A COMPACT SOFT X-RAY FREE-ELECTRON LASER FACILITY BASED ON A DIELECTRIC WAKEFIELD ACCELERATOR*

C. Jing#, P. Schoessow, A. Kanareykin, Euclid Techlabs LLC, Solon, OH 44139
J. G. Power, HEP Division, Argonne National Laboratory, Argonne, IL 60439
R. Lindberg, A. Zholents, APS, Argonne National Laboratory, Argonne, IL 60439
P. Piot, Northern Illinois University, Department of Physics, DeKalb, IL 60115

Abstract

X-ray free-electron lasers (FELs) are expensive instruments with the accelerator holding the largest portion of the cost of the entire facility. Using a high-energy gain dielectric wakefield accelerator (DWA) instead of the conventional accelerator may reduce the facility size and, significantly, its cost. We show that a collinear dielectric wakefield accelerator can, in principle, accelerate low charge and high peak current electron bunches to a few GeV energy with up to 100-kHz bunch repetition rate. Several such accelerators can share the same tunnel and cw superconducting linac (operating with a few-MHz bunch repetition rate), whose sole purpose is feeding the DWAs with wake producing low-energy, high-charge electron bunches with the desired periodicity. Then, ten or more x-ray FELs can operate independently, each using its own linac. In this paper, we present an initial case study of a single-stage 850-GHz DWA based on a quartz tube with a ~100-MV/m loaded gradient sufficient to accelerate a 50-pC main electron beam to 2.4 GeV at a 100-kHz bunch repetition rate in just under 30 meters. While the accelerated electron beam has a large energy chirp, we show that FEL gain can be maintained by appropriately tapering the undulator, although other schemes may also be possible.

INTRODUCTION

A number of FEL facilities are in operation around the world and more have been planned [1, 2]. While tremendously effective in providing extreme photon fluxes, these machines can only accommodate a small number of users at a time. To address this limitation, multi-user, multi-FEL facilities have been recently proposed (see, for example, [3]). The mainstream approach is to use a cw superconducting rf (SRF) linac to provide electron bunches of a few GeV at a 1-MHz repetition rate, which then feeds a switchyard of ten or more FELs at the end of the linac. Here, we propose a new approach where a 1-MHz repetition rate cw SRF linac of much lower energy (200 MeV) is used to feed a switchyard of ten DWA linacs, providing electron bunches with a few-GeV energy to ten FELs. As a result, the construction and operational costs of the facility, which are largely defined by the large SRF linac, can be significantly reduced.

A high-gradient DWA structure is the key to our proposal. Advanced accelerator studies aimed at a future high-energy collider have shown impressive results, achieving multiple-GV/m energy gradients in various wakefield acceleration schemes [4, 5].

A DWA SCHEME FOR FEL

In a collinear dielectric wakefield accelerator, the fields generated by a leading, high-charge drive bunch (either a single drive bunch or a train of drive bunches) is used to accelerate a trailing, low-charge main bunch that contains a relatively small amount of charge. The collinear configuration has the two beams traversing the accelerator along the same trajectory so that the energy is directly transferred from the drive bunch to the trailing main bunch. The DWA structure is very simple, consisting of a cylindrical dielectric tube with an axial vacuum channel inserted into a metal outer jacket. The dielectric constant and the inner and outer radii of the dielectric tube are chosen to adjust the fundamental monopole mode (TM_{01}) frequency excited by the passing beam. The phase velocity of the mode equals the beam velocity ~c.

The overall facility layout is shown in Figure 1. The scheme consists of 10 parallel DWAs (2.4-GeV each) sharing one drive beam. The 200-MeV, 1-MHz drive beam consisting of 1.6-nC, 3.3-ps shaped bunches is generated by an L-band cw SRF linac. The switchyard distributes electron bunches to the 10 DWAs on a rotating basis such that the repetition rate of each DWA is 100 kHz. This keeps the average rf power dissipation of the accelerating structure at a manageable level, as shown later. The drive bunch has a special double triangular current profile, shown in Figure 2. The current shape is tailored to maximize the transformer ratio, defined as the ratio of the maximum accelerating field behind the drive bunch over the maximum decelerating field inside the bunch. Current profile shaping of the drive bunch can be implemented via an emittance exchange (EEX) [6] or double EEX (DEEX) beamline, which is discussed later. Emittance exchange also can be used to shape the low-charge main bunch and fix it at a precisely defined distance trailing the drive bunch. The effective acceleration length for each beamline is 24 m, over which distance the drive beam has lost 83% of its energy.

*Work supported by U.S. Department of Energy, Office of Science, under Contract No. DE-AC02-06CH11357.
#jingchg@hep.anl.gov

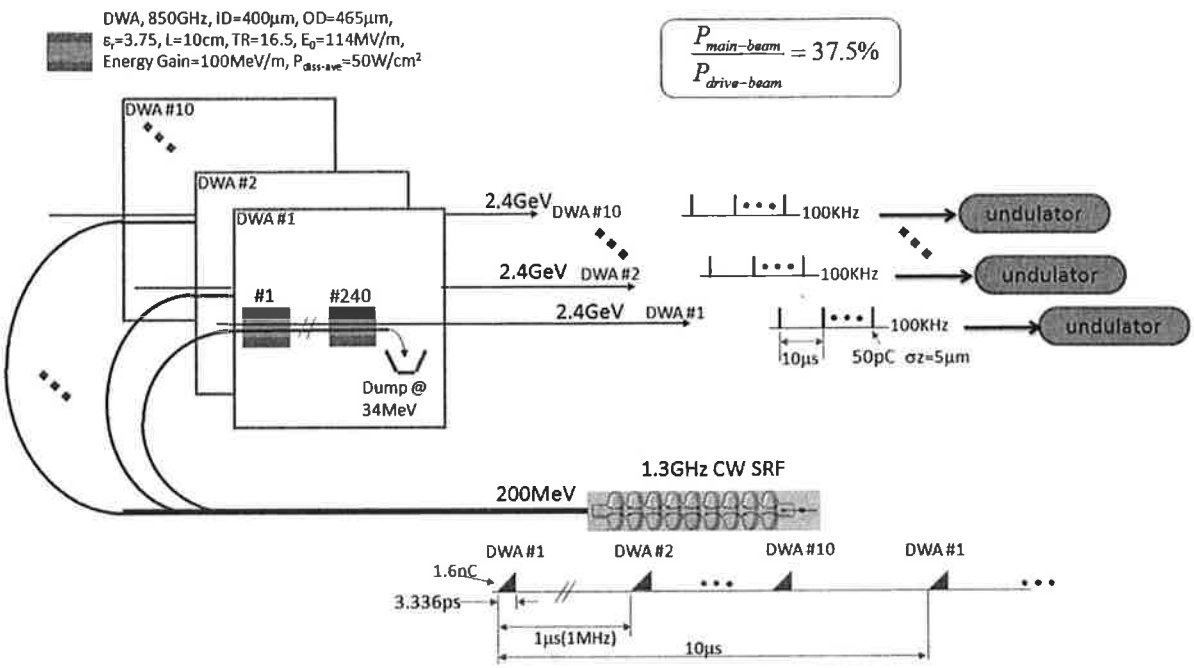

Figure 1. A schematic of the FEL light source facility showing a cw superconducting linac, transport lines, and an array of dielectric wakefield accelerators. Essential electron beam and accelerator parameters are shown.

Each DWA has 240 10-cm-long low-cost quartz-based tubes. (In principle, a tube can be any length as long as the electron beam control requirement is satisfied.) All tubes have a small 400-micron hole, critical for achieving a high-gradient acceleration using a relatively modest 1.6-nC drive bunch. Stable propagation of a drive bunch with energy varying from 200 MeV to 30 MeV along the linac through this hole presents a significant challenge that will be addressed in future publications and R&Ds.

In our strawman design, the transformer ratio is 16.5. The peak gradient behind the drive bunch is 114 MV/m, which means the drive bunch energy loss is 6.9 MeV/m. All the values are calculated using four longitudinal modes, but the dominant contribution is from the fundamental mode because the structure was intentionally designed to minimize the loss factor of all higher-order modes (HOMs). This is very important for the drive bunch to achieve a high transformer ratio, as well as to mitigate the self-wake inside the main beam (the main beam has much shorter bunch length and easily excites HOMs).

Two major factors were considered in the choice of the main bunch parameters: 1) to minimize beam loading, the main bunch charge was made as high as possible without causing a serious drop in the net gradient; 2) to maximize beam brightness, we kept the charge as low as possible to yield small energy spread and a short bunch. A short bunch makes it easier to minimize the energy spread. The self-wake from the main beam depends on the wakefield structure, is proportional to the charge, and is inversely proportional to the bunch length (approximate to some extent). In the collinear wakefield acceleration scheme, both the drive and main beam share the same beam channel. In general, while the structure is optimized to maintain a single-mode excitation for the drive beam, multiple modes are excited by the main beam due to its shorter bunch length. But the self-wake inside the main beam will converge very fast because the loss factor of HOMs becomes negligible after the first few modes. 50 pC/bunch and 5-m bunch length (rms) were chosen in the design to allow a reasonable beam loading and ~1.2-kA peak current. Another critical parameter of an accelerator is the efficiency. The overall efficiency of the proposed scheme can be roughly estimated by cascading the efficiencies of each subsystem; for example, klystron AC-to-rf efficiency, rf-to-drive beam efficiency, and drive-to-main beam efficiency. The drive-to-main beam efficiency, 37.5%, can be obtained using the parameters in the layout. If we assume 50% efficiency for each of the other subsystems, the overall wall plug efficiency is around 5-10% with consideration of power consumption from the cooling, magnets, infrastructure, etc.

BEAM LOADING AND ENERGY SPREAD

In a collinear wakefield acceleration scheme, the structure radius a is chosen to be small so that the drive beam excites a strong wakefield. However, because the bunch length of the main bunch is shorter than that of the drive, the main beam self-wake (deceleration) can be severe, so single bunch beam loading must be considered. Figure 2 shows the overall wakefield plot with a main bunch trailing behind the drive bunch at a precisely defined distance. Due to the main beam's short bunch length (~5 μm, rms), we included 10 modes in the calculation even though only the first five modes

significantly contribute to the beam wakefield. Based on calculations, a ~10-MeV centroid (i.e., mathematical expectation) energy loss per 10-cm-long structure is obtained, and the induced correlated energy spread on a mono-energetic injected beam is ~150 keV (rms). Considering the same rate of acceleration through the entire linac, we calculate ~1.5% relative energy spread in the bunch at the end of the linac. One way to reduce the beam loading and, thus, the correlated energy spread, would be to increase the bunch length of the main beam. For example, one could imagine accelerating a long bunch length main beam through the accelerator and then compressing it before the entrance to the FEL. Alternatively, a tapered undulator can be used to obtain a useful FEL gain. Details are discussed in the last section.

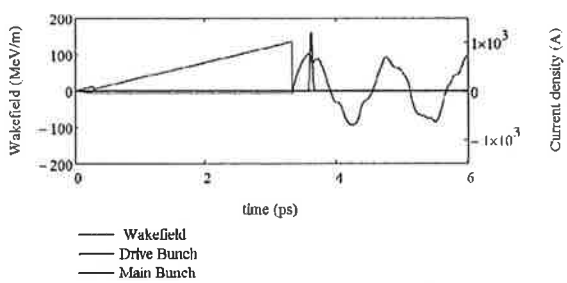

Figure 2. Wakefield due to drive and main bunches.

THERMAL ESTIMATION AND COOLING

Note that in the ~THz regime, the rf attenuation is very high so that the wakefield power will be exhausted in 10 cm. This means that the rf packet generated by a drive bunch is only 333 ps long. The single-layer DLA structure has a low electric field in the dielectric (Es ~$1.7Ea$, where Ea is the accelerating field) but a high magnetic field on the metal surface. Therefore, the rf power is mostly dissipated in the copper surface and dielectric losses contribute little for the low loss materials considered here.

Two types of thermal issues were considered when designing the accelerator: rf pulsed heating and average thermal heating. Pulse heating is an instantaneous phenomenon occurring primarily within a skin depth of the metallic surface. Rf pulse heating usually is not an issue for wakefield accelerators due to the short rf pulse length from the drive bunch. The temperature rise from rf pulse heating is only 17°C (conditions: 100 MV/m of loaded gradient corresponding to 680-kA/m peak magnetic field at the copper surface in a 100-ps square pulse to simulate the 333-ps heavily attenuated actual pulse).

Average thermal heating of the accelerating structure due to rf power dissipation is the main factor limiting repetition rate of the conventional S-band and X-band high-energy gradient accelerators. Fortunately, the problem is much less severe in the DWA because of the small amount of energy used to excite the wakefield and the short period of time that the wakefield remains in the structure; these properties result from the short structure length, high group velocity, and high loss in the THz range. Calculations based on the parameters in Figure 1 show that the average power load from the wakefield on the copper tube holding the dielectric is ~50 W/cm^2. This power can be easily handled by conventional water cooling, resulting in less than 0.5°C temperature rise across the copper, as seen in Figure 3.

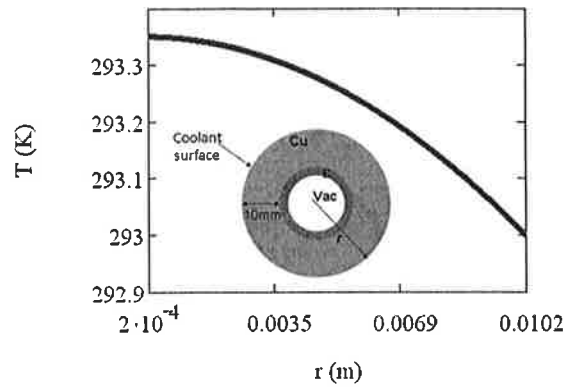

Figure 3. Estimated temperature rise from the dielectric surface to the 293 K cooling surface.

SHAPED DRIVE BUNCH GENERATION

One effective approach to enhance the transformer ratio is to use a shaped bunch, e.g., a bunch with triangular current profile. In this proposed DWA scheme, a new variation of the triangular bunch, a double-triangular bunch, is used to further enhance the transformer ratio to nearly twice that of the single-triangular bunch [7]. The current profile of the double-triangular bunch can be described as

$$I(t) = \begin{cases} I_0\omega t & 0 < t < \frac{\pi}{2\omega} \\ I_0\omega t - I_0 & \frac{\pi}{2\omega} < t < T \end{cases}, \quad (1)$$

where ω is the resonant frequency of the wake function, T is the bunch duration, and I_0 is a constant with units of current that scales with the total charge. Figure 2 illustrates the current profile of the double-triangular bunch and its corresponding longitudinal wakefield. The maximum achievable transformer ratio in theory is $R = \sqrt{1 + (\omega T - 1)^2}$ for a single-mode wakefield accelerator.

Currently, we consider shaping via a transverse-to-longitudinal phase space swap, by which shaping in the transverse dimension can be exchanged into a tailored longitudinal profile using the idea of emittance exchange [6-9]; this can be further improved with DEEX [10]. Preliminary calculations prepared for a proof-of-principle experiment at the AWA facility [11] with a 5-nC Gaussian bunch and plausible transverse and longitudinal emittances at 75 MeV are shown in Figure 4. After the transverse mask at the entrance of the EEX beamline, 34% of the beam is blocked, resulting in a bunch charge of 1.7 nC. The same mask is used to shape the main

beam. The focused beam can be contained within a 200 μm × 200 μm area (see Figure 4).

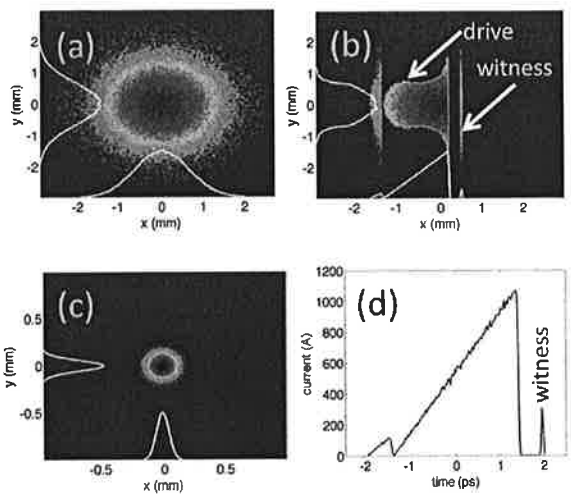

Figure 4. Double-triangular bunch generation using an EEX-based bunch shaper, where the transverse mask is tailored to generate a double-triangular bunch ["drive" in (b)] followed by a low-charge main bunch ["witness" in (b) and (d)]. (a) and (b) show the transverse beam density before and after the mask, respectively. (c) and (d) plot the beam density at the EEX exit, with (c) the transverse profile and (d) the current profile (positive time corresponds to the tail of the bunch).

BBU INVESTIGATION

The small aperture of the wakefield structures make them particularly vulnerable to the single-bunch breakup (BBU) instability driven by transverse modes (dominantly HEM_{11} mode for the DWA structures). Figure 5 shows an analytical calculation of the transverse wakefield generated by the double-triangular bunch in the DWA structure described in Figure 1. The maximum transverse wakefield of 2.5 MV/m for 1.6-nC drive bunch with 10-μm offset is large and must be mitigated by a strong external focusing, creating a large chirp in a frequency of betatron oscillations along the bunch similar to a BNS damping [12]. Specific design solutions and additional studies will be addressed in the follow-up publication.

FEL CONSIDERATIONS

As mentioned previously, the accelerated bunch also couples strongly to the dielectric structure, generating a strong wakefield that tends to decelerate the tail of the bunch with respect to the head of the bunch. We show the computed wakefield for several accelerated bunch lengths in Figure 6, which indicate that the wakefield produces a nearly linear energy chirp of the beam if the rms beam width $\sigma_e \leq 15\mu m$.

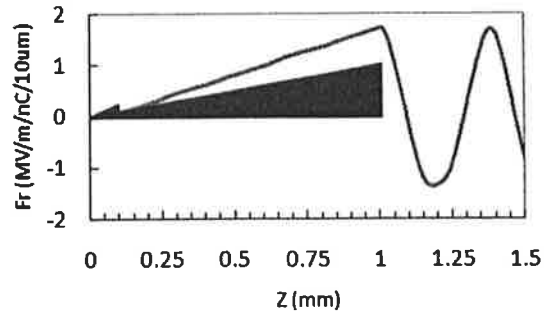

Figure 5. Analytical result of the transverse wakefield by a shaped bunch passing through the DWA structure described in Figure 1. Red area indicates the bunch shape and blue curve is transverse wakefield normalized to 1 nC and 10-μm offset.

The electron beam energy chirp will severely inhibit the FEL performance if the variation of the normalized energy variation over a coherence time $\sim 1/(\rho\omega_0)$ is comparable to FEL bandwidth ρ. More precisely, for a linear chirp over twice the rms width of the beam, the gain is not affected provided $\Delta\gamma/\gamma_0 \ll 4\pi\rho^2\sigma_e/\lambda_0$. As an example, the self-wake of the 5-μm accelerated bunch gives rise to a normalized energy variation $\Delta\gamma/\gamma_0 \approx 0.2$, while a soft x-ray FEL operating at 1 nm typically has $\rho \sim 10^{-3}$, in which case $4\pi\rho^2\sigma_e/\lambda_0 \approx 0.12$. Since this effect only gets worse at longer wavelengths, it may appear that such a beam is not suitable for FEL operation.

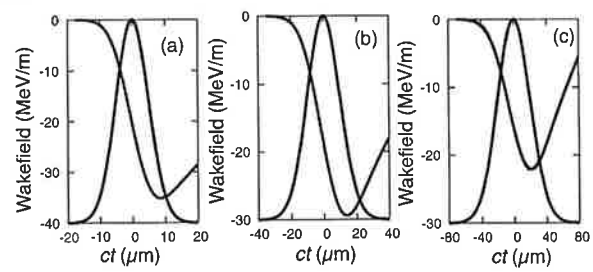

Figure 6. Wakefield across the accelerated bunch for (a) 5-μm, (b) 10 μm, and (c) 20-μm bunch lengths. The decelerating field is red, while the current profile is blue

However, one can overcome this limitation by varying either the undulator period or strength along its length to effectively counteract the energy chirp [13]. The basic physics can be understood by considering radiation initially generated at the middle of the bunch, which we assume satisfies the local FEL resonance condition at $z = 0$ given by $\lambda_0 = [1 + K(0)^2]\lambda_u(0)/2\gamma_0^2$. During subsequent evolution, the radiation slips forward toward the head of the bunch, where it interacts with electrons whose energy γ_+ is higher. Thus, to maintain resonance we require $\lambda_0 = [1 + K(z)^2]\lambda_u(z)/2\gamma_+^2$, meaning that we must increase the magnetic field strength K and/or the undulator period λ_u as a function of z. For a linear chirp, we have $\gamma_+ = \gamma_0 + (d\gamma/dt)dt$, so that using $K(z) = K(0) + (dK/dz)dz, \lambda_u(z) = \lambda_u(0) + (d\lambda_u/dz)dz$, and $dt/dz = -\lambda_0/\lambda_u$ (time decreases toward the bunch head), we find that resonance is maintained if

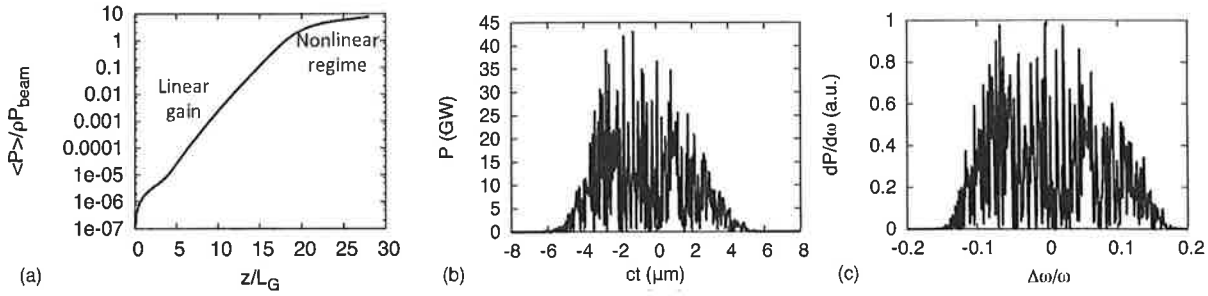

Figure 7. (a) FEL power evolution where the undulator parameters have been appropriately chirped so as to counteract the electron beam energy variation. (b) and (c) Temporal and spectral profiles, respectively, of the output radiation after 20 linear gain lengths, where nonlinear saturation has slowed the power growth.

$$\frac{d\lambda_u}{dz} + \frac{\lambda_u K}{1+K^2/2}\frac{dK}{dz} = -\frac{2\gamma_0}{\gamma_0 c}\frac{d\gamma}{dt}. \qquad (2)$$

Note that the DWA naturally produces a negative chirp $d\gamma/dt < 0$, and that over the entire undulator length the required fractional change of either the period $\Delta\lambda_u/\lambda_u \sim \Delta\gamma/\gamma_0$ or the normalized field strength $\Delta K/K \sim \Delta\gamma/\gamma_0$ (for K of order unity or larger) is approximately 20%.

In Figure 7(a) we plot power evolution along the undulator for a 5-μm accelerated bunch when the undulator is properly tapered to counteract the large energy chirp. We see that the exponential gain proceeds in a manner similar to that with no chirp in an untapered undulator. In Figures 7(b)-(c) we plot the output self-amplified spontaneous emission (SASE) power and spectrum after nonlinear saturation of the gain at $z = 20L_G$. The spectral power profile is quite broad, reflecting the wide variation of electron beam energy, but note that the spectrum is correlated with time.

Certain applications may not be able to use FEL output with this significant frequency chirp, while others may benefit from it. For example, the time-resolved x-ray Laue diffraction actually requires a relatively large spectral bandwidth, so that a DWA-based FEL would provide a suitable means of extending this currently third-generation light source technique to the higher photon flux available with FELs.

Alternatively, one might consider subsequent manipulations of the radiation to generate ultra-short soft x-ray pulses. Since time and wavelength are correlated, one can select a small temporal region of the radiation using a monochromator [13], albeit with a reduction in flux. A more ambitious plan would use a grating to temporally compress the field much like what is currently done in solid-state chirped pulse amplification schemes [14]. In this case, theoretically one might be able to compress the field to an ultra-short pulse of temporal extent $\Delta t \sim \frac{1}{\rho\omega_0}\frac{\gamma_0}{\Delta\gamma}\frac{4\pi\omega_e\rho^2}{\lambda_0}$, whose peak power is increased by that at saturation by a factor of order $\sigma_e/c\Delta t$. For the 1-nm case considered before, theoretically one could achieve pulse durations~250 as with a peak power in the TW regime. Whether such a compression scheme is feasible for high-power soft x-rays is yet to be determined. Regardless, one can still use a monochromator to reduce the pulse length, and since the energy chirp becomes even more significant at longer wavelengths, the temporal duration of the field can be potentially decreased to below that of the inverse FEL bandwidth $1/(\rho\omega_0)$ [13].

SUMMARY

It has been shown that the collinear dielectric wakefield accelerator is a viable candidate for an accelerator for a light source facility with multiple FELs. It is reasonably compact, inexpensive, and can support FELs producing up to 10^5 x-ray pulses per sec. The strawman design of the FEL facility consisting of ten FELs considered here established the feasibility of this approach, but left many unanswered questions for future R&D.

REFERENCES

[1] P. Emma et al., *Nature Photonics* **4,** 641-647 (2010).

[2] W. Ackermann et al., *Nature Photonics* **1**, 336-342 (2007).

[3] J. Corlett et al., Proc. PAC 2011, New York, p. 775 (2011); www.JACoW.org.

[4] I. Blumenfeld et al., *Nature* **445**, 741 (2007).

[5] M. C. Thompson et al., *Phys. Rev. Lett.* **100**, 214801 (2008).

[6] P. Emma, M. Cornacchia, *Phys. Rev. ST-AB* **9**, 100702 (2006).

[7] B. Jiang et al., *Phys. Rev. ST-AB* **15**, 011301 (2012).

[8] R. Filler, Fermilab report, Beam Docs 2271-v2 (2007).

[9] P. Piot et al., *Phys. Rev. ST- AB* **14**, 022801 (2011).

[10] A. Zholents and M. Zolotorev, APS Light Source Note ANL/APS/LS-327; presented at PAC 2011, New York (2011).

[11] See web page, www.hep.anl.gov/awa

[12] V. Balakin, A. Novokhatsky, V. Smirnov, Proc. of the 12th International Conference on High Energy Accelerators, Fermilab, Batavia, IL, 1983, p. 119.

[13] E.L. Saldin et al., *Phys. Rev. ST-AB* **6**, 050703 (2006).

[13] W.M. Fawley, *Nucl. Instrum. Methods Res. A* **593**, 111 (2008).