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

Coarse-grained hydrodynamics from correlation functions

Bruce Palmer

Phys. Rev. E **97**, 022106 — Published 6 February 2018

DOI: 10.1103/PhysRevE.97.022106

Coarse-Grained Hydrodynamics from Correlation Functions

Bruce Palmer
Computational Sciences and Mathematics Division
Pacific Northwest National Laboratory *
Box 999
Richland, WA 99352

Abstract

This paper will describe a formalism for using correlation functions between different grid cells as the basis for determining coarse-grained hydrodynamic equations for modeling the behavior of mesoscopic fluid systems. Configurations from a molecular dynamics simulation or other atomistic simulation are projected onto basis functions representing grid cells in a continuum hydrodynamic simulation. Equilibrium correlation functions between different grid cells are evaluated from the molecular simulation and used to determine the evolution operator for the coarse-grained hydrodynamic system. The formalism is demonstrated on a discrete particle simulation of diffusion with a spatially dependent diffusion coefficient. Correlation functions are calculated from the particle simulation and the spatially varying diffusion coefficient is recovered using a fitting procedure.

* Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under Contract DE-AC06-76RLO 1830.

I. INTRODUCTION

Mesoscale systems with dimensions on the order of nanometers are of considerable research interest but formidable challenges remain to understanding their behavior theoretically. From a molecular point of view, mesoscale systems are huge and direct simulation using atomistic techniques is prohibitively expensive. Not only is the number of molecules necessary for such a simulation large, but the time scales that need to be simulated increase with system size. Continuum approaches, such as computational fluid dynamics, are much more attractive in terms of speed and efficiency, but many of the assumptions behind continuum theories are breaking down in the mesoscale regime. The idea that the changes being modeled vary slowly on molecular dimensions may no longer be true and the discreteness of individual molecules is likely important for many nano-scale systems. It is fairly intuitive that in the mesoscale regime, conventional hydrodynamics will need to be either replaced or generalized in some way, but it is unclear how to go about this. It is also not clear how traditional hydrodynamic properties could be generalized to account for mesoscale behavior.

Simulations of fluids at interfaces and in pores strongly suggest that behavior will depart from that of the bulk fluid at scales on the order of a few molecules. Liquids become highly structured near many surfaces and this can be expected to alter properties such as diffusion, viscosity and thermal transport[1–4]. How to incorporate these changes into appropriate coarse grained equations and how to measure these properties is an unresolved question, but the indications from molecular simulations are that using hydrodynamics applicable for bulk systems is not appropriate.

This paper will investigate the use of correlation functions to map out the linear operators governing the fluid behavior in mesoscale systems. This is a generalization of methods that have been applied to the behavior of plane wave fluctuations and other correlation-based approaches to understanding hydrodynamic behavior[5–7]. Correlation functions can be formed from the basis functions commonly using in continuum calculations, such as indicator functions on rectangular grid cells. Based on general properties of correlation functions[7, 8], a simple relationship can be found between the time integral of these functions and linear operators defined on a grid that represents coarse-grained models of the system. This approach can provide a direct connection between mesoscale hydrodynamics and atomistic microscopic simulations. It is also an alternative to the usual coarse-graining approaches that are often used to derive mesoscale governing equations from microscopic models. Instead of reducing the number of degrees of freedom[9–11] or coarse-graining over a smoothing function to get mesoscopic equations[12], the coarse-grained governing equations can be derived from correlation functions.

The connection between fluctuations and hydrodynamic behavior has a long history, starting approximately with the work of Green[5, 6] in the early 1950's and subsequently extended and formalized by Kubo[13] and Zwanzig[14]. This work showed how autocorrelation functions formed from microscopic variables could be used to evaluate transport coefficients in continuum equations describing macroscopic fluid flows. Most of the practical calculations involving this formalism used Fourier transforms to analyze periodic systems, in effect, projecting fluctuations onto plane waves. The decay of these plane waves could then be used to evaluate the transport parameters in the governing equations. In addition to providing formulas for many transport coefficients, this work also lead to important advances in understanding light scattering in fluids[15] and microscopic behavior in many other physical systems[7].

Many simulation studies have looked at the behavior of hydrodynamic fluctuations projected onto plane waves. Some of the earlier work is summarized in Hoheisel and Vogelsang[16] and later studies have used parameterizations of decaying plane waves to evaluate transport coefficients[17–19]. More recent research has explored the connection between fluctuating plane waves seen in molecular dynamics simulations to fluctuating hydrodynamic models[20]. The focus on plane waves has restricted these studies to systems that are periodic and uniform in their properties. Mesoscale systems, such as a nanopore or colloid, are characterized by lower symmetry and spatially varying properties that do not map well to plane waves.

Localized functions restricted to finite regions, such as those used in continuum mechanics calculations, are much more attractive for nanoscale applications. Recent research by Español and coworkers has looked at the use of Delauney triangulations as basis functions for deriving macroscopic hydrodynamic equations from microscopic Hamiltonians[21–23]. However, to make progress in the derivations, this work makes use of approximations that are not applicable on length scales where there are variations in properties over lengths close to molecular scales. In particular, the requirement that individual grid cells hold very large numbers of molecules and that properties vary slowly on grid cell dimensions are not valid on the scale of a few molecules. For hydrodynamics, the additional requirements that fluctuations in the stress and heat currents decay on distances much shorter than grid cell dimensions are also extremely problematic. Relaxing these approximations to derive equations appropriate at the nano-scale would appear to lead to enormous complications in obtaining closed-form results. It may be possible to obtain information on the approximate forms for equations appropriate for nanoscale hydrodynamics using this approach but there is still a need to connect simulations to these approximate forms.

The work by Español *et al.* is a bottom-up approach to calculating hydrodynamics from an underlying Hamiltonian. The work presented in this paper is a top-down approach that assumes that measurable hydrodynamics exists, in the form of correlation functions, and that hydrodynamic operators can be inferred from these functions. The remainder of this paper will describe how correlation functions between cells in a numerical grid can be used to evaluate the linearized hydrodynamic evolution operator for the system. The formalism will be illustrated on a simple example with a spatially varying diffusion constant.

II. PROJECTION OF FLUCTUATIONS

This section will discuss a set of equations that can be used to relate the linear operator governing a system to a set of correlation functions describing the decay of fluctuations in grid cells overlaying the simulation. These relations are similar to many others in the literature[7, 8], but for the sake of completeness and to define a consistent set of notations and equations, a short derivation is presented here.

Starting with an atomistic simulation of a system, some set of basis functions is overlayed onto the simulation volume. Traditional approaches have usually looked at projecting system behavior onto plane waves, but there is no reason that other types of functions could not be used. It is possible to project fluctuations onto any complete set of functions $\{\phi_i\}$. The projection of a microscopic field of the form

$$a(\vec{r}) = \sum_n a_n \delta(\vec{r} - \vec{r}_n)$$

onto an orthonormal basis function ϕ_i is given by

$$\alpha_i = \int \phi_i(\vec{r})a(\vec{r})d\vec{r} = \sum_n a_n \phi(\vec{r}_n)$$

This is completely general and can be applied to any complete basis. If space is divided up into grid cells, then functions defined in one grid cell and zero outside the cell are orthogonal to functions in other grid cells. For a set of basis functions within a grid cell, it is likely that the first function in the basis is a constant. If only the first function is included, then the projection of the field in the region of support is just the average value of the field for all particles in that region. This allows a direct mapping of a particle simulation to a numerical grid with each grid cell representing the support of a set of basis functions. It is possible to include functions beyond the constant function in this expansion, at the expense of more algebra.

The goal of this analysis is to extract the discretized evolution operator from equilibrium fluctuations projected onto a numerical grid. Assume that the fluctuations are small enough that the corresponding evolution operator is linear and that it has no explicit or implicit time dependence. The continuous system is assumed to obey an equation of the form

$$\frac{\partial X}{\partial t} = L \cdot X \tag{1}$$

where X is some collection of continuum fields and L is some linear operator. The discretized form of equation (1) is

$$\frac{\partial \overline{X}}{\partial t} = \overline{\overline{L}} \cdot \overline{X} + \overline{B} \tag{2}$$

where $\overline{X}$ is a vector of coefficients and $\overline{\overline{L}}$ is a constant matrix corresponding to the linearized operator describing the behavior of small fluctuations. The vector $\overline{B}$ is a constant vector that may appear after applying boundary conditions to the discrete set of equations. Performing a Laplace transform on equation (2) and solving for $\overline{X}(z)$ gives the algebraic equation

$$\overline{X}(z) = \frac{1}{z - \overline{\overline{L}}} \cdot \overline{X}(t=0) + \frac{1}{z - \overline{\overline{L}}} \cdot \overline{B} \tag{3}$$

The real-time equivalent of equation (3) is

$$\overline{X}(t) = e^{t\overline{\overline{L}}} \cdot \overline{X}(t=0) + e^{t\overline{\overline{L}}} \cdot \overline{B} \tag{4}$$

Define the matrix $\overline{\overline{A}}(z)$ as

$$\overline{\overline{A}}(z) = \frac{1}{z - \overline{\overline{L}}} \tag{5}$$

In real time, $\overline{\overline{A}}(t)$ is the operator

$$\overline{\overline{A}}(t) = e^{t\overline{\overline{L}}} \tag{6}$$

Define $X_i(t)$ as being the ith component of $\overline{X}(t)$. If the $X_i(t=0)$ are taken from an ensemble of fluctuations in the system then the time correlation functions of the X_i are related to the matrix elements $A_{ik}(t)$ via the set of coupled equations[7, 8]

$$\langle X_i(t)X_j(t=0)\rangle = \sum_k A_{ik}(t)\langle X_k(t=0)X_j(t=0) + \sum_k A_{ik}(t)\langle B_k X_j(t=0)\rangle \tag{7}$$

Defining the equilibrium set of fluctuations as

$$\Delta_{ij} = \langle X_i(t=0) X_j(t=0) \rangle$$

and the time correlation functions as

$$C_{ij}(t) = \langle X_i(t) X_j(t=0) \rangle$$

then the time correlation functions, equilibrium fluctuations and the matrix $\overline{\overline{A}}(t)$ are all related via the expression

$$\overline{\overline{C}}(t) = \overline{\overline{A}}(t) \cdot \overline{\overline{\Delta}} \tag{8}$$

The terms containing the constant vector $\overline{B}$ in equation (7) drop out if $\overline{X}(t)$ represents fluctuations with zero mean. Taking the $z \to 0$ limit of the Laplace transform of equation (8) results in the expression

$$\overline{\overline{C}}(z=0) = -\overline{\overline{L}}^{-1} \cdot \overline{\overline{\Delta}} \tag{9}$$

where $\overline{\overline{C}}(z=0)$ is just the time integral of $\overline{\overline{C}}(t)$

$$\overline{\overline{C}}(z=0) = \int_0^\infty \overline{\overline{C}}(t) dt \tag{10}$$

In principal, equations (9) and (10) provide a path for determining $\overline{\overline{L}}$ from numerical experiments. The time correlation functions and equilibrium fluctuation correlations (Δ_{ij}) can be measured directly from a numerical simulation. The correlation functions can be integrated over time to get the $C_{ij}(z=0)$ and $\overline{\overline{L}}$ can be determined by evaluating the expression

$$\overline{\overline{L}} = -\overline{\overline{\Delta}} \cdot \overline{\overline{C}}^{-1} \tag{11}$$

In practice, direct application of this procedure is probably too costly to produce useful results, since all possible correlation functions in $\overline{\overline{C}}(t)$ would need to be calculated for long times and to high accuracy. However, equation (10) does provide a pathway to evaluating operators using a least squares fitting procedure to construct hydrodynamic models that can contain significant spatial variability in the evolution equations. This is further investigated below by considering a simple diffusion model with a spatially varying diffusion coefficient.

III. FITTING CONTINUUM MODELS

Equation (11) can be used as the basis for fitting transport operators if an approximate form of the operator $\overline{\overline{L}}$ is known. Suppose that $\overline{\overline{L}}$ depends on a collection of parameters $\overline{a}$, where the number of parameters in $\overline{a}$ is not necessarily related to the dimension of $\overline{\overline{L}}$. For hydrodynamics, the parameters $\overline{a}$ could be the spatially varying values of the transport coefficents in each grid cell, for non-local operators, $\overline{a}$ could contain spatially varying parameters in some kernel function. Suppose further, that some fraction of the correlation functions in the correlation matrix have been measured and integrated over time. Denote this set of correlations as $\{ij\}$ and the integrated value of a correlation measured from the simulation as $\tilde{C}_{ij}$. The value predicted from the operator $\overline{\overline{L}}(\overline{a})$ is denoted as C_{ij}. Then $\overline{a}$ can be determined using the following least squares objective function

$$\chi = \sum_{\{ij\}} (C_{ij} - \tilde{C}_{ij})^2 \tag{12}$$

If $\overline{e}_i$ is a vector that is 1 at element i and zero everywhere else, then equation (12) can be rewritten as

$$\chi = \sum_{\{ij\}} (-\overline{e}_i \cdot \overline{\overline{L}}^{-1}(\overline{a}) \cdot \overline{\overline{\Delta}} \cdot \overline{e}_j - \tilde{C}_{ij})^2 \tag{13}$$

Equation (13) accounts for the fact that not all correlation functions may have been measured and these are not included in the fit. The more obvious choice of objective function is

$$\chi = \left| \overline{\overline{L}}^{-1}(\overline{a}) \cdot \overline{\overline{\Delta}} - \tilde{\overline{\overline{C}}} \right|^2 \tag{14}$$

If equation (14) is used and not all correlation functions have been evaluated, then the fits are performed as though the values of those correlation functions were zero. This may not be the case. Although equation (13) accounts for missing values of C_{ij} it does not account for the fact that not all equilibrium fluctuations in $\overline{\overline{\Delta}}$ may have been measured. However, if the diagonal elements have been measured, they are likely to be the largest values. The off-diagonal elements representing separated grid cells should be much smaller if the separation distance is larger than a molecular correlation length and the effects of neglecting them can be expected to be minimal.

Making some sort of assumption about the form of $\overline{\overline{L}}$ can dramatically lower the number of correlation functions that must be evaluated. Using equation (10), the number of unknowns is on the order of N^2, where N is the number of grid cells. In the diffusion example discussed below, the assumption that $\overline{\overline{L}}$ has the form of a diffusion operator with a spatially varying diffusion coefficient lowers the number of unknowns to order N. Combined with the least squares fitting procedure, this can reduce the number of correlation functions that must be calculated in order to determine $\overline{\overline{L}}$.

Most efficient optimization algorithms require the gradient of the objective function, which in this case has the form

$$\nabla_{\overline{a}} \chi = -2 \sum_{\{ij\}} (-\overline{e}_i \cdot \overline{\overline{L}}^{-1} \cdot \overline{\overline{\Delta}} \cdot \overline{e}_j - \tilde{C}_{ij})(\overline{e}_i \cdot \overline{\overline{L}}^{-1}(\overline{a}) \cdot \nabla_{\overline{a}} \overline{\overline{L}}(\overline{a}) \cdot \overline{\overline{L}}^{-1}(\overline{a}) \overline{\overline{\Delta}} \cdot_j) \tag{15}$$

Even for a small grid, the number of terms in the gradient could be quite large. However, the evaluation of expressions involving $\nabla_{\overline{a}} \overline{\overline{L}}$ can be done quite efficiently in most cases using sparse matrix techniques. The practical aspects of evaluating $\overline{\overline{L}}$ and its gradient will be discussed in more detail in the next section.

IV. DIFFUSION EQUATION WITH A SPATIALLY VARYING DIFFUSION COEFFICIENT

Equation (13) is applied to a simple diffusion problem

$$\frac{\partial \phi}{\partial t} = \nabla \cdot \alpha(\vec{r}) \nabla \phi \tag{16}$$

where ϕ is a scalar field that relaxes diffusively and $\alpha(\vec{r})$ is a spatially varying diffusion coefficient. A discrete particle simulation of this equation is used to generate correlation functions. These correlation functions are then used to fit a model of the system to the simulations and recover the diffusion coefficient as a function of position.

The simulation of equation (16) is based on a finite volume discretization. Consider a 1D discretization of the region from $-L/2$ to $L/2$ into N uniform cells of size $h = L/N$. The concentration ϕ in cell i is ϕ_i, where i runs from 1 to N. The cells at 0 and $N+1$ are reserved for implementing boundary conditions. The value of the diffusion coefficent α_i is defined on the boundary between cell $i-1$ and cell i. The subscripts on α run from 1 to $N+1$ and the boundaries of cell i are associated with the values α_i and α_{i+1}. It is also possible to define the diffusion coefficient in the center of cells and then obtain values at the cell faces using a convention such as

$$\alpha_{i+1/2} = \frac{\alpha_i + \alpha_{i+1}}{2}$$

where the subscript $i+1/2$ refers to the boundary between cell i and $i+1$. While it leads to slightly more complicated algebra, it does not introduce any fundamental difficulties that would alter the discussion below.

Using the standard constitutive relation for a diffusion current and a first order approximation for the gradient, the flux into cell i from cell $i+1$ is

$$f_i^+ = \frac{\phi_{i+1} - \phi_i}{h}\alpha_{i+1}$$

and the flux from cell $i-1$ is

$$f_i^- = -\frac{\phi_i - \phi_{i-1}}{h}\alpha_i$$

If ϕ is conserved then the change in ϕ_i in a time interval Δt is

$$\begin{aligned} \Delta\phi_i &= \Delta t \frac{f_i^+ + f_i^-}{h} \\ &= \Delta t \left[\phi_{i+1}\frac{\alpha_{i+1}}{h^2} - \left(\frac{\alpha_i}{h^2} + \frac{\alpha_{i+1}}{h^2}\right)\phi_i + \phi_{i-1}\frac{\alpha_i}{h^2} \right] \end{aligned} \tag{17}$$

Equation (17) suggests a way of constructing a particle-based simulation of the equation (16). If Δt is chosen small enough that all the coefficients of the ϕ_i in equation (17) are less than 1, then they can be interpreted as probabilities (the coefficents are dimensionless). The requirement that the coefficients be less than 1 is equivalent to the requirement that the time step is small enough to resolve the particle motion. Too large a time step would be similar to using too large a time step in other numerical techniques.

If the value of ϕ_i is represented by the number of particles in cell i, then the probability that a given particle will move to cell $i+1$ on the next move is $\Delta t\alpha_{i+1}/h^2$ and the probability that it will move to cell $i-1$ is $\Delta t\alpha_i/h^2$. Based on this observation, a simulation can be constructed by looping over all the particles in each cell and generating a random number x on the interval [0,1]. If $0 \le x < \Delta t\alpha_i/h^2$, then the particle moves to cell $i-1$ on the next move, if $\Delta t\alpha_i/h^2 \le x < \Delta t(\alpha_i/h^2 + \alpha_{i+1}/h^2)$, then the particle moves to cell $i+1$. For $x > \Delta t(\alpha_i/h^2 + \alpha_{i+1}/h^2)$ the particle remains in cell i. If cell i and cell $i+1$ have the same number of particles on average, then the average net flux between them is zero, so the constant solution where all cells have the same number of particles is a possible solution of this system. The constant solution can be obtained by fixing the value of ϕ to be equal at $-L/2$ and $L/2$. In the simulation, this can be done by assuming that the cells at $i = 0$ and $i = N+1$ have a fixed number of particles. One time step is completed when a move

attempt on every particle in the system has been completed. This includes the fixed number of particles in cells 0 and N.

Equation (17) is also the route to constructing the matrix $\overline{\overline{L}}(\overline{a})$ and its gradient with the respect to $\overline{a}$. For this problem, the boundaries at both ends of the system are held at the same constant value $\phi_{0,N} = c_0$. This corresponds to a solution of (16) where $\phi(\vec{r})$ is a constant function with the value c_0 throughout the system. Using this boundary condition, the matrix $\overline{\overline{L}}$ has the form

$$L_{1j} = -\left(\frac{\alpha_1}{h^2} + \frac{\alpha_2}{h^2}\right)\delta_{1j} + \frac{\alpha_2}{h^2}\delta_{2j} \tag{18}$$

$$L_{ij} = \frac{\alpha_i}{h^2}\delta_{i-1,j} - \left(\frac{\alpha_i}{h^2} + \frac{\alpha_{i+1}}{h^2}\right)\delta_{ij} + \frac{\alpha_{i+1}}{h^2}\delta_{i+1,j} \quad (i > 1 \text{ and } i < N) \tag{19}$$

$$L_{Nj} = -\left(\frac{\alpha_N}{h^2} + \frac{\alpha_{N+1}}{h^2}\right)\delta_{Nj} + \frac{\alpha_{N+1}}{h^2}\delta_{N-1,j} \tag{20}$$

For finite c_0, the vector $\overline{B}$ has the form

$$B_i = \frac{\alpha_1}{h^2}c_0\delta_{1i} + \frac{\alpha_{N+1}}{h^2}c_0\delta_{Ni}$$

To evaluate correlation functions, only fluctuations are considered and in this case, $c_0 = 0$ so $\overline{B}$ vanishes as well. For this problem, the boundary conditions contribute no parameters to $\overline{\overline{L}}$, but this may not always true. For systems with complex interactions with surfaces, some of the parameters in $\overline{a}$ may result from applying boundary conditions.

The derivatives of $\overline{\overline{L}}$ with respect to the α_k can be used by most optimization algorithms and are given by

$$\frac{\partial L_{1j}}{\partial \alpha_k} = -\left(\frac{1}{h^2}\delta_{1k} + \frac{1}{h^2}\delta_{2k}\right)\delta_{1j} + \frac{\alpha_2}{h^2}\delta_{2j} \tag{21}$$

$$\frac{\partial L_{ij}}{\partial \alpha_k} = \frac{1}{h^2}\delta_{ik}\delta_{i-1,j} - \left(\frac{1}{h^2}\delta_{ik} + \frac{1}{h^2}\delta_{i+1,k}\right)\delta_{ij} + \frac{1}{h^2}\delta_{i+1,k}\delta_{i+1,j} \quad (i > 1 \text{ and } i < N) \tag{22}$$

$$\frac{\partial L_{Nj}}{\partial \alpha_k} = -\left(\frac{1}{h^2}\delta_{Nk} + \frac{1}{h^2}\delta_{N+1,k}\right)\delta_{Nj} + \frac{1}{h^2}\delta_{N+1,k}\delta_{N-1,j} \tag{23}$$

For each α_k, the corresponding gradient $\partial\overline{\overline{L}}/\partial\alpha_k$ is extremely sparse. No more than 4 matrix elements are non-zero for any given α_k. Evaluating the gradients can therefore be done very efficiently if the calculation is properly organized, although coding it may be complex.

For simulations with fixed concentrations at the boundaries, the total number of particles is not a conserved quantity and this simplifies the equations. If periodic boundary conditions were used, the total number of particles is conserved. This results in a zero eigenvalue appearing in $\overline{\overline{L}}$ that can be eliminated by removing a degree of freedom from the system of equations (21)-(23) using the conservation law. A similar approach can be used in systems that conserve momentum or energy.

V. NUMERICAL RESULTS

Using the algorithm described above, a discrete simulation of a system containing 10 cells was performed. The simulation is subject to fluctuations in the number of particles in

each grid cell, so it is possible to calculate correlation functions and use these to recover the diffusion coefficient $\alpha(x)$. The 10 cells correspond to 11 values of α. The cell size is set to $h = 1$, the time step is $\Delta t = 0.2$ and the diffusion coefficient is given by the expression

$$\alpha(x) = D(1 - Ae^{-x^2/\sigma^2}) \tag{24}$$

where $D = 1$, $A = 0.8$ and $\sigma = 2.5$. The units are arbitrary, but all results are reported consistently using the same units. The boundary conditions are applied by fixing the number of particles at the two ends of the system (cells 0 and 11) at 100 and using this number of particle to calculate the fluxes into cells 1 and 10. The average number of particles in each cell is 100. The system is equilibrated for 10^5 steps or, equivalently, 2×10^4 time units (elapsed time is the number of steps times Δt). The minimum value of D in equation (24) is 0.2 so if the relaxation time of the entire system is L^2/D, the maximum relaxation time evaluates to 500 time units. 10^5 steps should be more than sufficient to produce an equilibrated configuration. Data is collected for 0.1, 1, and 10 million steps. The different simulation lengths are used to give an estimate of how the quality of the correlation functions affects the fit. In addition, the 10 million step simulation was repeated 5 times with different random number seeds to provide additional information on uncertainty. Auto-correlation functions of the concentration fluctuations are calculated for each of the 10 cells and cross correlation functions are calculated for nearest neighbor cells. Correlation functions are evaluated out to a maximum of 500 steps (100 time units) and a sliding window algorithm[24] is used to extract the maximum amount of data from the time series of concentration values.

The autocorrelation functions for the first five cells are shown in figure (1) from one of the 10M step simulations. Comparing different simulations indicates that the uncertainty in the correlation functions is comparable to the line width of the plot. The second five autocorrelation functions are the same as the first five, due to mirror-image symmetry. As expected, the correlation functions closer to the center of the simulation decay more slowly due to to the smaller diffusion coefficient in this region. For comparison, the nearest neighbor correlation functions are shown in figure (2). The correlations functions are of the form $\langle \phi_i(0)\phi_{i+1}(t)\rangle$. These functions decay more slowly than the corresponding autocorrelation functions and also have an initial value of approximately zero. Since the particles in this simple model have no interaction, and there is no hard constraint on the total number of particles, initial concentrations of particles are expected to be uncorrelated. This turns out not to be completely true and some correlation appears near the boundary. If the total number of particles is conserved, as would be the case for periodic boundary conditions, then there would be a slight negative correlation at time 0 for all cross-correlation functions. Although the figures only show the correlation functions out to time $t = 25$, the functions are calculated out to $t = 100$, where the correlations have largely decayed to zero. The error in the correlation function integration due to truncation is unquantified, but given the small magnitude of the functions at $t = 100$ (the largest values are about a factor of 500 smaller than the maximum values) the errors should be small. To check, the integration interval was increased to 1000 steps ($t = 200$) and a simulation of 50 million steps was run (the increased simulation time was to improve the statistics in the tails of the correlation functions). The change in fitting parameters was comparable to the variation seen between different runs using 10 million steps.

The average number of particles in each simulation cell was checked to see that the boundary conditions were producing the correct behavior. The average value for all cells was approximately 100, as expected, and the deviations from 100 decreased as the simulation

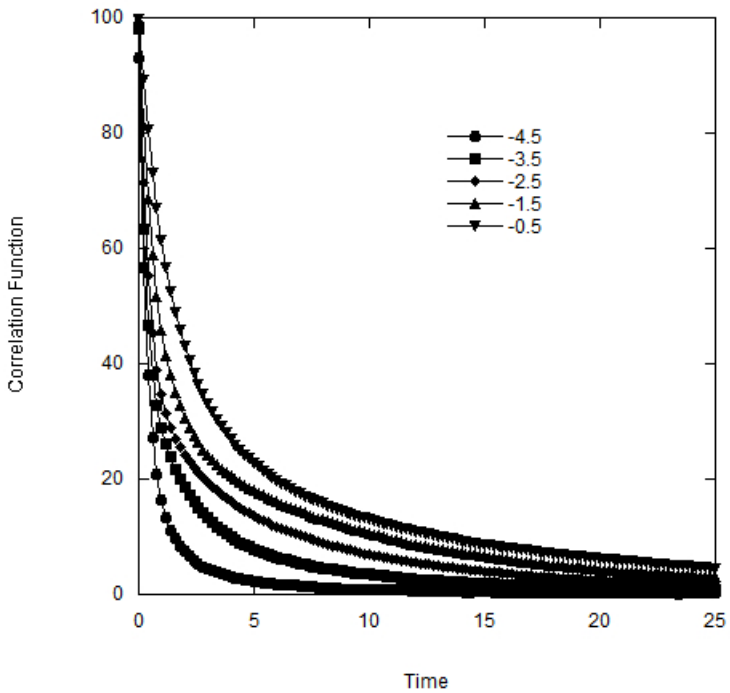

FIG. 1: Autocorrelation functions for concentration fluctuations in the first five simulation cells. Numbers refer to center location of cell.

length increased.

It is possible to compare the simulated correlation functions with expressions calculated using equation (8). The operator $\overline{\overline{L}}$ can be diagonalized and used to evaluate the time evolution operator in equation (6). The values of the matrix $\overline{\overline{\Delta}}$ are obtained from the simulation. A comparison of some autocorrelation functions from the 10M step simulation and equation (8) is shown in figure (3). Although the overall agreement is quite good, there are some small discrepancies. Some correlation functions between cells i and $i-1$ are shown in figure (4). Larger discrepancies for the cells near the boundary are clearly visible. Again, the correlation functions for 10M steps are reproducible in this time range to about the line thickness of the plots, so the differences in figure (4) are real. The diagonal elements of the fluctuation matrix $\overline{\overline{\Delta}}$, corresponding to the concentration fluctuations within cells, are shown in figure (5) for the simulations. As can be seen, the cells close to the edge of the simulation are different from those at the center, although the differences are relatively small compared to the average magnitude of the fluctuations. These differences, coupled with the discrepancies between the correlation functions and the expected exact results, suggests that the diffusion equation and the constant value boundary condition may not completely account for the dynamics of the discrete simulation in the vicinity of the boundary. To obtain a complete correspondence, a more subtle way of fixing the concentration at the boundaries for the simulation may be required.

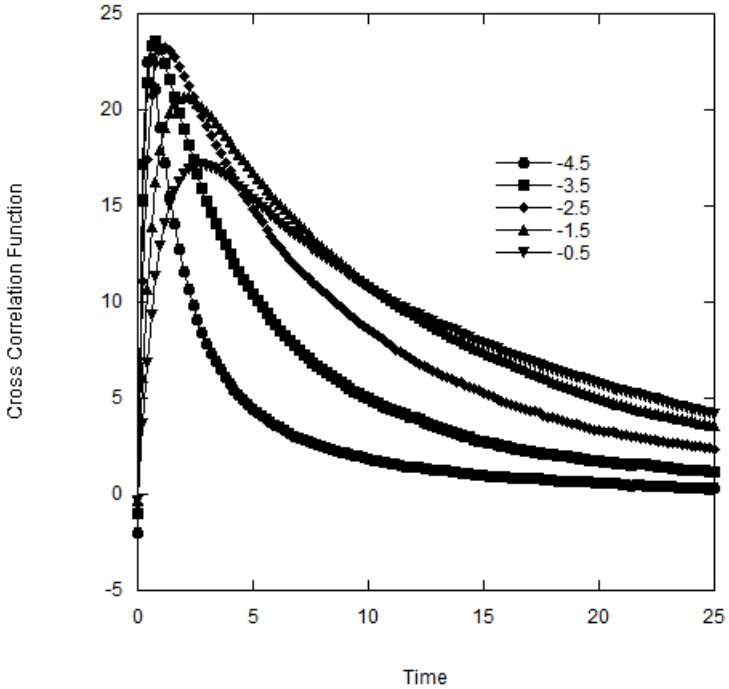

FIG. 2: Cross correlation functions for concentration fluctuations in the first five simulation cells and the next higher cell. Numbers refer to center location of the first cell.

The fits to the spatially varying diffusion coefficient are shown in figure (6). The results for the trajectory containing only 100K timesteps exhibit large errors. The results are arguably qualitatively similar to the correct profile, but clearly there are major deviations and the expected symmetry of the diffusion coefficient about the origin is missing. The situation improves markedly when the trajectory is extended to 1 million timesteps. In this case, the fitted diffusion coefficient is much closer to the true value. Increasing the simulation time to 10 million steps further improves the fit, although there are still some differences. The discrepancies in the fits track the behavior of the integrated correlation functions, shown in figure (7). Only the autocorrelation functions are shown, but similar results were obtained for the cross correlation functions. The 1 million and 10 million step trajectories give substantially the same results but the integrated values for the 0.1 million step show large variations. Given these differences, it is not surprising that the fitted values of $\alpha(x)$ are not very good for this trajectory.

A more precise picture of the uncertainties in the fitted results is presented in figure (8), which shows 5 fits from the 5 seperate simulations of 10M steps compared with the exact answer. The fits are clustered together fairly tightly, with more uncertainty at the edges of the system. However, the fitted values of $\alpha(x)$ appear to be consistently higher near the edges than the true answer, again suggesting that the proposed diffusion equation may not be completely accounting for the dynamics of this simulation.

The effect of the number of correlation functions included in the fit was also investigated.

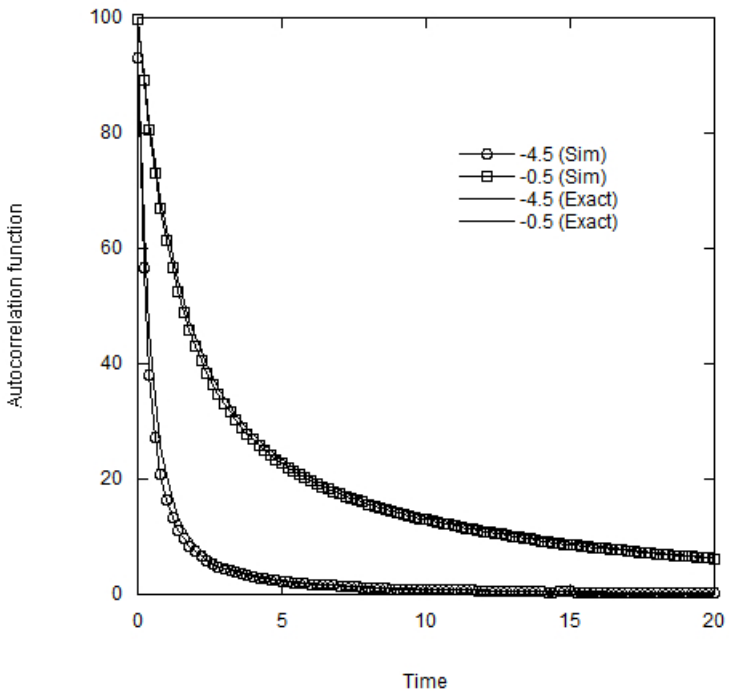

FIG. 3: Comparison of autocorrelation functions calculated using equation (8) and those from simulation. Autocorrelation functions for the cells centered at -4.5 and -0.5 are shown.

The next-nearest neighbor cross correlation functions, $\langle \phi_i(0)\phi_{i+2}(t)\rangle$, were calculated and included in the optimization, but had no appreciable effect on the quality of $\alpha(x)$.

VI. CONCLUSIONS

A method for fitting the structure of linear evolution operators from correlation functions obtained from projections of particle properties onto a numerical grid has been presented. Numerical results based on the behavior of a simple diffusion system show that this method is feasible. Although the test case presented here is somewhat artificial, there is nothing to suggest that this approach cannot be applied to more complicated systems. This is particularly promising for the nanoscale regime, where developing a precise picture of what equations to use to describe coarse-grained behavior is challenging. Different types of molecular dynamics calculations (e.g. classical MD or ab initio MD) can be used to generate correlation functions. These, in turn, can be used to probe the nature of operators governing the linear behavior of the systems. Given that there is currently little information on the structure of coarse-grained evolution equations, the correlation-based fits in this paper can be used to obtain more detailed information on transport near surfaces and nanochannels. The fits can also be used to determine the boundary conditions that apply in the nanoscale regime, since boundary conditions are incorporated into the discrete operator $\overline{\overline{L}}$.

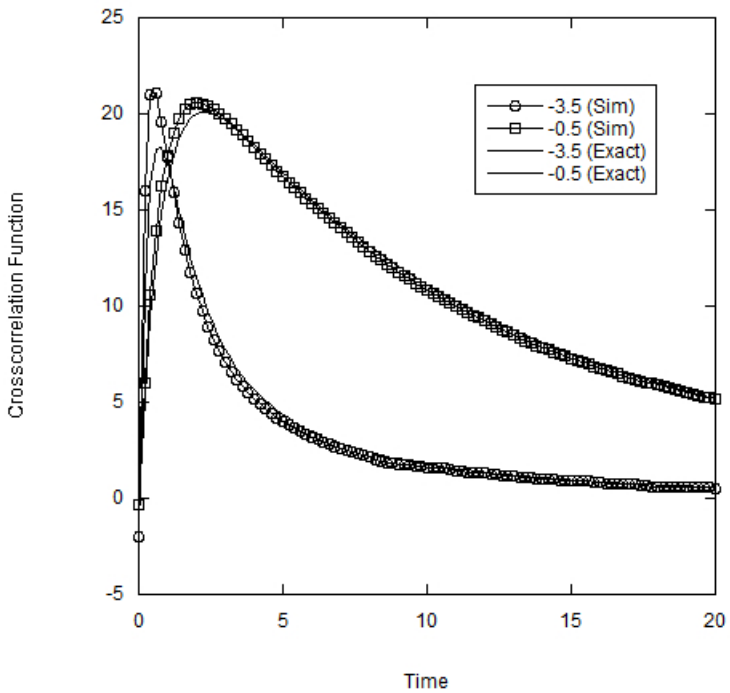

FIG. 4: Comparison of cross-correlation functions between cells i and $i-1$ calculated using equation (8) and those from simulation. Functions for i corresponding to the cells at positions -3.5 and -0.5 are shown.

Obvious generalizations of this work are to use it to investigate the behavior of more complicated molecular systems and include fields such as mass,momentum and energy densities as well as features such as charge density and ion concentration in the evolution operator. Incorporating memory effects, which leads to time dependencies in the operator $\overline{\overline{L}}$, is another area of future research. The chief challenges to using this approach are 1) to generate enough statistics to obtain reasonable quality correlation functions, 2) to code up analytical expressions for the linear operators and their gradients, and 3) to obtain fits to the global minimum for the linear operator. For realistic systems in three dimensions, the operator may have on the order of thousands of parameters and fitting these will therefore be more difficult than the relatively small number of variables used in this paper.

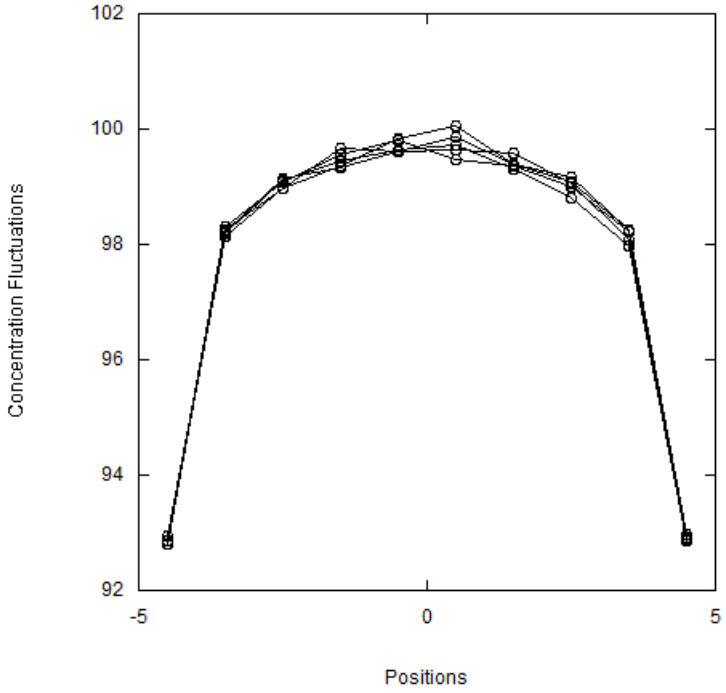

FIG. 5: Mean square concentration fluctuations within cells as a function of position for five separate simulations of 10 million steps.

Acknowledgements: A portion of this work was supported by the U.S. Department of Energy, Office of Science, Office of Advance Scientific Computing Research as part of the Collaboratory on Mathematics for Mesoscopic Modeling of Materials (CM4). Pacific Northwest Laboratory is operated by Battelle for the U.S. Department of Energy under contract DE-AC0576RLO1830.

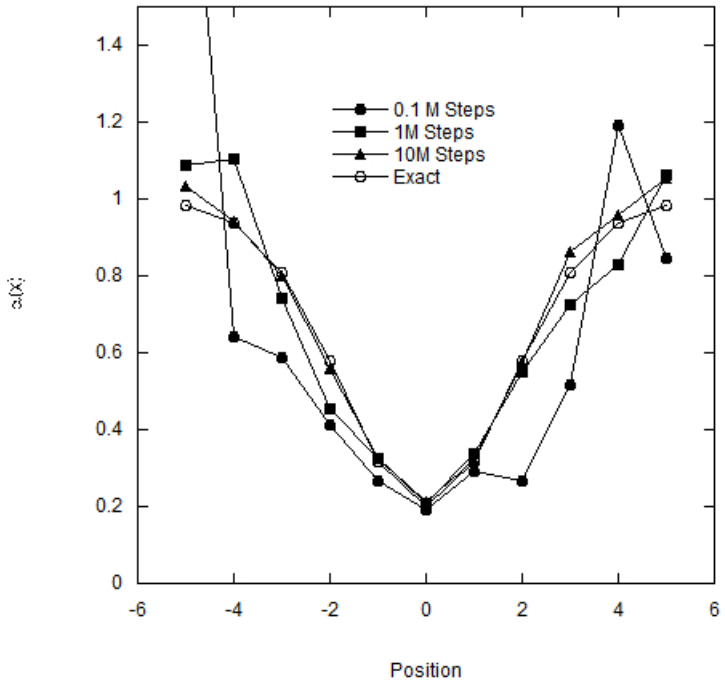

FIG. 6: Fits to the spatially varying diffusion coefficent $\alpha(x)$ for different length trajectories.

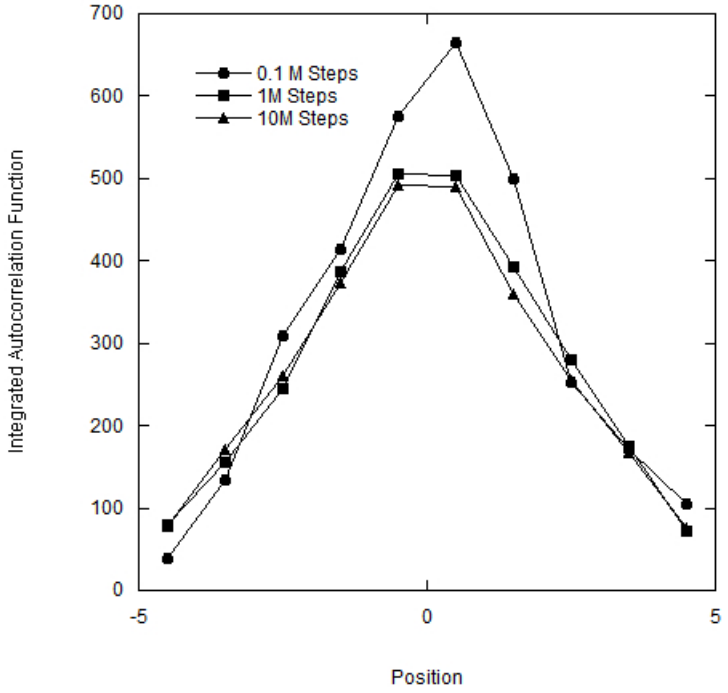

FIG. 7: Values of autocorrelation functions after integrating over time.

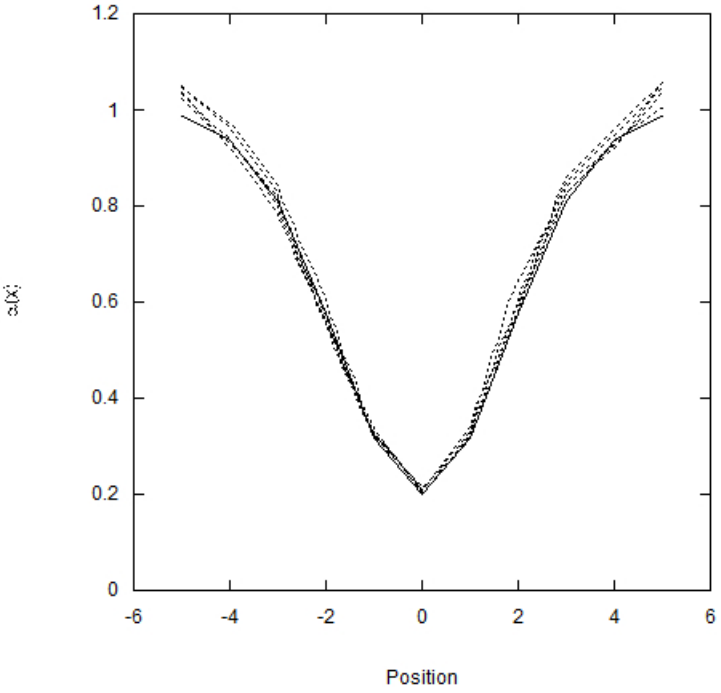

FIG. 8: Fits from five different simulations (dotted lines) compared with the exact expression for $\alpha(x)$ (solid line).

[1] D. Kim and E. Darve. Molecular dynamics simulation of electro-osmotic flows in rough wall nanochannels. *Phys. Rev. E*, 73:051203, 2006.

[2] Y. Li, J. Xu, and D. Li. Molecular dynamics simulation of nanoscale liquid flows. *Microfluid Nanofluid*, 9:1011–1031, 2010.

[3] A. Prakash J. Pfaendtner, J. Chun, and C. Mundy. Quantifying the molecular-scale aqueous response to the mica surface. *J. Phys. Chem. C*, 121:18496, 2017.

[4] B. Palmer. Direct simulation of hydrodynamic relaxation in microchannels. *J. Chem. Phys.*, 109:196, 1998.

[5] M. S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. *J. Chem. Phys.*, 20:1281–1295, 1952.

[6] M. S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. II. irreversible processes in fluids. *J. Chem. Phys.*, 22:398–413, 1954.

[7] D. Forster. *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions.* The Benjamin/Cummings Publishing Company, Inc., Reading, MA, 1975.

[8] C. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Springer-Verlag, New York, 1985.

[9] S. Isvekov and G. Voth. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.*, 123:134105, 2005.

[10] W. Noid, J.-W. Chu, G. Ayton, V. Krishna, S. Isvekov, G. Voth, A. Das, and H. Andersen. The multiscale coarse-graining method. I. a rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.*, 128:244114, 2008.

[11] P. Español and P. Warren. Statistical mechanics of dissipative particle dynamics. *Europhys. Lett.*, 30:191–196, 1995.

[12] J. Monaghan. Smoothed particle hydrodynamics. *Reports on Progress in Physics*, 68:1703–1759, 2005.

[13] R. Kubo. Statistical-mechanical theory of irreversible processes. I. general theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Japan*, 12:570–586, 1957.

[14] R. Zwanzig. Time-correlation functions and transport coefficients in statistical mechanics. *Annu. Rev. Phys. Chem.*, 16:67–102, 1965.

[15] J.-P. Boon and S. Yip. *Molecular Hydrodynamics.* Dover Publications, New York, 1980.

[16] C. Hoheisel and R. Vogelsang. Thermal transport coefficents for one- and two-component liquids for time correlation functions computed by molecular dynamics. *Comp. Phys. Rep.*, 8:1–70, 1988.

[17] B. Palmer. Transverse-current autocorrelation-function calculations of the shear viscosity for molecular liquids. *Phys. Rev. E*, 49:359–366, 1994.

[18] B. Palmer. Calculation of thermal-diffusion coefficients from plane-wave fluctuations in the heat-energy density. *Phys. Rev. E*, 49:2049–2057, 1994.

[19] B. Hess. Determining the shear viscosity of model liquids from molecular dynamics simulations. *J. Chem. Phys.*, 116:209–217, 2002.

[20] B. Shang, N. Voulgarakis, and J.-W. Chu. Fluctuating hydrodynamics for multiscale modeling and simulation: Energy and heat transfer in molecular fluids. *J. Chem. Phys.*, 137:044117, 2012.

[21] P. Español and I. Zúñiga. On the definition of hydrodynamic variables. *J. Chem. Phys.*,

131:164106, 2009.

[22] P. Español, J. Anero, and I. Zúñiga. Microscopic derivation of discrete hydrodynamics. *J. Chem. Phys.*, 131:244117, 2009.

[23] P. Español and A. Donev. Coupling a nano-particle with isothermal fluctuating hydrodynamics: Coarse-graining from microscopic to mesoscopic dynamics. *J. Chem. Phys.*, 143:234104, 2015.

[24] M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids.* Clarendon Press, Oxford, 1987.