Beam Dynamics with High Intensity

Alex Chao, SLAC

The Gauss law

Collective instabilities in accelerators mostly come from an intense charged-particle beam electromagnetically interacting with its vacuum chamber environment. As the beam interacts with its environment, it generates an electromagnetic field called the *wakefield*, and the wakefield acts back on the beam, disturbing its motion, and if the perturbation is strong enough, the beam becomes unstable.

To discuss the wakefields, we must start with its ultimate origin, the *Gauss law*, which states that each charged particle always is attached with it a definite amount of electric field lines. We can distort these field lines but we can never cut them loose from the charge under any circumstances. Furthermore, the amount of field lines attached to each charge can never be changed, neither increased nor decreased.

Gauss law is an amazing law. Mathematically, it reads

$$\nabla \cdot \vec{E} = 4\pi\rho$$

Physically it reads: Electric field lines are absolutely attached to the charges.

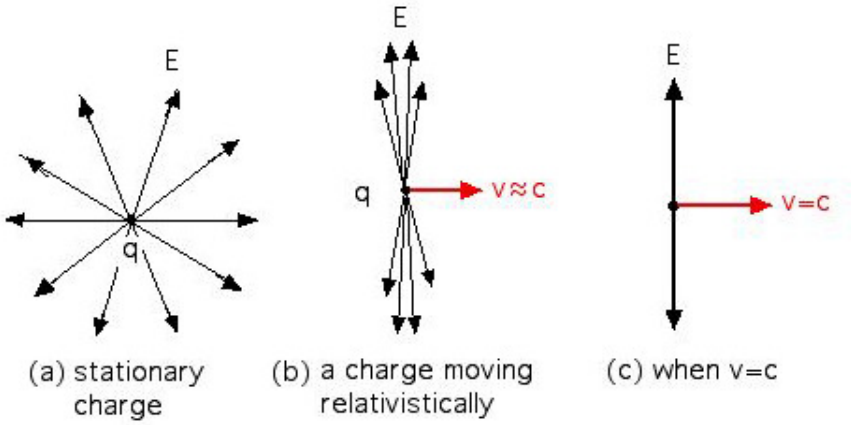
Integral form of the Gauss law:

$$\oint_S \vec{E} \cdot d\vec{S} = 4\pi Q$$

where Q is the total charge inside the volume enclosed by the surface S. It is amazing that this law holds no matter how the charges are moving – nonrelativistic, relativistic, or under acceleration, or whether they are embedded in any type of material. It also does not matter how close the charges might be immediately next to the surface S. The field integral will make a sudden change when a charge crosses the surface even infinitesimally.

A moving charge

If the charge is stationary and if it is in a free space, its field lines are as shown in fig.(a) below. For a moving charge, we see fig.(b). When v approaches c, we have fig.(c), when all the electric fields stay in an infinitely thin sheet as result of theory of relativity. For most accelerators, case (c) is closest to the case under consideration.



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

When the particle is moving, it also generates a magnetic field. This magnetic field also contracts to a thin pancake when $v = c$. Direction of the electric field is radial; direction of the magnetic field is azimuthal (right-hand rule).

$$\begin{aligned} E_r &= \frac{2q}{r}\delta(z-ct) \\ B_\theta &= \frac{2q}{r}\delta(z-ct) \end{aligned}$$

One observes that

$$B_\theta = E_r \quad \text{when} \quad v = c$$

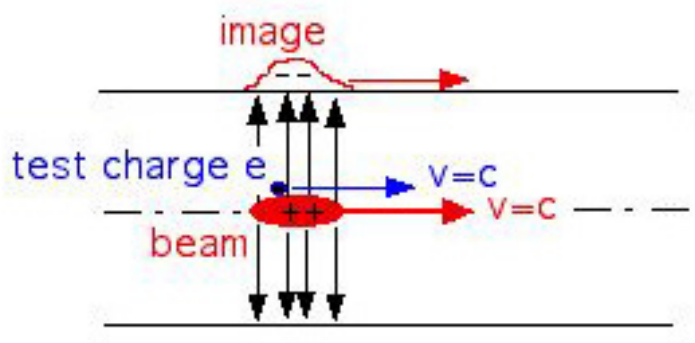
However, when $v = 0$, there is no magnetic field. When v increases, B_θ increases, but still weaker than E_r. Only when $v = c$, we have $B_\theta = E_r$. The fact that $B_\theta = E_r$ when $v = c$ has important consequences, as explained next.

The vacuum chamber

We now add the vacuum chamber. Consider a very smooth vacuum chamber beam pipe.[1] Consider the smooth pipe wall to be perfectly conducting.

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

[1] How smooth does the chamber have to be? A 1-mm discontinuity on the pipe is considered a potential problem. In some circumstances, a 1-μm roughness on the wall surface can have a significant effect



The electromagnetic fields are perfectly and cleanly 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 $v = c$ (except that this is phase velocity). The entire field pattern moves with the beam. There is no wakefield.

Is this beam stable? Consider a particular particle in the beam, the blue "test particle" e in the above figure. This test particle will see an electric force $e\vec{E}$ due to the electric field carried by 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 $(e/c)\,\vec{v} \times \vec{B}$. It is easily seen that this magnetic force is pointing towards the pipe axis.

We mentioned that when $v = c$, we have $E_r = B_\theta$. In the ultrarelativistic limit, therefore, the electric and the magnetic forces exactly cancel. The particles in the ultraralativistic beam see electric force and magnetic force, but they do not see a net force. The collective electromagnetic fields carried by the beam do not influence particle motion. There is no collective instabilities.

This cancellation between the electric and magnetic forces due to the beam's self fields is very fortunate and very important. Without this cancellation, no modern accelerators would have worked.

We conclude that there are three possible ways for a collective instability to occur:

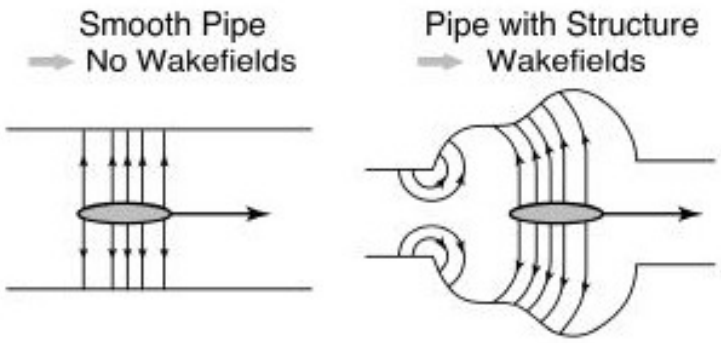
1. the beam is not relativistic enough
2. the vacuum chamber is too resistive
3. the vacuum chamber is not smooth enough

If all these conditions hold, the beam is stable as just illustrated. If any one of these conditions occurs, the exact cancellation of the electric and magnetic forces is disrupted, and the beam can encounter an instability.

We construct accelerators to be as close to the cancelation condition as possible. The electric and magnetic forces generally cancel to high accuracy by design. However, the cancellation is never perfect. The vacuum chambers made of copper or aluminum are not perfectly conducting. There will be many small necessary discontinuities along the vacuum chamber pipe, such as beam position monitors, vacuum pumping ports, etc. There are also big discontinuities known as rf cavities. As to the condition of $v = c$, it is never satisfied completely. So the cancellation of electric and magnetic forces are not perfect. And that residual non-cancelation leads to collective instabilities.

Wakefields due to discontinuities

When a beam traverses a discontinuity, an electromagnetic wakefield is generated. An intense beam will generate a strong wakefield. When the wakefield becomes too strong, the beam becomes unstable.



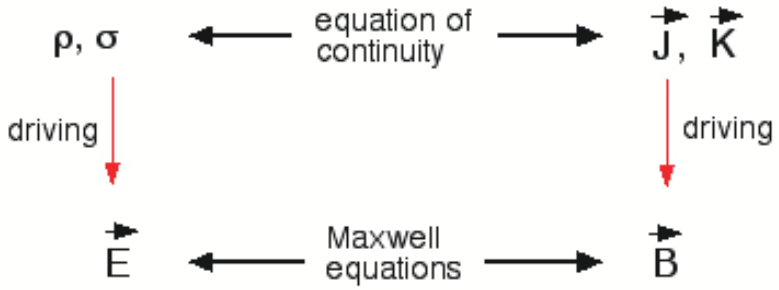
A wakefield is generated because the beam's image charges now have to move around a corner. Wakefields are the radiation fields of the image charges when their apparent trajectories are bent.[2]

Once accepting that wakefields are a result of radiation, then just like the case of any other radiation, it is natural to ask about 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 wide range, with wavelengths covering from microns to meters. To describe the frequency content of the wakefields, we introduce a quantity called *impedance*. Impedance is essentially the Fourier transform of wakefield.

Wakefield due to resistive wall

To discuss the resistive wall wakefield, let us first review the structure of electromagnetism by the following chart,

[2]These are apparent trajectories. Image charges do not physically move along the wall surface.



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

We then note a clear symmetry between the electric family and the magnetic family in the above chart. This symmetry, however, holds only in vacuum. It is lost when we consider a metal or an insulator. Metals break the symmetry by making a preference to the magnetic family $(\vec{B}, \vec{J})$, while insulators make a preference in favor of the electric family $(\vec{E}, \rho)$. No charges are allowed inside a metal while currents are allowed to penetrate. Inside a metal, therefore, there is more magnetic field than electric field. On the other hand, currents are not allowed inside an insulator, and there is more electric field than magnetic field.

In case of resistive wall, the wakefield is generated by the following physical process:

1. When the beam's image charges flow on the vacuum chamber wall, the electric field carried by the point charge will be terminated immediately by the image charges on the wall surface, while 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.

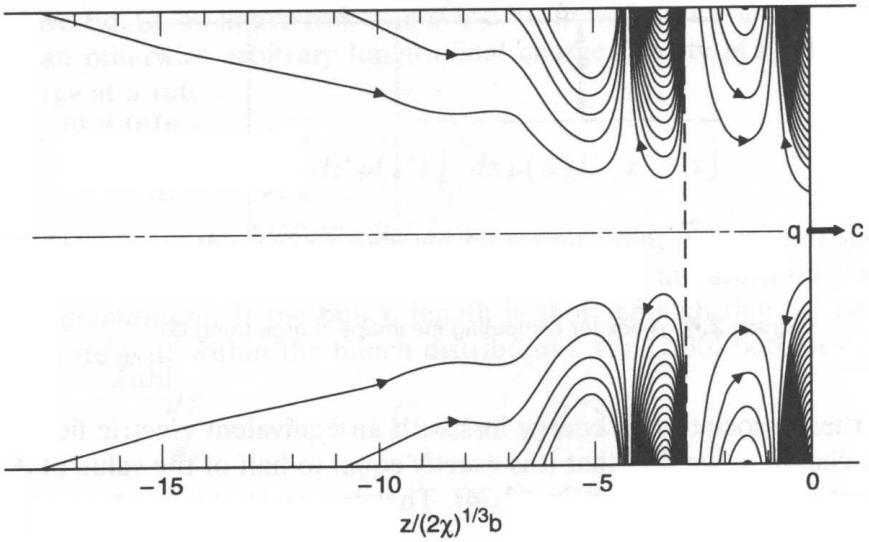
2. As the image current slowly re-surfaces after the point charge has past by, this re-surfacing image current drives new magnetic fields. These new magnetic fields occur after the point charge has left.

3. The resurfacing current and magnetic field will execute some transient behavior, and appear to oscillate a few times. After the initial transient, the resurfacing current and magnetic field decays away but at a very slow rate.

4. The re-surfacing changing magnetic field now drives an electric field by Maxwell equation. This yields some electric field inside the resistive wall after all, but this electric field is very weak.

In the case of a resistive wall pipe with circular cross section, an ultrarelativistic point charge q going down its axis will deposit a wakefield. The following

figure shows the electric component of the wakefield inside the vacuum chamber (outside of the wall material). Note that there is a matching magnetic field pattern following the leading point charge just like the electric field pattern does, and that both the electric and magnetic field patterns follow the leading point charge as a frozen pattern, indicating a phase velocity of c, but it is important to know that the Poynting vector does not indicate the field energy flows in the z-direction with the speed of light. The field energy does not propagate down the pipe. New field energy is deposited by the point charge as it moves down the pipe, while old energy gets transformed to heat by the resistivity on the pipe wall.



[The field line density is increased by a factor of 40 to the left of the dotted line.]

where χ is a small dimensionless parameter defined by

$$\chi = \frac{c}{4\pi\sigma b}$$

with b the vacuum pipe radiua, σ the conductivity of the pipe material. For example, if $b = 5$ cm and the wall is made of aluminum, we have $\chi = 1.5 \times 10^{-9}$.

As seen by the above figure, there is no wakefield ahead of the point charge, as causality would dictate. The wakefield pattern following the point charge is measured in distance z in units of $(2\chi)^{1/3}b$. Since $\chi \ll 1$, the resistive wall wakefield decays very quickly following the passage of the point charge.

On the other hand, after the quick initial decay, at long distances, the resistive wall wake starts to decay very slowly. This means the resistive wall

wakefield has a long tail. An intense beam bunch, for example, can leave a wakefield that lasts long enough to affect its motion when the bunch returns after making one complete circulation around a circular accelerator.

As will be shown later, the fact that the resistive wall generates both short-range and long-range wakefields is reflected by the fact that its corresponding impedance has an exceptionally wide spectrum, covering from very short to very long wavelengths.

What happens to particle motion when there are wakefields?

Consider a beam with distribution ψ in phase space $(\vec{q}, \vec{p})$. The dynamics of the evolution of ψ is described by the Vlasov equation (see later),

$$\frac{\partial \psi}{\partial t} + \frac{\vec{p}}{m} \cdot \frac{\partial \psi}{\partial \vec{q}} + \vec{f} \cdot \frac{\partial \psi}{\partial \vec{p}} = 0$$

where

$$\vec{f} = e\left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B}\right)$$
$$\vec{E} = \vec{E}_{\text{ext}} + \vec{E}_{\text{wake}}$$
$$\vec{B} = \vec{B}_{\text{ext}} + \vec{B}_{\text{wake}}$$

The wakefields are determined by the Maxwell equations where the source terms ρ and $\vec{j}$ are determined by the beam distribution ψ,

$$\rho = \int d^3p \ \psi, \qquad \vec{j} = \int d^3p \ \vec{v}\psi$$

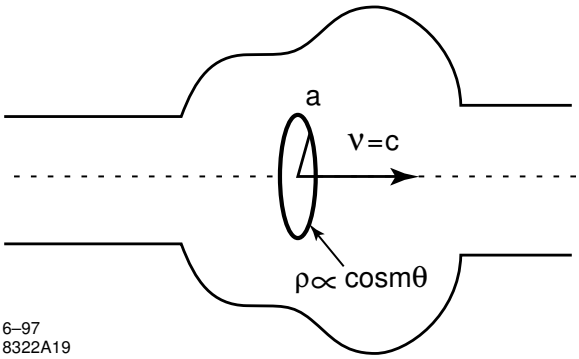
We therefore have the situation when the beam distribution is described by the Vlasov equation whose force terms are given by the electromagnetic fields, while the electromagnetic fields are described by the Maxwell equations whose source terms are given by the beam distribution. It is clear that a full treatment of the beam-wakefield system requires solving a coupled "Vlasov-Maxwell equation".

Beam-structure interaction is a difficult problem in general. Its solution often involves numerical solution using particle-in-cell codes with demanding boundary conditions. Applying PIC codes is reasonable for small devices such as electron guns and klystrons, but becomes impractical for large accelerators.

So, can we simplify the problem for our purpose while maintaining sufficiently accurate results? The answer is yes. For *high energy* accelerators, this complication can be avoided due to two simplifying approximations. These simplifications lead to the concepts of "wake function" and "impedance".

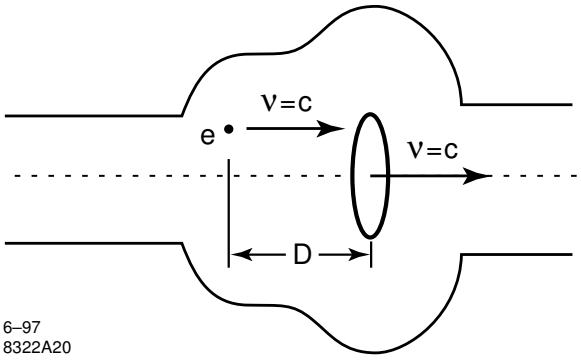
Rigid beam approximation The first simplification is the rigid beam approximation. At high energies, beam motion is little affected during the passage of a structure. This means one can calculate the wakefields assuming the beam shape is rigid and its motion is ultrarelativistic with $v = c$. In fact, we only need

to calculate the wakefields generated by a "rigid $\cos m\theta$ ring beam". Wakefield of a general beam can be obtained by superposition.



Impulse approximation The second simplification is the impulse approximation. We don't need to know $\vec{E}$ or $\vec{B}$ separately. We need only to know $\vec{f}$. For high energies, we don't even need the instantaneous $\vec{f}$. We only need the integrated impulse $\Delta\vec{p} = \int_{-\infty}^{\infty} dt\ \vec{f}$.

The following figure shows the configuration of a ring beam and a test charge that follows it. The ring beam generates a wakefield. The test charge receives a wake-induced impulse in the impulse approximation.



Panofsky-Wenzel Theorem

The instantaneous wakefields are complicated, but fortunately, $\Delta\vec{p}$ is much simpler and, at high energies, it is $\Delta\vec{p}$ that we need. The Panofsky-Wenzel theorem applies to $\Delta\vec{p}$. It is the basis of all beam instability analyses in high energy accelerators. In comparison, the PIC codes aim to calculate the instantaneous wakefields in all their details so in this regard is inefficient in their efforts.

Maxwell equations read

$$\begin{aligned}
&\nabla \cdot \vec{E} = 4\pi\rho \\
&\nabla \times \vec{B} - \frac{1}{c}\frac{\partial \vec{E}}{\partial t} = 4\pi\beta\rho\hat{z} \\
&\nabla \cdot \vec{B} = 0 \\
&\nabla \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} = 0
\end{aligned}$$

where we have made the important rigid beam approximation $\vec{j} = \rho\hat{v}$ and $\vec{v} = \beta c\hat{z}$.

The Lorentz force as seen by the rigid test charge e is given by

$$\vec{f} = e(\vec{E} + \beta\hat{z} \times \vec{B})$$

Both the beam and the test charge move with $\vec{v} = \beta c\hat{z}$. The impulse is

$$\Delta\vec{p}(x, y, D) = \int_{-\infty}^{\infty} dt\ \vec{f}(x, y, D + \beta ct, t)$$

Several important conditions can be found using the above Maxwell equations. One of them is the P-W theorem:

$$\nabla \times \Delta\vec{p} = \vec{0}$$

One can decompose the P-W theorem into a component parallel to $\hat{z}$ and a component perpendicular to $\hat{z}$ to obtain

$$\nabla \cdot (\hat{z} \times \Delta\vec{p}) = 0 \tag{1}$$

$$\frac{\partial}{\partial D}\Delta\vec{p}_\perp = \nabla_\perp \Delta p_z \tag{2}$$

The upper equation says something about the transverse components of $\Delta\vec{p}$. The lower equation says that the transverse gradient of the longitudinal wake impulse is equal to the longitudinal gradient of the transverse wake impulse.

Another important condition valid when $\beta = 1$ is

$$\nabla_\perp \cdot \Delta\vec{p}_\perp = 0 \tag{3}$$

It is clear that the Panofky-Wenzel theorem imposes strong constraints on the impulse received by a test charge from a relativistic beam.

Cylindrically symmetric pipe

In cylindrical coordinates, Eq.(1) gives

$$\begin{aligned}
&\nabla \cdot [\hat{z} \times (\Delta p_r \hat{r} + \Delta p_\theta \hat{\theta})] = 0 \\
\Longrightarrow \qquad &\frac{\partial}{\partial r}(r\Delta p_\theta) = \frac{\partial}{\partial \theta}\Delta p_r
\end{aligned}$$

Eq.(2) gives

$$\frac{\partial}{\partial D}(\Delta p_r \hat{r} + \Delta p_\theta \hat{\theta}) = \left(\hat{r}\frac{\partial}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial}{\partial \theta}\right)\Delta p_z$$

$$\Longrightarrow \quad \begin{cases} \frac{\partial}{\partial D}\Delta p_r = \frac{\partial}{\partial r}\Delta p_z \\ \frac{\partial}{\partial D}\Delta p_\theta = \frac{1}{r}\frac{\partial}{\partial \theta}\Delta p_z \end{cases}$$

Eq.(3) gives

$$\frac{1}{r}\frac{\partial}{\partial r}(r\Delta p_r) + \frac{1}{r}\frac{\partial}{\partial \theta}\Delta p_\theta = 0$$

$$\Longrightarrow \quad \frac{\partial}{\partial r}(r\Delta p_r) = -\frac{\partial}{\partial \theta}\Delta p_\theta \qquad (\beta = 1)$$

These results are surprisingly simple. They do not contain any beam source terms. Exact shape or distribution of the beam does not matter. Neither do they depend on the boundary conditions. The boundary can be perfectly conducting or resistive metal, or it can be dielectric. It does not have to be a sharply defined surface; it can for example be a gradually fading plasma surface. The only inputs needed are the Maxwell equations and the rigid-beam and the impulse approximations.

We are now ready to consider a $\cos m\theta$ ring beam with $\vec{v} = c\hat{z}$ as we set out to do by solving the three equations above. The solution can be expressed in terms of a function $W_m(D)$ such that

$$\begin{aligned} c\Delta\vec{p}_\perp &= -eI_m W_m(D) m r^{m-1}(\hat{r}\cos m\theta - \hat{\theta}\sin m\theta) \\ c\Delta p_z &= -eI_m W'_m(D) r^m \cos m\theta \end{aligned} \tag{4}$$

where I_m is the m-th multipole moment of the ring beam. $W_m(D)$ is the transverse wake function and $W'_m(D)$ the longitudinal wake function. The longitudinal wake function is simply derivative of the transverse wake function.

The solution (4) contains explicit dependences of r and θ. The fact that we can go so far without any specific details is surprising and shows the power of this line of analysis. The dependence on D is through $W_m(D)$ which can be obtained only if boundary conditions are invoked.

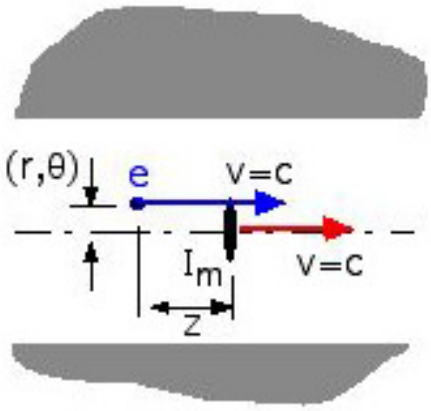
When the beam pipe is cylindrically symmetric, each m-multipole component of the beam excites a wake pattern according to (4). Different m's do not mix.

Decomposing wakefields into modes

Armed with the Panofsky-Wenzel theorem, to analyze the instability problem, we proceed as follows. We first consider the beam to be a δ-function in z. If the beam has a finite length, the result can be obtained by superposition.

We next decompose the transverse distribution into "modes" and consider a single transverse mode m. A general transverse distribution can be obtained by superposition with a summation over m.

So the problem is now reduced to finding the impulse integrated by a test charge that is trailing behind a beam slice with a transverse m-th moment I_m moving along the pipe axis. In this configuration, as shown below, I_m is the driving beam, e is the test charge, 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.



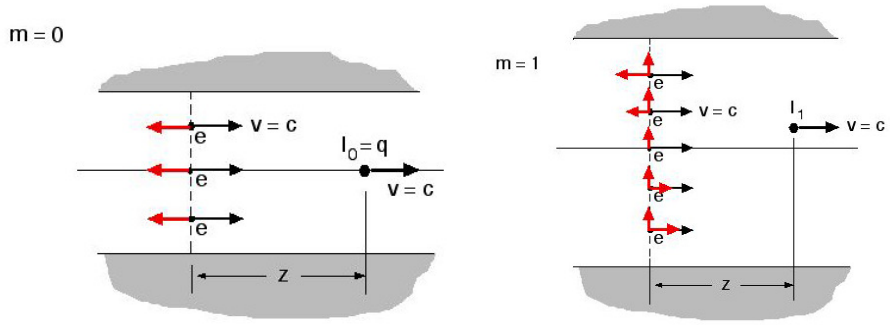
For a cylindrical pipe, the m-th multipole wakefield is driven when and only when the driving beam has an m-th moment.

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

In most applications, we care mostly about the $m = 0$ monopole mode when discussing longitudinal collective instabilities, and mostly about $m = 1$ when discussing transverse collective instabilities. Therefore, we mostly ask for

$W_0(z)$ and $W_1(z)$. The reason $W_0(z)$ is not relevant for transverse instabilities is because the transverse impulse vanishes when $m = 0$.

The wakefield impulses have simple patterns — the instantaneous wakefields do not share this simplicity. The $m = 0$ and $m = 1$ patterns are illustrated below:



Impedances

We mentioned that the wakefield wavelengths cover a wide range from ∼1 μm to ∼1 m. What characterize the frequency content of the wakefields are the impedances, the Fourier transforms of the wake functions,

$$Z_m^{\parallel}(\omega) = \int_{-\infty}^{\infty} \frac{dz}{c} e^{-i\omega z/c}\, W_m'(z)$$

$$Z_m^{\perp}(\omega) = i \int_{-\infty}^{\infty} \frac{dz}{c} e^{-i\omega z/c}\, W_m(z)$$

Since we have already discussed the wake functions, we consider this the definition of impedances.

Instead of wake functions, an accelerator designer therefore could alternatively ask about the impedance of the accelerator. The impedance is the quantity most directly related to the maximum beam current allowed by the accelerator. Inverting the Fourier transforms,

$$W_m'(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, e^{i\omega z/c}\, Z_m^{\parallel}(\omega)$$

$$W_m(z) = \frac{-i}{2\pi} \int_{-\infty}^{\infty} d\omega\, e^{i\omega z/c}\, Z_m^{\perp}(\omega)$$

The Panofsky-Wenzel theorem, which relates the longitudinal wake function to the derivative of the transverse wake function, also gives a relationship

between the longitudinal and transverse impedances for a given m,

$$Z_m^{\parallel}(\omega) = \frac{\omega}{c} Z_m^{\perp}(\omega)$$

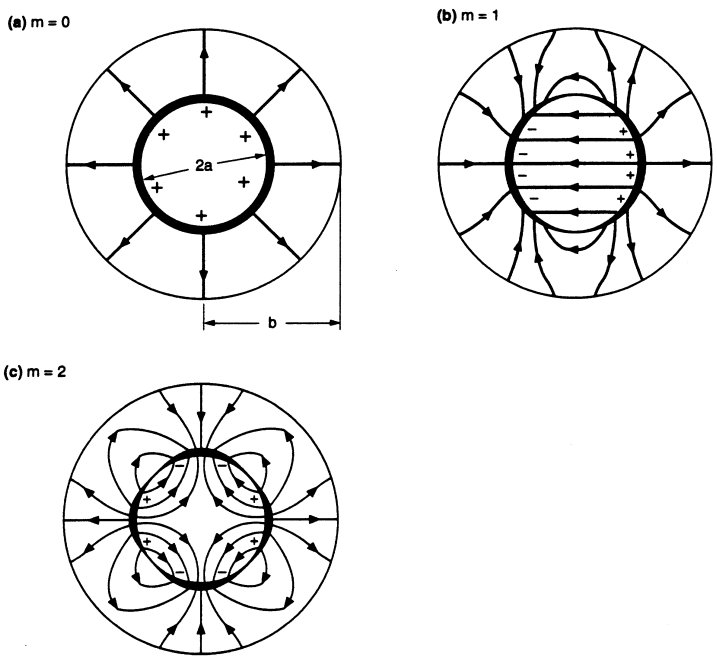
Some analytical examples of impedances and wake functions

We mentioned earlier that there are three ways when wakefields are generated.

1. the beam is not relativistic
2. the vacuum chamber is resistive
3. the vacuum chamber is not smooth

Three cases, each representing one of these three ways, that permit analytical expressions are given below.

Direct space charge This wakefield and impedance come about when the beam is not sufficiently relativistic. The following figure shows the space charge wakefields in the x-y plane driven by a ring-shaped, infinitely thin, $\cos m\theta$ beam.



The z-dependence is a δ-function. With a beam of radius a in a perfectly conducting round pipe of radius b and length L, we have

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

where $Z_0 = \sqrt{\mu_0/\epsilon_0} \approx 377\ \Omega$.[3] Due to the factor $1/\gamma^2$, space charge effects are most significant for low-to-medium energy proton or heavy ion accelerators.

The space charge impedance is purely imaginary, and is $\propto i\omega$ as if it is a pure inductance. However, its sign is as if it is a capacitance. By convention, we call it "capacitive".

Resistive wall Another case solvable analytically is for a round resistive pipe with radius b, conductivity σ_c, and length L. Defining the skin depth (change 0.066 to 0.086 for aluminum, and to 0.43 for stainless steel)

$$\delta_{\rm skin} = \sqrt{\frac{2c}{|\omega| Z_0 \sigma_c}}, \qquad \delta_{\rm skin}\,[\rm mm] = \frac{0.066}{\sqrt{f\,[\rm MHz]}} \text{ for copper}$$

one finds

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

The impedance is proportional to $(1-i)$, i.e. it is half resistive and half inductive.

The $|z|^{-1/2}$ dependence of $W_m(z)$ indicates that the resistive wall wakefield (especially its transverse component) decays slowly and typically lasts long after the beam passage, sometimes long enough for the beam to see its own wakefiled at its next revolution.

Slowly varying wall boundaries The third way to generate impedances is by discontinuities. Consider a case when the vacuum chamber (perfectly conducting) wall varies along the accelerator slowly, a perturbation technique can be applied. Specify the wall variation by $h(z)$ (cylindrically symmetric bump). At low frequencies $k = \omega/c < 1/(\text{bump length or width})$, the impedance is purely inductive — opposite in sign to space charge impedance,

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

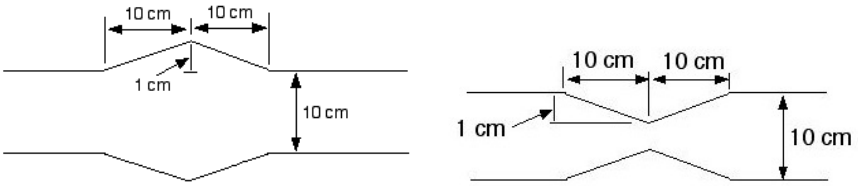
[3] Z_0 is the impedance of the vacuum. Yes, vacuum has impedance. An oscillating electromagnetic source will readily radiate into the vacuum. In fact, vacuum impedance is very large. A well designed accelerator will have impedance only a fraction of the vacuum impedance. Accelerators are poor radiators by design.

where

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

When the boundary varies rapidly, this formula breaks down. Numerical calculation has to be applied.

Homework Find the impedances of the following vacuum chamber geometries:



Beam energy spread in a linac

Consider a beam bunch traveling down the accelerator along the axis of the vacuum chamber pipe. The $m = 0$ wakefield excited by the beam produces a longitudinal force on particles in the beam. The main effect of this longitudinal force is a retarding voltage, causing energy changes of individual particles. As a result, there is a net energy loss of the beam to the wakefields. Furthermore, since not all particles in the bunch lose the same amount of energy, the wakefield also causes the beam to acquire an energy *spread*.

Consider first a one-particle model in which the beam bunch is a macroparticle of charge Ne. Traveling down the linac, it experiences the self-generated retarding longitudinal field and loses energy

$$\Delta E = -\frac{1}{2}Ne^2W_0'(0^-)$$

where the factor $\frac{1}{2}$ is due to the fundamental thoerem of beam loading.

Take the SLAC linac for example: $W_0'(0^-) = 7\ \text{cm}^{-1} \times L_0/L$, where $L_0 = 3$ km and $L = 3.5$ cm. We find $\Delta E = 2.2$ GeV for $N = 5 \times 10^{10}$.

This estimate can be improved by a two-particle model. The beam bunch is represented by two macroparticles, one leading and another trailing at a distance $|z|$ behind. The parasitic loss per particle in the leading macroparticle is 1.1 GeV due to its self-field. The trailing macroparticle loses, in addition to the 1.1 GeV due to self-field,

$$\Delta E = -\frac{1}{2}Ne^2W_0'(z)$$

due to the wakefield left behind by the leading macroparticle.

Take $z = -\sigma_z = -1$ mm, $N = 5\times10^{10}$, and $W_0'(-1\ \text{mm}) = 4.5\ \text{cm}^{-1}\times L_0/L$, each particle in the trailing macroparticle loses an additional 1.4 GeV. The net loss of a trailing particle is 2.5 GeV.

The one-particle model estimates a parasitic loss per particle of 2.2 GeV. The two-particle model estimates an average loss of $(1.1 + 2.5)/2 = 1.8$ GeV. The two-particle model has introduced an energy split of 1.4 GeV, or a 2.8% energy spread if the beam energy at the end of the linac is 50 GeV.

For linear colliders, this energy spread makes it difficult to focus the beam to a small spot at the collision point in a final focus system, and is to be avoided. Most of this spread can be removed by properly phasing the accelerating rf voltage relative to the beam.

One concern for a high-intensity linear collider can be described as follows. The energy spread at the end of the linac scales as

$$\frac{\Delta E}{E} \approx \frac{\frac{1}{2}Ne^2W_0'}{GL_0} \approx \frac{\frac{1}{2}Ne^2}{Gb^2}$$

where G is the acceleration gradient, and $W_0' \approx L_0/b^2$ is the longitudinal wake function, where b is the vacuum chamber radius characterizing the size of the accelerating cavities. On the other hand, the efficiency of energy extraction by the beam from the field energy U stored in the accelerating cavities [$U \approx \frac{1}{8\pi}(G/e)^2 \times \pi b^2 L_0$] is given by

$$\text{extraction efficiency} \approx \frac{NE}{U} \approx \frac{8Ne^2}{Gb^2}$$

which is equal to 16 times the energy spread. In other words, to improve the energy spread of the beam at the end of the linac necessarily requires sacrificing the energy extraction efficiency. One way to ameliorate this problem is to compensate $\Delta E/E$ by phasing the rf voltage. Another way is to send a *train* of M bunches per filling of the rf cavities. This will increase the energy extraction efficiency by a factor of M, although at the cost of having to deal with the multibunch interactions due to the long range wakefields.

We now depart from the simplified models and consider a bunch with a general longitudinal distribution $\rho(z)$. The energy change for a test charge e at longitudinal position z can be written as $eV(z)$, where

$$V(z) = -\int_z^\infty dz' \rho(z') W_0'(z - z')$$

or equivalently

$$V(z) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega e^{i\omega z/c} Z_0^{\parallel}(\omega)\tilde{\rho}(\omega)$$

A negative $V(z)$ means the test charge loses energy from the wakefield. An additional integration of $V(z)$ over the bunch then gives the total parasitic loss,

$$\Delta\mathcal{E} = \int_{-\infty}^{\infty} \rho(z)V(z)dz$$

For a bunch with Gaussian longitudinal distribution and uniform disk transverse distribution, for example, the energy spread due to space charge effect is

$$\begin{aligned} \frac{V(z)}{L} &= \sqrt{\frac{2}{\pi}}\frac{q}{\gamma^2\sigma_z^2}\left(\ln\frac{b}{a}+\frac{1}{2}\right) f\left(\frac{z}{\sigma_z}\right) \\ f(u) &= ue^{-u^2/2} \end{aligned}$$

Generally, particles in the front of the bunch ($z > 0$) lose energy due to wake fields, while particles in the back of the bunch ($z < 0$) may gain or lose energy, depending on the length of the bunch. This is not true for the special case of the space charge effect, for which particles in the front of the bunch gain energy, and particles in the back of the bunch lose energy. For the space charge effect, the energy gained by the bunch head is necessarily given up by the bunch tail so that the net energy of the bunch is unchanged.

Consider a numerical example of a 50 MeV proton transport line. If we take $q = 10^{10}e$, $\sigma_z = 3$ cm, $a = 2$ cm, and $b = 5$ cm, we obtain a longitudinal space charge force of ± 6 V/m for particles located at $z = \pm\sigma_z$. The net energy change of these particles after traveling 100 m of this transport line is $eV/\beta = \pm 2$ keV. The space charge induced beam energy spread is therefore $\pm 4 \times 10^{-5}$.

For a resistive wall, we have

$$\begin{aligned} \frac{V(z)}{L} &= \frac{q}{4b\sigma_z^{3/2}}\sqrt{\frac{c}{2\pi\sigma}} f\left(\frac{z}{\sigma_z}\right) \\ f(u) &= -|u|^{3/2}e^{-u^2/4}\left[(I_{-1/4}-I_{3/4})\,\mathrm{sgn}(u) - I_{1/4} + I_{-3/4}\right] \end{aligned}$$

with the Bessel functions $I_{\pm 1/4}$ and $I_{\pm 3/4}$ evaluated at $u^2/4$. Continuing the above numerical example, assuming an aluminum pipe, a particle located at $0.5\sigma_z$ ahead of bunch center loses an energy of 0.1 eV after traveling 100 m, and a particle located at $1.8\sigma_z$ behind the bunch center gains 0.04 eV.

Beam breakup in a linac

In the previous section, the beam was centered in the vacuum chamber pipe. There were no transverse wake forces ($m = 0$). In case the beam is executing a betatron oscillation, an $m = 1$ dipole wakefield is excited by the bunch head, which causes transverse deflection of the bunch tail. For a high-intensity beam, this leads to a transverse breakup of the beam. The first observation of beam breakup was made on the SLAC linac.

To proceed with a simplified macroparticle model, we first note that a one-particle model is not useful because a point charge does not exert a transverse wake force on itself. In the two-particle model, the leading macroparticle, unperturbed by its own transverse wakefield, executes a free betatron oscillation

$$y_1(s) = \hat{y}\cos k_\beta s$$

The trailing macroparticle, at a distance $|z|$ behind, sees a deflecting wakefield left behind by its leading partner,

$$\begin{aligned} y_2'' + k_\beta^2 y_2 &= -\frac{Ne^2 W_1(z)}{2EL}\, y_1 \\ &= -\frac{Nr_0 W_1(z)}{2\gamma L}\, \hat{y} \cos k_\beta s \end{aligned}$$

where $W_1(z)$ is the transverse wake function per cavity period L. We have ignored acceleration of the beam energy. For the SLAC linac, $k_\beta \approx 0.06$ m^{-1} and $k_\beta L \approx 0.002$.

The solution is

$$y_2(s) = \hat{y}\Big[\cos k_\beta s - \frac{Nr_0 W_1(z)}{4k_\beta \gamma L}\, s \sin k_\beta s\Big]$$

The first term describes the free oscillation and the second term is the resonant response to the driving wake force. The amplitude of the second term grows linearly with s. The mechanism of beam breakup is that particles in the tail of the beam are driven exactly on resonance by the oscillating wake left by the head of the beam.

At the end of the linac, the oscillation amplitude of the bunch tail relative to the bunch head is characterized by the dimensionless growth parameter

$$\Upsilon = -\frac{Nr_0 W_1(z) L_0}{4k_\beta \gamma L}$$

where L_0 is the total linac length.

For a beam bunch with realistic distribution, the bunch is distorted into a banana shape. The motion of the bunch head is $\cos k_\beta s$, while the deviation of the bunch tail relative to the bunch head is $s \sin k_\beta s$. When the bunch head is at a maximum displacement, the tail lines up with the bunch head, but when the bunch head displacement is zero, the tail swing is maximum. As the beam propagates down the linac, the swing amplitude of the flapping tail increases with s until the tail breaks up and particles are lost. Note that the sign of the tail swing shown is not arbitrary, because $\Upsilon > 0$.

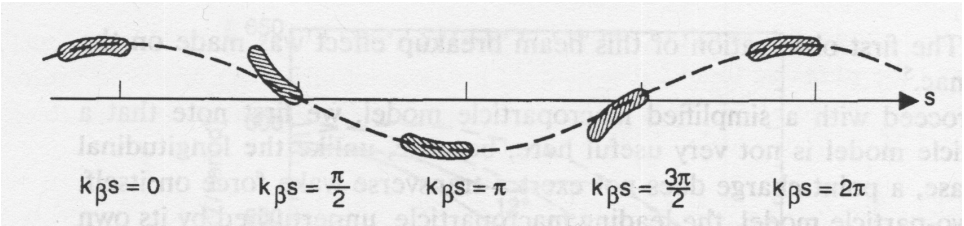
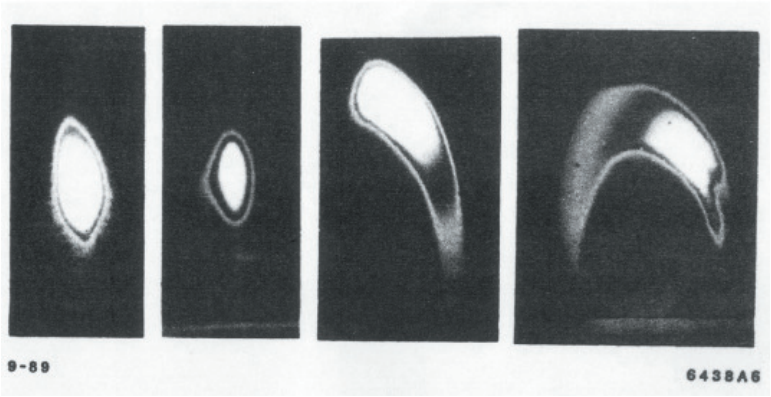


Figure below shows four transverse beam profiles observed at the end of the SLAC linac with $N = 2 \times 10^{10}$. The leftmost profile for a carefully steered

beam. When the beam was injected off center by 0.2, 0.5, and 1 mm, the beam profiles are as shown successively to the right. The beam sizes σ_x and σ_y were $\sim$120 μm.



(Courtesy John Seeman)

So far we have ignored beam acceleration, which has an important stabilizing effect because, as its energy increases, the beam becomes more rigid and less vulnerable to the wake fields. Repeating a similar analysis but taking into account of acceleration yields the growth parameter

$$\Upsilon = -\frac{Nr_0 W_1(z) L_0}{4k_\beta \gamma_f L} \ln \frac{\gamma_f}{\gamma_i}$$

which is basically simply replacing the factor L_0/γ by its integral counterpart $\int_0^{L_0} ds/\gamma(s)$. Due to acceleration, the tail amplitude thus grows logarithmically rather than linearly with s, and the growth parameter is much reduced. If the beam is accelerated in the SLAC linac from 1 to 50 GeV, the factor Υ becomes 14, instead of 180, which was calculated earlier for a beam coasting at 1 GeV.

The beam breakup instability described above is quite severe even with acceleration. To control it, the beam has to be tightly focused, rapidly accelerated, and carefully injected, and its trajectory carefully steered down the linac. Interestingly, the contribution from trajectory missteering can in principle be largely compensated by an intentional misinjection.

It turns out, however, that there is another interesting and effective method to ameliorate the situation. This method, known as the *BNS damping* after Balakin, Novokhatsky, and Smirnov, is described next.

Consider first the case without acceleration, where the leading macroparticle executes a free betatron oscillation. The idea of BNS damping requires introducing a slightly stronger betatron focusing of the bunch tail than the bunch head. The equation of motion of the tail particles can be written as

$$y_2'' + (k_\beta + \Delta k_\beta)^2 y_2 = -\frac{Nr_0 W_1(z)}{2\gamma L} \hat{y} \cos k_\beta s$$

The solution, assuming $|\Delta k_\beta/k_\beta| \ll 1$, is

$$y_2(s) = \hat{y}\cos(k_\beta + \Delta k_\beta)s + \frac{Nr_0W_1(z)}{4k_\beta\Delta k_\beta\gamma L}\,\hat{y}\big[\cos(k_\beta s + \Delta k_\beta s) - \cos k_\beta s\big]$$

Compared with the result without Δk_β, one observes that, by introducing a slightly different focusing strength for the bunch tail, the beam breakup mechanism of the bunch head resonantly driving the bunch tail is removed. A further inspection shows that there exists a magical condition for the bunch tail to follow the bunch head exactly for all s, namely

$$\frac{Nr_0W_1(z)}{4k_\beta\Delta k_\beta\gamma L} = -1$$

or equivalently,

$$\frac{\Delta k_\beta}{k_\beta} = -\frac{Nr_0W_1(z)}{4k_\beta^2\gamma L} = \frac{\Upsilon}{k_\beta L_0} \tag{5}$$

where Υ is defined before, and $k_\beta L_0$ is the total betatron phase advance of the linac. For short bunches, Υ and Δk_β are positive; the betatron focusing required to fulfill the BNS condition is therefore stronger at the bunch tail than at the bunch head.

Under the BNS condition, $y_2(s) = y_1(s) = \hat{y}\cos k_\beta s$, and the beam no longer breaks up.[4] Physically, this happens because the additional external focusing force introduced for the bunch tail has compensated for the defocusing dipole deflection force due to the wakefield left behind by the bunch head. Note that the BNS focusing has to be adjusted according to the beam intensity.

There are different ways to provide the BNS focusing. One is to introduce a radio frequency quadrupole whose strength changes as the bunch passes by so that the head and tail of the bunch see different quadrupole strengths. Another is to choose the location of the bunch relative to the acceleration rf voltage in such a way that the bunch tail acquires a lower energy than the bunch head. The energy spread across the bunch then causes a spread in betatron focusing according to

$$\frac{\Delta k_\beta}{k_\beta} = \xi\frac{\Delta E}{E}$$

where ξ is the chromaticity determined by the linac design. For a FODO lattice design, for example,

$$\xi = -\frac{2}{\mu}\tan\frac{\mu}{2}$$

where μ is the betatron phase advance per FODO cell. By properly choosing the phase of the rf voltage relative to the beam bunch, the betatron focusing required by the BNS condition can be obtained, provided the required $\Delta k_\beta/k_\beta$ is not excessive.

[4]The mechanism of BNS damping is not to be confused with that of Landau damping, to be discussed later. They have little in common other than the fact that both involve a frequency spread in the bunch population.

In case of an accelerated beam, the BNS condition is still given by Eq.(5), except that the parameter Υ is now that given by the case with acceleration. Take the SLAC linac, for example, and assume $\mu = 90^o$; then the energy deviation of the bunch tail from the bunch head required by the BNS condition is about -5.5%. BNS damping has been routinely employed to control the beam breakup instability in the SLAC linac operations.

Parasitic heating

When a beam bunch of charge q and line density $\lambda(t)$ traverses an impedance $Z_0^{\parallel}(\omega)$, it loses energy to the impedance. This *parasitic energy loss* (or HOM heating, HOM means "higher order mode") is

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

where $\kappa^{\parallel}$ is the *loss factor*, in units of V/pC,

$$\kappa^{\parallel} = \frac{1}{\pi}\int_0^{\infty} d\omega\, \mathrm{Re} Z_0^{\parallel}(\omega)\, |\tilde{\lambda}(\omega)|^2 \tag{6}$$

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

Only the real part of the impedance contributes to the parasitic loss. The space charge or the slowly varying wall impedances do not cause 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 the bunch head is recovered by the bunch tail.

For resistive wall,

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

where b is the pipe radius (assumed cylindrically symmetric). It shows explicitly that parasitic loss is more pronounced for short bunches.

Parasitic loss gives rise to heating of the vacuum chamber wall where there are impedances. In high intensity electron storage rings, the beam position monitors or bellows can heat up. This is especially serious for short bunches.

Most of the parasitic loss occurs as the beam traverses a discontinuity structure. Part of the wakefield gets trapped if the structure is cavity-like and if the wakefield frequency is below the cutoff frequency of the pipe. The 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 up and down the pipe and eventually dissipates on lossy material elsewhere in the vacuum chamber. For a cavity structure, $\kappa^{\parallel}$ is given by a sum over cavity modes below cut-off, plus a contribution above cut-off. Each cavity mode below cut-off

contributes a resonator impedance, with

$$\kappa^{\parallel} \approx \begin{cases} \dfrac{\omega_r R_s}{2Q_r} e^{-\omega_r^2\sigma^2} & \text{high-}Q\text{ resonator} \\ \dfrac{\omega_r R_s}{2Q_r} & \text{low-}Q\text{ resonator, short bunch } \omega_r\sigma \ll 1 \\ \dfrac{R_s}{4\sqrt{\pi}Q_r^2\omega_r^2\sigma^3} & \text{low-}Q\text{ resonator, long bunch } \omega_r\sigma \gg 1 \end{cases}$$

Above cut-off, the impedance per cavity can be represented by the diffraction model,

$$Z_0^{\parallel}(\omega) = [1 + \mathrm{sgn}(\omega)i]\frac{Z_0}{2\pi^{3/2}}\frac{1}{b}\sqrt{\frac{cg}{|\omega|}}$$

where g is the gap size of the cavity. This impedance has both real and imaginary parts.

For a single bunch in a circular accelerator, the integral in Eq.(6) is replaced by an infinite sum,

$$\kappa^{\parallel}(\sigma) = \frac{\omega_0}{2\pi}\sum_{p=-\infty}^{\infty} Z_0^{\parallel}(p\omega_0)\,|\tilde{\lambda}(p\omega_0)|^2$$

For short bunches in large machines ($\omega_0 \ll 1/\sigma$), the sum can be replaced by an integral, and the difference between single passes and multiple passes disappears.

The parasitic loss by the beam goes into wakefields. Typically, only a small fraction of the particle energy becomes 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 the ultimate culprit for the various collective instabilities.

The Vlasov equation

The Vlasov equation describes the collective behavior of a multiparticle system under the influence of electromagnetic forces. To construct the Vlasov equation, one starts with the single-particle equations of motion (assume 1-D)

$$\begin{aligned} \dot{q} &= f(q,p,t) \\ \dot{p} &= g(q,p,t) \end{aligned}$$

The state of a particle at a given time t is represented by a point in the phase space (q,p). The motion of a particle is described by the motion of its representative point in phase space.

In a conservative deterministic system, the particle trajectory in phase space is completely determined by the initial conditions (q_0,p_0) at time $t = t_0$. Two particles having the same initial conditions must have exactly the same trajectory in phase space. It follows that the only way for two trajectories to meet

at a given time is for them to coincide at all times. In other words, trajectories either completely coincide or never intersect.

Consider a distribution of particles occupying an area in the phase space. Because they cannot intersect with particles on the boundary of the distribution as the distribution evolves in time, particles inside cannot leak out of and particles outside can not enter the distribution.

If the system is conservative,

$$f = \frac{\partial H}{\partial p} \quad \text{and} \quad g = -\frac{\partial H}{\partial q}$$

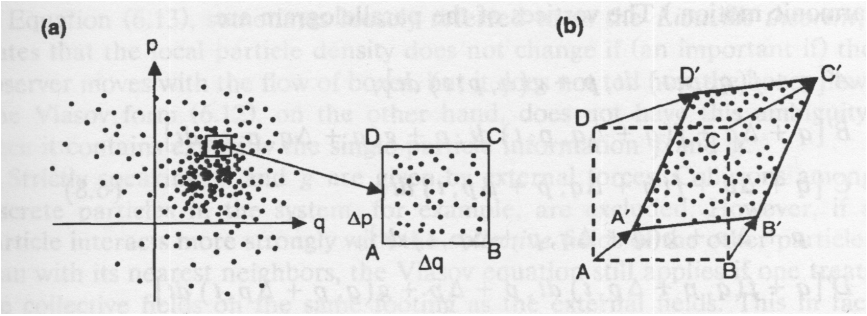
where H is the Hamiltonian. It follows that

$$\frac{\partial f}{\partial q} + \frac{\partial g}{\partial p} = 0$$

As will be seen later, this condition leads to an area preservation property: as the particle distribution evolves in the phase space, its shape may be distorted but its area remains constant. In fact, in a nonconservative system, $\frac{\partial f}{\partial q} + \frac{\partial g}{\partial p}$ has the physical meaning of the rate of area shrinkage.

Consider a distribution of a group of particles in the phase space at time t. A rectangular $\Delta q\, \Delta p$ box is drawn:

$$A(q, p),$$
$$B(q + \Delta q, p),$$
$$C(q + \Delta q, p + \Delta p),$$
$$D(q, p + \Delta p).$$



At a later time, $t + dt$, the box moves and deforms into a parallelogram $A'B'C'D'$ with the same area as $ABCD$. All particles inside the box move with the box. Let the number of particles enclosed by the box be

$$\psi(q, p, t)\, \Delta q\, \Delta p$$

where ψ is the phase space distribution density normalized by

$$\int_{-\infty}^{\infty} dq \int_{-\infty}^{\infty} dp\, \psi(q, p, t) = N$$

The vertices of the parallelogram are

$$
\begin{aligned}
&A'[q + f(q,p,t)\,dt, p + g(q,p,t)\,dt],\\
&B'[q + \Delta q + f(q+\Delta q,p,t)\,dt, p + g(q+\Delta q,p,t)\,dt],\\
&C'[q + \Delta q + f(q+\Delta q,p+\Delta p,t)\,dt, p + \Delta p + g(q+\Delta q,p+\Delta p,t)\,dt],\\
&D'[q + f(q,p+\Delta p,t)\,dt, p + \Delta p + g(q,p+\Delta p,t)\,dt].
\end{aligned}
$$

The condition that no particles leaking into or out of the box gives

$$\psi(q,p,t)\ \text{area}(ABCD) = \psi(q + f dt, p + g dt, t + dt)\ \text{area}(A'B'C'D')$$

For a Hamiltonian system, the area of the box is conserved:

$$
\begin{aligned}
\text{area}(A'B'C'D') &= |\overrightarrow{A'B'} \times \overrightarrow{A'D'}| \\
&= \Delta q\,\Delta p \left[1 + \left(\frac{\partial f}{\partial q} + \frac{\partial g}{\partial p}\right) dt\right] \\
&= \Delta q\,\Delta p = \text{area}(ABCD)
\end{aligned}
$$

We then have

$$
\begin{aligned}
\psi(q,p,t) &= \psi(q + f\,dt, p + g\,dt, t + dt) \\
&= \psi + \frac{\partial \psi}{\partial q} f\,dt + \frac{\partial \psi}{\partial p} g\,dt + \frac{\partial \psi}{\partial t}\,dt
\end{aligned}
$$

or, after canceling out ψ on both sides, we obtain the *Vlasov equation*

$$\frac{\partial \psi}{\partial t} + f\frac{\partial \psi}{\partial q} + g\frac{\partial \psi}{\partial p} = 0$$

Vlasov equation can also be put in a much more vague form

$$\frac{d\psi}{dt} = 0, \quad \text{or} \quad \psi = \text{const in time.}$$

Sometimes loosely referred to as the *Liouville theorem*, it states that the local particle density does not change if (an important if) the observer moves with the flow of boxes, but it does not tell how the boxes flow. The Vlasov form, in contrast, does not have this ambiguity, since it contains explicitly the single-particle information f and g.

Strictly speaking, f and g are given by external forces. Collisions among discrete particles in the system, for example, are excluded. However, if a particle interacts more strongly with the *collective* fields of the other particles than with its nearest neighbors, the Vlasov equation still applies if one treats the collective fields on the same footing as the external fields. This in fact forms the basis of treating the collective instabilities using the Vlasov technique.

One special case where the Vlasov equation can be solved exactly is when the system is described by a Hamiltonian $H(q,p)$ which does not have an explicit time dependence. A stationary solution is found to be

$$\psi(q,p) = \text{any function of } H(q,p)$$

In this system, individual particles stream along constant-H contours in the phase space in such a way that the overall distribution is stationary.

In the derivation of the Vlasov equation, we have assumed there are no diffusion or external damping effects. This is usually a good approximation for proton beams. For electron beams, synchrotron radiation contributes to both damping and diffusion, and one needs to apply the *Fokker-Planck equation*, a generalization of he Vlasov equation. However, when the instability occurs faster than the damping or diffusion times, the Vlasov treatment applies also to electrons.

Potential-well distortion

As a first application of the Vlasov technique, we study the effect of longitudinal wakefield on a distortion of the equilibrium shape of a beam bunch. The mechanism is a static one; no part of the beam bunch is executing collective oscillation. The extent of distortion depends on the beam intensity; higher beam intensities cause larger distortions.

Consider a bunched beam that travels along the axis of the vacuum chamber pipe in a circular accelerator. We assume the beam does not have any transverse dimension, i.e., the beam is an infinitesimally thin thread. Such a beam does not generate transverse wakefields; only the $m = 0$ wake is excited.

Consider a particle in the beam executing longitudinal synchrotron oscillation. The phase space coordinates q and p are

$$q = z \quad \text{and} \quad p = -\frac{\eta c}{\omega_s}\delta$$

where η is the slippage factor defined by the accelerator lattice, ω_s is the synchrotron oscillation frequency.

The single-particle equations of motion are

$$z' = -\eta\delta \quad \text{and} \quad \delta' = K(z)$$

We leave $K(z)$ open for now, except that we do know it cannot depend on δ, because the system is conservative.

The Vlasov equation reads

$$\frac{\partial\psi}{\partial s} - \eta\delta\frac{\partial\psi}{\partial z} + K(z)\frac{\partial\psi}{\partial\delta} = 0$$

where we will set $\partial\psi/\partial s = 0$, since we are looking for a stationary distribution. The general stationary solution is

$$\psi(z,\delta) \quad = \quad \text{any function of the Hamiltonian } H\,,$$

$$H = \frac{\eta^2 c^2}{\omega_s}\left[\frac{\delta^2}{2} + \frac{1}{\eta}\int_0^z K(z')\,dz'\right]$$

The second integral term in the Hamiltonian is the potential-well term. A simple harmonic system would have a parabolic potential well.

If the potential well is provided by an external rf voltage $V_{\rm rf}(z)$, we have

$$K(z) = \frac{eV_{\rm rf}(z)}{CE} = \frac{\omega_s^2}{c^2\eta V'_{\rm rf}(0)}V_{\rm rf}(z)$$

A practical case is given by $V_{\rm rf} = \hat{V}\sin(\omega_{\rm rf}z/c)$. The deviation of $V_{\rm rf}(z)$ from a linear dependence on z is a cause of potential-well distortion. The general stationary distribution is given by any function of the Hamiltonian

$$H = \frac{\eta^2 c^2}{2\omega_s}\delta^2 + \frac{\omega_s c^2}{\omega_{\rm rf}^2}\left[1 - \cos\left(\frac{\omega_{\rm rf}z}{c}\right)\right]$$

This Hamiltonian also describes the form of the rf bucket. A stationary distribution must conform to the contours of constant Hamiltonian inside the bucket. For small oscillation amplitudes, we have $K = \omega_s^2 z/\eta c^2$, the case of simple harmonic motion.

One noteworthy special case of the stationary beam distribution is that given by $\exp(-\text{const}\times H)$. This distribution is always Gaussian in δ. In case the bunch length is much shorter than the rf wavelength, ($z \ll c/\omega_{\rm rf}$) the familiar quadratic form of the Hamiltonian is reestablished, and the distribution is also Gaussian in z. As the bunch length increases, the bunch shape deviates from Gaussian; the potential well is distorted by the rf bucket, although the distribution remains Gaussian in δ.

There is another reason for the Hamiltonian to deviate from the quadratic form, and thus to cause potential-well distortion, namely, the wakefields. Consider a bunch that is short compared with the rf wavelength. Let the wakefields be $W'_0(z)$ integrated over the accelerator circumference, and assume that the wake has dissipated before the beam completes one revolution,

$$K(z) = \frac{\omega_s^2}{\eta c^2}z - \frac{r_0}{\gamma C}\int_z^\infty dz'\,\rho(z')W'_0(z-z')$$

The corresponding Hamiltonian is

$$H = \frac{\eta^2 c^2}{2\omega_s}\delta^2 + \frac{\omega_s}{2}z^2 - \frac{\eta c^2 r_0}{\omega_s\gamma C}\int_0^z dz''\int_{z''}^\infty dz'\,\rho(z')W'_0(z''-z')$$

The stationary solution to the Vlasov equation must be a function of H. The complication here is that the complicated z-dependence of H now involves the beam density ρ, which in turn is determined by the stationary distribution itself. Clearly some self-consistency requirement is to be imposed.

Continuing the Gaussian example, the stationary distribution maintains its gaussian distribution in δ,

$$\psi(z,\delta) = \frac{1}{\sqrt{2\pi}\sigma_\delta} \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right) \rho(z)$$

The Gaussian form and the value of σ_δ are arbitrary if the collective behavior is governed by the Vlasov equation, as in the case of a proton beam. However, if the beam behavior is governed, as for an electron beam, by the Fokker-Planck equation, then this Gaussian distribution with a specific value for σ_δ will be the unique solution of the stationary beam distribution.

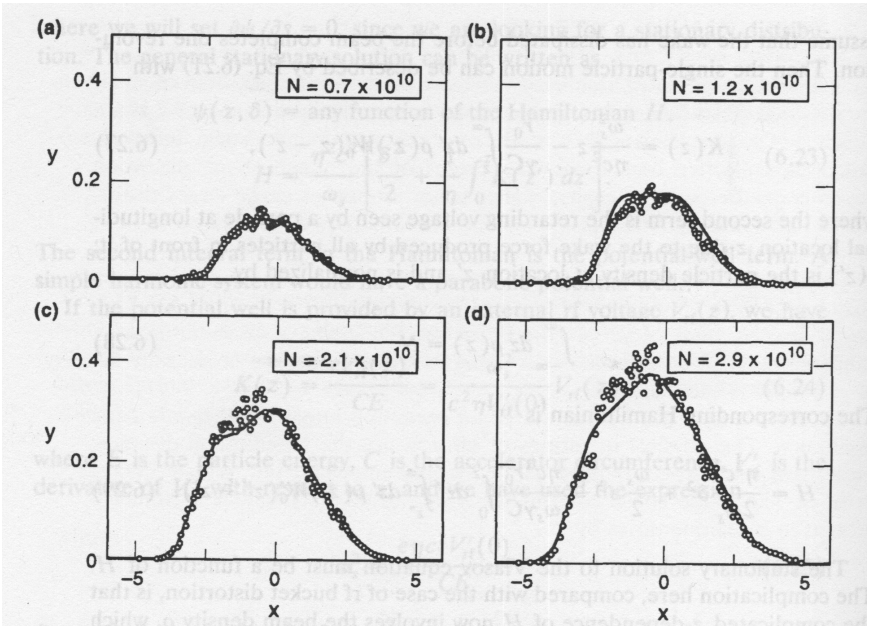
This distribution matches the stationary solution

$$\psi(z,\delta) \propto \exp\left(-\frac{\omega_s}{\eta^2 c^2 \sigma_\delta^2} H\right)$$

Self-consistency then imposes a transcendental equation for $\rho(z)$, called the*Haissinski equation*,

$$\rho(z) = \rho(0) \exp\left[-\frac{1}{2}\left(\frac{\omega_s z}{\eta c \sigma_\delta}\right)^2 + \frac{r_0}{\eta \sigma_\delta^2 \gamma C} \int_0^z dz'' \int_{z''}^\infty dz'\, \rho(z') W_0'(z'' - z')\right]$$

In the limit of zero beam intensity, the solution reduces to the bi-Gaussian form, where $\sigma_z = \eta c \sigma_\delta / \omega_s$. For high beam intensities, $\rho(z)$ deforms from Gaussian. The Haissinski equation is solved numerically for $\rho(z)$ once $W_0'(z)$ is known and σ_δ specified. Figure below shows the result for the electron damping ring for the SLAC Linear Collider. The bunch shape is Gaussian at low beam intensities, and it distorts as the beam intensity is increased. The calculations agree with the measurements.



The horizontal axis is $x = -z/\sigma_{z0}$, where σ_{z0} is the unperturbed rms bunch length. The vertical scale gives $y = 4\pi e\rho(z)/V'_{\rm rf}(0)\sigma_{z0}$. (Courtesy Karl Bane, 1992.)

Note that the distribution leans forward ($z > 0$) as the beam intensity increases. This effect comes from the parasitic loss of the beam bunch, and is a consequence of the real (resistive) part of the impedance. Since the SLC damping ring is operated above transition, the bunch moves forward so that the parasitic energy loss can be compensated by the rf voltage.

Note also that the bunch length increases as the beam intensity increases. The bunch shape distortion comes mainly from the imaginary part of the impedance. That the bunch lengthens is a consequence of the fact that the imaginary part of the impedance is mostly inductive.