Benchmarking Energy Calculations Using Formal Proofs

Ejike D. Ugwuanyi[1], Colin T. Jones[1], John Velkey[1], and Tyler R. Josephson[1,2]

[1]Department of Chemical, Biochemical, and Environmental Engineering,
University of Maryland, Baltimore County, Baltimore, MD, USA
[2]Department of Computer Science and Electrical Engineering,
University of Maryland, Baltimore County, Baltimore, MD, USA

April 28, 2025

Abstract

Traditional approaches for validating molecular simulations rely on making software open source and transparent, incorporating unit testing, and generally employing human oversight. We propose an approach that eliminates software errors using formal logic, providing proofs of correctness. We use the Lean theorem prover and programming language to create a rigorous, mathematically verified framework for computing molecular interaction energies. We demonstrate this in LeanLJ, a package of functions, proofs, and code execution software that implements Lennard-Jones energy calculations in periodic boundaries. We introduce a strategy that uses polymorphic functions and typeclasses to bridge formal proofs (about idealized real numbers) and executable programs (over floating point numbers). Execution of LeanLJ matches the current gold standard NIST benchmarks, while providing even stronger guarantees, given LeanLJ's grounding in formal mathematics. This approach can be extended to formally verified molecular simulations, in particular, and formally verified scientific computing software, in general.

Keywords: Formal verification, Lean 4, molecular simulations, functional programming.

1. Introduction

Molecular simulations serve as a cornerstone for understanding the behaviour of matter at the atomic scale, enabling predictions of properties and phenomena that underpin disciplines ranging from material science to biophysics. Central to these simulations is the accurate modelling of interactions between particles. Among the most widely used models are the Lennard-Jones potential and the Coulomb interaction, which capture the essential physics of van der Waals forces and electrostatic interactions [1] [2] [3] [4] [5]. The Lennard-Jones potential describes non-bonded interactions between particles, particularly in simple fluids and gases [6] [7]. It describes the interaction between a pair of neutral atoms or molecules based on their distance [8] [9], and efficiently captures the balance between attractive and repulsive forces [1] [10].

Practical simulations also incorporate periodic boundary conditions (PBC) to approximate an infinite system by wrapping particles around the edges of the simulation box [11] [12]. This ensures that no particle is artificially confined by the box itself. When computing interactions, the minimum image convention is applied, which means that each particle interacts only with the closest periodic copy of another particle [13]. Using the shortest physically meaningful (periodic) distance between particles avoids redundant calculations and reduces computational overhead. Long-range effects are approximately included via tail corrections.

Software tools like LAMMPS and Gromacs [14][15][16] allow users to simulate the dynamics of large molecular systems. However, the sheer complexity of these software packages and the systems they intend to model presents challenges in making simulations transparent, reproducible, useable by others and extensible (TRUE) [17]. For example, the SAMPL Challenges (Statistical Assessment of the Modelling of Proteins and Ligands) [18] and the Industrial Fluid Properties Simulation Challenges [19][20] task computational researchers to predict the solvation or binding free energies of small molecules or the thermophysical properties of fluids. Each year, researchers submit highly variable answers, reflecting differences in modelling choices by the researchers (e.g. force fields, simulation conditions, free-energy extrapolation strategies, etc.), as well as more hidden, subtle differences amongst software packages (e.g. default settings for managing Lennard-Jones cut-off and settings for Ewald summation). Projects such as the Molecular Simulation Design Framework (MoSDeF) [17][21] and the Molecular Sciences Software Institute (MolSSI) [22] address these issues by providing reproducible workflows for molecular simulation setup, and by teaching and promoting best practices in software development [23]. Simulation software can also be validated by comparing to benchmarks, such as those on the National Institute of Standards and Technology (NIST) Standard Reference Simulation Website (SRSW) [24].

Category of Error	**Example**	**Intervention**	**Lean**
Syntax	Not closing parentheses	Editor	Editor
Runtime	Accessing element in list that doesn't exist	Run program, program gives error message	Editor
Semantic	Missing a minus sign, transposing tensor indices	Human inspection of the code; test-driven development; observing anomalous behaviour	Editor
Floating-point/ Round-off	Subtracting small values from large values, ill-conditioned matrices	Modifying simulation methods, using double precision floats	–

Table 1: Errors in scientific computing software, and typical interventions. Our goal is to develop an approach to address syntax, runtime, and semantic errors in Lean at the "editor" stage, before code is compiled.

We propose an alternative paradigm for improving reliability of molecular simulations. To illustrate, consider the taxonomy of programming errors in Table 1. The simplest are syntax errors: these are addressed immediately because the code cannot compile, the editor highlights the mistake, and the programmer fixes it. Runtime errors occur during code execution, and may arise when users run the program under conditions not anticipated by the software developers. Nonetheless, runtime errors typically provide a helpful error message pointing toward the source of the issue. The deepest issues are semantic errors in the *meaning* of

the software: Python won't complain about misinterpreting a scientific principle or incorrectly transcribing math into code – it's simply not designed for that. Floating-point and round-off errors create numerical inaccuracies, since computers do not operate with infinite mathematical precision. These are addressed by judicious choices of simulation settings and algorithm choices, and by checking conditions like energy conservation after simulation completion [25].

In this work, we propose a strategy for catching syntax, runtime, and semantic errors at the "editor" stage, namely, before the code is compiled. Our approach stems from the *formal methods* community in computer science, which seeks to prove when software is correct by construction before it is run (also known as static program analysis), unlike traditional testing, which checks for errors by running a program with different inputs. This approach is handy in areas where even small errors can have significant consequences, such as hardware design and critical software systems. A prominent example is the Pentium FDIV bug in Intel processors in the early 1990s, the subject of a multi-million dollar recall stemming from a few misplaced bits in chip software [26]. Now, formal verification approaches prove the correctness of such arithmetic operations in manufactured chips [27]. Our approach most closely resembles that of Selsam, et al., who explored how formal methods can be applied to machine learning systems in Certigrad [28]. By proving the correctness of each step mathematically, this approach exposes errors that might otherwise slip through traditional empirical testing. They highlight the ability of theorem provers like Lean to eliminate entire classes of high-level errors that arise in complex software systems by enforcing correctness through formal reasoning. They demonstrate their approach by building a variational auto-encoder in Lean, proving properties about their implementation of stochastic gradient descent.

Most prior work on formal methods has focused on floating point operations [29]. In molecular simulations, these are typically insignificant, but they can lead to issues in certain settings, such as when programs are run with less precision to increase speed, or under extreme conditions. Tran and Wang [30], explored using interval arithmetic to model the propagation of these uncertainties in molecular dynamics simulations. Our work sets aside the imprecision of floating-point arithmetic, and instead focuses on verifying higher-level logic and mathematics. Incorporating interval arithmetic into our approach would in principle be possible, but these tools are currently in development [31].

Lean 4 is a theorem prover and functional programming language designed to write and verify mathematical proofs, as well as write formally-verified software [32]. Unlike traditional programming languages used for scientific computing (C, FORTRAN, Python, etc.), Lean provides a formally verified framework in which proofs of correctness can be explicitly constructed and checked [33]. We previously used Lean to formalize chemical physics [34]. Lean is also being used to formalize theories in high-energy physics [35]. We also recognize Tomáš Skřivan's ongoing SciLean project, which is working out methods for doing efficient, array-based computations in Lean [36]

In our work [34], we showed how theories in science can be rigorously encoded using the Lean theorem prover, proving the correctness of the derivations, grounding them in the foundations of mathematics. We formalized derivations of the Langmuir and BET adsorption models, meticulously defining assumptions and

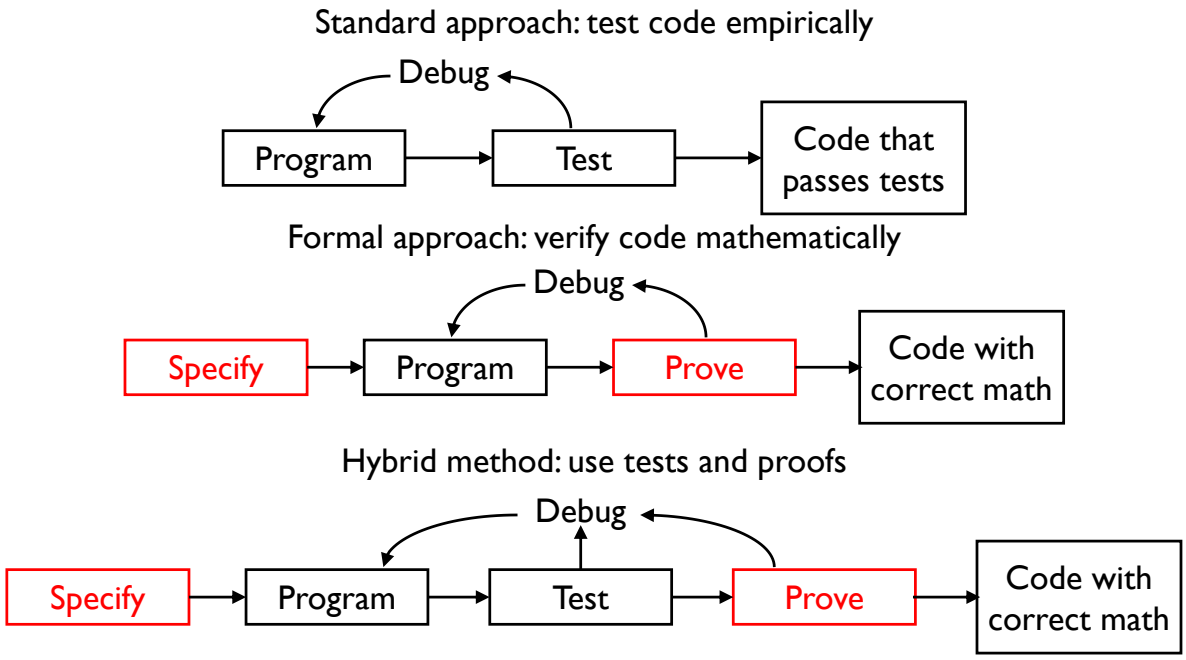


Figure 1: Comparison of code correctness approaches (adapted from [28]): the standard test-debug cycle, formal verification using proofs, and a hybrid method combining tests and formal proofs, that we adopt here.

derivations to ensure mathematical rigour. That work was limited to *proofs* in Lean – we extend that now to executable *programs* with formally-verified properties.

In this paper, we first present a familiar, informal description of Lennard-Jones energy calculations of periodic fluids (Section 3). We then highlight the proof components (especially definitions and theorems) for the formal implementation in Lean (Section 4). Section 5 describes how we implement these energy calculations in Lean, which requires novel approaches using functional programming, type polymorphism, and monads. Section 6 compares our calculations with the results from the NIST SRSW benchmarks [24].

2. Methods

We implemented this using Lean version 4.16.0-rc2, Mathlib 4 at commit `e1a3d4c`, and Visual Studio Code version 1.96. The source code is available in LeanLJ Repository.

3. Informal Description of the Molecular Simulation System

The Lennard-Jones system is modelled as a collection of N particles confined within a cubic simulation box of side length L. The position of each particle is represented as a vector in a three-dimensional space, $\mathbf{r}_i = (x_i, y_i, z_i)$, where $i = 1, 2, \ldots, N$. The interaction between particles is governed by the Lennard-Jones

potential:

$$V_{\mathrm{LJ}}(r_{ij}) = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{1}$$

where r_{ij} is the distance between particles i and j, ε represents the depth of the potential well, and σ is the characteristic length scale.

Periodic boundary conditions (PBCs) are applied to simulate an infinite system as shown in the equation for particle coordinates in the x, y, and z axes, respectively.

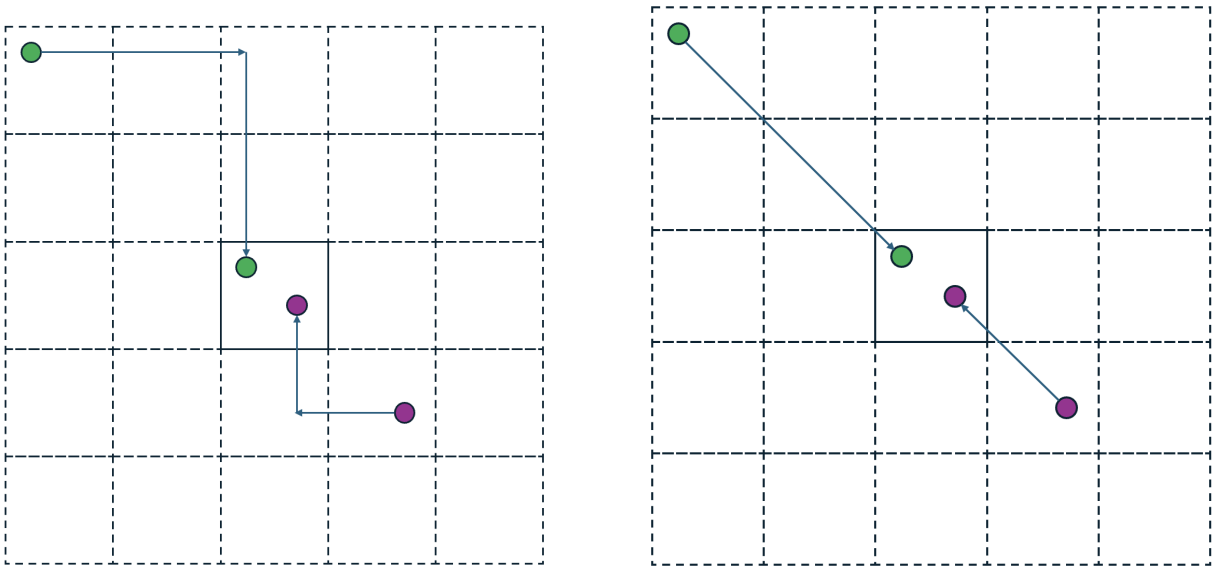
$$x_{i\_\mathrm{wrapped}} = x_i - L \cdot \mathrm{round}\left(\frac{x_i}{L} \right) \tag{2}$$

$$y_{i\_\mathrm{wrapped}} = y_i - L \cdot \mathrm{round}\left(\frac{y_i}{L} \right) \tag{3}$$

$$z_{i\_\mathrm{wrapped}} = z_i - L \cdot \mathrm{round}\left(\frac{z_i}{L} \right) \tag{4}$$

Because the LJ particles are in a system with PBCs, the distance between two particles is not the Euclidean distance, but the minimum image distance, the shortest pairwise distance considering the periodicity of the box as given in the equation below.

$$r_{ij} = \sqrt{\left(\Delta x - L \cdot \mathrm{round}\left(\frac{\Delta x}{L} \right) \right)^2 + \left(\Delta y - L \cdot \mathrm{round}\left(\frac{\Delta y}{L} \right) \right)^2 + \left(\Delta z - L \cdot \mathrm{round}\left(\frac{\Delta z}{L} \right) \right)^2} \tag{5}$$



(a) Periodic boundary conditions.

(b) Minimum image convention.

Figure 2: (a) Periodic boundary conditions: Particles outside the central cubic simulation box are wrapped back into it. Arrows represent the wrapping process along the directions. (b) Minimum image convention: Particles interact with the nearest periodic image, ensuring the shortest distance is used in calculations.

To improve computational efficiency, a cut-off radius r_{c} is introduced. Interactions are considered only for particle pairs that satisfy $r_{ij} \leq r_{\mathrm{c}}$, with contributions beyond this radius set to zero. This truncation neglects a

relatively minor contribution to the potential energy, depending on the cut-off radius r_c as shown in Figure 3.

$$V(r) = \begin{cases} V_{\mathrm{LJ}}(r), & r \leq r_c \\ 0, & r > r_c \end{cases} \tag{6}$$

The Lennard Jones potential function is defined in part: When $r \leq r_c$, the potential is calculated as $4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$, which captures both short-range repulsion and long-range attraction. For $r > r_c$, the potential is set to zero, reflecting the computational practice of truncating interactions beyond the cut-off to save resources. In addition, the inclusion of a cut-off distance makes the function practical for large-scale molecular systems.

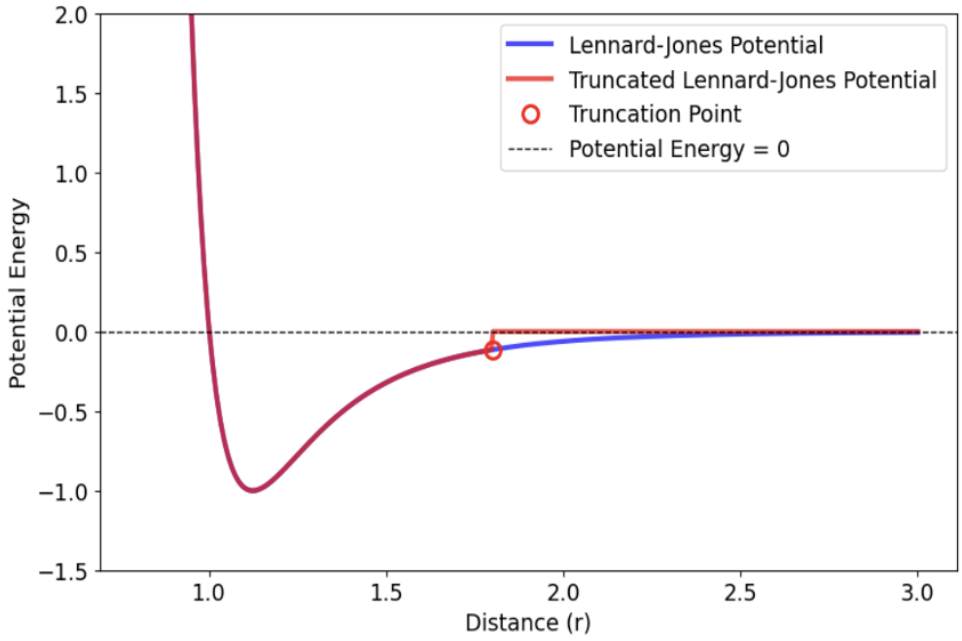


Figure 3: The Lennard-Jones potential, truncated at the cut-off. This plot was generated using Python and matplotlib, since data visualization in Lean is still experimental.

The total internal energy U_{pair} is calculated by summing the energies of the pairs of particles interacting. This is given by the following equation, where $V(r_{ij})$ is the simulated pair potential:

$$U_{\text{pair}} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} V(r_{ij}), \quad \text{where} \quad r_{ij} \leq r_{\text{c}}. \tag{7}$$

The neglected part of the Lennard-Jones potential can be approximately included by incorporating a “Long-Range Correction” (LRC), also known as “tail corrections". This incorporates the ensemble-averaged energy contribution of the particles beyond the cut-off radius, in a manner that only depends on the density of the system and does not require pairwise distance calculations [13]. The LRC is given by:

$$U_{LRC} = \frac{1}{2} 4\pi\rho \int_{r_c}^{\infty} r^2 V(r) dr, \tag{8}$$

where ρ is the density of the system, r_c is the cut-off radius, and $V(r)$ is the pairwise energy function.

When $V(r) = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)$, this integrates to:

$$U_{LRC} = \frac{1}{2}4\pi\rho\int_{r_c}^{\infty} r^2 V_{LJ}(r)dr \tag{9}$$

$$= \frac{1}{2}4\pi\rho\int_{r_c}^{\infty} r^2 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)dr \tag{10}$$

$$= \frac{8\pi\varepsilon\rho}{r_c^3}\left(\frac{\sigma^{12}}{9r_c^6} - \frac{\sigma^6}{3}\right) \tag{11}$$

4. Formally Defining the Mathematics

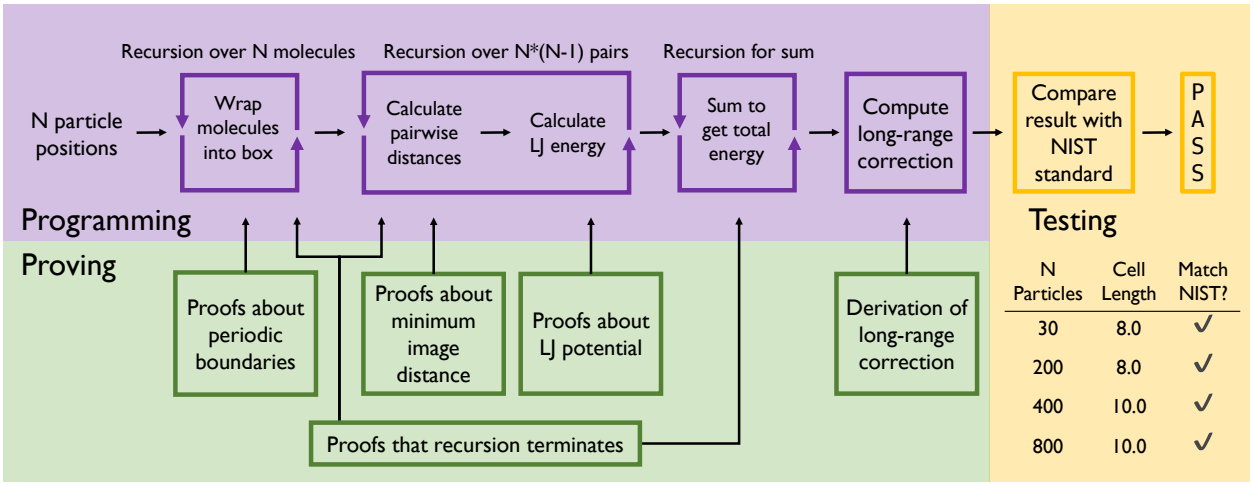


Figure 4: Workflow of the Lennard-Jones energy calculation using LeanLJ. The process involves recursive programming, formal proofs, and comparison to NIST benchmarks.

The previous section was an *informal* description of these concepts; now, we turn to a *formal* description, expressed as Lean code. Lean provides a structured framework to rigorously define the components of our system and prove their properties. Figure 4 illustrates our code and the means by which it is verified. In this section and the next, we describe the components of the system. We start by illustrating Lean's capabilities as a theorem prover.

4.1. Introduction to Lean syntax

Here are a few examples to illustrate the syntax of Lean 4. Lean's basic objects include types, tactics, definitions, and theorems; we don't introduce any custom types or tactics in this work, so we will focus on definitions and theorems. A definition has the following basic structure[1]:

```
def name_of_object (p1 : parameter1) ... : type_of_object := the_def_of_the_object
```

[1] This overview is inspired by the presentation in [35].

A theorem (or equivalently, a lemma) has the following basic structure:

```
theorem name_of_theorem (p1 : parameter1) ... (a1 : assumption1) ... :
     thing_to_be_proved := by
  proof
```

Lean's rich type system enables theorems to be stated and proved; while the user writes code, Lean effectively checks the types of the objects in the code for consistency. Type-checking a theorem object amounts to validating whether it is true. As the user writes the steps in a theorem's proof, Lean provides a concise overview of the current proof goal, as well as the current state of the assumptions and parameters. This information is presented in the "Lean Infoview' in VS Code, in what is known as a *tactic state*, which is organized as follows:

```
p1 : parameter1
...
a1 : assumption1
⊢ current_state_of_goal
```

To learn more about Lean, we highly recommend the textbooks "Mechanics of Proof" by Heather Macbeth [37] and "Functional Programming in Lean" by David Christiansen [33].

4.2. Lennard Jones Potential

We can write the Lennard-Jones potential energy function in multiple ways. In every case, we aim to formally define Eq. 6, using a function that takes four parameters (ε, σ, r, and r_c) and returns the energy between a pair of particles.

The first version of this is `lj`. In this version, all parameters are type $\mathbb{R}$, for the real numbers. Because real numbers are noncomputable, Lean requires this definition to be prefaced with the `noncomputable` keyword. This would be an issue if we were *executing* this function in our calculations, but we intend to execute other versions of this function – see functions `lj_float` and `lj_p` for computable LJ functions in Section 5.

```
noncomputable def lj (r r_c ε σ : ℝ) : ℝ :=
  if r ≤ r_c then
    4 * ε * ((σ / r) ^ 12 - (σ / r) ^ 6)
  else
    0
```

While `lj` may be a natural way to write Eq. 6, alternative formulations are typically used for efficient molecular simulations. For instance, r^{-3} can be computed first, which is then squared to obtain r^{-6}, which can be squared again to obtain r^{-12}. Our function `lj_real` reflects this idea, using intermediate variables

like r6 and r12. Because lj_real is also a function of real numbers, it is also noncomputable.

```
noncomputable def lj_real (r r_c ε σ : ℝ) : ℝ :=
  if r ≤ r_c then
    let r6 := (σ / r) ^ 6
    let r12 := r6 ^ 2
    4 * ε * (r12 - r6)
  else
    0
```

Lean allows us to formally prove the equivalence of these two forms, as shown in the theorem lj_eq, allowing us to use either representations confidently. This capability enables not only correctness, but also flexibility in implementing the most efficient forms for simulation. Keep in mind that we *don't* address floating-point or round-off errors; this guarantee holds only for idealized functions over real numbers, which have infinite precision. If a more-efficient version of a function is mathematically equivalent (over reals) to a base case, but leads to more round-off errors, that won't be detected in our formulation. A more-efficient version that is not mathematically equivalent (e.g. it invokes an approximation) would be shown to be distinct by this approach.

```
theorem lj_eq (r r_c ε σ : ℝ) : lj_real r r_c ε σ = lj r r_c ε σ := by
  unfold lj_real
  unfold lj
  simp
  ring_nf
```

The theorem lj_eq formally proves that lj_real r r_c ε σ = lj r r_c ε σ. This illustrates the syntax of Lean functions: unlike Python, which uses parentheses to denote function application (e.g., lj_real(r, r_c, epsilon, sigma)), Lean uses simple whitespace. In the expression lj_real r r_c ε σ, each argument is applied to the function from left to right, separated by spaces. Thus, lj_real r r_c ε σ represents "apply the function lj_real to these four arguments." This compact syntax is helpful in mathematical reasoning, where function application is so pervasive.

We can also prove various mathematical properties of our LJ function. Theorem cutoff_behavior states that for any $r > r_c$, the value of the Lennard-Jones potential is zero. (The way to read this theorem, is "for real numbers ε, σ, r, and r_c, assuming $r > r_c$, this function evaluates to zero"). This reflects practice of truncating the potential beyond the cut-off distance.

```
  theorem cutoff_behavior (ε σ r r_c : ℝ) (h : r > r_c) :
  lj_real ε σ r r_c = 0 := by
  unfold lj_real
  simp [if_neg (not_le_of_gt h)]
```

Theorem `ljp_eq_le` establishes that, in $0 < r \leq r_c$, the Lennard-Jones potential is $4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$. Lean can use logical operators like $\forall$ (for all) for defining properties of functions.

```
theorem ljp_eq_le {r_c ε σ : ℝ} : ∀ r ∈ {r | r > 0 ∧ r ≤ r_c },
 lj_real r r_c ε σ = 4 * ε * ((σ / r)^12 - (σ / r)^6) := by
  intro r hr
  have h_r_le_rc : r ≤ r_c := hr.2
  unfold lj_real
  rw [if_pos h_r_le_rc]
  ring
```

We also prove the continuity of the function in this range, in Theorem `ljp_continuous_closed_domain`. Continuity is essential in molecular dynamics simulations because forces are evaluated on the basis of energy gradients, and discontinuities can introduce artificial forces, destabilizing numerical integration [38]. Importantly, we do not, indeed we *cannot*, prove that this function is continuous for the whole domain of r; the LJ function diverges at $r = 0$ and undergoes a step change at $r = r_c$. Researchers have implemented alternative truncation methods for the LJ function, such as the truncated and shifted LJ function or the linear force shift function, which would be continuous for all $0 < r$ [39]. These properties could be formalized in Lean, but in this work, we have focused on the simple LJ function.

```
theorem lj_p_continuous_closed_domain (r_c ε σ : ℝ) :
  ContinuousOn (fun r => if r ≤ r_c then 4 * ε * (((σ / r) ^ 6) ^ 2 - (σ / r) ^ 6)
    else 0)
    {r | 0 < r ∧ r ≤ r_c} := by
  have subset_pos : {r | 0 < r ∧ r ≤ r_c} ⊆ {r | r > 0} := by
    intro r hr
    exact hr.1
  have base := (scale_continuous ε σ).mono subset_pos
  apply ContinuousOn.congr base
  intro r hr
  simp [if_pos hr.2]
  left
  ring
```

4.3. Periodic Boundaries

We follow the formulation in Allen and Tildesley [38] in defining functions for wrapping molecules according to periodic boundary conditions (PBCs) and calculating the minimum image distance. The periodic boundary function wraps a position from anywhere in space into the bounds of the simulation box. This function `pbc` takes in a one-dimensional position and box length and outputs a new position (all have type $\mathbb{R}$).

```
noncomputable def pbc_real (pos box_length : ℝ) : ℝ :=
   pos - box_length * round (pos / box_length)
```

We formally proved that the wrapped displacement produced by the periodic boundary condition function lies within the interval $[-L/2, L/2]$ for any real coordinate p and positive box length L. This ensures that particles always interact with the nearest periodic image, which is a key assumption in molecular dynamics simulations. The proof was constructed in Lean by expressing the wrapped position as $L \cdot \delta$, where $\delta = \frac{p}{L} - \text{round}\left(\frac{p}{L}\right)$, and rigorously showing that $|\delta| \leq \frac{1}{2}$, hence $|\text{pbc\_real}(p, L)| \leq \frac{L}{2}$.

```
theorem abs_pbc_le (p L : ℝ) (hL : 0 < L) :
    |pbc_real p L| ≤ L / 2 := by
  dsimp [pbc_real]
  let δ := (p / L) - round (p / L)
  have h_eq : p - L * round (p / L) = L * δ := by
    rw [mul_sub]
    field_simp [hL.ne']
  rw [h_eq, abs_mul, abs_of_pos hL]
  have hδ : |δ| ≤ 1 / 2 := abs_diff_round_le_half (p / L)
  trans L * (1 / 2)
  · exact mul_le_mul_of_nonneg_left hδ hL.le
  · field_simp
```

4.4. Minimum Image Distance

In defining the minimum image distance, we found it more convenient to first define the squared minimum image distance, and then take the square root of that to obtain the minimum image distance. The box length `box_length` and positions `posA` and `posB` are vectors in $\mathbb{R}^N$ (wtih $N = 3$) where each component corresponds to a coordinate in the respective dimension. This is specified using a vector type `Fin 3` $\to \mathbb{R}$. [2] The function iterates over each of the three dimension and computes a displacement, which is adjusted

[2] Lean can handle particularly rich mathematics through its use of *dependent types* – types that depend on a value. Vector is an example of this – it is a subtype of List that depends on a value, the length of the list, which in our case, is 3. This is one way in which Lean avoids runtime errors; before the code compiles, Lean can ensure that a function taking a vector of length `N` will always receive a vector of length `N`.

using the periodic boundary function `pbc_real`. The adjusted displacements are squared and summed over all dimensions. In our `squaredminImageDistance_real` function, the decide tactic is employed in each invocation of the vectors posB and posA, to prove to Lean that elements 0, 1, and 2 are in scope.

```
noncomputable def squaredminImageDistance_real (box_length posA posB : Fin 3 → ℝ) : ℝ
    :=
  let dx := pbc_real (posB ⟨0, by decide⟩ - posA ⟨0, by decide⟩) (box_length ⟨0, by
    decide⟩)
  let dy := pbc_real (posB ⟨1, by decide⟩ - posA ⟨1, by decide⟩) (box_length ⟨1, by
    decide⟩)
  let dz := pbc_real (posB ⟨2, by decide⟩ - posA ⟨2, by decide⟩) (box_length ⟨2, by
    decide⟩)
  dx^2 + dy^2 + dz^2
```

We can prove a neat property of how the periodic boundaries interact with the minimum image distance – that the minimum image distance between arbitrary points in space is equivalent to the minimum image distance between those points, after being wrapped into the simulation box. This is stated in theorem `squaredminImageDistance_theorem`, which requires an inline invoking a λ function to iterate over the box dimensions (for brevity, the proof steps are omitted here, but available on GitHub). This only holds for nonzero box lengths.

```
theorem squaredminImageDistance_theorem (box_length posA posB : Fin 3 → ℝ)
(hL : ∀ i, box_length i ≠ 0) squaredminImageDistance_real box_length posA posB =
   squaredminImageDistance_real box_length (λ i => pbc_real (posA i) (box_length i))
   (λ i => pbc_real (posB i) (box_length i)) := by
   ...
```

Finally, the function `minImageDistance_real` calls the `squaredminImageDistance` function, and takes the square root to obtain the minimum image distance.

```
noncomputable def minImageDistance_real (posA posB box_length : Fin 3 → ℝ) : ℝ :=
  (squaredminImageDistance_real posA posB box_length).sqrt
```

We can also prove that computed distances between particles are guaranteed to be non-negative in all applications of the minimum image convention; this can be useful when non-negativity is invoked in proofs about energy calculations.

```
theorem minImageDistance_real_nonneg ( posA posB box_length : Fin 3 → ℝ) :
  0 ≤ minImageDistance_real  posA posB box_length := by
  unfold minImageDistance_real
  apply Real.sqrt_nonneg
```

We also proved that the minimum image distance between a particle and itself is always zero (theorem `minImageDistance_self`).

```
theorem minImageDistance_real_self (pos box_length : Fin 3 → ℝ) :
  minImageDistance_real pos pos box_length = 0 := by
  unfold minImageDistance_real squaredminImageDistance_real
  have h0 : pbc_real (pos ⟨0, by decide⟩ - pos ⟨0, by decide⟩) (box_length ⟨0, by
    decide⟩) = 0 := by
    simp [pbc_real, sub_self, zero_div, round_zero, mul_zero, sub_zero]
  have h1 : pbc_real (pos ⟨1, by decide⟩ - pos ⟨1, by decide⟩) (box_length ⟨1, by
    decide⟩) = 0 := by
    simp [pbc_real, sub_self, zero_div, round_zero, mul_zero, sub_zero]
  have h2 : pbc_real (pos ⟨2, by decide⟩ - pos ⟨2, by decide⟩) (box_length ⟨2, by
    decide⟩) = 0 := by
    simp [pbc_real, sub_self, zero_div, round_zero, mul_zero, sub_zero]
  rw [h0, h1, h2]
  simp
```

While the above formulations of `pbc_real` and `minImageDistance_real` lead to valid computations and proofs, we are somewhat dissatisfied with the semantics. The `pbc_real` function operates on *particle positions* (i.e. x_i), wrapping them inside the box from outside. When this function is applied in the `minImageDistance` function, it is being applied to a *difference* between particle positions (i.e. $x_j - x_i$). Lean does not complain, because in both cases, these are just real numbers, and everything checks out, but a displacement is nonetheless not the same thing as a position. There may be a way to make this even more rigorous, by defining a custom type for positions and restricting the `pbc_real` function to only operate on such a type, but we kept our approach simpler for now.

4.5. Long-Range Corrections

The long-range correction, given in Eq. 8, is computed using the function `U_LRC`, which depends on ρ, ε, σ, and rc.

```
noncomputable def U_LRC_real (ρ ε σ rc : ℝ) : ℝ :=
  (8 * π * ρ * ε) * ((1/9) * (σ ^ 12 / rc ^ 9) - (1/3) * (σ ^ 6 / rc ^ 3))
```

We can prove that this function follows from the integral definition of U_{LRC}, Eq. 9. The integral `∫ (r :`

$\mathbb{R}$) in Set.Ioi rc is interpreted using measure theory, and refers to an integral over the set Set.Ioi rc, which is the open interval (r_c, ∞). We state the theorem here and omit the proof for brevity, the full proof is available in the LeanLJ Repository.

```
theorem long_range_correction_equality (hr : 0 < rc) (ρ ε σ : ℝ) :
   (2*π*ρ) * ∫ (r : ℝ) in Set.Ioi rc, 4*ε * (r^2 * (((σ / r)^12) - ((σ / r)^6))) =
   U_LRC ρ ε σ rc π := by
```

5. Code Execution

Combining formal proofs with numerical computation is central to this work. In this section, we elaborate on three aspects of programming in Lean. Subsection 5.1 introduces the function for energy summation; in Lean, this must be recursive instead of based on traditional for loops. Subsection 5.2 highlights our approach for bridging computations and proofs using polymorphic functions. Subsection 5.3 describes Lean's approach to input and output.

5.1. Functional Programming

Traditional molecular simulation software is implemented using imperative programming languages (like C and FORTRAN), but Lean is a functional programming language (like Haskell). Imperative programs are about "doing" (following a step-by-step procedure), while functional programs are about "being" (defining what a function *is*, which in Lean, ultimately enables proofs about it). Imperative programming is susceptible to "side effects" that are avoided in functional programming, reducing security risks and improving rigour. Functional programming avoids mutable data types; rather than updating (mutating) existing variables, such as assigning `x=x+1`, when new things must be computed, new variables are assigned. Lean 4 does support some imperative design patterns, but to get guarantees that come from proofs, writing code in a functional style is essentially required.

One of the most stark differences (and most relevant for molecular simulations) between imperative and functional programming is the use of for- and while-loops, which are not supported by Lean. Figure **??** illustrates the typical "double for-loop" used for pairwise energy calculations (using a Python example): first loop over particles i from 1 to N, then over particles j from $i+1$ to N. In Lean, this is done using recursive function calls, instead of for loops, as shown in the `total_energy` function. `total_energy` locally defines a recursive function `energy` and an accumulation variable `acc`. In the central function call, `energy (i+1) j (acc + lj_float r r_c ε σ)`, `energy` adds one LJ energy contribution to the value of `acc`, using particle indices $i+1$ and j to obtain the distance r. The remaining conditions handle increments on the edge cases.

Lean automatically checks functions for termination, which is quite important for recursive functions, lest they get trapped in an infinite loop. In developing this code, we also found execution and comparing to the NIST tests to be valuable, as we developed the logic of the double loop. This double-recursive function is

quite complicated; we highly recommend becoming familiar with singly-recursive functions (such as the factorial function, or a function for adding up items in a list) before tackling more complicated ones. We also incorporated tail recursion to facilitate efficient execution [33].

```python
def total_energy(positions, box_length, cutoff):
    energy = 0.0
    for i in range(len(positions)):
        for j in range(i + 1, len(positions)):
            r_ij = positions[i] - positions[j]
            r_ij = minimum_image_distance(r_ij, box_length)
            distance = (r_ij[0]**2 + r_ij[1]**2 + r_ij[2]**2) ** 0.5
            if distance <= cutoff:
                energy += lennard_jones_potential(distance)
    return energy
```

```
def total_energy (positions : List (Fin 3 → Float))
  (boxLength : Fin 3 → Float) (cutoff ε σ : Float) : Float :=
  let numAtoms := positions.length
  let rec energy : Nat → Nat → Float → Float
    | 0, _, acc => acc
    | i+1, 0, acc => energy i (i - 1) acc
    | i+1, j+1, acc =>
      let posI := positions.get! i
      let posJ := positions.get! j
      let r := minImageDistance posI posJ boxLength
      let e := lj_float r cutoff ε σ
      energy (i+1) j (acc + e)
  energy numAtoms (numAtoms - 1) 0.0
```

Figure 5: Top: Loop-based Python implementation for computing Lennard-Jones pairwise energy with periodic boundary conditions (top). Bottom: the equivalent recursive implementation in Lean 4, which mirrors the same logic using functional programming.

5.2. Polymorphism

In Lean, we can define functions specifically for real numbers ($\mathbb{R}$), which allows us to prove mathematical properties, or for floating-point numbers (`Float`), which enables efficient numerical computation. However, these separate implementations create a trade-off: the `real` version is *non-computable*, meaning it cannot be executed in actual simulations, while the `Float` version is not suitable for formal proofs, as floating-point arithmetic lacks the necessary mathematical structure (in the typical standard for floating point addition, IEEE 754, $0.1 + 0.2 \neq 0.3$). To bridge this gap, *polymorphic functions* are used, allowing the same definition to work for multiple types (Fig. 6). By introducing a generic type α that can subsume both `real` and `Float`, we ensure that our function can operate on both reals ($\mathbb{R}$) for proofs and floats (`Float`) for computations.

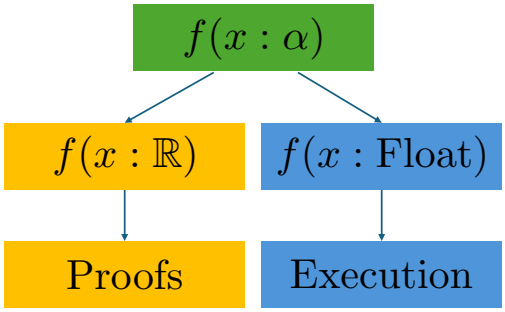


Figure 6: How polymorphic functions link proofs (over idealized real numbers) with execution (over floating point numbers). The polymorphic function f is defined for x with generic type α; proofs about f can be written when x is real, and computations with f can be executed when x is a float.

```
def pbc (position boxLength : α)
    [HSub α α α] [HMul α α α] [HDiv α α α] [HasRound α] : α :=
    position - boxLength * (HasRound.pround (position / boxLength))
```

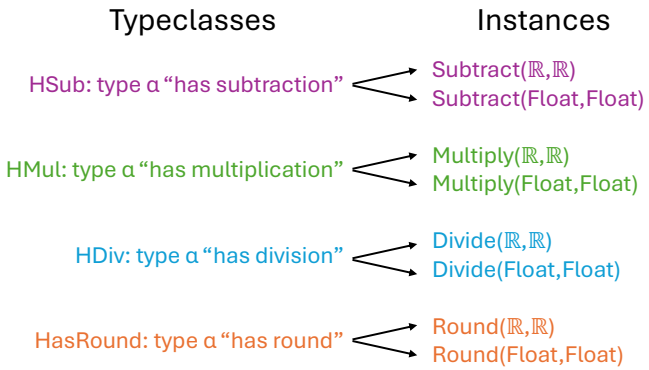


Figure 7: Explanation of the polymorphic `pbc` function. The function is defined over a generic type α, and the required operations—subtraction, multiplication, division, and rounding—are expressed through typeclasses: `HSub`, `HMul`, `HDiv`, and `HasRound`. Each typeclass specifies that the type α must support a given operation. For example, `HSub` α α α means α must support subtraction with two α inputs returning an α result. Concrete instances, such as `Float` and $\mathbb{R}$, implement these typeclasses to enable polymorphic behavior. This allows `pbc` to work with different numeric types, as long as they satisfy the required operations.

To illustrate, consider the function `pbc` (Fig. 7), which wraps a particle's position into the simulation box using periodic boundary conditions. Section 4 showed `pbc_real`, which operates on position and box length with type $\mathbb{R}$; `pbc` operates on position and box length with generic type α. We tell Lean more about what α can be, by using *typeclasses* and *instances*. Specially, `pbc` is defined for any type α that "knows how to" subtract, multiply, divide, and round. These capabilities are provided through the typeclasses `HSub`, `HMul`, `HDiv`, and `HasRound`. For example, the typeclass `HSub` $[\alpha\ \alpha\ \alpha]$ requires that there exists a definition of subtraction between two members of α that would output a third member of α. `HSub`, `HMul`, and `HDiv` are all defined in Mathlib; for rounding, we defined a custom typeclass, since Mathlib didn't already define that connection. This approach allows our definition of `pbc` to be used in two very different ways: with real numbers for formal proofs, and with floating-point numbers for actual simulations.

For some more examples, we provide the implementations for the Lennard-Jones potential in three forms: the *polymorphic* version (α), the *real number* version ($\mathbb{R}$), and the *floating-point* version (`Float`).

```
-- Polymorphic version: Works for both ℝ and Float
def lj_p {α : Type} [LE α] [DecidableLE α] [HDiv α α α] [HPow α α α] [HSub α α α]
 [HMul α α α] [OfNat α 2] [OfNat α 4] [OfNat α 6] [Zero α] (r r_c ε σ : α) : α :=
 if r ≤ r_c then
   let r6 := (σ / r) ^ (6 : α)
   let r12 := r6 ^ (2 : α)
   4 * ε * (r12 - r6)
 else
   0
```

```
-- Real number version: Allows formal proofs but cannot compute
noncomputable def lj_real (r r_c ε σ : ℝ) : ℝ :=
 if r ≤ r_c then
   let r6 := (σ / r) ^ 6
   let r12 := r6 ^ 2
   4 * ε * (r12 - r6)
 else
   0
```

```
-- Floating-point version: Can compute but lacks proof capabilities
def lj_float (r r_c ε σ : Float) : Float :=
 if r ≤ r_c then
   let r6 := (σ / r) ^ 6
   let r12 := r6 ^ 2
   4 * ε * (r12 - r6)
 else
   0
```

Ultimately, we define polymorphic functions for *all* executable functions in the overall execution flow (Fig. 4) and connect them to their real-value counterparts. Most often, we used typeclasses already in Mathlib, but we did need to define custom typeclasses for square roots (for distance calculations) and π (for the long-range correction).

5.3. Input and Output in Lean

Most of Lean is developed in terms of pure functions, whose behaviour can be guaranteed because the argument types limit the domain of the function inputs. By chaining pure functions with pure functions through-and-through, Lean guarantees there are no side effects. But input/output (IO) operations cannot

have the same guarantees. For instance, if one writes a molecular configuration file to disk, then reads it back in, one cannot guarantee that some other process modified it in the meantime.

But to be a useful programming language, Lean must nonetheless have IO. Lean separates this cleanly from its pure functions and math libraries, implementing it in the `IO` monad. This essentially serves as a bridge between the messy, "outside" world and the safe, pure functions inside Lean (Figure 8).

Monads are used to handle many kinds of computation patterns in a clean and consistent way, such as optional values, errors, and non-determinism. For example, the `Option` monad handles missing values, the `Except` monad deals with errors without crashing, and the `List` monad allows multiple possible results from a single computation. They are central in functional programming, but are encountered less often in imperative languages; the interested reader can learn more here [33].

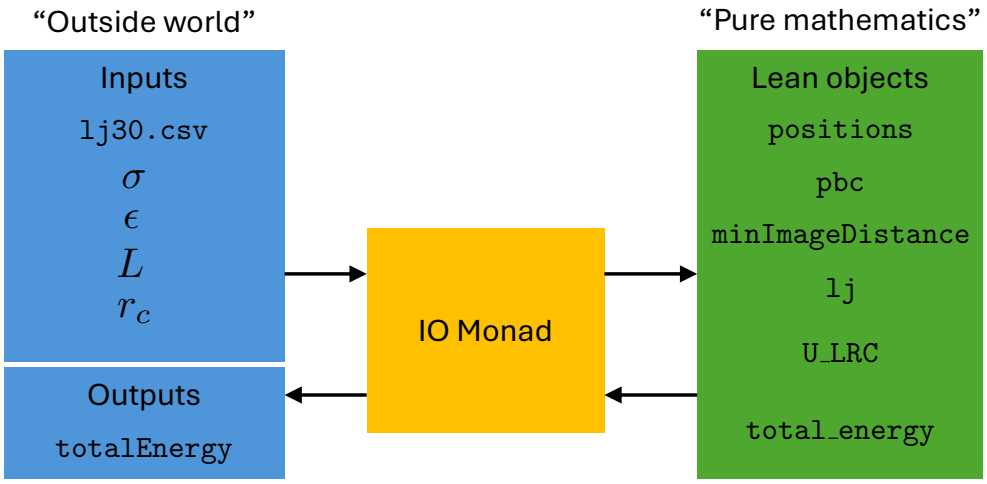


Figure 8: The IO Monad as a bridge that links the verified, pure functions in Lean with the messy real world, where data and simulation inputs reside. The CSV parser uses the IO Monad to read the particle coordinates from lj30.csv into the Lean object `positions`.

To import the configuration files from the NIST SRSW, we first saved them as comma-separated values (CSV) files. We adapted the CSV reader and used it to parse each configuration. In addition, users are asked to manually enter simulation parameters such as the cut-off radius, σ, ε, and the length of the box through the terminal. These user inputs and file reads are examples of interaction with the "outside world," and are handled explicitly in Lean using the IO monad. This makes it clear which parts of the program remain exposed to sources of error – our setup does *not* provide guarantees against sources of error on the "outside" of IO; if an incorrect value for σ were input, Lean would not catch it. This would be a form of semantic error (Table 1) that our current implementation does not avoid.

Proofs in Lean only provide guarantees about pure functions; errors in the I/O layer cannot be validated in this manner. This is why we advocate for both proofs and tests (Fig. 1). For instance, while developing this application, our first approach for reading the configuration failed to read all atoms, leading to incorrect energy calculations. The *proofs* do not catch bugs like these, but the tests do.

Table 2: Comparison of LRC and U_{pair} energy calculations from NIST SRSW [24] and LeanLJ for various particle counts. Energies are reported in scientific notation (reduced units).

Particles	U_{pair} **(Lean)**	U_{pair} **(NIST)**	**LRC (Lean)**	**LRC (NIST)**
30	−1.67903E+01	−1.6790E+01	−5.45166E−01	−5.4517E−01
200	−6.90004E+02	−6.9000E+02	−2.42296E+01	−2.42296E+01
400	−1.14666E+03	−1.1467E+03	−4.96222E+01	−4.9622E+01
800	−4.35154E+03	−4.3515E+03	−1.98488E+02	−1.9849E+02

6. Results

To evaluate our implementation, we compare the pairwise interaction energy (U_{pair}) and long-range correction (LRC) values computed using our Lean code with the NIST Standard Reference Simulation Website (SRSW) benchmark values [24] for `LJ` particles in a cubic box (Table 2). The results show exact agreement for all four systems, within the number of digits provided by NIST.

7. Conclusions and Outlook

In this study, we developed pairwise energy calculations in Lean and compared our results with the values provided by the NIST SRSW benchmark. Our calculations agree to machine precision with the NIST reference values. To be clear, our confidence in our system does *not* stem from its agreement with the NIST benchmark, rather from the theorems we have proved in Lean that certify that the functions in LeanLJ have those specified mathematical properties. For instance, the `pbc` function guarantees that all wrapped particles lie in the interval $[-L/2, L/2]$, and the derivation of the function computing long-range corrections is validated mathematically. We assert that LeanLJ is a more reliable benchmark than the NIST SRSW, at least for the components of the benchmark we have addressed. LeanLJ could be validated even further, by adding to the list of theorems proved about the current functions – we had wished to prove that the pairwise summation function operated over at most $N * (N-1)$ pairs of particles (such would be useful for validating energy summation in systems with polyatomic molecules), but we couldn't quite navigate that proof.

We consider it helpful to reflect on the remaining sources of uncertainty in our code – considering what we've verified, what could still be wrong? First, we are trusting in the axioms of mathematics, as expressed in Lean's core; errors here might compromise Mathlib, on which we depend. Second, our approach to polymorphism exposes us to mistakes in typeclasses. For example, we link the rounding functions for reals and floats – a subtle error in which one rounded 0.5 *down* while another rounded 0.5 *up* would not be detected by Lean (egregious errors like linking functions for round and floor would also be technically possible, so human oversight remains necessarily at this level). Third, we are still exposed to errors in input/output (Section 5.3), and in defining system-specific parameters, such as the force field parameters; these are mitigated by the testing, but do not prevent a user from inputting incorrect parameters for calculations outside the scope of the NIST benchmarks. A fourth source would be vulnerabilities in the broader operating system in which the code is executed. Nonetheless, traditional molecular simulation have far more possibilities for er-

rors, such that many of these concerns are not considered in typical conversations about software correctness.

More broadly, this work demonstrates how Lean can provide a new paradigm for computational molecular simulations, where the results and the entire computational process are provably correct. Logical steps to build on this framework include implementing support for triclinic simulation boxes, Ewald summation for Coulomb interactions, neighbour lists to improve computational efficiency, and of course, integrating Newtons equations of motion to evolve particle trajectories. Some of these are matters of implementation (triclinic cells), but others will involve grappling yet-unresolved questions of how to handle various approximations in a formal environment, such as how to precisely describe the conditions under which neighbour lists can be trusted.

In our previous work, we showed Lean's broader utility for formalizing derivations in science as math proofs [34], digitizing key results in absorption theory, thermodynamics, and kinematics. Joseph Tooby-Smith is also developing derivations in the high-energy physics field [35, 40]. These early works showcase Lean's rigour and versatility for building a library (or libraries) of formally-verified results in diverse areas of science, facilitating rigorous verification of scientific ideas in different disciplines.

LeanLJ demonstrates how *executable* scientific computing software can be tied to such proofs, using polymorphic functions. We believe this approach is quite general for reasoning about idealized real-valued functions in scientific *theories*, while linking these to floating-point executions in scientific computing software. Certigrad's [28] approach is also worth considering; this verifies the high-level mathematics in Lean, and then links high-level functions to unverified, but efficient, linear algebra libraries written in C. Compared to our approach, Certigrad's "bridge" between verified math and executable math consequently happens at a higher level; our polymorphic functions build this bridge at the level of individual math operators (e.g. addition, division) and constants (e.g. π).

Scientific computing benchmarks are typically based on human oversight and software best practices [23]; formal verification in Lean offers an even more rigorous alternative, enabling rigorous math proofs that implemented software is correct. This shift from empirical validation to formal proof introduces a new level of confidence in molecular simulations, setting the stage for more reliable and mathematically sound scientific computing.

Acknowledgements

The authors thank the members of the Lean Zulip for helpful discussions, especially Tomáš Skřivan for his insights into polymorphic functions. This material is based on work supported by the National Science Foundation (NSF) CAREER Award #2236769.

Declaration of Interest

The authors declare no competing financial or personal interests that could influence the work reported in this paper.

Data Availability

All code, proofs, and benchmark files are on the ATOMS Lab Github.

References

[1] Chenyang Sun, Yaning Zhang, Chaofeng Hou, and Wei Ge. A new correlation model for predicting the melting and boiling temperatures of the Lennard-Jones systems. *Physica Scripta*, 98(1):015702, January 2023. URL: https://iopscience.iop.org/article/10.1088/1402-4896/aca443, doi:10.1088/1402-4896/aca443.

[2] Peter Schwerdtfeger, Antony Burrows, and Odile R. Smits. The Lennard Jones Potential Revisited – Analytical Expressions for Vibrational Effects in Cubic and Hexagonal Close-Packed Lattices. *The Journal of Physical Chemistry A*, 125(14):3037–3057, April 2021. arXiv:2012.05413 [cond-mat]. URL: http://arxiv.org/abs/2012.05413, doi:10.1021/acs.jpca.1c00012.

[3] Peter Schwerdtfeger and David J. Wales. 100 Years of the Lennard-Jones Potential. *Journal of Chemical Theory and Computation*, 20(9):3379–3405, May 2024. Publisher: American Chemical Society. doi:10.1021/acs.jctc.4c00135.

[4] Koichi Fumino and Ralf Ludwig. Analyzing the interaction energies between cation and anion in ionic liquids: The subtle balance between Coulomb forces and hydrogen bonding. *Journal of Molecular Liquids*, 192:94–102, April 2014. URL: https://www.sciencedirect.com/science/article/pii/S0167732213002316, doi:10.1016/j.molliq.2013.07.009.

[5] Xipeng Wang, Simón Ramírez-Hinestrosa, Jure Dobnikar, and Daan Frenkel. The Lennard-Jones potential: when (not) to use it. *Physical Chemistry Chemical Physics*, 22(19):10624–10633, May 2020. Publisher: The Royal Society of Chemistry. URL: https://pubs.rsc.org/en/content/articlelanding/2020/cp/c9cp05445f, doi:10.1039/C9CP05445F.

[6] Jiapu Zhang. A Brief Review on Results and Computational Algorithms for Minimizing the Lennard-Jones Potential, December 2010. arXiv:1101.0039 [physics]. URL: http://arxiv.org/abs/1101.0039, doi:10.48550/arXiv.1101.0039.

[7] Junwen Wang and Shengfeng Cheng. Integrated Lennard-Jones Potential between a Sphere and a Thin Rod, May 2024. arXiv:2405.03944 [cond-mat]. URL: http://arxiv.org/abs/2405.03944, doi:10.48550/arXiv.2405.03944.

[8] Loup Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, 159(1):98–103, July 1967. URL: https://link.aps.org/doi/10.1103/PhysRev.159.98, doi:10.1103/PhysRev.159.98.

[9] A. Rahman. Correlations in the Motion of Atoms in Liquid Argon. *Physical Review*, 136(2A):A405–A411, October 1964. URL: https://link.aps.org/doi/10.1103/PhysRev.136.A405, doi:10.1103/PhysRev.136.A405.

[10] J.J. Nicolas, K.E. Gubbins, W.B. Streett, and D.J. Tildesley. Equation of state for the Lennard-Jones fluid. *Molecular Physics*, 37(5):1429–1454, May 1979. URL: https://www.tandfonline.com/doi/full/10.1080/00268977900101051, doi:10.1080/00268977900101051.

[11] Jerry Z. Yang and Xiantao Li. Comparative study of boundary conditions for molecular dynamics simulations of solids at low temperature. *Physical Review B*, 73(22):224111, June 2006. URL: https://link.aps.org/doi/10.1103/PhysRevB.73.224111, doi:10.1103/PhysRevB.73.224111.

[12] Luke Mizzi, Daphne Attard, Ruben Gatt, Krzysztof K. Dudek, Brian Ellul, and Joseph N. Grima. Implementation of periodic boundary conditions for loading of mechanical metamaterials and other complex geometric microstructures using finite element analysis. *Engineering with Computers*, 37(3):1765–1779, July 2021. URL: https://link.springer.com/10.1007/s00366-019-00910-1, doi:10.1007/s00366-019-00910-1.

[13] W. W. Wood and F. R. Parker. Monte Carlo Equation of State of Molecules Interacting with the Lennard-Jones Potential. I. A Supercritical Isotherm at about Twice the Critical Temperature. *The Journal of Chemical Physics*, 27(3):720–733, September 1957. URL: https://pubs.aip.org/jcp/article/27/3/720/204747/Monte-Carlo-Equation-of-State-of-Molecules, doi:10.1063/1.1743822.

[14] Steve Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics*, 117(1):1–19, March 1995. URL: https://linkinghub.elsevier.com/retrieve/pii/S002199918571039X, doi:10.1006/jcph.1995.1039.

[15] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, September 2015. URL: https://linkinghub.elsevier.com/retrieve/pii/S2352711015000059, doi:10.1016/j.softx.2015.06.001.

[16] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J. C. Berendsen. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, December 2005. URL: https://onlinelibrary.wiley.com/doi/10.1002/jcc.20291, doi:10.1002/jcc.20291.

[17] Matthew W. Thompson, Justin B. Gilmer, Ray A. Matsumoto, Co D. Quach, Parashara Shamaprasad, Alexander H. Yang, Christopher R. Iacovella, Clare McCabe, and Peter T. Cummings. Towards molecular simulations that are transparent, reproducible, usable by others, and extensible (TRUE). *Molecular Physics*, 118(9-10):e1742938, June 2020. URL: https://www.tandfonline.com/doi/full/10.1080/00268976.2020.1742938, doi:10.1080/00268976.2020.1742938.

[18] Anthony Nicholls, David L. Mobley, J. Peter Guthrie, John D. Chodera, Christopher I. Bayly, Matthew D. Cooper, and Vijay S. Pande. Predicting Small-Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. *Journal of Medicinal Chemistry*, 51(4):769–779, February 2008. URL: https://pubs.acs.org/doi/10.1021/jm070549%2B, doi:10.1021/jm070549+.

[19] Daniel G. Friend, David J. Frurip, Joseph W. Magee, and James D. Olson. Establishing benchmarks for the first industrial fluids simulation challenge. *Fluid Phase Equilibria*, 217(1):11–15, March 2004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378381203003571, doi:10.1016/S0378-3812(03)00357-1.

[20] Fiona Case, Anne Chaka, Daniel G Friend, David Frurip, Joseph Golab, Russell Johnson, Jonathan Moore, Raymond D Mountain, James Olson, Martin Schiller, and Joey Storer. The first industrial fluid properties simulation challenge. *Fluid Phase Equilibria*, 217(1):1–10, March 2004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378381203002085, doi:10.1016/S0378-3812(03)00208-5.

[21] Brad Crawford, Umesh Timalsina, Co D. Quach, Nicholas C. Craven, Justin B. Gilmer, Clare McCabe, Peter T. Cummings, and Jeffrey J. Potoff. MoSDeF-GOMC: Python Software for the Creation of Scientific Workflows for the Monte Carlo Simulation Engine GOMC. *Journal of Chemical Information and Modeling*, 63(4):1218–1228, February 2023. URL: https://pubs.acs.org/doi/10.1021/acs.jcim.2c01498, doi:10.1021/acs.jcim.2c01498.

[22] Jessica A. Nash, Mohammad Mostafanejad, T. Daniel Crawford, and Ashley Ringer McDonald. MolSSI Education: Empowering the Next Generation of Computational Molecular Scientists. *Computing in Science & Engineering*, 24(3):72–76, May 2022. URL: https://ieeexplore.ieee.org/document/9882989/, doi:10.1109/MCSE.2022.3165607.

[23] Matthew W. Thompson, Justin B. Gilmer, Ray A. Matsumoto, Co D. Quach, Parashara Shamaprasad, Alexander H. Yang, Christopher R. Iacovella, Clare McCabe, and Peter T. Cummings. Towards molecular simulations that are transparent, reproducible, usable by others, and extensible (TRUE). *Molecular Physics*, 118(9-10):e1742938, June 2020. URL: https://www.tandfonline.com/doi/full/10.1080/00268976.2020.1742938, doi:10.1080/00268976.2020.1742938.

[24] Vincent K. Shen, Daniel W. Siderius, William P. Krekelberg, and Harold W. Hatch, editors. *NIST Standard Reference Simulation Website*. Number 173 in NIST Standard Reference Database. National Institute of Standards and Technology, Gaithersburg, MD, 20899. URL: https://doi.org/10.18434/T4M88Q.

[25] Pascal T. Merz and Michael R. Shirts. Testing for physical validity in molecular simulations. *PLOS ONE*, 13(9):e0202764, September 2018. URL: https://dx.plos.org/10.1371/journal.pone.0202764, doi:10.1371/journal.pone.0202764.

[26] D. Price. Pentium FDIV flaw-lessons learned. *IEEE Micro*, 15(2):86–88, April 1995. URL: https://ieeexplore.ieee.org/document/372360/, doi:10.1109/40.372360.

[27] R. Kaivola and N. Narasimhan. Formal Verification of the Pentium® 4 Floating-Point Multiplier. In *Proceedings of the conference on Design, automation and test in Europe*, DATE '02, page 20, USA, March 2002. IEEE Computer Society.

[28] Daniel Selsam, Percy Liang, and David L. Dill. Developing Bug-Free Machine Learning Systems With Formal Mathematics. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3047–3056. PMLR, July 2017. ISSN: 2640-3498. URL: https://proceedings.mlr.press/v70/selsam17a.html.

[29] John Harrison. Floating-Point Verification Using Theorem Proving. In Marco Bernardo and Alessandro Cimatti, editors, *Formal Methods for Hardware Verification*, volume 3965, pages 211–242. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. Series Title: Lecture Notes in Computer Science. URL: http://link.springer.com/10.1007/11757283_8, doi:10.1007/11757283_8.

[30] Anh V. Tran and Yan Wang. Reliable Molecular Dynamics: Uncertainty quantification using interval analysis in molecular dynamics simulation. *Computational Materials Science*, 127:141–160, February 2017. URL: https://linkinghub.elsevier.com/retrieve/pii/S0927025616305171, doi:10.1016/j.commatsci.2016.10.021.

[31] Geoffrey Irving. girving/interval, March 2025. original-date: 2024-05-26T10:00:08Z. URL: https://github.com/girving/interval.

[32] Leonardo de Moura, Soonho Kong, Jeremy Avigad, Floris van Doorn, and Jakob von Raumer. The Lean Theorem Prover (System Description). In Amy P. Felty and Aart Middeldorp, editors, *Automated Deduction - CADE-25*, pages 378–388, Cham, 2015. Springer International Publishing. doi:10.1007/978-3-319-21401-6_26.

[33] David Thrane Christiansen. Functional Programming in Lean.

[34] Maxwell P. Bobbin, Samiha Sharlin, Parivash Feyzishendi, An Hong Dang, Catherine M. Wraback, and Tyler R. Josephson. Formalizing chemical physics using the Lean theorem prover. *Digital Discovery*, 3(2):264–280, 2024. URL: https://xlink.rsc.org/?DOI=D3DD00077J, doi:10.1039/D3DD00077J.

[35] Joseph Tooby-Smith. HepLean: Digitalising high energy physics, May 2024. arXiv:2405.08863 [hep-ph]. URL: http://arxiv.org/abs/2405.08863, doi:10.48550/arXiv.2405.08863.

[36] Tomáš Skřivan. lecopivo/SciLean, April 2025. original-date: 2021-09-27T21:50:10Z. URL: https://github.com/lecopivo/SciLean.

[37] The Mechanics of Proof — The Mechanics of Proof, by Heather Macbeth. URL: https://hrmacbeth.github.io/math2001/index.html.

[38] Michael Patrick Allen and Dominic J. Tildesley. *Computer simulation of liquids*. Oxford university press, Oxford, 2nd ed edition, 2017.

[39] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Number 1 in Computational science series. Academic Press, San Diego, 2nd ed edition, 2002.

[40] Joseph Tooby-Smith. Formalization of physics index notation in Lean 4, November 2024. arXiv:2411.07667 [cs]. URL: http://arxiv.org/abs/2411.07667, doi:10.48550/arXiv.2411.07667.