AN INSTRUMENTATION WISH LIST FOR HIGH POWER/HIGH BRIGHTNESS ERLS *

D. Douglas[#], Jefferson Lab, 12000 Jefferson Avenue, Newport News, VA 23606, U.S.A.

Abstract

The advent of the energy recovering linac (ERL) brings with it the promise of linac-quality beams generated with near storage ring efficiency. This potential will not, however, be fulfilled without overcoming a number of technical and operational challenges. We will review the basics of ERL dynamics and operation, and give examples of idiosyncratic ERL behavior and requirements posing particular challenges from the perspective of diagnostics and instrumentation. Beam performance parameters anticipated in next-generation ERLs will be discussed, and a "wish list" for the instrumentation of these machines presented.

"TRADITIONAL" ACCELERATORS

Energy Recovering Linacs [1] comprise a class of electron accelerators using novel architecture to generate outstanding beam quality with high wall plug efficiency. Initially conceived almost half a century ago [2] and first realized a decade later in a reflectively symmetric geometry [3], contemporary ERLs are most easily understood as an evolution of the recirculating linac [4].

"Traditional" accelerators typically fall into one of two classes: synchrotrons (including storage rings and cyclotrons for the purposes of this discussion), or linear accelerators. In synchrotrons, the accelerated beam passes many times through a modest radiofrequency (RF) accelerating system, gaining a small amount of energy on each pass. Most of the accelerator consists of the transport system, which draws little additional power. Storage rings can therefore generate extremely high beam powers using modest wall plug power. In such systems, a few megawatts of RF drive and similar levels of DC power (for the transport system) can be used to generate gigawatts of beam power (e.g. 0.5 A at 2 GeV). As the beam is circulated many times, it is however susceptible to degradation from error sources and beam dynamical effects such as synchrotron radiation. Storage rings, though efficient, thus provide only limited beam quality.

In contrast, linear accelerators (linacs) accelerate the beam rapidly through multiple RF structures using only limited beam transport. Most of the required power is therefore devoted to RF drive; moreover, in conventional (normal conducting) linacs, the final beam power is typically limited by wall losses in the accelerating structures. For example, a linac with a megawatt of RF drive (with very modest DC power for additional beam transport use) will produce at most a megawatt of beam power (e.g. 50 μA at 20 GeV). Linacs are thus electrically inefficient - but as the beam is only in the machine for a single pass though the accelerator (and there is little or no bending) the beam quality does not degrade: performance is source limited. It is thus possible to generate very high quality beams – an advantage that outweighs the power inefficiency in many applications. The high beam quality (brightness) provided by linacs has motivated efforts to make them more cost effective. As a consequence, numerous methods are available to control cost, improve efficiency, and enhance the performance of these systems.

RECIRCULATION, SUPERCONDUCTING RF, AND ENERGY RECOVERY

Recirculation provides a simple means of controlling linac cost while maintaining final energy: simply "reuse" the linac accelerating structure multiple times, thereby reducing the required length. In a "recirculating" linac, the beam is transported from the end of the linac back to the front, and reinjected in phase with the RF fields for further acceleration. This approach leverages the low cost of beam transport relative to RF, so that the cost/performance optimum for a system lies somewhere between "straight" and "circular". The result [5] (Figure 1) is a system of lower cost than a conventional linac, but which provides better beam quality than a storage ring.

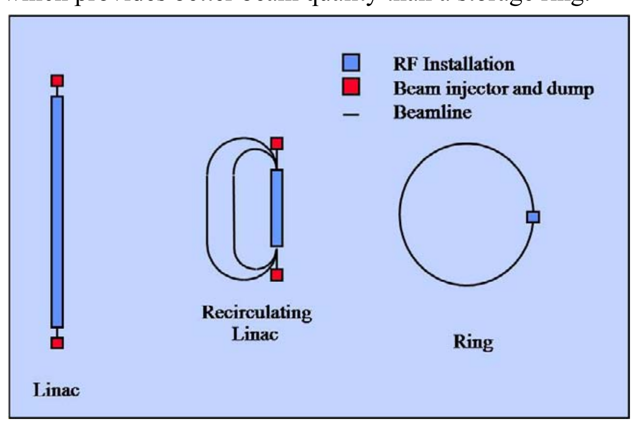

Figure 1: Evolution of conventional accelerator topology into that of the recirculating linac [6].

Superconducting RF (SRF) acceleration [7] provides a second means of linear accelerator cost/performance optimization. SRF cavities have essentially no wall losses and allow high gradient continuous-wave (CW) operation. This reduces the RF power required to reach a specific energy, increases average current, and significantly improves beam quality by avoiding the transients inherent in pulsed RF systems. The resulting reduction in RF power demand is sufficient to offset the cost of the

*Work supported by the Commonwealth of Virginia, the Air Force Research Lab, the Joint Technology Office, the Office of Naval Research, and by DOE Contract DE-AC05-060R23177.
[#]douglas@jlab.org

required cryogenic plant and to compensate for any additional RF system complexity.

Use of recirculation and SRF introduces additional beam handling and sensitivity to error sources, and thereby engenders susceptibility to instability and performance degradation. Application of these technologies in large-scale systems such as CEBAF [8] demonstrates, however, that it is possible manage such effects. Even so, RF power remains an obstacle to the use of recirculating linacs, as a simple example illustrates. A nuclear-science system such as CEBAF generates a 4 GeV, 200 μA beam, consequently requiring of order 800 kW of RF power – a manageable (affordable) level. A recirculating linac x-ray light source would, however, require 100 mA at 5 GeV (or ½ GW of RF power) to generate the necessary photon fluxes. Not only would a dedicated electrical plant would be needed to serve the accelerator, the disposal of a gigawatt of "waste" electrons presents an intractable design issue.

Such RF power-draw and waste-beam disposal issues can be addressed by use of energy recovery: after acceleration to – and use at – full energy, the electron beam is returned to the linac and reinjected – *180^o out of phase with the linac accelerating RF fields*. As the beam is "antisynchronous", is then *decelerated* as it traverses the accelerator structure. Beam power is deposited back into the RF cavities and is available to accelerate later bunches; after traversing the linac, the beam is returned to essentially injection energy – and power – levels, and thus disposal presents only a modest radiation issue. The use of SRF technology avoids wall losses and ensures the recovery process can proceed with excellent efficiency.

Energy recovery has been tested in high power [9] and large scale [10] systems, reliably reducing RF power demands while preserving beam quality and maintaining overall accelerator performance. Figure 2 characterizes RF power utilization in CEBAF-ER, a 1 GeV ERL operational mode of the CEBAF accelerator [11]. The red trace shows a low-level RF signal quantifying the required RF power without energy recovery over a 400 μsec interval. During a 250 μsec beam macropulse (followed by a short diagnostic pulse) RF power is required and the measured signal rises; turn-on and turn-off transients are manifested as exponential rise and fall of the signal value. The blue trace presents the same

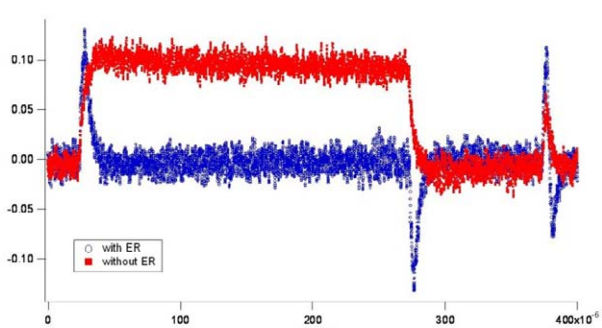

Figure 2: RF power draw during beam macropulse in CEBAF-ER without (red) and with (blue) energy recovery

signal with the use of energy recovery. Save for the turn-on and turn-off transients, *no* RF power (other than that needed to establish the fields in the SRF cavities) is needed: the recovered beam supplies all the power required for acceleration. Transient behavior during beam turn-on and turn-off is due to cavity beam loading and detuning, and constitutes a topic of considerable interest in the process of optimizing the RF drive system [12].

ERLs thus have both good acceleration efficiency and high beam quality. The use of SRF insures that there are no cavity wall losses; energy recovery returns essentially all beam power to the linac, so there is "no" beam loading: only the injected beam power must be provided. If a machine injects 0.1 A at 10 MeV (1 MW), accelerates it to 100 MeV, and energy recovers, the system requires 1 MW of power. If the system, in contrast, injects 0.1 A at 10 MeV (1 MW), accelerates to 10 GeV, and recovers, it *still requires only 1 MW*. As energy recovery recycles the "waste" (post-use) beam to drive the RF, it saves on RF costs and avoids high power radiation from the beam dump. Not only is it a natural "two-beam accelerator", the "recovered" beam is at low energy – and thus low *power* – making the design and fabrication of the beam dump much easier.

Energy recovery thus provides near-linac-quality beam at near-storage-ring efficiency. It is therefore an effective means of design cost optimization and – as the resulting systems are of novel architecture and operate in a non-traditional region of parameter space – allows the study of numerous physical phenomena in combinations that are typically inaccessible in traditional accelerator designs.

ERL PERFORMANCE

Existing ERLs have generated relatively high power (MW-class [13]) and high energy (GeV [14]) beams; the technology holds promise for orders-of-magnitude extension in both parameters so as to provide GW beam powers at GeV energy scales. Figure 3 presents an ERL "landscape" showing potential routes to both high energy and high power. To date, systems have been operated along the energy axis of this parameter space; the impending initial operation of the BNL ERL test-bed will explore increasingly high current behavior [15].

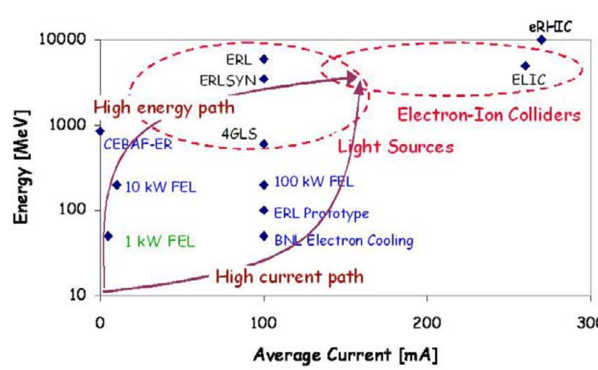

Figure 3: ERL Current/Energy landscape illustrating parameter set for existing (1 kW FEL, 10 kW FEL, CEBAF-ER) and proposed ERLs [16].

ERL applications tend to fall into two classes: use as FEL drivers, and as sources of synchrotron radiation or beams for investigating fundamental interactions.

FEL Driver ERLs

FEL drivers are defined by the electron beam required at the laser, which needs a high peak current and a beam transverse geometric emittance meeting the $\lambda_{FEL}/4\pi$ limit. The former is met at single bunch charges of 0.1 –1 nC by use of appropriate bunch length compression (see below). The latter is met by source normalized emittances of 1-10 mm-mrad and acceleration to high energy so as to adiabatically damp to the geometric emittance envelope of the FEL. For photons from THz to soft x-ray this motivates use of energies of tens of MeV to a few GeV. The FEL interaction is a coherent collective phenomenon; very high photon powers/fluxes can thus be obtained using modest CW repetition rates (100s of kHz – few MHz), beam currents of mA to tens of mA, and attendant beam powers at megawatt levels.

Light Sources

ERL-based light sources try to leverage the brightness advantage possessed by linacs over storage rings. Though linacs generally provide superior longitudinal emittances, modern ultra-low emittance rings provide smaller transverse emittances at energies below a few GeV (above which radiation excitation dominates). As a consequence, ERLs of this class are x-ray sources having electron beam energies lying in the 5-10 GeV range, and (to provide appropriate brightness) run 100s of mA currents in high repetition rate beams (GHz) with bunches of tens to 100s of pC. The lower charge allows for exceptional source performance (0.1 - 1 mm-mrad normalized emittance) and provides operational modes giving at least partial transverse coherence. Beam powers in such systems are extremely high, being near or at GW levels.

ERLS of either class hold advantages over traditional machines. They allow flexible time structure, spanning the range of single bunches to CW bunch trains. Moreover, they allow independent manipulation of various portions of the beam phase space essentially independently of other sub-spaces – they are fully six-dimensional systems, allowing transverse matching to desired spot sizes at specified locations, longitudinal matching to specific bunch length/energy spread combinations (via use of transverse-longitudinal coupling), and partial or full transverse phase space exchanges through the use of horizontal-vertical coupling. They can therefore generate beams of controlled peak current, specified momentum spread, and transverse emittances ranging from flat to round.

UNIQUE PROPERTIES/CHALLENGES

The operational flexibility and the beam power and brightness available from ERLs create potential for unique behavior. As beam powers are extremely high, the beam is continually renewed from the source, and the system never goes to equilibrium, halo is a monumental issue. This is best understood by contrast to storage rings: ERLs are quite unlike storage rings with respect to beam loss; they are instead equivalent to the *injection chain* of storage rings. Given the high CW beam power, losses must remain in the 10^{-5} – 10^{-6} range, equivalent to an injection efficiency of 99.999% or better: only one to ten of every million electrons provided by the source can be lost during acceleration, handling, recovery, and transport to the dump without potential machine damage.

Numerous other features distinguish ERLs from traditional classes of machine. ERLs do not have a closed orbit, nor are they even necessarily betatron stable. They might not, therefore, have uniquely defined Twiss parameters, but rather admit an infinite number of lattice function solutions. Physically, this is because the beam and the lattice are not the same; one cannot simply equate Twiss parameterizations, betatron functions, and lattice functions in an ERL, in stark contrast to their equivalence in an electron storage ring at equilibrium. ERLs do not, of course, go to quantum equilibrium (as the beam is in the machine much less than a damping or excitation time)

Unique as well is the fact that ERLs will have multiple beams (at least 2, but perhaps 4, 6, 8,...) in different focusing/accelerating structures – even while they are physically in the same devices. This is because different passes through linac are phased and or focused differently depending on energy. Depending on the machine configuration, multiple beams can also share common transport systems – though this will be nominally at a common energy and separated by (modulo) a half-RF period in phase. These features have significant impact on the process of multipass orbit correction (see example below) and the design of beam position monitors (BPMs) capable of distinguishing amongst the various passes and phases/timings of beam. The most obvious challenge is the dynamic range in current: from single bunch to high frequency CW. Less evident, though no less daunting, is the observation that RF-based diagnostics can be faced with the problem of discriminating between two CW bunch trains separated by only half an RF period. This task is often rendered more difficult by behavior similar to that in the accelerating cavities: the response of the diagnostic to the presence of one pass cancels out the response to the other (as they are out of phase with one another) and the device output is null.

ERLs utilize a variety of (nonlinear) longitudinal manipulations to compress the bunch in duration and/or energy. An example of this – bunch compression in an FEL driver – will be discussed below, and serves to highlight the type of beam and lattice longitudinal diagnostic of practical use in these systems. From the perspective of active beam stabilization, it is worthwhile to observe that ERL topology is similar to that of storage rings, and thereby allows feedback and feed-forward system implementation more easily than is typically possible in a conventional linac.

Given the novel machine architecture and the power advantages provided by energy recovery, it is unsurprising that ERLs face beam dynamics challenges common to all traditional classes of accelerator. One might cynically argue that they are a hybrid of storage rings and linacs and thus are susceptible to the problems of both. More realistically, because ERLs generate high power, high brightness beams, virtually all fundamental phenomena become relevant. These have been exhaustively discussed in the proceedings of recent workshops [17]; here we simply cite examples of major concern.

Beam Self-Interactions

This class of phenomena includes effects based on electrons interacting with one another via their charge or equivalently the bunch wake-field. Examples include space charge effects – both transverse and longitudinal (SC, LSC), coherent synchrotron radiation (CSR), and microbunching instabilities (MBI) mediated by either of the former (or other, external) effects.

Beam/Environment Interaction

Included here are beam interactions with modes in superconducting cavities and vacuum system components. Examples: beam break up (BBU), resistive wall, and the impact of environmental wakes and impedances of all kinds. Of particular interest for high-brightness ERL applications is the interaction of the beam with interceptive diagnostics such as view foils, for which coherent transition radiation (COTR) effects [18] can render high-precision measurements difficult.

Quantum Excitation

As in all electron transport systems, quantum excitation can lead to beam quality degradation. This is of particular concern in ERLs because they engender use of more bending than traditional linac configurations but unlike storage rings effectively have no compensatory damping mechanism and do not settle into an equilibrium.

Stray Power Deposition

As noted above, high power ERL beams preclude tolerance of beam loss at all but the lowest levels; halo is therefore a critical concern. In addition, ERL beams can generate unacceptably large power deposition when the beam current couples to the environment. Common concerns include RF heating and loss of power into propagating higher order modes (HOMs), resistive wall effects (as mentioned above), and CSR/THz power deposition. Existing systems have observed losses of tens to hundreds of Watts into such phenomena; proposed systems could – without proper design – experience tens of kW to MW losses.

Effects of Field Errors

ERLs invoke, by their very nature, transverse-longitudinal coupling to generate and manage the energy recovery process. This in turn creates a unique opportunity for DC magnetic field errors to lead to energy errors; field inhomogeneities thereby lead to energy spread after energy recovery and can impede the recovery process. This effect will be discussed in some detail in the next section.

EXAMPLES OF ERL IDIOSYNCRACIES

The challenges confronting ERLs are most easily contextualized by specific examples. At some point (typically, full energy) the beam will engage in a useful activity – interact with a target, generate light, drive an FEL – which will remove energy and degrade beam quality. Managing this degraded beam through energy recovery (in light of adiabatic antidamping) renders ERL operation a rather more complex matter than simply arranging for the beam to ride the RF crest up to full energy and resting in the trough of the waveform back down. We therefore offer five examples of the issues encountered in ERL design and operation.

Nonlinear Longitudinal Matching: FEL Drivers

FEL driver ERLs must satisfy two fundamental requirements: they have to deliver a bunch with high peak current and adequate beam quality *to* the FEL (to generate the required FEL gain), and they must recover the exhaust beam *from* the FEL (to provide the RF power needed for the acceleration of high average currents). The first of these involves performing a bunch compression with a high-brightness beam; the second is typically made difficult by the FEL-induced growth in beam energy spread, which can readily exceed 10% of the full beam energy after lasing. As a consequence, the system must not only recover the bunch energy centroid – in order to re-use the beam RF power – it must perform an energy compression during recovery so as to avoid beam loss in the linac back-end and at the beam dump. Figure 4 presents a schematic of the longitudinal matching scenario employed in such systems. The individual phase-energy plots indicate the orientation of the longitudinal phase space at key locations around the machine.

To avoid space-charge or LSC driven degradation of beam quality, the injector provides a long bunch with small momentum spread. This is accelerated off-crest in the linac, creating a phase-energy correlation (chirp). The momentum compaction of linac-to-wiggler recirculation transport rotates the "chirped" bunch upright, generating a short bunch at the wiggler. This bunching process is equivalent to producing a "parallel to point" image of the injected longitudinal phase space at the wiggler. Present transport systems (e.g. the Jefferson Lab (JLab) IR Upgrade FEL) typically produce bunch lengths of order 150 fsec rms at 100 MeV and peak currents of order 400 A [19]; recent design work motivated by the need to mitigate CSR effects, suggests shorter bunches and higher peak currents are accessible [20].

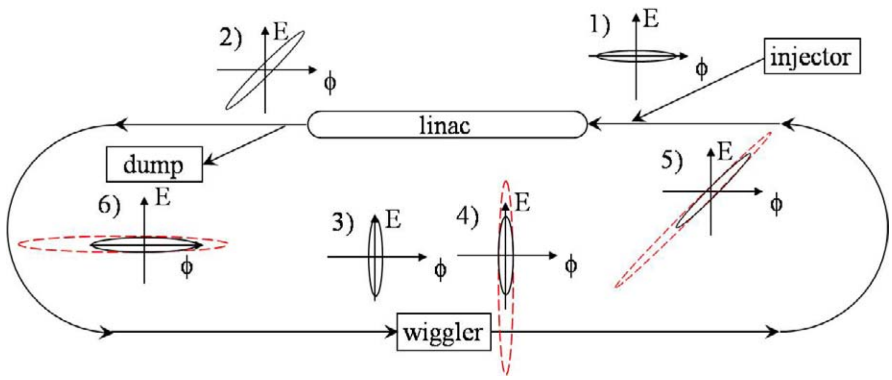

Figure 4: Longitudinal matching scheme for FEL driver ERL, showing phase/energy plots at critical locations: 1) injection; 2) end of linac (1st pass); 3) wiggler entrance; 4) wiggler exit (black: no lasing, dashed red, lasing); 5) reinjection, 6) beam dump. Centroid energy/phase and bunch length/energy spread change when laser turns off /on.

Unlike conventional linac FEL drivers, ERLs do not in general utilize harmonic RF to linearize bunch compression; reasons for this will be discussed below. Instead, the nonlinear longitudinal aberrations of the linac-to-wiggler transport are adjusted with sextupoles (or higher-order multipoles) to compensate for the curvature of the RF waveform, transport lattice aberrations, and phase space distortions due to CSR. Control of T_{566} is a standard practice in existing machines; higher order effects can be managed similarly [21].

When delivered to the wiggler the short bunch drives the FEL interaction, which does not affect the bunch length or transverse beam properties, but which shifts the central energy downward (as the FEL extracts power from the beam) and generates a very large momentum spread. This is depicted in the phase space diagrams adjacent to the wiggler in Figure 4; the growth in momentum spread is evident in Figure 5, which shows the beam in the JLab 115 MeV FEL driver at a dispersed point (η=0.7 m) in the recovery transport system downstream of the wiggler, without lasing (left image: full momentum spread ~2-3%, or 3 MeV) and with lasing (right image: full momentum spread ~10%, or ~12 MeV). This is, as well, indicative of the rather large momentum acceptance required of the recirculator.

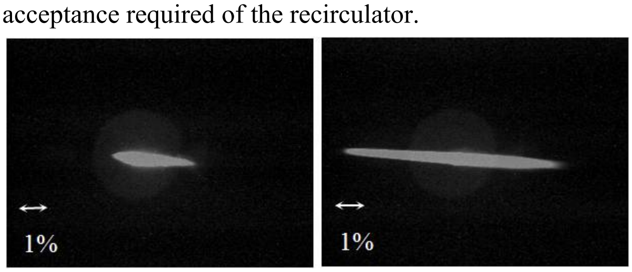

Figure 5: Synchrotron radiation image in recovery arc (dispersion of 0.7 m); left: no lasing, right: lasing

Recirculator path length is adjusted to control the phase of the reinjected beam and to select a recovery phase operating point providing sufficient RF gradient to allow for compression of the highest-energy components of the exhaust beam (Figure 6). We note that this phase is typically *not* 180^o away from the acceleration phase and as a consequence power balance in the linac is not exact; the energy recovery is “incomplete” [22].

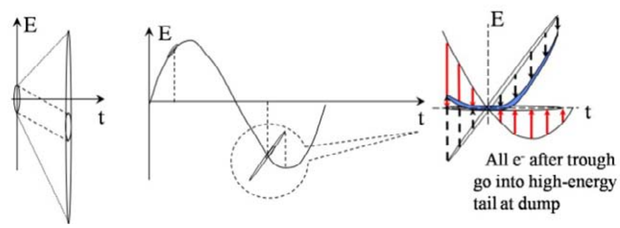

Figure 6: Choice of recovery phase. Left: lasing removes energy/increases energy spread; center, right: beam recovery phase must be far enough from trough to provide RF gradient needed for full energy compression.

Recirculator momentum compactions (M_{56}, T_{566}, W_{5666}) are used to rotate, curve, and torque the bunch so that an appropriate phase/energy correlation occurs at reinjection. This shaping is done using quadrupoles, sextupoles, and octupoles and fits the beam to the nonlinear variation of the RF waveform, insuring that a full energy compression occurs as the beam reaches the dump. In analogy to the bunching process described above, this is a “point to parallel” image of the longitudinal phase space from the wiggler to the dump. We thus note that as the laser turns off and on, the variation in the central energy of the exhaust beam will – as a consequence of this imaging process – be rotated into a phase variation through the linac; at the dump the energy does not change. The resulting phase transients in the high power beam as it is decelerated introduce fascinating beam loading effects and impose potentially severe constraints on the design of RF drive systems for high power ERLS [23].

As noted above, ERL drivers avoid use of harmonic RF for linearization. This is because the recovered beam can exceed 30^o in length at the RF fundamental (90^o at the harmonic) – so the linearization expansion converges poorly. Even harmonics cannot be used in energy recovered systems [24], so extending the technique to higher order requires fifth-harmonic RF, introduces prohibitively small apertures (intercepting halo) and high impedances (wake fields), effects that prove problematic in high power systems even at the third harmonic.

Practical implications of this process are shown in Figure 7, which presents "before and after" images of the JLab IR Demo high power beam dump. Limited initial understanding of the longitudinal matching process led to incomplete energy compression during recovery, resulting in scraping of halo and mechanical damage.

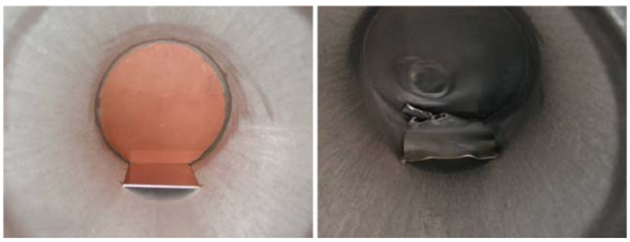

Figure 7: Pre- (left) and post-operations (right) images of JLab IR Demo dump showing damage from improperly-performed longitudinal match [25].

Multipass Orbit Correction

Recirculating linacs face special problems when correcting orbits because of the use of common elements for the transport of multiple beams at different energies [26]. ERLs share this issue; the energy of each pass differs from that of the (all) other(s) at any point along the linac(s), as a consequence the response to steering differs. The JLab IR Upgrade FEL [27] (shown in Figure 8) illustrates the problem. If the first pass beam is subjected to a closed orbit bump in the second two accelerating cryomodules, the recovered beam – at much lower energy at that location – is strongly deflected (Figure 9).

Because both common error sources and common transport elements – such as orbit steering correctors and focusing trims - act on different energies differently, such systems benefit from localized observation and correction of errors. This dictates a need for multipass BPMs capable of resolving the positions of multiple high-frequency CW bunch trains separated by as little as half an RF period. The same hardware could well be called upon for observation of single-shot pulse trains – or even individual bunch trajectories – in some applications.

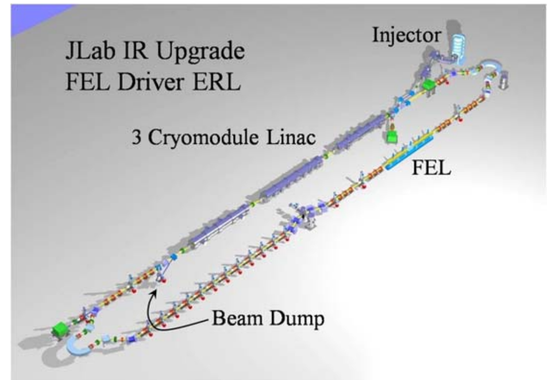

Figure 8: JLab IR Upgrade FEL; injector outside ring, at upper right; three-cryomodule linac on upper axis; FEL in lower axis. Beam dump to interior of ring at lower left.

Collective Effects

ERLs live to generate high brightness, high power beams; collective effects are a "logical consequence" of this lifestyle. The JLab ERL systems have been challenged by several effects [28], including

- LSC (degrading compressed bunch length [29])
- BBU (leading to current limitations [30])
- CSR (limiting deliverable beam quality [31])

ERL instrumentation and control systems must be able to observe, characterize, and quantify collective effects of these – and other – types. Figure 10 illustrates the use of streak camera data to assess the impact of LSC in the JLab IR Upgrade FEL. Measurements of the longitudinal phase space as a function of acceleration phase indicate that beam momentum spread increases while accelerating on the falling part of the waveform [32], this helps confirm simulations that provided guidance for mitigating the impact of the phenomenon [33]. Similarly, BBU control requires instrumentation to measure (and potentially feed back on) higher-order mode (HOM) power. CSR presents signatures such at beam steering as a function of compression [34] that can potentially be (and occasionally have been, at JLab!) misinterpreted as lattice dispersion errors due to magnetic field errors or transport lattice nonlinear aberrations.

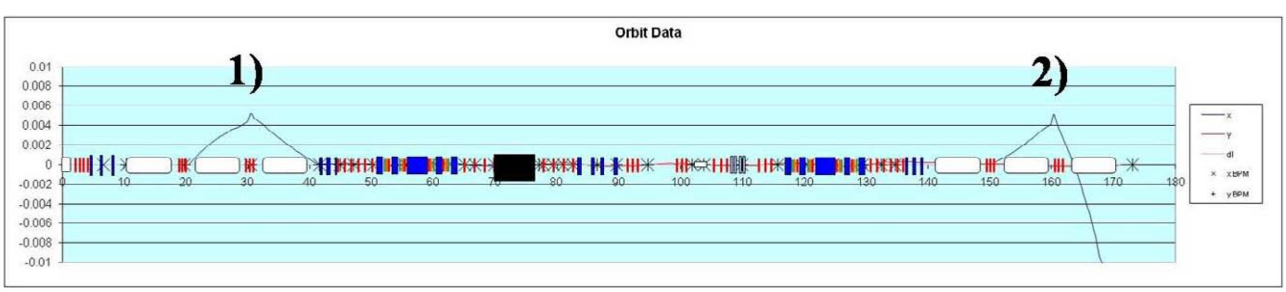

Figure 9: Simulation of orbit bump in common transport region of JLab IR Upgrade FEL. A closed orbit bump imposed on the first (acceleration) pass across the second two accelerating cryomodules (closed upon entry to the recirculation arc, left side of graphic, 1)) results in a severe deflection of the second (energy recovered) pass as it traverses the final cryomodule and approaches the beam dump (right side of graphic, 2)).

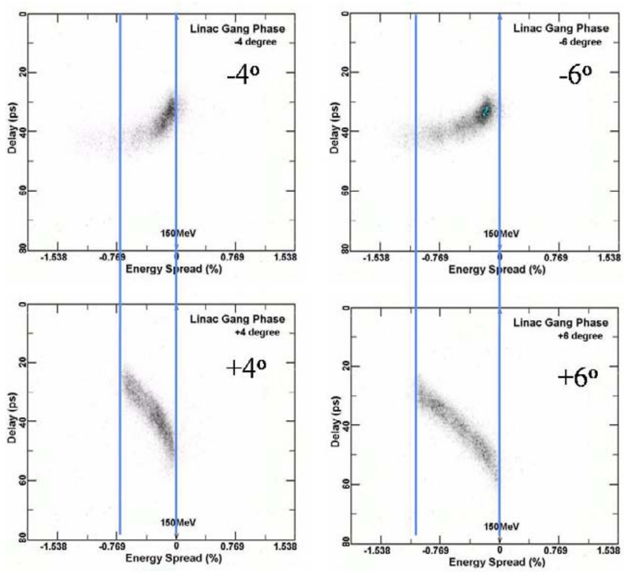

Figure 10: Streak camera data showing LSC effects. Top: acceleration on falling portion of RF waveform. Bottom: acceleration on rising portion. Top case has enlarged momentum spread and irregularity in density distribution.

Transverse-Longitudinal Coupling: The Impact of Magnetic Field Inhomogeneities

Imperfections in magnet field quality have long been recognized as a source of performance limitations in accelerators. In transport lines and linacs, field inhomogeneities lead to beam envelope mismatch, orbit-dependant optics, phase space distortion, and emittance degradation. The mechanism for such effects is simple – field deviations lead to position-dependent bending of some portions of the beam in a manner deviating from design and/or that experienced by other portions. The resulting unanticipated spread in beam angular distribution manifests itself as a focusing error (if linear) or a phase space distortion (if nonlinear).

In conventional accelerators, these errors are of primary significance in transverse phase space. In ERLs they can be longitudinally deleterious as well, by virtue of the six-dimensional nature of the beam dynamics ("synchro-betatron coupling" in synchrotron parlance). As ERLs rely – as described above – upon momentum compaction control to provide longitudinal phase space matching, the presence of angular errors can lead to unexpected changes in bunch length and RF phasing. These in turn alter the energy spread, limiting performance – either at full energy (*e.g.*, by generating excessive energy spread at user experiments) or after energy recovery (by yielding an unmanageably large energy spread at the beam dump).

We can assess the performance impact of field inhomogeneities by evaluating the energy spread after energy recovery caused by such errors. The application over a length l of an error field ΔB to a portion (hereafter, referred to as a "filament") of a beam leads to unintended bending of the filament through an angle $\delta\theta = \Delta Bl/B\rho$. If this occurs at a location from which the transfer matrix to reinjection has a nonzero dependence of path length on angle, or M_{52} (as is often the case, if the error is in a recirculation magnet), a path length error $\delta l = M_{52}\,\delta\theta$ will evolve. When reinjected, the beam filament will thus no longer be synchronous with the nominal RF phase; instead, it will experience a phase offset $\delta\phi = 2\pi\,\delta l/\lambda_{RF}$.

The energy recovered by deceleration through a linac with energy gain E_{linac} at phase ϕ is $E = E_{linac}\cos\phi$, the energy shift resulting from a phase offset $\delta\phi$ from a nominal phase set point ϕ_0 is therefore as follows.

$$\Delta E = E_{linac}\,[\cos(\phi_0+\delta\phi) - \cos(\phi_0)] \cong -E_{linac}\sin\phi_0\,\delta\phi$$

It is assumed that energy compression is desired during energy recovery, and that deceleration (and acceleration) therefore occurs off-trough (crest). (If the acceleration (energy recovery) are on crest (trough, $\phi_0 = 0$), the sine goes away and the energy offset is quadratic in $\delta\phi$.) Folding the preceding expressions together yields an expression relating the imposed field error to the energy offset generated after energy recovery.

$$\Delta E = -\left(\frac{2\pi M_{52}}{\lambda_{\mathrm{RF}}}\right)E_{linac}\sin\phi_0\left(\frac{\Delta Bl}{B\rho}\right)$$

The filament thus ends up at the "wrong" final energy, with error depending on the RF wavelength, the field quality (ΔBl), the details of the transport (M_{52}), and the linac energy gain (E_{linac}). Given that there will be variations in field error across the full beam, different filaments will be transported to different final energy offsets – thereby increasing the energy spread of the beam after energy recovery. By viewing ΔB as a bounding value or tolerance on the field variation and M_{52} as an average value for the path length term, we can then interpret ΔE as an estimate of the final energy spread.

This result can be rewritten to emphasize the effect of each term. M_{52} is constrained by the symplectic condition to be a combination of dispersive and betatron components of the matrix: $M_{52} = M_{22}M_{16} - M_{12}M_{26}$. Control of final momentum spread is therefore promoted by the use of small dispersion and betatron function values. Use of lower linac energy gain also reduces sensitivity. The impact of dipole field quality is made clear by noting $(\Delta Bl/B\rho) = (\Delta B/B)(l/\rho) = (\Delta B/B)\,\theta$, where θ is the bend angle. The induced energy offset is thus

$$\Delta E = -[(2\pi M_{52}/\lambda_{RF})\,E_{linac}\sin\phi_0\,](\Delta B/B)\,\theta\,,$$

which is directly proportional to the relative field error $\Delta B/B$. The total angle will typically be π (such as in the energy recovery transport of an FEL driver) or 2π (as in a generic ERL recirculator); to maintain a fixed energy spread after energy recovery, the relative field error must therefore *decrease* as the linac energy *increases*.

This latter observation implies yet another viewpoint. Note that the energy spread at the dump depends on $E_{linac}/B\rho$. For injection energies small relative to the linac energy the rigidity will however be approximately 33.3564 kg-m/GeV × E_{linac}. The energy error after energy recovery is thus *independent* of the full energy, and is influenced by only the lattice parameters, RF wavelength (M_{52}/λ_{RF}) and the absolute error integral ΔBl.

$$\Delta E = -\left(\frac{2\pi M_{52}}{\lambda_{\mathrm{RF}}}\right)\sin\phi_0\left(\frac{\Delta Bl}{33.3564\ \mathrm{kg\text{-}m/GeV}}\right)$$

The error field integral producing a specific energy error (energy spread) at the dump is thus *independent* of the linac energy. As the machine full energy increases (implying increased total field integral required to transport the higher energy beam), the tolerable relative error integral will *decrease*, as noted above.

The coupling of transverse steering errors into the longitudinal motion can potentially seriously constrain the performance of recirculating linacs and ERLs. This is not limited to the generation of unanticipated energy spread after energy recovery that is discussed above. Field errors in recirculation transport during acceleration can lead to bunch lengthening and result in growth of momentum spread in any recirculating linac. Synchrotron radiation excitation, with an associated shift in energy at dispersed or compactional locations of the beam transport, will drive the evolution of bunch length errors (longitudinal emittance) in a fashion similar to the degradation of transverse emittance, again leading to growth of momentum spread at full, as well as recovered, energy.

These effects all encourage the use of "better" magnets, lower linac energy gains (between transport system modules with compaction management [35]), and smaller M_{52} values (dispersions, beam envelopes). We note that the effect of poor magnet field quality is not "undoable" – the steering induced by field errors can be corrected or compensated; it does however require provision for this compensation. Diagnostics (for example, phase transfer function measurement systems [36]) and correction knobs (multipole correctors) should be made available if it is not possible to achieve the desired performance with the magnet field quality available within the constraints imposed by the system budget.

The preceding treatment mentioned only in passing the case of on-crest acceleration and energy recovery. In such cases, the resultant energy spread is linear in the linac energy gain and quadratic in the phase (or field integral) error, suggesting that the product (which goes as linac energy divided by the square of the beam rigidity) will *decrease* with increasing linac energy. This would imply that on-crest/in trough acceleration/energy recovery is quite desirable. This is in fact often the case in recirculating, non-energy recovering linacs such as CEBAF, which accelerates a short, small momentum spread bunch on crest so as to take maximum advantage of the available gradients and to limit growth in bunch length and momentum spread. However, for higher charge-state machines (such as FEL drivers and light sources) it may be preferable to accelerate and recover off crest or out of trough. Off-crest acceleration allows transport of longer bunches without undue excitation of wakefields and CSR; the bunch is compressed only where the time structure is needed. Extraction of large amounts of power from the beam (either in an FEL, as a CSR source, or as a source of incoherent synchrotron radiation) will typically lead to generation of energy spread, which, when decelerated, will adiabatically antidamp to large relative energy spreads after energy recovery.

Key to this discussion is the observation that between the limit of on-crest operation (with no energy compression and large final energy spread from adiabatic antidamping) and off-crest operation (with energy compression and potential for large final energy spread due to field-error induced bunch lengthening) there is an optimum at which the two sources of energy spread cross over, minimizing the total final spread. At this point, appropriate longitudinal diagnostics (for both lattice and beam) and accelerator tuning range can compensate for errors that could render the system otherwise inoperable.

Halo

As indicated in the discussion of "unique properties and challenges", halo is a serious concern. In that discussion, we did not, however, characterize what halo actually is. "Halo" includes all beam elements outside the beam core. Sources include stray/scattered drive laser light – and so called "ghost pulses" from incomplete suppression by electro-optical drive laser gating – in photocathode guns [37], cathode relaxation leading to temporal tails at bunch formation, beam dynamical effects (such as space charge) acting during the beam formation and capture process [38], and nonlinear collective effects (space charge, microbunching instabilities, CSR, environmental wake fields,...) during acceleration and beam handling.

Halo management is critical in ERL operation so as to avoid beam loss. Because of exceptionally high (GW) beam powers, losses must be – as noted – at 10^{-5} or 10^{-6} levels to avoid activating or damaging beamline components. This engenders some difficulty inasmuch as halo is characterized by emittances that can be significantly larger than those of the core beam and – recalling our observation that the lattice and the beam are not characterized by the same Twiss parameters – is not described by core betatron functions. Consequently, apertures cannot be specified by "rms beam sizes" – these are completely irrelevant to the halo, which is defined by different emittance and by envelopes that can be completely mismatched to the transport lattice focusing structure, which is typically optimized for the core. ERL beams are decidedly non-Gaussian (Figure 11).

Systems thus need a large "working aperture" allowance so as to accommodate these largely unknown and machine-dependent effects. Instrumentation with exceptional dynamic range – as much as six orders of magnitude [39] – must be developed to characterize halo properties, and controls devised that will allow management of halo without introduction of untoward effects on the beam core.

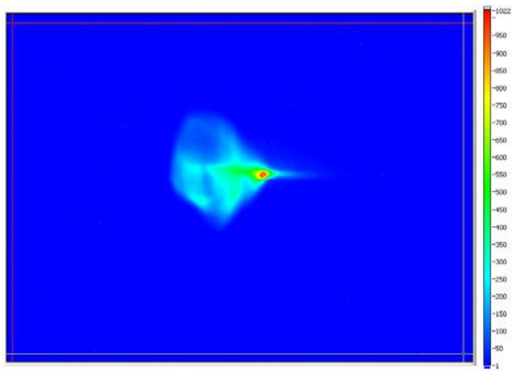

Figure 11: Non-gaussian "hummingbird" in JLab IR Upgrade FEL driver ERL (courtesy P. Evtushenko).

IMPLICATIONS FOR DIAGNOSTICS

These examples give guidance on ERL instrumentation requirements. There is a clear need for a full suite of longitudinal diagnostics, including means of measuring time of flight (for phasing and synchronism purposes), bunch length, and lattice properties such as linear and nonlinear path-length and energy correlations (e.g., in TRANSPORT notation, M_{55}, M_{56}, T_{566},...). Transverse diagnostics must measure positions (and potentially spot sizes) of multiple beams in common transport, in an operationally non-invasive manner. This is all the more challenging because beam time structure can be nearly infinitely variable (from single bunches to CW beam) and include virtually all timing patterns of bunch trains. In addition, accelerated and recovered beams will be separated by modulo one-half RF period – and may, in high power ERLs, be separated by *only* half an RF period during CW operation. The beam fields may then cancel in the resonant structures used for some types of diagnostic, suppressing available signal levels.

These basic requirements and concerns are aggravated by the need to separately characterize the accelerator transport lattice and the beam. As noted, the lattice need not be betatron stable, nor is there a closed orbit. As a result, lattice diagnostics cannot rely on traditional methods such as "pinging" the beam and looking for sidebands on the beam spectrum; there is no "tune" to serve as a reference. Similarly, "tune shifts as a function of focusing strength" do not provide "the" beta function: there *is* no unique betatron function. Instead, the system is characterized by separate (and potentially non-unique) lattice functions and beam envelopes.

ERL "lattices" are analyzed using difference-orbit measurements: the beam is treated as a macroparticle and steered in angle, time, or energy, and transfer matrix values determined from the response. As there is no closed orbit or requirement of betatron stability, phase advances are measured by counting oscillations. ERLs are thus at a disadvantage relative to storage rings, wherein each BPM can take numerous readings (beam positions at each of many turns) and thereby improve measurement resolution. In ERLs, precision is limited by the number and resolution of BPMs and the tolerable orbit excursion, as well as by magnet field quality, lattice nonlinearities, and potentially even collective effects like CSR, which – as mentioned above – may generate lattice-error-like effects during some operations. Instrumentation systems must provide a means of capturing the signatures of such phenomenon and extracting the required information.

The "beam" that is accelerated and recovered in the separately characterized lattice must be subjected to its own management (such as matching to the lattice) and tuning (to achieve optimum system performance), and thus requires additional appropriate diagnostics. First amongst these will be phase space tomography tools providing emittance and beam envelope (and coupling) data, to give insight into various lattice and beam dynamically induced phase space distortions.

A DIAGNOSTICS WISH LIST

A comprehensive diagnostics list was compiled during the 2009 ERL Workshop [40]. Experience suggests that the resolution of phase advance and matrix elements is a basic issue for lattice control, with limits set by the number of available diagnostics (particularly BPMs). The ability to resolve of beam properties is also limited by the number of available diagnostics (in multi-monitor measurements) and the dynamic operating range of the system, as it is in general not possible to run full power beam into an intercepting diagnostic, nor change focusing for a quad scan or tomographic measurement at high power. Finally, the diagnostic suite for an ERL must support excellent beam stability and synchronization.

The operation of existing machines gives specific guidance as to diagnostics that have (or would have) proven useful. These include high resolution multi-pass BPMs working over broad dynamic range (μA to 100s mA) and with flexible beam timing (single shot to CW), beam time of flight and phase transfer function measurement systems [41], and tomographic systems for the entire six-dimensional beam phase space. This requires high resolution/ large dynamic range profile monitors and bunch length measurement systems such as Martin-Puplett interferometers and THz (FTIR) spectrometers. As discussed above, halo is a serious concern in high power systems, so some type of "halo monitor" [42] is needed. Longitudinal manipulations are aided by an available streak camera. Finally, means of monitoring beam noise and stability [43] will be needed to meet the synchronism requirements in FEL drivers and beam stability demands in ERL-based light sources.

Energy recovering linacs are systems of novel architecture presenting great potential to serve as sources of uniquely bright, high power beams. This potential will however be realized only at the cost of meeting numerous beam dynamic and operational challenges. Extensive instrumentation and sophisticated controls are, therefore, absolutely necessary for successful operation of the next generation of ERLs – the diagnostic systems of which must be designed and implemented with full awareness of the idiosyncracies of these devices.

ACKNOWLEDGMENTS

We have presented results from work by many colleagues at Jefferson Lab to understand and operate both CEBAF and the IR/UV FEL. Financial support was provided by the Commonwealth of Virginia, Air Force Research Lab, Joint Technology Office, Office of Naval Research, and DOE Contract DE-AC05-060R23177.

REFERENCES

[1] L. Merminga *et al.*, "High-Current Energy-Recovering Electron Linacs," Annual Review of Nuclear and Particle Science 53 (2003) 387.
[2] M. Tigner, Nuovo Cimento 37 (1965) 1228.
[3] S.O. Schriber *et al.*, IEEE Trans. Nuc. Sci. NS-24 (1977) 1061.
[4] R.E. Rand, "Recirculating Electron Accelerators," Accelerators and Storage Rings, Vol. 3, Harwood Academic, New York (1984); JB Flanz and CP Sargent, IEEE Trans. Nuc. Sci. NS-32 (1985) 3213.
[5] *ibid*.
[6] L. Merminga *et al., op cit.*
[7] C.W. Leemann *et al.*, "The Continuous Electron Beam Accelerator Facility: CEBAF at the Jefferson Laboratory," Annual Review of Nuclear and Particle Science 51 (2001), 413.
[8] *ibid.*
[9] C. Tennant, "Progress at the Jefferson Laboratory FEL," to appear in PAC'09, Vancouver, May 2009, TH3PBI03, (2009); S. Benson *et al.,* "High Power Operation of the JLab IR FEL Driver Accelerator," PAC'07, Albuquerque, June 2007, p. 79 (2007).
[10] A. Bogacz *et al.,* "CEBAF Energy Recovery Experiment," PAC'03, Portland, May 2003, p.195 (2003).
[11] *ibid.*
[12] T. Powers and C. Tennant, "Implications of Incomplete Energy Recovery in SRF-Based Energy Recovery Linacs," ERL'07, Daresbury, May 2007, p.75 (2007).
[13] C. Tennant, *op. cit.,* S. Benson *et al., op. cit.*
[14] A. Bogacz, *op. cit.*
[15] D. Kayran *et al.,* "Status of High Current R&D Energy Recovery Linac at Brookhaven National Laboratory", LINAC'08, Victoria, October 2008, p. 453 (2007).
[16] L. Merminga *et al., op. cit.*
[17] See Proc. ERL'05, Newport News, March 2005, Nuc. Inst. Meth.-A 557 (2006) pp. 1-382; Proc. ERL'07, Daresbury, May 2007; Proc. ERL'09, Ithaca, June 2009; www.lepp.cornell.edu/Events/ERL09/; Proc. DoE BES Workshop "Accelerator Physics of Future Light Sources", Gaithersburg, September 2009, to appear in Nuc. Inst. and Meth.
[18] H. Loos *et al.,* "Observation of Coherent Optical Transition Radiation in the LCLS Linac," FEL'08, Gyeongju, August 2008, p. 485 (2008).
[19] S. Benson *et al., op. cit.,* C. Tennant, *op. cit.*
[20] D. Douglas, "Possibility of Recirculation Systems for High Brightness Beams," ERL'09, Ithaca, June 2009; www.lepp.cornell.edu/Events/ERL09/.
[21] D. Douglas, "The Jefferson Lab 1 kW IR FEL," LINAC'00, Monterey, August 2000, p. 716 (2000).
[22] S.V. Benson and D.R. Douglas, "Use of Incomplete Energy Recovery for the Energy Compression of Large Energy Spread Charged Particle Beams," US Patent #7166973, 23 January 2007.
[23] T. Powers and C. Tennant, *op. cit.*
[24] A half-period phase separation at the fundamental becomes a full period at an even harmonic, so the harmonic RF accelerates (or decelerates) *both* passes.
[25] K. Jordan, private communication.
[26] R.M. Bodenstein, M.G. Tiefenback, "Multipass Steering Protocols at Jefferson Lab," PAC'07 Albuquerque, June 2007, p. 3648 (2007).
[27] C. Tennant, *op. cit.*
[28] *ibid.*
[29] C. Hernandez-Garcia *et al.,* "Longitudinal Space Charge Effects in the JLab IR FEL SRF Linac", FEL'04, Trieste, August 2004, p. 363 (2004); S. Zhang *et al.,* "Temporal Characterization of Electron Beam Bunches With a Fast Streak Camera at JLab FEL Facility", FEL'05, Stanford, August 2005, p. 641 (2005).
[30] C. Tennant *et al.,* Phys. Rev. ST Accel. Beams **8**, 074403 (2005); C. Tennant *et al.*, Phys. Rev. ST Accel. Beams **9**, 064403 (2006).
[31] CSR effects are described in detail in K.L.F. Bane *et al.*, Phys. Rev. ST Accel. Beams 12, 030704 (2009).
[32] S. Zhang, *op. cit.*
[33] C. Henandez-Garcia, *op. cit.*
[34] K.L.F. Bane *et al., op. cit.*
[35] Use shorter linacs and multiple compaction managed recirculation arcs, in analogy with generating phase advance by using numerous, closely spaced, weaker quadrupoles rather than a few very strong ones.
[36] D. Hardy *et al.*, "Automated Path Length and M_{56} Measurements at Jefferson Lab", PAC'97, Vancouver, May 1997, p. 2265 (1997).
[37] When a drive laser beam passes through a window, it scatters, leading to halo production from a photocathode. This scattering can be at the 10^{-4} level (S. Benson, private communication).
[38] P. Evtushenko, unpublished simulation and machine study of JLab IR Upgrade injector.
[39] A.P. Freyberger, "Large Dynamic Range Beam Profile Measurements", PAC'03, Portland, May 2003, p. 2565 (2003).
[40] P. Evtushenko, "Instrumentation Needs from All Working Groups," ERL'09, Ithaca, June 2009; www.lepp.cornell.edu/Events/ERL09/
[41] D. Hardy *et al., op. cit.*
[42] A.P. Freyberger, *op. cit.*
[43] P. Evtushenko and D. Sexton, "Electron Beam Timing Jitter and Energy Modulation Measurements at the JLab ERL", BIW'08, Lake Tahoe, May 2008, p. 247 (2008).