Computers and Fluids 255 (2023) 105833

Contents lists available at ScienceDirect

Computers and Fluids

journal homepage: www.elsevier.com/locate/compfluid

HPC performance study of different collision models using the Lattice Boltzmann solver Musubi

Gregorio Gerardo Spinelli [a,*], Tobias Horstmann [b], Kannan Masilamani [a], Malav Mukesh Soni [c], Harald Klimach [a], Arthur Stück [a], Sabine Roller [a]

[a] German Aerospace Center (DLR), Institute of Software Methods for Product Virtualization, Dresden, Germany
[b] German Aerospace Center (DLR), Institute of Propulsion Technology, Berlin, Germany
[c] German Aerospace Center (DLR), Institute of Aerodynamics and Flow Technology, Braunschweig, Germany

ARTICLE INFO

Keywords:
Lattice Boltzmann method
Hybrid recursive regularized BGK
Cumulant
Turbulence modeling
Musubi

ABSTRACT

Over the past decades, the lattice Boltzmann method (LBM) has become increasingly popular thanks to its capabilities in the domain of Large-Eddy Simulations (LES). Different collision schemes have been proposed to extend the scope of application to higher Reynolds number flows. This study compares the accuracy and the performance of some of these schemes on a $D3Q27$ lattice, including the original Multiple Relaxation Times (MRT) model, the Hybrid Recursive Regularized Bhatnagar–Gross–Krook (HRR) operator, as well as the Projected Recursive Regularized Bhatnagar–Gross–Krook (PRR) operator and the parametrized Cumulant collision scheme. For this purpose, the above-mentioned schemes are implemented in the HPC LBM solver Musubi and tested on a well-documented test-case describing the flow past a circular cylinder at a Reynolds number of 3900. Three different subgrid scale (SGS) models are used to account for the unresolved turbulence, *i.e.* the Smagorinsky model, the Wall-Adapting Local Eddy-viscosity (WALE) model, and the Vreman model. The Cumulant scheme uses an Implicit LES (ILES) subgrid scale model and shows the best agreement with the experimental data followed by MRT with WALE, and HRR with Vreman. The examined collision models are able to capture the second peak at $f = 3f_{vs}$ of the power spectra density of the y velocity component first discovered in experiments.

With respect to performance, the collision models are compared in terms of MLUPs/node and parallel efficiency for a strong scaling analysis. Again the Cumulant scheme outperforms the other collision models even when they are run on the reduced $D3Q19$ stencil. All the collision schemes show a strong scaling parallel efficiency above 60% on up to 16384 cores in our implementation.

1. Introduction

Over the past years, the lattice Boltzmann method has further strengthened its position as a valuable tool in the field of Computational Fluid Dynamics (CFD) [1–4]. Due to its high parallel efficiency and ability to discretize complex geometries with little effort, it found its way into many industries and research organizations. In contrast to conventional methods that solve for a set of macroscopic conservation laws, this method populates a cartesian (quadtree/octree) mesh with discrete particle densities that hop between neighboring nodes during exactly one time-step. This so-called *streaming* step is followed (or rather preceded) by a local *collision*, which drives the distribution functions toward a thermodynamic equilibrium [5]. Mathematically, the *streaming* is an exact solution to the Lagrangian derivative of the populations along the discrete velocity directions defined by the lattice. This property, which computationally translates in a simple index shift in memory is one of the main reasons for the success of this method. However, the lack of numerical dissipation restricts the original method to the low Reynolds number regime.

A remedy may be provided by using a discretization in the Eulerian sense [6], which nevertheless compromises many of the method's advantages. More popular strategies to increase the stability limit address the collision operator. The simplest operator that respects the so-called *H-Theorem*, i.e. the increase in entropy during the collision, is the Bhatnagar–Gross–Krook (BGK) operator [7]. Here, the entire set of populations is relaxed with a Single Relaxation Time (SRT). An alternative approach transforms the populations into moment space before the collision, which then allows the use of Multiple Relaxation Times (MRT) [8–11] to specifically relax the different flow quantities

* Corresponding author.
E-mail address: gregoriogerardo.spinelli@dlr.de (G.G. Spinelli).

https://doi.org/10.1016/j.compfluid.2023.105833
Received 23 September 2022; Received in revised form 13 December 2022; Accepted 15 February 2023
Available online 17 February 2023

Nomenclature

Roman

a	Coefficient of the Hermite pol. exp.
$\boldsymbol{c}$, c	Lattice velocity, cumulant
c_s	Speed of sound
C_d	Drag coefficient
C_l	Lift coefficient
d	Number of dimensions
D	Diameter
f	Distribution function or frequency
f_{vs}	Vortex shedding frequency
$\boldsymbol{F}$, F	Force, Laplace transformation of f
$\mathcal{H}$	Hermite polynomial
L_r	Length of the recirculation bubble
m	Moment
$\mathbf{M}$	Moment transformation matrix
Ma	Mach number
n	General index
p	Pressure
q	Number of integration points
Re	Reynolds number
$\mathbf{S}$	Relaxation times matrix
St	Strouhal number
t	Time
T_{C}	Characteristic time
u	x-component of $\boldsymbol{V}$
v	y-component of $\boldsymbol{V}$
$\boldsymbol{V}$	Velocity
w	Weight associated with $\boldsymbol{c}$
$\boldsymbol{x}$	Position

Greek

α	General index
β	General index
γ	General index
ζ	Bulk viscosity
$\boldsymbol{\Xi}$	$\boldsymbol{c}$ in wave number space
ν	Kinematic viscosity
ρ	Density
σ	Blending coefficient
τ	Relaxation time
ω	Collision frequency
Ω	Collision operator

Superscript

$'$	Fluctuating quantity
-1	Inverse
eq	Equilibrium value
(n)	Order
neq	Non-equilibrium value
$\star$	Post-collision value

Subscript

0	Inflow quantity
i	Along ith direction
phy	In physical units
tot	Total quantity
tur	Turbulent quantity

(hydrodynamic moments). Contrary to the classical BGK scheme, the MRT allows having a variable bulk viscosity. For BGK the relation between kinematic and bulk viscosity [12] is locked *e.g.* in 3D $\zeta = 2/3\mu$. From the first mention of such an MRT model in the early nineties [8], many variants with improved stability properties have been proposed including different moment spaces (raw and central moments, cumulants) as well as different moment bases (orthogonal, tensor product, Hermite tensor product, etc.) [13–18]. An extensive overview is given by Coreixas et al. [19].

One of the main drawbacks of the MRT is the lack of a universal choice of relaxation parameters. The parameters related to non-hydrodynamics moments do not influence the physics of the problem as long as the continuum hypothesis is valid. However, they influence the stability and accuracy of the scheme. In some cases, these parameters are modified to improve the accuracy of a specific boundary condition [10,11]. Often an individual set of parameters only works well for a specific problem. To overcome this issue, Geier et al. [20] derived a formulation of the collision operator based on statistically independent observable quantities, the cumulants. The Cumulant scheme allows for curing the velocity-dependent errors in the transport coefficients, granting the Cumulant the tag of a Galilean-invariant scheme. This is not true in the case of the MRT scheme, where the moments are relaxed with different frequencies in a static reference frame [15].

Another stability-enhancing strategy introduces a regularization step prior to a BGK collision in function space. Similar to the MRT models, regularization attenuates the so-called ghost modes, which result from a disparity between the number of populations and the number of hydrodynamics moments. Instead of over-relaxing these ghost modes in moment space, the distribution functions are projected on Hermite polynomials.[1] This idea was first proposed by Latt and Chopard [21]. Improvements include the use of a recursive relation to construct the higher-order off-equilibrium moments [22] as well as a hybrid approach, where the second-order off-equilibrium moment (strain rate tensor) is reconstructed in a hybrid fashion using central differences [17]. It is worth mentioning that although the recursive regularization of the distribution functions, *i.e.* the inclusion of high-order velocity terms, helps to reduce the Galilean invariance defect [23], it introduces new non-hydrodynamic modes that might cause instabilities. The hybrid approach mitigates the defect, but does not cure it in a permanent fashion [18].

When the strain rate tensor is solely reconstructed by central differences, the model is referred to as Projected Recursive Regularized (PRR) [24] and constitutes one of the most dissipative variants of collision. As multifaceted as the possibilities are to model the collision process, little has been done so far to compare the different approaches and provide guidelines for their use. Only recently, Coreixas et al. [19] presented a consistent theoretical framework for existing collision models.

In addition, some numerical comparisons have been carried out. For instance, Coreixas et al. [25] compared different collision schemes on the D2Q9 stencil. They concluded that the Recursive Regularized BGK and the Cumulant are the best choices for weakly-compressible flows discretized in velocity space with the D2Q9 stencil. Nathen et al. [26] carried out a comparative study for direct numerical simulations of the Taylor Green vortex and the turbulent channel flow on a D3Q19 stencil.

[1] It should nevertheless be noted that regularization can be expressed as an MRT model with a Hermite tensor product basis, where the non-hydrodynamic moments are relaxed to their equilibrium value.

Table 1
The three-dimensional quadrature stencils commonly used to discretize the BE in the velocity space.

Stencil	i	$\boldsymbol{c}_i$	w_i	c_s
$D3Q19$	0	$(0,0,0)$	1/3	$1/\sqrt{3}$
	1 – 6	$(\pm1,0,0),(0,\pm1,0),(0,0,\pm1)$	1/18	
	7 – 18	$(\pm1,\pm1,0),(\pm1,0,\pm1),(0,\pm1,\pm1)$	1/36	
$D3Q27$	0	$(0,0,0)$	8/27	$1/\sqrt{3}$
	1 – 6	$(\pm1,0,0),(0,\pm1,0),(0,0,\pm1)$	2/27	
	7 – 18	$(\pm1,\pm1,0),(\pm1,0,\pm1),(0,\pm1,\pm1)$	1/54	
	19 – 26	$(\pm1,\pm1,\pm1)$	1/216	

They observed that the MRT scheme was not able to provide good convergence behavior. In particular, the results obtained with MRT on fine meshes produced nonphysical states. A similar comparison was conducted by Haussmann et al. [27] for the same stencil. To the best of the authors' knowledge, the literature lacks a comparison of these collision models for LES simulations in 3D. This lack of comparison was also pointed out by Coreixas et al. [25].

For this reason, we conducted a numerical comparison in 3D of two prominent, but conceptually different models: the parametrized Cumulant [28] and the hybrid recursive regularized BGK model. For the sake of comparison, we also showcase the results obtained with the PRR and MRT models. These models were implemented in the open-source LBM solver Musubi [29] and compared to experimental data provided by Parnaudeau et al. [30]. Besides their different nature of dealing with ghost modes, another crucial difference lies in the modeling of turbulence. While the Cumulant model, according to Geier et al. [31], intrinsically accounts for subgrid scale (SGS) turbulence, the HRR, PRR, and MRT models require the computation of a turbulent viscosity. Here three different SGS models are tested: Smagorinksy [32], Wall-Adapting Local Eddy-viscosity (WALE) [33], and Vreman [34].

The paper is organized as follows: Section 2 briefly presents the LBM framework and the collision models that are compared in this study. This is followed by a presentation of the LBM solver Musubi 3. The next two sections describe the results in terms of accuracy 4 and performance 5 for the considered collision models. A conclusion on the here presented comparison is provided in Section 6.

2. Lattice Boltzmann method

This section describes the lattice Boltzmann method and its equations. LBM aims to model fluid flows based on the Boltzmann equation (BE) [5]

$$\frac{\partial f}{\partial t} + \boldsymbol{c}\cdot\boldsymbol{\nabla} f + \frac{\boldsymbol{F}\cdot\boldsymbol{\nabla} f}{\rho} = \Omega(f). \tag{1}$$

The BE is discretized in space, time, and velocity space to obtain the force-free lattice Boltzmann equation (LBE) [5]

$$f_i(\boldsymbol{x}+\boldsymbol{c}_i\Delta t, t+\Delta t) = f_i(\boldsymbol{x},t) + \Delta t\Omega(f_i). \tag{2}$$

The quadrature stencil used to discretize the velocity space is indicated as $DdQq$, where d is the number of dimensions and q is the number of integration points. The commonly used stencils for three dimensions are $D3Q19$ and $D3Q27$, which are detailed in Table 1.

The original collision operator Ω (proposed by Boltzmann) is a nonlinear term, which can be simplified to the Bhatnagar–Gross–Krook (BGK) model [7], also referred to as the single relaxation time (SRT) model. Then Eq. (2) takes the following form in lattice units ($\Delta x = \Delta t = 1$)

$$f_i(\boldsymbol{x}+\boldsymbol{c}_i, t+1) = f_i(\boldsymbol{x},t) + \omega\left[f_i^{\mathrm{eq}}(\boldsymbol{x},t) - f_i(\boldsymbol{x},t)\right], \tag{3}$$

with ω being the relaxation frequency defined as

$$\omega = \frac{1}{\tau}, \tag{4}$$

where τ is the numerical relaxation time in lattice units. For standard lattices in three (and two) dimensions τ may be initialized as

$$\tau = 0.5 + \frac{3\nu_{\mathrm{phy}}}{\Delta x_{\mathrm{phy}} c_{s,\mathrm{phy}}}, \tag{5}$$

where ν_{phy} is the kinematic viscosity [35].

The two principal steps of an LBM algorithm are collision and streaming. Accordingly Eq. (3) is divided into a local collision (right-hand side) and a subsequent streaming step. Labeling the post-collision distribution functions as $f^\star$, Eq. (3) may then be presented as follows

$$f_i^\star(\boldsymbol{x},t) = f_i(\boldsymbol{x},t) + \omega\left[f_i^{\mathrm{eq}}(\boldsymbol{x},t) - f_i(\boldsymbol{x},t)\right], \tag{6}$$

$$f_i(\boldsymbol{x}+\boldsymbol{c}_i, t+1) = f_i^\star(\boldsymbol{x},t). \tag{7}$$

To recover the isothermal, weakly-compressible Navier–Stokes equations, the truncated equilibrium distribution function has to be expanded up to at least second-order, *i.e.*

$$f_i^{\mathrm{eq}} = w_i\rho\left[1 + \frac{\boldsymbol{c}_i\cdot\boldsymbol{V}}{c_s^2} + 0.5\frac{(\boldsymbol{c}_i\cdot\boldsymbol{V})^2}{c_s^4} - 0.5\frac{\boldsymbol{V}\cdot\boldsymbol{V}}{c_s^2} + \mathcal{O}(\mathrm{Ma}^3)\right]. \tag{8}$$

The macroscopic quantities are then retrieved (in lattice units) from the distribution functions as velocity moments such that

$$\rho = \sum_{i=0}^{q-1} f_i, \tag{9}$$

$$\boldsymbol{V} = \sum_{i=0}^{q-1} f_i\boldsymbol{c}_i, \tag{10}$$

$$p = c_s^2\rho. \tag{11}$$

2.1. Hybrid recursive regularized BGK

The hybrid recursive regularized scheme (HRR) is an advanced collision scheme based on the LBGK equation (3). Following the convention in literature to express the distribution function as an infinite sum of Hermite polynomials [18], the collision (Eq. (6)) is rewritten as

$$f_i^\star(\boldsymbol{x},t) = f_i^{\mathrm{eq}}(\boldsymbol{x},t) + (1-\omega)f_i^{\mathrm{neq}}(\boldsymbol{x},t), \tag{12}$$

with

$$f_i^{\mathrm{eq}}(\boldsymbol{x},t) = w_i\sum_{n=0}^{N}\frac{1}{n!}\boldsymbol{a}_0^{(n)} : \boldsymbol{\mathcal{H}}_i^{(n)}, \tag{13}$$

$$f_i^{\mathrm{neq}}(\boldsymbol{x},t) = w_i\sum_{n=2}^{N}\frac{1}{n!}\boldsymbol{a}_1^{(n)} : \boldsymbol{\mathcal{H}}_i^{(n)}. \tag{14}$$

The Hermite polynomials $\boldsymbol{\mathcal{H}}_i^{(n)}$ as well as the expression for the coefficients $\boldsymbol{a}_0^{(n)}, \boldsymbol{a}_1^{(n)}$ in the above expansions are given in [18]. In his PhD thesis, Coreixas [24] studied which coefficients are properly represented on a given stencil $DdQq$. The interested reader is invited to consult Appendix D of this thesis.

In order to increase the numerical stability, the second-order Hermite polynomial coefficient of f^{neq} is blended as follows

$$\boldsymbol{a}_1^{(2,\mathrm{HRR})} = \sigma\boldsymbol{a}_1^{(2)} + (1-\sigma)\boldsymbol{a}_1^{(2,\mathrm{PRR})}, \tag{15}$$

where $0 \le \sigma \le 1$. While the first expression of $\boldsymbol{a}_1^{(2)}$ on the RHS is the second-order velocity moment of f^{neq}, *i.e.*

$$a_{1,\alpha\beta}^{(2)} = \sum_{i=0}^{q-1} c_{i,\alpha}c_{i,\beta}f_i^{\mathrm{neq}}, \tag{16}$$

the second expression is derived from a Chapman–Enskog analysis of the standard LBGK equation up to second-order in the Knudsen number [12]

$$a_{1,\alpha\beta}^{(2),\mathrm{PRR}} = -\rho c_s^2\tau\left(\frac{\partial V_\beta}{\partial x_\alpha} + \frac{\partial V_\alpha}{\partial x_\beta}\right). \tag{17}$$

Evaluating the velocity gradient with a finite difference scheme [17] introduces a form of hyperviscosity that damps numerical instabilities near the Nyquist limit (checkerboard instability).

Higher order Hermite polynomial coefficients ($n > 2$) in Eq. (14) are computed via a recursive relation derived by Malaspinas [22] using the equilibrium coefficients $\boldsymbol{a}_0^{(n)}$. For the explicit formulations, the interested reader may refer to Appendix G in [24].

2.2. Projected Recursive Regularized BGK

The projected recursive regularized (PRR) [24] scheme is obtained from the HRR collision model. One needs to set σ to zero in Eq. (15) to obtain the following equation

$$\boldsymbol{a}_1^{(2,\mathrm{HRR})} = \boldsymbol{a}_1^{(2,\mathrm{PRR})}, \tag{18}$$

where $\boldsymbol{a}_1^{(2,\mathrm{PRR})}$ is evaluated as in Eq. (17). This model is stable, but highly dissipative due to the finite-difference formulation of the Hermite expansion coefficients.

2.3. Multi relaxation time

The multi-relaxation time (MRT) scheme was first introduced by d'Humières [8] in 1992. The main idea is to collide in moment space rather than velocity space. Therefore, Eq. (2) is transformed before the collision such that

$$f_i(\boldsymbol{x} + \boldsymbol{c}_i \Delta t, t + \Delta t) = f_i(\boldsymbol{x}, t) + \mathbf{M}^{-1}\mathbf{S}\left[m_i^{\mathrm{eq}}(\boldsymbol{x}, t) - m_i(\boldsymbol{x}, t)\right], \tag{19}$$

where $\boldsymbol{m} = \mathbf{M}\boldsymbol{f}$. $\boldsymbol{m}$ is a $q \times 1$ vector containing (amongst others) the relevant hydrodynamic moments and $\mathbf{M}$ is a $q \times q$ transformation matrix with q being the number of velocity directions of the lattice. The original transformation matrix had an orthogonal basis, but alternative bases such as simple tensor products and Hermite tensor products are possible as well [19]. The diagonal matrix $\mathbf{S}$ may take on different relaxation times for each moment m_i. Nevertheless, the frequencies related to the hydrodynamic moments are constrained to properly reproduce the Navier–Stokes dynamics. Non-physical modes may then be filtered explicitly by changing, *i.e.* increasing, the relaxation time associated with them. Different formulations for $\mathbf{M}$ with their suggested values for the relaxation matrix $\mathbf{S}$ are available [9,36,37].

2.4. Cumulant

The Cumulant collision model was introduced by Geier et al. in 2015 [20]. The MRT collision model has several drawbacks, such as a lack of a universal formulation of optimal collision rates, deficiency stemming from an arbitrary choice of moment space, and being non-Galilean invariant.[2] The Cumulant model was developed to overcome some of these drawbacks. This is achieved by using a formulation based on statistically independent observable quantities of the distribution functions, the cumulants $c_{\alpha\beta\gamma}$.

F is defined as the two-sided Laplace transformation of f, *i.e.*

$$F(\boldsymbol{\Xi}) = \mathcal{L}[f(\boldsymbol{c})] = \int_{-\infty}^{\infty} f(\boldsymbol{c}) \exp^{-\boldsymbol{\Xi}\cdot\boldsymbol{c}} \,\mathrm{d}\boldsymbol{c}, \tag{20}$$

where $\boldsymbol{\Xi} = (\Xi, Y, Z)$ is the particle velocity $\boldsymbol{c}$ in wave number space. Then, the cumulants are obtained as follows:

$$c_{\alpha\beta\gamma} = c^{-\alpha-\beta-\gamma} \frac{\partial^\alpha \partial^\beta \partial^\gamma}{\partial \Xi^\alpha \partial Y^\beta \partial Z^\gamma} \ln\left[F(\Xi, Y, Z)\right]\Big|_{\Xi=Y=Z=0}, \tag{21}$$

where $c^{-\alpha-\beta-\gamma}$ is the lattice velocity. Finally the cumulants are relaxed as follows

$$c^\star_{\alpha\beta\gamma} = c_{\alpha\beta\gamma} + \omega_{\alpha\beta\gamma}\left(c^{\mathrm{eq}}_{\alpha\beta\gamma} - c_{\alpha\beta\gamma}\right). \tag{22}$$

The Cumulant model solely works on the $D3Q27$ stencil and contains 27 moments or rather cumulants and therefore 27 possible collision frequencies. It is demonstrated in [38] that to attain the rotational invariance property some relaxation rates have to be identical, reducing the degrees of freedom to 10, i.e. $\omega_1, \omega_2, \ldots, \omega_{10}$. Only the first of these frequencies is related to the physical (kinematic) viscosity as in Eq. (5). Geier et al. [20] suggest to set $\omega_{2\text{–}10} = 1$.

In a later publication, Geier et al. [28] parametrized the Cumulant model to obtain a fourth-order accurate diffusion term in the Navier–Stokes equation. A limitation, which – at the same time – is beneficial to industrial applications is a constraint on the kinematic viscosity to be small. Appendix B of [28] introduces a further correction to the parametrized Cumulant model to obtain a fourth-order accurate advection operator. Nevertheless, the suggested correction has proven to be unstable at low viscosities and was therefore omitted in the present work. With this formulation ω_2 is chosen to set the bulk viscosity ζ of the flow [36]

$$\zeta_{\mathrm{phy}} = \frac{\Delta x^2_{\mathrm{phy}}}{\Delta t_{\mathrm{phy}}} \frac{5 - 9c^2_{s,\mathrm{lat}}}{9}\left(\omega_{2,\mathrm{lat}} - 0.5\right). \tag{23}$$

2.5. Boundary conditions

In this study, we use the non-equilibrium extrapolated boundary condition (BC) introduced by Guo et al. [39]. This BC is used to mimic the Dirichlet kind, where we assign a velocity at the inlet and a pressure at the outlet. The missing variables are extrapolated from inside the domain. Being an extrapolation method means that all distribution functions at the boundary nodes need to be reconstructed [5].

The Bouzidi BC is used to mimic a curved no-slip wall [40]. The distribution function of the links being crossed by the curved solid boundary is reconstructed via a combination of the bounce-back rule and a linear interpolation of neighboring values.

2.6. Subgrid scale modeling

Unresolved turbulent scales are here accounted for by an eddy-viscosity model adopted from the classical Navier–Stokes framework. The total viscosity of the flow is modified locally by adding a turbulent viscosity term ν_{tur} to the kinematic viscosity of the fluid ν_{phy}, *i.e.*

$$\nu_{\mathrm{tot}} = \nu_{\mathrm{phy}} + \nu_{\mathrm{tur}}. \tag{24}$$

Then ω or rather τ is evaluated via Eq. (5). An extensive introduction to the most frequently used turbulence models is given in chapter 2 of [41]. For this study, we adopt the standard Smagorinsky model [32], the Wall-Adapting Local Eddy-viscosity (WALE) model [33], and the Vreman model [34]. The coefficient of the Smagorinsky model is set to 0.17, while for the WALE model and the Vreman model it is set to 0.50 and 0.07, respectively.

The Cumulant model has an inherent numerical stabilization, therefore it does not require a turbulence model [31]. This collision scheme [28] relies on the time history of high-order cumulants. Geier et al. suggest a limiter for the third-order cumulants to reset the time history. It is worth mentioning that the error introduced by the limiters lies well below the leading error of the LBM. Thus, the fourth-order accuracy of the scheme is kept. The formula of $\omega_{3\text{–}5}$ is modified as in [28]. The limiters are set to a value of 10^{-2}.

[2] "Taking moments in a static frame of reference and relaxing them with different rates introduces violations of Galilean invariance not present in a single relaxation time model with the same velocity set and the same equilibrium function" [20].

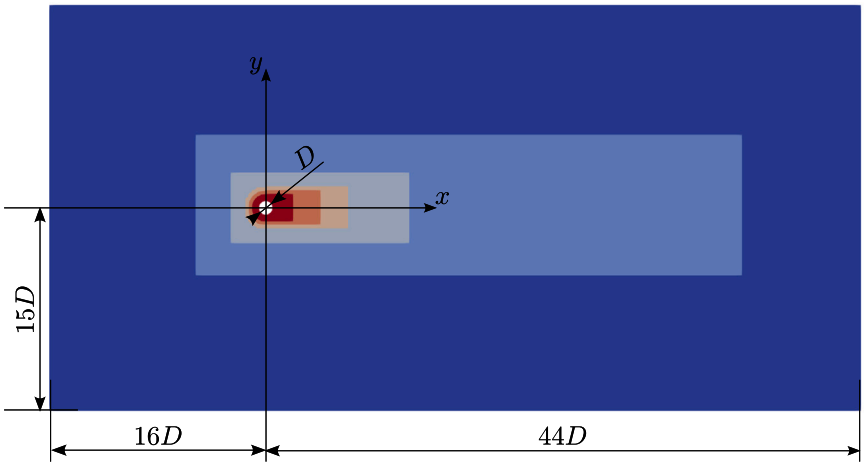

Fig. 1. Computational domain with its refinement levels, where red is the finest.

3. Musubi, a lattice Boltzmann solver

Musubi[3] [29] is an open-source lattice Boltzmann solver which is mostly written in Fortran 2003 and parallelized with the Message Passing Interface (MPI) [42]. Musubi offers a selection of different collision models, stencil layouts, and LES turbulence models, as well as the possibility to work on multi-level meshes. Since 2021, Musubi is adopted by the DLR and has been extended to support high Reynolds number flows in aerodynamic and aeroacoustic applications.

Musubi uses an octree-type mesh consisting of cubical cells generated by the mesh generator Seeder [43]. Seeder sorts the cells by depth-first order following a space-filling curve (Morton or Z curve), which maintains locality and reduces the communication between processors [44]. In the solver, the sorted list of cells is partitioned equally among processes, aiming to achieve optimal load balancing. However, in real applications, some cells are more computationally expensive than others due to grid refinements, boundary conditions, source terms, etc., resulting in load imbalances. This imbalance is resolved by deploying the load balancing algorithm SPartA [45], which redistributes the elements by using weights. These weights take into account the computational cost linked to each element. Musubi offers both, static and dynamic load balancing. Static load balancing can be used for fixed multi-level meshes. For instance, in a domain with one refinement level, fine cells are computationally two times more expensive than coarse cells due to the halved time-step. Therefore, the weights for the fine cells are doubled during mesh distribution. On the other hand, with dynamic load balancing, the weights of each cell are computed after every few iterations during runtime, and cells are redistributed according to the time spent on the different cells without the need of restarting the simulation.

The solver reads the sorted list of cells in a distributed manner, where each process requires minimal information about the mesh. Nevertheless, the explicit neighborhood information is necessary for the solver to pull the distribution function from the direct neighbor to the local cell according to the stencil definition (streaming step, as in Eq. (7)). A stencil is defined as a set of offset directions describing the relative position of the neighborhood of a cell. With the help of the space-filling curve and the topology of the octree mesh, a process can easily locate on which partition the neighbor of a cell is stored.

To achieve good performance with multi-level meshes, Musubi does computations on cells in a level-wise fashion. Thus, the sorted list of cells generated by the mesher is resorted by level in the solver. Then, ghost cells are used to interpolate values between different refinement levels. To decrease computational time state values are stored in a double buffer array at each level: one for reading and one for writing. Each buffer has a size of $N \times q$, where N is the number of elements in the level, and q is the number of velocity directions. Additionally, a connectivity array of the same size is stored for indirect addressing of adjacent neighbors in the state array. More arrays are used to store auxiliary values required for multi-level interpolation, computation of boundary conditions, source terms, and turbulent viscosities.

[3] The Musubi repository is publicly available at https://osdn.net/projects/apes/scm/hg/musubi/.

4. Results and discussions

To assess the accuracy and performance of the different collision models, we modeled the quasi 3D cylinder test-case at a Reynolds number of 3900, based on the cylinder diameter D. Although this is considered an academic test-case, it is commonly used to compare LES solvers as it is well documented in the literature [17,30,46,47].

Fig. 1 shows the computational domain with the cylinder located at the origin of the coordinate system. The rectangular domain stretches $60D$ in stream-wise (x-axis) and $30D$ in wall normal (y-axis) direction, respectively. Velocity inlet (left) and pressure outlet (right) boundary conditions are applied as mentioned in Section 2. The top and bottom sides of the domain are defined as slip boundaries. Periodic boundaries are used for the sides in spanwise direction (z-axis). The cylinder surface is modeled as a no-slip curved boundary according to [40]. To avoid any influence of the periodic boundaries on the flow field, the domain extends in spanwise direction by $4D$. Such a spanwise extent is compliant with the observations made by Breuer [48]. The inflow is a uniform velocity profile $\boldsymbol{V} = (u_0, 0, 0)$, with $u_0 = 20.077$ m s^{-1}. This leads to an inlet Mach number of $\mathrm{Ma}_0 = 5.850 \times 10^{-2}$. The pressure is fixed at the outflow and is obtained via Eq. (11), where $\rho_0 = 1.225$ kg m^{-3}.

To prevent spurious reflections at the boundaries that occur during the transient phase, both a rectangular-shaped viscous sponge layer and an absorbing layer [49] are placed on the boundaries in x and y directions. The width of the layers is set to $5D$ extending from the boundaries into the domain. The target quantities of the absorbing layer are set to $p = p_0$ and $\boldsymbol{V} = (u_0, 0, 0)$. The viscous sponge layer linearly increases the viscosity of the fluid to a factor of 10 at the boundaries. According to our simulations, this factor of 10 is not high enough to damp reflections coming from the inlet BC sufficiently when the Cumulant is used as collision scheme. To stabilize the Cumulant scheme, a numerical trial and error approach shows that a factor of 150 is required.

Five different uniform refinement regions are used to create a hierarchical Cartesian mesh of the domain (cf. Fig. 1). To properly resolve the boundary layer on the cylinder and the recirculating bubble

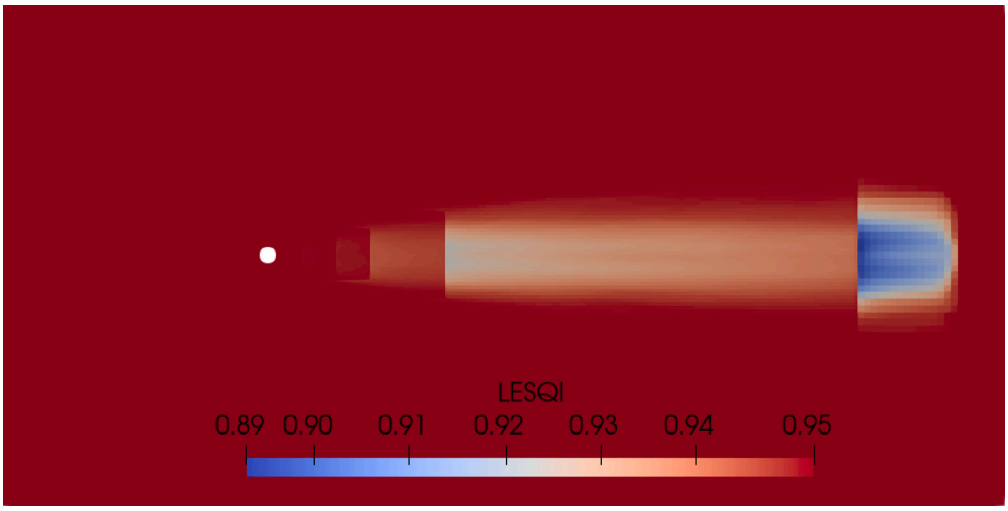

Fig. 2. Contour of the averaged LES quality index of Celik et al. [50] in the mid xy-plane for the combination of HRR with Vreman.

downstream of the cylinder, the diameter of the cylinder is discretized by 80 voxels in the finest region, leading to a maximum y^+ value of about 1 at 50° or rather 310°. The extents of the refinement regions are kept similar to the ones defined in Jacob et al. [17], albeit with a slight variation in the shape for the three finest levels. These regions (i.e. the three regions closest to the cylinder) have a radial offset of $D, 1.25D$, and $1.5D$ in the upstream direction and a rectangular offset of $2D, 4D$, and $6D$ in the wake direction, respectively. The other two refinement levels have dimensions of $[13D \times 5D]$ and $[40D \times 10D]$ and are located $2.5D$ and $5D$ away from the cylinder with respect to both the x-axis and the y-axis. This domain discretization results in approximately 12.7 million cells.

The statistics shown in the following are averaged in time with a total sampling time of $92\,T_C$ (≈20 vortex shedding periods), where $T_C = D/u_0$. The sampling is started after $t \geq 200\,T_C$. Finally, the statistics are spatially averaged in spanwise (z) direction.

The authors were not able to stabilize the MRT model to work on the $D3Q27$ stencil as suggested in [36]. For this reason, MRT results were only acquired for the $D3Q19$ stencil with a parametric setting as defined in [37]. On the other hand, the Cumulant scheme is only compliant with the $D3Q27$ stencil, hence results for the HRR and PRR schemes were also obtained on this stencil.

The quality of the LES solutions is computed according to the index of Celik et al. [50], as below

$$\mathrm{LESQI} = \frac{1}{1 + 0.05\left(1 + \nu_{\mathrm{tur}}/\nu_{\mathrm{phy}}\right)^{0.53}}. \tag{25}$$

An LES simulation is considered of good quality if LESQI ≥ 0.80. The same value of 0.80 is suggested by Pope [51] for a similar index. Fig. 2 shows the contour of the LES quality index of Celik et al. [50] in the mid-yz-plane for the combination of HRR with Vreman model. The minimum of the quality index is located in the wake in the coarsest patch. Noteworthy, the minimum is higher than 0.80, assuring the quality of the present LES result. The contours for the other combinations are similar and therefore omitted. Nevertheless, the minimum of the quality index for the considered combinations is reported in Table 2. The Cumulant scheme employs its ILES model. Thus, this quality index cannot be used for this scheme, $\nu_{\mathrm{tur}} = 0 \rightarrow \mathrm{LESQI} = 0.95$, which is the DNS limit. Nevertheless, the resolution of the finest region in the domain, located around the cylinder, is $\Delta x = 0.0125D$, which is lower than the one used by Parnaudeau et al. [30] ($\Delta x = 0.0210D$) and the same as the one used by Jacob et al. [17]. In general, this resolution is also in range with other LES studies such as Alkishriwi et al. [46] ($\Delta x = 0.0130D$) and Ouvrard et al. [47] ($\Delta x = 0.0170D$). Taking into account that the resolution in the finest mesh is appropriate to discretize the boundary layer, *i.g.* $y^+ \leq 1$, and that the quality index is higher than 0.94 in the region of interest, the mesh resolution is sufficiently high for this test-case.

Table 2
The minimum of the LES quality index of Celik et al. [50] for considered combinations of collision schemes and turbulence models.

Combination	HRR			MRT			PRR
	Smag.	Vrem.	WALE	Smag.	Vrem.	WALE	Vrem.
min(LESQI)	0.89	0.89	0.86	0.88	0.89	0.86	0.93

A first comparison is carried out between HRR, PRR, MRT, and Cumulant, where the first three collision models use the Vreman turbulence model, while the Cumulant uses the implicit subgrid scale model introduced in Section 2.6. Fig. 3 shows the average of the normalized x-velocity u (top), and the normalized Reynolds stress $u'u'/u_0^2$ (bottom) along the wake center line. The value of the maximum velocity deficit in the recirculation bubble is predicted accurately by the Cumulant model although the maximum is reached further downstream compared to the experiments. However, the length of the recirculation bubble agrees well with the experimental data. The HRR predicts well the maximum deficit and shows a slight overestimation of the downstream position where this occurs. For the MRT and PRR this trend is more pronounced so that it clearly deviates from the experimental data. Noteworthy is the recovery of the velocity in the far wake observed for all collision models. This implies that the domain size and the boundary conditions are adequately chosen to model this problem.

The trend of the Reynolds stress $u'u'$ along the streamwise direction, obtained by the experiment, shows two peaks in the recirculation bubble. Although both peaks are present in the numerical results, they do not match the experimental data correctly. The magnitude of the first and second peak is severely underestimated by the PRR. Only the Cumulant shows a fair agreement with respect to the position of these peaks. The other collision schemes shift the position of both peaks downstream. The MRT fails to predict the experimental trend when $x/D > 2$. The difference between the behavior of the schemes can be attributed to the physical assumptions of the models. The MRT collision matrix is generated from a moment base, which neglects high-order terms ($n > 3$) of velocity. The other schemes are generated by taking into account high-order terms of velocity (up to 6th-order) [19]. Therefore, as a natural consequence, the Cumulant and HRR have a reduced velocity-dependent error in the viscous stress tensor. Although the PRR has the same characteristic as HRR, its dissipative nature strongly affects the computation of the Reynolds stresses, leading to an underestimation.

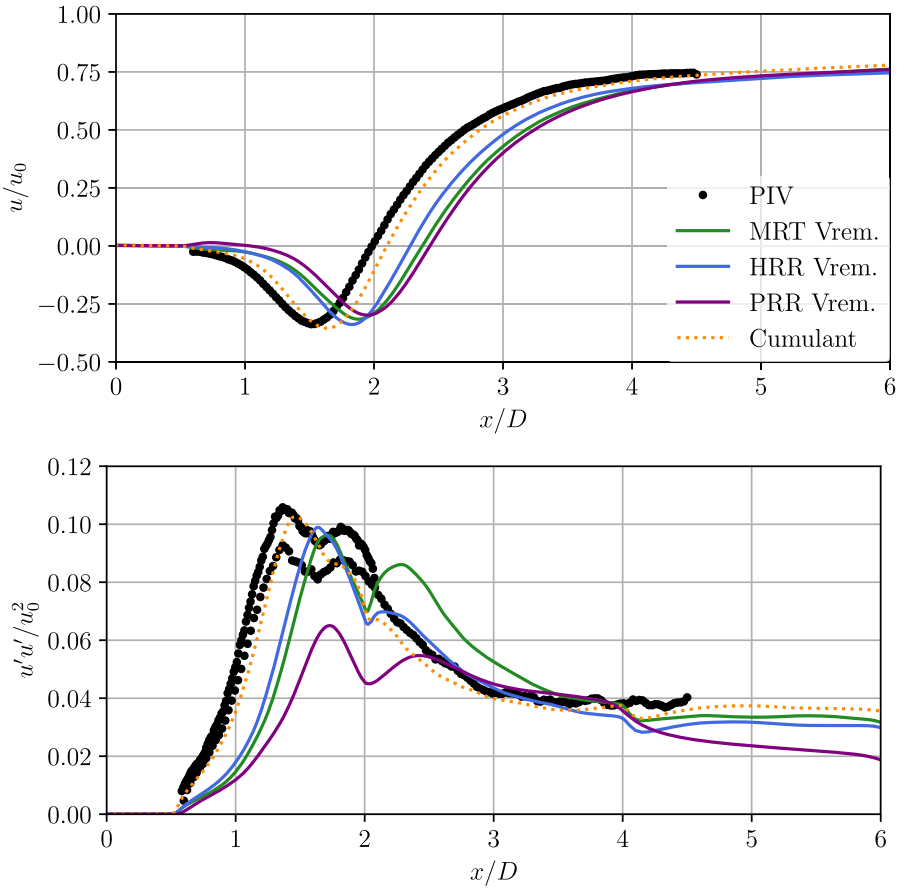

Fig. 3. Normalized average of the x-velocity u (top), and normalized Reynolds stress $u'u'$ (bottom) along the wake center line.

The same observation can justify the difference in the trends of the Reynolds stresses $u'u'$ in the yz plane at different positions downstream of the cylinder (cf. Fig. 4). At $1.06D$ downstream of the cylinder, all schemes capture the location of the peaks properly, but underestimate their magnitude, with PRR showing the worst and Cumulant showing the best agreement. The same behavior can be seen further downstream ($1.54D$), where the Cumulant is the only model that accurately predicts the magnitude of the peaks. At $2.02D$ all models but the PRR provide satisfactory results.

The next comparison is carried out to assess the impact of different turbulence models. Both, the MRT and the HRR schemes are coupled with the Smagorinsky, WALE, and Vreman subgrid model, respectively. Given the poor agreement of the results obtained with PRR compared to the experimental data, this collision model is excluded from the following comparisons. Fig. 5 shows the normalized average of the x-velocity u (top), and normalized Reynolds stress $u'u'/u_0^2$ (bottom) along the wake center line. It can be noticed that the combinations of collision and turbulence models that show the best agreement with the PIV data are MRT with WALE and HRR with Vreman. The minimum of the mean x-velocity is underpredicted by MRT with WALE, while the HRR with Vreman agrees well with the experimental data. The length of the recirculation bubble is well overpredicted by the MRT with Smagorinsky and HRR with WALE, while MRT with WALE shows the best agreement.

Regarding the evolution of the Reynolds stress $u'u'$ along the wake center line (Fig. 5 (bottom)), one can immediately observe that all the proposed combinations fail to predict the location of both, the first and second peak of the experimental trend, which is shifted further downstream. The magnitude of the first peak is in fair agreement with the experimental results for all combinations. The HRR with Vreman fails to predict the magnitude of the second peak, while the other combinations show a fair comparison. Another glaring difference is observed right downstream of the cylinder. The results obtained by the combinations of HRR with WALE and MRT with Smagorinsky show more discrepancy from the experimental result compared to other combinations. Noteworthy is the accurate prediction of $u'u'$ after $x/D > 2$ shown by the HRR with Vreman.

By analyzing Fig. 6, one can notice that in the near wake ($1.06D$) all models fail to correctly predict the Reynolds stresses. The combination of MRT with WALE shows the best agreement here, nevertheless the discrepancy with the experimental result is still noticeable. At the intermediate distance ($1.54D$), stresses are still under-predicted, whereas MRT with WALE shows the closest agreement with the experimental data. While both combinations of HRR with WALE and MRT with Smagorinsky show the farthest agreement. At the last downstream position ($2.02D$), all models show a satisfactory agreement with the Reynolds stresses.

Fig. 7 shows the ratio between the average eddy-viscosity and the physical viscosity of the fluid. The Vreman model generates the least amount of dissipation, followed by the WALE and Smagorinsky models. The Smagorinsky model contributes to the dissipation even in the finest patch of the mesh, $x/D \in [0, 2]$. However, the contribution of the other two models is nearly equal to zero. All the considered turbulence models rely on the gradient of velocity to estimate the eddy-viscosity. The HRR already uses this gradient during the collision step to generate dissipation and stabilize the scheme, see Eq. (17). For the modeled test-case, the HRR works well in conjunction with the Vreman model, which adds the least amount of dissipation among the considered models, cf. Fig. 7. A comparable plot is obtained for the MRT scheme, but it is omitted here for brevity. As the present study shows, the MRT delivers satisfactory results in conjunction with the WALE model.

Fig. 8 shows the normalized average of the x-velocity u (top), and normalized Reynolds stress $u'u'/u_0^2$ (bottom) along the wake center line

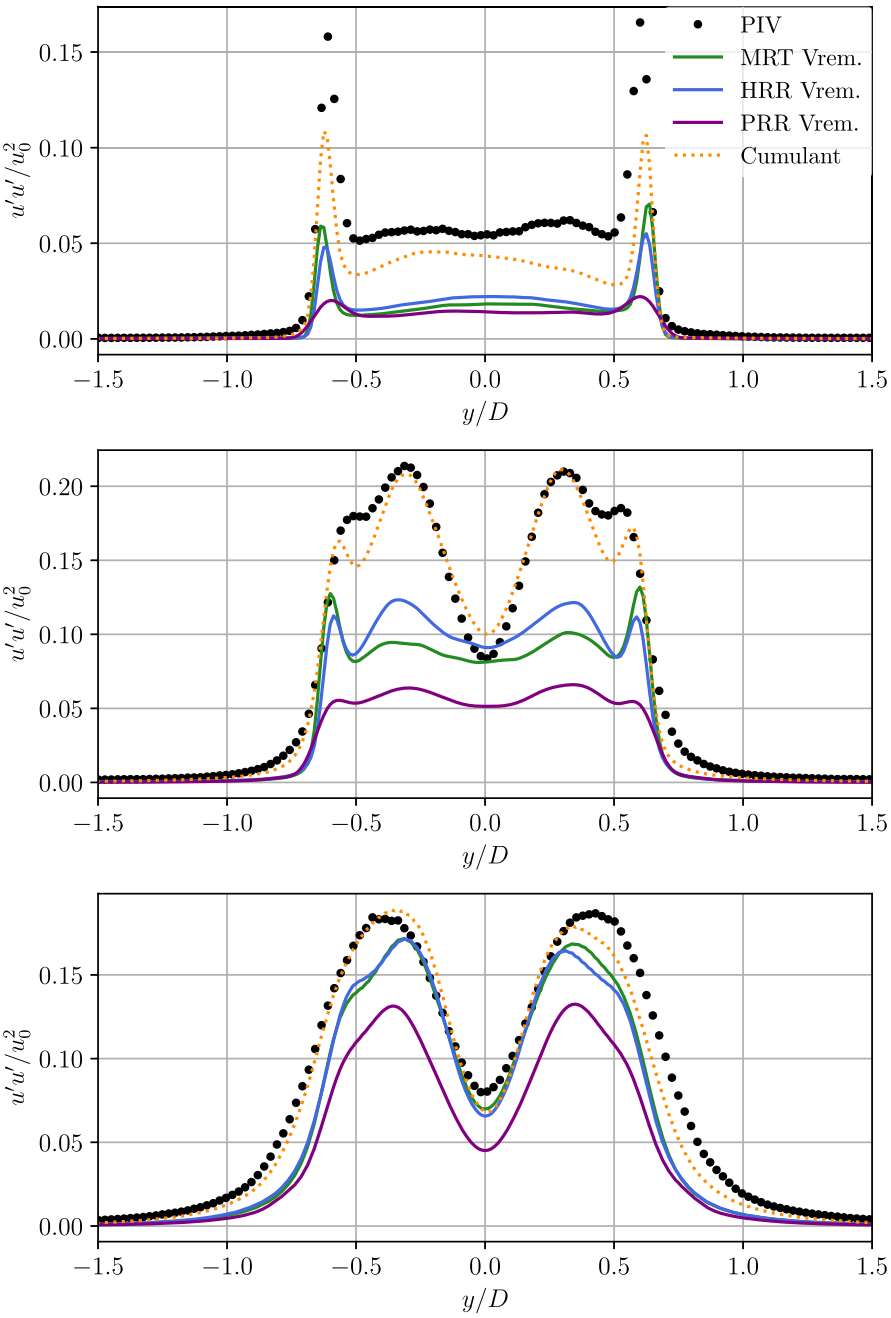

Fig. 4. Vertical profile of the normalized Reynolds stress $u'u'$ at $x = 1.06D$ (top), $x = 1.54D$ (middle), and $x = 2.02D$ (bottom).

for Cumulant with ILES, MRT with WALE, and HRR with Vreman. Compared to Fig. 3 only the MRT data is different. It can be seen that the WALE turbulence model leads to a higher velocity deficit but also a smaller recirculation bubble. One can notice that the result obtained by the MRT is located between the ones obtained by the HRR and Cumulant. The MRT model achieves a good prediction of the Reynolds stress but differs from the experimental trend in the interval $2 < x/D < 3$. Fig. 9 illustrates that the Cumulant model captures the Reynolds stresses in the detached shear layers (peaks) with the least deviation from the PIV data, followed by the combinations of MRT with WALE and HRR with Vreman. Noteworthy, the MRT with WALE approximates the peaks at the station $x/D = 1.06$ better than the Cumulant.

It is worth mentioning that while the MRT scheme is already written in a matrix form ($M^{-1}SM$), the HRR collision operator can be written in a matrix form as well. The former is generated by an orthogonal raw moment basis, while the latter is by a non-orthogonal Hermite tensor basis. On the other hand, this observation is not valid for the Cumulant scheme, which therefore generates results different than those obtained with HRR and MRT. Another important difference is the value of the bulk viscosity. For the HRR scheme, this value is fixed to $2/3\mu_{\text{phy}}$. The same value is used in the present work to evaluate the relaxation frequency which relaxes the kinetic energy moment of the MRT scheme. Regarding the parametrized Cumulant, the considered value would generate an error in the evaluation of ω_{3-5} attempting a division by zero [28]. Therefore, a value of $2\mu_{\text{phy}}$ is used to avoid computational error when evaluating the parametrized ω_{3-5}. In general, the effects of the bulk viscosity should be minimal, and mostly affects the sound propagation, which is not in the scope of this paper.

A quantitative comparison is shown in Table 3, where the force coefficients, the Strouhal number (St), and the length of the recirculation bubble L_r are compared with results available in the literature [17,30]. The results of Parnaudeau et al. [30] are experimental and only provide

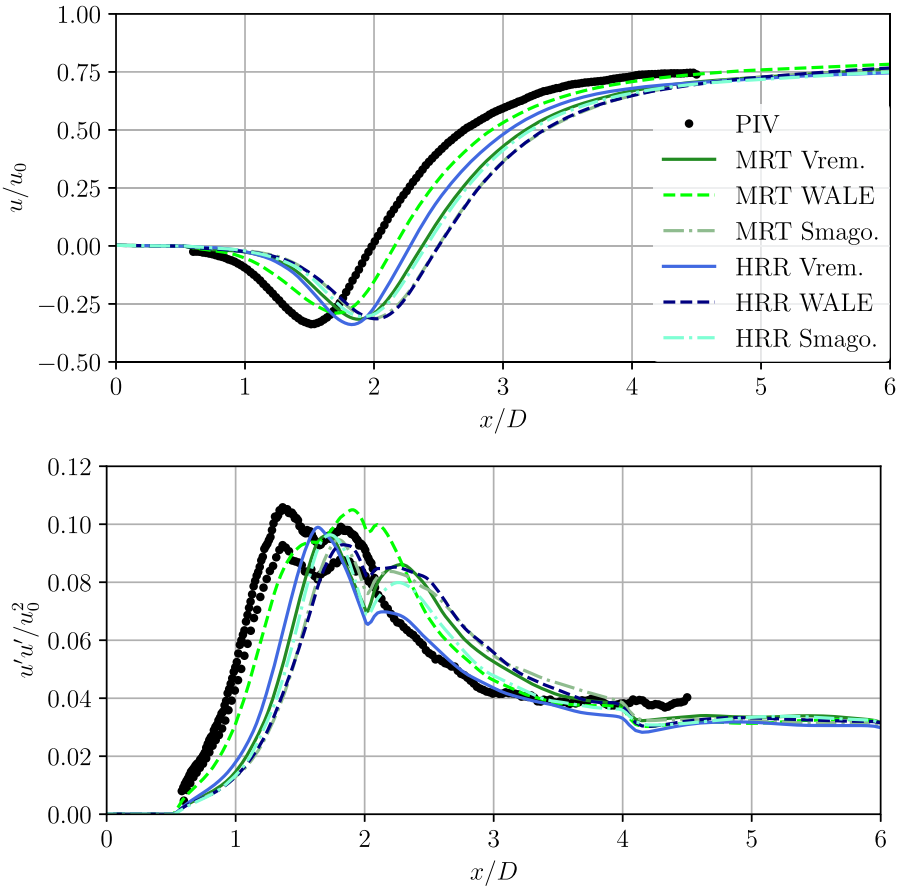

Fig. 5. Normalized average of the x-velocity u (top), and normalized Reynolds stress $u'u'$ (bottom) along the wake center line for different collision schemes and subgrid scale turbulence models.

Table 3
Comparison of the force coefficients, the Strouhal number and length of the recirculation bubble obtained by Musubi with the reference results [17,30]. Here C_l' indicates the root mean square of the lift coefficient.

Model	C_d	C_l'	St(v)	L_r/D
PIV [30]	–	–	0.208	1.51
ProLB (case 3) [17]	0.954	0.048	0.209	2.04
ProLB (case 4) [17]	1.047	0.165	0.212	1.43
Cumulant	1.073	0.335	0.212	1.60
HRR and Vreman	0.954	0.218	0.222	1.80
MRT and WALE	1.017	0.309	0.217	1.68

data for St and L_r. The results of Jacob et al. [17] are obtained with ProLB, an LBM based, commercial solver. The results of case 3 are obtained via a combination of HRR with Vreman model on a D3Q19 stencil. While, for the results of case 4, Jacob et al. utilize an implicit LES model instead of the Vreman model. At a first glance, we observe that all results obtained with Musubi are comparable with the reference data. The length of the recirculation bubble is generally overpredicted when compared to the PIV result, while the Cumulant shows the best agreement. This means that the transition to turbulence in the separated shear layer is delayed in the numerical simulations. The Strouhal number is best captured by the Cumulant model. With regard to the force coefficients, all models fail to predict the value obtained with ProLB case 4, while Cumulant and HRR with Vreman show the least deviation for C_d and C_l, respectively.

Fig. 10 presents the iso-contour of the normalized Q-criterion ($Q = 1$) colored by the normalized velocity magnitude. The snapshots show the Cumulant model with ILES (top-left), the HRR model with Vreman (top-right), and the MRT model with WALE (bottom). It can be observed that the flow is qualitatively well predicted in all cases. It is laminar around the cylinder, the separation of the boundary layer occurs in the laminar regime, and the transition to turbulent flow occurs in the wake. One can also notice the presence of large-scale and small-scale structures in the wake, which is typical of this flow regime. The difference in the length of the bubble is too subtle to be seen here.

The power spectra density of the y-component of the velocity at $(x, y, z) = (3.0D, 0.0, 0.0)$ is shown in Fig. 11. The sampling is done at each $10\Delta t$, for t in the range of 200–$300T_C$. One can see that the main peak is accurately predicted at $f = f_{vs}$. Experimentally, it was shown that there is a second peak at $f = 3f_{vs}$ [30]. This peak is also found in the numerical results. While Cumulant accurately predicts the peak at $f = 3f_{vs}$, the combinations of MRT with WALE and HRR with Vreman slightly shift its location to the right and to the left of $3f_{vs}$, respectively (see zoom-in in Fig. 11).

5. Performance

In this study, we performed computations on the CARA cluster of DLR. Each node is equipped with two AMD EPYC 7601 (32 cores; 2,2 GHz). Therefore, each node has 64 physical cores. The performance analysis is done for the HRR and PRR models on both, the $D3Q19$ and the $D3Q27$, for the MRT model on the $D3Q19$, and for the Cumulant scheme on the $D3Q27$ lattice stencil. All collision models but the Cumulant are coupled with the Vreman turbulence model. The Cumulant uses an implicit LES model. For the sake of clarification, the collision routines are optimized as follows.

MRT: vector matrix multiplication is explicitly written for all q directions. The null entries are omitted.

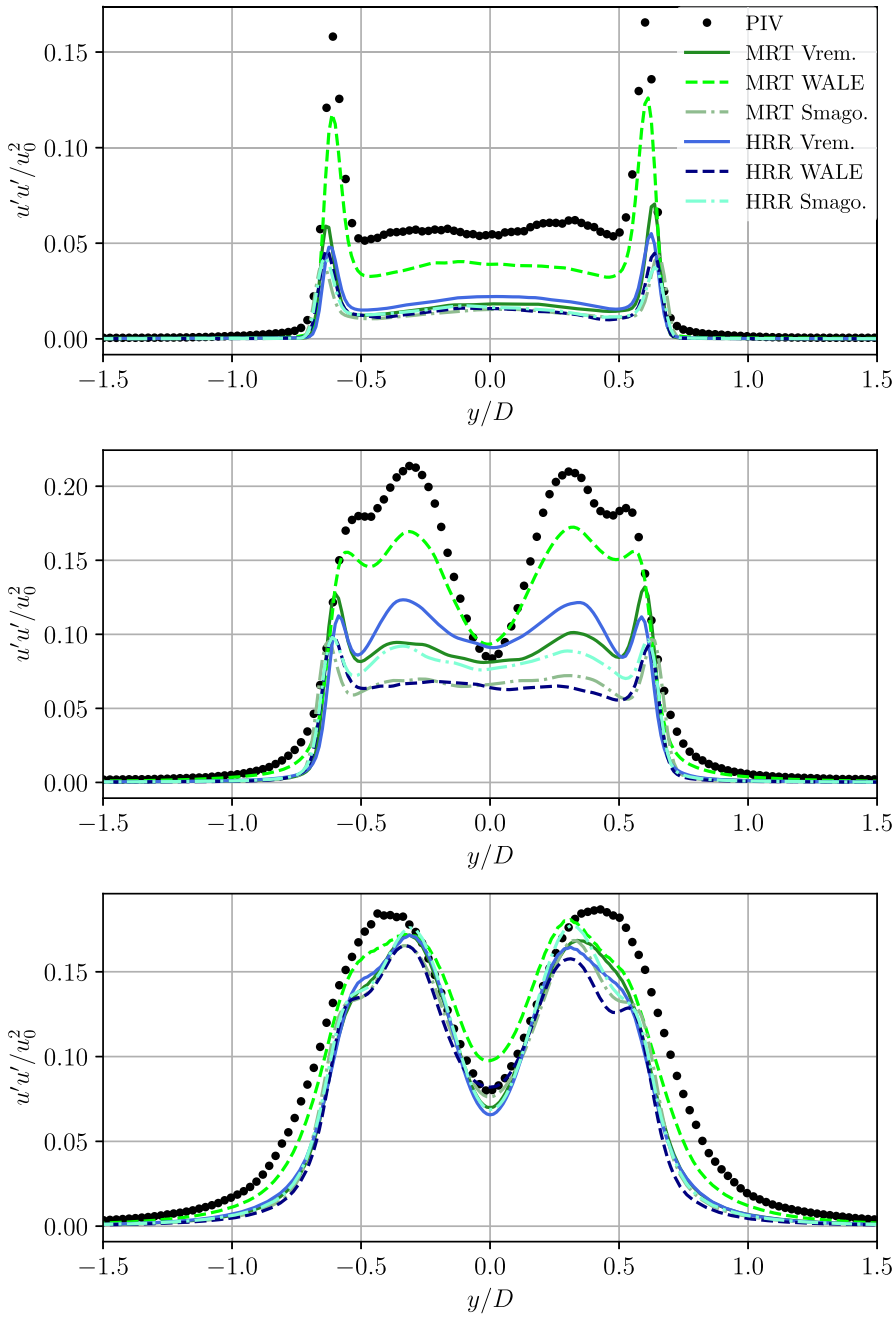

Fig. 6. Vertical profile of the normalized Reynolds stress $u'u'$ at $x = 1.06D$ (top), $x = 1.54D$ (middle), and $x = 2.02D$ (bottom).

HRR: the Hermite polynomials are explicitly computed and stored in memory during initialization rather than being calculated during run-time.

Cumulant: the weight coefficients are explicitly computed and stored in memory during initialization rather than being calculated during run-time. Furthermore, the fast Chimera transformation is employed as suggested by Geier et al. [28].

Although the following performance analysis is done with the same test-case as introduced before, the entire domain is discretized with $\Delta x_{\text{min}} = D/80$ leading to a uniform mesh of about 57 million cells.

Fig. 12 compares Million Lattice UPdates per second per node (MLUPs/node) on a range of node counts. It is noteworthy that the Cumulant is faster than any other model. In addition to the optimization explained above, the difference in performance is also affected by the turbulence model and by the locality of the accessed data. The Cumulant model solely uses local data, while PRR and HRR access the data from the neighbors to compute the gradients of the hybrid, second-order Hermite coefficient (cf. Eq. (17)). The same is true for the turbulence models. Smagorinsky, WALE, and Vreman utilize gradient calculation, while the implicit LES model of the Cumulant scheme only uses local data.

The $D3Q27$ and the $D3Q19$ stencils show a performance drop with 8 and 16–17 nodes, respectively. This drop only happens at these node counts specifically, as illustrated by the additional measurements with node counts of plus and minus 1 around the drop for Cumulant ($D3Q27$) and PRR with $D3Q19$. The reason might be a bad memory access pattern for this particular number of nodes, due to an unfortunate domain decomposition. To provide an explanation, further investigations are required.

The strong scaling analysis is shown in Fig. 13. As noticed before, there is an isolated sudden drop in performance. Aside from that drop,

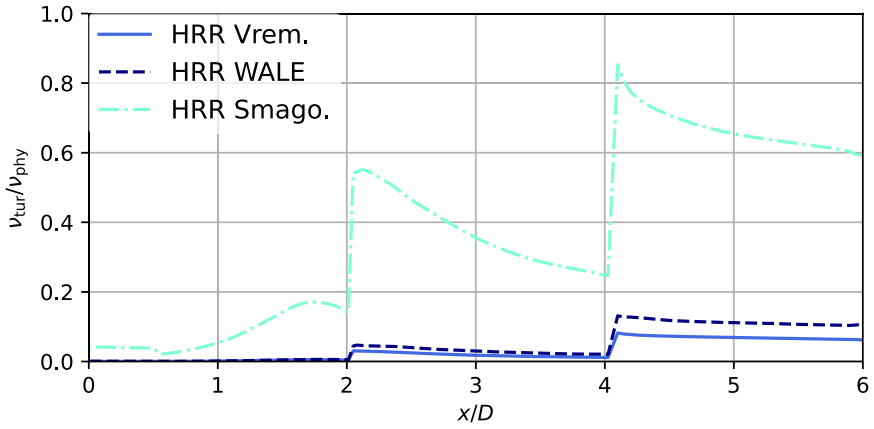

Fig. 7. Ratio between the averaged eddy-viscosity and the physical viscosity of the flow along the wake center line.

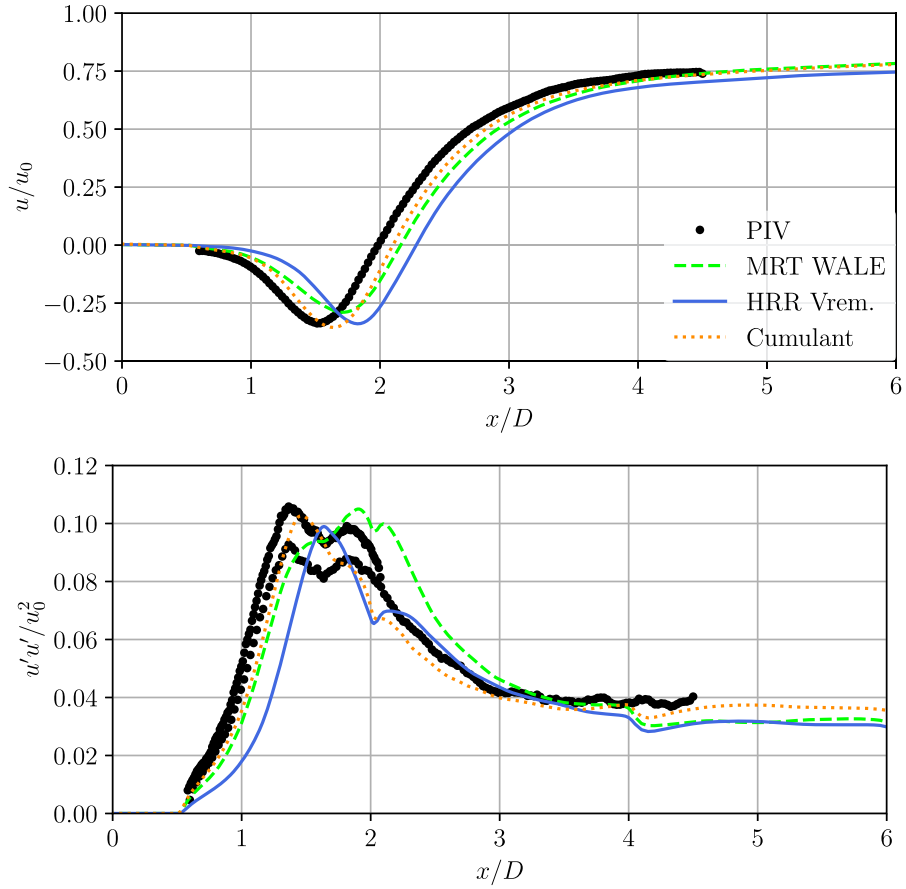

Fig. 8. Normalized average of the x-velocity u (top), and normalized Reynolds stress $u'u'$ (bottom) along the wake center line.

the general parallel efficiency nearly linearly decreases over the node counts for this setup with 57 million cells but yielding at least more than 60% on 256 nodes (16384 cores). Some variation in scalability can be observed between the schemes and the Cumulant approach yields the highest strong scaling parallel efficiency on 256 nodes with around 70%.

6. Conclusion

This study presents a comparison of different collision models of the lattice Boltzmann equation in 3D, including Cumulant, HRR, MRT, and PRR. For this purpose, the different models are implemented in the existing LBM framework Musubi, which has a proven record of achieving high scalability and performance on HPC infrastructures. Emphasis is laid on accuracy and performance. The velocity space is discretized with the $D3Q27$ stencil except for the MRT model, which is employed on a $D3Q19$ stencil due to stability reasons. Furthermore, the accuracy of these collision models is investigated in conjunction with three different subgrid scale turbulence models, namely the Smagorinsky, the WALE, and the Vreman model. The Cumulant has an implicit LES model and, therefore, does not require a separate model to account for turbulence. All the results in this study are obtained with the HPC LBM open-source solver Musubi.

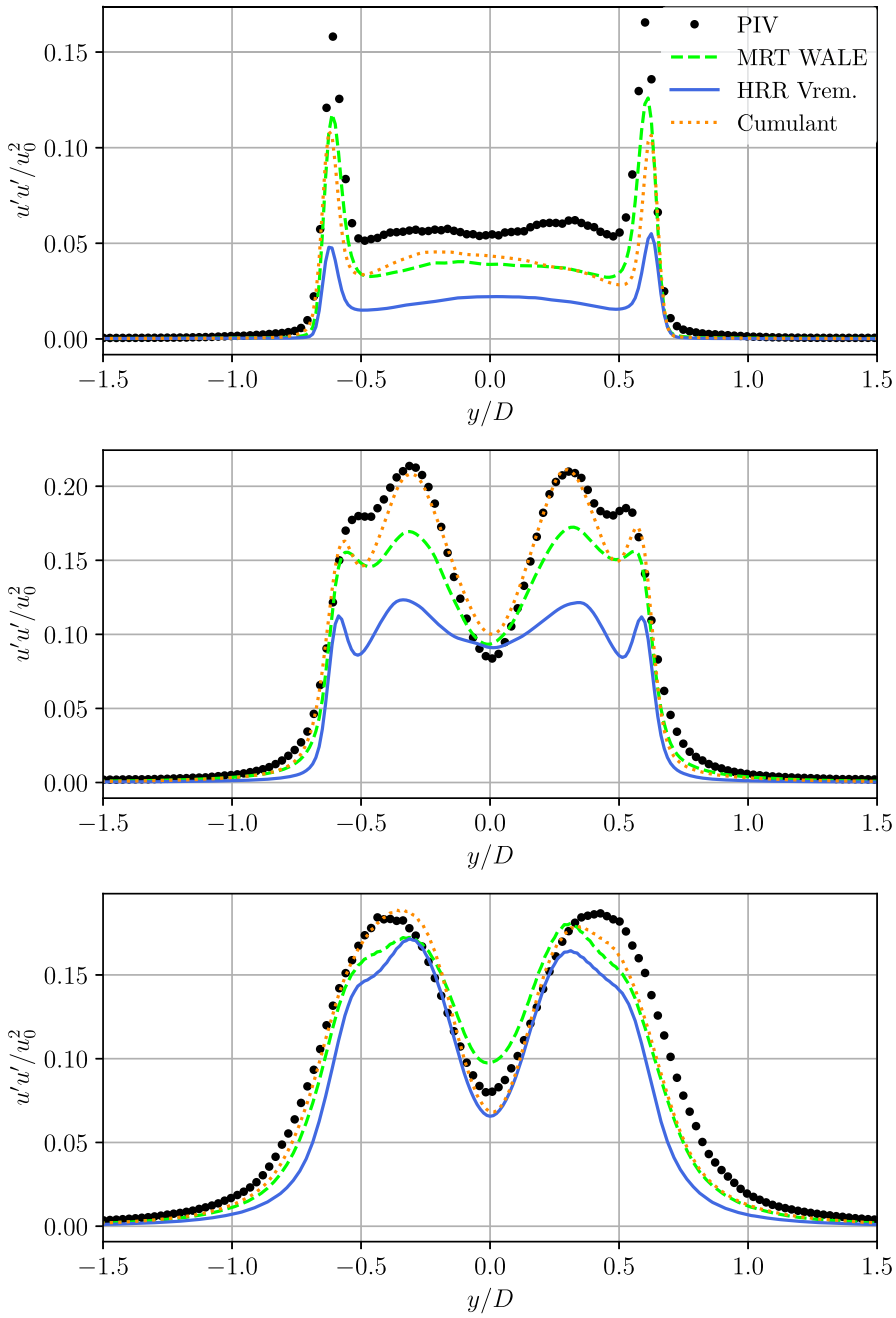

Fig. 9. Vertical profile of the normalized Reynolds stress $u'u'$ at $x = 1.06D$ (top), $x = 1.54D$ (middle), and $x = 2.02D$ (bottom).

The comparison is carried out by simulating a uniform flow past a circular cylinder at a Reynolds number of 3900. First, the effects of the mesh resolution on the results are studied. We quantified that the LES quality index is above 0.80 throughout the computational domain, which is the minimum value used in literature to label an LES solution as properly resolved. Given that also the resolution of our mesh in the regions of interest, *i.g.* boundary layer and wake, is in range with the ones available in the literature, and that the boundary layer is discretized with a mesh characterized by $y^+ \leq 1$, the effects of the mesh resolution on the results are deemed as minimal.

A first comparison is done by coupling HRR, MRT, and PRR with the Vreman model. We observe that the PRR fails to reproduce the trend of the Reynolds stresses obtained from the experiments. This observation agrees with the nature of the scheme, which is comparatively dissipative due to the finite-difference reconstruction of the second-order non-equilibrium Hermite coefficient. The results obtained via other schemes match the experimental trend, where the Cumulant shows the best agreement, followed by HRR and MRT. The difference is due to the high-order terms of velocity included in the equilibrium distributions. The MRT uses terms up to 2nd-order, while the other schemes use terms up to 6th-order. Thus, the macroscopic quantities show better accuracy due to the reduction of the magnitude of velocity-dependent errors.

In a second step, HRR and MRT are combined with different turbulence models. It is shown that results from HRR with Vreman, MRT with WALE, and the Cumulant with its ILES are in good agreement with the PIV data. Nevertheless, the best results are achieved with the Cumulant model. This becomes particularly clear when observing the trends of the Reynolds stresses. The HRR scheme does not cope well with the Smagorinsky and the WALE subgrid models. The reason for this relies on the velocity gradient. Both subgrid models use the value of the gradient to increase the amount of numerical dissipation (eddy-viscosity), while the same gradient is already used by the HRR

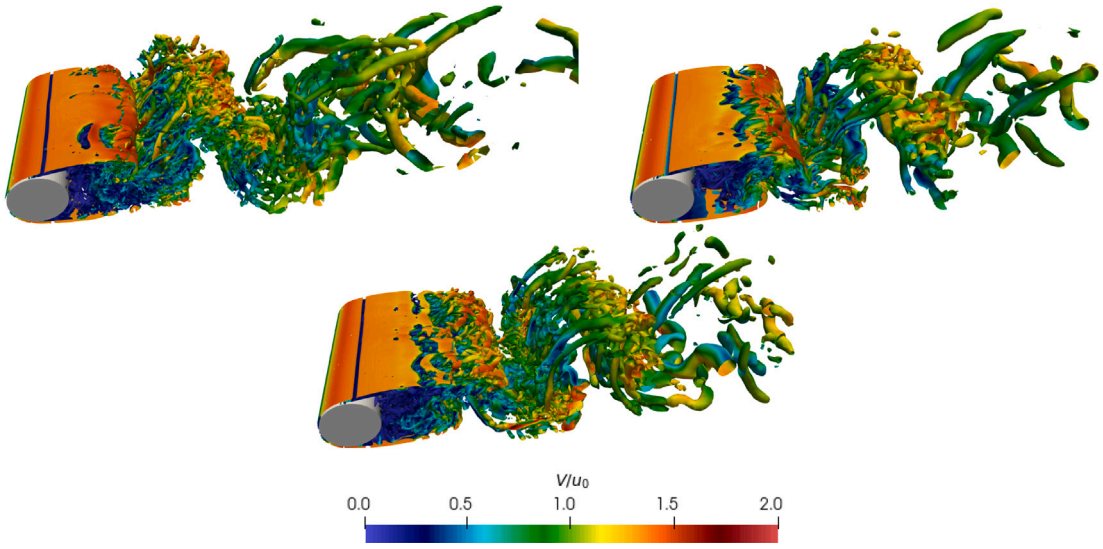

Fig. 10. Iso-contour of the normalized Q-criterion ($Q = 1$) colored by normalized velocity magnitude. Cumulant with ILES (top-left), HRR with Vreman (top-right), and MRT with WALE (bottom).

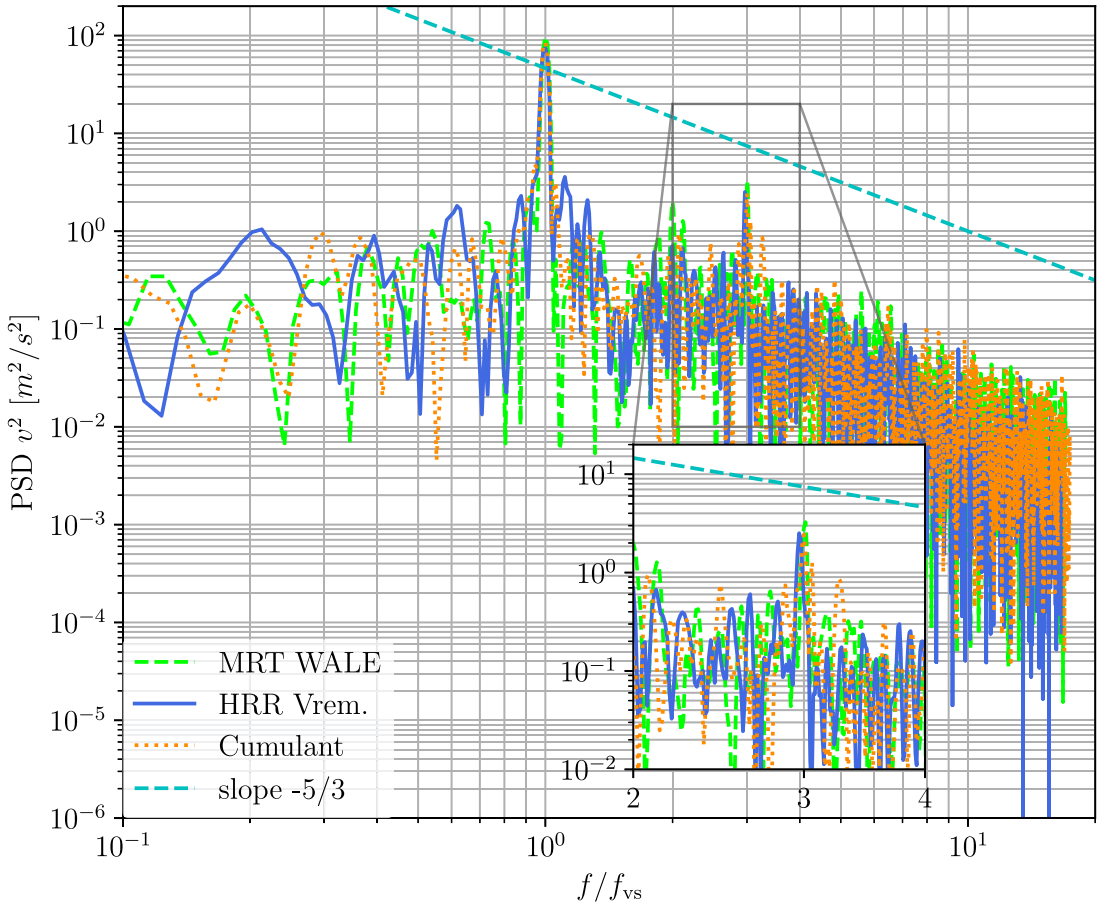

Fig. 11. Power spectra density of the y-component of the velocity at $(x, y, z) = (3.0D, 0.0, 0.0)$.

to generate dissipation during the collision step. The MRT shows a lack of dissipative behavior, which is corrected by the WALE model that introduces slightly more dissipation than the Vreman model.

A quantitative comparison is carried out by comparing drag and lift coefficients, the Strouhal number (St), and the length of the recirculation bubble (L_r). Apart from a general tendency to overpredict St and L_r, all results obtained with Musubi agree with both, numerical and experimental data available in the literature. The numerical benchmark was obtained with *ProLB*, a commercial CFD solver based on LBM. Once again, the Cumulant model stands out by best predicting the length of the recirculation bubble. By observing the power spectra density of the y-component of the velocity, all collision models are able to predict the second peak at $f = 3f_{vs}$, which was first observed experimentally.

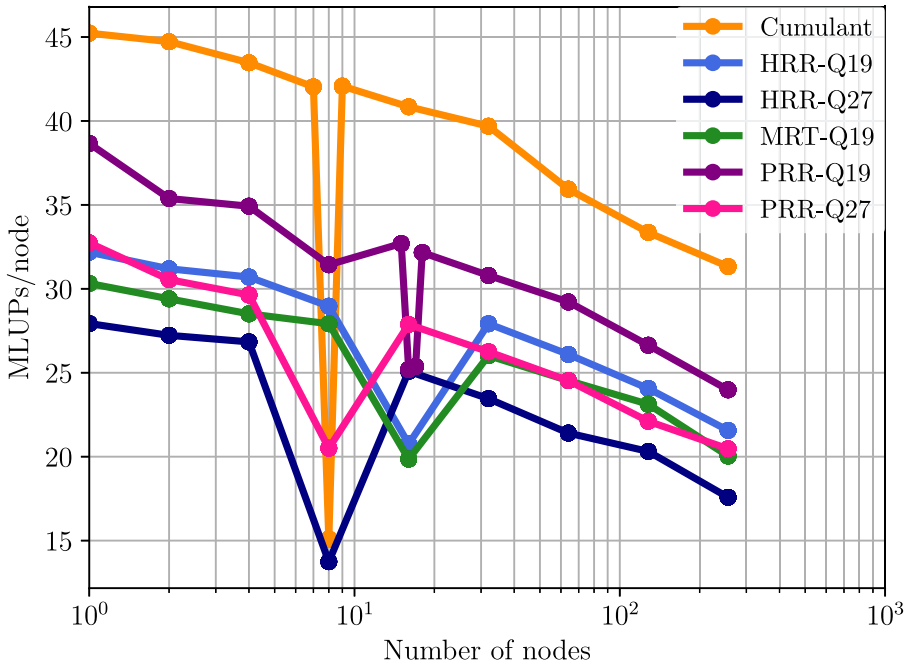

Fig. 12. Comparison of Million Lattice UPdates per second per node (MLUPs/node) with respect to the number of nodes for different collision models on either the $D3Q19$ or the $D3Q27$ lattice stencil. The Cumulant model works only on the $D3Q27$ stencil. For stability issues, the MRT runs only on the $D3Q19$ stencil.

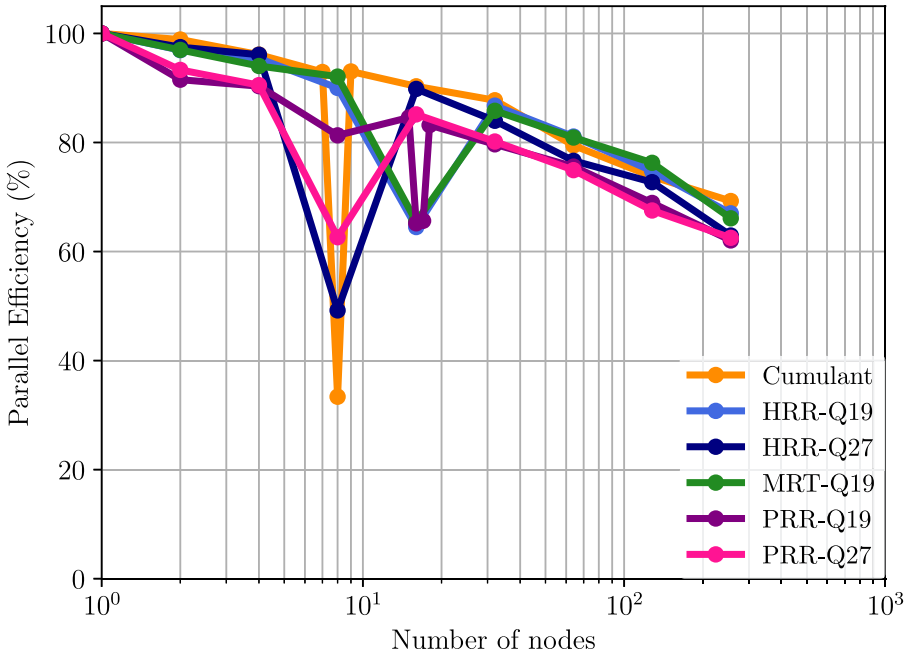

Fig. 13. Comparison of strong scaling for different collision models on either $D3Q19$ or $D3Q27$ lattice stencil. The Cumulant model works only on the $D3Q27$ stencil. For stability issues the, MRT runs only on the $D3Q19$ stencil.

The above test-case on a uniform mesh is used to investigate the performance of the different collision models. Results show that the Cumulant model, which was solely designed for the $D3Q27$ stencil, is faster than the other models independent of the $3D$ lattice stencil. The difference may be explained by the locality of the data. Regarding parallel efficiency, a strong scaling analysis for 57 million cells is carried out. We observe that the parallel efficiency remains higher than at least 60% on up to 16384 cores for all collision models.

In conclusion, the results confirm that the Cumulant model is the best choice in terms of accuracy and performance for the chosen test-case. Nevertheless, the implicit consideration of the subgrid scale turbulence is not intuitive and it is not clear to the authors how much turbulent viscosity is added when using the ILES model. To our best knowledge, there is no direct link between the turbulent viscosity and the limiter values used in the Cumulant model. How far these results and observed trends are transferable to other, more complex test-cases remains an open question that requires further investigation.

CRediT authorship contribution statement

Gregorio Gerardo Spinelli: Software, Validation, Formal analysis, Writing – original draft, Visualization, Investigation. **Tobias**

Horstmann: Validation, Writing – review & editing. **Kannan Masilamani:** Software, Validation, Investigation, Data curation. **Malav Mukesh Soni:** Validation. **Harald Klimach:** Software, Writing – review & editing. **Arthur Stück:** Funding acquisition, Supervision. **Sabine Roller:** Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The authors would like to thank Prof. Eric Lamballais from the University of Poitiers for providing the PIV data.

References

[1] Degrigny J, Pont G, Boussuge J-F, Sagaut P. Simulation of high-lift flows through IDDES in LBM. In: 55th 3AF international conferenceon applied aerodynamics, 23 — 25 March 2020, Poitiers – France. 2020.
[2] Konig B, Fares E, Murayama M, Ito Y, Yokokawa Y, Yamamoto K, Ishikawa K. Lattice-Boltzmann simulations of the JAXA JSM high-lift configuration. 2016, http://dx.doi.org/10.2514/6.2016-3721.
[3] Konig B, Duda BM, Laskowski GM. Lattice Boltzmann simulations for the 4th AIAA high-lift prediction workshop using powerflow. 2022, http://dx.doi.org/10.2514/6.2022-3433.
[4] Singh D, Konig B, Fares E, Murayama M, Ito Y, Yokokawa Y, Yamamoto K. Lattice-Boltzmann simulations of the JAXA JSM high-lift configuration in a wind tunnel. 2019, http://dx.doi.org/10.2514/6.2019-1333.
[5] Krüger T, Kusumaatmaja H, Kuzmin A, Shardt O, Silva G, Viggen EM. The lattice Boltzmann method. Springer Cham; 2017, http://dx.doi.org/10.1007/978-3-319-44649-3.
[6] Horstmann JT. Hybrid numerical method based on the lattice Boltzmann approach with application to non-uniform grids (Ph.D. thesis), Ecole Centrale Lyon; 2018, p. 143.
[7] Bhatnagar PL, Gross EP, Krook M. A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. Phys Rev 1954;94(3):511–25.
[8] D'Humières D. Generalized lattice-Boltzmann equations. Progr Astronaut Aeronaut 1992;159:450–8.
[9] d'Humières D. Multiple–relaxation–time lattice Boltzmann models in three dimensions. Philos Trans R Soc A 2002;360:437–51. http://dx.doi.org/10.1098/rsta.2001.0955.
[10] Ginzburg I, d'Humières D. Multireflection boundary conditions for lattice Boltzmann models. Phys Rev E 2003;68:066614. http://dx.doi.org/10.1103/PhysRevE.68.066614.
[11] Luo L-S, Liao W, Chen X, Peng Y, Zhang W. Numerics of the lattice Boltzmann method: Effects of collision models on the lattice Boltzmann simulations. Phys Rev E 2011;83:056710. http://dx.doi.org/10.1103/PhysRevE.83.056710.
[12] Dellar P. Bulk and shear viscosities in lattice Boltzmann equations. Phys Rev E 2001;64:1–10. http://dx.doi.org/10.1103/PhysRevE.64.031203.
[13] Fei L, Luo KH. Cascaded lattice Boltzmann method for incompressible thermal flows with heat sources and general thermal boundary conditions. Comput & Fluids 2018;165:89–95. http://dx.doi.org/10.1016/j.compfluid.2018.01.020.
[14] De Rosis A. Nonorthogonal central-moments-based lattice Boltzmann scheme in three dimensions. Phys Rev E 2017;95:013310. http://dx.doi.org/10.1103/PhysRevE.95.013310.
[15] Geier M, Greiner A, Korvink JG. Cascaded digital lattice Boltzmann automata for high Reynolds number flow. Phys Rev E 2006;73:066705. http://dx.doi.org/10.1103/PhysRevE.73.066705.
[16] Gruszczyński G, Łaniewski-Wołłk Ł. A comparative study of 3D cumulant and central moments lattice Boltzmann schemes with interpolated boundary conditions for the simulation of thermal flows in high Prandtl number regime. Int J Heat Mass Transfer 2022;197:123259. http://dx.doi.org/10.1016/j.ijheatmasstransfer.2022.123259.
[17] Jacob J, Malaspinas O, Sagaut P. A new hybrid recursive regularised Bhatnagar–Gross–Krook collision model for lattice Boltzmann method-based large eddy simulation. J Turbul 2018;19(11–12):1051–76. http://dx.doi.org/10.1080/14685248.2018.1540879.
[18] Feng Y, Boivin P, Jacob J, Sagaut P. Hybrid recursive regularized thermal lattice Boltzmann model for high subsonic compressible flows. J Comput Phys 2019;394:82–99. http://dx.doi.org/10.1016/j.jcp.2019.05.031.
[19] Coreixas C, Chopard B, Latt J. Comprehensive comparison of collision models in the lattice Boltzmann framework: Theoretical investigations. Phys Rev E 2019;100(3).
[20] Geier M, Schönherr M, Pasquali A, Krafczyk M. The cumulant lattice Boltzmann equation in three dimensions: Theory and validation. Comput Math Appl 2015;70:507–47. http://dx.doi.org/10.1016/j.camwa.2015.05.001.
[21] Latt J, Chopard B. Lattice Boltzmann method with regularized non-equilibrium distribution functions. Phys Fluid-Dyn 2005.
[22] Malaspinas O. Increasing stability and accuracy of the lattice Boltzmann scheme: recursivity and regularization. 2015, ArXiv e-Prints. arXiv:1505.06900.
[23] Dellar PJ. Lattice Boltzmann algorithms without cubic defects in Galilean invariance on standard lattices. J Comput Phys 2014;259:270–83. http://dx.doi.org/10.1016/j.jcp.2013.11.021.
[24] Coreixas C. High-order extension of the recursive regularized lattice Boltzmann method (Ph.D. thesis), Institut National Polytechnique de Toulouse; 2018.
[25] Coreixas C, Wissocq G, Chopard B, Latt J. Impact of collision models on the physical properties and the stability of lattice Boltzmann methods. Phil Trans R Soc A 2020;378:20190397. http://dx.doi.org/10.1098/rsta.2019.0397.
[26] Nathen P, Gaudlitz D, Krause MJ, Adams NA. On the stability and accuracy of the BGK, MRT and RLB Boltzmann schemes for the simulation of turbulent flows. Commun Comput Phys 2018;23(3):846–76. http://dx.doi.org/10.4208/cicp.OA-2016-0229.
[27] Haussmann M, Simonis S, Nirschl H, Krause MJ. Direct numerical simulation of decaying homogeneous isotropic turbulence — numerical experiments on stability, consistency and accuracy of distinct lattice Boltzmann methods. Internat J Modern Phys C 2019;30(09):1950074. http://dx.doi.org/10.1142/S0129183119500748.
[28] Geier M, Pasquali A, Schönherr M. Parametrization of the cumulant lattice Boltzmann method for fourth order accurate diffusion part I: Derivation and validation. J Comput Phys 2017;348:862–88. http://dx.doi.org/10.1016/j.jcp.2017.05.040.
[29] Hasert M, Masilamani K, Zimny S, Klimach H, Qi J, Bernsdorf J, Roller S. Complex fluid simulations with the parallel tree-based lattice Boltzmann solver musubi. J Comput Sci 2014;5(5):784–94.
[30] Parnaudeau P, Carlier J, Heitz D, Lamballais E. Experimental and numerical studies of the flow over a circular cylinder at Reynolds number 3900. Phys Fluids 2008;20:085101. http://dx.doi.org/10.1063/1.2957018.
[31] Geier M, Lenz S, Schönherr M, Krafczyk M. Under-resolved and large eddy simulations of a decaying Taylor–Green vortex with the cumulant lattice Boltzmann method. Theor Comput Fluid Dyn 2021;35(2):169–208.
[32] Smagorinsky J. General circulation experiments with the primitive equations I. The basic experiment. Mon Weather Rev 1963;91:99–164.
[33] Nicoud F, Ducros F. Subgrid-scale stress modelling based on the square of the velocity gradient tensor. Flow Turbul Combust 1999;62(3):183–200. http://dx.doi.org/10.1023/A:1009995426001, cited By 2446.
[34] Vreman B, Geurts B, Kuerten H. On the formulation of the dynamic mixed subgrid-scale model. Phys Fluids 1994;6(11):4057–9.
[35] Spinelli GG, Celik B. Leveling out interface temperature for conjugate heat transfer problems. Comput & Fluids 2020;210:104652. http://dx.doi.org/10.1016/j.compfluid.2020.104652.
[36] Suga K, Kuwata Y, Takashima K, Chikasue R. A D3Q27 multiple-relaxation-time lattice Boltzmann method for turbulent flows. Comput Math Appl 2015;69:518–29. http://dx.doi.org/10.1016/j.camwa.2015.01.010.
[37] Tölke J, Freudiger S, Krafczyk M. An adaptive scheme using hierarchical grids for lattice Boltzmann multi-phase flow simulations. Comput & Fluids 2006;35:820–30. http://dx.doi.org/10.1016/j.compfluid.2005.08.010.
[38] Geier M. Ab initio derivation of the cascaded lattice Boltzmann automaton (Ph.D. thesis), Albert-Ludwigs University Freiburg; 2006.
[39] Guo Z-L, Zheng C-G, Shi B-C. Non-equilibrium extrapolation method for velocity and pressure boundary conditions in the lattice Boltzmann method. Chin Phys 2002;11(4):366–74. http://dx.doi.org/10.1088/1009-1963/11/4/310.
[40] Bouzidi M, Firdaouss M, Lallemand P. Momentum transfer of a Boltzmann-lattice fluid with boundaries. Phys Fluids 2001;13:3452–9. http://dx.doi.org/10.1063/1.1399290.
[41] Uphoff S. Development and validation of turbulence models for lattice Boltzmann schemes (Ph.D. thesis), Braunschweig: Technische Universität Braunschweig; 2013.
[42] Message Passing Interface Forum. MPI: A message-passing interface standard version 4.0. 2021, URL https://www.mpi-forum.org/docs/mpi-4.0/mpi40-report.pdf.
[43] Masilamani K, Klimach H. Seeder mesh generator. 2013, [Online; accessed 01-August-2013] https://bitbucket.org/apesteam/seeder.
[44] Morton GM. A computer oriented geodetic data base; and a new technique in file sequencing. Tech. rep., Ottawa, Canada: IBM Ltd; 1966.
[45] Harlacher DF, Klimach H, Roller S, Siebert C, Wolf F. Dynamic load balancing for unstructured meshes on space-filling curves. Shanghai, China: IEEE; 2012, p. 1661–9. http://dx.doi.org/10.1109/IPDPSW.2012.207.

[46] Alkishriwi N, Meinke M, Schröder W. A large-eddy simulation method for low mach number flows using preconditioning and multigrid. Comput & Fluids 2006;35(10):1126–36. http://dx.doi.org/10.1016/j.compfluid.2005.06.002.

[47] Ouvrard H, Koobus B, Dervieux A, Salvetti MV. Classical and variational multi-scale LES of the flow around a circular cylinder on unstructured grids. Comput & Fluids 2010;39(7):1083–94. http://dx.doi.org/10.1016/j.compfluid.2010.01.017.

[48] Breuer M. Numerical and modeling influences on large eddy simulations for the flow past a circular cylinder. Int J Heat Fluid Flow 1998;19(5):512–21. http://dx.doi.org/10.1016/S0142-727X(98)10015-2.

[49] Xu H, Sagaut P. Analysis of the absorbing layers for the weakly-compressible lattice Boltzmann methods. J Comput Phys 2013;245:14–42. http://dx.doi.org/10.1016/j.jcp.2013.02.051.

[50] Celik IB, Cehreli ZN, Yavuz I. Index of resolution quality for large eddy simulations. J Fluids Eng 2005;127(5):949–58. http://dx.doi.org/10.1115/1.1990201.

[51] Pope SB. Turbulent flows. Cambridge University Press; 2000, http://dx.doi.org/10.1017/CBO9780511840531.