

Energy-conserving formulation of the two-fluid model for incompressible two-phase flow in channels and pipes

J.F.H. Buist^{a,b,*}, B. Sanderse^a, S. Dubinkina^c, R.A.W.M. Henkes^{b,d}, C.W. Oosterlee^e

^a Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands

^b Delft University of Technology, Delft, The Netherlands

^c Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

^d Shell Technology Centre Amsterdam, Amsterdam, The Netherlands

^e Utrecht University, Utrecht, The Netherlands

ARTICLE INFO

Keywords:

Two-fluid model
Energy conservation
Energy-conserving discretization
Incompressible flow

ABSTRACT

We show that the one-dimensional (1D) two-fluid model (TFM) for stratified flow in channels and pipes (in its incompressible, isothermal form) satisfies an energy conservation equation, which arises naturally from the mass and momentum conservation equations that constitute the model. This result extends upon earlier work on the shallow water equations (SWE), with the important difference that we include non-conservative pressure terms in the analysis, and that we propose a formulation that holds for ducts with an arbitrary cross-sectional shape, with the 2D channel and circular pipe geometries as special cases.

The second novel result of this work is the formulation of a finite volume scheme for the TFM that satisfies a discrete form of the continuous energy equation. This discretization is derived in a manner that runs parallel to the continuous analysis. Due to the non-conservative pressure terms it is essential to employ a staggered grid, which requires careful consideration in defining the discrete energy and energy fluxes, and the relations between them and the discrete model. Numerical simulations confirm that the discrete energy is conserved.

1. Introduction

The one-dimensional (1D) two-fluid model (TFM) is a dynamic model for stratified flow in channels and pipes. It simplifies the full three-dimensional multiphase flow problem by resolving only the cross-sectionally averaged quantities (hold-ups, velocities, and pressure), which are often of practical interest. There are many variants of the model, but the basic idea, of two interacting fluids whose behavior is cross-sectionally averaged to obtain a 1D model, was introduced by Wallis (1969) [1] and Ishii (1975) [2]. The model has among others applications in the oil and gas industry [3], in CO₂ transport and storage [4], and in nuclear reactor safety analysis [5].

An unsolved issue with the basic version of the TFM is that the initial value problem for the governing equations is only conditionally well-posed [6]. This means that it is well-posed for some flow configurations and ill-posed for others (e.g. when there is a large velocity difference between the two fluids). Conventionally, ill-posedness of the TFM is demonstrated by a linear stability analysis which shows an unbounded growth rate for the smallest wavelengths, when the values of the model variables are such that the eigenvalues are complex. In this case the solution is said to carry no physical meaning [7]. However,

when drawing conclusions on the well-posedness of the TFM, it is important to also consider its nonlinear aspects, and not only rely on a linearized analysis [8,9]. Examples of studies that have included nonlinear effects in the TFM analysis can be found in [10,11]. However, a complete nonlinear analysis, with implications for obtaining a robust discretization, is still missing.

In this work, we strive towards such a nonlinear analysis by presenting an expression for an energy which is conserved by the full (nonlinear) TFM, in its incompressible and isothermal form. This approach is motivated by the fact that for the incompressible Navier–Stokes equations such an analysis provides stability estimates [12,13], and that for compressible equations it is closely related to the concept of entropy stability [14]. Important to note is that such an energy is not the thermodynamic energy for which a separate conservation equation exists in the compressible TFM. Rather, the considered energy conservation is an inherent property of the mass and momentum conservation equations that constitute the incompressible TFM: the energy is a secondary conserved quantity of the model. Its physical meaning is therefore the mechanical energy of the system (kinetic plus potential energy).

* Corresponding author at: Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands.

E-mail address: jurriaan.buist@cwi.nl (J.F.H. Buist).

In order to derive this mechanical energy equation, we take the approach from [15], in which the dot product of the shallow water equations (SWE) and a vector of entropy variables is taken in such a way that a scalar energy equation results. However, an important difference with the SWE (and two-layer SWE [16]) is the presence of non-conservative pressure terms that are linked to the constraint that the fluid phases have to fill the cross section. Another important difference is that we consider arbitrary duct geometries, as opposed to the 1D SWE which in effect utilizes a planar channel geometry. Given these differences, the key challenge is thus to find a conserved energy and corresponding energy flux function for the TFM, and this will be the first main focus of this paper.

The second focus of this paper is to derive a spatial discretization which conserves a discrete version of the energy. Again, our approach is inspired by methods which have been developed for the SWE [17]. An important difference is that these methods are designed for collocated grids, while we will adapt them to a staggered grid. This is motivated by the presence of the (non-conservative) pressure terms in the TFM, which makes the use of a staggered grid much more convenient (similar to the case of the incompressible Navier–Stokes equations [12]). However, the staggered grid introduces new challenges, for example in terms of the definitions of the energy and energy fluxes. We will formulate a discretization method that tackles these issues and yields, in a constructive manner, a new set of numerical fluxes on a staggered grid that is energy-conserving. Our approach adheres to the conservative form of the model, and therefore yields a discretization that satisfies the proper shock conditions. This discretization can also be viewed as an extension (to the TFM) of the staggered grid SWE discretization found in [18], although that discretization is derived using a rather different approach (based on a comparison to the compressible Euler equations and arguments on skew-symmetric operators).

Since our discretization takes the form of a finite volume scheme, mass and total momentum are also conserved, along with energy. These discrete conservation properties match the properties of the continuous equations, and improve the physical fidelity of long-term numerical simulations, by preventing artificial (numerical) damping or amplification of the flow. This will be illustrated in this paper by a sloshing test case in a closed tank, which in absence of viscosity will constantly remain in motion, like an undamped pendulum.

This paper is set up as follows. First, in Section 2 we present the governing equations of the TFM. In Section 3 we discuss the conditions for energy conservation, and introduce an energy and energy flux that satisfies these conditions, providing local and global energy conservation equations for the continuous TFM. We outline how the equations are discretized in Section 4, while leaving open the specific form of the numerical fluxes. Then, in Section 5, we present the discrete versions of the continuous conditions for energy conservation, and propose a set of new conservative numerical fluxes. Finally, in Section 6 we present numerical results which demonstrate exact conservation of the aforementioned energy.

2. Governing equations

The 1D TFM, as considered in this work, describes the separated flow of a (heavier) lower fluid L and a (lighter) upper fluid U through a channel or pipe. It can be derived by applying a cross-sectional averaging procedure to the Navier–Stokes equations [9,19]. An important assumption made in the derivation of the model is that the streamwise length scale is much larger than the normal length scale (i.e. the pipe diameter), which is referred to as the long wavelength assumption. As a consequence, along the normal direction the flow is in hydrostatic balance. We will omit source terms, such as wall friction, since such terms are sources or sinks of energy, and we are interested in the energy conservation properties of the core model. Good discussions of the assumptions underlying the TFM are given by [20,21].

The cross-sectionally averaged equations can be written in the following concise form [22,23] (with $\mathbf{q} = \mathbf{q}(s, t)$):

$$\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{q})}{\partial s} + \mathbf{d}(\mathbf{q}) \frac{\partial p}{\partial s} = \mathbf{0}, \quad (1)$$

where \mathbf{q} constitutes the vector of ‘conserved’ variables,¹ namely the mass and momentum of each phase:

$$\mathbf{q}^T = [q_1 \quad q_2 \quad q_3 \quad q_4] = [\rho_U A_U \quad \rho_L A_L \quad \rho_U u_U A_U \quad \rho_L u_L A_L].$$

Here ρ_U and ρ_L are the densities, A_U and A_L are the cross sections, and u_U and u_L are the averaged velocities, all of the upper and lower fluids, respectively. The superscript T indicates a transpose. We consider the isothermal, incompressible case, so that ρ_U and ρ_L are constant.

The fluxes \mathbf{f} describe convection of mass and momentum and gradients in the interface level. In terms of \mathbf{q} they are given by

$$\mathbf{f}(\mathbf{q}) = \begin{bmatrix} q_3 \\ q_4 \\ \frac{q_3^2}{q_1} - \rho_U g_n \hat{H}_U \\ \frac{q_4^2}{q_2} - \rho_L g_n \hat{H}_L \end{bmatrix} = \begin{bmatrix} \rho_U u_U A_U \\ \rho_L u_L A_L \\ \rho_U u_U^2 A_U - \rho_U g_n \hat{H}_U \\ \rho_L u_L^2 A_L - \rho_L g_n \hat{H}_L \end{bmatrix}, \quad (2)$$

where $\hat{H}_U = \hat{H}_U(\mathbf{q})$ and $\hat{H}_L = \hat{H}_L(\mathbf{q})$ are geometric terms (to be discussed shortly), and g_n is the gravitational acceleration in the normal direction.

The fifth variable is the interface pressure p , and the non-conservative pressure terms are given by $\mathbf{d}(\partial p / \partial s)$ with

$$\mathbf{d}(\mathbf{q})^T = \begin{bmatrix} 0 & 0 & \frac{q_1}{\rho_U} & \frac{q_2}{\rho_L} \end{bmatrix} = \begin{bmatrix} 0 & 0 & A_U & A_L \end{bmatrix}.$$

The quantities $\hat{H}_U = \hat{H}_U(A_U(q_1, \rho_U))$ and $\hat{H}_L = \hat{H}_L(A_L(q_2, \rho_L))$ are geometry-dependent and are defined by

$$\hat{H}_U := \int_{a_U} (h - H_L) da, \quad \hat{H}_L := \int_{a_L} (h - H_L) da. \quad (3)$$

Here the difference between the coordinate h and the two-fluid interface height H_L is integrated over the area a_U occupied by the upper fluid and the area a_L occupied by the lower fluid, respectively. Using these general expressions, the model equations are valid for arbitrarily shaped cross sections. See Appendix A for evaluations of the integrals for the 2D channel and circular pipe geometries. The spatial derivatives of \hat{H}_U and \hat{H}_L that appear in the fluxes \mathbf{f} are known as the level gradient terms, which result from the hydrostatic variation of the pressure.

Since the upper and lower fluid together fill the pipe, the system is subject to the volume constraint

$$\frac{q_1}{\rho_U} + \frac{q_2}{\rho_L} = A. \quad (4)$$

The entire system therefore consists of four evolution equations plus one constraint, and four ‘conserved’ variables plus the pressure. In our incompressible setting, a derived constraint can be obtained by differentiating the constraint (4) and substituting the mass equations, leading to [23]:

$$\frac{\partial}{\partial s} \left(\frac{q_3}{\rho_U} + \frac{q_4}{\rho_L} \right) = 0, \quad (5)$$

which can be integrated in space to give that the volumetric flow Q is constant in space, and a function of time only:

$$Q(\mathbf{q}) := \frac{q_3}{\rho_U} + \frac{q_4}{\rho_L} = Q(t). \quad (6)$$

This derived constraint, termed the volumetric flow constraint, can be seen as the incompressibility constraint for the TFM.

¹ Note that the pressure term is not in conservative form, so q_3 and q_4 individually are not conserved, but $q_3 + q_4$ is.

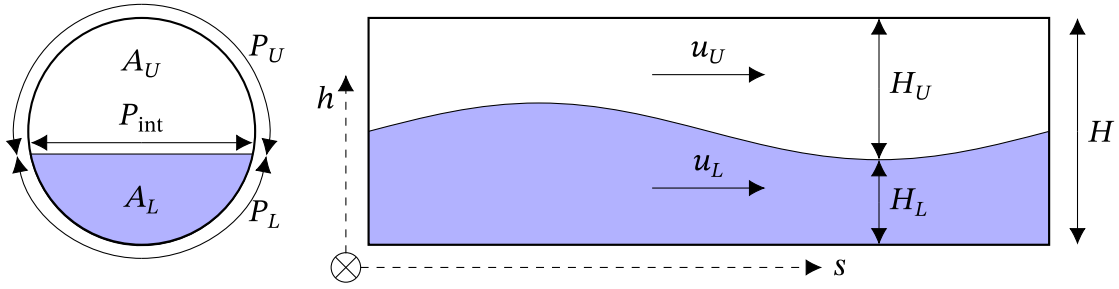


Fig. 1. A schematic of stratified two-fluid flow in ducts (a circular pipe segment is shown as an example) described by the 1D TFM.

We can use these constraints to set up an equation for the pressure. The pressure equation is obtained by summing the momentum equations [23]:

$$\mathbf{I}^T \mathbf{d} \frac{\partial \mathbf{p}}{\partial s} = -\mathbf{I}^T \left(\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial \mathbf{f}}{\partial s} \right), \quad \text{with } \mathbf{I}^T = \begin{bmatrix} 0 & 0 & \frac{1}{\rho_U} & \frac{1}{\rho_L} \end{bmatrix},$$

which can be expanded and rewritten with the definition of \mathbf{Q} to yield

$$\left(\frac{q_1}{\rho_U^2} + \frac{q_2}{\rho_L^2} \right) \frac{\partial p}{\partial s} = -\frac{d\mathbf{Q}}{dt} - \frac{\partial}{\partial s} \left(\frac{f_3}{\rho_U} + \frac{f_4}{\rho_L} \right).$$

Finally, taking the derivative of this equation to s and applying constraint (5) gives

$$\frac{\partial}{\partial s} \left(\left(\frac{q_1}{\rho_U^2} + \frac{q_2}{\rho_L^2} \right) \frac{\partial p}{\partial s} \right) = -\frac{\partial^2}{\partial s^2} \left(\frac{f_3}{\rho_U} + \frac{f_4}{\rho_L} \right). \quad (7)$$

This is a ‘Poisson-type’ equation for the pressure, which can be used in place of (4) to close the system of equations. In our numerical algorithm (discussed in Section 4) we apply a discrete version of (7) in this manner.

3. Energy conservation equation for the continuous two-fluid model

3.1. Outline: conditions for energy conservation

Having set-up the TFM governing equations, the first key objective of this paper is to prove *local* and *global* energy equalities that are *implied* by this equation set. This is similar to the energy analyses for e.g. the incompressible Navier–Stokes equations [12], the SWE [24], and the two-layer SWE [16]. In all these models, no energy conservation equation is included in the model, but energy conservation follows from the mass and momentum conservation equations alone. It can therefore be said that the energy is a secondary conserved quantity.

Our proof of global energy conservation follows the approach in [15,17] and starts by showing that a local energy conservation equation of the form

$$\frac{\partial e}{\partial t} + \frac{\partial}{\partial s} (h + j) = 0 \quad (8)$$

can be derived, purely based on manipulating the governing equations, given by (1). Here $e(\mathbf{q})$ is the local energy, and $h(\mathbf{q})$ and $j(\mathbf{q})$ are energy fluxes (to be detailed later); $h(\mathbf{q})$ is not be confused with the normal coordinate h shown in Fig. 1. If (8) holds, then it can be integrated in space to yield

$$\frac{dE}{dt} = -[h + j]_{s_1}^{s_2} = 0, \quad (9)$$

where the last equality ($= 0$) holds in case of periodic or closed boundaries, and the global energy $E(t)$ is defined as

$$E(t) := \int_{s_1}^{s_2} e \, ds. \quad (10)$$

The key is therefore to obtain the local energy conservation equation (8). To achieve this, one first postulates an energy $e(\mathbf{q})$ (typically

guided by physical considerations). Second, one calculates the vector of so-called entropy variables, defined as²

$$\mathbf{v}(\mathbf{q}) := \left[\frac{\partial e}{\partial \mathbf{q}} \right]^T.$$

Taking the dot product of the system (1) with \mathbf{v} leads to

$$\left\langle \mathbf{v}, \frac{\partial \mathbf{q}}{\partial t} \right\rangle + \left\langle \mathbf{v}, \frac{\partial \mathbf{f}}{\partial s} \right\rangle + \left\langle \mathbf{v}, \mathbf{d} \frac{\partial p}{\partial s} \right\rangle = 0, \quad (11)$$

in which we have ignored source terms (as indicated before), and the brackets denote a dot product over the vector elements:

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^T \mathbf{y}.$$

The time derivative term can be written as

$$\left\langle \mathbf{v}, \frac{\partial \mathbf{q}}{\partial t} \right\rangle = \left(\frac{\partial e}{\partial \mathbf{q}} \right) \frac{\partial \mathbf{q}}{\partial t} = \frac{\partial e}{\partial t},$$

so (11) becomes an equation for the time evolution of the energy.

Given an expression for e , the art is to find an energy flux h that satisfies

$$\left\langle \mathbf{v}, \frac{\partial \mathbf{f}}{\partial s} \right\rangle = \frac{\partial h}{\partial s}, \quad (12)$$

since then the second term in (11) can be written in the (locally) conservative form given by (8). In order to get a condition solely referring to the relations between different functions of \mathbf{q} (i.e. independent of s), the chain rule (valid for smooth solutions) is employed to convert (12) to:

$$\left\langle \mathbf{v}, \frac{\partial \mathbf{f}}{\partial \mathbf{q}} \right\rangle = \frac{\partial h}{\partial \mathbf{q}}. \quad (13)$$

This is the condition encountered in e.g. [15,17] for an energy flux h to conserve a given energy e (or, more generally: entropy function) of the SWE.

Likewise, we need to find a flux j such that the product of \mathbf{v} and the pressure gradient can be written in conservative form:

$$\left\langle \mathbf{v}, \mathbf{d} \frac{\partial p}{\partial s} \right\rangle = \frac{\partial j}{\partial s}. \quad (14)$$

The difference between h and j lies in the fact that h is responsible for the spatially conservative terms of the governing equations, whereas j takes the non-conservative part into account. Perhaps surprisingly, we will show that these non-conservative terms $\mathbf{d}(\partial p/\partial s)$ can indeed be written in conservative form in the energy equation. An alternative formulation of condition (14) is given by

$$\frac{\partial}{\partial s} (\langle \mathbf{v}, \mathbf{d} \rangle p) - p \frac{\partial}{\partial s} \langle \mathbf{v}, \mathbf{d} \rangle = \frac{\partial j}{\partial s}. \quad (15)$$

In order for the local energy to be conserved, there must exist a j (for the given e and resulting \mathbf{v}) such that this condition is satisfied.

An important difference between this derivation and the derivation for the SWE as found in e.g. [15] is the non-conservative pressure

² We take the convention that $\partial e/\partial \mathbf{q}$ is a row vector, making \mathbf{v} a column vector.

term. Although the two-layer SWE [25] also features a non-conservative term, in the TFM the non-conservative term depends on a variable for which there is no evolution equation (namely the pressure). This pressure term is instead linked directly to the volume constraint (4) and volumetric flow constraint (6) [23], which are not present in the SWE. For a system in conservative form without source terms, (13) is the only condition. This condition is emphasized in literature (e.g. [26]) as the condition for the existence of an entropy function. The derivation of energy conservation for the conservative part of the TFM system thus matches the derivation of an entropy condition for a conservative hyperbolic system.

In summary, the task is to find a set e , h and, j which satisfy conditions (12) and (14) for the current model with flux \mathbf{f} and pressure terms $\mathbf{d}(\partial p/\partial s)$. The alternative conditions (13) and (15) yield results more directly and will therefore be used in the following section. The result is the local energy conservation equation (8), and global energy conservation then follows directly.

3.2. Choice of energy and energy fluxes

We will show that the mechanical energy

$$e = \rho_U g_n \tilde{H}_U + \rho_L g_n \tilde{H}_L + \frac{1}{2} \frac{q_3^2}{q_1} + \frac{1}{2} \frac{q_4^2}{q_2} \quad (16)$$

$$= \rho_U g_n \tilde{H}_U + \rho_L g_n \tilde{H}_L + \frac{1}{2} \rho_U A_U u_U^2 + \frac{1}{2} \rho_L A_L u_L^2,$$

is conserved by the TFM (in absence of source terms). Here $\tilde{H}_U = \tilde{H}_U(A_U(q_1, \rho_U))$ represents the center of mass of the upper fluid multiplied by A_U and $\tilde{H}_L = \tilde{H}_L(A_L(q_2, \rho_L))$ represents the center of mass of the lower fluid multiplied by A_L (see Appendix A), so that the first two terms can be recognized as the potential energy of the upper and lower fluid, respectively. The third and fourth terms represent the kinetic energy of the upper and lower fluid, respectively. Therefore, this energy e has a clear physical interpretation. Note that just as for the incompressible Navier–Stokes equations, the pressure does not contribute to the energy.

The entropy variables are given by

$$\mathbf{v} = \left[\frac{\partial e}{\partial \mathbf{q}} \right]^T = \begin{bmatrix} -\frac{1}{2} \frac{q_3^2}{q_1} + g_n \frac{d\tilde{H}_U}{dA_U} \\ -\frac{1}{2} \frac{q_4^2}{q_2} + g_n \frac{d\tilde{H}_L}{dA_L} \\ \frac{q_3}{q_1} \\ \frac{q_4}{q_2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \frac{q_3^2}{q_1} + g_n(H - H_U) \\ -\frac{1}{2} \frac{q_4^2}{q_2} + g_n H_L \\ \frac{q_3}{q_1} \\ \frac{q_4}{q_2} \end{bmatrix}, \quad (17)$$

with $H_U = H_U(A_U(q_1, \rho_U))$ and $H_L = H_L(A_L(q_2, \rho_L))$ representing the fluid layer thickness of the upper and lower fluids, respectively (see Appendix A). It is important that in the energy and energy flux terms concerning the upper fluid we use $\hat{H}_U(A_U)$, $\tilde{H}_U(A_U)$, and $H_U(A_U)$, while for the lower fluid we use $\hat{H}_L(A_L)$, $\tilde{H}_L(A_L)$, and $H_L(A_L)$. It is possible to use the volume constraint to change this functional dependence, but our choice leads to an elegant form of the energy conservation conditions.

The task is to find h and j . We start with j : the pressure term needs to satisfy (15). Straightforward evaluation gives

$$\langle \mathbf{v}, \mathbf{d} \rangle = \frac{q_3}{q_1} \frac{q_1}{\rho_U} + \frac{q_4}{q_2} \frac{q_2}{\rho_L} = Q,$$

with Q the volumetric flow rate given by (6). Because of the volumetric flow constraint (6), the second term of (15) vanishes, so that the condition on the pressure gradient evaluates to

$$\frac{\partial}{\partial s} (Qp) = \frac{\partial j}{\partial s},$$

and (15) is satisfied with

$$j = Qp. \quad (18)$$

We note that p is the pressure that enforces incompressibility — it does not include a driving pressure gradient (which would appear as a source term in the TFM governing equations). Therefore j is periodic in space in the case of periodic boundaries. In the case of closed boundaries, Q must be zero, meaning that $j = 0$ throughout the domain. This means that when integrating (8) over a closed or periodic domain, the terms involving j vanish, and thus this definition for j is compatible with global energy conservation as described by (9).

Note that alternatively, global energy conservation of the pressure terms could be proved by integrating (14) over the domain, and equating it to the integral of the second term on the left-hand side of (15). This equality can be seen as a parallel to a well-known property of the incompressible Navier–Stokes equations, namely that, apart from the minus sign, the gradient operator applied to the pressure is the adjoint of the divergence operator applied to the velocity field [27]. For both models, the final step to proving global energy conservation of the pressure terms is substituting the incompressibility constraint (here given by $\partial Q/\partial s = 0$).

The next task is to find h . Based on the form of h for the SWE and condition (13), we propose the following choice

$$h = g_n q_3 (H - H_U) + g_n q_4 H_L + \frac{1}{2} \frac{q_3^3}{q_1^2} + \frac{1}{2} \frac{q_4^3}{q_2^2} \quad (19)$$

$$= g_n q_3 (H - H_U) + g_n q_4 H_L + \frac{1}{2} \rho_U A_U u_U^3 + \frac{1}{2} \rho_L A_L u_L^3,$$

which can be shown to satisfy condition (13) by computing:

$$\left[\langle \mathbf{v}, \frac{\partial \mathbf{f}}{\partial \mathbf{q}} \rangle \right]^T = \begin{bmatrix} -\frac{q_3^3}{q_1^2} - g_n \frac{d\tilde{H}_U}{dA_U} \frac{q_3}{q_1} \\ -\frac{q_4^3}{q_2^2} - g_n \frac{d\tilde{H}_L}{dA_L} \frac{q_4}{q_2} \\ \frac{3}{2} \frac{q_3^2}{q_1^2} + g_n \frac{d\tilde{H}_U}{dA_U} \\ \frac{3}{2} \frac{q_4^2}{q_2^2} + g_n \frac{d\tilde{H}_L}{dA_L} \end{bmatrix}, \quad \left[\frac{\partial h}{\partial \mathbf{q}} \right]^T = \begin{bmatrix} -\frac{q_3^3}{q_1^2} - \frac{g_n}{\rho_U} \frac{dH_U}{dA_U} q_3 \\ -\frac{q_4^3}{q_2^2} + \frac{g_n}{\rho_L} \frac{dH_L}{dA_L} q_4 \\ \frac{3}{2} \frac{q_3^2}{q_1^2} + g_n (H - H_U) \\ \frac{3}{2} \frac{q_4^2}{q_2^2} + g_n H_L \end{bmatrix}.$$

The last two entries in these vectors are equal because of relations (A.7), derived in Appendix A. The first two entries are equal due to the geometric relations (A.4), which we repeat here in terms of the conserved variables \mathbf{q} :

$$\frac{\rho_U}{q_1} \frac{d\hat{H}_U}{dA_U} = \frac{dH_U}{dA_U}, \quad \frac{\rho_L}{q_2} \frac{d\hat{H}_L}{dA_L} = -\frac{dH_L}{dA_L}. \quad (20)$$

These relations follow directly from the definitions of these geometric quantities and hold for arbitrary duct geometries. Note that, alternatively, condition (12) can be used (instead of (13)), which leads to the following conditions:

$$\frac{\rho_U}{q_1} \frac{\partial \hat{H}_U}{\partial s} = \frac{\partial H_U}{\partial s}, \quad \frac{\rho_L}{q_2} \frac{\partial \hat{H}_L}{\partial s} = -\frac{\partial H_L}{\partial s}, \quad (21)$$

which may also be shown to be satisfied directly via application of Leibniz' rule to the definitions of \hat{H}_U and \hat{H}_L . These last two conditions will play an important role in the discrete analysis in Section 5.

In conclusion, we have proposed a novel set of e , h , and j for the TFM and have shown that the local energy conservation equation (8) is satisfied.

3.3. Reformulation in terms of the entropy potential and conditions on fluxes

Conditions (12) and (14), or their alternatives (13) and (15), were used in the previous section to find a combination of e , h and j for the continuous TFM, given the fluxes \mathbf{f} from the governing equations. In Section 5, we will instead aim to find discrete flux functions, given

discretized versions of e , h and j (that are inspired by their continuous counterparts). Eq. (12) is not a very useful formulation to find such numerical flux functions, because it is a condition imposed on the jump in \mathbf{f} , rather than \mathbf{f} itself. Therefore, Eq. (12) is reformulated using the concept of the *entropy potential* [17,24].

The entropy potential is defined to be related to \mathbf{v} , \mathbf{f} , and h in the following manner:

$$\psi := \langle \mathbf{v}, \mathbf{f} \rangle - h. \quad (22)$$

With this definition, we can reformulate condition (12) using the product rule ($\frac{\partial}{\partial s} \langle \mathbf{v}, \mathbf{f} \rangle = \langle \frac{\partial \mathbf{v}}{\partial s}, \mathbf{f} \rangle + \langle \mathbf{v}, \frac{\partial \mathbf{f}}{\partial s} \rangle$) as:

$$\frac{\partial \psi}{\partial s} = \langle \frac{\partial \mathbf{v}}{\partial s}, \mathbf{f} \rangle. \quad (23)$$

The entropy potential can be directly calculated from its definition (22) and is given by:

$$\begin{aligned} \psi = \langle \mathbf{v}, \mathbf{f} \rangle - h &= -\rho_U g_n \hat{H}_U \frac{q_3}{q_1} - \rho_L g_n \hat{H}_L \frac{q_4}{q_2} \\ &= -\rho_U g_n \hat{H}_U u_U - \rho_L g_n \hat{H}_L u_L. \end{aligned} \quad (24)$$

Because this entropy potential is based on an h that satisfies (12), (23) is satisfied by construction. Nevertheless, we outline the details to convert (23) into conditions on the individual numerical fluxes, since they will be exactly mimicked by our discrete analysis in Section 5. We first introduce the following notation for the fluxes, and split them into the following components:

$$\mathbf{f} = \begin{bmatrix} f_1(q_3) \\ f_2(q_4) \\ f_{3,a}(q_1, q_3) + g_n f_{3,g}(q_1) \\ f_{4,a}(q_2, q_4) + g_n f_{4,g}(q_2) \end{bmatrix}. \quad (25)$$

Here $f_{3,a}$ and $f_{4,a}$ are the momentum advection terms, and $f_{3,g}$ and $f_{4,g}$ are the level gradient terms (divided by g_n). These fluxes and the definitions for \mathbf{v} (17) and ψ (24) can be substituted in (23). The resulting condition is first split into two conditions: one condition proportional to g_n , and one not proportional to g_n . This is done on the basis that the mass and momentum advection terms *do not* depend on g_n in the continuous case (see (2)), and *should not* depend on g_n in the discrete case. These two conditions are split again on the basis that f_1 and f_3 should not depend on q_2 and q_4 , and f_2 and f_4 should not depend on q_1 and q_3 . We obtain the following four conditions:

$$-\frac{\partial}{\partial s} \left(\frac{1}{2} \frac{q_3^2}{q_1^2} \right) f_1 + \frac{\partial}{\partial s} \left(\frac{q_3}{q_1} \right) f_{3,a} = 0, \quad (26a)$$

$$-\frac{\partial}{\partial s} \left(\frac{1}{2} \frac{q_4^2}{q_2^2} \right) f_2 + \frac{\partial}{\partial s} \left(\frac{q_4}{q_2} \right) f_{4,a} = 0, \quad (26b)$$

$$\frac{\partial}{\partial s} (g_n (H - H_U)) f_1 + \frac{\partial}{\partial s} \left(\frac{q_3}{q_1} \right) g_n f_{3,g} = -\frac{\partial}{\partial s} \left(\rho_U g_n \hat{H}_U \frac{q_3}{q_1} \right), \quad (26c)$$

$$\frac{\partial}{\partial s} (g_n H_L) f_2 + \frac{\partial}{\partial s} \left(\frac{q_4}{q_2} \right) g_n f_{4,g} = -\frac{\partial}{\partial s} \left(\rho_L g_n \hat{H}_L \frac{q_4}{q_2} \right). \quad (26d)$$

As mentioned, these equations are by construction satisfied by the flux vector (2). One important remark is that after we reformulate in terms of ψ , the geometric conditions (21) encountered in Section 3.2 still need to be satisfied in order for (26c) and (26d) to hold.

3.4. Comparison of the energy and energy fluxes to those of other models

Here we compare the expressions obtained for e and h to results from literature for other models, focusing on the case of a channel geometry. The expression (16) for e for the channel geometry can be obtained by substitution of the channel-specific evaluations of \tilde{H}_U and \tilde{H}_L (Appendix A):

$$e_{\text{ch}} = \rho_U g_n A_U (A - \frac{1}{2} A_U) + \frac{1}{2} \rho_L g_n A_L^2 + \frac{1}{2} \rho_U u_U^2 A_U + \frac{1}{2} \rho_L u_L^2 A_L. \quad (27)$$

For a single layer fluid, such as the single layer SWE, only the third and fifth terms remain, and they are consistent with the SWE entropy function as discussed in [15] (without channel inclination).

To compare with two-layer SWE theory, we rewrite (27) using the volume constraint (4) to obtain

$$e_{\text{AK}} = \frac{1}{2} \rho_U g_n A_U^2 + \rho_U g_n A_U A_L + \frac{1}{2} \rho_L g_n A_L^2 + \frac{1}{2} \rho_L u_L^2 A_L + \frac{1}{2} \rho_U u_U^2 A_U.$$

This is the expression presented by Abgrall and Karni (AK) [25] and Fjordholm [16] as an entropy function for the two-layer SWE. The energy found in the present study can therefore be seen as a generalization of the two-layer SWE energy to arbitrary duct geometries.

When comparing our energy flux h for the TFM to the one for the two-layer SWE, it should be realized that the two-layer SWE can be obtained from the TFM by the choice $p = \rho_U g_n H_U$. This means that the pressure flux j of the TFM needs to be added to h in order to compare with the SWE expressions. In our notation, the two-layer SWE entropy flux given by [25] is

$$\begin{aligned} h_{\text{AK}} &= \rho_U g_n u_U A_U^2 + \rho_U g_n (u_U + u_L) A_U A_L + \rho_L g_n u_L A_L^2 \\ &\quad + \frac{1}{2} \rho_U u_U^3 A_U + \frac{1}{2} \rho_L u_L^3 A_L. \end{aligned}$$

Our expression for h for a channel is given by

$$h_{\text{ch}} = \rho_U g_n u_U A_U (A - A_U) + \rho_L g_n u_L A_L^2 + \frac{1}{2} \rho_U u_U^3 A_U + \frac{1}{2} \rho_L u_L^3 A_L.$$

Upon adding $j = Qp = Q\rho_U g_n A_U$ to h_{ch} , and after some rewriting, we see that our TFM energy flux is consistent with the two-layer SWE entropy flux:

$$h_{\text{AK}} = h_{\text{ch}} + j.$$

To conclude, our proposed energy (16) and energy fluxes (18) and (19) can be seen as a generalization of the two-layer SWE energy and energy flux to arbitrary duct cross sections.

4. Discretization of the governing equations

4.1. Semi-discrete model equations

The system of equations (1) is discretized using a finite volume method on a uniform staggered grid, sketched in Fig. 2. This discretization naturally conserves mass for each fluid separately, and momentum for both fluids combined. The first two components of \mathbf{q} (the phase masses) and the pressure are defined at the centers of N_p pressure volumes, which have a cell size of $\Delta s = L/N_p$. The last two components of \mathbf{q} (the phase momenta) are defined at the centers of N_u velocity volumes.

On this staggered grid we define a local discrete vector of unknowns as follows:

$$\mathbf{q}_i(t) := \begin{bmatrix} q_{1,i}(t) \\ q_{2,i}(t) \\ q_{3,i-1/2}(t) \\ q_{4,i-1/2}(t) \end{bmatrix} = \begin{bmatrix} (\rho_U A_U \Delta s)_i \\ (\rho_L A_L \Delta s)_i \\ (\rho_U A_U u_U \Delta s)_{i-1/2} \\ (\rho_L A_L u_L \Delta s)_{i-1/2} \end{bmatrix} = \begin{bmatrix} \rho_U A_{U,i} \Delta s \\ \rho_L A_{L,i} \Delta s \\ \rho_U \bar{A}_{U,i-1/2} u_{U,i-1/2} \Delta s \\ \rho_L \bar{A}_{L,i-1/2} u_{L,i-1/2} \Delta s \end{bmatrix}.$$

The choice of using $i - 1/2$ in the definition of \mathbf{q}_i instead of $i + 1/2$ is arbitrary. Note that $q_{1,i}(t) \approx q_1(s_i, t)$ (and similar for the other entries); the notation is on purpose kept very close to the notation of the continuous model, but can be distinguished due to the extra index which the discrete variables carry. Another notable difference is that the cell sizes are included in the discrete unknowns, so that they have units of mass and momentum.

The last equality in the above equation describes the relations of the discrete conservative variables to the discrete primitive variables (cross-sections and velocities). Here we have introduced the following notation for interpolation operators [17]:

$$\bar{a}_{i-1/2} := \frac{1}{2} (a_{i-1} + a_i) \quad \bar{a}_i := \frac{1}{2} (a_{i-1/2} + a_{i+1/2}). \quad (28)$$

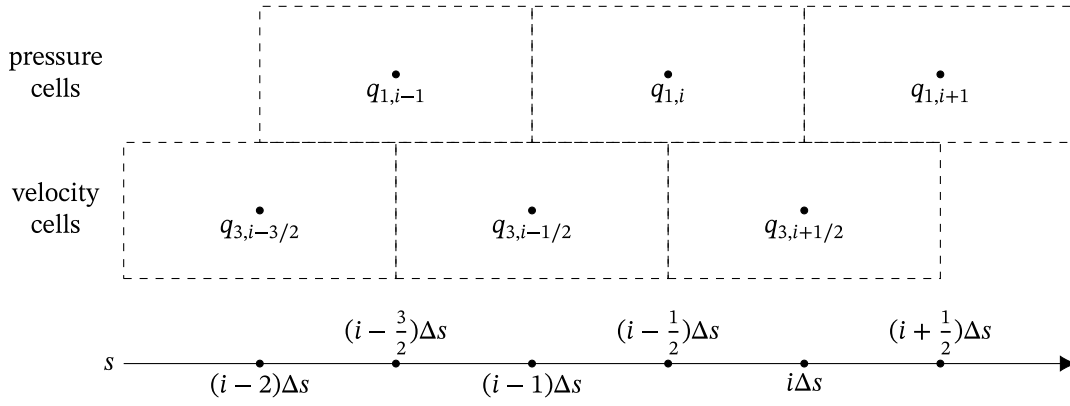


Fig. 2. Staggered grid layout.

The numerical scheme is implemented in terms of the conservative variables $q_{1,i}$ through $q_{4,i-1/2}$, but the primitive variables can be extracted in post-processing according to the given relations.

The notation with \mathbf{q}_i as a discrete local vector of unknowns allows us to write the discrete scheme in vector form as

$$\frac{d\mathbf{q}_i}{dt} + (\mathbf{f}_{i+1/2} - \mathbf{f}_{i-1/2}) + \mathbf{d}_i (p_i - p_{i-1}) = \mathbf{0}. \quad (29)$$

Here, we have defined $\mathbf{f}_{i-1/2}$ as

$$\mathbf{f}_{i-1/2}(\mathbf{q}_{i-2}, \mathbf{q}_{i-1}, \mathbf{q}_i) := \begin{bmatrix} f_{1,i-1/2}(\mathbf{q}_i) \\ f_{2,i-1/2}(\mathbf{q}_i) \\ f_{3,i-1}(\mathbf{q}_{i-2}, \mathbf{q}_{i-1}, \mathbf{q}_i) \\ f_{4,i-1}(\mathbf{q}_{i-2}, \mathbf{q}_{i-1}, \mathbf{q}_i) \end{bmatrix},$$

and

$$\mathbf{d}_i(\mathbf{q}_{i-1}, \mathbf{q}_i) := \begin{bmatrix} 0 \\ 0 \\ d_{3,i-1/2}(\mathbf{q}_{i-1}, \mathbf{q}_i) \\ d_{4,i-1/2}(\mathbf{q}_{i-1}, \mathbf{q}_i) \end{bmatrix}.$$

The numerical fluxes and numerical pressure terms are left undefined in this section, because we will define them based on the requirement of energy conservation, in Section 5.

The pressure terms are non-conservative and are not written as the difference between an inflow and an outflow of the finite volume cell. However, with the staggered grid employed here, one can see that $q_{3,i-1/2}$ and $q_{4,i-1/2}$ are directly and naturally connected to the pressure at the neighboring grid cells. Analogous to the incompressible (multi-dimensional) single-phase Navier–Stokes equations (for which staggered grids are known to lead to strong coupling), this pressure–velocity coupling is necessary to prevent checkerboard patterns, and would be much more difficult to achieve on a collocated grid.

The system is closed by the volume constraint (compare to (4)):

$$\frac{q_{1,i}}{\rho_U \Delta s} + \frac{q_{2,i}}{\rho_L \Delta s} = A, \quad (30)$$

which implies the volumetric flow constraint (compare to (5))

$$Q_{i+1/2} - Q_{i-1/2} := \frac{q_{3,i+1/2}}{\rho_U \Delta s} - \frac{q_{3,i-1/2}}{\rho_U \Delta s} + \frac{q_{4,i+1/2}}{\rho_L \Delta s} - \frac{q_{4,i-1/2}}{\rho_L \Delta s} = 0, \quad (31)$$

so that $Q_{i+1/2} = Q_{i-1/2} = Q(t)$, like in the continuous case. This step can only be made if we choose $f_{1,i-1/2} = q_{3,i-1/2}/\Delta s$ and $f_{2,i-1/2} = q_{4,i-1/2}/\Delta s$, and this will be used as a condition on the form of the numerical fluxes in Section 5.4.

Just as in the continuous case, these constraints are used to set up a Poisson equation for the pressure. The semi-discrete momentum equations are first summed to obtain

$$\frac{1}{\Delta s} \mathbf{l}^T \mathbf{d}_i (p_i - p_{i-1}) = -\frac{1}{\Delta s} \mathbf{l}^T \left(\frac{d\mathbf{q}_i}{dt} + (\mathbf{f}_{i+1/2} - \mathbf{f}_{i-1/2}) \right),$$

$$\text{with } \mathbf{l}^T = \begin{bmatrix} 0 & 0 & \frac{1}{\rho_U} & \frac{1}{\rho_L} \end{bmatrix}.$$

Expanding and substituting the definition of $Q_{i-1/2}$ yields

$$\begin{aligned} & \frac{1}{\Delta s} \left(\frac{d_{3,i-1/2}}{\rho_U} + \frac{d_{4,i-1/2}}{\rho_L} \right) (p_i - p_{i-1}) \\ &= -\frac{dQ_{i-1/2}}{dt} - \frac{1}{\Delta s} \left(\frac{f_{3,i} - f_{3,i-1}}{\rho_U} + \frac{f_{4,i} - f_{4,i-1}}{\rho_L} \right). \end{aligned}$$

After taking the difference between this equation and the same equation for index $i+1/2$, and applying (31), we obtain the discrete version of (7):

$$\begin{aligned} & \frac{1}{\Delta s^2} \left[\left(\frac{d_{3,i+1/2}}{\rho_U} + \frac{d_{4,i+1/2}}{\rho_L} \right) (p_{i+1} - p_i) \right. \\ & \quad \left. - \left(\frac{d_{3,i-1/2}}{\rho_U} + \frac{d_{4,i-1/2}}{\rho_L} \right) (p_i - p_{i-1}) \right] \\ &= -\frac{1}{\Delta s^2} \left(\frac{f_{3,i+1} - 2f_{3,i} + f_{3,i-1}}{\rho_U} + \frac{f_{4,i+1} - 2f_{4,i} + f_{4,i-1}}{\rho_L} \right). \quad (32) \end{aligned}$$

System (29) is discretized in time using the fourth-order semi-implicit Runge–Kutta method described in [23]. At each stage of the Runge–Kutta time step, a predictor–corrector algorithm is applied: the momentum equations are first solved without including the pressure terms, the discrete Poisson equation is solved for the pressure using these intermediate momenta, and the momenta are updated in a projection step using the calculated pressure. This ensures that the volume and volumetric flow constraints are satisfied at all stages. We solve (32) iteratively, using a preconditioned conjugate gradient method, which is run until the relative residual falls below the tolerance (10^{-12}), or until the maximum number of iterations is reached (50). The time integration method is fourth-order accurate for all variables, and requires a restriction to the CFL-number based on the eigenvalues of the TFM.

4.2. Boundary conditions

In the case of periodic boundaries, the domain is divided into N_p pressure volumes and $N_u = N_p$ velocity volumes. There are no special boundary points: the scheme as laid out in Section 4.1 applies everywhere, looping around the domain.

For closed boundaries, there are N_p interior pressure points and $N_u = N_p - 1$ interior velocity points. The first interior pressure node is located at $s = \Delta s/2$, the first interior velocity node is located at $s = \Delta s$, and similarly for the last nodes at the end of the domain [23]. For both the pressure and velocity grids, there are boundary points in addition to the interior points, one at each side of the domain. When calculating the discrete energy on the velocity grid (see Section 5.1), it is important to include the half-volumes between the boundary points and the first and last interior points.

At the boundary points, the mass fluxes ($\rho_U A_U u_U$ and $\rho_L A_L u_L$) are specified, and A_U and A_L follow via an analysis of the characteristics

corresponding to the incoming and outgoing waves at the boundary. In the case of closed boundaries, as used in this work, the mass fluxes are set to zero. Note that the characteristic analysis incorporates the volume constraint (30), and no boundary condition is needed for the pressure (the pressure at the boundaries has no influence on the solution in the interior). For more details on the implementation of the boundary conditions we refer to [23].

5. Energy-conserving spatial discretization of the two-fluid model

5.1. Outline: conditions for discrete energy conservation

In the discrete case, just as in the continuous case, we want to satisfy a local and a global energy equality. The use of a staggered grid instead of the commonly used collocated grid (e.g. [17,24]) makes it straightforward to obtain an energy-conserving discretization of the non-conservative pressure term, but introduces new challenges in terms of the definition of the discrete local energy, which is not unique anymore.

We choose to define the local energy at the velocity grid points, i.e. we choose $e_{i-1/2} = e(\mathbf{q}_{i-1}, \mathbf{q}_i)$, and are aiming for a discrete version of (8):

$$\frac{de_{i-1/2}}{dt} + (h_i - h_{i-1}) + (j_i - j_{i-1}) = 0, \quad (33)$$

with $h_i = h(\mathbf{q}_{i-1}, \mathbf{q}_i)$ and $j_i = j(p_i)$ as the numerical energy fluxes. This choice means that the potential energy terms and q_1 and q_2 in the kinetic energy terms need to be interpolated, but q_3 and q_4 do not require interpolation. With this choice, we obtain energy-conserving expressions for $f_{3,i}$ and $f_{4,i}$ in a constructive manner (after choosing advantageous expressions for $f_{1,i-1/2}$ and $f_{2,i-1/2}$). It is also possible to define the energy at the pressure grid points, and obtain an energy-conserving discretization, but in that case it is necessary to substitute trial solutions for $f_{3,i}$ and $f_{4,i}$, and interpolation of the pressure is required in the expression for j (see the remark at the end of Section 5.4). We would like to emphasize that (33) is *not* being solved as an additional equation; instead it will be shown to be a *consequence* of the discrete mass and momentum equations given in Section 4, if the numerical fluxes and \mathbf{d}_i are chosen appropriately.

If (33) holds, it can be summed over all finite volumes to yield

$$\frac{dE_h}{dt} = \sum_{i=1}^{N_u} \frac{de_{i-1/2}}{dt} = - \sum_{i=1}^{N_u} [(h+j)_i - (h+j)_{i-1}] = 0,$$

where the last equality should hold in the case of periodic or closed boundaries. Here we have defined the global discrete energy as the discrete counterpart of (10):

$$E_h(t) := \sum_{i=1}^{N_u} e_{i-1/2}.$$

Like in the continuous case, the art is to find expressions for $e_{i-1/2}$, h_i and j_i such that Eq. (33) is satisfied. In addition, the numerical flux $\mathbf{f}_{i-1/2}$ needs to be constructed. We will outline the steps to obtain these quantities in a manner parallel to the continuous derivation in Section 3.

First, we postulate an energy

$$e_{i-1/2} = e_{i-1/2}(\mathbf{q}_{i-1}, \mathbf{q}_i), \quad (34)$$

which will be based on the energy found for the continuous case. Note that the dependence could be expanded to additional grid points if required, but we will introduce an energy for which this is not necessary. Second, we calculate the vectors of entropy variables, defined as

$$\mathbf{v}_{i-1/2,i-1} := \left[\frac{\partial e_{i-1/2}}{\partial \mathbf{q}_{i-1}} \right]^T, \quad \mathbf{v}_{i-1/2,i} := \left[\frac{\partial e_{i-1/2}}{\partial \mathbf{q}_i} \right]^T.$$

Here the first index refers to the index of the energy, and the second index refers to the conservative variables to which derivatives are taken.

For the energy given by (34), the time derivative can be expressed as

$$\frac{de_{i-1/2}}{dt} = \langle \mathbf{v}_{i-1/2,i-1}, \frac{d\mathbf{q}_{i-1}}{dt} \rangle + \langle \mathbf{v}_{i-1/2,i}, \frac{d\mathbf{q}_i}{dt} \rangle, \quad (35)$$

Here the brackets represent dot products over the vectors (at a certain grid point), just as in the continuous case. The right-hand side of Eq. (35) follows by substituting Eq. (29) for i and $i-1$:

$$\begin{aligned} \frac{de_{i-1/2}}{dt} = & \langle \mathbf{v}_{i-1/2,i-1}, \llbracket \mathbf{f}_{i-1} \rrbracket \rangle + \langle \mathbf{v}_{i-1/2,i-1}, \mathbf{d}_{i-1} \rrbracket p_{i-3/2} \rrbracket \\ & + \langle \mathbf{v}_{i-1/2,i}, \llbracket \mathbf{f}_i \rrbracket \rangle + \langle \mathbf{v}_{i-1/2,i}, \mathbf{d}_i \rrbracket p_{i-1/2} \rrbracket, \end{aligned} \quad (36)$$

where we have introduced the following notation for jump operators [17]:

$$\llbracket a_{i-1/2} \rrbracket := a_i - a_{i-1}, \quad \llbracket a_i \rrbracket := a_{i+1/2} - a_{i-1/2}. \quad (37)$$

Comparing with (33) we see that the energy fluxes h_i and j_i need to satisfy

$$\langle \mathbf{v}_{i-1/2,i-1}, \llbracket \mathbf{f}_{i-1} \rrbracket \rangle + \langle \mathbf{v}_{i-1/2,i}, \llbracket \mathbf{f}_i \rrbracket \rangle = \llbracket h_{i-1/2} \rrbracket, \quad (38)$$

$$\langle \mathbf{v}_{i-1/2,i-1}, \mathbf{d}_{i-1} \rrbracket p_{i-3/2} \rrbracket + \langle \mathbf{v}_{i-1/2,i}, \mathbf{d}_i \rrbracket p_{i-1/2} \rrbracket = \llbracket j_{i-1/2} \rrbracket. \quad (39)$$

These conditions are analogous to (12) and (14) for the continuous case, with discrete jumps corresponding to derivatives with respect to s . Together, conditions (38) and (39) guarantee that (36) can be written as (33), thus proving conservation of the discrete local energy (34).

The challenge is to find the proper combination of discrete expressions for $e_{i-1/2}$, h_i , j_i , and $\mathbf{f}_{i-1/2}$, which are consistent approximations to their continuous counterparts, in such a way that the local energy conservation equation is satisfied. This is a difficult problem, since we have multiple degrees of freedom ($e_{i-1/2}$, h_i , j_i , and $\mathbf{f}_{i-1/2}$), and the solution might not be unique. To simplify the construction, we will use the concept of entropy potential introduced in Section 3.3: after choosing a certain $e_{i-1/2}$ and $\psi_{i-1/2}$, this yields straightforward conditions on the fluxes $\mathbf{f}_{i-1/2}$ to be energy-conserving.

5.2. Choice of discrete energy and energy fluxes

In this section we propose an energy $e_{i-1/2}$, and verify that this energy is conserved by the pressure terms of the discrete model (energy conservation for the flux terms is treated in Sections 5.3 and 5.4). Recalling the continuous energy (16), we define a discrete energy $e_{i-1/2} = e_{i-1/2}(\mathbf{q}_{i-1}, \mathbf{q}_i)$:

$$\begin{aligned} e_{i-1/2} = & \rho_U g_n \widetilde{H}_{U,i-1/2} \Delta s + \rho_L g_n \widetilde{H}_{L,i-1/2} \Delta s + \frac{1}{2} \frac{q_{3,i-1/2}^2}{\bar{q}_{1,i-1/2}} + \frac{1}{2} \frac{q_{4,i-1/2}^2}{\bar{q}_{2,i-1/2}} \\ = & \rho_U g_n \widetilde{H}_{U,i-1/2} \Delta s + \rho_L g_n \widetilde{H}_{L,i-1/2} \Delta s + \frac{1}{2} \rho_U \bar{A}_{U,i-1/2} u_{U,i-1/2}^2 \Delta s \\ & + \frac{1}{2} \rho_L \bar{A}_{L,i-1/2} u_{L,i-1/2}^2 \Delta s. \end{aligned} \quad (40)$$

Other choices are possible because on a staggered grid interpolation is required, and the interpolation may be carried out in various different ways.³ Our choice (40) is one of the most straightforward choices for the energy that is consistent with the continuous definition, when the energy is defined at the velocity grid points, and leads to an elegant form of the energy-conserving discretization (see also the remark at the end of Section 5.4).

³ In Appendix B we will show that the same results can be obtained with a global energy analysis, in which interpolation of the local potential energy to the velocity grid points is not needed.

We use identities given in [Appendix A](#) to calculate the \mathbf{v} vectors. They are given by

$$\mathbf{v}_{i-1/2,i-1} = \left[\frac{\partial e_{i-1/2}}{\partial \mathbf{q}_{i-1}} \right]^T = \begin{bmatrix} -\frac{1}{4} \frac{q_{3,i-1/2}^2}{\bar{q}_{1,i-1/2}^2} + \frac{1}{2} g_n (H - H_{U,i-1}) \\ -\frac{1}{4} \frac{q_{4,i-1/2}^2}{\bar{q}_{2,i-1/2}^2} + \frac{1}{2} g_n H_{L,i-1} \\ 0 \\ 0 \end{bmatrix}, \quad (41)$$

and

$$\mathbf{v}_{i-1/2,i} = \left[\frac{\partial e_{i-1/2}}{\partial \mathbf{q}_i} \right]^T = \begin{bmatrix} -\frac{1}{4} \frac{q_{3,i-1/2}^2}{\bar{q}_{1,i-1/2}^2} + \frac{1}{2} g_n (H - H_{U,i}) \\ -\frac{1}{4} \frac{q_{4,i-1/2}^2}{\bar{q}_{2,i-1/2}^2} + \frac{1}{2} g_n H_{L,i} \\ \frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \\ \frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} \end{bmatrix}, \quad (42)$$

and their sum is consistent with (17).

The pressure terms in (36) need to satisfy condition (39), which can be rewritten to obtain the discrete version of (15):

$$\begin{aligned} \llbracket j_{i-1/2} \rrbracket &= \overline{\langle \mathbf{v}_{i-1/2,i-1}, \mathbf{d}_{i-1} \rangle p_{i-3/2}} - \overline{\langle \mathbf{v}_{i-1/2,i-1}, \mathbf{d}_{i-1} \rangle p_{i-3/2}} \\ &+ \overline{\langle \mathbf{v}_{i-1/2,i}, \mathbf{d}_i \rangle p_{i-1/2}} - \overline{\langle \mathbf{v}_{i-1/2,i}, \mathbf{d}_i \rangle p_{i-1/2}}. \end{aligned} \quad (43)$$

On a staggered grid, it is straightforward to satisfy this condition by choosing for \mathbf{d}_i

$$\mathbf{d}_i = \frac{1}{\Delta s} \begin{bmatrix} 0 \\ 0 \\ \frac{\rho_U}{\bar{q}_{1,i-1/2}} \\ \frac{\rho_L}{\bar{q}_{2,i-1/2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \bar{A}_{U,i-1/2} \\ \bar{A}_{L,i-1/2} \end{bmatrix}, \quad (44)$$

since with this choice we have

$$\langle \mathbf{v}_{i-1/2,i-1}, \mathbf{d}_{i-1} \rangle = 0,$$

and

$$\langle \mathbf{v}_{i-1/2,i}, \mathbf{d}_i \rangle = \left(\frac{q_{3,i-1/2}}{\rho_U \Delta s} + \frac{q_{4,i-1/2}}{\rho_L \Delta s} \right) = Q_{i-1/2}.$$

Consequently, condition (43) can be written with the volumetric flow constraint (31) as

$$\llbracket j_{i-1/2} \rrbracket = \overline{\llbracket Q_{i-1/2} p_{i-1/2} \rrbracket} = \overline{\llbracket Q(t) p_{i-1/2} \rrbracket}, \quad (45)$$

so that (43) (and (39)) is satisfied when j is given by

$$j_i = Q(t) p_i. \quad (46)$$

Note that our constraint-consistent time integration method enforces that the volumetric flow constraint is satisfied up to machine precision [23].

5.3. Reformulation in terms of the entropy potential and conditions on numerical fluxes

The objective of finding energy-conserving numerical fluxes is better served by reformulating condition (38) in terms of the entropy potential, because this results in an alternative, constructive, condition for finding energy-conserving fluxes. The fluxes are then based on the entropy potential $\psi_{i-1/2}$ instead of the energy flux h_i . Similar to Section 3.3, we rewrite the left-hand side of (38) as:

$$\begin{aligned} &\langle \mathbf{v}_{i-1/2,i-1}, \llbracket \mathbf{f}_{i-1} \rrbracket \rangle + \langle \mathbf{v}_{i-1/2,i}, \llbracket \mathbf{f}_i \rrbracket \rangle \\ &= \overline{\langle \bar{\mathbf{v}}_{i-1/2,i-1}, \mathbf{f}_{i-1} \rangle} + \overline{\langle \bar{\mathbf{v}}_{i-1/2,i}, \mathbf{f}_i \rangle} \\ &- \overline{\langle \mathbf{v}_{i-1/2,i-1}, \mathbf{f}_{i-1} \rangle} - \overline{\langle \mathbf{v}_{i-1/2,i}, \mathbf{f}_i \rangle}, \end{aligned} \quad (47)$$

which can be interpreted as a discrete version of the product rule $\langle \mathbf{v}, \frac{\partial \mathbf{f}}{\partial s} \rangle = \frac{\partial}{\partial s} \langle \mathbf{v}, \mathbf{f} \rangle - \langle \frac{\partial \mathbf{v}}{\partial s}, \mathbf{f} \rangle$. We have made use of the following definitions:

$$\begin{aligned} \bar{\mathbf{v}}_{i,i-1/2} &= \frac{1}{2} (\mathbf{v}_{i-1/2,i-1} + \mathbf{v}_{i+1/2,i}) & \bar{\mathbf{v}}_{i,i+1/2} &= \frac{1}{2} (\mathbf{v}_{i-1/2,i} + \mathbf{v}_{i+1/2,i+1}), \\ \llbracket \mathbf{v}_{i,i-1/2} \rrbracket &= \mathbf{v}_{i+1/2,i} - \mathbf{v}_{i-1/2,i-1} & \llbracket \mathbf{v}_{i,i+1/2} \rrbracket &= \mathbf{v}_{i+1/2,i+1} - \mathbf{v}_{i-1/2,i}. \end{aligned}$$

These definitions are such that we only interpolate or take jumps between \mathbf{v} vectors with the same relative indices.

Instead of directly choosing ψ , it is more natural to use the last terms in (47) and define the jump in ψ (similar to (23)) as

$$\llbracket \psi_i \rrbracket = \langle \llbracket \mathbf{v}_{i,i-1/2} \rrbracket, \mathbf{f}_{i-1/2} \rangle + \langle \llbracket \mathbf{v}_{i,i+1/2} \rrbracket, \mathbf{f}_{i+1/2} \rangle, \quad (48)$$

since this leads to the following ‘implied’ definition of ψ :

$$\begin{aligned} \overline{\llbracket \psi_{i-1/2} \rrbracket} &= \overline{\llbracket \psi_{i-1/2} \rrbracket} = \frac{1}{2} \overline{\llbracket \psi_{i-1/2} \rrbracket} + \frac{1}{2} \overline{\llbracket \psi_i \rrbracket} \\ &= \overline{\langle \bar{\mathbf{v}}_{i-1/2,i-1}, \mathbf{f}_{i-1/2} \rangle} + \overline{\langle \bar{\mathbf{v}}_{i-1/2,i}, \mathbf{f}_i \rangle} - \overline{\llbracket h_{i-1/2} \rrbracket}, \end{aligned} \quad (49)$$

which is consistent with (22). The advantage of (48) over (38) is that we have a condition on the flux itself, rather than on the jump in the flux. Once $e_{i-1/2}$ and $\psi_{i-1/2}$ have been chosen and $\mathbf{f}_{i-1/2}$ has been derived, h_i can be determined from

$$h_i = \langle \bar{\mathbf{v}}_{i,i-1/2}, \mathbf{f}_{i-1/2} \rangle + \langle \bar{\mathbf{v}}_{i,i+1/2}, \mathbf{f}_{i+1/2} \rangle - \bar{\psi}_i. \quad (50)$$

We note that this expression is similar to the collocated grid setting, where one has $h_{i-1/2} = \langle \bar{\mathbf{v}}_{i-1/2}, \mathbf{f}_{i-1/2} \rangle - \bar{\psi}_{i-1/2}$ [17]. The difference lies in a shift in indices (because our energy is defined at $i-1/2$ instead of i), and in the way the term $\langle \mathbf{v}, \mathbf{f} \rangle$ is approximated.

We propose now the following discrete entropy potential for the equations:

$$\begin{aligned} \psi_{i-1/2}(\mathbf{q}_{i-1}, \mathbf{q}_i) &= -\rho_U g_n \widehat{H}_{U,i-1/2} \frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} - \rho_L g_n \widehat{H}_{L,i-1/2} \frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} \\ &= -\rho_U g_n \widehat{H}_{U,i-1/2} u_{U,i-1/2} - \rho_L g_n \widehat{H}_{L,i-1/2} u_{L,i-1/2}. \end{aligned} \quad (51)$$

This is a straightforward discretization of (24). Given the expressions for \mathbf{v} ((41) and (42)), condition (48) can now be evaluated to yield the numerical fluxes $\mathbf{f}_{i-1/2}$. In order to be able to derive from the (scalar) condition (48) multiple equations for the individual numerical fluxes, we split $f_{3,i-1}$ and $f_{4,i-1}$ into an advective component (denoted by subscript a) and a level gradient (or gravity) component (denoted by subscript g): $f_{3,i-1} = f_{3,i-1,a} + g_n f_{3,i-1,g}$ and $f_{4,i-1} = f_{4,i-1,a} + g_n f_{4,i-1,g}$.

As a consequence, condition (48) can be split into the following four separate conditions by collecting terms featuring g_n and those not featuring g_n , and by using the functional dependencies assumed for the fluxes:

$$-\left[\frac{1}{2} \frac{q_{3,i}^2}{\bar{q}_{1,i}^2} \right] \bar{f}_{1,i} + \left[\frac{q_{3,i}}{\bar{q}_{1,i}} \right] f_{3,i,a} = 0, \quad (52a)$$

$$-\left[\frac{1}{2} \frac{q_{4,i}^2}{\bar{q}_{2,i}^2} \right] \bar{f}_{2,i} + \left[\frac{q_{4,i}}{\bar{q}_{2,i}} \right] f_{4,i,a} = 0, \quad (52b)$$

$$\overline{\langle g_n (H - H_{U,i}) \rangle f_{1,i}} + \left[\frac{q_{3,i}}{\bar{q}_{1,i}} \right] g_n f_{3,i,g} = - \left[\rho_U g_n \widehat{H}_{U,i} \frac{q_{3,i}}{\bar{q}_{1,i}} \right], \quad (52c)$$

$$\overline{\langle g_n H_{L,i} \rangle f_{2,i}} + \left[\frac{q_{4,i}}{\bar{q}_{2,i}} \right] g_n f_{4,i,g} = - \left[\rho_L g_n \widehat{H}_{L,i} \frac{q_{4,i}}{\bar{q}_{2,i}} \right]. \quad (52d)$$

These conditions have been obtained analogously to their continuous equivalents (26). In the continuous case, the fluxes were known and these conditions were satisfied by construction. In the discrete case, these conditions will be used in the next section to determine the numerical fluxes.

5.4. Derivation of energy-conserving numerical fluxes for the TFM

System (52) is a system of four equations for six unknowns. This leaves the mass fluxes $f_{1,i-1/2}$ and $f_{2,i-1/2}$ free to be chosen arbitrarily, with the expressions for $f_{3,i}$ and $f_{4,i}$ depending on this choice. This observation has been made previously for the (isothermal) compressible Euler equations [28,29], of which the advective terms have a similar mathematical form to those of the TFM. We take the following choice for $f_{1,i-1/2}$ and $f_{2,i-1/2}$:

$$f_{1,i-1/2} = \frac{q_{3,i-1/2}}{\Delta s} \quad \text{and} \quad f_{2,i-1/2} = \frac{q_{4,i-1/2}}{\Delta s}, \quad (53)$$

which is motivated by the fact that it requires no interpolation (so that the mass and momentum equations are directly coupled), and moreover is such that the discrete Poisson equation (32) follows naturally from the discrete volumetric flow constraint (31) (which is used in our time integration method [23]).

Substituting $f_{1,i-1/2}$ in (52a) and $f_{2,i-1/2}$ in (52b) yields directly

$$f_{3,i,a} = \frac{1}{\Delta s} \left(\frac{q_{3,i}}{\bar{q}_{1,i}} \right) \bar{q}_{3,i} \quad \text{and} \quad f_{4,i,a} = \frac{1}{\Delta s} \left(\frac{q_{4,i}}{\bar{q}_{2,i}} \right) \bar{q}_{4,i}. \quad (54)$$

To get the gravity component of $f_{3,i}$ and $f_{4,i}$, substitution of $f_{2,i-1/2}$ in (52d) leads to

$$\begin{aligned} \frac{1}{2} \left[\left[H_{L,i-1/2} \right] \right] \frac{q_{4,i-1/2}}{\Delta s} + \frac{1}{2} \left[\left[H_{L,i+1/2} \right] \right] \frac{q_{4,i+1/2}}{\Delta s} + \left[\left[\frac{q_{4,i}}{\bar{q}_{2,i}} \right] \right] f_{4,i,g} \\ = \rho_L \bar{H}_{L,i-1/2} \frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} - \rho_L \bar{H}_{L,i+1/2} \frac{q_{4,i+1/2}}{\bar{q}_{2,i+1/2}}. \end{aligned}$$

After significant rewriting, this yields the following expression for the gravity component of $f_{4,i}$:

$$f_{4,i,g} = -\rho_L \hat{H}_{L,i} - \left[\left(\left[\frac{\hat{H}_{L,i}}{\bar{q}_{2,i}} \right] + \left[\frac{H_{L,i}}{\Delta s} \right] \right) q_{4,i} \right] \left[\left[\frac{q_{4,i}}{\bar{q}_{2,i}} \right] \right]^{-1}, \quad (55)$$

and a similar expression holds for $f_{3,i,g}$ (but with a minus sign instead of the plus sign).

The first term on the right-hand side is easily recognized as the discrete counterpart of $-\rho_L \hat{H}_L$. In order for the discrete expression to be practical and match the continuous expression, the second term must vanish, and we require the following conditions to be satisfied:

$$\begin{aligned} \left[\left[\hat{H}_{U,i-1/2} \right] \right] &= \frac{\bar{q}_{1,i-1/2}}{\rho_U \Delta s} \left[\left[H_{U,i-1/2} \right] \right], \\ \left[\left[\hat{H}_{L,i-1/2} \right] \right] &= -\frac{\bar{q}_{2,i-1/2}}{\rho_L \Delta s} \left[\left[H_{L,i-1/2} \right] \right]. \end{aligned} \quad (56)$$

In the continuous case a continuous version of these conditions, given by (21), is also required, and these can be shown to be satisfied exactly via manipulation of the continuous derivatives. The same manipulation is not possible with discrete jumps, so that in the discrete case these conditions are not satisfied in general, and the second term in (55) does not generally vanish. This means that we cannot obtain a practical energy-conserving discretization for arbitrary geometries (at least not with the conventional staggered-grid finite volume method that we have employed).

Even though conditions (56) are not generally exactly satisfied in the discrete case, we can show that they are approximately satisfied for arbitrary duct geometries, and that they are exactly satisfied for specific geometries such as a channel. This can be shown by evaluating both sides of (56) using Taylor series. We expand $\hat{H}_{L,i-1}$ and $H_{L,i-1}$ into Taylor series around $A_L = A_{L,i}$, and expand $A_{L,i-1}$ around $s = s_i$. These Taylor series are combined to obtain expressions for $\left[\left[\hat{H}_{L,i-1/2} \right] \right]$,

$\left[\left[H_{L,i-1/2} \right] \right]$, and $\left[\left[A_{L,i-1/2} \right] \right]$. With these expressions the left-hand side of (56) (for the lower fluid) evaluates to

$$\begin{aligned} \left[\left[\hat{H}_{L,i-1/2} \right] \right] &= - \left(\frac{d\hat{H}_L}{dA_L} \right)_i (A_{L,i-1} - A_{L,i}) \\ &\quad - \frac{1}{2} \left(\frac{d^2\hat{H}_L}{dA_L^2} \right)_i (A_{L,i-1} - A_{L,i})^2 \\ &\quad - \frac{1}{6} \left(\frac{d^3\hat{H}_L}{dA_L^3} \right)_i (A_{L,i-1} - A_{L,i})^3 + O(\Delta s^4), \end{aligned} \quad (57)$$

where $(\cdot)_i$ indicates (\cdot) evaluated at $A_{L,i}$. The right-hand side of (56) evaluates to

$$\begin{aligned} -\frac{\bar{q}_{2,i-1/2}}{\rho_L \Delta s} \left[\left[H_{L,i-1/2} \right] \right] &= \frac{1}{2} \left(\frac{dH_L}{dA_L} \right)_i (A_{L,i-1}^2 - A_{L,i}^2) \\ &\quad + \frac{1}{4} \left(\frac{d^2H_L}{dA_L^2} \right)_i (A_{L,i-1} + A_{L,i}) (A_{L,i-1} - A_{L,i})^2 \\ &\quad + \frac{1}{12} \left(\frac{d^3H_L}{dA_L^3} \right)_i (A_{L,i-1} + A_{L,i}) (A_{L,i-1} - A_{L,i})^3 + O(\Delta s^4). \end{aligned} \quad (58)$$

At this point we apply relation (A.4) from Appendix A to the discrete quantities used here:

$$\left(\frac{d\hat{H}_L}{dA_L} \right)_i = -A_{L,i} \left(\frac{dH_L}{dA_L} \right)_i,$$

and from this we can derive

$$\begin{aligned} \left(\frac{d^2\hat{H}_L}{dA_L^2} \right)_i &= - \left(\frac{dH_L}{dA_L} \right)_i - A_{L,i} \left(\frac{d^2H_L}{dA_L^2} \right)_i, \\ \left(\frac{d^3\hat{H}_L}{dA_L^3} \right)_i &= -2 \left(\frac{d^2H_L}{dA_L^2} \right)_i - A_{L,i} \left(\frac{d^3H_L}{dA_L^3} \right)_i. \end{aligned}$$

Substitution of these relations in (57), and comparison of the result to (58) yields

$$\begin{aligned} \left[\left[\hat{H}_{L,i-1/2} \right] \right] &= -\frac{\bar{q}_{2,i-1/2}}{\rho_L \Delta s} \left[\left[H_{L,i-1/2} \right] \right] \\ &\quad + \frac{1}{12} \left(\frac{d^2H_L}{dA_L^2} \right)_i (A_{L,i-1} - A_{L,i})^3 + O(\Delta s^4). \end{aligned} \quad (59)$$

This derivation can be carried out with similar results for the upper fluid.

These relations show that for arbitrary duct geometries, the geometric conditions (56) are satisfied only approximately in the discrete case. This stands in contrast to the continuous case, where the equivalent geometric conditions are satisfied exactly (for arbitrary geometries).

Fortunately, for a 2D channel geometry $dH_L/dA_L = 1$ and $d^2H_L/dA_L^2 = 0$, and all higher order derivatives are zero, so in this case (56) is exactly satisfied. This means that the 2D channel geometry is an important special case for which we obtain the following numerical fluxes:

$$f_{3,i,g} = -\rho_U \hat{H}_{U,i} \quad \text{and} \quad f_{4,i,g} = -\rho_L \hat{H}_{L,i}. \quad (60)$$

These fluxes are energy-conserving for other geometries with $d^2H_L/dA_L^2 = 0$, but not for geometries with curved sides, such as the pipe geometry.

The final collection of energy-conserving numerical fluxes is given by (53), (54), and (60). Of these, (53) and (60) are locally exact, and (54) involves second order accurate central interpolation. Together they form the numerical flux vector

$$\mathbf{f}_{i-1/2}(\mathbf{q}_{i-2}, \mathbf{q}_{i-1}, \mathbf{q}_i) = \begin{bmatrix} \frac{q_{3,i-1/2}}{\Delta s} \\ \frac{q_{4,i-1/2}}{\Delta s} \\ \frac{1}{\Delta s} \left(\frac{q_{3,i-1}}{\bar{q}_{1,i-1}} \right) \bar{q}_{3,i-1} - \rho_U g_n \hat{H}_{U,i-1} \\ \frac{1}{\Delta s} \left(\frac{q_{4,i-1}}{\bar{q}_{2,i-1}} \right) \bar{q}_{4,i-1} - \rho_L g_n \hat{H}_{L,i-1} \end{bmatrix}$$

$$= \begin{bmatrix} \rho_U \bar{A}_{U,i-1/2} u_{U,i-1/2} \\ \rho_L \bar{A}_{L,i-1/2} u_{L,i-1/2} \\ \rho_U \bar{u}_{U,i} \left(\bar{A}_{U,i-1} u_{U,i-1} \right) - \rho_U g_n \hat{H}_{U,i-1} \\ \rho_L \bar{u}_{L,i} \left(\bar{A}_{U,i-1} u_{U,i-1} \right) - \rho_L g_n \hat{H}_{L,i-1} \end{bmatrix}. \quad (61)$$

Here the flux is rendered in terms of primitive variables only for ease of interpretation; the implementation of the numerical flux (and of the discrete energy) in the numerical code is completely in terms of the conservative variables.

Remark 1. The difficulty to satisfy condition (56) for arbitrary cross-sectional geometries is not dependent on the choice of $\psi_{i-1/2}$, nor is it due to the interpolation of the potential energy to the velocity grid points (as needed on a staggered grid). This is shown in Appendix B by applying a global energy analysis.

Remark 2. The proposed discrete energy (40) is a consistent approximation to (16) which is conserved by the numerical fluxes given by (61). However, it is not unique. For example, an alternative definition is

$$e_i(\mathbf{q}_{i-1}, \mathbf{q}_i, \mathbf{q}_{i+1}) = \rho_U g_n \tilde{H}_{U,i} \Delta s + \rho_L g_n \tilde{H}_{L,i} \Delta s + \frac{1}{2} \left(\frac{q_{3,i}^2}{q_{1,i}} \right) + \frac{1}{2} \left(\frac{q_{4,i}^2}{q_{2,i}} \right). \quad (62)$$

In this formulation the energy is defined on the pressure grid, and the energy conservation conditions and local energy conservation equation can be adapted to accommodate for this. With a similar change in the entropy potential, it is again possible to derive a set of energy-conserving numerical fluxes, which turn out to be the same as those given by (61). As the issue of the geometric relations also persists with this choice, there seems no clear advantage over our proposed formulation.

Remark 3. It is possible to show that, with our discretization of the advective terms, the advective contribution to the global energy equation (given by a sum over the domain of the left-hand side of (38), minus gravitational terms) can be written as

$$\mathbf{u}_{U,h}^T \mathbf{C}_U \mathbf{u}_{U,h} + \mathbf{u}_{L,h}^T \mathbf{C}_L \mathbf{u}_{L,h},$$

where $\mathbf{u}_{U,h} = [u_{U,1/2} \dots u_{U,N_u-1/2}]^T$ and similar for $\mathbf{u}_{L,h}$, and with $\mathbf{C}_U = \mathbf{C}_U(\mathbf{f}_{1,h}, \mathbf{u}_{U,h})$ and $\mathbf{C}_L = \mathbf{C}_L(\mathbf{f}_{2,h}, \mathbf{u}_{L,h})$ being skew-symmetric matrices. From this it can be shown that the advective terms conserve global energy. In [18], this principle is used to derive a similar discretization of the advective terms for the SWE.

6. Numerical experiments

We perform numerical experiments for a 2D channel geometry, with the goal of verifying conservation of the discrete global energy, as discussed in Section 5.1:

$$\frac{dE_h}{dt} = 0.$$

The model for which we perform the experiments will not include source terms such as wall friction and interface friction, or diffusion, since these would lead to dissipation of energy in the continuous analysis. The test cases are chosen such that no discontinuities appear, for which the continuous analysis is invalid, since this would also necessitate dissipation of energy. Furthermore, the numerical experiments performed in this section will all be in the ‘well-posed regime’ of the TFM, meaning that the initial conditions are chosen such that the eigenvalues of the model are real, and remain so.

We use the discretization as outlined in Section 4, with the numerical fluxes given by (61). The vector \mathbf{d}_i of the pressure term is given by (44). We noted earlier that the scheme is spatially exactly energy-conserving, but not temporally. However, we can still obtain energy

Table 1

Parameters for the Gaussian perturbation test case.

Parameter	Symbol	Value	Units
Lower fluid density	ρ_L	1000	kg m ⁻³
Upper fluid density	ρ_U	780	kg m ⁻³
Acceleration of gravity	g	9.8	m s ⁻²
Channel inclination	ϕ	0	degrees
Domain length	L	1.83	m
Channel height	H	0.03	m
Initial lower fluid hold-up	$\alpha_{L,0}$	0.5	–
Initial lower fluid velocity	$u_{L,0}$	0	m s ⁻¹
Initial upper fluid velocity	$u_{U,0}$	0	m s ⁻¹

conservation by taking the time step sufficiently small. The difference between the initial energy E_h^0 and the final energy $E_h^{N_t}$ after N_t time steps should then be in the order of the machine precision, and we shall term this difference the ‘energy error’.

6.1. Gaussian perturbation in a periodic domain

We consider a test case with periodic boundaries, so that effectively we do not need to take the boundaries into account. We introduce a perturbation in the hold-up $\alpha_L = A_L/A$ of the form

$$\alpha_L(s) = \alpha_{L,0} + \Delta\alpha_L(s), \quad \Delta\alpha_L(s) = \Delta\hat{\alpha}_L \exp \left[-\frac{1}{2} \left(\frac{s - L/2}{\sigma} \right)^2 \right],$$

with $\Delta\hat{\alpha}_L = 0.2$ and $\sigma = L/10$, and L the length of the domain. This produces a Gaussian perturbation centered at the middle of the domain. The initial velocities are left at zero, which ensures exact initial satisfaction of the volumetric flow constraint (6) (in fact, $Q = 0$).

We use parameters similar to those used in the Thorpe experiment [30], as described by [31]. They are given by Table 1. The choice for a high upper fluid density is deliberate: it ensures that all terms in the expression for e , (40), are significant. Additionally, at large density ratios (i.e. low upper fluid density) the flow becomes more violent and the formation of shocks becomes more likely, the treatment of which is outside the scope of this paper.

We employ $N_p = N_u = 40$ finite volumes with $\Delta s = L/N_p$ and let the simulations run until $t = 30$ s, with $\Delta t = 0.001$ s. The perturbation splits symmetrically into a left-traveling and a right-traveling wave, which travel through the periodic boundaries, to eventually come together in the middle and reform the initial perturbation approximately. We show the evolution of the hold-up and velocity in Fig. 3, roughly up to the point that the waves meet at the boundaries of the domain.

In this test case we have a significant exchange between kinetic and potential energy, which can be seen in Fig. 4 (left panel). The total energy is conserved up to machine precision, as can be seen in the right panel of the figure. The mass of each phase and total momentum are also conserved, and the volume constraint and volumetric flow constraint are satisfied, up to machine precision (see also [23]). As time progresses, nonlinear effects start to play a role, leading to more irregular behavior of the potential and kinetic energy as a function of time. The sum of the two stays exactly constant, confirming our theoretical derivations, and showing that our newly proposed numerical fluxes for the TFM lead indeed to an energy-conserving discretization method.

We give further evidence that the energy is conserved exactly by the spatial discretization, and limited only by a temporal error, by plotting the convergence of the energy error with refinement of the time step, and with refinement of the grid. The first plot in Fig. 5 shows a fourth order convergence rate with Δt , in agreement with the fourth order accuracy of the Runge–Kutta time integration method. The convergence continues up to machine precision, which is reached around $\Delta t = 0.001$ s, as was used for the results in Fig. 4. The second plot in Fig. 5 shows that when using a small enough time step, the same (minimal) level of error is obtained, irrespective of spatial resolution.

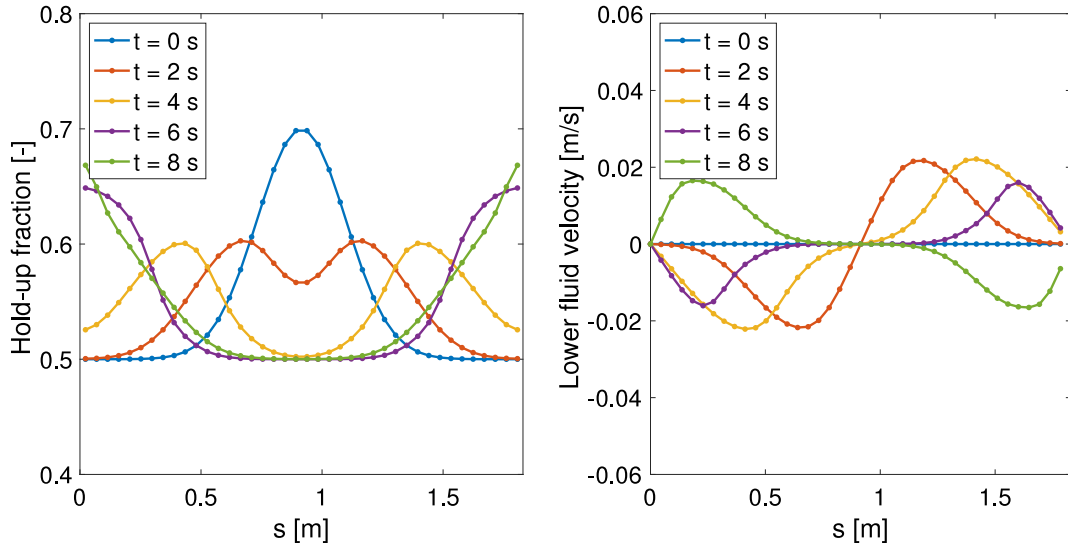


Fig. 3. The initial evolution of the Gaussian perturbation, up to the point that the boundaries are met. Left: lower fluid hold-up. Right: lower fluid velocity.

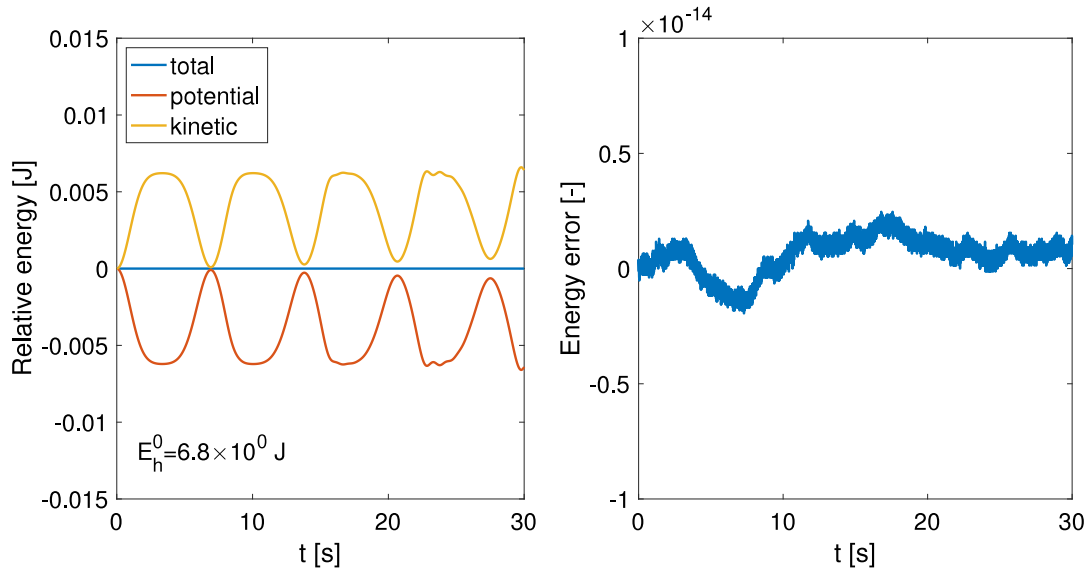


Fig. 4. Conserved quantities for the Gaussian perturbation test case. Left: potential, kinetic and total energy relative to their initial values. Right: $(E_h - E_h^0)/E_h^0$.

6.2. Sloshing in a closed tank

We now consider a test case with closed (solid-wall) boundaries, for which energy conservation is expected to hold because the fluxes h and j involve multiplication with q_3 and q_4 , which are zero at the boundaries. The test case features a closed rectangular tank in which the two fluids are brought out of equilibrium, so that sloshing occurs. The parameters are identical to those of the previous test case, see Table 1, except that the initial condition for the hold-up perturbation is different. It is given by

$$\alpha_L(s) = \alpha_{L,0} + \Delta\alpha_L(s), \quad \Delta\alpha_L(s) = \Delta\hat{\alpha}_L \frac{s - L/2}{L/2},$$

with $\Delta\hat{\alpha}_L = 0.2$. This yields a straight slanted interface, with $\alpha_L = 0.3$ at the left boundary and $\alpha_L = 0.7$ at the right boundary: see Fig. 6.

This is not a typical sloshing case, since the TFM was designed to model long-wavelength phenomena, and indeed we have taken $L \gg H$. Therefore we are not able to explicitly capture typical sloshing phenomena such as wave breaking. However, the effect of such small-scale phenomena on the averaged flow may be included in the model

via closure terms [32]. With accurate closure terms, the TFM can closely match DNS results, as shown in [33]. This is not included here, as this would lead to dissipation of energy and not allow us to show the energy-conserving properties of our proposed numerical discretization.

Like in the first test case, initially the total energy of the system consists of only potential energy. Under the influence of gravity (via the level gradient terms) the interface starts to flatten, which is achieved via a right-running and a left-running wave, that emanate from the left and right boundary, respectively. Around $t = 7$ s the interface is almost completely flat, and all potential energy has been converted into kinetic energy, and the interface starts to slant ('slosh') again in the opposite direction. Fig. 6 shows this behavior up to approximately the point that the lower fluid reaches its maximum height at the left boundary. Note that the evolution of the hold-up fraction is not exactly symmetric, amongst others because the wave speed in the 'deep' part is different from the wave speed in the 'shallow' part. Also in this test case, the mass of each phase is conserved up to machine precision, but there is a (physical) inflow of momentum at the boundaries, due to the level gradient terms.

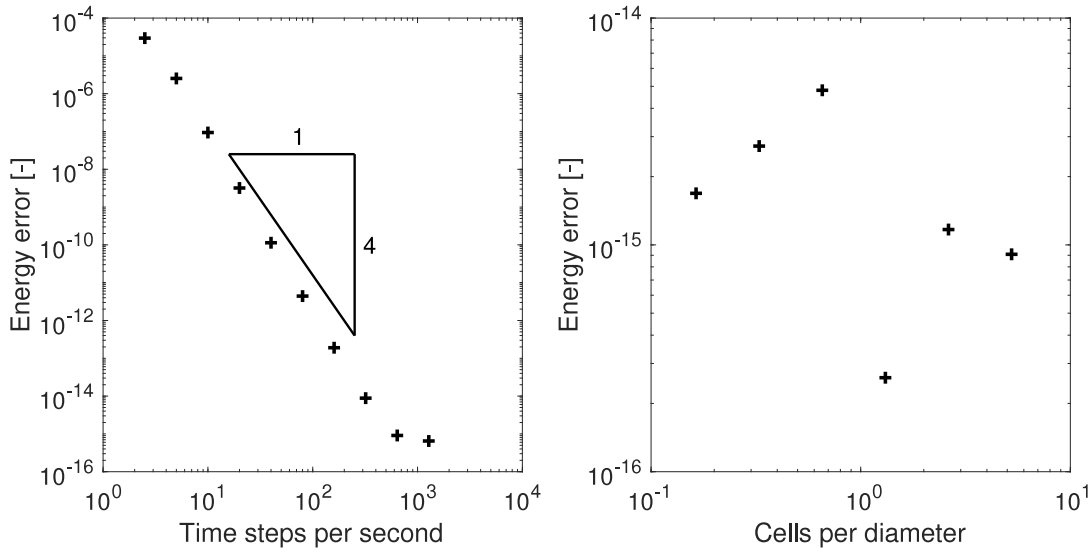


Fig. 5. Convergence of the energy error $(E_h^{N_t} - E_h^0)/E_h^0$ for the Gaussian perturbation test case. Left: constant $N_p = 40$. Right: constant $\Delta t = 0.0001$ s.

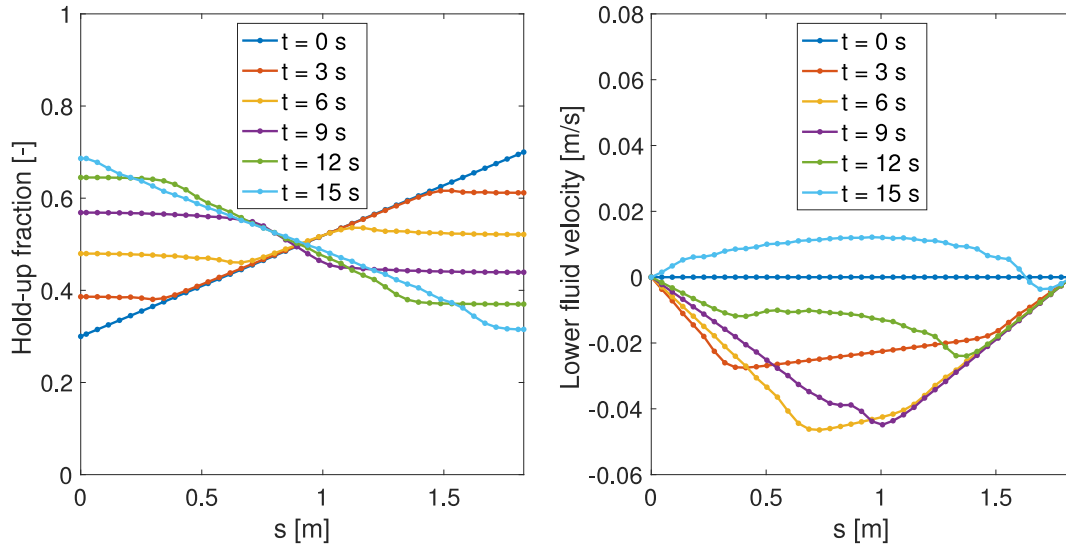


Fig. 6. The initial evolution of the sloshing simulation, approximately up to the point that the lower fluid reaches its maximum height at the left boundary. Left: lower fluid hold-up. Right: lower fluid velocity.

Fig. 7 shows the exchange of potential and kinetic energy as a function of time. Similar to the previous test case, exact energy conservation is achieved with our proposed spatial discretization, if the time step is fine enough (here $\Delta t = 0.005$ s, and $N_p = 40$). If the time step is not fine enough, a (small) energy error is made, which converges with fourth order upon time step refinement, as is shown in Fig. 8. The ability to conserve energy in this closed system is an important step in order to obtain fidelity in the simulation results. Non-energy-conserving schemes, e.g. schemes that dissipate energy, would introduce artificial (numerical) damping of the sloshing movement and incur a loss in the liquid height reached at the boundaries. In a way, the sloshing movement can be compared to a moving pendulum [34], for which it is well-known that conservation of the total energy (the Hamiltonian) is an important property that should be mimicked upon discretization in order to achieve realistic long-time behavior.

Fig. 9 shows results for a test case using the same parameters, but in a circular pipe geometry, with a diameter corresponding to the height of the 2D channel (0.03 m). A similar sloshing motion takes place, and the exchange of potential and kinetic energy is similar to that shown in Fig. 7 for the channel. However, for the pipe geometry, we observe

a spatial discretization error in the total energy, stemming from the extra terms in (59). Though the energy error comes out positive at $t = 30$ s, the long-term trend is towards dissipation. Fortunately, for this case with $N_p = 40$ and $\Delta t = 0.005$ s, the energy error incurred by our proposed discretization remains relatively small and the effect on the solution is limited.

6.3. Traveling wave

Finally, we perform a test case with a traveling wave in a periodic domain. The flow is uni-directional and stratified, with a velocity and density difference between the two fluids. We consider a steady base state, upon which a small periodic perturbation is introduced, of which we study the evolution in time. This case is similar to test cases examining the Kelvin–Helmholtz instability, such as in [7,23]. However, here the perturbation will be stable since the flow is inviscid and in the (linearly) well-posed regime.

Most of the parameters are again identical to those given by Table 1, but the initial conditions for the hold-up and the fluid velocities are

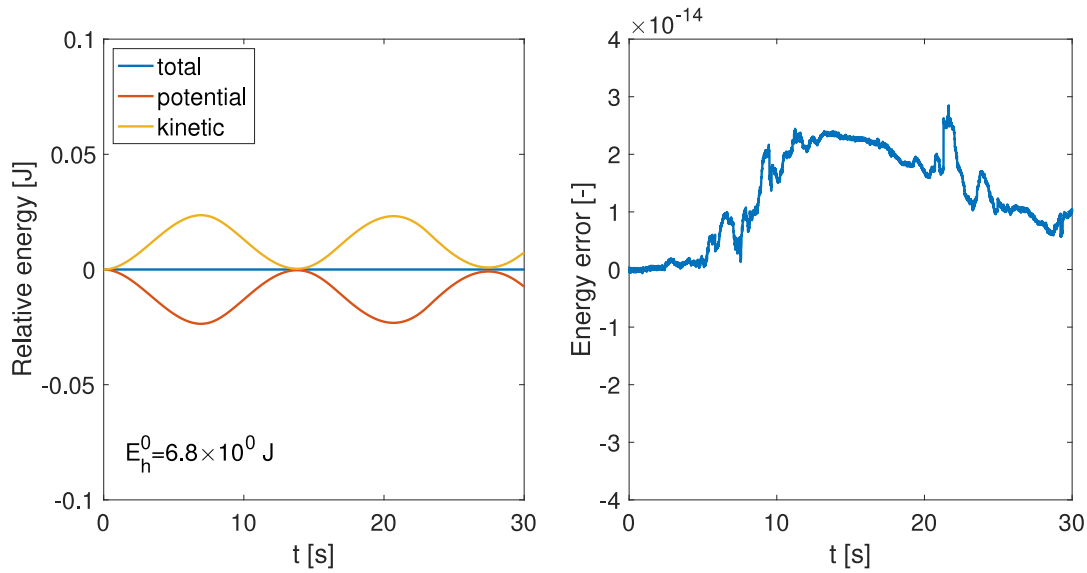


Fig. 7. Conserved quantities for the sloshing test case. Left: potential, kinetic and total energy relative to their initial values. Right: $(E_h - E_h^0)/E_h^0$.

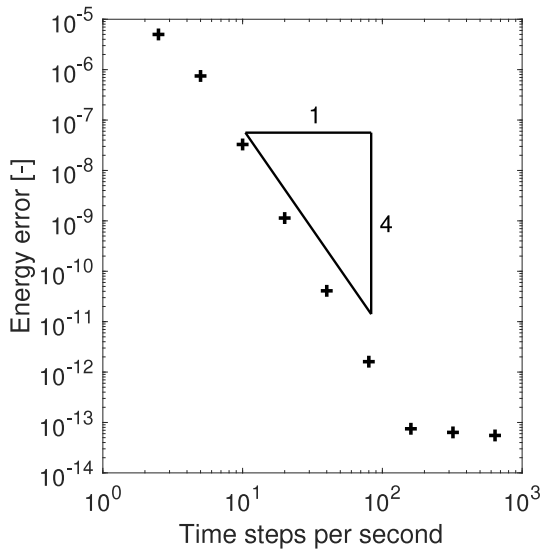


Fig. 8. Convergence of the energy error $(E_h^{N_t} - E_h^0)/E_h^0$ with time step, for the sloshing test case.

different. We set $\alpha_{L,0} = 0.4$ and $u_{L,0} = 1$. For u_U we take $u_{U,0} = 1.187$ (which is the value that would result in a steady flow with wall and interface friction⁴).

In order to construct an initial perturbation that results in a traveling wave, we conduct a linear stability analysis of the TFM [7]. The analysis is conducted in terms of its primitive variables in the form

$$\mathbf{w}^T = [\alpha_L \quad u_L \quad u_G \quad p]. \quad (63)$$

As exact solutions we obtain waves of the form

$$\Delta \mathbf{w} = \text{Re}(\Delta \hat{\mathbf{w}} \exp[i(\omega t - ks)]), \quad (64)$$

with $\Delta \hat{\mathbf{w}}$ the amplitude of the perturbation in each variable. The relative amplitudes in $\Delta \hat{\mathbf{w}}$ are such that $\Delta \hat{\mathbf{w}}$ is an eigenvector corresponding to one of two dispersion relations $\omega(k)$.

⁴ For this we take the Churchill friction model [35] with viscosities of $\mu_U = 1.5 \cdot 10^{-3} \text{ kg m}^{-1} \text{ s}^{-1}$ and $\mu_L = 1 \cdot 10^{-3} \text{ kg m}^{-1} \text{ s}^{-1}$.

The initial perturbation is defined as (64), with $t = 0$. We take a wavenumber of $k = 2\pi/L \text{ m}^{-1}$ and calculate the corresponding angular frequencies, of which one is selected. The chosen mode is

$$\omega = 3.982 \text{ s}^{-1}.$$

Setting $\Delta \hat{\alpha}_L = 1 \cdot 10^{-2}$, the amplitudes of the other variables are calculated so that $\Delta \hat{\mathbf{w}}$ is an eigenvector corresponding to this mode:

$$(\Delta \hat{\mathbf{w}})^T = [1.00 \cdot 10^{-2} \quad 3.99 \cdot 10^{-3} \quad 4.51 \cdot 10^{-3} \quad -2.30].$$

This ensures that the other mode is not present in the initial perturbation, so that we can study the isolated behavior of one mode. A projection step is then performed in order to make the initial condition satisfy the constraints (see Section 4).

The initial condition is shown in Fig. 10, along with its evolution in time, which is computed up to $t = 30 \text{ s}$. Setting the initial condition this way yields a wave traveling to the right at velocity $\omega/k = 1.16 \text{ m s}^{-1}$, which remains of approximately constant amplitude since the flow is inviscid and in the well-posed regime, so that ω has no imaginary component. The traveling wave can deform due to the nonlinear character of the governing equations, which is neglected in the linear stability analysis. This is made apparent by the snapshots of the solution shown in Fig. 10, which are separated by an integer number of wave periods: at the time of the last snapshot the wave has traveled through the domain 18 times. The solutions do not completely overlap and we see wave steepening taking place.

Fig. 11 shows the evolution of the energy. In this case, the exchange between kinetic and potential energy is small relative to the total energy of the base state. This is due to the fact that the wave is roughly constant in time, up to a displacement which does not change the energy.

The total energy can again be seen to remain constant up to a high precision. Like before, this is achieved by using a small time step ($\Delta t = 0.005 \text{ s}$), with a modest spatial resolution ($N_p = N_u = 40$). Fig. 12 shows how the energy converges with time step refinement. The convergence rate is fourth order over a wide range of time steps (matching the order of the time integration method), demonstrating that also for this test case, the spatial discretization conserves energy. While the solution moves away from the stable traveling wave predicted by linear analysis, its energy remains constant with time.

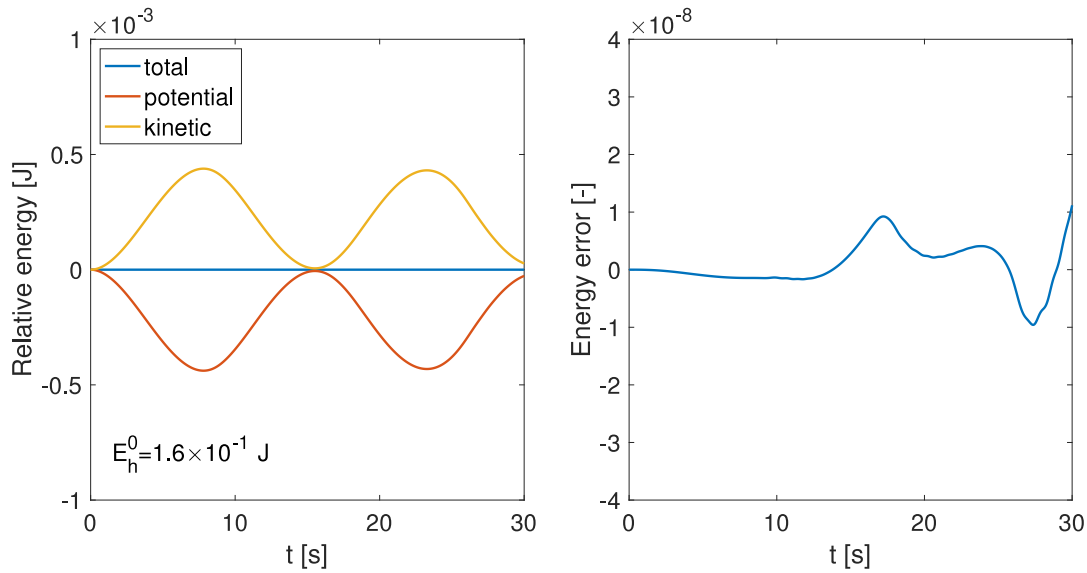


Fig. 9. Conserved quantities for the sloshing test case, but in a circular pipe geometry. Left: potential, kinetic and total energy relative to their initial values. Right: $(E_h - E_h^0)/E_h^0$.

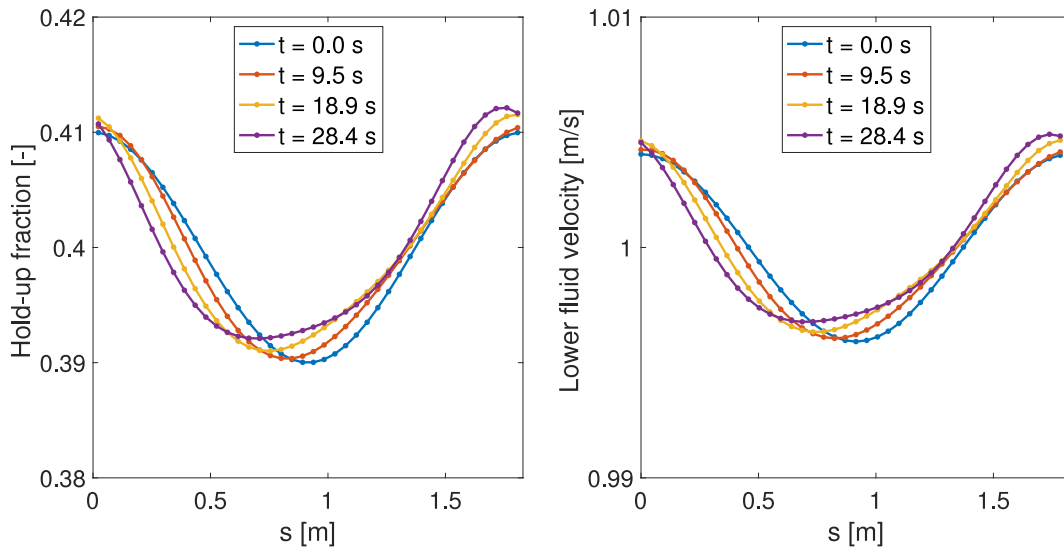


Fig. 10. The initial perturbation travels to the right with time. Consecutive snapshots are separated by a time interval of 6 wave periods. Left: lower fluid hold-up. Right: lower fluid velocity.

7. Conclusions

In this article, we have derived the result that the total mechanical energy (sum of kinetic and potential energy) is a secondary conserved quantity of the incompressible and isothermal TFM. This result is in line with the well-known fact that multi-dimensional incompressible frictionless flow equations conserve mechanical energy. Our novel insight is that this conservation statement still holds after averaging: the averaging procedure used to obtain the 1D TFM does not interfere with the energy conservation property. The approach was based on the formulation of entropy variables and an entropy potential, similar to what is commonly done for the SWE, but with two main differences: (i) we have included a non-conservative pressure term in our analysis, which is shown to be energy-conserving, and (ii) we have obtained our results independent of the duct geometry, which may be a 2D channel or a circular pipe, or any other closed cross-sectional duct shape.

The second novel result of this paper is a set of numerical fluxes that conserve a discrete form of the mechanical energy. A discretization on a staggered grid was proposed in order to keep the energy conservation

property of the non-conservative pressure terms in a discrete sense. Although the use of a staggered grid implies that the choice of a discrete energy and entropy potential is not unique, we were able to propose a combination which is such that the discrete analysis is consistent with and analogous to the continuous analysis. However, one important difference between the continuous and discrete cases remains, namely in the analysis of the level gradient terms (for arbitrary geometries). A geometric relation between the potential energy and the interface height is satisfied exactly in the continuous case, but only approximately in the discrete case. Fortunately, for the specific case of the 2D channel geometry, the condition is satisfied *exactly*, and the discrete level gradient reduces to a form which parallels the continuous form perfectly. For other geometries, such as the pipe, a small numerical energy error persists in the discrete analysis.

Our theoretical derivations are supported by numerical experiments, which show that the proposed energy is indeed exactly conserved by our new spatial discretization in both periodic and closed domains. Building on previous work [23], the discretization also conserves mass and momentum, has strong coupling between momentum

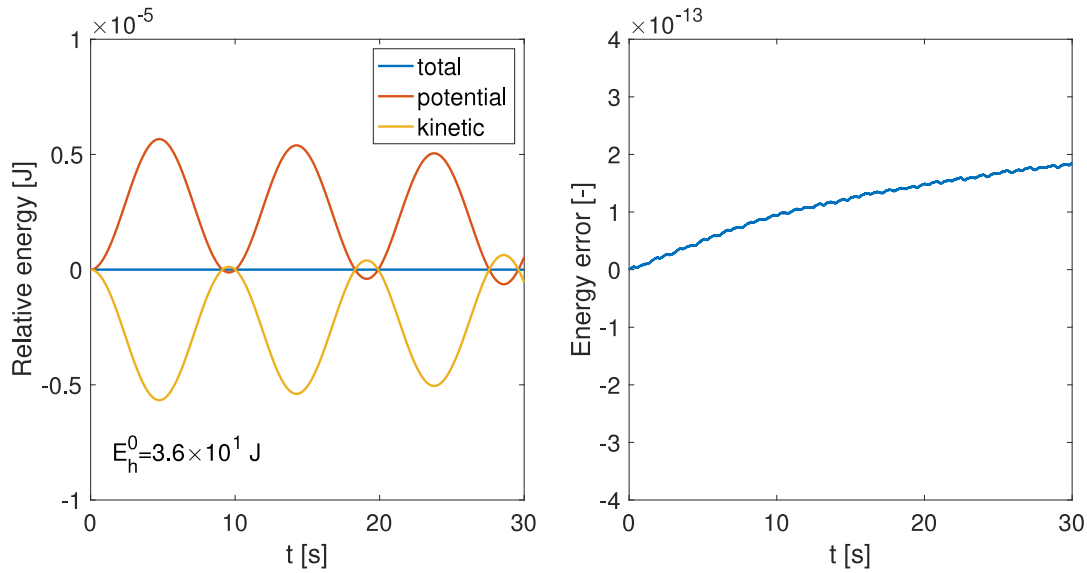


Fig. 11. Conserved quantities for the traveling wave test case. Left: potential, kinetic and total energy relative to their initial values. Right: $(E_h - E_h^0)/E_h^0$.

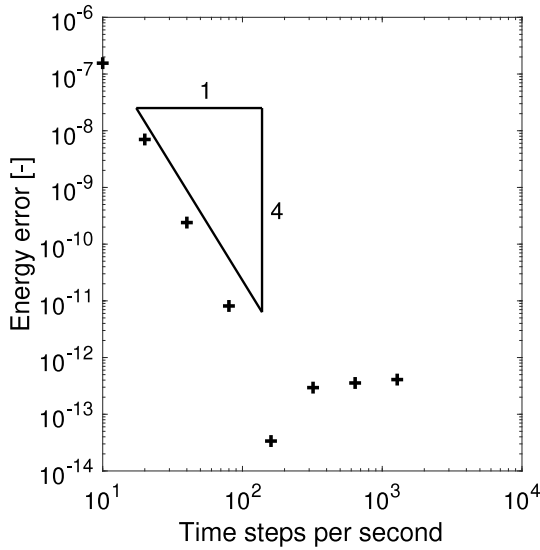


Fig. 12. Convergence of the energy error $(E_h^{N_t} - E_h^0)/E_h^0$ with time step, for the traveling wave test case.

and pressure, and is constraint-consistent. In these experiments the temporal error was negligible (due to a combination of high-order time integration and small time steps), but for future work it is suggested to also make the time integration method energy-conserving [13].

Our energy-conserving formulation of the TFM provides a foundation for investigating the nonlinear stability of the model. For related models, the energy acts as a norm or a convex entropy function of the solution, providing stability bounds, and it should be investigated if the TFM energy has similar implications. While we have only considered smooth solutions, in general cases discontinuities may appear, at which energy should be dissipated [36]. Failure of a numerical scheme to do so leads to numerical oscillations. As a first option, an energy-consistent discretization of the wall and interface friction, and streamwise diffusion due to molecular and turbulent viscosity [37,38], could be considered. Alternatively, a carefully chosen numerical diffusion operator could be added to the scheme to obtain the required dissipation at shocks [15,39].

CRediT authorship contribution statement

J.F.H. Buist: Conceptualization, Methodology, Software, Writing – original draft. **B. Sanderse:** Conceptualization, Methodology, Software, Writing – review & editing, Supervision. **S. Dubinkina:** Writing – review & editing, Supervision. **R.A.W.M. Henkes:** Writing – review & editing, Supervision. **C.W. Oosterlee:** Writing – review & editing, Supervision.

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.

Funding

This work was supported by the research program Shell-NWO/FOM Computational Sciences for Energy Research (CSER), project number 15CSER17.

Appendix A. Geometric relations

We treat the model equations in a way that is general to arbitrary duct geometries, using general geometric quantities which can be substituted for expressions that are specific to certain duct cross-sectional shapes. The most important general geometric terms are the H -variables, of which we have three for each fluid: H_U , \hat{H}_U , \tilde{H}_U , H_L , \hat{H}_L , \tilde{H}_L . We use H (implying something like a height) for each of these variables because they are all invertible functions only of A_U and A_L respectively, and these functions all depend only on the cross-sectional duct shape. They are all distinct though, and the relations between these geometric quantities (which hold for arbitrary geometries) are crucial to the results of this paper.

Two geometries of particular interest are the 2D channel and the circular pipe. For a 2D channel geometry, the following substitutions can be made in the equations:

$$\begin{aligned} H_L &= A_L, & H_U &= A_U, \\ P_L &= 1, & P_U &= 1, \end{aligned}$$

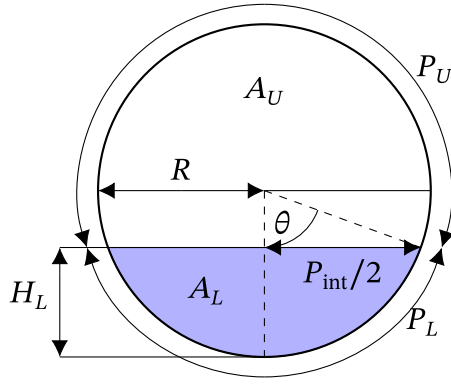


Fig. A.1. A schematic of a circular pipe cross-section.

$$A = H, \quad P_{\text{int}} = 1.$$

For a pipe geometry we have, as in [40],

$$\begin{aligned} H_L &= R(1 - \cos(\theta)), & H_U &= R(1 + \cos(\theta)), \\ P_L &= 2R\theta, & P_U &= 2R(\pi - \theta), \\ A &= \pi R^2, & P_{\text{int}} &= 2R \sin(\theta), \\ A_L &= R^2 \left(\theta - \frac{1}{2} \sin(2\theta) \right), & A_U &= R^2 \left(\pi - \theta + \frac{1}{2} \sin(2\theta) \right). \end{aligned}$$

In Fig. A.1 we show how the wetted angle θ is defined. If $\alpha_L = A_L/A$, then $\pi\alpha_L = \theta - \frac{1}{2} \sin(2\theta)$, and this equation must be solved iteratively in order to obtain θ from A_L , so that the remaining geometric quantities can be calculated.

The integrals (3) which appear in the governing equations of the two-fluid model are geometry-dependent:

$$\hat{H}_L := \int_{a_L} (h - H_L) da = \int_0^{H_L} (h - H_L) w(h) dh, \quad (\text{A.1})$$

$$\hat{H}_U := \int_{a_U} (h - H_L) da = \int_{H_L}^H (h - H_L) w(h) dh, \quad (\text{A.2})$$

with $w(h)$ the local width. Note that $w(H_L) = P_{\text{int}}$. For a 2D channel geometry, with $A_L = H_L$ and $A_U = H_U$, the width is given by $w(h) = 1$ and the integrals evaluate to

$$\hat{H}_L = -\frac{1}{2} A_L^2, \quad \hat{H}_U = \frac{1}{2} A_U^2,$$

where we have substituted $A_L = A - A_U$. For the pipe geometry, we make the transformation $h = R(1 - \cos(\theta^*))$, with θ^* the integration variable and θ the wetted angle, to get [22]

$$\hat{H}_L = \left[(R - H_L) A_L - \frac{1}{12} P_{\text{int}}^3 \right], \quad \hat{H}_U = - \left[(R - H_U) A_U - \frac{1}{12} P_{\text{int}}^3 \right].$$

The following derivatives of \hat{H}_L and \hat{H}_U are needed in order to calculate $\partial \mathbf{f} / \partial \mathbf{q}$:

$$\frac{d\hat{H}_L}{dA_L} = \frac{d\hat{H}_L}{dH_L} \frac{dH_L}{dA_L}, \quad \frac{d\hat{H}_U}{dA_U} = \frac{d\hat{H}_U}{dH_U} \frac{dH_U}{dA_U}. \quad (\text{A.3})$$

We use Leibniz' rule to calculate

$$\begin{aligned} \frac{d\hat{H}_L}{dH_L} &= \frac{d}{dH_L} \int_0^{H_L} (h - H_L) w(h) dh \\ &= (h(H_L) - H_L) w(H_L) \frac{dH_L}{dH_L} - (h(0) - H_L) w(0) \frac{d0}{dH_L} \\ &\quad + \int_0^{H_L} \frac{d}{dH_L} ((h - H_L) w(h)) dh \\ &= - \int_0^{H_L} w(h) dh = -A_L, \end{aligned}$$

and similarly

$$\frac{d\hat{H}_U}{dH_U} = \frac{d}{dH_U} \int_{H_L}^H h w(h) dh = A_U.$$

Substitution in (A.3) gives the following relations:

$$\frac{d\hat{H}_L}{dA_L} = -A_L \frac{dH_L}{dA_L}, \quad \frac{d\hat{H}_U}{dA_U} = A_U \frac{dH_U}{dA_U}, \quad (\text{A.4})$$

and the inverse of the derivatives appearing on the right-hand sides can also be evaluated using Leibniz' rule:

$$\frac{dA_L}{dH_L} = P_{\text{int}}, \quad \frac{dA_U}{dH_U} = P_{\text{int}}.$$

Besides \hat{H}_L and \hat{H}_U , the following geometric quantities are used in (16) and defined as:

$$\tilde{H}_L := \int_{a_L} h da = \int_0^{H_L} h w(h) dh = \hat{H}_L + H_L A_L, \quad (\text{A.5})$$

$$\tilde{H}_U := \int_{a_U} h da = \int_{H_L}^H h w(h) dh = \hat{H}_U + (H - H_U) A_U, \quad (\text{A.6})$$

which can be evaluated by substituting the expressions for \hat{H}_L and \hat{H}_U . In order to calculate \mathbf{v} as given by (17), we need the derivatives $d\tilde{H}_L/dA_L$ and $d\tilde{H}_U/dA_U$. They are found by differentiating (A.5) and (A.6), yielding

$$\frac{d\tilde{H}_L}{dA_L} = H_L, \quad \frac{d\tilde{H}_U}{dA_U} = H - H_U. \quad (\text{A.7})$$

Appendix B. Global energy analysis

The main text has described a way to derive the *local* semi-discrete energy conservation equation given by (33). In the case of periodic or closed boundaries, this can be integrated in space to yield global energy conservation. In this section, we directly derive the global energy conservation equation without the intermediate step of the local energy. This allows us to skip the step of choosing an entropy potential, which means that the derivation will contain less assumptions. On the other hand, the obtained conditions on the numerical fluxes are not constructive, because they are conditions for the 'jumps' of the numerical fluxes, rather than for a single numerical flux at one discrete point. Therefore, the global analysis is not used as a replacement, but as a validation of the local analysis.

The scheme (29) described in Section 4 for a certain pressure volume i and velocity volume $i - 1/2$ can be extended to describe the evolution of the entire state vector \mathbf{q}_h :

$$\frac{d\mathbf{q}_h}{dt} + \mathbf{f}_h + \mathbf{d}_h = \mathbf{0}, \quad (\text{B.1})$$

where $\mathbf{q}_h = [q_{1,1} \dots q_{1,N}, q_{2,1} \dots q_{2,N}, q_{3,1/2} \dots q_{3,N-1/2}, q_{4,1/2} \dots q_{4,N-1/2}]^T$, and similar expressions for \mathbf{f}_h and \mathbf{d}_h . For simplicity we only discuss periodic boundary conditions, for which $N_p = N_u = N$.

Similar to the local entropy variable \mathbf{v} we define the global entropy variable

$$\mathbf{v}_h := \left[\frac{dE_h}{d\mathbf{q}_h} \right]^T.$$

Taking the inner product of \mathbf{v}_h and (B.1), the first term yields

$$\langle \mathbf{v}_h, \frac{d\mathbf{q}_h}{dt} \rangle = \frac{dE_h}{dt}.$$

Thus, to obtain global discrete energy conservation, given by $\frac{dE_h}{dt} = 0$, we need the following conditions on \mathbf{f}_h and \mathbf{d}_h :

$$\langle \mathbf{v}_h, \mathbf{f}_h \rangle = 0, \quad (\text{B.2})$$

$$\langle \mathbf{v}_h, \mathbf{d}_h \rangle = 0. \quad (\text{B.3})$$

In order to evaluate \mathbf{v}_h , we note that

$$\frac{\partial E_h}{\partial \mathbf{q}_i} = \frac{\partial e_{i-1/2}}{\partial \mathbf{q}_i} + \frac{\partial e_{i+1/2}}{\partial \mathbf{q}_i} = \mathbf{v}_{i-1/2,i} + \mathbf{v}_{i+1/2,i}$$

$$= \begin{bmatrix} -\frac{1}{2} \left(\frac{q_{3,i}^2}{\bar{q}_{1,i}^2} \right) + g_n (H - H_{U,i}) \\ -\frac{1}{2} \left(\frac{q_{4,i}^2}{\bar{q}_{2,i}^2} \right) + g_n H_{L,i} \\ \frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \\ \frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} \end{bmatrix}, \quad (\text{B.4})$$

and \mathbf{v}_h follows by assembling this expression for all grid points (ordered by equation, like \mathbf{q}_h). The pressure condition (B.3) then evaluates to

$$\langle \mathbf{v}_h, \mathbf{d}_h \rangle = \sum_{i=1}^N Q_{i-1/2} (p_i - p_{i-1}) = \sum_{i=1}^N (Q_{i+1/2} - Q_{i-1/2}) p_i = 0,$$

and is thus satisfied because Q is uniform in space.

The flux condition (B.2) evaluates to

$$\begin{aligned} \langle \mathbf{v}_h, \mathbf{f}_h \rangle &= \sum_{i=1}^N \left(-\frac{1}{2} \left(\frac{q_{3,i}^2}{\bar{q}_{1,i}^2} \right) + g_n (H - H_{U,i}) \right) (f_{1,i+1/2} - f_{1,i-1/2}) \\ &\quad + \left(-\frac{1}{2} \left(\frac{q_{4,i}^2}{\bar{q}_{2,i}^2} \right) + g_n H_{L,i} \right) (f_{2,i+1/2} - f_{2,i-1/2}) \\ &\quad + \left(\frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \right) (f_{3,i} - f_{3,i-1}) + \left(\frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} \right) (f_{4,i} - f_{4,i-1}). \end{aligned}$$

We split this condition into two conditions: one proportional to g_n and one not proportional to g_n :

$$\langle \mathbf{v}_h, \mathbf{f}_h \rangle = \langle \mathbf{v}_h, \mathbf{f}_h \rangle_a + \langle \mathbf{v}_h, \mathbf{f}_h \rangle_g.$$

The advective condition is given by

$$\begin{aligned} \langle \mathbf{v}_h, \mathbf{f}_h \rangle_a &= \sum_{i=1}^N \left[-\frac{1}{2} \left(\frac{q_{3,i}^2}{\bar{q}_{1,i}^2} \right) (f_{1,i+1/2} - f_{1,i-1/2}) \right. \\ &\quad - \frac{1}{2} \left(\frac{q_{4,i}^2}{\bar{q}_{2,i}^2} \right) (f_{2,i+1/2} - f_{2,i-1/2}) \\ &\quad + \left(\frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \right) (f_{3,i,a} - f_{3,i-1,a}) \\ &\quad \left. + \left(\frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} \right) (f_{4,i,a} - f_{4,i-1,a}) \right]. \end{aligned}$$

Substituting (53) yields an equation that can be rewritten as

$$\begin{aligned} \langle \mathbf{v}_h, \mathbf{f}_h \rangle_a &= \sum_{i=1}^N \left[\frac{1}{2} \left(\frac{q_{3,i-1/2}^2}{\bar{q}_{1,i-1/2}^2} \frac{q_{3,i-1/2}}{\Delta s} - \frac{q_{3,i+1/2}^2}{\bar{q}_{1,i+1/2}^2} \frac{q_{3,i+1/2}}{\Delta s} \right) \right. \\ &\quad + \frac{1}{2} \left(\frac{q_{4,i-1/2}^2}{\bar{q}_{2,i-1/2}^2} \frac{q_{4,i-1/2}}{\Delta s} - \frac{q_{4,i+1/2}^2}{\bar{q}_{2,i+1/2}^2} \frac{q_{4,i+1/2}}{\Delta s} \right) \\ &\quad + \left(\frac{q_{3,i+1/2}}{\bar{q}_{1,i+1/2}} - \frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \right) \left(\frac{q_{3,i}}{\bar{q}_{1,i}} \right) \frac{\bar{q}_{3,i}}{\Delta s} \\ &\quad + \left(\frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \right) (f_{3,i,a} - f_{3,i-1,a}) \\ &\quad + \left(\frac{q_{4,i+1/2}}{\bar{q}_{2,i+1/2}} - \frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} \right) \left(\frac{q_{4,i}}{\bar{q}_{2,i}} \right) \frac{\bar{q}_{4,i}}{\Delta s} \\ &\quad \left. + \left(\frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} \right) (f_{4,i,a} - f_{4,i-1,a}) \right]. \end{aligned}$$

Here, the sum over the entries on the first two lines evaluates to zero, since each term has a matching term of opposite sign and index shifted by 1 (even the boundary terms, in case of periodic boundaries). In order for this to also hold for the terms in the remaining lines, we need to satisfy the condition

$$\left(\frac{q_{3,i+1/2}}{\bar{q}_{1,i+1/2}} - \frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \right) \left(\frac{q_{3,i}}{\bar{q}_{1,i}} \right) \frac{\bar{q}_{3,i}}{\Delta s} + \left(\frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \right) f_{3,i,a} = \left(\frac{q_{3,i+1/2}}{\bar{q}_{1,i+1/2}} \right) f_{3,i,a},$$

and similar for $f_{4,i,a}$. These are indeed satisfied with our choice (54).

The condition proportional to g_n , after substitution of (53), is given by

$$\begin{aligned} \langle \mathbf{v}_h, \mathbf{f}_h \rangle_g &= \sum_{i=1}^N \left[g_n (H - H_{U,i}) \left(\frac{q_{3,i+1/2}}{\Delta s} - \frac{q_{3,i-1/2}}{\Delta s} \right) \right. \\ &\quad + g_n H_{L,i} \left(\frac{q_{4,i+1/2}}{\Delta s} - \frac{q_{4,i-1/2}}{\Delta s} \right) \\ &\quad + g_n \left(\frac{q_{3,i-1/2}}{\bar{q}_{1,i-1/2}} \right) (f_{3,i,g} - f_{3,i-1,g}) \\ &\quad \left. + g_n \left(\frac{q_{4,i-1/2}}{\bar{q}_{2,i-1/2}} \right) (f_{4,i,g} - f_{4,i-1,g}) \right], \end{aligned}$$

and it can be rewritten as

$$\begin{aligned} \langle \mathbf{v}_h, \mathbf{f}_h \rangle_g &= \sum_{i=1}^N \left[g_n \frac{\Delta s (f_{3,i,g} - f_{3,i-1,g}) - (H - H_{U,i}) \bar{q}_{1,i-1/2} q_{3,i-1/2}}{\bar{q}_{1,i-1/2} \Delta s} \right. \\ &\quad + g_n \frac{\Delta s (f_{4,i,g} - f_{4,i-1,g}) - H_{L,i} \bar{q}_{2,i-1/2} q_{4,i-1/2}}{\bar{q}_{2,i-1/2} \Delta s} \\ &\quad \left. + g_n (H - H_{U,i}) \frac{q_{3,i+1/2}}{\Delta s} + g_n H_{L,i} \frac{q_{4,i+1/2}}{\Delta s} \right]. \end{aligned}$$

Now, in order for this to be conservative, we need the terms in the first two lines to be equal but opposite in sign to the terms in the third line (shifted in index by 1). This yields the following conditions:

$$\|f_{3,i-1/2,g}\| = -\frac{\bar{q}_{1,i-1/2}}{\Delta s} \|H_{U,i-1/2}\|, \quad (\text{B.5})$$

$$\|f_{4,i-1/2,g}\| = \frac{\bar{q}_{2,i-1/2}}{\Delta s} \|H_{L,i-1/2}\|,$$

which upon substitution of (60) reduce to the geometric conditions (56).

In conclusion, the results of the global discrete analysis are consistent with our local discrete analysis. The additional insight from the global analysis is that the geometric conditions (B.5) or (56) are *independent of the choice of the entropy potential*. Rather, they follow directly from our assumed form (60) of the numerical fluxes (which can be seen as a simplification of the full form (55), which was obtained using the entropy potential). This confirms that the choice of entropy potential does not limit the results.

Remark 4. The global energy analysis can also be performed without requiring interpolation of the potential energy to the velocity grid points, as needed in the definition of $e_{i-1/2}$ given by (40). Instead, one can directly define

$$E_h = \sum_{i=1}^{N_p} \left(\rho_U g_n \tilde{H}_{U,i} \Delta s + \rho_L g_n \tilde{H}_{L,i} \Delta s \right) + \sum_{i=1}^{N_u} \left(\frac{1}{2} \frac{q_{3,i-1/2}^2}{\bar{q}_{1,i-1/2}} + \frac{1}{2} \frac{q_{4,i-1/2}^2}{\bar{q}_{2,i-1/2}} \right).$$

It can be verified that this leads to the same \mathbf{v}_h as given by (B.4), and consequently the geometric condition (B.5) remains present.

References

- [1] Wallis GB. One-dimensional two-phase flow. New York: McGraw-Hill; 1969.
- [2] Ishii M. Thermo-fluid dynamic theory of two-phase flow. Paris: Eyrolles; 1975.
- [3] Goldszal A, Danielson TJ, Bansal KM, Yang ZL, Johansen ST, Depay G. LedaFlow 1D: Simulation results with multiphase gas/condensate and oil/gas field data. In: BHR group - 13th International conference on multiphase production technology. 2007, p. 15.
- [4] Aursand P, Hammer M, Munkejord ST, Wilhelmson Ø. Pipeline transport of CO2 mixtures: Models for transient simulation. Int J Greenhouse Gas Control 2013;15:174–85.
- [5] Berry RA, Zou L, Zhao H, Zhang H, Peterson JW, Martineau RC, Kadioglu SY, Andrs D. RELAP-7 theory manual. Technical Report INL/EXT-14-31366, Idaho National Laboratory; 2014.
- [6] Lyczkowski RW, Gidaspow D, Solbrig CW, Hughes ED. Characteristics and stability analyses of transient one-dimensional two-phase flow equations and their finite difference approximations. Nucl Sci Eng 1978;66(3):378–96.

- [7] Liao J, Mei R, Klausner JF. A study on the numerical stability of the two-fluid model near ill-posedness. *Int J Multiph Flow* 2008;34(11):1067–87.
- [8] Kreiss H-O, Yström J. A note on viscous conservation laws with complex characteristics. *BIT Numer Math* 2006;46(S1):55–9.
- [9] Stewart HB, Wendroff B. Two-phase flow: Models and methods. *J Comput Phys* 1984;56(3):363–409.
- [10] Keyfitz BL. Mathematical properties of nonhyperbolic models for incompressible two-phase flow. In: *Proceedings of the ICMF*; 2001.
- [11] López de Bertodano M, Fullmer W, Clausse A, Ransom VH. Two-fluid model stability, simulation and chaos. Cham, Switzerland: Springer International Publishing; 2017.
- [12] Coppola G, Capuano F, de Luca L. Discrete energy-conservation properties in the numerical simulation of the Navier–Stokes equations. *Appl Mech Rev* 2019;71(1).
- [13] Sanderse B. Energy-conserving Runge–Kutta methods for the incompressible Navier–Stokes equations. *J Comput Phys* 2013;233:100–31.
- [14] Tadmor E. Entropy stability theory for difference approximations of non-linear conservation laws and related time-dependent problems. *Acta Numer* 2003;12:451–512.
- [15] Fjordholm US, Mishra S, Tadmor E. Well-balanced and energy stable schemes for the shallow water equations with discontinuous topography. *J Comput Phys* 2011;230(14):5587–609.
- [16] Fjordholm US. Energy conservative and stable schemes for the two-layer shallow water equations. In: *Hyperbolic problems. Series in contemporary applied mathematics*, vols. 17 & 18, Co-Published with Higher Education Press; 2012, p. 414–21.
- [17] Fjordholm US, Mishra S, Tadmor E. Energy preserving and energy stable schemes for the shallow water equations. In: Cucker F, Pinkus A, Todd MJ, editors. *Foundations of computational mathematics*, Hong Kong 2008. Cambridge University Press; 2009, p. 93–139.
- [18] van 't Hof B, Veldman AEP. Mass, momentum and energy conserving (MaMEC) discretizations on general grids for the compressible Euler and shallow water equations. *J Comput Phys* 2012;231(14):4723–44.
- [19] Ishii M, Mishima K. Two-fluid model and hydrodynamic constitutive relations. *Nucl Eng Des* 1984;82(2–3):107–26.
- [20] Montini M. Closure relations of the one-dimensional two-fluid model for the simulation of slug flows (Ph.D. thesis), Imperial College London; 2011.
- [21] Munkejord ST. Analysis of the two-fluid model and the drift-flux model for numerical calculation of two-phase flow (Ph.D. thesis), Trondheim: Norwegian University of Science and Technology; 2006.
- [22] Sanderse B, Smith IE, Hendrix MHW. Analysis of time integration methods for the compressible two-fluid model for pipe flow simulations. *Int J Multiph Flow* 2017;95:155–74.
- [23] Sanderse B, Veldman AEP. Constraint-consistent Runge–Kutta methods for one-dimensional incompressible multiphase flow. *J Comput Phys* 2019;384:170–99.
- [24] Tadmor E, Zhong W. Energy-preserving and stable approximations for the two-dimensional shallow water equations. In: Munthe-Kaas H, Owren B, editors. *Mathematics and computation, a contemporary view*, vol. 3. Springer Berlin Heidelberg; 2008, p. 67–94.
- [25] Abgrall R, Karni S. Two-layer shallow water system: A relaxation approach. *SIAM J Sci Comput* 2009;31(3):1603–27.
- [26] Leveque RJ. *Finite volume methods for hyperbolic problems*. Cambridge University Press; 2002.
- [27] van 't Hof B, Vuik MJ. Symmetry-preserving finite-difference discretizations of arbitrary order on structured curvilinear staggered grids. *J Comput Sci* 2019;36:101008.
- [28] Jameson A. Formulation of kinetic energy preserving conservative schemes for gas dynamics and direct numerical simulation of one-dimensional viscous compressible flow in a shock tube using entropy and kinetic energy preserving schemes. *J Sci Comput* 2008;34(2):188–208.
- [29] Veldman AEP. A general condition for kinetic-energy preserving discretization of flow transport equations. *J Comput Phys* 2019;398:108894.
- [30] Thorpe SA. Experiments on the instability of stratified shear flows: Immiscible fluids. *J Fluid Mech* 1969;39(1):25–48.
- [31] Lopez de Bertodano M, Fullmer WD, Clausse A. One-dimensional two-fluid model for wavy flow beyond the Kelvin–Helmholtz instability: Limit cycles and chaos. *Nucl Eng Des* 2016;310:656–63.
- [32] Holmås H. Numerical simulation of transient roll-waves in two-phase pipe flow. *Chem Eng Sci* 2010;65(5):1811–25.
- [33] Buist J, Sanderse B, van Halder Y, Koren B, van Heijst GJ. Machine learning for closure models in multiphase flow applications. In: *Proceedings of the 3rd International conference on uncertainty quantification in computational sciences and engineering (UNCECOMP 2019)*, Crete, Greece; 2019, p. 379–99.
- [34] Ockendon H, Ockendon JR. How to mitigate sloshing. *SIAM Rev* 2017;59(4):905–11.
- [35] Churchill SW. Friction factor equation spans all fluid flow regimes. *Chem Eng* 1977;84:91–2.
- [36] Jameson A. The construction of discretely conservative finite volume schemes that also globally conserve energy or entropy. *J Sci Comput* 2008;34(2):152–87.
- [37] Fullmer WD, Ransom VH, Lopez de Bertodano MA. Linear and nonlinear analysis of an unstable, but well-posed, one-dimensional two-fluid model for two-phase flow based on the inviscid Kelvin–Helmholtz instability. *Nucl Eng Des* 2014;268:173–84.
- [38] Verstappen RWCP, Veldman AEP. Symmetry-preserving discretization of turbulent flow. *J Comput Phys* 2003;187(1):343–68.
- [39] Castro MJ, Fjordholm US, Mishra S, Parés C. Entropy conservative and entropy stable schemes for nonconservative hyperbolic systems. *SIAM J Numer Anal* 2013;51(3):1371–91.
- [40] Akselsen AH. Efficient numerical methods for waves in one-dimensional two-phase pipe flows (Ph.D. thesis), Norwegian University of Science and Technology; 2016.