The Pennsylvania State University
The Graduate School

FIRST-PRINCIPLE BASED MODEL REDUCTION METHOD WITH APPLICATIONS IN MOLECULAR DYNAMICS

A Dissertation in
Mathematics
by
Weiqi Chu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2019

The dissertation of Weiqi Chu was reviewed and approved* by the following:

Xiantao Li
Professor of Mathematics
Dissertation Advisor, Chair of Committee

Jinchao Xu
Verne M. Willaman Professor of Mathematics

John Harlim
Professor of Mathematics and Meteorology

Jorge O. Sofo
Professor of Physics and Materials Science and Engineering

Alexei Novikov
Professor of Mathematics
Chair of the Graduate Program, Department of Mathematics

*Signatures are on file in the Graduate School.

Abstract

In this dissertation, we introduce a first-principle based model reduction methodology with applications in molecular dynamics. We derive versatile orthogonal equations, using the idea of Mori-Zwanzig formalism. As applications, we study the heat conduction processes at nano-scale from a many-particle description and derive a group of equations in the form of the generalized Langevin equation for the coarse-grained variables.

We analyze the decay properties of the memory kernel function in the generalized Langevin equation quantitatively, with particular emphasis on the dependence of coarse-grained variables and levels of coarse-graining. We find that the kernel decays rapidly (exponentially) in space while the time decay is rather slow (in the power of $-1/2$).

By choosing the local energy as the coarse-grained variables, we observe fluctuating non-Fourier heat conduction phenomena. We also apply Markovian embedding techniques to eliminate the history dependence and introduced both additive and multiplicative Gaussian white noise to ensure the statistics of local energy. In sharp contrast to conventional energy transport models, this derivation yields stochastic dynamics models for the spatially averaged energy.

Beyond conventional projections, we come up with a new projection formalism by projecting the dynamics onto space spanned by nonlinear functions of coarse-grained variables, such as the driving force, entropy, etc. This procedure could address the challenge in modeling non-Gaussian statistics, such as energy with a lower bound, and give rise to nonlinear stochastic constitutive relations for the modeling of transient heat conduction processes at nano-scale.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest gratitude to my advisor Prof. Xiantao Li for his patience, support, and encouragement throughout my Ph.D. studies. Without his guidance and counsel, this dissertation would not have been possible. I could not imagine having a better advisor and mentor than him.

I would also like to thank the professors at Penn State who inspired and helped me, especially Prof. Chun Liu, Prof. John Harlim, Prof. Jinchao Xu, Prof. Jorge Sofo for teaching me so much through the years and their feedback on this dissertation.

I would like to thank my parents and family for their unconditional love and support over the years. Thank you for having confidence in me and being proud of me. You make me believe that I can do anything and everything in life.

The author was partially supported by National Science Foundation grants DMS-1522617 and DMS-1619661 during the writing of parts of this dissertation. The findings and conclusions of this dissertation do not necessarily represent the views of the National Science Foundation.

Dedication

To my parents.

Chapter 1 Introduction

1.1 Coarse-graining methodology

In many fields of applied sciences, one is often confronted with large-scale dynamical models, for which the computational cost is overwhelming. One example is molecular dynamical models, which are of great importance in understanding the structural properties of macromolecules and material systems [4]. Molecular dynamical models are written in the form of a Hamiltonian system for all particles, which requires enormous computations in direct numerical implementations.

On the other hand, of interest is a few coarse-grained variables that are sufficient to describe the overall collective properties of the system, such as locally averaged displacement, momenta, and energy. There is tremendous interest in reducing the enormous degrees of freedom associated with such models to a few coarse-grained variables that are able to capture and predict the properties people are interested in. Coarse-graining methodologies are essential in reducing the degrees of freedom, and remarkable progress has been made [5, 6]. In molecular dynamics (MD), the coarse-graining approach consists of replacing an atomistic description of biological molecules or materials with a lower-resolution model, with much fewer degrees of freedom, which averages or smooths away fine details.

Conceptually, the coarse-graining process can be described as follows. One starts with a large dynamical system, as shown in Figure 1.1. In general, the dynamics could be either deterministic or stochastic, in the form of ordinary partial differential

equations (ODEs), partial differential equations (PDEs) or stochastic differential equations (SDEs). The PDE system could be converted into an ODE system by standard procedures, such as Fourier transform, semi-discretization, finite element methods (FEM). The SDE system could also be converted into an ODE system by imposing random force and random initial values or working with the Fokker-Planck equation (FPE). The large dynamical system could also be discrete in time, where we could redefine the time derivative in the discrete case and write it as an ODE system.

The aim is to derive a reduced model that only involves a few variables $a = \phi(x)$ of interest, where $\dim(a) \ll \dim(x)$. In general, this procedure will lead to equations with memory effects, written abstractly in the form R, together with Q in the unresolved subspace, representing the *fluctuating* part of the equation,

$$\boxed{\dot{x}(t) = f(x(t))} \xrightarrow[a = \varphi(x)]{\text{Coarse Graining}} \boxed{\dot{a}(t) = R\left(\{a(t_1)\}_{t_1 \le t}\right) + Q.}$$

Figure 1.1: A diagram of the coarse-graining process.

Many coarse-graining methodologies have been developed and studied, both theoretically and algorithmically, as powerful tools to reduce the order of problems, including homogenization [7], reduced-order model [8, 9], scale-separation [10], etc. Enormous applications have been found in material science problems and biological problems [11–28]. Many coarse-grained models have been developed and they have shown great promise in reducing the computational cost and efficiently capturing the primary quantities of interest.

However, most existing coarse-grained molecular models focus on finding the effective potentials, known as the potential of mean forces, at a *constant* temperature. These existing coarse-grained models are in the similar form as the MD models with the possible addition of damping terms or random forces and typically take the coarse-grained variables as the linear combination of positions and velocities, describing only the time evolution of the averaged position and momentum. Among these methods, the Mori-Zwanzig formalism [29–31] has been developed as a systematic projection methodology for such dimension-reduction purposes.

1.2 Mathematical preliminaries

1.2.1 Many-particle description

Let us consider an ODE system suggestively written as,

$$\dot{x} = f(x), \quad x(0) = x_0. \tag{1.1}$$

The dot on the top represents the time derivative of variables. The dimension of x is a very large number, donated as $N = \dim(x)$. In the example of MD, (1.1) corresponds to an all-atom description, which embodies the detailed interactions among all the atoms in the system.

More specifically, let u_i and v_i be displacement and velocity of the i-th atom respectively. If atoms move in d-dimensional geographic space, then u_i and v_i are vectors in $\mathbb{R}^d$ and they follow the Newton's second law,

$$\begin{cases} \dot{u}_i = v_i, \\ m\dot{v}_i = -\dfrac{\partial V(u)}{\partial u_i}. \end{cases} \tag{1.2}$$

Here we assemble all displacements and velocities together as vectors u and v. $V(u)$ is the potential energy of the system. The above equation (1.2) can be expressed in a compact form in (1.1) with $x = (u, v)$. The dimension of the whole problem $N = 2dN_0$, with d being the geographic space dimension and N_0 being the total number of atoms.

Here, we assume the initial value $x_0 = (u_0, v_0)$ obeys a probability density function (PDF) ρ_0. If a mechanical system is in thermal equilibrium with a heat bath at a fixed temperature T, the statistical ensemble is called *canonical ensemble* with the density

$$\rho_0(x) = \frac{1}{\mathcal{Z}} e^{-\beta H}, \tag{1.3}$$

where $\beta = 1/k_B T$ and k_B is the Boltzmann constant, H is the Hamiltonian of the system and $\mathcal{Z}$ is a normalization factor.

1.2.2 Coarse-grained variables

However, due to the original physical system has a large number of degrees of freedom, solving (1.2) directly provides tremendous computational cost. Mostly, people are interested in collective behavior and properties of the system as a whole, rather than the instantaneous position and velocity of every particle in the fine-grained materials or systems. Reduced models are developed for these quantities of interest (QOI), which are of good use in describing the overall dynamics, especially in representing and predicting the longer time- and large-scale dynamics.

Next we consider a general case when the QOI is a function of x, namely, $a = \bar{\varphi}(x)$. Since x is a solution to (1.1) and it is important for our derivation to include the dependence on the initial configuration. We denote this relation as

$$a \overset{\text{def}}{=} \bar{\varphi}(x(t)) = \varphi(x_0, t). \tag{1.4}$$

In statistical mechanics, a is often referred to as *coarse-grained* (CG) variables. The selection of CG variables is problem-dependent. In principle, the dimension of CG variables should be much smaller than that of the original problem, *i.e.*, $\dim(a) \ll N$. Examples include local energy, averaged velocity, reaction coordinates, etc.

For a short notation, we denote a without parenthesis as the initial value, *i.e.*, $a = a(x_0, 0)$ and write $a(t) = \varphi(x_0, t)$ for the sake of simplicity. Even though we neglect the initial value x_0, there still exists the intrinsic dependence on it.

The goal is to derive a reduced equation for $a(t)$. We first define the Liouville operator (propagating operator) $\mathcal{L}$ as,

$$\mathcal{L} \overset{\text{def}}{=} f(x_0) \cdot \nabla_{x_0}. \tag{1.5}$$

Lemma 1.2.1. *Let $x(x_0, t)$ be the solution to* (1.1)*, then*

$$f(x(x_0, t)) = \nabla_{x_0} x(x_0, t) f(x_0). \tag{1.6}$$

Proof. First we define the difference of the above two terms as

$$F(x_0, t) \overset{\text{def}}{=} f(x(x_0, t)) - \nabla_{x_0} x(x_0, t) f(x_0). \tag{1.7}$$

It is not difficult to verify that

$$F(x_0, 0) = f(x_0) - \nabla_{x_0} x_0 f(x_0) = 0. \tag{1.8}$$

Direct computation yields,

$$\begin{aligned}
\partial_t F(x_0, t) =& \nabla_x f(x(x_0, t)) \partial_t x(x_0, t) - \nabla_{x_0} \partial_t x(x_0, t) f(x_0) \\
=& \nabla_x f(x(x_0, t)) f(x(x_0, t)) - \nabla_x \partial_t x(x_0, t) \nabla_{x_0} x(x_0, t) f(x_0) \\
=& \nabla_x f(x(x_0, t)) f(x(x_0, t)) - \nabla_x f(x(x_0, t)) \nabla_{x_0} x(x_0, t) f(x_0) \\
=& \nabla_x f(x(x_0, t)) \left[f(x(x_0, t)) - \nabla_{x_0} x(x_0, t) f(x_0) \right] \\
=& \nabla_x f(x(x_0, t)) F(x_0, t).
\end{aligned} \tag{1.9}$$

Combined with (1.8), one can see $F(x_0, t) \equiv 0$. □

With Lemma 1.2.1, we can write the dynamics of $a(t)$ in a compact form with the Koopman operator $e^{t\mathcal{L}}$ [32] that propagates an observable.

Lemma 1.2.2. *Let $a(t) = \varphi(x(x_0, t))$, then*

$$a(t) = e^{t\mathcal{L}} a, \tag{1.10}$$

Proof. To verify the expression, we first take a look at the initial value, $a(0) = a$. Using Lemma 1.2.1, direct computation yields,

$$\begin{aligned}
\dot{a}(t) &= \partial_t \varphi(x(x_0, t)) \\
&= \nabla_x \varphi(x(x_0, t)) f(x(x_0, t)) \\
&= \nabla_x \varphi(x(x_0, t)) \nabla_{x_0} x(x_0, t) f(x_0) \\
&= \nabla_{x_0} \varphi(x(x_0, t)) f(x_0) \\
&= f(x_0) \cdot \nabla_{x_0} \varphi(x(x_0, t)) \\
&= f(x_0) \cdot \nabla_{x_0} a(t) = \mathcal{L} a(t).
\end{aligned} \tag{1.11}$$

One can formally write the solution as $a(t) = e^{t\mathcal{L}} a$. □

Further, we assume that the component of a belongs to a Hilbert space $L^2(\cdot, \rho_0)$,

representing the L^2 inner-product space weighted by a density function ρ_0. We define the average $\langle\cdot\rangle$ and the correlation matrix $\langle\cdot,\cdot\rangle$ on $L^2(\cdot,\rho_0)$ component-wisely. For two coarse-grained variables with the same dimension, $a(t)=\varphi(x_0,t)$ and $b(t)=\psi(x_0,t)$,

$$\begin{aligned}\langle a(t)\rangle_i &\overset{\text{def}}{=} \int_{\mathbb{X}} \varphi_i(x_0,t)\rho_0(x_0)dx_0, && 1\le i\le n,\\ \langle a(t),b^{\intercal}(t)\rangle_{ij} &\overset{\text{def}}{=} \int_{\mathbb{X}} \varphi_i(x_0,t)\psi_j(x_0,t)\rho_0(x_0)dx_0, && 1\le i,j\le n.\end{aligned} \tag{1.12}$$

Here ρ_0 is the initial configuration density function of (1.1) and is compactly supported in $\mathbb{X}=\mathbb{R}^N$. The most popular choice is the canonical ensemble in (1.3), for example, by Mori [33]. However, in principle, ρ_0 can be more general, and even out-of-equilibrium. Several choices have been suggested by Zwanzig [34].

Chapter 2
Projection formalism and orthogonal equations

The *Mori-Zwanzig* (MZ) formalism is a systematic methodology that has shown great promise in reducing the dimension of a large system [30, 31, 35]. By selecting appropriate projection operators, MZ procedure yields an exact constitutive relation in the form of a *generalized Langevin equation* (GLE), where, unlike conventional homogenization approaches [7], memory effects are taken into consideration [5].

2.1 Dyson's formula and orthogonal equations

For two operators A and B, the exponentials e^{tA} and e^{tB} could be interpreted as C_0-semigroup, also known as strongly continuous one-parameter semigroup [32], generated by A and B respectively.

Lemma 2.1.1. *The following equation holds in the common domain of semi-groups,*

$$e^{t(A+B)} = e^{tA} + \int_0^t e^{(t-s)(A+B)} B e^{sA}\, ds. \tag{2.1}$$

Let $\mathcal{P}$ be a projection operator on $L^2(\cdot, \rho_0)$ and $\mathcal{Q}$ is the complementary operator, *i.e.*, $\mathcal{Q} = \mathcal{I} - \mathcal{P}$. One is able to split the dynamics of $a(t)$ at the level of operator.

Lemma 2.1.2.

$$\dot{a}(t) = e^{t\mathcal{L}}\mathcal{P}\mathcal{L}a + \int_0^t e^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}e^{s\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}a + e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}a, \tag{2.2}$$

Proof. According to Lemma 1.2.2, we have

$$\dot{a}(t) = e^{t\mathcal{L}}\mathcal{L}a = e^{t\mathcal{L}}\mathcal{P}\mathcal{L}a(t) + e^{t\mathcal{L}}\mathcal{Q}\mathcal{L}a(t). \tag{2.3}$$

Let us take $A = \mathcal{Q}\mathcal{L}$ and $B = \mathcal{P}\mathcal{L}$ in (2.1) and plug it into (2.3). Direct computation yields (2.2). □

2.2 Mori-Zwanzig formalism and its properties

A key step in the MZ formulation is a projection operator $\mathcal{P}$ that maps functions to a preassigned direction. Once the projection $\mathcal{P}$ is in place, (2.2) could be further simplified and more properties of the GLE would be revealed.

2.2.1 Mori's projection

Let us first recall the orthogonal projection suggested by Mori [30]. For each function $b(t) = \psi(x_0, t) \in L^2(\cdot, \rho_0)^n$, Mori's projection $\mathcal{P}$ is defined as follows,

$$\mathcal{P}b(t) \stackrel{\text{def}}{=} \langle b(t), a^\intercal \rangle M^{-1} a, \tag{2.4}$$

where M^{-1} is the inverse of $M = \langle a, a^\intercal \rangle$ which is the correlation function of a at initial state.

Note that the covariance matrix M only involves the one-point statistics of a and can be guaranteed to be nonsingular by carefully selecting the CG variables. In practice, this corresponds to the appropriate choice of φ so that the CG variables are not redundant. Even in the case when the CG variables are redundant, *e.g.*, when the sum of a_i is conservative and the matrix M becomes singular, the projection can still be well defined by interpreting M^{-1} as the pseudo-inverse.

The equation (2.2) reads

$$\dot{a}(t) = Sa(t) - \int_0^t \theta(t-s)a(s)ds + F(t), \tag{2.5}$$

where $S = \langle \mathcal{L}a, a^\intercal \rangle M^{-1}$ is a constant matrix, $\theta(t) = -\langle \mathcal{L}F(t), a^\intercal \rangle M^{-1}$ is the memory kernel and $F(t) = e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}a$ represents the noise term. Here, $e^{t\mathcal{Q}\mathcal{L}}$ will also be interpreted as a Koopman operator associated with the generator $\mathcal{Q}\mathcal{L}$. As $\mathcal{Q}\mathcal{L} = \mathcal{L} - \mathcal{P}\mathcal{L}$, it can be written as an integro-differential operator.

A very important issue is the choice of ρ_0. A natural choice is an equilibrium probability density: $\rho_0 = \rho_{eq}$, and ρ_{eq} needs to satisfy

$$\mathcal{L}^*\rho_{eq} \stackrel{\text{def}}{=} -\nabla_{x_0} \cdot (f(x_0)\rho_{eq}) = 0. \tag{2.6}$$

$\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$. The most common choice is the canonical ensemble for ρ_{eq} in (1.3). In this case, $a(t)$ is simply a stationary random process with zero average. The GLE (2.5) is still useful since it describes the fluctuation of the CG variable. When the system is near equilibrium, this serves as the first approximation. Further corrections can be made using the linear response approach [36]. In general, the initial density ρ_0 can be constructed using the maximum entropy principle, as discussed in the monograph [34]. For example, given the averages of the local energy, this approach yields a probability density that is similar to the canonical ensemble (1.3) but with non-uniform temperature.

Several properties can be deduced from the derivation. Some of them have been discussed in the original work of Mori [30]. They are summarized as follows.

Proposition 2.2.1. *Assuming that $\langle a \rangle = 0$, then the following properties hold,*

$$\begin{aligned} \langle F(t) \rangle &= 0, && \forall\, t \geq 0, \\ \langle F(t), a^\intercal \rangle &= 0, && \forall\, t \geq 0. \\ \theta(t_1 - t_2) &= \langle F(t_1), F(t_2)^\intercal \rangle M^{-1}, && \forall\, t_1, t_2 \geq 0 \text{ and } t_1 \geq t_2, \end{aligned} \tag{2.7}$$

Proof. Note that with the inner product defined above, the adjoint operator of $\mathcal{L}$ is $-\mathcal{L}$, and $\mathcal{P}$ and $\mathcal{Q}$ are self-adjoint, *i.e.*, $\langle \mathcal{L}b, c^\intercal \rangle = -\langle b, \mathcal{L}c^\intercal \rangle$ and $\langle \mathcal{P}b, c^\intercal \rangle = \langle b, \mathcal{P}c^\intercal \rangle$

for any $b, c \in L^2(\cdot, \rho_{eq})^n$.

Now for the first property, we proceed as follows,

$$\begin{aligned}\langle F(t)\rangle &= -\langle \mathcal{Q}\mathcal{L}e^{t\mathcal{Q}\mathcal{L}}a\rangle = \langle \mathcal{L}e^{t\mathcal{Q}\mathcal{L}}a\rangle\langle \mathcal{P}\mathcal{L}e^{t\mathcal{Q}\mathcal{L}}a^\intercal\rangle \\ &= \langle e^{t\mathcal{Q}\mathcal{L}}a\mathcal{L}1\rangle - \langle \mathcal{L}e^{t\mathcal{Q}\mathcal{L}}a, a^\intercal\rangle M^{-1}\langle a\rangle = 0.\end{aligned} \tag{2.8}$$

For the second property, since $\mathcal{P}a = a$ and $\mathcal{Q}F(t) = F(t)$, one can easily verify that,

$$\langle F(t), a\rangle = \langle \mathcal{Q}F(t), \mathcal{P}a^\intercal\rangle = \langle F(t), \mathcal{Q}\mathcal{P}a^\intercal\rangle = 0. \tag{2.9}$$

Finally, we recall the kernel θ after (2.5) and have

$$\begin{aligned}\theta(t_1 - t_2) &= -\langle \mathcal{L}F(t_1 - t_2), a^\intercal\rangle M^{-1} = \langle e^{(t_1-t_2)\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}a, \mathcal{L}a^\intercal\rangle M^{-1} \\ &= \langle \mathcal{Q}F(t_1), e^{t_2\mathcal{L}\mathcal{Q}}\mathcal{L}a^\intercal\rangle M^{-1} = \langle F(t_1), \mathcal{Q}e^{t_2\mathcal{L}\mathcal{Q}}\mathcal{L}a^\intercal\rangle M^{-1} \\ &= \langle F(t_1), F(t_2)^\intercal\rangle M^{-1}.\end{aligned} \tag{2.10}$$

□

The first and the third equations imply that $F(t)$ is a stationary random process in the wide sense [37], and the relation (2.10) between the random noise and the memory kernel is known as the second fluctuation-dissipation theorem (FDT) [38]. It is a necessary condition for the solution to have the correct variance. The second condition suggests that the random force and the initial value of a are uncorrelated. The last two properties have also been discussed in Mori's original paper [33].

2.2.2 An explicit representation of the random noise

In the general MZ formalism, the random noise has been expressed in a quite abstract form. The practical implementation is rather difficult in general. Here, we provide a more detailed characterization of the random noise, by embedding it in an infinite system of ordinary differential equations. Important properties can also be derived from these differential equations.

Suppose that $\{\mathcal{L}^j a\}_{j\geq 0}$ are linearly independent, by inspecting the first few terms

in the exponential operator, one finds that the random force term can be written as,

$$F(t) = \sum_{j\geq 0} C_j(t)\mathcal{L}^j a. \tag{2.11}$$

Together with the orthogonal dynamics, $\dot{F}(t) = \mathcal{Q}\mathcal{L}F(t)$, we can derive a set of equations for the coefficients,

$$\begin{aligned} \dot{C}_j(t) =& C_{j-1}(t), \quad j \geq 1, \\ \dot{C}_0(t) =& -\sum_{j\geq 0} C_j(t) M_{j+1} M^{-1}, \end{aligned} \tag{2.12}$$

where M_j are referred to as the *moments* associated with the statistics of a, defined as follows,

$$M_j \overset{\text{def}}{=} \left\langle \mathcal{L}^j a, a \right\rangle = \frac{d^j}{dt^j} \left\langle a(t), a \right\rangle |_{t=0}. \tag{2.13}$$

The initial condition is given by,

$$\begin{aligned} C_0(0) &= -M_1 M^{-1}, \\ C_1(0) &= I, \\ C_j(0) &= 0, \qquad \forall j \geq 2. \end{aligned} \tag{2.14}$$

Therefore, the random noise can be characterized via an infinite set of ordinary differential equations.

2.2.3 Properties of the kernel function

Thanks to the explicit representation of the random force and the FDT (2.7), certain values of the memory kernel can be reconstructed or approximated using the equilibrium properties.

Direct calculations yield,

$$\theta(t) = -\sum_{j\geq 0} C_j(t) M_{j+1} M^{-1}, \tag{2.15}$$

where $C_j(t)$ are the coefficients of $F(t)$ in the expansion (2.11) and are given by the

solution of (2.12). The derivatives of $\theta(t)$ at $t = 0$ can be written out explicitly, as shown by the following theorem, which can be proved by direct substitutions.

Theorem 2.2.2. *Let $\mathcal{M}_j$ be the normalized moments associated with a,*

$$\mathcal{M}_j \overset{def}{=} M_j M^{-1}, \tag{2.16}$$

and $C_0^{(-1)} \overset{def}{=} C_1$. $C_0^{(k)}(0)$ can be computed recursively, and they satisfy

$$C_0^{(k)}(0) = -\sum_{j=-1}^{k-1} C_0^{(j)}(0)\mathcal{M}_{k-j-1}, \quad k \geq 0. \tag{2.17}$$

Furthermore, the derivatives of the memory kernel are given by,

$$\theta^{(k)}(0) = -\sum_{j=-1}^{k} C_0^{(j)}(0)\mathcal{M}_{k-j+1}, \quad k \geq 0. \tag{2.18}$$

As a corollary, one can write down the first few derivatives of $\theta(0)$,

$$\begin{aligned}
\theta(0) =& -\mathcal{M}_2 + \mathcal{M}_1^2 \\
\theta'(0) =& -\mathcal{M}_3 + \mathcal{M}_2\mathcal{M}_1 + \mathcal{M}_1\mathcal{M}_2 - \mathcal{M}_1^3 \\
\theta''(0) =& -\mathcal{M}_4 + \mathcal{M}_3\mathcal{M}_1 + \mathcal{M}_2^2 + \mathcal{M}_1\mathcal{M}_3 \\
& - \mathcal{M}_2\mathcal{M}_1^2 - \mathcal{M}_1\mathcal{M}_2\mathcal{M}_1 - \mathcal{M}_1^2\mathcal{M}_2 + \mathcal{M}_1^4 \\
& \cdots\cdots
\end{aligned} \tag{2.19}$$

2.3 Oblique projection formalism

From observation in (2.5), modeling the non-Gaussian statistics is challenging in the Mori's approach since the CG variables are projected to their initial states, which yield a linear constitutive model of a. Driven by additive Gaussian white noise, the stochastic process $a(t)$ is supposed to be Gaussian, which makes it difficult to recover non-Gaussian statistics unless non-trivial approximations are introduced for the random noise.

Nonlinear models can be obtained from the Zwanzig's approach [31], by employing the conditional expectation as the projection. However, the memory term in Zwanzig's formulation cannot be written as a convolution, which leads to the difficulty in approximating the memory term. To alleviate these modeling difficulties, we have developed another projection formalism in order to reconcile the two approaches. This method ensures consistency with the true statistics and gives light to noise approximations. We will refer to this projection *oblique* projection.

More specifically, let variable b be a general function of a in the same dimension. For example, if a obeys a distribution $\rho(a) = \exp[-W(a)]$, $b = -\nabla W(a)$ can be interpreted as the driving force of a. We define the oblique projection as a Mori-type projection in a preselected direction. For $c \in L^2(\cdot, \rho_0)^n$,

$$\mathcal{P}c \stackrel{\text{def}}{=} \langle c, b^\intercal \rangle \langle b, b^\intercal \rangle^{-1} b, \tag{2.20}$$

which results in a GLE in the form of

$$\dot{a}(t) = Sb(t) - \int_0^t \theta(t-s)b(s) + F(t), \tag{2.21}$$

where $M = \langle b, b^\intercal \rangle$ is the new mass matrix and $S = \langle \mathcal{L}a, b^\intercal \rangle M^{-1}$ is a constant matrix, $\theta(t) = -\langle \mathcal{L}F(t), b^\intercal \rangle M^{-1}$ is the memory kernel and $F(t) = e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}a$ represents the noise term. The kernel function is proportional to the correlation of noises induced by a and b. This model collapses to Mori's approach if $b = a$ when the local entropy $W(a)$ is quadratic, but in general, the oblique projection yields a nonlinear stochastic model.

This routine allows us to express the values of the kernel function at $t = 0$ in terms of the equilibrium statistics of the CG variables. The approximation scheme in Chapter 4 will take advantage of these properties. Recall that $\theta(t) = -\langle \mathcal{L}e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}a, b \rangle M^{-1}$.

Direct computation yields,

$$\begin{aligned}
\theta(0) &= -\langle \mathcal{L}\mathcal{Q}\mathcal{L}a, b^{\intercal}\rangle M^{-1}, \\
\theta'(0) &= -\langle \mathcal{L}(\mathcal{Q}\mathcal{L})^2 a, b^{\intercal}\rangle M^{-1}, \\
\theta''(0) &= -\langle \mathcal{L}(\mathcal{Q}\mathcal{L})^3 a, b^{\intercal}\rangle M^{-1}, \\
&\cdots\cdots
\end{aligned} \tag{2.22}$$

Let $\mathcal{M}_j$ be the normalized moments associated with the statistics of a and b, *i.e.*,

$$\mathcal{M}_j \overset{\text{def}}{=} \langle \mathcal{L}^j a, b\rangle M^{-1}. \tag{2.23}$$

Using the notation of $\mathcal{M}_j$, we are able to further simplify the above derivatives. The first few terms are listed as follows,

$$\begin{aligned}
\theta(0) =& -\mathcal{M}_2 + \mathcal{M}_1^2, \\
\theta'(0) =& -\mathcal{M}_3 + \mathcal{M}_2\mathcal{M}_1 + \mathcal{M}_1\mathcal{M}_2 - \mathcal{M}_1^3, \\
\theta''(0) =& -\mathcal{M}_4 + \mathcal{M}_3\mathcal{M}_1 + \mathcal{M}_2^2 + \mathcal{M}_1\mathcal{M}_3 \\
&- \mathcal{M}_2\mathcal{M}_1^2 - \mathcal{M}_1\mathcal{M}_2\mathcal{M}_1 - \mathcal{M}_1^2\mathcal{M}_2 + \mathcal{M}_1^4, \\
&\cdots\cdots
\end{aligned} \tag{2.24}$$

Given statistics of initial states, the above quantities can be accurately calculated.

Chapter 3
Asymptotic analysis of the memory kernel

As discussed in the last chapter, the MZ formalism yields an *exact* equation, and with appropriate selection of the projection operators, the equation can be written as a generalized Langevin equation. The generalized Langevin equations have opened up a new paradigm in model reduction problems, *e.g.*, in molecular modeling. Numerous attempts have been made to derive the generalized Langevin equations from the full molecular models, and the models have been applied to solid-gas interface [39], protein dynamics [27, 40–43], crystalline solids [16, 44, 45], etc.

A distinct feature of the generalized Langevin equations is an integral that incorporates the history-dependence, and a random noise term, which satisfies the second fluctuation-dissipation theorem [38]. The memory kernel plays an essential role in describing the dynamics of coarse-grained equations. In general, little is known about the mathematical properties of the memory kernel since it involves the projected dynamics generated by $\mathcal{QL}$ and cannot be computed explicitly without solving the full dynamics (1.1) or calculations with the similar computational cost.

A very interesting experimental work [46, 47] has suggested that part of a protein dynamics can be described by generalized Langevin equations, in which the kernel function exhibits a decay in time that is proportional to $t^{-0.51}$.

On the other hand, it is of great practical interest to introduce approximations to the kernel function, to obtain an efficient method for solving the generalized Langevin equations. For instance, a brute force solution method would have to store the

solution for all previous steps, and then at every time step, the integral has to be evaluated, which further adds up to the overall computational cost. It is important to understand properties of the kernel functions so that appropriate approximations can be introduced.

Motivated by these observations, it is of great significance to answer the following questions:

1. How does the memory function depend on the choice of the CG variables?
2. How would the kernel function behave at different levels of coarse-graining?
3. How does the memory function decay in space and time?

To be able to quantify these aspects, we focus our analysis on a one-dimensional lattice model that has served as an excellent test model to study heat conduction, fracture, analysis of modeling errors, etc., e.g., in [48–55] to name a few and many related works.

3.1 The one-dimensional lattice dynamics model

We consider a one-dimensional infinite lattice model, in which the atoms are connected by linear springs, as shown in Figure 3.1.

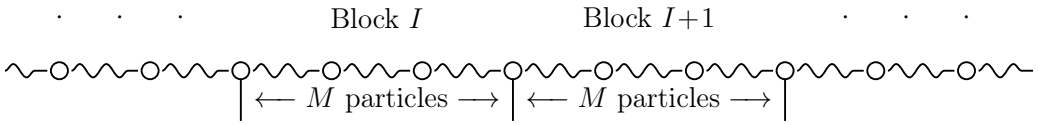

Figure 3.1: Diagram of a one-dimensional lattice chain. Particles are connected by linear springs and are divided evenly into blocks. One block contains M particles.

The dynamics of this one-dimensional chain can be described by (1.2),

$$m\ddot{r}_j = -\frac{\partial V(r)}{\partial r_j}, \quad j \in \mathbb{Z}, \tag{3.1}$$

where r_j is the position of i-th atom.

If we assume that particles have direct interactions among first and second neighbors, the potential energy can be written as

$$V(r) = \sum_{j\in\mathbb{Z}} \ell(r_{j+1} - r_j) + \ell(r_{j+2} - r_j), \tag{3.2}$$

where $\ell(r)$ is the potential energy associated with two interacting atoms of distance r. Atoms typically move around their equilibrium positions, denoted by reference positions $R_j = j\varepsilon_0$, and ε_0 is the equilibrium atom spacing (aka lattice constant). We define $u_j = r_j - j\varepsilon_0$ as the displacement and set $m = 1$ for simplicity.

Consider a first-order expansion of the right hand side of (3.1) around equilibrium positions R_j. This yields a linear dynamical system

$$\ddot{u}_j = -\mathcal{A}u_j, \tag{3.3}$$

where $\mathcal{A}$ is a finite difference operator, defined as

$$\mathcal{A}u_j \stackrel{\text{def}}{=} -\kappa_2 u_{j-2} - \kappa_1 u_{j-1} + \kappa_0 u_j - \kappa_1 u_{j+1} - \kappa_2 u_{j+2}, \tag{3.4}$$

where $\kappa_1 = \ell''(\varepsilon_0)$, $\kappa_2 = \ell''(2\varepsilon_0)$ and $\kappa_0 = 2(\kappa_1 + \kappa_2)$. This linear approximation of the atomic forces is known as the harmonic approximation [56], and it is a useful routine to obtain elastic parameters, the phonon dispersion relations, stability conditions, etc. It may seem as if this approximation is too crude, since it neglects nonlinear interactions entirely. But for crystalline solids, the accuracy of the harmonic approximation is quite reasonable, especially when the temperature is well below the melting point [56].

We assume that $\mathcal{A}$ is positive semidefinite in appropriate function spaces, which is pertinent to the stability of the lattice structure [57]. For the model (3.4), this amounts to the following phonon stability conditions:

$$\kappa_1 > 0 \text{ and } \kappa_1 + 4\kappa_2 > 0. \tag{3.5}$$

In fact, it is not difficult to verify that $\mathcal{A}$ is a discrete Laplacian operator.

We now discuss the reduction of the full dynamics model (3.3) and the emergence of the generalized Langevin equation. The first step in this effort is to map the atom

position and velocity to a set of CG variables. For this purpose, we will introduce some averaging operators. We let the one-dimensional infinite lattice be $\mathbb{L} = \varepsilon_0\mathbb{Z}$,

$$\mathbb{L} \stackrel{\text{def}}{=} \{0, \pm\varepsilon_0, \pm 2\varepsilon_0, \pm 3\varepsilon_0, \cdots\}. \tag{3.6}$$

We partition the entire lattice into blocks, each of which contains M atoms and use

$$\mathbb{L}_{\text{CG}} \stackrel{\text{def}}{=} M\mathbb{L} = \{0, \pm M\varepsilon_0, \pm 2M\varepsilon_0, \pm 3M\varepsilon_0, \cdots\} \tag{3.7}$$

to represent the coarse-grained space.

Let X be the space of functions representing the atomic displacement, i.e.,

$$X \stackrel{\text{def}}{=} \ell^2(\mathbb{L}) = \left\{ f : \mathbb{L} \to \mathbb{R},\ \sum_{x\in\mathbb{L}} f(x)^2 < \infty \right\}. \tag{3.8}$$

The CG variables will be defined with appropriate local spatial averaging. A natural idea is to extract the usual average from a block as the CG variable, whose one component is the arithmetic mean of the displacements of atoms in one block. In this case, the weights are called piecewise constant weighting functions: $1/M$ for atoms within the block, and zero otherwise, as shown in Figure 4.1.

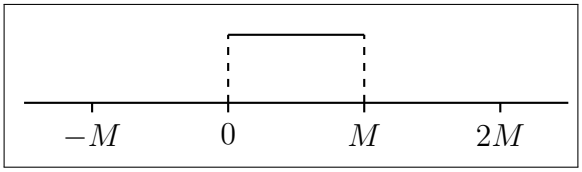

Figure 3.2: The constant weights for the averaging over one block.

On the other hand, one may choose non-uniform weighting functions, e.g., as shown in Figure 3.3, the weights are overlapping hat functions and the averaging is over two adjacent blocks. We call it piecewise linear averaging. This is motivated by the restriction operator in multigrid methods [58].

We could extend the definition of CG variables to other averaging schemes, but here we only focus on piecewise constant weights and piecewise linear weights for the averaging over blocks.

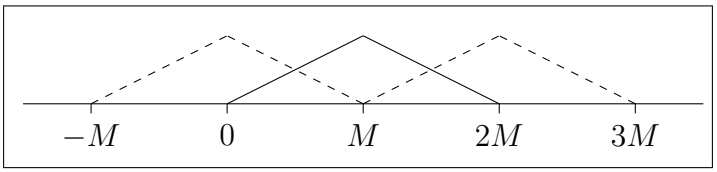

Figure 3.3: The piecewise linear weights for the averaging over blocks.

In either case, the averaging procedure is a mapping from X to X_{CG}, where X_{CG} is the function space at coarse-grained grids,

$$X_{\mathrm{CG}} \stackrel{\mathrm{def}}{=} \ell^2(\mathbb{L}_{\mathrm{CG}}) = \left\{ f : \mathbb{L}_{\mathrm{CG}} \to \mathbb{R},\ \sum_{x \in \mathbb{L}_{\mathrm{CG}}} f(x)^2 < \infty \right\}. \tag{3.9}$$

This averaging operator, denoted by $\Phi^\intercal$, when thought of as an infinite matrix, contains rows that are the weighting functions. For example, for the uniform (piecewise constant) averaging, $\Phi^\intercal$ is a block diagonal matrix,

$$\Phi^\intercal = \begin{bmatrix} \ddots & & & & \\ & (\frac{1}{M}, \cdots, \frac{1}{M}) & & & \\ & & (\frac{1}{M}, \cdots, \frac{1}{M}) & & \\ & & & \ddots & \end{bmatrix}. \tag{3.10}$$

The CG variables will be defined in a compact form using $\Phi^\intercal$. For $u, v \in X$,

$$\begin{aligned} q &= \Phi^\intercal u, \\ p &= \Phi^\intercal v, \end{aligned} \tag{3.11}$$

where $v = \dot{u}$ can be interpreted as the velocity. One can view this procedure as a local averaging, and the averaged values are defined in $\mathbb{L}_{\mathrm{CG}}$, *i.e.*, $p, q \in X_{\mathrm{CG}}$. In the context of multigrid theory, $\Phi^\intercal$ is a compression/restriction operator, and Φ corresponds to the elongation/interpolation operator.

The columns of Φ span a space $Y = \mathrm{range}(\Phi)$ and the space is a subspace of X. It is also useful to consider the orthogonal complement $Y^\perp$, which contains the 'extra'

degrees of freedom,

$$Y^{\perp} = \left\{ y^{\perp} \in X : y^{\perp} \cdot y = 0, \forall\ y \in Y \right\}, \tag{3.12}$$

where the dot represents ℓ^2 inner product in X. Let Ψ be the corresponding matrix, whose columns span $Y^{\perp}$.

Notice that we have the following unique decomposition: For each $u \in X$, $u = u_1 + u_2, u_1 \in Y$ and $u_2 \in Y^{\perp}$, and with Ψ, the decomposition can be written more explicitly as,

$$\begin{aligned} u &= \Phi q + \Psi q^{\perp}, \\ v &= \Phi p + \Psi p^{\perp}. \end{aligned} \tag{3.13}$$

Later we will see that the kernel function in the GLE does not depend on the specific choice of Ψ: It only depends on how CG variables q and p are defined, i.e., the choice of Φ.

Once the CG variables are selected, the MZ formalism can be invoked to derive a reduced model, in the form of the generalized Langevin equations,

$$\begin{cases} \dot{q} = p, \\ \dot{p} = \mathcal{S} q - \displaystyle\int_0^t \mathcal{K}(t-\tau) p(\tau) d\tau + \mathcal{F}(t). \end{cases} \tag{3.14}$$

In particular, the kernel function is given by [16, 59, 60],

$$\mathcal{K}(t) = \Phi^{\intercal} \mathcal{A} \Psi \cos(\Omega t) \Omega^{-2} \Psi^{\intercal} \mathcal{A} \Phi, \tag{3.15}$$

where $\Omega^2 = \Psi^{\intercal} \mathcal{A} \Psi$ and the cosine function is defined by the Taylor series,

$$\cos(\Omega t) = I - \frac{(\Omega t)^2}{2!} + \frac{(\Omega t)^4}{4!} - \frac{(\Omega t)^6}{6!} + \cdots. \tag{3.16}$$

Furthermore, $\mathcal{F}(t)$ is a stationary process. In the finite-dimensional case, it can be proved to be Gaussian [16, 59, 61]. The generalization of this property to an infinite dimensional system may require a Gaussian measure for the equilibrium distribution. Implicit in the equation (3.14), there also exists the relation between the kernel

function $\mathcal{K}$ and the random force for the second fluctuation-dissipation theorem, as shown in (2.7),

$$\left\langle \mathcal{F}(t), \mathcal{F}(t')^{\intercal} \right\rangle = k_B T \mathcal{K}(t - t'). \tag{3.17}$$

This condition is necessary for the system to equilibrate to its correct probability distribution. Although the random noise will not be under our consideration in this chapter, it is still important to recognize the fact that the kernel function is a covariance matrix. For the detailed derivations, see [16, 45, 59]. Similar derivations can be found under other settings [39, 60, 61].

3.2 Finite-dimensional treatment for the lattice structure

It is worthwhile to point out that we have selected the infinite lattice dynamics as the full model. This is because, for finite systems, the kernel function often exhibits a 'recurrence' phenomenon. Namely, the value of the kernel function first decays, and then after some time, the values go back up. The recurrence time increases as the system size increases, but it is difficult to have a decay property within a finite system.

Notice that the kernel function (3.15) represents an operator from $\ell^2\big(\mathbb{L}_{\mathrm{CG}}\big)$ to itself, which is an infinite dimensional space. As alluded beforehand, it is useful to understand how the kernel function decays in space and time, and also the dependence on the level of coarse-graining, i.e., M. The problem would become handy if the infinite dimensional system can be converted into a finite-dimensional one. From the definitions, operators $\mathcal{A}$ and Φ are Toeplitz forms [62] and exhibit translational symmetry. The Fourier transformation in terms of blocks is a useful tool to achieve this goal.

We split $\mathcal{A}$ into block matrices of size M-by-M. Since atoms only interact within the same block or neighboring blocks, overall, $\mathcal{A}$ has a tridiagonal block structure. In

fact, $\mathcal{A}$ is a Toeplitz form and can be expressed as

$$\mathcal{A} = \begin{bmatrix} \ddots & \ddots & \ddots & & & & \\ \cdots & A_1 & A_0 & A_{-1} & 0 & 0 & \cdots \\ \cdots & 0 & A_1 & A_0 & A_{-1} & 0 & \cdots \\ \cdots & 0 & 0 & A_1 & A_0 & A_{-1} & \cdots \\ & & & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{3.18}$$

The diagonal block, denoted by A_0, is a band matrix with bandwidth 5, containing the force constants,

$$A_0 = \begin{bmatrix} \kappa_0 & -\kappa_1 & -\kappa_2 & 0 & 0 & 0 & 0 & \cdots & 0 \\ -\kappa_1 & \kappa_0 & -\kappa_1 & -\kappa_2 & 0 & 0 & 0 & \cdots & 0 \\ -\kappa_2 & -\kappa_1 & \kappa_0 & -\kappa_1 & -\kappa_2 & 0 & 0 & \cdots & 0 \\ 0 & -\kappa_2 & -\kappa_1 & \kappa_0 & -\kappa_1 & -\kappa_2 & 0 & \cdots & 0 \\ 0 & 0 & -\kappa_2 & -\kappa_1 & \kappa_0 & -\kappa_1 & -\kappa_2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & -\kappa_2 & -\kappa_1 & \kappa_0 & -\kappa_1 & -\kappa_2 \\ 0 & 0 & \cdots & \cdots & 0 & -\kappa_2 & -\kappa_1 & \kappa_0 & -\kappa_1 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & -\kappa_2 & -\kappa_1 & \kappa_0 \end{bmatrix}. \tag{3.19}$$

The off-diagonal block, denoted by A_1, only has three non-zero entries,

$$\begin{aligned} A_1(1, M) &= -\kappa_1, \\ A_1(1, M-1) &= A_1(2, M) = -\kappa_2, \end{aligned}$$

and $A_{-1} = A_1^\intercal$.

In accordance with the partition of $\mathcal{A}$, we break $g \in X$ into blocks, with one block containing M elements and assemble blocks into $\boldsymbol{g}$, with $\boldsymbol{g}_n \in \mathbb{R}^M$ being the nth

component of $\boldsymbol{g}$, i.e.,

$$g = \begin{bmatrix} \vdots \\ \boldsymbol{g}_{n-1} \\ \boldsymbol{g}_n \\ \boldsymbol{g}_{n+1} \\ \vdots \end{bmatrix}, \quad \boldsymbol{g}_n = \begin{bmatrix} g_{nM} \\ g_{nM+1} \\ \vdots \\ g_{(n+1)M-1} \end{bmatrix}. \tag{3.20}$$

Now we can define the Fourier transform of g at the level of blocks,

$$\widehat{\boldsymbol{g}}(\xi) = \sum_n e^{-in\xi} \boldsymbol{g}_n, \qquad \xi \in (0, 2\pi], \tag{3.21}$$

with inverse given by,

$$\boldsymbol{g}_n = \frac{1}{2\pi} \int_0^{2\pi} \widehat{\boldsymbol{g}}(\xi) e^{in\xi} d\xi. \tag{3.22}$$

Similarly, one can also define the Fourier transform of Toeplitz forms in terms of its matrix representation. Suppose $\mathcal{G}$ is a Toeplitz form, generated by one of its columns G:

$$\begin{bmatrix} \vdots \\ G_{-1} \\ G_0 \\ G_1 \\ \vdots \end{bmatrix} \tag{3.23}$$

The Fourier transform of the operator G is defined as

$$\widehat{\mathcal{G}}(\xi) = \sum_n e^{-in\xi} G_n, \qquad \xi \in (0, 2\pi]. \tag{3.24}$$

Based on the block form, the Fourier transformation of $\mathcal{A}$ is simply defined as

$$\widehat{\mathcal{A}}(\xi) = A_0 + e^{-i\xi} A_1 + e^{i\xi} A_1^\intercal, \qquad \xi \in (0, 2\pi]. \tag{3.25}$$

From the perspective in solid state physics, grouping atoms in blocks is equivalent to building a complex lattice with M atoms in one primitive cell. The Fourier domain

$(0, 2\pi]$ is equivalent to the first Brillouin zone $(-\pi, \pi]$.

In terms of the Fourier transforms, it is not difficult to establish the following properties,

$$\widehat{\mathcal{A}f} = \widehat{\mathcal{A}}\widehat{f} \quad \text{and} \quad \widehat{\mathcal{G}\mathcal{H}} = \widehat{\mathcal{G}}\widehat{\mathcal{H}} \tag{3.26}$$

where $\mathcal{G}$ and $\mathcal{H}$ are Toeplitz operators with consistent domains and ranges.

We now apply Fourier transform to the kernel function in (3.15). With direct computation, we find that the (k, ℓ)-th entry of $\mathcal{K}(t)$ can be written as

$$\mathcal{K}_{k\ell}(t) = \frac{1}{2\pi} \int_0^{2\pi} e^{i(\ell-k)\xi} \widehat{\Phi}^* \widehat{\mathcal{A}} \widehat{\Psi} \cos(\widehat{\Omega} t) \left(\widehat{\Psi}^* \widehat{\mathcal{A}} \widehat{\Psi} \right)^{-1} \widehat{\Psi}^* \widehat{\mathcal{A}} \widehat{\Phi} \, d\xi, \tag{3.27}$$

where $\widehat{\Phi}, \widehat{\Psi}$ and $\widehat{\Omega}$ are defined based on the Fourier transformation in (3.24). Our asymptotic analysis will be established based on the representation of the kernel function in (3.27).

3.3 The dependence on the block size M

The kernel function depends on the choice of the CG variables, which can be defined based on the averaging methods within the blocks. The kernel function also depends on the block size M, which shows the level of coarse-graining.

Here, the dependence of the block size is discussed for two types of averaging, piecewise constant averaging and piecewise linear averaging, which are two most common and intuitive ways in finite element methods [58]. We first study the asymptotic behavior of $\mathcal{K}_{0,0}(0)$ in the kernel function as the block size M goes to infinity.

3.3.1 Piecewise constant weighting functions

Recall that Φ is a Toeplitz form, arising from the piecewise constant averaging. Φ can also be seen as an infinite dimensional block diagonal matrix, with diagonal block equal to

$$Q_1 = \frac{1}{\sqrt{M}} (1, 1, \cdots, 1)^\intercal, \tag{3.28}$$

and $Q_1 \in \mathbb{R}^M$. It is not difficult to find $Q_2 \in \mathbb{R}^{M\times(M\text{-}1)}$ whose column vectors are a set of orthonormal basis in the complementary subspace spanned by Q_1, i.e., $[Q_1\ Q_2] \in \mathbb{R}^{M\times M}$ is an orthonormal matrix. Ψ can be defined as a block diagonal form with diagonal block equal to Q_2. As a result, $\widehat{\Phi} = Q_1$ and $\widehat{\Psi} = Q_2$.

According to (3.27), in terms of Q_1 and Q_2, the diagonal entry of $\mathcal{K}$ can be written as

$$\mathcal{K}_{0,0}(0) = \frac{1}{2\pi}\int_0^{2\pi} Q_1^{\intercal}\widehat{\mathcal{A}}Q_2\left(Q_2^{\intercal}\widehat{\mathcal{A}}Q_2\right)^{-1}Q_2^{\intercal}\widehat{\mathcal{A}}Q_1\ d\xi. \tag{3.29}$$

On the right-hand side of (3.29), Q_1 comes from the weighting functions and is known to us. Provided the orthogonality of Q_1 and Q_2, the form of Q_2 is not unique. It is necessary to show that $\mathcal{K}_{0,0}(0)$ in (3.29) does not depend on the specific choice of Q_2. Another difficulty is that the expression contains the inverse of $Q_2^{\intercal}\widehat{\mathcal{A}}Q_2$ which is not easy to compute in general.

Fortunately, we have found that Q_1 and Q_2 have properties that allow us to resolve these issues. We begin with the following formula.

Lemma 3.3.1 (Inverses of 2-by-2 block matrices [63])**.** *Suppose C and D are both 2-by-2 block matrices. C_{11} and D_{11} belong to $\mathbb{C}^{n_1\times n_1}$ and C_{22} and D_{22} belong to $\mathbb{C}^{n_2\times n_2}$. If $CD = I$ and D_{11}, D_{22} are invertible, then the Shur complement $D_{22} - D_{21}D_{11}^{-1}D_{12}$ is invertible and*

$$C_{22} = \left(D_{22} - D_{21}D_{11}^{-1}D_{12}\right)^{-1}. \tag{3.30}$$

From direct calculations, the following lemma can be established.

Lemma 3.3.2. *Let (λ_j, w_j) be an eigenpair of $\widehat{\mathcal{A}}$, namely, λ_j is a scalar, $w_j \in \mathbb{C}^M$, and $\widehat{\mathcal{A}}w_j = \lambda_j w_j$. Then*

$$\lambda_j = 2(\kappa_1+\kappa_2) - 2\kappa_1\cos(\xi_j') - 2\kappa_2\cos(2\xi_j'), \tag{3.31}$$

$$w_j = \frac{1}{\sqrt{M}}\begin{bmatrix} 1 \\ e^{i\xi_j'} \\ e^{2i\xi_j'} \\ \vdots \\ e^{i(M-1)\xi_j'} \end{bmatrix}, \quad \xi_j' = \frac{-\xi + 2j\pi}{M}, \quad j = 1, 2, \cdots, M. \tag{3.32}$$

From Lemma 3.3.2 it can be seen that the stability condition (3.5) guarantees that $\widehat{\mathcal{A}}$ is positive semidefinite for any ξ in $(0, 2\pi]$.

Theorem 3.3.3. *If the CG variables are defined based on the piecewise constant averaging operator, then the following estimate holds,*

$$0 \leq \mathcal{K}_{0,0}(0) \leq \frac{2\kappa_1 + 4\kappa_2}{M}. \tag{3.33}$$

Proof. From Lemma 3.3.2, $\widehat{\mathcal{A}}$ has the zero eigenvalue only when $\xi = 2\pi$, and its corresponding eigenvector is equal to $\widehat{\Phi} = Q_1$. As a result, $\widehat{\Phi}^* \widehat{\mathcal{A}} \widehat{\Phi} \neq 0$ for $\xi \in (0, 2\pi)$. $\widehat{\Phi}$ is perpendicular to $\widehat{\Psi}$, so $\widehat{\Psi}^* \widehat{\mathcal{A}} \widehat{\Psi}$ is non-singular for $\xi \in (0, 2\pi]$.

For $\xi \neq 2\pi$, $\widehat{\mathcal{A}}$ is invertible and from the orthogonality of Q_1 and Q_2,

$$\left(\begin{bmatrix} Q_1^\intercal \\ Q_2^\intercal \end{bmatrix} \widehat{\mathcal{A}} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \right)^{-1} = \begin{bmatrix} Q_1^\intercal \\ Q_2^\intercal \end{bmatrix} \widehat{\mathcal{A}}^{-1} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}. \tag{3.34}$$

Applying Lemma 3.3.1 to the above matrices, we arrive at the following equality,

$$\left(Q_2^\intercal \widehat{\mathcal{A}} Q_2 \right)^{-1} = Q_2^\intercal \widehat{\mathcal{A}}^{-1} Q_2 - Q_2^\intercal \widehat{\mathcal{A}}^{-1} Q_1 \left(Q_1^\intercal \widehat{\mathcal{A}}^{-1} Q_1 \right)^{-1} Q_1^\intercal \widehat{\mathcal{A}}^{-1} Q_2 \ . \tag{3.35}$$

Substituting (3.35) into (3.29), and using the facts that $Q_1 Q_1^\intercal + Q_2 Q_2^\intercal = I$ and $\left(Q_1^\intercal \widehat{\mathcal{A}}^{-1} Q_1 \right)^{-1}$ is a scalar, one obtains

$$\mathcal{K}_{0,0}(0) = \frac{1}{2\pi} \int_0^{2\pi} Q_1^\intercal \widehat{\mathcal{A}} Q_1 - (Q_1^\intercal \widehat{\mathcal{A}}^{-1} Q_1)^{-1} \, \mathrm{d}\xi. \tag{3.36}$$

In light of (3.27) and the eigen decomposition in Lemma 3.3.2, we know that $\mathcal{K}_{0,0}(0)$, $Q_1^\intercal \widehat{\mathcal{A}} Q_1$ and $(Q_1^\intercal \widehat{\mathcal{A}}^{-1} Q_1)^{-1}$ are real and non-negative. Thus, it is sufficient to use the integral of the first term to bound $\mathcal{K}_{0,0}(0)$,

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} Q_1^\intercal \widehat{\mathcal{A}} Q_1 \, \mathrm{d}\xi &= \frac{1}{2\pi} \int_0^{2\pi} Q_1^\intercal \left(A_0 + e^{i\xi} A_1 + e^{-i\xi} A_1^\intercal \right) Q_1 \, \mathrm{d}\xi \\ &= Q_1^\intercal A_0 Q_1 = \frac{2\kappa_1 + 4\kappa_2}{M}. \end{aligned} \tag{3.37}$$

Consequently, $\mathcal{K}_{0,0}(0)$ can be bounded by the integral,

$$0 \leq \mathcal{K}_{0,0}(0) \leq \frac{1}{2\pi} \int_0^{2\pi} Q_1^{\intercal} \widehat{\mathcal{A}} Q_1 \, \mathrm{d}\xi = \frac{2\kappa_1 + 4\kappa_2}{M}. \tag{3.38}$$

□

The second fluctuation-dissipation theorem (3.17) indicates that $\mathcal{K}_{0,0}(0)$ is proportional to the variance of the random noise. This estimate shows that the magnitude of the noise is inverse proportional to the level of coarse-graining.

Due to (3.17) and Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathcal{K}_{I,J}^2(0) &= (k_B T)^{-2} \left\langle \mathcal{F}_I(0), \mathcal{F}_J(0) \right\rangle^2 \\ &\leq (k_B T)^{-2} \left\langle \mathcal{F}_I^2(0) \right\rangle \left\langle \mathcal{F}_J^2(0) \right\rangle \\ &= \mathcal{K}_{I,I}(0) \mathcal{K}_{J,J}(0). \end{aligned} \tag{3.39}$$

Since $\mathcal{K}$ is a Toeplitz form, we have $\mathcal{K}_{I,I} = \mathcal{K}_{J,J} = \mathcal{K}_{0,0}$. This implies that off-diagonal entries can be bounded by the diagonal entry and our estimate can be extended,

$$|\mathcal{K}_{I,J}(0)| \leq |\mathcal{K}_{0,0}(0)| = \mathcal{O}(M^{-1}). \tag{3.40}$$

Here we present a numerical experiment for the one-dimensional lattice model in Figure 3.4 to verify this estimate. The force constants are obtained from a Morse potential with $\kappa_1 = 12.2676$ and $\kappa_2 = 3.0628$. We generate a large enough system with periodic boundary conditions to imitate the one-dimensional infinite chain. When the number of atoms is large enough, the boundary conditions will have little effect on interior atoms within a relatively short period. It is reasonable to use this finite but large system to approximate the exact model.

We fix the total number of the atoms $N = 2^{13}$ in the simulation and vary the block size $M = 2^4, 2^5, \cdots, 2^{10}$, which represents the scale of averaging. We assemble M atoms as a block. From Figure 3.4, we observe that the estimate is quite sharp. The diagonal entry of the kernel function $\mathcal{K}_{0,0}(0)$ is observed to decay at the rate of $\mathcal{O}(M^{-1})$ as $M \to +\infty$.

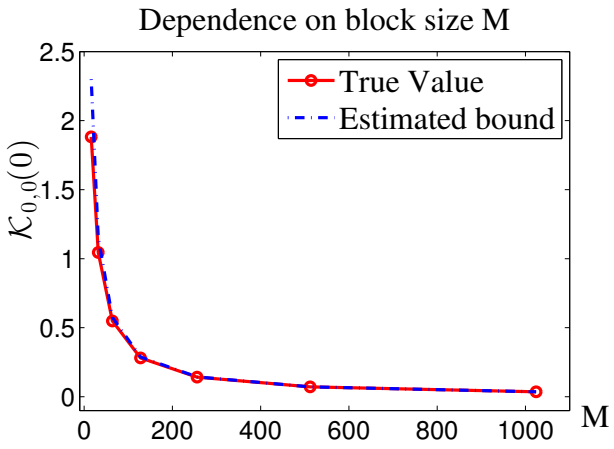

Figure 3.4: The dependence of the kernel function on the block size for the piecewise constant averaging operator.

3.3.2 Piecewise linear weighting functions

If the CG variables are defined by the piecewise linear weighting functions, as shown in Figure 3.3, the operator Φ is not block diagonal, and its Fourier transform is no longer a real-valued matrix. It becomes more difficult to simplify (3.27).

To deal with this issue, we first define two vectors h_1, $h_2 \in \mathbb{R}^M$ as follows,

$$h_1 = \frac{1}{c}\begin{bmatrix} 1 \\ 2 \\ \vdots \\ M \end{bmatrix} \quad \text{and} \quad h_2 = \frac{1}{c}\begin{bmatrix} M-1 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \tag{3.41}$$

where $c = \sqrt{\frac{1}{3}(2M^3 + M)}$ is a constant to normalize Φ. h_1 and h_2 represent the two sides of the hat function in Figure 3.3.

Notice that the operator Φ is lower block bidiagonal, and the diagonal block is

equal to h_1 and the first off-diagonal block is equal to h_2, i.e.,

$$\Phi = \begin{bmatrix} \ddots & & & & \\ \ddots & h_1 & & & \\ & h_2 & h_1 & & \\ & & h_2 & h_1 & \\ & & & \ddots & \ddots \end{bmatrix}. \tag{3.42}$$

We seek Ψ in a lower block bidiagonal form and denote its diagonal block as H_1 and the first off-diagonal block as H_2, and both H_1 and H_2 are in $\mathbb{R}^{M\times(M-1)}$. To make $[\Phi\ \Psi]$ a set of basis of $X = \ell^2(\mathbb{L})$, H_1 and H_2 should satisfy

$$\begin{cases} h_1^\intercal H_1 + h_2^\intercal H_2 = 0, \\ h_1^\intercal H_2 = 0, \\ h_2^\intercal H_1 = 0. \end{cases} \tag{3.43}$$

The above equations hold as long as columns of the matrix

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \tag{3.44}$$

belong to the solution set of the following linear equations

$$\begin{bmatrix} h_1^\intercal & h_2^\intercal \\ h_2^\intercal & 0 \\ 0 & h_1^\intercal \end{bmatrix} y = 0. \tag{3.45}$$

The dimension of the solution space is $2M - 2$.

Meanwhile, we need to ensure that the matrix H has full column rank, i.e., $\mathrm{rank}(H) = M - 1$, to guarantee that Ψ spans the whole complementary space of Φ. Then one has

$$\widehat{\Phi} = h_1 + e^{-i\xi} h_2, \quad \widehat{\Psi} = H_1 + e^{-i\xi} H_2. \tag{3.46}$$

Let $R \in \mathbb{C}^{(M\text{-}1)\times(M\text{-}1)}$ be the Cholesky factor of $\widehat{\Psi}^*\widehat{\Psi}$, i.e., R is an upper triangular

matrix and $R^*R = \widehat{\Psi}^*\widehat{\Psi}$. Let $\Phi_0 = \widehat{\Phi}(\widehat{\Phi}^*\widehat{\Phi})^{-\frac{1}{2}}$ and $\Psi_0 = \widehat{\Psi}R^{-1}$. It is not difficult to verify that $[\Phi_0\ \Psi_0]$ is an orthogonal matrix.

Direct computation yields that (3.27) can be simplified as follows,

$$\mathcal{K}_{0,0}(0) = \frac{1}{2\pi}\int_0^{2\pi} \left(\widehat{\Phi}^*\widehat{\Phi}\right) \Phi_0^*\widehat{\mathcal{A}}\Psi_0\left(\Psi_0^*\widehat{\mathcal{A}}\Psi_0\right)^{-1}\Psi_0^*\widehat{\mathcal{A}}\Phi_0 \,\mathrm{d}\xi. \tag{3.47}$$

Theorem 3.3.4. *If the CG variables are defined with piecewise linear weighting functions, then the following estimate holds,*

$$0 \le \mathcal{K}_{0,0}(0) \le \frac{2M\kappa_1 + 8M\kappa_2 - 6\kappa_2}{M(2M^2+1)/3}. \tag{3.48}$$

Proof. Applying Lemma 3.3.1 to $\left(\Psi_0^*\widehat{\mathcal{A}}\Psi_0\right)^{-1}$ and following similar steps in the proof of Theorem 3.3.3, we have

$$\mathcal{K}_{0,0}(0) = \frac{1}{2\pi}\int_0^{2\pi} \widehat{\Phi}^*\widehat{\mathcal{A}}\widehat{\Phi} - \left(\widehat{\Phi}^*\widehat{\mathcal{A}}^{-1}\widehat{\Phi}\right)^{-1}\left(\widehat{\Phi}^*\widehat{\Phi}\right)^2 \,\mathrm{d}\xi. \tag{3.49}$$

Again, with the positive semidefinite property, it suffices to estimate the integral of the first term. Direct calculations yield,

$$\begin{aligned}
\frac{1}{2\pi}\int_0^{2\pi} \widehat{\Phi}^*\widehat{\mathcal{A}}\widehat{\Phi} \,\mathrm{d}\xi &= \frac{1}{2\pi}\int_0^{2\pi} \left(h_1^\intercal + h_2^\intercal e^{-i\xi}\right)\left(A_0 + A_1 e^{i\xi} + A_1^\intercal e^{-i\xi}\right)\left(h_1 + h_2 e^{i\xi}\right) \,\mathrm{d}\xi \\
&= \frac{1}{2\pi}\int_0^{2\pi} h_1^\intercal A_0 h_1 + 2h_1^\intercal A_1^\intercal h_2 + h_2^\intercal A_0 h_2 \,\mathrm{d}\xi \\
&= \frac{2M\kappa_1 + 8M\kappa_2 - 6\kappa_2}{M(2M^2+1)/3}.
\end{aligned} \tag{3.50}$$

Then we have the estimate

$$0 \le \mathcal{K}_{0,0}(0) \le \frac{1}{2\pi}\int_0^{2\pi} \widehat{\Phi}^*\widehat{\mathcal{A}}\widehat{\Phi} \,\mathrm{d}\xi = \frac{2M\kappa_1 + 8M\kappa_2 - 6\kappa_2}{M(2M^2+1)/3}. \tag{3.51}$$

□

Interestingly, with this coarse-graining scheme, $\mathcal{K}_{0,0}(0)$, the variance of the random

noise, decays more rapidly with rate $\mathcal{O}(M^{-2})$, compared to the constant averaging.

With the same setup as in the previous numerical test, we tested the dependence of $\mathcal{K}_{0,0}(0)$ on M for piecewise linear averaging numerically. We examine the dependence of the kernel function on the scale of averaging for piecewise linear averaging with different force constants. One example is obtained from the Morse potential as used before. For the second one, we choose when parameters are close to the boundary of the stability conditions (3.5). The upper bounds are sharp for both case as shown in Figure 3.5.

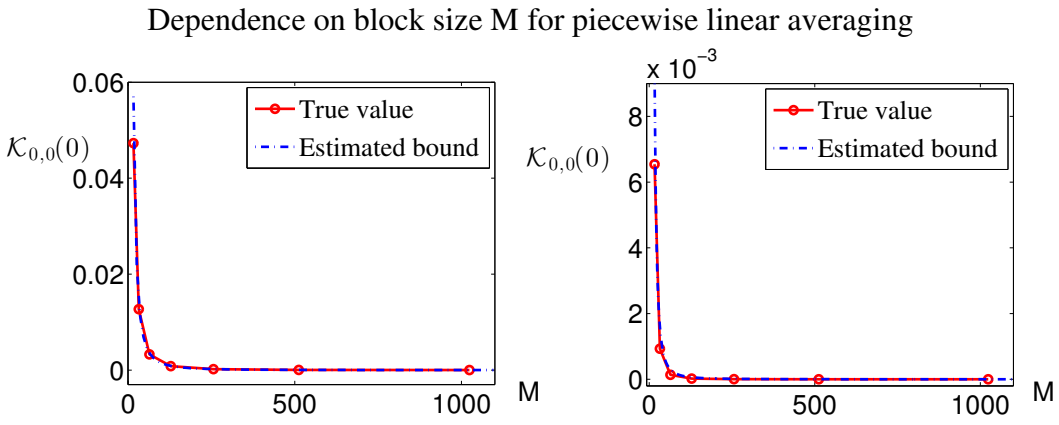

Figure 3.5: The dependence of the kernel function on the block size for the piecewise constant averaging operator. Left: $\kappa_1 = 12.2676$ and $\kappa_2 = 3.0628$. Right: $\kappa_1 = 12.2676$ and $\kappa_2 = -3$.

3.4 The spatial decay of $\mathcal{K}_{0,J}(0)$

Now let us turn to the analysis of the *off-diagonal* terms in $\mathcal{K}(0)$. In light of the second fluctuation-dissipation theorem (3.17), this entry indicates the spatial correlation of the random noise. In its matrix form, every column of Φ represents a weighting function centered at one block. The correlation between two blocks is expected to become smaller as the distance of the two blocks increases, which corresponds to the entry $\mathcal{K}_{0,J}(0)$.

For this analysis, we will fix the block size M, and focus on the behavior of the

kernel function as $|J|$ goes infinite. Only the piecewise constant averaging case is considered here.

Recall from (3.27) that we have

$$\mathcal{K}_{0,J}(0) = \frac{1}{2\pi}\int_0^{2\pi} e^{iJ\xi} Q_1^\intercal \widehat{\mathcal{A}} Q_2 \left(Q_2^\intercal \widehat{\mathcal{A}} Q_2\right)^{-1} Q_2^\intercal \widehat{\mathcal{A}} Q_1 \, \mathrm{d}\xi. \tag{3.52}$$

Lemma 3.4.1. *Let* $R(\xi) = \left(Q_1^\intercal \widehat{\mathcal{A}}^{-1} Q_1\right)^{-1}$ *on* $(0, 2\pi]$*, then* $R(\xi)$ *is a real-valued even function and it is* n *times continuously differentiable for any integer* n*, i.e.,* $R \in \mathcal{C}^\infty(0, 2\pi]$*.*

Proof. Since $\widehat{\mathcal{A}} = A_0 + e^{-i\xi}A_1 + e^{i\xi}A_1^\intercal$, it is clear that R is real-valued and even. For its differentiability, let us first consider $\xi \neq 0$ when $\widehat{\mathcal{A}}$ is invertible. From Cramer's rule, one has

$$R(\xi) = \det(\widehat{\mathcal{A}}) \left(Q_1^\intercal \mathrm{adj}(\widehat{\mathcal{A}}) Q_1\right)^{-1}, \tag{3.53}$$

where $\det(\widehat{\mathcal{A}})$ is the determinant of $\widehat{\mathcal{A}}$ and $\mathrm{adj}(\widehat{\mathcal{A}})$ is the adjugate matrix of $\widehat{\mathcal{A}}$. This expression also implies that $\xi = 2\pi$ is not a singularity of $R(\xi)$. From Lemma 3.3.2, we know that $\det(\widehat{\mathcal{A}})$ is smooth w.r.t ξ, and next we will prove $\left(Q_1^\intercal \mathrm{adj}(\widehat{\mathcal{A}}) Q_1\right)^{-1}$ is smooth as well.

From direct computations, we observe that

$$\left(Q_1^\intercal \mathrm{adj}(\widehat{\mathcal{A}}) Q_1\right)^{-1} = \left(C_0 - C_1\cos\xi - C_2\cos 2\xi\right)^{-1}, \tag{3.54}$$

where C_0, C_1 and C_2 are constants related to κ_1 and κ_2, and $C_0 > |C_1| + |C_2|$. This implies that $Q_1^\intercal \mathrm{adj}(\widehat{\mathcal{A}}) Q_1$ is positive and both $\left(Q_1^\intercal \mathrm{adj}(\widehat{\mathcal{A}}) Q_1\right)^{-1}$ and $R(\xi)$ belong to $\mathcal{C}^\infty(0, 2\pi]$. □

Theorem 3.4.2. *When the CG variables are defined by piecewise constant weighting functions, then*

$$|\mathcal{K}_{0,J}(0)| = \mathcal{O}\left(|J|^{-n}\right), \tag{3.55}$$

as $|J| \to +\infty$ *for any positive integer* n*.*

Proof. Similar to (3.36), we can simplify the expression (3.52) as follows,

$$\mathcal{K}_{0,J}(0) = \frac{1}{2\pi}\int_0^{2\pi} e^{iJ\xi}\left[Q_1^\intercal \widehat{\mathcal{A}} Q_1 - (Q_1^\intercal \widehat{\mathcal{A}}^{-1} Q_1)^{-1}\right]\ \mathrm{d}\xi. \tag{3.56}$$

Direct calculation yields,

$$\frac{1}{2\pi}\int_0^{2\pi} e^{iJ\xi} Q_1^\intercal \widehat{\mathcal{A}} Q_1\ \mathrm{d}\xi = \frac{\kappa_1 + 2\kappa_2}{\pi}\int_0^{2\pi} \cos(J\xi)\,(1-\cos(\xi))\ \mathrm{d}\xi. \tag{3.57}$$

Since $J \neq 0$, the integral is zero.

Notice that $\left(Q_1^\intercal \widehat{\mathcal{A}}^{-1} Q_1\right)^{-1}$ is even and $R(\xi)$ is periodic. We may simplify the above expression by integration by parts,

$$\begin{aligned}
\mathcal{K}_{0,J}(0) &= -\frac{1}{2\pi}\int_0^{2\pi} R(\xi)\cos(J\xi)\ \mathrm{d}\xi \\
&= \frac{1}{J}\left[\frac{1}{2\pi}\int_0^{2\pi} R'(\xi)\sin(J\xi)\ \mathrm{d}\xi\right] \\
&= \frac{1}{J^2}\left[\frac{1}{2\pi}\int_0^{2\pi} R''(\xi)\cos(J\xi)\ \mathrm{d}\xi\right] \\
&= \frac{1}{J^3}\left[-\frac{1}{2\pi}\int_0^{2\pi} R'''(\xi)\sin(J\xi)\ \mathrm{d}\xi\right] \\
&= \quad \cdots .
\end{aligned} \tag{3.58}$$

One can repeat the above steps and have the estimate $|\mathcal{K}_{0,J}(0)| \leq \mathcal{O}(|J|^{-n})$ for any integer order n. □

To verify the fast decay, we use the same model as the previous numerical tests, and plot the dependence of $\mathcal{K}_{0,J}(0)$ on J on a logarithmic scale in Figure 3.6. We fix the total number of atoms $N = 2^{15}$ and averaging block size $M = 2^5$. The x-axis indicates the matrix index J and the y-axis is the logarithms of $|\mathcal{K}_{0,J}(0)|$. We observe that $|\mathcal{K}_{0,J}(0)|$ shows an exponential decay as $|J| \to \infty$.

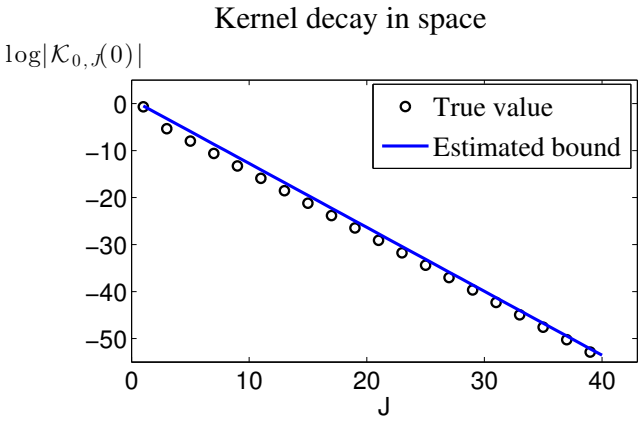

Figure 3.6: The spatial decay of $|\mathcal{K}_{0,J}|$ on the logarithmic scale for the piecewise constant averaging.

3.5 The temporal decay of $\mathcal{K}_{0,0}(t)$

Finally, we also analyze the dependence of the kernel functions on time t when other parameters are fixed, such as the weighting functions and the block size. $\mathcal{K}_{0,0}(t)$ decays over a long time period and we estimate its decay rate.

Finally, we analyze how the kernel function changes in time. Here we will focus on the time decay estimate when the averaging operators are piecewise constant. Our aim is to estimate the asymptotic behavior of

$$\mathcal{K}_{(0,0)}(t) = \frac{1}{2\pi}\int_0^{2\pi} \widehat{\Phi}^*\widehat{\mathcal{A}}\widehat{\Psi}\cos(\widehat{\Omega}t)\widehat{\Omega}^{-2}\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Phi}\ d\xi, \tag{3.59}$$

as $t \to +\infty$.

Theorem 3.5.1. *When the CG variables are defined by piecewise constant weighting functions, the kernel function decays in time with rate at least equal to* 0.5. *Namely,*

$$|\mathcal{K}_{0,0}(t)| \leq \mathcal{O}\left(t^{-1/2}\right) \text{ as } t \to +\infty. \tag{3.60}$$

Let $\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi} = X^*O^2X$ be the eigen decomposition of $\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi}$, where M is a real diagonal matrix and X is a unitary matrix. Then we are able to simplify (3.59) as

$$\mathcal{K}_{0,0}(t) = \frac{1}{2\pi}\int_0^{2\pi}\sum_{j=1}^{M}\frac{\cos(\mu_j t)|g_j|^2}{\mu_j^2}\,d\xi, \tag{3.61}$$

where $\mu_j = O_{j,j}$ is an eigenvalue of $\widehat{\Omega}$ and g_j is the jth component of $X\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Phi}$, and $\mathcal{K}_{0,0}(t)$ can be written as

$$\mathcal{K}_{0,0}(t) = \sum_{j=1}^{M}\frac{1}{2\pi}\int_0^{2\pi}\frac{|g_j|^2}{\mu_j^2}e^{i\mu_j t}\,d\xi. \tag{3.62}$$

This is an integral of Fourier type and we regard t as a large parameter. In this regime, the integrand oscillates severely and causes cancellation. The stationary phase approximation method can be used here to estimate the decay rate of (3.61).

The stationary phase method is an approach for estimating integrals with fast oscillation by evaluating the integrands in regions where they contribute the most. Let us recall the stationary phase approximation method [64] with the following lemma.

Lemma 3.5.2. *Consider the integral,*

$$I(t) = \int_a^b f(x)e^{i\omega(x)t}\,dx, \tag{3.63}$$

where f and ω are real-valued functions and ω is called phase function. Assume $\omega \in \mathcal{C}^2(a,b)$ and has one stationary point at $x_0 \in (a,b)$, where $\omega'(x_0) = 0$ and $\omega''(x_0) \neq 0$, then the integral can be approximated asymptotically by

$$I(t) \approx e^{i\omega(x_0)t}e^{sign(\omega''(x_0))\frac{i\pi}{4}}f(x_0)\left[\frac{2\pi}{t|\omega''(x_0)|}\right]^{\frac{1}{2}}, \tag{3.64}$$

as $t \to +\infty$.

If ω has more than one stationary points on (a,b), one can divide (a,b) into several intervals and add the estimate of each interval up, and as t goes infinite,

the decay rate of the integral will not change. If ω has one stationary point with $\omega'(x_0) = \omega''(x_0) = 0$, then the integral decays at a faster rate than $1/2$.

Up to now, we know that $\{\mu_j(\xi)\}_{j=1}^{M-1}$ are the eigenvalues of $\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi}$ given ξ pointwisely. Applying the stationary phase method would require the presence of stationary points. For this purpose, we need to label the eigenvalues with proper indices, so that the function $\mu_j(\xi)$ is continuous w.r.t ξ (or even smoother depending on the choice $\widehat{\mathcal{A}}$). μ_j will be called branches of eigenvalues. Figure 3.7 shows these eigenvalues and their interlacing property.

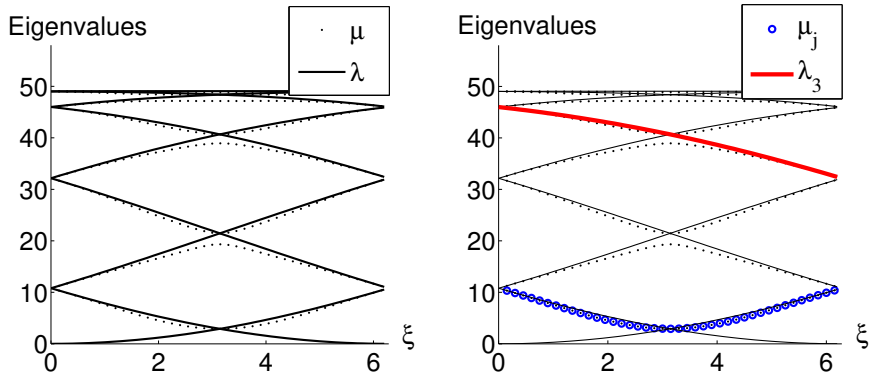

Figure 3.7: Eigenvalues of $\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi}$ and $\widehat{\mathcal{A}}$ on $(0, 2\pi]$, labelled with μ and λ respectively. Parameters: $M = 9$, and $\kappa_1 = 12.2676$ and $\kappa_2 = 3.0628$. The right figure highlights two branches of the eigenvalues of $\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi}$ and $\widehat{\mathcal{A}}$.

Now let us take a further look at the eigenvalues of $\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi}$. Let μ be an eigenvalue of $\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi}$ and x be its corresponding eigenvector, i.e.,

$$\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi}x = \mu x, \quad |x| = 1. \tag{3.65}$$

In the proof of Theorem 3.3.4, we know $\widehat{\Psi}^*\widehat{\mathcal{A}}\widehat{\Psi}$ is non-singular, so $\mu \neq 0$. In addition, for any two matrices $C \in \mathbb{C}^{k\times l}$ and $D \in \mathbb{C}^{l\times k}$, CD and DC have the same eigenvalues (counting multiplicities) except for the zero eigenvalue [65]. μ is a non-zero eigenvalues of $\widehat{\mathcal{A}}\widehat{\Psi}\widehat{\Psi}^*$, and the original eigenvalue problem (3.65) is equivalent to

$$\widehat{\mathcal{A}}\widehat{\Psi}\widehat{\Psi}^*y = \mu y, \quad \mu \neq 0, \quad |y| = 1. \tag{3.66}$$

Let $\widehat{\mathcal{A}} = V\Lambda V^*$ be the eigen decomposition of $\widehat{\mathcal{A}}$, then the eigenvalue problem is equivalent to solving μ in the equations as follows,

$$\left(\Lambda - \Lambda V^*\widehat{\Phi}\widehat{\Phi}^*V\right)\widetilde{y} = \mu\widetilde{y}, \quad \mu \neq 0, \tag{3.67}$$

where $\widetilde{y} = V^*y$.

The matrix determinant Lemma [66] states that if B is a complex square matrix and u, v are column vectors, then the following identity holds,

$$\det(B + uv^*) = \det(B) + v^*\mathrm{adj}(B)u. \tag{3.68}$$

To deal with the eigenvalue problem in (3.67), let the matrix $B = \Lambda - \Lambda V^*\widehat{\Phi}\widehat{\Phi}^*V$ and apply the matrix determinant Lemma to it. We would have the characteristic polynomial

$$f(\mu) = \det(\mu I - B) = \det(\mu I - \Lambda) - z^*\mathrm{adj}(\mu I - \Lambda)\Lambda z, \tag{3.69}$$

where $z = V^*\widehat{\Phi} = \frac{1-e^{-i\xi}}{M}\left(\frac{1}{1-e^{i\xi_1}}, \frac{1}{1-e^{i\xi_2}}, \cdots, \frac{1}{1-e^{i\xi_M}}\right)^*$. Setting the above characteristic polynomial equal to zero, we have

$$\prod_{j=1}^{M}(\mu - \lambda_j) - \sum_{j=1}^{M}\lambda_j|z_j|^2\left(\prod_{\substack{k=1\\k\neq j}}^{M}(\mu - \lambda_k)\right) = 0, \tag{3.70}$$

where $\lambda_j = 2\kappa_1(1 - \cos\xi_j') + 2\kappa_2(1 - \cos 2\xi_j')$ are the eigenvalues of $\widehat{\mathcal{A}}$.

From the fundamental Theorem of Algebra one can define $\mu(\xi)$ as a branch of non-trivial solutions to (3.70). Together with the explicit form of λ_j and z_j, we know $\mu(\xi)$ is smooth.

We may enforce an increasing order to label eigenvalues, for instance, $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_{M-1}$. However, when two branches intersect, the labelling may break down the smoothness of branches. Figure 3.8 further demonstrates one possible scenario.

Fortunately, we have the following lemma that shows eigenvalues of $\widehat{\Psi}^*\widehat{A}\widehat{\Psi}$ don't intersect.

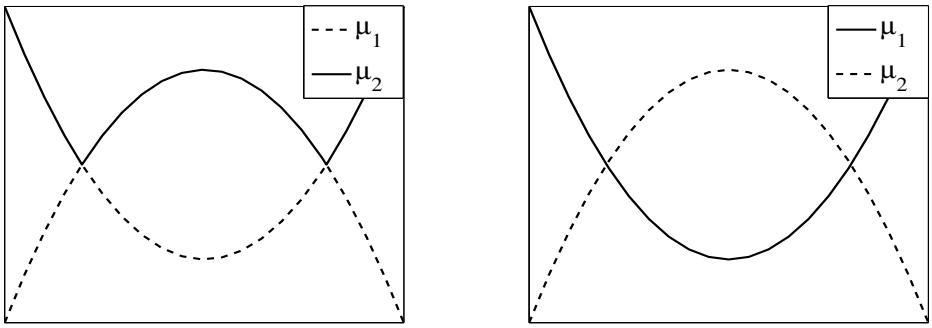

Figure 3.8: Both figures have the same two branches of eigenvalues but labelled in different ways. Eigenvalues in the left figure are labelled using the way described in the context satisfying $\mu_1 \leq \mu_2$ but it doesn't satisfy global smoothness. The right figure shows a more natural way to number eigenvalues and it preserves the global smoothness.

Lemma 3.5.3. *For $M > 4$ and $\xi \in (0, 2\pi)$, the matrix $\widehat{\Psi}^* \widehat{\mathcal{A}} \widehat{\Psi}$ does not have repeated eigenvalues.*

Proof. Let us prove it by contradiction. Suppose $\widehat{\Psi}^* \widehat{\mathcal{A}} \widehat{\Psi}$ has a repeated eigenvalue c at $\xi = \eta$. This means that for some j, $\mu_{j-1} = \mu_j = c$. From Cauchy interlacing theorem [67] for eigenvalues, one has

$$\lambda_{i_1} \leq \mu_1 \leq \lambda_{i_2} \leq \mu_2 \leq \cdots \leq \lambda_{i_{M-1}} \leq \mu_{M-1} \leq \lambda_{i_M}, \tag{3.71}$$

where λ_j are defined in Lemma 3.3.2 and $\{i_j\}_{j=1}^M$ is a permutation of $\{j\}_{j=1}^M$ subject to $\lambda_{i_{j-1}} \leq \lambda_{i_j}$. Then it is clear that $\lambda_{i_j} = c$ as well.

With a substitution of $\mu = c$ and $\xi = \eta$ into (3.70), one has

$$w_{i_j}(\eta) \prod_{\substack{k=1 \\ k \neq i_j}}^{M} (c - \lambda_k(\eta)) = 0, \tag{3.72}$$

where $w_j(\xi) = \lambda_j |z_j|^2 = \frac{2(1-\cos\xi)}{M^2}[\kappa_1 + 2\kappa_2(1 + \cos\xi'_j)]$.

$w_j(\eta)$ is a positive number, so there must be at least one term in the product that is equal to zero, and let us say $\lambda_l(\eta) = c$ $(l \neq i_j)$. One can factor $\mu - c$ out from

(3.70) and substitute $\mu = c$ and $\xi = \eta$ into it. One will have

$$(w_{i_j}(\eta) + w_l(\eta)) \prod_{\substack{k=1 \\ k \neq i_j, l}}^{M} (c - \lambda_k) = 0. \tag{3.73}$$

Due to the fact that $w_{i_j}(\eta)$ and $w_l(\eta)$ are positive, there exists another index h different from i_j and l s.t. $\lambda_h(\eta) = c$. Repeating above steps, one will have all eigenvalues are equal to c at η, which causes contradiction as explained next.

Let us write $\lambda_k = -4\kappa_2 \cos^2 \xi'_k - 2\kappa_1 \cos \xi'_k + 2\kappa_1 + 4\kappa_2$, which is the evaluation of a quadratic function

$$f(x) = -4\kappa_2 x^2 - 2\kappa_1 x + 2\kappa_1 + 4\kappa_2, \tag{3.74}$$

at $x = \cos \xi'_k = \frac{-\xi + 2k\pi}{M}$.

When the block size M is greater than 4, there must be at least 3 of $\{x_1, x_2, \cdots, x_M\}$ equal to the same number, d. However, the equation

$$\cos \xi'_j = \cos \frac{-\xi + 2j\pi}{M} = d, \tag{3.75}$$

has 2 different solutions at most when $\xi \in (0, 2\pi)$ and $j \in \{1, 2, \cdots, M\}$. This implies the contradiction.

□

With the help of Lemma 3.5.3, it is not difficult to prove Theorem 3.5.1.

Proof. From Lemma 3.5.3 and (3.70), one knows $\mu_j(\xi) \in \mathcal{C}^2(0, 2\pi)$ and does not intersect with another branch. We have $\mu_j(\xi) = \mu_j(2\pi - \xi)$ from the symmetry of (3.70), so $\mu'_j(\pi) = 0$ is a stationary point on the interval $(0, 2\pi)$. According to the stationary phase method, $|\mathcal{K}_{0,0}(0)|$ is bounded at least by $\mathcal{O}\left(t^{-1/2}\right)$ as $t \to +\infty$. □

Based on numerical observations, $\mu''_j(\pi)$ are always non-zero when $M = 9$ and $\kappa_1 = 12.2676, \kappa_2 = 3.0628$, so this estimate can not be further improved.

Numerical simulations about the temporal dependence of the kernel function are shown in Figure 3.9 with different force constants. $\mathcal{K}_{0,0}(t)$ is compared with functions

in the order of $t^{-1/2}$.

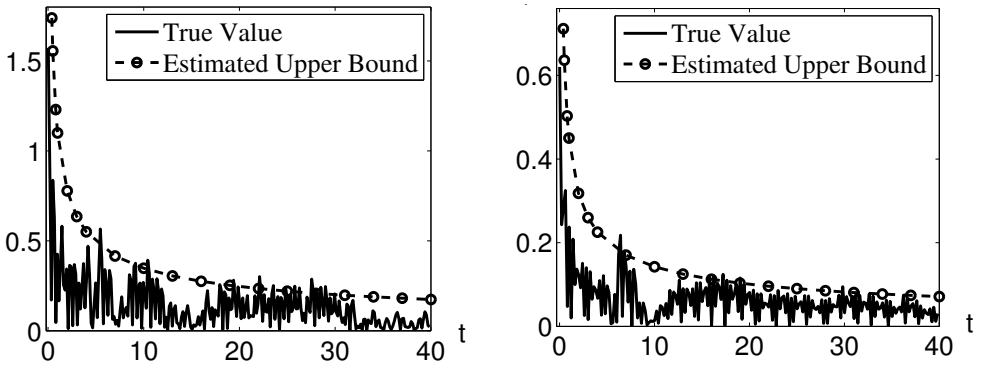

Figure 3.9: The decay behavior of $\mathcal{K}_{0,0}(t)$ as time approaches infinity with different force constants. 2000 atoms are taken into consideration and are divided into 20 blocks evenly. Left: $\kappa_1 = 12.2676$ and $\kappa_2 = 3.0628$. Right: $\kappa_1 = 12.2676$ and $\kappa_2 = -3$.

3.6 Discussions for the decay properties

Our analysis shows that the correlation between the CG variables decays rapidly with respect to the distance. This supports the typical practice where a cut-off radius is introduced in the modeling of the interactions of the CG variables. In sharp contrast to the spatial decay, the time decay is rather slow. In fact, the generalized Langevin equation derived by Adelman and Doll [39], as well as the model used to fit the experimental data [46], also have kernel functions with the similar decay rate.

The slow decay in the time implies that the random noise must be highly correlated. It also poses a rather challenging problem for the approximations of the kernel functions. For example, the early t-model approximation [5] will only be valid for a short period [68]. The Markovian approximation [69, 70] would capture the long-time statistics with a Green-Kubo type of approximations, but the accuracy at the transient time scale is not guaranteed. Beylkin and Monzón [71] proved a striking result that shows that a power law function can be approximated uniformly by a sum of exponentials for all $t > 0$. This type of approximations have been used in [72–75], but no specific error estimate has been provided so far.

The current problem represents one of the scenarios where the Mori-Zwanzig formalism leads to a memory term in the form of linear convolution. Another well-known case is when the full system is *nonlinear*, but the projection operator is chosen as an orthogonal projection, to a subspace spanned by the initial state of the coarse grain variables. This is known as the Mori's projection. In this case, the memory kernel is still proportional to the time correlation of the noise, $\mathcal{K}(t) \propto \langle \mathcal{R}(t), \mathcal{R}(0)^{\intercal} \rangle$, and the random noise can be expressed in terms of an exponential operator $e^{t\mathcal{Q}\mathcal{L}}$, where $\mathcal{Q}$ is the complementary projection operator. Using the fact that the projection is an orthogonal projection, it can be shown that the memory function can be written as solutions of an infinite set of linear differential equations [76]. As a result, a closed-form formula, just like the one in (3.15), also exists. This is true even when the full system is nonlinear. The analysis for the Mori's case will be published somewhere else.

The model considered here is a linearized one around a perfect lattice structure. As a result, the effect of nonlinear hydrodynamics is completely neglected. There has been a great deal of interest in the asymptotic behavior of time correlation functions of either conservative field variables or the corresponding current functions. Often of interest is the universal scaling properties. One-dimensional chain models with nonlinear interactions have been extensively studied. With modern day computing power, it is possible, but still a non-trivial, to probe the tails of correlation functions. See [77, 78] for discussions and some careful literature reviews. The coarse-grained equations considered in this dissertation only include the moment conservation, and the temperature is assumed to be constant. The kernel function would be different if the energy balance is included, and the thermal-mechanical interactions would also be interesting.

Chapter 4
Hierarchical approximations of the memory kernel

A well-known practical issue associated with the solution of the GLE is that computation of the memory term. Clearly, a direct evaluation of the integral requires the storage of the solutions from all previous steps, and such evaluations have to be carried out at every time step. A more delicate issue is that the kernel function $\theta(t)$,

$$\theta(t) = -\langle \mathcal{L} e^{t\mathcal{Q}\mathcal{L}} \mathcal{Q}\mathcal{L} a, b^{\intercal} \rangle \langle b, b^{\intercal} \rangle^{-1}, \tag{4.1}$$

is determined by the orthogonal dynamics $e^{t\mathcal{Q}\mathcal{L}}$ for $t > 0$. Except for very special cases, *e.g.*, linear dynamics as discussed in [79, 80], there is no explicit formula to compute $\theta(t)$ directly. Simulating the orthogonal dynamics is clearly not practical, either.

Since the memory kernel function in the time-domain is not accessible, a direct interpolation by a sum of exponential functions, *e.g.*, [72, 73], is not feasible. To alleviate such effort, and more importantly, to make connections to existing generalized heat conduction models, we will use the Markovian embedded technique [74, 81] and approximate the memory term via an extended system of equations [74]. Our formulation rests upon a rational interpolation, in which the required function values are statistics of the coarse-grain variables.

As Mori's projection in (2.4) is a special case of the oblique projection in (2.20), we use the notation in Section 2.3 for future references.

4.1 Extended dynamical equations

The idea of the hierarchical approximation is to incorporate the aforementioned values of the kernel function into the Laplace transform of θ. More specifically, we define

$$\Theta(\lambda) = \int_0^{+\infty} \theta(t) e^{-t/\lambda} dt. \tag{4.2}$$

As $\lambda \to 0+$, using integration by parts repeatedly, we find that

$$\Theta(\lambda) = \lambda\theta(0) + \lambda^2\theta'(0) + \lambda^3\theta''(0) + \lambda^4\theta'''(0) + \cdots. \tag{4.3}$$

Let us consider the rational approximation which is a generalization of approximating the memory kernel with sum of exponential functions [73, 82, 83]. In general, a rational function in k-th order is expressed in the following form,

$$R_{k,k} = \left[I - \lambda B_1 - \cdots - \lambda^k B_k\right]^{-1}\left[A_0 + \lambda A_1 + \lambda^2 A_2 + \cdots + \lambda^k A_k\right], \tag{4.4}$$

with coefficient matrices A_i's and B_i's to be determined. As we observe (4.3), $\Theta(\lambda)$ approaches to zero as $\lambda \to 0+$. A_0 has to be zero in the above form to make it consistent. The determination of other parameters will be elaborated in the next section.

At this point, if we define the memory term as $z(t)$ in (3.14), *e.g.*,

$$z(t) = \int_0^t \theta(t-s) b(s) ds, \tag{4.5}$$

and take the Laplace transform of it, we have the convolution decoupled as

$$\hat{z}(\lambda) = \Theta(\lambda)\hat{b}(\lambda), \tag{4.6}$$

where $\hat{z}$ and $\hat{b}$ are Laplace transforms of z and b,

$$\hat{z}(\lambda) = \int_0^{+\infty} z(t) e^{-t/\lambda} dt, \quad \hat{b}(\lambda) = \int_0^{+\infty} b(t) e^{-t/\lambda} dt. \tag{4.7}$$

By plugging (4.4) into (4.6), we have the following equation,

$$\frac{1}{\lambda}\hat{z} = A_1\hat{b} + B_1\hat{z} + \hat{z}_2, \tag{4.8}$$

where

$$\hat{z}_2 = (\lambda A_2 + \cdots + \lambda^{k-1}A_k)\hat{b} + (\lambda B_2 + \cdots + \lambda^{k-1}B_k)\hat{z}. \tag{4.9}$$

Since the inverse Laplace transform of (4.8) can be computed explicitly, we can have a system of extended ODEs of auxiliary variables by doing this procedure repeatedly. If we ignore the random effect and write $z_1 = z$, we have the following extended dynamics where the memory term is embedded in an extended dynamical system *without* memory,

$$\begin{cases} \dot{a} = Sb - z_1 \\ \dot{z}_1 = A_1 b + B_1 z_1 + z_2 \\ \dot{z}_2 = A_2 b + B_2 z_2 + z_3 \\ \qquad \cdots \\ \dot{z}_k = A_k b + B_k z_k. \end{cases} \tag{4.10}$$

As the order of the approximation k increases, we obtain a hierarchy of approximations for the memory term, which can be written as a larger extended system of equations [74, 79]. We could write the ODE system (4.10) in a more compact form,

$$\begin{cases} \dot{a} = Sb + e^{\intercal} z, \\ \dot{z} = Ab + Bz, \end{cases} \tag{4.11}$$

where $z = (z_1, z_2, \cdots, z_k)$ is the auxiliary variable and A, B are block matrices,

$$A = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ \cdots \\ A_k \end{pmatrix}, B = \begin{pmatrix} B_1 & I & 0 & \cdots & 0 \\ 0 & B_2 & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & B_{k-1} & I \\ 0 & 0 & \cdots & 0 & B_k \end{pmatrix}. \tag{4.12}$$

The idea of introducing auxiliary variables and extended dynamics has been

implemented in various previous works, *e.g.*, [72, 73, 82, 84–86].

4.2 Data-driven parameter estimation

Up to now, the parameters A_i's and B_i's in the approximation remain undetermined. Here we provide two possible methods to determine the parameters with appropriate interpolation conditions or matching requirement. We connect them with the statistics of coarse-grained variables in order to make them computable.

4.2.1 Two-point interpolation

As $R_{k,k}(\lambda)$ is an approximation of $\Theta(\lambda)$, it is natural to make them consistent by doing appropriate interpolations. In total, $2k+1$ parameters need to be determined which requires $2k+1$ interpolation conditions. We now incorporate the values of Θ from two regimes: $\lambda \to +\infty$ and $\lambda \to 0+$. Such two-sided approximations, which are similar to the Hermite interpolation problems, have demonstrated promising accuracy over both short and long time scales [74].

As λ goes to infinity, it embodies long-time behavior of the kernel function, where we match

$$\lim_{\lambda \to +\infty} R_{k,k}(\lambda) = \lim_{\lambda \to +\infty} \Theta(\lambda). \tag{4.13}$$

For this calculation, we write the correlation function as $M(t) = \langle b(t), b^\intercal \rangle$ and $N(t) = \langle a(t), b^\intercal \rangle$ which represent two-point statistics of a and b. We know that b is uncorrelated with the random noise F from the projection definition, so we multiply both sides by $b^\intercal$ in the GLE (3.14) and take the average yielding

$$\dot{N}(t) = SM(t) - \int_0^t \theta(t-s)M(s)ds. \tag{4.14}$$

Let $\widetilde{M}(\lambda)$ and $\widetilde{N}(\lambda)$ be the Laplace transform of $M(t)$ and $N(t)$ respectively, *i.e.*,

$$\begin{aligned} \widetilde{M}(\lambda) &= \int_0^\infty M(t)e^{-t/\lambda}dt, \\ \widetilde{N}(\lambda) &= \int_0^\infty N(t)e^{-t/\lambda}dt. \end{aligned} \tag{4.15}$$

After taking the Laplace transform of (4.14), we find

$$\frac{1}{\lambda}\widetilde{N}(\lambda) - N(0) = S\widetilde{M}(\lambda) - \Theta(\lambda)\widetilde{M}(\lambda), \tag{4.16}$$

which yields,

$$\begin{aligned} &\Theta(+\infty) = S + N(0)C^{-1}, \\ &C \stackrel{\text{def}}{=} \lim_{\lambda\to+\infty}\int_0^{+\infty} e^{-t/\lambda}M(t)dt. \end{aligned} \tag{4.17}$$

In practice, we keep the exponential penalty from the Laplace transform to ensure the integral is finite. It is well-known that [87] a necessary condition for the integral to be convergent is that the process $a(t)$ is ergodic, but it is not sufficient. Finally, for a scalar CG variable and the case of $b = a$, M can be identified as the correlation length. In this way, $\Theta(+\infty)$ is related to the statistics of $a(t)$, which can be obtained either from the full model or from experimental observations.

As λ approaches to $0+$, one imposes the $2k$ matching conditions to incorporate the short-time behaviors,

$$\begin{aligned} R_{k,k}(0) &= \Theta(0), \\ R'_{k,k}(0) &= \Theta'(0), \\ R''_{k,k}(0) &= \Theta''(0), \\ &\cdots\cdots \end{aligned} \tag{4.18}$$

The values of the right-hand side in (4.18) can be related to the derivatives of $\theta(0)$ via (4.3). They are computable using the statistics of a and b from (2.24).

The above interpolation conditions (4.13) and (4.18) give us $2k+1$ linear equations for A_i's and B_i's, which can be determined based on the values of the kernel functions and have explicit formulas with those statistics presented in (2.24).

Now let us turn to examples of this hierarchy approximation approach. When $k = 0$, we are led to a constant function, $R_{0,0} = \Gamma$, with equivalent approximation,

$$\int_0^t \theta(t-s)b(s)ds \approx \Gamma b(t), \tag{4.19}$$

where Γ we choose to be given by (4.17). In the time domain, this amounts to

approximating the kernel function by a delta function:

$$\theta(t) \approx \Gamma\delta(t). \tag{4.20}$$

This is often referred to as the *Markovian* approximation [70, 88].

When $k = 1$, we have $R_{1,1}(\lambda) = \big[I - \lambda B_1\big]^{-1}\big[A_0 + \lambda A_1\big]$. The matching conditions,

$$\begin{aligned} R_{1,1}(0+) &= \Theta(0+) \\ R'_{1,1}(0+) &= \Theta'(0+) \\ R_{1,1}(+\infty) &= \Theta(+\infty), \end{aligned} \tag{4.21}$$

lead to a group of linear equations, which yield

$$A_0 = 0,\ A_1 = \theta(0)\ \text{ and } B_1 = -A_1\Theta(+\infty)^{-1}. \tag{4.22}$$

In the time-domain, this corresponds to an approximation of the kernel function by an matrix exponential $e^{B_1 t}A_1$.

4.2.2 Moment consistency

Utilizing the fact that b is uncorrelated with the noise, we multiply $b^\intercal$ to (3.14) and take the Laplace transform of it, which yields,

$$\frac{1}{\lambda}\widetilde{N}(\lambda) - N(0) = S\widetilde{M}(\lambda) + R_{k,k}(\lambda)\widetilde{M}(\lambda), \tag{4.23}$$

If we assume that the short-time statistics can be extracted, *e.g.*, from equilibrium MD simulations,

$$N_\ell = \frac{d^\ell}{dt^\ell}\left\langle a(t), b^\intercal\right\rangle|_{t=0}, \quad M_\ell = \frac{d^\ell}{dt^\ell}\left\langle b(t), b^\intercal\right\rangle|_{t=0}. \tag{4.24}$$

The superscripts ℓ indicate the time derivative, $\ell \geq 0$.

With these statistics, the Laplace transforms $\widetilde{M}(\lambda)$ and $\widetilde{N}(\lambda)$ can be expanded

around 0+ [89],

$$\begin{aligned}\widetilde{M}(\lambda) &= \lambda M_0 + \lambda^2 M_1 + \lambda^3 M_2 + \cdots, \\ \widetilde{N}(\lambda) &= \lambda N_0 + \lambda^2 N_1 + \lambda^3 N_2 + \cdots.\end{aligned} \tag{4.25}$$

We use the same idea of incorporating the values of Θ from two regimes: $\lambda \to 0+$ and $\lambda \to +\infty$. This time we look at (4.23) as an equivalent equation for λ. The parameters are determined by matching the coefficients of λ's powers in the expansion around $\lambda = 0+$ as follows,

$$\begin{aligned} M_1 &= A_0 N_0 \\ \lambda(M_2 - B_1 M_1) &= \lambda(A_0 N_1 + A_1 N_0) \\ \lambda^2(M_3 - B_1 M_2 - B_2 M_1) &= \lambda^2(A_0 N_2 + A_1 N_1 + A_2 N_0) \\ &\cdots \end{aligned} \tag{4.26}$$

In general, we can match the first $2k$ coefficients and obtain $2k$ linear equations for the coefficients A_i's and B_i's.

The remaining condition is imposed at $\lambda \to +\infty$, which incorporates long-time statistics, yielding

$$M_0 = B_k^{-1} A_k N_\infty, \tag{4.27}$$

where

$$N_\infty = \lim_{\lambda \to +\infty} \widetilde{N}(\lambda) = \lim_{\varepsilon \to 0+} \int_0^{+\infty} e^{-\varepsilon t} N(t) dt. \tag{4.28}$$

As it turns out, without the long-time statistics, the resulting model will be wave-type of equations, with no dissipation.

4.3 Averaged dynamics and deterministic properties

Before we further look into the random noise term, let us first briefly discuss the deterministic GLE model (4.29) and examine the effectiveness of the models obtained from the previous approximations scheme.

By taking the average of (3.14), we can eliminate the fluctuating effect and have

a deterministic GLE

$$\langle \dot{a} \rangle(t) = S\langle b \rangle(t) - \int_0^t \theta(t-s)\langle b \rangle(s), \tag{4.29}$$

which is a set of integral-differential equations describing the time evolution of the average of the $a(t)$.

As a comparison with existing models, we revisit the heat conduction problem of a one-dimensional system. Let us consider a one-dimensional isolated chain consisting of N atoms in the system. Atoms are evenly divided into n blocks, each of which contains M atoms, as shown in Figure 4.1; $N = nM$. The spacing of two atoms is ε_0 at equilibrium. We will focus on the study of energy transport between these blocks.

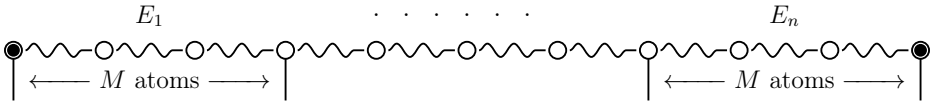

Figure 4.1: One-dimensional chain of atoms. M atoms are grouped into one block.

Let Ω_I be the index set of I-th block, labeled as, $\Omega_I = \{M(I-1)+1, \cdots, MI\}$, and $\ell(x_i - x_j)$ be the pairwise potential coming from interactions between the ith and jth atoms. Here, we use the Fermi-Pasta-Ulam (FPU) potential in the form of,

$$\ell(r) = \frac{r^2}{2} + c_0\frac{r^3}{3} + c_1\frac{r^4}{4}, \tag{4.30}$$

where c_0, c_1 are parameters in the potential. We assume $c_1 > 0$. If we only consider the nearest neighbor interactions, the potential energy in (1.2) is given by,

$$V(u) = \sum_{i=1}^{N} \ell(u_i - u_{i-1}), \tag{4.31}$$

where the periodic boundary condition is imposed, *i.e.*, $u_0 = u_N$ and $u_{N+1} = u_1$.

We define the locally averaged energy associated with the I-th block as follows,

$$E_I = \frac{1}{\varepsilon_0 M}\sum_{i\in\Omega_I} \frac{1}{2}mv_i^2 + \frac{1}{2}\ell(u_i - u_{i-1}) + \frac{1}{2}\ell(u_{i+1} - u_i). \tag{4.32}$$

The rate of change of the local energy can be attributed to the energy flux J,

$$\dot{E}_I = -\frac{1}{\varepsilon_0 M}\Big(J_{I+\frac{1}{2}} - J_{I-\frac{1}{2}}\Big), \tag{4.33}$$

where $J_{I+\frac{1}{2}}$ is the energy flux between the Ith and $(I+1)$th blocks. Direct calculation yields,

$$J_{I+\frac{1}{2}} = -\frac{1}{2}\ell'(u_{MI+1} - u_{MI})(v_{MI+1} + v_{MI}). \tag{4.34}$$

Notice that the energy flux only depends on the atoms next to the interfaces between two adjacent blocks. This expression has also been used in many numerical studies, *e.g.*, in [90–92].

The energy balance is in principle is exact, but the equation is not closed. The Mori-Zwanzig formalism provides a closure model with noise that takes into account the under-resolved dynamics. We define the CG variables as the centered averaged energy of blocks,

$$a_I(t) \stackrel{\text{def}}{=} E_I(t) - \langle E_I(0)\rangle. \tag{4.35}$$

The subtraction of the average is to ensure that $\langle a\rangle = 0$.

For the initial configuration, we assume that the local energy is given, denoted here by $\bar{E}$. One approach to set up the initial density ρ_0 is to maximize the entropy $-\int \rho(x_0)\log\rho(x_0)dx_0$, subject to the constraints, $\langle E_I\rangle = \bar{E}_I$. This approach yields the following density function [93],

$$\rho_0 \propto \exp -\sum_I \beta_I E_I, \tag{4.36}$$

with β_I being the Lagrange multiplier. It plays the role of local inverse temperature, and it can be determined based on the constraints $\langle E_I\rangle = \bar{E}_I$.

With this choice of ρ_0, we can show that in the GLE (3.14) $S \equiv 0$. In the approximation of the memory term, we will approximate the density $\rho_0 \approx \frac{1}{Z}\exp -\beta H$, with β being a reference temperature. The approximation considerably simplifies the

calculations. Direct computation yields, for $j \geq 0$,

$$\begin{aligned}\mathcal{M}_{2j+1} &= \langle \mathcal{L}^{2j+1} a, a^\intercal \rangle M^{-1} \\ &= \langle \mathcal{L}^{2j+1} E(0), E(0)^\intercal - \bar{E} \rangle M^{-1} - \langle \mathcal{L}^{2j+1} \bar{E}, E(0)^\intercal - \bar{E} \rangle M^{-1} \\ &= \langle \mathcal{L}^{2j+1} E(0), E(0)^\intercal \rangle M^{-1}.\end{aligned} \tag{4.37}$$

Notice that E is a function of x and v. From simple algebraic observations, $\mathcal{L}^{2j+1} E(0)$ is odd w.r.t. v_0 and $E(0)$ is even w.r.t. v_0. We integrate the product over the velocity domain weighted by a Gaussian distribution. Due to the symmetry, the integral is zero. This implies that the odd moments vanish in (2.24), *i.e.*,

$$\mathcal{M}_{2j+1} = 0, \quad \forall\, j \geq 0. \tag{4.38}$$

As a result, $S = 0$ and with the above results, we are able to further simplify the formulas (2.24) for the derivatives of θ at $t = 0$,

$$\begin{aligned}&\theta(0) = -\mathcal{M}_2, \\ &\theta'(0) = 0, \\ &\theta''(0) = -\mathcal{M}_4 + \mathcal{M}_2^2, \\ &\theta'''(0) = 0, \\ &\theta''''(0) = -\mathcal{M}_6 + \mathcal{M}_4\mathcal{M}_2 + \mathcal{M}_2\mathcal{M}_4 + \mathcal{M}_2^3, \\ &\qquad \cdots\cdots\end{aligned} \tag{4.39}$$

These quantities will be used later as interpolation conditions to determine the kernel function.

In the following of this Chapter, let us restrain ourselves with Mori's projection, *i.e.*, $b = a$. The zeroth order approximation (4.19), the memory effect is eliminated, with the kernel function approximated by a delta function, the GLE (4.29) is reduced to an ODE system,

$$\frac{d}{dt}\langle a \rangle(t) = -\Gamma \langle a \rangle(t). \tag{4.40}$$

For our energy transport problem (4.35), the matrix is given by,

$$\Gamma = M_0 C^{-1}, \tag{4.41}$$

with $M_0 = \langle a, a^\intercal \rangle$ and C defined in (4.17). In light of the periodic boundary conditions and the fact that the system is partitioned uniformly to define the local energy, it is reasonable to assume that a is a stationary process in space, implying that M_0 and C are both circulant matrices [94]. They can be generated from just one row of the matrices. This also implies that Γ is a circulant matrix as well.

We have observed from various numerical tests that $-\Gamma$ is proportional to a discrete Laplacian operator, *i.e.*,

$$\Gamma \propto -\nabla_h^2, \quad h = \varepsilon_0 M, \tag{4.42}$$

as suggested in the left figure in Figure 4.2.

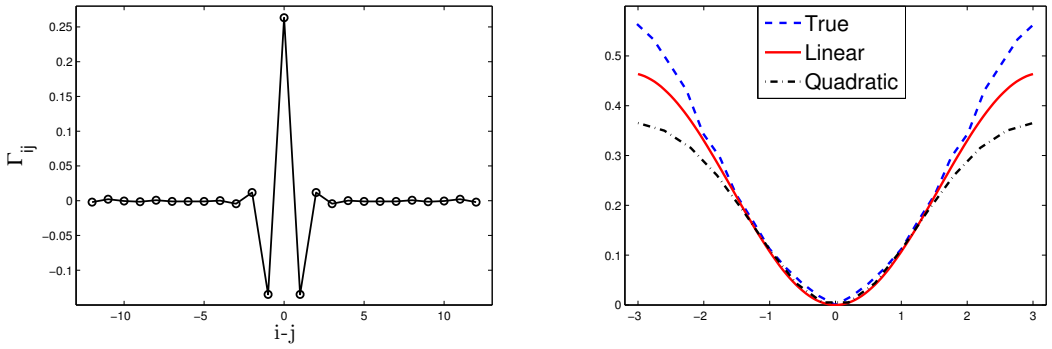

Figure 4.2: The left figure shows the entries of circulant Γ, with x-axis being indices $i - j$ and y-axis being Γ_{ij}. The right figure shows the true value of $\widehat{\gamma}$ and its linear and quadratic fitting of Laplace operator in Fourier space.

With the approximation that $\Gamma = -\alpha \nabla_h^2$, equation (4.40) coincides with a direct discretization of the one-dimensional heat equation,

$$\frac{d}{dt}\langle a \rangle(t) = \alpha \nabla^2 \langle a \rangle(t). \tag{4.43}$$

The parameter α represents the diffusion coefficient of the energy. To extract

the actual value of the parameter, we may expand Γ in terms of the finite-difference operator ∇_h^2:

$$-\Gamma \approx \alpha \nabla_h^2 + \nu \nabla_h^4 + \cdots . \tag{4.44}$$

A more transparent calculation, however, can be done using discrete Fourier transforms, thanks to the circulant structures. More specifically, a discrete Fourier transform can be applied to the generator of a circulant matrix [94]. It will diagonalize the matrix and yield its eigenvalues. We let $\widehat{\gamma}(k)$ and $\widehat{\lambda}(k)$ be the corresponding Fourier transforms of Γ and the discrete Laplacian operator, $k \in (-\frac{\pi}{h}, \frac{\pi}{h}]$. The wave numbers are uniformly distributed in the interval $(-\frac{\pi}{h}, \frac{\pi}{h}]$, and they can be viewed as k-points. In particular, $\widehat{\lambda}(k) = \frac{1}{h^2}\left[2 - 2\cos(k)\right]$. Similarly, we let $\widehat{m}(k)$ and $\widehat{c}(k)$ be the Fourier transform of the two matrices M_0 and C, respectively; then the matrix multiplications are turned into convolution, and we have,

$$\widehat{\gamma}(k) = \widehat{m}(k)/\widehat{c}(k). \tag{4.45}$$

The approximation in (4.44) corresponds to an expansion of $-\widehat{\gamma}(k)$ around $k = 0$,

$$\widehat{\gamma}(k) = \alpha\widehat{\lambda}(k) + \nu\widehat{\lambda}(k)^2 + \mathcal{O}\big(\widehat{\lambda}(k)^4\big), \quad |k| \ll 1, \tag{4.46}$$

as shown in the right picture in Figure 4.2. In particular, we can choose α as

$$\alpha = \lim_{k \to 0} \frac{\widehat{\gamma}(k)}{\widehat{\lambda}(k)}, \tag{4.47}$$

where the limit simply indicates a selection of a small value of k, *e.g.*, the k-point that is closest to the origin. We can not, however, choose $k = 0$, since both terms are zero at that point.

From the numerical observations, M_0 is close to being proportional to an identity matrix, and $\widehat{m}(k)$ is a constant that does not depend on k. It suffices to check $\widehat{c}(k)$ for small k.

Direct computation yields, for $0 < |k| \ll 1$,

$$\begin{aligned}\widehat{c}(k)\widehat{\lambda}(k) &= \frac{1}{n}\int_0^\infty \left\langle \widehat{a}(k,t), \widehat{a}(k,0)^* \right\rangle dt \frac{\left(e^{ik}-1\right)\left(e^{-ik}-1\right)}{h^2} \\ &= \frac{1}{n}\int_0^\infty \left\langle \frac{\left(e^{ik}-1\right)\widehat{a}(k,t)}{h}, \frac{\left(e^{-ik}-1\right)\widehat{a}(k,0)^*}{h} \right\rangle dt, \\ &= \frac{1}{n}\int_0^\infty \left\langle \widehat{\nabla_h a}(k,t), \widehat{\nabla_h a}(k,0)^* \right\rangle dt\end{aligned} \tag{4.48}$$

Interestingly, the right-hand side corresponds to the Fourier transform of the gradient of a. In fact, it is the Fourier transform of the covariance,

$$D \stackrel{\text{def}}{=} \int_0^{+\infty} \langle \nabla_h a(t), \nabla_h a^\intercal \rangle dt. \tag{4.49}$$

By combining (4.45), (4.47) and (4.48), we could determine the parameter α as

$$\alpha = \lim_{k\to 0} \frac{\widehat{m}(k)}{\widehat{d}(k)}, \tag{4.50}$$

where $\widehat{d}$ is the Fourier transform of D (4.49). It is clearly possible to include the $\widehat{\lambda}^4$ term in (4.45) and derive a higher order model.

By linearization of the average energy with respect to the temperature, $\langle a \rangle = c_p T$, with c_p being the specific heat, we can substitute the averaged energy by $c_p T$ in (4.43) and obtain

$$\frac{d}{dt}\langle a \rangle = \alpha c_p \nabla^2 T = -\nabla \cdot (-\alpha c_p \nabla T) \tag{4.51}$$

Combining the above equation with the conservation of energy, $\frac{d}{dt}\langle a \rangle = -\nabla \cdot \langle J \rangle$, we have

$$\langle J \rangle = -\alpha c_p \nabla T, \tag{4.52}$$

which implies that the thermal conductivity $\kappa_{\text{CG}} = \alpha c_p$. The heat capacity c_p can be determined from the statistics of the local energy [93]: $c_p = h\text{var}(E_I)/k_B T^2$, where h appears because E_I is averaged over blocks (energy per unit volume). This leads to a

formula for the thermal conductivity,

$$\kappa_{\mathrm{CG}} = \lim_{k\to 0} h\mathrm{var}(E_I)\widehat{m}(k)/(\widehat{d}(k)k_B T^2). \tag{4.53}$$

The expression on the right hand side correspond to the Fourier transform of the matrix

$$\frac{h}{k_B T^2}\langle a, a^{\intercal}\rangle^2 \Big[\int_0^{+\infty} \langle \nabla_h a(t), \nabla_h a^{\intercal}\rangle dt\Big]^{-1}. \tag{4.54}$$

Meanwhile, there are two conventional approaches to determine the heat conductivity. The first approach is based on the Green-Kubo formula, which is based on a linear response theory. The coefficient can be determined from equilibrium molecular dynamics simulations, which has been widely implemented [48, 95–97]. The formula is written in terms of the autocorrelation function of the heat current,

$$\kappa_{\mathrm{GK}} = \frac{1}{k_B T^2 N \varepsilon_0}\int_0^{\infty} \langle \mathbf{J}(t)\mathbf{J}(0)\rangle \, dt, \tag{4.55}$$

where $\mathbf{J} = \varepsilon_0 \sum_{I=1}^{n} J_{I+\frac{1}{2}}$.

The other traditional approach is based on non-equilibrium MD simulations with two different temperatures maintained at the boundary. The thermal conductivity can be estimated from the heat flux induced by a temperature gradient,

$$\langle j\rangle = -\kappa_{\mathrm{NEMD}}\nabla T. \tag{4.56}$$

We conducted several numerical tests and calculated the heat conductivity based on the three formulas. In our calculation, we consider 250 atoms, divided into 25 blocks equally. The mass of each atom is set to unity $m = 1$, and the spacing of two atoms is set to $\varepsilon_0 = 1$. For equilibrium MD simulations, we prepare the system with two Nosé-Hoover thermostats at temperature 1.5 to reach steady states and use periodic boundary conditions afterward. For non-equilibrium simulations, we impose the two Nosé-Hoover thermostats at temperatures of 1.4 and 1.6, at the left and right boundaries, respectively. The left- and right-most particles in the chain are fixed. For the integration in time, the 6-th order symplectic method is employed [98]. Each MD simulation consists of 10^7 time-steps with stepsize $\Delta t = 0.02$. We sample the

data every 5 steps. Motivated by the studies in [78], we tested both the symmetric ($c_0 = 0, c_1 = 1$) and the asymmetric ($c_0 = 1, c_1 = 1$) FPU potentials. The results are listed in Table 4.1 and show qualitative agreement.

thermal conductivity	CG models	Green-Kubo	NEMD
symmetric potential	31.82 ± 1.41	30.18 ± 1.34	29.16 ± 0.97
asymmetric potential	35.78 ± 1.40	34.39 ± 1.47	35.17 ± 0.89

Table 4.1: This table shows the numerical results of heat conductivity computed with three different methods. In the Green-Kubo formula, to truncate the integral, we have multiplied the correlation function with an exponential decay penalty $e^{-t/\lambda}$ with $\lambda = 10^3$ to eliminate the contribution from the long-time correlation, which may be subject to large numerical and sampling error. For NEMD, we run 5 copies and use the average to compute the thermal conductivity. For each method, we repeat the experiments 10 times to obtain the error bar.

We now examine the model obtained from the first order approximation after taking the average and compare it to existing continuum models. For the example of one-dimensional chain, we have

$$\begin{cases} \dfrac{d}{dt}\langle a\rangle(t) = -\langle z\rangle(t), \\ \dfrac{d}{dt}\langle z\rangle(t) = B_1\langle z\rangle(t) + A_1\langle a\rangle(t), \end{cases} \tag{4.57}$$

where A_1 and B_1 are determined in (4.22). The noise has been averaged out. The second interpolation condition in (4.22) yields $A_1 = -\mathcal{M}_2$, which according to (4.33) and (4.37), is proportional to a discrete Laplacian operator: $A_1 \propto \nabla_h^2$. Meanwhile, the third interpolation condition in (4.22) leads us to $B_1 = -A_1\Gamma^{-1}$. From the previous discussion, $\Gamma \approx -\alpha\nabla_h^2$. As a result, we have

$$A_1 = -c^2\nabla_h^2, \quad B_1 \approx \frac{c^2}{\alpha} I. \tag{4.58}$$

Altogether, this leads to a second order equation

$$\frac{d^2}{dt^2}\langle a\rangle(t) = -\frac{c^2}{\alpha}\frac{d}{dt}\langle a\rangle(t) + c^2\nabla_h^2\langle a\rangle(t). \tag{4.59}$$

Interestingly, this coincides with the Cattaneo and Vernotte (CV) model in heat transport, explicitly demonstrated in [99, 100], where c can be interpreted as the second sound speed and is determined from the relation. Higher order approximations will correspond to higher order relaxation models, similar to the models by Tzou [101].

The two cases presented here illustrate how the memory term can be approximated using the rational approximation in terms of the Laplace transform. In particular, it gives rise to deterministic (or drift) terms in the resulting approximate models. In this dissertation, we also consider the fluctuating effects in energy transport models, by introducing additive and multiplicative noises to the existing models, in the next chapter.

Chapter 5
Heat conduction at nano-scale

5.1 Background and motivation

During the past two decades, there has been a rapidly growing interest in modeling heat transport at the microscopic scale [95, 96, 102–104]. Such renewed interest has been driven by the progress in designing and manufacturing micromechanical and electrical devices, for which thermal conduction properties have significant influences on the performance and reliability. At small scales, electrical and mechanical devices have exhibited heat conduction properties that are quite different from the observations at the macroscopic level [101], which results in a rapidly growing interest in modeling heat transport at the microscopic scale [95, 102, 104].

As the size of electrical and mechanical devices is decreased to the micron and submicron scales, they often exhibit heat conduction properties that are quite different from the observations at the macroscopic level. See Figure 5.1. For example, there has been overwhelming evidence indicating the failure of the conventional Fourier's law (and therefore, the standard heat equation). Furthermore, heat pulses were observed in experiments [101], which are typical behavior of wave equations.

From a modeling viewpoint, a natural approach to incorporate (and predict) some of the observed behavior is to modify the traditional heat equation, *e.g.*, by introducing nonlocal terms or higher order derivatives [101, 105]. These generalized models have been successful in interpreting heat pulse propagation, and they are easy to work with, but they are quite ad hoc. For example, such a generalization

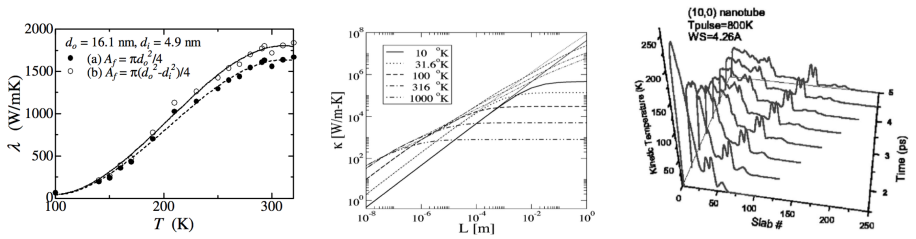

Figure 5.1: Examples of breakdown of the conventional Fourier's law and the standard heat equation of carbon nanotube. Left: dependence of thermal conductivity on temperature [1]. Middle: dependence of thermal conductivity on system length [2]. Right: observation of heat pulses [3].

inevitably introduces additional model parameters that are difficult to determine.

On the other hand, one may rely on the more fundamental description of phonons, which are the energy carriers in solids. The distribution of phonons is governed by the Peierles-Boltzmann (PB) equation [106–108], or simplified PB equations [109–113], where the cubic collision term is replaced by a relaxation term, similar to the Bhatnagar-Gross-Krook model [114] in kinetic theory. In general, however, the computation of the full PB model is still an open challenge due to the high dimensionality.

Thus far, the most popular approach to study heat conduction is direct molecular dynamics (MD) simulations, which are often designed to mimic experimental setup. Given an interatomic potential V, either empirically constructed or derived from more fundamental considerations, an MD model is typically expressed in terms of Newton's equations of motion. There are many well established computational techniques for MD simulations [115, 116].

Nevertheless, despite the many contributions that have recently appeared on the studies of heat conduction processes (*e.g.*, [97, 104, 117–127]), direct MD models have several serious limitations when applied to heat conduction problems. The first obvious limitation is the computational cost. One has to work with a large number of atoms, and the system needs to be integrated for millions or billions of steps.

Typical quantities of interest are expressed as ensemble averages or two-point correlations. For a non-equilibrium process such as the transient heat conduction process, the ensemble averages may not be replaced by time averages, at least very

little theory exists to support such a practice. Therefore, many copies, typically tens of thousands, need to be created to average out statistical fluctuations. Also, the size of the system (and time scale) that can be modeled by direct MD simulations is small, often comparable to or smaller than the mean free path of phonons.

As a consequence, most current MD studies are restricted to quasi one-dimensional systems, *e.g.*, nanowires [91, 104, 128–133], nanotubes [1, 126, 127, 134–139], and nanoribbons [140–142].

Further, while it is often straightforward to incorporate quantities such as displacement, velocity, temperature, and pressure into MD simulations as constraints. However, the temperature gradient is very difficult to impose. The temperature gradient that can be imposed is usually on the order of $10^8 - 10^9 K/m$, which is too large to model realistic systems. It is unclear whether results obtained in such MD simulations can be appropriately extrapolated to the correct regime.

This Chapter is strongly motivated by the above-mentioned issues, and the purpose is to present a coarse-grained (CG) model to alleviate these fundamental modeling difficulties. The CG procedure drastically reduces the number of degrees of freedom and offers a practical alternative.

5.2 Numerical observations of non-Gaussian statistics

We start with observations from some numerical experiments. The first example is a one-dimensional Fermi-Pasta-Ulam (FPU) model, with potential given by,

$$V(u) = \sum_j \ell(u_{j+1} - u_j), \quad \ell(u) = \frac{1}{2}u^2 + \frac{1}{4}u^4, \tag{5.1}$$

which has been studied extensively in statistical physics, *e.g.*, [48]. The second example is a single-wall carbon nanotube, with interactions modeled by the Tersoff potential [143]. In terms of the heat conduction process, both systems can be viewed as one-dimensional or quasi one-dimensional chain structures, even though atoms in the nanotube system move in the three-dimensional space.

In both examples, the chain structures are divided into blocks geographically $\{\Omega_i\}$ in order to define local energies (CG variables denoted by a) associated with them, as shown in Figure 5.2. The i-th component of a represents the energy associated with

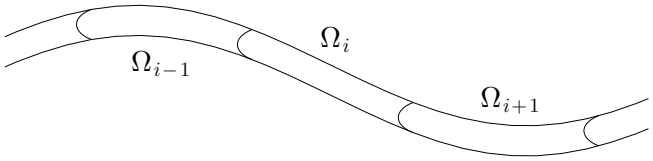

Figure 5.2: Diagram of a quasi one-dimensional chain system with partition. Ω_i indicates the geographic domain of i-th block.

i-th block, consisting of the sum of potential energy and kinetic energy of atoms in i-th block. When the potential energy involves two atoms from two different blocks, the energy will be separated equally between the two blocks. For example, in the nearest-neighbor interaction models, the local energy is defined as

$$a_i = \sum_{j\in\Omega_i} \frac{1}{2}m|v_j|^2 + \frac{1}{2}\ell(u_{j+1} - u_j) + \frac{1}{2}\ell(u_j - u_{j-1}), \tag{5.2}$$

where $u_j, v_j \in \mathbb{R}^d$ are the displacement and velocity of j-th atom and $\ell(\cdot)$ is the pair potential. In the FPU chain example, $d = 1$ and in the nanotube example $d = 3$. The general scheme for energy partition, especially for multi-body interactions, has been studied in [144].

We observe the statistics of the local energy generated from a direct MD simulation under the canonical ensemble (NVT), $f(u, v) = e^{-\beta E(u,v)}/Z$, where $\beta = 1/k_B T$ is the inverse temperature, and E is the energy of the whole system. Two ending blocks of the chain are connected with Nosé-Hoover thermostats [145] at the same temperature, in order to force the whole system to arrive in an NVT ensemble. For the numerical methods, a 6th order symplectic integrator is used to solve the Nosé-Hoover model, where the integrator is constructed based on the 2nd order operator-splitting method [146], followed by an extrapolation scheme [98]. After reaching a steady state, periodic boundary conditions are imposed. Unless otherwise mentioned, atomic units

are used throughout the Chapter.

For the FPU chain, we consider a system of 500 atoms with one block containing 10 atoms. The equilibrium temperature is set to be 1.5 a.u. For the nanotube example, the whole system has 1920 atoms with each block containing 16 atoms. The temperature is set to $300K$, which is converted to atomic units in the simulations. The displayed values of the energy follow the atomic units used in the Tersoff potential (eV).

We run MD simulations long enough to ensure that the system reaches a steady state and observe the probability density function (PDF) of the local energy a_i at equilibrium setting, as shown in Figure 5.3. The energy has been shifted to have zero mean.

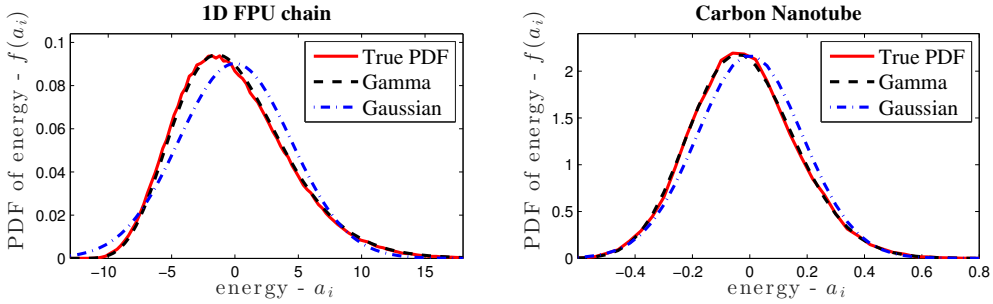

Figure 5.3: The PDF of shifted local energy of 5-th block, compared to Gaussian and Gamma distributions. We collected energy trajectories in a total of n_T steps, from direct MD simulations, after reaching its equilibrium state. The stepsize is $\Delta t = 0.02$ a.u. The true PDFs are normalized histograms of the data, which fits into Gamma $f(x) = \frac{1}{Z}(x+\mu)^{\alpha-1}\exp\left(-\eta(x+\mu)\right), \mu = \alpha/\eta$ and Gaussian $f(x) = \frac{1}{Z}\exp\left(-x^2/2\sigma^2\right)$ distributions with correct mean and variance. The left plot shows the FPU chain example, with a total number of data points $n_T = 10^7$. The right plot shows the nanotube example, with $n_T = 4 \times 10^6$. Similar results are observed for other blocks. For the FPU chain example, fitting parameters $\alpha = 9.8006, \eta = 0.7076$ and $\sigma = 4.4476$. For the carbon nanotube example, fitting parameters $\alpha = 62.6716, \eta = 43.0849$ and $\sigma = 0.1837$.

One interesting finding is that the statistics of the energy in such a regime is non-Gaussian. The PDF actually fits better to a shifted Gamma distribution,

$$\rho(a_i) \propto (a_i + \mu)^{\alpha-1} e^{-\eta(a_i+\mu)}. \tag{5.3}$$

It has also been observed, with two-dimensional histogram plots, that the local energy is almost independent among the blocks. A reasonable ansatz for the joint PDF is given by,

$$\rho(a) \propto \prod_{i=1}^{n_{\text{block}}} (a_i + \mu)^{\alpha-1} e^{-\eta(a_i+\mu)}. \tag{5.4}$$

Here (α, η) are the parameters in the Gamma-distribution which can be determined according to numerical data or empirical theories, and $\mu = \alpha/\eta$ represents the shift to ensure a zero mean of the energy. Independence among the local energy has also been observed for macromolecules [147].

For systems consisting of identical particles, as the block size increases, the local energies tend to exhibit a Gaussian property. Due to the central limit theorem, a non-Gaussian statistics emerges when the block size remains at the nano-mechanical scale. However, for complex systems with multiple types of particles, such as macromolecules, proteins, etc., non-Gaussian statistics have been observed at a larger scale [78].

Furthermore, we observe that the locally averaged energy obeys the Gamma distribution when the potential is either harmonic or asymmetric, as shown in Figure 5.4.

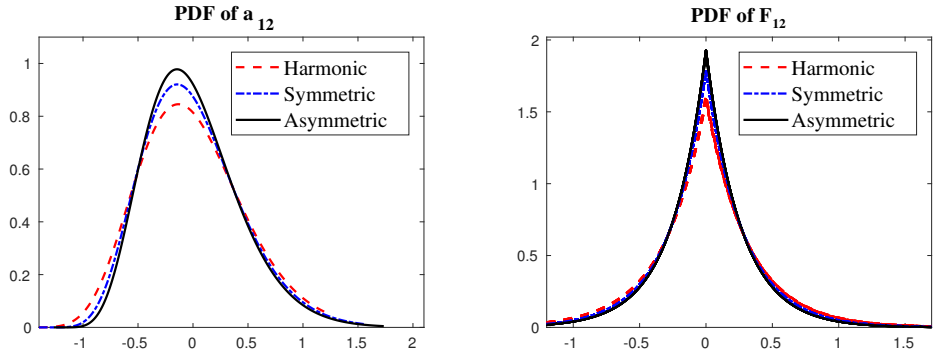

Figure 5.4: These two figures show the histograms of a_{12} and F_{12} at equilibrium from MD simulations, governed by different potential energies. The potentials are given by $\ell(r) = \frac{r^2}{2} + c_0\frac{r^3}{3} + c_1\frac{r^4}{4},$, with harmonic label $c_0 = 0, c_1 = 0$, symmetric label $c_0 = 0, c_1 = 1$ and asymmetric $c_0 = 1, c_1 = 1$. The block size and temperature are the same for three cases in MD simulations, which accounts for the similarity of the parameters.

The emergence of the non-Gaussian statistics can be understood heuristically as

follows. In the harmonic case, where the interactions are modeled by springs, we have the total energy given by,

$$E = \sum_i \frac{\sigma}{2}(u_{i+1} - u_i)^2 + \frac{v_i^2}{2}. \tag{5.5}$$

The parameter σ is the spring constant. Therefore the relative displacement $u_{i+1} - u_i$ can be viewed as independent normal random variables with variance $k_B T/\sigma$.

The local energy within each block can thus be written as,

$$E_I = \frac{1}{4M\varepsilon_0}\left[d_{MI}^2 - d_{M(I+1)}^2\right] + \frac{1}{2\varepsilon_0 M}\sum_{i\in S_I}\left[d_i^2 + v_i^2\right], \quad d_i = \sqrt{\sigma}(u_{i+1} - u_i). \tag{5.6}$$

For large M, the variable E_I will be mostly determined by the summation which is a Gamma distribution with parameters related to the block size.

On the other hand, with the harmonic approximation, the noise term becomes,

$$F_I = \mathcal{L}a = -\frac{1}{h}(J_{I+\frac{1}{2}} - J_{I-\frac{1}{2}}), \tag{5.7}$$

where

$$J_{I+\frac{1}{2}} = -\frac{\sigma}{2}(u_{MI+1} - u_{MI})(v_{MI+1} + v_{MI}) \tag{5.8}$$

from (4.34). Hence, it can be written as $F = \zeta_1\zeta_2 - \zeta_3\zeta_4$ where ζ_1, ζ_2, ζ_3, and ζ_4 are i.i.d. normal random variables, which, according to the well-known result [148], leads to the Laplace distribution.

This explains heuristically that when the interactions are nearly harmonic, such statistics are expected. However, the presence of anharmonic terms will affect the statistics, and it leads to different Gamma or Laplace distributions, as shown in Figure 5.4.

5.3 Models driven by additive Gaussian white noise

A natural (and most widely used) approximation is by a Gaussian white noise. For instance, for the zeroth order (Markovian) approximation (4.17), we are led to a

linear SDE,

$$\dot{a}(t) = -\Gamma a(t) + \sigma\zeta(t), \tag{5.9}$$

where $\zeta(t)$ is the standard Gaussian-white noise,

$$\langle \zeta_i(t_1), \zeta_j(t_2) \rangle = \delta_{ij}\delta(t_1 - t_2). \tag{5.10}$$

In order for the solution a to have the correct covariance $M_0 = \langle a, a^\intercal \rangle$, the parameter σ has to satisfy the Lyapunov equation [87, 149],

$$\Sigma \stackrel{\text{def}}{=} \sigma\sigma^\intercal = \Gamma M_0 + M_0\Gamma^\intercal. \tag{5.11}$$

For the next order rational approximation (4.21) of the kernel function, we may introduce noise via the second equation. Namely,

$$\begin{cases} \dot{a}(t) = -\, z(t), \\ \dot{z}(t) = B_1 z(t) + A_1 a(t) + \sigma\zeta(t), \end{cases} \tag{5.12}$$

where A_1 and B_1 are determined in (4.22). As demonstrated in the previous section, the deterministic part of the model coincides with the CV model. Here we discuss how the noise can be introduced to guarantee the correct statistics.

In (5.12), it is clear that the second equation can be solved explicitly and substituted into the first equation, which would yield a similar equation to the GLE (3.14). By choosing the initial condition $z(0)$ and Σ appropriately, the approximations to the memory kernel and the random noise can be made consistent, in terms of the second FDT (2.7).

Theorem 5.3.1. *Assuming the covariance of $z(0)$ is A_1, and*

$$B_1 A_1 + A_1 B_1^\intercal + \Sigma = 0, \tag{5.13}$$

then, the extended system is equivalent to approximating the kernel function by $\theta_1(t) = e^{tB_1}A_1$, and the approximate noise, denoted by $F_1(t)$, to $F(t)$ satisfies the

second FDT exactly. Namely,

$$\theta_1(t-t') = \langle F_1(t), F_1(t')^\intercal \rangle M_0^{-1}, \quad \forall\, 0 \le t' \le t.$$

The proof of this theorem can be found in [79]. In light of (4.22) and the FDT, the matrix A_1 is semi-positive definite, and it can be used as a covariance matrix.

The model (5.12) is a Langevin dynamics (by changing z to $-z$), which typically comes from the force (momentum) balance. It is therefore interesting that such models also arise in the balance equations for the energy and energy flux.

The approximation by additive noises inevitably leads to a Gaussian distribution for $a(t)$ [149]. To check the validity of this assumption, we solve the reduced models (5.9) and (5.12) directly and compare the results with true statistics obtained from MD simulations, including histograms of solutions at steady states and autocorrelation functions.

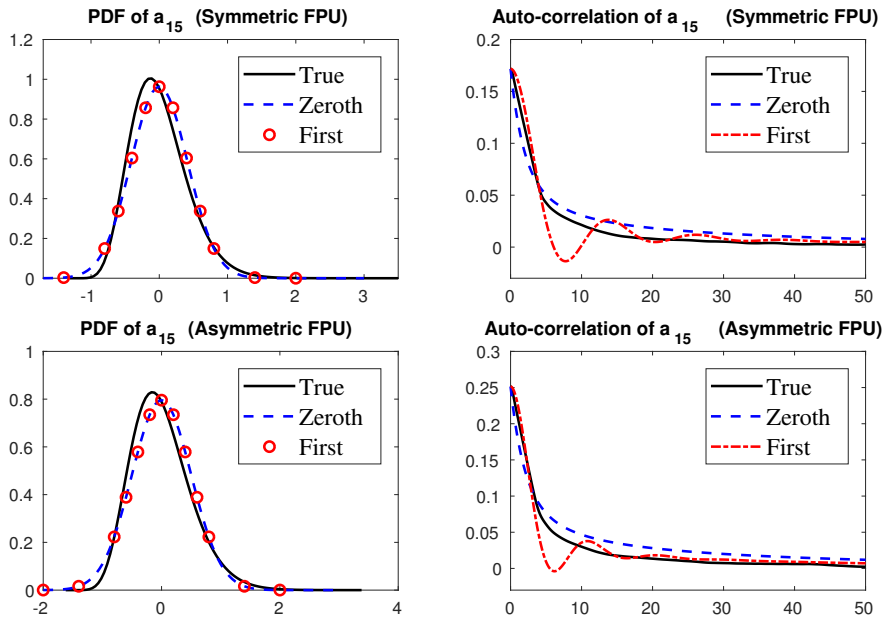

Figure 5.5: The left two figures show PDFs of a_{15} obtained from approximated models (5.9) and (5.12). The right figures show the auto time correlations $\langle a_{15}(t)a_{15}(0)\rangle$. Numerical results are compared with true statistics that come from direct MD simulations using a 6th order symplectic method.

From Figure 5.5, we observe that the autocorrelation of energy can be well captured from models driven by additive Gaussian noise, while the PDFs deviate from the true distributions. One obvious violation is that the solution of (5.9) is always Gaussian, which does not have a lower bound for energy and also unable to recover a non-symmetric PDF.

5.4 Models driven by multiplicative Gaussian white noise

As alluded to in the previous section, the approximate model driven by Gaussian additive white noise may not capture the correct PDF. In this section, we consider the multiplicative noise, intending to enforce the correct equilibrium statistics for the solution of the SDEs.

We start with a further observation that the energy of each block is almost independent of each other. Very interestingly, the same observations have been made for biomolecules [147]. It is difficult to prove the independence theoretically as atoms in block boundaries share pair potentials with their neighbors in adjacent blocks, and precisely the small correlation between blocks generates energy transport. However, when the block size is relatively large, the correlation is quite weak.

Therefore, we keep this as our main assumption and postulate the stationary PDF ($\rho(a)$) of the energy as a shifted multi-Gamma distribution with parameters k and η,

$$\rho(a) = \frac{1}{Z}\prod_{i=1}^{n}\left(a_i + \frac{k_i}{\eta_i}\right)^{k_i-1} e^{-\eta_i(a_i+k_i/\eta_i)},\ a_i \geq -\frac{k_i}{\eta_i}, \tag{5.14}$$

where

$$Z = \prod_{i=1}^{n} Z_i, \quad Z_i = \int_{-\frac{k_i}{\eta_i}}^{\infty}\left(a_i + \frac{k_i}{\eta_i}\right)^{k_i-1} e^{-\eta_i(a_i+k_i/\eta_i)}\, da_i. \tag{5.15}$$

Now we reconsider the zeroth order approximation, the Markovian approximation by $R_{0,0} = \Theta(+\infty) = \Gamma$. With a multiplicative noise, we are solving the following SDE,

$$\dot{a}(t) = -\Gamma a(t) + \sigma(a(t))\zeta(t), \tag{5.16}$$

where $\zeta(t)$ is the again standard Gaussian white noise. The SDE is interpreted in the Itô sense.

To derive a simple formula, we seek σ in a diagonal form. We aim to construct $D{=}2\sigma\sigma^\intercal$ to ensure the desired PDF given by (5.14). By simplifying the Fokker-Planck equation (FPE) that corresponds to (5.16), we obtain,

$$\frac{\partial D_{ii}\rho}{\partial a_i} = -\rho\left(\Gamma_{ii}a_i + \sum_{j=1,j\neq i}^{n}\Gamma_{ij}a_j\right). \tag{5.17}$$

By directly solving these differential equations, we obtain an explicit formula for the matrix D.

Theorem 5.4.1. *If Γ has only non-positive off-diagonal entries and Γ is semi-positive definite, then there exists a diagonal matrix D for which the multivariate Gamma distribution* (5.14) *is a steady state solution of the Fokker-Planck equation. The diagonals of D are given by a positive expression,*

$$D_{ii} = \frac{\Gamma_{ii}}{\eta_i}\left(a_i + \frac{k_i}{\eta_i}\right) - \sum_{j=1,j\neq i}^{m}\Gamma_{ij}\frac{a_j\int_{-\frac{k_i}{\eta_i}}^{a_i}\rho_i(x)dx + k_j/\eta_j}{\rho_i(a_i)}, \tag{5.18}$$

where ρ_i is the marginal PDF of a_i, $\rho_i = \frac{1}{Z_i}\left(a_i + \frac{k_i}{\eta_i}\right)^{k_i-1}e^{-\eta_i(a_i+k_i/\eta_i)}$.

Proof. Direct computation yields that the stationary PDF given in (5.14) satisfies the FPE of the first order model in (5.16), which proves this theorem. □

Let us turn to the model obtained by the first order approximation of the memory term. With multiplicative Gaussian white noise, the first order model can be written formally as follows,

$$\begin{cases} \dot{a}(t) = -\,z(t), \\ \dot{z}(t) = Aa(t) + Bz(t) + \sigma\big(a(t), z(t)\big)\zeta(t), \end{cases} \tag{5.19}$$

where $A = A_1$ and $B = B_1$ are given in (4.22). This is a Langevin equation, where we allow the diffusion coefficient to depend on both a and z.

Similar to the previous case, we are seeking an equilibrium distribution of a and z. Notice that $z(0) = F(0)$, which follows a Laplace distribution. We further assume that a and z are independent. These assumptions lead to the following ansatz for the joint PDF,

$$\rho(a, z) = \frac{1}{Z} \exp\Big[- \sum_{i=1}^{n} \big(W_i(a_i) + \gamma_i |z_i|\big)\Big]. \tag{5.20}$$

In order to find a reasonable expression of σ, we force σ to be diagonal and work with the steady state solution of the FPE, which writes as

$$\sum_{i=1}^{n} \frac{\partial}{\partial z_i} \left(\sum_{j=1}^{n} A_{ij} a_j \rho + \sum_{j=1}^{n} B_{ij} z_j \rho \right) = \sum_{i=1}^{n} \frac{\partial^2 D_{ii} \rho}{\partial z_i^2}, \tag{5.21}$$

where $D = 2\sigma^2$. Furthermore, if it satisfies

$$\frac{\partial D_{ii} \rho}{\partial z_i} = \sum_{j=1}^{n} A_{ij} a_j \rho + \sum_{j=1}^{n} B_{ij} z_j \rho, \tag{5.22}$$

then ρ is a solution to the FPE in (5.21).

Theorem 5.4.2. *Suppose diagonal entries of B are non-positive. ρ given in* (5.20) *is an equilibrium density of the FPE of first order model* (5.19) *if σ is a diagonal matrix and $\sigma_{ii} = (D_{ii}/2)^{1/2}$ for $i = 1, 2, \cdots, n$, where*

$$D_{ii} = -\frac{sgn(z_i)}{\gamma_i} \left(\sum_{j=1}^{n} A_{ij} a_j + \sum_{j=1, j\neq i}^{n} B_{ij} z_j \right) - B_{ii} \left(\frac{|z_i|}{\gamma_i} + \frac{1}{\gamma_i^2} \right) + \mathcal{D}_i e^{\gamma |z_i|}, \tag{5.23}$$

and

$$\mathcal{D}_i = \frac{1}{\gamma_i} \left(\sum_{j=1}^{n} |A_{ij}||a_j| + \sum_{j=1, j\neq i}^{n} |B_{ij}||z_j| \right). \tag{5.24}$$

Proof. Given ρ in (5.20), we could integrate (5.22) and have

$$D_{ii}\rho = -\frac{sgn(z_i)}{\gamma_i} \left(\sum_{j=1}^{n} A_{ij} a_j + \sum_{j=1, j\neq i}^{n} B_{ij} z_j \right) \rho - B_{ii} \left(\frac{|z_i|}{\gamma_i} + \frac{1}{\gamma_i^2} \right) \rho + D_{ii}(0) \rho(z_i = 0). \tag{5.25}$$

Here, $D_{ii}(0)\rho(z_i = 0)$ could be any expression that doesn't contain z_i. We pick $D_{ii}(0)$ deliberately to guarantee that D_{ii} is non-negative. Since $B_{ii} \leq 0$ from the assumption, it is sufficient to control the first term in (5.25). By letting $D_{ii}(0) = \mathcal{D}_i := \frac{1}{\gamma_i}\left(\sum_{j=1}^n |A_{ij}||a_j| + \sum_{j=1,j\neq i}^n |B_{ij}||z_j|\right)$, we have the theorem proved. □

We solved the SDEs (5.16) and (5.19) numerically and computed the statistics of steady solutions as verification of stochastic models. It is worthwhile to point out that the SDEs (5.16) and (5.19) contain an unbounded diffusion coefficient σ, which introduces a stiff problem for the numerical computations. The unboundedness reflects a mechanism for the energy to stay above a lower bound.

To resolve this numerical issue, we applied the implicit Taylor method [150], with time step size $\Delta t = 5 \times 10^{-4}$. Due to the uniform partition of the system, we expect the statistics to be the same for all the components of a. Here, we only exhibit the results from the locally averaged energy of the first bulk.

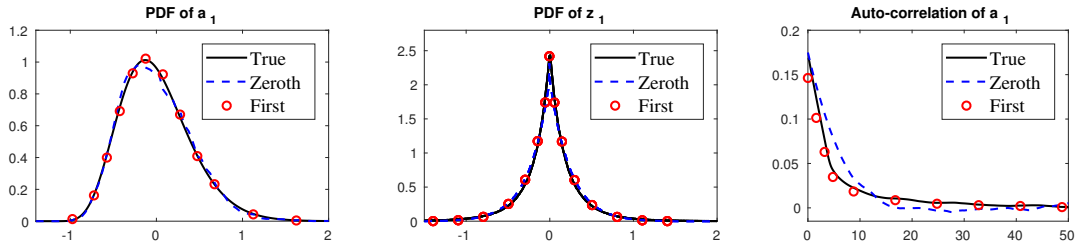

Figure 5.6: The figures show the approximate results of multiplicative model (5.16) and (5.19), in terms of the steady PDF of a_1, steady PDF of z_1 and time auto correlation of a_1.

The numerical results are displayed in Figure 5.6. One can see that the statistics of both a and z are recovered by both models (5.16) and (5.19). In addition, compared with models driven by additive Gaussian white noise, the multiplicative noise model has improved accuracy for the auto-correlation of a as well.

The author is not aware of any ergodicity result for these stochastic dynamics, especially when the coefficients are singular. Therefore, the numerical results here serve as a verification that the system will converge to the desired equilibrium measure.

5.5 Nonlinear heat conduction models

5.5.1 Applications of the oblique projection

In this section, we go back to the non-trivial oblique projection and introduce the driving force as the preassigned direction when we consider the projection.

To ensure the consistency with the true statistics, we first write the PDF formally as

$$\rho(a) = \frac{\exp[-W(a)]}{\int \exp[-W(a)]da}. \tag{5.26}$$

As observed in Figure 5.3, the Gamma distribution is proposed in (5.4). We can define the corresponding $W(a)$ accordingly as,

$$W(a) = \sum_{i=1}^{n_{\text{block}}} w(a_i), \tag{5.27}$$

where

$$w(\zeta) = \eta(\zeta + \mu) - (\alpha - 1)\ln(\zeta + \mu). \tag{5.28}$$

An example of $w(\cdot)$ that corresponds to the shifted Gamma distribution is depicted in Figure 5.7.

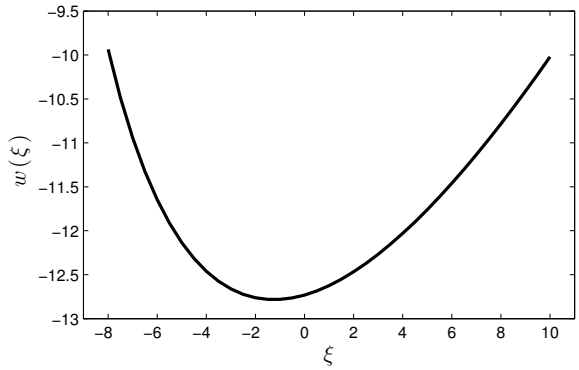

Figure 5.7: The function $w(\zeta)$ defined in (5.28), when $\alpha = 10, \eta = 0.8$ and $\mu = \alpha/\eta$.

We then define $b(t) \overset{\text{def}}{=} \psi(x_0, t) = \bar{\psi}\big(a(t)\big)$ as the *driving force* of energy transport

or heat conduction,

$$\bar{\psi}(a) \stackrel{\text{def}}{=} -\nabla W(a). \tag{5.29}$$

One can interpret $W(a) = -\ln \rho(a)$ as the local entropy and b as the inverse temperature.

Theorem 5.5.1. *The zeroth order model from the oblique projection reads*

$$\dot{a} = \Gamma \nabla W(a) + \sigma\zeta. \tag{5.30}$$

If we enforce $\sigma\sigma^\intercal = -2\text{symm}(\Gamma) = -(\Gamma + \Gamma^\intercal)$*, then the proposed density* $\rho(a)$ *in* (5.26) *is a stationary PDF of* (5.30)*.*

When $k = 1$, the first order include the rate of change of energy as an auxiliary variable $-z$, governed by an auxiliary equation,

$$\begin{cases} \dot{a} = -z, \\ \dot{z} = -A_1 \nabla W(a) + B_1 z + \sigma\zeta. \end{cases} \tag{5.31}$$

Theorem 5.5.2. *If* $\sigma\sigma^\intercal = 2symm(B_1 A_1)$ *and the matrix* A_1 *is symmetric, then the first order model* (5.31) *has a stationary PDF*

$$\rho_{eq}(a, z) = \exp\left[-\left(W(a) + \frac{1}{2} z^\intercal A_1^{-1} z\right)\right], \tag{5.32}$$

whose marginal PDF is exactly identical to $\rho(a)$ *in* (5.26)*.*

For the second order approximation in (4.4), the dynamics for auxiliary variables $z_1 (= z)$ and z_2 is written as

$$\begin{aligned} \dot{z}_1 &= A_1 b + B_1 z_1 + z_2, \\ \dot{z}_2 &= A_2 b + B_2 z_1. \end{aligned} \tag{5.33}$$

If we write it in a more compact form, by defining $z = (z_1, z_2)$,

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \text{ and } B = \begin{bmatrix} B_1 & I \\ B_2 & 0 \end{bmatrix}, \tag{5.34}$$

the GLE (1.1) reads as

$$\begin{cases} \dot{a} = e^{\intercal} z, \\ \dot{z} = -A\nabla W(a) + Bz + \sigma\zeta. \end{cases} \tag{5.35}$$

Theorem 5.5.3. *The equation* (5.35) *has an equilibrium density, whose marginal PDF of* a *is consistent with* (5.26)*,*

$$\rho(a, z) \propto \exp - \left[W(a) + \frac{1}{2} z^{\intercal} Q^{-1} z \right], \tag{5.36}$$

by choosing Q *according to the Lyapunov equation*

$$\sigma\sigma^{\intercal} = -(BQ + QB^{\intercal}). \tag{5.37}$$

The proof of the above theorems is standard by verifying the Fokker-Planck equation [149] associated with them. The analysis of stationary properties can be extended to higher order hierarchy of models.

5.5.2 Numerical schemes and results

We write all the reduced stochastic models in the following unified form,

$$\begin{aligned} \dot{a} =& E_1 z, \\ \dot{z} =& Ab + Bz + \sigma\zeta. \end{aligned} \tag{5.38}$$

Here $E_1 = (-I, 0, \cdots, 0)$ singles out z_1 from the auxiliary variables z. Then, the integrator is constructed by splitting the equations into two systems,

$$\begin{aligned} &(1) \qquad \dot{a} = E_1 z, \qquad \dot{z} = 0, \\ &(2) \qquad \dot{a} = 0, \qquad\quad \dot{z} = Ab + Bz + \sigma\zeta, \end{aligned} \tag{5.39}$$

both of which are linear by themselves, and the solutions can be written out explicitly. A second weak order method can be constructed by solving the first equation for half of the step, the second equation for one step, and then the first equation for half of

the step. Since this is by now a rather standard method in molecular simulations, we will refer the readers to the monographs for the explicit formulas [151].

As a simple validation of the choices of the coefficients of the noise, we compared the PDF obtained from the models (5.30) and (5.31), to that from the full MD simulation. As shown in Figure 5.8, the reduced models reproduce the correct non-Gaussian statistics.

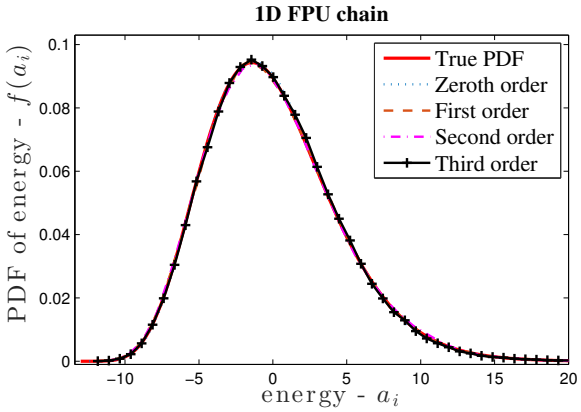

Figure 5.8: PDFs of a_5 in one-dimensioanl FPU chain example. Histograms of full MD simulations (True PDF) and histograms of solutions to reduced models (Zeroth, First, Second, Third orders) are compared. The time step is 0.001 for reduced models and the stochastic integration scheme is Milstein method.

We compute the two-point time correlations of the energy to examine dynamical properties,

$$C_i(t) = \langle a_i(t), a_i(0)^\intercal \rangle. \tag{5.40}$$

As shown in Figure 5.9, the two-point statistics of the energy is consistent for both the one-dimensional chain model (left) and the nanotube system (right). We observe the accuracy improved as we increase the order of the approximations. Interestingly, the one-dimensional model seems to be a bit harder to approximate: We had to extend the method to third order to obtain a good approximation of the time correlation. On the other hand, for the nanotube system, the first order model (5.31) already offers quite satisfactory results. For both systems, the zeroth order approximation (5.30), which is a nonlinear extension of the Fourier's law, exhibits large error.

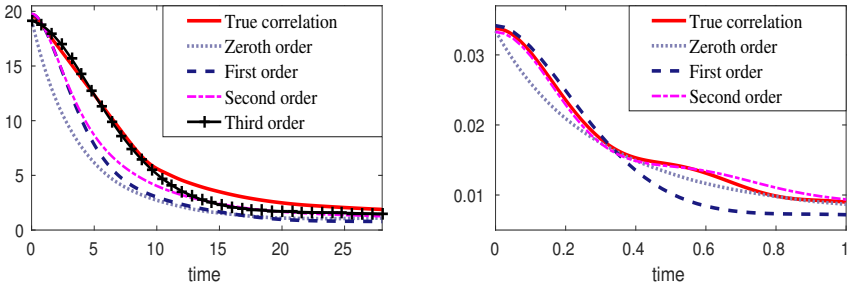

Figure 5.9: Two-point statistics of a_1 from the full MD simulations (True correlation) and reduced models. Left: one-dimensional FPU chain model; Right: single-wall nanotube.

In summary, we constructed models driven by additive Gaussian noise that are in agreement with the statistics of a. In principle, a could be in any type of distribution as long as the statistics are provided from data or empirical guessing.

5.5.3 Stochastic constitutive relations

The Zeroth Order Model. Now let us take a further look at the zeroth order model (5.30) and the first order model(5.31). In (5.30), $\Gamma = M_0 N_\infty^{-1}$ is associated with the statistics of energies in (4.27) and (4.28), which can be computed with the statistics of a.

Our numerical results suggest that Γ is a tri-diagonal matrix and proportional to a three-point discrete Laplacian operator,

$$\Gamma \approx -\kappa \nabla_h^2. \tag{5.41}$$

The subscription h is the width of each block. Here we use standard notations in finite-difference methods,

$$\nabla_h a_{j+1/2} \stackrel{\text{def}}{=} \frac{a_{j+1} - a_j}{h}, \quad \nabla_h \cdot q_j \stackrel{\text{def}}{=} \frac{q_{j+1/2} - q_{j-1/2}}{h}. \tag{5.42}$$

The subscript $j + 1/2$ in the first equation indicates that the finite difference is operated on a_j and a_{j+1}, which are quantities defined at the center of the blocks; The result of the gradient operator is a quantity defined at the interfaces of adjacent

blocks, here labeled by the index $j + 1/2$. Similarly, in the second equation, the divergence is operated on quantities at the interfaces, here denoted by $q_{j+1/2}$, *e.g.*, the heat flux, and the result is a quantity defined at the block centers. The Laplacian operator in (5.41) is defined based on these two operators,

$$\nabla_h^2 a_j \stackrel{\text{def}}{=} \nabla_h \cdot \nabla_h a_j = \frac{a_{j+1} - 2a_j + a_{j-1}}{h^2}. \tag{5.43}$$

Let us write b as a variational form $b = \frac{\delta W(a)}{\delta a}$, the GLE (3.14) is reduced to,

$$\dot{a} = \kappa \nabla_h^2 \frac{\delta W(a)}{\delta a} + \sigma \zeta(t). \tag{5.44}$$

The noise in the GLE (3.14) is approximated by standard Gaussian white noise $\zeta(t)$. Very interestingly, this coincides with the stochastic phase-field crystal model [152] as well as the general diffusion models [153]. The selection of σ is standard: $\sigma\sigma^\intercal = -2\kappa\nabla_h^2$ to make the stationary PDF of this model in the form of (5.26) we have observed from data. It can be viewed as a finite element method (FEM) discretization of the nonlinear heat equation with space-time white noise, which coincides with the stochastic phase-field crystal model [152] as well as the general diffusion models [153].

The local energy that we defined at the beginning satisfies the fundamental conservation law,

$$\dot{a}_j + \nabla_h \cdot q_j = 0, \tag{5.45}$$

with $q_{j+1/2}$ being the heat flux of between the adjacent j and $j + 1$ blocks. Explicit formulas can be derived for many MD models [144, 154].

Equations (5.44) and (5.45) suggest a stochastic constitutive relation for the heat flux,

$$q_{j+1/2} = -\kappa \nabla_h b_{j+1/2} + \sqrt{2\kappa}\zeta_{j+1/2}(t), \tag{5.46}$$

where $b_j = \frac{\delta W(a)}{\delta a_j}$ uses the conventional finite difference notation and $\zeta_{j+1/2}(t)$ is scalar standard Gaussian white noise. This implies a nonlinear, stochastic generalization of the Fourier's Law. It can be directly verified that the deterministic part of the nonlinear constitutive relation also obeys the second law of thermodynamics, since $\nabla_h b_{j+1/2} \cdot \nabla_h a_{j+1/2} \geq 0$.

(5.46) implies a nonlinear, stochastic generalization of the Fourier's Law. It can be directly verified that the deterministic part of the nonlinear constitutive relation also obeys the second law of thermodynamics.

The First Order Model. From the definition in (4.24), one can compute that $N_0 = -I$ and M_2 is proportional to a discrete Laplacian operator, *i.e.*, $M_2 = \gamma_1 \nabla_h^2$. The matching conditions (4.26) lead to an explicit formula for the coefficient matrix A_1,

$$A_1 = M_2 = \gamma_1 \nabla_h^2. \tag{5.47}$$

Now by incorporating the interpolation condition (4.27), we find that,

$$B_1 = -\gamma_1/\kappa I. \tag{5.48}$$

Thus the GLE (3.14) is simplified to,

$$\begin{cases} \dot{a} = -z, \\ \dot{z} = \gamma_1 \nabla_h^2 \dfrac{\delta W(a)}{\delta a} - \dfrac{\gamma_1}{\kappa} z + \sigma\zeta. \end{cases} \tag{5.49}$$

By solving the Lyapunov equation [149], σ is given by the formula,

$$\sigma\sigma^{\mathsf{T}} = -\frac{2\gamma_1^2}{\kappa} \nabla_h^2. \tag{5.50}$$

It is possible to select σ in other forms to ensure the stability, but by choosing σ in this way, the first order model (5.49) has the stationary PDF,

$$\rho(a, z) \propto \exp - \left[W(a) - \frac{1}{2\gamma_1} z^{\mathsf{T}} \nabla_h^{-2} z \right], \tag{5.51}$$

which can be verified by the stationary Fokker-Planck equation. Clearly, the marginal density associated with a is identical to the proposed PDF in (5.4). Again, together with the conservation law (5.45), this stochastic model (5.51) implies a constitutive relation for the heat flux,

$$\tau_1 \dot{q}_{j+1/2} + q_{j+1/2} = -\kappa \nabla_h b_{j+1/2} + \sqrt{2\kappa}\zeta_{j+1/2}(t). \tag{5.52}$$

This is an interesting generalization of the CV model [99]. Not only have we identified the origin of the relaxation parameter, $\tau_1 = \kappa/\gamma_1$, we also incorporated a nonlinear driving force and a random noise. The model can also be written as

$$\tau_1 \ddot{a}_j + \dot{a}_j = \kappa \nabla_h^2 \big[\frac{\delta W(a)}{\delta a}\big]_j + \tau_1 \sigma \zeta_j, \tag{5.53}$$

where σ behaves like a divergence operator $\nabla_h \cdot$ in light of (5.50). This equation corresponds to a discretization of a damped nonlinear wave equation with additive space-time white noise.

The Second Order Model. The second order approximation involves higher order statistical moments M_4 and N_2 in (4.24). Once the statistics are in place, we can observe the behavior of these parameters from numerical experiments. We find that

$$M_4 \approx -\gamma_2 \nabla_h^2, \quad N_2 \approx -\kappa_1 \nabla_h^2, \tag{5.54}$$

and $\gamma_2 \gg \gamma_1 \kappa_1$. After direct substitutions, we obtain an explicit form for the reduced model (5.35),

$$\begin{cases} \dot{a} = z_1, \\ \dot{z}_1 = \gamma_1 \nabla_h^2 \dfrac{\delta W(a)}{\delta a} - \dfrac{\gamma_2 \kappa}{\gamma_1^2} z_1 + z_2, \\ \dot{z}_2 = \dfrac{\gamma_2 \kappa}{\gamma_1} \nabla_h^2 \dfrac{\delta W(a)}{\delta a} - \dfrac{\gamma_2}{\gamma_1} z_1 + \sigma \zeta(t), \end{cases} \tag{5.55}$$

where $\sigma\sigma^{\intercal} = -2\frac{\gamma_2^2 \kappa}{\gamma_1^2} \nabla_h^2$. If we introduce a second relaxation parameter $\tau_2 = \gamma_1/\gamma_2$, the corresponding constitutive relation for the heat flux q reads for

$$\tau_2 \ddot{q}_{j+1/2} + \tau_1 \dot{q}_{j+1/2} + q_{j+1/2} = \kappa \nabla_h b_{j+1/2} + \frac{\tau_2 \kappa}{\tau_1} \nabla_h \dot{b}_{j+1/2} + \sqrt{2\kappa} \zeta_{j+1/2}(t). \tag{5.56}$$

This is again a generalization of the heat conduction model known as the Tzou model [105], with parameters linked to the statistics of the local energies.

We have also derived a third order model, and it can be regarded as a higher order relaxation model for the heat flux, but we omit the details here. Our derivations yield stochastic constitutive relations up to any order. With partial observations of the local energies, the coefficients in these reduced models are identified without running

full MD models.

The stochastic constitutive relations can be nonlinear, and the parameters are linked directly to the statistics of the local energy, making it possible to determine system-specific model parameters using statistical properties. Our parameter identification procedure uses statistics, which is made possible by the stochastic nature of the models. This is quite different from the parameter identification methods based on deterministic models [155]. The nonlinearity can be attributed to the non-Gaussian statistics of the local energies. Various models for the heat flux are obtained with approximations of the memory integral, and they can be viewed as nonlinear and stochastic extensions of the relaxation models from Cattaneo and Vernotte, and Guyer-Krumhansl models.

Bibliography

[1] Fujii, M., X. Zhang, H. Xie, H. Ago, K. Takahashi, T. Ikuta, H. Abe, and T. Shimizu (2005) "Measuring the thermal conductivity of a single carbon nanotube," *Physical Review Letters*, **95**(6), p. 065502.

[2] Mingo, N. and D. Broido (2005) "Length dependence of carbon nanotube thermal conductivity and the 'problem of long waves'," *Nano Letters*, **5**(7), pp. 1221–1225.

[3] Osman, M. A. and D. Srivastava (2005) "Molecular dynamics simulation of heat pulse propagation in single-wall carbon nanotubes," *Physical Review B*, **72**(12), p. 125413.

[4] Leach, A. R. (2001) *Molecular modelling: Principles and applications*, Pearson education.

[5] Chorin, A. and P. Stinis (2007) "Problem reduction, renormalization, and memory," *Communications in Applied Mathematics and Computational Science*, **1**(1), pp. 1–27.

[6] Flaschka, H., M. G. Forest, and D. McLaughlin (1980) "Multiphase averaging and the inverse spectral solution of the Korteweg-de Vries equation," *Communications on Pure and Applied Mathematics*, **33**(6), pp. 739–784.

[7] Pavliotis, G. and A. Stuart (2005) "Periodic homogenization for inertial particles," *Physica D: Nonlinear Phenomena*, **204**(3-4), pp. 161–187.

[8] Bai, Z., P. M. Dewilde, and R. W. Freund (2005) "Reduced-order modeling," *Handbook of Numerical Analysis*, **13**, pp. 825–895.

[9] Noid, W. G. (2013) "Perspective: Coarse-grained models for biomolecular systems," *The Journal of Chemical Physics*, **139**(9), p. 090901.

[10] Gramada, A. and P. E. Bourne (2011) "Coarse-graining the electrostatic potential via distributed multipole expansions," *Computer Physics Communications*, **182**(7), pp. 1455–1462.

[11] Rudd, R. E. and J. Q. Broughton (2005) "Coarse-grained molecular dynamics: Nonlinear finite elements and finite temperature," *Physical Review B*, **72**, p. 144104.

[12] Lehoucq, R. B. and S. A. Silling (2007) "Statistical coarse-graining of molecular dynamics into peridynamics," *Report SAND2007-6410, Sandia National Laboratories, Albuquerque, New Mexico*.

[13] Izvekov, S. and G. A. Voth (2006) "Modeling real dynamics in the coarse-grained representation of condensed phase systems," *The Journal of Chemical Physics*, **125**, pp. 151101–151104.

[14] Katsoulakis, M. A., A. J. Majda, and D. G. Vlachos (2003) "Coarse-grained stochastic processes for microscopic lattice systems," *Proceedings of the National Academy of Sciences*, **100**(3), pp. 782–787.

[15] Cockburn, B., S. Hou, and C.-W. Shu (1990) "The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV. The multidimensional case," *Mathematics of Computation*, **54**(190), pp. 545–581.

[16] Li, X. (2010) "A coarse-grained molecular dynamics model for crystalline solids," *International Journal for Numerical Methods in Engineering*, **83**(8-9), pp. 986–997.

[17] Baaden, M. and S. J. Marrink (2013) "Coarse-grain modelling of protein–protein interactions," *Current opinion in structural biology*, **23**(6), pp. 878–886.

[18] Gohlke, H. and M. Thorpe (2006) "A Natural Coarse Graining for Simulating Large Biomolecular Motion," *Biophysical Journal*, **91**(6), pp. 2115–2120.

[19] Golubkov, P. A. and P. Ren (2006) "Generalized coarse-grained model based on point multipole and Gay-Berne potentials," *The Journal of Chemical Physics*, **125**(6), p. 064103.

[20] Nielsen, S. O., R. E. Bulo, P. B. Moore, and B. Ensing (2010) "Recent progress in adaptive multiscale molecular dynamics simulations of soft matter," *Physical Chemistry Chemical Physics*, **12**(39), p. 12401.

[21] Noid, W. G., J. W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen (2008) "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models," *The Journal of Chemical Physics*, **128**, p. 244114.

[22] Poulain, P., A. Saladin, B. Hartmann, and C. Pravost (2008) "Insights on protein-DNA recognition by coarse grain modelling," *The Journal of Computational Chemistry*, **29**(15), pp. 2582–2592.

[23] Praprotnik, M., L. D. Site, and K. Kremer (2008) "Multiscale simulation of soft Matter: From scale bridging to adaptive resolution," *Annual Review of Physical Chemistry*, **59**(1), pp. 545–571.

[24] Riniker, S., J. R. Allison, and W. F. van Gunsteren (2012) "On developing coarse-grained models for biomolecular simulation: a review," *Physical Chemistry Chemical Physics*, **14**(36), p. 12423.

[25] Rudzinski, J. F. and W. G. Noid (2012) "The role of many-body correlations in determining potentials for coarse-grained models of equilibrium structure," *The Journal of Physical Chemistry B*, **116**(29), pp. 8621–8635.

[26] Shi, Q., P. Liu, and G. A. Voth (2008) "Coarse-graining in interaction space: an analytical approximation for the effective short-ranged electrostatics," *The Journal of Physical Chemistry B*, **112**(50), pp. 16230–16237.

[27] Stepanova, M. (2007) "Dynamics of essential collective motions in proteins: theory," *Physical Review E*, **76**(5), p. 051918.

[28] Zhang, Z., L. Lu, W. G. Noid, V. Krishna, J. Pfaendtner, and G. A. Voth (2008) "A Systematic Methodology for Defining Coarse-Grained Sites in Large Biomolecules," *Biophysical Journal*, **95**(11), pp. 5073–5083.

[29] Chorin, A. J., O. H. Hald, and R. Kupferman (2002) "Optimal prediction with memory," *Physica D: Nonlinear Phenomena*, **166**(3-4), pp. 239–257.

[30] Mori, H. (1965) "Transport, collective motion, and Brownian motion," *Progress of Theoretical Physics*, **33**(3), pp. 423–455.

[31] Zwanzig, R. (1973) "Nonlinear generalized Langevin equations," *Journal of Statistical Physics*, **9**(3), pp. 215–220.

[32] Koopman, B. O. (1931) "Hamiltonian systems and transformation in Hilbert space," *Proceedings of the National Academy of Sciences*, **17**(5), pp. 315–318.

[33] Mori, H. (1965) "A continued-fraction representation of the time-correlation functions," *Progress of Theoretical Physics*, **34**, pp. 399–416.

[34] Zwanzig, R. (2001) *Nonequilibrium statistical mechanics*, Oxford University Press.

[35] Chorin, A. J., O. H. Hald, and R. Kupferman (2000) "Optimal prediction and the Mori-Zwanzig representation of irreversible processes," *Proceedings of the National Academy of Sciences*, **97**(7), pp. 2968–2973.

[36] Toda, M., R. Kubo, and N. Hashitsume (1983) *Statistical physics II. Nonequilibrium statistical mechanics*, Springer.

[37] Chorin, A. J. and O. H. Hald (2009) *Stochastic tools in mathematics and science*, vol. 3, Springer.

[38] Kubo, R. (1966) "The fluctuation-dissipation theorem," *Reports on Progress in Physics*, **29(1)**, pp. 255 – 284.

[39] Adelman, S. A. and J. D. Doll (1974) "Generalized Langevin equation approach for atom/solid-surface scattering: Collinear atom/harmonic chain model," *The Journal of Chemical Physics*, **61**, p. 4242.

[40] Berkowitz, M., J. D. Morgan, D. J. Kouri, and J. A. McCammon (1981) "Memory kernels from molecular dynamics," *The Journal of Chemical Physics*, **75**, p. 2462.

[41] Curtarolo, S. and G. Ceder (2002) "Dynamics of an inhomogeneously coarse grained multiscale system," *Physical Review Letters*, **88**(25), p. 255504.

[42] Lange, O. F. and H. Grubmüller (2006) "Collective Langevin dynamics of conformational motions in proteins," *The Journal of Chemical Physics*, **124**, p. 214903.

[43] Oliva, B., X. Daura, E. Querol, F. X. Avilés, and O. Tapia (2000) "A generalized Langevin dynamics approach to model solvent dynamics effects on proteins via a solvent-accessible surface. The carboxypeptidase A inhibitor protein as a model," *Theoretical Chemistry Accounts*, **105**(2), pp. 101–109.

[44] Kauzlarić, D., J. T. Meier, P. Español, S. Succi, A. Greiner, and J. G. Korvink (2011) "Bottom-up coarse-graining of a simple graphene model: The blob picture," *The Journal of Chemical Physics*, **134**(6), pp. 064106–064106.

[45] Li, X. (2014) "Coarse-graining molecular dynamics models using an extended Galerkin projection method," *International Journal for Numerical Methods in Engineering*, **99**(3), pp. 157–182.

[46] Min, W., G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie (2005) "Observation of a power law memory kernel for distance fluctuation within a single protein molecule," in *SPIE Third International Symposium on Fluctuations and Noise*, International Society for Optics and Photonics, pp. 194–204.

[47] Kou, S. C. and X. S. Xie (2004) "Generalized Langevin equation with fractional Gaussian noise: Subdiffusion within a single protein molecule," *Physical Review Letters*, **93**, p. 180603.

[48] Lepri, S., R. Livi, and A. Politi (2003) "Thermal conduction in classical low-dimensional lattices," *Physics Reports*, **377**(1), pp. 1–80.

[49] Dobson, M. and M. Luskin (2009) "An optimal order error analysis of the one-dimensional quasicontinuum approximation," *SIAM Journal on Numerical Analysis*, **47**, pp. 2455–2475.

[50] Li, X. and P. Ming (2014) "On the effect of ghost force in the quasicontinuum method: Dynamic problems in one dimension," *Communications in Computational Physics*, **15**(3), pp. 647–676.

[51] Ming, P. and J. Yang (2009) "Analysis of a one-dimensional nonlocal quasicontinuum method," *Multiscale Modeling and Simulation*, **7**, pp. 1838–1875.

[52] Lee, M. H. (2016) "Local dynamics in an infinite harmonic chain," *Symmetry*, **8**(4), p. 22.

[53] Florencio Jr, J. and M. H. Lee (1985) "Exact time evolution of a classical harmonic-oscillator chain," *Physical Review A*, **31**(5), p. 3231.

[54] Español, P. (1996) "Dissipative particle dynamics for a harmonic chain: A first-principles derivation," *Physical Review E*, **53**(2), p. 1572.

[55] Kim, J. and I. Sawada (2000) "Dynamics of a harmonic oscillator on the Bethe lattice," *Physical Review E*, **61**(3), p. R2172.

[56] Ashcroft, N. and N. Mermin (1976) *Solid State Physics*, HRW international editions, Holt, Rinehart and Winston.

[57] E, W. and P. Ming (2007) "Cauchy–Born rule and the stability of crystalline solids: Static problems," *Archive for Rational Mechanics and Analysis*, **183**(2), pp. 241–297.

[58] Brenner, S. and R. Scott (2007) *The mathematical theory of finite element methods*, Texts in Applied Mathematics, Springer New York.

[59] Chen, M., X. Li, and C. Liu (2014) "Computation of the memory functions in the generalized Langevin models for collective dynamics of macromolecules," *The Journal of Chemical Physics*, **141**, p. 064112.

[60] Lee, C.-S., Y.-Y. Chen, C.-H. Yu, Y.-C. Hsu, and C.-S. Chen (2017) "Semi-analytical solution for the generalized absorbing boundary condition in molecular dynamics simulations," *Computational Mechanics*, **60**(1), pp. 23–37.

[61] Li, X. and E. Weinan (2007) "Variational boundary conditions for molecular dynamics simulations of crystalline solids at finite temperature: Treatment of the thermal bath," *Physical Review B*, **76**(10), p. 104107.

[62] Ciarlet, P. G. (1970) "Discrete maximum principle for finite-difference operators," *Aequationes mathematicae*, **4**, pp. 338–352.

[63] Lu, T.-T. and S.-H. Shiou (2002) "Inverses of 2×2 block matrices," *Computers and Mathematics with Applications*, **43**(1), pp. 119 – 129.

[64] Bleistein, N. and R. A. Handelsman (1975) *Asymptotic expansions of integrals*, Courier Corporation.

[65] Horn, R. and C. Johnson (2012) *Matrix Analysis*, Cambridge University Press.

[66] Harville, D. A. (1997) *Matrix algebra from a statistician's perspective*, Springer, New York.

[67] Hwang, S.-G. (2004) "Cauchy's Interlace Theorem for Eigenvalues of Hermitian Matrices," *The American Mathematical Monthly*, **111**(2), pp. 157–159.

[68] Zhu, Y., J. M. Dominy, and D. Venturi (2017) "Rigorous error estimates for the memory integral in the Mori-Zwanzig formulation," *arXiv preprint arXiv:1708.02235*.

[69] Hijón, C., P. Español, E. Vanden-Eijnden, and R. Delgado-Buscalioni (2010) "Mori-Zwanzig formalism as a practical computational tool," *Faraday discussions*, **144**, pp. 301–322.

[70] Hijón, C., M. Serrano, and P. Español (2006) "Markovian approximation in a coarse-grained description of atomic systems," *The Journal of Chemical Physics*, **125**, p. 204101.

[71] Beylkin, G. and L. Monzón (2010) "Approximation by exponential sums revisited," *Applied and Computational Harmonic Analysis*, **28**(2), pp. 131–149.

[72] Baczewski, A. D. and S. D. Bond (2013) "Numerical integration of the extended variable generalized Langevin equation with a positive Prony representable memory kernel," *The Journal of Chemical Physics*, **139**(4), p. 044107.

[73] Fricks, J., L. Yao, T. C. Elston, and M. G. Forest (2009) "Time-domain methods for diffusive transport in soft matter," *SIAM Journal on Applied Mathematics*, **69**(5), pp. 1277–1308.

[74] Lei, H., N. A. Baker, and X. Li (2016) "Data-driven parameterization of the generalized Langevin equation," *Proceedings of the National Academy of Sciences*, **113**(50), pp. 14183–14188.

[75] Ma, J. (2005) "Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes," *Structure (London, England:1993)*, **13**(3), pp. 373–80.

[76] Chu, W. and X. Li (In press) "The Mori-Zwanzig formalism for the derivation of a fluctuating heat conduction model from molecular dynamics," *Communications in Mathematical Sciences*.

[77] Mendl, C. B. and H. Spohn (2015) "Current fluctuations for anharmonic chains in thermal equilibrium," *The Journal of Statistical Mechanics: Theory and Experiment*, **2015**(3), p. P03007.

[78] Spohn, H. (2014) "Nonlinear fluctuating hydrodynamics for anharmonic chains," *Journal of Statistical Physics*, **154**(5), pp. 1191–1227.

[79] Ma, L., X. Li, and C. Liu (2016) "The derivation and approximation of coarse-grained dynamics from Langevin dynamics," *The Journal of Chemical Physics*, **145**(20), p. 204117.

[80] Chu, W. and X. Li (2018) "On the asymptotic behavior of the kernel function in the generalized Langevin equation: A one-dimensional lattice model," *Journal of Statistical Physics*, **170**(2), pp. 378–398.

[81] Katsoulakis, M., P. Plecháč, L. Rey-Bellet, and D. Tsagkarogiannis (2014) "Coarse-graining schemes for stochastic lattice systems with short and long-range interactions," *Mathematics of Computation*, **83**(288), pp. 1757–1793.

[82] Li, Z., X. Bian, X. Li, and G. E. Karniadakis (2015) "Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism," *The Journal of Chemical Physics*, **143**(24), p. 243128.

[83] Ottobre, M. and G. Pavliotis (2011) "Asymptotic analysis for the generalized Langevin equation," *Nonlinearity*, **24**(5), p. 1629.

[84] Ceriotti, M., G. Bussi, and M. Parrinello (2010) "Colored-Noise Thermostats à la Carte," *Journal of Chemical Theory and Computation*, **6**(4), pp. 1170–1180.

[85] Darve, E., J. Solomon, and A. Kia (2009) "Computing generalized Langevin equations and generalized Fokker-Planck equations," *Proceedings of the National Academy of Sciences*, **106**(27), pp. 10884–10889.

[86] Li, Z., H. S. Lee, E. Darve, and G. E. Karniadakis (2017) "Computing the non-Markovian coarse-grained interactions derived from the Mori-Zwanzig formalism in molecular systems: Application to polymer melts," *The Journal of Chemical Physics*, **146**(1), p. 014104.

[87] Pavliotis, G. A. (2014) *Stochastic processes and applications: Diffusion processes, the Fokker-Planck and Langevin equations*, vol. 60, Springer.

[88] Kauzlarić, D., P. Español, A. Greiner, and S. Succi (2012) "Markovian dissipative coarse grained molecular dynamics for a simple 2D graphene model," *The Journal of Chemical Physics*, **137**(23), p. 234103.

[89] Bleistein, N. and R. A. Handelsman (1986) *Asymptotic expansions of integrals*, Dover.

[90] Ladd, A. J., B. Moran, and W. G. Hoover (1986) "Lattice thermal conductivity: A comparison of molecular dynamics and anharmonic lattice dynamics," *Physical Review B*, **34**(8), p. 5058.

[91] Volz, S. G. and G. Chen (1999) "Molecular dynamics simulation of thermal conductivity of silicon nanowires," *Applied Physics Letters*, **75**(14), pp. 2056–2058.

[92] Yang, J. Z., X. Wu, and X. Li (2012) "A generalized Irving–Kirkwood formula for the calculation of stress in molecular dynamics models," *The Journal of Chemical Physics*, **137**(13), p. 134104.

[93] Li, X. (2014) "Heat conduction in nanoscale materials: A statistical-mechanics derivation of the local heat flux," *Physical Review E*, **90**(3), p. 032112.

[94] Davis, P. J. (2012) *Circulant matrices*, American Mathematical Society.

[95] Berber, S., Y.-K. Kwon, and D. Tománek (2000) "Unusually high thermal conductivity of carbon nanotubes," *Physical Review Letters*, **84**(20), p. 4613.

[96] Lepri, S., R. Livi, and A. Politi (1998) "On the anomalous thermal conductivity of one-dimensional lattices," *EPL (Europhysics Letters)*, **43**(3), p. 271.

[97] Schelling, P. K., S. R. Phillpot, and P. Keblinski (2002) "Comparison of atomic-level simulation methods for computing thermal conductivity," *Physical Review B*, **65**(14), p. 144306.

[98] Yoshida, H. (1990) "Construction of higher order symplectic integrators," *Physics Letters A*, **150**(5-7), pp. 262–268.

[99] Cattaneo, C. (1958) "A form of heat-conduction equations which eliminates the paradox of instantaneous propagation," *Comptes Rendus*, **247**, p. 431.

[100] Vernotte, P. (1958) "Paradoxes in the continuous theory of the heat equation," *CR Acad. Sci*, **246**(3), pp. 154–3.

[101] Tzou, D. Y. (1995) "Experimental support for the lagging behavior in heat propagation," *The Journal of Thermophysics and Heat Transfer*, **9**(4), pp. 686–693.

[102] Lepri, S., R. Livi, and A. Politi (1997) "Heat conduction in chains of nonlinear oscillators," *Physical Review letters*, **78**(10), p. 1896.

[103] Garrido, P. L., P. I. Hurtado, and B. Nadrowski (2001) "Simple one-dimensional model of heat conduction which obeys Fourier's law," *Physical Review Letters*, **86**(24), p. 5486.

[104] Donadio, D. and G. Galli (2009) "Atomistic simulations of heat transport in silicon nanowires," *Physical Review Letters*, **102**(19), p. 195901.

[105] Tzou, D. (2011) "Nonlocal behavior in phonon transport," *International Journal of Heat and Mass Transfer*, **15**, pp. 475–481.

[106] Callaway, J. (1959) "Model for lattice thermal conductivity at low temperatures," *Physical Review*, **113**(4), p. 1046.

[107] Peierles, R. E. (1955) *Quantum theory of solids*, Academic Press.

[108] Ziman, J. M. (2001) *Electrons and Phonons*, Oxford University Press.

[109] Chen, G. (2001) "Ballistic-diffusive heat-conduction equations," *Physical Review Letters*, **86**(11), p. 2297.

[110] Joshi, A. and A. Majumdar (1993) "Transient ballistic and diffusive phonon heat transport in thin films," *Journal of Applied Physics*, **74**(1), pp. 31–39.

[111] Majumdar, A. (1993) "Microscale heat conduction in dielectric thin films," *ASME Journal of Heat Transfer*, **115**, pp. 7–16.

[112] Sverdrup, P. G., Y. S. Ju, and K. E. Goodson (2001) "Sub-continuum simulations of heat conduction in silicon-on-insulator transistors," *Journal of Heat Transfer*, **123**(1), pp. 130–137.

[113] Xu, M. and X. Li (2012) "The modeling of nanoscale heat conduction by Boltzmann transport equation," *International Journal of Heat and Mass Transfer*, **55**(7-8), pp. 1905–1910.

[114] Bhatnagar, P. L., E. P. Gross, and M. Krook (1954) "A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems," *Physical Review*, **94**(3), p. 511.

[115] Allen, M. P. and D. J. Tildesley (1989) *Computer Simulation of Liquids*, Oxford University Press.

[116] Frenkel, D. and B. Smit (2002) *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed., Academic Press.

[117] Gill, S., Z. Jia, B. Leimkuhler, and A. Cocks (2006) "Rapid thermal equilibration in coarse-grained molecular dynamics," *Physical Review B*, **73**(18), p. 184304.

[118] Jolley, K. and S. Gill (2009) "Modelling transient heat conduction in solids at multiple length and time scales: A coupled non-equilibrium molecular dynamics/continuum approach," *Journal of Computational Physics*, **228**(19), pp. 7412–7425.

[119] Che, J., T. Cağom, W. Deng, and W. A. G. III (2000) "Thermal conductivity of diamond and related materials from molecular dynamics simulations," *The Journal of Chemical Physics*, **22**.

[120] Chen, J., G. Zhang, and B. Li (2010) "Molecular dynamics simulations of heat conduction in nanostructures: effect of heat bath," *Journal of the Physical Society of Japan*, **79**(7), p. 074604.

[121] Henry, A. S. and G. Chen (2008) "Spectral phonon transport properties of silicon based on molecular dynamics simulations and lattice dynamics," *Journal of Computational and Theoretical Nanoscience*, **5**(2), pp. 141–152.

[122] McGaughey, A. J. H. and M. Ka (2006) "Phonon Transport in Molecular Dynamics Simulations: Formulation and Thermal Conductivity Prediction," *Advances in Heat Transfer*, **39**, p. 169.

[123] Wang, J.-S. (2007) "Quantum thermal transport from classical molecular dynamics," *Physical Review Letters*, **99**(16), p. 160601.

[124] Wang, L., B. Hu, and B. Li (2012) "Logarithmic divergent thermal conductivity in two-dimensional nonlinear lattices," *Physical Review E*, **86**, p. 040101.

[125] McGaughey, A. J. H. and M. Kaviany (2003) "Thermal conductivity decomposition and analysis using molecular dynamics simulations. Part I. Lennard-Jones argon," *International Journal of Heat and Mass Transfer*, **47**, pp. 1783–1798.

[126] Pop, E., D. Mann, Q. Wang, K. Goodson, and H. Dai (2006) "Thermal conductance of an individual single-wall carbon nanotube above Room Temperature," *Nano Letters*, **6**, pp. 96–100.

[127] Volkova, A. N. and L. V. Zhigileib (2012) "Heat conduction in carbon nanotube materials: Strong effect of intrinsic thermal conductivity of carbon nanotubes," *Applied Physics Letters*, **101**, p. 043113.

[128] Li, D., Y. Wu, P. Kim, L. Shi, P. Yang, and A. Majumdar (2003) "Thermal conductivity of individual silicon nanowires," *Applied Physics Letters*, **83**(14), pp. 2934–2936.

[129] Dames, C. and G. Chen (2004) "Theoretical phonon thermal conductivity of Si/Ge superlattice nanowires," *Journal of Applied Physics*, **95**(2), pp. 682–693.

[130] Wang, S.-C., X.-G. Liang, X.-H. Xu, and T. Ohara (2009) "Thermal conductivity of silicon nanowire by nonequilibrium molecular dynamics simulations," *Journal of Applied Physics*, **105**(1), pp. 014316–014316.

[131] Lu, X., W. Shen, and J. Chu (2002) "Size effect on the thermal conductivity of nanowires," *Journal of Applied Physics*, **91**(3), pp. 1542–1552.

[132] Chen, K.-Q., W.-X. Li, W. Duan, Z. Shuai, and B.-L. Gu (2005) "Effect of defects on the thermal conductivity in a nanowire," *Physical Review B*, **72**(4), p. 045422.

[133] Chen, Y., D. Li, J. Yang, Y. Wu, J. R. Lukes, and A. Majumdar (2004) "Molecular dynamics study of the lattice thermal conductivity of Kr/Ar superlattice nanowires," *Physica B: Condensed Matter*, **349**(1), pp. 270–280.

[134] Chang, C.-W., D. Okawa, H. Garcia, A. Majumdar, and A. Zettl (2008) "Breakdown of Fourier's law in nanotube thermal conductors," *Physical Review Letters*, **101**(7), p. 075903.

[135] Kim, P., L. Shi, A. Majumdar, and P. McEuen (2001) "Thermal transport measurements of individual multiwalled nanotubes," *Physical Review Letters*, **87**, p. 215502.

[136] Hone, J., M. Whitney, C. Piskoti, and A. Zettl (1999) "Thermal conductivity of single-walled carbon nanotubes," *Physical Review B*, **59**(4), p. R2514.

[137] Maruyama, S. (2002) "A molecular dynamics simulation of heat conduction in finite length SWNTs," *Physica B: Condensed Matter*, **323**(1), pp. 193–195.

[138] Yu, C., L. Shi, Z. Yao, D. Li, and A. Majumdar (2005) "Thermal conductance and thermopower of an individual single-wall carbon nanotube," *Nano Letters*, **5**(9), pp. 1842–1846.

[139] Osman, M. A. and D. Srivastava (2001) "Temperature dependence of the thermal conductivity of single-wall carbon nanotubes," *Nanotechnology*, **12**(1), p. 21.

[140] Hu, J., X. Ruan, and Y. P. Chen (2009) "Thermal conductivity and thermal rectification in graphene nanoribbons: a molecular dynamics study," *Nano Letters*, **9**(7), pp. 2730–2735.

[141] Savin, A. V., Y. S. Kivshar, and B. Hu (2010) "Suppression of thermal conductivity in graphene nanoribbons with rough edges," *Physical Review B*, **82**(19), p. 195422.

[142] Evans, W. J., L. Hu, and P. Keblinski (2010) "Thermal conductivity of graphene ribbons from equilibrium molecular dynamics: Effect of ribbon width, edge roughness, and hydrogen termination," *Applied Physics Letters*, **96**(20), pp. 203112–203112.

[143] Tersoff, J. (1986) "New empirical model for the structural properties of silicon," *Physical Review Letters*, **56**, pp. 632 – 635.

[144] Wu, X. and X. Li (2015) "On consistent definitions of momentum and energy fluxes for molecular dynamics models with multi-body interatomic potentials," *Modelling and Simulation in Materials Science and Engineering*, **23**(1), p. 015003.

[145] Nosé, S. (1984) "A molecular dynamics method for simulations in the canonical ensemble," *Mol. Phys.*, **52**(2), pp. 255–268.

[146] Tuckerman, M. E., B. J. Berne, and G. J. Martyna (1990) "Reversible multiple time scale molecular dynamics," *The Journal of Chemical Physics*, **97**, pp. 1990–2001.

[147] Faure, G., R. Delgado-Buscalioni, and P. Español (2017) "The entropy of a complex molecule," *The Journal of Chemical Physics*, **146**(22), p. 224106.

[148] Kotz, S., T. Kozubowski, and K. Podgorski (2012) *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*, Springer Science and Business Media.

[149] Risken, H. (1989) "The Fokker-Planck equation. Methods of solution and applications," *Springer series in synergetics*, **301**.

[150] Tian, T. and K. Burrage (2001) "Implicit Taylor methods for stiff stochastic differential equations," *Applied Numerical Mathematics*, **38**(1-2), pp. 167–185.

[151] Leimkuhler, B. and C. Matthews (2015) *Molecular dynamics, with deterministic and stochastic numerical methods*, Springer, New York.

[152] Berry, J., K. Elder, and M. Grant (2008) "Melting at dislocations and grain boundaries: A phase field crystal study," *Physical Review B*, **77**(22), p. 224114.

[153] Donev, A., E. Vanden-Eijnden, A. Garcia, and J. Bell (2010) "On the accuracy of finite-volume schemes for fluctuating hydrodynamics," *Communications in Applied Mathematics and Computational Science*, **5**(2), pp. 149–197.

[154] Chen, Y. (2006) "Local stress and heat flux in atomistic systems involving three-body forces," *The Journal of Chemical Physics*, **124**, p. 054113.

[155] Singh, A. and E. B. Tadmor (2015) "Thermal parameter identification for non-Fourier heat transfer from molecular dynamics," *Journal of Computational Physics*, **299**, pp. 667–686.

Vita

Weiqi Chu

Weiqi Chu received her Bachelor of Science degree in mathematics from the School of Mathematical Sciences at Peking University, China, in 2014. That year she began her doctoral degree in computational mathematics at Pennsylvania State University.