OVERCOMING THE NUMERICAL SIGN PROBLEM IN THE WIGNER DYNAMICS VIA ADAPTIVE PARTICLE ANNIHILATION

YUNFENG XIONG∗ AND SIHONG SHAO†

Abstract. The infamous numerical sign problem poses a fundamental obstacle to particle-based stochastic Wigner simulations in high dimensional phase space. Although the existing particle annihilation via uniform mesh significantly alleviates the sign problem when dimensionality $D \leq 4$, the mesh size grows dramatically when $D \geq 6$ due to the curse of dimensionality and consequently makes the annihilation very inefficient. In this paper, we propose an adaptive particle annihilation algorithm, termed Sequential-clustering Particle Annihilation via Discrepancy Estimation (SPADE), to overcome the sign problem. SPADE follows a divide-and-conquer strategy: Adaptive clustering of particles via controlling their number-theoretic discrepancies and independent random matching in each group, and it may learn the minimal amount of particles that can accurately capture the non-classicality of the Wigner function. Combining SPADE with the variance reduction technique based on the stationary phase approximation, we attempt to simulate the proton-electron couplings in 6-D and 12-D phase space. A thorough performance benchmark of SPADE is provided with the reference solutions in 6-D phase space produced by a characteristic-spectral-mixed scheme under a $73^3 \times 80^3$ uniform grid, which fully explores the limit of grid-based deterministic Wigner solvers.

AMS subject classifications: 81S30; 60J85; 65C05; 62G09; 35Q40

Keywords: Wigner equation; branching random walk; negative particle method; sign problem; particle annihilation; Coulomb interaction

1. Introduction. During the past decades, it has burgeoned with a wide spectrum of applications of the Wigner quantum dynamics [1] in the fields of semiconductor devices [2, 3, 4, 5], nano-materials [6, 7, 8], high energy physics [9] and quantum tomography [10, 11] for its huge theoretical advantage in resolving the classical-quantum dichotomy [1, 12], as well as its experimental observability [11]. However, in contrast to the prosperity of both theoretical and experimental advances, there remains a huge gap in numerically solving the 6-D or higher dimensional Wigner quantum dynamics because of the well-known curse of dimensionality (CoD).

Grid-based deterministic Wigner solvers are able to produce highly accurate results owing to their solid mathematical theory and concise guiding principle [13, 14, 15], but both the computational cost and memory storage become extremely demanding when the dimensionality $D \geq 6$ due to their unfavorable scaling. Alternatively, one can recourse to particle-based stochastic methods, including particle affinity method [16], signed-particle Wigner Monte Carlo [12, 3, 6, 4], random cloud model [17, 18] and Wigner branching random walk (WBRW) [19, 20, 21], for their theoretical convergence of $N^{-1/2}$ with $N_0$ being the initial effective particle number (sample size), regardless of $D$. However, even the state-of-the-art stochastic algorithms are still restricted in 4-D phase space [20, 8] and few results have been reported for 6-D problems.

The formidable obstacle there turns out to be the notorious numerical sign problem [22, 23, 21], say, the exponential growth of both particle number and stochastic variance due to the increments of negative weights, which is generally believed to be NP-hard [24, 25]. This work follows the latter. Specifically, we propose an adaptive particle annihilation within the framework of WBRW, termed Sequential-clustering Particle Annihilation via Discrepancy Estimation (SPADE) [26], to overcome the sign problem.

∗CAPT, LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, China. Email addresses: xiongy@math.pku.edu.cn (Y. Xiong) sihong@math.pku.edu.cn (S. Shao).
†To whom correspondence should be addressed.
In our preceding work [21], we have pointed out that the sign problem is inherited in the widely used particle splitting technique [12] (the Hahn-Jordan decomposition) for the pseudodifferential operator (ΨDO), because it ignores the cancelation of particle trajectories with opposite sign and leads to a rapid growth of variances. The remedies are to fully utilize the near-cancellation of positive and negative weights. One approach directly aims at reducing stochastic variances, such as the semiclassical approximation [27], the fractional particle weights [20] and the stationary phase approximation (SPA) [21]. These methods are able to suppress the exponential growth of variances efficiently, albeit not completely eliminating it. The other approach is particle annihilation (PA), including PA via uniform mesh (PAUM for brevity) [12, 3, 6, 4, 20] and the particle resampling by filtering out the high-frequency components [28]. But the usage of existing PA methods to 6-D problems is highly non-trivial. For instance, the most popular PAUM is bothered by CoD as the mesh size grows exponentially in D, so that many particles are left uncanceled when the bin size largely exceeds the particle number [28, 29].

The proposed SPADE tries to ameliorate CoD by a two-step strategy: Adaptive clustering of particles via controlling their number-theoretic discrepancies, partially borrowing the pioneering idea in the non-parametric high-dimensional density estimation [30], and independent random matching among positive and negative particles in each cluster. In other words, SPADE gets rid of the reliance on a fixed grid mesh and overcomes the essential drawbacks in PAUM, and is capable to recover the “bottom line structure” [19], an indicator that describes the minimal amount of particles that can accurately capture the non-classicality, as well as the oscillation, of the Wigner function. Therefore, it may greatly facilitate realistic simulations, e.g., many-body problems in high-dimensional phase space. It deserves to mention that the calculation of the discrepancy of a sequence, as a pivotal step in adaptive clustering, is NP-hard in nature [31]. In a sense, SPADE resolves the numerical sign problem inherited from CoD by seeking efficient heuristic approximations to another NP-hard problem.

Combining SPADE and SPA together in WBRW, we succeed in simulating the proton-electron coupling, which is a typical non-equilibrium quantum dynamics under the Coulomb interaction [32, 8] and serves as the prototype for the Coulomb collisions [33, 9]. A thorough benchmark on 6-D simulations has been made to evaluate the performance of SPADE. For the sake of comparison, we endeavor to produce reference solutions by a massively parallel characteristic-spectral-mixed scheme [13], in which the Wigner function is represented as a tensor product of $75^3$ cubic spline basis in $x$-space and $80^3$ Fourier basis in $k$-space (with mesh size $73^3 \times 80^3 \approx 2 \times 10^{11}$) to attain high accuracy. Numerical results manifest that SPADE can be systematically improved by either increasing the sample size or deepening the partition. This constitutes the solid preparation for our attempt to obtain the first-principle solution to proton-electron coupling in 12-D phase space, where both proton and electron are treated quantum mechanically, and may potentially pave the way for the interlacement of kinetic theory and molecular dynamics in high energy density physics [9].

The rest is organized as follows. Section 2 briefly reviews the Wigner function formalism for quantum mechanics in phase space and illustrates the basic idea behind SPA. Section 3 details the WBRW-SPA model (i.e. using SPA in WBRW) for the Coulomb system. Section 4 focuses on the intuition and design of SPADE. Numerical simulations on 6-D and 12-D proton-electron couplings are reported in Sections 5 and 6, respectively. Finally, conclusions and discussions are drawn in Section 7.
2. Background. As a preliminary, we give a brief review of the Wigner quantum dynamics and the physical intuition behind SPA. The \( N \)-body Wigner function is defined by the Weyl-Wigner transform of density matrix \( \rho(x, y, t) \):

\[
\int_{\mathbb{R}^{Nd}} \rho(x - \frac{y}{2}, x + \frac{y}{2}, t) e^{-i k \cdot y} dy,
\]

with the spatial dimension \( d \) and the dimensionality of phase space \( D = 2Nd \). The quantum dynamics of the Wigner function is governed by the Wigner equation,

\[
\frac{\partial}{\partial t} f(x, k, t) + \frac{\hbar k}{m} \cdot \nabla_x f(x, k, t) = \Theta_V[f](x, k, t),
\]

where \( k/m \) denotes \( (k_1/m_1, \ldots, k_N/m_N) \) with \( k_i \) and \( m_i \) the wave vector and mass for the \( i \)-th body, respectively, \( \hbar \) is the reduced Planck constant, and \( \Psi_{DO} \) reads as

\[
\Theta_V[f](x, k, t) = \frac{1}{i\hbar(2\pi)^{Nd}} \int_{\mathbb{R}^{2Nd}} e^{-i(k-k') \cdot y} (V(x + \frac{y}{2}) - V(x - \frac{y}{2})) f(x', k', t) dy dk'.
\]

A profound advantage of the Wigner function is its simplicity in visualization in both position and momentum by projecting it onto a lower dimensional space [1]. For instance, the reduced Wigner function along the \( j \)-th coordinate reads that

\[
W_j(x, k, t) = \int_{\mathbb{R}^{Nd-1} \times \mathbb{R}^{Nd-1}} f(x, k, t) dx_{(j)} dk_{(j)},
\]

and \( x_{(j)} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{Nd}) \), \( k_{(j)} = (k_1, \ldots, k_{j-1}, k_{j+1}, \ldots, k_{Nd}) \). The spatial marginal distribution in the \( j \)-th coordinate is

\[
P_j(x, t) = \int_{\mathbb{R}^{Nd-1} \times \mathbb{R}^{Nd-1}} f(x, k, t) dx_{(j)} dk.
\]

2.1. Quantum Coulomb collision. The Coulomb interaction is of great interest in quantum science [9, 8]. Consider a two-body system composed of one electron and one proton and treat both quantum mechanically, with their coordinates in phase space denoted by \( (x, k) = (x_e, x_p, k_e, k_p) \) with \( x_e, x_p, k_e, k_p \in \mathbb{R}^3 \).

Under the atomic unit \( \hbar = m_e = e = 1 \), \( m_p \approx 1836 m_e \) and attractive Coulomb potential \( V(x) = -1/|x_e - x_p| \), one can directly obtain \( \Psi_{DO} \) for the Hydrogen model,

\[
\Theta_V[f](x, k, t) = \frac{1}{i\hbar c_{3,1}} \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) dk'\]

\[
- \frac{1}{i\hbar c_{3,1}} \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e + \frac{k'}{2}, k_p - \frac{k'}{2}, t) dk',
\]

with \( c_{n,\alpha} = \pi^{n/2} \Gamma(\frac{n+\alpha}{2}) / \Gamma(\frac{n-\alpha}{2}) \). Despite two-body Hydrogen wave functions are exactly solvable, the phase space solution from the first principle is less than straightforward or complete, especially for its non-equilibrium dynamics [34].

As a comparison, we shall also investigate the single-body electron Wigner equation under the adiabatic approximation [9], motivated by the fact \( m_p \gg m_e \) so that the proton moves much slower than the electron. Let us consider the marginal electron Wigner function by projecting \( f(x_e, x_p, k_e, k_p, t) \) onto \( (x_e, k_e) \) space

\[
P_e(x_e, k_e, t) = \int_{\mathbb{R}^3 \times \mathbb{R}^3} f(x_e, x_p, k_e, k_p, t) dx_p dk,
\]

and suppose the following two assumptions hold.
(1) The initial Wigner function is a product of uncorrelated electron Wigner function \( f_e \) and proton Wigner function \( f_p \),
\[
f(x_e, x_p, k_e, k_p, 0) = f_e(x_e, k_e, 0) f_p(x_p, k_p, 0).
\]

(2) There exists \( g(x_e, x_p, k_e, k_p, t) = O(1) \), a fixed \( x_A \in \mathbb{R}^3 \) and a small \( \varepsilon > 0 \) such that \( f(x_e, x_p, k_e, k_p, t) \) is localized in \( x_p \)-variable for \( t = O(1) \), say,
\[
f(x_e, x_p, k_e, k_p, t) = g(x_e, x_p, k_e, k_p, t) e^{-|x_p - \frac{\hbar k_p t}{m_p} - x_A|^2/\varepsilon}.
\]

Then we can formally derive the adiabatic approximation in the asymptotic sense,
\[
P_e(x_e, k_e, t) = f_e(x_e, k_e, t) + O(h m_p^{-1} \varepsilon^{3/2}),
\]
where \( f_e(x_e, k_e, t) \) satisfies the single-body electron Wigner equation
\[
\frac{\partial}{\partial t} f_e(x_e, k_e, t) + \frac{\hbar k_e}{m_e} \cdot \nabla_{x_e} f_e(x_e, k_e, t) = \Theta_V[f_e](x_e, k_e, t),
\]
and \( \Psi_{DO} \) reads that
\[
\Theta_V[f_e](x_e, k_e, t) = \frac{1}{16\pi^3} \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_A)}}{|k'|^2} \left( f_e(x_e, k_e - \frac{k'}{2}, t) - f_e(x_e, k_e + \frac{k'}{2}, t) \right) dk'.
\]

A brief derivation of Eq. (2.10) is given as below. Since
\[
\int d\mathbf{x} d\mathbf{p} \int_{\mathbb{R}^3} \frac{1}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) dk' = \int d\mathbf{x} d\mathbf{p} \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_A)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) dk' + \int d\mathbf{x} d\mathbf{p} \int_{\mathbb{R}^3} (e^{i k' \cdot (x_e - x_A)} - 1) \frac{e^{i k' \cdot (x_e - x_A)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) dk',
\]
the first term corresponds to \( \Theta_V[P_e](x_e, k_e, t) \). For the second term, as \( f \) is localized in \( x_p \)-space, using the Laplace asymptotics near \( x_A + \hbar k_p t / m_p \), it is governed by
\[
(2\pi \varepsilon)^3/2 \int d\mathbf{k} \int d\mathbf{k}_p \int_{\mathbb{R}^3} (e^{-i k' \cdot (x_e - x_A)} - 1) \frac{e^{i k' \cdot (x_e - x_A)}}{|k'|^2} g(x_e, x_A + \frac{\hbar k_p t}{m_p}, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) dk' = O(h m_p^{-1} \varepsilon^{3/2})
\]
the leading term of which is \( O(h m_p^{-1} \varepsilon^{3/2}) \) due to the factor \( \exp(-i k' \cdot \frac{\hbar k_p t}{m_p}) - 1 \).

Both \( \Psi_{DO} \) (2.6) and (2.12) have an intuitive scattering interpretation as the quantum interaction is given by averaging the inner-scattering state \( (x_e, x_p, k_e + \frac{k'}{2}, k_p + \frac{k'}{2}) \) for Eq. (2.6) or outer-scattering states \( (x_e, k_e \pm \frac{k'}{2}) \) for Eq. (2.12) weighted with the phase factor, with the Riesz potential \( |k'|^{-2} \) playing the role as the distribution [33, 12].

### 2.2. Stationary phase approximation
The quantum Coulomb interaction indeed decays as the two-body displacement \( |x_e - x_p| \) increases, because the phase factor becomes more and more oscillating. To characterize the decay property more precisely, we need to introduce a filter \( \lambda_0 \) and a decomposition of \( \Psi_{DO} \) (2.6),
\[
\Theta_V[f](x, k, t) = \Lambda^{<\lambda_0}[f](x, k, t) + \Lambda^{>\lambda_0}[f](x, k, t),
\]

where \( \Lambda^{<\lambda_0} \) and \( \Lambda^{>\lambda_0} \) are defined as
\[
\Lambda^{<\lambda_0}[f](x, k, t) = \int d\mathbf{x} d\mathbf{p} \int_{\mathbb{R}^3} \frac{1}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) dk',
\]

and
\[
\Lambda^{>\lambda_0}[f](x, k, t) = \int d\mathbf{x} d\mathbf{p} \int_{\mathbb{R}^3} \frac{1}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) dk'.
\]
where the low-frequency component \( \Lambda^{\leq \lambda_0}[f](x, k, t) \) reads that

\[
\Lambda^{\leq \lambda_0}[f](x, k, t) = \int_{\mathcal{B}(x, p)} \frac{\sin((x_e - x_p) \cdot k')}{\hbar c_3, 1 |k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) \, dk' - \int_{\mathcal{B}(x, p)} \frac{\sin((x_e - x_p) \cdot k')}{\hbar c_3, 1 |k'|^2} f(x_e, x_p, k_e + \frac{k'}{2}, k_p - \frac{k'}{2}, t) \, dk',
\]

and \( \mathcal{B}(r) \) is a ball centered at 0 with radius \( r \). When \( \lambda_0 \geq 1 \), the high-frequency component allows an asymptotic expansion \([21]\), in the light of SPA,

\[
(2.14) \quad \Lambda^{> \lambda_0}[f](x, k, t) = \Lambda^{> \lambda_0}[f](x, k, t) + \Lambda^{> \lambda_0}[f](x, k, t) + O(\lambda_0^{-3/2}),
\]

where two principal terms in the asymptotic expansion are

\[
\Lambda^{> \lambda_0}[f](x, k, t) = \frac{\pm 4 \pi}{\hbar c_3, 1} \int^{+\infty}_{-\infty} \frac{\sin(r|x_e - x_p|)}{r|x_e - x_p|} f(x, k_e \pm \frac{r\sigma_e(x)}{2}, k_p \pm \frac{r\sigma_e(x)}{2}, t) \, dr,
\]

and the critical point is parameterized by \( \sigma_e(x) = (\cos \theta^*, \sin \theta^* \cos \phi^*, \sin \theta^* \sin \phi^*) \),

\[
(2.15) \quad \theta^* = \arctan2(\sqrt{(x_{e,2} - x_{p,2})^2 + (x_{e,3} - x_{p,3})^2}, x_{e,1} - x_{p,1}),
\]

\[
\phi^* = \arctan2(x_{e,3} - x_{p,3}, x_{e,2} - x_{p,2}),
\]

with \( x_e = (x_{e,1}, x_{e,2}, x_{e,3}) \) and \( x_p = (x_{p,1}, x_{p,2}, x_{p,3}) \).

Now \( \PsiDO(2.6) \) decays asymptotically when \( |x_e - x_p| \) increases. For the low-frequency component, by scaling \( k' \to |x_e - x_p| \), it yields \( \Lambda^{\leq \lambda_0}[f](x, k, t) = O(|x_e - x_p|^{-1}) \) when \( |x_e - x_p| \) is sufficiently large. At the same time, the major contribution of the high-frequency component is determined by the scattering event along or opposite to the direction \( (x_e - x_p)/|x_e - x_p| \), while the contributions deviated from that line almost cancel out. By the integration by parts, the principal asymptotic terms behave like \( \Lambda^{> \lambda_0}[f](x, k, t) = O(|x_e - x_p|^{-1}) \) for large \( |x_e - x_p| \).

Similarly, the low-frequency component of single-body \( \PsiDO(2.12) \) reads that

\[
\Lambda^{\leq \lambda_0}[f_e](x_e, k_e, t) = \frac{1}{\hbar c_3, 1} \int_{\mathcal{B}(x_e - x, A)} \frac{\sin((x_e - x) \cdot k')}{|k'|^2} f_e(x_e, k_e - \frac{k'}{2}, t) \, dk' - \frac{1}{\hbar c_3, 1} \int_{\mathcal{B}(x_e - x, A)} \frac{\sin((x_e - x) \cdot k')}{|k'|^2} f_e(x_e, k_e + \frac{k'}{2}, t) \, dk',
\]

and SPA to the high-frequency component reads that

\[
\Lambda^{> \lambda_0}[f_e](x_e, k_e, t) = \pm \frac{4 \pi}{\hbar c_3, 1} \int^{+\infty}_{-\infty} \frac{\sin(r|x_e - x|)}{r|x_e - x|} f(x_e, k_e \pm \frac{r\sigma_e(x)}{2}, t) \, dr,
\]

where the critical point \( \sigma_e(x) \) is given in Eq. (2.15) by replacing \( x_p \to x_A \).
3. Numerical sign problem: Fundamental obstacle in negative particle method. The stochastic particle method for the deterministic Wigner equation \( \text{Eq. (2.2)} \) is based on its stochastic representation, which interprets the formal Neumann series expansion as the expectation of stochastic trajectories over Poisson jumps [12, 17, 20]. The relevant conceptual advance of the Wigner Monte Carlo is particle splitting [12], making quantum algorithm distinct from the Direct Simulation Monte Carlo [5], and the initial data \( \phi \) perspective, the Wigner Monte Carlo utilizes the fact that, for any test function \( \text{sign problem} \), and show how SPA can help alleviate such problem. The quantum Coulomb interaction, along with an illustrative description of numerical trajectories via SPA [21]. In the subsequent part, we will discuss WBRW-SPA for leads to a rapid growth of random noises. The palliative is to cancel out the stochastically [12]. Despite its vivid physical intuition and convenience in implementation, the direct splitting of \( \Psi DO \) ignores the cancelations of an oscillatory integral and leads to a rapid growth of random noises. The palliative is to cancel out the stochastic trajectories via SPA [21]. In the subsequent part, we will discuss WBRW-SPA for the quantum Coulomb interaction, along with an illustrative description of numerical sign problem, and show how SPA can help alleviate such problem.

3.1. Particle generation and variance reduction. From the mathematical perspective, the Wigner Monte Carlo utilizes the fact that, for any test function \( \varphi(x, k) \in L^2_{\text{loc}}(\mathbb{R}^d \times \mathbb{R}^d) \), an inner product \( \langle f, g \rangle = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x, k)g(x, k)\text{d}x\text{d}k \), and the initial data \( f_0 \in L^2(\mathbb{R}^d \times \mathbb{R}^d) \), it has

\[
\langle \varphi(x, k), f(x, k, t) \rangle = e^{-\gamma t} \langle \varphi(x(T - t), k), f_0(x, k) \rangle
\]

\[
- \int_T^0 \gamma_0 e^{-\gamma_0(t' - t)} \left( \langle \gamma_0^{-1} \Theta^{-1}_+[\varphi] - \gamma_0^{-1} \Theta^{-1}_-[\varphi], -\varphi \rangle(x(t' - t), k', t'), f(x, k, t') \right) \text{d}t',
\]

where the exponential distribution is introduced by adding \( \gamma_0 f \) on both sides of Eq. \( \text{Eq. (2.2)} \) and \( \langle x, k \rangle = x + \hbar k \tau / m \). One may expand \( \langle \varphi, f(t') \rangle \) and obtain an iterative integral related to a stochastic process. When \( f \) has a compact \( k \)-support, we have proven in [20, 21] that there exists a stochastic process \( X_t \) and constants \( C_1, C_2 > 0 \) such that

\[
\langle X_t, f_0 \rangle = \langle \varphi, f(t) \rangle, \quad \mathbb{E}|\langle X_t, f_0 \rangle - \langle \varphi, f(t) \rangle|^2 \leq C_1 \exp(C_2 t).
\]

In other words, the expectation of the stochastic model indeed solves the Wigner equation, whereas both the stochastic variance and particle number (taking \( \varphi = 1 \)) grow exponentially, thereby posing a formidable limitation to the particle method especially for long-time simulations. Such phenomenon is well known as the numerical sign problem for the negative particle method [22, 24, 28], stemming from the near-cancellation of positive and negative weights in sampling oscillatory functions. Because of a large variance of estimator, sample size must be large enough to obtain reliable results within a small relative uncertainty.

To alleviate the sign problem, we have suggested to replace the high-frequency component of \( \Psi DO \) by its principal asymptotic terms in Eq. \( \text{Eq. (2.14)} \), yielding another stochastic model with lower variance, termed WBRW-SPA [21]. That is, there exists a positive constant \( \alpha^* < 1 \) such that

\[
\mathbb{E}|\langle X_t, f_0 \rangle - \langle \varphi, f(t) \rangle|^2 \lesssim \exp(\alpha^* C_2 t),
\]

\[
\mathbb{E}|\langle X_t, f_0 \rangle - \langle \varphi, f(t) \rangle|^2 \leq C_1 \exp(C_2 t).
\]
which implies that SPA suppresses the exponential growth of both particle number and stochastic variances compared with all existing stochastic algorithms, at the cost of introducing a small asymptotic error term $O(\lambda_0^{-3/2})$. The implementation of WBRW-SPA for the Coulomb potential is illustrated in Algorithm 1, where single-body and two-body interactions are treated in a unified framework due to their strong resemblance, and more details can be referred to [21].

Algorithm 1 WBRW-SPA for two-body and single-body Coulomb systems

**Input parameters:** The initial time $t_l$ and final time $t_{l+1}$, the constant rate $\gamma_0$, the filter $\lambda_0$, $k$-domain $K$ and the upper band $r_{\text{max}} > 4|K|$.

**Sampling processes:** Suppose each particle in the branching particle system, carrying an initial weight $w$ either 1 or $-1$, starts at time $t_l$ at state $(x, k) = (x_c, x_p, k_e, k_p)$ for two-body system or $(x) = (x_c, k_e)$ for single-body system, and moves until $t_{l+1} = t_l + \Delta t$ according to the following rules.

1. **(Frozen)** Generate a random $\tau \sim \gamma_0 e^{-\gamma_0 t}$. For a particle at $(x, k)$ at instant $t \in [t_l, t_{l+1})$, if $t + \tau \geq t_{l+1}$, it becomes frozen at $(x + \frac{\hbar k(t_{l+1} - t)}{m}, k, t_{l+1})$.

2. **(Death)** If $\tau < \Delta t$, the particle is killed at shifted state $(x + \frac{\hbar k\tau}{m}, k, t + \tau)$.

3. **(Branching)** When the particle is killed at $(x + \frac{\hbar k\tau}{m}, k, t + \tau)$, it produces at most three offsprings at states $(x^{(1)}, k^{(1)}, t + \tau)$, $(x^{(2)}, k^{(2)}, t + \tau)$ and $(x^{(3)}, k^{(3)}, t + \tau)$. The third offspring is produced at state $(x^{(3)}, k^{(3)}) = (\bar{x}, \bar{k})$ with probability 1, carrying the weight $w$.
   - Two-body system: $\bar{x} = (\bar{x}_c, \bar{x}_p) = (x_c + \frac{\hbar k_{\gamma_0}}{m}, x_p + \frac{\hbar k_{\gamma_0}}{m})$ and $\bar{x}_p = x_{\text{A}}$.
   - Single-body system: $\bar{x} = x_c + \frac{\hbar k_{\gamma_0}}{m}$ and $\bar{k}_p = x_{\text{A}}$.

4. **(Scattering)** Generate a random number $r$ uniformly in $[0, r_{\text{max}}]$.
   - If $r < \lambda_0/|\bar{x} - \bar{x}_p|$, generate random $\theta$ uniformly in $[0, \pi]$ and $\phi$ uniformly in $[0, 2\pi]$, yielding a random vector $k' = (\cos \theta, \sin \theta \cos \phi, \sin \theta \sin \phi)$. Two offsprings are produced at states $(x^{(1)}, k^{(1)})$ and $(x^{(2)}, k^{(2)})$, $x^{(1)} = x^{(2)} = \bar{x}$, with probability $Pr(1)$, $Pr(2)$, endowed with weights $w_1$ and $w_2$, respectively.
     - **Probability:** $Pr(1) = Pr(2) = \frac{2\pi^3 r_{\text{max}}}{K_{\text{c}, 1, \gamma_0}} \left| \sin((\bar{x} - \bar{x}_p) \cdot k') \sin \theta \right|$.

   - **Random jump:**
     - Two-body system: $k^{(1)} = (k_e - \frac{k'}{2}, k_p + \frac{k'}{2})$, $k^{(2)} = (k_e + \frac{k'}{2}, k_p - \frac{k'}{2})$.
     - Single-body system: $k^{(1)} = k_e - \frac{k'}{2}$, $k^{(2)} = k_e + \frac{k'}{2}$.
     - **Update weight:** $w_i = (-1)^{i-1} w \cdot \frac{\sin((x^{(i)} - x_{\text{p}})) \sin \theta}{|\sin(x^{(i)} - x_{\text{p}}) \cdot k'| \sin \theta} \mathbb{1}_{\{k^{(i)} \in K\}}$, $i = 1, 2$.

   - If $r \geq \lambda_0/|\bar{x} - \bar{x}_p|$, two offsprings are produced with the probability $Pr(1)$, $Pr(2)$ at states $(x_1, k_1)$ and $(x_2, k_2)$, $x^{(1)} = x^{(2)} = \bar{x}$, $\sigma_+ = \sigma_+(\bar{x})$, endowed with updated weights $w_1$ and $w_2$, respectively.
     - **Probability:** $Pr(1) = Pr(2) = \frac{4\pi r_{\text{max}}}{K_{\text{c}, 1, \gamma_0}} \left| \sin((x_{\text{p}} - x_{\text{p}})) \right| \mathbb{1}_{\{k^{(i)} \in K\}}$, $i = 1, 2$.

   - **Random jump:**
     - Two-body system: $k^{(1)} = (k_c - \frac{r_{\gamma_0}}{2}, k_p + \frac{r_{\gamma_0}}{2})$, $k^{(2)} = (k_c + \frac{r_{\gamma_0}}{2}, k_p - \frac{r_{\gamma_0}}{2})$.
     - Single-body system: $k^{(1)} = k_c - \frac{r_{\gamma_0}}{2}$, $k^{(2)} = k_c + \frac{r_{\gamma_0}}{2}$.
     - **Update weight:** $w_i = (-1)^{i-1} w \cdot \frac{\sin((x_{\text{p}} - x_{\text{p}}))}{|\sin(x_{\text{p}} - x_{\text{p}}) \cdot k'|} \mathbb{1}_{\{k^{(i)} \in K\}}$, $i = 1, 2$.

5. **(Independence)** The offsprings continue to move independently.

**Termination condition:** All particles in the branching particle system are frozen.

Suppose the Wigner function is localized in $k$-space with a compact support $K$, so that the split Wigner kernels can be normalized. By taking average of many
realizations of WBRW-SPA, we obtain the particle estimator

\[
\langle \varphi(x, k), f(x, k, t) \rangle \sim \langle \varphi(x, k), \nu_t \rangle, \quad \nu_t = \frac{1}{N_0} \sum_{i=1}^{P(t)} \delta(x_i^+, k_i^+) - \frac{1}{N_0} \sum_{i=1}^{M(t)} \delta(x_i^-, k_i^-),
\]

where \( S^+ = \{(x_i^+, k_i^+)\}_{i=1}^{P(t)} \) and \( S^- = \{(x_i^-, k_i^-)\}_{i=1}^{M(t)} \) are positive and negative particles, carrying opposite particle weight \( \pm 1 \), respectively. The normalizing constant is the effective particle number \( N_0 = P(t) - M(t) \). Namely, the particle method seeks an empirical sign measure \( \nu_t \) to approximate the Wigner function in the weak sense. In particular, by taking \( \varphi(x, k) = \mathbb{1}_{X_\mu \times K_\nu}(x, k) \), where \( X_\mu \times K_\nu \) with volume \( |X_\mu||K_\nu| \) gives a uniform partition of \( \mathbb{R}^2 \) for \( 1 \leq \mu \leq N_x \) and \( 1 \leq \nu \leq N_k \), the reduced Wigner function \( W_1(x, k, t) \) can be reconstructed by a piecewise constant histogram,

\[
W_1(x, k, t) \approx \frac{1}{N_0} \sum_{\mu=1}^{N_x} \sum_{\nu=1}^{N_k} \frac{P(t)}{M(t)} \mathbb{1}_{X_\mu \times K_\nu}(x_i^+, k_i^+) - \mathbb{1}_{X_\mu \times K_\nu}(x_i^-, k_i^-) \frac{|X_\mu||K_\nu|}{|X_\mu||K_\nu|}.
\]

### 3.2. Demonstration of the numerical sign problem.

We provide an illustration of numerical sign problem by simulating the single-body 6-D Wigner equation (2.11). The motivation comes from the quantum optics as the coherent state is usually described by a Gaussian wavepacket.

**Example 1.** Suppose the initial electron Wigner function \( f_e(x, k, 0) \) is

\[
f_e(x, k, 0) = \pi^{-3} e^{-\frac{1}{2}((x_1-1)^2+x_2^2+x_3^2)\theta^{-2}(k_1^2+k_2^2+k_3^2)},
\]

interacting with a proton fixed at \( x_A = (0, 0, 0) \) under attractive Coulomb potential.

![Figure 1](image.png)

Fig. 1. **Numerical sign problem in stochastic Wigner simulations:** Variances grow in time due to the accumulation of negative particle weights. WBRW-SPA is able to suppress the growth of errors as it properly accounts for the decay of \( \Psi\text{DO} \) for large \( k \).

In order to measure the empirical variances of the particle simulations, we try to calculate the \( L^2 \)-error \( \mathcal{E}_2[W_1](t) \) (see Eq. (5.1)) between the histogram (3.5) of
$W_1(x, k, t)$ and the deterministic solutions produced by a characteristic-spectral-mixed scheme [13], as well as the deviation in total Hamiltonian $\mathcal{E}_H(t)$ (see Eq. (5.2)), with details put in Section 5. Aside from them, we also record the growth ratio of particle number, that is, the total particle number $N(t) = P(t) + M(t)$ divided by $N_0$.

In Figure 1, we first make a comparison between the Monte Carlo (MC) simulations with SPA ($\lambda_0 = 4$) and without SPA (direct splitting). The exponential growth of stochastic variance is clearly observed regardless of sample size $N_0$, and the convergence rate for the particle splitting deviates significantly from the theoretical order $N_0^{-1/2}$ as decay property of $\Psi_{DO}$ is ignored. By contrast, for WBRW-SPA, the exponential growth rate of variances can be suppressed to the large extent, and the convergence rate becomes closer to the theoretical order since it properly accounts for the decay for large wave vector $k$. A comparison of $W_1(x, k, t)$ is given in Figure 1(d).

The reduced Wigner function produced by direct particle splitting is very noisy, while the noises are significantly suppressed by SPA.

Fig. 2. Influence of $\lambda_0$ in WBRW-SPA: The first row is $N_0 = 1 \times 10^7$ and the second is $N_0 = 4 \times 10^7$. The filter $\lambda_0 = 4$ strikes the best balance in accuracy and efficiency. Too small $\lambda_0$ underestimate the contribution from the low-frequency part, while too large $\lambda_0$ fails to control the variances efficiently. This also coincides with our early observations made in [21].

1. From Figures 2(a), 2(b), 2(d) and 2(e), $\lambda_0 = 4$ strikes the best balance in accuracy and particle growth. Too small $\lambda_0$ underestimates the contribution from the low-frequency part, while too large $\lambda_0$ fails to control the variances efficiently. This also coincides with our early observations made in [21].

2. As shown in Figures 2(b) and 2(e), the particle grows two thousandfoldly up to $t = 4$ when the direct particle splitting is adopted (growth ratio is about $e^{1.910^7}$), thereby providing another evidence of the sign problem. Fortunately, it can be dramatically suppressed by using SPA to $\Psi_{DO}$.

A natural question arise: How does the filter $\lambda_0$ influence the numerical accuracy and growth of particle number? In Figure 2, we make a comparison of WBRW-SPA under $\lambda_0 = 2, 4, 8$. 

(a) MC errors, $N_0 = 1 \times 10^7$.  (b) Growth of particle number.  (c) Deviation in Hamiltonian.

(d) MC errors, $N_0 = 4 \times 10^7$.  (e) Growth of particle number.  (f) Deviation in Hamiltonian.
(3) For an energy-conserving system, the best choice of $\lambda_0$ can be determined by monitoring the deviation of total energy. It is observed in Figures 2(c) and 2(f) that WBRW-SPA under $\lambda_0 = 4$ achieves the smallest deviation. Based on these observations, we will adopt $\lambda_0 = 4$ for subsequent simulations.

4. Particle annihilation: A remedy for the numerical sign problem. Unfortunately, the numerical sign problem cannot be completely surmounted by SPA as demonstrated by Eq. (3.3) and Figure 1, because it is rooted in the oscillatory nature of the low-frequency component of $\Psi_{DO}$. Moreover, it is more probably to be aggravated as the dimensionality (system size) increases due to the enrichment of fine structures, such as alternating local maxima and minima in phase space.

To further alleviate the sign problem, particle annihilation (PA) turns out to be indispensable. For a given empirical signed measure of the form (3.4) (the dependence on time is omitted), PA intends to remove $N_A$ positive particles from $S^+$ and $N_A$ negative ones from $S^-$, and obtains another empirical signed measure $\tilde{\nu}$,

$$\tilde{\nu} = \frac{1}{N_0} \sum_{i=1}^{P-N_A} \delta(\tilde{x}_i^+, \tilde{k}_i^+) - \frac{1}{N_0} \sum_{i=1}^{M-N_A} \delta(\tilde{x}_i^-, \tilde{k}_i^-),$$

where $(\tilde{x}_i^+, \tilde{k}_i^+)$ can be either chosen as a subset of $S^+$, or be generated by certain operations of particles in $S^+$ like bootstrap filtering. The target of PA is to control the error function $E(\varphi) = |\langle \varphi, \nu \rangle - \langle \varphi, \tilde{\nu} \rangle|$ for suitable test functions $\varphi$. It is expected to annihilate two kinds of particles carrying opposite weights and cancelling out their contributions within a reasonable numerical accuracy. For this reason, PA is also named particle cancellation or particle resampling [28].

The prototype PAUM [12, 3, 6] borrows the idea from the histogram statistics [29], that is, using a uniform grid to divide particles into several clusters and annihilating the particles in the same bin. In spite of its simplicity and easy implementation, the cancellation might be very inefficient in high dimension since a large amount of particles are left uncanceled [28]. SPADE, our adaptive PA method, intends to get rid of the severe limitation of regular mesh and to overcome CoD in PAUM [26]. The intuition, design and implementation of SPADE are detailed below, and a thorough numerical comparison between PAUM and SPADE is left in Section 5, as well as in Appendix A.

4.1. PAUM: Particle annihilation via uniform mesh. A more general setting is considered with dimensionality denoted by $D = 2s$. The computational domain is a rectangular bin $\Omega = \prod_{m=1}^{s}[x_{\min}^{(i)}, x_{\max}^{(i)}] \times \prod_{m=1}^{s}[k_{\min}^{(i)}, k_{\max}^{(i)}]$ and our target is to annihilate the positive particles $S^+$ and negative particles $S^-$ located in $\Omega$.

A straightforward idea is to utilize a uniform mesh for dividing $\Omega$: $\Omega = \bigcup_{k=1}^{K} Q_k$, where $Q_k = X_{i_1} \times \cdots \times X_{i_s} \times K_{j_1} \times \cdots \times K_{j_l}$ is the tensor product of disjoint rectangular bins, with $X_{i_l} = [x_{\min}^{(l)} + (i_l - 1)\Delta x_l, x_{\max}^{(l)} + i_l \Delta x_l]$, $K_{j_l} = [k_{\min}^{(l)} + (j_l - 1)\Delta k_l, k_{\max}^{(l)} + j_l \Delta k_l]$, $l = 1, \ldots, s$, and $\Delta x_l$ and $\Delta k_l$ are spatial and momental spacings in the $l$-th dimension, respectively. Essentially, the uniform partition divides particles into $K$ clusters, namely, $S^+ = \bigcup_{k=1}^{K} S_k^+$, $S^- = \bigcup_{k=1}^{K} S_k^-$, and the particles in the same bin are assumed to almost equally contribute to an integration like (3.4).

More precisely, PAUM suggests to use a piecewise constant function $p(x, k)$ to approximate the Wigner function $f(x, k) \approx p(x, k) = \sum_{k=1}^{K} \frac{P_k - M_k}{N_0} \cdot \frac{1_Q_k(x, k)}{\text{vol}(Q_k)}$, where $K$ is total partition level, $P_k$ and $M_k$ count the positive and negative particles in $Q_k$, respectively, $1_Q_k$ denotes the indicator function and vol($Q_k$) is the Lebesgue measure.
of $Q_k$. Thus the particles carrying opposite signs are eliminated directly, and clearly the errors scale as $O(\prod_{i=1}^{\int} \Delta x_i \Delta b_i)$. For more details, the readers can refer to [3, 18].

It should be emphasized that PAUM requires a very subtle balance between partition level $K$ and the effective sample size $N_0$. When $K \gg N_0$, particles are divided into too many clusters and only a few are canceled [28]. The sharp deterioration in the efficiency of PAUM seems to be inevitable when the dimensionality increases, which is well known as the overfitting problem in statistics when dimensionality of feature space largely exceeds sample size [29]. In Section 5, we endeavor to test PAUM with a $61^3 \times 60^3$ uniform grid with $K = 4.9028 \times 10^{10}$ and find that $N_0$ should be at least $1 \times 10^9$, otherwise annihilation might be very inefficient. Actually, the huge uniform grid has to be distributed evenly in multiple nodes and is unfriendly to load balance, especially when there is no symmetry inside the physical problems.

### 4.2. SPADE: An adaptive particle annihilation.

Intuitively speaking, the key of SPADE to alleviate CoD is replacing the uniform mesh with an adaptive one via a divide-and-conquer strategy [26], partially borrowing the idea from the discrepancy-based density estimation [30]. First, it seeks an adaptive partition of $\Omega$ via the sequential binary splitting and controls the number-theoretic discrepancies of points in each group, so that the particles located in the same bin have almost equal contribution to the estimator (3.4). Once an adaptive partition $\Omega = \bigcup_{k=1}^{K} Q_k$ is obtained, it divides positive and negative particles into $K$ groups. Second, it seeks a random matching between the positive and negative particles in the same group independently, and the annihilation can be realized by removing the matched pairs.

SPADE can be implemented via a recursive binary splitting. A binary partition $P$ on a domain $Q = \Omega$ is the collection of sub-rectangles whose union is $Q$. Starting with $P_1 = \{Q\}$ at level 1, for $P_K = \{Q_1, \ldots, Q_K\}$ at level $K$, $P_{K+1}$ is produced by dividing one of the regions in $P_K$ along one of its coordinates and merging both sub-rectangles with the rest of regions in $P_K$. Two key points must be specified for a binary partition. One is whether to split and the other is where to split.

**Whether to split:** For the stopping criterion, we try to control the irregularity of points distribution in each bin, measured by the star discrepancy. The definition of the star discrepancy for a sequence $(x_1, \ldots, x_P) \subset [0,1]^{2s \times P'}$ reads that

$$D_P^s(x_1, \ldots, x_P) = \sup_{u \in [0,1]^{2s}} \left\{ \frac{1}{P} \sum_{i=1}^{P} \left| \int_{0}^{u} (x_i) - \text{vol}(0,u) \right| \right\}. \tag{4.2}$$

For the sequences $S_k^+$ and $S_k^-$ in $Q_k = [a_1^{(k)}, b_1^{(k)}] \times \cdots \times [a_{2s}^{(k)}, b_{2s}^{(k)}]$, the discrepancy can be defined by a linear scaling $t_k : S_k^+ \rightarrow [0,1]^{2s^*}$,

$$D_P^s(S_k^+) = D_P^s(t_j(x_1^{k,+}, k_1^{k,+}), \ldots, t_j(x_P^{k,+}, k_P^{k,+})), \tag{4.3}$$

$$D_M^s(S_k^-) = D_M^s(t_j(x_1^{k,-}, k_1^{k,-}), \ldots, t_j(x_M^{k,-}, k_M^{k,-})), \tag{4.4}$$

with $(x_1^{k,+}, k_1^{k,+}), \ldots, (x_P^{k,+}, k_P^{k,+})$ and $(x_1^{k,-}, k_1^{k,-}), \ldots, (x_M^{k,-}, k_M^{k,-})$ the positive and negative particles located in $Q_k$, respectively, and

$$t_k(x, k) = \begin{pmatrix} x_1 - a_1^{(k)} & x_2 - a_2^{(k)} & \cdots & x_s - a_s^{(k)} & k_1 - a_1^{(k)} & k_2 - a_2^{(k)} & \cdots & k_s - a_s^{(k)} \\ b_1^{(k)} - a_1^{(k)} & b_2^{(k)} - a_2^{(k)} & \cdots & b_s^{(k)} - a_s^{(k)} & b_1^{(k)} - a_1^{(k)} & b_2^{(k)} - a_2^{(k)} & \cdots & b_s^{(k)} - a_s^{(k)} \end{pmatrix}. \tag{4.5}$$

Then $k$-th bin $Q_k$ continues to be split until both discrepancy bounds are satisfied,

$$D_P^s(S_k^+) \leq \frac{\sqrt{N_0}}{\max(P_k, M_k)}, \quad D_M^s(S_k^-) \leq \frac{\sqrt{N_0}}{\max(P_k, M_k)}. \tag{4.6}$$


Fig. 3. The adaptive clustering via binary splitting and a decision tree. The binary partition of $Q$ is $P_0 = (Q_1, \ldots, Q_6)$, and each bin ceases to be split when both positive and negative particles satisfy the discrepancy bounds (4.6) simultaneously; otherwise it will be split further into two parts. In the mean time, the particles are divided into 6 groups.

where the sole parameter $\vartheta$ adjusts the depth of partition. The motivation of the bounds (4.6) is to control the contribution of positive and negative particles in the same bin to an integral (3.4). For the convergence issue of SPADE, one can refer to [26], where the Koksma-Hlawka inequality and concentration inequalities for sampling with or without replacement are exploited.

Where to split: For $Q_k = [a_1^{(k)}, b_1^{(k)}] \times \cdots \times [a_2^{(k)}, b_2^{(k)}]$, we select a node $c_j^{(k)}$ in the $j$-th dimension and split $Q_k$ into $Q_k^{(1)}$ and $Q_k^{(2)}$:

$$ Q_k^{(1)} = \prod_{i=1}^{j-1} [a_i^{(k)}, b_i^{(k)}] \times [a_j^{(k)}, c_j^{(k)}] \times \prod_{i=j+1}^{2^n} [a_i^{(k)}, b_i^{(k)}], \quad Q_k^{(2)} = Q_k \setminus Q_k^{(1)}. $$

Denote by $P_k^{(1)}$ and $M_k^{(1)}$ the counts of positive and negative particles in $Q_k^{(1)}$, respectively. It suggests to choose $c_j^{(k)}$ to optimize either the maximal gap [30]

$$ \max_{Q_k^{(1)}} \left( \frac{P_k^{(1)}}{P_k} - \frac{\text{vol}(Q_k^{(1)})}{\text{vol}(Q_k)} \right) \left| \frac{M_k^{(1)}}{M_k} - \frac{\text{vol}(Q_k^{(1)})}{\text{vol}(Q_k)} \right|, $$

or the difference gap [26], which is bounded by Eq. (4.8),

$$ \frac{1}{2} \max_{Q_k^{(1)}} \left( \frac{P_k^{(1)}}{P_k} - \frac{M_k^{(1)}}{M_k} \right) = \frac{1}{2} \max_{Q_k^{(2)}} \left( \frac{P_k^{(2)}}{P_k} - \frac{M_k^{(2)}}{M_k} \right). $$

The physical intuition behind is to dig out the nodal surfaces that divide positive and negative particles. When $P_k^{(1)}/P_k$ is much larger than $M_k^{(1)}/M_k$, positive particles
are concentrated in $Q^{(1)}$. At the same time, $P^{(2)}/P_k$ shall be smaller than $M^{(2)}/M_k$ so that negative particles are concentrated in $Q^{(2)}$. According to our experience, the maximal gap works better for $D \leq 6$, while the difference gap performs better for $D > 6$. Details are put in Appendixes B.4 and C.3.

A practical way to obtain a (sub)-optimal $c^{(k)}(j)$ is to pick up the $j$-th dimension and $m$ equidistant points $c^{(k)}(j) = a^{(k)}_j + \frac{l}{m}(b^{(k)}_j - a^{(k)}_j)$, $l = 1, \ldots, m - 1$, in $[a^{(k)}_j, b^{(k)}_j]$ to maximize the gap functions. For sufficiently large $m$, it may well approximate to the true gap. In practice, it suffices to take $m = 4$ for $D = 6$, and $m = 32$ for $D = 12$.

As a remark, the key ingredient of SPADE is to calculate of the star discrepancy, which is a NP-hard problem and essentially difficult to solve exactly. In a sense, SPADE tries to convert the NP-hard sign problem into another NP-hard combinatorial problem. Fortunately, the star discrepancy can be approximated by some heuristic algorithms, such as the improved version of threshold accepting algorithm (TA-improved), which gives an efficient approximation in moderately large dimension ($D \leq 60$) [31]. According to our tests, running TA-improved algorithm once can produce reliable approximation for a 6-D sequence under the iteration times $I = 64$, and $I = 128$ for a 12-D sequence (see Appendix D).

4.3. Demonstration of SPADE. An illustrative example is given to demonstrate our motivation behind SPADE.

Example 2. Suppose one draws samples according to a determinantal function,

\[(4.10) \quad \psi(x_1, x_2) = \det \begin{pmatrix} \psi_-(x_1) & \psi_+(x_1) \\ \psi_-(x_2) & \psi_+(x_2) \end{pmatrix} = \psi_-(x_1)\psi_+(x_2) - \psi_+(x_1)\psi_-(x_2),\]

where $\psi_\pm(x) = \left(2\pi\right)^{-\frac{1}{2}} e^{-\frac{1}{2}(x \pm 1)^2}$. One can draw samples from two Gaussian functions and take the minus sign as the particle weight (here we set $N_0 = 2000$). As seen in Figure 4(a), the overlap of two Gaussians almost cancels out. Thus for 2000 positive particles (red dot) and 2000 negative particles (blue circle) in Figure 4(b), we also want to cancel out the particles in the central region carrying opposite weights. Figures 4(c) and 4(e) plot the adaptive partitions under $\vartheta = 0.4$ and $\vartheta = 0.1$, respectively. Choosing a smaller $\vartheta$ leads to a refinement of partition, and consequently makes more particles left uncanceled (see Figures 4(d) and 4(f)).

It is clear that the partition is refined in the region where samples are concentrated and ceases to be split further when points are sparsely distributed. Such idea seems to be straightforwardly generalized to high dimensional problem, without a priori knowledge of underlying density or sparse structure. The sole parameter $\vartheta$ in the discrepancy bounds controls the partition level, as well as the accuracy of SPADE.

4.4. Particle annihilation in outer cells for mass conservation. For dynamical problems, particles that move outside the computational domain may result in loss of total mass, which is inconsistent with the conservative Hamiltonian system. To fix it, we need to design particle annihilation in the outer bins. A simple strategy is adopted here. An outer pointset $S_{\text{out}}$ is used to store the particles outside the computational domain $\Omega$. Once the particles in $S_{\text{out}}$ reach its maximal size, we make random matching among positive and negative particles and directly remove the redundant particles carrying opposite weights in pair. As the positive and negative particles are also generated in pair, the total mass can be rigorously conserved in the simulations (see Figure 6 below).
4.5. Implementations. SPADE has a one-to-one correspondence to a decision tree (see Figure 3), so that it requires to store all particles in a \((P+M) \times D\) real-valued matrix and an adaptive partition in a \(K \times (2D+2)\) real-valued matrix (including upper and lower bounds of bins and counts of particles), while PAUM requires to store two tensors in D-dimension, scaling as \(L^D\) with \(L\) the resolution in each direction. Since \(K\) is usually smaller than \(P+M\), the data storage in SPADE scales at most \((P+M) \times (3D + 2)\), which is significantly smaller than \(L^D\) as \(D\) increases.

Domain decomposition \(\Omega = \bigcup_{p=1}^{N_p} \Omega_p\) is a pretreatment for distributed-memory implementation. By dividing a tree into a forest composed of \(N_p\) independent trees, the adaptive partitions can be established independently in \(N_p\) processors. This also split particles into \(N_p\) batches, and \(S_{\text{out}} = \bigcup_{p=1}^{N_p} S_{\text{out}}^{(p)}\) with \(S_{\text{out}}^{(p)}\) particles in \(\Omega_p\) manipulated by the \(p\)-th processor. A relevant point is to strike a balance in overload. To this end, one shall keep the particle number in each \(\Omega_p\) more or less the same.
5. Particle simulations of 6-D Wigner-Coulomb dynamics. From this section, we are about to perform a series of benchmarks on simulating 6-D Wigner-Coulomb dynamics (see Example 1), which is a typical non-equilibrium quantum dynamics and the prototype for many realistic simulations [8, 9], while additional benchmarks on 4-D problems are given in Appendix B. Our performance evaluation is two-pronged: First, we make a thorough comparison between PAUM and SPADE (i.e., WBRW-SPA-PAUM v.s. WBRW-SPA-SPADE). Second, we investigate how the parameter \( \vartheta \) in SPADE and sample size \( N_0 \) influence the accuracy, energy conservation, particle number and the partition level \( K \). The latter is towards a comprehensive understanding of SPADE and a guiding principle for improving its accuracy systematically, and is pivotal to rigorous numerical analysis.

| Notation | What the notation stands for | Relation with other quantities |
|----------|-------------------------------|-------------------------------|
| \( N_0 \) | Initial effective sample size | \( N_0 \uparrow \Rightarrow \) MC errors \( \downarrow \) and \( K \uparrow \) |
| \( P(t) \) | Number of positive particles at time \( t \) | \( P(t) - M(t) = N_0 \) |
| \( M(t) \) | Number of negative particles at time \( t \) | \( P(t) - M(t) = N_0 \) |
| \( N(t) \) | Total particle number at time \( t \) | \( P(t) + M(t) = N(t) \) |
| \( K \) | Total partition level | \( K \uparrow \Rightarrow \) SPADE errors \( \downarrow \) |
| \( \vartheta \) | Parameter in discrepancy bounds | \( \vartheta \downarrow \Rightarrow \) partition is refined, \( K \uparrow \) |

Our main findings are summarized in Table 1. The depth \( K \) of partition, as well as the efficiency of SPADE, is controlled by the parameter \( \vartheta \). It is worth noting that too small \( \vartheta \) should NOT be suggested as it may lead to over-refinement of partition and make many particles uncanceled. This is known as the notorious overfitting problem [28, 29]. When the partition level largely exceeds the sample size, few particles are located in the same bin and thus cannot be annihilated efficiently. A direct consequence is the rapid growth of computational cost.

The reference solutions are produced by our recently developed characteristic-spectral-mixed scheme, where the Wigner function defined in a 6-D computational domain \([-10.8, 10.8]^3 \times [-4, 4]^3\) is expanded as the tensor product of 75³ cubic spline basis (with spacing \( \Delta x = 0.3 \)) and 80³ Fourier spectral basis (with spacing \( \Delta k = 0.1 \)) and integrated by the Lawson predictor-corrector scheme (with time step \( \Delta t = 0.025 \) a.u.) to ensure its accuracy [13]. It spends about 15 days to reach \( T = 15 \) a.u. under the platform: E5-2680 v4 (2.40 GHz, 14 Cores, 28 Threads) with 256GB Memory \( \times 16 \) (448 cores in total). To the best of our knowledge, this is also the first attempt to simulate the 6-D Wigner dynamics via the massively parallel deterministic solver.

For stochastic particle simulations, we adopt \( \gamma_0 = 50, \lambda_0 = 4 \) and a finite \( k \)-domain \([-3, 3]^3\) in Algorithm 1, and annihilate every particle every 1 a.u. The reduced Wigner function (2.4) and spatial marginal distribution (2.5), which can be readily obtained by histogram reconstruction (3.5) under a uniform grid mesh \([-9, 9] \times [-3, 3]\) with \( N_x = 61, N_k = 60 \) (with the same spacings \( \Delta x = 0.3, \Delta k = 0.1 \) as used in the deterministic simulations), allow both visualization of quantum Coulomb interaction and quantitative comparison with the deterministic counterparts. The performance metrics include the normalized \( l^2 \)-error \( E_2[W_1](t) \) to monitor the stochastic variances,

\[
(5.1) \quad E_2[W_1](t) = \left[ \frac{1}{N_x N_k} \sum_{i=1}^{N_x} \sum_{j=1}^{N_k} (W_{1, \text{ref}}(x_i, k_j, t) - W_{1, \text{num}}(x_i, k_j, t))^2 \right]^{1/2},
\]
where \( W_{1}^{\text{ref}} \) and \( W_{1}^{\text{num}} \) denote the reference and stochastic solution for \( W_{1} \), respectively, and the deviation of total Hamiltonian \( \mathcal{E}_{H}(t) \),

\[
(5.2) \quad \mathcal{E}_{H}(t) = |H(t) - H(0)|, \quad H(t) = \int_{\mathbb{R}^{3N}} \left( \frac{\hbar^{2} |k|^{2}}{2m} + V(x) \right) f(x, k, t) dx dk.
\]

In addition, the growth ratio of total particle \( \mathcal{N}(t)/N_{0} = (P(t) + M(t))/N_{0} \) is closely related to the computational complexity, while \( P(t) - M(t) \) is always conserved.

All simulations of WBRW-SPA-SPADE via our own Fortran implementations run on the High-Performance Computing Platform of Peking University with the platform: 2*Intel Xeon E5-2697A-v4 (2.60GHz, 40MB Cache, 9.6GT/s QPI Speed, 16 Cores, 32 Threads) with 256GB Memory \times 4. Each task uses 128 cores and is implemented via the Message Passing Interface (MPI). As a comparison, we also run WBRW-SPA-PAUM with \( N_{0} = 1 \times 10^{9} \) and a 61^3 \times 60^3 uniform grid mesh \( (K \approx 4.9 \times 10^{10} \text{ is fixed}) \). It spent 93.8 hours to reach \( T = 15 \text{a.u.} \) under 125 nodes \times 16 threads (2000 cores in total) via a mixture of MPI and OpenMP library. Such comparison is somehow unfair because the particle number in PAUM is much larger than those adopted in SPADE. But this is really our intention, as we would like to show that SPADE is able to control the growth of both stochastic variances and particle number much more efficiently than PAUM, especially when the sample size is not very large.

### 5.1. Comparison between PAUM and SPADE

Several groups of simulations are performed under \( N_{0} = 4 \times 10^{6}, 1 \times 10^{7}, 4 \times 10^{7} \) and \( \vartheta = 0.003 \) to 0.08. The stochastic errors are presented in Figure 5 by monitoring \( \mathcal{E}_{2}[W_{1}](t) \) and \( \mathcal{E}_{H}(t) \). The growth of particle number is plotted in Figure 6. Visualizations of reduced Wigner function \( W_{1}(x, k, t) \) and the spatial marginal distribution \( P_{1}(x, t) \) are provided in Figures 7 and 8, respectively. Based on them, we make the following observations.

**Stochastic variances:** In Figure 5, we set \( \mathcal{E}_{2}[W_{1}](t) \) under PAUM as the baseline (black solid line). Indeed, the variances can be suppressed by either increasing the sample size \( N_{0} \) or choosing smaller \( \vartheta \). An inspiring finding is that \( \mathcal{E}_{2}[W_{1}](t) \) under SPADE with \( \vartheta = 0.003 \) or 0.005 always outperform those under PAUM, even with \( N_{0} = 1 \times 10^{9} \gg 4 \times 10^{6} \). This actually manifests the huge advantage of adaptive partitioning over uniform one, especially with a relatively small sample size.

**Numerical energy:** The deviations of total energy are observed in all particle simulations since the cancelation of positive and negative particles may bring in some shifts in both kinetic energy and potential energy. Fortunately, it can be considerably ameliorated when \( N_{0} \) increases or the partition level \( K \) increases. This shall provide another evidence on the convergence of SPADE.

**Snapshots:** Particle simulations with either PAUM or SPADE can properly capture several important features of wavepackets, including the negative valley that manifest the uncertainty principle and the double-peak structure of wavepacket induced by the Coulomb collisions. Since \( W_{1}(x, k, t) \) and \( P_{1}(x, t) \) are reconstructed by a piecewise constant function (3.5) with spacings \( \Delta x = 0.3, \Delta k = 0.1 \), it may smooth the wavepackets and lead to some discrepancies in their crests (see Figure 8(b)). Nonetheless, the difference can be gradually compensated by either increasing the sample size or refining the adaptive partition, as clearly observed in Figure 8.

**Growth of particle number:** PAUM can work only when \( N_{0} \) is comparable to \( K \), otherwise many particles are still left uncanceled. From Figure 6, the particle number after PAUM grows to \( 8 \times 10^{8} \) at \( t = 15 \text{a.u.} \) \( (K \approx 4.9 \times 10^{10}) \). By contrast, SPADE is capable to annihilate particles more efficiently and keeps the total particle after annihilation at a stable level, thereby maintaining the computational complexity.
Fig. 5. A comparison of the $l^2$-errors of $E_2[W_1](t)$ to monitor the stochastic variances (left) and the deviation of total energy $E_H(t)$ (right). SPADE outperforms PAUM when $\vartheta = 0.003$ or $\vartheta = 0.005$. The group under $N_0 = 4 \times 10^6$, $\vartheta = 0.003$ is only simulated up to 7 a.u. because particle number grows too fast, leading to a rapid increase in complexity.

Fig. 6. Growth of total particle number: The effective particle number $P(t) - M(t)$ is always conserved. SPADE is superior to PAUM in efficiency. However, when $\vartheta$ is too small, it may lead to an over-refinement of partitioning and many particles are uncanceled.
Fig. 7. Snapshots of the reduced Wigner function \( W_1(x, k, t) \) by the deterministic scheme (left), WBRW-SPA-PAUM with \( N_0 = 10^9 \) (middle) and WBRW-SPA-SPADE with \( N_0 = 4 \times 10^7 \), \( \vartheta = 0.003 \) (right). The particle-based stochastic algorithms can properly capture the double-peak structure (Coulomb collision) and negative valley (uncertainty principle).
Overfitting: An exception is the group $N_0 = 4 \times 10^6$, $\vartheta = 0.003$ in Figure 6(a) as the particle number grows even faster than PAUM and the growth ratio reaches 48 at 6a.u., resulting in a rapid increase in computational time (see Table 2). Such phenomenon is called overfitting and may augment the complexity of SPADE. Despite the accuracy of the group $\vartheta = 0.003$ and $N_0 = 4 \times 10^6$, even outperforms that under PAUM with $N_0 = 1 \times 10^9$ (see Figure 5(a)), it is still NOT recommended to use too small $\vartheta$ for relatively small $N_0$ in consideration of efficiency.

5.2. Deep partition improves SPADE. The accuracy of SPADE can be systematically improved by deepening the partition. In Figure 10, we would like to emphasize that the bottom line structure of the total particle, or equivalently, the particle number after annihilation may be another indicator for measuring the accuracy and efficiency of SPADE. The physical motivation is that the Wigner function is bounded below and above and thus allows finite negative values [34], so that the size of signed particles reflecting non-classicality should be limited within a stable level.

Convergence with respect to the parameter $\vartheta$: In Figure 9, stochastic variances are compared when the parameter $\vartheta$ is fixed. It is observed that the accuracy of SPADE is almost comparable to that of PAUM when $\vartheta = 0.01$, and even outperforms PAUM when $\vartheta = 0.003$ or 0.005, regardless of $N_0$. Besides, SPADE largely alleviates the exponential growth of stochastic variances compared to pure Monte Carlo simulations. This again manifests the indispensability of adaptive particle annihilation.

Bottom line for particle number: The efficiency of SPADE can be characterized by its capability to maintain the accuracy, as well as the non-classicality, with minimal amount of particles. As shown in Figures 10(b) and 10(d), we have found that the particle number, although growing exponentially at each step, always returns
Fig. 9. The time evolution of the $l^2$-errors $E_2[W_1](t)$ with respect to sample size $N_0$, under different parameter $\vartheta$ (the depth of adaptive partition). SPADE significantly alleviates the rapid growth of stochastic variances, and outperforms PAUM when $\vartheta = 0.003$ or $\vartheta = 0.005$.

Fig. 10. Growth ratio of particle number before and after annihilation: A bottom line structure is observed for SPADE, say, the particle number after annihilation stays at a stable level. SPADE learns the minimal amount of particles that can accurately capture the non-classicality of the Wigner function under the prescribed sample size $N_0$. An exception is the group $N_0 = 4 \times 10^6, \vartheta = 0.003$, which is bothered by the overfitting problem.
back to a stable level after annihilation. This is called the bottom line structure as pointed out in [19]. Definitely, the bottom line should be larger than $N_0$ to properly account for the negative part of the Wigner function.

In Figures 10(a) and 10(c), the bottom line remains stable when $K$ is less or comparable to $N_0$. But it dramatically ascends when $K$ is much larger than $N_0$ (see $\vartheta = 0.003$ in Figure 10(a)). As more and more particles are generated without being canceled out, the computational complexity grows rapidly (see Table 2), thereby explaining the overfitting problem. Based on Figure 10(d) and Table 2, it is postulated that when the partition level $K$ is slightly smaller than $N_0$, SPADE attains a bottom line describing the minimal amount of particles that can accurately capture the non-classicality of the Wigner function under the prescribed sample size $N_0$, and it is able to strike the balance between efficiency and accuracy.

**Computational time:** The running time for SPADE up to 15a.u., which occupies more than 95% of total wall time, is recorded in Table 2. For 6-D problems, the advantage of particle-based stochastic simulations over grid-based deterministic counterparts becomes prominent. For $\vartheta \geq 0.01$, the time scales almost linearly on the sample size $N_0$, but it still grows rapidly when the overfitting problem happens.

**Table 2**

| $\vartheta$ | $N_0 = 4 \times 10^6$ | $N_0 = 1 \times 10^7$ | $N_0 = 4 \times 10^7$ |
|------------|---------------------|---------------------|---------------------|
|            | Time(h)  | Average $K$     | Time(h)  | Average $K$     | Time(h)  | Average $K$     |
| 0.003      | 70.5694  | (2.771$\times 10^7$) | 101.2576 | 1.772$\times 10^7$ | 95.7878 | 8.768$\times 10^6$ |
| 0.005      | 17.9169  | 5.408$\times 10^6$ | 17.7604  | 3.279$\times 10^6$ | 68.7162 | 3.288$\times 10^6$ |
| 0.01       | 3.2836   | 7.840$\times 10^5$ | 7.1684   | 8.574$\times 10^5$ | 31.8121 | 1.134$\times 10^5$ |
| 0.02       | 1.7982   | 2.503$\times 10^5$ | 4.6385   | 3.124$\times 10^5$ | 22.0516 | 4.859$\times 10^5$ |
| 0.04       | 1.2724   | 1.040$\times 10^5$ | 3.0352   | 1.410$\times 10^5$ | 17.081  | 2.362$\times 10^5$ |
| 0.08       | 0.9984   | 5.019$\times 10^4$ | 2.5116   | 7.085$\times 10^4$ | 13.203  | 1.232$\times 10^5$ |

(a) Smaller $\vartheta$ leads to larger $K$.

(b) Larger $K$ improves accuracy of SPADE.

**Fig. 11.** The relations among the average partition level $K$, the parameter $\vartheta$ and the stochastic variances measured by $E_2[W_1]$. The usage of small $\vartheta$ leads to the growth of partition level $K$, and consequently improves the accuracy of SPADE systematically.

In order to dig out the reasons behind, we plot $K$-$\vartheta$ curve in Figure 11(a) and investigate how the average partition level $K$ influences the growth of stochastic variances in Figure 11(b). The trend $\vartheta \downarrow \rightarrow K \uparrow \rightarrow E_2[W_1] \downarrow$ is clearly observed, implying
deepening the adaptive partition gradually improves the accuracy of SPADE. The price to pay is a sudden increase in complexity when $\vartheta$ becomes too small, as $K$ may grow exponentially and more particles are preserved (see Figure 6).

5.3. Large sample size improves SPADE. The above numerical results have already verified the potential of SPADE in alleviating the numerical sign problem. But the overfitting problem seems to be another bottleneck and might lead to a deterioration in efficiency. Fortunately, we would like to provide some evidences that the overfitting problem can be ameliorated when the effective sample size increases.

The convergence with respect to $N_0$: Numerical results validate the convergence of SPADE with respect to sample size $N_0$, although the convergence rate with respect to $N_0$ is much slower than $1/2$ due to the mixture of various error sources. As presented in Figure 9, when $\vartheta \geq 0.01$, the accuracy of SPADE can be improved by increasing $N_0$, while the improvement becomes less evident for smaller $\vartheta$.

Large sample size alleviates overfitting: SPADE may get rid of the overfitting problem under larger sample size. According to Figures 6 and 10, the particle number always remains at a stable level under $N_0 = 4 \times 10^7$ regardless of $\vartheta$, indicating that the over-partition is avoided as $N_0$ increases. From Table 2, we also see that the cost of the group $\vartheta = 0.003, N_0 = 4 \times 10^7$ is even smaller than that under $\vartheta = 0.003, N_0 = 1 \times 10^7$. These results uncover the opportunity to further improve SPADE by adopting small $\vartheta$, without a sacrifice in efficiency.

6. Particle simulations of 12-D Wigner quantum dynamics. The readers may be curious about whether SPADE is applicable in higher dimensional problem. Here we would like to demonstrate the potential of SPADE for $D = 12$ by solving the proton-electron Wigner equation in 12-D phase space.

Example 3. Consider a system composed of one proton and one electron interacting under the Coulomb potential, where the proton is strongly localized in $x$-space. The initial Wigner function is uncorrelated, say, $f_e(x_e, k_e, 0)f_p(x_p, k_p, 0)$ with

\[
f_e(x_e, k_e, 0) = \pi^{-3}e^{-\frac{1}{2}(x_{e,1}^2 + x_{e,2}^2 + x_{e,3}^2)}e^{-2(k_{e,1}^2 + k_{e,2}^2 + k_{e,3}^2)},
\]

\[
f_p(x_p, k_p, 0) = \pi^{-3}e^{-1002(x_{p,1}^2 + x_{p,2}^2 + x_{p,3}^2)}e^{-\frac{2}{1002}(k_{p,1}^2 + k_{p,2}^2 + k_{p,3}^2)},
\]

derived by the localized Hydrogen wavefunction in studying quantum tunneling [32].

In general, there is no exact solution to the non-equilibrium dynamics of proton-electron correlation [32, 34]. Fortunately, under conditions (2.8) and (2.9), it shall omit small terms $O(hn^{-1}e^{3/2})$ in the adiabatic approximation (2.10) for $t = O(1)$ as $m_p \approx 1836m_e$ and $\varepsilon = 2/1002 \ll 1$. This allows us to make a quantitative comparison between the projection of $f(x_e, x_p, k_e, k_p, t)$ onto $(x_e-k_e)$-plane and the single-body electron Wigner function, and consequently laying the foundation for our benchmarks on simulating 12-D Wigner dynamics.

For the sake of comparison, we still adopt $\gamma_0 = 50, \lambda_0 = 4$ in SPA, a finite $k$-domain $[-3, 3]^3 \times [-240, 240]^3$ and annihilate particles every 1 a.u. Each task uses 256 cores via MPI. The reduced Wigner functions $W_I(x, k, t)$ are obtained in $[-9, 9] \times [-3, 3]$ by histogram reconstruction (3.5) with the uniform meshes: $N_x = 100, N_k = 100$. Several groups under $N_0 = 1.6 \times 10^7, 4 \times 10^7, 1 \times 10^8$ and $\vartheta = 0.004, 0.005$ are performed. The $L^2$-error $E_2[W_I](t)$ and the deviation of total energy $E_H(t)$ are still adopted, which are plotted in Figures 12(a) and 12(b), respectively. The snapshots of $W_I(x, k, t)$ under $N_0 = 4 \times 10^7$ and $N_0 = 1 \times 10^8$ are plotted in Figure 13. The computational time, the average partition level $K$ and the growth ratio of particle
number $N(t)/N_0$ at $t = 9$ a.u. are recorded in Table 3. Aside from them, one can refer to Appendix C for more detailed comparison.

![Graphs showing evolutionary trends.](image)

(a) Evolution of stochastic variances. (b) Deviation in energy. (c) Evolution of partition levels. (d) Particle growth before and after SPADE.

Fig. 12. The time evolution of the $l^2$-errors $E_2[W_1](t)$ and the deviation in total energy in 12-D simulations, as well as the evolution of partition level and particle number. The accuracy can be improved by adopting larger sample size $N_0$ or smaller parameter $\vartheta$ in SPADE, resulted by the refinement in adaptive partition.

**Table 3**

Computational time in adaptive partitioning, average partition level $K$ and growth ratio of total particles for 12-D simulations up to 9 a.u.

| $N_0$    | $\vartheta = 0.004$ | $\vartheta = 0.005$ |
|----------|---------------------|---------------------|
|          | Time(h)  | Average K | $N(9)/N_0$ | Time(h)  | Average K | $N(9)/N_0$ |
| $1.6 \times 10^7$ | -       | -         | -           | 66.26   | 2.30$\times 10^7$ | 13.29       |
| $4 \times 10^7$   | 505.31   | 7.61$\times 10^7$ | 31.14       | 58.50   | 2.10$\times 10^7$ | 4.95        |
| $1 \times 10^8$   | 931.54   | 5.77$\times 10^7$ | 9.26        | 122.16  | 2.39$\times 10^7$ | 3.02        |

**Convergence:** The convergence of SPADE with respect to $N_0$ or $\vartheta$ is clearly presented in Figure 12(a). The accuracy of SPADE is significantly boosted via choosing a smaller $\vartheta$, and can be slightly improved under larger $N_0$. At the same time, the deviation of total energy can also be suppressed to a large extent when the partition is refined (see Figure 12(b)). These observations also support our findings in Table 1. From Figures 12(c) and 12(d), it is verified that the enhancement of accuracy is induced by the refinement in adaptive partition. The cost to afford is a rapid growth in computational time (see Table 3) since more particles remain uncancelled.

**Snapshots:** As expected in the adiabatic approximation, the electron-proton Wigner function behaves almost like a product of two uncorrelated wavepackets for $t = O(1)$. By the separation of variables, the single-body electron Wigner function is
approximately interacting with the nucleus fixed at \( x_A = (1, 0, 0) \), while the proton Wigner function is governed by the free advection. Numerical snapshots in Figure 13 indeed show that \( W_1(x, k, t) \) are almost comparable to their single-body counterpart produced by the deterministic solver, and the stochastic noises are evidently controlled when the sample size increases from \( N_0 = 4 \times 10^7 \) to \( 1 \times 10^8 \).

\[ \begin{array}{ccc}
\text{position} & \text{position} & \text{position} \\
(a) t = 3 \text{a.u.} & (b) t = 5 \text{a.u.} & (c) t = 7 \text{a.u.} \\
\end{array} \]

\[ \begin{array}{ccc}
\text{position} & \text{position} & \text{position} \\
(d) t = 9 \text{a.u.} & & \\
\end{array} \]

Fig. 13. Snapshots of the reduced electron Wigner function \( W_1(x, k, t) \) produced by the deterministic scheme (left), the 12-D stochastic simulations under \( N_0 = 4 \times 10^7 \) (middle) and \( N_0 = 1 \times 10^8 \) (right), with the parameter \( \theta = 0.004 \). The projection of two-body Wigner function seems to coincide with the single-body counterpart under the adiabatic approximation. The stochastic noises can be suppressed under larger sample size.

**Partition level and particle growth:** The bottom line structure is also observed in Figure 12(d) as the particle number remains stable after annihilation. This clearly manifests the capacity of SPADE to extract the minimal amount of negative
particles to present the non-classicality. Moreover, it is also observed that when larger sample size $N_0$ is adopted, the growth ratio of particle number after SPADE is always smaller, so that the overfitting problem can indeed be alleviated.

7. Conclusion and discussion. This paper discusses the adaptive particle annihilation algorithms to overcome the numerical sign problem in stochastic Wigner simulations. The Sequential-clustering Particle Annihilation via Discrepancy Estimation (SPADE) is proposed for breaking the curse of dimensionality (CoD) in existing particle annihilation via uniform mesh. By performing a series of benchmark tests on 6-D electron-proton coupling and a thorough comparison with our massively deterministic solver, we can conclude that (1) SPADE is able to learn the minimal amount of particles that capture the non-classicality of the Wigner function under arbitrary sample size $N_0$; (2) deepening the partition can systematically improve SPADE; (3) increasing the sample size $N_0$ can alleviate the overfitting problem. It follows by an attempt to simulate 12-D proton-electron Wigner dynamics. Experimental results demonstrate the potential of SPADE to overcome CoD in higher dimensional Wigner simulations. Our ongoing work is to boost the efficiency of adaptive partitioning and to explore the extension of WBRW-SPA-SPADE to the quantum BBGKY hierarchy [33, 9, 34], which paves a pivotal step for the interlacement of kinetic theory and molecular dynamics in high energy density physics [9], as well as lays the foundation for studying the Hydrogen tunneling via the Wigner approach [32].

Acknowledgement. This research was supported by the National Natural Science Foundation of China (No. 11822102, 1210010642), the Projects funded by China Postdoctoral Science Foundation (No. 2020TQ0011, 2021M690227) and the High-performance Computing Platform of Peking University. SS is partially supported by Beijing Academy of Artificial Intelligence (BAAI). The authors are sincerely grateful to the handling editor and referees for their patience and valuable suggestions. They would like also to thank Haoyang Liu and Shuyi Zhang at Peking University for their technical supports on computing environment, which greatly facilitate both stochastic and deterministic Wigner simulations.

REFERENCES

[1] E. Wigner. On the quantum corrections for thermodynamic equilibrium. Phys. Rev., 40:749-759, 1932.
[2] W. R. Frensley. Wigner-function model of a resonant-tunneling semiconductor device. Phys. Rev. B, 36:1570-1580, 1987.
[3] H. Kosina, V. Sverdlov, and T. Grasser. Wigner Monte Carlo Simulation: Particle Annihilation and Device Applications. 2006 International Conference on Simulation of Semiconductor Processes and Devices, 57-360, 2006.
[4] J. M. Sellier, M. Nedjalkov, I. Dimov, and S. Selberherr. A benchmark study of the Wigner Monte-Carlo method. Monte Carlo Methods Appl., 20:43-51, 2014.
[5] C. Jacoboni. Theory of Electron Transport in Semiconductors: A Pathway From Elementary Physics to Nonequilibrium Green functions. Springer Science & Business Media, 2010.
[6] M. Nedjalkov, D. Querlioz, P. Dollfus, and H. Kosina. Wigner Function Approach. In Nano-Electronic Devices. Springer, New York, 2011.
[7] L. Luca and V. Romano. Quantum corrected hydrodynamic models for charge transport in graphene. Ann. Phys., 406:30-53, 2019.
[8] M. Benam, M. Ballichia, J. Weinbub, S. Selberherr, and M. Nedjalkov. A computational approach for investigating Coulomb interaction using Wigner–Poisson coupling. J. Comput. Electron., 20:755-784, 2021.
[9] F. R. Graziani, J. D. Bauer, and M. S. Murillo. Kinetic theory molecular dynamics and hot dense matter: Theoretical foundations. Phys. Rev. E, 90(3):033104, 2014.
[10] C. Kurtsiefer, T. Pfau, and J. Mlynek. Measurement of the Wigner function of an ensemble of helium atoms. Nature, 386:150-153, 1997.

[11] B. I. Davies, R. Rundle, V. M. Dwyer, T. Todd, and M. Everitt. Visualizing spin degrees of freedom in atoms and molecules. Phys. Rev. A, 100:2469-9934, 2019.

[12] H. Kosina, M. Nedjalkov, and S. Selberherr. A Monte Carlo method seamlessly linking quantum and classical transport calculations. J. Comput. Electron., 2:147-151, 2003.

[13] Y. Xiong, Y. Zhang, and S. Shao. A characteristic-spectral-mixed scheme for six-dimensional Wigner-Coulomb dynamics. 2022. available at arXiv:2205.02380.

[14] O. Furtmaier, S. Succi, and M. Mendoza. Semi-spectral method for the Wigner equation. J. Comput. Phys., 305:1015-1036, 2016.

[15] M. L. Van de Put, B. Sorée, and W. Magnus. Efficient solution of the Wigner–Liouville equation using a spectral decomposition of the force field. J. Comput. Phys., 350:314-325, 2016.

[16] D. Querlioz and P. Dollfus. The Wigner Monte Carlo Method for Nanoelectronic Devices: A Particle Description of Quantum Transport and Decoherence. Wiley-ISTE, London, 2010.

[17] W. Wagner. A random cloud model for the Wigner equation. Kinet. Relat. Mod., 9:217-235, 2016.

[18] O. Muscato and W. Wagner. A class of stochastic algorithms for the Wigner equation. SIAM J. Sci. Comput., 38:A1483-A1507, 2016.

[19] S. Shao and J. M. Sellier. Comparison of deterministic and stochastic methods for time-dependent Wigner simulations. J. Comput. Phys., 300:167-185, 2015.

[20] S. Shao and Y. Xiong. A branching random walk method for many-body Wigner quantum dynamics. Numer. Math. Theor. Meth. Appl., 12:21-71, 2019.

[21] S. Shao and Y. Xiong. Branching random walk solutions to the Wigner equation. SIAM J. Numer. Anal., 58:2589-2608, 2020.

[22] T. C. Schmidt and K. Moehring. Stochastic path-integral simulation of quantum scattering. Phys. Rev. A, 48:R3418-R3420, 1993.

[23] M. Nedjalkov, H. Kosina, and S. Selberherr. The stationary Monte Carlo method for device simulation. II. Event biasing and variance estimation. J. Appl. Phys., 93:3564-3571, 2003.

[24] M. Troyer and U. J. Wiese. Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. Phys. Rev. Lett., 94:21-71, 2005.

[25] M. Iazzi, A. A. Soluyanov, and M. Troyer. Topological origin of the fermion sign problem. Phys. Rev. B, 93:15102, 2016.

[26] B. Yan and R. Caflisch. A Monte Carlo method with negative particles for Coulomb collisions. J. Comput. Phys., 298:711-740, 2015.

[27] M. Shah. A genetic algorithm approach to estimate lower bounds of the star discrepancy. Monte Carlo Methods Appl., 16:379-398, 2010.
Appendix A. Stochastic particle method for 4-D Wigner dynamics.

As another typical example, we consider the electron dynamics in 4-D phase space under the 2-D Morse potential \((N = 1, d = 2)\),

\[
V(x) = -2e^{-\kappa(|x-x_A|-r_0)} + e^{-2\kappa(|x-x_A|-r_0)},
\]

with \(x_A = (0, 0), r_0 = 0.5, \kappa = 0.5, \hbar = m = 1\). Under this potential, an equivalent form of \(\Psi\)DO reads that

\[
\Theta_V[\varphi](x, k, t) = -\frac{\kappa e^{\kappa r_0} c_2}{\hbar} \int_0^{2\pi} \int_0^{+\infty} \int_0^{+\infty} \frac{r \sin(2(x - x_A) \cdot k') \Delta_{r_{\sigma}}[\varphi](x, k, t)}{\sqrt{r^2 + (\kappa/2)^2}} r^2 + (\kappa/2)^2
\]

\[
+ \frac{\kappa e^{2\kappa r_0} c_2}{\hbar} \int_0^{2\pi} \int_0^{+\infty} \int_0^{+\infty} \frac{r \sin(2(x - x_A) \cdot k') \Delta_{r_{\sigma}}[\varphi](x, k, t)}{\sqrt{r^2 + \kappa^2}} r^2 + \kappa^2.
\]

where \(k' = r\sigma = (r \cos \theta, r \sin \theta)\) and \(c_2 = \Gamma(3/2)\pi^{-3/2} \approx 0.1592\), and the difference operator \(\Delta_{r_{\sigma}}[\varphi](x, k, t) = \varphi(x, k - r\sigma/2, t) - \varphi(x, k + r\sigma/2, t)\). Although the eigenfunctions of the quantum Hamiltonian operator can be solved, its phase space solution from the first principle is less than straightforward or complete, especially for its non-equilibrium dynamics [34].

To visualize the quantum dynamics in phase space, we adopt the reduced Wigner function

\[
W_1(x, k, t) = \int_{\mathbb{R}^2} f(x_1, x_2, k_1, k_2, t) dx_2 dk_2,
\]

and the spatial marginal distribution

\[
P(x_1, x_2, t) = \int_{\mathbb{R}^2} f(x_1, x_2, k_1, k_2, t) dk_1 dk_2.
\]

The initial Wigner function is a Gaussian wavepacket of a quantum mixed state,

\[
f_0(x_1, x_2, k_1, k_2, 0) = \frac{1}{\pi^2} e^{-0.5(x_1 - 8)^2 - 0.5(x_2 - 12)^2 - 2(k_1 - 0.5)^2 - 2(k_2 + 0.5)^2}.
\]

The snapshots of \(W_1(x, k, t)\) and \(P(x_1, x_2, t)\) up to \(t = 10\) a.u. are plotted in Figure A.1. One can clearly observe the negative components and oscillatory structure of the Wigner function in phase space. In the spatial direction, the Gaussian wavepacket is first attracted by the interacting body fixed at the origin, and then presents an oscillating pattern near the origin \(x_A\).

A complete flowchart of the stochastic Wigner simulations, as depicted in Figure A.2, consists of three cornerstones: Probabilistic (branching random walk) interpretation to the Wigner equation, sequential importance sampling and particle resampling (either particle annihilation via uniform mesh (PAUM) [3, 18, 35] or adaptive particle annihilation algorithm SPADE [26]).

The fundamental obstacle to the stochastic Wigner algorithms is the numerical sign problem, which is believed to be NP-hard in nature [24, 25]. In fact, we have proved that both particle number and stochastic variances in such branching particle system may grow exponentially in time, thereby dramatically hampering the efficiency of the sampling process [21]. Nevertheless, the sign problem can be alleviated to a large extent if one carefully cancels out the stochastic trajectories that contribute little to the expectation but amplify the variance.
1. **Probabilistic interpretation** represents the solution of the deterministic Wigner equation by the expectation of a branching random process. It defines the probabilistic rules of particle motions and generations.

2. **Sequential importance sampling** represents the Wigner function by a fi-
finite collection of weighted particles. First, the initial data is approximated by a weighted empirical measure. Second, particles move under the probabilistic rules, including deterministic motions, random jumps and random generations.

3. **Particle resampling** intends to adjust the particle weights by combinatorial techniques, such as clustering, matching and sampling with or without replacement. Specially, for a finite particle system involving both positive and negative particles, the particle resampling, also termed particle annihilation, cancels out those carrying opposite weights, thereby controlling the rapid growth of particle number and stochastic variances.

**A.1. The stationary phase approximation alleviates the sign problem.**
In our recent work [21], we have analyzed the bound of stochastic variances and found that the numerical sign problem is actually induced by the splitting technique. Although it gives a practical stochastic interpretation to $\Psi DO$, the splitting of the oscillatory integral ignores the near-cancelation of its high-frequency components, and consequently leads to an exponential increases of errors. The remedy is the stationary phase approximation (SPA) to $\Psi DO$. The leading terms of the asymptotic expansion capture the major contribution of the oscillatory integrals.

First, we introduce a filter $\lambda_0$ and a ball $B(r)$ with radius $r$, and try to replace the components outside the ball by an integral over a line,

$$\Theta^0_V[\varphi](x, k, t) = \int_{B(\frac{x_A - x}{\lambda_0})} e^{\frac{i}{\hbar}(x - x_A) \cdot k'} \psi(k')(\varphi(x, k - \frac{k'}{2}, t) - \varphi(x, k + \frac{k'}{2}, t)) dk'$$

$$+ 2 \int_{\frac{x_A - x}{\lambda_0}}^{+\infty} \Im \left( \sqrt{\frac{2\pi e^3 r|x - x_A|}{\sqrt{1r|x - x_A|}}} \right) r \psi(r \sigma_*)(\varphi(x, k - \frac{r \sigma_*}{2}, t) - \varphi(x, k + \frac{r \sigma_*}{2}, t)) dr,$$

where the amplitude function reads

$$\psi(k) = \frac{1}{i\hbar} \left[ -\frac{2\kappa e^{\kappa r_0} c_2}{(|k|^2 + \kappa^2)^{3/2}} + \frac{2\kappa e^{2\kappa r_0} c_2}{(|k|^2 + 4\kappa^2)^{3/2}} \right],$$

with $\sigma_* = (\cos \phi_*, \sin \phi_*)$ and $\phi_* = \text{atan}2(\frac{x_2 - x A_2}{x_1 - x A_1})$. One can prove

$$\Theta_V[\varphi](x, k, t) = \Theta^0_V[\varphi](x, k, t) + O(\lambda_0^{-1}),$$
in the sense that
\[ \| \Theta_V [\varphi](x, k, t) - \Theta_V^{\lambda_0}[\varphi](x, k, t) \|_{L^2_L L^2_k} \lesssim \lambda_0^{-1} \| \varphi(t) \|_{L^2_L L^2_k}, \]
and \[ \| \varphi(t) \|_{L^2_L H^1_k} = \| \varphi(t) \|_{L^2_L L^2_k} + \| \varphi(t) \nabla_k \|_{L^2_L L^2_k}. \] When \( \lambda_0 \) is larger than 1, the asymptotic error term \( O(\lambda_0^{-1}) \) decays as \( \lambda_0 \) increases.

Again, we take the 4-D Wigner equation under the Morse potential as an example. The implementation of the Wigner Branching Random Walk associated with SPA (termed WBRW-SPA for short) is illustrated in Algorithm A.1, starting from the initial instant \( t_l \) and stopping at the final instant \( t_{l+1} \). In this way, the particle method resolves the Wigner dynamics by simulating the deterministic motions, random jumps, random generation of superparticles in the phase space.

The remaining problem is how to choose the filter \( \lambda_0 \). From the theoretical results, \( \lambda_0 \) must not be too small, otherwise the asymptotic errors will be augmented. A visualization of the reduced Wigner function at \( t = 4 \) a.u., produced by WBRW-SPA under the initial effective sample size \( 1 \times 10^7 \), is presented in Figure A.3. It seems that \( \lambda_0 = 6 \) achieves the best performance in suppressing the random noises. Apparently, SPA under \( \lambda_0 = 1 \) or 2 fails to produce correct results, and the noises seems to be amplified when \( \lambda_0 \) is too large (\( \lambda_0 = 16 \)).

![Visualization of the reduced Wigner function](image)

**Fig. A.3.** The 4-D Morse system: A visualization of the reduced Wigner function \( W_1(x, k, t) \) produced by WBRW-SPA, under different filter \( \lambda_0 \).

To qualify the variances, we measure the \( l^2 \)-error of \( W_1(x, k, t) \) and \( P(x_1, x_2, t) \), the deviation of total energy \( \varepsilon_H(t) \) as defined in Eq. (5.2). According to Figure A.4, SPA under \( \lambda_0 = 6 \) indeed alleviates the exponential growth of particle number and variances simultaneously. The growth ratio of total particle with SPA is 157 at \( t = 4 \) a.u., compared to 193 without SPA. Too small \( \lambda_0 \) kills the accuracy due to the large asymptotic errors, while too large \( \lambda_0 \) may fail to kill redundant particles.

In practice, the choice of \( \lambda_0 \) can be determined by monitoring the deviation in Hamiltonian. In Figure 4(c), a large fluctuation of the total Hamiltonian is observed
Algorithm A.1 WBRW-SPA for the 2-D Morse system.

Input parameters: The initial time \( t_0 \) and final time \( t_{f+1} \), the constant rate \( \gamma_0 \), the filter \( \lambda_0 \), \( k \)-domain \( K \), and the upper band \( r_{\text{max}} > 4|K| \).

Sampling processes: Suppose each particle in the branching particle system, carrying an initial weight \( w \) either 1 or \(-1\), starts at state \((x, k)\) at time \( t_0 \) and moves until \( t_{f+1} = t_0 + \Delta t \) according to the following rules.

1. (Frozen) Generate a random \( \tau \sim \gamma_0 e^{-\gamma_0 t} \). For a particle at \((x, k)\) at instant \( t \in [t_0, t_{f+1}] \), if \( t + \tau \geq t_{f+1} \), it becomes frozen at \((x + \frac{\Delta k(t_{f+1} - t)}{m}, k, t_{f+1})\).
2. (Death) If \( \tau < \Delta t \), the particle moves to \((x + \frac{\Delta k}{m}, k, t + \tau)\) and is killed.
3. (Branching) When the particle is killed, it produces at most three offsprings at states \((x(1), k(1), t + \tau)\), \((x(2), k(2), t + \tau)\) and \((x(3), k(3), t + \tau)\). The third offspring is produced at state \((x(3), k(3)) = (x + \frac{\Delta k}{m}, k)\) with probability 1, carrying the weight \( w \).
4. (Scattering) Generate a random \( r \) from the Cauchy distribution \( \frac{1}{\pi r^2 + \kappa^2} \).
   (1) If \( r < \lambda_0/|x(3) - x_A| \), generate random numbers \( \phi \) uniformly in \([0, 2\pi]\), yielding vectors \( \sigma = (\cos\phi, \sin\phi) \), \( k' = r\sigma \). It produces two offsprings with probability \( \Pr(1) = \Pr(2) = \Pr \) at states \((x(1), k(1))\), \((x(2), k(2))\) endowed with updated weights \( w_1 \) and \( w_2 \), respectively.
   \[ Pr = \frac{|\psi(x(3), r, \sigma)|}{\gamma_0}, \]
   \[ x(1) = x(2) = x(3), \quad k(1) = k - \frac{k'}{2}, \quad k(2) = k + \frac{k'}{2}, \]
   \[ w_i w, \quad (-1)^{i-1} \psi(x(3), r, \sigma) \frac{1}{|\psi(x(3), r, \sigma)|} = \mathbb{I}(k(i) \in K), \quad i = 1, 2. \]
   (2) If \( r \geq \lambda_0/|x(3) - x_A| \), it produces two offsprings with the probability \( \Pr(1) = \Pr(2) = \Pr \) at states \((x(1), k(1))\), \((x(2), k(2))\) endowed with updated weights \( w_1 \) and \( w_2 \), respectively.
   \[ Pr = 2\gamma_0 \frac{\text{Im} \left[ e^{i\phi} \frac{2\pi}{|x(3) - x_A|}^{\frac{1}{2}} \right]}{\text{Im} \left[ e^{i\phi} \frac{2\pi}{|x(3) - x_A|}^{\frac{1}{2}} \right]} \]
   \[ x(1) = x(2) = x(3), \quad k(1) = k - \frac{r\sigma_s(x(3))}{2}, \quad k(2) = k + \frac{r\sigma_s(x(3))}{2}, \]
   \[ w_i = w, \quad \frac{\text{Im} \left[ e^{i\phi} \frac{2\pi}{|x(3) - x_A|}^{\frac{1}{2}} \right]}{\text{Im} \left[ e^{i\phi} \frac{2\pi}{|x(3) - x_A|}^{\frac{1}{2}} \right]} = \mathbb{I}(k(i) \in K), \quad i = 1, 2, \]
   for \( i = 1, 2 \), with \( \sigma_s(x(3)) = (\cos\phi_s, \sin\phi_s) \) and \( \phi_s = \tan^{-1} \left( \frac{x(3) - x_A}{x(3) - x_A} \right) \).
5. (Independence) The offsprings continue to move independently.

Termination condition: All particles in the branching particle system are frozen.

under \( \lambda_0 = 1, 2 \), while the deviation becomes very small under \( \lambda_0 = 6, 16 \). This is consistent with the trends in Figures 4(a) and 4(b).
Appendix B. Performance evaluation of SPADE in 4-D Wigner simulations.

From this section, we are about to make a thorough benchmark on SPADE by simulating the 4-D Wigner equation under the Morse potential, the purpose of which is is two-pronged. First, we make a thorough comparison between PAUM and SPADE. Second, we would like to investigate how the parameter $\vartheta$ in SPADE and sample size $N_0$ influence the numerical accuracy, energy conservation, particle number and the partition level $K$. The latter is towards a comprehensive understanding of SPADE and a guiding principle for improving accuracy systematically, and is pivotal to rigorous numerical analysis.

The reference solutions are produced by a high accurate deterministic advective-spectral-mixed scheme [36], where the Wigner function defined in a 4-D computational domain $[-12,12]^2 \times [-\frac{5\pi}{4},\frac{5\pi}{4}]^2$ is expanded as the tensor product of $163^2$ cubic spline basis and $128^2$ Chebyshev spectral basis (with $8 \times 8$ cells), with $\Delta y_\nu = \Delta y_\mu = 0.3$ and $y_\nu, y_\mu$ truncated at $Y = [-15,15]^2$. The three-step Lawson scheme is used for temporal integration, with time step $\Delta t = 0.02\text{a.u.}$

For stochastic particle simulations, we adopt $\gamma_0 = 2.59$, $\lambda_0 = 6$ and a finite $k$-domain $[-5,5]^2$ in Algorithm 1, and annihilate particles every $1\text{a.u.}$. The reduced Wigner function $W_1(x,k,t)$ and spatial marginal distribution $P(x_1,x_2,t)$ can be readily obtained by histogram reconstruction under a uniform grid mesh $[-12,12]^2 \times [-5,5]$ with $N_x = 161$, $N_k = 100$, $\Delta x = 0.15$, $\Delta k = 0.1$. For instance, for $x_\mu = [-12 + (\mu - 1)\Delta x,-12 + \mu \Delta x]$, $k_\nu = [-5 + (\nu - 1)\Delta k,-5 + \nu \Delta k]$, the reduced Wigner function $W_1(x,k,t)$ can be reconstructed by Eq. (3.5) As a comparison, we also perform the stochastic Wigner simulations and annihilate particles via PAUM with a $161^3 \times 100^3$
uniform grid mesh. The partition level is $K = 2.592 \times 10^8$, which is much larger than the sample size. We still adopt the $l^2$-error $\mathcal{E}_2[W_1](t)$ (5.1), the $l^2$-error for $\mathcal{E}_2[P_1](t)$ and the deviation of the total Hamiltonian (5.2) to measure the stochastic variances.

### B.1. Comparison between PAUM and SPADE

Now we begin to make a thorough comparison between SPADE and PAUM. We provide a visualization of the reduced Wigner function $W_1(x, k, t)$ and spatial distribution $P(x_1, x_2, t)$ at $t = 10$ a.u. in Figures B.2 and B.3, respectively, which vividly compares the results produced by deterministic scheme, PAUM and SPADE ($\vartheta = 0.003$ to 0.16). The time evolution of $l^2$-errors, as well as the deviations of total energy, are plotted in Figure B.5. The growth ratio of total particle number $P(t) + M(t)$ is plotted in Figure B.4. Based on these results, we have the following observations.

**Snapshots:** The snapshots demonstrate the capability of stochastic Wigner algorithm, with either PAUM or SPADE, to recover the fine oscillating structure of the Wigner equation when sample size is sufficiently large ($N_0 = 4 \times 10^7$). However, when the sample size decreases to $N_0 = 1 \times 10^7$, it is found that the result produced by PAUM is evidently more noisy than that by SPADE, as visualized in Figure B.1. What is worse, when $N_0 = 1 \times 10^6$, PAUM might fail to produce reliable results, while SPADE still works in this situation. This is because of the overfitting problem, as the partition level $K$ is much larger than the sample size in PAUM. As a result, many particles might be left uncanceled.

**Comparison between PAUM and SPADE:** According to Figure B.5, it is verified that the rapid growth of stochastic variances can be dramatically suppressed when the particle annihilation is used. The performance of PAUM outperforms SPADE when $\vartheta \geq 0.04$, while their accuracy seems to be comparable when $\vartheta = 0.02$. By further decreasing $\vartheta$ to 0.003 or 0.005, the accuracy of SPADE even outperforms PAUM, which coincides with the observation in Figure B.1. In fact, the performance of PAUM is sensitive to the sample-to-partition ratio $N_0/K$. When $N_0/K$ is too small, PAUM fails to kill redundant particles, so that might not suppress the random noises efficiently. This phenomenon has also been observed in our previous work [35]. By contrast, the average partition level in SPADE is $6.83 \times 10^5$ under $N_0 = 1 \times 10^7$ and $\vartheta = 0.005$, which ensures the efficiency of annihilation.

**Particle growth:** The particle growth is presented in Figure B.4. When $N_0 = 1 \times 10^6$, the particle number after PAUM reaches $4.5 \times 10^7$ until $t = 10$ a.u. (growth ratio is 11.3). Meanwhile, when $N_0 = 1 \times 10^7$, the particle number after PAUM reaches $7.4 \times 10^7$ (growth ratio is 7.4). This accounts for the reason that PAUM only works when $N_0$ is comparable to $K$ but becomes inefficient to suppress the growth of stochastic variances when $N_0$ is much smaller than $K$. By contrast, particle number after SPADE almost remains at a stable level because it can be automatically adapted to an appropriate partition level and avoid the “overfitting” phenomenon in PAUM. The exception is the group $N_0 = 4 \times 10^6$, $\vartheta = 0.003$ in Figure 4(a), where too small $\vartheta$ may lead to over-refined partition and hamper the efficiency of SPADE when sample size is not enough.

### B.2. Deep partition improves SPADE

Now we would like to demonstrate that the accuracy of SPADE can be systematically improved by deepening the partition, which is realized by choosing smaller $\vartheta$. To this end, we fix the sample size $N_0$ and evaluate the performance of SPADE under seven choices of particle number $\vartheta = 0.003, 0.005, 0.01, 0.02, 0.04, 0.08, 0.16$. The time evolutions of $l^2$-errors and deviation in energy are plotted in Figures B.6 and B.7, respectively. The partition level $K$ is recorded in Figure B.8. Based on the numerical results, we have the following
observations.

**Convergence with respect to \( \vartheta \):** According to Figure B.6, the numerical errors can be diminished by gradually decreasing \( \vartheta \) from 0.16 to 0.003, indicating that refinement in the adaptive partition can systematically improve the accuracy. The numerical energy may slightly increases due to the accumulation of SPADE errors, but the deviations can be alleviated when the partition is deepened.

**Partition level \( K \) with respect to \( \vartheta \):** In Figure B.8, it is clearly seen that the average partition level \( K \) increases along with the decrease in \( \vartheta \), and consequently leads to a reduction in stochastic variances. An exception is still the group \( N_0 = 4 \times 10^6, \vartheta = 0.003 \) due to the overfitting problem. From Figure 6(a), the over-refinement in partition may augment the numerical errors. It seems that \( K \) is inversely proportional to \( \vartheta \). This actually gives us a hint to postulate the partition level \( K \) by first performing some tests under relatively larger \( \vartheta \).
B.3. Large sample size improves SPADE. The accuracy of SPADE can also be improved by increasing the effective sample size $N_0$. Five groups of simulations are performed under the sample size $N_0 = 4 \times 10^5$, $1 \times 10^6$, $4 \times 10^6$, $1 \times 10^7$ and $4 \times 10^7$. The time evolution of $L^2$-errors is plotted in Figure B.9 and the partition level $K$ is recorded in Figure B.10. Based on the numerical results, we have the following observations.

Convergence with respect to $N_0$: According to Figure B.9, the numerical accuracy can be systematically improved by increasing $N_0$ from $2 \times 10^5$ to $4 \times 10^7$, which validate the convergence of stochastic Wigner algorithm. However, the convergence rate largely deviates from $-1/2$ as seen in Figure 10(a) due to the mixture of MC errors and SPADE errors. Again, the deviation in total energy can be suppressed when $N_0$ becomes larger.

Partition level $K$ with respect to $N_0$: As shown in Figure 10(b), it seems that the average partition level $K$ depends more on the parameter $\vartheta$, instead of $N_0$. That partially accounts for the reason that SPADE can work under a wide spectrum.
Fig. B.3. The 4-D Morse system: Visualization of the spatial marginal distribution $P(x_1, x_2, t)$ produced by the deterministic scheme (left), WBRW-SPA-PAUM under $N_0 = 4 \times 10^7$ (middle) and WBRW-SPA-SPADE under $N_0 = 4 \times 10^7$, $\vartheta = 0.003$ (right).

(a) $t = 2.5$ a.u.
(b) $t = 5$ a.u.
(c) $t = 7.5$ a.u.
(d) $t = 10$ a.u.

Fig. B.4. The 4-D Morse system: Growth ratio of particle number under PAUM and SPADE. The efficiency of particle annihilation may be hampered when the partition level is much larger than sample size as many particles are uncanceled, which is known as the overfitting problem.

(a) $N_0 = 4 \times 10^6$.
(b) $N_0 = 1 \times 10^7$.
(c) $N_0 = 4 \times 10^7$. 

1 2 3 4 5 6 7 8 9 10
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5
6
7
8
9
10
Fig. B.5. The 4-D Morse system: Evolution of errors (left: reduced Wigner function, middle: spatial distribution, right: deviations in energy). SPADE is inferior to PAUM when $\vartheta = 0.04$, $0.08$, $0.16$ but outperforms PAUM when $\vartheta = 0.003$, $0.005$. The efficiency of PAUM is ensured only when $N_0$ is sufficiently large.
Fig. B.6. The 4-D Morse system: $l^2$-errors under different $N_0$ (left: reduced Wigner function, right: spatial distribution). The accuracy of SPADE can be systematically improved by choosing smaller parameter $\vartheta$ and deepening the partitioning.

Fig. B.7. The 4-D Morse system: Deviation of total energy can be ameliorated by choosing small $\vartheta$ and deepening the partitioning.
Fig. B.8. The 4-D Morse system: Choosing small $\vartheta$ may lead to an exponential growth of partition level $K$, and consequently improve the accuracy of SPADE systematically. An exception is the group $\vartheta = 0.003$, $N_0 = 4 \times 10^6$ as too many redundant particles are not killed, which may hamper the accuracy of PA.

of sample sizes. Too small $\vartheta$ is NOT recommended to be used when sample size is not large, as the partition level may increase rapidly and even exceed the sample size, and consequently lead to the overfitting problem and hamper the efficiency of SPADE (see the group $N_0 = 4 \times 10^6$, $\vartheta = 0.003$ in Figure B.10).
Fig. B.9. The 4-D Morse system: $l^2$-errors under different $N_0$ (left: reduced Wigner function, right: spatial distribution). The accuracy of SPADE is improved as the sample size increases.

Fig. B.10. The 4-D Morse system: The convergence and the partition level with respect to $N_0$. 

(a) Convergence with respect to $N_0$. (b) The relation between $N_0$ and $K$.
B.4. How different splitting strategies influence SPADE. In practice, we have found that the splitting strategy in SPADE has a great influence on the numerical accuracy. The maximal gap works better for $D \leq 6$, while the difference gap performs better for $D > 6$.

Some numerical evidences are provided. When simulating 4-D Wigner equation under the Morse potential, one can realize that the difference gap gives very poor results and fails to capture several fine structures (such as the negative valley) of the Wigner equation, as visualized in Figure B.11. The rapid growth of $l^2$-errors, as presented in Figure B.11, also validates our observation.

Fig. B.11. The 4-D Morse system: A comparison of $\mathcal{E}_2[W_1](t)$ and $\mathcal{E}_2[P](t)$ under two kinds of splitting strategies, with $N_0 = 4 \times 10^7$, $\vartheta = 0.01$. The maximal gap produces much better numerical results.
Appendix C. Proton-electron Wigner equation in 12-D phase space.

Now we shall move to a realistic and much more challenging problem, say, the first-principle solution to non-equilibrium proton-electron coupling in 12-D phase space as given in Example 3.

C.1. Derivation of proton-electron Wigner dynamics. Consider one proton and one electron interacting under the Coulomb potential

$$V(x_e, x_p) = -\frac{1}{|x_e - x_p|},$$

the proton-electron Wigner equation in 12-D phase space reads that

$$\frac{\partial}{\partial t} f(x_e, x_p, k_e, k_p, t) + \frac{\hbar k_e}{m_e} \cdot \nabla_{x_e} f(x_e, x_p, k_e, k_p, t) + \frac{\hbar k_p}{m_p} \cdot \nabla_{x_p} f(x_e, x_p, k_e, k_p, t)$$

$$= \frac{1}{i\hbar (2\pi)^3} \int_{\mathbb{R}^6 \times \mathbb{R}^6} e^{-i(k_e - k'_e) \cdot y_e - i(k_p - k'_p) \cdot y_p} f(x_e, x_p, k'_e, k'_p, t) dy_e dy_p dk'_e dk'_p$$

$$- \frac{1}{i\hbar (2\pi)^3} \int_{\mathbb{R}^6 \times \mathbb{R}^6} e^{-i(k_e - k'_e) \cdot y_e - i(k_p - k'_p) \cdot y_p} f(x_e, x_p, k'_e, k'_p, t) dy_e dy_p dk'_e dk'_p.$$

By the conversion $y_e - y_p = \xi_1, \frac{y_e + y_p}{2} = \xi_2$, it yields that

$$\text{RHS} = \int_{\mathbb{R}^1} e^{-i(k_e - k'_e) \cdot y_e - i(k_p - k'_p) \cdot y_p} f(x_e, x_p, k'_e, k'_p, t) dy_e dy_p dk'_e dk'_p$$

Using the Fourier completeness relation,

$$\int_{\mathbb{R}^3} e^{-i(k_e - k'_e) \cdot \xi_2 - i(k_p - k'_p) \cdot \xi_2} d\xi_2 = (2\pi)^3 \delta(k_e - k'_e - k_p + k'_p),$$

it further yields that

$$\text{RHS} = \frac{1}{i\hbar (2\pi)^3} \int_{\mathbb{R}^6} e^{-i(k_e - k'_e) \cdot \xi_1} f(x_e, x_p, k'_e, k_p - k_e + k'_e, t) d\xi_1 dk'_e$$

$$- \frac{1}{i\hbar (2\pi)^3} \int_{\mathbb{R}^6} e^{i(k_e - k'_e) \cdot \xi_1} f(x_e, x_p, k'_e, k_p - k_e + k'_e, t) d\xi_1 dk'_e$$

$$= \frac{1}{i\hbar (2\pi)^3} \int_{\mathbb{R}^6} e^{-i(k'_e - k'_e) \cdot \xi_1} f(x_e, x_p, k'_e, k_p + k'_e, t) d\xi_1 dk'_e$$

$$- \frac{1}{i\hbar (2\pi)^3} \int_{\mathbb{R}^6} e^{-i(k'_e - k'_e) \cdot \xi_1} f(x_e, x_p, k'_e, k_p - k'_e, t) d\xi_1 dk'_e.$$

Finally, by changing the variables $\xi_1 \to \eta + 2x_p - 2x_e$ for the first line and $\xi_1 \to \eta - 2x_p + 2x_e$ for the second, it arrives at

$$\Theta_V[f] = \frac{2}{i\hbar (2\pi)^3} \int_{\mathbb{R}^6} e^{2i(k'_e \cdot (x_e - x_p) - i k'_e \cdot \eta)} f(x_e, x_p, k'_e, k_p + k'_e, t) d\eta dk'_e$$

$$- \frac{2}{i\hbar (2\pi)^3} \int_{\mathbb{R}^6} e^{-2i(k'_e \cdot (x_e - x_p) - i k'_e \cdot \eta)} f(x_e, x_p, k'_e, k_p - k'_e, t) d\eta dk'_e.$$
Since the Fourier conjugate of \( \frac{1}{|\eta|} \) is \( \frac{2\pi}{c_{n,\alpha}^2} \frac{1}{|k|^2} \), \( c_{n,\alpha} = \pi^{n/2} 2^{\alpha} \Gamma(\frac{\alpha}{2}) \), and \( k' \rightarrow \pm k'/2 \) for the first and second line, respectively, we obtain that

\[
\Theta_V[f](x, k, t) = \frac{1}{\hbar c_{3,1}} \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k}{2}, t) dk' \\
- \frac{1}{\hbar c_{3,1}} \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e + \frac{k'}{2}, k_p - \frac{k}{2}, t) dk'.
\]

### C.2. Adiabatic approximation.

Since \( m_p \approx 1836 m_e \), the proton moves much slower than the electron. This motivates us to consider the adiabatic approximation.

Starting from the Wigner equation

\[
\frac{\partial}{\partial t} f(x_e, x_p, k_e, k_p, t) + \frac{\hbar k_e}{m_e} \cdot \nabla_{x_e} f(x_e, x_p, k_e, k_p, t) + \frac{\hbar k_p}{m_p} \cdot \nabla_{x_p} f(x_e, x_p, k_e, k_p, t)
\]

\[
= \frac{1}{\hbar c_{3,1}} \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k}{2}, t) dk' \\
- \frac{1}{\hbar c_{3,1}} \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e + \frac{k'}{2}, k_p - \frac{k}{2}, t) dk',
\]

we can define the electron marginal Wigner function by projecting \( f(x_e, x_p, k_e, k_p, t) \) onto \((x_p, k_e)\) space

\[
P_e(x_e, k_e, t) = \int_{\mathbb{R}^3} f(x_e, x_p, k_e, k_p, t) dx_p dk_p.
\]

To illustrate its derivation of adiabatic approximation, we first start from

\[
\int_{\mathbb{R}^3} dx_p \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k}{2}, t) dk' \\
= \int_{\mathbb{R}^3} dx_p \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k}{2}, t) dk' \\
+ \int_{\mathbb{R}^3} dx_p \int_{\mathbb{R}^3} \left( e^{-i k' \cdot (x_e - x_p)} - 1 \right) \frac{e^{i k' \cdot (x_e - x_p)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k}{2}, t) dk'.
\]

For the first term, using the variable conversion \( k_p \rightarrow k_p - k'/2 \), it has that

\[
\int_{\mathbb{R}^3} dk_p \int_{\mathbb{R}^3} dx_p \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_A)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k}{2}, t) dk' \\
= \int_{\mathbb{R}^3} dk_p \int_{\mathbb{R}^3} dx_p \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_A)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p, t) dk' \\
= \int_{\mathbb{R}^3} \frac{e^{i k' \cdot (x_e - x_A)}}{|k'|^2} P_e(x_e, k_e - \frac{k'}{2}, t) dk'.
\]

For the second term, using the fact

\[
f(x_e, x_p, k_e, k_p, t) = g(x_e, x_p, k_e, k_p, t) e^{-|x_p - x_A|^2 / \alpha}
\]

43
and the Laplace asymptotic expansion, it yields that
\[
\Pi = \int_{\mathbb{R}^3} \mathrm{d}x_p \int_{\mathbb{R}^3} \left( e^{-ik'(x_p - x_A)} - 1 \right) \frac{e^{ik'(x_p - x_A)}}{|k'|^2} f(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) \mathrm{d}k' \\
= \int_{\mathbb{R}^3} e^{-ik_p \cdot \frac{\hbar x_p}{m_p} + \frac{\hbar |k'|^2}{2m_p}} \frac{\epsilon_p (x_p - x_A)}{|k'|^2} g(x_e, x_p, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) \mathrm{d}k' \\
\sim (2\pi \varepsilon)^{3/2} \int_{\mathbb{R}^3} \left( e^{-ik_p \cdot \frac{\hbar x_p}{m_p} + \frac{\hbar |k'|^2}{2m_p}} - 1 \right) \frac{e^{ik'(x_p - x_A)}}{|k'|^2} g(x_e, x_A, k_e - \frac{k'}{2}, k_p + \frac{k'}{2}, t) \mathrm{d}k' \sim O(\hbar m_p^{-1} \varepsilon^{3/2}).
\]

By further integrating in \( k_p \)-space, we have that
\[
\frac{\partial}{\partial t} P_e(x_e, x_p, t) + \frac{\hbar k_e}{m_e} \nabla_{x_e} P_e(x_e, x_p, t) \\
= \int_{\mathbb{R}^3} \frac{e^{ik'(x_p - x_A)}}{|k'|^2} \frac{1}{2m_p} P_e(x_e, k_e - \frac{k'}{2}, t) \mathrm{d}k' + O(\hbar m_p^{-1} \varepsilon^{3/2}).
\]

Finally, by omitting the asymptotic error terms and using the uncorrelated initial condition \( f(x_e, x_p, k_e, k_p, 0) = f_e(x_e, k_e, 0) f_p(x_p, k_p, 0) \), we arrive at Eq. (2.10).

**C.3. Particle simulations in 12-D phase space.** With the above preparation, it begins to simulate the proton-electron Wigner equation in the 12-D phase space and investigate the performance of SPADE.

![Graphs showing particle simulations in 12-D phase space](image)

(a) The growth of stochastic variances. (b) Deviation in Hamiltonian.

(c) Deterministic. (d) SPADE, maximal gap. (e) SPADE, difference gap.

**Fig. C.1. The 12-D proton-electron coupling: A comparison of the reduced Wigner function \( W_1(x, k, t) \) under two kinds of splitting strategy, with \( N_0 = 1.6 \times 10^7, \vartheta = 0.005 \). Apparently, the difference gap produces much better results.**

Unlike 4-D problem, the difference gap seems to outperform the maximal gap for 12-D problem. Figure C.1 gives a comparison between \( W_1(x, k, t) \) under different strategies. When the maximal gap is adopted, it seems that particle method might fail
to capture the negative parts properly, corresponding to the evidently larger $l^2$-error and deviation in total energy.

The snapshots of the spatial marginal density of electron

$$ P(x_1, x_2, t) = \int \int \int_{\mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3} f(x_1, x_2, x_3, x_p, k_e, k_p, t) dx_3 dx_p dk_e dk_p, $$

the reduced electron Wigner function

$$ W_1(x, k, t) = \int \int \int_{\mathbb{R}^2 \times \mathbb{R}^3 \times \mathbb{R}^2 \times \mathbb{R}^2} f(x_e, x_p, k_e, k_p, t) dx_{e, 2} dx_{e, 3} dx_p dk_{e, 2} dk_{e, 3} dk_p, $$

and the reduced proton Wigner function

$$ W_4(x, k, t) = \int \int \int_{\mathbb{R}^3 \times \mathbb{R}^2 \times \mathbb{R}^3 \times \mathbb{R}^2} f(x_e, x_p, k_e, k_p, t) dx_e dx_p dx_{p, 2} dx_{p, 3} dk_e dk_{p, 2} dk_{p, 3} $$

are visualized in Figures C.2, C.3 and C.4, respectively. The following observations are made from the results.

1. The particle-based stochastic algorithm can capture the pattern of spatial unharmonic oscillation of electron. The difference mainly lies at the peak of the wavepacket, which may be smoothed out by the piecewise constant reconstruction. Fortunately, from the comparison of $\int_{\mathbb{R}} P(x_1, x_2, t) dx_2$ on the right column of Figure C.2, the difference can be compensated by either increasing $N_0$ or refining the partition (choosing smaller $\vartheta$).

2. From Figures C.3 and C.4, it is clearly seen that the electron behaves almost like the single-body counterpart, while the proton is governed by the free advection, which coincides with the prediction of the adiabatic approximation. The main features of the electron Wigner function, including the double-peak structure induced by the Coulomb collisions and the negative valley that manifests the uncertainty, can be captured by the particle-based stochastic algorithm, albeit with slight stochastic noises.
Fig. C.2. The 12-D proton-electron coupling: Visualization of the marginal spatial distribution $P(x_1, x_2)$ produced by the deterministic scheme (left) and particle-based method (middle), as well as the one-dimensional projection $\int_2 P(x_1, x_2) dx_2$ (right). The projection of many-body Wigner function seems to coincide with the single-body counterpart under the adiabatic approximation. The stochastic noises can be suppressed by either increasing sample size $N_0$ or refining the partition (choosing smaller $\vartheta$).
Fig. C.3. The 12-D proton-electron coupling: Visualization of the reduced electron Wigner function $W_1(x,k,t)$ from $t = 1\, \text{a.u.}$ to $5\, \text{a.u.}$, produced by the deterministic scheme (left), the 12-D stochastic simulations under $N_0 = 1 \times 10^8$, $\vartheta = 0.004$ (middle), as well as the reduced proton Wigner function $W_4(x,k,t)$ (right). The projection of many-body Wigner function seems to coincide with the single-body counterpart under the adiabatic approximation, while the proton is governed by the free advection.
Fig. C.4. The 12-D proton-electron coupling: Visualization of the reduced electron Wigner function $W_1(x,k,t)$ from $t = 6$ a.u. to 9 a.u., produced by the deterministic scheme (left), the 12-D stochastic simulations under $N_0 = 1 \times 10^8$, $\vartheta = 0.004$ (middle), as well as the reduced proton Wigner function $W_4(x,k,t)$ (right). The projection of many-body Wigner function seems to coincide with the single-body counterpart under the adiabatic approximation, while the proton is governed by the free advection.
Appendix D. Calculation of the star discrepancy of a sequence.

In principle, the star discrepancy can be attained by searching all the critical boxes with the upper coordinate $y$, with its dimension drawn from that of all possible coordinates in $X$. However, the calculation of the star discrepancy is in general a NP-hard problem, say, there might not exist an algorithm that can attain the maximal value with polynomial complexity [37].

Several algorithms based on integer optimization are proposed, including the threshold accepting [37], the genetic algorithm [38] and the improved threshold accepting (TA-improved) method [31]. In particular, the threshold accepting with improved sampling strategy, although somehow heuristic, is shown to be efficient in moderate large dimensional problem (especially $d = 20-60$).

The threshold accepting algorithm, often referred to a refined local search algorithm, is an integer optimization heuristic. For a set $X$ of $N$ points, the collection of critical boxes is denoted by $\tilde{\Gamma}(X)$, we set a total number of iterations $I$ and the number of independent trials $N_{tr}$ and a non-positive threshold value $T$. During the iterations, $T$ will increase until it reaches zero. This procedure helps to get rid of being trapped in local minima. More details, e.g., searching a neighbor by sampling and setting the threshold $T$, can be found in [31]. Instead, a series of benchmarks has been made for the dimensionality ranging from $d = 6$ to $d = 360$. Here $N$, $I$ and $N_{tr}$ denote the number of points, number of iterations in threshold accepting and count of independent trials, respectively. The points are generated by random sampling from uniform distribution on $[0,1]^d$. Since the algorithms intends to solve an optimization problem, the maximal value attained in the experiments can be regarded as the references.

All the results are listed in Tables (D.1) to (D.8), which uncover several facts.

1. For low-dimensional points, under a small iteration time and a few trial times, TA-improved algorithm can yield satisfactory results.
2. For high-dimensional points, the number of iterations $I$ must be sufficiently large to ensure the efficiency of searching. Too small $I$ (such as $I = 16, 32$), even under a large trial time $N_{tr}$ (such as $N_{tr} = 100$), fails to give reasonable approximations.
3. When $I$ fixed, the computational time for searching points increases moderately as $d$ increases. However, when $d$ fixed, the computational time grows almost exponentially as $I$ becomes larger.
4. For randomly distributed points, a boundary effect is observed. When $N$ is small and $d$ is large, the discrepancy turns out to be near 1.
### Table D.1
Discrepancies of 6-D points up to $N = 10^4$ are calculated by TA-improved algorithm. The points are drawn from uniform distribution on $[0, 1]^6$. 

| $N_{tr}$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|---------|--------|--------|--------|--------|--------|
| 5       | 2.82E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 |
| 10      | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 |
| 15      | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 |
| 20      | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 |
| 50      | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 |
| 100     | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 | 2.84E-01 |

**time per trial (s)**  
5.23E-04  1.70E-03  5.34E-03  2.04E-02  7.84E-02

$N = 10^2$, Maximal value is 2.84E-01.

### Table D.2
Discrepancies of 12-D points up to $N = 10^4$ are calculated by TA-improved algorithm. The points are drawn from uniform distribution on $[0, 1]^d$. 

| $N_{tr}$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|---------|--------|--------|--------|--------|--------|
| 5       | 6.02E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 |
| 10      | 6.02E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 |
| 15      | 6.02E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 |
| 20      | 6.02E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 |
| 50      | 6.02E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 |
| 100     | 6.02E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 | 6.45E-02 |

**time per trial (s)**  
3.02E-03  9.89E-03  3.71E-02  1.44E-01  5.77E-01

$N = 10^2$, Maximal value is 6.45E-02.
Table D.3

Discrepancies of 24-D points up to $N = 10^d$ are calculated by TA-improved algorithm. The points are drawn from uniform distribution on $[0, 1]^d$.

| $N$  | $N_r$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|------|-------|----------|----------|----------|----------|----------|
| 1    |       | 3.847E-01| 3.827E-01| 3.821E-01| 3.773E-01| 3.723E-01|
| 5    |       | 3.827E-01| 3.821E-01| 3.821E-01| 3.773E-01| 3.723E-01|
| 10   |       | 3.821E-01| 3.821E-01| 3.821E-01| 3.773E-01| 3.723E-01|
| 15   |       | 3.821E-01| 3.821E-01| 3.821E-01| 3.773E-01| 3.723E-01|
| 20   |       | 3.821E-01| 3.821E-01| 3.821E-01| 3.773E-01| 3.723E-01|
| 50   |       | 3.821E-01| 3.821E-01| 3.821E-01| 3.773E-01| 3.723E-01|
| 100  |       | 3.821E-01| 3.821E-01| 3.821E-01| 3.773E-01| 3.723E-01|

Time per trial (s): 2.42E-01, 1.87E-02, 6.84E-03, 2.26E-01.

Table D.4

Discrepancies of 36-D points up to $N = 10^d$ are calculated by TA-improved algorithm. The points are drawn from uniform distribution on $[0, 1]^d$.

| $N$  | $N_r$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|------|-------|----------|----------|----------|----------|----------|
| 1    |       | 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01|
| 5    |       | 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01|
| 10   |       | 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01|
| 15   |       | 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01|
| 20   |       | 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01|
| 50   |       | 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01|
| 100  |       | 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01| 4.887E-01|

Time per trial (s): 3.26E-01, 2.12E-02, 6.99E-03, 2.07E-01.
Table D.5

Discrepancies of 60-D points up to $N = 10^4$ are calculated by TA-improved algorithm.
The points are drawn from uniform distribution on $[0, 1]^d$.

| $N_{tr}$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|----------|----------|----------|----------|----------|----------|
| 5        | 4.60E-01 | 5.03E-01 | 5.19E-01 | 5.28E-01 | 5.28E-01 |
| 10       | 4.62E-01 | 5.08E-01 | 5.19E-01 | 5.28E-01 | 5.28E-01 |
| 15       | 4.84E-01 | 5.08E-01 | 5.26E-01 | 5.28E-01 | 5.28E-01 |
| 20       | 4.84E-01 | 5.08E-01 | 5.28E-01 | 5.28E-01 | 5.28E-01 |
| 50       | 4.84E-01 | 5.18E-01 | 5.28E-01 | 5.28E-01 | 5.28E-01 |
| 100      | 4.84E-01 | 5.18E-01 | 5.28E-01 | 5.28E-01 | 5.28E-01 |
| time per trial (s) | 2.38E-03 | 7.82E-03 | 2.99E-02 | 1.17E-01 | 4.63E-01 |

Table D.6

Discrepancies of 120-D points up to $N = 10^4$ are calculated by TA-improved algorithm.
The points are drawn from uniform distribution on $[0, 1]^d$.

| $N_{tr}$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|----------|----------|----------|----------|----------|----------|
| 5        | 1.39E-01 | 1.67E-01 | 1.90E-01 | 1.93E-01 | 2.01E-01 |
| 10       | 1.39E-01 | 1.67E-01 | 1.90E-01 | 1.94E-01 | 2.02E-01 |
| 15       | 1.39E-01 | 1.73E-01 | 1.90E-01 | 1.96E-01 | 2.02E-01 |
| 20       | 1.42E-01 | 1.73E-01 | 1.90E-01 | 1.96E-01 | 2.02E-01 |
| 50       | 1.42E-01 | 1.76E-01 | 1.92E-01 | 1.99E-01 | 2.02E-01 |
| 100      | 1.51E-01 | 1.76E-01 | 1.92E-01 | 2.00E-01 | 2.02E-01 |
| time per trial (s) | 1.61E-02 | 6.77E-02 | 2.55E-01 | 9.92E-01 | 3.91E-00 |

| $N_{tr}$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|----------|----------|----------|----------|----------|----------|
| 5        | 3.64E-02 | 4.21E-02 | 5.33E-02 | 5.83E-02 | 6.13E-02 |
| 10       | 3.64E-02 | 4.60E-02 | 5.34E-02 | 5.83E-02 | 6.13E-02 |
| 15       | 3.64E-02 | 4.60E-02 | 5.34E-02 | 5.83E-02 | 6.13E-02 |
| 20       | 3.64E-02 | 4.60E-02 | 5.34E-02 | 5.85E-02 | 6.13E-02 |
| 50       | 3.64E-02 | 4.60E-02 | 5.34E-02 | 5.85E-02 | 6.15E-02 |
| 100      | 3.60E-02 | 4.74E-02 | 5.34E-02 | 5.98E-02 | 6.16E-02 |
| time per trial (s) | 3.31E-01 | 1.01E-00 | 3.51E+00 | 1.25E+01 | 4.62E+01 |

| $N_{tr}$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|----------|----------|----------|----------|----------|----------|
| 5        | 6.30E-01 | 6.84E-01 | 7.36E-01 | 7.48E-01 | 7.67E-01 |
| 10       | 6.30E-01 | 6.95E-01 | 7.48E-01 | 7.67E-01 | 7.67E-01 |
| 15       | 6.30E-01 | 6.97E-01 | 7.48E-01 | 7.67E-01 | 7.67E-01 |
| 20       | 6.30E-01 | 7.15E-01 | 7.48E-01 | 7.67E-01 | 7.67E-01 |
| 50       | 6.45E-01 | 7.19E-01 | 7.51E-01 | 7.67E-01 | 7.67E-01 |
| 100      | 6.66E-01 | 7.19E-01 | 7.52E-01 | 7.67E-01 | 7.67E-01 |
| time per trial (s) | 4.49E-03 | 1.50E-02 | 6.06E-02 | 2.32E-01 | 9.13E-01 |

| $N_{tr}$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|----------|----------|----------|----------|----------|----------|
| 5        | 1.64E-01 | 2.12E-01 | 2.33E-01 | 2.47E-01 | 2.53E-01 |
| 10       | 1.64E-01 | 2.16E-01 | 2.37E-01 | 2.47E-01 | 2.53E-01 |
| 15       | 1.71E-01 | 2.16E-01 | 2.37E-01 | 2.47E-01 | 2.56E-01 |
| 20       | 1.71E-01 | 2.19E-01 | 2.37E-01 | 2.47E-01 | 2.56E-01 |
| 50       | 1.73E-01 | 2.19E-01 | 2.37E-01 | 2.52E-01 | 2.57E-01 |
| 100      | 1.77E-01 | 2.19E-01 | 2.37E-01 | 2.53E-01 | 2.57E-01 |
| time per trial (s) | 3.88E-02 | 1.31E-01 | 5.01E-01 | 2.02E+00 | 9.60E+00 |

| $N_{tr}$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|----------|----------|----------|----------|----------|----------|
| 5        | 4.03E-02 | 5.09E-02 | 6.91E-02 | 8.25E-02 | 8.73E-02 |
| 10       | 4.17E-02 | 5.76E-02 | 6.91E-02 | 8.25E-02 | 8.73E-02 |
| 15       | 4.17E-02 | 6.08E-02 | 7.15E-02 | 8.25E-02 | 8.73E-02 |
| 20       | 4.17E-02 | 6.08E-02 | 7.15E-02 | 8.27E-02 | 8.73E-02 |
| 50       | 4.18E-02 | 6.08E-02 | 7.15E-02 | 8.27E-02 | 8.73E-02 |
| 100      | 4.42E-02 | 6.12E-02 | 7.29E-02 | 8.30E-02 | 8.73E-02 |
| time per trial (s) | 9.00E-01 | 2.61E+00 | 1.02E+01 | 4.69E+01 | 1.15E+02 |
Table D.7
Discrepancies of 240-D points up to $N = 10^4$ are calculated by TA-improved algorithm.
The points are drawn from uniform distribution on $[0, 1]^d$.

| $N$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|-----|----------|----------|----------|-----------|-----------|
| 5   | 3.000E-01| 8.062E-01| 8.657E-01| 9.025E-01| 9.077E-01|
| 10  | 7.000E-01| 8.062E-01| 8.657E-01| 9.025E-01| 9.077E-01|
| 15  | 7.000E-01| 8.062E-01| 8.657E-01| 9.025E-01| 9.077E-01|
| 20  | 7.000E-01| 8.062E-01| 8.775E-01| 9.025E-01| 9.077E-01|
| 50  | 7.159E-01| 8.169E-01| 8.856E-01| 9.070E-01| 9.077E-01|
| 100 | 7.225E-01| 8.265E-01| 8.947E-01| 9.070E-01| 9.077E-01|

Time per trial (s) 9.186E-03 3.157E-02 1.213E-01 4.870E-01 1.914E+00

Table D.8
Discrepancies of 360-D points up to $N = 10^4$ are calculated by TA-improved algorithm.
The points are drawn from uniform distribution on $[0, 1]^d$.

| $N$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|-----|----------|----------|----------|-----------|-----------|
| 5   | 2.462E-01| 3.076E-01| 3.521E-01| 3.689E-01| 3.773E-01|
| 10  | 2.462E-01| 3.076E-01| 3.521E-01| 3.701E-01| 3.788E-01|
| 15  | 2.662E-01| 3.086E-01| 3.569E-01| 3.701E-01| 3.791E-01|
| 20  | 2.662E-01| 3.086E-01| 3.569E-01| 3.701E-01| 3.791E-01|
| 50  | 2.702E-01| 3.091E-01| 3.569E-01| 3.715E-01| 3.791E-01|
| 100 | 2.702E-01| 3.091E-01| 3.569E-01| 3.715E-01| 3.791E-01|

Time per trial (s) 9.448E-02 3.128E-01 1.149E+00 4.562E+00 1.789E+01

Table D.9
Discrepancies of 360-D points up to $N = 10^4$ are calculated by TA-improved algorithm.
The points are drawn from uniform distribution on $[0, 1]^d$.

| $N$ | $I = 16$ | $I = 32$ | $I = 64$ | $I = 128$ | $I = 256$ |
|-----|----------|----------|----------|-----------|-----------|
| 5   | 6.816E-02| 8.787E-02| 1.143E-01| 1.324E-01| 1.444E-01|
| 10  | 6.816E-02| 8.787E-02| 1.143E-01| 1.332E-01| 1.448E-01|
| 15  | 6.816E-02| 8.787E-02| 1.143E-01| 1.332E-01| 1.448E-01|
| 20  | 6.816E-02| 9.158E-02| 1.143E-01| 1.332E-01| 1.448E-01|
| 50  | 7.070E-02| 9.381E-02| 1.143E-01| 1.332E-01| 1.451E-01|
| 100 | 7.070E-02| 9.381E-02| 1.143E-01| 1.332E-01| 1.451E-01|

Time per trial (s) 2.772E+00 1.004E+01 4.226E+01 1.348E+02 4.997E+02

53