Presented at the INS-Kikuchi Winter School
on Accelerators for Nuclear Physics,
January 29-February 2, 1984, JAPAN

TM-1259
0100.000

Fermilab

High Energy Hadron Colliders

L.C. Teng
Fermi National Accelerator Laboratory

January 1984

I. Beam Dynamics

A. Linear Lattice (Transverse)

1. Arcs

2. Insertions

3. Dispersions and compensations

4. Example - 20 TeV collider (SSC)

B. RF (Longitudinal)

C. Short Term Stability (Single Beam)

1. Low order resonances

2. Self-field effects (coherent instabilities)

3. Gas scattering and vacuum instability

4. Intrabeam scattering

D. Long Term Stability (Beam-Beam)

II. Superconducting Magnets

A. Superconductors and Cables

B. Magnet Configuration and Structure

C. Beam Quenching and Quench Protection

III. Beam Cooling and Antiproton Accumulation

A. Electron Cooling

B. Stochastic Cooling

C. Antiproton Accumulation

IV. Crossing Geometry, Luminosity and Tune-Shift

High Energy Hadron Colliders

L.C. Teng
Fermi National Accelerator Laboratory

January 1984

In this paper we discuss the more novel and important design considerations and features of high energy hadron colliders (pp or $p\bar{p}$). The paper does not attempt to be sufficient for making a complete design, but will contain enough references to other papers necessary for doing so. Formulas are generally given without derivation, and notations are not consistent from section to section. For most formulas the derivation is transparent although the mathematics may be lengthy. Whenever obscure, an explanation of the procedure for derivation will be given in physical terms. Detailed mathematical derivations are to be found in the references.

I. Beam Dynamics

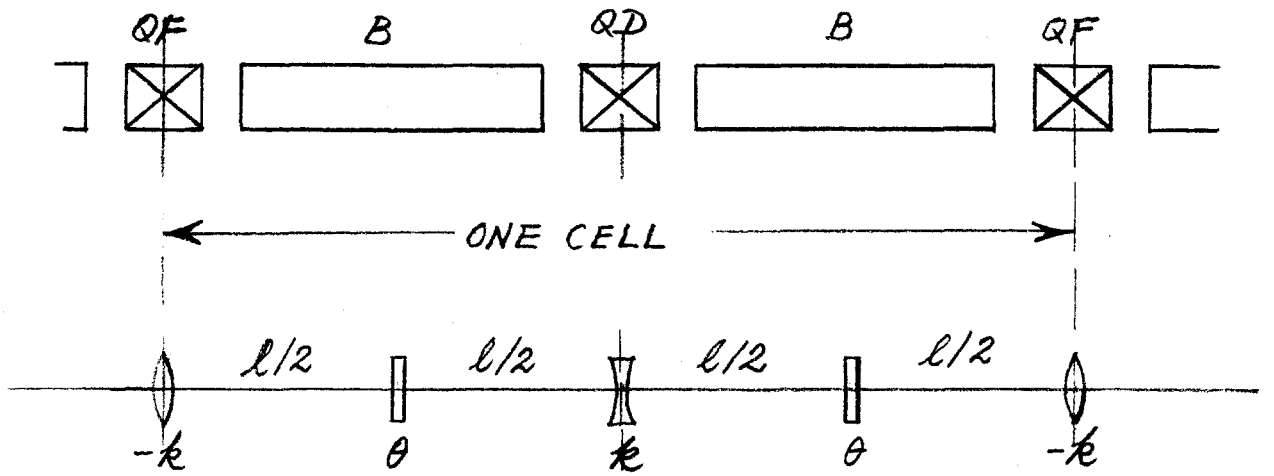
A. Linear Lattice (Transverse)

The magnet lattice is so designed that the beam of charged particles (p or $\bar{p}$ in this case) is stably confined by the magnetic forces. In this section the confinement considerations are discussed in the lowest and dominant order (linear). This will be extended to higher orders later. In addition to stability, the linear forces of the magnets control the size and the dispersion properties of the beam which in a collider must satisfy certain requirements. The linear lattice is designed to fulfill these requirements.

1. Arcs

The magnet lattice is designed to contain a number of crossing straight-sections in which the two counter rotating beams collide to produce high energy events to be studied. In between two crossing straights the beams are conducted in beam transport lines which are called "arcs"

and are designed to be as simple and economical as possible. The simplest beam transport is made up of FODO cells shown below.



At high energies the magnets can be approximated as evenly spaced δ-function elements (shown on the bottom part of the figure) with "strengths" given by

Dipole $\frac{B\ell_B}{B\rho} = \theta$ = bend angle

Quadrupole $\frac{|B'|\ell_Q}{B\rho} = \frac{1}{f} = k$ = focal strength

where

$B\rho$ = magnetic rigidity of particle

B, ℓ_B = field and length of dipole

B', ℓ_Q = gradient and length of quadrupole

($B'<0$ focusing, F; $B' > 0$ defocusing, D)

The orbital functions of such a FODO cell with "thin" elements can easily be calculated using transfer matrices and are expressed in terms of θ, k and

ℓ = half cell length. These are

$$2\sin\mu = k\ell, \qquad \mu = \text{half-cell phase advance}$$

$$|\alpha|_F = |\alpha|_D^{-1} = \sqrt{\frac{1+\sin\mu}{1-\sin\mu}} = \tan\left(\frac{\pi}{4}+\frac{\mu}{2}\right)$$

$$\beta_{F,D} = \frac{\ell}{\sin\mu}|\alpha|_{F,D}$$

$$|\eta'|_{\substack{F\\D}} = \left(\frac{1}{\sin\mu} \pm \frac{1}{2}\right)\theta$$

$$\eta_{F,D} = \frac{\ell}{\sin\mu}|\eta'|_{F,D} \qquad \eta_\theta = \frac{\ell}{\theta}|\eta'|_F\,|\eta'|_D$$

$$\Delta\ell/\frac{\Delta p}{p} = \ell\left(\frac{1}{\sin^2\mu} - \frac{1}{4}\right)\theta^2 = \theta\eta_\theta = \text{orbit length dispersion}$$

$$\Delta\mu/\frac{\Delta p}{p} = -\tan\mu = \text{phase advance dispersion}$$

where

α, β = Courant-Snyder amplitude functions

η, η' = closed orbit dispersion functions

and the subscripts F, D, and θ denote values at midpoints of QF, QD, and B. The exact orbital functions can be obtained by using computer programs such as SYNCH[1] or AGS[1]. Generally the following values of μ are used.

High, $\mu = 45°$ (4 cells per 360° phase advance)

Medium, $\mu = 36°$ (5 cells per 360° phase advance)

Low, $\mu = 30°$ (6 cells per 360° phase advance)

The gaps between dipoles and quadrupoles are needed to accommodate beam monitors; correction magnets; power, cryogenic and vacuum connections; etc. It is frequently advantageous to push the dipole to one end leaving a long gap at the other end for the functions mentioned.

2. Insertions

In long straight sections one frequently desires special orbital characteristics. So as not to perturb those in the arcs the sequence of dipoles and quadrupoles inserted between arcs are designed to yield the desired orbital characteristics in the middle and to yield orbit functions at the two ends matched to those of the arcs. The condition for matching is that the transfer matrix of the insertion must equal

$$\begin{pmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{pmatrix} \quad \text{with} \quad \begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} \eta_2 \\ \eta_2' \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_1' \end{pmatrix}$$

and

$$\begin{cases} a = \sqrt{\dfrac{\beta_2}{\beta_1}} \left(\cos\phi + \alpha_1 \sin\phi \right) \\ b = \sqrt{\beta_1 \beta_2} \sin\phi \\ c = -\dfrac{1}{\sqrt{\beta_1 \beta_2}} \left[(\alpha_2 - \alpha_1) \cos\phi + (1 + \alpha_1 \alpha_2) \sin\phi \right] \\ d = \sqrt{\dfrac{\beta_1}{\beta_2}} \left(\cos\phi - \alpha_2 \sin\phi \right) \end{cases}$$

where subscripts 1 and 2 denote arc-values at the two ends to be matched and where ϕ is the phase advance across the insertion and can be taken as a free parameter if there is no special requirement on its value. This matching must be accomplished in both the horizontal and the vertical planes. In general, this is very difficult and can only be done with computer programs such as TRANSPORT[1] or MAGIC. Experiences and proper intuitions help a great deal in the design of insertions. Even then, one is not always successful. Here we can only discuss some special simplified cases.

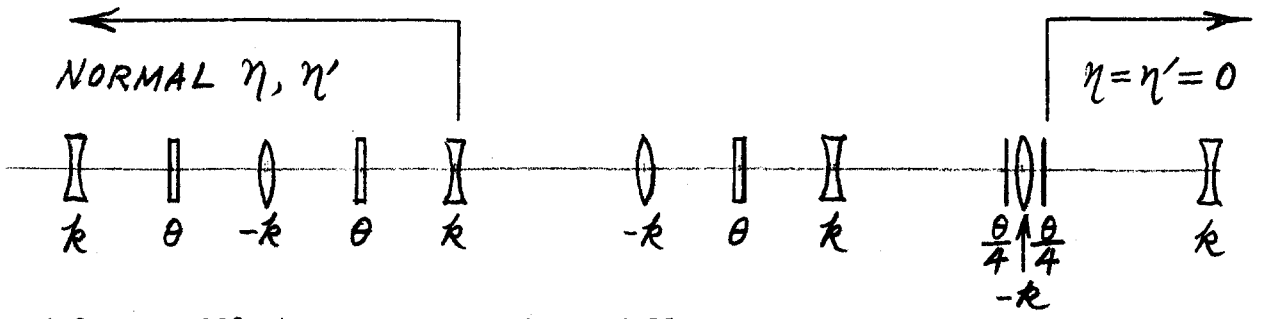
Most useful is an insertion that is straight, namely composed only of quadrupoles. For such an insertion e = f = 0 and the dispersion matching is automatically (although not necessarily) satisfied by $\eta_1 = \eta_1' = 0$, $\eta_2 = \eta_2' = 0$. Thus for a straight insertion usually one first brings the arc dispersion functions at the ends to zero then matches amplitude functions using 2x2 matrices. The dispersion function can be brought to zero in one cell by adjusting the strengths of the two quadrupoles. If the beam goes from F to D, the strengths should be k_F and k_D given by

$$\begin{cases} \ell k_F = -1 - \sin\mu \\ \frac{1}{2}\ell k_D = \frac{1}{\sin\mu}\left(\frac{1}{\sin\mu} + \frac{1}{2}\right) - 2 \end{cases}$$

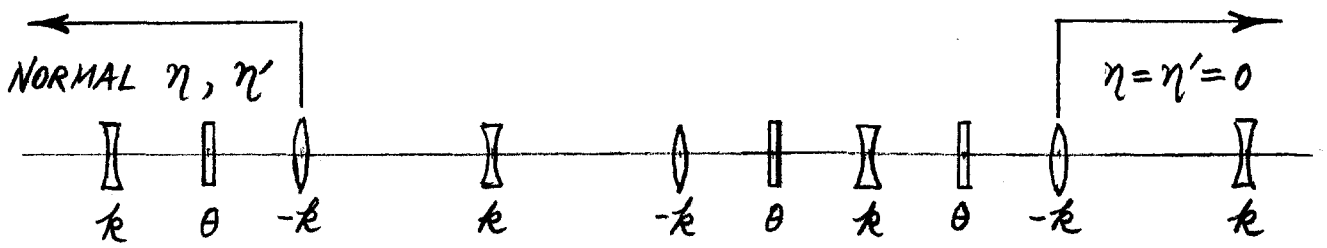
The dispersion will be zero after the dipole following k_D. If the beam goes from D to F the strengths are given by

$$\begin{cases} \ell k_D = -1 + \sin\mu \\ \frac{1}{2}\ell k_F = \frac{1}{\sin\mu}\left(\frac{1}{\sin\mu} - \frac{1}{2}\right) - 2 \end{cases}$$

and the dispersion will be zero after the dipole following k_F. Other arrangements involve adjusting the dipoles. Specifically for $\mu = 45°$, the following arrangement brings arc dispersion to zero:



and for $\mu = 30°$ the arrangement is as follows:

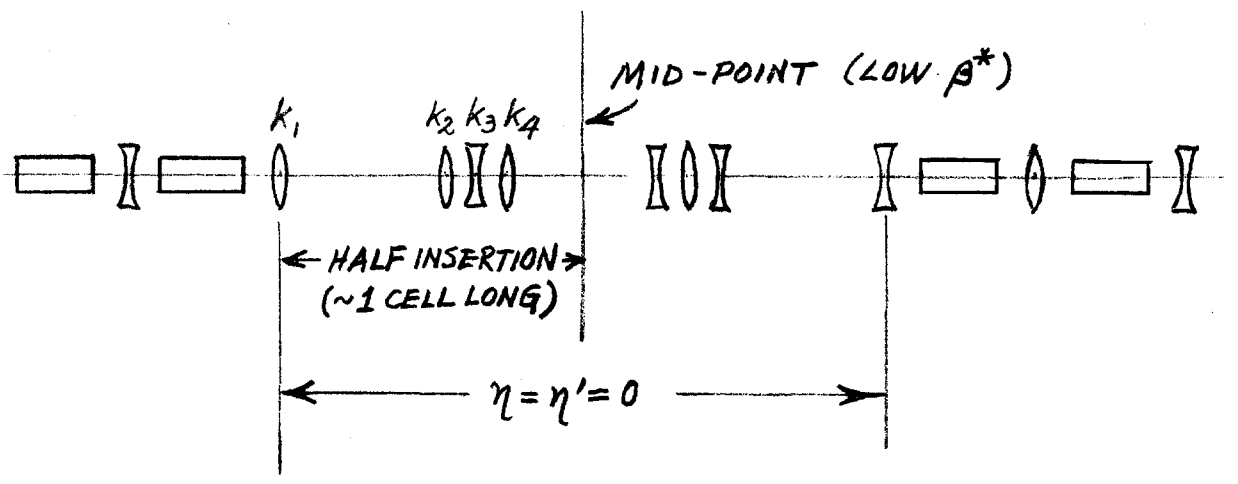


These dipole arrangements work also with all quadrupole polarities reversed. There are many more arrangements involving adjusting both the dipoles and the quadrupoles. One can invent one's own arrangement fairly easily after a little practice.

Matching the amplitude functions, namely matching both the horizontal and the vertical 2x2 transport matrices is much harder. If there is no requirement on phase advance matching, one needs 4 conditions: matching α and β in both planes or equivalently, matching 2 out of the 3 parameters specifying the transfer matrix in each plane (the 3rd parameter is the phase advance). To these one must add conditions specifying the desired orbit characteristics inside the insertion. All these conditions are transcendental equations involving trigonometric and hyperbolic functions. The solutions of these equations have very peculiar behaviors.

Symmetry helps. With a symmetric insertion, one which is unchanged when reflected about the mid-point (Of course, the end orbit functions to be matched must also be symmetric, namely $\beta_2 = \beta_1$, $\alpha_2 = -\alpha_1$.), $\alpha = 0$ at the mid-point and the end orbit function scales with the mid-point β. Thus we are left with only one matching condition (value of α_1/β_1) in each plane. With an antisymmetric insertion, one which when reflected about the mid-point is unchanged except for a sign reversal ($F \leftrightarrow D$), the matching in one plane automatically guarantees that of the other plane. We are also left with only 2 matching conditions.

The desired orbit characteristics inside the insertion appear in many different forms, but the most frequently occuring is the low-β insertion in which the mid-point β^* should be low. The simplest and most commonly used antisymmetric low-β^* insertion has the following form



The insertion is ~2 cell (4ℓ) long. The two beams collide at the mid-point where β^* has low values in both planes. This gives high luminosity.

Also for high luminosity the beams should collide head-on and be separated on either side away from the collision point. For pp the beams can be separated by a dipole magnet which acts on opposite currents in opposite directions. This separating dipole can be put inboard of the quadrupole triplet $k_2k_3k_4$.

In this case there could be two sets of triplets, one set for each of the two separated beams. Or, one can use one set of triplet commonly for both beams and separate the beams outboard of the triplet. For high energy colliders the very high strength required of the beam separating dipoles to impart sufficiently large separations within a short distance is always a problem.

For $\bar{p}p$ the beams can be separated only by an electric field. The beam separation that can be obtained is, thus, much smaller. Generally the two beams are not really separated, they are still contained in the same ring but only deflected by the electrostatic separator to go on obrits with opposite betatron oscillations. The beams are separated only if they are bunched and the bunches pass by each other near the crests of the betatron oscillations.

3. Dispersions and compensations

Because all beams have momentum spreads the dispersions of some of the linear lattice parameters are interesting. The following 3 are the most important.

a. Orbit length dispersion

This is important because it determines the transition energy and the longitudinal (phase) oscillation frequency. In a straight section the orbit length dispersion is zero. If the lattice is composed of only normal cells and straight sections the total orbit length dispersion is then

$$\frac{1}{\gamma_t^2} \equiv \frac{\Delta L}{L} \Big/ \frac{\Delta p}{p} = \frac{L_c}{2\pi R}\left(\frac{1}{\sin^2\mu} - \frac{1}{4}\right)\theta^2$$

where

L_c = total length of normal cells and

γ_t = transition energy in units of rest energy.

b. Tune dispersion (Chromaticity)

The tune dispersion is given by

$$\Delta\nu \Big/ \frac{\Delta p}{p} = \frac{1}{4\pi} \sum \beta \frac{B' \ell_Q}{B\rho}$$

where the summation is over all quadrupoles. For a normal cell this gives

$$\frac{\Delta\mu}{\pi} \Big/ \frac{\Delta p}{p} = \frac{1}{4\pi}\left(-k\beta_F + k\beta_D\right) = -\frac{1}{\pi} \tan\mu$$

If the ring is composed totally of N normal cells the chromaticity is

$$\xi \equiv \Delta\nu \Big/ \frac{\Delta p}{p} = N \frac{\Delta\mu}{\pi} \Big/ \frac{\Delta p}{p} = -\frac{N}{\pi} \tan\mu$$

(Beware that the chromaticity is sometimes defined as $\xi = \frac{\Delta\nu}{\nu} \Big/ \frac{\Delta p}{p}$.)
The density of quadrupoles in a straight section is generally much higher than that in the normal cell. Hence straight sections contribute a great deal to chromaticity. This is especially true for the low-β insertion because the quadrupoles on either side of the low-β point must be very strong and the β values there must be very large. The contribution from the low-β insertions is frequently larger than that from the normal cells.

Because of the unavoidable momentum spread in the beam, chromaticity must be controlled. To compensate or adjust the chromaticity, one places sextupoles in high dispersion (η) locations. The tune dispersion so introduced is given by

$$\Delta\nu \Big/ \frac{\Delta p}{p} = \frac{1}{4\pi} \sum \beta\eta \frac{B'' \ell_S}{B\rho} \quad , \qquad \ell_S = \text{length of sextupole}$$

where the summation is over all sextupoles. In general, two sets of sextupoles, one at high β_x locations and one at high β_y locations,

are needed to compensate for both the x and y chromaticities.

c. β dispersion

The value of β* is designed to be small to enhance luminosity. Large β* dispersion may make β* large for off momentum particles in the beam, thereby degrading the luminosity. The dispersion for the cell β values is

$$\frac{\Delta\beta}{\beta}\Big/\frac{\Delta p}{p} = -\tan^2\mu \qquad (\text{either } \beta_F \text{ or } \beta_D)$$

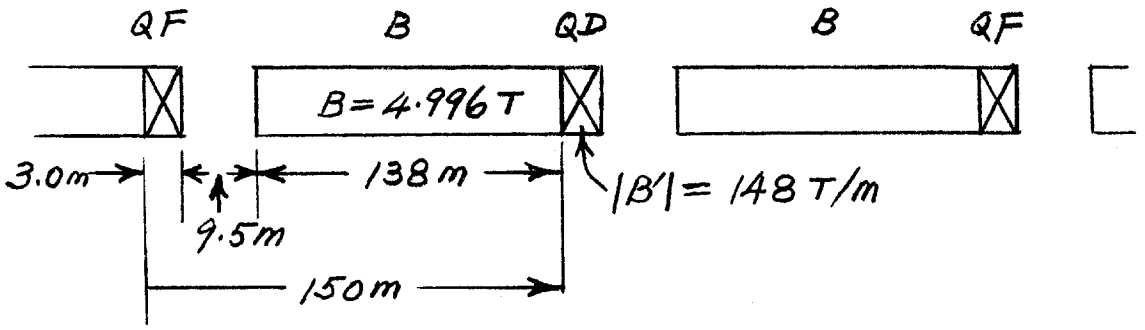
The dispersion of β* depends on the specific design of the insertion and is generally larger than that of the cell β. The exact value has to be computed using a computer. As a rule, the stronger are the insertion quadrupoles and the smaller is the β* value the greater is its dispersion.

4. Example — 20 TeV collider (Superconducting Super Collider)

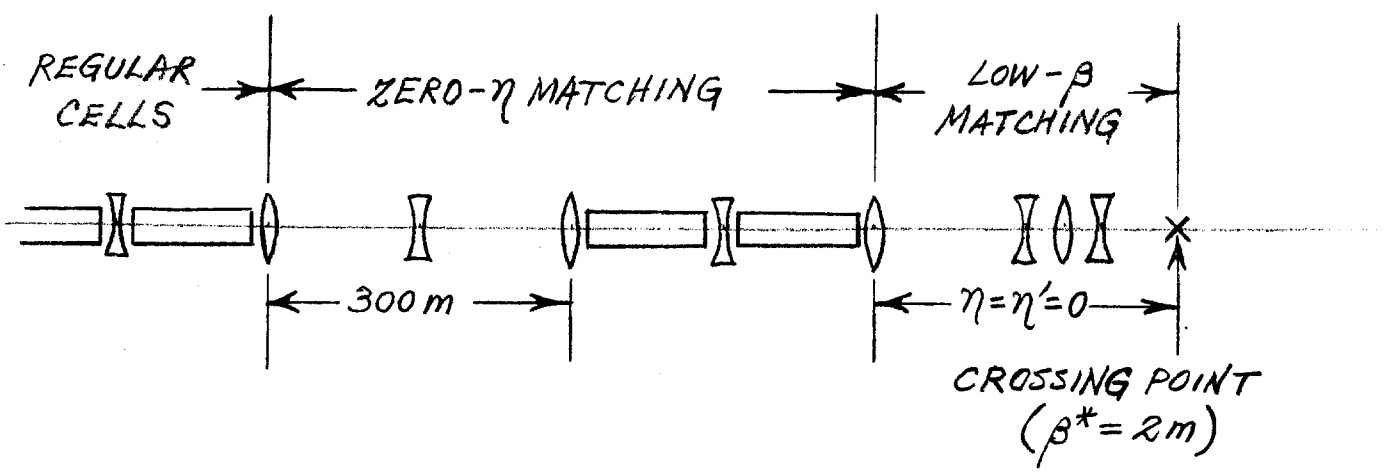
For the 20 TeV collider we assume $\mu = 30°$ FODO cells and eight symmetrically located beam-crossing matched straight-sections. The long straight-sections are matched for zero dispersion using the scheme of spacing out the dipoles and for low β* using triplets as described above. Some cutting and fitting is necessary. One set of consistent parameters is as follows:

Energy E	20 TeV (γ = 21316.6)
Rigidity $B\rho$	66716 Tm
Length of dipole ℓ_B	138 m
Number of dipole n_B	608
Total bending length $2\pi\rho = n_B\ell_B$	83904 m
Bending radius ρ	13354 m
Dipole field B	4.996 T
Dipole bend angle θ	10.334 mrad

Half cell length ℓ	150 m
Phase advance per half-cell μ	30°
Cell quadrupole focal length $B\rho/B'\ell_Q = f$	150 m
Cell quadrupole strength $B'\ell_Q$	445 T
Cell quadrupole length ℓ_Q	3.0 m
Cell quadrupole gradient B'	148 T/m



Number of cell quadrupoles n_Q		632
Number of low-β quadrupoles $n_Q{}^*$		64
Octant cell number		42
Regular cells	38	
Zero η cells (straight)	2	
Low β cells (straight)	2	
Octant length		12600 m
Circumference $2\pi R$		100800 m
Ring radius R		16043 m



Amplitude functions	
β_F	520 m
β_D	173 m
β^*	2 m
β_{max}	<2000 m
Dispersion functions	
η_F	7.75 m
η_D	4.65 m
η^*	0
Tune ν	58.67
Transition γ_t	52.5
Chromaticity ξ	-59 + "insertion" ~-130
Transverse emittance	
Normalized ε_n	10π mm-mrad
Unnormalized ε	0.00047π mm-mrad
Beam width	
Max. in cell	1.0 mm
Max. in insertion (β = 2000 m)	1.9 mm
At crossing (β^* = 2 m)	0.06 mm

B. RF (Longitudinal)

If bunched beams are desired the longitudinal confinement must be supplied by an rf system. The mathematical formulation of the phase motion is well known. We will only present here an unusual formulation which is patterned after that of the transverse motion and hence is easy to commit to memory. The proper coordinates are the deviation of the longitudinal position δz and the deviation of the longitudinal momentum δp from the central (reference) values. Instead of these we will use the dimensionless variables

$$\delta\phi \equiv \frac{1}{\lambda\!\!\!^{-}}\,\delta z\ , \qquad \delta\eta \equiv \frac{1}{mc}\,\delta p = \delta(\beta\gamma)$$

where mc is the rest momentum of the proton and $\lambda\!\!\!^{-} = \frac{R}{h} = \frac{\text{ring radius}}{\text{harmonic number}}$ is the rf wave length. Then the dimensions of the beam bunch, assumed to be much smaller than the rf bucket, are

$$\begin{cases} \delta\phi = 2\sqrt{\beta_\phi \dfrac{\varepsilon_\phi}{\pi}} = \text{bunch length} \\[2ex] \delta\eta = 2\sqrt{\dfrac{1}{\beta_\phi}\dfrac{\varepsilon_\phi}{\pi}} = \text{bunch width} \end{cases}$$

$$\nu_\phi = \frac{h\Lambda}{\eta}\,\frac{1}{\beta_\phi} = \quad \text{longitudinal oscillation wave number}$$

where ε_ϕ = longitudinal emittance $\equiv$ area in $(\delta\phi\delta\eta) = \frac{h}{mcR}$ [area in $(\delta z\delta p)$] and the longitudinal amplitude function β_ϕ is given by

$$\frac{1}{\beta_\phi^2} = \frac{1}{2\pi h}\,\frac{eV}{mc^2}\,\frac{\gamma}{\Lambda}\cos\phi_s \qquad \text{with} \qquad \Lambda \equiv \left|\frac{1}{\gamma_t^2} - \frac{1}{\gamma^2}\right|$$

with V = peak rf voltage and ϕ_s = synchronous phase. For the bucket we have

$$\begin{cases} \Delta\phi = \left[\phi_2(\phi_s) - \phi_1(\phi_s)\right] = \text{bucket length} \\ \Delta\eta = \dfrac{4}{\beta_\phi}\dfrac{\beta(\phi_s)}{\sqrt{\cos\phi_s}} = \text{bucket width} \\ A = \dfrac{16}{\beta_\phi}\dfrac{\alpha(\phi_s)}{\sqrt{\cos\phi_s}} = \text{bucket area} \end{cases}$$

where $\phi_1(\phi_s)$, $\phi_2(\phi_s)$, $\alpha(\phi_s)$, $\beta(\phi_s)$ are all listed in CERN/MPS-SI/Int. DL/70/4. For stationary bucket $\phi_s = 0$ and $\phi_2(0) = -\phi_1(0) = \pi$, $\alpha(0) = \beta(0) = 1$.

Several design considerations unique to the rf system of a high energy hadron collider should be mentioned.

1. Since, generally, high Q, high shunt impedance, fixed frequency cavities are used, one must be very careful in eliminating those parasitic modes which may excite longitudinal coherent instabilities.

2. Noise in the rf system blows up the longitudinal emittance of the beam bunches. This reduces the luminosity of the colliding beams and will eventually cause beam loss out of the rf bucket. Therefore, rf noise should be eliminated or, at least, reduced to a minimum.

3. The cavities should be located in straight sections where orbit dispersion is zero. This eliminates the first order coupling between the longitudinal and the transverse motion, and hence, synchro-betatron resonances such as

$$n_x\nu_x \pm n_y\nu_y \pm n_\phi\nu_\phi = m$$

with $n_\phi \neq 0$ and (n_x or n_y) $\neq 0$.

For the example of the 20 TeV collider SSC given above in A·4 with

Energy E	20 TeV (γ = 21316.6)
Circumference $2\pi R$	100.8 km
Longitudinal emittance ε_ϕ	
[area in ($\delta z \delta p$)	~3 eV sec]

we have, during storage at 20 TeV ($\phi_s = 0$)

Revolution frequency f_{rev}	2974.13 Hz
Harmonic number h	16800
R.F. frequency f_{rf}	49.9654 MHz
R.F. wave length $\lambda\!\!\!^{-} = R/h$	0.955 m
R.F. peak voltage V	1 MV
R.F. amplitude function β_ϕ	1.30
Bunch length $\delta\phi$	1.29
	($\delta\ell = \frac{R}{h}\,\delta\phi$ = 1.23 m)
Bunch width $\delta\eta$	0.99
	($\frac{\delta p}{p} = 0.46\times10^{-4}$)
Bucket area A	12.3

For acceleration from 1 TeV (at injection) to 20 TeV in 10 minutes we need

Average energy gain per turn $\left\langle\frac{\Delta E}{\Delta n}\right\rangle$	10.6 MeV
R.F. peak voltage V	13 MV
Synchronous phase ϕ_s	45° - 65°
Energy gain per turn $\frac{\Delta E}{\Delta n} = eV\sin\phi_s$	9.2 - 11.8 MeV
Bucket area A	1.62 - 1.67
R.F. amplitude function β_ϕ	1.91 - 0.554
Bunch length $\delta\phi$	1.56 - 0.84
	($\delta\ell$ = 1.49 - 0.80 m)
Bunch width $\delta\eta$	0.82 - 1.52
	($\frac{\delta p}{p} = 7.65 - 0.71\times10^{-4}$)

C. Short Term Stability (Single Beam)

1. Low order resonances

Construction errors are unavoidable. Error magnetic fields contain both non-linear components and deviations from ideal design linear components. Thus the lattice must be so designed as to avoid all low order resonances. The formulation of single resonances is well known.[2] All transverse resonances are of the form (assuming no coupling to the longitudinal)

$$n_x \nu_x \pm n_y \nu_y = m,$$

where n_x, n_y, m are all intergers and n_x, $n_y > 0$. The order of the resonance is defined by

$$\text{order} = n \equiv n_x + n_y.$$

These resonances fall into two major classes.

<u>Normal resonances</u> (n_y = even)

The nth order normal resonances are excited by the mth harmonic of the normal field coefficient

$$b_{n-1} \equiv \frac{1}{(n-1)!} \frac{\partial^{(n-1)} B_y}{\partial x^{(n-1)}}$$

Assuming that individual magnets are short (length ℓ) the driving amplitude of the above resonance can be written as

$$B_{n_x,n_y} = \frac{1}{2^n \pi B\rho} \binom{n}{n_y} \sum \left[(b_{n-1}\ell) \beta_x^{\frac{n_x}{2}} \beta_y^{\frac{n_y}{2}} \begin{matrix} \sin \\ \cos \end{matrix} (n_x \mu_x + n_y \mu_y) \right]$$

where n_y = even, $B\rho$ = magnetic rigidity of the particle, and β_x, β_y, μ_x, and μ_y are the amplitudes and phases of the linear oscillations. To trim out all n^{th} order normal resonance one needs n sets of normal 2n-pole

correction magnets installed in the ring and so adjusted as to make all n values of B_{nxny} zero.

Skew resonances (n_y = odd)

The nth order skew resonances are excited by the driving amplitudes

$$A_{n_x,n_y} = \frac{1}{2^n \pi B\rho} \binom{n}{n_y} \sum \left[(a_{n-1} l) \beta_x^{\frac{n_x}{2}} \beta_y^{\frac{n_y}{2}} \begin{matrix} \sin \\ \cos \end{matrix} (n_x \mu_x + n_y \mu_y) \right]$$

where

$$a_{n-1} \equiv \frac{1}{(n-1)!} \frac{\partial^{(n-1)} B_x}{\partial x^{(n-1)}}$$

to trim out all n^{th} order skew resonances one needs n sets of skew 2n-pole correction magnets so adjusted as to make the n values of A_{nxny} zero. A list of all resonances and driving amplitudes up to the 4th order (n=4) is given below:

	Normal Resonances			Skew Resonances		
Order	formula	driving amplitude	field coefficient	formula	driving amplitude	field coefficient
n=1 (Integer)	$\nu_x=m$	B_{10}	b_o	$\nu_y=m$	A_{01}	a_o
n=2 (Half Integer)	$2\nu_x=m$ $2\nu_y=m$	B_{20} B_{02}	b_1	$\nu_x \pm \nu_y=m$	A_{11}	a_1
n=3 (Third Integer)	$3\nu_x=m$ $\nu_x \pm 2\nu_y=m$	B_{30} B_{12}	b_2	$2\nu_x \pm \nu_y=m$ $3\nu_y=m$	A_{21} A_{03}	a_2
n=4 (Quarter Integer)	$4\nu_x=m$ $2\nu_x \pm 2\nu_y=m$ $4\nu_y=m$	B_{40} B_{22} B_{04}	b_3	$3\nu_x \pm \nu_y=m$ $\nu_x \pm 3\nu_z=m$	A_{31} A_{13}	a_3

These resonances have the following properties:

a. The normal linear field coefficients are to be interpreted as

$$b_o = \Delta B_y = \text{field error}$$
$$b_1 = \Delta \frac{\partial B_y}{\partial x} = \text{gradient error}$$

b. b_o(or B_{10}) leads to a horizontal closed orbit distortion, and a_o(or A_{01}) leads to a vertical closed orbit distortion. On the integer resonance (n=1) the closed orbit distortion goes to infinity unless the driving amplitude B_{10} or A_{01} is zero,

c. Only single-dimensional resonances and sum resonances may produce instabilities. Difference resonances lead only to coupling between x and y motions.

d. A half-integer resonance (n=2) has a stop band. Within the stop band all motions, however small, are unstable. This is because it is a linear resonance. Therefore it affects all motions, large or small, alike. The bandwidth depends on the driving amplitude and goes to zero when the driving amplitude goes to zero.

e. As one approaches a third-integer resonance (n=3) large amplitude oscillations become unstable first. Oscillations with infinitesimal amplitudes become unstable only when exactly on resonance. Thus, it does not have a stop band. The width is defined only in reference to the largest oscillation in the beam. Outside the resonance width all particles in the specific beam are stable.

f. For n>4 even when exactly on resonance there is still a region of stability around the closed orbit for oscillations with sufficiently small amplitudes, but generally this region is smaller than the area occupied by the phase points of the beam particles. Hence, generally a part of the beam will be unstable and be lost when exactly on these resonances.

g. For one single beam, resonances are excited only by field errors which are relatively weak. With two beams colliding the non-linear

electromagnetic forces due to one beam acting on particles in the other beam are much stronger and drive the resonances very hard. In storing a single beam it is generally sufficient to avoid resonances up to the 5th order, but for colliding beams one must avoid all resonances up to the 9th order or higher.

2. Self-field effects (coherent instabilities)

When the high current beam required for a collider travels down the conducting beam pipe it induces a voltage through an "impedance" of the beam pipe. This voltage can act back on the beam as positive feedback and make it unstable. Low frequency components of these coherent instabilities can be damped by negative electronic feedback systems, but high frequency components can only be checked by Landau damping derived from a spread in the natural frequencies of individual particles in the beam which causes the instability to lose coherence. The larger is the frequency spread and the smaller is the "impedance" the more stable is the beam.

The "impedance" depends and, therefore, imposes demands on the material, the structure, the shape and the size of the beam pipe. The tune spread $\Delta\nu$ is limited by non-linear resonances. The excitations of high order resonances by magnetic field errors are small and negligible beyond the octupole. But in colliders the excitation by beam-beam forces is large and resonances up to the 9th order must be avoided. This imposes a severe limitation on the allowable tune spread. This excitation is, however, independent of the orbit functions and hence makes no demand on the focusing strength. The dynamics of coherent instabilities of beams is a complex, multidimensional problem. To make our discussion understandable we will resort to using simplified semi-quantitative descriptions.

The condition for longitudinal stability (microwave modes) is at high energies[3]

$$\left\langle \frac{|Z_\ell|}{n} \right\rangle < F_\ell \frac{E/e}{I} \frac{1}{\nu^2} \left(\frac{\Delta p}{p}\right)^2$$

where we have used the approximation

$$\frac{1}{\gamma_t^2} - \frac{1}{\gamma^2} \cong \frac{1}{\gamma_t^2} \cong \frac{1}{\nu^2}$$

and where

γ_t = transition energy in units of rest energy

F_ℓ = beam distribution form-factor of order unity

Z_ℓ = longitudinal impedance

n = mode number = number of instability waves around the ring

< > denotes value weighed by the mode spectrum

E = energy of beam

I = peak current of beam

$\frac{\Delta p}{p}$ = FWHM of momentum spread.

The condition for transverse stability (high head-tail modes) is[4]

$$\langle |Z_t| \rangle < \pi F_t \frac{E/e}{I} \frac{\nu}{R} \Delta\nu$$

where

F_t = beam distribution form-factor of order unity

Z_t = transverse impedance

R = radius of ring

$\Delta\nu$ = tune spread in beam.

There are two main types of contribution to the impedance. The beam contribution depends on the energy and the dimensions of the beam and is non-zero even when the beam pipe is removed. This is generally small

for the range of parameters in consideration. The wall contribution is that due to the charge and current induced by the beam on the pipe wall and depends, therefore, on the material and the geometry of the beam pipe. The wall contributions of the longitudinal and the transverse impedances are related through the geometry of the pipe. For a circular beam pipe of radius b it is

$$Z_t = \frac{2R}{b^2}\frac{Z_\ell}{n}$$

There are two types of terms in the wall contribution to Z_ℓ/n. The "smooth" term, usually known as the resistive wall term, is that of a perfectly uniform and smooth pipe and depends on the size and the skin depth of the pipe. It is rich in low frequencies and is generally small. The most important is the "interruptions" term arising from discontinuities in the pipe and from various beam sensing and manipulating devices inserted in the pipe. This term in Z_ℓ/n is not sensibly dependent on the pipe size. For further discussion we shall consider the above equation as the approximate relation between the total contributions to the impedances.

Substituting this relation in the condition for Z_T we can rewrite the condition for transverse stability as

$$\left\langle \frac{|Z_\ell|}{n} \right\rangle < \frac{\pi}{2} F_t \frac{E/e}{I} \left(\frac{b}{R}\right)^2 \nu \Delta\nu .$$

Solving the two conditions on $\langle |Z_\ell|/n \rangle$ for ν we see that the choice of ν is hemmed in by longitudinal and transverse stability requirements as

$$B\left\langle \frac{|Z_\ell|}{n} \right\rangle < \nu < A\left\langle \frac{|Z_\ell|}{n} \right\rangle^{-\frac{1}{2}}$$

with

$$\begin{cases} A \equiv \left(\frac{E/e}{I}\right)^{\frac{1}{2}} \frac{\Delta p}{p} \propto \left(\frac{E}{I}\right)^{\frac{1}{2}} \frac{\Delta p}{p} \\ B \equiv \frac{2}{\pi} \frac{I}{E/e} \left(\frac{R}{b}\right)^2 \frac{1}{\Delta\nu} \propto \frac{I}{E} \left(\frac{R}{b}\right)^2 \end{cases}$$

where, consistent with the approximation, we have put $F_\ell = F_t = 1$. The available range for ν shrinks to zero when

$$\left\langle \frac{|Z_\ell|}{n} \right\rangle = \left(\frac{A}{B}\right)^{\frac{2}{3}} = \left(\frac{\pi}{2}\right)^{\frac{2}{3}} \frac{E/e}{I} \left(\Delta\nu \frac{\Delta p}{p}\right)^{\frac{2}{3}} \left(\frac{b}{R}\right)^{\frac{4}{3}}$$

At this point the single allowable value of ν is

$$\nu = (A^2 B)^{\frac{1}{3}} = \left[\frac{2}{\pi} \frac{1}{\Delta\nu} \left(\frac{\Delta p}{p}\right)^2 \left(\frac{R}{b}\right)^2\right]^{\frac{1}{3}}$$

This is a good value to choose for ν in any case because it allows the largest value of the impedance $\langle |Z_\ell|/n \rangle$.

For the 20 TeV collider we have:

Ring radius R	16043 m
Aperture radius b	0.025 m
Tune ν	58.67
Number of p per bunch n	2×10^{10}
Tune spread in beam $\Delta\nu$ (limited by resonances)	0.01

and at various stages of acceleration and storage we have

	Acceleration		Storage
E	1 TeV	20 TeV	20 TeV
$\delta\ell$	1.49 m	0.80 m	1.23 m
I	0.64 A	1.20 A	0.78 A
$\delta p/p$	7.65×10^{-4}	0.71×10^{-4}	0.46×10^{-4}
A	$953\ \Omega^{\frac{1}{2}}$	$291\ \Omega^{\frac{1}{2}}$	$\boxed{235\ \Omega^{\frac{1}{2}}}$
B	$\boxed{16.9\ \Omega^{-1}}$	$1.57\ \Omega^{-1}$	$1.026\ \Omega^{-1}$

we should use the smallest A and the largest B (boxed values) and obtain the conditions

$$(16.9\,\Omega^{-1})\left\langle\frac{|Z_\ell|}{n}\right\rangle < \nu < (235\,\Omega^{\frac{1}{2}})\left\langle\frac{|Z_\ell|}{n}\right\rangle^{-\frac{1}{2}}$$

The optimal tune is, then, $\nu = 97.7$ and the condition on impedance is $\langle|Z_\ell|/n\rangle < 5.8\ \Omega$ which is relatively easy to satisfy. Although the tune of the ring given in A·4 is not optimal, with

$$\nu = 58.67 \quad \text{and} \quad \gamma_t = 52.5$$

the original stability conditions become

$$\left\langle\frac{|Z_\ell|}{n}\right\rangle < \begin{cases} 20\ \Omega & \text{longitudinal at storage} \\ 3.5\ \Omega & \text{transverse at injection.} \end{cases}$$

We see that the condition for transverse stability at injection is most limiting. But even that is not bad.

Generally we would like to increase A and/or to decrease B. To do this we should:

a. Increase E/I. This extends the acceptable ν-range at both ends. This also shows that the tightest constraint occurs at injection when E is lowest. Reducing I helps, but the luminosity suffers.

b. Increase $\Delta p/p$. This raises the upper limit of the ν-range, but requires either blowing up the longitudinal emittance or a huge increase in rf voltage (as the 4th power of $\Delta p/p$). Neither alternative is attractive.

c. Increase b/R. Because of the squared dependence this is very effective in lowering the lower limit of the ν-range. Since the stored energy in the magnet ring is proportional to b^2R, to minimize the increase in stored energy it is more desirable to reduce R than to increase b.

3. Gas scattering and vacuum instability

The residual gas in the vacuum chamber causes two undesirable effects.

a. It is ionized by the beam and the positive ions are driven onto the chamber wall by the positive proton beam (say). These primary ions will knock out from the wall adsorbed gas molecules which will be further ionized by the beam and cause an avalanche.[5] The vacuum will go bad and the beam will be destroyed. This instability can be cured either by reducing the wall desorption or by improving the vacuum, namely reducing the residual gas.

b. The residual gas degrades the beam either by nuclear scattering or by multiple Coulomb scattering. The former kills the beam particles and the latter enlarges the beam, either will reduce the luminosity of the colliding beams. Again a high vacuum is required. Generally a vacuum of $<10^{-10}$ Torr is adequate.

4. Intrabeam scattering[6]

In the rest frame of a beam bunch the particles are confined in a 3-dimensional potential well. In addition, the particles interact via Coulomb scattering. This intrabeam scattering can be expected to cause growth in the 6-dimensional emittance. The growth rate is given by

$$\frac{1}{\tau} = \frac{\pi^2}{\gamma} \frac{c r_0^2 N (\log)}{\Gamma} \langle H(\lambda_1, \lambda_2, \lambda_3) \rangle$$

where

c = speed of light

$r_o = \frac{e^2}{mc^2}$ = classical proton radius

N = number of protons in bunch

log = Coulomb logarithm $\cong 20$.

The quantity $\Gamma = (2\pi\beta\gamma)^3 \varepsilon_h \varepsilon_v \varepsilon_\ell$ is the 6-dimensional normalized phase volume where

$$\varepsilon_q \equiv \frac{1}{mc\beta\gamma} \sigma_q \sigma_p \qquad \text{for} \qquad q = h, v, \ell$$

and

q,p = coordinate and momentum variables in the q degree-of-freedom

σ denotes the rms spread

$mc\beta\gamma$ = central momentum of beam.

The factor $\langle H(\lambda_1,\lambda_2,\lambda_3)\rangle$ is a dimensionless and homogeneous "momentum shape factor". $\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \frac{1}{\sqrt{\lambda_3}}$ measure the principal axes of the momentum ellipsoid of the beam bunch. $H = 0$ if $\lambda_1 = \lambda_2 = \lambda_3$, namely when the momentum spread is isotropic. $\langle\ \rangle$ denotes averaging around the ring. In an alternating gradient lattice the λ's cannot be equal everywhere. Hence the emittance will always grow.

In the general case, H can be expressed in terms of elliptic integrals, but in the special cases when $\lambda_1 > \lambda_2 \cong \lambda_3$, namely when the momentum distribution is an oblate circular ellipsoid with axis along the 1 direction, one can write

$$H = \frac{2(\lambda_1 + 2\lambda_2)}{\sqrt{\lambda_2(\lambda_1 - \lambda_2)}} \sin^{-1}\sqrt{\frac{\lambda_1 - \lambda_2}{\lambda_1}} - 6$$

Formulas, slightly more complicated, exist also for $\frac{1}{\tau_q}$, $(q = h,v,\ell)$,

namely growth rates for the individual horizontal, vertical and longitudinal emittances. For these the reader is referred to the reference.

For the type of lattices discussed above, it is usually slightly damping in the vertical plane ($\frac{1}{\tau_V} < 0$) and at high energies one can generally reduce the horizontal growth rate to less than 0.1 hr^{-1} and the longitudinal growth rate to an order-of-magnitude less than that.

D. Long Term Stability (Beam-Beam)

In section C·1 we discussed the effects of the non-linear error field on the particle motion. The error field has the following characteristics:

1. It arises from construction errors and is therefore weak.

2. Its non-linear components fall off steeply with increasing multipole order. Therefore only the low-order resonances are excited by error fields and only weakly. Low-order resonances are relatively widely spaced. With weak excitation, hence narrow widths, these resonances can be considered separated and can be treated individually by the single resonance formalism. This formalism concludes that as long as the motion stays outside the widths of individual low-order resonances it is stable.

This situation is totally different when the motion is acted on by the beam-beam forces. For colliding beams the electromagnetic forces exerted by one beam on a particle in the other beam is extremely non-linear in the transverse coordinates x and y and is like a δ-function in the longitudinal coordinate s, hence is extremely rich in harmonics. The non-linear resonances are strongly excited and the excitation remains strong and does not fall off with increasing resonance order. The high-order resonances are closely spaced and the motion necessarily straddles a near contimuum of high-order resonances. The gross behaviors of such an intrinsically non-linear motion are well known.[7] Such motions are complicated by the occurrence of two distinct regimes - the regular regime ("laminar" in phase space) and the stochastic regime ("turbulent" in phase space). Over the past 100 years or so a great deal of effort of many prominent mathematicians has been devoted to the study of the existence and the characteristics of these regimes of motion and the transition between them, and specifically to the determination of the stability of the stochastic regime. But, so far, we still do not have definitive

resolutions to these probems. These questions are clearly of crucial relevance to the long term stability of colliding beams. For designing colliders at the present we can only depend on semi-empirical models and scaling laws as guides. We give here a demonstration that "similarity parameters" can be established for such motions which can be used to specify the transition between the regimes. This is similar to the use of the Reynold's Number to specify the onset of turbulence in the motion of viscous fluids.

If the transverse force potential exerted by one beam bunch on a particle in the other beam is written as $V(x^2,y^2)\delta(s)$ [generally even in x and y] the Hamiltonian for the motion of the particle is

$$H = \frac{1}{2}(p_x^2 + K_x x^2) + \frac{1}{2}(p_y^2 + K_y y^2) - V(x^2, y^2)\,\delta(s)$$

The usual canonical transformation to angle/action variables $(\phi_x, J_x, \phi_y, J_y)$

$$\begin{cases} x = \sqrt{2\beta_x J_x}\,\cos\phi_x \\ p_x = -\sqrt{\dfrac{2J_x}{\beta_x}}\left(\sin\phi_x - \dfrac{\beta_x'}{2}\cos\phi_x\right) \end{cases} \qquad \text{(similar for y)}$$

where $\beta_x(s)$, $\beta_y(s)$ are the Courant-Snyder amplitude functions, and transformation to azimuthal inedpendent variable $\theta \equiv s/R$ with $2\pi R$ = circumference of ring, give for the transformed Hamiltonian

$$K = \nu_x J_x + \nu_y J_y - V(\beta_x J_x \cos^2\phi_x,\ \beta_y J_y \cos^2\phi_y)\,\delta(\theta)$$

and the corresponding canonical equations

$$\begin{cases} \dfrac{d\phi_x}{d\theta} = \dfrac{\partial K}{\partial J_x} = \nu_x - \dfrac{\partial V}{\partial J_x}\,\delta(\theta) \\ \dfrac{dJ_x}{d\theta} = -\dfrac{\partial K}{\partial \phi_x} = \dfrac{\partial V}{\partial \phi_x}\,\delta(\theta) \end{cases} \qquad \text{(similar for y)}$$

If we assume, say, a bi-Gaussian beam distribution

$$\rho = \frac{N}{2\pi\sigma_x\sigma_y}\,\delta(s)\,\exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right)$$

we get

$$V(x^2, y^2) = \frac{r_0 N}{\gamma}\int_0^\infty dt\,\frac{1-\exp\left[-\dfrac{x^2}{2(\sigma_x^2+t)} - \dfrac{y^2}{2(\sigma_y^2+t)}\right]}{\sqrt{(\sigma_x^2+t)(\sigma_y^2+t)}}$$

$$= \frac{r_0 N}{\gamma}\int_0^\infty d\tau\,\frac{1-\exp\left[-\dfrac{1+r}{1+r\tau}\,\dfrac{\beta_x J_x}{\sigma_x(\sigma_x+\sigma_y)}\cos^2\phi_x - \dfrac{1+r}{r+\tau}\,\dfrac{\beta_y J_y}{\sigma_y(\sigma_x+\sigma_y)}\cos^2\phi_y\right]}{\sqrt{\frac{1}{r}(1+r\tau)(r+\tau)}}$$

$$\equiv \frac{r_0 N}{\gamma} F_r\left(\mathcal{J}_x, \mathcal{J}_y, \phi_x, \phi_y\right)$$

where $\tau \equiv \dfrac{t}{\sigma_x\sigma_y}$, $r \equiv \dfrac{\sigma_y}{\sigma_x}$, $\mathcal{J}_x \equiv \dfrac{\beta_x J_x}{\sigma_x(\sigma_x+\sigma_y)}$, $\mathcal{J}_y \equiv \dfrac{\beta_y J_y}{\sigma_y(\sigma_x+\sigma_y)}$,

r_0 ≡ classical proton radius, and the subscript r on F indicates that F depends parametrically on and only on r. Now we can rewrite the canonical equations as

$$\begin{cases} \dfrac{d\phi_x}{d\theta} = \nu_x - \dfrac{r_0 N \beta_x}{\gamma\sigma_x(\sigma_x+\sigma_y)}\,\dfrac{\partial F_r}{\partial \mathcal{J}_x}\,\delta(\theta) \equiv \nu_x - 2\pi\xi_x\,\dfrac{\partial F_r}{\partial \mathcal{J}_x}\,\delta(\theta) \\ \dfrac{d\mathcal{J}_x}{d\theta} = \dfrac{r_0 N \beta_x}{\gamma\sigma_x(\sigma_x+\sigma_y)}\,\dfrac{\partial F_r}{\partial \phi_x}\,\delta(\theta) \equiv 2\pi\xi_x\,\dfrac{\partial F_r}{\partial \phi_x}\,\delta(\theta) \end{cases}$$

(similar for y)

Thus we see that the motion is uniquely specified by the 5 parameters

$$\nu_x\,, \quad \xi_x \equiv \frac{1}{2\pi}\frac{r_0 N \beta_x}{\gamma\sigma_x(\sigma_x+\sigma_y)}$$

$$\nu_y\,, \quad \xi_y \equiv \frac{1}{2\pi}\frac{r_0 N \beta_y}{\gamma\sigma_y(\sigma_x+\sigma_y)} \qquad \text{and} \qquad r \equiv \frac{\sigma_y}{\sigma_x}$$

It is easy to show that to the lowest order in x and y (quadratic in H) ξ_x and ξ_y are just the tune shifts.

If one is interested only in the onset of the stochastic regime ν_x, ν_y and r are all irrelevant. Transition from the regular to the stochastic regimes will occur when either one of the tune shifts ξ_x and ξ_y reaches the critical value, hence r is irrelevant. The tunes ν_x and ν_y enter only to relate the phases of the kicks given by $V(x^2,y^2)\delta(s)$. In the stochastic regime the kicks are random and ν_x and ν_y become also irrelevant. This argument also indicates that the tune shifts of multiple widely separated colliding points around the ring should not be straight-summed. They should at most be added stochastically (in quadrature).

It is generally believed that the beam-beam limit, namely the critical value of the tune shift beyond which the life time of the stored beam becomes too short to be useful, does not have to be the point at which the motion is stochastic over the entire phase space. Even far below this point the motion is already stochastic within thin layers along the separatrices of resonances. For motions in two degrees-of-freedom these stochastic layers in the shape of tori in the 4-dimensional phase space intersect one another and extend out to infinity. There are, therefore, continuous stochastic channels extending to infinity. Random, diffusion-like motions can follow these channels outward and become unstable. This is

known as Arnol'd diffusion. At the beam-beam limit the intersecting stochastic layers have gotten so thick that the Arnol'd diffusion rate, perhaps aided by real diffusion processes due to physical noises, have gotten so high that the beam life time becomes too short to be useful. Thus, for proper colliding beams operation ξ must stay below the beam-beam limit. Because of synchrotron radiation damping for electron beams, it is expected that the beam-beam limit would be higher for electron (positron) colliding beams than for proton (antiproton) colliding beams. No theory exists at the present to predict these limiting values. Experiences on existing colliders seem to show that the beam-beam limit is $\xi \cong 0.05$ for electrons and $\xi \cong 0.003$ for protons.

II. Superconducting Magnets

For high energy hadron colliders consideration of electric power consumption compels one to use superconducting magnets. Of course, one could simply use conventional iron yokes operating at some 1.8 T but energized with superconducting coils. Indeed, these "superferric magnets" have been proposed for some high energy hadron colliders. In addition to saving power, the higher current density that can be carried in superconductors enables one to obtain higher field intensities with superconducting coils. For these magnets the iron yokes function mainly as magnetic shields to keep the field from spreading all over and/or structure members to confine the coils against the exploding magnetic forces. We shall discuss here only these high field superconducting magnets. Combined with the well known design considerations of conventional magnets the discussion given here will also be adequate for the design of the superferric magnets.

A. Superconductors and Cables

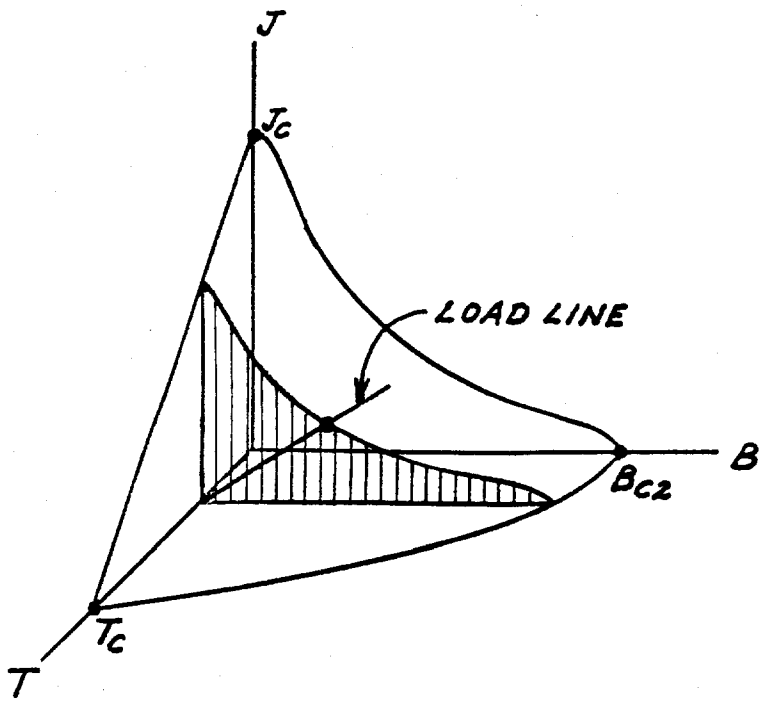
The characteristics of a Type II superconductor are usually represented by the triple diagram shown on the top part of next page where

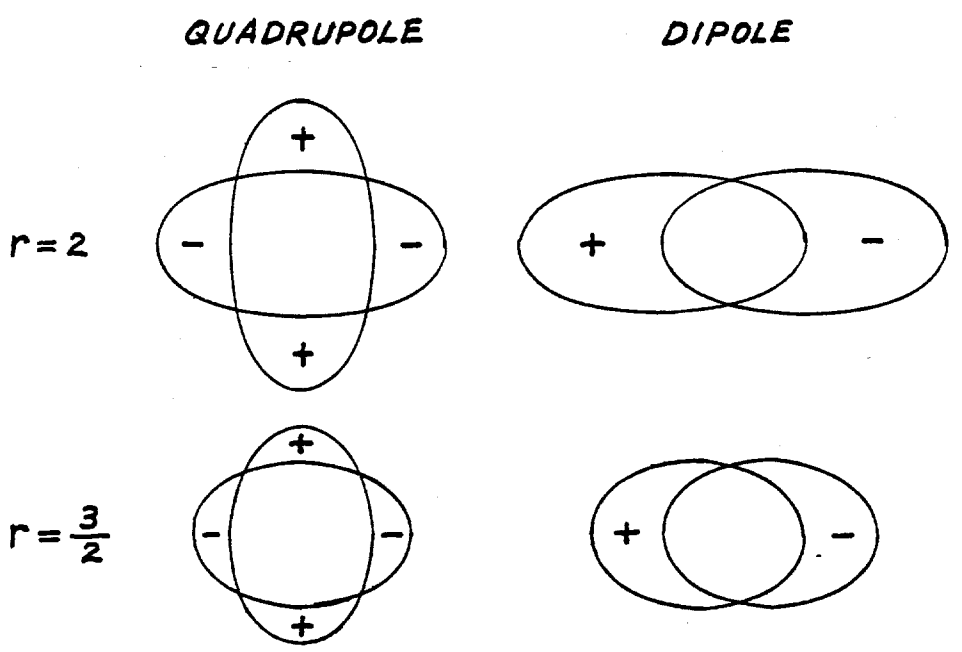
J_c = critical current density

T_c = critical temperature

B_{c2} = upper critical field

For a Type II superconductor at some fixed low values of J and T starting at high externally applied field B, the conductor is normal. As B decreases the resistivity starts to deviate from (fall below) the normal value at B_{c2} and goes to zero at the lower critical field B_c. For Type I superconductors such as pure Pb, $B_c = B_{c2}$ and the resistivity vanishes suddenly at B_c. T_c and B_{c2} are properties of the material. Simple theories[8]


J
J_c
LOAD LINE
B
B_{c2}
T_c
T


QUADRUPOLE
DIPOLE
r = 2
r = 3/2
+
−

relate T_c and B_{c2} through normal-state property values and show that the B vs T curve at J = 0 is a parabola. The following table gives T_c and B_{c2} for most of the known high-field superconductors.

Superconductors	T_c(K)	B_{c2}(T)
bcc Alloy		
V(40 atomic %) Ti	7.0	11
Nb(56 atomic %) Ti	9.0	14.1
Nb(25 atomic %) Zr	10.8	9.2
A-15 compound		
V_3Ga	14.8	25
V_3Si	16.9	24
Nb_3Sn	18.0	28
$Nb_3A\ell$	18.7	33
Nb_3Ga	20.2	34
Nb_3Ge	22.5	38
$Nb_3(A\ell_{0.7}Ge_{0.3})$	20.7	43.5
Ternary Sulfide		
$PbMo_{5.1}S_6$	14.4	60

Nb_3Ge has the highest T_c of ~23 K and $PbMo_{5.1}S_6$ has the highest B_{c2} of ~60 T. J_c and the J vs B curve depend on the geometry of the conductor and on the lattice defects in the material. Lattice defects such as dislocations, impurities or precipitates of a second phase act to "pin" the flux lines and prevent them from moving under the Lorentz force J x B. Hence JB is roughly constant in the middle part of the J vs B curve at fixed T. The value of JB depends crucially on the cold working and

annealing of the material. Recent efforts in Japan and in China have produced NbTi conductors with extremely high JB values, but in magnet design 6×10^9 T A/m^2 is generally considered a practical and safe value for NbTi at 4.2 K.

In practical application multi-filament conductors with NbTi filaments set in Cu matrix are used. The filaments are twisted to reduce the effects of flux jump. The cold Cu provides a low resistivity parallel current path so that if the superconductor locally goes normal the current is shunted across in the Cu, thereby prevented from damaging the superconductor. If the Cu to superconductor ratio is high, say >10 to 1 the heating of the Cu by the shunted current is small and can be carried away by the coolant (liquid He). The temperature of the cable will remain low and the superconductor can recover. Such a cable is called fully stabilized. The average current density in a coil wound with fully stabilized conductor is necessarily low and can only be ~1/30 of that in the superconductor. The minimum Cu to superconductor ratio needed to prevent damage is ~1 to 1. For such a cable the shunted current heats up the Cu so much that the heat cannot be carried away sufficiently fast by the coolant. The heating and the normal region will propagate and the whole magnet will quench. Nevertheless no damage will come to the coil. After the power is turned off the coil can be cooled down again and re-energized in a matter of minutes. The average current density in such a substabilized (sub-fully stabilized) coil could be as high as 1/3 of that in the superconductor.

For d.c. application the filaments could be quite thick, up to 50 μm, and a wire carrying a maximum current of 150-200 A is appropriate. For pulsed application the filament diameter should be smaller than ~10 μm and many strands of the 200 A wire, say 20 to 30, must be made into a cable

in which the strands are transposed, say every 5 cm, along the length to reduce the eddy current effects. A magnet wound with such a high current (~5000 A) cable will have reasonably small number of turns and hence, reasonably low inductance appropriate for pulsing.

The JB values for A-15 compounds are generally higher, being ~6×10^{10} TA/m^2 for Nb_3Sn at 4.2 K. Unfortunately all A-15 compounds are very brittle. Chemically sprayed Nb_3Sn on Cu tape or wire with a very large number (~5×10^4) of very fine (<5 μm diameter) Nb_3Sn filaments set in a Cu matrix are available for winding coils provided the bending curvature is not excessive.

B. Magnet Configuration and Structure

In an infinitely long solenoid one can get arbitrarily large field with any available current density by simply piling layers after layers of coil windings on the outside. This is not true for multipole magnets. The maximum field or its spatial derivatives are limited by the attainable current density. For coils with uniform current density over their cross-sections to produce dipole and quadrupole fields the proper cross-sectional shapes are given by intersections of ellipses as shown in the bottom figures on p. 33 where $r \equiv \frac{a}{b}$ is the ratio of the semi-major to semi-minor axes and where the aperture is adjusted to be roughly circular with radius $\cong$ b. This figure shows that for a dipole $r \sim \frac{3}{2}$ is reasonable and that for a quadrupole r can be ~2. In practice, the crescent shaped areas are approximated by some arrangement of turns of the conductor that is easy to wind. In the Fermiab Tevatron dipole the crescent area is approximated by two circular shells of windings.

For the intersecting-ellipses coils with uniform current density J the dipole field B and the quadrupole field gradient G are given by

$$B = \mu_0 J\left(\frac{r-1}{r+1}\right) 2b$$

$$r \equiv \frac{a}{b} .$$

$$G = \mu_0 J\left(\frac{r-1}{r+1}\right)$$

The highest fields on the coils are B_{coil} ~B for dipole and B_{coil} ~Gb for quadrupole. Therefore, if the limiting parameter is the Lorentz force JB we can rewrite the above equqtions as

$$B^2 = \mu_0 (JB)\left(\frac{r-1}{r+1}\right) 2b$$

$$G^2 = \mu_0 (JB)\left(\frac{r-1}{r+1}\right) \frac{1}{b}$$

For a fully stabilized NbTi coil at 4.2 K with $(JB)_{max}$ = (1/30) x 6×10^9 TA/m^2, b = 0.1 m and $\mu_0 = 4\pi \times 10^{-7}$ Vsec/Am we get

$$B_{max} = 3.2\ T \qquad \left(r = \frac{3}{2}\right)$$

$$G_{max} = 29\ T/m \qquad (r = 2)$$

Going to substabilized NbTi conductor we can gain a factor 10 in $(JB)_{max}$ and hence a factor 3 in B_{max} and G_{max}. These ideal values should, of course, be interpreted as the upper limits of what are attainable in practice. As mentioned before these high fields are available only with high current densities. At the same current density even if we go to $r = \propto$ (namely $a = \propto$) we get only an increase of a factor $\sqrt{5}$ in B and a factor $\sqrt{3}$ in G from those given above.

At these high fields the magnetic forces on the coil are very large, generally many tons per cm of length of magnet. These forces must be confined by some mechanical structure. Also, since one generally does not like the magnetic field to extend far away from the magnet all

coils should be magnetically shielded by an iron yoke. There are, then, two ways of providing the mechanical confinement.

1. Cold-iron structure

One could use the iron yoke to confine the forces on the coil. The yoke must then be in contact with the coil and be at the coil temperature. The cryostat must then enclose the entire coil and yoke. The mass to be cooled is very large. The cooldown time will be long. The iron next to the coil will be driven into magnetic saturation by the very strong field thereby distorting the field shape. The advantage, on the other hand, is that the iron yoke does provide a very simple and strong confinement structure.

2. Warm-iron structure

One could use a non-magnetic (say, stainless steel) collar to confine the forces on the coil, enclose only the coil and the collar in a smaller cryostat, and place the cryostat inside a room-temperature iron yoke or shield. The advantage is that the cooldown time is greatly reduced and that if the iron is far enough removed from the coil it will not be magnetically saturated and will not distort the field shape. The disadvantages are (a) the collar generally does not provide as strong a confinement structure as the heavy iron yoke, (b) the smaller cryostat generally results in greater heat leak, and (c) supporting the cryostat from the warm iron yoke is mechanically and thermally tricky. The Tevatron magnets use the warm-iron design.

Any movement of the coil conductor will create friction heating, thereby causing the conductor to go normal. Nearly all design difficulties are related to the requirements of strength and rigidity of the confinement structure and of thermal insulation of the cryostat. There is no problem with the superconducting behaviors of the conductors. Design concerns related

to the coil conductor are generally simple, mundane considerations such as avoiding large eddy current loops, avoiding breakage of the delicate superconducting filaments, providing good contact between conductor and coolant etc. As long as these fairly obvious requirements are satisfied the magnet will work reliably to the short sample limit. The mechanical and thermal requirements, although obvious, are, however, extremely demanding. It is likely that the ultimate performance of superconducting magnets will be limited by strength of material and mechanical engineering rather than superconducting properties of material.

C. Beam Quenching and Quench Protection

A substabilized superconducting magnet can quench (go normal) for a variety of reasons. The control system may malfunction, a current glitch may arise from the power supply, the coil conductor may move, the particle beam may strike the coil etc. The last mentioned is the most difficult to avoid. At liquid He temperature the heat capacity of the conductor is very small, less than $\frac{1}{2}$ mJ/g/K, and during the sub-millisecond time when it is struck by the beam, cooling by the coolant is totally ineffective. Depending on how close to the quench limit it is operating the magnet may quench when struck by very little beam. For example, $\sim\frac{1}{5}$ mJ/g ($\sim 2\times10^{7}$ GeV/cm^{3}) will give a temperature rise of $\sim\frac{1}{2}$ K and may cause a quench. The shower developed by one high energy proton will deposit some 5×10^{-3} GeV/cm^{3}. Therefore it takes a stray beam of only 4×10^{9} protons striking the superconductor within 1 cm^{2} and in a time <1 m sec to cause the magnet to quench. It is therefore essential that the superconducting magnets downstream of beam scatterers such as scrapers, setpa, etc. be well shielded, and good and reliable quench protection systems are incorporated in the design.

When the spot on the coil that is struck by beam quenches it heats up causing neighboring areas to quench in succession and the quench

propagates. If all the stored energy in the magnet is dumped at one spot the coil will surely be damaged. Thus we like the quench to propagate rapidly to the whole coil. To do this a heating strip is imbedded in the coil during construction. When the start of a quench is detected a current is sent through the heating strip to cause the whole coil to quench at once. In addition, an external circuit is triggered to (1) shunt the ring current around the quenched magnet, and (2) short the quenched magnet across a resistor. A large fraction of the stored energy is then absorbed by the resistor and the remaining small fraction is distributed over the whole coil thereby causing no damage. Experience shows that such a quench protection system can indeed be made effective and reliable.

III. Beam Cooling and Antiproton Accumulation

If only one ring is available one must collide p and $\bar{p}$. For this one needs a source of high phase-space density $\bar{p}$.

Antiprotons produced by a high energy proton beam striking a target have a very low phase-space density. There are two methods presently available to increase the phase-space density or, equivalently, to reduce the phase-space volume occupied by the beam - a process similar to cooling a volume of gas molecules.

A. Electron Cooling[9]

Because of the much lower mass, electron beams colder (lower random kinetic energy) than the $\bar{p}$ beam can be produced relatively easily. A cold electron beam with the same velocity is made to travel along and mix with the stored $\bar{p}$ beam in a straight section of the storage ring. In the rest frame this is just a 2-component plasma. Equipartition of energy between the 2 components through Coulomb interaction will cool the antiprotons and heat the electrons.

We give here a sketch of the derivation for the cooling rate in a simplified case. The velocity dependent "friction" force on a test charge (in this case the $\bar{p}$) moving in a spatially uniform electron distribution $f(\vec{v})$ is given by

$$\vec{F}(\vec{v}) = -\left(4\pi e^4 \frac{n_e^*}{m_e} \ln\Lambda\right) \nabla_v \phi$$

with the "potential" ϕ given by the Poisson-like equation in the velocity space

$$\nabla_v^2 \phi = -4\pi f(\vec{v})$$

where

e, m_e = charge, mass of electron

n_e^* = spatial density of electron distribution (*denotes value in rest frame)

$\ln\Lambda = \ln(12\pi\, n_e^*\, \lambda_D^3)$ = Coulomb logarithm

$$\lambda_D = \left(\frac{1}{4\pi r_e}\frac{\bar{\beta}_e^{\,2}}{n_e^*}\right)^{1/2} \quad = \text{Debye screening length}$$

$$r_e = \frac{e^2}{m_e c^2} = \text{classical electron radius}$$

$\bar{\beta}_e$ = rms value of the random part of $\frac{v}{c}$ in the rest frame of the electron distribution (related to the "temperature").

A swarm of $\bar{p}$'s will be cooled ("attracted" to the "velocity center" of the electron distribution) at the rate

$$\frac{1}{\tau^*} = \frac{1}{m}\nabla_v \cdot \vec{F} = \frac{1}{m}\left(4\pi e^4 \frac{n_e^*}{m_e}\ln\Lambda\right)\left(-\nabla_v^2\phi\right)$$

$$= (4\pi)^2 \frac{e^4}{m m_e} n_e^* \ln\Lambda\, f(\vec{v})$$

where m and $\vec{v}$ are the mass and the velocity of the $\bar{p}$. Assuming Maxwellian electron distribution and $v \ll v_e$ (Because of the much larger $\bar{p}$ mass this is generally valid even though the $\bar{p}$ beam is hotter - higher random energy.), and transforming to the lab-frame we get

$$\frac{1}{\tau} = \sqrt{32\pi}\, r_e r_p \ln\Lambda \left(\frac{1}{\beta\gamma^2}\frac{j_e/e}{\bar{\beta}_{ex}\bar{\beta}_{ey}\bar{\beta}_{ez}}\right)\eta$$

where

$r_p = \frac{e^2}{mc^2}$ = classical $\bar{p}$ radius

$\beta\gamma^2$ = relativistic kinematic factor of the $\bar{p}$ beam (identical for the e beam)

j_e/e = number current density of the electron beam

$\bar{\beta}_{ex}, \bar{\beta}_{ey}, \bar{\beta}_{ez}$ = standard deviations (rms) of the (Maxwellian) electron distribution in the rest-frame and in the 3 degrees-of-freedom indicated by the subscripts.

η = duty factor = fractional part of the $\bar{p}$ orbit overlapping with electron beam for cooling.

This shows that the cooling rate is higher for:

1. higher electron current density j_e
2. colder electron beam (smaller $\bar{\beta}_e$)
3. lower $\bar{p}$ beam energy (smaller $\beta\gamma^2$).

The realistically attainable cooling rate is unfortunately rather low for this application. Even for 200 MeV $\bar{p}$'s cooled over η = 5% of the circumference by a rather heroic electron beam of 1 A/cm^2 and as cold as $\bar{\beta}_{ex} = \bar{\beta}_{ey} = \bar{\beta}_{ez} = 10^{-3}$ (rms energy in each dimension $\cong \frac{1}{4}$ eV) we get

$$\frac{1}{\tau} = 2.1 \text{ sec}^{-1}.$$

Thus, for $\bar{p}$ accumulation electron cooling is too slow at GeV energies.

B. Stochastic Cooling[10]

Because the $\bar{p}$ beam is not a continuum but an ensemble of a finite number of individual particles a broad-band electronic feedback system could be used to cool the beam (Liouville theorem applies only to the mathematical phase-space volume, i.e. a continuum.). A pickup electrode senses the statistical fluctuation of the off-center displacement of the centroid of

a small sample of $\bar{p}$'s. The signal is amplified and applied to reduce the displacement by a kicker located downstream of the sensor. It is important that the sample remains more-or-less identifiable (not totally mixed with other $\bar{p}$'s) between the sensor and the kicker. It is also important that before returning to the sensor the sample should be mixed (at least partially) with other $\bar{p}$'s in the beam so that with each revolution new statistical fluctuations are presented to and reduced by the feedback electronics. The cooling rate is limited by the bandwidth, the noise, and the output power of the electronics. This scheme is particularly advantageous for longitudinal (momentum) cooling because in this degree-of-freedom the pickup signal is a frequency which can be easily cleaned up by a filter.

We give here without derivation the formula for the cooling rate and discuss its features and implications. The formula is

$$\frac{1}{\tau} = \frac{W}{2N}\left[2g - g^2\left(\frac{1}{n}\sum_{k=1}^{n}\frac{f_o}{w_k} + \eta\right)\right]$$

where

W = bandwidth of the feedback system

N = number of particles in beam

g = "gain" = fractional correction in one pass ($\Delta x = -gx$)

$\eta = \dfrac{\text{noise power}}{\text{signal power}}$

f_o = revolution frequency

$n = \dfrac{2W}{f_o}$ = number of Schottky bands

w_k = width of kth Schottky band.

We observe the following.

1. The factor $\frac{W}{N}$ is obvious. If $W = N$ one should be able to sample all the particles individually in 1 second.

2. For given $\frac{W}{N}$ the cooling rate has a maximum of

$$\left(\frac{1}{\tau}\right)_{max} = \frac{W}{2N}\left(\frac{1}{n}\sum_k \frac{f_o}{w_k} + \eta\right)^{-1}$$

at

$$g = \left(\frac{1}{n}\sum_k \frac{f_o}{w_k} + \eta\right)^{-1}$$

3. For perfect mixing each Schottky band has a width equal to the revolution frequency, namely $w_k = f_o$ and

$$\frac{1}{n}\sum_k \frac{f_o}{w_k} = \frac{1}{n}\sum_{k=1}^{n} 1 = \frac{n}{n} = 1$$

Thus, with zero noise ($\eta = 0$) and perfect mixing one should make $g = 1$ and obtain an optimum

$$\left(\frac{1}{\tau}\right)_{max} = \frac{W}{2N}$$

4. Noise and poor mixing limit the useable gain. Generally for longitudinal cooling we have

$$\frac{1}{n}\sum_k \frac{f_o}{w_k} + \eta \cong 10^3 \qquad \text{(order of magnitude)}$$

Hence the optimal "gain" is only $g = 10^{-3}$ and with a bandwidth of 2 GHz for a beam of 10^7 particles we get

$$\frac{1}{\tau} \cong \frac{2\times 10^9\ \text{sec}^{-1}}{2\times 10^7} \times 10^{-3} = \frac{1}{10\ \text{sec}}$$

Again the cooling is slow, but fortunately at such a low "gain" the ampliflier power required is attainable even at $\bar{p}$ energies of several GeV.

The electronic cooling scheme is particularly advantageous for longitudinal cooling because in this dimension the pickup signal is a frequency which can be easily cleaned up by a filter to give a much smaller η

C. Antiproton Accumulation

The spreads in the momentum variables p_x, p_y (transverse) and p_z (longitudinal) of the $\bar{p}$'s produced in the target are given by the production mechanism and hence are not under our control. The spreads in the coordinate variables x, y, and z are however, equal to those of the proton beam bunch incident on the target. To increase the $\bar{p}$ density in the 6-dimensional phase space (x, p_x, y, p_y, z, p_z) one should, therefore, reduce as much as possible the physical dimensions of the incident proton beam bunches. Thus, transversely the proton beam is focused to the smallest possible spot on the target and longitudinally the beam bunches are made as short as possible by rf manipulations. The $\bar{p}$ beam produced at some relatively low energy (a few GeV, for faster cooling) is collected by a large aperture (small f number) field-lens and stored in a large aperture ring so that a large bite of phase space volume is collected by taking large spreads in p_x, p_y and p_z. The phase space density is then increased sometime more than 10 decades, by one or both of the cooling mechanisms. Frequently the spread in p_z is first reduced by debunching the beam (stretching out the spread in z) since the cooling processes proceed faster for beams with smaller momentum spreads. The cooling may take a few seconds and every few seconds a batch of some 10^8 - 10^9 $\bar{p}$'s is produced, collected, cooled and stacked. The stacking rate could be as high as 10^{11} - 10^{12} $\bar{p}$/hour. The final stack of high phase density $\bar{p}$ beam is then accelerated to high energy and injected into the collider to collide with the counter-rotating p beam.

IV. Crossing Geometry, Luminosity and Tune-shift

For this section we shall assume bunched beams. With only minor modifications the discussions can be translated to apply to coasting (continuous) beams.

To minimize the beam-beam interaction the bunches of the two oppositely travelling beams should be kept separated except at the intended colliding points. Thus at the collision points the beams should cross at a small but finite angle θ. The condition that the beam bunches be separated far away from the collision point is

$$\theta > 2\sqrt{\frac{1}{\beta^*}\frac{\varepsilon}{\pi}}$$

where ε = emittance and β^* = low β at the collision point. On the other hand the condition that the beam bunches not be totally separated at the ends while crossing can be written as

$$\theta < \frac{2}{\ell}\sqrt{\left(\beta^* + \frac{\ell^2}{\beta^*}\right)\frac{\varepsilon}{\pi}}$$

where 2ℓ = bunch length. If one takes $\theta = 2\sqrt{2}\ \sqrt{\frac{1}{\beta^*}\frac{\varepsilon}{\pi}}$ to just satisfy the lower limit condition one gets the condition $\ell<\beta^*$. This gives either a lower limit of the low-β value or, for given β^*, an upper limit of the bunch length. This condition is also desirable from the consideration that when $\ell>\beta^*$ the luminosity is reduced because the beam bunches get too wide at the ends; but this is a much softer condition.

This beam separation is very small. For pp the two beams must be further separated by a common dipole far enough apart to travel in separate magnet rings. The common dipole can be placed on either side of the low-β matching quadrupoles. In either case, since the matching

quadrupoles must be located fairly close to the collision point, hence close to the common dipole, to keep the maximum β values in the quadrupoles not excessively large, it is likely that both beams must also be contained in common matching quadrupoles. One must therefore take into account the deflections of the orbits by these quadrupoles and the implication of common quadrupoles on the relative focusing orders (F or D) of the corresponding quadrupoles in the two rings. A common quadrupole has the same focusing actions on $\bar{p}p$ but opposite focusing actions on pp. The simplest arrangement is, thus, to keep these relative focusing orders all the way around the ring, viz. even for the parts of the rings which are totally separated.

For $\bar{p}p$ the two beams can only be separated by an electrostatic field and hence the separation can not be very large. Generally both beams must be contained in the aperture of the same ring and the electrostatic separators serve only to induce opposite betatron oscillations on the two beams. To ensure that the beam bunches are separated at all unwanted crossings the spacing between bunches in each beam must not be much smaller than the betatron wave length. This strongly limits the number of bunches in each beam and hence the attainable luminosity. Or one can make the crossing angle θ very large and the beam bunch length ℓ very small. If

$$\theta = 2M\sqrt{\frac{1}{\beta^*}\frac{\varepsilon}{\pi}} \qquad M \gg 1$$

one must have $\ell < \beta^*/M$. If one cannot make beam bunches so short one again takes a beating on luminosity.

In all these cases one can, of course, have the beam bunches collide head-on and depend on either the common dipoles (for pp) or the electrostatic separators (for $\bar{p}p$) to separate the beams. However, since the bunch

spacing must be greater than twice the distance between the collision point and the location where the beams are separated, it must then be rather large, generally much larger than that when the beams are separated by a crossing angle. With the minimal crossing angle ($M = \sqrt{2}$) and the short bunches ($\ell < \beta^*$) discussed above the luminosity per collision of two beam bunches is reduced only slightly from head-on collision. The total number of bunches per beam can, however, be made much larger, giving a big gain in the total luminosity.

For the example of the SSC we have

$$\ell = 0.61\ m < \beta^* = 2\ m$$

and

$$\theta > 2\sqrt{\frac{1}{\beta^*}\frac{\varepsilon}{\pi}} = 0.03\ mrad$$

Hence a crossing angle of $\theta = 0.05$ mrad is adequate. Since ℓ is much smaller than β^* the luminosity is essentially that of head-on collision and we have for the luminosity per collision of two p beam bunches of $n = 2\times10^{10}$ protons each

$$L = \frac{3}{2}\frac{\gamma n^2}{\beta^* \varepsilon_n} = 2\times10^{25}\ cm^{-2}$$

If the total pp cross-section at this energy is $\sigma \cong 200$ mbarn $= 2\times10^{-25}$ cm^2, the average number of events per collision of two bunches is, then, $\sigma L = 4$ which may be too high for present detectors to handle. Hopefully new and improved detectors will be able to handle more simultaneous events. Assuming a bunch spacing of $s = 40$ m we get a total luminosity of

$$\mathcal{L} = \frac{2\pi R}{s} f_{rev} L = 1.5\times10^{32}\ cm^{-2} sec^{-1}$$

The total number of protons per ring is

$$\frac{2\pi R}{s} n = 5 \times 10^{13}$$

and the beam-beam tune shift per crossing is

$$\xi = \frac{3}{2} r_p \frac{n}{\varepsilon_n} = 0.0015$$

which is believed to be alright.

References

1. J.S. Colonias, "Particle Accelerator Design: Computer Programs". Academic Press, New York and London (1974)

2. J. Moser, Nach. Akad. Wiss. Göttingen, IIA, No. 6, p. 87 (1955)
   W.P. Lysenko, Particle Accelerators, Vol. 5, pp. 1-21 (1973)

3. E. Keil and W. Schnell, CERN Report ISR-TH-RF/69-48 (1969)
   F. Sacherer, Proc. of the 1973 Part. Accel. Conf., San Francisco, IEEE Trans. Nucl. Sci., Vol. NS-20, No. 3, p. 825

4. F. Sacherer, Proc. of the 9th Int. Conf. on High Energy Accel., Stanford, CA (1974) p. 347

5. E. Fischer and K. Zankel, CERN Report CERN-ISR-VA/73-52 (1973)

6. J.D. Bjorken and S. Mtingwa, Particle Accelerators, Vol. 13, pp. 115-143 (1983)

7. S. Kheifets, in "Long Time Prediction in Dynamics", C.W. Horton Jr., L.E. Reichl, V.G. Szebehely, eds. (Wiley, New York, 1983), pp. 397-426
   J.F. Schonfeld, in "Physics of High Energy Particle Accelerators", M. Month, ed. (SLAC Summer School 1982) AIP Conf. Proc., No. 105, pp. 524-650 (1983)

8. D. Dew-Hughes, in "Introduction to Superconducting Materials", D. Dew-Hughes, T. Luhman, eds. Treatise on Mat. Sci. and Tech. Vol. 14, (Academic Press, 1979) pp. 1-46

9. T. Ogino and A.G. Ruggiero, Particle Accelerators, Vol. 10, pp. 197-205 (1980)

10. D. Möhl, G. Petrucci, L. Thorndahl and S. vanderMeer, Physics Reports, Vol. 58C, No. 2, pp. 73-119 (1980)