October 21, 1963

BEAM LOADING AND BEAM BLOWUP IN ELECTRON LINACS

J. E. Leiss
National Bureau of Standards

Our field is electron linear accelerators, where heavy beam loading is a common thing. Also, after we once get started, we have only a part of a phase oscillation. Thus in many ways, the situation with which we deal is quite different from that in the proton linac. However, there is some work we have been doing recently which is related to the things that are being discussed. I will direct my comments at some design similar to the Yale design, in order to try and give you an idea of the size of effects. If we talk about the mixed machine where we have an Alvarez structure followed by a travelling wave structure, we can look at some of the consequences.

We are talking about 1 mA of protons at about 1 GeV which is 1 MW of beam power, with a duty cycle of about 5%, so the instantaneous beam power is about 20 MW. There are some 60 sections with about 2 MW rf drive each, or 120 MW total peak power. At the frequencies that are being discussed, if one rides on the crest of the rf wave, and all things are optimized, the maximum power conversion efficiency to the beam is something like 50%. Now if you run roughly 30% down from the peak for phase stability reasons, it turns out that the power conversion efficiency varies as the square of the cosine of the phase difference, so that the maximum theoretical conversion efficiency is something like 25-30% and, as I will indicate, there are other factors that enter. The point is that the meson factories that are being discussed are very heavily beam-loaded in the travelling wave part of the accelerator, even by electron linac standards, and the consequences of this are going to be very important in the design of the machine.

I will try to survey a few of the various things that are known about beam loading that have probably not been covered to date in proton linac studies. There are a number of quite detailed calculations about the dispersive properties of linacs which I will not try to go through in any detail; however, I will summarize some of the results which seem pertinent.

The way people usually approach beam loading in electron linacs is to start with the power diffusion equation:

$$\frac{dP}{dz} = -2IP - iE \tag{1}$$

where P is the power density, I is the field attenuation factor, and i the current, thus iE is the term for the beam loading. You also need the shunt impedance, r:

$$r = \frac{E^2}{2IP} \qquad (2)$$

These can be converted to an equation in E

$$\frac{dE}{dz} = -IE - Iir \qquad (3)$$

where i is the instantaneous beam current averaged over the rf structure in the beam. From this, one gets the electric field, E, as a function of z

$$E(z) = E_o e^{-Iz} - ir(1 - e^{-Iz}) \qquad (4)$$

Equation (4) is a steady state equation. E_o is the peak electric field at the beginning of the waveguide. It is multiplied by an attenuation factor and reduced by a beam loading term. This can clearly be thought of as consisting of two fields: the usual term from the power source looks like $E_o e^{-Iz}$, the beam loading term is $ir(1 - e^{-Iz})$ and the sum of these fields is the field which the particles in the guide see. This assumes that the particles ride on the crest of the rf wave.

If you now try to put a non-zero beam phase, φ, into Eq. (4) you systematically come out wrong. If you work it out for the condition in which the klystron has been turned off, and ask what the phase of the rf is, it says that the klystron phase is known when the power isn't on, which is clearly nonsense. The point is that the wave which is caused by the beam loading is 180^o out of phase with the beam and there's nothing one can do about it. The analytic reason that one gets the wrong answer is that we started with the power equation, which is quadratic in E, and the superposition theorem can't be used in such an equation. You should start out with the equation in the field. If this is done correctly, you get a term in cos φ where φ is the bunch phase relative to the peak electric field. Likewise, the energy gain of a particle, obtained by simply integrating the E(z) equation over the length of the waveguide, gives the energy gain in one section:

$$V = E_o L \cos\varphi \left(\frac{1 - e^{-IL}}{IL} \right) - irL\left(1 - \frac{1 - e^{-IL}}{IL}\right) \qquad (5)$$

in which L is the length of the waveguide section. Now the phase is inserted correctly. If you ask what happens when the wave and the particle velocity are not synchronous, that is, if you are working off frequency a little bit, you get into very considerable difficulties trying to find reliable solutions. Using the same general approach, one can also work this out to give transient solutions to the problem.

Now, to give an idea of the effects, the energy gain can be written as a very simple expression:

$$V = V_o \left(1 - \frac{i}{2i_{max}}\right) \qquad (6)$$

This equation gives you a very good idea of the effect of beam loading. V_o is the energy gain for vanishing beam current, and i_{max} is the instantaneous beam current at maximum power transfer to the beam. In this case, since the currents we are talking about correspond to something on the order of 25 to 30% (or more) fully beam-loaded, energy gain changes in the waveguide sections of something like 20% or more will be expected in going from no beam to full beam. This is going to have large effects on the particle dynamics.

The above approach is based on the power diffusion equation. If one wants to determine the effects of subharmonic injection (for example, injection on every fourth or sixth rf cycle), then these equations cannot give an answer. The beam is tied to only one of the spatial modes. If one asks whether there is any beam loading caused by other spatial harmonics, the equations cannot solve the problem.

Therefore, we tried to approach the linac problem from the standpoint of filter theory. Consider, for example, a four-terminal filter. What follows is really quite general, because it follows from Floquet's theorem. If there is a field $V_o(\omega)$ in the first filter section and the system has a periodic structure, then in the q^{th} section we have $V_q(\omega)$. One can show, in general, that

$$V_q(\omega) = V_o(\omega)\, e^{-q\Gamma(\omega)} \qquad (7)$$

where $\Gamma(\omega)$ is the complex phase shift of the network (guide). This is assuming only the fundamental spatial mode.

The general nature of the dispersion diagram for the waveguide is indicated in Fig. 1. Here frequency, ω, is the ordinate, and β, the imaginary part of the complex phase shift, is the abcissa. In our filter language the β scale shown in Fig. 1 corresponds to two disks of the linac waveguide per filter section.

Referring again to Fig. 1, if $\Gamma_{o,o}(\omega)$ corresponds to the lowest frequency pass band and to the lowest spatial mode, then we know that

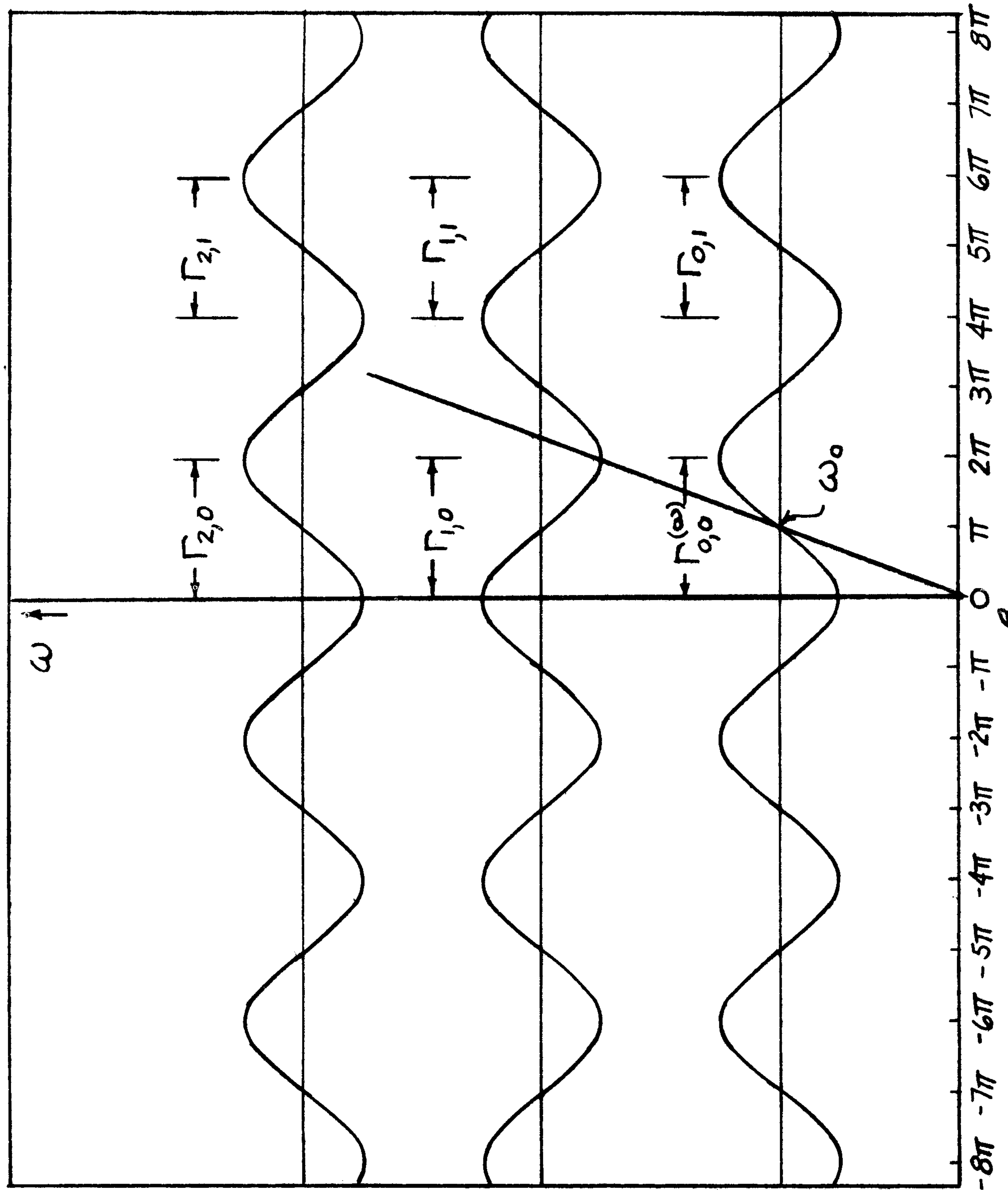
$$\Gamma_{o,n}(\omega) = \Gamma_{o,o}(\omega) + 4\pi\, ni \quad . \tag{8}$$

This also is a consequence of Floquet's theorem, and will hold, respectively, for each frequency passband. We can now say in general, including all spatial modes, that the transmission of rf in the waveguide for the n^{th} frequency passband is given by

$$V_q(\omega) = V_o(\omega) \left[\sum_{n=-\infty}^{\infty} A_{m,n}(\omega)\ e^{-q\Gamma_{m,n}(\omega)} \right] \tag{9}$$

where the $A_{m,n}(\omega)$ are structure dependent and are determined by the need to match boundary conditions on the walls of the waveguide.

This is a completely general description of the transmission of rf in the guide. It can be a standing or travelling wave structure, but it can be calculated



ω
β
ω_0
$\Gamma_{2,0}$
$\Gamma_{2,1}$
$\Gamma_{1,0}$
$\Gamma_{1,1}$
$\Gamma_{0,0}^{(\omega)}$
$\Gamma_{0,1}$
-8π -7π -6π -5π -4π -3π -2π -π 0 π 2π 3π 4π 5π 6π 7π 8π

FIG.1

in terms of these functions. The trick is, of course, to get the functions. For most of the remainder of this discussion I will talk only about the usual accelerating mode $\Gamma_{o,o}(\omega)$ and will refer to it as $\Gamma(\omega)$ for convenience, and will also assume $A_{o,o}(\omega) = 1$.

We wish to find solutions which correspond to the ones I wrote down initially from the power diffusion equation. These are dispersion-free solutions and for these, one type of expansion is convenient. Assume some frequency ω_a at which we expect to operate the machine. We then make an expansion of $\Gamma(\omega)$ around ω_a:

$$\Gamma(\omega) = \Gamma(\omega_a) + \Gamma'(\omega_a)\,(i\omega - i\omega_a) + \text{higher order terms} \quad (10)$$

If you keep just these two terms you will get solutions which in almost every respect give you the answers you get from the power diffusion equation. You can, however, now easily include such effects as particle transit time, off-frequency operation, sub-harmonic injection, excitation of other spatial modes, etc., to the extent that dispersion effects may be ignored. An expansion can be made by including higher order terms to show dispersion; however, then one has a terrible time getting a proper steady state solution.

For solutions including dispersive effects we found it much better to make an expansion around the mid-band frequency, at least, for the relatively symmetrical first passband. We forget about the higher passbands for the moment and deal only with the lowest passband so we have terms like

$$V_o(\omega)e^{-q\Gamma(\omega)} \quad .$$

Now, look at this formulation in terms of transforms. We see that if $V_o(t)$ is a δ function, and since the transform of the δ function is unity, we get a spectrum in ω flat with respect to all frequencies. If one goes into the inverse transform to get, for example, $V_q(t)$, it is convenient to call

$$e^{-q\Gamma(\omega)} = G_q(\omega)$$

and to define a function $G_q(t)$ which is the inverse transform of $G_q(\omega)$. $G_q(t)$ is thus the response of the q^{th} section of the waveguide to a δ function input. Since one can use the superposition theorem in a relativistic electron linac (and also in a proton linac if the velocity doesn't change too fast) then, by simple folding integrals, one can calculate virtually any other desired property of the linear accelerator, with or without dispersive effects.

If dispersive effects are ignored, we find that the solutions we get are essentially the same as one gets from the power diffusion equation, as long as one is talking about a simple machine in which you inject on every rf cycle, and so forth.

However, if we deal with the situation in which one injects every 4th or 6th pulse (which is the situation now planned in the transition from a drift tube accelerator to a disk-loaded accelerator) we find additional beam loading effects when the calculations are expanded to include dispersive effects. If injection occurs once every 5 rf cycles, it turns out that you have about 18% more beam loading than you might expect, and if once every 3 rf cycles, it is about 8% more. This is one specific qualitative beam-loading effect.

Because of the dispersive effects there are all kinds of oscillations developed on pulses, since the linac passband may be only 4 to 6 Mc/sec wide in an efficient 1000 Mc/sec structure. There are thus upwards of 100 cycles of rf in the guide at any one time, so that there is a substantial ringing of the pulses. However, for the long pulses we have discussed here, this is not a very critical point.

Another effect may be seen by considering the influence of different spatial modes of a structure. It is stated, for example, in Slater's article and other places, that only one spatial mode contributes and that the other spatial harmonics, although there is rf power in them, pass the electrons very rapidly and the net influence on the beam is negligible. This is certainly true for the case of rf power coming in from an external source, an electron or particle going past these higher modes sees a very rapidly oscillating field due to these modes.

However, in the case of heavy beam loading, this isn't quite as true as in the case of light beam loading, and in one particular case, that of the π mode, it is never true. The reason is that the sources of these beam-loading waves are the separate narrow bunches of electrons which move along the waveguide so that, although the wave and the beam particles may be crossing each other very rapidly, it isn't obvious that you couldn't run into a higher order synchronism such that each time the particle and the higher mode waves actually crossed each other they would, in fact, be in phase. In any case but the π mode there is some of this effect, and it can take as many as 200 rf cycles of injection for this effect to dissipate. It is a very short-pulse transient effect.

However, if you investigate the π mode, you find that this is no longer so, for the beam-loading waves are always in synchronism, and the beam loading turns out to be twice what you would get from normal theory. One might say that the shunt impedance which should be used to calculate beam loading is twice as big in a π-mode machine as the shunt impedance you use to calculate the energy gain from the external power source.

If you think about it, it is quite obvious that this must occur. We know that at the π-mode the forward-wave and backward-wave branches of the dispersion curve must coincide (see Fig. 1). The only way to accomplish this is for the amplitudes of the backward and forward waves to be equal in order to maintain the standing wave character of the beam loading wave as well as the external wave, (and this must be done to match boundary conditions on the waveguide.) Therefore, we must put as much power into one direction as we do into the other. This is true of every spatial mode in the waveguide for which the group velocity is zero. Thus, in designing the second part of the proton machine, the π mode is perhaps not the best choice. There will be substantially more beam loading than might have been anticipated.

The key to this point is that the beam is not continuous but is in discrete bunches. This coherence effect with higher spatial modes of the beam loading waves is a result of the fact that the beam is tightly bunched. Of course, if the beam were not bunched you would get no beam loading at all.

KNOWLES: Are your results based exclusively upon particles with $\varpi = 0$?

LEISS: No, but let me qualify our results. We have done two kinds of calculations. In one, we get a non-dispersive solution to the linac properties in terms of these $G_q(t)$ transfer functions. The second calculation includes dispersive properties to the extent that we match the $\beta - \omega$ diagrams. We do the latter by making an analytic power series fit around the center of the passbands. There are additional factors such as the amplitudes of the various spatial harmonics, which are actually functions of frequency. We don't know how to vary them with frequency and so we call them constants, which is an approximation.

GIORDANO: You mentioned the higher-order spatial harmonics of the backward-wave effect in the beam. Did you also include the effects of forward waves of higher order spatial harmonics?

LEISS: Yes, in the π-mode case, every one of those is synchronous with the beam for the fundamental passband.

GIORDANO: You have a forward and backward wave in the same phase.

LEISS: Yes, but those are forward and backward in the group velocity sense, and they add up to the same phase in their effect on the electron beam. The only way you can maintain the boundary conditions on the waveguide is for them all to be excited. We have also assumed that the coupling of the beam to the guide is independent of frequency, which again is clearly not correct. We have assumed it to be constant. Formally, we know that there is a $K(\omega)$ which expresses the coupling of the beam to the wave but we don't have any information on it. It requires a detailed calculation for any given waveguide structure, which one could do and then get the answers.

QUESTION: Could I ask about each of the harmonics having the same velocity as the beam, in the case of π mode?

LEISS: I didn't say they had the same velocity. I said that every time they see the bunches of electrons they see them in phase. This is difficult to explain -- I have used one analogy (which has its fallacies) but let me present it to demonstrate the idea. Imagine we have a wheel with an axle, rolling on a plane, and a stroboscope looking at it, as shown in Fig. 2.

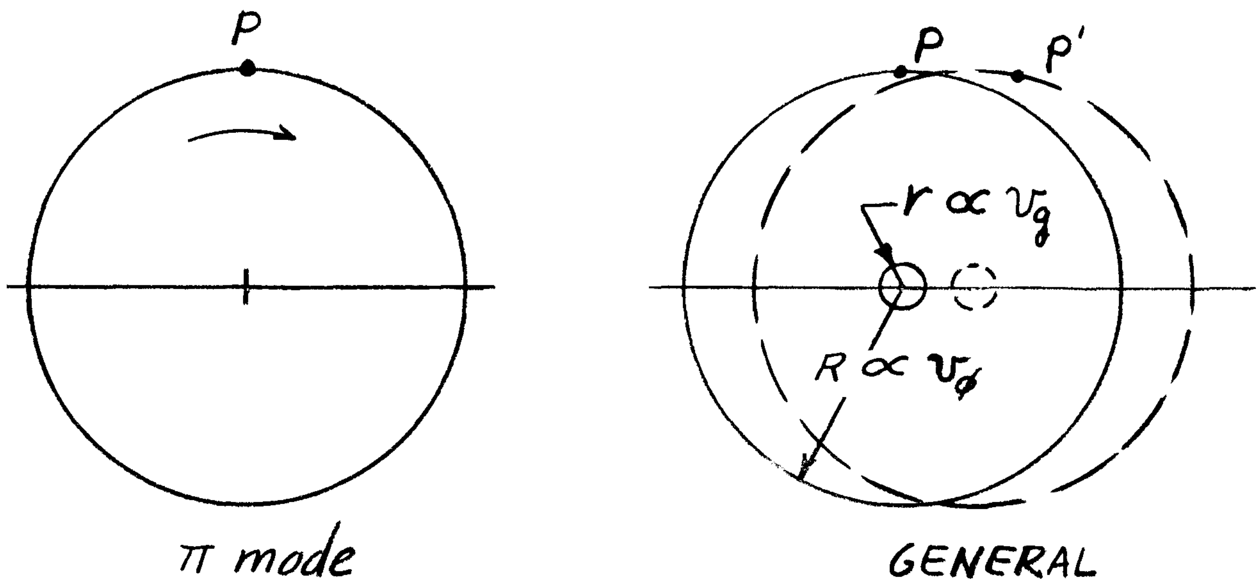


FIG. 2

The stroboscope represents the beam bursts and the ratio of the axle to the wheel circumference is the ratio of the group to the phase velocity. In the extreme case of a standing wave (π-mode) machine the axle has become a pivot. Now when a beam pulse comes along the light is flashed and finds a point p at the top. When the next light flash (beam pulse) appears, the wheel will have turned once and p would be at the top again, that is, in phase. This is the fundamental accelerating mode. However, if the wheel has turned around twice between light flashes the point p would still be in phase, and it might also have gone around backwards an integral number of times, with the same result. If now there is a finite axle on the wheel ($v_g \neq 0$) then a single rotation brings point p to p' and the machine is designed for this. But if it is rotated twice, then p moves twice as far and it will gradually drift out of synchronism. This corresponds to cases other than the π mode and only transient phenomenon will be seen in other modes, but not in the π mode.

PARKER: Do you assume zero group velocity for π mode operation?

LEISS: Yes, because it is not a superconducting linac, you never quite get there, but you can still get pretty close to it.

One can also include in the above treatment the possibility that the frequency may not be quite correct or that the particle velocity is changing through the section. I don't want to dwell very long on this because virtually everything is included in a report we have issued.[1]

I would now like to go to the second subject of this talk, the phenomena of beam blow-up. Consider the fundamental accelerating mode, operating at, say, ω_o (see Fig. 1). The vertical scale in the passband region is enormously exaggerated -- the true scale would represent the passband almost as a straight line so it really doesn't matter at which point in the lower mode one operates. The second passband is generally considered to be responsible for beam blowup and this band has been demonstrated (in the mass separator experiments at Stanford) to have the right field configuration. The mode in next higher passband has, let us say, a frequency ω_1. I don't intend to discuss the dynamical effects that occur in beam blowup. What is believed to happen (and people have calculated it) is that if a little bit of the ω_1 frequency is excited by some means, the wave, in general, turns out to be a backward wave (or if it isn't you will certainly find a spatial harmonic that is backward). This backward wave is a deflecting mode so that as it goes back it perturbs the beam going past, increases its rate of radial deflection, which generates more of the ω_1 mode, and pretty soon you have an oscillator. I think that this has been confirmed to be the mechanism of beam blowup. Now in making calculations of these phenomena, people have often ignored the coherence properties of these waves. In some six waveguides of different kinds on which we have made measurements, the ratio (ω/ω_o) was very close to 3/2, within about 1%.

GLUCKSTERN: Are these all $\beta = 1$ guides?

LEISS: These are all $\beta = 1$ guides, and this may make a difference, but I don't think this is important.

GLUCKSTERN: It probably doesn't make a big difference; you're talking about narrow bands anyway.

LEISS: Yes, it should make very little difference. The excitation of this type is clearly due to the harmonic structure of the beam.

If you calculate you will find that the excitation waves in this mode are of the form $e^{i\omega_1 t}$. Now generally the time t corresponds to the times at which particles are injected into the waveguides so t is of the form $t_k = 2\pi k/\omega_o$, in which k is some integer corresponding to phased injection of particle bunches with respect to ω_o. If we put the times t_k into the above expression we get

$$e^{2\pi i(\omega_1/\omega_o)k} \cong e^{3\pi ik} \tag{11}$$

in which we have used the measured 3/2 relationship between ω_1 and ω_o. This is the character of the excitation of this passband by successive beam pulses. Now, if k takes on all integral values (k = 1, 2, 3, . . .) then one gets an alternating +1 and -1 from $e^{3\pi ik}$ and the excitation of the ω_1 mode by the beam pulse is cancelling itself from pulse to pulse. Only fluctuations from random noise are going to get this mode started. However, if one injects every other pulse into the machine (k = 2,4,6, . . .) this function $e^{3\pi ik}$ is always either +1 or -1 and cancellation does not occur. When one injects on any even subharmonic of the beam, then a resonance has been hit. Now, how important the fact is that you are perhaps 1% off in ω_1 can be argued. I believe it to be very important that one is close and it is my feeling that if an even frequency ratio between the drift tube sections and the iris-loaded sections of the linac is chosen, one is going to have to worry very much about the excitations of the ω_1 mode, because the normal anticoherence is missing. I would predict that in such a machine you could experience enormous beam blowup at quite a small beam current because of this fact. An odd sub-harmonic would not have this effect in the ω_1 mode. However, this doesn't prove anything because there are many other passbands above that one and since now you're dealing with particles with $\beta < 1$, the possibilities of coupling to those modes is much greater than in the case of relativistic electrons. This will require a careful study to decide whether one can hit some accidental synchronism with a higher mode as apparently one does in a $\beta = 1$ electron linac.

GLUCKSTERN: In the electron linac that you have studied I gather that it is possible to suppress this ω_1 mode by putting some modifications in the structure. Does this conform with the blowup which could be related to higher passbands or don't they seem to be important in that case?

LEISS: Let me answer that in several ways. First, if one puts $\beta = 0.5$ (150 keV) unbunched electrons into a waveguide and looks at the frequency spectra he gets out of the guide, he can see anything he wants to see. There are hundreds of resonances in there. I haven't any idea what this means because the particles are nonrelativistic -- β is low but can be changed very easily by the fields that are excited. So I don't have any idea how serious this is, but there are certainly a great many resonances to couple to the beam.

The second point is that it is possible to design a waveguide to suppress many of these effects. The constant gradient type of structure, in which one changes the properties of the waveguide from section to section, is clearly very useful in this respect. There is, because the waveguide is tapered, a natural incoherence that starts as the wave moves down the waveguide. However, I suspect that, to be really safe, it does not taper fast enough. If I were trying to design one of these guides, I would taper it too fast, Let me expand upon this point: In a constant attenuation waveguide the field is a negative exponential function of the distance. In a constant gradient machine, you taper the guide to slow down the group velocity farther along the guide and keep the field constant. I would design such a structure so that the field comes up too far and then reverse it so that I had a very high fluctuating field gradient. This would keep the spatial coherence length of this higher mode as short as possible. I don't specifically know how to do it. However, one can get a pretty good idea of how much he should taper the field from the calculations that have been made on the beam blowup as a backward wave oscillator phenomenon. It's clear that the wave must travel backwards a certain distance in order to have a unity feedback factor. If the coherence is destroyed by tapering the guide fast enough so that for any useful current the feedback factor stays well below one, then clearly, the blowup mode won't be excited. If one has tapered fast enough to avoid exciting the first blowup mode he clearly has tapered fast enough to avoid exciting everything else.

KNOWLES: I talked to Dr. Chu at Stanford, and he suggested that the problem of feedback might be avoided in a π-mode machine because the group velocity of the ω_1 wave would then be forward, as can be seen from a diagram like Fig. 1.

LEISS: I don't agree. It's just that you're on another spatial harmonic, and it would take more current to cause blowup in this case because the coupling of that other spatial harmonic is smaller. However, in particular for the case of a proton machine where $\beta < 1$, you don't have to have a backward wave to excite the beam, although you do to get a backward wave oscillator.

GLUCKSTERN: I'd like to say that I agree with you because I don't think that the backward wave has anything to do with it.

CARNE: I think the fact that it's a backward wave is purely incidental to the effect.

BLEWETT: In a proton linac, there may be a strong effect if the beam is focussed by quadrupoles.

LEISS: It is very much tied up with some things we don't know; one of them is how far this backward wave has to go. If it really is only a portion of the waveguide, then quadrupoles won't help, and having a a resonant structure won't help. It's something that occurs on a very short time and distance scale. If it has to oscillate a few times in the waveguide and really build up over a long time, then perhaps these effects will be true. However, there's considerable evidence to the effect that it has to travel only a short distance. Of course, as the beam gains energy, it is harder to deflect and then the distance scale will increase.

Reference

(1) Transient and Beam-loading Phenomena in Linear Electron Accelerators. J. E. Leiss and R. A. Schrack, Internal Report, National Bureau of Standards, Wshington D.C., (October, 1962).