Collective Instabilities in Accelerators

Alex Chao 趙午
OCPA School 2010

1. Introduction

Design of an accelerator 加速器的設計

To design an accelerator, one first considers the motion of a single charged particle in the environment of magnets and RF cavities. The motion of this single particle in this environment must be stable.

For example, in a circular accelerator, the particle must stay inside of the accelerator vacuum chamber for many many revolutions, typically $\gg 10^{10}$ revolutions --- and much more than that of the lifetime of earth around the sun!

This is of course not an easy task. Accelerator physicists design accelerators with three basic elements:

Element	Function	Field	Focusing
Dipoles	Guide particle trajectory	Magnetic	weak focusing in x
Quadrupoles	Confine particle motion near the design trajectory	Magnetic	x,y
RF cavities	Keep particle energy near the design energy	Electric	z

All these are just to make sure that single-particle motion is stable. With these three elements arranged, the basic layout of an accelerator is determined.

Having provided a design trajectory, and made sure that there are focusing in x,y, and z, there seems to be nothing left to do. But that is not true. We still have to examine the stability of the beam particles in much more detail.

Single-particle stability 單粒子穩定性

This is one very important area of accelerator physics, i.e. *single-particle nonlinear dynamics* (李世元，郭錦城).

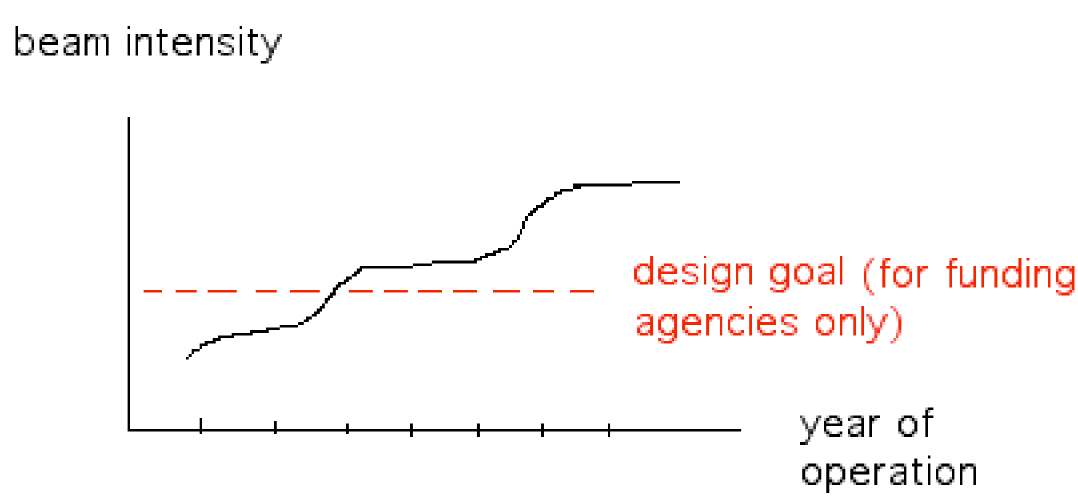
Multi-particle stability 多粒子穩定性

But there is a second significant part of accelerator physics. It is called *multi-particle collective beam instability* effects, sometimes also called collective beam instabilities, coherent beam instabilities, beam instabilities, or simply instabilities. 我們的主題。

After an accelerator is built, it is clear that the users will continue to ask for stronger and stronger beams. However, when the beam intensity is increased, the beam is bound to become unstable at some point. Beyond that point, the beam becomes unstable. So the first instinct is to operate the accelerator below that instability limit and never to exceed it.

But that is not the right way to think. What we actually do is to try to understand this instability mechanism (accelerator physics!), and cure it. After curing it, the beam intensity can be increased again.

But what happens is that then another instability mechanism will take over, limiting the beam intensity at some higher limit. We then understand and cure it again. This process repeats. So the net beam intensity ends up showing a behavior like this:



Note that collective instabilities are bound to occur. There is no way that an accelerator does not need to address questions on collective instabilities. This is because we will always try to push the beam intensity to its maximum. And at that maximum, by definition, there is a limiting instability. The study of collective instabilities cannot be avoided. 集體不穩定性的研究在任何加速器中都是不可避免的。

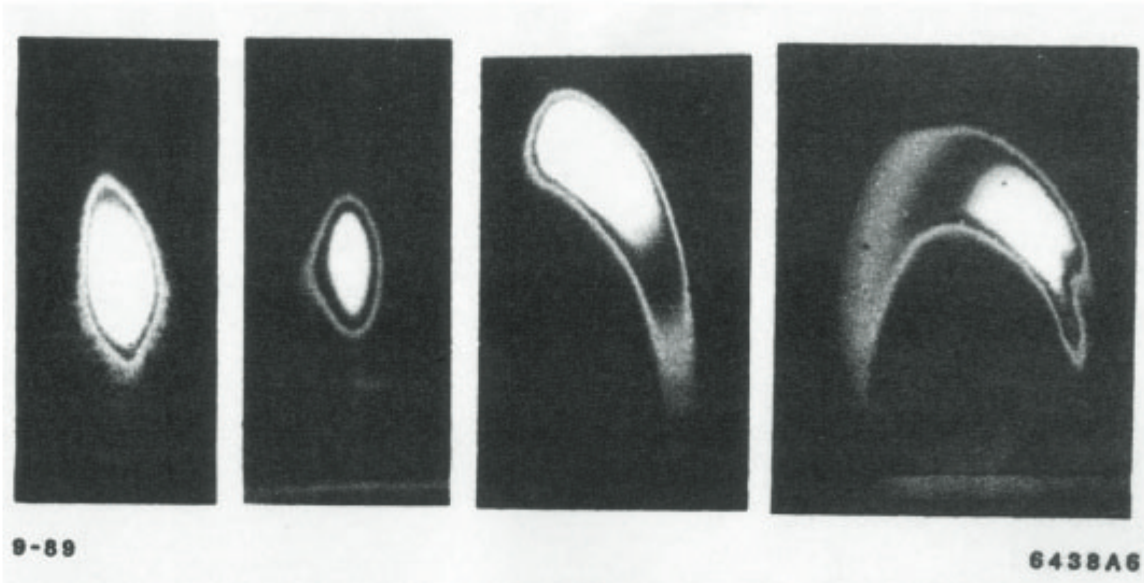
A side remark to be made here: Building an accelerator is not building a bridge 造加速器不是造橋鋪路. After a bridge is built, the job is done. The bridge then stays there the same way as it was built until one day it is demolished. In contrast, an accelerator is a living object. After it is built, an accelerator is constantly being improved. The concept of “design goal” of an accelerator applies only for the government funding agencies. A real accelerator always continues to grow, reaching the original design goal, and exceeding it. The “design goal” means nothing to an accelerator physicist, as illustrated in the above figure. An accelerator physicist must keep the right mind-set that his/her job is not done simply because an accelerator has been built or when its design goal has been met – an accelerator is not a bridge.

Note also that there is not just *one* instability. There are actually many types of instabilities. The beam can be stable or unstable depending on which instability mechanism one is talking about. For some instability mechanisms, the beam is always unstable but is fortunately stabilized by some stabilizing mechanisms. For some other instability mechanisms, the beam is stable below a certain intensity threshold, while unstable above it. Each instability occurs under some different conditions and at different beam intensities. But there are many of them, occurring at higher and higher intensities. So one would actually encounter a whole list of all kinds of possible instabilities. After encountering one instability mechanism and curing it, you will meet the next one.

Over the years, accelerator physicists have observed, explained, and (mostly) cured several intricate instability mechanisms:

negative mass instability 1959
resistive wall instability 1960
beam break up instability 1966
head-tail instability 1969
microwave instability 1969
beam-beam limit in colliders 1971
potential well distortion 1971
anomalous bunch lengthening 1974
transverse mode coupling instability 1980
coherent synchrotron radiation instability 1990
sawtooth instability 1993
electron beam-ion instability 1996
electron cloud instability 1997
microbunching instability 2005

Roughly, that is one instability mechanism every ~3 years. An example of beam break-up instability in an electron linac is shown below:



When the beam is well steered down the long 3-km linac at SLAC, the beam profile at the end looks nicely packed. When the beam trajectory is slightly mis-steered, the beam tail gets perturbed by the “wakefields” generated by the beam head. This instability, once understood, was cured by precision alignment of the linac, plus several trajectory control feedback systems.

History shows new instabilities are discovered as we overcome the older instabilities and push for higher and higher beam intensities. The latest are found in the most advanced accelerators, for example in factory-class collider storage rings and free electron laser linear accelerators. The process is still continuing as new accelerator concepts are invented, and as older accelerator concepts were pushed for higher performances.

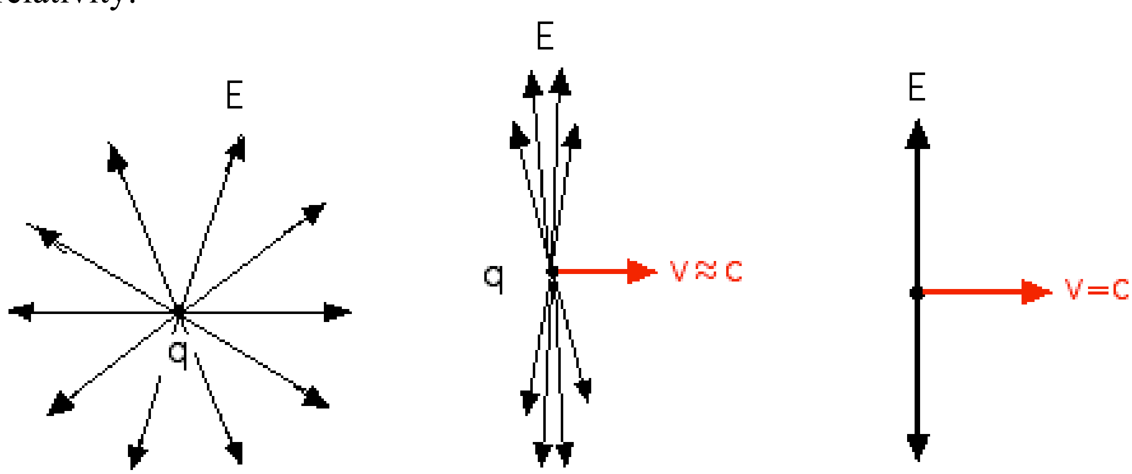
2. Wakefields and Impedances

A charged particle always carries electromagnetic fields

Collective instabilities mostly come from an intense beam interacting with its vacuum chamber environment in an accelerator. How does the interaction take place? One first has to know that each charged particle always is attached with it some electric field lines. You can distort these field lines but you can never cut them loose from the charge under any circumstances. This is called *Gauss's law*. (*) 奇妙的高斯定律

> (*) Gauss law is an amazing law. Mathematically, it reads $\nabla \bullet E = \rho/\varepsilon$. Physically it reads: Electric field lines are absolutely attached to the charges, no matter how violently you shake the particle trying to shake off its field lines.

If the charge is stationary and if it is in a free space, its field lines radiate radially outwards isotropically, as in fig.(a) below. For a moving charge, we see fig.(b). When v approaches c, the field lines get contracted into a thin "pancake". The pancake is attached to the particle and moves with it with velocity v. When we take the ultrarelativistic limit v=c, then the pancake reduces to infinitely thin sheet, as shown in fig.(c). All the electric fields stay in an infinitely thin sheet. This contraction is the result of theory of relativity.



(a) stationary charge (b) a charge moving relativistically (c) when v=c

Since in accelerators, the particles typically move with speed very close to c, we will now think of the picture in (c). In the TPS (Taiwan Photon Source), for example, electrons will move with v=0.99999999 c.

In addition to the electric field as shown in the above figure, when the particle is moving, it also generates a magnetic field. This magnetic field has the same distribution as the electric field, i.e. it contracts to a thin pancake when the particle's velocity approaches c, and into an infinitely thin sheet when v=c. Direction of the electric field is radial; direction of the magnetic field is azimuthal (right-hand rule).

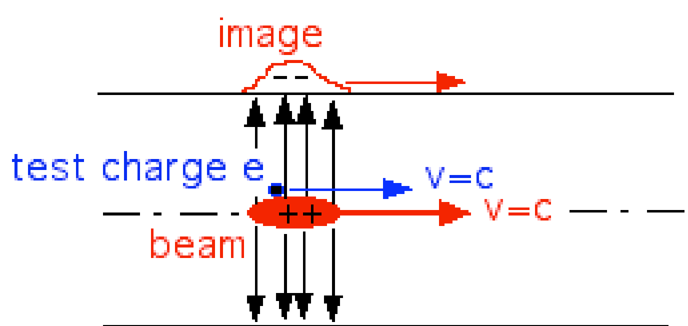
However, the magnetic field differs from the electric field on one important point. When v=0 as in fig.(a) above, there is electric field, but no magnetic field. When v increases, the magnitude of the magnetic field increases, but still weaker than the electric field. Only when v=c, the magnetic field increases to become the same magnitude as the electric field. The fact that the magnitudes of the electric and magnetic fields are equal when v=c has important consequences, as will be discussed later.

The vacuum chamber 真空盒

So far we have discussed a particle in free space. We now need to add the vacuum chamber.

Consider a very smooth cylindrical beam pipe. (How smooth does it have to be? It has to be so smooth that one small 1-mm discontinuity on the pipe is going to be a big deal. In some circumstances, even the very small 1-μm roughness on the wall surface can have a significant effect!) For now, let us also consider the smooth pipe wall to be perfectly conducting, i.e. no resistivity.

The ultrarelativistic beam going down the axis of the pipe, together with its electromagnetic field and the vacuum chamber look like this:



This is a complicated arrangement. Note first that the electromagnetic fields are perfectly terminated on the pipe wall. No fields penetrate into the wall because it is perfect conductor. The image charge on the wall is exactly equal and opposite to that of the beam, and it moves also with v=c in the same forward direction. As the beam moves forward, the entire field pattern moves with it. In particular, there are no electromagnetic fields left behind this pattern.

Now remember what we want to do: We want to examine what effect does the electromagnetic field carried by the beam has on the particles in the beam. So, let us now consider a particular particle in the beam, the blue charge e in the above figure, which of course moves with v=c with the beam. We call this particle the "test particle". This test particle will see an electric force eE due to the electric field of the beam. This force is easily seen to push e towards the vacuum chamber wall because the test charge e has the same sign as the charges of the beam.

But there is also a magnetic force. The magnetic field is in the azimuthal direction (right hand rule). The magnetic force is (e/c) v x B. It is easily seen that this magnetic force is pointing towards the pipe axis.

We mentioned that when v=c, the magnitude of E and magnitude of B are equal. In the ultrarelativistic limit, therefore, the electric and the magnetic forces exactly cancel! The particles in the ultraralativistic beam do see electric force and magnetic force, but they do not see a net force because they exactly cancel each other. The collective electromagnetic fields carried by the beam do not influence particle motion. If you think about this for a minute, it tells you that there can not be any collective instabilities!

Let us make a conclusion by stating the following theorem 定理:
There are no collective instabilities if the following conditions are all fulfilled: (a) the beam is ultrarelativistic, (b) the vacuum chamber is smooth, (c) the vacuum chamber wall is perfectly conducting.

Note that the pipe has to be smooth only in the z-direction. The pipe's cross section in the transverse plane can have any arbitrary shape, and the theorem remains valid.

Putting the same theorem in another way, we may say that there are three possible ways 三種情況 for a collective instability to occur:

(a) the beam is not relativistic enough,
(b) the vacuum chamber is not smooth enough,
(c) the vacuum chamber is too resistive.

If any one of these conditions occurs, the exact cancellation of the electric and magnetic forces is lost, and the beam can encounter an instability.

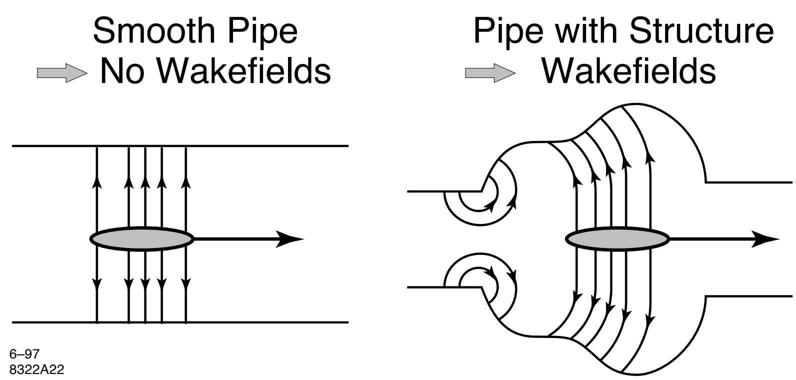
In reality, we try to avoid these three conditions as much as possible. In fact, we generally do such a good job in the design and the construction of accelerators that the electric and magnetic forces generally get to cancel to a high accuracy. This cancellation is a key to accelerators. Without it, basically no accelerators would work! 萬幸！

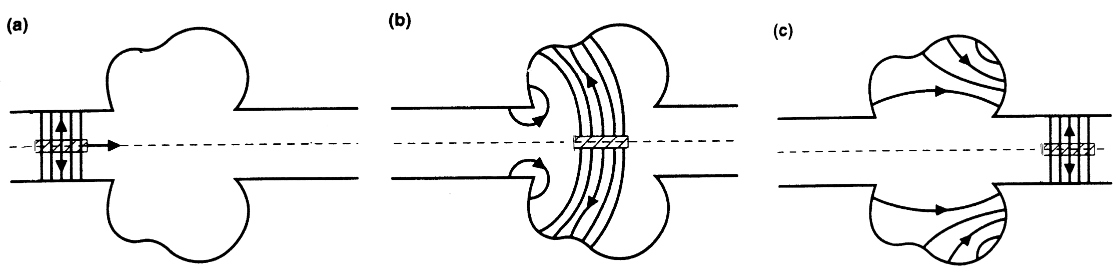
However, the cancellation is never perfect. The vacuum chamber is generally made of copper or aluminum, which are good conductors but not perfect. There will be many small necessary discontinuities along the vacuum chamber pipe, such as beam position monitors, vacuum pumping ports, etc. There are also those very *big* discontinuities known as RF cavities. As to the condition of v=c, even the TPS fails to satisfy it completely. So the cancellation of electric and magnetic forces are not perfect. And that leads to collective instabilities.

Wakefields 尾场due to discontinuities

When a charged particle beam traverses a discontinuity in the conducting vacuum chamber, an electromagnetic “wakefield” is generated. An intense beam will generate a strong wakefield. When the wakefield is strong enough, the beam becomes unstable.

Wakefields are generated by beam-structure interaction:





The reason a wakefield is generated when there is a discontinuity is because the image charges moving along the pipe now have to move around a corner. We all know that when a charge is bent in its trajectory, it radiates. Wakefields are really the radiation fields of the image charges when their trajectories are bent.

Once you learn that these wakefields are a result of radiation, it is natural to ask what frequencies are these radiation? What is the frequency content of these wakefields? The answer is that it depends on the details of the beam and the detailed geometry of the discontinuity. In general, it covers a very wide range, from micron wavelengths to long microwaves. To describe the frequency content of these wakefields, we will later introduce a quantity called impedance. Impedance is essentially the Fourier transform of wakefield.

Wakefield due to resistive wall

When the vacuum chamber is smooth but is resistive, there are also some wakefields generated. In the case of an ultrarelativistic point charge q going down the axis of a circular pipe with resistive wall, the wakefield looks like:

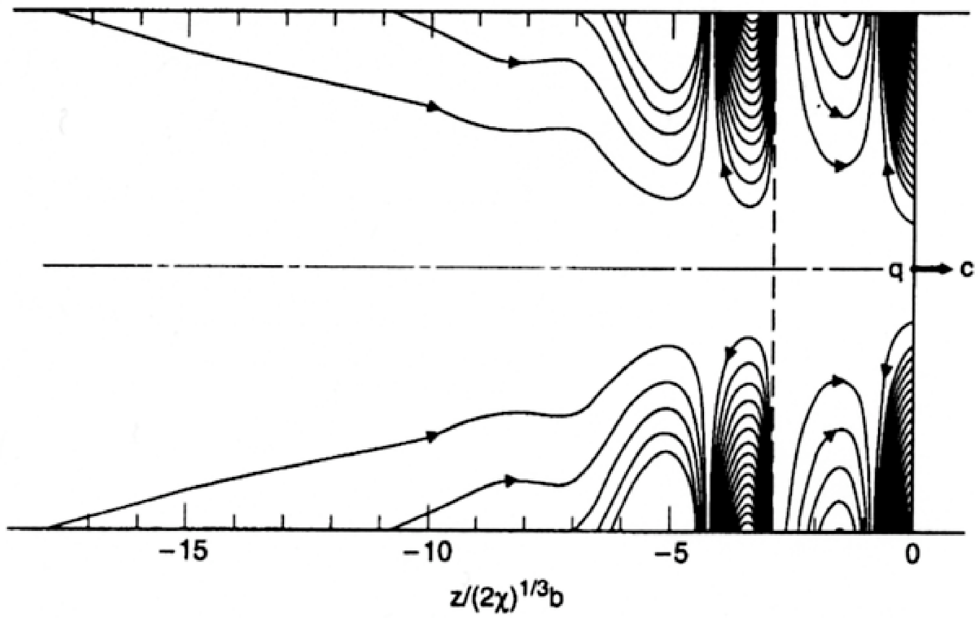
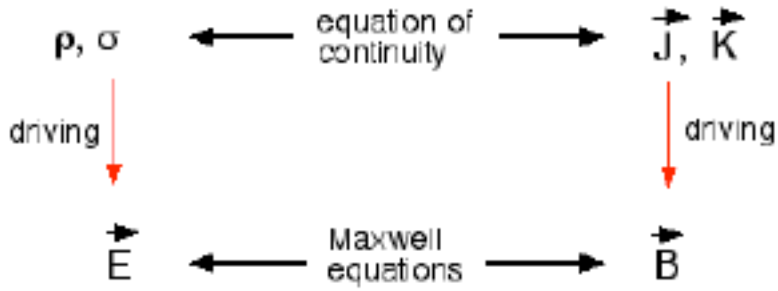


Figure 2.3. Wake electric field lines in a resistive wall pipe generated by a point charge q. The field pattern shows oscillatory behavior in the region $|z| \lesssim 5(2\chi)^{1/3}b$ (or $|z| \lesssim 0.35$ mm for an aluminum pipe with $b = 5$ cm). The field line density to the left of the dashed line has been magnified by a factor of 40. (Courtesy Karl Bane, 1991.)

Let us first review Maxwell equations below 電磁學綱要:



Definition of metal: $\rho = 0, \ \vec{J} = \sigma\vec{E}$

Definition of insulator: $\vec{J} = \vec{0}, \ \rho = \epsilon\nabla \cdot \vec{E}$

From Maxwell equations, we learned that

- electric field is driven by charge
- magnet field is driven by current
- electric and magnetic fields are connected by Maxwell equations
- charge and current are connected by equation of continuity.

We also learned these properties of a conductor:

- charges stay on the surface. They are not allowed inside.
- currents stay *near* the surface. They do penetrate into the conductor. The penetration depth is the skin depth.

For an insulator,

- there is no current inside
- but charges are allowed inside

In case of resistive wall, the wakefield is generated by the following physical process: When the beam's image charges flow on the vacuum chamber wall, the electric field is terminated by a surface charge on the wall surface, while magnetic field is cancelled by a surface current. However, electric and magnetic fields behave differently.

We conclude that

- the electric field carried by the point charge will terminate immediately by the image charges on the wall surface.
- the magnetic field carried by the point charge is mostly cancelled by the image current on the wall surface, but this cancellation is not exact because the current has penetrated into the wall by a skin depth.
- as the image current slowly re-surfaces after the point charge has past by, this re-surfacing image current drives new magnetic fields. These new magnetic fields occur *after* the point charge has left.
- the re-surfacing changing magnetic field now drives an electric field by Maxwell equation.

So you should now conclude that after the point charge has left, it leaves behind it a "wakefield". For the case of resistive wall, this wakefield is mainly a magnetic field contributing to transverse wake force, but there is also an electric field contributing to a longitudinal wake force.

What happens to particle motion when there are wakefields?

Earlier, we made this observation:

Higher beam intensity => stronger wakefields => instability

We now know there will be wakefields in the vacuum chamber after the beam passes by a discontinuity or a resistive wall. But we still need to explain how the wakefields affect particle's motion.

To address the question of beam instability, the particles we are interested in are those in the beam. These particles move with v=c together with the beam. One such particle was shown as the test particle e in an earlier figure.

As mentioned earlier, it is not the electric force or the magnetic force that are important. It is their sum, the net Lorentz force, that is important, and there is a strong tendency that the electric force and the magnetic force cancel each other. The cancellation is exact for the pancake fields, but the cancellation is lost for the wakefields.

This test particle sees a Lorentz force

$$F = e\,(E + v \times B\,/c)$$

where E and B are the wakefields. Because of the tendency of cancellation, the Lorentz force is much simpler quantity than the electric and the magnetic forces individually.

The wakefield effect on the test charge comes from the Lorentz force. But for this ultrarelativistic beam, the problem is further simplified because we are really only interested in the force *integrated over some distance*, i.e. we are interested only in *the impulse*,

$$\overline{\vec{F}} = \int_{-\infty}^{\infty} ds \vec{F}$$

$\overline{\vec{F}}$ is a function of the spacing between the test charge and the drive beam z. It also depends on (r, θ), the transverse coordinates of the test charge. So we have

$$\overline{\vec{F}} = \overline{\vec{F}}(z,r,\theta).$$

The quantity $\overline{\vec{F}}(z,r,\theta)$ is a much simpler and more elegant quantity to deal with than eE, ev x B/c, or F. In particular, it satisfies an amazing theorem called *Panofsky-Wenzel theorem*,

$$\nabla_\perp \overline{F}_\parallel = \frac{\partial}{\partial z} \overline{\vec{F}}_\perp$$

Here $\parallel$ denotes longitudinal and $\perp$ denotes transverse components.

Is seemed that there is not too much handle on the wakefields because they seem to have to depend on all kinds of details such as the geometry of the discontinuity, or the properties of the wall material. So it is quite amazing that on a very general ground, there is such a theorem like the Panofsky-Wenzel theorem, which relates the longitudinal and the transverse components of $\overline{\vec{F}}$. The proof of the theorem is omitted here, but it says that the transverse gradient of the longitudinal impulse is equal to the longitudinal gradient of the transverse impulse.

Once $\overline{\vec{F}}$ is calculated when the test particle traverses a discontinuity or a section of resistive wall, it receives a net impulse to its momentum as calculated here. If these impulses are too large, the subsequent motion of the test particle will be in question.

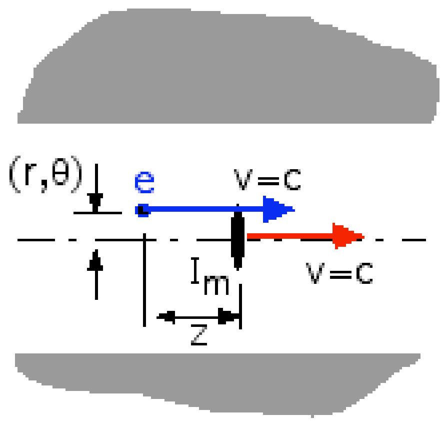
Decomposing wakefields into modes

Even with the Panofsky-Wenzel theorem, these wakefields are still very complicated in general. Accelerator physicists then proceed as follows.

The problem to analyze is what is the impulse received by the test charge e when it integrates the wakefield left behind by a particle beam -- both the test charge and the beam are moving down the pipe ultrarelativistically. To do so, they first consider the beam to be a delta-function in z, i.e. it is infinitely short in length. If the beam has any finite length, the result they obtain with the delta-function beam will serve as a Green's function, and a beam with any general longitudinal distribution can be analyzed simply by linear superposition.

The beam they consider now is an infinitely short beam with arbitrary transverse distribution. To break down the problem further, they next decompose the transverse distribution into "modes". Any general transverse distribution can be decomposed into a summation of transverse modes, the mode index is designated by m. They then consider a single transverse mode m. A general transverse distribution can be obtained again by superposition with a summation over m.

So the problem is now reduced to finding the impulse integrated by a test charge that is a distance z behind another larger beam; the beam is infinitely short in z, has a transverse m-th moment I_m, and is moving along the pipe axis. In this configuration, I_m is the "driving beam" (driving the wakefields), e is the test charge (integrating the wakefields), z is the longitudinal distance that e is trailing behind I_m, and (r,θ) is the transverse displacement of the test charge relative to the pipe axis. The impulse calculated by this configuration is going to be used as a Green's function when we analyze the beam instability problem later.



Consider a circular vacuum chamber pipe for simpler discussion. The wakefields can be decomposed into transverse modes:

m	mode	transverse distribution of wafefields	transverse moment of the driving beam
0	monopole	1	q
1	dipole	cos θ	q <x>
	skew dipole	sin θ	q <y>
2	quadrupole	cos 2θ	$q <x^2-y^2>$
	skew quadrupole	sin 2θ	q <2xy>
3	sextupole	cos 3θ	$q <x^3-3xy^2>$
	skew sextupole	sin 3θ	$q <3x^2y-y^3>$

Higher modes correspond to higher values of mode number m.

For a circular pipe, the m-th multipole wakefield is driven when and only when the driving beam has an m-th moment. For example, if the beam is transversely displaced from the pipe axis, then it contains an m=1 dipole moment. (If the beam is horizontally displaced, it contains a dipole moment. When it is vertically displaced, it contains a skew dipole moment.) Each moment then drives its corresponding wakefields. A beam with no skew quadrupole moment (i.e. a beam with q <2xy>=0), for example, will not drive ~ sin 2θ wakefields. A beam with m-th distribution moment I_m will generate a wakefield in the m-th mode that is proportional to I_m.

With mode decomposition, description of wakefields now becomes easier to handle.

In most applications, it turns out that we care mostly about the m=0 monopole mode when discussing longitudinal collective instabilities, and mostly about m=1 dipole and skew dipole modes when discussing transverse collective instabilities.

Wake functions

Things begin to get complicated. To get a handle of this, we now introduce a quantity called "wake functions".

We mentioned the Panofsky-Wenzel theorem earlier without proof. It turns out that the proof of this theorem contains a lot more information than just the theorem itself. In particular, let us consider again the configuration when a delta-function driving beam with transverse moment I_m going down the axis of the circular pipe. This beam will generate behind it a wakefield in the m-th mode. It can be shown that, for a test charge e following behind this I_m beam by a distance z and having a transverse displacement of (r, θ), the transverse and longitudinal components of the integrated wakefield impulse can be written as

$$\begin{aligned} \overline{\vec{F}}_{\perp}(r,\theta,z) &= -eI_m W_m(z)\, m r^{m-1}(\hat{r}\cos m\theta - \hat{\theta}\sin m\theta) \\ \overline{F}_{\parallel}(r,\theta,z) &= -eI_m W'_m(z)\, r^m \cos m\theta \end{aligned} \tag{1}$$

Here a prime denotes d/dz, $W_m(z)$ is called the *transverse wake function* and $W_m'(z)$ the *longitudinal wake function*. The longitudinal wake function is simply the z-derivative of the transverse wake function.

Equations (1) look rather complicated, and we have omitted its derivation, but they are also quite amazing and contain a wealth of information. First of all, you should be able to check explicitly that the Panofsky-Wenzel theorem is obeyed by these expressions. We then note that the fact that $\overline{\vec{F}}$ is proportional to e and I_m is straightforward and you would have guessed it. On the other hand, one sees that the dependences of $\overline{\vec{F}}$ on m, r, and θ have all been explicitly solved. And this is done even without you being told anything about the geometry of the vacuum chamber discontinuity or the chamber wall's resistivity!

Homework 1

Show that Eqs.(1) satisfy Panofsky-Wenzel theorem.

In Eqs.(1), the only remaining unknown is the *wake function* $W_m(z)$, which depends only on z. Furthermore, the transverse and the longitudinal wake potentials involve the *same* function $W_m(z)$.

So for each vacuum chamber discontinuity, no matter how complicated its geometry is, we have now reduced the wakefield problem to the wake functions $W_m(z)$. For each discontinuity along the vacuum chamber, we just ask the question, "What is the wake function of this discontinuity?" When these wake functions are calculated, we will know the impulse each particle in the beam receives from the collective wakefields generated by the beam. And when that is known, we can analyze the stability of the beam.

As mentioned, for most cases, we are interested only in m=0 for longitudinal beam instabilities, and m=1 for transverse instabilities. Therefore, for each discontinuity, we just ask for two functions: $W_0'(z)$ and $W_1(z)$, and we just calculate these two functions for most applications.

Why do we not worry about m=0 for transverse instabilities?

Answer: $\overline{\vec{F}}_\perp$=0 when m=0 (see the formula for $\overline{\vec{F}}_\perp$). So the leading transverse contribution comes from m=1.

Table below lists the two moments (first the normal moment and then the skew moment) of the driving beam and the associated transverse and longitudinal impulses seen by a test charge e with transverse coordinates (x,y) that follows at a distance z behind a beam which possesses an m-th moment.

Table 1

m	Distribution Moments of Beam	Longitudinal Wake Potential	Transverse Wake Potential
0	q	$-eq\, W_0'(z)$	0
1	$q\langle x\rangle$ $q\langle y\rangle$	$-eq\langle x\rangle x W_1'(z)$ $-eq\langle y\rangle y W_1'(z)$	$-eq\langle x\rangle W_1(z)\hat{x}$ $-eq\langle y\rangle W_1(z)\hat{y}$
2	$q\langle x^2-y^2\rangle$ $q\langle 2xy\rangle$	$-eq\langle x^2-y^2\rangle(x^2-y^2)W_2'(z)$ $-eq\langle 2xy\rangle 2xy\, W_2'(z)$	$-2eq\langle x^2-y^2\rangle W_2(z)(x\hat{x}-y\hat{y})$ $-2eq\langle 2xy\rangle W_2(z)(y\hat{x}+x\hat{y})$
3	$q\langle x^3-3xy^2\rangle$ $q\langle 3x^2y-y^3\rangle$	$-eq\langle x^3-3xy^2\rangle$ $\times(x^3-3xy^2)W_3'(z)$ $-eq\langle 3x^2y-y^3\rangle$ $\times(3x^2y-y^3)W_3'(z)$	$-3eq\langle x^3-3xy^2\rangle W_3(z)$ $\times[(x^2-y^2)\hat{x}-2xy\hat{y}]$ $-3eq\langle 3x^2y-y^3\rangle W_3(z)$ $\times[2xy\hat{x}+(x^2-y^2)\hat{y}]$

Homework 2

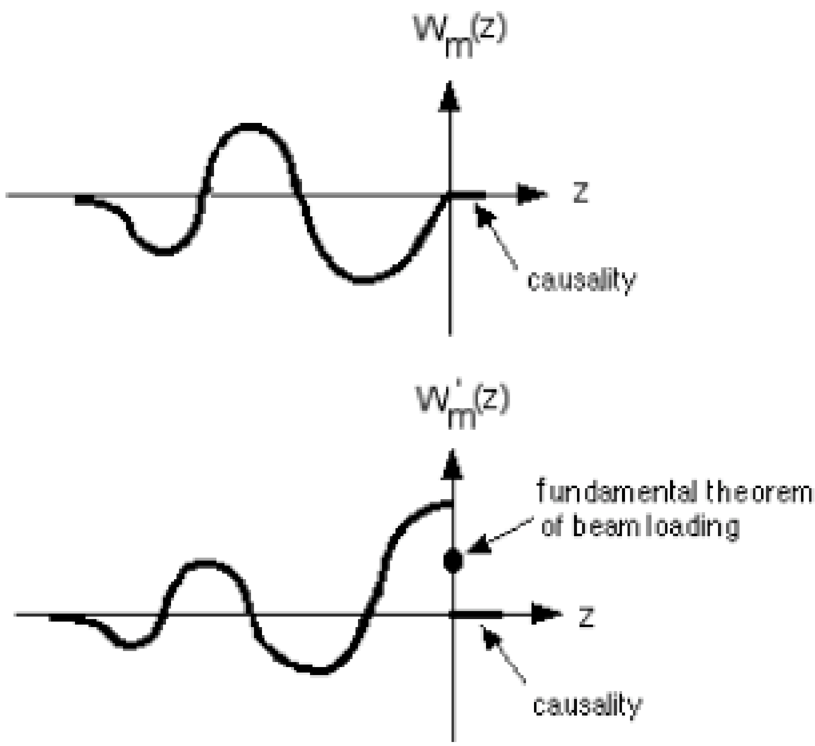
Table 1 is an important table. Follow the text and convince yourself of the results established in the table.

Properties of wake functions

There are many interesting and very general properties of the wake functions:

- $W_m(z) = 0$, $W'_m(z) = 0$ for $z > 0$ (causality).
- $W_m(z) \leq 0$, $W'_m(z) \geq 0$ for $z \to 0^-$.
- $W_m(0) = 0$ (in most cases, except space charge).
- $W'_m(0) = \frac{1}{2} W'_m(0^-)$ (fundamental theorem of beam loading).
- $W'_m(0^-) \geq |W'_m(z)|$ for all z.
- $\int_{-\infty}^{0} W'_m(z) dz \geq 0$.

In general, $W_m(z)$ is a sine-like function, while $W'_m(z)$ is a cosine-like function, as sketched below.



The lower curve is of course related to the upper one by taking derivative with respect to z. You should check all the properties listed above are satisfied by these curves.

As an illustration, let us prove the property $W_m'(0^-) > 0$ here. Immediately following the beam, we expect to see a longitudinal electric field that *retards* the beam, regardless of vacuum chamber properties. This is because the beam must not gain energy as it propagates down the pipe – otherwise we can create a perpetual moving machine. This means the quantity $j_z \boldsymbol{F}_{\parallel}$ must be negative, and in a few steps this proves $W_m'(0^-) > 0$.

Homework 3

(a) Make sure that you follow the steps in the above proof for $W_m'(0^-) > 0$.

(b) Then show $W_m(0^-) < 0$ using the fact that $W_m'(0^-) > 0$ and that $W_m(z>0) = 0$.

There is also the interesting property that $W_m'(0)$, evaluated exactly at z=0, is always equal to ½ times the value $W_m'(0^-)$ evaluated at z slightly less than 0. This is customarily referred to as the *fundamental theorem of beam loading* – even though the reason behind it is rather simple and not quite so fundamental.

One can also say something about the polarity of the transverse wake forces:

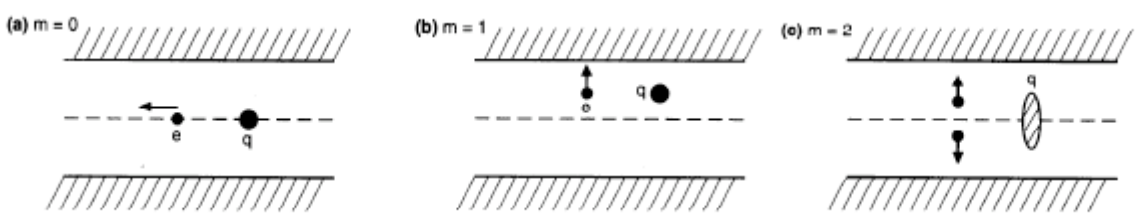


Figure 2.8. The polarity of the wake field always hurts a short beam. For $m = 0$, the longitudinal wake force is retarding. For $m = 1$, the transverse wake force further deflects the test charge e. For $m = 2$, the tail portion of an elliptical beam becomes further elongated. Arrows represent the wake force.

Conclusion: wakefields always do damages to short bunches!

Calculation of wakefields

Some analytically calculable examples of wakefields can be found later together with their corresponding impedances. For a general situation when there are no analytic results, one uses numerical calculations. There are several computer codes that calculate either the wakefields or the impedances. Some of these codes are commercially available, some are developed by individual researchers.

In general, a code that calculates wakefields requires a short driving beam that drives the electromagnetic fields in the vacuum chamber. The beam is then made to propagate down the beam pipe, the fields calculated, and the force on a trailing test particle is integrated (to get the impulse). Such an approach is called a *time-domain* approach.

There are also codes that calculate impedances instead of wakefields. In such codes, the driving beam is considered to be infinitely long and is sinusoidal with frequency ω, and all wakefields as well as the integrated voltage respond to the same frequency. The ratio of this voltage and the driving current then yields the impedance at frequency ω. Such an approach is called a *frequency-domain* approach.

The time-domain calculation of the wake functions and the frequency-domain calculation of the impedances are completely equivalent. They are related by Fourier transforms. For example, when a time-domain code is used and wake functions calculated, people typically make Fourier transforms and print the output of the impedances as a by-product.

Here is one example of a time-domain calculation:

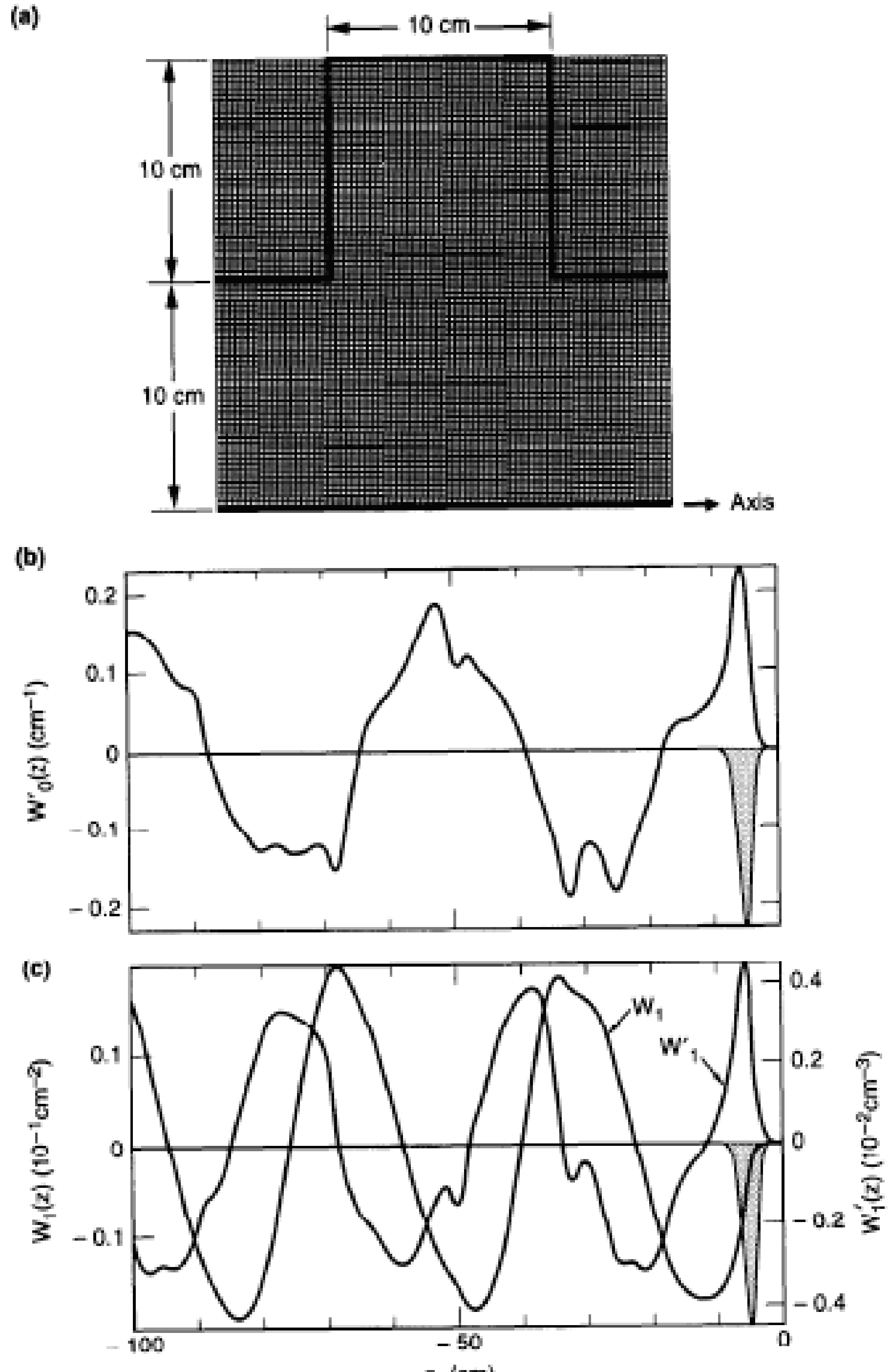


Figure 2.30. A time domain calculation of wake functions for a single cavity using the program TBCI. (a) The cavity geometry and the mesh used for the calculation. (b) The wake function $W_0'(z)$ calculated using a Gaussian driving beam (shaded curve) with $\sigma_z = 1$ cm. The peak of the driving beam is located at $z = -5$ cm. (c) Wake function $W_1(z)$ and $W_1'(z)$. (Courtesy Tom Weiland, Weiren Chou, and Bo Chen, 1991.)

In a time-domain calculation, the integrated wake impulse is numerically calculated for a test charge following a short Gaussian driving bunch. To calculate accurately, one should use as short a driving bunch as possible. If the driving bunch is not short enough, the short-range contents (or equivalently, the high frequency content) of the wakefield will be lost. It is difficult to calculate wakefields at short ranges (or equivalently, impedances at high frequencies). At the least, the number of mesh points will have to be increased sharply, and computing time becomes an issue.

Impedances 阻尼

We talked about frequency content of the wakfields and mentioned that its wavelengths cover a wide range from ~1 μm to ~1 m. The quantities that characterize the frequency content of the wakefields are the impedances, which are nothing but the Fourier transforms of the wake functions,

$$
\begin{aligned}
Z_m^{\parallel}(\omega) &= \int_{-\infty}^{\infty} \frac{dz}{v} e^{-i\omega z/v} W_m'(z) \\
Z_m^{\perp}(\omega) &= \frac{i}{v/c} \int \frac{dz}{v} e^{-i\omega z/v} W_m(z)
\end{aligned}
$$

Since we have already discussed the wake functions, we consider these equations simply the definition of impedances.

Instead of asking about wake functions, an accelerator designer therefore could alternatively ask: “What is the impedance of your accelerator?” The impedance is the quantity most directly related to the maximum beam current that can be accepted by the accelerator.

Properties of impedances

- $Z_m^{\parallel}(\omega) = \frac{\omega}{c} Z_m^{\perp}(\omega)$ (Panofsky-Wenzel theorem in frequency domain).

- $\begin{cases} Z_m^{\parallel *}(\omega) = Z_m^{\parallel}(-\omega) \\ Z_m^{\perp *}(\omega) = -Z_m^{\perp}(-\omega) \end{cases}$ (reality of wake functions).

- $\begin{cases} \int_0^\infty d\omega \, \mathrm{Im} Z_m^{\perp}(\omega) = 0 \\ \int_0^\infty d\omega \, \dfrac{\mathrm{Im} Z_m^{\parallel}(\omega)}{\omega} = 0 \\ \mathrm{Re} Z_m^{\parallel}(0) = 0 \end{cases}$ ($W_m(0) = 0$, in most cases).

- $\begin{cases} \mathrm{Re} Z_m^{\parallel}(\omega) = \dfrac{1}{\pi} \, \mathrm{PV} \displaystyle\int_{-\infty}^{\infty} d\omega' \, \dfrac{\mathrm{Im} Z_m^{\parallel}(\omega')}{\omega' - \omega} \\ \mathrm{Im} Z_m^{\parallel}(\omega) = -\dfrac{1}{\pi} \, \mathrm{PV} \displaystyle\int_{-\infty}^{\infty} d\omega' \, \dfrac{\mathrm{Re} Z_m^{\parallel}(\omega')}{\omega' - \omega} \end{cases}$ (causality, Hilbert transform) The same expressions apply to $Z_m^{\perp}$. PV means taking the principal value of the integral.

- $\begin{cases} \mathrm{Re} Z_m^{\parallel}(\omega) \geq 0 \text{ for all } \omega \\ \mathrm{Re} Z_m^{\perp}(\omega) \geq 0 \text{ if } \omega > 0, \quad \leq 0 \text{ if } \omega < 0 \end{cases}$

- $Z_1^{\perp} \approx \frac{2c}{b^2 \omega} Z_0^{\parallel}$, $Z_m^{\perp} \approx \frac{2c}{b^{2m} \omega} Z_0^{\parallel}$, $Z_m^{\parallel} \approx \frac{2}{b^{2m}} Z_0^{\parallel}$ These are approximate expressions relating transverse and longitudinal impedances, $b =$ pipe radius. They are exact for resistive round pipe.

Some expressions of impedances and wake functions

To find the impedance for a given vacuum chamber discontinuity, one needs to solve Maxwell equations for the electromagnetic fields produced in the vacuum chamber. Over the years, a large arsenal of techniques had been developed to calculate the impedances. Most results involve numerically solving the associated boundary value problems.

We mentioned that there are 3 ways when wakefields are generated. Three cases, each representing one of these 3 ways, that permit analytical expressions are given below. A lot more examples can be found in the Handbook.

<u>Direct space charge</u>

Impedances	Wake functions
$Z_0^{\parallel} = i\frac{Z_0 L\omega}{4\pi c\gamma^2}\left(1+2\ln\frac{b}{a}\right)$	$W_0' = \frac{Z_0 cL}{4\pi\gamma^2}\left(1+2\ln\frac{b}{a}\right)\delta'(z)$
$Z_{m\neq 0}^{\perp} = i\frac{Z_0 L}{2\pi\gamma^2 m}\left(\frac{1}{a^{2m}}-\frac{1}{b^{2m}}\right)$	$W_{m\neq 0} = \frac{Z_0 cL}{2\pi\gamma^2 m}\left(\frac{1}{a^{2m}}-\frac{1}{b^{2m}}\right)\delta(z)$

where $Z_0=(\mu_0/\varepsilon_0)^{1/2} = 377\ \Omega$ is the free-space impedance.

This case is in free space; there is no vacuum chamber, therefore no vacuum chamber discontinuity and no resistive wall. So why is there a wakefield and impedance? The answer is that the beam is not ultrarelativistic. Indeed, you should note that these wakefield and impedances are proportional to $1/\gamma^2$. For unltrarelativistic beam, they vanish.

Space charge effects are most significant only for low-to-medium energy proton or heavy ion accelerators. TPS, for example, will not suffer from space charge instability because γ is so large. This impedance is purely imaginary. By the sign of its imaginary part, we call this impedance "capacitive".

Resistive wall

The second way to produce wakefield is when the vacuum chamber wall is resistive:

Impedances	Wake functions
$Z_m^{\parallel} = \frac{\omega}{c} Z_m^{\perp}$	$W_m = -\frac{c}{\pi b^{m+1}(1+\delta_{m0})}\sqrt{\frac{Z_0}{\pi\sigma_c}}\frac{L}{\lvert z\rvert^{1/2}}$
$Z_m^{\parallel} = \frac{1-\mathrm{sgn}(\omega)i}{1+\delta_{0m}}\frac{L}{\pi\sigma_c\delta_{\mathrm{skin}}b^{2m+1}}$	$W_m' = -\frac{c}{2\pi b^{m+1}(1+\delta_{m0})}\sqrt{\frac{Z_0}{\pi\sigma_c}}\frac{L}{\lvert z\rvert^{3/2}}$

The impedance is proportional to 1-i, i.e. it is “half resistive and half inductive”.

Slowly varying wall boundaries

The third way is when the vacuum chamber has discontinuities, even though perfectly conducting. Consider a case when the vacuum chamber wall varies along the accelerator slowly, a perturbation technique can be used to calculate the impedances. Specify the wall variation by h(z) (1-dimensional cylindrically symmetric bump). At low frequencies k=ω/c < 1/(bump length or width), the impedance is purely inductive,

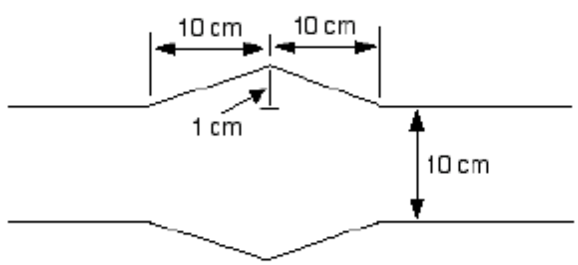
$$Z_0^{\|} = -\frac{2ikZ_0}{b}\int_0^{\infty} \kappa |\tilde{h}(\kappa)|^2 d\kappa$$

$$\text{where} \quad \tilde{h}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(z) e^{-ikz} dz$$

and k=ω/c.

Homework 4

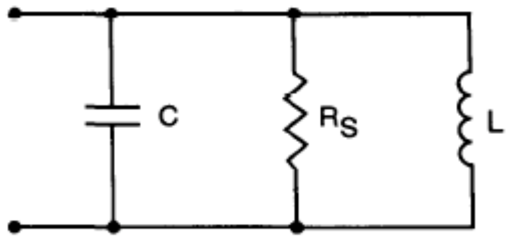
Calculate $Z_0^{\|}$ for



When the boundary varies rapidly, this impedance formula breaks down, and most likely numerical calculation has to be applied.

Resonator model

The longitudinal impedance can often be modeled by an equivalent parallel LRC resonator circuit,

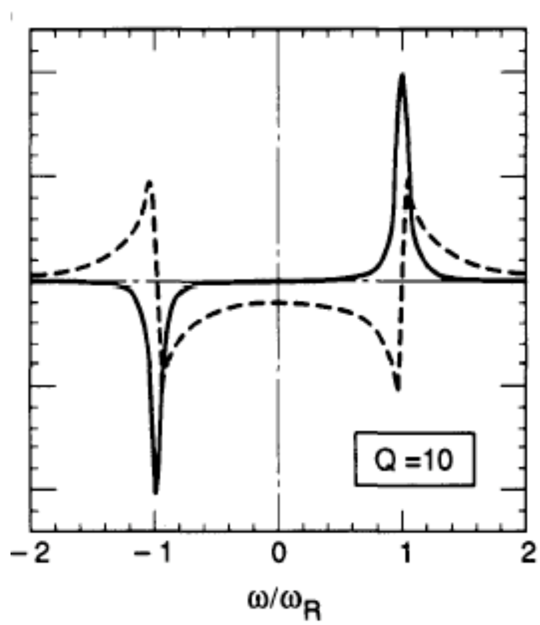


$$\frac{1}{Z_m^{\parallel}} = \frac{1}{R_S} + \frac{i}{\omega L} - i\omega C$$

or

$$Z_m^{\parallel} = \frac{R_S}{1 + iQ\left(\frac{\omega_R}{\omega} - \frac{\omega}{\omega_R}\right)}$$

where $Q = R_S\sqrt{C/L}$ is the quality factor and $\omega_R = 1/\sqrt{CL}$ is the resonant frequency. The width (half width at half maximum) of the resonance peak of Re $Z_m^{\parallel}(\omega)$ is about $\Delta\omega \approx \omega_R/2Q$ if Q>>1. A sharply peaked impedance has Q>>1, while a broad-band impedance has Q~1.



The corresponding wake function (Fourier transform) is

$$W'_m(z) = 2\alpha R_S\, e^{\alpha z/c}\left(\cos\frac{\overline{\omega} z}{c} + \frac{\alpha}{\overline{\omega}}\sin\frac{\overline{\omega} z}{c}\right)$$

$$\alpha = \omega_R/2Q \text{ and } \overline{\omega} = \sqrt{\omega_R^2 - \alpha^2}$$

At low frequencies $\omega \ll \omega_R$, $Z_m^{\parallel}(\omega) \approx -i\omega L$ is inductive. For $\omega \gg \omega_R$, we have $Z_m^{\parallel}(\omega) \approx i/\omega C$, which is capacitive. Around the resonant frequency $\omega \sim \omega_R$, the impedance $Z_m^{\parallel}(\omega) \approx R_S$ is mostly resistive.

The same resonator also contributes to a transverse impedance,

$$Z_m^{\perp} = \frac{c}{\omega}\,\frac{R_S}{1 + iQ\left(\dfrac{\omega_R}{\omega} - \dfrac{\omega}{\omega_R}\right)}$$

The corresponding transverse wake function is

$$W_m(z) = \frac{cR_S\omega_R}{Q\overline{\omega}}\, e^{\alpha z/c}\sin\frac{\overline{\omega} z}{c}$$

We mentioned earlier that the RF cavities often represent the largest discontinuities, and therefore sometimes the largest source of impedance. The accelerating mode (the fundamental mode) of an RF cavity typically has a Q-value ~10^4. This means the electromagnetic fields trapped in the fundamental mode of the cavity will make ~10^4 oscillations before they decay significantly. For most discontinuities on the vacuum chamber walls, however, the wakefields decay much faster. Many of them in fact can be modeled as a resonator model with Q~1.

For a low-Q object of the size of the order of the beam pipe radius b, a *broad-band resonator model* of its impedance would have the resonator parameters:

$R_S = Z_0//2\pi \sim 60\ \Omega$

$Q = 1$

$\omega_R = c/b$

The quantity Z/n

In some accelerators (most likely the earlier accelerators, such as an older synchrotron), beam bunches are long. For example, some have bunches longer than 1m, or at least several cm. In these cases, when the bunches are longer than the vacuum chamber pipe radius, it turns out that often it is the lower frequencies of the wakefields that dominate the collective instabilities.

We mentioned before that for longitudinal instabilities, it is the m=0 effects that dominates. This means we must pay most attention to $Z_0^{||}$.

From the properties of impedances given above, we know that $Z_0^{||}(\omega=0) = 0$. For small ω, this means Im($Z_0^{||}$) (ω) is proportional to ω. For older synchrotrons, the instability is therefore specified by the low-frequency slope of Im($Z_0^{||}$) (ω), i.e. we need a quantity Im($Z_0^{||}$) (ω) / ω at $\omega \to 0$. Such a quantity is called Z/n. We therefore ask the question, "what is the value of Z/n of your synchrotron?" By Z/n, we mean:

$$Z/n = \left(\text{ the low-frequency slope } \mathrm{Im}(Z_0^{||})(\omega) / \omega \right) \times \omega_0$$

where ω_0 is just the revolution frequency of the synchrotron.

Note that this Z/n quantity is a single-valued quantity, in units of Ohms. In comparison, the impedance is a much more complicated quantity because it is an entire function of frequency. By using Z/n, the entire problem of beam instability becomes a simple question of how large is the value of a single quantity Z/n. It is clear that this single value Z/n will not completely describe the instabilities, and it represents an oversimplification of the problem. But for older synchrotrons, it turns out not too bad an oversimplification.

It is an impressive accomplishment of the accelerator physicists to manage to zero in on a single key quantity to address such a complex physical problem. 抽丝剥茧 Their hard work 辛勤的耕耘 is summarized again in the following framework 别忘了整體的大架構:

A seemingly impossibly complex electromagnetic problem with 3-dimensional boundaries

→

Realizing that for ultrarelativistic beams, it is only the impulse that counts

→

Panofsky-Wenzel theorem

→

Wake functions $W_m(z)$

→

Impedances $Z_m(\omega)$

→

Z/n

The fact that this drastic simplification is even possible is a very lucky blessing from Mother Nature. 再一次的萬幸！

Careless limit of Z/n

Consider a cavity structure of the size of the pipe radius b on the vacuum chamber. Its impedance can be represented as a broad-band resonator model. In terms of Z/n, it will contribute

$$Z/n \text{ per cavity} \sim (Z_0/2\pi)(\omega_0/\omega_R) \qquad (2)$$

where $\omega_R = c/b$ and $\omega_0=c/R$ and $2\pi R$ is circumference of the circular accelerator.

If one now imagines a carelessly built accelerator in which the vacuum chamber is filled with all sorts of cavities and discontinuities of approximately the same size as the pipe radius, the total Z/n around the circumference is

$$(Z/n \text{ total}) \sim (Z/n \text{ per cavity})\, n_{cav}$$

where $n_{cav} = 2\pi R/2b$ is the total number of cavities around the circumference. This carelessly designed accelerator has

$$(Z/n \text{ total}) \sim (Z_0/2) \sim 160\ \Omega$$

This is the careless limit of impedance. You cannot do worse than that. Note that this limit is a fundamental constant, independent of the accelerator size R and the pipe size b.

In case a fraction ~f of the accelerator is filled with cavities, one has

$$Z/n \sim f \text{ X } (160\ \Omega)$$

In a typical modern accelerator, attempts are made to make Z/n less than 1 Ω or so. This means the vacuum chamber has to be sufficiently smooth to suppress the impedance by a factor of a few hundred compared with the careless limit.

Homework 5

Derive Eq.(2) for Z/n of a broad band resonator.

Impedance at high frequencies

For the more modern accelerators, however, particularly when the beam bunches gets shorter and peak intensity gets higher, the single value Z/n no longer provides a complete picture. The collective instability problem becomes more difficult. The single value of Z/n suffices for older synchrotrons, but not for these modern applications. For these applications, we need not only Z/n but also the entire impedance functions, particularly at high frequencies. Unfortunately, high frequency is also where impedance is most difficult to measure or to calculate.

The research on impedances 一些研究的方向 involves bench measurement of impedance components (electronics, rf techniques), analysis (electromagnetism problems with boundary conditions), computation physics (3-dimensional boundary value problems with very fine meshes, inverting $10^6 \times 10^6$ matrices!). High power computing is one important resource needed.

3. Collective Instabilities

We described wakefields and impedances. We still need to describe how to use these quantities to calculate beam instabilities. For example: Given the impedance, is the beam stable? If it is unstable, what happens to the beam? What is the instability growth rate?

As mentioned before, there are a large number of instability mechanisms. We will briefly describe three of them below:

- Parasitic heating
- Robinson instability
- Strong head-tail instability

Parasitic heating

When a beam bunch of charge q and line density $\lambda(t)$ traverses an impedance region in the vacuum chamber, it suffers some energy loss to the impedance. Let the longitudinal impedance be $Z_0^{\|}(\omega)$. This *parasitic energy loss* (sometimes also called HOM heating, HOM means "higher order modes") by the beam bunch is given by

$$\Delta\mathcal{E} = -\kappa^{\|} q^2$$

where $\kappa^{\|}$ is called the *loss factor*,

$$\kappa^{\|}(\sigma) = \frac{1}{\pi}\int_0^{\infty} d\omega\, \mathrm{Re}Z_0^{\|}(\omega)\, |\tilde{\lambda}(\omega)|^2$$

For a Gaussian bunch, we have

$\lambda = e^{-t^2/2\sigma^2}/(\sqrt{2\pi}\sigma)$, $\tilde{\lambda}(\omega) = e^{-\omega^2\sigma^2/2}$.

Only the real part (the resistive part) of the impedance contributes to the parasitic loss. Inductive impedances and the space charge or the slowly varying wall impedances do not introduce a net energy loss to the beam. However, this does not mean that individual particles do not change their energies. It only means that the energy loss by particles at the head of the bunch is recovered by particles in the tail of the bunch, so that there are only energy transfers but no net energy loss of the entire beam bunch.

In general, it happens that this beam energy loss becomes large for short bunches. Parasitic heating is mainly a problem for high intensity, and short, bunches. Substitute in the impedance of a resistive wall, for example, gives a formula

$$\frac{\kappa^{\|}(\sigma)}{L} = \frac{\Gamma(\frac{3}{4})c}{4\pi^2 b\sigma_z^{3/2}}\left(\frac{Z_0}{2\sigma_c}\right)^{1/2}, \qquad \Gamma(\frac{3}{4}) = 1.225$$

Parasitic loss gives rise to heating of the vacuum chamber wall where there are impedances. For example, in high intensity electron storage rings, the beam position monitors or bellows can easily heat up and get burned. This is especially serious when short bunches are required for the applications.

Most of the parasitic loss occurs as the beam traverses a discontinuous structure in the vacuum chamber pipe. Part of the wakefield gets trapped by the structure if the structure is cavity-like and if the wakefield frequency is below the cutoff frequency of the pipe. This trapped field energy is eventually deposited as heat on the cavity walls. The rest of the wakefield, with frequency higher than the cutoff frequency, propagates down the pipe and eventually deposits its energy on lossy material elsewhere in the vacuum chamber, much like heating a potato in a microwave oven.

The parasitic energy lost by the beam goes into wakefields. Typically, only a small fraction of the particle energy is depleted to produce the wakefields, and most of the energy stored in the wakefields ends up as heat on the vacuum chamber walls. But under unfavorable conditions, a small portion of the wakefield energy can be transferred systematically back to beam motion, causing beam instabilities. The parasitic loss, therefore, is ultimately responsible for the various collective beam instabilities. How the wakefields affect the beam dynamics and what are the mechanisms of the various collective beam instabilities are subjects to which we will have to study. The parasitic energy loss itself, of course, will have to be supplied back to the beam by an RF accelerating voltage.

Robinson instability

Robinson instability is one of the most basic instability mechanisms. It is a longitudinal instability that occurs in circular accelerators. The main contributor to this instability is the longitudinal impedance due to the RF accelerating cavities. These cavities are tuned to have a resonant frequency ω_R for its fundamental accelerating mode. This mode is where the klystrons feed into, but at the same time, it is also a big source of wakefield and impedance. Since we must have these modes in order to accelerate the beam, we must accept the existence of these very big wakefield and impedance and try to live with them.

In wakefield language, the fundamental mode is one of the m=0 modes with its electric field mainly in the longitudinal direction. In fact, it is the biggest m=0 mode in the entire accelerator. The real part of this impedance peaks at ω_R with a narrow width. The width is approximately given by $\Delta\omega_R/\omega_R \approx \pm 1/Q$, where Q is the Q-value of the RF cavity's accelerating mode. Typically, $Q \sim 10^4$ (or 10^9 for superconducting cavities). So this impedance is sharply peaked.

By design, ω_R is very close to an integer multiple of the revolution frequency ω_0 of the beam. This necessarily means that the wake field excited by the beam in the cavities contains a major frequency component near $\omega_R \approx h\omega_0$ or equivalently, the impedance $Z_0^{\parallel}$ has a sharp peak at $\omega_R \approx h\omega_0$, where h is an integer called the *harmonic number*.

As we will soon show, the exact value of ω_R relative to $h\omega_0$ is of critical importance for the stability of the beam. Above the transition energy, the beam will be unstable if ω_R is slightly above $h\omega_0$ and stable if slightly below. This instability mechanism was first analyzed by Robinson in 1964.

Kenneth Robinson (1925-1979)

To simplify the physical picture, let us consider a beam that is just a big charge Ne. It has no internal structures, and is just a big point charge. This of course is an over-simplification of a true beam bunch because internal structures can be important. Some instabilities – in fact, many instabilities -- involves internal instabilities. So this over-simplified picture will miss all those instabilities. However, this picture does allow descriptions of some important instabilities, for example the Robinson instability, and we will adopt this picture here. This picture is called a *one-macroparticle model* of the beam.

The advantage of this one-macroparticle model is that it allows simple analytical results. One can extend this idea and create a few two-macroparticle models. They also describe other instabilities, particularly those for which internal structures do play a role, and allow analytical results. But we will not address these models yet.

Since Robinson instability is a longitudinal effect, we now consider the longitudinal motion of this one-particle beam. Let z_n be the longitudinal displacement of the beam at the accelerating RF cavity in the n-th revolution, measured relative to the center of an idealized bunch unaffected by wakefields. The rate of change of z_n is related to the relative energy error $\delta_n = \Delta E/E$ of the beam in the same n-th revolution by（李世元)

$$\frac{d}{dn} z_n = -\eta C \delta_n$$

where η is the slippage factor, C is the accelerator circumference. A positive z_n means the beam arrives the RF cavity earlier than the idealized bunch.

The energy error also changes with time. Its equation of motion is

$$\frac{d}{dn} \delta_n = \frac{(2\pi \nu_s)^2}{\eta C} z_n$$

where ν_s is the synchrotron tune.

If we combine these two equations, we get a simple harmonic oscillation for both z_n and δ_n,

$$\frac{d^2 z_n}{dn^2} + (2\pi \nu_s)^2 z_n = 0$$

The oscillation has a phase advance of $2\pi\nu_s$ per revolution. This oscillation is just the *synchrotron oscillation* of the macroparticle beam. Typically, $\nu_s \ll 1$, i.e., synchrotron oscillation is slow and the beam does not execute much synchrotron motion during the time it completes one revolution.

But the above equation is valid only when the beam has a vanishing intensity. Otherwise, the pure simple harmonic oscillation is perturbed. For an intense beam, the energy variation also depends on the wake field generated by the beam. The longitudinal wakefield affects the energy equation of motion. The $d\delta_n/dn$ equation then acquires an additional term,

$$\frac{d}{dn}\delta_n = \frac{(2\pi\nu_s)^2}{\eta C} z_n - \frac{Nr_0}{\gamma} \sum_{k=-\infty}^{n} W_0'(kC - nC + z_n - z_k)$$

where W_0' is the longitudinal wake function accumulated over one turn of the accelerator. The summation over k is over the wakefields left behind by the beam from all revolutions prior to the n-th. The argument of the wake function is the longitudinal separation of beam positions between the n-th and the k-th revolutions.

The equation of motion now becomes

$$\frac{d^2 z_n}{dn^2} + (2\pi\nu_s)^2 z_n = \frac{Nr_0\eta C}{\gamma} \sum_{k=-\infty}^{n} W_0'(kC - nC + z_n - z_k)$$

We need now to solve this equation for z_n as a function of turn number n. To do so, we let z_n be written as

$$z_n \propto e^{-in\Omega T_0}$$

where $T_0=C/c=2\pi/\omega_0$ is the beam revolution period, and Ω is the mode frequency of the beam oscillation and is a key quantity yet to be determined.

Substituting into the equation of motion, we find an algebraic equation for Ω,

$$\Omega^2 - \omega_s^2 = -\frac{Nr_0\eta c}{\gamma T_0} \sum_{k=-\infty}^{\infty} (1 - e^{-ik\Omega T_0}) W_0''(kC)$$

where $\omega_s = \nu_s\omega_0$ is the synchrotron oscillation frequency.

Now the wake function can be expressed in terms of the longitudinal impedance by a Fourier transform. This yields

$$\Omega^2 - \omega_s^2 = -i\frac{Nr_0\eta}{\gamma T_0^2} \sum_{p=-\infty}^{\infty} \left[p\omega_0 Z_0^{\parallel}(p\omega_0) - (p\omega_0 + \Omega) Z_0^{\parallel}(p\omega_0 + \Omega) \right]$$

Given the impedance, this equation can in principle be solved for Ω. Note that Ω appears on both sides of the equation. Here, however, we will take a perturbative approach and assume Ω does not deviate much from ω_s for modest beam intensities. We thus replace Ω by ω_s on the right hand side of the equation. Quantity Ω is then easily solved.

In general, Ω is complex. The real part of Ω is the perturbed synchrotron oscillation frequency of the collective beam motion, and the imaginary part gives the growth rate (or damping rate if negative) of the motion. We then obtain a *mode frequency shift*,

$$\begin{aligned} \Delta\Omega &= \mathrm{Re}(\Omega - \omega_s) \\ &= \frac{Nr_0\eta}{2\gamma T_0^2 \omega_s} \sum_{p=-\infty}^{\infty} \left[p\omega_0 \,\mathrm{Im}\, Z_0^{\parallel}(p\omega_0) - (p\omega_0 + \omega_s) \mathrm{Im}\, Z_0^{\parallel}(p\omega_0 + \omega_s) \right] \end{aligned}$$

and an *instability growth rate*,

$$\tau^{-1} = \mathrm{Im}(\Omega - \omega_s) = \frac{Nr_0\eta}{2\gamma T_0^2 \omega_s} \sum_{p=-\infty}^{\infty} (p\omega_0 + \omega_s) \mathrm{Re}\, Z_0^{\parallel}(p\omega_0 + \omega_s)$$

It is the imaginary part of the impedance that contributes to the collective frequency shift and the real part that contributes to the instability growth rate.

Note that when we measure the synchrotron frequency in a real operation, the frequency we measure is not ω_s, but the shifted mode frequency Ω.

So far our results holds for arbitrary impedance. We now consider the resonator impedance for the fundamental cavity mode. The only significant contributions to the growth rate come from two terms in the summation, namely p=±h because the impedance is sharply peaked there. This gives

$$\tau^{-1} \approx \frac{N r_0 \eta h \omega_0}{2\gamma T_0^2 \omega_s}\left[\mathrm{Re}\, Z_0^{\parallel}(h\omega_0 + \omega_s) - \mathrm{Re}\, Z_0^{\parallel}(h\omega_0 - \omega_s)\right]$$

Beam stability requires $\tau^{-1} < 0$. That is, the real part of the impedance must be lower at frequency $h\omega_0+\omega_s$ than at frequency $h\omega_0+\omega_s$ if $\eta>0$ (above transition), and the other way around if $\eta<0$ (below transition). This condition gives the important *Robinson stability criterion* that, above transition, the resonant frequency ω_R of the fundamental cavity mode should be slightly detuned downwards from an exact integral multiple of ω_0. Below transition, stability requires ω_R be slightly higher than $h\omega_0$.

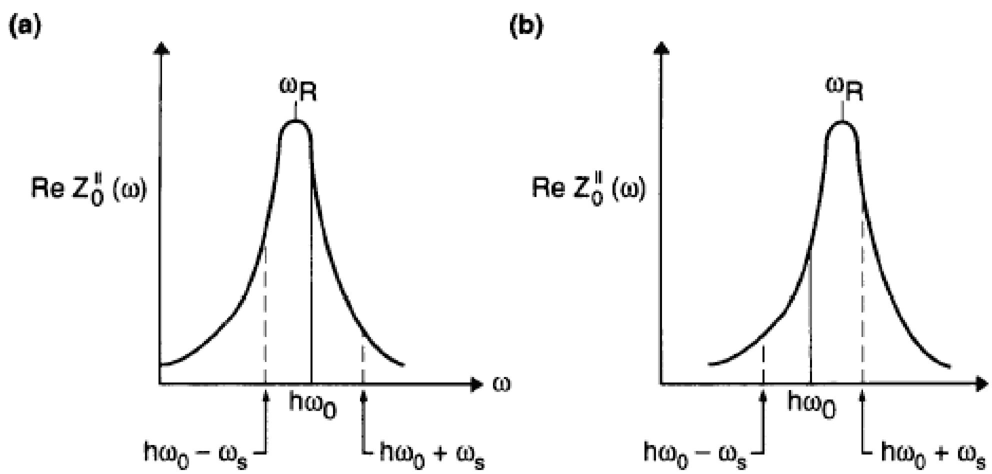


Figure 4.4. Illustration of the Robinson stability criterion. The rf fundamental mode is detuned so that ω_R is (a) slightly below $h\omega_0$ and (b) slightly above $h\omega_0$. (a) is Robinson damped above transition and antidamped below transition. (b) is antidamped above transition and damped below transition.

When $\tau^{-1}<0$, the Robinson mechanism leads to exponential damping of any synchrotron oscillations of the beam. When $\tau^{-1}<0$, the beam is unstable because any accidental small synchrotron oscillation of the beam would grow exponentially with the instability growth rate, eventually leading to the loss of the beam.

Robinson damping (or antidamping) can be rather strong. When the Robinson criterion is met, the synchrotron oscillation of the beam is "Robinson damped," and this damping will help stabilizing the beam against similar instabilities due to other impedance sources.

Physically, Robinson instability comes from the fact that the revolution frequency of an off-momentum beam is not given by ω_0 but by $\omega_0(1-\eta\delta)$. To illustrate the physical origin of the Robinson instability mechanism, consider a beam executing synchrotron oscillation above transition. Due to the energy error of the beam, the impedance samples the beam signal at a frequency slightly below $h\omega_0$ if $\delta>0$, and slightly above $h\omega_0$ if $\delta<0$. In order to damp this synchrotron oscillation of the beam, we need to let the beam lose energy when $\delta>0$ and gain energy when $\delta>0$. This can be achieved by having an impedance that decreases with increasing frequency in the neighborhood of $h\omega_0$. The Robinson stability criterion then follows.

Strong head-tail instability

The next topic is to introduce another instability mechanism, this time a transverse instability, called *strong head-tail instability*, and it is to be discussed using a two-macroparticle model. But here we will not elaborate on the analysis of this instability. Instead, we just mention that this transverse instability was first observed and analyzed at PEP. When intensity is above a certain threshold, the beam is unstable. Below it, the beam is stable but its motion is perturbed as seen below:

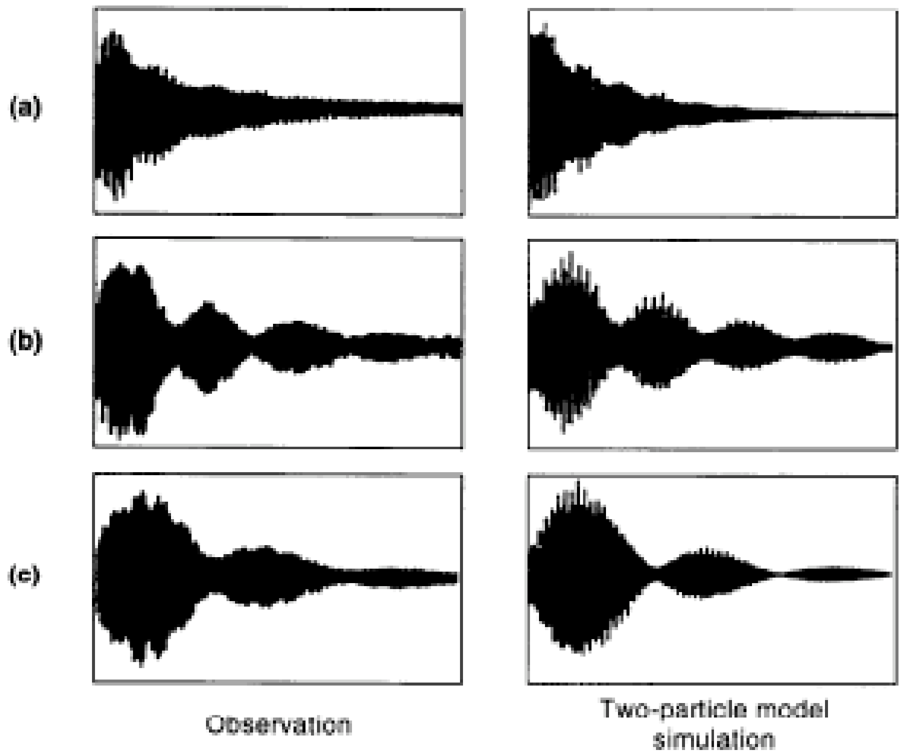
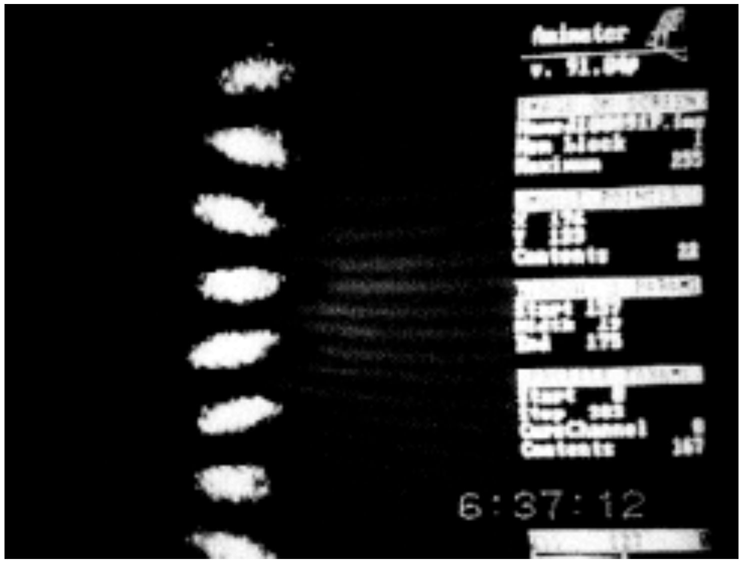


Figure 4.11. The beam-position-monitor signal as a function of time after the beam is kicked. On the left are the signals observed at the PEP storage ring: (a) is when the beam intensity N is 0.86 times the threshold intensity N_{th}. (b) $N / N_{th} = 0.93$, and (c) $N / N_{th} = 0.988$. On the right are the results of simulation using a two-particle model with (a) $\Upsilon / 2 = 0.77$, (b) $\Upsilon / 2 = 0.96$, and (c) $\Upsilon / 2 = 0.99$.

You see here also that the observation and the analysis using a two-particle model agree rather well.

The strong head-tail instability was also seen at LEP using a streak camera:



There are also many other instability mechanisms. Still another important topic would be to discuss an effect called *potential well distortion*, or potential well bunch lengthening. We will not cover all of them.

Further readings

A good fraction of the notes can be found in:
Handbook of Accelerator Physics and Engineering, ed. A. Chao and M. Tigner, World Scientific, 3rd print (2006).

For discussions emphasizing the physics principles, one may consult:
Physics of Collective Beam Instabilities in High Energy Accelerators, A. Chao, Wiley (1993).
(welcome to download: http://www.slac.stanford.edu/~achao/wileybook.html)

or

Physics of Intensity Dependent Beam Instabilities, K.Y. Ng, World Scientific (2006).

For a much more extensive discussion on impedances, one may consult:
Impedances and Wakes in High-Energy Particle Accelerators, B. Zotter and S. Kheifets, World Scientific (1997).