the distance of the curve to the origin is a measure of beam stability. If a feedback system is employed to stabilize the beam, its effect which should be to shift the curve towards the right side of the complex plane could also be evaluated.

Fig. 34 Evaluation of the beam stability with transfer function measurements
With very sensitive detectors and provided long analysis times are available (DC storage rings), beam transfer function is a very powerful technique, almost non-disturbing to the beam; it can also be used in a similar way for the longitudinal plane.

4.4 Bunched-beam transfer function

The main difference with respect to the unbunched beam case is that an excitation of the beam at a given frequency \( \omega \), not only results in a beam response at \( \omega \), but also at all frequencies \( n\omega_0 \pm \omega \). (This is because the bunched beam samples the \( \omega \) waveform at the revolution frequency \( \omega_0 \)). The process is therefore fundamentally nonlinear, and as a consequence, the beam transfer function is not defined in general, unless additional conditions are imposed\(^\text{12}\). For instance, if bunch to bunch coupling can be neglected, one can define unambiguously the beam transfer function of a single bunch, for a given mode of oscillation (dipole, quadrupole etc.), i.e. within an \( f_b \) frequency interval. Another interesting case is when the bunched beam behaves like an unbunched beam: many equal bunches, frequency range from DC up to \( f_{RF}/2 \) and negligible effects beyond.

In the transverse plane, the measurement of the machine tune is nothing but a beam transfer function measurement. Many descriptions of tune measurement systems exist in the literature; excitation can be sinusoidal or random (band limited noise) near a betatron line, or pulsed; beam measurement could be at the same or at different frequency. In general the machine impedance \( Z(\omega) \) cannot be measured directly, as a function of frequency; on the other hand if the shape of \( Z(\omega) \) is known (e.g. resistive wall) one can determine its magnitude by measuring the tune shift as a function of beam intensity.

The RF system and its associated feedback loops strongly perturbs the longitudinal transfer function of a bunched beam. This is particularly true for the dipole mode; fortunately the quadrupole mode is easier to analyse and can provide meaningful measurements of the machine impedance. Amplitude modulation of the RF waveform at around twice the synchrotron frequency excites the quadrupole mode of a single bunch; the quadrupole oscillation can be observed in a very simple way by peak detecting the bunch signal from a wide band longitudinal detector.

The measured beam transfer function, at low intensity shows a sharp phase discontinuity, at the bunch center, where the particle density is maximum, and a smooth phase curve near the bunch edge (Fig. 35a). This corresponds to the \( 1/B(\omega) \) plot in Fig.35b and provides a direct measurement of the center synchrotron frequency. At higher intensities, the inductive wall effect shifts the \( 1/B(\omega) \) curve along the imaginary axis (real frequency shift) and the phase curve of Fig. 35a shows a sharper transition. From those measurements, the magnitude of \( Z(\omega)/n \) for the inductive wall case can be determined over a frequency interval of the order of \( f_b \).
Fig. 35  Quadrupole-mode transfer function (bunched beam)

REFERENCES


3)  T. Linneccar. The high frequency longitudinal and transverse pick-ups used in the SPS. CERN SPS/ARF/78-17, (1978).


STOCHASTIC COOLING

D. Møhl
CERN, Geneva, Switzerland.

ABSTRACT
This paper describes the main analytical approaches to stochastic cooling. The first is the time domain picture in which the beam is rapidly sampled and a statistical analysis is used to describe the cooling behaviour. The second is the frequency domain picture, which is particularly useful since the observations made on the beam are mainly in this domain. This second picture is developed in detail to assess ingredients of modern cooling theory like mixing and signal shielding and to illustrate some of the diagnostic methods. Finally the use of a distribution function and the Fokker-Planck equation are discussed, which give the most complete description of the beam during the cooling.

1. INTRODUCTION

Beam cooling aims at reducing the size and the energy spread of a particle beam circulating in a storage ring. This reduction of size should not be accompanied by beam loss; thus the goal is to increase the particle density.

Since the beam size varies with the focusing properties of the storage ring, it is useful to introduce normalized measures of size and density. Such quantities are the (horizontal, vertical and longitudinal) emittances and the phase-space density. For our present purpose they may be regarded as the (squares of the) horizontal and vertical beam diameters, the energy spread, and the density, normalized by the focusing strength and the size of the ring to make them independent of the storage ring properties.

Phase-space density is then a general figure of merit of a particle beam, and cooling improves this figure of merit.

The terms beam temperature and beam cooling have been taken over from the kinetic theory of gases. Imagine a beam of particles going around in a storage ring. Particles will oscillate around the beam centre in much the same way that particles of a hot gas bounce back and forth between the walls of a container. The larger the mean square of the velocity of these oscillations in a beam the larger the beam size. The mean square velocity spread is used to define the beam temperature in analogy to the temperature of the gas which is determined by the kinetic energy $0.5 \text{mv}^2_{\text{TMS}}$ of the molecules.

Why do we want beam cooling? The resultant increase of beam quality is very desirable for at least three reasons:

i) Accumulation of rare particles

Cooling to make space available so that more beam can be stacked into the same storage ring. The Antiproton Accumulator (AA) at CERN is an example of this (see Fig. 1).
Fig. 1 The CERN Antiproton Accumulator (AA). Sketch and table of performance with number of particles, horizontal emittance, vertical emittance and momentum spread of incoming beam and of stack after 24 h of accumulation (design values).

ii) Improvement of interaction rate and resolution

Cooling to provide sharply collimated and highly mono-energetic beams for precision experiments with colliding beams or beams interacting with fixed targets. The Low Energy Antiproton Ring (LEAR) at CERN is an example of this (see Fig. 2).

iii) Preservation of beam quality

Cooling to compensate for various mechanisms leading to growth of beam size and/or loss of stored beam. Again LEAR is an example of this application.

Several cooling techniques are operative or have been discussed\(^1\)): Electron beams have a tendency to cool 'by themselves' owing to the emission of radiation as the orbit is curved. The energy radiated decreases very strongly with increasing rest mass of the particles. For (anti-)-protons and heavier particles, radiation damping is negligible at energies currently accessible in accelerators. 'Artificial' damping had therefore to be devised, and two such methods have been successfully put to work during the last decade: i) cooling of heavier particles by the use of an electron beam -- this is the subject of H. Poth's chapter in these proceedings; and ii) stochastic cooling by the use of a feedback system, which will be discussed later in this chapter.
(a) Momentum cooling at injection in LEAR; $\sqrt{\epsilon N/\epsilon p}$ displayed against momentum; $3 \times 10^9$ antiprotons, before and after 3 minutes of cooling. $\sigma_p/p$ is reduced by a factor 4.

(b) Comparison of the cooled beam extracted from LEAR to the low energy antiproton beams previously obtained in secondary beam lines

- Cooled beam from LEAR.
  - Several $10^5$ antiprotons/s
  - Typical beam size $5 \times 5$ mm$^2$

- 300 MeV/c beam from production target:
  - 200 antiprotons/s + several $10^4$ contaminants
  - Beam size $4 \times 4$ cm$^2$

Fig. 2 An example of momentum spread cooling and properties of the cooled beam from the CERN Low Energy Antiproton Ring (LEAR) compared to a secondary beam used before 1980

2. SIMPLIFIED THEORY, TIME-DOMAIN PICTURE

2.1 The basic set-up

The arrangement for cooling of the horizontal beam size is sketched in Fig. 3. Assume, for the moment, that there is only one particle circulating. Unavoidably, it will have been injected with some small error in position and angle with respect to the ideal orbit (centre of the vacuum chamber). As the focusing system continuously tries to restore the resultant deviation, the particle oscillates around the ideal orbit. Details of these 'betatron oscillations' are given by the focusing structure of the storage ring, namely by the distribution of quadrupoles and gradient magnets (and higher-order 'magnetic lenses') which provide a focusing force proportional to the particle deviation (and to higher-order powers of the deviation).

For the present purpose, we can approximate the betatron oscillation by a purely sinusoidal motion. The cooling system is designed to damp this oscillation. A pick-up electrode senses the horizontal position of the particle on each traversal. The error signal -- ideally a short pulse
with a height proportional to the particle's deviation at the pick-up -- is amplified in a broad-band amplifier and applied on a kicker which deflects the particle by an angle proportional to its error.

In the simplest case, the pick-up \(^3\) consists of a plate to the left of the beam and a plate to the right of it. If the particle passes to the left, the current induced on the left plate exceeds the current on the right one and vice versa. The difference between the two signals is a measure of the position error. The 'kicker' is, in principle, a similar arrangement of plates on which a transverse electromagnetic field is created which deflects the particle \(^3\).

![Diagram of beam, pick-up, amplifier, and kicker](image)

**Fig. 3** The principle of (horizontal) stochastic cooling. The pick-up measures horizontal deviation and the kicker corrects angular error. They are spaced by a quarter of the betatron wavelength \(\lambda_B\) (plus multiples of \(\lambda_B/2\)). A position error at the pick-up transforms into an error of angle at the kicker, which is corrected.

Since the pick-up detects the position and the kicker corrects the angle, their separation is chosen to correspond to a quarter of the betatron oscillation (plus an integer number of half wavelengths if more distance is necessary). A particle passing the pick-up at the crest of its oscillation will then cross the kicker with zero position error but with an angle which is proportional to its displacement at the pick-up. If the kicker corrects just this angle the particle will from thereon move on the nominal orbit. This is the most favourable situation (sketched as Case 1 in Figs. 4 and 5). A particle not crossing the pick-up at the crest of its oscillations will receive only a partial correction (Cases 2 and 3 in Figs. 4 and 5). As we shall see later, it will then take several passages to eliminate the oscillation.

![Graphs showing betatron phase](image)

**Fig. 4** The importance of betatron phase: Particle 1 crosses the pick-up with maximum displacement. Its oscillation is (ideally) completely cancelled at the kicker. Particle 2 arrives at an intermediate phase; its oscillation is only partly eliminated. Particle 3 arrives with the most unfavourable phase and is not affected by the system.
Fig. 5 Phase space representation of betatron cooling. The same as for Fig. 4 except that a 'polar diagram' $x' = f(x)$ is used to represent the betatron motion $x = x \sin [Q(s/R) + \psi_0]$, $x' = (R/Q) x' = \chi \cos [Q(s/R) + \psi_0]$. The undisturbed motion of a particle is given by a circle with the radius equal to the betatron amplitude $\chi$. Kicks correspond to a jump of $x'$. The cooling system tries to put particles onto smaller circles. Particles 1, 2 and 3 are sketched with the most favourable, the intermediate, and the least favourable initial phase, respectively. As the number of oscillations per turn is different from an integer or half-integer, particles come back with different phases on subsequent turns and all particles will be cooled progressively.

Another particularity of stochastic cooling is easily understood from the single particle model (Fig. 3): the correction signal has to arrive at the kicker at the same time as the test particle. Since the signal is delayed in the cables and the amplifier, whereas a high-energy particle moves at a speed close to the velocity of light, the cooling path has usually to take a short cut across the ring. Only at low and medium energy ($v/c < 0.5$) is a parallel path feasible.

We have thus familiarized ourselves with two constraints on the distance pick-up to kicker: taken along the beam, this distance is fixed, or rather quantized, owing to the required phase relationship of the betatron oscillation; taken along the cooling path this length is fixed by the required synchronism between particle and signal. A change of energy (particle velocity) and/or a change of the betatron wavelength will therefore require special measures. Incidentally, the first of these two conditions is due to the oscillatory nature of the betatron motion. For momentum spread cooling in a coasting beam, where the momentum deviation of a particle is constant rather than oscillatory, this constraint does not come into play and a greater freedom in the choice of pick-up-to-kicker distance exists.

It is now time to leave the one-particle consideration and turn our attention to a beam of particles which oscillate incoherently i.e. with different amplitudes and with random, initial phase.

By beam cooling we shall now mean a reduction with time of the amplitude of each individual particle. To understand stochastic cooling, we will next have a closer look at the response of the cooling system. This permits us to discern groups of particles -- so-called samples -- which will receive the same correcting kick during a passage through the system.
2.2 Notion of beam samples

To be able to analyze the response of the cooling system, let us start with an excursion into elementary pulse and filtering theory\(^1\). What we would like to take over is a bandwidth/pulse-length relation known as the Küpfmüller or Nyquist theorem\(^3\):

If a signal has a Fourier decomposition of band-width \(\Delta f = W\), then its ‘typical’ time duration will be

\[ T_s = \frac{1}{2W} \]

This is illustrated in Fig. 6, where we sketch the Fourier spectrum of a pulse and the resulting time-domain signal. Clearly the two representations are linked by a Fourier transformation, and this permits us to check the theorem.

For curiosity, note the difference between a pulse with a low-frequency and a high frequency spectrum (both cases are sketched in Fig. 6). In spite of the different shape of the time-domain signal, the ‘typical duration’ is in both cases \(1/(2W)\).

*Fig. 6 Illustration of the Küpfmüller-Nyquist relation: a signal whose Fourier decomposition \(S(f)\) has a bandwidth \(W\), has a typical time duration \(T_s = 1/(2W)\). Illustration for a 'low-pass' [case (a)] and a 'band-pass' signal [case (b)].

---

\(^1\) The bandwidth/pulse length relation was introduced by Nyquist and independently by Küpfmüller in 1928. This theorem is closely-related to the more general sampling theorem of communication theory: If a function \(S(t)\) contains no frequencies higher than \(W\) cycles per second, it is completely described by its value \(S(mT_s)\) at sampling points spaced by \(\Delta t = T_s = 1/(2W)\) (i.e. taken at the 'Nyquist rate' \(2W\); see, for example, J.A. Betts, Signal Processing and Noise (English Universities Press, London, 1970).
A corollary to the theorem is well known to people who design systems for transmitting short pulses:

When a short pulse is filtered by a low-pass or band-pass filter of bandwidth \( W \), the resulting pulse has a 'typical' time width (see Fig. 7)

\[
T_s = 1/(2W).
\]  
(2.1)

Fig. 7 Input and output signal \( S(t) \) of a low-pass system and 'rectangular' approximation to the output pulse \( S_q(t) \)

In this form, the theorem is directly applicable to our cooling problem, to which we now return. Passing through the pick-up, an off-axis particle induces a short pulse with a length given by the transit time. Owing to the finite bandwidth \( (W) \) of the cooling system, the corresponding kicker signal is broadened into a pulse of length \( T_s \). To simplify considerations, we approximate the kicker pulse by a rectangular pulse of total length \( T_s \) (Fig. 8).

A test particle passing the system at \( t_0 \) will then be affected by the kicks due to all particles passing during the time interval \( t_0 \pm T_s/2 \). These particles are said to belong to the sample of the test particle. In a uniform beam of length \( T \) (revolution time), there are \( N_s = T/T_s = 2WT \) equally spaced samples of length \( T_s \) with

\[
N_s = N/(2WT) \text{ particles per sample.}
\]  
(2.2)

Fig. 8 Pick-up signal of a particle and corresponding kicker pulse (idealized). The test particle experiences the kicks of all other particles passing within time \( -T_s/2 < \Delta t < T_s/2 \) of its arrival at the kicker. These particles are said to belong to the sample of the test particle. Cooling may be discussed in terms of the centre-of-gravity motion of samples.
Table 1
An example of samples corresponding to cooling at injection in LEAR\(^3\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of particles in the beam</td>
<td>(N \times 10^9)</td>
</tr>
<tr>
<td>Revolution time</td>
<td>(T \times 0.5 \mu s)</td>
</tr>
<tr>
<td>Transit time in one pick-up unit</td>
<td>(T_t \times 0.1 \text{ ns})</td>
</tr>
<tr>
<td>Cooling system bandwidth</td>
<td>(W \times 250 \text{ MHz})</td>
</tr>
<tr>
<td>Sample length</td>
<td>(T_s \times 2 \text{ ns})</td>
</tr>
<tr>
<td>No. of samples per turn</td>
<td>(I_s \times T/T_s \times 250)</td>
</tr>
<tr>
<td>No. of particles per sample</td>
<td>(N_s \times 4 \times 10^6)</td>
</tr>
</tbody>
</table>

### 2.3 Coherent and incoherent effects

The model of samples has allowed us to subdivide the beam into a large number of slices which are treated independently of each other by the cooling system. If the bandwidth can be made large enough so that there are no other particles in the sample of the test particle, then the single-particle analysis is still valid. However, to account for the reality of some million particles per sample, we have to go one step further and do some simple algebra. This will permit us to discern two slightly different pictures of the cooling process. In the 'test particle picture' we shall view cooling as the competition between: i) the 'coherent effect' of the test particle upon itself via the cooling loop; and ii) the 'incoherent effect', i.e. the disturbance to the test particle by the other sample members (see Fig. 9). In the 'sampling picture' we shall understand stochastic cooling as a process where samples are taken from the beam at a rate \(I_s\) per turn. By measuring and reducing the average sample error, the error of each individual particle will (on the average) slowly decrease.

A few simple equations will illustrate these pictures. Let us denote by \(x\) the error of the test particle and assume that the corresponding correction at the kicker is proportional to \(x\), say \(\lambda x\). With no other particles present, the error would be changed from \(x\) to a corrected

\[
x_c = x - \lambda x
\]  

(2.3)

i.e. the test particle receives a correcting kick,

\[
\Delta x = \lambda x
\]  

(2.4)

In reality the kicks \(-\lambda x_i\) of the other sample members have to be added, and the corrected error after one turn and the corresponding kick are
Fig. 9 Cooling system signals for the test particle picture. Signals at the instant of passage of the test particle are sketched. The upper trace gives the coherent correction signal due to the test-particle itself. The lower trace sketches the incoherent signal due to the other particles in the sample. The kick experienced is the sum of coherent and incoherent effects. If the amplification is not too strong and the sample population is small, the coherent effect which is systematic will predominate over the random heating by the incoherent signals.

\[
x_c = x - \lambda x - \sum_{s} \lambda_s x_s
\]

\[
\Delta x = -\lambda x - \sum_{s'} \lambda_{s'} x_{s'}
\]  

Equations (2.5) and (2.6) clearly exhibit the 'coherent' and the 'incoherent' effects mentioned above. The sum labelled s' includes all particles in the sample except the test particle. You may want to rewrite this sum including the test particle (this sum will be labelled s) and interpret it in terms of the average sample error (the sample centre of gravity if you like), which is by definition
\[ \langle x \rangle_s = \frac{1}{N_s} \sum_i^n x_i. \]  
(2.7)

Equations (2.5) and (2.6) then become

\[ x_c = x - (\lambda N_s) \langle x \rangle_s \equiv x - g \langle x \rangle_s \]  
(2.8a)

\[ \Delta x = - (\lambda N_s) \langle x \rangle_s \equiv -g \langle x \rangle_s. \]  
(2.8b)

This introduces the second picture. What the cooling system does is to measure the average sample error and to apply a correcting kick, proportional to \( \langle x \rangle_s \) to the test particle. Up to now the sample is defined with respect to a specific test particle; however, to the extent that any beam slice of length \( T_s \) has the same average error \( \langle x \rangle_s \) our considerations apply to any test particle. This is true on a statistical basis, as will become clear later.

A word about notation. It has become customary to write \( \lambda N_s = g \), and to call \( g \) the 'gain'. Remember that this \( g \) is proportional to the amplification (the electronic gain) of the system and proportional to \( N_s \). As from Eqs. (2.8), \(-g = \Delta x/\langle x \rangle_s \), a more precise (but longer) name is 'fraction of observed sample error corrected per turn'.

Now, we can again separate the coherent and incoherent effects and rewrite Eq. (2.6), by using the above notation:

\[ \Delta x = \frac{g}{N_s} x - g \langle x \rangle_s. \]  
(2.9)

coherent  incoherent
\( (\text{cooling}) \quad (\text{heating}) \)

Clearly, the problem is how to treat the incoherent term. The following approximations will be discussed:

First approximation: Neglect the incoherent term
Second approximation: Treat it as a fluctuating random term
Third approximation: Treat it as a fluctuating random term with some coherence due to imperfect mixing
Fourth approximation: Include additional coherence due to 'feedback via the beam'

2.3.1 First approximation

Neglecting completely the incoherent term in Eq. (2.9) we get a best performance estimate
\[ \Delta x = - \frac{q}{N_s} x . \]  

We expect an exponential form, \( x = x_0 e^{-t/\tau} \) for the amplitude of the test particle which gives the damping rate

\[ \frac{1}{\tau} = \frac{1}{x} \frac{dx}{dt} = \frac{1}{x} \frac{\Delta x}{\Delta t} = \frac{1}{x} \frac{\Delta x}{x} \text{ per turn}. \]  

Substituting into Eq. (2.11) from Eq. (2.10) gives

\[ \frac{1}{\tau} = \frac{g}{T} \frac{1}{N_s} . \]  

(2.12)

Interpreting \( g \) as the fractional correction, we intuitively accept that it is unhealthy to correct more than the observed sample error, i.e. we assume \( g < 1 \). Let us put \( g = 1 \) to make an estimate of the upper limit.

Finally it is convenient to express \( N_s \) in terms of the total number of particles, \( N \), in the beam and by the system's bandwidth \( W \), i.e. \( N_s = N(T_s/T) = N/2WT \) [see Eq. (2.2)]. We then obtain, a first useful approximation to the cooling rate:

\[ \frac{1}{\tau} = 2W \frac{1}{N} \text{ per second}. \]  

(2.13)

Amazingly enough, this simple relation overestimates the optimum cooling rate by only a factor of 2. However, to gain confidence, we have to justify some of our assumptions, especially the restriction of \( g < 1 \) and the neglect of the incoherent term. In fact, an evaluation of this term will clarify both assumptions and provide guidance on how to include other adverse effects such as amplifier noise.

2.3.2 Towards a better evaluation of the incoherent term

To be able to deal with the incoherent term, we make a detour into statistics to recall a few elementary 'sampling relations'\( ^6 \). Consider the following problem.

Given a beam of \( N \) particles characterized by an average \( \langle x \rangle = 0 \) and a variance \( \langle x^2 \rangle \equiv x^2_{\text{rms}} \) of some error quantity \( x \), suppose we take a random sample of \( N_s \) particles and do statistics on the sample population -- rather than on the whole beam -- to determine
i) the sample average $\langle x \rangle_s$;
ii) the sample variance $\langle x^2 \rangle_s$;
iii) the square of the sample average $[\langle x \rangle_s]^2$, i.e. the square of (i).

What are the most probable values [the expectation values, denoted by $E(\langle x \rangle_s)$, etc.] of these sample characteristics?

For random samples the most probable values are:

i) sample average + beam average;
ii) sample variance + beam variance;
iii) square of sample average + beam variance/sample population.

Or, in more mathematical language,

$$E(\langle x \rangle_s) = \langle x \rangle = 0$$  \hspace{1cm} (2.14a)

$$E(\langle x^2 \rangle_s) = \langle x^2 \rangle = x_{\text{rms}}^2$$  \hspace{1cm} (2.14b)

$$E[(\langle x \rangle_s)^2] = x_{\text{rms}}^2/N_s$$  \hspace{1cm} (2.14c)

Results (2.14a) and (2.14b) are in agreement with common sense, which expects that the sample characteristics are true approximations of the corresponding population characteristics. This is the basis for sampling procedures. Equation (2.14c) is more subtle as it specifies the error to be expected when one replaces the population average by the sample average.

$$[\langle x \rangle_s]^2 - (\langle x \rangle)^2 = x_{\text{rms}}^2/N_s$$

or symbolically

$$\langle x \rangle = \langle x \rangle_s \pm x_{\text{rms}}/\sqrt{N_s}$$

In other words: the larger the beam variance and the smaller the sample size ($N_s$), the more imprecise is the sampling. In this form, Eqs. (2.14) are used in statistics to determine the required sample size for given accuracy and presupposed values for the beam variance $x_{\text{rms}}^2$.

A slightly different interpretation is useful in the present context: suppose we repeat the process of taking beam samples and working out $\langle x \rangle_s$ many times. Although the beam has zero $\langle x \rangle$, the sample average will in general have a finite (positive or negative) $\langle x \rangle_s$. The sequence of sample averages will fluctuate around zero (around $\langle x \rangle$ in general) with a mean-square deviation $x_{\text{rms}}^2/N_s$. This is the fluctuation (or, if you prefer, the noise) of the sample average due to the finite particle number.
A simple example to illustrate the sampling relations and to familiarize us further with $\langle x^2 \rangle_S$ and $(\langle x \rangle_S)^2$ is given in Table 2. It is amusing to note that in this example 'the most probable values' 1/3 and 2/3 respectively [which agree with Eqs. (2.14)] never occur for any of the possible samples -- just another instance of statistics dealing with averages and being unjust to the individual.

Table 2

An example of the sampling relations

Assume a discrete distribution such that the values $x = -1, 0, 1$ occur with equal probability. Hence, beam average: $\langle x \rangle = 0$, and beam variance: $\langle x^2 \rangle = x_{\text{rms}}^2 = 1/3 \left[ (-1)^2 + 0^2 + 1^2 \right] = 2/3$.

Consider samples of size: $N_S = 2$. To work out the most probable values of the sample characteristics, write down all possible samples of size $N_S = 2$, determine $\langle x \rangle_S$, $(\langle x \rangle_S)^2$, and $\langle x^2 \rangle_S$, and take the average of these averages to find the expectations.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Sample averages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\langle x \rangle_S$</td>
</tr>
<tr>
<td>-1 -1</td>
<td>-1</td>
</tr>
<tr>
<td>-1 0</td>
<td>-0.5</td>
</tr>
<tr>
<td>-1 1</td>
<td>0</td>
</tr>
<tr>
<td>0 -1</td>
<td>-0.5</td>
</tr>
<tr>
<td>0 0</td>
<td>0</td>
</tr>
<tr>
<td>0 1</td>
<td>0.5</td>
</tr>
<tr>
<td>1 -1</td>
<td>0</td>
</tr>
<tr>
<td>1 0</td>
<td>0.5</td>
</tr>
<tr>
<td>1 1</td>
<td>1</td>
</tr>
</tbody>
</table>

Expectation = average of above values

$\langle x \rangle = \frac{1}{3} = \frac{2}{3}$

To conclude our detour, let us mention that the sampling relations (2.14) are a consequence of the more general 'central limit theorem' of statistics. For the present purpose we can quote this theorem as follows:

When a large number of random samples of size $N_S$ are taken from a population with statistics $\langle x \rangle = 0$ and $\langle x^2 \rangle = x_{\text{rms}}^2$ then the distribution of the sample averages is approximately Gaussian with a mean equal to the population mean and a standard deviation $\sigma = x_{\text{rms}} / \sqrt{N_S}$.

2.3.3 A better approximation of the cooling rate - second approximation

Returning to Eq. (2.8a), but re-expressing $\langle x \rangle_S$ in full we have,
\[ x_c = x - \frac{g}{N_s} \sum_{s} x_i . \]  

(2.15)

In order to profit from the sampling relations, it is more useful to evaluate the change \( \Delta(x^2) = x_c^2 - x^2 \) of the squared error rather than \( \Delta x \). Thus we obtain,

\[ \Delta(x^2) = -2g \frac{x}{N_s} \sum_{s} x_i + \left( \frac{g}{N_s} \sum_{s} x_i \right)^2 . \]  

(2.16)

The second term in Eq. (2.16) immediately gives

\[ \left( \frac{g}{N_s} \sum_{s} x_i \right)^2 = g^2 (\langle x \rangle)^2 + \frac{g^2}{N_s} x^2_{\text{rms}} , \]  

(2.17)

where we have used the sampling relation (2.14c) to express the expected variance of the sample average in terms of the beam variance \( x^2_{\text{rms}} \). To work out the first term we separate the test particle (once again) from the sum and write

\[ x \cdot \frac{1}{N_s} \sum_{s} x_i = \frac{x^2}{N_s} + \frac{x}{N_s} \sum_{s} x_i . \]

Next we apply the sampling relation (2.14a) to the remaining sum, i.e. we take

\[ E\left( \frac{1}{N_s-1} \sum_{s'} x_i \right) = E(\langle x \rangle_{s'}) = 0 \]

under the assumption that the sample (labelled \( s' \)) without the test particle is a random sample such that Eq. (2.14a) applies. Then

\[ E(x \cdot \frac{1}{N_s} \sum_{s} x_i) = \frac{x^2}{N_s} . \]  

(2.18)

Thus the first term in Eq. (2.16) has non-zero expectation. Clearly this is due to the fact that the \( x \) at the front "coheres" with the corresponding term inside the sum.

Putting together the terms, the expected change is then

\[ \Delta(x^2) = -2g \frac{x^2}{N_s} + \frac{g^2}{N_s} x^2_{\text{rms}} . \]  

(2.19)

Equation (2.19) applies to any test particle. Taking as typical a particle with an error equal to the beam r.m.s. we can write especially:
\[
\frac{1}{\tau_{x_{\text{rms}}}^2} \Delta(x_{\text{rms}}^2) = -\frac{1}{N} (2g - g^2).
\] (2.20)

This gives the cooling rate (per second) for the beam variance:

\[
\frac{1}{\tau_x} = -\frac{1}{T} \frac{\Delta(x_{\text{rms}}^2)}{x_{\text{rms}}^2} = \frac{1}{N} (2g - g^2) = \frac{2\Delta}{N} (2g - g^2) \text{.} \tag{2.21}
\]

Clearly the term 2g presents the coherent effect already identified. The \(-g^2\) term represents the incoherent heating by the other particles. The inclusion of this term is the improvement obtained in the statistical evaluation of this section.

It emerges quite naturally from Eq. (2.21) that g should not be too large! In fact, optimum cooling (maximum of 2g - g^2) is obtained with g = 1, and antidamping occurs if g > 2 (see Fig. 10).

It should be remembered that Eq. (2.21) gives the cooling rate 1/\(\tau_x^2\) for \(x^2\); the rate 1/\(\tau\) for \(x\) is half of this, as can be verified by comparing \(x^2 = x^2_0 \exp(-t/\tau_x^2)\) and \(x^2 = [x_0 \exp(-t/\tau)]^2\).

![Diagram](image-url)

**Fig. 10** Cooling or heating rate when considering the incoherent term as a random fluctuation.
2.3.4 Alternative derivation

For those who were not pleased with the way in which we separated the test particle from its sample and regarded the remainder as a random sample of size $N - 1$, we give yet another derivation of Eq. (2.21) which is due to Hereward (unpublished notes 1976, see also Ref. 7).

We restart from Eq. (2.16), which we write as

$$\Delta(x^2) = -2g x \cdot \langle x \rangle_s + g^2 \langle \langle x \rangle_s \rangle^2.$$  \hspace{1cm} (2.22)

This is the change for one test particle and one turn. We now take the average of this over the sample of the test particle (before, we took the average for one particle over many turns).

A slight complication arises from the fact that strictly speaking each particle defines its own sample, as sketched in Fig. 11. We can assume, however, that the long-term behaviour of any sample (i.e. any beam slice of length $T_s$) is the same, so that expectation values are independent of the choice of the sample.

![Fig. 11 Sample of the original test-particle (0) and of a particle passing earlier (i). Working out the average $\langle x_i \langle x \rangle_s \rangle_s$ of $x_i \langle x \rangle_s$ each particle has to be associated with its own sample. To the extent that all beam samples have the same statistical properties, all long-term averages are the same: $\langle x_i \langle x \rangle_s \rangle_s = \langle \langle x \rangle_s \rangle^2$.](image)

Then the only variable on the r.h.s. involved in averaging over the original sample is the $x$ in the first term, and we obtain

$$\langle \Delta(x^2) \rangle_s = -2g \langle \langle x \rangle_s \rangle^2 + g^2 \langle \langle x \rangle_s \rangle^2.$$  \hspace{1cm} (2.23)

Next we use the sampling relations (2.14b) and (2.14c). We include the fact that the correction (2.23) is applied to all beam samples once per turn. Thus,

$$\langle \Delta(x^2) \rangle_s = \Delta x^2_{\text{rms}},$$

$$\langle \langle x \rangle_s \rangle^2 = x^2_{\text{rms}} / N_s,$$
and the expected correction of beam variance per turn is

$$
\Delta(x_{rms}^2) = \frac{1}{N_s} (2g - g^2) x_{rms}^2
$$

i.e. exactly as assumed in Eq. (2.20).

This leads to the same cooling rate as that given by the previous approach, but the derivation lends itself to the following formulation of the 'sampling picture'.

Take a random beam sample of $N_s$ particles. Measure and correct its average error $\langle x \rangle_s$ by giving a kick $-g\langle x \rangle_s$ to all particles. Owing to the finite particle number, the beam variance appears as a fluctuation with 'noise' $(\langle x \rangle_s)^2 + x_{rms}^2 / N_s$ of the centre of gravity $\langle x \rangle_s$. By correcting $\langle x \rangle_s$ to $(1-g)$ of its value (i.e. to zero for full $g = 1$), one reduces the sample variance (on the average) by $1/N_s (2g-g^2)$. Repeat $N/N_s$ times per turn to reduce the beam variance by the same amount. Repeat for many turns.

| Table 3 |

'Simulation' of a one-turn correction (with $g = 1$) using the example of Table 2. We note down all possible samples of size $N_s = 2$ and reduce the sample errors to zero by applying the same correction to both sample numbers. This reduces the beam variance from 2/3 to 1/3, i.e. $\Delta x_{rms}^2 / x_{rms}^2 = 1/N_s = 1/2$.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Before correction</th>
<th>After correction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average $\langle x \rangle_s$</td>
<td>Variance $\langle x \rangle_s^2$</td>
</tr>
<tr>
<td>-1 -1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1 0</td>
<td>-0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>-1 1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0 0</td>
<td>-0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0 1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 -1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1 0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1 1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

'TBeam variance' (average of all sample variances) 2/3 1/3

Thus, rather than treating single particles, one measures and corrects the centres of gravity of beam samples. It is amusing (but not too surprising) to note that the total number of measurements, namely the number of turns $n = N_s$, required for reasonable cooling multiplied by the number $x_s = N/N_s$ of samples per turn, is $N$, as if we treated the $N$ particles individually.
It is easy to test this sampling prescription for simple distributions; in Table 3 we use the previous example (Table 2) to verify that the full correction (g = 1) reduces the variance by $1/N_5$ per turn. More generally, the sampling recipe can easily be simulated on a desk computer using a random number generator.

In the next two sections we will use the test particle and the sampling picture alternately to introduce two final ingredients, namely electronic noise of the amplifier and mixing of the samples due to the spread in revolution time.

2.3.5 A refinement to include system noise

A large amplification of the error signals detected by the pick-up is necessary to give the required kicks to the beam. Electronic noise in the preamplifiers then becomes important. In Table 4 we anticipate some typical numbers pertaining to transverse cooling of $10^5$ antiprotons in LEAR. This example should convince us of the necessity to rewrite the basic equations to include noise. It is convenient\(^7\) to represent noise by an equivalent sample error (denoted by $x_n$) as observed at the pick-up. We then regard the system sketched in Fig. 12 and write

$$x_c = x - g\langle x \rangle_s - gx_n.$$  \hspace{1cm} (2.24)

| Table 4 |
|-----------------|-----------------|
| **Pick-up signal** | 50 nA |
| **Preamplifier noise current** | 150 nA |
| **Kicker voltage per turn** | 1 V |
| **Corresponding current (into 50 Ω)** | 20 mA |
| **Power amplification** | $2 \times 10^{10}$ |

Signal, noise, and amplification of a cooling system; orders of magnitude for $10^5$ particles and 50 s cooling time

![Diagram](https://example.com/diagram.png)

**Fig. 12** Cooling loop including system noise. The noise is represented as an equivalent sample error $x_n(t)$ as observed at the pick-up.
Going once again through our basic procedure, taking random noise uncorrelated with the particles we obtain the expected cooling rate

\[
\frac{1}{\tau^2} = \frac{2W}{N} \left[ 2g - g^2 (1+U) \right],
\]

(2.25)

where \( U = E(x_n^2)/E(x_n^2) \) is the ratio of the expected noise to the expected signal power*, called the 'noise-to-signal power ratio' or noise-to-signal ratio for brevity.

This introduces the noise into our pictures: it increases the incoherent term by \((1+U)\). System noise and the disturbance caused by the other particles enter in much the same way; the latter is therefore also called particle noise.

Several things can be observed from Eq. (2.25). Cooling remains possible despite very poor signal-to-noise ratios \((1/U \ll 1)\). All we have to do is to choose \( g \) small enough \((g < g_0 = 1/(1+U) = 1/U)\), which unavoidably means slow cooling \((\tau > NU/2W)\). In other words, we have to be patient and give the system a chance to distill a signal out of the noise.

In the initial cooling experiment (ICE) with 200 circulating antiprotons the system worked with signal-to-noise ratios as low as \(10^{-6}\).

Secondly, \( U \) has a tendency to increase as cooling proceeds: namely the noise tends to remain the same, whereas the signal decreases as the beam shrinks. This is the case unless the pick-up plates are mechanically moved to stay close to the beam edge -- as will be done in the new antiproton collector ACOL\(^9\) to be built at CERN.

With changing \( U \), cooling is no longer exponential. Equation (2.25) gives a sort of instantaneous rate, and cooling stops completely \((1/\tau \to 0)\) when \( U \) has increased such that \((1+U) = 2/g\). In this situation, equilibrium is reached between heating by noise and the damping effect of the system. To avoid this 'saturation' it is sometimes advantageous to decrease \( g \) during cooling in order to work always close to the optimum gain [maximum of Eq. (2.25)] \( g_0 = 1/(1+U)\).

In all cases it is important to obtain a good signal-to-noise ratio. Frequently, this means having a large number of pick-ups as close as possible to the beam, as well as high quality, low-noise preamplifiers often working at cryogenic temperatures.

2.4 Mixing - third approximation

So far, all our considerations have been based on the assumption of random samples. This is a good hypothesis for an undisturbed beam. However, the cooling system is designed to correct the

* Stochastic cooling of heavy ions is becoming very important so we should note that \( U = U/Z^2 \) where \( Z \) is the charge number of the particle and \( U \) the noise to signal ratio calculated for singly charged particles.
statistical error of the samples. Just after correction, samples will no longer be random. For full correction the centre of gravity \( \langle x \rangle_s \) will be zero rather than \( \sqrt{x_{\text{rms}}} / N_s \) as expected for random conditions. Cooling will then stop as no error signal is observable.

Fortunately, owing to momentum spread, particles in a storage ring go round at slightly different speeds, and the faster ones continuously overtake the slower ones. Because of this mixing, the sample population changes and the sample error reappears, until ideally all particles have zero error. The dispersion of revolution time with momentum is governed by

\[
\frac{\Delta T}{T} = - \eta \frac{\Delta p}{p}
\]

(2.26)

where the off-momentum function \( \eta = \gamma^{-2} - \gamma^{-2}_t \) is given by the distance of the working energy \( (\gamma) \) from transition \( (\gamma_t) \).

If mixing is fast so that complete re-randomization has occurred on the way from kicker to pick-up then the assumption of random samples made in the previous sections is valid. If however, mixing is incomplete, cooling is slower. In fact, if it takes \( M \) turns for a particle of typical momentum error to move by one sample length with respect to the nominal particle \( (\Delta p / p = 0) \), then intuitively one expects an \( M \) times slower cooling rate.

A slightly different way of looking at imperfect re-randomization suggests itself in the frame of the test particle picture: bad mixing means that a particle stays too long -- namely \( M \) rather than 1 turns -- together with the same noisy neighbours. This increases the incoherent heating by the other particles by a factor \( M \).

We thus generalize the basic Eq. (2.25) (a rigorous derivation will be given later)

\[
\frac{1}{\tau_{\text{x}}^2} = \frac{2M}{N} \left[ 2g - g^2 (M+U) \right],
\]

(2.27)

and call \( M > 1 \) the mixing factor.

\( M \) is defined as the number of turns for a particle with one standard deviation in momentum to migrate by one sample length \( T_s \).

Equation (2.27) has the optimum,

\[
g = g_0 = 1/(M + U),
\]

\[
\tau = \tau_0 = \frac{N}{2M} (M + U).
\]

(2.28)

This underlines the importance of having good mixing -- \( M > 1 \) -- on the way from correction to the next observation, but ...
What about mixing between observation and correction? Surely if the sample as observed is very different from the sample as corrected, then adverse effects can happen. Let us again resort to the test particle description and try to imagine how the coherent and the incoherent effects change. As to the latter, we expect that it is to first order not affected. We can just assume that the perturbing kicks are due to a new sample which has the same statistical properties as the original beam 'slice'.

The coherent effect will, however, change because the system will be adjusted in such a way that the correction pulse will be synchronous with the nominal particle ($\Delta p/p = 0$). Particles that are too slow or too fast on the way from pick-up to kicker will therefore slip with respect to their self-induced correction (Fig. 13). In fact, in the rectangular response model used above, the coherent effect will be completely zero if the particle slips by more than half the sample length ($|\Delta T_{PK}| > T_S/2$). At this stage, it is more realistic to use a parabolic response model of the form $1 - (\Delta t/T_C)^2$, where $T_C$, the useful width of the correction pulse, is about equal to the sample length $T_S$ for a low-pass system. But $T_C$ is shorter than $T_S$ for a high-frequency band-pass system with $f_{\text{min}} > \tilde{f}_g$ with a response as sketched in Fig. 6(b); $\Delta t$ is the time-of-flight error of the particle between pick-up and kicker. Introducing the typical error $\Delta T_{PK}$ and calling $\Delta T_{PK}/T_C = 1/\tilde{M}$, we can modify the coherent term $g \rightarrow g[1 - \tilde{M}^{-2}]$ to account for unwanted mixing between observation and correction. In a regular lattice the flight time from pick-up to kicker is a fixed fraction of the time from kicker to pick-up, and the two mixing factors $M$ and $\tilde{M}$ are proportional to each other, $M = \alpha \tilde{M}$, with $\alpha$ being the ratio of the corresponding distances -- hence the interest in having a short beam path from pick-up to kicker.

![Diagram](image.png)

**Fig. 13** Synchronism between particles and their correcting pulse on their way from pick-up to kicker. The response of the cooling system to a particle (the 'coherent effect') is approximated by a 'parabola' $s(t) = 1 - (\Delta t/T_C)^2$ of width $\pm T_C$ instead of the 'rectangle' used in Figs. 7 and 8. A nominal particle (O) arrives at the kicker simultaneously with the correction kick. The particle f is much too fast and advances its correction pulse. The particle s is slightly too slow. Thus, the three particles receive full correction, no correction, or partial correction, respectively.

The reduced correction becomes,

\[
\Delta x_C = -2g[1 - (\frac{\Delta T_{PK}}{T_C})^2]x = -2g[1 - \tilde{M}^{-2}]x.
\]  

(2.29)
This will give us a slightly different form for the basic equation,

\[ \frac{1}{\tau x^2} = \frac{2M}{N} [2g(1-M^{-2}) - g^2(M+U)] \]  \hspace{1cm} (2.30)

Coherent Incoherent (cooling) (heating)

By a clever choice of the bending and focusing properties of the storage ring it is possible, in principle, to make \( \Delta T_b = 0 \) independent of momentum, and \( \Delta T_f \) large to approach the desired situation of \( M^{-2} = 0 \) and \( M = 1 \). But this complicates the storage ring lattice. The compromise adopted in existing designs is to sacrifice some of the desired re-randomisation in order to avoid too much unwanted mixing.

Following convention, we now return to the cooling rate for \( x \) rather than \( x^2 \) (using \( 1/\tau = 1/2 \)

\[ \frac{1}{\tau} = \frac{W}{N} [2g(1-M^{-2}) - g^2(M+U)] \]  \hspace{1cm} (2.31)

with \( M > 1, U > 0, M^{-2} < 1 \).

Equation (2.31) is the main result of our simple analysis. It exhibits some of the fundamental limitations of stochastic cooling. We first note that \( 1/\tau \) has a maximum characterized by

\[ g_0 = \frac{1 - \tilde{M}^{-2}}{M + U}, \]  \hspace{1cm} (2.32)

\[ \frac{1}{\tau_0} = \frac{W}{N} (\frac{1 - \tilde{M}^{-2}}{M + U})^2. \]  \hspace{1cm} (2.33)

As an example of relatively straightforward technology, we take \( W = 250 \text{ MHz} \). Then, in the best of all cases (\( M = 1, U = 0, \tilde{M}^{-2} = 0 \)) this gives

\[ 1/\tau = W/N = 2.5 \times 10^5/N \text{ [sec}^{-1}] \]  \hspace{1cm} (2.34)

i.e. \( \tau = 1 \text{ s at } 2.5 \times 10^5 \text{ p or } \tau = 1 \text{ day at } 10^{13} \text{ p.} \)

To include mixing, we assume that the time-of-flight dispersion between pick-up and kicker and between kicker and pick-up and the system response are such that the unwanted mixing \( \tilde{M} \) is one half of the wanted mixing, i.e. we put (as an example) \( \tilde{M}^{-1} = 0.5 \text{ M}^{-1} \). We further assume that the sensitivity and the number of pick-ups are such that \( U = 1 \) (little is gained in this example in going to more pick-ups, such that \( U \ll 1 \)). Then the best cooling, obtained with \( M = 1.5 \), is
This is about three times slower than the rate (2.34) with $\tilde{\omega}^{-2} + 0$, $U = 0$. We retain that over a wide range of parameters $1/s = a_0 W/N$.

From Fig. 14 we conclude that existing cooling systems follow a 'working line' with $1/s = 0.1$ to 0.3 W/N, i.e. $a_0 = 0.1-0.3$. A bandwidth of 250 to 500 MHz is (more or less) standard; 2 to 4 GHz will be used in the CERN-ACOL and the Fermilab antiproton sources. Bands of 4 to 8 GHz or higher have been contemplated for sources accumulating $10^{13}$ antiprotons in a few hours$^{10}$, as desirable for multi-TeV colliders (see Table 5).

Table 5

Parameters of present, future and 'ultimate' cooling systems. The quantity $a_0$, defined by $1/s = a_0 W/N$ describes the efficiency of solving the noise and mixing problems.

<table>
<thead>
<tr>
<th>Machine</th>
<th>Date</th>
<th>W (GHz)</th>
<th>$a_0$</th>
<th>$s[s]/N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achieved</td>
<td>1976</td>
<td>0.1</td>
<td>0.03</td>
<td>$1/3 \times 10^6$</td>
</tr>
<tr>
<td>AA</td>
<td>1980</td>
<td>0.25</td>
<td>0.1</td>
<td>$1/2 \times 10^7$</td>
</tr>
<tr>
<td>(precooling)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Future</td>
<td>Fermilab</td>
<td>1986</td>
<td>0.25</td>
<td>$1/5 \times 10^6$</td>
</tr>
<tr>
<td>ACOL</td>
<td>1987</td>
<td>2</td>
<td>0.25</td>
<td>$1/5 \times 10^6$</td>
</tr>
<tr>
<td>Ultimate</td>
<td>?</td>
<td>15</td>
<td>0.5</td>
<td>$1/7 \times 10^9$</td>
</tr>
</tbody>
</table>

2.5 Practical details

So far we have, in a general way, discussed a system for correcting 'some error $x$'.

In practice cooling is used to reduce the horizontal and/or vertical betatron oscillation and the momentum spread of the beam. Table 6 gives a summary of the corresponding hardware.

The simple time-domain approach can be directly applied to momentum cooling by the Palmer-Hereward method. This will be discussed in the next subsection. A discussion of the other momentum cooling methods and of betatron oscillation cooling will be deferred to later sections.
Fig. 14 Normalized cooling time versus intensity. The inclined lines represent the mixing limit. For low intensity the cooling time levels off because of noise. The points represent initial cooling in various machines. These points roughly follow a line with $\tau = 10$ N/W. During cooling, noise and/or mixing tend to become more important and the cooling time longer. Note that the vertical scale is normalized for 100 MHz bandwidth. Future systems (ACOL, Fermilab) aim at 2-5 GHz bandwidth.
Table 6
Stochastic cooling systems in use or proposed

<table>
<thead>
<tr>
<th>Type</th>
<th>Pick-up</th>
<th>Corrector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betatron cooling, horizontal or vertical</td>
<td>Difference pick-up</td>
<td>Transverse kicker</td>
</tr>
<tr>
<td>Momentum cooling, Palmer-Hereward type</td>
<td>Horizontal difference pick-up</td>
<td>RF gap (acceleration/ deceleration)</td>
</tr>
<tr>
<td>Momentum cooling, filter method</td>
<td>Longitudinal (sum) pick-up + comb filter</td>
<td>RF gap</td>
</tr>
<tr>
<td>Momentum cooling, transit time method</td>
<td>Longitudinal pick-up + differentiator or two longitudinal pick-ups</td>
<td>RF gap</td>
</tr>
</tbody>
</table>

2.6 Palmer cooling

A horizontal position pick-up is used to detect the horizontal orbit displacement $x = D \langle \Delta p/p \rangle_S$ concurrent with the momentum error of the sample; $D$ (also denoted by $\Delta p$ or $x_p$) is the value of the orbit 'dispersion function' at the pick-up as determined by the focusing properties of the storage ring. In addition to the momentum dependent displacement there are further contributions to the position error, especially the betatron oscillation ($x_B$) of the particles. We shall neglect this contribution, assuming that the pick-up is placed in a region of large dispersion so that $\langle D \Delta p/p \rangle_S$ dominates over $\langle x_B \rangle_S$. We are then in a situation where momentum cooling as envisaged by R. Palmer (private communication to L. Thorndahl and H.G. Hereward in 1975) is possible. At the RF gap the particle receives a 'kick' of momentum and hence a change of $x$ proportional to the detected error.

The basic one-passage equation (including noise) is written as

$$x_C = x - g \left[ \langle D \Delta p/p \rangle_S + x_n \right].$$

This is completely equivalent to Eq. (2.15), thus leading to the cooling rate (for $x^2$ and $\Delta p^2$, i.e. for the mean square of the momentum deviation):

$$\frac{1}{\tau_{\Delta p^2}} = \frac{2W}{N} \left[ 2g(1 - \bar{N}^{-2}) - g^2(\Delta U) \right],$$

(2.36)

where $U = E(x_n^2)/E[(D \Delta p/p)^2]$ is the noise-to-signal ratio; $x_n$ is the system noise expressed as the equivalent pick-up signal $D(\Delta p/p)$, and $E(x_n^2)$ the expectation (i.e. the long-term average) of $x_n^2$. Above we assumed that the orbit dispersion $x = D(\Delta p/p)$ dominates at the pick-up so that the betatron oscillation $x_B$ is negligible there. We also implied that at the kicker the dispersion function $D_k$, as well as its derivative, $D'_k$, are zero. Otherwise the momentum correction leads to
an excitation of horizontal betatron oscillations. The effect is that the momentum kick introduces
an abrupt change of the equilibrium orbit and the particle starts to oscillate around this new
displaced orbit.

The more realistic case where both $x$ and $x_\parallel$ are present at the pick-up and where $D_k$ is
non-zero at the kicker was analysed by Hereward. He showed$^{11}$ the mutual heating and at the same
time the possibility of using the Palmer system for simultaneous longitudinal and horizontal cooling
by a suitable choice of the pick-up to kicker distance.

3. A MORE DETAILED PRESENTATION OF BETATRON COOLING, FREQUENCY DOMAIN PICTURE

3.1 Betatron equation

Before entering into details, it is worth trying to establish a simple picture of betatron cooling in which the various phenomena can be identified.

Consider first the smooth sinusoidal approximation for the betatron motion$^{2}$ of a single
particle (subscript $i$) in a storage ring, with forcing terms on the right-hand side arising from its
proper motion, the motion of other particles (subscript $j$) and system noise.

$$\ddot{x}_i(t) + \omega^2 x_i(t) = G_{ii} x_i(t-t_p) + \sum_{n \neq j} G_{ij} x_j(t-t_p) + \text{'system noise'}$$  \hspace{1cm} (3.1)

<table>
<thead>
<tr>
<th>Coherent effect</th>
<th>Incoherent effect</th>
<th>Additional incoherent effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Mixing PU + K</td>
<td>- Mixing K + PU</td>
<td>- Enhancement of cooling</td>
</tr>
<tr>
<td>- Betatron phase errors</td>
<td>- Signal shielding</td>
<td></td>
</tr>
</tbody>
</table>

We interpret the left-hand side as the motion on entering the cooling kicker (K) and the forcing
terms on the right-hand side as being derived from the motion seen earlier (i.e. at $t-t_p$) in the
pick-up (PU). The characteristics of the pick-up, amplifier, transmission system and kicker enter
into both the coefficients $G_{ij}$ and the "system noise". $\Omega$ is the revolution frequency and $Q$ the
tune of the storage ring.

3.2 Simplified coherent effect

If we neglect the incoherent terms in Eq. (3.1) and make $G_{ii} =$ constant, we obtain a single-
particle cooling equation,

$$\ddot{x}(t) + \omega_p^2 x(t) = G_{ii} x(t-t_p)$$ \hspace{1cm} (3.2)

putting $\omega_p = \Omega Q$.

For a weak perturbation term, we can expect a solution of the form:

$$x(t) = \bar{x} e^{i\omega t}$$

$$x(t-t_p) = \bar{x} e^{i\omega(t-t_p)}.$$
Substituting into Eq. (3.2) gives,

\[
\begin{align*}
\omega^2 - \omega^2 &= G_{ii} e^{-i\mu} \\
-2\omega_0^2 \omega &= G_{ii} e^{-i\mu} \\
\Delta\omega &= \frac{-G_{ii} e^{-i\mu}}{2\omega_0^2}.
\end{align*}
\]  

(3.3)

This is the expected response of a feedback system. The real part of \(\Delta\omega\) is the frequency shift of the perturbed oscillation and the imaginary part of \(\Delta\omega\) gives the damping (or heating) of the oscillation.

\[
\frac{1}{\tau} = \text{Im}(\Delta\omega) = \text{Re} \left[ \frac{G_{ii}}{2\omega_0^2} e^{i(\mu - \pi/2)} \right].
\]  

(3.4)

Equation (3.4) would be exact if the observation and feedback on the beam were continuous, and \(G_{ii}\) constant, which are manifestly not the case. We must now, therefore, investigate the effects of periodic observation and correction.

3.3 Orbit equation for a constant localized kick

The orbit in a storage ring with constant kicks can be regarded as a betatron oscillation which closes onto itself by virtue of the angular discontinuities at the kicks. The closed orbit is given for any distribution of kicks by the well-known equation:

\[
\begin{align*}
\Delta B_z(s) + E_X(s) \frac{x(s)}{B_0 p_0 \beta} = x(s) + K(s)x(s) = F(s)
\end{align*}
\]  

(3.5)

where \(E_X\) is the transverse electric field [V/m]

- \(\Delta B_z \) is the transverse magnetic field [T] error (vertical field for horizontal orbit deformation and vice versa)
- \(B_0 p_0\) is the magnetic rigidity [T\cdot m] = 3.3356\times10^{-9} \ p [eV/c]
- \(p\) is the particle momentum [eV/c] and \(\beta = v/c\)
- \(s = \beta ct\) is the distance along the orbit.

The Eq. (3.5) is good for numerical calculations, but is inconvenient for analytical work, since it is defined piecewise in \(s\) as one proceeds around the machine through the various elements.

Using the well known Courant and Snyder transformation\(^2\) the equation can be rewritten as a "driven harmonic oscillator" with fixed frequency \(Q\) rather than with azimuthally varying \(K(s)\) as in Eq. (3.5):

\[
\ddot{\eta}(s) + Q^2 \eta = Q^2 \beta X(s) F(s)
\]  

(3.6)
where \( \eta = \alpha(s)\beta_x^{1/2}(s) \) is the normalised displacement, \( d\Phi = ds/Q\beta_x \) defines the Courant and Snyder angle which increases by \( 2\pi \) per turn, \( \beta_x \) is the betatron function of the storage ring and now indicates differentiation with respect to \( s \).

Equation (3.6) is quite general, but we are especially interested in a single narrow kick, which we can represent by a delta function, or rather by a periodic delta function \( \delta(s) \) with the revolution frequency (see Fig. 15).

![Diagram of a single narrow kick in a storage ring](image)

**Fig. 15** For a single narrow kick in a storage ring the particle sees a periodic influence

The kick as seen by a circulating particle is represented by,

\[
F(s) = (F_0\Delta s)\delta(s-n2\pi R) = (F_0\Delta s) \sum_{n=-\infty}^{\infty} \delta(s-n2\pi R)
\]  

(accepting that high frequency components, i.e. \( f > c/\lambda \) are not required).

Equation (3.7) is a good representation of a short kicker with a constant kick, but analytically it would be easier to manipulate if we could replace the discontinuous delta functions with continuous functions. This can readily be done by making a Fourier analysis of the quantities \( \beta_x^{3/2} F(s) \) which appears on the r.h.s. of Eq. (3.6).

\[
\beta_x^{3/2}(s)F(s) = \sum_{\lambda=\infty}^{\infty} f_\lambda(\theta) e^{i\lambda\theta}
\]  

where,

\[
f_\lambda = \frac{1}{2\pi} \int_0^{2\pi} \beta_x^{3/2} F e^{-i\lambda\theta} d\phi = \frac{1}{2\pi} \int_0^{2\pi} \beta_x^{3/2} F e^{-i\lambda\theta} \frac{ds}{Q\beta_x}.
\]

Since \( F \) is a delta function the integral (3.8) simply leads to Fourier coefficients which are all equal. (This is the advantage of the complex formulation, which avoids the \( f_0/2 \) coefficient of the real expansion.)

\[
f_0 = f_\lambda = \frac{F_0\Delta s}{2\pi Q} \beta_x^{1/2} \frac{\Theta}{2\pi Q} \beta_x^{1/2}
\]
where $\beta_k$ is the value of the beta function at the kicker ($s=0$) and

$$\theta = \left( \frac{\Delta B_k}{B_{0P_0}} + \frac{E_x}{\frac{p}{\beta}} \right) \Delta s$$

is the kick strength.

We can now rewrite Eq. (3.6) in terms of continuous functions

$$\bar{\eta}(\tau) + Q^2 \eta = \frac{R_{k}}{2\pi} \theta \sum_{\Omega} e^{i\Omega \tau}$$

This is in fact all we really need, but we can make two variable changes, which will make the equation more familiar. Firstly, since we prefer to think in terms of the time, we can introduce a 'time-like' variable $\bar{\phi}/\Omega$ i.e. the Courant and Snyder normalised phase, $\phi$, scaled by the revolution angular frequency, $\Omega$. Rather loosely we will still refer to this variable as $t$. In fact, this lack of rigour is not too serious, since $t$ will coincide with the true time at least once every revolution at the kicker ($s=0$), which is the one point where true time is important. In any case, in most lattices $\phi/\Omega$ will not stray far from true time at any point in the ring. Secondly, we like to think in the transverse deviation $x$, so we undo the normalisation of the variable $\eta$, but again we are only really interested in true deviations at the pickup, so we define a quasi-position variable $x_{PU}$ which gives true position once per turn at the pick-up and again we loosely call it $x$

$$t = \frac{\phi}{\Omega} = \text{true time} , \quad x = x_{PU}^{1/2} = \text{true position} .$$

Using (3.11) in (3.10), we find

$$X(t) + Q^2 x \frac{R_{k}}{2\pi} \theta \sum_{\Omega} e^{i\Omega \tau} .$$

Thus the betatron motion is driven by an infinite set of Fourier harmonics of equal amplitude and separated in frequency, one from the other, by the revolution frequency $\Omega$ (see Fig. 16).

---

**Fig. 16** Fourier spectrum of the single constant kick of Fig. 15
3.4 Transverse RF knockout

In the previous section we analysed the equation of betatron motion with a constant kick. If we now modulate this excitation we can still use the expansion (3.8) of the kick (3.7). We simply have to make the kick strength $\Theta$ of Eq. (3.9) a function of time. Just imagine $F$ in Eq. (3.7) multiplied by some modulation factor, say $e^{i\omega t}$. You can keep this factor separate, expand the rest as before and then multiply with the modulation factor. The result is that all Fourier coefficients $f_k$ (Eq. 3.9) are modulated by the same factor. This leads to the phenomenon known as transverse RF knockout. Let us rewrite Eq. (3.12) as:

$$x(t) + \omega_0^2 x = V_0 e^{i\omega t} \sum_{k=-\infty}^{\infty} \frac{1}{k!} \frac{d^k x}{dt^k}$$

where

$$V_0 = \frac{\Theta_0^2}{2\pi} \left( \sqrt{P_k} \right) \Theta_0,$$

$\Theta_0$ being the amplitude of the kick, $\Theta = \Theta_0 e^{i\omega t}$.

Note that $V_0 e^{i\omega t}$ uses true time i.e. the time at the kicker. Equation (3.13) may also be written as

$$\ddot{x}(t) + \omega_0^2 x = V_0 \sum_{k=-\infty}^{\infty} e^{i(k\Omega\omega)t}.$$

Equation (3.14) is expressed with negative and positive frequencies, which correspond to the slow and fast waves set up by the disturbing kick. Those not familiar with complex voltages and currents as used by electrical engineers may wonder about the significance of negative frequencies. In fact above we assumed a complex excitation $\Theta = \Theta_0 e^{i\omega t}$ of the kicker as this greatly simplifies the algebra. In real life we deal with cosine- rather than $e^{i\omega t}$-type of kicker fields. It is easy to go from the complex to the real world by taking the real parts of Eq. (3.14). Then the r.h.s. contains terms which can be written in the form $\cos(m \Omega t + \omega t)$, $\cos(m_1 \Omega t - \omega t)$ and - if $\omega > \Omega$ also $\cos(\omega t - m_2 \Omega t)$ - with positive frequencies only ($m$: any integer $> 0$, $m_1$ any integer $> \omega/\Omega$ and $m_2$ any integer $0 < m_2 < \omega/\Omega$).

Thus the particle 'sees' the frequencies $m\Omega \pm \omega$, i.e. two sidebands spaced by the kicker excitation frequency $\omega$ left and right of each revolution harmonic $m\Omega$. This is illustrated in Fig. 17 where the spectrum of the complex excitation [r.h.s. of Eq. (3.14)] and its reflection into the real world are sketched. Taking the real part simply corresponds to reflecting the negative frequencies into the positive $f$-plane.

The revolution sidebands at $m\Omega \pm \omega$ are very similar to the sidebands at $\omega_{osc} \pm \omega_{mod}$ of an amplitude modulated oscillator. For this simpler example the complex and the real analysis are once again summarized in Table 18.
Reflection of negative frequencies onto positive frequency side gives the true spectrum seen by the beam.

Fig. 17 Simple harmonic excitation on a short kicker, spectrum of excitation waveform and of waves seen by a particle.

**Table 18**

The spectrum of an amplitude modulated oscillator ($\omega_{RF} > \omega_m$)

<table>
<thead>
<tr>
<th></th>
<th>Complex notation</th>
<th>'Real world'</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modulated amplitude</strong></td>
<td>$a = A e^{i\omega_m t}$</td>
<td>$a = A \cos \omega_m t$</td>
</tr>
<tr>
<td><strong>Carrier</strong></td>
<td>$c = 0.5(e^{i\omega_{RF}t} + e^{-i\omega_{RF}t})$</td>
<td>$c = \cos \omega_{RF} t$</td>
</tr>
<tr>
<td><strong>Position of spectral lines of modulated signal $a \times c$</strong></td>
<td>$(\omega_{RF} + \omega_m)$ and $-(\omega_{RF} - \omega_m)$</td>
<td>$\omega_{RF} + \omega_m$ and $\omega_{RF} - \omega_m$</td>
</tr>
</tbody>
</table>
Returning to Eq. (3.14), clearly if the driving term has a harmonic at the natural frequency of response, $\omega_p$, then the beam will behave resonantly.

For resonance, $(\lambda^2 \omega)^2 = \omega_p^2$

i.e.

$$\omega_\pm = \lambda \Omega \pm \omega_p = (\lambda \pm ||Q|| \pm q)\Omega$$  \hspace{1cm} (3.15)

$$\omega_\pm = (n \pm q)\Omega$$  \hspace{1cm} (3.16)

where $q$ is the fractional part and $||Q||$ the integer part of the tune. Hence $n = \lambda \pm ||Q||$ is also an integer.

Equation (3.16) shows that the beam will respond resonantly at the "betatron sidebands" $(n\pm q)\Omega$ centred on the revolution frequency harmonics. Take as an example LEAR at 600 MeV/c with $f_0 = 1$ MHz, $q = 0.3$. Resonant beam response ("RF - knockout") will occur when the kicker is excited at: 0.3 MHz $[= (0 + q)f_0]$, or 0.7 MHz $[= (1 - q)f_0]$ or 1.3 MHz ... .

Figure 18 sketches how two voltages of different frequency ($\omega$ and $\omega Q$ respectively) on a short kicker can produce the same series of kicks as seen by a particle. Note the analogy to an RF accelerating cavity which can in principle work at any revolution harmonic.

Having established the beam response to a kicker we shall next analyse the reciprocal problem of the signal response of the pick-up to the betatron oscillation of a particle.

![V(t) frequency kicker excitation](image)

**Fig. 18** Example of how a beam can be driven in the same way by different frequencies applied from a short kicker. The bars are the kicker voltages at the moment the particle passes, i.e. the kick experienced. The low frequency ($\omega$) and the higher frequency ($\omega + \lambda\Omega$) excitation produce the same apparent kick.

3.5 Signals from a circulating particle

A circulating particle passes once per turn through the pick-up and induces a short pulse. This can be represented by a periodic delta function and we have this, now rather familiar, picture in Fig. 19.
The induced signal is given by,

\[ I_s(t) = e^{\frac{e}{\Delta t}}(t - t_p + nT) = e^{\sum_{n=-\infty}^{\infty} \delta(t - t_p + nT)} \]  

(3.17)

accepting that high frequency components, i.e. \( f > 1/\Delta t \), are not required. Here \( e \) is the charge per turn \( \int_0^T I(t) \, dt \) passing the pick-up, i.e. the charge of the particle.

As before we Fourier analyse the signal in order to replace the discontinuous delta functions by continuous functions

\[ I_s(t) = \sum_{m=-\infty}^{\infty} I_m e^{im\omega(t - t_p)} \]  

(3.18)

where

\[ I_m = \frac{e}{T} = e \frac{Q}{2\pi} . \]

The Fourier harmonic amplitudes are constant. (In practice as the harmonic frequency approaches \( 1/\Delta t \) the amplitudes will decrease.)

The pick-up we use for a betatron cooling system will need to be sensitive to the transverse beam position. This is achieved by placing electrodes on either side of the beam. Each plate will have a signal induced of the form of Eq. (3.18) and its amplitude will be proportional to the distance of the beam particle from the plate. From the difference signal between the two plates the particle's transverse position \( x \) is obtained. We write the difference signal as \( I(t) = I_s \cdot S_p \cdot x/h \), where \( I_s \) is the sum signal Eq. (3.18), \( h \) is the half aperture of the pickup and \( S_p \) a factor of the order of unity. This factor as well as the assumed linearity depend very much on the construction of the pick-up. For a particle performing betatron oscillations we obtain an induced signal modulated by the transverse betatron motion

\[ x = x \cdot e^{i(q\omega t - \mu \psi)} . \]  

(3.19)
This signal is:

\[ I_s(t) = \sum_{m'=-\infty}^{\infty} S_p \left( \frac{X}{\hbar} \right) \int_{-\infty}^{\infty} e^{imq(t-t_p)} \]

\[ I_s(t) = S_p \sum_{m'=-\infty}^{\infty} e^{i\frac{Q}{2\theta} t} \int_{-\infty}^{\infty} e^{im \omega(t-t_p) - im\mu} \]

Since only the real part of the betatron motion (3.19) and hence of the current (3.20) interests us, we may write Eq. (3.20) as

\[ I_s(t) = S_p \sum_{m'=-\infty}^{\infty} \frac{e^{i\frac{Q}{2\theta} t}}{\omega} \int_{-\infty}^{\infty} e^{im \omega(t-t_p) + \mu} \]

Looking at the exponent we find again that the particle induces signals at the sidebands,

\[ \omega_{m\mu} = (m \pm \left| Q \right| \pm q) \Omega = (n \pm q) \Omega \]

These are the same as the beam response frequencies (3.16). Thus the beam "responds" and "talks" at the same frequencies.

Figure 20 shows the time and frequency domain picture of the pick-up signal of a single particle. This signal [Eq. (3.20)] will be used in calculating the coherent effect of the test particle upon itself. The incoherent effect due to the other particles can be obtained by adding up their currents (3.20) with a proper distribution in amplitude and phase.

In a coasting beam of N particles with random initial betatron phase and random time of arrival these induced currents add in square to give noise like signals at the frequencies (3.22).

Fig. 20 Time and frequency domain signal of a particle performing a betatron oscillation. A position sensitive pick-up records a short pulse at each traversal modulated in amplitude by the betatron oscillation. The frequency spectrum contains lines at the two sideband frequencies \((m\pm Q)f_0\) of each revolution harmonic \(mf_0\).
As different particles have slightly different revolution and betatron frequencies, these signals occur in bands with a spectral power density

$$\frac{dN_{sc}}{d\omega} = \frac{1}{2} \left( \frac{e}{p} \right) \frac{x_{rms}^2}{\hbar} \frac{dN}{d\omega}$$  \hspace{1cm} (3.23)$$

where dN/d\omega is the fraction of particles with sideband frequencies in a range of width d\omega around \omega = (n \pm q)\Omega. These are the Schottky noise bands discussed in D. Boussard's chapter in these proceedings. With a dispersion of revolution and betatron frequencies (\Delta \Omega and \Delta q respectively) the width of the band at (n \pm q)\Omega is (n \pm q)\Omega \pm \Delta q. A spectrum analyser usually records current, i.e. the square root of the signal (3.23). A practical example is given in Fig. 21, where the spectrum analyser picture of the signal from a horizontal pick-up is shown. The frequency band of this 'Schottky scan' is centered around a revolution harmonic nf_0 and contains the two sidebands (n \pm q)f_0. Note that the height of these sidebands [root of Eq. (3.23)] is proportional to the r.m.s. betatron amplitude x_{rms} and thus decreases during cooling.

All pick-up currents discussed here are the 'induced currents'. To obtain the true output signal one has to include the response functions of the pick-up structure and the acquisition system. Usually one aims at making these response functions as flat as possible.

Fig. 21  Example of a horizontal Schottky scan in LEAR at 600 MeV/c. The central band, the harmonic n = 100 of the revolution frequency, is visible as wide the beam is not completely centred at the position pick-up. The right and left bands are the sidebands (90n)\Omega f_0 and (103\Omega)\Omega f_0 where 0 = 2.3. During emittance cooling the sidebands decrease. The difference between the base line of the trace and the bottom line (zero signal) is given by the noise of the pick-up system. The span covers (approximately) half a revolution interval f_0. During horizontal cooling the height of the sidebands decreases.
3.6 Coherent effect

We can now have a fresh look at the motion of particle \( i \), retaining for the moment only its "self-terms". We take a kicker voltage on the r.h.s. of Eq. (3.13) which is proportional to the pick-up signal [Eq. (3.20)] of the circulating particle. As before we refer to all signals at their time of arrival at the kicker.

The particle takes a nominal time \( t_p \) and the cooling signal a time \( t_c \) to travel from pick-up to kicker. The electronic delay \( t_c \) is in general frequency dependent. We include this by a phase factor \( \varphi(\omega_m) = \omega_c t_c(\omega_m) \) in the development (3.20) of the pick-up signal. The frequencies \( \omega_q \) contained in this signal are the \( (n \pm q)\omega \) betatron sidebands (3.22). In our complex notation (including positive and negative frequencies) they simply appear as

\[
\omega_m = (m + Q)\omega; \quad m = \pm \infty\to \infty. \tag{3.24}
\]

The point to retain is that all signal transmission occurs at the frequencies of Eq. (3.24).

In addition to the initial phase \( \mu_0 \) [Eq. (3.19)], we have to include the phase-advance \( \mu = \omega_q t_p \) of the betatron oscillation of the particle on its way from pick-up to kicker. Hence the exponential factor in Eq. (3.20) referred to the kicker is written as,

\[
e^{i[\omega_p t - \mu + \varphi(\omega_m) + \omega_m s(t - t_p)]}.
\]

To complete the driving term in the betatron equation (3.13) we introduce a transfer function \( G(\omega_m) \). It has to include the pick-up response \( Z_p(\omega) \) [i.e. the voltage output for the induced current (3.21)], the transfer function of the cooling loop between pick-up and kicker (with cables, amplifiers, filters etc. taken into account) as well as the kicker response. Let us for simplicity also absorb the constant factors

\[
\frac{Q^2 \Delta}{2\pi} \sqrt{\frac{\beta_p \rho_p u}{\rho_0}}
\]

of Eq. (2.13) and the factor

\[
S_p \ e^{u \frac{1}{2m h}}
\]

of Eq. (3.20) into this transfer function but keep \( x_i = x_i \ e^{i(\omega_q t + \mu_0)} \) separate. Hence we rewrite Eq. (3.14) as

\[
\dot{x}_i + \omega_p^2 x_i = x_i \sum_{m=\pm \infty} G_{ii}(\omega_m) \ e^{-i\mu + i\omega t + i\varphi(\omega_m) - i\omega t_p} \sum_{k=\pm \infty} \ e^{i\omega t} \tag{3.25}
\]

We take \( G_{ii}(\omega_m) \) as entirely real and include all phase shifts in \( \varphi(\omega_m) \). Note that the second sum in Eq. (3.25) is the 'sampling term' - appearing already in Eq. (3.13) - due to the fact that the particle passes the short kicker once per turn. The first sum clearly is due to the localised nature of the pick-up.
Equation (3.25) is almost the same as Eq. (3.2) except that we include frequency dependent 'gain' $G_{ii}(\omega) e^{i\Phi(\omega)}$ and localised pick-up and kicker. The product of the two sums can readily be converted into a double sum

$$\sum_m^{\infty} e^{im\Omega t} \cdot \sum_{\lambda} e^{i\lambda t} = \sum_{m} \sum_{\lambda} e^{i(m+\lambda)t} ,$$

noting that in general

$$\sum_m a_m \cdot \sum_{\lambda} b_\lambda = \sum_m a_1 \cdot b_\lambda + \sum_m a_2 \cdot b_\lambda \ldots = \sum_m \sum_{\lambda} a_m \cdot b_\lambda .$$

Equation (3.25) may now be interpreted as an oscillator with a frequency shift that varies in time. An approximate solution to such equations is obtained by taking the time average of the frequency shift only, i.e. if we retain terms with \( \lambda = -m \) in Eq. (3.26) and drop the rapidly oscillating frequency shifts. Using this approximation (3.25) becomes:

$$\ddot{x}_i + \omega_p^2 x_i = x_i \sum_{m=\infty}^{\infty} G_{ii}(\omega_m) e^{-i\mu + i\Phi(\omega_m) - im\Omega t_p} .$$

This defines a change of betatron frequency:

$$\Delta \omega_p \sim \frac{1}{2\omega_p} \Delta \omega_p^2 = \sum_{m=\infty}^{\infty} G_{ii}(\omega_m) e^{-i\mu + i\Phi(\omega_m) - im\Omega t_p} .$$

As we assumed $x = \tilde{x} e^{i\lambda t}$ the damping rate is

$$1/\varepsilon = \text{Im}(\Delta \omega_p) = \sum_{m=\infty}^{\infty} G_{ii}(\omega_m) e^{-i\mu - im\Omega t_p} .$$

Optimum cooling is given by

$$1/\varepsilon = \frac{1}{\Delta \omega_p} \sum_{m=\infty}^{\infty} |G_{ii}(\omega_m)|$$

and obtained for all $m$, if the phase factor is properly chosen such that $-\text{Im}(e^{i\lambda}) = 1$. This requires

$$\Phi(\omega_m) - \mu - m\Omega t_p = -\frac{\pi}{2} \text{ modulo } 2\pi .$$

Usually Eq. (3.29) is satisfied by putting the kicker at the 'proper' betatron phase advance from the pick-up, i.e.
\[ \mu = \frac{\pi}{2}, \text{ or } \frac{3}{2} \pi \] with signal inversion, or \( \frac{5}{2} \pi \) ...

and designing \( \psi(\omega_m) \) to be as close as possible to \( m \Delta \tau_p \):

\[ \psi(\omega_m) - m \Delta \tau_c = m \Delta \tau_p = 0. \quad (3.30) \]

Ideally this requires a signal delay of the cooling loop, \( t_c(\omega_m) \), equal to \( t_p \), the particle travelling time pick-up to kicker, independent of frequency. If the optimum spacing pick-up to kicker (\( \mu = \pi/2 \) mod. \( \pi \)) is not possible one can in principle include filters (as first proposed by Thorndahl) with a time delay characteristic \( t_c(\omega) \) such that Eq. (3.29) is still satisfied. This requires however "steep" filters with a phase delay \( \psi(\omega) \) varying by \( 2(\mu-\pi/2) \) from the \( n\pi \) to the \( n\pi \) betatron band. In fact writing Eq. (3.29) in the form

\[ \psi(\omega_m) - m \Delta \tau_c = (\mu - \frac{\pi}{2}) = \delta \mu \]

and noting that for any network \( \psi(\omega) = -\psi(\omega) \) we need for positive \( m \) (corresponding to the \( n+q \) sidebands):

\[ \psi(\omega) - m \Delta \tau_p = \delta \mu \]

and for negative \( m \) (corresponding to the \( n-q \) bands)

\[ \psi(\omega) - m \Delta \tau_p = -\delta \mu \]

thus requiring a phase difference \( 2\delta \mu \) between neighbouring bands.

Other effects can be identified from Eq. (3.28):

- The time-of-flight error \( \Delta t_p \) of a particle ("mixing between pick-up and kicker") as well as improper delay \( \Delta t_c \) of the cooling loop or improper pick-up to kicker betatron phase advance \( \delta \mu = \mu - \pi/2 \) appear as a phase factor in Eq. (3.28) which may be rewritten as

\[ \frac{1}{\tau} = \frac{1}{2 \omega \beta} \sum_{\chi=-\infty}^{\infty} G_{ii}(\omega) \cos (\delta \theta_m - \delta \mu) \quad (3.28b) \]

where \( \delta \theta_m = \psi(\omega_m) - m \Delta \tau_p = m \Delta (t_c - t_p) \).

- The fall-off of the pick-up-current spectrum at high frequency can be included in the expansion (3.21) and absorbed into the transfer function. The similar effect of the finite kicker length can be included in much the same manner via the expansion on the r.h.s. of Eq. (3.13). Finally we remark that Eq. (3.28) can be written in various other forms involving sums over positive \( m \) only which clearly reveal the \( (n\pi) \) bands. This is left as an exercise to those interested.

Having established the interaction of the particle with itself we next include the noise due to the other particles and the electronic system.
3.7 Noise

Noise will be treated in detail in G. Döme's chapter in these proceedings. For convenience we repeat the essentials here, which are useful to include the incoherent effect in the frequency-domain analysis of stochastic cooling.

Look at an oscilloscope picture like Fig. 22 which displays a pick-up signal \( u(t) \) when 'no beam is in the machine', i.e. the electronic noise of the system. It is customary to represent the mean square (averaged over a long enough time \( T \)) of such noisy voltages by a pseudo Fourier transformation

\[
\overline{u^2(t)} = \int \phi(\omega) \, d\omega.
\]

![Fig. 22 Noise signal on a pick-up](image)

The 'spectral power density function' \( \phi(\omega) \) is closely related to a Fourier development of \( u(t) \). In all practical applications the noisy voltage has been 'switched on' at some time \( t = 0 \) and we regard it up to \( t = T \). Outside this range the waveform is irrelevant, so, for the purpose of computation we can periodically continue it (Fig. 23). We then deal with a periodic function \( u(t \pm nT) = u(t) \) which we can Fourier-expand in the usual way

\[
u(t) = \sum_{m=-\infty}^{\infty} \tilde{u}_m e^{im\omega_0 t}; \quad \omega_0 = 2\pi/T.
\]

(3.31)

![Fig. 23 A noisy voltage u(t) observed from time t = 0 to t = T and its periodic continuation outside this range to permit a Fourier development](image)
The Fourier amplitudes:

\[ u_m = \frac{1}{T} \int_0^T u(t) e^{-i \omega_0 t} \, dt \quad (3.32) \]

are in general complex but for real \( u(t) \) \( u_m \) is the conjugate of \( u_m \). The mean square of Eq. (3.31) over the observation time \( T \) is by definition

\[ \overline{u^2(t)} = \frac{1}{T} \int_0^T u^2(t) \, dt \quad (3.33) \]

which yields after some calculation (transforming the square of the sum into a double sum similar to the analysis in conjunction with Eq. (3.26) and noting that averaged over a period all \( e^{i \omega_0 t} \) terms vanish except for \( k = 0 \))

\[ \overline{u^2(t)} = \sum_{m=0}^\infty \overline{u_m} \overline{u}_{-m} = \sum_{m=0}^\infty |u_m|^2 . \quad (3.34) \]

This is known as Perseval's equation in the theory of Fourier series; it applies to any Fourier development! Equation (3.34) presents the 'average noise power' \( u^2(t) \) as the sum of its spectral contributions at frequencies \( \omega_0 = n \omega / T \). Analyzed over shorter or longer time \( T \) the spectra are as sketched in Fig. 24.

![Fig. 24](image)

**Fig. 24** The power spectrum \( |u^2(\omega)| \) of a noisy voltage which is observed for a time interval \( T \) and periodically continued outside this interval. Increasing \( T \) the height of the spectral lines decrease proportionally to their spacing \( \omega_0 = 2 \pi / T \) so that the quantity \( |u^2(\omega)| / \omega_0 \) remains the same. In the limit \( T \to \infty \) one has a continuous spectrum where \( u^2(\omega) / \omega_0 \) is the spectral power density function of Eq. (3.30).

As the sum of the rays is \( \overline{u^2(t)} \) in both cases their height scales proportional to their spacing (for large \( T \)), i.e. \( u^2_m / \omega_0 = \text{constant} \).

For very large observation time \( T \to \infty \) the spectrum is practically continuous and the sum Eq. (3.34) is approximated by an integral of the form of Eq. (3.30):

\[ \overline{u^2(t)} = \sum_{m=0}^\infty |u_m|^2 = \int \frac{1}{\omega_0} \left| \tilde{u}(\omega) \right|^2 d\omega + \int \omega_0 \left| \tilde{u}_{-m}(\omega) \right|^2 d\omega . \quad (3.35) \]
Hence we identify for \( T \to \infty \)

\[
\frac{\hat{u}(\omega)}{\omega_0} \to \phi(\omega) .
\]  

(3.36)

This interpretation permits us to calculate (at least in simple cases) \( \phi(\omega) \) and to establish the following important theorem:

When noise with a spectral power density \( \phi_1(\omega) \) is transmitted through a linear system with a (complex) transfer function \( H(\omega) \) then the power spectrum at the output is

\[
\phi_2(\omega) = |H(\omega)|^2 \phi_1(\omega)
\]  

(3.37)

This follows immediately from the preceding noting that each of the components of the r.h.s. of Eq. (3.31) when transmitted through the network transforms according to Eq. (3.36):

\[ \tilde{u}_{m2} = \tilde{H}(m\omega_0) \cdot \tilde{u}_{m1} . \]

An example of the theorem (3.37) is the transformation of 'broadband noise' into band limited noise by a band pass filter (Fig. 25).

Fig. 25  An example of the theorem [Eq. (3.37)]: broadband noise transmitted through an ideal band pass filter is converted into a band-limited noise

This is as much as we need about noise for the purpose of this chapter. Many other fascinating facts will be discussed in Georges Döme's presentation.

3.8  Beam response to a noisy kicker

Consider a linear oscillator driven by a noisy excitation \( u(t) \) with power density \( \phi(\omega) \):

\[
\ddot{x} + \omega^2 x = u(t) .
\]  

(3.38)
This problem was treated (in a more general context) a quarter of a century ago in a classical paper by Hereward and Johnsen\cite{12}.

Their result, the "Hereward-Johnsen theorem" may - in our present case - be stated as follows:

For a particle injected at $t = 0$, the square of the amplitude $\tilde{x}$ of $x$, Eq. (3.38) expected (Fig. 26) at time $t$ (large) is

$$\tilde{x}^2 = \frac{2\pi}{\omega_0^2} \cdot \phi(\omega_0) \cdot t . \tag{3.39}$$

In words: the amplitude grows in a diffusion like manner ($\tilde{x} = \sqrt{t}$) at a rate which is determined by the spectral density of the noise at the resonance frequency $\omega_0$.

\begin{center}
\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig26.png}
\caption{Amplitude $\tilde{x}$ of betatron oscillation of a particle driven by a noisy kicker. The expectation value (i.e. the average) of $x^2$ grows linearly in time at a rate given by the spectral density of the noise at the resonance frequency. In addition to this average growth there is a fluctuating motion which is of little importance for the long term behaviour.}
\end{figure}
\end{center}

Equation (3.39) is for a simple harmonic oscillator. If we inject noise into the cooling loop (or directly onto the kicker) we have again to include the "sampling factor" $\sum e^{j\Omega t}$ of Eq. (3.13) because we use a short localized kicker. Thus we use

$$\tilde{x} + \omega_0^2 \tilde{x} = u(t) \sum_{\Omega=-\infty}^{\infty} e^{j\Omega(t-t_p)} . \tag{3.40}$$

The effect of each component, $e^{j\Omega t}$, is to 'shift' the frequency content of the driving force $\omega + \omega + \Omega$. In this sense we may interpret the r.h.s. of Eq. (3.40) as a sum of noisy driving forces with frequencies

$$\omega_d = \omega + \Omega.$$
We can apply the Hereward-Johnsen result to each of these bands (noting that in working out \( x^2(t) \) cross-terms between bands average to zero), hence the response of Eq. (3.40) to a noise \( u(t) \) is:

\[
\hat{x}^2 = \frac{2\pi}{\omega_p}\beta t \sum_{\lambda=-\infty}^{\infty} \phi(\omega_p + \lambda\Omega).
\]  

Equivalently if you prefer to work with positive frequencies only you may write [using \( \phi(-\omega) = \phi(\omega) \)], Eq. (3.41) as

\[
\hat{x}^2 = \frac{2\pi}{\omega_p}\beta t \left[ \sum_{\lambda=0}^{\infty} \phi(\lambda + q)\Omega + \sum_{\lambda=1}^{\infty} \phi(\lambda - q)\Omega \right].
\]  

This clearly presents the amplitude growth in terms of the spectral density of the noise at the betatron sidebands.

In working out the long term average of \( x^2(t) \) leading to Eq. (3.41) and Eq. (3.42) we have assumed that cross terms between different bands \( \lambda \) average to zero and also that \( \phi(-\omega) = \phi(\omega) \). A justification [which can be carried through, e.g. using an expansion of the type (3.31) for \( u(t) \) and the definition (3.36) of \( \phi(\omega) \)] is left as an exercise to those interested.

3.9 Back to beam Schottky noise and amplifier noise

Returning to a test-particle: apart from its self-term it will experience the 'beam Schottky noise' due to the presence of the other particles and the electronic noise of the preamplifier etc. The spectral density \( \phi(\omega) = dI^2_{SC}/du \) inside a Schottky band is determined by Eq. (3.23) and may be rewritten here (see D. Boussard's chapter on Schottky noise):

\[
\phi_{SC}(\omega) = \phi_0 M(\omega_k)
\]  

\[
\omega_k = (\lambda \pm q)\Omega, \quad \lambda = 1, 2, 3 \ldots
\]  

\[
\phi_0 = \frac{N e^2\Omega}{2N} \frac{x^2_{rms}}{\beta} \frac{\Omega}{(2\pi)^2 H^2}
\]

\[
M(\omega_k) = \begin{cases} 
\frac{Q}{2N} \frac{dN}{d\omega_k} = \frac{Q}{2\Delta\omega_k} & \text{if all bands are separated, i.e. } \Delta\omega_k < 2q\Omega \\
1 & \text{for complete band overlap, i.e. } \Delta\omega_k > \Omega.
\end{cases}
\]

Here \( x^2_{rms} \) is the mean-square-betatron amplitude of particles with a sideband frequency near \( \omega_k \), \( dN/d\omega_k \) is the number of particles with frequency in a range of width \( d\omega \) near \( \omega_k \).

The dimensionless quantity \( M(\omega) \) is closely related to mixing as will become clear later. It is customary to approximate \( dN/d\omega \) by a rectangular distribution \( N/\Delta\omega \) of total width.
\( \Delta \omega(x) = \delta(|x| \pm q) \omega = (|x| \pm q) \Delta \omega \pm \omega q = |x| \Delta \omega \) \hspace{1cm} (3.44)

where \( \Delta \omega(x) \) is the width of the sidebands at the \( x \)-th harmonic. This leads to the approximations for \( M(\omega_k) \) indicated under Eq. (3.43) and sketched in Fig. 27.

\[ M_i = \frac{\Omega}{\Delta \omega(i)}, \quad \Delta \omega(i) < 2q \]

\[ \Omega > \Delta \omega(i) > 2q \]

\[ M_i = \frac{\Omega}{\Delta \omega(i)} \]

\[ \Delta \omega(i) > \Omega \]

\[ M_i = 1 \]

Fig. 27 Transverse beam Schottky noise for separated bands (low frequency), partial overlap (intermediate frequency) and complete band overlap.

For high frequencies, bands completely overlap and the noise has a continuous power spectrum with density \( \phi(\omega) = \phi_0 \) as given by Eq. (3.43) with \( M = 1 \). Note that in this limit Eq. (3.43) is just the classical Schottky formula for the noise of a "DC-current" \( I_0 \). Defining the noise density \( \phi = dI^2/d\omega \) with respect to angular frequency \( \omega \) as we consistently do in the present chapter, this formula writes as:

\[ \phi(\omega) = \phi_0 \frac{e}{\pi} \cdot I_0. \] \hspace{1cm} (3.45)

In our case the circulating current is \( I_0 = \frac{N\phi}{2\pi} \) and \( 0.5 (S_p \chi_{\text{rms}}/h)^2 \) enters as we take the betatron oscillation signal from a difference pick-up.
The noise is transmitted through the cooling loop in the same way as the 'self-signal' of the test-particle. Hence the transfer function is the same except for phase factors due to the different arrival times of particles and due to different betatron phase. We denote this transfer function, which has the same modulus as $G_{ij}(\omega)$ and $G_{ij}(\omega)$, by $G(\omega)$.

By virtue of Eq. (3.37) the noise density on the kicker is therefore

$$\phi_\rho(\omega) = |G(\omega)|^2 \phi_0' M(\omega)$$  \hspace{1cm} (3.46)

$$\phi_0' = \frac{N}{\Omega} \chi^2_{\text{rms}}.$$

Here $\phi_0'$ is the same as $\phi_0$, Eq. (3.43) except for the factor $(S_p \rho Q/2\hbar)^2$ which was absorbed into $G^2(\omega)$ as before, see the discussion preceding Eq. (3.25).

We now turn to the electronic noise and assume that -- referred to at the entrance of the cooling loop (exit of the pick-up) -- it has a power spectrum $\phi_\delta(\omega)$. Let the transfer function from this point to the kicker be $H(\omega)$. Clearly this $H(\omega)$ is the same as $G(\omega)$ except for the pick-up response function. The noise seen by the test particle (j) (Fig. 28) is then

$$\phi(\omega_j) = |G(\omega_j)|^2 \phi_0^* M(\omega_j) + |H(\omega_j)|^2 \phi_\delta^*(\omega_j)$$  \hspace{1cm} (3.47)

Here $\phi_\delta^* = \phi_\delta/(S_p \rho Q/2\hbar)^2$ is the electronic noise reduced by the same factor as $\phi_0$ for consistency. To recover previous results it is useful to rewrite Eq. (3.47) as

$$\phi(\omega) = |G(\omega)|^2 \phi_0^* [M(\omega) + U(\omega)]$$  \hspace{1cm} (3.47a)

$$U(\omega) = \frac{|H(\omega)|^2 \phi_\delta^*(\omega)}{|G(\omega)|^2 \phi_0}.$$

Clearly $U(\omega)$ is the ratio of amplifier to beam noise (the latter in the high frequency limit where bands overlap). Using Eq. (3.41) we can write down the expectation value for the amplitude of the test particle as driven by the noise Eq. (3.47)

$$\chi^2_j(t) = \frac{2\pi}{\omega_0} \beta t \sum_{\chi=\omega_0} |G(\omega_j)|^2 [M(\omega_j) + U(\omega_j)]$$  \hspace{1cm} (3.48)

with $\phi_0$ as given under Eq. (3.46); $M(\omega)$ as given under Eq. (3.43) and $U(\omega_j)$ as given under Eq. (3.47a) and $\omega_0 = (t + q_j)\beta$ is now the sideband frequency of particle j.

Equation (3.48) determines the 'incoherent effect' as experienced by the test particle.
Fig. 28 The noise seen by a test particle is the Schottky noise due to the other particles and the electronic noise of the amplifier etc. At low frequency the Schottky noise occurs in bands with a density $M$ times higher than in the situation of complete overlap. This increase of noise density corresponds to enhanced 'heating of the test particle due to bad mixing'.

3.10 Cooling rate

We can now calculate the expected amplitude $\tilde{x}^2_j(t)$ of the test particle ($j$) by adding up the coherent and the incoherent effects. To use Eq. (3.28) for $d(x^2)/dt$ we note that in general

$$\frac{1}{x^2} \frac{d(x^2)}{dt} = \frac{2}{x} \frac{dx}{dt} = -\frac{2}{\tau}.$$

Hence we have from Eq. (3.28) for the coherent (damping) effect:

$$\frac{d(\tilde{x}^2_j)}{dt}_{\text{coh}} = -\tilde{x}^2_j \sum_{\omega_k} \Re \left[ G(\omega_k) e^{i\phi(\omega_k)} \right] \frac{1}{2}.$$  \hspace{1.0cm} (3.49)

For the incoherent (heating) effect we rewrite Eq. (3.48) substituting $\phi_0'$ from Eq. (3.46):

$$\frac{d(\tilde{x}^2_j)}{dt}_{\text{incoh}} = \tilde{x}^2_{\text{rms}} \sum_{\omega_k} \left| G(\omega_k) \right|^2 \left[ M(\omega_k) + U(\omega_k) \right] .$$ \hspace{1.0cm} (3.50)

The resultant cooling equation is:

$$\frac{d(x^2_j)}{dt} = \frac{d(x^2_j)}{dt}_{\text{coh}} + \frac{d(x^2_j)}{dt}_{\text{incoh}}.$$  \hspace{1.0cm} (3.51)
Equations (3.49) to (3.51) represent cooling as a sum of the contributions at the sideband frequency \( \omega_k = (\lambda + q)Q_j \) of the particle. In this form all frequency characteristics of the cooling loop can readily be included. This is especially handy for those who like to measure and calculate in the frequency domain. In addition we can rediscover and re-interpret the effects discussed before.

1. The influence of imperfect synchronisation of particle and cooling signal ('mixing pick-up to kicker'), \( \delta \Phi \neq 0 \). It enters as a phase error in the coherent term.

2. The influence of betatron phase errors (imperfect spacing) pick-up to kicker, \( \delta \mu \neq 0 \). It enters as another phase error in the coherent term.

3. Imperfect mixing on the way kicker to pick-up; expressed here as enhancement (\( M > 1 \)) of the heating by Schottky noise which is concentrated in bands and hence increased in density. Good mixing (\( M = 1 \)) corresponds to overlap of Schottky bands.

4. Amplifier (and other electronic) noise, \( U > 0 \).

Equation (3.49) so far is for any test particle. To obtain the damping rate for the mean square amplitude \( x_{\text{rms}}^2 \) we have to average Eq. (3.49) and Eq. (3.50) over the frequency distribution of the beam particles. In the simple case of perfect \( \delta \Phi = 0 \), \( \delta \mu = 0 \), \( M = 1 \) and constant \( G(\omega) = G \), \( U(\omega) = U \) inside the passband you can rediscover the familiar

\[
- \frac{1}{x_{\text{rms}}^2} \frac{d(x_{\text{rms}}^2)}{dt} = \frac{W}{N} \left[ 2g - g^2 (1 + U) \right] \tag{3.52}
\]

by calling

\[
\frac{G \cdot N}{\omega_p^2 \Omega/2\pi} = g. \tag{3.53}
\]

To work out the sums over \( \lambda \), note that with a passband of width \( \Delta f = W \) in the positive frequency plane the number of betatron lines contributing is \( 2 \lambda 2\pi W/Q \) as sketched in Fig. 29, namely \( 2\pi W/Q \) for positive \( \lambda \) (the \( n+q \) bands) and the same number again for negative \( \lambda \) (the \( n-q \) bands).

Fig. 29 Passband of cooling system in the \( f = -\) to \( f = + \) frequency plane. There are \( 2\pi W/Q = W/f_0 \) lines \( \omega_k = (\lambda + q)Q \) in the passband at negative frequencies (negative \( \lambda \)) and the same number in the passband at positive frequency (positive \( \lambda \)).
You can generalise Eq. (3.52) to include mixing factors, betatron phase errors and frequency dependence in $G$ and $U$ by interpreting the sums in Eq. (3.49) as averages over the passband. With this interpretation you may write Eq. (3.52) in various different forms useful for comparison with previous results, for instance:

$$\frac{d\langle\hat{x}^2\rangle}{dt} = \frac{W}{N} \langle -2\dot{x}^2 < \Re (g(\omega) e^{i(\delta \phi - \delta \mu)}) \rangle_{\text{passband}} + \hat{x}_{\text{rms}}^2 < |g^2(\omega)| \langle |M(\omega) + U(\omega)| \rangle_{\text{passband}} \rangle$$

(3.52a)

with

$$g(\omega) \equiv \frac{N \cdot G(\omega)}{\omega \cdot \bar{\rho}/2\pi}$$

$$\langle ... \rangle_{\text{passband}} = \frac{\omega}{2\pi} \int_{\omega - \omega}^{\omega + \omega} ...$$

3.11 Feedback via the beam and signal shielding

We shall now attempt to introduce a final ingredient of cooling theory known as 'feedback via the beam' or 'signal shielding'. Although this refinement will change our previous results by at most a factor of 2, the change of the beam Schottky signals when the cooling loop is closed has become an important diagnostics tool\textsuperscript{[13]}

Where did we miss out this effect in our treatment so far? In fact considering the test-particle equation

$$\ddot{x}_j + \omega_p^2 x_j = G_{jj} x_j + \sum_i G_{ij} x_i + \text{'system noise'}$$

\text{effect of particle upon itself: coherent term with zero average}

\text{effect of other particles: Schottky noise, fluctuating average}

\text{amplifier noise etc. fluctuating with zero average}

O.K. Not O.K. O.K.

We have described the effect of the other particles $\sum_i G_{ij} x_i$ as Schottky noise of an undisturbed beam, i.e. as a fluctuating term with zero time average. This assumption is not generally correct. F. Sacherer\textsuperscript{[14]} has pointed out that - in the case of poor mixing - $\sum_i G_{ij} x_i$ does lead to a coherent oscillation with finite average amplitude. The fluctuation occurs around this average amplitude and not around zero as it would be the case in an undisturbed beam. The reason is that part of the "modulation" imposed at the kicker is still present at the pick-up and re-enters the loop as sketched in Fig. 30. Thus the noise on a beam subject to cooling is different from the free beam noise. The feedback of the cooling signals via the beam changes all ingredients of the analysis, namely beam noise as well as the influence of the coherent term and the amplifier noise.
Fig. 30 Cooling system including the coherent beam modulation \( x_0 \) imposed at the kicker and partially preserved up to the pick-up due to imperfect mixing. The lower diagram shows Sacherer's equivalent feedback loop. Amplifier noise \( (x_N) \) and Schottky noise \( (x_s) \) are random noises whereas the coherent modulation is fed back via the beam from kicker to the pick-up. This feedback changes the open loop response to \( x_N + x_b \) by a complex transfer function \( f(\omega) \) which depends on the amplification (cooling strength) and the degree of mixing between kicker and pick-up.

Fortunately F. Sacherer has also shown the road to rescue our previous results. The way out he provides is a beautiful piece of accelerator theory.

As a pre-exercise: consider a system of \( N \) oscillators with a harmonic driving force and a 'collective force' proportional to the average displacement of the oscillators. Take for the \( g \)-th particle - sorry - oscillator

\[
\ddot{x}_j + \omega_j^2 x_j = \tilde{v}(\omega) e^{i\omega t} + \frac{\tilde{G}_k(\omega) x_k}{N}
\]

(3.55)

Here the term \( \langle \tilde{G}_k x_k \rangle \) may be interpreted as the weighted contribution of particle \( k \) to the average

\[
\langle \tilde{G}_k(\omega) x_k \rangle = \frac{1}{N} \sum \tilde{G}_k(\omega) x_k.
\]
A 'mechanical' and an 'electrical' analog of Eq. (3.55) are sketched in Figs. 31 and 32.

**Fig. 31** A 'mechanical analog' of Eq. (3.55). Person V tries to excite a system of oscillators (masses on springs) by shaking their point of suspension. Person G tries to damp the motion by observing the average displacement \( \langle x \rangle \) of the oscillators and applying a damping force \( G(x) \).

\[ \ddot{x}_j + \omega_j^2 x_j = V e^{i\omega t} + G(x) \]

**Fig. 32** An electrical analog of Eq. (3.55). A group of LC-resonators is driven by a voltage \( V e^{i\omega t} \). The sum \( I = \sum I_k \) of the currents through the resonators is fed back through an amplifier with gain \( G \) to add an input voltage \( G \sum I_k \).

\[ \dot{I}_j + \omega_j^2 I_j = \frac{L_j V}{L_j} + \frac{I_j G}{L_j} \langle I_k \rangle \]
To solve Eq. (3.55) we insert a trial solution

\[ x_j = \tilde{x}_j e^{i\omega t}; \quad \langle \tilde{a}_k(\omega) x_k \rangle = \langle \tilde{a}_k(\omega) \tilde{x}_k \rangle e^{i\omega t}. \]

To find

\[ \tilde{x}_j = \frac{1}{\omega_j^2 - \omega^2} [\tilde{v}(\omega) + \langle \tilde{a}_k \tilde{x}_k \rangle] \]

\[ \sim \]

multiply with \( G_j \) and average both sides. Call:

\[ \langle \tilde{a}_j \tilde{x}_j \rangle = \sim S(\omega) \]

(3.57)

and solve for the average

\[ \langle \tilde{a}_j \tilde{x}_j \rangle = \frac{1}{N} \sum_{j=1}^{N} G_j \tilde{x}_j = \langle \tilde{a}_k \tilde{x}_k \rangle \]

to find

\[ \langle \tilde{a}_k \tilde{x}_k \rangle = \frac{\tilde{S}(\omega)}{1 - \tilde{S}} \tilde{v}(\omega). \]

(3.58)

Thus we do have a finite coherent amplitude \( \langle \sim \rangle \). We can now use Eq. (3.58) to eliminate the 'collective force' term from Eq. (3.55). We find

\[ \ddot{x}_j + \omega_j^2 x_j = \tilde{v}(\omega) e^{i\omega t} \left[ 1 + \frac{\tilde{S}(\omega)}{1 - \tilde{S}} \right]. \]

(3.55a)

\[ \text{shielding factor} \]

Instead of treating the original Eq. (3.55) with a

\[ \text{r.h.s.} = [\text{driving force}] + [\text{weighted average displacement}] \]

we can therefore treat the same equation with the more convenient:

\[ \text{r.h.s.} = [\text{driving force}] \times [\text{shielding factor}]. \]

This is the essence of 'Sacherer's trick'. In the cooling equation we shall want to replace for each of the betatron bands involved:
r.h.s. = [coherent term] + N*[weighted average displacement] + [Schottky noise] + [amplifier noise]

by

r.h.s. = [coherent term + Schottky noise + amplifier noise]*[shielding factor].

A quantity of key importance is the 'dispersion function' \( \tilde{S}(\omega) \) entering into the shielding factor

\[
\tilde{T}(\omega) \equiv \left[ 1 + \frac{\tilde{S}}{1 - \tilde{S}} \right] = \left[ \frac{1}{1 - \tilde{S}} \right].
\]

(3.59)

For large \( N \) we have

\[
\tilde{S}(\omega) = \frac{1}{N} \int \frac{G(\omega_j)}{\omega_j - \omega^2} + \int \frac{G(\omega_j)n(\omega_j)}}{\omega_j - \omega^2} \, d\omega_j = \frac{1}{2\omega} \int \frac{G(\omega_j)n(\omega_j)}}{\omega_j - \omega^2} \, d\omega_j.
\]

(3.60)

Here \( n(\omega_j) \, d\omega_j \) is the fraction of particles with eigenfrequencies in a band of width \( d\omega_j \) near \( \omega_j \).

Dispersion integrals of the type (3.60) are treated in H.G. Hereward's chapter on Landau damping\(^{15}\)). For convenience some features are repeated in Appendix 2. Due to the pole, the integral has an imaginary part even if \( G(\omega_j) \) is real. Details depend on the distribution \( n(\omega_j) \) of eigenfrequencies and on \( G(\omega) \). A typical behaviour of \( \tilde{S}(\omega) \) is sketched in Fig. 33.

![Fig. 33 Frequency distribution and typical behaviour of the dispersion function Eq. (3.60) for a given function \( G(\omega) \) which is constant (or slowly varying) near the beam response frequency \( \omega_p \). This behaviour of \( G(\omega) \) is required for betatron cooling.](image-url)
A useful approximation is

\[ S(\omega) \approx \frac{\pi}{2\hbar \beta} \left[ 1 - i \frac{\omega - \omega_p}{\Delta \omega} \right] n(\omega) \]

\[ n(\omega) = \begin{cases} \frac{1}{\Delta \omega} & \text{for } |\omega - \omega_p| < \Delta \omega/2 \\ 0 & \text{for all other values} \end{cases} \]  

(3.61)

\[ \omega_p = \langle \omega_j \rangle = \frac{1}{N} \sum_{j=1}^{N} \omega_j \quad \text{average eigenfrequency} \]

To go one step further we now analyze a problem which is of some practical importance namely beam excitation by a single harmonic driving force on a kicker when the cooling loop is closed. We write the equation of motion of particle \( j \) as

\[ \ddot{x}_j + \omega_j^2 x_j = \left[ \sum_k (G(\omega)) \sum_m e^{im\Omega_k(t-t_k)-imk} e^{-\frac{1}{2}m^2}\right] \cdot \sum_{k \neq \omega} e^{-i\Omega_j(t-t_j-t_p)}. \]  

(3.62)

Here the first sum \( (k) \) is over the \( N \) beam particles, the sum over \( m \) is the 'sampling term' due to the localized pick-up and the sum over \( \lambda \) represents the harmonics of the localized kick, \( t_k \) is the arrival time of particle \( k \) at the pick-up, \( x_k = x_k e^{-i\Omega_k} \) presents the transformation of its oscillation from pick-up to kicker, \( \phi_m(\omega) = m\omega_t G(\omega) \) is the signal delay of the cooling loop, \( t_p \) is the travelling time of particle \( j \) from pick-up to kicker, hence its arrival time at the kicker is \( t_j + t_p \), \( V e^{i\omega t} \) is the external driving force, the term proportional to \( G(\omega) \) is the corresponding 'driving force' given by the response of the cooling loop to the beam oscillation.

Once again we drop all rapidly varying 'frequency shifts', i.e. we only take harmonics with \( m = \lambda \) in the first term on the r.h.s. of Eq. (3.62).

In the second term we only retain frequencies \( \omega \pm \Delta \Omega = \omega_p \pm \omega_j \) close to resonance. We assume that all bands are well separated so that only one \( \lambda \) leads to resonance. Thus we simplify Eq. (3.62) to

\[ x_j + \omega_j^2 x_j = \left[ \sum_k x_k G(\omega) e^{i\Omega_k(t-t_k)-imk} e^{i\phi(\omega)} + V e^{i\omega t} \right] e^{-i\Omega_j(t-t_j-t_p)}. \]  

(3.62a)

As response to the driving term \( V e^{i\omega t} \) we expect a solution of the form

\[ x_j \sim x_j e^{i(\omega - \Delta \omega_j)t + i\Omega_j(t_j + t_p)} \]

for any particle \( j \).
Upon substitution [using the corresponding expression for \( x_k \) on the r.h.s. of Eq. (3.62a)] we find:

\[
\bar{x}_j = \frac{1}{\omega_j^2 - (\omega - \omega_j)^2} \left( N \bar{G} + V \right)
\]  

(3.63)

where we define

\[
\bar{G} = \frac{1}{N} \sum_{k=1}^{N} \bar{G}_k(\omega)
\]

\[
\bar{G}_k(\omega) = G(\omega) e^{-i\phi_k}
\]

The quantity \( \delta \theta_k = \omega(\omega_j - \omega k) \) is the synchronisation error between particle \( k \) and the cooling signal. Let us denote the resonant driving frequency - as introduced already in Eq. (3.24) - by

\[
\omega = \omega_j + \omega Q_j = (Q_j + 1) \omega_j
\]

and use \( \omega^2 - (\omega - \omega_j)^2 = 2\omega_j [\omega_j - (\omega - \omega_j)] \).

From the preceding analysis we can now define a shielding factor for the present situation

\[
\bar{T}^{\omega_j}(\omega) = \left[ \frac{1}{1 - N \bar{S}(\omega_j)} \right]
\]

\[
\bar{S} = \frac{1}{2\omega_0} \int \frac{\bar{G}(\omega) n(\omega)}{(\omega_j + \omega Q_j) - \omega} d\omega
\]

(3.64)

\[
\omega_0 = \langle \omega_j \rangle : \text{beam average of betatron frequency.}
\]

Using this shielding factor we can rewrite Eq. (3.62a) as

\[
\ddot{x}_j + \omega_j^2 x_j = \bar{T}^{\omega_j}(\omega_j + V) e^{i\omega_j t - i\alpha_j}
\]

(3.62b)

where we use \( \alpha_j = -\omega Q_j (t_j + t_p) \) to denote the phase factor due to the arrival time of particle \( j \) at the kicker. Thus when the cooling loop is closed, the response of Eq. (3.62) to \( V e^{i\omega t} \) changes by \( \bar{T}(\omega_j) \). In this way \( \bar{T}(\omega_j) \) can be observed and \( \bar{G}(\omega) \) can be deduced from it. Usually these measurements are done using a network analyser to display the beam response to a swept sine wave (beam transfer function measurement) as sketched in Fig. 34. This permits us to adjust the characteristics of the cooling loop band by band.

To complete our analysis we return to Eq. (3.62) but now assume a general driving force represented by a Fourier series (or a Fourier integral) with a spectral density function \( V(\omega) \). We invoke superposition and resonant behaviour of the betatron equation at the frequencies \( \omega_k \). Thus we rewrite Eq. (3.62b) as
Fig. 34 Arrangement to measure beam transfer function. The frequency sweep of the network analyser is set to cover one or several betatron sidebands. The difference in beam response with cooling loop open and closed can be used to optimise the loop gain.

\[
\ddot{x}_j + \omega_j^2 x_j = \sum_{k=-\infty}^{\infty} \tilde{T}(\omega_k) V(\omega_k) e^{i\omega t - i\alpha_k}
\]  \hspace{1cm} (3.62b)

which presents the effect as the sum of the interaction at the sidebands \( \omega_k = (k+q)\Omega \). As a consequence of the beam feedback each band now has its proper shielding factor \( \tilde{T}(\omega_k) \), Eq. (3.64) (well separated bands, i.e. poor mixing assumed). The effect of the shielding factor is fully equivalent to introducing a transfer function \( \tilde{T}(\omega) \) between the driver and the kicker.

We can now generalize the cooling rate in Eqs. (3.49) and Eq. (3.50) to include shielding. We can interpret \( V(t) \), Eq. (3.65) as the cooling signals discussed before (namely the self-effect of the test particle, the Schottky noise due to the other particles and the amplifier noise). Since the beam feedback acts like a transfer function we simply include this into Eqs. (3.49) and Eq. (3.50) by substituting

\[
\tilde{G}(\omega_k) + \tilde{T}(\omega_k) \tilde{G}(\omega_k)
\]  \hspace{1cm} (3.65)

A typical behaviour of the shielding function is sketched in Fig. 35. Note that for small 'gain' (N.G small) and small \( S(\omega) \) the shielding factor is close to 1.

![Graph](image_url)  \hspace{1cm} Fig. 35 Typical behaviour of the shielding factor \( \tilde{T}(\omega) \) near a resonance frequency of the beam
To gain further insight we only look at particles near the centre of the distribution ($\omega_k = \omega_j + \kappa \omega_j$) and assume perfect betatron phase and perfect signal delay pick-up to kicker ($\delta \mu = 0, \delta \varphi = 0$). Then the gain function $G(\omega_k) = \frac{|G(\omega_j)|}{\omega_k \Omega/2\pi} e^{-\ii \kappa/2 - \delta \mu + \delta \varphi - \ii 1|G(\omega_j)|}$ becomes purely imaginary and $S(\omega_k)$ and $T(\omega_k)$ real in the centre of the band. Let us introduce the 'reduced gain'

$$g(\omega_k) = \frac{|G(\omega_j)| \star N}{\omega_k \Omega/2\pi} = g_\xi \tag{3.53b}$$

in analogy to Eq. (3.53) and recall the definition of the mixing factor $M_\xi = \Omega/2\Delta \omega_k$ for well separated bands (see under Eq. (3.43)).

Using the simplification Eq. (3.61) for the dispersion integral and Eq. (3.64) for $T(\omega_k)$ we have

$$\tilde{S}(\omega_k) = \frac{\pi}{2 \ii \omega_0} \frac{G(\omega_k)}{\Delta \omega_k} = \frac{-1}{N} g_\xi \frac{M_\xi/2}{\Delta \omega_k} \tag{3.66}$$

$$T(\omega_k) = \frac{1}{1 + g_\xi M_\xi/2} \cdot$$

The cooling rate equation for any particle is obtained from the expressions of section 3.10 by replacing $\tilde{S}(\omega) + \tilde{T}(\omega) G(\omega)$, Eq. (3.65). We obtain in the present case:

$$\frac{1}{\tau} = \frac{\Omega/2\pi}{2N} \sum_{k = \text{passband}} \frac{2 g_\xi}{1 + \frac{g_\xi M_\xi/2}{\Delta \omega_k}} - \frac{g_\xi^2}{(1 + \frac{g_\xi M_\xi/2}{\Delta \omega_k})^2} (M_\xi + U_\xi) \tag{3.67}$$

This is formally the same as Eq. (3.52) if we substitute

$$g_\xi + T_\xi g_\xi = \frac{g_\xi}{1 + g_\xi M_\xi/2}$$

Optimum cooling is obtained from Eq. (3.67) when:

$$T_\xi g_\xi (M_\xi + U_\xi) = 1$$

i.e. when for all bands

$$g_\xi = \frac{1}{M_\xi/2 + U_\xi}; \quad g_\xi = \frac{2}{M_\xi} \cdot \tag{3.68}$$
The limiting case (*) is for negligible amplifier noise, $U_A \ll M_A/2$. The optimum shielding factor corresponding to Eq. (3.68) is:

$$I = \frac{1}{1 + \frac{M_A}{(M_A + 2U_A)} \cdot \frac{1}{2}}$$  \hspace{1cm} (3.69)

and the optimum damping rate

$$\frac{1}{\tau} = \frac{U/2\pi}{2N} \cdot \frac{1}{M_A + U_A}.$$  \hspace{1cm} (3.70)

Thus in the situation of negligible amplifier noise, optimum cooling is obtained when the gain (at all bands involved) leads to signal reduction by a factor of about 2. By comparing open and closed loop signals (either Schottky noise or driven-beam response) the gain can thus be optimized band by band. An example of Schottky signal shielding of a band is given in Fig. 36. Note that the optimum gain Eq. (3.68) for $U = 0$ is twice the optimum Eq. (2.28) calculated without beam feedback. When the amplifier noise becomes important ($U \gg M$) then Eq. (3.68) and Eq. (3.69) yield the optimum $g = 1/U$, $T = 1$ as in the case without shielding.

Fig. 36 Reduction of a Schottky noise band when the cooling loop is closed. With negligible amplifier noise and well separated bands optimum gain of the cooling loop corresponds to a signal amplitude reduction by about 2 in the centre of the bands.

Thus the inclusion of beam shielding (which was done in an approximate manner here) leads to an improved expression for the cooling rate and - more importantly - to an adjustment criterion for the cooling system.

The analysis done here for betatron cooling can be repeated for momentum spread damping where similar gain adjustment criteria apply.
4. DISTRIBUTION FUNCTION EQUATIONS (FOKKER-PLANCK) AND MOMENTUM SCALING

4.1 Distribution functions and particle flux

To follow the details of the cooling process, we (may) want to know more than the evaluation of the mean-square beam size and the r.m.s. momentum spread -- the only quantities used up to now to characterize cooling. In fact, a beam profile monitor records the particle distribution with respect to transverse position (see Fig. 37 as an example), and a longitudinal Schottky scan such as Fig. 2 gives the (square root of the) momentum distribution. These pictures are rich in fine information on peak densities, densities in the tails, asymmetries, and other practical details which are overlooked if only the r.m.s. is regarded.

![a)](image1)

![b)](image2)

Fig. 37 Evolution of beam profile (number of particles vs. vertical position) during stochastic cooling test in "ICE". The scans were obtained with a profile monitor which records the position of electrons liberated by beam particles through collisions with the residual gas. a) Before cooling; b) after 4 min of cooling.

It is therefore challenging to find an equation which describes all that can be observed and that is of practical importance. Such an equation does in fact exist!
For stochastic cooling the problem was (to my knowledge) first tackled by Thorndahl\textsuperscript{16} who already in 1976 worked with a Fokker-Planck type of equation for the particle density. This line was followed by virtually all subsequent workers\textsuperscript{17}, and computer codes for solving the distribution function equations are extensively used in the design of stochastic cooling and stacking systems.

The basic ideas behind this 'distribution function analysis' are simple, so that also the beginner can get -- hopefully without too much pain -- some first degree of familiarity with this powerful tool of cooling theory. I will first give the recipe and then try to justify it.

Let $\psi(x)$ (Fig. 38) be the particle distribution with respect to the error $x$ (e.g. $x = \Delta p/p$). Define $\psi(x) = dN/dx$ so that $\psi(x) dx$ gives the number of particles with an error in the range $x$ to $x + dx$. During cooling we find different distributions $\psi(x)$, taking snapshots at different times (see Fig. 2 as an example). We characterize this by letting $\psi = \psi(x,t)$ be a function of time also. The partial differential equation which describes the dynamics of $\psi(x,t)$ can be written in the following form:

$$\frac{\partial \psi}{\partial t} = \frac{\partial}{\partial x} \left[ -F_0 + D \frac{\partial \psi}{\partial x} \right]. \quad (4.1)$$

![Graph showing particle distribution function $\psi(x)$ defining the number of particles $dN = \psi(x)dx$ with an error in the interval from $x$ to $x + dx$](image)

Fig. 38 A particle distribution function $\psi(x)$ defining the number of particles $dN = \psi(x)dx$ with an error in the interval from $x$ to $x + dx$

The cooling process is completely characterized by the two coefficients $F$ and $D$ (which describe the cooling system) and the initial conditions $\psi(x, t = 0)$ (which describe the distribution at the start). Particle loss due to walls or influx during stacking can be included via appropriate boundary conditions $\psi(x_1) = 0, (\Delta p/\Delta x)(x_1) = \text{const.}$, etc. Two representative examples of results obtainable with Eq. (4.1) are given in Fig. 39, taken from Ref. 18, and Fig. 40 from Ref. 13.

To analyse a given system we have to find its coefficients $F$ and $D$. These quantities are closely related to the coherent and incoherent effect, respectively, which we have identified before. In fact

$$F/F_0 = \langle \Delta \phi \rangle_t \quad (4.2)$$

![Graph showing distribution of particles](image)
Fig. 39 Momentum cooling at 600 MeV/c in LEAR computed using Eq. (8.1). (Curves taken from Ref. 18.)

Fig. 40 Evolution of the stack in the AA during stochastic accumulation. Curves computed using the distribution function equation with the boundary condition of constant particle influx simulating the new p added every 2.6 sec. (Curves taken from Ref. 19.)
is the expectation value (long-term average) of the coherent change \( \Delta x \) per turn of the error, and

\[
2D/f_0 = \langle (\Delta x)^2 \rangle_t
\]  

(4.3)

is the expectation of the square of this change. The quantities \( F \) and \( 2D \) alone are corresponding average changes per second. Note the difference between \( (\Delta x)^2 = (x_c - x)^2 \) used here, and \( \Delta(x^2) = x_c^2 - x^2 \) as frequently used before.

The important thing is that a distribution function Eq. (4.1) -- similar to the Fokker-Planck equation used in a variety of fields -- exists and that relatively simple prescriptions (4.2) and (4.3) permit us to establish the two coefficients \( F \) and \( D \) for any given stochastic cooling system. Incidentally, an equation similar to Eq. (4.1) had long been used (before 1976!) by the Novosibirsk Group to study the dynamics of electron cooling. Also the kinetic equations in plasma physics closely resemble our distribution equation.

Let us now try to follow a simple derivation of Eqs. (4.1)-(4.3). This derivation is due to Thorndahl\(^{16} \). It proceeds along the lines used in textbooks to derive the diffusion -- or heat transfer -- equations which resemble Eq. (4.1). Imagine a distribution function \( \psi(x) \) and calculate, for a particular value \( x_1 \) of \( x \), the number of particles per turn which are transferred from \( x \)-values below \( x_1 \) to values above \( x_1 \) (Fig. 16). If the correction per turn at the kicker is \( \Delta x \), then particles with an error between \( x_1 \) and \( x_0 = x_1 - \Delta x \) (cross-hatched area in Fig. 41) pass through \( x_1 \). Their number is

\[
\Delta N = \int_{x_0}^{x_1} \psi(x) \, dx.
\]  

(4.4)

![Fig. 41 A look at the distribution function Fig. 38 through a magnifying glass. When the error for particles with a value near \( x_1 \) is changed by \( \Delta x \), particles in the dark shaded area have the error value changed from values below to values above \( x_1 \), Eq. (4.6) expresses this area as the difference between the rectangle and the triangle sketched in the figure.](image)
Expanding $\phi$ at $x_1$,

$$\phi = \phi(x_1) + \frac{\partial \phi(x_1)}{\partial x}(x-x_1), \quad (4.5)$$

the integration yields

$$\Delta N = \phi(x_1) \cdot \Delta x - \frac{1}{2} \frac{\partial \phi(x_1)}{\partial x} (\Delta x)^2. \quad (4.6)$$

The first and second terms can be interpreted as the area of the rectangle and the triangle, respectively, sketched in Fig. 41.

We now define the (average) particle flux

$$\phi = \langle \phi \rangle_t$$

as the expected number of particles per second passing a given error value. Clearly, then, from Eq. (4.6), the instantaneous flux is:

$$\phi(x) = \langle \phi \rangle_t \cdot \phi(x) - \frac{\partial \phi}{\partial x} \langle (\Delta x)^2 \rangle_t \cdot \frac{\partial \phi}{\partial x}.$$

This gives the flux in terms of $F$ and $D$ as defined by Eqs. (4.2) and (4.3). The assumption has tacitly been made that the change $\Delta x$ per turn at the kicker is small and $\phi(x)$ smooth, so that higher expansion terms in Eq. (4.5) can be neglected.

Having found the flux we can immediately obtain Eq. (4.1) from the continuity equation

$$\frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial t} = 0. \quad (4.7)$$

It states that the change per second of the density is given by the 'gradient'-$\partial \phi/\partial x$ of the flux. This is similar to continuity considerations in other fields like, for instance, the charge conservation law of electrodynamics:

$$\frac{\partial j}{\partial x} + \frac{\partial \rho}{\partial t} = 0, \quad (4.8)$$

relating current density $j$ and charge density $\rho$.

Like other continuity equations, Eq. (4.7) can be obtained by looking at the flux going into and coming out of an element of width $dx$ in $\phi$, $x$-space (Fig. 42):
Incoming flux per second : \( \phi_1 \)

Outgoing flux per second : \( \phi_2 = \phi_1 + \frac{\partial \phi}{\partial x} \, dx \)

Surplus per second : \( \Delta \phi = \phi_1 - \phi_2 = -\frac{\partial \phi}{\partial x} \, dx \).

The resulting density increase (per second) in the element is thus

\[
\frac{\Delta \phi}{dx} = -\frac{\partial \phi}{\partial x},
\]

and conservation of the particle number requires a \( \Delta \phi/\Delta t \) equal to this.

\[\text{Fig. 42 The flux into and out of a narrow element of width } dx \text{ in } \psi-x \text{ space. An excess of incoming over outgoing flux leads to an increase with time of the density } \psi = \Delta N/dx \text{ of particles in the element.}\]

This completes the derivation. The resulting equation (4.1) agrees with observations made in the ISR and all subsequent machines using stochastic cooling. The reader who might have had some difficulty in appreciating the derivation may now be pleased to learn that the exact form of Eq. (4.1) has been a subject of discussion for quite some time. Looking at the derivation of the Fokker-Planck equation in textbooks\(^{20}\), one is tempted to put the coefficient D under the second derivative as is correct for a variety of other stochastic processes. In 1977 a machine experiment\(^{21}\) was performed at the ISR to clear up this question for cooling and diffusion problems in storage rings. The experiment clearly indicated that in the present case the diffusion term should be \( \delta/\delta x \cdot [D(\psi/\delta x)] \) as in Eq. (4.1) and not \((\delta^2/\delta x^2)(\delta \psi)\).

4.2 Example of asymptotic distributions and Palmer cooling

We may conclude from the preceding sections that it is relatively simple to determine the distribution equation pertaining to a given cooling problem. It is usually much more difficult to solve the equation. This is because in general the coefficients F and D are functions of \(x, t\), and \(\psi\) itself. Analytical solutions have therefore only been obtained in a few simple cases.
As an example, let us briefly look at Palmer cooling with the following simplifying assumption: No unwanted mixing, and Schottky noise negligible compared with amplifier noise. Denoting \( x = \langle \Delta p / p \rangle \), the correction per turn is

\[
\Delta x = -g [ \langle x \rangle_s + x_n ]
\]

as given by Eq. (2.24) in Section 2. In analogy to Eq. (2.18) in Section 2.3.3, we assume that the long-term average of \( \langle x \rangle_s = (1/N_s) \sum x_i \) is zero except for the contribution \( x/N_s \) of the test particle upon itself. The noise \( x_n \) has zero average. Hence

\[
\langle \Delta x \rangle_t = -g \frac{x}{N_s} = -g \frac{2N}{N_0} \cdot x.
\]

In a similar way (using the assumption that \( \langle x \rangle_t \gg \langle x^2 \rangle_t \)), i.e. amplifier noise dominating over Schottky noise

\[
\langle (\Delta x)^2 \rangle_t = g^2 \langle x^2 \rangle_t = g^2 x_n^{\text{rms}} = \text{const}.
\]

Hence in this simple case \( F = F_0' x \) and \( D = D_0 \), where \( F_0' = (2N/M)g \) and \( 2D_0 = f_0 g^2 \langle x_n^{\text{rms}} \rangle \) are constants. In this case, Eq. (4.1) is amenable to an analytic solution. Try

\[
\psi = \frac{N}{\sqrt{2\pi} \sigma(t)} \exp \left( -x^2/2\sigma(t)^2 \right)
\]

i.e. a Gaussian with \( \sigma \) changing in time. Upon substitution, one obtains an ordinary differential equation for the width, \( \sigma \), of the Gaussian:

\[
\frac{d\sigma}{dt} = -F_0' + D_0/\sigma^2.
\]

Special cases:

\[
D_0 = 0: \quad \sigma^2 = \sigma_0^2 e^{-2F_0't} \quad \text{(continuous cooling)};
\]

\[
F_0' = 0: \quad \sigma^2 = \sigma_0^2 + 2D_0 t \quad \text{(diffusion)}.
\]

General solution:

\[
\sigma^2 = \sigma_1^2 e^{-2F_0't} + D_0/F_0'.
\]

This describes cooling towards an asymptotic (Gaussian) distribution with \( \sigma = \sqrt{2F_0'/D_0} \). In this situation an equilibrium between heating and cooling is reached. A similar result is arrived at from the simple cooling equations [e.g. Eq. (25), Section 2] which suggest \( 1/\tau \to 0 \) when the signal \( \langle x^2 \rangle_s \) has decreased so much that \( g \mu x^2/\langle x^2 \rangle_s \to 0 \). The new information obtained from Eq. (4.1) is that the asymptotic \( \psi \) is Gaussian in the simple case considered.
The existence of asymptotic equilibrium distributions is a common feature also in more complicated cases of Eq. (4.1). The final distribution \( q_m \) can be obtained putting \( \partial q_\lambda / \partial t = 0 \), which converts Eq. (4.1) into a simpler ordinary differential equation:

\[
-f_{m} \frac{\partial q_{m}}{\partial x} = \text{const.} \tag{4.9}
\]

The constant is frequently zero (e.g. when \( F(x) = 0 \) and \( \partial q_i / \partial x = 0 \) for \( x = 0 \) as can often be inferred from the symmetry of the problem). Equation (4.9) is important as it indicates the limiting density which can be reached.

4.3 Momentum cooling by filter and transit time methods

These methods measure the revolution frequency of particles or the time of flight between pick-up and kicker in order to detect the momentum error. The filter method of Carron and Thorndahl\(^{22}\) (Fig. 43) uses a notch filter between the preamplifier and the power amplifier, with notches at all revolution harmonics in the passband (Fig. 44). In the simplest case the filter is a transmission line shorted at the far end (Fig. 44), with a length corresponding to half the revolution time of the particles in the storage ring. The notches are produced by \( \lambda / 2 \) resonances, where ideally the input impedance is zero and the phase changes sign. Because of these phase and amplitude characteristics, particles with a wrong revolution frequency are accelerated or decelerated until ideally all have 'fallen into the notches'. The filter method is important for the cooling of low-intensity beams, and in fact the whole antiproton complex at CERN would probably not have worked with stochastic cooling had this technique not been invented in due time. Sum pick-ups are used, and these produce a much larger signal than the difference devices that are necessary with other methods. The filter reduces not only the particle signals but also the preamplifier noise at the critical frequencies. This feature is important for fast cooling at low intensity. The price to pay for this is that all the Schottky bands used have to be well separated, so that particles 'know' the notches into which they have to fall. This means unavoidably imperfect mixing. However, this slight disadvantage could probably be circumvented by using the signal from a second pick-up -- rather than the reflection of the previous turn pulse via a cable -- to cancel signals of a particle with the correct time of flight between the two pick-ups and to accelerate/decelerate others\(^{23}\).

![Diagram](image)

Fig. 43 The basic set up for momentum cooling by the filter method. An advantage of this method is that a sum pick-up is used which is sensitive even to small beam signals. Secondly, Schottky and preamplifier noise are reduced by the filter.
Yet another time-of-flight method has been discussed at Fermilab\textsuperscript{24}). Essentially, the idea is to differentiate the pick-up pulse and apply this signal on the kicker with a delay so that particles with the correct time of flight between pick-up and kicker are not affected, whereas slow or fast ones get a correction.

![Diagram](image)

Fig. 44 A simple periodic notch filter namely a half-wave low-loss transmission line (used as a stub resonator). The (idealized) gain and phase characteristics are given by the half-wave resonances at the multiples of the revolution frequency. Additional elements are usually added to reduce the gain between the harmonics.

Both variants of the filter method are less efficient in noise suppression and have, therefore, not found applications so far.

We shall return to the time domain for a short moment to suggest slightly different explanations of the filter method: the pulse sent into the cooling system by a particle of nominal frequency will be cancelled by its pulse from the previous revolution reflected at the end of the line. For particles that are too slow or too fast, the cancellation is imperfect and acceleration or deceleration will result.

The filter method is usually analysed using the distribution Eq. (4.1). The coefficients $F$ and $D$ can be worked out theoretically and/or by measurements on the system. Usually, measurements and calculations are done harmonic by harmonic, including various ingredients such as imperfect mixing and signal suppression. All we want to do here is to write down the general form of the relevant coefficients $F$ and $D$ which, expanding up to second order in the error quantity $x = \Delta E/E$ take the following form:

\[
F = -G_0 x
\]

\[
D = G_2 \phi(x^2 + \kappa_0) + G_2 \mu(E^2 + \kappa_1),
\]

where $x$ is the relative energy error; $G_0$ (proportional to the 'gain' $g$), $G_2 (\propto g^2)$, and $G_2 (\propto g^2)$ are given by the ideal filter, $\kappa_0$ and $\kappa_1$ by the losses; and $\mu$ relates to the amplifier noise. The first term of $D$ (which is proportional to the density $\phi$) gives the Schottky noise filtered by the notches, and the second term the filtered preamplifier noise. For more details, the reader should consult the specialized literature.
REFERENCES

   H. Pouth, Electron cooling, these proceedings.
   K. Hübner, Synchrotron radiation, ibid. p. 239.


   D. Boussard, Schottky noise and beam transfer function diagnostics, these proceedings.


5) See Bibliography, 1980:9 and 10.


   See also Bibliography, 1980:1.


10) See Bibliography, 1983:2, 1984:9, 10, 11 and 12.


14) See Bibliography, 1978:3.


   S. van der Meer and L. Thorndahl, Computer programs for solving Eq. (4.1), private communication.


19) See Bibliography, 1978:15.


22) See Bibliography, 1978:2.

23) See Bibliography, 1984:3.

BIBLIOGRAPHY

Chronological list of internal reports, conference presentations, and other publications related to stochastic cooling

1838

1918

1956

1958

1972
I. S. Van der Meer, Stochastic damping of betatron oscillations, internal report CERN/ISR P0/72-31.
2. W. Schnell, About the feasibility of stochastic damping in the ISR, internal report CERN/ISR RF/72-46.

1973

1974

1975

1976
2. D. Cline, P. McIntyre, F. Mills and C. Rubbia, Collecting antiprotons in the Fermilab booster and very high energy proton-antiproton interactions, Fermilab internal report TM 689.
7. W. Hardt, Augmentation of damping rate for stochastic cooling by the additional use of non-linear elements, CERN/PS/DL 76-10.


8. L. Thorndahl, A differential equation for stochastic cooling of momentum spread with the filter method, technical note ISR-RF/LT/ps.


10. L. Jackson Laslett, Evolution of the amplitude distribution function for a beam subjected to stochastic cooling, Berkeley report LBL-6469.

1978

1. ICE Team, Initial Cooling Experiment progress reports Nos. 1 and 2, CERN-EP Div.

2. G. Carron and L. Thorndahl, Stochastic cooling of momentum spread with filter techniques, internal report CERN/ISR-RF/78-12 and ISR-RF/Note LT/ps.


6. F. Bonaudi et al., Antiprotons in the SPS, internal report CERN DG 2.


13. S. van der Meer, Stochastic cooling theory and devices, ibid.


1979
3. J. Bisognano, Kinetic equations for longitudinal stochastic cooling, Berkeley (LBL) internal report BECON-11, LBL-140.
8. S. van der Meer, Debunched π-π operation of the SPS, internal report CERN/PS/AA 79-42.

1980
2. S. van der Meer, A different formulation of the longitudinal and transverse beam response, internal report CERN/PS/AA 80-4.
8. D.E. Young, Progress on beam cooling at Fermilab, ibid., p. 800.
9. P. Lefevre et al., The CERN low energy antiproton ring (LEAR) project, ibid., p. 819.
10. Design study of a facility for experiments with low energy antiprotons (LEAR), internal report CERN/PS/DL 80-1.

1981


5. J. Hardek et al., ANL stochastic cooling experiments using the FNAL 200 MeV cooling ring, ibid., p. 2455.


1982


2. A.G. Ruggiero, Theory of signal suppression for stochastic cooling with multiple systems, Fermilab internal p note 193.

3. C. Kim, Design options for the fast betatron precooling systems in the debuncher or on the injection orbit, Berkeley (LBL) internal report BECON-25.


7. L. Tecchio, Comparison of electron and stochastic cooling for intermediate energy range, ibid.

8. F. Voelker, Electrodes for stochastic cooling of the FNAL antiproton source, ibid.

9. B. Autin, Fast betatron cooling in an antiproton collector, ibid.


11. K. Takayama, Effects of non-linear electrode on transverse cooling, ibid.

12. A.G. Ruggiero, 2-4 GHz tail stochastic cooling system with filters, ibid.

13. J. Marriner, Stochastic stacking without filters and accumulator gain profile, ibid.


15. C.R. Holt, A normal mode analysis of coupled betatron oscillations including cooling, ibid.

17. A. Ando and K. Takayama, RF stacking and tail cooling in the antiproton accumulator, Fermilab internal report TM 1103.

18. S. van der Meer, Gain adjustment criterion for betatron cooling in the presence of amplifier noise, internal report CERN/PS/AA Note 82-2.

19. B. Leskowar, Reliability considerations at travelling-wave tube and gallium-arsenide field effect transistor amplifiers, internal report LBL 1381.

1983


6. S. Chattopadhyay, On stochastic cooling of bunches in the colliding beam mode in high-energy \( \bar{p}p \) storage rings, ibid., p. 2334.


8. E. Peschardt and M. Studer, Stochastic cooling in the CERN ISR during \( \bar{p}p \) colliding beam physics, ibid., p. 2594.


10. W. Kells, Stability and signal suppression of Schottky signals from stochastically cooled beams, ibid., p. 2590.


12. A. Ruggiero and J. Simpson, Momentum precooling in the debuncher ring for the Fermilab Tevatron I project, ibid., p. 2596.

13. A. Ruggiero, Signal suppression analysis for momentum stochastic cooling with a multiple system, ibid., p. 2599.


15. S. Chattopadhyay, Vlasov theory of signal suppression for bunched beams interacting with a stochastic cooling feedback loop, ibid., p. 2646.

16. S. Chattopadhyay, Theory of bunched beam stochastic cooling, ibid., p. 2649.

17. S. Chattopadhyay, A formulation of transversely coupled betatron stochastic cooling of coasting beams, ibid., p. 2652.

18. B. Leskowar and C.C. Lo, Travelling-wave tube amplifier characteristics study for stochastic beam cooling experiments, ibid., p. 3423, preprint LBL 14142.

20. B. Autin, Technical developments for an antiproton collector at CERN, ibid., p. 393.
1984
4. C.S. Taylor, Stochastic cooling hardware, ibid., p. 163.
5. S. van der Meer, Optimum gain and phase for stochastic cooling, ibid., p. 183.
8. P. Bryant, Antiprotons in the CERN intersecting storage rings, ibid., p. 509.
1985
2. G. Carron et al., Status and future possibilities of the stochastic cooling system for LEAR, ibid., p. 27, preprint CERN/PS/LEA 85-12.


12. W.C. Barry, Suppression of propagating TE modes in the FNAL antiproton source stochastic beam cooling system, ibid., p. 2424.


14. H.G. Jackson et al., Preliminary measurements of gamma ray effects on characteristics of GaAs field effect transistor preamplifiers, LBL Berkeley report PUB-5141, LBL-21546.

1986


Workshops on beam cooling


Design reports of facilities using stochastic cooling

TEVATRON design study, 1983:27.
ACOL design study, 1983:28.

Summary reports and reviews

1977: 1 1981: 1, 3 and 12 1984: 1, 2, 3 and 4
1978: 3 and 13 1982: 1, 7 1986: 5
1980: 1 1983: 7

Special problems of cooling of heavy-ion beams are discussed in


Numbers refer to the Bibliography
Appendix 1

HISTORY

For a long time the principle of stochastic cooling was regarded as too far-fetched to be practical. A first experimental demonstration was tried only seven years after its invention and three years after the first publication of the idea (Table Al). The inventor, S. van der Meer, and the early workers had (mainly) emittance cooling of high-intensity beams in mind with a view to improving the luminosity in the CERN ISR.

A new era began in 1975 when P. Strolin, coming back from a visit to Novosibirsk, and L. Thorndahl realized the interest of stochastic cooling -- for both emittance and momentum -- of low-intensity $\bar{p}$ beams for the purpose of stacking. Stochastic cooling at low intensity is different from the original van der Meer cooling. The extension of the theory first done by H. Hereward and L. Thorndahl, and the design of the momentum cooling hardware (L. Thorndahl, G. Carron et al.), are perhaps as fundamental as the original invention and the earlier feasibility studies by S. van der Meer and W. Schnell.

Following upon this broadening of the field of application, in 1975 P. Strolin et al. worked out $\bar{p}\bar{p}$ collection schemes for the ISR using stacking in momentum space, and C. Rubbia et al. made their first proposals of the $\bar{p}\bar{p}$ scheme for the SPS using similar techniques of stochastic cooling and accumulation. This work gave new life to the idea at a time when the ISR was routinely stacking such high currents that proton beam-cooling became unnecessary -- or even impossible. Further milestones between 1975 and 1980 were the invention of the filter method of momentum cooling, the refinement of the stochastic stacking schemes, the results of the initial cooling experiment (ICE), and last, but not least, the success of the AA. The ICE ring was used to make a careful comparison of cooling theory with reality, including bunched beam cooling. By the middle of 1978 all systems worked so well that beam lifetimes at 2 GeV/c of the order of a week were reached, compared with lifetimes of a few hours without cooling. This permitted a measurement of the stability of the antiproton, and this experiment improved the lower limit in one go from 120 $\mu$s to 80 h.

One essential ingredient in this experiment was the technique developed at ICE to observe as few as 50 circulating particles in a non-destructive manner. This was made possible by stochastic cooling which reduced the momentum spread to $10^{-5}$ so that a resonant Schottky noise pick-up with the corresponding quality factor could be used.

Running-in of the AA started in the summer of 1980, and since 1981/82, stacks of several $10^{11}\bar{p}$ are routinely accumulated from batches of a few $10^6\bar{p}$ per second. The AA uses a total of seven cooling systems for longitudinal cooling of different 'regions' of the beam and for horizontal and vertical emittance cooling. Time constants are of the order of a second at up to $5 \times 10^5 \bar{p}$ or 30 minutes for $5 \times 10^{11} \bar{p}$, thus nearing design specifications. The AA is at the heart of CERN's antiproton programme, which culminated in the observation of the Intermediate Vector Bosons predicted by the unifying electroweak theory. As you know these new particles were produced in the collisions of dense proton- and antiproton bunches in the SPS as proposed by C. Rubbia and co-workers.
In the ISR stochastic 'post-cooling' of antiprotons from the AA was used (amongst other applications) to improve the beam lifetime and the resolution in conjunction with an internal hydrogen jet target. In this way charmonium states formed in proton antiproton interactions could be observed with high precision. This was another important achievement of the ISR just prior to its final shutdown. In fact the very last ISR beam was such an antiproton beam circulating in ring 2. It was finally dumped at 6.00 h on 25 June 1984 thus bringing to a definite end the glamorous career of a unique machine. The use of an internal target in conjunction with phase-space cooling of the circulating beam -- as proposed by the Novosibirsk group many years before -- had thus been put to work for the first time. This arrangement has stimulated much interest as an option for LEAR and as one of the basic techniques for the ion cooling rings now under design or construction.

The low-energy antiproton ring LEAR, which after ISR and SPS became the third customer of the AA, has given high quality \( \bar{p} \) beams to its 17 user groups on an operational basis since 1983. One particularity of LEAR is an 'adjustable' system which allows stochastic cooling at many different energies. In 1985 alone, a beam was extracted at 17 different momenta between 105 MeV/c (corresponding to 5.8 MeV kinetic energy) and 1.7 GeV/c (~ 1 GeV). Relatively fast cooling with time constants of 2 to 5 minutes for up to \( 4 \times 10^9 \) p works at 3 or 4 strategic energies; slow cooling with time constants of 10-20 minutes can be used at practically all momenta to keep the beam in shape during the one-hour extraction. In conclusion stochastic cooling has made a unique programme of antiproton physics possible at CERN.

From about 1978 onwards, other laboratories, especially the Novosibirsk group, who had pioneered electron cooling before, an ANL-LBL-Fermilab collaboration and, more recently, a group at the INS Tokyo, have done work (both experimental and theoretical) on stochastic cooling. This work has placed emphasis on various important aspects such as low-noise cryogenic amplifiers, very high frequency systems, cooling of heavy ions, or cooling bunches.

In 1983 the Tevatron-1 project at Fermilab and the \( \bar{p} \) collector ACOL to be added to the AA at CERN were approved. Both systems aim at stochastic cooling and stacking of antiprotons at a rate of several \( 10^7 \) \( \bar{p} \) per second. Some of the new features are cryogenically cooled components on the low-level side (amplifiers, terminating resistors and to some extent even cables and the pick-up plates themselves) to improve the signal-to-noise ratio and bandwidths in the Gigahertz range. The Fermilab antiproton source is now (Dec. 1986) in its final running-in phase; cooling and stacking rates close to design performance have already been obtained, at least in tests with protons. As in the CERN case the Fermilab source is entirely based on stochastic cooling: the 8 GeV debuncher ring uses fast emittance cooling (\( t = 2 \) sec at \( 7 \times 10^7 \) particles) of the \( \bar{p} \)-pulse from the production target. The 7.9 GeV accumulator ring combines stochastic cooling for stacking in momentum space with emittance cooling to improve the transverse density of the stack. In the accumulator alone there is a total of six specific cooling systems all working the 1 to 2 or 2 to 4 GHz range, each of them combining a large number of pick-up and kicker loops. The system for momentum cooling of the stack tail -- which is the largest system -- uses more than 100 pick-up and kicker units.

The ACOL ring at CERN, under construction since September 1986, is planned to come into operation in late summer 1987. Together with the AA, which is being modified to work in cascade with ACOL, it should improve the \( \bar{p} \) flux by an order of magnitude compared to that available with the AA alone. The number and the complexity of the cooling systems of AA-ACOL is as impressive as in the Fermilab case.
There has thus been a rapid development of stochastic cooling over the last decade and roughly one order of magnitude has been gained every four years in the cooling power, i.e. in the number of particles which can be cooled with a time constant of 1 s. This has become possible by making larger and larger bandwidths available. Probably an 'absolute' limit of the cooling power in the range of $10^9$ to $10^{10}$ particles per second will be reached unless bandwidths and frequencies much above 10 GHz can be used where (most) vacuum chambers transmit waveguide modes and where the beam size becomes comparable to the RF-wavelength.

Since about 1980, interest in cooling of heavy ion beams developed rapidly, and the combination of stochastic 'pre-cooling' with post-cooling by electrons looks attractive for some applications. A number of ion cooling rings with some resemblance to LEAR are being planned in the USA, Japan and Europe. All of them foresee electron cooling and many plan to use both electron- and stochastic damping. Three ion coolers: TARN II at the INS, Tokyo; the IUCF-cooler at Bloomington, Indiana and CELSIUS at Uppsala should come into operation in 1987 or early 1988. Six others, TSR at the MPI Heidelberg, ESR at GSI, Darmstadt, ASTRID at Aarhus, CRYRING at Stockholm, COSY at KFK Jülich, and the RNCP-cooler at Osaka are authorized or at least partly authorized projects. Other ion cooling rings are being planned at Oak Ridge, Berkeley and Brookhaven National Laboratory. Thus in the coming years phase space cooling -- both by electrons and by the stochastic method -- will be used to a very large extent at low and medium energy.

For the very highest energies, ideas on bunched beam cooling are being followed up, and a thorough study on stochastic cooling of bunches in the SPS collider has been carried out. It is being complemented by the study of some of the components needed such as the pick-ups and kickers as well as the system to transmit cooling signals over long distances.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liouville</td>
<td>1838</td>
<td>Invariance of phase space area</td>
</tr>
<tr>
<td>Schottky</td>
<td>1918</td>
<td>Noise in d.c. electron beams</td>
</tr>
<tr>
<td>MURA group (Lichtenberg, Mills, Sessler, Stahle, Synn)</td>
<td>1956-58</td>
<td>Liouville's theorem applied to particle storage rings</td>
</tr>
<tr>
<td>van der Meer</td>
<td>1968</td>
<td>Idea of stochastic cooling</td>
</tr>
<tr>
<td>ISR staff (Borer, Bramham, Hereward, Hübner, Schnell Thorndahl)</td>
<td>1972</td>
<td>Observation of proton beam Schottky noise in the ISR</td>
</tr>
<tr>
<td>van der Meer</td>
<td>1972</td>
<td>Theory of emittance cooling</td>
</tr>
<tr>
<td>Schnell</td>
<td>1972</td>
<td>Engineering studies</td>
</tr>
<tr>
<td>Hereward</td>
<td>1972-74</td>
<td>Refined theory, low-intensity cooling</td>
</tr>
<tr>
<td>Bramham, Carron, Hereward, Hübner, Schnell, Thorndahl</td>
<td>1975</td>
<td>First experimental demonstration of emittance cooling</td>
</tr>
<tr>
<td>Palmer (BNL), Thorndahl</td>
<td>1975</td>
<td>Idea of low-intensity momentum cooling</td>
</tr>
<tr>
<td>Stroin, Thorndahl</td>
<td>1975</td>
<td>$\bar{p}$ accumulation, schemes for the ISR using stochastic cooling</td>
</tr>
<tr>
<td>Rubbia</td>
<td>1975</td>
<td>$\bar{p}$ accumulation, schemes for the SPS</td>
</tr>
<tr>
<td>Thorndahl, Carron</td>
<td>1976</td>
<td>Experimental demonstration of momentum cooling</td>
</tr>
<tr>
<td>Thorndahl, Carron</td>
<td>1977</td>
<td>Filter method of momentum cooling</td>
</tr>
<tr>
<td>Sacherer, Thorndahl, van der Meer</td>
<td>1977-78</td>
<td>Refinement of theory; imperfect mixing, Fokker-Planck equations</td>
</tr>
<tr>
<td>ICE team</td>
<td>1978</td>
<td>Detailed experimental verification</td>
</tr>
<tr>
<td>Herr</td>
<td>1978</td>
<td>Demonstration of bunched beam cooling</td>
</tr>
<tr>
<td>AA team</td>
<td>1981-82</td>
<td>Accumulation of several $10^{11} \bar{p}$ from batches of several $10^8$, cooling times close to design specifications</td>
</tr>
<tr>
<td>LEAR team</td>
<td>1982</td>
<td>Stochastic cooling to permit loss-free deceleration of $\bar{p}$. Interest in combining stochastic and electron cooling.</td>
</tr>
<tr>
<td>Kiliin</td>
<td>1982</td>
<td>Stochastic cooling of heavy ions (proposal)</td>
</tr>
<tr>
<td>Novosibirsk group</td>
<td>1980/82</td>
<td>Stochastic cooling experiments, work on cooling theory</td>
</tr>
<tr>
<td>ANL-LBL-Fermilab group</td>
<td>1979/83</td>
<td>Theoretical and hardware studies, stochastic cooling experiments at FNAL</td>
</tr>
<tr>
<td>TARN group at INS Tokyo</td>
<td>1983/84</td>
<td>Stochastic cooling experiments in the TARN ring</td>
</tr>
<tr>
<td>Fermilab group</td>
<td>1983</td>
<td>Design report of the Tevatron-1 project using a debuncher ring and an accumulator ring for fast stochastic cooling and stacking of $4 \times 10^7 \bar{p}$ per second</td>
</tr>
<tr>
<td>AA team</td>
<td>1983</td>
<td>Design report of the $\bar{p}$ collector ACOL using fast momentum and emittance cooling and stacking of $4 \times 10^7 \bar{p}$ per second</td>
</tr>
<tr>
<td>Berkeley group</td>
<td>1983</td>
<td>$\bar{p}$ collection scheme for 20 TeV colliders. Collection of $5 \times 10^6 \bar{p}$ per second</td>
</tr>
<tr>
<td>SPS $\bar{p}$ team</td>
<td>1983</td>
<td>Feasibility study of bunched beam cooling in the SPS collider</td>
</tr>
<tr>
<td>Fermilab group</td>
<td>1985-86</td>
<td>Construction and running in of $\bar{p}$ source</td>
</tr>
<tr>
<td>AA-ACOL team</td>
<td>1986-87</td>
<td>Construction of ACOL</td>
</tr>
<tr>
<td>Groups in several different labs</td>
<td>1986-</td>
<td>Construction of (heavy) ion cooling rings</td>
</tr>
</tbody>
</table>
Appendix 2

DISPERSION INTEGRALS

In this appendix we wish to have a brief look at the dispersion integral Eq. (3.60) as required for the 'signal shielding' calculations. A more general discussion is given by H.G. Hereward\footnote{53} in the context of Landau damping.

To deal with the singularity of the integrand we assume that the eigenfrequency $\omega_j$ of the test particle has a small imaginary part, i.e. we take $\omega_j = \omega_j + i\alpha$ such that the free oscillation $a e^{i\omega_j t}$ corresponding to Eq. (3.55) is damped. Later we go for the limit $\alpha \to 0$.

With a complex eigenfrequency Eq. (3.60) becomes

$$\tilde{S}(\omega) = \frac{1}{2\pi} \int \frac{\tilde{G}(\omega_j) n(\omega_j) [(\omega_j - \omega) - i\alpha]}{(\omega_j - \omega)^2 + \alpha^2} \, d\omega_j.$$  

We are especially interested in the contribution due to the damping term $\alpha$:

$$\tilde{S}_1(\omega) = -\frac{1}{2\pi} \int \frac{\tilde{G}(\omega_j) n(\omega_j) i\alpha}{(\omega_j - \omega)^2 + \alpha^2} \, d\omega_j.$$  

The main contribution to this integral comes from the range $\omega_j = \omega \pm \alpha$ times $\alpha'$, near the pole. For small $\alpha$ we can usually assume that $G(\omega_j)$ and $n(\omega_j)$ are constant in this range and thus take the weighting function $G(\omega_j) n(\omega_j) = G(\omega) n(\omega)$ out of the integral. Integrating the rest from a minimum $\omega_j < \omega$ to a maximum eigenfrequency $\omega j > \omega$

$$\tilde{S}_1 = -\frac{i}{2\pi} \tilde{G}(\omega) n(\omega) \left[ \arctan \left( \frac{\omega - \omega_j}{\alpha} \right) - \arctan \left( \frac{\omega - \omega_j}{\alpha} \right) \right]$$  

and in the limit $\alpha \to 0$

$$\tilde{S}_1 = -\frac{i}{2\pi} \tilde{G}(\omega) n(\omega) \cdot \pi.$$  

Clearly this is the residuum due to the pole of the integrand. Due to the physics of the problem the value $-i\pi$ has to be retained.

The remaining part of the integral is the principal value. It can be expressed in terms of the Hilbert transform (see Erdélyi et al., tables of integral transforms Vol. 2, MacGraw Hill N.Y. 1954) defined by
\[ H[f(x)] = g(y) = \frac{1}{\pi} \int \frac{f(x)}{x-y} \, dx. \]

This transform has been tabulated for a large collection of functions. In terms of the Hilbert transform the principal value of Eq. (3.60) may be written as

\[ \tilde{S}_H(\omega) = \frac{\pi}{2\omega} \cdot H[n(\omega)\tilde{G}(\omega)]. \]

Further details depend on \( n(\omega) \) and \( G(\omega) \). For betatron cooling \( \tilde{G}(\omega) \) is ideally constant (and purely imaginary and negative) whereas \( n(\omega) \) is "bell shaped" around the average betatron frequency \( \omega_B \) (or the centre of the Schottky band). It is therefore convenient to work in terms of the deviation from \( \omega_B \) denoting \( x = \omega_j - \omega_B, \quad y = \omega - \omega_B \). Then

\[ \tilde{S}(y) = \frac{\pi}{2\omega} \left[ -iG(y)G(y) + H[G(x) \cdot n(x)] \right]. \]

Two distributions can serve as models and permit the construction of approximations like Eq. (3.61) above:

1) The semi-circular distribution (which models a distribution with sharp cut off)

\[ n(x) = \begin{cases} \frac{2}{\pi\Delta^2} \sqrt{\Delta^2 - x^2} & \text{if } |x| < \Delta \\ 0 & \text{elsewhere} \end{cases} \]

For constant \( \tilde{G} \) one obtains

\[ \tilde{S}(y) = -\frac{\tilde{G}}{\omega_B^2} (y + i\sqrt{\Delta^2 - y^2}). \]

2) The Lorentzian distribution (which models a distribution with important tails):

\[ n(x) = \frac{\delta}{\pi} \frac{\delta}{\delta^2 + x^2} \]

which for constant \( \tilde{G} \) yields

\[ \tilde{S}(y) = \frac{\tilde{G}}{2\delta_B} \left[ \frac{x + i\delta}{\delta^2 + x^2} \right]. \]
ELECTRON COOLING

H. Poth

ABSTRACT
A comprehensive introduction to electron cooling is given. After a brief explanation of electron cooling and its applications, the reader is introduced to theory by a simple approach. Next, experimental aspects of an electron cooling device are discussed. Then, the theory is discussed in more detail. This is followed by a summary of the electron cooling experiments and a comparison between theory and experimental results. At the end, future applications of electron cooling are presented.

1. INTRODUCTION
This lecture is intended to provide an introduction to electron cooling and its major applications in accelerators and other fields of physics. Intentionally, it is kept at a fundamental level and a rigorous theoretical treatment of the process is sometimes forfeited in favour of simple understanding and clarity.

Apart from the electron cooling theory the practical aspects are worked out and discussed in detail. Furthermore, a summary of past cooling experiments and an outlook on future systems is given. Possible experiments using the electron cooler as a device for atomic-physics experiments will also be discussed.

Electron cooling as a method to improve the properties of stored ion beams was proposed by G. Budker in 1966. He and his group built the first electron cooling device and did the first cooling experiments at Novosibirsk. They also laid down the theoretical framework. A little later, electron cooling experiments were also performed at CERN and at Fermilab. The general aim of these experiments was to test a technique which eventually could allow the accumulation of antiprotons. However, owing to the high energy of the production maximum for antiprotons, it turned out that the application of stochastic cooling was more appropriate. In spite of that, the spectacular results which came from the first pioneering experiments encouraged many physicists to use electron cooling in low-energy ion rings in order to improve luminosity and resolution in experiments with stored beams and internal targets.

2. WHAT ELECTRON COOLING IS
Electron cooling is a fast process to shrink the size, divergence, and energy spread of stored charged-particle beams without removing particles from the beam. Since the number of particles remains unchanged and the space coordinates and their derivatives are reduced, this means that the phase space occupied by the stored particles is compressed. It also entails the temperature of the beam -- if looked upon as a gas -- being reduced. We know of other cooling processes which achieve similar results. These are stochastic cooling, synchrotron radiation cooling, and laser cooling.

*) Visitor at CERN, Geneva, Switzerland
3. **WHY ELECTRON COOLING?**

The most frequent application of electron cooling being considered at present is the loss-free compression of ion beams. The reduction of beam size permits the luminosity to be considerably increased for colliding-beam experiments. For fixed-target experiments the small beam size provides an excellent definition of the interaction vertex. The reduction of momentum spread paves the way for high-resolution experiments with internal targets.

Loss-free compression of phase space allows the accumulation of species of charged particles not abundantly available, such as light to heavy ions or polarized protons and deuterons. Electron cooling very rapidly reduces the emittance of such beam pulses injected from a cyclotron or a Linac into a storage ring and so creates new space for subsequent pulses. Even with low primary-beam intensity high stored-beam currents could be achieved through stacking and accumulation.

The possibility of working with small beams has another important aspect, namely it allows one to keep the dimensions of the vacuum chambers and the magnet gap of storage rings small, which considerably reduces financial expense.

In addition to phase-space compression, electron cooling may be applied to compensate beam-heating effects. These are predominantly intrabeam, residual-gas, and internal-target scattering. The counteraction against intrabeam scattering permits, for instance, to keep even intense stored beams small and to have, at the same time, a small momentum spread.

With electron cooling, residual-gas scattering losses can be reduced from multiple-scattering (gradual growth of emittance until ring acceptance is reached) to single-scattering losses (only those particles are lost that surpass in a single scatter the ring acceptance angle), which considerably increases the stored beam lifetime. It relaxes stringent vacuum requirements in quite a number of cases. On the other hand, it also gives access to the operation of stored beams at low energies, even of heavy ions, with reasonably long lifetimes.

Compensation of energy loss and beam blow-up, coming from beam interaction with an internal target, permits the operation of targets at the maximum thickness and -- since the beam particles are continuously recycled -- the achievement of high-luminosity, well-defined interaction vertices, high-momentum resolution, and clean background conditions, particularly at low energy.

The electron cooling arrangement provides more than an exclusive cooling device. It is also, at the same time, an electron target with which the stored beam interacts. In particular, electron-ion interactions at low relative energies can be uniquely studied. Atomic-physics experiments investigating, for instance, radiative capture, dielectronic recombination, and Rydberg atoms find here nearly ideal conditions. Such a technique also finds applications in the eventual production of antihydrogen. Moreover, radiative recombination may be suited to the production of monoenergetic photons in the vacuum ultra-violet (VUV) and the X-ray regions. The electron cooling arrangement may hence become a very important tool for modern atomic physics.

4. **ION-BEAM AND STORAGE-RING PROPERTIES**

We consider an ion beam of nominal momentum \( p \) circulating with the fraction \( \beta \) of the speed of light \( c \). The beam properties are characterized by invariant quantities, which
describe the phase space occupied by the ions. The six-dimensional phase space is generated by the transverse and longitudinal coordinates and their derivatives. In the horizontal plane we have, for instance, as the coordinates the distance of an ion from the nominal orbit\( x_1 \) and its angle\( x_1' \) with respect to the nominal beam trajectory. In general, it is assumed that the particles of a beam are normally distributed in these coordinates according to

\[
f(x_1, x_1') = \frac{1}{8\pi \sigma_x \sigma_{x'}} \exp \left[ -\frac{(x_1 - x_1')^2}{2\sigma_x^2} \right] \exp \left[ -\frac{x_1'^2}{2\sigma_{x'}^2} \right].
\]

(1)

The product of the width of the spatial distribution (beam size)\( \sigma_x \) and the angular distribution (beam divergence)\( \sigma_{x'} \) is the beam emittance\( \epsilon \). This quantity divided by\( \beta \gamma \) is the normalized emittance. It remains constant in the absence of cooling and heating at any energy.

The transport of ions in a ring is described by the focusing functions\( \beta^*(s) \), where\( s \) is the coordinate along the nominal orbit. At any point in the ring it relates beam size, divergence, and emittance.

\[
\langle x_1 \rangle = \frac{\epsilon}{\pi \beta^*(s)}, \quad \langle x_1' \rangle = \frac{\epsilon}{\pi \beta^*(s)}.
\]

(2)

The quantities of\( \epsilon, \langle x_1 \rangle, \langle x_1' \rangle, \) and\( \beta^* \) are usually given in\( \text{mm mrad}, \text{mm mrad}, \text{and m}, \) respectively. A finite transverse beam size and divergence are due to betatron (transverse) oscillations of the ions around the nominal orbit, which is in sinusoidal approximation given by

\[
x_1(s) = \sqrt{\gamma} \cos \left( \frac{Q \phi}{R} + \delta \right),
\]

(3)

where the period of the betatron oscillation\( Q \) is called the tune. The quantity\( \delta \) is the initial phase of the ion and\( R \) is the ring circumference divided by\( 2\pi \).

In the longitudinal plane the beam is characterized by a distribution of the ion momentum around the nominal beam momentum. Only coasting beams are considered. Again we assume here a normal distribution. Ions with momenta different from the nominal beam momentum circulate on different orbits. Their radial difference\( r_1 \) from the nominal orbit at any point is related to their off-momentum\( \Delta \beta_1 \) by the dispersion function\( D(s) \),

\[
r_1 = D(s) \frac{\Delta \beta_1}{p},
\]

(4)

with\( D(s) \) usually given in\( m \).

At a given spatial position in the ring\( (s, x, z) \) we can define the local beam temperature in the rest frame in all three planes as

\[
T_h = m c^2 \beta^2 X_1^2 \langle x_1 \rangle^2, \quad T_v = m c^2 \beta^2 X_1^2 \langle x_1 \rangle^2, \quad T_\parallel = m c^2 \beta^2 \langle \Delta \beta_1/p \rangle^2, \quad T_\perp = T_h + T_v.
\]

(5)

In this definition the transverse beam temperature changes with the position in the ring. We may define an average transverse beam temperature by using the beam emittance [Eq. (2)] and the average\( \beta \)-function.
\begin{equation}
\langle \beta^* \rangle \geq \frac{R}{Q}
\end{equation}

\begin{equation}
T_{l} = mc^2 \beta^2 \gamma^2 e \left( \frac{1}{\langle \beta_h^* \rangle} + \frac{1}{\langle \beta_g^* \rangle} \right) = mc^2 \beta^2 \gamma^2 \epsilon \frac{e}{R} (Q_h + Q_g).
\end{equation}

**Beam lifetime**

Repeated small-angle scattering of stored ions from residual gas molecules leads to an emittance growth rate given by

\begin{equation}
\frac{d\epsilon}{dt} = \frac{10^6 \langle \beta^* \rangle P}{\beta^2 \gamma^2} = \frac{10^6 RP}{Q \beta^2 \gamma^2},
\end{equation}

where $P$ is the residual gas pressure in nitrogen equivalent. When giving $\langle \beta^* \rangle$ and $R$ in m and $P$ in Torr the dimension of $d\epsilon/dt$ is $\text{mm-mrad-s}^{-1}$. The emittance grows until the machine acceptance $\tilde{\Lambda}$ is reached and the beam gets lost. The beam lifetime $\tau_{ms}$ resulting from this process (multiple scattering), is given for a round vacuum chamber by

\begin{equation}
\tau_{ms} = \frac{0.85(\tilde{\Lambda}/\sigma) \beta^2 \gamma Q}{PR}.
\end{equation}

For a rectangular chamber $\tilde{\Lambda}$ has to be replaced by $2.5\tilde{\Lambda} \left( 1/\tilde{\Lambda} = 1/\tilde{\Lambda}_h + 1/\tilde{\Lambda}_g \right)$.

Whenever multiple scattering is counteracted by cooling, then only those particles get lost which undergo scatters larger than the machine acceptance angle $\theta_0$. The cross-section for that is given by

\begin{equation}
\Delta \sigma = 4\pi z^2 Z^2 \frac{m_e^2 \epsilon}{\beta^2 \gamma^2} \frac{1}{e^2} (\text{Rutherford scattering}).
\end{equation}

Here $z$ is the charge of the ion and $Z$ that of the gas nuclei, and $r_e$ is the classical electron radius; $m_e$ is the electron mass. The single scattering lifetime is given by

\begin{equation}
\tau_{ss} = \frac{A}{10^6 \rho_B \gamma C N_A} = \frac{P \Delta \theta_0^2}{4\pi e^2 \gamma \rho \beta^2 \gamma^2 \left( \frac{\beta e}{m_e \epsilon} \right)^2}.
\end{equation}

Here $A$ is the atomic weight and $\rho$ the density of the residual gas, $N_A$ is Avogadro's number and $\rho_\theta$ is the residual gas density at normal pressure $P_0 = 760$ Torr.

For a residual gas consisting of two components with relative abundances $\kappa$ and $1 - \kappa$, the following holds:

\begin{equation}
\frac{1}{\tau_{ss}} = \frac{1}{\tau_{ss_1}} + \frac{1 - \kappa}{\tau_{ss_2}}.
\end{equation}

5. **HOW ELECTRON COOLING WORKS IN PRINCIPLE**

In order to cool a stored ion beam with electrons, a nearly monochromatic and parallel electron beam is caused to overlap with the ion beam in one of the straight sections of a storage ring. The velocity of the electrons is made equal to the average velocity of the
ions (Fig. 1). A close-up view of the overlap region shows us the ions traversing, under different angles and different velocities, the stream of parallel electrons all moving with the same velocity. However, if we observe the situation from a frame moving with the velocity of the electrons, the latter will all be at rest, while the ions will pass through the electron gas from any direction with a variety of velocities, resembling the movement of particles in a hot gas (Fig. 2). The ions will undergo Coulomb scatters in that gas and will lose energy, which is transferred from the ions to the electrons through this Coulomb interaction reducing the motion of the ions as seen from the rest frame. The electrons are continuously renewed. In this picture the electron cooler can be understood as a heat exchanger.

We may also consider the electrons as being represented by a foil moving with the velocity \( v_0 \). Ions moving faster than the foil (electrons) will penetrate it and will lose energy along the direction of their momentum (dE/dx) during each passage until all transverse components are diminished and their longitudinal velocity is equal to the foil velocity (Fig. 3). For slower ions it is the same effect with the exception that they traverse the foil from the opposite side. Ideally, at the end, all ions will have the same longitudinal velocity as the foil and no transverse velocity component.

6. **INTRODUCTION TO ELECTRON COOLING THEORY (FOR PEDESTRIANS)**

We now want to derive the force which is responsible for the slowing down of the ions and the characteristic time which it takes. Our reference frame is still the system where the electrons are at rest. All quantities measured with respect to this system are marked with an asterisk. Let us consider first a single electron-ion collision. The ion moves with the velocity \( v_i \), and scatters from the electron at an impact parameter \( b \) (Fig. 4). The momentum transfer \( \Delta p \) from the ion to the electron is:
\[
\Delta p^* = \int \varphi_C \, dt = \int \frac{2e^2}{x^2 + b^2} \, dt,
\]

with \( \varphi_C \) being the Coulomb force. Since we consider times from negative to positive infinity, we can neglect the longitudinal part of the force and can replace \( \varphi_C \) by its transverse component \( \varphi_\perp = \varphi_C (b / \sqrt{x^2 + b^2}) \):

\[
\Delta p^* = \int \varphi_\perp \, dt = \frac{2e^2}{v_i} \int \frac{dx}{(x^2 + b^2)^{3/2}} = \frac{2Ze^2}{v_i b}.
\]

From this we can calculate the energy loss of the ion, which is the energy taken by the electron:

\[
\Delta E^*(b) = \frac{(\Delta p^*)^2}{2m} = \frac{2Ze^4}{m v_i^2 b^2}.
\]

So far we have considered a single collision. Now we extend the picture to the situation where the ion passes through a large number of electrons (Fig. 5). We have to integrate over all possible impact parameters to obtain the energy lost by the ion as it travels a length \( dx \) through the electron cloud of density \( n_e^* \):

\[
- \frac{dE^*}{dx} = 2\pi \int_{b_{\text{min}}}^{b_{\text{max}}} b \, db \, n_e^* \Delta E^*(b) = \frac{4\pi Ze^4}{m v_i^2} n_e^* \ln \frac{b_{\text{max}}}{b_{\text{min}}}.
\]

This is the frictional (or cooling) force \( F^* \) experienced by the ion: \( F^* = dE^*/dx \).

![Fig. 5 Illustration of electron-ion interaction in an electron gas](image)

The logarithmic ratio of maximal to minimal impact parameter is called the Coulomb logarithm:

\[
L_C(v_i^*) = \ln \frac{b_{\text{max}}}{b_{\text{min}}}.
\]

We have to find reasonable cut-offs for the impact parameters. The minimum impact parameter is determined by the maximum momentum transfer to the electron (classical head-on collision):

\[
\frac{2Ze^2}{v_i^* b_{\text{min}}} = \Delta p^* = \frac{2m}{2} v_i^* + b_{\text{min}} = \frac{Ze^2}{m v_i^*}.
\]
Using the classical electron radius \( r_e = \frac{e^2}{m_e c^2} \), we can write

\[
b_{\text{min}} = Z r_0 e^{-\frac{s^2}{r_D^2}}. \tag{19}\]

In a system of charged particles, we know that the Coulomb field is shielded and falls off exponentially within a characteristic radius \( r_D \), which is the Debye radius

\[
r_D = \sqrt{\frac{m_e e^2}{8\pi \varepsilon_0 n_e^2}}. \tag{20}\]

It is usually smaller than the electron beam radius. Otherwise, the latter has to be taken as the maximal impact parameter. So we can finally write for the cooling force:

\[
F^* = -\frac{4\pi e^2 \varepsilon_i}{m_e v_i^2} n_e L_c(v_i^*) = -4\pi e^2 \varepsilon_i r_n^* e n_e L_c(v_i^*) v_i^*. \tag{21}\]

Its inverse dependence on the ion velocity leads to a divergence as the latter approaches zero (Fig. 6).

![Fig. 6 Shape of the cooling force for a frozen electron beam](image)

Now we want to calculate the rate of velocity change \( \dot{v}^* \) at which the ion is slowed down in the electron gas:

\[
\dot{v}^* = -\frac{d v^*}{d t}. \tag{22a}\]

Using the following relations

\[
E^* = \frac{m_i}{2} v_i^*; \quad \frac{d v_i^*}{d t} = \frac{1}{m_i} \frac{d E}{d x} \frac{d x}{d t} = \frac{1}{m_i v_i^*} F^* v_i^*,
\]

the friction rate \( \dot{v}^* \) can be expressed as

\[
\dot{v}^* = -\frac{F^*}{m_i v_i^*} = -\frac{F^*}{F_i^*}. \tag{22b}\]
Usually the inverse of the friction rate (or damping decrement) is defined as the cooling time $\tau$:

$$
\tau^* = \frac{1}{\lambda^*} = -\frac{m_i v_i^*}{F^*}.
$$

(22c)

Hence we find for the cooling time, observing that $r_i = e^2/m_i c^2$,

$$
\tau^* = \frac{v_i^*}{4 \pi Z^2 c^4 r_i n_e L_C}.
$$

(23)

This is the cooling time in the electron rest frame. In the laboratory frame we observe a cooling time which is (noting that $n_e = \eta n_e$)

$$
\tau = \frac{\tau^*}{\eta} = \frac{\gamma v_i^*}{4 \pi r_i c^4 L C n_e e^2}.
$$

(24)

where $\eta$ is the ratio of the length of the cooling section to the ring circumference.

So far we have considered the electrons as being stationary. However, they have a finite temperature $T_e$ and hence a velocity distribution $f(v_e^*)$, which we consider to be Maxwellian, characterized by its velocity spread $\Delta_e$.

$$
f(v_e^*) = \frac{e^{-v_e^2/\Delta_e^2}}{\Delta_e^2 \sqrt{\pi}}, \quad \int f(v_e^*) d^3v_e^* = 1.
$$

(25)

and

$$
T_e = \frac{m_e \Delta_e^2}{2}.
$$

(26)

To account for this, we have to replace the ion velocity by the ion-electron velocity difference $u^* = v_i^* - v_e^*$ and to average over the electron velocity distribution $\mathcal{F} (\hat{u}) \times \langle \hat{F}^*, \hat{u} \rangle$, which gives

$$
\mathcal{F}^*(v_i^*) = -4 \pi Z^2 e^2 c^2 r_i n_e L_C \int L_C(v_i^*) f(v_e^*) \frac{u^*}{|u^*|^3} d^3 v_e^*.
$$

(27)

Since the variation of the Coulomb logarithm with the ion velocity is small we can put it in front of the integral

$$
\mathcal{F}^* = -4 \pi Z^2 e^2 c^2 r_i n_e L_C \int f(v_e^*) \frac{u^*}{|u^*|^3} d^3 v_e^*.
$$

(28)

This expression has an electrostatic analogy. The integral in real space resembles just the Coulomb force of a charge distribution acting on a test charge. It can be rewritten as

$$
\mathcal{F}^* = -4 \pi Z^2 e^2 c^2 r_i n_e L_C (v_i^*) \int f(v_e^*) \frac{u^*}{|u^*|^3} d^3 v_e^*.
$$

(29)
For a distribution

\[ f(v_e^*) = \begin{cases} \text{const} & \text{for } |v_e^*| < \Delta_e \\ 0 & \text{for } |v_e^*| > \Delta_e \end{cases} \]

it can be solved analytically and the shape of the cooling force is shown in Fig. 7.

For a Maxwellian velocity distribution it has to be evaluated numerically, although there exists a complete mathematical expression. The frictional force can, however, then be approximated (Fig. 8) by

\[ F^* = -\frac{v_e^*}{12\pi^2 e^2 c^2 r_e n_0 m_e c} \frac{v_i^*}{|v_i^*|^3 + 2\Delta_e^3} . \tag{30} \]

![Fig. 7 Shape of cooling force for an electron beam with rectangular velocity distribution.](image)

![Fig. 8 Shape of the cooling force for the Maxwellian electron velocity distribution (short dashes) and its approximation by a force of the form \( F^* \propto v_i^*/(|v_i^*|^3 + 2\Delta_e^3) \) (full line) and \( F^* \propto -v_i^*/(|v_i^*| + 2\Delta_e^3) \) (long dashes).](image)
Using this approximation the cooling time can be written as

\[ \tau = \frac{2}{\eta} \frac{v_i^*}{\Delta e} + \frac{2\Delta e}{12\pi^2 r c^4 n_e L_C} \]  

(31)

Generally the electron temperature \( T_e \) is independent of the beam energy and hence \( \Delta e \) is a constant of the apparatus. Therefore we can distinguish two domains of cooling:

i) Cooling of hot beams \( v_i^* \gg \Delta e \)

Here the cooling time is proportional to \( v_i^* \). It corresponds to the region where \( T_e \approx v_i^* \).

ii) Cooling of 'warm' beams \( v_i^* \ll \Delta e \)

In this domain the cooling time is practically independent of \( v_i^* \) since \( \Delta e \) is constant and one has an exponential damping. This simple model gives the following scaling behaviour of the cooling time:

\[ \tau \text{ is independent of } \beta \text{ for } n_e^* = \text{const} \]

\[ \tau \text{ is independent of ion beam intensity} \]

\[ \tau = \frac{2 \Delta e}{T_e} \]

\[ \tau = \frac{1}{n_e} \]

The scaling behaviour will be modified later when refining the model. For typical numbers (\( T_e = 0.2 \text{ eV}, \Delta e/c = 10^{-3}, v_i^* = \Delta e, n_e = 10^8 \text{ cm}^{-3}, L_C = 10, \eta = 0.04, \gamma = 1, Z = 1 \)) we find a cooling time of about 5 s.

Before we go into further details we will first discuss how electron cooling is experimentally realized.

7. EXPERIMENTAL REALISATION OF ELECTRON COOLING

7.1 Electron gun and acceleration

The efficient application of electron cooling on stored ion beams depends very much on the quality of the electron beam and the exact matching of both beams. We will now discuss the generation of a cold electron beam.

Electrons for cooling are produced in an electron gun, where they are accelerated electrostatically to the desired energy. We will discuss this using the Low-Energy Antiproton Ring (LEAR) gun (Fig. 9) as an example. A thermocathode which is heated resistively to a temperature above 1000°C serves as an electron source. Electrons leave the cathode in any direction, forming a cloud in front of it. The energy distribution of the electrons follows a Maxwellian distribution with an average velocity given by the cathode temperature \( T_e = kT_{\text{cath}} \) (\( k = \text{Boltzmann constant} \)). Electrons will be extracted from this cloud with the help of ring-shaped anodes and accelerated to the desired energy with which they enter into a drift region. Usually the cathode is at high negative potential and the anode potentials increase steadily to zero. In order to minimize transverse electric field
Fig. 9 The LEAR electron gun

1 Cathode W
2 Pierce shield Ta
3 Heat sink Mo
4 Gas-cooled base Cu
5 Cathode feedthrough Al₂O₃
6 Anode feedthrough Al₂O₃
7 Bellows s.s.
8 Anodes Ti
9 Anode Cu
10 Anode support Al₂O₃
11 Solenoid
components the cathode is surrounded by the Pierce shield, an electrode on cathode potential which is matched to such a shape as to nullify, together with the potential given by the electron cloud, the electric field on the cathode surface and to produce equipotential lines which are perpendicular to the beam axis. With the subsequent anodes one tries to maintain this situation as much as possible. However, at the end of the acceleration column transverse field components are unavoidable.

Electrons are emitted from the cathode because of their thermal energy, in any direction. Therefore the electron gun is embedded in a longitudinal magnetic field (solencoid), which has the effect that the transverse motions of the electrons are transformed into spirals about the magnetic field lines with the cyclotron frequency given by

$$\omega_c = \frac{eB}{mc}$$  \hspace{1cm} (32a)

The spiral radius is

$$\rho_c = \frac{v}{\omega_c}$$  \hspace{1cm} (32b)

where $v$ is the electron velocity transverse to the magnetic field. Moreover, the magnetic field value is chosen such that one revolution (or a multiple of it) is completed when the electrons enter into the drift region. This has the advantage that the effect of radial electric field components are minimized. Otherwise, it would lead to a continuous scalloping of the beam. This technique is called resonant acceleration (optics). The condition for resonant optics is that the time the electrons need to pass through the acceleration column is equal to a multiple $\nu$ of the inverse of the cyclotron frequency or that, in other words, the length of the acceleration region $\lambda_R$ is approximately

$$\lambda_R \approx \nu \frac{eB}{m_c} = \nu \frac{v}{\omega_c} \frac{m_c}{e} \gamma$$  \hspace{1cm} (33a)

This means the magnetic field should have the value

$$B = \nu B \frac{m_c}{e \lambda_R}$$

$$B [\text{kG}] = 3.4 \nu \frac{B \gamma}{\lambda_R \text{[cm]}}$$  \hspace{1cm} (33b)

So electrons are not extracted directly from the cathode, but from the space-charge cloud in front of it. Under these conditions the final electron current follows Child's law:

$$I = P U^{3/2}$$  \hspace{1cm} (34)

where the characteristic proportionality factor $P$ is called the pervance. The pervance is essentially determined by the ratio of beam radius $r_0$ and cathode-anode distance $d$:

$$P = 7.3 \mu \rho \left( \frac{r_0}{d} \right)^2$$  \hspace{1cm} (35)

where $\mu \rho = 10^{-6} \text{AV}^{-3/2}$. In our example here the gun pervance is $P = 0.5 \mu \rho$. This geometrical pervance can, however, be reduced by applying a smaller cathode-anode potential. The
electron current can also be reduced by heating the cathode less and thus reducing its electron emission. For such a temperature-limited gun Eq. (34) no longer holds.

In Fig. 9b the calculated electron trajectories in an electron gun are shown together with the electric potential lines. In the drift region the potential lines are parallel to the electron trajectories with increasing space between adjacent potential lines when going from the centre to the edge of the beam. The radial behaviour of the electric field is described by the potential of a homogeneous charge distribution with sharp boundaries.

$$\Delta U = \frac{m_e c^2}{e} n_e e r \rho r^2, \quad r < r_0.$$  \hfill (36)

Expressing $n_e$ through $\rho$:

$$n_e = \frac{I}{\pi r_0^2 e \sigma c} = \frac{\rho r_0^{3/2}}{\pi r_0^2 e \sigma c}$$  \hfill (37a)

and

$$eU = \frac{m_e c^2 \rho^2}{2} r$$  \hfill (37b)

one can rewrite the above formula

$$\frac{\Delta U}{U} = \frac{e \rho}{e c \gamma} \sqrt{\left(\frac{m_e c^2}{e\gamma} \right)^3 + \frac{r_0^2}{2 \gamma}} \quad \text{for } r < r_0.$$  \hfill (38)

One realizes that, for a constant perveance, the relative potential increase across the electron beam remains constant.

The change of the electric potential across the electron beam has two major consequences. Firstly, it leads to a radial electron velocity profile of the same parabolic form; and secondly, the $E \times B$ situation leads to an azimuthal drift of the electron beam with a drift velocity given by

$$v_d = \frac{r}{u_c} \frac{r}{2 \gamma n_e c^2}.$$  \hfill (39)

As mentioned before, the electrons emitted from the cathode have a three-dimensional Maxwellian velocity distribution given by Eq. (25). Applying a voltage $U$ accelerates them to an energy $E$:

$$E = eU + \frac{1}{2} m_e c^2$$  \hfill (40)

where the ripple of the high-voltage system $\Delta U$ and the cathode temperature $T_{\text{cath}}$ adds up to $T_{\text{eff}} = kT_{\text{cath}} + e \Delta U$ as the effective cathode temperature. Their final longitudinal energy spread is

$$\Delta E = \frac{m_e c^2}{2} = \frac{1}{2} E \frac{\Delta U}{U}$$  \hfill (41)

and the corresponding longitudinal temperature is, hence,
\[ T_{\text{eff}} = \frac{m_e e^2}{m_e c^2} \gamma^2 = \frac{\gamma}{\gamma_0} \left[ \frac{kT_{\text{cath}}}{\gamma E} + 2 \frac{\Delta U}{U} + \left( \frac{\Delta U}{U} \right)^2 \frac{E}{kT_{\text{cath}}} \right]. \]  

(42a)

For \( T_{\text{eff}} = kT_{\text{cath}} \) and \( T_{e_L} = kT_{\text{cath}} \), we have, hence,

\[ \frac{T_{e_L}}{T_{e_L}} = \frac{1}{4} \frac{\Delta e_L^2}{\Delta e_L^2} \quad \text{and} \quad \frac{\Delta e_L^2}{\Delta e_L^2} = \frac{1}{2} \frac{\Delta e_L^2}{\Delta e_L^2}. \]

(42b)

In the transverse direction no change takes place with respect to the situation before acceleration. Hence \( \Delta e_L^2 = \Delta e_L^2 \) and \( \Delta e_L^2 \ll \Delta e_L^2 \). Therefore, in the accelerated electron beam the ions (observer) find a longitudinally compressed (flattened) electron velocity distribution

\[ f(v_e^I) = \frac{\exp \left[ -\left( \frac{v_e^I}{\Delta e_L^2} \right)^2 \right]}{\sqrt{\pi} \Delta e_L^2 \Delta e_L^2} \quad \text{with} \quad \Delta e_L^2 \ll \Delta e_L^2. \]

(42c)

7.2 Electron beam transport and interaction region

From the gun exit onwards the magnetically confined electrons drift to the section where they are bent into the ion beam. This is achieved by a curved solenoid (toroid) -- see Fig. 10 -- and an additional magnetic dipole field. The latter is needed to compensate the centrifugal force experienced by the electrons and oblige them to follow the magnetic field lines. After the toroid the electrons enter the cooling region, where they overlap with ions also entering the toroid (Fig. 10).

The situation in the cooling section is illustrated in Fig. 11. It shows the parabolic electron-velocity profile and the straight line of the ion dispersion. Since the cooling force \( F \) is proportional to \(- (v_e - v_i)/v_i^2\), the ions are dragged along the straight line to point \( A \), except when they are at the right of point \( B \). In the latter case they are continuously accelerated and lost. Using the form of the cooling force given in Eq. (30) (we

![Fig. 10 The toroid section](image)

![Fig. 11 The electron velocity profile and the ion dispersion](image)
Fig. 12 The shape of the longitudinal cooling force taking into account electron-velocity profile and ion-momentum dispersion for a zero emittance beam.

consider here the longitudinal force component only and assume vanishing transverse velocities), one finds the shape of the longitudinal force as given in Fig. 12 when the parabolic electron velocity profile is taken into account. This means that the cooling force is enhanced between \( A \) and \( B \) and attenuated left of \( A \). In order to get stable conditions and efficient cooling the beams have to be very well aligned and the velocity of the electrons correctly chosen. We will come back to this point later.

At the end of the cooling section the electrons are separated from the ions again by a toroidal magnetic field and drift in a solenoid field to the collector.

7.3 Electron collector

The collector is a very important component. It has to reduce the power \( I_U \text{cath} \) stored in the electron beam to the lowest possible values. This is done by decelerating the electrons before they hit the collector. The remaining power \( I_U \text{coll} \) is dissipated in the water-cooled collector. The collector is usually a few thousand volts less negative than the cathode.

Another important task of the collector is to gather the electrons with very high efficiency. Secondary electrons created in the collector or electrons reflected at its entrance may bounce back and forth between gun and collector before they are lost somewhere on the grounded vacuum chamber walls at practically full energy. Apart from the gas load which these lost electrons produce, the corresponding loss current has to be provided by the high-voltage power supply. Since this has to be a highly stabilized power supply, the load should be as small as possible.
The deceleration of the electrons before the collector can be done with resonant optics similarly to the gun acceleration column, or merely by passing the electrons through a ring anode on the desired potential. When the electrons are decelerated, a space-charge cloud builds up forming a similar situation to that in front of the cathode. This charge has either to be compensated by opposite charges (positive ions) or the electrons have to be sucked away and distributed over a large volume; otherwise the electron beam is reflected from this virtual cathode. The latter is achieved by re-accelerating the electrons into the collector and reducing the magnetic field to zero. The vanishing magnetic field in the collector helps also to prevent the secondary electrons from leaving the collector and entering the cooler. Experience at Fermilab showed that it was quite useful to form an ion trap between the deceleration column and the collector. This was achieved by running a cylindrical collector anode a few hundred volts above the cathode potential, so that ions were trapped and compensated the electron space charge.

In previous collectors at Novosibirsk and at Fermilab, electron losses at the level of $10^4$ and below were achieved. In Fig. 13 the Novosibirsk collector is shown.

Finally, an example of a whole electron cooler assembly is shown in Fig. 14, which represents the LEAR electron cooler.

Fig. 13 The electron collector of the NAP-M electron cooler

Fig. 14 An electron cooling assembly (LEAR electron cooler)
7.4 High-voltage system

A typical schematic drawing for the high-voltage system of an electron cooler is shown in Fig. 15. An extremely well stabilized high-voltage power supply (ripple $< 10^{-4}$) provides a negative potential to a high-voltage platform. The cathode is directly connected to the platform. On the platform additional power supplies provide the bias for the collector and its electrodes. The potential for the acceleration (deceleration) nodes are either derived from a voltage divider or from auxiliary power supplies. This arrangement has the advantage that all the electron current essentially flows through the collector supply, and the high-voltage power supply $U_k$ has only to deliver the loss current (and the current through the voltage divider).

![Diagram of high-voltage system for electron coolers](image)

Fig. 15 Schema of high-voltage systems for coolers. The dashed outline shows the HT platform.

7.5 Vacuum system

Electron coolers will be used in storage rings operating under ultrahigh-vacuum conditions ($10^{-10}$–$10^{-12}$ Torr). The vacuum system of the electron cooler has to match that and hence it should be bakable. The main outgassing in an electron cooling device comes from

a) the hot cathode,

b) the collector,

c) the vacuum walls hit by lost electrons with full energy.

Mainly hydrogen and carbon monoxide are produced.

In order to keep the vacuum in the region which is traversed by the ion beam as low as possible, a differential pumping system has to be built between the gun and the cooling region, and the collector and the cooling region. This is difficult: since the whole system is contained in the solenoid and access is difficult. Suitable pumping systems are cryopumps and non-evaporable getter (NEG) pumps.
7.6 Magnetic field

The properties of the magnetic field are very important, firstly, to prevent the electron beam from being heated up and, secondly, to guarantee a good cooling efficiency. That means variations of the magnetic field should not take place over distances smaller than the spiral length of the electrons outside the cooling region (adiabatic). In the cooling region, however, the angle $\alpha_0$ between the magnetic field lines (electron trajectories) and the ion beam should everywhere be small compared to the average transverse angle of the electrons given by their transverse temperature

$$\alpha_0 \ll \frac{1}{2} \frac{e_0}{\beta c} = 0.5 \times 10^{-3} / \beta .$$

(43)

In Fig. 16 the magnetic field of the LEAR electron cooler before final correction is shown. Eventual magnetic field errors are usually compensated by suitable correction coils inside the main magnet. There are also steering coils to allow for a displacement of the electron beam.

![Graph](image)

Fig. 16 Plot of magnetic-field components for LEAR electron cooler (before field corrections).

7.7 Effects of the electron cooler on the ion beam

The major effects of the electron cooler on the ion beam are the deflection of the ion beam in the toroids, the focusing effect of the electron beam which produces a tune shift of the ion beam, and the coupling of the vertical and horizontal emittances in the solenoid. The latter also precesses the spin of the ion and the solenoid field has to be compensated if polarized ions are to be cooled.

The deflection of the ion beam in the vertically bent toroid is due to the rising and falling magnetic field there, causing a vertical dipole field. The deflection angle is

$$\theta \ [\text{rad}] = \frac{\int B \, dl}{B_0 \, \epsilon_0} = \frac{B R}{B_0 \, \epsilon_0} \ln \cos \phi_0 \quad (44b)$$

and the displacement of the beam is
\[
\Delta x = \frac{B_0 R^2}{B_0 v_0^2} \left| \frac{\beta}{v_0} \right| \tan \phi_0 \tag{44b}
\]

Here \(B\) is the field of the solenoid, \(B_0 v_0\) is the ion beam rigidity (\(B_0 v_0 [T \cdot m] = 3.3 \, p\) [with \(p\) in GeV\(\cdot c^{-1}\)]), \(R\) is the radius of the toroid and \(\phi_0\) its angular length (Fig. 10).

The tune shift produced by the electrons is given by

\[
\Delta \nu = 0.5 \left\langle \frac{\beta^2}{\nu_m} \right\rangle \frac{n_e}{p_e} \beta^2 \gamma^{-3} L \tag{44c}
\]

where \(L\) is the length of the cooling section. For an electron gun operating with constant permeance the tune shift remains unchanged when the beam energies are varied, since \(n_e = \beta^2\) [Eq. (37)].

The solenoidal magnetic field twists the ion beam by an angle which is given by

\[
\delta [\text{rad}] = \frac{L}{\beta c m_e} \frac{e}{c} \frac{m_e}{m_i} \tag{44d}
\]

This effect is minor unless one is working close to a machine resonance.

The solenoid rotates the spin by

\[
\phi [\text{rad}] = \gamma G \frac{B L}{B_0 v_0^2} \tag{44e}
\]

where \(G\) is the G-factor of the ion (\(G = 1.793\) for the proton). This has to be compensated by an additional solenoid; otherwise the beam would depolarize.

8. MORE ON THE THEORY OF ELECTRON COOLING

8.1 Cooling force

The force governing the electron cooling process was derived by various authors. One distinguishes two basic descriptions: the binary collision model by Derbenev and Skrinsky, and the plasma physical approach by Sørensen and Bonderup. Although the latter has some advantages compared to the binary model, we will follow the description of Derbenev and Skrinsky as it is more practical to do so here.

In Section 6 we derived a cooling force making the assumption of having free electrons with a spherical velocity distribution. In Section 7.1 we learnt that the electrons have, however, a flattened velocity distribution [Eq. (42c)] and that the electrons -- confined by a longitudinal field -- are performing rotations about the magnetic field lines.

Derbenev and Skrinsky have evaluated the cooling forces for a flattened distribution:

\[
F_i^T = -4\pi^2 e^2 c^2 \gamma_0 n_e^* \begin{cases} 
L_C(v_i^*) \frac{v_i^*}{v_i}, & v_i^* > \Delta e_L \\
\sqrt{\pi} L_C(\Delta e_L) \frac{v_i^*}{\Delta e_L}, & v_i^* \leq \Delta e_L
\end{cases}
\tag{45a}
\]

(here \(L_C\) is defined by Eq. (17) with \(b_{\text{max}} = r_C\), and
\[
F_{\parallel}^s = -4\pi e^2 c^2 \frac{m}{2} \left( \frac{v_{\perp}^*}{v_{\perp}} \right)^2 \left[ \frac{L_C(v_{\perp}^*)}{L_C(e_{\parallel})} - \frac{v_{\perp}^*}{\Delta e_{\parallel}} \right] \left( \frac{2}{\Delta e_{\parallel}} \right)^\frac{1}{2} L_C(e_{\parallel}) \left( \frac{2}{\Delta e_{\parallel}} \right)^\frac{1}{2} L_C(e_{\parallel}) \left( \frac{v_{\perp}^*}{\Delta e_{\parallel}} \right), \quad \frac{v_{\perp}^*}{\Delta e_{\parallel}} \ll \frac{v_{\perp}^*}{\Delta e_{\parallel}} (45b)
\]

Hence the flattened distribution has the essential consequence that the longitudinal cooling force falls off less rapidly for \( \frac{v_{\perp}^*}{\Delta e_{\parallel}} \) as in the case of a spherical distribution; that means the longitudinal cooling is faster than the transverse cooling for cool ion beams. This is easily understood. The influence of the magnetic field is more difficult.

When an ion scatters from a spiralling electron at impact parameters much larger than the cyclotron radius, and the collision time \( t = b/u^* \) is long compared to the cyclotron revolution frequency, the electron makes many rotations and only the longitudinal electron velocity has to be taken into account. Since the transverse electron motion is frozen, no transverse momentum is transferred in these slow collisions (large impact parameters + small ion velocity). This type of collision is therefore often called adiabatic collision. This fact has so far not been taken into account and the previous cooling force has to be complemented with a magnetic part.

Usually the collisions are divided into two types depending on the impact parameter \( b \):

- fast collisions: \( b_{\min} < b < r_c \)
- adiabatic collisions: \( r_c < b \).

correspondingly, the total cooling force is composed of a non-magnetic force \( F^s_{\parallel} \) [Eq. (45)] and a magnetic force \( F^m_{\parallel} \). The latter is also derived by Derbenev and Skrinsky:

\[
F^m_{\parallel} = -2\pi e^2 v_{\perp}^* c^2 \int \frac{v_{\perp}^*}{u_m} L_C^m \frac{\partial f(v_{\perp})}{\partial v_{\perp}} d^3 v_{\perp} \quad (46a)
\]

\[
F^m_{\parallel} = -2\pi e^2 c^2 \int \frac{v_{\perp}^*}{u_m} L_C^m \frac{\partial f(v_{\perp})}{\partial v_{\perp}} d^3 v_{\perp} \quad (46b)
\]

Here \( L_C^m \) is the adiabatic Coulomb logarithm with \( b^m_{\min} = r_c \) and \( b^m_{\max} = \min \{ r_c, u^* L/\beta c, u^*/\omega_p \} \), \( \omega_p \) being the electron plasma frequency \( \omega_p = \sqrt{4\pi e^2 n_c m} \).

For an infinitely narrow longitudinal electron velocity spread the integrals can be evaluated:
\[ F_{\perp}^{\text{magnetic}} = -2\pi Z^2 e^2 c^2 \rho_0 L_{\text{C}}^{\text{magnetic}} \frac{v_{\perp}^2}{v_1^2} \] (47a)

\[ F_{\parallel}^{\text{magnetic}} = -6\pi Z^2 e^2 c^2 \rho_0 L_{\text{C}}^{\text{magnetic}} \frac{v_{\perp}^2}{v_1^2} v_1 \] (47b)

The case \( v_1^\ast < \Delta_{e_\parallel} \) can be evaluated already for a finite longitudinal velocity spread \( \Delta_{e_\parallel} \):

\[ F_{\perp}^{\text{magnetic}} = -4\pi Z^2 e^2 c^2 \rho_0 L_{\text{C}}^{\text{magnetic}} \ln \left( \frac{\Delta_{e_\parallel}^0}{v_1^\ast} \right) \frac{v_{\perp}^\ast}{v_1^\ast} \] (48a)

\[ F_{\parallel}^{\text{magnetic}} = -4\pi Z^2 e^2 c^2 \rho_0 L_{\text{C}}^{\text{magnetic}} \Delta_{e_\parallel}^3 \frac{v_{\perp}^\ast}{v_1^\ast} \] (48b)

which is similar to the non-magnetic force with a spherical velocity distribution of the width \( \Delta_{e_\parallel}^0 \).

The effect of the magnetic cooling force is to enhance the cooling at low ion velocity considerably. This results in very short cooling times for the damping of small betatron oscillations and the reduction of small momentum spreads of the ion beam. However, Eq. (47a) shows that also transverse heating may occur if \( v_{\perp}^\ast < \sqrt{2} v_1^\ast \). Furthermore, it indicates that the transverse cooling is much slower than the longitudinal.

8.2 Cooling times

Cooling times are usually defined as the time it takes to damp betatron oscillation amplitudes or a momentum spread by a factor 1/e. This assumes exponential damping (constant cooling time) which is usually not the case. It is more appropriate to use a damping (cooling) rate. We will take the cooling time to be the inverse of the damping rate.

With the cooling forces given above, we observe the following behaviour of the cooling times:

\[ \tau^\ast = \frac{v_1^3}{v_{\perp}^\ast} \quad \text{for} \quad v_{\perp}^\ast > \Delta_{e_\parallel} \] (49a)

\[ \begin{align*}
\frac{1}{\tau^\ast} &= \begin{cases}
\text{const} & \text{non magnetic force} \\
-\frac{v_1^3}{v_{\perp}^\ast} & \text{magnetic force}
\end{cases} \\
&\text{for} \quad \Delta_{e_\parallel} < v_{\perp}^\ast < \Delta_{e_\parallel}
\end{align*} \] (49b)

\[ \tau^\ast = \text{const} \quad \text{for} \quad v_{\perp}^\ast < \Delta_{e_\parallel} \] (49c)

This is an essential correction to the formula given initially [Eqs. (24) and (31)].
The division of the collisions into two regimes is rather crude. Recently the theory developed by Sørensen and Bonderup overcame this problem in a natural way by deriving the frictional force from the polarization the ion induces in the magnetically confined electron gas.

8.3 Equilibrium

In the end phase of cooling, equilibrium is reached between the electron and the ion beam (under ideal conditions), which means $T_e = T_i$. Provided there is no coupling between transverse and longitudinal phase space and there are no other heating effects, the following holds:

$$T_{e_L} = T_{i_L} \quad \text{and} \quad T_{e_L} = T_{i_L}.$$

The beam temperature is related to the beam divergence by

$$T_{e_L} = \frac{m_e \beta^2 \gamma^2}{2}, \quad T_{i_L} = \frac{m_i \gamma \beta^2 c^2}{2}$$

and

$$T_{e_L} = \frac{m_e \beta^2 \gamma^2}{2}, \quad T_{i_L} = \frac{m_i (\Delta \rho/p) \beta^2 c^2}{2}.$$

Hence equal beam temperatures means

$$\theta_{i_L} = \sqrt{\frac{m_e}{m_i}} \theta_{e_L}, \quad \gamma \frac{\Delta \rho}{p} = \sqrt{\frac{m_e}{m_i}} \theta_{e_L}.$$

The ion beam divergence and the momentum spread can then become $\sqrt{m_i/m_e}$ times smaller than the corresponding values for the electron beam. One notes further that -- since $\Delta e_L = \text{const}$, $\Delta e_L = \beta \Delta e_L$, and $\Delta = \Delta \beta c$ -- ideally, small ion-beam divergences and momentum spreads are reached at high energies. For $\beta = 1$ one could ultimately get $\theta_{e_L} = 0.5 \text{ mrad}$ and $\theta_{e_L} \ll 10^{-5}$. Under these conditions one could contemplate reaching a flattened ion velocity distribution with a longitudinal temperature $T_{i_L} \leq 1^\circ \text{K}$. Gases at these low temperatures undergo phase transitions, and at a certain stage of crystallization will take place. Such effects are predicted to occur for ion beams cooled down to ultralow temperatures.

In practice, however, heating effects will prevent us, in many cases, from reaching this region. Dominant heating processes are the scattering of an ion from another ion in the beam (intrabeam scattering), residual gas scattering, machine imperfections, and misalignment of the beams.

9. RECOMBINATION

When positive ions are cooled by electrons, occasionally cooling electrons are radiatively captured by beam ions into atomic states with main quantum number $n$. The process

$$e^- + A^{v'} \rightarrow \left( A^{(v-1)} \right)^* + h\nu.$$  

(53)
is illustrated in Fig. 17. Its cross-section is

$$a_n = \frac{2s^2 r_e^2}{a} \frac{E_0}{nE_0 + n^2 E_e}.$$  \hspace{1cm} (54)

where $E_0 = (m_e c^2/2)z^2 a^2$ is the binding energy of the ground state and $n$ is the principle quantum number. One notes that it diverges for decreasing electron energy $E_e \to 0$. In electron cooling $E_e$ (in the rest frame) is very small. The velocity distribution of the electrons requires averaging as in the case of the cooling force. This average is called the recombination coefficient $a_r$:

$$a_r = \int f(v_e^*) u^* I a_n d^3 v_e^*.$$ \hspace{1cm} (55)

The evaluation of the integral for a flattened distribution and $v_1^* \ll v_e^*$ results in

$$a_r = 2s^2 a r_e^2 z^2 \frac{f_x}{\Delta_e} I \frac{1}{n},$$ \hspace{1cm} (56)

where $n_{\text{max}}$ indicates the state above which the ions are stripped in the motional electric field of the bending magnet and are recirculated. Let us assume here that $n_{\text{max}} = 4$; then $I 1/n = 2$.

The recombination rate per stored ion observed in the laboratory is

$$\lambda_{\text{rec}} = n_e a_r n_{\gamma}^{-2},$$ \hspace{1cm} (57a)

and using Eq. (56)

$$\lambda_{\text{rec}} = n_e n_{\gamma}^{-2} 4s^2 \frac{f_x}{\Delta_e} a r_e^2 z^2.$$ \hspace{1cm} (57b)

This can be compared with the cooling rate. Let us take, for instance, the inverse of Eq. (31) for $v_1^* \ll \Delta_e$; then we have
\[
\frac{\lambda_{\text{rec}}}{\lambda_{\text{cool}}} = \frac{2\pi/e}{3L_C} \pi \frac{e^2}{r_L c^2} = \frac{4\pi}{L_C} \frac{A_p}{m_e} \frac{T_e}{m_e c^2} = 4A \times 10^{-5}
\]

(58)

This shows that recombination is much slower than cooling even for heavy ions.

10. ELECTRON COOLING EXPERIMENTS

Pioneering electron cooling experiments were done in the NAP-M ring at Novisibirsk, the Initial Cooling Experiment (ICE) ring at CERN, and the Fermilab cooler ring. The experiments were performed with stored protons at 1.5, 35, 46, 85, 114, and 200 MeV. The cooling of coasting and bunched beams was studied, and the stacking and accumulation of proton pulses was tested.

The parameters of the storage rings are listed in Table 1 and those of the electron coolers are given in Table 2.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters of electron cooling storage rings</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Circumference [m]</td>
</tr>
<tr>
<td>Operation energy [MeV]</td>
</tr>
<tr>
<td>Stored beam intensity</td>
</tr>
<tr>
<td>Average ring vacuum [Torr]</td>
</tr>
<tr>
<td>Horizontal and vertical acceptance [(\pi \text{mm}\cdot\text{mrad})]</td>
</tr>
<tr>
<td>Longitudinal acceptance [%]</td>
</tr>
<tr>
<td>Fraction of cooling section of ring circumference</td>
</tr>
<tr>
<td>Working point Q_h, Q_v</td>
</tr>
<tr>
<td>Average horizontal (\beta)-function [m]</td>
</tr>
<tr>
<td>(\beta_h) in cooling section [m]</td>
</tr>
<tr>
<td>(\beta_v) in cooling section [m]</td>
</tr>
<tr>
<td>Dispersion in cooling section [m]</td>
</tr>
<tr>
<td>Beam lifetime without cooling [s]</td>
</tr>
<tr>
<td>Transition (\gamma_t)</td>
</tr>
</tbody>
</table>

\textsuperscript{a)} At 65 MeV.
Table 2
Parameters of electron coolers

<table>
<thead>
<tr>
<th></th>
<th>NAP-M</th>
<th>ICE</th>
<th>Fermilab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathode diameter [cm]</td>
<td>1, 2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Beam diameter [cm]</td>
<td>1, 2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Electron energy [keV]</td>
<td>0.7-46</td>
<td>26</td>
<td>62, 111</td>
</tr>
<tr>
<td>Electron current [A]</td>
<td>0.1-0.8</td>
<td>0.6, 1.3, 2.2</td>
<td>0.5-3</td>
</tr>
<tr>
<td>Electron density [10^4 cm^-3]</td>
<td>0.09-3.7</td>
<td>0.2, 0.4, 0.8</td>
<td>0.1-0.6</td>
</tr>
<tr>
<td>Magnetic field [kG]</td>
<td>1</td>
<td>0.5</td>
<td>0.7, 0.93</td>
</tr>
<tr>
<td>Toroidal angle [']</td>
<td>45</td>
<td>30</td>
<td>90</td>
</tr>
<tr>
<td>Length of cooling section [m]</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Gun-collector voltage [kV]</td>
<td>~ 1</td>
<td>~ 1.2</td>
<td>~ 1</td>
</tr>
<tr>
<td>Electron current losses</td>
<td>&lt; 10^-4</td>
<td>~ 2.5 x 10^-2</td>
<td>&lt; 10^-4</td>
</tr>
</tbody>
</table>

10.1 Cooling forces and cooling times

In these experiments transverse and longitudinal cooling times were measured under various conditions and equilibrium beam properties were determined. The principal results are shown in Figs. 18 and 19, where the measured cooling times and longitudinal frictional force are plotted against the proton velocity and betatron amplitude, respectively. One distinguishes in Fig. 18b the clear scaling of the transverse cooling time with the transverse proton velocity. The data scale approximately as \( \tau_L \propto v_p^2 \). From the previous section we would expect [Eq. (49)]:

\[
\tau_L = \begin{cases} \text{const} & \text{from non-magnetic force} \\ v_p^3 & \text{from magnetic force} \end{cases}
\]

for \( v_p < \Delta_e \approx v_c \)

and

\[ \tau \propto v_p^3 \]

for \( v_p > \Delta_e \).

The longitudinal cooling force rises rapidly with decreasing velocity (Fig. 19). There the ICE results show for small velocities a bend over and a rapid decrease in contrast to the Novosibirsk results. A possible explanation could be that in ICE a beam misalignment or a magnetic-field ripple prevented a further rise of the force to the point where it should then decay linearly [Eqs. (15b) and (18b)]. It also could indicate a considerable ripple on the high voltage, reducing the effect of the flattened distribution.
Fig. 18 Compilation of transverse cooling time measurements

Fig. 19 Compilation of longitudinal cooling time/force measurements
The Novosibirsk results show a continuous rise of the cooling force with decreasing \( v_p^* \) beyond \( v_p^* / \beta c = (\Delta e / \beta c) (\sim 2.5 \times 10^{-3}) \), which is a clear indication of the presence of the magnetic force. In none of the cooling experiments, however, was the region where \( v_p^* < \Delta e / \beta c \) checked for a linear dependence of the magnetic force on the proton velocity.

The Novosibirsk group has derived semi-empirical formulae, which describe the data rather well:

\[
\tau_{\perp} = \frac{1}{66 \pi r_e^2 \beta c^4 \eta n_e} \left( \frac{\alpha_0}{4} \beta^2 c^2 + v_{\perp}^2 + 11 v_{\parallel}^2 \right) (\Delta e / \beta c + v_{\perp}^2 + v_{\parallel}^2)^{-1/2} \quad (59a)
\]

\[
F_{\parallel} = \frac{1}{12 \pi r_e^2 \beta^2 c^4 n_e} \left[ \frac{\alpha_0}{4} \beta^2 c^2 + v_{\perp}^2 + v_{\parallel}^2 \right] \left( \frac{\Delta e}{\beta c} + v_{\perp}^2 + v_{\parallel}^2 \right)^{-1/2} \quad (59b)
\]

Here \( \alpha_0 \) accounts for a possible magnetic field ripple or a misalignment angle between the electron and the proton beam.

The longitudinal frictional force in Eq. (59b) scales like:

\[
F_{\parallel} \propto \frac{1}{v_p^*} \quad \text{for} \ v_{\parallel} > \Delta e / \beta c
\]

and

\[
F_{\parallel} \propto \frac{1}{\Delta e / \beta c v_p^*} \quad \text{for} \ v_{\parallel} < \Delta e / \beta c
\]

10.2 Equilibrium

The equilibrium proton beam properties were determined in the previous cooling measurements for various conditions. Final beam emittances well below 1 \( \mu \text{m} \text{mrad} \) were achieved. Depending on the value of the \( \beta \)-functions in the cooling section this yielded divergences of less than 0.1 mrad or beam sizes of a fraction of a millimetre. The equilibrium beam momentum spread was in most cases limited by intrabeam scattering blow-up and ranged between \( 10^{-6} \) and \( 10^{-7} \) depending on the beam intensity. However, at Novosibirsk, with low beam intensities, indications for a longitudinal ordering within the proton beam were found to point to a crystallization.

10.3 Recombination

The recombination of cooling electrons with circulating protons was observed and used as beam diagnostics. In particular, it served to measure the beam size and the overall electron beam temperature. The experimentally determined recombination coefficient \( \alpha_e \) ranged between \( 0.8 \times 10^{-12} \) and \( 2.3 \times 10^{-12} \ \text{cm}^3 \ \text{s}^{-1} \), giving neutral hydrogen rates between a few hundred and a few thousand per second.

10.4 Beam lifetime

In the absence of cooling and machine resonances the lifetime of the stored beam is governed by the emittance increase due to multiple small-angle scattering of ions on residual gas molecules. The beam emittance continuously increases until the machine acceptance
is reached and the beam gets lost. Cooling counteracts this beam blow-up and only those particles are lost which undergo a single scatter larger than the machine acceptance angle.

In all cooling experiments a considerable increase of beam lifetime was observed (approximately by a factor of 40). The calculated lifetimes, assuming single scattering losses [Eq. (11)], were in rather good agreement with the experimental observations. For instance, the lifetime of a 50 MeV proton beam stored in ICE was about 1 h for a vacuum of $2 \times 10^{-3}$ Torr and a residual gas composition of 50% $\text{H}_2$ and 50% $\text{N}_2$. At lower energies (1.5 MeV) a beam lifetime of about 7 s was determined in the Novosibirsk experiment (pressure ~ $10^{-10}$ Torr).

11. SIMULATION OF ELECTRON COOLING IN STORAGE RINGS

In many cases, the influence of the combined action of the cooling and heating processes, the initial ion beam properties, and the characteristics of the electron beam on the evolution of the beam emittance and momentum spread in time and on the equilibrium values cannot be predicted by a simple mathematical expression. Rather it is necessary to follow the fate of an ensemble of ions based on a realistic model for the storage ring, the electron cooler, the matching of both systems, and to implement a maximum of beam dynamics. This allows then to calculate emittance decrease rates and the reduction of momentum spread as a function of the various machine parameters, and hence permits the optimization of the process for various operating conditions.

A computer code was developed for this purpose in the past few years by the KIK group at CERN. Some examples are discussed here. Figure 20 shows the distribution of a sample of beam particles in the cooling region, before cooling is started. Also shown is the dispersion curve and the velocity profile of the electrons. The horizontal distance of the ions from the dispersion straight line is a measure of their betatron amplitude (a zero emittance beam would coincide with the dispersion line).

![Figure 20 Simulation of the electron cooling process; distribution of an ensemble of ions for various times after onset of cooling.](image-url)
Fig. 21 Time dependence of averaged proton beam properties during simulation. The emittances contain 63.2% of the beam particles for a Maxwellian beam profile.

In Fig. 21 the evolution of the horizontal emittance and the momentum spread in time is shown. One observes two regions with distinct damping rates and a region where the equilibrium is reached. In the region with the smaller damping rate essentially only the non-magnetic cooling force plays a role, while, after the initial compression of the phase space, the onset of the magnetic-force contribution leads to faster cooling.

12. ELECTRON COOLING DIAGNOSTICS

12.1 Electron-beam diagnostics

In order to optimize electron cooling, efficient diagnostics are needed. The important properties of the electron beam have to be measured. These are density distribution and the velocity profile across the electron beam as well as the longitudinal and transverse electron beam temperatures. For practical reasons it should be possible to determine the electron-beam position at various places in the cooler and the electron collection efficiency. Of course, the cathode temperature and the electron current should be easily determinable.

The cathode temperature can be measured pyrometrically once and then determined from the heating power. The electron current is essentially deduced from the collector current. The loss current has to be provided by the high-voltage supply and is hence known.

The beam position can be measured from electrostatic pick-up electrodes with an accuracy of a fraction of a millimetre. For this the beam current has to be modulated.

The density distribution can be determined by scanning across the electron beam with a small Faraday cup or crossed wires.

Temperatures and velocity profiles are difficult to measure. The overall transverse temperature of the electron beam can be determined from the microwave radiation spectrum
emitted by the electrons spiralling in the magnetic field of the solenoid. The level of this radiation can be measured and from that the transverse electron temperature be deduced if the coupling of the radiation field to the antenna and the characteristics of the detection system are known. This method was used in ICE to minimize the transverse electron temperature. It is also applied for the LEAR electron cooler.

Longitudinal electron temperature and velocity profile have never been determined from the electron beam alone. It was only from actual cooling experiments that this information was derived. However, it can be measured by scattering laser light from the electron beam. The back-scattered laser light is shifted in frequency owing to the relativistically moving electrons. Analysing the light in frequency allows for the determination of velocity profile and longitudinal temperature non-destructively. This method is being attempted at the LEAR electron cooler.

12.2 Electron cooling diagnostics

During the electron cooling process the position and alignment of the ion and the electron beam have to be known precisely. Equally important is the information on the evolution of the ion-beam emittance and momentum spread. For the measurement of the stored-beam properties usually the standard beam diagnostics, such as Schottky and electrostatic pick-ups, are applied. A fast beam profile monitor, which allows monitoring of the beam size during the cooling, is also very useful.

Much information can be extracted from the measurement of the recombination channel. If protons are to be cooled, recombination produces a neutral hydrogen beam which leaves the storage ring tangentially. The beam size can be measured with multiwire proportional chambers or solid-state detectors. It allows the shrinkage of the beam to be observed during the cooling process (Fig. 22) and provides information on the equilibrium beam size. The rate
can be measured with scintillation counters. It contains information on the average transverse electron temperature (see Section 9). For bunched beams the time structure of the down-charged ion beam can be measured, enabling the determination of bunch length and intensity distribution within the bunch.

Another interesting method of diagnostics is the laser-induced recombination (see Section 13.2.1). There the capture of an electron by an ion is stimulated by irradiating the whole system with laser light of suitable frequency. Measuring the recombination rate as a function of the laser frequency allows one to scan across the electron velocity distribution and to deduce the local transverse and longitudinal electron beam temperatures. The threshold of the recombination provides information on the energy of the beams.

If partially stripped ions are to be cooled, dielectronic recombination (Section 13.2.3) can be used to measure the electron velocity distribution and through that the temperatures. It also makes it possible to determine the energies of the beams precisely.

13. APPLICATIONS OF ELECTRON COOLING

So far electron cooling experiments have been done between $\beta = 0.05$ and 0.57. At present electron coolers are being built which will go up to $\beta = 0.78$ and eventually to $\beta = 0.86 \ (\gamma = 2)$. Moreover, studies for an electron cooling system which could go up to $\gamma = 6.9$ are under way. The major domain for electron cooling will, however, remain in the cooling of ion beams of velocities below 0.8c.

One task of electron cooling will then be the compression of the phase space of circulating beams at injection energy to allow the accumulation of pulses and the build-up of high stored beam intensities even with low-current injectors (rare ions, polarized particles).

13.1 Internal targets

The achievement of high-intensity stored ion beams makes the performance of internal experiments in the ring very attractive. Apart from colliding-beam experiments, the use of thin internal targets provides an efficient way to utilize the ions repeatedly. If the target thickness is kept small enough, multiple scattering beam blow-up and energy loss can be compensated by electron cooling; and beam losses are then only due to single scatters with angles larger than the machine acceptance angle at the target position, or to nuclear reactions. The admissible target thickness $\rho d$ can be estimated from the emittance growth rate

$$\frac{d\varepsilon}{dt} = \varepsilon_{\text{ms}} = \frac{19.2}{\beta^2 \gamma^2} \frac{\rho d}{\hbar \nu} [\text{r.m.s. mrad s}^{-1}] \quad (\rho d \text{ taken in g cm}^{-2}),$$

which is counteracted by electron cooling. For simplicity let us take a constant cooling time. The equilibrium emittance is then given by the solution of the differential equation

$$\varepsilon_{\text{ms}} - \frac{1}{2} \varepsilon = 0 ,$$

which is

$$\varepsilon_{\text{eq}} = \frac{1}{2} \varepsilon_{\text{ms}} .$$
Assuming a ring can be filled to its space-charge limit at injection energy

\[ I_{\text{max}} = \frac{A_{\text{beam}}}{2z} \frac{2\pi e}{r_p} \frac{q^2 \gamma^2 A Q}{2\pi \hbar}, \]  

(63)

luminosities of

\[ L = \frac{I_{\text{max}}}{2e} \frac{N_L}{A_{\text{targ}}} \qquad (N_L = \text{Avogadro's number}), \]  

(64)

can be achieved. The important point in this operation is that the ions are not lost after their passage through the target, but are recycled.

13.2 Atomic physics

A large fraction of future experiments with stored ions cooled by electrons will most likely be devoted to atomic-physics investigations, making direct use of the electron cooler as an electron target. This comes from the fact that electron-ion collisions can be made at very well defined energies which have at the same time a high resolution. The basic processes to be studied will be the recombination of an electron from the cooler with circulating ions.

13.2.1 Radiative recombination

We have already discussed the formation of hydrogen atoms during electron cooling of protons, which was an important diagnostic in previous experiments. In general the capture process is

\[ e^- + A^{+\nu} \rightarrow (A^{(\nu-1)}\*)^+ + h\nu \]  

(65)

This process is the gateway to an exciting new field. The reaction could be enhanced by irradiating the system with laser light of suitable frequency

\[ h\nu + e^- + A^{+\nu} \rightarrow (A^{(\nu-1)}\*)^+ + 2h\nu. \]  

(66)

Steering the frequency allows one to populate, in a well-defined manner, specific atomic levels. This, in principle, allows Rydberg atoms to be formed in a very clean way and their properties to be studied.

13.2.2 Antihydrogen production

The (stimulated) radiative recombination is at present the only promising way to form the never before observed anti-hydrogen atom, by replacing the electron with a positron and the proton with an antiproton in an arrangement similar to electron cooling. This is a very tantalizing application of electron cooling.

13.2.3 Dielectronic recombination

If partially stripped ions are caused to overlap with electrons the latter can be captured without the emission of a photon. The energy of the free electron is then dissipated through simultaneous excitation of the residual electron core. It happens only at electron
energies which match with ion excitation energies and it has resonance character. This reaction is called dielectronic recombination and could only be studied poorly so far. The electron cooling arrangement could provide here also a very powerful experimental tool.

There are many other aspects where the electron cooling arrangement provides directly or indirectly a cleaner and more precise approach to interesting questions in atomic, nuclear, and particle physics, which were already discussed in the literature or which will come up with the new generation of cooler rings at present under construction.

14. ELECTRON COOLING PROJECTS

Given the enormous experimental potential of storage rings equipped with electron cooling, the interest in this field has increased very much in the past few years. In Table 3 a list is given of the coolers which are at present operating, or under construction, or planned.

Table 3

<table>
<thead>
<tr>
<th>First cooling expected</th>
<th>First ion beam expected</th>
<th>Name, place, country</th>
<th>Stored particles a)</th>
<th>Status b)</th>
<th>Bending field [T.m]</th>
<th>Ring circ. [m]</th>
<th>Max. electron energy [keV]</th>
<th>Length of cooling section [m]</th>
<th>Max. permeance [μP]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>1982</td>
<td>LEAN, CERN, Switzerland</td>
<td>p, p, H</td>
<td>running</td>
<td>9</td>
<td>78</td>
<td>40 (100)</td>
<td>1.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1988</td>
<td>1987</td>
<td>JINR Cooler, Bloomington, USA</td>
<td>up to Li</td>
<td>DC</td>
<td>3.6</td>
<td>87</td>
<td>300 (500)</td>
<td>3</td>
<td>0.65</td>
</tr>
<tr>
<td>1988</td>
<td>1987</td>
<td>CELSIUS, Upsala, Sweden</td>
<td>L+H+</td>
<td>DC</td>
<td>6.25</td>
<td>82</td>
<td>300 (500)</td>
<td>3</td>
<td>0.1</td>
</tr>
<tr>
<td>1988</td>
<td>1988</td>
<td>ESR, GSI, Darmstadt, FRG</td>
<td>up to U</td>
<td>F</td>
<td>10</td>
<td>103</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>1988</td>
<td>TSR, MFT, Heidelberg, FRG</td>
<td>up to I</td>
<td>DC</td>
<td>1.5</td>
<td>35</td>
<td>7</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>1988</td>
<td>1988</td>
<td>ASTRID, Aarhus, Denmark</td>
<td>L+</td>
<td>DC</td>
<td>2</td>
<td>40</td>
<td>3</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>1988</td>
<td>1988</td>
<td>CRYRING, Stockholm, Sweden</td>
<td>L+</td>
<td>DC</td>
<td>1.5</td>
<td>29</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>1987</td>
<td>TARN II, Tokyo, Japan</td>
<td>L+</td>
<td>DC</td>
<td>7</td>
<td>76</td>
<td>120</td>
<td>1.5</td>
<td>1</td>
</tr>
<tr>
<td>1987</td>
<td>1987</td>
<td>CERN, Jülich, FRG</td>
<td>up to Ne</td>
<td>P</td>
<td>8</td>
<td>180</td>
<td>100 (300)</td>
<td>3</td>
<td>0.1</td>
</tr>
<tr>
<td>1987</td>
<td>1987</td>
<td>MISTRAP, Oak Ridge, USA</td>
<td>L+</td>
<td>DC</td>
<td>2</td>
<td>47</td>
<td>100</td>
<td>2</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Frascati, Italy
Fermilab, USA
Univ. Frankfurt, FRG

a) LI: Light ions
H: Heavy ions
DC: Under construction
F: Funded
P: Planned
BIBLIOGRAPHY

First ideas on electron cooling

Reviews on electron cooling in general

Topical conferences on electron cooling

Electron cooling at other international conferences

Electron cooling experiments
N.S. Dikansky et al., The study of fast electron cooling, INP Novosibirsk preprint 79-56 (1979).
V.I. Kudelainen et al., Temperature relaxation in magnetized electron flux, INP Novosibirsk preprint 82-78 (1982), to be published in Sov. Phys. JETP.

**Electron cooling theory**

**Electron capture**

**Simulation of electron cooling**

**Applications and special aspects of electron cooling**

**Vacuum systems of electron coolers**
C. Habfast et al., Das Ultrahochvakuum-System des Elektronenkühlers für LEAR, Karlsruhe report KfK 3816 (1985).
Magnetic field

Electron beams

Electron collectors

Diagnostics of electron cooling

Internal Targets
M. Giesch et al., Implications of an internal target for antineutron production at LEAR, PS/DL/LEAR Note 81-4, 1981.

Theses about electron cooling
P. Möller Petersen, Studies of electron cooling in the ICE storage ring at CERN, University of Aarhus (1982).
ELECTRON DYNAMICS WITH RADIATION AND NONLINEAR WIGGLERS

JOHN M. JOWETT

CERN, Geneva, Switzerland

ABSTRACT: The physics of electron motion in storage rings is described by supplementing the Hamiltonian equations of motion with fluctuating radiation reaction forces to describe the effects of synchrotron radiation. This leads to a description of radiation damping and quantum diffusion in single-particle phase-space by means of Fokker-Planck equations. For practical purposes, most storage rings remain in the regime of linear damping and diffusion; this is discussed in some detail with examples, concentrating on longitudinal phase space. However special devices such as nonlinear wigglers may permit the new generation of very large rings to go beyond this into regimes of nonlinear damping. It is shown how a special combined-function wiggler can be used to modify the energy distribution and current profile of electron bunches.

1. INTRODUCTION

In this lecture we shall present some mathematical tools which are particularly useful in the study of electron or positron dynamics in storage rings and apply them to some important problems. However the emphasis is on understanding the physical content rather than the mathematics itself. Accordingly, some more technical material has been placed in appendices.

It is assumed that the reader has some familiarity with general accelerator theory and has, in particular, been introduced to the phenomenon of synchrotron radiation and how it affects electron motion. Excellent introductions to these topics were given in the first School of this series. Among other introductions, the classic lectures of Sands are especially worth reading. Texts in Hamiltonian dynamics and classical and quantum electrodynamics provide the physical background while books on the theory of stochastic processes will give more details of some of the techniques employed here. Since a rigorous mathematical discussion of the latter would get us irrecoverably sidetracked, we shall adopt a formal approach, trusting to intuition for the meaning of words like “random” and “noise”.

The last part of this lecture is devoted to the special topic of nonlinear wigglers which give rise to new dynamical phenomena amenable to description in terms of a Fokker-Planck equation and ideas from the stability theory of dissipative systems. Their use is limited to very large storage rings such as LEP and no such wiggler has yet been operated. However they open up an interesting new range of possibilities for controlling the parameters of the beams.

Radiation effects on betatron motion are not discussed in detail here because most of the mathematical techniques can be illustrated in connexion with synchrotron motion and the most interesting effects of nonlinear wigglers are longitudinal. We have taken the opportunity to provide a Hamiltonian formulation of synchrotron motion, specially geared for electron machines with localised RF cavities. This framework allows a natural development of longitudinal chromatic effects and the important notion of the damping aperture. Betatron motion was treated in Refs. 2, 10, 11 and, with the approach used here, in Ref. 13.
2. THE DYNAMICS OF ELECTRONS IN A STORAGE RING

We shall formulate the equations of motion for electrons (or positrons) in a storage ring. The Hamiltonian description of particle motion in a circular accelerator, regarded as a special configuration of external electric and magnetic fields, is familiar to the reader from other lectures in this School. Since it provides the shortest route from the general equations of motion of a charged particle in an electromagnetic field to the specific forms which these equations take in a storage ring (Hill equations for betatron motion etc.), we shall employ it freely to avoid a long recapitulation of basic accelerator physics.

On the other hand, our main interest here is the effect of synchrotron radiation on electron dynamics. And this cannot be described solely in the context of Hamilton's equations. We must add dissipative terms to describe the energy loss through radiation. Moreover, these terms must somehow include the essentially random nature of the photon emission process. Appropriate mathematical tools are found in the theory of stochastic processes, notably in stochastic differential equations and the associated Fokker-Planck equations. Here we shall follow the treatment outlined in Refs. 15 and 13 although the applications will be somewhat different; some alternative approaches to the mathematical side of the problem may be found in Refs. 16, 17 and 18.

2.1 Coordinate system and Hamiltonian

We shall use the curvilinear coordinate system of Courant and Snyder and follow the conventions of most of the standard optics programs (MAD, TRANSPORT, etc.).

![Diagram of reference orbit and Courant-Snyder coordinate system]

Fig. 1 The reference orbit and Courant-Snyder coordinate system

For simplicity, we assume that the magnets are perfectly aligned in the sense that there exists a closed planar reference curve $r_0(s)$, passing through all their centres which is also the closed orbit for a hypothetical reference particle of constant momentum $p_0$ which neither radiates nor couples to the RF...
accelerating fields. The position of a real particle of kinematic momentum \( \mathbf{p} \) is then described by giving the azimuthal position \( s \) of the closest point on this curve and its radial and vertical deviations \( x \) and \( y \) from that point, as shown in Fig. 1.

If we neglect edge effects in the magnets and use the Coulomb gauge then, in many important cases, the fields can be described in the electromagnetic fields can be derived from a single scalar function,\(^9\) the canonical vector potential\(^4\)

\[
A_s(x, y, t, s) = A \cdot e_s (1 + G(s)x) \\
= -\frac{p_0 e}{c} \left\{ xG(s) \left( 1 + G(s)\frac{x}{2} \right) + \frac{1}{2} K_1(s)(x^2 - y^2) + \frac{1}{6} K_2(s)(x^3 - 3xy^2) + \cdots \right\} \\
+ \sum_k \frac{e\hat{V}_k}{\omega_{rf}} \delta_C(s - s_k) \cos(\omega_{rf}t + \phi_k).
\]  
(2.1)

This includes only the fundamental accelerating mode of a set of RF cavities with peak voltages \( \hat{V}_k \) located at positions \( s_k \); \( \delta_C \) is a \( \delta \)-function, periodic on the circumference \( C = 2\pi R \). Our assumption about the closed orbit of the reference particle can only hold if the normalised dipole field strength is equal to the curvature of the reference orbit:

\[
G(s) \overset{\text{def}}{=} \frac{K''(s) \cdot x_0(s)}{|x_0(s)|}
\]  
(2.2)

(where primes denote derivatives with respect to \( s \)). Following Courant and Snyder,\(^9,^3\) we take \( s \) as an independent variable so that \((x, y, t)\) may be taken as canonical coordinates and \((p_x, p_y, -E)\) as canonical momenta. The Hamiltonian of a particle with kinematic momentum \( \mathbf{p} \) is

\[
H_s(x, y, t, p_x, p_y, -E; s) = (p + (e/c)A) \cdot e_s (1 + G(s)x) \\
= -\frac{(e/c)A_s(x, y, t, s)}{1 + G(s)x} \\
- (1 + G(s)x) \sqrt{E^2/c^2 - m^2c^2 - p_x^2 - p_y^2}.
\]  
(2.3)

Before proceeding further, it is convenient to replace the energy, \( E \), by the magnitude of the total momentum \( \mathbf{p} \); this requires a canonical transformation of one pair of variables,

\[
(t, -E) \mapsto (zt, p),
\]  
(2.4)

effected by means of the generating function

\[
F_z(p, t) = -ct \sqrt{p^2 + m^2c^2},
\]  
(2.5)

and resulting in the new Hamiltonian

\[
H(x, y, zt, p_x, p_y, p; s) = -\frac{e}{c} A_s(x, y, t(zt, p), s) - (1 + G(s)x) \sqrt{p^2 - p_x^2 - p_y^2}.
\]  
(2.6)

\(^4\) This recently introduced\(^9\) term saves a good deal of circumlocution.
The new canonical variables are related to the old by

\[ p = \sqrt{E^2/c^2 - m^2c^2}, \]
\[ z_t = -ct\sqrt{1 - m^2c^2/E^2} = -(\text{instantaneous velocity}) \times (\text{time particle passes s}). \]  

i.e. \( z_t \) is not immediately related to the path length, except while the energy is constant. The explicit form of Hamilton’s equations in these variables is

\[ x' = (1 + Gx) \frac{p_x}{\sqrt{p^2 - p_x^2 - p_y^2}} \approx (1 + Gx) \frac{p_x}{p} \]
\[ y' = (1 + Gx) \frac{p_y}{\sqrt{p^2 - p_x^2 - p_y^2}} \approx (1 + Gx) \frac{p_y}{p} \]
\[ z'_t = -(1 + Gx) \frac{p}{\sqrt{p^2 - p_x^2 - p_y^2}} \approx -(1 + Gx) \]
\[ p'_x = -G(p - p_0) - p_0(G^2 + K_1)x - \frac{1}{2}p_0K_2(x^2 - y^2) + \cdots \]
\[ p'_y = p_0K_1y + \frac{1}{2}p_0K_2xy + \cdots \]
\[ p' = -\sum_k \frac{e}{c} \delta_0(s - s_k) \sin(\omega t z_t/c + \phi_k). \]  

With these variables, the reference momentum \( p_0 \) factors out of all terms in \( H \) except those describing the cavities. Thus, particle motion in magnets will be “geometric”, depending only on \( \delta = (p - p_0)/p_0 \) and not on the mass or absolute value of \( p \). The price paid comes through the more complicated expression of the time-dependence of \( A_s \) in (2.6). There, \( t(z_t, p) \) denotes the solution of (2.7) for \( t \) in terms of \( z_t \) and \( p \). This does not matter much since the motion of high energy electrons is extreme-relativistic and the third argument of \( A_s \) in (2.6) may be set equal to \(-z_t/c\). From this also follows the excellent approximation

\[ t' \approx (1 + Gx)/c \]

For later convenience, let us define the normalised magnetic field strength \( b \) through

\[ \mathbf{B}(z, y, s) = \nabla \times \mathbf{A} = \frac{p_0c}{e} b(x, y, s) \]

and also write \( b(x, y, s) \) for \(|b(x, y, s)|\).

2.2 Statistical properties of incoherent synchrotron radiation

We review the essential facts about incoherent synchrotron radiation\(^{10,13,5,6}\) and recast them in a notation suited for our present purposes.

As a particle is accelerated transversely in a magnetic field, it emits photons. Because this is a quantum-mechanical phenomenon, the emission times and the quanta of energy carried away by the
photons are random quantities. However certain average quantities such as the mean emission rate and the mean radiated power may be calculated with good accuracy* within classical electrodynamics. In the classical picture, the accelerated particle emits a continuous beam of radiation in a narrow cone around its momentum vector. Quantum mechanically, the momentum vector of each photon is almost collinear with the particle's momentum.

Since the orbital quantum numbers of electrons in typical storage rings are very large, we may use classical arguments to construct the equations of motion provided we do not attempt to describe the emission process itself. In fact it will be represented simply as an instantaneous jump in energy. This is acceptable since, for an electron of energy \( E = \gamma mc^2 \) in a magnetic field \( B = E/e\rho \), photon emission occurs within a time

\[
\tau_\gamma \simeq \frac{\rho}{\gamma_\epsilon} \ll \frac{1}{\Omega},
\]

where \( \Omega \) is a frequency characteristic of betatron or synchrotron oscillations. In addition, the fact that

\[
\frac{1}{\tau_\gamma} \ll \omega_\epsilon,
\]

where \( \omega_\epsilon \) is the frequency corresponding to the critical energy (defined below), means that the frequency (or energy) spectrum of the photons is locally well-defined.²

Let us now develop these ideas formally.

An individual photon emission event, in which a photon of energy \( u_j \) is emitted at \( s = s_j \), is specified by the ordered pair of random variables \((u_j, s_j)\). The distribution function of \((u_j, s_j)\) depends only on the local magnetic field and the particle's momentum. Since these conditions vary as the particle moves, there is no very meaningful way of relating time averages (along the trajectory of a given particle) and ensemble averages (over many hypothetical particles experiencing the same conditions).

For definiteness, let us define the expectation value of a dynamical variable \( \mathcal{A}(x, y, z, p_x, p_y, p; s) \), associated with the instantaneous state of the particle, to be the average of \( \mathcal{A} \) over all possible realisations of \((u_j, s_j)\), that is to say, all the ways in which the particle's photon emission history might occur, weighted appropriately. We denote such ensemble averages by

\[
\langle \mathcal{A}(x, y, z, p_x, p_y, p; s) \rangle \quad \text{or, more briefly and generally,} \quad \langle \mathcal{A} \rangle_{X s}
\]

as it suits us; here \( X \) is a shorthand notation for the set of (usually canonical) variables describing the instantaneous state of the particle in whatever representation we happen to be using. The averaging is understood to be taken while the azimuthal position \( s \) and the phase-space coordinates \( X \)—and thereby the magnetic field felt by the particle—are supposed fixed. Parameters characterising the synchrotron radiation may be regarded as dynamical variables of the particle since they too are determined by \( X \) and \( s \). For example, the critical energy,

\[
u_\epsilon \stackrel{\text{def}}{=} \frac{3}{2} \frac{\hbar c p_0}{m c^2} p^2 b(x, y, s),
\]

may be thought of as a parameter determining the overall scale of the distribution in energy of the photons which the particle has a propensity to emit.

---

* Provided the energies and magnetic fields are not too high.
The exact density of a given realisation (holding $X$ fixed) in $(u, s)$ space is

$$\Omega_X(u, s) = \sum_j \delta(s - s_j)\delta(u - u_j),$$  \hspace{1cm} (2.15)

where the sum is taken over all events which actually take place. Its expectation value is the distribution function of $(u_j, s_j)$ (more correctly termed the probability density function)

$$(\Omega_X(u, s)) = N_X(s)f_X(u; s)/c,$$  \hspace{1cm} (2.16)

which factorises, reflecting the statistical independence of $s_j$ and $u_j$. Here,

$$N_X(s) = \frac{5\sqrt{3}}{6} \frac{e^3}{m^2 c^2 \hbar} |B(x, y, s)| = \frac{5\sqrt{3}}{6} \frac{e r_c p_0}{\hbar} b(x, y, s)$$  \hspace{1cm} (2.17)

is the distribution function of $s_j$, or the average photon emission rate, and is independent of the particle’s momentum; $r_c = e^2/mc^2$ is the classical electron radius.

![Fig. 2 Distribution of photons in energy](image)

The distribution of photon energies, $f_X(u; s)$, is closely related to the classical frequency spectrum of synchrotron radiation

$$f_X(u; s) = \frac{5\sqrt{3}}{8} \frac{(mc)^3}{\hbar c p_0} \frac{F(u/u_c)}{p^2 b(x, y, s)},$$  \hspace{1cm} (2.18)

where we follow the standard definition

$$F(\xi) \equiv S(\xi)/\xi, \quad S(\xi) \equiv \frac{9\sqrt{3}}{8\pi} \int_0^\infty K_{5/3}(\xi d\xi)$$  \hspace{1cm} (2.19)

and $K_{5/3}$ is a modified Bessel function (see e.g. Refs. 2, 5, 10 for derivations of (2.17) and (2.18)). The universal functions $F(\xi)$ and $S(\xi)$ are plotted in Fig. 2; note that a non-algebraic dependence of $f_X(u; s)$ on the momentum and magnetic field arises through the factor $u_c$ in the argument of $F$. 
The instantaneous radiated power is

\[ P_X(s) = 4 \int_0^\infty \Omega_X(u, s) \, du = 4 \sum_j \Omega_{su}(s - s_j). \] (2.20)

A typical realisation of \( P_X(s) \), obtained by simulation, is shown in Fig. 3; the parameters are such that the expectation value of the number of photons emitted in one revolution period is

\[ N_X(s)T_0 = 1000 \] (2.21)

Counting the peaks, we find 1049, a 1.55σ deviation. In preparing the figure, \( \tau_p \) was been taken to be fixed and equal to the width of a line on the printer so that the distribution of the heights of the peaks, displayed in units of the critical energy \( \epsilon_c \), is given by the function \( f_X(u; s) \) defined in (2.18). It is worth remarking that, although half the energy is carried away by photons with energies greater than \( \epsilon_c \), there are few such photons. In fact 91% of the photons have \( u < \epsilon_c \) and 50% have \( u < 0.1 \epsilon_c \).

The expectation value of \( P_X(s) \) is the classical power, given by the relativistic Larmor formula. Using the Lorentz force equation in 4-vector form \( p^\alpha = (E/c, \mathbf{p}) \) and proper time \( \tau \), for a particle in a purely magnetic field with \( \mathbf{p} \cdot \mathbf{B} = 0 \)

\[ \frac{dp^\mu}{d\tau} = \frac{e}{mc} (0, \mathbf{p} \times \mathbf{B}), \] (2.22)

we express this in terms of the canonical variables

\[ \langle P_X(s) \rangle = \frac{2}{3} \frac{e^2}{m^2 c^3} \frac{dp^\mu}{d\tau} dp^\mu = \frac{2}{3} \frac{e^4}{m^4 c^3} (\mathbf{p} \times \mathbf{B})^2 \]

\[ = \frac{2}{3} \frac{e^2 r_p}{(mc)^3} \left| \mathbf{B}(x, y, s) \right|^2 = \frac{2}{3} \frac{r_p^2}{m^3 c} \mathbf{p}^2 (x, y, s)^2 \] (2.23)

\[ = N_X(s)(u)X_s \]
The last form uses the mean photon energy

$$
\langle u \rangle_{X_s} = \int_0^\infty u f_X(u; s) \, du = \frac{4 \hbar c p_0}{5 \sqrt{3} (mc)^3} u^2 b(x, y, s) = \frac{8}{15 \sqrt{3}} u_c.
$$

Similarly, the mean-square quantum energy is the second moment of $f_X(u; s)$:

$$
\langle u^2 \rangle_{X_s} = \int_0^\infty u^2 f_X(u; s) \, du = \frac{11 (\hbar c p_0)^2}{12 (mc)^6} p^4 b(x, y, s)^2.
$$

The exact power (2.20) may be split into its mean and fluctuating parts:

$$
P_X(s) = \langle P_X(s) \rangle + \hat{P}_X(s),
$$

where $\hat{P}_X(s)$ is just the difference between the classical power (2.23) and the instantaneous power in a given realisation.

The two-time correlation function of such a quantity is given by (a generalised version of) Campbell's Theorem,\textsuperscript{21,9} closely related to the well-known Schottky formula,

$$
\langle \hat{P}_X(s) \hat{P}_X(s') \rangle = c N_X(s) \langle u^2 \rangle_{X_s} \delta(s - s')
= \frac{55}{24 \sqrt{3}} \frac{r_s \hbar c p_0^3}{(mc)^6} p^4 b(x, y, s)^2 \delta(s - s').
$$

The $\delta$-function expresses the fact that $\hat{P}_X(s)$ and $\hat{P}_X(s')$ are uncorrelated when $s \neq s'$. This really means $|s - s'| \gg c r_t$ with $r_t$ as defined in (2.11).

Let us introduce a unit noise source, $\xi(s)$, known technically\textsuperscript{8,7} as a centred, Gaussian Markov process. It is defined to have the formal properties

$$
\langle \xi(s) \rangle = 0, \quad \langle \xi(s) \xi(s') \rangle = \delta(s - s').
$$

with respect to our ensemble-averaging operation $\langle \ldots \rangle$. With this, we can correct a formal representation of the stochastic power which reproduces the essential properties derived above, namely (2.23) and (2.27),

$$
P_X(s) = p^2 c^2 |c_1 b(x, y, s)^2 + \sqrt{c_2} |b(x, y, s)|^{3/2} \xi(s)|
$$

where the constants\textsuperscript{*} $c_1$ and $c_2$ are

$$
c_1 \overset{\text{def}}{=} \frac{2 r_s \hbar p_0^2}{3 (mc)^3}, \quad c_2 \overset{\text{def}}{=} \frac{55 r_s \hbar p_0^3}{24 \sqrt{3} (mc)^6}.
$$

Statistically, there is no way of distinguishing our conceptual model of discrete random photon emission and this formal representation.

Noting that $c_2 \propto \hbar$, we see that, in this formalism, the classical radiation power has been corrected by a stochastic term of order $\sqrt{\hbar}$. We also observe that, in general, the average radiation power and its quantum fluctuations depend nonlinearly on the particle's coordinates through the spatial dependences of the magnetic field.

\textsuperscript{*} These are not "fundamental" constants because they still depend on the absolute value of the bending field in the ring through the reference momentum $p_0$. 
2.3 Radiation reaction forces

Now that we know the distribution of photons, we can include their effect on the motion of the electrons by adding radiation reaction forces to Hamilton’s equations (2.8).

A single photon emission of energy $u_j/c$ at azimuth $s = s_j$ (when $t = t_j$, say) will produce an abrupt (since $r_\gamma$ is short) change in the momentum but leave the spatial position of the particle unchanged. At high energy the opening angle of the beam of radiation is

$$\theta_{\text{max}} \approx \frac{1}{2\gamma} \approx 0.26 \text{ mrad} \quad \text{at} \quad E_0 = 1 \text{ GeV}. \quad (2.31)$$

It is therefore an excellent approximation to take the photon’s 3-momentum vector $u_j/c = u_j p/c$ to be collinear with the momentum $p$ and apply momentum conservation to evaluate the changes in the canonical momenta:

$$p \rightarrow p - u_j/c,$$

$$p_x \rightarrow p_x - \frac{u_j}{c} p_x = p_x - \frac{u_j}{c} \frac{x'}{\sqrt{1 + (x')^2 + (y')^2}} = p_x - \frac{u_j}{c^2} \frac{x'}{t'},$$

$$p_y \rightarrow p_y - \frac{u_j y'}{c^2} \frac{1}{t'}.$$

If we consider a time-interval surrounding the moment of photon emission, which is so short that the probability of more than one photon being emitted can be neglected, then we know that the energy of the emitted photon is equal to the time-integral of the fluctuating radiation power (2.20),

$$u_j = \int_{s_j - \varepsilon}^{s_j + \varepsilon} P_X(s) \frac{ds}{c} \quad (\text{with probability} \rightarrow 1 \text{ as} \ \varepsilon \rightarrow 0^+), \quad (2.33)$$

and we may reinterpret (2.32) as stochastic differential equations

$$dp = -P_X(s) \frac{dt}{c} = -P_X(s) t' \frac{ds}{c} = -P_X(s) \frac{dz}{c^2} + O(\gamma^{-2}) \frac{ds}{c^2},$$

$$dp_x = -P_X(s)(x' + t') \frac{dt}{c^2} = -P_X(s) x' \frac{ds}{c^2},$$

$$dp_y = -P_X(s)(y' + t') \frac{dt}{c^2} = -P_X(s) y' \frac{ds}{c^2}. \quad (2.34)$$

To complete the equations of motion, we must restore the forces due to the direct action of the external fields given by Hamilton’s equations, $x' = \partial H/\partial p_x$, etc. In this way we find

$$x' = \frac{\partial H}{\partial p_x}, \quad p_x' = -\frac{\partial H}{\partial x} - \frac{P_X(s) \partial H}{c^2 \partial p_x},$$

$$y' = \frac{\partial H}{\partial p_y}, \quad p_y' = -\frac{\partial H}{\partial y} - \frac{P_X(s) \partial H}{c^2 \partial p_y},$$

$$z' = \frac{\partial H}{\partial p_z}, \quad p_z' = -\frac{\partial H}{\partial z} + \frac{P_X(s) \partial H}{c^2 \partial p_z}. \quad (2.35)$$

The Hamiltonian part of these equations has already been written out explicitly in (2.8), but it is instruct-

---

* We are neglecting a very, very small increase in $z_i$ due to the small reduction in the velocity of the electron; see (2.7). It is easy to check that this is utterly negligible.
tive to write out the radiation terms in detail:

\[
\begin{align*}
p_x' &= -\frac{\partial H}{\partial x} - (1 + G z) p y_p c_1 b(x, y, s)^2 + \sqrt{c_2 b(x, y, s)^2} \xi(s) \overset{\text{def}}{=} -\frac{\partial H}{\partial x} + \frac{p_0}{c} \Pi_x, \\
p_y' &= -\frac{\partial H}{\partial y} - (1 + G z) p x_p c_1 b(x, y, s)^2 + \sqrt{c_2 b(x, y, s)^2} \xi(s) \overset{\text{def}}{=} -\frac{\partial H}{\partial y} + \frac{p_0}{c} \Pi_y, \\
p' &= -\frac{\partial H}{\partial z_t} - (1 + G z) p^2 c_1 b(x, y, s)^2 + \sqrt{c_2 b(x, y, s)^2} \xi(s) \overset{\text{def}}{=} -\frac{\partial H}{\partial y} - \frac{p_0}{c} \Pi_t, \\
\end{align*}
\]  

(2.36)

This also serves to define the radiation coupling functions \( \Pi_x, \Pi_y \) and \( \Pi_t \).

Notice the dependence on the canonical momenta—this is at the root of Robinson’s Theorem on the damping partition numbers.\(^{22,2}\)

3. NORMAL MODES AND OPTICAL FUNCTIONS

In principle, the equations (2.35) completely describe electron motion under the combined influences of the applied electromagnetic fields and synchrotron radiation. At a fundamental level their physical content is manifest but they are not in a form suitable for many practical calculations. Other lectures in this School have shown how useful it is to describe particle trajectories first of all in terms of the three normal modes of linearised motion around the closed orbit in the 6-dimensional phase space and then in terms of the optical functions which characterise the storage ring lattice and determine the frequencies of these modes. In a planar ring with \( x-y \) coupling terms, such as we have assumed, these are the familiar modes of linearised betatron and synchrotron motion.

In the coordinates \((x, y, z_t)\), the Hamiltonian contains linear coupling terms between \( x \) and \( p \) due to the spectrometer effect of the (horizontal) bending magnets, but these may be eliminated by introducing the dispersion functions.

The remainder of this section may be skimmed by the reader who does not wish to be convinced of each step in the introduction of the dispersion, functions and the concepts of synchrotron and betatron motion as they emerge in the Hamiltonian formulation used here. He will be familiar with these notions from other lectures in the School. The following sections are included principally to cover certain aspects peculiar to electron machines.

Behind the formalism, however, there lie a few key physical ideas which are essential to the understanding of what follows. In particular the reader should be aware of the distinction between the two components of the momentum deviation \( \delta_s \) and \( \epsilon \) (to be introduced) and the way in which the energy loss by synchrotron radiation is coupled into the transverse oscillations through the dispersion functions.

3.1 Synchrotron motion

Let us simplify the Hamiltonian (2.6) by neglecting higher order kinematic terms in the transverse momenta and in the momentum deviation \( \delta \). Accordingly, we approximate the square root term by

\[
(1 + G z) \sqrt{p^2 - p_x^2 - p_y^2} \approx (1 + G z) p - \frac{p_x^2 + p_y^2}{2p} + \cdots
\]

(3.1)

Before writing down the Hamiltonian, we perform a simple rescaling of variables which makes all the
momenta dimensionless:

\[ H \mapsto H_1 = H/p_0, \quad p_x \mapsto P_x = p_x/p_0, \quad p_y \mapsto P_y = p_y/p_0, \quad p \mapsto P = p/p_0. \]  

(3.2)

The canonical coordinates and the independent variable all have dimensions of length and the Hamiltonian is

\[
H_1(x, y, z_1, p_x, p_y, p; s) \simeq -Gx(P - 1) + \frac{P_x^2 + P_y^2}{2P} \\
+ \frac{G^2x^2}{2} + \frac{1}{2}K_1(x^2 - y^2) + \frac{1}{6}K_2(x^3 - 3xy^2) + \cdots \\
- \sum_k \frac{eV_k}{p_0\omega k} \delta_C(s - s_k) \cos(\omega k s_1/c + \phi_k). 
\]  

(3.3)

The dispersion functions \( \eta \) and \( \zeta \) are designed to eliminate the linear coupling appearing in the first term. Some higher-order couplings can be eliminated at the same time by allowing these functions to depend on momentum\(^n\) and this approach is often used in nonlinear optics studies when synchrotron motion is neglected.\(^*\) On the other hand, for electron rings, where the value of \( p \) oscillates relatively rapidly and certainly must be included as a dynamical variable, it appears at first sight that the simplest approach would consist in defining \( \eta \) and \( \zeta \) with respect to the reference momentum \( p_0 \). This avoids having a Hamiltonian which depends on a canonical momentum through functions which have to be calculated (and, eventually, differentiated) numerically.\(^\dagger\)

However we can do a little better than this if we recognize that, depending on the precise value of the RF frequency, the equilibrium momentum of the beam may not be equal to \( p_0 \); synchrotron oscillations will then take place around a slightly different value of the momentum which we shall denote as \( p_0(1 + \delta_s) \), with \( \delta_s \ll 1 \). With this in mind, the dispersions may be introduced by means of a canonical transformation

\[
(x, y, z_1, P_x, P_y, P) \mapsto (x\beta, y, z, p_x, p_y, p)
\]  

(3.4)

whose generating function is

\[
F_\beta(p_0, p_1, p, x, y, z_1) = p_0 \left[ x - \eta(\delta_s, s)(\delta_s + \epsilon) \right] + x\zeta(\delta_s, s)(\delta_s + \epsilon) + p_y y \\
+ (1 + \delta_s + \epsilon) \left[ z_1 + Z_0(s) \right].
\]  

(3.5)

In this expression, \( Z_0(s) \), \( \eta(\delta_s, s) \) and \( \zeta(\delta_s, s) \) are as yet unspecified functions of \( s \); natural choices for them will emerge in the following. In order to take proper account of chromatic effects in cases where the equilibrium value of \( p \) is other than \( p_0 \), they have also been allowed to depend parametrically on the constant \( \delta_s \). Later we shall show how the value of \( \delta_s \) is determined naturally by the RF frequency. When the equations of motion are constructed from the new Hamiltonian there is no need to differentiate \( \eta \) or \( \zeta \) with respect to \( \delta_s \).

---

* E.g. for hadron colliders where the synchrotron oscillation frequency is very low.

\(^\dagger\) Despite what is said in the following paragraphs, it may yet prove convenient to define the dispersion in this way because it provides a measure of how much the equilibrium orbit differs from the reference orbit, presumed to pass through the reference points of the beam position monitors at the centres of the magnet apertures. In practice a value quoted for the dispersion will almost always be this one, denoted below as \( \eta(0,s) \).
The new coordinates and momenta are given by
\[
x_\beta = \frac{\partial F_1}{\partial \rho_\beta} = x - \eta(\delta_s, s)(\delta_s + \epsilon), \quad P_\beta = \frac{\partial F_1}{\partial \epsilon} = p_\beta + \zeta(\delta_s + \epsilon),
\]
\[
y = \frac{\partial F_2}{\partial \rho_y} = y, \quad P_y = \frac{\partial F_2}{\partial \rho_y} = p_y,
\]
\[
z = \frac{\partial F_2}{\partial \epsilon} = z_1 + Z_0(s) - \eta p_\rho + \zeta x_\beta + \eta \zeta(\delta_s + \epsilon), \quad P = \frac{\partial F_2}{\partial z_1} = 1 + \delta_s + \epsilon,
\]
\[
H_2 = H_1 + \frac{\partial F_2}{\partial s} = H - p_\beta(\delta_s + \epsilon) \eta' + [z_\beta + \eta(\delta_s + \epsilon)](\delta_s + \epsilon) \zeta'.
\]
(3.7)

The splitting of \( x \) into its betatron and "energy" components should be familiar. It is perhaps less well-known that, in order to preserve the canonical structure (symplecticity), one must also use a new longitudinal coordinate \( z \) which takes account of local changes in the length of the particle's orbit due to its betatron oscillations. Since there is no vertical bending in our perfect machine, there is no vertical dispersion and the \( y \) transformations are trivial.

Expressing the new Hamiltonian \( H_2 \) in terms of the new coordinates, we can eliminate coupling terms linear in \( x_\beta \) by imposing the conditions
\[
\frac{\partial H_2}{\partial x_\beta} = \frac{\partial H_2}{\partial p_\beta} = 0 \quad \text{for} \quad x_\beta = p_\beta = y = p_y = z = \epsilon = 0.
\]
(3.8)

Writing these out explicitly, we find that \( \eta \) and \( \zeta \) must satisfy first-order differential equations and a periodicity condition,
\[
\eta' = \frac{\zeta}{1 + \delta_s}, \quad \zeta' = G - (K_1 + G^2)\eta - \frac{1}{2} K_2 \eta^2 \delta_s, \quad \eta(\delta_s, s + 2\pi R) = \eta(\delta_s, s),
\]
(3.9)

which are nothing but the familiar equations defining the dispersion function;** we emphasise again that \( \delta_s \) appears as a parameter and that primes denote differentiation with respect to \( s \). A common practical means of determining these functions for a range of values of \( \delta_s \) is to expand them as
\[
\eta(\delta_s, s) = \eta_0(s) + \eta_1(s) \delta_s + \ldots, \quad \zeta(\delta_s, s) = \zeta_0(s) + \zeta_1(s) \delta_s + \ldots,
\]
(3.10)

and equate coefficients of \( \delta_s \) in the equations (3.9). Then each function may be evaluated once and for all, independently of \( \delta_s \).

From (2.36) and (3.6) it is straightforward to work out† the new equations of motion
\[
x_\beta' = \frac{\partial H_2}{\partial \rho_\beta} + \Pi_\epsilon \eta / c, \quad p_\beta' = -\frac{\partial H_2}{\partial x_\beta} + (\Pi_\tau + \Pi_\epsilon \zeta) / c,
\]
\[
y' = \frac{\partial H_2}{\partial \rho_y}, \quad p_y' = -\frac{\partial H_2}{\partial y} + \Pi_y / c,
\]
\[
z' = \frac{\partial H_2}{\partial \epsilon} - \Pi_\epsilon \eta / c - \Pi_\zeta \zeta / c, \quad \epsilon' = -\frac{\partial H_2}{\partial \epsilon} - \Pi_\epsilon / c.
\]
(3.11)

Defining an effective quadrupole gradient‡ for a particle with the reference momentum which happens

** Sometimes \( \eta \) is denoted \( D, \psi \) or \( \alpha_\eta(s) \) but there are no other notations for the conjugate function \( \zeta(\delta_s, s) \); we have introduced it to exhibit the fact that (3.9) can themselves be derived from a Hamiltonian.

† Alternatively we could use the transformation theory of Appendix B.15

‡ Had we included octupole fields, it would also have been natural to define an effective sextupole gradient in a similar fashion.
to be at the position of the off-momentum orbit\footnote{A further step might be to divide the Hamiltonian and all magnetic field terms by $(1 + \delta_s)$ so that the rôle of the original reference momentum $p_0$ would be taken over by $p_0(1 + \delta_s)$. This is sometimes convenient. However since it is often useful to take advantage of the properties of the separated-function lattice which is used for most electron machines, we shall refrain from taking this step. The interested reader may consult Ref. 20.}

\begin{equation}
  k_1(\delta_s, s) = K_1(s) + \frac{1}{2} K_2(s) \eta(\delta_s, s) \delta_s
\end{equation}

we find, after a good deal of algebra, exploiting the cancellations implied by (3.8) and dropping several constant terms, that the Hamiltonian is

\begin{equation}
  H_2(x, y, p_x, p_y, \varepsilon; s) = \frac{1}{2} \left( \varepsilon^2 - (G^2 + k_1) \eta^2 \right) (\delta_s + \varepsilon)^2 + \left[ Z_0(s) - 1 \right] \varepsilon
\end{equation}

\begin{equation}
  + \frac{1}{6} K_2(e - \delta_s/2)(\delta_s + \varepsilon)^3
\end{equation}

\begin{equation}
  + \frac{p_x^2 + p_y^2}{2(1 + \delta_s)} + \frac{G^2 x^2}{2} + \frac{1}{2} k_1(x_y^2 - y^2) + \frac{1}{6} K_2(x_y^2 - 3x_y y^2)
\end{equation}

\begin{equation}
  - \sum_k \frac{eV_k}{p_0 \omega_{rf}} \delta_C(s - s_k) \cos \left\{ \frac{\omega_{rf}}{c} \left[ z - Z_0(s) + \eta p_y - \zeta x_p - \zeta \eta(\delta_s + \varepsilon) + \phi_k \right] \right\}.
\end{equation}

Although the terms describing betatron motion are simplified, the local formulation of synchrotron motion appears fairly complicated. However we recall that, in order to avoid dangerous synchro-betatron coupling effects, storage rings are almost always designed so that the dispersion functions vanish at the locations of the RF cavities:

\begin{equation}
  \eta(s_k) = \zeta(s_k) = 0, \quad \text{for each } k.
\end{equation}

[In practice, of course, imperfections will usually create some horizontal and vertical dispersion in the cavities.] Then, thanks to the $\delta$-functions, the phase of the cosine describing the RF waveform simplifies to

\begin{equation}
  \Phi(z, s) = \frac{\omega_{rf}}{c} \left[ z - Z_0(s) + \phi_k \right],
\end{equation}

and all coupling effects between the longitudinal and transverse motions have been eliminated. We remain free to choose the function $Z_0(s)$ to our best advantage. A formal analogy with (3.8) prompts us to demand that

\begin{equation}
  \frac{\partial H_2}{\partial \varepsilon} = 0 \quad \text{for} \quad x_p = p_y = y = p_y = z = \varepsilon = 0.
\end{equation}

The physical interpretation of this condition is clear if we notice that (in the absence of radiation effects) the change in $z$ around an orbit will be

\begin{equation}
  z(2\pi R) - z(0) = \int_0^{2\pi R} z'(ds) = \int_0^{2\pi R} \frac{\partial H}{\partial \varepsilon} ds,
\end{equation}

and the condition (3.16) determines a shift in the origin of phase space to a fixed point of the mapping which describes the evolution of the phase space coordinates over one turn. Moreover we are insisting that this hold true at every point on the circumference. Alternatively, we can describe this as a canonical transformation to a reference frame moving with the synchronous particle.
Working out (3.16) explicitly, we find that the effects of the sextuple terms in the longitudinal part of the Hamiltonian cancel, leaving us with

\[
\left\{ \frac{s^2}{1 + \delta_s} - (G^2 + k_1)\eta^2 \right\} \delta_s + Z_0'(s) - 1 = 0, \tag{3.18}
\]

which is straightforwardly integrated to yield

\[
Z_0(s) = -z_s + s - \delta_s \int_0^s \Gamma(\delta_s, s) \, ds, \tag{3.19}
\]

where \(z_s\) is a constant, related to the stable phase angle and we defined the local path length slippage function by

\[
\Gamma(\delta_s, s) \overset{\text{def}}{=} \frac{s^2}{1 + \delta_s} - (G^2 + k_1)\eta^2. \tag{3.20}
\]

In Appendix C we give the details of the Fourier analysis of this function, showing that its average value is just the negative of the momentum compaction factor, \(\alpha_s\), and how one can construct the smooth approximation to synchrotron motion from this local description.

Neglecting unimportant sextuple terms, the Hamiltonian for local synchrotron motion is now

\[
H_s(z, \epsilon, s) = \frac{1}{2} \Gamma(\delta_s, s) \epsilon^2 - \sum_k \frac{eV_k}{p_0 c \omega_{rf}} \delta_C(s - s_k) \cos \left\{ \frac{\omega_{rf}}{c} (z + z_s) - \frac{\omega_{rf}}{c} \int_0^s \Gamma(\delta_s, \sigma) \, d\sigma + \phi_k \right\}. \tag{3.21}
\]

Finally the requirement that this Hamiltonian be periodic in \(s\),

\[
H_s(z, \epsilon, s + 2\pi R) = H_s(z, \epsilon, s), \tag{3.22}
\]

means that the argument of the cosine must advance by an integer multiple \(2\pi\) per revolution. In the limit \(\delta_s \to 0\), the RF frequency has to be an integer multiple of the revolution frequency on the reference orbit:

\[
\omega_{rf} = 2\pi f_{rf} = 2\pi \frac{h \epsilon}{R}, \quad (\delta_s = 0), \tag{3.23}
\]

where \(h\) is called the harmonic number. From this it follows that (3.22) is equivalent to

\[
- \frac{\omega_{rf}}{c} (2\pi R) + \frac{\omega_{rf}\delta_s}{c} \int_0^{2\pi R} \Gamma(\delta_s, s) \, ds = -2\pi h \tag{3.24}
\]

\[
\Leftrightarrow \quad \delta_s = -\frac{1}{\alpha_s(\delta_s)} \left( 1 - \frac{hc}{\omega_{rf} R} \right)
\]

The equilibrium momentum of the beam may be determined by small shifts of the RF frequency. Writing
$f_{\text{RF}} = f_0 + \Delta f_{\text{RF}}$, we may write the familiar linearised version of this relationship,

$$\delta_s \approx -\frac{1}{\alpha_c(0)} \frac{\Delta f_{\text{RF}}}{f_{\text{RF}}},$$  \hspace{1cm} (3.25)

and exhibit the dependence of the average radius of the equilibrium orbit on the momentum compaction factor:

$$R(\delta_s) \overset{\text{def}}{=} \frac{hc}{\omega_{\text{RF}}} = R(1 + \alpha_c(\delta_s)\delta_s), \quad \frac{dR}{\delta_s} = \alpha_c(\delta_s).$$  \hspace{1cm} (3.26)

In Appendix C, the details of the phasing of the RF cavities are worked out and it is shown that, in smooth approximation, one may replace (3.13) with the simplified Hamiltonian

$$H_3(x, y, z, p_x, p_y, \varepsilon; s) = -\frac{\alpha e^2}{2} - \frac{e \gamma}{2 \pi \rho c \hbar} \cos \left( \frac{\hbar (z + z_s)}{R} \right)$$

$$+ \frac{p_y^2 + p_y^2}{2(1 + \delta)} + \frac{G^2 x^2}{2} + \frac{1}{2} k_1(x^3 - y^3) + \frac{1}{6} K_2(x^3 - 3xy^3).$$  \hspace{1cm} (3.27)

Including the radiation reaction effects, the equations of motion are given by (3.11). To make the radiation terms explicit, we have to work forward through the chain of variable substitutions from the original forms of $\Pi_s$ and $\Pi_t$ as functions of $(x, p, s)$. These operations are deferred to the next section.

In this formulation of synchrotron motion it might appear that we are always above transition energy and have somehow neglected the possibility of the revolution frequency’s increasing with momentum as it does below transition. This appearance is only a consequence of our having used $s$, and not time $t$, as independent variable. Transition energy does indeed occur when the increase in time taken to cover a greater orbit length due to a momentum deviation is exactly compensated by the greater velocity of the particle on that orbit. If the transformation of independent variable is made, and if higher order terms are included in (2.9), the velocity (and hence the familiar $\gamma^{-2}$ factor) enters explicitly. A canonical transformation then restores variable sign in the coefficient of $e^2$.

4. RADIATION DAMPING

The deterministic parts of the equations (3.11) show how the transverse momenta are damped directly by photon emission and how, moreover, the dispersion function couples the longitudinal damping into the radial phase space. As mentioned in the Introduction, we shall not discuss betatron motion any further than this.

4.1 Damping in longitudinal phase space

Let us set $x = p = y = p_y = 0$ and study the effects of radiation on the dynamics of longitudinal phase space in smooth approximation. Carrying through the changes of variables, we find from (2.36),
that the deterministic parts of the equations of synchrotron motion are

\[
\varepsilon' = \frac{e\hat{V}}{2\pi R(\delta_s)p_0c} \sin \left( h(x + z_s)/R \right) - \frac{1}{2\pi R} \int_0^{2\pi R} ds \langle \Pi_t/c \rangle,
\]

\[
z' = -\alpha_s \varepsilon - \frac{1}{2\pi R} \int_0^{2\pi R} ds \langle \Pi_s \eta/c - \Pi_\zeta \zeta/c \rangle.
\]

(4.1)

The only dissipative term which does not average out is that in the equation for \( \varepsilon' \) and is simply related to the average energy loss. Integrating the definition of \( \Pi_t \) contained in (2.35) and (2.36) we find that

\[
p_0 \int_0^{2\pi R} \Pi_t ds = \int_0^{2\pi R} P_X(s)t' ds = U(\delta_s + \varepsilon)
\]

(4.2)

is the energy loss per turn of a particle with total momentum \( p_0(1 + \delta_s + \varepsilon) \).

The normalised magnetic field strength at a displacement \( z \) in the median plane is

\[
b(x, 0, s) = G(s) + K_1(s)x + \frac{1}{2} K_2(s)x^2 + \cdots.
\]

(4.3)

Taking into account the energy lost in the lattice dipoles and quadrupoles, we can write out the first few terms in the expansion of the expectation value in powers of \( \delta_s \) and \( \varepsilon \):

\[
U(\delta_s + \varepsilon) \approx p_0^2 c \int_0^{2\pi R} ds \left\{ 1 + (2 + G(s)\eta)(\delta_s + \varepsilon) \right\} c_1 b(\eta(\delta_s + \varepsilon), 0, s)^2
\]

(4.4)

\[
= c_1 p_0^2 c \left\{ I_2 + \delta_s(2I_2 + I_4) + \varepsilon(2I_2 + I_4) + (\delta_s^2 + 2\varepsilon\delta_s)(I_2 + 2I_4 + I_8) + O(\varepsilon^2) \right\}.
\]

The arguments of \( \eta \) have been suppressed and the definitions of the synchrotron radiation integrals \( I_2, I_4 \) and \( I_8 \) will be found in Appendix D.

A fixed point of the equations (4.1) is a point where \( \varepsilon' = z' = 0 \) and it is easy to see that one exists on the line \( \varepsilon = 0 \) in the phase plane. The still undetermined constant \( z_s \) can now be chosen according to

\[
e^\hat{V} \sin(hz_s/R) = U(\delta_s) = c_1 p_0^2 c \left\{ I_2 + (2I_2 + I_4)\delta_s + \delta_s^2(I_2 + 2I_4 + I_8) \right\}
\]

(4.5)

so as to move the origin to this natural position.

In the case of a stable fixed point, \( \phi_s = hz_s/R \) is called the stable phase angle. A particle which maintains this phase relationship with the RF wave will find that its acceleration just balances its average energy loss by synchrotron radiation.
Introducing the damping partition number$^{2,22,25}$

$$J_{\epsilon}(\delta_{a}) = \frac{d \log U(\delta_{a})}{d\delta_{a}} = 2 + \frac{I_{4}}{I_{2}} + \delta_{a} \left( \frac{2I_{6}}{I_{2}} - \frac{I_{3}^{2}}{I_{2}} - 2 \right) + O(\delta_{a}^{2}), \tag{4.6}$$

e negligent

we can linearise (4.1) (still neglecting the fluctuation terms) to find

$$z' = -\alpha_{c} c, \quad \epsilon' = \frac{eV \cos \phi}{2\pi R^{3}h_{p0}c} z - J_{\epsilon}(\delta_{a}) \frac{U(\delta_{a})}{p_{0}c^{2}} f_{0} \epsilon. \tag{4.7}$$

These, of course, are the equations of a damped linear oscillator with natural frequency $\Omega_{e}$ given by

$$\alpha_{c}^{2} \Omega_{e}^{2} = \frac{\alpha_{c} eV}{2\pi R^{3}h_{p0}c} \cos \phi_{a}, \tag{4.8}$$

and are equivalent to a single second-order differential equation

$$\ddot{z} + 2\alpha_{c} \dot{z} + \Omega_{e}^{2} z = 0, \tag{4.9}$$

where the damping rate $\alpha_{c}$ and the damping time $\tau_{e}$ are defined by

$$\alpha_{c} \overset{\text{def}}{=} \frac{1}{\tau_{e}} \overset{\text{def}}{=} \frac{J_{\epsilon}(\delta_{a})}{2\tau_{e}} \overset{\text{def}}{=} \frac{J_{0}(\delta_{a})U(\delta_{a})}{2p_{0}c} f_{0} = \frac{\tau_{e} f_{0}}{3} \left( \frac{p_{0}}{m_{c}} \right)^{3} \left( 2I_{2} + I_{4} \right). \tag{4.10}$$

The damping rate $\alpha_{c}$ coincides with the quantity $\alpha_{c}$ when $J_{\epsilon} = 2$. This case is a useful reference point, as will be explained below.

### 4.2 Damping partition numbers and damping aperture

Although we have not given the derivations here, damping partition numbers analogous to (4.6) also exist for the radial and vertical betatron oscillations.$^{10}$ While the damping is linear, they satisfy the sum rule known as Robinson's Theorem:

$$J_{e}(\delta_{a}) + J_{y}(\delta_{a}) + J_{y}(\delta_{a}) = 4. \tag{4.11}$$

for all values of $\delta_{a}$ such that an off-momentum closed orbit given by $\eta(\delta_{a}, z)$ exists. This holds even if one of the partition numbers is negative. In most lattice designs, $J_{y} = 1$.

By varying the RF frequency, and thereby $\delta_{a}$, it is therefore possible to redistribute the damping between the longitudinal and radial modes. In a storage ring, one must ensure that each damping partition number remains positive. The range of values of $\delta_{a}$ in which this is true is called the damping aperture, and is determined by the values of the synchrotron integrals $I_{2}$, $I_{4}$ and $I_{6}$:

$$- \frac{2I_{2} + I_{4}}{2I_{6}} < \delta_{a} < \frac{I_{2} - I_{4}}{2I_{6}}. \tag{4.12}$$

Together with (3.25), this translates directly into an allowable range of variation of RF frequencies or an allowable displacement of the equilibrium orbit. For further details, including the use of Robinson wigglers to shift the damping aperture, the reader may consult Ref. 10.

In small storage rings, the physical aperture is usually smaller than the damping aperture.

The damping aperture is easily measured by varying the RF frequency and watching for beam blow-up on a synchrotron light monitor.
5. QUANTUM FLUCTUATIONS AND FOKKER-PLANCK EQUATIONS

When certain conditions are satisfied, sets of stochastic ordinary differential equations can be replaced by a partial differential equation for a distribution function on phase space. In the limit of vanishing correlation time of the random terms, this partial differential equation takes the form of a Fokker-Planck equation. Its physical meaning and precise relation to the stochastic equations are discussed in Appendix D. Fokker-Planck equations have been applied to several problems in accelerator physics; for some examples, see Refs. 12 (several articles), 14, 17, 23 and the reference lists which they contain.

5.1 Quantum fluctuations in longitudinal phase space

We now reintroduce the fluctuating part of the radiation power into the longitudinal equations of motion. To prepare the ground for writing down the Fokker-Planck equation, we write them in the form

\[ z' = K_x(z, \varepsilon) + Q_x(z, \varepsilon) \xi(s), \quad \varepsilon' = K_\varepsilon(z, \varepsilon) + Q_\varepsilon(z, \varepsilon) \xi(s) \]  

(5.1)

where the \( K \)- and \( Q \)-functions are

\[ K_x(z, \varepsilon) = -\alpha_c \varepsilon, \quad K_\varepsilon(z, \varepsilon) = \frac{(\Omega_s/c)^2 \varepsilon - J_s \alpha_c \varepsilon}{c}, \]
\[ Q_x(z, \varepsilon) = 0, \quad Q_\varepsilon(z, \varepsilon) = -p_0 \sqrt{c_2 G(s)^3}. \]  

(5.2)

Strictly speaking, these are in something of a hybrid form since the smooth approximation has not yet been applied to the fluctuation terms. These, by their nature, must be approximated in a root-mean-square fashion, rather than directly. This is easier to understand in terms of the Fokker-Planck equation. Since

\[ \frac{\partial Q_x}{\partial \varepsilon} = \frac{\partial Q_\varepsilon}{\partial z} = 0 \]  

(5.3)

the recipe for writing it down (see Appendix A) simplifies by virtue of the lack of “spurious drift” terms and we find

\[ \frac{\partial F(z, \varepsilon, s)}{\partial s} = \alpha_c \varepsilon \frac{\partial F(z, \varepsilon, s)}{\partial \varepsilon} - \frac{(\Omega_s/c)^2 \varepsilon}{c} \frac{\partial F(z, \varepsilon, s)}{\partial \varepsilon} + \frac{J_s \alpha_c}{c} \frac{\partial}{\partial \varepsilon} [\varepsilon F(z, \varepsilon, s)] + \frac{c_2 p_0^2 J_s}{2(2\pi R)} \frac{\partial^2 F(z, \varepsilon, s)}{\partial \varepsilon^2}, \]  

(5.4)

where \( F(z, \varepsilon, s) \) is the distribution function in longitudinal phase space and we have made the smooth approximation of the diffusion term (i.e. the one with second derivatives) in terms of the synchrotron integral \( J_s \) defined in Appendix D.

This equation can be solved completely in terms of its Green function or by eigenfunction expansions but we can simplify it further by making a phase-mixing assumption

\[ \langle z \rangle = \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} d\varepsilon \, z F(z, \varepsilon, s) = 0 \]  

(5.5)

which will be true in many situations, including that of equilibrium. Then we can integrate (5.4) over \( z \)
to get an equation for the reduced distribution function

\[ F(\varepsilon, s) = \int_{-\infty}^{\infty} F(x, \varepsilon, s) \, dx, \]  

(5.6)

namely

\[ \frac{\partial F(\varepsilon, s)}{\partial s} = J_2 \alpha_2 \frac{\partial}{\partial \varepsilon} \left[ \varepsilon F(\varepsilon, s) \right] + \frac{c_2 p_0^2 I_3}{2(2\pi R)} \frac{\partial^3 F(\varepsilon, s)}{\partial \varepsilon^3}. \]

(5.7)

To find the equilibrium solution \( F_0(\varepsilon) \), we simply set \( \partial_s F = 0 \) and integrate once to find

\[ -\varepsilon F_0(\varepsilon) = \frac{c_2 p_0^2 I_3 f_0}{2J_2 \alpha_2} F'_0(\varepsilon). \]

(5.8)

When we integrate this again, choosing the constant of integration to normalise the distribution to unity, we find the familiar gaussian distribution of momentum (or energy deviations)

\[ F_0(\varepsilon) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{\varepsilon^2}{2\sigma^2} \right), \]

(5.9)

where \( \sigma_\varepsilon \) is the r.m.s. energy spread in the beam for a linear damping rate determined by the value of \( J_2 \):

\[ \langle \varepsilon^2 \rangle \equiv \sigma^2 = \frac{2\sigma^2}{J_2} = \frac{55 \hbar}{32\sqrt{3} mc} \left( \frac{p_0}{mc} \right)^2 \frac{I_3}{J_2 I_2}. \]

(5.10)

The quantity \( \sigma_\varepsilon \), which can be regarded as a measure of the strength of quantum excitation, is defined to be the energy spread for the reference case \( J_\varepsilon = 2^* \).

Since \( I_2/I_3 \approx \rho \), the bending radius, in an isomagnetic ring there is very little which can be done, beyond varying \( J_\varepsilon \), to reduce the energy spread of an electron storage ring. Moreover, the energy spread is directly proportional to \( E_0 \). A very small decrease of \( \sigma_\varepsilon \) can in principle be achieved with wiggler magnets but their usual effect is to increase it.

This is an important limitation since it determines the energy resolution of particle physics experiments which may be trying, for example, to detect, or measure the widths of, narrow resonances in the mass spectrum. In fact, since design considerations for colliding beam rings usually imply \( \rho \propto E_0^2 \), it almost always turns out that \( \sigma_\varepsilon \approx 0.1\% \) at the top energy of a given ring. Nevertheless \( e^+ e^- \) rings still provide a much finer energy resolution than any foreseeable linear collider or hadron collider and there remains the possibility of enhancing it still further with the so-called "monochromator" insertions.

The gaussian distribution of energy deviation is by no means inevitable—nonlinear terms (dissipative or conservative) may well change it, especially in the tails (as we shall see shortly). Arguments based on the "Central Limit Theorem" should only be applied in linear approximation and the analogy with the Maxwell-Boltzmann velocity distribution in a gas is not a complete one.

---

* Often no notational distinction is made between \( \sigma_\varepsilon \) and \( \sigma_\varepsilon \), so one should always be careful to understand which is meant.
5.2 Fokker-Planck equation in action-angle variables

We transform to action-angle variables of linearized synchrotron motion and make a rescaling to variables \((x, I)\) with

\[
z = \frac{1}{\kappa^2} \sqrt{2I} \cos(\kappa^2 x), \quad \varepsilon = \sqrt{2I} \sin(\kappa^2 x), \quad \kappa^2 \overset{\text{def}}{=} \frac{\Omega_0}{\alpha_c c}.
\]  

(5.11)

The constant \(\kappa^2\) can be thought of as a conversion factor between energy deviation and length units:

\[
\sqrt{\langle \varepsilon^2 \rangle} = \kappa \sqrt{\langle z^2 \rangle} = \frac{Q_s}{\alpha_c R} \sqrt{\langle z^2 \rangle}
\]

(5.12)

and \(Q_s = \Omega_s/2\pi f_0\) is the synchrotron tune. With these variables, the longitudinal Hamiltonian reduces to

\[
H_z = -\alpha_c I
\]

(5.13)

and, by applying the results of Appendix B (or otherwise), we can derive stochastic equations of motion equivalent to (5.4)

\[
\begin{align*}
    x' &= K_x(x, I) + Q_x(x, I) \xi(s), \quad I' = K_I(x, I) + Q_I(x, I) \xi(s)
\end{align*}
\]

(5.14)

where

\[
\begin{align*}
    K_x(x, I) &= -\alpha_c - \frac{J_0(\kappa^2 x)}{2c \kappa^2 c} \sin(2\kappa^2 x), \quad & K_I(x, I) &= -\frac{J_0(\kappa^2 x)}{c} \left[ 1 - \cos(2\kappa^2 x) \right],
    \\
    Q_x(x, I) &= -\frac{2\alpha_c}{\kappa^2 \sqrt{2c I}} \cos(\kappa^2 x), \quad & Q_I(x, I) &= -2\alpha_c \frac{\alpha_c}{2c I} \sqrt{\langle z^2 \rangle} \sin(\kappa^2 x).
\end{align*}
\]

(5.15)

Now \(Q_I\) and \(Q_x\) depend on \(I\) and \(x\) and their derivatives will contribute spurious drift terms and some algebraic complications to the Fokker-Planck equation

\[
\frac{\partial F(I, x, s)}{\partial s} = -\frac{\partial}{\partial I} [D_I F(I, x, s)] - \frac{\partial}{\partial x} [D_x F(I, x, s)]
\]

\[
+ \frac{1}{2} \frac{\partial^2}{\partial I^2} [Q_I^2 F(I, x, s)] + \frac{\partial^2}{\partial I \partial x} [Q_I Q_x F(I, x, s)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [Q_x^2 F(I, x, s)].
\]

(5.16)

Here, the complete drift terms are

\[
\begin{align*}
    D_I &= K_I + \frac{1}{2} \frac{\partial Q_I}{\partial I} Q_I + \frac{1}{2} \frac{\partial Q_I}{\partial x} Q_x = -\frac{\alpha_c}{c} (J_0 I - 2\alpha_c^2) + \frac{J_0 \alpha_c I}{c} \cos(2\kappa^2 x),
    \\
    D_x &= K_x + \frac{1}{2} \frac{\partial Q_x}{\partial I} Q_I + \frac{1}{2} \frac{\partial Q_x}{\partial x} Q_x = -\alpha_c - \frac{\alpha_c}{\kappa^2 c} \left( J_0 + \frac{\alpha_c^2}{2I} \right) \sin(2\kappa^2 x),
\end{align*}
\]

(5.17)

and the diffusion terms are

\[
\begin{align*}
    Q_I^2 &= \frac{4\alpha_c \alpha_c^2 I}{c} \left[ 1 - \cos(2\kappa^2 x) \right], \quad & Q_I Q_x &= \frac{2\alpha_c \alpha_c^2}{\kappa^2 c} \sin(2\kappa^2 x), \quad & Q_x^2 &= \frac{\alpha_c \alpha_c^2}{\kappa^2 c} \left[ 1 + \cos(2\kappa^2 x) \right].
\end{align*}
\]

(5.18)

In action-angle variables, it is of course easy to apply the averaging method * and write down an averaged Fokker-Planck equation for the action variable which will be valid on time-scales longer than that of a

* This is a step beyond the smooth approximation.
synchrotron oscillation, notably the damping time scale:

$$\tau_{\epsilon} \frac{\partial F}{\partial t} = \frac{\partial}{\partial I} \left( [-J_{\epsilon} I + 2\sigma_2^2 F] + c(\epsilon \sigma_2^2 \partial F / \partial \chi) + 2\sigma_2^2 \partial^2 F / \partial I^2 (IF) + \frac{\sigma_2^2}{2\kappa_2^2} \frac{\partial^2 F}{\partial \chi^2} \right).$$

(5.19)

The absence of a damping term for the phase \( \chi \) guarantees that the phase diffusion term superposed on the rapid oscillatory phase advance will lead to a uniform distribution in \( \chi \) on \([0, 2\pi]\). Moreover, we have not included the dependence of the synchrotron frequency on amplitude which produces a filamentation effect, further accelerating the phase-mixing.

We can solve (5.19) for the stationary distribution

$$F_0(I) = \frac{1}{\sigma_2^2} \exp \left( -\frac{I}{\sigma_2^2} \right)$$

(5.20)

from which we can evaluate the longitudinal emittance (in these units!)

$$\langle I \rangle = \int_0^\infty IF_0(I) dI = \sigma_2^2 = \left( \frac{\epsilon^2 + \kappa_2^2 \epsilon^2}{2} \right).$$

(5.21)

Fig. 4 Gaussian longitudinal phase space distribution and its projection

The distribution (5.20) is equivalent to a joint gaussian in \( z \) and \( \epsilon \):

$$F_0(z, \epsilon) = F_0(\sigma_2^2 z^2/2 + \epsilon^2/2) = \frac{1}{2\pi \sigma_2 \sigma_z^2} \exp \left( -\frac{\epsilon^2}{2\sigma_z^2} - \frac{z^2}{2\sigma_z^2} \right).$$

(5.22)

where the natural (or zero-current) bunch length is

$$\sigma_z = \sigma_\epsilon / \kappa_2.$$

(5.23)

This distribution is shown in Fig. 4 for some typical values of \( J_\epsilon \) and \( \sigma_\epsilon \); the meaning of the parameters \( b \) and \( R \) will be explained in a later section. In this, and similar plots to be shown later, we also show the
projection of the action distribution along one axis, given by the integral

\[
\lambda(z) = \int_{-\infty}^{\infty} F_0(z, \varepsilon) \, d\varepsilon = \int_{0}^{\infty} \frac{F_0(I)}{\sqrt{2I - \kappa^2 z^2}} \, dI. \tag{5.24}
\]

By virtue of the special properties of the gaussian, this is just the same as (5.9). In general it need not be, as we shall see later. Such a projection corresponds to the energy distribution or the longitudinal current density profile of the bunch. It is not the shadow of the phase space distribution. The second equality in (5.24) holds only for rotationally symmetric distributions.

6. NONLINEAR WIGGLERS

In small and medium-sized $e^+e^-$ storage rings, it is an excellent approximation to assume that the radiation damping and quantum excitation effects remain linear to large amplitudes. On the other hand, the new generation of large rings (such as TRISTAN, LEP or the HERA electron ring) begin to enter a regime where these effects can develop amplitude-dependences which may have to be included in a realistic calculation. Such nonlinear effects will tend to produce equilibrium distributions whose cores are fatter and whose tails decay more slowly than predicted by the linear theory of the previous sections. Such effects may generally be expected to be detrimental to beam stability and lifetime.

More optimistically, we might regard the existence of such effects as an opportunity to favourably influence the distribution function by means of intentional dissipative nonlinearities. In this section, we shall introduce the idea of a nonlinear wiggler which allows one some freedom to shape the energy distribution in an $e^+e^-$ ring. Such wigglers have been studied in the context of the LEP design as a means of reducing the severity of certain collective instabilities and, possibly, depolarizing effects.

Nonlinear wigglers are special combined-function magnets which modify the low-intensity particle distribution in longitudinal phase space. The original idea for a nonlinear wiggler in LEP was a combined function dipole-octupole magnet. Although this works, it produces a large additional energy loss through its contribution to the integral $I_2$. One, and only one, other multipole combination exists which produces the same nonlinear damping effect with negligible additional energy loss.

6.1 Quadrupole-sextupole wiggler

We shall consider the quadrupole-sextupole wiggler in which, as the name suggests, there are superposed quadrupole and sextupole fields. The vertical component of $\mathbf{B}$ in the $y = 0$ plane is

\[
B_\parallel(x) = \pm \frac{p_0 c}{e} (K_1 w x + K_2 w x^3/2), \tag{6.1}
\]

alternating in sign between adjacent blocks of the device, so that the integrated quadrupole and sextupole field components vanish. For simplicity, let us assume that $K_1$ and $K_2$ have constant magnitude in the blocks, whose total length is $L_w$. To correct this approximation for a real wiggler, one must first find

* Beyond the variation of damping partition numbers.
the correct gradient profiles, either numerically or by measurements on real magnets. Each term in the energy loss given below is of the form of a product of $L_w$ and powers of field-gradants and dispersion functions and simply has to be replaced by the corresponding synchrotron radiation integral.

### 6.2 Nonlinear damping

Particles with a momentum deviation $p_0 \varepsilon$ from the synchronous value $p_0 (1 + \delta_s)$ will pass through the wiggler with a horizontal displacement $\eta_w (\delta_s + \varepsilon)$ off the axis. The additional contribution to the total energy loss of such particles due to the wiggler is equal to

$$
U_w (\delta_s + \varepsilon) = c_1 p_0^2 c L_w \left[ K_1^2 \eta_w^2 \delta_s^2 + (2 K_1^2 \eta_w^2 + K_1 K_2 \eta_w^3) \delta_s^3 \right]
$$

\begin{align*}
\text{(i)} & \quad \left( 2 K_1^2 \eta_w^2 \delta_s + 3(2 K_1^2 \eta_w^2 + K_1 K_2 \eta_w^3) \delta_s^2 \right) \varepsilon \\
\text{(ii)} & \quad \left( K_1^2 \eta_w^2 + 3(2 K_1^2 \eta_w^2 + K_1 K_2 \eta_w^3) \delta_s \right) \varepsilon^2 \\
\text{(iii)} & \quad \left( 2 K_1^2 \eta_w^2 + K_1 K_2 \eta_w^3 \right) \varepsilon^3 + \cdots \\
\text{(iv)} & \quad \frac{3}{I_2} \int \frac{3}{L_w} \left( 2 K_1^2 \eta_w^2 + K_1 K_2 \eta_w^3 \right) ds
\end{align*}

(6.2)

Each term in this expression has a different physical significance and must be examined in its turn:

(i) These terms are independent of $\varepsilon$ and simply add to the total value of $U(\delta_s)$; they include a contribution to the integral $I_8$, reducing the damping aperture somewhat.

(ii) The coefficient of $\varepsilon^1$ is related to (i) and adds to the linear damping rate.

(iii) Being proportional to $\varepsilon^2$ these average out over the phase of the oscillation.

(iv) With these terms we find a qualitatively new effect; being proportional to $\varepsilon^3$, they provide a damping proportional to $I^2$. The most significant contribution comes from the quadrupole-sextupole cross term which only exists by virtue of the fact that the quadrupole and sextupole fields are spatially superposed. Building a wiggler with separate quadrupole and sextupole blocks will not produce the same effects, no matter how closely the blocks are spaced.

One can verify numerically that the nonlinear quantum excitation due to the wiggler is much less important than the nonlinear damping.

Taking (6.2) into account, the equation of motion for $I$ is given by (5.14) where $Q_I$ is given by (5.15) but now

$$
K_I (\chi, I) = - \frac{J_c (\delta_s) c \pi}{e} \left[ 1 - \cos (2 \kappa \chi) \right] + \left( \frac{b}{2} \right)^2 \frac{8}{3} \sin (\kappa \chi) \right) I^4,
$$

where the nonlinear damping coefficient for a quadrupole-sextupole wiggler is defined by

$$
b \overset{\text{def}}{=} \frac{3}{I_2} \left( 2 K_1^2 \eta_w^2 + K_1 K_2 \eta_w^3 \right) ds \approx \frac{3 L_w}{I_2} \left( 2 K_1^2 \eta_w^2 + K_1 K_2 \eta_w^3 \right)
$$

(6.4)

Since the average of $(\sin \theta)^4$ is $3/8$ the nonlinear term makes an important contribution after phase-averaging.
6.3 Negative $J_\varepsilon$ and birth of a limit cycle

From (6.3) we can see that, even when $J_\varepsilon$ is negative, so that small amplitude synchrotron oscillations are anti-damped, the nonlinear terms generated by the quadrupole-sextupole wiggler can restore positive damping at larger amplitudes. It is possible to choose $\delta_\varepsilon$ so that the central momentum of the beam lies outside the damping aperture on the side of negative $J_\varepsilon$.

Returning to cartesian coordinates in phase space, it is easy to show that the deterministic equation (4.9) should be replaced by a van der Pol equation\(^{14}\)

$$\ddot{z} + \alpha_z (J_\varepsilon + b\dot{z}^2 z^2) \dot{z} + \Omega_z^2 z = 0. \quad (6.5)$$

![Diagram](image)

Fig. 5 Approximate solutions of the Van der Pol equation

In Fig. 5, we show solutions of this equation which are obtained analytically (for $\alpha_z/\Omega_z \ll 1$) by an application of the averaging method. They are equivalent to integrating (6.3) for two qualitatively distinct cases. For clarity the damping time has been artificially shortened to a few times the synchrotron period. Two cases are plotted:

(a) Here $J_\varepsilon = 2$, $b = 0$ and the origin of phase space is a simple attracting point. All orbits within the separatrix of the RF bucket (not shown here) are attracted to it. If $b$ is given a positive value no qualitative change occurs but particles with large amplitudes are damped more rapidly.

(b) Now $J_\varepsilon = -1$ and $b = 5 \times 10^5$; small amplitudes are anti-damped but positive damping is restored at larger amplitudes. The linear anti-damping and nonlinear damping balance at the value

$$I = \dot{I} = -\frac{2J_\varepsilon}{b}, \quad (6.6)$$

corresponding to a limit cycle of (6.5), clearly visible as a periodic orbit which attracts particles from both larger and smaller amplitudes. The fixed point at the origin has become unstable via the so-called Hopf bifurcation.
6.4 Fokker-Planck equation with nonlinear wiggler

To see how this new phase space structure affects the distribution function, we construct the averaged Fokker-Planck equation in the action variable by generalizing (5.19) and integrating over phase

$$\tau_\varepsilon \frac{\partial F(I,t)}{\partial t} = -\frac{\partial}{\partial I} \left[ -\left( \frac{J_\varepsilon}{2} I^2 - 2\sigma_\varepsilon^2 \right) F(I,s) \right] + 2\sigma_\varepsilon^2 \frac{\partial^2}{\partial I^2} [IF(I,s)]. \quad (6.7)$$

6.5 Equilibrium solution

Integrating (6.7), we find a non-gaussian equilibrium distribution,

$$F_0(I) = Z(J_\varepsilon, b, \sigma_\varepsilon)^{-1} \exp \left( -\frac{J_\varepsilon}{2\sigma_\varepsilon^2} I^2 - \frac{b}{8\sigma_\varepsilon^2} I^2 \right). \quad (6.8)$$

where the normalisation constant $Z(J_\varepsilon, b, \sigma_\varepsilon)$ will be discussed in detail below.

For $b > 0$, the tails of this distribution decay very much faster than gaussian ones | as $\exp(-\varepsilon^4)$ rather than $\exp(-\varepsilon^2)$ | and this can considerably improve the lifetime and stability of the particle beam. Even when $J_\varepsilon$ is made negative, the balance between linear anti-damping at small amplitudes and nonlinear damping at larger amplitudes results in a stable distribution, i.e., one which can be normalised.

By adjusting the two free parameters $J_\varepsilon$ and $b$ we find that we have an additional degree of freedom in moulding the longitudinal profile of the bunch. In addition, our freedom to vary $J_\varepsilon$ is extended by the possibility of moving it onto the negative real axis.

It can be shown\(^\dagger\) that distributions with the same value of the dimensionless parameter

$$R = -\frac{J_\varepsilon}{\sqrt{2b\sigma_\varepsilon}} \quad (6.9)$$

are geometrically similar.
Each of the Figs. 6–11 is analogous to Fig. 4 but includes the influence of a nonlinear wiggler of a certain strength. To ease comparison, all three scales of Figs. 6–11 are the same as in Fig. 4.

In this sequence of figures, we can follow what happens as a single parameter, \( J_e \), is varied from positive to negative values. All other relevant parameters, namely the wiggler strength \( b \) and the quantum excitation \( \sigma_e \) are held constant.

In Figs. 6 and 7 the distribution is similar in form to the gaussian Fig. 4 when \( b = 0 \) except that the tails decay faster. Few particles lie in the region of phase space where the wiggler has much influence and therefore the r.m.s. energy spread is only very slightly reduced. However if off-momentum particles are responsible for unwanted effects (e.g. depolarization), such a distribution may be very beneficial.
When $J_\varepsilon$ is decreased to zero, Fig. 8 there is no quadratic term in the exponent of (6.8); the distribution remains stable but has spread out considerably since small amplitude particles are hardly damped at all.

When $J_\varepsilon$ goes negative (Figs. 9–11) the peak at the centre of the phase space distribution becomes a crater. The maximum density then occurs approximately above the attracting limit cycle of the deterministic equations of motion (4.7). For sufficiently small negative values of $J_\varepsilon$, (Fig. 9) the profile of the bunch in real space still contains only one hump because the crater is so shallow that it is wiped out in the integration across phase space.

This state of affairs persists until $J_\varepsilon$ has passed through a special value:  

$$J_\varepsilon < -0.5409\sqrt{2\sigma_\varepsilon}.$$  

(6.10)
\[ J_\varepsilon = -0.6 \]
\[ b = 5.65 \]
\[ \sigma_\varepsilon = 0.7 \times 10^{-3} \]
\[ R = 0.8571 \]
\[ \sqrt{\langle \varepsilon^2 \rangle} = 1.68 \times 10^{-3} \]

Fig. 11 Two peaks appear in current profile

Around this transition value (Fig. 10), the profile is very flat. Beyond it two humps appear in the current profile (Fig. 11).

Such a flattening-out of the current distribution can be of considerable utility in the attempts to increase the amount of stored current in a storage ring. The peak current in the bunch is lowered and the energy spread and bunch length are increased, tending to reduce the wake-fields.

At first sight, it may seem that a much larger RF voltage would be necessary to accommodate the larger energy spread with reasonable quantum lifetime. Generally however this is not a serious problem for two reasons:

(i) In the case of a storage ring which is also an accelerator, the bunch-lengthening effect would be needed mainly in the lower part of the energy range of the storage ring, notably at injection energy and, there, there ought to be RF voltage to spare.

(ii) For a given energy spread and RF voltage, the quantum lifetime for a distribution such as (6.8) is much longer than that of (5.20) because of the much faster decay of its tails. In other words one can fill up a much larger proportion of the RF bucket with particles without increasing the loss rate across the separatrix.

6.6 Partition function and moments

The distribution (6.8) is normalised to unity by means of the partition function

\[ Z(J_\varepsilon, b, \sigma_\varepsilon) = \int_0^\infty \exp \left( -\frac{J_\varepsilon I}{2\sigma_\varepsilon^2} - \frac{b}{8\sigma_\varepsilon^2} I^2 \right) \, dI = 2\sigma_\varepsilon \sqrt{\frac{\pi}{2b}} w \left( \frac{iJ_\varepsilon}{\sqrt{2b\sigma_\varepsilon}} \right), \tag{6.11} \]

where \( w \) is the error function for complex arguments (sometimes known as the plasma dispersion function). Considered as a function of the parameters \( J_\varepsilon \), \( b \) and \( \sigma_\varepsilon \), the partition function contains a lot of information
about the global properties of the distribution and is a convenient tool for calculation. In this respect, it is analogous to the partition functions of equilibrium statistical mechanics.

As an aid to physical understanding, it is particularly useful to make two distinct asymptotic expansions of the reciprocal:

\[
Z(J, b, \sigma) = \frac{J}{4\pi \sigma^2} + \frac{b^2}{4\pi J} + \frac{b^2 \sigma^2}{2\pi J^2} + \ldots \quad (b \to 0^+, \quad J > 0),
\]

(6.12)

which is useful as we consider the transition from positive linear damping to nonlinear damping (small values of \(b\)), and

\[
Z(J, b, \sigma) = \sqrt{\frac{b}{(2\pi)^3 \sigma^2}} + \frac{J}{2\pi^2 \sigma^2} + \frac{1}{\pi} - \frac{1}{4} \frac{J^2}{\pi \sqrt{2\pi b \sigma^2}} + \ldots \quad (J \to 0, \quad b > 0),
\]

(6.13)

which describes the neighbourhood of the bifurcation as \(J\) changes sign. For comparison with (5.21), the longitudinal emittance is given by

\[
\langle I \rangle = \int_{0}^{\infty} I F_0(I) \, dI = (-2\sigma_z^2) \frac{\partial \log Z}{\partial J} = \frac{8\pi \sigma^2}{b Z(J, b, \sigma)} - \frac{2Z(J, b, \sigma)}{b}.
\]

(6.14)

Evaluating this in the limit of small \(b\), with the help of (6.12), shows that it does indeed reduce to (5.21) when the nonlinear wiggler is turned off

\[
\langle I \rangle \sim \frac{2\sigma_z^2}{J} \left(1 - \frac{2b \sigma_z^2}{J}\right), \quad (b \to 0^+, \quad J > 0).
\]

(6.15)

More generally, all the moments of the equilibrium distribution can be found from

\[
\langle I^n \rangle = \int_{0}^{\infty} I^n F_0(I) \, dI = (-2\sigma_z^2)^n Z^{-1} \frac{\partial^n Z}{\partial J^n}.
\]

(6.16)

6.7 Quantum lifetime

The longitudinal quantum lifetime of the beam is the inverse of the loss rate of particles across the separatrix due to quantum fluctuations. Other loss mechanisms\textsuperscript{11} may also contribute to determine the net lifetime of the beam.

To calculate the quantum lifetime, one interprets the Fokker-Planck equation as a continuity equation in phase space and identifies the diffusive component of the particle flux across the separatrix. This component constitutes the loss rate. At the separatrix, it is not balanced by a flux of particles damping down from larger amplitudes. The details of such a calculation for the gaussian distribution were given
in Ref. 11 and can be generalized for (6.8); the details of this part may be found in Ref. 14. The result is

$$r_{\epsilon} = \frac{\tau_{\epsilon} Z(J_{\epsilon}, b, \sigma_{\epsilon}) \exp \left( \frac{\xi^{2} b \tau_{\epsilon}^{2}}{2} + \frac{\xi^{2} b^{2}}{2} \right)}{\pi \xi \sigma_{\epsilon}^{2}(\xi \sigma_{\epsilon}^{2} b + 2 J_{\epsilon})}$$  (6.17)

where the parameter $\xi$ is half the squared bucket half-height in units of $\sigma_{\epsilon}$:

$$\xi = \frac{r_{\epsilon}^{2}}{2 \sigma_{\epsilon}^{2}}.$$  (6.18)

Except for the fact that we have taken $J_{\epsilon} = 2$ as a reference case, this is precisely the same definition as introduced in Ref. 2. The energy aperture $\epsilon_{\text{max}}$ is usually thought of as arising from the RF voltage limitation but it may also arise from a limitation of the vacuum chamber aperture in a dispersive region or even a reduced dynamic aperture at large momentum deviation.

The formula (6.17) includes the result for a gaussian distribution$^{2,11}$ as a special case.

### 6.8 Practical aspects

A study of the available gradients and apertures of combined-function quadrupole-sextupole magnets would be out of place here. We only mention that, to make the nonlinear damping effect noticeable, we need $2 b \sigma_{\epsilon}^{2} \sim 1$. From (6.4), we see that, for given gradients, $b \propto p_{0}^{-2}$ while from (6.10), we have $\sigma_{\epsilon}^{2} \propto p_{0}^{3}$, so that $b \sigma_{\epsilon}^{2}$ is independent of energy for a given set of wiggles and storage ring. In addition the dependence on the bending radius cancels out from the product. Because the dispersion function is of the order of 1 m in almost all lattice designs, it follows that the length of wiggler required is roughly independent of the size of the storage ring and the energy at which it is operated. Since several tens of metres of wiggler are required, we can contemplate installing them only in the largest rings.

Taking the example of LEP, we find that if a set of quadrupole-sextupole wiggles were installed with effective parameters

$$\frac{\partial B_{y}}{\partial x} = 10 \text{ tesla m}^{-1}, \quad \frac{\partial^{2} B_{y}}{\partial x^{2}} = 400 \text{ tesla m}^{-2}, \quad L_{w} = 40 \text{ m}, \quad \eta_{w} = 1.8 \text{ m}. \quad (6.19)$$

then, at a beam energy of 20 GeV, where $\sigma_{\epsilon} = 0.3 \times 10^{-3}$ we would have $b \simeq 5 \times 10^{5}$, the value we used in Figs. 6–11. The energy spread used in the figures would correspond to a beam energy of around 45 GeV in LEP.

The only serious adverse effect of a quadrupole-sextupole wiggler seems to be the reduction in damping aperture due to its contribution to the integral $I_{\epsilon}$. This requires a slightly more elaborate control of the RF frequency to achieve the desired value of $J_{\epsilon}$ since the variation of $J_{\epsilon}$ will be coupled with the excitation of the wiggler.
ACKNOWLEDGEMENTS

I acknowledge the influence of colleagues too numerous to mention in CERN, other laboratories and several universities; some, by no means all, of their names are among the references.

The task of writing up a paper containing so many formulae and graphics has been wonderfully eased by Donald E. Knuth's program \TeX; I am grateful to SLAC for access to it during a visit.

* * *

REFERENCES


3. J.S. Bell, Hamiltonian mechanics, these proceedings.


11. A. Piwinski, Beam losses and lifetime in Reference 1.


27. A. Renieri, *Possibility of achieving very high energy resolution in electron-positron storage rings*, Frascati Preprint /INF-75/6(R), (1975).
31. C.N. Lashmore-Davies, *Kinetic theory and the Vlasov equation, these proceedings*.
APPENDIX A: Physical meaning of the Fokker-Planck equation

After some discussion of its relationship to Liouville's Theorem, we state a recipe for writing down the Fokker-Planck equation corresponding to a given set of stochastic differential equations. A variety of derivations of this relationship may be found in the literature on stochastic processes\textsuperscript{7,8} and one tailored to the present problems and notations has been given previously.\textsuperscript{13}

Consider some vector of coordinates $X(t) = (X_1, \ldots, X_N)$ (e.g. in a near-Hamiltonian system, the canonical coordinates and momenta $(q_1, \ldots, q_n, p_1, \ldots, p_n)$, where $N = 2n$). It evolves in time according to a set of first-order stochastic differential equations with a single\textsuperscript{*} noise source $\xi(t)$ satisfying

$$\langle \xi(t) \rangle = 0, \quad \langle \xi(t) \xi(t') \rangle = \delta(t - t').$$ \hfill (A1)

The stochastic differential equations are taken to be

$$\dot{X}(t) = K(X, t) + Q(X, t)\xi(t).$$ \hfill (A2)

Note that putting (2.35) in this form requires $K$ to include the terms in $\langle F_X(s) \rangle$.

When we speak of a realisation of $\xi(t)$ we mean just one of the ensemble of possible histories of the stochastic process. Of course this ensemble would contain much more information than we could possibly cope with so we bear in mind that we must afterwards average over all realisations with an appropriate weighting.

To each realisation of $\xi(t)$ there corresponds a realisation of the solution $X(t)$. Let us introduce the exact phase-space density for this solution

$$F(X, t) = \delta(X - X(t)) = \prod_i \delta(X_i - X_i(t))$$ \hfill (A3)

where $X$ is a free variable and $X(t)$ is the realisation of the solution of (A2) starting from initial conditions $X(0)$. The continuity equation for $F$ is

$$\partial_t F(X, t) + \nabla_X \cdot [\dot{X}(t) F(X, t)] = 0.$$ \hfill (A4)

or, less ambiguously,

$$\partial_t F(X, t) + \nabla_X \cdot [K(X, t) F(X, t)] + \nabla_X \cdot [Q(X, t) \xi(t) F(X, t)] = 0.$$ \hfill (A5)

and is completely equivalent to (A2). Since $X(t)$ depends on the values of the fluctuations $\xi(t)$, $F(X, t)$ is also a fluctuating quantity:

$$F(X, t) = \langle F(X, t) \rangle + \hat{F}(X, t).$$ \hfill (A6)

and it would be just as difficult to deal with—let alone find—the ensemble of its solutions as it would be to deal with that of (A2).

\textsuperscript{*} The generalisation to several noise sources is easy; we only need one in our application since the changes in the three canonical momenta of the electron due to a particular photon emission are correlated.
In the particular case of a Hamiltonian system, where

\[ K = \left( \frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_n}, -\frac{\partial H}{\partial q_1}, \ldots, -\frac{\partial H}{\partial q_n} \right), \quad Q = 0, \]  
(A7)

the gradient operator \( \nabla_x = (\partial/\partial q, \partial/\partial p) \) in (A5) may be commuted to the right, past the \( \dot{X}(t) \), giving Liouville's theorem.

To derive the Fokker-Planck equation describing the time evolution of the average phase-space density of a system, subject to deterministic drift \( K \) and diffusion \( Q \), one must average (A5) but take account of the way in which the fluctuations make the sharply-peaked phase space density (A3) spread out to become, upon averaging, a smooth function.\(^{13}\) In this process, information about the fine details of the distribution function is lost and we make the transition to an irreversible description of the time-evolution. Reinterpreting \( F(X, t) \) to mean this average or smoothed-out function, the final result is

\[ \partial_t F = -\sum_i \frac{\partial}{\partial X_i} \left\{ \left[ K_i + \frac{1}{2} \sum_j \frac{\partial Q_i}{\partial X_j} Q_j \right] F \right\} + \frac{1}{2} \sum_i \sum_j \frac{\partial}{\partial X_i} \frac{\partial}{\partial X_j} [Q_i Q_j F]. \]  
(A8)

The second term inside the curly brackets is known as the “spurious drift” and is absent when the diffusion \( Q(X, t) = Q(t) \) independently of amplitude. Because it often vanishes, this term is sometimes overlooked although it is essential for a complete description—see e.g. its rôle in section 5.

It is possible to avoid this apparent complication by writing stochastic differential equations which incorporate the spurious drift terms into the definition of \( K \). One must then work with the so-called \( \text{Itô calculus} \), in which the ordinary rules of differentiation are replaced by rules which bring in extra terms designed precisely to maintain the simple relationship between the differential equations and the Fokker-Planck equation. A change of variables is then rather more work than it is when one works in the Stratonovich interpretation as we have done in this lecture. What you gain on the swings of Fokker-Planck equations, you lose on the roundabouts of variable transformations. We spent more time on the roundabouts so the Stratonovich interpretation was the better choice. In this scheme, Markovian stochastic differential equations are written down as limits of equations governing real processes, as was implicit in the formulation of the electron equations of motion (2.35). \( \text{Itô equations of motion can be written down to describe the same physical system but, since the Fokker-Planck equation has to be the same, they look different from (2.35).} \)

For a system, like the electron in a storage ring, which is close to being Hamiltonian, (A2) would take the form

\[
\begin{align*}
\dot{q} &= \frac{\partial H(q, p, t)}{\partial p} + \epsilon K_d(q, p, t) + \sqrt{\epsilon} Q_d(q, p, t) \xi(t), \\
\dot{p} &= -\frac{\partial H(q, p, t)}{\partial q} + \epsilon K_p(q, p, t) + \sqrt{\epsilon} Q_p(q, p, t) \xi(t).
\end{align*}
\]  
(A9)

Here the ordering of terms with respect to the small parameter \( \epsilon \) has been chosen so that the dissipative terms have an effect proportional to \( \epsilon \Delta t \) in a short time interval \( \Delta t \). That this is so is clearer when,
following (A8), we write down the Fokker-Planck equation corresponding to (A9)

\[
\partial_t F = -\frac{\partial H}{\partial p_i} \frac{\partial F}{\partial q_i} + \frac{\partial H}{\partial q_i} \frac{\partial F}{\partial p_i} \\
- \epsilon \frac{\partial}{\partial q_i} \left\{ \left[ K_{ii} + \frac{1}{2} \frac{\partial Q_{ii}^\mu}{\partial q_i} Q_{ij} + \frac{1}{2} \frac{\partial Q_{ij}^\mu}{\partial p_j} Q_{ji} \right] F \right\} \\
- \epsilon \frac{\partial}{\partial p_i} \left\{ \left[ K_{pi} + \frac{1}{2} \frac{\partial Q_{pi}^\mu}{\partial q_i} Q_{ij} + \frac{1}{2} \frac{\partial Q_{ij}^\mu}{\partial p_j} Q_{pi} \right] F \right\} \\
+ \frac{\epsilon}{2} \frac{\partial}{\partial q_i} \frac{\partial}{\partial q_j} [Q_{ij} Q_{kj} F] + \frac{\epsilon}{2} \frac{\partial}{\partial q_i} \frac{\partial}{\partial p_j} [Q_{ij} Q_{pj} F] + \frac{\epsilon}{2} \frac{\partial}{\partial q_i} \frac{\partial}{\partial q_j} [Q_{ij} Q_{ij} F]
\]  
(A10)

where summation over repeated indices is to be understood. The first two terms represent the incompressible flow in phase space described by Liouville's Theorem; the second pair represents an at least partially irreversible drift (e.g. damping); finally the last three terms describe diffusion.

Since this School has included a course on plasma kinetic theory\(^{31}\) it may be helpful to briefly discuss the relationship of the Fokker-Planck equation described there to that given here.

Mathematically speaking, these equations are not of the same form, but they are nevertheless related. Physically, of course, they describe quite different phenomena—this is one resemblance between high energy electron bunches and plasmas which turns out to be rather superficial.

Here, the most notable difference is that (A8) is always linear, while in plasma physics, the name is commonly applied to an equation (sometimes also called the Landau equation) in which the distribution function appears quadratically in order to describe collisions between particles of the same, or different, species. Thus, while we have been exclusively concerned with single particles subject, in effect, to random external forces, the collision terms of the "plasma" Fokker-Planck equation describe forces acting between pairs of particles. This richer physical content makes it considerably more difficult to solve. Yet, when the distribution is linearised about the equilibrium Maxwell-Boltzmann form, an equation of the form (A8) is recovered. This, of course, is not surprising since, in that case, we may think of the equation as describing the distribution function of a single test particle subjected to random forces arising from the thermal background plasma. In this limit, the Rosenbluth potentials\(^{31}\) become independent of the perturbation of the distribution function and are related to the functions \(K\) and \(Q\) used here.
APPENDIX B: Canonical transformations for dissipative systems

The title of this appendix verges on the self-contradictory. Of course, we do not suggest that dissipative systems can be made canonical, only that, when their equations of motion contain a Hamiltonian part, upon which it is convenient to make canonical transformations, we may extend the formalism of generating functions to transform the dissipative parts too. This is of obvious utility in electron storage ring theory because the radiation terms in the equations of motion are small on average compared with those describing the applied electromagnetic fields.

Let us consider the case of a free \((i.e.\) such that old and new coordinates are independent) transformation from \((q, p) = (q_1, \ldots, q_n, p_1, \ldots, p_n)\) to new variables \((Q, P) = (Q_1, \ldots, Q_n, P_1, \ldots, P_n)\).

Call the old Hamiltonian \(H(q, p, s)\) and the new \(K(Q, P, s)\). Both in the old and new variables, Hamilton’s equations have to be supplemented by terms \((a, b)\) and \((A, B)\) which describe the dissipation:

\[
\begin{align*}
q' &= \frac{\partial H}{\partial p} + a(q, p, s) \\
p' &= -\frac{\partial H}{\partial q} + b(q, p, s)
\end{align*} \quad \iff \quad \begin{align*}
Q' &= \frac{\partial K}{\partial P} + A(Q, P, s) \\
P' &= -\frac{\partial K}{\partial Q} + B(Q, P, s).
\end{align*}
\]

Since the transformation is canonical, there exists a generating function \(S(Q, q, s)\), depending on the old and new coordinates and no momentum, such that the relationship between old and new coordinates and momenta is obtained by solving

\[
p = \frac{\partial S}{\partial q}, \quad P = -\frac{\partial S}{\partial Q}, \quad K = H + \frac{\partial S}{\partial s}.
\]

In a system with 3 degrees of freedom, 6 different types of generating function are necessary to generate all possible canonical transformations, and in practice each of them comes up sooner or later. The other cases can be worked by analogy to this one; for a different example, see Ref. 13.

The equivalence of the two descriptions of the Hamiltonian system is guaranteed by these relationships. Given that the two sets of equations in (B1) are supposed to describe the same dissipative system, we need to know how to calculate the new dissipative terms \((A, B)\).

[As in Appendix A, we must now acknowledge that any stochastic terms are to be interpreted in the Stratonovich, not the Itô, sense. The following results would be much more complicated in the Itô calculus.]

To avoid cumbersome notations, let us denote partial derivatives by subscripts where convenient; thus, for example, we may write a vector or a matrix

\[
S_q = \begin{pmatrix} \frac{\partial S}{\partial q_1} \\
\frac{\partial S}{\partial q_2} \\
\vdots \\
\frac{\partial S}{\partial q_n} \end{pmatrix}, \quad S_{QQ} = \begin{pmatrix} \frac{\partial^2 S}{\partial q_1 \partial q_1} & \frac{\partial^2 S}{\partial q_1 \partial q_2} & \cdots & \frac{\partial^2 S}{\partial q_1 \partial q_n} \\
\frac{\partial^2 S}{\partial q_2 \partial q_1} & \frac{\partial^2 S}{\partial q_2 \partial q_2} & \cdots & \frac{\partial^2 S}{\partial q_2 \partial q_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 S}{\partial q_n \partial q_1} & \frac{\partial^2 S}{\partial q_n \partial q_2} & \cdots & \frac{\partial^2 S}{\partial q_n \partial q_n} \end{pmatrix}.
\]

\footnote{This type of generating function, which is useful in the transformation to action-angle variables, is sometimes denoted \(F_1(Q, q, s)\).}
By hypothesis, the arguments of the generating function provide a unique labelling of points in extended phase space.\(^*\) Hence, their differentials form a basis for the vector space of differentials of dynamical variables and any such differential can be expressed as a linear combination of them. In particular, from (B2), we may use the ordinary chain rule to calculate

\[ dp = S_{QQ} dq + S_{Qq} dQ + S_{qQ} ds. \] (B4)

When the postulated equations of motion (B1), are used to project this identity along the local direction of time-evolution of the system, we find

\[
\begin{align*}
   dp &= S_{QQ}(H_p + a) ds + S_{QQ}(K_p + A) dQ + S_{qQ} ds \\
   &= -H_q ds + b ds
\end{align*}
\]

\[
\begin{aligned}
   \{ S_{QQ} H_p + S_{QQ} K_p + S_{qQ} &= -H_q \\
   S_{qq} a + S_{Qq} A &= b. \}
\end{aligned}
\] (B5)

where we separately identified the Hamiltonian and dissipative parts of the equality.

In an analogous way, we can calculate \( dp \) and conclude

\[
\begin{align*}
   S_{QQ} H_p + S_{QQ} K_p + S_{qQ} &= K_p, \\
   -S_{qq} a - S_{Qq} A &= B.
\end{align*}
\] (B6)

Since the canonical transformation is free, the Jacobian determinant of either set of canonical variables with respect to the independent coordinates \((Q, q)\) cannot vanish, \( i.e. \)

\[
\frac{\partial (p, q)}{\partial (Q, q)} = \left| \begin{array}{cc} \frac{\partial p}{\partial Q} & \frac{\partial p}{\partial q} \\ \frac{\partial q}{\partial Q} & \frac{\partial q}{\partial q} \end{array} \right| = \det \frac{\partial p}{\partial Q} = \det S_{QQ} \neq 0. \] (B7)

Hence the matrix \( S_{QQ} \) is invertible and it follows that the last members of (B5) and (B6) may be solved simultaneously to yield a linear relationship between the dissipative terms in the old and new representations:

\[
\begin{align*}
   A &= [S_{QQ}]^{-1}(b - S_{QQ} a) \\
   B &= -S_{qq} a - S_{QQ}[S_{QQ}]^{-1}(b - S_{QQ} a). \quad \text{(B8)}
\end{align*}
\]

By means of these formulae, the dissipative terms can be transformed in parallel with the canonical transformation; (B8) may be regarded as a convenient recipe for carrying out complicated transformations. It should be remembered that, in the course of this work, all derivatives of the generating function must be expressed in terms of the variables \((Q, q)\) and the expressions for \( q(Q, P) \) and \( p(Q, P) \) should only be substituted afterwards. Of course, exactly the same constraint applies while transforming the Hamiltonian.

\* When these terms have fluctuating parts this property is not at all trivial. It applies only as a consequence of our use of the Stratonovich interpretation.
APPENDIX C: Local synchrotron motion and smooth approximation

A single synchrotron oscillation takes many turns of the machine; in fact the largest synchrotron tunes are realised in large machines like LEP and are of the order of 0.1. It is therefore natural to simplify the description of this motion by making a Fourier decomposition of the Hamiltonian on the circumference and looking only at the most slowly-varying terms.

Let us consider the terms in (3.21) individually.

C1 Momentum compaction

From (3.21) the "kinetic energy" of synchrotron oscillations is given by the term

$$\frac{1}{2} \Gamma(\delta_s, s) e^2 = \frac{1}{2} \left\{ \frac{\varepsilon^2}{1 + \delta_s} - (G^2 + k) \eta^2 \right\} e^2 = \frac{1 + \delta_s}{2} \left\{ (\eta^*)' \right\} - G\eta \right\} e^2,$$

where we used (3.9). We make a Fourier analysis of the function

$$\Gamma(\delta_s, s) = \sum_{n=-\infty}^{\infty} \Gamma_n(\delta_s) e^{in\pi/R},$$

where the coefficients will be given by

$$\Gamma_n(\delta_s) = \frac{1}{2\pi R} \int_0^{2\pi R} ds \Gamma(\delta_s, s) e^{-in\pi/R}.$$

Integrating by parts twice and using periodicity arguments, it is not difficult to show that

$$\Gamma_n(\delta_s) = \Gamma_{-n}(\delta_s) = -\frac{1 + \delta_s}{2\pi R} \int_0^{2\pi R} ds \eta(\delta_s, s) \left[ \frac{n^2 \eta(\delta_s, s)}{2 R^2} + G(s) \right] e^{in\pi/R} ds.$$

In particular, the constant term is just

$$\Gamma_0(\delta_s) = \frac{1 + \delta_s}{2\pi R} \int_0^{2\pi R} ds \left\{ \frac{\varepsilon^2}{1 + \delta_s} - (G^2 + K_1) \eta^2 \right\} = -\frac{1 + \delta_s}{2\pi R} \int_0^{2\pi R} ds G\eta = -\alpha_c(\delta_s),$$

where $\alpha_c(\delta_s)$ is called the momentum compaction factor which, like the dispersion functions may be expanded in powers of $\delta_s$. Generally, $\alpha_c \ll 1$.

Neglect of the higher harmonics of $\Gamma(\delta_s, s)$ (and the RF voltage terms) gives the smooth approximation. In addition, simple scaling arguments show that the first term in the integral (C4) is much smaller than the second.

To simplify the analysis of the RF voltage distribution we define a function

$$\Sigma(\delta_s, s) = -\int_0^s \Gamma(\delta_s, \sigma) d\sigma,$$

which gives the increase in path length per unit momentum deviation $\delta_s$ in the sector of the ring between azimuths 0 and $s$. 

- 607 -
C2 Effective RF voltage

A similar Fourier analysis of the RF voltage term in (3.21), may be effected by substituting the identity

\[ \delta_C(s) = \frac{1}{2\pi R} \sum_{n=-\infty}^{\infty} e^{ins/R} \]  

and expanding the cosine into complex exponentials. The result is

\[ -\frac{1}{2\pi R} \sum_k \frac{e^{i\hat{\nu}_k}}{P_{\omega R}} \sum_{n=-\infty}^{\infty} \left\{ \exp \left[ \frac{\sin(s - s_k)}{R} + \frac{h(z + s_k)}{K(\delta_s)} - \frac{h}{K(\delta_s)} [s + \Sigma(\delta_s, s)\delta_s] + \phi_k \right] \right. \]

\[ \left. \exp \left[ \frac{\sin(s - s_k)}{R} - \frac{h(z + s_k)}{K(\delta_s)} + \frac{h}{K(\delta_s)} [s + \Sigma(\delta_s, s)\delta_s] - \phi_k \right] \right\}. \]  

(Bearing in mind the rôle of the RF harmonic number, we recognize that nearly all the terms in the expansion are rapidly oscillating functions of the independent variable \( s \) and will not produce significant average effects on the beam. The terms which do count are those in which the \( s \)-dependence can be made to cancel from the arguments of the exponentials. From (3.23), it follows that there are precisely two of these, namely the term with \( n = +h \) in the first group and that with \( n = -h \) in the second. Combining them, we can reassemble a slowly-varying cosine function

\[ -\frac{1}{2\pi R} \sum_k \frac{e^{i\hat{\nu}_k}}{P_{\omega R}} \cos \left\{ h \left( \frac{s}{R} - \frac{s}{K(\delta_s)} - \frac{\Sigma(\delta_s, s)\delta_s}{K(\delta_s)} \right) - \frac{hs_k}{R} + \frac{h(z + s_k)}{K(\delta_s)} + \phi_k \right\}. \]  

We separate the function \( \Sigma \) into a contribution from the \( n = 0 \) term in (C2) and a remainder:

\[ \Sigma(\delta_s, s) = \alpha_c(\delta_s)s + \tilde{\Sigma}(\delta_s, s) \]  

where, by virtue of (3.26), (C1) and (C2),

\[ \tilde{\Sigma}(\delta_s, s) = \sum_{n=1}^{2\pi R} \frac{1}{n\pi R} \int_0^{2\pi R} \Gamma(\delta_s, \sigma) \sin \left[ \frac{n(s - \sigma)}{R} \right]. \]  

Then (treating the two terms in (C10) differently with respect to the original \( \delta \)-functions) the \( s \)-dependence cancels and the argument of the cosine becomes

\[ -\frac{hs_k}{R} + \frac{h(z + s_k)}{K(\delta_s)} - \frac{h\delta_s}{K(\delta_s)} \tilde{\Sigma}(\delta_s, s_k) + \phi_k. \]  

It is clear that for the most efficient use of the RF system* one should choose the relative phases of the}

* We are ignoring all collective effects here.
APPENDIX D: Common synchrotron radiation integrals

In the following table we collect the definitions of some of the more important synchrotron radiation integrals together with indications of the contexts in which they occur.

<table>
<thead>
<tr>
<th>Definition</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1 = \int_0^{2\pi R} G\eta , ds$</td>
<td>momentum compaction factor</td>
</tr>
<tr>
<td>$I_2 = \int_0^{2\pi R} G^2 , ds$</td>
<td>energy loss, energy spread, damping times, emittances, damping partition numbers</td>
</tr>
<tr>
<td>$I_3 = \int_0^{2\pi R}</td>
<td>G</td>
</tr>
<tr>
<td>$I_{3a} = \int_0^{2\pi R} G^3 , ds$</td>
<td>polarization level</td>
</tr>
<tr>
<td>$I_4 = \int_0^{2\pi R} (G^2 + 2K_1)G\eta , ds$</td>
<td>energy spread, emittances, damping partition numbers</td>
</tr>
<tr>
<td>$I_5 = \int_0^{2\pi R}</td>
<td>G</td>
</tr>
<tr>
<td>$I_{6a} = \int_0^{2\pi R} K_1^2 \beta_z , ds$</td>
<td>energy loss in quadrupoles, nonlinear radiation damping</td>
</tr>
<tr>
<td>$I_6 = \int_0^{2\pi R} K_1^2 \eta^2 , ds$</td>
<td>damping partition number variation</td>
</tr>
</tbody>
</table>

The function $\mathcal{H}$ is defined\textsuperscript{2,11} by

$$\mathcal{H}(s) = (\eta(s)^2 + (\beta(s)\gamma(s) + \alpha(s)\eta(s))^2)/\beta(s).$$  \hfill (D1)

Integrals 1–5 were defined in Ref. 32 which also describes useful algorithms for evaluating them. These are implemented in several computer programs (e.g. BEAMPARAM, COMFORT, PATRICIA ...). Further information on the use of these integrals will be found in Refs. 2, 10, 11, 13, 17, 25, 26 and, especially, 32.
BEAM BREAK UP
J. Le Duff
Laboratoire de l'Accélérateur Linéaire, Orsay, France

ABSTRACT
After a brief review of the experimental evidence for the beam break up instability in electron linacs, emphasis is given to the transverse deflection which arises from radio frequency fields. Typical transverse RF modes in circular iris loaded waveguides are then described. Finally two types of beam break up are discussed: the regenerative BBU which occurs in a single accelerating section and the cumulative BBU which is a multi-section effect.

1. EXPERIMENTAL EVIDENCE

The beam break up (BBU), also known as beam blow up, is a transverse instability observed in RF electron linacs and in induction linacs. As early as 1957 the phenomenon was observed on short linacs operating with long pulses (> 1 µs) in the range of 500 mA peak. Above the current threshold the beam pulse length, as observed at the output of the accelerator, is shortened\(^1\). This suggests induced fields by the head of the pulse which act back on the tail to make it unstable. This mechanism which can occur in a single accelerating structure was called regenerative beam break up.

Later on, with longer linacs made of many successive accelerating structures, the same pulse shortening effect could be observed but at a much lower threshold, of the order of 10 mA peak, suggesting a cumulative effect from all the structures. This second manifestation of the beam break up has been intensively studied on the Stanford Linear Accelerator, two-miles long, since it appeared as the main limiting effect\(^2\). The experimental observations can be summarized as follows:

a) At any location along the accelerator the typical pictures of the beam pulses below and above threshold are illustrated in Fig. 1. The shortening is more pronounced as the current from the injector is increased.

![Oscillograms of beam pulses below and above beam break up threshold](image)

Fig. 1 Oscillograms of beam pulses below and above beam break up threshold
b) Above the threshold the amount of transmitted charge along the accelerator decreases erratically, and if the current is further increased the losses will appear earlier along the accelerator path (Fig. 2).

![Transmitted charge along the accelerator for different injected currents](image)

**Fig. 2** Transmitted charge along the accelerator for different injected currents

c) Above the threshold the beam cross section as observed at the end of the accelerator is randomly increased in both transverse directions, suggesting a transverse instability.
d) A suitable external magnetic focusing system, as provided for instance by quadrupole magnets, can improve the BBU threshold.
e) The BBU effect strongly depends on misalignment of the accelerating structures (off-axis beam) and on the noise level from the RF power sources.

The beam break up, as described above, either regenerative or cumulative, has been identified as a beam interaction with parasitic deflecting modes which can propagate in the accelerating structures. Such single modes can have relatively high shunt impedances and high Q values; hence the beam-induced wake field has a long memory which can affect the tail of long beam pulses. In these cases the equivalent impedance of the surroundings, which causes deflection, is a narrow-band type resonator.

In more recent times, linacs are being operated with very short, high peak current, pulses. This is for instance the case at SLAC, where the SLC project (SLAC Linear Collider) requires acceleration of single, high density, RF bunches. In that case a new type of beam break up has been identified in which the head of the bunch also influences the tail\(^3\). However since the pulse is very short the mechanism needs a much faster induced wake field. Such a fast wake field can be generated by the equivalent low-Q, wide-band transverse impedance of the accelerating structures, corresponding to an average effect of all the narrow-band parasitic modes up to quite high frequencies. Since the frequency spectrum of very short bunches is wide, the excitation of this impedance is made possible. The effect is very similar to the head-tail effect in storage rings, although here there is no synchrotron oscillation to enhance the amplification mechanism. In a linear accelerator the head of the bunch does not allow deflection, only the tail is affected, leading to a banana shape as shown in Fig. 3 a. However this corresponds to the most simple oscillating mode (dipole mode) for the tail motion under deflecting
forces. Figure 3 b for instance shows a quadrupole oscillating mode for which the center of gravity of the tail does not move. In both cases the effective transverse emittance is increased.

According to the similarity of the banana effect with the head tail, it will not be treated further in this lecture.

2. TRANSVERSE DEFLCTION OF CHARGED PARTICLES IN RADIO-FREQUENCY FIELDS

Consider an electron travelling parallel to the axis of an accelerating structure with a velocity $v < c$. If the structure develops an electromagnetic field having transverse components, the transverse force applied to the electron is:

$$F_\perp = e \left[ \mathbf{E}_\perp + \mathbf{v} \times \mathbf{B}_\perp \right] e^{i \theta_0}$$

where $\mathbf{v} = v \mathbf{u}$, $\mathbf{u}$ being the unit vector along the longitudinal axis oz, $\mathbf{B}_\perp = \mu \mathbf{J}_\perp$ and $\theta_0$ is the initial phase difference between the particle and the longitudinal component of the wave.

First considering a travelling wave propagating along oz one has:

$$\mathbf{E}_\perp = \mathbf{E}_\perp(x, y) \exp i(\omega t - \beta z)$$
$$\mathbf{B}_\perp = \mathbf{B}_\perp(x, y) \exp i(\omega t - \beta z)$$

where the phase velocity is defined as $v_\perp = \omega / \beta$

Hence:

$$F_\perp = e \left[ \mathbf{E}_\perp + v \mathbf{u} \times \mathbf{H}_\perp \right] \exp i(\omega t - \beta z + \theta_0)$$

Analysis of Maxwell's equations leads to the following identity relating the transverse components of the field to the longitudinal electric component:

$$\mu (\mathbf{u} \times \mathbf{E}_\perp) = -\frac{1}{v_\perp} \mathbf{E}_\perp + i \frac{\omega}{v_\perp} \mathbf{E}_\perp \times \mathbf{E}_z$$

and the force now becomes:

$$F_\perp = e \left[ (1 - \frac{v}{v_\perp}) \mathbf{E}_\perp + i \frac{\omega}{v_\perp} \mathbf{E}_\perp \times \mathbf{E}_z \right] \exp i(\omega t - \beta z + \theta_0)$$

Assuming the particle travels in synchronism with the wave, $v = v_\perp$, and since $t = z/v$ one gets $\omega t = \omega z/v = \beta z$. Hence:

$$F_\perp = \frac{e}{\beta} \mathbf{E}_\perp \mathbf{E}_z$$

having chosen $\theta_0 = -\pi/2$ which means that the particle is in phase with the transverse components of the travelling wave.
The previous expression is general and can be applied to all types of travelling waves. It shows that for a synchronous wave the combined effect from the transverse electric and magnetic fields is proportional to the transverse gradient of the longitudinal electric field component. Applied to classical waves one can conclude that:

- for TE waves, since $E_z = 0$, there is no transverse deflection whatever the particle velocity is, provided the synchronism condition is satisfied. In other words the transverse magnetic field exactly compensates the transverse electric field for this case.

- for TM waves, synchronous with the particle, the deflecting force is finite but decreases as $v = v_\phi$ increases. For $v_\phi$ approaching $c$, it can be seen from Maxwell's equations that $V_x E_z$ tends to zero; hence the transverse deflection from synchronous TM waves goes to zero for ultra-relativistic particles.

As a first conclusion, one does not expect any transverse deflection from classical travelling waves, synchronous with ultra-relativistic particles. Notice that above a few MeV electrons can be considered as ultra-relativistic ($v = c$).

Let us consider now the case of classical standing-wave modes, knowing that although an accelerating structure has been designed to propagate a peculiar accelerating mode without reflection, higher-order standing-wave modes can develop locally for instance where mechanical transitions occur (change in iris diameter for tapered iris-loaded structures to keep the accelerating gradient constant). Deflecting properties of standing-wave modes can be emphasized with a very simple example. Consider for instance a cylindrical cavity in which a TM$_{110}$ mode is excited; the field components are:

$$E_z = E_0 J_1(kr) \cos \phi$$

$$H_r = -\frac{E_0}{\rho} \left[ J_1(kr)/kr \right] \sin \phi$$

$$H_\phi = -i \frac{E_0}{\rho} J'_1(kr) \cos \phi$$

with $\rho = \sqrt{\mu_0/\varepsilon_0}$, $k = 2\pi/\lambda_0 = \omega/c$, where $\lambda_0$ is the free space wavelength.

Expanding the Bessel functions and assuming the wave is polarized in the horizontal plane, one gets near the axis:

$$E_z = \left( \frac{2E_0}{3\pi} \right) r$$

$$H_y = -i \frac{E_0}{\rho} = -i \left( \frac{2E_0}{3\pi} \right) \frac{\hat{y}}{r} \cdot$$

A relativistic electron traversing the cavity of length $L$ near the axis will receive a transverse momentum impulse from the $H_y$ component:

$$\Delta p_x = e \int_0^L \frac{(v_0 H_y)}{v} \, dz = i e \int_0^L \left( \frac{2E_0}{3\pi} \right) \frac{\hat{y}}{r} \, dz$$

showing that a standing wave TM mode can deflect relativistic particles. In fact this finite deflection comes from the interaction of the particle with the backward wave, which of course is not synchronous with the particle, and can only have a limited interaction length.
At this point one should mention that the off-axis particle will also interact with the longitudinal electric component of the deflecting mode, leading to an energy increment:

$$\Delta W = e \int_{0}^{L} \left( \frac{\partial E_z}{\partial x} \right) x \, dz$$

Depending on the sign of this quantity the particle can either get more energy or lose a fraction of its initial energy. In the latter case the particle gives energy to the deflecting mode.

A similar treatment in the case of standing wave TE mode would show that no deflection is expected in that case, which in fact is quite obvious since both the backward and forward waves have no $E_z$ component.

3 DEFLECTING MODES IN CIRCULAR IRIS LOADED WAVEGUIDES

Up to now only TM and TE modes have been considered, and for which little trouble can be expected, since for relativistic particles no synchronous deflection can occur. However, these modes happen to be independent solutions of Maxwell’s equations only in the case of simplified structures such as smooth waveguides and closed boxes. In practice there must be holes, for instance in a resonant cavity, for the beam passage and if one considers a TM mode in such a real cavity, the magnetic component of the field, which normally lies in a plane perpendicular to the axis, will be distorted in the neighbourhood of the holes (Fig. 4) resulting in an additional axial magnetic field component. Thus the presence of the end holes results in a mode which is no longer a pure TM mode, but a TM like mode with an associated longitudinal magnetic field. Such a mode is called a hybrid mode.

In order to satisfy Maxwell’s equations it can be shown that in the general case two independant hybrid modes are found; they are called HE (hybrid electric) and HM (hybrid magnetic) and they become TE and TM modes in the special case of simple boundary conditions. Hybrid modes are also very often called HEM (hybrid electro magnetic) modes in the literature.

To accelerate relativistic electrons one mainly uses travelling iris loaded waveguides (Fig. 5) which change the phase velocity to that of light.

![Diagram](image)

Fig. 4 Fringing field in the cut-off of a resonant cavity

![Diagram](image)

Fig. 5 Iris loaded structure
The analysis of TE and TM modes shows that they tend to become plane waves in the limiting case where \( v_\phi = c \); thus they are no longer independent solutions, and hence hybrid modes are necessary to satisfy completely Maxwell’s equations. More generally the irises of a loaded waveguide will have an effect similar to the end holes of a cavity since they will distort the field lines and introduce additional field components.

The first hybrid deflecting mode is the HEM\(_{11}^*\) since it can be shown that the HEM\(_{01}^*\) tends to split into two independent TM\(_{01}^*\) and TE\(_{01}^*\), the former being used for acceleration.

The general expressions for the components of this deflecting hybrid mode are complicated. However, in the limiting case where the phase velocity \( v_\phi \) is equal to the light velocity \( c \), one gets simple algebraic terms for an iris loaded structure \( (k/\alpha) \):

\[
\begin{align*}
E_r &= -iE_o \left[ \left( \frac{1}{2}ka \right)^2 + \left( \frac{1}{2}kr \right)^2 \right] \exp(i\phi) \exp(ikz-i\omega t) \\
E_\phi &= E_o \left[ \left( \frac{1}{2}ka \right)^2 - \left( \frac{1}{2}kr \right)^2 \right] \exp(i\phi) \exp(ikz-i\omega t) \\
E_z &= 2E_o \left[ \left( \frac{1}{2}kr \right)^2 \right] \exp(i\phi) \exp(ikz-i\omega t) \\
Z_{0H,r} &= E_o \left[ 1 - \left( \frac{1}{2}ka \right)^2 - \left( \frac{1}{2}kr \right)^2 \right] \exp(i\phi) \exp(ikz-i\omega t) \\
Z_{0H,\phi} &= iE_o \left[ 1 - \left( \frac{1}{2}ka \right)^2 + \left( \frac{1}{2}kr \right)^2 \right] \exp(i\phi) \exp(ikz-i\omega t) \\
Z_{0H,z} &= i2E_o \left[ \left( \frac{1}{2}kr \right)^2 \right] \exp(i\phi) \exp(ikz-i\omega t)
\end{align*}
\]

where \( Z_0 = \mu_0 c \) is the free space wave impedance, \( k = 2\pi/\lambda_0 \) is the free space propagation constant, and \( a \) is the iris radius.

Since the transverse force vector is given by

\[
F_\perp = E_\perp \times B \quad (k = \omega/c = \beta)
\]

it can be seen that in the present case the magnetic and electric forces no longer compensate each other at light velocity. The components of the transverse force are:

\[
\begin{align*}
F_r &= eE_o \sin\phi \exp(ikz-i\omega t) \\
F_\phi &= eE_o \cos\phi \exp(ikz-i\omega t)
\end{align*}
\]

Note that space harmonics also exist in order to satisfy the periodic boundary conditions due to the irises, but they do not contribute to the deflection since in general they are not in synchronism with the particle.

Hybrid modes can exhibit some curious properties; depending on the choice of the waveguide parameters, such as 2\( a \) and 2\( b \) (see Fig. 5), the group velocity can be either positive or negative.

The group velocity is given by:

\[
v_g = P/\omega_S
\]

where \( \omega_S \) is the stored energy per unit length and \( P \) the time average power transmitted across the waveguide, for instance the closed surface \( S \) defined by the iris hole:

\[
P = \mathcal{R}_e \frac{1}{2} \int_S \left[ E_\perp \times H^*_\perp \right] dS.
\]

Since

\[
dS = rdr d\phi
\]

one gets:
Fig. 6a  Lower Passband

Fig. 6b  Upper passband

Fig. 7  Brillouin (or $\omega-\phi$) diagram for the HEM$_{11}$ mode at five locations along SLAC ten-foot accelerator section
\[ P = \mathcal{R}_e \left( \frac{1}{2} \int_0^a \int_0^{2\pi} \left[ E_x H^+_\phi - E_y H^+ \phi \right] r \, dr \, d\phi \right) . \]

From the field components set previously for the HEM$_{11}$ mode one gets finally:

\[ P = \pi a^2 \frac{e^2}{\varepsilon_0} \left( \frac{(ka)^2}{16} \right) \left[ \frac{(ka)^2}{6} - 1 \right] \]

showing that the group velocity is negative if \( ka < \sqrt{6} \).

For a standard iris loaded waveguide, such as the SLAC one at 3 GHz, the first hybrid mode exhibits a negative group velocity (Fig. 6 a), but not the next one (Fig. 6 b).

As will be seen later the first hybrid mode, which frequency is roughly 1.5 times the accelerating mode frequency, is the deflecting mode of interest for BBU because of its negative group velocity.

However, since most of the iris-loaded structures used for electron acceleration are tapered to keep the accelerating gradient constant, or quasi constant, the Brillouin diagram will show as many dispersion curves as there are different iris diameters along the structure. This is for instance illustrated in Fig. 7 for the SLAC case.

Figures of merit are also defined for deflecting modes, such as \( r, Q \) and \( r/Q \). The expression for \( r_\perp \) relevant in calculating the transverse deflection is given by:

\[ r_\perp = \left[ \frac{1}{B} \left( \frac{\partial E_z}{\partial t} \right) \right]^2 \left( \frac{dP}{dE} \right) \]

which takes account of the fact that \( E_z = 0 \) on the axis but not apart from the axis (\( V_\perp E_z \neq 0 \)).

4 REGENERATIVE BEAM BREAK UP

Regenerative BBU occurs in one accelerating section and is due to the deflecting HEM$_{11}$ wave travelling in the direction of the electron motion with a phase velocity slightly lower than the light velocity so that approximate synchronism is possible. The small difference in velocity will make the electron slip ahead. Depending on its initial phase the electron can be deflected in one or the other direction (Fig. 8).

---

**Fig. 8** Deflecting HEM$_{11}$ mode in an iris-loaded structure
If, for instance, the electron enters the structure with an initial phase such that the transverse deflecting force has its maximum value, then the longitudinal electric component of the mode is zero, but since the electron slips ahead it will enter in a longitudinally decelerating field, off-axis, and give energy to the mode. As a consequence a noise-generated HEM_{11} wave can be amplified by the beam itself as soon as it has been brought off-axis.

This can be better understood by using a schematic representation of the Lorentz force near the axis, in a system co-moving with the HEM_{11} wave (Fig. 9).

Electrons entering the structure at phase points 1 and 3, corresponding to maximum deflecting forces, will move off axis and since they travel faster than the wave they will enter in a retarding field and thus will transfer energy to the field. If these electrons leave the structure at points 1' and 3' corresponding to a n-slippage they will have reached the maximum deflection. Electrons which enter the structure at intermediate phases, such as point 2 will in general also transfer a positive (or zero) amount of energy to the field. The optimum phase slip, giving a maximum deflection, depends on the initial electron phase relative to the wave, and it can be as much as 180° as seen before.

In a dispersive structure there will be in general some frequency at which the phase slip is optimized and it is near this frequency that beam break up is most likely to occur (Fig. 7).

In addition to the amplification mechanism which has been described, regeneration (or enhancement of the amplification) can occur due to the backward wave characteristic of the HEM_{11} mode ($v_g < 0$). As a matter of fact the energy deposited flowing back upstream will reinforce the original deflecting field.

Finally if the corresponding generated power exceeds the power losses into the walls, both the field and the deflection will grow exponentially leading to a transverse instability.
The starting current, or instability threshold, for the backward wave oscillation has been estimated by several authors\(^{1,6}\) by equating the power generated by the beam to the power propagating into the wave and lost into the walls.

The energy given by the beam to the deflecting mode can be written as:

\[
\Delta W = \int_0^L q E_z \, dz = - \int_0^L q \left( \frac{\partial E_z}{\partial x} \right)_b \, dz
\]

where \(q\) is the accelerated charge, \(E_z\) the longitudinal electric component of the deflecting mode, \(x\) the beam transverse displacement relative to the axis. The index "b" means that the integral must be performed along the beam path.

The generated power, time averaged over an RF cycle, in complex notation, becomes:

\[
P_b = -\frac{1}{2} \mathcal{R}_e \int_0^L \left[ \frac{\partial E_z^*}{\partial x} x \right] \, dz
\]

where \(I\) is the beam current, \(L\) the length of the accelerating structure.

The first term in the bracket is obtained from the expressions of the field components of the HEM₁₁ mode:

\[
\frac{\partial E_z}{\partial x} = k E_0 e^{i(\omega t - k z)}
\]

where \(t = z/c\) assuming the particle velocity remains constant and equal to \(c\) and where \(\beta\) is assumed to be slightly different from \(k = \omega/c\). Expanding \(\beta\) near the point \(\beta_0 = \omega/c\) where the straight line \(v_\phi = c\) crosses the dispersion curve (see Fig. 6 a) one gets:

\[
\beta = \beta_0 - (\omega - \omega_0) / |v_\phi|
\]

Introducing the following quantity:

\[
\delta \beta = (\omega - \omega_0) \left[ \frac{1}{c} + \frac{1}{v_\phi} \right]
\]

one gets:

\[
\frac{\partial E_z}{\partial x} = k E_0 e^{i \delta \beta z}
\]

The second term in the bracket is obtained as follows; since:

\[
P_x = \frac{d}{dz} \int_0^z p_x \, dz
\]

then:

\[
x(z) = \int_0^z \frac{p_x}{P_z} \, dz
\]

Using the results of Section 2 for the transverse momentum:

\[
P_x = \frac{1}{kc} \int_0^z \frac{\partial E_z}{\partial x} \, dz
\]
one finally gets:

$$P_b = -\frac{1}{2} \frac{e}{p \epsilon c} \int k E_0^2 \int_0^L dz \int_0^z dz' \int_0^{z''} e^{i \delta \beta (z'' - z')} dz''$$

having assumed $P = p = \text{cte}$.

Integration of the previous expression leads to:

$$P_b = 2k(L/\pi)^3 (e/p \epsilon) I E_0^2 g(\alpha)$$

with

$$\alpha = \frac{1}{L} \delta \beta$$

$$g(\alpha) = \frac{1}{2} \left( 1 - \cos \alpha - \frac{1}{2} \alpha \sin \alpha \right)/\left( \alpha/\pi \right)^3$$

It is found that there is an optimum value for the phase-slip parameter $\alpha$ which gives the maximum effect:

$$\alpha = 2.65 \quad \Rightarrow \quad g(\alpha) = 1.04$$

The power loss can be deduced from the definition of the transverse shunt impedance per unit length:

$$r_\perp = \frac{(1/k)^2 (\partial E_0^2/\partial x)^2}{dP/dz}$$

Since:

$$Q = \frac{\omega w_\perp}{dP/dz}$$

$$w_\perp = \frac{P}{v}$$

where $w_\perp$ is the stored energy per unit length, then:

$$P = \frac{v_0}{\omega} \frac{1}{r_\perp k^2} \left( \frac{\partial E_0}{\partial x} \right)^2 = \frac{v_0}{\omega} \frac{Q}{r_\perp} E_0^2$$

Equating the power loss to the generated power leads to the current threshold for regenerative BBU:

$$I_\text{th} = \frac{\pi}{g} \frac{v_0}{c} \frac{Q}{r_\perp \lambda_0} \frac{pc}{e (\lambda_0^3)}$$

where $\lambda_0$ is the free-space wavelength.

**Numerical application**

Consider a 1 meter long S-band structure operating at 2.8 GHz in the $\pi/2$ mode, with an accelerating gradient of 15 MeV/m. The estimated characteristics of the first deflecting mode are:

- Frequency $= 4.25$ GHz
- $v_0/c = 0.02$
- $r_\perp \lambda_0 / Q = 200 \Omega$

Hence, one gets for the threshold, $I_\text{th} = 66$ mA. It must be noticed that in similar structures BBU has been observed at the level of 100 mA, after several microseconds.
5 CUMULATIVE BEAM BREAK UP

The cumulative beam break up, or multisection beam break up, differs considerably from the previous one. Here each section acts like an amplifier which provides a small increase in the amplitude of the transverse deflecting wave (Fig. 10). Even though the gain per stage is very small the total gain in a long accelerator can be very large. At each amplifying cavity there is a transverse displacement modulation and a transverse momentum modulation on the beam.

![Diagram of Local HEM_{II} Oscillations and Electron Bunches](image)

**Fig. 10** Multisection HEM_{II} transverse deflection

The transverse displacement modulation excites the cavity through the interaction with the off-axis $E_z$ field component of the HEM_{II} mode, and the resulting $H_z$ field component provides an additional momentum kick to the beam.

In the drift space between cavities the transverse momentum is converted into additional displacement.

This modulation further excites the resonant field in the downstream cavities which in turn deflect the next bunches even more until finally they scrape the accelerator walls.

The effect manifests itself at beam currents well below the threshold for regenerative BBU. It was first observed and extensively studied at SLAC^{6}).

The model for cumulative BBU assumes that the effect of an entire accelerator section is equivalent to an impulse at a single point (Fig. 11). This description applies particularly to machines which use tapered sections for which the synchronous length at any frequency is very short.

![Diagram of Schematic Representation for Cumulative BBU](image)

**Fig. 11** Schematic representation for cumulative BBU
Hence the whole system can be considered as a transport system:

\[
\begin{bmatrix}
    x \\
    p_x
\end{bmatrix}
= 
\begin{bmatrix}
    m_{11} & m_{12} \\
    m_{21} & m_{22}
\end{bmatrix}
\begin{bmatrix}
    1 & 0 \\
    \Delta \rho_{n-1} & 1
\end{bmatrix}
\begin{bmatrix}
    x \\
    p_{x, n-1}
\end{bmatrix}
\]

where the matrix \( m_{ij} \) can include external focusing lenses.

Obviously cumulative BBU will depend on the magnetic focusing channel along the accelerator which gives additional transverse momentum kicks that can compensate partially the cavity deflections.

Since the information is transferred from one unit to the next by the beam itself, a slight change in the geometry of each structure, which also changes the deflecting mode frequency, will provide a detuning effect that can lower the efficiency of cumulative beam break up. This trick has been used on recent linacs and is very similar to the one which consists of tapering a single structure to lower the regenerative effect.

The standard model for cumulative BBU does not use the field configuration for the mode, but rather assumes that the mode is characterized by a vector potential \( A_\lambda \) and makes use of the fact that the transverse kick is directly related to that vector potential\(^{11}\):

\[
\Delta p_x = \frac{ie}{kc} \int \frac{\partial E_z}{\partial x} \, dz = e \int \frac{\partial A_z}{\partial x} \, dz .
\]

The deflecting mode is considered as a single cavity eigen mode and is characterized by a vector potential \( A_\lambda \) that obeys the Helmholtz equation:

\[
\nabla^2 A_\lambda + (\omega/c)^2 A_\lambda = 0 .
\]

However, since a beam travels along the cavity the actual time dependant vector potential \( A(r,t) \) must obey the wave equation:

\[
\nabla^2 A - \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} = - \mu_0 J .
\]

Assuming the following expansion\(^{12,13,14}\):

\[
A(r,t) = \sigma(t) A_\lambda(r)
\]

one gets:

\[
(\frac{\partial}{\partial t} + \omega_\lambda^2 \sigma) A_\lambda = \frac{1}{\mu_0} J .
\]

If the current is flowing in the z direction, \( J_x = J_y = 0 \) and the component \( A_{\lambda z} \) remains.

Taking the scalar product of the previous equation with \( A_\lambda \) and integrating over the volume occupied by the field leads to:

\[
\bar{\sigma} + \omega_\lambda^2 \sigma = \frac{1}{\mu_0} \int_V \frac{J_z A_{\lambda z} \, dV}{\int_V A_\lambda^2 \, dV} .
\]
Assuming $x$ is independent of $z$ in the active region and if in addition $A_{\lambda x}$ varies linearly with $x$, one can write:

\[
\int J_z A_{\lambda x} \, dV = I \times \int \frac{2A_{\lambda x}}{\lambda x} \, dz.
\]

The integral in the denominator is related to the stored energy:

\[
V_s = \frac{e_0}{2} \int E^2_\lambda \, dV = \frac{e_0 \omega^2}{2} \int A^2_{\lambda x} \, dV.
\]

Mixing previous formulae leads to:

\[
\left[ \frac{d^2}{dt^2} + \omega^2 \right] \Delta p_x = \frac{e I x \omega}{2V_s} \left[ \int \frac{2A_{\lambda x}}{\lambda x} \, dz \right]^2.
\]

The total transverse shunt impedance of the cavity is such that:

\[
\frac{R_1}{Q} = \left[ \frac{1}{k} \int \frac{2E_\lambda}{\lambda x} \, dz \right]^2 = \left[ \frac{c}{\omega} \int \frac{2A_{\lambda x}}{\lambda x} \, dz \right]^2 \quad \text{(in ohm)}.
\]

Hence one finally gets:

\[
\left[ \frac{d^2}{dt^2} + \omega^2 \right] \Delta p_x = \frac{e I x \omega^2}{2c^2} \left( \frac{R_1}{Q} \right).
\]

Losses in the cavity can be taken into account in the bracket by adding a term $(\omega/\omega_0) d/dt$, and defining the damping factor as $\alpha = \omega/\omega_0$.

Integrating the previous expression by applying the Green-function method gives the solution relating $\Delta p_n$ at the $n^{th}$ unit to $x_n$ at the same unit:

\[
\Delta p_{x_n} = \frac{e \omega^2}{2c^2} \left( \frac{R_1}{Q} \right) \int_0^t I(t') \, x_n(t') \, e^{-\alpha(t-t')} \sin \omega(t-t') \, dt'.
\]

having introduced the initial conditions such that $\Delta p = 0$ and $d(\Delta p)/dt = 0$ at $t = 0$. The integral shows that at time $t$, in cavity number $n$, the effect depends on the sum of the displacements of each part of the beam which has already passed the cavity. The displacement of each part of the beam depends on the momentum kick which was given to that part of the beam in the previous cavity and can be computed from the transfer matrix. A computer code would divide the beam in elementary portions corresponding to transit times $\Delta t'$ and replace the integral by a sum. In fact since the beam in a linac is bunched these portions could be taken corresponding to the RF micro bunches. A sophisticated theory\(^{14}\) has included the bunching in the starting assumptions giving the final result in cavity number $N$ in terms of a summation over the bunch number $M$.

Attempts to find analytic solutions of the equations of cumulative BBU have been made by several authors\(^{6,11,13,14}\).
It is found that the beam break up can be characterized by three regimes. The first corresponds to an exponential increase of bunch displacement with time (or RF bunch number). The second corresponds to the maximum displacement, while the third is the steady state regime.

In the exponential growth regime, for a coasting beam and no focusing the e-folding factor is given by:

\[
F_e(t) = \frac{3}{25/3} \left( \frac{z^2 I_0 c t \pi^2 R_L}{L V_0 \lambda^2 Q} \right)^{1/3}
\]

where \( L \) is the distance between the input of two successive cavities, \( e V_0 = \gamma m_0 c^2 \), \( \lambda \) the wavelength of the active mode, \( z \) the position along the accelerator.

The maximum steady state displacement arising from an initially modulated beam has an e-folding factor which can be written, with no focusing:

\[
F_e(c \nu) = (3)^{3/4} \left( \frac{2 \pi I_0 R_L}{2L V' \lambda} \right)^{1/2}
\]

valid for an accelerated beam with energy much greater than the initial energy. \( V' = dV/dz \) represents a uniform accelerating gradient.

In the case where a smooth focusing is introduced and if the accelerating gradient and the focusing strength are constant along the accelerator the e-folding factor becomes:

\[
F_e' = F_e \left( 1 - \frac{k_\beta^2 z^2}{F_e^2} \right)
\]

Here \( k_\beta \) is the betatron wave number of the focusing system and \( C \) is equal to 1/2 for the steady state case and 3/4 for the transient case.

Compensation of cumulative beam break up can be partly attained by good design of the focusing system. Improvement can also be made by minimizing the positioning errors (beam off-axis) and the noise from RF sources.

* * *

* * *
REFERENCES


4) Y. Garault, CERN 64-43 (1964).


10) B.W. Montague, Particle Separation at High Energies : II Radio frequency separation, Progress in Nuclear Techniques and Instrumentation, 3, (1968), North Holland Publ. Co.


BEAM LOADING

D. Boussard
CERN, Geneva, Switzerland

ABSTRACT
Beam loading on RF cavities may seriously limit the performance of high-intensity circular accelerators or storage rings. The RF power requirements to correct for beam loading will be first examined in several typical cases (lepton and hadron machines). Then, the methods to control the RF system (feedback and feedforward) and achieve stability under heavy beam loading conditions will be reviewed.

1. INTRODUCTION

In accelerator language, beam loading usually refers to the effects induced by the passage of the beam in the radio frequency cavities. As such, it could be considered to be one particular example of the more general problem of the beam interaction with its surroundings, in this case the cavity impedance.

However the beam loading problem deserves a special treatment, for several reasons. Firstly, the RF cavities are very often the largest contributor to the total ring impedance (in the following we shall concentrate on circular machines) and, consequently, power considerations play a very important rôle in beam-loading problems. Secondly, contrary to many other machine elements, the RF cavities are well known items being carefully designed and measured, from the RF point of view, and are easily accessible from the outside world via the RF power amplifier. Dedicated correction techniques can therefore be used where not only the cavity but also its associated RF amplifier are included.

In the following we shall first consider the stationary situation established in the beam-cavity system, the two extreme cases being when the bunches are wide apart and when every bucket is filled. Travelling-wave cavities with their inherent advantages as far as beam loading is concerned will be examined in this context.

Before settling to the stationary situation, the beam-cavity system undergoes a transient phase which may be very harmful to the beam, especially for hadron machines without natural damping. To circumvent this problem, it will be shown that RF power must be available. Finally the various methods used to control the RF amplifier-cavity combination in order to suppress beam-loading effects will be reviewed.
2. **SINGLE-BUNCH PASSAGE IN A CAVITY**

When the distance between bunches is very large compared to the filling time of the cavity, the fields induced by the previous bunches, or the previous bunch passages of the same bunch, have decayed sufficiently and can be neglected. Consequently, before the bunch passage the RF waveform is a pure sinewave produced by the RF generator (Fig. 1a).

The effect of the bunch passage is to excite an additional field in the cavity (Fig. 1b). For a short bunch (short compared to the RF period) and considering only the fundamental resonance of the cavity, the excited waveform is an exponentially decaying sinewave oscillating at the resonant frequency of the cavity \( \omega_c \).

Combining the generator driven and beam driven waveforms, one obtains the total voltage \( V(t) \) at the cavity gap (Fig. 1c).

![Fig. 1 Single-bunch passage in a cavity](image)

In vector representation the power delivered to the beam by the RF generator is simply:

\[
P = -\hat{V}_g \cdot \hat{i}_b
\]  

where \( \hat{V}_g \) is the generator driven voltage before the bunch passage and \( \hat{i}_b \) is the
fundamental component of the beam current. When crossing the gap the charge q, \(i_b = q/T_b\); \(T_b\) being the bunch distance) induces the voltage \(V_{bo}\), and loses a fraction of its energy which is finally transformed into heat in the cavity walls before the next bunch passage.

In the transient phase (short time scale compared to the RF period), the cavity gap impedance can be represented by a single capacitor \(C\) related to the cavity parameters by:

\[
\frac{1}{C} = \frac{R}{Q_o \omega_c}
\]

where \(\omega_c\) is the resonant frequency of cavity, \(Q_o\) the unloaded cavity quality factor and \(R\) the shunt impedance of the cavity (circuit convention). Obviously \(V_{bo} = q/C\), and the energy lost by the bunch and stored in the cavity just after the bunch passage amounts to:

\[
W = \frac{1}{2} C V_{bo}^2 = \frac{1}{2} q V_{bo}^2
\]

The net power received by the beam \(P\) is simply, remembering that \(\dot{i}_b\) and \(\dot{V}_{bo}\) are in phase:

\[
P' = -\dot{V}_g \cdot \dot{i}_b - \frac{1}{2} \dot{V}_{bo} \cdot \dot{i}_b
\]

\[
P' = -(\dot{V}_g + \frac{1}{2} \dot{V}_{bo}) \cdot \dot{i}_b = -\dot{V}_g \cdot \dot{i}_b = -(\dot{V}_g + \dot{V}_b) \cdot \dot{i}_b
\]

Here \(V\) is the effective RF voltage, delivering the net power \(P'\) to the beam.

In other words, the beam "sees" only one-half of its own induced voltage: \(V_b' = 1/2 V_{bo}\). This result is sometimes quoted as "the fundamental theorem of beam loading," and can be demonstrated more generally (P. Wilson\(^1\)) using linearity and superposition. Similarly, it is easy to show that, in fact, \(V_{bo}\) represents the sum of all beam induced voltages for all cavity modes. Equation (5) leads to the vector diagram of Fig. 2, which shows the voltages before and after the bunch passage and their relations with the bunch current. Obviously the voltage \(V_g\) to be delivered by the generator is higher for the same effective voltage \(V\), than in the case of no beam loading. The excess power can be easily computed from the cavity shunt resistance and beam current.

![Fig. 2 Vector diagram - Single-bunch passage in a cavity](image-url)
3. **MULTIPLE-BUNCH PASSAGES**\(^1,2\)

We look for a stationary solution, when an infinite train of bunches, spaced by \(h_b\) RF periods, crosses the cavity gap. Following P. Wilson's analysis\(^1\) we should replace \(V_g\) in Eq. (5), which represents the voltage just before the bunch passage, by the combination of the generator-driven voltage and the voltage resulting from all previous bunch passages. The decay of the voltage between two successive bunch passages is simply \(\delta = T_b/T_f\) where \(T_f\) is the cavity time constant \((T_f = 2Q_L/\omega_c, Q_L: \text{loaded cavity quality factor})\), and the phase shift with respect to the RF generator amounts to \(\psi = \omega_c T_b - 2\pi h_b\).

The relation \(\frac{V}{V_g} + 1/2V_{bo}\) will therefore transform into:

\[
\frac{V}{V_g} = V_{bo} \left( e^{-\delta} e^{j\psi} + e^{-2\delta} e^{2j\psi} + \ldots \right) + \frac{1}{2} V_{bo} = \frac{V}{V_g} + \frac{1}{2} V_{bo}
\]

in which the term in brackets represents the contributions of all previous bunch passages, whereas the last one reflects the effect of the bunch on itself (Fig. 3).

![Fig. 3 Vector diagram - Multiple-bunch passages](image)

Using the sum of the geometric series:

\[
V_{bo} \left( 1 + e^{-\delta} e^{j\psi} + e^{-2\delta} e^{2j\psi} + \ldots \right) = \frac{V_{bo}}{1 - e^{-\delta} e^{j\psi}}
\]

one obtains:

\[
\frac{V_{bo}}{2} \left( 1 + \frac{1}{1 - e^{-\delta} e^{j\psi}} - \frac{1}{2} \right)
\]

which, when separating real and imaginary parts leads to:

\[
V_b = V_{bo} \left[ F_1(\delta, \psi) + j F_2(\delta, \psi) \right]
\]

with:

\[
F_1(\delta, \psi) = \frac{1 - e^{-2\delta}}{2(1 - 2e^{-\delta} \cos \psi + e^{-2\delta})}
\]

\[
F_2(\delta, \psi) = \frac{e^{-\delta} \sin \psi}{1 - 2e^{-\delta} \cos \psi + e^{-2\delta}}
\]
If we introduce now the more usual cavity parameters:

\[ \tan \phi_c \text{ (detuning angle)} = \frac{2Q_L}{\omega_c} \frac{\omega - \omega_c}{\omega} \]  

(11)

\[ \beta \text{ (coupling coefficient); } Q_L = Q_o \frac{1}{1+\beta} \]  

(12)

and \( \delta_o = T_b/T_{fo} \) \( (T_{fo} \text{ being the filling time of the unloaded cavity}) \), Eq. (9) becomes:

\[ V_b = 2 \frac{i}{i_o} \Re \left\{ F_1(\beta, \phi_c) + jF_2(\beta, \phi_c) \right\} \]  

(13)

where \( i_o \) is the DC beam current.

The functions \( F_1 \) and \( F_2 \) are given by:

\[ F_1(\beta, \phi_c) = \frac{1-e^{-\delta_o (1+\beta)}}{2\delta_o} \]  

(14)

\[ F_2(\beta, \phi_c) = \frac{e^{-\delta_o (1+\beta)} \sin \delta_o (1+\beta) \tan \phi_c}{D} \]  

(15)

\[ D = 1 - 2e^{-\delta_o (1+\beta)} \cos[\delta_o (1+\beta) \tan \phi_c] + e^{-2\delta_o (1+\beta)} \]  

(16)

From these expressions, it is possible to calculate the generator power needed to produce a given accelerating voltage \( V \). For a generator which is assumed to be matched, by using, for instance, a circulator between generator and cavity, one obtains \(^1\):

\[ P_g = \frac{(1+\beta)^2 V^2}{4\beta} \frac{1}{2R} \left\{ \frac{1}{2} \left( A^2 + B^2 \right) \right\} \]  

(17)

where \( A \) and \( B \) are complicated functions of cavity and beam parameters \(^1\). Numerical computations are required to optimize the various parameters in order to minimize \( P_g \).

4. **LIMITING CASE \( \delta_o = 0 \)**

When the bunch distance \( T_b \) is short compared to the unloaded cavity filling time (Fig. 4), Eqs.(10) simplify to:

\[ F_1(\delta_o, \beta) = \frac{-1}{\delta_o^2 (1+\beta) (1+\tan^2 \phi_c)} \]  

(18)
Fig. 4 Case $\delta_0 = 0$. The RF waveform is a quasi sinusoid

$$F_2(\delta_0, \beta) = \frac{-\tan \varphi}{\delta_0 (1 + \beta) (1 + \tan^2 \phi_c)}.$$  \hspace{3em} (19)

Combined with (13), one obtains:

$$V_b = \frac{1}{2} \frac{R}{1 + \beta} \frac{1}{1 - j \tan \phi_c}.$$  \hspace{3em} (20)

In this case the cavity gap waveform is approximately sinusoidal (Fig. 4), and the equivalent circuit of Fig. 5, where the beam current is represented by its component at the RF frequency ($i_b = 2i_0$ for short bunches), can be used. There the coupling coefficient $\beta$ is simply related to the cavity and generator shunt resistances by: $\beta = R/R_g$. Obviously, $V_b$ given by Eq. (20) is the cavity voltage (sinusoidal in the approximation $\delta_0 = 0$) developed when $i_g = 0$.

Fig. 5 Equivalent circuit for the case $\delta_0 = 0$

In the vector diagram of Fig. 6a, the total current $i_L = i_g + i_b$ drives the RLC circuit and produces the gap voltage $V$. For a given $V$, the vector $i_L$ follows the dotted line in Fig. 6a, when the detuning angle $\phi_c$ is varied. This is because the admittance of the equivalent RLC circuit has a constant real part.

If we again assume a generator connected to the cavity via a circulator, the required RF power:

$$P_g = \frac{1}{2} R_g i_g^2.$$  \hspace{3em} (21)
is a minimum for given \( V, i_b \) and \( \phi_s \), if the two conditions:

\[
\tan \phi_{cm} = -\frac{R \cos \phi_s}{i_b (1+\beta) V}
\]

(22)

\[
\beta_m = 1 + \frac{R \sin \phi_s}{i_b V}
\]

(23)

are fulfilled. The minimum RF power for \( \phi_c = \phi_{cm} \) and \( \beta = \beta_m \) is given by:

\[
P_{gm} = \frac{V^2}{2R} + V i_b \sin \phi_s
\]

(24)

the first term corresponding to the cavity losses and the second to the power delivered to the beam. For the optimum conditions where no power is reflected towards the circulator, it is easy to see from Eq. (22) that \( \phi_c = \phi_{cm} \) corresponds to \( i_b \) and \( V \) being in phase (Fig. 6b). Usually there is a servo-tuner which measures the phase difference between RF drive and gap voltage, and controls the cavity tune via a mechanical tuner or ferrite bias, for instance. At equilibrium of the servo-tuner, Eq. (22) is automatically satisfied.

On the contrary, the cavity coupling is usually fixed by construction, and can only be optimized for a given value of \( i_b \) and \( \phi_s \). However for a hadron storage ring, where \( \phi_s = 0 \), the critical coupling (\( \beta = 1 \)) corresponds to the optimum situation.

\[\text{Fig. 6 Vector diagrams for the case } \phi_o = 0. \text{ Optimum tuning in (b).}\]
5. THE CASE OF A TRAVELLING-WAVE STRUCTURE

It is known that in a long chain of coupled resonators travelling waves can propagate within some frequency limits i.e. passbands of the structure. In the travelling mode of operation, the structure is terminated by its characteristic impedance and behaves like a transmission line (Fig. 7). At synchronism, the phase velocity \( v_p \) of the wave equals the particle velocity \( v_p \), giving maximum voltage seen by the beam, like an RLC circuit at resonance.

![Fig. 7 Schematics of a travelling-wave structure](image)

For a single-bunch passage, it is usually possible to neglect the cavity coupling as the energy transfer from cell to cell is much slower than the bunch velocity \( (v \ll v_p; v_p \text{: group velocity}) \). The previous analysis can therefore be applied to the quasi uncoupled resonators. It is generally applied also for the standing-wave mode of operation of multicell cavities, which are non-terminated structures\(^3\). However, for a repetitive train of many bunched, the RLC equivalent circuit model would fail in the travelling-wave mode because the waves excited by previous bunch passages also propagate along the structure.

For instance, at exact synchronism \( (v = v_p) \), the waves excited in each cell by the beam passage add linearly in the forward direction, and, on average, cancel in the reverse direction, in a forward travelling wave structure. In other words, the decelerating electric field \( E_z \) is simply proportional to the distance along the structure counted from the feed point.

If the synchronism is not perfect, we must introduce a phase factor \( \exp j(\omega t - \beta z) \) for each individual wave, where \( z = v_p t \) and \( \beta \) is the wave propagation constant, with the result that the induced field \( E_z(z) \) is proportional to the integral:

\[
E_z(z) \sim \int_0^z \exp j(\omega t - \beta \psi) \, dz = \int_0^z \exp j\theta \, dz .
\]  

Expanding \( \theta \) around the synchronous point \( (\omega_0, \beta_0) \) one obtains:

\[
\theta = \omega t - \beta_0 z = \left( \frac{\omega_0 + \Delta \omega}{v_p} - \beta_0 - \Delta \beta \right) z = \left( \frac{\Delta \omega}{v_p} - \Delta \beta \right) z .
\]
Introducing $v_s = \Delta w/\Delta \theta$ and the phase slip angle $\tau$ defined by:

$$\tau = L \frac{\Delta w}{v_s} \left( 1 - \frac{v_A}{v_p} \right)$$  (27)

$L$ being the structure length, one obtains:

$$\theta = - \tau \frac{E}{L}$$  (28)

and:

$$\int_0^L \exp j\theta \, dz = \frac{1 - \exp(-j\frac{\tau}{L}z)}{j\frac{\tau}{L}}.$$  (29)

In particular, for $z = 0$, the integral vanishes: the beam induced field is zero on the upstream end of the structure (generator side). This is a very important result as it shows that, for a travelling-wave structure, there is no beam loading effect seen by the RF generator, which always remains matched without the need for a circulator. In the case of a backward-wave structure, where the generator is connected to the downstream end of the structure, this result is still valid. Beam loading only changes the field on the load side: not all the generator power goes into the load, some fraction is transferred to the beam.

The total voltage $V_b$ seen by the beam is obtained by integrating the electric field, given by (29) along the structure:

$$V_b = \int_0^L \frac{1 - \exp(-j\frac{\tau}{L}z)}{j\frac{\tau}{L}} \, dz.$$  (30)

It gives finally:

$$V_b = i_b R_2 \frac{L^2}{8} \left[ \left( \frac{\sin \tau}{\tau/2} \right) ^2 - 2j \frac{\tau}{\tau^2} \left. \frac{x - \sin \tau}{\tau^2} \right| \right]$$  (31)

where the proportionality factor $R_s$, called the series impedance of the structure, is characteristic of its geometry\(^4\). Figure 8 shows a plot of equation (31) in the complex impedance plane.
6. **TRANSIENT CORRECTION**

Consider again the case of a cavity represented by its RLC equivalent circuit. Even in the case $\delta_0 = 0$ (quasi sinusoids) the stationary solution of Section 4, where only the RF frequency component is considered, cannot describe transient situations, when $V$ or $i_b$ change rapidly.

The worst case situation corresponds to a sudden change of $V$ (e.g. transition) or $i^*_b$ (injection of a prebunched beam, fast ejection of part of the beam). The resulting unwanted transient must of course be damped for the stationary solution described above to settle down properly, but it must also be short compared with the synchrotron period $T_s$. This condition will ensure that the effects on the beam such as mismatch and subsequent blow-up, or even loss of particles, will be minimum, or in other words that beam loading will be properly corrected.

We shall now consider the example of a prebunched beam $i_b$ injected into an empty machine. Before injection the servo-tuning keeps $i^*_t = i^*_b$ and $V$ in phase. Immediately after injection the new vector $i^*_b$ destroys the equilibrium, and $V$ changes by a large amount until the tuning loop returns the cavity to a different value. Unless one uses very fast tuners \(^5\), which may lead to multiloop stability problems \(^6\), it will take more than a small fraction of a synchrotron period for the tuning loop to settle at its new value, the result being a strong distortion of the longitudinal phase plane.

The only way to maintain $V$ constant during the transient phase of the tuner is to act via the RF power generator which provides a fast control of $V$. The obvious solution (Fig. 9) is to change $i^*_g$ into $i^*_g'$ when the beam is injected. If we make:

![Fig. 8 Impedance seen by the beam of a travelling-wave structure](image-url)
$$i_g' = i_g - i_b$$

Fig. 9 Correction of beam-loading transient with the power generator

$$i_g' = i_g - i_b$$  \hspace{1cm} (32)

the total current in the cavity does not change and, at constant tuning, \( V \) stays constant.

In the simple case of no acceleration, the amplitude of the peak current \( i_g' \), which must be delivered by the RF power tube during the transient phase of the tuner, is given by:

$$\left| i_g' \right|^2 = \left| i_g \right|^2 + \left| i_b \right|^2 .$$  \hspace{1cm} (33)

This extra current must be delivered in a non-matched load in this simplified example. With a circulator inserted between the RF amplifier and the cavity (Fig. 10), the generator is always matched and the extra current also means extra power. Again for \( \Phi_s = 0 \) a similar analysis can be made; it gives the peak power \( \hat{P} \) needed during the transient phase of the tuner:

$$P = P_o \left[ 1 + \left( \frac{\left| V \right| \left| i_b \right|}{4 P_o} \right)^2 \right]$$  \hspace{1cm} (34)

where \( P_o \) is the power for no beam (matched cavity); the excess power \( \hat{P} - P_o \) is simply wasted into the load to keep \( V \) constant. One can optimize \( \hat{P} \) by selecting the best cavity impedance \( R_{opt} = 2V/i_b \) and obtain the simple result:

$$\hat{P}_{opt} = 2P_o = \left| V \right| \left| i_b \right|/2 .$$  \hspace{1cm} (35)

Fig. 10 A circulator to match the RF power generator
Remember, nevertheless, that this is the worst case situation and in certain cases it is possible to minimize the required peak power or peak current. In particular, by pretuning the cavity before injection, one can make the two powers, before and after injection, equal and obtain in this case $V_{opt} = \frac{|V||I_b|}{4}$ (for $\phi_s = 0$). With superconducting cavities, usually without variable tuners, the peak power can even be reduced to $|V|/|I_b|/8^9$. One can also reduce the transient on $I_b$ with multiple injections of smaller currents, or by adjusting the bunching factor of the injected beam.

In the above analysis, we assumed the filling time of the cavity to be long compared to the revolution period $T_o = 1/f_o$ but small with respect to $T_g$, which means that all bunches are submitted to the same RF voltage. If this is not the case ($Q_L < h$; $h$: harmonic number), unequal filling of the ring will give a modulation of $V$ at $f_o$ and its multiples. The same analysis applies here: at each "batch" passage transient beam loading must be corrected to make all bunches see the same RF voltage. This effect is particularly important in large machines not only at injection but also at transition. As before, condition $[34]$ is valid in the worst case situation, $I_b$ being now the batch current.

7. RF DRIVE GENERATION

During the transient phase of the tuner, we must synthesize $I_g$ to meet condition $i_g$ (32) and correct for the effect of beam loading. It obviously implies that $I_g$ (or the corresponding power $P$) is available from the RF generator, otherwise transient beam loading cannot be corrected completely. Various techniques used to generate the proper $I_g$ will now be examined.

7.1 Amplitude and phase servo loops

The synthesis of $I_g$ in order to keep $V$ constant irrespective of the beam loading can be done with two servo loops (Fig.11): the first acting on the amplitude of $I_g$ (amplitude loop) controls $|V|$, and the second maintains the relative phase of $V$ and $I_b$ constant through the control of the phase of $i_g$ (phase loop). The cut-off frequency $f_c$ of the loops must be much larger than the synchrotron frequency $f_s$, which means very strong damping of beam oscillations. This justifies the simplified stability analysis$^6$ in which the beam transfer function is neglected. The cut-off frequency $f_c$ is obviously limited by the delays in the system, including the cavity bandwidth, but more fundamentally by the revolution frequency $f_o$. The simple configuration of Fig. 11 with high loop gains cannot correct transient beam loading at $f_o$ and its multiples.

Steady beam loading with its associated cavity detuning could excite mode $n = 0$ (Robinson instability$^{10}$) if it were not heavily damped by the phase loop. However, mode $n = 1$ (one wavelength per turn) which is not damped may show up also due to cavity
detuning and must be suppressed by dedicated feedback circuitry acting through the RF cavity itself.

Independent amplitude and phase control of $V$ is a well known technique for proton machines. It works satisfactorily for relatively small beam currents, i.e. when the gap voltage is predominantly determined by the generator current (typically $|i_b| < |i_g|$).

For higher beam currents, a variation of the amplitude of $i_b$, for instance, not only results in a variation of the amplitude of $V$ but also of its phase. In other words, the two loops, which were independent at low beam currents, become coupled together and an unstable behaviour of the system results above a certain beam current threshold.

Pedersen's detailed analysis,\(^6\) confirmed by experiments on the CERN PS booster, lead to the generalized Robinson stability criterion, valid for $\Phi_s = 0$:

$$
\frac{|i_b|}{|i_g|} < \sqrt{2 + \frac{f_a}{f_a} + \frac{f_T}{f_T} + \frac{f_p}{f_p} + \frac{f_T}{f_T} + \frac{f_p}{f_p} + \frac{f_T}{f_T}}
$$

(36)

where $f_a$, $f_p$, and $f_T$ are the unity gain frequencies of the loops (amplitude, phase and tuning respectively). Although the threshold is weakly dependent on the loop cut-off frequencies, it might be dangerous in this configuration to increase the servo-tuner bandwidth.

Although it is in principle possible to compensate loop coupling by an additional decoupling circuitry so increasing the instability threshold, a much simpler solution is offered by feedforward correction.

7.2 Feedforward correction

With a pick-up electrode followed by a filter centered at $f_{RP}$, one can obtain a signal proportional to $-i_b$ independently from the RF system, and generate $i_g'$ (RF
drive with beam) according to (32) with a simple adder. Applied to the amplitude and phase servo loops described in section 7.1, the method consists of injecting into the input of the RF amplifier the pick-up signal, with proper amplitude and phase \( \phi \), to generate the \(-i_b\) current at the gap (Fig. 11). The amplitude and phase loops now act on the quantity \( i_g \), corresponding to no beam loading, instead of \( i'_g \), and the cross couplings between loops are removed, as can be shown analytically and experimentally \(^{12}\) (Fig. 12). As a result, the instability threshold can be considerably increased and, for instance, stable operating conditions have been observed in the CERN PS for \( |i_b|/|i_g| = 8 \) to 10.

Fig. 12 Transient response of amplitude loop with (b) and without (a) feedforward correction (CERN PS machine). The loop response becomes oscillatory at high intensity (bottom trace) without feedforward correction.

The signal corresponding to \(-i_b\) does not need to be synthesized with the ultimate precision as it only removes the loop couplings and restores stability. For a varying RF frequency, the pick-up to cavity delay must be continuously adjusted, and the variations in gain and phase of the RF power amplifier (assumed linear) corrected. In the CERN PS, a coarse feedforward correction (cavity compensation) covers the whole RF frequency swing during acceleration, but more precise settings are possible at a few critical (fixed frequency) points.

Feedforward correction can also be considered as a means to reduce the effective impedance of the cavity seen by the beam. At the RF frequency, the beam induced voltage on the cavity amplifier combination is zero for a perfect correction. From this point of view, high amplitude and phase loop gains at \( f_g \) are no longer required to correct beam loading as \( V \) is automatically kept constant by the feedforward compensation. Application of this technique (low loop gains) was, for instance, used on the Brookhaven AGS during adiabatic capture.

It is interesting to mention a variant of the feedforward technique derived from the Alvarez linear accelerator technology. If the generator is a grided tube (tetrode or triode), its output impedance is high if maximum RF power is to be extracted from the tube. When connected to the cavity by a long line, it fully reflects the beam loading wave travelling from the cavity to the generator. One can choose the length of the line to make the reflected wave cancel the beam induced voltage at the gap, the high impedance
of the generator is then transformed into a quasi-short circuit at the cavity. Note that, even with no voltage induced on the gap, the generator sees a mismatched load with beam and must be able to deliver the current under this condition. This technique is in use on the CERN PS 200 MHz RF system, with trombones inserted on the feeder lines of the fixed-tune cavities.

If the pick-up to cavity delay is adjusted to be exactly one turn \( T_0 \), beam loading cancellation can be achieved, not only at \( f_{RP} \), but also at frequencies \( f_{RP} \pm n f_0 \). This is relatively easy at fixed RF frequency, for example in the CERN ISR 12), but with modern sampled or digital filters and variable delays it is also possible to follow a varying RF frequency. The overall result is a rapidly changing impedance, ideally zero at frequencies \( n f_0 \), but twice as large at intermediate frequencies, \((n+\frac{1}{2})f_0\), where there are no beam current components (Fig. 13). With a one turn delay and perfect cancellation, the voltage perturbation only lasts \( T_0 \) which is small compared with \( T_s \) since \( Q_s = \left( f_0 T_s \right)^{-1} \) is usually \( \ll 1 \). In other words the reduction of the magnitude of the cavity impedance at the synchrotron satellites \( n f_0 \pm m f_s \) is also large (factor \((2\sin mQ_s)^{-1}\)) for a small \( Q_s \).

![Fig. 13 Residual impedance at synchrotron sidebands for a one turn delay feedforward correction](image)

### 7.3 RF feedback around the power amplifier

We can consider the cavity itself as a beam pick-up tuned at \( f_{RP} \) and obtain the \( -i_b \) signal from the gap itself. This leads to the configuration of Fig. 14 in which one obviously recognizes a feedback loop built around the RF power amplifier. From the loop equations one obtains:

\[
\frac{i_g'}{i_g} = 1 - \frac{GZ i_b}{1 + GZ}
\]

which, for \( GZ \gg 1 \) (\( GZ \): loop gain, \( Z \): cavity impedance) reduces to equation (32):

\[
\frac{i_g'}{i_g} = 1 - i_b
\]

\( i_g \) being here the generator current with no beam.
Fig. 14 RF feedback around the power amplifier

The feedback loop automatically generates the correct compensating signal, which is another way of saying that it keeps the controlled parameter V constant. One can consider RF feedback as a means to reduce the output impedance of the RF amplifier, a well known design being the cathode follower with its low output impedance which shunts the cavity. However, stability of the cathode follower with a reactive load needs careful study.\textsuperscript{13}

Even simpler, but of limited efficiency, is the use of a triode instead of a tetrode as the RF power tube, the internal plate to grid feedback reducing the output impedance. In the same way pulsing the DC current of the RF tube or powering second tube, in parallel\textsuperscript{14}, has been used to reduce the output impedance of the RF amplifier for short periods.

In the case of Fig. 14, the cavity parameters (pole at $f_{RF}/2Q_c$) and the total delay of the feedback path determine the loop stability. The preamplifiers which are selected for the shortest propagation delay must be located very close to the power amplifier-cavity combination. As an example Table 1 gives the parameters for the CERN PS booster second harmonic system, operating between 6 and 16 MHz:\textsuperscript{15}

<table>
<thead>
<tr>
<th>Feedback parameters of the CERN PSB second-harmonic system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preamplifier gain</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Power</td>
</tr>
<tr>
<td>Propagation delay</td>
</tr>
<tr>
<td>Impedance reduction factor:</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

For a varying RF frequency one could, in principle, adjust the delay of the return path to keep the 180° phase condition at $f_{RF}$. However, in many designs, for instance the second harmonic PS booster and the future PS RF system, a wide bandwidth preamplifier is used to keep the total delay short enough to ensure stability over the entire RF
frequency range, without programming the phase. In this case it is extremely important to
damp the higher resonances of the cavity or to reject the corresponding signals in order
to avoid parasitic oscillations of the feedback system at high frequencies.

The RF feedback technique is very attractive since it reduces the effective impedance
of the cavity not only at the RF frequency but also over a large bandwidth. This feature
is particularly helpful to avoid self-bunching instabilities in storage rings for
debunched beams and is used at the CERN ISR and AA for instance.

In such a conventional feedback system the total phase slip should be less than about
\( \pm \pi/4 \) over the unity gain bandwidth \( 2 \Delta \omega \) of the system, giving the condition:

\[
\Delta \omega = \frac{\pi}{4T}
\]

where \( T \) is the overall delay in the feedback path. For a fixed tuned cavity and a small
detuning angle, the cavity impedance (RLC approximation) far from the \( \omega_c \) resonance is
given by \( Z = R/2jQ_L (\Delta \omega/\omega_c) \). The overall loop gain, \( GZ \), at the \( \pm \Delta \omega \) points is
of the order of unity: this gives an upper limit for \( GZ \) and a minimum value of the
impedance seen by the beam, \( R_{\text{min}} \), given by:

\[
R_{\text{min}} = \frac{2}{\pi} \frac{R \omega_c}{Q_L}
\]

Equation (39) shows that the ultimate performance of wideband RF feedback only depends
on \( T \) and the cavity geometry \( (R/Q_L \) parameter).

If applicable, i.e. if \( T \) can be made small enough, this is the best solution to the
problem of beam loading since it provides wide band coverage and avoids the need for
critical adjustments. Very large impedance reduction factors of several orders of
magnitude have been achieved at low RF voltages in the CERN AA for instance using fixed
cavity tune without a servo loop. If, however, a servo tuner is used, it may be necessary
to control it by the normalized reactive power of the amplifier\(^{16}\).

7.4 The RF feedback with long delay\(^{17}\)

In large RF systems, the CERN SPS for instance, long delays may be unavoidable and
the conventional RF feedback would have a too restricted bandwidth, much smaller than the
cavity bandwidth itself in the SPS case. Transient beam loading at multiples of \( f_0 \)
would not be corrected, leading to phase oscillations of fractions of the beam and
possibly coupled bunch instabilities.

In order to solve the problem, we observe that a large gain \( G \) is only needed in the
vicinity of the revolution frequency harmonics where beam current components exist.
Outside these bands, the phase rotation due to the excessive delay will be unimportant if
\( G \) can be made small enough. With a return path transfer function having a comb-filter
shape with maxima at every \( f_0 \) harmonic, this condition can be satisfied. In addition, the overall delay of the system must be extended to exactly one machine turn\( (T_0) \) to ensure a zero phase at the \( f_\text{RF} + n f_0 \) frequencies.

The comb filter transfer function (Fig. 15) is of the form:

\[
H(j\omega) = \frac{G_0}{1 - K \exp(-j\omega T_0)}
\]

(40)

where \( G_0 \) and \( K \) are constants \((0 < K < 1)\).

![Graph of the comb filter transfer function](image)

**Fig. 15 Comb-filter transfer function** \( K = 7/8 \) \( N = 462 \)

Combined with the one turn delay (transfer function: \( \exp(-j\omega T_0) \)), the overall open loop transfer function becomes:

\[
G(j\omega)Z(j\omega) = \frac{G_0 Z(j\omega)}{\exp(j\omega T_0) - K}
\]

(41)

represented in the complex plane by a circle for a slowly varying \( Z(j\omega) \). The complex plane origin is encircled and therefore the gain of the system is limited by the stability condition. In the vicinity of the cavity resonance, where \( Z \) is maximum and real, (note that for a travelling wave structure \( Z \) is always real\(^4\)), the circle crosses the negative real axis at a distance \(-G_0 Z/(1+K) \) from the origin.

![Complex plane diagram](image)

**Fig. 16 Open-loop transfer function for RF feedback with long delay**
Stability obviously requires that $|G_Z|<1+K$, and it can be shown that this condition is also sufficient even outside resonance for an RF cavity approximated by a single RLC equivalent circuit.

Again for $Z$ real, the apparent impedance of the cavity $Z'$:

$$Z' = Z \frac{\exp(j\omega T_0) - K}{\exp(j\omega T_0) - K - G_0 Z}$$

(42)

is real for frequencies:

$$f_{RF} + n f_o \quad Z' = Z \frac{1 - K}{1 - K + G_0 Z} \ll Z$$

(43)

and:

$$f_{RF} + (n+\frac{1}{2}) f_o \quad Z' = Z \frac{1 + K}{1 + K - G_0 Z}$$

(44)

To stay at a reasonable distance from the stability limit, take for instance $G_0 Z = (1+K)/2$. This gives, at frequencies $f_{RF} + (n+\frac{1}{2}) f_o$, $Z' = 2Z$ as in the case of feedforward correction, whereas for the revolution frequency harmonics one obtains:

$$Z' = Z (1-K)$$

(16)

for $(1-K)\ll 1$.

By making $K$ close to unity, RF feedback approaches the theoretical performance of the feedforward correction but with all the inherent advantages of closed loop systems in particular no critical adjustments are needed. Similarly, the time response of the RF feedback is entirely determined by the one turn delay as in the feedforward case. Note that the unity gain frequency of the servo in this case is of the order of $f_o/2$.

The residual impedance at the synchrotron sidebands is approximately the same as for a one turn delay feedforward correction (for $K=1$ and $G_0 Z=1$); its phase changes sign at each $n f_o$ harmonic resulting in a rotation of the complex synchrotron frequency shift curve. The coupled-bunch, cavity-driven, instability thresholds must be obtained numerically.\(^{17}\)

Except for relatively small machines with fixed RF frequency, long delay feedforward or feedback techniques could only be envisaged with the help of modern signal processing technology, i.e. sampled or digital filters. The digital comb filter is derived from the well known first-order low-pass recursive filter shown in Fig. 17. With a sampling frequency $N f_o$ locked to a subharmonic of the RF frequency, the theoretical bandwidth of the filter is $N f_o/2$, corresponding to $N/2$ maxima in the comb filter response ($N = 462$ in the SPS design). Implementation of the one turn delay is straightforward in digital technology with a memory (R.A.M. or first-in first out-type).
The speed of the various elements, limited by the cycle time $(T_0/N)$, may become very critical requiring the fastest A-D converters (flash converters), memories and multipliers (parallel multipliers). For this reason the number of bits is limited, 8 bits in the ADC and 12 bits in the multiplier array in the case of the SPS, but no adverse effects from the quantization errors can be observed. However, $K$ cannot be made very close to unity with a small number of bits and the residual impedance $Z^*$ at the revolution harmonics is essentially determined by this technological limitation $(1-K = 1/8$ for the SPS case).

The RF signals may have to be translated in frequency to be conveniently processed. Coherent mixing with separate channels for in-phase and in-quadrature components is necessary to reject the unwanted image frequencies, (measured rejection >35dB), and to make the overall electronic chain look a linear system. For a varying RF frequency the correct phase can even be maintained with an artificial delay inserted between the output and input local oscillators as in Fig. 18.
REFERENCES

1) P. Wilson, CERN ISR-TH/78-23 (1978).
Polarization in Electron and Proton Beams

J. Buon

Laboratoire de l'Accélérateur Linéaire, 91405 ORSAY, France

Abstract

One first introduces the concept of polarization for spin 1/2 particle beams and discusses properties of spin kinetics in a stationary magnetic field. Then the acceleration of polarized protons in synchrotrons is studied with emphasis on depolarization when resonances are crossed and on the methods of reducing it. Finally, transverse polarization of electrons in storage rings is discussed as an equilibrium between polarizing and depolarizing effects of synchrotron radiation. Means for obtaining longitudinal polarization are also treated.

Introduction

Spin is an important feature of nuclei and subnuclear particles, as well as their mass and electric charge. In general, interactions between them depend on their spin. The experimental study of these interactions with unpolarized beams and targets cannot investigate this spin dependence and is incomplete. Polarization experiments are able to reveal important and new aspects of Nature. There have been in the past many examples of unexpected results obtained in such experiments, the most famous one being the discovery of parity violation in $\beta$-decay. One then could ask why so few polarization experiments are done in Nuclear and Subnuclear Physics. The reason is that these experiments are generally more difficult and more delicate. In particular, polarized beams are more elaborate to produce than unpolarized beams. Usually their intensity and their reliability are lower. If it was not so, all experimentalists would ask for polarized beams! Surely progress in the development of polarized beams would be valuable.

The physics of polarized beams is a wide topic, not often familiar to accelerator physicists. It is difficult to cover it completely in a limited time. We will restrict ourselves to the acceleration of polarized protons in synchrotrons and to the polarization of electrons in storage rings, i.e. to the most common high-energy polarized beams. We will not consider other polarized beams like secondary beams, electron beams in linear accelerators and synchrotrons, muon beams, deuteron beams, ... We will concentrate on the spin kinetics of electron and proton polarized beams and we will not study other aspects like polarized-ion sources and polarization monitoring.

The aim is to explain the physics of spin kinetics in these polarized beams to accelerator physicists. No attempt will be made to use less familiar mathematical formalisms (like the SU(2) representation of spin rotations), to derive the basic formulae (such as the Thomas-BMT and Froissart-Stora equations or the formulae of Sokolov-Ternov and Derbenev-Kondratenko), or to treat particular details or more advanced topics, reserving these developments to specialists. We prefer to limit ourselves to an analysis of the physical contents of the basic equations and of their consequences, illustrated by experimental results.
We will not try to quote in references all the authors and contributors in the field of polarized beams. We limit the bibliography to a few general and recent reports which were models for preparing this lecture and which can be recommended to the non-specialist reader. The latter will find in them all the relevant references.

This lecture is divided into three parts. The first one is devoted to generalities on the physics of polarized beams which are useful for understanding the behaviour of polarized protons and electrons in accelerators, especially circular accelerators. We first remember the concept of spin and we extensively discuss the meaning of polarization for spin 1/2 particle beams. Some knowledge of Quantum Mechanics is not really needed as we will essentially take a semiclassical point of view, apart from two particular points which can possibly be omitted by the reader. Then the kinetics of spin motion in a stationary magnetic field is extensively studied, starting from the Thomas-BMT equation of spin motion, and with emphasis on the spin-orbit coupling. A general discussion of depolarization resonances is based on the consequences of spin-orbit coupling. Finally the great similarity with Nuclear Magnetic Resonance is stressed, recognizing that the basic features of spin kinetics are the same.

In the second part the acceleration of polarized protons in synchrotrons is studied with emphasis on depolarization when resonances are crossed and on the cures for reducing it. In particular spin kinetics in a ring equipped with "Siberian Snakes" is qualitatively discussed as "Siberian Snakes" appear essential for very high energies.

The third and last part is devoted to the polarization of electrons in storage rings, which has very different aspects. As in beam dynamics, the synchrotron radiation dominates spin kinetics in electron storage rings. Synchrotron radiation provides a polarizing mechanism (the Sokolov-Ternov effect) and enhances also beam depolarization. We qualitatively discuss both polarization and depolarization phenomena induced by synchrotron radiation and how to manage with them for obtaining a high degree of polarization. At the end we briefly discuss spin rotators for obtaining longitudinal polarization and indicate two particular and important problems: depolarization enhancement by large energy spread of beams at high energies and depolarization by the beam-beam interaction.
I. GENERALITIES ON POLARIZATION AND SPIN MOTION

I.1 Spin and magnetic moment of a particle

The spin $\mathbf{S}$ of a particle (electron, proton, ...) is an internal degree of freedom which behaves like angular momentum. It is an axial vector with quantized values of its modulus $|\mathbf{S}|^2$ and of its component $S_z$ on any axis $Oz$:

$$|\mathbf{S}|^2 = s(s + 1) \hbar^2$$

$$\frac{1}{\hbar} S_z = -s, -s+1, ..., s-1, s$$

where $\hbar$ is the Planck constant divided by $2\pi$.

The spin value $s$ is a half-integer for Fermions (1/2 for electrons, muons, protons, neutrons, ...) and an integer for Bosons (0 for $\pi$ and $K$ mesons, 1 for photons and deuterons)

The component $S_z$ can take $2s+1$ values from $-s\hbar$ to $+s\hbar$, i.e. two values $\pm \hbar$ for spin 1/2 particles.

Charged particles have a magnetic moment $\mathbf{\mu}$ proportional to their spin $\mathbf{S}$:

$$\mathbf{\mu} = g \frac{e}{2m_0} \mathbf{S}$$

where $e$ and $m_0$ are the electric charge and the rest mass of the particle, respectively ($\mathbf{\mu}$ is parallel to $\mathbf{S}$ for a proton and antiparallel for an electron according to the sign of their electric charge).

The gyromagnetic ratio $g$ is 2 for point-like Fermions in the Dirac theory. There are corrections and the deviation from 2 is measured by the gyromagnetic anomaly $a = (g-2)/2$ (very often also designated by $G$ in the literature):

$$\begin{align*}
\text{electron} & \quad 1.1596 \times 10^{-3} & \quad 1.1659 \times 10^{-3} & \quad 1.7928 & \quad -0.1430
\end{align*}$$

A charged particle placed in a magnetic induction $\mathbf{B}$ has a magnetic energy $\mathcal{W}$ given by:

$$\mathcal{W} = - \mathbf{\mu} \cdot \mathbf{B}$$

Here we will consider spin 1/2 particles (electrons and protons) which have two states of magnetic energy only.

I.2 Polarization of spin 1/2 particles

A bunch of spin 1/2 particles is polarized if their spins have a preferred direction. This situation is characterized by a polarization vector $\mathbf{\hat{P}}$ pointing in this direction. The length $|\mathbf{\hat{P}}|$ is the degree of polarization.
Here we will define the polarization vector $\vec{P}$ in the most general case. Essentially the polarization vector is a classical quantity (following a classical equation of motion) which determines completely any spin state of a spin 1/2 particle ensemble. These two properties justify the semiclassical description of polarization for spin 1/2 particles, which is based on the evolution of the polarization vector $\vec{P}$. This semiclassical description is totally equivalent to a purely quantum-mechanical description. The proof of these two properties which needs some knowledge in Quantum Mechanics, is given in sections I.2.2 and I.2.3. They have been put in appendix at the end of this paper such that they can be omitted if one is not familiar with Quantum Mechanics.

### I.2.1 Definition of the polarization vector $\vec{P}$

In a pure spin state of an individual particle the direction of spin $\vec{S}$ is the direction along which the spin component takes the maximum value $(+h/2)$ with probability 1. The polarization vector $\vec{P}$ is defined as the unit vector in this spin direction.

Now, for a bunch of $N$ particles with different polarization vectors $\vec{P}_i$ $(i = 1, N)$, the polarization vector $\vec{P}$ is defined as the barycentre of all the individual $\vec{P}_i$: \[ \vec{P} = \frac{1}{N} \sum_{i=1}^{N} \vec{P}_i. \] (I.2.1)

The degree of polarization $|\vec{P}|$ varies from 0 to 1 depending on the relative directions of the vectors $\vec{P}_i$.

When all the spins are parallel to OZ, the component $P_z$ of the polarization vector $\vec{P}$ for one particle is +1 if its spin is "up" ($S_z = +h/2$) and -1 if it is "down" ($S_z = -h/2$). If $N_+$ and $N_-$ are the numbers of particles with spin "up" and "down" respectively, the polarization vector $\vec{P}$ is parallel to OZ and its component $P_z$ measures the asymmetry in the populations of these two spin states: \[ P_z = \frac{N_+ - N_-}{N_+ + N_-}. \] (I.2.2)

A bunch of $N$ particles is unpolarized ($P_z = 0$) when $N_+ = N_-$, and completely polarized ($P_z = \pm 1$) when either $N_+$ or $N_-$ vanishes, i.e. when all the spins are either parallel or antiparallel to OZ.

Finally, abrupt transitions (spin-flip) between "up" and "down" states may occur, as for electrons radiating in a static magnetic field. The spin state of such an electron is mixed: a statistical mixture of the two spin states "up" and "down" with probabilities $q$ and $1-q$ respectively. In this case the polarization vector $\vec{P}$ is defined as the statistical average of the polarization vectors $\vec{P}_\pm$ of the two possible states: \[ \vec{P} = q\vec{P}_+ + (1-q)\vec{P}_-. \]

This statistical average is equivalent to an ensemble average with $q = N_+/(N_+ + N_-)$ and this case does not need to be distinguished in the following.
I.3 Spin precession in static electromagnetic fields

Following the tradition we will, from now on, use the expression "spin vector $\vec{s}$" for designating the polarization vector $\vec{f}$ of an individual particle. The expression "polarization vector $\vec{f}$" will be reserved to the case of a particle ensemble.

In this section we will study the classical motion of spin vector $\vec{s}$ in static electromagnetic fields. We will not consider the effect of emission (or absorption) of electromagnetic radiation, which happens when electrons radiate. Since these radiative effects occur in very short times, we will describe spin motion only between two consecutive radiative effects. For protons these effects are normally negligible and can be ignored.

The classical equation of spin motion will be first written down in the non-relativistic and relativistic cases (sections I.3.1 and I.3.2). Then the general properties included in the relativistic equation will be emphasized (section I.3.3). Finally, it will be explicitly shown that the classical equation is strictly equivalent to the Schrödinger equation for spinors in a quantum-mechanical formalism. Again this last part (section I.3.4) has been put in appendix and can be omitted if one is not familiar with Quantum Mechanics.

I.3.1 Non-relativistic particles

The spin-vector motion of an individual particle is given by the interaction of its magnetic moment $\vec{\mu}$ with the magnetic induction $\vec{B}$:

$$\frac{d\vec{s}}{dt} = \vec{\mu} \times \vec{B}$$  \hspace{1cm} (I.3.1)

This equation of motion can be rewritten, using formula (I.1.1):

$$\frac{d\vec{s}}{dt} = \vec{\omega}_L \times \vec{s}$$  \hspace{1cm} (I.3.2)

with $\vec{\omega}_L = - \frac{eB}{2\mu_0} \vec{B} = - (1+a) \frac{e}{\mu_0} \vec{B}$.

The motion is a precession around the field $\vec{B}$ at the "Larmor" frequency $\frac{e}{\mu_0}B$ times the gyromagnetic ratio $g$.

This precession is similar to the velocity rotation in a magnetic field:

$$\frac{d\vec{v}}{dt} = \frac{e}{\mu_0} \vec{v} \times \vec{B} = \vec{\omega}_C \times \vec{v}.$$  

with the "cyclotron" frequency $|\vec{\omega}_C| = \frac{e}{\mu_0}B$.

The relative frequency $Q_a$ of spin and velocity precessions is proportional to the gyromagnetic anomaly $a$:

$$Q_a = \vec{\omega}_L - \vec{\omega}_C = a \vec{\omega}_C.$$  \hspace{1cm} (I.3.3)

The measurement of $Q_a$ is the basis of all the "g-2" experiments which intend to measure this gyromagnetic anomaly.
I.3.2 Relativistic particles

The equation of spin-vector motion in an electric and magnetic field becomes:

$$\frac{d\hat{S}}{dt} = \hat{\omega}_{\text{BMT}} \times \hat{S},$$

(I.3.4)

with

$$\hat{\omega}_{\text{BMT}} = -\frac{e}{m_0 c} \left[(1+\gamma a) \hat{B}_\perp + (1-a) \hat{B}_\| - \left(a + \frac{1}{1+\gamma a}\right) \gamma \hat{B} \times \frac{\hat{S}}{c}\right],$$

where $\hat{B}_\perp (\hat{B}_\|)$ is the transverse (longitudinal) component of the induction field $\hat{B}$ relative to the particle velocity; $\gamma$ is the relativistic Lorentz factor and $\hat{B}$ the ratio of the velocity $\hat{V}$ to the light velocity $c$ (all quantities in MKS units).

In this "Thomas-Bargman-Michel-Telegdi" equation, referred to as Thomas-BMT equation, the fields $\hat{E}$ and $\hat{B}$, and the time $t$, are calculated in the laboratory frame, but the spin vector $\hat{S}$ is calculated in the rest frame of the particle for avoiding complicated Lorentz transformation of spin.

For comparison the velocity rotation in a transverse magnetic field $\hat{B}_\perp$ is given by:

$$\frac{d\hat{V}}{dt} = \hat{\omega}_c \times \hat{V}$$

with the relativistic "cyclotron" frequency $|\omega_c| = \frac{eB_\perp}{m_0 c}$.

I.3.3 General properties of spin precession

The spin-vector motion as given by the Thomas-BMT equation (I.3.4) is a rotation about the rotation vector $\hat{\omega}_{\text{BMT}}$ with an angular frequency

$$\frac{d\hat{S}}{dt} = |\hat{\omega}_{\text{BMT}}|$$

and with the following general properties:

i) The effect of an electric field $E$ has nearly the same amplitude as the effect of a magnetic field $B = E/c$. Therefore an electric field of $3 \times 10^6$ V/m is comparable to a magnetic field of one Tesla. The electric fields normally found in accelerators have then a negligible effect as compared to the magnetic fields and these electric fields will now be ignored.

ii) The spin rotating power of a longitudinal field $B_\|$, is inversely proportional to the particle momentum $\hat{p}$, exactly like the velocity rotating power of a transverse field $B_\perp$. More precisely the longitudinal-field integral $\int B_\| ds$, needed for rotating the spin by one radian, is:

$$\int B_\| ds \text{ (Tm/rad)} = \frac{10.479}{\hat{p}} \cdot \frac{P(\text{GeV/c})}{1 + a}. $$

(I.3.5)

This integral becomes very large at high energies.
iii) The relative frequency $\Omega_a$ of spin and velocity precessions in a transverse magnetic field $B_\perp$:

$$\hat{\Omega}_a - \hat{\Omega}_{BMT} = \Omega_c \left( -\frac{e}{m_0} \hat{z}_\perp \right),$$  \hspace{1cm} (I.3.6)

is exactly independent of the particle energy, and is $\gamma a$ larger than the cyclotron frequency $\Omega_c$:

$$\hat{\Omega}_a = \gamma a \hat{\Omega}_c.$$ \hspace{1cm} (I.3.7)

The vector $\hat{\Omega}_a$ is the spin rotation vector with respect to a frame following the particle motion (usually named orbit frame) as this frame rotates at cyclotron frequency $\Omega_c$.

The transverse-field integral, needed for rotating the spin by one radian in the orbit frame, $\int B_\perp ds$ is:

$$\int B_\perp ds \,(\text{Mrad}) = \frac{5.484}{\gamma} \cdot \frac{P_c}{E} \quad \text{for a proton},$$ \hspace{1cm} (I.3.8)

$$\int B_\perp ds \,(\text{Mrad}) = \frac{4.618}{\gamma} \cdot \frac{P_c}{E} \quad \text{for an electron},$$

where $E$ is the total relativistic energy of the particle. In a given transverse field, a proton and an electron with the same velocity have nearly the same spin rotation, as the larger mass of the proton is compensated by its larger gyromagnetic anomaly (formula I.1.2).

iv) At high energies, i.e. when $\gamma a > 1$, the spin rotating power of a longitudinal field becomes much smaller than the power of a transverse field. Therefore transverse fields are usually preferred for spin manipulations at high energies. Moreover the absolute precession frequency $\Omega_{BMT}$ in a transverse field becomes nearly energy-independent and spin rotation appears to be easier to realize than trajectory bending.

v) In a circular accelerator with distributed bending magnets, the spin motion is a succession of rotations. In one turn the mapping of spin is a rotation: product of all the successive rotations in individual magnets. This mapping is characterized by a precession axis $\hat{n}$ and angle $\Psi$, which play an important role for the spin kinetics in circular accelerators.

I.4 Spin-orbit coupling

According to the Thomas-BMT equation (I.3.4), the spin motion at a given energy is determined by the magnetic fields encountered by the particles. These fields depend on the individual trajectories followed by the particles. The spin motion is coupled to the orbital motion.

For instance, in an ideally planar ring, the reference orbit lies in the horizontal plane. All along this orbit the magnetic field is vertical and spin precesses around the vertical line.
On the other hand, along a vertical betatron trajectory, radial fields $B_r$ proportional to vertical displacement are experienced in quadrupoles. A small longitudinal field is also experienced in bending magnets where the trajectory has a vertical slope.

In general spin-orbit coupling is responsible for depolarizing effects since particles in a beam have slightly different trajectories and energies. Their spin vectors $\vec{S}$ rotate about different fields with different speeds. They tend to spread out in all directions and the polarization vector decreases in length. These depolarizing effects are the main concern for spin motion in accelerators. Their most general aspects will be studied in section I.6. The only exception where spin-orbit coupling does not cause depolarization is a horizontally flat beam, polarized in the direction of the vertical bending field, for instance in an ideally planar ring. Whatever the particle energy and motion in the horizontal plane, spin precession is about the vertical line and the vertically aligned spin vectors do not rotate at all. In this ideal situation the ring is said to be "spin-transparent". In general for reducing depolarization one tries to approach spin-transparency as much as possible.

On the other hand, the reverse coupling, an orbit perturbation depending on spin state, is expected as in Stern-Gerlach experiments. However this coupling is very weak at accelerator energies as the magnetic energy given by the Hamiltonian $\hat{H}$ (formula I.3.11 in appendix) is at most of the order of $a(e\hbar/2m)B = 10^{-14}$ MeV and is very much smaller than the kinetic energy. An effect of the Stern-Gerlach type cannot be observed in practice.

Now, taking account of the spin-orbit coupling, the question may be raised whether spin manipulations are at all possible since the trajectory optics in a circular accelerator is almost completely determined by many imposed constraints.

In particular, one could argue that, in one turn of a ring, the overall spin precession would be strictly proportional to the velocity rotation since spin rotation in a transverse field is nearly $v$ times the velocity rotation (formula I.3.7). Spin precession could not then be changed without great modification of beam optics. This argument is wrong at high energy because velocity rotations in magnets are relatively small and nearly commute between themselves, and at the same time spin rotations are large and do not commute. Then the result of successive rotations for spin and for velocity can be very different from a simple proportionality rule.
In other words spin manipulations are possible at high energy due to the non-commutativity of rotations. An example of this possibility is a spin rotator made of several transversely bending magnets, which bends the spin by $90^\circ$ but not the trajectory (see section III.6):

In conclusion, at high energy, in spite of the spin-orbit coupling, spin motion can be considered as a new degree of freedom to some extent which allows spin manipulations.

I.5 Spin closed solution and spin tune

Hereafter we will restrict our considerations to spin motion in circular accelerators. In this section we only consider on-momentum particles circulating on a reference closed orbit in a ring with non-uniform magnetic field such that the reference orbit is not necessarily planar.

The one-turn mapping, starting at azimuth $s$, is a rotation $T(s)$ with a precession axis $\hat{n}(s)$ and a precession angle $\psi$ (in general different from $2\pi$ for avoiding depolarization resonance). We will prove the following theorem:

**Theorem:** The one-turn precession axis $\hat{n}(s)$ is the periodic solution of spin motion, named the spin closed solution, and any spin-vector direction rotates by $2\pi \nu$ about $\hat{n}(s)$ in one turn, where $\nu$ is the spin tune and is independent of the initial azimuth $s$.

**Proof:** the one-turn rotations $T(s)$ and $T(s_o)$, starting at azimuth $s$ and $s_o$ respectively, can be related by:

$$T(s) = R(s_o, s) \cdot T(s_o) \cdot R^{-1}(s_o, s)$$

where $R(s_o, s)$ is the spin rotation between these two azimuths. The spin direction $\hat{n}(s)$ is the eigenvector of the rotation $T(s)$, corresponding to the eigenvalue 1:

$$\hat{n}(s) = T(s) \cdot \hat{n}(s)$$

or

$$\hat{n}(s) = R(s_o, s) \cdot T(s_o) \cdot R^{-1}(s_o, s) \cdot \hat{n}(s)$$

then:

$$R^{-1}(s_o, s) \cdot \hat{n}(s) = T(s_o) \cdot R^{-1}(s_o, s) \cdot \hat{n}(s)$$.
This relation shows that \( R^{-1}(s_0, s) \hat{n}(s) \) is also an eigenvector of the one-turn rotation \( T(s_0) \), for the same eigenvalue \( 1 \). The unicity of this eigenvector for a rotation, different from the identity, leads to:

\[
R^{-1}(s_0, s) \hat{n}(s) = \hat{n}(s_0)
\]
or

\[
\hat{n}(s) = R(s_0, s) \hat{n}(s_0)
\]
showing that \( \hat{n}(s) \) is effectively a solution of spin motion. This solution is periodic as, in a second turn, the spin mapping is the same, the reference orbit being periodic too.

For proving the second part of the theorem, let one consider another spin-vector direction \( \hat{\ell}(s) \) at azimuth \( s \), orthogonal to the spin closed solution \( \hat{n}(s) \). After one turn this direction \( \hat{\ell}(s) \) is mapped into \( \hat{\ell}'(s) \) and the one-turn precession angle is:

\[
\psi = (\hat{\ell}(s), \hat{\ell}'(s))
\]

Again the mapping between \( \hat{\ell}(s) \) and \( \hat{\ell}'(s) \) : \( \hat{\ell}'(s) = T(s) \hat{\ell}(s) \)
can be written:

\[
\hat{\ell}'(s) = R(s_0, s) T(s_0) \hat{\ell}(s)
\]
or

\[
R^{-1}(s_0, s) \hat{\ell}'(s) = T(s_0) R^{-1}(s_0, s) \hat{\ell}(s)
\]
showing that \( \hat{\ell}(s_0) = R^{-1}(s_0, s) \hat{\ell}(s) \) is mapped into \( \hat{\ell}'(s_0) = R^{-1}(s_0, s) \hat{\ell}'(s) \) in one turn starting at azimuth \( s_0 \). It follows that the spin precession angle \( (\hat{\ell}(s_0), \hat{\ell}'(s_0)) \) is also \( \psi \) as the directions \( \hat{\ell}(s_0), \hat{\ell}'(s_0) \) and \( \hat{\ell}(s), \hat{\ell}'(s) \) are related respectively by the same rotation \( R^{-1}(s_0, s) \) which conserves the angle between them. The spin tune \( \nu = \psi/2\pi \) is then independent of the azimuth \( s \).

**Exercise**: Prove this theorem by using the Floquet theorem.

(note: a rotation by an angle \( \psi \) has three eigenvalues: \( 1, e^{i\psi} \) and \( e^{-i\psi} \))

Consequently, it is often convenient to look at spin motion as a rotation about the spin closed solution \( \hat{n}(s) \), since the angle between the spin-vector direction and \( \hat{n}(s) \) is conserved for particles circulating on the reference orbit. This rotation has a \( 2\pi \nu \) phase advance per turn.

In an ideal ring with a uniform vertical field, the spin closed solution \( \hat{n}(s) \) is vertical everywhere and the spin tune \( \nu \), in the orbit frame, is given by:

\[
\begin{align*}
\text{Electrons} & : \nu = \frac{E_{(\text{GeV})}}{.44065} \\
\text{Protons} & : \nu = \frac{E_{(\text{GeV})}}{.52335} \\
\text{Deuterons} & : \nu = \frac{E_{(\text{GeV})}}{13.13}
\end{align*}
\]

(1.5.1)
as function of the total relativistic energy \( E \).

According to formula (I.3.6), the spin precession frequency \( \Omega_\perp \), relative to the orbit frame, is then:

\[
\Omega_\perp = \nu \Omega_c
\]

In a ring with a non-uniform bending field, the spin tune is in general different from \( \nu \). The most famous example is a ring equipped with a "Siberian Snake" (see section II.5), where the spin tune is \( 1/2 \) whatever the energy.
I.6 Resonant perturbations of spin motion

Normally polarized particles circulate in a circular accelerator with their spins pointing in the direction of the spin closed solution \( \hat{n}(s) \), which is the only stable direction of polarization as will be seen in part II for protons and part III for electrons. This direction \( \hat{n}(s) \) corresponds to on-momentum particles circulating on the reference orbit.

However, particle motion and energy slightly differ from these references, due to closed-orbit distortions, betatron and synchrotron oscillations. These perturbations of orbital motion lead to a perturbed spin motion, via the spin-orbit coupling. They produce a perturbing magnetic field \( \vec{B} \) which bends the spin vector \( \vec{S} \) away from the spin closed solution \( \hat{n}(s) \). Only the perturbing field component orthogonal to \( \hat{n}(s) \) needs to be considered here.

These perturbations are small and rapidly varying in time. Usually they tend to cancel out on average. Therefore a large spin deviation from \( \hat{n} \) can only occur if there is some piling-up of small perturbations. Such a coherent effect is observed when the perturbing field \( \vec{B}(s) \) has a component precessing about \( \hat{n} \) at the same frequency as the spin vector \( \vec{S} \). This is illustrated in the following figure which shows their orientations at different times as seen in the plane transverse to \( \hat{n} \) (the dotted arrows show the direction in which \( \vec{S} \) will be tilted).

For analyzing in frequency the perturbing field \( \vec{B}(\theta) \) let us define the complex quantity:

\[
b(\theta) = b_1 + ib_2 \quad (\theta = s/R, \ R = \text{average radius})
\]

where \( b_1 \) and \( b_2 \) are the two components of \( \vec{B} \) on two axes \( \hat{e}_1 \) and \( \hat{e}_2 \), orthogonal to \( \hat{n} \) (\( \hat{e}_1 \) lying in the transverse plane \( x0z \)). The frequency spectrum of \( b(\theta) \):

\[
b(\theta) = \sum_j b_j e^{i\nu_j \theta}
\]

involves a set of tunes \( \nu_j \) with the general expression:

\[
\nu_j = k_o + k_x q_x + k_z q_z + k_s q_s
\]

\((k_o, x, z, s : \text{integer})\)

since the perturbing field \( \vec{B} \) results from closed-orbit distortions with integer harmonics, betatron oscillations with \( q_x \) and \( q_z \) tunes, and synchrotron oscillations with \( q_s \) tune.
One notices that \( b_j e^{ij \theta} \) represents a transverse component of \( \vec{S} \) precessing about \( \hat{n} \) at the frequency \( \nu_j \). This component and the spin vector \( \vec{S} \) precess at the same frequency when the resonant condition:

\[
\nu = k_o + k_x \theta_x + k_z \theta_z + k_s \theta_s
\]

(I.6.1)
is fulfilled. Then large deviation of \( \vec{S} \) from \( \hat{n} \) will occur, which depends on particle energy and oscillation amplitudes, and which in general leads to some depolarization justifying the name: "depolarization resonance".

### I.6.1 General classification of depolarization resonances

Linear resonances are mostly produced by transverse quadrupole fields:

\[
\vec{B} = \vec{B}_x \frac{\partial}{\partial x} z + \vec{B}_z \frac{\partial}{\partial z} x
\]

(I.6.2)

where \( x \) and \( z \) are the trajectory displacements in the direction of the radial \( \hat{x} \) and vertical \( \hat{z} \) unit vectors respectively, and \( \frac{\partial}{\partial x}, \frac{\partial}{\partial z} \) are the corresponding field gradients.

These linear resonances are classified into the following families:

1) **Vertical betatron resonances** (also named intrinsic resonances)

\[
\nu = k_o \pm \theta_z
\]

These are produced by a vertical betatron oscillation:

\[
z = a_z \sqrt{\beta_z} \cos (\theta_z + \phi_z)
\]

whenever the spin closed solution \( \hat{n} \) is not pointing in the radial direction \( \hat{x} \). This is in general the case as \( \hat{n} \) is vertical for an accelerator ring lying in an horizontal plane.

Normally the integer \( k_o \) is a multiple of the ring superperiodicity \( P \) (\( k_o = kP \)), as it results from harmonics of the periodic functions \( \beta_z, \theta_z \) and \( \partial \beta_z / \partial \theta_z \) when analyzing the perturbing field \( \vec{B} \) in frequency. However, in a ring with gradient errors, \( k_o \) may be not superperiodic (\( k_o \neq kP \)).

2) **Horizontal betatron resonances**

\[
\nu = k_o \pm \theta_x
\]

These are produced by a horizontal betatron oscillation:

\[
x = a_x \sqrt{\beta_x} \cos (\theta_x + \phi_x)
\]

every time the spin closed solution \( \hat{n} \) is not vertical, as happens in a non-planar ring (very often due to small imperfections). They can also be produced by an \( x-z \) coupling:

\[
z = \varepsilon \cos (\theta_x + \phi_x),
\]

when \( \hat{n} \) is not radial.
iii) **Integer resonances** (also named imperfection resonances)

\[ \nu = k_o \]

These are produced either by a vertical closed-orbit distortion \( (x_{c0} \neq 0) \) when \( \hat{n} \) is not radial, or by a horizontal closed-orbit distortion \( (x_{c0} \neq 0) \) when \( \hat{n} \) is not vertical. These closed-orbit distortions are due to magnet imperfections. For random imperfections, all integer harmonics are present and integer resonances are separated by one unit in spin tune, i.e. by 440 MeV for electrons, 523 MeV for protons and 13.1 GeV for deuterons (formula I.5.1). For systematic imperfections, superperiodic resonances \( (k_o = k_P) \) are produced also.

iv) **Synchrotron resonances**

\[ \nu = k_o \pm q_s \]

These are produced by synchrotron oscillations:

\[ y = D_y \frac{\Delta P}{P} \cos (Q_s \theta + \phi_s), \quad y = x, z \]

in the \( \hat{x} \) or \( \hat{z} \) direction, proportionally to the corresponding dispersion \( D_y \) and to the amplitude \( \Delta P/P \) of energy deviation.

It is worth noting the absence of parametric resonances \( \nu = k/2 \) (k integer). Moreover, for polarized beams in storage rings, a half-integer spin tune is generally the best operating point, which is midway between depolarization resonances.

Nov., **nonlinear resonances** are produced by higher-order multipole fields:

\[ b(\theta) \propto x^p z^q \quad (p + q > 1) \]

The frequency analysis of \( b(\theta) \) leads to a resonant condition:

\[ \nu = k_o + k_x q_x + k_z q_z \]

with \( |k_x| < p \) and \( |k_z| < q \).

For instance a sextupole field will drive nonlinear resonances with \( |k_x| + |k_z| = 2 \). The beam-beam interaction in storage rings will also drive series of nonlinear resonances.

Moreover, large-amplitude synchrotron oscillations cause a large frequency modulation of spin tune which is normally proportional to particle energy. Similarly to frequency modulation in RF-waves, several synchrotron satellite-lines appear in the frequency spectrum of spin motion:

\[ \nu_o = k_s q_s \quad (|k_s| > 1) \]

where \( \nu_o = \nu \alpha \) is the usual spin tune for vanishing synchrotron amplitude. Then several synchrotron satellites:
\[ v = \nu_R + k_s \Omega_S \quad (|k_s| > 1) \]

appear on each side of any depolarization resonance \((v = \nu_R)\) of the previous types.

**I.6.2 Similarity with Nuclear Magnetic Resonance (NMR) phenomena**

Summarizing the mechanism of an isolated depolarization resonance, a perturbing field \(\vec{b}\) rotates about the spin closed solution \(\hat{n}\) at frequency \(\nu_R\). The spin vector \(\vec{S}\) is precessing about \(\hat{n}\) at frequency \(v\). Resonant spin motion occurs when both frequencies are equal \((v = \nu_R)\).

In a standard NMR experiment, the magnetization vector \(\vec{M}\) of a nuclear magnetic substance is precessing about a stationary magnetic field \(\vec{B}_0\) at the Larmor frequency \(\omega_L\). A transverse RF-field \(\vec{B}_{RF}\) is superimposed with a frequency \(\omega_R\). This \(\vec{B}_{RF}\) field can be decomposed into two fields \(\vec{B}\) and \(\vec{B}'\) rotating in opposite directions. Nuclear magnetic resonance occurs when both frequencies are equal \((\omega_L = \omega_R)\), \(\vec{B}\) and \(\vec{M}\) staying in phase. This resonance is experimentally observed as a signal in RF-energy absorption by the substance, corresponding to the population of higher-energy states.

The motions of \(\vec{M}\) and \(\vec{S}\) look similar and appear simpler within a new frame:

Let us consider a frame rotating about the closed solution \(\hat{n}\), at the resonance frequency \(\nu_R\). With respect to this rotating frame, the perturbing field \(\vec{b}\) is at rest and the spin vector \(\vec{S}\) precesses about \(\hat{n}\) at the frequency \(v - \nu_R\).

On resonance, the spin precession about \(\hat{n}\) is vanishing and one is left with only the spin precession about the stationary perturbing field \(\vec{b}\). The spin vector becomes "up" and "down" periodically, explaining the population of the two spin states.

In fact, an experiment of NMR type is used in electron storage rings for a very accurate \((\approx 10^{-5})\) energy calibration. When the applied RF-field enters into resonance with spin precession a sharp depolarization occurs. Then the frequency of the RF-field is equal to the spin tune \((v = \gamma a)\) and is proportional to the beam energy.
II. ACCELERATION OF POLARIZED PROTONS IN SYNCHROTRONS

Polarized proton beams have been accelerated successfully in several synchrotrons: the ZGS at Argonne (up to 12 GeV), the AGS at Brookhaven (at present up to 16.5 GeV), Saturne at Saclay (up to 3 GeV) and the KEK PS at Kyoto (at present in the 0.5 GeV booster).

The scheme for accelerating polarized protons in a synchrotron is nearly the same for all machines. One can take the example of the AGS scheme, shown in Fig. 1. It involves a polarized ion source delivering a 25 μA H\(^+\) beam which is accelerated up to 200 MeV in a linac. After injection with electron stripping, about \(10^{10}\) protons per pulse are accelerated in the synchrotron ring. After reaching the top energy (16.5 GeV in 1984, 26 GeV planned for 1985), protons are extracted and transported to experimental areas for bombarding fixed targets (possibly polarized targets). The degree of polarization is measured by polarimeters at several stages of the acceleration process: a 200 MeV polarimeter at the end of the linac, an internal polarimeter inside the main ring and an external high-energy polarimeter in front of experiments. The measurement of polarization is based on the asymmetry in the scattering of polarized protons through a thin target.

During acceleration in the main ring, the spin tune increases linearly with time and energy, and several depolarization resonances are crossed. Depolarization is reduced by correction devices of two types: dipoles correcting harmonics of the vertical closed orbit distortion when imperfection resonances are crossed, and pulsed quadrupoles driven by special power supplies enabling rapid jumping of intrinsic resonances. In the AGS five intrinsic and 31 imperfection resonances are crossed when accelerating up to 16.5 GeV. Final polarization is about 40% and represents nearly 60% of the polarization at injection into the main ring. In Saturne, only four intrinsic and six imperfection resonances are crossed when accelerating up to 3 GeV. A remarkably high degree of polarization, about 80%, is currently obtained.

The low intensity (\(10^{10}\) ppp) is the price to be paid for obtaining polarized protons. However, present developments in polarized-ion sources and injection techniques support
the hope that in the near future polarized-beam intensity will reach present values (> $10^{12}$ pps) of unpolarized-beam intensity.

The crossing of depolarization resonances is the main problem to study. In the following sections we investigate the mechanism of depolarization when crossing an isolated resonance and the cures for low-energy synchrotrons as well as the proposed "Siberian Snakes" at higher energies.

II.1 Depolarization resonances in proton synchrotrons

We consider only planar rings lying in a horizontal plane, as usual up to now. The spin closed solution is then vertical, i.e. parallel to the magnetic field in bending magnets.

For understanding depolarization phenomena in proton rings, the most important feature to consider is the effect of beam energy spread. As the spin tune is proportional to energy ($\omega = \frac{1}{\gamma}$ in a planar ring with vertical bending field), energy spread leads to spin tune spread. Spin vectors of particles with different energies precess at different rates and will rapidly get out of phase. For instance two 100 MeV protons, differing in total energy by $10^{-4}$, reach a spin phase shift of $2\pi$ after only 5000 turns.

Due to this spin phase mixing, any horizontal component of the polarization vector vanishes rapidly. Only its vertical component $P_z$ can survive. Consequently the beam must be injected in the ring with the polarization vector pointing in the vertical direction.

Now, any perturbation of spin motion, due to a perturbation of orbital motion, will, on resonance, lead to large deviations of the spin vector $\mathbf{S}$ away from the vertical Oz. The vertical component $P_z$ decreases and depolarization is observed. It is worth noting that amplitudes of orbital-motion perturbation may differ from one proton to another; the spin vector $\mathbf{S}$ will deviate more for large amplitudes than for small ones. An averaged amplitude must be taken for calculating the amount of depolarization.

The most important perturbations, and the resonances they drive, are of two types:

i) Vertical betatron oscillations are responsible for radial fields along the trajectories, which bend the spin vector away from Oz. They drive vertical betatron resonances, named intrinsic resonances. They are caused by the finite vertical emittance of the injected beam.

ii) Vertical closed-orbit distortions are responsible for radial fields also. They drive integer resonances, named imperfection resonances, and they are caused by field errors and magnet misalignments.

II.2 Resonance strength and width

As stressed in section II.6.2, the resonance phenomenon looks simpler when seen in a rotating frame. This frame rotates at the resonance frequency $\nu_R$, relatively to the orbit frame, about the vertical line.
In this frame, according to formula I.3.6, spin vector \( \mathbf{3} \) precesses about the vertical at the frequency \( \delta = (\nu - \nu_c) \omega_c \), where \( \nu_c \) is the cyclotron frequency in the field \( B \) of the ring magnets. On the other hand it also precesses about the stationary component \( \mathbf{B}_R \) of the perturbing radial field at the frequency \( \epsilon B_c \), where

\[
\epsilon = (1 + \gamma a) \frac{b_R}{B_c}
\]

is the resonance strength. Globally, spin vector \( \mathbf{3} \) precesses about the resulting rotation vector \( \mathbf{3} \): \( \mathbf{3} = (\delta \mathbf{3} + \epsilon \hat{1}) \omega_c \)

where \( \hat{1} \) is a unit vector in the direction of \( \mathbf{B}_R \). The deviation of \( \mathbf{3} \) from \( 0z \) is proportional to the resonance strength \( \epsilon \). The rotation vector \( \mathbf{3} \) is also the spin closed solution in the rotating frame.

On top of resonance the precession about the vertical vanishes (\( \delta \omega \)) and the rotation vector \( \mathbf{3} \) is transverse, parallel to \( \mathbf{B}_R \). The spin rotation frequency is \( \Omega = \epsilon \omega_c \), showing that the strength \( \epsilon \) is the ratio of the spin rotation angle \( \psi \) to the velocity rotation angle \( \alpha \):

\[
\epsilon = \frac{d\psi}{d\alpha}.
\]

In other words, the resonance strength \( \epsilon \) is the spin rotation angle per radian of velocity rotation, and is dimensionless.

The resonance strength \( \epsilon \) can also be considered as the resonance width since the angle of the rotation vector \( \mathbf{3} \) with \( 0z \) is larger than \( \pi/4 \) in the \( (\nu - \epsilon, \nu + \epsilon) \) spin tune interval.

A rough estimate of the resonance strength \( \epsilon \) can easily be obtained by considering only the radial quadrupole fields as given by formula (I.6.2):

\[
\mathbf{3}(\theta) = \frac{ab}{\beta^2} z
\]

where \( z \) is the vertical displacement due to either a vertical betatron oscillation for an intrinsic resonance or a vertical closed-orbit distortion for an imperfection resonance. The stationary component \( \mathbf{3}_R \) is the one-turn average of \( \mathbf{3}(\theta) \) in the rotating frame (noted \( \hat{1}, \hat{m}, \hat{n} \) frame with \( \hat{n} \) parallel to \( 0z \)), given by:

\[
b_R = \frac{1}{2\pi} \int \mathbf{3}(\theta + i\hat{l}) \cdot \mathbf{3}(\theta) d\theta
\]

in complex notation. The resulting complex expression of the resonance strength \( \epsilon \) is:

\[
\epsilon = (1 + \gamma a) \frac{b_R}{2\pi B_c} \int \hat{m} \cdot d\hat{l}
\]

where \( K \) is the quadrupole strength : \( K = \frac{1}{2\pi} \frac{ab}{\beta^2} \).

For imperfection resonances, the resonance strength \( \epsilon \) scales linearly with the vertical closed-orbit distortion and the total proton energy. It is in the \( 10^{-4} - 10^{-3} \) range in the AGS at Brookhaven, and would reach \( 10^{-1} \) at most in the Tevatron at Fermilab.
For intrinsic resonances, \( \epsilon \) scales as the square root of the vertical invariant emittance and of energy. It is in the \( 10^{-3} - 10^{-2} \) range in the AGS and would reach \( 10^{-1} \) at most in the Tevatron. The strength \( \epsilon \), considered here, is an average over all the betatron amplitudes in the proton beam. However, when considering a single particle, the strength depends on its betatron amplitude and will vary from one particle to another.

II.3 Linear crossing of an isolated resonance

For small resonance width as in the AGS, the distance in energy between resonances is very large compared to their widths. Each resonance, crossed during acceleration, can be considered as isolated.

Then a simple picture of spin-vector \( \vec{z} \) motion, when crossing an isolated resonance, can be obtained in the rotating frame again.

Far below the resonance energy, the rotation vector \( \vec{\Omega} \) is vertical and downward (\( \delta \ll \epsilon \)). When approaching the resonance, \( \vec{z} \) starts to deviate from \( \vec{0z} \), and becomes exactly horizontal (\( \delta = 0 \)) on top of the resonance. Above the resonance, \( \vec{z} \) moves symmetrically and becomes vertical, in the upward direction, at the end (\( \delta = \epsilon \)). Globally, the rotation vector \( \vec{\Omega} \) undergoes a complete reversal of direction when the resonance is crossed.

Does the spin vector \( \vec{z} \) of an individual particle, which precesses about \( \vec{\Omega} \), follow it during its reversal? If yes, the spin vector \( \vec{z} \), assumed vertical initially, will also be reversed as \( \vec{z} \) is, and there is an adiabatic spin flip. If all the particles do the same, the polarization vector \( \vec{f} \) is only reversed. Initially vertical, it becomes vertical again after crossing, but pointing in the opposite direction. There is no depolarization.

The adiabaticity condition for spin flip is a spin precession about \( \vec{\Omega} \) much faster than the motion of \( \vec{z} \) itself. More precisely, assuming a linear variation of energy with time, i.e. a linear variation of spin tune \( \nu \) with azimuth \( \theta \):

\[
\nu = \nu_0 + \alpha \theta,
\]

the crossing "time" \( \Delta \theta \) is about:

\[
\Delta \theta = \frac{2 \epsilon}{\alpha}.
\]

During this time the spin precession angle \( \psi \) is:

\[
\psi \approx \epsilon \Delta \theta = \frac{2 \epsilon^2}{\alpha}.
\]

The adiabaticity condition (\( \psi \approx 1 \)) can be written:

\[
\frac{\epsilon^2}{\alpha} \approx 1.
\]
On the contrary, for very fast crossing:

\[ \frac{\varepsilon^2}{\alpha} \ll 1, \]

the spin vector \( \vec{S} \) has not enough time for starting to move during the resonance crossing. In this case the vertical direction of \( \vec{S} \) is not changed. There is no change in polarization either, assuming fast crossing for all particles.

What happens between these two extreme cases? One expects an incomplete spin flip, with \( \vec{S} \) finally pointing in a non-vertical direction. Consequently the vertical component \( |S_z| \) of \( \vec{S} \), which initially was unity, has decreased at the end. The vertical component \( P_z \) of the polarization vector has also decreased and some depolarization has resulted.

A quantitative estimate of the final vertical component \( S_z \) compared to its initial value, is given by the Froissart-Stora formula:

\[ \frac{S_z \text{ final}}{S_z \text{ initial}} = 2 e^{-\frac{\varepsilon^2}{2\alpha}} - 1 \]

which includes the two extreme cases of adiabatic spin flip \( (S_z \text{ final} = -S_z \text{ initial}) \) and of fast crossing \( (S_z \text{ final} = +S_z \text{ initial}) \), as well as the intermediate cases.

The effect of a particular resonance depends on its strength \( \varepsilon \) as compared to the acceleration rate \( \alpha \). Moreover, for intrinsic resonances, particles with very small betatron amplitude will experience a weak resonance and their spin vector will not be reversed. On the contrary particles with large amplitude will experience a strong resonance and their spin will be reversed. For this type of resonance the amount of depolarization is given by an average over the betatron amplitudes among the particles.

As an example, Fig. 2 shows the variation of the polarization after crossing the imperfection resonance \( \nu = 3 \) in Saturne as a function of a dipole correction which changes the strength \( \varepsilon \) of this resonance. The observed maximum corresponds to a total compensation

![Graph showing polarization P after crossing the imperfection resonance \( \nu = 3 \) in Saturne versus correction of vertical closed-orbit harmonics.](image)
of its natural strength by the dipole correction. With a correction, either null or opposite in sign, the polarization has the opposite value, indicating a successful adiabatic spin flip. Such an almost perfect spin flip is observed when crossing five imperfection resonances and two intrinsic resonances in Saturne, explaining the high degree of polarization (about 80%) maintained during the acceleration cycle up to the top energy (3 GeV).

II.4 Cures for low-energy synchrotrons

Very often, crossed resonances have a strength which is harmful to polarization as the spin vector is bent away from the vertical upon crossing. The understanding of the depolarizing mechanism indicates four methods for reducing depolarization. Two of these methods have been successfully applied, for instance in the AGS synchrotron.

1) Decrease the resonance strength $\varepsilon$. This is a correction method (also named harmonic spin matching) which aims to cancel out the resonance strength. For imperfection resonances, this method uses several dipole correctors which control harmonics of the vertical closed-orbit distortion. By varying the cosine and sine components of the most important harmonics one can compensate the driving field of a particular resonance. In general this is done after orbit correction and the needed harmonic correction is sufficiently small for not causing any trouble to the closed orbit.

The signal for monitoring this correction is the polarization $P$ itself. Total correction is achieved when polarization after resonance crossing is maximum as shown in Fig. 3.

This method has been successfully used for correcting about thirty imperfection resonances in the AGS, the corresponding corrections being turned on successively during resonance crossing.

In principle a similar correction method, using quadrupole correctors, can be used for intrinsic resonances, but is limited to rather weak resonances as in the case of two non-superperiodic intrinsic resonances in Saturne.

2) Increase the crossing rate $\alpha$. This method (named resonance jumping) aims to realize a fast crossing during which the spin vector has no time for moving away from the vertical. This method is essentially designed for intrinsic resonances.

During acceleration the spin tune increases linearly with time. When approaching an intrinsic resonance, at time $t_0$, the vertical betatron tune $Q_z$ is abruptly decreased such that the resonance is crossed in a very short time. Thereafter, the initial betatron tune is restored more slowly.
This method is applied in the AGS for crossing four strong intrinsic resonances. A set of pulsed quadrupoles, powered by special power supplies, is used for decreasing the vertical spin tune by $\Delta q_z = 0.25$ with a risetime $\Delta t$ of 1.6 $\mu$s. The crossing rate $\alpha$, which is normally $3.10^{-4}$ per turn, is increased by two orders of magnitude such that the resonance is crossed in less than one turn. Figure 4 shows the polarization after crossing as a function of the time $t_0$ at which pulsed quadrupoles are fired. One observes a polarization maximum when the time $t_0$ is properly set for crossing the resonance during the rise time of the pulsed quadrupoles (the observed secondary maximum could be an artefact).

Finally, the other two methods aim to achieve complete adiabatic spin flip by either increasing the resonance strength $c$ or decreasing the crossing rate $\alpha$. They are not commonly used.

All these four methods seem to be limited to low-energy synchrotrons; the limit in energy may well be of the order of the AGS top energy. There are two reasons: i) the strength and width of resonances increase with energy (see section II.2) making them more difficult to compensate or to jump; ii) the number of resonances to be crossed increases linearly with energy, requiring higher efficiency for curing each resonance in order to obtain a useful degree of polarization at the top energy. One could imagine arriving at a complete spin flip for most resonances. However, they become wide and can overlap, and new harmful effects are expected when overlapping occurs.

II.5. "Siberian Snakes"

A very different method, which would work at higher energies, has been proposed for avoiding depolarization on resonance crossing. The idea is to equip the synchrotron ring with one or several magnetic devices, named "Siberian Snakes".

In principle a Siberian Snake rotates the spin vector $\mathbf{S}$ by a $\pi$ angle about an axis $\hat{u}$ lying in the horizontal plane of the ring.

The spin motion in this horizontal plane is illustrated in Fig. 5 for a ring equipped with a single Siberian Snake. Starting at the point 0 opposite the Snake, after one turn a horizontal spin direction (1) is transformed into the direction (4), which is symmetric to direction (1) with respect to the axis $\hat{u}$, as seen in the orbit frame. In particular the direction $\hat{u}$ at point 0 is transformed into itself and then coincides with the spin closed solution $\hat{n}$ at this point. This spin closed solution lies in the horizontal plane at any point in the ring. Moreover, the above symmetry property of directions (1) and (4) in the horizontal plane shows that they are connected by a $\pi$-rotation about the direction $\hat{u}$. The one-turn spin mapping is a $\pi$-rotation about the spin closed solution $\hat{n}$ and the spin tune is $1/2$. 
Fig. 5  a) Spin motion in the horizontal plane of a ring equipped with a Siberian Snake (SS).
   b) Successive horizontal spin directions seen in the orbit frame.
   (1) Initial direction at point 0 of a test spin.
   (2) Spin direction at the Snake entrance.
   (3) Spin direction at the Snake exit.
   (4) Final spin direction at point 0 again.

The essential feature of a ring equipped with Siberian Snakes is that the 1/2 spin tune is independent of energy contrary to the usual linear dependence. Then beam energy spread does not lead to any spin tune spread, and there is no spin phase mixing, at least for an ideal Siberian Snake producing an exact π-rotation for all particles. One can expect a large reduction of depolarizing effects.

Spin motion in a ring equipped with a snake is analogous to a well-known NMR phenomenon, named Spin Echo. In a Spin Echo "thought" experiment (Fig. 6), a nuclear magnetic substance is magnetized such that the magnetization vector \( \vec{M} \) precesses about a stationary field \( \vec{H}_0 \) in a transverse plane. Due to local field inhomogeneities, magnetic moments \( \vec{u}_1,2,3 \) of different nuclei 1,2,3 precess at slightly different frequencies. If they were aligned in the same direction originally, they spread out after and magnetization decreases. At time \( T \) a transient field is applied which rotates all the magnetic moments by \( \pi \) about the axis \( \vec{u} \). The fastest moment \( (\vec{u}_1) \), which was the former, becomes the latter after this π-rotation. Then at time 2\( T \) the magnetic moments are aligned together again and magnetization is restored.

Fig. 6  Scheme of an NMR Spin Echo experiment.
   a) precession of three magnetic moments \( \vec{u}_1,2,3 \) about magnetic field \( \vec{H}_0 \)
      with π-rotation at time \( T \).
   b) variation of magnetization \( \vec{M} \) with time \( t \).
In a ring equipped with a Snake, the spin vectors of a particle bunch with some energy spread, have exactly the same behaviour. The Snake plays the role of the transient field in the Spin Echo experiment. Starting at the opposite point 0 in the ring, with all the spin vectors aligned in the same direction, they will be realigned together after one turn in spite of their different precession frequencies. They have the same precession angle and spin tune whatever their energies are.

Now, another popular scheme is a two-snake ring, i.e. a ring equipped with two opposite Siberian Snakes which rotate the spin vector by π about two orthogonal and horizontal axes \( \hat{u} \), \( \hat{v} \). It has the theoretical advantage of a more stable spin closed solution, with an energy-independent vertical orientation: downward in one half-ring between the Snakes and upward in the other half-ring.

The one-turn mapping (Fig. 7) of a spin direction (1), lying in the horizontal plane, is just obtained by adding a second π-rotation about \( \hat{v} \) axis which transforms direction (4) into direction (5). This final direction (5) is opposite to the initial direction (1), showing that the one-turn mapping is effectively a π-rotation and that the spin tune is 1/2 again.

![Diagram](image)

**Fig. 7** a) Spin motion in the horizontal plane of a two-snake ring. b) Successive horizontal spin directions seen in the orbit frame.

1. Initial spin direction.
2. Spin direction at 1st Snake (SS1) entrance.
3. Spin direction at 1st Snake (SS1) exit.
4. Spin direction at 2nd Snake (SS2) entrance.
5. Final spin direction at 2nd Snake (SS2) exit.

In a two-snake ring one can study what happens when the particle energy coincides with one of previous resonances (\( \gamma a = k \) or \( k + 0 \)), i.e. when a perturbing field component is resonant with spin precession in the arcs of the ring.

The one-turn mapping is perturbed. The spin closed solution \( \hat{\mathbf{n}} \) slightly deviates from the vertical and the spin tune is not exactly 1/2.

During resonance crossing, instead of being reversed, the spin closed solution \( \hat{\mathbf{n}} \) only undergoes a transient excursion away from the vertical. The original vertical direction (upward in one arc) is restored after crossing. This is in contrast with the complete reversal in a ring without snakes.
If the evolution of \( \hat{n} \) is sufficiently slow, the spin vector \( \hat{s} \), which rotates about \( \hat{n} \), will adiabatically follow it in its motion, \( S_z \text{ final} = S_z \text{ initial} \), and there is no depolarization. Simulation shows that this is the case for a two-snake ring, even for strong resonances, confirming the initial idea of depolarization suppression by Snakes. However, for very strong intrinsic resonances \( (\varepsilon > 0.2) \) lack of adiabaticity has been observed in simulation, indicating that such depolarization resonances are still harmful.

Finally, how can a Siberian Snake be realized in practice? The simplest idea is to use a solenoid for obtaining a \( \pi \)-rotation about the longitudinal axis. According to formula (I.3.5) one needs a field integral of 3.752 Tm per GeV/c in the solenoid. For instance, for crossing the first strong intrinsic resonance \( \nu = 0.2 = 8.75 \) in the AGS with a Snake, the needed field integral amounts to 16.8 Tm. Obviously, when going to higher energies, the field integral becomes rapidly too large. One is forced to consider a Snake made of transverse field magnets. There is a large variety of possibilities. Figure 8 shows one attractive scheme with six vertically bending magnets and six horizontally bending magnets. According to formula (I.3.7), such a snake works at a nearly fixed field for maintaining the \( \pi \)-rotation at all energies. The 22 Tm overall field integral for the twelve magnets is modest, as compared to the bending-field integral needed in the arcs at high energy. The excursion of the vertical and horizontal beam bumps inside the Snake decreases when ramping in energy. This leads to a variable geometry of the beam line. This excursion is largest at the injection energy and the magnet aperture limits the lowest possible energy.

![Diagram](image)

**Fig. 8** Sketch of a Siberian Snake made of six bending magnets. Arrows indicate spin orientation and numbers indicate spin rotation angles.

In conclusion, Siberian Snakes are very attractive solutions for avoiding depolarization of proton beams during acceleration to high energies. However, there could still be an upper limit in energy, about 1-10 TeV, where depolarization resonances would become so strong that Siberian Snakes become inefficient.
III. POLARIZATION OF ELECTRONS IN STORAGE RINGS

The behaviour of electron beams in high-energy storage rings is very different from proton beams. This is due to synchrotron radiation which causes fluctuations and damping in particle oscillations. For instance, the emittance of an electron beam is fully determined by synchrotron radiation. On the contrary, emittance of a proton beam depends on its value at injection.

The same difference between electrons and protons appears in polarization. Protons must be injected polarized and final polarization is at most equal to its initial value. Electrons become transversely polarized in situ and do not need to be injected polarized. This is due to a polarizing effect of synchrotron radiation. On the other hand, fluctuations and damping induced by synchrotron radiation also causes depolarization which may even overcome the polarizing effect at high energies. Correction procedures are needed for reducing this depolarization. Finally, longitudinally polarized electrons are more interesting in colliding beam experiments, and some spin manipulation is needed for changing a transverse polarization into a longitudinal one at interaction points.

III.1 Sokolov-Ternov polarizing effect

Electrons, stored in a ring, radiate in the magnetic field of bending magnets. The synchrotron radiation power is given by the classical expression:

$$W_{cl} = \frac{2}{3} \frac{c^2 e}{\rho^2} \frac{m c^2 \gamma^4}{\gamma},$$

where $r_e$ is the electron classical radius and $\rho$ the bending radius of trajectory in the magnetic field.

There are also quantum aspects in synchrotron radiation. Their magnitudes depend on the ratio $h \omega_c / E$ of the emitted-photon critical energy ($\omega_c = \frac{3}{2} \frac{c}{\rho} \gamma^3$) to electron energy $E$. This ratio scales like $E^7 / \rho$ and is of the order of $10^{-6}$ for all the storage rings. Quantum effects are then small. However, their smallness can be compensated by the very high radiation rate. This happens for some spin effects involved in quantum emission of synchrotron radiation.

When an electron emits a photon, its spin state may either not change (non-spin-flip emission) or be reversed (spin-flip emission). Moreover, the probability of emission depends on the initial electron spin state: either "up" (↑), i.e. parallel to the field, or "down" (↓), i.e. antiparallel to the field. This gives an asymmetry in the radiated power between these two spin states.
The non-spin-flip asymmetry in the radiated power of the two states is small:
\[
\frac{W_{\uparrow\downarrow} - W_{\downarrow\uparrow}}{W_{\uparrow\uparrow} + W_{\downarrow\downarrow}} = 2 \frac{\hbar \omega}{E}.
\]
On the other hand the spin-flip asymmetry is quite large:
\[
\frac{W_{\uparrow\downarrow} - W_{\downarrow\uparrow}}{W_{\uparrow\uparrow} + W_{\downarrow\downarrow}} = \frac{35}{64} \sqrt{3} = 0.95,
\]
although its relative intensity is very low:
\[
\frac{W_{\uparrow\uparrow} + W_{\downarrow\downarrow}}{W_{\uparrow\downarrow} + W_{\downarrow\uparrow}} = \frac{4}{3} \left( \frac{\hbar \omega}{E} \right)^2.
\]
Consequently, the spin-flip transition from "down" state to "up" state is very rare and the "down" state gradually becomes more populated. The electron beam is polarized in the direction antiparallel to the magnetic field which in fact corresponds to the lowest energy state. (The electron magnetic moment is then parallel to the field). This polarizing mechanism is called the Sokolov-Ternov effect. In the same way a positron beam is polarized in the direction parallel to the field. However the polarization build-up rate is low as compared to the total radiation rate, nevertheless usually much faster than the rate of particle losses in the stored beam, and polarization can be observed if one waits a sufficient time after beam injection.

III.2 Polarization build-up in magnetic fields

Let us study the polarization build-up first in a storage ring with uniform magnetic field and then with a non-uniform magnetic field.

III.2.1 Uniform magnetic field

According to formula (1.2.2) the populations \( N_\pm \) of the two spin states can be written:
\[
N_\pm = \frac{N}{2} (1 \pm P), \quad \text{(III.2.1)}
\]
where \( N = N_+ + N_- \) is the total number of stored electrons and \( P \) is the degree of polarization.

The asymmetry \( A \) in the transition rates \( \lambda_\pm \) from these two spin states is written:
\[
A = \frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-},
\]
from where:
\[
\lambda_\pm = \frac{\lambda}{2} (1 \pm A), \quad \text{(III.2.2)}
\]
with \( \lambda = \lambda_+ + \lambda_- \).

Now, the rates of change in their population are given by:
\[
\frac{dN_+}{dt} = - \frac{dN_-}{dt} = N_+ \lambda_- - N_- \lambda_+.
\]
Using equations (III.2.1) and (III.2.2) one obtains:
\[
\pm \frac{d\mathbf{N}}{dt} = -\frac{N\lambda}{Z} (\mathbf{P} + \mathbf{A}) ,
\]
and one derives the polarization rate:
\[
\frac{d\mathbf{P}}{dt} = -\lambda (\mathbf{P} + \mathbf{A}) .
\]
By integration, assuming no polarization at time \( t=0 \), the evolution of polarization is:
\[
P(t) = A \left( e^{-\lambda t} - 1 \right) .
\]
This equation of evolution gives the ultimate degree of polarization:
\[
\left| P(\infty) \right| = A = \frac{3}{5} \frac{8}{\sqrt{3}} = 0.92376
\]
and the polarization time \( \tau_p \):
\[
\tau_p^{-1} = \frac{5}{3} \frac{\hbar}{E} \cdot \left( \frac{\hbar \omega_c}{E} \right)^2 \cdot \frac{W_c}{\hbar \omega_c} = \frac{5}{3} \frac{\hbar}{E} \frac{\lambda_c \gamma_e}{\rho} \frac{\sqrt{\gamma}}{\rho} ,
\]
where \( \lambda_c \) is the Compton wavelength divided by \( 2\pi \).

In a storage ring with straight sections, in which electrons do not radiate, the polarization time is longer by the ratio of the ring average radius \( R \) to the bending radius \( \rho \) in magnets. Numerically:
\[
\tau_p(s) = 98.66 \frac{\rho^2(m)}{E^5(\text{GeV})} \cdot R(m).
\]
The ultimate degree of polarization is high, although slightly lower than 100% due to the residual spin-flip probability from "down" state to "up" state.

The polarization time decreases very rapidly when energy is increased. This is due to the very fast increase of radiation rate which compensates the small intensity of spin-flip transition.

This polarization build-up by the Sokolov-Ternov effect has been observed in all the electron storage rings where it has been sought. The following table gives for some of these rings the maximum energy \( E \) at which polarization has been observed, the polarization time \( \tau_p \) and the measured degree of polarization \( P \).

<table>
<thead>
<tr>
<th>VEPP2-M</th>
<th>ACO</th>
<th>SPEAR</th>
<th>VEPP4</th>
<th>DORIS II</th>
<th>CESR</th>
<th>PETRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>E (GeV)</td>
<td>.625</td>
<td>.536</td>
<td>3.7</td>
<td>5</td>
<td>5</td>
<td>4.7</td>
</tr>
<tr>
<td>( \tau_p ) (min)</td>
<td>70</td>
<td>160</td>
<td>15</td>
<td>40</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>P %</td>
<td>90</td>
<td>90</td>
<td>&gt;70</td>
<td>80</td>
<td>80</td>
<td>30*</td>
</tr>
</tbody>
</table>

* after 120 min.
** after optimization
III.2.2 Non-uniform magnetic field

The polarization build-up has been theoretically studied in a general magnetic field configuration. The key-point is that the Sokolov-Ternov effect is a very slow process compared to spin precession about the field, and also compared to fluctuations in spin precession induced by quantum emissions and spin-orbit coupling. Therefore spin alignment by the Sokolov-Ternov effect can only, on a long time interval, appear along a direction which is stable against fast spin precession and its random fluctuations. The only such direction is the spin closed solution \( \hat{n}(s) \). Polarization is built up along \( \hat{n}(s) \) which then is the equilibrium spin direction.

At first, one can consider only the case where spin precession fluctuations give negligible effects on average, i.e. where the depolarizing effects produced are negligible. This situation is encountered either at low energies, far away from depolarization resonances, or at higher energies when the storage ring is sufficiently spin-transparent. A quantitative calculation gives then the expressions of the ultimate polarization degree \( P(\infty) \) and of the polarization time \( \tau_p \):

\[
P(\infty) = \frac{8}{5\sqrt{3}} \frac{\langle \rho^{-3} | \hat{n} \cdot \hat{n} \rangle}{\langle \rho^{-3} | \left( 1 - \frac{2}{g} \beta \hat{n} \right)^2 \rangle},
\]

\[
\tau_p^{-1} = \frac{5 \sqrt{3}}{3} c \kappa c \gamma^5 \gamma \langle \rho^{-3} | \left( 1 - \frac{2}{g} \beta \hat{n} \right)^2 \rangle,
\]

(III.2.5)

(III.2.6)
where $\hat{b}$ is a unit vector along the reference orbit and $\hat{b}$ a unit vector along the transverse field component. The brackets $\langle \rangle$ indicate an average along the ring circumference, and the absolute value of bending radius $\rho$ is taken for including all possibilities of field orientation.

One could imagine increasing the polarization by increasing $\hat{b} \cdot \hat{n}$ in the denominator of formula (III.2.5). However, $\hat{n}$ cannot be simultaneously parallel to $\hat{b}$ and $\hat{b}$ which are orthogonal. Therefore maximum polarization is obtained when $\hat{n} \cdot \hat{b}$ is large and $\hat{b} \cdot \hat{n}$ small. The latter cannot give more than a very few percent increase of polarization. In most cases non-uniform fields will then give a lower polarization than the maximum 92.4 $\lambda$ expected in a planar ring, as they are not parallel to the equilibrium spin direction $\hat{n}$ everywhere.

One notices in formula (III.2.5) that magnets in which the field is orthogonal to $\hat{n}$ do not contribute to the polarization. This is so because the spin-flip asymmetry in synchrotron radiation vanishes for a spin direction orthogonal to the field. However, spin-flip probability does not vanish and becomes depolarizing in these magnets.

Wiggles with alternating sign of magnetic field also give no contribution to polarization. However, asymmetric wigglers, with higher field of one sign, can give a large contribution due to the strong polarization dependence on the field intensity (the $\rho^{-3}$ factor in formula III.2.5). They will also speed-up the polarization (formula III.2.6) as foreseen in LEP at 50 GeV where the polarization time is too long due to the small field intensity in the arc magnets (Fig. 10).

### III.3 Resonant spin diffusion

The energy jump caused by a quantum emission of synchrotron radiation in a bending magnet excites a synchrotron oscillation and also a horizontal or vertical betatron oscillation if the horizontal or vertical dispersion does not vanish in that magnet.

In an ideally planar ring, only horizontal oscillations are produced, neglecting angular distribution of the emitted photons at high energies. The stored beam is flat, polarized in the vertical direction, and experiences only vertically bending fields. There is no perturbation of spin motion and the ring is completely spin-transparent.

However, in a real ring with imperfections and possibly with vertical bends, vertical oscillations are produced also and the equilibrium spin direction $\hat{n}$ is not vertical all around the ring. Then, the quantum excitation of vertical and horizontal betatron oscillations, as well as synchrotron oscillations, perturbs the spin-motion, via the spin-orbit coupling. The spin-transparency of the ring is destroyed.

Within a few damping times following a quantum emission, the excited synchrotron and betatron oscillations will disappear again, and the particle will be left with a modified spin orientation. This perturbation becomes very significant on resonance, i.e. when the spin tune fulfills a resonant condition (formula I.6.1) corresponding to a perturbing field driven by the excited oscillations.
The result of successive random quantum emissions is a resonant random diffusion of spin vectors away from the equilibrium spin direction. This spin diffusion competes with the spin alignment of the Sokolov-Ternov effect and leads to an equilibrium degree of polarization lower than the ultimate value given by formula (III.2.5).

It is worth noting that spin diffusion is analogous to particle diffusion in phase space induced by quantum emissions. Following this point of view, the Sokolov-Ternov effect looks like a damping mechanism analogous to the damping of particle oscillations. The equilibrium polarization is then analogous to the equilibrium beam emittance.

It must be emphasized that the discrimination between spin diffusion and spin alignment is only made possible by the very different characteristic times of quantum emission (<10⁻¹⁰s), oscillation damping (10⁻² - 10⁻⁷s) and polarization build-up (>10⁷s).

### III.4 Spin-orbit coupling vector and equilibrium polarization

Spin diffusion is characterized by the rate \( \dot{N} \) of quantum emissions and by the deviation:

\[
\frac{d}{dt} \frac{\Delta E}{E}
\]

of the spin vector \( \hat{S} \) away from the spin closed solution \( \hat{n} \), produced by one energy jump \( \Delta E \) and measured after damping of the excited oscillations. In a first order and linear approximation, this deviation is proportional to \( \frac{\Delta E}{E} \) and the deviation \( \hat{d}(s) \) per unit of energy variation is called the "spin-orbit coupling vector". It depends on the details of ring optics and, in general, varies with the azimuth \( s \) of the quantum emission around the ring.

Due to spin diffusion, the average decrease of spin component along the equilibrium direction \( \hat{n} \) is per unit time:

\[
\frac{1}{\Delta n} \frac{ds}{dt} = \frac{1}{2} N \left| \frac{\Delta E}{E} \right|^2
\]

or

\[
\frac{1}{\Delta n} \frac{ds}{dt} = \frac{11}{18 \tau_p} \left| \hat{d} \right|^2
\]

where the average \( < > \) is taken over the energy jump \( \Delta E/E \) and the azimuth \( s \) around the circumference. This is also the rate \( \tau_d^{-1} \) of depolarization by spin diffusion, which is in balance with the polarization rate \( \tau_p^{-1} \) of the Sokolov-Ternov effect, leading to an approximate polarization decrease:

\[
\Delta P/P = \tau_p/\tau_d^{-1} = \frac{11}{18} \left| \hat{d} \right|^2
\]

More precisely the equilibrium degree \( P \) of polarization is given by the Derbenev-Kondratenko formula:

\[
P = \frac{8}{5\sqrt{3}} \frac{\left< \left| n^{-3} \cdot \hat{b} \cdot (\hat{n} - \hat{d}) \right|^2 \right>}{\left< \left| r^{-3} \left[ 1 - \frac{2}{9} (\hat{B} \cdot \hat{n})^2 + \frac{11}{18} \left| \hat{d} \right|^2 \right] \right|^2 \right>}
\]

(III.4.1)

which replaces formula (III.2.5) when spin diffusion is taken into account.
The origin of the linear $\mathcal{d}$ term in the numerator of this formula (III.4.1) is not discussed here. It is normally a small term as the spin-orbit coupling vector must be much smaller than unity for obtaining a high degree of polarization.

In a small-amplitude linear model of orbital motion, the result of a first-order calculation of the spin-orbit coupling vector $\mathcal{d}(s)$ is:

$$
\mathcal{d}(s) = -\text{Im} \left[ (\hat{m} + i\hat{c})^* \left( \Delta_x + \Delta_{x,x} + \Delta_z + \Delta_{z,z} + \Delta_{y} + \Delta_{y,y} \right) \right],
$$

(III.4.2)

where $\hat{m}$ and $\hat{c}$ are two orthogonal unit vectors, solutions of the spin motion and orthogonal to the spin closed solution. $\Delta_{x,x,z,z}$ represent the contribution of the radial and vertical betatron and synchrotron oscillations respectively:

$$
\Delta_{x,x,z}(s) = \frac{(\nu + 1)}{2\nu(\nu \pm 1)} \frac{e^{i\Phi_{x,x,z}}}{\sqrt{\beta_{x,z}}} J_{x,x,z}(s)
$$

(III.4.3.a)

$$
\Delta_{x,z}(s) = \frac{(\nu + 1)}{2\nu(\nu \pm 1)} \frac{e^{i\Phi_{x,z}}}{\sqrt{\beta_{x,z}}} J_{x,z}(s),
$$

(III.4.3.b)

where $D,D'$ are the dispersion and its slope, and $\Phi_{x,x,z,s}$ is the phase of betatron or synchrotron oscillations.

The first factor in $\Delta$ is a resonant factor which follows from piling-up of successive perturbations of spin motion, turn by turn during oscillations. It causes the spin-orbit coupling vector to become very large whenever the beam energy is such that the spin tune $\nu$ approaches any of the values:

$$
\nu = k \mp Q_{x,z,s} \quad (k \text{ integer}).
$$

Depolarization mainly occurs in the vicinity of these betatron or synchrotron resonances.

The second factor (in brackets) in formula (III.4.3.a) expresses that betatron oscillations and the induced perturbations of spin motion are proportional to the dispersion at the azimuth where quantum emission occurs. This quantity is familiar for calculating the beam emittance. For vertical betatron oscillations, it differs from zero only in presence of vertical orbit distortions.

But the real keys for obtaining high polarization are the spin-orbit coupling integrals that appear in (III.4.3) as the last factor $J_{x,x,z}(s)$:

$$
J_{x,x,z}(s) = \int_{s}^{s+c} (\hat{m} + i\hat{c}) \cdot \hat{e}_z K \left[ \beta_{x,z} e^{i\Phi_{x,z}} \right] ds'
$$

(III.4.4)

$$
J_{x,z}(s) = \int_{s}^{s+c} (\hat{m} + i\hat{c}) \cdot \hat{e}_x K \left[ \beta_{x,z} e^{i\Phi_{x,z}} \right] ds'
$$

$$
J_{s}(s) = \int_{s}^{s+c} (\hat{m} + i\hat{c}) \cdot (\hat{e}_{x,x} + \hat{e}_{x,z}) K \beta_{x,z} ds' ,
$$
where \( c \) is the ring circumference, \( K(s') \) is the quadrupole strength and \( \hat{e}_{x,z} \) the unit vector in radial or vertical direction. For each type of oscillation, the corresponding integral is proportional to the effective spin rotation away from the equilibrium direction \( \hat{n} \) during one revolution around the ring, starting at the azimuth \( s \) of the quantum emission.

The scalar product \((\hat{m} + i\hat{\ell}).\hat{e}_z\) vanishes in a planar ring without magnet errors, where \( \hat{n}(s') = \hat{e}_z \) everywhere. In a real ring with vertical alignment errors of quadrupoles, however, the beam will be subjected to small vertical kicks which may cause the equilibrium spin direction to deviate from the vertical. In this case, the coupling integrals \( J_x \) and \( J_z \) do not vanish and give a finite contribution from radial betatron and synchrotron oscillations to depolarization.

Similarly, the vertical dispersion caused by vertical orbit distortions will give a finite contribution from vertical betatron and synchrotron oscillations to depolarization, since the coupling integrals \( J_x \) and \( J_z \) containing the scalar product \((\hat{m} + i\hat{\ell}).\hat{e}_x\) do not vanish in general.

A computer code (named SLIM) has been written for calculating the spin-orbit coupling vector \( \hat{C} \) and the equilibrium polarization in a ring with given imperfections and vertical bends. It uses an \( 8 \times 8 \) matrix formalism which includes spin motion in addition to transverse and longitudinal motions. The matrix formalism is based on linearization of the orbital motion for small amplitudes. This linear formalism cannot account for non-linear resonances.

In conclusion, the spin-orbit coupling vector \( \hat{C} \) becomes large on resonance where polarization is strongly reduced. It scales linearly with beam energy, since the spin precession increases faster than the particle velocity rotation when energy is increased (see the \( \gamma a + 1 \) factor in formulae III.4.3). Finally its amplitude is determined by the spin-orbit coupling integrals.

* In the integral \( J_z \), the synchrotron phase factor \( e^{\pm i\phi_s} \) has been neglected, assuming a very small synchrotron tune \( (Q_s \ll 1) \).
Figure 11 shows an experimental scan of polarization as a function of beam energy in the SPEAR storage ring, which shows several depolarization resonances, in particular nonlinear ones. Between these resonances polarization nearly reaches the maximum polarization (92.4%) for a planar ring.

![Graph showing polarization vs. energy](image)

**Fig. 11** Relative polarization ($P/P_0 = 92.4\%$) versus beam energy in SPEAR. Occurrence of resonance tune values ($k_{x}, k_{y}, k_{0_{x}}, k_{0_{y}}, k_{0_{z}}$) are indicated above the figure, in particular synchrotron satellites of $\omega_{8}$ and $\omega_{3}0_{x}$ resonances.

III.5 Spin matching

In a real ring, at high energies (> 5 GeV), the spin-orbit coupling vector $\mathbf{\hat{d}}$ is never negligible, even outside resonances, and polarization is lower than allowed by the Sokolov-Ternov effect. The main problem is to reduce this depolarization for obtaining a usefully high degree of polarization. Procedures, generally named spin-matching, have been invented for doing so, and one of them has been successfully applied in the PETRA storage ring.

III.5.1 Global spin-matching

In principle, according to formulae (III.4.2) and (III.4.3), $\mathbf{\hat{d}}$ vanishes if the five spin-orbit coupling integrals $J_{x}$, $J_{y}$, $J_{z}$, $J_{x}$, $J_{y}$ are made to vanish in every magnet. These integrals are proportional to spin rotation away from equilibrium spin direction $\mathbf{n}$ as produced by radial, vertical and synchrotron oscillations respectively in one revolution around the ring. When these integrals vanish, the spin orientation is again along $\mathbf{n}$ after each turn. The ring is perfectly spin-transparent and no spin diffusion can occur. All depolarization resonances are suppressed at the same time.

However, such a global spin-matching cannot easily be achieved in any situation. When spin-transparency is lacking due to imperfections distributed all around the ring, one has ten conditions to satisfy at each magnet in the ring (the real and imaginary parts of each integral must be cancelled out). The total number of conditions is ten times the number of
magnets. Moreover, the value of the $J_{2x}$ and $J_s$ integrals depend on the amplitudes of imperfections which are normally unknown. For all these reasons global spin-matching is not practicable in general.

Nevertheless, if lack of spin-transparency is only due to a few vertical bends, such as those of spin rotators (see section III.6), the spin-matching conditions degenerate into a much smaller number. They can be added to the other matching conditions of the beam optics and one only needs a few more focusing elements for satisfying them.

The spin-orbit coupling integrals depend on beam energy, since the $(\hat{n} + i\hat{e})$ complex vector in formulae (III.4.4) precesses proportionally to beam energy. Therefore spin-matching conditions are energy-dependent and spin-matching procedure must be repeated at each operating energy.

III.5.2 Harmonic spin-matching

In fact, the largest contribution to depolarization comes from the excitation of a few resonances closest to the spin tune, as this contribution is inversely proportional to the squared distance from spin tune to the top of the resonance. It would be enough to compensate these nearby resonances, and the number of conditions to fulfill would then be reduced. This is the only practical possibility for reducing depolarization due to ring imperfections and this is essential at high energies where this depolarization is large.

Similarly to the case of polarized protons (see Section II.2), the strength of a depolarization resonance is proportional to the spin rotation caused by a resonant component of the perturbing field. In the case of an electron ring with imperfections, this perturbing field is produced by the oscillations following quantum emissions and is proportional to the integrand $\omega_y(s)$ of the spin-orbit coupling integral $J_y(s)$ for each type of oscillation ($y = \pm x, \pm z, s$):

$$J_y(s) = \int_0^{s+c} \omega_y(s') ds'.$$

The resonant component is then obtained by a frequency analysis of the perturbing field amplitude $\omega_y(s)$. According to formulae (III.4.4), $\omega_y$ has a $2\pi(v \pm Q_y)$ phase advance per turn* and its frequency decomposition is:

$$\omega_y(s) = \sum_p \epsilon_{p, y} e^{2\pi i (v-p \pm Q_y) s/c}$$  (p integer)

$$\epsilon_{p, y} = \int_0^{s+c} \omega_y(s) e^{-2\pi i (v-p \pm Q_y) s/c} ds/c$$

* $(\hat{n} + i\hat{e})$ and $e^{\pm i\hat{n}_x \hat{z}}s$ have $2\pi v$ and $2\pi n_x \hat{z} s$ phase advance respectively, neglecting the $e^{\pm i\hat{s}}$ phase factor in $J_s$.)
Compensation of the $v = p \neq 0_y$ resonance consists of cancelling its strength $e_{p,y}$. One has only two conditions to fulfill for each resonance to be compensated, as $e_{p,y}$ is a complex quantity representing a perturbing field orthogonal to the equilibrium spin direction $\hat{n}$. This compensation can be realized experimentally by using correctors which counteract ring imperfections and which are chosen to act on the nearby depolarization resonances.

The concept of harmonic spin-matching has been extended to any type of harmonic correction used to optimize the degree of polarization. For instance the procedure successfully applied in PETRA is based on polarization optimization by varying some harmonics of the vertical closed-orbit distortion. As explained in section III.4, the spin-orbit coupling integrals $J_x$ and $J_y$ are sensitive to any deviation of the equilibrium spin direction from the vertical. Such a deviation is due to vertical closed-orbit distortion, and in particular to the distortion harmonics closest to the spin tune, since the perturbing field corresponding to these harmonics is nearly in phase with spin precession.

Experimentally eight orbit correctors have been used for varying the sine or the cosine component of the closest harmonics ($37^{th}$ and $38^{th}$ harmonics at 16.5 GeV). The degree of polarization was measured with a polarimeter and optimized by varying the amplitude of these harmonics components (Fig. 12).

![Fig. 12](image)

This variation in harmonic amplitude has no visible effect on the vertical closed-orbit distortion, as the $37^{th}$ and $38^{th}$ harmonics are far away from the vertical betatron tune ($Q_z = 23.3$), and have a very small amplitude.

This method needs a measurable degree of polarization at the beginning. For this reason the closed orbit must have previously been carefully corrected by the usual correction procedures.

![Fig. 13](image)
After applying harmonic spin matching at a certain energy, the polarization is largest (60–80%) in a small range around this energy (Fig. 13), and decreases when the energy is shifted, since the coupling integrals $J_{X,Z,s}$ are energy-dependent through the spin tune. Moreover, when approaching depolarization resonances residual values of these integrals still lead to large depolarization.

III.6 Spin rotators

The experimental study of electroweak effects in $e^+e^-$ and $e^+p$ collisions calls for longitudinally polarized electrons of both helicities at interaction points. The importance of polarization for testing interaction models at very high energies ($E_{\text{beam}} > 25$ GeV) enhances the interest for polarized beams in electron storage rings, which was poor at lower energies. However, electrons become transversely polarized, instead of longitudinally, by the Sokolov-Ternov effect. Several schemes (90° spin rotators) have been proposed for rotating the vertical polarization in the arcs of a storage ring into a longitudinal polarization at the interaction points.

Let us describe schematically the spin rotator which will be built for the ep HERA collider. It follows from a compromise between the geometrical and optical constraints of this ring in one hand, and the need for a high degree of polarization on the other. Again, the existence of a good compromise is an illustration of the relative freedom in spin manipulations allowed by the spin-orbit coupling (see Section I.4).

![Fig. 14 Sketch of a HERA mini rotator pair (arrows indicate spin direction)](image)

A pair of 90° spin rotators, of a so-called "mini rotator" type, will be installed around each of the interaction regions, which turns the spin vector, after leaving the arc, into the beam direction and then back into the vertical before entering the following arc. Each mini rotator consists of three horizontally bending magnets, interleaved with three vertically bending magnets which superimpose a vertical beam bump to the horizontal beam deflection (Fig. 14). A pair of such mini rotators is symmetric in the horizontal plane with respect to the interaction point, but antisymmetric in the vertical plane. Both helicities are obtained by inverting the sign of the vertical beam bump in each mini rotator.
The mini rotator is laid out for a chosen design energy (29.8 GeV) at which it simultaneously provides the correct horizontal beam geometry and spin rotation (Fig. 15). When varying the beam energy, horizontal beam geometry is maintained by ramping the horizontal rotator magnets in synchronism with all the other ring magnets. For keeping the spin direction vertical in the arcs and longitudinal at interaction points, one needs two parameters to vary. One is the amplitude of the vertical beam bump in the mini rotator; the other is the amplitude of a superimposed horizontal beam bump which vanishes at the design energy. The mini rotator can thus be operated in a range going from 27 GeV to about 35 GeV. The increase of energy from the lower to the upper limit of this range has the advantage of reducing the polarization time from 40 min to 12 min, and also to increase from 80% to 85% the maximum degree of polarization allowed by the Sokolov-Ternov effect.

![Fig. 15 Simplified sketch of the mini rotator geometry. H: horizontally bending magnets, V: vertically bending magnets (arrows indicate spin direction; modifications required by the head-on ep collision scheme and superimposed horizontal beam bump are not shown)]

In principle rotator pairs, which are antisymmetric in both the horizontal and vertical planes, would have the advantage of restoring the vertical spin direction in the arcs at any energy, since the spin transformation, in an antisymmetric rotator pair, is the identity whatever the energy. However, such antisymmetric schemes have the drawback of either a smaller maximum degree of Sokolov-Ternov polarization, or a more space-consuming geometry, and were therefore not chosen for HERA.

The most important property characterizing a rotator is its effects on polarization. The mini rotator as chosen results from a minimization of its depolarization, for an acceptable length of about 56 m.

The first effect is a reduction of the Sokolov-Ternov degree of polarization, as the spin direction is not antiparallel to the field in the rotator magnets (see formula III.2.5). For a given spin rotation in these magnets, the reduction is inversely proportional to the squared bending radius and scales like the inverse of the squared rotator length. The spin rotation angles in these magnets and their lengths have been chosen for minimizing this depolarizing effect, together with other minor aspects. Depolarization amounts to only 8% at 35 GeV for eight 56m-long mini rotators in HERA, justifying the name of "mini rotator".
The second effect is a breakdown of spin-transparency. The rotators generate non-vanishing spin-orbit coupling integrals and lead to large depolarization by spin diffusion. A global spin-matching procedure has been applied for restoring spin-transparency (see Section III.5.1). The relatively small length of the mini rotators permits having no focusing elements in them. Thus the arcs and the straight sections between rotators can separately be made spin-transparent. Only two conditions for the arcs and three conditions for the straight sections are required. For instance, one condition expresses the spin-transparency for horizontal betatron oscillation in the straight section between rotators of a pair where the spin vector is longitudinal. The overall spin rotation about the vertical, as produced by this oscillation, must vanish in this section. As only rotations about the vertical are involved here, spin and velocity rotations are globally proportional, and the slope of a betatron trajectory must be the same at the ends of this section. This condition is automatically realized for a sine-like betatron trajectory due to the optical symmetry about the interaction point (IP) when the straight section is symmetric with respect to this point. It suffices to impose this condition for a cosine-like trajectory.

A calculation (with the SLIM code) of the spin-orbit coupling vector $\mathbf{j}$ shows that the equilibrium polarization reaches 84% at 35 GeV for HERA with four rotator pairs and spin-transparent optics, and without imperfections.

III.7 Nonlinear depolarizing effects

Two nonlinear depolarizing effects are expected to be important at high energies. However, the information available on them is scarce.

III.7.1 Depolarization enhancement by energy spread

Beam energy spread increases quadratically with energy and reaches a r.m.s. value of about 50 MeV at 50 GeV. On the other hand, the spacing between depolarizing resonances is constant with energy and each particular type of resonance occurs repeatedly with a 440 MeV separation.

Particles with sufficiently large energy deviation, i.e. with large synchrotron amplitude, may then approach a resonance energy even if the central beam energy is set as far as possible away from nearby resonances. Larger depolarization than expected for small oscillations in linear theory can then occur.

However, this picture of a particle approaching a resonance is not very consistent as it mixes the time domain and the frequency domain. A more correct way of considering this energy spread effect is to take into account the frequency modulation of spin precession.
produced by synchrotron oscillations, as spin tune is proportional to energy. This situation is very similar to the well-known effect of frequency modulation in RF-waves. It leads to the appearance of satellites around the central frequency. Similarly for the spin motion, synchrotron satellites are generated around any depolarizing resonance. These satellites are regularly spaced by the synchrotron tune $\Omega_s$. If the energy spread is sufficiently large, involving large synchrotron amplitudes, then several satellites are excited and the energy range in which depolarization by a particular resonance occurs is widened. This is equivalent to saying that particles with large energy spread are approaching resonance.

Already at 3.7 GeV in SPEAR higher-order synchrotron satellites of a betatron resonance ($\nu = 3 + Q_x$) have been observed (see Fig. 11). More satellites and larger depolarization are expected at higher energies. Analytic models predict that, at 50 GeV in LEP, depolarization will be enhanced by about a factor five due to the energy spread effect. The only cure would be to reach higher efficiency in achieving spin-transparency by harmonic spin-matching.

III.7.2 Beam-beam depolarization

In storage ring colliders, beam-beam interaction is responsible for beam blow-up which limits the performances. This beam-beam interaction should also perturb spin motion, via spin-orbit coupling. Due to the non-linearity of the space charge field, non-linear depolarization resonances are expected to be excited. Beam-beam depolarization should particularly be important at the beam-beam limit.

Experimentally, over 70% polarization has been observed in high-luminosity $e^+e^-$ collisions at $2 \times 3.7$ GeV in SPEAR, and has been used for high-energy physics experiments at this energy. However, this polarization in collision mode could not be reproduced at a later stage of SPEAR development.

Polarization has also been observed in collision mode at 16.5 GeV in PETRA. Figure 16 shows the decrease of polarization near the beam-beam limit where the beams are blown-up.

Although the beam-beam depolarization mechanism is not well understood, these scarce experimental results support a moderate optimism for the future of $e^+e^-$ or ep experiments with polarized beams.
ACKNOWLEDGEMENTS

I am indebted to J.P. Koutchouk, B. W. Montague and K. G. Steffen for carefully reading the final draft of this paper and for making valuable comments.

BIBLIOGRAPHY

- For an elementary introduction to the concept of polarization in particle beams, see: J.S. BELL, Report CERN 75-11 (1975).

- For a general review on polarization in high-energy rings, see: B.W. Montague, Physics Reports, Vol. 113 (1984).


A.1.1 The polarization vector behaves classically.

In Quantum Mechanics any pure spin state $\psi$ of one particle is completely specified by a ket vector $|\psi\rangle$, following the Dirac notation. The two "up" and "down" spin states along the Oz axis form a basis for the ket vectors of a spin 1/2 particle. On this basis a ket vector $|\psi\rangle$ is represented by a column vector with two components $f$ and $g$, called a spinor:

$$|\psi\rangle = \begin{pmatrix} f \\ g \end{pmatrix}$$

where $f$ and $g$ are complex numbers, normalized such that $|f|^2 + |g|^2 = 1$.

The two basic "up" and "down" states are represented by the spinors $|\uparrow\rangle$ and $|\downarrow\rangle$ respectively. According to the Superposition Principle any spin state is a linear superposition of these basic states, represented by the linear superposition of their spinors:

$$|\psi\rangle = f |\uparrow\rangle + g |\downarrow\rangle$$

It is shown in Quantum-Mechanics textbooks that the three components of the unit vector $\vec{p}$ in the direction of the spin $\hat{S}$ for a pure state are given by:

$$\begin{align*}
P_x &= fg^* + f^*g \\
P_y &= i(fg^* - f^*g) \\
P_z &= |f|^2 - |g|^2
\end{align*}$$

as a function of the spinor components $f$ and $g$. In particular one can easily check that this formula gives the correct spin direction in the case of the "up" and "down" states.

Now, from quantum indeterminacy the measurement of the spin components $S_x$, $S_y$, $S_z$ along the three axes Ox, Oy, Oz will not always give the same result in general. If one considers their average values over many measurements, they are given by the three components of the spin operator $\hat{S}^{op}$ on the spin state $|\psi\rangle$:

$$\langle\psi|\hat{S}^{op}|\psi\rangle,$$

which is named the quantum average of the spin operator $\hat{S}^{op}$.

On the previous basis this operator is represented by three $2 \times 2$ matrices, proportional to the Pauli matrices $\sigma_x$, $\sigma_y$, $\sigma_z$, and acting on the spinor $|\psi\rangle$:

$$\hat{S}^{op} = -\frac{h}{2} \sigma$$

where $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ and:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

With the use of this basis the calculation of the quantum average $\langle\psi|\hat{S}^{op}|\psi\rangle$ is straightforward and the result is:

$$\langle\psi|\hat{S}^{op}|\psi\rangle = \frac{h}{2} \vec{p}.$$ (A.3)

Finally, according to the Ehrenfest theorem, such a quantum average must behave classically. It means that the evolution of the polarization vector $\vec{p}$ is governed by a classical equation of motion, like the Thomas-BMT equation introduced later in section A.1.3.
For a bunch of particles, the polarization vector, which is the statistical average of individual polarization vectors, has the same property.

For radiating electrons, which are in a mixed state, the polarization vector behaves classically too. However, it can only describe the spin state evolution after averaging over the possible spin-flip transitions. It cannot describe individual transitions.

A.1.2 Any spin state is fully determined by the polarization vector.

This statement means that the average value \( \langle A \rangle \) of any observable quantity \( A \), as obtained in a measurement, is completely determined by the polarization vector only. This average value \( \langle A \rangle \) involves a quantum average for each particle and an ensemble average over all particles of a bunch.

Firstly, the quantum average \( \langle \psi_k | A | \psi_k \rangle \) for a particle \( \# k \) in a pure state \( | \psi_k \rangle \) can be calculated with the basis used in the previous section, knowing the matrix elements \( A_{++}, A_{+-}, A_{-+}, A_{--} \) of the corresponding operator \( A_{\text{op}} \) between the basic states:

\[
\langle \psi_k | A | \psi_k \rangle = | f_{kk}^+ |^2 A_{++} + | f_{kk}^- |^2 A_{--} + f_{kk}^+ f_{kk}^- A_{+-} + f_{kk}^- f_{kk}^+ A_{-+} .
\]

The spinor components are then expressed in terms of the polarization vector components of the particle, given by formulae (I.2.3):

\[
\langle \psi_k | A | \psi_k \rangle = \frac{1}{2} \left[ (1 + P_{2k}) A_{++} + (1 - P_{2k}) A_{--} + (P_{xk} + iP_{yk}) A_{+-} + (P_{xk} - iP_{yk}) A_{-+} \right].
\]

Secondly, one has to take the ensemble average over all the particles in a bunch (or among all the possible states of a mixed state):

\[
\langle A \rangle = \frac{1}{N} \sum_k \langle \psi_k | A | \psi_k \rangle .
\]

The result:

\[
\langle A \rangle = \frac{1}{2} \left[ (1 - P_{2k}) A_{++} + (1 - P_{2k}) A_{--} + (P_{x} + iP_{y}) A_{+-} + (P_{x} - iP_{y}) A_{-+} \right], \tag{A.4}
\]

shows that the average value \( \langle A \rangle \) depends only on the components of the polarization vector.

This is a characteristic feature of spin 1/2 particles, which cannot be generalized to particles of higher spin, although a polarization vector can still be defined and plays an important role in the description of polarized states.

The result presented here is usually derived from a simple relationship between the polarization vector and the density matrix representing the spin state of a particle ensemble. For simplicity we have preferred to give a more elementary proof.

A.1.3 Equivalence of a quantum-mechanical description of spin precession.

Very often in the literature, spin precession is studied in a quantum-mechanical formalism instead of solving the Thomas-BMT equation (formula I.3.4). Then one has to solve the Schrödinger equation for the spinor \( | \Psi \rangle \), which characterizes the spin state of one particle:
\[ i\hbar \frac{d}{dt} \psi = H |\psi\rangle \]  

where $H$ is the Hamiltonian operator for spin motion.

The equivalence between both methods is obtained by verifying that the Thomas-BMT equation follows from the Schrödinger equation with a suitable choice of the Hamiltonian $H$.

From the Schrödinger equation (formula A.5) and from the conjugate equation:

\[ -i\hbar \frac{d}{dt} \langle \psi | H = \langle \psi | H |\psi\rangle , \]

one derives an equation for the quantum average of the spin operator $\hat{S}_{\text{op}}$:

\[ i\hbar \frac{d}{dt} \langle \psi | \hat{S}_{\text{op}} |\psi\rangle = \langle \psi | [\hat{S}_{\text{op}}, H] |\psi\rangle , \]  

(A.6)

where $[\hat{S}_{\text{op}}, H] = \hat{S}_{\text{op}} H - H \hat{S}_{\text{op}}$ is the commutator of these two operators.

The Hamiltonian $H$ is a scalar quantity, defined up to an arbitrary additive constant, and is an operator represented by a $2 \times 2$ matrix. Such a matrix must be a linear combination of the three Pauli matrices. The Hamiltonian $H$ can then be written:

\[ H = \hat{S}_{\text{op}} \cdot \vec{u} , \]

where $\vec{u}$ is a vector to be determined.

The commutator $[\hat{S}_{\text{op}}, H]$ is easily calculated by using the matrix representation (formula I.2.4) for $\hat{S}_{\text{op}}$ and knowing the commutation rules of Pauli matrices:

\[ [\hat{S}_{\text{op}}, \hat{S}_{\text{op}} \cdot \vec{u}] = i\hbar \vec{u} \times \hat{S}_{\text{op}} \]

The equation (I.3.10) becomes:

\[ i\hbar \frac{d}{dt} \langle \psi | \hat{S}_{\text{op}} |\psi\rangle = i\hbar \vec{u} \times \langle \psi | \hat{S}_{\text{op}} |\psi\rangle , \]

which is identical to the Thomas-BMT equation (I.3.4), when choosing $\vec{u} = \hat{S}_{\text{BMT}}$.

The Hamiltonian operator $H$ for spin motion is then:

\[ H = \hat{S}_{\text{op}} \cdot \hat{S}_{\text{BMT}} \]  

(A.7)

One can easily check that this Hamiltonian gives the correct magnetic energy (formula I.1.3) at the non-relativistic limit, through the Correspondence Principle.

Finally, it is worth noting that the equivalence between the Schrödinger equation and the Thomas-BMT equation, as shown here, is just a proof of the Ehrenfest theorem, which was invoked in section I.2.2, for the particular case of spin motion.
PARTICLE TRACKING

H. Mais, G. Ripken and A. Wrulich*

DESY, Notkestraße 85, D-2000 Hamburg 52

F. Schmidt

II. Physikal. Inst. der Univ. Hamburg, D-2000 Hamburg

ABSTRACT

After a brief description of typical applications of particle tracking in storage rings and after a short discussion of some limitations and problems related with tracking we summarize some concepts and methods developed in the qualitative theory of dynamical systems. We show how these concepts can be applied to the proton ring HERA.

1. Introduction

The aim of this chapter is to discuss some applications and limitations of particle tracking in storage rings\(^1\),\(^2\),\(^3\).

Although collective phenomena, as for example instabilities, are very important for accelerators we restrict ourselves to the single particle dynamics, i.e. we study the equations of motion of a single charged ultrarelativistic \((v = c)\) particle under the influence of external electromagnetic fields. In general, these equations are nonlinear. The main nonlinearities are due to the beam-beam interaction, due to nonlinear cavity fields or due to transverse multipole fields. These multipole fields are either introduced artificially e.g. by sextupoles which compensate the natural chromaticity or they occur naturally as deviations from linear fields due to errors. Since the beam-beam interaction will be treated in extra seminars we shall not consider it here. We shall also not consider effects which are induced by radiation such as radiation damping and quantum excitations which are very important for light particles like electrons and positrons. In proton storage rings these effects can approximately be neglected. The radiation losses of a proton in HERA for example are a factor \(10^{-7}\) less than the losses of the electron.

2. Hamiltonian description of the proton motion

The starting point for the proton dynamics is the following relativistic Lagrangian for a charged particle under the influence of an electromagnetic field described by a vector potential \(A(r, t)\)\(^4\)

\[
L = - m_0 c^2 \sqrt{1 - \frac{\dot{r}^2}{c^2}} + \frac{e}{c} \cdot \dot{r} A(r, t) .
\]

\(^*)\) Present address: SSC, LBL, Univ. Res. Assoc., University of California, Berkeley, USA.
Usually, one changes to a Hamiltonian description of motion and one introduces the curvilinear coordinate system depicted in Fig. 1.

\[ H = -(1 + \kappa x) \cdot \left\{ (1 + p_\perp)^2 - \left( \frac{e}{E_0} A_\perp \right)^2 \right\}^{1/2} - (1 + \kappa x) \left( \frac{e}{E_0} \right) A_\perp + (1 + p_\perp) \]  \hspace{1cm} (2)

with the corresponding equations of motion

\[ \begin{aligned}
\frac{dx}{ds} &= \frac{\partial H}{\partial p_x} ; \\
\frac{dp_x}{ds} &= -\frac{\partial H}{\partial x} \\
\frac{dz}{ds} &= \frac{\partial H}{\partial p_z} ; \\
\frac{dp_z}{ds} &= -\frac{\partial H}{\partial z} \\
\frac{d\sigma}{ds} &= \frac{\partial H}{\partial p_\perp} ; \\
\frac{dp_\perp}{ds} &= -\frac{\partial H}{\partial \sigma}
\end{aligned} \hspace{1cm} (3)

and \( \mathbf{A} = (A_\perp, A_x, A_z) \) satisfying Maxwell's equations.

By expanding the square root in equation (2) and the vector potential \( \mathbf{A} \) into a Taylor series various examples for nonlinear motion can be investigated.
Example 1: Nonlinear cavity fields

\[ H = \frac{1}{2} p_x^2 + \frac{1}{2} p_z^2 + \frac{g_0}{2} (x^2 - z^2) + \]
\[ + \frac{1}{2} \kappa x x^2 - \kappa x p_0 + V(s) \cos \sigma \]  

(4)

with \( g_0 = \frac{e}{E_0} \left( \frac{\partial B_z}{\partial x} \right)_0 \), \( V(s) \equiv \text{cavity voltage} \).

Introducing the dispersion function \( D \) defined by

\[ D'' = -(\kappa^2 + g_0) D + \kappa \left( \frac{d}{ds} \right) \]  

(5)

via the canonical transformation\(^5,6,7\)

\[ F_z = \frac{p_x}{D_X} (x - \frac{p_0}{D} D) + \frac{p_0}{D} D' x + \frac{p_0}{D} \sigma + \frac{p_z}{D} z - \frac{1}{2} DD' \frac{p_0}{D} z \]  

(6)

one obtains

\[ \bar{H} = \frac{1}{2} \bar{p}_x^2 + \frac{1}{2} \left( g_0 + \kappa^2 \right) \bar{x}^2 + \]
\[ + \frac{1}{2} \bar{p}_z^2 - \frac{1}{2} g_0 \bar{z}^2 - \]
\[ - \frac{1}{2} \kappa D \bar{p}_0^2 + V(s) \cos(\sigma + D \bar{p}_x - D' \bar{x}) \]  

(7)

If there is no dispersion in the cavity region (\( V(s) D(s) \equiv 0 \)) the synchrotron motion \((\sigma, p_0)\) is completely decoupled from the betatron motion \((x, p_x, z, p_z)\)\(^8\). In the case of a small dispersion one can write

\[ \bar{H} = \frac{1}{2} \bar{p}_x^2 + \frac{1}{2} \left( g_0 + \kappa^2 \right) \bar{x}^2 + \]
\[ + \frac{1}{2} \bar{p}_z^2 - \frac{1}{2} g_0 \bar{z}^2 - \]
\[ - \frac{1}{2} \kappa D \bar{p}_0^2 + V(s) \cos \sigma - \]
\[ - V(s) \cdot (D \bar{p}_x - D' \bar{x}) \sin \sigma \]  

(8)

Example 2: As a second example of nonlinear motion we consider the influence of transverse multipole fields with the following Hamiltonian:

\[ H = \frac{1}{2} p_x^2 + \frac{1}{2} p_z^2 - \frac{e}{E_0} A_t(x, z) \]  

(9)
The equations of motion are given by

\[\begin{align*}
x' &= p_x \\
px' &= e \frac{\partial A_r}{\partial x} = -\frac{e}{E_0} B_z(x, z) \\
z' &= p_z \\
p_z' &= e \frac{\partial A_r}{\partial z} = \frac{e}{E_0} B_x(x, z)
\end{align*}\]

with \( (B_z + iB_x) = B_0 \sum_{n=2}^{\infty} (b_n + i a_n)(x + iz)^{n-1} \).

The equations of motion in these two examples are highly nonlinear, and in general they cannot be solved analytically.

3. Dynamic aperture

One of the most important topics in accelerator physics one has to study is the dynamic aperture. This is an effective aperture of particle motion, beyond which the particle motion becomes unstable due to the nonlinear magnetic field. Figure 2 shows the ideal case where the dynamic aperture is almost the same as the physical aperture defined mainly by the size of the vacuum chamber.

![Dynamic aperture, physical aperture](image)

Among the questions for study are:

i) Is it possible to calculate and predict the dynamic aperture and how can this be done?
ii) How does it depend on the nonlinearities (multipole distribution, spatial distribution)?

iii) How does it depend on tunes? closed orbit distortions?

Tracking codes have been widely used to investigate these problems.

4. Particle tracking

The main idea of these codes is to track particles over many revolutions in a realistic model of the storage ring and to observe the amplitude of the particle at a special point $s_0$. Given the initial amplitude $\chi(s_0) = (x(s_0), p_x(s_0), z(s_0), p_z(s_0), c(s_0), p_y(s_0))$ one needs to know $\chi(s_0 + nL)$ ($L =$ circumference of the accelerator) for $n$ of the order of $10^5$ (corresponding to a storage time of a particle of about 10 hours in HERA). Different methods and codes have been developed to evaluate $\chi(s_0 + nL)$. Among others there are MARYLIE$^9$, TRANSPORT$^{10}$, RACETRACK$^{11}$ and PATRICIA$^{12}$. The last two codes are kick codes where the nonlinear elements are replaced by $\delta$-kicks according to:

$$a_{rm}(s) x^n y^m \longrightarrow \bar{a}_{rm} x^n y^m \cdot \delta(s - s_0) . \quad (11)$$

In all cases mentioned the problem is reduced to the study of nonlinear symplectic mappings of the form:

$$\chi(s_0 + nL) = \mathcal{I}(\chi(s_0 + (n - 1) \cdot L)) \quad (12)$$

or in shorthand notation

$$\chi(n) = \mathcal{I}(\chi(n - 1)) \quad (12a)$$

The dimension of the mapping (dimension of $\chi$) can vary from two to six according to the effects one has included (pure x- or z-motion, coupled betatron (x-z) motion, completely coupled synchro-betatron motion).

As an example for a kick code we briefly describe RACETRACK$^{11}$, which is a fast computer code to treat transverse magnetic multipole fields up to 20 poles. Several additional features, such as linear optics calculations, chromaticity adjustment, tune variation, orbit adjustment and inclusion of synchrotron oscillations are available. A schematic flow diagram is shown in Fig. 3.

Typical examples for the dynamic aperture of HERA obtained with RACETRACK are shown in Figs. 4 and 5$^3$ (four-dimensional coupled betatron case).

The main problems with tracking codes are the unavoidable rounding errors of the computers and the limited CPU-time. The rounding errors depend on the number system used by
Fig. 3 Flow diagram of RACETRACK

Fig. 4 Stable amplitude area

Fig. 5 Stable initial amplitude as a function of fractional tune
the compiler and they can destroy the symplectic structure of the nonlinear mappings. Thus, these rounding errors can simulate non-physical damping effects\textsuperscript{11}. In order to estimate the order of magnitude of these effects one can switch to a higher precision structure in the computer hardware or software and observe the differences. Another way is to compare the differences between forward tracking of the particle and backward tracking\textsuperscript{13}. The limited CPU-time restricts the number of revolutions one can track to about $10^6$ ($10^6$ revolutions in HERA with multipole errors require a CPU-time in the order of days on an IBM 3081 K).

Besides these technical problems there are also some physical problems related with the evaluation and interpretation of the tracking data. For example, fast instabilities with an exponential increase of amplitudes beyond a certain boundary can easily be detected whereas slow, diffusion like processes which become dangerous only after $10^5$ or $10^6$ revolutions are much more difficult to detect.

Nevertheless tracking is the only way to obtain realistic estimates for the dynamic aperture up to $10^5$ - $10^6$ revolutions, but it is very difficult to extrapolate these data to longer times ($10^9$ revolutions and more).

In order to get maximum information out of these numerical simulations and for a better understanding of the underlying physics one should also apply analytical (perturbation) methods\textsuperscript{14}. To understand how nonlinear systems might develop one should also know some of the results of the qualitative theory of dynamical systems.

5. Qualitative theory of dynamical systems

Although there are excellent review articles on this field\textsuperscript{14,15,16,17,18} we summarize some important results in order to make this chapter as self-contained as possible.

The reduction of a Hamiltonian system to a nonlinear mapping as done by tracking codes has been a well-known procedure since Poincaré (1890). Consider for example a two-dimensional Hamiltonian system without explicit time ($s$-) dependence $H(q_1,q_2,p_1,p_2)$. The corresponding phase space is four dimensional, and since $H$ itself is a constant of the motion the physically accessible phase space is three dimensional. Consider a surface $\Sigma$ in this three-dimensional space as depicted for example in Fig. 6.

The bounded particle motion induced by the Hamiltonian $H$ will generally intersect this surface in different points ($P_0 \rightarrow P_1 \rightarrow P_2 \rightarrow \ldots$). If one is not interested in the fine details of the orbit but only in the behaviour over longer time scales it is sufficient to consider the consecutive points $P_0 \rightarrow P_1 \rightarrow P_2 \rightarrow \ldots$ of intersection. These contain complete information on the Hamiltonian system. In this sense one has reduced the Hamiltonian dynamics to a mapping of $\Sigma$ to itself which is in general nonlinear (Poincaré surface of section technique). Similar mappings can also be derived for Hamiltonian systems with explicit periodic time-$(s)$-dependence (this is normally the case in storage rings).
Another important fact and, after the work of Chirikov\textsuperscript{18}), one of the few beacons among an otherwise still dense mist of diverse phenomena is the KAM-theorem (Kolmogorov, Arnold, Moser; see for example Ref. 14). We will only illustrate this theorem in the two-dimensional case and instead of concentrating on mathematical rigour we will discuss its physical implications. Consider first the bounded motion of a two-dimensional autonomous (no explicit t- (s-) dependence) Hamiltonian system which is integrable. Roughly speaking, an n-dimensional system $H(q_1 \ldots q_n, p_1 \ldots p_n)$ is integrable if there exists a canonical transformation to action-angle variables $(I_1 \ldots I_n, \theta_1 \ldots \theta_n)$ such that the transformed Hamiltonian depends only on the $n$ (constant) action variables $I_1 \ldots I_n$. For the considered two-dimensional case this implies that the motion is restricted to a two-torus parametrized by the two angle variables $\theta_1$ and $\theta_2$ as depicted in Fig. 7.
As surface of section one can choose the \((l_1 - \theta_1)\)-plane for \(\theta_2 = \text{const.}\). In this surface of section which may be chosen to be just the plane of the page the motion of the integrable two-dimensional system looks very simple.

During the motion around the torus from one crossing of the plane to the next the radius of the torus (action variable) does not change,

\[
I_1(n) = I_1(n-1)
\]

and the angle \(\theta_1\) changes according to

\[
\theta_1(n) = \theta_1(n-1) + \omega_1 \cdot T
\]

where \(T\) is just the revolution time in \(\theta_2\)-direction from one intersection of the plane to the next

\[
T = \frac{2\pi}{\omega_2}.
\]

Thus one obtains for an integrable system

\[
I_1(n) = I_1(n-1)
\]
\[
\theta_1(n) = \theta_1(n-1) + 2\pi \alpha I_1(n) .
\]  

(13)

The term \(\alpha\) is the so-called winding number. It is the ratio of the two frequencies of the system and it generally depends on \(I_1\). If \(\alpha\) is irrational the \(\theta_1(n)\) form a dense circle while if \(\alpha\) is rational the \(\theta_1(n)\) close after a finite sequence of revolutions (periodic orbit). Thus, there are invariant curves (circles) under the mapping which belong to rational and irrational winding numbers. What happens now if a perturbation is switched on, i.e. if

\[
I_1(n) = I_1(n-1) + \epsilon f(I_1(n), \theta_1(n-1))
\]
\[
\theta_1(n) = \theta_1(n-1) + 2\pi \alpha I_1(n) + \epsilon g(I_1(n), \theta_1(n-1)) .
\]  

(14)

In particular, can one still find invariant curves? The KAM-theorem says that this is indeed the case if the following conditions are fulfilled (together with some requirements of differentiability and periodicity for \(f\) and \(g\); for more details see for example Ref. 14):

i) The perturbation must be weak

ii) \(\alpha = \frac{\omega_1}{\omega_2}\) must be sufficiently irrational, i.e. \(|\alpha - \frac{p}{q}| > \frac{K(\epsilon)}{q^{2+\delta}}\).

Under these assumptions most of the unperturbed tori survive the perturbation although in distorted form.

The rational and some nearby tori however are destroyed, only a finite number of fixed points of the rational tori survive - half of them are stable (elliptic orbits
around this fixed point, half of them are unstable (hyperbolic orbits). The hyperbolic fixed points are the source of chaotic motion in phase space, i.e. motion which is extremely sensitive to the variation of initial conditions. The motion around the elliptic fixed points can be considered as motion around a torus with smaller radius and the arguments used till now can be repeated on this smaller scale giving rise to the schematic picture shown below.

![Diagram of perturbation of an integrable system]

**Fig. 8** Perturbation of an integrable system

Thus, the phase-space pattern of a weakly perturbed integrable two-dimensional system looks extremely complicated. There are regular orbits confined to tori and among them are distributed chaotic trajectories in a delicate manner. One should point out at this stage that there are no analytical methods for calculating these chaotic orbits - perturbation theories diverge.

6. *Studies of chaotic behaviour in HERA caused by transverse magnetic multipole fields*

Now we would like to present numerical results using RACETRACK with special emphasis on finding and investigating chaotic trajectories in phase space\(^{19,20}\). The calculations have been performed on a 370 E Emulator and the IBM 3081 K. The number of revolutions was varied between 30000 and 300000 using a HERA proton optics with a fixed realistic multipole distribution of the kind resulting from nonlinear field errors in the superconducting magnets.

At first, we have studied purely horizontal motion (i.e., without coupling to the vertical betatron motion) which of course leads to a two-dimensional nonlinear mapping. Fig. 9 shows a \(p_x-x\) plot of a particle trajectory near the dynamic aperture.

In an enlarged scale one clearly sees the island structure around elliptic fixed points and the chaotic (area filling) behaviour near the hyperbolic fixed points (see Fig. 10).

In this two-dimensional case the dynamic aperture could be identified with the largest existing KAM-circle. There exist well-known methods for investigating the break-up of these border lines\(^{21,22}\) whose disappearance with increasing perturbation would lead to
a kind of global chaos, a situation one naturally wants to avoid in storage ring physics. In addition, two-dimensional systems are special in that the existence of KAM-circles implies exact stability. Since chaotic trajectories cannot escape without intersecting these invariant surfaces, they are forever trapped between these tori if they indeed exist.

This is not true for higher dimensional systems where the KAM-theorem predicts three-tori \((S_1 \times S_1 \times S_1)\) in six-dimensional phase space, four-tori in eight-dimensional phase space etc.
Here chaotic trajectories can in principle always escape although their motion can be obstructed strongly by these tori. Chaotic regions can even form a connected web along which the particle can diffuse as has been demonstrated by Arnold for a special example (Arnold diffusion, see for example\textsuperscript{14}).

As a next step we consider the fully coupled $x - z$ motion in HERA under the influence of the nonlinear multipole fields. There are several possibilities for displaying four-dimensional phase space trajectories. The simplest way is to draw projections onto the different planes $(x,p_x)$, $(z,p_z)$, $(x,z)$, $(p_x,p_z)$, $(x,p_z)$ and $(z,p_x)$ but one can also use three-dimensional projections and colour to represent the fourth variable\textsuperscript{20}.

In this higher-dimensional case one cannot simply use the area filling property for distinguishing chaotic trajectories from regular ones, one needs some other characteristic features. One property of chaotic motion is the exponential separation of two phase-space points which initially have been close together. Formally this can be described by the characteristic Lyapunov exponent\textsuperscript{14})

$$\lambda = \lim_{t \to \infty} \frac{1}{t} \ln \frac{|d(t)|}{|d(o)|}$$

where $d(t)$ describes how the (Euclidean) distance between two adjacent phase space points evolves with time and $d(o)$ is the initial distance. Non-zero Lyapunov exponents are a quantitative measure for stochasticity of the considered trajectories.

Typical examples for regular and chaotic trajectories for HERA are shown in Figs. 11 to 22. We show the projections of these orbits onto the different planes.

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{fig11.png} \hspace{0.05\textwidth} \includegraphics[width=0.4\textwidth]{fig12.png}
\caption{Fig. 11 $P_z$ versus $z$ (regular trajectory) \hspace{0.5\textwidth} Fig. 12 $P_z$ versus $z$ (chaotic trajectory)}
\end{figure}
Fig. 13 $x$ versus $z$ (regular trajectory)

Fig. 14 $x$ versus $z$ (chaotic trajectory)

Fig. 15 $P_x$ versus $x$ (regular trajectory)

Fig. 16 $P_x$ versus $x$ (chaotic trajectory)

Fig. 17 $x$ versus $P_z$ (regular trajectory)

Fig. 18 $x$ versus $P_z$ (chaotic trajectory)
Figures 23 and 24 show how the distance between two adjacent phase space points evolves with time, first for a regular trajectory (linear increase) and second for a chaotic orbit (exponential increase).
7. Summary

Thus, HERA shows all the features which are characteristic for nonintegrable Hamiltonian systems. However, because of the possibility of Arnold diffusion the existence of tori does not imply global stability in the four-dimensional case (coupled betatron motion) contrary to the uncoupled case. Until now, these chaotic trajectories have been observed only near the dynamic aperture. However, it is not clear whether this is also true for the case of coupled synchro-betatron motion (six-dimensional mappings) and how relevant these chaotic regions are in practice. Further investigations in this direction and more computer experiments are certainly needed for a better understanding. In addition, the application of perturbation methods might be helpful in suggesting directions for further investigations and how to design these numerical experiments.

Recently interesting attempts have also been made to compare the theoretical and tracking predictions with machine experiments.

For future work it is also desirable to extend these investigations to include collective effects and spin effects. Promising attempts have been made already but many questions are still open.

Acknowledgements

It is a pleasure to thank our colleagues at DESY Dr. D.P. Barber, Dr. R. Brinkmann and Dr. F. Willeke for many helpful discussions.

* * *

References

2) E. Keil, CERN 84-01, 1984.
3) A. Wrulich, DESY HERA 84-07, 1984.
5) H. Mais, G. Ripken, DESY Report to be published.
8) A. Piwinski, these proceedings.
12) H. Wiedemann, PEP Note 220, SLAC, 1976.
20) F. Schmidt, private communication and PhD-thesis Univ. of Hamburg, to be published.
THE RADIOFREQUENCY QUADRUPOLE LINEAR ACCELERATOR

M. Puglisi
University of Pavia, Department of Theoretical and Nuclear Physics, Pavia, Italy

ABSTRACT

The seminar is aimed to give a comprehensive picture of an RFQ. After a short description of the accelerating structure the T-K expansion is treated and the fundamental formula for the potential is derived. The vane tips shaping, completed to first order is followed by the physics of the machine where the most important parameters are listed and illustrated. Since the RFQ is essentially a cavity resonator this topic has been given particular attention. Design and other technical considerations complete the picture, while in the last section the new ideas are briefly outlined.

1. INTRODUCTION

The RFQ is a linear accelerator for ions that uses electric fields to simultaneously focus, bunch and accelerate a beam of heavy particles.

While, in principle, the RFQ can accept, focus and accelerate to the decided energy any kind of charged particle this machine is particularly convenient for accepting and accelerating an intense, low velocity beam from a continuous dc injector. In this case the main advantages of the RFQ can be summarized as follows: small size, low voltage dc injection, compatibility with complex ion sources, bunching with high efficiency, high beam current capacity, high output beam quality, easy operation. Conversely, at high energy (2-3 MeV/amu) most of the above advantages become scarcely significant and the standard linacs are preferable.

Before going further we should recall that the RFQ was invented by Lapchinskii and Tepliakov in 1970, the first Russian test was in 1975. Later on the work began at Los Alamos (1978) and subsequently in many other places as for instance Berkeley, Brookhaven, CERN, Chalk-River, GSI, Frankfurt, Saclay, Tokyo. The so-called Proof of Principle "POP" was given at Los Alamos in 1980. Since that time many RFQ have been successfully realized and the interest in this machine is no longer limited only to the high energy physicist because many industries are now planning to use the RFQ for ion implantation, testing of materials, medical purposes. Finally the RFQ can play an important role as a heavy-ion accelerator for the inertial fusion.

2. THE ACCELERATING STRUCTURE

Basically an RFQ (Radio Frequency Quadrupole) is made by four equal electrodes symmetrically placed around the beam axis, excited by an appropriate radio frequency voltage and contained in a vacuum tank with highly conducting walls. Depending upon their shape the electrodes are named as
"vanes" or "rods" and in Fig. 2 a group of four idealized vanes is sketched in order to show the geometry of the arrangement and to give a rough idea of the vane tip shaping that is needed for creating the appropriate accelerating fields.

Actually the whole beam dynamics of the machine depends upon the shape of the vane tips and it is rather evident that the vane tips shaping of a physical machine will be determined by a compromise among many conflicting requirements. Nevertheless for the sake of clarity the physics of the machine will be discussed on the basis of the sketch already seen.
In Fig. 3 only a horizontal and a vertical vane have been sketched together with the appropriate reference axes. The peaks for each tip are the nearest points to the $z$ axis and the valleys are the points that, being on the coordinate planes, are the most distant from the $z$ axis. The distance between two adjacent peaks or valleys changes along the beam axis according to the beam dynamics and the structure is non-periodic.

![Schematic view of two adjacent vanes](image)

Fig. 3 Schematic view of two adjacent vanes

We consider now the assembly of four vanes and two planes normal to the $z$ axis passing through two adjacent peaks of the horizontal vane. The space limited by the two planes is called an elementary unit each made up of two cells. In Fig. 4 again one horizontal and one vertical vane are represented with the vertical vane rotated 90$^\circ$ into the same plane as the horizontal vane. The horizontal vanes are supposed to be held at a dc potential equal to $+V/2$ while the vertical ones are supposed to be held at a potential equal to $-V/2$.

![Two adjacent vanes are shown on the same horizontal plane. The potential between the vanes is sketched.](image)

Fig. 4 Two adjacent vanes are shown on the same horizontal plane. The potential between the vanes is sketched.
It is now rather evident that a positively charged particle passing through the first cell (with length equal to CL) will gain energy, while the same particle will lose energy in going through the second due to the potential along the beam axis.

If now we imagine exciting the four vanes with an RF voltage then the whole structure can be accelerating if the particle takes a half period of the RF voltage to pass through each cell. It should be noted that under the previous hypothesis when one cell is focusing on the vertical plane (and necessarily defocusing on the horizontal) then the following cell will be focusing on the horizontal plane while defocusing on the vertical. The net result can be focusing as predicted by the theory for the alternate gradient machines.

Since both the transverse and the longitudinal focusing critically depend on the shape of the vane tips, a detailed knowledge of the fields in the beam region is required. Unfortunately, employing the geometry already indicated, the calculation of the electromagnetic field distribution is very complicated and some simplifying hypotheses need to be sought. Actually, if one attempts to design a physically realizable (and useful) structure then any kind of calculation (even the simplest one) shows that the important part of the RFQ cross section has very small dimensions if compared with the free-space wavelength of the accelerating field. For that reason M. Kapchinskii and V.A. Tepliakov did the calculations assuming the electrostatic distribution for the field inside the beam region. The subsequent experience proved that this analysis is sufficiently accurate.

3. OUTLINE OF THE T-K EXPANSION

The distribution of the electrostatic field due to the vanes is normally known as the T-K expansion\(^1,2\). This is easily obtained solving the Laplace equation, written in cylindrical coordinates, with the technique of the separation of variables in our cylindrical reference frame. The z axis is the beam axis and the origin is located as in Fig. 3. The \( \psi \) coordinate is equal to zero on the positive section of the x axis.

If \( U = U(r,z,\psi) \) is the unknown potential then the Laplace equation is as follows:

\[
\frac{\partial^2 U}{\partial r^2} + \frac{1}{r} \frac{\partial U}{\partial r} + \frac{1}{r^2} \frac{\partial^2 U}{\partial \psi^2} + \frac{\partial^2 U}{\partial z^2} = 0. \tag{1}
\]

In order to separate the variables we assume that:

\[
U = f(r,\psi) \cdot \phi(z). \tag{2}
\]

Upon substituting Eq. (2) into Eq. (1) and separating we obtain the system:

\[
\begin{cases}
\frac{\partial^2 f(r,\psi)}{\partial r^2} + \frac{1}{r} \frac{\partial f(r,\psi)}{\partial r} + \frac{1}{r^2} \frac{\partial^2 f(r,\psi)}{\partial \psi^2} = \frac{\hbar^2 f(r,\psi)}{h} \quad \text{in } \psi \\
\frac{\partial^2 \phi(z)}{\partial z^2} = -\frac{\hbar^2 \phi(z)}{h^2} \quad \text{in } z
\end{cases} \tag{3}
\]

where \( h^2 \) is an arbitrary constant to be determined using the boundary conditions.
If we assume that \( h^2 \) is a positive number and that \( A \) and \( B \) do not depend upon \( z \) then
\[
\phi(z) = A \cos(hz) + B \sin(hz)
\]  
(5)

is a solution of Eq. (4). Because of the linearity of the previous equation then any linear combination of functions like (5) is a solution of Eq. (4) that could obey the boundary conditions.

At this point we assume that our structure is periodic. In this case the potential we are looking for should be a periodic even function of \( z \) with period equal to \( L \). It follows that:
\[
\Phi(z) = \sum_{n=1}^{\infty} A_n \cos \left( \frac{2\pi nz}{L} \right)
\]
(6)
is a solution of Eq. (4) that can fit the actual boundary conditions of the particular problem if the appropriate values are given to \( A_n = A_n (r, \psi) \).

From Eq. (6) the values of the separatrix constant are also known because \( h \) should be equal to \( 2\pi n / L \). In this continuation the quantity \( 2\pi L \) will be set equal to \( k \) (normally known as the phase constant). Now for each value of \( h \) we have a solution of Eq. (3). In order to satisfy the fundamental relationship already assumed, \( U = f(r, \psi) \). \( \phi(z) \) each solution of Eq. (3) should be multiplied by the corresponding \( \phi(z) \). Consequently the quantities appearing in Eq. (6) should be those functions which are solutions of Eq. (3); each solution \( A \) being determined by the eigenvalue \( nk \) that pertains to the corresponding \( A_n(z) \).

The method used above can be employed for solving Eq. (3). Again we separate the variables assuming the product solution:
\[
f(r, \psi) = R(r) \cdot \Theta(\psi)
\]
where \( R \) and \( \Theta \) are respectively functions of \( r \) and \( \psi \) only. Substituting and separating the variables we obtain:
\[
\begin{cases}
\Theta''(\psi) = -m^2 \Theta(\psi) \\
R''(r) + \frac{1}{r} R'(r) - \left[ \frac{(nk)^2}{r^2} + \frac{(m)^2}{r} \right] R(r) = 0
\end{cases}
\]
(7)
(8)

where \( m \) is a new arbitrary constant to be determined by the boundary conditions. If we assume that \( m \) is a positive number a solution of Eq. (7) is as follows:
\[
\Theta(\psi) = p \cos(m\psi) + q \sin(m\psi)
\]
(9)
where \( p \) and \( q \) depend neither on \( z \) nor on \( \psi \).
The electrical excitation of the structure (the vertical vanes are in parallel as the horizontal ones) requires that the potential $U$ should be a periodic even function of the variable $\psi$ with period equal to $\pi$. This means that to meet the above requirements we should have:

$$q = 0, \quad m = 2s, \quad s = 1, 2, 3$$

and, consequently, the $\Theta$ functions should have the form:

$$\Theta(\psi) = A_s \cos(2s\psi).$$

Again each value of $m$ must be substituted into Eq. (8) in order to obtain a radial function that depends upon both $n$ and $m$. The general solution will be written term by term assuming that $m$ ranges from 0 to $\infty$ and that for each value of $n$ the index $s$ can assume all the values that fit the boundary conditions.

In general the boundary conditions mentioned above can be summarized as follows:

1) For $r = 0$ the potential should remain finite.
2) For $kz = \pi/2$ and $kz = 3\pi/2$ each unit should exhibit a four pole symmetry. (The potential on the axis is equal to zero.)

Taking into account the above conditions for $n = 0$ the contribution $U_0$ to the total potential $U$ is as follows:

$$U_o = \sum_{s} A_s r^{2s} \cos(2s\psi)$$

where because of the four pole symmetry (independently of $z$) we must have:

$$J = 2s + 1 \quad s = 1, 2, \ldots$$

For $n \neq 0$, Eq. (8) is solved by the modified Bessel functions of order $2s$ and as a contribution to the total potential we obtain:

$$U_n = \left( \sum_s A_s I_{2s}(nk) \cdot \cos(2s\psi) \right) \cos(nkz)$$

where $(n + s)$ should be odd in order to fit the four pole symmetry that the structure periodically exhibits. (The Neumann functions are excluded because, as already said, the potential must remain finite for $r = 0$.)

Adding the various contributions we obtain the well known T-K expansion:

$$U(r, z, \psi) = U_o + \sum_{n} U_n$$

and the electrostatic problem is virtually solved.
4. THE VANE TIPS SHAPING

As was demonstrated in the previous paragraph the T-K expansion is very complicated and consequently if the vane tip geometry is assigned then a sufficiently accurate description of the electrostatic field in the beam region could be a very difficult problem. On the other hand an adequate study of the beam dynamics inside the machine can be done only if the field in the beam region is well known. A reasonable procedure\(^3\) for overcoming the problem can be to shape the vane tips in such a way as to coincide with the equipotentials described by a few terms of the T-K expansion. In other words we select a reasonable form for the potential in the beam region. When the potential is known the field is known, and the beam dynamics associated with the selected potential can be defined completely. If the calculated beam dynamics is satisfactory then the vane tips must have the form of the equipotentials that limit the beam region.

The simplest function that describes a potential consistent with a boundary condition of the beam region is obtained maintaining only the lowest-order terms of the T-K expansion. Accordingly, if we set \( n = 1 \) and \( s = 0 \) we find that the simplest form for the potential in each unit is as follows:

\[
F(r, z, \psi) = A_1 r^2 \cos(2\psi) + A_2 I_0(kr)\cos(kz)
\]  

(13)

where \( A_1 \) and \( A_2 \) are two constants to be defined and \( k = 2\pi/L \) is the phase constant. Since the four vanes are powered by an RF voltage with period \( T \) (radian frequency \( \omega \) equal to \( 2\pi/T \)) the complete form of the quasi-static potential is:

\[
U(r, z, \psi, t) = F(r, z, \psi) \cdot \sin(\omega t + \phi)
\]  

(14)

where \( \phi \) is the phase of the RF voltage when the charged particle enters the unit.

From the physical picture of the accelerator we know that the "synchronous" particle should pass through the unit exactly in one period of the radio frequency voltage. Consequently if \( \beta \) is the average normalized velocity of the synchronous particle we can write:

\[
k = \frac{2\pi}{L} - \frac{2\pi}{3cT} - \frac{2\pi}{3\lambda}
\]  

(15)

where \( \lambda \) is the free-space wavelength of the applied field.

\( A_1 \) and \( A_2 \) indicate two constants which depend on the geometry of the unit and this means that with this choice we can input only two boundary conditions. It should be noted that those boundary conditions are not as arbitrary as could be thought because once \( A_1 \) and \( A_2 \) are given the resulting structure should be physically realizable and electrically compatible.

Accordingly, with the scheme given in Fig. 3 our boundary conditions at \( z = 0 \) are as follows:
\[ U = \frac{V}{2} [ X \left( \frac{r}{a} \right)^{2 \cos(2\psi)} + A_{0} I_{0}(kr) \cos(kz) ] \]  

and it should be remembered that the inter-vane voltage \( V \) is equal to \( V_{0} \sin(\omega t + \phi) \).
Now we assume that the potential $U$ is the potential actually existing in one unit and we want to shape the vane tips in such a way as to realize this potential distribution. Once $a$, $m$ and $k$ are given, then the shape of the unit is uniquely determined and one of the simplest ways to arrive at the vane profile is to determine a reasonable number of vane cross sections along the $z$ axis. These cross sections can be determined by solving, numerically, Eq. (17) in which $U$ is made equal to the potential of the considered vane, the $z$ coordinate is an input for each cross section and a series of values is given to $\phi$. For each value of $\phi$ the corresponding value of $r$ is calculated. To calculate the profile of the cross section of a horizontal vane the potential must be set equal to $+V/2$. For a chosen value of $z = z_n$ we select a series of values for $\phi(0 < \phi < \pi/2)$ and find the corresponding values of $r$ solving Eq. (18):

$$1 = \frac{X r^2 \cos(\zeta \psi)}{a^2} + A_l(kr) \cos(kz_n)$$

(18)

An identical procedure has to be followed to find the cross section of a vertical vane. The potential must be set equal to $-V/2$ and $\pi/2 < \phi < \pi/2$. Consequently Eq. (19) is the one to be solved.

$$-1 = \frac{X r^2 \cos(\zeta \psi)}{a^2} + A_l(kr) \cos(kz_n)$$

(19)

The procedure outlined above cannot be followed blindly and some remarks are in order. A unit entirely generated using Eqs. (18) and (19) might lead to a physically unrealizable structure. This means that once the pole tip profiles are found then the remainder of the vanes have to be determined using different criteria. Moreover for $z = 0$ and $z = \beta \lambda$, the unit has identical cross sections while in an accelerating unit the initial and final sections are different. If the particles are not relativistic their velocity increases during the acceleration. This means that the distance travelled by the synchronous particle during one first half-cycle of the accelerating field should be shorter than the one travelled during the second half of the same cycle. The distance travelled by the synchronous particle during a half cycle of the accelerating voltage is called the unitary cell length and is obviously equal to $\beta \lambda/2$ where $\beta$ is the average normalized velocity of the transit particle.

A portion of a horizontal vane that contains two adjacent cells is sketched (exaggerating for the sake of clarity) in Fig. 5. The very nature of the RFQ accelerator requires that if, for instance, the cell "n" begins at a peak of a horizontal vane then the "n + 1" cell begins where a peak on the vertical vane occurs. This means that we could make use of the symmetrical expansion of a function (the vane profile) pretending that the distance $\beta \lambda$ is just twice the spatial period of a non-physically existing cell (with even-symmetry properties) that as regards the first half does coincide with the actual one. Consequently the shape of an actual cell can be determined with the same procedure outlined above but a new cell begins every time $z = \beta \lambda/2$. In other words once each value is assigned to the four parameters $k$, $a$, $m$, $\beta$, then the detailed procedure for finding the profile of the vanes is listed below.
Fig. 5 Portion of a horizontal vane that is part of two adjacent cells

1. The $p^{th}$ cross section of the horizontal vane of the $j$ cell is obtained by solving Eq. (18) written as follows:

$$1 = X \left( \frac{r}{a_j} \right)^2 \cos(2\psi) + A_1 \cos(k_z r)$$

where $z_p$ is the $p$ portion of the cell length $l_j$ and the $\psi$ coordinate is varied from zero to $\pi/4$.

2. The $p^{th}$ cross section of the vertical vane of the same $j$ cell is obtained from the same equation where $+1$ is substituted by $-1$, and the $\psi$ coordinate varies from $\pi/4$ to $\pi/2$ (again $k_j z$ will vary from 0 to $\pi$).

3. The next cell ($j + 1$) now begins with a radius $a_{j+1} m_{j+1}$ on the horizontal vane ($U = +\pi/2, \psi = 0$). On the other hand the cell length is now $l_{j+1}$ and this length should be divided into many intervals as above but the argument of the cosine should vary from $\pi$ to $2\pi$. This means that the quantity $\pi$ should be added to the argument of the cosine.

From the above arguments it is evident that there is no continuity between adjacent cells and the previous procedure should be modified in such a way as to obtain continuity between adjacent cells as shown in Fig. 5. For instance $a$ and $m$ can be made linear functions of $z$. On the other hand even the modified procedure could generate cells in which the curvatures along $z$ and in the $X$-$Y$ planes may create serious mechanical and electrical problems. In Fig. 6 the assembly of four physical vanes is sketched together with some cross sections of the whole machine.

The above procedure can be followed and the vane tip profiles exhibit a very complicated shape that should be machined with a high degree of accuracy. Moreover, especially at the low energy end, the electric field may be too much enhanced. A better mechanical solution is obtained if the pole tip cross-section may have a constant curvature radius. This can be achieved by introducing higher order multipoles into the potential function. Starting from an optimized two-term potential structure at every cell one can try a formula with more than two terms in order to minimize the
Fig. 6 Possible shape for the four vanes. Assembly in the container and cross sections.

deviation of the pole cross section from a circle of constant radius. Obviously many other important
manipulations can be done to cope with the particular performance required, but the procedure
outlined above remains substantially the same.
5. PHYSICAL CONSIDERATIONS

The lowest order potential function depends upon three parameters: a, m and k and we have seen that each cell is completely determined whenever the value of those parameters is specified. From the gradient (changed in sign) of the potential function (17) we obtain the fields inside the beam region as follows.

\[
\begin{align*}
E_r &= \frac{XV}{a^2} r \cos(2\psi) - \frac{kAV}{2} I_I(kr) \cos(kz) \\
E_\psi &= \frac{XV}{a^2} r \sin(2\psi) \\
E_z &= \frac{kAV}{2} I_I(kr) \sin(kz)
\end{align*}
\]

where \( V = V_0 \sin(\omega t + \phi) \) is the intervane voltage.

It is immediately clear, by inspection, that a particle travelling on the \( z \) axis does not experience any transverse force while for this particle (charge \( q \) and mass \( m_0 \)) the accelerating force \( f_z \) becomes:

\[
f_z = q \frac{kAV}{2} \sin(kz) - \frac{\pi q AV_o}{(2\pi)} \sin \left( \frac{2\pi}{2} z \right) \sin \left( \omega t + \phi \right).
\]

This means that the quantity \( X \) is related to the transverse focusing force while \( A \) is connected with the voltage gain per cell. In fact the potential difference \( \Delta U \) that exists on the axis between the beginning and the end of each cell is equal to \( AV \) as can be easily verified upon substitution.

\[
\Delta U = \frac{U}{2} - \frac{V}{2} \left\{ A - A \cos \left( \frac{2\pi}{2} \right) \right\} = AV
\]

Another important element for the design of the accelerator is the dynamic gain of energy per cell, \( \Delta E \). Taking into account that in an RFQ the accelerated particles are not relativistic, the motion of a particle travelling on the \( z \) axis is as follows:

\[
\frac{\dot{z}}{m_0} = \frac{kAV}{(2\pi m_0)} \sin(kz) \sin(\omega t + \phi)
\]

where \( \beta \) is the normalized average velocity of the transit particle.
It has been demonstrated (3) that under very broad conditions the above equation can be solved analytically and consequently the quantity $qE$ can be calculated with a high degree of accuracy. Nevertheless in a "normal" RFQ the relative variation of velocity per cell $\Delta \beta/\beta$ is always small and this means that the average and the instantaneous velocity of the transit particle are rather close.

In this case we can write $\omega t = kz$ and consequently the force $f_x$ becomes:

$$f_x = \frac{\pi qAV}{2\beta\lambda} \left[ \cos(\phi) - \cos(2kz + \phi) \right].$$

(23)

Integrating over the cell length we obtain:

$$\Delta E = \frac{\pi}{4} qAV \cos \phi.$$  

(24)

The numerical computations and the experience on the actually existing RFQ have proved that (24) is sufficiently accurate (in the above formula $\pi/4$ is the value of the transit time factor for a longitudinal field with space variations equal to sin kz).

In addition to Eq. (22), which describes the motion of a particle travelling on the $z$ axis, we need the equations for the motion in the transverse planes. While the general theory of the transverse motion is very complicated, it is rather easy to define the parameters which determine the stability of the beam on the transverse planes. For instance, using the expression of the gradients [Eq. (20)], we can write the differential equation of the displacement along $x$ as follows:

$$\ddot{x} = \frac{qE_x}{m_x} = \frac{q}{m_x} \left[ \frac{XV}{a^2} x - \frac{kAV}{2} I_1(kx) \cos(kz) \right].$$

(25)

The modified Bessel function $I_1(kx)$ can be expanded and for small values of $x$ we obtain $I_1(kx) = kx/2$. Substituting into Eq. (25) and using the explicit formula for $V$ and $k$ we obtain:

$$\ddot{x} = \left[ \frac{qXV}{a^2 m_x} \sin(\omega t + \phi) \right] x - \left[ \frac{\pi^2 AVq}{\beta^2 \lambda^2} \sin(2\omega t + \phi) \right] x.$$

Again we can make the hypothesis $kz = \omega t$ and after a little algebra we obtain:

$$\ddot{x} = \left[ \frac{qXV}{a^2 m_x} \sin(\omega t + \phi) \right] x +$$

$$- \left[ \frac{\pi^2 AVq}{2 m_x \beta^2 \lambda^2} \sin \phi \right] x +$$

$$- \left[ \frac{\pi^2 AVq}{2 m_x \beta^2 \lambda^2} \sin(2\omega t + \phi) \right] x.$$

(26)
Now in order to obtain dimensionless coefficients we multiply by \( T^2/\lambda \) both sides of Eq. (26) and change the variables as follows:

\[
\lambda x = \xi, \quad \lambda t = \tau .
\]

Upon substituting into Eq. (26) we obtain:

\[
\frac{d^2\xi}{d\tau^2} = -\left[ B \sin(2\pi \kappa_0 + \phi) + \Delta \right] \xi - 8 \xi \cos(4\pi \kappa_0 + \phi)
\]  
(27)

where, following the nomenclature used at Los Alamos,  

\[
B = X \frac{gV_0}{m_0c^2} \left( \frac{\lambda}{a} \right)^2 \quad \Delta = \frac{\pi^2 \frac{AVq}{2} \sin \phi}{\beta \gamma m_0 c^2}  
\]  
(28)

and \( \gamma = \sqrt{1 - v^2/c^2} \).

It should be noted that \( B \) and \( \Delta \) can be interpreted as normalized forces. \( B \) is responsible for the focusing effect while a defocusing effect corresponds to \( \Delta \) when \( \phi \) is negative, as is the case in a linac.

The solutions of the above equation can be convergent or divergent depending upon the numerical value of the parameters \( B \) and \( \Delta \). Since \( \Delta \) is always very small (typically \( \approx 0.05 \) at the injection) then a good degree of stability is obtained if \( B \) is larger than a few units and smaller than \( 15 \).

A more general analysis of the transverse stability will not be undertaken here because it requires the use of techniques too specialized for a general seminar on the RFQ. Nevertheless it should be emphasized that the general theory of the radial stability, valid for a linear accelerator, is applicable to the RFQ.

Before leaving the problem of the radial stability it is important to consider the role of the radius \( r_0 \) (the being the distance of the vane tips from the axis) that occurs for any cell when \( \cos (kz) = 0 \). In fact if the above condition is fulfilled then Eq. (17) reduces to:

\[
U = \frac{V}{2} X \left( \frac{r}{a} \right)^2 \cos(2\psi)
\]  
(29)

It follows that both for \( \phi = 0 \) and \( \phi = \pi/2 \) the distance \( r \) from the axis of the pole tips is as follows:

\[
r = r_0 = \frac{a}{\sqrt{X}},
\]  
(30)
This particular value of the radius is the so-called four-pole radius because on the plans for which \( kz = \pi/2 \) the vanes show perfect four-polar symmetry with hyperbolic cross sections defined by the equations:

\[
\begin{align*}
& r = \frac{r_0}{\sqrt{\cos(2\psi)}} & (0 \leq \psi \leq \pi/4) & \text{horizontal vane}, \\
& r = \frac{r_0}{\sqrt{-\cos(2\psi)}} & (\pi/4 \leq \psi \leq \pi/2) & \text{vertical vane}.
\end{align*}
\]

Moreover a very simple calculation can show that when the radius is equal to \( r_0 \), the radius of curvature at the pole tips is also equal to \( r_0 \) (on the X-Y plane).

Returning to the transverse focusing we observe that if \( V \) is constant, keeping the focusing strength at a fixed value requires that the quantity \( X/a^2 \) remains constant along the machine and this means [Eq. (30)] that the radius \( r \) should remain constant. Moreover a fixed value of \( r_0 \) can be expected to minimize variations in the vane-to-vane capacitance and should facilitate the design of an RFQ in which the pole tip voltage distribution is required to be flat over its entire length. For the above reason the quantity \( r_0 \) can be regarded as a characteristic average radius of the RFQ pole tips that affects all the design of the machine.

6. THE STRUCTURE OF AN RFQ

The accelerating and focusing fields depend upon the voltage applied to the four vanes. This is normally obtained via a cavity resonator where the vanes are a fundamental part of the whole structure. More specifically the RFQ cavity (vanes and container) should be designed for resonating, at the working frequency, in such a way as to create the desired voltage on the vanes. This is a problem that requires some knowledge of the microwave technique. Because the RFQ is essentially a radio frequency device where the microwave techniques play a major role it may be useful to give, in the following, a short outline of this topic.

An electromagnetic field that depends upon the time as a sine wave can exist and propagates inside a hollow cylindrical pipe with a perfectly conducting wall if certain conditions are met. A short way to permit calculations to be made is to assume that both the fields \( E \) and \( H \) depend upon \( z \) and \( t \) as follows:

\[
F(r,\psi,z,t) = f(r,\psi)e^{i(\omega t - k_0 z)}
\]  

(31)

where \( F \) and \( f \) stand both for \( E \) and \( H \) and \( \omega \) is the radian frequency of the field.

Moreover assuming that no currents are contained in the bounded volume then the Maxwell equation that we need can be written as follows:
\[
\begin{align*}
\text{and } & +\gamma E_y + j \omega \mu H_x = \frac{i}{r} \frac{\partial E_x}{\partial r} \\
-\gamma E_r & = 0 \\
-j \omega \epsilon E_r & = 0 \\
0 & -j \omega \epsilon E_y - \gamma H_r \\
0 & = -\frac{i}{r} \frac{\partial H_y}{\partial r}
\end{align*}
\]

Solving with the Kramer rule and using the normal notation:

\[
K_e^2 = \omega^2 \epsilon \mu + \gamma^2
\]

we obtain the transverse components of the field as functions of the longitudinal components \(E_z\) and \(H_z\):

\[
\begin{align*}
E_r & = -\frac{1}{K_e^2} \left[ \gamma \frac{\partial E_x}{\partial r} + j \omega \mu \frac{\partial H_y}{\partial r} \right] \\
E_y & = -\frac{1}{K_e^2} \left[ \frac{\delta}{r} \frac{\partial E_x}{\partial \psi} - j \omega \mu \frac{\partial H_y}{\partial r} \right] \\
H_r & = \frac{1}{K_e^2} \left[ -\frac{j \omega \epsilon}{r} \frac{\partial E_x}{\partial \psi} + \gamma \frac{\partial H_y}{\partial r} \right] \\
H_y & = -\frac{1}{K_e^2} \left[ \frac{j \omega \epsilon}{r} \frac{\partial E_x}{\partial \psi} + \frac{\delta}{r} \frac{\partial H_y}{\partial \psi} \right].
\end{align*}
\]

This means that if the longitudinal components of the field are known then the transverse ones can be obtained by derivation. Moreover if \(E_z\) is always zero then \(H_z = 0\) (because otherwise the whole field is zero) and we have the family of the so-called \(\text{TE}\) modes, where \(\text{TE}\) is an abbreviation for transverse electric mode. If \(H_z = 0\) and consequently \(E_z \neq 0\) we have the family of the \(\text{TM}\) modes (transverse magnetic).

Since the above system is linear then the superposition principle applies and any sinusoidal field can be reduced to a linear combination of \(\text{TE}\) and \(\text{TM}\) modes\(^1\). It is now important to

\(^1\) In the literature concerned with particle accelerators the \(\text{TM}\) modes are called "accelerating modes" while the \(\text{TE}\) modes are called "deflecting modes".
recognize that the structure containing the vanes should be excited with a TE mode. In fact only a TE mode can create the four-polar focusing field that, on the other hand, cannot be accelerating. The modulation on the vane tips introduces the local perturbation that is adequate for creating, locally, the accelerating field. Consequently we are naturally led to finding a possible solution for $H_z$ in the structure already described.

Under the previous hypotheses the Maxwell equations are as follows:

\[ \nabla \cdot E = 0 \]
\[ \nabla \cdot H = 0 \]
\[ \nabla \times E = -j\omega \mu H \]
\[ \nabla \times H = -j\omega \epsilon E \]

Taking the curl of the last equation, substituting $\nabla \times E$ and recalling that:

\[ \nabla \times \nabla \times H = \nabla (\nabla \cdot H) - \nabla^2 H \]

we obtain the familiar wave equation for the vector $H$. Since the same procedure applied to the curl of $E$ gives, formally, the same result we can write:

\[ \nabla^2 \begin{bmatrix} E \\ H \end{bmatrix} + \omega^2 \mu \epsilon \begin{bmatrix} E \\ H \end{bmatrix} = 0 \]  \hspace{1cm} (33)

Expanding the above equation and retaining the longitudinal $z$ component of $H$ we obtain:

\[ \frac{\partial^2 H_z}{\partial r^2} + \frac{1}{r} \frac{\partial H_z}{\partial r} + \frac{1}{r^2} \frac{\partial^2 H_z}{\partial \psi^2} + K_c^2 H_z = 0 \]  \hspace{1cm} (34)

Equation (34) can be solved with the same technique that has been used for the T-K expansion. We can assume that $H_z = R(r) \cdot \phi(\psi)$ where $R = R(r)$ is a function of $r$ and $\phi = \phi(\psi)$ is a function of $\psi$. Substituting and manipulating we obtain:

\[ \frac{r^2 R''}{R} + \frac{r R'}{R} + K_c^2 r^2 = \frac{\phi''}{\phi} \]  \hspace{1cm} (35)

The left side is a function of $r$ alone, the right of $\phi$ alone. Consequently if both sides are to be identical for all values of $r$ and $\phi$ then both sides must be equal to the same constant: for instance $\nu^2$ (assumed positive).
By substitution we obtain:

\[
\begin{cases}
R'' + \frac{1}{\rho} R' + \left( K_c^2 - \frac{\nu^2}{\rho^2} \right) R = 0 \\
\Theta'' = -\nu^2 \Theta 
\end{cases}
\]

The first equation is solved with the Bessel and the Neumann functions of order \(\nu\) whereas the second is solved with sinusoids.

For \(r = 0\), \(H_z\) cannot be infinite and this means that the Neumann function does not fit this boundary condition, on the other hand the field should be the same every time we vary \(\phi\) by a multiple of \(2\pi\). This means that \(\nu\) must be an integer. Moreover a proper selection of the origin for the coordinate will allow us to use either the sine or the cosine in the trigonometric part of the solution.

Consequently we obtain:

\[
H_z = H_0 J_\nu(K_c r) \cdot \cos(\nu \psi). \tag{36}
\]

The field described by Eq. (36) is parallel to the conducting wall and automatically obeys the boundary conditions. Conversely, from the second of the Eqs. (32) and using (36), we obtain \(E\) that, being parallel to the boundary, must be zero on the perfectly conducting walls. This means that on the boundary \((r = a)\) the derivative of \(J_\nu(K_c r)\) must be zero and we obtain:

\[
J_\nu'(K_c a) = 0 \tag{37}
\]

where \(a\) is the inner radius of the cylinder while \(J_\nu'(k_c a)\) indicates the value of the derivative of the Bessel function of order \(\nu\) for \(r = a\). Equation (37) determines the infinite series of the \(K_c\), and for each \(K_c\) we have a particular solution indicated as the \(TE_{\nu\lambda}\) mode. Specifically, \(\nu\) indicates the number of variations along \(\phi\) and \(\lambda\) indicates the order of the zero which determines the particular solution.

If we are looking for a transverse field with four-pole symmetry and no variations along \(z\) (as required from the assumed form of the potential) we have to set:

\[
\gamma = 0; \quad \nu = 2
\]

and from Eq. (37), selecting the first zero, we obtain:

\[
K_c a = \omega \sqrt{\varepsilon \mu} a = 3.05424 \quad \text{or} \quad f_c = \frac{145.8}{\xi} MH \tag{38}
\]

where \(f_c\) is the so-called cut-off frequency of the selected mode (Fig. 7). (In a waveguide the cut-off frequency is always the one for which \(\gamma = 0\).)
Fig. 7 Electric lines of force for TE mode in the cross section of a uniform cylindrical waveguide

This means that an infinitely long cylindrical lossless pipe, with inner radius equal to a, can support the axially uniform four-pole mode. The relationship between frequency and mode being defined by Eq. (38).

An infinitely long wave guide is not a practical device but it is possible to build a physical structure where, for a long portion of the axis, the field has four-pole symmetry and is adequately uniform (this structure is the RFQ resonant cavity where the cylindrical wall, the vanes and the end sections are fundamental parts of the whole structure).

In order to have some ideas about the cylindrical cavity resonators we imagine short circuiting, with a conducting wall normal to the axis, both ends of our hollow pipe leaving a clearance equal to L between the short circuits. Now, as well as the above conditions on the cylindrical wall [Eq. (37)], the electric field of a TE mode should be zero on the short circuiting surfaces (which are parallel both to $E_r$ and to $E_\theta$) and we have a third condition that enters into the determination of the cavity resonant frequency. It is nearly obvious that this condition is fulfilled if the distance L is an integer multiple of the half wavelength of the field as measured inside the pipe.

Let us call $R_{zL}$ the value of the argument that satisfies Eq. (37). Consequently we have:

$$\gamma^2 + \omega^2 \mu \varepsilon = \left( \frac{R_{zL}}{a} \right)^2$$

and we see that $\omega$ and $\gamma$ can be given any value consistent with Eq. (39).

In order to build up a stationary field we need propagation in both directions of the z axis. This means that $\gamma$ should be imaginary and we put $\gamma = j\delta$. If $\lambda_0$ is the wavelength inside the pipe (the so-called guide wavelength) then it is rather obvious that $\delta = 2\pi\lambda_0$. In fact when we pass through a distance equal to $\lambda_0$ the field has to repeat itself because the argument of $e^{j\delta z}$ changes by $2\pi$. 
The degeneracy already seen allows zero value for the last index of a TM mode. This cannot happen for a TE mode because the transverse electric field must be always zero on the short circuiting walls at the end of the cavity. Therefore if no variations are allowed along \( z \) (last index equal to zero) then the whole field should go to zero.

Up to this point we considered only the elementary cylindrical resonator where the fields \( E_z \) or \( H_z \) are completely described with only their eigenfunctions and the resonant frequency is the corresponding eigenvalue. However the cavity resonators used as accelerators, even maintaining the cylindrical symmetry, are often much more complicated and, in order to satisfy the boundary conditions dictated by a technical resonator, the complete set of the cylindrical eigenfunctions is normally required.

Substituting in Eq. (37) and recalling that \( \omega^2_{eM} = (2\pi/\lambda)^2 \), where \( \lambda \) is the free space wavelength of the field, we obtain:

\[
\left(\frac{2\pi}{\lambda}\right)^2 = \left(\frac{R_{xy}}{a}\right)^2 + S^2 = \left(\frac{R_{xy}}{a}\right)^2 + \left(\frac{2\pi}{\lambda_y}\right)^2.
\]

Rearranging and introducing the third condition that the resonator length \( L \) can be equal only to an integer number, say \( p \), of half guide wavelengths we obtain:

\[
\lambda = \frac{2L}{\sqrt{p^2 + (2L)^2 \cdot (R_{xy}/2\pi a)^2}}
\]

where \( \lambda \) is the free space wavelength for a cylindrical resonator of radius \( a \) and length \( L \) operating in the \( TE_{\nu\lambda p} \) mode.
At this point a very short outline of the TM modes for a cylindrical cavity seems in order. Equation (33) can be solved for \( E_z \) and following step by step the outlined procedure we obtain:

\[
E_z = E_o J_v(K_c r) \cdot \cos(\nu \psi) .
\]  

(42)

The \( E_z \) component is, by definition, parallel to the perfectly conducting wall and consequently \( E_z \) must be zero for \( r = a \). This condition is verified if:

\[
J_v(K_c a) = 0 .
\]  

(43)

Now it is rather evident that since \( E_z \) is always normal to the short circuit at the ends of the cylindrical cavity then an infinite series of modes can exist with no variations along \( z \) (TM_{v0} modes). As a consequence it happens that a cylindrical cavity can support any TM_{v0} mode independently of its length.

In addition to the above degenerate modes, a cylindrical cavity can exhibit a TM resonance if the cavity length \( L \) is equal to an integer number of half wavelengths measured inside the cavity. Again, following the procedure outlined for the TE modes, we find that the resonant wavelength of a TM mode is given by the formula (41) where now \( R_{v0} \) is the zero of order 1 of the Bessel function of order \( v \). In Fig. 8 examples of resonant modes are illustrated.

Even a simple outline of the general theory would go beyond the purposes of this rather intuitive treatment. Many powerful computer programs are now available for analyzing, with good accuracy, practically any useful cylindrical resonator\(^5\).

The cavity for an RFQ originates from a TE_{211} cylindrical resonator which is loaded with four V-shaped vanes symmetrically connected to the cylindrical wall as shown in Fig. 9. The vanes terminate at some distance from the short circuiting wall and consequently the central vane section is symmetrically coupled to the two end sections. (We should observe, in passing, that this resonator is no longer uniform along the abscissa.) The boundary conditions provided by the end sections allows the whole cavity to resonate in a very complicated manner where the fields are nearly uniform inside a large portion of the vane section. More specifically this condition is obtained if the TE_{21} cut-off frequency of the uniform guide represented by the vane section is slightly below the operating frequency of the whole cavity.

Fig. 9 Simplified axial and longitudinal cross sections of an RFQ
Figure 10 shows one of the four pole sections of the BNL RFQ. Since in this machine the focusing force is held constant then it follows that all the four pole sections should be equal.

As explained above, the choice of "constant r_0" minimizes the difference of the electrostatic capacity between different portions of the same structure but those differences are non-vanishing. This means that some "distributed tuning" along the structure would help in obtaining the good uniformity of the field that is really needed. For this reason each vane has been loaded with two bars tapered along the z axis and, in fact, a very good uniformity of the field has been achieved after a careful adjustment of the bars.

The "distributed" tuning mentioned above eliminates the field distortions that a series of lumped tuners would certainly introduce. Nevertheless the vane tips modulation, the unavoidable tuners at the end sections, the devices for feeding the power, and many other mechanical complications, always make the spurious modes which are near the wanted TE_{211} very strong. Figure 11 shows the electric lines of force between two adjacent pole tips of an RFQ for the quadrupole (TE_{211}) and dipole (TE_{111}) modes.

In a uniform cavity the dipole mode is always below the quadrupole mode and the same should happen in a well balanced RFQ cavity. In this case, as the whole cavity should be tuned just above the cut-off frequency of the guide corresponding to the vane section of the cavity, then it follows that dipole mode is enhanced. For this reason mode suppressing special techniques are required.
Fig. 11 Lines of force of the electric field between adjacent pole tips for the $\text{TE}_{211}$ and $\text{TE}_{111}$ mode

7. DESIGN AND TECHNICAL CONSIDERATIONS

A full technical description of the machine, together with practical design considerations would go beyond the purposes of this seminar. Nevertheless some of the problems concerning the whole RFQ will be illustrated in order to improve the general picture of the machine.

7.1 Tuning and excitation of the cavity

Figure 12 shows an idealized section of an RFQ where the horizontal vanes have been removed for simplicity. From the drawing it is evident that the electrodes placed on the end sections load both ends of each vane with an adjustable capacity to ground that greatly helps in balancing the vanes and tuning the whole cavity.

**TERMINATION OF UNIFORM FIELD REGION**

Fig. 12 Schematic axial section of an RFQ. The tuners placed on the end section are shown.
Any cavity resonator can be fed in many ways. Electrodes capacitively coupled with the vanes and connected to the RF generator are not favoured because they tend to arc in case of temporary mismatch (poor vacuum, multipacting, detuning ...). The loop coupling, with one or more excitation loops placed near the end sections and coupled with the magnetic fields which exist between the vanes, is much more used.

In addition to the "lumped" devices for coupling to the RF power source, many other "distributed" coupling methods can be used. These methods, which are well known and widely used in microwave techniques, have been used at Los Alamos since the beginning and later on were adopted in many other laboratories. The solution proposed by Los Alamos is for a large portion of the RFQ cavity to be symmetrically inserted into a shorter cylindrical cavity so that the new structure can be considered as a coaxial cable shorted at both ends, where the surface of the inner conductor coincides with the outer boundary of the RFQ cavity. A coaxial cable shorted at both ends resonates, in a transverse electromagnetic mode, when its length is equal to half the free-space wavelength of the exciting RF field. If this cavity is made exactly equal to \( \lambda/2 \) and some coupling slots are opened on the outer wall of the RFQ cavity, then the excitation of the coaxial resonator also excites the RFQ cavity. This coaxial cavity which matches the RF power generator to the RFQ cavity is known as the coaxial manifold.

![Fig. 13 Example of coaxial manifold (Los Alamos)](image)

It should be noted that this technique allows the RFQ cavity to be excited from many positions uniformly distributed along the outer wall of the machine. Moreover this distributed excitation is obtained without introducing electrodes in the regions between the vanes as shown in Fig. 13.
7.2 Suppression of the spurious modes

As was seen in the previous paragraph, an ideal resonant cavity can oscillate in an infinite number of modes. Actually an RFQ always exhibits a large number of strong resonances that very often are randomly bunched into very small intervals of frequency. Those modes reduce the amount of power that could excite the required one (the TE$_{211}$) and, by distorting severely the field, impair the calculated beam dynamics. Particularly dangerous is the dipole mode already seen.

For the above reasons many useful devices have been invented in order to eliminate as many spurious modes as possible, at least in the neighbourhood of the working frequency. Two different methods will be quoted here to give an idea of the problem.

With the first method the vanes of the same polarity are electrically connected with conducting rings as shown in Fig. 14. It is interesting to note that the same technique was successfully used at the dawn of the microwave tubes when the eight resonant cavities of the magnetron were synchronized by connecting "with a conducting wire" the homologous edges of two adjacent cavities (the so-called strapped magnetron).

![Fig. 14 Technique for mode suppression (Berkeley)](image)

The second method proposes the insertion of loops coupled with both modes TE$_{211}$ and TE$_{111}$. The loops are connected in such a way as to short circuit the TE$_{111}$ mode while allowing the existence of the TE$_{211}$ mode. A practical device based on this criterion was realized at BNL where the RF power is fed to the RFQ through two groups of loops (four for each group) placed inside the two end sections of the machine. The eight loops are excited in parallel and are coupled to the H field that exists among the vanes. By selecting the proper orientation for each loop it is possible to short circuit the TE$_{111}$ mode. Figure 15 shows a picture of the power splitter connected with the eight coaxial cables.

7.3 Design considerations

The operating frequency is a very important design parameter. Since the cell length is equal to $\lambda/2$, it follows that the higher is the frequency the shorter is the machine. On the other hand for
Fig. 15 The power splitter used at BNL for feeding the RFQ from eight places and simultaneously suppressing the TE_{111} mode

very high frequencies the length of the cells becomes too short at the low energy end of the machine. Moreover the working frequency determines the radius of the RFQ cavity and too low frequencies demand a very large diameter.

Another important parameter is the voltage between the adjacent vanes. As a general rule this voltage should be as high as possible, obviously avoiding the risk of sparking.

If the ion species with their initial and final energies are specified, and f, the frequency, and intervane voltage are given, then the RFQ design is determined when the three independent functions a(z), m(z), q(z) are given, where z is the axial distance along the accelerator. Two different ways for arriving at the above functions are indicated in Ref. 3. The methods used at Los Alamos can be better understood with the aid of Fig. 16, which shows a functional block diagram of an RFQ where, beside the acceleration, the greatest attention was paid to limit the growth of the radial emittance of the beam.

```
| RADIAL MATCHING SECTION | SHAPER | GENTLE BUNCHER | ACCELERATOR |
```

Fig. 16 Functional diagram of an RFQ
As indicated in the figure, the first section accomplishes the transition from a beam having time independent characteristics to one that has the proper variations with time (in this section the profile of the vanes is smooth). In order to obtain high capture efficiency the bunching and the energy of the beam should be slowly varying functions of $z$. This is achieved in two different sections of the machine. Typically the quantity $A$ increases very little in the shaper, while it undergoes a significant change in the gentle buncher. In the last section, since the bunching is nearly completed, both the synchronous phase and the value of $A$ are held constant. Figure 17 shows the suggested variations of the parameters along the machine. It is important to recognize that when $\lambda$, $\phi$ and $k$ are given then $a$ and $m$ are consequences of the assigned values for $A$ and $X$.

**RFQ ACCELERATION CYCLE**

![Diagram](image)

Fig. 17 A possible choice for the function $A(z)$, $\phi(z)$ and $B(z)$

Another important parameter is the maximum value of the electric field $E_s$ that should always be kept below the sparking limit. Analytical and numerical calculations show that the maximum field $E_s$ occurs around the middle of each separation. A good approximation for $E_s$ can be as follows:

$$E_s = \frac{\alpha}{E_0}$$  \hspace{1cm} (44)

where $\alpha$, the enhancing factor, is near 1.4 and obviously depends on the pole tip shaping.

If the operating wavelength $\lambda$, the normalized focusing force $B$ and the maximum field $E_s$ are assigned, then combining the equations (28), (39) and (44) we obtain the value for $r_0$ and $V_0$ as follows:

$$r_0 = \frac{\lambda}{\alpha B} \frac{q\lambda E_s}{m_e c^2} ; \hspace{1cm} V_0 = E_s \frac{\lambda}{\alpha} \frac{q\lambda E_s}{\epsilon^2 B m_e c^2}$$  \hspace{1cm} (45)
If we call $E_0$ the average wave of the amplitude of the electric field in each cell we can write:

$$E_0 = \frac{AV_0}{\beta\lambda/2} = \frac{E_s}{2q\lambda E_s} \frac{2Aq\lambda E_s}{\omega^2 B\beta m_0 c^2}$$

(46)

and this means that once the value of $E_0$ is assigned (according to the selected energy gain per cell) then the value of $A$ is determined. From Eq. (16) and taking into account that $X$ and $r_0$ are correlated [Eq. (30)] we can calculate (numerically) the values for $m$ and $a$. Values for $m$ equal to one at the low energy end, and near to two at the high energy end, normally produce a good compromise between acceleration and focusing efficiency.

8. RECENT DEVELOPMENTS

The RFQ described is a very complicated radio frequency resonator which has the purpose of creating the special RF fields capable of focusing and accelerating a beam. Following the first proposal from Kapchinski and Tepliakov, it was clear that any device capable of exciting four suitably shaped electrodes could be used; the outstanding solution studied and realized at Los Alamos was successfully adopted in many laboratories and lasted until new developments were presented at the Santa Fe conference on "Particle Accelerators" in 1982.

The leading idea that was very simple was for a non-uniform transmission line, made with four bars with circular cross section, be used for creating the special field needed in an RFQ. Figure 18 shows a very simple and effective arrangement. It is rather evident that each bar can be turned on a lathe, while for shaping the vane pole tips the very complicated and expensive tridimensional milling machine was mandatory. Moreover the reciprocal position of the bars can be easily adjusted without interfering with the container that, on the other hand, can have a cross section independent of the working frequency.

![Fig. 18 Transmission line formed with four bars. The indicated shaping produces the focusing and accelerating field (Frankfurt Univ.)](image)

While the mechanical advantages obtained with the four bars are really enormous, there are some doubts about the electrical efficiency of this structure. The surface offered by the four bars to the RF currents is always smaller than the one offered by the equivalent vane and the corresponding four vane RFQ exhibits a larger shunt impedance. The choice of the best way for designing an RFQ
cannot be decided on theoretical basis. Only the purpose for which each machine is designed can indicate what is more important; the mechanical simplicity or the RF power consumption. Figure 19 shows a sketch of the fundamental structure of an RFQ realized with the bars. Only two bars, of opposite polarity are shown for the sake of clarity. The U-shaped support can be considered as a piece of uniform transmission line made from two parallel metallic tapes. One end of the transmission line (the bottom) is short circuited while the other end is loaded with the four bars that, as a first approximation, behave as a "distributed" capacity.

Let $Z_0$ be the characteristic impedance of the transmission line and $c$ the loading capacity of the bars. Then, if losses and radiation are neglected, the structure will exhibit an infinite impedance at the open end if the length of the support, the radian frequency $\omega$ and the loading capacitance obey the well known relation:

$$\frac{1}{\omega c} = z_0 \tan \left( \frac{\omega L}{v} \right)$$  \hspace{1cm} (47)

where $v$ is the phase velocity that, in our case, can be set equal to the speed of light in vacuum.

If the above condition is verified then the metallic support does not perturb the bars (the so-called $\lambda/4$ support). The whole structure of the machine can be realized by supporting, periodically, the bars with resonating supports. From the first proposal many different resonating supports have been invented$^7,8$ and a large variety of devices have been tested. It is important to note that each resonating support is magnetically coupled, at least, with the neighbouring one. Taking advantage of this situation it is possible to arrange that all the elements of the structure resonate in phase, independently of the physical length of the bars. Consequently the amplitude of the voltage which excites the bars is constant.
The four bars and the supports should be contained in an appropriate metallic tank in order to prevent radiation escaping from the structure but, in this case, the container is not part of the fundamental structure as for the vane RFQ.

ACKNOWLEDGEMENTS

The author is indebted to C. Rossi for his cooperation and to C. Guida and G. Bonaschi for correcting the present work. The drawings and the photographs obtained from my friends in Los Alamos and Brookhaven have been essential for this seminar.

***

REFERENCES


6) S. Abbott et al., Lawrence Berkeley Lab. LBL 14624 (1982).


8) S.O. Schriber, Present status of RFQ, Los Alamos National Laboratory, At-Do Ms H811 - Los Alamos, NM 87545.
FUNDAMENTAL FEATURES OF SUPERCONDUCTING CAVITIES FOR HIGH ENERGY ACCELERATORS

H. Piel
Department of Physics, University of Wuppertal, Wuppertal, West Germany

ABSTRACT
Superconducting accelerating systems are presently under design, test or construction for electron positron storage rings and linear accelerators for nuclear physics research. This seminar tries to give an introduction to superconducting accelerating cavities which are the main elements of these instruments. The fundamental features of superconductors in rf fields namely the surface resistance and the fundamental limits of the accelerating field are discussed and the design principles of superconducting cavities for high energy accelerators are outlined. Special attention is paid to the anomalous losses in these resonators which are responsible for the performance limitations observed today. Diagnostic techniques, defects, thermal stability, high purity niobium and, last but not least, electron loading are the key words in this context. Current projects in different laboratories and the important parameters and achievements of experiments directed towards the application of superconducting cavities in high energy accelerators are reviewed briefly.

1. INTRODUCTION

It is now more than 20 years ago since the first electrons were accelerated in a superconducting lead-plated resonator at Stanford \(^1\). Between 1968 and 1970 very successful experiments with X-band resonators fabricated from bulk niobium \(^2\) laid the ground for large scale systems built thereafter.

In the beginning of the 70's construction of the Stanford Superconducting Recyclotron \(^3\), the Illinois Microtron using a superconducting accelerating section \(^4\) and the CERN-Karlsruhe s.c. Particle Seperator \(^5\) was started. In 1974 a superconducting resonator successfully accelerated an electron beam to 4 GeV in the CORNELL Synchrotron \(^6\) and in 1976 the construction of the Argonne s.c. Heavy Ion Post-Accelerator \(^7\) was begun. In 1977 the first Free Electron Laser was operated using the high brightness beam of the Stanford Superconducting Accelerator \(^8\). Several of these devices have now been operated for many thousands of hours reliably and under routine conditions. It was shown that the drastic reduction of the rf surface resistance in s.c. cavities could be achieved even in complex resonators. The early expectations however, to reach the very high electric accelerating or deflecting fields promised by the elementary theory of superconductors in radio frequency fields were not fulfilled. In analysing the performance of the s.c. resonators it is necessary to consider their geometry. Superconducting structures for proton or heavy ion accelerators therefore have to be discussed separately from accelerating structures for electrons.
According to the title of this seminar I want to focus on velocity of light structures. This is done to concentrate on one important aspect of the application of the s.c. cavities. Other important applications of s.c. cavities are in the field of heavy ion accelerators (for a review see Ref. 9). The experiments with the Single Atom Maser 10 and the Superconducting Cavity Stabilized Oscillator 11 are examples of the successful application of s.c. cavities outside accelerators in atomic physics and metrology.

Although the accelerating fields of 2 to 3 MV/m achieved in the first operating s.c. accelerators were about 10 to 15 times lower than expected from BCS-theory, the early results at X-band and recent experiments at L- and S-band frequencies show that there are no other fundamental limitations. Research and development work in rf superconductivity should therefore be rewarding.

This seminar tries to give an introduction to the fundamental features of superconducting cavities and is organized as follows: In the following section the concept of coupled resonators which form an accelerator module and important quantities like the rf surface resistance, the cavity Q and the shunt impedance are introduced. Section 3 discusses the fundamentals of rf superconductivity. A short introduction to superconductivity is given. The surface resistance of a superconductor in an rf field is explained in the frame of a two-fluid model and the critical rf magnetic surface field is introduced. The fourth section gives design considerations for superconducting cavities and addresses the problem of electron multipacting. In section five the importance of anomalous losses in s.c. cavities is outlined. The diagnostic method, microscopic defects, thermal stability and high purity niobium as well as the progress in electron field emission studies are described. Section six deals with cavities covered with superconducting thin films. Niobium sputtered onto a copper cavity and niobium cavities with a Nb₃Sn surface are the two subjects. The last section gives a brief review of achievements in present experiments directed towards the application of rf superconductivity to high energy accelerators.

For additional reading on the subject of this seminar the references 9, 15 and 16 are suggested.

2. SOME CAVITY FUNDAMENTALS

2.1 Coupled cavities

The heart of each high energy accelerator is the rf accelerating section which generally is composed of a number of accelerating modules each of which is a chain of coupled rf resonators. For educational purposes we want to assume that such a module is a string of weakly-coupled pill-box cavities as shown in Fig. 1a each of which is excited in the TM₁₀ mode. This mode has a longitudinal electric field on the axis of the cavity which is surrounded by the circular field lines of the magnetic field which reaches its maximum at
the cylindrical wall of the cavity. The accelerating module of Fig. 1a is a chain of
coiled oscillators very much like the coupled pendula shown in Fig. 1b. The resonant fre-
quency of the free pendulum corresponds to the resonant frequency (ω₀ = 2πf₀)

\[
\omega_0 = 2.405 c/a \\
c = \text{velocity of light} \\
a = \text{radius of pill-box cavity}
\]  

of the TM_{010}-mode of the pill-box cavity. The coupling spring between the pendula is equi-
alent to the coupling electric flux through the small iris openings connecting the indi-
vidual cavities.

In classical normal conducting linear accelerators such a module consists of many cavi-
ties and is generally operated in a travelling wave mode. The rf power is coupled into the
first cavity of the string, travels down the structure and is absorbed strongly by the rf
losses in the cavity walls. In superconducting accelerating modules these losses are re-
duced by many orders of magnitude and a travelling wave operating mode is inappropriate. A
superconducting accelerator module is therefore operated as a chain of N coupled resonators.
Such a module is then excited in one of its N eigenmodes. By solving the characteristic
equations of such a coupled oscillator system one obtains for the resonant frequencies ω_q
of the eigenmodes and the axial electric field E_n (q,t) of the n-th cavity the following
relations

\[
\omega_q^2 = \omega_0^2 (1 + K(1-\cos ω_q))
\]  

\[
E_n (q,t) = E_o \sin \left(\frac{2n-1}{2} a_q \right) \cos ω_q t
\]  

E_o = maximum axial electric field

a_q = q \cdot \frac{2}{N} \quad q = 1,2 \ldots N

N = number of coupled cavities

K = coupling factor between cavities.
In the so called π-mode \((q = N \text{ or } a_q = \pi)\) the accelerating module oscillates in its highest frequency and normalized to the acc. field has the smallest rf losses in its walls. This is the reason why the π-mode is a favourite mode of operation for accelerating modules. In this mode the accelerating fields are equal in magnitude and opposite in direction in each pair of cavities as shown in Fig. 1a and seen from Eq. (3). A velocity of light electron which enters the first cavity at time 0 will enter the second cavity after a time \(t = \frac{d}{c}\).

If this time equals half the rf period \((\pi/\omega_N)\), then the electron will receive a maximum of acceleration in the accelerating module. The length \(d\) of one cavity of the module is then equal to \(\pi c/\omega_N\) where \(\omega_N\) equals the π-mode frequency of the module according to Eq. (2). A disadvantage of the π-mode is its sensitivity to mechanical tuning errors of the individual cells of a module which scales with \(N^2\). The average accelerating field \(E_a\) referred to frequently in this seminar is given as \(E_a = V/l\), where \(V\) is the voltage gain of the electron after traversing an accelerating module of length \(l = N \cdot d\). \(E_a\) is directly proportional to \(F_0\).

2.2 Surface resistance, cavity Q and shunt impedance

In normal conducting cavities fabricated from high conductivity copper the electromagnetic field penetrates into the cavity wall by the skin depth \(\delta\) with

\[
\delta = \left(\frac{2}{\sigma \omega}\right)^{1/2}
\]

\(\sigma = \text{electrical conductivity (for copper at room temperature 5.80\cdot10^7/(\Omega m)})\)

\(\mu = \text{magnetic permeability of the cavity wall.}\)

At 500 MHz this skin depth is about 3.0 μm. The rf losses per unit surface area \(P_s\) produced in this thin layer can be expressed as

\[
P_s = \frac{1}{2} R_s H_s^2
\]

where \(H_s\) is the magnetic surface field and \(R_s\) is the surface resistance. \(R_s\) has the dimension of Ohms and for a normal conducting cavity is given by

\[
R_s = \left(\frac{\mu \omega}{20}\right)^{1/2} = \frac{1}{\sigma^2}.
\]

This gives at 500 MHz a surface resistance of 5.8 mΩ. In very pure metals, \(\sigma\) which is proportional to the mean free path \(k\) of the conduction electrons can be increased by more than four orders of magnitude if the conductor is cooled to the temperature of liquid helium.

The rf surface resistance however decreases only by a factor of about five. This behaviour is not explained by (6) and is due to the anomalous skin effect which has to be considered when \(k\) becomes comparable to the classical skin depth \(\delta\). In the limit of \(k >> \delta\) the surface resistance is given by (7):

\[
R_s = \frac{8}{9} \left(\frac{\sqrt{3}}{16\pi} \cdot \frac{\mu_0 e^2 k^2}{\sigma}\right)^{1/3}.
\]
$R_s$ scales like $\omega^{2/3}$ and becomes independent of $\lambda$. It is therefore of no benefit to cool a normal conducting cavity to low temperatures.

The quality factor $Q$ of a cavity is directly related to its surface resistance. $Q$ is defined as the ratio of the energy stored in a cavity to the energy lost per rf period. Energy can be transferred to the particle beam, it can be radiated out of the cavity through openings or antennas and it is dissipated and converted to heat by the rf losses in the cavity wall. If only the losses $P_w$ from the unavoidable Joule heating of the cavity wall are taken into account one arrives at the unloaded $Q$ of a cavity:

$$Q_o = \frac{\omega U}{P_w}. \quad (8)$$

The stored energy $U$ is proportional to the cavity volume and to the square of the average accelerating field. $P_w$ scales with the surface resistance, the area of the cavity wall and is also proportional to $E_a^2$. It therefore can be shown that (8) reduces to

$$Q_o = G/R_s \quad (9)$$

where $G$ is the so called geometry constant of the cavity. It is independent of the cavity frequency and for resonators like the ones shown in Figs. 1a and 4 is approximately 270 to 300 $\Omega$.

If $P$ is the rf power per unit length necessary to maintain an accelerating field $E_a$ in an unloaded cavity, then the shunt impedance (per unit length) $r$ of the accelerator cavity is defined by

$$P = E_a^2/r. \quad (10)$$

The shunt impedance is proportional to $Q_o$. The specific shunt impedance is defined as the ratio $r/Q_o$. For a single cell cavity, shaped like the cells of the accelerator structure in Fig. 4, $r/Q_o \cdot d$ is about 150 $\Omega$. The length $d$ of one cell in an accelerator structure for highly relativistic electrons operated in the $\pi$-mode is $\pi c/\omega$ as already mentioned in section 2.1. One therefore obtains from Eqs. (9) and (10)

$$P = \pi c R_s E_a^2/150 \cdot \omega \cdot G. \quad (11)$$

From this one concludes that the rf power necessary to maintain a given accelerating field per unit length is, for normal conducting cavities, proportional to $\omega^{-1/2}$. High frequencies are therefore favoured for the operation of normal conducting linear accelerators. At 500 MHz, which is a frequency typical for storage ring cavities, Eq. (11) gives for copper at room temperature and for $E_a = 5$ MV/m a dissipated rf power of 1.0 MW/m.

This very high power dissipated in normal conducting accelerator cavities is the main reason for the interest in rf superconductivity.
3. SUPERCONDUCTING CAVITIES

3.1 Short introduction to superconductivity

It is well known that many metals and alloys become superconducting below a certain critical temperature $T_C$ which is characteristic for the specific material. The highest $T_C$ known today is about 23 Kelvin and therefore all superconducting devices have to be cooled by liquid helium. Only three years after the first liquefaction of helium, Heike Kammerlingh Onnes 18) found in Leyden 1911 that mercury lost its resistivity completely below 4.15 K. A great many but not all metals become superconducting and it took almost 50 years before Bardeen, Cooper and Schrieffer 19) could explain the mechanism behind this phenomenon in their theory often referred to as the BCS theory. It would be beyond the scope of this seminar to give an account of this beautiful theory but it may be useful to extract some ingredients in order to explain the two-fluid model of a superconductor given by H. London already in 1934 20). This model is very useful for understanding the basic features of a superconductor in an rf field.

It turns out that due to the interaction of the conduction electrons in a metal with the vibrations of the atoms in the lattice there is a very small net attraction between electrons. As a result of this conduction electrons can form into pairs, the so called Cooper pairs. The energy of pairing $2\Delta(T)$ ($\Delta(T)$ would be the pairing energy per electron) is very weak and in the BCS theory given at $T=0$ to be

$$\Delta(0) = \alpha kT_C$$

$$\alpha = 1.75$$

$$k = \text{Boltzmann constant}.$$  

Only a very small thermal energy is needed to ionize a Cooper pair back into two "normal" electrons. At $T=0$ all conduction electrons are paired but at finite temperatures there is always a probability that a pair is broken up. This probability is given by the Boltzmann factor $\exp(-\Delta(T)/kT)$ and for the ratio of the densities of normal electrons ($n_e$) and Cooper pairs ($n_c$) we find:

$$n_e/n_c = e^{-\Delta(T)/kT}.$$  

At temperatures below $T_C/2$, $n_c$ and $\Delta$ are very close to their values $n_o$ and $\Delta(0)$ at $T=0$. The "two fluids" therefore are the superfluid of Cooper pairs of density $n_o$ and the normal fluid of conduction electrons of density $n_e(T)$ with

$$n_e(T) = 2\cdot n_o \cdot e^{-\frac{T}{T_C}}$$  

for $T < T_C/2$.  

R.F. Feynman 21) gives a very instructive explanation as to why the fluid of Cooper pairs can carry an electric current without any losses. Contrary to normal electrons, Cooper pairs
are Bose particles. When there are many Bosons in a given state then there is an especially large probability for the other Bosons to go into the same state. So nearly all Cooper pairs will be locked down at the lowest energy in exactly the same state and it will not be easy to get one of them out of this state. The probability to go into this state is by a factor \( \sqrt{n_0} \) higher than into any other state and \( n_0 \) is a very large number. Therefore all Cooper pairs move in the same quantum state. Resistivity comes from knocking on electrons and transferring energy to the lattice but this becomes impossible because they are all bound into Bosons.

Cooper pairs can be ionized by electromagnetic radiation if the frequency is high enough. The energy of the photons has to be

\[
\hbar \omega = 2 \Delta(T)
\]

which in the case of niobium (\( 2\Delta(0) = 3.12 \text{ meV} \)) results in a frequency of about 700 GHz.

It should be noted that Cooper pairs are not closely bound like, for example, the nucleus and its electrons in an atom. Cooper pairs are ordered states in momentum space with the two electrons having opposite but equal momenta and opposite spins. For our purpose however it is qualitatively acceptable, although somewhat superficial, to consider a Cooper pair as a bound state with a rather large extension for which the coherence length \( \xi \) gives a good measure. \( \xi \) is a material constant and ranges typically between 38 nm (niobium) and 1600 nm (aluminium). The distance between Cooper pairs is therefore considerably smaller than their "size".

A sufficiently strong magnetic field will destroy superconductivity. The critical value of the applied field is denoted by \( H_c(T) \) and exhibits a temperature dependence given by

\[
H_c(T) = H_c(0) \left( 1 - (T/T_c) \right)^2.
\]

Meissner and Ochsenfeld \(^{22}\) found that in a superconductor which is cooled in an external field smaller than \( H_c \), below \( T_c \) the magnetic field is completely expelled. The interior of the superconductor is screened by currents which flow in a very thin skin layer. The external magnetic field exponentially decays in this surface layer and its decay length is called the London penetration depth \( \lambda \). It ranges between 15 and 110 nm and is material dependent.

There are two classes of superconducting materials denoted as type I and type II superconductors. There is no difference in the fundamental mechanism of superconductivity between them. They differ from each other only by a completely different Meissner effect. A good type I superconductor excludes a magnetic field until superconductivity is destroyed abruptly at \( H_c \) and then the magnetic field penetrates completely. A good type II superconductor expels the field only for relatively weak external fields smaller than \( H_{c1} \). Above \( H_{c1} \) the field partially penetrates into the superconductor which remains superconducting. At a much higher \( H_{c2} \) field, sometimes 100 kOe or more, the flux penetrates completely
and the superconductivity vanishes. The so called thermodynamical critical field is then
approximately the geometric mean of the lower and upper critical magnetic field:
\[ H_c = (H_{c1} \cdot H_{c2})^{1/2}. \]  (17)

3.2 Basic characteristics of a superconducting cavity

3.2.1 The rf resistance of a superconducting surface

In the case of a normal conducting rf resonator the electromagnetic field penetrates
by the skin depth into the cavity wall. In a superconducting cavity the equivalent "super-
conducting skin depth" is approximately equal to the London penetration depth and there-
fore about two orders of magnitude smaller than \( \delta \). In contrast to the zero resistivity for
dc electric currents there are losses if the superconductor is exposed to a high frequency
field. This can be explained by the two-fluid model. The time varying magnetic surface
field, \( H_s \), penetrates into the superconductor and induces in the "s.c. skin depth" an
electric field. The amplitude of this field will therefore be proportional to \( \omega H_s \). The
electric field accelerates the Cooper pairs which transport this part of the surface
current without losses. It will also accelerate the normal electrons which can interact
with the lattice and produce losses according to the anomalous skin effect. The power dissis-
pated in the wall of the s.c. cavity per unit area \( P_s^t \) (the index \( t \) denotes the two fluid
model) can therefore be expressed as
\[ P_s^t = n_e(T) \omega^2 H_s^2. \]  (18)

Using Eq. (14) one arrives at
\[ P_s^t = n_e \omega^2 e^{-\frac{\alpha}{T}} H_s^2 \]  (19)
\[ T < T_c/2. \]

Comparing Eqs. (5) and (19) one gets for the surface resistance in the two-fluid model for
frequencies well below the ionization limit and for \( T < T_c/2 \):
\[ R_s^t = A \omega^2 e^{-\frac{\alpha}{T}} \]  (20)

where \( A \) may depend on material parameters like \( \lambda, \xi, \ell \) and \( v_F \). For frequencies below 10 GHz
and for \( T < T_c/2 \) the experimental data are in fact described well by the relation
\[ R_s = A(\lambda, \xi, \ell, v_F) \frac{\omega^2}{T} e^{-\frac{\alpha}{T}} + R_{res}. \]  (21)

The residual resistance \( R_{res} \) which is temperature independent and not related to the super-
conducting surface is easily separated. The first term in Eq. (21), which is often referred
to as the BCS resistance, agrees remarkably well with the result of the two-fluid model.
Expressions for the surface resistance which are based on the BCS theory have been derived by Mattis and Bardeen \(23\) and Abrikosov, Gorkov and Khalatnikov \(24\). Computations of the surface resistance based on these rather complex expressions have been performed by Halbritter \(25\) and Turneaure \(26\). A further refinement of the BCS theory in regard to rf superconductivity has been achieved by R. Blaschke \(27\) by including the anisotropy of the pairing energy which is induced by the anisotropy of a crystal lattice. This modification removed a long existing discrepancy between experiment and theory in respect to the frequency dependence of the surface resistance. The quadratic dependence reflected by the two-fluid model has to go into a \(\omega^{2/3}\) behaviour as the frequency approaches the ionisation limit.

Figure 2 compares experimental data on the surface resistance of niobium at 4.2 K by U. Klein \(28\) and G. Müller \(29\) with the computational results of Blaschke. The agreement is excellent. The two-fluid model describes the frequency dependence below 10 GHz quite well but cannot account for the change of slope at very high frequencies.

Figure 3 shows the temperature dependence of the surface resistance of a single cell niobium cavity at a frequency of 3 GHz. The exponential temperature dependence explained by the two-fluid model is nicely demonstrated as well as the existence of a residual resistance which is characterized by its temperature independence. Extracting \(a\) from the data of Fig. 3 and many other experiments at other frequencies, one finds a very near to 1.85 for frequencies below 10 GHz. This value is close to the prediction of the BCS theory.
At 500 MHz and at a temperature of 4.2 K the BCS surface resistance of niobium is 70 nΩ compared to the 5.8 mΩ of copper at room temperature. For an accelerating field of 5 MW/m (example in section 2.2) the dissipated power in a superconducting accelerating module will be only 12 W. This power is absorbed at 4.2 K and has therefore to be corrected for the Carnot and technical efficiency of a 4.2 K refrigerator. This brings the 12 W to 5.5 kW which is two hundred times lower than the power dissipated in an equivalent copper structure.

Another important difference between a superconducting and normal conducting cavity becomes apparent if one combines the equations (11) and (21) neglecting the residual resistance $R_{res}$. One obtains then for the power dissipated in a superconducting cavity:

$$ P_{sc} = \frac{\pi c w}{150 \pi} \frac{T_C}{T} \frac{T}{E_a^2}. $$

One sees that, contrary to the case of normal conductivity, low frequencies are preferred in superconducting cavities. Presently the validity of this statement is limited to surface resistances larger than 50 to 100 nΩ. The reason for this is the residual resistance $R_{res}$ which, to our present knowledge, is not a property of a superconducting surface in an rf field. It is caused by anomalous losses which are described in more detail in section 5. They are critically dependent on the purity of the cavity surface. Chemical etching, electropolishing, rinsing with ultra pure water and methanol and very high temperature treatment (up to about 1800°C) in a UHV furnace are final preparation steps for superconducting cavities fabricated from niobium. Normal conducting residues left on the cavity surface by these procedures can contribute significantly to the residual resistance. Achieved residual resistances of 1 nΩ or, more typically, 10 nΩ correspond to only about 0.1 to 1 ppm of normal conducting surface area. It is therefore obvious that superconducting cavities have to receive their final surface preparation and assembly in a clean room environment.

### 3.2.2 Fundamental field limitations

All the considerations given above are valid only if the superconducting cavity is in a true Meissner state. On first sight this can only be the case if the maximum magnetic rf surface field $H_{B \text{max}}$ is smaller than $H_{C1}$ or $H_{C2}$ in a type I or a type II superconductor respectively. This statement however may hold only in equilibrium condition and may therefore not apply to microwave cavities. The transition from the superconducting to the normal conducting state is a phase transition. Such a transition needs nucleation centers and it is therefore possible that there may be a metastable or superheated state before the superconductor returns to its normal conducting state. The maximum field up to which this transition state may persist is called the critical superheating field $H_{sh}$. In type I superconductors like lead for example, $H_{sh}$ is higher than $H_{C1}$. For type II superconductors ($\text{Nb}_3\text{Sn}$ for example) the superconducting state persists beyond $H_{C2}$ but $H_{sh}$ stays below the thermodynamical critical field $H_{C1}$. Hatricon and James have calculated the dependence of $H_{sh}$ on $\kappa = \lambda/\xi$ by solving the Ginzburg Landau equations which are based on a phenomenological theory of superconductivity. Their results have the limiting form $16$.
\[ H_{sh} = 0.75 H_C \text{ for } \kappa >> 1 \]
\[ H_{sh} = 1.2 H_C \text{ for } \kappa = 1 \]
\[ H_{sh} = \frac{1}{\kappa} H_C \text{ for } \kappa << 1 \]

\( H_{sh} \) shows a smooth behaviour as \( \kappa \) passes through the interesting value of \( 1/\sqrt{2} \) which separates type I from type II superconductors.

The persistence of the Meissner state may be very stable in rf fields. This is expected because the nucleation time of flux lines is around \( 10^{-6} \) s compared to the \( 10^{-8} \) s typical for the rf period of microwave cavities. Experimentally the rf critical field has been studied for type I superconductors like In, Sn and Pb near their critical temperature \( T_c \). The results of these experiments are in agreement with theory \(^{32}\). For a typical type II superconductor like Nb\(_3\)Sn the lower critical field \( H_{c1} \) has also been surpassed \(^{33}\). One therefore presently assumes that the fundamental limit for \( H_{sh} \) is given by the critical superheating field. Table 1 gives some material parameters for Pb, Nb and Nb\(_3\)Sn.

### Table 1

Transition temperature and critical fields of the most frequently used materials in rf superconductivity:

<table>
<thead>
<tr>
<th>Material</th>
<th>( T_c ) [K]</th>
<th>( H_C ) [Oe]</th>
<th>( \kappa )</th>
<th>( H_{sh} ) [Oe]</th>
<th>( H_{sh}^{exp} ) at ( T=2K ) [Oe]</th>
<th>( E_a^{max} ) at ( T=2K ) [MV/m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pb</td>
<td>7.2</td>
<td>804</td>
<td>0.5</td>
<td>1050</td>
<td>900(^{32})</td>
<td>22</td>
</tr>
<tr>
<td>Nb</td>
<td>9.2</td>
<td>2000</td>
<td>1.0</td>
<td>2400</td>
<td>1590(^{34})</td>
<td>50</td>
</tr>
<tr>
<td>Nb(_3)Sn</td>
<td>18.2</td>
<td>5400</td>
<td>20</td>
<td>4000</td>
<td>1060(^{33})</td>
<td>88</td>
</tr>
</tbody>
</table>

The maximum magnetic surface field in a cavity excited in the \( T_{100} \)-mode is close to its equator and a good rule of thumb is \( H_s^{max} / E_a = 45 \) Oe/MW/m. It is the ratio which is used in Table 1 to compute the maximum accelerating field \( E_a^{max} \) from \( H_{sh}^{'}(0) \). The comparison between \( H_{sh} \) and \( H_s^{exp} \) (compare Eq. (16)) shows that even the latest experimental results do not yet attain the theoretical expectations. This however, from experimental evidence, is due to the anomalous and point like losses described in section 5. Today we do not know of any fundamental limitation which prevents us from reaching the limiting fields given in Table 1. The high values for the accelerating field promised especially for niobium and Nb\(_3\)Sn cavities make it worthwhile to continue the experimental efforts.

### 4. CAVITY DESIGN

The main design criterion for a normal conducting accelerator cavity is the minimization of the rf power necessary to maintain a given accelerating field. The surface
resistance of such a cavity is fixed by the choice of the most suitable material, high conductivity copper. Therefore \( \frac{r}{Q_0} \) has to be optimized. In superconducting cavities \( R_s \) can be reduced by orders of magnitude by changing the operating temperature and can be made almost arbitrarily small for practical purposes if one succeeds in controlling the residual resistance. Therefore \( \frac{r}{Q_0} \) is an almost free design parameter and other important design criteria can be considered.

4.1 Electron multipacting and the spherical cavity shape

The resonant multiplication of free electron currents (called electron multipacting) was a very annoying field limitation in practically all superconducting cavities before 1979. This phenomenon was analysed and virtually eliminated by work done at Stanford 12), Genoa 13) and Wuppertal 14) in 1977 to 1979. Cavity shapes 14) and, in special cases, grooving of the cavity surface 36) were proposed which later proved to suppress multipacting up to the highest fields reached so far. It is because of this that today all s.c. accelerator structures are of the spherical or elliptical 36) design as can be seen from the examples shown in Figs. 4 and 5.

---

**Fig. 4** 350 MHz niobium cavity foreseen for the energy upgrade of LEP 37).

The whole unit has a length of 2.4 m.

---

**Fig. 5** Five-cell 3 GHz cavity for the Darmstadt 130 MeV Superconducting Recyclotron 38) showing the rf power coupler (1), frequency tuning system with motor driven coarse tuner (2) and piezoelectrically driven fine tuner (3).
Although electron multipacting appears not to limit the performance of superconducting cavities any longer, a short account of this avalanche phenomenon will be given. "One point multipacting", which is the multipacting variety which has plagued s.c. cavities of the old design, comes about by the following process:

An electron of a few eV may be released from the cavity wall into the rf field for example as a knock-on electron from a cosmic ray event. It is accelerated by the local electric field $E_{\perp}$ which is perpendicular to the cavity surface and bent backwards to its origin by the tangential magnetic surface field $H_{\parallel}$ as shown in Fig. 6.

![Fig. 6 Cross section of a π-mode accelerator cavity of a design typical before 1979. The circle indicates an area where one-point multipacting preferably takes place. In the magnified view of this area multipacting trajectories of first, second and third-order are displayed.](image)

If the local configuration of the time dependent electromagnetic field is such that the electron returns approximately to its starting point after one rf period, it can produce other electrons by secondary emission. If the secondary emission coefficient is larger than one, this process leads to a resonant multiplication and an avalanche develops. This avalanche absorbs all the excess rf power delivered to a cavity in order to increase the accelerating field. Therefore the field is limited at a so-called multipacting threshold.

If, in a very simplified picture, one assumes that the electron moves on "cyclotron orbits" in the magnetic surface field $B_s$, then its round-trip frequency $\omega$ would be $eB_s/m$. As the rf period has to be a multiple of the "cyclotron period" one would find the resonance condition $B_s = (1/n) (m\omega/e)$ (with $n = 1, 2, ...$). One denotes $n$ as the order of the multipacting trajectory.
In an accelerator cavity the time dependent local electromagnetic surface field can only be calculated with computer codes like SUPERFISH $^{39}$ or URMEL $^{40}$. An analysis of multipacting trajectories and their resonance conditions can therefore only be achieved by numerical integration. This has been done successfully at Stanford in 1977 $^{12}$. Later, a similar computer code was developed at Wuppertal. For the resonance condition one found in Stanford approximately:

$$ B_s = \frac{1}{n} \varepsilon_1 \frac{m}{e} \omega $$

with $\varepsilon_1 = 0.64 \pm 0.08$.

Because of (24) it was firmly believed before 1979 that, despite the $a^2$-dependence of the surface resistance, high frequencies should be preferred for s.c. cavities in order to achieve high accelerating fields. The resonance condition (24) however is not sufficient for a multipacting barrier. The impact energy $T$ of the "returning electron" has to be high enough for the secondary emission coefficient $\delta$ to be larger than 1 (see Fig. 7).

![Diagram](image)

Fig. 7 Secondary electron emission coefficient $\delta$ of a niobium surface after different surface treatments $^{41}$ as a function of the energy $T$ of the impacting electrons

Under normal conditions the threshold energy for $\delta$ to become larger than 1 is about 30 to 50 eV. By argon discharge cleaning this threshold can be increased to about 150 eV. This explains why it is frequently possible to pass through a multipacting barrier.
If an electron released from the cavity wall is accelerated by the electric surface field, $E_{\perp} \sin \omega t$, away from the cavity wall for half an rf period, then turned around (by the magnetic field during this half period) and accelerated back to the wall it gains an energy $T$ of $e^2 E_{\perp}^2 / 2m \omega^2$. The Stanford computer calculations show that this simple estimation leads in fact to the right order of magnitude and the correct result is:

$$T = \varepsilon_2 \frac{e^2}{2m} \frac{E_{\perp}^2}{\omega^2}$$

(25)

with $\varepsilon_2 = 4 \pm 1$.

Since $T$ has to be only about 40 eV to get $\delta > 1$, which is fulfilled quite easily for typical $E_{\perp}$, (24) is a much more stringent condition for the occurrence of multipacting.

Experiments in Genoa 13) were performed in 1978 and surprisingly high fields were obtained, well above the threshold placed by (24). U. Klein and D. Proch 14) noticed this peculiarity and found that, in cavities of "spherical shape" like the ones shown in Figs. 4 and 5, stable multipacting trajectories were not to be found. This is attributed to the fact that in a spherical cavity $E_{\perp}$ is quite large everywhere away from its zero crossing at the equator. A multipacting trajectory therefore drifts after only two or three impacts to the equator where $E_{\perp} = 0$ and the multipacting electrons cannot gain energy.

Many experiments with spherical or elliptical cavities have been performed since 1979 but no one-point multipacting threshold could be identified. As those thresholds scale with frequency, the 10.8 MV/m reached in a 350 MHz cavity at CERN 69) is, so far, the best evidence for the fact that spherical cavities are virtually free of multipacting.

A very special variety of a two-point multipacting in a very close vicinity of the cavity equator was found and analysed by W. Weingarten 42). This type of multipacting can only take place on a very contaminated surface.

4.2 Higher-order modes and cavity design

An accelerator cavity is always operated by two power sources. One is the rf generator which supplies harmonic power to the cavity in order to make up for the power dissipated in the walls and absorbed by the accelerated beam. The other power source is the bunched particle beam. The latter is by no means harmonic. It has a discrete but sometimes very dense Fourier spectrum. Each cavity of an accelerator module has, apart from its fundamental TM$_{010}$ accelerating mode, an infinite number of eigenmodes at higher frequencies, the so-called higher-order modes. One of the Fourier lines of the beam may coincide with one of the HOM's of the module. Then a high field is built up which may lead to a beam instability but certainly to an unwanted Joule heating of the cavity wall or to an excitation of the cavity to its critical rf field. To avoid such circumstances the external Q of dangerous higher-order modes have to be reduced to low values. This is done by special antennas, the so-called HOM couplers. The fundamental mode and higher-order mode coupler should be located
at the beam tubes of an accelerator module (as in Figs. 4 and 5) in order not to disturb the spherical geometry of the cells. A sufficient loading of HOM's by appropriate coupling antennas then requires a good cell-to-cell coupling for as many modes as possible. A large iris opening can be helpful but may also be dangerous. Each application of a s.c. cavity needs its special optimization which can be performed today with computer programs like URMEL and TBCI. Such an optimization was carried out for the superconducting cavity for LEP and resulted in the design shown in Fig. 4. The almost free choice for \( r/Q_0 \) very much supports such optimization as already mentioned.

A bunch of charged particles which enters a cavity will produce a wake field which can react with the bunch itself and, for more than a certain threshold charge, may lead to a disruption of the bunch. Such threshold currents can be increased significantly when the metallic wall which surrounds the beam, and thereby its mirror charge, is at a large distance. Large iris openings (resulting in low \( r/Q_0 \)) and cavities of low frequencies are therefore preferred. Superconducting cavities are well suited to fulfill these requirements.

5. ANOMALOUS LOSSES

5.1 Temperature mapping and microscopic defects

The origin of field limitations well below \( H_{sh} \) and the causes of the residual resistance are the main areas of interest for the research on superconducting cavities. Several diagnostic techniques have been developed to study these questions. In the framework of this seminar only one, namely the "temperature mapping in subcooled helium" will be described.

As each energy loss mechanism will finally lead to an increase of the temperature of the cavity wall, temperature measurements are of prime importance to identify causes for field and \( Q \)-limitations. C. Lyneis at Stanford was in 1972 the first to use a chain of rotating carbon resistors mounted a few millimeters from a cavity wall to detect the location of a thermal instability.

This method has since been used by many groups working in this field. The carbon thermometers (see Fig. 8) used are 56 or 100 \( \Omega \) (1/8 W or 1/4 W) Allen Bradley resistors, the bakelite insulation of which is often ground off to increase their sensitivity. Different electric schemes have been used to read the resistance value of the many thermometers generally used on one cavity, either an oscilloscope display or an automatic data acquisition system. During a quench all the energy stored in a cavity is set free and a substantial heat flux develops which leads to film boiling and a marked increase of the temperature of the helium film close to the quench area. This can be detected easily even in superfluid helium and if the resistor is not in contact with the cavity wall. The detection of quench areas is certainly a most useful diagnostic procedure. A temperature map of the surface of a cavity well below the breakdown field however, will reveal even more information about the nature of high-loss areas. Temperature mapping can only be done for bath
temperatures above the $\lambda$-temperature. The main obstacle for a temperature mapping experiment is the fact that only the temperature of the outside of the cavity wall can be measured which is very effectively cooled by the surrounding liquid helium. In an experiment performed at CERN in 1978 \cite{44} it was shown that temperature mapping can be carried out quite well in a subcooled helium bath (favourable subcooled condition: bath temperature slightly above $T_\lambda$ and bath pressure = 1000 mb). In a subcooled bath, bubbles are absent and therefore the microconvection produced by bubbles rising from the heated surface is avoided. This reduces the cooling capability of liquid helium substantially and increases the heat transfer resistance between the niobium surface and the helium.

Fig. 8 Cross section through carbon thermometer used for temperature mapping in subcooled helium

A = copper tube housing
B = bakelite insulation
C = carbon body of resistor
D = gap filled with conduction silver
E = copper beryllium spring

The first set up used for the temperature mapping of a 500 MHz spherical cavity \cite{47} is shown in the photographs of Fig. 9.

Fig. 9 General configuration and close up view of the carbon thermometer scanning system \cite{47}
Thirty nine carbon thermometers (100 Ω, 1/4 W Allen Bradley) slide under spring tension on the cavity wall and can be rotated around the cavity. The resistor voltages and their angular position are read by a computer controlled data acquisition system. Figure 10 shows one of the first 3-dimensional temperature maps of a superconducting, 500 MHz, niobium cavity operated at an effective accelerating field of 3.2 MV/m \(^4\)). This measurement was done in a subcooled helium bath at a temperature of 2.3 K. On the x-axis the distance along one circle of constant latitude around the spherical cavity is plotted. The y-axis shows the number of carbon thermometers (with resistor 1 corresponding to the top of the cavity and resistor 39 to the bottom of the resonator). The vertical axis displays the temperature increase \(\Delta T\) detected by the carbon resistor. The residual resistance of this cavity was rather poor (\(R_{\text{res}} \approx 330\) nΩ). It can be attributed to the very non-uniform high-loss area at the top of the cavity. In this early experiment at CERN the clean-room handling was not as well developed as today and already at an accelerating field of 3.2 MV/m one observes strong non-resonant electron loading.

![Temperature map](image)

Fig. 10 Early temperature map of a CERN 500 MHz cavity at \(E_a = 3.2\) MV/m with line like regions of increased temperature due to the impact of electrons field emitted by point sources \(^4\))

After these first measurements the technique of temperature mapping was refined considerably \(^4\,4\)). The temperature increase of the intermediate helium layer at the outside cavity surface was calibrated against the heat flux density and the dependence of this calibration on the bath temperature was experimentally determined. The relation between the measured temperature increase and the heat flux density is very dependent on the "hydrodynamics" of the flow of the local convection stream in the subcooled helium bath. All these effects have to be considered carefully. The necessary calibration experiments can be carried out at higher temperatures (for example at a frequency of 3 GHz at 3 K) where
the well known BCS losses of the s.c. surface determine solely the heat flux through the cavity wall.

The spatial distribution of the heat flux density on a 20-cell superconducting accelerator module for the Darmstadt Recyclotron is shown in Fig. 11. This map is typical for the present day diagnostic technique and for the specific losses observed in a s.c. cavity. Spikes in the heat flux density are seen on the flat background of uniform losses which are expected from the BCS part of the surface resistance. Similar spikes have already been observed in the very first temperature maps at CERN. They are produced by high loss areas on the rf surface which must be smaller than a few millimeters in diameter. They are in fact found to be microscopically small and in most cases invisible to the naked eye.

![Spatial distribution of the heat flux density on a 20-cell superconducting accelerator module for the Darmstadt 130 MeV Recyclotron at $E_a = 4.8$ MV/m](image)

*Fig. 11 Spatial distribution of the heat flux density on a 20-cell superconducting accelerator module for the Darmstadt 130 MeV Recyclotron at $E_a = 4.8$ MV/m*

In a few very important experiments at CERN, such defects were detected in 3 GHz single-cell cavities by temperature mapping, then cut out of the cavity and analysed with a scanning electron microscope. Four of the photographs obtained are displayed in Figs. 12 to 15.
Fig. 12 Tungsten inclusion on a cavity weld, probably embedded during TIG-welding. Quench field $E_a = 4.5 \text{ MV/m}$.

Fig. 13 Nb sphere (presumably originating from a welding bead).
Diameter: $80 \text{ \mu m}$.
Quench field $E_a = 6.8 \text{ MV/m}$.

Fig. 14 Microscopic hole in a weld of a 3 GHz cavity, causing a thermal instability close to $E_a = 8 \text{ MV/m}$.

Fig. 15 Chemical residue (drying mark).
Diameter: $400 \text{ \mu m}$
Quench field $E_a = 3.4 \text{ MV/m}$.

Foreign material inclusions, beads from electron beam welding, holes in welds and chemical residues were found. During the mounting of the cavity to the vacuum system, to rf couplers or other parts, or during rapid pump downs particles can fall onto the s.c. surfaces. They can heat up in the cavity field to very high temperatures, emit light, cause thermal electron emission so leading to an excessive heating of their environment and thereby induce quenching. If a quench location is detected during temperature mapping a later inspection of the cavity often shows a dark spot composed of a central region and a halo as in Fig. 16. One can assume that this halo is produced by material from the "dust particle" evaporated during high field operation. In cavities mounted horizontally such particles would fall onto the equatorial surface, where the electric field is small. They would not give rise to electron loading and would initiate quenches only if they were
of metallic nature. At CERN it was in fact observed that horizontally mounted cavities reach higher quench fields at the first attempt and also show a much reduced electron loading compared to cavities mounted vertically into the test cryostat\textsuperscript{73}).

All the above observations lead to the adoption of very careful cleaning and mounting procedures for superconducting accelerator cavities. Chemical treatment of the cavities with clean chemicals, the final rinsing procedures carried out with demineralized and dust-filtered water, and the mounting of the cavities to the test facility in clean rooms have improved the reliability with which low residual resistances and high accelerating fields can be presently obtained.

5.2 Thermal instabilities and the virtues of high purity niobium

Defect induced thermal instabilities and electron field emission from point sources (see section 5.3) are the main mechanisms which limit the performance of s.c. cavities. A defect on a cavity surface like the ones shown in Figs. 12 to 15 is heated in the rf field and the dissipated energy is transferred to the helium bath. The temperature gradient produced across the cavity wall may lift the temperature of the defect's environment above the critical temperature of the niobium and a sudden dissipation of the energy stored in the cavity will result. The threshold field of such a thermal instability can be increased if the thermal conductivity of niobium can be improved\textsuperscript{51}). In standard, commercial, reactor-grade niobium the interstitial impurities O, C and N determine the poor thermal conductivity\textsuperscript{52}). These impurities can be controlled to a large extent during electron beam melting of the raw niobium and the consecutive manufacturing steps of the sheet material. The residual resistivity ratio (RRR) of niobium is proportional to its electronic thermal conductivity. Typical RRR values of standard niobium range between 20 and 40. Due to a refinement in production techniques, niobium of RRR values between 80 and 160 is commercially available since the end of 1983. This advance was achieved mainly by improving the vacuum condition and the procedure during the multiple electron beam melting of the niobium ingots. The progress in cavity performance compared to the status of 1983 can be attributed mainly to this improvement of the thermal conductivity\textsuperscript{53,54}). Not only the obtainable fields have increased, but also the reliability with which the present design
fields of 5 MV/m can be reached.

An effective procedure to clean niobium of the most critical impurity, oxygen, is the evaporation of yttrium onto the niobium surface developed at CORNELL\textsuperscript{55}. During this process the surfaces of a niobium cavity are brought into the proximity of an yttrium foil at a pressure of about $10^{-5}$ Torr at 1250°C for several hours. A vapor deposited film of several μm thickness traps the oxygen which diffuses rapidly from the bulk to the surface. The oxygen enriched surface layer of yttrium is then dissolved chemically. Starting from standard material (RRR = 30) the RRR value and thereby the thermal conductivity can be improved by about a factor of three (depending on the initial oxygen content). Starting from high purity commercial niobium, RRR values of up to about 500 were obtained at CORNELL. The same technique has been tried experimentally at CORNELL\textsuperscript{56} and KEK\textsuperscript{57} using much cheaper titanium foils at slightly higher temperatures, with similar success. The KEK results on a single-cell, 500 MHz cavity in Table 2 were obtained that way.

Figure 17 shows the measured temperature dependence of the thermal conductivity $\lambda$\textsuperscript{58} of niobium samples of different purity. Curve a) represents the status until 1983 and curve b) shows the quality of niobium which is now commercially available. Curve c) gives the thermal conductivity of a niobium sample which was yttrium treated at CORNELL.

![Graph showing temperature dependence of thermal conductivity](image)

**Fig. 17** The temperature dependence of the thermal conductivity of niobium samples of different purity characterized by its residual resistivity ratio (RRR)\textsuperscript{58}.

- a) RRR = 36,  
- b) RRR = 152,  
- c) RRR = 360.
A very instructive display is shown in Fig. 18. There the performance of cavities fabricated from niobium of different purity is compared. The measurements were carried out with single-cell, 3 GHz cavities of spherical shape excited in the $\text{TM}_{010}$ mode. Each measurement is made after a new chemical treatment which dissolves more than 20 $\mu$m of the cavity surface and therefore creates a completely new surface as far as the shallow penetration depth of the rf field is concerned. Measurements of two laboratories (CERN and Wuppertal) are contained in the data. The dependence of the cavity performance on the purity of the niobium or its RRR is clearly seen.

![Graph showing performance of s.c. 3 GHz single-cell cavities fabricated from niobium of different purity]({})

Fig. 18 Performance of s.c. 3 GHz single-cell cavities fabricated from niobium of different purity $^{53,59}$

The $Q_0$ versus $E_a$ dependence of the very high purity niobium cavity, yttrium treated at CORNELL and built and tested at Wuppertal is shown in Fig. 19. This cavity is not limited any more by defect induced thermal instability but its field is limited by electron field emission.

![Graph showing $Q_0$ versus $E_a$ dependence of a RRR = 300 single-cell cavity which was obtained by the yttrification of an RRR = 80 cavity. The cavity was built and tested at Wuppertal and yttrified at CORNELL $^{59}$]({})

Fig. 19 $Q_0$ versus $E_a$ dependence of a RRR = 300 single-cell cavity which was obtained by the yttrification of an RRR = 80 cavity. The cavity was built and tested at Wuppertal and yttrified at CORNELL $^{59}$.
5.3 Progress in field emission studies

Resonant electron loading has been overcome in superconducting cavities as is described in section 4.1. Also, the improving ability to avoid lossy defects on niobium surfaces, and the progress in thermal stability of s.c. cavities, have allowed surface electric fields of more than 25 MV/m at all frequencies suitable for accelerating structures. At such surface fields, field-emission induced electron loading is observed and constitutes an important field limitation. Already one of the very first temperature maps obtained at CERN in 1980 and displayed in Fig. 11 gave evidence for the existence of point-like electron sources which emit at anomalously low electric surface fields. The measured emission currents from the point-like sources seen in s.c. resonators do not correspond to predictions by the Fowler-Nordheim theory \(^{74}\) applied to an ideal niobium surface. The origin of this anomaly is still unknown but it can be assumed that the field emission in rf cavities is related to the dc field emission from broad area cathodes. At the University of Geneva, experiments are underway to study the field emission properties of niobium samples prepared in a similar way to cavity surfaces \(^{60}\). The measurements are carried out in a commercial Vacuum Generators "SCALAB" UHV System including a scanning electron gun producing a beam of 0.5 μm in diameter, a 157° spherical sector electron analyser, a secondary emission detector and an argon gun. Niobium samples of 1.4 cm diameter can be fixed to a purpose built manipulator which permits the cathode x-y-z-movement necessary for the field emission scans. The anode holder can accommodate several units, for example a 1 mm diameter flat anode and a pointed tungsten anode which has been electrolytically etched to a micron size tip radius (Fig. 20).

![Fig. 20](image)

Tungsten tip of the pointed anode of the "field emission scanning microscope" set up at the University of Geneva together with an emitting particle on a niobium surface.

Using this anode a high electric field can be produced on a very small area of the niobium sample. Peak surface fields of 500 MV/m have been measured locally. By moving the cathode the anode is scanned automatically across the sample with a 1 μm setting precision. Figure 21 shows a scanning image of 1 cm\(^2\) of a niobium surface at different scanning fields and after different treatments of the sample. The scan along each line of one image is carried out at a constant field. When a field emission site is encountered, the electric field is electronically reduced to hold the emission current below a fixed limit. These field reductions result in vertical deflections on the plotted lines. After localizing an
emitting site it is investigated with the built-in scanning electron microscope (Fig. 20) and its chemical composition is determined by Auger analysis. So far the most important results of Ø. Fischer's group at the University of Geneva are as follows:

In general one can say that broad area cathodes seem to show the same kind of anomalous field emission as observed in s.c. cavities. In detail the following statements seem to be valid: The emissions are most certainly not coming from metallic protrusions with a static electric field enhancement. The emission sites are usually associated with micron-size particles, some of them sitting probably rather loosely on the surface. The elemental composition of these particles is not unique. In a minority of cases no particle was seen down to a resolution of 0.5 μm. The emission from micron-size particles underlines the importance of the clean room techniques applied to the final cavity treatments before assembly. Another result of the Geneva group is shown in Fig. 21.

![Field emitting sites on a niobium surface and their sensitivity to bakeouts.](image)

- **Chemically Polished**
- **Heating**
  - 800°C, 30 mins
  - 1400°C, 30 mins

**50 MV/m**

**100 MV/m**

Fig. 21 Field emitting sites on a niobium surface and their sensitivity to bakeouts.
- Upper row: Scanning field 50 MV/m.
- Lower row: Scanning field 100 MV/m.
- From left to right: No bakeout, 800°C, 1400°C (30 mins each).
- All scans are performed with an anode of 1 mm diameter.

In the ESCALAB System the samples can be moved under UHV conditions to a station where they can be baked out at temperatures up to 2000°C. Figure 21 shows a series of field emission scans of one and the same sample which prior to each scan was baked out for 30 minutes at a given temperature. The number of emitting sites is reduced considerably after bakeout at a temperature of more than 800°C. This interesting observation certainly asks for more studies but it may already be seen as a hint to apply high temperature firing under UHV and clean room conditions for s.c. cavities to surmount the field emission barrier.
A very recent and quite spectacular result shows that, if a sample which was treated at temperatures of at least 1400°C and on which the emitters at a given surface field have been reduced as shown in Fig. 21 is heat treated again at 800°C, a very large number of emitters reappear. This is attributed by the Geneva group to microscopic segregations which dissolve in the niobium at high temperatures and which segregate again at a temperature specific for the impurity. This observation may indicate that microscopic impurities are of much greater importance than presently assumed.

6. CAVITIES COVERED WITH SUPERCONDUCTING THIN FILMS

Because of the very small penetration depth of an electromagnetic field into a superconductor, it is very reasonable to deposit thin superconducting films of special properties onto a cavity built from an appropriate supporting material. Superconducting lead-plated copper resonators were among the first cavities tested for accelerator applications. The Stony Brook post accelerator for heavy ions is successfully based on this technology. In more recent times one has started a programme at CERN to deposit niobium onto a copper cavity. The experiments at the University of Wuppertal on Nb3Sn covered niobium resonators constitute another example of rf work with superconducting thin films.

6.1 Copper cavities sputter-coated with a niobium film

It would be very desirable to have a reliable technique by which a film of pure niobium of a few µm could be deposited onto a copper cavity. This would not only improve the thermal stability at high fields but also give the possibility to produce a niobium layer virtually free of foreign material inclusions. A feasibility study towards this goal was started at CERN in 1980. A method was developed to coat a 500 MHz cavity made of OFHC copper with a thin niobium film by dc bias sputtering. Figure 22 shows a schematic view of the sputtering arrangement. Three properly shaped niobium cathodes are rotated inside the copper cavity at a potential of −1400 V and an argon pressure of 5·10\(^{-2}\) Torr. In order to confine the sputtering discharge to the front of the cathodes, the latter are surrounded at the back and the sides by a shield at 4 mm distance. This shield is biased at +80 V with reference to the cavity wall which is at ground potential. During 24 h a sputtered film between 1.1 µm (equator) and 3.7 µm (iris) is grown. The results obtained in first experiments with 500 MHz cavities in 1984 were quite encouraging. A maximum accelerating field of 8.6 MV/m was reached which is comparable to the best results from niobium cavities. The observed reduction of the cavity Q with increasing field however required a further improvement of the experimental procedure to produce a niobium layer free of defects. In very recent experiments an accelerating field of more than 10 MV/m (Table 2) was achieved in one case and, in another case, a low field Q of 3.7·10\(^9\) which reduced to 2·10\(^9\) at the 5 MV/m design field of the superconducting LEP cavities. Experiments with sputter-coated copper cavities are continuing at CERN to further investigate the field dependence of the cavity Q and the technology of the niobium sputter deposition on a four-cell 350 MHz cavity.
6.2 \( \text{Nb}_3\text{Sn} \) coated accelerator cavities

The \( Q_0 \) and thereby the shunt impedance of a s.c. accelerating structure increases exponentially with the critical temperature \( T_C \) of the superconducting material (see Eq. (2)). Therefore niobium, the element with the highest \( T_C \), is the material most frequently used for s.c. cavities. Among the A15-materials, characterized by high critical temperatures and critical thermodynamic magnetic fields (\( H_C \)), \( \text{Nb}_3\text{Sn} \) gained early attention. Its \( T_C \) of 18.2 K, \( a \) of 2.2 and \( H_C \) of 5400 Oe make it a promising material for superconducting cavities. The brittleness of this compound is of no disadvantage in this application. A \( \text{Nb}_3\text{Sn} \) layer of typically 5 \( \mu \)m is formed on a niobium cavity by the vapor diffusion process \(^{63}\). The cavity is processed in a vacuum furnace at around 1100°C in a tin atmosphere with a partial pressure of a few \( 10^{-3} \) Torr. Recently, work with \( \text{Nb}_3\text{Sn} \) resonators has been resumed at Wuppertal \(^{64}\). A single-cell and a five-cell cavity (3 GHz) have been covered with a \( \text{Nb}_3\text{Sn} \) layer. For the first time a depth profile of the \( \text{Nb}_3\text{Sn} \) layer was measured on a niobium sample which was treated by the vapor diffusion process together with a five-cell cavity (Fig. 23). This measurement was carried out using dispersive X-ray analysis in a scanning electron microscope of 0.2 \( \mu \)m resolution at CERN. The tin content near the surface slightly exceeds that of stoichiometric \( \text{Nb}_3\text{Sn} \) but is still below the upper limit of the stable \( \text{Nb}_3\text{Sn} \) phase \(^{65}\). It is observed that removing the first 0.5 \( \mu \)m of the \( \text{Nb}_3\text{Sn} \) surface by oxipolishing \(^{66}\) significantly reduces the residual resistance of a \( \text{Nb}_3\text{Sn} \) layer.
Therefore, all cavities were oxidized by this amount, rinsed with demineralized and filtered water and dust-free methanol before they were mounted in the test system.

To learn more about the seemingly high residual resistance of Nb₃Sn and its significant field dependence, and about field limitations specific to Nb₃Sn, the temperature mapping technique was applied to single and multicell cavities. One component of the residual resistance was found to be dependent on the cool down cycle. The Q₀ versus Eₐ curve (Fig. 24) clearly shows the significant difference between the residual losses after a fast and a slow cool down of the cavity. A careful study of the temperature maps taken in both cases indicates that even the residual losses after a slow cool down are, at least in part, caused by the same mechanism. The origin of these losses is unclear. At present it is assumed that frozen-in magnetic flux produced by thermoelectric currents and excited at the Nb₃Sn-Nb interface is responsible.

Fig. 23
Depth profile of the Nb₃Sn layer on a niobium sample which was treated by the vapor-diffusion process together with a five cell cavity.

Fig. 24
Dependence of the cavity Q₀ on the accelerating field and on the cool down procedure.
a) after fast cool down
b) after slow cool down.
Measurements at 20 GHz, 8 GHz and 3 GHz performed at Wuppertal show that the minimum residual resistance increases with frequency. The lowest residual resistance found so far was measured in a five-cell 3 GHz structure and was 27 nΩ. Scaling this with $\sqrt{\omega}$ to 1.5 GHz would result in a cavity $Q_0$ of about $1.5 \times 10^{10}$ at 4.2 K. In fact 1.5 GHz is the design frequency of the superconducting 4 GeV electron accelerator presently considered for construction at CEBAF in Newport News, Virginia. This accelerator will be composed of niobium cavities of the CORNELL design 67) and has to be operated at 2 K. A later conversion to Nb$_3$Sn covered resonators appears feasible and makes a further investigation of Nb$_3$Sn cavities worthwhile. At 4.2 K, a theoretical $Q_0$ of about $9 \times 10^{10}$ is expected for Nb$_3$Sn accelerating resonators at this frequency.

The accelerating fields obtained in Nb$_3$Sn cavities are comparable to results from cavities fabricated from low purity niobium. Temperature maps taken on a five-cell Nb$_3$Sn 3 GHz cavity at different field levels shown in Fig. 25, show the existence of microscopic regions of weak superconductivity. Already at low surface fields (10 mT) these regions switch to a high loss state and lead to thermal instabilities. At present one can only speculate about the nature of these switching defects. Impurity inclusions in the niobium base material which disturb the uniform Nb$_3$Sn layer and which become weak superconductors by the proximity effect are one explanation. The use of the new high purity niobium for the production of Nb$_3$Sn cavity is therefore a next experimental step.

![Diagram](image)

Fig. 25 Spatial distribution of the rf losses in a Nb$_3$Sn coated five-cell cavity taken at 2.2 K in subcooled He at $E_a = 2.55$ MV/m (a), 2.6 MV/m (b) and 3.9 MV/m (c). With increasing field a few presumably microscopic regions switch into high loss areas ($Q$ in arbitrary units).
As a conclusion of this seminar a brief summary of experiments shall be given which are directed towards the application of superconducting cavities to high energy accelerators. Table 2 displays important parameters and accomplishments of these projects. Five projects span the frequency range from 350 MHz to 3 GHz. This is nicely documented by the photograph in Fig. 26 where a 3 GHz single-cell cavity is compared to 350 MHz single-cell cavity.

**Fig. 26** The CERN 350 MHz s.c. cavity in comparison with a 3 GHz single cell. The cavity is mounted horizontally on its vacuum system. The rotating arm with 60 gliding carbon thermometers for temperature mapping can be seen.

The most ambitious programme is underway at CERN. There, has started the prototype work for the production of eight four-cell cavities of 352 MHz to be installed in the LEP storage ring in order to assess their performance under real conditions \(^{68}\). If this proves
Table 2

Best performances of cavities from present high energy accelerator projects

<table>
<thead>
<tr>
<th>LABORATORY</th>
<th>CERN</th>
<th>KEK</th>
<th>DESY</th>
<th>CORNELL</th>
<th>DARMSTADT/WUPPERTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCELERATOR</td>
<td>LEP</td>
<td>TRISTAN</td>
<td>PETRA/HERA</td>
<td>CESR</td>
<td>130 MEV RECYCLotron</td>
</tr>
<tr>
<td>MATERIAL</td>
<td>Nb</td>
<td>Nb</td>
<td>Nb on Cu</td>
<td>Nb</td>
<td>Nb</td>
</tr>
<tr>
<td>FREQUENCY (MHz)</td>
<td>350</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td>TEMPERATURE (K)</td>
<td>4.2</td>
<td>4.2</td>
<td>4.2</td>
<td>4.2</td>
<td>4.2</td>
</tr>
<tr>
<td>SINGLE CELL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E_a$ (MV/m)</td>
<td>10.8</td>
<td>13.0*</td>
<td>10.8</td>
<td>7.6</td>
<td>5.5</td>
</tr>
<tr>
<td>$Q$ at $E_a \cdot 10^3$</td>
<td>1.8</td>
<td>0.7</td>
<td>0.4</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>MULTICELL</td>
<td>4-CELLS</td>
<td>5-CELLS</td>
<td>-</td>
<td>3-CELLS</td>
<td>9-CELLS</td>
</tr>
<tr>
<td>$E_a$ (MV/m)</td>
<td>7.5*</td>
<td>5.0</td>
<td>-</td>
<td>5.8</td>
<td>5.5</td>
</tr>
<tr>
<td>$Q$ at $E_a \cdot 10^3$</td>
<td>2.2</td>
<td>0.7</td>
<td>-</td>
<td>0.6</td>
<td>0.5</td>
</tr>
</tbody>
</table>

*) Cavities fabricated from high thermal conductivity niobium

**) Yttrified niobium

to be successful then the energy of LEP will be upgraded to about 100 GeV per beam using a s.c. accelerating system. The first prototype of such a four-cell unit is shown in Fig. 27.

Fig. 27 Prototype of the four-cell 352 MHz superconducting cavities foreseen for the energy upgrade of LEP

It had its first test in 1985. The design field of 5 MV/m and the desired $Q_o$ of $3 \cdot 10^9$ were obtained at the first attempt 69). Meanwhile a long term test has been carried out with this cavity mounted in a horizontal cryostat. The cavity was operated without problems for 2400 hours at a field of 7.2 MV/m and the design $Q_o$ of $3 \cdot 10^9$. The research and development work at CERN concentrates at present on the development of the sputter-coating technique as described in section 6.1, and on the assembly and testing of a complete prototype superconducting accelerator module for LEP.
In 1984, a 500 MHz three-cell cavity was tested at KEK in Japan in the TRISTAN Accumulation Ring. This was to demonstrate the feasibility of installing several tens of s.c. cavities to the TRISTAN Main Ring. With an accelerating field of 4.3 MV/m an electron current of 1mA was successfully accelerated to 5 GeV.

In March 1985, the most recent test of a s.c. system was carried out at DESY. A nine-cell 1 GHz cavity, operated at 4.5 K with an accelerating field of 2.5 MV/m, stored an electron beam of 2mA at 7 GeV in PETRA. This experiment is still going on and will try to answer the important question of the long term behaviour of a s.c. resonator in a storage ring. In addition, a development programme has been started recently to design, build and test a 500 MHz superconducting rf module for HERA. This is in preparation for a possible energy upgrade of the electron ring.

At CORNELL, a 1.5 GHz s.c. cavity system was tested in CESR in the last months of 1984. There, the accelerating fields in the multicell structures were the highest achieved so far. The 15.3 MV/m of the laboratory test of a fully equipped five-cell cavity show the virtues of the new high-thermal-conductivity niobium and they come close to the parameters

Fig. 28 Present status of the superconducting 130 MeV Recyclotron for electrons at the Technische Hochschule Darmstadt. The small quadrupole and bending magnets of the recirculation system are seen in the foreground, behind which the first cryostats of the superconducting linac and injector are assembled. The injector cryostat (background) contains one five-cell and two twenty-cell superconducting cavities (see Table 2) which are expected to have their first system test in the first half of 1986.
necessary for superconducting colliders. The research and development programme at CORNELL is aiming at this next step.

At SLAC tests are conducted with superconducting lead, niobium and Nb$_3$Sn cavities at 2.85 GHz, and which are driven by very short (~1 μs) rf pulses. The results obtained indicate that in this mode of operation fields close to the critical values can be consistently reached. The advantages and disadvantages of the pulsed method have been recently discussed in an excellent review).

At CEBAF in Newport News, Virginia, one is presently planning to convert the normal conducting design of a 4 GeV c.w. electron accelerator to a superconducting facility. Approval and funding of this project is expected and will have a very positive influence on the further development of rf superconductivity.

At the Technische Hochschule Darmstadt, a 130 MeV superconducting Recyclotron for electrons is under construction by a Darmstadt-Wuppertal Collaboration. The refrigeration system and the cryostat for the linear accelerator, consisting of a 10 MeV injector and 40 MeV accelerating section, is presently set up and shown in Fig. 28. Five accelerating structures for this accelerator have been tested so far. Two of them reached more than 5 MV/m, the design Q of $3 \times 10^9$ being obtained in all cases. They were fabricated in 1984 from stock niobium with low thermal conductivity and a purity typical of niobium sheet material produced before 1984. The first test of the 10 MeV injector is expected in the first half of 1986. Parallel to this programme, work in Wuppertal is presently concentrated onto the improvement of cavity performance with regard to the limiting fields, the development of Nb$_3$Sn resonators and to metrology.

* * *

REFERENCES


22) W. Meissner and R. Ochsenfeld, Naturwiss. 21, 787 (1933).
31) V.L. Ginzburg and L.D. Landau, Zh. Ekspert. i. Theor. Fig. 20, 1064 (1950).
33) H. Pfister, Cryogenics 16, 17 (1976).


40) T. Weiland, DESY 82-015 and DESY M-82-24.


42) W. Weingarten, ibid. ref. 9, p. 551.

43) T. Weiland, DESY 83-005 and DESY M-83-17.


45) H. Piel, ibid. ref. 15, p.85


49) D. Huppelsberg, Diplomarbeit, University of Wuppertal, WUD-85-3 (1985).


51) H. Padamsee, ibid. ref. 15, p. 145.


54) H. Padamsee, ibid. ref. 53, p. 149.


60) Ph. Niedermann, N. Sankararaman, R.J. Noer, Ø. Fischer, ibid. ref. 19, p.583.


63) G. Arnolds-Mayer, ibid. ref. 9, p. 643.


72) I.E. Campisi, ibid. ref. 53, p. 134.


HIGH-FIELD ELECTRON LINACS

J. Le Duff
Laboratoire de l'Accélérateur Linéaire, Orsay, France

ABSTRACT
High-field electron linacs are considered as potential candidates to provide very high energies beyond LEP. For almost twenty years little improvement has been made on linac technologies as they have mostly kept at low and medium energies to be used as injectors for storage rings. Today, both their efficiency and their performances are being reconsidered, and for instance the pulse compression scheme developed at SLAC and introduced to upgrade the energy of that linac is a first step towards a new generation of linear accelerators. However this is not enough in terms of power consumption and more development is needed to improve both the efficiency of accelerating structures and the performances of RF power sources.

1. INTRODUCTION

After introducing briefly the needs for higher gradient electron linear accelerators by showing that simple extrapolation of present technologies would fail in trying to reach much higher energies, we shall review different ways of improving these conventional techniques and their related problems. Since high gradient means also high RF power we shall also present and discuss a new type of RF power source based on the double aim of reaching much higher peak power with short pulses and having much higher efficiencies.

The present report will not deal with totally new accelerating techniques such as laser-plasma and wake-field accelerators, since they are taken up by other lecturers.

2. EXTRAPOLATION OF PRESENT TECHNOLOGIES

There are three large electron/positron linacs operating in the world (Table 1) as injectors for storage rings (although LAL and SLAC were initially built as high-energy physics facilities).

<table>
<thead>
<tr>
<th>Laboratory</th>
<th>Energy [GeV]</th>
<th>Accelerating length [m]</th>
<th>overall length [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAL/ORSAY</td>
<td>2.3</td>
<td>230</td>
<td>360</td>
</tr>
<tr>
<td>KEK/TSUKUBA</td>
<td>2.5</td>
<td>320</td>
<td>400</td>
</tr>
<tr>
<td>SLAC/STANFORD</td>
<td>24</td>
<td>3050</td>
<td>3350</td>
</tr>
</tbody>
</table>

Table 1 Large linacs in the world

The accelerating gradient lies in the range of 8 to 14 MeV/m. Recently the SLAC linac has been upgraded to 33 GeV and soon is expected to reach 50 GeV with new klystrons followed by a pulse compression system\(^1,2\). In the last mode of operation the accelerating gradient will
be as much as 17 MeV/m. Two bunches, electrons and positrons, will be simultaneously accelerated, then transferred in the two arms of a circular transport system in such a way that they will collide once at a given location. This will be the first linear collider (SLC) coming into operation in the world, at an energy level comparable with LEP stage 1. It will serve as a test bed for future linear colliders as well as for studying the intermediate boson $Z_0$.

In order to reach many hundred GeV or a few TeV in the center of mass with electrons and positrons, it appears that linacs are better suited than storage rings since circular machines would lead to enormous power being radiated in the bends; remember that in LEP, operating at 100 GeV per beam, each particle will loose 2.8 % of its energy per turn. Clearly the SLC scheme, with a single linac plus a circular transport system, will be also avoided for higher energies, and future linear colliders will consist of two linacs firing against each other.

Consider now a first step in the linac energy, by roughly one order of magnitude, using present SLAC technology for the accelerating sections. Table 2 gives the resulting constraints and three possible schemes have been considered, knowing that:

$$E_{\text{acc}} = \frac{1}{2} \left( \frac{P_{\text{input}}}{f_{\text{rep}} N_K n_K P_K} \right)^{1/2}$$

where $E_{\text{acc}}$ is the accelerating gradient, $P_{\text{input}}$ the input RF power at the structure, $f_{\text{rep}}$ the linac repetition frequency, $N_K$, $P_K$, $n_K$ being respectively the total number of klystrons, the peak power and the efficiency of each klystron.

<table>
<thead>
<tr>
<th></th>
<th>SLAC</th>
<th>Super SLAC (1)</th>
<th>Super SLAC (2)</th>
<th>Super SLAC (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E [GeV]</td>
<td>30</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>L [Km]</td>
<td>3</td>
<td>30</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$N_K$</td>
<td>240</td>
<td>2400</td>
<td>240</td>
<td>960</td>
</tr>
<tr>
<td>$P_K$ [MW]</td>
<td>38</td>
<td>38</td>
<td>3800</td>
<td>950</td>
</tr>
<tr>
<td>$E_{\text{acc}}$ [MV/m]</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$f_{\text{rep}}$ [Hz]</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td>180</td>
</tr>
<tr>
<td>sections/klystron</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>pulse length [$\mu$s]</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>$P_{\text{a.c.}}$ [MW]</td>
<td>13.7</td>
<td>137</td>
<td>1370</td>
<td>1370</td>
</tr>
</tbody>
</table>

Table 2 Extrapolation of SLAC up to 300 GeV

Clearly a higher gradient keeps the linac length to a reasonable value but introduces strong constraints on the power. For instance new power sources need to be developed at the level of 1 GW peak. The state of the art in klystrons is 50 MW peak with 5 $\mu$s RF pulse
length which can be compressed to increase
the effective peak power by a factor 4. A
150 MW, 1 μs klystron is also under develop-
ment at SLAC\textsuperscript{3}).

Consider another example of a high gradient
linear collider using the CERN site. On
Fig. 1 two linacs have been drawn along a
LEP diameter as if the circumference of LEP
were the ultimate possible size for a new
accelerator (this is not a statement). Table 3
shows the possible stages in energy that can
be considered for this example\textsuperscript{4}). The first
case uses conventional technologies such as
iris loaded travelling wave structures and
50 MW, 5 μs klystrons followed by a pulse compression system. However the power consumption

<table>
<thead>
<tr>
<th>Gradient MV/m</th>
<th>Energy/beam GeV</th>
<th>a.c. Power/linac MW</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>100</td>
<td>200</td>
<td>conventional</td>
</tr>
<tr>
<td>125</td>
<td>500</td>
<td>1500</td>
<td>need for new power sources</td>
</tr>
<tr>
<td>250</td>
<td>1000</td>
<td>-</td>
<td>need for new power sources and new structures</td>
</tr>
<tr>
<td>1250</td>
<td>5000</td>
<td>-</td>
<td>new accelerating methods</td>
</tr>
</tbody>
</table>

Table 3 Possible stages for a linear collider along a LEP diameter

has been minimized by matching the accelerating structures to the pulse compressor (see
next section), but even so it is too high to compete with the LEP storage ring at the
same energy. The second case uses an accelerating gradient of 125 MeV/m which has been
already reached on experimental structures. However it can only work if one uses 1 GW, short
pulse RF power sources which do not exist yet. In both cases a linac repetition rate of
1000 Hz has been considered.

Improvements on the power consumption may come from improvements in the efficiency of
accelerating structures and also from some tricks such as for instance the use of pulse
current trains that can lower the repetition rate for the same luminosity\textsuperscript{5}).
3. RF COMPRESSION SCHEME

3.1 Present situation

Although this scheme can be considered as an already existing technique, it is worthwhile to recall the principles since one can expect to improve the efficiency of the system in the near future.

The pulse compression is schematically represented on Fig. 2; the first part of the

![Diagram](image)

Fig. 2 Pulse compression scheme

long pulse from the klystron is stored in a couple of low loss cavities. At a given time \( t_1 \), the input signal to the klystron is rapidly shifted so that the energy is now reflected at the entrance of the storage cavities and directly goes to the structure. In addition the stored energy flows out of the cavities and also goes to the structure making the peak energy during the time interval \([t_1, t_2]\) much higher than it would have been from a direct feed of the structure.

The method can be either used to increase the energy of an existing linac, or to save on the total number of power sources for a given output energy:

**SLED (SLAC Energy Doubler)**

Storage cavities are placed between the klystrons and the accelerating structures. The RF pulse length is 5 µs and since the filling time of the structure is .8 µs one can adjust the switching time such that \( t_2 - t_1 = .8 \) µs. However the compression scheme has a poor efficiency and the maximum improvement factor on the peak power that one can expect is about 3 leading to an improvement factor \( \sqrt{3} \) on the accelerating gradient.

**LIPS (LEP Injector Power Saver)**

The scheme is used to reduce the number of klystrons by a factor 2 as shown on Fig. 3. Instead of feeding two structures from a single klystron, one can feed 4 structures with the
same total beam energy if the system is adjusted to increase the effective peak power by a factor 2. This improvement factor is lower than the previous one and can be obtained from a 3.5 μs klystron pulse length, the filling time of the LEP injector Linac (LIL) structures being 1.2 μs. It will be seen later that the improvement factor mainly depends on the initial klystron pulse length when all other parameters, such as for instance the filling time of the accelerating structures, have been optimized.

![Diagram](image)

**Fig. 3** The LIPS scheme

Up to here the maximum improvement factor one can reach with a conventional electron linac, by adding a compression system, is of the order of 1.7 and from there existing linacs might be able to operate with an accelerating gradient of the order of 35 MV/m, taking into account also the up-to-date klystrons and assuming that each klystron feeds a single structure. As will be seen in the next section, an additional improvement factor can be obtained by optimizing the parameters of the accelerating structures to match properly the compression system.

### 3.2 - Optimization of TW accelerating structures for SLED operation

The RF pulse shape due to the compression scheme, worked out by Z.D. Parkas et al. 2), is shown on Fig. 4. There are two regions of interest: region 1 which corresponds to a continuous increase of the stored energy in the cavities and region 2 which, after a π phase shift at the klystron input occurring at time $t_1$, corresponds to a relatively fast decay of the stored energy in these cavities to the benefit of the accelerating field. **Fig. 4** Pulse shape due to compression

For a unit rectangular klystron pulse the combined field entering the accelerating section is:

$$E_1(t) = (\alpha - 1) - \alpha e^{-t/\tau_c} \quad \text{for} \quad 0 < t < t_1$$
\[ E_2(t) = \gamma e^{-\frac{t-t_1}{\tau_c}} - (\alpha - 1) \quad \text{for } t_1 < t < t_2 \]

where

- \( \tau_c = \frac{2Q_c}{\omega(1+\delta)} \) is the filling time of the storage cavities
- \( Q_c \) is the unloaded quality factor of the storage cavities
- \( \delta \) is the coupling coefficient of the storage cavities
- \( \alpha = 2\delta/(1+\delta) \)
- \( \gamma = \alpha(2 - e^{-t_1/\tau_c}) \).

For a given peak power \( P \) from the klystron (rectangular pulse) the accelerating field at the input of the section would be:

\[ E = E_0 \cdot E_{1,2} \]

\[ E_0 = \left( \frac{\omega \tau_o}{P \cdot v_g \cdot Q} \right)^{1/2} \]

where \( \tau_o, v_g, Q \) are respectively the shunt impedance per unit length, the group velocity and the quality factor of the first accelerating cell.

In a constant impedance structure all the cells are identical, and hence \( r, v_g, Q \) will remain constant along the structure.

Due to power dissipation in the cells the amplitude of the propagating field will decrease exponentially. At a given azimuth \( z \) the field becomes:

\[ E(z, t) = E_{1,2} \left[ t - \Delta t(z) \right] e^{-(\omega/2v_g Q)z} \]

where index 1,2 refers to the two different time intervals as previously defined. Here again the expression needs to be multiplied by \( E_0 \) for a given peak power \( P \) from the klystron.

\( \Delta t(z) \) is the wave propagation time from the origin up to \( z \):

\[ \Delta t(z) = \int_0^z \frac{dz}{v_g(z)} = \frac{z}{v_g} \]

It looks interesting to use the normalized variable \( z' = \frac{z}{L} \) where \( L \) is the length of the structure. Then:

\[ \Delta t = \tau_a z' \]

with

\[ \tau_a = \frac{L}{v_g} \]

Depending whether the time \( t - \Delta t \) appears to be below or above \( t_1 \), the field \( E_1 \) or \( E_2 \) should be used. That tells us that a field discontinuity will appear at some location \( z_1 \) in the structure such that:

\[ z_1 = \tau_a \]
\[ t - \Delta t = t_1 \]
or \[ t - t_1 = \tau_a z'_1 \].

If \( z'_1 < 1 \) the energy gain along the structure is the contribution of two field integrals:

\[
V(t) = \int_{0}^{z'_1} E_2(t - \Delta t(z')) e^{\frac{-\omega t}{2Q} z'} dz' + \int_{z'_1}^{1} E_1(t - \Delta t(z')) e^{\frac{-\omega t}{2Q} z'} dz'
\]

where now \( t \) represents the time at which the particle traverses the structure (the transit time of the particle is negligible compared to the filling time of the structure).

Let us call \( V_1 \) and \( V_2 \) the integrals relative to \( E_1 \) and \( E_2 \). One gets:

\[
V = V_1 + V_2
\]

\[
V_1(z'_1) = -(a - 1) \tau_1 \frac{T_1}{\tau_a} \left[ e^{\frac{-\tau_a}{\tau_1}} - e^{\frac{-\tau_a}{T_1} z'_1} \right] - a \tau_2 \frac{T_2}{\tau_a} e^{\frac{-\tau_a}{T_2}} \left[ e^{\frac{-\tau_a}{T_2} z'_1} - e^{\frac{-\tau_a}{T_1} z'_1} \right]
\]

\[
V_2(z'_1) = (a - 1) \tau_1 \frac{T_1}{\tau_a} \left[ e^{\frac{-\tau_a}{T_1} z'_1} - 1 \right] + \gamma \tau_2 \frac{T_2}{\tau_a} \left[ e^{\frac{-\tau_a}{T_2} z'_1} - e^{\frac{-\tau_a}{T_2} z'_1} \right]
\]

with:

\[
\frac{1}{T_1} = \frac{\omega}{2Q}
\]

\[
\frac{1}{T_2} = \frac{\omega}{\tau_c} - \frac{\omega}{2Q}
\]

It is interesting to look at the behaviour of the function \( V(z'_1) \) in the interval \( 0 < z'_1 < 1 \). It can be shown numerically that for each value of \( z'_1 \) there is a value of \( \beta \) which maximizes the energy gain. This has been taken into account in the plots of Fig. 5, where it appears that the maximum energy gain corresponds to \( z'_1 = 1 \), which means that the beam should enter the structure at time \( t = t_2 = t_1 + \tau_a \) and that the width of the compressed pulse must be equal to the filling time of the structure.

The study will continue by considering only this optimum case for which:

\[
V_1 = 0
\]

\[
V = V_2(z'_1 = 1)
\]
Fig. 5 Multiplication factor for a constant impedance structure as a function of the beam timing

Fig. 6 Multiplication factor versus the filling time of a constant impedance structure
This leads to:

\[ V_M = (\alpha - 1) \frac{T_1}{T_a} \left[ \frac{T_a}{T_1} - 1 \right] + \gamma \frac{T_2}{T_a} \left[ \frac{T_a}{T_1} - \frac{T_a}{T_c} \right] \, , \]

However this is not the exact energy multiplication factor since for a unit pulse entering a constant impedance structure the energy gain over a unit length is:

\[ V_O = \frac{T_1}{T_a} \left( 1 - e^{-\frac{T_a}{T_1}} \right) \, . \]

Hence the real multiplication factor is the ratio \( V_M / V_O \). For each value of \( T_a \), there is a value of \( \beta \), hence a value of \( T_c \), which maximizes this multiplication factor as seen on Fig. 6.

A similar treatment for the case of a constant gradient structure would show that the efficiency of this type of structure is either the same, at low filling time, or slightly less, at high filling time, than the efficiency of a constant impedance structure. Hence we shall proceed with constant impedance structures in what follows.

It has been seen that for a given structure length there was an ensemble of optimum values for \( \beta \), \( T_c \) and \( T_a \) which realize the correct matching between the SLED pulse and the accelerating structure. It is interesting to look in more detail at the performances of these structures versus different parameters, like the parameter setting of the storage cavities \( Q_c, \beta \), the length and the aperture of the accelerating structures, the width of the direct peak power pulses from the klystrons.

For a constant impedance structure, fed by a klystron peak power pulse \( P \), \( T_2 \), through a couple of storage cavities with a \( \pi \) phase shift at time \( t_1 = t_2 - T_1 \), the energy gain is:

\[ V = \left( \frac{p \cdot B \cdot w}{Q} \frac{T_a}{T} \right)^{1/2} \frac{V}{T_a} \left( \frac{1}{T_c} - \frac{\omega}{2Q} \right)^{-1} \left[ - \frac{\omega T_a}{2Q} - \frac{T_a}{T_c} \right] + (a - 1) \frac{2Q}{\omega T} \left( - \frac{\omega T_a}{2Q} - 1 \right) \, , \]

where \( R = rL \) is the total shunt impedance of the structure, and \( T_a = L/\nu \) its filling time.

The fact that for a given length there is an optimum value for \( T_a \) means that there is an optimum value for \( v_q \), hence for the iris aperture \( 2a \) of the structure. To illustrate this point let us consider the cell characteristics of the LEP injector linacs (LIL)\(^8\) which operate at 3 GeV in the \( 2\pi/3 \) mode:

\[ \begin{align*}
Q & = 15200 \\
r & = 86 - 3.6 \times (2a)^2 \\
v_q/c & = (2a)^{1.23}/891
\end{align*} \]

where \( 2a \), the iris diameter, is expressed in cm while the shunt impedance \( r \) is in M\(\Omega\)/m.
Figure 7 shows the evolution of the RF performances versus the iris diameter, for different structure lengths. As the length decreases the iris diameter also decreases in order to get the maximum gain corresponding to the right matching value for $\tau_a$. In all cases $\beta$ and $\tau_c$ have been optimized.

The maximum energy gains obtained for each structure length are plotted on Fig. 8 as well as the corresponding values of $\tau_a$ and $\tau_c$ which clearly remain constant.

A systematic study of the energy gain as a function of the other parameters, like $t_2$, $Q_C$ and $Q(\tau=\text{cte})$ leads to the following conclusions:

- neither $Q$ nor $Q_C$ have influence on the optimum value of $\tau_a$. Both give a little effect on the optimum energy gain. The optimum value of $\tau_c$ changes with $Q_C$.

- the optimum value of $\tau_a$ changes with the width $t_2$ of the direct klystron wave. For long pulses one can hold a longer filling time, but that means a smaller aperture for a fixed structure length. An important increase of the energy gain follows an increase of $t_2$.

- one of the most important features, considering the results plotted on Fig. 8, is that the total energy gain from one klystron source will be higher if the power is shared between smaller structures, for the same total length. This fact is illustrated on Fig. 9, assuming no power losses in the RF networks, and knowing that the energy gain follows the square root of the input power. Of course smaller structures, when optimized, will have smaller apertures and the interesting result is that the minimum structure length will directly depend on the beam aperture requirement. For instance a minimum aperture of 1.8 cm would lead to a design length of 1.8 m for LIL type cells, according to Fig. 10.

In order to design an optimum linac structure under SLED operation, it is useful to draw design curves having the main design parameters, $P_{\text{klystron}}$, $Q$, $Q_C$ and $t_2$. Such a design example is shown on Fig. 11. If one introduces a design constraint such as $(2a)_{\text{min}} = 2.0\,\text{cm}$ one gets directly the remaining design parameters which in the present case are:

$$\begin{align*}
\tau_c &= 2.12 \, \mu\text{s} \\
L &= 2.5 \, \text{m} \\
\tau_a &= 0.8 \, \mu\text{s} \\
\beta &= 8 \\
t_1 &= 4.2 \, \mu\text{s}
\end{align*}$$

If a single structure is fed by one klystron the average accelerating gradient becomes 52 MeV/m. A smaller value for $(2a)$ would lead to a higher gradient, for instance $2a = 1.65\,\text{cm}$ gives 75 MV/m and the corresponding structure length is 1.3 meters.

Finally with short constant impedance structures optimized to match the SLED conditions, and commercially available klystrons one can get close to 100 MeV/m in a short term future.
Fig. 7  Energy gain of a constant impedance structure versus the iris diameter and the structure length

Fig. 8  Maximum energy gain as a function of the structure length

Fig. 9  Total energy gain from a single klystron as a function of the number of structures, for a given total length

Fig. 10  Optimum aperture of an iris loaded structure versus the structure length
4. ULTIMATE ACCELERATING GRADIENTS IN CONVENTIONAL STRUCTURES

A question can be raised now: can we reach in practice the gradient previously mentioned and can we go even further?

The answer to the first part of the question is mainly related to breakdown limits in warm structures and will be treated in this section. If no limitation occurs one way to go further consists of improving both the efficiency of accelerating structures and the peak power of RF sources (their efficiency too) and this will be treated in the next two sections.

4.1 - The Kilpatrick criterion

Breakdown phenomena may occur at high field level on the walls of accelerating structures and they are not very well understood at microwave frequencies. The study done by Kilpatrick\(^9\) is one of the few investigations of breakdown phenomena and was in the past very often referred to by accelerator designers. He empirically derived a relationship between frequency and maximum electric field:

\[ f = 1.643 \frac{E_{\text{max}}^2}{\exp(-8.5/E_{\text{max}})} \]

where \( f \) is the RF frequency in MHz and \( E_{\text{max}} \) is the maximum electric field in MV/m. At \( f = 3000 \text{ MHz} \) this relation predicts \( E_{\text{max}} = 46.8 \text{ MV/m} \).

The corresponding maximum accelerating field now depends on the type of structure. For instance disk loaded waveguides have a ratio \( E_{\text{wall}}/E_{\text{acc}} \) of the order of 2, hence the maximum expected gradient would be 23 MV/m. This could be one reason why accelerating gradients have been kept below this value for a long time, but certainly another good reason was that the klystron peak power was still low and long accelerating structures were making a better use of this power in terms of maximum beam energy per klystron (no SLED). The overall linac length was not a big worry at that time.
Accelerating structures with higher shunt impedances would lead to lower maximum accelerating gradients since, as a matter of fact, the ratio $E_{\text{max}}/E_{\text{acc}}$ increases when the shunt impedance increases.

Since recently the need for higher gradients became more and more obvious and new checks of the Kilpatrick criterion became of real concern at a few places. As a result it is now believed that the Kilpatrick criterion is pessimistic, at least under pulsed RF conditions.

4.2 - The experiment at Varian$^{10}$

The experimental set up is shown on Fig. 12, where a single nose cone cell is fed by a magnetron (2.6 MW, 4.4 µs).

![Diagram](image_url)

**Fig. 12** A cross-sectional view of the cavity test system

In this experiment the repetition rate could be varied between 70 and 300 pps while the output peak power could be varied from 0.2 to 2.6 MW.

The type of cavity which is used has a high shunt impedance, as much as 130 Ω/m, at 3 GHz, and the corresponding $E_{\text{wall}}/E_{\text{acc}}$ ratio is of the order of 3. The observed breakdown limit corresponded to an accelerating field of 30 MV/m and a maximum field of 240 MV/m on the inner surface of the cell. With different geometries corresponding to different $E_{\text{wall}}/E_{\text{acc}}$ ratios the maximum field was roughly the same.

It was also observed that above a certain level of wall polishing there was no effect on the breakdown limit. The limit also was found to be independent of the repetition frequency in the range previously mentioned.
From this experiment one can conclude that the maximum surface electric field can be at least as high as five times the limit predicted by Kilpatrick. Extrapolating this result to disk loaded cavities one can expect at least accelerating gradients close to 120 MV/m.

4.3 - Experiments at SLAC\textsuperscript{11,12)}

The first high gradient test at SLAC was done on the normal SLAC accelerating structures. In order to increase the gradient two klystrons operating in the SLED mode were combined, so that each of the four sections normally fed by one klystron could receive an input peak power as high as 87 MW (Fig. 13).

The corresponding SLED field in the sections was then up to 32 MV/m on the axis and 65 MV/m on the walls. At this level no breakdown occurred in the sections.

![Fig. 14 Resonant structure used for the second experiment. Test points indicate locations of thermocouples.](image)

In order to increase the gradient a second experiment was set up in which a short disk-loaded structure was designed to operate in the $2\pi/3$ S.W. mode (Fig. 14).

The cavity fed by 30 MW RF peak power did not show breakdown problems after a short processing. The maximum equivalent travelling wave accelerating and surface fields in these conditions were respectively 133 MV/m and 259 MV/m.

However it should be noticed that in this experiment considerable X-ray radiation was detected around the section corresponding to a strong field emission.

5. A SURVEY OF ACCELERATING STRUCTURES

Previous experiments tell us that accelerating gradients of the order of 100 MV/m can be achieved with conventional disk loaded structures, but this will need very high peak
power and correspondingly high average power to fit the luminosity requirements in a linear collider.

Efforts have already been made to improve the efficiency of accelerating structures and at least four types of accelerating structures, either operating in L-band or in S-band, have been developed for the acceleration of electrons (Fig. 15). One can make the following remarks:

- The disk loaded structure is very well known since it has been used for a long time in linac design. It has a relatively low shunt impedance but a very good ratio $E_{\text{max}}/E_{\text{acc}}$.

![Diagram of disk loaded structure](image)

a) Disk loaded

![Diagram of jungle gym or cross bar](image)
b) Jungle Gym or cross bar

c) Disk and washer

![Diagram of disk and washer](image)
d) Side coupled

Fig. 15 Different types of accelerating structures

- The jungle gym structure has been first developed at low frequency. Since it has no revolution symmetry it is hard to study this structure with computer codes, and hence it needs more prototype work. However, it is expected to get from this structure an improved shunt impedance with a high group velocity.

- The disk and washer structure is an open structure, as the previous one, which makes the wall losses smaller and correspondingly leads to a higher shunt impedance. It has also a higher $Q$ but not a higher $r/Q$.

- The side coupled structure has a very high shunt impedance but a very bad ratio $E_{\text{max}}/E_{\text{acc}}$. 
The last two structures are quite complicated to build, and up to now they have been mostly considered in the S.W. mode according to their high shunt impedance.

From the power consumption point of view it is well recognized that for a given type of structure, operation in the S.W. mode is less efficient (although not obvious when considering small linear accelerators) than operation in the T.W. mode if correct matching of the source is made in both cases. Hence it is still preferable to consider T.W. structures for very high energy linacs and in that case the parameters of real importance are $r/Q$, $E_{\text{max}}/E_{\text{acc}}$, and $v/c$. For these reasons it is believed that the jungle gym structure may become a good possibility but still needs more development. In the meantime the old disk loaded structure will remain a good candidate.

Another advantage of the T.W. accelerating structure comes from the fact that it can be used in the SLED mode. Moreover, if the group velocity is high the klystron pulse can be made very short and correspondingly the peak power can be increased which is the right direction to follow in the non SLED case.

Table 4 compares the performances of several structures in the T.W. mode, at different frequencies. The disk and washer is also shown for comparison.

<table>
<thead>
<tr>
<th></th>
<th>$r$ (mil/m)</th>
<th>$Q$</th>
<th>$v/c$</th>
<th>$L$ (m)</th>
<th>$\tau_a$ ($\mu$s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2856 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disk-Loaded ($a=1.16$ cm)</td>
<td>56</td>
<td>13,300</td>
<td>.012</td>
<td>3</td>
<td>.83</td>
</tr>
<tr>
<td>Disk-Loaded ($a=1.50$ cm)</td>
<td>46</td>
<td>13,000</td>
<td>.035</td>
<td>6</td>
<td>.57</td>
</tr>
<tr>
<td>Disk and Washer</td>
<td>76</td>
<td>32,000</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Jungle Gym ($\pi/2$)</td>
<td>51</td>
<td>9,000</td>
<td>.20</td>
<td>6</td>
<td>.10</td>
</tr>
<tr>
<td>Jungle Gym ($\pi/3$)</td>
<td>60</td>
<td>9,000</td>
<td>.10</td>
<td>6</td>
<td>.20</td>
</tr>
<tr>
<td>4040 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jungle Gym ($\pi/2$)</td>
<td>61</td>
<td>7,500</td>
<td>.20</td>
<td>6</td>
<td>.10</td>
</tr>
<tr>
<td>Jungle Gym ($\pi/3$)</td>
<td>71</td>
<td>7,500</td>
<td>.10</td>
<td>6</td>
<td>.20</td>
</tr>
<tr>
<td>5712 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jungle Gym ($\pi/2$)</td>
<td>72</td>
<td>6,500</td>
<td>.20</td>
<td>6</td>
<td>.10</td>
</tr>
<tr>
<td>Jungle Gym ($\pi/3$)</td>
<td>85</td>
<td>6,500</td>
<td>.10</td>
<td>6</td>
<td>.20</td>
</tr>
</tbody>
</table>

Table 4 comparison of structure for a collider

It is still worthwhile developing short disk loaded structures in the frame of improved power sources.

6. RF POWER SOURCE: THE LASERTRON

Up to now pulsed klystrons have been used to provide high RF peak power to electron accelerating structures. Peak powers up to 50 MW within 5 $\mu$s pulse length have already been achieved but the efficiency of these devices is still below 50%. A higher peak power klystron, about 150 MW, at a shorter pulse length, about 1 $\mu$s, is under development at SLC.
and it seems very difficult to go much higher. As a matter of fact high beam power either requires a higher beam current which would create strong space charge effects, or a higher accelerating voltage which would reduce the bunching efficiency, hence the amount of power contained in the fundamental and the extraction efficiency.

To overcome these difficulties, although not proven to be fundamental limitations, a new microwave RF power tube has been recently proposed\(^{15,16}\) in which a photocathode, illuminated by a modulated laser, emits short, dense current pulses which, after being accelerated, traverse an output cavity where the RF beam modulation is extracted (Fig. 16).

Here, a high accelerating voltage is necessary to compensate for the space charge forces which otherwise would distort the emitted short bunches and reduce the extraction efficiency. Since in principle the laser can provide a train of such bunches with a given repetition rate, the accelerating voltage can be d.c.

![Schematic of a photocathode microwave power source](image.png)

**Fig. 16** Schematic of a photocathode microwave power source

Considering the fast pulsed photo-emission it is believed that the maximum charge which can be extracted per pulse from the photocathode is equal to the superficial charge\(^{17}\):

\[
Q = \varepsilon_o E A S = \frac{\varepsilon}{d} S V = CV
\]

where \(E_A\) is the accelerating field at the photocathode, \(S\) the useful area of the photocathode, \(d\) the distance between the cathode and the anode, \(C\) the gun capacitance and \(V\) the accelerating voltage.

This maximum charge is twice the space charge limit, showing that the limitation of such a tube is very different from that of a klystron. As a matter of fact, if \(f_{RF}\) is the repetition frequency of the laser pulses, the average current per laser burst is:

\[
I_o = f_{RF} CV
\]

and the beam power:

\[
P_b = f_{RF} CV^2
\]

while in a klystron the maximum current is related to the voltage through a parameter \(k\) called perveance:

\[
k = \frac{I_o}{y^{3/2}}
\]

\[
P_b = k y^{5/2}
\]

A 2-D simulation of the lasertron has been already performed\(^{18}\) which, for a given accelerating voltage, shows an increase of the energy spread and an increase of the bunch length above a certain average beam current, or beam power (Fig. 17).
A corresponding decrease of the ratio $I_1/I_0$, where $I_1$ is the amplitude of the first harmonic of the current modulation, is shown on Fig. 18, leading to a decrease of the efficiency and to a saturation of the extracted RF power. This is shown on Fig. 19 for the case of the prototype under consideration at SLAC\(^{19}\). Improvement of the saturation level would follow an increase of the accelerating voltage.

The main parameters of the SLAC prototype\(^{20}\) are given in Table 5. The power level is comparable to the peak power of the best klystron, and this is a first step to check the lasertron principle before envisaging much higher peak power.
Prototypes are also under consideration in Japan and in France.

Among the difficulties encountered in designing a lasertron it is worthwhile mentioning the high current photocathode. Remembering the poor efficiency of lasers, the photocathode must have a very good quantum efficiency. Unfortunately it happens that efficient cathodes, like AsGa for instance, show poor lifetime. On the contrary, metallic cathodes are robust but with a poor quantum efficiency.

Modulated lasers at S-band or C-band frequencies, with long pulse trains and high repetition rate have to be developed also, with optical frequencies either in the visible (green) or in the UV.

Figure 20 taken from reference 20) gives a good idea of the lasertron geometry as well as the technologies involved.
REFERENCES

1) SLAC Linear Collider (SLC) : Conceptual design report, SLAC-Report 229 (1980).


5) F. Bulos et al., Physics with linear colliders in the TeV CM energy region. SLAC-PUB-3002. Also, contribution to the Proceedings of the Summer Study on Elementary Particle Physics and Future Facilities, 1982, Snowmass, CO.


8) R. Belbeoch et al., Rapport d'études sur le projet des linacs injecteurs de LEP (LIL); LAL/RT/82-01, Orsay (1982).

9) W.D. Kilpatrick, Criterion for vacuum sparking designed to include both RF and DC. UCRL-2321, (1953).


19) P.B. Wilson, Private communication.

FREE ELECTRON LASERS: A SHORT REVIEW OF THE THEORY AND EXPERIMENTS

G. Dattoli, A. Renieri and A. Torre*)
DIP. TIB, Divisione Fisica Applicata, Centro Ricerche Energia Frascati,
C.P. 65 - 00044 Frascati, Rome (Italy).

ABSTRACT
This note is devoted to a short review of the theoretical and experimental aspects of Free Electron Lasers (FEL). We will discuss both recirculated and single-passage FELs and their relevant design problems.

1. INTRODUCTION

Since laser sources have been experimentally demonstrated the concept of a "universal" or "radio-like" coherent light source has been recognized as a powerful tool for a large number of potential applications.

The concept of fully tunable lasers is therefore as old as laser Physics.

The research activity in this field, developed through the years, is summarized in Fig. 1 where we have plotted the power against the wavelength (and wavelength range) of the commonly considered tunable conventional sources. Many of the light sources of Fig. 1 are far from being real tunable lasers. It is self-evident that the ultimate tunable laser has not been developed, but its desired performance can be easily outlined:

a) stability
b) long life
c) easily manageable
d) high power
e) easily tunable via external settings to any selected frequency.

These and many other "science-fiction" performances will be the characteristics of a real tunable laser. The areas of application of these kinds of sources are as wide as their versatility and include such different fields as spectroscopy, remote detection, photochemistry etc.1).

To give an example, conventional tunable lasers like excimers, dyes and harmonic generation have been used for UV spectroscopy, while color center lasers have been exploited in the middle IR spectroscopy and have provided substantial advances. The use of real tunable sources in connec-

*) Dept. of Physics and Astronomy, Dartmouth College, Hanover 03755 N.H. (USA)
Fig. 1 Comparative chart between FEL and conventional coherent sources. Average power vs \( \lambda \). Curves FEL (SP): single passage FEL average power vs \( \lambda \) 1st and 3rd harmonic respectively, maximum electron beam (e.b.) power 20 MW, \( K=1 \), \( \lambda_u=5 \) cm, \( N=50 \), \( L_c=6 \) m, \( \tau_M=12 \mu s \), \( \delta=5\% \). Straight curves: FEL storage ring average power vs \( \lambda \) without (continuous) and with (dashed) Touschek effect respectively. \( I=100 \) A and operating parameter of LEDA-F, see Ref. 2. \( L_c \) cavity length, \( \delta \) duty cycle.

...tion with photoacoustic spectroscopy and photothermal radiometry will result in more useful and flexible detection techniques \(^1\). In the field of photochemistry the tunable sources will allow more reliable analytic techniques and laser-based chemical processing ranging from controlled thermal chemistry to laser initiated radical reactions.

The above list of potential applications of tunable sources may be complemented with more technological subjects such as mechanical processing, information transfer and communication. The noticeable interest in the field is therefore fully justified. The state of the art of the tunable sources has been discussed in Ref. 2 where a comprehensive review of the experimental results and of the literature has been given. As a comment to Fig. 1 we notice that among the solid state lasers the most reliable is the Alexandrite, with a tuning bandwidth of about 1000 A, operating at room temperature. The color-center lasers have a more significant tuning range, but require improvements in terms of stability and lifetime. The most limiting factor of dye lasers is represented by the stability, even though their tunability over the years has been improved, to cover the range between the visible and the near IR. However, the power limitation beyond 10 \( \mu m \) is evident. In this region and in the short wavelength region (VUV, X-rays) tunable non-conventional sources are provided by the Free Electron Lasers (FEL) \(^3\).
In Fig. 2 the tunability curve of FELs is shown, the continuous line ranges from VUV to the microwaves. Although the basic mechanism of FEL allows a wide tunable range, this device, as it stands, does not provide the universal laser we are talking about. A fully tunable FEL requires, indeed, a "universal" accelerating electron machine able to provide an electron beam with continuous varying energies from MeV to GeV region. Even this kind of machine does not exist, but its characteristics for FEL application can be easily listed:

a) easy energy tunability  
b) modest size  
c) high beam power (average and peak)  
d) good beam qualities (small energy spread and emittances).

An overview of the design characteristics of an accelerating electron device dedicated to FEL has been presented in Ref. 4. In that paper the performances of storage rings, diode machines, induction linacs, electrostatic devices and R.F. accelerating machines have been discussed within the framework of their relevance to FEL. In Fig. 3 we have summarized the relative range of energy and current of accelerating devices. These are impressive, from MeV and kA to GeV and hundreds of mA.

* LBL-LLNL (LIVERMORE)  ○ STANFORD  
▽ UCSB (SINGLE STAGE)  ▼ NOVOSIBIRSK  
▲ UCSB (TWO STAGES)  ▲ ORSAY  
● AT&T-BELL  ◇ FRASCATI (INFN)  
△ FRASCATI (ENEA)  △ BROOKHAVEN  
□ UK PROJECT  ☉ MK III-STANFORD  
★ TRW STANFORD  ◎ YEREVAN  
▲ LANL (LOS ALAMOS)  ▼ SOR-RING  
Θ MSNW-BAC

Fig. 2 FEL scenario
Fig. 3 Current vs. Energy for existing accelerators

Such a flexible accelerator is very far from the present technological capabilities. Nevertheless, exploiting such different tools as relative energy tunability, higher-harmonics emission, undulator gap variations etc. an FEL can provide a range of tunability much larger than that of the conventional sources.

In the next sections we will briefly summarize the main features of the FEL theory with particular emphasis on the design criteria of both recirculated and linear devices. The final section is devoted to concluding remarks where we complete the comparison with the conventional lasers.

2. FEL: THEORY AND DESIGN CRITERIA

In an FEL a beam of ultrarelativistic electrons interacts with an undulator magnet (UM) where it undergoes transverse oscillations and emits radiation at a fixed wavelength; the radiation is stored in an optical cavity, it reinteracts with the copropagating e-beam and is amplified.

An undulator magnet is a spatial array of magnets arranged as in
Fig. 4, with spatial period $\lambda_u$ and it was originally proposed as a tool to enhance the brightness of the synchrotron light 5).

We will give a simple heuristic explanation of the central emission frequency in an undulator. Madey 6) has shown that for ultrarelativistic e-beam the undulator field can be treated as a radiation field with wavelength

$$\lambda^* = 2\lambda_u \tag{1}$$

and density number of pseudophotons 7)

$$\bar{n} = \frac{\alpha}{4} \frac{\lambda_u^2}{\lambda_u^2 - \kappa^2}, \quad \kappa = \frac{e \tilde{B}}{2\pi m_0 c^2} \tag{2}$$

where $\alpha$ is the fine structure constant, $r_o$ is the classical electron radius, $\kappa$ is the undulator parameter and $\tilde{B}$ (= $B_o$ for helical undulator, $B_o/\sqrt{2}$ for linear undulator) is the average on axis field ($B_o$ is the on axis field).

The relation (1) can be understood as follows. The vector potential of the undulator field can be written as

$$A = \{ \Re \{ B_o \lambda_u \exp(iz/\lambda_u) \} \} \hat{y} \tag{3}$$

As a consequence the magnetic field is given by

$$B = \nabla \times A = \Re \{-iB_o \exp(iz/\lambda_u) \} \hat{x} \tag{4}$$

($\hat{x}, \hat{y}, \hat{z}$ = unit vectors).

From (4) we get also

$$\nabla \times B = \nabla \times \nabla \times A = -\nabla^2 A = 1/\kappa^2 \lambda_u A \tag{5}$$

![Diagram of Undulator Magnet Geometry](image-url)
which clearly contradicts the Maxwell equation. The last term on the right hand side of (5) can however be reinterpreted, according to Ref. 3, as a kind of "photon mass". This is indeed the price to be paid when one treats from the very beginning the undulator vector potential in the transverse gauge. Furthermore, transforming the vector potential to that seen by a relativistic electron moving through the UM we have

$$A' = \Re\{B_o \chi_u \exp[i \gamma/\lambda_u (z' + \beta ct')]\} \hat{y}$$

$$\beta = v_z/c, \quad \gamma^2 = (1 - \beta^2)^{-1}$$

(6)

thus yielding for the magnetic and electric field the expressions

$$B' = \Re\{-i\gamma B_o \exp[i \gamma/\lambda_u (z' + \beta ct')]\} \hat{x}$$

$$E' = \Re\{-i\gamma \beta B_o \exp[i \gamma/\lambda_u (z' + \beta ct')]\} \hat{y}$$

(7)

and finally obtaining the "Maxwell equation"

$$\nabla' \times B' = \frac{1}{c} \frac{\partial E'}{\partial t'} + \frac{1}{\lambda_u} A'$$

$$\nabla' \times E' = -\frac{1}{c} \frac{\partial B'}{\partial t'}$$

(8)

From the first of equations (8) it follows that the "mass term" is neglected once $\gamma >> 1$. Furthermore, since $\beta \sim 1$ the fields (7) approximate closely to those of a radiation field. Since a light pulse must be a radiation field in all the reference frames, the field defined by (7) must remain a "radiation" field in the laboratory frame too. Transforming back its wavelength to that frame one finds

$$\lambda' = (1 + \beta) \lambda_u$$

(9)

thus getting for $\beta \sim 1$ the relation (1). Finally the equation (2) is a straightforward consequence of the definition of the field energy density.

Since the undulator field can be treated as an ordinary electromagnetic wave, its interaction with the electron can be viewed as a scattering. According therefore to the well known formula of the Compton scattering, the wavelength of the scattered light at an angle $\theta$ is given by (see Fig.5)

$$\phi$$

$$\lambda_u$$

Fig. 5 Compton scattering diagram
\[
\lambda = 2\lambda u \frac{1 - \beta \cos \theta}{1 + \beta} .
\]  
(10)

Expanding for small angles and ultrarelativistic energies one finds

\[
\lambda = \frac{\lambda u}{2\gamma^2} (1 + k^2 + \gamma^2 \delta^2) .
\]  
(11)

The corrective term \( k^2 \) is due to the transverse electron motion and it is analogous to the effect suggested by Brown and Kibble in the analysis of the electron motion in an intense laser wave, where an intensity dependent contribution to the Compton wavelength shift was found (see Ref. 3 for further comments).

One of the most peculiar characteristics of the light emitted by a charged particle running in a UM is the bandwidth. This quantity can be easily evaluated according to the following simple argument:

1) The duration of the emitted light pulse is linked to the difference between electron and photon flight times (see Fig. 4)

\[
\Delta t = \frac{N\lambda u}{c} (1 - \beta) \equiv \frac{N\lambda}{c} .
\]  
(12)

(N is the number of undulator periods).

2) According to the indetermination principle the bandwidth can be easily evaluated from

\[
\Delta \omega \Delta t \equiv \kappa, \quad \Delta \omega \equiv \frac{\kappa c}{N\lambda} .
\]  
(13)

3) Combining both (11) and (13) we get the relative homogeneous bandwidth

\[
\left( \frac{\Delta \omega}{\omega} \right)_0 = \frac{1}{2N} .
\]  
(14)

The denomination "homogeneous" and inhomogeneous derives (with the same meaning) directly from the standard theory of the photon emission by atoms or molecules. As is well known, the spectrum of an N-period light pulse has a width given by (14), while its shape is

\[
f(\omega) \propto \left( \frac{\sin \nu/2}{\nu/2} \right)^2, \quad \nu = 2\pi N \frac{w_0 - \omega}{\omega_0} \left( \omega_0 = \frac{2\pi c}{\lambda} \right) .
\]  
(15)

The spectrum given by Eq. (15) plays a fundamental role in the FEL theory and
it has been plotted in Fig. 6 together with two experimental spectra (Stanford 8) and Orsay 9). The above brief comments are relevant to the spontaneous emission. For a recent and detailed analysis the reader is referred to Ref. 4. We are now interested in stimulated emission and gain. By the former we mean emission in the presence of other e.m. modes and variation of the intensities of the those modes. A rigorous analysis of the gain can be found in Ref. 3, but here we will evaluate the gain function in a rather direct way. The gain mechanism can be understood as a balance between an absorption and an emission photon process. The gain function will be therefore given by the difference between the probabilities of emitting and absorbing a photon. The functional form of emitting or absorbing a photon is the same as Eq. (15), the only distinguishing feature is the

---

Fig. 6 Undulator magnet forward emission spectrum: (a) Theoretical spectrum; (b) Experimental spectrum (Stanford, Ref. 8); (c) Experimental spectrum (Orsay, Ref. 9).
electron recoil, so that
\[ g(\omega) = f(\omega - \varepsilon) - f(\omega + \varepsilon) \propto \frac{\sin v/2}{\alpha (\frac{v}{2})^2} - \left(\frac{\sin 4/2}{\frac{v}{2}}\right)^2 \]

(16)

\[ v_\pm = 2\pi N \frac{\omega_0 (1 \pm (\omega \lambda_e/c \gamma)) - \omega}{\omega_0} , \quad \lambda_e = \frac{\lambda}{mc} . \]

At the lowest order in the small quantity \((\omega \lambda_e/c \gamma)\) we get \(^3\)

\[ g(\omega) = -\frac{4\pi^2}{\gamma} \frac{\lambda L}{\Sigma E} \frac{I}{I_0} \mathcal{F} \frac{k^2}{1+k^2} (\frac{\Delta \omega}{\omega_0})^2 \frac{d}{dv} \frac{\sin v/2}{(\frac{v}{2})^2} \]

(17)

where \(\Sigma_E\) is the e-beam cross section, \(\mathcal{F}\) is the filling factor

\[ \mathcal{F} = \begin{cases} \frac{1}{\Sigma_L / \Sigma_E} \quad & \text{if } \Sigma_E > \Sigma_L \\ \Sigma_E / \Sigma_L \quad & \text{if } \Sigma_E < \Sigma_L \end{cases} \]

(18)

\(\Sigma_L\) is the laser-beam cross section, \(L\) is the undulator length, \(I\) the e-beam peak current and \(I_0 = ec/r_0(= 1.7 \times 10^4 A)\) is the Alfvén current. Since emission at higher harmonics occurs in undulator magnets on and off axis, an analogous gain formula for higher harmonics can be derived (see Ref. 10 for a general formulation).

In particular, linear undulators allow odd-harmonic emission on axis and the relevant gain can be written as

\[ g_n(\omega) = -g_n^0 \frac{\pi}{n} \frac{d}{dv} \frac{\sin n\omega/2}{[n\omega/2]^2} , \quad n = 1,2,... \]

(19)

\[ v_n = 2\pi N \frac{n \omega_0 - \omega}{n\omega_0} \]

\[ g_n^0 = \frac{4\pi^2}{\gamma} \frac{\lambda L}{\Sigma E} \frac{I}{I_0} \mathcal{F} \frac{\Delta \omega}{\omega_0}^2 , \quad \lambda_n = \frac{\lambda}{n} \]

\[ \mathcal{F}_n = n-\text{th harmonic filling factor} \]

\[ F_n(\xi) = \xi[n (\frac{J_{n+1}(n\xi)}{J_n(n\xi)} - \frac{J_{n-1}(n\xi)}{J_n(n\xi)})]^2 , \quad \xi = \frac{1/2}{k^2 + k^2} \]

\((J_n(\cdot) \equiv n-\text{th cylindrical Bessel function})\).
The analysis we have developed so far is relevant to a small-signal, single-mode, homogeneously-broadened FEL operation. The strong signal and multimode behaviour will be treated below. By homogeneous broadening we mean an FEL operating with an electron beam whose energy spread and emittances produce negligibly small effects.

It is well known that those beam qualities produce both a broadening of the emission line and a reduction of the gain \(^3\). The value of the inhomogeneous linewidth in terms of the beam emittances and energy spread is \(^3\)

\[
\left( \frac{\Delta \omega}{\omega} \right) = \left( \frac{\Delta \omega}{\omega} \right)_0 \sqrt{1 + \mu_x^2 + \mu_y^2 + \mu_z^2}. \tag{20}
\]

The \(\mu\)-coefficients are the ratio between the inhomogeneous and homogeneous widths and have played a crucial role in the design and optimisation of an FEL device \(^1\).

In particular,

\[
\mu = 4N \sigma_e, \quad \sigma_e \equiv \text{r.m.s. energy spread}
\]

\[
\mu = \frac{4N \sigma_e^2}{1 + k^2} \left[ \frac{1}{2 \sigma_x^2} \frac{\varepsilon_x \varepsilon_y}{2 \pi} + 2 \left( \frac{k_\parallel}{\gamma_\parallel \nu} \right)^4 h_{x,y} \sigma_x^2 \sigma_y^2 \right] \tag{21}
\]

where \(\varepsilon_x, \varepsilon_y\) are the radial and vertical emittances, \(\sigma_{x,y}\) the transverse e-beam dimensions, \(h_{x,y}\) are coefficients depending on the undulator geometry, namely \(h_x = h_y = 1\) for helical undulators and \(h_x = 6, h_y = 2+6 (6<1)\) for the linear case, with polarization along the y-axis. Physically \(\delta\) is the magnitude of the sextupolar term along the x-direction \(^3\).

It is worth noting that the inhomogeneous broadening due to the emittances consists of two distinct contributions; the first due to the angular divergence, the second to the finite beam size which explores regions of different magnetic field strength.

The expression Eq.(21) suggests that one can choose an optimum \(\sigma_{x,y}\) to minimize the effect of the inhomogeneous broadening \(^3,\ast\) i.e.

\[
\sigma_{x,y} = \frac{1}{\pi} \left[ \frac{1}{2 |h_x, y|} \right]^{1/4} \left( \frac{\lambda u \varepsilon_x \varepsilon_y}{2 k} \right)^{1/2}. \tag{22}
\]

Therefore one finally gets

\[\ast\] Since, \(\sigma_{x,y} = \sqrt{(\varepsilon_x \varepsilon_y \beta_x \beta_y)/2\pi}\) where \(\beta_{x,y}\) is the beta amplitude function, Eq.(22) amounts to \(\beta_{x,y} = [1/2 |h_x, y|]^{1/2} (\gamma_\parallel \nu)/k\pi\).
$$\mu_{x,y} = 2N \sqrt{2} |h_{x,y}| \frac{k}{1+k^2} \frac{\gamma e_{x,y}}{\lambda_u}.$$  \hspace{1cm} (23)

We must remark, however, that while the choice of Eq. (22) minimizes the effect of the inhomogeneous broadening it may, at the same time, create difficulty with the filling factor. Therefore, inserting the auxiliary condition that the e.b. cross section is of the order of the laser mode waist, we get

$$\varepsilon_{x,y} \equiv \sqrt{2} |h_{x,y}| n \left( \frac{k}{\gamma} \right) N \lambda.$$  \hspace{1cm} (24)

A further condition on the emittance can be obtained requiring that $\mu_{x,y} \leq 1$ and thus

$$\varepsilon_{x,y} \leq \frac{\lambda}{\sqrt{2} |h_{x,y}|} N \kappa.$$  \hspace{1cm} (25)

Combining Eq. (23) and (24) one can also get a condition on the sextupolar terms $|h_{x,y}|$.

When a very small beam section is required we can neglect in Eq. (21) the inhomogeneous broadening induced by the undulator field inhomogeneities. With the requirement that the beam cross section is of the order of the laser waist and that (23) be less than unity we find the conditions

$$\varepsilon_{x,y} \leq \lambda, \quad \beta_{x,y} \sim N \lambda_u.$$  \hspace{1cm} (26)

To give an idea of the $\mu$ parameters on the spontaneous emission and the gain, we have plotted in Fig. 7 those functions against $\nu$ for different values of $\mu$. It is evident that with increasing values of the inhomogeneous parameters the curves are both widened and reduced.

Longitudinal mode locking arises for FEL operation with bunched e-beams. It has indeed been shown that in this situation a natural phase locking is induced by the FEL interaction, and the strength of the coupling between the modes is given by the further parameter

$$\mu_c = \frac{N \lambda}{\sigma_z}.$$  \hspace{1cm} (27)

where $\sigma_z$ is the electron bunch r.m.s. longitudinal length. The larger is $\mu_c$, the greater is the number of coupled modes. The bunched e-beam structure is also responsible for the so called FEL lethargic behaviour, i.e. the slow down of the light pulse due to the interaction and the necessity to shorten the cavity length with respect to the nominal round-trip period to keep the synchronization between light and electron bunches 3). The im-
Fig. 7 Inhomogeneous broadened gain: (a) $\mu_x = \mu_y = 0$; (b) $\mu_x = 1, \mu_y = 0$; (c) $\mu_x = 0, \mu_y = 1, \mu_y = 0$; (d) $\mu_x = \mu_y = 1$.

The importance of this effect for short-pulse operating FEL devices will be discussed below.

The above notions are the minimal theoretical background to understand the FEL operation; in the next two subsections we will describe in some detail FEL devices in storage rings and single-passage devices.

3. FEL STORAGE RING OPERATION

We have seen that the practical realization of a FEL requires an e-beam with good qualities, namely large peak current and relatively low energy spread and emittance. A storage ring (SR) provides a very good e-beam for the FEL operation.

In these devices the e-beam is continuously recirculated through the interaction region and as a consequence, the energy spread and the emittances increase. Therefore, according to the arguments presented so far, the increase of the inhomogeneous broadening reduces the FEL amplification. This dynamical behaviour is peculiar to storage ring FELs. A correct description of the storage ring FEL operation requires indeed the self-consistent analysis, turn by turn, of both the laser and the electron beams. Storage ring FELs have been suggested for laser operation in the short wavelength region from visible down to VUV and x-ray (see Fig. 2). Just to start these introductory remarks we show in Fig. 8 the layout of an SR designed for FEL operation. The machine has a twofold symmetry (problems arise for non-symmetrical structures). The symmetry is provided by the insertion of two, long, low-field undulators (for the FEL operation) and two, short, high-field undulators to enhance the synchrotron radiation...
emission (this fact will be clarified below).

The focalization is provided by an alternate distribution of horizontal focusing and defocusing quadrupole magnets (F and D in Fig. 8). Also inserted in each quarter of the machine are two bending magnets and two sextupoles to minimize the dependence of the transverse oscillation frequency on the particle energy \(^{12}\). The free space between the quadrupoles of the long straight section is utilized for the injection of the electrons into the machine and to insert the RF accelerating system which accelerates the electron to higher energies than the injection one and supplies the energy lost by synchrotron emission in the bending magnets.

Since the particles emit synchrotron radiation the off-energy particles tend to reduce, turn by turn, the energy shift from the synchronous ones with a damping time

\[
\tau_d \sim \frac{TE_0}{U_0}, \quad U_0 = -\frac{4\pi}{3} \frac{r_0}{\rho} \gamma^4 m_0 c^2
\]  

(28)

where \(E_0\) is the machine nominal energy, \(T\) the revolution period, \(U_0\) the energy radiated per turn and \(\rho\) the bending magnet radius (assumed identical for all the magnets).
The betatron motion too is damped, with damping times

\[ \tau_x = \tau_y = 2\tau_z \quad . \quad (29) \]

The above expressions are only approximate. The correct ones involve the so-called damping partition numbers \( J_i \), namely

\[ \frac{T_i}{\tau_i} = J_i \frac{U_0}{2E_0} \quad (i = x, y, z) \quad . \quad (30) \]

According to the Robinson Theorem \(^{13}\) the \( J \) numbers obey the following identity

\[ \sum_{i} J_i = 4 \quad . \quad (31) \]

In any case, Eqs. (29) and (30) are good approximations for a typical plain machine \( (J_y = 2, J_x = 1, J_y = 1) \). The exact expression should contain a small correction to take into account the (eventual) radial gradient in the bending magnets.

After these few remarks on SR physics let us briefly discuss what are the achievable e-beam qualities.

3.1. Emittances

The smallest emittances in an S.R. are achieved with the magnetic lattice suggested by Chasman, Green and Rowe \(^{14}\). Such a magnetic structure consists of \( M \) achromatic bends and, according to Krinsky \(^{15}\), Sommer \(^{16}\) and Potaux \(^{17}\), the minimum horizontal emittance can be written as

\[ \epsilon_{\text{min}} = (7.7 \times 10^{-13} \text{ m.rad}) \frac{\gamma^2}{J_x M^3} \quad (32) \]

which is very accurate for \( M \geq 4 \quad ^{18} \). In actual storage ring design it is difficult to achieve the minimum value given by (32) and a more realistic estimate is \(^{15}\)

\[ \epsilon_x = 2 \epsilon_{\text{min}} = (1.5 \times 10^{-12} \text{ m.rad}) \frac{\gamma^2}{J_x M^3} \quad . \quad (33) \]

3.2 Energy Spread

According to the explanations given so far, it may be thought that, due to damping, the e-beam becomes point like. This is not the case. The synchrotron radiation is, indeed, emitted in quanta of discrete energy.
which generate a kind of noise. As a consequence the electrons undergo a diffusion mechanism, counteracting the damping, the resulting energy spread is 3)

\[ \sigma_e = \sqrt{C \frac{1}{q}} \gamma \sim \sqrt{\frac{C}{2q}} \gamma \]  

(34)

where

\[ C_q = \frac{55}{32\sqrt{3}} \frac{m_e}{m_o} \equiv 3.84 \times 10^{-13} \text{ m} \].  

(35)

Beside the quantum excitation two other effects may cause beam heating, namely the "Touschek effect" and the "anomalous bunch lengthening". We will discuss them within the framework of the current-limiting factors.

3.3 e-beam current

Current limitation in SR's is due to such reasons as beam-gas interaction, ion trapping, intra-beam scattering, anomalous bunch lengthening etc. For a more complete description of these effects the interested reader may like to see the paper by Le Duff in Ref. 19. In this note we will briefly discuss the Touschek and bunch lengthening effects.

When two particles performing transverse oscillations collide, a part of their transverse momentum is transformed into a longitudinal momentum change. As a consequence, one particle gains and the other loses momentum. If the momentum variation is larger than the momentum acceptance of the SR both particles are lost\(^*\). This is the Touschek effect and its consequence is a reduction of the beam life-time. The probability of intrabeam scattering increases with the beam density. Therefore, the effect is also a limiting factor of both emittance and current density. Calculations of the maximum achievable current and beam life-time have been given and can be found in Refs. 20.

Let us now discuss a little more quantitatively, the so called anomalous bunch lengthening. The phenomenological model 21) predicts that when the bunch current exceeds a certain threshold value the energy spread and the bunch length will both increase with the stored current in the bunch.

\(^*\) Strictly speaking this is a single-particle Touschek effect. However, another effect takes place namely the multiple Touschek effect, in which the energy transfer between particles does not lead to particle losses but appears as a noise source for the particle motion. The obvious consequence is an increase of the energy spread in the beam.
It can be shown that, under specific conditions the following two equalities hold 15,20):

\[ \frac{\sigma_Z}{R} = \frac{\alpha}{v_s} \sigma_e, \quad R = \text{machine radius} \]

\[ \sigma_e^3 = \frac{1}{\sqrt{2\pi}} \frac{e v_s I}{\alpha E_0 \left| \frac{Z_n}{n} \right|} \]

where \( \sigma_Z \) is the longitudinal bunch length, \( \alpha \) the momentum compaction, \( v_s \) the machine tune, \( I \) the average current and \( |Z_n/n| \) the characteristic impedance. The peak current is given to

\[ \hat{I} = \frac{\sqrt{2\pi} R}{\sigma_Z} I . \]

Combining both (37) and (36) yields 15)

\[ e \hat{I} \left| \frac{Z_n}{n} \right| = 2\pi E_0 \alpha \sigma_e^2 . \]

We now have all the most important parameters to write down the gain for an SR FEL. We must underline that Eq.(38) has a particularly interesting meaning. It shows that the energy spread is not only an undesirable feature in the sense that it reduces the gain but, since larger energy spread allows larger peak currents, a suitable balance between the two competitive effects may give rise to an "optimum" energy spread for FEL operation. The explicit expression of the gain function is \((\Sigma_L \sim L \lambda)\)

\[ g(\omega) = \frac{2\pi^3}{c} \frac{\sigma F_1(\xi)}{|Z_n/n|} \mu_e^2 f(\nu; \mu_C, \nu_X, \mu_\xi) \]

where \( f(...) \) is the inhomogeneous gain function and reduces to the ordinary gain function when all the \( \mu \)'s are zero. The presence of the coupling coefficient \( \mu_C \) in Eq.(39) is due to the bunched beam operation and, therefore, to the longitudinal phase locking. Within this framework it is not an extra independent variable but, according to Eqs.(27) and(36) may be written as

\[ \mu_C = \frac{4N^2\lambda}{\alpha R} \frac{v}{\sigma_e} \mu_\xi . \]

Once the emittance is fixed by Eqs.(32-37) and by the further condition of Eq.(26) one can find the optimum \( \sigma_e \) by looking at the maximum of the function \( \mu_e^2 f(...) \) against \( \mu_\xi \).

Analogous optimization criteria can be found for the current limita-
tion due to the Touschek effect, but we will not discuss this case since
the optimisation procedure closely follows that developed above. Let us
now quickly discuss the power achievable with an SR FEL (for a complete
analysis the interested reader is referred to Refs. 19-20).

At the beginning of this subsection we have briefly outlined the SR
FEL dynamics and saw that the FEL interaction acts as a kind of noise which,
in the chosen hypothesis\(^{22}\), induces a diffusion counteracted only by the
damping due to the synchrotron emission in the bending and undulator mag-
nets. We can, therefore, expect that the average laser power \( P_L \) will be
related to the synchrotron emission one. The relationship can be stated
more quantitatively as follows. The laser process itself degrades the
e-beam qualities, then the gain decreases and the laser is switched off.
We must wait a time of the order of the damping time to have a new
laser pulse. The average laser power is therefore approximatively given by\(^{22}\)

\[
P_L \approx \frac{N_e \Delta E}{\tau_z} = \frac{N_e}{\tau_z} E \frac{\Delta E}{E} \leq \frac{1}{2N} P_S
\]  \( \tag{41} \)

where \( \Delta E \) is the maximum energy variation, \( N \) is the number of particles in
the beam and \( P_S \) is the synchrotron radiation power given by (see Eq. 28)

\[
P_S = \frac{N_e U_0}{T} = \frac{N_e E}{\tau_z}
\]  \( \tag{42} \)

Using the above scaling law and the design parameters of the LEDA-F ma-
chine\(^{12} \) we have plotted in Fig. 1 the FEL-SR power levels against the
wavelength and have also included the current limitation due to the
Touschek effect.

4. \textbf{SINGLE-PASSAGE FEL OPERATION}

The first FEL operation was accomplished with the Stanford supercon-
ducting Linac. This electron source was characterized by extremely good
beam qualities which made it an almost unique tool for the first experi-
mental attempts.

Even though an ideal machine for FEL operation, the superconducting
Linac has long been considered an impracticable solution for FELs owing
to its technological complexity and large operational costs\(^4\). However,
recent progress in superconducting cavity technology made the machine
operation less critical, in principle, and reduced considerably the costs.
It is therefore desirable that, with its special characteristics, this
accelerating device be carefully reconsidered for FEL operation. Single-passage FEL's using more conventional sources have been proposed and up to now have operated with an induction Linac \cite{23}, an RF Linac \cite{24}, a Van der Graaf machine \cite{25} and a microtron \cite{26}.

The chart of existing experiments is shown in Fig. 2. A rather detailed review of the low energy accelerators dedicated to FEL operation has been made in Ref. 4. Here we will briefly discuss a few of the characterizing features of each e-beam source.

4.1 Linacs

The most comprehensive review on Linacs has been given in Ref. 26. There are essentially two types, namely the RF and the induction. The first can provide currents of the order of hundreds of mA and (for FEL operation) an energy of hundreds of MeV. Induction Linacs can furnish e-beams of tens of kA and tens of MeV \cite{4}; the advanced test accelerator is indeed designed to provide a beam of 10 kA at 50 MeV \cite{27}.

It is clear that conventional RF Linacs can be dedicated to Compton regime FELs, whilst induction Linacs FEL operate in the so called high gain collective regime \cite{3}. RF Linacs have been operated around two frequencies, 3 GHz (S-band) and 1.3 GHz (L-band). The main limitation of these machines is the large energy spread whose main sources are the variation in RF accelerating field along the length of the bunch and single-bunch beam loading \cite{28}. A typical energy spread is 1% at 50 MeV which can result in a too large inhomogeneous broadening to ensure laser action. A typical measure to overcome this limitaton is to compress the energy by an order of magnitude.

The typical longitudinal microbunch length varies from 3-4 ps for superconducting Linacs, to 6 and 15 ps for S and L band normal Linacs respectively. According to Eq. (28) S-band Linacs (as well as superconducting ones) may create problems for FEL operation with long undulators at FIR wavelengths. The use of the energy compression mechanism may solve this problem too, but inevitably creates difficulties with the peak current. A criterion to optimize bunch length and peak current has been discussed in Ref. 11. Large peak currents can, in principle, be obtained with a Linac, but limitations arise for the average current. Using a subharmonic bunching technique 100 A peak current in 30 ps has been obtained at Los Alamos \cite{24}. At Osaka University \cite{30} the same method is used to obtain very high single bunch current (3kA in 16 ps).

Let us now briefly discuss the problem of emittance in Linacs. In Fig. 9 we have plotted the product of the normalized emittances against the average currents of the most representative sample of existing low-energy accelerators \cite{11}. These two quantities are roughly correlated by an empirical relationship \cite{11,*}.
Fig. 9  Current vs emittances for existing accelerating devices without radiative damping

\[ \frac{\bar{I}}{\beta^2 y^2 x^2 y} \approx 1.1 \times 10^4 \]  \hspace{1cm} (43)

which has been exploited to get scaling relationships for the FEL gain and power \(^{11}\). However, since no fundamental limit other than the cathode emission exists, one may expect that with careful research the limit of Eq. 43 can be improved by an order of magnitude or more \(^{31}\).

Induction Linacs, as already mentioned, can provide very high beam currents, but are, at the moment, limited in energy (tens of MeV) and pulse duration (tens of ns). As to the energy spread it can be derived from the following relationship \(^{28}\)

\[ \sigma_x \sim 10^{-3} \sqrt{1 + 3.6 \times 10^{-3} \frac{I[A]}{\beta_y}^2} \]  \hspace{1cm} (44)

Typical values are around a few \(^\%\). As far as the emittance is concerned, the above considerations relevant to RF Linacs also hold for induction ones.

*) The emittances are expressed in cm.rad.
4.2 Microtrons

A review of FEL experiments in progress with microtrons and their relevant technology has been made in Ref. 4. The microtrons offer with respect to the Linacs the important advantage of a much lower energy spread while the pulse length is longer than that of the S-band Linacs. These two effects can be explained by the energy compression mechanism which is automatically set up by the microtron operating principle

Typical values of energy spread are

\[ \sigma_e \sim 1.5 \times 10^{-2} \frac{E_r}{E_0} \]  

(45)

where \( E_r \) is the resonant energy gain per orbit and \( E_0 \) is the nominal machine energy.

From Eq. (11) and the above relationship, we find a condition on the resonant energy gain to avoid energy spread problems, namely

\[ E_r \leq \frac{10^2}{6} \frac{E_0}{N} \]  

(46)

Taking into account typical operating FEL microtron parameters and the fact that \( E_r \leq 1 \) MeV, the condition of Eq. (46) is largely satisfied. Typical values of the bunch length are around 20-30 ps. Limitations can arise for FEL operation at long wavelengths (FIR) or with long undulators \(^*)\). For the microtron emittances the conclusions arrived at for the Linacs \(^11\) also hold.

The most serious disadvantage of a microtron is the peak current which can reach only a few Amperes. The intrinsic limits are the amount of power which can be pumped into the cavity, and the cathode geometry. The race-track microtrons having a separate injector section can be used to overcome this difficulty and enjoy both the advantages of linacs and conventional microtrons \(^4\).

4.3 Van der Graaf accelerators

Finally we mention the Van der Graaf accelerators. A machine of this type has been already exploited as an e-beam source for the UCSB FEL experiment \(^25\). The beam of a Van der Graaf accelerator is characterized

\(^*)\) We must however underline that the FEL optimization is a rather complicated process, which should be carried out with regard to the various effects contributing to the gain. Taking these effects into account separately may lead to misleading conclusions.
by extremely good qualities. For example, the UCSB accelerator has furnished a beam with an emittance (normalized) at 2.5 MeV of about 7.5x10^-6 m.rad. Furthermore the beam has a continuous structure and can reach average currents of a few Amperes with maximum energy of tens of MeV 4,33).

4.4 General Features of Single-Pass FELs

We have discussed so far the accelerator performances rather than those relevant to the laser. We will now briefly discuss the main characteristics of the single-pass FEL gain and saturation.

Assuming that the main limitation on the beam current is the RF power, we can write FEL single passage gain as follows 3)

\[
g_h = g^\circ_h S_{eq_y} [\theta ; \mu_x, \mu_y, \mu_e] \quad h \equiv \text{helical}
\]

\[
g_{\parallel, n} = g^\circ_{\parallel, n} S_{eq_y} [\theta_n ; \mu_x, \mu_y, \mu_e], \quad \parallel \equiv \text{linear}
\]

\[
g^\circ_h = 88 \times 10^{-4} N^2 \left( \frac{k}{1+k^2} \right)^2 \frac{P_{[MW]}}{\delta} \frac{\lambda}{\lambda_u}, \quad (47)
\]

\[
g^\circ_{\parallel, n} = 88 \times 10^{-4} n^2 N^2 \left( \frac{k}{1+k^2} \right)^2 \frac{P_{[MW]}}{\delta} \frac{\lambda}{\lambda_u} \left[ J_{(n-1)/2} - J_{(n+1)/2} \right]^2,
\]

where \( \delta \) is the machine duty cycle, \( P_{[MW]} \) is the e-beam power in megawatts and \( \theta \) and \( \theta_n \) are the "delay-parameters" given by

\[
\theta = \frac{4N}{\pi} \frac{w \delta T}{g^\circ_h}, \quad \theta_n = \frac{4N}{\pi} \frac{w \delta T}{g^\circ_{\parallel, n}} \quad (48)
\]

where \( \delta T = T_C - T \), \( T_C \) the cavity round trip period and \( T \) the bunch-bunch time (see Fig. 10).

The quantity \( S_{eq_y} \) represents the maximum value of the multimode gain function and contains also the dependence on the different parameters entering the process.

The typical behaviour of \( S_{eq_y} \) against \( \theta \) is shown in Fig. 11, together with the dimensionless laser power \( \chi \) 3). It should be noticed that the maximum gain and the maximum output laser power do not correspond to the same value of \( \theta \). Therefore optimization of the gain does not result in the maximum output laser power.

The average laser power can be evaluated according to the following formula 34)

\[
P_L [MW] = P_{[MW]} f [Hz] \chi(\theta) (\tau_M [\mu s] - \tau_R [\mu s]) \quad (49)
\]
Fig. 10 e-beam structure from an RF machine: \( \tau_b \equiv \) microbunch time duration, \( \tau_M \equiv \) macropulse time duration, \( T \equiv \) microbunch time separation, \( f \equiv \) repetition frequency.

Fig. 11 Gain function and dimensionless laser power vs \( \phi \)

where \( f \) is the machine repetition frequency, \( \tau_M \) is the e-beam macropulse duration (see Fig. 10) and \( \tau_R \) is the pulse rise-time linked to the gain by the following expression

\[
\tau_R[\mu s] \approx 0.14 \frac{L_c[m]}{g - \gamma_T} \tag{50}
\]

where \( g \) is the gain as a function of the above parameters, \( \gamma_T \) is the cavity loss and \( L_c \) is the length of the cavity. In Fig. 1 the curves represent the average output power of an FEL operating at the 1st and 3rd harmonic respectively, with an e-beam power of 20 MW. It is evident that in the region \( 10 < \lambda(\mu m) < 100 \) the FEL, in principle, may generate larger power than the conventional sources.
5. CONCLUSIONS

In this note we have presented a review of both the storage rings and single-pass FELs. We have emphasized the problems relevant to the electron sources and laser light output but no mention has been made of the cavity and undulator technology which are dealt with more completely in Ref. 4. We have also stressed that the future development of FELs as a "workhorse" for tunable applications strongly depends on the reliability of the electron source.

The use of the FEL for industrial applications will depend on its cost being relatively modest. In Ref. 34 a comparative cost analysis of FEL with other lasers was carried out and the results are summarized in Fig. 12. It is clear that the FEL is competitive when it is operating with a high efficiency extraction system (in the figure with an efficiency of 10%).

As a concluding remark we would like to stress that the goal of this paper has been twofold, namely to give a review of the basic ideas and problems underlying the FEL physics, and to indicate how this new laser device must be realistically considered within the framework of tunable sources.

![Diagram showing FEL cost/watt vs λ with efficiencies of 1% and 10%](image)

Fig. 12 FEL cost/watt vs λ with efficiencies of 1% and 10%

*) The analysis has been limited to single-pass operating FELs where a market analysis for the electron source can be carried out. For the SR the technology is so specific that a market analysis makes no sense.
REFERENCES

2) B.D. Guenther and R.G. Buser in Ref. 1 p. 1179.


ISIS, THE ACCELERATOR BASED NEUTRON SOURCE AT RAL

D A Gray and G H Rees
Rutherford Appleton Laboratory, UK

1. INTRODUCTION

During the Oxford Accelerator School a tour and description were arranged of the Rutherford Appleton Laboratory's new neutron source. Subsequently, on the last day of the school, a seminar was given on the high-intensity performance of the source's rapid cycling synchrotron. Details of the talk and seminar are repeated here.

The design specification for the pulsed neutron source called for peak fluxes of thermal and epithermal neutrons \( > 10^{16} \text{ n cm}^{-2} \text{ sec}^{-1} \) in pulses of duration \( \sim 10 \mu s \) at a repetition frequency of 50 Hz. To achieve this goal at RAL the method adopted has been the construction of a 50 Hz, 800 MeV proton synchrotron to provide \( 2.25 \times 10^{13} \) protons per pulse at a heavily shielded target of depleted uranium 238. The initial reaction in the target is the production of fast neutrons by spallation and fission. This is followed by the slowing down of the neutrons to thermal and epithermal energies by associated moderators.

Such a spallation neutron source allows significant advances compared with existing high flux reactor sources. The effective flux is much greater than that available from reactors for the higher energies of the neutron spectrum. This increase in neutron flux will be a major benefit to a wide range of condensed matter studies, especially for the case of the epithermal neutrons at energies of several electron volts.

2. LINAC AND SYNCHROTRON

The synchrotron injector is a 70.4 MeV \( \text{H}^- \) linac with 4 Alvarez tanks operating at 202.5 MHz. The pre-injector is a 665 kV Cockroft-Walton set with a medium gradient accelerating column, using glass insulators. The \( \text{H}^- \) ion source is of the Penning type and uses a mixture of hydrogen gas and caesium vapour. It is a direct extraction source with a duty cycle of 2.5% (50 Hz, 500 \( \mu s \)).

A transfer line takes the 70.4 MeV beam line from outside to inside the synchrotron magnet ring. Here it includes a 202.5 MHz debuncher cavity to control the input beam momentum spread; there is also a septum magnet, for injection from an inside machine radius. Diagnostics are provided for emittance and momentum spread measurements.

The injection straight section of the synchrotron is approximately 5 m in length. It houses 4 septum-type dipole magnets for creating a localised bump of the closed orbit. The first of the bump magnets lies adjacent to the injection septum magnet and the region between the 2 central magnets is used to house the foil which strips \( \text{H}^- \) ions to protons.
The synchrotron is divided into 10 superperiods with each superperiod containing a pair of doublet quadrupoles, a long straight section, a singlet quadrupole, a combined function gradient-bending magnet and a medium length straight section. The quadrupoles and bending magnets are connected in a 50 Hz, series resonant circuit together with associated capacitors and a common choke with 10 secondary windings (ex-NINA). The superperiod and magnet designs were chosen for compatibility with the stored energy of the existing choke and capacitors. In principle, the magnet current may be locked in frequency to a fixed 50 Hz frequency or to the 50 Hz mains; in practice it is necessary to lock to the fixed frequency for adequate stability.

Vacuum system components are entirely of metal or ceramic apart from the ferrite of the injection and ejection magnets. The design has aimed for simplicity and reliability to minimize the maintenance in the high radiation environment. Each superperiod has 3 triode titanium sputter ion pumps of capacity 400 l sec$^{-1}$ with additional pumps at the ferrite locations. The system is pumped down to $10^{-3}$ Torr via a roughing line which passes through the ring shielding to external carbon vane, sorption pump units and a turbo-pump. The ion pumps reduce the pressure to $5 \times 10^{-7}$ Torr within 10 hours, ultimately reaching $10^{-8}$ Torr. There are no sector valves in the ring but all-metal valves are included in the injection and extraction beam lines.

Eddy currents preclude the use of solid metal chambers within the large aperture, rapid-cycling magnets; ceramic chambers are used for both the main and the correction magnets. Sections of chamber are formed, typically 300 mm in length, by isostatically pressing 97.6% pure alumina powder in a mould, machining the pressed powder, firing at high temperature and subsequently grinding the external surfaces to a high tolerance. Final chambers are formed from the individual sections by glazing and dovelling the mating surfaces and then heating the free-standing assembly to 1100 °C in a furnace to allow the 0.25 mm glass layer to fuse into the ceramic, bonding the surfaces. End flanges are also of ceramic and are sealed to adjacent flanges by re-usable indium T-seals.

Special radio frequency shields are inserted within the ceramic chambers to reduce the coupling impedance of the proton beam to its environment. The shields are made as rectangular chambers of stainless steel rods, supported in insulating frames, with the rods lying parallel to the beam direction. In the bending magnets the side rods are replaced by 2 mm thick solid, stainless steel plates, standing vertically. Each rod and side plate is connected to the adjacent straight section by compact, ceramic capacitors which present a high impedance at 50 Hz but a low one at and above 100 kHz.

Acceleration from 70.4 to 800 MeV is achieved via 6, double-gapped, ferrite-tuned cavities, operating from 1.347 to 3.09 MHz at harmonic number 2. The net peak accelerating voltage has to vary smoothly from 3 to 160 kV per turn in the 10 ms acceleration period and the design intensity corresponds to a high level of beam loading. At present four of the cavities are in operation and are adequate for acceleration to 550 MeV. The final two are to be installed in mid-1986. Each cavity is powered by its own
RF amplifier, the final stage of which contains two 250 kW tetrodes in parallel. For the present stage of running, only one of the tetrodes is included per amplifier but the second must be added to control the maximum beam loading levels. The power amplifiers have been designed for ease of removal with electrical and coolant connections made via quick disconnect terminations in a region that may be shielded if it is found necessary.

Extraction is achieved by three fast kicker magnets and an extraction septum magnet. The plane for extraction is vertical and a closed orbit bump at the septum location reduces the kicker requirements. Each kicker is of a push pull design with the ferrite split at the mid-point by an electrical ground plane. Each half magnet is powered via a pulse forming network and coaxial thyatron switch. The voltage on the system is 40 kV, the peak current 5000 A and the required kick rise-time 225 ns. The septum unit lies above the synchrotron and encloses its curved (21°) vacuum chamber which is joined to the top of the straight section beneath. The septum chamber is non-magnetic but the adjacent straight section is of mild steel to reduce the septum leakage field at the beam. Parameters of the magnet are field level 1 T, bending length 1.8 m and septum thickness 10 - 15 mm. In the extraction straight are a number of beam loss protection units; the low energy units are made of copper and graphite and the high energy units of stainless steel.

There is a long (150 m) beamline from the extraction point to the target station. It consists almost entirely of ex-NIMROD, large aperture quadrupoles and dipoles. The power dissipation in the line is high, over 800 kW, because of the large bend angles involved. The downstream end of the line lies in the neutron experimental hall and in this region it is heavily shielded. In the future an intermediate target station is to be installed in the line. This is to feed a powerful muon spin resonance beam line and experimental station. At the end of the main beamline the proton beam is focussed to a 70 mm diameter spot at the incident end of the uranium target. Beam profiles along the line are measured using strip secondary emission detectors.

3. TARGET STATION

The target consists of a cooled uranium target, associated moderators, reflectors and decouplers, a bulk shield and shutter system and a remote handling facility for dealing with spent and replacement targets.

Target material is depleted uranium 238 and, because of its poor thermal conductivity, it is segmented into a number of plates with intermediate parallel cooling channels. Heavy water is used for the cooling and the uranium plates are clad with Zircaloy-2 to avoid corrosion and to contain fission products.

The interaction of the 800 MeV proton beam with the target material is a combination of spallation and nuclear excitation. Fast neutrons and other secondary particles result directly from spallation and subsequently from fission and evaporation after nuclear de-excitation. Secondary particles undergo further interactions leading to
a particle cascade. Uranium is used in preference to other heavy metals as there is a further factor of 2 increase in the neutron yields due to the fissions.

The target array is 340 mm long which is 20% longer than the range length for 800 MeV protons. The uranium discs are 90 mm in diameter and they are mounted in rectangular picture frames contained in a stainless steel vessel. Neutron production is approximately 26 neutrons escaping the target per incident proton; the average neutron energy is 2 MeV and there are about 10% of the neutrons with an energy greater than 15 MeV. Power dissipation in the target is 200 kW for 180 μA of incident protons. A choice of 90 mm is made for the target diameter as a compromise between fast neutron production and coupling to the associated moderators.

There has to be a metallurgical bond between the uranium plates and the cladding of Zircaloy-2 and this is accomplished by hot isostatic pressing. The cooling channel between the individual plates is 1.75 mm and subsequent swelling of the uranium may reduce this to 1 mm. Heavy water is the coolant rather than light water as the presence of the heavy water in the extensive cooling manifolds acts as a neutron reflector. For safety there are secondary and tertiary light water cooling circuits coupled via heat exchangers to the primary heavy water circuit.

Fast neutrons are slowed to epithermal and thermal energies by moderators. Two are sited just above the target and two below, all in wing geometry. A typical size of moderator is 100 x 100 x 50 mm³, small enough to restrict the neutron pulse duration to 5 - 100 μs (dependant on λ), a requirement for the time of flight: experimental stations. All moderators have external decouplers on all faces except the exit face and the whole array is contained within a reflector. The decouplers prevent neutrons thermalised outside the moderator from entering it and degrading the output pulse. The reflector scatters back fast neutrons into the moderator, enhancing the output yield by a factor of 3.

Each moderator is designed to optimise its performance over a particular range of the neutron energy spectrum. Two types contain ambient-temperature light water, one type liquid methane and one para-hydrogen at 20 °K. The lowest temperature moderator provides the longest wavelength neutrons, 4 - 10 Å. Enclosing the target and moderators is the reflector which contains beryllium and heavy water. Total power deposited in the reflector is 7.2 kW, that in the moderators is 1 kW and that in the decouplers is 9 kW.

There is a large bulk shield surrounding the target which reduces the radiation level in accessible areas to < 7.5 μSv hr⁻¹. Additional shielding is used around the 18 neutron beam ports in the shield and around the 18 neutron beam tubes and detectors. The overall shield height is 7 m, the thickness 4.3 m and the outer 0.25 m layer is concrete loaded with 1% boron. In the forward direction, the shielding is extended towards the remote handling cell.

Within the shielding is a target void vessel, a shutter system, shielding inserts
and a plinth and shield door. A cylindrical pressure vessel with 18 double aluminium windows is used to contain a helium atmosphere around the target and a closed cooling circuit for the helium removes about 5 kW from the vessel walls. The shutter system consists of two shutter vessels, each containing nine, 22 tonne shutters. These are made of concrete and iron and are used to isolate an individual neutron beam line. The shielding inserts are prefabricated steel boxes with removable shielding blocks packed around the neutron beam tubes so that each beam tube may be readily re-designed. The concrete includes two caverns. The final component of the shield is the 90 tonne, 4.5 m thick door at the downstream end of the target, ahead of the remote handling cell and the services region. There is a seal between the door and the shield to contain the atmosphere of helium. The target-moderator assembly is cantilevered from the shield door.

When it is necessary to obtain access to the target assembly, the shield door is rolled backwards on rails until the target assembly is in the remote handling cell; the door then completes the back wall of the cell. Four master-slave manipulators are used in the cell for removing a spent target, replacing with a new one and for any maintenance on the components of the assembly. To facilitate use of the manipulators, the cell is provided with two large, zinc bromide windows. Viewing is supplemented by TV cameras. Irradiated targets may be stored in any of three storage wells in the floor of the cell and be removed via an access hatch.

4. EXPERIMENTAL FACILITIES

A wide range of instruments are used with the eighteen neutron beam lines; at present eight are in operation and in the future up to twenty five may be accommodated. In addition to neutron scattering science, neutrino physics will be undertaken at ISIS in a large neutrino blockhouse adjacent to the target. A brief description only is given of the present experimental facilities.

A spectrometer named IRIS is used to measure quasi-elastic scattering in processes such as diffusional motion in liquids and rotational and translational dynamics of molecules. Good time of flight resolution is achieved with a 40 m cold neutron guide and with energy analysis by crystal analysers in back reflection. The resolution of the instrument is in the range around 50 μev.

LAD is a liquids' and amorphous materials' diffractometer. It is used to study the structure factors of non-crystalline materials and also as a medium-resolution, high intensity diffractometer. This instrument had been previously tested on the Harwell pulsed neutron source.

HRPD is a high resolution powder diffractometer and includes a 96 m thermal neutron guide. It allows a large number of structural parameters to be determined from powder measurements and it may also be used to study phase transitions and line broadening effects.
TXFA is a time-focussed crystal analyser; HET is a high energy transfer inelastic spectrometer; LOQ is a low-Q diffractometer; EVS is an electron volt spectrometer for epithermal neutrons and POLARIS is for study as a polarisation spectrometer. POLARIS uses a neutron polarising filter, Sm 149, in the incident neutron beam and it will be developed for performing inelastic polarisation experiments for investigating electronic and nuclear magnetism.

Several mechanical chopper systems have been developed for use with the neutron instruments. A magnetic bearing has been incorporated in the chopper for the HET spectrometer; there are three rotors allowing peak transmission at 0.25, 0.5 and 1.0 eV energies for 1 μs pulses.

The neutrino facility, KARMEN, has been initiated by the Karlsruhe Laboratory, FRG. It consists of a 5500 tonne iron shield, housing two detector systems, and is located 14 m from the ISIS target station. The inside dimensions of the blockhouse are 10 x 4 x 3 m³. One detector is a total energy calorimeter using liquid scintillator and the other is a high precision tracking device for measuring neutrino-electron scattering.

5. HIGH INTENSITY PERFORMANCE OF THE ISIS SYNCHROTRON

The main features involved in high intensity operation of the synchrotron are the performance of the linac injector, the efficiency of the H⁻ injection process, the bunch formation in the ring, the trapping efficiency, the heavy beam loading, the crossing of betatron resonances, the possibility of instabilities, the extraction efficiency and the activation of machine components.

At the time of the Oxford meeting, the maximum beam injected at low repetition frequency had been 10¹³ protons per pulse. The best performance had been 9 μA on target, corresponding to 4.5 x 10¹² protons accelerated per pulse (for 5.5 x 10¹² H⁻ ions injected) at 12.5 Hz. Since that time the performance has been improved to 40 μA on target corresponding to 5 x 10¹² protons per pulse at 50 Hz.

A schematic lay-out of ISIS is shown in Fig. 1, the scale of which may be judged from the 52 m synchrotron diameter and the 150 m beam line to the target station. The linac is shown in the forefront of the figure.

The injector has not yet met its design specification. Typical performance figures have been output currents of 5 mA for pulse lengths up to 200 μs at 25 and 50 Hz and currents of 3.5 mA for pulse lengths up to 450 μs at lower repetition frequencies. These figures are set both by the H⁻ ion source and by the fact that, as the average current is raised, there is increased frequency of breakdown of the 665 kV accelerating column. The mechanism of breakdown is not understood but the present performance has been achieved only after improved pumping at the high voltage end of the column and after installing inter-electrode shields in the column to intercept ions before they reach the glass insulators. Future plans include a thorough cleaning and check of the column and the
development of a new ion source. There are no funds available for the development of an RFQ.

The output beam emittances from the linac have been found to be as expected with 95% of the beam within transverse (un-normalised) emittances of 20 π μ rad m. At currents up to 3 mA, momentum spread measurements indicate 95% of the beam within Δp/p values of ± 1.2 10⁻³ and a debuncher cavity is used routinely to reduce the spread to < ± 5 10⁻⁴.

5.1 H⁻ charge exchange injection

Up to 300 turns have been injected into the synchrotron with high efficiency by appropriate filling of horizontal and vertical betatron phase space following the stripping of H⁻ ions to protons. Large aluminium oxide stripping foils, 120 mm × 40 mm, have been developed within the laboratory and have proved highly satisfactory in operation. They have a thickness of 0.25 μ.

Over 98% of the input H⁻ ions are stripped to protons and about 1½% to H⁺ particles. There is a separation of the H⁺ beam from the protons after passage through the injection bump magnet just downstream of the foil (the third of the four bump magnets in the long injection straight). A non-destructive monitor of the injected beam is obtained by using an internal scintillator and an external TV camera to view the separated H⁺ beam. Fluctuations of the injected beam are readily seen on this monitor. A second internal scintillator has been used to observe the injected beam after one revolution in the ring. The two monitors have been used together to obtain correct vertical alignment of the injected beam.

Injection occurs over intervals of up to 450 μs, commencing 550 μs before the minimum of the biased, sinusoidal guide field of the main ISIS magnets. Stacking in horizontal phase space is automatically obtained by holding all bump magnet fields constant while the main guide field falls during injection. The input distribution may be altered by programming of the bump magnets. For vertical filling, the beam is swept vertically in the injection line with a correlation between large vertical betatron and small horizontal betatron oscillations in the ring and vice-versa.

Beam loss during injection is observed by radiation monitors adjacent to the injection straight but no injection loss is seen on radiation monitors adjacent to the other straight sections. In Fig. 2 some features of the injection loss are shown. The time base is 200 μs cm⁻¹ and the injection interval is 90 μs; the upper trace is for the radiation monitor near the injection straight, the centre trace for the monitor near the end of the injection line and the lower trace is the first trace repeated but with the bump magnets switched off 150 μs earlier. Loss continues after injection, decreases and then increases again even though the equilibrium orbit in the machine is moving away from the injection septum. The loss ceases once the orbit bump is reduced. The total loss corresponding to the upper trace is ~4.5%, of which ~2% is the stripping loss and the
late loss is thought to correspond to horizontal beam growth of \( \sim 5 \) mm in a time of
\( \sim 100 \) \( \mu \)s, for an injected beam of \( 3 \times 10^{12} \) protons. Further studies of the effect are
needed to identify the growth mechanism.

5.2 Bunch formation

The RF is switched on 145 \( \mu \)s before the guide field minimum (\( T = 0 \)) and is held at
constant frequency and constant volts/turn until \( T = 0 \). Subsequently, the frequency is
raised to keep the beam centred in the aperture and the volts/turn are rapidly increased
from 3 kV to 80 kV by \( T = 1 \) ms and to 112 kV by \( T = 5 \) ms. This is the mode of operation
for acceleration to 550 MeV. When two further cavities are added, the volts/turn will be
raised to 156 kV by mid-cycle and acceleration will be to 800 MeV.
The most efficient operation has been with the debuncher cavity powered and a narrow momentum spread injected, $\Delta p/p = \pm 5 \times 10^{-4}$. Particles undergo a quarter of a synchrotron oscillation by $T = 0$, at which time two smooth bunch shapes have developed. Later motion is non-adiabatic with filamentation present and the development of non-equilibrium bunch distributions.

The shape of both bunches is double-humped by $T = 100 \mu s$ and periodically returns to this form, but with more complex forms at intermediate times. Typical patterns at low intensity (6 $10^{11}$ protons per bunch) are shown sequentially in Fig. 3 for $T = 100, 175, 225$ and $300 \mu s$. As the intensity is increased, the shapes become smoother due to the enhanced effect of the longitudinal space charge forces. A one-dimensional longitudinal space charge tracking code has been developed to study the bunch development and it is of interest to see if the code continues to predict the motion at increased intensity.

During the early commissioning, the pronounced double-humped bunches led to complications for the dipole-mode beam control loop. Incorrect limiting led to spurious phase detector signals and the formation of narrow, dense bunches oscillating through the main bunch distribution. Filtering of the bunch signals before limiting removed the spurious effects.

The bunches become progressively smoother as acceleration proceeds and appear stable at the present maximum levels of accelerated beam ($3 \times 10^{12}$ protons per bunch). The trapping efficiencies are typically 85 to 90% but decreasing at the highest level of beam loading observed.

Bunch areas appear larger than that corresponding to the injected momentum spread. Some of the increase is due to scattering of the beam as it circulates through the foil prior to RF switch-on. This introduces a tail in the momentum distribution but it is not large enough to explain the effect observed.

5.3 Beam loading

Present intensities have been obtained only with the aid of feed-forward beam loading compensation. Each of the four cavities has been powered by its own Class B power amplifier and feed-forward signals have been introduced in the first 2 ms of the 10 ms acceleration period. At the start of acceleration the RF voltage is low and then it is advantageous to keep the cavities tuned to resonance and not to detune them for reactive beam loading compensation.

Future plans include the installation of two further cavities and the addition of a Class A power stage in parallel with each Class B stage to provide greater linearity for the feed-forward compensation. Also, there will be some reduction in the shunt impedance of the cavities and some stabilisation of the gain of the feed-forward signals.
5.4 Betatron resonances

The betatron Q-values have been measured throughout acceleration and found to be \( \sim 1\% \) lower than the values predicted from magnetic measurements. Trim quadrupole correction magnets have been powered to adjust the tunes, both for chromatic correction during the injection period when the beam is spiralling towards the centre of the aperture and subsequently. The performance is, in general, not sensitive to the tune correction apart from in an interval late in the acceleration cycle when a slow vertical orbit bump is introduced to reduce the kicker magnet requirements for fast extraction.

The only betatron resonance effect observed is associated with the slow orbit bump for extraction. The effect has been observed on a scintillator which may be inserted at the input of the extraction septum magnet. With the Q-values uncorrected, a beam spot is observed on the scintillator with a dense core and four prominent wings, characteristic of a fourth order resonance. It is thought to be the coupling resonance \( 2q_h - 2q_v = 1 \). The effect is removed by adjusting the Q-values away from \( q_h = 4.26 \), \( q_v = 3.76 \), after which there is very little loss of beam during fast extraction.

5.5 Instabilities

No instabilities are observed during acceleration for the maximum beam intensities achieved, \( 6 \times 10^{12} \) protons per cycle. There is, however, the growth of the circulating beam following injection, described in section 5.1.

Each ceramic vacuum chamber in the ring includes a special RF shield, designed for a low beam coupling impedance. Quadrupole and bending magnet chambers are shown in Fig. 4 and the experimental set-up for comparing the longitudinal impedance of the shield with that of a solid chamber in Fig. 5. The resistive component of the shield is restricted by ensuring good RF contact between the shield and the neighbouring vacuum chamber components. The reactive component of the longitudinal impedance is lowered by arranging for the shield wires to approximately follow the low energy beam profile. Even with this care, the space charge contribution to the longitudinal coupling impedance, \( Z/n \), is \(-j\ 700\ \Omega\) at low energy and \(-j\ 170\ \Omega\) at high energy.

The impedance values are such that the design currents are above the threshold levels for longitudinal and transverse instabilities. For the coaxing beam longitudinal microwave instability, the predicted initial growth time for the residual 202.5 MHz linac bunch structure is of order 100 \( \mu \)s and it is planned to look for this effect while operating the synchrotron as a 70.4 MeV storage ring. It is believed that an initial tail in \( \Delta p/p \) may develop which inhibits further growth. Possibly this is contributing to the horizontal beam growth observed after injection.

The synchrotron operates below transition and with the natural negative values of the chromaticities. Transverse coherent instabilities may arise for higher intensities at the end of the injection interval; the most likely frequency is 140 kHz for the lowest
Fig. 4  Ceramic Vacuum Chambers

Fig. 5  Impedance Measurements
vertical coherent mode and the next most likely frequencies are for modes near 100 MHz. Space has been left in the magnet lattice for the inclusion of a set of octopole lenses to combat transverse instabilities, if necessary.

5.6 Collection of lost beam

A horizontal 70 - 100 MeV beam loss collection system has been installed which is designed to localise much of the loss in one long, straight section of the synchrotron. The purpose of the system is to restrict radiation damage and activation of machine components while operating at high intensity. The major loss is expected to be the beam loss during trapping.

The primary intercepting unit is placed at an inside machine radius, near the upstream end of a long, straight section. This is just after an F quadrupole so that the dispersion is high and the radial motion of untrapped beam near a maximum. Thus the location gives a good interception efficiency and also a high probability of capturing outscattered beam on further collectors located downstream in the straight section.

The collectors are mainly of graphite but the primary collector has a lip of copper at the downstream end to enhance the angle of outscatter. The use of the graphite is to reduce the resulting activation levels.

* * *

REFERENCES


2. S Koscielniak, private communication.