# The Navier-Stokes system with temperature and salinity for free surface flows. Numerical scheme and validation

Léa Boittin, François Bouchut, Marie-Odile Bristeau, Anne Mangeney, Jacques Sainte-Marie, Fabien Souillé

# The Navier-Stokes system with temperature and salinity for free surface flows - Numerical scheme and validation

L. Boittin[1,4], F. Bouchut[2], M.-O. Bristeau[1], A. Mangeney[1,3], J. Sainte-Marie[1], and F. Souillé[1,4]

[1]Inria Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12 and Sorbonne Université, Univ. Paris Diderot, SPC, CNRS, Laboratoire Jacques-Louis Lions, LJLL, F-75005 Paris
[2]Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), CNRS, Univ. Gustave Eiffel, UPEC, F-77454, Marne-la-Vallée, France
[3]Univ. Paris Diderot, Sorbonne Paris Cité, Institut de Physique du Globe de Paris, Seismology Group, 1 rue Jussieu, Paris F-75005, France
[4]Risk Management Solutions, Peninsular House, 30 Monument Street, London, EC3R 8NB, UK
[5]EDF R&D LNHE - Laboratoire National d'Hydraulique et Environnement, 6 quai Watier, F-78400 Chatou

September 23, 2021

**Abstract**

In this paper, we propose a numerical scheme for the layer-averaged Euler with variable density and the Navier-Stokes-Fourier systems presented in part I (Boittin et al., 2020). These systems model hydrostatic free surface flows with density variations. We show that the finite volume scheme presented is well balanced with regards to the steady state of the lake at rest and preserves the positivity of the water height. A maximum principle on the density is also proved as well as a discrete entropy inequality in the case of the Euler system with variable density. Some numerical validations are finally shown with comparisons to 3D analytical solutions and experiments.

*Keywords:* Navier-Stokes equations, free surface flows, variable density flows, layer-averaged formulation, finite volume scheme

# Contents



# 1 Introduction

In this paper we present a numerical scheme for the 3D incompressible Navier-Stokes-Fourier system with variable density and free surface, as well as numerical test cases. This model describes variable density flows with free surface, the density variations coming from differences in temperature and/or salinity. The model is presented in the companion paper (Boittin et al., 2020), in which a layer-averaged formulation is also given. The layer-averaged formulation suppresses the need for moving meshes (Decoene

and Gerbeau, 2009), (Donea et al., 2004). It allows to perform 3D simulations with a 2D fixed mesh.

Variable density flows are frequently studied by oceanographers. Different systems of coordinates exist, among which terrain-following coordinates and isopycnal coordinates. For a discussion of the advantages and disadvantages of the various coordinates frequently used in ocean models, the reader is referred to (Griffies et al., 2000) and (Song and Hou, 2006). The layer-averaged model presented here is not a terrain-following coordinate model. Though the layer thicknesses are defined as fractions of the total water height, it is not a $\sigma$-coordinate system. The model also differs from isopycnal coordinate models because in the layer-averaged formulation, the layers exchange mass between themselves, which means that the internal layer boundaries are actually not physical.

For the Euler part of the Navier-Stokes-Fourier system with variable density, a finite-volume formalism is adopted. The hydrostatic reconstruction technique is used (Audusse et al., 2004). Therefore, the topography is accurately represented and the scheme is well-balanced. Yet the discretization of the nonconservative pressure terms demands special care. For the viscosity terms, we use finite elements as in (Allgeyer et al., 2019).

In (Audusse et al., 2011), a similar model was studied and simulated. The scheme presented in (Audusse et al., 2011) was a 2D $(x - z)$ scheme and relied on a kinetic interpretation. In the present work we present a fully 3D scheme which is more flexible in a certain sense, because the kinetic flux is only one of the possible choices for the numerical flux. With any flux consistent with the semi-discrete in time Euler system, the resulting scheme is well-balanced and preserves the nonnegativity of the water depth. A maximum principle on the density is satisfied. In order to prove an in-cell entropy inequality, we adopt a kinetic flux, already used in the context of the Shallow Water equations in (Perthame and Simeoni, 2001), (Audusse et al., 2011). The entropy inequality is satisfied for a constant topography and includes third-order rest terms. Moreover, the unknowns in (Audusse et al., 2011) were not the same, which resulted in a complicated numerical scheme - nonlinear systems were solved at each step and a Newton fixed-point method was used. The present scheme is simpler and does not involve nonlinear systems. Finally, the present scheme is more stable than the scheme in (Audusse et al., 2011), the CFL condition of which could actually degenerate and give a time step equal to zero. With the proposed scheme, the computational cost of the simulation of a non-Boussinesq flow is not greater than that of the simulation of a Boussinesq flow.

The proposed numerical scheme is validated on three test cases simulated with the Freshkiss3d code Freshkiss3d (2020). Each of these test cases allows to validate an aspect of the numerical scheme: wet/dry interface treatment, buoyancy terms, diffusion effects, second order extensions (space and time)... A real application of hydrodynamics in a river with chlorides entries will be presented in a forthcoming paper (the research report is available (Souillé et al., 2017), in French).

The first test is a convergence test towards an analytical solution (Bristeau et al., 2020) for the Euler system with variable density. In the second test, a lock exchange simulation is performed and the results are compared with experimental data available

from the literature (Adduce et al., 2012). Finally, in two diffusion cases, the differences between the Navier-Stokes-Fourier with variable density and Boussinesq models are evidenced.

The paper is organized as follows. In section 2, the layer-averaged Navier-Stokes-Fourier system and the Euler with variable density introduced in (Boittin et al., 2020) are recalled. A numerical scheme for the layer-averaged Euler model with variable density is presented in Section 3, its properties are studied. An extension of this scheme for the layer-averaged Navier-Stokes-Fourier model with variable density is presented in section 4. The numerical test cases are presented in section 5.

# 2 The layer-averaged models

minder_models⟩ From now on, we just call the Euler system, the Euler system with variable density. We briefly recall the features of the multilayer models presented in (Boittin et al., 2020) and studied here. The multilayer Navier-Stokes-Fourier model is a layer-averaged version of the incompressible, hydrostatic Navier-Stokes-Fourier system

$$\nabla \cdot \mathbf{U} = -\frac{\rho'(T)}{\rho^2 c_p}\left(\nabla\cdot(\lambda\nabla T) + \mu|\nabla_{x,y}\mathbf{u}|^2 + \mu\left|\frac{\partial \mathbf{u}}{\partial z}\right|^2\right), \tag{1}$$

eq:NSF_1_p2

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{U}) = 0, \tag{2}$$

?eq:NSF_2_p2?

$$\frac{\partial(\rho\mathbf{u})}{\partial t} + \nabla_{x,y}\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \frac{\partial(\rho\mathbf{u}w)}{\partial z} + \nabla_{x,y}\int_z^\eta \rho g dz_1 = \mu\Delta_{x,y}\mathbf{u} + \mu\frac{\partial}{\partial z}\left(\frac{\partial \mathbf{u}}{\partial z}\right), \tag{3}$$

eq:NSF_3_p2

where $\mathbf{U}(t,x,y,z) = (u,v,w)^T$ is the velocity, $\mathbf{u} = (u,v)^T$ is the horizontal velocity vector and $\rho$ is the density. The notation $\nabla$ denotes $\nabla = \left(\frac{\partial}{\partial x}, (\frac{\partial}{\partial y}, (\frac{\partial}{\partial z}\right)^T$, $\nabla_{x,y}$ corresponds to the projection of $\nabla$ on the horizontal plane i.e. $\nabla_{x,y} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)^T$ and the quantity $|\nabla_{x,y}\mathbf{u}|^2$ means $|\nabla_{x,y}\mathbf{u}|^2 = (\nabla_{x,y}\mathbf{u}) : (\nabla_{x,y}\mathbf{u})^T$.

The temperature $T$ is linked to the density $\rho$ via the state equation $T = T(\rho)$. The viscosity coefficient is denoted by $\mu$, the heat conductivity by $\lambda$ and the specific heat capacity at constant pressure by $c_p$.

The energy balance for model (1)-(3) is

$$\frac{\partial}{\partial t}\left(\rho\frac{|\mathbf{u}|^2}{2} + \rho e\right) + \nabla_{x,y}\cdot\left(\mathbf{u}\left(\rho\frac{|\mathbf{u}|^2}{2} + \int_z^\eta \rho g dz_1 + \rho e\right) - \mu\nabla\frac{|\mathbf{u}|^2}{2}\right) = \nabla.(\lambda\nabla T),$$

with $e$ is the internal energy of the fluid governed by

$$\frac{\partial(\rho e)}{\partial t} + \nabla_{x,y}\cdot(\rho e\mathbf{u}) = \Big(\int_z^\eta \rho g dz_1\Big)\nabla_{x,y}\cdot\mathbf{u} + \mu|\nabla_{x,y}\mathbf{u}|^2 + \mu\left|\frac{\partial \mathbf{u}}{\partial z}\right|^2 + \nabla.(\lambda\nabla T).$$

We consider a free surface flow, therefore we assume

$$z_b(x,y) \le z \le \eta(t,x,y) := h(t,x,y) + z_b(x,y),$$

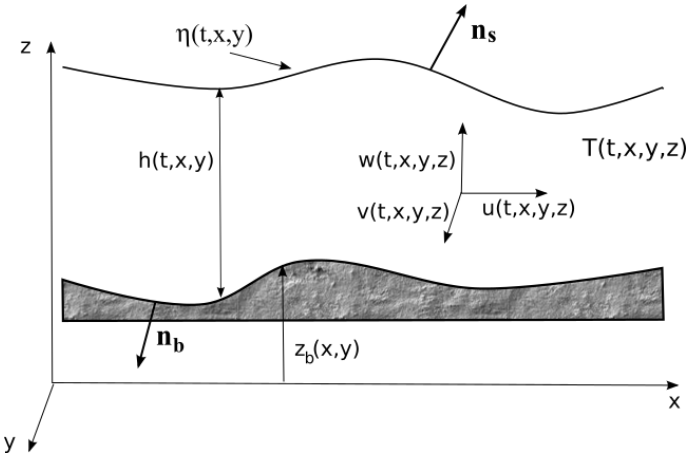


Figure 1: Flow domain with water height $h(t,x,y)$, free surface $\eta(t,x,y)$ and bottom $z_b(x,y)$.

tations3d_rho⟩

with $z_b(x,y)$ the bottom elevation and $h(t,x,y)$ the water depth, see Fig. 1.

Let $\mathbf{n}_b$ and $\mathbf{n}_s$ be the unit outward normals at the bottom and at the free surface respectively defined by

$$\mathbf{n}_b = \frac{1}{\sqrt{1+|\nabla_{x,y} z_b|^2}} \begin{pmatrix} \nabla_{x,y} z_b \\ -1 \end{pmatrix}, \quad \text{and} \quad \mathbf{n}_s = \frac{1}{\sqrt{1+|\nabla_{x,y}\eta|^2}} \begin{pmatrix} -\nabla_{x,y}\eta \\ 1 \end{pmatrix}.$$

On the bottom we prescribe an impermeability condition

$$\mathbf{U}.\mathbf{n}_b = 0, \tag{4}$$

eq:bottom_p2

and a friction condition given e.g. by a Navier law

$$\mu\sqrt{1+|\nabla_{x,y} z_b|^2}\frac{\partial \mathbf{u}}{\partial \mathbf{n}_b} = -\kappa \mathbf{u}, \tag{5}$$

?eq:fric_p2?

with $\kappa$ a Navier coefficient.

On the free surface, the kinematic boundary condition

$$\frac{\partial \eta}{\partial t} + \mathbf{u}(t,x,y,\eta)\cdot\nabla_{x,y}\eta - w(t,x,y,\eta) = 0, \tag{6}$$

eq:free_surf_p

is satisfied, along with the no stress condition

$$\mu\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{n}_s} = 0, \tag{7}$$

?eq:bound3ns_p

with $\tilde{\mathbf{u}} = (\mathbf{u}, 0)^T$.

On solid walls, we prescribe a slip condition

$$\mathbf{U}\cdot\mathbf{n} = 0, \tag{8}$$

?eq:slip?

coupled with an homogeneous Neumann boundary condition

$$\mu \frac{\partial \mathbf{u}}{\partial \mathbf{n}} = 0,$$

$\mathbf{n}$ being the outward normal to the considered wall.

Boundary conditions for the temperature also have to be considered, we can choose either Neumann or Dirichlet conditions namely at the bottom

$$\lambda \frac{\partial T}{\partial \mathbf{n}_b} = \phi_T|_{z_b}, \tag{9}$$

BC:neumann_bot

or

$$T_b = T_b^0 \tag{10}$$

BC:dirichlet_b

and at the free surface

$$\lambda \frac{\partial T}{\partial \mathbf{n}_s} = \phi_T|_{\eta} \tag{11}$$

BC:neumann_sur

or

$$T_s = T_s^0 \tag{12}$$

BC:dirichlet_s

where $\phi_T|_{z_b}$, $\phi_T|_{\eta}$ are two given temperature fluxes and $T_b^0$, $T_s^0$ are two given temperatures. Since $T = T(\rho)$, the boundary conditions for $\rho$ naturally ensue from the boundary conditions for $T$.

The system is completed with some initial conditions

$$h(0,x,y) = h^0(x,y), \quad \rho(0,x,y) = \rho^0(x,y), \quad \mathbf{U}(0,x,y,z) = \mathbf{U}^0(x,y,z).$$

The system (1)-(3) was derived from the compressible Navier-Stokes system in (Boittin et al., 2020). More specifically, the derivation consisted in performing the incompressible limit. This model respects the second principle of thermodynamics (non-decreasing entropy).

## 2.1 The layer-averaged Navier-Stokes-Fourier model

f_multilayer⟩? We consider a discretization of the fluid domain by layers (see Fig. 2) where the layer $\alpha$ contains the points of coordinates $(x,y,z)$ with $z \in L_\alpha(t,x,y) = (z_{\alpha-1/2}, z_{\alpha+1/2})$ and $\{z_{\alpha+1/2}\}_{\alpha=1,\dots,N}$ is defined by

$$\begin{cases} z_{\alpha+1/2}(t,x,y) = z_b(x,y) + \sum_{j=1}^{\alpha} h_j(t,x,y), \quad \alpha \in [0,\dots,N], \\ h_\alpha(t,x,y) = z_{\alpha+1/2}(t,x,y) - z_{\alpha-1/2}(t,x,y) = l_\alpha h(t,x,y), \end{cases} \tag{13}$$

eq:layer

and $\sum_{\alpha=1}^{N} l_\alpha = 1$.

Using the notations (13), let us consider the space $\mathbb{P}_{0,h}^{N,t}$ of piecewise constant functions defined by

$$\mathbb{P}_{0,h}^{N,t} = \left\{ \mathbb{1}_{z \in L_\alpha(t,x,y)}(z), \quad \alpha \in \{1,\dots,N\} \right\}, \tag{14}$$

?eq:P0_space?

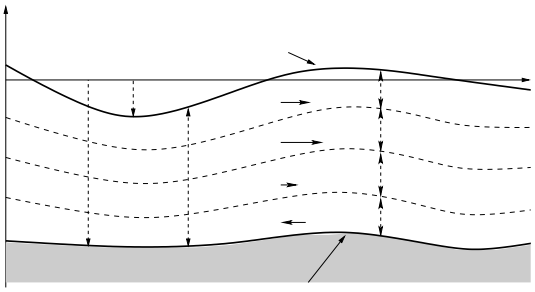

Figure 2: Notations for the layerwise discretization.

⟨fig:free_p2⟩

where $\mathbb{1}_{z\in L_\alpha(t,x,y)}(z)$ is the characteristic function of the layer $L_\alpha(t,x,y)$. Using this formalism, the projection of $\rho$, $u$, $v$ and $w$ on $\mathbb{P}^{N,t}_{0,h}$ is a piecewise constant function defined by

$$X^N(t,x,y,z,\{z_\alpha\}) = \sum_{\alpha=1}^{N} \mathbb{1}_{z\in L_\alpha(t,x,y)}(z) X_\alpha(t,x,y), \tag{15}$$ ?eq:ulayer?

for $X \in (\rho, u, v, w)$. When the quantities $\{\rho_\alpha(t,x,y)\}_{\alpha=1,...,N}$ are known, if the function $T = T(\rho)$ is known, it is possible to recover the temperature using the formula

$$T^N(t,x,z,\{z_\alpha\}) = \sum_{\alpha=1}^{N} \mathbb{1}_{z\in L_\alpha(t,x,y)}(z) T(\rho_\alpha(t,x,y)).$$

The layer-averaged Navier-Stokes-Fourier system introduced in (Boittin et al., 2020) reads

$$\frac{\partial h}{\partial t} + \sum_{\alpha=1}^{N} \nabla_{x,y}\cdot(h_\alpha \mathbf{u}_\alpha) = -\sum_{\alpha=1}^{N} \frac{\rho'(T_\alpha)}{\rho_\alpha^2 c_p}(\mathcal{S}_{T,\alpha} - \mathcal{S}_{\mu,\alpha}), \tag{16}$$ eq:massesvml

$$\frac{\partial \rho_\alpha h_\alpha}{\partial t} + \nabla_{x,y}\cdot(\rho_\alpha h_\alpha \mathbf{u}_\alpha) = \rho_{\alpha+1/2} G_{\alpha+1/2} - \rho_{\alpha-1/2} G_{\alpha-1/2}, \qquad \alpha = 1,...,N, \tag{17}$$ ?eq:massesvm

$$\begin{aligned}\frac{\partial \rho_\alpha h_\alpha \mathbf{u}_\alpha}{\partial t} + \nabla_{x,y}\cdot(\rho_\alpha h_\alpha \mathbf{u}_\alpha \otimes \mathbf{u}_\alpha) + \nabla_{x,y}\big(h_\alpha p_\alpha\big) &= p_{\alpha+1/2}\nabla_{x,y} z_{\alpha+1/2} - p_{\alpha-1/2}\nabla_{x,y} z_{\alpha-1/2} \\ &+\mathbf{u}_{\alpha+1/2}\rho_{\alpha+1/2}G_{\alpha+1/2} - \mathbf{u}_{\alpha-1/2}\rho_{\alpha-1/2}G_{\alpha-1/2} + \nabla_{x,y}\cdot(\mu h_\alpha \nabla_{x,y}\mathbf{u}_\alpha) \\ &+\Gamma_{\alpha+1/2}(\mathbf{u}_{\alpha+1} - \mathbf{u}_\alpha) - \Gamma_{\alpha-1/2}(\mathbf{u}_\alpha - \mathbf{u}_{\alpha-1}) - \kappa_\alpha \mathbf{u}_\alpha, \qquad \alpha = 1,...,N,\end{aligned} \tag{18}$$ eq:mvtsvml_

with

$$G_{\alpha+1/2} = -\sum_{j=1}^{N}\Big(\sum_{p=1}^{\alpha} l_p - \mathbb{1}_{j\le\alpha}\Big)\nabla_{x,y}\cdot(h_j\mathbf{u}_j) + \sum_{j=1}^{\alpha}\frac{\rho'(T_j)}{\rho_j^2 c_p}(S_{T,j} - S_{\mu,j}), \tag{19}$$ ?eq:Qalphabis_

$$\kappa_\alpha = \begin{cases}\kappa & \text{if} \quad \alpha = 1 \\ 0 & \text{if} \quad \alpha \neq 1\end{cases},$$

$$\mathcal{S}_{T,\alpha} = \left( \lambda \nabla_{x,y} \cdot (h_\alpha \nabla_{x,y} T_\alpha) + 2\lambda_{\alpha+1/2} \frac{T_{\alpha+1} - T_\alpha}{h_{\alpha+1} + h_\alpha} - 2\lambda_{\alpha-1/2} \frac{T_\alpha - T_{\alpha-1}}{h_\alpha + h_{\alpha-1}} \right), \quad (20)$$ S_T_alpha

$$\lambda_{\alpha+1/2} = \lambda \quad \text{for} \quad \alpha = 1, \ldots, N-1,$$
$$T_\alpha = T(\rho_\alpha).$$

For $\alpha = 0$, $2\lambda_{\alpha+1/2} \frac{T_{\alpha+1} - T_\alpha}{h_{\alpha+1} + h_\alpha} = \phi_T|_{z_b}$ if the Neumann boundary condition (9) is chosen, or $h_0 = h_1$, $T_0 = T_b^0$ if the Dirichlet boundary condition (10) is chosen. Likewise, for $\alpha = N$, $2\lambda_{\alpha+1/2} \frac{T_{\alpha+1} - T_\alpha}{h_{\alpha+1} + h_\alpha} = \phi_T|_\eta$ with the boundary condition (11), or $h_{N+1} = h_N$, $T_{N+1} = T_s^0$ with the boundary condition (12). The dissipation term due to the viscous effects is

$$\mathcal{S}_{\mu,\alpha} = -h_\alpha \mu |\nabla_{x,y} \mathbf{u}_\alpha|^2 - \Gamma_{\alpha+1/2} \frac{|\mathbf{u}_{\alpha+1} - \mathbf{u}_\alpha|^2}{2} - \Gamma_{\alpha-1/2} \frac{|\mathbf{u}_\alpha - \mathbf{u}_{\alpha-1}|^2}{2} - \kappa_\alpha |\mathbf{u}_\alpha|^2, \quad (21)$$ S_mu_alpha

with

$$\Gamma_{\alpha+1/2} = \frac{2\mu_{\alpha+1/2}}{h_{\alpha+1} + h_\alpha}, \quad (22)$$ Gamma_alpha_p

$$\mu_{\alpha+1/2} = \begin{cases} 0 & \text{if} \quad \alpha = 0 \\ \mu & \text{if} \quad \alpha = 1, \ldots, N-1 \\ 0 & \text{if} \quad \alpha = N. \end{cases} \quad (23)$$ ?mu_alpha_plus

The term $|\nabla_{x,y} \mathbf{u}_\alpha|^2$ actually denotes

$$\begin{aligned} |\nabla_{x,y} \mathbf{u}_\alpha|^2 &= (\nabla_{x,y} \mathbf{u}_\alpha) : (\nabla_{x,y} \mathbf{u}_\alpha)^T \\ &= \left( \frac{\partial u_\alpha}{\partial x} \right)^2 + \left( \frac{\partial u_\alpha}{\partial y} \right)^2 + \left( \frac{\partial v_\alpha}{\partial x} \right)^2 + \left( \frac{\partial v_\alpha}{\partial y} \right)^2. \end{aligned}$$

The velocities $\mathbf{u}_{\alpha+1/2}$ and the densities $\rho_{\alpha+1/2}$ at the interfaces are defined by

$$v_{\alpha+1/2} = \begin{cases} v_\alpha & \text{if} \ \ G_{\alpha+1/2} \leq 0 \\ v_{\alpha+1} & \text{if} \ \ G_{\alpha+1/2} > 0 \end{cases} \quad (24)$$ ?eq:upwind_uT_

for $v = \mathbf{u}, \rho$.
The pressure terms $p_\alpha$, $p_{\alpha+1/2}$ are given by

$$p_\alpha = g \left( \frac{\rho_\alpha h_\alpha}{2} + \sum_{j=\alpha+1}^{N} \rho_j h_j \right) \qquad \text{and} \qquad p_{\alpha+1/2} = g \sum_{j=\alpha+1}^{N} \rho_j h_j. \quad (25)$$ ?eq:palpha1_p2

The pressure is hydrostatic. The terms $G_{\alpha+1/2}$ represent the mass exchanges between the layers. Notice that some of the viscous and diffusion terms have been simplified, see (Boittin et al., 2020). Since the right-hand side of the total height conservation equation (16) is nonzero, we expect to observe dilatation and contraction due to the temperature diffusion and to the viscosity.

We recall here the following result, obtained in (Boittin et al., 2020).

y_balance_ns⟩? **Proposition 2.1** *The system* (16)-(18) *completed with the equation*

$$\frac{\partial}{\partial t}(\rho_\alpha h_\alpha e_\alpha) + \nabla_{x,y}\cdot(\rho_\alpha h_\alpha \mathbf{u}_\alpha e_\alpha) = \rho_{\alpha+1/2}e_{\alpha+1/2}G_{\alpha+1/2} - \rho_{\alpha-1/2}e_{\alpha-1/2}G_{\alpha-1/2} + p_\alpha \frac{\rho'(T_\alpha)}{\rho_\alpha^2 c_p}(\mathcal{S}_{T,\alpha} - \mathcal{S}_{\mu,\alpha}) + \mathcal{S}_{T,\alpha} - \mathcal{S}_{\mu,\alpha}$$

*admits, for smooth solutions, the energy balance*

$$\begin{aligned}
&\frac{\partial}{\partial t}E_\alpha + \nabla_{x,y}\cdot\left(\mathbf{u}_\alpha\left(E_\alpha + h_\alpha p_\alpha - \mu h_\alpha \nabla_{x,y}\mathbf{u}_\alpha\right)\right)\\
&+\Gamma_{\alpha+1/2}\frac{|\mathbf{u}_{\alpha+1}|^2 - |\mathbf{u}_\alpha|^2}{2} - \Gamma_{\alpha-1/2}\frac{|\mathbf{u}_\alpha|^2 - |\mathbf{u}_{\alpha-1}|^2}{2}\\
&\quad= \left(\rho_{\alpha+1/2}\frac{\mathbf{u}_{\alpha+1/2}^2}{2} + g\rho_{\alpha+1/2}z_{\alpha+1/2}\right)G_{\alpha+1/2} + p_{\alpha+1/2}\big(G_{\alpha+1/2} - \frac{\partial z_{\alpha+1/2}}{\partial t}\big)\\
&\quad\quad- \left(\rho_{\alpha-1/2}\frac{\mathbf{u}_{\alpha-1/2}^2}{2} + g\rho_{\alpha-1/2}z_{\alpha-1/2}\right)G_{\alpha-1/2} - p_{\alpha-1/2}\big(G_{\alpha-1/2} - \frac{\partial z_{\alpha-1/2}}{\partial t}\big)\\
&\quad\quad-\frac{1}{2}\left(\rho_{\alpha+1/2}(\mathbf{u}_{\alpha+1/2} - \mathbf{u}_\alpha)^2 + gh_\alpha(\rho_{\alpha+1/2} - \rho_\alpha)\right)G_{\alpha+1/2}\\
&\quad\quad+\frac{1}{2}\left(\rho_{\alpha-1/2}(\mathbf{u}_{\alpha-1/2} - \mathbf{u}_\alpha)^2 - gh_\alpha(\rho_{\alpha-1/2} - \rho_\alpha)\right)G_{\alpha-1/2} + \mathcal{S}_{T,\alpha},
\end{aligned} \tag{26}$$

eq:energy_mcl

*with*

$$E_\alpha = \rho_\alpha\frac{h_\alpha|\mathbf{u}_\alpha|^2}{2} + \frac{\rho_\alpha g h_\alpha z_\alpha}{2} + e_\alpha. \tag{27}$$

?eq:energ_al_N

Note that in (26), we use the notation

$$\mathbf{u}_\alpha\nabla_{x,y}\mathbf{u}_\alpha = \begin{pmatrix} u_\alpha\frac{\partial u_\alpha}{\partial x} + v_\alpha\frac{\partial v_\alpha}{\partial x}\\ u_\alpha\frac{\partial u_\alpha}{\partial y} + v_\alpha\frac{\partial v_\alpha}{\partial y}\end{pmatrix} = \nabla_{x,y}\frac{|\mathbf{u}_\alpha^2|}{2}.$$

The sum of Eqs. (26) for $\alpha = 1,\dots,N$ gives

$$\begin{aligned}
&\frac{\partial}{\partial t}\sum_{\alpha=1}^N E_\alpha + \sum_{\alpha=1}^N \nabla_{x,y}\cdot\mathbf{u}_\alpha\left(E_\alpha + h_\alpha p_\alpha\right)\\
&\quad= -\sum_{\alpha=1}^N \rho_{\alpha+1/2}\frac{|\mathbf{u}_{\alpha+1} - \mathbf{u}_\alpha|^2}{2}|G_{\alpha+1/2}| + \sum_{\alpha=1}^N \mathcal{S}_{T,\alpha}\\
&\quad\quad-\frac{g}{2}\sum_{\alpha=1}^N\left(h_\alpha(\rho_{\alpha+1/2} - \rho_\alpha) + h_{\alpha+1}(\rho_{\alpha+1/2} - \rho_{\alpha+1})\right)G_{\alpha+1/2}.
\end{aligned}$$

The sum of $\mathcal{S}_{T,\alpha}$ over the layers gives

$$\sum_{\alpha=1}^N \mathcal{S}_{T,\alpha} = \lambda\sum_{\alpha=1}^N \nabla_{x,y}\cdot(h_\alpha\nabla_{x,y}T_\alpha) - \nabla T|_s\cdot\mathbf{n}_s + \nabla T|_b\cdot\mathbf{n}_b.$$

As explained in (Boittin et al., 2020), the terms on the last line of the right-hand side are third-order terms.

## 2.2 The layer-averaged Euler system

What we refer to hereafter as the layer-averaged Euler system is the system (16)-(18) without viscosity and without diffusion terms, i.e.

$$\frac{\partial h}{\partial t} + \sum_{\alpha=1}^{N} \nabla_{x,y} \cdot (h_\alpha \mathbf{u}_\alpha) = 0, \tag{28}$$

eq:massesvm

$$\frac{\partial \rho_\alpha h_\alpha}{\partial t} + \nabla_{x,y} \cdot (\rho_\alpha h_\alpha \mathbf{u}_\alpha) = \rho_{\alpha+1/2} G_{\alpha+1/2} - \rho_{\alpha-1/2} G_{\alpha-1/2}, \qquad \alpha = 1, ..., N, \tag{29}$$

?eq:massesvm

$$\frac{\partial \rho_\alpha h_\alpha \mathbf{u}_\alpha}{\partial t} + \nabla_{x,y} \cdot (\rho_\alpha h_\alpha \mathbf{u}_\alpha \otimes \mathbf{u}_\alpha) + \nabla_{x,y}\big(h_\alpha p_\alpha\big) = p_{\alpha+1/2} \nabla_{x,y} z_{\alpha+1/2} - p_{\alpha-1/2} \nabla_{x,y} z_{\alpha-1/2}$$
$$+ \mathbf{u}_{\alpha+1/2} \rho_{\alpha+1/2} G_{\alpha+1/2} - \mathbf{u}_{\alpha-1/2} \rho_{\alpha-1/2} G_{\alpha-1/2}, \qquad \alpha = 1, ..., N. \tag{30}$$

eq:mvtsvml_p

The quantity $G_{\alpha+1/2}$ (resp. $G_{\alpha-1/2}$) corresponds to mass exchange accross the interface $z_{\alpha+1/2}$ (resp. $z_{\alpha-1/2}$) and $G_{\alpha+1/2}$ is defined here by

$$G_{\alpha+1/2} \;=\; \sum_{j=1}^{\alpha} \left( \frac{\partial h_j}{\partial t} + \nabla_{x,y} \cdot (h_j \mathbf{u}_j) \right) = - \sum_{j=1}^{N} \Big( \sum_{p=1}^{\alpha} l_p - \mathbb{1}_{j \leq \alpha} \Big) \nabla_{x,y} \cdot (h_j \mathbf{u}_j), \tag{31}$$

eq:Qalphabis_p

for $\alpha = 1, \ldots, N$. Notice that the mass conservation for the layer $\alpha$ writes

$$\frac{\partial h_\alpha}{\partial t} + \nabla_{x,y} \cdot (h_\alpha \mathbf{u}_\alpha) = G_{\alpha+1/2} - G_{\alpha-1/2}, \tag{32}$$

eq:exchange_te

and the sum (total or partial) of the previous equations with $G_{1/2} = G_{N+1/2} = 0$ (see (Boittin et al., 2020, Prop. 3.1)) gives Eqs. (28) and (31).

The energy balance verified by the system (28)-(30) is given in (Boittin et al., 2020). It is very similar to the balance (26), obviously without the viscosity and diffusion terms. In the balance for the Euler system, the internal energy $e_\alpha$ does not intervene. It is actually equal to 0 because there is no volume variation.

# 3 Numerical scheme for the layer-averaged Euler system

sec:euler_num

In this section, a numerical scheme for the layer-averaged Euler system is designed and analyzed. It extends the work done by some of the authors in (Allgeyer et al., 2019; Audusse et al., 2011). Before specifying the scheme for the Euler system, a common strategy for the time discretization of the Navier-Stokes-Fourier and Euler systems is presented. The discretization of the diffusion terms does not present any additional difficulty, however including these terms considerably lengthens the equations. This is why it seems preferable to explain the numerical scheme for the Euler system first.

The advantages of the numerical scheme are the following

- it gives a 3D approximation of the Navier-Stokes-Fourier system, while only 2D situations were considered in (Audusse et al., 2011)

- it can be implemented with any flux that is consistent with the homogeneous Saint-Venant system; the kinetic flux is used only for the discrete entropy property stated for a constant topography
- the scheme is endowed with strong stability properties (well-balanced, positivity of the water depth)
- convergence curves towards a 3D non-stationary analytical solution with wet-dry interfaces are obtained, see paragraph 5.1.

## 3.1 Strategy for the time discretization

The system (16)-(18) has the form

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla_{x,y} \cdot F(\mathbf{U}) = \mathcal{S}_p(\mathbf{U}, z_b) + S_e(\mathbf{U}, \partial_t \mathbf{U}, \partial_x \mathbf{U}) + S_{v,f}(\mathbf{U}), \tag{33}$$ eq:glo_p2

where the vector of unknowns is

$$\mathbf{U} = \left(h, \rho_1 h_1, \ldots, \rho_N h_N, q_{x,1}, \ldots, q_{x,N}, q_{y,1}, \ldots, q_{y,N}\right)^T,$$

with $q_{x,\alpha} = \rho_\alpha h_\alpha u_\alpha$, $q_{y,\alpha} = \rho_\alpha h_\alpha v_\alpha$. We denote by $F(\mathbf{U}) = (F_x(\mathbf{U}), F_y(\mathbf{U}))^T$ the fluxes of the conservative part and by

$$\mathcal{S}_p(\mathbf{U}, z_b) = \left(0, \ldots, p_{3/2}\frac{\partial z_{3/2}}{\partial x} - p_{1/2}\frac{\partial z_{1/2}}{\partial x}, \ldots, p_{3/2}\frac{\partial z_{3/2}}{\partial y} - p_{1/2}\frac{\partial z_{1/2}}{\partial y}, \ldots\right)^T,$$

the non-conservative part of the pressure terms. The source terms are $S_e(\mathbf{U}, \partial_t \mathbf{U}, \partial_x \mathbf{U})$ and $S_{v,f}(\mathbf{U})$, representing respectively the mass and momentum exchanges and the viscous and friction effects. Notice that, as a consequence of the layer-averaged discretization, the system (33) is made of only 2d $(x, y)$ partial differential equations with source terms. Hence, the spacial approximation of the considered PDEs is performed on a 2d planar mesh.

We consider discrete times $t^n$ with $t^{n+1} = t^n + \Delta t^n$. For the time discretisation of the layer-averaged Navier-Stokes-Fourier system (33) we adopt the following scheme

$$\mathbf{U}^{n+1} = \mathbf{U} - \Delta t^n \left(\nabla_{x,y} \cdot F(\mathbf{U}) - \mathcal{S}_p(\mathbf{U}, z_b)\right) + \Delta t^n S_e^{n+1} + \Delta t^n S_{v,f}^{n+l}, \tag{34}$$ eq:glo_dis

where the integer $l = 0, 1/2, 1$ will be precised below. In (34) and wherever there is no ambiguity the superscript $^n$ has been omitted.

## 3.2 Semi-discrete (in time) scheme

From now on and until the end of this section, the system considered is the Euler system. Similarly to (Allgeyer et al., 2019), the semi-discrete in time scheme (34) with $S_{v,f}^{n+l} = 0$ writes

$$h_\alpha^{n+1/2} = h_\alpha - \Delta t^n \nabla_{x,y}(h_\alpha \mathbf{u}_\alpha) \quad (35)$$ eq:nscont_laye

$$(\rho_\alpha h_\alpha)^{n+1/2} = \rho_\alpha h_\alpha - \Delta t^n \nabla_{x,y} \cdot (\rho_\alpha h_\alpha \mathbf{u}_\alpha) \quad (36)$$ eq:nsml22_dbis

$$\begin{aligned}(\rho_\alpha h_\alpha \mathbf{u}_\alpha)^{n+1/2} = {} & \rho_\alpha h_\alpha \mathbf{u}_\alpha - \Delta t^n \Big( \nabla_{x,y} \cdot (\rho_\alpha h_\alpha \mathbf{u}_\alpha \otimes \mathbf{u}_\alpha) + \nabla_{x,y}\big(h_\alpha p_\alpha\big) \\ & - p_\alpha \nabla_{x,y} h_\alpha + \rho_\alpha g h_\alpha \nabla_{x,y} z_\alpha \Big), \end{aligned} \quad (37)$$ eq:nsml22_d

$$h^{n+1} = h^{n+1/2} = \sum_{\alpha=1}^{N} h_\alpha^{n+1/2} = h - \Delta t^n \sum_{\alpha=1}^{N} \nabla_{x,y} \cdot (h_\alpha \mathbf{u}_\alpha), \quad (38)$$ eq:nsml11_d

$$(\rho_\alpha h_\alpha)^{n+1} = (\rho_\alpha h_\alpha)^{n+1/2} + \Delta t^n \Big( \rho_{\alpha+1/2}^{n+1} G_{\alpha+1/2} - \rho_{\alpha-1/2}^{n+1} G_{\alpha-1/2} \Big), \quad (39)$$ eq:nsml33bis_d

$$(\rho_\alpha h_\alpha \mathbf{u}_\alpha)^{n+1} = (\rho_\alpha h_\alpha \mathbf{u}_\alpha)^{n+1/2} + \Delta t^n \Big( \mathbf{u}_{\alpha+1/2}^{n+1} \rho_{\alpha+1/2}^{n+1} G_{\alpha+1/2} - \mathbf{u}_{\alpha-1/2}^{n+1} \rho_{\alpha-1/2}^{n+1} G_{\alpha-1/2} \Big) \quad (40)$$ eq:nsmlbis_d

with

$$G_{\alpha+1/2} = -\sum_{j=1}^{N} \Big( \sum_{p=1}^{\alpha} l_p - \mathbb{1}_{j \leq \alpha} \Big) \nabla_{x,y} \cdot (h_j \mathbf{u}_j). \quad (41)$$ eq:nsf_6d

Since the relations

$$p_{\alpha \pm 1/2} = p_\alpha \mp \frac{\rho_\alpha g h_\alpha}{2}, \quad z_\alpha = \frac{z_{\alpha+1/2} + z_{\alpha-1/2}}{2}$$

hold, in (37) we often use the identity

$$p_{\alpha+1/2} \nabla_{x,y} z_{\alpha+1/2} - p_{\alpha-1/2} \nabla_{x,y} z_{\alpha-1/2} = p_\alpha \nabla_{x,y} h_\alpha - \rho_\alpha g h_\alpha \nabla_{x,y} z_\alpha. \quad (42)$$ eq:S_P_alpha

Notice that $h_\alpha^{n+1} = l_\alpha h^{n+1}$ and $h_\alpha^{n+1} \neq h_\alpha^{n+1/2}$, see Eq. (32).

The first three equations (35)-(37) consist in an explicit time scheme where the horizontal fluxes and the pressure terms are taken into account whereas in Eqs. (39)-(40) an implicit treatment of the exchange terms between layers is proposed. The implicit part of the scheme requires to solve a linear problem, see paragraph 3.4.3 and Allgeyer et al. (2019).

Following (34), Eqs. (35)-(38) also writes

$$\mathbf{U}^{n+1/2} = \mathbf{U} - \Delta t^n \left( \nabla_{x,y} \cdot F(\mathbf{U}) - \mathcal{S}_p(\mathbf{U}, z_b) \right), \quad (43)$$ eq:glo_dis1

these computations being detailed in paragraphs 3.4.1 and 3.4.2. Equations. (38)-(40) can be reformulated under the form

$$\mathbf{U}^{n+1} = \mathbf{U}^{n+1/2} + \Delta t^n S_e^{n+1}, \quad (44)$$ ?eq:glo_dis2?

this implicit step being precised in paragraph 3.4.3.

### 3.3 Eigenvalues for the advection and pressure part

In order to propose a finite volume discretisation for Eqs. (35)-(38), an estimation of the eigenvalues is necessary and given below.

Without the exchange terms, the system (28)-(30) writes

$$\frac{\partial h}{\partial t} + \sum_{\alpha=1}^{N} \nabla_{x,y} \cdot (h_\alpha \mathbf{u}_\alpha) = 0, \tag{45}$$

$$\frac{\partial (\rho_\alpha h_\alpha)}{\partial t} + \nabla_{x,y} \cdot (\rho_\alpha h_\alpha \mathbf{u}_\alpha) = 0, \tag{46}$$

$$\frac{\partial (\rho_\alpha h_\alpha \mathbf{u}_\alpha)}{\partial t} + \nabla_{x,y} \cdot (\rho_\alpha h_\alpha \mathbf{u}_\alpha \otimes \mathbf{u}_\alpha) + p_\alpha \nabla_{x,y} h_\alpha - \rho_\alpha g h_\alpha \nabla_{x,y} z_\alpha = 0, \tag{47}$$

with

$$p_\alpha \nabla_{x,y} h_\alpha - \rho_\alpha g h_\alpha \nabla_{x,y} z_\alpha = g h_\alpha \left( \sum_{j=\alpha+1}^{N} \nabla_{x,y} (\rho_j h_j) + \frac{1}{2} \nabla_{x,y} (\rho_\alpha h_\alpha) \right) + \frac{\rho_\alpha g h_\alpha}{2} \nabla_{x,y} h_\alpha + \rho_\alpha g h_\alpha \left( \sum_{j=1}^{\alpha-1} \nabla_{x,y} h_j + \nabla_{x,y} z_b \right).$$

This system is the continuous version of system (35)-(38).

We rewrite the system (45)-(47) under the form

$$\frac{\partial h}{\partial t} + \sum_{j=1}^{N} (l_j h \nabla_{x,y} \cdot \mathbf{u}_j) + \sum_{j=1}^{N} (l_j \mathbf{u}_j) \cdot \nabla_{x,y} h = 0, \tag{48}$$

$$\frac{\partial \rho_\alpha}{\partial t} + \mathbf{u}_\alpha . \nabla_{x,y} \rho_\alpha = 0, \tag{49}$$

$$\frac{\partial \mathbf{u}_\alpha}{\partial t} + (\mathbf{u}_\alpha \cdot \nabla_{x,y}) \mathbf{u}_\alpha + \frac{1}{\rho_\alpha h_\alpha} \left( p_\alpha \nabla_{x,y} h_\alpha - \rho_\alpha g h_\alpha \nabla_{x,y} z_\alpha \right) = 0, \tag{50}$$

and the quasilinear form of the system (48)-(50) writes

$$\frac{\partial \tilde{\mathbf{U}}}{\partial t} + A(\tilde{\mathbf{U}}) \nabla_{x,y} \tilde{\mathbf{U}} = s_b(\tilde{\mathbf{U}}), \tag{51}$$

with

$$\tilde{\mathbf{U}} = (h, \mathbf{u}_1, \ldots, \mathbf{u}_N, \rho_1, \ldots, \rho_N)^T,$$

and

$$A(\tilde{\mathbf{U}}) = \begin{pmatrix} A_1(\tilde{\mathbf{U}}) & A_2(\tilde{\mathbf{U}}) \\ A_3(\tilde{\mathbf{U}}) & A_4(\tilde{\mathbf{U}}) \end{pmatrix},$$

$$A_1(\tilde{\mathbf{U}}) = \begin{pmatrix} \sum_{j=1}^N l_j u_j & l_1 h & \dots & \dots & \dots & l_N h \\ \tilde{p}_1 & u_1 & 0 & \dots & \dots & 0 \\ \tilde{p}_2 & 0 & u_2 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & u_j & \ddots & 0 \\ \vdots & 0 & \ddots & 0 & \ddots & 0 \\ \tilde{p}_N & 0 & 0 & \dots & 0 & u_N \end{pmatrix},$$

$$A_2(\tilde{\mathbf{U}}) = \begin{pmatrix} 0 & \dots & \dots & 0 \\ \frac{gh_1^2}{2} & 0 & \ddots & 0 \\ gh_2h_1 & \frac{gh_2^2}{2} & 0 & 0 \\ \vdots & 0 & \frac{gh_j^2}{2} & 0 \\ gh_Nh_1 & 0 & 0 & \frac{gh_N^2}{2} \end{pmatrix},$$

$$A_3(\tilde{\mathbf{U}}) = \mathbf{0}_{N\times N+1}, \qquad A_4(\tilde{\mathbf{U}}) = \mathrm{diag}(u_j),$$

$$\tilde{p}_j = \frac{g}{\rho_j}\left(\rho_j l_j + \rho_j \sum_{i=1}^{j-1} l_i + \sum_{i=1}^{j-1} \rho_i l_i\right).$$

For the sake of simplicity, the expression of the matrix $A(\tilde{\mathbf{U}})$ given below corresponds to the 1D case i.e. for $v_i = 0$, $i = 1, \dots, N$. Notice that $A_2(\tilde{\mathbf{U}})$ and $A_3(\tilde{\mathbf{U}})^T$ are rectangular matrices with $N+1$ rows and $N$ columns.

The following proposition holds, consisting in a version of the Cauchy's interlace theorem (Hwang, 2004) in the case of non symmetric matrix.

**Proposition 3.1** *The system* (51) *is strictly hyperbolic for* $h > 0$ *and the eigenvalues of* $A(\tilde{\mathbf{U}})$ *belong to the interval* $(\lambda_{min}, \lambda_{max})$ *with*

$$\lambda_{min} = \min_j\{u_j, v_j\} - \frac{\max\{\rho_j\}}{\min\{\rho_j\}}\sqrt{gh},$$
$$\lambda_{max} = \max\{u_j, v_j\} + \frac{\max\{\rho_j\}}{\min\{\rho_j\}}\sqrt{gh}.$$

⟨prop:hyper⟩

**Proof of prop. 3.1** *Since* $A(\tilde{\mathbf{U}})$ *is a block-matrix, its eigenvalues consist in the eigenvalues of* $A_1(\tilde{\mathbf{U}})$ *completed with the set* $\{u_i\}_{i=1}^N$. *Writing the characteristic polynomial of* $A_1(\tilde{\mathbf{U}})$ *under the form (development e.g. along the first row)*

$$P_{A_1} = \Pi_{i=1}^N(\lambda - u_i)\left(\lambda - \sum_{j=1}^N l_j u_j - \sum_{i=1}^N \frac{l_i h \tilde{p}_i}{\lambda - u_i}\right),$$

*the eigenvalues of* $A_1(\tilde{\mathbf{U}})$ *satisfy*

$$\lambda - \sum_{j=1}^N l_j u_j = Q_{A_1}(\lambda),$$

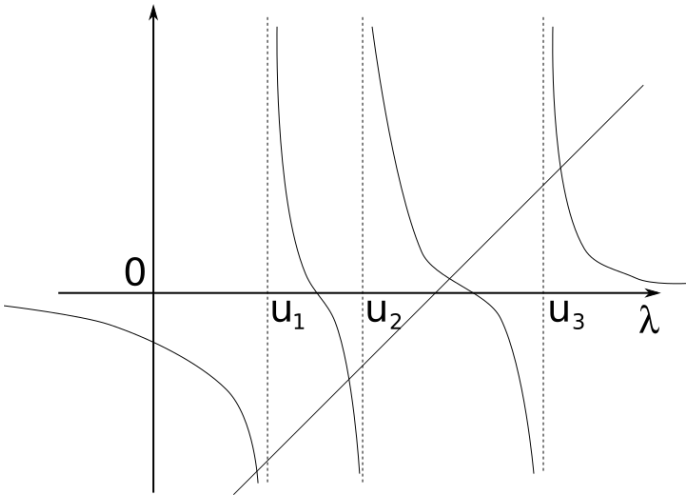


Figure 3: The two functions $\lambda \mapsto Q_{A_1}(\lambda)$ and $\lambda \mapsto \lambda - \sum_{j=1}^{N} l_j u_j$, each intersection of the two curves is an eigenvalue of $A_1(\tilde{\mathbf{U}})$.

⟨fig:pol_carac⟩

*with* $Q_{A_1}(\lambda) = \sum_{i=1}^{N} \frac{l_i h \tilde{p}_i}{\lambda - u_i}$. *For* $N = 3$, *the functions* $\lambda \mapsto Q_{A_1}(\lambda)$ *and* $\lambda \mapsto \lambda - \sum_{j=1}^{N} l_j u_j$ *are depicted over Fig. 3 and it is easy to see that the four eigenvalues* $\lambda_i$ *exists with the interlacing*

$$\lambda_1 < u_1 \leq \lambda_2 \leq \ldots \leq u_3 < \lambda_4.$$

*Moreover we have*

$$Q_{A_1}(\lambda_{max}) = \sum_{i=1}^{N} \frac{l_i h \tilde{p}_i}{\lambda_{max} - u_i} \leq \frac{\min\{\rho_j\}}{\max\{\rho_j\}} \sum_{i=1}^{N} \frac{l_i h \tilde{p}_i}{\sqrt{gh}} \leq \frac{\max\{\rho_j\}}{\min\{\rho_j\}} \sqrt{gh} \leq \lambda_{max} - \sum_{j=1}^{N} l_j u_j,$$

*and likewise*

$$Q_{A_1}(\lambda_{min}) = \sum_{i=1}^{N} \frac{l_i h \tilde{p}_i}{\lambda_{min} - u_i} \geq -\frac{\min\{\rho_j\}}{\max\{\rho_j\}} \sum_{i=1}^{N} \frac{l_i h \tilde{p}_i}{\sqrt{gh}} \geq -\frac{\max\{\rho_j\}}{\min\{\rho_j\}} \sqrt{gh} \geq \lambda_{min} - \sum_{j=1}^{N} l_j u_j,$$

*therefore the eigenvalues* $\{\lambda_i\}_{i=1}^{N}$ *of* $A_1(\tilde{\mathbf{U}})$ *satisfy*

$$\lambda_{min} \leq \lambda_i \leq \lambda_{max}, \qquad i = 1, \ldots N,$$

*proving the result.*

■

## 3.4 Finite volume formalism for the Euler part

⟨subsec:fv⟩ In this paragraph, we propose a space discretization for the model (35)-(40) completed with (41). We first recall the general formalism of finite volumes on unstructured meshes. Let $\Omega$ denote the computational domain with boundary $\Gamma$, which we assume is polygonal.

Let $T_h$ be a triangulation of $\Omega$ for which the vertices are denoted by $P_i$ with $S_i$ the set of interior nodes and $G_i$ the set of boundary nodes. The dual cells $C_i$ are obtained by joining the centers of mass of the triangles surrounding each vertex $P_i$. We use the following notations (see Fig. 4):

- $K_i$, set of subscripts of nodes $P_j$ surrounding $P_i$,
- $|C_i|$, area of $C_i$,
- $\Gamma_{ij}$, boundary edge between the cells $C_i$ and $C_j$,
- $L_{ij}$, length of $\Gamma_{ij}$,
- $\mathbf{n}_{ij}$, unit normal to $\Gamma_{ij}$, outward to $C_i$ ($\mathbf{n}_{ji} = -\mathbf{n}_{ij}$).

If $P_i$ is a node belonging to the boundary $\Gamma$, we join the centers of mass of the triangles adjacent to the boundary to the middle of the edge belonging to $\Gamma$ (see Fig. 4) and we denote

- $\Gamma_i$, the two edges of $C_i$ belonging to $\Gamma$,
- $L_i$, length of $\Gamma_i$ (for sake of simplicity we assume in the following that $L_i = 0$ if $P_i$ does not belong to $\Gamma$),
- $\mathbf{n}_i$, the unit outward normal defined by averaging the two adjacent normals.

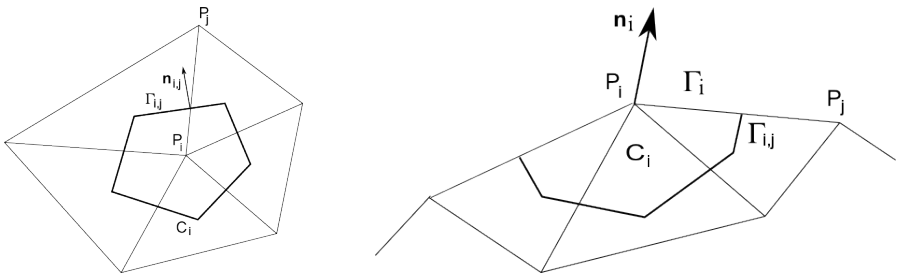


Figure 4: *(a)* Dual cell $C_i$ and *(b)* Boundary cell $C_i$.

⟨fig:mesh⟩

We define the piecewise constant functions $\mathbf{U}^n(x, y)$ on cells $C_i$ corresponding to time $t^n$ as

$$\mathbf{U}^n(x, y) = \mathbf{U}_i^n, \quad \text{for } (x, y) \in C_i,$$

with $\mathbf{U}_i^n = (h_i^n, \rho_{1,i}^n h_{1,i}^n, \ldots, \rho_{N,i}^n h_{N,i}^n, q_{x,1,i}^n, \ldots, q_{x,N,i}^n, q_{y,1,i}^n, \ldots, q_{y,N,i}^n)^T$ i.e.

$$\mathbf{U}_i^n \approx \frac{1}{|C_i|} \int_{C_i} \mathbf{U}(t^n, x, y) dx dy.$$

For the topography, we choose a piecewise constant approximation under the form

$$z_i \approx \frac{1}{|C_i|}\int_{C_i} z_b(x,y)dxdy.$$

We will also use the notation

$$\mathbf{U}^n_{\alpha,i} \approx \frac{1}{|C_i|}\int_{C_i} \mathbf{U}_\alpha(t^n,x,y)dxdy,$$

with $\mathbf{U}_\alpha$ defined by

$$\mathbf{U}_\alpha = (h_\alpha, \rho_\alpha h_\alpha, \rho_\alpha h_\alpha u_\alpha, \rho_\alpha h_\alpha v_\alpha)^T. \tag{52}$$ eq:u_alpha

### 3.4.1 The horizontal fluxes and the pressure terms

ntal_discrete⟩ A finite volume scheme for solving the system (35)-(38) is a formula of the form

$$\mathbf{U}_i^{n+1/2} = \mathbf{U}_i - \sum_{j\in K_i}\sigma_{i,j}\mathcal{F}_{i,j} - \sigma_i\mathcal{F}_{e,i} + \sum_{j\in K_i}\sigma_{i,j}\mathcal{S}_p(\mathbf{U}_i,\mathbf{U}_j,z_{b,i},z_{b,j}), \tag{53}$$ eq:upU0

where using the notations of (34)

$$\sum_{j\in K_i} L_{i,j}\mathcal{F}_{i,j} \approx \int_{C_i}\nabla_{x,y}\cdot F(\mathbf{U})dxdy, \tag{54}$$ eq:flux_dis1

with

$$\sigma_{i,j} = \frac{\Delta t^n L_{i,j}}{|C_i|}, \quad \sigma_i = \frac{\Delta t^n L_i}{|C_i|}.$$

Here we consider first-order explicit schemes where

$$\mathcal{F}_{i,j} = \begin{pmatrix} F^h_{i,j} \\ F^{\rho_1 h_1}_{i,j} \\ \vdots \\ F^{\rho_N h_N}_{i,j} \\ F^{\mathbf{u}_1}_{i,j} \\ \vdots \\ F^{\mathbf{u}_N}_{i,j} \end{pmatrix}. \tag{55}$$ eq:flux_def

and

$$F^h_{i,j} = \sum_{\alpha=1}^N F^{h_\alpha}_{i,j}, \tag{56}$$ eq:flux

and for the boundary nodes

$$\mathcal{F}_{i,e} = \begin{pmatrix} F^h_{i,e} \\ F^{\rho_1 h_1}_{i,e} \\ \vdots \\ F^{\rho_N h_N}_{i,e} \\ F^{\mathbf{u}_1}_{i,e} \\ \vdots \\ F^{\mathbf{u}_N}_{i,e} \end{pmatrix}. \tag{57}$$ eq:fluxbis

The fluxes $F_{i,j}^{h_\alpha}$, $F_{i,j}^{\rho_\alpha h_\alpha}$, $F_{i,j}^{\mathbf{u}_\alpha}$ appearing in expressions (55),(56), (57) are numerical fluxes such that

$$F_{i,j}^{m_\alpha} = F^{m_\alpha}(\mathbf{U}_{\alpha,i}, \mathbf{U}_{\alpha,j}, \mathbf{n}_{i,j}),$$

with $m = h, \rho h, \mathbf{u}$, $\alpha = 1, \ldots, N$.

Relation (53) tells how to compute the values $\mathbf{U}_i^{n+1/2}$ knowing $\mathbf{U}_i$ and discretized values $z_{b,i}$ of the topography. Following (54), the term $\mathcal{F}_{i,j}$ in (53) denotes an interpolation of the normal component of the flux $F(\mathbf{U}).\mathbf{n}_{i,j}$ along the edge $C_{i,j}$. The functions $F(\mathbf{U}_i, \mathbf{U}_j, \mathbf{n}_{i,j}) \in \mathbb{R}^{2N+1}$ are the numerical fluxes, see (Bouchut, 2004).

Until now, the expression for the numerical fluxes is not detailed since any numerical fluxes (Rusanov, HLL,...) can be used (Bouchut, 2004). In paragraph 3.5 we define $\mathcal{F}(\mathbf{U}_i, \mathbf{n}_{i,j})$ using kinetic fluxes and we prove a discrete entropy inequality for the system. The computation of the value $\mathbf{U}_{i,e}$, which denotes a value outside $C_i$ (see Fig. 4-*(b)*), defined such that the boundary conditions are satisfied, and the definition of the boundary flux $F(\mathbf{U}_i, \mathbf{U}_{e,i}, \mathbf{n_i})$ are described in (Allgeyer et al., 2019). Notice that we assume a flat topography on the boundaries i.e. $z_{b,i} = z_{b,i,e}$.

For the discretization of the pressure source term $\mathcal{S}_p(\mathbf{U}, z_b)$, we adopt a strategy defined below.

### 3.4.2 The hydrostatic reconstruction technique

⟨subsubsec:HR⟩ The hydrostatic reconstruction scheme (HR scheme for short) for the Saint-Venant system has been introduced in (Audusse et al., 2004) in the 1d case and described in 2d for unstructured meshes in (Audusse and Bristeau, 2005). The HR in the context of the kinetic description for the Saint-Venant system has been studied in (Audusse et al., 2016).

In order to take into account the topography variations and to preserve relevant equilibria, the HR leads to a modified version of (53) under the form

$$U_i^{n+1/2} = U_i - \sum_{j \in K_i} \sigma_{i,j} \mathcal{F}_{i,j}^* - \sigma_i \mathcal{F}_{i,e} + \sum_{j \in K_i} \sigma_{i,j} \mathcal{S}_{p,i,j}^*, \tag{58}$$ eq:upU0_HR

where

$$\begin{aligned} \mathcal{F}_{i,j}^* &= F(U_{i,j}^*, U_{j,i}^*, \mathbf{n}_{i,j}), & (59)\ \text{?eq:flux\_HR?} \\ \mathcal{S}_{p,i,j}^* &= S_p(U_i, U_{i,j}^*, z_{b,i}, z_{b,j}, \mathbf{n}_{i,j}) \\ &= \begin{pmatrix} 0 \\ \tilde{p}_{1,i,j}^*(h_{1,i,j} - h_{1,i})\mathbf{n}_{i,j} - g\rho_{1,i}\tilde{h}_{1,i,j}^*(z_{1,i,j} - z_{1,i})\mathbf{n}_{i,j} \\ \vdots \\ \tilde{p}_{\alpha,i,j}^*(h_{\alpha,i,j} - h_{\alpha,i})\mathbf{n}_{i,j} - g\rho_{\alpha,i}\tilde{h}_{\alpha,i,j}^*(z_{\alpha,i,j} - z_{\alpha,i})\mathbf{n}_{i,j} \\ \vdots \end{pmatrix} & (60)\ \text{eq:Sp\_HR} \end{aligned}$$

with

$$
\begin{array}{l}
z^*_{b,i,j} = \max(z_{b,i}, z_{b,j}), \quad h^*_{i,j} = \max(h_i + z_{b,i} - z^*_{b,i,j}, 0), \\
U^*_{i,j} = (h^*_{i,j}, \rho_{1,i} l_1 h^*_{i,j}, \ldots, \rho_{N,i} l_N h^*_{i,j}, \rho_{1,i} l_1 h^*_{i,j} u_{1,i}, \ldots, \rho_{N,i} l_N h^*_{i,j} u_{N,i}, \ldots)^T, \\
z^*_{\alpha,i,j} = z^*_{b,i,j} + \left( \frac{l_\alpha}{2} + \sum_{j=1}^{\alpha-1} l_j \right) h^*_{i,j}, \\
z_{\alpha,i,j} = z_{\alpha,j,i} = \frac{z^*_{\alpha,i,j} + z^*_{\alpha,j,i}}{2}, \\
h_{\alpha,i,j} = h_{\alpha,j,i} = \frac{h^*_{\alpha,i,j} + h^*_{\alpha,j,i}}{2}, \\
\tilde{h}^*_{\alpha,i,j} = \frac{h_{\alpha,i} + h^*_{\alpha,i,j}}{2}, \\
\tilde{p}^*_{\alpha,i,j} = \frac{p_{\alpha,i} + p^*_{\alpha,i,j}}{2},
\end{array}
\tag{61}
$$

eq:state_HR

and

$$
z_\alpha = z_b + \left( \frac{l_\alpha}{2} + \sum_{j=1}^{\alpha-1} l_j \right) h, \qquad p_\alpha = \frac{\rho_\alpha g h_\alpha}{2} + \sum_{j=\alpha+1}^{N} \rho_j g h_j.
$$

Throughout this work, the $^*$ refers to the HR technique.

**Remark 3.2** *Since the quantity $\mathcal{S}(\mathbf{U}, z_b)$ appearing in* (33) *contains non conservative terms, its integration over the cell $C_i$ is not straightforward and we have used the result proposed by Bouchut (Bouchut, 2004, Proposition 5.3) to obtain the expression* (60)*.*

### 3.4.3 The vertical exchange terms

sec:exchanges

We give the fully discrete expression of the step for the vertical exchanges, described by equations (39)-(40). The step for the vertical exchanges consists in

$$
U_i^{n+1} = U_i^{n+1/2} + \Delta t^n \mathcal{G}_i^{n+1}, \tag{62}
$$

eq:upU1_HR

with

$$
\mathcal{G}_i^{n+1} = \begin{pmatrix}
0 \\
\rho^{n+1}_{3/2,i} G_{3/2,i} \\
\rho^{n+1}_{5/2,i} G_{5/2,i} - \rho^{n+1}_{3/2,i} G_{3/2,i} \\
\vdots \\
\rho^{n+1}_{N-1/2,i} G_{N-1/2,i} - \rho^{n+1}_{N-3/2,i} G_{N-3/2,i} \\
-\rho^{n+1}_{N-1/2,i} G_{N-1/2,i} \\
u^{n+1}_{3/2,i} \rho^{n+1}_{3/2,i} G_{3/2,i} \\
u^{n+1}_{5/2,i} \rho^{n+1}_{5/2,i} G_{5/2,i} - u^{n+1}_{3/2,i} \rho^{n+1}_{3/2,i} G_{3/2,i} \\
\vdots \\
u^{n+1}_{N-1/2,i} \rho^{n+1}_{N-1/2,i} G_{N-1/2,i} - u^{n+1}_{N-3/2,i} \rho^{n+1}_{N-3/2,i} G_{N-3/2,i} \\
-u^{n+1}_{N-1/2,i} \rho^{n+1}_{N-1/2,i} G_{N-1/2,i}
\end{pmatrix}.
$$

The mass conservation for the layer $\alpha$ is governed by Eq. (31) and its discretization gives

$$
h^{n+1}_{\alpha,i} = h^{n+1/2}_{\alpha,i} + \Delta t (G_{\alpha+1/2,i} - G_{\alpha-1/2,i}).
$$

The previous equation multiplied by $\rho_{\alpha,i}^{n+1}$ from equation (39) gives the relation

$$\rho_{\alpha,i}^{n+1}h_{\alpha,i}^{n+1/2} = \rho_{\alpha,i}^{n+1/2}h_{\alpha,i}^{n+1/2} + \Delta t((\rho_{\alpha+1/2,i}^{n+1} - \rho_{\alpha,i}^{n+1})G_{\alpha+1/2,i} - (\rho_{\alpha-1/2,i}^{n+1} - \rho_{\alpha,i}^{n+1})G_{\alpha-1/2,i}).$$

Hence, relation (62) can be rewritten under the form

$$(H_{N,i}^{n+1/2} + \Delta t G_{N,i})\rho_i^{n+1} = (\rho_i h_i)^{n+1/2}, \tag{63}$$ eq:matG

where $H_{N,i}^{n+1/2}$ is a diagonal matrix of size $N$ with coefficients $(H_{N,i_j})_{1\leq j\leq N} = h_j^{n+1/2}$ and the matrix $G_{N,i}$ is given by

$$G_{N,i} = \begin{pmatrix} \frac{|G_{3/2,i}|_+}{h_{1,i}^{n+1/2}} & -\frac{|G_{3/2,i}|_+}{h_{1,i}^{n+1/2}} & 0 & 0 & \dots & 0 \\ \frac{|G_{3/2,i}|_-}{h_{1,i}^{n+1/2}} & \ddots & \ddots & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & 0 \\ \vdots & 0 & \frac{|G_{\alpha-1/2,i}|_-}{h_{\alpha,i}^{n+1/2}} & \frac{|G_{\alpha+1/2,i}|_+}{h_{\alpha,i}^{n+1/2}} - \frac{|G_{\alpha-1/2,i}|_-}{h_{\alpha,i}^{n+1/2}} & -\frac{|G_{\alpha+1/2,i}|_+}{h_{\alpha,i}^{n+1/2}} & 0 \\ \vdots & \ddots & 0 & \ddots & \ddots & -\frac{|G_{N-1/2,i}|_+}{h_{N,i}^{n+1/2}} \\ 0 & \dots & 0 & 0 & \frac{|G_{N-1/2,i}|_-}{h_{N,i}^{n+1/2}} & -\frac{|G_{N-1/2,i}|_-}{h_{N,i}^{n+1/2}} \end{pmatrix}$$

If $h_{\alpha,i}^{n+1/2} = 0$ for all $\alpha$, then we trivially have $\rho_{\alpha,i}^{n+1} = 0$ for all $\alpha$. Let us now assume that there exists a layer $\alpha$ such that $h_{\alpha,i}^{n+1} > 0$. Then the matrix $H_{N,i}^{n+1/2} + \Delta t G_{N,i}$ is a strictly diagonally dominant matrix. Therefore, it is invertible and the entries of its inverse are all nonnegative - see the proof made in Audusse et al. (2018).

This system and its numerical resolution are studied afterwards, in section 3.4.5.

### 3.4.4 The variable update

Hence the sum of relations (58) and (62) gives

$$U_i^{n+1} = U_i - \dots, \tag{64}$$ eq:upU_HR

The space discretization of the system (38)-(40) and its numerical resolution allow to determine the quantities

$$h_i^{n+1}, \rho_{\alpha,i}^{n+1}, (\rho_\alpha h_\alpha \mathbf{u}_\alpha)_i^{n+1},$$

for any $i \in I$ knowing the quantities $\{h_j, (\rho_\alpha h_\alpha)_j, (\rho_\alpha h_\alpha \mathbf{u}_\alpha)_j\}_{j\in I}$.

Thus, it is possible to recover $h_{\alpha,i}^{n+1}$, $\rho_{\alpha,i}^{n+1}$ and $\mathbf{u}_{\alpha,i}^{n+1}$ from

$$h_{\alpha,i}^{n+1} = l_\alpha h_i^{n+1}, \qquad \rho_{\alpha,i}^{n+1} = \frac{(\rho_\alpha h_\alpha)_i^{n+1}}{l_\alpha h_i^{n+1}}, \qquad \mathbf{u}_{\alpha,i}^{n+1} = \frac{(\rho_\alpha h_\alpha \mathbf{u}_\alpha)_i^{n+1}}{(\rho_\alpha h_\alpha)_i^{n+1}}.$$

Concerning the temperature $T_{\alpha,i}^{n+1}$, we use the formula

$$T_{\alpha,i}^{n+1} = \rho^{-1}(\rho_{\alpha,i}^{n+1}).$$

### 3.4.5 Properties of the numerical scheme

ppopumumschemme⟩ **Proposition 3.3** *Consider a consistent numerical flux $\mathcal{F}$ for the homogeneous problem ( (35)-(38) i.e. Eq. (43) with $\mathcal{S}_p(\mathbf{U}, z_b) = 0$) that preserves the non-negativity of the water depth $h_i$ under the corresponding CFL condition, then the finite volume scheme (58)-(62)*

*(i) preserves the non-negativity of the water depth,*

*(ii) preserves the steady state of a lake at rest,*

*(iii) is consistent with the system (28)-(30),(31).*

n_properties⟩?

**Proof** *(i) For $\alpha = 1, \ldots, N$, $h_\alpha^{n+1/2}$ is obtained using the fully discrete version of (35), that is to say,*

$$h_{\alpha,i}^{n+1/2} = h_{\alpha,i} - \sum_{j \in K_i} \sigma_{i,j} \mathcal{F}_{i,j}^{h_\alpha}$$

*Since the flux $\mathcal{F}$ preserves the positivity for the shallow water equations, then $h_{\alpha,i}^{n+1/2}$ is positive. As a sum of positive terms, $h^{n+1}$ is positive. This proves (i).*

*(ii) In the constant density case and with a non-flat topography, (ii) is proved in (Audusse et al., 2011). In the variable density case with flat topography, the proof is trivial because at equilibrium, the density is constant in each of the layers. In the general case, the continuous equations describing the static equilibrium are*

$$\begin{aligned} u_\alpha &= 0, \quad \alpha = 1, \ldots, N, \\ \nabla_{x,y}\eta &= 0, \\ \nabla_{x,y}(h_\alpha p_\alpha) &= p_{\alpha+1/2}\nabla_{x,y} z_{\alpha+1/2} - p_{\alpha-1/2}\nabla_{x,y} z_{\alpha-1/2}, \quad \alpha = 1, \ldots, N. \end{aligned} \tag{65}$$

eq:static_eq_p

*While the first two equations are easy to write at the discrete level, the third one is not. There is no simple discrete formulation of (65). Yet, for*

$$\mathbf{u}_{\alpha,i} = 0, \quad \forall \alpha, \forall i, \qquad \text{and} \qquad \eta_i = \eta_{eq}, \quad \forall i \tag{66}$$

eq:static_equi

*with $\eta_{eq}$ a constant, due to the use of the HR technique, the proposed numerical scheme verifies*

$$\begin{aligned} h_{\alpha,i}^{n+1/2} &= h_{\alpha,i}, \\ (\rho h)_{\alpha,i}^{n+1/2} &= (\rho h)_{\alpha,i}, \\ (\rho h)_{\alpha,i}^{n+1} &= (\rho h)_{\alpha,i}^{n+1/2}, \\ (\rho h \mathbf{u})_{\alpha,i}^{n+1} &= (\rho h \mathbf{u})_{\alpha,i}^{n+1/2}, \end{aligned}$$

*for all $\alpha$ and all $i$. Therefore, starting from a situation described by (66) and a discrete version of (65), the discrete equilibrium is preserved. Starting from a situation near the equilibrium, the system will evolve towards equilibrium. We have empirically checked*

*that a system initially far from equilibrium will reach an equilibrium, see (Audusse et al., 2011).*

*(iii) Since the flux $\mathcal{F}$ is consistent with the homogeneous ($\mathcal{S}_p(U, z_b) = 0$), Eq. (60) is a consistent discretization of the non-conservative pressure terms, and (62) is a consistent discretization of the vertical exchange terms, then (iii) is true.*

■

**Proposition 3.4** *Consider a consistent numerical flux $\mathcal{F}$ for the homogeneous problem (35)-(38) i.e. Eq. (43) with $\mathcal{S}_p(\mathbf{U}, z_b) = 0$ that preserves nonnegativity of $h_i(t)$ and such that*

$$F_{i,j}^{\rho_\alpha h_\alpha} = \rho_{\alpha,i,j} F_{i,j}^{h_\alpha}, \tag{67}$$ eq:flux_less_c

*with*

$$\rho_{\alpha,i,j} = \left\{ \begin{array}{ll} \rho_{\alpha,i} & \text{if } F_{i,j}^{h_\alpha} \geq 0 \\ \rho_{\alpha,j} & \text{if } F_{i,j}^{h_\alpha} \leq 0 \end{array} \right\} \tag{68}$$ eq:upwind_rho

*then the numerical scheme (58)-(62) satisfies a maximum principle on the density i.e.* prop:max_prin *for any $\alpha, i$ one has*

$$\rho_{\alpha,i}^{n+1} \leq \max_\alpha \{\rho_{\alpha,i}, \rho_{\alpha,j}\}, \qquad \forall j \in K_i.$$

**Remark 3.5** *The formula (67) was initially proposed in (Larrouturou, 1991) see also (Bouchut, 2004).*

**Proof of prop 3.4** *Let us first deal with the horizontal exchanges. Due to the choice of $F_{i,j}^{\rho_\alpha h_\alpha}$, the numerical discretization of (36) can be decomposed into*

$$(\rho_\alpha h_\alpha)_i^{n+1/2} = \rho_{\alpha,i} \left( h_{\alpha,i}^n - \sum_{j \in K_i} \sigma_{i,j} |F_{i,j}^{h_\alpha}|_+ \right) - \sum_{j \in K_i} \sigma_{i,j} \rho_{\alpha,j} |F_{i,j}^{h_\alpha}|_-$$

*The right hand side of this expression is positive. Indeed, $h_{\alpha,i}^n - \sum_{j \in K_i} \sigma_{i,j} |F_{i,j}^{h_\alpha}|_+$ is positive due to the CFL condition and $-\sum_{j \in K_i} \sigma_{i,j} \rho_{\alpha,j} |F_{i,j}^{h_\alpha}|_-$ is positive because $|F_{i,j}^{h_\alpha}|_-$ is negative. The right hand side is factorized by $\max\{\rho_{\alpha,i}, \rho_{\alpha,j}\}$*

$$(\rho_\alpha h_\alpha)_i^{n+1/2} \leq \max\{\rho_{\alpha,i}, \rho_{\alpha,j}\} \left( h_{\alpha,i}^n - \sum_{j \in K_i} \sigma_{i,j} |F_{i,j}^{h_\alpha}|_+ - \sum_{j \in K_i} \sigma_{i,j} |F_{i,j}^{h_\alpha}|_- \right).$$

*Therefore, dividing by $h_{\alpha,i}^{n+1/2}$ positive, we obtain that*

$$\rho_{\alpha,i}^{n+1/2} \leq \max\{\rho_{\alpha,i}, \rho_{\alpha,j}\}, \qquad \forall j \in K_i. \tag{69}$$ eq:maxpple_hor

*We deal next with the vertical exchanges, see paragraph 3.4.3.*

*Note that the matrix $H_{N,i}^{n+1/2} + \Delta t G_{N,i}$ defined by (63) is an $M$-matrix. Let $\mathbf{1}$ be the vector the entries of which are all equal to 1. We notice that*

$$(H_{N,i}^{n+1/2} + \Delta t G_{N,i}) \mathbf{1} = h_i^{n+1/2},$$

*so we also have* $\mathbf{1} = (H_{N,i}^{n+1/2} + \Delta t G_{N,i})^{-1} h_i^{n+1/2}$. *Let* $(\rho_i h_i)^{n+1/2}$ *be the vector the entries of which are* $(\rho_i h_i)_\alpha^{n+1/2} = \rho_{\alpha,i}^{n+1/2} h_{\alpha,i}^{n+1/2}$, $\alpha = 1, \dots, N$. *Then*

$$||(H_{N,i}^{n+1/2} + \Delta t G_{N,i})^{-1} (\rho_i h_i)^{n+1/2}||_\infty \leq ||\rho_i^{n+1/2}||_\infty ||(H_{N,i}^{n+1/2} + \Delta t G_{N,i})^{-1} h_i^{n+1/2}||_\infty,$$

*which is exactly*

$$||\rho_i^{n+1}||_\infty \leq ||\rho_i^{n+1/2}||_\infty. \tag{70}$$ maxpple_vertic

*To conclude, relationship* (69) *is applied to* $\rho_{\alpha,i}^{n+1/2}$ *for all* $\alpha$. *Combining with* (70) *gives the maximum principle on the density.* ■

mark:max_prin⟩ **Remark 3.6** *Let us present a more accurate result for the maximum principle on the vertical exchanges. Let* $\alpha, \alpha_0$ *and* $\alpha_1$ *such that* $1 \leq \alpha_0 < \alpha < \alpha_1 \leq N$ *and*

$$G_{\alpha_j+1/2,i} < 0, \quad G_{\alpha_j-1/2,i} > 0 \qquad \text{for } j \in \{0, 1\}.$$

*The coefficients of the lines* $\alpha_0$ *and* $\alpha_1$ *of matrix* $G_{N,i}$ *are respectively* $(G_{N,i})_{\alpha_0,j} = \delta_{\alpha_0,j}$ *and* $(G_{N,i})_{\alpha_1,j} = \delta_{\alpha_1,j}$ *with* $\delta_{k,l}$ *the Kronecker symbol, which means that* $\rho_{\alpha_0,i}^{n+1} = \rho_{\alpha_0,i}^{n+1/2}$ *and* $\rho_{\alpha_1,i}^{n+1} = \rho_{\alpha_1,i}^{n+1/2}$. *The system can be solved using forward elimination and backward substitution. Denoting by* $\rho_{k-l,i}^{n+1/2}$ *the vector* $(\rho_{k,i}^{n+1/2}, \rho_{k+1,i}^{n+1/2}, \dots, \rho_{l,i}^{n+1/2})^T$, *we have the following results*

$$||\rho_{1-\alpha_0,i}^{n+1}||_\infty \leq ||\rho_{1-\alpha_0,i}^{n+1/2}||_\infty, \quad ||\rho_{\alpha_0-\alpha_1,i}^{n+1}||_\infty \leq ||\rho_{\alpha_0-\alpha_1,i}^{n+1/2}||_\infty, \quad ||\rho_{\alpha_0-N,i}^{n+1}||_\infty \leq ||\rho_{\alpha_0-N,i}^{n+1/2}||_\infty.$$

*The layers receiving no mass from the layers above and below them separate groups of layers which exchange mass between themselves.*

**Remark 3.7** *For the following semi-implicit scheme for the vertical exchanges, the maximum principle on the density of proposition 3.4 is also verified:*

$$(\rho_{\alpha,i} h_{\alpha,i})^{n+1} = (\rho_{\alpha,i} h_{\alpha,i})^{n+1/2} + \frac{\Delta t}{2}(\rho_{\alpha+1/2,i}^{n+1} G_{\alpha+1/2,i} - \rho_{\alpha-1/2,i}^{n+1} G_{\alpha-1/2,i}) + \frac{\Delta t}{2}(\rho_{\alpha+1/2,i}^{n+1/2} G_{\alpha+1/2,i} - \rho_{\alpha-1/2,i}^{n+1/2} G_{\alpha-1/2,i}).$$

*The more accurate maximum principle presented in remark 3.6 is verified as well. However, in the rest of the paper, we work only with the fully implicit scheme for the vertical exchanges. More specifically, proposition 3.10 is stated only for the fully implicit scheme.*

**Remark 3.8** *We have presented a first order in space discretization of the system. In practice, we apply a formally second order extension in space and time presented in (Audusse and Bristeau, 2005) and (Allgeyer et al., 2019). More specifically, the modified Heun scheme which is used is presented in (Allgeyer et al., 2019). For the obtained numerical scheme we are able to prove the consistence, the well-balancing and the non-negativity of the water depth. But a discrete entropy inequality such as the one in proposition 3.10 has not yet been obtained. The proof of proposition 3.10 cannot be adapted for the second order scheme; a different strategy would be needed.*

**Remark 3.9** *At each time step, to advance*

- *from* $h_i^n$ *to* $h_i^{n+1/2}$
- *from* $(\rho_{\alpha,i}h_{\alpha,i})^n$ *to* $(\rho_{\alpha,i}h_{\alpha,i})^{n+1/2}$

*convex combinations are used, which gives the scheme a certain stability. Then, to advance from* $(\rho_{\alpha,i}h_{\alpha,i})^{n+1/2}$ *to* $(\rho_{\alpha,i}h_{\alpha,i})^{n+1}$*, the fact that the matrix of the system (matrix* $H_{N,i}^{n+1/2} + \Delta t G_{N,i}$*, defined in the proof of proposition 3.4) is an* $M$*-matrix gives stability to the computation.*

## 3.5 Kinetic fluxes

ec:kin_fluxes⟩ Whereas the proposed numerical scheme can be adapted to any finite volume solver for the classical Saint-Venant system, in Section 5, the numerical simulations are performed using a kinetic solver and hence the numerical fluxes in (55) in the kinetic context now are specified.

To define the numerical fluxes, we introduce the functions $\chi_0$, $M_\alpha$

$$\chi_0(z_1, z_2) = \frac{1}{4\pi}\mathbb{1}_{z_1^2+z_2^2\leq 4},$$
$$M_\alpha = M(U_\alpha, \xi, \gamma) = \frac{h_\alpha}{c_\alpha^2}\chi_0\left(\frac{\xi - u_\alpha}{c_\alpha}, \frac{\gamma - v_\alpha}{c_\alpha}\right),$$

with $c_\alpha = \sqrt{p_\alpha/\rho_\alpha}$, $U_\alpha$ defined by (52) and where $(\xi, \gamma) \in \mathbb{R}^2$. We also define the quantity $M_\alpha^\rho$ by

$$M_\alpha^\rho = \rho_\alpha M_\alpha.$$

The quantity $M_\alpha^\rho$ satisfies the following moment relations

$$\int_{\mathbb{R}^2} \begin{pmatrix} 1 \\ \xi \end{pmatrix} \rho_\alpha M(U_\alpha, \xi, \gamma)\, d\xi d\gamma = \begin{pmatrix} \rho_\alpha h_\alpha \\ \rho_\alpha h_\alpha u_\alpha \\ \rho_\alpha h_\alpha v_\alpha \end{pmatrix}, \tag{71}$$ eq:kinmom1

$$\int_{\mathbb{R}^2} \begin{pmatrix} \xi^2 \\ \xi\gamma \\ \gamma^2 \end{pmatrix} \rho_\alpha M(U_\alpha, \xi, \gamma)\, d\xi d\gamma = \begin{pmatrix} \rho_\alpha h_\alpha u_\alpha^2 + h_\alpha p_\alpha \\ \rho_\alpha h_\alpha u_\alpha v_\alpha \\ \rho_\alpha h_\alpha v_\alpha^2 + h_\alpha p_\alpha \end{pmatrix}. \tag{72}$$ ?eq:kinmom2?

Hence in the context of the kinetic description, the fluxes appearing in (55) have the expressions

$$F_{i,j}^{h_\alpha} = \int_{\mathbb{R}^2} M_{\alpha,i,j}\zeta_{i,j} d\xi d\gamma, \quad F_{i,j}^{\rho_\alpha h_\alpha} = \rho_{\alpha,i,j} F_{i,j}^{h_\alpha}, \quad F_{i,j}^{\mathbf{u}_\alpha} = \rho_{\alpha,i,j} \int_{\mathbb{R}^2} \begin{pmatrix} \xi \\ \gamma \end{pmatrix} M_{\alpha,i,j}\zeta_{i,j} d\xi d\gamma \tag{73}$$ eq:kin_fluxes

with

$$M_{\alpha,i,j} = M_{\alpha,i,j}^* \mathbb{1}_{\zeta_{i,j}\geq 0} + M_{\alpha,j,i}^* \mathbb{1}_{\zeta_{i,j}\leq 0}, \tag{74}$$ eq:kin_dis22

where $M^*_{\alpha,i,j} = M(U^*_{\alpha,i,j}, \xi, \gamma)$, $U^*_{\alpha,i,j} = (l_\alpha h^*_{i,j}, l_\alpha h^*_{i,j} u_{\alpha,i}, l_\alpha h^*_{i,j} v_{\alpha,i})^T$. The $^*$ refers to the HR technique, see (61). The density $\rho_{\alpha,i,j}$ is defined by (68) and

$$\zeta_{i,j} = \begin{pmatrix} \xi \\ \gamma \end{pmatrix} . \mathbf{n}_{i,j}.$$

We give here some details about Proposition 3.3 in the case of the kinetic flux. We consider the equation

$$f^{n+1/2-}_{\alpha,i} = \quad M_{\alpha,i} - \sum_{j \in K_i} \sigma_{i,j} \zeta_{i,j} \mathbb{1}_{\zeta_{i,j} \geq 0} M^*_{\alpha,i,j} - \sum_{j \in K_i} \sigma_{i,j} M^*_{\alpha,j,i} \zeta_{i,j} \mathbb{1}_{\zeta_{i,j} \leq 0} \tag{75}$$

eq:kin_horizon

Using (71), we see that integrating equation (75) for $\alpha = 1, \ldots, N$ with respect to $\xi, \gamma$ gives the HR scheme for (35).

Let us now give the CFL condition for the kinetic flux. There exists a velocity $v_m \geq 0$ such that for all $\alpha, i$

$$\left| \frac{\xi - u_{\alpha,i}}{c_{\alpha,i}} \right| \geq v_m \text{ or } \left| \frac{\gamma - v_{\alpha,i}}{c_{\alpha,i}} \right| \geq v_m \Rightarrow M(U_{\alpha,i}, \xi, \gamma) = 0.$$

A CFL condition strictly less than one is considered

$$\tilde{\sigma}_i \big( |u_{\alpha,i}| + |v_{\alpha,i}| + v_m c_{\alpha,i} \big) \leq \beta < \frac{1}{2} \quad \text{for all } i, \alpha,$$

where $\tilde{\sigma}_i = \Delta t^n \Sigma_{j \in K_i} L_{i,j} / |C_i|$, and $\beta$ is a given constant. More precisely, the CFL condition writes

$$\Delta t^n \leq \frac{1}{2} \min_{\alpha,i} \frac{|C_i|}{\Sigma_{j \in K_i} L_{i,j} (|u_{\alpha,i}| + |v_{\alpha,i}| + v_m c_{\alpha,i})}.$$

Under this CFL condition, the kinetic function $f^{n+1/2-}_{\alpha,i}$ remains non-negative, i.e.

$$f^{n+1/2-}_{\alpha,i} \geq 0, \quad \forall (\xi, \gamma) \in \mathbb{R}^2, \forall i, \forall \alpha.$$

The proof can be found in (Allgeyer et al., 2019). Therefore, the water depth $h^{n+1/2}$ is non-negative. Note that the CFL condition does not depend on the vertical exchange terms because they are treated implicitly.

## 3.6 Discrete entropy inequality

In this paragraph, a discrete entropy inequality is proved in the case of a flat topography. The crucial point of the numerical scheme is the treatment of the pressure source term $\mathcal{S}_{p,\alpha}$ written under the form (42). Indeed the other terms appearing in the numerical scheme are either conservative – and hence easily incorporated in the numerical fluxes – or similar to terms appearing in the constant density case, see (Allgeyer et al., 2019). The term $\mathcal{S}_{p,\alpha}$ is an extension of the topography term for the Saint-Venant system but in a far more complex setting. Proposition 3.10 is interesting since until now, the properties

satisfied by the numerical scheme detailed in paragraph 3.4.5 concern Eqs. (35)-(41) except Eq. (37). Hence, for the momentum equation (37), only the equilibrium at rest is proved.

In the context of the kinetic description, the relation between the mass and momentum fluxes is simple (see (73)) and it is possible to slightly modify the definitions (60) in order to obtain an in cell discrete entropy inequality. The authors do not claim that only the kinetic fluxes allow to obtain such a result but, as in (Audusse et al., 2016), it is not clear whether the result holds with other numerical fluxes in particular the definition (77) is related to the kinetic description and not easily available for others finite volume solvers (Rusanov, HLL,...).

We consider in this paragraph a discrete form of (42) that is slightly different from (60) and defined by

$$\mathcal{S}_{p,\alpha,i,j} = p_{\alpha,i}(\hat{h}_{\alpha,i,j} - h_{\alpha,i})\mathbf{n}_{i,j} - g\rho_{\alpha,i}h_{\alpha,i}(z_{\alpha,i,j} - z_{\alpha,i})\mathbf{n}_{i,j}, \tag{76}$$

eq:Sp_HR_mod

with

$$\begin{aligned}
z_{\alpha+1/2,i} &= z_{b,i} + \sum_{l=1}^{\alpha} h_{l,i}, \quad \text{with } z_{b,i} = z_{b,j} = cst,\ \forall j,\\
z_{\alpha,i} &= \frac{z_{\alpha+1/2,i} + z_{\alpha-1/2,i}}{2},\\
z_{\alpha,i,j} &= \frac{z_{\alpha,i} + z_{\alpha,j}}{2},\\
\hat{h}_{\alpha,i,j} &= \int_{\mathbb{R}^2} (M_{\alpha,i}\mathbb{1}_{\zeta_{i,j}\leq 0} + M_{\alpha,j}\mathbb{1}_{\zeta_{i,j}\geq 0})d\xi d\gamma, = \int_{\mathbb{R}^2} M_{\alpha,i,j}d\xi d\gamma \qquad (77)\\
p_{\alpha,i} &= \frac{\rho_{\alpha,i}}{2}\frac{\int_{\mathbb{R}^2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix} - \mathbf{u}_{\alpha,i}\right|^2 M_{\alpha,i}d\xi d\gamma}{\int_{\mathbb{R}^2} M_{\alpha,i}d\xi d\gamma}. \qquad (78)
\end{aligned}$$

eq:h_alpha_i_j

eq:state_HR_mo

Since we consider a flat topography, the notations associated with the HR do not appear in the previous definitions. The discretization (76) is very similar to the discretization (60) written for a flat topography, but a bit of upwinding with respect to the advection is included in (76). Note that this kind of upwinding could help to reduce the error term due to the topography in (Audusse et al., 2016). For a non-flat topography, the HR induces upwinding with respect to the topography, not with respect to the advection.

The main interest of the following proposition is to justify the discretization (76) but for three reasons it is a partial result:

- it only concerns flat topography situations whereas it is well known that the numerical treatment of topography source terms is a very difficult issue,
- the extension of the expression (76) to the situation of a non flat topography is not natural since in the simple case of a single layer with a constant density i.e. the classical Saint-Venant system, the definition (76) does not exactly match with previous works of some of the authors (Audusse et al., 2016) (whereas definition (60)

exactly reduces to the scheme studied in (Audusse et al., 2016) in the Saint-Venant case),

- the numerical tests carried out with the two possible discretizations of $\mathcal{S}_{p,\alpha,i,j}$, namely (60) and (76) give very similar results especially similar convergence curves, see paragraph 5.1.

For these reasons and even if the result proposed in the following proposition is an interesting stability property, the numerical simulations presented in Section 4 have been obtained using the discretization (60).

crete_entropy⟩ **Proposition 3.10** *When considering a flat bottom, the scheme* (58),(62) *with the fluxes defined by* (73) *satisfies an in cell fully discrete entropy inequality having the form*

$$
\begin{aligned}
&E_{\alpha,i}^{n+1} - E_{\alpha,i} + \sum_{j\in K_i} \sigma_{i,j} \int_{\mathbb{R}^2} \left( g z_{\alpha,i,j} + \frac{\xi^2+\gamma^2}{2} - \frac{1}{2} \left| \begin{pmatrix} \xi \\ \gamma \end{pmatrix} - \tilde{\mathbf{u}}_{\alpha,i,j} \right|^2 \right) \rho_{\alpha,i,j} M_{\alpha,i,j} \zeta_{i,j} d\xi d\gamma \\
&-\Delta t^n \rho_{\alpha+1/2,i}^{n+1} \left( \frac{|\mathbf{u}_{\alpha+1/2,i}^{n+1}|^2}{2} + g z_{\alpha+1/2,i} \right) G_{\alpha+1/2,i} + \Delta t^n \left( \rho_{\alpha-1/2,i}^{n+1} \frac{|\mathbf{u}_{\alpha-1/2,i}^{n+1}|^2}{2} + g z_{\alpha-1/2,i} \right) G_{\alpha-1/2,i} \\
&- p_{\alpha+1/2,i}^{n*} \big( \Delta t^n G_{\alpha+1/2,i} - z_{\alpha+1/2,i}^{n+1} + z_{\alpha+1/2,i} \big) + p_{\alpha-1/2,i}^{n*} \big( \Delta t^n G_{\alpha-1/2,i} - z_{\alpha-1/2,i}^{n+1} + z_{\alpha-1/2,i} \big) \\
&\qquad = \; d_{\alpha,i} + e_{\alpha,i} + f_{\alpha,i}
\end{aligned}
$$

*where $E_{\alpha,i}$ is the discrete energy*

$$
E_{\alpha,i} = \rho_{\alpha,i} h_{\alpha,i} \frac{|\mathbf{u}_{\alpha,i}|^2}{2} + \rho_{\alpha,i} g h_{\alpha,i} z_{\alpha,i},
$$

*and where $d_{\alpha,i}$ is a sum of non-positive and hence dissipative terms whereas $e_{\alpha,i}$ (resp. $f_{\alpha,i}$) contains errors terms of magnitude $\mathcal{O}(diam(C_i)^3)$ (resp. $\mathcal{O}(\Delta t^n)^2$). The expressions of $d_{\alpha,i}$ and $e_{\alpha,i}$ are given by*

$$
\begin{aligned}
d_{\alpha,i} \;=& -\sum_{j\in K_i} \sigma_{i,j} \frac{|\tilde{\mathbf{u}}_{\alpha,i,j} - \mathbf{u}_{\alpha,i}|^2}{2} \int_{\mathbb{R}^2} \rho_{\alpha,i} M_{\alpha,i} |\zeta_{i,j}| d\xi d\gamma \\
&+ \sum_{j\in K_i} \sigma_{i,j} \frac{|\mathbf{u}_{\alpha,i} - \mathbf{u}_{\alpha,j}|^2}{4} \int_{\{\zeta_{i,j}\le 0\}} \rho_{\alpha,j} M_{\alpha,j} \zeta_{i,j} d\xi d\gamma \qquad (79)
\end{aligned}
$$

def

$$
\begin{aligned}
e_{\alpha,i} \;=& \sum_{j\in K_i} \sigma_{i,j} g \int_{\mathbb{R}^2} \left( \rho_{\alpha,i} h_{\alpha,i} \mathbf{u}_{\alpha,i} - \rho_{\alpha,i,j} M_{\alpha,i,j} \begin{pmatrix} \xi \\ \gamma \end{pmatrix} \right) (z_{\alpha,i,j} - z_{\alpha,i}) \cdot \mathbf{n}_{i,j} d\xi d\gamma \\
&- \sum_{j\in K_i} \sigma_{i,j} p_{\alpha,i} \int_{\mathbb{R}^2} (M_{\alpha,i,j} - M_{\alpha,i}) (\tilde{\mathbf{u}}_{\alpha,i,j} - \mathbf{u}_{\alpha,i}) \cdot \mathbf{n}_{i,j} d\xi d\gamma \\
&- \sum_{j\in K_i} \sigma_{i,j} (\tilde{\mathbf{u}}_{\alpha,i,j} - \mathbf{u}_{\alpha,i}) \cdot \int_{\{\zeta_{i,j}\le 0\}} \left[ \left( \begin{pmatrix} \xi \\ \gamma \end{pmatrix} - \mathbf{u}_{\alpha,i} \right) \rho_{\alpha,i} M_{\alpha,i} - \left( \begin{pmatrix} \xi \\ \gamma \end{pmatrix} - \mathbf{u}_{\alpha,j} \right) \rho_{\alpha,j} M_{\alpha,j} \right] \zeta_{i,j} d\xi d\gamma
\end{aligned}
$$

$$-\sum_{j\in K_i}\sigma_{i,j}\frac{|2\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}-\mathbf{u}_{\alpha,j}|^2}{4}\int_{\{\zeta_{i,j}\leq 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma. \tag{80}$$ def

*The quantity $f_{\alpha,i}$ is defined by*

$$\begin{aligned}
f_{\alpha,i} &= \Delta t^n\frac{gh_{\alpha,i}}{2}\left((\rho_{\alpha,i}^{n+1}-\rho_{\alpha+1/2,i}^{n+1})G_{\alpha+1/2,i}+(\rho_{\alpha,i}^{n+1}-\rho_{\alpha-1/2,i}^{n+1})G_{\alpha-1/2,i}\right)\\
&\quad +g\left((\rho_{\alpha,i}h_{\alpha,i})^{n+1}-\rho_{\alpha,i}^{n+1}h_{\alpha,i}\right)(z_{\alpha,i}^{n+1}-z_{\alpha,i})+\frac{(\rho_{\alpha,i}h_{\alpha,i})^{n+1}}{2}|\mathbf{u}_{\alpha,i}^{n+1}-\mathbf{u}_{\alpha,i}|^2\\
&\quad -\Delta t^n\rho_{\alpha+1/2,i}^{n+1}\frac{\mathbf{u}_{\alpha,i}-\mathbf{u}_{\alpha,i}^{n+1}}{2}.(\mathbf{u}_{\alpha,i}^{n+1}-2\mathbf{u}_{\alpha+1/2,i}^{n+1}+\mathbf{u}_{\alpha,i})G_{\alpha+1/2,i}\\
&\quad +\Delta t^n\rho_{\alpha-1/2,i}^{n+1}\frac{\mathbf{u}_{\alpha,i}-\mathbf{u}_{\alpha,i}^{n+1}}{2}.(\mathbf{u}_{\alpha,i}^{n+1}-2\mathbf{u}_{\alpha-1/2,i}^{n+1}+\mathbf{u}_{\alpha,i})G_{\alpha-1/2,i},
\end{aligned}$$

*where the first line is the discrete version of the term appearing in the last line of the continuous energy balance given in Eq.* (26) *(see also (Boittin et al., 2020)) and the other lines come from the explicit time scheme and vanish in the semi-discrete (in space) case. The velocity $\tilde{\mathbf{u}}_{\alpha,i,j}$ is defined by*

$$\tilde{\mathbf{u}}_{\alpha,i,j}=\frac{\displaystyle\int_{\mathbb{R}^2}M_{\alpha,i,j}\begin{pmatrix}\xi\\ \gamma\end{pmatrix}d\xi d\gamma}{\displaystyle\int_{\mathbb{R}^2}M_{\alpha,i,j}d\xi d\gamma}, \tag{81}$$ eq:utilde

*where $M_{\alpha,i,j}$ is defined by* (74) *and the pressure $p_{\alpha\pm1/2,i}^{n*}$ is defined by*

$$p_{\alpha\pm1/2,i}^{n*}=p_{\alpha,i}\mp\rho_{\alpha,i}^{n+1}g\frac{h_{\alpha,i}}{2}.$$

:entropy_ineq⟩

**Remark 3.11** *Since we use an explicit time scheme it is natural to have error terms $f_{\alpha,i}$ of order $\mathcal{O}(\Delta t^n)^2$. Concerning the error terms $e_{\alpha,i}$ due to the space discretization, we point out that they are of order $\mathcal{O}(diam(C_i)^3)$ i.e. smaller than residuals with second terms.*

**Proof of prop. 3.10** *The proof of this proposition is long but only contains simple computations. The authors have not found a simpler presentation.*

*Starting from the set of discrete equations* (64)*, the energy balance for the cell $i$ at the layer $\alpha$ is obtained by performing the sum of the two following quantities:*

- *the mass conservation equation over the layer $\alpha$ in* (64) *multiplied by $gz_{\alpha,i}-|\mathbf{u}_{\alpha,i}|^2/2$*
- *the momentum equation in* (64) *over the layer $\alpha$ multiplied by $\mathbf{u}_{\alpha,i}$.*

*In other words, the energy balance comes from a rewriting of the quantity*

$$\begin{aligned}\mathcal{E}_{\alpha,i} \;:=\; & \big(gz_{\alpha,i}-\frac{|\mathbf{u}_{\alpha,i}|^2}{2}\big)\Big((\rho_{\alpha,i}h_{\alpha,i})^{n+1}-\rho_{\alpha,i}h_{\alpha,i}+\sum_{j\in K_i}\sigma_{i,j}\rho_{\alpha,i,j}F_{i,j}^{h_\alpha}\\ & +\Delta t^n\Big(\rho^{n+1}_{\alpha+1/2,i}G_{\alpha+1/2,i}-\rho^{n+1}_{\alpha-1/2,i}G_{\alpha-1/2,i}\Big)\Big)\\ & +\mathbf{u}_{\alpha,i}.\left((\rho_{\alpha,i}h_{\alpha,i}\mathbf{u}_{\alpha,i})^{n+1}-\rho_{\alpha,i}h_{\alpha,i}\mathbf{u}_{\alpha,i}+\sum_{j\in K_i}\sigma_{i,j}(F_{i,j}^{h_\alpha\mathbf{u}_{\alpha,i}}-\mathcal{S}_{p,\alpha,i,j})\right.\\ & \left.-\Delta t^n\Big(\rho^{n+1}_{\alpha+1/2,i}\mathbf{u}^{n+1}_{\alpha+1/2,i}G_{\alpha+1/2,i}-\rho^{n+1}_{\alpha-1/2,i}\mathbf{u}^{n+1}_{\alpha-1/2,i}G_{\alpha-1/2,i}\Big)\right). \qquad (82)\end{aligned}$$

eq:energ_init

*Since the manipulations necessary to obtain the result are, to some extent tedious, we proceed as follows: first we consider the terms involving time derivatives then those involving the horizontal fluxes and finally, we consider the vertical exchange terms.*

*The discrete time derivatives The terms appearing in* (82) *writes*

$$\mathcal{E}^t_{\alpha,i}:=\big(gz_{\alpha,i}-\frac{|\mathbf{u}_{\alpha,i}|^2}{2}\big)\Big((\rho_{\alpha,i}h_{\alpha,i})^{n+1}-\rho_{\alpha,i}h_{\alpha,i}\Big)+\mathbf{u}_{\alpha,i}.\left((\rho_{\alpha,i}h_{\alpha,i}\mathbf{u}_{\alpha,i})^{n+1}-\rho_{\alpha,i}h_{\alpha,i}\mathbf{u}_{\alpha,i}\right),$$

*and can be rewritten under the form*

$$\begin{aligned}\mathcal{E}^t_{\alpha,i}=(\rho_{\alpha,i}h_{\alpha,i})^{n+1}\frac{|\mathbf{u}^{n+1}_{\alpha,i}|^2}{2}-(\rho_{\alpha,i}h_{\alpha,i})\frac{|\mathbf{u}_{\alpha,i}|^2}{2}+(\rho_{\alpha,i}gh_{\alpha,i}z_{\alpha,i})^{n+1}-(\rho_{\alpha,i}gh_{\alpha,i}z_{\alpha,i})\\ -(\rho_{\alpha,i}gh_{\alpha,i})^{n+1}(z^{n+1}_{\alpha,i}-z_{\alpha,i})-\frac{(\rho_{\alpha,i}h_{\alpha,i})^{n+1}}{2}|\mathbf{u}^{n+1}_{\alpha,i}-\mathbf{u}_{\alpha,i}|^2. \qquad (83)\end{aligned}$$

eq:time_d

*The last term in* (83) *is classical in the context of explicit in time schemes and give rise to a non-negative term in the energy balance.*

*The horizontal fluxes The quantities we are now considering are*

$$\begin{aligned}\mathcal{E}^{xy}_{\alpha,i,j}:=\big(gz_{\alpha,i}-\frac{|\mathbf{u}_{\alpha,i}|^2}{2}\big)\rho_{\alpha,i,j}\int_{\mathbb{R}^2}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma\\ +\mathbf{u}_{\alpha,i}.\left(\rho_{\alpha,i,j}\int_{\mathbb{R}^2}\begin{pmatrix}\xi\\ \gamma\end{pmatrix}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma-\mathcal{S}_{p,\alpha,i,j}\right).\end{aligned}$$

*And using* (76) *we rewrite* $\mathcal{E}^{xy}_{\alpha,i,j}$ *under the form*

$$\begin{aligned}\mathcal{E}^{xy}_{\alpha,i,j} \;:=\; & \int_{\mathbb{R}^2}\left(gz_{\alpha,i,j}+\frac{\xi^2+\gamma^2}{2}-\frac{1}{2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\tilde{\mathbf{u}}_{\alpha,i,j}\right|^2\right)\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma\\ & +\int_{\mathbb{R}^2}\frac{1}{2}\left(\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\tilde{\mathbf{u}}_{\alpha,i,j}\right|^2-\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right|^2\right)\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma\\ & -p_{\alpha,i}(\hat{h}_{\alpha,i,j}-h_{\alpha,i})\mathbf{u}_{\alpha,i}.\mathbf{n}_{i,j}\\ & +g\int_{\mathbb{R}^2}\rho_{\alpha,i,j}M_{\alpha,i,j}\big(z_{\alpha,i,j}-z_{\alpha,i}\big)\big(\mathbf{u}_{\alpha,i}.\mathbf{n}_{i,j}-\zeta_{i,j}\big)d\xi d\gamma \qquad (84)\end{aligned}$$

{?}

$$+g\int_{\mathbb{R}^2}(\rho_{\alpha,i}h_{\alpha,i}-\rho_{\alpha,i,j}(z_{\alpha,i,j}-z_{\alpha,i})\mathbf{u}_{\alpha,i}\cdot\mathbf{n}_{i,j}d\xi d\gamma. \tag{85}$$ eq:E_ij_xy

*The last two lines of the previous equation reduce to*

$$-\sum_{j\in K_i}\sigma_{i,j}p_{\alpha,i}\int_{\mathbb{R}^2}(M_{\alpha,i,j}-M_{\alpha,i})(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i})\cdot\mathbf{n}_{i,j}d\xi d\gamma.$$

*Now, using the definition* (81)*, we rewrite the second line of* (85)*, denoted* $\mathcal{E}^{2,xy}_{\alpha,i,j}$*, under the form*

$$\begin{aligned}
\mathcal{E}^{2,xy}_{\alpha,i,j} &= -\frac{1}{2}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}\int_{\mathbb{R}^2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right|^2\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma\\
&\quad-\int_{\mathbb{R}^2}\frac{1}{2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right|^2\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma\\
&\quad+\int_{\mathbb{R}^2}\frac{1}{2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\tilde{\mathbf{u}}_{\alpha,i,j}\right|^2\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma\\
&\quad+\frac{1}{2}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}\int_{\mathbb{R}^2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right|^2\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma\\
&= -p_{\alpha,i}\int_{\mathbb{R}^2}M_{\alpha,i}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}d\xi d\gamma\\
&\quad-(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\mathbb{R}^2}\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\frac{\tilde{\mathbf{u}}_{\alpha,i,j}+\mathbf{u}_{\alpha,i}}{2}\right)\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma\\
&\quad+(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}\int_{\mathbb{R}^2}\frac{1}{2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right|^2\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma,
\end{aligned}\tag{86}$$ eq:e2_xy_fin

*where the definition* (78) *has been used. Let* $\mathbf{n}_{i,j}=(n_{1,i,j},n_{2,i,j})^T$*, because the Gibbs equilibrium* $M_{\alpha,i}$ *is an even function of the variables* $\xi-u_{\alpha,i}$ *and* $\gamma-v_{\alpha,i}$*, we get*

$$\begin{aligned}
\frac{1}{2}\int_{\mathbb{R}^2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right|^2\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma n_{1,i,j} &= \int_{\mathbb{R}^2}\frac{(\xi-u_{\alpha,i})^2+(\gamma-v_{\alpha,i})^2}{2}\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma n_{1,i,j}\\
&= \int_{\mathbb{R}^2}(\xi-u_{\alpha,i})^2\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma n_{1,i,j}=\int_{\mathbb{R}^2}(\xi-u_{\alpha,i})\xi n_{1,i,j}\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma\\
&= \int_{\mathbb{R}^2}(\xi-u_{\alpha,i})\zeta_{i,j}\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma,
\end{aligned}\tag{87}$$ eq:even_odd_1

*where we have used that since the function* $(\xi-u_{\alpha,i})M_{\alpha,i}$ *is even then*

$$\int_{\mathbb{R}^2}(\xi-u_{\alpha,i})\gamma n_{2,i,j}\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma=0.$$

*Likewise, we obtain*

$$\frac{1}{2}\int_{\mathbb{R}^2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right|^2\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma n_{2,i,j}=\int_{\mathbb{R}^2}(\gamma-v_{\alpha,i})\zeta_{i,j}\rho_{\alpha,i}M_{\alpha,i}d\xi d\gamma. \tag{88}$$ eq:even_odd_2

*And therefore, using Eqs.* (87),(88), *we can rewrite the last line of relation* (86) *under the form*

$$(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\mathbb{R}^2}\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right)\rho_{\alpha,i}M_{\alpha,i}\zeta_{i,j}d\xi d\gamma,$$

*leading to the following expression for* $\mathcal{E}^{2,xy}_{\alpha,i,j}$

$$\begin{aligned}\mathcal{E}^{2,xy}_{\alpha,i,j} &= -p_{\alpha,i}\int_{\mathbb{R}^2}M_{\alpha,i}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}d\xi d\gamma\\ &\quad-(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\mathbb{R}^2}\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\frac{\tilde{\mathbf{u}}_{\alpha,i,j}+\mathbf{u}_{\alpha,i}}{2}\right)\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma\\ &\quad+(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\mathbb{R}^2}\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right)\rho_{\alpha,i}M_{\alpha,i}\zeta_{i,j}d\xi d\gamma.\end{aligned}$$

*We rewrite the second line of* $\mathcal{E}^{2,xy}_{\alpha,i,j}$ *under the form*

$$\frac{|\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}|^2}{2}\int_{\mathbb{R}^2}\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma-(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\mathbb{R}^2}\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right)\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma,$$

*or equivalently*

$$\begin{aligned}\frac{|\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}|^2}{2}\int_{\mathbb{R}^2}\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma-(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\mathbb{R}^2}\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i,j}\right)\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma\\ +(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\mathbb{R}^2}(\mathbf{u}_{\alpha,i}-\mathbf{u}_{\alpha,i,j})\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma.\end{aligned}$$

*Using the previous identities and considering the cases* $\zeta_{i,j}\geq 0$ *and* $\zeta_{i,j}\leq 0$, *simple computations give the following expression for* $\mathcal{E}^{2,xy}_{\alpha,i,j}$

$$\begin{aligned}\mathcal{E}^{2,xy}_{\alpha,i,j} &= -p_{\alpha,i}\int_{\mathbb{R}^2}M_{\alpha,i}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}d\xi d\gamma\\ &\quad+\frac{|\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}|^2}{2}\int_{\{\zeta_{i,j}\geq 0\}}\rho_{\alpha,i}M_{\alpha,i}\zeta_{i,j}d\xi d\gamma\\ &\quad+(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\{\zeta_{i,j}\leq 0\}}\left[\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right)\rho_{\alpha,i}M_{\alpha,i}-\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,j}\right)\rho_{\alpha,j}M_{\alpha,j}\right]\zeta_{i,j}d\xi d\gamma\\ &\quad+\frac{1}{2}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\{\zeta_{i,j}\leq 0\}}\big(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}-2(\mathbf{u}_{\alpha,j}-\mathbf{u}_{\alpha,i})\big)\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma,\end{aligned}$$

*and the last line of the previous expression can be written under the form*

$$\begin{aligned}-\frac{1}{2}|\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}|^2\int_{\{\zeta_{i,j}\leq 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma\\ +(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,j})\int_{\{\zeta_{i,j}\leq 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma,\end{aligned}$$

*or equivalently*

$$
\begin{aligned}
&-\frac{1}{2}|\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}|^2\int_{\{\zeta_{i,j}\leq 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma\\
&-\frac{|\mathbf{u}_{\alpha,i}-\mathbf{u}_{\alpha,j}|^2}{4}\int_{\{\zeta_{i,j}\leq 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma+\frac{|2\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}-\mathbf{u}_{\alpha,j}|^2}{4}\int_{\{\zeta_{i,j}\leq 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma,
\end{aligned}
$$

*where the identity* $ab=(a+b)^2/4-(a-b)^2/4$ *has been used. Hence, the final expression for* $\mathcal{E}^{2,xy}_{\alpha,i,j}$ *is given by*

$$
\begin{aligned}
\mathcal{E}^{2,xy}_{\alpha,i,j} \;=\;& -p_{\alpha,i}\int_{\mathbb{R}^2}M_{\alpha,i}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}d\xi d\gamma\\
&+\frac{|\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}|^2}{2}\int_{\mathbb{R}^2}\rho_{\alpha,i}M_{\alpha,i}|\zeta_{i,j}|d\xi d\gamma-\frac{|\mathbf{u}_{\alpha,i}-\mathbf{u}_{\alpha,j}|^2}{4}\int_{\{\zeta_{i,j}\leq 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma\\
&+(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\int_{\{\zeta_{i,j}\leq 0\}}\left[\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,i}\right)\rho_{\alpha,i}M_{\alpha,i}-\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix}-\mathbf{u}_{\alpha,j}\right)\rho_{\alpha,j}M_{\alpha,j}\right]\zeta_{i,j}d\xi d\gamma\\
&+\frac{|2\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}-\mathbf{u}_{\alpha,j}|^2}{4}\int_{\{\zeta_{i,j}\leq 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma,
\end{aligned}
$$

*where in the previous expression, the second line is nonnegative and the last two lines are third order terms (i.e. of order* $\mathcal{O}(diam(C_i)^3)$*) when considering Lipschitz continuous solutions.*

*Then for the third line of* (85)*, performing simple manipulations we have*

$$
\begin{aligned}
\mathcal{P}^{xy}_{\alpha,i,j} =& -p_{\alpha,i}(\hat{h}_{\alpha,i,j}-h_{\alpha,i})\mathbf{u}_{\alpha,i}.\mathbf{n}_{i,j}\\
=& -p_{\alpha,i}\int_{\mathbb{R}^2}(M_{\alpha,i,j}-M_{\alpha,i})d\xi d\gamma\mathbf{u}_{\alpha,i}.\mathbf{n}_{i,j}\\
=& -p_{\alpha,i}\int_{\mathbb{R}^2}(M_{\alpha,i,j}\tilde{\mathbf{u}}_{\alpha,i,j}.\mathbf{n}_{i,j}-M_{\alpha,i}\mathbf{u}_{\alpha,i}.\mathbf{n}_{i,j})d\xi d\gamma\\
&+p_{\alpha,i}\int_{\mathbb{R}^2}M_{\alpha,i,j}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}d\xi d\gamma\\
=& -p_{\alpha,i}\int_{\mathbb{R}^2}(M_{\alpha,i,j}\zeta_{i,j}-M_{\alpha,i}\mathbf{u}_{\alpha,i}.\mathbf{n}_{i,j})d\xi d\gamma\\
&+p_{\alpha,i}\int_{\mathbb{R}^2}M_{\alpha,i,j}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}d\xi d\gamma\\
=& -p_{\alpha,i}\int_{\mathbb{R}^2}(M_{\alpha,i,j}\zeta_{i,j}-M_{\alpha,i}\mathbf{u}_{\alpha,i}.\mathbf{n}_{i,j})d\xi d\gamma\\
&+p_{\alpha,i}\int_{\mathbb{R}^2}M_{\alpha,i}(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}d\xi d\gamma\\
&+p_{\alpha,i}\int_{\mathbb{R}^2}(M_{\alpha,i,j}-M_{\alpha,i})(\tilde{\mathbf{u}}_{\alpha,i,j}-\mathbf{u}_{\alpha,i}).\mathbf{n}_{i,j}d\xi d\gamma \qquad (89)
\end{aligned}
$$

eq:p1_cons

*where the definitions* (81),(77) *have been used. Notice that for the second term in the first line of* (89)*, we have*

$$\sum_{j\in K_i} p_{\alpha,i}\int_{\mathbb{R}^2} M_{\alpha,i}\mathbf{u}_{\alpha,i}.\mathbf{n}_{i,j}d\xi d\gamma = p_{\alpha,i}\int_{\mathbb{R}^2} M_{\alpha,i}\mathbf{u}_{\alpha,i}d\xi d\gamma.\sum_{j\in K_i}\mathbf{n}_{i,j} = 0,$$

*and using the discrete form of the continuity equation, for the first term in the first line of* (89) *we get*

$$-\sum_{j\in K_i}\sigma_{i,j}p_{\alpha,i}\int_{\mathbb{R}^2} M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma = p_{\alpha,i}\big(h_{\alpha,i}^{n+1} - h_{\alpha,i} - \Delta t^n(G_{\alpha+1/2,i} - G_{\alpha-1/2,i})\big).$$

*The vertical exchange terms It remains to examine the contribution of the vertical exchange terms over the energy balance, namely in Eq.* (82) *the quantity*

$$\begin{aligned}\mathcal{V}_{\alpha,i} \;:=\; & \Delta t^n\big(gz_{\alpha,i} - \frac{|\mathbf{u}_{\alpha,i}|^2}{2}\big)\Big(\rho^{n+1}_{\alpha+1/2,i}G_{\alpha+1/2,i} - \rho^{n+1}_{\alpha-1/2,i}G_{\alpha-1/2,i}\Big)\\ & +\Delta t^n\mathbf{u}_{\alpha,i}.\Big(\rho^{n+1}_{\alpha+1/2,i}\mathbf{u}^{n+1}_{\alpha+1/2,i}G_{\alpha+1/2,i} - \rho^{n+1}_{\alpha-1/2,i}\mathbf{u}^{n+1}_{\alpha-1/2,i}G_{\alpha-1/2,i}\Big).\end{aligned}$$

*And we write*

$$\begin{aligned}\mathcal{V}_{\alpha,i} \;=\; & \Delta t^n\rho^{n+1}_{\alpha+1/2,i}\frac{|\mathbf{u}^{n+1}_{\alpha+1/2,i}|^2}{2}G_{\alpha+1/2,i} - \Delta t^n\rho^{n+1}_{\alpha-1/2,i}\frac{|\mathbf{u}^{n+1}_{\alpha-1/2,i}|^2}{2}G_{\alpha-1/2,i}\\ & +\Delta t^n g\Big(\rho^{n+1}_{\alpha+1/2,i}z_{\alpha+1/2,i}G_{\alpha+1/2,i} - \rho^{n+1}_{\alpha-1/2,i}z_{\alpha-1/2,i}G_{\alpha-1/2,i}\Big)\\ & -\Delta t^n\rho^{n+1}_{\alpha+1/2,i}\left(\frac{|\mathbf{u}^{n+1}_{\alpha+1/2,i} - \mathbf{u}^{n+1}_{\alpha,i}|^2}{2} + \frac{\mathbf{u}_{\alpha,i} - \mathbf{u}^{n+1}_{\alpha,i}}{2}.(\mathbf{u}^{n+1}_{\alpha,i} - 2\mathbf{u}^{n+1}_{\alpha+1/2,i} + \mathbf{u}_{\alpha,i})\right)G_{\alpha+1/2,i}\\ & +\Delta t^n\rho^{n+1}_{\alpha-1/2,i}\left(\frac{|\mathbf{u}^{n+1}_{\alpha-1/2,i} - \mathbf{u}^{n+1}_{\alpha,i}|^2}{2} + \frac{\mathbf{u}_{\alpha,i} - \mathbf{u}^{n+1}_{\alpha,i}}{2}.(\mathbf{u}^{n+1}_{\alpha,i} - 2\mathbf{u}^{n+1}_{\alpha-1/2,i} + \mathbf{u}_{\alpha,i})\right)G_{\alpha-1/2,i}\\ & -\Delta t^n g\frac{h_{\alpha,i}}{2}\Big(\rho^{n+1}_{\alpha+1/2,i}G_{\alpha+1/2,i} + \rho^{n+1}_{\alpha-1/2,i}G_{\alpha-1/2,i}\Big).\end{aligned} \tag{90}$$

eq:exchang

***Remark 3.12*** *When $G_{\alpha+1/2,i}\le 0$, then the third line of* (90) *is nonnegative namely,*

$$-\Delta t^n\rho^{n+1}_{\alpha+1/2,i}\frac{|\mathbf{u}_{\alpha,i} - \mathbf{u}^{n+1}_{\alpha,i}|^2}{2}G_{\alpha+1/2,i} = \mathcal{O}\big((\Delta t^n)^3\big),$$

*whereas for $G_{\alpha+1/2,i} > 0$*

$$\begin{aligned}&\frac{|\mathbf{u}^{n+1}_{\alpha+1/2,i} - \mathbf{u}^{n+1}_{\alpha,i}|^2}{2} - \frac{\mathbf{u}_{\alpha,i} - \mathbf{u}^{n+1}_{\alpha,i}}{2}.(\mathbf{u}^{n+1}_{\alpha,i} - 2\mathbf{u}^{n+1}_{\alpha+1/2,i} + \mathbf{u}_{\alpha,i})\\ &\qquad = \frac{\mathbf{u}^{n+1}_{\alpha+1,i} - \mathbf{u}^{n+1}_{\alpha,i}}{2}(\mathbf{u}^{n+1}_{\alpha+1,i} - 2\mathbf{u}_{\alpha,i} + \mathbf{u}^{n+1}_{\alpha,i}) = \mathcal{O}\big((\Delta t^n)^2\big).\end{aligned}$$

*Hence, when it is not a dissipative term, the third line of* (90) *is a $\mathcal{O}\big((\Delta t^n)^3\big)$ term. The same result holds for the fourth line of* (90)*.*

*All the contributions*

*Now, summarizing the computations carried out in the previous paragraphs, we sum all the contributions, namely* $\mathcal{E}^t_{\alpha,i}$, $\mathcal{E}^{xy}_{\alpha,i,j}$ *and* $\mathcal{V}_{\alpha,i}$ *leading to a new expression for* $\mathcal{E}_{\alpha,i}$ *under the form*

$$E^{n+1}_{\alpha,i} - E_{\alpha,i} + \sum_{j\in K_i}\sigma_{i,j}\int_{\mathbb{R}^2}\left(gz_{\alpha,i,j} + \frac{\xi^2+\gamma^2}{2} - \frac{1}{2}\left|\begin{pmatrix}\xi\\ \gamma\end{pmatrix} - \tilde{\mathbf{u}}_{\alpha,i,j}\right|^2\right)\rho_{\alpha,i,j}M_{\alpha,i,j}\zeta_{i,j}d\xi d\gamma$$
$$-\Delta t^n\rho^{n+1}_{\alpha+1/2,i}\left(\frac{|\mathbf{u}^{n+1}_{\alpha+1/2,i}|^2}{2} + gz_{\alpha+1/2,i}\right)G_{\alpha+1/2,i} + \Delta t^n\left(\rho^{n+1}_{\alpha-1/2,i}\frac{|\mathbf{u}^{n+1}_{\alpha-1/2,i}|^2}{2} + gz_{\alpha-1/2,i}\right)G_{\alpha-1/2,i}$$
$$= d_{\alpha,i} + e_{\alpha,i} + F_{\alpha,i}$$

*where* $d_{\alpha,i}$ *are non-positive and hence dissipative terms whereas* $e_{\alpha,i}$ *are errors terms.* $d_{\alpha,i}$ *and* $e_{\alpha,i}$ *are given by* (79) *and* (80) *respectively. The quantity* $F_{\alpha,i}$ *is defined by*

$$\begin{aligned}
F_{\alpha,i} &= (\rho_{\alpha,i}gh_{\alpha,i})^{n+1}(z^{n+1}_{\alpha,i} - z_{\alpha,i}) - p_{\alpha,i}\big(h^{n+1}_{\alpha,i} - h_{\alpha,i} - \Delta t^n(G_{\alpha+1/2,i} - G_{\alpha-1/2,i})\big)\\
&\quad -\Delta t^n\rho^{n+1}_{\alpha+1/2,i}g\frac{h_{\alpha,i}}{2}G_{\alpha+1/2,i} - \Delta t^n\rho^{n+1}_{\alpha-1/2,i}g\frac{h_{\alpha,i}}{2}G_{\alpha-1/2,i} + \frac{(\rho_{\alpha,i}h_{\alpha,i})^{n+1}}{2}|\mathbf{u}^{n+1}_{\alpha,i} - \mathbf{u}_{\alpha,i}|^2\\
&\quad -\Delta t^n\rho^{n+1}_{\alpha+1/2,i}\left(\frac{|\mathbf{u}^{n+1}_{\alpha+1/2,i} - \mathbf{u}^{n+1}_{\alpha,i}|^2}{2} + \frac{\mathbf{u}_{\alpha,i} - \mathbf{u}^{n+1}_{\alpha,i}}{2}.(\mathbf{u}^{n+1}_{\alpha,i} - 2\mathbf{u}^{n+1}_{\alpha+1/2,i} + \mathbf{u}_{\alpha,i})\right)G_{\alpha+1/2,i}\\
&\quad +\Delta t^n\rho^{n+1}_{\alpha-1/2,i}\left(\frac{|\mathbf{u}^{n+1}_{\alpha-1/2,i} - \mathbf{u}^{n+1}_{\alpha,i}|^2}{2} + \frac{\mathbf{u}_{\alpha,i} - \mathbf{u}^{n+1}_{\alpha,i}}{2}.(\mathbf{u}^{n+1}_{\alpha,i} - 2\mathbf{u}^{n+1}_{\alpha-1/2,i} + \mathbf{u}_{\alpha,i})\right)G_{\alpha-1/2,i},
\end{aligned}$$

*and the first three lines of* $F_{\alpha,i}$ *can be rewritten under the form*

$$\begin{aligned}
F^1_{\alpha,i} &= (\rho_{\alpha,i}gh_{\alpha,i})^{n+1}(z^{n+1}_{\alpha,i} - z_{\alpha,i}) - p_{\alpha,i}\big(h^{n+1}_{\alpha,i} - h_{\alpha,i}\big) + \frac{(\rho_{\alpha,i}h_{\alpha,i})^{n+1}}{2}|\mathbf{u}^{n+1}_{\alpha,i} - \mathbf{u}_{\alpha,i}|^2\\
&\quad +\Delta t^n\Big(p^{n*}_{\alpha+1/2,i}G_{\alpha+1/2,i} - p^{n*}_{\alpha-1/2,i}G_{\alpha-1/2,i}\Big)\\
&\quad +\Delta t^n\frac{gh_{\alpha,i}}{2}\left((\rho^{n+1}_{\alpha,i} - \rho^{n+1}_{\alpha+1/2,i})G_{\alpha+1/2,i} + (\rho^{n+1}_{\alpha,i} - \rho^{n+1}_{\alpha-1/2,i})G_{\alpha-1/2,i}\right). \qquad (91)
\end{aligned}$$

eq:f_alpha

*Now, we rewrite the first two terms of the first line of* (91) *under the form*

$$\begin{aligned}
F^2_{\alpha,i} &= \frac{(\rho_{\alpha,i}gh_{\alpha,i})^{n+1}}{2}(z^{n+1}_{\alpha+1/2,i} + z^{n+1}_{\alpha-1/2,i} - z_{\alpha+1/2,i} - z_{\alpha-1/2,i})\\
&\quad -p_{\alpha,i}\big(z^{n+1}_{\alpha+1/2,i} - z^{n+1}_{\alpha-1/2,i} - z_{\alpha+1/2,i} + z_{\alpha-1/2,i}\big)\\
&= -p^{n*}_{\alpha+1/2,i}(z^{n+1}_{\alpha+1/2,i} - z_{\alpha+1/2,i}) + p^{n*}_{\alpha-1/2,i}(z^{n+1}_{\alpha-1/2,i} - z_{\alpha-1/2,i})\\
&\quad +g\left((\rho_{\alpha,i}h_{\alpha,i})^{n+1} - \rho^{n+1}_{\alpha,i}h_{\alpha,i}\right)(z^{n+1}_{\alpha,i} - z_{\alpha,i}).
\end{aligned}$$

*Hence, we have for* $F_{\alpha,i}$

$$F_{\alpha,i} = p^{n*}_{\alpha+1/2,i}\big(\Delta t^nG_{\alpha+1/2,i} - z^{n+1}_{\alpha+1/2,i} + z_{\alpha+1/2,i}\big)$$

$$
\begin{aligned}
&-p^{n*}_{\alpha-1/2,i}\big(\Delta t^n G_{\alpha-1/2,i} - z^{n+1}_{\alpha-1/2,i} + z_{\alpha-1/2,i}\big)\Big)\\
&-\Delta t^n \rho^{n+1}_{\alpha+1/2,i}\left(\frac{|\mathbf{u}^{n+1}_{\alpha+1/2,i} - \mathbf{u}^{n+1}_{\alpha,i}|^2}{2} + \frac{\mathbf{u}_{\alpha,i} - \mathbf{u}^{n+1}_{\alpha,i}}{2}.(\mathbf{u}^{n+1}_{\alpha,i} - 2\mathbf{u}^{n+1}_{\alpha+1/2,i} + \mathbf{u}_{\alpha,i})\right) G_{\alpha+1/2,i}\\
&+\Delta t^n \rho^{n+1}_{\alpha-1/2,i}\left(\frac{|\mathbf{u}^{n+1}_{\alpha-1/2,i} - \mathbf{u}^{n+1}_{\alpha,i}|^2}{2} + \frac{\mathbf{u}_{\alpha,i} - \mathbf{u}^{n+1}_{\alpha,i}}{2}.(\mathbf{u}^{n+1}_{\alpha,i} - 2\mathbf{u}^{n+1}_{\alpha-1/2,i} + \mathbf{u}_{\alpha,i})\right) G_{\alpha-1/2,i}\\
&+\Delta t^n \frac{g h_{\alpha,i}}{2}\left((\rho^{n+1}_{\alpha,i} - \rho^{n+1}_{\alpha+1/2,i})G_{\alpha+1/2,i} + (\rho^{n+1}_{\alpha,i} - \rho^{n+1}_{\alpha-1/2,i})G_{\alpha-1/2,i}\right)\\
&+g\left((\rho_{\alpha,i}h_{\alpha,i})^{n+1} - \rho^{n+1}_{\alpha,i}h_{\alpha,i}\right)(z^{n+1}_{\alpha,i} - z_{\alpha,i}) + \frac{(\rho_{\alpha,i}h_{\alpha,i})^{n+1}}{2}|\mathbf{u}^{n+1}_{\alpha,i} - \mathbf{u}_{\alpha,i}|^2,
\end{aligned}
$$

*the first two lines being conservatives terms and the four following ones being error terms corresponding to $f_{\alpha,i}$. The fifth line is the discrete version of the term appearing in the last line of the continuous energy balance given in Eq.* (26) *(see also (Boittin et al., 2020)) and the last line of the previous relation comes from the time scheme and is of order $\mathcal{O}\big((\Delta t^n)^2\big)$.*

*In order to conclude the proof, it remains to prove that all the quantities appearing in $e_{\alpha,i}$ are third order terms i.e. of magnitude $\mathcal{O}(diam(C_i)^3)$. The terms in each of the sums in $e_{\alpha,i}$ are obviously second-order terms. Since the sum is made on the faces, gradients of second-order terms appear. These gradients are indeed third-order terms.* ■

If one wishes to introduce even more upwinding in the discretization of $S_{p,\alpha}$ by using the discretization

$$
\mathcal{S}_{p,\alpha,i,j} = p_{\alpha,i}(\hat{h}_{\alpha,i,j} - h_{\alpha,i})\mathbf{n}_{i,j} - g\rho_{\alpha,i,j}\hat{h}_{\alpha,i,j}(z_{\alpha,i,j} - z_{\alpha,i})\mathbf{n}_{i,j},
$$

Proposition 3.10 still holds but the expression of the rest term $e_{\alpha,i}$ becomes

$$
\begin{aligned}
e_{\alpha,i} \;=\; & \sum_{j\in K_i}\sigma_{i,j} g\int_{\mathbb{R}^2}\rho_{\alpha,i,j}M_{\alpha,i,j}\big(z_{\alpha,i,j} - z_{\alpha,i}\big)(\mathbf{u}_{\alpha,i} - \mathbf{u}_{\alpha,i,j})\cdot\mathbf{n}_{i,j}d\xi d\gamma\\
&+\sum_{j\in K_i}\sigma_{i,j} g\int_{\mathbb{R}^2}\rho_{\alpha,i,j}M_{\alpha,i,j}\big(z_{\alpha,i,j} - z_{\alpha,i}\big)\big(\mathbf{u}_{\alpha,i,j}\cdot\mathbf{n}_{i,j} - \zeta_{i,j}\big)d\xi d\gamma\\
&-\sum_{j\in K_i}\sigma_{i,j}p_{\alpha,i}\int_{\mathbb{R}^2}(M_{\alpha,i,j} - M_{\alpha,i})(\tilde{\mathbf{u}}_{\alpha,i,j} - \mathbf{u}_{\alpha,i})\cdot\mathbf{n}_{i,j}d\xi d\gamma\\
&-\sum_{j\in K_i}\sigma_{i,j}(\tilde{\mathbf{u}}_{\alpha,i,j} - \mathbf{u}_{\alpha,i})\cdot\int_{\{\zeta_{i,j}\le 0\}}\left[\left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix} - \mathbf{u}_{\alpha,i}\right)\rho_{\alpha,i}M_{\alpha,i} - \left(\begin{pmatrix}\xi\\ \gamma\end{pmatrix} - \mathbf{u}_{\alpha,j}\right)\rho_{\alpha,j}M_{\alpha,j}\right]\zeta_{i,j}d\xi d\gamma\\
&-\sum_{j\in K_i}\sigma_{i,j}\frac{|2\tilde{\mathbf{u}}_{\alpha,i,j} - \mathbf{u}_{\alpha,i} - \mathbf{u}_{\alpha,j}|^2}{4}\int_{\{\zeta_{i,j}\le 0\}}\rho_{\alpha,j}M_{\alpha,j}\zeta_{i,j}d\xi d\gamma.
\end{aligned}
$$

This new expression of $e_{\alpha,i}$ also contains only third-order error terms, thus the result is not deteriorated.

# 4 Numerical scheme for the layer-averaged Navier-Stokes-Fourier system

⟨sec:NS_num⟩ The discretization of the full layer-averaged Navier-Stokes-Fourier is presented. The main difficulties have already been tackled in section 3.

## 4.1 Semi-discrete (in time) scheme

The semi-discrete in time scheme (34) yields the following system

$$
\begin{aligned}
h^{n+1} &= h^{n+1/2} = h - \Delta t^n \sum_{\alpha=1}^{N} \nabla_{x,y}\cdot(h_\alpha \mathbf{u}_\alpha) - \sum_{\alpha=1}^{N} \frac{\rho'(T_\alpha)}{\rho_\alpha^2 c_p}(S_{T,\alpha} - S_{\mu,\alpha}), && (92)\ \boxed{\texttt{eq:NSF\_sd\_1}}\\
(\rho_\alpha h_\alpha)^{n+1/2} &= \rho_\alpha h_\alpha - \Delta t^n \nabla_{x,y}\cdot(\rho_\alpha h_\alpha \mathbf{u}_\alpha), && (93)\ \texttt{?eq:NSF\_sd\_2?}\\
(\rho_\alpha h_\alpha \mathbf{u}_\alpha)^{n+1/2} &= \rho_\alpha h_\alpha \mathbf{u}_\alpha - \Delta t^n \Big( \nabla_{x,y}\cdot(\rho_\alpha h_\alpha \mathbf{u}_\alpha \otimes \mathbf{u}_\alpha) + \nabla_{x,y}\big(h_\alpha p_\alpha\big)\\
&\quad - p_\alpha \nabla_{x,y} h_\alpha + \rho_\alpha g h_\alpha \nabla_{x,y} z_\alpha \Big), && (94)\ \texttt{?eq:NSF\_sd\_3?}\\
(\rho_\alpha h_\alpha)^{n+1} &= (\rho_\alpha h_\alpha)^{n+1/2} - \Delta t^n \Big( \rho^{n+1}_{\alpha+1/2} G_{\alpha+1/2} - \rho^{n+1}_{\alpha-1/2} G_{\alpha-1/2} \Big), && (95)\ \texttt{?eq:NSF\_sd\_4?}\\
(\rho_\alpha h_\alpha \mathbf{u}_\alpha)^{n+1} &= (\rho_\alpha h_\alpha \mathbf{u}_\alpha)^{n+1/2} - \Delta t^n \Big( \mathbf{u}^{n+1}_{\alpha+1/2} \rho^{n+1}_{\alpha+1/2} G_{\alpha+1/2} - \mathbf{u}^{n+1}_{\alpha-1/2} \rho^{n+1}_{\alpha-1/2} G_{\alpha-1/2}\\
&+ \nabla_{x,y}\cdot(\mu h^{n+1}_\alpha \nabla_{x,y} \mathbf{u}^{n+l}_\alpha) + \Gamma^{n+1}_{\alpha+1/2}(\mathbf{u}^{n+1}_{\alpha+l} - \mathbf{u}^{n+l}_\alpha)\\
&- \Gamma^{n+1}_{\alpha-1/2}(\mathbf{u}^{n+l}_\alpha - \mathbf{u}^{n+l}_{\alpha-1}) - \kappa_\alpha \mathbf{u}^{n+l}_\alpha \Big), && (96)\ \boxed{\texttt{eq:NSF\_sd\_5}}
\end{aligned}
$$

where $l = 0, 1$ and with

$$
G_{\alpha+1/2} = -\sum_{j=1}^{N}\Big(\sum_{p=1}^{\alpha} l_p - \mathbb{1}_{j\le\alpha}\Big)\nabla_{x,y}\cdot(h_j \mathbf{u}_j) + \sum_{j=1}^{\alpha} \frac{\rho'(T_j)}{\rho_j^2 c_p}(S_{T,j} - S_{\mu,j}).
$$

and $\Gamma_{\alpha+1/2}$, $S_{T,\alpha}$, and $S_{\mu,\alpha}$ respectively defined by (22),(20) and (21).

Note that the definition of the mass exchange terms $G_{\alpha+1/2}$ is different from the definition of the mass exchange terms for the Euler system given in (41).

## 4.2 Spatial discretization of the diffusion terms

The Euler part of the system is discretized in space as in section 3.4. We present here only the discretization of the diffusion terms. For the discretization of the viscosity terms in the momentum equation, we refer to (Allgeyer et al., 2019). A classical $\mathbb{P}_1$ finite element type approximation with mass-lumping is used. Let us define the number of cells $N_x$ as well as the vector of unknowns in layer $\alpha$

$$
\begin{aligned}
\mathbf{U}_\alpha = (&h_{\alpha,1}, \dots, h_{\alpha,N_x}, (\rho_\alpha h_\alpha)_1, \dots, (\rho_\alpha h_\alpha)_{Nx},\\
&(\rho_\alpha h_\alpha u_\alpha)_1, \dots, (\rho_\alpha h_\alpha u_\alpha)_{Nx}, (\rho_\alpha h_\alpha v_\alpha)_1, \dots, (\rho_\alpha h_\alpha v_\alpha)_{Nx})^T
\end{aligned}
$$

and the vector containing the temperatures in layer $\alpha$

$$\mathbf{T}_\alpha = (T_{\alpha,1}, \dots, T_{\alpha,N_x})^T.$$

The discretization of $\nabla_{x,y} \cdot (\mu h_\alpha^{n+1} \nabla_{x,y} \mathbf{u}_\alpha^{n+1}) + \Gamma_{\alpha+1/2}^{n+1}(\mathbf{u}_{\alpha+1}^{n+1} - \mathbf{u}_\alpha^{n+1}) - \Gamma_{\alpha-1/2}^{n+1}(\mathbf{u}_\alpha^{n+1} - \mathbf{u}_{\alpha-1}^{n+1})$ reads

$$-\mathcal{K}_{\mu,\alpha}\mathbf{U}_\alpha + \mathcal{M}_{\mu,\alpha+1/2}(\mathbf{U}_{\alpha+1} - \mathbf{U}_\alpha) - \mathcal{M}_{\mu,\alpha+1/2}(\mathbf{U}_\alpha - \mathbf{U}_{\alpha-1}),$$

with the $4N_x \times 4N_x$ block matrices

$$\mathcal{K}_{\mu,\alpha} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \mathcal{K}'_{\mu,\alpha} & 0 \\ 0 & 0 & 0 & \mathcal{K}'_{\mu,\alpha} \end{pmatrix}, \quad \mathcal{M}_{\mu,\alpha+1/2} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \mathcal{M}'_{\mu,\alpha+1/2} & 0 \\ 0 & 0 & 0 & \mathcal{M}'_{\mu,\alpha+1/2} \end{pmatrix}$$

where the non-zero coefficients are given by

$$(\mathcal{K}'_{\mu,\alpha})_{j,i} = \frac{3}{A_j} \frac{\mu}{(\rho_\alpha h_\alpha)_i} \int_\Omega h_\alpha \nabla_{x,y}\varphi_i \cdot \nabla_{x,y}\varphi_j dxdy,$$
$$(\mathcal{M}'_{\mu,\alpha+1/2})_{j,i} = \frac{\mu}{(\rho_\alpha h_\alpha)_i} \frac{\delta_{i,j}}{h_{\alpha+1,i} + h_{\alpha,i}}.$$

The $\varphi_i$ are the basis functions and $\delta_{i,j}$ is the Kronecker symbol. The area $A_j$ is the area of the support of the test function $\varphi_j$.
The discretization of the terms $S_{T,\alpha}$ due to temperature diffusion is similar. To discretize $\frac{\rho'(T_\alpha)}{\rho_\alpha^2 c_p} S_{T,\alpha}$, we propose the simplification

$$\frac{\rho'(T_{\alpha,i})}{\rho_{\alpha,i}^2 c_p} (S_{T,\alpha})_i.$$

The term $S_{\mu,\alpha}$ ensures that the energy of the layer-averaged system is consistent, see (Boittin et al., 2020). Indeed, the following identity holds

$$\begin{aligned} \mathbf{u}_\alpha \cdot \nabla_{x,y} \cdot (\mu h_\alpha \nabla_{x,y} \mathbf{u}_\alpha) = \nabla_{x,y} \cdot (\mu h_\alpha \mathbf{u}_\alpha \nabla_{x,y} \mathbf{u}_\alpha) \\ + \Gamma_{\alpha+1/2} \frac{|\mathbf{u}_{\alpha+1}|^2 - |\mathbf{u}_\alpha|^2}{2} - \Gamma_{\alpha-1/2} \frac{|\mathbf{u}_\alpha|^2 - |\mathbf{u}_{\alpha-1}|^2}{2} + S_{\mu,\alpha}. \end{aligned}$$

It is important to ensure the same consistency at the discrete level, i.e. an analogous identity should be verified at the discrete level. This gives guidelines for the discretization of $S_{\mu,\alpha}$. A similar problem is studied in (Grapsas et al., 2016) in the case of the compressible Navier-Stokes equations. However the numerical analysis of the scheme with the dissipation and diffusion terms has not been done yet. It will be performed in a future work.

# 5 Numerical validation

ec:validation⟩ In this section, we confront the proposed numerical scheme to three test cases. In the first one we present convergence results in the case of a 3d analytical solution for the Euler system. This test case is 3d, non-stationary with wet/dry interfaces and density variations and hence it is a challenging problem to obtain convergence curves towards the analytical solution. The second test deals with the simulation of the lock exchange phenomenon and the comparison with experimental results available in the literature. The third test consists in two simple diffusion cases for which we present a validation with an analytical solution and comparisons between the Navier-Stokes-Fourier and Boussinesq models. All the presented simulations have been obtained with the numerical code Freshkiss3d (2020).

## 5.1 Parabolic bowl

ubsec:thacker⟩ In (Bristeau et al., 2020), the authors have proposed an analytic solution for the Euler system i.e. the system (1)-(3) with $\lambda = 0, \mu = 0$. It is an extension of the Thacker' analytical solution (Thacker, 1981), corresponding to a periodic oscillation in a parabolic bowl.

The following proposition holds, its proof is detailed in (Bristeau et al., 2020).

**Proposition 5.1** *For any nonnegative function $s \mapsto \rho(s)$ and for some $(a, \alpha, \eta, h_0) \in \mathbb{R}^3 \times \mathbb{R}_+$, let us consider the functions $h, u, v, w, p, \phi$ defined for $(x, y) \in [-L/2, L/2]^2$, $t \geq t_0$ by*

$$
\begin{aligned}
&h(t,x,y) = \max\left\{0, h_0 - \alpha \frac{\left(x - \eta\cos(\omega t)\right)^2 + \left(y - \eta\sin(\omega t)\right)^2}{2}\right\},\\
&u(t,x,y,z) = -\eta\omega\sin(\omega t),\\
&v(t,x,y,z) = \eta\omega\cos(\omega t),\\
&w(t,x,y,z) = -\alpha\eta\omega\left(x\sin(\omega t) - y\cos(\omega t)\right),\\
&p(t,x,y,z) = p^a(t) + \int_z^{h+z_b} \rho(T(t,x,y,z_1))dz_1,\\
&T(t,x,y,z) = a\left(h + z_b - z\right),
\end{aligned}
$$

*with $\omega = \sqrt{\alpha g}$ and with a bottom topography defined by $z_b(x,y) = \frac{\alpha}{2}(x^2 + y^2)$, then $h, u, v, w, p, \phi$ as defined previously satisfy the 3D hydrostatic Euler system with variable density (system (1)-(3) with $\lambda = 0, \mu = 0$) completed with the kinematic boundary conditions (4), (6).*

NS2d_var_dens⟩ For the numerical validation the parameters are set to $\eta = 0.1$, $h_0 = 0.1$, $a = 10$, $\alpha = 1$ and $L = 4$ and we consider a simplified state law given by $\rho(T) = \rho_0 + \beta T$ with $\rho_0 = 1000$ and $\beta = 10$. The free surface is plotted at different times in Figure 5 highlighting the planar motion of the fluid in the bowl. On Figure 6, the density in the

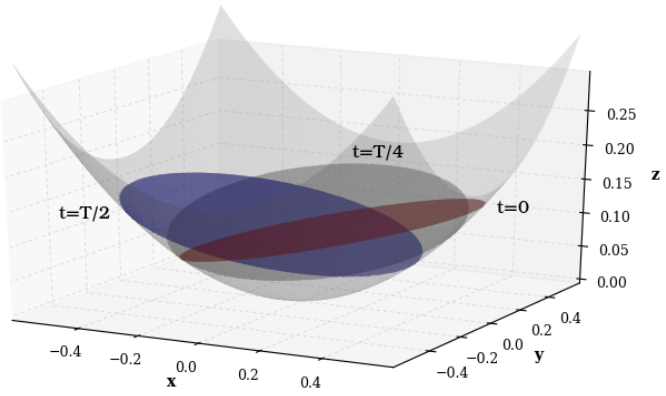


Figure 5: Analytical solution of prop. 5.1, 3D planar surface in a parabolic bowl: free surface at $t = 0$ (red), $t = \tau/4$ (dark grey), $t = \tau/2$ (blue), with the period $\tau$ defined by $\tau = 2\pi/\omega$.

r2d_xz_slice2⟩

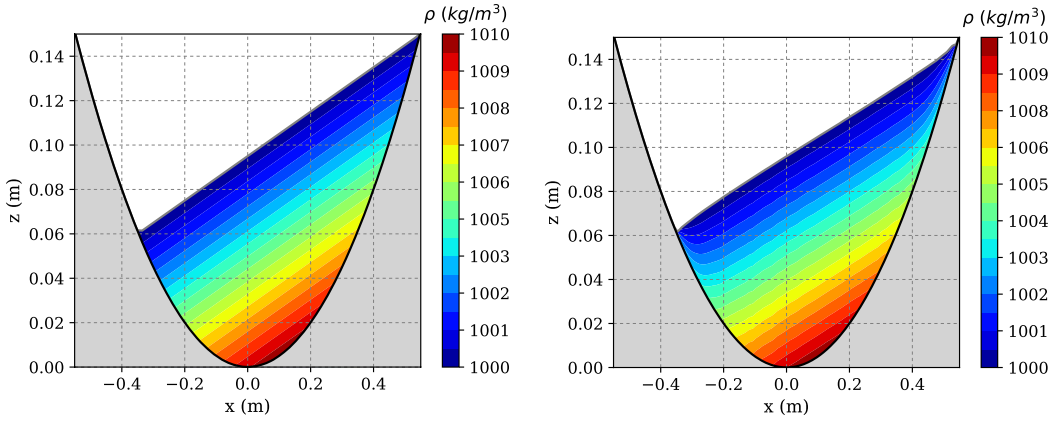


Figure 6: Numerical result of the parabolic bowl with variable density. Free surface and density contour in the slice plane (x, y=0, z) at initial time (left) and at time $\tau = 2\pi/\omega$ with first order scheme (right).

density_slice⟩

slice plane (x, y=0, z) after a period for a mesh with 31316 triangles and 30 layers is plotted.

The convergence towards the analytic solution is assessed by plotting the logarithm of the cumulative error (in $L^2$-norm) at time $\tau = 2\pi/\omega$ versus the space discretization (i.e. $\log(l_0/l_i)$ where $l_0$ is the average edge length of the mesh $i$ and $l_0$ the average edge length of the coarsest mesh) for several unstructured meshes with 934, 2194, 4020, 6408, 9066, 12674 triangles. More precisely, the error in $L^2$-norm is computed by summing on all the nodes of the mesh at each time step, for each layer. Then, the cumulative error is obtained by summing the errors at each time step, normalized by the time step.

Figures 7 and 8 show the cumulative errors obtained with a constant number of layers equal to 10 (the mesh is refined in the horizontal direction but not in the vertical direc-

tion). The analytic solution being non-stationary, errors in time and space accumulate over time and the theoretical rate of convergence is thus hard to obtain.

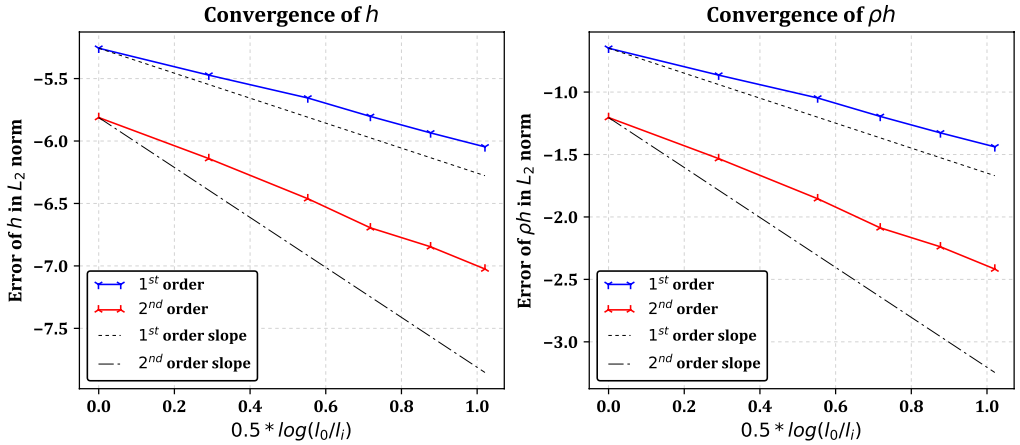


Figure 7: Parabolic bowl: convergence of $h$ and $\rho h$ in $L^2$-norm towards the analytical solution, constant number of layers.

ergence_cstNL⟩

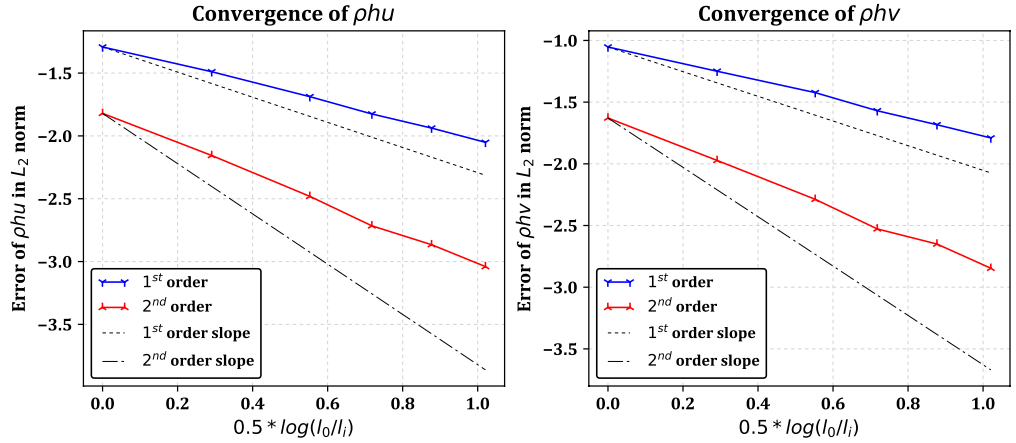


Figure 8: Parabolic bowl: convergence of $\rho hu$ and $\rho hv$ in $L^2$-norm towards the analytical solution, constant number of layers.

ergence_cstNL⟩

Figures 9 and 10 show the cumulative error in $L^2$ norm for meshes with 934, 2194, 4020, 6408, 9066, 12674 triangles and 10, 12, 14, 16, 18, 20 layers respectively. Increasing the number of layers while refining the horizontal mesh is a reasonable idea because by doing so, the proportions of the 3D wedge cells are preserved. A super-convergence phenomenon can be observed for $\rho h$ when the number of layers is increased as the mesh is refined. A rate of convergence higher than the theoretical rate is also observed for $\rho hu$ and $\rho hv$. The faster one refines the mesh in the vertical direction, the higher the convergence rates obtained. The results shown on figures 9 and 10 prove the stability of the numerical scheme for the Euler system.

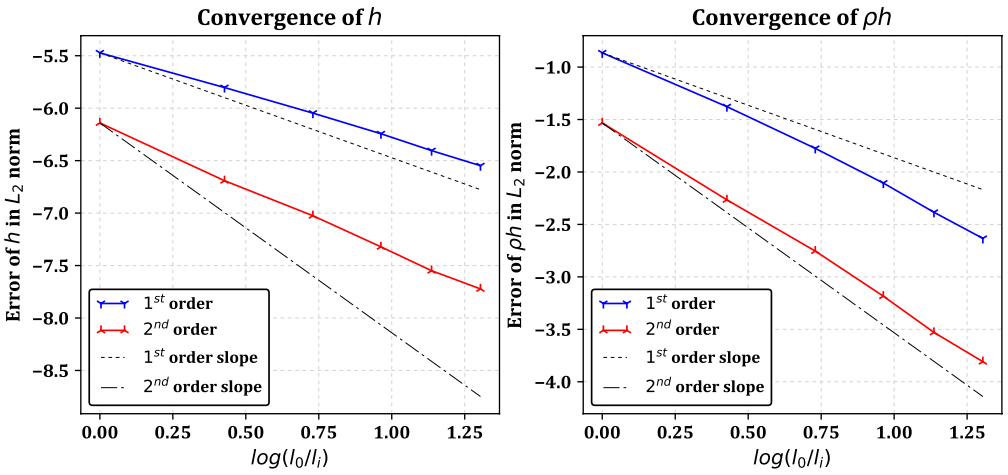


Figure 9: Parabolic bowl: convergence of $h$ and $\rho h$ in $L^2$-norm towards the analytical solution, increasing number of layers.

h_convergence⟩

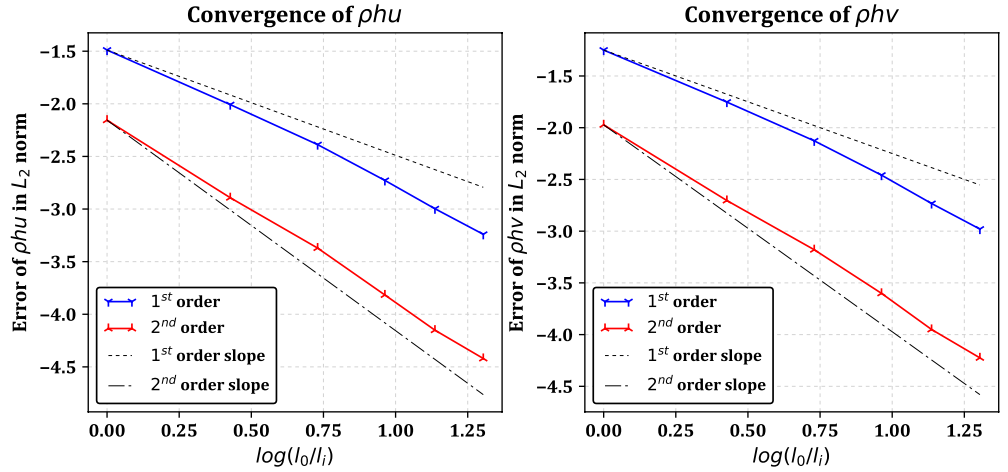


Figure 10: Parabolic bowl: convergence of $\rho hu$ and $\rho hv$ in $L^2$-norm towards the analytical solution, increasing number of layers.

v_convergence⟩

To summarize the results on the convergence of the schemes and to overcome the difficulty of interpreting the super-convergence cases, we propose another way of plotting the results in the case of multi-layer models. The error is plotted as a function of the horizontal space step $l_i$ and of the vertical proportion $l_p = 1/N$ at the same time. Figure 11 shows the horizontal convergence rates of the first- and second-order schemes, as well as the first order vertical convergence. Note that the vertical discretization proportion does not correspond to the vertical space step because the water depth varies ($h_\alpha = l_p h$).

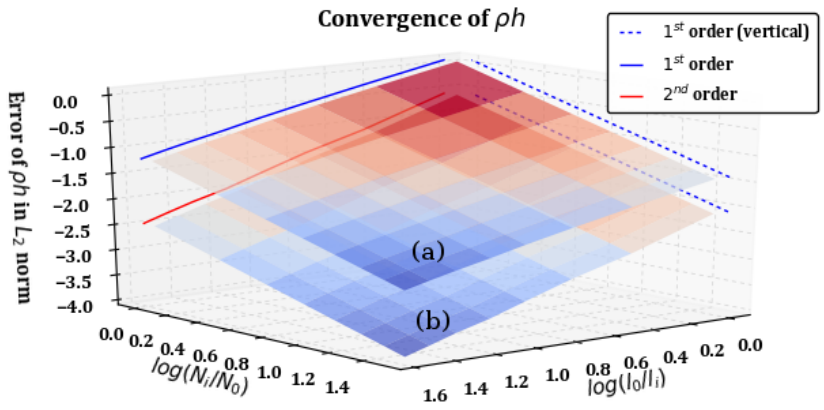


Figure 11: Parabolic bowl: error of $\rho h$ in $L^2$-norm as a function of vertical and horizontal discretization for first order (a) and second order (b) numerical schemes.

ergence_plane⟩

## 5.2 Lock exchange

Gravity currents triggered by lock-exchanges are encountered in many applications and their numerical simulation is a challenge. In this section we show the ability of our numerical scheme to properly simulate the propagation of lock-exchange induced density currents. The computed front position is compared to the experiments carried out by Adduce et al. (2012). The results presented were obtained with the Navier-Stokes-Fourier model where the dissipation terms $S_{\mu,\alpha}$ due to the viscosity were neglected.

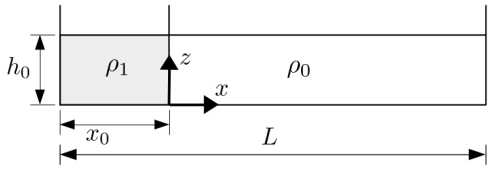


Figure 12: Fluid domain of the lock-exchange test case

fluid_domain⟩?

Initially, fluids of different densities, $\rho_1 = 1090kg.m^{-3}$ and $\rho_0 = 1000kg.m^{-3}$ respectively, are at rest and separated by a wall located at $x_0 = 0.3m$. When the vertical barrier is removed, the denser fluid flows under the lighter one due to the difference in the hydrostatic pressure. The initial water height is $h_0 = 0.3m$ and the length of the domain is $L = 3m$. The reduced gravity is defined by $g^* = g\,(1-\gamma)$ where $\gamma = \rho_0/\rho_1$ is the density ratio. We also define the buoyancy velocity as $u_b = \sqrt{g^* h_0}$. The Navier-Stokes equations are usually made dimensionless using the Grashof number defined by

$$Gr = \left(\frac{u_b h_0}{\nu}\right)^2, \tag{97}$$ {?}

with $\nu$ the kinematic viscosity. Simulations have been carried out with a Grashof number of $Gr = 2.53 \times 10^8$ on different meshes. The evolution of the density and the position of the front are presented in figures 13 and 14 respectively. The initial opening of the

gate has been delayed ($1s$) to take into account the time of opening in the experiment. The figure 14 shows the convergence of the numerical scheme and we observe a good matching between the numerical simulation and the experimental data of Adduce et al. (2012).

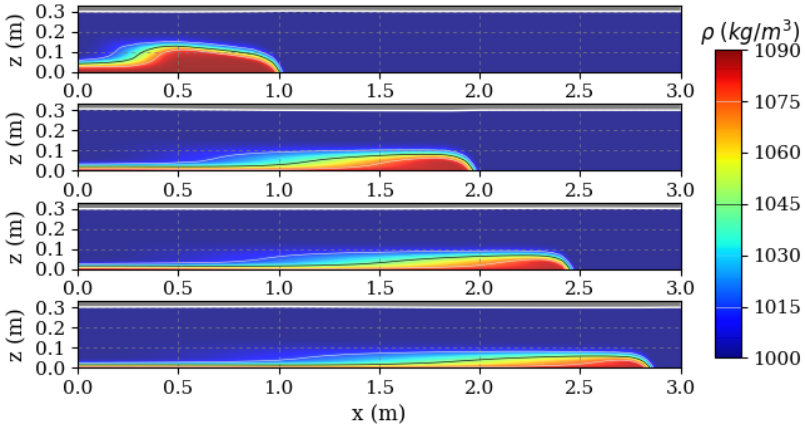


Figure 13: Lock exchange: computed density with the most refined mesh in the slice plane $(x, y = 0, z)$ with $Gr = 2.53 \times 10^8$ at times $t = 3$, 7, 9 and $11s$ (from top to bottom).

sults_density⟩

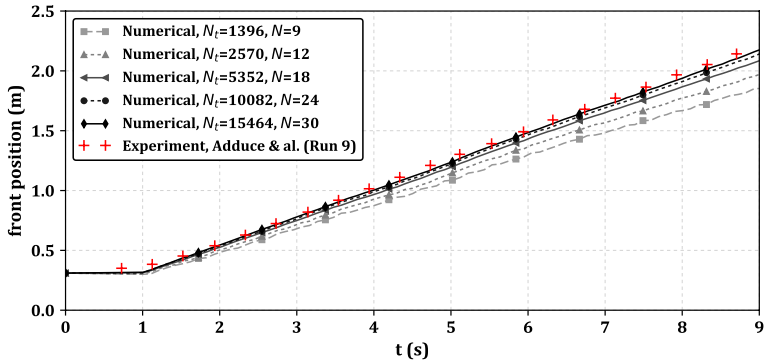


Figure 14: Lock exchange: front position as a function of time for different meshes with comparison to Adduce et al. (2012) experimental results (where $N_t$ is the number of triangles and $N$ is the number of layers).

results_front⟩

## 5.3 Comparison with the Boussinesq assumption

In this section we mainly test the effects of the temperature diffusion and we compare the solutions obtained with the model (1)-(3) to the solutions with the Boussinesq assumption defined below. So, we introduce the system

$$\nabla \cdot \mathbf{U} = 0, \tag{98}$$

eq:div_boussi

$$\rho_0 c_p \frac{\partial T}{\partial t} + \nabla \cdot (T\mathbf{U}) = \nabla \cdot (\lambda \nabla T), \tag{99}$$ ?eq:temp_bouss

$$\rho_0 \left( \frac{\partial \mathbf{u}}{\partial t} + \nabla_{x,y} \cdot (\mathbf{u} \otimes \mathbf{u}) + \frac{\partial (\mathbf{u} w)}{\partial z} \right) + \nabla_{x,y} \int_z^\eta \rho g dz = 0, \tag{100}$$ eq:mom_boussi

i.e. where the Boussinesq assumption is made - $\rho_0$ is a constant.

We consider simple diffusion cases in which a basin is initially at rest i.e. $\mathbf{U} = (0,0,0)^T$ and the free surface and bottom are flat and equal to $h(t_0,x,y) = h_0$ and $z_b(t_0,x,y) = 0$ respectively. In the basin, the temperature is initially distributed such that $T(t_0,x,y,z) = T_0(z)$, where $T_0 = T_0(z)$ is a given function.

Starting from the initial conditions described above, it is easy to see that the velocity and density for $t \geq t_0$ are given by $\mathbf{U} = (0,0,0)^T$ and $\rho = \rho(T)$, where $T = T(t,z)$ is governed by the heat equation:

$$\begin{cases} \dfrac{\partial T}{\partial t} = \mathcal{D}_0 \dfrac{\partial^2 T}{\partial z^2}, \\ T(t_0,z) = T_0(z). \end{cases} \tag{101}$$ eq:heat_bouss

The coefficient $\mathcal{D}_0$ is the diffusivity defined by $\mathcal{D}_0 = \frac{\lambda}{\rho_0 c_p}$.

Now considering the system (1)-(3) with $\mu = 0$ i.e. without the Boussinesq assumption and starting also from the initial conditions described above, it is easy to see that the velocity and temperature for $t \geq t_0$ is given by $\mathbf{U} = (0,0,\overline{w})^T$ and $T = T(\rho)$, where $\overline{w}$ is defined by:

$$\bar{w} = -\frac{\lambda}{c_p} \int_{z_b}^z \frac{\rho'}{\rho^2} \frac{\partial^2 T}{\partial z^2} dz,$$

where $\rho' = \frac{\partial \rho}{\partial T}$ is deduced from the state law and $\rho = \rho(t,z)$ is governed by the equation:

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \overline{w} \dfrac{\partial \rho}{\partial z} = \dfrac{\lambda}{c_p} \dfrac{\rho'}{\rho} \dfrac{\partial^2 T}{\partial z^2}, \\ \rho(t_0,z) = \rho_0(z). \end{cases} \tag{102}$$ eq:heat_non_bo

The velocity $\overline{w}$ is the result of the fluid dilatation. When $\overline{w} = 0$ the two models are almost identical, the only difference being the diffusivity $\mathcal{D}$. It is either a constant in the Boussinesq model or defined by $\mathcal{D} = \frac{\lambda}{\rho c_p}$ in the case of the Navier-Stokes-Fourier model. In the following examples we show that for common water state laws, $\overline{w}$ is sufficiently small to have no noticeable effect on $T$ and $\rho$. However, this term is essential in order to obtain rigorous mass conservation.

### 5.3.1 Temperature diffusion

In this test, the initial temperature $T_0$ (and density $\rho_0$) is constant in the domain and a Dirichlet boundary condition is applied at the bottom with a temperature of $T_b = 0$ (cf. figure 15). At the free surface we impose a homogeneous Neumann boundary condition ($\phi_T|_\eta = 0$, where $\phi_T$ is the thermal flux).

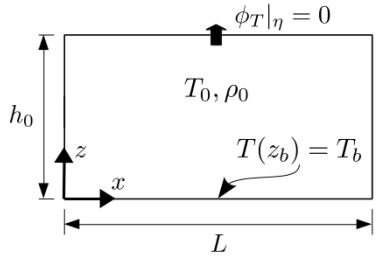


Figure 15: Fluid domain of the diffusion test case with Dirichlet boundary condition at the bottom.

n_diffusion_1⟩

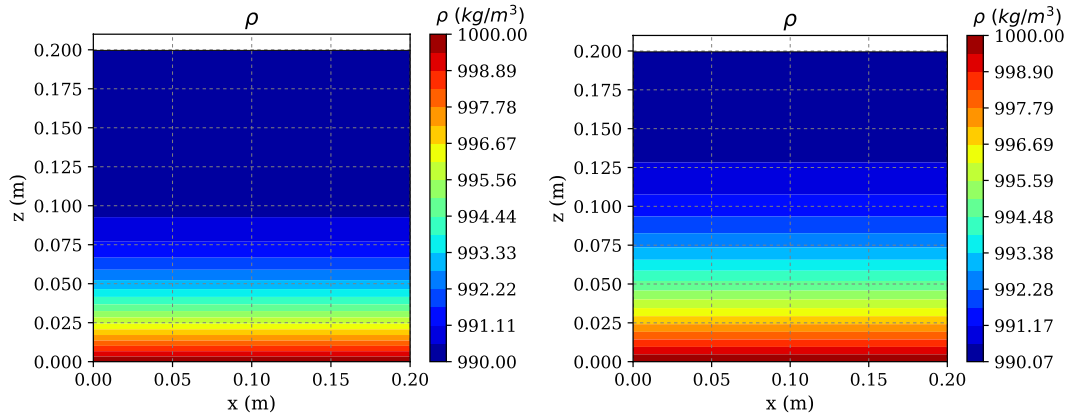


Figure 16: Temperature diffusion: evolution of the density in the slice plane (x, y=0, z) with the Navier-Stokes-Fourier model at time $\tilde{t} = 0.03$ (left) and $\tilde{t} = 0.06$ (right).

ain_diffusion⟩

In Fig. 16, we show the density distribution in the slice plane (x, y=0, z) obtained with the Navier-Stokes-Fourier model at two different times $\tilde{t} = 0.03$ and $\tilde{t} = 0.06$, $\tilde{t}$ being defined below.

Dimensionless parameters are defined such that $\tilde{z} = z/h_0$, $\tilde{t} = \frac{\mathcal{D}_0 t}{h_0^2}$ and $T = T_b + (T_0 - T_b)\tilde{T}$. From Eq. (101) we obtain the dimensionless heat equation

$$\frac{\partial \tilde{T}}{\partial \tilde{t}} = \frac{\partial^2 \tilde{T}}{\partial \tilde{z}^2}, \tag{103}$$

?eq:heat_bouss

whose analytical solution is given (for the boundary conditions $\tilde{T}_0 = 1$ and $\tilde{T}_b = 0$) by

$$\tilde{T}_{an}(\tilde{z}, \tilde{t}) = \frac{2}{\sqrt{\pi}} \int_0^{\frac{\tilde{z}}{2\sqrt{\tilde{t}}}} exp(-\xi^2) d\xi. \tag{104}$$

eq:heat_equat

We compare the temperature obtained by a numerical resolution of the system (92)-(96) (with a simplified state law given by $T(\rho) = \frac{\rho_0 - \rho}{\beta}$ with $\rho_0 = 1000$ and $\beta = 10$) to the analytical solution (104). Notice that from Eq. (102) we obtained a slightly different dimensionless heat equation

$$\frac{\partial \tilde{T}}{\partial \tilde{t}} - \frac{\rho(T_0 - T_b)}{h_0} \frac{\partial \tilde{T}}{\partial \tilde{z}} \int_{z_b}^{z} \frac{\rho'}{\rho^2} \frac{\partial \tilde{T}}{\partial \tilde{z}} dz = \frac{\partial^2 \tilde{T}}{\partial \tilde{z}^2}.$$

We obtain a good matching between the numerical results and the analytical solution both with the Navier-Stokes-Fourier system (with $\mu = 0$) and the Boussinesq system (98)-(100) which validates the numerical treatment of the diffusion. The results are plotted on Figure 17 in the case of the Navier-Stokes-Fourier model. However, there is no noticeable difference between the analytical solution and the Navier-Stokes-Fourier numerical solution even though $\rho' = -\beta$ has been chosen higher than its real physical value (in the case of water $0.03 < \rho' < 0.13$).

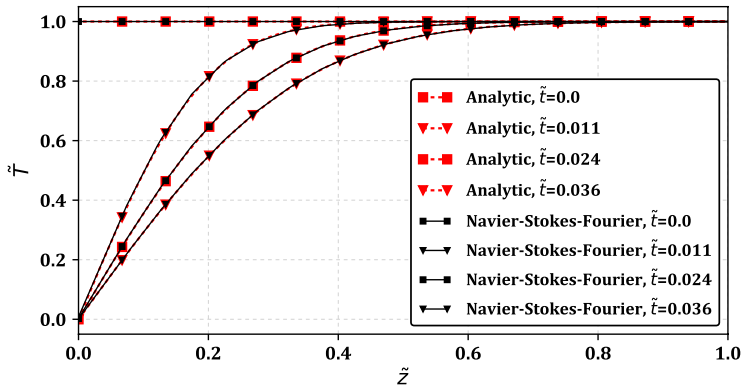


Figure 17: Temperature diffusion: dimensionless temperature $\tilde{T}$ as a function of $z/h_0$ at different times and comparison between analytical solution and numerical simulation with a number of layers equal to $N = 20$.

tion_analytic⟩

Note that the mass is strictly conserved in the Navier-Stokes-Fourier model, even though the density varies. Dilatation effects induce a variation of the water height such that the integral of the density over the total volume of fluid stays constant over time. This is not the case with the Boussinesq model, where the volume is not affected by temperature variation and mass conservation is violated (cf. Figure 18).

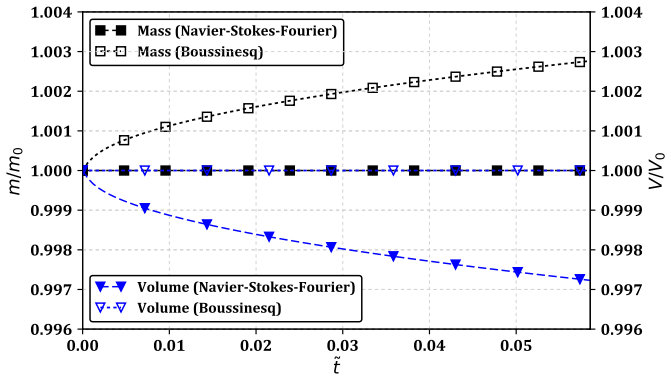


Figure 18: Evolution of the mass ratio $(m/m_0)$ and volume ratio $(V/V_0)$ for the Navier-Stokes-Fourier and Boussinesq models.

ass_vs_volume⟩

### 5.3.2 Thermal equilibrium

In this second test a well stratified fluid is considered. No exterior forcing is applied and zero thermal flux boundary conditions are considered so that only the internal diffusion due to gradients of temperature affects the evolution of the density in the fluid (cf. Figure 19). Given a non null thermal conductivity, the density converges towards a stationary and uniform solution.

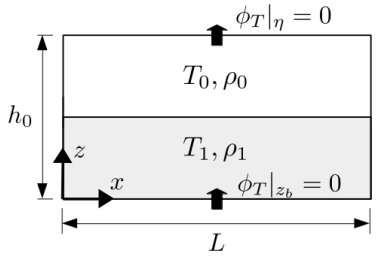


Figure 19: Thermal equilibrium: fluid domain for the diffusion test case

n_diffusion_2⟩

For the numerical simulation the parameters have been set to $h_0 = 2m$, $\rho_0 = 995.52kg.m^{-3}$ and $\rho_1 = 999.76kg.m^{-3}$. We now use a more realistic water state law defined by $T(\rho) = 4 + \sqrt{\frac{\rho_0 - \rho}{\beta \rho_0}}$ with $\beta = 6.63 \times 10^{-6}$ and $\rho_0 = 1000kg.m^{-3}$. This gives the following initial temperatures $T_0 = 30°C$ and $T_1 = 10°C$. We define the dimensionless $z$ coordinate and time as $\tilde{z} = z/h_0$ and $\tilde{t} = \frac{a_m t}{h_0^2}$, where $a_m$ is the initial mean diffusivity defined by $a_m = \frac{\lambda}{\rho_m c_p}$ with $\rho_m = (\rho_0 + \rho_1)/2$.

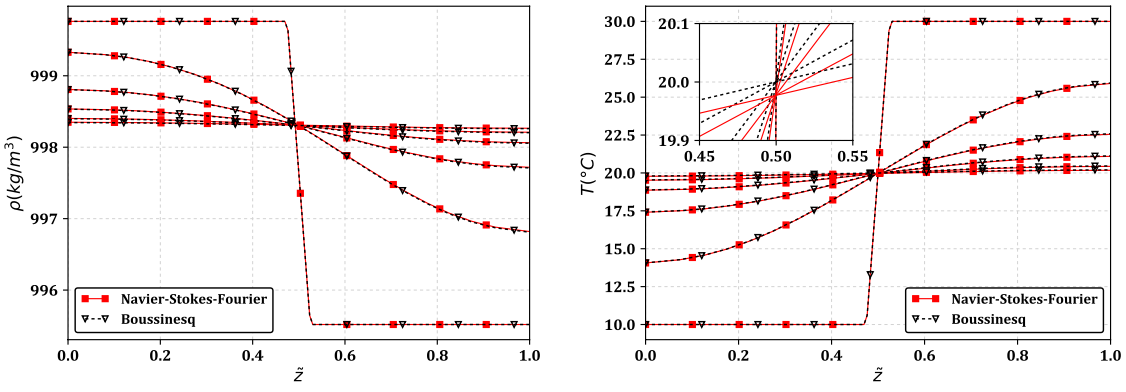


Figure 20: Thermal equilibrium: density (left) and temperature (right) against $z/h_0$ with the Navier-Stokes-Fourier and Boussinesq models at times $\tilde{t} = 0$, 0.07, 0.16, 0.24, 0.33 and 0.42 with $N = 20$

_2_comp_T_Rho⟩

The evolution of density and temperature obtained with both Navier-Stokes-Fourier and Boussinesq is plotted in Fig. 20. The temperature converges rigorously towards an equilibrium temperature of $T_{eq} = T_m = (T_0 + T_1)/2 = 20°C$ in the case of the

Boussinesq model. For the Navier-Stokes-Fourier model, the equilibrium temperature is shifted below $T_m$ due to dilatation effects and is equal to $T_{eq} = 19.977°C$. Because of the diffusion effects, the equilibrium temperature is lower than $T_m$. Note that for both models the density does not converge towards $\rho_m$ as the state law is not linear.

As in Fig. 18, Fig. 21 emphasizes the effects induced by the Boussinesq assumption over mass and volume variations.

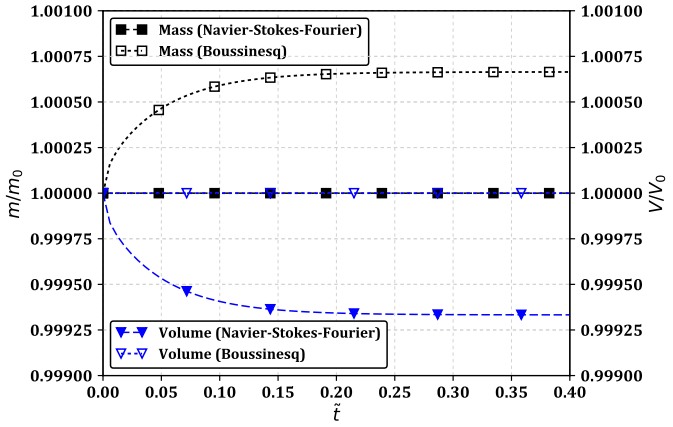


Figure 21: Thermal equilibrium: evolution of the mass ratio ($m/m_0$) and volume ratio ($V/V_0$) for the Navier-Stokes-Fourier and the Boussinesq models.

ass_vs_volume⟩

# 6 Conclusion

In this work, we have proposed and analyzed a finite-volume scheme to solve the Navier-Stokes-Fourier equations, which describe free-surface variable density flows. With any flux consistent with the semi-discrete in time Euler system, the proposed scheme is well-balanced and preserves the non-negativity of the water depth. In the case of a kinetic flux, a discrete entropy balance is proved for a flat topography. The numerical scheme is validated. The confrontation is made with results obtained with the simulation of the Boussinesq system. Notably, in a simple thermal diffusion case, the equilibrium temperature is not the same for the two systems.

A discrete entropy balance for a non-flat topography has yet to be obtained. Another challenge is the design and analysis of a numerical scheme for a system with a non-Newtonian rheology and in such a situation, we expect to be able to simulate complex interactions between the rheology terms and the temperature fluxes. Finally, simulations of the Navier-Stokes-Fourier system could be performed to investigate the propagation of internal waves in a stratified ocean.

# Acknowledgments

The authors acknowledge the Inria Project Lab "Algae in Silico" for its financial support. This research is also supported by the ERC SLIDEQUAKES ERC-CG-2013-PE10-617472.


# References


adduce [1] C. Adduce, G. Sciortino, and S. Proietti. Gravity Currents Produced by Lock Exchanges: Experiments and Simulations with a Two-Layer Shallow-Water Model with Entrainment. *J. Hydraul. Eng.*, 138:111–121, 2012.

art_3d [2] S. Allgeyer, M.-O. Bristeau, D. Froger, R. Hamouda, V. Jauzein, A. Mangeney, J. Sainte-Marie, F. Souillé, and M. Vallée. Numerical approximation of the 3d hydrostatic Navier-Stokes system with free surface. *ESAIM: M2AN*, 53(6):1981–2024, 2019. doi: https://doi.org/10.1051/m2an/2019044.

bristeau [3] E. Audusse and M.-O. Bristeau. A well-balanced positivity preserving second-order scheme for Shallow Water flows on unstructured meshes. *J. Comput. Phys.*, 206(1): 311–333, 2005.

bristeau1 [4] E. Audusse, F. Bouchut, M.-O. Bristeau, R. Klein, and B. Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for Shallow Water flows. *SIAM J. Sci. Comput.*, 25(6):2050–2065, 2004.

JSM_JCP [5] E. Audusse, M.-O. Bristeau, M. Pelanti, and J. Sainte-Marie. Approximation of the hydrostatic Navier-Stokes system for density stratified flows by a multilayer model. Kinetic interpretation and numerical validation. *J. Comp. Phys.*, 230:3453–3478, 2011. doi: 10.1016/j.jcp.2011.01.042.

JSM_entro [6] E. Audusse, F. Bouchut, M.-O. Bristeau, and J. Sainte-Marie. Kinetic entropy inequality and hydrostatic reconstruction scheme for the Saint-Venant system. *Math. Comp.*, 85(302):2815–2837, 2016. ISSN 0025-5718. doi: 10.1090/mcom/3099. URL `http://dx.doi.org/10.1090/mcom/3099`.

kin_entro [7] E. Audusse, M.-O. Bristeau, and J. Sainte-Marie. Kinetic entropy for layer-averaged hydrostatic Navier-Stokes equations. *submitted*, 2018.

nsf_partI [8] L. Boittin, F. Bouchut, M.-O. Bristeau, A. Mangeney, J. Sainte-Marie, and F. Souillé. Low-Mach type approximation of the Navier-Stokes system with temperature and salinity for free surface flows. working paper or preprint, Mar. 2020. URL `https://hal.inria.fr/hal-02510711`.

uchut_book [9] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources.* Birkhäuser, 2004. ISBN 3764366656.

nal_euler [10] M.-O. Bristeau, B. Di Martino, A. Mangeney, J. Sainte-Marie, and F. Souillé. Various analytical solutions for the incompressible Euler and Navier-Stokes systems with free surface. Accepted for publication in J. Fluid Mechanics, 2020. URL https://hal.archives-ouvertes.fr/hal-01831622.

e_telemac [11] A. Decoene and J.-F. Gerbeau. Sigma transformation and ALE formulation for three-dimensional free surface flows. *Internat. J. Numer. Methods Fluids*, 59(4):357–386, 2009.

ale:2004 [12] J. Donea, A. Huerta, J.-P. Ponthot, and A. Rodríguez-Ferran. *Encyclopedia of Computational Mechanics*, volume 1, chapter 14, Arbitrary Lagrangian–Eulerian Methods. John Wiley and Sons Ltd., 2004.

eshkiss3d [13] Freshkiss3d. home page. https://freshkiss3d.gitlabpages.inria.fr/freshkiss3d/applications.html, 2020.

rbin:2016 [14] D. Grapsas, R. Herbin, W. Kheriji, and J.-C. Latché. An unconditionally stable staggered pressure correction scheme for the compressible Navier-Stokes equations. *SMAI Journal of Computational Mathematics*, 2:51–97, 2016.

griffies [15] S. M. Griffies, C. Böning, F. Bryan, E. Chassignet, R. Gerdes, H. Hasumi, A. Hirst, A.-M. Treguier, and D. Webb. Developments in ocean climate modelling. *Ocean Modelling*, 2(3):123 – 192, 2000. ISSN 1463-5003. doi: https://doi.org/10.1016/S1463-5003(00)00014-7. URL http://www.sciencedirect.com/science/article/pii/S1463500300000147.

interlace [16] S. Hwang. Cauchy's interlace theorem for eigenvalues of hermitian matrices. *The American Mathematical Monthly*, 111(2):157–159, 2004. ISSN 00029890, 19300972. URL http://www.jstor.org/stable/4145217.

rrouturou [17] B. Larrouturou. How to preserve the mass fractions positivity when computing compressible multi-component flows. *J. Comp. Phys.*, 95(1):59 – 84, 1991. ISSN 0021-9991. doi: https://doi.org/10.1016/0021-9991(91)90253-H. URL http://www.sciencedirect.com/science/article/pii/002199919190253H.

simeoni [18] B. Perthame and C. Simeoni. A kinetic scheme for the Saint-Venant system with a source term. *Calcolo*, 38(4):201–231, 2001.

song:2006 [19] T. T. Song and T. Y. Hou. Parametric vertical coordinate formulation for multiscale, Boussinesq, and non-Boussinesq ocean modeling. *Ocean Modelling*, 11:298–332, 2006.

-01691949 [20] F. Souillé, M.-O. Bristeau, and J. Sainte-Marie. Remontées de chlorures dans la Vilaine. Research report, Inria Paris, Dec. 2017. URL https://hal.inria.fr/hal-01691949.

thacker [21] W. C. Thacker. Some exact solutions to the non-linear shallow-water wave equations. *J. Fluid Mech.*, 107:499–508, 1981.