SLAC-PUB-5397
January 1991
(T/E)

B-FACTORY STORAGE RING DESIGN*

John R. Rees

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309

Contributed to the 18th Annual SLAC Summer Institute on Particle Physics, Stanford, CA, July 16-27, 1990.

⋆ Work supported by Department of Energy contract DE-AC03-76SF00515.

1. The Challenges of a B-Factory

In the past few years a good deal of enthusiasm has arisen in the US, Europe and Asia for B-Factories. In these machines electrons and positrons are collided with center-of-mass energies at or near the $\Upsilon(4s)$ resonance, with unprecedentedly high luminosities, to produce copious fluxes of B-mesons. The object is to make high-precision studies of the CP non-conserving B decays. Various colliding-beam configurations have been suggested including both linear colliders and storage rings, but one scheme has emerged as generally preferable to the others. It is the asymmetric storage ring system—asymmetric in the sense that the two beam energies are markedly different and the center of mass is moving in the direction of the higher energy beam. With this arrangement the decaying B-mesons fly off from the interaction region in the same direction, and the time-order of their decays can be deduced from the locations of their decay vertices.

These B-Factories present the accelerator builder with two main challenges: to achieve luminosity far beyond that attained in existing storage rings and to do it in the unexplored arena of unequal beam energies. The highest luminosity reached up to now in electron-positron storage rings has been a little over 10^{32} cm^{-2} s^{-1} attained in Cornell's CESR. Although several machines have been designed to deliver luminosities of 10^{32} cm^{-2} s^{-1}, that goal has proven very difficult to reach in practice. A B-Factory requires luminosities thirty to one hundred times higher than that figure to support the CP-violation experiments. And on top of that difficulty, we must deal with colliding beams of unequal energies. That may turn out to introduce no punitive limitations, but it surely represents an uncertainty that aggravates the problems of attaining high luminosity. On the whole, we can fairly conclude that all of our knowledge and ingenuity will have to be brought to bear in designing these machines and building them.

Fortunately the means of meeting these challenges appear to be in hand on the basis of our present understanding of the accelerator physics of colliding-beam storage rings. The problems have been studied in several laboratories in Europe, Japan, the US and the USSR, and the solutions devised in those studies have converged in their general features. A B-Factory will consist of two separate storage rings with a common collision region; each ring will carry what is, by today's standards, high circulating beam currents, and as a consequence, the vacuum chambers will be very well-cooled and strongly vacuum-pumped; and each beam will be comprised of many bunches. It appears that the optical and mechanical designs of the interaction regions will be quite complicated, but also quite feasible.

2. An Existing B-Factory: CESR

There are in fact two storage rings running now as B-Factories: CESR at Cornell and DORIS at DESY. While they are not capable of the fluxes of B-mesons necessary for precision studies of CP violation, they are the sources for a lot of the B physics done to date. As I have mentioned before, CESR holds the luminosity record for electron-positron storage rings, and a description of the machine will be useful to illuminate the difficulties in reaching high luminosities.

CESR is a single storage ring with a circumference of 768 m. Some of its primary parameters are given in Table 1. The beams of course have equal energies which are tuned to half the energy of the $\Upsilon(4s)$ resonance. The luminosity has been referred to before. The maximum tune shift is a figure of merit that we shall discuss in some detail later. The 77 mA stored in each beam radiate 80 kW as synchrotron radiation, but since the source is distributed around an appreciable fraction of the circumference (radiation is emitted in the bending magnets), the power density is not prohibitive.

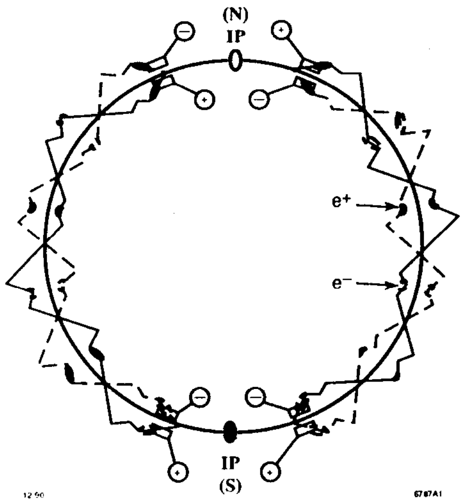

Figure 1: A schematic depiction of the "pretzel orbits" in the CESR storage ring. The circle represents the central orbit; the other solid line is the positron orbit; and the dashed line is the electron orbit. The electron and positron bunches avoid each other, because they are deflected onto different orbits by the electric fields situated on either side of the interaction region.

Each beam has seven bunches and, therein lies the secret of CESR's high luminosity. Originally built as an 8-GeV machine with one bunch in each beam and a small beam aperture, its performance was stringently curtailed by the tune-shift limit. More on this later. The Cornell accelerator physicists overcame this obstacle by creating "pretzel orbits" that prevent the counter rotating bunches from colliding with one another except at the interaction regions. High-voltage electric plates make strong electric fields in the vacuum chamber transverse to the beam axis and deflect the electron and positron beams oppositely onto their separate pretzel orbits. The technique is shown schematically in Figure 1.

Returning to Table 1, we should note for later reference that the bunch frequency, the frequency at which the bunches of either beam pass the interaction point (or any other fixed point on the circumference, for that matter) is 2.7 MHz. The vertical beta-function at the interaction region is 1.5 cm. We shall discuss the significance of the beta-function later; suffice it to say that the smaller it is, the better. CESR is a flat-beam storage ring, so the transverse aspect ratio of the beam at the IR (interaction region) is small. The longitudinal impedance is a parameter we shall discuss at some length later; again the smaller, the better.

To push CESR to still higher luminosity more bunches would have to be added with a corresponding increase in circulating currents and radiated power and, possibly, a reduction in longitudinal impedance. A reduction in the IR beta-function would help too, but there is not much to be hoped for there. These measures are very difficult in a single ring.

3. Constraints Imposed by Beam Dynamics

Why can't we get all the luminosity we want? We know that there are what we might think of as engineering problems, *e.g.,* handling high heat fluxes and pumping large gas loads. Nevertheless, if we are willing to spend the necessary effort and money, shouldn't we be able to overcome the problems? In fact, the answer is, "Not necessarily." Some of the problems are beyond the power of engineering solutions. Those problems are the subject of this chapter. The physics of particle beams imposes strict limitations on what we can accomplish. The phenomena that harass us at every turn in our pursuit of high luminosity are beam instabilities. Either the individual particles go into unstable motion or whole bunches do. In either case the beams become useless for producing luminosity. These instabilities can be classified in several ways; for our purposes we shall classify them into two categories: *beam-beam instabilities* and *single-beam instabilities.*

We shall see that single-beam instabilities will prevent us from increasing bunch populations much beyond present practice, while the demand for high luminosity will force us to higher total beam currents. As a result we shall be forced to choose a double-ring system with many bunches in each ring. Then we shall find that the circumferences of the two rings must stand in the ratio of small integers, optimally 1/1, to avoid coherent beam-beam instabilities.

Happily in the end, we shall find a set of design choices that can deliver the luminosity we need for a B-Factory.

3.1. Beam-Beam Instabilities

Storage rings are designed so that the motion of a typical stored particle is comfortably stable. A particle in orbit by itself would stay in the ring forever if it experienced no other forces than the magnetic guide forces. In actuality any particle emits quanta of synchrotron radiation and interacts with the occasional residual gas molecule, so after a time, typically hours, its life in the storage ring comes to an end. But those influences do not vitiate its usefulness to us. We can use particles that stay in the ring for hours.

The forces that cause the main trouble when we try to increase the beam current (which we must do to raise the luminosity) are the electromagnetic forces between the stored particles. The particles of a bunch together create a collective field that moves along with the bunch. The metallic vacuum chamber establishes boundary conditions and thereby shapes the field of the bunch, an effect we can describe in terms of image charges that also move along with the bunch. This is the collective field, and it acts on each individual particle in the bunch that creates it. It also acts (impulsively) on each particle of each bunch that collides with the bunch that creates it. For these particles, the forces of the collective field add to the guide-field forces and can change the particle's motion from

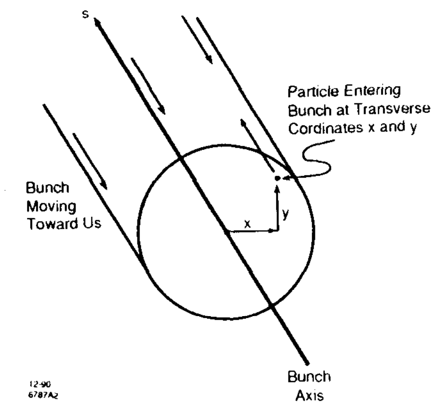

Figure 2: A particle entering a counter-moving bunch at transverse coordinates x and y. The bunch is idealized as a cylinder of charge.

stable to unstable. In this section we shall discuss the particular effects of these collective fields between colliding bunches.

Consider a particle of one beam as it passes though a counter-moving bunch. It is acted on by both the electric and magnetic collective fields of that bunch; the two forces are almost equal, they add and their dependencies on the transverse coordinates of the particle are identical. More importantly, that dependence is *very non-linear* in both coordinates. If the particle penetrates the center of the bunch, there is no force. As the line of penetration moves out from the center, the force first increases and then decreases. These are the properties of the electromagnetic field of a thin tube of space charge current.

As the particle passes through the opposing bunch, it accumulates a transverse impulse that creates a corresponding transverse deflection. The beam particle coordinates are measured transverse to the central closed orbit of the storage ring and are customarily called x (horizontal) and y (vertical). See Figure 2. The slopes are therefore x' and y', where primes denote differentiation

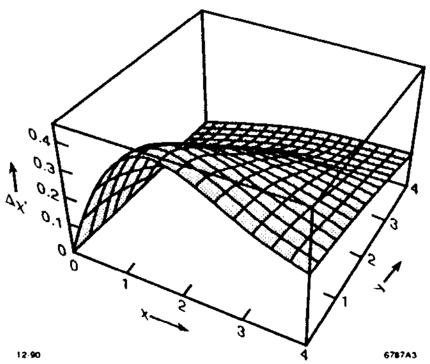

Figure 3: The transverse deflection, $\Delta x'$, of a particle due to the collective field of the oncoming bunch, plotted against the coordinates at which the particle passes through the bunch. In this example the bunch is distributed as a round Gaussian, and the units of x and y are the radial standard deviation.

with respect to the longitudinal (along-the-orbit) coordinate s. The effect of a transverse impulse is to change x' and y'. Figure 3 is a three-dimensional plot showing a typical relationship between the horizontal impulse $\Delta x'$, plotted along the vertical axis, and the particle's coordinates, plotted along the other two. The warping of the surface implies a high degree of non-linearity. This particular example is of a round beam, but the non-linearity is typical of all beams. The round beam of Figure 3 has a Gaussian radial density distribution. The units of x and y are the radial standard deviation of the distribution, the characteristic transverse scale length (size). The feature that the impulse peaks and reverses slope at a distance of the order of the transverse size is also typical of all beams.

3.1.1. The Incoherent Beam-Beam Limit

Why is this non-linearity such a bad thing? To answer that question, we need to delve a bit into beam dynamics. In particular, we need to discuss betatron resonances—henceforth, just resonances.

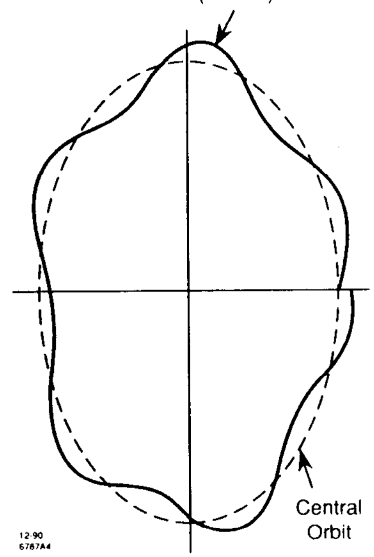

Figure 4: A betatron oscillation.

As a first step in the design of a storage ring, the guide field is laid out with linear magnetic-optical elements (quadrupoles, bending magnets and drift spaces) only. At the next stage of design, non-linear elements such as sextupoles are introduced, but these elements are weak in the sense that the motions of particles in the guide field are determined almost entirely by the linear elements

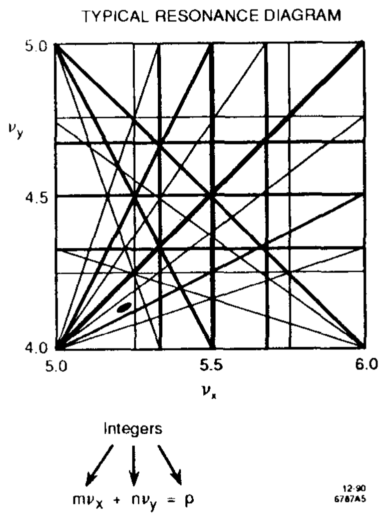

Figure 5: A typical resonance diagram showing some strong resonance lines. The lines are of the form $m\nu_x + n\nu_y = p$. The small island in the lower left-hand corner represents a beam distribution. It must fit between the lines!

and almost negligibly by the non-linear ones. Thus the beam dynamics of the storage ring are *almost* the beam dynamics of the linear lattice (the sequence of optical elements). Providentially the beam dynamics of the linear lattice are remarkably benign. If the ring were linear and perfectly constructed, all particle motions in it would be stable. That is a property of well-designed linear lattices. In general, the particles oscillate about the central orbit, executing the same number of oscillations each time they make a complete circuit. See Figure 4. The number of oscillations is called the betatron frequency or *tune* and is customarily denoted ν (in the US). There is one betatron frequency for x-motion (ν_x) and another, generally different, for the y-motion (ν_y). The motion is stable for any values of the ν's.

Real storage rings are not perfectly constructed; they have field inhomogeneities (non-linearities), misalignments, stray fields, *etc.* Indeed, in terms of linear beam dynamics, the sextupoles mentioned above constitute a deliberate introduction of such imperfections. The general effect of introducing these imperfections is to impose narrow bands in the $\nu_x\nu_y$-plane in which the particle motions are no longer stable. We call these regions resonances, because the unstable motion is caused by a resonant process. At values of the tunes lying in these bands, the phase shift of the (unperturbed) oscillation is such that the deleterious effects of the non-linear fields keep reinforcing each other from turn to turn. Figure 5 shows a portion of the $\nu_x\nu_y$-plane criss-crossed by a typical set of strong resonance lines. In a real storage ring the tunes must be kept away from these forbidden lines, or the beam particles will literally be lost; in other words, the tunes must lie securely in between the resonance lines.

Now we turn to the effect of the beam-beam interaction. We may think of it as equivalent to a highly non-linear lens acting on the particles. It shifts the frequencies of the oscillations, but more importantly it shifts them by amounts that depend strongly on their amplitudes. Thus different particles with different amplitudes acquire different frequencies, and a scatter plot of the frequencies of the particles in a bunch would show them to be smeared out on an island in the $\nu_x\nu_y$-plane. An example is shown in the lower left portion of Figure 5. Obviously the stronger the beam-beam interaction, the bigger the island, and if the island is big enough to lap over any of the resonance lines, particles will be lost. Therefore the size of the island is limited. The size of the island is measured by the two tune spreads, that in ν_x and that in ν_x. According to the resonance picture, they are both stringently bounded.

On the other hand, in all storage rings built to date, the beams at the IR's have been flat (as in CESR, for example), and in the case of a flat beam, we are concerned primarily with the vertical tune spread. While the horizontal tune shift is presumably limited, it proves comparatively easy to avoid any deleterious consequences of that fact by adjusting the details of the optical design. All of the current B-Factory designs call for flat beams too, so we shall not be far wrong to deal only with the vertical tune-shift limit, and that is what we shall do.

How do we estimate the tune spread? Let us consider first a particle undergoing a very large amplitude y oscillation, large that is, compared to the vertical size scale of the counter-moving beam. On the average (and speaking anthropomorphically), the particle scarcely feels the non-linear force due to the counter-moving beam, so its vertical tune is almost unshifted. In other words, it oscillates at its normal frequency. Now let us consider a small-amplitude oscillation, one in which the y-coordinate stays well within the size of the counter-moving beam. In that region of y, the beam-beam force is very nearly linear; it causes a tune shift that is independent of amplitude and one that is the largest of tune shifts for any amplitude. All other amplitudes give rise to tune shifts that lie between these two extremes, so the small-amplitude tune shift, which we shall call ξ_y, is a good estimate of the tune spread:

$$\xi_y^{\pm} = \frac{N^{\mp} r_e \beta_y^{\pm}}{2\pi\gamma^{\pm}\sigma_x\sigma_y(1+r)} \quad , \tag{1}$$

where

$\beta_y^{\pm}$ = IR beta functions (vertical)

N^{+} = positron bunch population

N^{-} = electron bunch population

r = beam aspect ration at IR

r_e = classical radius of electron

$\sigma_{x,y}$ = transverse r.m.s. bunch dimensions.

This, then, is the parameter that is limited by the beam-beam limit, and we shall refer to it as a tune-shift limit, although it is really a tune-spread limit. From the picture we have drawn of the limiting process, we would expect that the value ξ_y can reach in a particular storage ring should be extremely sensitive to the location of the working point in the $\nu_x\nu_y$-plane, and indeed it is. We would further expect that if measures were taken to weaken or eliminate resonance lines in the vicinity of the working point, the allowed tune shift would increase, and indeed it does. In all existing and past storage rings—and they all have had flat beams—the maximum allowable tune shift, after all ameliorating measures had been taken, has been

$$0.01 \le \xi_y \le 0.06 \quad .$$

The maximum allowable tune-shift is called the *incoherent beam-beam limit.* Certain machines are characterized by the lower values of this limit, while others are characterized by the higher values. In other words, the limit is, to a large extent, characteristic of a ring. On the other hand, we cannot claim to understand very well the reasons for these variations from machine to machine. Therefore, to base a new storage-ring design on a value of ξ_y as high as 0.06 would appear to be optimistic, while to base it on a value of 0.03 might be regarded as prudent.

The physical interpretation of Equation (1) is clear. The combination, $N/\sigma_x\sigma_y$, measures the peak current density of the bunch. The deflection will be proportional to that and inversely proportional to the particle's γ. The optical function, β_y, appears in the numerator, telling us that the bigger β_y is, the bigger, *i.e.*, the worse, is the tune shift. In other words, β_y is a sensitivity parameter. The reason for that lies in the role of the beta-function in describing the betatron oscillations: its square root is the "envelope function" that describes the variations of the size of the beam along the central orbit. Where β_y is large, the beam is large vertically, and where β_y is small, the beam is small. Since the extent of the beam in phase space is preserved (Liouville's Theorem), where the beam is small its angular divergence is large and *vice versa.* Where the angular divergence is large, an unwanted deflection such as the beam-beam interaction is less troublesome than it is where the divergence is small, so a small β_y at the IR is good.

Now let us proceed to the question, how does the beam-beam limit affect the achievable luminosity? The luminosity is given by the expression,

$$\mathcal{L} = \frac{N^+ N^- f_B}{4\pi\sigma_x\sigma_y} \tag{2}$$

where f_B is the bunch frequency defined earlier in the section on CESR. In writing this expression for the luminosity, we have restricted ourselves to the case in which the dimensions of the two colliding beams at the IR are matched:

$$\sigma^+_{x,y} = \sigma^-_{x,y} \quad .$$

It is not hard to see that this is the optimal case, and there is nothing to be gained by departing from it. If we solve Equation (1) for the bunch populations,

$$N^{\mp} = \frac{2\pi\gamma^{\pm}\sigma_x\sigma_y(1+r)\xi^{\pm}_y}{r_e\beta^{\pm}_y} \quad , \tag{3}$$

we can substitute them into Equation (2) and express the luminosity directly in terms of the tune shift.

$$\mathcal{L} = \frac{\pi(\gamma^{\pm}\xi^{\pm}_y)^2(f_B\sigma_x\sigma_y)(1+r)^2}{r_e\beta^{\pm}_y} \quad . \tag{4}$$

A storage ring system can have only one luminosity, so we are assuming that we have done the right thing: that we have somehow contrived to make the

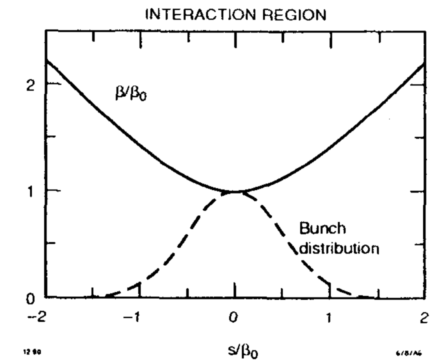

Figure 6: The longitudinal bunch distribution overlayed on a graph of the beta-function in the vicinity of the IR. In the case depicted, many particles lie in the wings of the beam waist.

two expressions on the right-hand side of this equation (represented by the two choices of the superscript sign) equal. In single-ring systems, this is accomplished more or less automatically.

Our task is to find ways of getting the highest luminosity we can. Aside from the constant π, the first factor in the numerator of Equation (4) involves the tune shifts and the energies; the tune shifts are bounded and the γ's are fixed by experimental requirements. The next factor is the combination $(f_B\sigma_x\sigma_y)$, and we shall return to that later. The final factor in the numerator is $(1+r)^2$ which suggests that making the beam round at the IR ($r = 1$) could gain us something in luminosity. Unfortunately, the correctness of Equation (1), and therefore of Equation (4), in the case of $r = 1$, has not been tested, and there is plenty of room for doubt. More importantly, the problems of suppressing background in the detector have forced the designers away from a round beam and toward a flat beam. We therefore take $r \ll 1$.

The denominator of Equation (4) urges us to make β_y as small as possible. That is, anyway, common storage ring design practice for single rings as well as double rings. How small is possible? We cannot make β_y smaller than the bunchlength; indeed for safety, it should be several times the bunchlength. The reason lies in the tune-shift formula, Equation (1). The interaction region is a place where the beam is tightly focused to a waist. That means that the β-function is parabolic there, and the functional form is simply

$$\beta_y(s) = \beta_0 \left[1 + \left(\frac{s}{\beta_0} \right)^2 \right] \quad ,$$

where s is measured from the center of the IR. The longitudinal distribution of a bunch is Gaussian, and what I have called the bunchlength is just the standard deviation, σ_s. Consider a particle riding one standard deviation behind the center of its bunch. See Figure 6. That particle passes the center of the counter-moving bunch a distance, $\beta_0/2$ beyond the center of the IR. At that point, β_y is substantially higher than it is at the center of the IR. The question is this: what value of β_y should we use in the tune-shift formula for that particle, or rather for the average particle? Obviously it is a bigger value than β_0 unless we restrict β_0 as follows,

$$\beta_y \geq \sigma_s \quad , \tag{5}$$

and so we do. The effect of this restriction, taken together with the realities of radio frequency system design, is to limit the usable values of β_y at the interaction region to a minimum of about 1 cm. In practice, given the optical complexities of the interaction region for a double-ring system, even that small a value may prove hard to achieve.

Let us return to the multiplicative factor, $(f_B \sigma_x \sigma_y)$, in the luminosity formula, Equation (4). That factor is the product of the beam cross sectional area at the IR and the bunch frequency, a sort of area per unit time. Suppose we choose to increase the cross sectional area greatly. In that case, Equation (3) tells us that we would also have to increase the bunch populations proportionally. In the next chapter, we shall learn that increasing the bunch population is a bad thing. That leaves us with f_B as our last remaining handle for increasing the luminosity.

Since the bunches move at the speed of light, increasing f_B is exactly equivalent to decreasing the distance between bunches in each beam. We saw in the example of CESR that the possibilities for doing that in a single ring are woefully limited. The fact is that if we want to use a lot of bunches, we must use two separate rings for the two beams. Even with two rings there are practical difficulties that prevent us from putting the bunches too close together, difficulties associated with separating the beams as they leave the interaction region after colliding. So in the end, the B-Factory designer is forced to put the bunches as close together as he can, given the IR design, and thus raise the bunch frequency as high as he practically can; and he is still hard pressed to achieve the luminosity required of a B-Factory.

We shall conclude our discussion of incoherent beam-beam constraints on the performance of B-Factories with a comment on their implications for the radiofrequency accelerating system. Replacing only one of the N's in Equation (2) with Equation (3), and noting that the dc beam current of one beam is $I = eNf_B$, we get

$$\mathcal{L} = \frac{(I/e)\xi_y \gamma}{2 r_e \beta_y} \quad ,$$

where the current, the tune shift, the energy, *etc.* refer consistently to one beam or the other. Once we have chosen the luminosity, there is very little latitude in the beam current, none in fact after we have minimized β_y. This formula applies to the high-energy beam in particular, and that is the beam that suffers the larger radiation loss, call it U_0. We see immediately that the radiated power from the high-energy beam,

$$P_{rad} = \left(\frac{I}{e}\right) U_0 \quad ,$$

will be proportional to the luminosity. We shall also see when we take a look at some B-Factory designs that P_{rad} is measured in megawatts.

3.1.2. Beam-Beam Coherent Instabilities

In the last section we dwelt on the effects of different particles moving differently—with different frequencies, for example—but how about the effects of different particles moving similarly? If we consider a cluster of particles close together in phase space as they approach the oncoming field of the opposing bunch, we shall see that despite the non-linearity of the field, they will respond more or less together; they will exhibit coherent motion.

We can think profitably about this kind of motion by simplifying our picture of the transverse inter-bunch force to a linear approximation. We can get away with that stratagem, because most of the particles in the bunch are concentrated near its axis, and we are concerned only with the motions of the center of mass (charge). In that approximation we have a lot of coupled oscillators, each bunch oscillating transversely in the guide-field restoring force and interacting with one or more opposing bunches at the interaction region. Such a system has the following properties:

1. There are as many modes as there are bunches.
2. The coupling is proportional to the bunch populations, and almost any mode will go unstable at sufficiently high coupling.
3. The most benign arrangement is to have equal numbers of bunches in the two rings, and in that case, these instabilities can be avoided at the required bunch populations by judicious choices of the tunes.

We can see how these instabilities arise by considering the simple case of equal-size rings. Each bunch of one ring interacts with only one bunch of the other ring. The normal modes of the bunch pair are the so-called Π- and O-modes. In the latter there is no inter-bunch restoring force, but in the Π-mode there is, so in the Π-mode we can think of the interaction as a gradient error, like a quadrupole magnet stuck into the lattice where it does not belong. If it is strong enough, a gradient error will drive a lattice into a nearby "stopband," which is beam-dynamical *argot* for instability.

We can make a more homely, but also more perspicuous, analogy to a pair of simple classical oscillators—masses on springs—coupled together by a third spring which has a negative spring constant. Negative springs are rare, but there is no difficulty in dealing with them mathematically, and the reader is invited to set up the equations of motion for the system. That system has the same two modes, Π and O. The coupling spring plays no role in the O-mode, because the separation of the masses remains constant, but it shifts the Π-mode frequency downward. If the coupling spring is strengthened more and more, eventually the restoring force supplied by the supporting springs is overcome, and there is no longer a restoring force. The response of the system to a Π-mode disturbance is an exponential flight from equilibrium.

The forces at play in these instabilities are very large, and feedback systems are of little avail against them. Therefore it is fortunate that we can avoid these instabilities by wise choices of tunes.

Returning to point 3 above, if we delved into the details of the coherent beam-beam instabilities, we would learn that rings of unequal circumference *can* be used without narrowing the allowable choices of tunes catastrophically provided their circumferences stand in the ratio of small integers (1/2 or 1/3, for example). Equal circumferences for the two rings is not a hard rule, just a conservative choice. So it comes down to a cost-benefit trade-off, and at this point we enter the treacherous area of tunnel economics. Despite some bitter experience in this field, I shall not utter any generalities on the subject, except to say that every case is different, depending on the economic health of the building industry, the weather and other unpredictables. Suffice it to say that a laboratory that already owns a suitable tunnel ought to take a different view of the matter than one that does not.

3.2. Single-Beam Instabilities

We turn now to beam instabilities of a different sort: instabilities which arise from the effects of the collective field of a bunch *on the particles of the same bunch or of the same beam.* The fields are the same ones we dealt with in the last section, but we must be more careful about the exact values of the electric and magnetic fields. The reason is that while the electric and magnetic forces on a counter-moving particle add, those on a co-moving particle nearly cancel, and the difference between them is very sensitive to the exact values of the fields.

As an example of this near cancellation, consider a highly relativistic bunch of particles travelling through space. The magnitudes of the magnetic and electric fields stand in the ratio, v/c, where v is the speed of the particles. The corresponding forces on a co-moving particle stand in the ratio, $(v/c)^2$, and since they oppose one another, the net force is proportional to $[1-(v/c)^2]=\gamma^{-2}$. In a storage ring, the presence of the metal vacuum chamber, and particularly of steps and pillboxes in its contours, upsets the delicate cancellation. As a result, the force on a co-moving particle is stronger than in the free-space example, and it is not proportional to γ^{-2}.

The dominant process for establishing these fields in storage rings is the excitation of local electromagnetic fields in the vicinity of abrupt changes in the vacuum envelope. A good example is a pillbox such as that formed by a radiofrequency accelerating cavity. It is a resonator, and the passage of a bunch along its axis naturally excites many of its normal modes. The fields are like the wake of a ship: they carry a disturbance long after the bunch has passed. Indeed we often call them wakefields. Each of the modes extracts some energy from the bunch and stores it as oscillating electromagnetic field energy. As the modal field rings, it dissipates its energy in the resonator walls, but it also acts on any particle that happens to pass. The lowest-frequency modes have periods that are long compared to the temporal bunchlength and decay times that are many periods long; they affect not only the particles in the exciting bunch but those in the following bunches. At the opposite extreme, the highest-frequency modes that are significantly excited have periods that are shorter than the temporal bunchlength and decay times that are comparable to it, so they affect only particles of the exciting bunch. If we are concerned with multibunch motions, we shall be interested in the lower frequencies, and if we are concerned with intra-bunch motions, we shall be interested in the higher frequencies. Impedance was invented to deal with cases like these.

In electrical circuits the impedance relates the voltage between a pair of terminals to the current flowing through it. It is a function of frequency and it applies to steady-state sinusoidal excitations. The apparatus of Fourier transformation enables us to combine these excitations to deal with complicated and even non-periodic currents. We shall take the same approach in our beam dynamical problems, although we shall not be hidebound about the dimensions of our impedances, and we shall make definitions that are appropriate to the problems.

Before tackling the single-beam instabilities, we had better get an idea of what range of frequencies we shall be dealing with: What does a typical bunch

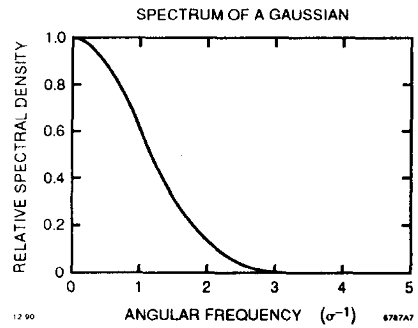

Figure 7: The spectrum of a Gaussian pulse. The ordinate is the angular frequency in units of the standard deviation (in time) of the pulse.

current spectrum look like? The bunch is a needle-like object. Typically its transverse dimensions, although they vary as the bunch moves along its orbit, are measured in fractions of a millimeter. Its length, which does not vary, is about 1 cm. The shape and size of the bunch then are much like that of a particularly long printer's dash, in fact just like this one: ——. Well, not the shape really, because the distribution of charge in each dimension is very nearly Gaussian, but the ratio of length to width is well represented. Since its spatial length is about 1 cm, its temporal length is about 30 ps. Viewed at a fixed location on the circumference of the ring, it is a Gaussian pulse of current. As Figure 7 reminds us, the spectrum of a Gaussian pulse contains no significant frequency components above the reciprocal of the r.m.s. pulselength, so the spectrum of the bunch reaches up to about 30 GHz. The lowest frequency present is f_B if all the bunches have the same populations, but that frequency is usually far below the lowest modal frequency of any vacuum-chamber discontinuity.

Now we know the spectrum of the driving current, but at what frequencies do we expect to find the resonances in the vacuum chamber? A good rule of thumb is that for the lowest-frequency resonance, the free-space wavelength (c/f) corresponding to the resonant frequency is approximately equal to a typical dimension of the cavity. For example, a rectangular box 30 cm on a side has its lowest resonance at 1.4 GHz which corresponds to a wavelength of 21 cm; and a pillbox with a diameter of 30 cm has its lowest resonance at 0.77 GHz which corresponds to a wavelength of 39 cm. In addition to these lowest modes there are many higher modes, infinite numbers of them in fact. Fortunately they are not all pernicious. Some do not couple to the beam; they don't have a longitudinal electric field on the orbit, for example. Others are damped by some mechanism that quickly dissipates or conducts away the stored energy. Nonetheless, many modes remain, and their effects add together to make the impedance function.

3.2.1. Longitudinal Microwave Instability

As our first example of single-beam instabilities, we treat the longitudinal microwave instability, a process driven by the highest frequencies present in the wakefields, frequencies that correspond to free-space wavelengths that are shorter than the bunch, *i.e.*, millimeter waves. The bunch spectrum is weak in that region of frequencies, as we have seen; nevertheless these components cannot be ignored. The electromagnetic modes in question have longitudinal electric fields on the central orbit through which the bunch excite the modes and through which the modes act back on the particles of the bunch. Different portions of the bunch are accelerated and decelerated. The subsequent longitudinal motion of the particles within the bunch redistributes the current in the bunch and therefore its high-frequency spectrum which, in turn, changes the amplitude of the mode—feedback. There are also influences present, such as non-linearities in the motion, that damp the modes, so there is a threshold current for the longitudinal microwave instability. If the *peak* current exceeds the threshold, many modes typically go unstable, and the resulting longitudinal motion is turbulent, lengthening the bunch and increasing the energy spread among the particles in the bunch.

We cannot tolerate bunch lengthening in B-Factories for reasons we have discussed in Chapter 3. According to Equation (5), we cannot make β_y at the IR smaller than the bunchlength, but according to Equation (4), we must make it as small as we can. Therefore, we adjust β_y carefully just larger than the bunchlength. If then the bunch lengthens, our efforts are vitiated and the luminosity drops. We cannot accept extra energy spread either, because the Υ resonance is narrow. The conclusion is that we must avoid the longitudinal microwave instability.

The impedance we use to describe this kind of beam-chamber interaction is called the longitudinal impedance, $Z_{\|}(\omega)$. It is the ratio of the longitudinal electric field on the axis (multiplied by the circumference of the orbit to create a voltage) to the current driving it (assumed sinusoidal). It is just the sum of all the local impedances around the orbit. If $\mathcal{E}$ is the sum of the electric fields,

$$Z_{\|}(\omega) = \frac{\mathcal{E}L}{I} \quad .$$

The threshold current for the longitudinal microwave instability, $I_{\|}$, is

$$I_{\|} = \frac{2\pi(E/e)\alpha\sigma_\delta^2}{|Z_{\|}(n\omega_0)/n|} \quad ,$$

where α is the momentum compaction factor of the ring, σ_δ is the fractional energy spread in the beam and ω_0 is the angular orbital frequency. The number, n, is called the mode number; it counts the number of wavelengths of the field contained on the circumference, L, and according to what we have said it is a very high number indeed for the microwave instability. The bunch is unstable if the peak bunch current exceeds $I_{\|}$, *i.e.*, if

$$\frac{eNc}{\sigma_s} > I_{\|} \qquad \text{unstable}$$

where σ_s is the r.m.s. bunchlength. The left-hand member is just an estimate of the peak bunch current.

An interesting and important fact is that the quantity, $Z_{\|}(n\omega_0)/n$, is nearly independent of n for the n's of interest in the vacuum chambers of storage rings like CESR. It is a constant, although one that differs somewhat from machine to machine, depending on the care with which the vacuum chamber was designed.

How can we combat this instability? After we have lowered the impedance as far as it is feasible to do in a practical vacuum chamber, can we do anything else—feedback for instance? Unfortunately the very high frequencies of the modes are beyond the reach of current broad-band feedback practice.

Fortunately, we have a way to build B-Factories without confronting this instability: we can increase the number of bunches without increasing the bunch current. We have been building storage rings for a couple of decades, and we have learned how low we can keep the impedance in a practical vacuum chamber and therefore how high a peak current we can hope to store. If we do not ask for higher bunch currents than those in today's rings, we shall be on sound ground, and that is the reason referred to in the preceding chapter for choosing to increase the total beam current by adding many bunches rather than trying to increase the populations of the individual bunches.

3.2.2. Transverse Mode-Coupling Instability

The transverse mode-coupling instability is, as its name suggests, a transverse instability. The unstable motion is betatron motion as opposed to the longitudinal motion of the microwave instability, but the cause of instability is the same. The bunch, once set into motion by noise, excites wakefields which in turn act back on the particles in the bunch and amplify the motion—another feedback mechanism. This instability is sometimes called the "fast head-tail" instability, because the wakefield of the head of the bunch drives the tail of the bunch and the resulting growth is very fast as instability growth rates go. Longitudinal motion, even though it is not driven unstable, plays a crucial role in the transverse mode-coupling instability. Since wakes can only follow their sources, not lead them, the head *always* drives the tail and the tail *never* drives the head. Therefore if the particles in the head at a particular time remained there forever, and if the same were true of the particles in the tail, there would be no feedback, but they don't. The particles of the bunch constantly circulate longitudinally from head to tail and back. Thus head particles influence the motion of tail particles which then become head particles and influence the original head particles which have become tail particles. What?

The characteristic wavelengths of the electromagnetic field modes that drive the transverse mode-coupling instability range from a few bunchlengths, *i.e.*, a few centimeters, to the longest modal wavelengths that exist in the chamber. The modes have transverse magnetic fields that are roughly constant over the transverse dimensions of the bunch and longitudinal electric fields that are zero on the axis and vary linearly with transverse position. Such a mode is shown in Figure 8. The mode is excited in proportion to the deviation of the center of charge of the head, and the tail particles are deviated by the magnetic field.

The most useful impedance-like function for such modes is a spatial derivative of an impedance. For the mode shown in Figure 8, the appropriate definition is

$$Z'_\perp = \frac{d}{dy}\left(\frac{\mathcal{E}L}{I}\right) \quad .$$

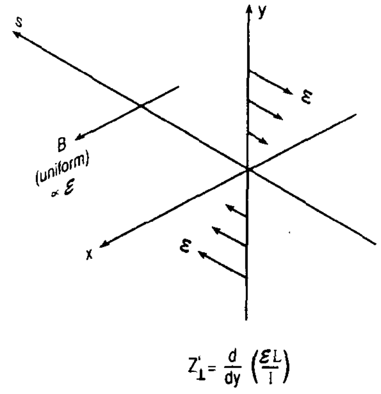

Figure 8: Fields of a mode that drives the transverse mode-coupling instability. Near the axis, the magnetic field is transverse and nearly constant and the electric field is longitudinal and linearly dependent on a transverse coordinate.

We define a threshold current for the transverse mode-coupling instability.

$$I_\perp = 4\pi \frac{(E/e)\,\nu_s\,(\sigma_s/R)}{\langle\beta\rangle\ \mathrm{Im}\left[Z'_{\perp eff}\right]} \quad ,$$

where ν_s is the longitudinal oscillation frequency in units of the orbital frequency and $\langle\beta\rangle$ denotes the average value of the appropriate beta-function, horizontal or vertical, around the whole ring. The "radius," R, is just the circumference of the ring divided by two pi. The impedance in the denominator, $Z'_{\perp\ eff}$, is an average over the impedance function from the lowest mode up to those whose wavelengths correspond to the bunchlength. To avoid this instability the bunch current should not exceed $I_\perp$. Again the condition for stability is

$$\frac{eNc}{\sigma_s} \leq I_\perp \qquad \text{stable} \quad .$$

Like the microwave instability, the mode-coupling instability admonishes us to avoid increasing the bunch population. It is also beyond the reach of today's feedback technology.

3.2.3. Coupled-Bunch Instabilities

The microwave instability is mediated by millimeter-wavelength fields. The transverse mode-coupling instability is driven by the whole spectrum of fields from long wavelengths to centimeter wavelengths. Because they can propagate in the vacuum chamber, millimeter- and centimeter-wavelength modes die out very rapidly. They are dead before the next bunch arrives. But what of the longer-wavelength, lower-frequency modes? Such modes will exist in the rf cavities, for example, and since they cannot propagate in the vacuum chamber, they are trapped. Consequently they can have high Q-values and ring for a long time. We know that they act within the bunch, contributing to the mode-coupling instability, but they must also act from one bunch to the next and *couple the coherent motions of the different bunches.* The strengths of these modes determine the low-frequency values of the impedance functions, and if those values are too high, coupled-bunch instabilities will destroy the beam. These modes have not troubled today's single rings, because the bunches in those machines are so far apart in time that even the these high-Q modes decay away between bunches, but they are extremely dangerous to B-Factories.

Just as the intra-bunch instabilities occur in two varieties, longitudinal and transverse, so do the inter-bunch instabilities. Furthermore they are driven by electromagnetic modes that have the same characters as those that drive the intra-bunch instabilities in every respect save frequency. The difference in frequency, however, is of paramount importance. It opens the door to feedback systems to control these instabilities.

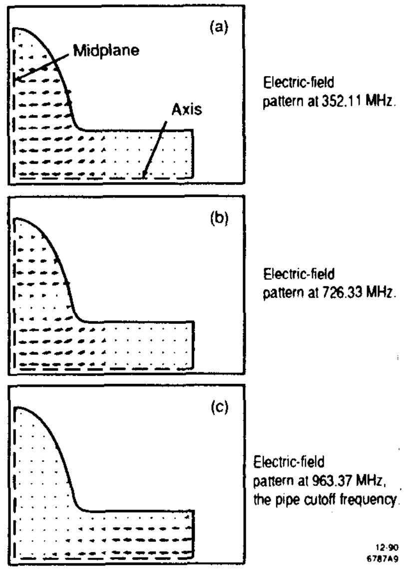

Figure 9: One quadrant of an rf cavity with only two trapped modes. The mode at 352 MHz is the accelerating mode. The mode at 963 MHz propagates down the pipe. Only the mode at 726 MHz remains to create unwanted parasitic impedance.

Another feature of these modes is that they can exist only in fairly large boxes—remember the relationship between dimensions and low-lying modes—and the mode spectrum is not very dense. In other words, we know where to look for the offending parts of the chamber (mainly the rf cavities) and we know that only a few modes will plague us. Consequently, we can hope to "swamp" these modes, to damp them selectively. Research and development on highly damped cavities that display only one beam-interactive mode is in progress in several centers.

How do we do this? Rf cavities, after all, are designed to present a high impedance to the rf generator, *e.g.,* the klystron, so that a limited amount of rf power can produce the highest possible voltage. A cavity's fundamental (accelerating) mode must be a low-loss mode. But its higher modes are obnoxious;

they drive beam instabilities. Thus we need to learn to build cavities that, practically speaking, have only one high-Q mode. That seems a tall order, but we are learning to do it by devising cavity shapes in which only the accelerating mode is "trapped." The higher modes can propagate out through the beam pipes and consequently do not store field energy locally. Figure 9 shows one quadrant of a spheroid cavity with only a single parasitic energy-trapping mode.

In spite of all the impedance-suppressing measures we can take, multibunch instabilities will still arise, and we must combat them with wide-band beam feedback systems. These systems work by sensing beam motion and deflecting the individual bunches to damp the motion. Such feedback systems are in common use in accelerators and storage rings, but B-Factories will require particularly powerful ones. The bandwidth called for is half the bunch frequency, $f_B/2$, and the power is measured in kilowatts. In the SLAC/LBL design, the figures are 90 MHz and 2.6 kW.

Between the two countermeasures, feedback and mode damping, it appears quite feasible to keep coupled-bunch instabilities under control in B-Factories.

3.2.4. Ion-Beam Instability

While a positron beam does not collect ions, an electron beam does, and if the ions produced by beam-gas collisions are allowed to collect in the beam, their fields will drive the beam unstable. We can see how this phenomenon works in a qualitative way by observing that a swarm of nearly stationary ions sitting in the volume of the electron beam acts like a continuous focusing lens—focusing because the ions attract the electrons. As the ions collect and their density builds up, the betatron tune of the beam shifts higher and higher until it approaches a strong resonance. As it approaches the resonance, the electron beam develops scallops that grow larger and larger, and it becomes useless for colliding beams.

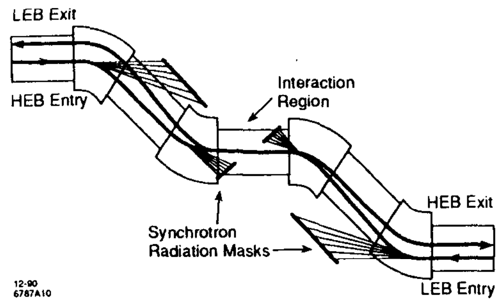

Figure 10: Horizontal S-bend interaction region design. The two beams enter and exit along separate lines; they follow a common line through the collision point at the center where they collide head-on.

Mercifully a bunched beam can be made to drive out ions rather than collect them, at least in linear-approximation theory. We have some experience with this process in storage rings devoted to producing synchrotron radiation, but we cannot claim at this time to understand everything we have observed. Understanding these phenomena is one of our challenges.

4. The Interaction Region

In single-ring colliding-beam storage rings, head-on collisions are automatic if no transverse electric fields are introduced in the vacuum chamber. Furthermore the optics for the two beams are the same. These things are not true for beams in two separate rings. When the beams are stored in different rings, a special transport system must be devised to carry the two beams from their respective rings to the collision point and to deliver them back into their respective rings and to trim their optics so that the beams are properly matched at the collision point. In cases where the two beam energies are the same, we have to use electric fields to combine and separate the beams, but in a B-Factory the beams have different energies and that circumstance actually helps us. Since particles of different energies follow different orbits in magnets, the combination and separation can, in principle, be accomplished with magnets alone. Several schemes have been developed, and one of them is shown in Figure 10 in which horizontal bending exclusively is used to bring the beams together and separate them.

All schemes in which the paths of the two beams are *colinear at the interaction point* share a common disadvantage. The paths cannot begin to diverge until they enter the first bending magnet, a feature that acts to the detriment of using lots of bunches. The point is related to the "distant collisions" of counter-moving bunches at other points than the interaction point, encounters in which the centers of charge are well separated laterally so that each bunch feels the "$1/r$" field of the other. Even a "$1/r$" field is non-linear, so while we want them to collide at the IP, we do not want the bunches to feel each other's fields anywhere else, because any other encounters, would be disruptive and would probably lower the attainable luminosity.

To understand this matter more clearly, let us follow a bunch out from the IP after the main, luminosity-producing collision. The bunches in both rings are separated by some common bunch separation, let's call it S_B, equal to the circumference divided by the bunch number. After our sample bunch has progressed by $S_B/2$, it passes the next incoming bunch; when it reaches S_B, it passes the following incoming bunch; and so on. Ideally we would like to shield the bunches from each other in these encounters by having them in separate metal pipes, but we cannot do it until the two orbits have separated enough to insert a metal wall between them. In this way schemes in which the paths of the two beams are colinear at the interaction point are disadvantageous.

From this point of view, a scheme in which the two orbits cross at an angle would be preferable. In that case, the orbits begin separating immediately as they leave the IP. Unfortunately, there is a catch. Since the needle-like bunches

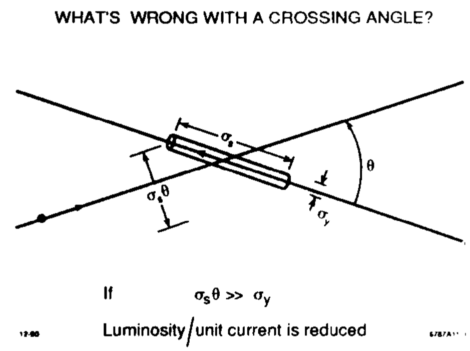

Figure 11: Collisions with a crossing angle. Think of one bunch as the target for the other one. The projected area of the target bunch is increased by the crossing angle, so the target density is reduced.

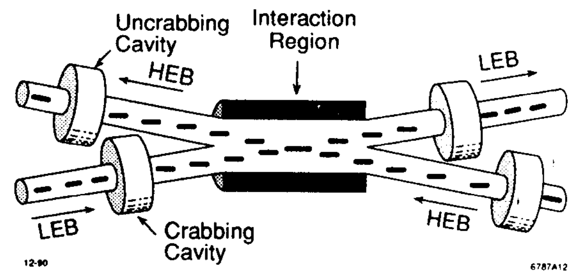

Figure 12: A diagram of crab crossing which shows how, by tilting the bunches (by half the crossing angle), crossing is made to appear "head-on." To the particles of one bunch, the other bunch seems to have the same target density it would have in a head-on collision, and no luminosity is sacrificed. After being tilted on one side of the IR, the bunches need to be un-tilted on the other side. Powerful rf cavities are needed to do that.

are aligned along the orbits, and the orbits cross each other at an angle, the bunches plough through each other at an angle, and the luminosity is degraded. Figure 11 shows how it happens.

A remedy for this defect has been proposed: it is called *crab crossing*. In the crab-crossing scheme, illustrated schematically in Figure 12, the orbits cross at an angle, but the bunches are tilted relative to the their orbits on one side of the IR and un-tilted on the other side, so that, in effect, they collide head-on, and no luminosity is lost. If one observes the collision from a frame of reference moving upward through the IP, the collision *is* head-on.

The crab-crossing scheme has the merit that it eases the problems of separating the bunches of the opposing beams, but it carries with it some drawbacks too. The chief drawback is that powerful rf cavities must be installed on either side of the interaction region to accomplish the bunch-tilting; these cavities bring with them additional impedances to aggravate instabilities.

5. Selected *B*-Factory Designs

As an epilogue, we offer a sampler of current *B*-Factory designs. The art of designing *B*-Factories is comparatively young—less than five years old—and naturally it is developing and becoming more sophisticated. Therefore, we can expect that the *B*-Factory designs of the future will differ in some ways from those of today. Still the *B*-Factory designs we have in hand today are remarkably alike in many respects, which connotes a certain degree of maturity. Table 2, Table 3, and Table 4 present some salient characteristics of three current designs: the SLAC/LBL design, the Cornell design for a future asymmetric system and a design carried out cooperatively by CERN and PSI, Darmstadt.

Table 1.

Beam energy, E (GeV)	5.3
Max. luminosity, $\mathcal{L}$ (cm^{-2} s^{-1})	1×10^{32}
Max. tune shift, ξ_y	0.017
Radiated power/beam (MW)	0.08
Beam current, I (A)	0.077
Number of bunches, k_B	7
Bunch frequency, f_B (MHz)	2.7
I.R. beta-function, β_y (cm)	1.5
Beam aspect ratio, r	$\ll 1$
Circumference (m)	768
Long. impedance, $\lvert Z_{\parallel}(n\omega_0)/n\rvert$ (Ohms)	1

Table 1: Parameters of Cornell's CESR storage ring in service as a *B*-Factory.

Table 2.

Beam energy, E (GeV)	9.0/3.1
Max. luminosity, $\mathcal{L}$ (cm^{-2} s^{-1})	3×10^{33}
Max. tune shift, ξ_y	0.03
Radiated power/beam (MW)	5.5/2.7
Beam current, I (A)	1.48/2.14
Number of bunches, k_B	1746/1746
Bunch frequency, f_B (MHz)	238
I.R. beta-function, β_y (cm)	3.0/1.5
Beam aspect ratio, r	$\ll 1$
Circumference (m)	2200
Long. impedance, $\lvert Z_{\parallel}(n\omega_0)/n\rvert$ (Ohms)	1.5

Table 2: The SLAC/LBL (PEP-based) *B*-Factory design.

Table 3.

Beam energy, E (GeV)	8.0/3.5
Max. luminosity, $\mathcal{L}$ (cm^{-2} s^{-1})	3×10^{33}
Max. tune shift, ξ_y	0.03
Radiated power/beam (MW)	3.2/0.3
Beam current, I (A)	0.6/1.4
Number of bunches, k_B	210/210
Bunch frequency, f_B (MHz)	82
I.R. beta-function, β_y (cm)	1.0/1.0
Beam aspect ratio, r	$\ll 1$
Circumference (m)	768
Long. impedance, $\lvert Z_{\parallel}(n\omega_0)/n \rvert$ (Ohms)	1.8/0.14

Table 3: The Cornell Asymmetric B-Factory design of May 1990.

Table 4.

Beam energy, E (GeV)	8.0/3.5
Max. luminosity, $\mathcal{L}$ (cm^{-2} s^{-1})	1×10^{33}
Max. tune shift, ξ_y	0.03
Radiated power/beam (MW)	3.1/0.4
Beam current, I (A)	0.6/1.3
Number of bunches, k_B	80/80
Bunch frequency, f_B (MHz)	25
I.R. beta-function, β_y (cm)	3.0/3.0
Beam aspect ratio, r	$\ll 1$
Circumference (m)	963
Long. impedance, $\lvert Z_{\parallel}(n\omega_0)/n \rvert$ (Ohms)	0.3

Table 4: The CERN/PSI B-Factory of March 1990.