This is the accepted manuscript made available via CHORUS. The article has been published as:

Diagnostics for plasma-based electron accelerators

M. C. Downer, R. Zgadzaj, A. Debus, U. Schramm, and M. C. Kaluza
Rev. Mod. Phys. 90, 035002 — Published 8 August 2018
DOI: 10.1103/RevModPhys.90.035002
Diagnostics for plasma-based electron accelerators

M. C. Downer* and R. Zgadzaj

Department of Physics,
University of Texas at Austin,
Austin, TX 78712,
USA

A. Debus and U. Schramm

Helmholtz-Zentrum Dresden-Rossendorf,
Institute for Radiation Physics, 01328 Dresden,
Germany

M. C. Kaluza

Institute of Optics and Quantum Electronics,
Friedrich-Schiller-University, 07743 Jena,
Germany

Helmholtz Institute Jena, 07743 Jena,
Germany

(Dated: May 3, 2018)

Plasma-based accelerators that impart energy gain as high as several GeV to electrons or positrons within a few centimeters have engendered a new class of diagnostic techniques very different from those used in connection with conventional radio-frequency (RF) accelerators. The need for new diagnostics stems from the micrometer scale and transient, dynamic structure of plasma accelerators, which contrasts with the meter scale and static structure of conventional accelerators. Because of this micrometer source size, plasma-accelerated electron bunches can emerge with smaller normalized transverse emittance ($\epsilon_n < 0.1 \text{ mm} \cdot \text{mrad}$) and shorter duration ($\tau_b \approx 1 \text{ fs}$) than bunches from RF linacs. We review single-shot diagnostics that determine such small $\epsilon_n$ and $\tau_b$ non-invasively and with high resolution from wide-bandwidth spectral measurement of electromagnetic radiation the electrons emit: $\epsilon_n$ from x-rays emitted as electrons interact with transverse internal fields of the plasma accelerator or with external optical fields or undulators; $\tau_b$ from THz to optical coherent transition radiation emitted upon traversing interfaces. The duration of $\sim 1 \text{ fs}$ bunches can also be measured by sampling individual cycles of a co-propagating optical pulse or by measuring the associated magnetic field using a transverse probe pulse. Because of their luminal velocity and micrometer size, the evolving structure of plasma accelerators, the key determinant of accelerator performance, is exceptionally challenging to visualize in the laboratory. Here we review a new generation of laboratory diagnostics that yield snapshots, or even movies, of laser- and particle-beam-generated plasma accelerator structures based on their phase modulation or deflection of femtosecond electromagnetic or electron probe pulses. We discuss spatiotemporal resolution limits of these imaging techniques, along with insights into plasma-based acceleration physics that have emerged from analyzing the images, and comparing them to simulated plasma structures.

CONTENTS

I. Introduction 1

II. Properties of plasma accelerator structures and beams 3

A. General properties of plasma electron accelerators 3

1. Ponderomotive and Coulomb forces 3

2. Wake structures 3

3. Accelerating and focusing fields 5

4. Plasma density range 6

5. De-phasing, pump depletion and transformer ratio 6

6. Atomic composition of the plasma 7

B. Plasma accelerator configurations. 7

III. Diagnostics of Plasma-Accelerated Electron Bunches 12

A. Radiation from plasma-accelerated electrons 13

1. Synchrotron radiation 13

2. Undulator and Thomson backscatter radiation 13

3. Betatron radiation 15

4. Transition radiation 16

B. Bunch charge and energy measurement 20

1. Total and spatially-resolved charge measurement 20

2. MeV and GeV magnetic spectrometers 22

3. Spectroscopy of on-axis undulator radiation 23
I. INTRODUCTION

Energetic electron and positron beams from accelerators probe the fundamental structure of sub-atomic matter, irradiate cancerous tumors, and produce bright x-rays that sterilize food and medical equipment and irradiate cancerous tumors, and produce bright x-rays. In all cases, the transient plasma cavity accelerates the captured particles further with GV/cm field strength until either they dephase from it or the drive pulse (bunch) depletes.

Since Tajima and Dawson first proposed the concept of plasma-based, laser-driven electron acceleration in 1979 (Tajima and Dawson, 1979), wakefield acceleration has grown into an international research and development effort encompassing laboratory demonstrations, theoretical modeling and computer simulations of rapidly increasing sophistication and scope (Esarey et al., 2009). Within the last few years, two laboratories demonstrated single-stage laser-driven wakefield accelerators (LWFAs) that captured ambient plasma electrons and accelerated them to 2 to 4 GeV energy with few percent energy spread and up to 0.1 nC charge within an acceleration distance of only a few centimeters (Leemans et al., 2014; Wang et al., 2013). Many laboratories have converted LPAs into femtosecond x-ray sources based on betatron radiation from electrons accelerating within the plasma structure, synchrotron radiation in external undulators, or Thomson backscatter of laser light from accelerated electrons outside the LWFA (Corde et al., 2013), while the pursuit of tabletop x-ray free electron lasers based on LWFA beams is a forefront research challenge (Gruener et al., 2007; Nakajima, 2008). Dual-stage LWFAs using one (Kim et al., 2013a) or two independent (Steinke et al., 2016) laser drive pulses, a critical step toward multi-stage LWFA systems, further required for colliders at the energy frontier (Leemans and Esarey, 2009), were also recently demonstrated.

Meanwhile, single-stage plasma accelerators driven by relativistic electron or positron bunches (Chen et al., 1985) — usually called simply plasma wakefield accelerators (PWFA) — have imparted energy gain of several GeV within an acceleration distance of approxi-
mately a meter to independent, externally injected, co-
propagating electron (Litos et al., 2014) and positron
(Corde et al., 2015) witness bunches of 0.1 nC charge
while maintaining high beam quality. The possibility of
accelerating trailing electrons to twice the energy of the
drive electron bunch in a single stage has been demon-
strated in principle (Blumenfeld et al., 2007), opening
the prospect of compact, plasma-based, energy-doubling
afterburners for conventional electron accelerators (Har-
riss, 2016; Lee et al., 2002). Computer simulations have
shown that PWFA driven by bunches of relativistic pro-
tons can in principle accelerate electrons or positrons to
the energy frontier in one sub-kilometer stage (Caldwell
et al., 2009; Lotov, 2010), a possibility that the AWAKE
project at CERN is now beginning to explore in the lab-
boratory (Caldwell et al., 2016; Geschwendtner et al., 2016).

Despite rapid recent progress with LWFA and PW-
FAs, concern remains about whether plasma wakefield
techniques can be incorporated into practical accelerat-
ors useful for high-energy physics or other areas of dis-
covery science, in view of stringent requirements for nar-
row energy spread, emittance, beam stability and con-
trol, and brightness preservation (HEPAP, 2015). This
concern brings the subject of this review — diagnos-
tics for plasma-based electron accelerators — to cen-
ter stage. Diagnostics link the micron-scale structure
and femtosecond-scale dynamics of plasma wakes to the
key beam properties — bunch duration, transverse emi-
tance, charge and energy spread — that govern the per-
formance of a collider or light source. They also link ob-
servable accelerator properties to theory and computer
simulation output. However, diagnostics in widespread
use with conventional RF accelerators have, by and large,
proven insufficient for characterizing plasma-based elec-
tron accelerators.

There are two reasons for this. First, because of the mi-
crometer scale of plasma accelerator structures, plasma-
accelerated electron bunches can emerge with shorter du-
ration ($\sigma_t/c \sim 1 \text{ fs}$), and smaller transverse beam size
($0.1 \mu m \lesssim \sigma_r \lesssim 0.1 \mu m$) than bunches from meter-scale
RF accelerator structures. Thus, beams from plasma-
based accelerators can have smaller normalized trans-
verse emittance ($\epsilon_n < 0.1 \text{ mm mrad}$) than beams from con-
ventional accelerators. Here $\epsilon_n$ is the product of a
beam’s geometric emittance (roughly $\sigma_r^0\gamma_n^0$ at a beam
waist, where $\sigma_r^0$ is its angular divergence) and its Lorentz
factor $\gamma_n$ that is conserved in an ideal beam transport sys-
tem. Small $\epsilon_n$ is potentially a key advantage over conve-
tional accelerators, because it enables high luminosity in
collider interactions and wide tuning range in free elec-
tron lasers (FELs). Together with short bunch length $\sigma_z$, it is also essential for achieving high peak bright-
ness from Thomson backscatter and FEL light sources.

Section II reviews these unique properties of plasma-
accelerated electron bunches, and the laser-plasma con-
ditions that optimize them, while Section III reviews new
methods now emerging to measure them with the reso-
lation required to operate practical plasma-based accel-
erators. Second, plasma accelerator structures, in con-
trast to fixed, stationary conventional accelerator struc-
tures, are evolving and transient as they propagate at
luminal velocity, and must be re-created with high fi-
delity for each bunch. Accelerator performance depends
sensitively on details of the plasma structure and dy-
namics. For example, “bubble”-like electron density cav-
ties created via highly nonlinear interaction of drive
pulses/bunches with a plasma yield more mono-energetic
beams than sinusoidal wakes created via linear inter-
actions (Pukhov and Meyer-ter-Vehn, 2002). Moreover,
evolution of the bubble structure during propagation gov-
erns self-injection of plasma electrons (Kalmykov et al.,
2009). Accurate single-shot visualization of these plasma
structures and their dynamics in the laboratory is essen-
tial not only for operating practical plasma accelerators,
but for validating individual stages of holistic computer
simulation output, and for understanding LWFA physics
detail. Section IV reviews the many innovative di-
agnostic methods that have been, and continue to be,
invented for capturing images of these plasma structures
in the act of accelerating electrons and positrons to rela-
tivistic energy. This aspect of diagnostics has no counter-
part in conventional RF acceleration, and is likely to be
a continuing source of innovation and discovery. Sec. V
presents our conclusions, and our assessment of impor-
tant future directions in plasma acceleration diagnostics.

II. PROPERTIES OF PLASMA ACCELERATOR
STRUCTURES AND BEAMS

A. General properties of plasma electron accelerators

We refer the reader to (Esarey et al., 2009; Hooker,
2013; Leemans and Esarey, 2009; Malka, 2012; Malka
et al., 2008; Norreys, 2009) for reviews of the physics and
applications of laser-driven plasma-based electron accel-
erators, and to (Caldwell et al., 2016; Hogan et al., 2010;
Muggli, 2016; Muggli and Hogan, 2009) for reviews of
particle-beam-driven plasma wakefield acceleration. Here
we summarize basic features of LWFA and PWFA struc-
tures and beams that are needed for the subsequent dis-
cussion of diagnostics.

1. Ponderomotive and Coulomb forces

The simplest LWFA consists of a single intense laser
pulse focused into a confined gas or pre-ionized under-
dense plasma. For gas targets, the leading edge of the
pulse ionizes the gas, creating plasma. The intense por-
tion of the pulse then creates a light-speed accelerat-
ing structure by expelling plasma electrons longitudi-
nally and radially from within its envelope via “pon-
derivative" pressure, which is equivalent to the gradient \( \vec{\nabla} (\varepsilon_0 E_r^2 / 2) \) of the pulse’s cycle-averaged electromagnetic energy density. Here, \( E_r \) is the local amplitude of the laser electric field in V/m. When each electron’s quiver motion in the field \( E_r \) is non-relativistic, the ponderomotive force \( \vec{F}_p \) on each electron (mass \( m_e \)) is (Esarey et al., 2009; Krueer, 1988)

\[
\vec{F}_p = -m_e c^2 \vec{\nabla} (a^2 / 2),
\]

where \( a = \varepsilon_0 E_r / m_e c \omega \) is the local dimensionless normalized vector potential, equal to the ratio of momentum \( \varepsilon_0 E_r / \omega \) that the laser field of frequency \( \omega \) imparts to an electron in an optical cycle to \( m_e c \) as a function of position within the pulse profile. Thus \( a \approx 1 \) is a soft boundary between non-relativistic electron undulation and linear laser-plasma interaction at \( a < 1 \), and relativistic electron motion and laser-plasma interaction at \( a > 1 \). We can relate \( a \) to a local intensity \( I = E_r^2 / 2 Z_0 \) of a pulse of wavelength \( \lambda \) by

\[
I \approx 2 \left( 4 \pi n_e \right)^{0.3} a^{0.7} / \lambda^{0.7} \equiv a^{0.7} / \lambda^{0.7} Z_0 \approx 10^{18} \text{W/cm}^2.
\]

(2)

For highly relativistic \( (a \gg 1) \) pulses, the ponderomotive force can be written \( \vec{F}_p = -m_e c^2 \vec{\nabla} \gamma c \), where \( \gamma \equiv 1 + (p_e / m_e c)^2 \approx 1 + a^2 \) is the Lorentz factor associated with the electron’s quiver motion and \( p_e \) is the oscillating electron’s momentum. Ions also experience \( \vec{F}_p \), but respond much more slowly than electrons. Consequently charge separates, creating longitudinal (accelerating) and transverse (focusing) fields, or wakefields, that can trap and accelerate leptons.

In the PWFA, the Coulomb force of a bunch of density \( n_b \) replaces the ponderomotive force (1) as plasma wake driver. For relativistic bunches, the Lorentz-contracted Coulomb electric field is essentially transverse \((\vec{E} \approx E_r \hat{e}_r)\), and the bunch’s internal space-charge force \( e n_b E_r \) is reduced by a factor \( \gamma c^{-2} \) compared to non-relativistic bunches (Muggli, 2016). Consequently, transverse bunch dynamics over meter-scale path lengths is dominated by emittance and external focusing forces, rather than space-charge (Hogan et al., 2003; Muggli et al., 2008b).

The boundary between linear (Chen et al., 1985) and nonlinear (Rosenzweig et al., 1991) wake excitation occurs at \( n_b \sim n_e \) for PWFA, analogous to \( a_0 \sim 1 \) for LWFA. Unlike laser drivers, particle bunch drivers can either “blow out” or “suck in” plasma electrons, depending on whether they are negatively or positively charged. Indeed, PWFA experiments have been (and are being) performed with electron (Blumenfeld et al., 2007; Litos et al., 2014), positron (Blue et al., 2003; Corde et al., 2015), and proton (Caldwell et al., 2016; Geschwendtner et al., 2016) drive bunches.

2. Wake structures

The shape and dynamics of a wake’s electron density profile \( n_e (r, z, t) \) depend on the duration, focus and energy (and for PWFAs, charge; for LWFAs, \( \lambda \)) of the drive pulse, and on the density, composition and pre-formed structure of the plasma target. As a simple example, a 1D laser pulse of longitudinal duration \( \tau_L \) (FWHM) less than than a plasma period \( \tau_p \equiv f_p^{-1} \equiv 2 \pi / \omega_p \) and amplitude \( a_0 \ll 1 \) propagating at group velocity \( v_g \approx c \) in a uniform, underdense plasma of unperturbed electron density \( n_e \) linearly excites a sinusoidal 1D electron density wake (Tajima and Dawson, 1979)

\[
\delta n_e (r, z, t) = \delta n_{e,0} \sin k_p r \cos \omega_p t.
\]

(3a)

Here, \( \delta n_{e,0} \) is the local density perturbation of amplitude \( \delta n_{e,0} = \omega_p / v_g \) is the plasma wavenumber, \( \omega_p = (\varepsilon_0 c^2 / e_0 m_e)^{1/2} \) is the plasma frequency for collective electron density oscillations, \( v_g = \omega_p / k_p = v_g \) is the plasma wave phase velocity set by \( v_g \) and \( \xi = z - v_g t \). As \( \tau_L \) varies, \( \delta n_{e,0} \) exhibits a broad resonant peak at \( \tau_L \sim \tau_p / 2 \).

In 3D with cylindrical symmetry, a sub-relativistic \((a_0 \ll 1)\) pulse of duration \( \tau_L \ll \tau_p / 2 \), focused to Gaussian transverse profile \( a_0 \exp(-r^2 / w_0^2) \) of width \( w_0 \sim \lambda_p \), linearly excites a wake of the form

\[
\delta n_e (r, z, t) = \delta n_z (r, z, t) + \delta n_r (r, z, t) = A e^{-r^2 / w_r^2} \left( 1 + f(r) \right) \sin k_p r \cos \omega_p t.
\]

(3b)

that can be calculated analytically from cold fluid equations (Esarey et al., 1989; Gorbunov and Kisranov, 1987). Here, \( A \) depends on the pump’s peak power and its Rayleigh length \( z_R = \pi w_0^2 / \lambda \), and \( f(r) \equiv (\lambda_p^2 / \pi \lambda z_R) (1 - r^2 / w_r^2) \). The first (second) term in square brackets corresponds to the contribution \( \delta n_z \) (\( \delta n_r \)) arising from the longitudinal (radial) ponderomotive force. Such 2D linear laser wakes were the first to be observed directly in the laboratory with sub-\( \lambda_p \) resolution, using ultrashort laser probe pulses (see Sec.IV.B.2). Researchers took advantage of tight focusing to create \( \delta n_r \sim n_0 \) over \( \lesssim 0.1 \) mm path with sub-terawatt (sub-TW) lasers (Marquès et al., 1996; Siders et al., 1996b), before multi-TW lasers (Backus et al., 1998) became widely available.

With advances in laser technology, excitation of wakes with short \( (\tau_L \lesssim \tau_p / 2) \), mildly relativistic \((a_0 \sim 1)\) pulses over multi-mm paths became possible. In this regime, numerical calculations are required to describe wake excitation in 3D, although analytic 1D solutions that exhibit the main physical effects are possible with simplifying assumptions such as a non-evolving driver and a quasi-static plasma structure (Sprangle et al., 1990a,b).
The wake develops non-sinusoidal features attributable to the relativistic mass increase of the strongly-driven plasma electrons. In 1D, the main new features are steepened wavefronts and lengthened wake period (Berezhiani and Murusidze, 1990; Bulanov et al., 1989; Sprangle et al., 1990a,b). In 3D, these effects depend on \( r \). A mildly relativistic drive laser pulse that is peaked on axis creates a period-lengthened, steepened wake on axis, but a linear wake off axis. As a result, wave fronts curve, with curvature increasing with distance behind the driver (Andreev et al., 1997b; Bulanov et al., 1995; Decker and Mori, 1994; Sprangle et al., 1992). Reduction of \( n_e \) on axis (e.g. by ponderomotive channeling) can further lengthen the on-axis period, and accentuate the curvature. Fig. 1a shows a computer simulation of a wake excited in this regime. Ultimately transverse wave-breaking can occur (Bulanov et al., 1997b). Secs. IV.C describes single-shot diagnostic experiments in which these features of mildly relativistic plasma wakes were first observed in the laboratory (Dong et al., 2010a; Matlis et al., 2006).

![Color online. Particle-in-cell simulation of evolving laser-driven plasma bubble. Grey scale: local electron density \( n_e \); color scale: local strength of drive laser profile centered at \( y = z - v_g t = 0 \). Gas jet: He with 1% N, with linear entrance ramp (0 < \( z < 0.55 \) mm), plateau (0.55 < \( z < 2.55 \) mm) at \( n_e = 4.4 \times 10^{18} \) cm\(^{-3} \) after ionization, exit ramp (2.55 < \( z < 3.1 \) mm). Laser pulse: \( \lambda = 800 \) nm, \( \tau_L = 30 \) fs, beam waist \( w_0 = 19 \mu m \) at \( z = 4.1 \) mm. (a) \( z = 1.72 \) mm, \( a_0 \approx 1 \), mildly nonlinear wake; (b) \( z = 2.15 \) mm, \( a_0 \approx 3 \), bubble has formed, ionized inner-shell electrons from N-dopant injected and trapped; (c) \( z = 2.37 \) mm, \( a_0 \approx 6 \), bubble evolves, trapped electrons advance; (d) \( z = 2.8 \) mm, \( a_0 \approx 6 \), bubble lengthens in down ramp. Courtesy R. Pausch.](image)

As the drive pulse intensifies further to \( a_0 \gtrsim 3 \), it can evacuate electrons completely from its immediate wake (Fig. 1b). This strongly nonlinear LWFA regime, first discovered in computer simulations (Pukhov and Meyer-ter-Vehn, 2002), leaves behind a nearly spherical bare ion cavity bounded by a thin, dense electron “wall”. Fig. 1b-d show simulations of the self-consistent evolution of the drive laser pulse (red) and trailing ion cavity excited in this strongly nonlinear regime, which researchers began calling the “bubble” regime (Kostyukov et al., 2004). Using 3D simulations, (Lu et al., 2007) comprehensively documented properties of bubble-regime LWFA over a wide range of laser-plasma parameters. Because of their special importance for plasma accelerator science in general, and diagnostic development in particular, we review properties of bubble-regime LWFA separately in Sec. II.C. Sec. III reviews a new generation of single-shot electron bunch diagnostics developed primarily to meet the challenge of measuring the unusually narrow (\( \sigma_r \ll 1 \mu m \)), short (\( \sigma_z \sim 1 \mu m \)),”quasi-monoenergetic” electron bunches that closely resemble those described above for LWFA. (Chen et al., 1985) developed the linear theory of the standard PWFA. Predictions of steepened wavefronts and period lengthening in 1D nonlinear theory of PWFA (Rosenzweig, 1987), and of wavefront curvature in 2D (Rosenzweig et al., 1991), preceded, but closely parallel, corresponding predictions for LWFA. Sec. IV.B.1 describes diagnostic experiments in which linear and mildly nonlinear PWFAs were first observed in the laboratory. (Rosenzweig et al., 1991) also developed the theory of the strongly nonlinear \( (n_b = n_e) \) PWFA regime, usually called the “bubble” regime in the context of PWFA, which produces wake structures similar to those shown in Fig. 1b-d. The uniform ion column that an electron driver creates in this regime guides drive and trailing accelerating bunches over many initial beam “beta-function” lengths, analogous to Rayleigh lengths of a laser driver. However, electron (Blumenfeld et al., 2007; Litos et al., 2014) and positron (Corde et al., 2015) drivers produce different wake structures, a distinction that does not arise with LWFA. Direct observation of these differences, using probing techniques developed for LWFA (see Secs. IV.C, IV.D) is an important opportunity for future diagnostics research.

3. Accelerating and focusing fields

The wake’s electron density change \( \delta n_e(r, z, t) \) is the source of its enormous internal accelerating \( [\hat{E}_z E_z(r, z, t)] \) and focusing \( [\hat{E}_r E_r(r, z, t)] \) fields. In 1D, Poisson’s equation \( \partial E_z / \partial z = -\epsilon \delta n_e(z, t) / e_0 \) yields 1D solutions for the accelerating field \( E_z(z, t) = E_0 \cos k_p z \) corresponding to
Eq. (3a). Since the maximum density perturbation is of order $\delta n_e^{\text{(max)}} \approx \bar{n}_e$, the maximum accelerating field is
\begin{equation}
E_0 = m_e c p / e 
\approx 0.96\sqrt{\bar{n}_e}\left(\text{cm}^{-3}\right) \text{ V/cm.}
\end{equation}

Eq. (4a) is the so-called cold nonrelativistic wave breaking field (Dawson, 1959). In the form Eq. (4b), it provides a simple estimate of the maximum accelerating field achievable in a plasma of density $\bar{n}_e$. Drive fields of amplitude $a_0 \gtrsim 1$ are required to reach $E_0$. For $a_0 \gtrsim 2$, the wake becomes strongly nonlinear (see Fig. 1b-d). 3D computer simulations show that the maximum accelerating field is then (Lu et al., 2007):
\begin{equation}
E_z^{\text{(max)}} \approx E_0 \sqrt{a_0}.
\end{equation}

Eqs. (4) can be compared with accelerating fields in conventional RF accelerators, which are currently limited to $\sim 10^6$ V/cm. The plasma accelerator in Fig. 1c ($\bar{n}_e = 4.4 \times 10^8 \text{ cm}^{-3}$, $a_0 \approx 6$), in contrast, has maximum accelerating field $E_0 \sqrt{a_0} \approx 5 \times 10^6$ V/cm.

In 3D, both accelerating and focusing fields pervade the plasma wake. Fig. 2 shows (a) accelerating and (b) focusing fields inside the wake in Fig. 1c. In the back half of the positively-charged bubble, $E_z$ accelerates electrons forward from their internal injection point near the back wall (see Fig. 1b) toward the bubble’s center. Simultaneously $E_r$ focuses these electrons toward the propagation axis, maintaining low emittance. The strong $E_r$ of bubbles in longitudinally shaped plasmas can potentially provide emittance-preserving beam transport between LWFA stages, or between an LWFA and conventional accelerator or FEL undulator (Xu et al., 2016).

Field structures in blowout-regime electron-beam-driven PWFA’s resemble those in Fig. 2, and those of conventional RF linacs (Rosenzweig et al., 1998). With PWFA’s, unlike LWFA’s, acceleration of positrons, an essential element for a collider, has been investigated extensively through theory (Lee et al., 2001), simulation and experiments (Blue et al., 2003; Corde et al., 2015; Hogan et al., 2003) because of the availability of a relativistic positron bunch injector at SLAC’s Facility for Advanced Accelerator Experimental Tests (FACET) (Hogan et al., 2010) and its predecessors. $E_z$ accelerates positrons forward from an external injection point just in front of the bubble’s center, towards the front of the bubble. Unfortunately, $E_r$ defocuses these positrons away from the propagation axis, causing emittance growth. A possible solution for preserving the emittance is to excite the PWFA in a hollow channel surrounded by an annular plasma (Chiou and Katsouleas, 1998; Gessner et al., 2016; Kimura et al., 2011; Schroeder et al., 1999b). Alternatively, (Corde et al., 2015) demonstrated that a positron driver can draw a quasi-static reservoir of plasma electrons to the bubble’s axis that compensates the defocusing fields. Similar issues will arise when LWFA’s accelerate positrons. They highlight the importance of developing laboratory diagnostics of internal fields of plasma lepton accelerators.

Whereas ultrashort optical pulses mostly probe electron-density structure of plasma wakes, electron bunches directly probe their internal fields. The above-cited experiments of (Rosenzweig et al., 1988, 1989), detailed in Sec. IV.B.1, used sub-\(\lambda_p\) electron “witness” bunches to probe $E_z(r, z, t)$ and $E_r(r, z, t)$, rather than $\delta n_e(r, z, t)$ of linear and non-linear PWFA’s. Sec. IV.C.4 details more recent experiments in which few-fs electron witness bunches — derived, ironically, from a diagnostic bubble-regime LWFA — probed the internal electric fields of a subject bubble-regime LWFA (Zhang et al., 2017). An important exception to the role of optical probes is ultrafast Faraday rotation probes of a plasma wake’s internal magnetic fields, created by the current of the accelerating electron bunch and/or by the displacement current of the wake’s dynamic electric fields (Buck et al., 2011; Kaluza et al., 2010). Secs. III.D.1.c and IV.C.3.a detail these diagnostic experiments.

4. Plasma density range

A reasonable criterion for plasma accelerators to provide significant advantage over conventional RF accelerators is that $E_0$ be at least 100× their breakdown field $\sim 10^6$ V/cm. Eq. (4b) then dictates a lower limit $\bar{n}_e > 10^{16} \text{ cm}^{-3}$ on plasma density. In fact, this is about the lowest density at which laser-driven electron acceleration has been reported in the laboratory (Amiranoff et al., 1998; Clayton et al., 1993). Self-injection of plasma electrons into a LWFA becomes inefficient as $\bar{n}_e$ decreases (Froula et al., 2009), and guiding of the drive pulse, either via its own relativistic Kerr effect (Sprangle et al., 1987) or via self- (Sprangle et al., 1992) or pre-formed (Durfee III et al., 1995) plasma waveguides, becomes in-
creasingly difficult at low $\bar{n}_e$. Thus the above-cited experiments, and others up to $\bar{n}_e \approx 10^{17}$ cm$^{-3}$ (Amiranoff et al., 1995; Kitagawa et al., 2004), required either external injection or a capillary waveguide. For these reasons most LWFA experiments have used $\bar{n}_e > 3 \times 10^{17}$ cm$^{-3}$.

PWFAs have used density as low as $\bar{n}_e \approx 10^{13}$ cm$^{-3}$ (Rosenzweig et al., 1988, 1989), but then $E_0$ approaches that of conventional accelerators. Recent PWFAs, however, have used $\bar{n}_e$ from $5 \times 10^{16}$ cm$^{-3}$ (Litots et al., 2014) to $8 \times 10^{16}$ cm$^{-3}$ (Corde et al., 2015).

An upper limit on $\bar{n}_e$ for LWFAs comes from the requirement that the plasma be underdense — i.e. $\omega_p < \omega$. When it is not, a laser pulse penetrates only a skin depth $(c/\omega_p \lesssim 10^{-2}$ cm) into the plasma, and reflects. When it is, the laser pulse can propagate over distances 0.1 cm $< z < 10$ cm required for its wake to accelerate electrons to 0.1 GeV $< E_e < 10$ GeV with gradient $E_0 \approx 10^9$ V/cm. Stated equivalently, $\bar{n}_e$ must be less than the critical plasma density

$$n_{cr} = \frac{c_0 m_e e^2}{e^2} = \frac{1.1 \times 10^{21}}{[\lambda(\mu m)]^2} \text{cm}^{-3}.$$

Thus LWFAs driven by near-infrared $(0.8 < \lambda < 1.\mu m)$ solid-state lasers, the dominant technology for reaching $a_0 \geq 1$ since the 1990s, are limited to $\bar{n}_e < 10^{21}$ cm$^{-3}$. CO$_2$ lasers ($\lambda \approx 10 \mu m$), which drove the earliest LWFAs at $a_0 \ll 1$ in the early 1990s (Clayton et al., 1992), and which, with recent advances in chirped pulse amplification (Polyanskiy et al., 2015), promise to drive future LWFAs at $a_0 \geq 1$ (Pogorelsky and Ben-Zvi, 2014), are limited to driving LWFAs at $\bar{n}_e < 10^{19}$ cm$^{-3}$.

For particle-bunch drivers there is no counterpart to the critical frequency (5). Nevertheless, for both LWFAs and PWFAs, the pulse (bunch) duration $\tau_L$ ($\tau_b$) available at power (bunch density) needed to drive a high-amplitude wake, together with the resonant criterion $\tau_{L,b} \lesssim \omega_p^{-1}$, set a practical upper limit on $\bar{n}_e$. Multi-TW laser pulses at $\lambda \approx 0.8\mu m$ are currently limited to $\tau_L \gtrsim 10$ fs, limiting resonant LWFA to $\bar{n}_e \lesssim 3 \times 10^{19}$ cm$^{-3}$ $\approx 0.2n_{cr}$. Chicane compressors can provide n bunches with $\tau_b \gtrsim 100$ fs, limiting resonant PWFA to $\bar{n}_e \lesssim 3 \times 10^{17}$ cm$^{-3}$. Future multi-TW CO$_2$ laser pulses ($\lambda \approx 10 \mu m$) are likely to be limited to $\tau_b \gtrsim 500$ fs (Pogorelsky and Ben-Zvi, 2014), limiting resonant LWFA at this $\lambda$ to $\bar{n}_e \lesssim 10^{16}$ cm$^{-3} \approx 0.03n_{cr}$. So-called “self-modulated” LWFAs and PWFAs, discussed in Sec.II.B, use non-resonant excitation ($\tau_{L,b} \gg \omega_p^{-1}$). Then $\bar{n}_e$ can be higher than the limits stated above.

Diagonstics should be versatile enough to probe wake structure over a wide $\bar{n}_e$ range. While many optical probe methods discussed in Secs.IV.C and IV.D were developed to visualize wakes in $\bar{n}_e \sim 10^{19}$ cm$^{-3}$ plasma, diagnostic electron bunches have probed wakes at $\bar{n}_e \sim 10^{17}$ cm$^{-3}$ (Sec.IV.C.4) or lower (Sec.IV.B.1) density. Sec.IV.E discusses density scaling of wake diagnostics.

5. De-phasing, pump depletion and transformer ratio

Within the practical LWFA density range $3 \times 10^{17}$ cm$^{-3}$ $\lesssim \bar{n}_e \lesssim 3 \times 10^{19}$ cm$^{-3}$ for a $\lambda \approx 0.8 \mu m$ drive pulse, useful acceleration length and achievable singe-stage energy gain $\Delta W_e$ vary widely. There are two main reasons for this. First, the group velocity $v_g = c(1 - \bar{n}_e/\bar{n}_{cr})^{1/2}$ of the drive pulse (and thus the phase velocity of the plasma accelerating structure) decreases with increasing $\bar{n}_e$, dropping by a fraction $(c - v_g)/c \approx 0.015$ below $c$ at the upper limit $\bar{n}_e/\bar{n}_{cr} \approx 0.03$ of the LWFA density range. Consequently, a relativistic electron propagating at $\sim c$ in the laboratory frame drifts at velocity $c - v_g$ through the wake’s accelerating cavity, which has length of order $\sim \lambda_p/2$. Dephasing between electron and driver limits acceleration time to $\sim \lambda/2(c - v_g)$ and acceleration distance in the lab frame to dephasing length $L_d \approx \lambda_p/2(c - v_g)$, or in common laboratory units

$$L_d(\mu m) \approx \lambda_p^3/\lambda^2 \approx \frac{3.7}{\bar{n}_e(10^{18} \text{cm}^{-3})^{3/2}[\lambda(\mu m)]^2}.$$

Thus for $\lambda = 1 \mu m$ and uniform $\bar{n}_e = 0.03n_{cr} = 3 \times 10^{19}$ cm$^{-3}$, $L_d \approx 0.02$ cm, limiting electron energy gain to $\Delta W_e = e_0E_0L_d \approx 100$ MeV. $\Delta W_e \propto (\bar{n}_e)^{-1}$ at other densities. Introduction of a gradual density up-ramp $(\bar{n}_e/dz > 0)$ along the drive pulse propagation path can, in principle, compensate dephasing, since the accelerating cavity then shrinks in proportion to $\sim \lambda_p/\epsilon$ as it propagates, keeping the accelerating electron bunch at its rear (Bulanov et al., 1997b; Katsouleas, 1986; Sprangle et al., 2001). However, laser-driven tapered plasmas in the laboratory (Abuaoum et al., 2012; Kaganovich et al., 1999; Kim et al., 2013b; Rittershofer et al., 2010) have not yet accelerated electrons significantly beyond the limit given by Eq. (6).

In contrast, for PWFA, a drive bunch with Lorentz factor $\gamma_b$ propagates at velocity $v_b = (1 - \gamma_b^{-2})^{1/2}$ independent of $\bar{n}_e$. Thus a highly relativistic driver and its wake propagate vanishingly close to $c$ — e.g. $(c - v_b)/c \approx 10^{-11}$ for 20 GeV ($\gamma_b = 4 \times 10^8$) drive bunches used in (Corde et al., 2015; Litots et al., 2014) — effectively eliminating dephasing, an advantage of PWFA over LWFA.

Secondly, dephletion of the drive laser pulse energy increases with increasing $\bar{n}_e$. Various laser-plasma instabilities can contribute to depletion, depending on the intensity and duration of the drive pulse. In the strongly nonlinear ($a_0 > 1$), short pulse ($\omega_p\tau_L < 1$) regime of greatest interest for electron acceleration, erosion of the leading edge of the drive pulse due to diffraction, scattering and photon deceleration by the density spike at the leading edge of the plasma wake dominates pump depletion (Decker et al., 1996). 3D simulations in this regime show that pump depletion limits effective acceleration length in the lab frame to $L_{pd} \approx (n_{cr}/\bar{n}_e)c\tau_L$ (Lu
et al., 2007), or in common laboratory units

\[
L_{pd}(\text{cm}) \approx \frac{0.03 \tau (\text{fs})}{n_e(10^{18} \text{cm}^{-3})[\lambda(\mu\text{m})]^2}.
\] (7)

Thus a \(\tau_L \approx 30 \text{ fs}\), \(\lambda = 1 \mu\text{m}\) pulse driving a strongly nonlinear wake in plasma of density \(\bar{n}_e = 3 \times 10^{19} \text{ cm}^{-3}\) depletes within \(L_{pd} \approx 0.03 \text{ cm}\), similar to the dephasing limit for a uniform plasma. Pump depletion fundamentally limits LWFA. No methods exist to compensate it. Pump depletion and dephasing together limit practical laser-plasma electron accelerators to \(\bar{n}_e < 0.03n_{cr}\).

PWFA is subject to beam-plasma instabilities (Deng et al., 2006; Dodd et al., 2002; Huang et al., 2007). Moreover, maximum energy gain per stage is limited to the product of the drive bunch kinetic energy and a transformer ratio (Ruth et al., 1984), which is equivalent to the pump depletion limit of LWFA. Thus, for example, a PWFA driven by an electron bunch of energy \(E_{drive}\) is limited to accelerating trailing electrons to \(\sim 2E_{drive}\). Special shaping of the drive bunch can extend this limit (Chen et al., 1986).

6. Atomic composition of the plasma

In addition to density, atomic composition of the plasma must be chosen carefully. Nearly all LWFA experiments use targets comprised mostly of H\(_2\) or He. This is because field strengths in the range \(0.01 < a_0 < 0.1\), realized in the leading edge of relativistic \((a_0 \gtrsim 1)\) drive pulses or in separate pre-ionizing pulses, can field-ionize these low-Z atoms completely over a wide footprint. This avoids complicating the wake-forming laser-plasma interaction at \(a_0 \sim 1\) with delayed ionization of inner-shell electrons, which occurs at \(a_0 \gtrsim 1\) in higher-Z atoms. On the other hand, few-percent admixtures of high-Z gases (e.g. N\(_2\), Ar) into H\(_2\) or He targets can facilitate injection of electrons into a laser-driven wake (McGuffey et al., 2010; Pak et al., 2010), as discussed in Sec. II.C.1.

For PWFA, lithium is a common choice of target gas (Muggli et al., 1999) because the drive bunch, or a synchronized laser pulse, can ionize it over a multi-meter path. Self-ionization can degrade accelerator performance (Deng et al., 2003; O’Connell et al., 2006) so pre-ionization is preferred (Green et al., 2014). As with LWFAs, high-Z admixtures can stimulate injection from within the plasma, known as “Trojan horse” injection (Hidding et al., 2012) in the context of PWFA.

B. Plasma accelerator configurations.

LWFAs (PWFAs) can be excited with one, two, or multiple drive pulses (bunches). (Esarey et al., 2009) reviews the various configurations for LWFAs and (Caldwell et al., 2016; Muggli, 2016; Muggli and Hogan, 2009) for PWFAs. Here we summarize configurations important for the ensuing discussion of diagnostics.

1. Standard LWFA and PWFA.

Sec. II.A described general properties of simple “standard” LWFAs and PWFAs, driven, respectively, by one laser pulse (see Figs. 1, 2) or one particle bunch of duration \(\tau \lesssim \omega_p^{-1}\). Wake excitation in linear \((a_0 \ll 1, n_b \ll \bar{n}_e)\), weakly nonlinear \((a_0 \sim 1, n_b \sim \bar{n}_e)\) or strongly nonlinear \((a_0 \gg 1, n_b \gg \bar{n}_e)\) regimes results in very different lepton bunch properties.

Most standard LWFA experiments in linear and weakly nonlinear regimes produced no self-injected electrons [an exception was (Kitagawa et al., 2004)]. Some experiments in this regime accelerated electrons injected from a linac (Amiranoff et al., 1998; Bernard et al., 1999; Dewa et al., 1998).

Standard strongly nonlinear LWFAs can capture, trap and accelerate electrons from surrounding plasma (see Fig. 1b) through various mechanisms. They can produce bunches of a few pC to few hundred pC charge, few-% energy spread, and kA peak current without external injection. Because of their special importance in motivating beam diagnostic development reviewed in Sec. III, we discuss them separately in Sec. II.C.

Standard PWFAs producing high-quality bunches have so far required externally-injected witness bunches whether in linear (Rosenzweig et al., 1988), weakly nonlinear (Rosenzweig et al., 1989), or blowout (Corde et al., 2015; Litos et al., 2014) regimes, although internal injection is an active field of research (Hidding et al., 2012; Wittig et al., 2015). For example, (Litos et al., 2014) matched the injected electron bunch sufficiently well that 74 pC of injected charge extracted wake energy with up to 30% efficiency while gaining \(\Delta E_e > 1\) GeV and maintaining energy spread as low as 0.7%. Emissitance, duration, energy spread of the accelerated bunches were determined by the external conventional injector rather than the plasma physics of the bubble, and were measurable by standard beam diagnostics. Standard PWFAs, and linear and mildly nonlinear standard LWFAs have provided the context for many innovations in wake structure diagnostics (see Sec. IV).

2. Self-modulated LWFA and PWFA.

Starting in 1995, LWFA experiments using “long” \((\tau_L \sim 1\text{ps} \gg \omega_p^{-1})\) energetic \((\sim 1\) J) laser pulses to drive \(\bar{n}_e \sim 10^{19}\) \text{cm}^{-3}\) plasma yielded copious, self-injected, tens-of-MeV electrons with thermal energy distribution (Coverdale et al., 1995; Modena et al., 1995; Nakajima et al., 1995). Strong wake generation and energetic electron production occurred when the peak power \(P\) of the
drive pulse exceeded the critical power (Andreev, 1992; Antonsen and Mora, 1992; Sprangle et al., 1992)
\[ P_c[GW] = 17 \left( n_{e\tau}/\bar{n}_e \right), \]
for relativistic self-focusing (RSF). RSF, favored at high \( \bar{n}_e \), enabled the drive pulse to focus to, and self-guide at (Wagner et al., 1997), higher \( a_0 \) inside the plasma than it reached at the plasma entrance. This enabled it to drive self-modulation and forward Raman instabilities efficiently over \( \sim \) nm paths (Andreev, 1992; Antonsen and Mora, 1992; Sprangle et al., 1992). These instabilities broke up the incident pulse into a train of subpulses of length \( \epsilon \tau \lesssim \lambda_p \) spaced by \( \lambda_p \). Simultaneously a wake grew, and Stokes and anti-Stokes sidebands at \( \pm n\omega_p \) \( (n = 1, 2, 3, \ldots) \) appeared on the transmitted drive pulse spectrum, sometimes out to multiple orders, signifying a high-amplitude wake \( (E_c \rightarrow E_0 \sim GV/cm) \). An extensive literature, summarized by (Esarey et al., 2009), developed around such “self-modulated” (SM) wakes. Sec. IV.A describes time-resolved light-scattering experiments that diagnosed SM-LWFAs under conditions of high (Gordon et al., 1998; Le Blanc et al., 1996) and moderate (Ting et al., 1996) accelerated charge.

SM-LWFA experiments yielded energetic electron beams more simply than standard LWFA experiments, requiring no external injector or waveguide, and generated much higher charge \( [\text{e.g. } Q = 0.5 \text{nC at } E_c > 1 \text{ MeV (Wagner et al., 1997)}] \). As a result, SM-LWFA dominated LWFA science in the decade following 1995. Breaking of the high-amplitude SM-wave injected plasma electrons indiscriminately throughout the wake, yielding wide energy and angular \( [\text{e.g. } \sim 8^\circ (Wagner et al., 1997)] \) spread. High \( \bar{n}_e \), restricted \( L_d \) to \( \sim 0.1 \text{ cm (see Eq. (6))} \), and thus energy gain to \( eE_La_d \lesssim 100 \text{ MeV} \). These electron bunch properties posed no special challenges for, and stimulated no significant advances in, beam diagnostics.

The self-modulation beam-plasma instability (Bret et al., 2010) has emerged as a key first step in re-shaping \( \tau_0 \sim 300 \text{ ps, TeV proton bunches from the CERN Super Proton Synchrotron (SPS) or Large Hadron Collider (LHC) into } \lambda_p-\text{spaced multi-bunch trains that can excite high-gradient } (E_0 \gtrsim 10^8 \text{ V/cm, } \bar{n}_e \gtrsim 10^{16} \text{ cm}^{-3}) \text{ plasma wakes efficiently (Caldwell and Lotov, 2011). Compressing CERN proton bunches to single } \tau_0 \sim 300 \text{ fs bunches needed to drive high-gradient wakes resonantly would be prohibitively expensive using conventional techniques. Effective seeding of the instability is needed to create a stable bunch train, and to avoid parasitic instabilities. Initial experiments in the AWAKE project (Caldwell et al., 2016; Geschwendtner et al., 2016) are using an ionization front created by a short laser pulse that co-propagates in the front part of the proton drive bunch. Single-shot wake diagnostics (Sec. IV.C.D) scaled to the appropriate \( \bar{n}_e \) (Sec. IV.E) can potentially play a key role in evaluating these seeding strategies.}

3. Multi-Pulse LWFA and PWFA.

LWFAs can be driven by two or more laser pulses. In the plasma heat-wave accelerator (PBWA), two long pulses of frequencies \( \omega_1 \) and \( \omega_2 \) resonantly excite a plasma wake when \( \Delta \omega \equiv \omega_1 - \omega_2 \approx \omega_p \). The first laser-driven plasma electron accelerators (Amiranoff et al., 1995; Clayton et al., 1993; Kitagawa et al., 1992) utilized PBWA, and provided the context for extensive plasma wave diagnostic development based on collective Thomson scattering (Clayton et al., 1985). Sec. IV.A discusses later applications of these techniques to SM-LWFA. See (Clayton, 2009; Esarey et al., 2009) for reviews of PBWA.

Related to PBWA and SM-LWFA is excitation of LWFA with optimized trains of short \( (\tau < \omega_p^{-1}) \) pulses, \textit{i.e.} multiple-pulse (MP) LWFA (Hooker et al., 2014; Umstadter et al., 1994, 1995). (Esarey et al., 2009) and (Hooker et al., 2014) cite earlier theoretical work. In a simple MP-LWFA, \( m \) identical pulses of energy \( E \), field strength \( a_0 \lesssim 1 \) separated by \( \lambda_p \), each adds coherently to the wake, ultimately generating a wake equivalent to that generated by a single identically-shaped pulse of energy \( mE \). Nevertheless, the weaker pulses can potentially be generated by lasers capable of higher wall-plug efficiency and repetition rate than lasers that generate single joule-class pulses. The flexibility to tailor inter-pulse spacing, or shape and amplitude of individual pulses, offers additional potential advantages. In the nonlinear regime \((\omega_0 \gtrsim 1)\), an optimized pulse train can excite a stronger wake than the equivalent-energy single pulse (Umstadter et al., 1994, 1995). In addition, an appropriately timed trailing pulse can remove the wake behind the primary accelerating bucket, recover its energy by blue-shifting, and avoid unnecessary plasma heating (Hooker et al., 2014). Recently (Cowley et al., 2017) reported the first MP-LWFA experiments, using diagnostic techniques described in Sec. IV.C.

PWFA can benefit similarly from excitation via optimized particle-bunch trains. Sec. IV.B.1 reviews diagnostic experiments on the multi-bunch PWFA (Kallos et al., 2008; Muggli et al., 2008a).

C. Electron beams from strongly nonlinear LWFA

Three reports of “bubble” regime LWFA that produced relativistic electron bunches with \( \Delta E/e \) ranging from \( \sim 0.02 \) (Geddes et al., 2004; Mangles et al., 2004) to 0.24 (Faure et al., 2004) transformed laser-plasma accelerator science in 2004. Reported charge within the “quasi-monoenergetic” peak ranged from 22 pC (Mangles et al., 2004) to > 100 pC (Faure et al., 2004; Geddes et al., 2004). The transformation was so complete that today most LWFA operate in the “bubble” regime.

The peaked electron energy distributions highlighted in these reports originated from a unique process that, al-
TABLE I Properties of electron bunches from strongly nonlinear LWFA, determined by diagnostic methods reviewed in Sec. III. Best reported values (boldface type) are not achieved simultaneously. \( Q \) = charge within \( \Delta E_e/E_e \); \( E_e \) = electron energy gain; \( \Delta E_e/E_e \) = fractional energy spread; \( \varepsilon_n \) = normalized transverse emittance; \( \tau_n \) = bunch duration.

<table>
<thead>
<tr>
<th>Bunch property</th>
<th>Typical range</th>
<th>See Sections...</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q ) (nC)</td>
<td>( 0.01 - 0.5^a )</td>
<td>H.C.1, H.B.1</td>
</tr>
<tr>
<td>( E_e ) (GeV)</td>
<td>( 0.01 - 4^b )</td>
<td>H.II.B.2</td>
</tr>
<tr>
<td>( \Delta E_e/E_e ) (FWHM)</td>
<td>( 0.01^e )</td>
<td>H.II.C.1, H.II.B.2-4</td>
</tr>
<tr>
<td>( \varepsilon_n ) (mm mrad)</td>
<td>( \sim 0.1^d )</td>
<td>H.C.2, H.I.C.3</td>
</tr>
<tr>
<td>( \tau_n ) (fs)</td>
<td>( \sim 1.6^e )</td>
<td>H.C.3, II.C.6</td>
</tr>
</tbody>
</table>

\( ^a \) \( E_e \approx 0.3 \) GeV, \( \Delta E_e/E_e \approx 0.15 \) (Couperus et al., 2017)
\( ^b \) \( Q \approx 0.006 \) nC, \( \Delta E_e/E_e \approx 0.06 \) (Lee et al., 2014)
\( ^c \) \( Q \approx 0.01 \) nC, \( E_e \approx 0.2 \) GeV (Rechatin et al., 2009); \( Q \sim 0.02 \) nC, \( E_e \approx 0.06 \) GeV (Gallacher et al., 2009); \( 0.01 < Q < 0.08 \) nC, \( 0.2 < E_e < 0.6 \) GeV (Wang et al., 2016)
\( ^d \) \( Q \sim 0.001 \) nC, \( E_e \approx 0.4 \) GeV (Plateau et al., 2012)
\( ^e \) \( Q \approx 0.015 \) nC, \( E_e \approx 0.085 \) GeV (Lundh et al., 2011)

though highly nonlinear, injected electrons precisely into a small fraction of the bubble’s \( \sim \lambda_0^2 \) volume. Simulations (Pukhov and Meyer-te-Verne, 2002) showed that, as a bubble formed, an electron density spike built up at its rear (see Fig. 1b) and eventually broke, injecting electrons. Since these electrons were localized in space and time (see black dot at back of bubble in Fig. 1b), they experienced nearly the same field, and thus accelerated with small energy spread.

This injection mechanism had two corollaries beyond the question of energy spread. First, the small injection volume could lead to unusually small normalized transverse emittance \( \varepsilon_n \), as discussed in Sec. II.C.2. Second, the small injection volume could also lead to unusually short bunch duration \( \tau_n \), as discussed in Sec. II.C.3. In 2004, researchers could only speculate about the values of \( \varepsilon_n \) and \( \tau_n \). Simulations provided only rough guidance, and methods for measuring them did not exist. Sec. III reviews new diagnostic methods that emerged in the past decade specifically to address the challenge that bubble-regime LWFA posed in 2004. Table II summarizes reported bunch properties that these methods determined, with forward references to appropriate subsections of Sec. III. The properties of beams accelerated in PWFA, on the other hand, are, to a large extent, governed by the conventional accelerator that injected them. Thus PWFA beam diagnostics are closer to those used widely at conventional accelerators. We refer the reader to (Green et al., 2017; Li and Hogan, 2011) for an overview of beam diagnostics at the FACET project.

1. Charge and energy spread

The 2004 results were based on measurements with standard magnet spectrometers and charged particle sensors, instruments that had not previously been used to measure few-MeV-wide spectral peaks from LWFA, nor to evaluate charge within these peaks. Unlike conventional accelerators, LWFA produced copious poorly-characterized low-energy electrons and background radiation. Moreover, their beams fluctuated in pointing from shot to shot, and diverged in a few-mrad angle cone. Questions about the accuracy of absolute values of \( Q \) and \( \Delta E_e/E_e \) reported under these conditions in early bubble-regime LWFA papers emerged [e.g. (Glinec et al., 2006)]. Sec. III.B details how researchers met the diagnostic challenges that bubble-regime LWFA posed for absolute charge and energy measurements.

Efforts to improve upon the tens-to-hundreds pC charge, “quasi”-monoenergetic energy spread, and shot-to-shot fluctuations of self-injected LWFA beams by micro-controlling the injection process are a highlight of post-2004 research. FELs and colliders demand bunches with < 1% energy spread, nC charge, and high reproducibility. (Faure, 2017) has reviewed several “controlled injection” schemes applied to bubble-regime LWFA. The right-hand column of Fig. 3 depicts them schematically, while the main panels summarize results of 21 studies. For consistency, \( \Delta E_e^{(FWHM)} \) values in Fig. 3 represent raw electron spectrometer output, without corrections for instrument resolution. Thus, in some cases, a plotted value exceeds the value reported in the corresponding publication, after such corrections were applied. With this uniform criterion we can compare the relative effectiveness of different injection methods for producing high \( Q \) and/or small \( \Delta E_e/E_e \) objectively.

Experiments that relied on self-injection (gold data points in Fig. 3a,b) yielded the most widely-varying (a) \( \Delta E_e/E_e \) and (b) \( Q \) results. The former include some of the widest (50 – 100 MeV) and one of the narrowest (3 MeV) reported \( \Delta E_e^{(FWHM)} \) values. After deconvolving instrument broadening, (Gallacher et al., 2009) reported \( \Delta E_e^{(FWHM)}/E_e \approx 0.01 \) for the last result (plotted at \( E_e = 65 \) MeV in Fig. 3a), a milestone in LWFA research. The wide variation of \( \Delta E_e^{(FWHM)} \) shows the sensitivity of self-injection to different experimental conditions, and the difficulty of controlling it. Reported \( Q \) varied from a few pC to \( \sim 50 \) pC (Fig. 3b).

(Faure et al., 2006), based on theory by (Esarey et al., 1997), controlled injection with a second pulse (amplitude \( 0.1a_0 \approx a_1 \approx 0.4a_0 \) that collided with the LWFA drive pulse (\( a_0 \)) of similar wavelength \( \lambda \) (see Fig. 3, right column, second schematic). The resulting interference introduced a ponderomotive force wave of period \( \lambda \), amplitude proportional to \( 2a_0a_1/\lambda_0 \), and near-zero phase velocity. When positioned near the back of the bubble, this wave could induce injection with spatial precision \( \sim \lambda \) into a LWFA operating below its self-injection threshold (Faure et al., 2006), or control phase space volume for trapping within a self-injected bubble, thus controlling
accelerated charge and its energy spread (Rechatin et al., 2009). Adjusting delay between the pulses controlled the collision point location within the jet, and thus acceleration length. Adjustment of \( a_1 \) fine-tuned \( Q \) and \( \Delta E_e \). Experiments using colliding-pulse injection (green data points, Fig.3) have consistently yielded \( \Delta E_e^{(\text{FWHM})} \) between 10 and 25 MeV, showing improved control compared to self-injection. \( E_e \) and \( Q \) tuning ranges 60-200 MeV and 6-60 pC, respectively, have been achieved. After deconvolving instrument response, (Rechatin et al., 2009) reported \( \Delta E_e^{(\text{FWHM})}/E_e \approx 0.01 \), also a milestone. Use of two or more injection pulses may yield further improvements (Esarey et al., 1997).

Another group of experiments (blue and black data points, Fig.3) controlled injection by sculpting the plasma's longitudinal density profile \( \bar{n}_e(z) \). A density downramp \( \partial \bar{n}_e/\partial z \) < 0 along a wake's propagation direction decreases its phase velocity, encouraging injection (Bulanov et al., 1998; Suk et al., 2001). Downramp injection can occur below the self-injection threshold, enabling better control. (Geddes et al., 2008) demonstrated longitudinal momentum spread as small as 0.17 MeV/c for < 1 MeV electrons emerging from a LWFA generated in a downramp of length > \( \lambda_p \). Later (Buck et al., 2013; Schmid et al., 2010) introduced a more abrupt downramp of scale length < \( \lambda_p \) (Fig.3, right column, third schematic) by inserting a knife edge into a supersonic gas flow to create a local shock front. This simple approach has proven robust, enabling consistent output of bunches with \( \Delta E_e^{(\text{FWHM})} \) < 10 MeV as \( E_e \) tuned up to \( \sim 200 \) MeV (blue points, Fig.3a). Most results yielded \( Q \approx 30 \) pC (blue points, Fig.3b), although \( Q \approx 90 \) pC was achieved with more laser energy.

The success and simplicity of downramps prompted researchers to tailor more sophisticated density distributions to improve flexibility (Guillaume et al., 2015). Using one laser pulse, (Gonsalves et al., 2011; Wang et al., 2016) drove tandem, differently-sloped ramps, each performing a separate function: injection, rephasing, beam focusing. Recent work with this approach yielded more energetic (250 – 500 MeV) and higher \( Q \) (30 – 70 pC) bunches than single down-ramps, while maintaining \( \Delta E_e^{(\text{FWHM})} \approx 10 \) MeV (solid black triangles, Fig.3). After deconvolving instrument resolution, (Wang et al., 2016) reported \( \Delta E_e^{(\text{FWHM})}/E_e \approx 0.01 \), the smallest yet reported for LWFAs.

For controlled injection of \( Q > 100 \) pC bunches, researchers have driven bubble-regime LWFAs in H\(_2\) or He carrier gas doped with higher-Z gas (e.g. N\(_2\)). The leading edge of the drive pulse fully ionizes the carrier \( e.g. \) He I, II form at \( a_0 \approx 0.03, 0.1 \), respectively (Augst et al., 1991) and outer shell electrons of the dopant. From these electrons, the main part of the drive pulse \( (a_0 > 1) \) forms the bubble, and ionizes K-shell electrons of the dopant \textit{inside} the bubble. The bubble's internal fields can then trap and accelerate these electrons (see Fig.3, right column, fourth schematic), even below the self-injection threshold. Early “ionization-injected” LWFAs yielded 45-250 MeV electrons (open red diamonds, Fig.3a) with < 10 pC charge (open red diamonds, Fig.3b), and energy spread from \( \sim 30 \) MeV (McGuffey et al., 2010; Pak et al., 2010) to \( \sim 300 \) MeV (Clayton et al., 2010). Subsequently, (Liu et al., 2011; Pollock et al., 2011) reduced energy spread by confining dopant gas to a short initial stage, followed by a longer acceleration stage without dopant. (Mirzaie et al., 2015) achieved a similar goal by focusing 100 TW pulses into uniformly-doped gas in an unmatched geometry. Subsequent co-evolution of laser pulse and bubble self-truncated ionization injection, yielding accelerated bunches with \( Q \sim 10 \) (50) pC, \( E_e > 1 \) GeV (< 1 GeV), \( \Delta E_e^{(\text{FWHM})} < 100 \) MeV.
(Couperus et al., 2017) accelerated $Q \sim 0.5$ nC bunches to $E_e \sim 300$ MeV, $\Delta E_e \sim 45$ MeV, exploiting beam-loading to improve beam quality. Solid red diamonds in Fig. 3 show a selection of these later results.

Two broad conclusions emerge from this brief overview of controlled LWFA injection. First, controlled injection has improved consistency of LWFA output, compared to self-injection. Each method generates bunches within a narrower $\Delta E_e$ range than self-injection, although methods differ from each other: e.g. colliding-pulse and down-ramp methods yield $\Delta E_e < 10$ MeV over wide $E_e$ and $Q$ ranges, while ionization injection yields $\Delta E_e \sim 50$ MeV with higher charge. Secondly, a few milestone results (Gallacher et al., 2009; Rechatin et al., 2009; Wang et al., 2016) notwithstanding, the goal of producing $\Delta E_e/E_e \ll 0.01$ has proven difficult to realize consistently with single stage LWFAs. Although research continues, and although $\Delta E_e/E_e$ may improve in multi-stage LWFAs, emphasis in FEL design (Huang et al., 2012) and diagnostics (Lin et al., 2012) has shifted from overall $\Delta E_e$ to correlated energy spread of longitudinal slices within a bunch profile, which can be smaller than $\Delta E_e$. Secs. III.B.4 and III.D.1.a discuss how the plasma accelerator community has met the diagnostic challenge of measuring slice energy spread of LWFA electron bunches.

2. Transverse emittance.

Electron bunches that emerge from bubble-regime plasma accelerators have potentially outstanding properties for compact light sources and colliders: ultra-small transverse ($\sigma_x \sim 0.3 \mu m$) and longitudinal ($\sigma_z \sim 1 \mu m$) size, which when combined with moderate charge ($Q \sim 0.3$ nC), could yield “condensed matter” charge density ($Q/e \sigma_x \sigma_z \sim 2 \times 10^{22} \text{ cm}^{-3}$) and 100 kA peak currents. Here $r(s)$ denote transverse (longitudinal) beam coordinates. To be useful, however, LWFA bunches must be transported to the usage point without losing these outstanding properties. Conventional particle transport lines consisting of magnetic solenoids and quadrupoles were designed for beams with not only larger source size ($\sigma_x \sim 1$ mm, dictated by cathode size), but smaller divergence $\sigma_x' \ll 1$ mrad and energy spread $\Delta E_e/E_e = \Delta \gamma_e/\gamma_e < 10^{-3}$ than LWFA bunches, for which $\sigma_x' \gtrsim 1$ mrad and $\Delta \gamma_e/\gamma_e \gtrsim 10^{-2}$ are typical. The different $\sigma_x'$ and $\Delta \gamma_e/\gamma_e$ originate from contrasting conditions under which LWFA and conventional bunches are “born” (Antici et al., 2012). LWFA bunches form within focusing and accelerating fields of magnitude $E_0$ (see Eq. (4a)), which impart initial momenta $p_x(0) \sim p_z(0) \sim eE_0/\omega_p \sim m_e c \sim \text{MeV}/c$ and uncorrelated energy spread $\sim \text{MeV}$ to the bunch electrons. Subsequent acceleration of a bunch to $\sim \text{GeV}$ energy over distance $L$ yields $\sigma_x' \sim p_x(0)/p_x(L) \sim 10^{-3}$ rad at the accelerator exit, as observed. Correlated energy spread also grows, due to the non-uniform accelerating field. In contrast, electrons emerge from conventional cathodes with momenta $p_x \sim p_z \sim eV/c$. In ideal uniform extraction fields, subsequent acceleration of a collimated bunch to $\text{MeV}/c$ would yield $\sigma_x' \sim 10^{-3}$ mrad and $\Delta \gamma_e/\gamma_e \sim 10^{-6}$. Because of their larger $\sigma_x'$, $\Delta \gamma_e/\gamma_e$ and initial energy, LWFA beams are poorly matched to conventional particle transport systems (Antici et al., 2012). Innovative capture and transport methods are therefore needed to realize LWFA applications (Dornmair et al., 2015; Steinke et al., 2016; van Tilborg et al., 2015; Xu et al., 2016).

Here we focus on diagnostic challenges of measuring LWFA beam transport properties. Large $\sigma_x'$ and $\Delta \gamma_e/\gamma_e$ render many conventional diagnostics unsuitable for LWFA beams (Cianchi et al., 2013). Secs. III.C.2 and III.C.3 describe new diagnostics researchers have developed in response to this challenge. Emittance $\varepsilon$ is the area of transverse phase space (units: mm mrad) occupied by a beam of particles, each with spatial and angular coordinates $x$ and $x'$, $dx' = dx/s = p_x/p_s = p_x/(\beta\gamma_m c)$. Some authors express $\varepsilon$ as the area of an ellipse in phase space, using units $\pi \text{ mm mrad}$. Here we omit the factor $\pi$. Under conditions of the Liouville Theorem, $\varepsilon$ is motion invariant, like the wavelength of a laser beam in an optical transport system. Geometric rms emittance of a relativistic ($\beta \approx 1$) beam is defined by

$$\varepsilon^2 = \langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2$$

(9)

where $\langle \rangle$ is an average over the particle ensemble. For a beam of given $\varepsilon$, local spread in particle position and trajectory slope is described by Twiss or Courant-Snyder beam envelope parameters (Lee, 2004)

$$\beta_T = \langle x^2 \rangle / \varepsilon, \quad \gamma_T = \langle x'^2 \rangle / \varepsilon, \quad \alpha_T = -\langle xx' \rangle / \varepsilon.$$  \hspace{1cm} (10)

Beta-function $\beta_T(s)$ describes the rate at which local beam size $\sigma_x(s) = \sqrt{x^2(s)} = \sqrt{\beta_T(s)}\varepsilon$ changes along $s$ in free space, analogous to the Rayleigh range of a laser beam. $\gamma_T$ describes local divergence. $\alpha_T$ is the correlation between $x$ and $x'$; $\alpha_T = 0$ at a beam waist.

Analysis of a collimated ($\alpha_T = 0$) electron bunch — which e.g. “matched” propagation in a fully blown out plasma channel can produce — illustrates use of Eqs. (10), and estimates $\sigma_x(L)$ of a bunch emerging from a strongly nonlinear LWFA. Transverse forces within the bubble confine the matched (subscript $m$) bunch to constant radius $\sigma_x,m = \beta_T,m \varepsilon$, leading to beta-function (Krall and Joyce, 1995; Reiser, 2008)

$$\beta_{T,m} = \frac{1}{k_B} = \frac{\sqrt{2} \varepsilon}{k_p}$$

(11a)

$$\approx 7 \times 10^6 \sqrt{\gamma_e/n_c} \text{ cm}^{-1} \text{ mm},$$ \hspace{1cm} (11b)

where $k_B$ ($k_p$) denote betatron (plasma) wavenumbers. From (11b), $\beta_{T,m} < 1$ mm for LWFAs, in contrast to
An electron bunch’s out-coupling from a LWFA influences its downstream propagation. Fig. 4a shows the phase-space profile of a beam that exited the LWFA non-adiabatically through a steep density gradient, preserving its shape. In contrast, the profile in Fig. 4b results from exiting the LWFA adiabatically, with focusing strength changing slowly within a betatron wavelength. This conserves $\varepsilon_n$ and rotates the ellipse to favor smaller $\sigma'_s$ (Floettmann, 2014; Sears et al., 2010a; Weingartner et al., 2012). Consequently this beam diverges more slowly. Fig. 4c shows this beam’s ellipse after 10 cm ideal ($\sigma_{\gamma_n} = 0$) free-space propagation, illustrating rapid conversion of angular spread into correlated divergence, necessitating careful matching of downstream acceleration stages (Dornmair et al., 2015; Xu et al., 2016). Non-zero $\sigma_{\gamma_n}$ accelerates beam divergence in both cases.


Bunch duration $\tau_b$ determines peak current, a critical parameter for LWFA-based FELs and colliders. No single prediction of the duration of bunches emerging from strongly nonlinear LWFA exists. Different laser-plasma conditions can yield different $\tau_b$, $\tau_{\gamma_n}$ can change with propagation, and numerical instabilities can arise when simulating electron dynamics that vary strongly over ultra-small space and time scales (Lehe et al., 2013). Nevertheless, a few simulation studies have addressed the question of bunch duration near the exit of a nonlinear LWFA. Continuous injection into a continuously evolving bubble can lead to $\tau_b$ as long as half a plasma period (Kalmykov et al., 2012). At the opposite extreme, (Li et al., 2013a) predict that a broad ($w_0 \sim 20\lambda$), intense ($a_0 \sim 6$) laser pulse passing from a density up-ramp to a density plateau can inject an electron sheet as short as tens of attoseconds into a subsequent wake. In between these extremes, simulations of colliding-pulse (Fubiani et al., 2004; Schroeder et al., 1999a), down-ramp (Fubiani et al., 2006) and ionization-induced (Li et al., 2016) injection have predicted $\tau_b$ to be a small fraction of a plasma period, generally $1\,\text{fs} \lesssim \tau_b \lesssim 5\,\text{fs}$ for the specified conditions. Sec. III.D reviews new diagnostic methods that researchers have developed over the past decade to measure $\tau_b$ in this range.

III. DIAGNOSTICS OF PLASMA-ACCELERATED ELECTRON BUNCHES

Unconventional methods have become necessary to diagnose the few-fs duration, initially sub-$\mu$m radius, kA-peak-current, mrad-divergence, few-percent energy spread LWFA electron bunches described in Sec. II.C, both within the accelerator, and in the downstream world of applications. As an added challenge, when rapid feed-
back is needed, repetition rate is low, or shot-to-shot fluctuations are significant, single-shot diagnosis is essential.

This section reviews new beam diagnostics that have emerged to meet these challenges. Electromagnetic radiation from THz to γ-rays that electron bunches emit both within the accelerator and at downstream instruments is central to many of these diagnostics. Thus, as a prelude, Sec. III.A reviews the theory of radiation from plasma-accelerated electrons. Therein we summarize short-wavelength Thomson, undulator (Sec. III.A.2) and betatron (Sec. III.A.3) radiation briefly for completeness, but refer the reader to (Corde et al., 2013) for a more in-depth summary. On the other hand, we describe the theory of longer-wavelength transition radiation (TR) at greater length (Sec. III.A.4), because TR diffraction and spectroscopy are emerging as primary beam diagnostics for plasma accelerators. Moreover, other reviews of TR in this context are lacking. Subsequent subsections review experimental procedure and results for diagnosing bunch charge and energy distribution (Sec. III.B), transverse emittance (Sec. III.C) and bunch length (Sec. III.D).

(Clayton, 2009) reviewed the state of beam diagnostics for plasma-based accelerators approximately a decade ago. Here we emphasize developments since then. (Green et al., 2017; Li and Hogan, 2011) reviewed the comprehensive suite of e-beam diagnostics used in PWFA experiments at FACET. Thus we do not review them here. We review plasma structure diagnostics at FACET in Sec. IV.D.3.

### A. Radiation from plasma-accelerated electrons

Relativistic electrons that oscillate transversely emit forward Doppler-upshifted radiation into a relativistically contracted solid angle cone (width ~ 1/γe) in the laboratory frame (Corde et al., 2013; Esarey et al., 1993; Ride et al., 1995). Insertion devices with alternating magnetic dipoles (undulators or wigglers), counter-propagating electromagnetic radiation (Thomson-backscatter), or focusing plasma wake fields (betatron radiation) provide the fields needed to stimulate such radiation.

#### 1. Synchrotron radiation

The description of radiation from plasma-accelerated electrons begins naturally with synchrotron radiation (SR), which provides a foundation for describing all other classical radiation effects based on electron trajectories. Classical radiation is emitted when charges accelerate. The total power $P$ that one relativistic electron of normalized velocity $\beta$ [Lorentz factor $\gamma_e = (1 - \beta^2)^{-1/2}$] radiates is given by the Lorentz-invariant Larmor formula

$$ P = \frac{e^2 c}{6 \pi \varepsilon_0} \left[ \left( \frac{d(\gamma_e \beta)}{d\tau} \right)^2 - \frac{1}{e^2} \left( \frac{d\gamma_e}{d\tau} \right)^2 \right], $$

(13)

where $\tau$ denotes Lorentz-invariant proper time. For a charge moving with constant $\beta$ in a circle of radius $\rho$, as in a synchrotron, Eq. (13) becomes

$$ P = \frac{e^2 c \gamma_e^2}{6 \pi \varepsilon_0 \rho^2} \left( \frac{d(\gamma_e \beta)}{d\tau} \right)^2 = \frac{e^2 c \gamma_e^2}{6 \pi \varepsilon_0 \rho^2}. $$

(14)

Eq. (14) displays the strong $\gamma_e^2$ electron energy scaling of SR explicitly. Any centripetal force yields an effective instantaneous radius $\rho$. For example $\rho = m_e c^2 \beta \gamma_e / eB$ in a constant magnetic field of magnitude $B$.

![FIG. 5 Color online. Spectral power of synchrotron radiation (SR), given by Eq. (16) of text.](image)

The far-field spectral intensity that one electron with acceleration $\dot{\beta}$ radiates into solid angle $d\Omega$ in direction $\hat{n}$ is (Jackson, 1999)

$$ \frac{d^2 I}{d\omega d\Omega} = \frac{e^2}{16 \pi^3 \varepsilon_0 c} \left| \int_{-\infty}^{+\infty} f(\vec{n}, \vec{\beta}) e^{i \omega (\vec{n} \cdot \vec{\beta} t/c)} dt \right|^2, $$

(15)

where $f(\vec{n}, \vec{\beta}) \equiv \vec{n} \times \left[ (\vec{n} - \vec{\beta}) \times \vec{\beta} \right] / (1 - \beta \cdot \vec{n})^2$. The far-field approximation is valid for interaction volumes with dimensions much smaller than the distance to the observer. Integrating (15) over all angles and one revolution yields spectral power

$$ \frac{dW_{SR}}{d\omega} = \frac{\sqrt{3} e^2}{4 \pi \varepsilon_0 c} \gamma_e \omega \int_{\omega/\omega_c}^{+\infty} d\xi K_{5/3}(\xi), $$

(16)

plotted in Fig. 5. $K_{5/3}$ is a modified Bessel function of the second kind, and

$$ \omega_c = 3 c \gamma_e^3 / 2 \rho $$

(17)

is the critical frequency. Intuitively, $\omega_c$ is related to the reciprocal of the duration $\Delta t = 2/\omega_c$ of a single ultrashort “lighthouse” burst of radiation, with corresponding broadband spectrum (Fig. 5), that a fixed distant
observer in the orbital plane sees from the circulating electron. Formally, \( \omega_e \) is the median frequency of SR spectral intensity that this observer sees — i.e., half the total energy is emitted above, half below \( \omega_e \) (Clarke, 2004; Jackson, 1999; Onuki and Elleaume, 2002). The angular energy distribution observed at angle \( \theta \) from the particle’s orbital plane is

\[
\frac{dW_{\text{SR}}}{d\Omega} = \frac{7e^2}{96\pi \varepsilon_0 c} \frac{\omega_e \gamma_e^2}{(1 + \gamma_e^2 \theta^2)^{5/2}} \left( 1 + \frac{5}{7} \frac{\gamma_e^2 \theta^2}{1 + \gamma_e^2 \theta^2} \right). \tag{18}
\]

Each electron emits light in a directional \( 1/\gamma_e \) angle cone centered on its instantaneous velocity, polarized predominantly in the plane of its orbit (power ratio \( P_\parallel/P_\perp = 7 \)). Relativistic electrons generally emit x-rays. Since the 0.1 – 1\( \mu \)m dimension of plasma-accelerated electron bunches exceeds an x-ray wavelength, x-ray SR from such bunches is spatially and temporally incoherent.

2. Undulator and Thomson backscatter radiation

From Eq. (15), \( f(\vec{n}, \vec{\beta}) \) (and thus \( d^2I/d\omega d\Omega \)) is maximized when the denominator \((1 - \vec{\beta} \cdot \vec{n})^2 \) is minimized — i.e., when \( \beta \approx 1 \) and \( \vec{\beta} \parallel \vec{n} \) — and when the numerator \((\vec{n} - \vec{\beta}) \times \vec{\beta} \) is maximized — i.e., when \( \vec{\beta} \perp \vec{n} \). These two conditions are realized for an observer along the axis of a wiggler or undulator, in which periodic external fields perturb relativistic (\( \beta \approx 1 \)) electrons propagating toward the observer (\( \beta \parallel \vec{n} \)) transversely \( \vec{\beta} \perp \vec{n} \). (Clarke, 2004; Onuki and Elleaume, 2002) have reviewed the principles behind, and characteristics of, wiggler and undulator radiation. Here we summarize the main points required for subsequent discussion.

When electrons pass through \( N_u \) periods of alternating magnetic fields (strength \( B_0 \), period \( \lambda_u \)), they bend into a sinusoidal trajectory with maximum deflection angle \( \phi_e = K/\gamma_e \), where

\[
K = e\lambda_u B_0/(2\pi m_e c) \approx 0.93\lambda_u [\text{cm}] B_0 [\text{T}], \tag{19}
\]

is the normalized undulator parameter. \( K \) is the ratio of the deflection angle \( \phi_e \) to the emission cone angle \( \gamma_e^{-1} \), and distinguishes wiggler \((K > 1)\) from undulator \((K < 1)\) modes. Here we focus on the latter, for which emission cones of consecutive oscillations of one electron overlap, and thus superpose coherently.

Locally, a portion of this electron’s trajectory within one undulator period can be approximated by a portion of a circle of radius \( r \). Eqs. (14)-(18) and Fig. 5 then describe the properties of this single radiation burst approximately. Repetition over \( N_u \) periods sends to a distant observer a train of Doppler-shifted bursts separated by time \( \Delta t_u(\lambda_u, \theta, \beta) \) determined by the Lorentz-contracted \( \lambda_u \), the observer’s angle \( \theta \) from the electron beam axis, and \( \beta \). These bursts interfere in the frequency domain, modulating the broadband SR spectrum (Fig. 5) at spectral period \( \omega_u = 2\pi/\Delta t_u \). provided \( \omega_u < \omega_e \). When the latter condition is realized (generally for \( K \rightarrow 1 \)), the observer sees radiation at a fundamental frequency \( \omega_u(\lambda_u, \theta, \beta) \) and its harmonics \( n\omega_u \) \((n = 1, 2, 3, \ldots)\), each with bandwidth \( \Delta \omega /\omega_u = 1/N_u \), as in Fig. 6c. This is analogous to the train of high-order harmonics observed from atoms excited near the ionization threshold by a multi-cycle laser pulse (Protopapas et al., 1997). When \( \omega_u > \omega_e \) (generally for \( K \ll 1 \)), the observer sees only a single spectral peak at \( \omega_u(\lambda_u, \theta, \beta) \), as in Fig. 6a. The detected frequencies are (Clarke, 2004; Corde et al., 2013; Onuki and Elleaume, 2002)

\[
\omega_{\text{sc}} = \frac{2\gamma_e^2}{1 + K^2/2 + \gamma_e^2 \theta^2} n\omega_u, \tag{20}
\]

where the term \( K^2/2 \) takes into account the reduction of longitudinal electron velocity caused by transverse quiver motion. The frequency up-shift (20) is the combined result of Lorentz contraction of \( \lambda_u \) seen by the electron and Doppler shift of the electron’s oscillation frequency viewed in the lab frame. Emission from an electron bunch of dimensions greater than a radiated wavelength is incoherent, although pinhole spatial filtering, when possible, can recover spatial coherence (Attwood et al., 1999).

In the past decade, researchers have studied undulator radiation from LWFA electrons at visible (Gallacher et al., 2009; Schlenvoigt et al., 2008), UV (Lambert et al., 2012), and XUV (Anania et al., 2014; Fuchs et al., 2009; Shaw et al., 2012) wavelengths. Compact, short-period undulators custom-designed for LWFA have emerged (Eichner et al., 2007). This radiation underlies single-shot, non-intercepting diagnostics of LWFA bunch energy and energy spread (Sec. III.C.3) and transverse emittance (Sec. III.C.3). Moreover, these results are widely viewed as first steps toward a LWFA-driven XFEL (Gruener et al., 2007; Nakajima, 2008). Achieving gain, however, will require beams with \( \varepsilon_n \) and slice energy spread that challenge current LWFA capabilities (Maier et al., 2012; Seggebrock et al., 2013). This highlights the need for improved diagnostics and control of these quantities.

The oscillating electric field of a linearly-polarized laser field (frequency \( \omega_0 \), field strength \( a_0 \)) that “collides” with electrons can serve as an optical undulator of period \( \lambda_u \sim 1 \mu \text{m} \). Undulator radiation from one electron — now known as Thomson backscatter — is emitted at frequencies (Bardsley et al., 1989; Brown and Kibble, 1964; Esarey et al., 1993)

\[
\omega_{\text{sc}} = \frac{2\gamma_e^2(1 - \cos \varphi_{\text{coll}})}{1 + a_0^2/2 + \gamma_e^2 \theta^2} n\omega_0, \tag{21}
\]

equivalent to Eq. (20), with \( a_0 \) playing the role of \( K \), and \( \omega_0 \) the role of \( \omega_u \). Here, \( \varphi_{\text{coll}} \) is the collision angle. For head-on collisions (\( \varphi_{\text{coll}} = 180^\circ \)), exact back-scatter (\( \theta = 0 \)), and \( a_0 \ll 1 \), Eq. (21) reduces to \( \omega_{\text{sc}} \approx 4\gamma_e^2 \omega_0 \).
Thus Thomson backscatter of near-infrared \( (h\omega_0 \approx 1 \text{ eV}) \) light from \( \gamma_c \gtrsim 500 \) electrons—a range available from LWFAs—provides tunable, directional MeV photons.

The upper limit of total photon yield from linear \( (a_0 \ll 1) \) Thomson scatter is (Esarey et al., 1993)

\[
N_{sc} = 2\pi\alpha_f N_u N_e a_0^2 \left( \Delta\omega_{sc}/\omega_{sc} \right) \tag{22}
\]

where \( \alpha_f = 1/137 \) is the fine-structure constant, \( N_u \) the number of laser oscillations, and \( N_e \) the number of electrons in the overlap region. Actual yield can be much smaller for non-ideal beams and laser pulses (Debus et al., 2009; Rykovovanov et al., 2014). (Hartemann et al., 2005) calculated that peak brightness of Thomson backscatter of a laser pulse \( (N_{hv}, \text{overlapping photons}) \) colliding head-on with an electron bunch of normalized emittance \( \varepsilon_n \), duration \( \tau_b \) scales as \( \gamma_c^{-2} N_e N_{hv}/\varepsilon_n^2 \tau_b \). The small \( \varepsilon_n \), \( \tau_b \) bunches available from LWFAs are thus advantageous for high Thomson backscatter yield. Explicit calculations (Hartemann et al., 2007) for backscatter of a laser pulse \( (21 \text{ fs}, h\omega_0 = 1.5 \text{ eV}, a_0 \approx 0.3) \) from a LWEA electron bunch \( (N_e \approx 3 \times 10^9, \gamma_c = 340, \varepsilon_n \approx 4 \text{ mm mrad}) \) predicted \( \sim 10^7 \) Thomson photons at \( h\omega_{sc} = 0.7 \text{ MeV} \), comparable to the limit (22) for these conditions. (Debus et al., 2010b) have suggested that Thomson yields \( N_{sc} \approx 10^{10} \) are achievable with a similar LWFA using a laser pulse with tilted front in a side-scatter geometry, which enables a longer laser-electron interaction length (i.e., higher \( N_u \)) than a head-on, backscatter geometry.

In the past decade, starting with (Schwoerer et al., 2006), researchers have developed many LWFA-based Thomson x-ray sources, most using the near-head-on, backscatter geometry. (Umstadter, 2015) has reviewed developments through 2015. Developments since 2015 include linear \( (a_0 < 1) \) backscatter from GeV electrons, resulting in \( \gamma \)-ray photons up to \( h\omega_{sc} \approx 85 \text{ MeV} \) (Shaw et al., 2017), and nonlinear \( (a_0 > 1) \) backscatter from sub-GeV electrons, resulting in high energy tails up to \( \sim 20 \text{ MeV} \) (Yan et al., 2017). Like undulator radiation, Thomson backscatter underlies diagnostics of LWFA bunch energy and energy spread (Sec. III.B.3) and transverse emittance (Sec. III.C.3).

3. Betatron radiation

(Esarey et al., 2002; Kostyukov et al., 2003; Wang et al., 2002) predicted that electrons accelerating in plasma wakes would emit betatron radiation when they undulate transversely in response to the wake’s radial field \( E_r \) (see Fig. 7a). The ion cavity can act as either an undulator or wiggler, depending on how far from the axis electrons are injected. Soon thereafter, (Rousse et al., 2004) observed betatron radiation in the laboratory. Since then betatron radiation has become a versatile ultrashort broadband x-ray source (Albert et al., 2013, 2008; Kneip et al., 2010, 2008; Phuoc et al., 2006; Schnell et al., 2013) as well as an important electron diagnostic (see Sec. III.C.2). (Corde et al., 2013) has reviewed betatron (and other x-ray) radiation from plasma accelerators comprehensively, and introduced its basic physics based on original work by (Esarey et al., 2002; Kostyukov et al., 2003; Thomas, 2010). Here we focus on properties of betatron radiation required for diagnostics.

As for wigglers, one can assign betatron oscillations a period \( \lambda_b = 2\pi/k_\beta \approx 2\pi\sqrt{2\gamma_c}/k_p = \lambda_p\sqrt{2\gamma_c} \) where \( \gamma_c \) is averaged over an oscillation and \( k_p (\lambda_p) \) is plasma wavenumber (wavelength). Thus, for example, in an accelerator producing \( \gamma_c = 200 \) electrons in \( n_e = 10^{19} \text{ cm}^{-3} \) plasma \( (\lambda_p = 10 \mu m) \), we get \( \lambda_b \approx 200 \mu m \). We can also assign oscillation amplitude \( r_\beta \), which injection dynamics determine. The product of \( \gamma_c \), \( k_\beta \) and \( r_\beta \) defines dimen-
Electron energy loss per unit distance is

$$W_{\beta}^{(\text{loss})} = \frac{e^2}{48\pi e_0} \gamma_e^2 k^4 p e_{\beta}$$

$$\simeq 1.5 \times 10^{-45} (\gamma_e n_e [\text{cm}^{-3}] r_{\beta} [\mu m])^2 \text{ MeV/cm}$$

(27)

$$W_{\beta}^{(\text{loss})} \approx 0.006 \text{ MeV/cm}$$ for the above example. Thus one electron emits on average

$$\langle N_{\omega_{\beta}} \rangle = \frac{2\pi}{9 \hbar c (4\pi e_0)} N_{\beta} K \simeq 5.6 \times 10^{-3} N_{\beta} K$$

(28)

photons of mean energy $\hbar \omega_{\beta}$ over $N_{\beta}$ oscillation periods. For our example, each electron emits $\langle N_{\omega_{\beta}} \rangle / N_{\beta} \approx 10^{-3}$ photons per oscillation. Thus a 100 pC bunch emits $\sim 10^7$ photons per oscillation.

Betatron radiation differs from wiggler radiation in that radiating electrons simultaneously accelerate longitudinally. Thus $\gamma_e$ increases along the acceleration path, which implies $\lambda_{\beta} \propto \gamma_e^{1/2}$ (see Fig. 7b). Moreover, it can be shown (Corde et al., 2013) that $r_{\beta} \propto \gamma_e^{1/4}$ (see again Fig. 7b), implying $K \propto \gamma_e^{1/4}$ [via Eq. (23)], and $\omega_{\beta} \propto \gamma_e^{7/4}$ [via Eq. (26)]. Because of these scalings, the conceptually simple spectrum consisting of a fundamental frequency $\omega_{\beta}$ and discrete harmonics of order $N_{\beta} \approx 3K^2/4$ (Corde et al., 2013) smears into a broad continuum. In addition, betatron power increases nonlinearly with $\gamma_e$ [see Eqs. (25),(18)]. Thus most radiation, especially frequencies beyond $\omega_{\beta}$, is emitted at the highest $\gamma_e$, near the end of the accelerator (see Fig. 7c). The source can be further localized longitudinally for electron bunches undergoing collective, high $K$ oscillation, for which radiation is generated mostly at extrema of the electron trajectory (see Fig. 7a). If $N_{\beta}$ is small, most betatron radiation can be generated at the final extremum, within the time scale of an electron bunch duration.

Variations in $K$, $\gamma_e$ during acceleration also render betatron radiation of even a single electron temporally incoherent. Betatron radiation from electron bunches is both temporally and spatially incoherent. This is because a typical critical energy $\hbar \omega_e \approx 10$ keV corresponds to $\lambda \approx 1.24$ angstrom, much smaller than the spatial extent of a $\mu$m-scale LWFA electron bunch. Nevertheless, (Kneip et al., 2010; Shah et al., 2007) observed interference fringes in the shadow of an atomically sharp knife edge inserted into a bright betatron beam at distance $l$ from the source. The knife-edge selected radiation from a small angle range $\theta \ll K/\gamma_e$. The limiting condition $L_{\text{trans}} \Delta k = 1$ then yielded transverse coherence length $L_{\text{trans}} \approx \lambda / 4\pi \sigma_{\pi}$. Such transverse coherence properties provide one diagnostic of $\sigma_{\pi}$ via betatron radiation. The betatron spectrum provides another (see Sec. III.C.2). A similar measurement of the penumbra of a mask (Schnell et al., 2013) can set an upper limit on $\sigma_{\pi}$. (Litos and Corde, 2012) proposed that observations of the profile and spectrum of betatron radiation emitted.
by a PWFA drive bunch could diagnose its proximity to matched beam propagation.

4. Transition radiation

Transition radiation (TR) is emitted when a relativistic electron passes suddenly from one medium into another with a different refractive index (Frank, 1966; Ginzburg, 1979; Ginzburg and Frank, 1946; Ter-Mikaelian, 1972). Within a material, the transiting electron repels surrounding electrons, exciting time-varying radial currents (in metals, Fig. 8a) or polarization waves (in dielectrics) that radiate. However, in the bulk, as long as the phase velocity of the radiation differs from the velocity of the relativistic electron — i.e. there is no Čerenkov radiation — various contributions to the radiation field interfere destructively on volume average. For every plane wave excited at one position there is another with opposite phase. Bulk absorption can also suppress residual radiation when cancellation is incomplete. However, an interface or free surface breaks volume symmetry, enabling net radiation. Moreover, net radiation into vacuum is not absorbed. In contrast to synchrotron, Compton and betatron radiation, here the medium, rather than the relativistic electron, radiates.

FIG. 8 Color online. Transition radiation (TR). (a) Relativistic electron bunch passing through metal foil induces transient surface currents that radiate radially polarized light. (b) Electric field lines (thin red arrows) emanating from relativistic electron (blue circle enclosing “−” sign) emerging from conductor, and converging upon its image charge (red circle enclosing “+” sign) inside conductor. The electron’s previously shielded field expands at c as it emerges from conductor. The electromagnetic shock front (solid red arc) that terminates this expanding field, and bends field lines back to the surface, is the source of TR.

Several authors have derived rigorous expressions for TR from Maxwell’s macroscopic equations (Pafomov, 1971; Schroeder et al., 2004; Sütterlin et al., 2007; Ter-Mikaelian, 1972). Here one applies Maxwell’s interface conditions \( \mathbf{n}_1 \times (\mathbf{E}_2 - \mathbf{E}_1) = 0 \) and \( (\mathbf{D}_2 - \mathbf{D}_1) \cdot \mathbf{n}_1 = \rho_0 \) to the Coulomb field of a relativistic electron passing through an interface of given geometry, then identifies the radiating part of the field. Alternatively, the image charge (or “annihilation radiation”) picture (see Fig. 8b) describes TR microscopically and intuitively (Bolotovskii and Serov, 2009; Carron, 2000; Ter-Mikaelian, 1972). As a relativistic electron propagating inside a conductor reaches the surface at time \( t = 0 \), its previously shielded Coulomb field expands into vacuum at the speed of light, and combines with the field of its image charge receding into the metal. At time \( t \), this field vanishes beyond an expanding sphere of radius \( ct \) centered where the electron emerges from the metal. Since electric field lines terminate only at charges, field lines at this electro-magnetic shock front bend back to the conductor surface. This shock front, which for an ideal conductor with an infinite planar surface is infinitesimally thin, is the source of broadband TR.

\[ a. \text{TR from one electron.} \quad \text{The angular (Ω) distribution of TR spectral power } dW_e/d\omega \text{ from a single electron (e) transiting the step-like surface of an ideal, semi-infinite conductor at normal incidence provides a foundation for describing many basic observable characteristics of TR. It is given by the Ginzburg-Frank formula (Schroeder et al., 2004; Ter-Mikaelian, 1972)} \]

\[
d^2W_e d\omega d\Omega = \frac{r_e m_e c}{\pi^2} \frac{\beta^2 \sin^2 \theta}{(1 - \beta^2 \cos^2 \theta)^2},
\]

where \( r_e = e^2/(4\pi\varepsilon_0 m_e c^2) \) is the classical electron radius and \( \theta \) the angle between the observation direction and the electron propagation direction (see Fig. 8a). The spectral power distribution (29) vanishes on axis (\( \theta = 0 \)), and for highly relativistic (\( \gamma_e \gg 1 \)) electrons, peaks at \( \theta \approx \gamma_e^2 \), and falls off rapidly at larger \( \theta \) (see Fig. 9a). The angular width of this cone is thus a signature of electron energy. The power distribution is axially symmetric, and the TR field is linearly polarized within a plane defined by the electron trajectory and the observation direction. Thus the entire TR beam can be described as radially polarized, consistent with it vanishing on-axis. A lens centered on the electron propagation axis focuses the TR cone to a radially polarized, “donut”-shaped intensity profile (see Fig. 9b) with FWHM \( \approx \sqrt{2} \lambda \) (Artru et al., 1998b; Castellano and Verzilov, 1998). Integrating Eq. (29) over \( \Omega \) for \( \gamma_e \gg 1 \) yields single-electron TR spectral power

\[
\frac{dW_e}{d\omega} = (2r_e m_e c/\pi) \ln(\gamma_e),
\]

which depends only weakly on \( \gamma_e \). A relativistic electron generates backward (reflected) and forward (transmitted) TR cones, respectively, upon entering and exiting a foil (see Fig. 9c).

Eqs. (29) and (30) are frequency-independent, a consequence of the assumption of a semi-infinite, perfectly conducting planar foil. In reality, radiator size and conductivity \( \sigma(\omega) \) are finite, and \( \sigma(\omega) \) and foil relative permittivity \( \varepsilon(\omega) \) depend on frequency. These factors lead to frequency-dependent TR power spectrum, including low-
and high-frequency cutoffs. The effect of finite radiator size stems from the expansion of the electron’s relativistic Coulomb field envelope at angle ∼1/γ_e in the plane perpendicular to its propagation direction (see Fig. 9d). As a result, the passing electron perturbs metal electrons at distance ρ from its path over longitudinally extent ∼ρ/γ_e, which in turn determines the TR wavelength. Thus TR of wavelength λ originates at characteristic distance ∼λγ_e from the electron’s path. A small radiator thus emits less energy at long wavelengths, and diffracts these wavelengths more. Both effects combine to create a low-frequency cutoff of the TR spectrum. Conversely, the critical cut-off frequency ω_{crit} = γ_eω_p of the foil sets a high-frequency limit on TR. Above ω_{crit}, both metals and dielectrics become transparent. Consequently the discontinuity in index of refraction that underlies TR disappears, and spectral power dW_{TR}/dω drops as rapidly as ω^{-4} (Dolgoshein, 1993). For Al, hω_p = 32.8eV. Generally, ω_{crit} lies in the EUV to soft X-ray range for solid foils. (Schroeder et al., 2004), however, observed much lower ω_{crit} for low-energy electrons passing through an underdense plasma-vacuum interface.

Practical e-beam diagnostics often use foils that are tilted, thin or rough, introducing additional TR characteristics beyond Eqs. (29) and (30). A foil with surface normal oriented obliquely to the electron’s path generates a backward TR cone centered on the specular “back-reflected” direction (see Fig. 10a). The forward “transmitted” TR cone is still centered around the electron propagation direction. Rotational symmetry of both cones is broken (see Fig. 10b). In addition, a TR radiator must be thicker than the formation length L_f = βc/(ω [1 − β√ε cos θ]) (Carron, 2000; Dolgoshein, 1993; Ter-Mikaelian, 1972) over which TR (observed at angle θ from the electron’s trajectory) accumulates. Physically, this is the distance over which the Coulomb field of the relativistic electron and the emitted TR drift by one wavelength from one another. TR is greatly diminished for media thinner than L_f. In a transparent medium, such as underdense plasma, L_f can be ∼100μm. In metals, ε ≃ ω/ε_0ω is imaginary, and L_f reduces to ∼δ/2√2 cos θ, where skin depth δ = 2/ωμ_0σ is typically in the nm range. Thus silver-coated Kapton foils can be a good flat TR source.

A related quantity is vacuum formation length L_{f,0}, which defines a coherence length over which TR fields from two or more spatially separated sources interfere. For example, when forward TR from one foil reflects from a second tilted foil, separated by distance L_{sep} < L_{f,0}, in which the same e-bunch generates back-reflected TR, the two TR fields (not intensities) add coherently. Diagnostics based on the Wartski interferometer (Fiorito and Rule, 1994; Fiorito et al., 2006; Wartski et al., 1975) and TR radiators using multiple interfaces (Artru et al., 1975; Dolgoshein, 1993) exploit this effect experimentally. It can also impact the resolution of imaged TR (Artru et al., 1998a,b). In the frequency-domain this can lead to intensity-modulations, whose frequency-width Δω = (ω_2 − ω_1) is defined by L_{sep}. For some natural number l, L_{sep} = l : L_{f,0}(ω_1) for the lower frequency and (l + 1) · L_{f,0}(ω_2) for the higher frequency. As a numerical example, L_{f,0} = γ_e^2c/ω for ε = 1 and θ = 0 and equals ∼1 cm for TR of wavelength λ ≈ 1μm generated by 100 MeV electrons. Generally L_{f,0} greatly exceeds the RMS height ⟨σ_{RMS}⟩ of surface roughness features. Nevertheless roughness can impact TR significantly if surface variations exceed λ within a characteristic disc of radius γ_eλ. In this case, various, sometimes conflicting, effects of roughness on TR have been reported: increased flux due to increased surface area; decreased flux due to transverse shielding, which can also lead to depolarization and disappearance of the central intensity.
minimum; and a speckled intensity pattern (Arutyunyan et al., 1979; Baghiyan, 2001, 2004; Reiche and Rosenzweig, 2001). Roughness affects forward and backward TR in the same way.

b. TR from electron bunches. TR is a useful diagnostic to the extent that it can reveal internal structure of relativistic electron bunches containing $N_e$ electrons. To describe TR from a bunch, one must superpose the TR fields of the individual electrons. At wavelengths $\lambda$ much longer than longitudinal ($\sigma_z$) and transverse ($\sigma_r$) bunch dimensions, these fields differ negligibly in phase, and add coherently. The spectral energy

$$\frac{d^2W_{\text{CTR}}}{d\omega d\Omega} = N_e^2 \cdot \frac{d^2W_e}{d\omega d\Omega}$$

(31a)

of such coherent transition radiation (CTR) scales with $N_e^2$. On the other side, at wavelengths much smaller than the bunch dimensions, the TR fields of different electrons differ by $\geq 2\pi$ in phase, and thus add incoherently. The spectral energy

$$\frac{d^2W_{\text{ITR}}}{d\omega d\Omega} = N_e \cdot \frac{d^2W_e}{d\omega d\Omega}$$

(31b)

of such incoherent transition radiation (ITR) scales linearly with $N_e$. Longitudinal and transverse coherence influence TR in different ways. For $\lambda \gg \sigma_z$, forward TR is fully coherent, and thus $\sim N_e \times$ stronger than for $\lambda \ll \sigma_z$. The result is a high-frequency cut-off in the bunch TR spectrum that has no counterpart in single-electron TR theory. Similarly, bunch radius $\sigma_r$ governs transverse coherence. Transverse coherence is maintained for $\sigma_r \ll \gamma_0 \lambda$ for a collimated bunch, or $\sigma_r \ll \lambda/\Delta \psi$ for a bunch with divergence $\Delta \psi \ll 1$. For this reason CTR foils are often placed close to LWFA, before their beams diverge to larger radii.

Micro-bunching or complex longitudinal bunch profiles complicate the CTR spectrum. A train of two or more bunches separated by delay $\Delta t > \sigma_z/c$ create a corresponding train of TR bursts. The power spectrum of the train is intensity-modulated with period $2\pi/c\Delta t$ in the frequency domain. The spectrum of a complex micro-bunched format with multiple $\Delta t$ contains multiple modulation periods, and is not easily distinguished from a single bunch with complex internal structure.

The total energy loss $W_{\text{tot}}$ of an electron bunch to CTR can be estimated by integrating Eq. (30) over all frequencies and multiplying by $N_e^2$. For an electron incident normally on a TR foil, and for $W_{\text{tot}}$ much less than total bunch kinetic energy $N_e(\gamma - 1)m_e c^2$, the result is (Schroeder et al., 2004)

$$W_{\text{tot}} \simeq (4e m_e c^2) N_e^2 \ln(\gamma_e)/\lambda_{\text{min}}$$

$$\simeq 3.6 \times 10^{-2} (Q[uC] )^2 \ln(\gamma_e)/\lambda_{\text{min}} [\mu m] ,$$

(32a)

$$W_{\text{tot}} \simeq 3.6 \times 10^{-2} (Q[uC] )^2 \ln(\gamma_e)/\lambda_{\text{min}} [\mu m] ,$$

(32b)

where $Q$ denotes bunch charge. This energy loss criterion determines limits on maximum $Q$, radiated bandwidth and time-resolution for diagnostics based on CTR.

Weak ITR at short $\lambda$ is useful for calibrating degree of coherence at longer $\lambda$, and for testing detector dynamic range. Collimated ITR from small sources can reveal spatial bunch characteristics through statistical analysis of intensity noise in its spectrum (Sannibale et al., 2009).

The ideal conductor approximation usually suffices for modeling TR from realistic electron bunches quantitatively (Casalbuoni et al., 2008, 2005a,b; Schroeder et al., 2004; van Tilborg et al., 2004). The most general expression for TR from a bunch of $N_e$ electrons transiting an infinitely wide ideal conductor surface is an integral over individual electron contributions:

$$\frac{d^2W_{\text{ITR}}}{d\omega d\Omega} = \frac{e^2 N_e}{(4\pi \varepsilon_0)^{1/2}} \left\{ \left[ \int d^3\vec{p} g(\mathcal{E}_\parallel^2 + \mathcal{E}_\perp^2) \right]_{\text{ITR}} + N_e \left[ \left| \int d^3\vec{p} g\varepsilon \mathcal{F}_\parallel \right|^2 + \left| \int d^3\vec{p} g\varepsilon \mathcal{F}_\perp \right|^2 \right]_{\text{CTR}} \right\} ,$$

(33a)

where $g \equiv g(\vec{p})$ is electron momentum distribution, $N_e' \equiv N_e - 1$, normalized TR field amplitudes are

$$\mathcal{E}_\parallel = \frac{\left( u \cos \psi [u \sin \psi \cos \phi - (1 + u^2)^{1/2} \sin \theta] \right)}{N'(\theta, u, \psi, \phi)}$$

(33b)

$$\mathcal{E}_\perp = \frac{u^2 \cos \psi \sin \phi \cos \theta}{N'(\theta, u, \psi, \phi)}$$

(33c)

with denominator

$$N'(\theta, u, \psi, \phi) = [(1 + u^2)^{1/2} - u \sin \psi \cos \phi \sin \theta]^2 - u^2 \cos^2 \psi \cos^2 \theta ,$$

(33d)

where $u = p/mc = \gamma_0 \beta$ is normalized momentum, and the form factor is

$$F = \frac{1}{g(\vec{p})} \int q^2 r_{\perp} e^{-i k_{\perp} \cdot r_{\perp}}$$

$$\times \int dz e^{-i(z - k_{\perp} \cdot r_{\perp})/v_z h(\vec{r}, \vec{p})} ,$$

(33c)

where $v_z$ is electron velocity projected along $z$. Quantities with subscript “||” are oriented along the normal $\vec{z}$ to the foil surface; “\perp” quantities lie in this surface ($xy$-plane). Polar angle $\psi$ and azimuthal angle $\phi$ denote electron directions with respect to $\vec{z}$. Without loss of generality, the observation angle $\theta$ with respect to $\vec{z}$ lies in the $x, z$-plane. The 6D frequency distribution $h(\vec{r}, \vec{p})$, with corresponding $g(\vec{p})$, describes electron phase-space. Both $h$ and $g$ are normalized to yield unity when integrated over all of their respective coordinates.

From the general Eqs. (33), one can generate simpler, approximate expressions useful for analyzing many experiments. As one example, when transverse electron momentum is negligible ($\psi \ll 1$), electrons are highly
relativistic ($\gamma_e \gg 1$), and electron position $\vec{r}$ and momentum $\vec{p}$ are uncorrelated, one can write $h(\vec{r}, \vec{p}) = \rho(\vec{r}) g(\vec{p})$. The form factor (33c) then simplifies to

$$F(\omega, \theta) = \int d\vec{r} \rho(\vec{r}) e^{-i\vec{k} \cdot \vec{r}}, \quad (34)$$

i.e. the Fourier transform of the normalized 3D bunch charge distribution $\rho(\vec{r})$. This is the quantity of interest in most beam characterization experiments. If additionally the bunch is cylindrically symmetric ($\phi$-independent) and incident normally $(\langle \psi \rangle = 0)$ on a TR foil, Eqs. (33) simplify to (Schroeder et al., 2004)

$$\frac{d^2W_{\text{TR}}}{d\omega d\Omega} = \left\langle \frac{d^2W_e}{d\omega d\Omega} \right\rangle \left[ N_e + N_e^2 |F(\omega, \theta)|^2 \right], \quad (35)$$

where $\left\langle \frac{d^2W_e}{d\omega d\Omega} \right\rangle$ is the weighted average of (29) over the bunch’s separately measured electron energy distribution, and equals (29) for a mono-energetic bunch. For fully coherent (incoherent) TR, $F = 1$ ($F = 0$) and (35) reduces to Eq. (31a) (Eq. (31b)). If additionally the transverse and longitudinal bunch charge distributions are uncorrelated, we can write $\rho = \rho_{\perp}(\vec{r}_{\perp}) \rho_{\parallel}(z)$ and

$$F = F_{\perp}(\vec{k}_{\perp}) F_{\parallel}(k_z) \quad (\text{Schroeder et al., 2004; van Tilborg et al., 2004}).$$

For a Gaussian bunch with $h(\vec{r}, \vec{p}) = g(\vec{p}) [(2\pi)^{3/2} \sigma_\gamma^2 \sigma_\parallel^2]^{-1} \exp(-r_{\perp}^2/2\sigma_\gamma^2) \exp(-z^2/2\sigma_\parallel^2)$,

$$F_{\perp} = e^{-(1/2)(\omega/c)^2 \sigma_\gamma^2 \sin^2 \theta} \quad (36a)$$

$$F_{\parallel} = e^{-(1/2)(\omega/c)^2 \sin^2 \theta}. \quad (36b)$$

If $F_{\perp}$ is characterized independently, then CTR spectral intensity measured over a wide bandwidth directly yields $|F_{\parallel}(\omega, \theta)|^2$ via Eq. (35). This is the basis of frequency-domain measurements of ultrashort bunch length (Heigoldt et al., 2015). However, the longitudinal bunch profile $\rho_{\parallel}$ does not follow directly from $|F_{\parallel}(\omega, \theta)|$ via Eq. (34) because phase information is lacking. Additional information and assumptions are needed to extract $\rho_{\parallel}$ (Bajlekov et al., 2013; Bakalli Taheri et al., 2016).

As a second example, Eqs. (33) can be adapted to evaluate TR from a radiator with lateral structure, or finite lateral extent, instead of an infinite foil. This example underlies several important diagnostic applications. These include evaluation of TR in the far field, which is defined by observation distances that are large with respect to finite source size. A far-field description is essential, in turn, for modeling focused TR (see Fig. 9b) from electron bunches (Artru et al., 1998b; Castellano and Verzilov, 1998), which diagnoses the bunch’s lateral density profile, and for modeling time-domain TR, which diagnoses bunch duration (van Tilborg et al., 2004). This example is also relevant for TR from an electron bunch passing through a hole of characteristic size $\sim \gamma_e \lambda$ in a foil, a configuration that avoids electron scattering within the foil, while sensitively diagnosing variations in electron bunch pointing and transverse structure (Fiorito, 2001), and for modeling TR generated as electrons exit a µ-scale plasma channel or wakefield structure (van Tilborg et al., 2004). When the radiator is transversely confined or structured, diffraction of TR comes into play. When diffraction is significant, (C)TR is then often called (Coherent) Diffraction Radiation, or (C)DR (Bolotovski and Voskresensiki, 1966; Karlovetz and Potylitsyn, 2008; Pafmov, 1971; Ter-Mikaelian, 1972). A circular disc radiator of radius $\rho_0$ is an important special case, for which Eq. (33a) becomes (Schroeder et al., 2004)

$$d^2W_{\text{CDR}} \frac{d\omega d\Omega}{d\omega d\Omega} = \frac{e^2}{\pi^2 \epsilon_0} N_e (N_e - 1) \sin^2 \theta \times$$

$$\left| \int du g(t) F(\theta, u) \frac{u(1 + u^2)^{1/2}}{1 + u^2 \sin^2 \theta} D(\rho_0, u, \theta) \right|^2, \quad (37a)$$

where we suppressed the dependence of $F$ and $D$ on $k = \omega/c$ for brevity, and $D(\rho_0, u, \theta)$ is a diffraction factor

$$D(\rho, u, \theta) = D(b, u \sin \theta)$$

$$= 1 - J_0(b u \sin \theta) \left[ b K_1(b) + \frac{b^2}{2} K_0(b) \right] - \frac{b^2}{2} K_0(b) J_2(b u \sin \theta), \quad (37b)$$

where $b \equiv k \rho_0/u$, and $J_\nu (K_\nu)$ are ordinary (modified) Bessel functions. Diffraction is important for long-wavelength TR, specifically when $\gamma_e \lambda \gtrsim \rho_0$ (or equivalently $b \lesssim 2 \pi/\beta$). In the opposite limit, $D(\rho_0, u, \theta)$ is close to unity for all $\theta$ within the TR radiation cone. $d^2W_{\text{CDR}}/d\omega d\Omega$ for an electron bunch passing through a circular hole of radius $\rho_0$ can be derived from (37) using Babinet’s principle (Jackson, 1999). Other important analytic CDR results for different $D$ include the donut-shaped intensity distribution

$$I(k, \rho) = \frac{e^2 k^2 c}{8 \pi^2 \epsilon_0 u^2} \int_0^1 d\zeta \left| \frac{\zeta^2}{\zeta^2 + (\beta \gamma)^2} J_1(k \rho \zeta) \right|^2$$

of CTR/CDR imaged from a foil surface to a detector (Fig. 9b), and the CTR field observed at distance $R$ from a foil of dimensions $\ll R$ (van Tilborg et al., 2004)

$$E(\vec{x}, t) = - \frac{e N}{\pi R (4 \pi \epsilon_0)} \frac{dE}{d\omega d\Omega} \int d\vec{k} \langle \mathcal{E}(\theta, u) D(\omega, u, \theta, \rho) \rangle \times F(\omega, u, \theta) \rho e^{-i k (ct - R)}, \quad (39)$$

which is the basis of time-domain measurements of the electron bunch duration. For more complex CTR/CDR sources, the general integral (33) can often be solved numerically (Shkvarunets and Fiorito, 2008).

**B. Bunch charge and energy measurement**

Single-shot measurement of total charge $Q = N_e e$ of each $N_e$-electron bunch, and of the distribution
\[ \frac{dQ}{dE} = \epsilon \frac{dN_e}{dE} \] of charge with electron energy \( E \), are among the most important, beam diagnostics required for any accelerator. Event rate at a collider interaction point or brightness of undulator radiation depend directly on \( Q \), while narrow energy spread is essential for exciting meaningful particle collisions or for driving an FEL. Plasma electron acceleration science has adopted some standard charge and energy spectrum diagnostics from conventional accelerators. Yet unique properties of plasma-accelerated bunches have necessitated re-design or re-calibration of these standard instruments. Sec. III.B.1 reviews re-designed integrating current transformers (ICTs) that measure absolute \( Q \) non-invasively in a noisy environment, and cross-calibrated scintillating screens and imaging plates that measure spatially-resolved electron charge at the sub-pC/mm² level in one shot. Sec. III.B.2 reviews magnet spectrometer designs that accommodate unique electron energy distributions and pointing fluctuations typical of plasma-accelerated beams. Remaining sections review emerging diagnostics with origins inside the plasma accelerator community: non-intercepting \( E \) and \( d\mathcal{N}_e/dE \) diagnostics based on undulator radiation spectroscopy (Sec. III.B.3), and diagnostics of correlated “slice” energy spread based on CTR imaging and optical or plasma-wave micro-deflectors (Sec. III.B.4).

Fig. 11 schematically overviews a LWFA beam diagnostic system. Non-invasive ICTs are placed early in the diagnostic chain. Absorbing spatially-resolved charge detectors at the magnetic spectrometer’s focal plane comprise the end of the beam-line. Beam dumping and radiation protection are also needed. Charge detectors are shown in the context of betatron x-ray, TR and undulator radiation diagnostics.

1. Total and spatially-resolved charge measurement

Faraday cups are a classic method to measure total \( Q \), but intercept the beam, and can become unacceptably bulky when they must capture GeV electrons with large stopping distances. Nevertheless, they have been tested for low-MeV LWFA (Hidding et al., 2007). Nuclear activation provided invasive, cumulative total \( Q \) measurements in early LWFA experiments (Leemans et al., 2001; Reed et al., 2007). However, integrating current transformers (ICTs), which measure integrated current \( \int I(t) dt \) that a bunch induces in a coil upon passing through it, have become the instrument of choice for measuring total \( Q \) from plasma-based accelerators, because they are compact, non-invasive, single-shot and energy-independent over a wide range (Bergoz et al., 1991; Unser, 1989). The constant of proportionality between \( Q \) and \( \int I(t) dt \) depends on the ICT’s geometry and electrical characteristics, and must be calibrated. ICTs were used to report \( Q \) in many early “bubble”-regime LWFA experiments (see Sec. II.C.1).

The major problem that LWFAs posed for ICTs was strong background radiation that unavoidably accompanied LWFA electron bunches, and contaminated ICT signals. The charge of interest from LWFAs is usually contained within a primary quasi-monoenergetic peak (see e.g. \( \sim 270 \) MeV peak in Fig. 11a,b). Often, however, lower-energy electrons (see \( \lesssim 5 \) MeV feature in Fig. 11a,b), and uncharacterized additional electrons below the spectrometer’s detection edge, outnumber electrons in the main high-energy group. Unless shielded, ICTs are sensitive to all of these electrons, and thus can overestimate \( Q \) of the energetic peak. LWFAs are also prolific sources of electromagnetic pulses (EMP) e.g. from electronic devices that drive high-power lasers, supersonic gas jets or high-voltage capillary discharges, or from the laser-plasma interaction itself. An ICT, or its connecting cables, can pick up prompt EMP from these sources that obscures the electron signal. Interaction of the drive laser or accelerated bunch with gas...
cell, alignment apertures, or the ICT itself can create prompt secondary particles or x-rays that also distort the signal. The first calibrations of ICTs against EMP-insensitive detectors [scintillating screens (Glinec et al., 2006); imaging plates (Hidding et al., 2007)] that were independently calibrated at RF accelerators, showed that in noisy LWFA environments, ICTs overestimated $Q$ by factors ranging from $3 - 4 \times$ (Hidding et al., 2007) to $> 10 \times$ (Glinec et al., 2006). This called into question $Q$ values reported in previous LWFA experiments.

(Nakamura et al., 2011) investigated the reasons for these discrepancies by cross-calibrating an ICT with a scintillating screen (Kodak Lanex) and a nuclear activation measurement (Leemans et al., 2001). ICT and Lanex were first cross-calibrated using electrons of energy up to 1.5 GeV from an RF accelerator. Then all three detectors were cross-calibrated at a LWFA outputting electrons over the same energy range. In the LWFA environment, special care was taken to shield the ICT from EMP, secondary radiation, and < 1 MeV electrons. All three diagnostics agreed within ±8% for $Q > 5$ pC bunches, showing that ICTs could measure $Q$ from LWFA accurately with proper shielding. Further research showed that residual EMP within a narrow frequency bandwidth was the primary limit on ICT sensitivity for $Q < 5$ pC. Frequency filtering of the ICT’s signal processing electronics improved sensitivity to the sub-pC level, while improving noise immunity and retaining excellent cross-calibration with other detectors (Nakamura et al., 2016). The filtered ICT has been marketed, an example of commercial product development spurred by plasma accelerator diagnostics research.

Spatially-resolved absolute charge measurement underlies beam profile monitors and magnetic spectrometers. Plasma accelerator researchers have used a variety of detectors to resolve accelerated electrons spatially, including cloud chambers and surface barrier detectors (Clayton et al., 1994), scintillators with photomultipliers (Umonster et al., 1996), scintillating fibers (Gahn et al., 2000), radiochromic film (Galimberti et al., 2005; Giulietti et al., 2002), imaging plates (IPs) (Mangles et al., 2004; Wang et al., 2013), and scintillating screens with cameras. Scintillating screens and IPs have become the most popular choices because they combine wide area detection, high spatial resolution, insensitivity to EMP, wide dynamic range, re-usability and low cost with sensitivity good enough to detect spatially-dispersed few-pC electron bunches in one shot. Nevertheless, the sensitivity of these detectors to X-ray radiation has to be considered in designing detector assemblies.

The working principle and read-out concept of scintillating screens and image plates (IPs) differ considerably from each other. Scintillating screens are based on prompt cathodoluminescence, i.e. rapid conversion of deposited electron (or x-ray) energy into light. Transparent rare-earth oxysulfide host crystals (e.g. Gd$_2$O$_2$S) doped with luminescing rare-earth ions (e.g. Tb$^{3+}$) were developed as medical x-ray phosphors in the 1970s (Wickersheim et al., 1970). As a powder embedded in urethane binder, they comprise the active layer of Kodak “Lanex” screens, used widely at conventional and plasma-based electron accelerators (Buck et al., 2010). The high density ($7.44$ g/cm$^3$) and high average $\bar{Z}$ of Gd$_2$O$_2$S favor strong electron energy deposition [~0.1 MeV per incident 1 MeV $\lesssim E_e \lesssim$ 1GeV electron (Glinec et al., 2006)] via impact excitation. This, combined with high [~16% (Glinec et al., 2006)] intrinsic conversion efficiency of deposited electron energy to excited states of Tb$^{3+}$, and thence to light emission (typically green), results in high electron detection sensitivity. For example, (Buck et al., 2010) reported a lower detection limit of ~0.5 pC in a spot with 11 mm FWHM (i.e. ~10 IC/mm$^2$) for $E_e = 40$ MeV, $\tau_b = 2$ ps electron bunches incident on KODAK Biomax MS screens. However, this limit depends on the type of scintillator and the optical detection system, and may depend also on $E_e$ and $\tau_b$. The efficiency of imaging system and camera can be calibrated absolutely using reference light sources in the plane of the screen (Buck et al., 2010; Kurz et al., 2018; Nakamura et al., 2011). Light emission increases linearly with charge density over 3 to 4 orders of magnitude, with saturation setting in typically at 10–100 pC/mm$^2$ (Buck et al., 2010). Saturation corrections, however, extend their usable range to many 100 pC/mm$^2$. The short (~1 ns) luminescence lifetime is well-suited to high-repetition-rate electron sources. Recently, the consortium responsible for the scintillating screen calibration of (Buck et al., 2010) updated the work under improved conditions, and will soon report revised calibrations (Kurz et al., 2018). They also identified signs of permanent aging at accumulated charge densities of only few 10 nC/mm$^2$, which may necessitate frequent re-placement or scrolling of screens that monitor ~100 pC bunches at high repetition rate.

IPs, also developed for medical radiography (Amemiya and Miyahara, 1988; Miyahara et al., 1986), are based on photo-stimulated luminescence (PSL) (Iwabuchi et al., 1994). Incoming electrons, positrons, x-rays or ions deposit energy in fine Eu$^{2+}$-doped phosphor crystals (e.g. BaFBr), embedded in flexible plastic, by converting Eu$^{2+}$ ions to Eu$^{3+}$. Color centers in the host crystal trap the freed electrons, storing the deposited energy and “exposing” the plate. Sequential visible (typically $\lambda = 0.632\mu$m) irradiation of 50 – 200µm “pixels” (defined by the focused light, not by material boundaries) in a calibrated off-line scanner (manufactured by Fuji-film) releases the trapped electrons. These recombine with Eu$^{3+}$ ions to form excited Eu$^{2+}$ ions, which luminesce (typically $\lambda_{PSL} = 0.39\mu$m) with intensity proportional to the deposited energy. A photomultiplier tube detects the PSL. Exposed IPs fade over several hours (Tanaka et al., 2005; Zeil et al., 2010), are eraseable by exposure to intense light, and re-usable almost indefi-
nately. IP sensitivity to relativistic electrons is high, and nearly energy-independent. For example, (Tanaka et al., 2005) reported detection of $10^3$ electrons with $E_e > 80$ MeV within a $\sim 100\mu m$ pixel — i.e. $\sim 10^{-2}$ pC/mm$^2$. However, detection limits vary depending on the noise floor from co-exposure to x-rays, cosmic rays and other background radiation. IPs provide large dynamic range ($>10^3$) — (Zeil et al., 2010) report no observable saturation effects — but multiple, time-consuming, scans are required to reach the same signal. IPs are insensitive to EMP (Tanaka et al., 2005; Zeil et al., 2010). The measured signal in a well-tested (Hidding et al., 2007; Tanaka et al., 2005; Zeil et al., 2010) range is absolute and universal. Since exposed IPs must be scanned offline, they are best suited for low-repetition-rate systems or for cross-calibration of scintillating screens.

A decade ago, measurements of total $Q$ from LWFA's were plagued with order-of-magnitude uncertainties (Glinec et al., 2006), while convenient spatially-resolved charge detectors, despite decades of use as slowly integrating medical x-ray detectors, remained uncalibrated as single-shot detectors of intense, sub-ps, relativistic electron bunches. The plasma accelerator community has now transformed this situation. Shielded, filtered ICTs that measure sub-pC, absolute $Q$ accurately in noisy LWFA environments are now commercially available. Scintillating screens and IPs are now extensively cross-calibrated at conventional ($\tau_0 \sim$ few ps) and LWF ($\tau_b$ ~ few fs) electron accelerators for matching bunch structure and energy. Ongoing research is examining the response of these detectors to high average charge flux (Kurz et al., 2018), and the dependence of sensitivity, saturation, and linearity on $\tau_0$ in the sub-ps range.

2. MeV and GeV magnetic spectrometers

The dipole magnet spectrometer (Brown, 1975) is the workhorse for single-shot measurement of electron energy spectra both in conventional (Brown, 1975) and plasma-based (Glinec et al., 2006; Nakamura et al., 2008; Sears et al., 2010b) accelerators. Plasma-based electron accelerators, however, pose two special challenges. First, their electrons usually span a wide energy range (as much as low-MeV to multi-GeV). Even in “quasi-monoenergetic” plasma accelerators, low-energy electrons usually accompany the main high-energy peak (see Fig.11a,b). Accurate, simultaneous measurement of both is essential to diagnosing accelerator performance. In contrast, conventional accelerator beams usually need to be characterized accurately only over a narrow energy band. Secondly, plasma accelerators usually launch electrons over a few-mrad range of angles into the magnetic spectrometer, due to betatron oscillations or to pointing fluctuations of the driver. While a well-designed spectrometer can bring angularly-dispersed low-MeV electrons of a given $E_e$ to a common focus at the detector, this is often not feasible with GeV electrons. Special measures are therefore needed to characterize GeV electron trajectories through a spectrometer to avoid errors in evaluating $dN_e/dE_e$.

In contrast, electron beams in conventional accelerators usually enter the spectrometer highly collimated.

Magnetic spectrometers for plasma accelerators can cover MeV to multi-GeV energies in one setup, albeit with varying resolution. The bending radius $\rho$ of a relativistic electron in homogeneous magnetic field $B$

$$\rho = \frac{m_e c \beta \gamma_e}{e B} \approx 1.7 \times 10^{-3} \frac{\gamma_e}{B[T]} \text{ m} \quad (40)$$

defines the scale of the spectrometer — e.g. $\rho \approx 1$ m for $\gamma_e = 600$ with a practical upper limit $B \sim 1$ T.

![FIG. 12 Energy read-out error of dipole magnetic spectrometer for various electron entrance angles (-6 to 6 mrad) and energies (0 < $E_e$ < 600 MeV): (a) low electron energy, detector in focal plane of magnet (corresponding to $E_{1,2}$ in Fig.11), and (b) high electron energy, detector in forward direction before focal plane (corresponding to $E_3$ in Fig.11). Adapted from (Schramm et al., 2017).](image-url)
detector as shown in Fig. 11. In addition, the influence of fringe fields becomes energy-dependent. Nevertheless, relative energy resolution $\Delta E_e/E_e$ on the percent level is usually possible without additional magnet design modifications (Savran et al., 2010), as shown in Fig. 12a. Mapping of the field distribution (including fringe fields) and numerical analysis of beam trajectories is needed to define image plane and resolution precisely.

For energy $E_3$, reached when $\rho$ equals roughly twice the magnet length, focal plane detection must be abandoned in the near forward direction. This regime includes the critical quasi-monoenergetic peak of many plasma-based electron accelerators. Declining spectrometer dispersion $d^2\phi_{\text{def}}/dE_e$ at high $E_e$ by itself tends to decrease energy resolution in this regime. With loss of focal plane detection, uncompensated beam divergence further compromises resolution. The detection point of an electron becomes sensitive to beam pointing jitter, as shown in Fig. 12b, impairing energy calibration. Special measures are then needed to ensure accurate energy measurement. If space is available, additional beam optics can recover focal plane detection (Litots et al., 2014; Weingartner et al., 2011). Alternatively, hard apertures can collimate the beam at the expense of charge. Reference grids, transmission beam pointing monitors (Cha et al., 2012) or feedback from secondary radiation (Shaw et al., 2017) can monitor beam pointing, allowing shot-to-shot correction for jitter. (Clayton et al., 2010; Pollock et al., 2009) introduced a correction scheme with two tandem scintillating screens. By correlating common electron spectral features on each screen, they deduced the complete electron trajectory through the spectrometer, including launch angle. For greater precision, (Wang et al., 2013) employed two tandem grids of 125$\mu$m tungsten wires to cast sharply defined shadows on dispersed electron signals recorded on one IP. Analysis enabled measurement of $E_e$ up to 2 GeV, and associated launch angles, with only $\pm 5\%$ uncertainty at the $2\sigma$ level using a 1.1 T magnetic field within a 4 x 4 cm central plateau. This is a low-cost solution for accurately calibrating $dN_e/dE_e$ up to low-GeV energy. Alternatively, much larger magnets can be used — e.g. a circular 1.25 T field of 40 cm diameter was used to characterize 4 GeV LWFA electrons (Leemans et al., 2014; Nakamura et al., 2008).

Dipole magnet spectrometers have accurately characterized plasma-accelerated electrons from low-MeV to multi-GeV energies (Blumenfeld et al., 2007; Leemans et al., 2014; Litots et al., 2014), resolving peaks as narrow as $\Delta E_e/E_e \approx 0.01$ (Rechatin et al., 2009). With the highest-performing plasma accelerators currently producing bunches with few-% energy spread, the field has not yet challenged magnetic spectrometer technology to achieve higher energy resolution. With scintillating screens, read-out sensitivities of 0.1 pC/MeV can be achieved. Scintillating fibers (Sears et al., 2010b) enable higher sensitivity, at greater cost and with digitized energy resolution. Since beam propagation remains undisturbed in the non-bending plane, the spectrometer’s detector simultaneously measures beam divergence.

3. Spectroscopy of on-axis undulator radiation

A non-intercepting alternative to magnet spectrometers is spectroscopy of forward-directed light that relativistic electron bunches emit on passing through an undulator or Thomson backscatter light field (see Sec. III.A.2). For diagnostics, low values of undulator parameter $K$ (Eq. 19) or field strength $a_0$ (Eq. 2) are preferred to avoid perturbing the electron bunch or complicating the spectrum of emitted radiation (see Fig. 6). Such a noninvasive approach will be essential for monitoring $E_e$, $\Delta E_e$, angular divergence and other beam characteristics in early stages of a multi-stage plasma accelerator (Steink et al., 2016). It will be especially valuable for characterizing few-GeV electrons, a range in which dipole electron spectrometers have difficulty providing image plane diagnostics (see Sec. III.B.2) and require large expensive magnets.

![FIG. 13 Photon energies of undulator and Thomson backscatter radiation from LWFA electrons. Lower right, from low to high $\omega_s$: radiation (open diamonds) from undulators of period $\lambda_0 = 20$ mm (orange, $1 < \omega_s < 2$ eV) (Schlenvoigt et al., 2008); 15 mm (red, $5 < \omega_s < 8$ eV) (Anania et al., 2014); 5 mm (magenta, $40 < \omega_s < 70$ eV) (Fuchs et al., 2009). Upper left: Thomson backscatter radiation (filled diamonds) generated by ultrashort $\lambda_0 = 0.8\mu$m laser pulses of fields strength $a_0 \sim 0.8$ (blue, $6 < \omega_s < 40$ keV) (Khrenikov et al., 2015b); 0.3 (black, $80 < \omega_s < 1000$ keV) (Powers et al., 2014). Note red-shift of higher $a_0$ data. Black circle: data from Fig. 16 for comparison).

Measurement of maximum frequency $\omega_{sc}^{(\text{max})} \propto \left[ \gamma_{\text{max}}^{(\text{max})} \right]^2$ of light emitted on-axis directly determines the maximum electron Lorentz factor $\gamma_{\text{max}}^{(\text{max})}$, which in quasi-monoenergetic plasma accelerators corresponds to the primary spectral peak of interest. The lower right portion of Fig. 13 summarizes results of correlated measurements of Lorentz factors of single-stage LWFA electrons in the
range 100 < \( \gamma \) < 400 (made with magnetic spectrometers) and of the photon energy of undulator radiation from visible (1.8 eV) (Gallacher et al., 2009; Schlenvoigt et al., 2008) to EUV (83 eV) (Anania et al., 2014; Fuchs et al., 2009) that the electrons generated in undulators of periods 0.5 cm < \( \lambda \) ≤ 2.0 cm. The measurements agreed well with the relation \( \omega_{sc} = 2\gamma^2 \omega \) (solid orange, red, magenta lines), from Eq. (20) for \( K < 1 \), \( \theta \rightarrow 0 \). In this photon energy range, and probably up to \( \omega_{sc} \sim \) few keV (corresponding to \( E_e \sim \) few GeV), optical to soft x-ray spectrometers can measure \( \omega_{sc} \) at least as accurately as magnetic spectrometers can measure \( \gamma_e \).

The emergence of miniature, short-period undulators designed specifically for LWFAs (Eichner et al., 2007) will help to make such non-invasive \( \gamma_e \) measurement compact and cost-effective.

The upper left portion of Fig. 13 summarizes corresponding results for Thomson backscatter of \( \lambda = 0.8 \mu m \) light from LWFA photons of similar \( \gamma_e \) range. Since \( \omega_0 \gg \omega \) photon energies now range from 60 keV to 1 MeV (Golovin et al., 2016; Khrennikov et al., 2015b; Powers et al., 2014; Tsai et al., 2015), and agree well with the relation \( \omega_{sc} = 4\gamma^2 \omega_0 \) (solid blue, black lines), from Eq. (21) for \( a_0 < 1 \), \( \theta \rightarrow 0 \), \( \varphi_{coll} = 180^\circ \). Experiments with 1-2 GeV (2000 < \( \gamma_e < 4000 \) electrons (not shown) extended the range to \( \omega_{sc} \sim 80 \) MeV (Shaw et al., 2017). In this photon energy range, the challenge shifts to accurately spectroscopy of such hard x-rays (Gahn et al., 1998; Golovin et al., 2016; Günter et al., 2011; Horst et al., 2015; Sarri et al., 2014). If adequate resolution can be achieved, the possibility of realizing Thomson backscatter from low-MeV (Ta Phuoc et al., 2012) to multi-GeV (Shaw et al., 2017) electrons simply by inserting a plasma mirror (foil) after the plasma to back-reflect the spent drive laser pulse (see Sec. III.A.2) offers an exceptionally low-cost electron energy diagnostic.

In addition to \( \gamma_e \) spectroscopy of undulator/Thomson radiation can also diagnose spread \( \Delta \gamma_e \) of a peaked electron energy distribution if additional beam parameters are known (Brown and Hartemann, 2004; Chouffani et al., 2006; Golovin et al., 2016; Jochmann et al., 2013; Krämer et al., 2018; Schlenvoigt et al., 2008) (see Sec. III.C). Relative on-axis photon energy spread (for \( K \ll 1 \), \( a_0 \ll 1 \)) is determined not only by relative electron energy spread \( \Delta \gamma_e / \gamma_e \), but by electron angular spread \( \sigma_r \), laser bandwidth \( \Delta \omega_0 \) (for Thomson backscatter only), number of periods \( N_u \), and solid angle \( \Omega_{det} \) covered by the on-axis detector:

\[
\left( \frac{\Delta \omega_{sc}}{\omega_{sc}} \right)^2 = \left( \frac{2 \Delta \gamma_e}{\gamma_e} \right)^2 + \left( \gamma_e^2 \sigma_r^2 \right)^2 + \left( \frac{2 \Delta \omega_0}{\omega_0} \right)^2 + \left( \frac{1}{N_u} \right)^2 + \left( \gamma_e^2 \Delta \Omega_{det}^2 \right). \tag{41}
\]

Electron beam optics can reduce divergence (Anania et al., 2014; Becker et al., 2009; Fuchs et al., 2009), but then chromatic selectivity distorts the spectrum.

Absolute \( E_e \) and \( \Delta E_e / E_e \) of LWFA beams of energies (60 MeV < \( E_e \) < 120 MeV) and divergences (1 mrad < \( \sigma_r' \) < 3.5 mrad) have been resolved with undulator radiation (\( N_u \sim 50 \) ranging from near-IR (Gallacher et al., 2009) to vacuum ultraviolet (Anania et al., 2014). After deconvolving the \( \sigma_r' \) contribution, (Gallacher et al., 2009) obtained \( \Delta \gamma_e / \gamma_e \) as small as \( \sim 0.01 \), in agreement with independent magnet spectrometer measurements, equivalent to the smallest LWFA energy spread observed with magnetic spectrometer alone (Rechatin et al., 2009). When magnetic quadrupoles pre-collimated the LWFA beam, \( \Delta \gamma_e \) of the selected main peak became the dominant contribution to \( \Delta \omega_{sc} \). However, discrepancies between \( \Delta \gamma_e \) determined from undulator and magnetic spectrometer were observed (Anania et al., 2014). Extension of such measurements to GeV electrons will be a priority for future research.

4. Slice energy spread

More than a decade after first demonstrations of “quasi-monenergetic” plasma accelerators, electron bunches with overall energy spread \( \Delta E_e / E_e \) below 0.01 have not been observed. Yet \( \Delta E_e / E_e \sim 10^{-3} \) and kA peak current are vital for short-wavelength conventional (Behrens et al., 2014; Röhrs et al., 2009) and LWFA-driven (Gruener et al., 2007) free electron lasers (FELs). \( \Delta E_e / E_e \) of high quality plasma-accelerated bunches can, however, be dominated by correlated spread. This means that \( \Delta E_e / E_e \) is the sum of spreads of longitudinal slices of the bunch, each of which is individually much lower than \( \Delta E_e / E_e \). Relative slice energy spread \( 10^{-3} \) could enable FEL gain with plasma accelerators (Huang and Kim, 2007). Bunch decompression reduces energy spread over the slice at the expense of decreasing the peak current (Seggebrock et al., 2013). Dispersively matched transverse gradient undulators could then operate with different energy slices (Huang et al., 2012) simultaneously. Diagnosis of slice energy spread is thus a key component of current plasma accelerator research.

(Lin et al., 2012) deduced slice properties of LWFA bunches indirectly by imaging visible CTR several meters downstream from an LWFA. If \( \Delta E_e / E_e \sim 0.05 \) were present throughout the bunch’s longitudinal profile, sub-µm density and momentum modulations, a prerequisite for non-zero form factor \( F \) [Eq. (34)] and thus for generating coherent TR at visible wavelengths [Eq. (35)], would disappear within a few centimeters of propagation (Glince et al., 2007; Lundh et al., 2011). (Lin et al., 2012) created such modulations by generating wakes in sufficiently dense (\( n_i > 10^{19} \) cm\(^{-3} \)) plasma that the trailing edge of the drive laser pulse modulated the accelerating bunch. Yet several meters downstream of the LWFA, they observed two clear signatures of CTR: (1) intensity
up to $10^3 \times$ stronger than calculated from Eq. (35) with $F = 0$; (2) imaged TR spatial profiles containing “hot spots” much smaller than the overall beam profile. Persistence of coherent features over such long propagation lengths required $\Delta E_e/E_e|_{\text{slice}} \lesssim 0.005$, well below the total energy spread.

A method for measuring $\Delta E/E|_{\text{slice}}$ directly in the time domain was developed for FEL drivers and FACTET (Dolgashev et al., 2014): an X-band (typically 11.4 or 12 GHz) RF transverse deflection cavity stretches the bunch. Energy-resolving monitored of the streaked transverse profile recovers longitudinally-distributed $dN_e/dE_e$. A slice resolution of 70 fs was demonstrated at SLAC for the 20 GeV FACET beam ($\epsilon_n \sim 40 \mu m$) where for LCLS operating conditions (energy $< 14$ GeV and $\epsilon_n \sim 0.5 \mu m$) values of $< 4$ fs can be reached, illustrating the influence of transverse beam parameters on the resolution. However, RF cavities cannot resolve $\Delta E/E|_{\text{slice}}$ of few-fs LWFA bunches. New approaches employing light fields (Zhang et al., 1996) derives complete expressions for $\epsilon_i$ and $\epsilon$. Generalization to 2D (Fig. 14b) is straightforward. With sufficient charge, the pepper-pot technique can measure $\epsilon$ in one shot. Analogous pin-hole or lenslet arrays characterize transverse phase structure of optical beams (Platt and Shack, 2001).

**C. Transverse emittance measurement**

Early measurements of transverse emittance of LWFA electron bunches adopted mask and focus-scan techniques developed for conventional accelerator beams (Sec. III.C.1). But resolution and $E_e$ range limits motivated development of new methods based on betatron x-ray spectroscopy (Sec. III.C.2), and undulator and transition radiation (Sec. III.C.3).

1. Conventional mask and focus-scan methods

a. Masks. Techniques using beam-intersecting masks (subscript $m$) with 1D (2D) arrays of slits (pinholes) – i.e. “pepper-pots” – were designed to characterize emittance of low-MeV, space-charge-dominated electron beams (Lejeune and Aubert, 1980; Mostacci et al., 2008), although (Delerue, 2011; Thomas et al., 2013) recently extended their use to RF-accelerated GeV electrons. Fig. 14a shows a schematic 1D geometry. A beam impinges at centroid position $\bar{x}$ with average x-momentum $\bar{x}'$ on an array of slits, each of width $d$, centered at $x_{i,m}$. The $i$th slit transmits a beamlet of well-defined origin and low enough charge $e_n$ that it becomes “emittance-dominated” – i.e. positions $x_i$ and transverse momenta $x'_i$ of electrons passing through the slit ($j = 1, 2, ..., n_i$), rather than Coulomb repulsion, dominate beamlet propagation. A screen (subscript $s$) at distance $l$, often thin Ce:YAG to avoid grain-size resolution limits of Lanex, records the centroid position $x_{i,s}$ (yielding average x-momentum $\bar{x}' = (\langle x'_i \rangle, \langle x'^2_i \rangle, \langle x'_m \bar{x}_m \rangle) / l$, spatial profile (diameter $D_{s,i}$), and (if linear in response) relative electron number $n_i$ of each expanded beamlet. As long as each beamlet’s angular spread $\Delta x_i'$ is not too large, the averages $\langle x'^2_i \rangle, \langle x'^2_i \rangle, \langle x'_m \bar{x}_m \rangle$ contribute to its x-emittance $\epsilon_i$ [see Eq. (9)] are related straightforwardly to observed beamlet $(n_i, x_{i,m}, \bar{x}_{i,s}, d_{s,i})$ and whole beam $(\bar{x}, \bar{x'})$ quantities.

A sum of $\epsilon_i$ over slits then yields emittance $\epsilon$ of the subset of incident electrons that passed through the slits, assuming negligible space-charge. (Zhang, 1996) derives analogous pin-hole or lenslet arrays characterize transverse phase structure of optical beams (Platt and Shack, 2001).

FIG. 14 Color online. Pepper-pot measurement of beam emittance. (a) Opaque mask with slits or holes transmits diverging beamlets across beam profile. Thin downstream Ce:YAG screen at distance $l$ detects expanded beamlet profiles (width $D_{s,i}$, location $x_{i,s}$) with typically $\lesssim 10 \mu m$ resolution. (b) Screen image of 125 MeV LWFA beam 30 cm downstream of source, yielding $\epsilon_n \sim 2.2 \pm 0.7 \operatorname{mm} \operatorname{mrad}$ (resolution $\sim 1.2 \operatorname{mm} \operatorname{mrad}$), divergence $\sim 3 \operatorname{mrad}$. Panel b) from (Brunetti et al., 2010).

(Cianchi et al., 2013) pointed out limits of the pepper-pot method for resolving $\epsilon$ of LWFA beams, which is dominated by divergence rather than space-charge. While for typical injector beams, the ratio of divergence to initial beam size is $\sim 1 \operatorname{mrad/mm}$, for LWFA beams it is $\sim 10^3 \operatorname{mrad/mm}$ (e.g. $\sigma_r(0) \approx 1 \mu m$, $\sigma_x' \approx 1 \operatorname{mrad}$). Because of their relatively strong divergence, the phase-space profile of LWFA beams flattens after a few cm of free-space propagation, with $x'$ strongly correlated with $x$ (see Fig. 4c). Consequently, a slit at $x = x_i$ sees a very small uncorrelated spread $\Delta x'_i = (D_{s,i} - d \cdot x_{s,i} / x_i) / l$ of $x'$ values, and has difficulty resolving it above the uninformative geometric slit projection $d(x_{s,i} / x_i) / l$. Narrowing slit width $d$ to compensate scatters the beam. Thus pepper-pots tend to under-sample transverse phase space of LWFA beams.

Nevertheless, pepper-pot measurements placed first-generation upper limits on LWFA beam emittance before other measurements existed. Experiments for quasi-monochromatic, $\sigma_x' \sim$ few-mrad LWFA beams in the energy range 15 to 200 MeV with scanning pin-hole (Fritzler et al., 2004), scanning slit (Sears et al., 2010a),...
and pepper-pot masks (Brunetti et al., 2010; Manahan et al., 2014) yielded upper-limit $\varepsilon_n$ estimates between 1.3 mm mrad with $\sim 30\%$ relative errors. These results were based on estimated upper limit initial beam size $\sigma_r(0) \sim \lambda_p \sim 10 \mu m$. However, (Cianchi et al., 2013) later showed that for $\sigma_r(0) \sim 1 \mu m$ (see e.g. Fig. 4a,b), the phase space ellipse of a $\sigma_r' \sim$ few mmrad beam flattens so quickly that pepper-pots with realistic pin-hole diameters (tens of $\mu m$) cannot resolve their emittance. Indeed betatron x-ray spectroscopy shows $\sigma_r' < 1 \mu m$ (see Sec. III.C.2). On the other hand, mask techniques remain useful for characterizing LWFA beams transmitted through emittance-increasing optics (Manahan et al., 2014), or for evaluating transport conditions between acceleration stages (Dormair et al., 2015; Xu et al., 2016).

b. Focus Scans. Conventionally, $\varepsilon$ of non-space-charge-dominated beams is often characterized by measuring their transverse profile after focusing optics, e.g. at a fixed location while changing focusing strength (quadrupole scan), or at multiple locations with fixed optics (multi-screen measurements) (Minty and Zimmermann, 2013; Rees and Rivkin, 1984). To recover three Twiss parameters [Eq. (10)], at least three measurements, usually over multiple shots, are needed. (Mostacci et al., 2012) succinctly reviewed this technique, and analyzed new measurement uncertainties that come into play for LWFA and other beams with $\gtrsim$ few-% energy spread (see Fig. 3a), compared to $\sim 0.1\%$ for conventional focus-scan measurements, and with large $\sigma_r$ (due to strong divergence) at the first quadrupole in a sequence. The resulting chromatic effects can not only introduce systematic errors into measurement of $\varepsilon$, but can degrade $\varepsilon$ of the beam under test. The severity of these effects depends on the specific quadrupole configuration and on incident beam properties. For example, to characterize a beam with $E_e \sim 150$ MeV, $\Delta E/e/E_e \sim 0.01$ and $\varepsilon_n \sim 1$ mm mrad using a single quadrupole requires sub-mm spot size to avoid strong emittance dilution (Cianchi et al., 2013).

(Weingartner et al., 2012) turned energy spread into a diagnostic advantage by inserting an energy-dispersing magnetic dipole between a fixed high-gradient quadrupole doublet (Becker et al., 2009) and a Ce:YAG screen located 10 cm and $\sim 2$ m, respectively, after a $\sim 250$ MeV LWFA. Screen luminescence displayed energy-resolved beam size as quadrupole position scanned over multiple shots, avoiding chromatic errors in determining $\varepsilon$. Alternatively, this configuration enabled a rare single-shot focus-scan measurement by exploiting the quadrupole focal length’s dependence on $E_e$. Thus the beam’s natural energy spread mapped onto an equivalent focus-scan at the detector without the need for moving or adjusting the field strength of the quadrupole. Both multi- and single-shot measurements yielded $\varepsilon_n = 0.2$ mm mrad with 5-10% uncertainty, well below the upper limits set by pepper-pot measurements. Active lensing in plasma discharge cavities provides a complementary means for performing focus scans close to the plasma accelerator, ensuring small spot size (van Tilborg et al., 2015). Recently (Barber et al., 2017) applied an energy-resolved focus-scan to determine the influence of injection scheme on beam emittance. Results showed that shock-induced down-ramp injection yielded a factor of two lower normalized emittance ($\varepsilon_n \sim 1$ mm mrad) than ionization injection at equal charge densities ($dQ/dE \sim 2$ pC/MeV).

2. Betatron x-ray spectroscopy

In contrast to conventional beam-intercepting emittance diagnostics outside the accelerator, betatron spectroscopy (see Sec. III.A.3) emerged as a non-invasive, albeit indirect, alternative for diagnosing $\varepsilon_n$ within, and near the end of, the accelerator, where most betatron radiation is emitted (see Fig. 7c). Measured quantities are electron spectrum $dQ/dE_e$ (Fig. 11b), betatron x-ray spectrum $dW_\beta^{(\text{meas})}/d\omega$ (Fig. 15) and $\bar{\varepsilon}_e$. A simple analysis fits $dW_\beta^{(\text{meas})}/d\omega$ to a calculated single-electron betatron power spectrum $dW_\beta^{(\text{meas})}/d\omega$ [Eqs. (16), (24); see green curve, Fig. 15], and extracts the critical frequency $\omega_c$ [Eq. (26)], typically $\sim 10$ keV for LWFA forming few-hundred-MeV electrons. Then via the relation $\omega_c \propto \gamma_0^3 k^2 r_\beta \propto \gamma_0^2 \omega_0^2 r_\beta$ from Eqs. (23), (26)], one obtains the betatron oscillation amplitude $r_\beta$ of a single electron, using $\gamma_0$ and $\omega_0$ from the quasi-monoenergetic peak of $dQ/dE_e$ and the $\bar{\varepsilon}_e$ measurement, respectively. The extracted $\bar{\varepsilon}_e$ can be interpreted as an ensemble average amplitude $\bar{r}_\beta$, then equated with (Plateau et al., 2012), or related with help from auxiliary measurements to (Schnell et al., 2012), intra-LWFA bunch radius $\sigma_r$. Combining $\sigma_r$ with measured beam divergence $\sigma_r'^{(\text{meas})}$ outside the accelerator – e.g. from the width of the electron spectrometer signal orthogonal to the dispersion plane – yields the estimate

$$
\varepsilon_n \approx \gamma_0 \sigma_r \sigma_r'^{(\text{meas})}
$$

(42) for the uncorrelated component of normalized emittance [i.e. first term on right-hand side of Eq. (12a)]. However, Eq. (42) neglects correlations between $x$ and $x'$ [second term on right-hand side of Eq. (12a)]. Moreover, $\sigma_r'^{(\text{meas})}$ can misrepresent $\langle \gamma_0^2 x^2 \rangle$ inside the LWFA if the plasma-vacuum boundary re-shapes the beam’s phase-space ellipse (see Fig. 4a,b), thereby altering its downstream divergence (Sears et al., 2010a; Weingartner et al., 2012).

(Köhler et al., 2016; Plateau et al., 2012; Schnell et al., 2012) used this simple analysis method to estimate $\varepsilon_n$ from quasi-monoenergetic LWFA under a variety of conditions. They measured $dW_\beta^{(\text{meas})}/d\omega$ in a single shot...
They assumed bunches were radially symmetric and did not interact with the tail of the drive pulse. Accordingly, they confined analysis to shots that produced radially symmetric betatron x-ray profiles, and to conditions for which bubble radius significantly exceeded drive pulse length. With these simplifications, they solved the inverse problem to recover a complete intra-LWFA radial beam profile \( P(r) \), instead of just an average \( \sigma_r \), and a complete distribution \( \Theta(\theta_d) \) of transverse angles \( \theta_d = dr/dz \), including correlations \( r(\theta_d) \). This enabled determination of \( \varepsilon_n = 0.6 \pm 0.1 \) mm mrad for a \( \gamma_e = 640\pm6 \) beam, including correlations \( \langle x\gamma_e x' \rangle^2 \), solely from \( dW_{\beta}^{(\text{meas})}/d\omega \) and \( dQ/dE_e \) measurements, without invoking \( \sigma_r^{(\text{meas})} \). Nevertheless, calculated \( \sigma_r' \) agreed with \( \sigma_r^{(\text{meas})} \) for the conditions investigated. On the other hand, estimating emittance of the same beam from Eq. (42), using \( \sigma_r^{(\text{meas})} \), yielded \( \varepsilon_n = 1.5 \pm 0.3 \) mm mrad, demonstrating that correlations contribute significantly to \( \varepsilon_n \) inside a plasma bubble.

Accelerating electrons with non-planar trajectories — i.e. angular momentum — can generate radially asymmetric betatron x-ray intensity profiles (Phuoc et al., 2006). Although a linearly-polarized drive laser pulse imparts no net angular momentum to its wake, if it has a radially asymmetric spatial profile, it creates a plasma bubble with symmetric focusing forces \( F_x \neq F_y \). Accelerating electrons and their heavier plasma cavity can then acquire equal and opposite, cyclically-evolving angular momenta, while total angular momentum remains constant. (Thaury et al., 2013) observed such periodic cycling of betatron x-ray asymmetry by controlling acceleration length with colliding-pulse injection (see Sec. II.C.1). Since angular momentum, over and above emittance, influences downstream beam propagation, its accurate single-shot diagnosis is essential. Additionally, observation of a non-uniformly polarized betatron x-ray profile signifies a preferred oscillation direction, which can arise from bunch interaction with the rear of the drive laser pulse (Cipiccia et al., 2011; Curcio et al., 2015; Mangles et al., 2006; Németh et al., 2008).

3. Undulator and transition radiation diagnostics

Radiation that electron beams emit outside an accelerator — e.g. undulator/Thomson-scatter (Sec. III.A.2) or TR (Sec. III.A.4) — can also characterize their emittance indirectly. (Leemans et al., 1996) and (Chouffani et al., 2006; Johmann et al., 2013; Krämer et al., 2018) developed basic principles for diagnosing beam divergence \( \sigma_{x,y} \) using 90° and 180° Thomson scattering, respectively. These studies used tens-of-MeV electron bunches of negligible energy spread \( (\Delta E_e/\gamma_e \lesssim 0.003) \) from conventional linacs. In the 90° geometry, a ~100 fs laser pulse scattered from a longitudinal slice of a 10-15 ps bunch, selected by changing bunch-laser delay. The energy-integrated transverse Thomson x-ray intensity profile was measured on a phosphor screen, then fit to a single-
electron angular power distribution $dW_{\text{el}}/d\Omega$ integrated over a Gaussian distribution of electron propagation angles to extract $\sigma_x' \sigma_y'$. The 180° geometry integrated over bunch length, but $\theta$-dependent Thomson-scatter spectra — given by Eq. (21) with $\varphi_{\text{coll}} = 180^\circ$ for one electron — were measured with high spectral/angular resolution using either Si(Li) and PIN detectors (Choufani et al., 2006) or a pixelated x-ray CCD (Jochmann et al., 2013) similar to those used in betatron x-ray spectroscopy (see Sec. III.C.2). Fig. 16 shows typical results from (Jochmann et al., 2013). The peak of the x-ray photon energy distribution $dN(\theta)/dE$ (red swath) closely tracked Eq. (21) (black dotted curve), except for a $\sim 5\%$ red-shift near $\theta = 0$, attributable to the fraction of the diverging electron ensemble that deviated from $\varphi_{\text{coll}} = 180^\circ$. The width and asymmetry of $dN(\theta)/dE$ (black solid curves) changed dramatically with increasing observation angle $\theta$. Numerical fits to these distinctive features enabled extraction of electron angular distribution with unprecedented accuracy.

However, the measurement is then destructive. As a non-invasive alternative, (Golovin et al., 2016) obtained single-shot, $\theta$-resolved spectra equivalent to Fig. 16 by Thomson-scattering a 40 fs ($\Delta \omega_0/\omega_0 \sim 0.03$) laser pulse at $\varphi_{\text{coll}} = 170^\circ$ from a $\sim 60$ MeV LWFA beam with $\Delta \gamma_e/\gamma_e \sim 0.1$. Under these conditions $\Delta \gamma_e, \sigma_x'$ and $\Delta \omega_0$ all contributed to the observed spectral width. They then simulated Thomson scatter using a beam of variable $\gamma_e, \Delta \gamma_e$ and $\sigma_x'$ to achieve the best global fit to the $\theta$-dependent spectrum. The fit correctly recovered independently-measured $\gamma_e$ and $\Delta \gamma_e$, and output the beam’s local $\sigma_x'$ at the collision point. The latter was observed approximately to double as the beam propagated from the LWFA exit to a point 40 cm downstream, a signature of emittance growth due to space charge.

In addition to $\sigma_x'$, a measurement of $\sigma_r$ is needed to estimate $\varepsilon_n$ via Eq. (42). (Golovin et al., 2016), like (Kneip et al., 2012; Köhler et al., 2016; Schnell et al., 2012) earlier (see Sec. III.C.2), used x-ray knife-edge shadowing, which has difficulty resolving $\sigma_r \lesssim 1\mu$m beams expected near a LWFA exit. A potential minimally invasive alternative is TR imaged from a thin foil surface to a detector. For $\sigma_r \gg 1\mu$m beams, this image is simply a replica of the transverse beam profile. For $\sigma_r \lesssim 1\mu$m beams, TR images formed from visible light no longer directly resolve this profile. On the other hand, the imaged profile transforms to the annular point spread function of a single electron [Fig. 9b, Eq. (38)]. Properties of the central minimum, which has been observed from extremely low $\varepsilon_n$ conventional beams (Karataev et al., 2011), can potentially diagnose $\sigma_r$ in this range. (Bourgeois et al., 2012) recovered transverse LWFA beam profiles from CTR images using an iterative algorithm. Near an LWFA exit, a second upstream foil is needed to deflect the intense drive laser pulse.

**FIG. 16** Color online. Thomson-backscattering spectra $dN/dE$ vs. observation angle $\theta$. Electron beam: $E_e = 22.5$ MeV, $\Delta E_e/E_e \approx 0.0025$, $\tau_0 \approx 4$ ps, $Q = 77$ pC linac bunches; laser: $\lambda = 800$ nm, $\Delta \lambda \approx 20$ nm, $a_0 \approx 0.05$. Black solid curves: experimental data; dotted curve: Eq. 21; linear color scale: numerical model of $dN(\theta)/dE$ with blue minimum, red maximum. From (Jochmann et al., 2013).

The accuracy of the above measurements relied on small relative energy spread $\Delta \gamma_e/\gamma_e$, just as undulator-based measurements of $\Delta \gamma_e/\gamma_e$ relied conversely on small angular spread (see Sec. III.B.3). Small detector angle ($\Omega_{\text{det}}$) and small laser bandwidth ($\Delta \omega_0/\omega_0, \Delta \omega_0$) were also essential. Eq. (41) summarized the tradeoffs (Krämer et al., 2018). The principal new challenge that LWFA beams presented for undulator/Thomson-based emittance characterization was their comparatively large $\Delta \gamma_e/\gamma_e$ (see Sec. II.C.1, Fig. 3). (Fuchs et al., 2009) chromatically focused an LWFA beam into an undulator, thereby selecting a narrow energy band.

**D. Bunch length measurement**

Conventional radio-frequency electron accelerators based on photocathodes illuminated with short laser pulses generally produce electron bunches as short as a few picoseconds, limited by energy spread, and peak currents up to $\sim 100$ A. The advent of compact XUV and x-ray free-electron lasers drove development of magnetic chicane compressors capable of reducing these durations to $\sim 100$ fs, while increasing peak current to $>1$ kA (Dohlus et al., 2005). Such sources have provided drive and injected bunches for recent electron-beam-driven PWFA experiments (Corde et al., 2015; Litos et al., 2014). Time-domain diagnostic methods such as transverse deflection structures (TDSs) (Behrens et al., 2012) and electro-optic (EO) methods (Berden et al., 2007) are well-suited, and widely used, for characterizing bunch durations in this range. LWFA operating in the “bubble” regime, on the other hand, are capable of
producing electron bunches of only a few femtoseconds duration. Measurement of such short bunch durations, and their internal profiles, is one of the greatest new, ongoing diagnostic challenges that plasma electron accelerators have posed. Some researchers have addressed this challenge with creative extensions of time-domain TDS (Zhang et al., 2016a) and EO (Debus et al., 2010a; van Tilborg et al., 2006) methods. (Buck et al., 2011) introduced a time-domain magneto-optic (MO) method to resolve LWFA bunch durations in the few-fs range (Secs. III.D.1.c and IV.C.3). Most recently, frequency-domain, wide-bandwidth OTR methods have successfully characterized few-fs bunch profiles (Heigoldt et al., 2013). Preliminary CDR (Castellano et al., 2001; Fiorito, 2001) and Smith-Purcell radiation (Andrews et al., 2014a) results also appear promising.

1. Time-domain methods

Purely electronic time-domain bunch characterization methods, such as integrating current transformers (Nakamura et al., 2016), resolve bunch duration at best down to the nanosecond level. TDS and EO methods, on the other hand, provide sub-ps characterization.

a. Transverse deflecting structures. A TDS, analogous to a streak camera, imparts a rapidly time-varying transverse momentum kick to an electron bunch that deflects electrons at different longitudinal positions within a bunch to different transverse locations on a downstream detector. Microwave transverse deflecting cavities have resolved longitudinal features of bunches from conventional rf accelerators down to the hundreds of fs scale (Behrens et al., 2012; Ding et al., 2011; Röhrs et al., 2009; Xiang et al., 2011). However, microwave TDS’s have not yet been applied to LWFA beams.

The problem of characterizing few-fs LWFA bunches has spurred the proposal (Bettoni et al., 2016; Dornmair et al., 2016; Xiang and Huang, 2007) and laboratory demonstration (Bettoni et al., 2016; Zhang et al., 2016a) of new TDS configurations that use near-IR light fields (Xiang and Huang, 2007; Zhang et al., 2016a) or wake fields (Bettoni et al., 2016; Dornmair et al., 2016) with few-fs periods to deflect electrons, instead of microwaves with ns periods. (Bettoni et al., 2016) proposed and demonstrated a passive deflector based on the self-transverse wakefield interaction of the bunch passing off-axis through a dielectric-lined or corrugated waveguide. (Dornmair et al., 2016) proposed to propagate the subject electron bunch obliquely through the zero-crossing between focusing and defocusing fields in a linear LWFA, where fields vary rapidly enough to deflect different longitudinal slices of a few-fs bunch in substantially different directions. (Xiang and Huang, 2007) proposed to propagate the subject bunch through a small aperture in a CTR foil oriented at 45° to its propagation direction. Radiation recoil from back-reflected near-IR CTR (see Fig. 10a) deflects electrons in proportion to the product of bunch charge and form factor (33e), from which bunch length can be estimated.

The experiment of (Zhang et al., 2016a) used electric fields in the trailing edge of a near-IR LWFA drive pulse to transversely deflect electrons in an energy-chirped bunch accelerating inside a LWFA, thereby locally enhancing their betatron motion in opposite directions every half-cycle (see Fig. 17a). The technique could equally well be implemented with a separate probe pulse overlapping and co-propagating with the accelerating bunch. (Zhang et al., 2016a) relied on energy chirp within the accelerating bunch to map longitudinal position within the bunch onto energy. Thus when the modulated bunch exited the accelerator, and passed through a slit that blocked its most deflected parts (Fig. 17b-1), a periodic series of minima appeared in the energy distribution observed at the detection plane of a magnetic spectrometer (Fig. 17b-3). Simply counting these minima yielded the bunch duration in units of optical half-cycles. By ana-
lyzing energy-dependent modulation amplitudes, (Zhang et al., 2016a) reconstructed the longitudinal shape of the ~4 fs bunch (Fig. 17c). Although invasive and reliant on energy chirp in its current form, the method of (Zhang et al., 2016a) is an important first demonstration of the promise of advanced TDS methods for characterizing few-fs bunches.

b. Electro-optic methods. EO methods convert either the Lorentz-contracted Coulomb field of the electron bunch itself (Casalbuoni et al., 2005a) or a sub-cycle, THz-frequency CTR pulse that the bunch generated (Berden et al., 2007, 2004; Jamison et al., 2003; Schmidt, 2006; Steffen et al., 2009; Wilke et al., 2002) into an optical signal of similar duration. The temporal envelope $|\mathcal{E}(t)|$ of the bunch or CTR field approximates the bunch’s longitudinal profile. The quasi-static $|\mathcal{E}(t)|$ overlaps a time-synchronized, co-propagating optical probe pulse in a transparent EO crystal (e.g. ZnTe, GaP), and modulates the probe’s polarization via the Pockels effect. A polarization analyzer filters out this modulated portion of the probe (see Fig. 18a), which carries information about the duration $\tau_E$ and profile of the $|\mathcal{E}(t)|$ impulse. Direct EO detection of the fields of LWFA bunches near the LWFA exit (where they are shortest) is challenging because of strong background signals from the drive laser and irradiated plasma. The CTR approach avoids this background by propagating a CTR THz pulse generated near the accelerator to remote detectors, and has been used for most EO measurements of LWFA bunch length.

Time resolution of EO measurements has been discussed extensively in the literature. A conceptually simple EO measurement of $|\mathcal{E}(t)|$ would use a transform-limited probe pulse with duration $\tau_{pr}^{(0)} < \tau_E$, then record intensity of the EO-modulated probe vs. probe-CTR delay $\Delta t$. This procedure maps $|\mathcal{E}(t)|$ with resolution $\tau_{pr}^{(0)}$, but requires multiple shots. For single-shot EO measurements, (Sun et al., 1998) chirped the probe to duration $\tau_{pr}(ch) > \tau_E$, so that it overlapped the entire $|\mathcal{E}(t)|$ profile. The THz field then encoded its waveform onto the probe spectrum (see Fig. 18a). Measurement of the modulated probe spectrum then decoded this waveform, but degraded resolution to $[\tau_{pr}^{(0)} \tau_{pr}(ch)]^{1/2}$, because the THz pulse modulated only a portion of the probe spectrum (Sun et al., 1998). However, (Berden et al., 2004) showed that by instead reconstructing the time-domain field of the EO-modulated probe using standard single-shot cross-correlation (Salin et al., 1987), one could recover resolution $\sim \tau_{pr}^{(0)}$ in single-shot bunch profile measurements. They thereby recovered the profiles of $\sim 275$ fs RMS (650 fs FWHM), 50 MeV bunches from a radio-frequency accelerator with $\sim 50$ fs resolution. EO measurements of shorter LWFA bunches, however, encounter further limits from THz transverse optical phonon resonances of common EO crystals (e.g. 5.3 THz for ZnTe, 11 THz for GaP), which absorb THz light and cause group-velocity walk-off of the THz signal from the optical probe. These effects can severely distort fs-duration profiles (Casalbuoni et al., 2005a; Gallot et al., 1999). By using thin (100-300 µm) ZnTe crystals, (van Tilborg et al., 2006) minimized these distortions, and obtained an upper limit of $\sim 50$ fs RMS (120 fs FWHM) on LWFA bunch duration from a multi-shot EO measurement.

(Debus et al., 2010a) reduced this upper limit further to 13 fs RMS (32 fs FWHM) in a single-shot EO measurement that took advantage of the CTR signal from a relatively long (0.71 ps) low-energy background electron bunch with thermal energy distribution ($kT_e = 6$ MeV) that emerged from a LWFA (45 fs, 0.5 J drive pulse, $\bar{n}_e = 1.5 \times 10^{19}$ cm$^{-3}$ plasma) along with the main quasi-monoenergetic [40±7 (RMS) MeV, $\sim 30$ pC] bunch.
At a CTR foil 5 mm downstream of the accelerator, the centroid of the background bunch trailed the main bunch by 0.36 ps, with its leading edge overlapping (see Fig. 18b, top). Consequently the two contributions to the CTR signal formed a time-domain interference pattern (Fig. 18b, bottom) that encoded the duration and phase of the ultrashort portion of the signal, and that a cross-correlator measured. (Debus et al., 2010a) exploited the strong THz dispersion of the ZnTe crystal to re-shape the ultrashort CTR component in a way that enhanced visibility of the interference fringes. A theoretical fit (Fig. 18c, green curve) of the measured time-domain interference pattern (Fig. 18c, bold black curve) yielded the bunch duration cited above. This milestone notwithstanding, EO measurements remain limited by probe pulse bandwidth and by the accuracy with which EO crystal dispersion can be modeled. Thus they have provided only an upper limit on LWFA bunch duration.

2. Frequency-domain methods

Spectral bunch length measurements are based on analyzing TR over a bandwidth that can extend from far infrared ($\lambda \sim 30\mu m$) to ultraviolet ($\lambda \sim 0.3\mu m$), including both coherent and incoherent emission (see Fig. 19). In principle, spectral methods can measure bunch length practically anywhere along an electron beam line with sub-fs resolution in a single shot, while disturbing the beam minimally. The principal challenges lie in calibrating TR spectral intensity accurately and with high spectral resolution over a wide wavelength and dynamic range, and in solving the difficult inverse problem of retrieving a (sometimes complicated) bunch profile uniquely from spectral intensity measurements. Bunches whose shortest features are of duration $\tau_b \gtrsim 100$ fs generate TR predominantly at far-IR wavelengths ($\lambda \gtrsim c\tau_b \sim 30\mu m$), a range where sensitive detectors with high spectral resolution are lacking. Time-domain EO methods are thus superior in this $\tau_b$ range (see Fig. 19). On the other hand, bunches with $\tau_b \lesssim 10$ fs — i.e., those from strongly nonlinear LWFA — can be accurately characterized via their mid-IR to UV TR, a range that is amenable to broad-band, high-resolution spectral detection.

![FIG. 19 Color online. Overview of time and spectral scales for single-shot profiling of longitudinal LWFA bunch profile, including (right to left) sub-structures, main bunch profile, and background pedestal. Solid blue curve: TR spectrum of 200 MeV, 20 pC (Gaussian) bunch with $\tau_b = 10$ fs (FWHM). Horizontal black lines: ranges best suited for time-domain EO (left) and frequency-domain CTR (right) techniques.](image-url)
foil damaged by the LWFA drive pulse after each shot.

(Glinec et al., 2007), used a foil placed only 1.5 mm beyond a quasi-monoenergetic LWFA, and spectrally analyzed CTR in a single shot over a range 400 < λ < 850 nm (see Fig. 20a). This ∼1-octave bandwidth was too narrow to resolve τb, or internal bunch structure. However, (Glinec et al., 2007) observed strong frequency-domain interference fringes (discussed further in Sec.IV.B.2) with period ∆ω = (∆t)^{-1}, consistent with CTR from two sources separated longitudinally by ∆t = 74 fs. A model of the CTR spectra suggested that these sources were a ∼10 fs bunch in the first wakefield bucket, followed by a few-fs bunch – or one with substructure on this scale – in the second (see Fig. 20a, inset). Later (Lundh et al., 2013) used similar spectral interference fringes to diagnose bunch distribution among multiple buckets.

(Lundh et al., 2011) analyzed CTR and ITR spectra from a foil 15 mm downstream of a quasi-monoenergetic (85 MeV) LWFA over a 3.3-octave range extending from mid-IR (λ = 5.5μm) to visible (λ = 0.55μm) wavelengths (see Fig. 20b). They combined an optical spectrometer with an infrared monochromator that could only acquire the IR portion of the spectra over multiple shots. Nevertheless, a fit based on Eq. (39) to the measured TR spectra, which exhibited a coherent threshold at λ ≈ 1μm, yielded a most probable ensemble duration 1.4 - 1.8 fs (RMS). The researchers could not determine whether this was the duration of the entire bunch or the shortest feature in a longer bunch, nor whether their use of colliding optical pulse injection influenced the bunch duration. Nevertheless, this remains the shortest τb reported from a LWFA.

(Bajlekov et al., 2013) and (Heigoldt et al., 2015) demonstrated the first single-shot high-resolution spectroscopic bunch length measurements by distributing > 4 octaves of CTR bandwidth amongst visible (0.4−1.1μm), near-IR (1.1−1.8μm) and mid-IR (1.7−7.1μm) spectrometers. In these experiments, a ∼50 TW laser pulse drove a nonlinear LWFA in \( n_e = 3.9 \times 10^{19} \text{ cm}^{-3} \) plasma in a cell that tuned in length \( L \) from 3 to 14 mm in 1 mm increments. As \( L \) increased from 3 mm, (Heigoldt et al., 2015) observed electron energy increased to a maximum ∼650 MeV at \( L = 9 \text{ mm} \), consistent with the pump depletion length \( L_{pd} \). For \( L < 9 \text{ mm} \), they observed smooth CTR spectral profiles, and correspondingly reconstructed longitudinal charge profiles \( \rho(t) \) consisting of single bunches of duration ≤ 5 fs (FWHM) (see Fig. 20c). For \( L > 9 \text{ mm} \), on the other hand, they observed CTR spectra that were modulated at a single dominant frequency (as in Fig. 20a). Correspondingly, the reconstructed \( \rho(t) \) included an additional bunch trailing by \( \Delta t \approx 50 \text{ fs} \) (Fig. 20c), slightly less than a plasma period (2π/ωp = 56 fs). As \( n_e \) changed, \( \Delta t \) tracked, but remained less than, 2π/ωp. The authors conjectured that the trailing bunch was injected within the first LWFA cavity in response to a transition from LWFA to beam-driven PWFA starting at \( L \approx L_{pd} \). Indeed, such tunable bunch pairs are of interest as drive-witness pairs for tabletop PWFAs, of interest in turn for compact electron acceleration free of dephasing. These results demonstrate the ability of multi-octave-bandwidth CTR to reconstruct \( \rho(t) \) simultaneously on the few-fs scale of a single bunch and the tens-of-fs scale of separated bunches. They also demonstrate quantitative diagnosis of bunch evolution with propagation through a LWFA.

In general, all CTR spectral intensity measurements...
described above diagnose electron bunch profiles indirectly. As discussed in Sec. III.A.4, these yield the magnitude $|F_{\|}(\omega, \theta)|$ of the form factor via Eq. (35), but not its spectral phase, which is required to extract $\rho_{\|}(z)$ directly via Eq. (34). In principle, the spectral phase of $F_{\|}$ could be determined from the spectral phase of the CTR field (via Eq. (39)), measured e.g. by CTR interferometry. However, this has so far proven impractical to do with high resolution over a multi-octave bandwidth. Hence the inverse problem of reconstructing $\rho_{\|}(z)$ from CTR spectra is ill-posed, analogous to determining the structure of a molecule from a diffraction pattern generated by coherent x-rays. Iterative algorithms — see (Bajlekov et al., 2013; Bakkali Taheri et al., 2016; Heigoldt et al., 2015) and references therein — are used to reconstruct $\rho_{\|}(z)$ — subject to physical constraints — from spectral intensity measurements alone. Briefly, most reconstruction algorithms are variants of the following approach: one starts with an initial guess $\tilde{\rho}_{\|}(0)(z)$, calculates its complex Fourier transform $|G(k)|\exp[i\psi(k)]$, then replaces the amplitude $|G(k)|$ with the experimental $|F_{\|}(k)|$. An inverse Fourier transform then yields a revised estimate $\tilde{\rho}_{\|}(0)(z)$. Finally, one forcibly adjusts $\tilde{\rho}_{\|}(0)(z)$ to satisfy physical constraints — e.g. it must be real, positive-definite, and non-zero only within a realistic temporal range — yielding $\rho_{\|}(1)(z)$, to complete the first iteration. One then re-iterates until, hopefully, the solution converges. Challenges include ensuring convergence, demonstrating uniqueness, and quantifying uncertainty of the result based on uncertainties of measured inputs. As with many inverse-problems, simply searching for one “needle” — i.e. a best fit to available data — in a figurative haystack of possibilities is not enough. One must exhaust a sizable portion of the haystack to ensure there is only one, or no, needle left (Tarantola, 2006).

Current bunch profile reconstruction algorithms solve a 1D problem of reconstructing $\rho_{\|}(z)$ from $\Omega$-integrated CTR power spectra. Expanding retrieval algorithms and CTR spectral data to 2 and 3 dimensions, analogous to x-ray scattering from molecules, will be an important future direction. Adding an angular dimension to spectral data not only enables simultaneous access to $\rho_{\perp}(\vec{r}_{\perp})$, but can reduce ambiguities in reconstructed $\rho_{\|}(\vec{r}_{\perp})$ that are fundamental in the corresponding 1D problem. Diffraction radiation (DR) from bunches passing through non-interactive slits and apertures adds such a dimension (Karlovets and Potylitsyn, 2008), and has led to reconstructions of transverse and longitudinal profiles of $\sim$ps bunches that benchmark successfully with independent diagnostics (Castellano et al., 2001; Fiorito, 2001). Its non-invasiveness is especially attractive for $\sim$μm-scale LWFA bunches. DR thus appears ripe for extension to few-fs LWFA bunches, but will require small apertures to avoid short-wavelength/short-time-scale cutoffs, and hence good pointing stability. Similar remarks apply to Smith-Purcell radiation, emitted when a relativistic beam passes over a grating — i.e. a regular array of diffractive radiators placed within a vacuum formation length (Brownell et al., 1998; Karlovets and Potylitsyn, 2006; Kesar, 2010; Smith and Purcell, 1953). The freedom to tailor grating dimension, spacing, blazing angle, groove shape and material to a specific wavelength/time-scale range makes Smith-Purcell radiation a versatile bunch length diagnostic. Indeed, the RF-accelerator community has demonstrated Smith-Purcell bunch length measurement down to sub-ps (Andrews et al., 2014a,b; Blackmore et al., 2009; Korby et al., 2006), even few-fs (Bartolini et al., 2012) resolution. Extending these methods to plasma-accelerated beams should have high priority in future research.

IV. DIAGNOSTICS OF PLASMA ACCELERATOR STRUCTURES

Diagnostics of electron and positron beams only indirectly characterize the plasma structure that captured and accelerated them. Beam diagnostics alone rarely pinpoint the cause of sub-optimal performance, or provide clear guidance on corrective action. Direct observation of the plasma structure in the laboratory, and comparison of laboratory images with simulations, then become essential. This is challenging for four reasons:

(i) The structures are microscopic. For the $\bar{n}_e$ range of interest for particle acceleration (see Sec. II.A.4) accelerator cavities have overall dimension $100\mu$m $\gtrsim \lambda_p \gtrsim 10\mu$m. Moreover, nonlinear wakes possess sharp micron-scale sub-structures (see Fig. 1). Diagnostic probes must resolve these spatial scales.

(ii) The structures have low optical contrast, for wavelengths $\lambda \ll \lambda_p$ that resolve structural details. For example, for a $\lambda = 1\mu$m optical probe, an evacuated “bubble” in $\bar{n}_e = 10^{18}$ cm$^{-3}$ plasma differs in refractive index by only $\Delta n \approx n_e/2\pi n_{cr} = .0005$ from surrounding plasma.

(iii) The structures propagate at $\sim c$ over mm to meter distances. To minimize image blurring, a probe should co-propagate with the structure. However, even if temporal slip between probe and wake is negligible, wakes can evolve on a ps time scale in ways that are important to diagnose. A co-propagating probe integrates over this evolution. Thus diagnostics must combine longitudinal and transverse probing.

(iv) The structures are transient, and prone to shot-to-shot variation. Single-shot probes are desirable.

These challenges have no counterpart in conventional RF acceleration, which uses macroscopic, stationary, permanent structures. This section reviews plasma wake diagnostics and results that have emerged uniquely with development of plasma-based electron accelerators. They address the above challenges using wide bandwidth op-
TABLE II Properties of plasma wakes, and methods developed to diagnose them. Key to wake properties: Plasma frequency ($\omega_p$), wave-vector ($\vec{k}_p$), wavelength ($\lambda_p$); $\delta n_c(z, \xi, r)$ = wake electron density profile as function of driver propagation distance ($z$), distance behind driver ($\xi$), and distance $r$ from propagation axis; $\vec{E}(\xi, r)$ = wake electric field profile. In column 3, the probe is described as optical (o) or electron (e); propagating longitudinally (L), transversely (T), or obliquely (O) to $\vec{k}_p$; having duration much less than ($<\ll$), less than ($<$), similar to ($\approx$), or greater than ($>\gg$) a plasma period $\tau_p$; and diagnosing the wake in single- (s) or multi- (m) shot.

<table>
<thead>
<tr>
<th>Wake property</th>
<th>Diagnostic method</th>
<th>Probe properties</th>
<th>See...</th>
</tr>
</thead>
<tbody>
<tr>
<td>wave-breaking, $\omega_p, \vec{k}_p$</td>
<td>CTS (Hamster et al., 1994, 1993)</td>
<td>o, L or T</td>
<td>IV.A.2</td>
</tr>
<tr>
<td>$\delta n_c(\xi)$</td>
<td>quasi-static: FDI (Clayton, 2009), FDS* (Siders et al., 2010); sub-$\lambda_p$: FDSC, FDT (Marquès et al., 2014b)</td>
<td>o, L, &lt;, m</td>
<td>IV.B.2</td>
</tr>
<tr>
<td>$\delta n_c(z, \xi, r)$ structures</td>
<td>Faraday rotation (Hamster et al., 2010), transverse sheath shadowgraphy* (Buck et al., 2011)</td>
<td>o, T, &lt;, s</td>
<td>IV.C.3.a</td>
</tr>
<tr>
<td>$\delta n_c(z, \xi, r)$ structures</td>
<td>Faraday rotation (Hamster et al., 2010), transverse sheath shadowgraphy* (Buck et al., 2011)</td>
<td>o, T, &lt;, s</td>
<td>IV.C.3.b</td>
</tr>
<tr>
<td>evolving sub-$\lambda_p$ structures</td>
<td>shadowgraphic movies* (Kaluza et al., 2010), MOPI* (Pauk et al., 2011)</td>
<td>o, O, &lt;, m</td>
<td>IV.D.1</td>
</tr>
<tr>
<td>$\delta n_c(z, \xi, r)$ structures</td>
<td>delayed witness bunch (Sävert et al., 2014a); transverse radiography** (Dong et al., 2014b)</td>
<td>e, L, &lt;, m</td>
<td>IV.B.1</td>
</tr>
<tr>
<td>profiles</td>
<td>longitudinal radiography** (Zhang et al., 2017)</td>
<td>e, L, &gt;, s</td>
<td>IV.C.4</td>
</tr>
</tbody>
</table>

** (Hamster et al., 1994; Helle et al., 2010; Thomas et al., 2007)  
* Collective Thomson Scattering (Clayton, 2009)  
† Frequency-Domain Interferometry (Marquès et al., 1997; Siders et al., 1996b)  
‡ Frequency-Domain Holography (Matlis et al., 2006)  
§ Frequency-Domain Shadowgraphy (Dong et al., 2010b)  
¶ (Kaluza et al., 2010)  
‖ (Buck et al., 2011)  
¶† (Sävert et al., 2015)  
¶‡ Frequency-Domain Streak Camera (Li et al., 2014a)  
¶§ Frequency-Domain Tomography (Li et al., 2014b)  
¶& Multi-Object-Plane Imaging (Li et al., 2013b)  
¶† Applied to PWFA (Kallo et al., 2008; Rosenzweig et al., 1989)  
¶‡ Applied to PWFA (Clayton et al., 2016)  
¶§ Applied to PWFA (Clayton et al., 2016)

A. Light emission and scattering from plasma waves

1. Plasma self-emission

Just as electromagnetic radiation from plasma-accelerated electrons (Sec. III.A) underlies multifarious beam diagnostics (Secs. III.B.3, III.C.2, III.C.3, III.D), electromagnetic emission from the plasma wave itself helps to diagnose wake structure and dynamics. (Hamster et al., 1994, 1993) first detected laser wakefields by observing far-infrared radiation (FIR) that they emit, using a liquid-helium-cooled bolometer in conjunction with a Fourier transform spectrometer. As $\bar{n}_e$ changed from $\sim 10^{15}$ to $\sim 10^{19}$ cm$^{-3}$, FIR frequency closely tracked $\omega_p$ in the few-THz range where it was measurable, while FIR intensity peaked at the resonant condition, which the authors defined as $\omega_p \tau = 2$, corresponding to $n_e \approx 7 \times 10^{16}$ cm$^{-3}$ and laser pulse duration $\tau = 120$ fs. This showed that the FIR originated from collective charge density oscillations of a standard LWFA (see Sec. II.B.1). The observed time duration and angular distribution of FIR qualitatively diagnosed the wake’s lifetime and 3D structure, respectively.

In strongly nonlinear wakes, light emission is useful for diagnosing ultrafast, small-scale phenomena beyond the resolution of impinging probe pulses. One example is “wave-breaking” radiation (Thomas, 2010; Thomas et al., 2007), a broadband light flash accompanying electron injection into a plasma bubble. Used in conjunction with transverse optical probes, it can pinpoint injection spatially and temporally within the larger wake formation and acceleration process. Fig. 35 and accompanying text (Sec. IV.D.1) will show an example. A second example is radiation emitted at the second-harmonic of the drive pulse frequency from the ultra-thin electron sheath surrounding a plasma bubble (see Fig. 1b-d). Here the source current is contained within a region smaller than the emitted wavelength, while fulfilling a Cherenkov-angle condition for the laser second-harmonic (Gordon et al., 2008). The resulting electro-optic shock is emitted in a characteristic ring pattern (Helle et al., 2010).

The distinctive spectral, angular and temporal features of wake self-emission have potential to become quantitative diagnostics with advances in simulation capabilities. Recent work aims to calculate quantitative features of observed emission, such as angle- and time-resolved spectra, within PIC simulations, taking into account all simulated particles (Pausch et al., 2014a,b).

2. Light scattering

Among the earliest diagnostics of plasma accelerator structures were experiments that scattered probe pulses of duration $\tau_p r \gg \omega_p^{-1}$ from laser-driven electron plasma waves (EPWs). The theory (Froula et al., 2011) and ex-
Experimental approach (Clayton, 2009; Villeneuve et al., 1991) resemble those for light scattering from ultrasonic waves in transparent media (Born and Wolf, 1980). In these experiments, a probe pulse of frequency $\omega_{pr} \gg \omega_p$ and wave vector of magnitude $|\vec{k}_{pr}| \gg |\vec{k}_p|$ impinging on a plasma wave (frequency $\omega_p$, wave-vector $\vec{k}_p$) either collinearly ($\vec{k}_{pr}|\vec{k}_p$), transversely ($\vec{k}_{pr} \perp \vec{k}_p$) or obliquely. Uncorrelated individual plasma electrons scatter probe light at frequency $\omega_{pr}$ in a dipole pattern, due to their individual light-driven oscillatory motion, a process known as linear Thomson scattering (Jackson, 1999). However, a collective electron density oscillation $\delta n_e(x,t)$ driven above the level of thermal fluctuations — i.e. $k_p \lambda_D < 1$, where $\lambda_D$ is the plasma Debye length (Froula et al., 2011) — appears to the probe as a refractive index grating

$$\eta(x,t) - \eta_0 = \left\{1 - \left[\frac{\delta n_e(x,t)}{n_{cr}}\right]\right\}^{1/2} - \eta_0 \approx \frac{\delta n_e(x,t)}{2n_{cr}} \tag{43}$$

moving at phase velocity $\omega_p/|\vec{k}_p|$. Here, $\eta_0 = \left[1 - \left(\frac{\delta n_e}{n_{cr}}\right)^2\right]^{1/2}$ is the refractive index of unperturbed plasma, and the second line is valid for $\omega_{pr} \gg \omega_p$. This grating imprints a moving sinusoidal phase-modulation on the probe, resulting in scattered light at frequency $\omega_{pr} \pm \omega_p$ and wave vectors $\vec{k}_{pr} \pm \vec{k}_p$, and over and above the linear Thomson scattering background. This is known as linear collective Thomson scatter (CTS) (Slusher and Surko, 1980; Villeneuve et al., 1991). For $\tau_{pr} \gg \omega_p^{-1}$, the probe bandwidth can be much less than $\omega_p$. If so, the CTS spectrum consists of discrete Stokes/anti-Stokes sidebands well outside the incident probe bandwidth, making it easily distinguishable in a spectrometer from background Thomson-scattered probe light. Moreover in the same limit, a collimated probe has wave-vector spread much smaller than $|\vec{k}_p|$. When such a probe interacts transversely with an EPW, CTS light scatters at discrete angles $\theta \approx |\vec{k}_{pr}|/|\vec{k}_p|$ (typically a few degrees) from $\vec{k}_{pr}$, well outside the transmitted probe diffraction cone, providing additional spatial discrimination. Thus frequencies $\omega_{pr} \pm \omega_p$ and wave-vectors $\vec{k}_{pr} \pm \vec{k}_p$ of CTS light are the main observables. Analysis yields the frequency $\omega_p$, wave vector $\vec{k}_p$, and local amplitude $\delta n_e$ of the EPW. When $\tau_{pr}$ is less than the EPW lifetime, the time evolution of $\delta n_e$ behind the driver can also be measured by varying pump-probe delay $\Delta t$ (Le Blanc et al., 1996; Ting et al., 1996). Nonlinear EPWs generate harmonics $\pm m\omega_p$ ($\pm m\vec{k}_p$), where $m = 1, 2, 3, ...$, of the frequency-(wave-vector) shift of scattered light as an additional diagnostic (Everett et al., 1995). When one or more harmonics is (are) present, analysis of harmonic ratios (Lal et al., 1997) enables more accurate estimates of absolute $\delta n_e$ than estimates based on $m = \pm 1$ sidebands alone, which are subject to uncertainties in estimating propagation distance and transverse dimension of the EPW. Light scattering experiments, however, do not resolve internal sub-$\lambda_p$ structure of the EPW. (Clayton, 2009) has reviewed light scattering from plasma accelerator structures in detail. Thus here we only discuss two examples.

**FIG. 21** Color online. Collective Thomson scatter from self-modulated LWFA. (a) Left: schematic of EPW (blue rectangle) and EPW light scattered in direction $\vec{k}_{pr} + \vec{k}_p$ (3.4°±1.9° from $\vec{k}_p$) from EPW localized at $-2 \text{mm} < z < -1 \text{mm}$, acquired in one shot along with background unshifted probe light ($\Delta \lambda = 0$, attenuated 4×). Inset: $x$-lineout of anti-Stokes scattered light, and thus of EPW amplitude. Panel b) adapted from (Gordon et al., 1998).

Figure 21a (left) schematically illustrates experiments by (Le Blanc et al., 1996; Ting et al., 1996) in which a sub-ps probe pulse ($\lambda_{pr} = 0.53\mu m$, green rectangle) copropagated at delay $\Delta t$ behind a drive pulse (peak power $P = 1 - 2 \text{TW}$, $\tau = 0.4 \text{ps}$, $\lambda_{pu} = 1.053\mu m$, red rectangle) in a gas jet (grey) that the drive pulse ionized to density $0.75 \times 10^{18} \text{ cm}^{-3} < \bar{n}_e < 3 \times 10^{18} \text{ cm}^{-3}$ (Le Blanc...
et al., 1996) or $\eta_e \approx 10^{19}$ cm$^{-3}$ (Ting et al., 1996). When $P$ exceeded the critical power $P_{cr} = 17(\omega^2/\omega_p^2)$ GW for relativistic self-focusing, the drive pulse efficiently drove a self-modulated (SM) LWFA (blue rectangle, see Sec. II.B.2). Under conditions of (Le Blanc et al., 1996), the wake produced a strong collimated $\sim 2$ MeV electron beam (Umstadter et al., 1996). In both experiments, the probe pulse overlapped 10-20 periods of this wake. A spectrometer, aided by a notch filter, isolated and recorded spectra of the transmitted wake-modulated probe as $\Delta t$ varied over multiple shots. Results showed $\Delta t$-dependent Stokes and anti-Stokes sidebands (Fig. 21a, right), with amplitudes proportional to $\delta n_e(\Delta t)$. Using this data, (Le Blanc et al., 1996) measured EPW growth rate: side-bands appeared at $\Delta t \approx -0.5$ ps, then grew to a maximum corresponding to $\delta n_e/\eta_e \sim 0.1$ at $\Delta t \approx 0.3$ ps. This wake, however, started earlier, and grew more slowly, than a 2D theory of forward Raman scattering instability in uniform pre-formed plasma predicted (Decker et al., 1996; Tseng et al., 1996). (Le Blanc et al., 1996) thus concluded that plasma noise generated at the ionization front ($\Delta t \approx -0.7$ ps) seeded the instability earlier than expected, and that a 3D theory (Andreev et al., 1995; Esarey et al., 1994) was required to explain its growth.

Both (Le Blanc et al., 1996) and (Ting et al., 1996) measured EPW decay rate, with different results. (Le Blanc et al., 1996) found that sidebands (and thus $\delta n_e$) decayed to undetectable levels within 1.5 ps (Fig. 21a, right). They attributed the rapid decay to efficient conversion of collective EPW energy into electron beam energy. (Ting et al., 1996), on the other hand, working at a higher $\eta_e$ with less efficient electron beam production, found that probe sidebands persisted out to $\Delta t \approx 5 - 7$ ps. They explained the slower decay rate via conversion of the EPW into ion acoustic waves (IAWs), a subject of considerable prior theoretical work (Mora et al., 1988; Zakharov, 1972). They confirmed this by observing a sharp increase in scatter of the central probe spectral peak at angle $\sim 30^\circ$ from the pump-probe propagation axis over time interval $5 \lesssim \Delta t \lesssim 30$ ps, correlated with decay of Stokes/anti-Stokes sidebands. They attributed this rising signal to CTS from growing EPW-fed, slow IAWs, which they observed to decay subsequently over 30 ps $\lesssim \Delta t \lesssim 100$ ps. The physics of EPW $\rightarrow$ IAW conversion is re-surfacing in recent theoretical (Lotov et al., 2014; Sahai et al., 2016) and experimental (Zgadzaj et al., 2016) wake diagnostic work because it is a strong plasma heating mechanism (see Sec. IV.D.3).

(Filip et al., 2004) collinearly probed PBWAs (see Sec. ??) driven by a 2-color CO$_2$ laser in plasma of density $10^{16}$ cm$^{-3}$ $\lesssim \eta_e \lesssim 10^{17}$ cm$^{-3}$. At such low $\eta_e$, special techniques were required to distinguish Stokes, anti-Stokes sidebands from the unshifted probe spectrum (Filip et al., 2003). CTS showed that high $\delta n_e$ wakes were achievable even when the beat frequency was far from $\omega_p$. With emergence of TW, ps CO$_2$ lasers (Polyanskiy et al., 2015), researchers are now also beginning to probe CO$_2$-laser-driven SM-LWFA, first studied theoretically by (Andreev et al., 2003), using collinear CTS.

Figure 21b (left) schematically illustrates a complementary “transverse” CTS geometry in which a probe pulse crossed the path of a EPW of transverse width $\lesssim \lambda_p$ at $\sim 90^\circ$ (i.e. $\vec{k}_{pr} \perp \vec{k}_p$). This geometry has characterized PBWAs (Clayton et al., 1993) and decay of laser-driven EPW into a manifold of daughter waves (Everett et al., 1995). Fig. 21b (right) shows an example of transverse CTS data (Gordon et al., 1998) from a SM-LWFA that produced copious collimated electrons up to 94 MeV energy when driven in $\eta_e = 1.4 \times 10^{19}$ cm$^{-3}$ plasma by tightly focused drive pulses of similar wavelength and duration ($\lambda_p = 1.053\mu m$, $\tau = 1$ ps), but higher power ($P \approx 20$ TW), than those used in the experiments of (Le Blanc et al., 1996; Ting et al., 1996). The duration ($\tau_{pr} \approx 20$ ps) and radius ($w_0 \approx 0.5$ cm) of the transverse probe pulse ($\lambda_{pr} = 0.53\mu m$) were adjusted so that it illuminated the wake during its entire transit through the $\sim 0.1$ cm long gas jet. (Gordon et al., 1998) used anti-Stokes ($\Delta \lambda \approx -35$ nm) light scattered at $\vec{k}_{pr} + \vec{k}_p$ ($\sim 3.4^\circ$ from $\vec{k}_{pr}$) to image the EPW propagation path onto the slit of an imaging spectrometer. In Fig. 21b (right), this anti-Stokes image is visible at $\Delta \lambda \approx -35$ nm, from only a $\sim 1$ mm section ($-2 \text{ mm} < z < -1 \text{ mm}$) of the gas jet, along with background unshifted probe light ($\Delta \lambda = 0$) that scattered from the jet’s entire length. This showed that the wake had significant amplitude only in this 1mm section, because of tight pump focus there. For looser pump foci, on other other hand, wakes persisted for longer distances (up to the entire jet), but yielded lower energy electrons. Transverse CTS provides this information about the wake’s trajectory in a single shot, information unavailable from co-propagating CTS because it integrates over the main pulse’s propagation direction. The complementarity of diagnostic information available from co-propagating (Sec. IV.B, C.1.2) vs. transverse (or oblique) probes (Sec. IV.C.3,4, D,E) is a recurring theme throughout the remainder of this section.

B. Multi-shot sub-$\lambda_p$ probes

Optical diagnosis of sub-$\lambda_p$ structure of plasma accelerators requires probes of bandwidth $\Delta \omega_{pr} > \omega_p$. This precludes CTS, since Stokes and anti-Stokes shifts would be less than $\Delta \omega_{pr}$. In this sub-section, we describe experiments in which optical or electron probes of duration $\tau_{pr} < \omega_p^{-1}$ co-propagate behind the wake driver with controlled time delay $\Delta t$. The probe overlaps a sub-$\lambda_p$ longitudinal slice of the structure, which modifies the probe. Photon frequency or electron energy analysis, or optical interferometry detects these modifications. In addition, the wake’s transverse fields or density gradients deflect
probe electrons or photons sideways. These transverse probe profile modifications are also detected downstream of the accelerator. Thus by varying $\Delta t$ over multiple shots, these methods can map out both longitudinal and transverse wake structures with sub-$\lambda_p$ resolution, if the structure is stable from shot to shot. The co-propagating geometry also maximizes the probe’s interaction length with the slice of the structure that it overlaps, thus optimizing sensitivity to low-contrast sub-structures.

1. Electron witness bunches

Beam-driven PWFAs (Chen et al., 1985) were not only the first plasma-based electron accelerators to be demonstrated in the laboratory, but the first to be spatially and temporally diagnosed. The Advanced Accelerator Test Facility at Argonne National Laboratory (ANL) measured wake fields of an electron drive bunch ($E_d = 21$ MeV, 2 nC, $Q \leq 4$ nC, $\tau_b = 16$ ps) through their effect on a 15 MeV “witness” bunch created by degrading the energy of part of the drive bunches with a carbon target, and splitting them away with a dipole magnet. Separate beam transport lines delivered the two synchronized bunches colinearly to the plasma, the witness leg containing an adjustable “trombone” section that varied witness delay $\Delta t$ (see Fig. 22a). A dipole magnet spectrometer after the plasma measured witness bunch energy with $\sim .01$ MeV resolution, and transverse witness bunch deflection perpendicular to the spectrometer dispersion plane, as $\Delta t$ varied. (Rosenzweig et al., 1988, 1989) used this facility to drive and probe wakes in $\sim 30$ cm-long plasma of density $n_e \approx 10^{13}$ cm$^{-3}$. Available bunch duration $\tau_b \sim 16$ ps (FWHM) dictated this choice of $n_e$, which enabled these bunches to resonantly drive and resolve individual plasma oscillations of period $\omega_p^{-1} \sim 30$ ps (see Fig. 22b,c).

Initial experiments drove the plasma with 2 nC bunches of longitudinal ($s$) and radial ($r$) half-widths $\sigma_s \approx \sigma_r \approx 2.1$ mm, yielding bunch density $n_b \approx .015n_e$ (Rosenzweig et al., 1988). Since $n_b \ll n_e$, these excited linear wakes with accelerating field $E_d^{(max)} < E_0 \sim 300$ MV/m (see Eq. 4a) and density perturbation $\delta n_e \ll n_e$ (Lu et al., 2005). These wakes modulated witness bunch centroid energy $E_c$ sinusoidally as a function of $\Delta t$, with amplitude $\Delta E_c^{(max)} \approx \pm .05$ MeV. The quotient of $\Delta E_c^{(max)}$ and acceleration length yielded effective gradient $E_z^{(eff)} \sim .05$ MeV/0.3 m $\approx 0.16$ MV/m. However, fits of the data to a 2D linear theory of plasma wakes (Chen, 1985) showed that fields as high as $E_z^{(max)} \approx 1$ MV/m $\approx .03E_0$ were generated. Radial averaging over wake and witness bunch profiles accounted for the discrepancy.

In follow-up experiments, (Rosenzweig et al., 1989) delivered 5-fold denser bunches (4 nC, $\sigma_r \approx 1.4$ mm) to the plasma. These self-pinched, increasing peak current further. They excited nonlinear wakes in which $E_z^s (z - ct)$ oscillated in a sawtooth waveform (see Fig. 22b). Fourier analysis showed that the waveform contained harmonics of $\omega_p$, as expected for a nonlinear wake. Transverse deflections of the witness bunch also oscillated with $\Delta t$ in phase with $E_z$ (see Fig. 22c), simultaneously profiling radial fields $E_r^s(z - ct)$. 2D modeling suggested peak $E_z$ up to 5 MV/m was achieved, with $E_z^{(eff)}$ smaller as before. Two features of these results were surprising. First, wake oscillations, despite their distinct nonlinearity, were observed out to 18 periods behind the driver with little degradation in form or amplitude. Second, the relativistic increase in plasma period $\sqrt{\gamma}$ $\omega_p^{-1}$ expected (Rosenzweig, 1987) for such steepened plasma waves was not observed. Instead a slight decrease in period was observed as wake amplitude increased. Later, (Marqués et al., 1997) re-visited, and partially explained, similar features of nonlinear laser-driven plasma wakes (see Sec IV.B.2). Still later, (Matlis et al., 2006), in a study of high-amplitude laser-driven wakes (see Sec IV.C.1), finally observed the expected relativistic increase in $\omega_p^{-1}$.

Continuing experiments (Barov et al., 2000) scaled bunch density into the strongly nonlinear “blowout” regime ($n_b > n_e$), discussed in Secs II.A.2 and II.A.3.

**FIG.** 22 Color online. Measurements of local electric field $\vec{E}(r, z - ct)$ of e-bunch-generated nonlinear plasma wave in $n_e = 0.7 \times 10^{13}$ cm$^{-3}$ plasma. (a) Schematic experimental setup. Carbon target decreased energy of part of incoming 21 MeV electron drive bunches (left) to create 15 MeV witness bunch, which dipole magnets (cross-hatched boxes) separated, guided through paths of variable relative length, and recombined collinearly with controlled time delay $\Delta t$. Upper right inset: schematic of plasma wave potential (dashed curve) and right-propagating drive (solid red, right) and witness (solid blue, left) bunches. Adapted from (Rosenzweig et al., 1988). (b) Energy change $\Delta E$ and (c) transverse deflection $\Delta y$ perpendicular to energy dispersion plane of witness bunch vs. $\Delta t$. Panels b) and c) adapted from (Rosenzweig et al., 1989).
This work produced drive and witness bunches more compactly by exciting a photocathode with tandem laser pulses. However, they were higher in emittance, longer, and less reproducible than expected. Thus, although this work observed $E_z^{(eff)} \sim 25\, \text{MV/m}$, and inferred $E_z \sim 60\, \text{MV/m}$, amongst its major conclusions was the need for improved methods of generating drive and witness bunches. Simulations by (Serafini, 1996) analyzed the work observed and inferred maximum witness bunch energy gain $\Delta W_e \approx 0.9\, \text{MeV}$ at $\bar{n}_e \approx 10^{16}\, \text{cm}^{-3}$, corresponding to $\Delta t = 1.5\lambda_e/c$. This corresponded to $E_z^{(eff)} = 0.9\, \text{MeV}/0.06\, \text{m} = 150\, \text{MV/m}$, which matched the peak simulated $E_z$. Thus the witness bunch was sufficiently compressed and focused to observe the maximum $E_z$ directly, without mathematical deconvolution of radial averaging.

(Muggli et al., 2008a) refined the bunch-splitting technique by inserting a mask at a position in the dogleg where the beam’s energy was transversely chirped. The incident bunch could then be split into a train of sub-ps micro-bunches of controllable number, length and spacing by adjusting beam and mask parameters. An analogous technique generates controlled trains of ultrashort optical pulses (Weiner, 2000). This enabled production not only of a witness bunch, but of a train of drive bunches. Strategic adjustment of their spacing, shape and charge can increase transformer ratio (Laziev et al., 1988) and energy extraction efficiency (Maeda et al., 2004) of a PWFA, in theory, by more than an order of magnitude (Farmer et al., 2015; Nakajima, 1989) compared to a single-bunch driver (Lotov, 2013; Ruth et al., 1984). (Muggli et al., 2011) demonstrated acceleration of a witness bunch in a wake driven by two mask-generated drive bunches. However, experiments have not yet realized full predicted capabilities of the multi-bunch PWFA. Direct 2D mapping of a multi-bunch PWFA with an electron or optical witness bunch as drive parameters change is a promising future diagnostic experiment.

Recent experiments at SLAC’s Facility for Advanced Accelerator Experimental Tests (FACET) (Hogan et al., 2010) used co-propagating electrons to diagnose internal fields of strongly blown out PWFAs in $\bar{n}_e \approx 2 \times 10^{17}\, \text{cm}^{-3}$ plasma. (Clayton et al., 2016) mapped these fields in one shot, using electrons in the trailing portion of a drive bunch itself as witnesses. This is discussed with other single-shot experiments in Sec. IV.C.4. When the charge of a separate witness bunch becomes large enough to perturb the wake in which it is accelerating, it ceases to act purely as a diagnostic “witness” bunch. Such “beam loading” becomes beneficial when the accelerating bunch flattens local gradients in the wake field, helping the bunch to accelerate mono-energetically (Lu et al., 2006; Tzoufras et al., 2008). Recent PWFA experiments realized this beam-loaded regime, and imparted energy gains of several GeV to high-charge electron (Litos et al., 2014) and positron (Corde et al., 2015) bunches. 2D optical profiling of beam-loaded PWFAs is a promising future diagnostic experiment (Zgadzaj et al., 2016).

![FIG. 23 Color online. Measurements of peak longitudinal electric field $E_z$ of e-bunch-generated linear plasma wave in $\bar{n}_e = 10^{16}\, \text{cm}^{-3}$ plasma. (a) Energy gain of witness bunch at fixed $\Delta t = 500\, \text{fs}$ behind drive bunch as $\bar{n}_e$ changes. Inset: simulated plasma wave potential (dashed blue) and right-propagating drive and witness bunches (solid green). (b) Measured and (c) simulated energy spectra of witness and drive bunches without (dashed blue) and with (solid red) plasma of density $\bar{n}_e = 10^{16}\, \text{cm}^{-3}$. Adapted from (Kallos et al., 2008).](image-url)
2. Laser probe pulses

Delayed, co-propagating electromagnetic probe pulses of duration \( \tau_{pr} < \omega_p^{-1} \) have also diagnosed plasma wake structure. For laser-driven wakes, inexpensive beam-splitters trivially separate perfectly synchronized probe(s) from the driver while preserving durations \( \tau \lesssim 30 \, \text{fs} \). For beam-driven wakes, state-of-the-art electronic techniques can synchronize an e-beam and independent laser probe with < 1 fs jitter (Xin et al., 2017).

The direct optical analog of e-beam diagnostics described in Sec. IV.B.1 is “photon acceleration” (Bulanov et al., 1993; Esarey et al., 1990; Wilks et al., 1989) – i.e., blue- or red-shifts \( \Delta \omega_{pr} \) of probe pulses that co-propagated with a longitudinal slice of the wake at which \( n_e \) locally increases \( (dn_e/d\Delta t > 0) \) or decreases \( (dn_e/d\Delta t < 0) \), respectively, with increasing \( \Delta t \). From Poisson’s equation, maximum \( |dn_e/d\Delta t| \) correspond to strongest \( |E_z| \). For co-propagation distance \( L \),

\[
\Delta \omega_{pr}(r, \Delta t) = -\left( \omega_{pr}/c \right) \int_0^L (dn_e/d\Delta t) \, dz \quad (44a)
\]

\[
\approx (\omega_{pr} L/2n_{cr}c) \, dn_e/d\Delta t. \quad (44b)
\]

Here, \( \eta \) denotes local refractive index \( \eta(r, \Delta t; z) = (1 - n_e(r, \Delta t; z)/n_{cr})^{1/2} \), and (44b) holds when \( dn_e/d\Delta t \) is \( z \)-independent and \( n_e \ll n_{cr} \). Both expressions assume \( L \) is short enough that the probe remains collimated. Thus a longitudinal wake slice that maximally accelerates an electron also maximally blue-shifts a probe pulse. Multi-shot pump-probe blue-shift experiments have diagnosed ionization front structure in atmospheric density gases (Wood et al., 1991). However, despite in-depth theoretical analyses (Dias et al., 1998; Kasim et al., 2015) and initial experiments (Trines et al., 2009), photon acceleration has not yet probed detailed plasma wake structure \( n_e(r, \Delta t) \). This can be attributed to the small magnitude of \( |dn_e/d\Delta t| \) in linear wakes in sub-atmospheric density gases, and to the wide bandwidth of probe pulses capable of resolving sub-\( \lambda_p \) features, making subtle spectral centroid shifts difficult to observe. Photon deceleration (red-shift) of wakefield drive pulses, which characterizes their energy transfer to plasma waves, has been observed (Murphy et al., 2006; Shiraishi et al., 2013), but does not measure detailed wake structure.

Researchers have had greater success diagnosing wake structure by analyzing phase shift \( \Delta \phi_{pr}(r, \Delta t) \) that a wake imprints on a co-propagating probe. Unlike \( \Delta \omega_{pr} \), \( \Delta \phi_{pr} \) can be measured interferometrically with high accuracy even for a wide-bandwidth probe. A collimated probe of duration \( \tau_{pr} < \omega_p^{-1} \) experiences phase shift

\[
\Delta \phi_{pr}(r, \Delta t) = \left( \omega_{pr}/c \right) \int_0^L \eta(r, \Delta t; z) \, dz \quad (45a)
\]

\[
\approx (\omega_{pr} L/2n_{cr}c) n_e(r, \Delta t), \quad (45b)
\]

uniformly over its longitudinal profile. Here, (45b) holds in the same limit as (44b). Thus \( \Delta \phi_{pr}(r, \Delta t) \) is proportional to the local density \( n_e(r, \Delta t) \), rather than the local field \( E_z(r, \Delta t) \), at which the probe propagates. By varying \( \Delta t \) and imaging the transverse probe profile, \( n_e(r, \Delta t) \) can be mapped over multiple shots.

**FIG. 24** Color online. Measurements of laser-generated electron density waves \( n_e(r, z - ct) \) in \( 25 < \bar{n}_e < 3 \times 10^{17} \, \text{cm}^{-3} \) plasma. (a) Schematic frequency-domain interferometry (FDI) setup: BS = beam splitter; SHG = second-harmonic generation; MI = Michelson interferometer; DM = dichroic mirror; L = lens; SPEC = spectrometer; CCD = charge-coupled device. Top-center inset: schematic of plasma wave density (dashed curve) created by right-propagating drive laser pulse (tan, red), with probe (trailing) and reference (leading) diagnostic pulses (short, blue). Top-right inset: schematic FDI interferogram. Adapted from (Marquès et al., 1996). (b) Optical probe phase change \( \Delta \phi_{pr}(\Delta t) \) vs. pump-probe delay \( \Delta t \) for fully-ionized \( \bar{n}_e = 0 \) (top), 1.7 (middle) and 3 (bottom) \( \times 10^{17} \, \text{cm}^{-3} \) He plasmas excited by 100 fs, 0.8\( \mu \)m, 10 mJ laser pulses focused with f/4. Adapted from (Siders et al., 1996b). 4.8 (2.7) Torr data was acquired with standard (differential) FDI (c) 2D map \( \Delta \phi_{pr}(r, \Delta t) \) of left-propagating nonlinear plasma wave for \( \bar{n}_e = 0.25 \times 10^{17} \, \text{cm}^{-3} \). Panel c) adapted from (Marquès et al., 1997).

Frequency-domain interferometry (FDI) has mapped \( \Delta \phi_{pr}(r, \Delta t) \) in a variety of multi-shot pump-probe experiments (Geindre et al., 1994; Reynaud et al., 1989; Tokunaga et al., 1992). Fig. 24a shows a schematic FDI
setup. A diagnostic pulse is split from the drive pulse, then shifted in frequency, rotated in polarization, or both to help discriminate it from scattered pump light after the interaction. The diagnostic pulse is sub-divided into a reference pulse $E_{ref}^{(0)}(t)$, and a probe pulse $E_{pr}^{(0)}(t-T)$ that trails the reference by time $T$ (e.g. by a Michelson interferometer with unequal arm lengths). These recombine with the pump (e.g. at a dichroic or polarizing mirror), and all three pulses co-propagate through the interaction region without overlap.

Different 3-pulse sequences can be used. In “standard” FDI, $E_{ref}^{(0)}(t)$ leads, and $E_{pr}^{(0)}(t-T)$ trails, the pump with $T$ fixed (see Fig. 24a, center inset). The pump-induced index change $\eta(r, \Delta t)$ then affects only the probe, phase-shifting it to $E_{pr}(t-T) = E_{pr}^{(0)}(t-T)e^{i\Delta \phi_{pr}(r,\Delta t)}$. This configuration has two limits. First, the requirement that the reference lead the pump limits probe-pump delay to $0 < \Delta t < T$. Second, if the pump ionizes neutral gas before generating a wake, both static plasma (density $\bar{n}_e$) and wake oscillations (amplitude $\delta n_e$) contribute to $\eta(r, \Delta t)$. But if $\delta n_e \ll \bar{n}_e$, and $\bar{n}_e L$ fluctuates from shot to shot, the fluctuations can mask the wake. “Differential” FDI, in which reference and probe both trail the pump, separated by a half-integer number of plasma periods, can then determine the $\Delta t$ range and discriminate small wake oscillations more effectively (Marquès et al., 1997, 1996; Siders et al., 1996b). In this configuration, the reference pulse shifts to $E_{ref}(t) = E_{ref}^{(0)}(t)e^{i\Delta \phi_{ref}(r,\Delta t)}$, and the relative phase shift $\Delta \phi \equiv \Delta \phi_{pr} - \Delta \phi_{ref}$ is measured. The static contribution to $\Delta \phi$ cancels, while the oscillatory component doubles in amplitude compared to $\Delta \phi_{pr}$ in the standard configuration. In both configurations, lens L2 (see Fig. 24a) images $E_{ref}(r,t)$ and $E_{pr}(r,t-T)$ to the entrance slit of an imaging spectrometer. The slit selects a slice along a direction hereafter called “y”. Spectral dispersion temporally broadens both pulses, causing them to overlap at the spectrometer’s array detector, which records their combined spectral intensity

$$I(y, \omega) = |F[E_{ref}(y,t) + E_{pr}(y,t-T)]|^2 \approx |E_{ref}(y,\omega) + E_{pr}(y,\omega)e^{-i\omega T}|^2$$

(46a)

$$= S(y,\omega) + E_{ref}(y,\omega)E_{pr}^*(y,\omega)e^{i\omega T} + E_{pr}(y,\omega)E_{ref}(y,\omega)e^{-i\omega T}.$$  

(46b)

Here, $F$ denotes Fourier transform, $S(y,\omega) = |E_{ref}(y,\omega)|^2$  +  $|E_{pr}(y,\omega)|^2$, and $E_{ref}(y,\omega) = E_{ref}^{(0)}(y,\omega)e^{i\Delta \phi_{ref}(y,\omega)}$ (j = ref, pr) are perturbed FD fields expressed in terms of unperturbed fields $E_{j}^{(0)}(y,\omega)$. For $E_{ref}^{(0)} = E_{pr}^{(0)} = E$ with no losses, refraction or diffraction in the plasma, Eq. (46) becomes (Tokunaga et al., 1992)

$$I(y, \omega) = 2 |E(y,\omega)|^2 \{1 + \cos[\omega T - \Delta \phi(y,\omega)]\}.$$  

(46c)

Eq. (46c), an oscillating function of frequency with period $2\pi/T$, is the “frequency-domain interferogram” that encodes probe-reference phase shifts (Reynaud et al., 1989). In the absence of an interaction ($\Delta \phi = 0$), straight (y-independent) fringes appear on the detector. In the presence of a y-dependent interaction, fringes distort in pump-excited y-regions (see Fig. 24a, right inset). Straight fringes in unexcited y-regions then serve as the reference null interferogram, from which the fringe shift $\Delta \phi(y,\omega)$ in excited y-regions is extracted for each $\Delta t$. If $\tau_{pr} \ll \omega_{pr}^{-1}$, then $n_e(y,\Delta t)$, $\eta(y,\Delta t)$ and $\Delta \phi(y,\Delta t)$ are constant in time at each $y$ over the probe longitudinal profile. If, in addition, probe bandwidth $\Delta \omega_{pr} \ll \omega_{pr}$, then $\eta(y,\Delta t)$ and $\Delta \phi(y,\Delta t)$ are also constant in frequency over the probe bandwidth. In this “FDI approximation”, $\Delta \phi(y,\omega) = \Delta \phi(y,\Delta t)$, both quantities being frequency-independent at each $y$ and $\Delta t$. All fringes at each $y$ and $\Delta t$ then shift by the same amount, and the distinction between spectral and temporal phase disappears. Eq. (46c) then becomes

$$I(y, \omega) = 2 |E(y,\omega)|^2 \{1 + \cos[\omega T - \Delta \phi(y,\omega)]\}.$$  

(46d)

$\Delta \phi(y,\Delta t)$ can now be extracted directly from the measured fringe shift between “signal” and null interferograms. Accurate extraction requires only that the pixel density of the spectrometer’s 2D array detector be high enough for a given $T$ (typically $\geq 10$ pixels per fringe) to resolve the fringe shift. The FDI approximation remains valid even when the Taylor expansion $\Delta \phi(y,\omega) \approx \Delta \phi(y,\Delta t) + \partial_e \Delta \phi(y,t)|_{e=\Delta t}$ (t-T) includes linear temporal variations within the probe longitudinal profile, since $\partial_e \Delta \phi(y,t)|_{e=\Delta t}$ is simply an overall probe centroid frequency shift (photon acceleration) (Siders et al., 1996a). Only when quadratic and higher-order temporal variations within the probe become significant is the equivalence of $\Delta \phi(y,\omega)$ and $\Delta \phi(y,\Delta t)$ lost, necessitating more sophisticated Fourier analysis of the interferograms. They then become FD “holograms” (see Sec. IV.C).

The first FDI experiments to characterize plasma wakes (Marquès et al., 1996; Siders et al., 1996b) used first-generation sub-TW Ti:S ($\lambda = 0.8\mu m$) chirped-pulse amplified (CPA) laser technology (Backus et al., 1998), for which pulse duration was limited to $\tau \sim 100$ fs, pump energy to 10 mJ $\lesssim \mathcal{E} \lesssim 30$ mJ. This $\tau$ limited $n_e$ to $\lesssim 3 \times 10^{17}$ cm$^{-3}$ ($\lambda_p$ to $\geq 60 \mu m$), in order to excite wakes resonantly and resolve sub-$\lambda_p$ features, while this $\mathcal{E}$ necessitated focusing to strongly sub-$\lambda_p$ spot sizes ($3.5\mu m < \omega_0 < 6\mu m$) to reach field strengths ($0.35 < \alpha_0 < 0.5$) sufficient to excite observable wakes. An advantage of such tight focus was that the drive pulse’s radial ponderomotive force dominated wake excitation, producing larger $\delta n_e/\bar{n}_e$ (up to $\sim 1$) and $\Delta \phi_{pr}$ (up to $\sim 30$ mrad) than its longitudinal ponderomotive force alone would have produced. Eq. (3b) quantifies this advantage. On axis ($r = 0$), the ratio $\delta n_e/\bar{n}_e = (\lambda_p/\pi\omega_0)^2$ of radial to longitudinal wake contributions ranged from 30 to 300 in the experiments of (Marquès et al., 1996; Siders et al., 1996b). Moreover, the contribution of the longitudinal
(radial) component of $\delta n_e$ to $\Delta \phi_{pr}$ is proportional to $\delta n_e z_R$ ($\delta n_e z_R$). From Eq. (3b), $\delta n_e z_R$ is independent of pump focus, whereas $\delta n_e z_R$ increases in proportion to $1/z_R$ as focus tightens. Thus with available lasers, tight focus was critical to initial FDI wake observation.

A disadvantage of tight focus was that the probe pulse averaged over radial wake profiles as it transited the interaction region. For example, the data in Fig. 24b correctly shows $\Delta \phi_{pr}(\Delta t)$ oscillating longitudinally at $\omega_p$ for each of two $\bar{n}_e$, but does not directly convey radial wake structure (Siders et al., 1996b). Instead these researchers inferred radial structure indirectly by calculating it via Eq. (3b), computing $\Delta \phi_{pr}(\Delta t)$ induced on a focused (Siders et al., 1996b) or collimated (Marqués et al., 1996) probe, and confirming that the computed $\Delta \phi_{pr}(\Delta t)$ oscillation amplitude agreed with the measured value. The relationship between $\Delta \phi_{pr}(\Delta t)$ and $\delta n_e$ was thus more complicated than Eq. (45b), which assumed longitudinally invariant wake structure and collimated probe. Limited direct radial information was obtained by imaging the transverse profile of $\Delta \phi_{pr}(y, \Delta t)$ with the probe at a peak or valley (Marqués et al., 1996), or by observing increased $\Delta \phi_{pr}(\Delta t)$ as the spectrometer slit narrowed (Siders et al., 1996b).

In a follow-up study, (Marqués et al., 1998, 1997) mapped $\Delta \phi_{pr}(y, \Delta t)$ behind a tightly focused drive pulse in greater detail. Fig. 24c illustrates these expanded results. The left part (“1”) of the image, acquired with standard FDI, shows $\Delta \phi_{pr}(y, \Delta t)$ primarily from plasma formation. Regions of doubly- (red) and singly-ionized (green) He, integrated along the laser axis, can be seen. The right part (“2”) of the image, acquired with differential FDI, shows only wake oscillations, which are localized within the doubly-ionized region. The axial lineout below this image shows damped oscillation of the form $\Delta \phi_{pr} \propto \exp(-\gamma \Delta t) \sin(\omega_p \Delta t)$ behind the pump.

Analysis led to two new discoveries. First, $\omega_p$ of the first few oscillations was $\sim 5\%$ higher than later oscillations, the frequency of which matched $\omega_p$ for doubly-ionized helium. This finding resembled the temporary increase in $\omega_p$ that (Rosenzweig et al., 1989) reported for a nonlinear PWFA (Sec. IV.B.1). Plasma heating, which increases electron thermal velocity $v_{th}$ and thus plasma frequency via $[\omega_p(k)]^2 = \omega_p^2 + 3k^2 v_{th}^2$, could account for $< 1\%$ increase in $\omega_p$, and would not relax within a few oscillation periods. Relativistic electron mass increase in high-amplitude electron oscillations, if present, would decrease, rather than increase, $\omega_p$ (Rosenbluth and Liu, 1972). Instead a simple electrostatic mechanism unique to high-amplitude radial wakes appeared to dominate: the radial displacement $\delta r$ of an electron away from its initial position $r_0$ produced greater charge density at the center of symmetry, and thus a stronger restoring force, than in a planar wake, resulting in fractional increase $\Delta \phi_{pr}/\omega_p \approx (\delta r/r_0)^2/12$ in $\omega_p$ (Dawson, 1959). Computer simulations reproduced the observed increase, as well as its temporal relaxation. Second, damping rate $\gamma$ was faster than expected from mechanisms expected in a uniform plasma — e.g. fine-scale mixing of orbits of electrons with amplitude-dependent oscillation frequencies (Dawson, 1959), or thermal convection. Instead radial electron excursions that cross the $\text{He}^{2+}/\text{He}^+$ boundary and de-phase from the wake provided the best quantitative explanation. These discoveries illustrate how high-resolution wake structure diagnostics, in concert with simulations, advance plasma wakefield physics.

C. “Snapshots” of wake structures

Multi-shot techniques diagnose wakefield structure in microscopic detail, but require lengthy data acquisition, and thus cannot provide rapid feedback. Moreover, the data is subject to shot-to-shot fluctuations, especially when wakes are excited nonuniformly. These considerations motivated development of diagnostics that recover plasma wake structure with equivalent detail in a single shot. Single-shot diagnostics developed most rapidly with optical, rather than electron, probes because of the relative ease with which optical pulses can be stretched, compressed, expanded and imaged and with which their internal phase and amplitude structure can be measured. Nevertheless, single-shot, high-resolution electron radiography of internal fields of plasma-based electron accelerators was recently demonstrated (see Sec. IV.C.4).

The development of optical single-shot plasma wake diagnostics drew upon three established optical technologies, which have no counterpart in particle beam technology. The first of these was holography. Holographic wake diagnostics drew from conventional holography the concepts of a coherent “object” (or probe) pulse that illuminates the entire object of interest simultaneously, and of a mutually coherent “reference” pulse with which the “object” pulse interferes on a recording medium. New challenges for holographic wake diagnostics included adapting these concepts to a near light-speed object, and developing methods for “reading” the “hologram” quickly and accurately (Sec. IV.C.1). The second underlying technology was ultrafast optics. By the year 2000, chirped-pulse-amplified (CPA) lasers producing terawatt pulses of tens of fs duration were widely available, and techniques for stretching, manipulating and re-compressing such pulses while maintaining their coherence had matured (Backus et al., 1998). Moreover, pulse retrieval algorithms such as frequency-resolved optical gating (Trebin et al., 1997) and spectral shearing interferometry (Iaconis and Walmsey, 1998) for characterizing the internal amplitude and phase structure of individual laser pulses had been developed, and advanced to single-shot implementation (Dorrer et al., 1999; O’Shea et al., 2001). Single-shot optical wake diagnostics developed naturally from these techniques. A third underlying technology was computerized
tomography (CT), developed as an x-ray-based internal medical diagnostic in the 1970s (Kak and Slaney, 1998). Tomographic wake diagnostic methods are just beginning to emerge (Li et al., 2014b). They draw from established CT the concept of reconstructing multi-dimensional images of difficult-to-access objects from multiple projections. They also utilize elements of established reconstruction algorithms. New challenges include adapting these concepts to an evolving light-speed object, recovering its picosecond evolution, and achieving high spatial and temporal resolution simultaneously.

1. Frequency-domain holography

(Siders et al., 1996a) first proposed extending FDI to a single-shot plasma wake diagnostic. One way of doing this is simply to time $N$ probe pulses at different delays $\tau_{pr}^{(i)}$ ($i = 1, 2, ..., N$) behind the pump (see Fig. 25a). A multi-armed Michelson interferometer, for example, has produced 16-pulse trains with high throughput (Siders et al., 1998). Since each probe-reference delay $T^{(i)} = \tau_{ref} - \tau_{pr}$ corresponds to a different oscillation period of the resulting multi-period FD interferogram (Fig. 25a, inset), Fourier analysis (Takeda et al., 1982) would then yield the time-domain phase shift $\Delta \phi(\tau^{(i)})$ on each probe in one shot. More simply, one can replace this multiplexed probe pulse train with a single continuous long pulse (see Fig. 25b). (Siders et al., 1996a) called the latter configuration “frequency-domain holography” (FDH), and envisioned creating the long probe pulse by inserting a flat-phase bandpass filter in the probe arm of a two-armed Michelson interferometer. This broadens the probe temporally, but maintains its phase coherence with the (still short) reference pulse. (LeBlanc et al., 2000) realized this goal in the laboratory by inserting phase-matched frequency-doubling crystals of different thickness into each arm of the interferometer (see Fig. 25b, inset). The thicker crystal (2 mm LiIO$_3$) had a narrow phase-matching bandwidth that generated a temporally long (1 ps) probe pulse, while the thinner crystal (150$\mu$m KDP) generated a temporally short (70 fs) reference pulse. The unperturbed probe and reference pulses formed FD fringes only within the narrower bandwidth of the probe pulse. However, after interacting with the ultrafast pump-induced index transient, the probe acquired new frequency components that interfered with the broader reference spectrum. The latter bandwidth determined the temporal resolution of the phase reconstruction. (LeBlanc et al., 2000) recovered laser-induced Kerr index transients in fused silica and ionization fronts in air over $\sim 1$ ps range with 70 fs time resolution, and 1D transverse spatial profiling, in a single shot.

(Chien et al., 2000; Geindre et al., 2001) implemented FDH using probe and reference pulses that were both broad-bandwidth and both temporally broadened simply by linearly chirping them, without narrowing their spectra (see Fig. 25c). In a chirped pulse, frequency components are distributed in a monotonic time sequence $\omega(t)$ within the stretched pulse. (Chien et al., 2000; Geindre et al., 2001) introduced chirp by passing the diagnostic pulse through a transparent, linear dispersive material before a Michelson interferometer split it into probe and reference pulses. This simplified FDH by eliminating alignment-sensitive transmissive optics inside the

![FIG. 25 Color online. Early development of frequency-domain holography (FDH) concept based on right-propagating pump (pu), probe (pr) and reference (ref) pulses, and wake (dashed curve). (a) Multiplexed FDI with several temporally short probe pulses (Siders et al., 1996a). Inset: Spectral intensity $I(\omega)$ of 3 probes and 1 reference pulse at detector. (b) FDH with single temporally long, transform-limited probe and temporally short, transform-limited reference (Siders et al., 1996a). Inset: laboratory implementation (LeBlanc et al., 2000). (c) FDH with equivalent chirped, wide-bandwidth probe and reference pulses (Chien et al., 2000; Geindre et al., 2001; Kim et al., 2002b). Inset: laboratory implementation (Matlis et al., 2006): incident 800nm diagnostic pulse split from pump up-converts to 400 nm via second-harmonic generation (SHG) in thin crystal, creating probe; glass plate group-delays (GDs) probe by $\sim 2$ ps from transmitted fundamental, which up-converts in second thin SHG crystal, creating reference; both chirp to $\sim 1$ ps duration in thick dichroic mirror (DM), which recombines them collinearly with pump.]
interferometer. These authors demonstrated single-shot recovery of the dynamics of laser-induced air ionization (Chien et al., 2000) and plastic target breakdown (Geindre et al., 2001), also with 1D transverse spatial profiling.

Chirped-pulse FDH had the additional advantage of lending itself to wide bandwidth supercontinuum probe and reference pulses (Kim et al., 2002b), which provide high longitudinal time resolution, wide temporal range, and (when generated in a frequency band near the pump) smaller group-velocity walk-off from the pump than frequency-doubled pulses. This approach yielded high-resolution measurements of e.g. double ionization in He (Kim et al., 2002a) and Kerr effect and plasma generation in various gases (Chen et al., 2007; Wahlstrand et al., 2011). However, nonlinear laser-wakefield excitation itself produces copious forward-directed chirped supercontinuum (Ralph et al., 2009), which interferes with co-propagating diagnostic pulses close to the drive pulse frequency, complicating image recovery. To avoid this, (Matlis et al., 2006) returned to frequency-doubled diagnostic pulses to obtain the first FDH “snapshots” of laser wakefields. The inset of Fig. 25c shows the compact probe-reference generator, based on a linear Fabry-Perot interferometer configuration, that (Matlis et al., 2006) introduced. This configuration proved more robust against vibrations and alignment errors than Michelson interferometers, while avoiding phase noise from background pump-generated supercontinuum. Visible supercontinuum pulses may nevertheless prove useful in imaging plasma wakes driven by particle-beams or mid-to long-wavelength-infrared laser pulses, for which this background is absent or at much longer wavelengths.

Using chirped pulses offers the intuitively attractive possibility of mapping the evolving pump-induced index transient \( \eta(t) \) directly onto the probe spectrum \( \omega(t) \) (Chien et al., 2000). In a simplified analysis, one divides the FD interferogram into \( N \) frequency bands \( \Delta \omega^{(i)}(t^{(i)}) \) \( (i = 1, 2, ..., N) \), then measures the fringe shift within each band to find the corresponding instantaneous index \( \eta(t^{(i)}) \). However, (Geindre et al., 2001; Kim et al., 2002b) pointed out that the narrow bandwidth of each bin limits \( \eta(t) \) recovery to slowly-varying index transient that satisfies \( \partial \eta / \partial t \mid_{t^0} < \Delta \omega^{(i)} \). More rapidly varying index transients shift spectral content out of bin \( \Delta \omega^{(i)} \) into neighboring bins, causing inaccurate recovery of \( \eta(t) \). Research on THz pulse modulation of chirped optical pulses has resulted in various strategies for addressing the distorted frequency-time relationship that the imprinting process causes (Fletcher, 2002; Peng et al., 2008; Yellampalle et al., 2005).

To recover rapidly-varying \( \eta(\Delta t) \) by FDH, limited only by the bandwidth \( \Delta \omega \) of the entire reference pulse, (Geindre et al., 2001; Kim et al., 2002b; LeBlanc et al., 2000) introduced a holistic FDH signal reconstruction procedure valid for either transform-limited (Fig. 25b) or chirped (Fig. 25c) probes. Figs. 26 and 27 illustrate the main steps, using data from (Matlis et al., 2006) as an example. For the data shown, a 0.3 J, 30 fs (10 TW) Ti:S pump pulse focused to beam waist \( w_0 \approx 50 \mu \text{m} \) in a 2-mm-long He gas jet created plasma of average density \( n_e \sim 10^{18} \text{ cm}^{-3} \) and a wake of wavelength \( \lambda_p \sim 25 \mu \text{m} \) in the doubly-ionized region. Fig. 26a shows a raw FD hologram, acquired using the “standard” chirped pulse FDH configuration shown in Fig. 25c. Both ionization front (not shown in Fig. 25c) and wake oscillations contributed to the fringe distortions in Fig. 26a. Fig. 26b shows a line-out of the hologram \( I(y_0, \omega) \) at \( y = y_0 \). Formally, \( I(y_0, \omega) \) is given by Eq. (46), and for identical unperturbed probe and reference pulses, by Eq. (46c). However, Eq. (46d) does not apply, since the FD phase \( \Delta \phi_{pr}(y, \omega) \) is no longer equivalent to, nor simply related to, the desired pump-induced TD phase shift \( \Delta \phi(y, \Delta t) \) as in FDI.

To recover \( \Delta \phi(y, \Delta t) \), (Geindre et al., 2001; Kim et al., 2002b; LeBlanc et al., 2000) began by isolating the com-
plete frequency-domain probe electric field

\[ E_{pr}(y, \omega) = |E_{pr}(y, \omega)|e^{i[\Delta \phi_{pr}(y, \omega) + \phi_{ch}(\omega)]} \]  \hspace{1cm} (47) \]

from the FD hologram, using a computer-based fringe-pattern analysis method introduced by (Takeda et al., 1982). Here \( \phi_{ch}(\omega) \) is the FD phase due to the probe chirp. We now suppress the argument \( y \) for brevity. To isolate \( E_{pr}(\omega) \), they first inverse-Fourier-transformed the recorded FD hologram \( I(\omega) \), given by Eq. (46)b or Fig. 26b, at each \( y \):

\[ S(t) = \tilde{F}[I(\omega)] \]  \hspace{1cm} (48) \]

This operation electronically “reads” the recorded hologram with a plane wave, analogous to physically reading a conventional film-recorded hologram with the reference wave (Siders et al., 1996a). The resulting complex time-domain function \( S(t) \), of which Fig. 26c shows the amplitude, has the form \( S(t) = 2h(t) + H(t + T) + h(t - T) \) with peaks at \( t = 0, -T \) and \( T \). These peaks correspond, respectively, to \( \tilde{F} \) of the 3 terms in Eq. (46)b. The central \( (t = 0) \) peak is the intensity autocorrelation of the reference and probe pulses, but lacks phase information. Only the side (cross-correlation) peaks at \( t = \pm T \) encode the desired phase of \( E_{pr} \). (Siders et al., 1996a) gives complete expressions for all 3 peaks. The side peak at \( t = T \) has the general form

\[ h(t - T) = \tilde{F}[E_{ref}^*(y, \omega)E_{pr}(y, \omega)e^{-i\omega T}]. \]  \hspace{1cm} (49a) \]

For unperturbed linearly-chirped, Gaussian pulses \( E_{pr}(\omega) = E_{ref}(\omega) = E_0 \exp(-(1/2)(1 + i\sigma)|\omega - \omega_0|/\Delta \omega|^2 \), where \( \sigma \) represents the FD chirp, it has the specific form (Matlis et al., 2016)

\[ h(t - T) = \Delta \omega \left( \frac{E_0^2}{\sqrt{2}} \right) e^{-i\omega_0(t - T)}e^{-\frac{1}{2}(t/\delta t)^2}. \]  \hspace{1cm} (49b) \]

where \( \delta t \approx 1/\Delta \omega \) is the coherence time of the probe and reference pulses. When a temporal phase shift \( \Delta \phi_{pr}(\Delta t) \) modulates the probe, \( h(t - T) \) develops informative substructure, shown for \( 0.95 \times 10^{18} \text{ cm}^{-3} < n_e < 6 \times 10^{18} \text{ cm}^{-3} \) in the magnified view of the base of the \( t = T \) peak in the inset of Fig. 26c. The step-function-like ionization front, which in the FD blue-shifts frequency components that it overlaps, appears as a shoulder on the \( t < T \) side of \( h(t - T) \). This shoulder becomes more prominent at higher \( n_e \). A sinusoidal wake oscillation, which in the FD creates Stokes and anti-Stokes sidebands at \( \pm \omega_p \), appears in the TD as sidebands at \( T \pm \delta T \). Here \( \delta T \), which is \( \sigma \omega_p / \Delta \omega^2 \) for linearly-chirped probe and reference pulses (Matlis et al., 2016), increases as \( \sqrt{n_e} \), and exceeds \( \delta t \) throughout the \( n_e \) range shown. A nonlinear wake, which contains harmonics of \( \omega_p \), would create additional higher-order TD sidebands at \( T \pm m\delta T \), where \( m = 2, 3, \ldots \) (not shown in the inset of Fig. 26c). For cases of interest, these ionization- and wake-induced spectral shifts are smaller than \( \Delta \omega \), and thus are not visible directly in the FD. Chirped pulse FDH, however, encodes them in the TD, and by Fourier-transforming \( I(\omega) \), converts them to temporal shifts \( \delta T \) greater than the temporal width \( \delta t \) of the reconstructed \( h(t - T) \) peak. (Matlis et al., 2016) called this process temporally-encoded spectral shifting (TESS). TESS analysis enables subtle wake-induced phase modulations to be separated from the often much larger phase modulations that ionization fronts imprint, e.g. by analyzing the wake-induced sideband on the \( t > T \) side of \( h(t - T) \). This gives FDH a capability analogous to “differential” FDI mode (see Sec. IV.B.2). Recently (Cowley et al., 2017) used TESS to characterize plasma wakes driven by optimized trains of laser pulses.

To complete the extraction of \( E_{pr}(\omega) \), (Geindre et al., 2001; Kim et al., 2002b; LeBlanc et al., 2000) multiplied \( h(t - T) \) by a soft-edged apodizing window centered at \( t = T \) to isolate this peak (indicated by dashed rectangle in Fig. 26c). The window must be wide enough to include all sub-structure that the interaction introduced. Fourier-transformation back to the FD yields

\[ |E_{ref}(\omega)||E_{pr}(y, \omega)|e^{i\Delta \phi_{pr}(y, \omega)} \]  \hspace{1cm} (50) \]

for the \( t = T \) peak. Finally, one divides Eq. (50) by the independently-measured reference pulse power spectrum \( |E_{ref}(\omega)|^2 \) and augmented it with the FD chirp phase \( \phi_{ch} \), measured by standard single-shot pulse characterization methods (Dorror et al., 1999; O’Shea et al., 2001), to complete reconstruction of the field Eq. (47).

The final step in recovering \( \Delta \phi(y, \Delta t) \) is inverse Fourier transformation of \( E_{pr}(y, \omega) \) to get the complete time-domain probe electric field

\[ E_{pr}(y, \Delta t) = \tilde{F}[E_{pr}(y, \omega)] \]  \hspace{1cm} (51) \]

After subtracting the independently characterized time-domain chirp \( \phi_{ch}(y, \Delta t) \) from its phase, we recover \( \Delta \phi(y, \Delta t) \) at one \( y \). Repeating this analysis for each \( y \) yields the 2D phase map shown in Fig. 27a.

Fig. 27b compares this phase map with simulated plasma density perturbations \( \delta n_e(y, \Delta t) \) near the center of the gas jet (\( z = 1 \text{ mm} \)) for the same conditions. The overall structures of \( \delta n_e(y, \Delta t) \) and \( \Delta \phi_{pr}(y, \Delta t) \) match closely: boundaries of \( \text{He}^{2+} \) and \( \text{He}^+ \) regions in Fig. 27b appear at \( y \) values similar to those of sharp \( \Delta \phi_{pr} \) boundaries in Fig. 27a; plasma density oscillations in the \( \text{He}^{2+} \) region in Fig. 27b have the same wavelength and radial structure as corresponding \( \Delta \phi_{pr} \) oscillations in Fig. 27a. (Matlis et al., 2006) found similar correspondence for shots over a wide \( n_e \) range, showing that the approximation of a collimated probe pulse, discussed in connection with FDI in Sec. IV.B.2, remained valid here. Nevertheless, (Matlis et al., 2006) found the amplitude of phase oscillations in Fig. 27a was significantly smaller than expected from probing the density perturbations in Fig. 27b.
over a uniform jet. Group velocity walk-off of 400 nm probe from 800 nm pump did not explain the discrepancy. Instead it arose from longitudinal density nonuniformity of the gas jet, which caused the pump to generate, and probe to sample, wakes of different frequencies as they co-propagated through the 2 mm jet. Remarkably, despite this averaging, the recovered hologram Fig. 27b preserved the main features of the wake near the center of the jet. Longitudinally uniform gas targets are thus preferred to obtain accurate wake amplitudes from FDH.

(Dong et al., 2010a; Matlis et al., 2006) found that imaging of nonlinear wakes generated by 1 J, 30 fs (30 TW) pump pulses enabled independent in situ calibration of absolute wake oscillation amplitude. FDH images of wakes excited under these conditions had curved wavefronts (Fig. 28a), a signature of strongly driven, nonlinear laser-plasma interaction (Andreev et al., 1997b; Decker et al., 1994). Wavefronts that were flat immediately behind the pump evolved into curved horse-shoe profiles after several periods. Simultaneously, the amplitude of the phase-shift oscillations increased over the same interval. A particle-in-cell simulation of the wake (Fig. 28b) showed that plasma density oscillations δn_e(y, Δt)/n_e near the center of the gas jet also exhibited both of these features. Both are relativistic in origin. The wavefronts curve because as |δn_e(r = 0, Δt)| approaches unity on axis (r = 0), electrons making up the wave oscillate relativistically (γ > 1), causing ω_p(r = 0) to decrease by √γ relative to its off-axis value (Andreev et al., 1997b; Decker et al., 1994). Simulations showed that the reciprocal radius of curvature ρ−1 = gΔt grows linearly with Δt, where growth rate g depends sensitively on |δn_e(r = 0, Δt)/n_e|. Analysis of ρ−1 in the FDH image in Fig. 28a showed |δn_e(r = 0, Δt)/n_e| was ∼ 0.5 immediately behind the pump, then grew steadily over 6 cycles. (Dong et al., 2010a) explained this growth by analogy with amplitude growth observed in simulations of wakes generated in plasma channels with parabolic radial density profiles (Andreev et al., 1997a; Shvets and Li, 1999). Here the radial relativistic γ(r)ω_p profile played the role of the channel density profile. With increasing Δt, trajectories of radially neighboring electron fluid elements, oscillating at slightly different frequencies, approach. In a process akin to optical pulse compression, interaction amongst fluid elements spanning the γ(r)ω_p bandwidth steepen and narrow the plasma wavefronts, causing the observed amplitude growth. As trajectories cross, the waves can eventually break, although the image and simulation in Fig. 28 stop before this happens.

FIG. 27 Color online. (a) Composite “snapshot” of right-propagating wake formed by stacking temporal pump-induced phase structure ∆ϕ_{pr}(y, Δt) extracted from cross-correlation peaks at each y. (b) Simulated wake density profile n_e(y, Δt) near center of gas jet for conditions matching the experimental data. Adapted from (Matlis et al., 2006).

FIG. 28 Color online. Strongly-driven right-propagating wake with curved wavefronts. (a) FDH phase profile ∆ϕ_{pr}(y, Δt) of wake (colored surface) that a 30 TW pump pulse generated in He^{2+} plasma with n_e = 2.2 × 10^{18} cm^{-3}. The grey-scale image is a projection onto a plane. DC phase shift from surrounding plasma profile has been subtracted to highlight wake oscillations. (b) Simulated wake density profile n_e(y, Δt) near the gas jet center, showing growth of wavefront curvature and amplitude with increasing Δt as in (a). Adapted from (Dong et al., 2010a): © Deutsche Physikalische Gesellschaft, reproduced by permission of IOP Publishing (CC-BY-NC-SA).
Correlated growth of wave curvature and amplitude, effects never previously observed in the laboratory, are thus precursors of wave breaking and electron injection. The example illustrates the new wakefield physics that is accessible from dialog between in-situ plasma structure diagnostics and simulations.

2. Longitudinal optical shadowgraphy

The images in Figs. 27a and 28a were reconstructed entirely from accumulated phase shift \( \Delta \phi_{pr}(r, \Delta t, z_{exit}) \) on the wake-modulated probe pulse at the exit plane of the accelerator. Corresponding changes in the probe amplitude \( |E_{pr}(r, \Delta t, z_{exit})| \) were negligible under these conditions. As \( n_e L \) increases, plasma structures with high index contrast not only re-shape the drive pulse (Decker et al., 1996), but refract a co-propagating FDH probe as well. As an example, the top row of Fig. 29 shows a PIC simulation using the code WAKE (Mora and Antonson, 1997) of guiding and compression of an initially 30 fs, 30 TW drive pulse (outlined in red) as it generates nonlinear wake profile \( n_e(r, \Delta t, z) \) (grey scale) upon propagating from \( z = 0.1 \) mm (left) to its depletion distance \( z = L_d = 1.8 \) mm (right) in plasma of average density \( n_e = 8 \times 10^{18} \) cm\(^{-3}\). The second row of Fig. 29 shows corresponding changes in the amplitude envelope \( |E_{pr}(r, \Delta t, z)| \) of a co-propagating 400 nm FDH probe pulse. By \( z = 0.5 \) mm, the pump compresses to 20 fs (top middle) as it blows out electrons from the first wake bucket. This high-index (\( \eta = 1 \)) plasma bubble also focuses probe light inside it (bottom middle), and, by \( z = 1.8 \) mm, compresses it near the front of the bubble to dimensions \( \lambda_{pr}^3 \sim (10 \mu\text{m})^3 \). Formation of such 3D-confined light packets correlates closely with, and non-invasively diagnoses, bubble formation.

(Dong et al., 2010b) observed bubble-formed light packets, which they called “optical bullets”, in the laboratory by recovering the amplitude profile \( |E_{pr}(r, \zeta, z_{exit})| \) of an FDH probe pulse co-propagating with a laser-generated bubble. Fig. 30 shows sample results using a \( \sim 2 \)-mm-thick He gas jet, and laser parameters as in the simulations above. Fig. 30a shows a reference reconstruction of the unaltered incident probe pulse, acquired with the gas jet turned off. For Fig. 30b, the pump created plasma of density \( n_e = 1.2 \times 10^{19} \) cm\(^{-3}\), and an optical bullet (highlighted by vertical arrow), signifying bubble formation, appeared near the probe leading edge. Yet no electrons were produced in this case, showing that bubbles can form below the threshold for spontaneous electron injection. With further increase of density (\( n_e = 3.2 \times 10^{19} \) cm\(^{-3}\)), a smaller, brighter optical bullet formed (Fig. 30c-1) and nearly mono-energetic electrons were produced (Fig. 30c-2). Supporting simulations confirmed the observed bubble formation and injection thresholds (Dong et al., 2010a,b).

FIG. 29 Color online. WAKE simulations for \( n_e = 8 \times 10^{18} \) cm\(^{-3}\) showing formation of plasma wake (top row), with pump \( e^{-2} \) isointensity contours outlined by solid red curves, and refraction of 400 nm FDH probe to form optical bullet inside the bubble (bottom row). Pump and probe pulses propagate from gas jet entrance (left column) to \( z = 0.5 \) mm (middle column) to the pump depletion and bullet formation length \( z_{exit} \) (right column) near jet exit. Grey scale indicates electron density (top) or probe intensity (bottom). Adapted from (Dong et al., 2010a).

FIG. 30 Color online. Longitudinal shadowgraphy of strongly nonlinear LWFA. Probe amplitude profile \( |E_{pr}(r, \zeta, z_{exit})| \) is reconstructed using FDH methods (a) with no gas jet (undistorted profile), and after co-propagating left to right with wake in doubly-ionized He plasma of density \( n_e[10^{19} \text{ cm}^{-3}] = 1.2 \) (b) or 3.2 (c-2), showing optical bullets (highlighted with vertical arrows) trapped inside “bubble”. (c-2): electron energy spectrum. Adapted from (Dong et al., 2010a).

The FDH analysis procedure also outputs probe phase profiles \( \Delta \phi_{pr}(r, \zeta, z_{exit}) \) in the bullet formation regime. However, they are no longer as simply related to plasma structure as in the quasi-linear regime of (Matlis et al., 2006). This is in part because \( \Delta \phi_{pr} \) often exceeds \( 2\pi \), creating phase jumps that are difficult to unwrap (Ghiglia and Romero, 1994), and in part because refraction distorts the radial distribution of \( \Delta \phi_{pr}(r, \zeta, z_{exit}) \). \( |E_{pr}(r, \zeta, z_{exit})| \) profiles alone lend themselves to clear physical interpretation in highly refractive plasmas.
Since they resemble shadowgraphs projected on a virtual screen at \( z_{\text{exit}} \), FDH can be called “FD shadowgraphy” (FDS) in this regime.

3. Transverse optical probing

In discussing scattering of long \((\tau_{pr} > \omega_p^{-1})\) probe pulses from plasma waves (Sec. IV.A), we noted that co-propagating and transverse probes yielded complementary diagnostic information. The same is true of ultra-short \((\tau_{pr} < \omega_p^{-1})\) probe pulses. Co-propagating CTS (Sec. IV.A), FDI (Sec. IV.B.2), FDH (Sec. IV.C.1), and FDS (Sec. IV.C.2) probes integrate longitudinally over the wake’s evolution as it propagates. Yet such evolution is an essential part of nonlinear wakefield acceleration. Its diagnosis requires a probe with a velocity component transverse to \( \vec{k}_{pr} \). Wake-induced alteration of an ultra-short transverse probe pulse can yield a snapshot related to the wake’s internal plasma density (Sävert et al., 2015) or magnetic field (Kaluza et al., 2010) profile, or both (Buck et al., 2011), with sub-\( \lambda_p \) resolution at time \( \Delta t \). When shot-to-shot-fluctuations are small, a \( \Delta t \)-sequence of such images from successive shots forms a movie of the evolving wake (see Sec. IV.D). Figure 31 shows a transversely probed plasma accelerator schematically.

\[ \nabla \times \vec{B}(\vec{r},t) = \mu_0 \left( \vec{j}(\vec{r},t) + \varepsilon_0 \frac{\partial \vec{E}(\vec{r},t)}{\partial t} \right) \]  

(52)

that reach kilo-Tesla strength in plasma surrounding a bubble. A transverse linearly-polarized probe impinging on the wake thus “sees” \( \vec{B} \) components that are both parallel and perpendicular to \( \vec{k}_{pr} \) (see Fig. 31, lower left inset) that alter its polarization by the Faraday or Cotton-Mouton effects, respectively. The Faraday effect \((\vec{k}_{pr}||\vec{B}(\vec{r},t))\) locally rotates probe polarization, which remains linear. The Cotton-Mouton effect \((\vec{k}_{pr} \perp \vec{B}(\vec{r},t))\) locally induces polarization ellipticity. However, a transverse probe experiences equivalent components \( \vec{B} \perp \vec{k}_{pr} \) of opposite sign on the entrance and exit sides of the azimuthal field profile. Thus, the Cotton-Mouton effect approximately cancels out. In contrast, probe rays propagating above or below the central axis of the plasma wave experience components \( \vec{B} \parallel \vec{k}_{pr} \) that retain their direction along each ray’s entire path. Consequently, Faraday polarization rotation accumulates up to angle

\[ \phi_{\text{rot}} = \frac{e}{2m_e c} \int_{\text{plasma}} \frac{n_e(\vec{r})}{\langle \gamma \rangle \cdot n_{\text{cr}}} \vec{B}(\vec{r}) \cdot d\vec{s}, \]  

(53)

where the integration is along the path of each probe ray through the magnetized plasma, \( n_{\text{cr}} \) is the critical density at \( \lambda_{pr} \) (see Eq. (5)), and \( \langle \gamma \rangle \) is the time-averaged Lorentz factor of streaming magnetized background plasma electrons (not of the accelerated electron bunch) that induce Faraday rotation. (Buck et al., 2011; Kaluza et al., 2010) observed typical Faraday rotations \( \phi \sim 1^\circ \).

The polarization of probe rays propagating above or below the plasma wave’s axis rotates in opposite directions, since \( \vec{B}(\vec{r}) \cdot d\vec{s} \) changes sign. A linear polarizer converts these local Faraday rotations into intensity modulations that a CCD camera detects. To distinguish subtle Faraday modulations from background probe intensity variations, (Buck et al., 2011; Kaluza et al., 2010) employed differential detection with two CCD cameras imaging the same region of plasma, as shown in Fig. 31. They detuned polarizers in front of each CCD in opposite directions from the blocking angle for the initial probe polarization. Consequently, when they divided the two raw images Figs. 32a,b, intensity variations unrelated to polarization rotation canceled out, whereas those induced by Faraday rotation doubled. Magnetized plasma regions then stood out clearly, as shown in Fig. 32c,d,e.

The transverse size \((\sim 55 \times 35 \, \mu m^2)\) of the Faraday-rotated region in Fig. 32c (see dotted ellipse) was determined mostly by imperfect imaging and by \( \sim 100 \) fs probe pulse duration (Kaluza et al., 2010). With improved imaging resolution \((\sim 1 \, \mu m)\) and a shorter \( (8.5 \, fs) \) probe, the imaged signal shrank (see Fig. 32f) to a size limited by the transverse diameter of the plasma wave and the duration of the accelerating electron bunch, rather than the

![Figure 31](image-url)
measurement system. Analysis of the horizontal extent of this improved Faraday signal yielded electron bunch duration $\tau_c = 5.8^{+1.9}_{-2.1}$ fs (see Sec. III.D). As $\Delta t$ changed, this signal visualized formation and acceleration of the electron bunch (Buck et al., 2011).

b. Transverse shadowgraphy of plasma wakes. Shadowgraphic images of the laser-driven plasma that (Kaluza et al., 2010) recorded on each individual detector using $\tau_{pr} \approx 100$ fs probe pulses (see Fig. 32a,b), despite clear intensity variations, showed no evidence of a wake structure oscillating with the period $\lambda_p \approx 5\mu m$ expected in $n_e = 4 \times 10^{19}$ cm$^{-3}$ plasma. This is because $c\tau_{pr}/\lambda_p \approx 6$
cycles of the light-speed wake passed while the probe illuminated it transversely, washing out any signature of individual cycles in the shadowgraphic image. (Buck et al., 2011) found that only by using probe pulses compressed to $c\tau_{pr} \lesssim 0.5\lambda_p$ did a periodic wake appear in the shadowgraph, as shown in Fig. 32f for $c\tau_{pr} \approx 0.43\lambda_p$ ($\tau_{pr} = 8.5$ fs, $\lambda_p \approx 6\mu m$, $n_c = 3.2 \times 10^{19}$ cm$^{-3}$). The period scaled with $n_c^{-1/2}$, confirming that it originated from plasma waves. Superposition of Faraday rotation and shadowgraphic images derived from a common detection system (Fig. 32f) then localized the magnetized region surrounding the electron bunch with respect to the wake.

(Buck et al., 2011) split 8.5 fs probe pulses from $\sim 65$ mJ (< 10 TW) pump pulses that were also compressed to 8.5 fs. Such short drive pulses had only enough energy to generate mildly nonlinear wakes and tens of MeV electrons. Later (Schwab et al., 2013) spectrally broadened split-off probes in a gas-filled hollow core fiber before compressing them to the required few-cycle duration, thereby decoupling probe from drive pulse duration. This enabled transverse probing of strongly nonlinear wakes driven by more powerful (> 30 TW), albeit longer (~ 30 fs), pulses. Moreover, they generated diagnostic pulses as short as $\tau_{pr} = 5.9 \pm 0.4$ fs. A wider range of wakes could thereby be transversely imaged.

FIG. 32 Color online. Faraday rotation measurements of quasi-linear LWFA. (a, b) Two CCD camera images of the same interaction region in $n_e = 4 \times 10^{19}$ cm$^{-3}$ plasma taken with 100 fs transverse probe light through polarization analyzers detuned in opposite directions from the blocking orientation. (c) Ratio of images (a)/(b), to highlight intensity variations caused by +/- Faraday rotation (blue, upper/red, lower). (d) Lineout of intensity ratio along vertical line connecting single red arrows at top and bottom of (c); (e) deduced rotation angle. Panels a) thru e) from (Kaluza et al., 2010). f) Similar result in $n_e = 3.2 \times 10^{19}$ cm$^{-3}$ plasma using 8.5 fs transverse probe and higher resolution imaging, showing Faraday rotation signal (color at $z - ct = y = 0$) induced by magnetic fields from laser-accelerated electron bunch of $5.8^{+1.9}_{-2.1}$ fs duration, superposed on periodic structure (grey scale) observed in single detector, from probe refraction by plasma wave. Panel f) adapted from (Buck et al., 2011).

FIG. 33 Color online. Simulated and measured shadowgraphs of strongly nonlinear wake. (a) Simulated electron density relative to ambient $n_e = n_0 = 1.7 \times 10^{19}$ cm$^{-3}$, driven by 36 fs, 810 nm pulse focused to intensity $2.5 \times 10^{18}$ W/cm$^2$. 18.8$\mu$m spot (FWHM) after propagating $v_g t \approx 1.2$ mm. (b) Simulated shadowgraph for this distribution, and (c) corresponding experimental shadowgraph using $\lambda_p = 0.75\mu m$. Probe intensity change $\Delta I_{pr}$ relative to incident probe intensity $I_0$ is plotted. Brown dotted lines indicate lengths of first ($\lambda_b$) and second ($\lambda_p$) plasma wave periods. Adapted from (Siminos et al., 2016).
Figure 33c shows a shadowgraph of a right-propagating bubble-regime wake acquired with this versatile system (Sävert et al., 2015). At the head of the wake, two dark, oppositely-curved, arcs define an ellipse with major (minor) axis 10 (7) µm. This feature originates from the first, directly-laser-driven period of the wake. Behind are several smaller (radius ∼ 5µm), lower-contrast dark circles that originate from subsequent periods of the wake. While it is tempting to correlate these features directly with electron density, in reality dark regions form when probe rays refract in density gradients. In addition, the wake moves distance λ_e during transit of even a δ-function probe pulse across the wake, causing temporal blurring. Optical aberrations and the finite light collection angle also influence the image.

To relate light intensity distribution in the shadowgraph to electron density distribution in the wake, (Siminos et al., 2016) carried out 3D PIC simulations using the code EPOCH (Arber et al., 2015) that included a transverse probe pulse. Fig. 33a shows the simulated wake density distribution for conditions and propagation distance (z = 1.2 mm) corresponding to the shadowgraph in Fig. 33c. Fig. 33b shows the simulated shadowgraph, taking into account realistic imaging optics. The simulation showed that the probe refracted most strongly as it crossed the propagation axis, where longitudinal density gradients ∂n_e/∂z were strongest. This explained how sharp sub-λ_p features appeared in the shadowgraphs, even though the wake propagated ∼ λ_p during a probe transit. The length λ_b of the simulated, fully-evacuated leading bubble (Fig. 33a) is ∼ 20% smaller than the major axis of the simulated (Fig. 33b) and measured (Fig. 33c) leading shadowgraph ellipse. This is the result of transverse deflection of probe light at the dense front and back walls of the leading bubble. On the other hand, the diameters of near-circular trailing shadow closely match the length λ_p of trailing buckets of the simulated wake (Fig. 33a). These trailing buckets are less fully evacuated, and have less dense walls, than the leading bucket, and thus deflect probe rays less. (Sävert et al., 2015; Siminos et al., 2016) attributed lengthening of the first bucket to a rapid increase of drive pulse intensity and of the associated relativistic mass of plasma electrons that immediately precedes, and prompts, self-injection of electrons into this bucket. Transverse shadowgraphy enabled direct visualization of this critical, and otherwise elusive, stage of wakefield physics.

4. Electron radiography

Transverse electron radiography complements transverse optical shadowgraphy by probing internal electromagnetic fields of a dynamic wake. Relativistic electron bunches probe wakes in the ray optics regime, limited by their energy bandwidth, transverse emittance and duration rather than by refraction or diffraction. Hence they are subject to different resolution limits than optical probes. Moreover, as with longitudinal electron witness bunches (see Sec. IV.B.1), electron probes are sensitive to lower density plasma structures than optical probes, an advantage for diagnosing GeV plasma accelerators. Electron probes are also insensitive to quasi-neutral plasma and gas surrounding a wake, which can imprint unwanted background phase shift on a transverse optical probe.

(Williams, 1995; Williams et al., 1990a,b) simulated interaction of long (τ > ω_p⁻¹), transverse electron probes with wakes in 10¹⁶ cm⁻³ ≲ n_e ≲ 10¹⁷ cm⁻³ plasma. As for long optical probes (see Sec. IV.A), the simulations showed that scatter of low-emittance bunches could characterize global wake structure, but not sub-λ_p structure. (Fainberg et al., 1998, 1996) proposed to resolve sub-λ_p structure of wakes in n_e ∼ 10¹¹ cm⁻³ plasma via picosecond transverse electron radiography. At such low n_e, however, accelerating fields would be smaller than in conventional RF accelerators. Electron radiography of sub-λ_p structure of wakes in n_e ∼ 10¹⁷ cm⁻³ plasma requires bunches of few femtosecond duration.

LWFAs themselves provide bunches of this duration (see Sec. III.D), and of very small ε_n (see Sec. III.C). (Schumaker et al., 2013) exploited these properties to probe evolving magnetic fields in a laser-excited solid target with sub-picosecond time- and micrometer space-resolution. (Zhang et al., 2017, 2016b) used ultrashort, low-ε_n, 60-80 MeV electron bunches from one LWFA to probe transversely the internal electromagnetic fields of a second LWFA (λ_p ≈ 65µm) driven by a split-off portion of the same laser pulse (see Fig. 34a). The diverging probe, after expanding to ∼ 800µm diameter, irradiated a multi-λ_p section of the subject wake, which deflected probe electrons transversely. A scintillating screen at distance L_{scr} from the interaction recorded the 2D profile I(x,y) of the transmitted electron bunch (see Fig. 34a). The wide field of view enabled observation of variations in wake structure along its length. For example, a radiograph of a wake within a density ramp n_e(z) (Fig. 34b,c) revealed changing λ_p along the ramp.

As with transverse optical shadowgraphs (see Sec. IV.C.3), the electron radiograph I(x,y), though not a simple projection, is closely related to the wakefield that produced it. For best wake reconstruction, L_{scr} should be small enough to avoid electron trajectory crossing, which loses information. For wakefield amplitude given by Eq. (4a), this condition leads to L_{scr} ≲ Mγ_prλ_p/10, where M is a geometric magnification factor such that M = 1 (> 1) for a collimated (diverging) probe, and γ_pr = (1 − β²_pr)⁻¹/² is the Lorentz factor (β_pr = normalized probe velocity). Optimal L_{scr} can range from a few to hundreds of millimeters. In the limit of an extremely short (τ_pr ≪ ω_p⁻¹) bunch probing a quasi-static linear wake that perturbs
incident electron momentum $p_{pr}$ only slightly, the effects of the radial $[E_r(r,z)]$ and longitudinal $[E_z(r,z)]$ wake electric fields on the probe decouple via the Panofsky-Wenzel theorem (Panofsky, 1956). $E_{r,z}(r,z)$ can then be recovered exactly by solving two Abel transforms

$$\frac{\partial I}{\partial z} = \kappa K_m \nabla^2 \int_{-s}^{s} E_r'(r,z)dx$$

$$\frac{\partial I}{\partial y} = \kappa K_m \nabla^2 \int_{-s}^{s} E_z'(r,z)dx,$$

where $\kappa = eL_m/\beta_{pr}c_{pr}$, $K_m = \exp\left[-\left(h_p\sigma_E\right)^2/2\right]$ is an averaging factor caused by wake motion that depends on wake width $\sigma_E$, and $E_{r,z}'$ denote static fields with the same form as $E_{r,z}$. Reconstruction becomes less accurate, but still possible, for strongly nonlinear wakes. The wake’s magnetic field then contributes significantly to electron deflection. Unlike an optical probe, the electron probe senses the electric (as well as magnetic) field of the accelerating electron bunch inside the wake. This is because the bunch’s electric field, unlike its refractive index, is not suppressed by a factor $\gamma_e^{-1}$, which makes a highly relativistic bunch practically invisible to an optical probe, except via Faraday rotation (see Sec. IV.C.3). Simultaneous sensitivity to the electric and magnetic fields of highly relativistic accelerating bunches and of tenuous plasma structures will make transverse electron radiography an attractive choice for characterizing multi-GeV laser- and beam-driven plasma accelerators.

(Clayton et al., 2016), working at SLAC’s FACET (Hogan et al., 2010), mapped longitudinal variation of fields within a strongly blown out PWFA in one shot, using electrons in the trailing portion of the drive bunch itself ($\sim 20$ GeV, $\sigma_z \sim 25\mu$m) as witnesses. The energy spectrum of these electrons, imaged from the PWFA exit plane, exhibited a series of energy peaks and transverse bunch size modulations originating from their transverse oscillations in the bubble’s radial $E_r$ fields as they accelerated in its longitudinal $E_z$ fields. Analysis of this structure enabled reconstruction of $E_r$ along the bubble’s length, and showed that it was longitudinally uniform to within $\pm 3\%$, as expected for a nearly fully blown out bubble. From the Panofsky-Wenzel theorem, the authors inferred comparable radial uniformity of $E_z$, a key requirement for emittance preservation. Such field maps can help to optimize placement and shape of separate, shorter, higher charge witness bunches.

D. “Movies” of wake evolution

1. Multi-shot transverse probes

Transverse optical (Sec. IV.C.3) or electron (Sec. IV.C.4) probing records wake shadowgraphs or radiographs at fixed delay $\Delta t$ between drive and probe pulses. If shot-to-shot variations can be neglected, a sequence of projections recorded over multiple shots with varying $\Delta t$ yields a “movie” of the wake’s evolution as it propagates through the plasma.

Fig. 35 shows a $\Delta t$-sequence of six optical shadowgraphs of a strongly nonlinear wake driven by 35-fs, 750 mJ pulses in $n_e = 1.65 \times 10^{19}$ cm$^{-3}$ plasma (Sävert et al., 2015). They illustrate several stages of wake evolution. Early in the laser-plasma interaction [Fig. 35(a)], successive dark regions in the shadowgraph (positions highlighted by white vertical lines beneath) were spaced nearly equally, indicating a linear plasma wave. Subsequently, contrast between dark and light regions increased [Fig. 35(b)], signifying increased wave amplitude. In Fig. 35(c), the first plasma period lengthened, signaling onset of strongly nonlinear laser-plasma interaction (see Sec. IV.C.3.b). Simultaneously, $\sim 65\mu$m ahead of the

FIG. 34 Color online. Transverse electron radiography of laser wakefield. (a) Schematic experimental setup. The electron probe bunch (4 fs FWHM, 2–10 pC, 7 mrad FWHM divergence) from one plasma accelerator (2mm gas jet, $\bar{n}_e = 8 \times 10^{16}$ cm$^{-3}$, 95% He, 5% N$_2$) driven by a 25 TW, 40 fs FWHM Ti:S laser pulse, propagates transversely through the “subject” plasma accelerator (3mm gas jet, with synchronized 4 TW, 100 fs FWHM Ti:S drive pulse) located 11 cm away, then exposes a Ce:YAG scintillator further downstream. Magnetic spectrometer can measure probe bunch energy with scintillator removed. From (Zhang et al., 2017). Below: scintillator images at probe delays (b) $t_0$, (c) $t_0 + 2.3$ ps, and (d) $t_0 + 4.6$ ps of wake generated in tapered density up-ramp ($\bar{n}_e \approx 2.5 \times 10^{17}$ cm$^{-3}$ at $z = 0$ to $3.5 \times 10^{17}$ cm$^{-3}$ at $z = 665\mu$m), yielding decreasing plasma period evident in panels (b), (c). Panels b) thru d) from (Zhang et al., 2018).
wake, bright plasma emission, spectrally much broader (600-1000 nm) than the drive pulse, was observed, consistent with “wave breaking” radiation (see Sec. IV.A), a signature of onset of self-injection (Thomas et al., 2007). Continuing increase in shadow contrast at the beginning of the wave train [Figs. 35(d)-(f)] signifies increasing density gradient at the front of the bubble as the wave becomes highly nonlinear. In Fig. 35(f), the direction of curvature of the trailing shadowgraph wave periods even reverses. These features are closely linked to transverse wave breaking (Bulanov et al., 1997a).

Two extensions of the results in Fig. 35 offer rich possibilities for future diagnostic development. First, the shadowgraphic movie of the evolving plasma structure can be coordinated with a Faraday rotation movie of the evolving electron bunch, as (Buck et al., 2011) already did for single frames [see Fig. 32(f)]. The magnetic signature of injected electrons should appear with wave-breaking radiation as the primary bubble lengthens [Fig. 35(c)], and evolve in subsequent frames. Second, one could acquire the movie in one shot by multiplexing the probe pulse. The possibilities parallel those considered in developing single-shot FDH from multi-shot FDI (see Fig. 25). One is to split the probe into N replicas, each backlighting a different section of the interaction region, and projecting its shadowgraph onto a separate detector. Another possibility (cf. FDH) is to chirp the probe pulse. This relieves the requirement of maintaining few-cycle pulse duration, and maps probe arrival time at the wake onto probe frequency (Siminos et al., 2016). One could then distribute frequency bands \( \Delta \omega_i \) \((i = 1, 2, \ldots, N)\) of the wake-diffracted probe to separate detectors, each recording a color-coded 2D shadowgraph of a different stage of wake evolution. However, as discussed in connection with a similar suggestion for chirped longitudinal probes (see Sec. IV.C.1), the wake diffractions the probe over time scale \( t \ll \Delta \omega_i^{-1} \), imprinting new frequency components on each band that cause them to contaminate neighboring bands. Instead a holistic algorithm, analogous to FDH, is needed to deconvolve a time sequence of 2D images from the shadowgram of a chirped pulse. (Nakagawa et al., 2014) recently demonstrated an all-optical “motion picture femto-photography” method that recovered images of propagating lattice vibration waves or expanding plasmas with few-\( \mu \)m spatial resolution and frame interval ~200 fs. Extension of such methods to transverse wake shadowgraphy is a promising direction for future research.

2. Single-shot frequency-domain streak camera

(Li et al., 2014b, 2010) introduced an alternative way to record single-shot “movies” using temporally broad, chirped probe pulses propagating obliquely to the pump pulse, as shown in Fig. 36a,b. This approach generalized FDH, and was insensitive to slight temporal broadening that occurred in manipulating the probe pulses.

In Fig. 36a,b a chirped probe pulse crosses the path of a pump-driven object (“wake”) at angle \( \alpha \) inside a medium (“jet”) of thickness \( L \). In the probe frame, the evolving object sweeps across the probe pulse profile, imprinting a phase shift “streak” of length vector

\[
\vec{L}_{\text{streak}} = L[(\cos \alpha - v_{pr}/v_{ob})\hat{z}_{pr} + \sin \alpha \hat{y}_{pr}],
\]

that makes projection angle

\[
\phi = \tan^{-1} \left( \frac{v_{pr}v_{ob}/c}{v_{ob} - v_{pr}\cos \alpha} \right)
\]

with the object’s propagation direction \( \hat{z}_{ob} \). Here \( v_{ob} \) and \( v_{pr} \) denote lab frame group velocities of object and probe, respectively, in the medium. The streak is recovered by interfering the probe spectrally with a temporally advanced, equivalently chirped reference pulse (not shown in Fig. 36a,b), as in FDH (see Sec. IV.C.1). Figure 36c-e shows streaks of the nonlinear index of a pump transiting
a glass Kerr medium, recorded simultaneously on probe pulses propagating at 3 different angles. A series of line-outs perpendicular to each streak axis $L_{\text{streak}}$ constitutes a time sequence of the wake’s projections at angle $\phi$.

The streak resolves $N$ stages of the object’s evolution, where $N$ is the number of separated objects that can be lined up sequentially along $L_{\text{streak}}$ (e.g. $N = 5$ for the situation in Fig. 36b). Phase streaks with $\phi = 0$ (Fig. 36c), obtained with co-propagating pump and probe with $v_{pr} \neq v_{ob}$, reveal evolution of the object’s transverse profile, as it drifts longitudinally along the probe profile. This drift, a disadvantage in FDH because it blurs the recovered image, here provides valuable dynamic information. Streaks at $\phi = \pi/2$ (approximated by Fig. 36e) record evolution of the object’s longitudinal profile as it drifts sideways across the probe profile. Streaks at intermediate $\phi$ (Fig. 36d) record evolution of a diagonal profile. (Li et al., 2010) called this type of measurement a “frequency-domain streak camera” (FDSC).

By probing an evolving object simultaneously at several discrete angles, (Li et al., 2014b) reconstructed a time sequence of its entire 2D profile using tomographic algorithms. For example, Fig. 36f-j shows a 5-frame “movie” of the nonlinear index profile of a 10µJ, 30 fs laser pulse propagating through glass, reconstructed in one shot using five probes interfering with one reference pulse. Frames f-i show positive nonlinear index shift intensifying as the pump self-focuses, while in frame j a compensating negative shift appears in the center of the profile as a multi-photon-excited plasma appears. (Li et al., 2014b) called this extension of FDSC “frequency-domain tomography” (FDT). (Matlis et al., 2012) demonstrated a complementary spectrally-multiplexed tomography (SMT), that probed an object at a continuous range of angles by angularly dispersing the frequency components of an ultrafast probe pulse. After measuring the object-induced phase shift of each component by FDI, they reconstructed the position and cross-sectional structure of two quasi-stationary elliptical plasma filaments in one shot. Now that FDT and SMT have been demonstrated in such test experiments, their application to visualizing plasma wakes and plasma channels is a promising future direction.

FIG. 36 Color online. Imprint of phase-shift streak on chirped probe pulse propagating obliquely to evolving laser-driven wake. Beginning (a) and end (b) of imprint, showing lab (α) and projection (φ) angles. Middle: 3 phase streaks from pump pulse transit through glass Kerr medium, recorded simultaneously for $\alpha (\phi) = 0.1^\circ (15^\circ)$ (c); $1.4^\circ (27^\circ)$ (d); $9.7^\circ (68^\circ)$ (e). Pump entrance into medium corresponds to right-most end of each streak. Bottom: 5-frame “movie” of pump-induced index change profile $\Delta n$ at normalized propagation distances $z_{ob}/L = 0.17$ (f), 0.33 (g), 0.49 (h), 0.65 (i), 0.81 (j) into Kerr medium, reconstructed tomographically from 5 streaks. Adapted from (Li et al., 2014b).

FIG. 37 Color online. FDSC study of nonlinear LWFA dynamics, showing data from shot (a) (left column) and (b) (right) at $n_c = 2 \times 10^{19}$ cm$^{-2}$. Top row: electron spectra, showing quasi-monoenergetic 80 MeV electrons (a-1), and no electrons (b-1). Middle, bottom rows: reconstructed probe phase shift $\Delta \phi_{pr}$. Pump drifts from lower right to upper left as it transits jet, imprinting nearly uniform background plasma-induced phase-shift $\Delta \phi$ (yellow-orange area) as it ionizes gas. But only (a-2) shows narrow phase-shift dip $\Delta \phi$ (yellow streak, highlighted by white arrow in (a-2) and line-outs in (a-3)) imprinted by evolving bubble. Dashed lines in (a-3) and (b-3) show drift trajectory of center of pump pulse. Adapted from (Li et al., 2014a).

Meanwhile, (Li et al., 2014a) used single-probe FDSC
to characterize formation, propagation, and lengthening of a laser-generated plasma bubble as it accelerated electrons quasi- monoenergetically to $\sim 100$ MeV. Fig. 37 summarizes results for 2 contrasting shots, one (left column) that produced a high-quality electron bunch (a-1), the other (right) no relativistic electrons (b-1). During both shots, a loosely-focused chirped probe pulse of bandwidth $\Delta \lambda_{pr} \approx 10$ nm crossed the path of a tightly-focused drive pulse in a He gas jet of length $L \approx 3$ mm at angle $\alpha = 8.6^\circ$. At this $\alpha$, plasma structures drifted across the probe profile at $\phi \approx 90^\circ$. For both shots, the quasi-static, pump-ionized plasma column imprinted a wide uniform background phase shift $\Delta \phi_p \sim 10$ mrad (orange areas in Fig. 37, middle row). Uniquely on shot (a), a dynamic bubble of diameter $\lambda_p \sqrt{\alpha_0 / \pi} \sim 10 \mu$m (Lu et al., 2007) in the pump’s immediate wake generated a narrow phase-shift streak $\Delta \phi_b \sim -2$ mrad of opposite sign, due to the absence of free electrons inside the bubble and the dearth of electrons in the bubble’s side (relative to its front and rear end) walls. Thus, after the pump blew out a fully-formed bubble, which occurred at $z \approx 2L/3$ for the case shown in Fig. 37a, $\Delta \phi_b$ carved a narrow dip (yellow streak in Fig. 37a-2, highlighted by arrow, lineouts in a-3) into the broader $\Delta \phi_p$ profile, exhibiting a time sequence of longitudinal projections of the plasma bubble. On shot (b), no $\Delta \phi_b$ phase dip was observed, i.e. no bubble formed, consistent with no relativistic electrons.

(Li et al., 2014a) observed variations in the length, depth and shape of the $\Delta \phi_b$ streak, and in corresponding electron spectra, as $\bar{n}_e$ (and thus $P/P_{cr}$) changed. For example, at 15% lower $\bar{n}_e$ than in Fig. 37a, the bubble formed fully only near the end of the jet ($z \approx 5L/6 \approx 2.5$ mm) due to weaker self-focusing, leaving acceleration length ($L/6 \approx 0.5$ mm) shorter than the dephasing length $L_d$. Consequently acceleration was incomplete, and broad low-energy electron spectra were observed. Conversely, at 10% higher $\bar{n}_e$ than in Fig. 37a, the bubble formed at $z < L/2$, leaving acceleration length $> 4L_d$. This also yielded a broad electron energy spectrum and poor beam quality, due to strong dephasing. Stable, nearly mono-energetic electron beams were observed only in a narrow range $\bar{n}_e = 2.0 \pm 0.1 \times 10^{19}$ cm$^{-3}$, for which FDSC showed acceleration length $\sim 1.5L_d$, assuming injection coincided with full bubble formation. Simulations confirmed this optimal acceleration length, somewhat greater than $L_d$. Evidently moderate dephasing helps to compress the electron spectrum by decelerating the fastest electrons just enough for the slowest electrons to catch up (Yi et al., 2010). Among other LWFA physics, (Li et al., 2014a) also observed that bubbles lengthened as they finished accelerating electrons, a signature of beam loading, and that bubbles were only partially evacuated at the injection threshold.

Simulations show that FDSC can access additional LWFA physics by using wider bandwidth probe and reference pulses. As an example, Fig. 38a shows the 3D wake electron density profile at $z = 2L/3 = 2$ mm for conditions corresponding to data in Fig. 37a, simulated using the PIC code Virtual Laser Plasma Lab (VLPL) (Pukhov, 1999). The primary accelerating cavity, with injected electrons at its rear, and two trailing cavities separated by electron-density sheaths of $\sim 1 \mu$m thickness, are visible. Fig. 38b shows a $z = 2L/3$ lineout of the $\Delta \phi_b$ streak of a simulated FDSC experiment, using a 400 nm probe of unlimited bandwidth (dashed red curve) and with finite bandwidths $\Delta \lambda_{pr} = 100$ nm (dashed blue) and 10 nm (solid blue). The last curve agrees well with the measured FDSC line-out (black curve) from Fig. 37a. However, both average over sharp electron density spikes separating the three cavities. Direct observation of these sheaths is an important future diagnostic goal, since subtle sheath thickness variations govern electron injection (Yi et al., 2010). On the other hand, a simulated $\Delta \lambda_{pr} = 100$ nm probe resolves these spikes well (dashed blue curve). Visible supercontinuum probe-reference pulses (Kim et al., 2002a) provide the necessary bandwidth, and should resolve $\mu$m-size structures in plasma accelerators that do not themselves produce strong background supercontinuum $\alpha$ (Ralph et al., 2009) at the probe angle and wavelength range. Otherwise, ultraviolet supercontinuum probe-reference pulses and/or large $\alpha$ will be necessary for high-resolution FDSC.

3. Single-shot imaging of meter-long wakes

FDSC studies described in the previous sub-section relied on the assumption that a phase modulation $\Delta \phi_{pr}(z_0)$ of width $w_{ob}$ that an object imprinted locally on a probe pulse at position $0 < z_0 < L$ within a medium preserved its shape until it reached the medium’s exit plane ($z = L$). An imaging system then relayed it to a detector. If, however, the distance $\Delta z \equiv L - z_0$ had exceeded diffraction length $\Delta z_{dif f} \approx \pi w_{ob}^2 / \lambda_{pr}$, then the imprinted phase modulation would have diffracted before reaching the exit. Thus e.g. a $\lambda_{pr} = 0.4 \mu$m probe can faithfully record the evolving shape of a plasma bubble of radius...
\( w_{ob} = 10 \mu m \) (see Fig. 38a) over propagation distance \( \Delta z \sim 1 \text{ mm} \), a limit that the experiment depicted in Fig. 37 barely satisfied, since the bubble formed \( \sim 1 \text{ mm} \) from the jet exit. This places an additional limit on resolving e.g. bubble sheath structures, over and above the limit set by probe bandwidth.

The depth of field \( \Delta z_{dof} \) of the imaging system that relays the wake-modulated probe(s) from \( z = L \) to the detector must similarly be matched to \( \Delta z_{diff} \) of the smallest sub-structure one wishes to resolve, otherwise portions of the imprinted streak will be out of focus at the detector. FDSC of plasma accelerators of length \( \Delta z > \Delta z_{diff} \) encounters depth-of-field limits. Examples include multi-GeV LPAs of multi-cm acceleration length (Leemans et al., 2014; Wang et al., 2013), meter-scale e-beam-driven PWFAs (Blumenfeld et al., 2007), and proton-driven PWFAs hundreds of meters long (Caldwell et al., 2009; Geschwendtner et al., 2016).

![Diagram](image_url)

**FIG. 39** Color online. Visualization of multi-cm plasma accelerator structures. (a) Top: schematic MOPI setup. Adapted from (Li et al., 2013b). Middle, triple panel: intensity modulation of 100 fs probe that propagated at angle 4\(^\circ\), delay \( \Delta t \approx 100 \text{ ps} \) behind a 20 GeV, 2 nC, 100 fs electron bunch in hydrogen plasma of density \( n_e = 3 \times 10^{17} \text{ cm}^{-3} \), imaged from 3 object planes to each of 3 CCD cameras. Lower panel: image of unstable portion of electron trajectory captured by CCD1. (b) Plot of continuous phase shift \( \Delta \phi_{pr}(x, y, z) \) induced by \( n_e \approx 10^{16} \text{ cm}^{-3} \) plasma over 10 cm path on 100 fs probe pulse that propagated at angle 1\(^\circ\), \( \Delta t \approx 1.7 \text{ ps} \) behind a 3 mJ laser pulse in air, reconstructed from 4 CCD images using Gerchberg-Saxton algorithm. Panel b) adapted from (Li et al., 2013b).

Multiplexed transverse optical shadowgraphy (Sec. IV.B.2.3) or electron radiography (Sec. IV.B.4) provides one potential solution to single shot imaging of plasma structures without depth-of-field limits, but will require maintaining few-fs duration of multiple replicas of the incoming probe. Alternatively, (Li et al., 2013b) introduced multi-object-plane imaging (MOPI) of a small-\( \alpha \) probe, illustrated in Fig. 39. MOPI multiplexes detection of a single probe after, rather than before, it interacts with the plasma structure. Fig. 39a (top) schematically depicts the setup. A collimated probe pulse (gray) crossed the path of a synchronized wake driver (red) at angle \( \alpha \), with beam waist \( w_0 \), illuminating wake structures over propagation distance \( z \sim w_0/\alpha \). In contrast to FDSC, (Li et al., 2013b) used a probe pulse compressed to \( \sim 100 \text{ fs} \). Thus it characterized only a \( \sim 100 \text{ fs} \) longitudinal slice of the pump-generated structure. The illuminated slice swept transversely across the compressed probe profile, mapping its index profile at propagation distance \( z \) onto transverse position \( x \) on the probe profile via the relation \( x = \alpha z \). Tilting the probe intensity front by angle \( \alpha/2 \) with a prism prevented the phase streak from walking off of the compressed probe profile, extending the propagation distance that could be probed. After the interaction, copies of the phase-modulated probe created by beam splitters were imaged through a 4\( f \) system from MOPs along the pump path to corresponding image planes. Phase-contrast imaging (PCI), which (Li et al., 2013b) implemented by placing a thin Kerr medium at the Fourier plane of the 4\( f \) system (see Fig. 39a), converted phase modulations to easily detectable amplitude modulations. CCD cameras recorded “bow-tie”-shaped images (Fig. 39a, triple panel), in which “knots” corresponded to object planes, wider wings to nearby out-of-focus regions.

Transverse line-outs of the “knots” of recorded bow-tie intensity patterns straightforwardly yield phase shifts \( \Delta \phi_{pr}(x_{ob}, z_i) \) that the plasma structure induced on the probe at selected object planes \( z_i \). (Li et al., 2013b) reconstructed phase shifts at intervening values of \( z \) by iteratively fitting the complete diffraction patterns of overlapping bow-ties using a Gerchberg-Saxton algorithm (Fienup, 1982). Fig. 39b shows an example of a continuous phase-shift streak \( \Delta \phi_{pr}(x_{ob}, z) \) over a 10-cm propagation path, thus reconstructed from four MOPI bow-tie images of a probe that propagated at \( \Delta t = 1.7 \text{ ps} \) behind a pump pulse propagating in air. The main feature in this reconstructed phase profile originates from plasma of density \( n_e \approx 10^{16} \text{ cm}^{-3} \), demonstrating high sensitivity of MOPI + PCI to tenuous plasma structures.

(Zgadzaj et al., 2016) have begun to use MOPI to image plasma wake structures driven by 20 GeV, 2 nC electron bunches over meter-scale propagation distances at SLAC’s FACET. The bottom panel in Fig. 39a shows a “bow-tie” image of a 15-cm segment of the unstable propagation path of a 20 GeV e-bunch through laser-ionized hydrogen plasma, observed at \( \Delta t = 100 \text{ ps} \) after passage of the bunch. At this late delay, the original electron wake had transferred most of its energy into an “ion wake” (Lotov et al., 2014; Sahai et al., 2016; Vieira et al., 2012), that nevertheless preserved some features of the
E. Scaling of wake probes with plasma density

Most single-shot optical wake diagnostics reviewed here were demonstrated on MeV plasma accelerators at $n_e \sim 10^{19}$ cm$^{-3}$. However, future plasma accelerator research will focus on multi-GeV accelerators at $10^{17}$ cm$^{-3} < n_e < 10^{18}$ cm$^{-3}$. Here we consider how these diagnostic methods scale as $n_e$ decreases.

For transverse optical probes, signal strength is proportional to plasma structure width $L_\perp$, which scales as $\lambda_p \sim n_e^{-1/2}$. For shadowgraphs, plasma structure visibility is proportional to local probe deflection angle

$$\Delta \theta_{def} = \int_0^{L_\perp} \frac{dn}{dz} ds \approx \frac{dn}{dz} L_\perp \propto \frac{n_e}{n_{cr}},$$

where $dn/dz$ is the local gradient of refractive index $\eta = (1 - n_e/n_{cr})^{1/2} \approx 1 - n_e/2n_{cr}$ along the wake propagation axis $z$, approximated as a constant over $L_\perp$ in the penultimate expression of Eq. (57). For local index change $\Delta \eta = \Delta n_e/2n_{cr}$ over distance $\Delta z$, $\Delta n_e \propto n_e$ and $\Delta \eta \propto n_e^{-1/2}$ as $n_e$ changes. Taking $dn/dz \propto \Delta n_e/\Delta z$, and scaling $L_\perp$ as noted above, we obtain the density scaling in the final expression of Eq. (57): equivalent shadowgraph contrast corresponds to constant $n_e/n_{cr}$. Since $n_{cr} \propto \lambda_p^{-2}$, equivalent contrast is maintained by scaling $\lambda_p \sim n_e^{-1/2}$ as $n_e$ changes. Fig. 40 presents simulated shadowgraphs of wakes in plasma of density $n_e = (a) 1.7 \times 10^{19}$ cm$^{-3}$ and (b) $0.48 \times 10^{19}$ cm$^{-3}$ that confirm (57). Probe wavelength $\lambda_{pr} = 0.75 \mu m$ yields a high-contrast shadowgraph (c) at the higher $n_e$, but low contrast at the lower $n_e$ (d). Shifting to $\lambda_{pr} = 1.4 \mu m$ recovers high contrast at the lower $n_e$ (e).

![Fig. 40 3D PIC simulations showing scaling of transverse optical shadowgraphy with plasma density $n_e$. Top row: electron density profiles of plasma wakes driven by $\lambda_{na} = 0.8 \mu m$, $\sim 0.7$ J, 30 fs pulses in $n_e = (a) 1.7 \ (\lambda_p = 9 \mu m)$ and (b) $0.48 \times 10^{19}$ cm$^{-3} \ (\lambda_p = 17 \mu m)$ plasma. Second and third rows: simulated transverse shadowgraphs using $\lambda_{pr} = 0.75 \mu m$ [panels (c), (d)], and $1.4 \mu m$ [panel (e)]. Courtesy E. Siminos.](image-url)

The condition (57) implies constant $\lambda_{pr}/\lambda_p$. Thus spatial imaging resolution relative to feature size remains constant, even though absolute resolution scales with $\lambda_p$. Comparison of panels (c) and (e) of Fig. 40 confirms that constant $n_e/n_{cr}$ preserves feature resolution.

Similar arguments applied to Eq. (53) show that the sensitivity of Faraday rotation also depends on $n_e/n_{cr}$, although the strength of $\vec{B}$ remains an independent scaling parameter. For an interferometric probe propagating at angle $\alpha$ with respect to the wake (e.g. FDSC, MOPI), signal strength is given by local probe phase shift

$$\Delta \phi_{pr} = \frac{2\pi}{\lambda_{pr}} \int_0^{L_\perp/\sin \alpha} [\eta(s) - 1] ds \approx \frac{\pi}{\sin \alpha} \frac{L_\perp}{\lambda_{pr}} \frac{n_e}{n_{cr}}$$

integrated over the probe’s oblique path across the wake, along which the last expression in (58) assumes constant $n_e \propto n_e$. The factor $L_\perp/\lambda_{pr}$ now appears in addition to $n_e/n_{cr}$. However, constant $n_e/n_{cr}$ implies constant $L_\perp/\lambda_{pr}$, and thus constant $\Delta \phi_{pr}$. The above remarks about resolution carry over without change. Thus $n_e/n_{cr}$ is a universal scaling factor for transverse optical probes.

For longitudinal optical interferometric probes (e.g. FDI, FDH), the phase shift is

$$\Delta \phi_{pr} = \frac{2\pi}{\lambda_{pr}} \int_0^{L_\parallel} [\eta(s) - 1] ds \approx \frac{\pi L_\parallel}{\lambda_{pr}} \frac{n_e}{n_{cr}},$$

where again the last expression assumes constant $n_e \propto n_e$ along the probed length $L_\parallel$. Here the limiting value of $L_\parallel$ determines $n_e$ scaling. Several possibilities can arise. If the gas cell (or pump Rayleigh) length $L$ limits the interaction and remains constant as $n_e$ changes, then
Eq. (59) yields $\Delta\phi_{pr} \propto \bar{n}_e$ for fixed $\lambda_{pr}$, $\Delta\phi_{pr} \propto \bar{n}_e^{1/2}$ for fixed $\bar{n}_e/n_{cr}$, or $\Delta\phi_{pr}$ constant for $\lambda_{pr} \propto \bar{n}_e^{-1}$. If $L$ can be adjusted as $\bar{n}_e$ changes, then often longitudinal drift of probe-driver delay due to group-velocity mismatch $\Delta v_g = v_g^{\text{(pr)}} - v_g^{\text{(dr)}}$, which must be limited to a small fraction $f \ll 1$ of $\lambda_{pr}$, to avoid washing out longitudinal wake structure, determines $L_{\parallel}$. This limit is common for LWFAs, where $\lambda_{pr}$ must differ significantly from $\lambda_{pu}$ so that it can be spectrally filtered from forward-scattered pump light, but would also arise for PWFAs, where the velocity ($\approx c$) of a relativistic particle bunch driver would always exceed $v_g^{\text{(pr)}}$. In this case the allowable probe-driver drift occurs after propagation time $f \lambda_{pr}/\Delta v_g$, yielding $L_{\parallel} = cf\lambda_{pr}/\Delta v_g$. For LWFAs in which $\lambda_{pr}$ and $\lambda_{pu}$ remain fixed as $\bar{n}_e$ changes, $\Delta v_g \propto \bar{n}_e$ and $L_{\parallel} \propto \bar{n}_e^{-3/2}$, equivalent to the $\bar{n}_e$ scaling of electron dephasing length $L_d$ (see Eq. (6)). Then $\Delta\phi_{pr} \propto \bar{n}_e^{-1/2}$ — *i.e.* signal strength increases with decreasing density. (Dias et al., 1998; Kasim et al., 2015) discuss other cases that can arise if $\lambda_{pr}$ and/or $\lambda_{pu}$ scale systematically with $\bar{n}_e$, or when $L_d$ or $L_{pd}$ (Eq. (7)) determine $L_{\parallel}$.

**V. CONCLUSION**

Approximately 10% of the budget of any accelerator is devoted to diagnostics. No accelerator can operate without them. The advent of plasma accelerators with cavities of dimensions $10 \mu m \lesssim \lambda_p \lesssim 100 \mu m$ that accelerate lepton bunches of dimensions $\sigma \ll \lambda_p$ has posed unprecedented diagnostic challenges. Not only are these cavities and bunches much smaller than their conventional counterparts, but they evolve rapidly during acceleration. Moreover, plasma cavities are transient and light-speed, and thus extraordinarily difficult to visualize accurately in the laboratory. Theory and simulations have provided essential, yet incomplete, guidance in understanding plasma accelerators. Predictions of bunch size varied widely before accurate measurements were carried out, while the dynamics of *e.g.* 3D wave-breaking, electron self-injection, and highly nonlinear wake evolution remained incompletely solved theoretical problems. Innovative laboratory diagnostics have filled many gaps in our understanding, and helped transform plasma electron acceleration from a fringe empirical activity to a quantitative science at the center of international planning for next-generation light sources and colliders.

The coming transition from prototype acceleration experiments to operating accelerators will place new demands on stability, energy spread and emittance of electron bunches, and thus on diagnostics. Early designs of plasma-accelerator-based light sources (Maier et al., 2012) and colliders (Adli et al., 2013; Leemans and Esarey, 2009) have already made clear that control of emittance growth and charge loss when transporting bunches from LWFA to undulator, or between collider stages, will be major challenges. Development of versatile, accurate, noninvasive bunch diagnostics in transport regions will be as important in future research as intrastage diagnostics in past research. For colliders, diagnostics will have to be replicated over hundreds of stages. This will favor those that are simple, reliable and low-cost while still achieving high spatial and temporal resolution. Moreover, integration of these diagnostics into machine learning systems and genetic algorithms (He et al., 2015; Wu et al., 2017) for feedback control and optimization of accelerator performance will become a focus of research.

The greatest bunch diagnostic challenges that plasma accelerators posed were measurements of ultrasmall transverse emittance ($\varepsilon_n < 1 \text{ mm rad}$) and ultrashort duration ($\tau_b \sim \text{ few fs}$) of electron bunches that strongly nonlinear LWFAs could uniquely produce (Secs. II.C.2, II.C.3). Ultrasmall bunch dimensions were a source of numerical instabilities in simulations (Lehe et al., 2013), requiring small grids and time-steps as well as algorithmic advances for realistic results. Laboratory diagnostics therefore became the primary source of accurate information. Early efforts to measure $\varepsilon_n$ of nonlinear LWFA bunches adapted the conventional pepper-pot method. However, it sampled phase space too sparsely, was limited in electron energy range, and was invasive (Sec. III.C.1.a). Similarly, early efforts to measure $\tau_b$ adapted EO methods in wide use for measuring compressed ps-duration bunches from conventional accelerators. However, strong THz dispersion of EO crystals limited even the most innovative measurements to $> 10 \text{ fs}$ RMS resolution when applied to nonlinear LWFA bunches (Sec. III.D.1). Thus these conventional beam diagnostics, pushed to their limits, managed only to set upper bounds on $\varepsilon_n$ and $\tau_b$.

The most important breakthroughs in beam diagnostics that plasma accelerators spurred in the past decade were noninvasive methods for resolving $\varepsilon_n$ and $\tau_b$ in one shot. Betatron x-ray spectroscopy first found evidence of bunch radii $\sigma_r \sim 0.1 \mu m$ inside the LWFA. Full trace-space analysis methods reconstructed complete $\varepsilon_n$, including the correlation term, in one shot without invoking a downstream divergence measurement, and found values up to an order of magnitude smaller than pepper-pot methods had resolved (Sec. III.C.2). New spectrally-resolved approaches to traditional quadrupole focus-scan methods avoided chromatic distortions, enabled single-shot measurements, and resolved $\varepsilon_n$ values *outside* the accelerator as small as those that betatron spectroscopy measured inside (Sec. III.C.1.b).

Meanwhile, new methods for resolving few-fs $\tau_b$ emerged. MO measurements of Lorentz-contraction magnetic fields of relativistic bunches in plasma resolved $\tau_b$ down to $\sim 5 \text{ fs}$ (Buck et al., 2011), free of phonon-dispersion limits of EO crystals. Combined with inverse optical shadowgraphy (Sec. IV.C.3.b), MO meth-
ods localize the accelerating bunch within an LWFA, but require probe pulses $\lesssim \tau_0$ in duration (Secs. III.D.1.c and IV.C.3.a). The use of co-propagating infrared light pulses as transverse deflectors (TDs) of period $\lambda \sim 1\mu m$ has, in combination with a magnetic spectrometer, yielded single-shot measurements of internal energy-time structure of few-fs chirped LWFA bunches (Zhang et al., 2016a). Though technically challenging, such micro-TD experiments directly measure slice emittance and slice energy spread, key parameters for LWFA-driven FELs and colliders (Sec. III.D.1). Finally, researchers transformed transition radiation (TR), a beam diagnostic of few-fs LWFA bunches (Bakkali et al., 2016). TR spectroscopy complements time-domain MO and TD methods by characterizing $\tau_0$, outside the accelerator without probe pulses (Sec. III.D.2).

The biggest future challenge facing plasma accelerator beam diagnostics is to develop single-shot methods for recovering sub-$\mu m$ transverse, as well as few-fs longitudinal, beam profiles simultaneously and with high resolution throughout beam transport lines, and over multiple accelerator stages. This challenge will increase as electron energies reach multi-GeV. CTR methods show the greatest promise for this task because of their demonstrated ability to profile bunches both longitudinally and transversely with high resolution outside the accelerator, and because of the relatively low cost of the components. Combinations of CDR and Smith-Purcell methods with multi-octave spectroscopy appear promising for bringing about the required marriage of spatial and temporal diagnostics. In addition, all diagnostics must ultimately be applied to plasma-accelerated positron bunches. The challenges that plasma accelerators posed for plasma structure diagnosis differed from those for beam diagnostics. Theory predicted the $\lambda_p$ dimension of wake structural units with certainty from the beginning, and analytic 2D theories of the formation and internal morphology of linear and mildly nonlinear wakes were available early in the development of plasma electron accelerators (Chen, 1985; Esarey et al., 1989; Gorbunov and Kirsanov, 1987). The main structural features of wakes that 3D simulations predicted even in the strongly nonlinear broken-wave regime (Pukhov and Meyer-ter-Vehn, 2002) were never in serious doubt, and most were large enough to be optically resolvable. The challenge lay rather in 4D visualization — i.e. 3D structure plus time evolution — a photography problem that Edwaerd Muybridge solved for galloping horses in the 19th century (Clegg, 2007), but which remained unsolved for light-speed objects in the 21st century (Li et al., 2014b; Pleasants, 2014). Divergent spatial-temporal scales — i.e. internal wake structural dynamics measured in $\mu m$ and fs, propagation in m and ns — together with low optical contrast and demand for single-shot visualization heightened the methodological challenge. The physics challenges were to observe difficult-to-simulate nonlinear features such as bubble formation and evolution, waveform curvature, wave-breaking, self-injection and beam-loading dynamics. A practical challenge was to monitor shot-to-shot variations in wake structure and dynamics arising from non-ideal drivers or targets, which are often not fully included in idealized simulations.

Early wake diagnostic experiments drew upon established methods from other fields. Collective light scatter, a long-established diagnostic of holistic properties ($\omega_p, k_p, \delta n_e/\bar{n}_e$) of electron plasma waves (Froula et al., 2011), helped early researchers discover laser wakefield growth and decay processes that remain important today, but did not resolve sub-$\lambda_p$ wake structure (Sec. IV.A). Likewise pump-probe experiments using fully compressed, co-propagating optical or electron pulses, a staple of ultrafast science, first resolved sub-$\lambda_p$ structure of plasma wakes, and observed variations in the length $\lambda_p$ and amplitude $\delta n_e/\bar{n}_e$ of individual periods within a long wake, but required multiple shots and long, painstaking data acquisition (Sec. IV.B).

The singular new advances in wake diagnostics that plasma accelerators spurred were single-shot methods that resolved detailed sub-$\lambda_p$ structure of plasma wakes. Frequency-domain holography (FDH), which recovers temporal phase $\phi_{pr}(t)$ of long, but spectrally broad, co-propagating probes pulses, captured “snapshots” of relativistic curvature (Matlis et al., 2006) and amplitude spiking (Dong et al., 2010a) of plasma wavefronts within mildly nonlinear wakes, and the variation of these effects from period to period (Sec. IV.C.1). Two generalizations of FDH expanded its diagnostic functionality: frequency-domain shadowgraphy (FDS), which recovers the probe’s temporal amplitude $|E_{pr}(t)|$, captured snapshots of strongly nonlinear plasma wakes (Dong et al., 2010b) (Sec. IV.C.2); and frequency-domain streak camera (FSC), in which the probe propagates obliquely, captured “movies” of projections of a laser-driven bubble forming, deepening and expanding as it propagated (Li et al., 2014a) (Sec. IV.D.2). Synchronized transversely-propagating optical probes of few-fs duration projected shadowgraphs related to the wake’s instantaneous density profile (Buck et al., 2011) onto a camera (Sec. IV.C.3.b). A change of pump-probe delay over successive shots revealed structural changes during propagation, and correlated these changes with injection of electrons (Sivert et al., 2015), rendered visible by Faraday rotation (Kaluza et al., 2010) (Sec. IV.D). Electron probes profiled the wake’s instantaneous electric field profile (Clayton et al., 2016; Zhang et al., 2017, 2016b), and proved sensitive at densities down to $\bar{n}_e \lesssim 10^{17} \text{cm}^{-3}$ (Sec. IV.C.4). Computer simulations benchmarked im-
ages obtained with this diverse and unprecedented suite of single-shot probes in detail, opening a rich new line of communication between experimental and theoretical plasma accelerator science.

Future challenges for plasma wake diagnostics lie in three main directions. One is full realization of the potential for single-shot 4D visualization — i.e. experimental output of a propagation (\(z\)) sequence of instantaneous wakefield density \(n_e(x,y,z-v_g t,z)\) or field \(E_e(x,y,z-v_g t,z)\) profiles. Such outputs are similar to snapshots from PIC simulations, and thus might be called “PIC-tures”. Frequency-Domain Tomography (FDT) (Li \textit{et al.}, 2014b) and Spectrally-Multiplexed Tomography (SMT) (Matlis \textit{et al.}, 2012) are promising approaches that have been demonstrated on test index objects, but must be extended to plasma wakes using wide-bandwidth probes (Sec. IV.D.2). Another promising path is multiplexing of transverse optical probes, enabling capture of a \(z\)-sequence of transverse shadowgraphs and Faraday-rotation images in one shot. This approach uniquely observes the wake’s shadow, the accelerated electron bunch, and wave-breaking radiation with a common apparatus, enabling correlation of diverse features of the plasma acceleration process (Secs. IV.C.3).

A second future direction is incorporating advanced diagnostics and their results directly into simulations. This is essential not only for data analysis, but for making testable predictions. One frontier is “synthetic diagnostics”: simulating interactions of diagnostic probes with, or radiation from, the accelerator, not by post-processing a sub-set of previously simulated results, but \textit{during} a simulation, when all physics quantities are accessible in memory. Another frontier is computational solution of inverse problems. Already well advanced in photon science, astronomy or geophysics, this approach uses diagnostic data from diverse sources to reconstruct quantities of interest — \textit{e.g.} reconstructing electron bunch profiles using several diagnostics, even though part of the information \((\textit{e.g. spectral phase})\) is not directly measurable (see Sec. III.D). This approach is especially powerful when analyzing diagnostics that are related by known physics, such as laser or electron beam profiles separated by propagation through vacuum.

A third future direction is expanding the application of single-shot wake diagnostics to a wider range of drivers, densities and contexts. Even though beam-driven PWFA were among the first to be diagnosed with sub-\(\lambda_p\) precision (Sec.IV.B.1), researchers are now only beginning to apply modern single-shot optical wake diagnostics to strongly nonlinear PWFA (Zgadzaj \textit{et al.}, 2016). This application brings the added methodological challenges of imaging sub-mm-wide wakes over multi-meter-long acceleration paths, synchronizing optical and electron probes with e-beam drivers, and probing wakes at lower density plasma (\(n_e < 10^{17} \text{ cm}^{-3}\)) than in many past experiments. (Sec. IV.D.3). New physics challenges include visualization of differences among electron-, positron- and proton-driven wakes, characterization of beam-plasma instabilities, and probing of ion wakes on a ns time scale after excitation. Pulses from emerging TW-peak-power CO\(_2\) lasers (Polyanskiy \textit{et al.}, 2015), because of the large ponderomotive force Eq. (1) they exert on a plasma for given pulse energy, duration and spot size, will similarly be able to drive bubble-regime wakes of hundreds of microns diameter in \(n_e < 10^{17} \text{ cm}^{-3}\) plasma. Such bubbles “write large” offer the possibility not only of precisely injecting synchronized, low \(\Delta E_e/E_e\), low-\(\varepsilon_n\) bunches from conventional linacs into plasma accelerators, but of probing their internal structures and evolution with higher resolution than in past experiments. Such challenges will continue to spur creativity and innovation in diagnostics for plasma electron accelerators for years to come.

\textbf{Acknowledgments}

M. C. D. acknowledges support from the Alexander von Humboldt Foundation during preparation of this manuscript, as well as support from the U. S. Department of Energy (grants DE-SC0011617, DE-SC0014043, DE-SC0014043), National Science Foundation (PHY-1734319), Air Force Office of Scientific Research (grants FA9550-16-1-0013, FA9550-14-1-0045), and the Robert Welch Foundation (grant F-1038).
REFERENCES


