
MPF-Bench : A Large Scale Dataset for SciML of Multi-Phase-Flows: Droplet and Bubble Dynamics

Mehdi Shadkhah¹, Ronak Tali^{1,*}, Ali Rabeh^{1,*}, Cheng-Hau Yang^{1*}, Abhishek Upadhyaya², Adarsh Krishnamurthy¹, Chinmay Hegde², Aditya Balu¹, Baskar Ganapathysubramanian¹

¹Iowa State University
{mehdish,rtali,arabeh,chenghau,
adarsh,baditya,baskarg}@iastate.edu

²New York University
{au2216,chinmay.h}@nyu.edu

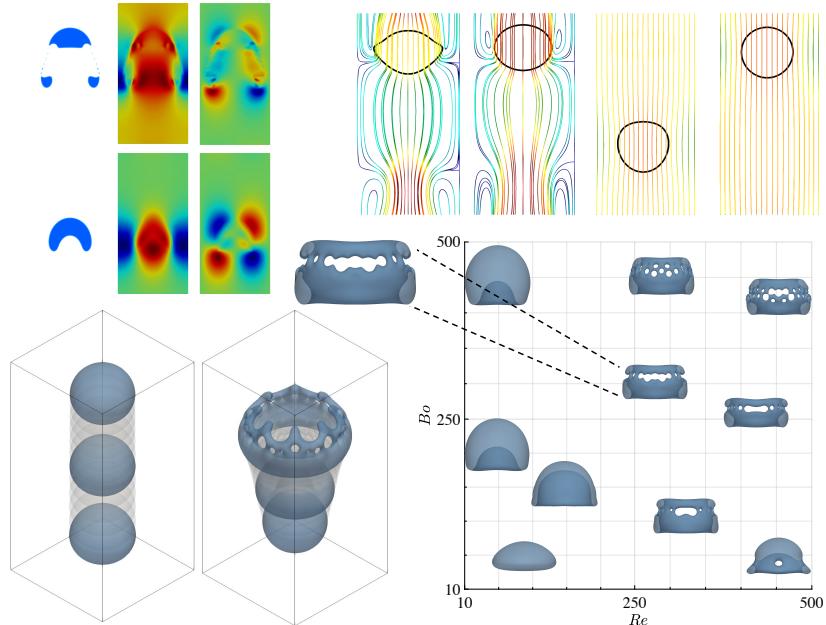


Figure 1: *MPF-Bench* includes 11000 two-phase flow simulations. The top left panel displays velocity results in both x and y directions and bubble shapes for two randomly selected 2D bubble rising simulations. The top right panel presents 2D droplet falling simulations, depicting streamlines colored by velocity magnitude and droplet shapes. The bottom left panel illustrates 3D simulation outputs, highlighting time series outputs from our dataset. The bottom right panel shows 3D bubble shapes for different Reynolds and Bond number combinations.

Abstract

The dynamics of falling droplets and rising bubbles involve complex multiphase fluid physics crucial for many industrial applications. However, traditional multiphase solvers are computationally expensive due to the need to resolve complex phenomena like instabilities, wave patterns, and bubble breakup. Scientific machine learning (SciML) is a promising strategy for designing multiphase flow solvers, but it requires a large dataset. This paper presents a comprehensive dataset created from 11,000 simulations, both in 2D and 3D, using the Lattice Boltzmann method (LBM). The dataset ([here](#)) captures intricate physics by varying factors such as surface tension, density, and viscosity of fluids. These simulations, comprising 1 million time

*These authors contributed equally to this work.

snapshots, provide extensive data on two-fluid behavior. By making this dataset publicly available, we aim to encourage SciML research and its applications in complex fluid systems, facilitating the creation of more precise and efficient SciML frameworks for multiphysics applications. Our dataset spans multiple orders of magnitude of Reynolds and Bond numbers, density, and viscosity ratios, making it exceptionally rich and valuable for understanding multiphase flows.

1 Introduction

Flow behavior in multiphase flow is crucial for many industrial and chemical applications. In drug delivery, two-phase flow can be used to create uniform drug-loaded microspheres or microcapsules. These microcapsules can provide controlled and sustained release of drugs, improving therapeutic outcomes [1, 2]. Two-phase flows are also essential for rapid diagnostics and biochemical applications in lab-on-a-chip technologies [3, 4]. Discrete phase bubbles in microchannels, generated via T-junctions [5], co-flowing systems [6], or flow-focusing techniques [7], have a high surface-to-volume ratio, enhancing reaction efficiency and sensitivity. The shearing forces of the continuous phase precisely control bubble size and formation, which is crucial for device performance. By thoroughly understanding gas-liquid or liquid-liquid interactions, engineers can optimize mixing conditions [8–10] to enhance reaction rates, improve product consistency, and reduce energy consumption.

Bubbles (*light fluid volumes moving in a denser fluid medium*) and droplets (*heavier fluid volumes moving in a lighter fluid medium*) play an integral role in applications such as drug delivery and lab-on-a-chip technologies. The dynamics of droplets and bubbles exhibit significant complexity, primarily due to phenomena such as breakup, deformation, and surface tension. Firstly, the breakup of droplets and bubbles is a highly nonlinear and complex process governed by factors such as viscosity ratio, density ratio, and surface tension. For example, for high inertia flows, the fast and irregular breakup results in smaller and widely-distributed droplets; at low Reynolds numbers, laminar flow leads to a more even breakup and larger droplets [11]. Secondly, droplets can be deformed by shear and pressure forces. Various studies have shown that the Capillary number [12, 13], Atwood number [14, 15], and Reynolds number [16–18] all have a significant impact on droplet deformation.

To better understand multiphase phenomena (both droplets and bubbles), researchers often perform a canonical simulation/experiment called the bubble rising case [19–25], where a bubble is placed in a higher density fluid so that the bubble moves up due to buoyancy. Conversely, using a droplet of higher density causes the droplet to fall down due to gravity [26, 27]. This canonical study is essential since it provides insights into bubble dynamics and shape evolution, which are critical factors for optimizing industrial processes and improving numerical models in fluid dynamics research. Nonetheless, capturing the bubble-rising or droplet-falling phenomenon is a multiscale problem with forces acting at different scales, ranging from microscale molecular interactions to macroscale fluid dynamics. Therefore, high-fidelity simulations are essential to accurately resolve these interactions, particularly at the thin interfaces where precise capturing of surface tension and interfacial dynamics is critical.

Scientific Machine Learning (SciML) represents a powerful approach for addressing complex multiphase flow problems. SciML leverages the inherent physics within the system to develop models that can learn from limited data while still producing reliable predictions. By integrating physics-based knowledge directly into the loss function, SciML can be an attractive approach to predict multiphase flow dynamics [28–30]. A key ingredient to training and accessing SciML multiphase solvers is a comprehensive dataset and evaluation metrics. [MPF-Bench](#) is such a benchmark dataset. This dataset captures the intricate physics of multi-phase flows by varying factors such as the Bond number - a dimensionless number that indicates the strength of the gravitational field versus surface tension forces, and the Reynolds number, which indicates the ratio of inertial to viscous forces. It includes wave patterns, bubble and droplet dynamics, and breakup. [MPF-Bench](#) has three major features:

- **Extensive amount of data:** Our dataset includes 11000 simulations in 2D and 3D with over 1 million time-series snapshots. This extensive volume of data allows for robust training of SciML models, which will help in advancing the development of accurate and reliable SciML models for multiphase flow dynamics.
- **Multiphysics simulations:** We conduct simulations of rising bubbles and falling droplets, solving the Navier-Stokes equations coupled with the Allen-Cahn equation. This approach captures considerable physical phenomena, including breakup and deformation.

- **Wide range of properties:** The dataset varies key parameters such as density ratio, viscosity ratio, Reynolds number, and Bond number. This diversity ensures the models trained on this data can potentially generalize across diverse flow conditions and interface dynamics.

Our Contributions: We summarize our main contributions below:

- Our dataset features 11,000 simulations and over 1 million time-series snapshots, with variations in density ratio, viscosity ratio, Reynolds number, and Bond number. This extensive dataset encompasses many phenomena, ranging from subtle surface deformations in bubble oscillations to full bubble breakups driven by surface tension and density ratio variations. The richness and breadth of this dataset offer deep insights into the intricate dynamics of multiphase flows, making it a valuable resource for advancing research in this field. We provide our [dataset](#) as a benchmark for others interested in developing and evaluating SciML models. Additional details can also be found in our [website](#).
- Two neural operators were trained on our data i.e., predicting constitutive fields at a certain time step using the corresponding field values at previous time steps. To our knowledge, no study has evaluated the performance of neural operators on multiphase flow datasets.

2 Related Work

The Stanford Multiphase Flow Database (SMFD) used in [31], the flow experiment dataset [32], and the BubbleML dataset [29] are resources for understanding multiphase flow dynamics.

The SMFD features 5659 measurements across a range of gas and liquid properties, pipe characteristics, and operational conditions. This dataset, derived from laboratory and field sources, supports various pipe inclinations and flow patterns. SMFD covers different flow regimes, including stratified, slug, and annular flows. However, it does not appear publicly available, so we cannot identify the number of individual snapshots in this dataset.

The flow experiment dataset [32] focuses on the effects of density, viscosity, and surface tension on two-phase flow regimes and pressure drops in horizontal pipes. The 2904 measurements from air-liquid system experiments provide insights into fluid properties' influence on flow regimes and pressure drops. This dataset's development of flow regimes and pressure contour maps enhances the understanding of fluid behavior in horizontal two-phase flows.

Additionally, the BubbleML dataset [29] is a data collection focused on multiphysics phase change phenomena generated through physics-driven simulations, providing ground truth information for various boiling scenarios, including nucleate pool boiling, flow boiling, and sub-cooled boiling. We summarize these and other databases alongside our dataset in [Table 1](#).

Table 1: Comparison of public Multiphase Flow Datasets

Name	Samples	Snapshots	Scope	Sources	Ranges of material properties
Flow Experiment Dataset	2904	2904	Horizontal pipes, effects of density, viscosity, surface tension	Controlled lab environment	ρ : [1, 1.5] gm/cc, μ : [1, 3.1] cP, Surface tension = [32, 70] mN/m
BubbleML	79	7641	pool boiling, flow boiling, and sub-cooled boiling	2D and 3D Numerical simulations based on Flash-X	$Re = 0.0042$, $\rho^* = 0.0083$, $\mu^* = 1$, $Pr = 8.4$, $We = 1$, $Fr = [1, 100]$
MPF-Bench	11000	> 1 million	Droplet and bubble dynamics	2D and 3D Simulations using LBM	$\rho^* : [10, 1000]$, $\mu^* : [1, 100]$, $Bo : [10, 500]$, $Re : [10, 1000]$

3 Multi-phase flow (MPF) Bench

We present the **MPF-Bench** dataset, encompassing 5500 bubble rise and 5500 droplet flow simulations, with each simulation containing 100 time-snapshots, making it, to our knowledge, two orders of magnitude larger – in terms of number of time-snapshots – than any existing multiphase flow dataset. This dataset features 2D and 3D transient simulations, capturing a spectrum of flow behaviors influenced by surface tension and density/viscosity ratios. **MPF-Bench** includes scenarios from bubble oscillations with minor surface deformations to complete bubble breakup, offering a comprehensive resource for studying bubble rise and droplet fall dynamics.

3.1 Problem Definition: Initial and Boundary Conditions, and outputs

We consider 2D and 3D simulations of bubble rise and droplet fall simulations using the lattice Boltzmann method. The domain sizes for 2D and 3D are [256, 512] and [128, 256, 128] lattice units, respectively. For 2D simulations, the bubble is initially centered at (64, 64) and the droplet is centered at (128, 384). In 3D, the bubble is centered at (64, 64, 64) while the droplet is centered at (128, 384, 64). The initial diameter D_0 for both problems is set to 128 lattice units in 2D and 64 lattice units in 3D. The boundary conditions are set to free-slip on the side walls and periodic at the top and bottom as illustrated in [Figure 2](#). This problem is driven mainly by the density and viscosity ratio of the two phases in addition to the Reynolds and Bond numbers. The Reynolds number measures the ratio of inertial forces to viscous forces, while the Bond number measures the ratio of gravitational forces to surface tension forces. Below is the definition of these four dimensionless numbers:

$$\rho^* = \frac{\rho_h}{\rho_l}, \quad \mu^* = \frac{\mu_h}{\mu_l}, \quad \text{Re}_h = \frac{\sqrt{g_y \rho_h (\rho_h - \rho_l)} D^3}{\mu_h}, \quad \text{Bo} = \frac{g_y (\rho_h - \rho_l) D^2}{\sigma} \quad (1)$$

where h and l indices refer to the heavy and light fluids, respectively. We have selected random, dimensionless numbers uniformly to ensure the entire defined range is covered. The outputs of the simulations are the interface indicator (c), velocity components (u, v, w), pressure (p), and density (ρ), which provide insights into the dynamics of multiphase flow and the interactions between the phases.

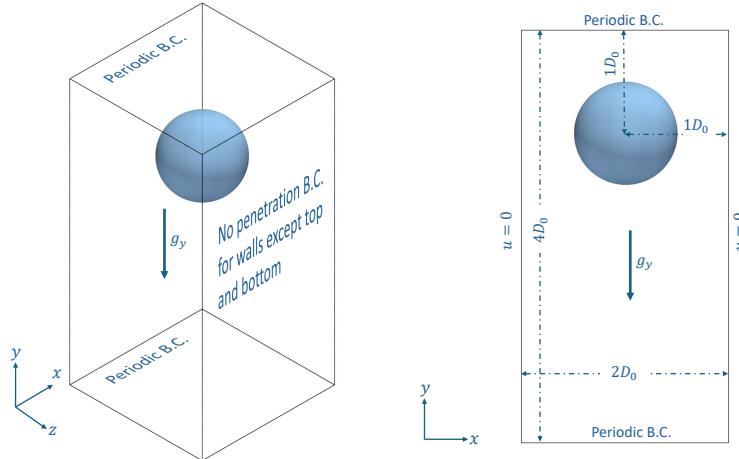


Figure 2: Boundary conditions for the simulation of a falling droplet. The left panel illustrates the 3D case, while the right panel illustrates the 2D case.

3.2 Simulation Framework and Compute Effort

Our simulation framework employs a highly parallel, in-house Lattice Boltzmann code, utilizing one of the most accurate two-phase models, the phase field model, to capture the complexities of the interface. The code has been rigorously tested across various problems, with validation results provided in [Section A.3](#). For 2D simulations, we used a uniform lattice grid of 256×512 , while for 3D simulations, the domain was set to $128 \times 256 \times 128$. We achieved high parallelization by distributing the computation across 12 Nvidia A100-SXM4 80GB GPUs. The total computational

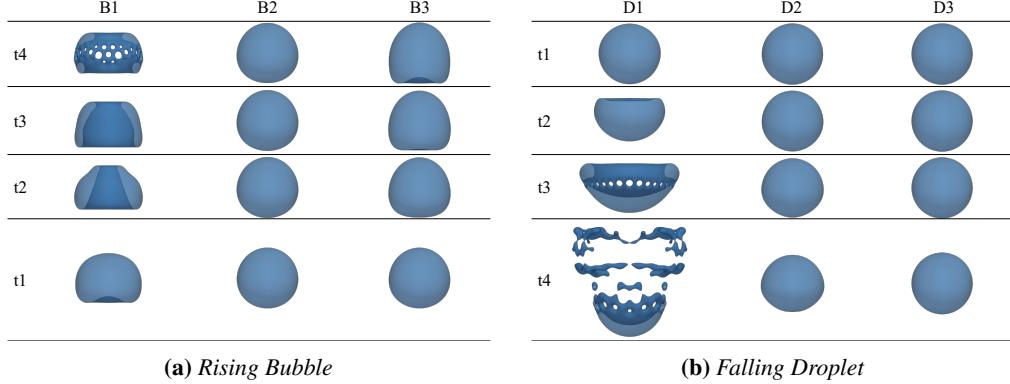


Figure 3: (a) Snapshot of a 3D rising bubble and (b) snapshot of a 3D falling droplet. The properties of the fluids for each case are detailed in [Table 3](#).

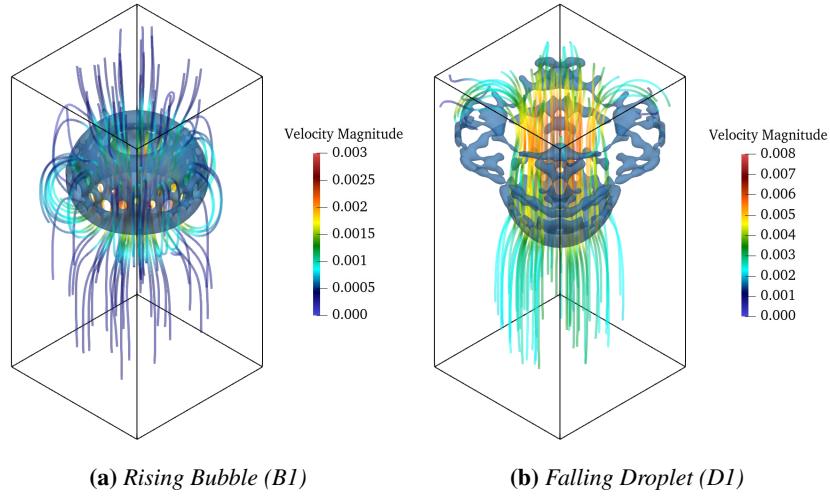


Figure 4: Streamlines of a 3D rising bubble (a) and a 3D falling droplet (b), with colors indicating the magnitude of velocity. The properties of the fluids for each case are detailed in [Table 3](#).

cost for 2D and 3D cases was approximately 4,000 GPU hours. We use the ParaView tool [33] to visualize and understand our dataset.

3.3 MetaData

Input Fields: We have provisioned the following dimensionless quantities as inputs to our study as defined in [Section 3.1](#). These are the Density Ratio (ρ^*), Viscosity Ratio (μ^*), Bond Number (Bo), and Reynolds Number (Re). Since these are scalar values, we feed them to the neural network by creating a constant field with a dimension consistent with the number of samples, in this case, 10,000 in 2D and 1000 in 3D.

Output Fields: In analyzing multiphase flow problems, we are interested in solving the governing PDEs to obtain solutions at every point in the domain's interior for certain cardinal fields. For a 2D solution domain, these are: c - interface indicator, u - velocity in x direction, v - velocity in y direction, p - pressure. Additionally, because this is a time-dependent problem, we have these cardinal fields or a sequence of these fields distributed uniformly over time (100 time steps).

Resolution: We maintained the original resolution of our datasets, matching the Lattice Boltzmann simulation domain. This ensures the complete physics is presented to the Neural Operator and allows direct comparison with Lattice Boltzmann method simulations. Our datasets are published at 256×512 resolution for 2D and $128 \times 256 \times 128$ for 3D simulations.

Dataset Format: For both the bubble and droplet datasets, we have released a single file for each sample. This decision was taken with the view to allow for maximum flexibility to the end user in

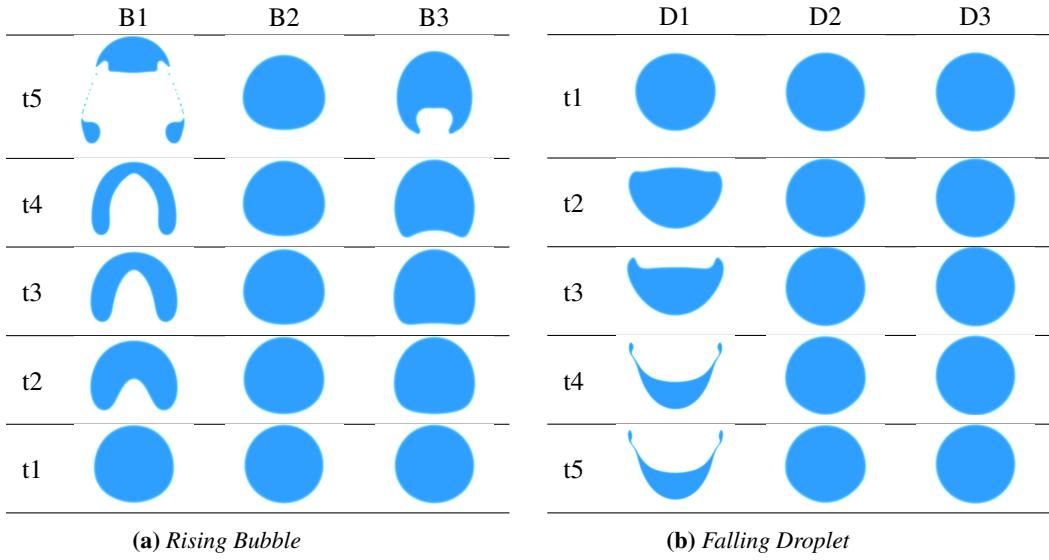


Figure 5: (a) Snapshot of a 2D rising bubble and (b) snapshot of a 2D falling droplet. The properties of the fluids for each case are detailed in [Table 3](#).

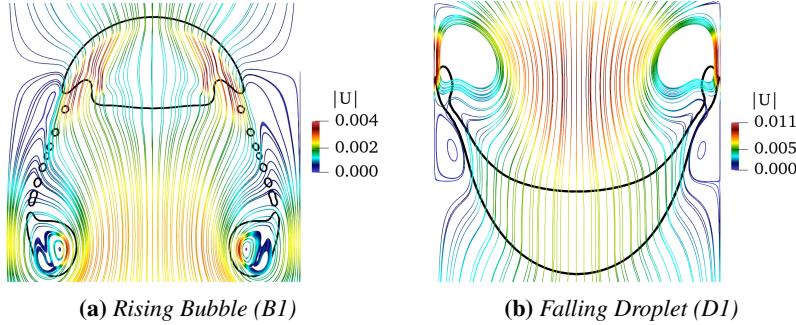


Figure 6: Streamlines of a 2D rising bubble (a) and a 2D falling droplet (b), with colors indicating the magnitude of velocity.

Table 2: Formulaic description of the input and output tensors. 5000 - sample size for the dataset. 101 - number of time steps in the simulation. x, y - The x, y dimension of a field. E.g., $Y[0, 100, 1, :, :]$ indicates the pointwise v velocity over the entire grid of size 256×512 for the first sample at time step 100.

Dataset	Dim.	Input Tensor	Output Tensor
Droplet	2	$X[5000][\rho^*, \mu^*, Bo, Re]$	$Y[5000][101][c, u, v, p, \rho][y][x]$
Bubble	2	$X[5000][\rho^*, \mu^*, Bo, Re]$	$Y[5000][101][c, u, v, p, \rho][y][x]$
Droplet	3	$X[500][\rho^*, \mu^*, Bo, Re]$	$Y[500][51][c, u, v, w, p, \rho][z][y][x]$
Bubble	3	$X[500][\rho^*, \mu^*, Bo, Re]$	$Y[500][51][c, u, v, w, p, \rho][z][y][x]$

deciding what and how many time steps they want to use to train their models, as these time-dependent problems often take the shape of sequence to sequence formulations. In 2D, the resulting .npz files take the form:

`[number_of_time_steps][number_of_channels][resolution_y][resolution_x]`

whereas in 3D, `[resolution_z]` incorporated as an additional dimension for depth. In this study, we have released a total of 11,000 samples spread across two families of datasets. [Table 2](#) provides a detailed formulaic description of the packaging of the input and output numpy tensors for both these families:

Table 3: Material properties and nondimensional numbers of three bubble rise simulations (B1, B2, B3) and three droplet fall simulations (D1, D2, D3). The table shows the density ratio, viscosity ratio, Reynolds number, and Bond number of all six simulations.

case	B1	B2	B3	D1	D2	D3
Density Ratio (ρ^*)	10^3	10^3	10^3	10	10^3	10^3
Viscosity Ratio (μ^*)	10^2	10^2	10^2	1	10^2	10^2
Re	5×10^2	10	10	10^3	10^3	10
Bo	5×10^2	10	5×10^2	5×10^2	5×10^2	10

3.4 Evaluation Metrics and Test Dataset Analysis

Evaluation metrics: We evaluate the performance of the neural operators trained on the dataset on two sets of metrics ($M1, M2$):

- $M1$: *Shape metrics*: The Mean Squared Error (MSE) of the indicator function, c , over the entire domain is a good primary metric to measure the accuracy of the predicted shape. This metric is useful to evaluate how accurately the neural operators can capture topological changes while ensuring that total mass is conserved.
- $M2$: *Flow metrics*: The Mean Squared Error (MSE) over the entire domain is a good metric to measure the accuracy of predicted velocity fields, reflecting how closely the model’s predictions match the true values.

Test Datasets: We recommend evaluating trained models on two test scenarios

- *Standard*: This is the standard 80-20 random split of the data.
- *Extreme cases*: We report average performance across three cases for bubbles(B1, B2, B3) and three cases for droplets (D1, D2, and D3). Each case is characterized by the Bond number, Reynolds number, and density ratio of the bubble to the surrounding fluid as illustrated in [Table 3](#). These parameters significantly influence the bubble’s velocity, surface deformation, or possible breakup dynamics. Our simulations illustrate each case’s complex and nonlinear bubble dynamics, as seen in the time snapshots in [Figure 3](#). Case B1, characterized by high Bond and Reynolds numbers, exhibits significant deformation and eventual rupture due to the high inertial forces and weak surface tension, meaning that inertia can easily rupture the droplet surface. Also, due to the high-density ratio, we experience flattening and increased circulation at the bubble’s rear, as illustrated by [Figure 4a](#) since the surrounding fluid cannot push back against the bubble. In contrast, case B3, with a lower Reynolds number, shows crown-shaped deformation without rupture since it experiences less inertial forces to deform its surface. Regarding case B2, the lower Bond number indicates a higher surface tension, causing the bubble to retain its circular shape especially since inertial is low. Here, the dominance of viscous forces stabilizes the bubble, preventing drastic shape changes or breakups. This stability is consistent with forming ellipsoidal caps with closed and laminar wakes at low Reynolds numbers, reflecting minimal surface oscillations [34]. These three cases demonstrate how varying Reynolds and Bond numbers influence bubble dynamics and breakup in complex fluid dynamics. [Figure 3b](#) shows the dynamics of three droplet falling cases. A distinct drop-falling behavior for case D1 is characterized by a low-density ratio and high Reynolds and Bond numbers. In this case, the droplet breaks apart and dissociates into smaller droplets. This phenomenon is called the satellite effect due to capillary instability [35, 36]. The interplay between strong inertial and buoyancy forces and weak surface tension induces notable deformation and eventual rupture of the droplet, causing the satellite effect like that observed in bubble rise dynamics. Videos of the bubble rise and drop fall flows are provided in the [MPF-Bench website](#).

4 Experiments

Neural Operators are a new class of deep learning frameworks attuned to learning functional solutions to PDEs. A distinguishing characteristic of Neural Operators is that instead of learning a specific solution to a PDE given a fixed set of parameters, they are able to learn general solutions to the PDE applicable over a complete range of PDE parameters. While these operators have shown considerable success in terms of learning solutions to single-phase flows, there is a lack of evidence regarding their ability to model complex fluid flows such as two-phase flows. The physics of multi-phase flows are significantly more complex because of factors such as surface tension, varying density, and viscosity in the flow domain. As outlined in previous sections, these give rise to chaotic phenomena such as

the break up of a bubble or droplet. Against this backdrop, we wanted to investigate how well we can capture these phenomena using two of the recently proposed neural operators.

Fourier Neural Operator (FNO) [37] is a recently proposed neural operator design that acts on the data on a Fourier basis, thus allowing the inputs and outputs of a PDE to be represented in a continuous space. While this is sufficient to approximate linear operator solutions to a PDE, to approximate non-linear operator solutions, FNO applies an additional layer of non-linearity. Theoretically, this mechanism can approximate a general operator solution to any arbitrary PDE. While FNOs demonstrate significant advancement in the field of SciML, the issue of discretization error has been raised [38] and potentially breaking the principle of continuous-discrete equivalence[39].

Convolutional Neural Operator (CNO) [39] is another recently proposed neural operator that generalizes the very popular convolutional neural network (CNN) architecture to estimate the solution operators to PDEs. CNOs have desirable characteristics, such as structure preservation through continuous-discrete equivalence, i.e., map bandlimited functions to bandlimited functions, overcoming issues such as aliasing errors [39].

Our SciML problem formulation takes the following two (sequence of tensors to a single tensor) alternate structural forms in tensor notation:

$$C1 : Y[:, 5, :, :, :] \sim \underbrace{Y[:, (1 \cdots 4), :, :, :]}_{Seq. \ of \ initial \ solutions}$$

Note: All the neural operators learn a space-time smooth solution to the PDE. This is in contrast to traditional numerical techniques where the time element is merely serves as a way to step forward the space solution in time. The other structure that we modeled in this study is:

$$C2 : Y[:, 9, :, :, :] \sim Y[:, (1 \cdots 8), :, :, :]$$

It is known that physics evolves relatively slowly during the initial time steps. Moreover, phenomena such as bubble breaking are absent during this period. Such relatively benign conditions allow us to create reproducible benchmark studies. We hope the community will take these benchmark results as a starting point and go on to fit models at more advanced time steps where the physics becomes very challenging to replicate through the means of a ML framework.

The results of fitting FNO and CNO (purely on data and not explicitly embedding the flow physics) to our chosen datasets are summarized below. We used the MSE metric to train both operators. All reported MSEs were calculated on a validation sample held out at 20% of the full dataset for a given problem. In addition we used the same python packages/code base released by the creators of FNO and CNO.

We conducted all training for these experiments on a NVIDIA A100 GPU - 80GB. We trained uniformly at a batch size of 5,200 epochs, and a learning rate of 1e-3. We noticed that CNO and FNO converged satisfactorily to validation losses of the order 1e-6. Performance on both benchmarks shows good convergence for both the bubble and droplet datasets.

Table 4: Performance of CNO and FNO neural operators on the bubble rise and droplet datasets. All Mean Squared Errors (MSE) are reported on a validation dataset. M1 is the mean squared error of the interfacial indicator, i.e., c , and M2 is the mean squared error of the velocity fields, i.e., u, v .

Dataset	Metric	FNO	CNO
C1 - bubble	M1 (c)	8.647×10^{-6}	1.651×10^{-6}
	M2 (u, v)	$(7.586 \times 10^{-6}, 3.874 \times 10^{-6})$	$(3.672 \times 10^{-6}, 1.164 \times 10^{-6})$
C2 - bubble	M1 (c)	4.253×10^{-6}	2.144×10^{-7}
	M2 (u, v)	$(7.256 \times 10^{-6}, 8.647 \times 10^{-7})$	$(3.505 \times 10^{-6}, 8.068 \times 10^{-8})$
C1 - droplet	M1 (c)	1.609×10^{-6}	8.690×10^{-7}
	M2 (u, v)	$2.896 \times 10^{-6}, 8.642 \times 10^{-7}$	$4.102 \times 10^{-6}, 3.050 \times 10^{-7}$
C2 - droplet	M1 (c)	8.133×10^{-6}	5.216×10^{-7}
	M2 (u, v)	$2.132 \times 10^{-6}, 5.831 \times 10^{-7}$	$2.016 \times 10^{-6}, 4.030 \times 10^{-7}$

Table 5: Performance of CNO and FNO neural operators on bubble rise and droplet extreme cases. All Mean Squared Errors (*MSE*) are reported on a validation dataset. *M1* is the mean squared error of the interfacial indicator i.e., c , and *M2* is the mean squared error of the velocity fields, i.e., u, v .

Case	Metric	FNO	CNO
B1	M1 (c)	4.628×10^{-5}	3.383×10^{-6}
	M2 (u, v)	$6.855 \times 10^{-5}, 8.351 \times 10^{-5}$	$8.162 \times 10^{-6}, 4.861 \times 10^{-6}$
B2	M1 (c)	3.625×10^{-5}	4.391×10^{-7}
	M2 (u, v)	$8.651 \times 10^{-5}, 3.931 \times 10^{-5}$	$8.781 \times 10^{-6}, 4.013 \times 10^{-7}$
B3	M1 (c)	4.581×10^{-5}	4.341×10^{-7}
	M2 (u, v)	$3.836 \times 10^{-5}, 7.616 \times 10^{-5}$	$3.814 \times 10^{-6}, 6.711 \times 10^{-8}$
D1	M1 (c)	4.310×10^{-5}	6.911×10^{-6}
	M2 (u, v)	$8.322 \times 10^{-6}, 8.013 \times 10^{-5}$	$2.955 \times 10^{-6}, 3.741 \times 10^{-6}$
D2	M1 (c)	3.318×10^{-6}	2.841×10^{-6}
	M2 (u, v)	$7.132 \times 10^{-6}, 2.418 \times 10^{-5}$	$8.120 \times 10^{-6}, 4.831 \times 10^{-6}$
D3	M1 (c)	2.183×10^{-6}	9.081×10^{-7}
	M2 (u, v)	$2.172 \times 10^{-5}, 1.839 \times 10^{-5}$	$8.044 \times 10^{-6}, 2.427 \times 10^{-6}$

An analysis of [Table 4](#) reveal that given the stability of the physics modeled by C1 and C2, we are obtaining a rapid convergence to the ground truth for both CNO and FNO. We observe that CNOs are marginally better. The key differentiator, however, is the small number of tuning parameters in the case of CNO that makes the process of training models much easier.

Referring to [Table 5](#), we observe that the Neural Operators produce the least error in predicting the behavior of bubble B2 and the most for B1 because of low deformation of B2 compared to B1 and captured in [Figure 5a](#). In a similar vein, the ordering of MSE errors for droplets is in ascending order D3, D2, and D1, as evident in [Table 5](#). This is inline with our expectations as captured in [Figure 5b](#), i.e., D1 undergoes the most deformation and D3 the least, therefore, it is easier for the model to predict the behavior of D3 as opposed to D1.

5 Conclusions

In conclusion, we present an extensive time series dataset encompassing 10,000 simulations in 2D and 1,000 simulations in 3D, focused on bubble rise and droplet fall dynamics. Our dataset is rich in two-phase flow dynamics, including simulations with density ratios up to 1000, Reynolds numbers up to 1000, and Bond numbers up to 500.

Limitations: Our dataset has the following limitations:

- **Limited 3D Simulations:** Due to the high computational cost, we could only conduct a small number of 3D simulations. Consequently, our dataset includes a more limited sample of 3D cases.
- **Model fitting with a large number of time steps:** Owing to GPU constraints, we have limited the number of time steps we could fit in your CNO model. Therefore, we had to resort to staggered time steps to fit the model tensors within a single GPU. Since we are using the code made available directly by the creators of CNO, we were limited to using a single GPU.

References

- [1] Sophie Hernot and Alexander L Klibanov. Microbubbles in ultrasound-triggered drug and gene delivery. *Advanced drug delivery reviews*, 60(10):1153–1166, 2008.
- [2] Amirmohammad Sattari, Pedram Hanafizadeh, and Mina Hoorfar. Multiphase flow in microfluidics: From droplets and bubbles to the encapsulated structures. *Advances in Colloid and Interface Science*, 282:102208, 2020.
- [3] Stefan Haeberle and Roland Zengerle. Microfluidic platforms for lab-on-a-chip applications. *Lab on a Chip*, 7(9):1094–1110, 2007.
- [4] Daniel Mark, Stefan Haeberle, Günter Roth, Felix Von Stetten, and Roland Zengerle. Microfluidic lab-on-a-chip platforms: requirements, characteristics and applications. *Microfluidics based microsystems: fundamentals and applications*, pages 305–376, 2010.
- [5] Todd Thorsen, Richard W Roberts, Frances H Arnold, and Stephen R Quake. Dynamic pattern formation in a vesicle-generating microfluidic device. *Physical review letters*, 86(18):4163, 2001.
- [6] Carsten Cramer, Peter Fischer, and Erich J Windhab. Drop formation in a co-flowing ambient fluid. *Chemical Engineering Science*, 59(15):3045–3058, 2004.
- [7] Shelley L Anna, Nathalie Bontoux, and Howard A Stone. Formation of dispersions using “flow focusing” in microchannels. *Applied physics letters*, 82(3):364–366, 2003.
- [8] Norbert Schwesinger, Thomas Frank, and Helmut Wurmus. A modular microfluid system with an integrated micromixer. *Journal of Micromechanics and Microengineering*, 6(1):99, 1996.
- [9] Abraham D Stroock, Stephan KW Dertinger, Armand Ajdari, Igor Mezic, Howard A Stone, and George M Whitesides. Chaotic mixer for microchannels. *Science*, 295(5555):647–651, 2002.
- [10] Joshua D Tice, Helen Song, Adam D Lyon, and Rustem F Ismagilov. Formation of droplets and mixing in multiphase microfluidics at low values of the reynolds and the capillary numbers. *Langmuir*, 19(22):9127–9133, 2003.
- [11] Jens Eggers and Emmanuel Villermaux. Physics of liquid jets. *Reports on progress in physics*, 71(3):036601, 2008.
- [12] Anja Vananroye, Pieter JA Janssen, Patrick D Anderson, Peter Van Puyvelde, and Paula Moldenaers. Microconfined equiviscous droplet deformation: Comparison of experimental and numerical results. *Physics of Fluids*, 20(1), 2008.
- [13] Zhaomiao Liu, Fanming Cai, Yan Pang, Yanlin Ren, Nan Zheng, Rui Chen, and Siyu Zhao. Enhanced droplet formation in a t-junction microchannel using electric field: A lattice boltzmann study. *Physics of Fluids*, 34(8), 2022.
- [14] Abbas Fakhari and Mohammad Hassan Rahimian. Investigation of deformation and breakup of a moving droplet by the method of lattice boltzmann equations. *International journal for numerical methods in fluids*, 64(8):827–849, 2010.
- [15] Satyvir Singh. Role of atwood number on flow morphology of a planar shock-accelerated square bubble: A numerical study. *Physics of Fluids*, 32(12), 2020.
- [16] Konstantinos Vontas, Cristina Boscariol, Manolia Andredaki, Anastasios Georgoulas, Cyril Crua, Jens Honoré Walther, and Marco Marengo. Droplet impact on suspended metallic meshes: Effects of wettability, reynolds and weber numbers. *Fluids*, 5(2):81, 2020.
- [17] Zhikun Xu, Tianyou Wang, and Zhizhao Che. Droplet deformation and breakup in shear flow of air. *Physics of Fluids*, 32(5), 2020.
- [18] Drue Seksinsky and Jeffrey S Marshall. Droplet impingement on a surface at low reynolds numbers. *Journal of Fluids Engineering*, 143(2):021304, 2021.
- [19] D. Bhaga and M. E. Weber. Bubbles in viscous liquids: shapes, wakes and velocities. *Journal of Fluid Mechanics*, 105:61–85, 1981. doi: 10.1017/S002211208100311X.
- [20] Jinsong Hua and Jing Lou. Numerical simulation of bubble rising in viscous liquid. *Journal of Computational Physics*, 222(2):769–795, 2007. ISSN 0021-9991. doi: <https://doi.org/10.1016/j.jcp.2006.08.008>. URL <https://www.sciencedirect.com/science/article/pii/S0021999106003949>.

- [21] Shu-Ren Hysing, Stefan Turek, Dmitri Kuzmin, Nicola Parolini, Erik Burman, Sashikumaar Ganesan, and Lutz Tobiska. Quantitative benchmark computations of two-dimensional bubble dynamics. *International Journal for Numerical Methods in Fluids*, 60(11):1259–1288, 2009.
- [22] Luz Amaya-Bower and Taehun Lee. Single bubble rising dynamics for moderate reynolds number using lattice boltzmann method. *Computers & Fluids*, 39(7):1191–1207, 2010. ISSN 0045-7930. doi: <https://doi.org/10.1016/j.compfluid.2010.03.003>. URL <https://www.sciencedirect.com/science/article/pii/S004579301000054X>.
- [23] Sebastian Aland and Axel Voigt. Benchmark computations of diffuse interface models for two-dimensional bubble dynamics. *International Journal for Numerical Methods in Fluids*, 69(3):747–761, 2012.
- [24] H. Z. Yuan, Z. Chen, C. Shu, Y. Wang, X. D. Niu, and S. Shu. A free energy-based surface tension force model for simulation of multiphase flows by level-set method. *Journal of Computational Physics*, 345:404–426, 2017. ISSN 10902716. doi: 10.1016/j.jcp.2017.05.020.
- [25] Makrand A. Khanwale, Kumar Saurabh, Masado Ishii, Hari Sundar, James A. Rossmanith, and Baskar Ganapathysubramanian. A projection-based, semi-implicit time-stepping approach for the cahn-hilliard navier-stokes equations on adaptive octree meshes. *Journal of Computational Physics*, 475:111874, 2023. ISSN 0021-9991. doi: <https://doi.org/10.1016/j.jcp.2022.111874>. URL <https://www.sciencedirect.com/science/article/pii/S0021999122009378>.
- [26] Junxiang Yang, Darae Jeong, and Junseok Kim. A fast and practical adaptive finite difference method for the conservative allen-cahn model in two-phase flow system. *International Journal of Multiphase Flow*, 137:103561, 2021. ISSN 0301-9322. doi: <https://doi.org/10.1016/j.ijmultiphaseflow.2021.103561>. URL <https://www.sciencedirect.com/science/article/pii/S0301932221000094>.
- [27] M. Jalaal and K. Mehravaran. Fragmentation of falling liquid droplets in bag breakup mode. *International Journal of Multiphase Flow*, 47:115–132, 2012. ISSN 0301-9322. doi: <https://doi.org/10.1016/j.ijmultiphaseflow.2012.07.011>. URL <https://www.sciencedirect.com/science/article/pii/S0301932212001140>.
- [28] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. ISSN 25225820. doi: 10.1038/s42254-021-00314-5. URL <https://doi.org/10.1038/s42254-021-00314-5>.
- [29] Sheikh Md Shakeel Hassan, Arthur Feeney, Akash Dhruv, Jihoon Kim, Youngjoon Suh, Jaiyoung Ryu, Yoonjin Won, and Aparna Chandramowlishwaran. BubbleML: A multiphase multiphysics dataset and benchmarks for machine learning. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023. URL <https://openreview.net/forum?id=0Wmglu8zak>.
- [30] Rômulo M Silva, Malú Grave, and Alvaro LGA Coutinho. A pinn-based level-set formulation for reconstruction of bubble dynamics. *Archive of Applied Mechanics*, pages 1–16, 2024.
- [31] Majdi Chaari, Abdennour C Seibi, Jalel Ben Hmida, and Afef Fekih. An optimized artificial neural network unifying model for steady-state liquid holdup estimation in two-phase gas–liquid flow. *Journal of Fluids Engineering*, 140(10):101301, 2018.
- [32] Ala S Al-Dogail and Rahul N Gajbhiye. Effects of density, viscosity and surface tension on flow regimes and pressure drop of two-phase flow in horizontal pipes. *Journal of Petroleum Science and Engineering*, 205:108719, 2021.
- [33] Utkarsh Ayachit. *The paraview guide: a parallel visualization application*. Kitware, Inc., 2015.
- [34] Roland Clift, John R Grace, and Martin E Weber. *Bubbles, drops, and particles*. Courier Corporation, 2013.
- [35] Jens Eggers and Todd F Dupont. Drop formation in a one-dimensional approximation of the navier–stokes equation. *Journal of fluid mechanics*, 262:205–221, 1994.
- [36] Christophe Clanet, Cédric Béguin, Denis Richard, and David Quéré. Maximal deformation of an impacting drop. *Journal of Fluid Mechanics*, 517:199–208, 2004.
- [37] Zongyi Li, Nikola Kovachki, Kamyr Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations, 2021.

- [38] Samuel Lanthaler, Andrew M. Stuart, and Margaret Trautner. Discretization error of fourier neural operators, 2024.
- [39] Bogdan Raonić, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of pdes, 2023.
- [40] Oliver Penrose and Paul C. Fife. Thermodynamically consistent models of phase-field type for the kinetic of phase transitions. *Physica D: Nonlinear Phenomena*, 43(1):44–62, may 1990. ISSN 01672789. doi: 10.1016/0167-2789(90)90015-H. URL <https://linkinghub.elsevier.com/retrieve/pii/016727899090015H>.
- [41] David Jacqmin. Calculation of Two-Phase Navier-Stokes Flows Using Phase-Field Modeling. *Journal of Computational Physics*, 155(1):96–127, oct 1999. ISSN 00219991. doi: 10.1006/jcph.1999.6332. URL <https://linkinghub.elsevier.com/retrieve/pii/S0021999199963325>.
- [42] Samuel M. Allen and John W. Cahn. Mechanisms of phase transformations within the miscibility gap of Fe-rich Fe-Al alloys. *Acta Metallurgica*, 24(5):425–437, may 1976. ISSN 00016160. doi: 10.1016/0001-6160(76)90063-8. URL <https://linkinghub.elsevier.com/retrieve/pii/0001616076900638>.
- [43] Abbas Fakhari, Martin Geier, and Diogo Bolster. A simple phase-field model for interface tracking in three dimensions. *Computers & Mathematics with Applications*, 78(4):1154–1165, aug 2019. ISSN 08981221. doi: 10.1016/j.camwa.2016.08.021. URL <http://dx.doi.org/10.1016/j.camwa.2016.08.021><https://linkinghub.elsevier.com/retrieve/pii/S0898122116304758>.
- [44] Pao-Hsiung Chiu and Yan-Ting Lin. A conservative phase field method for solving incompressible two-phase flows. *Journal of Computational Physics*, 230(1):185–204, jan 2011. ISSN 00219991. doi: 10.1016/j.jcp.2010.09.021. URL <http://dx.doi.org/10.1016/j.jcp.2010.09.021><https://linkinghub.elsevier.com/retrieve/pii/S0021999110005243>.
- [45] Y. Y. Yan and Y. Q. Zu. A lattice Boltzmann method for incompressible two-phase flows on partial wetting surface with large density ratio. *Journal of Computational Physics*, 227(1):763–775, nov 2007. ISSN 10902716. doi: 10.1016/j.jcp.2007.08.010. URL <https://linkinghub.elsevier.com/retrieve/pii/S0021999107003580>.
- [46] Hang Ding, Peter D.M. Spelt, and Chang Shu. Diffuse interface model for incompressible two-phase flows with large density ratios. *Journal of Computational Physics*, 226(2):2078–2095, oct 2007. ISSN 00219991. doi: 10.1016/j.jcp.2007.06.028. URL <https://linkinghub.elsevier.com/retrieve/pii/S0021999107002793>.
- [47] Q. Li, K. H. Luo, Y. J. Gao, and Y. L. He. Additional interfacial force in lattice Boltzmann models for incompressible multiphase flows. *Physical Review E*, 85(2):026704, feb 2012. ISSN 1539-3755. doi: 10.1103/PhysRevE.85.026704. URL <https://link.aps.org/doi/10.1103/PhysRevE.85.026704>.
- [48] Didier Jamet, David Torres, and J. U. Brackbill. On the theory and computation of surface tension: The elimination of parasitic currents through energy conservation in the second-gradient method. *Journal of Computational Physics*, 182(1):262–276, oct 2002. ISSN 00219991. doi: 10.1006/jcph.2002.7165.
- [49] DAVID JACQMIN. Contact-line dynamics of a diffuse fluid interface. *Journal of Fluid Mechanics*, 402:57–88, jan 2000. ISSN 0022-1120. doi: 10.1017/S0022112099006874. URL https://www.cambridge.org/core/product/identifier/S0022112099006874/type/journal_article.
- [50] Michael C. Sukop and Daniel T. Thorne. *Lattice boltzmann modeling: An introduction for geoscientists and engineers*. Springer Berlin Heidelberg, 2006. ISBN 3540279814. doi: 10.1007/978-3-540-27982-2. URL <http://link.springer.com/10.1007/978-3-540-27982-2>.
- [51] Haibo Huang, Michael C. Sukop, and Xi Yun Lu. *Multiphase Lattice Boltzmann Methods: Theory and Application*. John Wiley & Sons, Ltd, jul 2015. ISBN 9781118971338. doi: 10.1002/9781118971451. URL <http://doi.wiley.com/10.1002/9781118971451>.

- [52] Timm Krüger, Halim Kusumaatmaja, Alexandre Kuzmin, Orest Shardt, Goncalo Silva, and Erlend Magnus Viggen. *The Lattice Boltzmann Method*, volume 58 of *Graduate Texts in Physics*. Springer International Publishing, 2017. ISBN 978-3-319-44647-9. doi: 10.1007/978-3-319-44649-3. URL <http://link.springer.com/10.1007/978-3-319-44649-3>.
- [53] Zhaoli Guo, Chuguang Zheng, and Baochang Shi. Discrete lattice effects on the forcing term in the lattice Boltzmann method. *Physical Review E*, 65(4):046308, apr 2002. ISSN 1063-651X. doi: 10.1103/PhysRevE.65.046308. URL <https://link.aps.org/doi/10.1103/PhysRevE.65.046308>.
- [54] Martin Geier, Abbas Fakhari, and Taehun Lee. Conservative phase-field lattice Boltzmann model for interface tracking equation. *Physical Review E*, 91(6):063309, jun 2015. ISSN 1539-3755. doi: 10.1103/PhysRevE.91.063309. URL <https://link.aps.org/doi/10.1103/PhysRevE.91.063309>.
- [55] Andrea Prosperetti. Motion of two superposed viscous fluids. *Physics of Fluids*, 24(7): 1217 – 1223, 1981. doi: 10.1063/1.863522. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0019376759&doi=10.1063%2f1.863522&partnerID=40&md5=eb53d8f4f5117a9f1224ba3ed1d61239>. Cited by: 149.
- [56] D. Bhaga and M. E. Weber. Bubbles in viscous liquids: shapes, wakes and velocities. *Journal of Fluid Mechanics*, 105:61–85, 1981. doi: 10.1017/S002211208100311X.
- [57] Monica Gumulya, Jyeshtharaj B. Joshi, Ranjeet P. Utikar, Geoffrey M. Evans, and Vishnu Pareek. Bubbles in viscous liquids: Time dependent behaviour and wake characteristics. *Chemical Engineering Science*, 144:298–309, 2016. ISSN 0009-2509. doi: <https://doi.org/10.1016/j.ces.2016.01.051>. URL <https://www.sciencedirect.com/science/article/pii/S0009250916300318>.
- [58] Feng Ren, Baowei Song, Michael C. Sukop, and Haibao Hu. Improved lattice boltzmann modeling of binary flow based on the conservative allen-cahn equation. *Phys. Rev. E*, 94: 023311, Aug 2016. doi: 10.1103/PhysRevE.94.023311. URL <https://link.aps.org/doi/10.1103/PhysRevE.94.023311>.
- [59] Y. Q. Zu and S. He. Phase-field-based lattice boltzmann model for incompressible binary fluid systems with density and viscosity contrasts. *Phys. Rev. E*, 87:043301, Apr 2013. doi: 10.1103/PhysRevE.87.043301. URL <https://link.aps.org/doi/10.1103/PhysRevE.87.043301>.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default [TODO] to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section 3. The dataset is CC BY NC. The code used to simulate the two-phase flows is proprietary.

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
 - (a) Do the main claims made in the abstract, and introduction accurately reflect the paper's contributions and scope? [Yes]
 - (b) Did you describe the limitations of your work? [Yes] We describe the limitations of our work in **Section 5**
 - (c) Did you discuss any potential negative societal impacts of your work? [N/A] Since the data generated involves fundamental science, the negative impact it has on the society is of similar nature to any fundamental science itself.
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? [N/A]
 - (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments (e.g. for benchmarks)...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] **We provide the code, data, and instructions to reproduce the main experimental results on the MPF-Bench website.**
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] **The dataset itself is generated deterministically from simulations. Where possible, we ran multiple experiments and will report the error.**
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] **We mention the simulation efforts in Section 3.2**
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? [Yes] **We cite the code used for SciML in Section 4**
 - (b) Did you mention the license of the assets? [Yes]
 - (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] **We included all material on our MPF-Bench website**
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

- (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
- (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Details of the CFD simulation framework

Our computational framework employs the CUDA platform to implement the algorithms necessary for the Lattice Boltzmann Method (LBM). We achieve significant computational performance enhancements by leveraging CUDA's parallel processing capabilities. The primary performance bottleneck in GPU architectures is often the data transfer between GPU memory and unified CPU memory. To mitigate this, we minimize such data transfers, conducting them only when necessary for convergence checks or final output retrieval.

We utilize a single one-dimensional array in conjunction with macro functions to handle the substantial data volumes intrinsic to LBM simulations. This method optimizes memory usage and computational efficiency on the GPU, ensuring that we fully exploit the GPU's computational power and memory bandwidth. This strategy allows for the high-performance execution of LBM algorithms, crucial for large-scale simulations and complex fluid dynamics problems.

A.1 Formulation of Navier Stokes and Allen Cahn equations

Several lattice Boltzmann models, such as the Cahn-Hilliard and Allen-Cahn models, utilize interface tracking equations and are thus categorized as phase-field models [40, 41]. These models describe multiphase flows using a diffuse interface, with the Allen-Cahn equation commonly employed for this purpose [42]. In some studies, this approach is called the conservative phase-field LB model [43]. The phase-field variable, ϕ , which tracks the interface, ranges from 0 to 1, leading to the following expression for the phase-field equation [44]:

$$\frac{\partial \phi}{\partial t} + \nabla \cdot (\phi u) = \nabla \cdot \left[M(\nabla \phi - \frac{1 - 4(\phi - \phi_0)^2}{\xi} \hat{n}) \right], \quad (2)$$

where t represents time, u is the velocity, M denotes a positive constant for the mobility parameter, ξ is the interfacial thickness, and $\phi_0 = \frac{\phi_H + \phi_L}{2}$. ϕ_H and ϕ_L represent the interface indicator values for the heavy and light fluids, respectively, set to 1.0 for the heavy fluid and 0.0 for the light fluid. The unit normal vector \hat{n} for the interface can be defined as:

$$\hat{n} = \frac{\nabla \phi}{|\nabla \phi|}. \quad (3)$$

Note, the interface location at x_0 is initialized as [45]:

$$\phi(x) = \phi_0 \pm \frac{\phi_H - \phi_L}{2} \tanh\left(\frac{|x - x_0|}{\xi/2}\right). \quad (4)$$

According to the phase-field model, the following equations exist for incompressible multiphase flows [46, 47]:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0, \quad (5a)$$

$$\rho \left(\frac{\partial u}{\partial t} + u \cdot \nabla u \right) = -\nabla p + \nabla \cdot (\mu [\nabla u + (\nabla u)^T]) + F_s + F_b. \quad (5b)$$

In Equation 5a, ρ represents the density of fluids, p denotes the macroscopic pressure, F_b is the body force, and F_s corresponds to the surface tension force. The equation for calculating the surface tension force term is also expressed as [48]:

$$F_s = \mu_\phi \nabla \phi, \quad (6)$$

where

$$\mu_\phi = 4\beta\phi(\phi - 1)(\phi - 1/2) - \kappa\nabla^2\phi, \quad (7)$$

denotes the chemical potential equation utilized for binary fluids [49]. Equation 8 establishes a

relation between the coefficients β and κ , interface thickness ξ , and surface tension σ , as;

$$\beta = 12\sigma/\xi, \quad \kappa = 3\sigma\xi/2. \quad (8)$$

A.2 Lattice Boltzmann Method

Given that interfaces are typically of mesoscopic scale, the kinetic-based Lattice Boltzmann Method (LBM) presents a more effective approach for simulating multiphase flows compared to the traditional Navier-Stokes solvers [50, 51]. The Chapman-Enskog analysis validates the consistency between the LBM and the Navier-Stokes equations [52]. In this study, we investigate hydrodynamic properties such as velocity and pressure using the standard form of the Lattice Boltzmann equation as outlined in [53]:

$$f_a(x + e_a \delta t, t + \delta t) = f_a(x, t) + \Omega_a(x, t) + F_a(x, t), \quad (9)$$

In this context, f_a denotes the velocity-based hydrodynamic distribution function for incompressible fluids, Ω_a represents the collision operator, and F_a signifies the force term. This study employs the two-dimensional nine-velocity (D2Q9) model for 2D simulations and the three-dimensional nineteen-velocity (D3Q19) model for 3D simulations.

To define the interface between phases, we employed the following Lattice Boltzmann Equation (LBE) to accurately determine the interface between fluid phases [54]:

$$g_a(x + e_a \delta t, t + \delta t) = g_a(x, t) - \frac{g_a(x, t) - \bar{g}_a^{eq}(x, t)}{\tau_\phi + 1/2} + F_a^\phi(x, t). \quad (10)$$

Here, g_a represents the distribution function for the phase-field, and τ_ϕ denotes the dimensionless phase-field relaxation time. The forcing term is calculated as follows:

$$F_a^\phi(x, t) = \delta t \frac{\left[1 - 4(\phi - \phi_0)^2\right]}{\xi} \omega_a e_a \cdot \frac{\nabla \phi}{|\nabla \phi|}. \quad (11)$$

In Equation 11, ω_a and e_a denote the weight coefficient and the mesoscopic velocity set, respectively. Here, ξ denotes the thickness of the interface. As illustrated in Figure 7, we carefully selected this parameter to ensure adequate lattice nodes within the interface. This choice is critical for accurately capturing the complex physics in the rapid change of material properties across the interface. The appropriate selection of ξ ensures that the computational mesh can effectively represent the gradients and variations within the interface, thus enhancing the overall stability and accuracy of the simulation.

A.3 Validation

In this section, we validate our numerical model through benchmark tests covering a range of two-phase flow phenomena. We include four distinct validation cases to comprehensively assess the accuracy and robustness of our approach: (1) the capillary wave problem, which evaluates the model's capability to handle surface tension-driven flows; (2) the bubble rising dynamics, which tests the interaction between buoyancy and viscous forces; (3) the falling droplet dynamics, which examines the breakup mechanisms of liquid droplets under gravity; and (4) the Rayleigh-Taylor instability, which explores the interfacial instability between fluids of differing densities under gravitational influence. Each subsection compares our simulation results and established experimental or numerical data, demonstrating the model's fidelity across various flow regimes.

A.3.1 Capillary Wave

To validate our Lattice Boltzmann Method (LBM) simulations of two-phase flow, we focus on the dynamic behavior of capillary waves at the interface between two immiscible fluids. In our study, a sinusoidal perturbation with a small amplitude η_0 and wave number k is applied to the initially quiescent interface. This setup provides a rigorous test for the LBM framework, as it has a well-established analytical solution for cases with identical kinematic viscosities ν but differing densities of the two fluids. The temporal evolution of the interface amplitude $\eta(t)$ is utilized as a benchmark

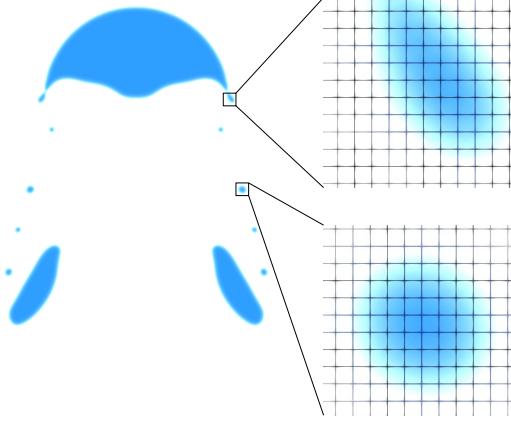


Figure 7: Illustration of the interface region captured by the computational mesh. The magnified views show the distribution of lattice nodes within the interface, ensuring precise resolution of interfacial dynamics and transitions. The careful selection of the interface thickness parameter ξ ensures that the mesh adequately represents the gradients and variations in the interface region.

for our simulations. The analytical expression for the decay of the wave amplitude, $\eta(t)$, is given by [55]:

$$\frac{\eta(t)}{\eta_0} = \frac{4(1-4\gamma)\nu^2 k^4}{8(1-4\gamma)\nu^2 k^4 + \omega_0} \operatorname{erfc}(\sqrt{\nu k^2 t}) + \sum_{i=1}^4 \frac{z_i}{Z_i} \frac{\omega_0^2}{z_i^2 - \nu k^2} e^{(z_i^2 - \nu k^2)t} \operatorname{erfc}(z_i \sqrt{\nu t}) \quad (12)$$

where $\omega_0 = \sqrt{\frac{\sigma k^3}{\rho_H + \rho_L}}$ is the angular frequency, $\gamma = \frac{\rho_H \rho_L}{(\rho_H + \rho_L)^2}$ and $Z_i = \prod_{\substack{1 \leq j \leq 4 \\ j \neq i}} (z_j - z_i)$. The evaluation of the complementary error function $\operatorname{erfc}(z_i)$ can be done by solving the following algebraic equation:

$$z^4 - 4\gamma\sqrt{\nu k^2}z^3 + 2(1-6\gamma)\nu k^2 z^2 + 4(1-3\gamma)(\nu k^2)^{3/2}z + (1-4\gamma)\nu k^2 + \omega_0^2 = 0. \quad (13)$$

Our validation involves analyzing the propagation of capillary waves, an inherently transient process that tests the model's ability to accurately capture key physical parameters such as density and viscosity ratios, along with surface tension effects. By varying these parameters and the wavelength, we compare the simulation results with predictions from linear theory. According to Figure 8, the lighter fluid with density ρ_L overlays the heavier fluid with density ρ_H , with the initial interface described by $y = L + \eta_0 \cos(2\pi x)$, where η_0 is the initial perturbation amplitude. The decay of this wavy profile to a flat interface, driven by viscosity and surface tension, without external forces like gravity, serves as a critical validation test for our LBM approach. The computational domain is discretized into a grid of 256 by 512 lattice nodes. Free-slip boundary conditions are applied in the direction of wave propagation, while no-slip conditions are imposed at the top and bottom boundaries. The simulation parameters are set as follows: $\eta_0 = 0.02$, $\sigma = 10^{-4}$, $\xi = 4$, and $M_\phi = 0.02$. Since the interface may not align exactly with the grid points, the values of $\eta(t)$ are interpolated from ϕ values using the following relationship:

$$\eta(t) = y - \frac{\phi(x_{L0/2}, y)}{\phi(x_{L0/2}, y) - \phi(x_{L0/2}, y-1)}, \quad \phi(x_{L0/2}, y)\phi(x_{L0/2}, y-1) < 0. \quad (14)$$

The length (η) and time scales (t) are normalized by the initial amplitude a_0 and the angular frequency ω_0 , respectively, denoted as $\eta^* = \eta/\eta_0$ and $t^* = t\omega_0$.

It is worth noting that angular frequency is crucial for any wave system. It depends on surface tension, viscosity, wave number, and density values. The equation is derived assuming that both fluids have the same viscosity, set to $\nu = 0.005, 0.0005$. Note that the wavelength magnitude matches the grid size $L_0 = 256$.

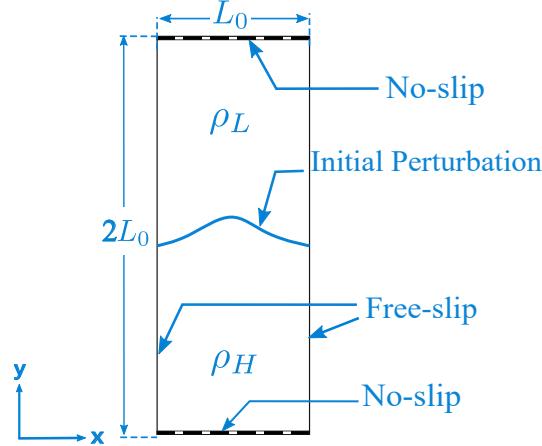


Figure 8: Schematic diagram of the capillary wave problem setup.

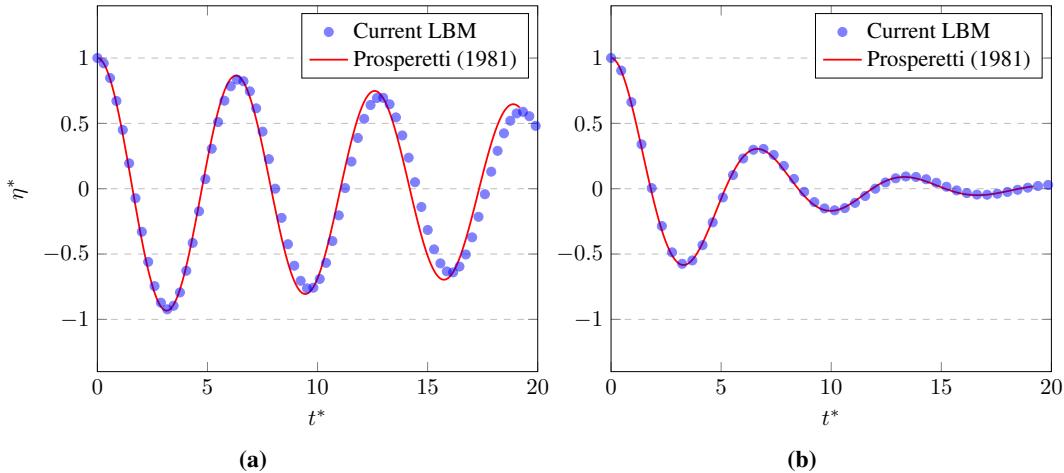


Figure 9: Comparison of the normalized interface amplitude η^* as a function of normalized time t^* between the current LBM simulation and the analytical solution by Prosperetti (1981). (a) corresponds to a viscosity of $\nu = 0.0005$, and (b) corresponds to a viscosity of $\nu = 0.005$. The LBM results (blue circles) closely match the analytical results (red line).

A.3.2 Rise of a single bubble in quiescent fluid

The dynamics of a rising bubble have been extensively studied due to their significance in various natural and industrial processes. When a bubble rises through a liquid, it is subjected to several forces, including buoyancy, drag, and surface tension, which influences its shape, velocity, and trajectory [19, 20, 22, 25]. Our investigation focuses on the dynamics of a bubble rising within a rectangular channel. The simulation begins with a circular bubble of diameter $D = L_0/5$ placed at the coordinates $(L_0/2, L_0/2)$ within a domain with a length of L_0 and a height of $4L_0$. Boundary conditions are set such that the no-slip is applied at the top and bottom, while free-slip boundary conditions are used for the lateral boundaries. The fluids experience a volumetric buoyancy force $F_b = -(\rho_h - \rho_l)g_y\mathbf{j}$, where g_y represents the gravitational acceleration in the y -direction. This study highlights four crucial dimensionless parameters: the density ratio ρ_h/ρ_l , the viscosity ratio μ_h/μ_l , the gravity Reynolds number, and the Eötvös (Bond) number.

The gravity Reynolds number is defined as:

$$\text{Re}_h = \frac{\sqrt{g_y\rho_h(\rho_h - \rho_l)D^3}}{\mu_h} \quad (15)$$

The Eötvös (Bond) number is defined as:

$$\text{Eo} = \frac{g_y(\rho_h - \rho_l)D^2}{\sigma} \quad (16)$$

In many studies, the Morton number is also considered, defined as:

$$Mo = \frac{g_y(\rho_h - \rho_l)\mu_h^4}{\sigma^3\rho_h^2} \quad (17)$$

The dimensionless time is also defined by:

$$t^* = t\sqrt{\frac{g_y}{D}} \quad (18)$$

The reference velocity scale needed in the Péclet number can be chosen for gravity-driven flows as $U_0 = \sqrt{g_y D}$. Four sets of simulations are conducted at Four different Eötvös and Morton numbers. The density and viscosity ratios are fixed at 1000 and 100, respectively. The numerical parameters are $L_0 = 512$, $Pe = 25$ and $Cn = 0.010$, and the LBM simulation results are shown in Figure 10.

To evaluate the accuracy and reliability of the proposed LBM, a comparison is made between the results obtained from the LBM approach and those from the experiments and FVM, as illustrated in Figure 10. In the spherical regime, surface tension dominates, resulting in small bubbles that maintain a nearly spherical shape due to the strong cohesive forces at the interface. As the bubble size increases, the shape transitions to an ellipsoidal form. In this ellipsoidal regime, the inertial forces become more significant, causing the bubble to deform. This deformation is influenced by the surrounding liquid's viscosity and the interface's surface tension. The dynamics of this regime can be described using correlations that account for the balance between inertial and surface tension forces [22]. In the spherical cap regime, the bubbles are large enough that inertia forces dominate, leading to further deformation into a cap shape. This regime is characterized by a significant increase in terminal velocity, which is proportional to the size of the bubble [56]. These patterns are consistent among all results.

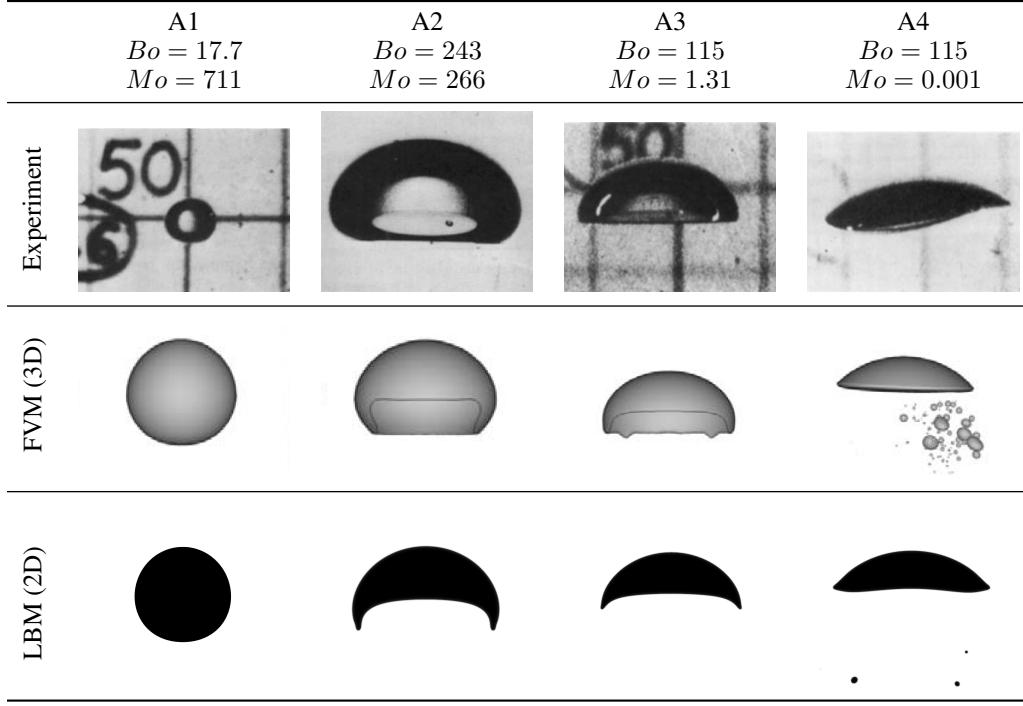


Figure 10: Comparison of bubble shapes at constant rise velocity: Experimental results by Bhaga and Weber (1981) [56], LBM results, and FVM results by Gumulya et al. (2016) [57] for various Bond numbers (Bo) and Morton numbers (Mo).

A.3.3 Falling Droplet

The dynamics of a falling droplet under gravity is another fascinating two-phase flow phenomenon that has been extensively studied in the literature [26, 27]. In this study, a liquid droplet with diameter

$D = L_0/5$ is initially placed at $(L_0/2, 6L_0/2)$ within a rectangular computational domain of length L_0 and height $3L_0$. The same boundary conditions are applied as in the bubble rising simulations: the no-slip boundary condition is applied at the top and bottom, while free-slip boundary conditions are imposed at the lateral boundaries. The volumetric buoyancy force $F_b = -(\rho_h - \rho_l)g_y\mathbf{j}$, where \mathbf{j} is unit vector in y -direction and g_y represents the gravitational acceleration in the y -direction, acts on the fluids.

The dimensionless analysis identifies several key parameters that characterize the flow: the density ratio ρ_h/ρ_l , the viscosity ratio μ_h/μ_l , the gravity Reynolds number, and the Eötvös (Bond) number. The gravity Reynolds number is defined as:

$$\text{Re}_h = \frac{\sqrt{g_y\rho_h(\rho_h - \rho_l)}D^3}{\mu_h} \quad (19)$$

Similarly, the Eötvös number, which represents the ratio of gravitational forces to surface tension forces, is given by:

$$\text{Eo} = \frac{g_y(\rho_h - \rho_l)D^2}{\sigma} \quad (20)$$

Another important dimensionless group in the literature is the Morton number, which characterizes the fluid properties affecting the bubble and droplet dynamics:

$$\text{Mo} = \frac{g_y(\rho_h - \rho_l)\mu_h^4}{\sigma^3\rho_h^2} \quad (21)$$

The Ohnesorge number (Oh) is a dimensionless number that characterizes the relative importance of viscous forces compared to inertial and surface tension forces in a fluid. It is particularly relevant in the study of droplet dynamics and is defined as:

$$\text{Oh} = \frac{\mu_h}{\sqrt{\rho_h\sigma D}} \quad (22)$$

The simulation is conducted at a moderate density ratio to capture the breakup mechanisms of the falling droplet, allowing for comparisons with the VOF model. The simulation considers an Eötvös number: $Eo = 288$, with density and viscosity ratios fixed at 10 and 1, respectively, and the Oh number set to 0.05. The numerical parameters are $Pe = 5$ and $Cn = 0.010$. As mentioned in [Section A.3.2](#), the reference velocity scale needed for the Péclet number can be chosen as $U_0 = \sqrt{g_y D}$ for gravity-driven flows. Also, dimensionless time can be defined by:

$$t^* = t\sqrt{\frac{g_y}{D}} \quad (23)$$

Our simulation results exhibit excellent agreement with the findings of Jalaal and Mehravar [27]. As shown in [Figure 11](#), the comparison of the deformation of a liquid drop using both the Lattice Boltzmann Method (LBM) in 2D and the Volume of Fluid (VOF) method in 3D demonstrates that the evolution of the drop shapes over time is remarkably similar. For instance, at $t^* = 0.1647$, both methods capture the formation of a curved interface, and at $t^* = 0.3575$, the drop breakup into smaller droplets is observed in both approaches. This consistency across different numerical methods, with parameters set at $Eo = 288$, $Oh_h = Oh_l = 0.05$, and $\rho^* = 10$, validates the robustness and accuracy of our LBM simulations in replicating complex two-phase flow phenomena.

Overall, the dynamics of falling droplets involve complex interactions between buoyancy, inertia, and surface tension forces, leading to various deformation and breakup patterns, such as forming bags, ligaments, and secondary droplets. These phenomena are influenced significantly by the Eötvös number, with higher values leading to more pronounced deformations and faster breakup processes [27].

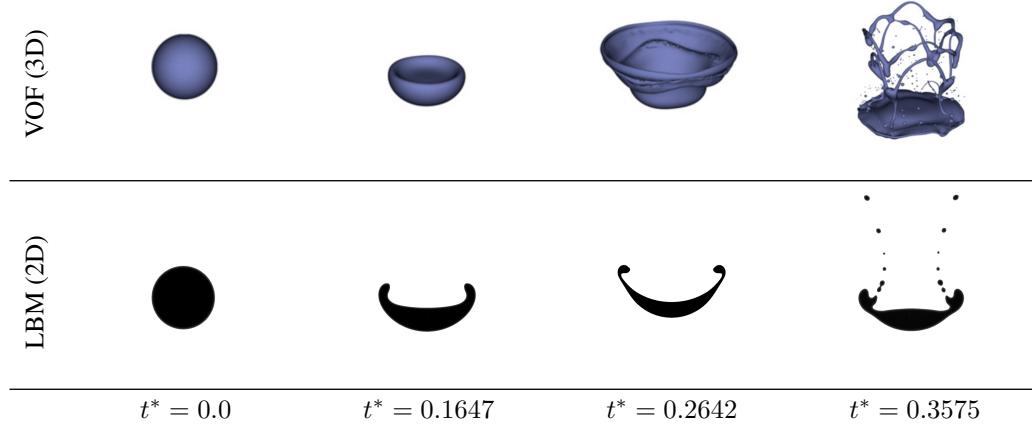


Figure 11: Comparison of the deformation of a liquid drop using the LBM and VOF methods: Current results and those of Jalaal et al. (2012) [27] for $Eo = 288$, $Oh_h = Oh_l = 0.05$, and $\rho^* = 10$.

A.3.4 Rayleigh-Taylor Instabilities

The Rayleigh-Taylor instability (RTI) arises when a denser fluid is positioned above a less dense fluid in the presence of a gravitational field, causing the interface between the two fluids to become unstable. This phenomenon has been extensively studied due to its relevance in various natural and engineering contexts [25, 58, 59].

We consider a computational domain of size $[0, L_0] \times [0, 4L_0]$ with $L_0 = 256$ for our simulations. The initial interface is defined as $y_0(x) = 2L_0 + 0.1L_0 \cos(2\pi x/L_0)$. Periodic boundary conditions are applied on the left and right boundaries, while no-slip conditions are enforced at the top and bottom boundaries. The dimensionless numbers characterizing the RTI include the Atwood number, Reynolds number, Capillary number, and Peclet number:

$$At = \frac{\rho_H - \rho_L}{\rho_H + \rho_L}, \quad (24)$$

$$Re = \frac{\rho_H U_0 L_0}{\mu_H}, \quad (25)$$

where $U_0 = \sqrt{g_y L_0}$,

$$Ca = \frac{\mu_H U_0}{\sigma}, \quad (26)$$

$$Pe = \frac{U_0 L_0}{M}. \quad (27)$$

In our study, we used a density ratio $\rho^* = 3$, viscosity ratio $\mu^* = 1$, Reynolds number $Re = 128$, Atwood number $At = 0.5$, Peclet number $Pe = 744$, and interface width $\xi = 5$. The results are compared with the findings from Ren et al. [58] and Zu and He [59]. The dimensionless time is defined as $t^* = t/t_0$, where $t_0 = \sqrt{L_0/(gAt)}$.

Snapshots of the interface evolution for the 2D Rayleigh-Taylor instability at different times are shown in Figure 12. Initially, the interface undergoes a symmetrical penetration of the heavier fluid into the lighter fluid, forming counter-rotating vortices. As time progresses, the heavier fluid rolls up into mushroom-like shapes, and secondary vortices form at the tails of the roll-ups. Our simulations' interface patterns and vortex structures are consistent with those reported in previous studies [58, 59].

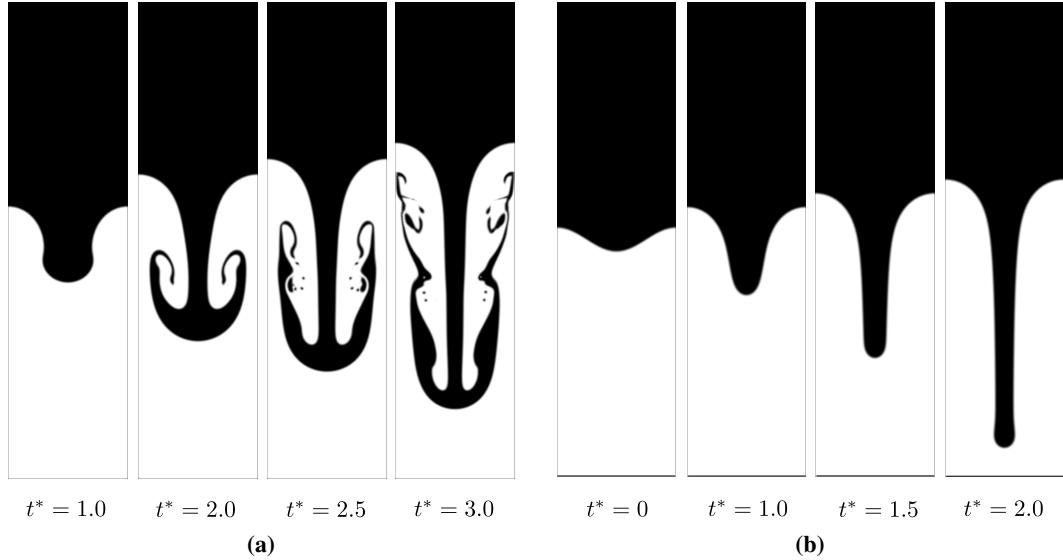


Figure 12: Evolution of the interface pattern of the 2D Rayleigh-Taylor instability for two scenarios: (a) $\rho^* = 3$, $\mu^* = 1$, $Re = 128$, $At = 0.500$, $Pe = 744$, $\xi = 5$; (b) $\rho^* = 1000$, $\mu^* = 100$, $Re = 3000$, $At = 0.998$, $Pe = 200$, $Ca = 8.7$, $\xi = 5$.

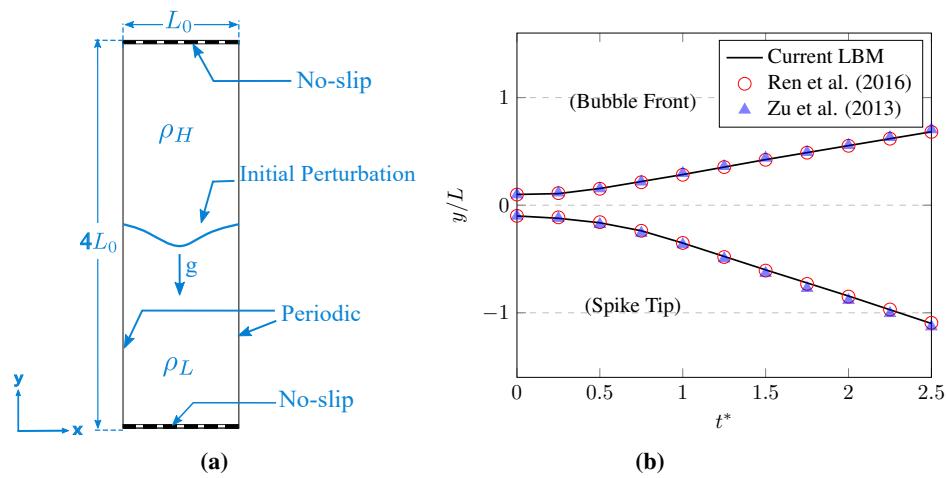


Figure 13: (a) Schematic of the initial setup for the Rayleigh-Taylor instability simulation, showing the boundary conditions and initial perturbation. (b) Comparison of the bubble front and spike tip positions over time for the Rayleigh-Taylor instability case with parameters $\rho^* = 3$, $\mu^* = 1$, $Re = 128$, $At = 0.500$, $Pe = 744$, and $\xi = 5$. The current LBM results (solid line) are compared with the results of Ren et al. [58] (red circles) and Zu and He [59] (blue triangles), showing excellent agreement in capturing the evolution of the instability.