eRHIC/11
C-A/AP/#290
September 2007

Free Electron Lasers and High-Energy Electron Cooling

Vladimir N. Litvinenko (BNL)
Yaroslav S. Derbenev (TJNAF)

Collider-Accelerator Department
Brookhaven National Laboratory
Upton, NY 11973

FREE ELECTRON LASERS AND HIGH-ENERGY ELECTRON COOLING*

Vladimir N. Litvinenko, BNL, Upton, Long Island, NY, USA#
Yaroslav S. Derbenev, TJNAF, Newport News, VA, USA)

Abstract

Cooling intense high-energy hadron beams remains a major challenge in modern accelerator physics. Synchrotron radiation of such beams is too feeble to provide significant cooling: even in the Large Hadron Collider (LHC) with 7 TeV protons, the longitudinal damping time is about thirteen hours. Decrements of traditional electron cooling decrease rapidly as the high power of beam energy, and an effective electron cooling of protons or antiprotons at energies above 100 GeV seems unlikely. Traditional stochastic cooling still cannot catch up with the challenge of cooling high-intensity bunched proton beams - to be effective, its bandwidth must be increased by about two orders-of-magnitude.

Two techniques offering the potential to cool high-energy hadron beams are optical stochastic cooling (OSC) and coherent electron cooling (CEC) – the latter is the focus of this paper. In the early 1980s, CEC was suggested as a possibility for using various instabilities in an electron beam to enhance its interaction with hadrons (i.e., cooling them). The capabilities of present-day accelerator technology, Energy Recovery Linacs (ERLs), and high-gain Free-Electron Lasers (FELs), finally caught up with the idea and provided the all necessary ingredients for realizing such a process.

In this paper, we discuss the principles, and the main limitations of the CEC process based on a high-gain FEL driven by an ERL. We also present, and summarize in Table 1, some numerical examples of CEC for ions and protons in RHIC and the LHC.

INTRODUCTION

In contrast with electron- and positron-beams, hadron beams in storage rings (colliders) do not have strong loss mechanism (such as synchrotron radiation for leptons) and, therefore, do not have a natural damping mechanism to reduce their energy spreads and emittances.

There are several reasons why cooling high-energy hadron beams, mostly at the top energy range of a collider, is strongly desirable.

First, any increases in the longitudinal- and transverse-emittances of a hadron beam accumulated during multi-stage acceleration from a source to the store energy (collision) remain in the beam. Any instability causing the growth of emittance (for example, that of the electron cloud that is the limiting factor for most modern hadron colliders) may entail the need to discard accelerated beams and start the process again. In any case, present-day high-energy hadron colliders do not have control of beam emittances at the collision energy, and are forced to use beams as they are; this is not always the optimum approach.

The main figure of merit of any collider is its average luminosity, i.e., its average productivity for an appropriate branch of physics. Cooling hadron beams at top energy may further this productivity.

For a round beam, typical for hadron colliders, the luminosity is given by a simple expression:

$$\mathcal{L} = f_c \frac{N_1 N_2}{4\pi\beta^* \varepsilon} \cdot h\left(\frac{\sigma_s}{\beta^*}\right) \tag{1}$$

where N_1, N_2 are the number of particles per bunch, f_c is their collision frequency, β^* is the transverse β-function at the collision point, ε is the transverse emittance of the beam, σ_s is the bunch length, and $h \leq 1$ is a coefficient accounting for the so-called hourglass effect [1]:

$$h(x) = \frac{\sqrt{\pi}}{x} e^{1/x^2} erfc(1/x).$$

The hourglass effect is caused by variations in the beam's size $\sigma_r^2 = \beta^* \varepsilon \left(1 + s^2/\beta^{*2}\right)$ along the length of the collision region, which is defined by the bunch-length, σ_s. Accordingly, for $h > 0.75$, β^* should be limited to values $\beta^* \geq \sigma_s$. Hence, longitudinal cooling of hadron beam may allow reduction of β^* and increase the colliders' luminosity.

The effect of transverse emittance cooling on the collider's luminosity is less straightforward. For beams with limited intensities, the luminosity (1) grows as the transverse emittance decreases. But luminosity in many colliders is limited by the beam-beam effects, and hence, a reduction of emittance can be either counterproductive or luminosity-neutral. In this case, the role of transverse cooling is reduced to one of counteracting intra-beam scattering or/and cooling large amplitudes of oscillation to lower the detector's backgrounds.

The presence of strong damping in electron-positron colliders allows the use of significantly higher bunch intensities compared with hadron beams of the same energy: the beam-beam tune shift can be about factor of four larger in the best lepton colliders than in best hadron colliders. The accelerator community discussed the conjecture that effective (strong) cooling of hadron beams might increase luminosity in such colliders via increasing the allowable beam-beam tune-shift. We emphasize that *this assertion*, though worth mentioning, *is very speculative* and should not be accepted without serious skepticism.

The electron-hadron collider, eRHIC, however, is one type of collider where cooling the hadron beam offers dramatic advantages [2]. In eRHIC, polarized electrons accelerated in an ERL will collide with hadrons stored in the RHIC's storage ring. In this case, a reduction in the transverse emittance of the hadron beam engenders a proportional reduction of the electron beam's intensity

* Work performed under the auspices of the U.S. Department of Energy.

while maintaining its ultimate luminosity constant [3]. Reduction of the electron beam's current has multiple advantages: reducing the strain on the polarized electron source, proportionally lowering synchrotron radiation (the main source of the detector's background); and, offering the possibility of increasing the electron beam's energy.

Plans are to use a non-zero crossing angle at the Large Hadron Collider (LHC) at CERN [4]. In this case, reducing the bunch's length would directly contribute to increasing the luminosity and eliminating the necessity of having a crab-crossing system [5].

Hence, high-energy hadron cooling may play important role in increasing the performance of high-energy hadron and lepton-hadron colliders, including RHIC and eRHIC at Brookhaven National Laboratory, BNL, the Tevatron at Fermilab, and the LHC at CERN.

Electron cooling proved to be very efficient method of cooling intense hadron- and ion-beams at low and medium energies [6]. The electron cooler of 9 GeV antiprotons in the Fermilab recycler represents state-of-the-art technology [7]. Development of the ERL-based electron cooler at BNL promises effective cooling of gold ions with energies of 100 GeV per nucleon [8]. Nevertheless, the effectiveness of electron cooling is weaker for protons than it is for ions (it scales with Z^2/A, where Z is the change number of an ion and A its atomic number, viz., $Z^2/A=1$ for protons and $Z^2/A=31.7$ for $_{197}Au^{79}$). It also falls sharply with the beam's energy (for RHIC it falls as $\gamma^{7/2}$ [9], where $\gamma=E/mc^2$ is the relativistic factor of a particle). Hence, traditional electron-cooling of protons with energies from about 100 GeV (RHIC) to a few TeV (LHC) with conventional techniques is hardly possible within the realm of present accelerator technology.

The idea of coherent electron cooling (CEC) [10,11] encompasses various possibilities of using collective instabilities in the electron beam to enhance the effectiveness of the interaction between hadrons and electrons. In this paper, we focus on a specific case of using a high gain FEL (driven by an ERL) for CEC. CEC combines the advantages of electrostatic interaction with the broad band of FEL-amplifiers: examples in this paper span from tens of THz to hundreds of PHz. Such systems are naturally fit into a straight section of modern high-energy hadron colliders. The proposed CEC method has some potential advantages compared with the concept of optical stochastic cooling [12]:

a) it may not entail significant modifications to the lattice of the hadron machine ;
b) it uses electrostatic interaction instead of very inefficient radiation and interaction with TEM waves by protons in a wiggler;
c) it is not limited to few potential choices of laser frequencies and their bandwidths in THz range.

Similar to other coherent cooling techniques, the CEC's cooling rate is limited by the cross-talk of neighboring hadrons (and the short noise in the electron beam). Thus, the cooling rate is limited by an effective number of particles in a coherent sample, which is inversely proportional to the amplifier's bandwidth. In the CEC scheme, the FEL frequency can be chosen appropriately to match the energy of the electron beam. Consequently, for LHC energies the FEL wavelength naturally extends into the soft-X-ray range (nm), where frequencies are measured in ExaHertzs (10^{18} Hz). Even a tiny fraction of this frequency extends far beyond the bandwidth of any other amplifier.

Table 1. Comparison of estimations for various cooling mechanisms in RHIC and LHC colliders. The sign ∞ is used to indicate helplessly long damping times.

Machine	Species	Energy GeV/n	Synchrotron radiation, hrs	Electron cooling, hrs	**CEC, hrs**
RHIC	Au	100	20,961 ∞	~ 1	**0.03**
RHIC	protons	250	40,246 ∞	> 30	**0.8**
LHC	protons	450	48,489 ∞	> 1,600	**0.95**
LHC	protons	7,000	13 (energy)/26 (transverse)	∞∞	**< 2**

To estimate electron cooling in LHC we used an energy scaling $\gamma^{7/2}$ typical for RHIC's electron cooler design [8,9], i.e., cooling protons in LHC at 7 TeV is ~10^{10} harder that cooling antiprotons in the Fermilab recycler [7]. Hence, our usage of ∞∞ in an appropriate column.

1. PRINCIPLES OF CEC AT HIGH ENERGIES

Figure 1 shows two (of many) possible layouts of a longitudinal coherent electron cooler. As in a regular electron-cooling scheme, electrons and hadrons should have the same relativistic factor in the CEC:

$$\gamma_o = \frac{E_e}{m_e c^2} = \frac{E_h}{m_h c^2}. \qquad (2)$$

The simplest (and most economical) version of the CEC allows electrons and hadrons to co-propagate along the same straight section. It has a small, weak chicane at the end of the FEL section for adjusting the timing between the electron-beam's modulation and that of the hadron. This scheme imposes limitations on the value of the wiggler parameter, a_w (see discussion in the following sections).

A more elaborate scheme (which also is more flexible and complicated) separates the hadron- and electron-beam so each can be individually manipulated.

For simplification, let us initially consider longitudinal (energy) cooling of the hadron beam. As shown in Section 3, this cooling can be redistributed to include transverse cooling. Otherwise, the principles of the FEL-

based coherent cooling remain the same. It has three parts: The Modulator, the FEL Amplifier/ Dispersion, and the Kicker.

Many processes discussed in this paper are easier to describe in a co-moving (CMS) frame propagating with beam velocity along the straight sections:

$$\beta_o = \frac{v_o}{c} = \sqrt{1-\gamma_o^{-2}} \qquad (3)$$

For high-quality ultra-relativistic (γ_o>>1) hadron- and electron-beams of interest for this paper, the motion of the particles in the CMS frame usually is non-relativistic (***v*** **<<** ***c***, where ***c*** is the speed of the light). In addition, the velocity distribution function is highly anisotropic with the root-mean-square (RMS) velocity spread in the longitudinal direction:

$$\sigma_{V//,CSM} \cong c \cdot \sigma_\delta \qquad (4)$$

(where $\delta = E/E_o - 1$ is the relative energy deviation of a particle) that is much smaller compared with that in the transverse direction

$$\sigma_{V\perp,CSM} \cong c \cdot \gamma_o \sigma_\theta, \qquad (5)$$

where σ_θ is the angular spread of electron beam in the laboratory -frame.

Let us next consider the principle of CEC using this simple model, which is discussed in more detail in the following sections.

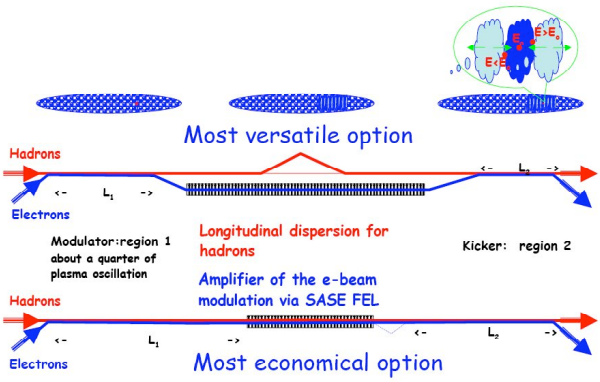

Fig. 1. Schematic layout of the Coherent Electron Cooler with three sections: a) A modulator, where the electron beam is polarized (density modulated) by presence of hadrons; b) an FEL, where density modulation in the electron beam is amplified / longitudinal dispersion for hadrons; c) a kicker, where the longitudinal electrostatic field in the electron beam accelerates or decelerates hadrons. The cooling mechanism is based upon longitudinal dispersion in the hadron beam, i.e., dependence of the time-of-flight on their energy.

In the modulator, individual hadrons attract* electrons and create local density (and velocity) modulation centers at the position of individual hadrons. The process is a linear one, and density modulation on the ensemble of the hadrons is the direct superposition of density modulations induced by individual hadrons. Because of the flat velocity-distribution, the shape of the charge-density modulation resembles that of a flat pancake, with longitudinal extent significantly smaller that the transverse size. When translated into the lab-frame, the longitudinal extent of the pancake shrinks by a factor of γ_o into the nanometer range (see next sub-section). If the length of modulator is chosen to allow a quarter of the plasma oscillation to occur within the electron beam, then, at the end of this section, the electron beam density has a pancake-like distortion with a total excess charge of ***–Ze*** centered at the location of the hadron.

In a FEL-amplifier this modulation of charge density in the electron beam is amplified via the well-known mechanism of exponential FEL growth [13]. Maximum gain in the optical power of such a FEL amplifier is limited by saturation [14,15] to about few millions. Thus, a linear FEL power gain of $\leq 10^6$ and a corresponding amplitude gain, $G_{FEL} \leq 10^3$, are conceivable. In this case, at the exit of the FEL, the individual charge pancake will become a wave-packet (stack) of such pancakes separated by the FEL's resonant wavelength

$$\lambda_o = \frac{\lambda_w}{2\gamma_o^2}(1+a_w^2), \qquad (6)$$

(where λ_w and a_w, respectively, are the wiggler period and wiggler parameter), and, most importantly, with a G_{FEL} -times larger charge in the pancake. The duration of such a wave-packet (i.e., the thickness of the individual pancake stack) is equal to the coherence length of SASE FEL radiation [14,15], and can be as short as a few or a few tens of FEL wavelengths (see the corresponding section below). This pancake stack of charge-density modulation will generate a periodic longitudinal electrostatic field with period of the FEL wavelength: $k_o = 2\pi/\lambda_o$,

$$E(z) = E_o \cdot \sin(k_o z + \varphi) \ . \qquad (7)$$

Hadrons' time of flight *through the dispersion section* depends on the hadrons' energy:

$$(t - t_o)v_o = -D \cdot \delta, \qquad (8)$$

where t_o is time of flight of a hadron with ideal energy and longitudinal dispersion; in general cases, it is a combination of velocity dependence (in drifts) and pass-length (in chicanes) on the hadrons' energy:

$$D = D_{free} + D_{chicane};$$
$$D_{free} = \frac{L}{\gamma^2}; \quad D_{chicane} = l_{chicane} \cdot \langle \vartheta^2 \rangle, \qquad (9)$$

where ϑ is the angle of trajectory in the chicane.

The pass-time of hadron with ideal energy should be equal to that of the space-charge wave-packet. This crude synchronization can be achieved by properly choosing

* Repel, in the case of antiprotons or negatively charged ions

their trajectories (or a_w and small chicane for straight-pass option – see details below). We note that charge-density modulation induced by a hadron propagates with the average longitudinal speed of the electron beam everywhere except at the FEL. There, the wave-packet of charge-modulation propagates with the group velocity of the FEL's optical wave-packet, which falls in between the speed of light and the longitudinal velocity of electrons [16]:

$$v_g \cong c\left(1-\frac{1+a_w^2}{3\gamma_o^2}\right). \qquad (10)$$

Fine tuning the chicane provides for synchronization between the space-charge wave-packet induced by a hadron in such away that the hadron with central energy, E_o, arrives at <u>*the kicker section*</u> just on the top of the pancake of increased electron density (induced by the hadron)[†], wherein the longitudinal electric field is zero. Hadrons with higher energy will arrive at the kicker ahead of their respective pancake in the electron beam, and will be pulled back (decelerated) by the coherent field of the electron beam; we note that positively charged hadrons are attracted to high-density pancakes of electrons. Similarly, a hadron with lower energy falls behind and, as a result will be dragged forward (accelerated) by the clump of electron density. While propagating in a kicker section of length, L_2, the hadrons will experience an energy kick of

$$\Delta\mathbf{E} = -eZ\cdot E_o\cdot L_2\cdot\sin\{kD\delta\} \quad , \qquad (11)$$

where Ze is the hadron's charge ($Z=1$ for protons and $Z=79$ for Au ions). Thus, hadrons with energy deviation within the $|\delta| < \pi/kD$ range will experience a coherent cooling, strength of which is proportional to FEL gain. We dedicated the next section to calculations/estimations of the CEC strength. In Section 4, we discuss the effect of surrounding particles on the cooling mechanism.

Table 2: Main parameters of BNL's ERL

Parameter	**Units**	**Value**
Energy	MeV	54.34
RF Frequency	MHz	703.75
Charge per bunch	nC	5
Normalized transverse emittance	mm mrad	3.2
Relative energy spread		$3.6\ 10^{-4}$
Bunch length	cm	0.76

[†] Again, for negatively charged particle like antiprotons or negatively charged ions, it will be low-density valley. The synchronization conditions are independent of the sign of the hadrons' charge.

2. DETAILS OF THE CEC MECHANISM

A complete and detailed theoretical description of CEC is deeply embedded in early publications on CEC [10, 11, 17] and extends well beyond the scope and size of this paper. Here we focus on discussing a specific scheme of CEC using the ERL-driven high-gain FEL.

For estimating the values we will adopt the parameters of the ERL's electron beam designed at BNL for electron cooling of RHIC [18,19], but scaled to the appropriate energy of interest. Table 2 lists the main beam parameters of this ERL.

Modulator

The modulator in the CEC scheme represents the easiest, but all-important section, from where information about the position and charge of each individual hadron (with charge Ze) is imprinted into the co-moving electron beam. It is desirable to choose the length of the modulator section equal ~ 1/4 of the electron-beam's plasma oscillation, which occurs with frequency of

$$\omega_{pe} = \sqrt{\frac{4\pi n_e e^2}{\gamma_o m_e}} \qquad (12)$$

where n_e is the electron beam's density in the laboratory-frame. After a quarter of a plasma oscillation [20], each hadron will be surrounded by a cloud of electrons with total excess charge of –Ze [‡] It is important to note that screening of the hadron is a dynamic process, i.e., electron cloud screening of a hadron moving with a velocity v_h with respect to the electron plasma also will move with the same velocity, while individual electrons will have very small drag velocity. The process is similar to waves in the water, where the water molecules do not move in average. It also is important to note that the hadron's velocity will affect the size of its surrounding electron cloud. Thus, the velocity spreads of both the electrons and hadrons will determine the size of the Debye ellipsoid. The longitudinal extent of the Debye ellipsoid (electron cloud) in the CSM frame will be a simple product of the plasma's period and the RMS velocity spread:

$$r_{D//,CSM} \sim T_{1/4}\cdot\sigma_{V//,CSM}\,. \qquad (13)$$

In the lab-frame, this dimension is Lorenz-contracted to

$$r_{D//,lab} \sim \frac{\pi c}{2\omega_p}\cdot\frac{\sigma_\delta}{\gamma_o}. \qquad (14)$$

With ω_p~5 10^9 Hz being typical for the electron- beam's parameters, we plan to use for CEC (by keeping n_e/γ ~ const), $r_{D//,lab}$ ~ 750 nm for 100 GeV/u Au ions and $r_{D//,lab}$ ~ 170 nm for 250 GeV proton in RHIC. For 7

[‡] for very cold electron plasma waiting 1/2 of the period will make this value to be –2Ze, which is not very important improvement and which also diminishes in warm plasma.

TeV protons in LHS, $r_{D//,lab}$ ~0.17 nm lies in the soft X-ray range.

Similarly, the transverse Debye radius will be given by

$$r_{D\perp} \sim \frac{\pi c}{2\omega_p} \cdot \gamma\sigma_\theta, \qquad (15)$$

which is frame-independent and covers the range from 0.1 mm to 3 mm for the range of parameters we consider in this paper.

Overall, there are no problems with assuming density modulation in electron beam to be point-like longitudinally and flat transversely, i.e., perfectly suited for amplification in a high-gain short-wavelength FEL.

The only problem with the modulator length required for 1/4 of the plasma oscillation that it is proportional to γ because of the Lorenz time-contraction in the CMS frame (see Table 3). As a practical outcome from such dependence, the length of the modulator for LHC energies becomes impractical. Fortunately, it is possible to employ velocity modulation of the electron along with a buncher similar to that in an optical klystron [21] to effectively modulate the electron beam's density even for LHC energies.

Table 3: Plasma oscillations in CEC's modulator

Machine, Energy, Hadrons, GeV/u	**Electron peak current, A**	**Length of 1/4 plasma oscillation in lab frame, m**
RHIC, 40 GeV, Au	50	4.8
RHIC, 100 GeV, Au, p	50	8.5
RHIC, 250 GeV, p	100	30
LHC, 7 TeV, p	100	1325

A detailed expansion of the theory of this process will not fit into this short paper; hence, we consider the following simple case where hydrodynamics equations can be used, but which contain the most important underlying physics. Let us consider a hadron in rest[§] in cold, uniform electron plasma and its density perturbation caused by the hadron $\tilde{n}(\vec{r},t)$ obeying the equation:

$$\frac{\partial^2 \tilde{n}}{\partial t^2} + \omega_e^2 \tilde{n} = Z\omega_e^2 \delta(\vec{r}). \qquad (16)$$

For amplification in the FEL, we are interested in longitudinal density modulation

$$\rho(z) = \iint \tilde{n}(\vec{r})dxdy, \qquad (17)$$

and, to be exact, in its spatial harmonic on FEL with k-vector $k_{cms} = k_o / 2\gamma\beta_z$:

$$\frac{\partial^2 \rho_k}{\partial t^2} + \omega_e^2 \rho_k = \frac{Z\omega_e^2}{2\pi}. \qquad (18)$$

Equation (18) has a straightforward solution

$$\rho_k = \frac{Z}{2\pi}\left(1 - \cos\omega_e t\right), \qquad (19)$$

which indicates that initially the growth of density modulation $\sim(\omega_e t)^2$ is very inefficient for $\omega_e t << 1$, i.e., for a short modulator, as in the LHC case. Fortunately, velocity modulation grows linearly and for the case of $t = L_1/\gamma_o v_o << 1/\omega_e$, one can easily calculate the average longitudinal CMS velocity field induced by presence of a hadron in electron beam with a radius a and the resulting energy modulation in the lab-frame:

$$\left\langle \frac{\delta E}{E} \right\rangle \cong -2Z\frac{r_e}{a^2} \cdot \frac{L_1}{\gamma} \cdot \left(\frac{z}{|z|} - \frac{z}{\sqrt{a^2/\gamma^2 + z^2}} \right). \qquad (20)$$

Propagating such electron distribution through a simple buncher with longitudinal dispersion $D = \lambda_o / 2\pi\sigma_{\delta_e}$ [21] will create maximum density modulation with wave-vector k_o.

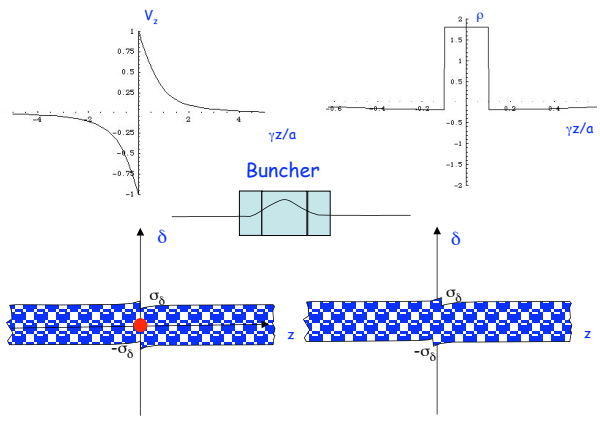

Fig. 2. Longitudinal velocity map (left upper graph) and resulting density modulation (right upper corner graph) after propagation through a simple 3-pole buncher (depicted in the middle). The simulations were done with Mathematica [22].

The two pictures in the bottom of Fig. 2 illustrate the process in the case of a simple phase-space distribution: an initial continuous electron beam (very long, compared with a a/γ_o), with energy spread $\pm\ \sigma_\delta$ that experiences energy modulation in the presence of a hadron (left picture). In a buncher, electrons with higher energies slip forward and those with lower energies slip backwards to create density modulation at the location of the velocity rift. In short, proper use of buncher for electron beam[**] facilitates the generation of density modulation

[§] Obviously we are using CMS fame of electron beam. Generalization for a case of moving hadron is straightforward: $\frac{\partial^2 \tilde{n}}{\partial t^2} + \omega_e^2 \tilde{n} = Z\omega_e^2 \delta(\vec{r} - \vec{v}t)$ and does not change physics of the process in any significant manner

[**] in kinematics terminology buncher is nothing else but a devise reducing effective longitudinal mass and increasing electrons mobility in z-direction

proportional to first order of the phase of plasma oscillations, $\Delta Q \sim -eZ \cdot (\omega_e t)$. In the case CEC for 7 TeV LHC protons, the use of a short modulator section (by LHC standards) of 100 m and a buncher provides for very short cooling times of less than one hour.

FEL amplifier for electrons and the dispersion for hadrons

FEL amplifier: It is well known that initial modulation (and noise) in the electron beam density will be amplified. The resulting TEM field, energy modulation, and density modulation, all evolve together in FEL. If the FEL gain is limited to below G~10^3 (to stay away from saturation), the resulting amplitudes represent the simple linear superposition of amplified signals from each hadron plus noise from the electrons [14]. In particular, density modulation in electron beam at the exit of FEL (and the entrance of the kicker section) can be expressed as

$$\rho = -k_o e \begin{pmatrix} Z \sum_{hadrons} K_{FEL}\left(t - t_h - T_f\right) \\ + \sum_{electrons} K_{FEL}\left(t - t_e - T_f\right) \end{pmatrix}. \qquad (21)$$

where K_{FEL} is the Green function of FEL in response to a δ-function charge perturbation, and T_f is the time-of-flight of electron "wave-packet" from the end of modulator section to the beginning of the kicker section. The main features of the Green function are that it has carrying frequency at FEL wavelength $\omega_o = k_o c$, has a duration of so-called correlation length (i.e., the number of oscillation is ~ number of wiggler periods per FEL gain length). The inverse duration of this "wave-packet" defines the FEL amplifier's bandwidth.

In addition, this "wave-packet" propagates with the group velocity of FEL signal (10), i.e., the peak of its amplitude moves forward with respect to electron beam. As the result, for $a_w^2 < 1/2$, the group velocity in FEL is larger that that of the hadron.

This situation affords an excellent opportunity to use simple co-propagation of electrons and hadrons through the FEL with a small chicane to delay the electrons, which are lighter and easier to manipulate. This delay for few FEL wavelengths puts the hadron on the top of the charge density "wave-packet" that it induces itself in the modulator.

In the case of $a_w^2 > 1/2$, such simple scenario is not applicable, and the electrons and hadrons should be separated. The hadrons should be delayed for the equivalent of a few FEL wavelengths so to overlap with the maximum of the "wave-packet"– the peak of K_{FEL}.

In both cases, hadrons with designed energy should finish in a zero electric field. This will require sub-FEL-wavelength adjustments, which do not interfere with large-scale timing. The process is similar to maximizing gain in an optical klystron by tiny variation of longitudinal dispersion in its buncher [21].

We take a simple approach for estimating FEL parameters. We assume that wiggler period of 5 cm with reasonable a_w works for all cases, and that selected FEL wavelength is significantly longer that $r_{D//,lab}$. We use analytical formulae derived by Ming Xie [23] to estimate the lengths of the FEL gain,

$$L_{G3D} = L_{G1D}(1+\Lambda); \quad L_{G1D} = \frac{\lambda_w}{4\pi\rho\sqrt{3}}; \qquad (22)$$

and the necessary FEL length for a desirable gain. Table 4 gives some resulting parameters.

The optimum longitudinal dispersion for hadrons can be chosen from two considerations: a) Maximum cooling decrement for the typical RMS energy spread δ: $\sin kD\delta = 1$; b) the requirements for cooling all particles within the energy spread {-δ_{max},+δ_{max}}: $\sin kD\delta_{\max} = 0$. Reasonable compromises can be chosen, viz., $D = 1/k\sigma_\delta$, which we will use for our further estimations.

Table 4: Main parameter of FELs for CEC

Machine, Energy, Hadrons, GeV/u	**λ_o**	**FEL Length, m**	**FEL Gain, *amplitude***
RHIC, 40 GeV/n, Au	18 μm	6.5	100
RHIC, 100 GeV, Au, p	3 μm	9	200
RHIC, 250 GeV, p	500 nm	15	100
LHC, 7 TeV, p	10 nm	25	500

Kicker section

In the kicker section, we have sinusoidal modulation of the electron beam's density, which, in the CMS frame is

$$\tilde{\rho} = \frac{k}{2\gamma_o} \frac{G_{FEL} \cdot Z \cdot e}{B} \cdot \sin\left(kz/2\gamma_o\right) \qquad (23)$$

where G_{FEL} is the FEL amplification (for amplitude!) and $B = 2\pi\beta_\perp \varepsilon_n / \gamma_o$ is the beam's transverse area, ε_n is the normalized transverse emittance of electron beam, and $\beta_\perp$ is the transverse beta-function. In the CMS frame the field is electrostatic and it is sufficient to solve a trivial $div\vec{E} = 4\pi\tilde{\rho}$ to find it. Since the electron beam in the ERL has close to a uniform distribution and the transverse size of the beam is significantly larger the modulation wavelength, the field near the beam's axis is predominantly longitudinal. Because Lorentz transformation does not change the amplitude of a longitudinal field component, we have in the lab-frame:

$$E_z \cong \frac{2G_{FEL} \cdot Ze}{\beta_\perp \varepsilon_n} \gamma_o \cdot \sin\left(k_o\left(z - \mathrm{v}_o t\right)/\beta_o\right). \qquad (24)$$

This field will exist for about a quarter of the plasma's oscillation and, therefore, for a short kicker length, L_2, one can finally write the change of the hadron's energy

per turn in CEC ($r_p = m_p c^2 / e^2$ is the classical radius of proton, and A is atomic number of hadron):

$$\Delta\delta = -G_{FEL} \cdot \frac{r_p \cdot L_2}{\beta_\perp \varepsilon_n} \frac{Z^2}{A} \cdot \sin(kD\delta) \quad . \qquad (25)$$

For the natural choice of dispersion $D = 1/k\sigma_\delta$, the cooling decrement for small energy deviations per turn will be

$$J_{CEC} = 2G \cdot \frac{r_p}{\sigma_\delta \varepsilon_n} \cdot \frac{L_2}{\beta_\perp} \cdot \frac{Z^2}{A}. \qquad (26)$$

It worth noting that this expression is energy independent, i.e., it promises to be effective in high-energy colliders like the LHC.

Cooling heavy ions using CEC is easier than cooling protons because of $Z^2/A > 1$ (factor 31.7 for $_{197}Au^{78}$ ions in RHIC). It is also important to point out that because of synchrotron oscillations, the value of the CEC decrement (26) should be divided by 2, i.e., the standard ratio for an oscillator.

Table 5: Main parameter of kickers for CEC

Machine, Energy, Hadrons, GeV/u	**L_2, m**	**Hardon bunch length, nsec**	$\frac{\sigma_{s,e}}{\sigma_{s,h}}$
RHIC, 40 GeV, Au	3	2.5	100
RHIC, 100 GeV, Au, p	5	1	200
RHIC, 250 GeV, p	10	1	100
LHC, 7 TeV, p	50	0.25	500

In practice, hadron bunches are usually much longer than electron bunches. In this case, the decrement (26) should be multiplied by the ratio of the bunch lengths:

$$\langle J_{CEC} \rangle = J_{CEC} \cdot \frac{\sigma_{s,e}}{\sigma_{s,h}} \qquad (27)$$

These values, listed in Table 5, were used for calculating the cooling times in Table 1.

3. TRANSVERSE COOLING

Similarly to re-distribution of the damping increments of synchrotron radiation, it is possible to re-distribute decrements of CEC within the boundaries of the "sum of decrement" theorem:

$$J_s + J_x + J_y = J_{CEC}. \qquad (28)$$

One easy way of providing cooling of the horizontal emittance is to use a chromatic chicane, as shown in Fig.3 below. In a chromatic chicane, symplectic conditions couple the arrival time of these wave fronts with the transverse coordinates and angles, which provides an energy gain in the kicker's dependence of

$$\Delta \mathbf{E} = -eZ^2 \cdot E_o \cdot L_2 \cdot \sin\{k(D\delta + R_{16}x' - R_{26}x + R_{36}y' + R_{46}y)\}; \qquad (29)$$

where R_{nm} are elements of transport matrix through chicane (indexes 1,2,3,4,5,6 belong to canonical variables x,x',y'y', -ct and δ).

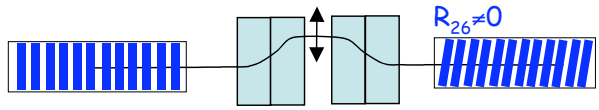

Fig. 3 Chromatic chicane tilts the wave-fronts of charge-density modulation in an electron beam and couples the energy kick with the horizontal position.

The energy change causes a local change in the closed orbit, and hence, a change in the horizontal betatron displacement that is proportional to horizontal dispersion η: (*sin* in eq. (29) is linearized for simplicity)

$$\Delta x = -\eta \cdot eZ^2 \cdot E_o \cdot L_2 \cdot kR_{26}x + ... \qquad (29)$$

The typical decrement in horizontal cooling for such system will be

$$J_x \sim \frac{\eta \sigma_\varepsilon}{\sigma_x} J_{CEC}.$$

Cooling of both transverse directions will require at least local transverse coupling (see eqs. (29) and (30)) and non-zero projection of the dispersion vector on both eigen vectors of the betatron modes. A detailed description of such scheme goes beyond the scope of this paper.

4. COLLECTIVE EFFECTS AND LIMITATIONS ON THE CEC DECREMENT

As we discussed earlier, the electric field in the electron beam is a linear superposition of "wave-packets" excited by all hadrons and electrons, which carry information about their origin, i.e., the position of each hadron and electron in the beam (note that hadrons with a non-unit charge have an additional Z in front of the identical field-function, E_o).

$$E_z = Z \cdot \sum_{i,hadrons} E_o(v_o t - z + z_i) \cdot \sin k(v_o t - z + z_i)$$
$$- \sum_{j,electrons} E_o(v_o t - z + z_j) \cdot \sin k(v_o t - z + z_j)$$

Hence, the energy kick for each hadron will contain kick from all hadron and electrons within the extent of the ''wave-packet". The action of the self-induced field, which we already considered separately in the previous section, is coherent and is repeated turn after turn. At the same time, the relative phases of hadrons at an optical (μn and nm) scale are completely random, and any correlations are completely eliminated after each turn (similar to the storage ring in FELs [24]). Let us define $\tilde{N} = \tilde{N}_h + \tilde{N}_e / Z^2$, where $\tilde{N}_h$ and $\tilde{N}_e$ is the number of

electrons in the FEL coherence length[††] $L_c = N_c \lambda_o$, which "communicate with each other". Writing the equation of evolution of RMS energy and averaging <sin^2> of the random phase, we get the standard equation for RMS spread (analogous to stochastic cooling calculations):

$$\frac{d\sigma_E^2}{dn} = -2\Delta \frac{kD}{\mathbf{E}_o}\sigma_E^2 + \frac{1}{2}\Delta^2 \tilde{N}; \quad \Delta = eZ^2 \cdot L_2 \cdot E_o,$$

with two important consequences. First is that maximum cooling rate can not be larger that 1/Ñ per turn. This limitation is taken into account by properly selecting the FEL gain for cooling rates shown in Table 1.

Table 6: Other parameters for the CEC

Machine, Energy, Hadrons, GeV/u	**Nc**	**Lc, μm**	N_h per bunch	Ñ
RHIC, 40 GeV, Au	20	360	$1\ 10^9$	$6\ 10^5$
RHIC, 100 GeV, Au	20	60	$2\ 10^9$	$5\ 10^5$
RHIC, 250 GeV, p	10	5	$2\ 10^{11}$	$7\ 10^6$
LHC, 7 TeV, p	10	0.2	$1.15\ 10^{11}$	$0.4\ 10^6$

Second consequence is that the stationary (self-consistent) energy spread is proportional to the number of particles in the coherence length (i.e., it is desirable to operate the FEL at as short wavelength as possible!), and also to the cooling rate:

$$\frac{\sigma_E^2}{\mathbf{E}^2{}_o} = \frac{1}{4kD} \cdot \frac{\Delta}{\mathbf{E}_o} \cdot \tilde{N} \qquad (30)$$

Finally, we mention that throughout the entire suite of numerical examples used in this paper we ensured that emittances of both the electron- and hadron- beams did not exceed the FEL wavelength divided by 4π, This action ensures the preservation of phase correlations at the scale of the FEL wavelength.

5. CONCLUSIONS

In this paper, we showed that coherent electron cooling (CEC) can be very promising method of cooling beams in high-energy hadron colliders. The CEC takes full advantage of high-gain FELs based on high brightness ERLs. Cooling of 100 GeV/n ions and 250 GeV protons in RHIC seems to be straightforward with this method.

Cooling protons in LCH at 7 TeV also seems to be possible, but may require a slightly more elaborate scheme (such as a buncher and separating the beams).

Proof-of-principle experiment to cool Au ions in RHIC at ~ 40 GeV/n is feasible using the existing R&D ERL, which is under construction in BNL's Collider-Accelerator Department (C-AD). Commissioning this ERL is planned for early 2009. Moving this accelerator into RHIC tunnel may take from two to three years.

In this paper we assumed standard statistical shot-noise in the electron beam and did not relied on any possible noise suppression, which is being actively discussed by the electron- cooling community.

[††] again, for simplicity of we used N_c=10 and 20 for numerical example in this paper. There is a rigorous way of defining the correlating length for a specific FEL by using correlation length of its Green function K:

$$L_c = \frac{1}{|K_{\max}|^2} \int_{-\infty}^{\infty} |K(z)|^2 dz$$

ACKNOWLEDGEMENTS

We want to thank Ilan Ben Zvi and Thomas Roser (BNL) for encouraging discussions and well-pointed questions and suggestions, which lead to some of the technical solutions presented in this paper. We are grateful to Anatoly Kondratenko for his wiliness to share his knowledge of processes in high-gain FELs. The first author also would like to thank C-AD's electron cooling and R&D ERL groups for being a test-bed for our initial discussion of ideas used for the FEL-based CEC, and for generating several valuable suggestions.

REFERENCES

[1] M. A. Furman, Proc. 1991 Particle Accelerator Conf., San Francisco, May 6-9, 1991, p. 422.
[2] V. Ptitsyn et al, http://www.bnl.gov/cad/eRhic/Documents/AD_Position_Paper_2007.pdf
[3] V.N. Litvinenko et al., Proc. PAC 2005 http://cern.ch/AccelConf/p05/PAPERS/TPPP043.PDF
[4] http://www-td.fnal.gov/LHC/USLARP/References/IR_Upgrades.pdf
[5] R.Calaga et al., http://care-hhh.web.cern.ch/care-hhh/Literature/TUPAS089.pdf
[6] I.Meshkov, Phys. Part. Nucl., 25, 6 (1994) 631
[7] S.Nagaitsev et al., Phys. Rev. Lett. **96**, 044801 (2006)
[8] Ilan Ben-Zvi, et. al., NIM **A532**, 177, (2004)
[9] A.V.Fedotov et al, New J. of Physics **8** (2006) 283.
[10] Ya. S. Derbenev, Proc. of 7th All-Union Conference on Charged Particle Accelerators, 14-16 October 1980, Dubna, USSR, p. 269 (in Russian)
[11] Ya.S. Derbenev, Preprint, UMHE91-28, 1991
[12] A. Mikhalichenko, M. Zolotorev, Phys. Rev. Lett., **71**, p.4146 (1993).
[13] A.M.Kondratenko, E.L.Saldin, Part. Acc. **10** (1980) 207
[14] E.L.Saldin, E.A.Schneidmiller, M.V.Yurkov, The Physics of Free Electron Lasers, Springer, 1999
[15] C.Pelegrinni, NIM **A475**, 177, (2001) p.1
[16] see Ref. [14], page 365
[17] Ya.S. Derbenev, Dissertation, Novosibirsk, 1978
[18]http://www.bnl.gov/cad/ecooling/docs/PDF/Electron_Cooling.pdf
[19] D. Kayran et al., Proc. of PAC 2007 http://accelconf.web.cern.ch/AccelConf/p07/PAPERS/THPAS096.PDF
[20] see for example, C.L. Longmire, Elementary plasma physics, Interscience publishers, 1963
[21] N.A.Vinokurov, A.N.Skrinsky, Preprint of Institute of Nuclear Physics, INP 77-67, 1977, Novosibirsk
[22] Mathematica, © Wolfram Research, Inc
[23] M. Xie, NIM **A445**, 59 (2000).
[24] V.N.Litvinenko et al, NIM **A475**, 177, (2001) p.65