



# On Thermodynamically Compatible Finite Volume Methods and Path-Conservative ADER Discontinuous Galerkin Schemes for Turbulent Shallow Water Flows

Saray Busto<sup>1</sup> · Michael Dumbser<sup>1</sup> · Sergey Gavriluk<sup>2,3</sup> · Kseniya Ivanova<sup>4</sup>

Received: 14 February 2021 / Revised: 29 April 2021 / Accepted: 8 May 2021 / Published online: 12 June 2021  
© The Author(s) 2021

## Abstract

In this paper we propose a new reformulation of the first order hyperbolic model for unsteady turbulent shallow water flows recently proposed in Gavriluk et al. (J Comput Phys 366:252–280, 2018). The novelty of the formulation forwarded here is the use of a new evolution variable that guarantees the trace of the discrete Reynolds stress tensor to be always non-negative. The mathematical model is particularly challenging because one important subset of evolution equations is nonconservative and the nonconservative products also act across genuinely nonlinear fields. Therefore, in this paper we first consider a thermodynamically compatible *viscous extension* of the model that is necessary to define a proper vanishing viscosity limit of the inviscid model and that is absolutely fundamental for the subsequent construction of a thermodynamically compatible numerical scheme. We then introduce two different, but related, families of numerical methods for its solution. The first scheme is a provably *thermodynamically compatible* semi-discrete finite volume scheme that makes direct use of the *Godunov form* of the equations and can therefore be called a *discrete Godunov formalism*. The new method mimics the underlying continuous viscous system *exactly* at the semi-discrete level and is thus consistent with the conservation of total energy, with the entropy inequality and with the vanishing viscosity limit of the model. The second scheme is a general purpose high order path-conservative ADER discontinuous Galerkin finite element method with a posteriori subcell finite volume limiter that can be applied to the inviscid as well as to the viscous form of the model. Both schemes have in common that they make use of path integrals to define the jump terms at the element interfaces. The different numerical methods are applied to the inviscid system and are compared with each other and with the scheme proposed in Gavriluk et al. (2018) on the example of three

---

✉ Saray Busto  
saray.busto@unitn.it

✉ Michael Dumbser  
michael.dumbser@unitn.it

<sup>1</sup> Department of Civil, Environmental and Mechanical Engineering, University of Trento, Via Mesiano 77, 38123 Trento, Italy

<sup>2</sup> Aix-Marseille Univ and CNRS UMR 7343 IUSTI, 5 rue Enrico Fermi, 13453 Marseille, France

<sup>3</sup> Lavrentyev Institute of Hydrodynamics, 15 Lavrentyev Ave., Novosibirsk, Russia 630090

<sup>4</sup> WSL-Institut für Schnee- und Lawinenforschung SLF, Flüelastrasse 11, 7260 Davos Dorf, Switzerland

Riemann problems. Moreover, we make the comparison with a fully resolved solution of the underlying viscous system with small viscosity parameter (vanishing viscosity limit). In all cases an excellent agreement between the different schemes is achieved. We furthermore show numerical convergence rates of ADER-DG schemes up to sixth order in space and time and also present two challenging test problems for the model where we also compare with available experimental data.

**Keywords** Godunov form of hyperbolic equations · Discrete Godunov formalism · Thermodynamically compatible finite volume schemes · Vanishing viscosity limit · Path-conservative ADER discontinuous Galerkin schemes · Unsteady turbulent shallow water flows · Realizable Hyperbolic turbulence model

## 1 Introduction

In the last decades, a lot of work has been devoted to the study of shallow water flows. When dispersive effects are negligible (this is the case, for example, for the modelling of hydraulic jumps for large Froude numbers, or tsunami waves), one usually employs the classical Saint-Venant (SV) or shallow water equations. The underlying hypothesis in the derivation of the Saint-Venant equations is the flow potentiality. The horizontal vorticity (parallel to the bottom) in the shallow water approximation is related with the horizontal velocity shear:  $\omega_{||} \approx \mathbf{V}_z$ , where  $\mathbf{V}$  is the instantaneous (non-averaged) horizontal velocity, and the index  $z$  means the derivative in the vertical direction. The absence of the vorticity means the absence of the horizontal velocity shear. The shallow water equations are hyperbolic, see e.g. [116]. When discontinuous solutions to the SV equations are studied, one uses the conservation of mass and momentum at the shocks, but the energy equation is always used as the entropy inequality. The reason for this is the following. The fluid flow ahead of the jump front is supercritical with respect to the front and almost potential, while behind the front it is highly turbulent: large vortex structures are usually formed. The energy is not conserved because a part of this energy is transformed into the energy of vortexes which is not taken into account in the SV model. An ideal model of free surface shallow flows which takes into account shear effects was recently derived by Teshukov [112]. The governing equations are obtained by depth averaging of the multi-dimensional Euler equations [100,101,112]. The hypothesis of smallness of the horizontal vorticity (the hypothesis of weakly sheared flows) allows us to keep the second order depth averaged correlations in the governing equations but neglect the third order correlations, and thus to close the governing system in the dissipationless limit. To apply the model to the study of real flows (formation of roll waves and hydraulic jumps) the model was complemented by dissipative source terms, see [69,80,100,101].

The corresponding multi-dimensional model of shear shallow water flows is a hyperbolic system of equations which is reminiscent of the Reynolds-averaged Euler equations for barotropic compressible turbulent flows. The model has three families of characteristics corresponding to the propagation of surface waves, shear waves and waves propagating with the average flow velocity. The main difficulty in studying such a system is the highly non-conservative nature of the governing equations: for six unknowns (the fluid depth, two components of the depth averaged horizontal velocity, and three independent components of the symmetric Reynolds stress tensor) one has only five conservation laws: conservation of mass, momentum, energy and mathematical “entropy”. The last one determines the evolution of the determinant of the Reynolds stress tensor. The non-conservative nature of

the multi-dimensional equations represents an enormous difficulty from the mathematical and numerical point of view. The definition and computation of discontinuous solutions for non-conservative hyperbolic equations is a challenging problem, see e.g. [5,23,85]. A numerical method (based on a splitting procedure) was recently developed for solving this non-conservative system [69,80]. The essential ingredient was the use of the energy conservation. It allowed, in particular, creation of vorticity once the jump appears. The splitting procedure was as follows. First, a geometric splitting was applied consisting in solving the governing equations first in  $x$  and then in  $y$  direction. Second, each one-dimensional system was also split into two subsystems, each of which contained only one ‘sound’ speed: the velocity of surface waves for the first sub-system, and the velocity of shear waves for the second sub-system. Each subsystem admitted its own energy conservation law, and its own ‘entropy’. However, such an operator splitting could be also a source of numerical errors. This is why it is very important to develop also different numerical methods for solving this challenging non-conservative system, like the two new unsplit schemes proposed in this paper, namely a completely new thermodynamically compatible unsplit finite volume scheme, as well as a slightly modified general-purpose high order ADER discontinuous Galerkin finite element method.

In order to put the new unsplit thermodynamically compatible finite volume scheme presented in this paper into the proper context, let us briefly review the main ideas on which it is based. Exactly sixty years ago, in 1961, Godunov published his groundbreaking paper *An interesting class of quasilinear systems* [70], in which he discovered the connection between *symmetric hyperbolicity* in the sense of Friedrichs [62] and *thermodynamic compatibility* (SHTC), ten years before Friedrichs and Lax [63], who independently rediscovered the same connection again in 1971. In a subsequent series of papers, Godunov and Romenski carried out further research on this link between symmetric hyperbolicity and thermodynamic compatibility and generalized the seminal idea of Godunov to the more general SHTC framework of symmetric hyperbolic and thermodynamically compatible systems, which includes not only the compressible Euler equations of gasdynamics, but also the magnetohydrodynamics (MHD) equations [71] and the equations of nonlinear hyperelasticity [74,75,77]. The findings of Godunov and Romenski on nonlinear hyperelasticity were subsequently further employed and extended in [1,6,14,17,47,57,59,67,72,81,87,88,91,93]. A very general class of symmetric hyperbolic and thermodynamically compatible systems was presented by Romenski in [102], which is able to describe the interaction of moving-dielectric solids with electromagnetic fields, the dynamics of superfluid helium and also contains a hyperbolic model for heat conduction. An extension of this class of models to compressible multi-phase flows was forwarded in [103,104,106]. The SHTC framework remains valid even in the context of special and general relativity, see [76,105]. Recently, a connection between the class of symmetric hyperbolic and thermodynamically compatible systems and Hamiltonian mechanics was rigorously established in [92]. SHTC systems go also beyond classical continuum mechanics, see e.g. [94] for an SHTC formulation of continuum mechanics with torsion. Despite the mathematical beauty and rigor of the SHTC framework, up to now it was never carried over to the discrete level. So far, most papers on thermodynamically compatible schemes are based on the ideas of the seminal work of Tadmor [108], in which a discrete extra conservation law for the *entropy* is obtained as a consequence of the discretization of all other equations (including the energy conservation, which is explicitly discretized). Instead, in the new scheme presented in this paper we are *not* discretizing the energy equation explicitly, but are rather looking for a thermodynamically compatible scheme in which a discrete *total energy* conservation law is obtained as direct *consequence* of the compatible discretization of all the other equations. For an interesting application of entropy compatible schemes for

the discretization of non-conservative equations, see [3,60]. The ideas presented there are related to the new compatible scheme introduced in the present paper, though [3,60] deal with much simpler equation systems. Recently, convergence of entropy-stable schemes was proven in [26]. For extensions of entropy-compatible schemes to high order discontinuous Galerkin methods, see [28,36,66,78,84] and references therein. While most of the aforementioned schemes are thermodynamically compatible only at the semi-discrete level, a fully discrete entropy-stable scheme has been recently presented in [95]. We also would like to point out that a very general framework for the construction of numerical schemes satisfying additional extra conservation laws has been recently forwarded by Abgrall in [2].

As already stated above, the *major difference* of the thermodynamically compatible scheme proposed in this paper with respect to previous thermodynamically compatible schemes is its discrete compatibility with the conservation of total energy as a consequence of all equations and *not* the conservation of entropy as a consequence. In other words, the thermodynamically compatible finite volume scheme presented in this paper *never* explicitly discretizes the energy equation, but total energy conservation is obtained as a mere consequence of a thermodynamically compatible discretization of the other equations, including a compatible discretization of the numerical viscosity.

The second unsplit scheme proposed in this paper is a fully-discrete one-step high order ADER discontinuous Galerkin method (ADER-DG). Explicit discontinuous Galerkin schemes for hyperbolic equations have been put forward by Reed and Hill in [96] introducing the use of discontinuous polynomials in a Galerkin framework to allow the jump of the discrete solution across cell boundaries.

Then, the first extensions to multidimensional and non linear hyperbolic systems were presented in the series of papers by Cockburn and Shu [27,30–33]. Parabolic terms have been considered for the first time in [9,10,34,35]. The severe time step restriction induced by the inclusion of higher order derivatives, [83,122,123], and nonlinear dispersive equations, [51,53,54], has driven to the development of fully implicit approaches, [44], whose major disadvantage is the solution of the resulting ill-conditioned algebraic systems. An alternative approach recently proposed is the use of hyperbolic reformulations of dispersive models which allow for more efficient discretizations, [7,8,18,52].

Regarding high order methods, it is important to remark that while attaining high order in space is straightforward for DG methodologies, there are different possibilities concerning high order time discretizations. The original DG schemes of Cockburn and Shu employed high order Runge-Kutta schemes in time, leading to the family of RKDG schemes. An alternative consists in the family of fully implicit and semi-implicit space-time DG methods, see e.g. [19,82,98,99,109–111,120,121]. Another different option that leads to high order explicit fully-discrete one-step schemes, and which is followed in this paper, combines ideas of the ADER approach of Toro and Titarev, originally developed within the finite volume framework [20,114,117,118], with space-time DG methods. This methodology, based on the ideas outlined in [41,43], makes use of an element-local space-time DG predictor, thus avoiding the cumbersome Cauchy-Kovalevskaya procedure of classical ADER schemes and thus allowing also the solution of complex PDEs in multiple space dimensions. Some examples of the wide range of applicability for this approach include the compressible Euler and Navier-Stokes equations, [40,41], compressible multi-phase flows [45], the Godunov-Peshkov-Romenski model of continuum mechanics, [17,47,93]. Discontinuous Galerkin schemes for hyperbolic PDE systems with non-conservative products have been proposed for the first time in [42,97], based on the ideas of path conservative schemes [22–24,85,89], which will be also a key point for the development of the numerical schemes proposed in this paper.

The rest of this paper is organized as follows: in Sect. 2 we present the original model [69] and a novel reformulation based on a decomposition of the specific Reynolds stress tensor  $\mathbf{P}$  as  $\mathbf{P} = \mathbf{Q}\mathbf{Q}^T$ . We furthermore introduce a viscous extension of the governing PDE system in order to define a rigorous and thermodynamically compatible vanishing viscosity limit of the model. We finally recall the Godunov formalism of thermodynamically compatible systems and prove that the proposed viscous system is thermodynamically compatible with the energy conservation law and with the entropy inequality. In Sect. 3 we present a novel thermodynamically compatible finite volume scheme, which mimics the aforementioned viscous extension of the system *exactly* at the semi-discrete level. In Sect. 4 a high order ADER discontinuous Galerkin method with a posteriori subcell limiter (MOOD) is presented for the new reformulation of the model proposed in this paper, including its viscous extension. Special care is taken concerning the conservation of total energy. Numerical results are shown in Sect. 5, where first a numerical convergence study is presented for third to sixth order ADER-DG schemes in space and time; subsequently, different schemes are compared with each other for three Riemann problems, discussing in particular the discretization of the non-conservative terms of the model in the context of thermodynamically compatible systems. The end of Sect. 5 contains numerical results for some challenging test problems for which experimental reference data are available, such as supercritical roll waves and the circular shock instability developing in the SWASI experiment, see [61]. The paper is rounded-off with some concluding remarks and an outlook to future work in Sect. 6.

## 2 Governing Equations

We consider the following *overdetermined* hyperbolic model for turbulent shear shallow water flows in multiple space dimensions, which has been recently proposed in [69] and which was also applied and studied in [11,25,80]:

$$\partial_t h + \nabla \cdot (h\mathbf{v}) = 0, \quad (1)$$

$$\partial_t (h\mathbf{v}) + \nabla \cdot \left( h\mathbf{v} \otimes \mathbf{v} + \frac{1}{2}gh^2\mathbf{I} + h\mathbf{P} \right) + gh\nabla b = -C_f \|\mathbf{v}\| \mathbf{v}, \quad (2)$$

$$\partial_t \mathbf{P} + \mathbf{v} \cdot \nabla \mathbf{P} + \nabla \mathbf{v} \mathbf{P} + \mathbf{P} \nabla \mathbf{v}^T = -2\frac{\alpha}{h} \mathbf{P}, \quad (3)$$

$$\partial_t b = 0, \quad (4)$$

with the gravity constant  $g$ . The physical (primitive) state variables in (1)–(4) are the following:  $h = h(\mathbf{x}, t)$  is the water depth,  $b = b(\mathbf{x})$  is the known bottom topography,  $\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$  is the depth-averaged flow velocity and  $\mathbf{P} = \mathbf{P}(\mathbf{x}, t)$  is the specific Reynolds stress tensor. For shallow water systems it is convenient to include the stationary bottom profile  $b(\mathbf{x})$  in the set of state variables. The reason is that this allows to represent stationary free surface waves associated with bottom jumps and to obtain well-balanced numerical schemes, see e.g. [21,22,64,89,90] for more details. Via straightforward calculations it can be shown that the system (1)–(4) admits the following extra conservation law

$$\partial_t (hE) + \nabla \cdot \left( \mathbf{v}(hE) + \left( \frac{1}{2}gh^2\mathbf{I} + h\mathbf{P} \right) \mathbf{v} \right) = -C_f \|\mathbf{v}\|^3 - \alpha \operatorname{tr} \mathbf{P}, \quad (5)$$

with the total energy defined as  $hE = \frac{1}{2}gh^2 + hgb + \frac{1}{2}h\|\mathbf{v}\|^2 + \frac{1}{2}h \operatorname{tr} \mathbf{P}$ .

The bottom friction is taken into account by a coefficient  $C_f$  and the dissipation function  $\alpha$  is given according to [69] as

$$\alpha = \max \left( 0, C_f \frac{\text{tr} \mathbf{P} - \varphi h^2}{(\text{tr} \mathbf{P})^2} \|\mathbf{v}\|^3 \right). \quad (6)$$

## 2.1 Reformulation of the Model in Terms of a New Evolution Variable

The above model requires  $\text{tr} \mathbf{P} \geq 0$  for hyperbolicity. In order to guarantee this property also at the *discrete level* for all times, we propose the following novel *reformulation* of the system (1)–(5). For this, we consider first the homogeneous part of equation (3) for the symmetric tensor  $\mathbf{P}$ :

$$\dot{\mathbf{P}} + \mathbf{L}\mathbf{P} + \mathbf{P}\mathbf{L}^T = 0. \quad (7)$$

Here for shortness, for any  $f$ ,  $\dot{f}$  means the material time derivative:  $\dot{f} = f_t + \mathbf{v} \cdot \nabla$ , and  $\mathbf{L} = \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \nabla \mathbf{v}$ . Let us replace  $\mathbf{P}$  by  $\mathbf{P} = \mathbf{Q}\mathbf{Q}^T$ . What is the equation for  $\mathbf{Q}$ ? One obtains from (7):

$$(\dot{\mathbf{Q}} + \mathbf{L}\mathbf{Q})\mathbf{Q}^T + \mathbf{Q}(\dot{\mathbf{Q}} + \mathbf{L}\mathbf{Q})^T = 0. \quad (8)$$

If

$$\dot{\mathbf{Q}} + \mathbf{L}\mathbf{Q} = \mathbf{B}(\mathbf{Q}^T)^{-1} \quad (9)$$

with an antisymmetric tensor  $\mathbf{B} = -\mathbf{B}^T$ , the equation for  $\mathbf{P}$  will be obviously satisfied. Thus, the equation for  $\mathbf{Q}$  is defined up to an antisymmetric tensor  $\mathbf{B}$  taking into account a proper rotation of the Reynolds tensor (for details, see [68]). We hypothesize that friction forces will drastically reduce the influence of this proper rotation, i.e. we take  $\mathbf{B} = 0$ . Such a class of solutions is not equivalent to all solutions governed by the equation for  $\mathbf{P}$ , but is able, as we will show, to describe complex flow configurations.

What is a geometrical sense of such a decomposition  $\mathbf{P} = \mathbf{Q}\mathbf{Q}^T$ ? Let us recall first the definition of the *Gram matrix*  $\mathbf{G}$  (in the 2D case). Consider two vectors  $\mathbf{w}_i$ ,  $i = 1, 2$ . The Gram matrix is defined as

$$\mathbf{G} = \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{w}_1 & \mathbf{w}_1 \cdot \mathbf{w}_2 \\ \mathbf{w}_1 \cdot \mathbf{w}_2 & \mathbf{w}_2 \cdot \mathbf{w}_2 \end{pmatrix}. \quad (10)$$

The ‘dot’ here is for the scalar product of vectors. It can be also written as

$$\mathbf{G} = \mathbf{Q}\mathbf{Q}^T \quad (11)$$

with

$$\mathbf{Q} = \begin{pmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \end{pmatrix}, \quad (12)$$

i.e. the line vectors  $\mathbf{w}_i$  are the lines of  $\mathbf{Q}$ . Let us recall that in our case  $\mathbf{P}$  is the correlation tensor expressed in terms of the velocity pulsations (see [112] for details) as:

$$\mathbf{P} = \begin{pmatrix} \overline{v_1'^2} & \overline{v_1'v_2'} \\ \overline{v_1'v_2'} & \overline{v_2'^2} \end{pmatrix} \quad (13)$$

Here the averaging operation, denoted by a “bar”, is the depth averaging. The tensor  $\mathbf{P}$  is positive definite due to the Cauchy–Schwarz inequality. Let us show that  $\mathbf{P}$  can be presented in the form (10), i.e. there exist vectors  $\mathbf{w}_i$ :

$$\mathbf{P} = \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{w}_1 & \mathbf{w}_1 \cdot \mathbf{w}_2 \\ \mathbf{w}_1 \cdot \mathbf{w}_2 & \mathbf{w}_2 \cdot \mathbf{w}_2 \end{pmatrix}$$

For this we take

$$\mathbf{w}_1 = \sqrt{v_1'^2} (\cos \theta_1, \sin \theta_1)^T, \quad \mathbf{w}_2 = \sqrt{v_2'^2} (\cos \theta_2, \sin \theta_2)^T$$

with

$$\cos(\theta_1 - \theta_2) = \frac{\overline{v_1' v_2'}}{\sqrt{v_1'^2 v_2'^2}}.$$

The last relation is well defined due to the Cauchy–Schwarz inequality:  $|\overline{v_1' v_2'}| \leq \sqrt{v_1'^2 v_2'^2}$ .

With  $P_{ik} = Q_{im} Q_{km}$  written in terms of the new evolution variable  $\mathbf{Q}$  and the notation  $\partial_m = \partial/\partial x_m$  the above system can be rewritten again as an overdetermined PDE system as follows:

$$\partial_t h + \partial_m (h v_m) = 0, \quad (14)$$

$$\partial_t (h v_i) + \partial_k \left( h v_i v_k + \frac{1}{2} g h^2 \delta_{ik} + h P_{ik} \right) + g h \partial_i b = -C_f \|\mathbf{v}\| v_i, \quad (15)$$

$$\partial_t Q_{ik} + v_m \partial_m Q_{ik} + (\partial_m v_i) Q_{mk} = -\frac{\alpha}{h} Q_{ik}, \quad (16)$$

$$\partial_t b = 0, \quad (17)$$

with the conservative evolution variables  $h = h(\mathbf{x}, t)$ ,  $h\mathbf{v} = h\mathbf{v}(\mathbf{x}, t)$ ,  $\mathbf{Q} = \mathbf{Q}(\mathbf{x}, t)$  and the stationary bottom profile  $b = b(\mathbf{x})$ .

It is easy to see that (3) is a consequence of (16) by simply multiplying (16) with  $\mathbf{Q}^T$  from the right and summing the transpose of (16) multiplied by  $\mathbf{Q}$  from the left. It can be easily checked that also the new system (14)–(17) admits an extra energy conservation law

$$\partial_t (hE) + \partial_i \left( (hE) v_i + \left( \frac{1}{2} g h^2 \delta_{ik} + h Q_{im} Q_{km} \right) v_k \right) = -C_f \|\mathbf{v}\|^3 - \alpha \operatorname{tr} \mathbf{P}, \quad (18)$$

which can be obtained as a consequence of (14)–(17). In terms of  $\mathbf{Q}$  the total energy reads  $hE = \frac{1}{2} g h^2 + h g b + \frac{1}{2} h v_i v_i + \frac{1}{2} h Q_{ij} Q_{ij}$ , which for flat bottom  $b = 0$  is a strictly convex function in the variables  $h, h v_i$  and  $S_{ij} = h Q_{ij}$ . It is also a convex function of  $(h, h v_i, Q_{ij})$ , if the turbulent energy is small compared to the gravitational potential energy (see “Appendix A” for details). Also note that due to  $\operatorname{tr} \mathbf{P} = Q_{ij} Q_{ij} \geq 0$  the use of  $\mathbf{Q}$  instead of  $\mathbf{P}$  automatically guarantees a non-negative trace of  $\mathbf{P}$  by construction, and hence also at the discrete level for all times. In this sense, system (14)–(18) is analogous to a so-called *realizable* turbulence model. At this point we emphasize that the thermodynamically compatible scheme proposed later in this paper will consider only the case of a flat bottom with  $b = 0$ .

Last but not least, we would like to point out the *difference* in the only apparently similar structure of PDE (16) and the governing PDE for the distortion field  $A_{ik}$  in nonlinear hyperelasticity [47, 93], which reads

$$\partial_t A_{ik} + v_m \partial_m A_{ik} + A_{im} (\partial_k v_m) = -\frac{1}{\theta(\tau)} E_{A_{ik}}. \quad (19)$$

As one can easily see, the order of the matrix-product in the third term on the left hand side of (16) and (19) is exchanged. It is well-known that for hyperelasticity there is an additional conservation law associated with the determinant of the distortion field  $A_{ik}$  and in the following we will show that the same applies to the determinant of the field  $Q_{ik}$ . The time derivative of the determinant of  $\mathbf{Q}$  can be easily obtained via the Jacobi formula, which expresses the derivatives of the determinant of a matrix in terms of the inverse of the matrix and the derivatives of the matrix itself:

$$\partial_t |Q| = |Q| Q_{ki}^{-1} \partial_t Q_{ik}, \quad \partial_m |Q| = |Q| Q_{ki}^{-1} \partial_m Q_{ik}, \quad (20)$$

where  $Q_{ki}^{-1}$  is a compact notation for  $(Q^{-1})_{ki}$ . Applying (20) to (16) yields

$$\partial_t |Q| + |Q| Q_{ki}^{-1} v_m \partial_m Q_{ik} + |Q| Q_{ki}^{-1} (\partial_m v_i) Q_{mk} = -\frac{\alpha}{h} |Q| Q_{ki}^{-1} Q_{ik}, \quad (21)$$

from which one obtains

$$\partial_t |Q| + v_m \partial_m |Q| + |Q| (\partial_m v_i) \delta_{mi} = -\frac{\alpha}{h} |Q| \delta_{kk}, \quad (22)$$

and therefore the sought additional balance law for the determinant  $|Q|$ ,

$$\partial_t |Q| + \partial_m (v_m |Q|) = -\frac{\alpha}{h} |Q| \delta_{kk}, \quad (23)$$

which for  $\alpha = 0$  has the same structure as the mass conservation equation (14). As such, we can assume that for  $h > 0$  also  $|Q| > 0$  holds.

Via straightforward calculations it can be shown that for smooth solutions the conservation law (23) for the determinant  $|Q|$  is equivalent to the conservation law

$$\partial_t (h\psi) + \partial_m (v_m h\psi) = -\frac{4\alpha}{h^3} (P_{11} P_{22} - P_{12}^2), \quad \psi = \frac{|\mathbf{P}|}{h^2} = \frac{|\mathbf{Q}\mathbf{Q}^T|}{h^2} \quad (24)$$

already found in [69]. Assuming  $\alpha = 0$  it reduces to

$$\partial_t (h\psi) + \partial_m (v_m h\psi) = 0. \quad (25)$$

## 2.2 Eigenstructure of the Reformulation

The eigenvalues of the homogeneous part of (14)–(17) in  $x_1$  direction are

$$\lambda_{1,7} = v_1 \mp c, \quad \lambda_{2,6} = v_1 \mp \sqrt{P_{11}}, \quad \lambda_{3,4,5} = v_1, \quad \lambda_8 = 0, \quad (26)$$

with  $c^2 = gh + 3P_{11}$ . The associated right eigenvectors read

$$\begin{aligned} \mathbf{r}_{1,7} &= (hK, h(v_1 \mp c)K, h(v_2 K \mp 6cP_{12}), Q_{11}K, Q_{12}K, 6Q_{11}P_{12}, 6Q_{12}P_{12}, 0)^T, \\ \mathbf{r}_{2,6} &= (0, 0, \mp\sqrt{P_{11}}h, 0, 0, Q_{11}, Q_{12}, 0)^T, \\ \mathbf{r}_3 &= (-2hQ_{11}, -2hv_1Q_{11}, -2hv_2Q_{11}, gh + P_{11}, 0, Q_{11}^{-1}(2Q_{12}|Q| + \Pi_1), 0, 0)^T, \\ \mathbf{r}_4 &= (-2hQ_{12}, -2hv_1Q_{12}, -2hv_2Q_{12}, gh + P_{11}, 0, Q_{11}^{-1}(2Q_{12}P_{12} - \Pi_2), 0, 0)^T, \\ \mathbf{r}_5 &= (0, 0, 0, 0, 0, -Q_{11}^{-1}Q_{12}, 1, 0)^T, \\ \mathbf{r}_8 &= (hM, 0, h(2v_1P_{12} + v_2M), Q_{11}M, Q_{12}M, -2P_{12}Q_{11}, -2P_{12}Q_{12}, \Pi_3)^T, \end{aligned} \quad (27)$$

with  $K = 2c^2 + gh$ ,  $M = P_{11} - v_1^2$ ,  $\Pi_1 = Q_{21}(P_{11} - gh)$ ,  $\Pi_2 = Q_{22}(P_{11} + gh)$  and  $\Pi_3 = g^{-1}M(u^2 - c^2)$ . All eigenvalues are real since  $h > 0$  and  $P_{11} = Q_{11}^2 + Q_{12}^2 \geq 0$  and there exists a full set of eigenvectors, hence the system is hyperbolic.

## 2.3 The Godunov Form of Nonlinear Systems of Hyperbolic Conservation Laws

In order to define the vanishing viscosity limit of system (14)–(17) and in order to introduce the new thermodynamically compatible finite volume schemes developed later in this paper, which are exactly compatible with the vanishing viscosity limit, it is necessary to recall the Godunov form [70] of hyperbolic PDE systems. We first consider only hyperbolic systems of conservation laws in two space dimensions of the type

$$\mathbf{q}_t + \partial_k \mathbf{f}_k = 0, \quad (28)$$

with flux tensor  $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2)$ , that admit the following parametrization according to Godunov [70]

$$(L\mathbf{p})_t + \partial_k ((v_k L)\mathbf{p}) = 0, \quad (29)$$

with the extra conservation law of the form

$$\mathcal{E}_t + \partial_k F_k = 0, \quad (30)$$

where  $F_k$  is the total energy flux in the  $k$ -th coordinate direction. Equations (29) and (30) are in the following called the *Godunov form* of the conservation law (28) and constitute an *overdetermined* system of PDE. The system is *thermodynamically compatible* if the following relations hold:

$$\mathbf{q} = L\mathbf{p}, \quad \mathbf{p} = \mathcal{E}_{\mathbf{q}}, \quad \mathbf{f}_k = (v_k L)\mathbf{p}, \quad F_k = \mathbf{p} \cdot \mathbf{f}_k - v_k L. \quad (31)$$

Here,  $L$  is the so-called *generating potential* and  $\mathcal{E}$  is the total energy density, which are the Legendre transforms of each other and thus satisfy

$$L = \mathbf{p} \cdot \mathbf{q} - \mathcal{E}, \quad \mathcal{E} = \mathbf{p} \cdot \mathbf{q} - L. \quad (32)$$

We assume  $L$  and  $\mathcal{E}$  to be strictly convex functions of their arguments, hence the transformation matrices between  $\mathbf{p}$  and  $\mathbf{q}$  variables, which are the Hessian matrices of  $L$  and  $\mathcal{E}$ , respectively, verify

$$\frac{\partial \mathbf{p}}{\partial \mathbf{q}} = \mathcal{E}_{\mathbf{q}\mathbf{q}} > 0, \quad \frac{\partial \mathbf{q}}{\partial \mathbf{p}} = L_{\mathbf{p}\mathbf{p}} > 0, \quad L_{\mathbf{p}\mathbf{p}} = (\mathcal{E}_{\mathbf{q}\mathbf{q}})^{-1}, \quad (33)$$

$$L_{\mathbf{p}\mathbf{p}} = L_{\mathbf{p}\mathbf{p}}^T, \quad \mathcal{E}_{\mathbf{q}\mathbf{q}} = \mathcal{E}_{\mathbf{q}\mathbf{q}}^T. \quad (34)$$

It is easy to check that (30) is a consequence of (29), since scalar multiplication of (29) with  $\mathbf{p} = \mathcal{E}_{\mathbf{q}}$  yields

$$\begin{aligned} \mathbf{p} \cdot (L\mathbf{p})_t + \mathbf{p} \cdot \partial_k \mathbf{f}_k &= \mathcal{E}_t + \partial_k (\mathbf{p} \cdot \mathbf{f}_k) - (\partial_k \mathbf{p}) \cdot \mathbf{f}_k \\ &= \mathcal{E}_t + \partial_k (\mathbf{p} \cdot \mathbf{f}_k) - \partial_k \mathbf{p} \cdot (v_k L)\mathbf{p} \\ &= \mathcal{E}_t + \partial_k (\mathbf{p} \cdot \mathbf{f}_k) - \partial_k (v_k L), \\ &= \mathcal{E}_t + \partial_k F_k = 0, \end{aligned} \quad (35)$$

which is the sought form of the total energy conservation law (30). For details on the class of symmetric hyperbolic and thermodynamically compatible (SHTC) systems and their application, see [17,47,70,71,73–75,93,102]. The shallow water subsystem for flat bottom

$$\partial_t h + \partial_k (h v_k) = 0, \quad (36)$$

$$\partial_t (h v_i) + \partial_k \left( h v_i v_k + \frac{1}{2} g h^2 \delta_{ik} \right) = 0, \quad (37)$$

$$\partial_t \mathcal{E} + \partial_k \left( \mathcal{E} v_k + \frac{1}{2} g h^2 v_k \right) = 0, \quad (38)$$

contained in (14)–(18) falls into the class of PDE (28)–(30). The corresponding potentials are

$$\mathcal{E} = \frac{1}{2} g q_1^2 + \frac{1}{2} \frac{q_2^2 + q_3^2}{q_1} \quad (39)$$

and

$$L = \frac{1}{2g} \left( p_1 + \frac{1}{2} (p_2^2 + p_3^2) \right)^2, \quad (40)$$

with the vectors  $\mathbf{q} = (h, h v_1, h v_2)^T$  and  $\mathbf{p} = (gh - \frac{1}{2}(v_1^2 + v_2^2), v_1, v_2)^T$ . The associated Hessian matrices are

$$\mathcal{E}_{\mathbf{q}\mathbf{q}} = \frac{1}{h} \begin{pmatrix} gh + v_1^2 + v_2^2 & -v_1 & -v_2 \\ -v_1 & 1 & 0 \\ -v_2 & 0 & 1 \end{pmatrix} \quad (41)$$

and

$$L_{\mathbf{p}\mathbf{p}} = \frac{1}{g} \begin{pmatrix} 1 & v_1 & v_2 \\ v_1 & gh + v_1^2 & v_1 v_2 \\ v_2 & v_1 v_2 & gh + v_2^2 \end{pmatrix}. \quad (42)$$

It is easy to see that with (40) and the flux tensor  $\mathbf{F} = (h v_k, h v_i v_k + \frac{1}{2} g h^2 \delta_{ik})^T$  the energy fluxes (31) in (30) are

$$F_k = \mathbf{p} \cdot \mathbf{f}_k - v_k L = \mathcal{E} v_k + \frac{1}{2} g h^2 v_k, \quad (43)$$

which corresponds to the energy flux in (38).

## 2.4 Thermodynamically Compatible Vanishing Viscosity Limit

In order to define weak solutions for system (14)–(17), we define an associated *thermodynamically compatible viscous system* that satisfies at the same time an entropy-type inequality, as well as the total energy conservation law. In this section we assume a flat bottom with  $b = 0$  for simplicity, as well as  $\alpha = C_f = 0$ , while a small parabolic dissipation term with dissipation coefficient  $\varepsilon > 0$  is added to the equations. In order to guarantee exact total energy conservation, a non-negative production term  $T_{ik}$  must be added to the governing PDE for  $\mathbf{Q}$ :

$$\partial_t h + \partial_m(hv_m) = \partial_m \varepsilon \partial_m h, \quad (44)$$

$$\partial_t(hv_i) + \partial_k \left( hv_i v_k + \frac{1}{2} g h^2 \delta_{ik} + h P_{ik} \right) = \partial_m \varepsilon \partial_m(hv_i), \quad (45)$$

$$\partial_t Q_{ik} + v_m \partial_m Q_{ik} + (\partial_m v_i) Q_{mk} = \partial_m \varepsilon \partial_m Q_{ik} + T_{ik}, \quad (46)$$

$$\partial_t \mathcal{E} + \partial_i \left( (\mathcal{E}_1 + \mathcal{E}_2) v_i + \left( \frac{1}{2} g h^2 \delta_{ik} + h P_{ik} \right) v_k \right) = \partial_m \varepsilon \partial_m \mathcal{E}, \quad (47)$$

with the total energy  $\mathcal{E} = hE = \mathcal{E}_1 + \mathcal{E}_2$  that can be decomposed into two contributions with  $\mathcal{E}_1 = hE_1 = \frac{1}{2} g h^2 + \frac{1}{2} h v_i v_i$  and  $\mathcal{E}_2 = hE_2 = \frac{1}{2} h Q_{ik} Q_{ik}$ . Here,  $\mathcal{E}_1$  is the total energy potential of the shallow water subsystem (36)–(37) and  $\mathcal{E}_2$  is the total energy associated with the new object  $Q_{ik}$ . In what follows, we will denote the inviscid part of the total energy flux in (47) by

$$F = (\mathcal{E}_1 + \mathcal{E}_2) v_i + \left( \frac{1}{2} g h^2 \delta_{ik} + h P_{ik} \right) v_k = F_G + \mathcal{E}_2 v_i + h P_{ik} v_k, \quad (48)$$

with the abbreviation

$$F_G = \mathcal{E}_1 v_i + \frac{1}{2} g h^2 v_i \quad (49)$$

that will be used later and which corresponds to the energy flux related to the shallow water subsystem, see also (43).

The production term,  $T_{ik}$ , which is needed to achieve the consistency of (44)–(46) with the total energy conservation law (47) reads

$$T_{ik} = \varepsilon \frac{Q_{ik}}{h \operatorname{tr} \mathbf{P}} \partial_m q_i (\mathcal{E}_{q_i q_j}) \partial_m q_j. \quad (50)$$

The consistency with physics and experimental observations requires total energy conservation, see [68, 69, 80, 100, 101] for a more detailed discussion. In (50) the vector  $\mathbf{q} = q_i = (h, hv_i, Q_{ik})$  indicates the vector of primary state variables and  $\mathcal{E}_{q_i q_j}$  is the Hessian matrix of the total energy potential with respect to these state variables. One can show that the Hessian matrix is positive definite for small turbulent kinetic energy  $Q_{ij} Q_{ij}$  compared to  $gh$ , see “Appendix A” for details.

**Theorem 1** (Energy conservation) *The energy conservation law (47) is a consequence of equations (44)–(46).*

**Proof** The shallow water subsystem (36)–(37) related to  $\mathcal{E}_1$ , which are the black terms in (44)–(45), directly falls into the general class of PDE (29)–(30) found by Godunov, hence the compatibility of the shallow water subsystem with the energy conservation law with energy potential  $\mathcal{E}_1$  is obvious. It is therefore enough to consider only the remaining terms associated with the quantity  $Q_{ik}$  (red) and the viscous terms on the right hand side (blue).

We first show compatibility of the red terms: Since  $(\mathcal{E}_2)_h = E_2 = \frac{1}{2} Q_{ik} Q_{ik} = \frac{1}{2} \operatorname{tr} \mathbf{P}$ ,  $\mathcal{E}_{h v_i} = v_i$ ,  $\mathcal{E}_{Q_{ik}} = (\mathcal{E}_2)_{Q_{ik}} = h (E_2)_{Q_{ik}} = h Q_{ik}$  summation of (44)–(46) with the thermodynamic dual variables and considering only new contributions that are not yet contained in the Godunov-form yields

$$\begin{aligned}
& E_2 (\partial_t h + \partial_m (h v_m)) + v_i \partial_k (h P_{ik}) + h Q_{ik} (\partial_t Q_{ik} + v_m \partial_m Q_{ik} + (\partial_m v_i) Q_{mk}) \\
& = E_2 \partial_t h + h \partial_t \left( \frac{1}{2} Q_{ik} Q_{ik} \right) + E_2 \partial_m (h v_m) + h v_m \partial_m \left( \frac{1}{2} Q_{ik} Q_{ik} \right) \\
& \quad + v_i \partial_k (h Q_{im} Q_{km}) + h Q_{ik} Q_{mk} \partial_m v_i \\
& = \partial_t (h E_2) + \partial_m (h v_m E_2) + \partial_k (v_i h Q_{im} Q_{km}) \\
& = \partial_t (h E_2) + \partial_m (v_m \mathcal{E}_2) + \partial_k (v_i h P_{ik}) .
\end{aligned} \tag{51}$$

After simple renaming of indices this proves the thermodynamic compatibility of the red terms contained in the left hand side of (44)–(46) with the red terms on the left hand side of the energy equation (47).

We now consider the right hand side (blue terms): We define a viscous flux tensor  $\mathbf{g}_k$  as

$$\mathbf{g}_m = \varepsilon \partial_m \mathbf{q} \tag{52}$$

and a production term  $\mathbf{T}$  that is equal to zero for all PDE apart from the non-zero production term  $T_{ik}$  in the PDE for  $Q_{ik}$ , see (46) and (50). Summation of the right hand sides of (44)–(46) with the thermodynamic dual variables  $\mathbf{p} = \mathcal{E}_q$  yields

$$\begin{aligned}
\mathcal{E}_q \cdot \partial_m \mathbf{g}_m + \mathcal{E}_q \cdot \mathbf{T} &= \mathcal{E}_q \cdot \partial_m \varepsilon \partial_m \mathbf{q} + \mathcal{E}_q \cdot \mathbf{T} \\
&= \partial_m (\varepsilon \mathcal{E}_q \cdot \partial_m \mathbf{q}) - \varepsilon \partial_m \mathcal{E}_q \cdot \partial_m \mathbf{q} + \mathcal{E}_q \cdot \mathbf{T} \\
&= \partial_m \varepsilon \partial_m \mathcal{E} - \varepsilon (\mathcal{E}_{qq} \partial_m \mathbf{q}) \cdot \partial_m \mathbf{q} + \mathcal{E}_q \cdot \mathbf{T} \\
&= \partial_m \varepsilon \partial_m \mathcal{E} - \varepsilon \partial_m q_i (\mathcal{E}_{q_i q_j}) \partial_m q_j + \mathcal{E}_q \cdot \mathbf{T} \\
&= \partial_m \varepsilon \partial_m \mathcal{E},
\end{aligned} \tag{53}$$

where  $-\varepsilon \partial_m q_i (\mathcal{E}_{q_i q_j}) \partial_m q_j + \mathcal{E}_q \cdot \mathbf{T} = 0$  since  $\mathcal{E}_{Q_{ik}} = h Q_{ik}$  and

$$\mathcal{E}_q \cdot \mathbf{T} = h Q_{ik} T_{ik} = \varepsilon \frac{h Q_{ik} Q_{ik}}{h \operatorname{tr} \mathbf{P}} \partial_m q_i (\mathcal{E}_{q_i q_j}) \partial_m q_j = \varepsilon \partial_m q_i (\mathcal{E}_{q_i q_j}) \partial_m q_j. \tag{54}$$

The combination of the right hand sides of (44)–(46) therefore yields the right hand side of (47), which completes the proof.  $\square$

**Theorem 2** (Entropy-type inequality) *A direct consequence of the PDE (46) without the parabolic dissipative term, i.e. of the equation*

$$\partial_t Q_{ik} + v_m \partial_m Q_{ik} + (\partial_m v_i) Q_{mk} = T_{ik}, \tag{55}$$

*is an entropy-type inequality*

$$\partial_t |Q| + \partial_m (v_m |Q|) = \varepsilon \frac{|Q| \delta_{kk}}{h \operatorname{tr} \mathbf{P}} \partial_m q_i (\mathcal{E}_{q_i q_j}) \partial_m q_j \geq 0, \tag{56}$$

with  $\{i, j\} \in \{1, 2, 3\}$ .

**Proof** To see that the entropy inequality is a direct consequence of (55) we apply the Jacobi identity (20) to (55), which leads to

$$\partial_t |Q| + |Q| Q_{ki}^{-1} v_m \partial_m Q_{ik} + |Q| Q_{ki}^{-1} (\partial_m v_i) Q_{mk} = |Q| Q_{ki}^{-1} T_{ik}, \tag{57}$$

from which one obtains

$$\partial_t |Q| + \partial_m (v_m |Q|) = |Q| Q_{ki}^{-1} T_{ik}. \tag{58}$$

With

$$|Q|Q_{ki}^{-1}T_{ik} = \varepsilon \frac{|Q|Q_{ki}^{-1}Q_{ik}}{h \operatorname{tr} \mathbf{P}} \partial_m q_i \mathcal{E}_{q_i q_j} \partial_m q_j = \varepsilon \frac{|Q|\delta_{kk}}{h \operatorname{tr} \mathbf{P}} \partial_m q_i \mathcal{E}_{q_i q_j} \partial_m q_j \geq 0 \quad (59)$$

one obtains the following entropy-type inequality associated with system (44)–(46):

$$\partial_t |Q| + \partial_m (v_m |Q|) = \varepsilon \frac{|Q|\delta_{kk}}{h \operatorname{tr} \mathbf{P}} \partial_m q_i \mathcal{E}_{q_i q_j} \partial_m q_j \geq 0, \quad (60)$$

where  $\{i, j\} \in \{1, 2, 3\}$ .  $\square$

Throughout this paper, we will consider entropy solutions of (14)–(18) that satisfy (44)–(47) in the limit  $\varepsilon \rightarrow 0$ . As shown later, the thermodynamically compatible scheme proposed in Sect. 3 of this paper is provably compatible with this vanishing viscosity limit, since it mimics the above viscous system *exactly* at the semi-discrete level. In the section containing the numerical results, we provide numerical evidence that also the high order ADER-DG schemes proposed in Sect. 4 of this paper as well as the numerical scheme already developed in [69] are compatible with this vanishing viscosity limit.

The meaning of Theorem 2 is the following. The evolution equation for the tensor  $\mathbf{P}$  (or for  $\mathbf{Q}$ ) is responsible for the vorticity transport, dissipation and production. While the transport and dissipation terms are clearly identified in previous works, it is not the same for the production terms. In the one-dimensional case the energy equation is equivalent to the ‘entropy’ equation and the ‘entropy’ (or vorticity) production is a consequence of the energy conservation. In the multi-dimensional case, the situation is completely different because the governing equations cannot be written in conservative form (the proof is given in [69]). So, the definition of weak solutions for such a non-conservative hyperbolic system which is compatible with the entropy production, should be given. In particular, such a definition was proposed in [69]. The Theorem 2 can be seen as a compatible alternative approach for the definition of weak solutions: the ‘viscous’ terms playing a major role in shocks guarantee the vorticity production. Moreover, the ‘viscous’ terms are consistent with the energy conservation law (Theorem 1) that is a necessary condition for all physically reasonable mathematical models.

### 3 Thermodynamically Compatible Finite Volume Scheme

In order to derive our new thermodynamically compatible finite volume scheme for system (14)–(18) that mimics the structure of the viscous system (44)–(47) *exactly* at the semi-discrete level, we proceed in a similar way as on the continuous level. First, a compatible scheme for the shallow water subsystem (36)–(38) is derived, based on a semi-discrete version of the Godunov form of (29)–(30). This corresponds to the discretization of the black terms in (44)–(47). Then, numerical viscosity together with an appropriate entropy production term is added to the scheme, which corresponds to the discrete analogue of the blue terms in (44)–(47). Last but not least the discretization of the red terms in (44)–(47) is discussed. To keep the presentation simple, we restrict ourselves to the one-dimensional case, but the generalization to multiple space dimensions is straightforward. To avoid confusion in the notation throughout this section we will use the lower case subscripts  $i, j, k, l, m$  for tensor indices and the lower case superscript  $r$  for the spatial discretization index. We emphasize again that the scheme proposed in this section is only valid for the flat bottom case with  $b = 0$ .

### 3.1 Compatible Schemes Without Dissipation Applied to the Godunov Form

A *semi-discrete* conservative finite volume scheme for system (28) in one space dimension based on the spatial control volume  $\Omega^r = [x^{r-\frac{1}{2}}, x^{r+\frac{1}{2}}]$  reads

$$\frac{d}{dt} \mathbf{q}^r = - \frac{\mathbf{f}^{r+\frac{1}{2}} - \mathbf{f}^{r-\frac{1}{2}}}{\Delta x}. \quad (61)$$

By adding and subtracting  $\mathbf{f}^r = \mathbf{f}(\mathbf{q}^r)$  we get

$$\frac{d}{dt} \mathbf{q}^r = - \frac{(\mathbf{f}^{r+\frac{1}{2}} - \mathbf{f}^r) - (\mathbf{f}^{r-\frac{1}{2}} - \mathbf{f}^r)}{\Delta x}. \quad (62)$$

We now try to obtain a *discrete form* of the total energy conservation law (30) also as a *consequence* of the discrete equations (62). For this purpose, we multiply (62) with  $\mathbf{p}^r = \mathcal{E}_{\mathbf{q}}(\mathbf{q}^r)$  from the left and get

$$\mathbf{p}^r \cdot \frac{d}{dt} \mathbf{q}^r = \frac{d}{dt} \mathcal{E}^r = - \mathbf{p}^r \cdot \frac{(\mathbf{f}^{r+\frac{1}{2}} - \mathbf{f}^r) + (\mathbf{f}^r - \mathbf{f}^{r-\frac{1}{2}})}{\Delta x} := - \frac{1}{\Delta x} \left( D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r-\frac{1}{2},+} \right), \quad (63)$$

where the fluctuations  $D_{\mathcal{E}}^{r+\frac{1}{2},-} = \mathbf{p}^r \cdot (\mathbf{f}^{r+\frac{1}{2}} - \mathbf{f}^r)$  and  $D_{\mathcal{E}}^{r-\frac{1}{2},+} = \mathbf{p}^r \cdot (\mathbf{f}^r - \mathbf{f}^{r-\frac{1}{2}})$  have been introduced for convenience. Obviously,  $D_{\mathcal{E}}^{r+\frac{1}{2},+} = \mathbf{p}^{r+1} \cdot (\mathbf{f}^{r+1} - \mathbf{f}^{r+\frac{1}{2}})$ . We now compute the temporal rate of change of the sum of the total energy in cell  $r$  and  $r+1$ , which yields

$$\Delta x \frac{d}{dt} (\mathcal{E}^r + \mathcal{E}^{r+1}) = - \left( D_{\mathcal{E}}^{r-\frac{1}{2},+} + D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r+\frac{1}{2},+} + D_{\mathcal{E}}^{r+\frac{3}{2},-} \right). \quad (64)$$

It is clear that in order to obtain a *flux conservative* form of the discrete energy conservation equation we must require that the contribution of the left and the right fluctuation at the interface  $r + \frac{1}{2}$  is a *flux difference*, i.e.

$$D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r+\frac{1}{2},+} := F^{r+1} - F^r, \quad (65)$$

where  $F^r$  must be a consistent approximation of the total energy flux  $F$ . Inserting the definitions of the fluctuations into (65) yields

$$\begin{aligned} & \mathbf{p}^r \cdot (\mathbf{f}^{r+\frac{1}{2}} - \mathbf{f}^r) + \mathbf{p}^{r+1} \cdot (\mathbf{f}^{r+1} - \mathbf{f}^{r+\frac{1}{2}}) = \\ & -\mathbf{f}^{r+\frac{1}{2}} \cdot (\mathbf{p}^{r+1} - \mathbf{p}^r) + \mathbf{p}^{r+1} \cdot \mathbf{f}^{r+1} - \mathbf{p}^r \cdot \mathbf{f}^r := F^{r+1} - F^r. \end{aligned} \quad (66)$$

Using the parametrization (29) and the associated relations (31) we get

$$\begin{aligned} & -(vL)_{\mathbf{p}}^{r+\frac{1}{2}} \cdot (\mathbf{p}^{r+1} - \mathbf{p}^r) + \mathbf{p}^{r+1} \cdot \mathbf{f}^{r+1} - \mathbf{p}^r \cdot \mathbf{f}^r := \\ & \mathbf{p}^{r+1} \cdot \mathbf{f}^{r+1} - (vL)^{r+1} - \mathbf{p}^r \cdot \mathbf{f}^r + (vL)^r, \end{aligned} \quad (67)$$

with  $F^r = \mathbf{p}^r \cdot \mathbf{f}^r - (vL)^r$  and thus the sought relation that the numerical flux  $\mathbf{f}^{r+\frac{1}{2}}$  must satisfy is

$$\mathbf{f}^{r+\frac{1}{2}} \cdot (\mathbf{p}^{r+1} - \mathbf{p}^r) = (vL)_{\mathbf{p}}^{r+\frac{1}{2}} \cdot (\mathbf{p}^{r+1} - \mathbf{p}^r) = (vL)^{r+1} - (vL)^r. \quad (68)$$

The condition (68) above is like a Roe-type property, but only for the vector  $\mathbf{f}^{r+\frac{1}{2}}$  instead of an entire Roe matrix. Based on the ideas of path-conservative schemes of Castro and Parés

[22,89] we thus define the numerical flux via a *path-integral* in phase-space, since by the fundamental theorem of calculus we have

$$(vL)^{r+1} - (vL)^r = \int_{\mathbf{p}^r}^{\mathbf{p}^{r+1}} (vL)_{\mathbf{p}} \cdot d\mathbf{p} = \int_0^1 (vL)_{\mathbf{p}} \cdot \frac{\partial \boldsymbol{\psi}}{\partial s} ds \quad (69)$$

for any path  $\boldsymbol{\psi} = \boldsymbol{\psi}(s)$  connecting  $\mathbf{p}^r$  with  $\mathbf{p}^{r+1}$ , see also the pioneering work of Tadmor [108] for a similar construction of an entropy-conservative flux at the aid of a path integral. The last equality in (69) means a concrete parametrization of the chosen integration path using integration by substitution and a dimensionless integration parameter  $s$  in the range  $0 \leq s \leq 1$ . In the following we choose two different parametrizations based on the simple straight-line segment path. Note that the choice of the path is arbitrary, hence we are free to choose a path that is somehow *convenient* for our purposes.

1. Segment path in the  $\mathbf{p}$  variables ( $\mathbf{p}$ -scheme). In the  $\mathbf{p}$ -scheme, the path between  $\mathbf{p}^r$  and  $\mathbf{p}^{r+1}$  is directly given by the straight line segment

$$\boldsymbol{\psi}(s) = \mathbf{p}^r + s(\mathbf{p}^{r+1} - \mathbf{p}^r), \quad 0 \leq s \leq 1. \quad (70)$$

We thus obtain

$$\frac{\partial \boldsymbol{\psi}}{\partial s} = \mathbf{p}^{r+1} - \mathbf{p}^r, \quad (71)$$

and therefore relation (69) results in

$$\begin{aligned} (vL)^{r+1} - (vL)^r &= \int_{\mathbf{p}^r}^{\mathbf{p}^{r+1}} (vL)_{\mathbf{p}} \cdot d\mathbf{p} = \int_0^1 \mathbf{f}(\boldsymbol{\psi}(s)) \cdot \frac{\partial \boldsymbol{\psi}}{\partial s} ds \\ &= \left( \int_0^1 \mathbf{f}(\boldsymbol{\psi}(s))^T ds \right) \cdot (\mathbf{p}^{r+1} - \mathbf{p}^r). \end{aligned} \quad (72)$$

By comparison with (68) we find that the thermodynamically compatible numerical flux of the  $\mathbf{p}$ -scheme is therefore given by

$$\mathbf{f}_{\mathbf{p}}^{r+\frac{1}{2}} = \int_0^1 \mathbf{f}(\boldsymbol{\psi}(s)) ds, \quad (73)$$

which by construction satisfies  $(\mathbf{p}^{r+1} - \mathbf{p}^r) \cdot \mathbf{f}_{\mathbf{p}}^{r+\frac{1}{2}} = (vL)^{r+1} - (vL)^r$  and thus condition (68). The problem with the  $\mathbf{p}$ -scheme is that it requires  $\mathbf{f}$  in terms of  $\mathbf{p}$  variables, which in general is very cumbersome, since usually  $\mathbf{f}$  is easier known in terms of  $\mathbf{q}$  rather than in terms of  $\mathbf{p}$ .

2. Segment in the  $\mathbf{q}$  variables ( $\mathbf{q}$ -scheme). To avoid the above-mentioned problem, in the  $\mathbf{q}$ -scheme the path between  $\mathbf{p}^r$  and  $\mathbf{p}^{r+1}$  is now defined in terms of a straight line segment in the  $\mathbf{q}$  variables, which means in terms of  $\mathbf{p}$  variables the path is in general *not* a segment. We set

$$\tilde{\boldsymbol{\psi}}(s) = \mathbf{p}(\mathbf{q}^r + s(\mathbf{q}^{r+1} - \mathbf{q}^r)), \quad 0 \leq s \leq 1. \quad (74)$$

Here, we only use the notation  $\tilde{\psi}(s)$  to avoid confusion with the path used before in the  $\mathbf{p}$ -scheme. We therefore have

$$\frac{\partial \tilde{\psi}}{\partial s} = \frac{\partial \mathbf{p}}{\partial \mathbf{q}} \cdot (\mathbf{q}^{r+1} - \mathbf{q}^r) = \mathcal{E}_{\mathbf{q}\mathbf{q}} \cdot (\mathbf{q}^{r+1} - \mathbf{q}^r), \quad (75)$$

and thus condition (69) results in

$$\begin{aligned} (vL)^{r+1} - (vL)^r &= \int_{\mathbf{p}^r}^{\mathbf{p}^{r+1}} (vL)_{\mathbf{p}} \cdot d\mathbf{p} = \int_0^1 \mathbf{f}(\tilde{\psi}(s)) \cdot \frac{\partial \tilde{\psi}}{\partial s} ds \\ &= \left( \int_0^1 \mathcal{E}_{\mathbf{q}\mathbf{q}} \mathbf{f}(\tilde{\psi}(s))^T ds \right) \cdot (\mathbf{q}^{r+1} - \mathbf{q}^r). \end{aligned} \quad (76)$$

If we now check again condition (68) we still need to transform the jump in  $\mathbf{p}$  variables into a jump in  $\mathbf{q}$  variables. For that purpose, we define a Roe-type matrix  $\tilde{L}_{\mathbf{q}\mathbf{q}}$  that satisfies the Roe property

$$\tilde{L}_{\mathbf{p}\mathbf{p}}^{r+\frac{1}{2}} \cdot (\mathbf{p}^{r+1} - \mathbf{p}^r) = \mathbf{q}^{r+1} - \mathbf{q}^r, \quad (77)$$

which can be easily achieved by construction by the means of a path integral. In practice we first define the *inverse* of the Roe matrix  $\tilde{L}_{\mathbf{p}\mathbf{p}}^{r+\frac{1}{2}}$  as

$$\tilde{\mathcal{E}}_{\mathbf{q}\mathbf{q}}^{r+\frac{1}{2}} = \int_0^1 \mathcal{E}_{\mathbf{q}\mathbf{q}}(\tilde{\psi}(s)) ds, \quad (78)$$

which is again a Roe matrix, but which is easy to compute, and which can be checked to satisfy

$$\tilde{\mathcal{E}}_{\mathbf{q}\mathbf{q}}^{r+\frac{1}{2}} \cdot (\mathbf{q}^{r+1} - \mathbf{q}^r) = \mathbf{p}^{r+1} - \mathbf{p}^r \quad (79)$$

by construction. We thus obtain

$$\tilde{L}_{\mathbf{p}\mathbf{p}}^{r+\frac{1}{2}} = \left( \tilde{\mathcal{E}}_{\mathbf{q}\mathbf{q}}^{r+\frac{1}{2}} \right)^{-1} = \left( \int_0^1 \mathcal{E}_{\mathbf{q}\mathbf{q}}(\tilde{\psi}(s)) ds \right)^{-1}, \quad (80)$$

which finally yields the desired thermodynamically compatible numerical flux of the  $\mathbf{q}$  scheme as

$$\mathbf{f}_{\mathbf{q}}^{r+\frac{1}{2}} = \tilde{L}_{\mathbf{p}\mathbf{p}}^{r+\frac{1}{2}} \int_0^1 \mathcal{E}_{\mathbf{q}\mathbf{q}} \mathbf{f}(\tilde{\psi}(s)) ds = \left( \int_0^1 \mathcal{E}_{\mathbf{q}\mathbf{q}}(\tilde{\psi}(s)) ds \right)^{-1} \left( \int_0^1 \mathcal{E}_{\mathbf{q}\mathbf{q}} \mathbf{f}(\tilde{\psi}(s)) ds \right). \quad (81)$$

Note that if  $\mathbf{f}$  is only easily known in terms of  $\mathbf{q}$  variables, one can directly plug the straight segment path in terms of the  $\mathbf{q}$  variables into the function  $\mathbf{f}$  and into the Hessian  $\mathcal{E}_{\mathbf{q}\mathbf{q}}$ , without needing to compute  $\mathbf{p}(\mathbf{q})$  at all!

In practical calculations, we approximate all path integrals by *numerical quadrature*, which can be done up to any desired level of accuracy, see also [42,45,48,49] where this strategy has already been successfully used.

### 3.2 Compatible Scheme with Dissipation Applied to the Godunov Form

The above schemes are compatible with the parametrization (29) of the system (28) and also satisfy the extra conservation law (30). However, to obtain a *dissipative scheme*, we still need to add a *compatible* numerical dissipation. For that purpose we write a *dissipative* scheme for (28) of the form

$$\frac{d}{dt} \mathbf{q}^r + \frac{\mathbf{f}^{r+\frac{1}{2}} - \mathbf{f}^{r-\frac{1}{2}}}{\Delta x} = \frac{\mathbf{g}^{r+\frac{1}{2}} - \mathbf{g}^{r-\frac{1}{2}}}{\Delta x} + \mathbf{T}^r, \quad (82)$$

with the compatible flux  $\mathbf{f}^{r+\frac{1}{2}}$  as defined before and the additional *dissipative numerical flux*  $\mathbf{g}^{r+\frac{1}{2}}$  given by

$$\mathbf{g}^{r+\frac{1}{2}} = \mu^{r+\frac{1}{2}} \frac{\mathbf{q}^{r+1} - \mathbf{q}^r}{\Delta x} = \mu^{r+\frac{1}{2}} \frac{\Delta \mathbf{q}^{r+\frac{1}{2}}}{\Delta x}, \quad (83)$$

where  $\mu^{r+\frac{1}{2}} \geq 0$  is a *scalar* numerical dissipation. Henceforth we simply set

$$\mu^{r+\frac{1}{2}} = \frac{1}{2} \left( 1 - \varphi^{r+\frac{1}{2}} \right) \Delta x s_{\max}^{r+\frac{1}{2}} \geq 0, \quad (84)$$

with  $s_{\max}^{r+\frac{1}{2}}$  an estimate for the maximum signal speed at the interface. For  $\varphi^{r+\frac{1}{2}} = 0$  this choice corresponds to a classical first order Rusanov-type scheme, see [107, 115]. To reduce numerical dissipation in smooth regions, a TVD *minbee flux limiter*  $\varphi^{r+\frac{1}{2}}$  is employed, which is defined as follows, see the second order TVD SLIC scheme described by Toro in [115],

$$\varphi^{r+\frac{1}{2}} = \min \left( \varphi_-^{r+\frac{1}{2}}, \varphi_+^{r+\frac{1}{2}} \right), \quad \text{with} \quad \varphi_{\pm}^{r+\frac{1}{2}} = \max \left( 0, \min \left( 1, \rho_{\pm}^{r+\frac{1}{2}} \right) \right), \quad (85)$$

with the ratios of subsequent slopes of the total energy potential defined as

$$\rho_-^{r+\frac{1}{2}} = \frac{\mathcal{E}^r - \mathcal{E}^{r-1}}{\mathcal{E}^{r+1} - \mathcal{E}^r}, \quad \text{and} \quad \rho_+^{r+\frac{1}{2}} = \frac{\mathcal{E}^{r+2} - \mathcal{E}^{r+1}}{\mathcal{E}^{r+1} - \mathcal{E}^r}. \quad (86)$$

Note that in regions of  $\varphi^{r+\frac{1}{2}} = 1$  the scheme exhibits *no numerical viscosity at all*. The production term  $\mathbf{T}^r$  will be defined later. Computing the dot product of (82) with  $\mathbf{p}^i$  yields

$$\frac{d}{dt} \mathcal{E}^r + \frac{F^{r+\frac{1}{2}} - F^{r-\frac{1}{2}}}{\Delta x} = \mathbf{p}^r \cdot \frac{\mathbf{g}^{r+\frac{1}{2}} - \mathbf{g}^{r-\frac{1}{2}}}{\Delta x} + \mathbf{p}^r \cdot \mathbf{T}^r, \quad (87)$$

where we denote the inviscid numerical flux for the total energy by  $F^{r+\frac{1}{2}}$ . Since the dissipation-free scheme has already been shown to be compatible with the energy conservation law, in what follows, the explicit expression for  $F^{r+\frac{1}{2}}$  is not needed, but is given here for completeness:

$$F^{r+\frac{1}{2}} = D_{\mathcal{E}}^{r+\frac{1}{2},-} + F^r = \mathbf{p}^r \cdot (\mathbf{f}^{r+\frac{1}{2}} - \mathbf{f}^r) + F^r. \quad (88)$$

We now rewrite the right hand side of (87) as

$$\begin{aligned}
 & \mathbf{p}^r \cdot \frac{\mathbf{g}^{r+\frac{1}{2}} - \mathbf{g}^{r-\frac{1}{2}}}{\Delta x} + \mathbf{p}^r \cdot \mathbf{T}^r \\
 &= \mathbf{p}^r \cdot \mathbf{T}^r + \frac{1}{\Delta x} \left( \frac{1}{2} \mathbf{p}^r \cdot \mathbf{g}^{r+\frac{1}{2}} + \frac{1}{2} \mathbf{p}^{r+1} \cdot \mathbf{g}^{r+\frac{1}{2}} + \frac{1}{2} \mathbf{p}^r \cdot \mathbf{g}^{r+\frac{1}{2}} - \frac{1}{2} \mathbf{p}^{r+1} \cdot \mathbf{g}^{r+\frac{1}{2}} \right) \\
 &\quad - \frac{1}{\Delta x} \left( \frac{1}{2} \mathbf{p}^r \cdot \mathbf{g}^{r-\frac{1}{2}} + \frac{1}{2} \mathbf{p}^{r-1} \cdot \mathbf{g}^{r-\frac{1}{2}} + \frac{1}{2} \mathbf{p}^r \cdot \mathbf{g}^{r-\frac{1}{2}} - \frac{1}{2} \mathbf{p}^{r-1} \cdot \mathbf{g}^{r-\frac{1}{2}} \right) \\
 &= \mathbf{p}^r \cdot \mathbf{T}^r + \frac{1}{2} \frac{\mathbf{p}^{r+1} + \mathbf{p}^r}{\Delta x} \cdot \mathbf{g}^{r+\frac{1}{2}} - \frac{1}{2} \frac{\mathbf{p}^r + \mathbf{p}^{r-1}}{\Delta x} \cdot \mathbf{g}^{r-\frac{1}{2}} \\
 &\quad - \frac{1}{2} \frac{\mathbf{p}^{r+1} - \mathbf{p}^r}{\Delta x} \cdot \mathbf{g}^{r+\frac{1}{2}} - \frac{1}{2} \frac{\mathbf{p}^r - \mathbf{p}^{r-1}}{\Delta x} \cdot \mathbf{g}^{r-\frac{1}{2}} \\
 &= \mathbf{p}^r \cdot \mathbf{T}^r + \frac{1}{2} \frac{\mathbf{p}^{r+1} + \mathbf{p}^r}{\Delta x} \cdot \mu^{r+\frac{1}{2}} \frac{\mathbf{q}^{r+1} - \mathbf{q}^r}{\Delta x} - \frac{1}{2} \frac{\mathbf{p}^r + \mathbf{p}^{r-1}}{\Delta x} \cdot \mu^{r-\frac{1}{2}} \frac{\mathbf{q}^r - \mathbf{q}^{r-1}}{\Delta x} \\
 &\quad - \frac{1}{2} \frac{\mathbf{p}^{r+1} - \mathbf{p}^r}{\Delta x} \cdot \mu^{r+\frac{1}{2}} \frac{\mathbf{q}^{r+1} - \mathbf{q}^r}{\Delta x} - \frac{1}{2} \frac{\mathbf{p}^r - \mathbf{p}^{r-1}}{\Delta x} \cdot \mu^{r-\frac{1}{2}} \frac{\mathbf{q}^r - \mathbf{q}^{r-1}}{\Delta x}. \quad (89)
 \end{aligned}$$

The total energy flux including the dissipative terms thus reads as follows

$$\begin{aligned}
 F_d^{r+\frac{1}{2}} &= F^{r+\frac{1}{2}} - \frac{1}{2} (\mathbf{p}^{r+1} + \mathbf{p}^r) \cdot \mu^{r+\frac{1}{2}} \frac{\Delta \mathbf{q}^{r+\frac{1}{2}}}{\Delta x} \\
 &\approx F^{r+\frac{1}{2}} - \mu^{r+\frac{1}{2}} \frac{\Delta \mathcal{E}^{r+\frac{1}{2}}}{\Delta x}, \quad (90)
 \end{aligned}$$

since the expression  $\frac{1}{2} (\mathbf{p}^{r+1} + \mathbf{p}^r) \cdot \Delta \mathbf{q}^{r+\frac{1}{2}}$  is an *approximation of the path integral*

$$\int_{\mathbf{q}^r}^{\mathbf{q}^{r+1}} \mathbf{p} \cdot d\mathbf{q} = \int_{\mathbf{q}^r}^{\mathbf{q}^{r+1}} \mathcal{E}_{\mathbf{q}}^T \cdot d\mathbf{q} = \mathcal{E}^{r+1} - \mathcal{E}^r := \Delta \mathcal{E}^{r+\frac{1}{2}} \quad (91)$$

using the simple trapezoidal quadrature rule. Making again use of the *symmetric* Roe matrix  $\tilde{\mathcal{E}}_{\mathbf{q}\mathbf{q}}^{r+\frac{1}{2}}$ , which satisfies  $\tilde{\mathcal{E}}_{\mathbf{q}\mathbf{q}}^{r+\frac{1}{2}} (\mathbf{q}^{r+1} - \mathbf{q}^r) = \mathbf{p}^{r+1} - \mathbf{p}^r$ , the semi-discrete total energy conservation law takes the form

$$\begin{aligned}
 \frac{d}{dt} \mathcal{E}^r + \frac{F_d^{r+\frac{1}{2}} - F_d^{r-\frac{1}{2}}}{\Delta x} &= \mathbf{p}^r \cdot \mathbf{T}^r - \\
 &\quad - \frac{1}{2} \mu^{r+\frac{1}{2}} \frac{\mathbf{q}^{r+1} - \mathbf{q}^r}{\Delta x} \cdot \tilde{\mathcal{E}}_{\mathbf{p}\mathbf{p}}^{r+\frac{1}{2}} \frac{\mathbf{q}^{r+1} - \mathbf{q}^r}{\Delta x} - \frac{1}{2} \mu^{r-\frac{1}{2}} \frac{\mathbf{q}^r - \mathbf{q}^{r-1}}{\Delta x} \cdot \tilde{\mathcal{E}}_{\mathbf{q}\mathbf{q}}^{r-\frac{1}{2}} \frac{\mathbf{q}^r - \mathbf{q}^{r-1}}{\Delta x}. \quad (92)
 \end{aligned}$$

By requiring that

$$\mathbf{p}^r \cdot \mathbf{T}^r := \frac{1}{2} \mu^{r+\frac{1}{2}} \frac{\Delta \mathbf{q}^{r+\frac{1}{2}}}{\Delta x} \cdot \tilde{\mathcal{E}}_{\mathbf{q}\mathbf{q}}^{r+\frac{1}{2}} \frac{\Delta \mathbf{q}^{r+\frac{1}{2}}}{\Delta x} + \frac{1}{2} \mu^{r-\frac{1}{2}} \frac{\Delta \mathbf{q}^{r-\frac{1}{2}}}{\Delta x} \cdot \tilde{\mathcal{E}}_{\mathbf{q}\mathbf{q}}^{r-\frac{1}{2}} \frac{\Delta \mathbf{q}^{r-\frac{1}{2}}}{\Delta x}, \quad (93)$$

we finally obtain the sought *conservation form* of the discrete total energy equation (87):

$$\frac{d}{dt} \mathcal{E}^r + \frac{F_d^{r+\frac{1}{2}} - F_d^{r-\frac{1}{2}}}{\Delta x} = 0. \quad (94)$$

The term  $\mathbf{T}^r$  is set identically to zero in all its components, apart from the equations that are needed for the entropy inequality, which are the nonconservative evolution equations for  $Q_{ik}$ . Therefore, the term  $\mathbf{T}^r$  will be discussed later.

To summarize, the thermodynamically compatible *dissipative* numerical flux of the  $\mathbf{q}$  scheme for (28) reads

$$\mathbf{f}_{\mathbf{q},d}^{r+\frac{1}{2}} = \left( \int_0^1 \mathcal{E}_{\mathbf{q}\mathbf{q}}(\tilde{\psi}(s)) ds \right)^{-1} \left( \int_0^1 \mathcal{E}_{\mathbf{q}\mathbf{q}}\mathbf{f}(\tilde{\psi}(s)) ds \right) - \frac{1}{2} s_{\max}^{r+\frac{1}{2}} \left( 1 - \varphi^{r+\frac{1}{2}} \right) (\mathbf{q}^{r+1} - \mathbf{q}^r). \quad (95)$$

### 3.3 Thermodynamically Compatible Discretization of the Terms Related to $Q_{ik}$

We now present the discretization of the Reynolds stress tensor  $R_{ik} = h P_{ik}$  in the momentum equation that is thermodynamically compatible with the term  $(\partial_m v_i) Q_{mk}$  in (46) and the term  $h P_{ik} v_k$  in the energy equation. To ease notation, we present the discretization only for the one-dimensional case in  $x_1$  direction. Extension to multiple space dimensions is straightforward. For the term  $(\partial_m v_i) Q_{mk}$  we a priori choose the following discretization:

$$\Delta x (\partial_1 v_i) Q_{1k} \approx Q_{1k}^{r+\frac{1}{2}} (v_i^{r+1} - v_i^r), \quad \text{with} \quad Q_{1k}^{r+\frac{1}{2}} = \frac{1}{2} (Q_{1k}^r + Q_{1k}^{r+1}). \quad (96)$$

Multiplication of the momentum equation (45) with the dual variable  $\mathcal{E}_{hv_i} = v_i$  and of PDE (46) with the dual variable  $\mathcal{E}_{Q_{ik}} = h Q_{ik}$  and requiring thermodynamic compatibility with total energy equation leads to the following requirement that needs to be fulfilled by the yet unknown discretization of the Reynolds stress tensor  $R_{i1}^{r+\frac{1}{2}}$ :

$$\begin{aligned} & v_i^r \left( R_{i1}^{r+\frac{1}{2}} - R_{i1}^r \right) + v_i^{r+1} \left( R_{i1}^{r+1} - R_{i1}^{r+\frac{1}{2}} \right) \\ & + h^r Q_{ik}^r \frac{1}{2} Q_{1k}^{r+\frac{1}{2}} (v_i^{r+1} - v_i^r) + h^{r+1} Q_{ik}^{r+1} \frac{1}{2} Q_{1k}^{r+\frac{1}{2}} (v_i^{r+1} - v_i^r) \\ & = h^{r+1} Q_{im}^{r+1} Q_{1m}^{r+1} v_i^{r+1} - h^r Q_{im}^r Q_{1m}^r v_i^r. \end{aligned} \quad (97)$$

Using  $R_{ik} = h Q_{im} Q_{km}$  and collecting terms leads to

$$- R_{i1}^{r+\frac{1}{2}} (v_i^{r+1} - v_i^r) + Q_{1k}^{r+\frac{1}{2}} \frac{1}{2} (h^r Q_{ik}^r + h^{r+1} Q_{ik}^{r+1}) (v_i^{r+1} - v_i^r) = 0. \quad (98)$$

Since  $Q_{1k}^{r+\frac{1}{2}} = \frac{1}{2} (Q_{1k}^r + Q_{1k}^{r+1})$ , we obtain the following *compatible discretization* for the *numerical flux* of the Reynolds stress tensor in  $x_1$  direction:

$$R_{i1}^{r+\frac{1}{2}} = \frac{1}{2} (Q_{1k}^r + Q_{1k}^{r+1}) \frac{1}{2} (h^r Q_{ik}^r + h^{r+1} Q_{ik}^{r+1}), \quad (99)$$

which needs to be *added* to the compatible dissipative flux (95) in the semi-discrete momentum equation, which then takes the form

$$\frac{d}{dt} (h v_i) = - \frac{1}{\Delta x} \left( \mathbf{f}_{hv_i,d}^{r+\frac{1}{2}} - \mathbf{f}_{hv_i,d}^{r-\frac{1}{2}} \right) - \frac{1}{\Delta x} \left( R_{i1}^{r+\frac{1}{2}} - R_{i1}^{r-\frac{1}{2}} \right), \quad (100)$$

where  $\mathbf{f}_{hvi,d}^{r+\frac{1}{2}}$  is the part of the dissipative flux in  $x_1$  direction (95) that refers to the momentum equation.

The last term in (46) that needs to be discretized is the convective term  $v_m \partial_m Q_{ik}$ , which requires compatibility with the mass conservation law (44) and the energy conservation (47). To achieve such a compatible discretization, the mass conservation equation needs to be multiplied with the remaining contribution  $\mathcal{E}_{2,h} = E_2$  and the PDE for  $Q_{ik}$  is again multiplied with  $\mathcal{E}_{Q_{ik}}$ , and the following condition must be satisfied:

$$\begin{aligned} E_2^r \left( (hv)_1^{r+\frac{1}{2}} - (hv)_1^r \right) + E_2^{r+1} \left( (hv)_1^{r+1} - (hv)_1^{r+\frac{1}{2}} \right) \\ + h^r Q_{ik}^r \frac{1}{2} \tilde{v}_1^{r+\frac{1}{2}} \left( Q_{ik}^{r+1} - Q_{ik}^r \right) + h^{r+1} Q_{ik}^{r+1} \frac{1}{2} \tilde{v}_1^{r+\frac{1}{2}} \left( Q_{ik}^{r+1} - Q_{ik}^r \right) \\ = (hv)_1^{r+1} E_2^{r+1} - (hv)_1^r E_2^r, \end{aligned} \quad (101)$$

with the yet unknown average velocity  $\tilde{v}_1^{r+\frac{1}{2}}$  at the cell interface. Note that the numerical mass flux  $(hv)_1^{r+\frac{1}{2}}$  is the *known* compatible *inviscid* mass flux of the numerical flux  $\mathbf{f}_{\mathbf{q}}^{r+\frac{1}{2}}$  of the  $\mathbf{q}$ -scheme according to the semi-discrete Godunov formalism, see (81). Collecting terms leads to

$$(hv)_1^{r+\frac{1}{2}} \left( E_2^{r+1} - E_2^r \right) = \tilde{v}_1^{r+\frac{1}{2}} \left( h^{r+1} E_2^{r+1} - h^r E_2^r - \frac{1}{2} Q_{ik}^r Q_{ik}^{r+1} (h^{r+1} - h^r) \right), \quad (102)$$

from which we obtain the sought expression for the average velocity at the interface as

$$\tilde{v}_1^{r+\frac{1}{2}} = \frac{(hv)_1^{r+\frac{1}{2}} \left( E_2^{r+1} - E_2^r \right)}{h^{r+1} E_2^{r+1} - h^r E_2^r - \frac{1}{2} Q_{ik}^r Q_{ik}^{r+1} (h^{r+1} - h^r)}. \quad (103)$$

In case the denominator is zero, we simply set the velocity to the arithmetic average  $\tilde{v}_1^{r+\frac{1}{2}} = \frac{1}{2} (\tilde{v}_1^r + \tilde{v}_1^{r+1})$ .

In order to get compatibility with the total energy conservation law also in the presence of numerical viscosity, we need to add the *discrete production term* to the PDEs of  $Q_{ik}$  at the right and left element interface, according to the condition (93) already derived before:

$$T_{ik}^{r+\frac{1}{2},-} = \mu^{r+\frac{1}{2}} \frac{1}{2} \frac{Q_{ik}^r}{(h \text{tr} \mathbf{P})^r} \frac{\Delta \mathbf{q}^{r+\frac{1}{2}}}{\Delta x} \cdot \tilde{\mathcal{E}}_{\mathbf{qq}}^{r+\frac{1}{2}} \frac{\Delta \mathbf{q}^{r+\frac{1}{2}}}{\Delta x} \quad (104)$$

and

$$T_{ik}^{r-\frac{1}{2},+} = \mu^{r-\frac{1}{2}} \frac{1}{2} \frac{Q_{ik}^r}{(h \text{tr} \mathbf{P})^r} \frac{\Delta \mathbf{q}^{r-\frac{1}{2}}}{\Delta x} \cdot \tilde{\mathcal{E}}_{\mathbf{qq}}^{r-\frac{1}{2}} \frac{\Delta \mathbf{q}^{r-\frac{1}{2}}}{\Delta x}. \quad (105)$$

The physical entropy production is always non-negative, since we assume  $\mu^{r+\frac{1}{2}} \geq 0$  and  $\tilde{\mathcal{E}}_{\mathbf{qq}}^{r+\frac{1}{2}} \geq 0$ . It is obvious that (105) and (104) are discrete analogues of the continuous production term (50).

The final semi-discrete scheme for  $Q_{ik}$  in one space dimension reads:

$$\begin{aligned} \frac{d}{dt} Q_{ik}^r = -\frac{1}{2} \tilde{v}_1^{r+\frac{1}{2}} \frac{Q_{ik}^{r+1} - Q_{ik}^r}{\Delta x} - \frac{1}{2} \cdot \frac{Q_{1k}^r + Q_{1k}^{r+1}}{2} \cdot \frac{v_i^{r+1} - v_i^r}{\Delta x} \\ + \frac{\mu^{r+\frac{1}{2}}}{\Delta x} \cdot \frac{Q_{ik}^{r+1} - Q_{ik}^r}{\Delta x} + T_{ik}^{r-\frac{1}{2},+} + T_{ik}^{r+\frac{1}{2},-}. \end{aligned} \quad (106)$$

### 3.4 Summary of the Scheme and Stability Proof

For completeness, we now gather together all equations of the thermodynamically compatible scheme, thus obtaining

$$\frac{dh^r}{dt} = -\frac{1}{\Delta x} \left( D_h^{r+\frac{1}{2},-} + D_h^{r-\frac{1}{2},+} \right) + \frac{1}{\Delta x} \left( g_h^{r+\frac{1}{2}} - g_h^{r-\frac{1}{2}} \right), \quad (107)$$

$$\begin{aligned} \frac{dhv_i^r}{dt} = & -\frac{1}{\Delta x} \left( D_{hv_i}^{r+\frac{1}{2},-} + D_{hv_i}^{r-\frac{1}{2},+} \right) - \frac{1}{\Delta x} \left( R_{i,1}^{r+\frac{1}{2},-} - R_{i,1}^{r-\frac{1}{2},+} \right) \\ & + \frac{1}{\Delta x} \left( g_{hv_i}^{r+\frac{1}{2}} - g_{hv_i}^{r-\frac{1}{2}} \right), \end{aligned} \quad (108)$$

$$\begin{aligned} \frac{dQ_{ik}^r}{dt} = & -\frac{1}{\Delta x} \left( D_{Q_{ik}}^{r+\frac{1}{2},-} + D_{Q_{ik}}^{r-\frac{1}{2},+} \right) + \frac{1}{\Delta x} \left( g_{Q_{ik}}^{r+\frac{1}{2}} - g_{Q_{ik}}^{r-\frac{1}{2}} \right) \\ & + T_{ik}^{r+\frac{1}{2},+} + T_{ik}^{r+\frac{1}{2},-}, \end{aligned} \quad (109)$$

with the fluctuations

$$D_q^{r+\frac{1}{2},-} = f_q^{r+\frac{1}{2}} - f_q^r, \quad \text{and} \quad D_q^{r+\frac{1}{2},+} = f_q^{r+1} - f_q^{r+\frac{1}{2}}, \quad (110)$$

where  $f_q^r$  denotes the physical flux evaluated in cell  $r$  and  $f_q^{r+\frac{1}{2}}$  is the compatible flux for depth and momentum, i.e. for  $q \in \{h, hv_i\}$ . Recall that the flux vector in the previous notation reads

$$\mathbf{f}_q^{r+\frac{1}{2}} = \left( f_h^{r+\frac{1}{2}}, f_{hv_i}^{r+\frac{1}{2}}, 0 \right) \quad (111)$$

and is computed according to the  $q$ -scheme of the semi-discrete Godunov formalism presented previously. We have also introduced the fluctuations

$$R_{i,1}^{r+\frac{1}{2},-} = \left( R_{i,1}^{r+\frac{1}{2}} - R_{i,1}^r \right), \quad R_{i,1}^{r+\frac{1}{2},+} = \left( R_{i,1}^{r+1} - R_{i,1}^{r+\frac{1}{2}} \right), \quad (112)$$

with  $R_{i,1}^{r+\frac{1}{2}}$  being the compatible discretization of the Reynolds stress tensor in  $x_1$  direction given in (99), and

$$D_{Q_{ik}}^{r+\frac{1}{2},\pm} = \frac{1}{2} \tilde{v}_1^{r+\frac{1}{2}} \left( Q_{1k}^{r+1} - Q_{1k}^r \right) + \frac{1}{2} \tilde{Q}_{1k}^{r+\frac{1}{2}} \left( v_i^{r+1} - v_i^r \right). \quad (113)$$

Let us also recall that the dissipative fluxes,  $g_q^{r\pm\frac{1}{2}}$ , have been defined in (83).

**Theorem 3** *The semi-discrete scheme (107)–(109) admits the semi-discrete energy conservation equation*

$$\frac{d\mathcal{E}^r}{dt} = -\frac{1}{\Delta x} \left( D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r-\frac{1}{2},+} \right) + \frac{1}{\Delta x} \left( g_{\mathcal{E}}^{r+\frac{1}{2}} - g_{\mathcal{E}}^{r-\frac{1}{2}} \right) \quad (114)$$

with

$$D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r+\frac{1}{2},+} = F^{r+1} - F^r. \quad (115)$$

As a result the scheme is energy conserving and therefore marginally stable in the energy norm, i.e. the scheme satisfies

$$\int_{\Omega} \frac{d\mathcal{E}}{dt} dx = \sum_r \Delta x \frac{d\mathcal{E}^r}{dt} = 0. \quad (116)$$

**Proof** We first demonstrate that (114) is a direct consequence of (107)–(109). To this end we proceed as in the continuous case, i.e. we start by considering the time derivatives and we sum the contributions coming from equations (107)–(108) multiplied by  $\mathcal{E}_h^r$ ,  $\mathcal{E}_{hv_1}^r$  and  $\mathcal{E}_{Q_{ik}}^r$ , respectively, thus obtaining

$$\mathcal{E}_h^r \frac{dh^r}{dt} + \mathcal{E}_{hv_1}^r \frac{dhv_i^r}{dt} + \mathcal{E}_{Q_{ik}}^r \frac{dQ_{ik}^r}{dt} = \mathcal{E}_{\mathbf{q}}^r \cdot \frac{d\mathbf{q}^r}{dt} = \frac{d\mathcal{E}^r}{dt}. \quad (117)$$

For the convective terms, the Reynolds stress tensor and the PDE related to  $\mathbf{Q}$  we define

$$\begin{aligned} D_{\mathcal{E}}^{r+\frac{1}{2},-} &:= \mathcal{E}_{1,h}^r D_h^{r+\frac{1}{2},-} + \mathcal{E}_{2,h}^r D_h^{r+\frac{1}{2},-} + \mathcal{E}_{hv_1}^r \left( D_{hv_1}^{r+\frac{1}{2},-} + R_{i,1}^{r+\frac{1}{2},-} \right) + \mathcal{E}_{Q_{ik}}^r D_{Q_{ik}}^{r+\frac{1}{2},-}, \\ D_{\mathcal{E}}^{r+\frac{1}{2},+} &:= \mathcal{E}_{1,h}^{r+1} D_h^{r+\frac{1}{2},+} + \mathcal{E}_{2,h}^{r+1} D_h^{r+\frac{1}{2},+} + \mathcal{E}_{hv_1}^{r+1} \left( D_{hv_1}^{r+\frac{1}{2},+} + R_{i,1}^{r+\frac{1}{2},+} \right) + \mathcal{E}_{Q_{ik}}^{r+1} D_{Q_{ik}}^{r+\frac{1}{2},+}. \end{aligned} \quad (118)$$

Let us remark that the definitions of the fluctuations in (118) differ from the ones given in Sect. 3.1, where the terms related to the total energy  $\mathcal{E}_2$  associated with  $Q_{ik}$  were not yet included. Finally, the dot product of  $\mathbf{p}^r$  by the vector of the blue terms in (107)–(109) yields

$$\mathbf{p}^r \cdot \frac{\mathbf{g}^{r+\frac{1}{2}} - \mathbf{g}^{r-\frac{1}{2}}}{\Delta x} + \mathbf{p}^r \cdot \mathbf{T}^r = \mathbf{p}^r \cdot \frac{\mu^{r+\frac{1}{2}} \Delta \mathbf{q}^{r+\frac{1}{2}} - \mu^{r-\frac{1}{2}} \Delta \mathbf{q}^{r-\frac{1}{2}}}{\Delta x^2} + \mathbf{p}^r \cdot \mathbf{T}^r. \quad (119)$$

Taking into account (104)–(105) in  $\mathbf{p}^r \cdot \mathbf{T}^r = \mathbf{p}^r \cdot \mathbf{T}^{r-\frac{1}{2},+} + \mathbf{p}^r \cdot \mathbf{T}^{r+\frac{1}{2},-}$  we get (93). Substitution of this result in (119) and making use of the developments in (89) gives

$$\begin{aligned} \mathbf{p}^r \cdot \frac{\mathbf{g}^{r+\frac{1}{2}} - \mathbf{g}^{r-\frac{1}{2}}}{\Delta x} + \mathbf{p}^r \cdot \mathbf{T}^r &= \frac{1}{\Delta x^2} \left( \mu^{r+\frac{1}{2}} \Delta \mathcal{E}^{r+\frac{1}{2}} - \mu^{r-\frac{1}{2}} \Delta \mathcal{E}^{r-\frac{1}{2}} \right) \\ &= \frac{1}{\Delta x} \left( g_{\mathcal{E}}^{r+\frac{1}{2}} - g_{\mathcal{E}}^{r-\frac{1}{2}} \right). \end{aligned} \quad (120)$$

Gathering (117), (120) and (118), we get (114):

$$\frac{d\mathcal{E}^r}{dt} + \frac{1}{\Delta x} \left( D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r-\frac{1}{2},+} \right) = \frac{1}{\Delta x} \left( g_{\mathcal{E}}^{r+\frac{1}{2}} - g_{\mathcal{E}}^{r-\frac{1}{2}} \right). \quad (121)$$

Let us now consider the discrete equation for the total energy, (114). Integrating it on a computational domain,  $\Omega$ , we get

$$\begin{aligned} \int_{\Omega} \frac{d\mathcal{E}}{dt} dx &= \sum_r \int_{\Omega^r} \frac{d\mathcal{E}^r}{dt} dx = \sum_r \Delta x \frac{d\mathcal{E}^r}{dt} \\ &= - \sum_r \frac{\Delta x}{\Delta x} \left( D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r-\frac{1}{2},+} \right) + \sum_r \left( g_{\mathcal{E}}^{r+\frac{1}{2}} - g_{\mathcal{E}}^{r-\frac{1}{2}} \right). \end{aligned} \quad (122)$$

Recalling that  $D_{\mathcal{E}}^{r\pm\frac{1}{2},\mp}$  represent the jumps of the energy flux at the interfaces and assuming the solution on the boundaries of  $\Omega$  to tend to a constant value, the jumps of  $\mathbf{q}$  are zero at the

boundaries of  $\Omega$  and then the boundary contributions of  $D_{\mathcal{E}}$  and also the contribution of the dissipation terms vanish. Hence, reordering the pairs of  $D_{\mathcal{E}}^{r+\frac{1}{2},\mp}$  to consider couples related to the interfaces instead of pairs corresponding to the cells yields

$$\int_{\Omega} \frac{d\mathcal{E}}{dt} dx = - \sum_r \left( D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r+\frac{1}{2},+} \right) + \sum_r \left( g_{\mathcal{E}}^{r+\frac{1}{2}} - g_{\mathcal{E}}^{r-\frac{1}{2}} \right). \quad (123)$$

Note that the summation over the blue dissipative fluxes is obviously a telescopic sum that vanishes. On the other hand, we can also prove that the contributions of the fluctuations at the interfaces reduce to a flux difference of the form

$$D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r+\frac{1}{2},+} = F^{r+1} - F^r. \quad (124)$$

To this end, we simply develop the fluctuations related to the total energy

$$\begin{aligned} D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r+\frac{1}{2},+} &= \mathcal{E}_h^r D_h^{r+\frac{1}{2},-} + \mathcal{E}_h^{r+1} D_h^{r+\frac{1}{2},+} + \mathcal{E}_{hv_i}^r D_{hv_i}^{r+\frac{1}{2},-} + \mathcal{E}_{hv_i}^{r+1} D_{hv_i}^{r+\frac{1}{2},+} \\ &\quad + \mathcal{E}_{hv_1}^r R_{i,1}^{r+\frac{1}{2},-} + \mathcal{E}_{hv_1}^{r+1} R_{i,1}^{r+\frac{1}{2},+} + \mathcal{E}_{Q_{ik}}^r D_{Q_{ik}}^{r+\frac{1}{2},-} + \mathcal{E}_{Q_{ik}}^{r+1} D_{Q_{ik}}^{r+\frac{1}{2},+} \\ &= \mathcal{E}_{1,h}^r \left( f_h^{r+\frac{1}{2}} - f_h^r \right) - \mathcal{E}_{1,h}^{r+1} \left( f_h^{r+\frac{1}{2}} - f_h^{r+1} \right) \\ &\quad + \mathcal{E}_{hv_i}^r \left( f_{hv_i}^{r+\frac{1}{2}} - f_{hv_i}^r \right) + \mathcal{E}_{hv_i}^{r+1} \left( f_{hv_i}^{r+\frac{1}{2}} - f_{hv_i}^{r+1} \right) \\ &\quad + \mathcal{E}_{2,h}^r D_h^{r+\frac{1}{2},-} + \mathcal{E}_{2,h}^{r+1} D_h^{r+\frac{1}{2},+} + \mathcal{E}_{hv_1}^r \left( R_{i,1}^{r+\frac{1}{2}} - R_{i,1}^r \right) \\ &\quad + \mathcal{E}_{hv_1}^{r+1} \left( R_{i,1}^{r+1} - R_{i,1}^{r+\frac{1}{2}} \right) + \mathcal{E}_{Q_{ik}}^r D_{Q_{ik}}^{r+\frac{1}{2},-} + \mathcal{E}_{Q_{ik}}^{r+1} D_{Q_{ik}}^{r+\frac{1}{2},+}. \end{aligned} \quad (125)$$

Reordering black terms and using (97) and (113) for the red ones, we obtain

$$\begin{aligned} D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r+\frac{1}{2},+} &= - \left( \mathcal{E}_{1,h}^{r+1} - \mathcal{E}_{1,h}^r \right) f_h^{r+\frac{1}{2}} + \mathcal{E}_{1,h}^{r+1} f_h^{r+1} - \mathcal{E}_{1,h}^r f_h^r \\ &\quad - \left( \mathcal{E}_{hv_i}^{r+1} - \mathcal{E}_{hv_i}^r \right) f_{hv_i}^{r+\frac{1}{2}} + \mathcal{E}_{hv_i}^{r+1} f_{hv_i}^{r+1} - \mathcal{E}_{hv_i}^r f_{hv_i}^r \\ &\quad + E_2^r D_h^{r+\frac{1}{2},-} + E_2^{r+1} D_h^{r+\frac{1}{2},+} \\ &\quad + h^{r+1} Q_{im}^{r+1} Q_{1m}^{r+1} v_i^{r+1} - h^r Q_{im}^r Q_{1m}^r v_i^r \\ &\quad + h^r Q_{ik}^r \frac{1}{2} \tilde{v}_1^{r+\frac{1}{2}} \left( Q_{1k}^{r+1} - Q_{1k}^r \right) + h^{r+1} Q_{ik}^{r+1} \frac{1}{2} \tilde{v}_1^{r+\frac{1}{2}} \left( Q_{1k}^{r+1} - Q_{1k}^r \right) \\ &= - \left( \mathbf{p}^{r+1} - \mathbf{p}^r \right) \cdot \mathbf{f}_q^{r+\frac{1}{2}} + \mathbf{p}^{r+1} \cdot \mathbf{f}^{r+1} - \mathbf{p}^r \cdot \mathbf{f}^r \\ &\quad + E_2^r D_h^{r+\frac{1}{2},-} + E_2^{r+1} D_h^{r+\frac{1}{2},+} \\ &\quad + h^{r+1} Q_{im}^{r+1} Q_{1m}^{r+1} v_i^{r+1} - h^r Q_{im}^r Q_{1m}^r v_i^r \\ &\quad + h^r Q_{ik}^r \frac{1}{2} \tilde{v}_1^{r+\frac{1}{2}} \left( Q_{1k}^{r+1} - Q_{1k}^r \right) + h^{r+1} Q_{ik}^{r+1} \frac{1}{2} \tilde{v}_1^{r+\frac{1}{2}} \left( Q_{1k}^{r+1} - Q_{1k}^r \right). \end{aligned} \quad (126)$$

Finally, taking into account (101) with (103) in the above expression and the discretization of  $\mathbf{f}_q^{r+\frac{1}{2}}$  given in (68), (81), we conclude

$$\begin{aligned}
 D_{\mathcal{E}}^{r+\frac{1}{2},-} + D_{\mathcal{E}}^{r+\frac{1}{2},+} &= -(vL_1)^{r+1} + (vL_1)^r + \mathbf{p}^{r+1} \cdot \mathbf{f}^{r+1} - \mathbf{p}^r \cdot \mathbf{f}^r \\
 &\quad + (hv)_1^{r+1} E_2^{r+1} - (hv)_1^r E_2^r \\
 &\quad + h^{r+1} Q_{im}^{r+1} Q_{1m}^{r+1} v_i^{r+1} - h^r Q_{im}^r Q_{1m}^r v_i^r \\
 &= (\mathbf{p}^{r+1} \cdot \mathbf{f}^{r+1} - (vL_1)^{r+1}) - (\mathbf{p}^r \cdot \mathbf{f}^r - (vL_1)^r) \\
 &\quad + (hv)_1^{r+1} E_2^{r+1} - (hv)_1^r E_2^r \\
 &\quad + h^{r+1} Q_{im}^{r+1} Q_{1m}^{r+1} v_i^{r+1} - h^r Q_{im}^r Q_{1m}^r v_i^r \\
 &= F_G^{r+1} - F_G^r + (hv)_1^{r+1} E_2^{r+1} - (hv)_1^r E_2^r \\
 &\quad + h^{r+1} Q_{im}^{r+1} Q_{1m}^{r+1} v_i^{r+1} - h^r Q_{im}^r Q_{1m}^r v_i^r \\
 &= F^{r+1} - F^r,
 \end{aligned} \tag{127}$$

which is the sought total energy flux difference. Therefore, under the hypothesis that the total energy fluxes on the boundary are zero, we have

$$\int_{\Omega} \frac{d\mathcal{E}}{dt} dx = \sum_r \Delta x \frac{d\mathcal{E}^r}{dt} = - \sum_r (F^{r+1} - F^r) + \sum_r \left( g_{\mathcal{E}}^{r+\frac{1}{2}} - g_{\mathcal{E}}^{r-\frac{1}{2}} \right) = 0, \tag{128}$$

hence the scheme is marginally stable in the energy norm.  $\square$

## 4 Path-Conservative ADER Discontinuous Galerkin Schemes

In this section we briefly recall ADER-DG schemes on rectangular equidistant Cartesian grids with a posteriori subcell finite volume limiter (SCL). The governing PDE system (14)–(17) can be cast into the following general form

$$\frac{\partial \mathbf{q}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{q}, \nabla \mathbf{q}) + \mathbf{B}(\mathbf{q}) \cdot \nabla \mathbf{q} = \mathbf{S}(\mathbf{q}, \nabla \mathbf{q}), \tag{129}$$

with  $\mathbf{x} = (x_1, x_2) \in \Omega$  the coordinate vector in the two-dimensional domain  $\Omega \subset \mathbb{R}^2$ ,  $t \in \mathbb{R}_0^+$  the time, the state vector  $\mathbf{q} \in \Omega_{\mathbf{q}} \subset \mathbb{R}^m$ , the state space or phase-space  $\Omega_{\mathbf{q}} \subset \mathbb{R}^m$ , the flux tensor  $\mathbf{F}(\mathbf{q}, \nabla \mathbf{q}) = (\mathbf{f}, \mathbf{g})$ , the nonconservative product  $\mathbf{B}(\mathbf{q}) \cdot \nabla \mathbf{q} = \mathbf{B}_1(\mathbf{q}) \partial_x \mathbf{q} + \mathbf{B}_2(\mathbf{q}) \partial_y \mathbf{q}$  and the source term  $\mathbf{S}(\mathbf{q}, \nabla \mathbf{q})$ , which may also depend on gradients of the state vector. The general structure (129) is needed if we want to discretize also directly the underlying thermodynamically compatible *viscous system* (44)–(47), including the production term  $T_{ik}$ .

In order to solve very general nonlinear time-dependent PDE systems like (129) numerically, in this paper we employ the family of high order accurate fully-discrete path-conservative one-step ADER discontinuous Galerkin schemes supplemented with an a posteriori subcell finite volume limiter, see e.g. [17,41,42,50,56,124]. In the next sections we provide a brief description of the method and the reader is referred to the above references for more details. Concerning more details on the framework of a posteriori limiting (MOOD), the reader is referred to [29,38,39].

#### 4.1 Unlimited High Order ADER-DG Schemes

The system (129) is discretized on a domain  $\Omega$  making use of a uniform Cartesian grid with elements  $\Omega_i = [x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2}] \times [y_i - \frac{\Delta y}{2}, y_i + \frac{\Delta y}{2}]$ . Here,  $\mathbf{x}_i = (x_i, y_i)$  is the barycenter of  $\Omega_i$  and  $\Delta x$  and  $\Delta y$  are the mesh spacings in the  $x$  and in the  $y$  direction, respectively. The numerical solution of (129) is defined in the space of piecewise polynomials of degree  $N$  and is denoted by  $\mathbf{u}_h(\mathbf{x}, t^n)$ . For each element  $\Omega_i$  it is sought under the form

$$\mathbf{u}_h(\mathbf{x}, t^n) = \varphi_l(\mathbf{x}) \hat{\mathbf{u}}_{l,i}^n, \quad \mathbf{x} \in \Omega_i. \quad (130)$$

Here,  $\varphi_l(\mathbf{x}) = \varphi_{l_1}(\xi)\varphi_{l_2}(\eta)$  are the basis or ansatz functions, which are tensor products of one-dimensional ansatz functions  $\varphi_{l_m}(\chi)$  on the unit reference element  $\chi \in \Omega_{\text{ref}} = [0, 1]$ . The mapping from the reference element to the physical one reads  $x = x_i - \frac{1}{2}\Delta x + \xi\Delta x$  and  $y = y_i - \frac{1}{2}\Delta y + \eta\Delta y$  with  $0 \leq \xi, \eta \leq 1$ . The multi-index  $l = (l_1, l_2)$  refers to the one-dimensional basis functions  $\varphi_{l_m}$  that are employed in the tensor product. The basis functions on the reference element are defined as the Lagrange interpolation polynomials that pass through the Gauss-Legendre quadrature nodes of a Gaussian quadrature formula with  $N + 1$  quadrature points. This choice automatically leads to an *orthogonal* nodal basis.

Multiplication of (129) by test functions  $\varphi_k$ , which according to the Galerkin approach are chosen identical to the ansatz functions, and integration over  $\Omega_i \times [t^n, t^{n+1}]$  leads to

$$\int_{t^n}^{t^{n+1}} \int_{\Omega_i} \varphi_k (\partial_t \mathbf{q} + \nabla \cdot \mathbf{F}(\mathbf{q}, \nabla \mathbf{q}) + \mathbf{B}(\mathbf{q}) \cdot \nabla \mathbf{q}) \, d\mathbf{x} \, dt = \int_{t^n}^{t^{n+1}} \int_{\Omega_i} \varphi_k \mathbf{S}(\mathbf{q}, \nabla \mathbf{q}) \, d\mathbf{x} \, dt. \quad (131)$$

Using (130) and integration by parts yields

$$\begin{aligned} & \left( \int_{\Omega_i} \varphi_k \varphi_l \, d\mathbf{x} \right) (\hat{\mathbf{u}}_{l,i}^{n+1} - \hat{\mathbf{u}}_{l,i}^n) + \int_{t^n}^{t^{n+1}} \int_{\partial\Omega_i} \varphi_k (\mathcal{G}(\mathbf{q}_h^-, \mathbf{q}_h^+) + \mathcal{D}(\mathbf{q}_h^-, \mathbf{q}_h^+)) \cdot \mathbf{n} \, dS \, dt \\ & - \int_{t^n}^{t^{n+1}} \int_{\Omega_i} \nabla \varphi_k \cdot \mathbf{F}(\mathbf{q}_h, \nabla \mathbf{q}_h) \, d\mathbf{x} \, dt + \int_{t^n}^{t^{n+1}} \int_{\Omega_i^\circ} \varphi_k \mathbf{B}(\mathbf{q}_h) \cdot \nabla \mathbf{q}_h \, d\mathbf{x} \, dt = \\ & \int_{\Omega_i} \varphi_k \mathbf{S}(\mathbf{q}_h, \nabla \mathbf{q}_h) \, d\mathbf{x} \, dt, \end{aligned} \quad (132)$$

where  $\mathbf{n}$  is the outward-pointing unit normal vector at the cell boundary  $\partial\Omega_i$ , and  $\mathbf{q}_h$  is a local space-time predictor, the computation of which will be briefly explained later. Since in the DG framework the discrete solution is allowed to jump between two neighboring cells, a numerical flux is required on the boundary. For an exhaustive overview of numerical fluxes and Riemann solvers, see [115]. In this paper, we use the simple Rusanov-type flux

$$\mathcal{G}(\mathbf{q}_h^-, \mathbf{q}_h^+) \cdot \mathbf{n} = \frac{1}{2} (\mathbf{F}(\mathbf{q}_h^+, \nabla \mathbf{q}_h^+) + \mathbf{F}(\mathbf{q}_h^-, \nabla \mathbf{q}_h^-)) \cdot \mathbf{n} - \frac{1}{2} s_{\max} \mathbf{I} (\mathbf{q}_h^+ - \mathbf{q}_h^-), \quad (133)$$

with  $s_{\max} = \max(|\lambda_k(\mathbf{q}_h^-)|, |\lambda_k(\mathbf{q}_h^+)|) + \varepsilon(2N + 1)/\Delta x$  being an estimate of the maximum signal speed at the interface, including also the viscous terms with viscosity coefficient  $\varepsilon$ , see [65]. In (133)  $\mathbf{q}_h^-$  and  $\mathbf{q}_h^+$  denote the boundary-extrapolated values of the space-time predictor from within the element and its neighbor, respectively. The non conservative products are

discretized via a path conservative scheme, as forwarded by Castro, Parés and collaborators in [21–24,86,89] and which are based on the theory established in [85]. The term  $\mathcal{D}(\mathbf{q}_h^-, \mathbf{q}_h^+)$  contains the jump in the non-conservative product and is computed at the aid of a path integral in phase space between the states  $\mathbf{q}_h^-$  and  $\mathbf{q}_h^+$ . Using the simple segment path

$$= (\mathbf{q}_h^-, \mathbf{q}_h^+, s) = \mathbf{q}_h^- + s(\mathbf{q}_h^+ - \mathbf{q}_h^-), \quad s \in [0, 1], \quad (134)$$

the path integral reduces to

$$\mathcal{D}(\mathbf{q}_h^-, \mathbf{q}_h^+) \cdot \mathbf{n} = \frac{1}{2} \left( \int_0^1 \mathbf{B}((\mathbf{q}_h^-, \mathbf{q}_h^+, s)) \cdot \mathbf{n} ds \right) \cdot (\mathbf{q}_h^+ - \mathbf{q}_h^-). \quad (135)$$

The integral (135) is approximated via a simple trapezoidal quadrature rule. The use of path integrals based on the straight-line segment path is the common point between the path-conservative ADER-DG scheme presented here and the thermodynamically compatible semi-discrete finite volume method presented in the previous section.

Following [17,41,43,50] the predictor  $\mathbf{q}_h(\mathbf{x}, t)$  is obtained at the aid of a weak formulation of (129) in space-time, which allows to completely avoid the Cauchy-Kovalevskaya procedure that was originally employed in ADER schemes, see [20,113,114,118,119].

The predictor solution is defined at the aid of space-time ansatz functions  $\theta_l = \theta_l(\mathbf{x}, t) = \varphi_{l_0}(\tau)\varphi_{l_1}(\xi)\varphi_{l_2}(\eta)$ , which are again tensor products of the 1D basis functions  $\varphi_{l_m}(\chi)$  and where now an additional temporal basis function is included, with  $t = t^n + \tau \Delta t$ :

$$\mathbf{q}_h(\mathbf{x}, t) = \theta_l(\mathbf{x}, t) \hat{\mathbf{q}}_{l,i}^n, \quad (136)$$

Multiplication of (129) by  $\theta_k$  and integration over  $\Omega_i \times [t^n, t^{n+1}]$  yields

$$\begin{aligned} & \int_{t^n}^{t^{n+1}} \int_{\Omega_i} \theta_k \partial_t \mathbf{q}_h d\mathbf{x} dt + \int_{t^n}^{t^{n+1}} \int_{\Omega_i} \theta_k \nabla \cdot \mathbf{F}(\mathbf{q}_h, \nabla \mathbf{q}_h) d\mathbf{x} dt \\ & + \int_{t^n}^{t^{n+1}} \int_{\Omega_i^\circ} \theta_k \mathbf{B}(\mathbf{q}_h) \cdot \nabla \mathbf{q}_h d\mathbf{x} dt = \int_{\Omega_i} \theta_k \mathbf{S}(\mathbf{q}_h, \nabla \mathbf{q}_h) d\mathbf{x} dt \end{aligned} \quad (137)$$

and after integration by parts one obtains the final weak form in space-time:

$$\begin{aligned} & \int_{\Omega_i} \theta_k(\mathbf{x}, t^{n+1}) \mathbf{q}_h(\mathbf{x}, t^{n+1}) d\mathbf{x} - \int_{\Omega_i} \theta_k(\mathbf{x}, t^n) \mathbf{u}_h(\mathbf{x}, t^n) d\mathbf{x} - \int_{t^n}^{t^{n+1}} \int_{\Omega_i} \partial_t \theta_k \mathbf{q}_h d\mathbf{x} dt + \\ & \int_{t^n}^{t^{n+1}} \int_{\Omega_i} \theta_k \nabla \cdot \mathbf{F}(\mathbf{q}_h, \nabla \mathbf{q}_h) d\mathbf{x} dt \end{aligned}$$

$$\begin{aligned}
& + \int_{t^n}^{t^{n+1}} \int_{\Omega_i^o} \theta_k \mathbf{B}(\mathbf{q}_h) \cdot \nabla \mathbf{q}_h \, d\mathbf{x} \, dt = \\
& \int_{\Omega_i} \theta_k \mathbf{S}(\mathbf{q}_h, \nabla \mathbf{q}_h) \, d\mathbf{x} \, dt.
\end{aligned}
\tag{138}$$

Equation (138) is a nonlinear element-local algebraic system in the unknowns  $\hat{\mathbf{q}}_{l,i}^n$ , while the coefficients  $\hat{\mathbf{u}}_{l,i}^n$  are the known from  $\mathbf{u}_h(\mathbf{x}, t^n)$  at the previous time. The solution of (138) is obtained by an iterative algorithm, whose convergence was proven in [17] for the case of hyperbolic conservation laws without non-conservative products and without source terms. Concerning the choice of a suitable initial guess for the unknown space-time coefficients  $\hat{\mathbf{q}}_{l,i}^n$ , the reader is referred to [55, 79]. This completes the description of the unlimited ADER-DG scheme.

## 4.2 A Posteriori Subcell Finite Volume Limiter

The numerical scheme presented above is high order accurate and *linear* in the sense of Godunov, hence it will inevitably generate spurious oscillations in the vicinity of shock waves and discontinuities according to the well-known Godunov theorem. In [12, 46, 50, 124] a new a posteriori subcell limiter was introduced for ADER-DG schemes, using the ideas of the MOOD paradigm forwarded in [29, 38, 39] for finite volume schemes.

At the beginning of each time step, the *unlimited* scheme described in the previous section is run on the entire computational domain. This produces a so-called *candidate solution*, in the following denoted by  $\mathbf{u}_h^*(\mathbf{x}, t^{n+1})$ . Next, the candidate solution is a posteriori checked against different numerical and physical detection criteria, such as the positivity of the water depth and of the determinant of  $\mathbf{Q}$ . Furthermore, the absence of floating point errors (NaN) is required and we also require a discrete maximum principle (DMP) to be satisfied, see [50]. If any of these numerical or physical detection criteria is violated, a high order DG cell is marked as troubled and is scheduled for the a posteriori subcell finite volume limiting.

The cells  $\Omega_i$  that have been scheduled for subcell finite volume limiting are now split into  $(2N + 1)^d$  finite volume subcells, which are denoted by  $\Omega_{i,s}$  with  $\Omega_i = \bigcup_s \Omega_{i,s}$ . This subdivision of a high order DG element into many small finite volume subcells does *not* reduce the time step of the DG scheme because the CFL number of explicit discontinuous Galerkin schemes scales with  $1/(2N + 1)$ , while for the finite volume scheme used on the subgrid cells, the maximum Courant number allowed is of the order of unity. At time  $t^n$  the numerical solution in the finite volume subcells  $\Omega_{i,s}$  is represented as usual via *piecewise constant* cell averages denoted by  $\bar{\mathbf{u}}_{i,s}^n$  and which are obtained from the high order DG polynomials  $\mathbf{u}_h(\mathbf{x}, t^n)$  as

$$\bar{\mathbf{u}}_{i,s}^n = \frac{1}{|\Omega_{i,s}|} \int_{\Omega_{i,s}} \mathbf{u}_h(\mathbf{x}, t^n) \, d\mathbf{x}.
\tag{139}$$

These subcell averages are now evolved in time at the aid of a second order MUSCL-Hancock-type TVD finite volume scheme with minmod limiter, which is also a predictor-corrector method and thus looks quite similar to the ADER-DG scheme. The main difference is that now the test function is unity, hence the volume integral over the flux term disappears, and the spatial control volumes  $\Omega_i$  are replaced by the sub-volumes  $\Omega_{i,s}$ :

$$\begin{aligned}
& |\Omega_{i,s}| \left( \bar{\mathbf{u}}_{i,s}^{n+1} - \bar{\mathbf{u}}_{i,s}^n \right) + \int_{t^n}^{t^{n+1}} \int_{\partial\Omega_{i,s}} \left( \mathcal{G}(\mathbf{q}_h^-, \mathbf{q}_h^+) + \mathcal{D}(\mathbf{q}_h^-, \mathbf{q}_h^+) \right) \cdot \mathbf{n} \, dS \, dt \\
& + \int_{t^n}^{t^{n+1}} \int_{\Omega_{i,s}^p} (\mathbf{B}(\mathbf{q}_h) \cdot \nabla \mathbf{q}_h) \, d\mathbf{x} \, dt = \int_{t^n}^{t^{n+1}} \int_{\Omega_{i,s}} \mathbf{S}(\mathbf{q}_h, \nabla \mathbf{q}_h) \, d\mathbf{x} \, dt, \quad (140)
\end{aligned}$$

where the local space-time predictor  $\mathbf{q}_h$  is now easily obtained from the Cauchy-Kovalevskaya procedure, see [115]. Once the cell averages  $\bar{\mathbf{u}}_{i,s}^{n+1}$  of all subcells contained within cell  $\Omega_i$  have been computed at the new time  $t^{n+1}$  according to equation (140), the limited DG polynomial  $\mathbf{u}'_h(\mathbf{x}, t^{n+1})$  at time  $t^{n+1}$  can be simply obtained via a constrained least squares reconstruction. For this we require that

$$\frac{1}{|\Omega_{i,s}|} \int_{\Omega_{i,s}} \mathbf{u}'_h(\mathbf{x}, t^{n+1}) \, d\mathbf{x} = \bar{\mathbf{u}}_{i,s}^{n+1} \quad \forall \Omega_{i,s} \in \Omega_i, \quad (141)$$

and

$$\int_{\Omega_i} \mathbf{u}'_h(\mathbf{x}, t^{n+1}) \, d\mathbf{x} = \sum_{\Omega_{i,s} \in \Omega_i} |\Omega_{i,s}| \bar{\mathbf{u}}_{i,s}^{n+1}. \quad (142)$$

The constraint (142) means conservation of the solution within the element  $\Omega_i$ . In addition to the coefficients  $\bar{\mathbf{u}}_{i,l}^{n+1}$  of the limited DG polynomial, in all limited DG cells we also keep in memory the finite volume subcell averages  $\bar{\mathbf{u}}_{i,s}^{n+1}$ , since they serve as initial condition for the subcell finite volume limiter in the case when a cell is troubled also in the next time step, see [50]. This completes the description of the a posteriori subcell finite volume limiter. For more details, the reader is referred to [46, 50, 124].

### 4.3 Renormalization of $\mathbf{Q}$

In order to maintain a strict compatibility of the discrete energy conservation law (18) with the discrete trace of  $\mathbf{P}$ , for the *inviscid* case  $\varepsilon = 0$  we proceed as follows: at the end of each time step, we compute in each degree of freedom of the DG scheme and in each control volume of the subcell finite volume limiter the trace of  $\mathbf{P}$  from the total energy and subsequently rescale  $\mathbf{Q}$  according to

$$(\text{tr} \mathbf{P})_l^{n+1} = 2(hE)_l^{n+1}/h_l^{n+1} - gh_l^{n+1} - \|\mathbf{v}_l^{n+1}\|^2, \quad (143)$$

$$\tilde{\mathbf{Q}}_l^{n+1} = \mathbf{Q}_l^{n+1} \sqrt{\frac{(\text{tr} \mathbf{P})_l^{n+1}}{\text{tr}(\mathbf{Q}\mathbf{Q}^T)_l^{n+1}}}, \quad (144)$$

where the subscript  $l$  denotes a generic degree of freedom,  $\mathbf{Q}_l^{n+1}$  is the preliminary value as computed from the numerical scheme described previously and  $\tilde{\mathbf{Q}}_l^{n+1}$  is the final result of  $\mathbf{Q}$  after rescaling at the end of each time step.

We stress that the above renormalization (143)–(144) is *not performed* for the case of a *viscous system*, i.e. when  $\varepsilon > 0$ , since for sufficiently fine meshes (well-resolved viscous flow), the compatibility with the energy conservation law (47) must hold *automatically* up to the order of accuracy of the numerical scheme, since Theorem 1 establishes the compatibility of (44)–(46) with (47) at the continuous level.

## 5 Numerical Tests

Throughout this section the gravity constant is set to  $g = 9.81$ . Moreover, we will use SI units :  $m$ ,  $s$ , etc. without writing them explicitly.

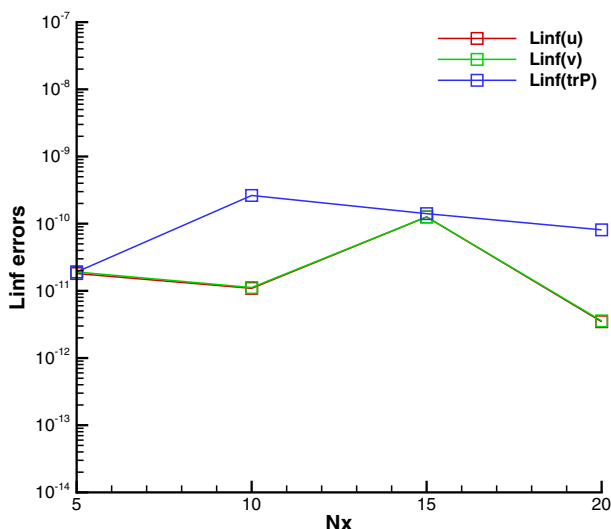
### 5.1 Test Problem with Exact Solution

In the following we solve a test problem suggested in [69], which has an exact solution of the PDE system (1)–(5) that reads

$$h(\mathbf{x}, t) = \frac{h_0}{1 + \beta^2 t^2}, \quad \mathbf{v}(\mathbf{x}, t) = \frac{\beta}{1 + \beta^2 t^2} \begin{pmatrix} +y + \beta t x \\ -x + \beta t y \end{pmatrix}, \quad (145)$$

$$\mathbf{P}(\mathbf{x}, t) = \frac{1}{(1 + \beta^2 t^2)^2} \begin{pmatrix} \lambda + \gamma \beta^2 t^2 & (\lambda - \gamma) \beta t \\ (\lambda - \gamma) \beta t & \gamma + \lambda \beta^2 t^2 \end{pmatrix}. \quad (146)$$

In our setup we choose the above exact solution at time  $t = 0$  as initial condition for  $h$  and  $\mathbf{v}$ , while  $\mathbf{Q}$  is initialized as  $\mathbf{Q} = \text{diag}(\sqrt{P_{11}}, \sqrt{P_{22}})$ . Our numerical simulations are run until a final time of  $t = 1$  in the computational domain  $\Omega = [-1, 1]^2$  with  $\beta = \lambda = \gamma = 0.1$  and  $h_0 = 1$ . A third order ADER-DG scheme ( $N = 2$ ) is run on a sequence of successively refined meshes of  $N_x = 5, 10, 15, 20$  elements. Since the exact solution is a global polynomial of degree one in space, the spatial discretization is *exact* for this test problem. Since time variations in this test problem are also quite small and since the high order DG schemes require a rather small time step for stability, the overall error that is observed on all meshes, see Fig. 1, is of the order of machine accuracy, as expected.



**Fig. 1** Test case with exact solution: observed errors in  $L_\infty$  norm for the velocity components  $u$  and  $v$  and for  $\text{trP}$  obtained on different meshes

## 5.2 Numerical Convergence Study

Aiming at assessing the behaviour of the ADER-DG methodology, we now consider a manufactured solution test given by

$$\begin{aligned} h(\mathbf{x}, t) &= h_0, \quad h\mathbf{v}(\mathbf{x}, t) = \begin{pmatrix} \sin(x) \cos(y) \cos(t) \\ -\cos(x) \sin(y) \cos(t) \end{pmatrix}, \\ \mathbf{Q}(\mathbf{x}, t) &= q_0 \begin{pmatrix} \sin(x) \cos(y) \cos(t) & -\sin(x) \cos(y) \cos(t) \\ -\cos(x) \sin(y) \cos(t) & \cos(x) \sin(y) \cos(t) \end{pmatrix} \end{aligned} \quad (147)$$

with corresponding total energy,

$$hE(\mathbf{x}, t) = g \frac{h_0^2}{2} + \left( \frac{2}{h_0} + h_0 q_0^2 \right) (\sin^2(x) \cos^2(y) + \cos^2(x) \sin^2(y)) \cos^2(t). \quad (148)$$

Moreover, the former expression for  $\mathbf{Q}(\mathbf{x}, t)$  yields a stress tensor of the form

$$\mathbf{P}(\mathbf{x}, t) = 2q_0^2 \begin{pmatrix} \sin^2(x) \cos^2(y) \cos^2(t) & -\sin(x) \cos(y) \cos^2(t) \cos(x) \sin(y) \\ -\sin(x) \cos(y) \cos^2(t) \cos(x) \sin(y) & \cos^2(x) \sin^2(y) \cos^2(t) \end{pmatrix}. \quad (149)$$

To complete the definition of the problem we set  $h_0 = 1$ ,  $q_0 = 0.5$  and  $C_f = C_r = 0$ . Let us remark that to get the sought solution, (147)–(148), a set of analytical source terms, calculated by substitution of (147)–(148) in (14)–(17), must be added to the right hand side of the original system. The simulation is run until  $t = 0.25$  using ADER-DG schemes of polynomial degrees  $N \in \{2, 3, 4, 5\}$ . The errors in  $L^2$  norm obtained for  $h$ ,  $u$  and  $Q_{11}$  are reported in Table 1. Overall, the expected order of accuracy is reached, see bold numbers in Table 1.

## 5.3 Riemann Problems

It is well-known that the numerical discretization of nonconservative hyperbolic PDE is notoriously difficult, see e.g. [4, 23] for a more detailed discussion. This is particularly true when the nonconservative product is acting across genuinely nonlinear waves. This situation is usually the case in the nonconservative equation (16), which makes its numerical discretization particularly difficult. In [69] a special split scheme was developed, splitting the original system (1)–(5) into two quasi-conservative subsystems and during the solution of each of the subsystems, energy conservation was rigorously enforced. In this paper, two new unsplit schemes have been proposed for the discretization of (14)–(18). In the case of the path-conservative ADER-DG scheme described in Sect. 4, the total energy conservation law is explicitly discretized and for the *inviscid case*  $\varepsilon = 0$  the object  $Q_{ik}$  is *renormalized* at the end of each time step according to (143)–(144), in order to maintain *discrete compatibility* with total energy conservation for each degree of freedom of the DG scheme and for each subcell average in case of the subcell FV limiter. Instead, when the path-conservative ADER-DG scheme is applied to the viscous system with  $\varepsilon > 0$  then *no renormalization* is carried out and the compatibility with the total energy conservation law must be guaranteed by the high order of accuracy of the scheme in combination with a sufficiently fine mesh alone, without any renormalization of  $Q_{ik}$ . As such, the fully resolved direct numerical solution (DNS) of the viscous system (44)–(46), which also includes the production term, with small but not vanishing  $\varepsilon > 0$ , constitutes the highest level of fidelity concerning the discretization

**Table 1**  $L^2$  errors and convergence rates for the manufactured test obtained using the ADER-DG method with  $N \in \{2, 3, 4, 5\}$ . The simulations were run on Cartesian meshes of  $N_x \times N_x$  elements up to time  $t = 0.25$ 

$N_x = N_y$	$L^2(h)$	$\mathcal{O}(h)$	$L^2(u)$	$\mathcal{O}(u)$	$L^2(Q_{11})$	$\mathcal{O}(Q_{11})$
ADER-DG $N = 2$						
8	1.5643E-03		9.2605E-03		6.8172E-03	
16	1.9587E-04	<b>3.00</b>	1.5629E-03	<b>2.57</b>	1.3400E-03	<b>2.35</b>
32	2.3442E-05	<b>3.06</b>	2.8002E-04	<b>2.48</b>	2.4678E-04	<b>2.44</b>
64	3.1021E-06	<b>2.92</b>	5.4213E-05	<b>2.37</b>	4.3402E-05	<b>2.51</b>
ADER-DG $N = 3$						
8	8.8166E-05		2.5480E-04		1.1146E-04	
16	5.6738E-06	<b>3.96</b>	1.5191E-05	<b>4.07</b>	5.4435E-06	<b>4.36</b>
24	1.1747E-06	<b>3.88</b>	2.9993E-06	<b>4.00</b>	1.0004E-06	<b>4.18</b>
32	4.0746E-07	<b>3.68</b>	1.0391E-06	<b>3.68</b>	3.3417E-07	<b>3.81</b>
ADER-DG $N = 4$						
8	3.8841E-06		2.6818E-05		1.8500E-05	
16	1.2852E-07	<b>4.92</b>	1.1450E-06	<b>4.55</b>	9.0778E-07	<b>4.35</b>
24	1.6424E-08	<b>5.07</b>	1.9105E-07	<b>4.42</b>	1.4928E-07	<b>4.45</b>
32	3.7032E-09	<b>5.18</b>	5.4539E-08	<b>4.36</b>	4.1229E-08	<b>4.47</b>
ADER-DG $N = 5$						
4	1.2340E-05		6.1955E-05		2.5864E-05	
8	1.5933E-07	<b>6.28</b>	9.8471E-07	<b>5.98</b>	2.2857E-07	<b>6.82</b>
12	1.1618E-08	<b>6.46</b>	8.5429E-08	<b>6.03</b>	1.4666E-08	<b>6.77</b>
16	2.0895E-09	<b>5.96</b>	1.4195E-08	<b>6.24</b>	2.2897E-09	<b>6.46</b>

of the non-conservative product since the solution can be considered as *smooth* and therefore there are no ambiguities concerning the proper definition of the non-conservative product at all. Instead, the compatible HTC scheme, which implements a semi-discrete Godunov formalism, does *not* directly discretize the energy equation at all, but total energy conservation is a mere consequence of all the other equations at the semi-discrete level. In this case, the discretization of the nonconservative products, of the Reynolds stress tensor and of the viscous terms (the red and blue terms in (44)–(46)) is carried out in such a manner that at the semi-discrete level total energy conservation is automatically ensured.

In the following we solve three Riemann problems on the domain  $\Omega = [0, 1] \times [0, 0.5]$  with initial condition

$$\mathbf{q}(\mathbf{x}, 0) = \begin{cases} \mathbf{q}^L & \text{if } x \leq 0.5, \\ \mathbf{q}^R & \text{if } x > 0.5. \end{cases} \quad (150)$$

The left and right initial states, as well as the final simulation times  $t_{\text{end}}$ , are summarized in Table 2. The values of the state variables not indicated in the table are set to zero, i.e.  $v_1 = 0$ ,  $Q_{12} = 0$  and  $Q_{21} = 0$ . The simulations are run with four different numerical schemes S1–S4:

- (S1) The split scheme of Gavrilyuk *et al.* [69], using a very fine mesh in one space dimension, which serves as a reference solution. This scheme makes explicit use of the total energy conservation law in each subsystem used in the splitting approach.
- (S2) A high order unsplit ADER-DG scheme (132) with polynomial approximation degree  $N = 3$  and 11,  $200 \times 4$  elements, applied to the *viscous system* (44)–(46) with small

**Table 2** Initial left and right states for the Riemann problems RP1–RP3

Test	$h^L$	$h^R$	$v_2^L$	$v_2^R$	$Q_{11}^L$	$Q_{11}^R$	$Q_{12}^L$	$Q_{12}^R$	$Q_{22}^{L,R}$
RP1	0.02	0.01	0	0	0.01	0.01	0	0	$10^{-4}$
RP2	0.01	0.01	+0.01	−0.01	0.02	0.02	0	0	0.02
RP3	0.01	0.01	0	0	0.01	0.01	0.02	0.01	0.01

but positive viscosity parameter  $\varepsilon > 0$  (vanishing viscosity limit). Since  $\varepsilon > 0$  we do *not* solve the energy equation (47) explicitly and apply *no* renormalization to  $Q_{ik}$ . To obtain the discrete compatibility with the energy conservation law (47) and for the proper definition of the nonconservative products, only a sufficiently fine mesh is needed in combination with the high order DG scheme (fully resolved DNS). This approach serves to generate an additional and totally independent reference solution.

- (S3) The new unsplit thermodynamically compatible HTC scheme based on the semi-discrete Godunov formalism described in Sect. 3. This scheme is by construction exactly compatible with the viscous system (44)–(46) at the semi-discrete level and therefore the semi-discrete energy conservation law is a *direct consequence* of the discretization of all the other equations and thus does *not* need to be discretized explicitly again.
- (S4) The high order unsplit ADER-DG scheme applied to the *inviscid* system, setting  $\varepsilon = 0$  in (44)–(47). In this case, the energy equation (47) is explicitly discretized and the object  $Q_{ik}$  is *renormalized* at the end of each timestep according to (143)–(144).

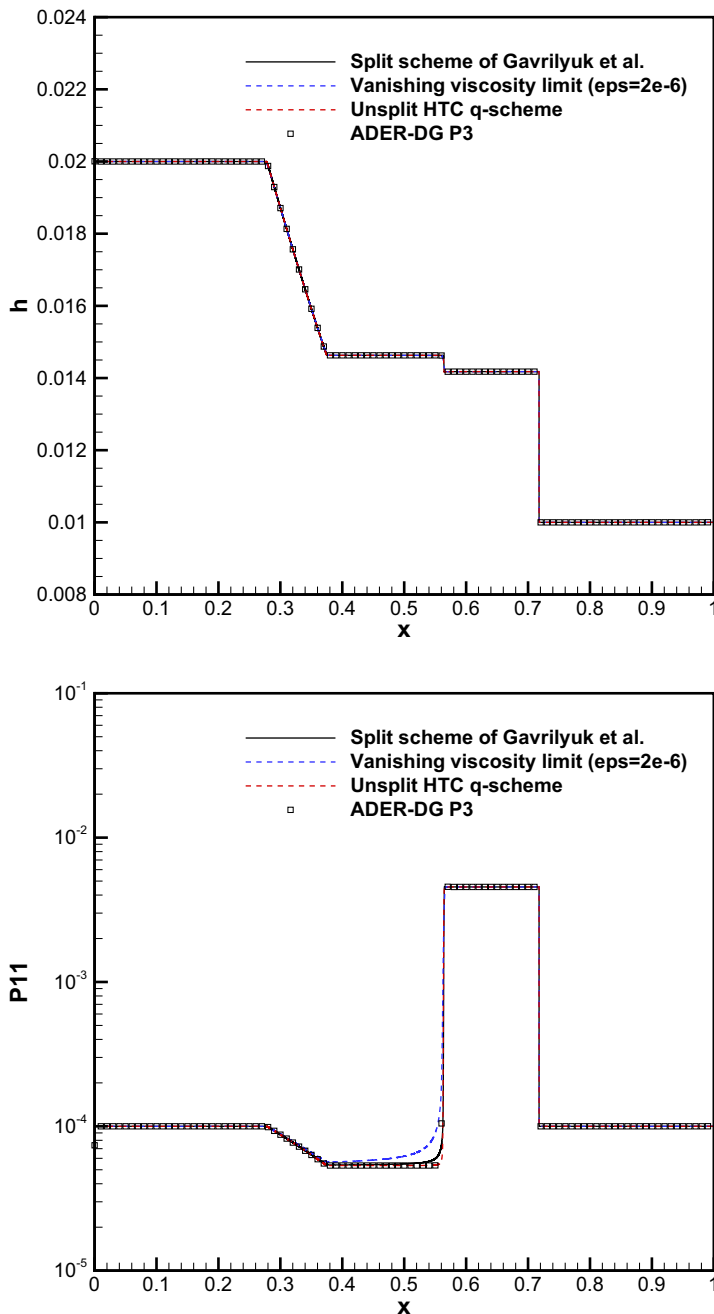
We emphasize that in all four schemes, discrete compatibility with the total energy conservation law is always assured in one way, or in another: either by directly using the total energy conservation equation within the numerical scheme (S1 and S4), or by achieving compatibility exactly at the discrete level (S3). In S2 the compatibility is merely achieved at the aid of negligible discretization errors by using a very high order scheme applied to the viscous system with  $\varepsilon > 0$  and using a sufficiently fine mesh (fully resolved DNS).

The computational results obtained for the three Riemann problems are shown in Figs. 2, 3 and 4, where also the exact mesh resolution is given for each scheme, together with the choice of the viscosity parameter  $\varepsilon$  in the case of the simulation of the vanishing viscosity limit. For all Riemann problems one can note an excellent agreement between the numerical solutions obtained with all four schemes (S1–S4) listed above.

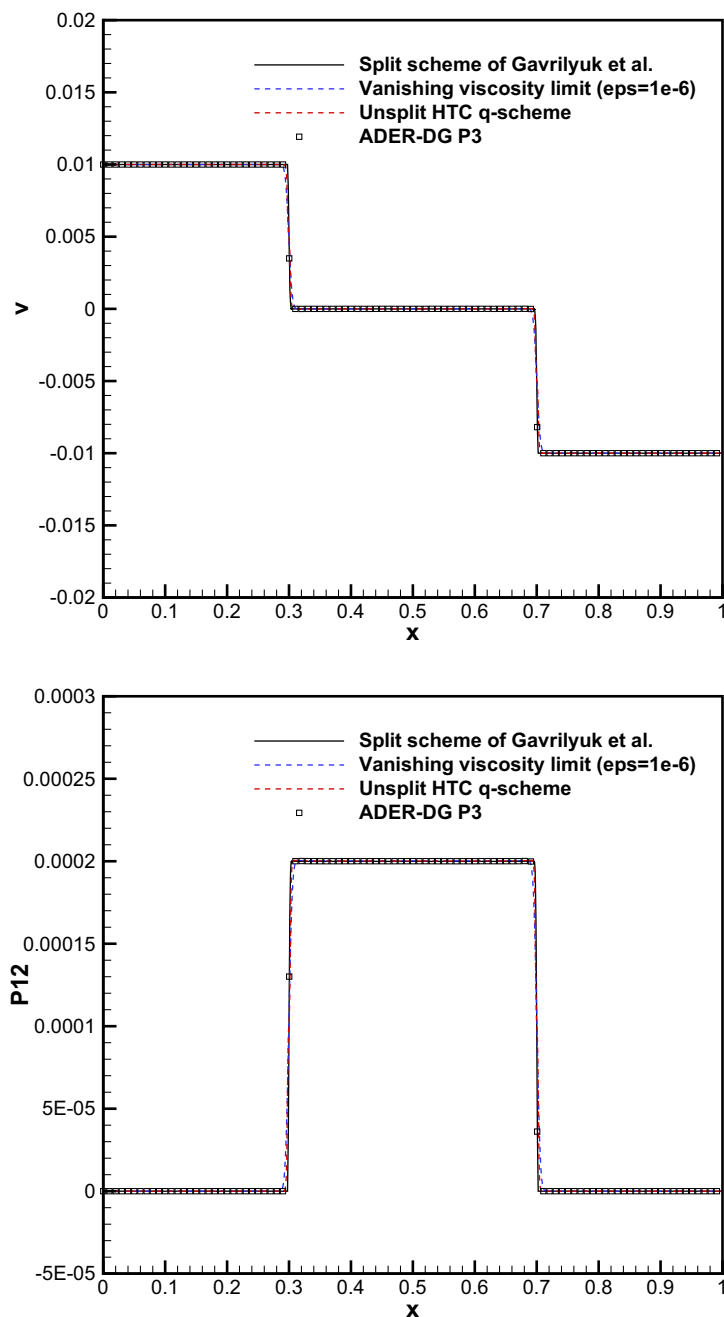
This clearly highlights that *discrete thermodynamic compatibility* is a *key feature* for the correct discretization of nonconservative products that are acting across genuinely nonlinear fields, like the ones present in (16).

## 5.4 One Dimensional Brock Profile

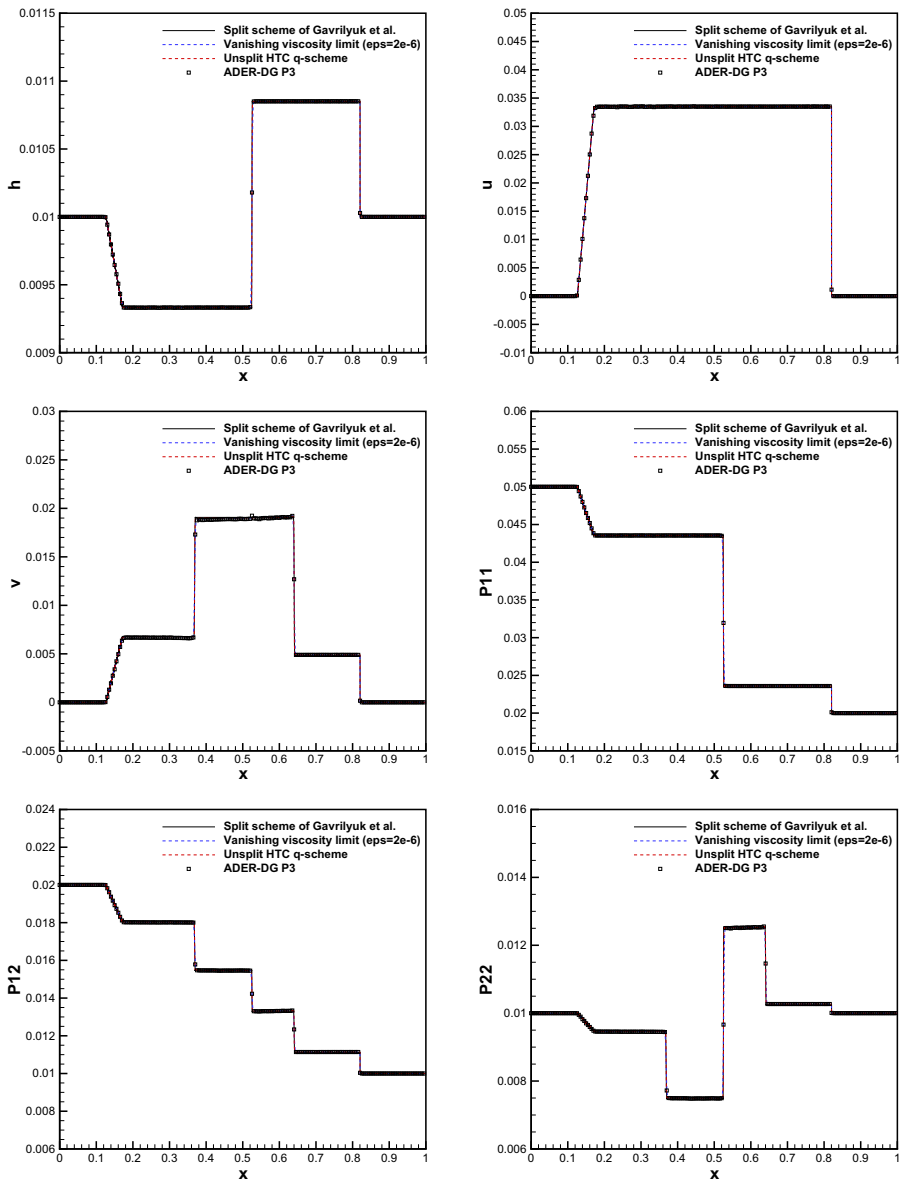
Here we repeat the numerical experiment concerning one-dimensional roll waves proposed in [69] and compare the obtained numerical results with the experimental data provided by Brock in [15, 16]. The initial condition is given according to [69] by  $h = h_0(1 + a \sin(2\pi x/L))$  with  $a = 0.05$ ,  $v_1 = \sqrt{gh_0 \tan \theta / C_f}$ ,  $v_2 = 0$  and  $\mathbf{Q} = \sqrt{\frac{1}{2}\varphi h^2} \mathbf{I}$ . The bottom slope angle of this test problem is  $\theta = 0.05011$  with  $\partial_x b = \tan(\theta)$ , the still water depth is set to  $h_0 = 0.00798$ , the bottom friction coefficient is chosen as  $C_f = 0.0036$ , the parameter  $C_r$  is set to  $C_r = 0.00035$  and  $\varphi = 22.76$ , see [69].



**Fig. 2** Numerical solution of the Riemann problem RP1 obtained with different numerical schemes at time  $t = 0.5$ : split scheme of [69] on 250,000 elements (S1, solid black line); vanishing viscosity limit of the viscous system (44)–(47) with  $\varepsilon = 2 \cdot 10^{-6}$  using a fourth order ADER-DG scheme ( $N = 3$ ) on 11,200 elements (S2, dashed blue line); unsplit thermodynamically compatible  $q$ -scheme on 56,000 elements (S3, dashed red line); fourth order ADER-DG scheme ( $N = 3$ ) applied to the inviscid model (14)–(18) using 1,400 elements (S4, squares) (Color figure online)



**Fig. 3** Numerical solution of the Riemann problem RP2 obtained with different numerical schemes at time  $t = 10$ : split scheme of [69] on 100,000 elements (S1, solid black line); vanishing viscosity limit of the viscous system (44)–(47) with  $\varepsilon = 1 \cdot 10^{-6}$  using a fourth order ADER-DG scheme ( $N = 3$ ) on 10,200 elements (S2, dashed blue line); unsplit thermodynamically compatible  $q$ -scheme on 28,000 elements (S3, dashed red line); fourth order ADER-DG scheme ( $N = 3$ ) applied to the inviscid model (14)–(18) using 1,000 elements (S4, squares) (Color figure online)



**Fig. 4** Numerical solution of the Riemann problem RP3 obtained with different numerical schemes at time  $t = 0.5$ : split scheme of [69] on 250,000 elements (S1, solid black line); vanishing viscosity limit of the viscous system (44)–(47) with  $\varepsilon = 2 \cdot 10^{-6}$  using a fourth order ADER-DG scheme ( $N = 3$ ) on 10,200 elements (S2, dashed blue line); unsplit thermodynamically compatible  $q$ -scheme on 56,000 elements (S3, dashed red line); fourth order ADER-DG scheme ( $N = 3$ ) applied to the inviscid model (14)–(18) using 1,400 elements (S4, squares) (Color figure online)

The computational domain is  $\Omega = [0, L] \times [0, 0.5]$  with  $L = 1.3$  and is discretized with  $104 \times 20$  ADER-DG elements of polynomial approximation degree  $N = 3$ . Periodic boundary conditions are applied in  $x_1$  and  $x_2$  direction. Simulations are run for system (14)–(16) until a final time of  $t = 12.5$ . For this test, the bottom slope term is simply implemented as an algebraic source term in order to be compatible with the periodic boundary conditions. The numerical results obtained with the path-conservative ADER-DG scheme and the experimental profile of Brock are depicted in the left panel of Fig. 5. Overall, we can note a very good agreement between the numerical results and the experimental reference data. In the right panel of Fig. 5 a visualization of the a posteriori subcell limiter is shown (red cells are highlighted in red, while unlimited cells are plotted in blue). It can be noticed that the limiter is only active at the shock wave.

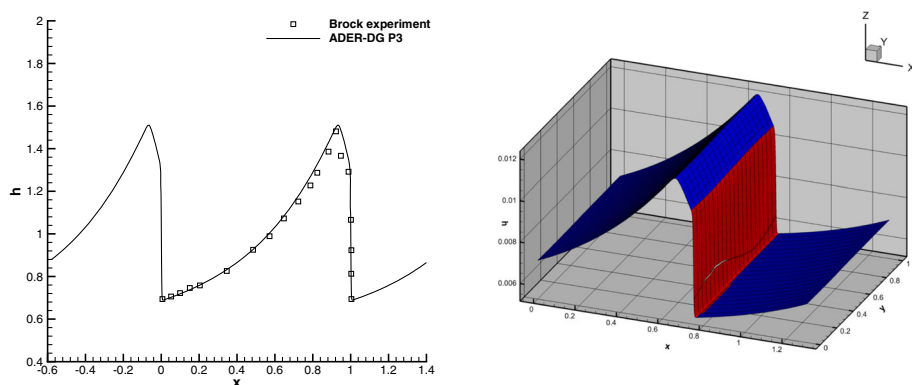
## 5.5 Numerical Simulation of the SWASI Experiment

In this last and most complex numerical test we carry out the simulation of the SWASI experiment proposed by Foglizzo et al. in [61] and which was numerically investigated in [80]. The flow field is turbulent and turbulence is responsible for the developing flow structures. The computational domain is  $\Omega = [-1, +1]^2$  and is discretized with a uniform Cartesian mesh composed of  $256 \times 256$  ADER-DG elements of polynomial approximation degree  $N = 3$ .

The bottom topography of this test is given according to [80] by

$$b(r) = \begin{cases} \frac{A}{L_1^4} \left( (r - R^- - L_1)^2 - L_1^2 \right)^2 & \text{if } R^- \leq r \leq 2L_1 + R^-, \\ (r - R^- - 2L_1) \tan \beta & \text{if } r > 2L_1 + R^-, \end{cases} \quad (151)$$

with  $r = \|\mathbf{x}\|$ ,  $L_1 = 0.02$ ,  $A = 0.005$ ,  $\beta = 0.07$ ,  $R^- = 0.08$  and  $R_1 = \sqrt{2}$ . The model parameters for this test are set to  $C_f = 0.0036$ ,  $C_r = 1$  and  $\varphi = 2$ . The reference inflow discharge is chosen as  $q_0 = 1.2 \cdot 10^{-3}$ , while the reference water depth at  $R_1$  is set to  $h_0 = 0.003$ .



**Fig. 5** Two-dimensional numerical simulation of the roll wave experiment of Brock [15,16] at time  $t = 12.5$ . Left: comparison of a 1D cut through the numerical simulation with the experimental profile. Right: computational grid with troubled cells highlighted in red and unlimited cells colored in blue (Color figure online)

In contrast to [80], in this paper the initial velocity field is chosen to be the *stationary equilibrium* between bottom slope and bottom friction and which satisfies the following ODE in radial direction:

$$\frac{du_r}{dr} = u \frac{C_f |u|^3 r^3 - u g q_0 r^2 \tan \beta - q_0^2 g}{q_0 r (r u^3 + q_0 g)} \quad (152)$$

with initial condition  $u_r(R_1) = -q_0/(h_0 R_1)$  for both regions,  $r \leq R_1$  and  $r > R_1$ . This ODE is solved once at the beginning of the simulation at the aid of a classical fourth order Runge-Kutta scheme. Once the radial velocity  $u_r$  is known, the water depth can be easily computed as  $h = q_0/(r u_r)$  and the final velocity field is given by the equilibrium solution plus a sinusoidal perturbation as follows:

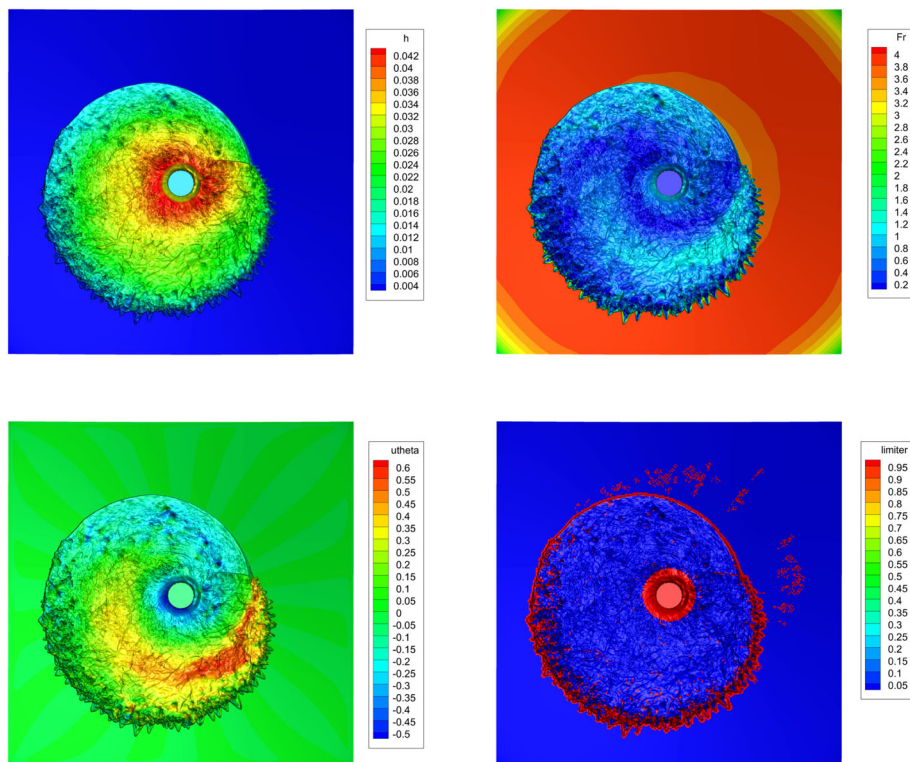
$$v_1 = u_r \cos \theta (1 + d \sin(16\theta)), \quad v_2 = u_r \sin \theta (1 + d \sin(16\theta)), \quad (153)$$

with  $d = 0.005$ . We furthermore set  $\mathbf{Q} = \sqrt{\varphi h_0^2} \mathbf{I}$  as initial condition for the object  $\mathbf{Q}$ . The outflow through the central hole is generated by a sink term, setting  $h = 10^{-2}$ ,  $\mathbf{v} = 0$  and  $\mathbf{Q} = 10^{-5} \mathbf{I}$  for  $r < 0.075$  at all times.

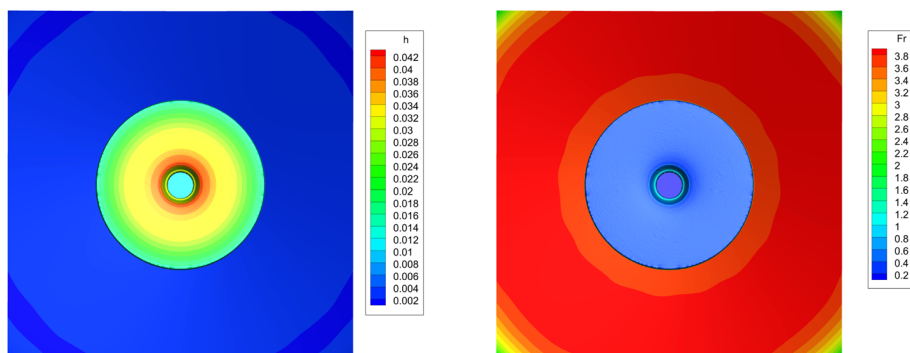
The numerical simulations are carried out until a final time of  $t = 57$ . Figure 6 shows the computational results obtained for the water depth, the Froude number, the angular velocity and the instantaneous a posteriori subcell limiter map, in which red cells are highlighted in red, while unlimited cells are plotted in blue. The limiter is essentially activated along the moving shock front. The obtained numerical results agree qualitatively very well with experimental observations made in [61] and with the numerical results previously presented in [80]. In particular, one can note the characteristic cusp in the shock front that is visible in both, the experiment [61] as well as in the numerical results of [80]. At this point we would like to emphasize that the model formulation as well as the numerical scheme used in this paper are *completely different* compared to the model formulation and the scheme employed in [80]. While in [80] the problem was solved after rewriting the governing PDE system in polar coordinates, here we solve the problem directly in Cartesian coordinates on a Cartesian mesh. Furthermore, the model used in [80] was based on the temporal evolution equation of  $\mathbf{P}$ , while here we use the new model formulation in terms of the object  $\mathbf{Q}$  that guarantees  $\text{tr} \mathbf{P} \geq 0$  by construction. Last but not least, in [80] a split finite volume scheme was used, while the present paper employs an unsplit high order ADER-DG scheme. The fact that the numerical results obtained with different models and different schemes agree well with each other and with experimental observations shows the validity of the different mathematical model formulations as well as of the chosen numerical discretizations. As already pointed out in [80], the same simulation run with the same model parameters ( $C_f = 0.0036$ ) and the same numerical method on the same mesh applied to the simple shallow water equations leads only to a steady circular shock wave, without developing any shock instability and without showing the typical cusp of the SWASI experiment, see Fig. 7.

## 6 Conclusion

In this paper we have introduced a new reformulation of the first order hyperbolic model for unsteady turbulent shallow water flows introduced and studied in [11, 69, 80]. The main idea of the model reformulation proposed in this paper is the decomposition of the specific Reynolds stress tensor  $\mathbf{P}$  at the aid of a new object  $\mathbf{Q}$  so that  $\mathbf{P} = \mathbf{Q} \mathbf{Q}^T$ . This guarantees that  $\text{tr} \mathbf{P} \geq 0$  by construction also at the discrete level for all times, since in terms of  $\mathbf{Q}$  the trace of the Reynolds



**Fig. 6** Numerical simulation of the SWASI experiment with a fourth order ADER-DG scheme at time  $t = 57$  s applied to the model for unsteady turbulent shallow water flows (14)–(18). Water depth (top left), Froude number (top right), angular velocity (bottom left) and limiter map with limited cells highlighted in red and unlimited cells plotted in blue (bottom right) (Color figure online)



**Fig. 7** Numerical simulation of the SWASI experiment with a fourth order ADER-DG scheme at time  $t = 57$  s applied to the classical shallow water equations. Water depth (left) and Froude number (right). With the classical shallow water model no shock wave instability develops

stress tensor, i.e. the turbulent kinetic energy, can be written as  $\text{tr}\mathbf{P} = Q_{ij}Q_{ij} \geq 0$ . Compared to the previous model used in [11,69,80] we also add a thermodynamically compatible viscous flux and an associated entropy production term that together guarantee the compatibility of the viscous system with the total energy conservation law and with the entropy inequality, which in the new reformulation can be simply expressed in terms of an extra conservation law for the determinant of  $\mathbf{Q}$ . Based on the *Godunov form* of hyperbolic conservation laws found by Godunov in his groundbreaking work *An interesting class of quasilinear systems* [70], we have derived a new thermodynamically compatible semi-discrete finite volume scheme that mimics the Godunov form of the inviscid conservative part of the system *exactly* at the semi-discrete level. The proposed schemes can therefore be called a *discrete Godunov formalism*, or a *hyperbolic and thermodynamically compatible* (HTC) finite volume scheme. Subsequently, also a thermodynamically compatible viscous extension of the scheme has been proposed, together with the thermodynamically compatible discretization of the remaining nonconservative terms and of the Reynolds stress tensor, which do not fit into the original Godunov formalism. At this point we stress again that the proposed scheme mimics the underlying viscous system of the mathematical model *exactly* at the semi-discrete level and as such also falls into the class of *structure-preserving schemes*, since all properties of the thermodynamic structure of the governing PDE system are properly maintained by the numerical scheme. The paper also considers high order path-conservative fully-discrete one-step ADER discontinuous Galerkin schemes with a posteriori subcell limiter that can be applied to both, the viscous and the inviscid form of the mathematical model. The performance and accuracy of all schemes is carefully assessed at the aid of three Riemann problems, where also a direct comparison with the scheme introduced in [69] has been shown. An excellent agreement between all different methods was observed in all cases. For the high order ADER-DG schemes a numerical convergence study was carried out at the aid of a manufactured solution, since the analytic solution used in [69] was too simple for a high order DG scheme. The new model was applied to the simulation of roll waves, comparing with the experimental data of Brock and obtaining an excellent level of agreement between the numerical and the experimental results. As a last test problem we have carried out a numerical simulation of the SWASI experiment of Foglizzo *et al.* [61], using the computational setup proposed in [80]. Our simulations show the same cusp in the moving shock front that was already observed in the experiments and in the numerical simulations shown in [80].

In the future we plan to extend the new family of thermodynamically compatible schemes to the equations of nonlinear hyperelasticity [14,67,75,77,87,102] and to the unified hyperbolic model of continuum mechanics [13,17,47,93,102], as well as to hyperbolic reformulations of dispersive systems [7,18,37,58]. Further work will also concern the extension of the discrete Godunov formalism presented in this paper to higher order semi-discrete discontinuous Galerkin finite element schemes, see e.g. [36].

Another open challenge remains the development of thermodynamically compatible schemes like those presented in this paper that also maintain curl and divergence involution constraints *exactly* at the semi-discrete level, similar to the structure-preserving semi-implicit method recently proposed in [13], but which was not thermodynamically compatible.

**Acknowledgements** The research presented in this paper has been financed by the European Union's Horizon 2020 Research and Innovation Programme under the project *ExaHyPE*, Grant No. 671698 (call FETHPC-1-2014). S.B. and M.D. are both members of the INdAM GNCS group and acknowledge the financial support received from the Italian Ministry of Education, University and Research (MIUR) in the frame of the Departments of Excellence Initiative 2018–2022 attributed to DICAM of the University of Trento (Grant L. 232/2016) and in the frame of the PRIN 2017 project *Innovative numerical methods for evolutionary partial differential equations and applications*. S.B. was also funded by INdAM via a GNCS grant for young researchers and by

an *UniTN starting grant* of the University of Trento. S.G. has been partially funded by the Excellence Initiative of Aix-Marseille University A\*Midex, a French Investissements d'Avenir programme AMX-19-IET-010, and the Russian Science Foundation (Project 20-11-20189).

**Funding** Open access funding provided by Università degli Studi di Trento within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Convexity of the Energy

Consider the total energy as a function of  $\mathbf{s} = (h, r_i = hv_i, S_{ij} = hQ_{ij})^T$ :

$$\mathcal{E} = \frac{gh^2}{2} + \frac{r_1^2 + r_2^2}{2h} + \frac{S_{11}^2 + S_{12}^2 + S_{21}^2 + S_{22}^2}{2h}.$$

We denote  $\mathcal{E}_{s_i s_j}$  the Hessian matrix with respect to  $\mathbf{s}$ :

$$\begin{pmatrix} g + \frac{r_1^2 + r_2^2}{h^3} + \frac{S_{11}^2 + S_{12}^2 + S_{21}^2 + S_{22}^2}{h^3} & -\frac{r_1}{h^2} & -\frac{r_2}{h^2} & -\frac{S_{11}}{h^2} & -\frac{S_{12}}{h^2} & -\frac{S_{21}}{h^2} & -\frac{S_{22}}{h^2} \\ -\frac{r_1}{h^2} & \frac{1}{h} & 0 & 0 & 0 & 0 & 0 \\ -\frac{r_2}{h^2} & 0 & \frac{1}{h} & 0 & 0 & 0 & 0 \\ -\frac{S_{11}}{h^2} & 0 & 0 & \frac{1}{h} & 0 & 0 & 0 \\ -\frac{S_{12}}{h^2} & 0 & 0 & 0 & \frac{1}{h} & 0 & 0 \\ -\frac{S_{21}}{h^2} & 0 & 0 & 0 & 0 & \frac{1}{h} & 0 \\ -\frac{S_{22}}{h^2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{h} \end{pmatrix}. \quad (154)$$

Using Sylvester's criterion, one can easily show that  $\mathcal{E}_{s_i s_j}$  is positive definite, if  $h > 0$ . Hence, the energy is a convex function of  $\mathbf{s}$ .

Let us remark that if one considers  $\mathcal{E}$  as a function of  $\mathbf{q} = (h, r_i = hv_i, Q_{ij})^T$ , the energy

$$\mathcal{E} = \frac{gh^2}{2} + \frac{r_1^2 + r_2^2}{2h} + h \frac{Q_{11}^2 + Q_{12}^2 + Q_{21}^2 + Q_{22}^2}{2}$$

is not, a priori, a convex function of  $\mathbf{q}$ . Indeed, in this case the Hessian matrix  $\mathcal{E}_{q_i q_j}$  reads

$$\begin{pmatrix} g + \frac{r_1^2 + r_2^2}{h^3} & -\frac{r_1}{h^2} & -\frac{r_2}{h^2} & Q_{11} & Q_{12} & Q_{21} & Q_{22} \\ -\frac{r_1}{h^2} & \frac{1}{h} & 0 & 0 & 0 & 0 & 0 \\ -\frac{r_2}{h^2} & 0 & \frac{1}{h} & 0 & 0 & 0 & 0 \\ Q_{11} & 0 & 0 & \frac{1}{h} & 0 & 0 & 0 \\ Q_{12} & 0 & 0 & 0 & \frac{1}{h} & 0 & 0 \\ Q_{21} & 0 & 0 & 0 & 0 & \frac{1}{h} & 0 \\ Q_{22} & 0 & 0 & 0 & 0 & 0 & \frac{1}{h} \end{pmatrix}. \quad (155)$$

We now use again Sylvester's criterion. The first three principal minors are positive. One can show that if the determinant of  $\mathcal{E}_{q_i q_j}$  is positive, the other principal minors are also positive.

It is equivalent to the inequality  $gh - Q_{11}^2 - Q_{12}^2 - Q_{21}^2 - Q_{22}^2 > 0$ . Thus the determinant is positive if the ‘turbulent’ energy is small compared to  $gh$ . In practice, it is always the case. In the following, for convenience, we consider the energy as a function of  $\mathbf{q} = (h, hv_i, Q_{ij})^T$ .

## References

1. Abbate, E., Iollo, A., Puppo, G.: An asymptotic-preserving all-speed scheme for fluid dynamics and nonlinear elasticity. *SIAM J. Sci. Comput.* **41**, A2850–A2879 (2019)
2. Abgrall, R.: A general framework to construct schemes satisfying additional conservation relations. Application to entropy conservative and entropy dissipative schemes. *J. Comput. Phys.* **372**, 640–666 (2018)
3. Abgrall, R., Bacigaluppi, P., Tokareva, S.: A high-order nonconservative approach for hyperbolic equations in fluid dynamics. *Comput. Fluids* **169**, 10–22 (2018)
4. Abgrall, R., Karni, S.: Computations of compressible multifluids. *J. Comput. Phys.* **169**, 594–623 (2001)
5. Abgrall, R., Karni, S.: A comment on the computation of non-conservative products. *J. Comput. Phys.* **229**, 2759–2763 (2010)
6. Barton, P.T., Drikakis, D., Romenski, E., Titarev, V.A.: Exact and approximate solutions of Riemann problems in non-linear elasticity. *J. Comput. Phys.* **228**, 7046–7068 (2009)
7. Bassi, C., Bonaventura, L., Busto, S., Dumbser, M.: A hyperbolic reformulation of the Serre–Green–Naghdi model for general bottom topographies. *Comput. Fluids* **212**, 104716 (2020)
8. Bassi, C., Busto, S., Dumbser, M.: High order ADER-DG schemes for the simulation of linear seismic waves induced by nonlinear dispersive free-surface water waves. *Appl. Numer. Math.* **158**, 236–263 (2020)
9. Bassi, F., Rebay, S.: A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier–Stokes equations. *J. Comput. Phys.* **131**, 267–279 (1997)
10. Baumann, C.E., Oden, T.J.: A discontinuous hp finite element method for the Euler and the Navier–Stokes equations. *Int. J. Numer. Methods Fluids* **31**, 79–95 (1999)
11. Bhole, A., Nkonga, B., Gavriluk, S., Ivanova, K.: Fluctuation splitting Riemann solver for a non-conservative modeling of shear shallow water flow. *J. Comput. Phys.* **392**, 205–226 (2019)
12. Boscheri, W., Dumbser, M.: Arbitrary-Lagrangian–Eulerian discontinuous Galerkin schemes with a posteriori subcell finite volume limiting on moving unstructured meshes. *J. Comput. Phys.* **346**, 449–479 (2017)
13. Boscheri, W., Dumbser, M., Ioriatti, M., Peshkov, I., Romenski, E.: A structure-preserving staggered semi-implicit finite volume scheme for continuum mechanics. *J. Comput. Phys.* **424**, 109866 (2021)
14. de Brauer, A., Iollo, A., Milcent, T.: A Cartesian scheme for compressible multimaterial hyperelastic models with plasticity. *Commun. Comput. Phys.* **22**, 1362–1384 (2017)
15. Brock, R.: Development of roll-wave trains in open channels. *J. Hydraul. Div.* **95**, 1401–1428 (1969)
16. Brock, R.: Periodic permanent roll waves. *J. Hydraul. Div.* **96**, 2565–2580 (1970)
17. Busto, S., Chiocchetti, S., Dumbser, M., Gaburro, E., Peshkov, I.: High order ADER schemes for continuum mechanics. *Front. Phys.* **8**, 32 (2020)
18. Busto, S., Dumbser, M., Escalante, C., Gavriluk, S., Favrie, N.: On high order ADER discontinuous Galerkin schemes for first order hyperbolic reformulations of nonlinear dispersive systems. *J. Sci. Comput.* **87**, 48 (2021)
19. Busto, S., Tavelli, M., Boscheri, W., Dumbser, M.: Efficient high order accurate staggered semi-implicit discontinuous Galerkin methods for natural convection problems. *Comput. Fluids* **198**, 104399 (2020)
20. Busto, S., Toro, E., Vázquez-Cendón, E.: Design and analysis of ADER-type schemes for model advection–diffusion–reaction equations. *J. Comput. Phys.* **327**, 553–575 (2016)
21. Castro, M., Gallardo, J., López, J., Parés, C.: Well-balanced high order extensions of Godunov’s method for semilinear balance laws. *SIAM J. Numer. Anal.* **46**, 1012–1039 (2008)
22. Castro, M., Gallardo, J., Parés, C.: High-order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. Applications to shallow-water systems. *Math. Comput.* **75**, 1103–1134 (2006)
23. Castro, M., LeFloch, P., Muñoz-Ruiz, M., Parés, C.: Why many theories of shock waves are necessary: convergence error in formally path-consistent schemes. *J. Comput. Phys.* **227**, 8107–8129 (2008)
24. Castro, M.J., Fernández, E., Ferriero, A., García, J.A., Parés, C.: High order extensions of Roe schemes for two dimensional nonconservative hyperbolic systems. *J. Sci. Comput.* **39**, 67–114 (2009)
25. Chandrashekar, P., Nkonga, B., Meena, A.M., Bhole, A.: A path conservative finite volume method for a shear shallow water model. *J. Comput. Phys.* **413**, 109457 (2020)

26. Chatterjee, N., Fjordholm, U.: Convergence of second-order, entropy stable methods for multi-dimensional conservation laws. *ESAIM Math. Model. Numer. Anal.* **54**(4), 1415–1428 (2020)
27. Chavent, G., Cockburn, B.: The local projection  $p^0 - p^1$  discontinuous Galerkin finite element method for scalar conservation laws. *Math. Model. Numer. Anal.* **23**, 565–592 (1989)
28. Cheng, T., Shu, C.: Entropy stable high order discontinuous Galerkin methods with suitable quadrature rules for hyperbolic conservation laws. *J. Comput. Phys.* **345**, 427–461 (2017)
29. Clain, S., Diot, S., Loubère, R.: A high-order finite volume method for systems of conservation laws—multi-dimensional optimal order detection (MOOD). *J. Comput. Phys.* **230**(10), 4028–4050 (2011)
30. Cockburn, B., Hou, S., Shu, C.W.: The Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws IV: the multidimensional case. *Math. Comput.* **54**, 545–581 (1990)
31. Cockburn, B., Lin, S.Y., Shu, C.: TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws III: one dimensional systems. *J. Comput. Phys.* **84**, 90–113 (1989)
32. Cockburn, B., Shu, C.W.: TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws II: general framework. *Math. Comput.* **52**, 411–435 (1989)
33. Cockburn, B., Shu, C.W.: The Runge–Kutta local projection P1-Discontinuous Galerkin finite element method for scalar conservation laws. *Math. Model. Numer. Anal.* **25**, 337–361 (1991)
34. Cockburn, B., Shu, C.W.: The local discontinuous Galerkin method for time-dependent convection diffusion systems. *SIAM J. Numer. Anal.* **35**, 2440–2463 (1998)
35. Cockburn, B., Shu, C.W.: Runge–Kutta discontinuous Galerkin methods for convection-dominated problems. *J. Sci. Comput.* **16**, 173–261 (2001)
36. Derigs, D., Winters, A.R., Gassner, G., Walch, S., Böhm, M.: Ideal GLM-MHD: about the entropy consistent nine-wave magnetic field divergence diminishing ideal magnetohydrodynamics equations. *J. Comput. Phys.* **364**, 420–467 (2018)
37. Dhaouadi, F., Favrie, N., Gavriluk, S.: Extended Lagrangian approach for the defocusing nonlinear Schrödinger equation. *Stud. Appl. Math.* **2018**, 1–20 (2018)
38. Diot, S., Clain, S., Loubère, R.: Improved detection criteria for the multi-dimensional optimal order detection (MOOD) on unstructured meshes with very high-order polynomials. *Comput. Fluids* **64**, 43–63 (2012)
39. Diot, S., Loubère, R., Clain, S.: The MOOD method in the three-dimensional case: very-high-order finite volume method for hyperbolic systems. *Int. J. Numer. Methods Fluids* **73**, 362–392 (2013)
40. Dumbser, M.: Arbitrary high order PNPM schemes on unstructured meshes for the compressible Navier–Stokes equations. *Comput. Fluids* **39**, 60–76 (2010)
41. Dumbser, M., Balsara, D., Toro, E., Munz, C.: A unified framework for the construction of one-step finite-volume and discontinuous Galerkin schemes. *J. Comput. Phys.* **227**, 8209–8253 (2008)
42. Dumbser, M., Castro, M., Parés, C., Toro, E.: ADER schemes on unstructured meshes for non-conservative hyperbolic systems: applications to geophysical flows. *Comput. Fluids* **38**, 1731–1748 (2009)
43. Dumbser, M., Enaux, C., Toro, E.: Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws. *J. Comput. Phys.* **227**, 3971–4001 (2008)
44. Dumbser, M., Facchini, M.: A local space-time discontinuous Galerkin method for Boussinesq-type equations. *Appl. Math. Comput.* **272**, 336–346 (2016)
45. Dumbser, M., Hidalgo, A., Castro, M., Parés, C., Toro, E.: FORCE schemes on unstructured meshes II: non-conservative hyperbolic systems. *Comput. Methods Appl. Mech. Eng.* **199**, 625–647 (2010)
46. Dumbser, M., Loubère, R.: A simple robust and accurate a posteriori sub-cell finite volume limiter for the discontinuous Galerkin method on unstructured meshes. *J. Comput. Phys.* **319**, 163–199 (2016)
47. Dumbser, M., Peshkov, I., Romenski, E., Zanotti, O.: High order ADER schemes for a unified first order hyperbolic formulation of continuum mechanics: viscous heat-conducting fluids and elastic solids. *J. Comput. Phys.* **314**, 824–862 (2016)
48. Dumbser, M., Toro, E.F.: On universal Osher-type schemes for general nonlinear hyperbolic conservation laws. *Commun. Comput. Phys.* **10**, 635–671 (2011)
49. Dumbser, M., Toro, E.F.: A simple extension of the Osher Riemann solver to non-conservative hyperbolic systems. *J. Sci. Comput.* **48**, 70–88 (2011)
50. Dumbser, M., Zanotti, O., Loubère, R., Diot, S.: A posteriori subcell limiting of the discontinuous Galerkin finite element method for hyperbolic conservation laws. *J. Comput. Phys.* **278**, 47–75 (2014)
51. Engsig-Karup, A., Hesthaven, J., Bingham, H., Warburton, T.: DG-FEM solution for nonlinear wave-structure interaction using Boussinesq-type equations. *Coastal Eng.* **55**, 197–208 (2008)

52. Escalante, C., Dumbser, M., Castro, M.: An efficient hyperbolic relaxation system for dispersive non-hydrostatic water waves and its solution with high order discontinuous Galerkin schemes. *J. Comput. Phys.* **394**, 385–416 (2019)
53. Eskilsson, C., Sherwin, S.: An unstructured spectral/hp element model for enhanced Boussinesq-type equations. *Coastal Eng.* **53**, 947–963 (2006)
54. Eskilsson, C., Sherwin, S.: Spectral/hp discontinuous Galerkin methods for modelling 2D Boussinesq equations. *J. Comput. Phys.* **212**, 566–589 (2006)
55. Fambri, F., Dumbser, M., Köppel, S., Rezzolla, L., Zanotti, O.: ADER discontinuous Galerkin schemes for general-relativistic ideal magnetohydrodynamics. *Mon. Not. R. Astron. Soc.* **477**, 4543–4564 (2018)
56. Fambri, F., Dumbser, M., Zanotti, O.: Space-time adaptive ADER-DG schemes for dissipative flows: compressible Navier–Stokes and resistive MHD equations. *Comput. Phys. Commun.* **220**, 297–318 (2017)
57. Favrie, N., Gavriluk, S.: Diffuse interface model for compressible fluid—compressible elastic-plastic solid interaction. *J. Comput. Phys.* **231**, 2695–2723 (2012)
58. Favrie, N., Gavriluk, S.: A rapid numerical method for solving Serre–Green–Naghdi equations describing long free surface gravity waves. *Nonlinearity* **30**, 2718–2736 (2017)
59. Favrie, N., Gavriluk, S., Saurel, R.: Solid–fluid diffuse interface model in cases of extreme deformations. *J. Comput. Phys.* **228**, 6037–6077 (2009)
60. Fjordholm, U., Mishra, S.: Accurate numerical discretizations of non-conservative hyperbolic systems. *ESAIM: Math. Model. Numer. Anal.* **46**(1), 187–206 (2012)
61. Foglizzo, T., Masset, F., Guilet, J., Durand, G.: Shallow water analogue of the standing accretion shock instability: experimental demonstration and a two-dimensional model. *Phys. Rev. Lett.* **108**(5), 051103 (2012)
62. Friedrichs, K.: Symmetric positive linear differential equations. *Commun. Pure Appl. Math.* **11**, 333–418 (1958)
63. Friedrichs, K., Lax, P.: Systems of conservation equations with a convex extension. *Proc. Nat. Acad. Sci. USA* **68**, 1686–1688 (1971)
64. Gallardo, J., Parés, C., Castro, M.: On a well-balanced high-order finite volume scheme for shallow water equations with topography and dry areas. *J. Comput. Phys.* **227**, 574–601 (2007)
65. Gassner, G., Lörcher, F., Munz, C.: A contribution to the construction of diffusion fluxes for finite volume and discontinuous Galerkin schemes. *J. Comput. Phys.* **224**, 1049–1063 (2007)
66. Gassner, G., Winters, A., Kopriva, D.: A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations. *Appl. Math. Comput.* **272**, 291–308 (2016)
67. Gavriluk, S., Favrie, N., Saurel, R.: Modelling wave dynamics of compressible elastic materials. *J. Comput. Phys.* **227**, 2941–2969 (2008)
68. Gavriluk, S., Gouin, H.: Geometric evolution of the Reynolds stress tensor. *Int. J. Eng. Sci.* **59**, 65–73 (2012)
69. Gavriluk, S., Ivanova, K., Favrie, N.: Multi-dimensional shear shallow water flows: problems and solutions. *J. Comput. Phys.* **366**, 252–280 (2018)
70. Godunov, S.: An interesting class of quasilinear systems. *Dokl. Akad. Nauk SSSR* **139**(3), 521–523 (1961)
71. Godunov, S.: Symmetric form of the magnetohydrodynamic equation. *Numer. Methods Mech. Contin. Medium* **3**(1), 26–34 (1972)
72. Godunov, S., Peshkov, I.: Thermodynamically consistent nonlinear model of elastoplastic Maxwell medium. *Comput. Math. Math. Phys.* **50**(8), 1409–1426 (2010)
73. Godunov, S., Romenski, E.: Nonstationary equations of the nonlinear theory of elasticity in Euler coordinates. *J. Appl. Mech. Tech. Phys.* **13**, 868–885 (1972)
74. Godunov, S., Romenski, E.: Thermodynamics, conservation laws, and symmetric forms of differential equations in mechanics of continuous media. In: *Computational Fluid Dynamics Review* 95, pp. 19–31. Wiley, NY (1995)
75. Godunov, S., Romenski, E.: *Elements of Continuum Mechanics and Conservation Laws*. Kluwer Academic/Plenum Publishers, Dordrecht (2003)
76. Godunov, S.K.: Thermodynamic formalization of the fluid dynamics equations for a charged dielectric in an electromagnetic field. *Comput. Math. Math. Phys.* **52**, 787–799 (2012)
77. Godunov, S.K., Romenskii, E.I.: Nonstationary equations of nonlinear elasticity theory in Eulerian coordinates. *J. Appl. Mech. Tech. Phys.* **13**(6), 868–884 (1972)
78. Hennemann, S., Rueda-Ramírez, A., Hindenlang, F., Gassner, G.: A provably entropy stable subcell shock capturing approach for high order split form DG for the compressible Euler equations. *J. Comput. Phys.* **426**, 109935 (2021)

79. Hidalgo, A., Dumbser, M.: ADER schemes for nonlinear systems of stiff advection–diffusion–reaction equations. *J. Sci. Comput.* **48**, 173–189 (2011)
80. Ivanova, K., Gavriluk, S.: Structure of the hydraulic jump in convergent radial flows. *J. Fluid Mech.* **860**, 441–464 (2019)
81. Jackson, H., Nikiforakis, N.: A unified Eulerian framework for multimaterial continuum mechanics. *J. Comput. Phys.* **401**, 109022 (2019)
82. Klaij, C., der Veeg, J.V., der Ven, H.V.: Space-time discontinuous Galerkin method for the compressible Navier–Stokes equations. *J. Comput. Phys.* **217**, 589–611 (2006)
83. Levy, D., Shu, C., Yan, J.: Local discontinuous Galerkin methods for nonlinear dispersive equations. *J. Comput. Phys.* **196**, 751–772 (2004)
84. Liu, Y., Shu, C., Zhang, M.: Entropy stable high order discontinuous Galerkin methods for ideal compressible MHD on structured meshes. *J. Comput. Phys.* **354**, 163–178 (2018)
85. Maso, G.D., LeFloch, P., Murat, F.: Definition and weak stability of nonconservative products. *J. Math. Pures Appl.* **74**, 483–548 (1995)
86. Muñoz, M., Parés, C.: Godunov method for nonconservative hyperbolic systems. *Math. Model. Numer. Anal.* **41**, 169–185 (2007)
87. Ndanou, S., Favrie, N., Gavriluk, S.: Criterion of hyperbolicity in hyperelasticity in the case of the stored energy in separable form. *J. Elast.* **115**, 1–25 (2014)
88. Ndanou, S., Favrie, N., Gavriluk, S.: Multi-solid and multi-fluid diffuse interface model: applications to dynamic fracture and fragmentation. *J. Comput. Phys.* **295**, 523–555 (2015)
89. Parés, C.: Numerical methods for nonconservative hyperbolic systems: a theoretical framework. *SIAM J. Numer. Anal.* **44**, 300–321 (2006)
90. Parés, C., Castro, M.: On the well-balance property of Roe’s method for nonconservative hyperbolic systems. Applications to shallow-water systems. *Math. Model. Numer. Anal.* **38**, 821–852 (2004)
91. Peshkov, I., Boscheri, W., Loubère, R., Romenski, E., Dumbser, M.: Theoretical and numerical comparison of hyperelastic and hypoelastic formulations for Eulerian non-linear elastoplasticity. *J. Comput. Phys.* **387**, 481–521 (2019)
92. Peshkov, I., Pavelka, M., Romenski, E., Grmela, M.: Continuum mechanics and thermodynamics in the Hamilton and the Godunov-type formulations. *Contin. Mech. Thermodyn.* **30**(6), 1343–1378 (2018)
93. Peshkov, I., Romenski, E.: A hyperbolic model for viscous Newtonian flows. *Contin. Mech. Thermodyn.* **28**, 85–104 (2016)
94. Peshkov, I., Romenski, E., Dumbser, M.: Continuum mechanics with torsion. *Contin. Mech. Thermodyn.* **31**, 1517–1541 (2019)
95. Ranocha, H., Dalcin, L., Parsani, M.: Fully discrete explicit locally entropy-stable schemes for the compressible Euler and Navier–Stokes equations. *Comput. Math. Appl.* **80**(5), 1343–1359 (2020)
96. Reed, W., Hill, T.: Triangular mesh methods for neutron transport equation. Tech. Rep. LA-UR-73-479, Los Alamos Scientific Laboratory (1973)
97. Rhebergen, S., Bokhove, O., van der Veeg, J.: Discontinuous Galerkin finite element methods for hyperbolic nonconservative partial differential equations. *J. Comput. Phys.* **227**, 1887–1922 (2008)
98. Rhebergen, S., Cockburn, B.: A space-time hybridizable discontinuous Galerkin method for incompressible flows on deforming domains. *J. Comput. Phys.* **231**, 4185–4204 (2012)
99. Rhebergen, S., Cockburn, B., van der Veeg, J.J.: A space-time discontinuous Galerkin method for the incompressible Navier–Stokes equations. *J. Comput. Phys.* **233**, 339–358 (2013)
100. Richard, G.L., Gavriluk, S.L.: A new model of roll waves: comparison with Brock’s experiments. *J. Fluid Mech.* **698**, 374–405 (2012)
101. Richard, G.L., Gavriluk, S.L.: The classical hydraulic jump in a model of shear shallow-water flows. *J. Fluid Mech.* **725**, 492–521 (2013)
102. Romenski, E.: Hyperbolic systems of thermodynamically compatible conservation laws in continuum mechanics. *Math. Comput. Model.* **28**(10), 115–130 (1998)
103. Romenski, E., Belozarov, A.A., Peshkov, I.M.: Conservative formulation for compressible multiphase flows. *Q. Appl. Math.* **74**(1), 113–136 (2016)
104. Romenski, E., Drikakis, D., Toro, E.: Conservative models and numerical methods for compressible two-phase flow. *J. Sci. Comput.* **42**, 68–95 (2010)
105. Romenski, E., Peshkov, I., Dumbser, M., Fambri, F.: A new continuum model for general relativistic viscous heat-conducting media. *Philos. Trans. R. Soc. A* **378**, 20190175 (2020)
106. Romenski, E., Resnyansky, A., Toro, E.: Conservative hyperbolic formulation for compressible two-phase flow with different phase pressures and temperatures. *Q. Appl. Math.* **65**, 259–279 (2007)
107. Rusanov, V.V.: Calculation of interaction of non-steady shock waves with obstacles. *J. Comput. Math. Phys. USSR* **1**, 267–279 (1961)

108. Tadmor, E.: The numerical viscosity of entropy stable schemes for systems of conservation laws I. *Math. Comput.* **49**, 91–103 (1987)
109. Tavelli, M., Dumbser, M.: A staggered, space-time discontinuous Galerkin method for the three-dimensional incompressible Navier–Stokes equations on unstructured tetrahedral meshes. *J. Comput. Phys.* **319**, 294–323 (2016)
110. Tavelli, M., Dumbser, M.: A pressure-based semi-implicit space-time discontinuous Galerkin method on staggered unstructured meshes for the solution of the compressible Navier–Stokes equations at all Mach numbers. *J. Comput. Phys.* **341**, 341–376 (2017)
111. Tavelli, M., Dumbser, M.: Arbitrary high order accurate space-time discontinuous Galerkin finite element schemes on staggered unstructured meshes for linear elasticity. *J. Comput. Phys.* **366**, 386–414 (2018)
112. Teshukov, V.M.: Gas dynamic analogy for vortex free-boundary flows. *J. Appl. Mech. Tech. Phys.* **48**, 303–309 (2007)
113. Titarev, V., Toro, E.: ADER: arbitrary high order Godunov approach. *J. Sci. Comput.* **17**(1–4), 609–618 (2002)
114. Titarev, V., Toro, E.: ADER schemes for three-dimensional nonlinear hyperbolic systems. *J. Comput. Phys.* **204**, 715–736 (2005)
115. Toro, E.: *Riemann Solvers and Numerical Methods for Fluid Dynamics*, 2nd edn. Springer, Berlin (1999)
116. Toro, E.: *Shock-Capturing Methods for Free-Surface Shallow Flows*. Wiley, New York (2001)
117. Toro, E., Millington, R., Nejad, L.: Towards very high order Godunov schemes. In: Toro, E. (ed.) *Godunov Methods. Theory and Applications*, pp. 905–938. Kluwer/Plenum Academic Publishers, Dordrecht (2001)
118. Toro, E., Titarev, V.: Solution of the generalized Riemann problem for advection-reaction equations. *Proc. R. Soc. Lond.* **458**, 271–281 (2002)
119. Toro, E.F., Titarev, V.A.: Derivative Riemann solvers for systems of conservation laws and ADER methods. *J. Comput. Phys.* **212**(1), 150–165 (2006)
120. van der Vegt, J.J.W., van der Ven, H.: Space-time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows I. General formulation. *J. Comput. Phys.* **182**, 546–585 (2002)
121. van der Ven, H., van der Vegt, J.J.W.: Space-time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows II. Efficient flux quadrature. *Comput. Methods Appl. Mech. Eng.* **191**, 4747–4780 (2002)
122. Yan, J., Shu, C.: A local discontinuous Galerkin method for KdV type equations. *SIAM J. Numer. Anal.* **40**, 769–791 (2002)
123. Yan, J., Shu, C.: Local discontinuous Galerkin methods for partial differential equations with higher order derivatives. *J. Sci. Comput.* **17**, 27–47 (2002)
124. Zanutti, O., Fambri, F., Dumbser, M., Hidalgo, A.: Space-time adaptive ADER discontinuous Galerkin finite element schemes with a posteriori sub-cell finite volume limiting. *Comput. Fluids* **118**, 204–224 (2015)