Method to manage integration error in the Green-Kubo method

Laura de Sousa Oliveira* and P. Alex Greaney†
Mechanical Engineering Department, University of California, Riverside, California, USA
(Received 25 September 2016; revised manuscript received 20 December 2016; published 21 February 2017)

The Green-Kubo method is a commonly used approach for predicting transport properties in a system from equilibrium molecular dynamics simulations. The approach is founded on the fluctuation dissipation theorem and relates the property of interest to the lifetime of fluctuations in its thermodynamic driving potential. For heat transport, the lattice thermal conductivity is related to the integral of the autocorrelation of the instantaneous heat flux. A principal source of error in these calculations is that the autocorrelation function requires a long averaging time to reduce remnant noise. Integrating the noise in the tail of the autocorrelation function becomes conflated with physically important slow relaxation processes. In this paper we present a method to quantify the uncertainty on transport properties computed using the Green-Kubo formulation based on recognizing that the integrated noise is a random walk, with a growing envelope of uncertainty. By characterizing the noise we can choose integration conditions to best trade off systematic truncation error with unbiased integration noise, to minimize uncertainty for a given allocation of computational resources.

DOI: 10.1103/PhysRevE.95.023308

I. INTRODUCTION

Transport properties are ubiquitous in materials science and engineering. Heat sinks and thermal barrier coatings are two obvious examples where thermal conductivity is paramount for materials' performance, but there are also a huge number of materials applications in which transport properties are folded in with a number of other properties to dictate performance. Nanofluids are a promising new material for numerous applications [1–3] that include heat dissipation [2,3] for which, in addition to thermal transport, viscosity calculations are necessary to better our understanding of heat transfer mechanisms. Moreover, the rheological characterization of fluid materials has numerous engineering applications beyond cooling (e.g., lubrication [4], sheathing [5], or hydraulics [6]), as well as applications in other fields (e.g., medicine [7], geophysics [8]). Viscous ionic electrolytes in batteries are an example where viscosity, diffusion, and ionic conductivity [9] all play an important role in the materials' eventual performance. In short, the ability to reliably predict transport properties is essential in the search for new materials for a wide variety of applications. Molecular dynamics (MD) simulations provide a powerful approach for quickly obtaining atomistic level insight into the physics of mass, momentum, or energy transport processes in materials. Two approaches are possible: MD can be used to (1) simulate systems in equilibrium or (2) perturb and drive systems out of equilibrium to then measure their response.

Equilibrium molecular dynamics (EMD) calculations are performed using the well-established Green-Kubo formalism [10,11], which relates transport quantities to the duration of fluctuations in a microscopic state of the system—the underlying principle is that the processes that dissipate small local fluctuations are the same that are responsible for a material's feedback to a stimulus. Mathematically this is achieved by integrating the current autocorrelation function, as is shown in the general expression for the Green-Kubo method:

$$\gamma = \alpha \int_0^\infty \langle \mathbf{A}(t)\mathbf{A}(t+\tau)\rangle \, d\tau, \tag{1}$$

where γ is the transport property of interest and $\mathbf{A}$ is the current that drives it. The expression $\langle \mathbf{A}(t)\mathbf{A}(t+\tau)\rangle$ is the autocorrelation function of quantity $\mathbf{A}$ and α is a temperature-dependent coefficient. For instance, for thermal conductivity, κ, the Green-Kubo expression becomes

$$\kappa = \frac{V}{3k_B T^2} \int_0^\infty \langle \mathbf{J}(t)\mathbf{J}(t+\tau)\rangle \, d\tau, \tag{2}$$

where k_B is Boltzmann's constant, T is the temperature, V is the volume of the simulated region, $\mathbf{J}$ is the heat flux, and $\langle \mathbf{J}(t)\mathbf{J}(t+\tau)\rangle$ is the non-normalized heat current autocorrelation function (HCACF). This method is widely used by materials scientists, chemists, and physicists. In addition to thermal conductivity calculations [12–15], it has been used to calculate viscosity [4,8,16], diffusivity [17,18], and ionic conductivity [9] for a wide range of materials, by integrating the pressure tensor, velocity, and ionic flux autocorrelation functions (ACFs), in that order.

There are clear advantages to using an equilibrium approach: while both equilibrium and nonequilibrium methods suffer from size artifacts, the use of periodic boundary conditions in EMD allows for a smaller system size; for anisotropic systems, one EMD simulation suffices to compute the full transport tensor; and EMD can be used irregardless of the linearity of the transport regime with system size. There is, however, also one major pitfall. Fully converging the autocorrelation function requires very long simulation times and often a compromise has to be made between including the contribution of slow processes and introducing a random error, or excluding these processes and introducing a systematic truncation error. In this paper, by recognizing that the integrated ACF error mimics a random walk, we propose a method that allows researchers to evaluate this trade-off on the fly and make better informed decisions about where to truncate the ACF and how to optimize computational resources. In the remainder of the paper, we will focus exclusively on thermal

*Also at the Mechanical Engineering Department, University of California, Riverside, CA, USA.
†agreaney@engr.ucr.edu

transport. It is left for the reader to draw the obvious parallels with other transport properties. The next paragraphs concern the origin of the oscillations, existing approaches to integrate the autocorrelation function, and the introduction of the concept of a random walk in the HCACF. Our proposed method and its implementation to an example data set are described next, followed by the discussion and conclusion remarks.

A. The oscillatory behavior of the autocorrelation function

The HCACF, $\langle \mathbf{J}(t)\mathbf{J}(t+\tau)\rangle$, can be numerically computed as

$$\langle \mathbf{J}_n \mathbf{J}_{n+m}\rangle \equiv \sum_{n=0}^{N-m} \frac{\mathbf{J}_n \mathbf{J}_{n+m}}{N-m}, \tag{3}$$

where $\mathbf{J}_n$ is the value of $\mathbf{J}$ at the nth time step, i.e., $\mathbf{J}_n = \mathbf{J}(t_n)$, for $n = 0,1,2,\ldots,N$, and $\mathbf{J}_{n+m}$ is $\mathbf{J}$ at the $(n+m)$th time step, or $\mathbf{J}(t_n+\tau_m)$, for $m = 0,1,2,\ldots,M$. N and M are, respectively, the maximum numbers of steps in the simulation and in the HCACF. Analytically, the autocorrelation function is computed as the inverse Fourier transform of the same transform of the current multiplied by its complex conjugate, averaged over $N-m$. It follows that to obtain good statistical averaging M must be significantly less than N, and that the error associated with the HCACF increases over time for fixed N. This is applicable to other transport properties. For a system in equilibrium, the average current of any property is zero, and the ACF is expected to decay to zero given sufficient time. Instead, large oscillations with a significant contribution to the integral have been observed [15,19–22]. Figure 1(a) depicts an example of fluctuating HCACFs and the growing error in the corresponding integrals, and Fig. 1(b) shows the longevity of the fluctuations.

If we were able to sample an infinite system for infinite time, we should find the system's *true* ACF and thus a fixed true transport quantity. It follows that, for the thermal transport example we have been using, κ, computed with Eq. (2), is

$$\begin{aligned}\kappa &= \kappa_{\text{true}} \pm \Delta\kappa \\ &= \alpha \lim_{t\to\infty} \int_0^{\infty} \langle \mathbf{J}(t)\mathbf{J}(t+\tau)\rangle \, d\tau \pm \alpha \int_0^{\tau_{\max}} \eta(\tau)\, d\tau, \end{aligned} \tag{4}$$

where $\tau_{\max}$ is the maximum time for which the HCACF is computed, and $\alpha = \frac{V}{3k_B T^2}$. The first term in the equation is the *true* integrated HCACF, and the second term is the integral of the HCACF noise that comes about due to insufficient averaging. As shall be discussed more thoroughly in due course, at least two sets of different frequency oscillations can be distinguished that mirror the fast and slow fluctuations in the heat current.

Accurately predicting the ACF is critical for transport predictions using the Green-Kubo method. Notwithstanding, there is little consensus in the literature as to what approach to take to mitigate the noise, and the cumulative quality of the integrated noise has seldom been used to inform the choice of ACF integration approach. The next paragraphs reference some of the most common ACF integration approaches and a few less common strategies found in the literature. While it has been shown that the Green-Kubo approach can be successfully used with quantum-based calculations [23,24], simulation size and length present a major difficulty in using EMD approaches within *ab initio*, and other methods [25,26] continue to offer greater advantages. However, as computers become faster, density functional theory (DFT) MD transport calculations could become more common, and error estimation more important. Within classical MD, the evolution of computing means averaging large enough systems for longer will become less of an issue, thus reducing or even eliminating the error from these calculations. However, there is an increasing trend to develop high-throughput approaches for the rapid screening of materials, which in turn require quick, on-the-fly

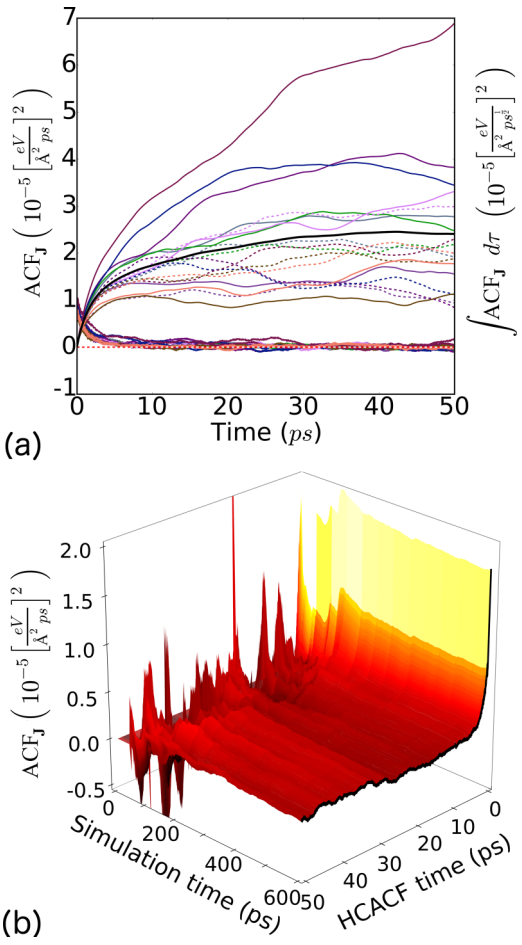

FIG. 1. Panel (a) shows the HCACFs (the decaying functions) plotted along side their integrals (the curves that rise to a plateau) computed from nine separate simulations of a 10 648-atom, perfectly crystalline, and periodically contiguous block of graphite. The data were taken from a study to determine the influence of Wigner defects on thermal transport in graphite [22]. The dashed lines correspond to the heat flux along the $[2\bar{1}\bar{1}0]$ direction and the solid lines correspond to the heat flux along $[01\bar{1}0]$. The system was found to be converged for size, and κ is expected to be the same in both directions along the basal plane. This plot illustrates the increasingly diverging noise of the HCACF integrals, present even after 50 ps. To the eye, the ACFs look nicely converged after 10–15 ps. Plot (b) shows the gradual convergence of the HCACF with increasing averaging time during a single simulation. The amplitude of the fluctuations in the tail of the HCACF decays over time, but it is notable that continued averaging does not remove the pattern of the fluctuations.

approaches for uncertainty quantification. The method introduced herein meets these requirements.

B. Common autocorrelation function integration approaches

A common strategy to reduce the noise in the ACF is to fit an exponential to the first few picoseconds ($\tau < 10$) [20,21]. The system depicted in Fig. 1 exhibits a rapid decay associated with high-frequency phonons and a slower decay associated with lower frequency phonons; similar two- or three-stage decay is observed in many single-element materials and different authors have modeled κ by fitting the HCACF to the sum of two or more exponentials [20,21,27]. This approach captures multiple relaxation processes and is therefore more physically meaningful than a single exponential fit, but it is ineffective when the HCACF cannot be represented by an exponential fit [12,22,28] and it forces a behavior description of the HCACF that might not be accurate. The same is true of shear relaxation times in viscosity calculations. For ionic liquid calculations, authors have also fit the pressure tensor autocorrelation function to Kohlrausch's law [29,30] and/or applied weighing factors to their fits [31,32]. Fits to the frequency domain are also a solution, depending on the resulting ACF for given data [12,28]. Some strategies include direct integration of the ACF truncated to various cutoffs. Whether direct integration is performed or a fit is applied, the cutoffs are oftentimes arbitrarily selected [33–35]. They can also be more systematically determined, for instance by taking the running mean of the integrated autocorrelation at its plateauing region [36,37]. Recently, Chen *et al.* have proposed a noise-sensitive mathematical approach: to truncate the HCACF when the scale of the fluctuations becomes the same as the mean, i.e., when $|\frac{\sigma}{E}| > 1$, where σ is the standard deviation and E is the expected value of the HCACF in an interval $(\tau, \tau + \delta\tau)$ [38]. Chen *et al.* further suggest including a fixed offset term, Y_0, to the exponential fitting approach (e.g., $A_1 e^{-\tau/t_1} + A_2 e^{-\tau/t_2} + Y_0$) to the normalized HCACF. In a study concerning thermal transport in irradiated graphite, we implemented and compared this and other methods [22]. The method of Chen *et al.* is a useful, systematic approach, but it neglects the growing nature of the uncertainty that results from integrating over the noise. Other approaches that acknowledge the incremental error of the HCACF integral have been proposed [31,39]. For instance, Zhang *et al.* [31] use a time-decomposition method to compute a growing standard deviation to which they suggest fitting a power law, and from which a cutoff can be selected based on a desired percentage error. With the insight gained from the graphitic systems studied, we develop here another approach to quantify and mitigate the noise introduced with the Green-Kubo. This approach is based on recognizing that the ACF fluctuations around zero integrate into Brownian noise; i.e., for each simulation a random walk is effectively added to the integral of the *true* ACF. Before proceeding, it is perhaps useful to briefly introduce the notion of a random walk and how it relates to the noise in the HCACF.

C. Random walk

A random walk is a succession of Markovian (uncorrelated) random steps. This has the property that the expected root mean square (rms) displacement after N steps is $\langle x_N \rangle = \sigma_d \sqrt{N}$, where σ_d is the standard deviation of the magnitude of the steps (i.e., the displacement). Here we argue that the noise in the HCACF has the statistical properties of a stream of uncorrelated fluctuations or excursions from zero. Although these fluctuations have a characteristic duration, the time integral of a fluctuation equates to one jump in a random walk. If one determines the time scale over which the HCACF noise is uncorrelated (the jump frequency δt) and the typical integrated excursion (jump magnitude, d) then one can equate the accumulation of noise integration error to the rms displacement of the equivalent random walk. The equivalence of the HCACF to a stream of uncorrelated fluctuations that when integrated yield a random walk is demonstrated in Figs. 2(a)–2(c). In these simulations, the average step size is $\sigma_v \delta t$, where σ_v is the standard deviation of the noise *velocity* $(d/\delta t)$, i.e., the *velocity* at which the random walk occurs through time. The standard deviation of the *velocity* (σ_v) is effectively that of the steps. The total number of Markovian steps over time t is $N = t/\delta t$, and so the expected uncertainty $U(t)$ after integrating to time t is given by

$$U(t) = \delta t \sigma_v \sqrt{\frac{t}{\delta t}} = \sigma_v \sqrt{t \delta t}. \tag{5}$$

In this relationship computing σ_v is straightforward, and so the remaining challenge is to determine the uncorrelated fluctuation time δt.

By characterizing the integrated HCACF noise as a random walk, or as a sum of random walks, in terms of δt and σ_v, we propose that one can use Eq. (5) to compute an *uncertainty envelope* that informs on how quickly the integrated noise error in a single simulation grows. From the uncertainty envelope of a *single* simulation one can compute the expected uncertainty in the average of any number of simulations. The crucial point is that information about the distribution of error in many simulations can be obtained from a first, short (a few hundred picoseconds) simulation, and thus after the first simulation has been performed, one can decide on an optimal computational strategy for minimizing uncertainty.

Upon quick inspection, the HCACFs shown in Fig. 1 appear to be converged by 20 ps. Figure 2(e) shows the result of integrating random fluctuations in the 20–50 ps interval of the HCACF tail. To parallel Figs. 2(a) and 2(b), which depict an example of fluctuations [in Fig. 2(a)] that give rise to a random walk [in Fig. 2(b)], a single HCACF tail is depicted in Fig. 2(d), but the integrals of 18 HCACFs' tails are plotted in Fig. 2(e). The noise in this data [Fig. 2(b)] is *not* uncorrelated from point to point along the data stream but instead has some memory of itself. To predict the uncertainty from this noise we must compute the lifetime for this memory to find the time scale at which the noise becomes uncorrelated. Instead of a jump (or walk) at every interval in the autocorrelation, jumps are better described by (some of) its peaks [see the line in magenta in Fig. 2(d)]. The distribution in Fig. 2(f) corresponds to the compound HCACF tails for the 18 simulations. Figure 2(g) was obtained from the peaks as exemplified in Fig. 2(d). A normal distribution with the standard deviation for each case and mean zero is shown in red, and the distributions with the correct mean are in black and magenta for the whole set of tails

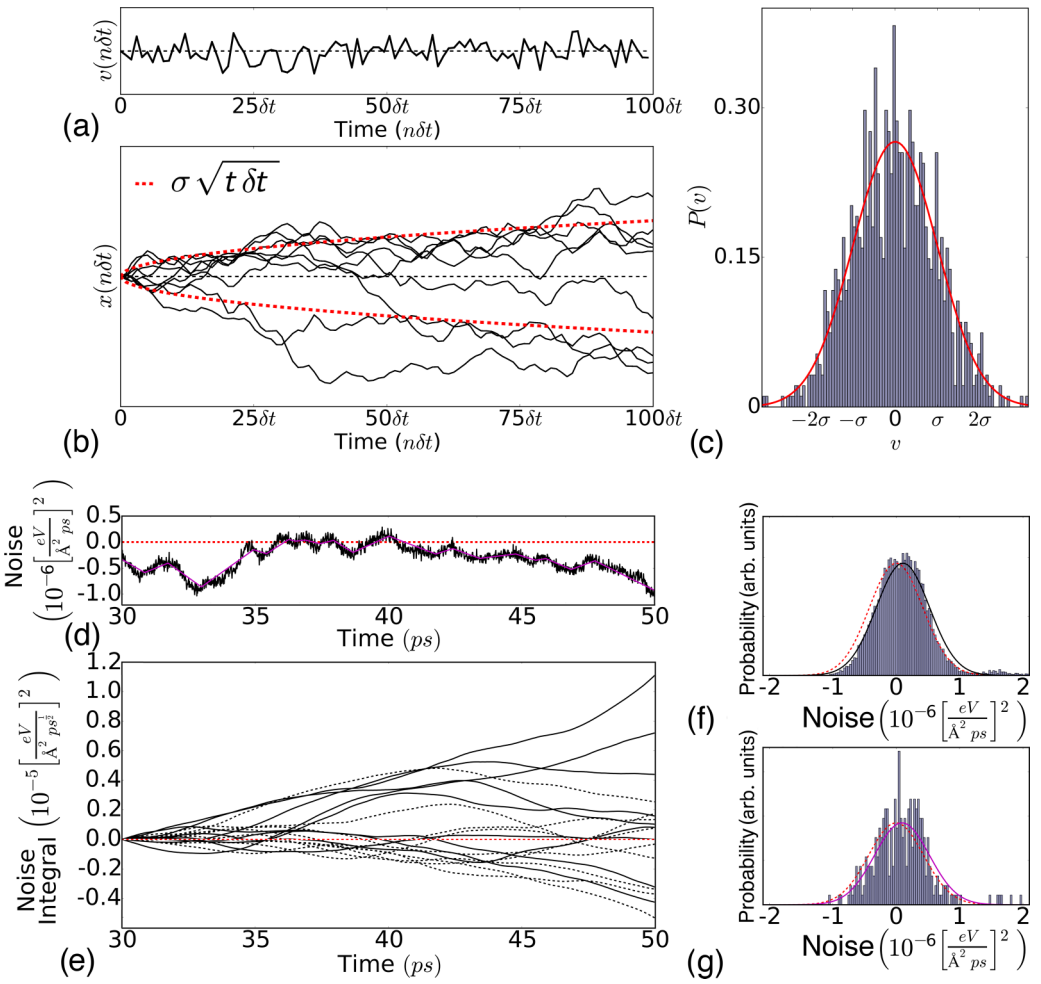

FIG. 2. Panel (a) corresponds to the step or velocity fluctuations that give rise to a random walk; in panel (b) a set of 10 random walks is shown in black and the expected root mean square translation distance at time t is plotted in red; and panel (c) is the distribution of the random walks shown in panel (b). Panel (d) corresponds to the tail of a HCACF, depicting the noise fluctuations that integrate to a large error akin to a random walk, shown in panel (e) for all HCACF tails. The black lines correspond to heat-flux measurements along the x direction ($[2\bar{1}\bar{1}0]$), and the blue ones are along the y direction ($[01\bar{1}0]$). Both values were measured along the basal plane, and this distinction should not matter. The data set is explained in the Methods section. For the selected 20–50 ps interval, the distribution of all data points across the multiple simulation tails is shown in panel (f). A 1-ps moving average was used along with a peak find algorithm to plot major peaks in the HCACF tails, as shown in panel (d), in magenta. The peak distribution for all data is offered in panel (g). The dashed red lines in panels (f) and (g) correspond to a normal distribution with the standard deviation of each of the distributions and mean zero. A normal distribution with the mean for each of the data sets is shown in the solid lines for each case.

and peaks, respectively, in Figs. 2(f) and 2(g). The distributions will again be addressed in the Results section.

The method developed to quantify the uncertainty that results from the Green-Kubo approach by treating the noise in the autocorrelation function as a random walk is introduced in the Methods section, but not before a more detailed explanation of the data set used for Figs. 1 and 2 is offered.

II. METHODS

All simulations used to perform error analysis were obtained with the large-scale equilibrium classical molecular dynamics software LAMMPS [40], using the adaptive intermolecular reactive empirical bond-order (AIREBO) potential function formulated by Stuart *et al.* [41]. The simulations correspond to a size-converged $11 \times 11 \times 11$ perfectly crystalline graphite supercell with 10 648 atoms and a $27.05 \times 46.86 \times 73.79$ Å^3 volume in the x, y, and z directions, respectively. Previous work has shown that this system is large enough to be size converged for thermal conductivity [22]. We use data from nine simulations that were relaxed and equilibrated in the microcanonical ensemble (NVE), using a standard Velocity-Verlet quadrature scheme, for 50 ps after being given a thermal energy equivalent to 300 K before starting to record the HCACF. Each of the nine runs was simulated for an additional 0.6 ns with a 0.2-fs time step and periodic boundary conditions. Because κ can be computed in all lattice directions from a single simulation using the Green-Kubo formalism, there are 18 HCACFs along the basal plane of the graphite supercell with which to perform data analysis (nine each along x and y, that is, $[2\bar{1}\bar{1}0]$ and $[01\bar{1}0]$). These data were obtained for a previous publication on the thermal conductivity of irradiated

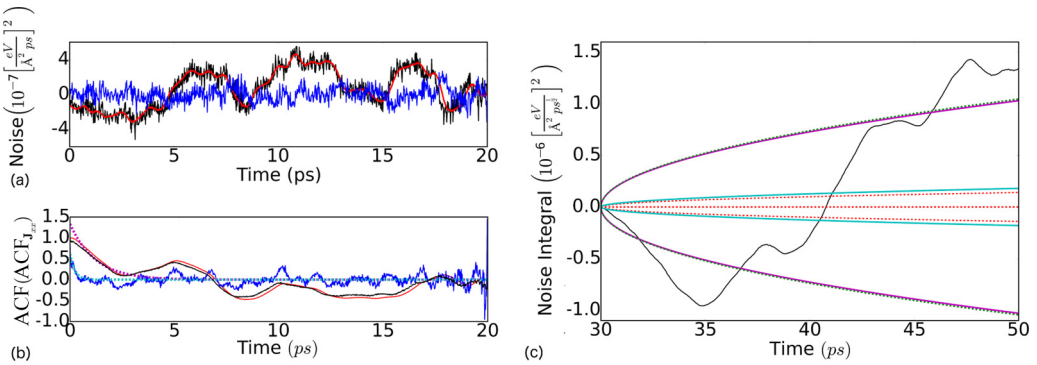

FIG. 3. The noise of a HCACF tail in the 30–50 ps interval is shown (a) decomposed into high-frequency (blue) and low-frequency (red) noise. The autocorrelations of the noise (black) and the high-frequency (blue) and low-frequency (red) components of the noise are shown in panel (b), along with fits through the high-frequency (cyan) and the low-frequency (magenta) autocorrelations. In panel (c) the integrated tail appears in black and the uncertainty envelope for δt equal to the interval of the HCACFs is shown in dashed red; the uncertainty envelopes corresponding to the high-frequency and low-frequency noise are in cyan and magenta, respectively. The dashed black line that follows along the magenta is the combined uncertainty envelope of the high- and low-frequency noises, i.e., the square root of the sum of their squares.

graphite [22]. A longer 8.0-ns simulation with a 0.4-fs time step was also performed, under the same conditions. Based on the premise that the noise of the integrated HCACF is akin to a random walk, we can use Eq. (5) to compute the root mean squared of the noise integrated up to time $\tau_{\max}$. This is the expected deviation (or error) from the mean for each random walk, and we can thus compute the standard deviation of said error at time $\tau_{\max}$ in an average of N random walks, with the same characteristic δt and σ_v, as $S_N = \sigma_v\sqrt{\frac{\tau_{\max}\delta t}{N}}$.

Decomposing the noise into uncorrelated fluctuations is the first step required to discern between a single random walk or the sum of varying frequency random walks. Then, to characterize the random walks, one must determine the standard deviation of these fluctuations and the average interval between them. If the random fluctuations occurred at the same interval that the HCACF is recorded, the expected noise uncertainty envelope would be as indicated in Fig. 3(c), in the dashed red line. This largely underestimates the integrated noise. A moving average low-pass filter with a 0.4-ps window applied to the noise reveals that at least two distinct sets of noise frequencies are present [see Fig. 3(a)]. This indicates that instead of a single random walk with the same time step as that of the HCACF, the noise is best described by the sum of different frequency random walks. Finding the contribution of each random walk to the expected error can be difficult, but a series of frequency passes (see Fig. 4) can help examine the contribution of varying frequencies in the noise to the expected error. The subsequent analysis is performed with the separate sets of noise identified as having the largest contribution to the expected error and shown in Fig. 3(a). While the noise behaves similarly to a random walk, the system has a memory of itself and the fluctuations should be correlated with each other. The correlation time obtained from the autocorrelation function of the noise gives the average time interval, δt, at which the fluctuations are Markovian. This method is applied to a single simulation as detailed in the following steps, with the aid of Fig. 3:

(i) The first step is to isolate the noise from the data. This is easily done by selecting a portion of the tail; if it is clear the HCACF is converged after some time. Otherwise, a fit could be used to extract the noise. Using the tail of the HCACF to analyze the noise is generally preferable to using a fit, as it removes the uncertainty that arises from guessing the behavior of the HCACF. The choice of interval (30–50 ps) to characterize the noise is explained in the Results section.

(ii) The second step is to filter the noise for different frequencies. This step is exemplified in Fig. 3(a). A low-pass filter allows us to distinguish two main sets of oscillations, in red and in blue. While only one pass, separating frequencies below and above 2.5 THz, is illustrated in Fig. 3(a), more could be applied (see Fig. 4) to gain a better understanding of the noise. This is discussed more thoroughly in the Results section. The contribution of each set of data is considered as described next.

(iii) The third step consists in computing the autocorrelation of the different frequency noise components. For the low- and high-frequency noise found in step (ii) and depicted in

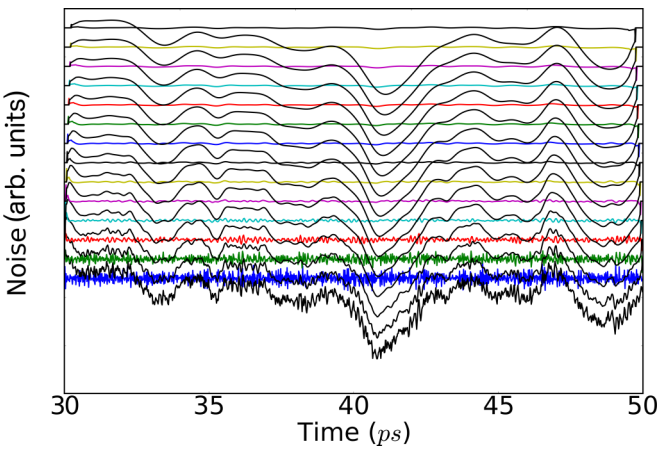

FIG. 4. This graph shows the application of multiple pass filters to isolate existing frequencies in the HCACF noise. The first filter applied selects out data below a 0.04-ps interval (the blue high-frequency line at the bottom of the graph) and leaves the remaining frequencies. The next filter has a 0.08-ps window and is used to filter the low-frequency data remnant from the first pass. This procedure is performed for 0.04-ps intervals up to a filter with a 0.56-ps window.

Fig. 3(a), the ACFs are shown in red and blue, respectively, in Fig. 3(b).

(iv) The fourth step is to fit a single exponential $a_i e^{-\frac{t}{\tau}}$ to each of the above autocorrelations. The fits are shown in magenta and cyan, for the low- and high-frequency cases, in that order. The fitting parameter τ provides an estimate of the interval of our near-random walk noise. The autocorrelation of the low-frequency noise (in red) is comparable to that of the whole system (in black). It is already clear that the contribution of the low-frequency HCACF noise explains most of the random walk uncertainty.

(v) The fifth step is to compute the standard deviation, σ, of each of the noise contributions.

(vi) The sixth and final step is to compute the uncertainty envelope by using the calculated τ and σ in Eq. (5). In Fig. 3(c), the magenta uncertainty envelope corresponds to the low-frequency oscillations, and the cyan envelope corresponds to the contribution of the high-frequency noise. As anticipated, the high-frequency noise envelope is not much greater than the envelope calculated with the HCACF interval (in dashed red). The combined error of high- and low-frequency noise (in dashed black) is barely distinguishable from that of the low-frequency noise (in magenta). As expected, the contribution of low-frequency oscillations largely explains the noise.

τ and σ are all that is necessary to characterize the random walk. This means a simulation could be undergoing and its data used to evolve the uncertainty envelope on the fly. An example of this is shown in the results. For the present data set, the low-frequency oscillations explain nearly all of the noise, and it would suffice to consider the autocorrelation of the whole, unfiltered noise, to obtain an estimate for the integrated noise envelope. A more thorough discussion of the filtering is offered in the Results section. Also in the Results section, this approach is applied to the 18 HCACFs, thus allowing us to obtain an error estimate of the *uncertainty envelope*. We also show that a frequency decomposition analysis similar to that applied to the HCACF can be used directly on the heat flux to determine a suitable simulation time step to optimize HCACF convergence.

III. RESULTS

We applied steps (i)–(vi) to all HCACFs. The second step involves identifying different noise frequencies. It is worthwhile to remark on the difficulty of extricating individual random walks from a sum of random walks. For instance, applying a filter (as in Fig. 4) can syphon out data that belongs to a lower frequency random walk. In Fig. 4, frequency filters are applied with windows ranging between 0.04 and 0.56 ps at a 0.04-ps interval. Each time, the data filtered are removed from the overall noise. One might be tempted to say, from evaluation of Fig. 4, that there are multiple high-frequency random walks, with time fluctuations $\tau = 0.04$, 0.08, and 0.12 ps, for instance, and that might be correct or the sets of filtered data might belong to a single random walk. If the former is true, the contribution of the independent sets of high-frequency data were calculated to be negligible compared to the low-frequency data, in the same way the high-frequency data obtained with a single (0.4-ps) filter, as shown in Fig. 3(a), does not significantly contribute to the overall noise [see Fig. 3(c)]. Similarly, the low-frequency noise could be considered as the sum of its parts, but this would remove the underlaying characteristics of the noise. For this reason, having identified distinct frequency ranges in the noise, and having determined that their contribution is remarkably unequal, we proceed with the analysis performed as described in steps (i)–(vi).

For all simulations, τ was computed as to minimize the standard deviation, with the caveat that the maximum allowed value for τ was limited by the lowest intercept with zero between all noise autocorrelation functions. This is because we fit to the natural logarithm of the noise autocorrelation. This does befit us, however, in that we aim to calculate the effect of the fast rate of decay of the systems' memory reflected in the noise. Moreover, a similar argument to there being a true autocorrelation function for the heat flux can be made with regards to the noise. If the frequency of the noise is the same across samples, there is one true autocorrelation function that describes the interval for which the noise is correlated,

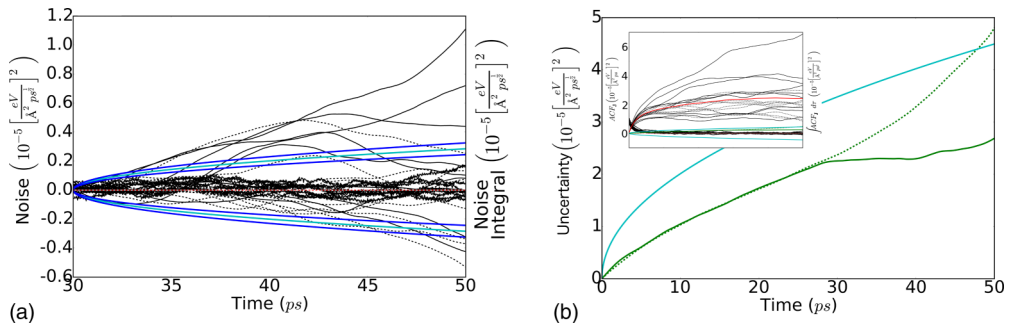

FIG. 5. In panel (a) the tail of the HCACFs, their integral, the uncertainty envelope (cyan) calculated as described in the text, and its error (blue) are all plotted. In the inset in panel (b), instead of only considering the noisy tails of the HCACFs, the whole HCACFs are represented. In both panel (a) and the inset in panel (b) the solid black lines correspond to results along the y direction, and the dashed black lines correspond to results along the x direction. The bold red line in the inset in panel (b) is the integral of the average of the HCACFs; the solid green line is the standard error computed for the 18 HCACF integrals; and the dashed green line is the standard error of the 216 50-ps HCACF integrals that can be obtained from the 18 sets of data with 600 ps each. These lines are shown in the inset in panel (b) for perspective, but also in the larger plot in panel (b) for a clearer distinction between them and the cyan line, which shows the uncertainty calculated as described in the text, using the random walk approach.

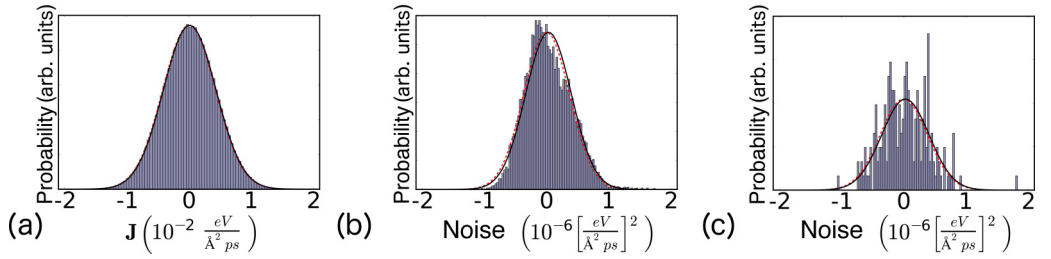

FIG. 6. Panel (a) is the normal distribution over all **J**. Panel (b) is the distribution of the noise from the tails in the 30–50 ps interval. Panel (c) is the distribution of the peaks fit to the noise from the tails in the 30–50 ps interval, as shown in Fig. 2(d).

i.e., before it becomes random. For the high-frequency noise, $\tau_H = 0.27 \pm 0.02$ ps and is one order of magnitude greater than the interval of the HCACF ($\delta t = 0.02$ ps), but, as depicted in Fig. 3(c) for the calculated uncertainty envelope of a single HCACF tail, it has a low impact in the overall uncertainty envelope. For the low-frequency noise, τ_L is 4.6 ± 0.78 ps. The standard deviation for the high-frequency noise, σ_H, is $8.06 \pm .11 \times 10^{-8}\,\mathrm{eV}^2/\mathrm{\AA}^4\,\mathrm{ps}^2$ and for the low-frequency noise σ_L is $2.89 \pm 16 \times 10^{-7}\,\mathrm{eV}^2/\mathrm{\AA}^4\,\mathrm{ps}^2$. Figure 5(a) shows how the noise integrals compare to the envelope (in cyan) computed from the mean τ_L and σ_L obtained from the 18 HCACF tails, using Eq. (5), including the error (in blue) obtained by propagating the standard error of each quantity; the above stated uncertainties for τ_H, τ_L, σ_H, and σ_L are the standard error. In Fig. 5(b) in the inset the envelope is compared with the full HCACF integrals. The standard error computed over of the 18 HCACF integrals is also depicted in Fig. 5(b) (in solid green), including in the inset, as is the standard error computed over the set of 216 sets of 50-ps HCACF integrals to which the 18 sets can be reduced (in dashed green) by splitting each 600-ps set of **J** values in 12 sets of 50 ps. This method of splitting the heat current data into many small parcels and computing the HCACF independently for each parcel means that the individual HCACF's are more noisy, but there are more data sets from which to infer the standard error in the integral. This method predicts an uncertainty slightly smaller that the random walk method. The approach is appealing because it is simple and it appears to provide a narrow estimate of uncertainty. Unfortunately, the tails of HCACFs computed from neighboring data windows are found to be correlated and so the approach underestimates the error, providing a false degree of certainty. It can be seen in Fig. 5(b) that nearing 30 ps the error defined as the standard error of the HCACF integrals becomes more ill defined. Again, this is because over time each of the HCACFs has less data to average over. The possibility that **J** is still correlated after the length of the HCACF implies that, unlike the method proposed herein, a correct noise estimate with the standard error approach requires multiple simulations with differing starting points. As seen in Fig 10, with the *random walk* approach a few hundred picoseconds suffice to characterize the error and obtain an *uncertainty envelope*.

In Figs. 2(f) and 2(g) it can be observed that for the 20–50 ps interval selected the HCACF tails have a nonzero mean. This suggests that the HCACFs might not have been fully relaxed by 20 ps. In Fig. 6 we consider the distributions of **J** [Fig. 6(a)], the noise in the 30–50 ps interval for the entire data [Fig. 6(b)], and for the case where the peaks are computed from a moving average with a 1-ps interval [Fig. 6(c)] as shown in Fig. 2(d). Figures 6(b) and 6(c) correspond to Figs. 2(e) and 2(f) for the smaller interval. Figure 6 reassures us that over all simulations the system is close to relaxed by 30 ps. However, not all individual simulations seem to have converged by 30 ps. While the distribution of **J** for each simulation reveals a consistently normal distribution with mean zero, the mean of the distribution of individual HCACF tails fluctuates around but is not consistently at zero. This is not an issue because the random walk approach to estimate the uncertainty of the Green-Kubo method is largely insensitive to prevailing steady deviations from zero and it considers these variations as real slow decay processes.

Figure 7 evidences that the random walk method is robust to slow decay processes affecting the characterization of the

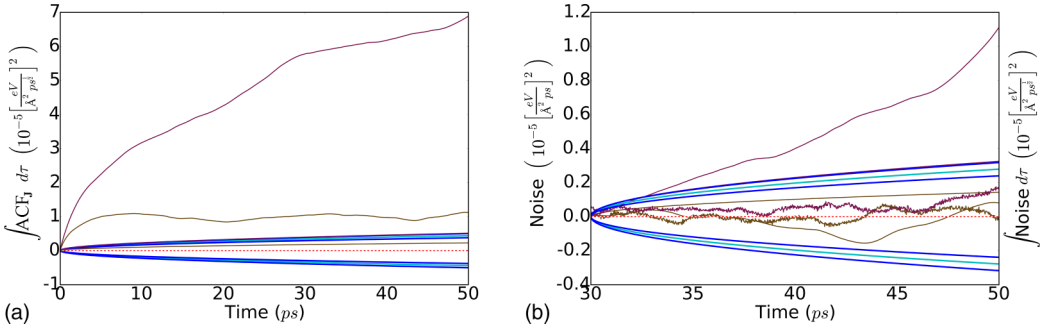

FIG. 7. Panel (a) shows two extremes both in terms of their total integrated value and the interval, τ_L, of their low-frequency oscillations. The uncertainty envelope for the integrated HCACF in purple is slightly above the maximum standard error (blue), whereas that of the HCACF integral in brown is below. The corresponding noise and noise integrals for these extrema are shown in panel (b).

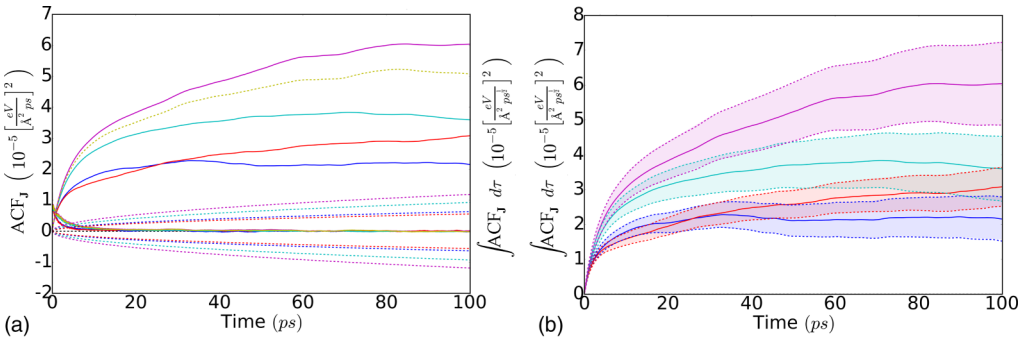

FIG. 8. Panel (a) shows the averaged HCACFs for all simulations along x (cyan) and y (magenta), the HCAFCs for x (blue) and y (red) for the large, 8-ns, simulation and the corresponding integrals in the same color. To observe the effect of a single outlier, all HCACFs except the *purple* one (see Fig. 7) are averaged. The resulting HCACF and integral are plotted in dashed yellow. Panel (b) shows the integrals [using the same color scheme as in panel (a)] with the corresponding uncertainty envelope around them.

noise. Upon first impression the integral in purple, in Figs. 5(b) and 7(a), stands out as having a large noise—it is well above the mean of all integrals [shown in red in the inset in Fig. 5(b)]. Yet, since its value is large, the error is a smaller fraction of the total integral value. There are possibly three factors at play here. (1) A random walk is, well, random, and the uncertainty envelope is merely an estimate of the expected value of any random walk for a given σ and τ. (2) Figure 7(a) includes the individual uncertainty envelopes computed with the random walk approach for each simulation. In both cases $\sigma_L \approx 2.9 \times 10^{-7}$ eV^2 Å^{-4} ps^2. However, τ_L is 1.24 ps for the simulation in brown, and 6.32 ps for the simulation in purple, so some of the error does seems to be due to a lower noise frequency and it is accounted for in the envelope. (3) A closer look at this HCACF reveals that it is not yet converged [see Fig. 7(b)]. In this particular case, the noise due to the random walk is not the main cause for the discrepancy between this HCACF integral and the remainder. This is in agreement with the above discussion of the individual simulations' distribution. The uncertainty envelope for this simulation being below the integrated HCACF is thus consistent with the random walk method being broadly agnostic to slow decay processes. To reinforce this idea, we computed τ_L after displacing the HCACF tail by the mean so it oscillates around zero and it equals 6.28 ps, not noticeably different from $\tau_L = 6.32$ ps as calculated above. In other words, because we are interested in the rapid decay process of the HCACFs, slow-rate processes in the HCACF are not mistaken for noise.

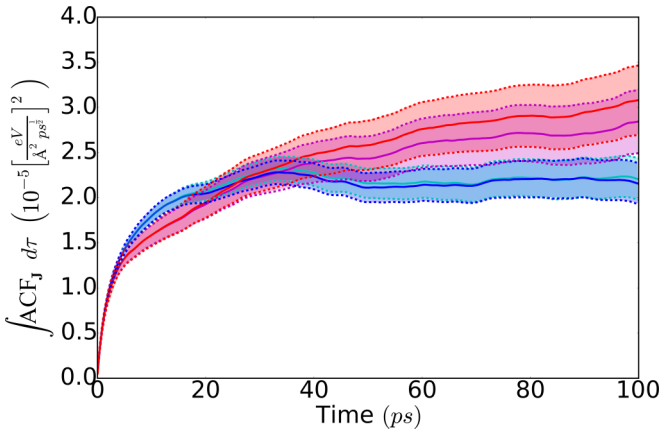

FIG. 9. This shows the integrated HCACF average for all simulations along x (cyan) and y (magenta) for the subset of 800-ps simulations resulting from the 8-ns simulation, the integrated HCAFCs for x (blue) and y (red) for the large, 8-ns, simulation and the corresponding uncertainty envelope around them.

The random walk uncertainty quantification approach could be a valuable tool for guiding researchers on how the noise varies over time or across simulations. To test this, a simulation of the same system was performed along x and y for 8.0 ns. For the 8.0-ns simulation data was collected at 0.04-ps intervals. The set of 18 simulations of 600 ps each adds to 10.8 ns, or 5.4 ns if we consider the x and y independently, with data collected every 0.02 ps. A total of 200 000 data points are available for averaging over the single simulation, and 270 000 for a nine-simulations set. As expected, the final HCACF for the 8.0-ns simulation is much smoother than any of the HCACFs from the 600-ps simulations, but as shown in Fig. 8 it continues to retain some of its oscillatory features. In Fig. 8, the integrated mean HCACFs for x and y for each of the two sets of nine simulations are compared to the x and y HCACF integrals obtained from the 8.0-ns simulation and their corresponding uncertainty envelopes. Figure 8(a) also shows the impact of a single outlier on the integrated HCACF average. Strikingly, the noise obtained from a single large simulation with fewer data points is lower than that obtained by averaging multiple simulations over a greater number of data points.

Recall that each simulation was performed from scratch by replicating a unit cell and conferring each system a temperature using individual seeds for each simulation. To determine if the discrepancy between the cross-autocorrelation averaging and the single-simulation autocorrelation averaging was maintained over a similar simulation length for the same seed, we subdivided the 8-ns simulation into a set of 10 800 ps simulations and averaged over them (see Fig. 9). Cross-simulation averaging with the same amount of data actually seems to reduce the error slightly. Most importantly, the smaller interval selected for a larger simulation is a worthwhile trade-off.

An example of an on-the-fly application of the suggested approach is given in Fig. 10(a), which shows the running mean of the evolving random walk uncertainty envelope as

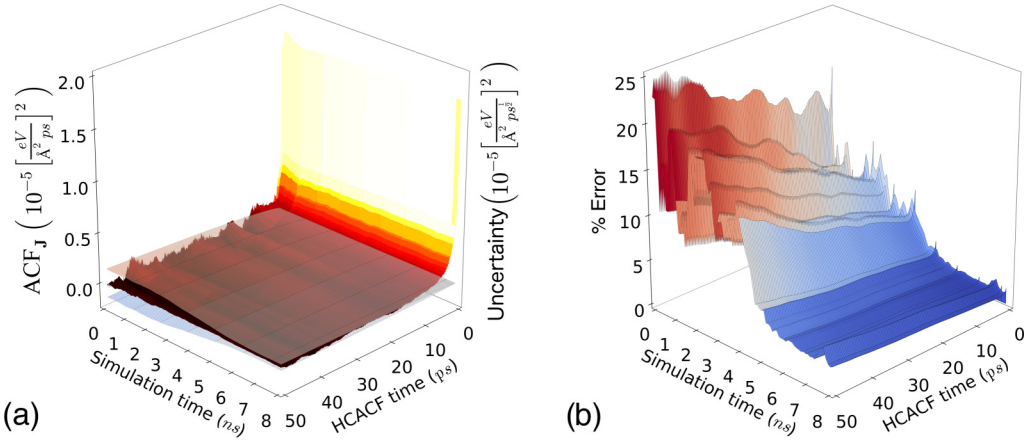

FIG. 10. In panel (a), in addition to the HCACF, the moving average of the uncertainty envelope computed using the random walk approach is also propagated through the simulation time. In panel (b) the percentage error is computed as the uncertainty envelope over the total integral.

the simulation progresses. The correlation (R) between τ and the evolving envelope is 0.52, and that between σ and the envelope is 0.56, both with a zero P value. This indicates a strong dependence of the envelope variance on both variables. The percentage error is computed throughout the simulation as the ratio between the envelope and the integral of the HCACF [see Fig. 10(b)]. It is interesting to notice that around 4 ns there is a steep decrease in the expected HCACF integrated noise, after which point the variation in the uncertainty diminishes.

To determine if there was an apparent direct correspondence between the system's Lyapunov memory and the system's energy fluctuation memory, we computed the Lyapunov instability, λ, which was found to be around 0.55 THz. Several simulation intervals for the system size were considered, including the 0.2-fs interval used for our simulations. The systems lose coherence between 15–20 ps. The distance, $d(t)$, between systems was computed as $|(X)_A - (X)_B|$, where $(X)_A$ are the coordinates of system A, started an approximate 10^{-5} Å distance away from system B.

To evaluate the hypothesis that the origin of the noise in the tails results from larger peaks in **J** that have not been averaged out due to insufficient data, we performed an autocorrelation through **J** with both a gradual and a rough cutoff of these peaks [see Fig. 11(a)]. The results obtained [see Figs. 11(b)–11(e)] indicate otherwise. A cut, soft—i.e., such that the value of **J** is reduced by a higher fraction the further away from zero **J** it is—or abrupt—i.e., removing peaks above and below a cutoff—through **J** reveals the importance of the peaks to set the shape of the HCACF [see Fig. 11(b)], but it provides evidence contrary to our hypothesis that the correlation between a few wider peaks were at the origin of the random-walk-type noise. If we consider a moving average (in red) through **J**, we find that it perfectly captures the trend of the HCACF [see Fig. 11(b)]. The normalized HCACF obviates that the trend of the data is more acutely captured by the moving average. The

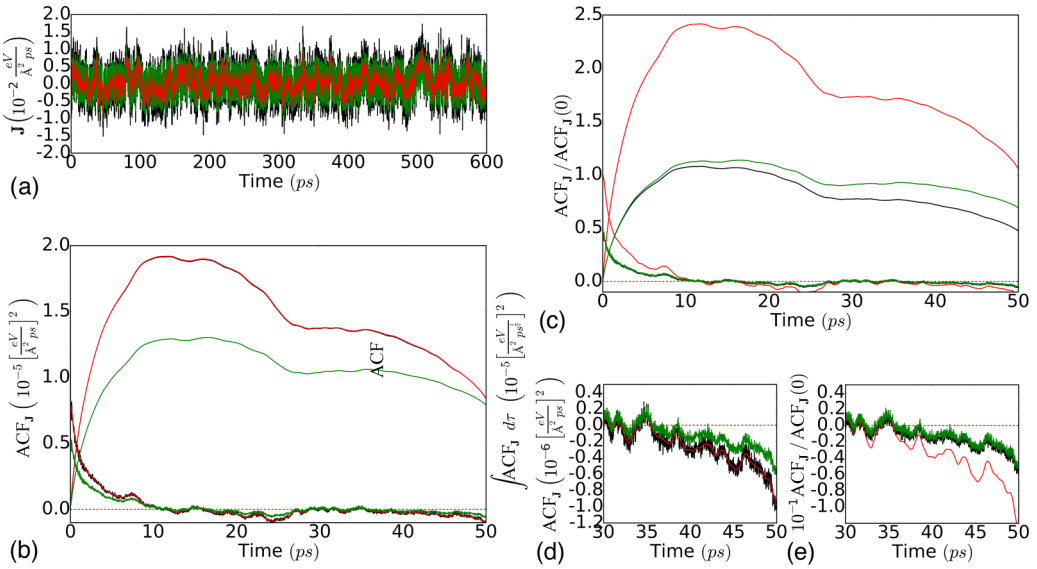

FIG. 11. The heat flux (black), **J**, a 0.4-ps moving average of **J** (red), and a gradual cutoff of the higher peaks of **J** (green) are shown in panel (a). The HCACF and integral for each of the above cases is shown in panel (b) as is, and is normalized in panel (c). Panels (d) and (e) are close-ups of panels (b) and (c), respectively. The color coding is maintained throughout the figures.

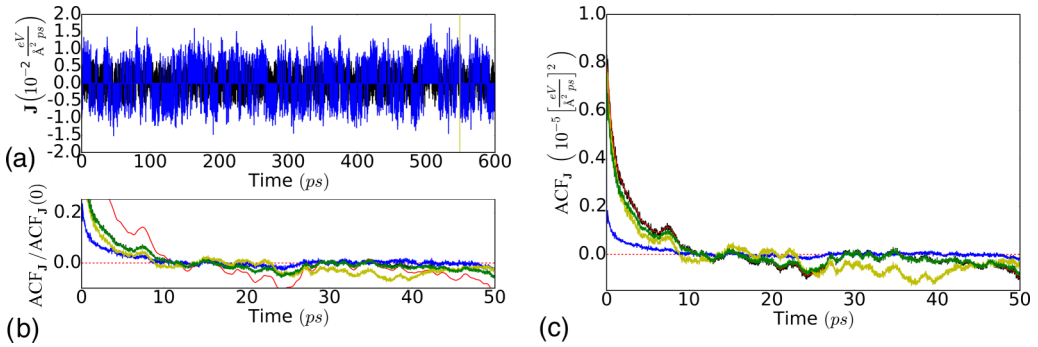

FIG. 12. Panel (a) shows **J** (black), a transform on **J** that keeps its higher peaks and replaces data between the peaks with a zero value (blue), and a line at 550 ps representing a cutoff of the **J** data above it. Panel (b) shows the normalized HCACF for the above cases, including those depicted in Fig. 11(a). The HCACF as is shown in panel (c). The color code is kept constant between Figs. 11 and 12.

normalized HCACF discrepancy between the moving average and the actual data could be omitted by normalizing the moving average autocorrelation function by the first element of the true HCACF.

If we, conversely, only consider the data from the highest peaks, setting all other data to zero (in blue in Fig. 12), some of the noise fades away, but so does the overall trend of the HCACF. A cut through the data increases the noise as expected (in yellow in Fig. 12), by reducing the amount of data to average over. In as far as we can ascertain, the noise is coupled to the overall fluctuations of **J**.

IV. CONCLUSION

In this paper we propose a method for quantifying the uncertainty of the autocorrelation function and thus that of transport properties computed using the Green-Kubo approach. This method is based on the premise that the noise of the autocorrelation function is akin to discrete white noise and it integrates into a random walk. The value of this method goes beyond estimating the error of a single simulation and it can be used to determine the minimum duration of a simulation to achieve a desired error threshold, as evidenced in Fig. 10. Most valuably, for a stipulated error, this method can be used to determine the optimal simulation time on the fly. While we have not found conclusive evidence for the origin of the noise, we have determined it is coupled to the overall trend of the measured flux and that the error is largely the result of fluctuations at frequencies below terahertz. Moreover, our results indicate that it is preferable to trade off a smaller time step for a longer total simulation time with a wider time step, to smooth the long-term oscillatory behavior of the HCACF, provided the time step is large enough to account for the relevant physics of the simulated system. Transport properties computed with equilibrium MD can be optimized by combining (1) performing a single simulation to determine the minimum required simulation time to reach a desired Markovian error with (2) performing multiple independent simulations with which to obtain a robust average autocorrelation function and standard error. The suggested approach can also be used to determine if slow decay processes are present in the autocorrelation by comparing the noise distribution to a normal with the mean and standard deviation found to characterize the noise. The method herein is suitable for high-throughput approaches for which expeditious simulations and uncertainty quantification are paramount.

ACKNOWLEDGMENTS

L.d.S.O. thanks Daniel McCoy and Trevor Howard for useful discussions. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant No. OCI-1053575, this work was supported in part by the National Science Foundation under Award No. 1403423.

[1] V. V. Chaban and O. V. Prezhdo, ACS Nano **8**, 8190 (2014).
[2] H. Kang, Y. Zhang, M. Yang, and L. Li, J. Nanotechnol. Eng. Med. **3**, 021001 (2012).
[3] P. C. Mishra, S. Mukherjee, S. K. Nayak, and A. Panda, Int. Nano Lett. **4**, 109 (2014).
[4] S. T. Cui, P. T. Cummings, and H. D. Cochran, Mol. Phys. **93**, 117 (1998).
[5] F. Léonforte, J. Servantie, C. Pastorino, and M. Müller, J. Phys.: Condens. Matter **23**, 184105 (2011).
[6] J. A. Thomas and A. J. McGaughey, Nano Lett. **8**, 2788 (2008).
[7] N. Zhang, P. Zhang, W. Kang, D. Bluestein, and Y. Deng, J. Comput. Phys. **257**, 726 (2014).
[8] Y. Zhang, G. Guo, and G. Nie, Phys. Chem. Miner. **27**, 164 (2000).
[9] R. E. Jones, D. K. Ward, and J. A. Templeton, J. Chem. Phys. **141**, 184110 (2014).
[10] M. S. Green, J. Chem. Phys. **22**, 398 (1954).
[11] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
[12] L. Han, M. Budge, and P. A. Greaney, Comput. Mater. Sci. **94**, 292 (2014).
[13] S. G. Volz and G. Chen, Phys. Rev. B **61**, 2651 (2000).

[14] C. Hooker, A. Ubbelohde, and D. Young, Proc. Royal Soc. London, Ser. A **284**, 17 (1965).
[15] D. P. Sellan, E. S. Landry, J. E. Turney, A. J. H. McGaughey, and C. H. Amon, Phys. Rev. B **81**, 214305 (2010).
[16] O. Borodin, G. D. Smith, and H. Kim, J. Phys. Chem. B **113**, 4771 (2009).
[17] E. Kaxiras and K. C. Pandey, Phys. Rev. Lett. **61**, 2693 (1988).
[18] H. Ohta and S. Hamaguchi, Phys. Plasmas **7**, 4506 (2000).
[19] L. de Sousa Oliveira and P. A. Greaney, in *MRS Proceedings* (Cambridge University Press, Cambridge, UK, 2013), Vol. 1543, pp. 65–70.
[20] J. Che, T. Çağın, W. Deng, and W. A. Goddard III, J. Chem. Phys. **113**, 6888 (2000).
[21] J. Li, L. Porter, and S. Yip, J. Nucl. Mater. **255**, 139 (1998).
[22] L. de Sousa Oliveira and P. A. Greaney, Comput. Mater. Sci. **103**, 68 (2015).
[23] D. Alfè and M. J. Gillan, Phys. Rev. Lett. **81**, 5161 (1998).
[24] A. Marcolongo, P. Umari, and S. Baroni, Nat. Phys. **12**, 80 (2016).
[25] J. T. K. Wan, T. S. Duffy, S. Scandolo, and R. Car, J. Geophys. Res. **112** (2007).
[26] Y. Cai, J. Lan, G. Zhang, and Y.-W. Zhang, Phys. Rev. B **89**, 035438 (2014).
[27] A. J. H. McGaughey and M. Kaviany, Phys. Rev. B **69**, 094303 (2004).
[28] F. Saiz and C. H. Amon, in *ASME 2015 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems collocated with the ASME 2015 13th International Conference on Nanochannels, Microchannels, and Minichannels* (American Society of Mechanical Engineers, Burlingame, United States, 2015), p. V002T06A002.
[29] G.-J. Guo, Y.-G. Zhang, K. Refson, and Y.-J. Zhao, Mol. Phys. **100**, 2617 (2002).
[30] E. R. Meyer, J. D. Kress, L. A. Collins, and C. Ticknor, Phys. Rev. E **90**, 043101 (2014).
[31] Y. Zhang, A. Otani, and E. J. Maginn, J. Chem. Th. Comput. **11**, 3537 (2015).
[32] B. Hess, J. Chem. Phys. **116**, 209 (2002).
[33] B. Huang, A. McGaughey, and M. Kaviany, Int. J. Heat Mass Transf. **50**, 393 (2007).
[34] M. Mouas, J.-G. Gasser, S. Hellal, B. Grosdidier, A. Makradi, and S. Belouettar, J. Chem. Phys. **136**, 094501 (2012).
[35] A. Mondal and S. Balasubramanian, J. Chem. Eng. Data **59**, 3061 (2014).
[36] T. Chen, B. Smit, and A. T. Bell, J. Chem. Phys. **131**, 246101 (2009).
[37] J.-F. Danel, L. Kazandjian, and G. Zérah, Phys. Rev. E **85**, 066701 (2012).
[38] J. Chen, G. Zhang, and B. Li, Phys. Lett. A **374**, 2392 (2010).
[39] D. Alfe, G. Kresse, and M. J. Gillan, Phys. Rev. B **61**, 132 (2000).
[40] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).
[41] S. J. Stuart, A. B. Tutein, and J. A. Harrison, J. Chem. Phys. **112**, 6472 (2000).