New density estimation methods for charged particle beams with applications to microbunching instability

Balsˇa Terzic´ *
Jefferson Lab, Newport News, Virginia 23606, USA

Gabriele Bassi†
Brookhaven National Laboratory, Upton, New York 11973, USA

(Received 13 May 2011; published 8 July 2011)

In this paper we discuss representations of charge particle densities in particle-in-cell simulations, analyze the sources and profiles of the intrinsic numerical noise, and present efficient methods for their removal. We devise two alternative estimation methods for charged particle distribution which represent significant improvement over the Monte Carlo cosine expansion used in the 2D code of Bassi et al. [G. Bassi, J. A. Ellison, K. Heinemann, and R. Warnock, Phys. Rev. ST Accel. Beams 12, 080704 (2009); G. Bassi and B. Terzic´, in Proceedings of the 23rd Particle Accelerator Conference, Vancouver, Canada, 2009 (IEEE, Piscataway, NJ, 2009), TH5PFP043], designed to simulate coherent synchrotron radiation (CSR) in charged particle beams. The improvement is achieved by employing an alternative beam density estimation to the Monte Carlo cosine expansion. The representation is first binned onto a finite grid, after which two grid-based methods are employed to approximate particle distributions: (i) truncated fast cosine transform; and (ii) thresholded wavelet transform (TWT). We demonstrate that these alternative methods represent a staggering upgrade over the original Monte Carlo cosine expansion in terms of efficiency, while the TWT approximation also provides an appreciable improvement in accuracy. The improvement in accuracy comes from a judicious removal of the numerical noise enabled by the wavelet formulation. The TWT method is then integrated into the CSR code [G. Bassi, J. A. Ellison, K. Heinemann, and R. Warnock, Phys. Rev. ST Accel. Beams 12, 080704 (2009)], and benchmarked against the original version. We show that the new density estimation method provides a superior performance in terms of efficiency and spatial resolution, thus enabling high-fidelity simulations of CSR effects, including microbunching instability.

DOI: 10.1103/PhysRevSTAB.14.070701

PACS numbers: 07.05.Tp, 52.65.Rr, 02.60.Gf, 41.75.−i

I. INTRODUCTION

When a charged particle beam travels along a curved trajectory (e.g., bending magnet), it emits synchrotron radiation. If the wavelength of the synchrotron radiation is longer than the bunch length, the resulting radiation is coherent. Therefore, coherent synchrotron radiation (CSR) is the low frequency part of the radiation power spectrum, and is more pronounced when the beam bunch is short. Short beam bunch lengths are desirable in many different contexts—free electron lasers (FELs), energy recovering linacs, B-factories—and their importance is only expected to grow in the future. This necessitates a development of a trustworthy code which is capable of properly simulating CSR and its effects, such as an appreciable degradation of the beam’s emittance and its fragmentation. For a recent study of CSR effects and its modeling see [1], where measurements of CSR-induced energy loss and transverse emittance growth have been done for the two bunch compressors of the Linac Coherent Light Source and compared with particle tracking codes. Despite the good agreement found between measurements and simulations, more studies are needed to validate numerical codes, especially in the study of the microbunching instability, where CSR can be emitted also at wavelength shorter than the bunch length if the beam has high-frequency density modulations caused by instability itself. This effect was first observed in numerical simulations with the code ELEGANT and led to a serious concern for its impact on FEL performance [2]. Several theoretical studies advanced our understanding of the microbunching instability [3–5] leading to the development of various suppression schemes [6], while simple numerical methods have been successful in simulating its basic mechanism [7]. Encouraged by these advancements, the FEL community set out to organize a series of workshops [8] to further our theoretical understanding and numerical modeling of the microbunching instability, as well as improve experimental observations and diagnostic techniques. In addition to having detrimental effects in
bunch compressors, the microbunching instability is a concern for manipulation schemes such as high harmonic generation for FELs that require accurate modeling of very small-scale structures.

The numerical method adopted in this paper to model CSR effects is based on a Green’s function approach. Its use faces two major difficulties: (i) storing the history of the beam bunch is very memory intensive for simulations with reasonable spectral resolution; and (ii) the manner in which the computation of the beam’s self-interaction scales.

The numerical study of the CSR effects can be made computationally feasible by introducing different approximations/simplifications. The line-charge (1D) approximation [3] of a beam bunch is one such approximation, in which the CSR effects are likely overrated by the collapse of the transverse distribution onto a rigid line. Point-to-point approaches mimic the self-interaction of the beam by direct interaction among N macroparticles which represent the beam charge density [9,10]. The drawback here is that the computation of the self-interaction scales as the square of the number of macroparticles, which gets prohibitively expensive rather quickly, thus limiting N to the order of a thousand. This, in turn, yields poor spectral resolution unable to account for small-scale structures present in CSR simulations. The mean-field (mesh, grid, particle-in-cell) approach alleviates this problem by replacing direct interaction of the particles with the interaction of the particles with the mean field computed from depositing the particles onto a finite grid. This results in an algorithm which scales better with the number of particles (linear instead of square in the point-to-point approach) and allows for a superior spatial resolution.

In this study we improve the 2D CSR mean-field code developed by Bassi et al. [11]. The code is based on integration of the retarded potential for a 2D charge distribution (no vertical size), considering the shielding effects by infinite parallel plates. In the original implementation of [11], the particle distribution sampled by N point-charge particles is first approximated by the cosine expansion, then evaluated on the finite grid and stored for computation of retarded potentials. We improve the code by employing two new alternative, grid-based techniques for representing a 2D particle distribution. Both of them are extremely efficient in comparison to the traditional Monte Carlo cosine expansion—they are about 3 orders of magnitude faster. Each of the new methods starts with a simple gridded distribution and uses different approaches to denoise it, thus improving the accuracy of the gridded approximation.

The first alternative technique—truncated fast cosine transform (TFCT)—uses fast cosine transforms to transform the gridded distribution to Fourier space, where it truncates higher-order frequencies by retaining only a fraction of the cosine coefficients. This nondiscriminatory removal of the small-scale structure results in smoothing of the distribution, and represents an overly simplistic method for noise removal. In terms of accuracy, the TFCT approximation is equivalent to the Monte Carlo cosine expansion.

The second alternative technique uses discrete fast wavelet transforms to transform the gridded distribution to wavelet space, where it performs wavelet thresholding, thereby largely removing the numerical noise intrinsic in numerical simulations. Transforming back to the physical space yields a truncated wavelet transform (TWT) approximation to the particle distribution which is appreciably more accurate than the cosine expansion.

In Sec. II we outline the original algorithm behind the 2D CSR code developed in [11]. In Sec. III we first analyze the origin and profile of numerical noise associated with charge deposition using various charge deposition schemes (CDS). We then present the two alternative methods in detail. Special attention is devoted to the wavelet-denoised approximation, along with the discussion of how wavelets can be used to remove noise from particle-in-cell (PIC) simulations. In Sec. IV we use typical particle distributions from simulation of the microbunching instability in [11] to compare the spatial resolution, accuracy, smoothness, and efficiency of the two alternative methods to the original Monte Carlo cosine approximation. In Sec. V we apply two versions of the code—one with the original and the other with the new wavelet-denoised density estimation algorithm—to the simulation of the FERMI@Elettra first bunch compressor system, and compare the results. Finally, in Sec. VI we discuss the significance and use of the new approach and outline its future applications.

II. ORIGINAL ALGORITHM WITH MONTE CARLO COSINE DENSITY ESTIMATION

We outline the algorithm for density estimation currently implemented in [11] where the equations of motion are solved in beam frame with phase-space coordinates $\xi = (z, p_z, x, p_x)$ and independent variable s. The quantities $\rho(z, x, s) = \int_{B} dp_z dp_x f(z, x, s)$ and $\tau_g(z, x, s) = \int_{B} dp_z dp_x f(z, x, s) dp_z dp_x$ must be estimated for the calculation of the self-fields. To this end, using the grid transformation $(z, x) \mapsto (x_1, x_2) \in [0, 1] \times [0, 1]$ and denoting $\rho_g$, $\tau_g$ in the grid coordinates $(x_1, x_2)$, we expand $\rho_g (x_1, x_2, s)$ and $\tau_g (x_1, x_2, s)$ in a finite Fourier series

$$\rho_g (x_1, x_2, s) = \sum_{i=0}^{l} \sum_{j=0}^{j} \theta_{ij}(s) \phi_i(x_1) \phi_j(x_2),$$

$$\tau_g (x_1, x_2, s) = \sum_{i=0}^{l} \sum_{j=0}^{j} \Theta_{ij}(s) \phi_i(x_1) \phi_j(x_2),$$

where

$$\theta_{ij}(s) = \int_A dx_1 dx_2 \phi_i(x_1) \phi_j(x_2) \rho_g (x_1, x_2, s),$$

$$\Theta_{ij}(s) = \int_A dx_1 dx_2 \phi_i(x_1) \phi_j(x_2) \tau_g (x_1, x_2, s).$$
Here \( \{ \phi_i \} \) is the orthonormal basis:
\[
\phi_i(x) = \begin{cases} 
\frac{1}{\sqrt{2}} \cos(i \pi x) & i = 0, \\
\frac{1}{\sqrt{2}} \sin(i \pi x) & i \neq 0,
\end{cases}
\]
for \( x \in [0, 1] \).

Since \( \rho_g \) is a probability density, the Fourier coefficients \( \theta_{ij} \) may be written as the expected value \( E \) of \( \phi_i(X_1) \phi_j(X_2) \) with respect to \( \rho_g(\cdot, s) \):
\[
\theta_{ij}(s) = E[\phi_i(X_1) \phi_j(X_2)] = \int_A dx_1 dx_2 \phi_i(x_1) \phi_j(x_2) \rho_g(x_1, x_2, s),
\]
where \( X = (X_1, X_2) \) is the random variable with probability density \( \rho_g \). To estimate \( \tau_g \), which is not a probability density, we notice that the Fourier coefficients \( \Theta_{ij} \) may be written as the expected value \( E \) of \( \phi_i(X_1) \phi_j(X_2) P_X \) with respect to \( f_g(\cdot, s) \):
\[
\Theta_{ij}(s) = E[\phi_i(X_1) \phi_j(X_2) P_X] = \int_A dx_1 dx_2 \int_{\mathbb{R}^2} dp_x dp_s \phi_i(x_1) \phi_j(x_2) p_x \times f_g(x_1, p_z, x_2, p_y, s),
\]
where \( X = (X_1, P_2, X_2, P_X) \) is the random variable with probability density \( f_g(\cdot, s) \).

It follows that the natural estimate of \( E \) is the sample mean, i.e., we have the following two Monte Carlo formulas:
\[
\theta_{ij}(s) = \frac{1}{N} \sum_{n=1}^N \phi_i(X_{1n}) \phi_j(X_{2n}),
\]
\[
\Theta_{ij}(s) = \frac{1}{N} \sum_{n=1}^N \phi_i(X_{1n}) \phi_j(X_{2n}) P_{Xn},
\]
where a realization of the random variable \( X = (X_1, P_2, X_2, P_X) \) is obtained from beam frame scattered phase-space points \( z_i, p_z, x_i, p_x \) at \( s \), \( i = 1, \ldots, N \) [via the transformation: \( (z_i, p_z, x_i, p_x) \rightarrow (x_1, p_z, x_2, p_x) \)]. This is a density estimation used in statistical estimation (e.g., see [12]).

### III. PARTICLE-IN-CELL METHOD: ALTERNATIVE DENSITY ESTIMATION METHODS

The code developed in [11] is not a classic PIC code, since the point-charge particles are not directly deposited on a grid, but instead are used to compute a cosine approximation to the distribution which is then evaluated at predefined grid points. This yields a smooth particle distribution, which, in turn, produces a well-behaved integrand in the evaluation of the retarded fields, at the price of a costly cosine approximation evaluation.

Here we propose to do away with the cosine approximation and directly deposit particles on the grid, in the spirit of traditional PIC codes. Such direct deposition yields a noisy gridded approximation to the particle distribution that is less smooth than the cosine approximation (see Sec. IV). The gridded approximation can then be denoised/smoothed out in several different ways: (i) applying the higher-order CDS; (ii) truncating high frequencies from the Fourier expansion; and (iii) thresholding the wavelet expansion. In the remainder of this section we discuss each of these approaches and argue that only the wavelet thresholding provides a natural setting for judicious noise removal, while others systematically remove physical small-scale structure along with small-scale fluctuations due to numerical noise. This is reflected in the accuracy of the wavelet-denosed representation being clearly superior to the others.

#### A. Error in particle deposition schemes

In PIC simulations, the beam’s charge distribution sampled by particles is given by the Klimontovich particle distribution,
\[
f(z, p_z, x, p_x) = q_0 \sum_{i=1}^N \delta(z - z_i) \delta(p_z - p_{zi}) \times \delta(x - x_i) \delta(p_x - p_{xi}),
\]
where all \( N \) particles carry the same charge \( q_0 \).

To get to the charge density and current distributions at an arbitrary point in \((x, z)\) space, the Klimontovich particle distribution must be convolved with a \( k \)th order CDS \( d_k \) over some neighborhood \( V_k \) around it:
\[
\rho_g(z, x) = \iint_{V_k} \int_{\mathbb{R}^2} f(z - \tilde{z}, p_z, x - \tilde{x}, p_x) dp_x dp_z d_k(\tilde{z}) \times d_k(\tilde{x} d \tilde{z} d \tilde{x}),
\]
\[
\tau_g(z, x) = \iint_{V_k} \iint_{\mathbb{R}^2} p_x f(z - \tilde{z}, p_z, x - \tilde{x}, p_x) dp_x dp_z d_k(\tilde{z}) \times d_k(\tilde{x} d \tilde{z} d \tilde{x}).
\]

The particle density distribution above is evaluated at the specific set of equally spaced points—a rectangular grid. In our 2D case here, the grid is given by the set of grid points \((z_j = ih_{zj}, x_j = jh_{xj})\) with \( j = 1, \ldots, N_z \) and \( i = 1, \ldots, N_x \), with the total number of grid points \( N_{grid} = N_z N_x \). In any realization sampled by \( N \) particles, \( n_j \) is the number of particles inside a \( V_k \) neighborhood of the \( j \)th grid point, where \( V_k = \left[ (-kh_{1j}/2, kh_{1j}/2), (-kh_{2j}/2, kh_{2j}/2), \ldots, (-kh_{Dj}/2, kh_{Dj}/2) \right] \), and \( D \) is the dimensionality of the problem.

CDS \( d_k \) can be classified according to the number of grid points to which the total charge is deposited in 1D (also one more than the order of the local polynomial which describes its shape) [13].

(i) \( k = 1 \) is the nearest grid point (NGP) CDS, in which all of particle’s charge is deposited onto the nearest grid point (see Fig. 1):
\[
d_1(x) = \begin{cases} 
1 & \text{if } |x| \leq \frac{1}{2} \\
0 & \text{otherwise},
\end{cases}
\]

\[070701-3\]
where $x$ is a general coordinate and $h$ is the spacing in that coordinate.

(ii) $k = 2$ is the cloud-in-cell (CIC) CDS, in which each particle’s charge is deposited to all vertices of the cell it occupies (two vertices in 1D, four in 2D, eight in 3D), with the weight inversely proportional to its distance from the vertex (see Fig. 1):

$$d_2(x) = h^{-1}\begin{cases} 1 - \frac{|x|}{h} & \text{if } |x| \leq h \\ 0 & \text{otherwise.} \end{cases}$$

(iii) $k = 3$ is the triangular shaped cloud (TSC) CDS, which deposits the charge onto three nearest cells (three vertices in 1D, nine in 2D, 27 in 3D) using a quadratic (see Fig. 1):

$$d_3(x) = h^{-1}\begin{cases} \frac{3}{4} - \left(\frac{x}{h}\right)^2 & \text{if } |x| \leq \frac{h}{2} \\ \frac{1}{2}\left[\frac{3}{2} - \frac{|x|}{h}\right]^2 & \text{if } \frac{h}{2} \leq |x| \leq \frac{3h}{4} \\ 0 & \text{otherwise.} \end{cases}$$

Other higher-order CDS are obtained by convolving the previous CDS with a CIC CDS [13]. They are, however, rarely used because of their computational cost: the next higher-order CDS than TSC increases the number of points to which the charge is deposited from $3^D$ to $4^D$, which in 3D is from 27 to 64—an increase of more than a factor of 2. Furthermore, the smoothing of the particle distribution obtained by using higher-order CDS comes at the expense of spectral resolution—if the domain of the CDS is too wide (the order of the CDS too high), small-scale structures of the order of the size of the domain or smaller will be smoothed out. This is why in practice most PIC simulations use either CIC or TSC CDS. In the present study, we also, for the same reasons, deal only with $k = 1, 2, 3$ CDS.

In analyzing the origin of the numerical noise associated with CDS, we follow and expand on the discussion of [14].

For the NGP CDS ($k = 1$), the probability of a particle being deposited into the $j$th grid point is given by the binomial distribution

$$P(n_j = n|p_j) = \frac{N!}{n!(N-n)!} p_j^n (1 - p_j)^{N-n},$$

where $p_j = \bar{n}_j/N$ and $\bar{n}_j$ is the expected (exact) value of the particle distribution at the location of the $j$th grid point. In $N$-body simulations, $N$ is quite large, in which case the binomial distribution converges to the Poisson distribution with mean $\bar{n}_j$:

$$P(n_j = n|p_j) = \frac{\bar{n}_j^p e^{-\bar{n}_j}}{n!},$$

where $n$ is an integer.

One of the ways to quantify noise in a PIC simulation is to define a “global” measure of the error associated with noise induced by sampling and particle deposition, defined as the algebraic average of variances:

$$\sigma^2 = \frac{1}{N_{\text{grid}}} \sum_{i=1}^{N_{\text{grid}}} \text{Var}(Q_i) = \frac{1}{N_{\text{grid}}} \sum_{i=1}^{N_{\text{grid}}} \text{Var}(\tilde{Q}_i),$$

where $Q_i$ is the random variable quantifying the charge in the $V$ neighborhood of the $i$th grid point, and $\tilde{Q}_i$ the expected (exact) value of the particle distribution at the location of the $i$th grid point (see Appendix A).

For the CDS considered in this paper ($k = 1, 2, 3 \ldots$), their values are (see Appendix A)

$$\sigma_k^2 = \frac{a_1^2 q_0^2 N}{N_{\text{grid}}} = \frac{a_1^2 N q_0}{N N_{\text{grid}}},$$

where

$$a_1 = 1, \quad a_2 = \left(\frac{2}{3}\right)^{D/2}, \quad a_3 = \left(\frac{11}{20}\right)^{D/2},$$

and $q_0 = N q_0$ is the total charge of the distribution. Therefore, the noise distribution for the higher-order CDS ($k = 2, 3 \ldots$) is a contracted Poissonian, given by Eq. (16) with $n_j \rightarrow a_k n_j$. This also means that the signal obtained by depositing the particle distribution onto a grid using a higher-order CDS is $a_k^{-1}$ more accurate than the NGP ($k = 1$). Table I shows the values of $a_k$ for dimensions $D = 1, 2, 3$.

The global variance as defined above is closely related to the integrated mean square error (IMSE), or $L_2$ error, on the grid:

| Table I. Values of $a_k$ for different order CDS. |
|----------------------|----------------------|----------------------|
| $k$                  | $k = 1$              | $k = 2$              | $k = 3$              |
| $D$                  | $1$                  | $\left(\frac{2}{3}\right)^{D/2}$ | $\left(\frac{11}{20}\right)^{D/2}$ |
| $D = 1$              | $1$                  | $0.82$               | $0.74$               |
| $D = 2$              | $1$                  | $0.67$               | $0.55$               |
| $D = 3$              | $1$                  | $0.54$               | $0.41$               |
\[ e = h_z h_x \sum_{i=1}^{N_{\text{grid}}} (q_i - \bar{q}_i)^2 = \frac{L_z L_x}{N_{\text{grid}}} \sum_{i=1}^{N_{\text{grid}}} (q_i - \bar{q}_i)^2 \]

where \( h_z \) and \( h_x \) are the resolution and \( L_z \) and \( L_x \) the size of the grid in z and x directions, respectively, \( q_i \) is the single Monte Carlo realization of the random variable \( Q_i \), the charge deposited onto the \( i \)th grid point, and \( \bar{q}_i = \bar{Q}_i \) is the expected (exact) value of the distribution. The sum of differences in a single Monte Carlo realization \( \sum (q_i - \bar{q}_i)^2 \) is an asymptotically accurate estimate of the sum of variances over all Monte Carlo realizations \( \sum \text{Var}(Q) \). The IMSE is often used to quantify the accuracy of the approximations of the analytically known particle distribution.

Equations (18) and (20) imply that the quantities \( \sigma^2_i / a_i^2 \) and \( \varepsilon / a_i^2 \) are the same for all CDS, and that, for a fixed resolution (\( N_{\text{grid}} = \text{const} \)), they both scale as \( \propto N^{-1} \). This is clearly demonstrated in Fig. 2. Therefore, depositing point-charge particles over several grid points—i.e., using higher-order CDS—essentially cuts off the highest order frequencies of the resulting distribution. Such smoothened distributions are closer—as measured by the IMSE—to the smooth distribution function being sampled, at the expense of reduced spatial resolution.

The algebraic average of variances of noise in a PIC simulation \( \sigma^2_i \) depends sensitively on the parameters of the simulation (number of particles and the grid resolution) and on the CDS (through \( a_i \)), but only very weakly on the particle distribution (see Appendix A of [14]). After a systematic study of the many types of distributions of particles arising in beam dynamics, this weak dependence appears to be negligible (see [14] for a more detailed discussion). This global measure of numerical noise in PIC simulations is needed to compute the noise threshold which is used in wavelet thresholding (see Sec. III C).

**B. Alternative method 1: Truncated fast cosine transform**

A finite cosine expansion of a charge and current density distribution is given by

\[ \rho(z, x) = \sum_{n=0}^{M} \sum_{m=0}^{M} \theta_{nm} \phi_n(z) \phi_m(x), \]

\[ \tau(z, x) = \sum_{n=0}^{M} \sum_{m=0}^{M} \Theta_{nm} \phi_n(z) \phi_m(x), \]

where

\[ \theta_{nm} = \int_0^1 \int_0^1 \rho(z, x) \phi_n(z) \phi_m(x) dz dx, \]

\[ \Theta_{nm} = \int_0^1 \int_0^1 \tau(z, x) \phi_n(z) \phi_m(x) dz dx, \]

and the basis functions \( \phi \) given in Eq. (5). Both the Monte Carlo and the TFCT evaluate the coefficients from Eq. (22) using the charge density distribution function \( \rho_{\delta}(z, x) \) from Eq. (11) with different CDS.

(i) The Monte Carlo cosine expansion uses a “zeroth” order CDS,\(^1\) the Dirac delta function, which “deposits” the entire charge of the particle onto a point at which it is located:

\[ d_{0}(x) = \delta(x). \] (23)

Upon substituting Eq. (23) in Eq. (11), the resulting charge and current density distributions become

\[ \rho(z, x) = q_0 \sum_{i=1}^{N} \delta(z - z_i) \delta(x - x_i), \]

\[ \tau(z, x) = q_0 \sum_{i=1}^{N} p_{zi} \delta(z - z_i) \delta(x - x_i). \] (24)

When inserted in the expression for the coefficients (22), this charge density distribution yields

\[ \theta_{nm} = q_0 \sum_{i=1}^{N} \phi_n(z_i) \phi_m(x_i), \]

\[ \Theta_{nm} = q_0 \sum_{i=1}^{N} p_{zi} \phi_n(z_i) \phi_m(x_i). \] (25)

(ii) The TFCT uses higher (nonzero) order CDS \( d_k \) (Sec. III A) evaluated on a grid of resolution \( (N_x, N_y) \). The coefficients are then computed from this gridded distribution via fast cosine transform, which is just an optimized way to solve Eq. (22). Finally, the full cosine

---

\(^1\)This is a CDS only in a loose sense. It is meaningless to talk about properties of this CDS as we did for the higher-order ones \( (\sigma_0, a_0) \), because they are not defined.
expansion containing \((N_z, N_x)\) coefficients is truncated to retain only the lowest \((M_z, M_x)\).

When the resolution of the grid is fine, the \(V_k\) neighborhood is small and the CDS are narrow. The narrowest CDS of nonzero order is NGP, which, as the resolution grows, becomes increasingly a better approximation to the delta function, and hence closer to the zeroth order CDS. Therefore, we expect the cosine coefficients computed from a grid resulting from a NGP CDS to be very close to those of the Monte Carlo cosine, given in Eq. (25). In Appendix B, we prove that the two are the same to within a second-order approximation. Figure 3 shows this: the IMSE as a function of the number of particles for the two expansions overlap.

The full (nontruncated) fast cosine expansion generated solving Eq. (25) with \(M_z = N_z\) and \(M_x = N_x\) is as accurate as the approximation generated using NGP CDS, because neither of the two lowest CDS (zeroth order delta function or the first order NGP) smoothen the contributions of individual point charges over space: in both cases the charges are deposited onto a single point, unlike in the case of the higher-order CDS. As discussed earlier, smoothing by a higher-order CDS decreases the IMSE (Fig. 3) by removing the highest order frequencies (thereby reducing its spectral resolution). However, more powerful smoothing is accomplished by truncation of the cosine representation (see Sec. IV B).

We introduce the alternative cosine expansion—TFCT—which is equivalent to the one currently used in [11], yet considerably faster. The algorithm can be outlined as follows.

(i) Deposit particles on the \((N_z, N_x)\) grid.

(ii) Apply 2D fast cosine transform on the grid, thus yielding \((N_z, N_x)\) cosine coefficients.

(iii) Remove the high-frequency contribution to the density by truncating coefficients higher than \(M_z\) and \(M_x\) in \(z\) and \(x\) coordinates, respectively. This removal of the high-frequency components results in a smoother distribution, but it also removes small-scale structures that may not be due to noise. Therefore, the truncation of the Fourier coefficients also restricts the spatial resolution of the representation (see Sec. IV B).

(iv) Apply 2D inverse fast cosine transform on the grid, to obtain the smoothed distribution in physical space.

C. Alternative method 2: Thresholded wavelet transform

The details of the wavelet thresholding in the context of the PIC simulations of beams have been outlined in [14]. Here we highlight the main points of their discussion.

Since their inception in the 1980’s [15–19], wavelets and wavelet transforms have been used to denoise signals contaminated with noise. Just like the discrete Fourier transform (DFT), the discrete wavelet transform (DWT) is a fast, linear operation on data sets with size of integer power of 2 in each dimension, yielding a transformed data set of the same size. The most important difference between DWT and DFT is that the DWT provides localization in both frequency space and configuration space, whereas the DFT is only localized in frequency space. The reason for this is that, while still providing a hierarchy of frequencies, the basis functions of the DWT have compact support (finite range over which they are nonzero); in contrast, the basis functions of the DFT have infinite range. This dual localization renders a number of operators and functions sparse in the wavelet space. Essential background on wavelets and wavelet transforms is available in literature [15,18,19].

Upon transformation of the noisy data to wavelet space, the signal is generally given by a few large coefficients, while most of the noise is mapped onto many small wavelet coefficients. Wavelet thresholding is a process which alters the contribution of the wavelet coefficients deemed to represent noise from the signal. After judiciously adopting a noise threshold \(T\), one of the two common thresholding techniques can be applied.

(i) **Hard thresholding** sets the coefficients with magnitudes below the noise threshold \(T > 0\) to zero:

\[
\tilde{w}_i = \begin{cases} 
  w_i & \text{if } |w_i| > T \\
  0 & \text{if } |w_i| \leq T.
\end{cases}
\]

(ii) **Soft thresholding** sets the coefficients with magnitudes below the noise threshold \(T > 0\) to zero, and contracts coefficients above it by \(T\):

\[
\tilde{w}_i = \begin{cases} 
  \text{sign}(w_i)|w_i| - T & \text{if } |w_i| > T \\
  0 & \text{if } |w_i| \leq T.
\end{cases}
\]

Thresholding the wavelet coefficients therefore removes the smallest-scale fluctuations usually associated with numerical noise. However, the noise threshold must be chosen judiciously so as to avoid uncontrolled “smoothing out” over the physical fine-scale details that serve as a seed for the onset of global features.

FIG. 3. Accuracy (measured by the IMSE) for the gridded distribution deposited using NGP (green line), CIC (blue), TSC (black) CDS, and the full (nontruncated) Monte Carlo cosine expansion (red) as a function of the number of particles \(N\) for the sinusoidally modulated flattop distribution used in Sec. IV.
The discussion on choosing the noise threshold is extensively discussed in literature [14,20–26]. The most commonly used noise threshold is given in terms of the standard deviation $\sigma$ of the noise as

$$T = \sqrt{2 \log N_{\text{grid}}} \sigma. \quad (28)$$

This is a universal threshold for signals with Gaussian white noise, i.e., better noise removal cannot be achieved for all signals in all wavelet bases. It results in a denoised signal which is within a small factor of ideal denoising [22,23].

Most, if not all, of the work on wavelet denoising deals with distributions (signals) polluted with additive (distribution-independent) Gaussian noise [22,23,25]. In Sec. III A, however, we demonstrate that the noise in PIC simulations is Poisson distributed (multiplicative) and weakly distribution dependent. We assure that the wavelet thresholding theory outlined in earlier work directly applies to PIC simulations by transforming a signal with Poisson-distributed noise into one for which the noise is approximately Gaussian distributed via a variance-stabilizing transformation due to Anscombe [27]:

$$X_G = 2 \sqrt{X_P + \frac{3}{8}} \quad (29)$$

The transformed signal $X_G$ has unit variance ($\sigma = 1$) and mean

$$m_G = \sqrt{m_P + \frac{3}{8} - \frac{1}{8m_P^{1/2} + \frac{1}{64m_P^{3/2}}}}. \quad (30)$$

where $m_P$ is the mean of the Poissonian signal. Applying the Anscombe transformation leads to a bias in the data [26,27], which is removed by ensuring that the denoised and noisy data have the same mean (in PIC simulations, this is equivalent to enforcing conservation of total charge).

Application of the Anscombe transformation to a signal contaminated with Poisson noise as in Eq. (16) yields a signal which is approximately normally distributed, with variance

$$\sigma_{\text{NGP}} = 1 \text{ for the NGP CDS. For the higher-order CDS, the data distribution is a contracted Poissonian, given by Eq. (16) with } n_j \to a_k n_j, \text{ which implies that the resulting variance will be appropriately contracted by a factor } a_k, (\sigma_{\text{CIC}} = a_2, \sigma_{\text{TSC}} = a_3).$$

Combining these with Eq. (28) yields optimal noise threshold for the Anscombe-transformed (variance-stabilized) data obtained by depositing a distribution using a $k$th order CDS:

$$T_k = \sqrt{2 \log N_{\text{grid}} a_k}. \quad (31)$$

The analytical distribution used is the sinusoidally modulated flattop distribution described in Sec. IV. Figure 4 shows the IMSE as a function of the threshold $T$. 

FIG. 4. Accuracy of the approximation (measured by the IMSE) as a function of the wavelet threshold $T$ for the sinusoidally modulated flattop distribution used in Sec. IV. The resolution of the grid is fixed at $1024 \times 128$ in $(\zeta, x)$ space. Number of particles sampling the distribution is given in the title of each panel. The four curves correspond to the four different wavelets from Daubechies family: order 2 (red), order 3 (green), order 6 (blue), and order 10 (purple). The light blue horizontal line represents the accuracy of the grid deposition and black the cosine expansion. The vertical line is the analytical value of $T_2$ given in Eq. (31), because CIC CDS ($k = 2$) is used.
The limits are $T \to 0$—no thresholding is done, the signal is unaltered, yielding the original gridded distribution; and $T \to \infty$—all wavelet coefficients are set to zero, and there is no remaining signal. We notice that the threshold $T_2$ is consistently at or near the minimum IMSE. This has also been observed in our earlier study with different kinds of distributions [14].

1. Choosing a wavelet family

Wisely choosing a wavelet family should yield a compact representation of the signal. Donoho and Johnstone [22,23] showed that for a given application there exists the "ideal" wavelet basis in which the optimal compression is achieved. This means that the closer the basis is to the ideal basis, the better the compression that results, and vice versa.

Figure 5 shows the fraction of wavelet coefficients of the charge density retained after thresholding versus the number of particles sampling the particle distribution (at the fixed resolution). In the present study, we use two types of wavelets: orthogonal—represented by Daubechies family of wavelets of order 2, 3, 6, and 10, symlets of order 4, 6, 8, and coiflets of order 1, 2, 3; and biorthogonal—represented by bior (4,4) and bior (6,8) [17].

As the order of the wavelet family increases, so does the width of its compact support. Better localization in space is obtained through wavelets with small compact support (lower order), while better localization in frequency is achieved by higher-order ones.

In this study, the more important of the two localizations is the localization in frequency, which is why we use Daubechies wavelets of order 10. From Fig. 4, it seems to be the best-performing wavelet in its family.

2. Combining wavelet denoising with different particle deposition schemes

The TWT method is grid based, so it is natural to ask which of the CDS investigated in Sec. III A, when combined with wavelet thresholding, will yield the most accurate approximation.

Figure 6 shows the IMSE for the three CDS—NGP ($k = 1$), CIC ($k = 2$), and TSC ($k = 3$)—before (thin lines) and after wavelet thresholding (thick lines). Wavelet-denoised CIC and TSC representations are clearly more accurate than the wavelet-denoised NPG. For small and intermediate $N$, the CIC and TSC CDS are quite similar in accuracy. For large $N$, however, the lower-order CDS CIC yields a better approximation than the TSC, because of the added smoothing used in the higher CDS. This sends a clear message that, if wavelet denoising is employed, there is no need to use CDS higher than CIC ($k = 2$).

It is because of these considerations that we only use the CIC CDS in the remainder of the present study.

3. Outline of the thresholded wavelet method

The wavelet-denoised algorithm for estimating the particle density can be outlined as follows: (i) deposit particles on the $(N_z, N_x)$ grid; (ii) apply Anscombe transformation to convert the resulting gridded signal polluted by the Poissonian noise (the gridded particle distribution) to the signal with Gaussian noise; (iii) apply 2D DWT on the grid, thus yielding $(N_z, N_x)$ wavelet coefficients; (iv) perform wavelet thresholding on the wavelet coefficients in order to remove numerical noise and smoothen the particle distribution. Unlike the filtering of the cosine coefficients in Fourier space, this judicious noise removal

![Image 1](image1.png)

**FIG. 5.** Sparsity of the wavelet-denoised representation of the flattop distribution used in Sec. IV, quantified by the fraction of nonzero wavelet coefficients after thresholding. Wavelet families used are: Daubechies of order 2, 3, 6, 10; symlets of order 4, 6, 8; coiflets of order 1, 2, 3; and biorthonormal (4,4) and (6,8).

![Image 2](image2.png)

**FIG. 6.** Accuracy of the approximation (measured by the IMSE) as a function of the number of particles $N$ for the three lowest-order CDS: NGP (thin red line), CIC (thin green), and TSC (thin blue) for the sinusoidally modulated flattop distribution used in Sec. IV. The thick lines correspond to the wavelet-denoised distribution for each.
does not restrict the spatial resolution of the distribution: small-scale structures which are not deemed to be noise during wavelet thresholding (i.e., the amplitude of the corresponding wavelet coefficients is larger than the threshold) are retained in the denoised distribution; (v) apply 2D inverse DWT on the grid, to obtain the smoothed distribution in physical space; and (vi) apply inverse Anscombe transformation.

IV. COMPARISON OF THE DENSITY ESTIMATION METHODS

We now compare the TFCT and the TWT estimation methods in the following areas: spatial resolution, accuracy, smoothness, and efficiency. The simulations in this section are based on an analytical model charge distribution, carefully chosen to represent small-scale structure in the beam. Knowing the underlying analytical distribution allows computation of the IMSE for the two methods. It is our expectation that the comparison presented here, which is based on the analytically known distribution, extends to general particle distributions arising in realistic simulations.

A. Analytical model charge distribution:
Sinusoidally modulated flattop

In order to study the microbunching instability, Bassi et al. [11] designed a simulation in which a flattop distribution, modulated with a single sinusoidal frequency, is passed through the FERMI@Elettra first bunch compressor system. The resulting amplification of the initial modulation, denoted as the gain factor (see Sec. V), is a good indicator as to how small-scale structures may respond to bunch compression. It is therefore important to show how the gain factor varies as a function of the frequency (wavelength) of the modulations.

In what follows, we investigate only the approximations to the particle charge distribution. However, the analysis also applies to approximations to the current density, because the two have the same form [see Eq. (21)].

Figure 7 shows the flattop charge distribution in physical space:

\[
\rho(x, z) = \left[ 1 + A \cos \left( \frac{2\pi z}{\lambda} \right) \right] \mu(z) g(x),
\]

where

\[
\mu(z) = \frac{1}{4a} \left[ \tanh \left( \frac{z + a}{b} \right) - \tanh \left( \frac{z - a}{b} \right) \right],
\]

\[
g(x) = \frac{1}{\sqrt{2\pi\sigma_{z0}}} e^{-\left( x^2 / 2\sigma_{z0}^2 \right)}.
\]

Parameters \( a \) and \( b \) determine the width and the steepness of the flattop in \( z \) coordinate; \( \sigma_{z0} \) the width in \( x \) coordinate, and \( A \) and \( \lambda \) are the magnitude and the wavelength of the modulation, respectively. The parameters and variables are nondimensionalized in the manner given in Table II. The charge distribution above is a particular case of the full phase-space distribution function used in the code, given by Eq. (48).

B. Spatial resolution

In general, the smallest structure representable by a finite cosine approximation (including the TFCT method) is determined by the highest order basis function retained in the expansion. The highest order basis function in \( z \) coordinate is \( M_z \) and \( M_x \) in \( x \). The arguments of the basis functions [given in Eq. (5)] are

\[
M_z \pi \bar{z} = \frac{M_z \pi z}{L_z} + \phi_z, \quad \text{in } z \text{ coordinate},
\]

\[
M_x \pi \bar{x} = \frac{M_x \pi x}{L_x} + \phi_x, \quad \text{in } x \text{ coordinate},
\]

where \( L_z \) and \( L_x \) are the size of the computational box around the beam in \( z \) and \( x \) coordinates, and \( \phi_z \) and \( \phi_x \) are dimensionless constants (and therefore just simple phase shifts). Therefore, the smallest wavelengths representable by this expansion are

\[
\lambda_{z, \text{min,cos}} = \frac{2L_z}{M_z}, \quad \text{in } z \text{ coordinate},
\]

\[
\lambda_{x, \text{min,cos}} = \frac{2L_x}{M_x}, \quad \text{in } x \text{ coordinate}.
\]

In the microbunching simulations, we use a sinusoidally modulated flattop distribution to study the evolution of the small-scale structures. The wavelength of these small-scale longitudinal structures is given by \( \lambda \) in Eq. (32). The cosine approximation will be able to approximate accurately only the particle distributions with small-scale structure with wavelengths \( \lambda \) larger than the smallest wavelength.
representable by the cosine expansion $\lambda_{\text{min,cos}}^z$. This imposes a limit on usability of the cosine expansion:

$$\lambda > \lambda_{\text{min,cos}}^z = \frac{2L_z}{M_z} \quad (36)$$

and

$$M_z > \frac{2L_z}{\lambda} = \frac{7200 \, \mu\text{m}}{\lambda}. \quad (37)$$

Figure 8 shows the accuracy in approximating the gain factor of a distribution sampled with $N = 10^9$ particles by a cosine approximation with $M_z = 100$ and $M_\perp = 40$ and the finite grid (both with and without wavelet denoising) of resolution $N_\perp = 1024$ and $N_z = 128$. The relation (37) predicts that the precipitous drop in accuracy (increase in error) for the cosine approximation with $M_z = 100$ should occur for

$$\lambda < \frac{7200 \, \mu\text{m}}{M_z} = 72 \, \mu\text{m}, \quad (38)$$

which is exactly what is observed. This general constraint is manifested in Fig. 9, where the IMSE is plotted as a function of the longitudinal modulation wavelength $\lambda$ and number of cosine basis functions in the longitudinal direction ($M_z$). The steep dropoff in accuracy of cosine approximation, as outlined by the contour plot, coincides with the line $\lambda M_z = 7200 \, \mu\text{m}$, given by Eq. (38). It is also interesting to note that, for the range where cosine expansion provides a good approximation to the particle distribution function ($\lambda M_z > 7200 \, \mu\text{m}$), increasing $M_z$ does not yield a more accurate approximation, as evidenced by an upward sloping surface of the error $\varepsilon$ as a function of $M_z$. All of this essentially means that the optimal simulation with cosine expansion—one that would provide the best accuracy at the lowest cost—should have the number of longitudinal cosine basis function just slightly larger than 7200 $\mu\text{m}/\lambda$.

### TABLE II. Dimensional and nondimensionalized parameters of the simulation.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Physical</th>
<th>Unit</th>
<th>Nondimensional</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal beam size $\sigma_{\text{z,0}}$</td>
<td>$9 \times 10^{-4}$</td>
<td>m</td>
<td>$\sigma_{\text{z,0}}$</td>
<td>1</td>
</tr>
<tr>
<td>Transverse beam size $\sigma_{\text{x,0}}$</td>
<td>$2.3 \times 10^{-4}$</td>
<td>m</td>
<td>$\sigma_{\text{x,0}}$</td>
<td>1</td>
</tr>
<tr>
<td>Modulation amplitude $A$</td>
<td></td>
<td></td>
<td>$A$</td>
<td>0.05</td>
</tr>
<tr>
<td>Modulation wavelength $\lambda$</td>
<td></td>
<td></td>
<td>$\lambda = \lambda/\sigma_{\text{z,0}}$</td>
<td>$\lambda/900$</td>
</tr>
<tr>
<td>Beam parameter $a$</td>
<td>$1.18 \times 10^{-3}$</td>
<td>m</td>
<td>$\tilde{a} = a/\sigma_{\text{z,0}}$</td>
<td>1.311111</td>
</tr>
<tr>
<td>Beam parameter $b$</td>
<td>$1.5 \times 10^{-4}$</td>
<td>m</td>
<td>$\tilde{b} = a/\sigma_{\text{z,0}}$</td>
<td>0.166666</td>
</tr>
<tr>
<td>Longitudinal coordinate $z$</td>
<td>$[-2\sigma_{\text{z,0}}, 2\sigma_{\text{z,0}}]$</td>
<td>m</td>
<td>$\tilde{z} = z/\sigma_{\text{z,0}}$</td>
<td>$[-2, 2]$</td>
</tr>
<tr>
<td>Transverse coordinate $x$</td>
<td>$[-4\sigma_{\text{x,0}}, 4\sigma_{\text{x,0}}]$</td>
<td>m</td>
<td>$\tilde{x} = x/\sigma_{\text{x,0}}$</td>
<td>$[-4, 4]$</td>
</tr>
<tr>
<td>Longitudinal box size $L_z$</td>
<td>$4\sigma_{\text{z,0}}$</td>
<td>m</td>
<td>$\tilde{L}<em>z = L_z/\sigma</em>{\text{z,0}}$</td>
<td>4</td>
</tr>
<tr>
<td>Transverse box size $L_x$</td>
<td>$8\sigma_{\text{x,0}}$</td>
<td>m</td>
<td>$\tilde{L}<em>x = L_x/\sigma</em>{\text{x,0}}$</td>
<td>8</td>
</tr>
</tbody>
</table>

FIG. 8. Accuracy of FFT-based cosine expansion (red line), simple grid deposition (green line), grid-based truncated FCT (blue), and a wavelet-denoised grid deposition (purple) for the flattop distribution modulated with a sinusoidal perturbation with $\lambda = 100 \, \mu\text{m}$ and $A = 0.05$ on a grid with resolution $N_\perp = 1024$, $N_z = 128$, and $N = 10^9$ particles. Vertical lines denote minimal wavelengths representable with grid and cosine approximations (from left to right), as predicted by Eqs. (39) and (35). The cosine expansion has $M_z = 100$ and $M_\perp = 40$ coefficients. The wavelet family used is Daubechies of order 10.

FIG. 9. Accuracy (measured by the IMSE) for the cosine approximation of the modulated flattop distribution as a function of the longitudinal modulation wavelength $\lambda$ and number of cosine basis functions in the longitudinal direction ($M_z$), for $N = 10^9$, $N_\perp = 1024$, $N_z = 128$. The number of transverse basis functions is fixed at $M_\perp = 40$. The contour curve denoting the dropoff in accuracy of about 2 orders of magnitude corresponds to $\lambda M_z = 7200 \, \mu\text{m}$, as predicted by Eq. (35).
The smallest structure representable on the grid has the wavelength of only four grid spacings, describing the "sawtooth" structure:

$$\lambda_{z, \text{min,grid}} = 4h_z = \frac{4L_z}{N_z}. \quad (39)$$

This means that the requirement for the accurate representation of the distribution on the finite grid is then given by

$$\lambda > \lambda_{z, \text{min,grid}} = \frac{4L_z}{N_z}, \quad N_z > \frac{4L_z}{\lambda} = \frac{3600 \, \mu m}{\lambda}. \quad (40)$$

For the simulations in Fig. 8, where $N_z = 1024$, the relation (40) predicts that for

$$\lambda < \frac{3600 \, \mu m}{N_z} = 14 \, \mu m, \quad (41)$$

the grid approximation will be inaccurate, as evidenced by the increase in error of the grid approximation.

Therefore, the spatial resolution for the cosine expansion is determined by the number of expansion coefficients $M_z$ and $M_x$, while for the grid approximation it is given by the number of grid points $N_z$ and $N_x$. In the original implementation of the code in [11], the longitudinal resolution is fixed so that the number of grid points over the period of the highest retained basis function is fixed to 20. After using Eq. (39), we obtain

$$\lambda_{z, \text{min,cos}} = 20h_z = 5\lambda_{z, \text{min,grid}}, \quad (42)$$

and, with the help of Eq. (35),

$$N_z \approx 10M_z. \quad (43)$$

In other words, the smallest structure representable by the truncated cosine expansion (Monte Carlo or TFCT on the grid) is 5 times larger than the smallest structure representable on the grid.

Thresholding of wavelet coefficients does not automatically remove highest-frequency components in the manner of the TFCT method. Instead, only the coefficients whose magnitude is below a certain predetermined threshold are set to zero. The process of wavelet thresholding, including the Anscombe transform, removes only the coefficients judged to be noise, regardless of their spatial resolution. As a consequence of this, small-scale structures which are not classified as noise remain a part of the denoised signal. This means that the wavelet thresholding on a grid does not further limit the spatial resolution. Therefore, the spatial resolution of the TWT is the same as that of the grid (see Fig. 8).

C. Accuracy

Figure 10 shows the IMSE for the grid, Monte Carlo cosine, TFCT and TWT approximations as a function of the number of particles $N$ used in the simulation. It is evident that the IMSE for all approximations scales as $N^{-1}$, consistent with the well-known result that the signal quality (as measured by the square root of the normalized inverse of the IMSE, also known as the signal-to-noise ratio) scales as the square root of the number of particles. As we have shown in Appendix B, the Monte Carlo cosine expansion and TFCT are equivalent to within second order in the grid spacing. It is also clear that the wavelet denoising significantly improves the accuracy of the approximation, surpassing that of the cosine expansion. The increase in accuracy when using the denoised grid approximation is even more pronounced for other values of the modulation wavelength $\lambda$ (see Fig. 8).

As we mentioned earlier, we require that the simulations are set up in such a way that the highest-frequency basis function retained is sampled by at least 15–20 grid points, in order to ensure smooth finite differencing required for the numerical computation of derivatives. Also, in order to maintain the same signal quality, as represented by the distribution of point charges sampling the underlying distribution deposited on a finite grid, we maintain the same number of particles per grid for the entire simulation. The parameters for the simulation are given in Table III. Figure 11 shows the accuracy of the grid, TFCT and TWT representations (top), as well as the number of coefficients retained in the ideal representation, normalized to the number of grid points (bottom). It is evident that both truncated representations are significantly more accurate

![Figure 10](image)

**FIG. 10.** Accuracy (measured by the IMSE) as a function of the number of particles $N$ for the approximations of the flattop distribution with $\lambda = 100 \, \mu m$ modulations: Monte Carlo cosine (red), grid (green), FCT on a grid (blue), and grid with wavelet denoising, using Daubechies wavelets of order 10 (purple). The resolution is $N_z = 1024$, $N_x = 128$, and the number of cosine basis functions is 100 and 40 in $z$ and $x$ directions, respectively.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$N_z$</th>
<th>$N_x$</th>
<th>$M_z$</th>
<th>$M_x$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25–50</td>
<td>4096</td>
<td>128</td>
<td>400</td>
<td>40</td>
<td>$2 \times 10^6$</td>
</tr>
<tr>
<td>60–90</td>
<td>2048</td>
<td>128</td>
<td>200</td>
<td>40</td>
<td>$10^8$</td>
</tr>
<tr>
<td>100–200</td>
<td>1024</td>
<td>128</td>
<td>100</td>
<td>40</td>
<td>$5 \times 10^7$</td>
</tr>
</tbody>
</table>

**TABLE III.** Values for the parameters used in simulations in Figs. 11, 12, and 23. The number of particles per cell is fixed at 382.
than the gridded representation alone, and that the TWT is appreciably more accurate than the TFCT. Furthermore, TWT is about an order of magnitude more sparse, which reduces memory requirements and leads to a more efficient algorithm.

D. Smoothness

In order to estimate the smoothness of the particle distribution, we compute the IMSE of its derivatives obtained using a finite difference scheme on the grid. Figure 12 compares the smoothness of the three gridded approximations—(nonsmoothened) grid, TFCT, and TWT—measured by the IMSE of the $z$ derivatives computed using a third-order-accurate four-point asymmetric stencil (this is how the fields are computed from the gridded particle distribution in [11]). Both smoothening schemes—TFCT and TWT—improve the accuracy of the first derivative by more than 3 orders of magnitude.

A detailed analysis of the accuracy of derivatives computed using a finite difference scheme is given in Appendix C.

E. Efficiency

Figure 13 shows how execution times of different methods and their constituent parts scale with the number of particles $N$. The Monte Carlo-based computation of cosine coefficients requires integration over all of the particles in

\[ n_x = 128, \quad n_z = 1004, \quad M = 100, \quad M_c = 40. \]

Wavelet family used is Daubechies of order 10. $\lambda = 100\ \mu\text{m}$ and $A = 0.05$. The simulation was executed on a dual Intel(R) Xeon(TM) 3.06 GHz CPU Pentium with 2 Gb memory, running Linux 2.6.20.
the distribution, therefore scaling as \( \propto O(M_z M_x N) \). Using cosine coefficients to approximate the particle distribution on the grid requires summing over all the cosine coefficients, thus scaling as \( \propto O(M_z M_x N) \). In realistic simulations \( N \gg N_{\text{grid}} \), so the computation of cosine coefficients will be by far more time consuming of the two.

Both grid-based methods use a charge deposition, which scales as \( \propto O(N) \), and fast transforms: FCT \( \propto O(N_{\text{grid}} \log(N_{\text{grid}})) \) and DWT \( \propto O(KN_{\text{grid}}) \), where \( K \) is the width of the wavelet family. This means that, for large numbers of particles \( N \) used in realistic simulations, the most time-consuming part of the approximation is the charge deposition (Fig. 13). This is also why the execution times of grid methods become quite similar for large \( N \) (Fig. 14).

Figure 14 shows the execution time for the two alternative methods normalized with the execution time of the Monte Carlo cosine approximation used in the original code as a function of the number of particles \( N \). It is evident that the ratio of execution times asymptotically approaches \( 1/(M_z M_x) \), because Monte Carlo cosine scales as \( \propto O(M_z M_x N) \) and the grid-based methods (for large \( N \) used in realistic simulations) as \( \propto O(N) \).

An honest comparison of the efficiency of Monte Carlo cosine expansion and the expansions with TFCT and TWT has to take into account the optimum number of basis functions in the cosine expansion. The execution times for the computation of cosine coefficients and grid-based methods are given by

\[
t_{\text{MC cos}} = k_1 M_z M_x N, \quad t_{\text{grid}} = k_2 N, \quad (44)
\]

where \( k_1 \) and \( k_2 \) are some constants which depend on the computer executing the simulation. Here we assume that for large \( N \), as required by realistic simulations, the dominant cost is due to charge deposition, and not fast transforms (see Fig. 13). Therefore, the ratio of the execution times is simply

\[
r = \frac{t_{\text{MC cos}}}{t_{\text{grid}}} = \frac{k_1 M_z M_x}{k_2 N}.
\]

which means that the cost effectiveness of both TFCT and TWT over the Monte Carlo cosine expansion will be even more pronounced for simulations of small-scale perturbations, as expected. For example, this means that for \( M_z = 40 \), the increase in computational efficiency for \( \lambda < 100 \) \( \mu m \) should be by at least a factor of about 2880, and double that at \( \lambda < 50 \) \( \mu m \).

V. SIMULATION COMPARISON: TFCT VS TWT

As an application, we study the microbunching instability in the first bunch compressor system of the FERMI@Elettra FEL. We discuss the numerical results and compare the performance of the TWT and TFCT methods.

In order to reach the desired electron peak current capable of inducing the collective FEL instability in FELs, the pulse length of a low-emittance electron bunch generated from the photocathode rf gun is magnetically compressed in the linear accelerator by more than an order of magnitude. Numerical and theoretical investigations of high-brightness electron bunch compression lead to a microbunching instability driven by CSR that can significantly degrade the beam quality [6].

The mechanism for microbunching instability can be understood as follows. A high-brightness electron beam with a small amount of longitudinal density modulation can create self-fields that lead to beam energy modulation. Since a magnetic bunch compressor (usually a chicane) introduces path length dependence on energy, the induced energy modulation is then converted to additional density
modulation that can be much larger than the initial density modulation. This amplification process (the gain in microbunching) is accompanied by a growth of energy modulation if significant energy modulation is induced in a dispersive region such as the chicane. Thus, the instability can be harmful to FEL performance, which depends critically on the high quality of the electron beam [6]. We study the microbunching instability with parameters for the first bunch compressor system of FERMI@Elettra. The layout of the system is shown in Fig. 15 and the chicane and beam parameters at first dipole are shown in Table IV. The phase-space density on entrance to the chicane is a smooth function $a_0(z, p_z, x, p_x)$ modulated by a factor $1 + A \cos(2\pi z/\lambda)$, where $A$ is a small amplitude and $\lambda$ is the perturbation wavelength.

The function $a_0$ contains the $z - p_z$ correlation that is necessary for bunch compression, called energy chirp.

A standard approach to study the microbunching instability consists of calculating a gain factor for a given initial modulation wave number $k_0$. This gain factor is defined as

$$G(k_0) = |\frac{\rho[k_0C(s_f), s_f]}{\rho(k_0, 0)}|,$$

for a given initial modulation wavelength of $\lambda = 2\pi k_0$. Here $\rho_L(z, s) = \int dx \rho(z, x, s)$ is the longitudinal spatial density, $C(s) = 1/[1 + hR_{56}(s)]$ is the compression factor of the chicane, $s$ is the arclength along the reference orbit, $s_f$ is the value of $s$ at the exit of the chicane, and $h$ is the linear chirp parameter. This definition is motivated by the fact that without self-fields an initial modulation at wave number $k_0$ will appear at wave number $k_0C(s_f)$ at the end of the chicane due to the bunch compression. We calculate the gain factor numerically and compare it with the analytical formula derived in [4]. The derivation of the analytical gain factor is based on a Vlasov approach with a 1D steady state CSR model and with a coasting beam form of $a_0$ in Eq. (49). The gain factor is obtained by linearizing the Vlasov equation and solving it with an iterative procedure. The numerical gain factor is calculated with the 2D model of CSR [11] and with a 1D field approximation to the 2D model [28]. Because of the several assumptions made in the derivation of the analytical gain factor, we do not expect a detailed agreement between analytical and numerical results. In what follows, we sketch out the 2D model and 1D field and discuss the numerical results.

### A. Self-consistent 2D model

We briefly outline the two-dimensional mean-field treatment of CSR effects discussed in [11]. We use Frenet-Serret coordinates with respect to the reference curve and in addition to the lab system we have two coordinate systems, which we call “beam system 1” and “beam system.” The former uses $u = ct$ as independent variable where path length $s$ is a dependent variable, and the latter

![Figure 15](image-url)  
**FIG. 15.** First bunch compressor of FERMI@Elettra. The parameters are listed in Table IV. The curve in blue is the reference curve in the laboratory system. The regions in gray represent the magnets.

### TABLE IV. Chicane parameters and beam parameters at first dipole.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy reference particle</td>
<td>$E_r$</td>
<td>233</td>
<td>MeV</td>
</tr>
<tr>
<td>Peak current</td>
<td>$I$</td>
<td>120</td>
<td>A</td>
</tr>
<tr>
<td>Bunch charge</td>
<td>$Q$</td>
<td>1</td>
<td>nC</td>
</tr>
<tr>
<td>Normalized transverse emittance</td>
<td>$\gamma\epsilon_0$</td>
<td>1</td>
<td>$\mu$m</td>
</tr>
<tr>
<td>Alpha function</td>
<td>$\alpha_0$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Beta function</td>
<td>$\beta_0$</td>
<td>10</td>
<td>m</td>
</tr>
<tr>
<td>Linear energy chirp</td>
<td>$h$</td>
<td>-27.5</td>
<td>1/m</td>
</tr>
<tr>
<td>Uncorrelated energy spread</td>
<td>$\sigma_E$</td>
<td>2</td>
<td>KeV</td>
</tr>
<tr>
<td>Momentum compaction</td>
<td>$R_{56}$</td>
<td>0.025</td>
<td>m</td>
</tr>
<tr>
<td>Radius of curvature</td>
<td>$r_0$</td>
<td>5</td>
<td>m</td>
</tr>
<tr>
<td>Magnetic length</td>
<td>$L_b$</td>
<td>0.5</td>
<td>m</td>
</tr>
<tr>
<td>Distance 1st−2nd, 3rd−4th bend</td>
<td>$L_1$</td>
<td>2.5</td>
<td>m</td>
</tr>
<tr>
<td>Distance 2nd−3rd bend</td>
<td>$L_2$</td>
<td>1</td>
<td>m</td>
</tr>
</tbody>
</table>
uses $s$ is the independent variable and $u$ is a dependent variable [29]. The equations of motion in the beam system are

$$
z' = -\kappa(s)x, \quad x' = p_z,
$$

$$
p_z' = \frac{q}{P_{r,c}}[t(s) + p_zn(s)] \cdot E_{||}(\hat{R}, s),
$$

$$
p_{x}' = \kappa(s)p_x + \frac{q}{P_{r,c}}[n(s) \cdot E_{||}(\hat{R}, s) - cB_y(\hat{R}, s)],
$$

where $t = \frac{d}{ds}t$ and $\hat{R} = R_r(s) + M(s)r$. Here $R_r(s) = \begin{bmatrix} Z_r(s)X_r(s) \end{bmatrix}^T$ gives the reference curve in the lab system, $t = \begin{bmatrix} Z_r'X_r' \end{bmatrix}^T$, $n = \begin{bmatrix} -X_r'Z_r' \end{bmatrix}^T$, $M = \begin{bmatrix} t, n \end{bmatrix}$, $r = \begin{bmatrix} z, x \end{bmatrix}^T$, and $P_r$ is the momentum of the reference particle.

The self-fields are retarded solutions of Maxwell’s equations, given by

$$
\left(\begin{array}{c}
E_{||} \\
B_y
\end{array}\right)(R, u) = -\frac{1}{4\pi} \int_0^u \int_0^{\pi} S[\hat{R}(\theta, v)] d\theta dv,
$$

where $\hat{R}(\theta, v) = R + (u - v)e(\theta)$ and $S$ is the lab system source term related to the charge/current density $\rho_L$, $J_L$ in the lab system. These densities are determined from the phase-space density $f_B$ in the beam system [29]. For more details see [11,29].

The numerical integration of Eq. (56) is done as follows. The inner $\theta$ integration is done with the trapezoidal rule on a uniform mesh with $N_\theta$ grid points and the outer $v$ integration is done with an adaptive Gauss-Kronrod algorithm with $N_v$ evaluations of the integrand. The computational effort of the field calculation on the 2D mesh with $N_z \times N_x$ grid points is therefore $\mathcal{O}(N_zN_xN_vN_\theta)$.

FIG. 16. Comparison of the mean power calculated with the 2D model [11] and the 1D field approximation scheme [28] for the FERMI@Elettra first bunch compressor. The 1D field approximation scheme is in very good agreement with the 2D model. The simulations have been done with the TFCT method.

FIG. 17. The same as in Fig. 16 for the $x$ emittance. The agreement is quite satisfactory. The final values of the $x$ emittance are almost the same.

B. 1D field approximation scheme

We now discuss a 1D field approximation to our 2D model. This approximation is based on the fact that the beam system spatial density is almost stationary in the $(\hat{z}, \hat{x})$ coordinates given by Eq. (38) of [11]. The 1D field approximation takes $\hat{x} = 0$ and the $\hat{R}$ in Eq. (55) becomes $\hat{R} = R_r(s) + M(s)[1 + hR_{s0}(s), hD(s)]^T \hat{z}$. The numerical results shown in Figs. 16–19 have been obtained with the TFCT method. The accuracy of the scheme is shown in Figs. 16 and 17. The final value of the transverse emittance is 1.44 $\mu$m, to be compared with 1.45 $\mu$m of the exact 2D model. The calculation is considerably faster than the 2D calculation as it reduces the computational cost by a factor proportional to the number of grid points in $\hat{x}$, i.e., the computational effort is $\mathcal{O}(N_zN_xN_\theta)$.

FIG. 18. Plot of $F_{z1}(\hat{z}, \hat{x}, s)$ in normalized coordinates at $s = 5$ m (exit third magnet) of the FERMI@Elettra first bunch compressor. The 1D field approximation scheme uses $F_{z1}(\hat{z}, 0, s)$ (blue line). It is evident here that the 1D field approximation scheme is considerably faster than the 2D calculation as it reduces the computational cost by a factor proportional to the number of grid points in $\hat{x}$.
relies on a weak dependency of $F_{z1}$ on its second argument for fixed values of the first, as illustrated in Figs. 18 and 19 where we have plotted $F_{z1}(\tilde{z}, \tilde{x}, s) = (q/P_{c})E_{0}(\tilde{R}, s) \cdot t(s)$ in $(\tilde{z}, \tilde{x})$ coordinates that put the $5\sigma$ range of the tilde variables on the square $[-1, 1]^2$. This weak dependence is satisfied to good approximation where the CSR force has its maximum strength, namely inside the magnets, with the worst case scenario for $s = 5 \text{ m}$ illustrated by Fig. 18.

C. Gain factor calculation

We calculate the gain factor for $25 \mu\text{m} \leq \lambda \leq 200 \mu\text{m}$ with the 1D approximation scheme and compare the results with the 2D model and analytical theory. The computational cost of the field calculation in the 2D model is $O(N_x N_y N_z)$, while in the 1D field approximation scheme is $O(N_x N_y)$. The computational cost for the density estimation based on TWT and TFCT methods is
dominated, for large particle numbers $N$, by the cost in the charge deposition cosine scheme which is $O(N)$, while the Monte Carlo cosine approximation has a computational load of $O(M_{f}N)$. According to Table III, for the calculation of the gain factor at $\lambda = 100 \, \mu m$ we have $N_{c} = 1024$, $N_{s} = 128$, $M_{c} = 100$, $M_{s} = 40$, and $N = 5 \times 10^{7}$. In typical simulations $N_{c} = 1000$, and, given that $N_{s} = N_{c}$, we have that the computational cost of the field calculation in the 2D model and in the Monte Carlo cosine approximation are both $O(10^{11})$. Therefore the use of both the TWT and TFCT methods reduces the computational load by a factor of 2 for $\lambda = 100 \, \mu m$. Notice that the computational cost of the field calculation in the 1D approximation scheme is $O(10^{9})$, still dominating over the cost in the charge deposition scheme which is $O(10^{7})$. The numerical simulations discussed in this paper have been done on the parallel cluster NERSC. The simulation for $\lambda = 100 \, \mu m$ was run on the NERSC Cray XT4 system Franklin and took approximately 8 h with the 2D model and approximately 5 min with the 1D field approximation scheme.

In Fig. 20 we compare the longitudinal density for $\lambda = 100 \, \mu m$ at the end of the chicane calculated with TWT and TFCT. We see clearly that TFCT with $N_{s} = 2048$ gives the same accuracy of TWT with $N_{c} = 1024$, thus proving that TWT is more efficient than TFCT in its ability to resolve small-scale structures. It is encouraging to see, however, that the results from the two methods converge as the grid resolution is increased.

In Fig. 21 we plot the gain factor as a function of the initial modulation wavelength $\lambda$. The blue line gives the results with the 1D field approximation scheme obtained with the TWT method, compared to the results with the 2D model obtained with the TFCT method, as discussed in [30]. Simulations akin to those in Fig. 20 show that the gain factor calculation with the two methods are quite close to each other. The 1D and 2D numerical results differ at short wavelengths from the analytical formula and agree quite well with each other, showing a decrease of the gain factor for the range $\lambda \approx 70–100 \, \mu m$. Figure 20 (bottom right) shows the 2D spatial density at the end of the chicane for $\lambda = 100 \, \mu m$. Figure 22 shows the longitudinal density at the end of the chicane for a range of values of $\lambda$; it clearly demonstrates (top left) that the modulation wavelength $\lambda = 25 \, \mu m$ leads to a mild microwave instability.

The discussion so far has been focused on the calculation of the gain factor, that by definition gives the amplification factor of an initial perturbation at a certain wavelength determined by the initial modulation and compression factor of the chicane. A more detailed study would consider an initial modulation with a more general spectral content, to study coupling between modes, and would analyze the amplification process in the full spectrum of the longitudinal density. Such a study has been performed in [11] with the 2D model. Moreover, in design studies of FELs, the distribution at the entrance of the bunch compressor is obtained by numerically tracking the beam from the injector to the bunch compressor. A more realistic modeling of the microbunching instability would therefore consist of the study of the evolution of an arbitrary distribution at the entrance of the bunch compressor. These studies are on our agenda.

A crucial parameter that affects the microbunching instability is the uncorrelated energy spread. An important prediction of the gain factor formula from [4] is that

![FIG. 21. Left: Analytical gain factor (green line) and numerical gain factor computed with the 2D model with the TFCT method (red line) and the 1D field approximation scheme with the TWT method (blue line). We notice that the 1D and 2D numerical results differ at short wavelengths from the analytical formula and agree quite well with each other, showing a decrease of the gain factor for the range $\lambda \approx 70–100 \, \mu m$. The derivation of the analytical gain factor formula is based on a linearized Vlasov treatment with a steady state CSR wake, therefore we do not expect close agreement with our numerical simulations. Right: Longitudinal density for $\lambda = 200 \, \mu m$ at $s = 0$ and $s = s_f$ without self-fields showing compression.](14, 070701 (2011) 070701-17)
increasing the uncorrelated energy spread reduces the gain factor. This led to a proposal, the laser heater, to increase the uncorrelated energy spread within FEL tolerance in order to damp the microbunching instability without degrading the FEL performance. An analysis of this effect together with the study of the evolution of an arbitrary distribution at the entrance of the bunch compressor will be discussed in a forthcoming paper.

FIG. 22. Longitudinal density at the exit of the chicane with and without self-fields for $\lambda = 25 \mu$m (top left), $\lambda = 50 \mu$m (top right), $\lambda = 75 \mu$m (center left), $\lambda = 100 \mu$m (center right), $\lambda = 150 \mu$m (bottom left), and $\lambda = 200 \mu$m (bottom right). The results are obtained with the TWT method. We clearly see (top left) that the modulation wavelength $\lambda = 25 \mu$m leads to a mild microwave instability.
VI. DISCUSSION AND CONCLUSION

We presented an approximation of the charged distribution function which provides a superior alternative to the cosine expansion currently in use in Bassi et al. [11] code. The new method deposits the particles on the finite grid—an approximation to the charged particle distribution which is intrinsically plagued by numerical noise—and then uses wavelet thresholding to remove some of this noise. The resulting wavelet-based grid approximation is appreciably more accurate—as quantified by the IMSE—and more than 3 orders of magnitude more efficient. As a result of this drastic computational speedup in the density estimation, the computational load for the simulations with the 2D model is reduced by a factor of 2, allowing a more efficient study of the evolution of the small-scale structures in the beam. Resolving these small-scale structures requires simulations with a very large number of point particles, which is prohibitively expensive with the original Monte Carlo method that scales as \( \propto M_r M_c N \). This improvement is even more dramatic for the simulations with the 1D field approximation scheme, where the computational load is reduced by 2 orders of magnitude, allowing for a very fast calculation of the microbunching instability.

One of the major computational efforts in the present implementation of Bassi’s code is the storage of the history of the beam on a 3D grid (2D in space and 1D in time). In the future, it will be beneficial to explore possibilities of further optimizing computational efficiency of the new wavelet-based approach by exploiting the sparsity of the charge representation (see Fig. 5). Storing the entire history of the charge distribution as a small set of sparse wavelet coefficients instead of the 3D full grid could provide substantial savings in memory and processor communication, thus significantly reducing simulation times. This study is currently underway.

ACKNOWLEDGMENTS

We gratefully acknowledge R. Ryne and P. Spentzouris for an account on NERSC. B. T. was partially supported by ONR Contract No. N00014-36-1-0587 while at Northern Illinois University. This work was supported under U.S. DOE Contract No. DE-AC05-06OR23177.

APPENDIX A: VARIANCE OF NOISE FOR THE PARTICLE DEPOSITION SCHEMES

In this Appendix we detail the process of sampling a continuous charge density distribution by \( N \) particles, with subsequent charge deposition on a grid. The detailed discussion of the two lowest-order CDS—NGP (\( k = 1 \)) and CIC (\( k = 2 \))—is presented in Appendix A of [14]. Here we extend their discussion to TSC (\( k = 3 \)) and present the generalized approach for an arbitrary order CDS (\( k > 3 \)).

We calculate the expectation and variance of \( Q(0) \), the total charge assigned to the \( i \)th node of the lattice, for the general CDS (arbitrary \( k \)). Each particle carries the same charge \( q_0 = Q_{tot}/N \).

The total charge \( Q(i) \) deposited onto the grid point \( i \) can be interpreted as the sum of \( N \) independent and identically distributed random variables \( \{Q_1, \ldots, Q_N\} \):

\[
Q(i) = \sum_{k=1}^{N} Q_k;
\]

one may consider this sampling process as particles being added one at the time. All \( \{Q_k\} \) are then distributed as the “prototype” random variable \( Q \), the charge assigned to the grid point \( i \) when a new particle is added to the sample. We also introduce an auxiliary variable \( I \), which assumes value 1 if the sampled position of a \( Q_k \) is within the domain \( \Omega(i) \) of the CDS function. For the \( k \)th order CDS, \( \Omega(i) \) is defined to be a \( D \)-dimensional cube of edge length \( kh \), centered on the node \( i \).

The working assumption here is that the grid is sufficiently fine for the probability density function to be approximated by a linear function on \( \Omega(i) \). After adopting this, one can readily show:

\( k = 1 \):

\[
E[Q^2|I = 1] = \frac{q_0^2}{2D}, \quad E[Q|I = 1] = \frac{q_0}{2},
\]

\[
\text{Var}(Q|I = 1) = \left[ \left( \frac{1}{2} \right)^D - \left( \frac{1}{4} \right)^D \right] q_0^2,
\]

\( k = 2 \):

\[
E[Q^2|I = 1] = \frac{q_0^2}{3D}, \quad E[Q|I = 1] = \frac{q_0}{3},
\]

\[
\text{Var}(Q|I = 1) = \left[ \left( \frac{1}{3} \right)^D - \left( \frac{1}{9} \right)^D \right] q_0^2.
\]

\( k = 3 \):

\[
E[Q^2|I = 1] = \frac{q_0^2(11)^D}{60}, \quad E[Q|I = 1] = \frac{q_0}{3D},
\]

\[
\text{Var}(Q|I = 1) = \left[ \left( \frac{11}{60} \right)^D - \left( \frac{1}{9} \right)^D \right] q_0^2.
\]

\( k > 3 \):

\[
E[Q^2|I = 1] = q_0^2 \left[ \frac{1}{kh} \int_{-(k/2)h}^{(k/2)h} d^D k(x) dx \right]^D = q_0^2 \tilde{a}_k^2,
\]

\[
E[Q|I = 1] = \frac{q_0}{k^D},
\]

\[
\text{Var}(Q|I = 1) = \left[ \tilde{a}_k^2 - \left( \frac{1}{k^2} \right)^D \right] q_0^2.
\]

\( E[Q|I = 0] = 0 \) and \( \text{Var}(Q|I = 0) = 0 \) for all CDS.

We now define a new random variable \( K \) as the number of particles for which \( I = 1 \). In a given realization, \( K \) will
have a value \( k \) between 0 and \( N \) (\( 0 \leq k \leq N \)). For any given \( k \),
\[
E[Q^{(i)}|K = k] = kE[Q|I = 1] \tag{A10}
\]
and
\[
\text{Var}(Q^{(i)}|K = k) = k\text{Var}(Q|I = 1), \tag{A11}
\]
because the \( \{Q_1, \ldots, Q_N\} \) are independent and identically distributed (as \( Q \)). \( K \) itself is a random variable, so the expectation and variance of \( Q^{(i)} \) has to be calculated using the double expectation theorem:
\[
E[Q^{(i)}] = E[E(Q^{(i)}|K)] = E[K]E[Q|I = 1]. \tag{A12}
\]
\[
\text{Var}(Q^{(i)}) = E[\text{Var}(Q^{(i)}|K)] + \text{Var}(E[Q^{(i)}|K])
= E[K]\text{Var}(Q|I = 1) + (E[Q|I = 1])^2\text{Var}(K). \tag{A13}
\]
The calculation of \( E[K] \) and \( \text{Var}(K) \) that enters the expressions above is aided by the realization that \( K \) is the number of “successes” \( (I = 1) \) in a series of \( N \) trials. It is, therefore, distributed according to a binomial distribution with “probability of success” \( P_i \) equal to the continuous probability density function integrated over the \( \Omega^{(i)} \):
\[
E[K] = NP_i, \tag{A14}
\]
\[
\text{Var}(K) = NP_i(1 - P_i). \tag{A15}
\]
Analogously to \( P_i \), we define \( p_i \) as the value of the integral of the continuous probability density over \( \omega^{(i)} \), a \( D \)-dimensional cube of edge length \( h \) centered on the node \( i \). With this notation:

\( k = 1 \):
\[
P_i \approx 2^D p_i. \tag{A16}
\]
\( k = 2 \):
\[
P_i \approx 2^D p_i. \tag{A17}
\]
\( k = 3 \):
\[
P_i \approx 3^D p_i. \tag{A18}
\]
\( k > 3 \):
\[
P_i \approx k^D p_i. \tag{A19}
\]
After combining Eqs. (A2), (A4), (A6), (A8), (A10), (A12), and (A14) and (A16)–(A19), we obtain
\[
E[Q^{(i)}] = q_0^2 N p_i \tag{A20}
\]
for all CDS, and, finally, \( k = 1 \):
\[
\text{Var}(Q^{(i)}) = q_0^2 NP_i \left(1 - \frac{P_i}{4^D}\right) = q_0^2 NP_i (1 - p_i). \tag{A21}
\]
\( k = 2 \):
\[
\text{Var}(Q^{(i)}) = q_0^2 NP_i \left(1 - \frac{P_i}{4^D}\right) = q_0^2 NP_i \left(\frac{2^D}{3} - p_i\right). \tag{A22}
\]
\( k = 3 \):
\[
\text{Var}(Q^{(i)}) = q_0^2 NP_i \left(1 - \frac{P_i}{4^D}\right) = q_0^2 NP_i \left(\frac{11}{20} - p_i\right). \tag{A23}
\]
\( k > 3 \):
\[
\text{Var}(Q^{(i)}) = q_0^2 NP_i \left(1 - \frac{P_i}{4^D}\right) = q_0^2 NP_i \left(k^D a_k^2 - p_i\right).
\]
We introduce a global measure of the error associated with noise due to sampling and charge deposition onto a discrete grid
\[
\sigma^2 \equiv \frac{1}{N_{\text{grid}}^2} \sum_{i=1}^{N_{\text{grid}}} \text{Var}(Q^{(i)}). \tag{A24}
\]
After substituting Eqs. (A21)–(A24) into Eq. (A24), we observe that the second term is bounded,
\[
\frac{1}{N_{\text{grid}}} \leq \sum_{i=1}^{N_{\text{grid}}} p_i^2 \leq 1. \tag{A25}
\]
The lower bound is reached for the uniform distribution, while the upper bound is attained when the whole distribution is assigned to a single grid point. The charge density distribution in a typical PIC simulation very rarely possesses highly prominent, strongly localized peaks; furthermore, the grid is always adjusted so as to minimize (or, at least, greatly reduce) the number of grid points in the regions of zero density. This means that the value of the term will be much closer to its lower end, thus yielding it essentially negligible in computing \( \sigma^2 \) using Eq. (A24). Using this simplification, for the CDS schemes considered here we obtain
\[
\sigma_k^2 = \frac{q_0^2 \delta N}{N_{\text{grid}}} \approx \frac{q_0^2 \delta N_{\text{tot}}}{NN_{\text{grid}}}, \tag{A26}
\]
where \( a_k \) are given in Eq. (19). With this formalism, one can compute \( \sigma_k \) for the CDS \( d_k(x) \) of any order \( k \). As we mentioned before, one can easily compute the \( k \)th order CDS by convolving the \( k - 1 \)th order CDS with the second-order CDS \( d_2(x) \) [13].

**APPENDIX B: NEAR-EQUIVALENCE OF MONTE CARLO COSINE AND TFCT EXPANSIONS FROM A NGP GRID**

The cosine coefficients computed from a NGP grid are given by combining Eqs. (22) and (24):

\[
070701-20
\]
\[ \theta_{nm} = q_0 \sum_{i=1}^{N} \int_{0}^{1} \int_{-h/2}^{h/2} \frac{\delta(z - z_i - \frac{\varepsilon}{2})(x - x_i - \frac{\varepsilon}{2})d_1(\varepsilon)d_1(x)\phi_n(z)\phi_m(x)dx d\varepsilon dz \]

\[ = q_0 \sum_{i=1}^{N} \int_{-h/2}^{h/2} \int_{-h/2}^{h/2} \int_{0}^{1} \delta(z - z_i - \frac{\varepsilon}{2})(x - x_i - \frac{\varepsilon}{2})\phi_n(z)\phi_m(x)dx d_1(\varepsilon)d_1(z)d\varepsilon dz \]

\[ = q_0 \sum_{i=1}^{N} \int_{-h/2}^{h/2} \phi_n(z_i)\phi_m(z_i + \frac{\varepsilon}{2})d_1(\varepsilon)d\varepsilon d\varepsilon dx \]

\[ = q_0 \sum_{i=1}^{N} \phi_n(z_i)\phi_m(x_i) + O(h^2) + O(h^2). \quad (B1) \]

The sum \( q_0 \sum_{i=1}^{N} \phi_n(z_i)\phi_m(x_i) \) is the expression for the coefficients in the Monte Carlo expansion, which can be readily seen from Eq. (8). The same treatment can easily be extended to the coefficients of the current density \( \theta_{nm} \). In the previous calculation, we first switched the order of integration, then integrated over \( z \) and \( x \), after which we expanded in Taylor series \( \phi(z + \varepsilon) \) around \( z_i \) and \( \phi(x + \varepsilon) \) around \( x_i \) (because \( z_i \gg h_z \) and \( x_i \gg h_x \)). In the Taylor expansion, the linear term vanishes since it integrates an odd function over an even interval. Therefore, the final result is second-order accurate. It should come as no surprise that

\[ \lim_{h_z, h_x \to 0} \theta_{nm} = q_0 \sum_{i=1}^{N} \phi_n(z_i)\phi_m(x_i). \quad (B2) \]

because \( \lim_{h \to 0} d_1(x) = \delta(x) \), which is the “deposition function” used in Monte Carlo cosine expansion.

**APPENDIX C: ERROR IN USING FINITE DIFFERENCE SCHEME TO EVALUATE DERIVATIVES**

Finite difference schemes are used to evaluate derivatives of a gridded distribution. The order of the derivative approximation depends on the number of points used in the differencing, i.e., the size of the stencil. The \( p \)th order accurate \( \ell \)th \( z \) derivative computed using a differencing scheme can be written as

\[ D_{z}^{(\ell)} f(z_j, x_j) = \sum_{n=0}^{M} \sum_{m=0}^{M} \theta_{nm} \phi_m(x_j) \sum_{k} c_k \phi_n(z_j + k h_z). \quad (C1) \]

\( c_k \) are the relative weights on the points of the stencil (see Table V). The range of \( k \) is over the entire stencil centered at \( z \). Earlier, we defined the IMSE as the sum of the square of differences between the numerically computed solution, such as the one given in Eq. (C1), and the exact solution

\[ \partial_{x}^{(\ell)} f(z_i, x_j) = \sum_{n=0}^{M} \sum_{m=0}^{M} \theta_{nm} \phi_n(z_i) \phi_m(x_j). \quad (C2) \]

where \( \theta_{nm} \) are the exact coefficients of the distribution. The exact and the numerically computed coefficients are related by \( \theta_{nm} = \tilde{\theta}_{nm} + \delta \theta_{nm} \).

It can be readily shown that

\[ \sum_{k} c_k \phi_n(z_i + k h_z) = \phi_n(z_i) + \delta \phi_n. \quad (C3) \]

where

\[ \delta \phi_n \propto O(h^p) = O(N_z^{-p}). \]

It follows that the IMSE for the \( \ell \)th \( z \) derivative is

\[ e_{z} = h_z h_x \sum_{\ell=1}^{N_z} \sum_{j=1}^{N_x} \left[ D_{z}^{(\ell)} f(z_i, x_j) - \partial_{x}^{(\ell)} f(z_i, x_j) \right] \]

\[ = h_z h_x \sum_{\ell=1}^{N_z} \sum_{j=1}^{N_x} \sum_{m=0}^{M} \sum_{n=0}^{M} \theta_{nm} \delta \phi_n + \delta \theta_{nm} \phi_n^{(\ell)}(z_i) \]

\[ + \delta \theta_{nm} \phi_n^{(\ell)} \quad (C4) \]

The first term in the bracket is due to the error of the finite difference scheme, and depends only on the type (order) of the scheme. The second is due to the error in computing

**TABLE V. Coefficients \( h_z c_k \) for \( p \)th order first derivative difference operator in \( z \) coordinate.**

<table>
<thead>
<tr>
<th>( p )</th>
<th>( -3 )</th>
<th>( -2 )</th>
<th>( -1 )</th>
<th>( 0 )</th>
<th>( 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>3</td>
<td>-1/3</td>
<td>3/2</td>
<td>-3</td>
<td>11/6</td>
<td></td>
</tr>
</tbody>
</table>
expansion coefficients, and depends on the number of particles (N) and the grid resolution parameters \((N_z, N_x)\).

The third term is due to both error in computing expansion coefficients and in the finite difference scheme. This term is of the second order, and therefore not a significant contributor to the overall error, which is why we ignore it in our analysis.

From our earlier discussion, the IMSE for the distribution scales as \(N^{-1}\), and can be written as

\[
\varepsilon = h_z h_x \sum_{i=1}^{N_z} \sum_{j=1}^{N_x} \left( f(z_i, x_j) - \tilde{f}(z_i, x_j) \right)^2
\]

\[
= \sum_{n=0}^{M_z} \sum_{m=0}^{M_x} \delta \theta_{nm}^2 \left[ h_z h_x \sum_{i=1}^{N_z} \sum_{j=1}^{N_x} \phi_n^2(z_i) \phi_m^2(x_j) \right]
\]

\[
\approx \sum_{n=0}^{M_z} \sum_{m=0}^{M_x} \delta \theta_{nm}^2 \propto N^{-1}.
\]  

The last line follows from the fact that in the limit of increasing resolution \((h_z, h_x \to 0)\), the bracketed term converges to the integral

\[
\int_0^1 \int_0^1 \phi_n^2(z) \phi_m^2(x) dx dz = 1.
\]

Therefore, the total IMSE in the approximation of derivatives consists of two main constituent errors which scale as \(O(N^{-2p})\) and \(O(N^{-1})\); their relative importance changes with the parameters of the simulation.

Figure 23 shows the accuracy of estimation of the first derivative in \(z\) as a function of the length scale \(\lambda\) for the grid, TFCT, TWT, and the exact distribution [the first term in the brackets of Eq. (C4)], evaluated with a difference scheme of the first (top), second (middle), and third (bottom) order. It is evident that the error of the derivative estimate is dictated by the inaccuracy in estimating expansion coefficients, and that using the higher-order difference scheme for evaluation of derivatives does not lead to more accurate estimates. We also notice that the estimates of the first derivatives using TFCT and TWT are quite similar.