



Fluid Properties Extraction in Confined Nanochannels with Molecular Dynamics and Symbolic Regression Methods

Dimitrios Angelis 🗅, Filippos Sofos *🔍, Konstantinos Papastamatiou 🕒 and Theodoros E. Karakasidis 🕒

Condensed Matter Physics Laboratory, Department of Physics, University of Thessaly, 35100 Lamia, Greece; dimangelis@uth.gr (D.A.); kpapmath@gmail.com (K.P.); thkarak@uth.gr (T.E.K.) * Correspondence: fsofos@uth.gr

Abstract: In this paper, we propose an alternative road to calculate the transport coefficients of fluids and the slip length inside nano-conduits in a Poiseuille-like geometry. These are all computationally demanding properties that depend on dynamic, thermal, and geometrical characteristics of the implied fluid and the wall material. By introducing the genetic programming-based method of symbolic regression, we are able to derive interpretable data-based mathematical expressions based on previous molecular dynamics simulation data. Emphasis is placed on the physical interpretability of the symbolic expressions. The outcome is a set of mathematical equations, with reduced complexity and increased accuracy, that adhere to existing domain knowledge and can be exploited in fluid property interpolation and extrapolation, bypassing timely simulations when possible.

Keywords: nanochannels; molecular dynamics; diffusion coefficient; shear viscosity; thermal conductivity; slip length



Citation: Angelis, D.; Sofos, F.; Papastamatiou, K.; Karakasidis, T.E. Fluid Properties Extraction in Confined Nanochannels with Molecular Dynamics and Symbolic Regression Methods. *Micromachines* 2023, 14, 1446. https://doi.org/ 10.3390/mi14071446

Academic Editors: Yahui Xue and Jun Lu

Received: 1 July 2023 Revised: 16 July 2023 Accepted: 18 July 2023 Published: 19 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

The utilization of machine learning (ML) techniques in current physics and engineering problems is anticipated to expand across all fields that involve numerical data. The idea of creating new knowledge via predictions based on historical simulation or experimental data is recently trending. Modern computational approaches try to bind techniques with traditional frameworks, taking advantage of their adaptability and ability to provide faster and, oftentimes, accurate predictions in scientific and technological applications. To mention a few, ML-assisted frameworks have been incorporated for atomic-scale force field extraction [1], fluid flow estimation [2], materials science [3–5], construction [6,7], energy [8,9], and chemical industry applications [10–12].

A major challenge that drives the evolution of new and existing ML algorithms is the transparency of the methods employed and interpretability of the results. Towards this perspective, symbolic regression (SR) methods have emerged. Symbolic regression is an ML method that derives mathematical equations in closed form based on available data using Genetic Programming (GP) principles, without prior knowledge of the system under investigation. A pool of candidate equations is stochastically extracted through crossover and mutation operations [13] and the user has to select those that best fit the physical problem. Unlike black-box models, SR allows us to uncover the underlying mechanisms of the studied system, clarify ambiguous relationships between variables, and provide a more physical understanding [14].

In such data-driven approaches, the main factor that determines the performance of each model is based upon the quality and quantity of the input dataset. Keeping in mind that oftentimes experimental measurements are hard to obtain, the common data acquisition method is the utilization of simulation or synthetic data to train an ML model. Since the presence of walls within nano-conduits significantly affects the system, continuum theory is not applicable, and the calculation of transport coefficients becomes ambiguous, making equilibrium methods insufficient [15]. The macroscale approach to fluid dynamics is mainly based on the Navier–Stokes (NS) equations. However, the fundamental hypothesis of the no-slip condition fails to anticipate phenomena occurring at the micro- and nano-scale [16].

In bulk systems, static and dynamic property calculation is carried out at the microand nano- scale with relations usually coming from statistical mechanics. In nano- and macro-devices, confinement between solid surfaces, the implied boundary conditions, as well as the interaction between fluid and wall atoms lead to fundamentally different mechanics of their mass and energy transport [17], affect fluid properties [18], and make their estimation harder [19]. Molecular dynamics (MD) simulations seem to be the most prominent solution for their investigation, which involves calculating particle positions under a given potential, incorporating Newton's second law [20–24]. This characteristic allows the calculation of transport phenomena accurately via equilibrium MD (EMD) and non-equilibrium MD (NEMD) frameworks.

Due to the complex nature of the investigated system, e.g., geometrical discontinuities, molecule adsorption on walls, or even the extension of the wall force field inside neighboring fluid layers [25], EMD approaches are less applicable and NEMD has to be employed [26]. Nevertheless, both in EMD and NEMD, the calculation of the diffusion coefficient, shear viscosity, and thermal conductivity requires complex relationships, such as the Green–Kubo (GK) equations, which are computationally intensive and rely on expensive experimental procedures [27]. This complex nanochannel environment is further amplified by the breakdown of the no-slip condition [19], which gives rise to a fluid/surface property known as the slip length [28,29]. This has an impact on material surface properties and the rate of mass flow. Specifically, surfaces that have been engineered to achieve specific properties can generate desirable slip lengths to control flow rates in various applications.

In this paper, we propose an alternative approach that relies on data derived from MD simulations. In the sections that follow, we discuss conventional computational approaches, present MD simulation details, and argue for SR applicability (Section 2). Furthermore, we present the generated analytical expressions and compare their accuracy characteristics (Section 3), dive deep into the physics hidden behind the obtained mathematical operators (Section 4), and lastly, propose directions for further research that could be incorporated in designing micro-devices at small scales (Section 5).

2. Computational Methods

Next, we present all theoretical relations and parameters employed in the simulation model, as well as the SR techniques and methods incorporated to harness the dataset.

2.1. Mathematical Relations

2.1.1. Particle Interaction Potentials

A well-established particle interaction potential for MD simulations is the Lennard– Jones 12-6 (LJ) potential, $u_{r_{ii}}^{LJ}$, defined between the *i* and *j* particles as [30]

$$u_{r_{ij}}^{LJ} = \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right], & r_{ij} < r_{c} \\ 0, & r_{ij} \ge r_{c} > \sigma \end{cases}$$
(1)

where r_{ij} denotes the *i* and *j* particle distance, ϵ and σ are model energy and size parameters, respectively, and r_c is the cutoff radius. The LJ potential is capable of modeling bulk or multi-particle systems and ϵ and σ parameters are computed accordingly (see Ref. [31] for more details). In nanochannel flows, where fluid particles coexist with solid, wall particles, the wall may be considered either as static or as an array of particles that fluctuate around their equilibrium positions by applying an elastic spring potential [32],

$$u_{wall}(|r(t) - r_{eq}|) = \frac{1}{2}K(|r(t) - r_{eq}|)^2,$$
(2)

where r(t) refers to the particle's position at time t, r_{eq} is its initial lattice position, and K is the spring constant, with a high value of K denoting a rigid wall.

In the case of polar liquids, coulombic long-range interactions must also be taken into account, with the electric potential being [32]

$$E = \frac{Cq_i q_j}{er} \quad r < r_c . \tag{3}$$

Here, q_i and q_j are atoms' charges, *C* is the energy conversion constant, and *e* is the dielectric constant.

2.1.2. Transport Properties

The self-diffusion coefficient (D) can be calculated by either the Einstein equation [31],

$$D = \lim_{t \to \infty} \frac{1}{2dNt} \left\langle \sum_{j=1}^{N} \left[\mathbf{r}_{j}(t) - \mathbf{r}_{j}(0) \right]^{2} \right\rangle, \tag{4}$$

or the GK relation,

$$D = \frac{1}{3N} \int_0^\infty \left\langle \sum_{j=1}^N \mathbf{v}_j(0) \cdot \mathbf{v}_j(t) \right\rangle,\tag{5}$$

where \mathbf{r}_j denotes the j^{th} particle's position vector and \mathbf{v}_j its velocity vector, the dimensionality of the system is d, and N is the total number of particles. The former diffusion coefficient equation is generally incorporated in equilibrium systems; however, by omitting drift from the fluid flow, one could apply them in non-equilibrium configurations as well [33].

Similarly, the GK formalization for thermal conductivity (λ) is given by

$$\lambda = \frac{1}{Vk_B T^2} \int_0^\infty dt < J_q^x(t) \cdot J_q^x(0) >, \tag{6}$$

with J_q being the microscopic heat flow

$$\mathbf{J}_{q} = \frac{1}{2} \sum_{i=1}^{N} m_{i}(v_{i})^{2} \mathbf{v}_{i} - \sum_{i=1}^{N} \sum_{j>1}^{N} \left[\mathbf{r}_{ij} : \frac{\partial u(r_{ij})}{\partial \mathbf{r}_{ij}} - \mathbf{I} \cdot u(r_{ij}) \right] \cdot \mathbf{v}_{i}$$
(7)

where \mathbf{v}_i is the *i*th atom's velocity vector and **I** is the unitary matrix.

Finally, the respective GK equation for shear viscosity, (η) , is

$$\eta = \frac{1}{Vk_BT} \int_0^\infty dt < J_p^{xy}(t) \cdot J_p^{xy}(0) >,$$
(8)

where the off-diagonal elements of the microscopic stress tensor denoted by J_p^{xy} are described by

$$J_{p}^{xy} = \sum_{i=1}^{N} m_{i} v_{i}^{x} v_{i}^{y} - \sum_{i=1}^{N} \sum_{j>1}^{N} r_{ij}^{x} \frac{\partial u(r_{ij})}{\partial r_{ij}^{y}} , \qquad (9)$$

with $u(r_{ij})$ indicating the LJ potential of Equation (1); the distance between *i* and *j* particles is $r_{i,j}$, and v_i^j is the *i*th particle velocity for j = x, y, z directions. We have to note that thermal conductivity (Equation (6)) and shear viscosity (Equation (8)) are derived only for equilibrium conditions or, at least, for systems close to equilibrium [34].

2.1.3. Slip Length

Fluid confinement investigation requires proper designing of the modeling system, as density appears to be non-homogeneous at the wall's vicinity [35] due to wall characteristics (e.g., degree of wettability and mass of each particle) [36], topology of the surface, and

thermal, atomic, or geometrical roughness [37]. As already mentioned, in the macroscale hypothesis of no-slip breakdown [28,38], the slip length (L_s) has to be taken into account, from the following equation [32]:

$$L_s = u_w \left/ \frac{du_{w,z}}{dz} \right|_w,\tag{10}$$

with u_w being fluid velocity at the wall. In a nanochannel of height h, the dimensionless slip length L_s^* , is given by the slip length-to-channel height ratio L_s/h . Its values can be calculated by projecting the fluid's velocity profile until the point where it vanishes inside the wall, as shown in Figure 1.



Figure 1. Molecular dynamics model, with all parameters involved in flows between two infinite plates, as shown in Tables 1 and 2. The mechanisms of diffusion, *D*, viscosity, η , thermal conductivity, λ , and the slip length, *L*_s, are abstractly presented, along with a characteristic velocity and density profile at the nanoscale.

2.2. Simulation Model and Dataset Creation

A Poiseuille-based simulation model is considered for dataset creation [39]. More specifically, a monatomic LJ liquid flows between two infinite, solid plates, in various cases of flat or grooved walls (Figure 1). Next, we refer to fluid quantities with index f and w for walls. The distance between the two plates in the *z*-direction is h, groove height and length are depicted as h_g and h_l , respectively, while periodic boundary conditions are set in *x*- and *y*-directions. A cutoff radius equal to $r_c = 2.5\sigma$ has been considered for the LJ potential (Equation (1)), while ϵ and σ resemble those of argon (Ar), i.e, $\sigma_f = \sigma_w = 0.3405$ nm, $\epsilon_f/k_B = 119.8$ K (k_B : Boltzmann constant). The particle mass has been set equal to $m_{Ar} = 39.95 a.u.$

To account for various surface wettability properties, the ratio $\epsilon_{wf}/\epsilon_{ff}$ corresponds to "hydrophobic" walls when $\epsilon_{wf}/\epsilon_{ff}$ is close to zero and "hydrophilic" walls when $\epsilon_{wf}/\epsilon_{ff}$ approaches unity (see Ref. [40] for details). Fluid particles flow due to an external force F_{ext} applied to every fluid particle, small enough to remain close to the linear regime [41,42]. The application of Nosé–Hoover thermostats at the walls [43,44] keeps the system at the NVT ensemble, with constant temperature. To achieve enhanced thermalization of the wall particles, two distinct thermostats have been used, one for the upper wall and another for the lower wall. The processes of self-diffusion, viscosity, and thermal conductivity, as well as the slip length, are briefly summarized in Figure 1. These properties are affected by confinement, which leads to fluid particle ordering near the walls (as shown in the density profile in Figure 1).

All parameters of the MD simulation, along with the derived transport properties and the slip length, are given in Tables 1 and 2.

Parameters	Diffusion Coefficient	Shear Viscosity	Thermal Conductivity	
Channel height (h)	2.64-100.44	2.64-100.44	2.64–100.44	
External force (F_{ext})	0.0004-0.0743	0.0004-0.0743	0.0004-0.0743	
Energy ratios $(\epsilon_{wf}/\epsilon_{ff})$	0.1–5.0	0.1–5.0	0.1–5.0	
Transport property (D_c, η_c, λ_c)	1.227-10.3575	0.7479-3.1946	1.73–3.1163	
Number of observations	54	54	54	

Table 1. Transport property dataset value range.

Table 2. Slip length dataset value range.

Parameters	Min	Max
Channel height (h)	1.049869	210.0
Groove length to channel height (h_l/h)	0.0119	1.0
Groove height to channel height (h_d/h)	0.0	2.05
Wall-to-fluid energy interaction ratio $(\epsilon_{wf}/\epsilon_{ff})$	0.1	2.236
Wall-to-fluid particle size ratio $(\sigma_{wf} / \sigma_{ff})$	1.0	3.0
Wall-to-fluid particle mass ratio (m_w/m_f)	0.663	20.0
External force (F_{ext})	0.0	4.9
Wall spring constant (K^*)	57.15	10,000.0
Reduced temperature (T^*)	0.8333	2.59
Reduced density (ρ^*)	0.0468	1.303
Slip length-to-channel height ratio (L_s/h)	0.0	7.677928
Number of observations		343

2.3. Symbolic Regression

Symbolic regression is a supervised ML algorithm that constructs a mathematical function based on a given dataset. It is differentiated by other regression instances (e.g., linear, multivariate, polynomial, etc.) as both structure and constitutive parts of the expression are being selected on the fly. This feature diminishes the need to include some kind of prior knowledge into the procedure and allows a completely data-driven evolution. In contrast to the traditional black-box ML models (e.g., neural networks), SR presents a clearer linkage between the system's variables and, at the same time, reveals hidden dynamics.

The common SR implementation is based on GP principles, an Evolutionary Algorithm (EA) subset that approaches the optimal solution systematically in a procedure that resembles Darwin's theory of evolution. First, a pre-defined number of random expressions is created in a large set (population), where each equation (individual) can be visualized as a tree (see Figure 2). Secondly, the accuracy of each individual is estimated by a measure of fitness, such as the Mean Squared Error (MSE, Equation (11)). Then, "strong" equations that achieve small errors are tabulated and reproduced in a subsequent population by employing the GP operations of crossover and mutation (see Figure 2). Step after step, more robust expressions are extracted, and finally, several expressions are exported by balancing accuracy and complexity characteristics (Pareto frontier). The complexity term that usually accounts for the number of nodes of the tree structure can facilitate a general understanding of the given network. More details on the method can be found in [14,45].

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 , \qquad (11)$$

The application of SR in current physical science problems is gaining ground [14], opening a new route to re-evaluate traditional empirical and approximate equations. Examples include the prediction of fluid properties [11,46–48], fatigue life [49,50], modeling plastic deformation [51], and data-driven proof of physical and chemical laws [52], among others.



Figure 2. The processes of (**a**) mutation and (**b**) crossover in genetic programming, shown in a tree structure form. In mutation, new equation nodes or branches can substitute less accurate ones, while in crossover, branches can be swapped between different trees.

3. Results

Four distinct datasets that provide the self-diffusion coefficient (*D*), thermal conductivity (λ), shear viscosity η , and slip length (L_s) are considered, created by MD simulations and literature review (Tables 1 and 2). An SR framework based on PySR [53,54] and our own Python code follows to obtain a pool of candidate equations. Since finding an appropriate expression is a multi-objective optimization problem, more than one optimal solution is produced.

For the selection of final expressions, several factors have been taken into consideration. At first, since the establishment of a direct linkage between the MD structure variables in current theoretical approaches is scarce, we focused on repeating forms at the generated expressions. Moreover, as the formation of an equation via GP principles is a stochastic procedure, an intermittent re-appearance of several configurations could be an indication of capturing part of the system's dynamics. Secondly, their ability to physically describe the system and their inherent complexity (*Compl.*) was taken into account. The final criterion was their prediction accuracy, as revealed by measures such as the R-squared (R^2), mean absolute error (*MAE*), (*MSE*), and root MSE (*RMSE*) values.

After a significant number of parallel SR calculations, the dominant equations have been selected for the self-diffusion coefficient (D_c^*) , thermal conductivity (λ_c^*) , shear viscosity (η_c^*) , and slip length (L_s^*) , as follows:

$$D_c^* = \sqrt{h + 2\log(h) - 2\log(\epsilon_{wf}/\epsilon_{ff})} - \sqrt{F_{ext} h}, \qquad (12)$$

$$\lambda_c^* = \log\left(F_{ext} h + u_1 \sqrt{u_2 h + e^{\epsilon_{wf}/\epsilon_{ff}}} + u_3\right), \tag{13}$$

$$\eta_c^* = \left[e^{\epsilon_{wf}/\epsilon_{ff}(-u_1 F_{ext}+1)} e^{u_2} \right]^{\frac{1}{h}}, \qquad (14)$$

$$L_s^* = \frac{w_1(h_l/h) + w_2(m_w/m_f)^2 e^{-w_3(\epsilon_{wf}/\epsilon_{ff})^4} + w_4\sqrt{(h_d/h)}}{h^2},$$
(15)

where $u_1 = 1.6$, $u_2 = 0.39$, $u_3 = 2.47$, $v_1 = 0.732$, $v_2 = 0.882$, $w_1 = 0.13$, $w_2 = 1.3$, $w_3 = w_4 = 4$.

Although the application of the SR algorithm in small datasets [55] may lead to overfitting, the above equations achieve relatively high accuracy scores. The identity plots for the generated expressions and a zoomed figure of the dense regions can be found in Figure 3. In addition, their complexity values and overall accuracy across several random states, setting a partitioning factor of 70% train and 30% test, are shown in Table 3.



Figure 3. Identity plots for the generated expressions for (a) D_c^* , (b) λ_c^* , (c) η_c^* , and (d) L_s^* . The 45° line is a guide to the eye, denoting perfect prediction when data points lie on it.

Table 3. Generated expressions specification	ons
--	-----

Equation	R	<i>R</i> ²		MSE		RMSE		AE	Compl.
-	Train	Test	Train	Test	Train	Test	Train	Test	-
D_c^* (Equation (12))	0.94	0.91	0.18	0.17	0.42	0.41	0.28	0.27	15
λ_c^* (Equation (13))	0.92	0.90	0.01	0.01	0.10	0.10	0.07	0.07	14
η_c^* (Equation (14))	0.95	0.93	0.02	0.02	0.14	0.13	0.10	0.10	12
L_s^* (Equation (15))	0.86	0.83	0.11	0.11	0.33	0.33	0.20	0.21	19

It should be noted that D_c^* , λ_c^* , and η_c^* final expressions (Equations (12)–(14)) incorporate every input parameter of the dataset. In contrast, the slip length formula L_s^* (Equation (15)) spots only the important parameters and ends up with the wall roughness parameters h_l/h and h_d/h , wall-to-fluid interaction and mass ratio $\epsilon_{wf}/\epsilon_{ff}$ and m_w/m_f , respectively, and the channel height h. The slip length equation reports a complexity value of Compl = 19. One could opt for different equations than those presented with more or less advanced complexity; however, a trade-off between complexity and accuracy has to be retained [14]. From a statistical point of view, transport property expressions achieve fine accuracy. This is further supported by minor residual errors observed in the respective identity plots (Figure 3). On the contrary, slip length accuracy measures are smaller. A point worth mentioning is the fact that training and testing accuracy values are balanced, and this is evidence of the absence of over-fitting. In other words, the SR algorithm trained on the available data may be incorporated for predictions outside (extrapolation) and between (interpolation) the available data points, making this framework both transparent and generalizable.

4. Discussion

The inherent computational complexity of MD simulations is mainly focused on calculating the interactions between the system's particles. These interactions are typically calculated for millions of particles, for every time step (in the order of fs) considered. On the other hand, the proposed ML framework can provide an alternative to conventional research methods by approaching a possible solution at only a fraction of the initial computational demand. However, first we have to ensure that the proposed equations are closely connected to the physics of the investigating system. The generated expressions are purely data-driven and able to provide a direct estimation of the targeted quantity (i.e., the three transport properties D_c^* , λ_c^* , and η_c^* and the slip length L_s^*) as functions of the atomic-scale parameters of the confined system.

Equation (12) for D_c^* presents a strong dependence on channel height, *h*. This trend agrees with simulation results [31], where it has been shown that the self-diffusion coefficient decreases in nanochannels of small width and approaches its bulk value in wider channels, where the wall effect is minimum. It is also observed that a square root relation between D_c^* , $\epsilon_{wf}/\epsilon_{ff}$, and F_{ext} exists; however, it is difficult to decompose their effect on the self-diffusion coefficient. It seems that D_c^* is affected by $\epsilon_{wf}/\epsilon_{ff}$ in small channel heights.

On the other hand, the physical explanation is clearer for thermal conductivity, as λ_c^* in Equation (13) seems to be a logarithmic function, where h, F_{ext} , and $\epsilon_{wf}/\epsilon_{ff}$ positively affect its value. As far as shear viscosity is concerned, Equation (14) presents a complex exponential behavior, where η_c^* is affected mainly by $\epsilon_{wf}/\epsilon_{ff}$ and h. Here, as $h \to 0$, shear viscosity increases, for a given $\epsilon_{wf}/\epsilon_{ff}$. Moreover, an increase in $\epsilon_{wf}/\epsilon_{ff}$ leads to an increase in η_c^* . The external force in most of our simulation cases lies in the range $0.01 \le F_{ext} \le 0.5$ (some extreme values have been used only for testing), which means that the term $u_1 F_{ext} + 1 > 0$ only slightly decreases η_c^* , especially in narrower channels (small h).

In order to obtain a clear view of each input parameter effect on transport properties in confined channels, we provide a quantified evaluation of the SR expressions' responses in different scenarios in Figure 4. Here we have omitted the effect of the external force, as it has been found that it does not affect the outcome significantly, since its effect on flow is subtracted from the final calculations in GK approaches (see Ref. [31] for more details), and argue for the effect of $\epsilon_{wf}/\epsilon_{ff}$ and *h*.

The proposed SR-derived equation for the self-diffusion coefficient in hydrophobic and hydrophilic nanochannels (Figure 4a,b, respectively) presents a clear increasing behavior as *h* increases. In narrower channels, the effect of $\epsilon_{wf}/\epsilon_{ff}$ becomes significant. Hydrophobic nanochannel walls ($\epsilon_{wf}/\epsilon_{ff} \rightarrow 0$) lead to higher self-diffusion values compared to hydrophilic ones ($\epsilon_{wf}/\epsilon_{ff} > 1$). This illustrates the fact that, in hydrophilic cases, where particles are attracted to the surface, fluid particles find it difficult to diffuse. On the contrary, in a hydrophobic case, where atoms are repelled from the wall surface, fluid diffusion is facilitated. It is also shown that, for a nanochannel of constant height, e.g., $h = 2.64\sigma$, where the wall effect is strong due to the small height, D_c^* decreases as the walls become more hydrophilic (Figure 4c).

Another important remark is that shear viscosity from the SR-derived Equation (14) for both hydrophobic and hydrophilic channels, shown in Figure 4g,h, respectively, clearly obtains its bulk value for $h > 20\sigma$. This is in fine agreement with our previous numerical

investigation in [15]. Larger η_c^* values are obtained for hydrophilic walls, and this is also shown in Figure 4i, for the case of $h = 2.64\sigma$. Furthermore, thermal conductivity SR Equation (13) reveals a λ_c^* increase vs. $\epsilon_{wf}/\epsilon_{ff}$ and h, in Figure 4d,e. Higher λ_c^* values refer to more hydrophilic walls. It is worth noting that thermal conductivity inside narrow channels, i.e., for $h = 2.64\sigma$, increases as the walls become more hydrophilic (Figure 4f).



Figure 4. Examining the behavior of the proposed SR equations for D_c^* vs. *h* in (**a**) hydrophobic and (**b**) hydrophilic channels, and (**c**) D_c^* vs. $\epsilon_{wf}/\epsilon_{ff}$ for $h = 2.64\sigma$, for λ_c^* vs. *h* in (**d**) hydrophobic and (**e**) hydrophilic channels and (**f**) λ_c^* vs. $\epsilon_{wf}/\epsilon_{ff}$ for $h = 2.64\sigma$, and for η_c^* vs. *h* in (**g**) hydrophobic and (**h**) hydrophilic channels and (**i**) η_c^* vs. $\epsilon_{wf}/\epsilon_{ff}$ for $h = 2.64\sigma$.

In case of slip length prediction from the proposed Equation (15), we observe that L_s^* is inversely proportional to the squared channel height, h^2 . This is physically correct, since in channels of large heights, the no-slip condition applies [19]. Moreover, geometrical parameters of the walls $(h_l/h, h_d/h)$ and particle masses (m_w/m_f) seem to have a positive effect on the outcome. We attribute this to the fact that rough nanochannels have been widely incorporated to achieve larger slip lengths at the nanoscale [37,39,56]. Conversely, the energy ratio $\epsilon_{wf}/\epsilon_{ff}$ has a negative exponential effect on L_s^* . This is also an expected behavior, since as $\epsilon_{wf}/\epsilon_{ff}$ increases, the walls become more hydrophilic, imposing difficulties in fluid movement, and slip length decreases.

Slip length behavior is graphically presented in Figure 5. Each one of the independent variables has been held constant at the mean value $(\overline{h_l/h} = 0.7, \overline{h_d/h} = 0.07, \overline{m_w/m_f} = 3.0, \overline{h} = 16.96)$ and the effect of one variable is depicted in the respective sub-figure. Firstly, we observe that groove length has a minor influence on the outcome (Figure 5a), since slip length values remain constant when increasing the h_l/h ratio. On the other hand, slip length is highly affected by the depth of the groove (Figure 5b). Apparently, the SR procedure has identified that groove depth overwhelms the groove length impact, especially when wall/fluid interactions constitute a super hydrophobic environment. This can also be seen by the proposed equation weights on each of the corresponding variables in Equation (15), where the groove depth weight (w_4) is sufficiently larger than the weight of groove length (w_1).

Similarly, the slip length increases as the wall/fluid particle mass ratio m_w/m_f increases (Figure 5c). Larger wall particle masses compared to fluid particles would pose an atomic roughness wall, which, along with strong hydrophobic interactions, increases the

slip length. As far as the slip length and channel width, h, are concerned (Figure 5d), the former assumption of the inverse proportionality is evident and, ultimately, leads to the restoration of the no-slip assumption as bulk behavior is approached.

A common characteristic in every subfigure in Figure 5 is that the wall-to-fluid energy interaction ratio ($\epsilon_{wf}/\epsilon_{ff}$) has a major impact at the nanoscale, as far as the slip length is concerned. For hydrophobic cases where fluid atoms become subject to higher repulsion by the wall atoms, a large slip length value can be expected due to the reduction in the friction between fluid and wall as a result of the wall-fluid interaction area minimization [57]. Conclusively, the proposed Equation (15), derived from the data-driven SR method, seems to capture the physical interpretation of the slip mechanism in confined flows.



Figure 5. Slip length equation dynamics for various $\epsilon_{wf}/\epsilon_{ff}$ vs. (a) h_l/h , (b) h_d/h , (c) m_w/m_f , and (d) h.

5. Conclusions

Recently, data-based methodologies have progressively entered and enhanced conventional approaches by concurrently establishing computational efficiency and economy. Toward this end, an ML architecture has been trained on data derived from NEMD simulations and found broad applicability in the construction of micro- and nano-devices, i.e., the self-diffusion coefficient, thermal conductivity, shear viscosity, and the slip length. The outcome is a set of closed form equations that perform fine in several cases investigated and are able to decode hidden mechanics, even with no prior understanding of the system.

The SR-based self-diffusion coefficient equation has identified the tendency of fluid particles to stick onto hydrophilic channel walls and achieves lower values as the wall-to-fluid interaction ratio increases. In the same way, thermal conductivity and viscosity equations are increased inside hydrophilic nanochannels, especially in narrow channels. On the other hand, the suggested slip length equation applies only in channels of small height, *h*, and agrees with the macroscale hypothesis of the no-slip condition for $h > 20\sigma$.

We believe that symbolic regression techniques have matured and new computational approaches will emerge that are able to bind these data-driven methods to classical simulations at all scales. Special emphasis should be placed on the physical explanation of system dynamics, which, in this paper, has been found to agree with the established domain knowledge. The area of fluids under confinement investigation and fluid-to-surface interaction dynamics present fundamentally different behavior than the bulk, and novel, interpretable machine learning techniques could aid traditional approaches in addressing current computational obstacles.

Author Contributions: Conceptualization, F.S.; methodology, F.S., D.A. and K.P.; software, F.S., D.A. and K.P.; validation, F.S. and T.E.K.; formal analysis, F.S. and T.E.K.; investigation, F.S., D.A. and K.P.; resources, F.S. and T.E.K.; data curation, D.A.; writing—original draft preparation, D.A. and F.S.;

writing—review and editing, F.S. and T.E.K.; visualization, D.A.; supervision, F.S. and T.E.K.; funding acquisition, F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Center of Research, Innovation and Excellence of University of Thessaly, funded by the Special Account for Research Grants of University of Thessaly (No.5600.03.08.03).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data may be available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Zeni, C.; Rossi, K.; Pavloudis, T.; Kioseoglou, J.; de Gironcoli, S.; Palmer, R.E.; Baletto, F. Data-driven simulation and characterisation of gold nanoparticle melting. *Nat. Commun.* **2021**, *12*, 6056. [CrossRef] [PubMed]
- Morimoto, M.; Fukami, K.; Zhang, K.; Fukagata, K. Generalization techniques of neural networks for fluid flow estimation. Neural Comput. Appl. 2022, 34, 3647–3669. [CrossRef]
- Chowdhury, M.A.; Hossain, N.; Ahmed Shuvho, M.B.; Fotouhi, M.; Islam, M.S.; Ali, M.R.; Kashem, M.A. Recent machine learning guided material research—A review. *Comput. Condens. Matter* 2021, 29, e00597. [CrossRef]
- 4. Wei, J.; Chu, X.; Sun, X.Y.; Xu, K.; Deng, H.X.; Chen, J.; Wei, Z.; Lei, M. Machine learning in materials science. *InfoMat* 2019, 1, 338–358. [CrossRef]
- 5. Sajid, S.; Haleem, A.; Bahl, S.; Javaid, M.; Goyal, T.; Mittal, M. Data science applications for predictive maintenance and materials science in context to Industry 4.0. *Mater. Today Proc.* **2021**, *45*, 4898–4905. [CrossRef]
- 6. DeRousseau, M.; Laftchiev, E.; Kasprzyk, J.; Rajagopalan, B.; Srubar, W. A comparison of machine learning methods for predicting the compressive strength of field-placed concrete. *Constr. Build. Mater.* **2019**, *228*, 116661. [CrossRef]
- Nguyen, H.; Vu, T.; Vo, T.P.; Thai, H.T. Efficient machine learning models for prediction of concrete strengths. *Constr. Build. Mater.* 2021, 266, 120950. [CrossRef]
- Zhong, S.; Yap, B.K.; Zhong, Z.; Ying, L. Review on Y6-Based Semiconductor Materials and Their Future Development via Machine Learning. *Crystals* 2022, 12, 168. [CrossRef]
- 9. Zhang, L.; He, M.; Shao, S. Machine learning for halide perovskite materials. Nano Energy 2020, 78, 105380. [CrossRef]
- 10. Artrith, N.; Butler, K.T.; Coudert, F.X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best practices in machine learning for chemistry. *Nat. Chem.* **2021**, *13*, 505–508. [CrossRef]
- 11. Karakasidis, T.E.; Sofos, F.; Tsonos, C. The Electrical Conductivity of Ionic Liquids: Numerical and Analytical Machine Learning Approaches. *Fluids* **2022**, *7*, 321. [CrossRef]
- He, W.; Bagherzadeh, S.A.; Shahrajabian, H.; Karimipour, A.; Jadidi, H.; Bach, Q.V. Controlled elitist multi-objective genetic algorithm joined with neural network to study the effects of nano-clay percentage on cell size and polymer foams density of PVC/clay nanocomposites. J. Therm. Anal. Calorim. 2020, 139, 2801–2810. [CrossRef]
- 13. Koza, J.R. A genetic approach to econometric modeling. In Proceedings of the Sixth World Congress of the Econometric Society, Barcelona, Spain, 22–28 August 1990; Volume 27.
- 14. Angelis, D.; Sofos, F.; Karakasidis, T.E. Artificial Intelligence in Physical Sciences: Symbolic Regression Trends and Perspectives. *Arch. Comput. Methods Eng.* **2023**, *30*, 3845–3865. [CrossRef] [PubMed]
- 15. Giannakopoulos, A.; Sofos, F.; Karakasidis, T.; Liakopoulos, A. Unified description of size effects of transport properties of liquids flowing in nanochannels. *Int. J. Heat Mass Transf.* **2012**, *55*, 5087–5092. [CrossRef]
- Tomy, A.M.; Dadzie, S.K. Diffusion-Slip Boundary Conditions for Isothermal Flows in Micro- and Nano-Channels. *Micromachines* 2022, 13, 1425. [CrossRef]
- 17. Zhao, Z.; Zhou, R.; Sun, C. Molecular dynamics study of water diffusivity in graphene nanochannels. *Int. J. Thermophys.* 2020, 41, 79. [CrossRef]
- Leverant, C.J.; Greathouse, J.A.; Harvey, J.A.; Alam, T.M. Machine Learning Predictions of Simulated Self-Diffusion Coefficients for Bulk and Confined Pure Liquids. *J. Chem. Theory Comput.* 2023, 19, 11. [CrossRef]
- 19. Sofos, F.; Karakasidis, T.E.; Liakopoulos, A. Parameters affecting slip length at the nanoscale. *J. Comput. Theor. Nanosci.* 2013, 10, 648–650. [CrossRef]
- 20. Sun, C.; Lu, W.Q.; Bai, B.; Liu, J. Transport properties of Ar–Kr binary mixture in nanochannel Poiseuille flow. *Int. J. Heat Mass Transf.* 2012, *55*, 1732–1740. [CrossRef]
- 21. Nan, Y.; Li, W.; Jin, Z. Slip length of methane flow under shale reservoir conditions: Effect of pore size and pressure. *Fuel* **2020**, 259, 116237. [CrossRef]
- 22. Zhu, H.; Wang, Y.; Fan, Y.; Xu, J.; Yang, C. Structure and Transport Properties of Water and Hydrated Ions in Nano-Confined Channels. *Adv. Theory Simulations* **2019**, *2*, 1900016. [CrossRef]
- Leverant, C.J.; Harvey, J.A.; Alam, T.M.; Greathouse, J.A. Machine Learning Self-Diffusion Prediction for Lennard-Jones Fluids in Pores. J. Phys. Chem. 2021, 125, 25898–25906. [CrossRef]

- 24. Shahshahani, S.; Shahgholi, M.; Karimipour, A. The thermal performance and mechanical stability of methacrylic acid porous hydrogels in an aqueous medium at different initial temperatures and hydrogel volume fraction using the molecular dynamics simulation. *J. Mol. Liq.* **2023**, *382*, 122001. [CrossRef]
- Rabani, R.; Merabia, S.; Pishevar, A. Conductive heat transfer through nanoconfined argon gas: From continuum to free-molecular regime. *Int. J. Therm. Sci.* 2023, 192, 108391. [CrossRef]
- Sofos, F.; Karakasidis, T.; Liakopoulos, A. Variation of Transport Properties Along Nanochannels: A Study by Non-equilibrium Molecular Dynamics. In Proceedings of the IUTAM Symposium on Advances in Micro- and Nanofluidics, Dresden, Germany, 6–8 September 2007; Ellero, M., Hu, X., Fröhlich, J., Adams, N., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. 67–78.
- Pas, M.F.; Zwolinski, B.J. Computation of the transport coefficients of dense fluid neon, argon, krypton and xenon by molecular dynamics. *Mol. Phys.* 1991, 73, 471–481. [CrossRef]
- 28. Thompson, P.; Troian, S.A. A general boundary condition for liquid flow at solid surfaces. Nature 1997, 389, 360–362. [CrossRef]
- 29. Barrat, J.L.; Bocquet, L. Large slip effect at a nonwetting fluid-solid interface. *Phys. Rev. Lett.* **1999**, *82*, 4671. [CrossRef]
- 30. Sofos, F. A Water/Ion Separation Device: Theoretical and Numerical Investigation. *Appl. Sci.* 2021, *11*, 8548. [CrossRef]
- 31. Sofos, F.; Karakasidis, T.; Liakopoulos, A. Transport properties of liquid argon in krypton nanochannels: Anisotropy and non-homogeneity introduced by the solid walls. *Int. J. Heat Mass Transf.* **2009**, *52*, 735–743. [CrossRef]
- 32. Sofos, F.; Karakasidis, T.E. Nanoscale slip length prediction with machine learning tools. Sci. Rep. 2021, 11, 12520. [CrossRef]
- Karniadakis, G.E.; Beskok, A.; Gad-el Hak, M. Micro flows: Fundamentals and simulation. *Appl. Mech. Rev.* 2002, 55, B76. [CrossRef]
- 34. Todd, B.D.; Evans, D.J.; Daivis, P.J. Pressure tensor for inhomogeneous fluids. *Phys. Rev. E* 1995, *52*, 1627–1638. [CrossRef] [PubMed]
- 35. Bhadauria, R.; Aluru, N.R. A quasi-continuum hydrodynamic model for slit shaped nanochannel flow. *J. Chem. Phys.* **2013**, 139, 74109. [CrossRef] [PubMed]
- Polster, J.W.; Acar, E.T.; Aydin, F.; Zhan, C.; Pham, T.A.; Siwy, Z.S. Gating of Hydrophobic Nanopores with Large Anions. ACS Nano 2020, 14, 4306–4315. [CrossRef] [PubMed]
- 37. Priezjev, N.V. Effect of surface roughness on rate-dependent slip in simple fluids. J. Chem. Phys. 2007, 127, 144708. [CrossRef]
- 38. Eral, H.B.; van den Ende, D.; Mugele, F.; Duits, M.H.G. Influence of confinement by smooth and rough walls on particle dynamics in dense hard-sphere suspensions. *Phys. Rev. E* 2009, *80*, 61403. [CrossRef]
- 39. Sofos, F.; Karakasidis, T.E. Machine Learning Techniques for Fluid Flows at the Nanoscale. Fluids 2021, 6, 96. [CrossRef]
- Sofos, F.; Karakasidis, T.E.; Liakopoulos, A. Surface wettability effects on flow in rough wall nanochannels. *Microfluid. Nanofluid.* 2012, 12, 25–31. [CrossRef]
- Binder, K.; Horbach, J.; Kob, W.; Paul, W.; Varnik, F. Molecular dynamics simulations. J. Phys. Condens. Matter 2004, 16, S429. [CrossRef]
- 42. Hurst, J.; Wen, J. Computation of shear viscosity: A systems approach. In Proceedings of the Proceedings of the 2005, American Control Conference, Portland, OR, USA, 8–10 June 2005; Volume 3, pp. 2028–2033. [CrossRef]
- 43. Evans, D.J.; Holian, B.L. The Nose–Hoover thermostat. J. Chem. Phys. 1985, 83, 4069–4074. [CrossRef]
- 44. Holian, B.L.; Voter, A.F.; Ravelo, R. Thermostatted molecular dynamics: How to avoid the Toda demon hidden in Nosé-Hoover dynamics. *Phys. Rev. E* 1995, 52, 2338–2347. [CrossRef]
- 45. Koza, J.R. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* **1994**, *4*, 87–112. [CrossRef]
- Papastamatiou, K.; Sofos, F.; Karakasidis, T.E. Machine learning symbolic equations for diffusion with physics-based descriptions. *AIP Adv.* 2022, 12, 025004. [CrossRef]
- 47. Sofos, F.; Charakopoulos, A.; Papastamatiou, K.; Karakasidis, T.E. A combined clustering/symbolic regression framework for fluid property prediction. *Phys. Fluids* **2022**, *34*, 062004. [CrossRef]
- Alam, T.M.; Allers, J.P.; Leverant, C.J.; Harvey, J.A. Symbolic regression development of empirical equations for diffusion in Lennard-Jones fluids. J. Chem. Phys. 2022, 157, 014503. [CrossRef]
- 49. Gan, L.; Wu, H.; Zhong, Z. Integration of symbolic regression and domain knowledge for interpretable modeling of remaining fatigue life under multistep loading. *Int. J. Fatigue* **2022**, *161*, 106889. [CrossRef]
- 50. Ren, J.; Zhang, L.; Zhao, H.; Zhao, Z.; Wang, S. Determination of the fatigue equation for the cement-stabilized cold recycled mixtures with road construction waste materials based on data-driven. *Int. J. Fatigue* **2022**, *158*, 106765. [CrossRef]
- 51. Kabliman, E.; Kolody, A.H.; Kronsteiner, J.; Kommenda, M.; Kronberger, G. Application of symbolic regression for constitutive modeling of plastic deformation. *Appl. Eng. Sci.* 2021, *6*, 100052. [CrossRef]
- 52. Neumann, P.; Cao, L.; Russo, D.; Vassiliadis, V.S.; Lapkin, A.A. A new formulation for symbolic regression to identify physicochemical laws from experimental data. *Chem. Eng. J.* 2020, *387*, 123412. [CrossRef]
- Cranmer, M. PySR: Fast & Parallelized Symbolic Regression in Python/Julia. 2020. Available online: https://zenodo.org/record/ 4041459 (accessed on 1 May 2023).
- 54. Cranmer, M. Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. arXiv 2023, arXiv:2305.01582.
- 55. Wilstrup, C.; Kasak, J. Symbolic regression outperforms other models for small data sets. *arXiv* **2021**, arXiv:2103.15147.

- 56. Galea, T.M.; Attard, P. Molecular Dynamics Study of the Effect of Atomic Roughness on the Slip Length at the Fluid-Solid Boundary during Shear Flow. *Langmuir* 2004, 20, 3477–3482. [CrossRef] [PubMed]
- 57. Bocquet, L.; Barrat, J.L. Flow boundary conditions from nano- to micro-scales. Soft Matter 2007, 3, 685–693. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.