Fermi National Accelerator Laboratory

FERMILAB-Conf-87/160

A Practical Guide to Modern High Energy Particle Accelerators*

Stephen D. Holmes
Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

October 1987

*Lectures given at the Theoretical Advanced Summer Institute, Santa Fe, New Mexico, July 6-24, 1987

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

A PRACTICAL GUIDE TO MODERN HIGH ENERGY PARTICLE ACCELERATORS

Stephen D. Holmes *
Fermi National Accelerator Laboratory
Batavia, Il. 60510

I. INTRODUCTION

Particle accelerators are the basic experimental tool used by high energy physicists in testing current understanding of the laws of nature. The number of existing high energy facilities in the world numbers in the single digits and it is expected that this situation will remain unchanged in the foreseeable future. The purpose of these lectures is to convey to you, a group of young theoretical physicists, an understanding of how these machines work, and why they look the way they do. The approach taken will be physically intuitive rather than mathematically rigorous. It is hoped that you will develop a feeling as to what determines the scale of eixisting and proposed accelerators, as well as some insight into the fundamental limitations of these machines.

The emphasis will be on the description of **proton circular** accelerators and colliders. Linear accelerators (linacs) will be mentioned only in passing as sources of protons for higher energy rings. Electron accelerators/storage rings and antiproton sources will be discussed only by way of brief descriptions of the features which distinguish them from proton accelerators.

The first lecture will be devoted to explaining the basics of how generic accelerators work. The discussion will focus on descriptions of what sets the overall scale, single particle dynamics and stability, and descriptions of the phase space of the particle beam. In the second lecture we will use what we have learned to design a Superconducting Super Collider (SSC).

II. OVERVIEW

Let us first look at Fermilab as being prototypical of a modern proton accelerator. At Fermilab protons pass through four distinct accelerators on their way to the peak energy of 900 GeV. The series of steps by which the protons are accelerated is shown in Figure 1. The

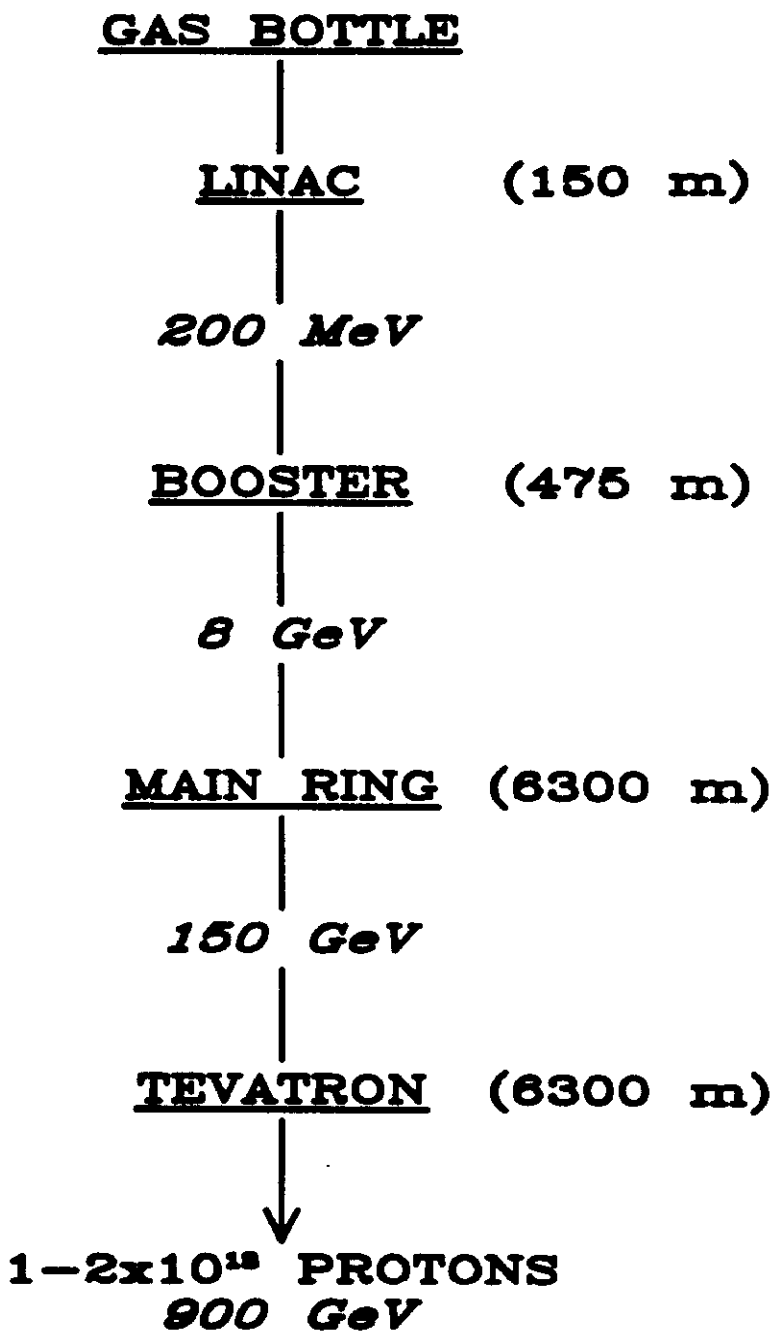

Figure 1. The Fermilab Acceleration String

protons originate in a Hydrogen gas bottle. They are accelerated to a kinetic energy of 200 MeV in a linac of length 150 meters. (For reasons which time does not allow us to discuss, the protons are actually accelerated with two electrons attached to form H^- ions.) Bare protons are then transferred into a circular accelerator of circumference 475 meters, called the Booster, where they attain an energy of 8 GeV. This is followed by a 6300 meter circumference ring called the Main Ring (the old 400 GeV accelerator) where their energy increases to 150 GeV. Finally, the protons are injected into the new superconducting accelerator, the Tevatron, where they reach their ultimate energy of 900 GeV. Between 1 to 2×10^{13} protons can be accelerated at once and the entire acceleration process takes about 60 seconds to complete. (A little arithmetic shows that a single gram of Hydrogen gas is sufficient for operating Fermilab for thousands of years.) Once the protons are at 900 GeV they can be used for a least two purposes: 1)They can be extracted from the Tevatron and sent through beamlines to collide with stationary targets for the benefit of experimenters doing "fixed-target" physics; or 2)They can be made to collide with 900 GeV antiprotons counter-circulating in the Tevatron to be observed by experimenters doing "collider" physics.

You should be starting to wonder at this point why this particular scheme has been concocted. In particular you might ask: "Why are circular accelerators used almost exclusively rather than linear accelerators?"; "How are the protons accelerated?"; and "Why are a series of accelerators used rather than just one?". We discuss below the answers to these questions. A word of warning however--Please keep in mind that the arguments given below are really only appropriate to a discussion of proton accelerators. The phenomonon of synchrotron radiation completely distorts the arguments once one tries to design an electron accelerator or storage ring. A short discussion of electron accelerators will be included at the end of these lectures.

II.1 Why Circular Accelerators?

All accelerators use electric fields to accelerate charged particles. (You remember, of course, from Freshman Physics that a magnetic field cannot change the kinetic energy of a charged particle.) The most straight forward manner in which to accelerate a charged particle is to set up an electric potential difference between two points and let the particle traverse the space between the two points. Current techniques allow us to create average electric fields of about 5×10^6 Volts/meter in proton linear accelerators. So in the case of a proton linac we have an approximate

relationship between the energy of the proton beam we would like to produce and the length of the linac,

$$\text{Energy(MeV)} = 5 \times \text{Length(meters)}. \qquad (1)$$

If we wanted to accelerate protons to 1 TeV, according to (1) we would need a linac 124 miles long! This is clearly an immense engineering task. I do not mean to imply here that linacs are useless, just that they are not a very efficient way to get protons to very high (>1 GeV) energies. Linacs do indeed make excellent sources of low energy protons for injection into higher energy circular accelerators.

The problem with linacs exhibited in equation (1) is that they represent an inefficient use of the available electric fields. If one were able to recirculate protons through the same electric field many times, high energies can be attained without producing electric fields extended over large regions of space. Specifically, if one can produce an electric potential difference, V, between two points in space, and conspire to have a charged particle traverse this region n times, then the energy of the paricle can be made quite large even if V is small,

$$\text{Energy(MeV)} = \text{V(MV)} \times \text{n}. \qquad (2)$$

As an example, in the Fermilab Tevatron a voltage, V, of about 1 MV is provided in a region which extends over about 20 meters in space. By traversing this region 850,000 times the protons are accelerated from 150 GeV up to 1 TeV. You are probably now wondering why the Fermilab Tevatron needs to be 6300 meters in circumference if only 20 meters are used to provide the electric field necessary to accelerate the beam. The reason has to do with the means by which we bring the protons back around to recirculate through the region of electric fields. The recirculation is done using magnetic fields and it is our ability to create high magnetic fields which sets the scale of proton circular accelerators.

Note the contrast between the design bottleneck of the proton linac and the proton circular accelerator. In the linac the problem is constructing the largest average electric field possible in order to keep the physical size (length) of the accelerator as small as possible. In the proton circular accelerator the problem becomes producing the largest possible magnetic field possible for recirculation, again in order to keep the physical size (circumference) small. As you have probably surmised, in practice the circular accelerator becomes the more efficient way of reaching very high energies.

II.2 Confinement and Acceleration in Circular Accelerators

As we said above the recirculation, or confinement, function in circular accelerators is provided by magnetic fields. (The use of electric fields for

confinement is not really technically feasible. The equivalent impulse on a charged particle is produced by either a 10 kGauss magnetic field or a 300 MV/meter electric field, and the 10 kGauss magnetic field is infinitely easier to make.) We know that the motion of a charged particle in a uniform magnetic field is a circle of radius R, where R is given by,

$$R(\text{meters}) = \frac{P(\text{GeV/c})}{.03B(\text{kGauss})} \qquad (3)$$

where P is the momentum of the particle, B is the magnetic field, and we assume the particle has a charge q=e. As a numerical example, for a field of 44 kGauss and a momentum of 1 TeV/c, the radius of curvature becomes 760 meters. This is comparable to the radius of the Fermilab Tevatron, 1000 meters, with the difference accounted for by the regions of the circumference of the accelerator which are not filled with a uniform magnetic field. In the earliest circular accelerators (cyclotrons) particles were accelerated in a region of fixed magnetic field, so that the radius of the orbit of a particle increased with its momentum. All modern high energy proton accelerators are **synchrotrons**. In the synchrotron the magnetic field is increased proportionally to the momentum so that the particle orbits always stay in the same place. This allows us to build accelerators which have a circumference of 3.9 miles (such as Fermilab) without having to cover 850 acres with magnetic field.

Acceleration in a circular accelerator is provided by radio frequency (RF) electric fields. The use of DC fields to accelerate the beam is precluded by Maxwell's Equations. (See if you can guess which one.) Radio frequencies (i.e. those in the 1 MHz to 1 GHz range) are used for a number of reasons. First, it is easier to get high electric fields at high frequencies. And second, very high frequencies (>1 GHz) are generally precluded (in the case of protons) because the inverse relation between the RF wavelength and the charge density produced tends to limit the beam intensity one can produce due to beam instabilities. At Fermilab acceleration is provided by electric fields which oscillate at 53 MHz.

Let me reiterate what I said earlier with regard to the scale of proton accelerators. ***The scales of all proton accelerators in existence today, as well as those contemplated in the remainder of the twentieth century, are set by the maximum magnetic field that we can create to make the beam go around in a circle.***

II.3 Why are Accelerators Cascaded?

In Figure 1 we showed an acceleration process which was completed in steps using different accelerators. You might wonder why we don't save some money and effort by simply injecting the 200 MeV protons coming

out of the Fermilab Linac directly into the Tevatron for acceleration to 900 GeV. We don't do this for two reasons. First, since there is a certain range of magnetic fields over which we feel comfortable operating, we don't like to build accelerators where the ratio p_{out}/p_{in} is too large. And second, the behavior of the phase space of accelerated proton beams results in a much more efficient use of aperture in a cascaded system.

In general it becomes difficult to operate accelerators built with conventional iron core/copper conductor magnets at magnetic fields of less than a few hundred Gauss. The reason is that the residual magnetization, or **remnant fields**, in the iron result in a deterioration in the field quality of the magnet at low fields. Since the highest magnetic field we are able to create in a conventional magnet is somwhere between 15 and 20 kGauss, the result is that conventional accelerators tend to have ratios p_{out}/p_{in} in the range 10-40.

An analogous effect occurs in superconducting magnets although the physical mechanism is different. Eddy currents are induced in the superconductor itself during the process of accelerating the beam and reducing the magnetic fields back to the value correpsonding to the injection momentum. Since the superconductor offers no resistance to these currents they persist for long periods of time, whence the name **persistent currents**, and create contributions to the magnetic field which adversely affect the magnetic field quality at low fields. The size of the effect is related to the physical dimensions of the superconducting filaments. In the Fermilab Tevatron this effect becomes important at fields below about 6 kGauss, while at the SSC (with its smaller filament size) it is expected that the effect will become important below 3 kGauss. Since the maximum field currently obtainable in superconducting accelerator magnets is about 60 kGauss this effectively constrains the ratio of p_{out}/p_{in} to lie in the range 10-20.

Even in the absence of lower limits on allowable magnetic fields we would probably build cascaded accelerator systems in order to make efficient use of available aperture. An important fact of accelerator life is that the largest component of proton accelerator cost is in the magnets, and that the cost of a magnet increases very rapidly with its transverse size. Coupled with this is the fact that the phase space of proton beams behaves in such a manner that during the acceleration process the physical transverse beam size decreases with increasing momentum as,

$$\text{Transverse Size} \propto 1/\sqrt{p} \ .$$

This means, for example, that if we were to try to inject beam into the Fermilab Tevatron at 8.9 GeV/c rather than at 150 GeV/c we would

have to be prepared to accept a beam with four times the transverse physical dimensions. The cost of building an accelerator to go from 8.9 GeV/c to 150 GeV/c would probably be less than the cost of enlarging the Tevatron magnet aperture. (This was a moot point at Fermilab since the Main Ring already existed.)

II.4 Phase Space

We have already started talking about the phase space of the beam implicitly in the preceding section. Beams are made up of collections of particles, each of which is located at a particular point in a 6-dimensional phase space:

$$(x,x';y,y';t,E).$$

In a perfect world all particles would have phase space coordinates (0,0,0,0,0,0), that is all particles would be exactly where we had intended and would have exactly the intended momentum. In our discussion we will assume that the six variables can be treated as three independent sets of two variables. Traditionally (x,x') are the horizontal position and angle, (y,y') are the vertical position and angle, and (t,E) are the azimuthal position around the ring (translated to time units) and the energy relative to their ideal values. The total area in phase space occupied by the beam is called the **emittance, ϵ**. Units of emittance are commonly millimeter-milliradians (mm-mr) for the two transverse planes and electron volt-seconds (eV-sec) for the longitudinal plane. ***The physical dimensions of the beam are related to the emittance and the optical properties of the accelerator.*** An example is shown in Figure 2 where I have shown two beams occupying identical areas in transverse phase space but with different physical extents.

II.5 Acceleration

We stated earlier without proof that the physical beam size in a proton accelerator shrinks as the energy is increased. The physical mechanism by which this happens is shown in Figure 3. A particle enters an RF cavity with some position and angle (x,x'). The angle at which the particle is travelling is a reflection of the ratio of its transverse to longitudinal momentum. After traversing the cavity its longitudinal momentum is increased (as part of the acceleration process) and its transverse momentum remains constant. Thus the angle at which the particle is directed is reduced. In the language of classical mechanics the conjugate variables which define a preserved phase space are (x,P_x), not (x,x'). It is easy to see from the picture that the emittance as we have

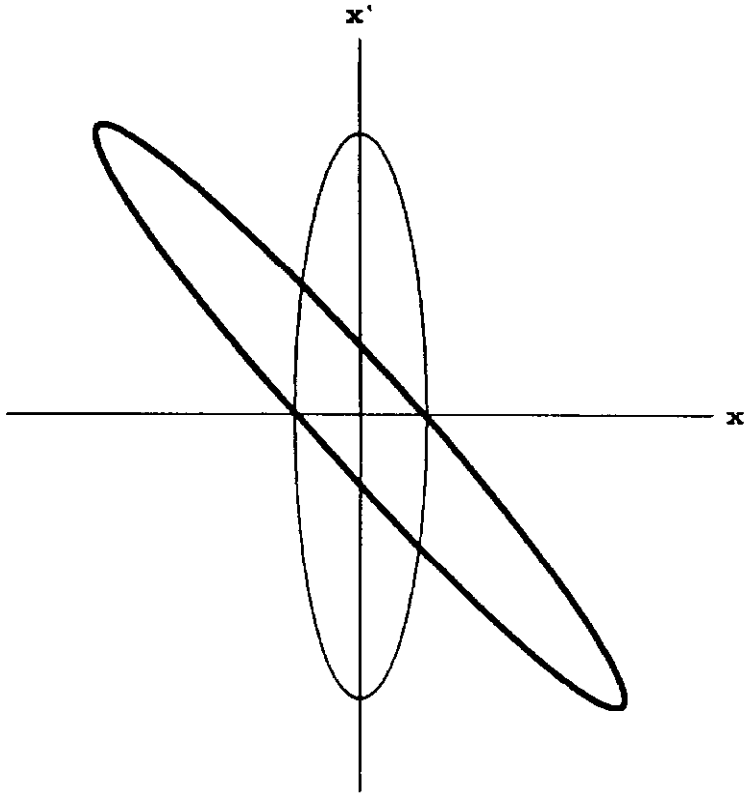

Figure 2. Examples of two contours enclosing equal phase space areas, but with different physical extents.

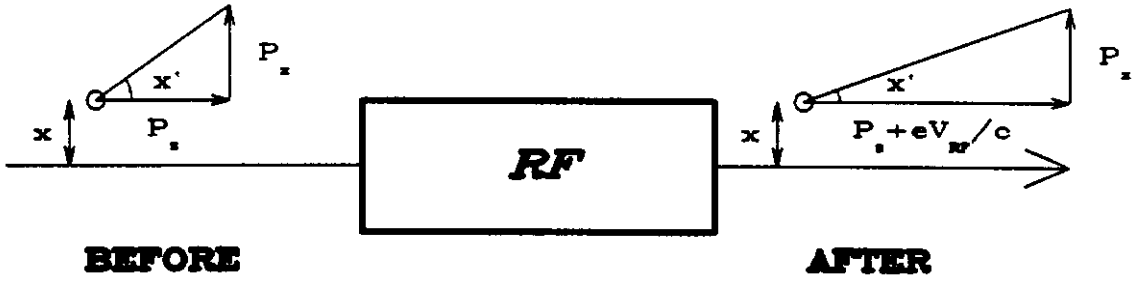

Figure 3. Reduction of the angle of a particle through application of a longitudinal impulse.

defined it is proportional to 1/P. For this reason we often define the **normalized transverse emittance** as

$$\epsilon_N = \epsilon(\beta\gamma)_{rel} \qquad (4)$$

where β and γ are the standard relativistic factors. The normalized emittance is preserved during the acceleration process. Typical values of ϵ_N at existing and planned proton accelerators are 20π at Fermilab and 6π mm-mr at the SSC.

III. TRANSVERSE PHASE SPACE

All modern high energy circular accelerators are **alternating gradient synchrotrons.** This means they use magnetic **dipole** fields for confinement and **quadrupole** fields for focusing. (A dipole field is uniform in space while a quadrupole field varies linearly with displacement from a particular axis.) The alternating gradient principle was invented in 1952 by Courant, Livingston, and Snyder[1], and independently by Christofilos[2]. They showed that stability could be obtained in <u>both</u> transverse planes simultaneously with linear focusing by alternating focusing and defocusing elements. Prior to this time very weak, constant gradient focusing was used to provide stability in accelerators. The reason the invention of the alternating gradient synchrotron was so important was that it allowed one to reduce by orders of magnitude the physical dimensions associated with a given emittance. The reason that it had not been thought of before was presumably because of the fact that magnetic quadrupoles (B_x=B'y, B_y=B'x) do not focus in both transverse planes simultaneously.

We will be discussing for the next few minutes the formalism which has been used for the last thirty five years to describe the phase space and optical characteristics of alternating gradient synchrotrons[3]. The culmination of this discussion will be a definition of the prescription for relating emittances to physical beam sizes in a synchrotron.

III.1 Single Particle Dynamics

We begin by examining the motion of individual particles under the influence of the magnetic fields, both dipole and quadrupole, from which a synchrotron is constructed. We need a coordinate system in order to describe the position of a beam particle at any particular time. The coordinate system we are going to use is shown in Figure 4.

We start out by defining a **reference orbit.** The reference orbit is simply defined as the trajectory which we intend an ideal particle to follow when we design the accelerator. In general the reference orbit is

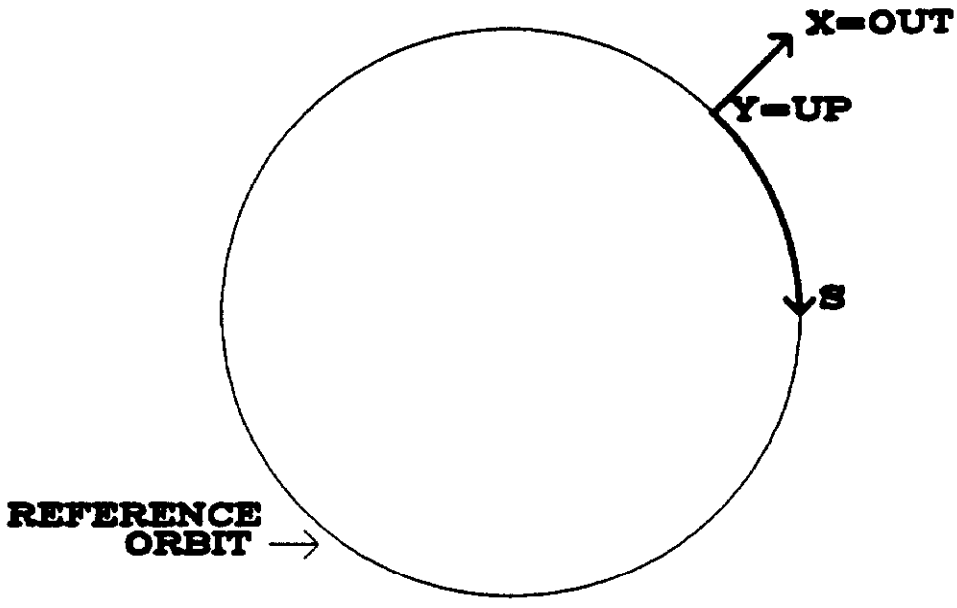

Figure 4. Coordinate system as viewed from above.

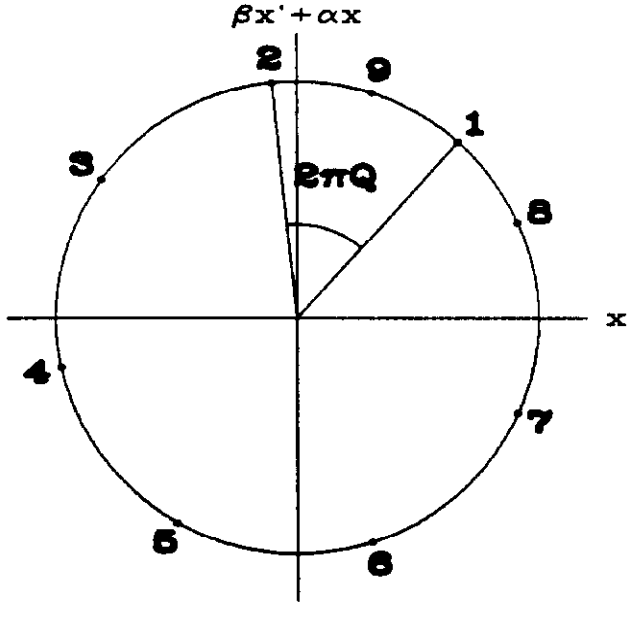

Figure 5. The position and angle of a particle as seen on consecutive turns at a fixed azimuth, s.

made up of arcs of circles in the dipole magnets and straight line segments elsewhere. The reference orbit closes on itself and goes directly through the centers of the quadrupole magnets. To simplify the discussion we will assume that the reference orbit lies in a horizontal plane, that there is no coupling between the horizontal and vertical motion, and we will not consider particles whose momenta deviate in any way from the design momentum. The coordinate system used consists of two transverse coordinates, x and y, and a longitudinal coordinate, s. The convention will be that x is the horizontal and y is the vertical displacement relative to the reference orbit. We define a useful quantity called the **magnetic rigidity**, $(B\rho)$, which is the momentum of the particle (with q=e) in units of kGauss-meters: $(B\rho)(kG\text{-}m)=33.3p(GeV/c)$.

The equations of motion can be written using the azimuthal coordinate, s, as the independent variable:

$$\frac{d^2x}{ds^2} + \left(\frac{1}{\rho(s)^2} - K(s)\right)x = 0 \qquad (5a)$$

$$\frac{d^2y}{ds^2} + K(s)y = 0 \qquad (5b)$$

where $\rho(s)$ is the local radius of curvature of the reference orbit ($\rho=0$ except in dipole magnets), and K(s) is related to the local quadrupole strength,

$$K(s) = -\frac{1}{(B\rho)}\frac{dB}{dx}y \ . \qquad (6)$$

Equations (5a) and (5b) are of the general form,

$$\frac{d^2x}{ds^2} + k(s)x = 0 \qquad (7)$$

where k(s) is periodic with the circumference, L, of the accelerator: k(s+L)=k(s). Equation (7) is just the equation of motion of a harmonic oscillator with a periodic, time varying spring constant. You should not be surprised then to learn that the solution to (7) may be written as a wave with time varying amplitude and wavelength,

$$x(s) = A\sqrt{\beta(s)}\cos(\phi(s)-\phi_0) \ . \qquad (8)$$

A and ϕ_0 are the two arbitrary constants associated with the second order equation (7). Two words of warning at this point: 1)β(s) is <u>not</u> v/c. It is however a single valued function of the azimuthal position, s; and 2)ϕ(s) is <u>not</u> a single valued function of s, i.e. $\phi(s+L)\neq\phi(s)$. In fact ϕ is related to β by

$$\frac{d\phi}{ds} = \frac{1}{\beta} \ . \qquad (9)$$

The particle motion described by equation (8) is called **betatron motion.**

Before we talk about how one finds $\beta(s)$ and $\phi(s)$ for a particular arrangement of magnets in an accelerator I would like to look at the behavior of x and x' as a beam particle circulates in an accelerator. The phase space motion of a particular particle in one of the transverse planes is given by,

$$\begin{cases} x = A\sqrt{\beta}\cos(\phi-\phi_0) \\ x' = \frac{dx}{ds} = -\frac{\alpha x}{\beta} - \frac{A}{\sqrt{\beta}}\sin(\phi-\phi_0) \end{cases} \tag{10}$$

where we define α as $\alpha=-1/2(d\beta/ds)$. One thing we can see right away from (10) is that there is a constant of the motion which is analogous to the total energy in a simple hormonic oscillator,

$$\frac{x^2 + (\beta x'+\alpha x)^2}{\beta} = A^2 \ . \tag{11}$$

There is also a quantity analogous to the frequency of a simple harmonic oscillator called the **tune.** The tune is usually designated either **Q** or ν and is defined as,

$$Q = \frac{\phi(s+L) - \phi(s)}{2\pi} \ . \tag{12}$$

For Q≠integer the transverse position of a particle is not the same on subsequent revolutions of the accelerator. In Figure 5 I show the phase space position of a beam particle at a particular azimuthal location in an accelerator with a non-integer tune for five subsequent circuits of the ring. Note that the points all lie on a circle of radius-squared βA^2. The tune is one of the most important parameters describing an accelerator. We will talk a little later about how one goes about choosing a tune when designing an accelerator.

III.2 Lattice Functions and the Matrix Formulation

The quantities β, α, and ϕ are called **lattice functions.** They are all functions of the azimuthal position, s, and specify the motion of individual particles as they travel around the ring according to equation (10). As I implied in the discussion above, the lattice functions are a reflection of the particular way dipole and quadrupole magnets are arranged around the circumference of the accelerator. I would like to describe to you here a means of calculating the lattice functions resulting from an arbitrary arrangement of magnets. The easiest way to do this is to use a transfer matrix formalism to represent the transport of a particle between two points in the accelerator. The matrix formulation takes advantage of the

fact that $\rho(s)$ and K(s) are piecewise constant in real accelerators, and is only possible because the equations of motion, (5), are linear.

Let us start by looking at the motion of a particle as it traverses a quadrupole. The quadrupole has a field,

$$\begin{cases} B_x = B'y \\ B_y = B'x \end{cases}$$

with B' constant over the length of the quadrupole, l. Equation (5) can be solved over the region of the quadrupole using the position and angle of the particle at the entrance of the magnet to specify the two arbitrary constants. The solution gives the position and angle at the exit of the magnet as linear combinations of the position and angle at the entrance:

$$\begin{bmatrix} x \\ x' \end{bmatrix}_1 = \begin{bmatrix} \cos(\sqrt{k}l) & \frac{1}{\sqrt{k}}\sin(\sqrt{k}l) \\ -\sqrt{k}\sin(\sqrt{k}l) & \cos(\sqrt{k}l) \end{bmatrix} \begin{bmatrix} x \\ x' \end{bmatrix}_0 \quad (13)$$

where $k=B'/(B\rho)$ and $B'>0$. A similar expression is obtained for the vertical motion except the trigonometric functions are replace by hyperbolic functions. Several properties of the transfer matrix shown in (13) are worth noting: 1) The determinant of the matrix is 1. This is what results in the existence of the invariant quantity A in (11); 2) One plane is focusing (the one with cos and sin) and the other is defocusing (the one with cosh and sinh). Which is which depends upon the sign of B'; 3) Dipoles and empty spaces also have transfer matrix representations; and 4) In the limit $l\to 0$ with kl finite the transfer matrix becomes,

$$M = \begin{pmatrix} 1 & 0 \\ -kl & 1 \end{pmatrix} .$$

You may recognize this from optics as the transfer matrix of a lens with focal length,

$$f = \frac{1}{kl} = \frac{(B\rho)}{B'l} .$$

To make the connection between transfer matrices and lattice functions we go back to equation (10) and complete the same exercise of writing the position and angle at a point s_2 in terms of the position and angle at s_1 thus,

$$\begin{pmatrix} x \\ x' \end{pmatrix}_2 = M(1,2) \begin{pmatrix} x \\ x' \end{pmatrix}_1 . \quad (14)$$

The result is,

$$M(1,2) = \begin{bmatrix} \sqrt{\frac{\beta_2}{\beta_1}}(\cos\Delta\phi + \alpha_1\sin\Delta\phi) & \sqrt{\beta_1\beta_2}\sin\Delta\phi \\ -\frac{(1+\alpha_1\alpha_2)\sin\Delta\phi + (\alpha_1-\alpha_2)\cos\Delta\phi}{\sqrt{\beta_1\beta_2}} & \sqrt{\frac{\beta_1}{\beta_2}}(\cos\Delta\phi - \alpha_2\sin\Delta\phi) \end{bmatrix} \quad (15)$$

where $\Delta\phi$ is the phase advance between locations s_1 and s_2. At first glance it does not appear that we have made much progress. The transfer matrix written in terms of the lattice functions contains five unknowns. What it does show us however, is that if we knew the lattice functions at the point s_1 and the transfer matrix between s_1 and s_2 we could use (15) to calculate the lattice functions at s_2. The way we proceed now is to write down equation (15) for the special case $s_2=s_1+L$. In this case we know that $\beta_1=\beta_2$, $\alpha_1=\alpha_2$, and $\Delta\phi=2\pi Q$. So,

$$M(s,s+L) = \begin{pmatrix} \cos 2\pi Q+\alpha\sin 2\pi Q & \beta\sin 2\pi Q \\ -\frac{(1+\alpha^2)\sin 2\pi Q}{\beta} & \cos 2\pi Q-\alpha\sin 2\pi Q \end{pmatrix} \qquad (16)$$

We know have a complete prescription for calculating the lattice functions everywhere assuming we know the arrangement of dipole and quadrupole magnets in the accelerator. First we calculate the single turn transfer matrix at some point, s, by multiplying together the transfer matrices (such as given in (13)) of all the individual components of the accelerator. We then make the identification (16) to solve for $\beta(s)$, $\alpha(s)$, and the tune, Q. We can then use the individual transfer matrices from the point s to any other position in the ring to solve for the lattice functions everywhere! This process is relatively efficient when programmed on a computer and is the method by which virtually all accelerator lattice calculating programs work. Remember that since the transfer matrices in the horizontal and vertical planes are generally not the same, this process must be carried out independently in both planes and will lead to different lattice functions in the two planes.

I leave you with one word of caution before we move on to our next topic. Any random arrangement of quadrupoles and dipoles in an accelerator will <u>not</u> generally result in the existence of stable orbits. If there is no stability in a particular arrangement the symptom is a single turn transfer matrix with a trace either greater than 2 or less than -2. We can see from (16) that if the absolute value of the trace is greater than 2, then the tune must be imaginary and the subsequent application of M(s,s+L) over many revolutions will lead to unbounded motion.

III.3 Emittance

So far I have shown you that the motion of individual particles is given by equation (10) and have told you how to calculate the lattice functions referred to in (10). In general the particle beam will be made up of a lot of particles with different A's and ϕ_0's. For a particular value of A, the motion of a particle in phase-space as viewed over many circuits of

the accelerator draws out the ellipse defined by equation (10) and shown in Figure 6. We will define the **transverse emittance, ϵ, as the area of the ellipse which encompasses 95% of the beam particles.** One can show with a little work that for a Gaussian distribution of beam particles the root-mean-square beam size is then related to the emittance and the lattice function, β by,

$$\sigma_{x\beta} = \sqrt{\frac{\epsilon\beta}{6\pi}} = \sqrt{\frac{\epsilon_N\beta}{6\pi(\beta\gamma)_{rel}}} \qquad (17)$$

Equation (17) displays explicitly the promised relationship between the physical beam size, the emittance, and optical properties of the accelerator. Note that since β is a function of azimuthal location around the accelerator ring, the beam size varies around the ring while the emittance is invariant. While we said that equation (10) describes the motion of individual particles circulating in an accelerator you might be wondering whether the beam as a whole undergoes the same motion. In general the answer is no--while the individual particle are moving around in the beam, the beam as a whole remains in the same place and has the same shape when viewed at a particular place over many revolutions.

IV. LONGITUDINAL PHASE SPACE

In Section III we described the motion of beam particles which were displaced from the reference orbit by an amount (x,x';y,y') and showed that in a well-designed accelerator the motion of these particle was stable and bounded. I would now like to describe the motion in the longitudinal phase space, (t,E). I know that in the real world the particles making up the beam will not all have exactly the same momenta (although the spread in momenta is generally smaller than 1%). So I would like to answer the question "What happens if a particle does not have exactly the momentum it needs to follow the reference orbit in the presence of the existing magnetic fields?"

I told you earlier that the accelerating electric fields in modern accelerators are provided by Radio Frequency (RF) systems. This means that there is an electric field which looks like,

$$V_{RF} = V\sin\omega t$$

where ω is the frequency at which the RF system operates. The phase space coordinate t is thus the time at which a particle traverses the RF cavity. In what follows we will assume that the time required for a beam particle to traverse the RF cavity is very small so that we can treat the

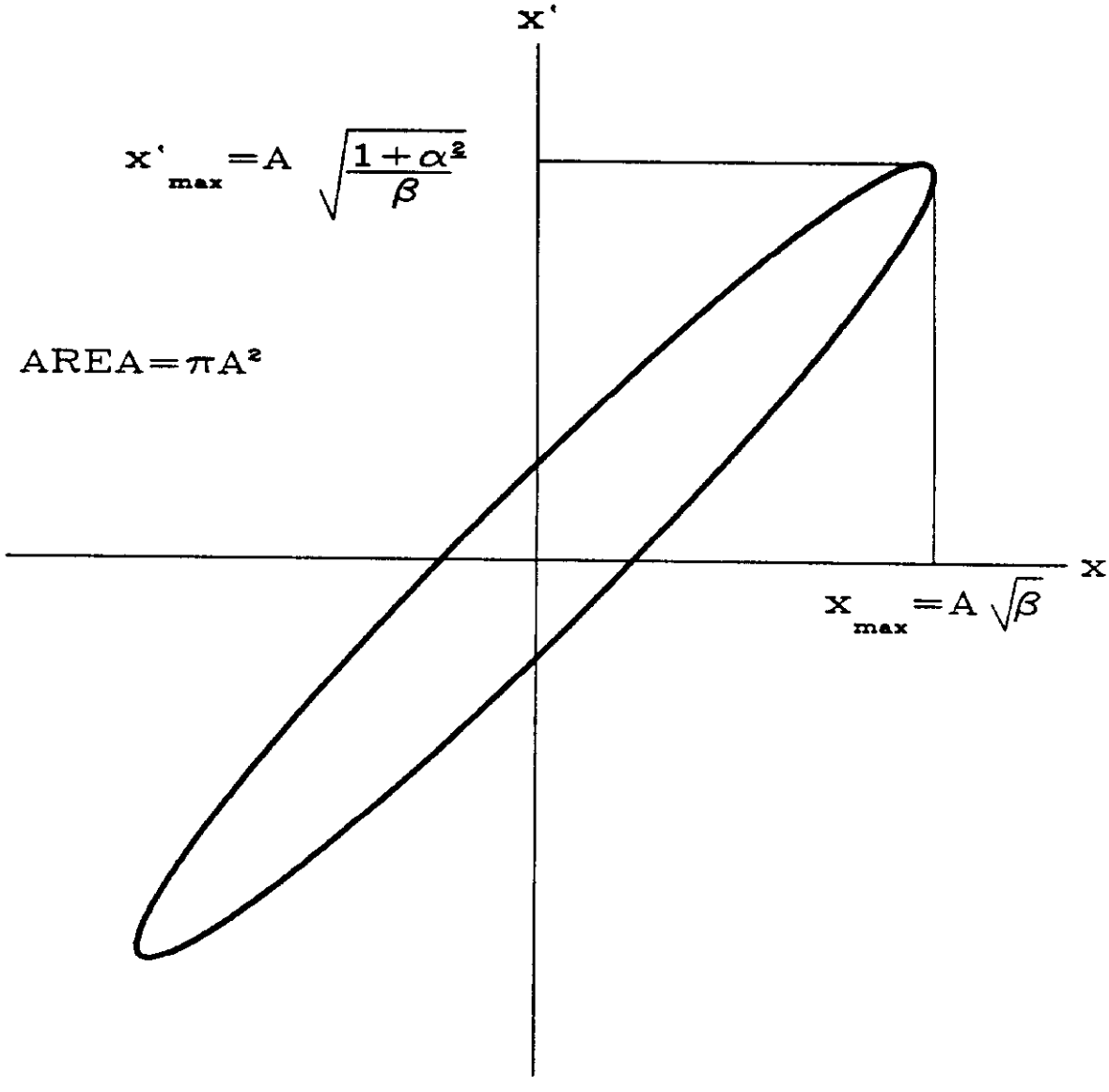

Figure 6. Phase space trajectory for equation (10)

energy gain as being instantaneous. This assumption is generally valid in all existing circular accelerators. Let me define a few useful quantities at this point:

1. T is the **revolution period** of the accelerator. Since the circumference of the accelerator is fixed, T will vary as the beam is accelerated to the extent that the velocity of the particle changes with momentum.

2. h is called the **harmonic number** of the RF system. It is equal to $\omega T/2\pi$ and must equal an integer.

3. P_s is the **synchronous momentum.** It is the momentum which gives a revolution period, T, satisfying $\omega T/2\pi$=integer for the magnetic field present in the accelerator. It is usually also equal to the design momentum of the ring and a particle with the synchronous momentum will follow the reference orbit.

4. ϕ_s is the **synchronous phase.** It is the angle for which $V\sin\phi_s$ is equal to the desired energy gain/turn.

5. η is the **frequency dispersion.** It relates the change in revolution period to the change in momentum for a **fixed magnetic field:**

$$\eta = \frac{(\Delta T/T)}{(\Delta p/p)} \ . \qquad (18)$$

η is a function both of the velocity of the beam particles and of the optics of the accelerator, i.e. it can change as the beam is accelerated. η can also be either greater than, less than, or equal to zero. We will describe how to calculate η later.

Before writing down and solving the longitudinal equations of motion I would like to take you through a qualitative description of the longitudinal motion of the beam. Figure 7 shows the variation of the RF voltage with time and two locations in time where the energy gain/turn is the desired value. Let us ask what happens to a slightly off-energy particle situated at either these locations for the case $\eta>0$. Near the point labelled 1, we see that if the energy is larger than desired it takes longer (since $\eta>0$) than desired for the particle to come back around to the RF station. As a result the particle picks up more energy than it should and the discrepency between its actual and desired momentum becomes greater.

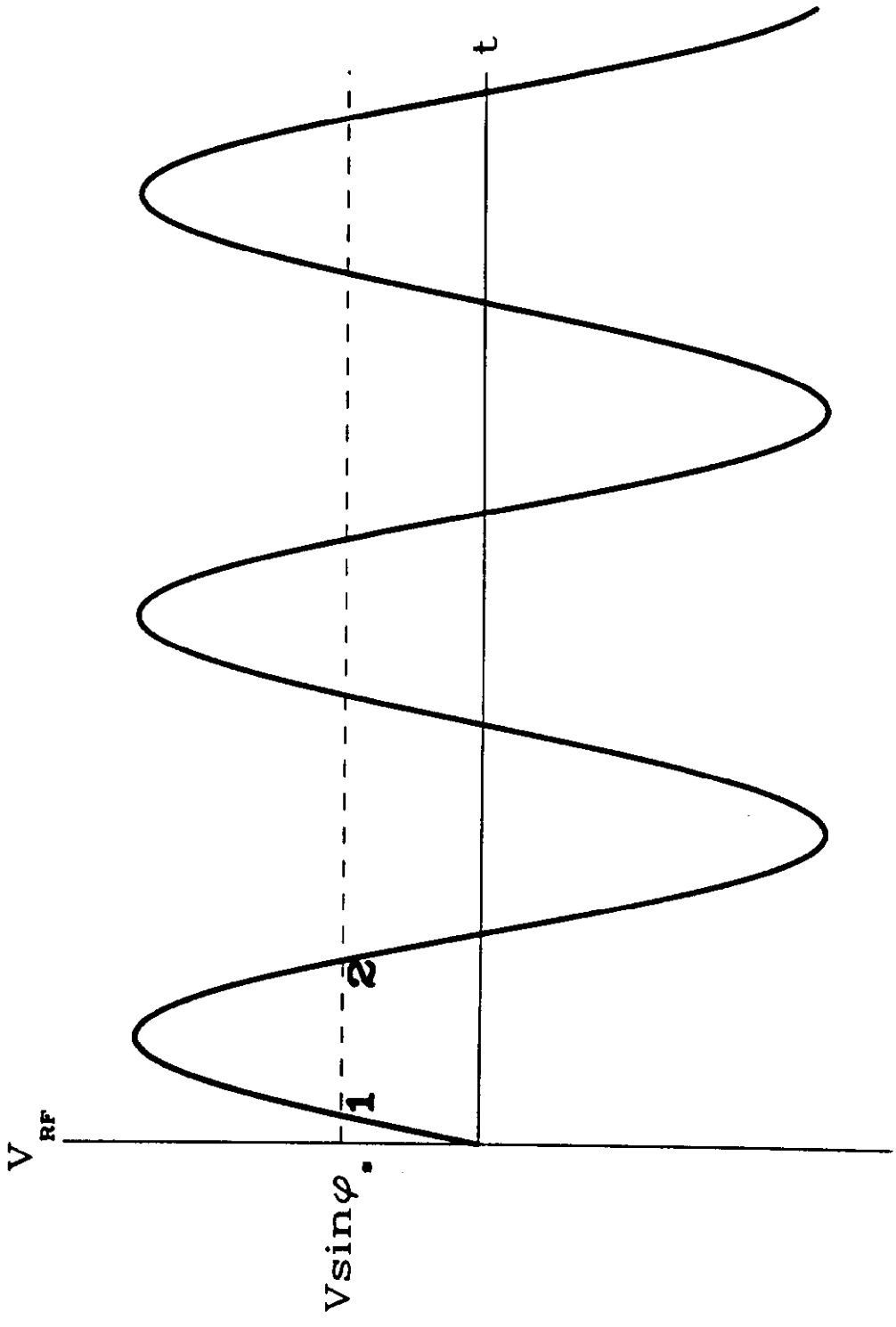

Figure 7. The RF waveform.

Near the point labelled 2 the opposite happens. The higher energy makes the particle again arrive late, but in this case the energy gain is less than desired so the initial energy error is partially corrected. So particles at point 1 are unstable, and particles at point 2 are potentially stable with respect to small perturbations. You can easily show that the situation is reversed if η is less than zero.

IV.1 Single Particle Dynamics

I can quantify the above argument by writing down and solving the equations of motion. Before doing so I will switch from (t,E) to a more convenient set of variables (ϕ,Δ), where $\phi=\omega t$ measures the phase of the RF wave at the arrival time of the particle and $\Delta=(p-p_s)/p_s$ is the relative difference between the actual momentum and the synchronous momentum. The equations of motion in the longitudinal plane are then,

$$\begin{cases} \dfrac{d\Delta}{dn} = \dfrac{eV}{p_s\beta c}(\sin\phi - \sin\phi_s) \\ \dfrac{d\phi}{dn} = 2\pi h\eta\Delta \end{cases} \tag{19}$$

where n is the revolution number. The top equation reflects the fact that the momentum error changes if the particle enters the cavity when the phase of the RF wave is not equal to the synchronous phase, while the lower equation reflects the change in the RF phase at the particle's arrival time due to the correlation between momentum and revolution period.

The equations of motion are not linear as they were in the transverse case (5). This means that the motion will not be bounded over the entire phase space. It also means that we are not able to write down a general analytic solution to (19). However, there is still a wealth of information buried in (19).

IV.2 The Separatrix

It is easy to show that there is a constant of the motion described by (19). You can work it out yourself, and what you will find is that

$$\begin{cases} H = \Delta^2 + \dfrac{eV}{p_s\beta c\pi h\eta}(\cos\phi+\phi\sin\phi_s) \\ \dfrac{dH}{dt} = 0 \ . \end{cases} \tag{20}$$

The equations of motion, (19), also possess two fixed points. These points, (Δ,ϕ) equals $(0,\phi_s)$ and $(0,\pi-\phi_s)$, are the two points we looked at on Figure 7. For $\eta>0$ the first fixed point is unstable and the second is stable. The roles are interchanged if η is negative. Associated with the

unstable fixed point is a contour of constant H, called the **separatrix.** Phase space trajectories lying outside the area enclosed by the separatrix are unbounded. Thus the separatrix encompasses the stable area in phase space. For $\eta>0$ the separatrix is defined by,

$$H = \Delta^2 + \frac{eV}{p_s\beta c\pi h\eta}(\cos\phi+\phi\sin\phi_s) = \frac{eV}{p_s\beta c\pi h\eta}(\cos\phi_s+\phi_s\sin\phi_s) \quad (21)$$

As you might guess the area enclosed by the separatrix depends on the RF voltage, the frequency dispersion, the harmonic number, and the synchronous phase. Figure 8 shows examples of two separatrices with all parameters from equation (21) identical with the exception of ϕ_s. As you can see, changing the synchronous phase from 0° to 34° substantially reduces the stable area available. In Figure 9 you can see examples of phase space trajectories generated by numerical iteration of (19) for the same parameters as in Figure 8. In general the choice of ϕ_s in a real accelerator represents a compromise between the desire to accelerate the beam quickly (large ϕ_s), and the desire to proved a large stable area to accomodate the beam (small ϕ_s).

IV.3 Synchrotron Oscillations

For small deviations around the stable fixed point the equations (19) can be linearized to give harmonic motion. If we approximate,

$$\sin\phi-\sin\phi_s \simeq \cos\phi_s(\phi-\phi_s)$$

then the equations of motion can be combined into a single second order differential equation thus,

$$\frac{d^2\Delta}{dn^2} = \left[\frac{2\pi h\eta eV\cos\phi_s}{p_s\beta c}\right]\Delta \ . \quad (22)$$

Equation (22) is just the simple harmonic oscillator equation. (Remember we are dealing with the case $\eta>0$, $\cos\phi_s<0$.) We can read off the oscillation frequency, converting from revolution units to time units,

$$(\Omega_s T)^2 = -\frac{2\pi h\eta eV\cos\phi_s}{p_s\beta c} \ . \quad (23)$$

For small displacements from the stable fixed point particles undergo simple harmonic motion with a frequency, Ω_s, which is called the **synchrotron frequency.** Figure 10 is a picture of particle bunches undergoing synchrotron motion in the Fermilab 8 GeV Booster. Time is progessing vertically in the picture and the history of three different bunches is shown. The time covered by the picture is about 1 millisecond.

IV.4 Emittance

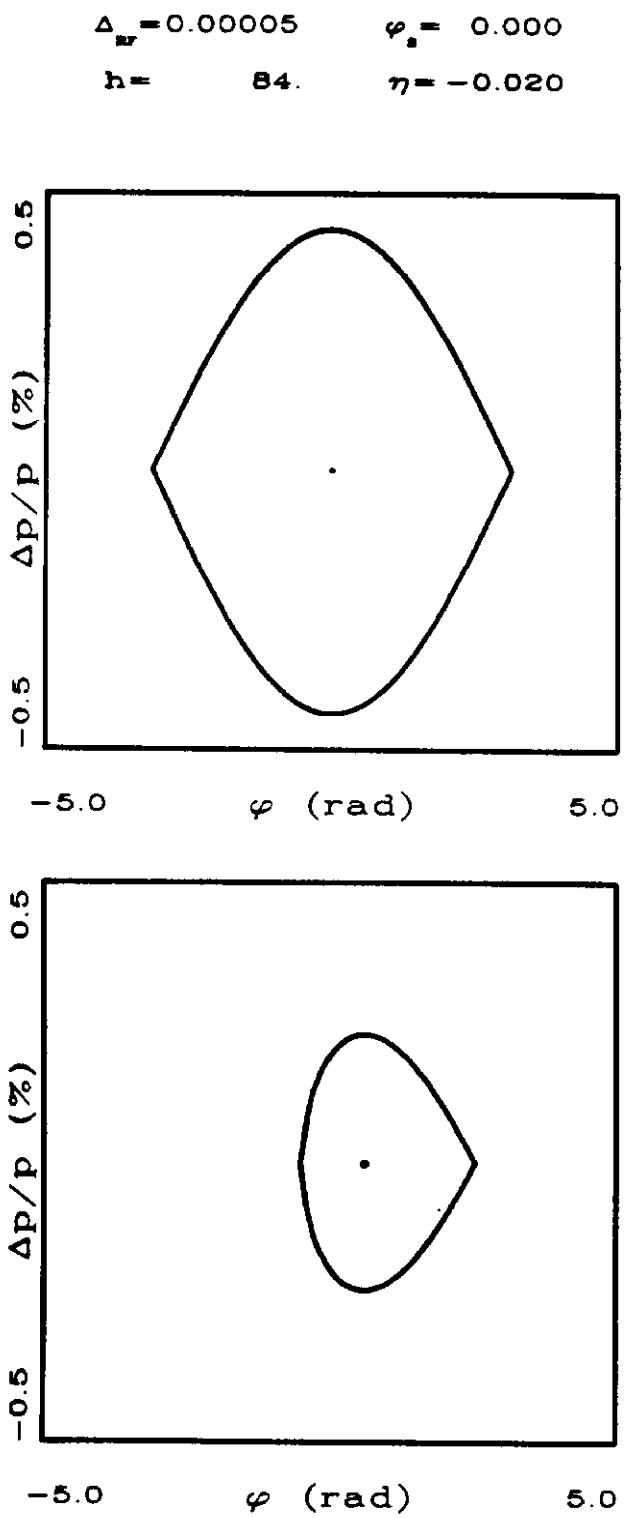

Figure 8. Separatrices for φ_s=0° (top) and 34° (bottom).

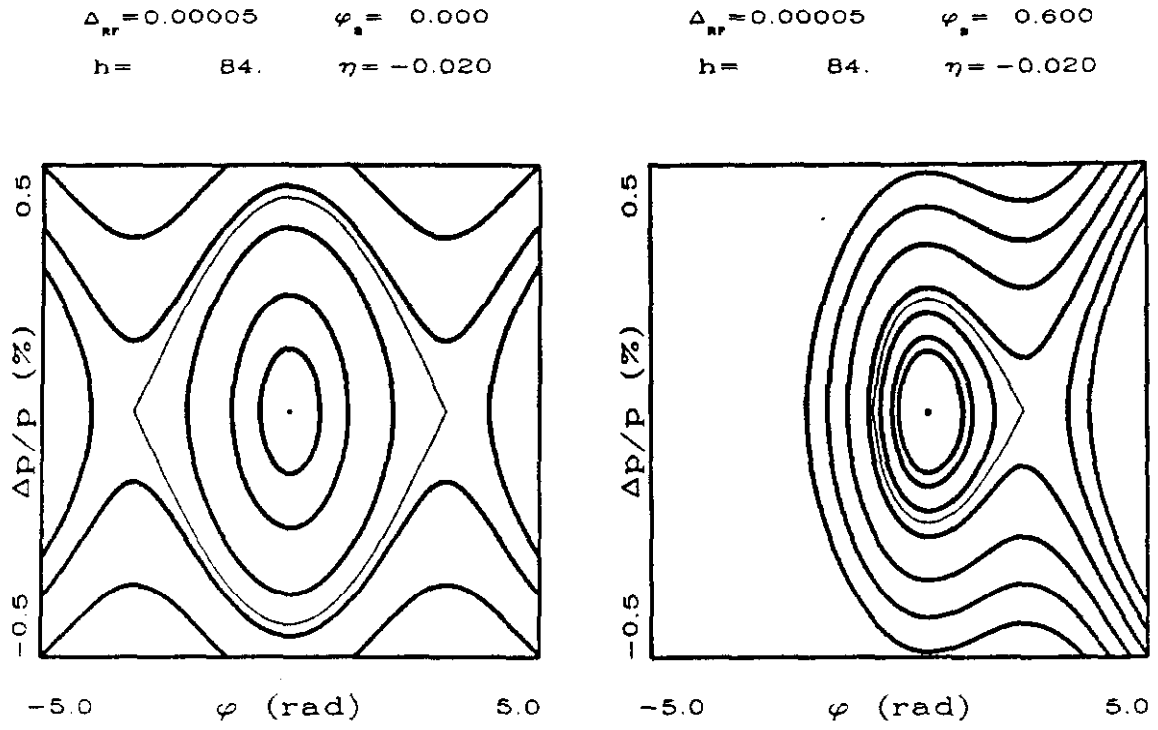

Figure 9. Longitudinal phase space trajectories obtained by iterating (19). The conditions are the same as in Figure 8.

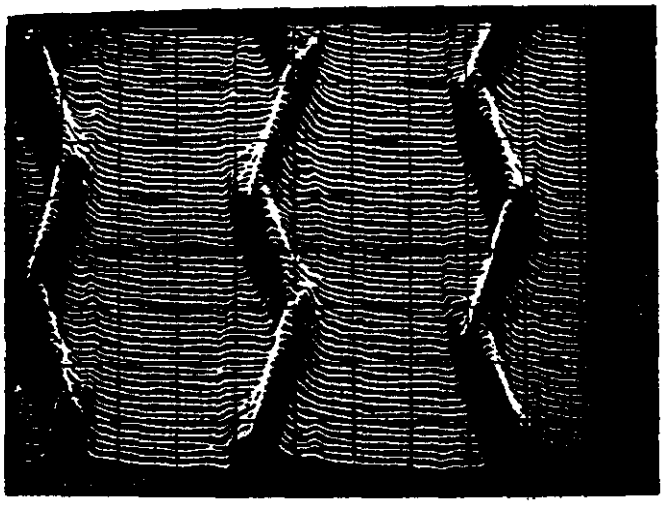

Figure 10. Bunches undergoing synchrotron motion.

The beam in an accelerator is localized in the stable areas of the longitudinal phase space. There is a stable region associated with each stable fixed point, and since the RF voltage is a periodic function the number of such stable regions around the azimuth of the accelerator is just equal to the harmonic number, h. The beam particles contained in any one of these stable regions are referred to as a **bunch.** The number of bunches in an accelerator can thus be any number up to h. (Sometimes some of the stable regions are left empty).

The area of longitudinal phase space occupied by a bunch is called the **longitudinal emittance,** ϵ_L. It is defined in an analogous manner to the transverse emittance: You draw the contour (21) which encompasses 95% of the beam particles and the area enclosed by this contour is ϵ_L. Typical values of the longitudinal emittance we deal with at Fermilab are 0.3 eV-sec. The longitudinal emittance is preserved during acceleration.

IV.5 Transition

I would like to conclude our discussion of longitudinal phase space by telling you what I meant by my earlier comment that the quantity η was not necessarily constant during the acceleration process. As you remember from (18) η relates changes in revolution period to changes in momentum for a fixed magnetic field. The revolution period is related to the total path length, L, around the accelerator and the velocity of the beam particle by $T=L/\beta c$. Two distinct effects contribute to the change in revolution period with momentum--the change in path length with momentum, and the change in velocity with momentum:

$$\frac{dT}{T} = \frac{dL}{L} - \frac{d\beta}{\beta} . \qquad (24)$$

The change in path length around the accelerator is simply related to the fact that as the momentum of a particle increases it travels on a circle of larger radius. The relation is parameterized by

$$\frac{dL}{L} = \frac{1}{\gamma_t^2} \frac{dp}{p} \qquad (25)$$

where γ_t is called the **transition gamma.** The transition gamma is so-named for a reason which will become obvious shortly. It is a property of the optics of the accelerator and does not depend on the energy of the beam. I will tell you how to calculate γ_t at the end of this lecture. You all learned how to calculate the increase in the velocity of a particle with momentum many years ago:

$$\frac{d\beta}{\beta} = \frac{1}{\gamma^2} \frac{dp}{p} .$$

So we can substitute into equation (24) to get our result,

$$\frac{dT}{T} = \left(\frac{1}{\gamma_t^2} - \frac{1}{\gamma^2}\right) \frac{dp}{p}$$

or,

$$\eta = \frac{1}{\gamma_t^2} - \frac{1}{\gamma^2} \quad . \tag{26}$$

The momentum dependence of η is now displayed explicitly and the designation 'transition gamma' becomes apparent: For γ less than γ_t η is negative, while for γ greater than γ_t η is positive. In many proton accelerators, including the Booster and Main Ring at Fermilab, the beam passes through transition during the acceleration process. There are two important consequences of passing through transition. First, the stable phase of the RF wave shifts from ϕ_s to π-ϕ_s. Since the beam cannot move discontinuously in time, it is necessary to change the phase of the RF voltage as the beam passes through transition. The second effect is that if you were to look at the invariant of the longitudinal motion, (20), you would find that right at transition the momentum spread of the beam becomes infinite while the bunch length (i.e. the extent over ϕ) goes to zero. That this cannot be allowed to happen forces one to make sure that the passage through transition is performed fast enough to be non-adiabatic. In this case transition is not really a practical problem.

V. MACHINE ERRORS

So far we have established a reference orbit in our accelerator, with individual beam particles executing stable oscillations in both the transverse and longitudinal planes around that reference orbit. What I would like to look at now are the consequences of building a non-perfect accelerator. We will examine three types of errors:

1. All dipole magnets do not have exactly the same values of their magnetic fields.

2. All quadrupoles are not exactly aligned with their centers on the reference orbit.

3. All quadrupoles do not have exactly the same magnetic field gradients.

As we shall see below, 1. & 2. lead to a distorted **closed orbit**, while 3. leads to **tune shifts** and **stop-bands.**

V.1 The Closed Orbit

We will suppose that we have a distribution of dipole fields around our ring, $\Delta B(s)$, which were not intended to be present when the accelerator was designed. This can happen because of either of the first two effects listed above. Our equation of motion, (7), then becomes,

$$\frac{d^2x}{ds^2} + k(s)x = \frac{\Delta B(s)}{(B\rho)} \ .$$

This is an inhomogenous differential equation which has a solution $x(s)=x_\beta(s)+x_c(s)$, where x_β is a solution to the homogenous equation as described earlier, and x_c is the particular solution to the inhomogenous equation with $x_c(s+L)=x_c(s)$. The solution to the inhomogenous equation is,

$$x_c(s) = \beta^2(s)\sum_n \frac{Q^2}{Q^2 - n^2} b_n e^{in\phi(s)}$$

where,

$$\frac{\Delta B(s)}{(B\rho)} = \sum_n b_n e^{in\phi(s)} \ .$$

X_c is called the closed orbit. It differs from the reference orbit if $\Delta B(s)$ differs from zero for any value of s. This means that in a real accelerator with real construction errors, the transverse motion described by equations (10) is really measured with respect to the closed orbit, not the reference orbit. The closed orbit formula can be written in a more convenient alternate form if the field errors are approximated by a discrete set of error elements,

$$\frac{\Delta B(s)}{(B\rho)} = \sum_i \frac{(\Delta Bl)}{(B\rho)}_i \delta(s-s_i) \ .$$

The index i refers to a contributing error field at the position s_i. In this case the expression for the closed orbit can be written,

$$x_c(s) = \frac{\sqrt{\beta(s)}}{2\sin\pi Q} \sum_i \frac{(\Delta Bl)}{(B\rho)} i\sqrt{\beta_i}\cos(|\phi(s)-\phi_i|-\pi Q) \ . \qquad (27)$$

Both forms of x_c show that **values of the tune, Q, near an integer need to be avoided.** Equation (27) is the form in which accelerator designers tend to think of the closed orbit because it explicitly shows the effect of localized errors, and because it is easily programable. Closed orbit errors are inevitable in a real machine. They are typically in the range 1-10 mm and are controlled by providing adjustable correction dipoles. Figure 11 shows a typical closed orbit measured in the Fermilab Booster.

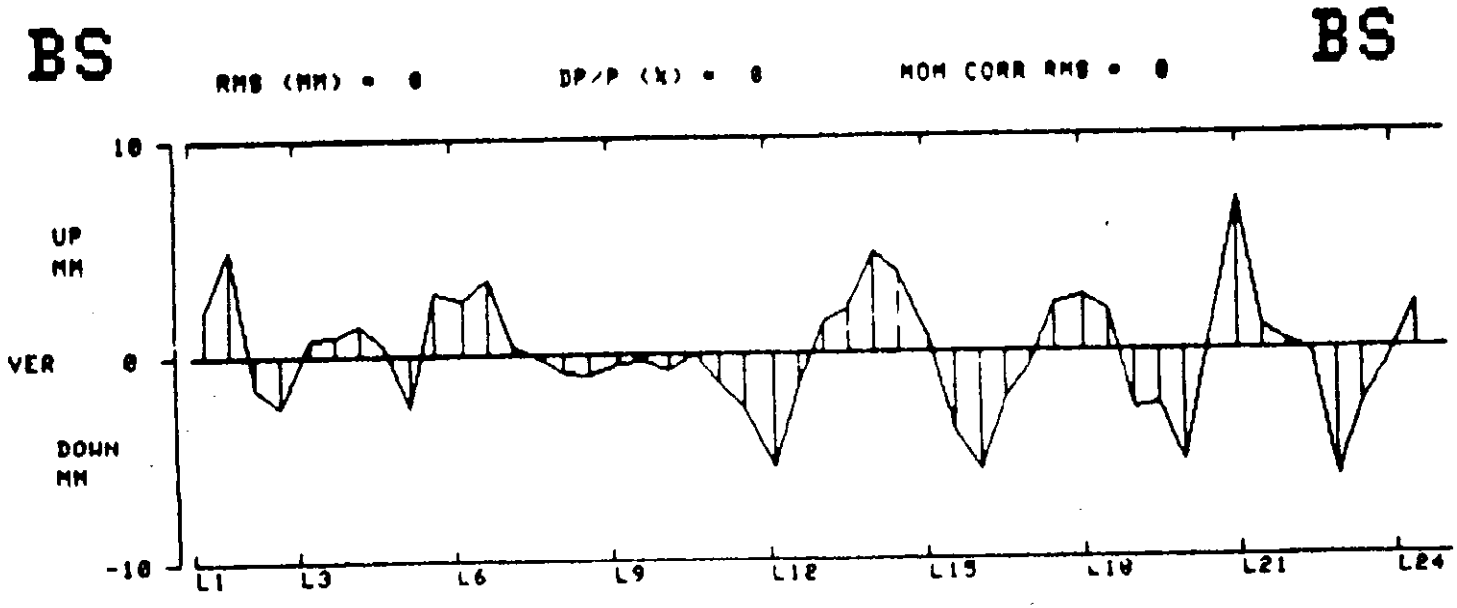

Figure 11. A vertical closed orbit measured in the Fermilab Booster. The horizontal axis is the azimuth, s; the vertical axis is the orbit in mm

V.2 Dispersion

Up until this point we have only obliquely referred to the effect of a momentum error on the transverse position of a beam particle, and the beam momentum spread on the transverse beam size. We did this when we wrote down equation (25) relating the increase in path length to the change in momentum. We did not specify at the time how one might calculate the value of γ_t, a situation which we will now correct.

I would like to start by considering the closed orbit of a particle whose momentum differs from the reference momentum by an amount Δp. To first order the particle sees reduced bending in each dipole by an amount,

$$\frac{\Delta(B\rho)}{(B\rho)} = \frac{\Delta p}{p} .$$

So the effect is exactly the same as if there were an additional error field at each dipole of,

$$(\Delta Bl)_i = (Bl)_i \frac{\Delta p}{p} .$$

We can substitute this expression into (27) to find the closed orbit for an off-momentum particle:

$$x_c(s) = \frac{\sqrt{\beta(s)}}{2\sin\pi Q} \sum_i \frac{(Bl)}{(B\rho)} i\sqrt{\beta_i}\cos(|\phi(s)-\phi_i|-\pi Q) \frac{\Delta p}{p} \qquad (28)$$

or,

$$x_c(s) = \alpha_p(s) \frac{\Delta p}{p} . \qquad (29)$$

The index, i, in (28) runs over all dipole magnets in the accelerator. $\alpha_p(s)$ is called the **dispersion.** It is a function of azimuthal position in the ring, and for accelerators which lie in a horizontal plane is non-zero only in the horizontal dimension. Typical values of the dispersion in proton accelerators range up to 10 meters and tend to be positive.

We now have a complete formulation for the transverse position, with respect to the reference orbit, of a beam particle undergoing betatron motion, with a momentum offset, in the presence of closed orbit errors:

$$x(s) = x_\beta(s) + \alpha_p(s)\frac{\Delta p}{p} + x_c(s) . \qquad (30)$$

Note that **the betatron motion discussed earlier, $x_\beta(s)$, actually takes place relative to a momentum dependent closed orbit.** We can also modify the expression for the physical beam size, (17), to account for any momentum spread present in the beam (assuming the transverse and longitudinal phase spaces are uncorrelated):

$$\sigma_x = \sqrt{\frac{\epsilon\beta}{6\pi} + (a_p \frac{\Delta p}{p})^2} \quad . \tag{31}$$

As you might guess, the transition gamma of an accelerator has something to do with the dispersion throughout the accelerator. I will state the result here and leave verification as an exercise for the student:

$$\frac{1}{\gamma_t}2 = \frac{1}{L} \sum_i \theta_i a_p(s_i) \quad . \tag{32}$$

The sum is over all dipoles in the accelerator, and θ_i is the nominal bending angle of the i^{th} dipole. As a rule of thumb, the transition gamma is approximately equal to the horizontal tune.

V.3 Quadrupole Gradient Errors

We want to look at the effect of a quadrupole gradient error, Δkl, at some point in the accelerator. We will denote by M and M' the unperturbed and perturbed single turn transfer matrices, (16), around the ring, and by Q and Q' the unperturbed and perturbed tunes. In the limit that the length of the offending error is small (compared to a betatron wavelength), M' may be written as,

$$M'(s,s+L) = M(s,s+L) \begin{pmatrix} 1 & 0 \\ -\Delta kl & 1 \end{pmatrix} \quad .$$

Substituting for M from equation (16) gives us,

$$M' = \begin{pmatrix} \cos 2\pi Q + \alpha \sin 2\pi Q - \Delta kl \beta \sin 2\pi Q & ***** \\ ***** & \cos 2\pi Q - \alpha \sin 2\pi Q \end{pmatrix}$$

where I have not bothered to write the off-diagonal terms since we are only interested in the perturbed tune. As you remember, the tune is related to the trace of the single turn transfer matrix via $2\cos 2\pi Q = \mathrm{Trace}(M)$. So the perturbed tune is given by,

$$\cos 2\pi Q' = \cos 2\pi Q - \frac{\Delta kl \beta}{2} \sin 2\pi Q \quad . \tag{33}$$

The β in the expression is the lattice function at the location of the gradient error. The functional dependence of the perturbed tune, Q', on the unperturbed tune, Q, is shown in Figure 12 for $\Delta kl\beta/4\pi = .05$. There are several features of the figure I would like you to take note of: First, there is a region of unperturbed tunes near to $Q=1/2$ for which there is no solution for Q'. In this region the single turn transfer matrix has a trace less than -2 signifying the lack of stability in the transverse plane as we discussed earlier. The region around $Q=1/2$ devoid of stable orbits is called a **stop-band.** The width of the stop-band is related to the strength of the gradient error. Clearly, **we would like to avoid designing an**

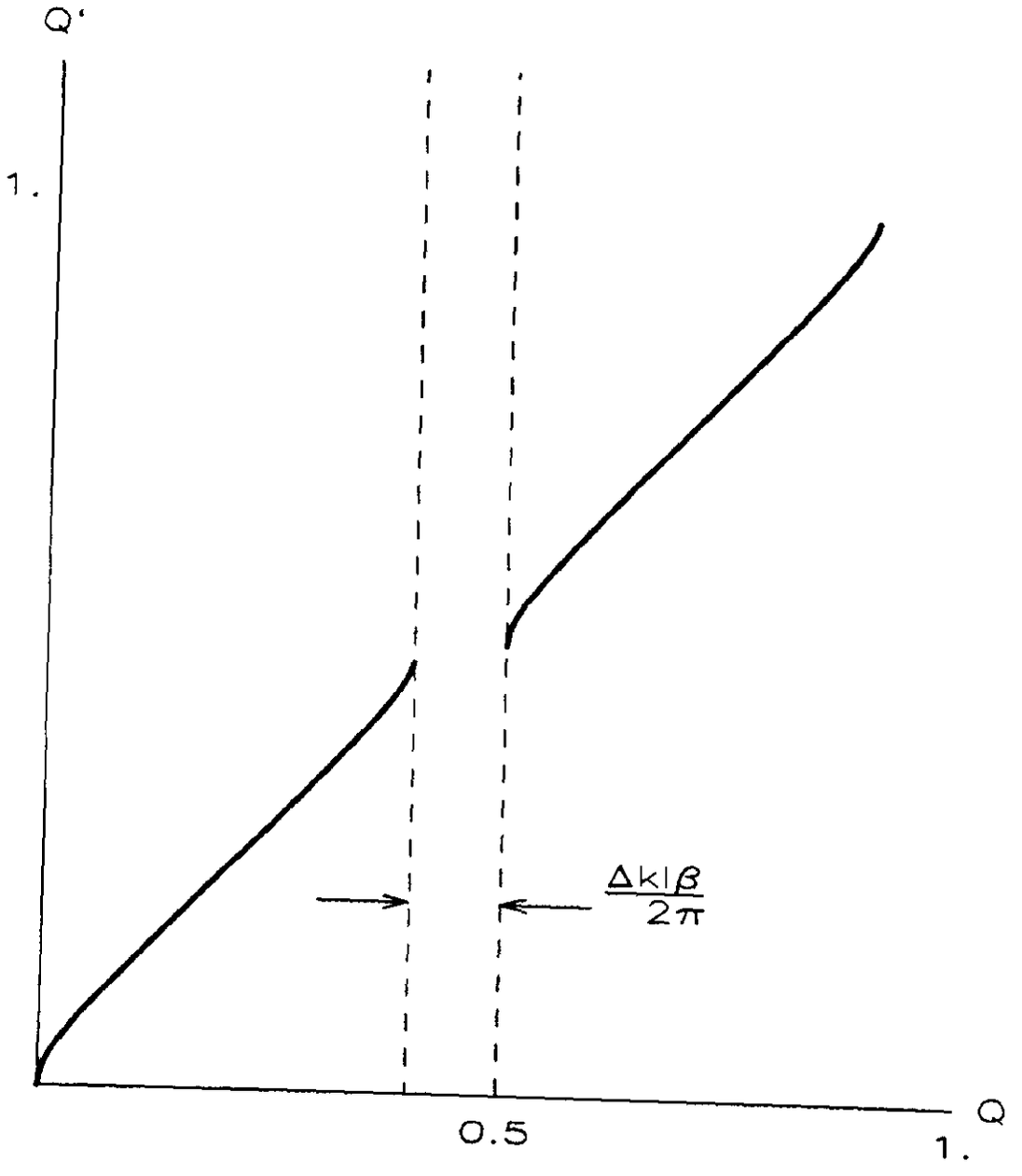

Figure 12. The perturbed tune, Q', resulting from a gradient error $\Delta k l \beta = .2\pi$.

accelerator with a tune near 1/2, if we wish to minimize our susceptibility to having an accelerator which does not possess stable orbits. The second thing to notice is that for tunes not too close to 1/2, and small gradient errors the change in tune, ΔQ (=Q'-Q), is given by.

$$\Delta Q = \frac{\Delta k l \beta}{4\pi} \ . \tag{34}$$

The half-integer stop-band is not a problem in a well designed accelerator. If a problem is expected, however, adjustable correction quadrupole magnets can be distributed around the ring to tune it out.

V.4 Chromaticity

We just got through showing how a change in quadrupole strength is related to a change in the tune. Since the effective focusing strength of a quadrupole, k, depends on the momentum as $(\Delta k/k)=(\Delta p/p)$, a beam particle with a momentum which differs from the reference momentum by an amount Δp, will have a tune,

$$\Delta Q = -\sum \frac{(kl)_i \beta_i}{4\pi} \frac{\Delta p}{p} \ . \tag{35}$$

The sum in (35) is over all quadrupoles in the accelerator and must be performed separately in each of the transverse planes. The proportionality constant relating ΔQ to $\Delta p/p$ is called the **chromaticity**, ξ, and is defined by,

$$\begin{cases} \Delta Q_x = \xi_x \frac{\Delta p}{p} \\ \Delta Q_y = \xi_y \frac{\Delta p}{p} \ . \end{cases} \tag{36}$$

The chromaticity of a machine containing only dipoles and quadrupoles is always negative due to the decrease in focusing with increasing momentum. As a rule, the chromaticity is about equal to -Q. The chromaticity introduces a tune spread into any beam which possesses an intrinsic momentum spread. This is bad because of the existence of resonances which need to be avoided (we will come to this in a minute). It can be corrected by the introduction of sextupole magnets into the ring at locations of non-zero dispersion.

V.5 Resonances

We have shown in our discussions that there are at least two sets of values of the tune which need to be avoided in the construction of an accelerator: Q=integer and 2Q=integer. The first was related to dipole and the second to quadrupole error fields. You might be developing the

suspicion that this result can be generalized to higher multipoles of the magnetic fields. If so, your suspicions are not unfounded. It turns out that if fields of the form

$$B(x) \propto B_0 x^{(m+n-1)}$$

are present, then one needs to avoid tunes satisfying,

$$mQ_x+nQ_y=k \qquad (37)$$

where m, n, and k are integers. You might be thinking that this means there are no acceptable tunes since (37) rules out all rational numbers. Fortunately, in practice it turns our that avoiding resonances up to $m+n\simeq 5$ is sufficient for an accelerator and up to $n+m\simeq 12$ is sufficient for a storage ring.

To get a quantitative picture of what happens when close to a resonance with $m+n\geq 3$, one has to go back and rewrite the equations of motion, (5), including the higher order multipole field. This gives a non-linear differential equation for which mankind has not yet discovered a general solution. We can get a picture of what is happening, however, by using numerical methods to look at the case $m+n=3$, with the field confined to a single azimuth. We will iterate the difference equations,

$$\begin{cases} x_{n+1} = (\cos 2\pi Q)x_n + (\sin 2\pi Q)x'_n \\ x'_{n+1} = -(\sin 2\pi Q)x_n + (\cos 2\pi Q)x'_n + Ax^2_{n+1} \ . \end{cases} \qquad (38)$$

From (37) we expect some sort of anomalous behavior near $Q=1/3$. Figure 13 shows the results of iterating equation (38) for two values of the tune near 1/3, and two values of the strength of the non-linear term, A. For $Q=.38$, $A=0$ we get phase space trajectories which are circles, as described by (10), and stable orbits exist for all regions of the phase space. For $Q=.38$, $A=1$ we see trajectories which are still circles near the origin, become distorted away from the origin, and then cease to exist if we move far enough away from the origin. Finally, for a tune closer to 1/3, $Q=.35$, and $A=1$ we see a similar plot except with the stable area reduced relative to $Q=.38$.

Figure 13 displays the most common characteristics of higher order (≥ 3) resonances. **In the presence of higher order magnetic multipoles the area of the phase space over which the motion of individual beam particles remains stable is restricted. The size of the available stable area depends both on the strength and configuration of the higher order field, and on the difference between the tune and the resonant tune.** The contour enclosing the stable area in phase space is called the **dynamic aperture**. Except for the case of motion near a <u>single</u> resonance, the

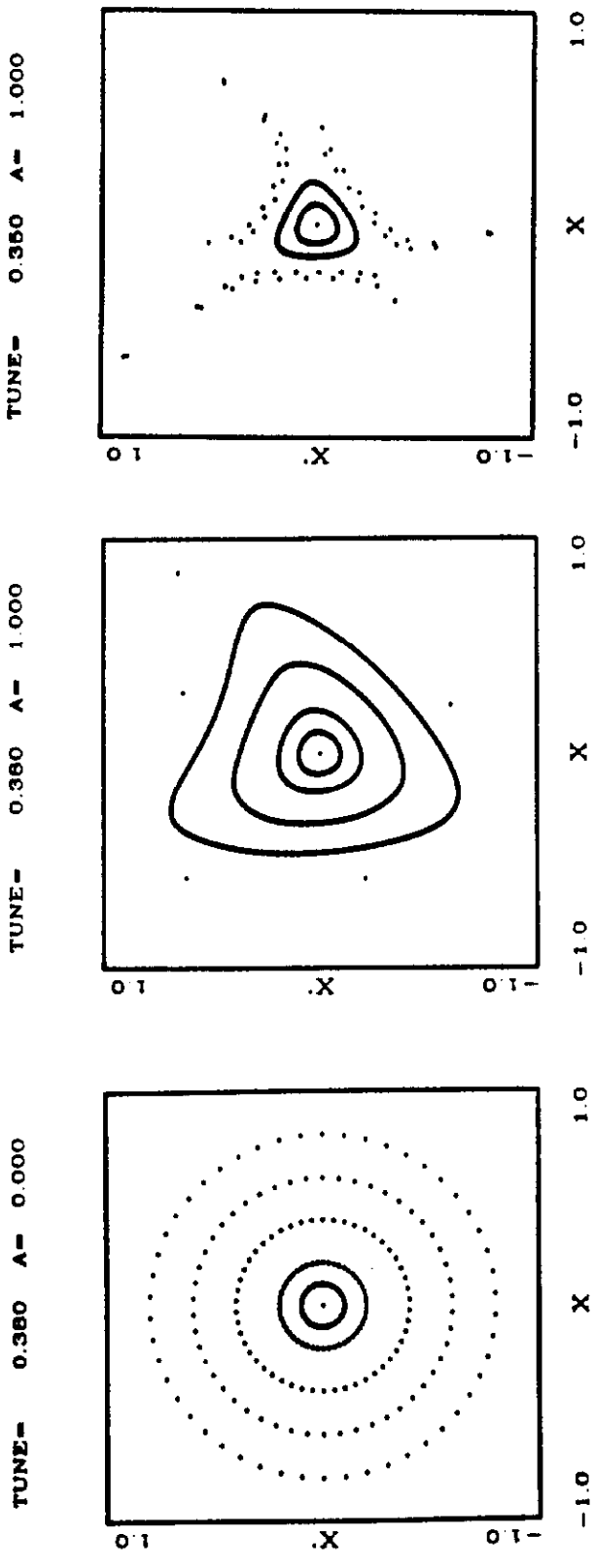

Figure 13. Particle trajectories near a third order resonance

calculation of the separatrix defining the dynamic aperture cannot be done analytically. Particle accelerator designers have started using tracking codes to calculate dynamic apertures. As in the case of the integer and half-integer resonances, higher order resonances can be controlled through the use of adjustable higher order correction elements. As far as I know, no existing accelerator is limited in performance by its dynamic aperture.

This concludes our discussion of the basic principles of acclerator design and operation. We now have the knowledge required to make a conceptual SSC design.

VI. DESIGNING THE SUPERCONDUCTING SUPER COLLIDER (SSC)

We will proceed with our design of the SSC based on the principles illuminated in the preceding lecture along with a few practical points that I will introduce as needed. The design of the SSC as put together by the SSC Central Design Group is contained in the Conceptual Design Report[4] (CDR) published in March of 1986. As we go through our design we will in many ways be approximating the thought processes which went into the creation of the CDR, and we will make reference to that document.

Before we design the SSC we have to know what it is supposed to do. To this end the Department of Energy sets up advisory panels and the high energy physics community organizes workshops in an effort to develop a consensus on the specification for the machine. After years of wrangling everyone agrees on the need for a hadron collider meeting the following specification:

1. The total energy available in the center-of-mass, $\mathbf{E_{cm}}$, should be **40 TeV.**

2. The luminosity, **L**, should be at least $\mathbf{10^{33}\ cm^{-2}sec^{-1}}$. (Luminosity is the proportionality constant relating event rates to cross sections.)

3. On the average there should be **less than 1.4 interactions per bunch crossing.**

4. There should be at least $\mathbf{1.6x10^{-8}}$ **sec. between bunch crossings.**

5. We want **proton-proton** collisions.

6.There should be **six interaction points** available for the installation of experiments.

Within the framework of the specification the only feasible strategy is to build two proton storage rings with individual proton bunches brought into collision at the specified number of locations around the ring.

A storage ring is simply an accelerator capable of holding a particle beam at a fixed energy for a long time. For collisions between counter-rotating beams each having N particles/bunch, with transverse rms beam sizes σ (assumed equal in the horizontal and vertical planes), and with collisions between subsequent bunches occuring at a rate f, the luminosity is written,

$$L = \frac{N^2 f}{4\pi\sigma 2} = L_0 f \quad . \tag{39}$$

L is the luminoisty and has dimensions of $cm^{-2}sec^{-1}$, while L_0 is the luminosity per crossing with has units of cm^{-2}.

Specification 3. tells us what L_0 the experimenters want if we know the total p-p cross section at 40 TeV. We will guess $\sigma(pp)=9x10^{-26}cm^2$ at this energy. We must have $L_0\sigma(pp)$ less than or equal to 1.4, so

$$\mathbf{L_0=1.6x10^{25}\ cm^{-2}.}$$

Specification 2. now constrains the choice of bunch separations to give the total desired luminosity: $f=L/L_0$, or

$$\mathbf{f=62.5\ MHz.}$$

This means that collisions between bunches occur every **16 nsec**, which is consistent with 3. So at least we have not been handed an internally inconsistent specification!

Next we need to do some work on choosing appropriate values of N and σ to get the desired L_0. What we need is,

$$1.6x10^{25}cm^{-2} = \frac{1}{4\pi}\left[\frac{N}{\sigma}\right]^2. \tag{40}$$

We will write σ at the interaction point in terms of the normalized transverse emittance, which we will denote simply as ϵ for the remainder of these lectures, and the lattice function at the interaction point, β^*, assuming we can design the lattice to give $\alpha_p=0$ at the collision point. The mean-square beam size at the collision point is then,

$$\sigma^2 = \frac{\beta^*\epsilon}{6\pi(\beta\gamma)_{rel}}.$$

In general we would prefer to make σ small rather than making N large. Let us ask what a reasonable goal is. We could say that Fermilab runs

with $\epsilon=20\pi$ mm-mr in their collider and we ought to be able to do three times better. So we will assume we can produce a normalized emittance of $\epsilon=6\pi$ **mm-mr.** We could also say that Fermilab has built a ring with a β^* of 1 meter and we ought to be able to do two times better. So we will assume $\beta^*=0.5$ **meters.** If you work it out you will see that this gives an rms beam size at the collision point of

$$\sigma=4.8\text{x}10^{-6} \text{ meters},$$

and according to (40) the number of protons required per bunch is $N=6.8\text{x}10^{9}$.

We have now effectively learned what beam parameters are required to satisfy the physicists' specification without making wild extrapolations from what we know how to do now. There are several minor refinements to our parameters which have to be made. First, we should check that the parameters chosen do not violate the 'beam-beam tune shift criterion'. This criterion is just a limit, based on experience, of how strong the macroscopic electromagnetic interactions between the colliding beams can become before the beams self-destruct. It turns out that we are well below that limit.

The second refinement is a little more subtle, and is based on our desire to keep the counter-rotating proton beam separated everywhere except at the collision point. Because of the short (16nsec=4.8meters) separation between adjacent bunches it is necessary to have the proton beams cross through each other at an angle if a particular proton bunch is to avoid colliding with multiple proton bunches in the other beam as it passes through the interaction region--that is the collisions cannot be exactly head-on. This can be a problem because it can cause a significant decrease in the luminosity if the crossing angle times the bunch length becomes comparable to the transverse beam size. The problem is combatted in the CDR by making the bunch length short through the use of a higher harmonic RF system. The CDR proposes $f_{RF}=375$ MHz. Since this is six times f(=62.5 MHz) only one out of every six RF buckets is filled. This system produces a bunch length of about 6 cm which, when coupled with the crossing angle of $7.5\text{x}10^{-5}$ radians, results in a 15% degradation in the luminsoity. This degradation is made up by increasing the total number of protons per bunch to $7.3\text{x}10^{9}$.

Let us summarize the beam parameters we have specified in order to meet the collider requirements given to us by the physicists:

1. $N = 7.3\text{x}10^{9}$ **protons/bunch.**
2. $f = 62.5$ **MHz.**
3. $\epsilon = 6\pi$ **mm-mr.**

4. $\beta^* = 0.5$ meters.

5. $\alpha_p = 0.$ at the collision point(s).

We now need to design an accelerator which can produce a beam with these characteristics.

VI.1 General Layout of the Collider

We are going to need two identical accelerators to accomodate the two proton beams required, and an associated complex to provide protons at a proper energy for injection into the collider rings. Generically, each accelerator will be made up of **'Arcs'** and **'Straight Sections'**. The function of the arcs is simply transport of the beam between the straight sections. The bending dipoles required to bring the beam around in a circle are, for the most part, located in the arcs. The straight sections are the special purpose units of the accelerator. They provide room for the RF systems, for injection and abort systems, and for the collision regions. In the arcs we will situate the two accelerators on top of each other, while in the collision straight sections we must conspire to bring the two beams together. For logisitical reasons we will try to put the straight sections in close proximity to one another as shown in Figure 14.

The scale of the collider is set by the energy-magnetic field relationship, (3). If we were to use Fermilab collider style magnets (i.e. B=44 kGauss) then the bend radius in the dipoles would be 15000 meters. And if we were to assume that 80% of the total length of the rings were filled with dipoles, then we would be left with a collider of about 120 km in circumference. We would hope that modern magnet designers could raise the magnetic field by perhaps 50%.

VI.2 The Arcs

We are going to make the arcs out of what are called **FODO cells.** When concatenated, FODO cells are the simplest, most efficient means of transporting a beam from one location to another. The generic FODO cell is shown in Figure 15. It consists of alternating focusing and defocusing (in the horizontal plane by convention) quadrupoles interspersed with dipole magnets. The length, L, is measured between the midpoints of the focusing quadrupoles. The **matched** lattice functions, i.e. the lattice functions you obtain if you were to create an entire ring out of these cells, can be calculated from equation (15). We give the result here written in terms of the length of the cell, L, and the betatron phase advance per cell, ϕ.

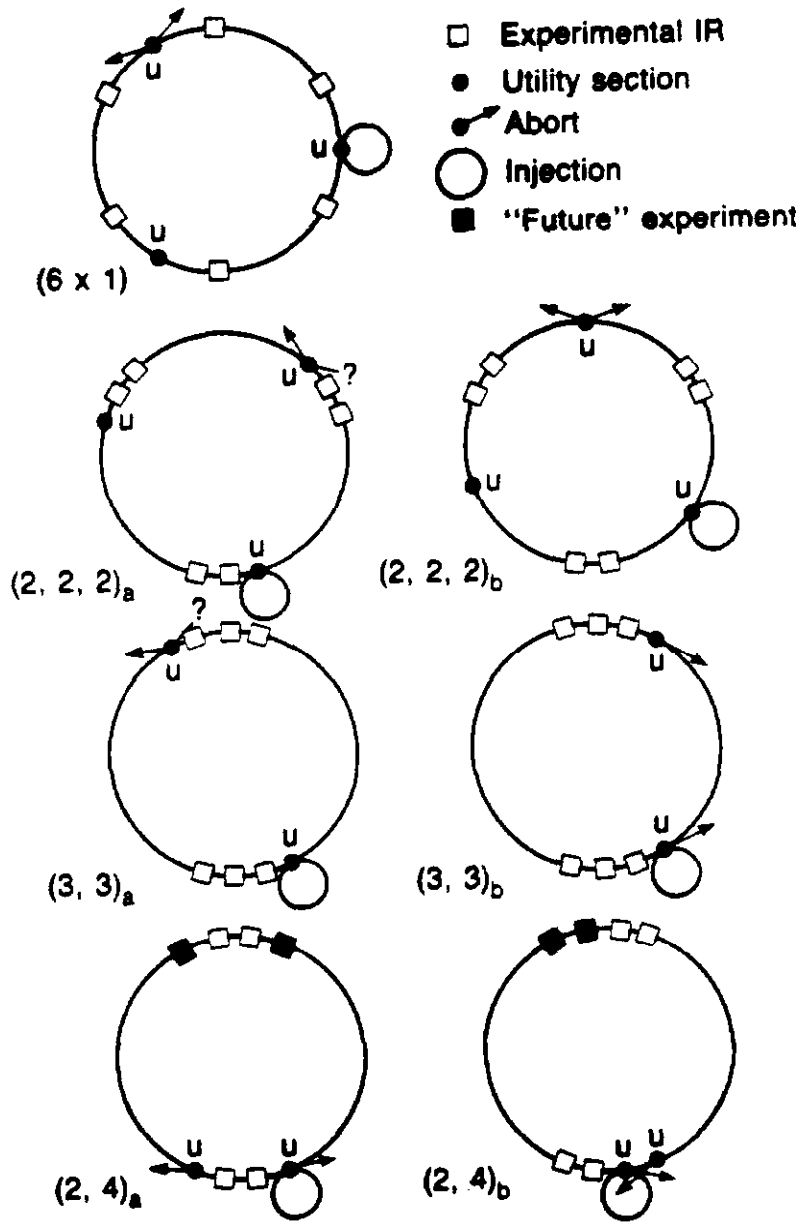

Figure 14. Possible clustering arrangements for the SSC. The arrangement at the lower right was chosen.

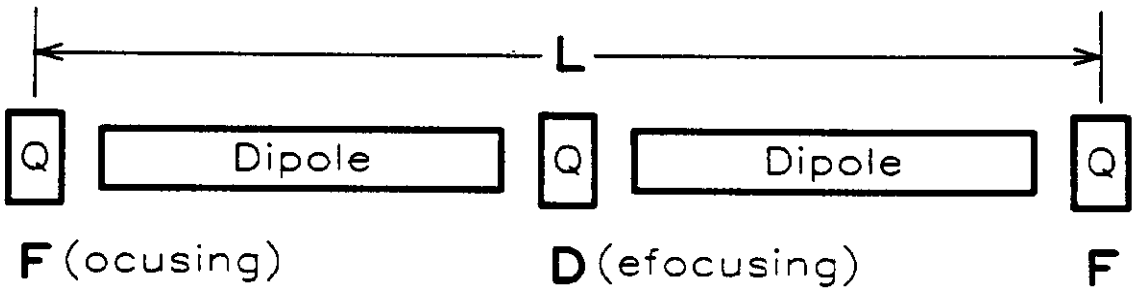

Figure 15. The Generic FODO cell.

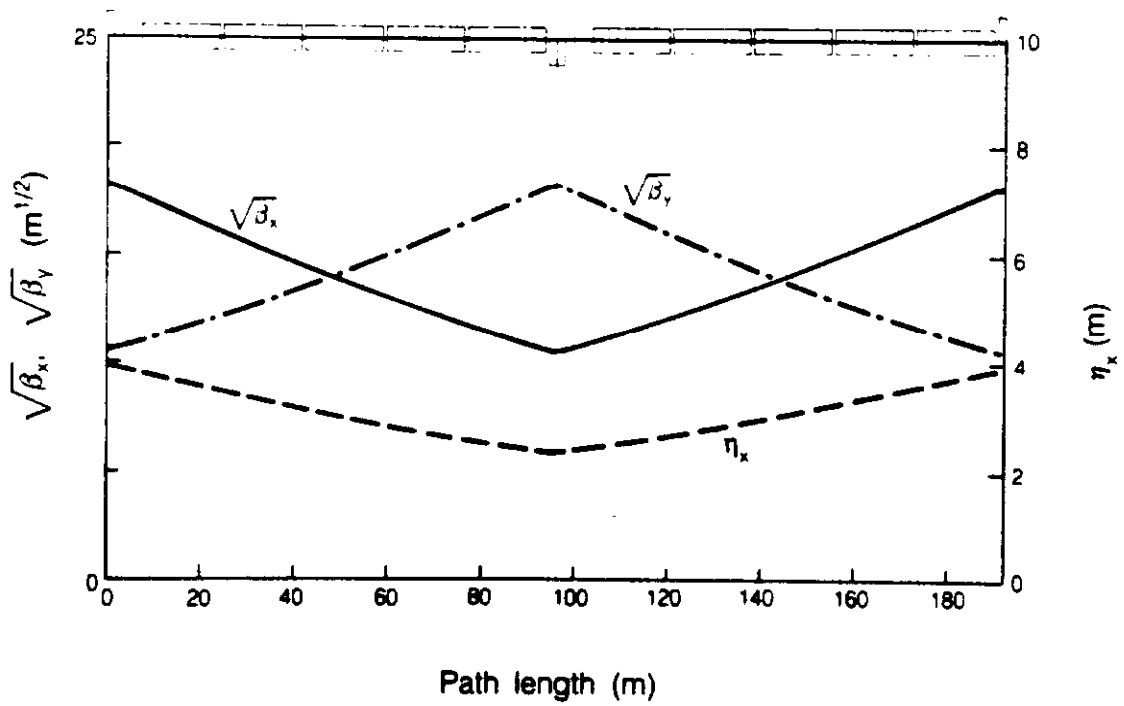

Figure 16. The SSC FODO cell with lattice functions.

$$\beta_F = \frac{L}{\sin\phi}(1+\sin(\phi/2))$$
$$\beta_D = \frac{L}{\sin\phi}(1-\sin(\phi/2))$$
$$\alpha_{pF} = \frac{L\theta(1+1/2\sin(\phi/2)}{\sin^2(\phi/2)} \qquad (41)$$
$$\alpha_{pD} = \frac{L\theta(1-1/2\sin(\phi/2)}{\sin^2(\phi/2)}$$

where θ is the bending angle in one of the dipole magnets. The F and D subscripts refer to locations at the center of the focusing and defocusing quadrupoles. The lattice functions listed are in the horizontal plane. By symmetry the vertical lattice functions are obtained by interchanging F and D indices on the β functions and setting $\alpha_p=0$ in the absence of vertical bending dipoles. Also by symmetry, the lattice function α is equal to zero at the midpoints of all quadrupoles. The quadrupole strength required in this cell is

$$\frac{B'l}{(B\rho)} = \frac{4}{L}\sin(\phi/2) \quad . \qquad (42)$$

If we can define some reasonable criteria for choosing L and ϕ we will have the FODO cell, and hence the arcs of the machine, specified.

The choice of L will represent a compromise between two competing demands. First, we know from experience that the dipole magnets are going to be the single most costly element of the collider, and we also know from experience that **the cost of a superconducting magnet increases very rapidly with the transverse aperture.** This means there will be a premium on **keeping L short to keep β small.** On the other hand we would like to make as few cells as possible to minimize the total number of discrete elements we have to produce, and to keep the dipole packing fraction high. This drives us toward **making L long.** We are going to compromise by using the criterion that **the beam size at injection does not dominate the dipole aperture requirement.**

Since the physical beam size is largest at injection into the collider, the magnet aperture requirement arises from the need to be able to accept the injected beam. The aperture, however, is required to accomodate more than just the physical beam size. It must provide for the possiblity of injection errors, closed orbit errors, and other miscellaneous effects which are estimated to consume about 7 mm of aperture. Providing an injected beam size of much less than this does not gain us anything, while providing a much larger beam would needlessly enlarge the aperture requirement. We are going to specify that the injected beam size, $\pm 4\sigma$, be

4 mm. This will give us a β_F for the assumed normalized emittance of 6π mm-mr once we know the injection energy. We will anticipate a later discussion in stating that the injection energy into the collider will be 1 TeV. Then,

$$\beta_F = \frac{6\pi\sigma^2(\beta\gamma)}{\epsilon} = \frac{6\pi(.0005\text{m})^2(1066)}{6\pi\text{x}10^{-6}\text{m}} = 267 \text{ meters.}$$

Choosing the phase advance per cell, ϕ, is at some level a matter of personal taste. Some people prefer 90° while others like 60°. There isn't really that much to choose as far as I am concerned. In fact, over the range 60° to 90° the choice of phase advance has almost no effect on the length L. The Central Design Group chose 60° for the cell in the CDR. According to (41) this would give us L=154 meters for our choice of β.

To determine how many cells are needed in the arcs we need to know what magnetic bending field we can provide. We ask the magnet gurus, who tell us '66 kGauss'. This gives us a bending radius in the dipoles of 10.1 km, and assuming that we can fill 90% of the length of the cell with dipole magnets gives a total arc length of about 67 km. The number of cells required is then about 67000/154, or 430.

So we have finished designing the arcs. We show below the comparison between the arcs that we have designed and those designed by the Central Design Group. They are really very similar.

Property	Us	CDR
Length of Cell	154 m	192 m
Phase Advance	60 degrees	60 degrees
Number of Cells	430	332
β_F	267 m	325 m
Tune of the Arcs	71.7	55.3

The FODO cell from the Conceptual Design Report is shown in Figure 16. (Note that the CDR denotes dispersion by η_x.)

VI.3 Designing the Straight Sections

Straight sections need to be included into the collider because the FODO cells do not have properties which make them useable for certain special purposes. For example, we said that at the collision point we would like to have $\beta^* = 0.5$ m, $\alpha_p = 0$, and the experimenters also require "Lots of room for the detector". The cell shown in Figure 16 does not come close to meeting any of these requirements.

In designing any special purpose straight section the trick is to design it in such a way that it does not disturb the natural lattice functions in the arcs. Creation of such a design is called **matching**, and a straight section which satisfies this requirement is called a **matched insertion.** In order to qualify as a matched insertion the straight section must satisfy one of two properties: 1) The transfer matrix of the insertion is the unit matrix; or 2) The transfer matrix has the property that it transforms,

$$\beta(\text{in}) \rightarrow \beta(\text{out})$$
$$\alpha_p(\text{in}) \rightarrow \alpha_p(\text{out})$$
$$\alpha(\text{in}) \rightarrow -\alpha(\text{out})$$
$$\alpha_p'(\text{in}) \rightarrow -\alpha_p'(\text{out}).$$

Condition 1) simply makes the straight section invisible to the rest of the ring. It is a sufficient but not necessary condition for a matched insertion. Condition 2) is the less restrictive necessary and sufficient condition. It works because of the inherent symmetry of the situation in which the straight section is bounded on each side by identical FODO cells. Condition 2) is the one accelerator designers usually apply. Numerous computer programs have been written to calculate transfer matrices which satisfy 2) while at the same time providing the desired optical properties within the straight section.

An example of a special purpose straight section is shown in Figure 17, taken from the SSC CDR. It is one of the utility straight sections housing the RF, injection, and the abort. You can notice the FODO cells at both the left and right hand side of the figure, with a big empty space in between. The empty space in between has properties which are conducive to the specialized needs: moderately large β functions with lots of space for injection and extraction, and dispersion free regions for RF. The total length of the straight section is about a mile.

VI.4 Designing the Interaction Region

Designing the straight sections accomodating injection, extraction, and RF is generally pretty straightforward. Associated with the design of the straight section in which collisions are to take place are several special problems which makes this job a little more challenging.

The first challenge is to bring the beams together. (I might note that in a proton-antiproton collider in which the beams circulate within a common ring this is a non-problem as long as the number of bunches is small.) A series of vertical bending magnets is used to bring the beam together as shown in Figure 18. Note that this will create a vertical dispersion which must be matched in order to keep it localized.

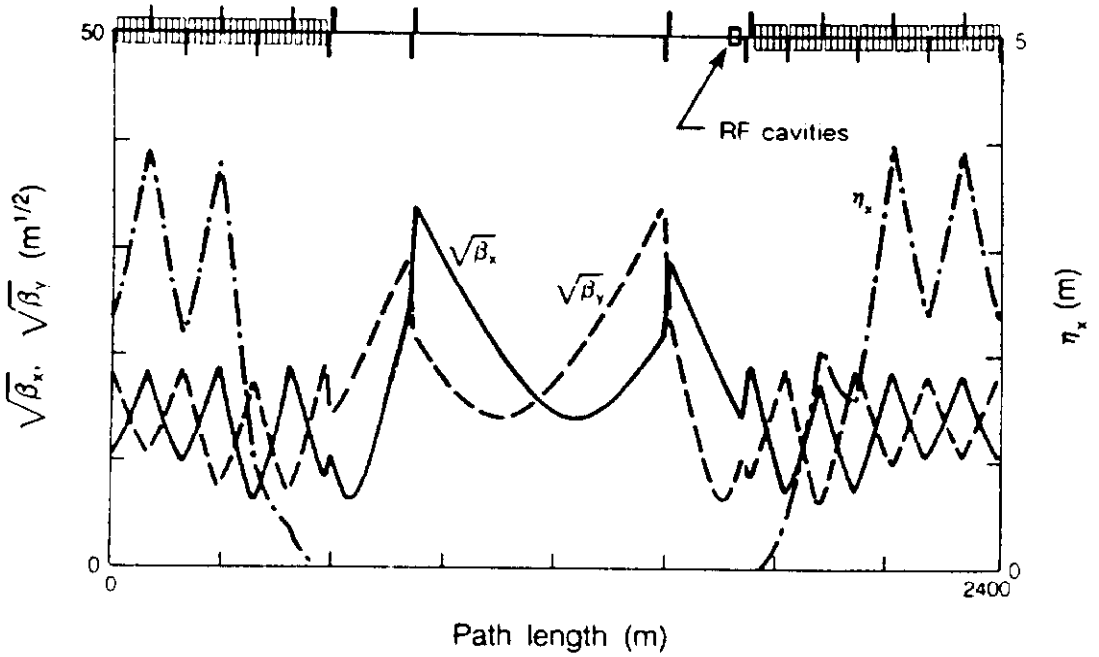

Figure 17. The SSC utility straight section housing RF injection, and aborts

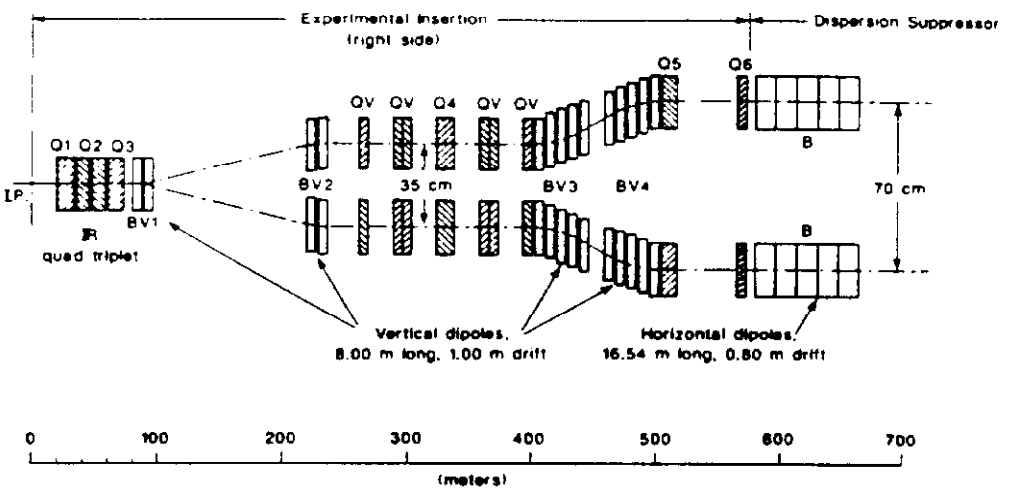

Figure 18. Side view of the geometrical layout of the SSC interaction region.

The more serious problems are related to the optics needed to produce the low β^* required for high luminosity, while at the same time providing sufficient free space for the experimenters to install their detector. The problem arises because the β function grows quadratically with distance from the collision point, s, according to,

$$\beta(s) = \beta^* + \frac{s^2}{\beta^*} .$$

The low β^* is usually provided by a very strong quadrupole doublet or triplet. (A single quadrupole cannot be used because it could not provide simultaneous focusing in both planes.) As s increases and/or β^* decreases (both desirable from the experimenters' point of view), the β function at the end of the quadrupole triplet grows. The situation is even worse than you might think because this quadrupole at the end of the triplet is going to defocus the beam further in one of the two planes. As a rule of thumb, the maximum β that is produced is related to β^* and the length of the free space, l, by

$$\beta^* \beta_{max} \simeq 10\ l^2.$$

This effect can be seen in Figure 19 (top) where the SSC interaction region is displayed. The FODO cells are on the left and the interaction point is at the far right of the figure. One half of the straight section is shown with mirror symmetry around the interaction point. Associated with the β^* of 0.5 meters and the free space of 20 meters is a β_{max} of 8000 meters.

The large β_{max} creates two sorts of problems. First, it makes the accelerator much more susceptible to machines errors (c.f. equations (27) and (33)). And second, it produces a very large beam in the quadrupole triplet during injection. In fact, the beam in the quadrupole triplet is so large during injection that it is impossible to contemplate building these quadrupoles with the required aperture. The problem is solved by using different optics during injection, as shown in Figure 19 (bottom), and then retuning the straight section after the beam has been accelerated and is smaller. This is not an easy maneuver to complete without killing the beam, but there is experience with following just such a procedure in the Fermilab collider.

VI.4 The SSC Design

We have essentially completed the optical design of the collider. With the addition of the straight sections shown in Figures 18 and 19 (two utility straights and four interaction region straights) we have increased the tune of the accelerator(s) to 78.3 and the circumference to 82.94 km.

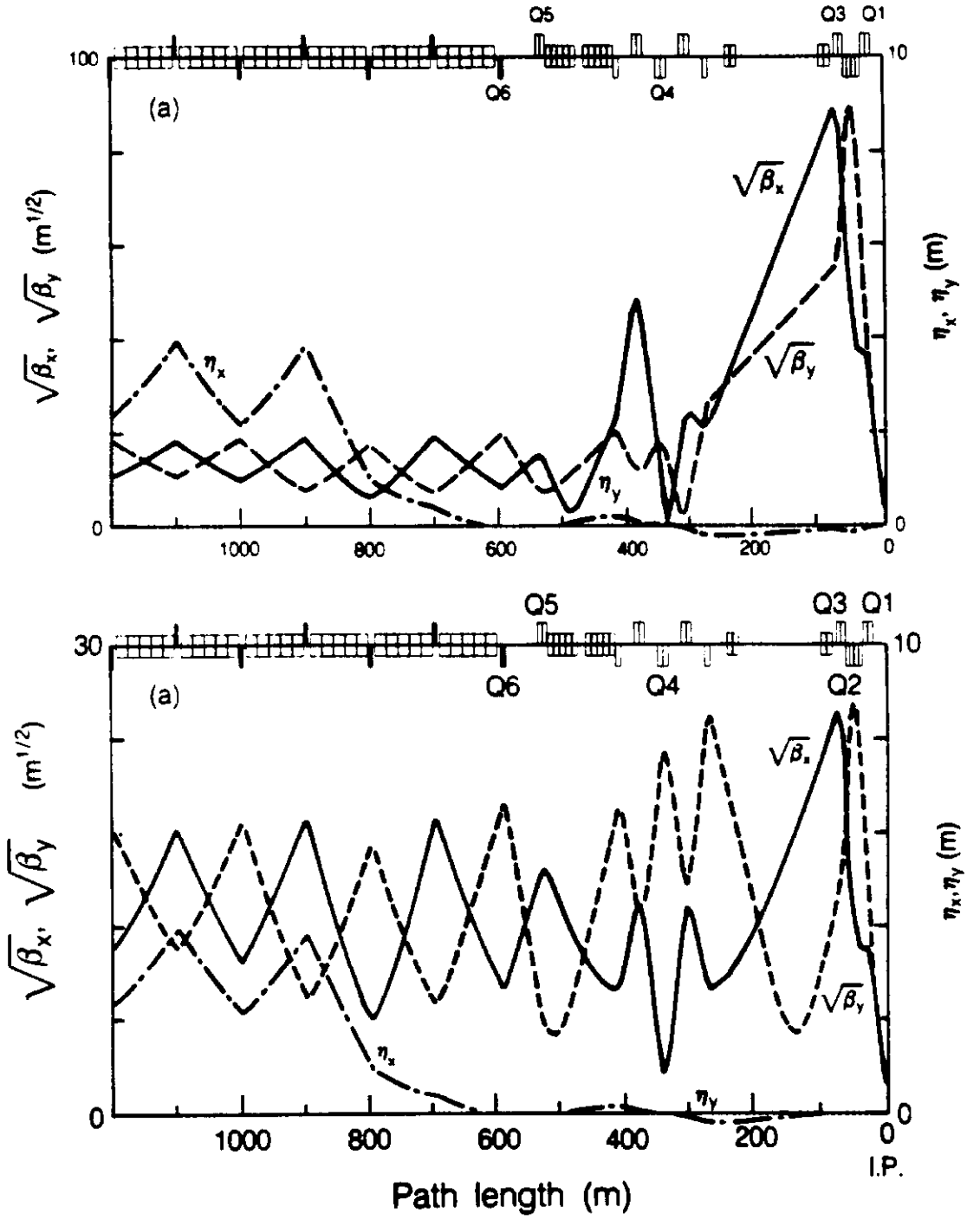

Figure 19. Interaction straight section with lattice functions tuned for collisions (top) and for injection (bottom).

All that remains to be done is to specify the magnet 'good field aperture' requirements for the magnet builders, and to specify the injector requirements to the injector builders. We have already discussed the magnet aperture requirements. What is needed is a good field aperture of 10 mm. The primary injector requirement is for an accelerator complex capable of providing the design beam intensity and phase space density at an enery of 1 TeV. The injection energy is determined by the requirement that persistent current effects in the collider magnets be manageable as discussed in the previous lecture.

We now have it!--An SSC design which is shown in Figure 20. It looks good on paper, so let's hope somebody buys it.

VII. OTHER SORTS OF ACCELERATORS

I want to leave you with a least an inkling of the differences between the proton accelerators/colliders which I have been describing here and the other sorts of high energy accelerators used for physics research. These include electron storage rings, antiproton sources, and electron linear colliders. We turn first to electrons.

Electron storage rings have totally different scales of energy, circumference, and emittance than proton accelerators. The differences are a reflection of the dominant role played by **synchrotron radiation** in electron machines. In completing a single revolution of an accelerator an electron loses energy in an amount,

$$U(\text{MeV}) = .088\frac{E^4(\text{GeV})}{\rho(\text{m})} \quad . \tag{43}$$

The most immediate consequence of (43) is that it gets very hard to pay the power bill as the energy goes up. For example, at the Large Electron Positron Collider (LEP) at Cern an 80 GeV electron will lose 600 MeV of energy every time it goes around the ring--energy which has to be replenished by the RF system! The economics connected with (43) are such that **the circumferences of electron accelerators tend to scale as the Energy^2** rather than with the Energy as in proton rings. This results in the interesting feature that the highest energy electron rings use magnets with the lowest magnetic fields! One attempt to defeat the consequences of (43) is the Stanford Linear Collider in which electrons are accelerated in a straight line before being brought into collision and subsequently discarded. It is interesting to note that protons do produce synchrotron radiation, but at a rate reduced from (43) by $(m_e/m_p)^4$. The SSC will be the first

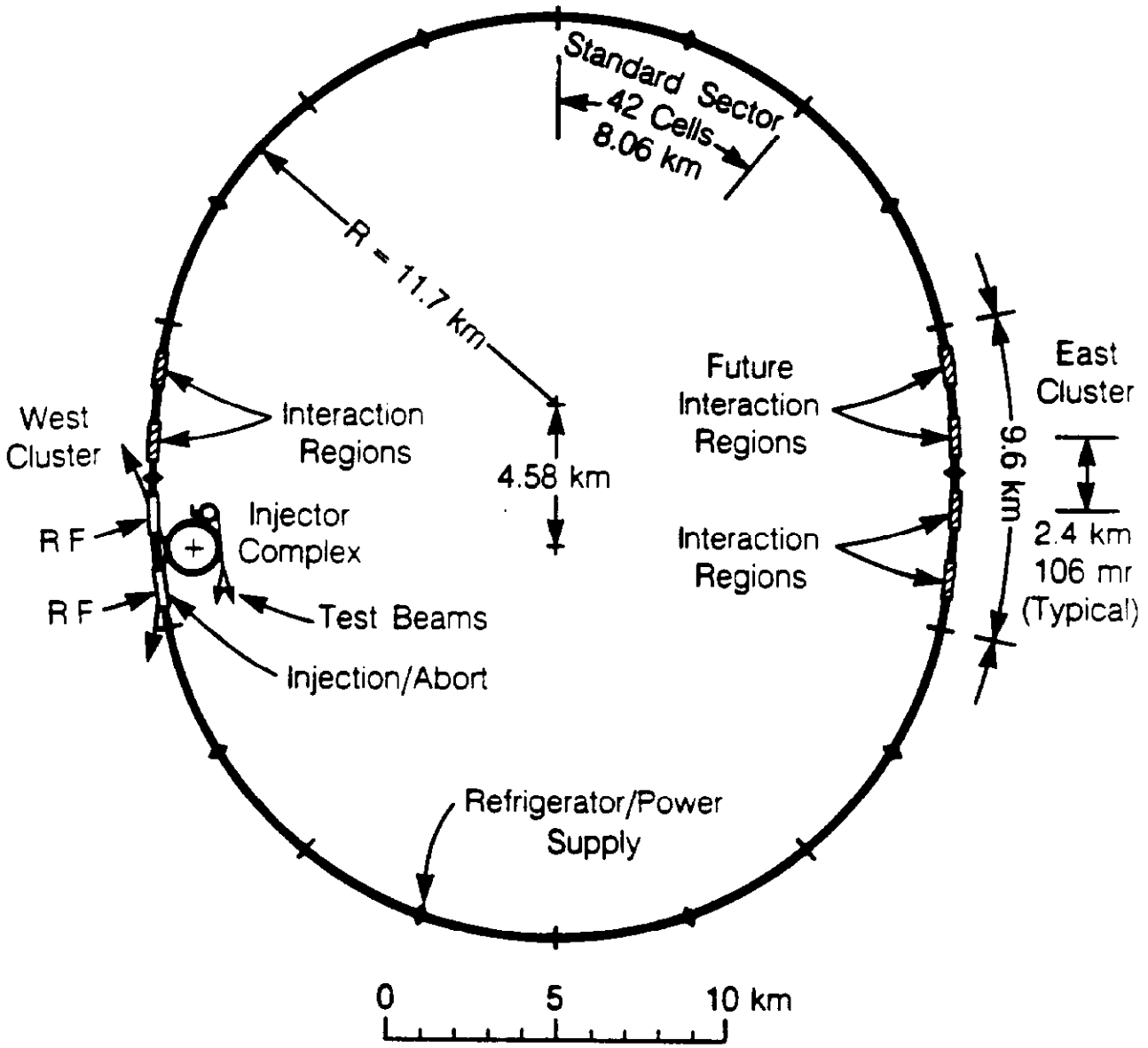

Figure 20. The layout of the SSC.

proton accelerator to have measureable synchrotron radiation effects--primarily an increased heat load on the magnet refrigeration system.

A second consequence of synchrotron radiation is the presence of **radiation damping** in electron storage rings. Since the RF system is providing energy to the electrons even while they are not being accelerated, the mechanism shown in Figure 3 works to reduce the beam emittance even without acceleration. At the same time, the energy dependence of the synchrotron radiation insures that the longitudinal phase space is damped also. The beam emittance is shrunk to a level at which the damping and quantum fluctuations in the radiation process come into equilibrium. The result is that **the emittance of an electron beam in a circular accelerator is an intrinsic property of the accelerator itself,** i.e. it depends on the details of the lattice functions. While the emittance behaves very differently in electron and proton rings, the entire formalism we described for treating the transverse and longitudinal motion of individual particles remains unchanged.

Antiproton sources are specialized storage rings used to provide antiprotons for proton-antiproton colliders. The problem they have to cope with is the difficulty in collecting large enough numbers of antiprotons into a small enough phase space to produce a decent luminosity in a collider. Antiprotons are produced by striking a metal target with high energy protons, and collecting all the negatively charged particles which emanate at some energy into some transverse acceptance. The yields are very low. At Fermilab, using 120 GeV protons, into a momentum acceptance of 3% at 8 GeV, and into a transverse acceptance of 20π mm-mr the yield is about one antiproton for every million protons striking the target.

Not only are there not very many of them, but the antiprotons are also widely dispersed in phase space. They are accumulated over many hours and simultaneously their emittance is reduced using a **stochastic cooling** system. The stochastic cooling system reduces the otherwise invariant emittance by sensing the position of individual antiprotons at a **pickup** located at a specified position around the ring, and then sending a signal to a **kicker** located downstream which applies a correcting impulse to the particle. The cooling rate is related to the bandwidth (i.e. time resolution) of the cooling system, W, and the total number of antiprotons in the ring. In a perfect cooling system the cooling time is

$$\tau_c = \frac{N}{W} .$$

In the Fermilab Antiproton Source the number of antiprotons ranges up to 5×10^{11} and the bandwidth of the cooling systems is 2 GHz. This

results in an optimum cooling time of 4 minutes. The highest antiproton accumulation rate recorded at Fermilab (or at CERN) is 1.2×10^{10}/hour. At this rate several hours are needed to accumulate a useful number of antiprotons. You might find it interesting to calculate how many grams of antimatter are being produced per year at this rate.

VIII. SUMMARY

I hope that you come away from these lectures with an understanding that particle accelerators are rational devices which obey the laws of physics. We discovered that the macroscopic beam characteristics can be understood in terms of the motions of individual beam particles, which are in turn described by simple harmonic motion. We found that higher order nonlinear effects are a potential fly in the ointment which cannot always be ignored. And we saw how the scales of accelerators are set both for proton and electron accelerators. Finally, we learned enough of the jargon to understand that whereas the language of accelerator physicists may not be completely understandable to the uninitiated, the underlying principles are simple and easily understood.

There is a multitude of topics I did not cover here, either because of lack of expertise or because of time restrictions. These include linacs, accelerator diagnostics, injection and extraction, beamlines, limits on accelerator performance and beam instabilities, and further details of electron accelerators and antiproton sources. There are many interesting problems waiting to be solved in many of these areas.

References

* Operated by Universities Research Association under contract to the United States Department of Energy

1. E.D. Courant, M.S. Livingston, and H.S. Snyder, "The Strong-Focusing Synchrotron--A New High Energy Accelerator",Physical Review, 88, 1190 (1952)

2. N.C. Christofilos, Unpublished Report (1950)

3. E.D, Courant and H.S. Snyder, "Theory of the Alternating Gradient Synchrotron", Annals of Physics **3**, 1 (1958).

4. SSC Central Design Group, "Superconducting Super Collider Conceptual Design", SSC-SR-2020 (1986)