Thermodynamically consistent viscoelastic phase separation: numerical analysis and simulation

Dissertation

zur Erlangung des Grades
Doktor der Naturwissenschaften

am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität Mainz

Paul Jonathan Strasser

geboren in Wiesbaden

Mainz, Februar 2023

Datum der mündlichen Prüfung: 27. Oktober 2023

Abstract

Soft matter is a significant topic of modern research, since especially polymers and liquid crystals are important in a wide range of technological applications. In this context, it is essential to understand the thermodynamics of polymer solutions. In this work, we present and validate new, efficient, thermodynamically consistent numerical schemes for the simulation of phase separation of polymer-solvent mixtures. The proposed mathematical models are based on a viscoelastic (non-Newtonian) phase-field model by *Zhou, Zhang and E (Physical Review E 73, 2006)*. It consists of the Cahn-Hilliard equation, describing the dynamics of a diffusive interface separating polymer and solvent phase, and extended Oldroyd-B equations for the complex hydrodynamics of a polymer solution. This macroscopic model is isothermal and dissipates energy over time. Therefore, it is consistent with the second law of thermodynamics. Further, it is the first thermodynamically consistent model which reproduces all essential features of experimentally observed viscoelastic phase separation. The main goal of this dissertation is to derive energy-stable numerical schemes for such a complex phase-field model, which are both accurate and computationally efficient. Thus, the proposed schemes shall satisfy the conservation of mass and preserve the thermodynamic consistency of the model equations while suitably linearizing all nonlinear terms. To this end, several problem-specific time and space discretizations will be proposed, and their properties will be discussed. Furthermore, various numerical experiments will be conducted, including experimental convergence tests, to verify the reliability of the proposed numerical schemes. Additionally, to investigate the quality of our numerical solutions describing the physics of viscoelastic phase separation, we perform a comparison to computationally vastly more expensive simulation results of a thoroughly validated mesoscopic model describing the same physical problem. The latter is realized through our collaboration with the Max Planck Institute for Polymer Research in Mainz.

Kurzfassung

Weiche Materie ist ein wichtiges Thema moderner Forschung, denn insbesondere Polymere und Flüssigkristalle sind für vielfältige technologische Anwendungen von Bedeutung. In diesem Zusammenhang ist es essenziell die Thermodynamik von Polymerlösungen zu verstehen. In dieser Arbeit präsentieren und validieren wir neue, effiziente, thermodynamisch konsistente numerische Verfahren zur Simulation von Phasenseparation von Polymer-Lösungsmittel-Gemischen. Die vorgeschlagenen mathematischen Modelle basieren auf einem viskoelastischen (nichtnewtonschen) Phasenfeldmodell von *Zhou, Zhang und E (Physical Review E 73, 2006)*. Dieses besteht aus der Cahn-Hilliard-Gleichung, welche die Dynamik einer diffusiven Grenzfläche zwischen Polymer- und Lösungsmittelphase beschreibt, und erweiterten Oldroyd-B-Gleichungen für die komplexe Hydrodynamik von Polymerlösungen. Dieses makroskopische Modell ist isotherm und gibt Energie über die Zeit ab. Daher steht es im Einklang mit dem zweiten Hauptsatz der Thermodynamik. Außerdem ist es das erste thermodynamisch konsistente Modell, welches alle wesentlichen, experimentell beobachteten Merkmale der viskoelastischen Phasenseparation reproduziert. Das primäre Ziel dieser Dissertation ist die Herleitung von energiestabilen numerischen Verfahren für solch ein komplexes Phasenfeldmodell, welche sowohl akkurat als auch recheneffizient sind. Die vorgeschlagenen Verfahren sollen folglich die Massenerhaltung und die thermodynamische Konsistenz der Modellgleichungen bewahren, während sie alle nichtlinearen Terme geeignet linearisieren. Zu diesem Zweck werden einige problemspezifische Zeit- und Ortsdiskretisierungen vorgeschlagen und ihre Eigenschaften diskutiert. Darüber hinaus werden verschiedene numerische Simulationen durchgeführt, einschließlich experimenteller Konvergenztests, um die Verlässlichkeit der vorgeschlagenen numerischen Verfahren zu verifizieren. Zusätzlich, um zu untersuchen, wie gut unsere numerischen Lösungen die Physik der viskoelastischen Phasenseparation beschreiben, führen wir einen Vergleich mit rechentechnisch wesentlich aufwändigeren Simulationsergebnissen eines gründlich validierten mesoskopischen Modells durch, welches dasselbe physikalische Problem beschreibt. Letzteres wird durch unsere Zusammenarbeit mit dem Max-Planck-Institut für Polymerforschung in Mainz realisiert.

Contents

List of Figures

List of Tables

1 Introduction

Phase separation in binary fluids is a fundamental process in condensed-matter physics. For Newtonian fluids, the phenomenon of spinodal decomposition is reasonably well understood in terms of the so-called *model H*, see, e.g., *Hohenberg and Halperin, Onuki, Bray* [10, 44, 63]. Here, the hydrodynamic equations of motion for mass and momentum conservation are coupled to an advection-diffusion equation for the concentration. The thermodynamic behavior of the concentration, or in general the volume fraction φ, gives rise to a driving force and is described by a free energy functional $E(\varphi)$, see the original paper by *Cahn and Hilliard* [15] or, e.g., *Gurtin, Abels, Depner and Garcke* [1, 41]. A major part of this energy originates from a double-well potential that represents the tendency of a system to have two different stable phases. There are two common choices here, the *Flory-Huggins potential* [32, 46], which is based on the thermodynamics of polymer solutions, and the *Ginzburg-Landau potential* [35]. In the special case of diffuse interface phase-field models, the interface between two phases is a thin layer of finite thickness, across which φ varies continuously. This diffusive interface generates the other major part of the free energy. A big advantage of such models is that interfaces are defined implicitly and do not need to be tracked.

The physics – and therefore, also the mathematics and numerics – becomes more involved if (at least) one component is a macromolecular compound. In this case, the large molecular relaxation time gives rise to a dynamic coupling between intramolecular processes and the demixing on experimentally relevant time scales, with interesting new phenomena, for which the term *viscoelastic phase separation* has been coined. Here, the construction of physically sound dynamic equations with suitable constitutive relations to describe the viscoelasticity is already a challenge in itself. *Helfand and Fredrickson* [43] made a first attempt in this direction, followed by further investigations by, e.g., *Onuki, Doi, Milner, Tanaka and Araki* [58, 60, 62, 68, 69]. However, their proposed models do either not reproduce all the essential features of viscoelastic phase separation observed experimentally or can not be shown to obey the second law of thermodynamics. Therefore, *Zhou, Zhang and E* [74] provided a modified set of the equations by *Tanaka and Araki* [68, 69], which satisfies both.

The essential features of viscoelastic phase separation observed experimentally read, according to *Tanaka* [68]:

- the existence of a frozen period, which is an incubation time for nucleation of the solvent holes,

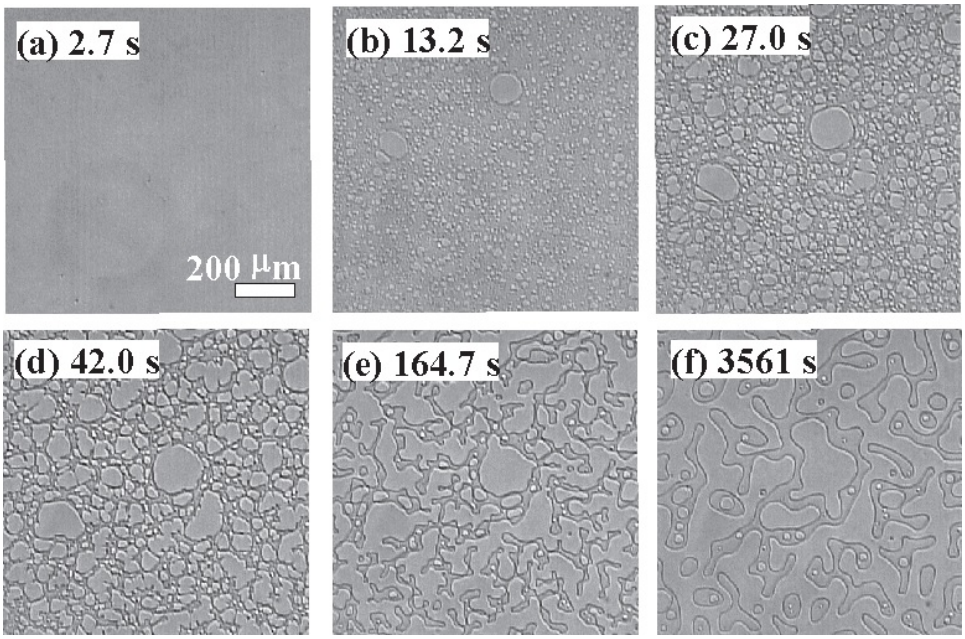


Figure 1.1: A phase separation process observed with phase-contrast microscopy. The polymer solution consists of 6.78% (percentage by mass) polystyrene and diethyl malonate. The time shown in the figure is the elapsed time after the homogeneous mixture is rapidly cooled to 9.3 °C, which is below the *upper critical temperature* of 16.5 °C for the phase separation of this solution. Republished with permission of IOP Publishing Ltd, from Journal of Physics: Condensed Matter, Institute of Physics (Great Britain), volume 12, issue 15, 2000; permission conveyed through Copyright Clearance Center, Inc.

- the volume shrinking of the polymer-rich phase,
- the resulting formation of a networklike structure, and finally,
- the phase inversion caused by relaxation of the pattern dominated by the elastic energy to that dominated by interfacial tension.

These essential features are visible in observations from lab experiments from *Tanaka* [68], see Figures 1.1 and 1.2.

According to the second law of thermodynamics, the change in free energy of a process with constant temperature and volume must be non-positive. And in fact, the above-mentioned models are all isothermal, since they use double-well potentials based on a fixed temperature. Therefore, the total energy of these models has to satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t} E_{total} \leq 0$$

at all times $t \in [0, \infty)$ to fulfill the second law of thermodynamics and thus, to be thermodynamically consistent. Note that the case of energy conservation, i.e., $\frac{\mathrm{d}}{\mathrm{d}t} E_{total} = 0$, typically arises at the end of a phase separation process. Assuming a fixed temperature in the dynamic equations reflects the experimental situation, where usually a polymer solution being in one

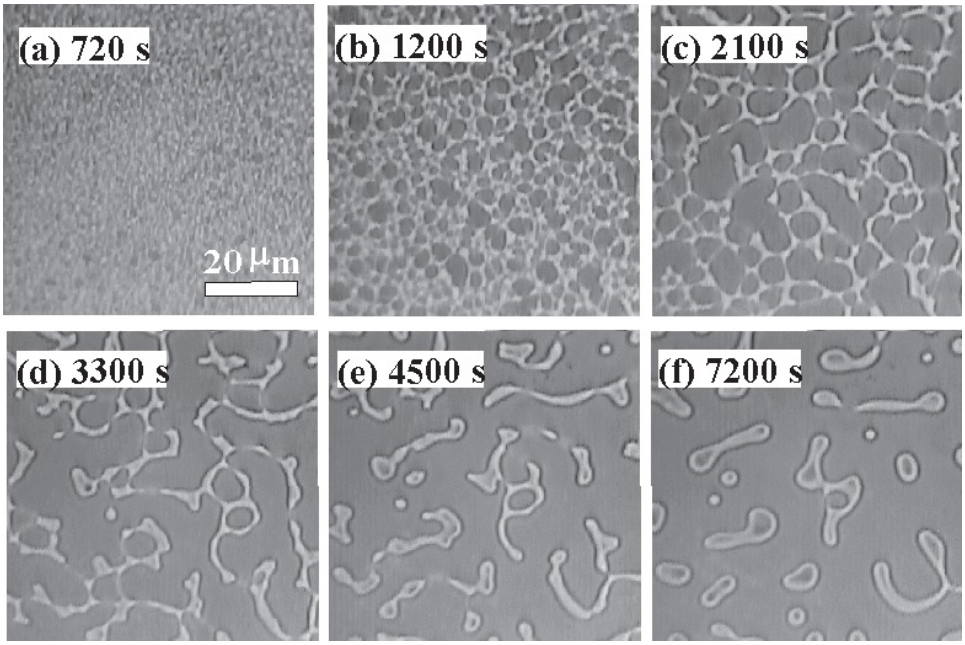


Figure 1.2: Another phase separation process observed with phase-contrast microscopy. The mixture consists of 20% (percentage by mass) polystyrene and polyvinyl methyl ether. The time shown in the figure is the elapsed time after the temperature jump. Republished with permission of IOP Publishing Ltd, from Journal of Physics: Condensed Matter, Institute of Physics (Great Britain), volume 12, issue 15, 2000; permission conveyed through Copyright Clearance Center, Inc.

thermodynamic phase is cooled (or heated) to a fixed temperature below the *upper critical temperature* (or above the *lower critical temperature*) in order to separate into two phases.

In order to describe the dynamics of a complex polymer-solvent mixture, *Zhou et al.* [74] couple the Cahn-Hilliard equation for the phase-field evolution with an extended Oldroyd-B model, which consist of the momentum equations for the velocity field, the continuity equation, and the rheological equations for time evolution of the elastic shear stress tensor and the elastic bulk stress. This model is thermodynamically consistent because it is derived through the variational principle as a minimizer of a total free energy. The authors also present simulation results, which confirm the capabilities of the model to reproduce all the essential features of viscoelastic phase separation mentioned above. These simulation results are based on a space discretization by finite volumes and a purely explicit solver in time. However, this explicit solver introduces positive terms to the time derivative of the free energy on a discrete level. Thus, the thermodynamic consistency of the model equations is not necessarily conserved.

A main part of this thesis is devoted to the development of more suitable runtime-efficient numerical schemes to solve the viscoelastic phase-field model by *Zhou et al.* [74]. Here, more suitable means that the schemes carry forward the significant thermodynamic consistency of the model equations. This task has already been tackled by us in a preliminary fashion in *Lukáčová-Medviďová, Dünweg, Strasser and Tretyakov* [56] and in more detail in *Strasser, Tierra, Dünweg and Lukáčová-Medviďová* [67]. Therefore, parts of this thesis, which will be

indicated, are based on these publications.

In the literature, one can already find several numerical methods that have been used for the numerical approximation of the diffuse interface phase-field model by Cahn and Hilliard, see, e. g., [16, 37, 48, 53, 71] and the references therein. Typically for diffuse interface models, the interfacial region separating the two fluids is very narrow. Thus, a high spatial resolution is required to accurately capture the interface dynamics. Additionally, in order to conserve the thermodynamic consistency and to reduce time step restrictions, an at least partially implicit time discretization is necessary.

There are also numerous analytical as well as numerical results available in the literature for the Oldroyd-B model, see, e. g., [5, 28, 31, 57]. The main challenges in this field are to obtain a stable approximate numerical solution for large Weissenberg numbers and for rapidly changing velocities. The dimensionless Weissenberg number represents elastic effects; it is large when the molecular relaxation time is comparable to the time scale of the flow, or even exceeds it significantly. In the present work, we consider the non-critical regime of Weissenberg numbers and only small velocity changes naturally caused by the demixing process.

Even though the system by *Zhou et al.* [74] reproduces all the essential features of viscoelastic phase separation and is thermodynamically consistent, a fully satisfactory macroscopic solution of the problem is probably still missing. For this reason, we carefully investigate how well our numerical solutions to this system describes the physics, i.a. by comparing it with and linking it to results of computer experiments that are based upon a mesoscopic (coarse-grained molecular dynamics) model, that can be considered as physically sound beyond a reasonable doubt. Note that the latter naturally requires vastly more computational effort.

This thesis is organized as follows. We begin by introducing the notation and essential identities in Chapter 2. In Chapter 3, we present several mathematical models for two component phase-fields and discuss their thermodynamic consistency. The most challenging one being the model for polymer-solvent mixtures by *Zhou et al.* [74], consisting of the Cahn-Hilliard equation for the phase-field dynamics and the extended Oldroyd-B equations for the viscoelastic hydrodynamics. We also consider a simplification, modeling only the phase-field dynamics of a polymer-solvent mixture without any hydrodynamic effects. Chapter 4 is devoted to problem-specific numerical methods for the above-mentioned models. We introduce schemes with truncation errors of first and second order in space and time that are linear and thermodynamically consistent. Starting with suitable semi-discretizations in time for the Cahn-Hilliard equation, with special attention being paid to the double-well potentials beforehand, before successively moving forward up to the most challenging model. For this aim we use implicit, explicit and implicit-explicit (IMEX) Euler methods, the Crank-Nicolson method, the two-step Adams-Bashforth method and some Taylor approximations. Most considered models are highly coupled and demand an at least partially implicit discretization. Therefore, the coupled nature of the model gets easily carried forward to its discretization. Since we focus on linear and thus efficient schemes, we also introduce splitting methods in order to decouple the calculations as far as possible and thereby save even more computational effort. Thereafter, we derive finite difference and finite volume space discretizations for two and three dimensions and present some full discretizations. Numerical experiments discussed in Chapter 5 confirm

the schemes robustness and reliability to simulate (viscoelastic) phase separation and include parameter and experimental order of convergence studies, as well as comparisons with results from the aforementioned coarse-grained molecular dynamics simulations. Finally, Chapter 6 provides our conclusion and a brief outlook.

2 Preliminaries

This chapter provides some useful preliminaries needed for the upcoming introduction and analysis of (viscoelastic) two-fluid models and their successive numerical discretization. We start with the notation and recall the integration by parts rules and some estimates and equalities for symmetric positive matrices used throughout the thesis afterwards.

Notation

Let $d \in \{2, 3\}$ denote the space dimension and let $\boldsymbol{x} = (x_1, ..., x_d)^T \in \mathbb{R}^d$ be a column vector. Let φ and q be scalar-valued functions, $\boldsymbol{u} = (u_1, ..., u_d)^T$ and $\boldsymbol{v} = (v_1, ..., v_d)^T$ be vector-valued functions and $\boldsymbol{C} = \{C_{ij}\}$, $\boldsymbol{D} = \{D_{ij}\}$, $i, j = 1, ..., d$, be tensor-valued (or matrix-valued) functions. Then, we use the following notation in which $i, j, k = 1, ..., d$

$$
\begin{aligned}
(\nabla\varphi)_i &= \frac{\partial\varphi}{\partial x_i}\,, & (\nabla\boldsymbol{u})_{ij} &= \frac{\partial u_i}{\partial x_j}\,, & (\nabla\boldsymbol{C})_{ijk} &= \frac{\partial C_{ij}}{\partial x_k}\,,\\
& & \nabla\cdot\boldsymbol{u} &= \sum_{i=1}^{d}\frac{\partial u_i}{\partial x_i}\,, & (\nabla\cdot\boldsymbol{C})_i &= \sum_{j=1}^{d}\frac{\partial C_{ij}}{\partial x_j}\,,\\
\Delta\varphi &= \sum_{i=1}^{d}\frac{\partial^2\varphi}{\partial x_i^2}\,, & (\Delta\boldsymbol{u})_i &= \sum_{j=1}^{d}\frac{\partial^2 u_i}{\partial x_j^2}\,, & (\Delta\boldsymbol{C})_{ij} &= \sum_{k=1}^{d}\frac{\partial^2 C_{ij}}{\partial x_k^2}\,,\\
& & \boldsymbol{u}\cdot\boldsymbol{v} &= \sum_{i=1}^{d} u_i v_i\,, & \boldsymbol{C}\,:\,\boldsymbol{D} &= \sum_{i,j=1}^{d} C_{ij}D_{ij}\,,\\
(\boldsymbol{u}\cdot\nabla)\varphi &= \sum_{i=1}^{d} u_i\frac{\partial\varphi}{\partial x_i}\,, & ((\boldsymbol{u}\cdot\nabla)\boldsymbol{v})_i &= \sum_{j=1}^{d} u_j\frac{\partial v_i}{\partial x_j}\,, & ((\boldsymbol{u}\cdot\nabla)\boldsymbol{C})_{ij} &= \sum_{k=1}^{d} u_k\frac{\partial C_{ij}}{\partial x_k}\,,\\
& & (\boldsymbol{u}\otimes\boldsymbol{v})_{ij} &= u_i v_j\,, & (\boldsymbol{u}\otimes\boldsymbol{C})_{ijk} &= u_k C_{ij}\,.
\end{aligned}
$$

The above introduced *inner product* (or *dot product*)

$$
\boldsymbol{u}\cdot\boldsymbol{v} = \sum_{i=1}^{d} u_i v_i
$$

induces the *Euclidean norm*

$$|\boldsymbol{u}| = (\boldsymbol{u} \cdot \boldsymbol{u})^{1/2} = \left(\sum_{i=1}^{d} u_i^2 \right)^{1/2} .$$

Further, $\boldsymbol{C}^T$, $\mathrm{tr}\boldsymbol{C}$ and $\det \boldsymbol{C}$ denote the *transpose*, the *trace* and the *determinant* of tensor $\boldsymbol{C}$, respectively. The *Frobenius inner product* (or *double dot product*) introduced above holds

$$\boldsymbol{C} : \boldsymbol{D} = \sum_{i,j=1}^{d} C_{ij} D_{ij} = \mathrm{tr}(\boldsymbol{C}\boldsymbol{D}^T) = \mathrm{tr}(\boldsymbol{C}^T \boldsymbol{D})$$

and induces the *Frobenius norm*

$$|\boldsymbol{C}| = (\boldsymbol{C} : \boldsymbol{C})^{1/2} = \left(\sum_{i,j=1}^{d} C_{ij}^2 \right)^{1/2} = \left(\mathrm{tr}(\boldsymbol{C}\boldsymbol{C}^T) \right)^{1/2} .$$

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with the (piecewise) *Lipschitz continuous* boundary $\partial\Omega$. We denote by $L^2(\Omega)$ the *Lebesgue space* of all measurable functions whose second powers are Lebesgue-integrable in Ω. Then, we can introduce the associated L^2-norm

$$\|\boldsymbol{u}\|_{L^2(\Omega)} = \left(\int_\Omega |\boldsymbol{u}|^2 \, \mathrm{d}\boldsymbol{x} \right)^{1/2} .$$

Green's first identity

Let us assume φ and $\boldsymbol{u}$ are continuously differentiable, $\boldsymbol{e}_i$ is the i-th *standard basis vector* for $i = 1, ..., d$ and $\boldsymbol{n} \in \mathbb{R}^d$ is the *outer normal vector*. Then, integration by parts reads

$$\int_\Omega u_i \frac{\partial \varphi}{\partial x_i} \, \mathrm{d}\boldsymbol{x} = \int_{\partial\Omega} u_i \varphi \boldsymbol{e}_i \cdot \boldsymbol{n} \, \mathrm{d}S - \int_\Omega \frac{\partial u_i}{\partial x_i} \varphi \, \mathrm{d}\boldsymbol{x} . \tag{2.1}$$

Summing over i yields the higher dimensional integration by parts formula

$$\int_\Omega \boldsymbol{u} \cdot \nabla\varphi \, \mathrm{d}\boldsymbol{x} = \int_{\partial\Omega} \varphi \boldsymbol{u} \cdot \boldsymbol{n} \, \mathrm{d}S - \int_\Omega \varphi \nabla \cdot \boldsymbol{u} \, \mathrm{d}\boldsymbol{x} . \tag{2.2}$$

Assuming q is twice continuously differentiable, the case $\boldsymbol{u} = \nabla q$ is known as *Green's first identity*

$$\int_\Omega \nabla q \cdot \nabla\varphi \, \mathrm{d}\boldsymbol{x} = \int_{\partial\Omega} \varphi \nabla q \cdot \boldsymbol{n} \, \mathrm{d}S - \int_\Omega \varphi \Delta q \, \mathrm{d}\boldsymbol{x} . \tag{2.3}$$

Symmetric positive definite matrices

The following lemma and its proof are based on an analogous version by *Mizerová* [61].

Lemma 2.1. *(Jacobi's formula)*
Any symmetric positive definite matrix function $\boldsymbol{C}(t) \in \mathbb{R}^{d\times d}$, which is continuously differentiable for all $t \in K \subseteq \mathbb{R}$, holds

$$\frac{\mathrm{d}\boldsymbol{C}}{\mathrm{d}t} : \boldsymbol{C}^{-1} = \operatorname{tr}\left(\boldsymbol{C}^{-1}\frac{\mathrm{d}\boldsymbol{C}}{\mathrm{d}t}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\ln\boldsymbol{C} = \frac{\mathrm{d}}{\mathrm{d}t}\ln\det\boldsymbol{C}\,, \tag{2.4}$$

where the matrix logarithm is defined as follows.

Let $\boldsymbol{C} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$ be an eigendecomposition into the orthogonal matrix $\boldsymbol{Q}$, consisting column by column of the eigenvectors of $\boldsymbol{C}$, and the diagonal matrix $\boldsymbol{\Lambda}$, having the eigenvalues $\lambda_i, i = 1, ..., d$, of $\boldsymbol{C}$ on its diagonal. Then,

$$\ln\boldsymbol{C} = \boldsymbol{Q}\ln\boldsymbol{\Lambda}\boldsymbol{Q}^T\,,$$

where $\ln\boldsymbol{\Lambda}$ is still diagonal with the logarithm of each entry of $\boldsymbol{\Lambda}$ on its diagonal.

Proof. The first equality of (2.4) holds due to the symmetry of $\boldsymbol{C}$.
Let us decompose $\boldsymbol{C}$ into

$$\boldsymbol{C} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T\,,$$

where the orthogonal matrix $\boldsymbol{Q}$ and the diagonal matrix $\boldsymbol{\Lambda}$ are continuously differentiable. Then, we have the following derivatives

$$\begin{aligned}
\frac{\mathrm{d}\boldsymbol{C}}{\mathrm{d}t} &= \frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t}\boldsymbol{\Lambda}\boldsymbol{Q}^T + \boldsymbol{Q}\frac{\mathrm{d}\boldsymbol{\Lambda}}{\mathrm{d}t}\boldsymbol{Q}^T + \boldsymbol{Q}\boldsymbol{\Lambda}\frac{\mathrm{d}\boldsymbol{Q}^T}{\mathrm{d}t}\,,\\
\frac{\mathrm{d}\ln\boldsymbol{C}}{\mathrm{d}t} &= \frac{\mathrm{d}\left(\boldsymbol{Q}\ln\boldsymbol{\Lambda}\boldsymbol{Q}^T\right)}{\mathrm{d}t} = \frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t}\ln\boldsymbol{\Lambda}\boldsymbol{Q}^T + \boldsymbol{Q}\frac{\mathrm{d}\ln\boldsymbol{\Lambda}}{\mathrm{d}t}\boldsymbol{Q}^T + \boldsymbol{Q}\ln\boldsymbol{\Lambda}\frac{\mathrm{d}\boldsymbol{Q}^T}{\mathrm{d}t}\,.
\end{aligned}$$

Since $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$ and since the trace is linear and invariant under cyclic permutations, it holds

$$\begin{aligned}
\operatorname{tr}\frac{\mathrm{d}\ln\boldsymbol{C}}{\mathrm{d}t} &= \operatorname{tr}\left(\ln\boldsymbol{\Lambda}\boldsymbol{Q}^T\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t}\right) + \operatorname{tr}\left(\boldsymbol{Q}^T\boldsymbol{Q}\frac{\mathrm{d}\ln\boldsymbol{\Lambda}}{\mathrm{d}t}\right) + \operatorname{tr}\left(\ln\boldsymbol{\Lambda}\frac{\mathrm{d}\boldsymbol{Q}^T}{\mathrm{d}t}\boldsymbol{Q}\right)\\
&= \operatorname{tr}\left(\ln\boldsymbol{\Lambda}\left[\boldsymbol{Q}^T\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t} + \frac{\mathrm{d}\boldsymbol{Q}^T}{\mathrm{d}t}\boldsymbol{Q}\right]\right) + \operatorname{tr}\frac{\mathrm{d}\ln\boldsymbol{\Lambda}}{\mathrm{d}t} = \operatorname{tr}\left(\ln\boldsymbol{\Lambda}\frac{\mathrm{d}\left(\boldsymbol{Q}^T\boldsymbol{Q}\right)}{\mathrm{d}t}\right) + \operatorname{tr}\frac{\mathrm{d}\ln\boldsymbol{\Lambda}}{\mathrm{d}t}\\
&= \operatorname{tr}\frac{\mathrm{d}\ln\boldsymbol{\Lambda}}{\mathrm{d}t} = \operatorname{tr}\frac{\mathrm{d}\boldsymbol{\Lambda}}{\mathrm{d}t}\boldsymbol{\Lambda}^{-1}\,.
\end{aligned}$$

Thus,

$$\begin{aligned}
\operatorname{tr}\left(\boldsymbol{C}^{-1}\frac{\mathrm{d}\boldsymbol{C}}{\mathrm{d}t}\right) = \frac{\mathrm{d}\boldsymbol{C}}{\mathrm{d}t} : \boldsymbol{C}^{-1} &= \left(\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t}\boldsymbol{\Lambda}\boldsymbol{Q}^T + \boldsymbol{Q}\frac{\mathrm{d}\boldsymbol{\Lambda}}{\mathrm{d}t}\boldsymbol{Q}^T + \boldsymbol{Q}\boldsymbol{\Lambda}\frac{\mathrm{d}\boldsymbol{Q}^T}{\mathrm{d}t}\right) : \boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T\\
&= \operatorname{tr}\left(\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t}\boldsymbol{Q}^T + \frac{\mathrm{d}\boldsymbol{\Lambda}}{\mathrm{d}t}\boldsymbol{\Lambda}^{-1} + \frac{\mathrm{d}\boldsymbol{Q}^T}{\mathrm{d}t}\boldsymbol{Q}\right)\\
&= \operatorname{tr}\left(\frac{\mathrm{d}\left(\boldsymbol{Q}^T\boldsymbol{Q}\right)}{\mathrm{d}t} + \frac{\mathrm{d}\ln\boldsymbol{\Lambda}}{\mathrm{d}t}\right)\\
&= \operatorname{tr}\frac{\mathrm{d}\ln\boldsymbol{C}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\ln\boldsymbol{C}\,.
\end{aligned}$$

The last equality of (2.4) holds, since

$$\operatorname{tr} \ln \boldsymbol{C} = \operatorname{tr}\left(\boldsymbol{Q} \ln \boldsymbol{\Lambda} \boldsymbol{Q}^T\right) = \operatorname{tr}\left(\boldsymbol{Q}^T \boldsymbol{Q} \ln \boldsymbol{\Lambda}\right) = \operatorname{tr} \ln \boldsymbol{\Lambda} = \sum_{i=1}^{d} \ln \lambda_i = \ln\left(\prod_{i=1}^{d} \lambda_i\right) = \ln \det \boldsymbol{C}\,.$$

Note that the last equation also proofs (2.5a). □

The following lemma is from *Boyaval, Lelièvre and Mangoubi* [7]. Let us refer to their publication for the proof.

Lemma 2.2. *Let $\boldsymbol{C} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{D} \in \mathbb{R}^{d\times d}$ be two symmetric positive definite matrices. We have:*

$$\operatorname{tr} \ln \boldsymbol{C} = \ln \det \boldsymbol{C}, \tag{2.5a}$$

$$\boldsymbol{C} - \ln \boldsymbol{C} - \boldsymbol{I} \text{ is symmetric positive semidefinite and thus } \operatorname{tr}(\boldsymbol{C} - \ln \boldsymbol{C} - \boldsymbol{I}) \geq 0, \tag{2.5b}$$

$$\boldsymbol{C} + \boldsymbol{C}^{-1} - 2\boldsymbol{I} \text{ is symmetric positive semidefinite and thus } \operatorname{tr}(\boldsymbol{C} + \boldsymbol{C}^{-1} - 2\boldsymbol{I}) \geq 0, \tag{2.5c}$$

$$\operatorname{tr}(\boldsymbol{C}\boldsymbol{D}) = \operatorname{tr}(\boldsymbol{D}\boldsymbol{C}) \geq 0, \tag{2.5d}$$

$$\operatorname{tr}\left((\boldsymbol{C} - \boldsymbol{D})\boldsymbol{D}^{-1}\right) = \operatorname{tr}\left(\boldsymbol{C}\boldsymbol{D}^{-1} - \boldsymbol{I}\right) \geq \ln \det\left(\boldsymbol{C}\boldsymbol{D}^{-1}\right) = \operatorname{tr}(\ln \boldsymbol{C} - \ln \boldsymbol{D}), \tag{2.5e}$$

$$\operatorname{tr}\big((\ln \boldsymbol{C} - \ln \boldsymbol{D})\boldsymbol{C}\big) \geq \operatorname{tr}(\boldsymbol{C} - \boldsymbol{D}). \tag{2.5f}$$

3 Mathematical models

In this chapter, we give a detailed overview of Cahn-Hilliard based phase-field models and analyze if the time evolutions of their (free) energies are in accordance with the second law of thermodynamics. We start from the pure Cahn-Hilliard equation for phase-fields with diffusive interface. Then, we add hydrodynamics in form of the well-known Navier-Stokes equations. Afterwards, we introduce and refine the effects of viscoelasticity using the Oldroyd-B model in order to finally be able to macroscopically model the phase separation of polymer-solvent mixtures in line with physics. Closing the chapter with some noteworthy model modifications and extensions, as well as a brief overview of existence and uniqueness results.

3.1 Cahn-Hilliard equation

Let $\Omega \subset \mathbb{R}^d$ be a computational domain with (piecewise) Lipschitz continuous boundary $\partial\Omega$. The phase-field variable, which is a scalar function of space $\boldsymbol{x} \in \Omega$ and time $t \in [0, \infty)$, is denoted by $\varphi \cong \varphi(\boldsymbol{x}, t)$. There are two major ways to define φ. One variant is that $\varphi \in [0, 1]$ indicates the volume fraction of one component such that $1 - \varphi$ is the volume fraction of the other. Thus, $\bar{\varphi}_1 = 0$ and $\bar{\varphi}_2 = 1$ are the pure phases. The other variant is that $\varphi = \frac{c_1 - c_2}{c_1 + c_2}$ indicates the difference of the two concentrations $c_1 \geq 0$ and $c_2 \geq 0$. Here, $\bar{\varphi}_1 = -1$ and $\bar{\varphi}_2 = 1$ are the pure phases. According to *Cahn and Hilliard* [15], the free energy functional of a diffusive phase-field is given by

$$E_{mix}(\varphi) = \int_\Omega \left(\frac{\lambda}{2} |\nabla \varphi|^2 + F(\varphi) \right) \mathrm{d}\boldsymbol{x} \,, \tag{3.1}$$

where λ is a positive constant controlling the thickness of the diffusive interface between the phases, and $F(\varphi)$ is a double-well potential that represents the tendency of a system to have two different stable phases, see Subsection 3.1.1 for details. We refer to (3.1) as mixing energy. Note that the mixing energy is not necessarily non-negative, but always bounded from below.

The following standard derivation of the Cahn-Hilliard equation is based on a microforce balance, see, e.g., *Gurtin* [41] for more details. We start from the law of conservation of mass

$$\frac{\partial \varphi}{\partial t} + \nabla \cdot \boldsymbol{J} = 0 \,, \tag{3.2}$$

where the mass flux $\boldsymbol{J}$ is related to the chemical potential μ by the constitutive equation

$$\boldsymbol{J} = -M(\varphi) \nabla \mu \,. \tag{3.3}$$

The coefficient function $M(\varphi)$ is the so-called mobility, which is inversely related to the friction $\zeta(\varphi) > 0$. It reads

$$M(\varphi) = \frac{1}{\zeta(\varphi)} \left((-\bar{\varphi}_1 + \varphi)(\bar{\varphi}_2 - \varphi)\right)^n \geq 0\,, \tag{3.4}$$

where $n \geq 0$. If $n > 0$, the mobility is called degenerate (or degenerative) since in this case it decreases towards the pure phases. The chemical potential μ is given by the variational derivative of the mixing energy (3.1) with respect to φ

$$\mu = \frac{\delta E_{mix}(\varphi)}{\delta \varphi}\,. \tag{3.5}$$

Theorem 3.1. *Assuming suitable boundary conditions, i.e., periodic or* $\frac{\partial \varphi}{\partial \boldsymbol{n}}|_{\partial\Omega} = 0$*, the variational derivative of the mixing energy satisfies*

$$\frac{\delta E_{mix}(\varphi)}{\delta \varphi} = -\lambda \Delta \varphi + f(\varphi)\,, \tag{3.6}$$

where

$$f(\varphi) = F'(\varphi)\,.$$

Proof. Assuming that the mixing energy has the form (3.1), its variation reads

$$\begin{aligned}
\delta E_{mix}(\varphi) &= \delta \left[\int_\Omega \left(\frac{\lambda}{2} |\nabla \varphi|^2 + F(\varphi) \right) \mathrm{d}\boldsymbol{x} \right] \\
&= \int_\Omega \left[\frac{\partial}{\partial \nabla \varphi} \left(\frac{\lambda}{2} |\nabla \varphi|^2 \right) \delta \nabla \varphi + \frac{\partial}{\partial \varphi} \left(F(\varphi) \right) \delta \varphi \right] \mathrm{d}\boldsymbol{x} \\
&= \int_\Omega \left[\lambda \nabla \varphi \cdot \delta \nabla \varphi + F'(\varphi)\, \delta \varphi \right] \mathrm{d}\boldsymbol{x}\,.
\end{aligned}$$

Applying integration by parts (*Green's first identity* (2.3)) on the first term and defining $f(\varphi) = F'(\varphi)$ yields

$$\delta E_{mix}(\varphi) = -\int_\Omega \lambda \Delta \varphi\, \delta \varphi \, \mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} \lambda (\nabla \varphi \cdot \boldsymbol{n})\, \delta \varphi \, \mathrm{d}S + \int_\Omega f(\varphi)\, \delta \varphi \, \mathrm{d}\boldsymbol{x}\,,$$

where $\boldsymbol{n} \in \mathbb{R}^d$ is the outward-pointing normal vector and $\nabla \varphi \cdot \boldsymbol{n} = \frac{\partial \varphi}{\partial \boldsymbol{n}}$ the normal derivative of φ. Assuming suitable boundary conditions, i.e., periodic or $\frac{\partial \varphi}{\partial \boldsymbol{n}} = 0$ on $\partial\Omega$, the boundary integral vanishes and we obtain

$$\begin{aligned}
\delta E_{mix}(\varphi) &= \int_\Omega \left[-\lambda \Delta \varphi\, \delta \varphi + f(\varphi)\, \delta \varphi \right] \mathrm{d}\boldsymbol{x} \\
&= \int_\Omega \left[-\lambda \Delta \varphi + f(\varphi) \right] \delta \varphi \, \mathrm{d}\boldsymbol{x}\,.
\end{aligned}$$

Further, it follows from [33, p. 18] that

$$\delta E_{mix}(\varphi) = \int_\Omega \frac{\delta E_{mix}(\varphi)}{\delta \varphi}\, \delta \varphi \, \mathrm{d}\boldsymbol{x}\,.$$

In conjunction with the previous computation this yields the desired result

$$\frac{\delta E_{mix}(\varphi)}{\delta \varphi} = -\lambda \Delta \varphi + f(\varphi)\,.$$

□

Summarized, we have the Cahn-Hilliard equation

$$\frac{\partial \varphi}{\partial t} = \nabla \cdot \left(M(\varphi) \nabla \mu \right), \tag{3.7a}$$

$$\mu = -\lambda \Delta \varphi + f(\varphi)\,. \tag{3.7b}$$

Remark 3.2. *In order to have an initial-boundary-value problem, we complement the Cahn-Hilliard equation (3.7) with suitable initial and boundary conditions. Our initial conditions read*

$$\varphi|_{t=0} = \varphi^0 \quad \text{in } \Omega\,.$$

Suitable boundary conditions include

- *periodic boundary conditions, and*
- *the homogeneous Neumann boundary conditions*

$$\frac{\partial \varphi}{\partial \boldsymbol{n}}\Big|_{\partial\Omega} = \frac{\partial \mu}{\partial \boldsymbol{n}}\Big|_{\partial\Omega} = 0 \quad \text{on } \partial\Omega \times (0, \infty)\,.$$

Note that these two options are the most common choices in the literature studying the Cahn-Hilliard equation and that we will rely on periodic boundary conditions for our numerical experiments in Chapter 5. Further, note that both above-mentioned choices are suitable for the derivation of energy laws throughout this thesis.

Remark 3.3. *Assuming it exists a well-known free energy of a system of differential equations, the time evolution of this free energy is usually computed through the multiplication of each differential equation with a smooth test function ψ and the integration over the whole computational domain Ω. Then, assuming suitable boundary conditions, utilizing integration by parts (Green's identities), choosing appropriate substitutes for the test functions for each equation and summing up yields us an energy law.*

Theorem 3.4. *Assuming suitable boundary conditions, see Remark 3.2, the Cahn-Hilliard equation (3.7) satisfies the following energy law*

$$\frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} = -\int_\Omega M(\varphi)|\nabla \mu|^2 \mathrm{d}\boldsymbol{x} \leq 0\,, \tag{3.8}$$

since $M(\varphi) \geq 0$. Recalling that the Cahn-Hilliard equation is based on a constant temperature, this energy law implies that it fulfills the second law of thermodynamics and therefore that it is thermodynamically consistent.

Proof. Multiplying (3.7a) by the smooth test function ψ, integrating over the computational domain Ω and applying integration by parts, we obtain

$$\begin{aligned} 0 &= \int_\Omega \left(\frac{\partial \varphi}{\partial t} - \nabla \cdot (M(\varphi)\nabla\mu) \right) \psi \, \mathrm{d}\boldsymbol{x} \\ &= \int_\Omega \frac{\partial \varphi}{\partial t} \psi \, \mathrm{d}\boldsymbol{x} - \int_\Omega \nabla \cdot (M(\varphi)\nabla\mu)\, \psi \, \mathrm{d}\boldsymbol{x} \\ &= \int_\Omega \frac{\partial \varphi}{\partial t} \psi \, \mathrm{d}\boldsymbol{x} + \int_\Omega M(\varphi)\nabla\mu \cdot \nabla\psi \, \mathrm{d}\boldsymbol{x} - \int_{\partial\Omega} M(\varphi)\nabla\mu\psi \cdot \boldsymbol{n} \mathrm{d}S \,, \end{aligned}$$

where $\boldsymbol{n}$ is the outer-pointing normal. Taking, in particular, $\psi = \mu$ yields

$$\begin{aligned} 0 &= \int_\Omega \frac{\partial \varphi}{\partial t} \mu \mathrm{d}\boldsymbol{x} + \int_\Omega M(\varphi)\nabla\mu \cdot \nabla\mu \mathrm{d}\boldsymbol{x} - \int_{\partial\Omega} M(\varphi)\nabla\mu\mu \cdot \boldsymbol{n} \mathrm{d}S \\ &= \int_\Omega \frac{\partial \varphi}{\partial t} \frac{\delta E_{mix}(\varphi)}{\delta \varphi} \mathrm{d}\boldsymbol{x} + \int_\Omega M(\varphi)|\nabla\mu|^2 \mathrm{d}\boldsymbol{x} - \int_{\partial\Omega} M(\varphi)\mu \frac{\partial \mu}{\partial \boldsymbol{n}} \mathrm{d}S \\ &= \frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \int_\Omega M(\varphi)|\nabla\mu|^2 \mathrm{d}\boldsymbol{x} - \int_{\partial\Omega} M(\varphi)\mu \frac{\partial \mu}{\partial \boldsymbol{n}} \mathrm{d}S \,. \end{aligned}$$

Assuming suitable boundary conditions, see Remark 3.2, the boundary integral vanishes and we have the desired energy law (3.8).

□

For the sake of simplicity, we will leave out the symbols $\mathrm{d}\boldsymbol{x}$ of the space integrals and $\mathrm{d}S$ of the boundary integrals in the following.

3.1.1 Double-well potentials

A double-well potential represents the tendency of a system to have two different stable phases. A simple potential that satisfies this condition is the Ginzburg-Landau potential

$$F_{GL}(\varphi) = \frac{1}{4}(\varphi^2 - 1)^2 \,, \tag{3.9}$$

which is defined on the whole real axis, and whose two minima occur at the pure phase values $\varphi = \pm 1$, see Figure 3.1. This potential is quite often used in the mathematical literature studying the Cahn-Hilliard equation, see, e. g., *Elliot and Zheng* [25] or *Elliot and Garcke* [23, 24]. Note that we can extend the Ginzburg-Landau potential to a general formulation for polynomial double-well potentials,

$$F_{polyn}(\varphi) = \beta(\varphi - \alpha_1)^2(\varphi - \alpha_2)^2 \,, \tag{3.10}$$

where the minima occur at the pure phase values α_1 and α_2, which can be chosen arbitrarily in $\mathbb{R}$, while $\beta > 0$ is used to control the slope.

From a physical point of view, the Flory-Huggins potential [32, 46] describes polymer-solvent phase separation more accurately, as it is derived from the mean field theory for polymer systems. It is usually given by

$$F_{FH}(\varphi) = \frac{1}{n_p}\varphi \ln \varphi + \frac{1}{n_s}(1-\varphi)\ln(1-\varphi) + \chi\varphi(1-\varphi) \,, \tag{3.11}$$

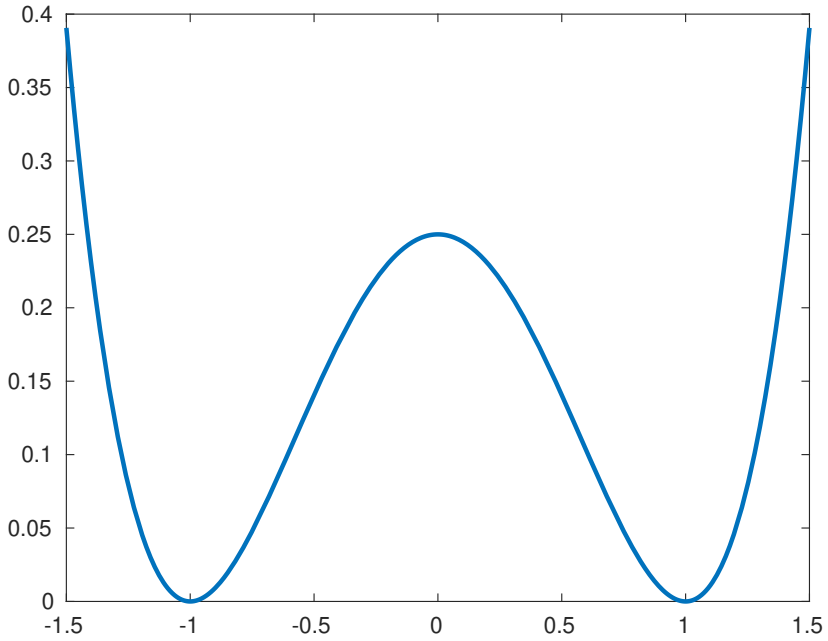


Figure 3.1: Ginzburg-Landau potential (3.9) for $\varphi \in [-1.5, 1.5]$.

which is defined on the interval $[0, 1]$ and has up to two minima within. In equation (3.11), n_p and n_s denote the degrees of polymerization of the two components. In viscoelastic phase-field models usually $n_p \geq 1$, depending on the length of the polymer-chains, and $n_s = 1$ for a pure solvent, resulting in an asymmetric double-well potential, see Figure 3.2. Furthermore, $\chi > 0$ is the polymer-solvent interaction parameter, also called the Flory-Huggins interaction, which is inversely proportional to the constant temperature of the system (the model is isothermal as pointed out before). It controls the regime of the potential, i.e., whether it separates the phases or mixes them. If the temperature is above a certain value (the so-called upper critical temperature) the potential only has one well, see Figure 3.3. Note that the Flory-Huggins potential is usually measured in units of the thermal energy $k_B T$, where k_B is the Boltzmann constant and T the absolute temperature, but it is non-dimensionalized here.

Note that the Flory-Huggins potential can also be formulated for arbitrary intervals. $[-1, 1]$ is a not uncommon choice due to the analogy to the pure phase values of the Ginzburg-Landau potential (3.9).

For purposes of the derivation of linear numerical schemes that provably conserve the thermodynamic consistency of the Cahn-Hilliard model which they approximate, it is indispensable for the potential to have a uniformly bounded second derivative, i.e.,

$$\sup_{\varphi \in K} |F''(\varphi)| \leq const.,$$

where $K \subset \mathbb{R}$ is chosen suitably. Since neither of the two aforementioned potentials fulfills this demand, we suggest suitable modifications of them in the following.

For example, in [38, 72], the authors use an appropriate modification of the Ginzburg-Landau potential (3.9), where the steep increase $\sim \varphi^4$ outside $[-1, 1]$ is replaced by a weaker quadratic

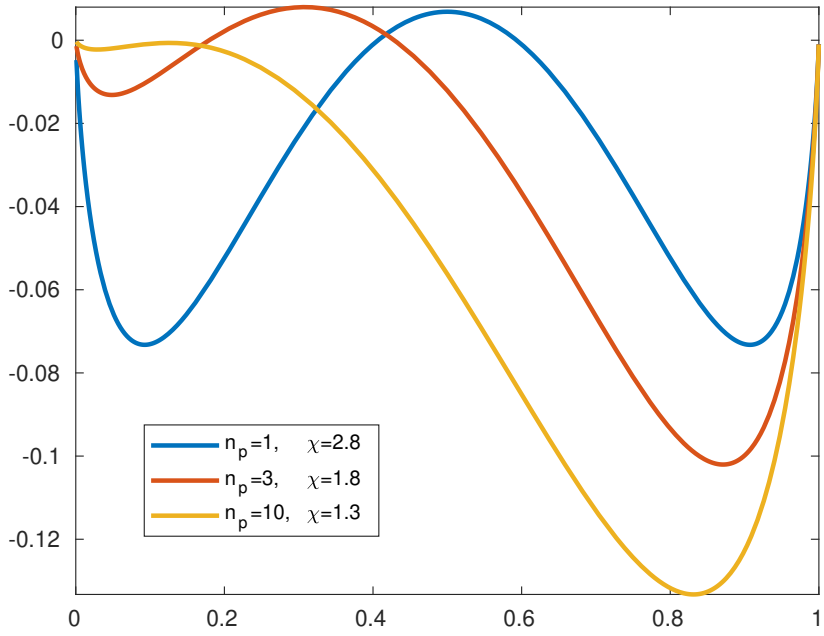


Figure 3.2: Flory-Huggins potential (3.11) for different degrees of polymerization n_p while $n_s = 1$. The Flory interaction χ has to be adapted to the polymerization such that we stay in the regime where phase separation occurs (below the upper critical temperature of the system).

increase

$$\tilde{F}_{GL}(\varphi) = \begin{cases} (\varphi+1)^2 & \varphi < -1\,, \\ \frac{1}{4}(\varphi^2-1)^2 & \varphi \in [-1,1]\,, \\ (\varphi-1)^2 & \varphi > 1\,. \end{cases} \tag{3.12}$$

This modified potential is twice continuously differentiable on the whole real axis with

$$\tilde{F}''_{GL}(\varphi) = \begin{cases} 2 & \varphi < -1\,, \\ 3\varphi^2-1 & \varphi \in [-1,1]\,, \\ 2 & \varphi > 1\,. \end{cases}$$

Here, $\tilde{F}''_{GL}(-1) = 2 = \tilde{F}''_{GL}(1)$, i.e., the second derivative of the Ginzburg-Landau potential is cut and continuously extended constantly outside $[-1,1]$. Therefore,

$$\sup_{\varphi\in\mathbb{R}} |\tilde{F}''_{GL}(\varphi)| = 2\,,$$

i.e., (3.12) has a bounded second derivative.

We can modify the general polynomial potential (3.10) analogously to (3.12) outside the

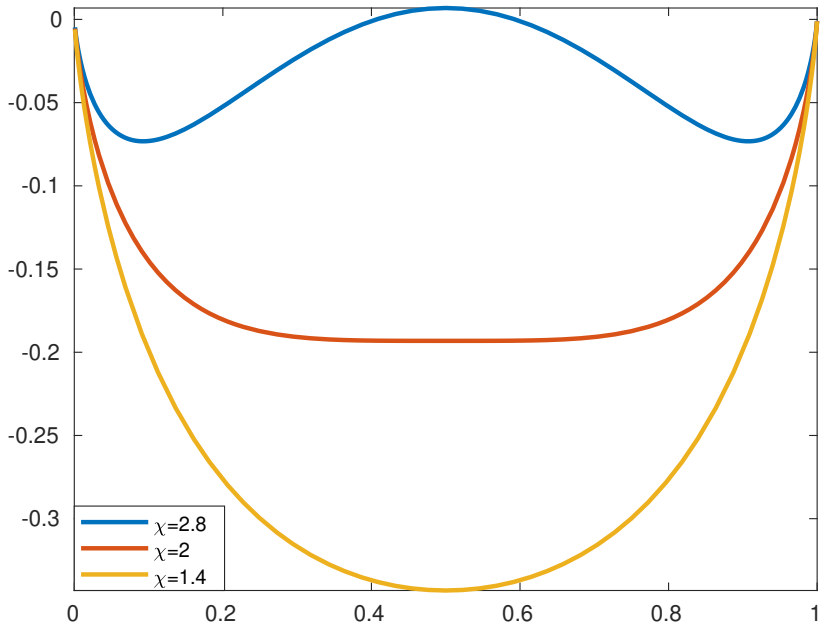


Figure 3.3: Flory-Huggins potential (3.11) for different Flory interactions χ and $n_p = 1 = n_s$. If the temperature, which is inversely proportional to χ, is above the upper critical value, we are in a regime where no separation occurs. Therefore, the wells reduce from two to one for low values of χ.

interval between the two minima, yielding

$$\tilde{F}_{polyn}(\varphi) = \begin{cases} \beta(\alpha_1 - \alpha_2)^2(\varphi - \alpha_1)^2 & \varphi < \alpha_1\,, \\ \beta(\varphi - \alpha_1)^2(\varphi - \alpha_2)^2 & \varphi \in [\alpha_1, \alpha_2]\,, \\ \beta(\alpha_1 - \alpha_2)^2(\varphi - \alpha_2)^2 & \varphi > \alpha_2\,, \end{cases} \tag{3.13}$$

where $\beta > 0$ and $-\infty < \alpha_1 < \alpha_2 < \infty$ in order that $F(\alpha_1)$ and $F(\alpha_2)$ are the two minima of $\tilde{F}_{polyn}$. This general potential is also twice continuously differentiable on the whole real axis with

$$\tilde{F}''_{polyn}(\varphi) = \begin{cases} 2\beta(\alpha_1 - \alpha_2)^2 & \varphi < \alpha_1\,, \\ 2\beta\left[(\varphi - \alpha_1)^2 + 4(\varphi - \alpha_1)(\varphi - \alpha_2) + (\varphi - \alpha_2)^2\right] & \varphi \in [\alpha_1, \alpha_2]\,, \\ 2\beta(\alpha_1 - \alpha_2)^2 & \varphi > \alpha_2\,. \end{cases}$$

Here, $\tilde{F}''_{polyn}(\alpha_1) = 2\beta(\alpha_1 - \alpha_2)^2 = \tilde{F}''_{polyn}(\alpha_2)$ is used for the constant continuous extension analogously to $\tilde{F}''_{GL}(\varphi)$. Since $\tilde{F}''_{polyn}(\varphi)$ is convex on $[\alpha_1, \alpha_2]$ with $\min(\tilde{F}''_{polyn}(\varphi)) = -\beta(\alpha_1 - \alpha_2)^2$

$$\sup_{\varphi \in \mathbb{R}} |\tilde{F}''_{polyn}(\varphi)| = 2\beta(\alpha_1 - \alpha_2)^2 < \infty\,,$$

i.e., (3.13) also has a bounded second derivative for $K = \mathbb{R}$. In order that the two minima occur within the interval $[0, 1]$, we can specify $0 \leq \alpha_1 < \alpha_2 \leq 1$.

Due to its more accurate description of polymer-solvent phase separation, let us also introduce a suitable modification of the Flory-Huggins potential (3.11). It is achieved by the same technique, yielding

$$\tilde{F}_{FH}(\varphi) = \begin{cases} \beta_1(\varphi-\alpha_1)^2 + F_{FH}(\alpha_1) & \varphi < \alpha_1\,, \\ \frac{1}{n_p}\varphi\ln\varphi + \frac{1}{n_s}(1-\varphi)\ln(1-\varphi) + \chi\varphi(1-\varphi) & \varphi \in [\alpha_1,\alpha_2]\,, \\ \beta_2(\varphi-\alpha_2)^2 + F_{FH}(\alpha_2) & \varphi > \alpha_2\,, \end{cases} \tag{3.14}$$

where $\beta_{1/2} > 0$ and $0 < \alpha_1 < \alpha_2 < 1$ such that

$$\tilde{F}'_{FH}(\alpha_1) = 0 = \tilde{F}'_{FH}(\alpha_2)\,.$$

Thus, $\tilde{F}_{FH}$ has the same two minima $F_{FH}(\alpha_1)$ and $F_{FH}(\alpha_2)$ of the original Flory-Huggins potential (3.11). Potential (3.14) is also twice continuously differentiable on the whole real axis with

$$\tilde{F}''_{FH}(\varphi) = \begin{cases} 2\beta_1 & \varphi < \alpha_1\,, \\ \frac{1}{n_p\varphi} + \frac{1}{n_s(1-\varphi)} - 2\chi & \varphi \in [\alpha_1,\alpha_2]\,, \\ 2\beta_2 & \varphi > \alpha_2\,. \end{cases}$$

In order that $\tilde{F}''_{FH}(\varphi)$ is continuous in $\mathbb{R}$, we can define

$$2\beta_1 := \tilde{F}''_{FH}(\alpha_1) \quad\Leftrightarrow\quad \beta_1 := \frac{1}{2n_p\alpha_1} + \frac{1}{2n_s(1-\alpha_1)} - \chi$$

and

$$2\beta_2 := \tilde{F}''_{FH}(\alpha_2) \quad\Leftrightarrow\quad \beta_2 := \frac{1}{2n_p\alpha_2} + \frac{1}{2n_s(1-\alpha_2)} - \chi\,.$$

Hereby, we also continuously extend the second derivative of the Flory-Huggins potential constantly outside $[\alpha_1,\alpha_2]$.

Note that we can also cut and continuously extend the second derivative of a double-well potential constantly at a certain distance from the interval between its minima, i.e.

$$\tilde{F}''(\varphi) = \begin{cases} F''(\delta_1) & \varphi < \delta_1\,, \\ F''(\varphi) & \varphi \in [\delta_1,\delta_2]\,, \\ F''(\delta_2) & \varphi > \delta_2\,, \end{cases} \tag{3.15}$$

where $-\infty < \delta_1 < \alpha_1 < \alpha_2 < \delta_2 < \infty$ for polynomial potentials and $0 < \delta_1 < \alpha_1 < \alpha_2 < \delta_2 < 1$ for the Flory-Huggins potential (3.11). This second derivative is also uniformly bounded, since it is convex in the interval $[\delta_1,\delta_2]$ and constant outside. The corresponding potential $\tilde{F}(\varphi)$ can be reconstructed by integrating (3.15) with suitable integration constants, such that $\tilde{F}'$ and $\tilde{F}$ are continuously differentiable. Note that the further away from the minima we cut, the higher the bound of the resulting second derivative of the double-well potential.

Further, note that we do not leave the interval $[\alpha_1,\alpha_2]$ by more than a small margin in our numerical experiments, see Chapter 5. Therefore, we can use the boundedness of the second derivative (3.15) for the discrete energy laws in Chapter 4 without the necessity to implement any modification of the original potentials (3.9) and (3.11).

3.2 Cahn-Hilliard-Navier-Stokes model

We start from the Cauchy momentum equation, which describes the momentum transport $\boldsymbol{u} \in \mathbb{R}^d$ of a fluid in a continuum. For more details, see, e.g., *Landau and Lifshitz*[49]. In Lagrangian (i.e., convective) form, the equation reads

$$\rho \frac{D\boldsymbol{u}}{Dt} = -\nabla p(\rho) + \nabla \cdot \mathbf{T} + \rho \boldsymbol{g}\,, \tag{3.16}$$

where $\frac{D}{Dt} = \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla$ is the material derivative, $\rho > 0$ the density, $p(\rho) \in \mathbb{R}$ the pressure, $\mathbf{T} \in \mathbb{R}^{d\times d}$ the stress tensor and $\boldsymbol{g} \in \mathbb{R}^d$ represents body accelerations acting on the continuum.

To complete the conservation form of the equations of motion we add the mass continuity equation (3.2) for density ρ with mass flux $\boldsymbol{J} = \rho\boldsymbol{u}$, reading

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0\,. \tag{3.17}$$

For viscous Newtonian fluids the stress tensor is given by

$$\mathbf{T} = 2\eta(\rho)\boldsymbol{D}(\boldsymbol{u})\,, \tag{3.18}$$

where $\eta(\rho) \geq 0$ is the viscosity, which varies according to the density ρ, and $\boldsymbol{D}(\boldsymbol{u}) \in \mathbb{R}^{d\times d}$ is the rate of deformation or rate of strain tensor, which equals the symmetric part of the velocity gradient

$$\boldsymbol{D}(\boldsymbol{u}) = \frac{1}{2}\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right]. \tag{3.19}$$

Now, combining (3.16) and (3.17) with the just mentioned stress definition yields the compressible Navier-Stokes equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0\,, \tag{3.20a}$$

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u}\right) = -\nabla p(\rho) + \nabla \cdot \left(\eta(\rho)\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right]\right) - \rho \boldsymbol{g}\,. \tag{3.20b}$$

Note that they reduce to the compressible Euler equations for an inviscid fluid where $\eta = 0$.

Assuming the Mach number to be small, i.e., the flow to be slow in relation to the speed of sound, the flow can be considered to be approximately incompressible. Thus, the density ρ is considered to be constant in $(\boldsymbol{x}, t)$, reducing the mass continuity equation (3.17) to

$$\nabla \cdot \boldsymbol{u} = 0\,, \tag{3.21}$$

i.e., the velocity field is divergence free. Furthermore, the viscosity η becomes constant, too.

Theorem 3.5. *For a constant density and viscosity, it holds*

$$\nabla \cdot 2\eta \boldsymbol{D}(\boldsymbol{u}) = \eta \nabla \cdot \left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right] = \eta \Delta \boldsymbol{u}\,.$$

Proof. Since the calculations are analogous for two and three space dimensions, we only show the two-dimensional case. Let $\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\boldsymbol{u} = \begin{pmatrix} u \\ v \end{pmatrix}$ and $u_x := \frac{\partial u}{\partial x}$ and $u_y := \frac{\partial u}{\partial y}$ are the partial derivatives in x and y direction, then

$$\nabla \cdot \nabla \boldsymbol{u} = \nabla \cdot \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} = \begin{pmatrix} u_{xx} + u_{yy} \\ v_{xx} + v_{yy} \end{pmatrix} = \Delta \boldsymbol{u}$$

and

$$\begin{aligned} \nabla \cdot (\nabla \boldsymbol{u})^T &= \nabla \cdot \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}^T \\ &= \nabla \cdot \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} = \begin{pmatrix} u_{xx} + v_{xy} \\ u_{xy} + v_{yy} \end{pmatrix} = \begin{pmatrix} (u_x + v_y)_x \\ (u_x + v_y)_y \end{pmatrix} = \nabla(\nabla \cdot \boldsymbol{u}) \,. \end{aligned}$$

Since the velocity is divergence free, it holds

$$\nabla(\nabla \cdot \boldsymbol{u}) = \nabla 0 = \boldsymbol{0} \,.$$

□

Thus, the incompressible Navier-Stokes equations read

$$\rho \left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \right) = -\nabla p + \eta \Delta \boldsymbol{u} - \rho \boldsymbol{g} \,, \tag{3.22a}$$

$$\nabla \cdot \boldsymbol{u} = 0 \,. \tag{3.22b}$$

Note that $\nabla p \neq \nabla p(\rho)$ in the incompressible case. Here, p is a Lagrange multiplier enforcing the incompressibility of the system. Therefore, ∇p can be redefined by the addition of arbitrary gradients of scalar functions, which we will use later.

Let us consider a mixture of two fluids with different velocities $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Let φ be the volume fraction of fluid one and $1 - \varphi$ of fluid two. Hence, the total volume is a conserved quantity and therefore, the fluids obey the continuity equations

$$\frac{\partial \varphi}{\partial t} + \nabla \cdot (\varphi \boldsymbol{u}_1) = 0 \,, \tag{3.23a}$$

$$-\frac{\partial \varphi}{\partial t} + \nabla \cdot ((1 - \varphi)\boldsymbol{u}_2) = 0 \,. \tag{3.23b}$$

The volume-averaged velocity $\boldsymbol{u}$ of both fluids is given by

$$\boldsymbol{u} = \varphi \boldsymbol{u}_1 + (1 - \varphi)\boldsymbol{u}_2 \,.$$

Summing up the continuity equations (3.23a) and (3.23b) yields

$$\nabla \cdot (\varphi \boldsymbol{u}_1 + (1 - \varphi)\boldsymbol{u}_2) = 0 \,,$$

i.e., the volume-averaged velocity $\boldsymbol{u}$ is incompressible.

The dynamics of $\boldsymbol{u}$ follow the aforementioned momentum equation (3.16). It can be shown, using the least-action principle and the mixing energy (3.1), that the total stress of the two-fluid model reads

$$\mathbf{T} = 2\eta(\varphi)\boldsymbol{D}(\boldsymbol{u}) + \boldsymbol{\Pi}(\varphi)\,,$$

where the viscosity η now depends on the volume fraction in case that the two fluids have different viscosities. $\boldsymbol{\Pi}(\varphi) \in \mathbb{R}^{d\times d}$ is the osmotic tensor, also called Korteweg stress, which is extra elastic stress induced by the interfacial surface tension between the fluids and reads

$$\boldsymbol{\Pi}(\varphi) = \frac{\partial E_{mix}}{\partial \nabla\varphi} \otimes \nabla\varphi = -\lambda \nabla\varphi \otimes \nabla\varphi\,.$$

Then, by replacing the time derivative in the Cahn-Hilliard equation (3.7) with the material derivative and by neglecting the body accelerations $\boldsymbol{g}$ in the momentum equation (3.16), we have the incompressible Cahn-Hilliard-Navier-Stokes model

$$\frac{\partial \varphi}{\partial t} + \boldsymbol{u}\cdot\nabla\varphi = \nabla\cdot\big(M(\varphi)\nabla\mu\big)\,, \tag{3.24a}$$

$$\mu = -\lambda\Delta\varphi + f(\varphi) \tag{3.24b}$$

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\right) = -\nabla p + \nabla\cdot\left\{\eta(\varphi)\Big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\Big]\right\} - \nabla\cdot(\lambda\nabla\varphi\otimes\nabla\varphi)\,, \tag{3.24c}$$

$$\nabla\cdot\boldsymbol{u} = 0\,. \tag{3.24d}$$

Its total energy consists of the mixing energy and the kinetic energy and reads (for $\rho = const.$)

$$E_{total}(\varphi,\boldsymbol{u}) = E_{mix}(\varphi) + E_{kin}(\boldsymbol{u}) = \int_\Omega \left(\frac{\lambda}{2}|\nabla\varphi|^2 + F(\varphi)\right) + \int_\Omega \frac{1}{2}\rho|\boldsymbol{u}|^2\,. \tag{3.25}$$

The original model which led to the Cahn-Hilliard-Navier-Stokes model (3.24) is the so-called *Model H* by *Hohenberg and Halperin* [44].

In order to prove that the free energy of a solution to system (3.24) is non-increasing in time, we need the following lemma.

Lemma 3.6. *The divergence of the interfacial elastic stress satisfies*

$$\nabla\cdot(\lambda\nabla\varphi\otimes\nabla\varphi) = -\mu\nabla\varphi + \nabla\left(\frac{\lambda}{2}|\nabla\varphi|^2 + F(\varphi)\right)\,. \tag{3.26}$$

Proof. For $\boldsymbol{e}_i$ the i-th unit vector and d the spatial dimension the following relation holds true

$$\begin{aligned}
\nabla \cdot (\lambda \nabla\varphi \otimes \nabla\varphi) &= \lambda \sum_{i,j=1}^{d} \frac{\partial}{\partial x_j} \left(\frac{\partial \varphi}{\partial x_i} \frac{\partial \varphi}{\partial x_j} \right) \boldsymbol{e}_i \\
&= \lambda \sum_{i,j=1}^{d} \left(\frac{\partial \varphi}{\partial x_i} \frac{\partial^2 \varphi}{\partial x_j^2} + \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \frac{\partial \varphi}{\partial x_j} \right) \boldsymbol{e}_i \\
&= \lambda \sum_{j=1}^{d} \frac{\partial^2 \varphi}{\partial x_j^2} \sum_{i=1}^{d} \frac{\partial \varphi}{\partial x_i} \boldsymbol{e}_i + \frac{\lambda}{2} \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \sum_{j=1}^{d} \left(\frac{\partial \varphi}{\partial x_j} \right)^2 \boldsymbol{e}_i \\
&= \lambda \sum_{j=1}^{d} \frac{\partial^2 \varphi}{\partial x_j^2} \sum_{i=1}^{d} \frac{\partial \varphi}{\partial x_i} \boldsymbol{e}_i + \frac{\lambda}{2} \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \sum_{j=1}^{d} \left(\frac{\partial \varphi}{\partial x_j} \right)^2 \boldsymbol{e}_i + \left(\frac{\partial F(\varphi)}{\partial \varphi} - f(\varphi) \right) \sum_{i=1}^{d} \frac{\partial \varphi}{\partial x_i} \boldsymbol{e}_i \\
&= \left(\lambda \sum_{j=1}^{d} \frac{\partial^2 \varphi}{\partial x_j^2} - f(\varphi) \right) \sum_{i=1}^{d} \frac{\partial \varphi}{\partial x_i} \boldsymbol{e}_i + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \boldsymbol{e}_i \left(\frac{\lambda}{2} \sum_{j=1}^{d} \left(\frac{\partial \varphi}{\partial x_j} \right)^2 + F(\varphi) \right) \\
&= \big(\lambda \Delta\varphi - f(\varphi) \big) \nabla\varphi + \nabla \left(\frac{\lambda}{2} |\nabla\varphi|^2 + F(\varphi) \right) \\
&= -\mu \nabla\varphi + \nabla \left(\frac{\lambda}{2} |\nabla\varphi|^2 + F(\varphi) \right).
\end{aligned}$$

□

Since the Cahn-Hilliard-Navier-Stokes model (3.24) is incompressible, the pressure is a Lagrange multiplier like in the incompressible Navier-Stokes equations (3.22). Therefore, we can introduce a modified pressure term

$$\tilde{p} = p + \frac{\lambda}{2} |\nabla\varphi|^2 + F(\varphi).$$

Thus, by using relation (3.26), we can rewrite system (3.24) to

$$\frac{\partial \varphi}{\partial t} + \boldsymbol{u} \cdot \nabla\varphi = \nabla \cdot \big(M(\varphi) \nabla\mu \big), \tag{3.27a}$$

$$\mu = -\lambda \Delta\varphi + f(\varphi) \tag{3.27b}$$

$$\rho \left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \right) = -\nabla\tilde{p} + \nabla \cdot \left\{ \eta(\varphi) \left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T \right] \right\} + \mu \nabla\varphi, \tag{3.27c}$$

$$\nabla \cdot \boldsymbol{u} = 0. \tag{3.27d}$$

Remark 3.7. *In order to have an initial-boundary-value problem, we complement the Cahn-Hilliard-Navier-Stokes model (3.27) with suitable initial and boundary conditions. Our initial conditions read*

$$(\varphi, \boldsymbol{u})|_{t=0} = (\varphi^0, \boldsymbol{u}^0) \quad \text{in } \Omega.$$

Suitable boundary conditions include

- *periodic boundary conditions and*

- *homogeneous Neumann boundary conditions for φ and μ*

$$\frac{\partial \varphi}{\partial \boldsymbol{n}}\Big|_{\partial\Omega} = \frac{\partial \mu}{\partial \boldsymbol{n}}\Big|_{\partial\Omega} = 0 \quad \text{on } \partial\Omega \times (0,\infty)\,,$$

coupled with homogeneous Dirichlet or no-slip boundary conditions for $\boldsymbol{u}$, both yielding

$$\boldsymbol{u} \cdot \boldsymbol{n} = 0 \quad \text{on } \partial\Omega \times (0,\infty)\,.$$

Note that these are the most common choices in the literature studying the Cahn-Hilliard-Navier-Stokes model and that we will rely on periodic boundary conditions for our numerical experiments in Chapter 5. Further, note that all above-mentioned choices are suitable for the derivation of energy laws throughout this thesis.

Theorem 3.8. *Assuming suitable boundary conditions, see Remark 3.7, the Cahn-Hilliard-Navier-Stokes model (3.27) satisfies the following energy law*

$$\frac{\mathrm{d}E_{total}(\varphi,\boldsymbol{u})}{\mathrm{d}t} = -\int_\Omega M(\varphi)|\nabla\mu|^2 - \int_\Omega 2\eta(\varphi)|\boldsymbol{D}(\boldsymbol{u})|^2\,. \tag{3.28}$$

Since the mobility $M(\varphi)$ and the viscosity $\eta(\varphi)$ are non-negative, model (3.27) is thermodynamically consistent.

Proof. Analogously to the derivation of energy law (3.8), equation (3.27a) is multiplied by μ and integrated over Ω. Assuming suitable boundary conditions, see Remark 3.7, we obtain

$$\frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \int_\Omega M(\varphi)|\nabla\mu|^2 + \int_\Omega \boldsymbol{u}\cdot\nabla\varphi\mu = 0\,.$$

Taking the inner product (dot product) of (3.27c) and $\boldsymbol{u}$ and integrating over Ω yields

$$\begin{aligned}
0 &= \int_\Omega \rho\frac{\partial \boldsymbol{u}}{\partial t}\cdot\boldsymbol{u} + \int_\Omega \rho(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\cdot\boldsymbol{u} - \int_\Omega \left(\nabla\cdot\left\{\eta(\varphi)\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right]\right\}\right)\cdot\boldsymbol{u} \\
&\quad + \int_\Omega \nabla\tilde{p}\cdot\boldsymbol{u} - \int_\Omega \mu\nabla\varphi\cdot\boldsymbol{u} \\
&= \int_\Omega \rho\frac{1}{2}\frac{\partial|\boldsymbol{u}|^2}{\partial t} + \int_\Omega \rho(\boldsymbol{u}\cdot\nabla)\frac{1}{2}|\boldsymbol{u}|^2 + \int_\Omega \left\{\eta(\varphi)\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right]\right\} : \nabla\boldsymbol{u} \\
&\quad - \int_{\partial\Omega}\left(\left\{\eta(\varphi)\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right]\right\}\boldsymbol{n}\right)\cdot\boldsymbol{u} - \int_\Omega \tilde{p}(\nabla\cdot\boldsymbol{u}) + \int_{\partial\Omega}\tilde{p}\boldsymbol{u}\cdot\boldsymbol{n} - \int_\Omega \mu\nabla\varphi\cdot\boldsymbol{u} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{1}{2}\rho|\boldsymbol{u}|^2 - \int_\Omega \rho(\nabla\cdot\boldsymbol{u})\frac{1}{2}|\boldsymbol{u}|^2 + \int_{\partial\Omega}\rho(\boldsymbol{u}\cdot\boldsymbol{n})\frac{1}{2}|\boldsymbol{u}|^2 + \int_\Omega \eta(\varphi)\left[|\nabla\boldsymbol{u}|^2 + \mathrm{tr}\left((\nabla\boldsymbol{u})^2\right)\right] \\
&\quad - \int_{\partial\Omega}\left(\left\{\eta(\varphi)\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right]\right\}\boldsymbol{n}\right)\cdot\boldsymbol{u} - \int_\Omega (\nabla\cdot\boldsymbol{u})\tilde{p} + \int_{\partial\Omega}\tilde{p}\boldsymbol{u}\cdot\boldsymbol{n} - \int_\Omega \boldsymbol{u}\cdot\nabla\varphi\mu \\
&= \frac{\mathrm{d}E_{kin}(\boldsymbol{u})}{\mathrm{d}t} - \int_\Omega(\nabla\cdot\boldsymbol{u})(\frac{1}{2}\rho|\boldsymbol{u}|^2 + \tilde{p}) + \int_\Omega \frac{\eta(\varphi)}{2}\sum_{i,j=1}^{d}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)^2 \\
&\quad + \int_{\partial\Omega}\left(\frac{1}{2}\rho|\boldsymbol{u}|^2\boldsymbol{n} + \tilde{p}\boldsymbol{n} - \left\{\eta(\varphi)\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right]\right\}\boldsymbol{n}\right)\cdot\boldsymbol{u} - \int_\Omega \boldsymbol{u}\cdot\nabla\varphi\mu\,.
\end{aligned}$$

Taking into account the incompressibility (3.27d), $\int_\Omega (\nabla \cdot \boldsymbol{u})(\frac{1}{2}\rho|\boldsymbol{u}|^2 + \tilde{p})$ vanishes. Assuming suitable boundary conditions, see Remark 3.7, the boundary integral vanishes, too. Furthermore,

$$\int_\Omega \frac{\eta(\varphi)}{2} \sum_{i,j=1}^{d} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)^2 = \int_\Omega \frac{\eta(\varphi)}{2} |\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T|^2 = \int_\Omega 2\eta(\varphi)|\boldsymbol{D}(\boldsymbol{u})|^2 \,.$$

Thus, summing up the above relations, we obtain

$$\frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \int_\Omega M(\varphi)|\nabla\mu|^2 + \int_\Omega \boldsymbol{u} \cdot \nabla\varphi\mu + \frac{\mathrm{d}E_{kin}(\boldsymbol{u})}{\mathrm{d}t} + 2\int_\Omega \eta(\varphi)|\boldsymbol{D}(\boldsymbol{u})|^2 - \int_\Omega \boldsymbol{u} \cdot \nabla\varphi\mu = 0 \,,$$

which is equivalent to the desired energy law (3.28). □

3.3 Viscoelastic two-fluid models

Now, let us discuss complex fluids, in particular polymer solutions. We are interested in the phase behavior of polymer-solvent mixtures, which usually consist of a Newtonian fluid filled with a certain amount of polymer chains. Polymer chains are often coarse grained to elastic bead and spring dumbbells. Such dumbbells induce the inclusion of the elastic stress $\boldsymbol{\sigma} \in \mathbb{R}^{d\times d}$ in the aforementioned total stress tensor $\mathbf{T}$. For now, let us neglect the phase behavior, i.e., we assume the mixture to be homogeneous. We define the total stress of a viscoelastic fluid by

$$\mathbf{T} = 2\eta_s \boldsymbol{D}(\boldsymbol{u}) + \boldsymbol{\sigma} \,, \tag{3.29}$$

where $2\eta_s \boldsymbol{D}(\boldsymbol{u})$ is the solvent part, which is equal to stress tensor (3.18), with η_s being the viscosity of the solvent. To determine the time evolution of the polymeric part $\boldsymbol{\sigma}$, we use the upper-convected Maxwell model

$$\boldsymbol{\sigma} + \tau \overset{\triangledown}{\boldsymbol{\sigma}} = 2\eta_p \boldsymbol{D}(\boldsymbol{u}), \tag{3.30}$$

where τ is the relaxation time of the polymers, η_p is their viscosity and $\overset{\triangledown}{\boldsymbol{\sigma}}$ is the upper-convected time derivative of the elastic stress, which reads

$$\overset{\triangledown}{\boldsymbol{\sigma}} = \frac{\partial \boldsymbol{\sigma}}{\partial t} + \boldsymbol{u} \cdot \nabla\boldsymbol{\sigma} - \left[(\nabla \boldsymbol{u})\boldsymbol{\sigma} + \boldsymbol{\sigma}(\nabla \boldsymbol{u})^T \right] . \tag{3.31}$$

Together, these are the rheological equations for an Oldroyd-B fluid, see, e.g., *Larson* [50]. To determine the dynamics of an Oldroyd-B fluid, we supplement the rheological equations by a generalization of the incompressible Navier-Stokes equations (3.22), wherein the viscous stress (3.18) is replaced by the viscoelastic stress (3.29). Summarized, the Oldroyd-B model for incompressible fluids reads

$$\rho \left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \right) = -\nabla p + \nabla \cdot \left\{ \eta_s \left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T \right] \right\} + \nabla \cdot \boldsymbol{\sigma} - \rho \boldsymbol{g} \,, \tag{3.32a}$$

$$\nabla \cdot \boldsymbol{u} = 0 \,, \tag{3.32b}$$

$$\tau \left(\frac{\partial \boldsymbol{\sigma}}{\partial t} + \boldsymbol{u} \cdot \nabla\boldsymbol{\sigma} - \left[(\nabla \boldsymbol{u})\boldsymbol{\sigma} + \boldsymbol{\sigma}(\nabla \boldsymbol{u})^T \right] \right) = -\boldsymbol{\sigma} + \eta_p \left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T \right] . \tag{3.32c}$$

Now, in order to analyze the phase behavior of polymer-solvent mixtures at a fixed temperature, we extend the Cahn-Hilliard-Navier-Stokes model (3.24) to a viscoelastic phase-field model. This has been done at first by *Helfand and Fredrickson* [43], also by considering the elastic stress $\boldsymbol{\sigma}$ in addition to the viscous stress $2\eta_s \boldsymbol{D}(\boldsymbol{u})$ and the Korteweg surface stress $\boldsymbol{\Pi}(\varphi)$. Thus, for viscoelastic fluids with a diffusive interface, the total stress reads

$$\mathbf{T} = 2\eta_s \boldsymbol{D}(\boldsymbol{u}) + \boldsymbol{\sigma} + \boldsymbol{\Pi}(\varphi)\,. \tag{3.33}$$

The viscoelastic two-fluid model by *Helfand and Fredrickson* is further investigated by, e.g., *Onuki, Doi and Milner* [58, 60, 62], and can be written

$$\frac{\partial \varphi}{\partial t} = -\nabla \cdot (\boldsymbol{u}\varphi) + \nabla \cdot \frac{\varphi(1-\varphi)^2}{\zeta(\varphi)} \left[\varphi \nabla \mu - \nabla \cdot \boldsymbol{\sigma}\right], \tag{3.34a}$$

$$\rho \frac{D\boldsymbol{u}}{Dt} = -\nabla p + \eta_s \Delta \boldsymbol{u} + \nabla \cdot \boldsymbol{\sigma} + \mu \nabla \varphi\,, \tag{3.34b}$$

$$\nabla \cdot \boldsymbol{u} = 0\,, \tag{3.34c}$$

with the assumption that one may set $\frac{D\boldsymbol{u}}{Dt} \cong \frac{\partial \boldsymbol{u}}{\partial t}$. Like in the Oldroyd-B model (3.32), the classical example of a constitutive equation determining the stress tensor $\boldsymbol{\sigma}$ is used, the upper-convected Maxwell model

$$\frac{\partial \boldsymbol{\sigma}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{\sigma} = (\nabla \boldsymbol{u})\boldsymbol{\sigma} + \boldsymbol{\sigma}(\nabla \boldsymbol{u})^T - \frac{1}{\tau(\varphi)}\boldsymbol{\sigma} + G_S(\varphi)\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right], \tag{3.35}$$

where the relaxation time τ is now dependent on the local polymer volume fraction. Following the description of the Oldroyd-B model above, the so-called shear (relaxation) modulus or rigidity $G_S(\varphi)$ is given by the polymer viscosity $\eta_p(\varphi)$ divided by the relaxation time $\tau(\varphi)$.

Since numerical solutions of system (3.34), (3.35) do not reproduce all the essential features of viscoelastic phase separation observed experimentally, *Tanaka and Araki* [68, 69] suggest incorporating the effect of volume change into the stress tensor. For this aim, they include the usually neglected bulk relaxation modulus $G_B(\varphi)$ as a controllable parameter, and separate the elastic stress into what they call shear stress $\boldsymbol{\sigma}_S$ and bulk stress $\boldsymbol{\sigma}_B$. *Tanaka and Araki* propose the separate calculation of both stress parts by the similar upper-convected Maxwell equations

$$\frac{\partial \boldsymbol{\sigma}_w}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{\sigma}_w = (\nabla \boldsymbol{u})\boldsymbol{\sigma}_w + \boldsymbol{\sigma}_w(\nabla \boldsymbol{u})^T - \frac{1}{\tau_w(\varphi)}\boldsymbol{\sigma}_w + G_w(\varphi)\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right], \tag{3.36}$$

with $w \in \{B, S\}$, followed by the following final corrections $\boldsymbol{\sigma}_w^f$.

$$\boldsymbol{\sigma}_S^f = \boldsymbol{\sigma}_S - \frac{1}{d}\mathrm{tr}(\boldsymbol{\sigma}_S)\boldsymbol{I}\,, \tag{3.37}$$

where $\boldsymbol{I} \in \mathbb{R}^{d\times d}$ is the identity (or unit tensor), i.e., the final shear stress is traceless as expected of a shear stress tensor.

$$\boldsymbol{\sigma}_B^f = \frac{1}{d}\mathrm{tr}(\boldsymbol{\sigma}_B)\boldsymbol{I}\,, \tag{3.38}$$

i.e., the final bulk stress is diagonal with identical values. Hence, the total elastic stress reads

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_S^f + \boldsymbol{\sigma}_B^f = \boldsymbol{\sigma}_S - \frac{1}{d}\mathrm{tr}(\boldsymbol{\sigma}_S)\boldsymbol{I} + \frac{1}{d}\mathrm{tr}(\boldsymbol{\sigma}_B)\boldsymbol{I}\,. \tag{3.39}$$

The numerical solutions of model (3.34), (3.36), (3.39) shown in [68, 69] approve that such a model reproduces all essential features of viscoelastic phase separation observed in lab experiments.

However, the respective authors do not show if the two above-mentioned viscoelastic two-fluid models are thermodynamically consistent. Other than the mixing energy and the kinetic energy, the free energy of the elastic stress is not definitely determined. It is regularly supposed to be given by

$$E_{el}(\boldsymbol{\sigma}) = \int_\Omega \frac{1}{2}\mathrm{tr}\boldsymbol{\sigma}\,, \tag{3.40}$$

see, e.g., [45, 74]. Thereby, the total energy of the two viscoelastic two-fluid models (3.34), (3.35) and (3.34), (3.36), (3.39) reads

$$E_{total}(\varphi, \boldsymbol{u}, \boldsymbol{\sigma}) = E_{mix}(\varphi) + E_{kin}(\boldsymbol{u}) + E_{el}(\boldsymbol{\sigma}) = \int_\Omega \left(\frac{\lambda}{2}|\nabla\varphi|^2 + F(\varphi)\right) + \int_\Omega \frac{1}{2}\rho|\boldsymbol{u}|^2 + \int_\Omega \frac{1}{2}\mathrm{tr}\boldsymbol{\sigma}\,,$$

where for the second model $\mathrm{tr}\boldsymbol{\sigma} = \mathrm{tr}(\boldsymbol{\sigma}_B^f)$, since $\boldsymbol{\sigma}_S^f$ is traceless.

Theorem 3.9. *We assume that the free energy of the elastic stress tensor is given by (3.40) and that suitable boundary conditions are used, see Remark 3.7 again, since no specific boundary conditions for $\boldsymbol{\sigma}$ are required for what follows. Then, the viscoelastic two-fluid model (3.34), (3.35) satisfies the following energy law*

$$\begin{aligned}\frac{\mathrm{d}E_{total}(\varphi, \boldsymbol{u}, \boldsymbol{\sigma})}{\mathrm{d}t} =& -\int_\Omega \frac{(\varphi(1-\varphi))^2}{\zeta(\varphi)}|\nabla\mu|^2 - \int_\Omega \eta_s|\nabla\boldsymbol{u}|^2 \\ &- \int_\Omega \frac{1}{2\tau(\varphi)}\mathrm{tr}\boldsymbol{\sigma} + \int_\Omega \frac{\varphi(1-\varphi)^2}{\zeta(\varphi)}(\nabla\cdot\boldsymbol{\sigma})\cdot\nabla\mu\,.\end{aligned} \tag{3.41}$$

Obviously, the last integral can be positive, i.e.

$$\frac{\mathrm{d}E_{total}(\varphi, \boldsymbol{u}, \boldsymbol{\sigma})}{\mathrm{d}t} \leq 0$$

does not necessarily hold.

Energy law (3.41) is also satisfied by model (3.34), (3.39), (3.36). Here, the elastic dissipation term becomes

$$-\int_\Omega \frac{1}{2\tau_B(\varphi)}\mathrm{tr}(\boldsymbol{\sigma}_B^f)\,,$$

since $\boldsymbol{\sigma}_S^f$ is traceless, i.e., $\mathrm{tr}(\boldsymbol{\sigma}_S^f) = 0$.

Proof. Analogously to the derivation of energy law (3.28), equation (3.34a) is multiplied by μ and integrated over Ω, yielding

$$\begin{aligned}0 =& \int_\Omega \frac{\partial\varphi}{\partial t}\mu + \int_\Omega \nabla\cdot(\boldsymbol{u}\varphi)\mu - \int_\Omega \nabla\cdot\frac{\varphi(1-\varphi)^2}{\zeta(\varphi)}\left[\varphi\nabla\mu - \nabla\cdot\boldsymbol{\sigma}\right]\mu \\ =& \int_\Omega \frac{\partial\varphi}{\partial t}\frac{\delta E_{mix}(\varphi)}{\delta t} + \int_\Omega (\nabla\cdot\boldsymbol{u})\varphi\mu + \int_\Omega \boldsymbol{u}\cdot\nabla\varphi\mu \\ &+ \int_\Omega \frac{\varphi(1-\varphi)^2}{\zeta(\varphi)}\left[\varphi\nabla\mu - \nabla\cdot\boldsymbol{\sigma}\right]\cdot\nabla\mu - \int_{\partial\Omega} \boldsymbol{n}\cdot\frac{\varphi(1-\varphi)^2}{\zeta(\varphi)}\left[\varphi\nabla\mu - \nabla\cdot\boldsymbol{\sigma}\right]\mu\,.\end{aligned} \tag{3.42}$$

Assuming suitable boundary conditions, see Remark 3.7, the boundary integral vanishes. Furthermore, using the incompressibility condition $\nabla \cdot \boldsymbol{u} = 0$, (3.42) reduces to

$$0 = \frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \int_\Omega \boldsymbol{u} \cdot \nabla\varphi\mu + \int_\Omega \frac{(\varphi(1-\varphi))^2}{\zeta(\varphi)}|\nabla\mu|^2 - \int_\Omega \frac{\varphi(1-\varphi)^2}{\zeta(\varphi)}(\nabla \cdot \boldsymbol{\sigma}) \cdot \nabla\mu\,. \tag{3.43}$$

Also analogously to the derivation of energy law (3.28), we take the inner product of equation (3.34b) and $\boldsymbol{u}$ and integrate over Ω, yielding

$$\begin{aligned} 0 &= \int_\Omega \rho\frac{\partial \boldsymbol{u}}{\partial t} \cdot \boldsymbol{u} + \int_\Omega \nabla p \cdot \boldsymbol{u} - \int_\Omega \eta_s(\Delta \boldsymbol{u}) \cdot \boldsymbol{u} - \int_\Omega (\nabla \cdot \boldsymbol{\sigma}) \cdot \boldsymbol{u} - \int_\Omega \mu\nabla\varphi \cdot \boldsymbol{u} \\ &= \int_\Omega \rho\frac{1}{2}\frac{\partial|\boldsymbol{u}|^2}{\partial t} - \int_\Omega p(\nabla \cdot \boldsymbol{u}) + \int_{\partial\Omega} p\boldsymbol{u} \cdot \boldsymbol{n} + \int_\Omega \eta_s\nabla\boldsymbol{u} \cdot \nabla\boldsymbol{u} - \int_{\partial\Omega} \eta_s\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}} \cdot \boldsymbol{u} \\ &\quad + \int_\Omega \boldsymbol{\sigma} : \nabla\boldsymbol{u} - \int_{\partial\Omega} (\boldsymbol{\sigma}\boldsymbol{n}) \cdot \boldsymbol{u} - \int_\Omega \boldsymbol{u} \cdot \nabla\varphi\mu\,. \end{aligned}$$

Due to the incompressibility and suitable boundary conditions, see Remark 3.7, the equation reduces to

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{1}{2}\rho|\boldsymbol{u}|^2 + \int_\Omega \eta_s|\nabla\boldsymbol{u}|^2 + \int_\Omega \boldsymbol{\sigma} : \nabla\boldsymbol{u} - \int_\Omega \boldsymbol{u} \cdot \nabla\varphi\mu\,. \tag{3.44}$$

Now, we take the Frobenius inner product (double dot product) of (3.35) and $\frac{1}{2}\boldsymbol{I}$, where $\boldsymbol{I}$ is the identity, and integrate over Ω, yielding

$$\begin{aligned} 0 &= \int_\Omega \frac{\partial \boldsymbol{\sigma}}{\partial t} : \frac{1}{2}\boldsymbol{I} + \int_\Omega (\boldsymbol{u} \cdot \nabla)\boldsymbol{\sigma} : \frac{1}{2}\boldsymbol{I} - \int_\Omega \left[(\nabla\boldsymbol{u})\boldsymbol{\sigma} + \boldsymbol{\sigma}(\nabla\boldsymbol{u})^T\right] : \frac{1}{2}\boldsymbol{I} \\ &\quad + \int_\Omega \frac{1}{\tau(\varphi)}\boldsymbol{\sigma} : \frac{1}{2}\boldsymbol{I} - \int_\Omega G(\varphi)\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right] : \frac{1}{2}\boldsymbol{I}\,. \end{aligned}$$

Recalling that for all matrices $\boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{n\times n}, n \in \mathbb{N}$,

$$\boldsymbol{C} : \boldsymbol{D} = \mathrm{tr}(\boldsymbol{C}\boldsymbol{D}^T)\,, \qquad \boldsymbol{C} : \boldsymbol{I} = \mathrm{tr}(\boldsymbol{C}\boldsymbol{I}) = \mathrm{tr}(\boldsymbol{C})\,,$$

and that the elastic stress $\boldsymbol{\sigma}$ is a symmetric tensor, we get

$$\begin{aligned} 0 &= \int_\Omega \frac{1}{2}\mathrm{tr}\left(\frac{\partial \boldsymbol{\sigma}}{\partial t}\right) + \int_\Omega \frac{1}{2}(\boldsymbol{u} \cdot \nabla)\mathrm{tr}\boldsymbol{\sigma} - \int_\Omega \frac{1}{2}\left(\nabla\boldsymbol{u} : \boldsymbol{\sigma}^T + \boldsymbol{\sigma} : \nabla\boldsymbol{u}\right) \\ &\quad + \int_\Omega \frac{1}{2\tau(\varphi)}\mathrm{tr}\boldsymbol{\sigma} - \int_\Omega G(\varphi)\mathrm{tr}(\nabla\boldsymbol{u}) \\ &= \int_\Omega \frac{1}{2}\frac{\partial \mathrm{tr}\boldsymbol{\sigma}}{\partial t} - \int_\Omega \frac{1}{2}(\nabla \cdot \boldsymbol{u})\mathrm{tr}\boldsymbol{\sigma} + \int_{\partial\Omega} \frac{1}{2}(\boldsymbol{u} \cdot \boldsymbol{n})\mathrm{tr}\boldsymbol{\sigma} - \int_\Omega \boldsymbol{\sigma} : \nabla\boldsymbol{u} \\ &\quad + \int_\Omega \frac{1}{2\tau(\varphi)}\mathrm{tr}\boldsymbol{\sigma} - \int_\Omega G(\varphi)(\nabla \cdot \boldsymbol{u})\,. \end{aligned}$$

Again due to the incompressibility and assuming the same suitable boundary conditions for $\boldsymbol{u}$ as above, the equation reduces to

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{1}{2}\mathrm{tr}\boldsymbol{\sigma} - \int_\Omega \boldsymbol{\sigma} : \nabla\boldsymbol{u} + \int_\Omega \frac{1}{2\tau(\varphi)}\mathrm{tr}\boldsymbol{\sigma}\,. \tag{3.45}$$

Summing up (3.43), (3.44) and (3.45) yields energy law (3.41). □

In the thematically still recent paper [74], *Zhou, Zhang and E* propose a modified model for polymer-solvent phase-fields. It is based on the model by *Tanaka and Araki* but supposedly thermodynamically consistent.

Zhou et al. derive their model from the total energy

$$\begin{aligned} E_{total}(\varphi, \boldsymbol{u}, q, \boldsymbol{\sigma}) &= E_{mix}(\varphi) + E_{kin}(\boldsymbol{u}) + E_{bulk}(q) + E_{el}(\boldsymbol{\sigma}) \\ &= \int_\Omega \left(\frac{\lambda}{2} |\nabla \varphi|^2 + F(\varphi) \right) + \int_\Omega \frac{1}{2} |\boldsymbol{u}|^2 + \int_\Omega \frac{1}{2} q^2 + \int_\Omega \frac{1}{2} \mathrm{tr} \boldsymbol{\sigma} \,, \end{aligned} \tag{3.46}$$

where $q\boldsymbol{I}$ corresponds to the bulk stress tensor $\boldsymbol{\sigma}_B^f$. Note that q being scalar is reasonable, since $\boldsymbol{\sigma}_B^f$ is diagonal with identical values as discussed above. According to *Zhou et al.*, the free energy of the bulk stress $E_{bulk}(q)$ is derived "from the chain conformational entropy of polymer molecules which can be regarded as some penalty function since the migration of polymer molecules will cause much more decrease in entropy compared with the solvents". $E_{el}(\boldsymbol{\sigma})$ is the same elastic energy of polymer molecules as given by (3.40) above, with $\boldsymbol{\sigma}$ supposedly being related to the (non-final) shear stress tensor $\boldsymbol{\sigma}_S$, which is not necessarily traceless. Therefore, we refer to it as elastic stress and not as shear stress. The kinetic energy $E_{kin}(\boldsymbol{u})$ does not contain the density parameter ρ here, i.e., the latter is set to one.

Starting from the total energy (3.46), *Zhou et al.* obtain by the principle of virtual work a modified model for polymer-solvent phase-fields, which reads

$$\frac{\partial \varphi}{\partial t} + \boldsymbol{u} \cdot \nabla \varphi = \nabla \cdot \left\{ \varphi(1-\varphi) \frac{1}{\zeta(\varphi)} \left[\varphi(1-\varphi) \nabla \mu - \nabla (G_B(\varphi) q) \right] \right\} , \tag{3.47a}$$

$$\frac{\partial q}{\partial t} + \boldsymbol{u} \cdot \nabla q = -\frac{1}{\tau_B(\varphi)} q - G_B(\varphi) \nabla \cdot \left\{ \frac{1}{\zeta(\varphi)} \left[\varphi(1-\varphi) \nabla \mu - \nabla (G_B(\varphi) q) \right] \right\} , \tag{3.47b}$$

$$\frac{\partial \boldsymbol{\sigma}}{\partial t} + (\boldsymbol{u} \cdot \nabla) \boldsymbol{\sigma} = (\nabla \boldsymbol{u}) \boldsymbol{\sigma} + \boldsymbol{\sigma} (\nabla \boldsymbol{u})^T - \frac{1}{\tau_S(\varphi)} \boldsymbol{\sigma} + G_S(\varphi) \left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T \right] , \tag{3.47c}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla) \boldsymbol{u} = -\nabla p + \nabla \cdot \left\{ \eta(\varphi) \left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T \right] \right\} - \nabla \cdot (\lambda \nabla \varphi \otimes \nabla \varphi) + \nabla \cdot \boldsymbol{\sigma} \,, \tag{3.47d}$$

$$\nabla \cdot \boldsymbol{u} = 0 \,, \tag{3.47e}$$

where the chemical potential reads $\mu = -\lambda \Delta \varphi + f(\varphi)$ like above, with $f(\varphi)$ being the derivative of the Flory-Huggins potential (3.11) specifically.

Based on the assumption that $\varphi \in (0, 1)$, with φ being the polymer volume fraction and thus $1 - \varphi$ being the solvent volume fraction, *Zhou et al.* have set the physical quantities of model (3.47) in the following way:

- The relaxation times read, following *Tanaka* [68]

  $$\tau_B(\varphi) = \tau_B^0 \varphi^2 \,, \quad \tau_S(\varphi) = \tau_S^0 \varphi^2 \,,$$

  with their coefficients τ_B^0 and τ_S^0 being positive constants.

- The shear relaxation modulus reads

  $$G_S(\varphi) = G_S^0 \varphi^2 \,,$$

  with G_S^0 also being a positive constant.

- The bulk relaxation modulus is given by

$$G_B(\varphi) = G_B^0 \left[1 + \tanh\left(\frac{\cot(\pi\varphi^*) - \cot(\pi\varphi)}{\varepsilon} \right) \right] + G_B^1 ,$$

  which is a smoothed step function from G_B^1 to G_B^0, where the smoothed jump occurs around φ^* and its width is controlled by ε, which hence should be small ($<< 1$). Here, φ^* is the critical concentration for polymers to crosslink.

Remark 3.10. *There are a few physical quantities in the model of Zhou et al., which we want to discuss in more detail. According to Zhou et al., the bulk stress q is derived from the polymer crosslinking, which does not occur noticeably below the critical concentration φ^*. Thus, it is reasonable to assume that the bulk modulus $G_B(\varphi)$ vanishes for $\varphi < \varphi^*$. For this purpose, G_B^1 has to be zero. Tanaka [68] uses a non-smoothed step function here, which also jumps from zero to a positive value.*

Note that the total elastic stress is given by

$$\boldsymbol{\sigma}_{total} = \boldsymbol{\sigma} + q\boldsymbol{I} .$$

Therefore, the elastic stress coupling term in the Navier-Stokes equation (3.47d) originally reads

$$\nabla \cdot (\boldsymbol{\sigma} + q\boldsymbol{I}) = \nabla \cdot \boldsymbol{\sigma} + \nabla q .$$

At this point, the property of p being a Lagrange multiplier (due to the incompressibility) is used. Thereby, ∇p contains the bulk stress coupling term ∇q in this model.

Furthermore, the deformation rate tensor in the Navier-Stokes equation (3.47d) is written in its general form, with the viscosity $\eta(\varphi)$ being a function of the volume fraction. This is reasonable, since $\boldsymbol{u}$ is a volume averaged velocity in this model. In the Oldroyd-B model (3.32), $\boldsymbol{u}$ is the solvent velocity, i.e., η is the constant viscosity of the solvent as long as we assume incompressibility.

However, we can transfer from the Oldroyd-B model that the shear modulus $G_S(\varphi)$ equals the fraction of polymer viscosity η_p and relaxation time τ_S. And according to Le Bris and Lelièvre [51], the polymer viscosity of an Oldroyd-B fluid is given by

$$\eta_p = n_p \tau_S k_B T .$$

Here, n_p is either the local polymer count or the total polymer count in the computational domain, with the latter being constant for appropriate boundary conditions. Thus, transferred to the model of Zhou et al., we have

$$G_S(\varphi) = \frac{\eta_p(\varphi)}{\tau_S(\varphi)} = \frac{n_p(\varphi)\tau_S(\varphi)k_B T}{\tau_S(\varphi)} = k_B T n_p(\varphi) ,$$

where $k_B T$ is constant, since the model is isothermal. Thus, a constant polymer count n_p results in a constant shear modulus G_S.

3.3.1 Simplified model (without hydrodynamics)

Zhou et al. [74] also consider the special case of model (3.47) without hydrodynamic transport, i. e.

$$\boldsymbol{u} = \varphi \boldsymbol{u}_p + (1-\varphi)\boldsymbol{u}_s = \boldsymbol{0}\,.$$

The resulting simplified model reads

$$\frac{\partial \varphi}{\partial t} = \nabla \cdot \left\{ \varphi(1-\varphi)\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big] \right\}, \tag{3.48a}$$

$$\frac{\partial q}{\partial t} = -\frac{1}{\tau_B(\varphi)}q - G_B(\varphi)\nabla \cdot \left\{ \frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big] \right\}. \tag{3.48b}$$

For the sake of simplicity, we denote model (3.47) as the full model and its simplification (3.48) as the simplified model in the following.

Remark 3.11. *In order to have an initial-boundary-value problem, we complement the simplified model (3.48) with suitable initial and boundary conditions. Our initial conditions read*

$$(\varphi, q)|_{t=0} = (\varphi^0, q^0) \quad \text{in } \Omega\,.$$

Suitable boundary conditions include

- *periodic boundary conditions, and*
- *the homogeneous Neumann boundary conditions*

$$\frac{\partial \varphi}{\partial \boldsymbol{n}}\Big|_{\partial\Omega} = \frac{\partial \mu}{\partial \boldsymbol{n}}\Big|_{\partial\Omega} = \frac{\partial q}{\partial \boldsymbol{n}}\Big|_{\partial\Omega} = 0 \quad \text{on } \partial\Omega \times (0,\infty)\,.$$

Note that these two options are the most common choices in the literature studying Cahn-Hilliard models and that we will rely on periodic boundary conditions for our numerical experiments in Chapter 5. Further, note that both above-mentioned choices are suitable for the derivation of energy laws throughout this thesis.

The total energy of the simplified model consists of the mixing energy and the bulk energy, reading

$$E_{total}(\varphi, q) = E_{mix}(\varphi) + E_{bulk}(q) = \int_\Omega \left(\frac{\lambda}{2}|\nabla\varphi|^2 + F(\varphi)\right) + \int_\Omega \frac{1}{2}q^2\,. \tag{3.49}$$

Theorem 3.12. *Assuming suitable boundary conditions, see Remark 3.11, problem (3.48) satisfies the following energy law*

$$\frac{\mathrm{d}E_{total}(\varphi, q)}{\mathrm{d}t} = -\int_\Omega \frac{1}{\tau_B(\varphi)}q^2 - \int_\Omega \frac{1}{\zeta(\varphi)}\Big|\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big|^2\,. \tag{3.50}$$

The bulk relaxation time $\tau_B(\varphi)$ and the friction $\zeta(\varphi)$ are both positive. Therefore, the simplified model (3.48) is thermodynamically consistent.

Note that for $\tau_B(\varphi) = \tau_B^0\varphi^2$ and $\varphi \neq 0$

$$\int_\Omega \frac{1}{\tau_B(\varphi)}q^2 = \frac{1}{\tau_B^0}\int_\Omega \left(\frac{q}{\varphi}\right)^2 = \frac{1}{\tau_B^0}\left\|\frac{q}{\varphi}\right\|^2_{L^2(\Omega)}\,.$$

Proof. Multiplying (3.48a) by μ, integrating over the computational domain Ω and applying integration by parts, we obtain

$$\begin{aligned}
0 &= \int_\Omega \frac{\partial \varphi}{\partial t}\mu - \int_\Omega \nabla\cdot\left\{\varphi(1-\varphi)\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\right\}\mu \\
&= \int_\Omega \frac{\partial \varphi}{\partial t}\frac{\delta E_{mix}(\varphi)}{\delta\varphi} + \int_\Omega \left\{\varphi(1-\varphi)\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\right\}\cdot\nabla\mu \\
&\quad - \int_{\partial\Omega} \boldsymbol{n}\cdot\varphi(1-\varphi)\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\mu \\
&= \frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \int_\Omega \frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\cdot\Big[\varphi(1-\varphi)\nabla\mu\Big] \\
&\quad - \int_{\partial\Omega}\frac{1}{\zeta(\varphi)}\left[\varphi(1-\varphi)\frac{\partial\mu}{\partial\boldsymbol{n}} - \frac{\partial(G_B(\varphi)q)}{\partial\boldsymbol{n}}\right]\varphi(1-\varphi)\mu\,.
\end{aligned}$$

Multiplying (3.48b) by q and integrating over Ω yields

$$\begin{aligned}
0 &= \int_\Omega \frac{\partial q}{\partial t}q + \int_\Omega \frac{1}{\tau_B(\varphi)}q^2 + \int_\Omega G_B(\varphi)\nabla\cdot\left\{\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\right\}q \\
&= \int_\Omega \frac{1}{2}\frac{\partial q^2}{\partial t} + \int_\Omega \frac{1}{\tau_B(\varphi)}q^2 + \int_\Omega \nabla\cdot\left\{\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\right\}G_B(\varphi)q \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_\Omega \frac{1}{2}q^2\right) + \int_\Omega \frac{1}{\tau_B(\varphi)}q^2 - \int_\Omega \frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\cdot\nabla(G_B(\varphi)q) \\
&\quad + \int_{\partial\Omega}\boldsymbol{n}\cdot\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]G_B(\varphi)q \\
&= \frac{\mathrm{d}E_{bulk}(q)}{\mathrm{d}t} + \int_\Omega \frac{1}{\tau_B(\varphi)}q^2 + \int_\Omega \frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\cdot\Big[-\nabla(G_B(\varphi)q)\Big] \\
&\quad + \int_{\partial\Omega}\frac{1}{\zeta(\varphi)}\left[\varphi(1-\varphi)\frac{\partial\mu}{\partial\boldsymbol{n}} - \frac{\partial(G_B(\varphi)q)}{\partial\boldsymbol{n}}\right]G_B(\varphi)q\,.
\end{aligned}$$

Summing up the above relations yields

$$\begin{aligned}
&\frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \frac{\mathrm{d}E_{bulk}(q)}{\mathrm{d}t} + \int_\Omega \frac{1}{\tau_B(\varphi)}q^2 \\
&+ \int_\Omega \frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\cdot\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big] \\
&- \int_{\partial\Omega}\frac{1}{\zeta(\varphi)}\left[\varphi(1-\varphi)\frac{\partial\mu}{\partial\boldsymbol{n}} - \frac{\partial(G_B(\varphi)q)}{\partial\boldsymbol{n}}\right]\Big(\varphi(1-\varphi)\mu - G_B(\varphi)q\Big) = 0\,,
\end{aligned}$$

where the second normal derivative of the boundary integral satisfies

$$\frac{\partial(G_B(\varphi)q)}{\partial\boldsymbol{n}} = \frac{\partial G_B(\varphi)}{\partial\boldsymbol{n}}q + G_B(\varphi)\frac{\partial q}{\partial\boldsymbol{n}} = \frac{\partial G_B(\varphi)}{\partial\varphi}\frac{\partial\varphi}{\partial\boldsymbol{n}}q + G_B(\varphi)\frac{\partial q}{\partial\boldsymbol{n}}\,.$$

Thus, assuming suitable boundary conditions, see Remark 3.11, the boundary integral vanishes, and we obtain

$$\frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \frac{\mathrm{d}E_{bulk}(q)}{\mathrm{d}t} + \int_\Omega \frac{1}{\tau_B(\varphi)}q^2 + \int_\Omega \frac{1}{\zeta(\varphi)}\Big|\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big|^2 = 0\,,$$

which is equivalent to the desired energy law (3.50). □

3.3.2 Full model (with hydrodynamics)

Let us now discuss the energy properties of the full two-fluid model for viscoelastic phase separation (3.47). For this aim, we rewrite the model, using the pressure term $\tilde{p} = p + \frac{\lambda}{2}|\nabla\varphi|^2 + F(\varphi)$, introduced in Section 3.2, together with Lemma 3.6, yielding

$$\frac{\partial\varphi}{\partial t} + \boldsymbol{u}\cdot\nabla\varphi = \nabla\cdot\left\{\varphi(1-\varphi)\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\right\}, \tag{3.51a}$$

$$\frac{\partial q}{\partial t} + \boldsymbol{u}\cdot\nabla q = -\frac{1}{\tau_B(\varphi)}q - G_B(\varphi)\nabla\cdot\left\{\frac{1}{\zeta(\varphi)}\Big[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big]\right\}, \tag{3.51b}$$

$$\frac{\partial\boldsymbol{\sigma}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{\sigma} = (\nabla\boldsymbol{u})\boldsymbol{\sigma} + \boldsymbol{\sigma}(\nabla\boldsymbol{u})^T - \frac{1}{\tau_S(\varphi)}\boldsymbol{\sigma} + G_S(\varphi)\Big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\Big], \tag{3.51c}$$

$$\frac{\partial\boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \nabla\cdot\left\{\eta(\varphi)\Big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\Big]\right\} - \nabla\tilde{p} + \mu\nabla\varphi + \nabla\cdot\boldsymbol{\sigma}, \tag{3.51d}$$

$$\nabla\cdot\boldsymbol{u} = 0. \tag{3.51e}$$

Remark 3.13. *In order to have an initial-boundary-value problem, we complement the full model (3.51) with suitable initial and boundary conditions. Our initial conditions read*

$$(\varphi, q, \boldsymbol{\sigma}, \boldsymbol{u})|_{t=0} = (\varphi^0, q^0, \boldsymbol{\sigma}^0, \boldsymbol{u}^0) \quad \text{in } \Omega.$$

Suitable boundary conditions include

- *periodic boundary conditions, and*
- *homogeneous Neumann boundary conditions for φ, μ and q*

  $$\frac{\partial\varphi}{\partial\boldsymbol{n}}\Big|_{\partial\Omega} = \frac{\partial\mu}{\partial\boldsymbol{n}}\Big|_{\partial\Omega} = \frac{\partial q}{\partial\boldsymbol{n}}\Big|_{\partial\Omega} = 0 \quad \text{on } \partial\Omega\times(0,\infty),$$

  coupled with homogeneous Dirichlet or no-slip boundary conditions for $\boldsymbol{u}$, both yielding

  $$\boldsymbol{u}\cdot\boldsymbol{n} = 0 \quad \text{on } \partial\Omega\times(0,\infty).$$

Note that these are the most common choices in the literature studying Cahn-Hilliard-Navier-Stokes models and that we will rely on periodic boundary conditions for our numerical experiments in Chapter 5. Further, note that all above-mentioned choices are suitable for the derivation of energy laws throughout this thesis and that no specific boundary conditions on $\boldsymbol{\sigma}$ are required for the latter.

Theorem 3.14. *Assuming suitable boundary conditions, see Remark 3.13, system (3.51) obeys the following energy law*

$$\begin{aligned}\frac{\mathrm{d}E_{total}(\varphi, q, \boldsymbol{\sigma}, \boldsymbol{u})}{\mathrm{d}t} = &-\int_\Omega \frac{1}{\tau_B(\varphi)}q^2 - \int_\Omega \frac{1}{\zeta(\varphi)}\Big|\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\Big|^2 \\ &- \int_\Omega \frac{1}{2\tau_S(\varphi)}\mathrm{tr}\boldsymbol{\sigma} - 2\int_\Omega \eta(\varphi)|\boldsymbol{D}(\boldsymbol{u})|^2.\end{aligned} \tag{3.52}$$

The bulk relaxation time $\tau_B(\varphi)$, the friction $\zeta(\varphi)$ and the viscosity $\eta(\varphi)$ are all positive. Consequently, the full model (3.51) is guaranteed to be thermodynamically consistent if the elastic stress satisfies $\mathrm{tr}\boldsymbol{\sigma} \geq 0$ for all $t \geq 0$, which is fulfilled under certain conditions, see Lemma 3.17 for details.

Proof. Analogously to the derivation of energy law (3.50), (3.51a) is multiplied by μ and (3.51b) by q and both are integrated over Ω. By the same calculations as above, we obtain

$$\begin{aligned}
&\frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \frac{\mathrm{d}E_{bulk}(q)}{\mathrm{d}t} + \int_\Omega \frac{1}{\tau_B(\varphi)} q^2 + \int_\Omega \frac{1}{\zeta(\varphi)} \left| \varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q) \right|^2 \\
&- \int_{\partial\Omega} \frac{1}{\zeta(\varphi)} \left[\varphi(1-\varphi)\frac{\partial\mu}{\partial \boldsymbol{n}} - \frac{\partial(G_B(\varphi)q)}{\partial \boldsymbol{n}} \right] \left(\varphi(1-\varphi)\mu - G_B(\varphi)q \right) \\
&+ \int_\Omega \boldsymbol{u} \cdot \nabla\varphi\mu + \int_\Omega \boldsymbol{u} \cdot \nabla q q = 0\,,
\end{aligned}$$

where the last term satisfies

$$\int_\Omega \boldsymbol{u} \cdot \nabla q q = \int_\Omega \boldsymbol{u} \cdot \nabla \frac{1}{2} q^2 = \int_{\partial\Omega} \boldsymbol{u} \cdot \boldsymbol{n} \frac{1}{2} q^2 - \int_\Omega (\nabla \cdot \boldsymbol{u}) \frac{1}{2} q^2 .$$

Due to the incompressibility condition $\nabla \cdot \boldsymbol{u} = 0$, the last term vanishes. Assuming suitable boundary conditions, see Remark 3.13, the boundary integrals vanish as well, and we obtain

$$\begin{aligned}
&\frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \frac{\mathrm{d}E_{bulk}(q)}{\mathrm{d}t} + \int_\Omega \frac{1}{\tau_B(\varphi)} q^2 \\
&+ \int_\Omega \frac{1}{\zeta(\varphi)} \left| \varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q) \right|^2 + \int_\Omega \boldsymbol{u} \cdot \nabla\varphi\mu = 0\,.
\end{aligned} \tag{3.53}$$

Now, analogously to the derivation of energy law (3.41), we take the Frobenius inner product of (3.51c) and $\frac{1}{2}\boldsymbol{I}$, and integrate over Ω, yielding

$$\begin{aligned}
&\int_\Omega \frac{1}{2} \frac{\partial \mathrm{tr}\boldsymbol{\sigma}}{\partial t} - \int_\Omega \frac{1}{2} (\nabla \cdot \boldsymbol{u}) \mathrm{tr}\boldsymbol{\sigma} + \int_{\partial\Omega} \frac{1}{2} (\boldsymbol{u} \cdot \boldsymbol{n}) \mathrm{tr}\boldsymbol{\sigma} \\
&- \int_\Omega \boldsymbol{\sigma} : \nabla\boldsymbol{u} + \int_\Omega \frac{1}{2\tau_S(\varphi)} \mathrm{tr}\boldsymbol{\sigma} - \int_\Omega G_S(\varphi)(\nabla \cdot \boldsymbol{u}) = 0\,.
\end{aligned}$$

Assuming suitable boundary conditions for $\boldsymbol{u}$, see Remark 3.13, the boundary integral vanishes, and due to the incompressibility of $\boldsymbol{u}$, we can further reduce the last equation to

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \frac{1}{2} \mathrm{tr}\boldsymbol{\sigma} - \int_\Omega \boldsymbol{\sigma} : \nabla\boldsymbol{u} + \int_\Omega \frac{1}{2\tau_S(\varphi)} \mathrm{tr}\boldsymbol{\sigma} = 0\,. \tag{3.54}$$

Taking the dot product of (3.51d) and $\boldsymbol{u}$ and integrating over Ω yields

$$\begin{aligned}
0 &= \int_\Omega \frac{\partial \boldsymbol{u}}{\partial t}\cdot\boldsymbol{u} + \int_\Omega (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\cdot\boldsymbol{u} - \int_\Omega \left(\nabla\cdot\left\{\eta(\varphi)\left[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^T\right]\right\}\right)\cdot\boldsymbol{u} \\
&\quad + \int_\Omega \nabla\tilde{p}\cdot\boldsymbol{u} - \int_\Omega \mu\nabla\varphi\cdot\boldsymbol{u} - \int_\Omega (\nabla\cdot\boldsymbol{\sigma})\cdot\boldsymbol{u} \\
&= \int_\Omega \frac{1}{2}\frac{\partial|\boldsymbol{u}|^2}{\partial t} + \int_\Omega (\boldsymbol{u}\cdot\nabla)\frac{1}{2}|\boldsymbol{u}|^2 + \int_\Omega \left\{\eta(\varphi)\left[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^T\right]\right\} : \nabla\boldsymbol{u} \\
&\quad - \int_{\partial\Omega}\left(\left\{\eta(\varphi)\left[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^T\right]\right\}\boldsymbol{n}\right)\cdot\boldsymbol{u} + \int_\Omega \nabla\tilde{p}\cdot\boldsymbol{u} - \int_\Omega \mu\nabla\varphi\cdot\boldsymbol{u} - \int_\Omega (\nabla\cdot\boldsymbol{\sigma})\cdot\boldsymbol{u} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{1}{2}|\boldsymbol{u}|^2 - \int_\Omega (\nabla\cdot\boldsymbol{u})\frac{1}{2}|\boldsymbol{u}|^2 + \int_{\partial\Omega}(\boldsymbol{u}\cdot\boldsymbol{n})\frac{1}{2}|\boldsymbol{u}|^2 + \int_\Omega \eta(\varphi)\left[|\nabla\boldsymbol{u}|^2 + \operatorname{tr}\left((\nabla\boldsymbol{u})^2\right)\right] \\
&\quad - \int_{\partial\Omega}\left(\left\{\eta(\varphi)\left[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^T\right]\right\}\boldsymbol{n}\right)\cdot\boldsymbol{u} - \int_\Omega \tilde{p}(\nabla\cdot\boldsymbol{u}) + \int_{\partial\Omega}\tilde{p}\boldsymbol{u}\cdot\boldsymbol{n} \\
&\quad - \int_\Omega \boldsymbol{u}\cdot\nabla\varphi\mu + \int_\Omega \boldsymbol{\sigma} : \nabla\boldsymbol{u} - \int_{\partial\Omega}(\boldsymbol{\sigma}\boldsymbol{n})\cdot\boldsymbol{u}\,.
\end{aligned}$$

Again, assuming suitable boundary conditions for $\boldsymbol{u}$, see Remark 3.13, and using the incompressibility, the last equation can be reduced to

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{1}{2}|\boldsymbol{u}|^2 + 2\int_\Omega \eta(\varphi)|\boldsymbol{D}(\boldsymbol{u})|^2 - \int_\Omega \boldsymbol{u}\cdot\nabla\varphi\mu + \int_\Omega \boldsymbol{\sigma} : \nabla\boldsymbol{u} = 0\,. \tag{3.55}$$

Summing up equations (3.53), (3.54) and (3.55), we obtain the desired energy law (3.52). □

Remark 3.15. *We consider a divergence free velocity $\boldsymbol{u}$. Therefore, we may rewrite the advection and convection terms as follows*

$$\boldsymbol{u}\cdot\nabla\omega = \boldsymbol{u}\cdot\nabla\omega + \nabla\cdot\boldsymbol{u}\omega = \nabla\cdot(\boldsymbol{u}\omega)\,,\quad \omega\in\{\varphi, q, u_i, \sigma_{ij}\},\quad i,j\in\{1,\dots,d\}\,.$$

For example, the full model (3.51) with rewritten advection terms reads

$$\frac{\partial\varphi}{\partial t} + \nabla\cdot(\boldsymbol{u}\varphi) = \nabla\cdot\left\{\varphi(1-\varphi)\frac{1}{\zeta(\varphi)}\left[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\right]\right\}, \tag{3.56a}$$

$$\frac{\partial q}{\partial t} + \nabla\cdot(\boldsymbol{u}q) = -\frac{1}{\tau_B(\varphi)}q - G_B(\varphi)\nabla\cdot\left\{\frac{1}{\zeta(\varphi)}\left[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\right]\right\}, \tag{3.56b}$$

$$\frac{\partial\boldsymbol{\sigma}}{\partial t} + \nabla\cdot(\boldsymbol{u}\otimes\boldsymbol{\sigma}) = (\nabla\boldsymbol{u})\boldsymbol{\sigma} + \boldsymbol{\sigma}(\nabla\boldsymbol{u})^T - \frac{1}{\tau_S(\varphi)}\boldsymbol{\sigma} + G_S(\varphi)\left[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^T\right], \tag{3.56c}$$

$$\frac{\partial\boldsymbol{u}}{\partial t} + \nabla\cdot(\boldsymbol{u}\otimes\boldsymbol{u}) = \nabla\cdot\left\{\eta(\varphi)\left[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^T\right]\right\} - \nabla\hat{p} - \varphi\nabla\mu + \nabla\cdot\boldsymbol{\sigma}\,, \tag{3.56d}$$

$$\nabla\cdot\boldsymbol{u} = 0\,. \tag{3.56e}$$

Note that the φ-coupling term in the Navier-Stokes equation (3.56d) is rewritten as well, such that it corresponds to the rewritten advection term of the Cahn-Hilliard equation (3.56a) in the energy law. Consequently, the pressure term reads $\hat{p} = \tilde{p} - \varphi\mu$ here, since

$$\nabla\hat{p} + \varphi\nabla\mu = \nabla\tilde{p} - \nabla\varphi\mu - \varphi\nabla\mu + \varphi\nabla\mu = \nabla\tilde{p} - \mu\nabla\varphi\,.$$

The introduced modifications of p are still Lagrange multipliers, enforcing the incompressibility. Therefore, we will just write p instead of $\tilde{p}$ and $\hat{p}$ in the following for the sake of simplicity.

3.3.3 The conformation tensor

We should point out that the elastic stress tensor $\boldsymbol{\sigma}$ does not necessarily need to be positive definite. Thus, $\mathrm{tr}(\boldsymbol{\sigma})$ in the energy law (3.52) is not necessarily positive and could therefore interfere with the thermodynamic consistency of the full model. To control this, we introduce the so-called conformation tensor

$$\boldsymbol{C} = \frac{\tau}{\eta_p}\boldsymbol{\sigma} + \boldsymbol{I}\,. \tag{3.57}$$

To avoid conflicts in what follows, we only consider the case

$$\frac{\tau}{\eta_p} = const.,$$

which means that the shear modulus $G = \eta_p/\tau$ is based on the total polymer count of the computational domain as discussed in Remark 3.10 above.

The conformation tensor is a physical quantity related to the elongation of polymer chains. It is well-known to satisfy symmetric positive definiteness and this property is propagated in time as stated by Lemma 3.16 below.

First, let us rewrite the Oldroyd-B model (3.32) to the conformation tensor notation. Rearranging equation (3.57) to $\boldsymbol{\sigma}$ reads

$$\boldsymbol{\sigma} = \frac{\eta_p}{\tau}(\boldsymbol{C} - \boldsymbol{I})\,. \tag{3.58}$$

Then, we substitute (3.58) in the evolution equation for the elastic stress (3.32c), yielding

$$\begin{aligned}
\boldsymbol{0} = \tau &\left(\frac{\eta_p}{\tau}\frac{\partial(\boldsymbol{C}-\boldsymbol{I})}{\partial t} + (\boldsymbol{u}\cdot\nabla)\frac{\eta_p}{\tau}(\boldsymbol{C}-\boldsymbol{I}) - \left[(\nabla\boldsymbol{u})\frac{\eta_p}{\tau}(\boldsymbol{C}-\boldsymbol{I}) + \frac{\eta_p}{\tau}(\boldsymbol{C}-\boldsymbol{I})(\nabla\boldsymbol{u})^T\right]\right) \\
&+ \frac{\eta_p}{\tau}(\boldsymbol{C}-\boldsymbol{I}) - \eta_p\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right] \\
= \eta_p &\frac{\partial \boldsymbol{C}}{\partial t} + \eta_p(\boldsymbol{u}\cdot\nabla)\boldsymbol{C} - \eta_p\left[(\nabla\boldsymbol{u})(\boldsymbol{C}-\boldsymbol{I}) + (\boldsymbol{C}-\boldsymbol{I})(\nabla\boldsymbol{u})^T\right] - \eta_p\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right] \\
&+ \frac{\eta_p}{\tau}(\boldsymbol{C}-\boldsymbol{I}) \\
= \eta_p &\frac{\partial \boldsymbol{C}}{\partial t} + \eta_p(\boldsymbol{u}\cdot\nabla)\boldsymbol{C} - \eta_p\left[(\nabla\boldsymbol{u})\boldsymbol{C} + \boldsymbol{C}(\nabla\boldsymbol{u})^T\right] + \frac{\eta_p}{\tau}(\boldsymbol{C}-\boldsymbol{I})\,.
\end{aligned} \tag{3.59}$$

Further, we substitute (3.58) in the Navier-Stokes part of the Oldroyd-B model (3.32a), which reads

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \nabla\cdot\left\{\eta_s\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right]\right\} - \nabla p + \nabla\cdot\frac{\eta_p}{\tau}(\boldsymbol{C}-\boldsymbol{I})\,. \tag{3.60}$$

Together with the divergence-freeness of $\boldsymbol{u}$ and equation (3.59), divided by the (positive) polymer viscosity η_p, we have the Oldroyd-B model in conformation tensor notation, reading

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \nabla\cdot\left\{\eta_s\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right]\right\} - \nabla p + \frac{\eta_p}{\tau}\nabla\cdot\boldsymbol{C}\,, \tag{3.61a}$$

$$\nabla\cdot\boldsymbol{u} = 0\,, \tag{3.61b}$$

$$\frac{\partial \boldsymbol{C}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{C} = (\nabla\boldsymbol{u})\boldsymbol{C} + \boldsymbol{C}(\nabla\boldsymbol{u})^T - \frac{1}{\tau}(\boldsymbol{C}-\boldsymbol{I})\,. \tag{3.61c}$$

Lemma 3.16. *Let $\boldsymbol{C}$ be a smooth solution of system (3.61). Then, if the initial condition $\boldsymbol{C}|_{t=0}$ is symmetric positive definite everywhere in Ω, the solution $\boldsymbol{C}$ remains so at all times $t \in (0,\infty)$ and for all $\boldsymbol{x} \in \Omega$.*

Proof. See *Hulsen* [47], where sufficient conditions for a rather general model are derived. □

Note that the following Lemma and its proof are based on [45] by *Hu and Lelièvre*, and that this Lemma is essential for the thermodynamic consistency of the full model (3.51).

Lemma 3.17. *Let us assume that* $\det \boldsymbol{C}|_{t=0} \geq 1$. *Then, we have* $\det \boldsymbol{C} \geq 1$ *for all* $t \geq 0$, *which yields* $\mathrm{tr}\boldsymbol{\sigma} \geq 0$ *for all* $t \geq 0$.

Proof. Taking the double dot product of the evolution equation for the conformation tensor (3.61c) and $\boldsymbol{C}^{-1}$ reads

$$\begin{aligned}
0 &= \left(\frac{\partial \boldsymbol{C}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{C} - \left[(\nabla\boldsymbol{u})\boldsymbol{C} + \boldsymbol{C}(\nabla\boldsymbol{u})^T\right] + \frac{1}{\tau}(\boldsymbol{C}-\boldsymbol{I}) \right) : \boldsymbol{C}^{-1} \\
&= \mathrm{tr}\left(\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial t} \right) + \mathrm{tr}(\boldsymbol{C}^{-1}(\boldsymbol{u}\cdot\nabla)\boldsymbol{C}) - \mathrm{tr}\Big(\boldsymbol{C}^{-1}(\nabla\boldsymbol{u})\boldsymbol{C} + \boldsymbol{C}^{-1}\boldsymbol{C}(\nabla\boldsymbol{u})^T \Big) \\
&\quad + \frac{1}{\tau}\mathrm{tr}(\boldsymbol{C}^{-1}\boldsymbol{C} - \boldsymbol{C}^{-1}\boldsymbol{I}) \\
&= \mathrm{tr}\left(\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial t} \right) + \mathrm{tr}(\boldsymbol{C}^{-1}(\boldsymbol{u}\cdot\nabla)\boldsymbol{C}) - \mathrm{tr}\Big(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T \Big) + \frac{1}{\tau}\mathrm{tr}(\boldsymbol{I} - \boldsymbol{C}^{-1}) .
\end{aligned}$$

Using Jacobi's formula, see Lemma 2.1, we can rewrite the above equation to

$$\begin{aligned}
0 &= \frac{\partial}{\partial t}\ln\det\boldsymbol{C} + (\boldsymbol{u}\cdot\nabla)\ln\det\boldsymbol{C} - \mathrm{tr}\Big(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T \Big) + \frac{1}{\tau}\mathrm{tr}(\boldsymbol{I} - \boldsymbol{C}^{-1}) \\
&= \frac{\partial}{\partial t}\ln\det\boldsymbol{C} + (\boldsymbol{u}\cdot\nabla)\ln\det\boldsymbol{C} - 2\mathrm{tr}(\nabla\boldsymbol{u}) + \frac{1}{\tau}\mathrm{tr}(\boldsymbol{I} - \boldsymbol{C}^{-1}) \\
&= \frac{\partial}{\partial t}\ln\det\boldsymbol{C} + (\boldsymbol{u}\cdot\nabla)\ln\det\boldsymbol{C} + \frac{1}{\tau}\mathrm{tr}(\boldsymbol{I} - \boldsymbol{C}^{-1}) ,
\end{aligned}$$

where the last equality holds true, since $\mathrm{tr}(\nabla\boldsymbol{u}) = \nabla\cdot\boldsymbol{u} = 0$. Since any symmetric positive definite matrix $\boldsymbol{M} \in \mathbb{R}^{d\times d}$ satisfies

$$(\det \boldsymbol{M})^{\frac{1}{d}} \leq \frac{1}{d}\mathrm{tr}\boldsymbol{M} , \tag{3.62}$$

we have

$$\frac{\partial}{\partial t}\ln\det\boldsymbol{C} + (\boldsymbol{u}\cdot\nabla)\ln\det\boldsymbol{C} = \frac{1}{\tau}(\mathrm{tr}\boldsymbol{C}^{-1} - d) \geq \frac{d}{\tau}\left((\det\boldsymbol{C})^{-\frac{1}{d}} - 1 \right) ,$$

which we can rewrite to

$$\begin{aligned}
& d\left(\frac{\partial}{\partial t}\ln(\det\boldsymbol{C})^{\frac{1}{d}} + (\boldsymbol{u}\cdot\nabla)\ln(\det\boldsymbol{C})^{\frac{1}{d}} \right) \geq \frac{d}{\tau}\left((\det\boldsymbol{C})^{-\frac{1}{d}} - 1 \right) \\
\Leftrightarrow\ & \frac{d}{(\det\boldsymbol{C})^{\frac{1}{d}}}\left(\frac{\partial}{\partial t}(\det\boldsymbol{C})^{\frac{1}{d}} + (\boldsymbol{u}\cdot\nabla)(\det\boldsymbol{C})^{\frac{1}{d}} \right) \geq \frac{d}{\tau}\left((\det\boldsymbol{C})^{-\frac{1}{d}} - 1 \right) \\
\Leftrightarrow\ & \tau\left(\frac{\partial}{\partial t}(\det\boldsymbol{C})^{\frac{1}{d}} + (\boldsymbol{u}\cdot\nabla)(\det\boldsymbol{C})^{\frac{1}{d}} \right) \geq 1 - (\det\boldsymbol{C})^{\frac{1}{d}} .
\end{aligned} \tag{3.63}$$

Now, let us assume $(\det \boldsymbol{C}|_{t=0})^{\frac{1}{d}} \geq 1$. If $(\det \boldsymbol{C})^{\frac{1}{d}}$ does not remain greater than 1, let us consider the first time t_0, where $(\det \boldsymbol{C}|_{t=t_0})^{\frac{1}{d}} = 1$ (in case $\det \boldsymbol{C}|_{t=0} = 1$, obviously $t_0 = 0$). Inserted in equation (3.63), we have

$$\frac{D}{Dt}(\det \boldsymbol{C}|_{t=t_0})^{\frac{1}{d}} = \frac{\partial}{\partial t}(\det \boldsymbol{C}|_{t=t_0})^{\frac{1}{d}} + (\boldsymbol{u} \cdot \nabla)(\det \boldsymbol{C}|_{t=t_0})^{\frac{1}{d}} \geq 0\,.$$

Therefore, $(\det \boldsymbol{C})^{\frac{1}{d}}$ can not decrease any further. This yields $(\det \boldsymbol{C})^{\frac{1}{d}} \geq 1$ if $(\det \boldsymbol{C}|_{t=0})^{\frac{1}{d}} \geq 1$ and thus $\det \boldsymbol{C} \geq 1$ if $\det \boldsymbol{C}|_{t=0} \geq 1$.

Inequality (3.62) with $(\det \boldsymbol{C})^{\frac{1}{d}} \geq 1$ yields $\frac{1}{d}\mathrm{tr}\boldsymbol{C} \geq 1$, i.e., $\mathrm{tr}\boldsymbol{C} \geq d$. And since $\boldsymbol{\sigma} = \frac{\eta_p}{\tau}(\boldsymbol{C} - \boldsymbol{I})$,

$$\mathrm{tr}\boldsymbol{\sigma} = \frac{\eta_p}{\tau}(\mathrm{tr}\boldsymbol{C} - \mathrm{tr}\boldsymbol{I}) \geq \frac{\eta_p}{\tau}(d - d) = 0\,.$$

Thus, $\mathrm{tr}\boldsymbol{\sigma} \geq 0$. □

Rewriting the full model (3.56) to conformation tensor notation analogously to above yields

$$\frac{\partial \varphi}{\partial t} + \boldsymbol{u} \cdot \nabla \varphi = \nabla \cdot \left\{ \varphi(1-\varphi)\frac{1}{\zeta(\varphi)}\left[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\right]\right\}, \tag{3.64a}$$

$$\frac{\partial q}{\partial t} + \boldsymbol{u} \cdot \nabla q = -\frac{1}{\tau_B(\varphi)}q - G_B(\varphi)\nabla \cdot \left\{ \frac{1}{\zeta(\varphi)}\left[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\right]\right\}, \tag{3.64b}$$

$$\frac{\partial \boldsymbol{C}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{C} = (\nabla \boldsymbol{u})\boldsymbol{C} + \boldsymbol{C}(\nabla \boldsymbol{u})^T - \frac{1}{\tau_S(\varphi)}(\boldsymbol{C} - \boldsymbol{I})\,, \tag{3.64c}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = \nabla \cdot \left\{ \eta(\varphi)\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right]\right\} - \nabla p + \mu \nabla \varphi + \nabla \cdot [G_S(\boldsymbol{C} - \boldsymbol{I})]\,, \tag{3.64d}$$

$$\nabla \cdot \boldsymbol{u} = 0\,. \tag{3.64e}$$

The elastic energy in conformation tensor notation is according to *Hu and Lelièvre* [45] given by

$$E_{el}(\boldsymbol{C}) = \int_\Omega \frac{G_S}{2}\mathrm{tr}(\boldsymbol{C} - \ln \boldsymbol{C} - \boldsymbol{I})\,,$$

where $\mathrm{tr}(\boldsymbol{C} - \ln \boldsymbol{C} - \boldsymbol{I}) \geq 0$ due to estimate (2.5b). Thus, the total energy of the full model (3.64) in conformation tensor notation reads

$$\begin{aligned} E_{total}(\varphi, q, \boldsymbol{C}, \boldsymbol{u}) &= E_{mix}(\varphi) + E_{bulk}(q) + E_{el}(\boldsymbol{C}) + E_{kin}(\boldsymbol{u}) \\ &= \int_\Omega \left(\frac{\lambda}{2}|\nabla\varphi|^2 + F(\varphi)\right) + \int_\Omega \frac{1}{2}q^2 + \int_\Omega \frac{G_S}{2}\mathrm{tr}(\boldsymbol{C} - \ln \boldsymbol{C} - \boldsymbol{I}) + \int_\Omega \frac{1}{2}|\boldsymbol{u}|^2\,. \end{aligned}$$

Theorem 3.18. *Assuming suitable boundary conditions, see Remark 3.13, model (3.64) satisfies the following energy law*

$$\begin{aligned} \frac{\mathrm{d}E_{total}(\varphi, q, \boldsymbol{C}, \boldsymbol{u})}{\mathrm{d}t} = &-\int_\Omega \frac{1}{\tau_B(\varphi)}q^2 - \int_\Omega \frac{1}{\zeta(\varphi)}\left|\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\right|^2 \\ &- \int_\Omega \frac{G_S}{2\tau_S(\varphi)}\mathrm{tr}(\boldsymbol{C} + \boldsymbol{C}^{-1} - 2\boldsymbol{I}) - 2\int_\Omega \eta(\varphi)|\boldsymbol{D}(\boldsymbol{u})|^2 \leq 0\,, \end{aligned} \tag{3.65}$$

since $\mathrm{tr}(\boldsymbol{C} + \boldsymbol{C}^{-1} - 2\boldsymbol{I}) \geq 0$ *due to estimate* (2.5c).

Proof. Analogously to the derivation of energy law (3.52), equation (3.64a) is multiplied by μ, (3.64b) by q and (3.64d) with the dot product by $\boldsymbol{u}$, and all three are integrated and summed up. Assuming suitable boundary conditions, see Remark 3.13, and using the incompressibility, we obtain

$$\begin{aligned}&\frac{\mathrm{d}E_{mix}(\varphi)}{\mathrm{d}t} + \frac{\mathrm{d}E_{bulk}(q)}{\mathrm{d}t} + \frac{\mathrm{d}E_{kin}(\boldsymbol{u})}{\mathrm{d}t} + \int_\Omega \frac{1}{\tau_B(\varphi)} q^2 \\ &+ \int_\Omega \frac{1}{\zeta(\varphi)} \left| \varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q) \right|^2 + 2\int_\Omega \eta(\varphi)|\boldsymbol{D}(\boldsymbol{u})|^2 - \int_\Omega \big(\nabla\cdot[G_S(\boldsymbol{C}-\boldsymbol{I})]\big)\cdot\boldsymbol{u} = 0\,,\end{aligned} \tag{3.66}$$

where the last integral satisfies

$$\begin{aligned}-\int_\Omega \big(\nabla\cdot[G_S(\boldsymbol{C}-\boldsymbol{I})]\big)\cdot\boldsymbol{u} &= -\int_\Omega \big(\nabla\cdot(G_S\boldsymbol{C})\big)\cdot\boldsymbol{u} \\ &= \int_\Omega G_S\boldsymbol{C} : (\nabla\boldsymbol{u}) - \int_{\partial\Omega} G_S(\boldsymbol{C}\boldsymbol{n})\cdot\boldsymbol{u} \\ &= \int_\Omega G_S\boldsymbol{C} : (\nabla\boldsymbol{u})\,.\end{aligned}$$

Here, the boundary integral vanishes in the last equality due to the assumed boundary conditions for $\boldsymbol{u}$. Then, taking the double dot product of (3.64c) and $\frac{G_S}{2}\boldsymbol{I}$, and integrating over Ω, we obtain

$$\begin{aligned}0 &= \int_\Omega \frac{\partial\boldsymbol{C}}{\partial t} : \frac{G_S}{2}\boldsymbol{I} + \int_\Omega (\boldsymbol{u}\cdot\nabla)\boldsymbol{C} : \frac{G_S}{2}\boldsymbol{I} - \int_\Omega \Big[(\nabla\boldsymbol{u})\boldsymbol{C} + \boldsymbol{C}(\nabla\boldsymbol{u})^T\Big] : \frac{G_S}{2}\boldsymbol{I} \\ &\quad + \int_\Omega \frac{1}{\tau_S(\varphi)}(\boldsymbol{C}-\boldsymbol{I}) : \frac{G_S}{2}\boldsymbol{I} \\ &= \int_\Omega \frac{G_S}{2}\mathrm{tr}\left(\frac{\partial\boldsymbol{C}}{\partial t}\right) + \int_\Omega \frac{G_S}{2}(\boldsymbol{u}\cdot\nabla)\mathrm{tr}\boldsymbol{C} - \int_\Omega \frac{G_S}{2}\Big[\nabla\boldsymbol{u} : \boldsymbol{C}^T + \boldsymbol{C} : \nabla\boldsymbol{u}\Big] \\ &\quad + \int_\Omega \frac{G_S}{2\tau_S(\varphi)}\mathrm{tr}(\boldsymbol{C}-\boldsymbol{I}) \\ &= \int_\Omega \frac{G_S}{2}\frac{\partial\mathrm{tr}\boldsymbol{C}}{\partial t} - \int_\Omega \frac{G_S}{2}(\nabla\cdot\boldsymbol{u})\mathrm{tr}\boldsymbol{C} + \int_{\partial\Omega} \frac{G_S}{2}(\boldsymbol{u}\cdot\boldsymbol{n})\mathrm{tr}\boldsymbol{C} - \int_\Omega G_S\boldsymbol{C} : \nabla\boldsymbol{u} \\ &\quad + \int_\Omega \frac{G_S}{2\tau_S(\varphi)}\mathrm{tr}(\boldsymbol{C}-\boldsymbol{I}) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{G_S}{2}\mathrm{tr}\boldsymbol{C} - \int_\Omega G_S\boldsymbol{C} : \nabla\boldsymbol{u} + \int_\Omega \frac{G_S}{2\tau_S(\varphi)}\mathrm{tr}(\boldsymbol{C}-\boldsymbol{I})\,.\end{aligned} \tag{3.67}$$

Additionally, we take the double dot product of (3.64c) and $-\frac{G_S}{2}\boldsymbol{C}^{-1}$ and integrate, yielding

$$\begin{aligned}0 &= -\int_\Omega \frac{\partial\boldsymbol{C}}{\partial t} : \frac{G_S}{2}\boldsymbol{C}^{-1} - \int_\Omega (\boldsymbol{u}\cdot\nabla)\boldsymbol{C} : \frac{G_S}{2}\boldsymbol{C}^{-1} + \int_\Omega \Big[(\nabla\boldsymbol{u})\boldsymbol{C} + \boldsymbol{C}(\nabla\boldsymbol{u})^T\Big] : \frac{G_S}{2}\boldsymbol{C}^{-1} \\ &\quad - \int_\Omega \frac{1}{\tau_S(\varphi)}(\boldsymbol{C}-\boldsymbol{I}) : \frac{G_S}{2}\boldsymbol{C}^{-1} \\ &= -\int_\Omega \frac{G_S}{2}\mathrm{tr}\left(\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{C}}{\partial t}\right) - \int_\Omega \frac{G_S}{2}\mathrm{tr}\left(\boldsymbol{C}^{-1}(\boldsymbol{u}\cdot\nabla)\boldsymbol{C}\right) + \int_\Omega \frac{G_S}{2}\mathrm{tr}\Big(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\Big) \\ &\quad - \int_\Omega \frac{G_S}{2\tau_S(\varphi)}\mathrm{tr}(\boldsymbol{I}-\boldsymbol{C}^{-1})\,.\end{aligned}$$

Rewriting the first two terms using Lemma 2.1, we get

$$\begin{aligned}
0 &= -\int_\Omega \frac{G_S}{2}\frac{\partial}{\partial t}\mathrm{tr}\ln \boldsymbol{C} - \int_\Omega \frac{G_S}{2}(\boldsymbol{u}\cdot\nabla)\mathrm{tr}\ln \boldsymbol{C} + \int_\Omega G_S \mathrm{tr}(\nabla \boldsymbol{u}) + \int_\Omega \frac{G_S}{2\tau_S(\varphi)}\mathrm{tr}(\boldsymbol{C}^{-1}-\boldsymbol{I}) \\
&= -\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{G_S}{2}\mathrm{tr}\ln \boldsymbol{C} + \int_\Omega \frac{G_S}{2}(\nabla\cdot\boldsymbol{u})\mathrm{tr}\ln \boldsymbol{C} - \int_{\partial\Omega}\frac{G_S}{2}(\boldsymbol{u}\cdot\boldsymbol{n})\mathrm{tr}\ln \boldsymbol{C} \\
&\quad + \int_\Omega G_S \nabla\cdot\boldsymbol{u} + \int_\Omega \frac{G_S}{2\tau_S(\varphi)}\mathrm{tr}(\boldsymbol{C}^{-1}-\boldsymbol{I}) \\
&= -\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{G_S}{2}\mathrm{tr}\ln \boldsymbol{C} + \int_\Omega \frac{G_S}{2\tau_S(\varphi)}\mathrm{tr}(\boldsymbol{C}^{-1}-\boldsymbol{I})\,. \qquad (3.68)
\end{aligned}$$

Finally, summing up equations (3.66), (3.67) and (3.68) yields energy law (3.65). □

3.3.4 Noteworthy model extensions

The above introduced phase-field models for two-phase flows are restricted to the case of matched densities of the two fluids, since they consider the density ρ of the mixture to be constant in $(\boldsymbol{x}, t)$. However, there are various approaches in the literature considering different densities for the two fluids. *Lowengrub and Truskinovsky* derive in [55] one of, if not the first thermodynamically consistent Cahn-Hilliard-Navier-Stokes model with different densities, but the resulting model is quasi-incompressible, i.e., the averaged velocity field is no longer divergence free. In [9], *Boyer* derives an incompressible Cahn-Hilliard-Navier-Stokes model with different densities and viscosities and even proposes a numerical scheme for it, but leaves out respective energy laws. *Shen and Yang* do take care of the thermodynamic consistency of their proposed model and the matching numerical scheme in [64], but the convection term of their model is not frame invariant. The model by *Abels, Garcke and Grün* derived in [4] has the same flaw. However, they also derive a frame indifferent, thermodynamically consistent two-phase flow model with different densities in [3], reading

$$\frac{\partial \varphi}{\partial t} + \boldsymbol{u}\cdot\nabla\varphi = \nabla\cdot\big(M(\varphi)\nabla\mu\big)\,, \qquad (3.69a)$$

$$\rho\frac{\partial \boldsymbol{u}}{\partial t} + \big((\rho\boldsymbol{u}+\boldsymbol{J})\cdot\nabla\big)\boldsymbol{u} = -\nabla p + \nabla\cdot\left\{\eta(\varphi)\Big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\Big]\right\} - \nabla\cdot(\lambda\nabla\varphi\otimes\nabla\varphi)\,, \qquad (3.69b)$$

$$\nabla\cdot\boldsymbol{u} = 0\,, \qquad (3.69c)$$

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{u}+\boldsymbol{J}) = 0\,, \qquad (3.69d)$$

where

$$\boldsymbol{J} = -\frac{\rho_1-\rho_2}{2}M(\varphi)\nabla\mu$$

is specified as a relative mass flux related to the diffusion of the two components, which have the different densities ρ_1 and ρ_2. Numerical schemes and simulations for model (3.69) are presented by *Guillén-González and Tierra* in [40]. Another thermodynamically consistent model having a variable density is derived by *Liu, Shen and Yang* in [54], using the energetic variational approach. They also propose numerical schemes and validate them and the model by experiments, using large density and viscosity ratios.

For our own model, let us use an idea originally mentioned for an Allen-Cahn/Navier-Stokes system by *Giga, Kirshtein and Liu* in [34]. It is sufficient to rewrite the material derivative of the Navier-Stokes equation (3.24c) as follows to incorporate different densities

$$\frac{\partial \varphi}{\partial t} + \boldsymbol{u} \cdot \nabla \varphi = \nabla \cdot \left(M(\varphi) \nabla \mu \right), \tag{3.70a}$$

$$\frac{1}{2}\left[\rho(\varphi)\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\right) + \frac{\partial \rho(\varphi)\boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)(\rho(\varphi)\boldsymbol{u})\right] = \nabla \cdot \mathbf{T}, \tag{3.70b}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{3.70c}$$

where

$$\mathbf{T} = -p\boldsymbol{I} + \eta(\varphi)\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right] - \lambda \nabla \varphi \otimes \nabla \varphi,$$

i.e., the right hand sides of the Navier-Stokes equations (3.70b) and (3.24c) are identical. Note that both variable density models (3.69) and (3.70) have the same free energy (3.25) as the constant density model (3.24) and satisfy the same energy law (3.28). Further, note that the full model for viscoelastic phase separation (3.47) should be modifiable like model (3.70) in order to incorporate different densities, which is left for future work. The derivation of appropriate compressible models for two-phase polymer-solvent mixtures is also part of ongoing research in the SFB-TRR 146 subproject C3. The compressible model naturally contains a variable density, but obviously complicates the analysis and the numerics considerably.

A second model extension concerns the Oldroyd-B model (3.32). As mentioned above, it is derived from bead and spring dumbbells. In the derivation, one assumes similar velocities for both beads of one dumbbell. However, in order to allow for a slight variation between these velocities, one can add noise in terms of diffusion to the evolution equation of the stress. The resulting diffusive version of the stress evolution equation (3.35) reads

$$\frac{\partial \boldsymbol{\sigma}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{\sigma} = (\nabla \boldsymbol{u})\boldsymbol{\sigma} + \boldsymbol{\sigma}(\nabla \boldsymbol{u})^T - \frac{1}{\tau_S(\varphi)}\boldsymbol{\sigma} + G_S(\varphi)\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right] + \varepsilon_S \Delta \boldsymbol{\sigma}, \tag{3.71}$$

where the diffusion term $\varepsilon_S \Delta \boldsymbol{\sigma}$ is related to the center-of-mass diffusion of the polymer chain, with $\varepsilon_S > 0$, but small. We can regularize the bulk stress equation (3.47b) as well, reading

$$\frac{\partial q}{\partial t} + \boldsymbol{u}\cdot\nabla q = -\frac{1}{\tau_B(\varphi)}q - G_B(\varphi)\nabla \cdot \left\{ \frac{1}{\zeta(\varphi)}\left[\varphi(1-\varphi)\nabla\mu - \nabla(G_B(\varphi)q)\right]\right\} + \varepsilon_B \Delta q, \tag{3.72}$$

where also $\varepsilon_B > 0$, but small. These generalizations lead to a fully parabolic system, instead of a parabolic-hyperbolic one, which enables the exploration of analytical properties like the existence.

3.4 Existence and uniqueness

In order for a differential equation model to be mathematically well-posed, it is crucial that a solution to the problem exists and that this solution is unique and stable. The required analysis is not a part of this thesis, but due to its importance, we will give some references. For the sake of clarity, we list the known existence and uniqueness results for two and three

space dimensions in Tables 3.1 and 3.2. The abbreviations used in these tables are explained in the following. Note that the existence results listed in Table 3.1 are in general for weak solutions. Here, we distinguish between the degenerate (deg) and the non-degenerate Cahn-Hilliard model. The degenerate model is named after its degenerate mobility, i.e., equation (3.4) with $n > 0$. It uses a logarithmic potential like the Flory-Huggins potential (3.11). The non-degenerate model has a bounded, positive mobility, i.e., equation (3.4) with $n = 0$, and uses a polynomial potential like the Ginzburg-Landau potential (3.9). Since the Navier-Stokes equations and the Oldroyd-B model in and of itself are independent of the potential and the mobility, the existence results are the same with or without "deg". Regarding the uniqueness results listed in Table 3.2, we separate into the uniqueness of weak solutions and the weak-strong (w-s) uniqueness, with one exception being for strong solutions (str). To the best of our knowledge, there are no uniqueness results for the degenerate Cahn-Hilliard model, therefore, this case is left out. Some uniqueness results only hold locally (loc), i.e., for $t \in [0, T)$, with T depending on the initial conditions and usually small, while others, having different additional assumptions, are conditional (con). A bracketed reference means that the results are originally for a greater system, but directly transferable to this model. A checkmark without a reference indicates that we are aware of existing proofs that are not published yet.

Elliott and Garcke [26] investigate both variants of the Cahn-Hilliard model (3.7) (CH), while *Brunk, Egger, Habrich and Lukáčová-Medviďová* [11] add weak-strong uniqueness results for the non-degenerate model. The incompressible Navier-Stokes equations (3.22) (NS) are analyzed, e.g., by *Boyer and Fabrie* [59] and with respect to their weak-strong uniqueness by *Fefferman, Robinson and Rodrigo Diez* [30]. Note that the existence of smooth solutions in three space dimensions is an open millenium problem, see the official problem description [29]. In [19], *Constantin and Kliegl* discuss the diffusive Oldroyd-B model (3.71) (OB), since the non-diffusive one is presumably only locally unique. Note that their uniqueness results are for strong solutions. The Cahn-Hilliard-Navier-Stokes model (3.24) (CH-NS) is, e.g., investigated by *Abels, Depner, Garcke, Boyer and Fabrie* [2, 8, 59]. A viscoelastic two-fluid model similar to (3.34), (3.35) (CH-OB) is covered by *Chupin* [18]. And finally, in [12, 13, 14], *Brunk and Lukáčová-Medviďová* investigate a viscoelastic Peterlin two-fluid model (full model), which is similar to the full model in conformation tensor notation (3.64), but with diffusion and further generalizations to the Oldroyd-B model, yielding the diffusive Peterlin model. For details on the latter, see, e.g., *Mizerová* [61]. Note that the results from *Brunk and Lukáčová-Medviďová* are directly transferable to the Cahn-Hilliard-Navier-Stokes model (3.24) and to the simplified model (without hydrodynamics) (3.48) (simp model). Further, they should be transferable to the full model (3.51) as well, if we add diffusion to the shear and bulk stress evolution equations as done in equations (3.71) and (3.72).

Table 3.1: Existence of weak solutions

	CH	NS	OB	CH-NS	CH-OB	full model	simp model
2D	[26]	[59]	[19]	[8]	[18]	[13]	([13])
2D deg	[26]	[59]	[19]	[2, 59]		[14]	([14])
3D	[26]	[59]		[59]	[18]	✓	(✓)
3D deg	[26]	[59]		[2, 59]		✓	(✓)

Table 3.2: Uniqueness of weak(-strong) solutions

	CH	NS	OB	CH-NS	CH-OB	full model	simp model
2D	[26]	[59]	[19] str	[8]	[18]	✓	(✓)
2D w-s	[11]	[30]		([12])		[12]	([12])
3D		[59] loc		[59] loc	[18] loc		
3D w-s	[11]	[30]		([12] con)		[12] con	([12] con)

4 Numerical schemes

In this chapter, we propose numerical schemes for several phase-field models introduced in Chapter 3, and analyze their properties. For the numerical time integration of these models, we primarily use linear low-order Runge-Kutta methods, and for the space discretization finite difference and finite volume methods.

We start with semi-discretizations in time, since they contain the main challenges regarding the thermodynamic consistency. These challenges will be outlined in Section 4.1. Section 4.2 deals exclusively with suitable discretizations for the derivative of the double-well potential. Then, in Section 4.3, we introduce our first numerical schemes, focusing on the Cahn-Hilliard equation. In Section 4.4, we add the Navier-Stokes equations and introduce a first splitting scheme. In Section 4.5, we propose schemes for the simplified model (3.48) and in Sections 4.6 and 4.7 for the full model for viscoelastic phase separation in its original form (3.51) and in conformation tensor notation (3.64), respectively.

Thereafter, we introduce our spatial discretization methods in Section 4.8, and also some full discretizations. Here, we successively derive finite differences for one, two and three space dimensions, such that we can fully discretize the Cahn-Hilliard equation and the simplified model afterwards. Then, we derive finite volume methods, again for one to three space dimensions, to finally fully discretize the Cahn-Hilliard-Navier-Stokes model and subsequently the full model for viscoelastic phase separation.

We conclude the chapter with a brief overview of time step restrictions and adaptive time-stepping in Section 4.9.

4.1 Challenges of appropriate time discretizations

Fundamental for numerical schemes is the conservation of as many properties of the continuum model as possible. The major advantage of the full two-fluid model for viscoelastic phase separation (3.47) from *Zhou et. al.* [74] is that it is thermodynamically consistent, since its free energy is non-increasing in time, while also being able to reproduce almost all the essential features of viscoelastic phase separation. Thus, the primary focus of appropriate numerical schemes is the conservation of the thermodynamic consistency.

Definition 4.1. *(energy-stability)*
We call a numerical scheme energy-stable, if it conserves the thermodynamic consistency of

the original partial differential equation (system), which it solves approximately, for suitable boundary conditions. Thus, if for all $n \in \mathbb{N}$

$$E_{total}(\boldsymbol{\omega}^{n+1}) \leq E_{total}(\boldsymbol{\omega}^{n}),$$

where $\boldsymbol{\omega}^{n+1} \approx \boldsymbol{\omega}(t_{n+1})$ *and* $\boldsymbol{\omega}^{n} \approx \boldsymbol{\omega}(t_{n})$ *are the numerical solution vectors at times* t_{n+1} *and* t_n*, with* $t_{n+1} \geq t_n$*.*

Since the Cahn-Hilliard equation is derived from the law of conservation of mass, see Section 3.1, it is obviously mass-conservative. Therefore, it is desirable that our numerical schemes conserve mass as well.

Definition 4.2. *(mass conservation)*
A numerical scheme for a Cahn-Hilliard model is mass-conservative, if for suitable boundary conditions and all $n \in \mathbb{N}$

$$\int_{\Omega} \varphi^{n} = \int_{\Omega} \varphi^{0},$$

where $\varphi^{n} \approx \varphi(t_n)$ *is the numerical solution of the Cahn-Hilliard equation at time* t_n*.*

Even though the Navier-Stokes equations are derived from the law of conservation of momentum, momentum is only conserved for periodic boundary conditions. Additionally, conservation of momentum is in general very difficult to preserve numerically. Therefore, we neglect it when constructing our numerical schemes.

Another crucial property of numerical schemes is efficiency. The introduced continuum models are highly nonlinear, therefore, their fully implicit discretizations are nonlinear as well. Solving nonlinear systems of equations necessitates an iteration in each time step with the Newton method or a fixed-point method, which can substantially increase the computational cost. Thus, we focus on time discretizations which linearize the continuum models, such that we obtain fully linearized schemes. To achieve this goal, we use suitable mixed low-order IMEX (implicit-explicit) Runge-Kutta methods, i.e., the implicit and explicit Euler method and the Crank-Nicolson method. Further, we apply a simple multistep method in order to construct numerical schemes with a truncation error of second order in time, which is the two-step Adams-Bashforth method. Furthermore, we use Taylor approximations and stabilization terms to treat the derivative of the double-well potential, since its approximation is anything but trivial. We also introduce some more stabilization terms in order to be able to decouple some calculations, which significantly reduces the size of the systems of linear equations, which have to be solved, and therefore leads to a reduction of computing time.

Remark 4.3. *Summarized, we aim for the following three major properties when constructing our numerical schemes:*

- *linearity of the discretized systems,*
- *energy-stability, and*
- *mass conservation.*

Further, in order to be efficient, the possibility to use a comparatively large time step size is always advantageous and the decoupled calculation of coupled equations can be favorable. On the other hand, in order to have a good approximate solution, a small truncation error is preferable, as well as keeping deviations from the continuous energy law small in its discrete version, or if possible, having none at all.

We will summarize the fulfillment of these properties in the form of a table at the end of subsections, where it is relevant.

Zhou et al. [74] state to use the explicit Euler method, a purely explicit solver, to discretize the full model (3.47) and the simplified model (3.48) in time.

Let us consider the Cahn-Hilliard equation

$$\frac{\partial \varphi}{\partial t} = \nabla \cdot (M(\varphi) \nabla(-\lambda \Delta \varphi + f(\varphi)) , \tag{4.1}$$

initial and boundary conditions from Remark 3.2, and a uniform partition of the time interval $[0, T]$ into N time steps with constant time step size Δt. Further, let φ^n be the numerical approximation of the solution $\varphi(t_n)$, where $t_n = n\Delta t, n \in \{1, ..., N\}$. Then, the explicit Euler method for the Cahn-Hilliard equation (4.1) reads

$$\begin{aligned} & \frac{\varphi^{n+1} - \varphi^n}{\Delta t} = \nabla \cdot (M(\varphi^n) \nabla(-\lambda \Delta \varphi^n + f(\varphi^n)) \\ \iff \quad & \varphi^{n+1} = \varphi^n + \Delta t \nabla \cdot (M(\varphi^n) \nabla(-\lambda \Delta \varphi^n + f(\varphi^n)) . \end{aligned} \tag{4.2}$$

Thus, given initial data $\varphi^0 = \varphi(t_0)$, one can iteratively calculate $\varphi^n, n = 1, ..., N$, without the necessity to solve any linear equation system.

We will show in Theorem 4.4 that the explicit Euler method for the full model (3.51) by *Zhou et al.* [74] is in general not energy-stable. Thus, it does not necessarily conserve the thermodynamic consistency of the continuum model. This result is transferable to any kind of Cahn-Hilliard model, which will be obvious from the theorem and its proof.

Another letdown of purely explicit schemes is their time step restriction. In order to be numerically stable, the time step size Δt has to correlate to the spatial step size h to the power of the degree of the highest explicitly discretized spatial derivative of the underlying partial differential equation (system). Thus, if the fourth derivative in the Cahn-Hilliard equation is calculated explicitly, the time step size has to be as small as

$$\Delta t \in \mathcal{O}\left(h^4\right) .$$

We will discuss time-stepping in more detail in Section 4.9.

The explicit Euler method for the full model (3.51) reads

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^n \cdot \nabla \varphi^n \tag{4.3a}$$
$$- \nabla \cdot \left\{ \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)} \left[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\, q^n) \right] \right\} = 0\,,$$
$$\frac{q^{n+1}-q^n}{\Delta t} + \boldsymbol{u}^n \cdot \nabla q^n + \frac{1}{\tau_B(\varphi^n)} q^n \tag{4.3b}$$
$$+ G_B(\varphi^n) \nabla \cdot \left\{ \frac{1}{\zeta(\varphi^n)} \left[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\, q^n) \right] \right\} = 0\,,$$
$$\frac{\boldsymbol{\sigma}^{n+1}-\boldsymbol{\sigma}^n}{\Delta t} + (\boldsymbol{u}^n \cdot \nabla)\boldsymbol{\sigma}^n - (\nabla \boldsymbol{u}^n)\cdot \boldsymbol{\sigma}^n - \boldsymbol{\sigma}^n \cdot (\nabla \boldsymbol{u}^n)^T \tag{4.3c}$$
$$+ \frac{1}{\tau_S(\varphi^n)} \boldsymbol{\sigma}^n - G_S(\varphi^n) \left[\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T \right] = \boldsymbol{0}\,,$$
$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^n - \nabla \cdot \left\{ \eta(\varphi^n) \left[\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T \right] \right\} \tag{4.3d}$$
$$+ \nabla p^n - \mu^n \nabla \varphi^n - \nabla \cdot \boldsymbol{\sigma}^n = \boldsymbol{0}\,,$$
$$\nabla \cdot \boldsymbol{u}^n = 0\,, \tag{4.3e}$$

where

$$\mu^n = -\lambda \Delta \varphi^n + f(\varphi^n)\,.$$

Note that the semi-discrete free energy is given by the same equation as the continuous free energy (3.46). Thus, at time $t_n = n\Delta t$, the semi-discrete free energy reads

$$\begin{aligned} E_{total}(\varphi^n, \boldsymbol{u}^n, q^n, \boldsymbol{\sigma}^n) &= E_{mix}(\varphi^n) + E_{kin}(\boldsymbol{u}^n) + E_{bulk}(q^n) + E_{el}(\boldsymbol{\sigma}^n) \\ &= \int_\Omega \left(\frac{\lambda}{2} |\nabla \varphi^n|^2 + F(\varphi^n) \right) + \int_\Omega \frac{1}{2} |\boldsymbol{u}^n|^2 + \int_\Omega \frac{1}{2} (q^n)^2 + \int_\Omega \frac{1}{2} \mathrm{tr}(\boldsymbol{\sigma}^n)\,. \end{aligned} \tag{4.4}$$

Consequently, a numerical scheme is energy-stable, i.e., the thermodynamic consistency of the original model is conserved, if for all $n \in \mathbb{N}$

$$\frac{E_{total}(\varphi^{n+1}, q^{n+1}, \boldsymbol{\sigma}^{n+1}, \boldsymbol{u}^{n+1}) - E_{total}(\varphi^n, q^n, \boldsymbol{\sigma}^n, \boldsymbol{u}^n)}{\Delta t} \leq 0\,. \tag{4.5}$$

Theorem 4.4. *The numerical scheme (4.3) is linear and, assuming suitable boundary conditions, see Remark 3.13, it satisfies the following discrete version of energy law (3.52)*

$$\frac{E_{total}(\varphi^{n+1}, q^{n+1}, \boldsymbol{\sigma}^{n+1}, \boldsymbol{u}^{n+1}) - E_{total}(\varphi^n, q^n, \boldsymbol{\sigma}^n, \boldsymbol{u}^n)}{\Delta t} \tag{4.6a}$$
$$= - \int_\Omega \frac{1}{\zeta(\varphi^n)} \left| \varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\, q^n) \right|^2 - \frac{1}{\tau_B^0} \left\| \frac{q^n}{\varphi^n} \right\|^2_{L^2(\Omega)} \tag{4.6b}$$
$$- \int_\Omega \frac{1}{2\,\tau_S(\varphi^n)} \mathrm{tr}(\boldsymbol{\sigma}^n) - \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^n)|^2 \tag{4.6c}$$
$$- \int_\Omega \left(f(\varphi^n) \frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} \right) \tag{4.6d}$$
$$+ \frac{1}{2\,\Delta t} \left(\lambda \|\nabla \varphi^{n+1} - \nabla \varphi^n\|^2_{L^2(\Omega)} + \|q^{n+1} - q^n\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^{n+1} - \boldsymbol{u}^n\|^2_{L^2(\Omega)} \right)\,, \tag{4.6e}$$

where (4.6a)-(4.6c) are analogous to energy law (3.52), while (4.6d) and (4.6e) consist of numerical dissipation terms caused by the explicit time discretization. Here, the double-well potential integral (4.6d) can be either positive or negative, while the three norms in (4.6e) are positive. Therefore, scheme (4.3) is in general not energy-stable.

Proof. It is clear that the proposed scheme is linear. Analogously as for the continuous model, we multiply each semi-discrete partial differential equation with a suitable test function, integrate over the computational domain Ω, and apply integration by parts in order to calculate a discrete energy law. A suitable test function for (4.3a) is μ^n, giving

$$
\begin{aligned}
0 &= \int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t}\mu^n + \int_\Omega \boldsymbol{u}^n \cdot \nabla\varphi^n \mu^n \\
&\quad - \int_\Omega \nabla \cdot \left\{ \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)} \Big[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\, q^n) \Big] \right\} \mu^n \\
&= \int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t} \left(-\lambda\Delta\varphi^n + f(\varphi^n)\right) + \int_\Omega \boldsymbol{u}^n \cdot \nabla\varphi^n \mu^n \\
&\quad + \int_\Omega \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)} \Big[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\, q^n) \Big] \cdot \nabla\mu^n \\
&\quad - \int_{\partial\Omega} \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)} \Big[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\, q^n) \Big] \mu^n \cdot \boldsymbol{n} \\
&= \int_\Omega \frac{\lambda}{\Delta t}\nabla(\varphi^{n+1}-\varphi^n)\cdot\nabla\varphi^n - \int_{\partial\Omega} \frac{\varphi^{n+1}-\varphi^n}{\Delta t}\lambda\frac{\partial\varphi^n}{\partial\boldsymbol{n}} + \int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t} f(\varphi^n) + \int_\Omega \boldsymbol{u}^n\cdot\nabla\varphi^n\mu^n \\
&\quad + \int_\Omega \frac{1}{\zeta(\varphi^n)} \Big[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\, q^n) \Big] \cdot \left[\varphi^n(1-\varphi^n)\nabla\mu^n\right] \\
&\quad - \int_{\partial\Omega} \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)} \left[\varphi^n(1-\varphi^n)\frac{\partial\mu^n}{\partial\boldsymbol{n}} - \frac{\partial(G_B(\varphi^n)\, q^n)}{\partial\boldsymbol{n}} \right] \mu^n .
\end{aligned}
\tag{4.7}
$$

Assuming suitable boundary conditions, see Remark 3.13, the boundary integrals vanish. Since we want to calculate the discrete time difference of the semi-discrete mixing energy, which reads

$$
\begin{aligned}
\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} &= \frac{1}{\Delta t}\int_\Omega \left(\frac{\lambda}{2}|\nabla\varphi^{n+1}|^2 + F(\varphi^{n+1}) \right) - \frac{1}{\Delta t}\int_\Omega \left(\frac{\lambda}{2}|\nabla\varphi^n|^2 + F(\varphi^n) \right) \\
&= \int_\Omega \left(\frac{\lambda}{2\Delta t}\left(|\nabla\varphi^{n+1}|^2 - |\nabla\varphi^n|^2 \right) + \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} \right),
\end{aligned}
$$

we rewrite the first integral of (4.7) as follows

$$\begin{aligned}
&\int_\Omega \frac{\lambda}{\Delta t}\nabla(\varphi^{n+1}-\varphi^n)\cdot\nabla\varphi^n \\
&= \int_\Omega \frac{\lambda}{\Delta t}\left(\nabla\varphi^{n+1}\cdot\nabla\varphi^n - |\nabla\varphi^n|^2\right) \\
&= \int_\Omega \frac{\lambda}{\Delta t}\left(-\frac{1}{2}|\nabla\varphi^{n+1}-\nabla\varphi^n|^2 + \frac{1}{2}|\nabla\varphi^{n+1}|^2 - \frac{1}{2}|\nabla\varphi^n|^2\right) \\
&= -\frac{\lambda}{2\Delta t}\|\nabla\varphi^{n+1}-\nabla\varphi^n\|^2_{L^2(\Omega)} + \int_\Omega \frac{\lambda}{2\Delta t}\left(|\nabla\varphi^{n+1}|^2 - |\nabla\varphi^n|^2\right) \\
&\quad + \int_\Omega \left(\frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t} - \frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t}\right) \\
&= -\frac{\lambda}{2\Delta t}\|\nabla\varphi^{n+1}-\nabla\varphi^n\|^2_{L^2(\Omega)} + \frac{E_{mix}(\varphi^{n+1})-E_{mix}(\varphi^n)}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t}\,.
\end{aligned}$$

Summarized, (4.7) becomes

$$\begin{aligned}
0 = &-\frac{\lambda}{2\Delta t}\|\nabla\varphi^{n+1}-\nabla\varphi^n\|^2_{L^2(\Omega)} + \frac{E_{mix}(\varphi^{n+1})-E_{mix}(\varphi^n)}{\Delta t} \\
&+ \int_\Omega \left(\frac{\varphi^{n+1}-\varphi^n}{\Delta t} f(\varphi^n) - \frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t}\right) + \int_\Omega \boldsymbol{u}^n\cdot\nabla\varphi^n\mu^n \\
&+ \int_\Omega \frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\,q^n)\right]\cdot\left[\varphi^n(1-\varphi^n)\nabla\mu^n\right]\,.
\end{aligned} \tag{4.8}$$

A suitable test function for (4.3b) is q^n, yielding

$$\begin{aligned}
0 = &\int_\Omega \frac{q^{n+1}-q^n}{\Delta t}q^n + \int_\Omega (\boldsymbol{u}^n\cdot\nabla q^n)q^n + \int_\Omega \frac{1}{\tau_B(\varphi^n)}q^nq^n \\
&+ \int_\Omega G_B(\varphi^n)\nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\,q^n)\right]\right\}q^n \\
= &\int_\Omega \frac{1}{\Delta t}\left(q^{n+1}q^n - (q^n)^2\right) + \int_\Omega \boldsymbol{u}^n\cdot\nabla\left(\frac{1}{2}(q^n)^2\right) + \int_\Omega \frac{(q^n)^2}{\tau_B^0\,(\varphi^n)^2} \\
&+ \int_\Omega \nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\,q^n)\right]\right\}G_B(\varphi^n)q^n \\
= &\int_\Omega \frac{1}{2\Delta t}\left(-(q^{n+1}-q^n)^2 + (q^{n+1})^2 - (q^n)^2\right) + \frac{1}{\tau_B^0}\left\|\frac{q^n}{\varphi^n}\right\|^2_{L^2(\Omega)} \\
&- \int_\Omega (\nabla\cdot\boldsymbol{u}^n)\frac{1}{2}(q^n)^2 + \int_{\partial\Omega}(\boldsymbol{u}^n\cdot\boldsymbol{n})\frac{1}{2}(q^n)^2 \\
&- \int_\Omega \left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\,q^n)\right]\right\}\cdot\nabla\left(G_B(\varphi^n)q^n\right) \\
&+ \int_{\partial\Omega}\boldsymbol{n}\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\,q^n)\right]\right\}G_B(\varphi^n)q^n\,.
\end{aligned} \tag{4.9}$$

Assuming suitable boundary conditions, see Remark 3.13, using the incompressibility condition

(4.3e) and substituting the semi-discrete elastic bulk stress energy, (4.9) becomes

$$\begin{aligned} 0 = &- \frac{1}{2\Delta t}\|q^{n+1} - q^n\|^2_{L^2(\Omega)} + \frac{E_{bulk}(q^{n+1}) - E_{bulk}(q^n)}{\Delta t} + \frac{1}{\tau_B^0}\left\|\frac{q^n}{\varphi^n}\right\|^2_{L^2(\Omega)} \\ &- \int_\Omega \left\{ \frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^n - \nabla(G_B(\varphi^n)\, q^n)\right]\right\} \cdot \nabla\left(G_B(\varphi^n) q^n\right) . \end{aligned} \tag{4.10}$$

Taking the Frobenius inner product of (4.3c) and $\frac{1}{2}\boldsymbol{I}$ and integrating over the computational domain Ω yields

$$\begin{aligned} 0 = &\int_\Omega \frac{\boldsymbol{\sigma}^{n+1} - \boldsymbol{\sigma}^n}{2\Delta t} : \boldsymbol{I} + \int_\Omega (\boldsymbol{u}^n \cdot \nabla)\boldsymbol{\sigma}^n : \frac{1}{2}\boldsymbol{I} - \int_\Omega \left((\nabla \boldsymbol{u}^n)\cdot \boldsymbol{\sigma}^n + \boldsymbol{\sigma}^n \cdot (\nabla \boldsymbol{u}^n)^T\right) : \frac{1}{2}\boldsymbol{I} \\ &+ \int_\Omega \frac{1}{2\tau_S(\varphi^n)}\boldsymbol{\sigma}^n : \boldsymbol{I} - \int_\Omega G_S(\varphi^n)\left[\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T\right] : \frac{1}{2}\boldsymbol{I} \\ = &\int_\Omega \frac{\mathrm{tr}(\boldsymbol{\sigma}^{n+1}) - \mathrm{tr}(\boldsymbol{\sigma}^n)}{2\Delta t} + \int_\Omega \frac{1}{2}\mathrm{tr}\left((\boldsymbol{u}^n \cdot \nabla)\boldsymbol{\sigma}^n\right) - \int_\Omega \frac{1}{2}\left(\nabla \boldsymbol{u}^n : (\boldsymbol{\sigma}^n)^T + \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^n\right) \\ &+ \int_\Omega \frac{1}{2\tau_S(\varphi^n)}\mathrm{tr}(\boldsymbol{\sigma}^n) - \int_\Omega G_S(\varphi^n)\mathrm{tr}(\nabla \boldsymbol{u}^n) \\ = &\frac{E_{el}(\boldsymbol{\sigma}^{n+1}) - E_{el}(\boldsymbol{\sigma}^n)}{\Delta t} - \int_\Omega \frac{1}{2}\mathrm{tr}\left((\nabla \cdot \boldsymbol{u}^n)\boldsymbol{\sigma}^n\right) + \int_{\partial\Omega} \frac{1}{2}\mathrm{tr}\left((\boldsymbol{u}^n \cdot \boldsymbol{n})\boldsymbol{\sigma}^n\right) - \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^n \\ &+ \int_\Omega \frac{1}{2\tau_S(\varphi^n)}\mathrm{tr}(\boldsymbol{\sigma}^n) - \int_\Omega G_S(\varphi^n)(\nabla \cdot \boldsymbol{u}^n) . \end{aligned} \tag{4.11}$$

Assuming suitable boundary conditions for $\boldsymbol{u}$, see Remark 3.13, and using the incompressibility, (4.11) becomes

$$0 = \frac{E_{el}(\boldsymbol{\sigma}^{n+1}) - E_{el}(\boldsymbol{\sigma}^n)}{\Delta t} - \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^n + \int_\Omega \frac{1}{2\tau_S(\varphi^n)}\mathrm{tr}(\boldsymbol{\sigma}^n) . \tag{4.12}$$

Taking the inner product of (4.3d) and $\boldsymbol{u}^n$ and integrating over Ω yields

$$\begin{aligned} 0 = &\int_\Omega \frac{\boldsymbol{u}^{n+1} - \boldsymbol{u}^n}{\Delta t} \cdot \boldsymbol{u}^n + \int_\Omega (\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^n \cdot \boldsymbol{u}^n \\ &- \int_\Omega \left(\nabla \cdot \left\{\eta(\varphi^n)\left[\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T\right]\right\}\right) \cdot \boldsymbol{u}^n \\ &+ \int_\Omega \nabla p^n \cdot \boldsymbol{u}^n - \int_\Omega \mu^n \nabla \varphi^n \cdot \boldsymbol{u}^n - \int_\Omega (\nabla \cdot \boldsymbol{\sigma}^n) \cdot \boldsymbol{u}^n \\ = &\frac{1}{\Delta t}\int_\Omega \boldsymbol{u}^{n+1} \cdot \boldsymbol{u}^n - |\boldsymbol{u}^n|^2 + \int_\Omega (\boldsymbol{u}^n \cdot \nabla)\frac{1}{2}|\boldsymbol{u}^n|^2 \\ &+ \int_\Omega \left\{\eta(\varphi^n)\left[\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T\right]\right\} : \nabla \boldsymbol{u}^n - \int_{\partial\Omega} \left\{\eta(\varphi^n)\left[\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T\right]\right\} \boldsymbol{u}^n \cdot \boldsymbol{n} \\ &- \int_\Omega p^n(\nabla \cdot \boldsymbol{u}^n) + \int_{\partial\Omega} p^n \boldsymbol{u}^n \cdot \boldsymbol{n} - \int_\Omega \mu^n \nabla \varphi^n \cdot \boldsymbol{u}^n + \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^n - \int_{\partial\Omega} \boldsymbol{\sigma}^n \boldsymbol{u}^n \cdot \boldsymbol{n} \\ = &\frac{1}{2\,\Delta t}\int_\Omega -|\boldsymbol{u}^{n+1} - \boldsymbol{u}^n|^2 + |\boldsymbol{u}^{n+1}|^2 - |\boldsymbol{u}^n|^2 - \int_\Omega (\nabla \cdot \boldsymbol{u}^n)\frac{1}{2}|\boldsymbol{u}^n|^2 + \int_{\partial\Omega} (\boldsymbol{u}^n \cdot \boldsymbol{n})\frac{1}{2}|\boldsymbol{u}^n|^2 \\ &+ \int_\Omega \eta(\varphi^n)\left[|\nabla \boldsymbol{u}^n|^2 + \mathrm{tr}\left((\nabla \boldsymbol{u}^n)^2\right)\right] - \int_{\partial\Omega} \left\{\eta(\varphi^n)\left[\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T\right]\right\} \boldsymbol{u}^n \cdot \boldsymbol{n} \\ &- \int_\Omega p^n(\nabla \cdot \boldsymbol{u}^n) + \int_{\partial\Omega} p^n \boldsymbol{u}^n \cdot \boldsymbol{n} - \int_\Omega \boldsymbol{u}^n \cdot \nabla \varphi^n \mu^n + \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^n - \int_{\partial\Omega} \boldsymbol{\sigma}^n \boldsymbol{u}^n \cdot \boldsymbol{n} . \end{aligned} \tag{4.13}$$

Using the incompressibility, (4.13) can be reduced to

$$\begin{aligned} 0 = &- \frac{1}{2\,\Delta t}\|\boldsymbol{u}^{n+1} - \boldsymbol{u}^n\|^2_{L^2(\Omega)} + \frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \int_\Omega \frac{\eta(\varphi^n)}{2} \sum_{i,j=1}^{d} \left(\frac{\partial u_i^n}{\partial x_j} + \frac{\partial u_j^n}{\partial x_i} \right)^2 \\ &- \int_\Omega \boldsymbol{u}^n \cdot \nabla\varphi^n \mu^n + \int_\Omega \boldsymbol{\sigma}^n \,:\, \nabla \boldsymbol{u}^n \\ &+ \int_{\partial\Omega} \left(\frac{1}{2}|\boldsymbol{u}^n|^2 \boldsymbol{u}^n + p^n \boldsymbol{u}^n - \left\{ \eta(\varphi^n) \left[\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T \right] \right\} \boldsymbol{u}^n - \boldsymbol{\sigma}^n \boldsymbol{u}^n \right) \cdot \boldsymbol{n} \,. \end{aligned} \tag{4.14}$$

Furthermore, with

$$\int_\Omega \frac{\eta(\varphi^n)}{2} \sum_{i,j=1}^{d} \left(\frac{\partial u_i^n}{\partial x_j} + \frac{\partial u_j^n}{\partial x_i} \right)^2 = \int_\Omega \frac{\eta(\varphi^n)}{2} |\nabla \boldsymbol{u}^n + (\nabla \boldsymbol{u}^n)^T|^2 = \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^n)|^2 \,,$$

and assuming suitable boundary conditions for $\boldsymbol{u}$ like above, (4.14) becomes

$$\begin{aligned} 0 = &\frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} - \frac{1}{2\,\Delta t}\|\boldsymbol{u}^{n+1} - \boldsymbol{u}^n\|^2_{L^2(\Omega)} \\ &+ \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^n)|^2 - \int_\Omega \boldsymbol{u}^n \cdot \nabla\varphi^n \mu^n + \int_\Omega \boldsymbol{\sigma}^n \,:\, \nabla \boldsymbol{u}^n \,. \end{aligned} \tag{4.15}$$

Summing up the four discrete energy laws (4.8), (4.10), (4.12) and (4.15), the integral forms of the coupling terms cancel out thanks to the chosen test functions and we have the claimed discrete total energy law (4.6). □

To validate thermodynamic consistency, it is crucial that we either eliminate the three positive norms at the end of energy law (4.6), or that we can at least substitute them by negative terms, e.g., by changing their signs. This can be achieved by an at least partially implicit discretization, as we will show in the following sections. Since the integral including the double-well potential can be either positive or negative, we have to control it somehow as well.

In the following, we start with suitable discretizations for the derivative of the double-well potential $F(\varphi)$, before we introduce thermodynamically consistent schemes for the Cahn-Hilliard equation and successively move forward up to the full model.

4.2 Approximations for $f(\varphi)$

The choice of a suitable linear approximation $f(\varphi^{n+1}, \varphi^n)$ for $f(\varphi) = F'(\varphi)$ strongly depends on the given double-well potential $F(\varphi)$. We present several approximations and their applicability in the following and summarize their properties in Remark 4.9 at the end of this section.

4.2.1 Midpoint approximation

The only way to eliminate the numerical dissipation term

$$\mathrm{ND}_{pot}^{n+1} := \int_\Omega f(\varphi^{n+1}, \varphi^n) \frac{\varphi^{n+1} - \varphi^n}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} \tag{4.16}$$

in the discrete energy law is achieved by using the midpoint approximation

$$f(\varphi^{n+1}, \varphi^n) = \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\varphi^{n+1} - \varphi^n}. \tag{4.17}$$

Approximation (4.17) is nonlinear, since both the Ginzburg-Landau and the Flory-Huggins potential and all of their modifications introduced in Subsection 3.1.1 are nonlinear. Thus, since we focus on linear schemes, we cannot eliminate the numerical dissipation ND_{pot}^{n+1}. Instead, we introduce approximations in the following, which either result in positive dissipation or in dissipation whose dimension is of the same or a higher order as the error in time of our numerical schemes.

4.2.2 Linear Eyre approximation

In [27] Eyre proposes a linear semi-implicit time discretization for the derivative $f(\varphi)$ of the Ginzburg-Landau potential (3.9), which ensures the positivity of the numerical dissipation term ND_{pot}^{n+1} for certain intervals of φ. Let us extend the Ginzburg-Landau potential to

$$F_{GL}(\varphi) = \frac{1}{4}(\varphi^2 - 1)^2 = \frac{1}{4}(\varphi^4 + 2\beta\varphi^2 - 2(\beta+1)\varphi^2 + 1), \tag{4.18}$$

where $\beta > 0$, and split it into

$$F_1(\varphi) = \frac{1}{2}\beta\varphi^2 \quad \text{and} \quad F_2(\varphi) = \frac{1}{4}(\varphi^4 - 2(\beta+1)\varphi^2 + 1). \tag{4.19}$$

Then, the linear Eyre approximation reads

$$\begin{aligned} f(\varphi^{n+1}, \varphi^n) &= F_1'(\varphi^{n+1}) + F_2'(\varphi^n) \\ &= (\varphi^n)^3 - (\beta+1)\varphi^n + \beta\varphi^{n+1} \\ &= f(\varphi^n) + \beta(\varphi^{n+1} - \varphi^n). \end{aligned} \tag{4.20}$$

Theorem 4.5. *Approximation (4.20) satisfies for any potential*

$$\mathrm{ND}_{pot}^{n+1} \in \mathcal{O}(\Delta t) \qquad \forall n \in \mathbb{N},$$

and for the Ginzburg-Landau potential (3.9) additionally

$$\mathrm{ND}_{pot}^{n+1} \geq 0,$$

if for all $\boldsymbol{x} \in \Omega$ and some $\zeta \in (0,1)$ with $\varphi^{n+\zeta} = \varphi^n + \zeta(\varphi^{n+1} - \varphi^n) = \zeta\varphi^{n+1} + (1-\zeta)\varphi^n$

$$\frac{2\beta+1}{3} \geq (\varphi^{n+\zeta})^2.$$

Proof. Replacing $F(\varphi^{n+1})$ in the numerical dissipation term (4.16) by the following *Taylor expansion* [70]

$$F(\varphi^{n+1}) = F(\varphi^n) + (\varphi^{n+1} - \varphi^n) f(\varphi^n) + \frac{(\varphi^{n+1} - \varphi^n)^2}{2} f'(\varphi^{n+\zeta})$$

yields

$$\begin{aligned}\mathrm{ND}_{pot}^{n+1} &= \int_\Omega f(\varphi^{n+1},\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t} \\ &= \int_\Omega f(\varphi^{n+1},\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - f(\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - f'(\varphi^{n+\zeta})\frac{(\varphi^{n+1}-\varphi^n)^2}{2\Delta t}\,.\end{aligned}$$

Using approximation (4.20), we have

$$\begin{aligned}\mathrm{ND}_{pot}^{n+1} &= \int_\Omega \left(f(\varphi^n)+\beta(\varphi^{n+1}-\varphi^n)\right)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - f(\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - f'(\varphi^{n+\zeta})\frac{(\varphi^{n+1}-\varphi^n)^2}{2\Delta t} \\ &= \int_\Omega \beta\frac{(\varphi^{n+1}-\varphi^n)^2}{\Delta t} - f'(\varphi^{n+\zeta})\frac{(\varphi^{n+1}-\varphi^n)^2}{2\Delta t} \\ &= \int_\Omega \frac{\Delta t}{2}\left(-f'(\varphi^{n+\zeta})+2\beta\right)\left(\frac{\varphi^{n+1}-\varphi^n}{\Delta t}\right)^2\,.\end{aligned}$$

Then, substituting the discrete time derivative

$$\partial_t\varphi^{n+\frac{1}{2}} = \frac{\varphi^{n+1}-\varphi^n}{\Delta t} \tag{4.21}$$

yields

$$\mathrm{ND}_{pot}^{n+1} = \frac{\Delta t}{2}\int_\Omega \left(-f'(\varphi^{n+\zeta})+2\beta\right)\left(\partial_t\varphi^{n+\frac{1}{2}}\right)^2\,. \tag{4.22}$$

Thus, it holds

$$\mathrm{ND}_{pot}^{n+1} \in \mathcal{O}(\Delta t) \qquad \forall n\in\mathbb{N}\,.$$

Further, using the Ginzburg-Landau potential (3.9), which fulfills

$$f'(\varphi^{n+\zeta}) = 3(\varphi^{n+\zeta})^2 - 1\,,$$

equation (4.22) satisfies

$$\mathrm{ND}_{pot}^{n+1} = \frac{\Delta t}{2}\int_\Omega \left(-3(\varphi^{n+\zeta})^2+2\beta+1\right)\left(\partial_t\varphi^{n+\frac{1}{2}}\right)^2\,.$$

Therefore, it holds

$$\mathrm{ND}_{pot}^{n+1} \geq 0\,,$$

if for all $\boldsymbol{x}\in\Omega$

$$-3(\varphi^{n+\zeta})^2+2\beta+1\geq 0 \qquad \Leftrightarrow \qquad \frac{2\beta+1}{3}\geq(\varphi^{n+\zeta})^2\,.$$

□

The original discretization introduced by *Eyre* in [27] considers $\beta=2$ and therefore satisfies

$$\mathrm{ND}_{pot}^{n+1}\geq 0 \quad \text{if} \quad \forall\boldsymbol{x}\in\Omega:\ (\varphi^{n+\zeta})^2\leq\frac{5}{3}\,.$$

Setting, e.g., $\beta = 1$ yields less numerical dissipation, but satisfies $\mathrm{ND}_{pot}^{n+1} \geq 0$ only if for all $\boldsymbol{x} \in \Omega : (\varphi^{n+\zeta})^2 \leq 1$. Let us point out that no matter how we chose β, it is not clear that always $\mathrm{ND}_{pot}^{n+1} \geq 0$, since the Cahn-Hilliard equation does not obey a maximum principle.

However, we can fix this issue by, e.g., using the modified version (3.12) of the Ginzburg-Landau potential, which has the bounded second derivative

$$\|f'\|_{L^\infty(\mathbb{R})} = \|f'\|_{L^\infty(-1,1)} = 2\,.$$

Setting $\beta = \frac{1}{2}\|f'\|_{L^\infty(\mathbb{R})}$ in the linear Eyre approximation (4.20), we get the general approximation

$$f(\varphi^{n+1}, \varphi^n) = f(\varphi^n) + \frac{1}{2}\|f'\|_{L^\infty(\mathbb{R})}(\varphi^{n+1} - \varphi^n)\,, \tag{4.23}$$

which was originally proposed in this form by *Guillén-González, Rodríguez-Bellido and Tierra* in [38].

Theorem 4.6. *Using a potential with a bounded second derivative, approximation* (4.23) *satisfies for all* $n \in \mathbb{N}$:

$$\mathrm{ND}_{pot}^{n+1} \in \mathcal{O}(\Delta t) \qquad \textit{and} \qquad \mathrm{ND}_{pot}^{n+1} \geq 0\,.$$

Proof. The first part

$$\mathrm{ND}_{pot}^{n+1} \in \mathcal{O}(\Delta t) \qquad \forall n \in \mathbb{N}$$

directly follows from Theorem 4.5. Replacing $F(\varphi^{n+1})$ in the numerical dissipation term (4.16) by the Taylor expansion

$$F(\varphi^{n+1}) = F(\varphi^n) + (\varphi^{n+1} - \varphi^n) f(\varphi^n) + \frac{(\varphi^{n+1} - \varphi^n)^2}{2} f'(\zeta)\,,$$

where $\zeta \in (\varphi^n, \varphi^{n+1})$ or $\zeta \in (\varphi^{n+1}, \varphi^n)$, yields

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega f(\varphi^{n+1}, \varphi^n)\frac{\varphi^{n+1} - \varphi^n}{\Delta t} - f(\varphi^n)\frac{\varphi^{n+1} - \varphi^n}{\Delta t} - f'(\zeta)\frac{(\varphi^{n+1} - \varphi^n)^2}{2\Delta t}\,.$$

Using approximation (4.23), we have

$$\begin{aligned}\mathrm{ND}_{pot}^{n+1} &= \int_\Omega \frac{1}{2}\|f'\|_{L^\infty(\mathbb{R})}\frac{(\varphi^{n+1} - \varphi^n)^2}{\Delta t} - f'(\zeta)\frac{(\varphi^{n+1} - \varphi^n)^2}{2\Delta t} \\ &= \frac{1}{2\Delta t}\int_\Omega \left(\|f'\|_{L^\infty(\mathbb{R})} - f'(\zeta)\right)\left(\varphi^{n+1} - \varphi^n\right)^2\,,\end{aligned}$$

which satisfies

$$\mathrm{ND}_{pot}^{n+1} \geq 0 \qquad \forall n \in \mathbb{N}\,,$$

since $\|f'\|_{L^\infty(\mathbb{R})} \geq f'(\zeta)$ for all $\zeta \in (\varphi^n, \varphi^{n+1})$ or $\zeta \in (\varphi^{n+1}, \varphi^n)$ and $n \in \mathbb{N}$. □

Note that approximation (4.23) and Theorem 4.6 hold for any potential with a bounded second derivative. But the resulting numerical dissipation rises with the size of the upper bound of its second derivative, which can become quite large for the proposed modifications of the Flory-Huggins potential, see Subsection 3.1.1. The further away from the minima the cut is applied, the higher the upper bound of the second derivative, because the original potential and

its derivatives have singularities at 0 and 1. Even if we cut directly at the minima as proposed in modification (3.14), the bound is much higher than that of the modified Ginzburg-Landau potential (3.12).

Nevertheless, if using a modified Flory-Huggins potential, we suggest using the following extension

$$\begin{aligned} F_{FH}(\varphi) &= \frac{1}{n_p}\varphi\ln\varphi + \frac{1}{n_s}(1-\varphi)\ln(1-\varphi) + \chi\varphi(1-\varphi) \\ &= \frac{1}{n_p}\varphi\ln\varphi + \frac{1}{n_s}(1-\varphi)\ln(1-\varphi) + (\chi+\frac{1}{2}\beta)\varphi(1-\varphi) - \frac{1}{2}\beta\varphi(1-\varphi)\,, \end{aligned} \tag{4.24}$$

which can be split into

$$F_1(\varphi) = -\frac{1}{2}\beta\varphi(1-\varphi)\,, \tag{4.25a}$$

$$F_2(\varphi) = \frac{1}{n_p}\varphi\ln\varphi + \frac{1}{n_s}(1-\varphi)\ln(1-\varphi) + (\chi+\frac{1}{2}\beta)\varphi(1-\varphi)\,. \tag{4.25b}$$

Using this splitting, the approximation

$$\begin{aligned} f(\varphi^{n+1},\varphi^n) &= F_1'(\varphi^{n+1}) + F_2'(\varphi^n) \\ &= -\frac{1}{2}\beta + \beta\varphi^{n+1} + \frac{1}{n_p}(1+\ln\varphi^n) - \frac{1}{n_s}(1+\ln(1-\varphi^n)) + (\chi+\frac{1}{2}\beta)(1-2\varphi^n) \\ &= f(\varphi^n) + \beta(\varphi^{n+1}-\varphi^n) \end{aligned}$$

is analogous to the linear Eyre approximation (4.20).

4.2.3 Optimal dissipation 2 approximation

As a universal linear second order approximation for any potential, we suggest using the second order accurate Taylor expansion

$$f(\varphi^{n+1},\varphi^n) = f(\varphi^n) + \frac{\varphi^{n+1}-\varphi^n}{2}f'(\varphi^n)\,. \tag{4.26}$$

This is called the *optimal dissipation 2* (OD2) approximation, see *Guillén-González and Tierra* [37], because the generated numerical dissipation is of second order in time.

Theorem 4.7. *Approximation* (4.26) *satisfies*

$$\mathrm{ND}_{pot}^{n+1} \in \mathcal{O}\left((\Delta t)^2\right) \qquad \forall n \in \mathbb{N}\,.$$

Proof. Substituting the following Taylor expansion, where $\zeta \in (\varphi^n,\varphi^{n+1})$ or $\zeta \in (\varphi^{n+1},\varphi^n)$,

$$F(\varphi^{n+1}) = F(\varphi^n) + (\varphi^{n+1}-\varphi^n)f(\varphi^n) + \frac{(\varphi^{n+1}-\varphi^n)^2}{2}f'(\varphi^n) + \frac{(\varphi^{n+1}-\varphi^n)^3}{6}f''(\zeta)$$

and approximation (4.26) into the numerical dissipation term (4.16) yields

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega -\frac{1}{6\Delta t}f''(\zeta)\left(\varphi^{n+1}-\varphi^n\right)^3\,.$$

Then, substituting the discrete time derivative (4.21), we have

$$\mathrm{ND}_{pot}^{n+1} = (\Delta t)^2 \int_\Omega -\frac{1}{6} f''(\zeta) \left(\partial_t \varphi^{n+\frac{1}{2}}\right)^3 . \tag{4.27}$$

Thus, it holds

$$\mathrm{ND}_{pot}^{n+1} \in \mathcal{O}\left((\Delta t)^2\right) \qquad \forall n \in \mathbb{N} .$$

□

Note that this approximation is not able to control the sign of ND_{pot}^{n+1}. Nevertheless, we only use schemes with truncation error up to second order in time. One may assume that an approximation causing dissipation which is of similar or smaller scale as the truncation error of the whole scheme does not violate the thermodynamic consistency as long as the time step size is sufficiently small. And indeed, our numerical experiments presented in Chapter 5 validate this.

4.2.4 Linear stabilized second order approximation

We can also construct a linear second order accurate approximation for the modified Ginzburg-Landau potential (3.12), which can provably ensure the energy-stability of a resulting numerical scheme, even though it does not guarantee the positivity of the numerical dissipation term. Using the same splitting as in (4.19) with $\beta = 2$, $F_1(\varphi) = \varphi^2 =: F_{vex}(\varphi)$ is convex and $F_2(\varphi) = \frac{1}{4}(\varphi^4 - 6\varphi^2 + 1) =: F_{cave}(\varphi)$ is concave in $[-1, 1]$. Then, the approximation, which was originally proposed by *Wu, van Zwieten and van der Zee* in [72], reads

$$\begin{aligned} f(\varphi^{n+1}, \varphi^n) =& f_{vex}(\varphi^{n+1}) - \frac{\varphi^{n+1} - \varphi^n}{2} f'_{vex}(\varphi^{n+1}) + f_{cave}(\varphi^n) + \frac{\varphi^{n+1} - \varphi^n}{2} f'_{cave}(\varphi^n) \\ & - \Delta t \alpha \Delta \left(\varphi^{n+1} - \varphi^n\right) , \end{aligned} \tag{4.28}$$

which is a combination of two second order accurate Taylor expansions and a stabilization term. Note that the coefficient α of the stabilization term has to satisfy the following estimate

$$\alpha \geq \frac{\|M\|_{L^\infty(\mathbb{R})} \left(\|f'_{vex}\|_{L^\infty(\mathbb{R})} + \| - f'_{cave}\|_{L^\infty(\mathbb{R})}\right)^2}{16}$$

in order to get a provably energy-stable numerical scheme, see [72] for more details and the proof. Further, note that M in the above estimate is the mobility function (3.4), which in case of not being bounded has to be cut and continuously extended analogously to the second derivatives of the modified potentials in order to be bounded. The truncation error of the stabilization term is also of second order in time, since

$$\Delta t \alpha \Delta \left(\varphi^{n+1} - \varphi^n\right) = (\Delta t)^2 \alpha \Delta \left(\frac{\varphi^{n+1} - \varphi^n}{\Delta t}\right) = (\Delta t)^2 \alpha \Delta \left(\partial_t \varphi^{n+\frac{1}{2}}\right) .$$

Approximation (4.28) is linear, since the implicitly discretized derivative f_{vex} of the convex part of the potential is linear and its second derivative f'_{vex} is constant.

Theorem 4.8. *Approximation (4.28) satisfies*

$$\mathrm{ND}_{pot}^{n+1} \in \mathcal{O}\left((\Delta t)^2\right) \qquad \forall n \in \mathbb{N}\,.$$

Proof. Substituting the Taylor expansions

$$F_{cave}(\varphi^{n+1}) = F_{cave}(\varphi^n) + (\varphi^{n+1}-\varphi^n) f_{cave}(\varphi^n) + \frac{(\varphi^{n+1}-\varphi^n)^2}{2} f'_{cave}(\varphi^n) + \frac{(\varphi^{n+1}-\varphi^n)^3}{6} f''_{cave}(\zeta)\,,$$
$$F_{vex}(\varphi^n) = F_{vex}(\varphi^{n+1}) - (\varphi^{n+1}-\varphi^n) f_{vex}(\varphi^n) + \frac{(\varphi^{n+1}-\varphi^n)^2}{2} f'_{vex}(\varphi^{n+1}) - \frac{(\varphi^{n+1}-\varphi^n)^3}{6} f''_{vex}(\zeta)$$

and approximation (4.28) into the numerical dissipation term

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega \left(f(\varphi^{n+1},\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \frac{F_{cave}(\varphi^{n+1}) + F_{vex}(\varphi^{n+1}) - F_{cave}(\varphi^n) - F_{vex}(\varphi^n)}{\Delta t} \right)$$

yields

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega \left(-\Delta t \alpha \Delta(\varphi^{n+1}-\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \frac{1}{6\Delta t}(f''_{cave}(\zeta) + f''_{vex}(\zeta))(\varphi^{n+1}-\varphi^n)^3 \right).$$

Using partial integration and assuming suitable boundary conditions, see Remark 3.2, we have

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega \left(\alpha |\nabla(\varphi^{n+1}-\varphi^n)|^2 - \frac{1}{6\Delta t} f''(\zeta)(\varphi^{n+1}-\varphi^n)^3 \right).$$

Then, substituting the discrete time derivative (4.21) yields

$$\mathrm{ND}_{pot}^{n+1} = (\Delta t)^2 \int_\Omega \left(\alpha \left| \nabla\left(\partial_t \varphi^{n+\frac{1}{2}}\right)\right|^2 - \frac{1}{6} f''(\zeta)\left(\partial_t \varphi^{n+\frac{1}{2}}\right)^3 \right).$$

Thus, it holds

$$\mathrm{ND}_{pot}^{n+1} \in \mathcal{O}\left((\Delta t)^2\right) \qquad \forall n \in \mathbb{N}\,.$$

□

Approximation (4.28) is also applicable for the modified Flory-Huggins potential (3.14), if we use the above-mentioned splitting (4.25). But note that the proof of energy-stability in [72] is based on the fact that the splitting is convex-concave, with $F_1(\varphi)$ being convex and having a linear derivative and $F_2(\varphi)$ being concave. While $F_1(\varphi)$ given by (4.25a) is indeed convex for $\beta > 0$ and has a linear derivative, the unconditional concavity of $F_2(\varphi)$ given by (4.25b) can only be guaranteed for a suitably large $\beta \gg 0$. Further, the size of the numerical dissipation of this approximation is again related to the upper bound of the second derivative of the potential, which is directly related to β. And as mentioned above, this upper bound for modified Flory-Huggins potentials is much higher than that of the modified Ginzburg-Landau potential (3.12).

Remark 4.9. *Summarized, see Table 4.1, we have introduced some linear potential derivative approximations, which do not void the thermodynamic consistency under certain conditions, and / or if used for the modified Ginzburg-Landau potential (mod. GL) or the modified Flory-Huggins*

Table 4.1: Approximation summary for $f(\varphi)$

	error and ND_{pot} size	$ND_{pot} \geq 0$ unmod. pot.	mod. GL	mod. FH
Eyre (4.20)	$\mathcal{O}(\Delta t)$	conditional	conditional	conditional
mod. Eyre (4.23)	$\mathcal{O}(\Delta t)$	inapplicable	✓	✓
OD2 (4.26)	$\mathcal{O}\left((\Delta t)^2\right)$	-	-	-
Stabilized (4.28)	$\mathcal{O}\left((\Delta t)^2\right)$	inapplicable	- (but energy-stable)	- (but energy-stable)

potential (mod. FH). But they can introduce large dissipation, since they include stabilization terms relying on parameters which may have to be large. We have also introduced the OD2 approximation, which is always applicable and causes small dissipation, since it does not include stabilization terms, but can theoretically void the thermodynamic consistency. Despite this, we preferably use OD2 for our numerical experiments in Chapter 5 due to its flexibility and universally small dissipation. We then always check the energy-stability of our simulations.

4.3 Schemes for the Cahn-Hilliard equation

As mentioned above, an at least partially implicit discretization is necessary to prove thermodynamic consistency for a semi-discretization in time of the Cahn-Hilliard equation (3.7). We present several suitable discretizations in the following and summarize their properties in Remark 4.17 at the end of this section.

4.3.1 Partially implicit Euler scheme

An intuitive approach for a thermodynamically consistent scheme for the Cahn-Hilliard equation results from an implicit discretization of the fourth derivative of the volume fraction φ, which results from the second derivative of the chemical potential μ. We will show in Theorem 4.10 that this partially implicit Euler discretization changes the sign of the related L^2-Norm $\frac{\lambda}{2\Delta t}\|\nabla\varphi^{n+1} - \nabla\varphi^n\|^2_{L^2(\Omega)}$ in the resulting discrete energy law. We consider a uniform partition of the time interval $[0,T]$ with a constant time step Δt. Given φ^n from the previous time step, we compute φ^{n+1} such that

$$\frac{\varphi^{n+1} - \varphi^n}{\Delta t} - \nabla \cdot \left(M(\varphi^n)\nabla\mu^{n+1}\right) = 0\,, \tag{4.29a}$$

$$\mu^{n+1} = -\lambda\Delta\varphi^{n+1} + f(\varphi^{n+1}, \varphi^n)\,, \tag{4.29b}$$

where $f(\varphi^{n+1}, \varphi^n)$ is a suitably linearized approximation of $f(\varphi) = F'(\varphi)$. A reasonable choice for the approximation $f(\varphi^{n+1}, \varphi^n)$ is a topic on its own and therefore outlined in Section 4.2 above.

Note that we do not need to compute φ^{n+1} and μ^{n+1} in a coupled way, since we can insert relation (4.29b) in equation (4.29a), before solving the latter numerically.

Theorem 4.10. *Let $f(\varphi^{n+1}, \varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.29) is linear and, assuming suitable boundary conditions, see Remark 3.2, it satisfies the following discrete version of energy law (3.8)*

$$\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} = -\int_\Omega M(\varphi^n)|\nabla\mu^{n+1}|^2 - \mathrm{ND}_{pot}^{n+1} - \frac{\lambda}{2\Delta t}\|\nabla\varphi^{n+1} - \nabla\varphi^n\|^2_{L^2(\Omega)}, \tag{4.30}$$

where

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega f(\varphi^{n+1}, \varphi^n)\frac{\varphi^{n+1} - \varphi^n}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t}.$$

Proof. It is clear that the proposed scheme is linear if $f(\varphi^{n+1}, \varphi^n)$ is linear. Multiplying equation (4.29a) with μ^{n+1}, integrating over the computational domain Ω and applying integration by parts results in

$$\begin{aligned}
0 &= \int_\Omega \frac{\varphi^{n+1} - \varphi^n}{\Delta t}\mu^{n+1} - \int_\Omega \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+1}\right)\mu^{n+1} \\
&= \int_\Omega \frac{\varphi^{n+1} - \varphi^n}{\Delta t}\left(-\lambda\Delta\varphi^{n+1} + f(\varphi^{n+1}, \varphi^n)\right) + \int_\Omega M(\varphi^n)\nabla\mu^{n+1}\cdot\nabla\mu^{n+1} \\
&\quad - \int_{\partial\Omega} M(\varphi^n)\nabla\mu^{n+1}\mu^{n+1}\cdot\boldsymbol{n} \\
&= \int_\Omega \frac{\lambda}{\Delta t}\nabla(\varphi^{n+1} - \varphi^n)\cdot\nabla\varphi^{n+1} - \int_{\partial\Omega}\frac{\varphi^{n+1} - \varphi^n}{\Delta t}\lambda\frac{\partial\varphi^{n+1}}{\partial\boldsymbol{n}} + \int_\Omega \frac{\varphi^{n+1} - \varphi^n}{\Delta t}f(\varphi^{n+1}, \varphi^n) \\
&\quad + \int_\Omega M(\varphi^n)|\nabla\mu^{n+1}|^2 - \int_{\partial\Omega} M(\varphi^n)\frac{\partial\mu^{n+1}}{\partial\boldsymbol{n}}\mu^{n+1}.
\end{aligned} \tag{4.31}$$

Assuming suitable boundary conditions, see Remark 3.2, the boundary integrals vanish and since the first integral satisfies

$$\begin{aligned}
\int_\Omega \frac{\lambda}{\Delta t}\nabla(\varphi^{n+1} - \varphi^n)\cdot\nabla\varphi^{n+1} &= \int_\Omega \frac{\lambda}{\Delta t}\left(|\nabla\varphi^{n+1}|^2 - \nabla\varphi^n\cdot\nabla\varphi^{n+1}\right) \\
&= \int_\Omega \frac{\lambda}{\Delta t}\left(\frac{1}{2}|\nabla\varphi^{n+1} - \nabla\varphi^n|^2 + \frac{1}{2}|\nabla\varphi^{n+1}|^2 - \frac{1}{2}|\nabla\varphi^n|^2\right) \\
&= \frac{\lambda}{2\Delta t}\|\nabla\varphi^{n+1} - \nabla\varphi^n\|^2_{L^2(\Omega)} + \int_\Omega \frac{\lambda}{2\Delta t}\left(|\nabla\varphi^{n+1}|^2 - |\nabla\varphi^n|^2\right),
\end{aligned}$$

equation (4.31) becomes

$$\begin{aligned}
0 = \frac{\lambda}{2\Delta t}\|\nabla\varphi^{n+1} - \nabla\varphi^n\|^2_{L^2(\Omega)} + \int_\Omega \frac{\lambda}{2\Delta t}\left(|\nabla\varphi^{n+1}|^2 - |\nabla\varphi^n|^2\right) \\
+ \int_\Omega \frac{\varphi^{n+1} - \varphi^n}{\Delta t}f(\varphi^{n+1}, \varphi^n) + \int_\Omega M(\varphi^n)|\nabla\mu^{n+1}|^2.
\end{aligned} \tag{4.32}$$

Further, since the discrete mixing energy at time t_n reads

$$E_{mix}(\varphi^n) = \int_\Omega \left(\frac{\lambda}{2}|\nabla\varphi^n|^2 + F(\varphi^n)\right),$$

we sum equation (4.32) with

$$0 = \int_\Omega \left(\frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} - \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} \right),$$

to get

$$\begin{aligned} 0 &= \frac{\lambda}{2\Delta t} \|\nabla\varphi^{n+1} - \nabla\varphi^n\|^2_{L^2(\Omega)} + \int_\Omega \left(\frac{\lambda}{2\Delta t} \left(|\nabla\varphi^{n+1}|^2 - |\nabla\varphi^n|^2 \right) + \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} \right) \\ &\quad + \int_\Omega \left(\frac{\varphi^{n+1} - \varphi^n}{\Delta t} f(\varphi^{n+1}, \varphi^n) - \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} \right) + \int_\Omega M(\varphi^n) |\nabla\mu^{n+1}|^2 \\ &= \frac{\lambda}{2\Delta t} \|\nabla\varphi^{n+1} - \nabla\varphi^n\|^2_{L^2(\Omega)} + \frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} \\ &\quad + \int_\Omega \left(\frac{\varphi^{n+1} - \varphi^n}{\Delta t} f(\varphi^{n+1}, \varphi^n) - \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} \right) + \int_\Omega M(\varphi^n) |\nabla\mu^{n+1}|^2 . \end{aligned}$$

Rearranging yields the discrete energy law (4.30). □

As discussed in Section 4.2, ND_{pot}^{n+1} represents numerical dissipation caused by the discretization of the derivative of the double-well potential. Depending on the approximation considered for $f(\varphi^{n+1}, \varphi^n)$, we obtain different numerical schemes with different discrete energy laws.

Theorem 4.11. *Assuming suitable boundary conditions, see Remark 3.2, scheme (4.29) satisfies*

$$\int_\Omega \varphi^{n+1} = \int_\Omega \varphi^0 \qquad \forall n \in \mathbb{N} .$$

Thus, it is mass-conservative.

Proof. Multiplying equation (4.29a) by the smooth test function ψ, integrating over the computational domain Ω and applying integration by parts, we obtain

$$\begin{aligned} \int_\Omega \frac{\varphi^{n+1} - \varphi^n}{\Delta t} \psi &= \int_\Omega \nabla \cdot \left(M(\varphi^n) \nabla\mu^{n+1} \right) \psi \\ &= - \int_\Omega M(\varphi^n) \nabla\mu^{n+1} \cdot \nabla\psi + \int_{\partial\Omega} M(\varphi^n) \nabla\mu^{n+1} \psi \cdot \boldsymbol{n} \\ &= - \int_\Omega M(\varphi^n) \nabla\mu^{n+1} \cdot \nabla\psi + \int_{\partial\Omega} M(\varphi^n) \frac{\partial\mu^{n+1}}{\partial \boldsymbol{n}} \psi . \end{aligned}$$

Then, setting $\psi = 1$ and assuming suitable boundary conditions, see Remark 3.2, we have

$$\int_\Omega \frac{\varphi^{n+1} - \varphi^n}{\Delta t} = 0 \qquad \Longleftrightarrow \qquad \int_\Omega \varphi^{n+1} = \int_\Omega \varphi^n .$$

Thus, it follows by induction that

$$\int_\Omega \varphi^{n+1} = \int_\Omega \varphi^0 \qquad \forall n \in \mathbb{N} .$$

□

4.3.2 Crank-Nicolson type scheme

An improved thermodynamically consistent scheme for the Cahn-Hilliard equation is achieved using a midpoint approximation for the fourth derivative of φ, which completely eliminates the L^2-Norm

$$\frac{\lambda}{2\,\Delta t}\|\nabla\varphi^{n+1}-\nabla\varphi^{n}\|^2_{L^2(\Omega)}$$

in the resulting discrete energy law, see Theorem 4.12 and, e.g., *Guillén-González and Tierra* [37]. We consider a uniform partition of the time interval $[0,T]$ with a constant time step Δt. Given φ^n from the previous time step, we compute φ^{n+1} such that

$$\begin{aligned}\frac{\varphi^{n+1}-\varphi^{n}}{\Delta t}-\nabla\cdot\left(M(\varphi^{n})\nabla\mu^{n+\frac{1}{2}}\right)&=0\,,\\ \mu^{n+\frac{1}{2}}&=-\lambda\Delta\varphi^{n+\frac{1}{2}}+f(\varphi^{n+1},\varphi^{n})\,,\end{aligned}\tag{4.33}$$

where

$$\varphi^{n+\frac{1}{2}}:=\frac{\varphi^{n+1}+\varphi^{n}}{2}$$

is a midpoint approximation in time, i.e., scheme (4.33) is of Crank-Nicolson [21] type.

Theorem 4.12. *Let $f(\varphi^{n+1},\varphi^{n})$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.33) is linear and, assuming suitable boundary conditions, see Remark 3.2, it satisfies the following discrete version of energy law (3.8)*

$$\frac{E_{mix}(\varphi^{n+1})-E_{mix}(\varphi^{n})}{\Delta t}=-\int_\Omega M(\varphi^{n})|\nabla\mu^{n+\frac{1}{2}}|^2-\mathrm{ND}^{n+1}_{pot}\,,\tag{4.34}$$

where

$$\mathrm{ND}^{n+1}_{pot}=\int_\Omega f(\varphi^{n+1},\varphi^{n})\frac{\varphi^{n+1}-\varphi^{n}}{\Delta t}-\int_\Omega\frac{F(\varphi^{n+1})-F(\varphi^{n})}{\Delta t}\,.$$

Proof. It is clear that the proposed scheme is linear if $f(\varphi^{n+1},\varphi^{n})$ is linear. Multiplying equation (4.33) by $\mu^{n+\frac{1}{2}}$, integrating over the computational domain Ω and applying integration by parts, we obtain

$$\begin{aligned}0&=\int_\Omega\frac{\varphi^{n+1}-\varphi^{n}}{\Delta t}\mu^{n+\frac{1}{2}}-\int_\Omega\nabla\cdot\left(M(\varphi^{n})\nabla\mu^{n+\frac{1}{2}}\right)\mu^{n+\frac{1}{2}}\\&=\int_\Omega\frac{\varphi^{n+1}-\varphi^{n}}{\Delta t}\left(-\lambda\Delta\varphi^{n+\frac{1}{2}}+f(\varphi^{n+1},\varphi^{n})\right)+\int_\Omega M(\varphi^{n})\nabla\mu^{n+\frac{1}{2}}\cdot\nabla\mu^{n+\frac{1}{2}}\\&\quad-\int_{\partial\Omega}M(\varphi^{n})\nabla\mu^{n+\frac{1}{2}}\mu^{n+\frac{1}{2}}\cdot\boldsymbol{n}\\&=\int_\Omega\frac{\lambda}{\Delta t}\nabla(\varphi^{n+1}-\varphi^{n})\cdot\nabla\varphi^{n+\frac{1}{2}}-\int_{\partial\Omega}\frac{\varphi^{n+1}-\varphi^{n}}{\Delta t}\lambda\frac{\partial\varphi^{n+\frac{1}{2}}}{\partial\boldsymbol{n}}+\int_\Omega\frac{\varphi^{n+1}-\varphi^{n}}{\Delta t}f(\varphi^{n+1},\varphi^{n})\\&\quad+\int_\Omega M(\varphi^{n})|\nabla\mu^{n+\frac{1}{2}}|^2-\int_{\partial\Omega}M(\varphi^{n})\frac{\partial\mu^{n+\frac{1}{2}}}{\partial\boldsymbol{n}}\mu^{n+\frac{1}{2}}\,.\end{aligned}\tag{4.35}$$

Assuming suitable boundary conditions, see Remark 3.2, the boundary integral vanishes, and since

$$\nabla(\varphi^{n+1}-\varphi^{n})\cdot\nabla\varphi^{n+\frac{1}{2}}=(\nabla\varphi^{n+1}-\nabla\varphi^{n})\cdot\frac{1}{2}(\nabla\varphi^{n+1}+\nabla\varphi^{n})=\frac{1}{2}\left(|\nabla\varphi^{n+1}|^2-|\nabla\varphi^{n}|^2\right)\,,$$

equation (4.35) becomes

$$0 = \int_\Omega \frac{\lambda}{2\Delta t}\left(|\nabla\varphi^{n+1}|^2 - |\nabla\varphi^n|^2\right) + \int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t} f(\varphi^{n+1},\varphi^n) + \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 . \tag{4.36}$$

Further, since the discrete mixing energy at time t_n reads

$$E_{mix}(\varphi^n) = \left(\int_\Omega \frac{\lambda}{2}|\nabla\varphi^n|^2 + F(\varphi^n)\right),$$

we sum equation (4.36) with

$$0 = \int_\Omega \left(\frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} - \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t}\right),$$

to get

$$\begin{aligned} 0 =& \int_\Omega \left(\frac{\lambda}{2\Delta t}\left(|\nabla\varphi^{n+1}|^2 - |\nabla\varphi^n|^2\right) + \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t}\right) \\ &+ \int_\Omega \left(\frac{\varphi^{n+1}-\varphi^n}{\Delta t} f(\varphi^{n+1},\varphi^n) - \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t}\right) + \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 \\ =& \frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} \\ &+ \int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t} f(\varphi^{n+1},\varphi^n) - \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t} + \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 . \end{aligned}$$

Rearranging yields the discrete energy law (4.34). □

Theorem 4.13. *Assuming suitable boundary conditions, see Remark 3.2, scheme (4.33) satisfies*

$$\int_\Omega \varphi^{n+1} = \int_\Omega \varphi^0 \qquad \forall n \in \mathbb{N} .$$

Thus, it is mass-conservative.

Proof. Analogous to the proof of Theorem 4.11. □

Remark 4.14. *Note that in general, the explicit and the implicit Euler method have a truncation error of first order in time, since they correspond to the right and left finite difference, respectively, see Subsection 4.8.1 for the derivation. The Crank-Nicolson method corresponds to the central finite difference and is therefore second order accurate in time.*

Note that even though the Crank-Nicolson method is second order accurate in time, scheme (4.33) is of first order in time as long as the mobility $M(\varphi)$ is not constant, since we discretize the latter explicitly in order to have a linear scheme.

4.3.3 Second order scheme

In order to introduce a linear second order numerical scheme for variable mobilities, we use the explicit multistep method by *Adams and Bashforth* [6] to extrapolate $M(\varphi^n)$ by $M(\varphi^{n-\frac{1}{2}})$, where

$$\varphi^{n-\frac{1}{2}} := \frac{3\varphi^n - \varphi^{n-1}}{2}, \tag{4.37}$$

yielding a two-step numerical scheme, i.e., a scheme using the information from two previous time steps.

Theorem 4.15. *The explicit extrapolation* (4.37) *is of second order in time.*

Proof. Let us consider the two central approximations

$$\varphi^{n+\frac{1}{2}} = \frac{\varphi^{n+1} + \varphi^n}{2}, \tag{4.38}$$

like in the Crank-Nicolson method, and similarly

$$\varphi^n = \frac{\varphi^{n+1} + \varphi^{n-1}}{2}, \tag{4.39}$$

which both yield second order in time schemes, since they are midpoint approximations that correspond to central finite differences, see Subsection 4.8.1 for the derivation. Rewriting (4.39) to

$$\varphi^{n+1} = 2\varphi^n - \varphi^{n-1}$$

and substituting it into (4.38) reads

$$\varphi^{n+\frac{1}{2}} = \frac{2\varphi^n - \varphi^{n-1} + \varphi^n}{2} = \frac{3\varphi^n - \varphi^{n-1}}{2} = \varphi^{n-\frac{1}{2}}.$$

□

The proposed two-step numerical scheme reads

$$\begin{aligned} \frac{\varphi^{n+1} - \varphi^n}{\Delta t} - \nabla \cdot \left(M(\varphi^{n-\frac{1}{2}}) \nabla \mu^{n+\frac{1}{2}} \right) &= 0, \\ \mu^{n+\frac{1}{2}} &= -\lambda \Delta \varphi^{n+\frac{1}{2}} + f(\varphi^{n+1}, \varphi^n). \end{aligned} \tag{4.40}$$

Theorem 4.16. *Let $f(\varphi^{n+1}, \varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme* (4.40) *is linear and, assuming suitable boundary conditions, see Remark 3.2, it is mass-conservative and satisfies the following discrete version of energy law* (3.8)

$$\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} = -\int_\Omega M(\varphi^{n-\frac{1}{2}}) |\nabla \mu^{n+\frac{1}{2}}|^2 - \mathrm{ND}_{pot}^{n+1}, \tag{4.41}$$

where

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega f(\varphi^{n+1}, \varphi^n) \frac{\varphi^{n+1} - \varphi^n}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t}.$$

Proof. It is clear that the proposed scheme is linear if $f(\varphi^{n+1}, \varphi^n)$ is linear. The mass conservation is calculated analogously to the proof of Theorem 4.11 and the discrete energy law analogously to the proof of Theorem 4.12. □

In order to compute φ^1 from φ^0, a second order one-step nonlinear scheme can be considered. We overcome this by setting $\varphi^{-1} := \varphi^0$, i.e., we solve the first order scheme (4.33) in the first time step. As long as the initial data is sufficiently smooth, the influence to the final solution φ^N is usually negligible for $T \gg 0$, see the experimental convergence results presented in Chapter 5.

Table 4.2: Summary of the schemes for the Cahn-Hilliard equation

	linear	energy-stable	mass-cons.	truncation error	Δt
explicit Euler (4.2)	✓	-	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^4)$
implicit Euler	-	-	✓	$\mathcal{O}(\Delta t)$	-
implicit (4.29)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
Crank-Nicolson (4.33)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
second order (4.40)	✓	✓	✓	$\mathcal{O}\left((\Delta t)^2\right)$	$\mathcal{O}(h^2)$

Remark 4.17. *Summarized, see Table 4.2, the second order Crank-Nicolson type scheme (4.40) delivers the best properties while fulfilling all our necessities and is therefore our method of choice for solving the Cahn-Hilliard equation numerically. Note that the two Crank-Nicolson type schemes are not only energy-stable but conserve the continuous energy laws up to the dissipation term* ND_{pot} *resulting from the linearized potential.*

Other common thermodynamically consistent approaches to solve the Cahn-Hilliard equation numerically are BDF schemes, see, e.g., *Yan, Chen, Wang, Wise* [73]. Their main disadvantage, compared to the above introduced methods, is that they require modifications to the free energy functional E_{mix} in order to be energy-stable.

4.4 Schemes for the Cahn-Hilliard-Navier-Stokes model

Adding the Navier-Stokes equations introduces additional challenges to the derivation of suitable linear numerical schemes. One being the added advection term $\boldsymbol{u} \cdot \nabla\varphi$ in the Cahn-Hilliard equation, which couples it to the Navier-Stokes equations. We have to ensure that it cancels out in the discrete energy law with the φ-coupling term in the Navier-Stokes equations, analogously to the continuous case (3.24) and to the explicit Euler scheme (4.3) for the full model. Another one being the (modified) pressure p, which again acts as a Lagrange multiplier due to the incompressibility condition $\nabla \cdot \boldsymbol{u} = 0$.

4.4.1 Mixed scheme

An often used approach for solving the Navier-Stokes equations numerically is the implicit Euler method. We combine it with the Crank-Nicolson type scheme (4.33) for the Cahn-Hilliard

equation to get a coupled scheme to solve the Cahn-Hilliard-Navier-Stokes model (3.24). We consider a uniform partition of the time interval $[0,T]$ with a constant time step Δt. Given $(\varphi^n, \boldsymbol{u}^n)$ from the previous time step, we compute $(\varphi^{n+1}, \boldsymbol{u}^{n+1}, p^{n+1})$ such that

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^{n+1}\cdot\nabla\varphi^n - \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right) = 0\,, \tag{4.42a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,$$

$$\rho\left(\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+1}\right) - \nabla\cdot\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1} + (\nabla\boldsymbol{u}^{n+1})^T\right]\right\} + \nabla p^{n+1} - \mu^{n+\frac{1}{2}}\nabla\varphi^n = \boldsymbol{0}\,, \tag{4.42b}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,. \tag{4.42c}$$

Note that the numerical scheme (4.42) is fully coupled through the implicit coupling terms $\boldsymbol{u}^{n+1}\cdot\nabla\varphi^n$ and $\mu^{n+\frac{1}{2}}\nabla\varphi^n$ and the implicit discretization ∇p^{n+1}. Thus, one has to solve all equations simultaneously, which may cause increased computational effort in comparison to solving one equation after the other.

Theorem 4.18. *Let $f(\varphi^{n+1},\varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.42) is linear and, assuming suitable boundary conditions, see Remark 3.7, it satisfies the following discrete version of energy law (3.28)*

$$\begin{aligned}\frac{E_{total}(\varphi^{n+1},\boldsymbol{u}^{n+1}) - E_{total}(\varphi^n,\boldsymbol{u}^n)}{\Delta t} = &-\mathrm{ND}_{pot}^{n+1} - \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 \\ &- \frac{\rho}{2\,\Delta t}\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|^2_{L^2(\Omega)} - \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,,\end{aligned} \tag{4.43}$$

where like above

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega f(\varphi^{n+1},\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t}\,.$$

Proof. It is clear that the proposed scheme is linear. Analogously to the proof of Theorem 4.12, we multiply (4.42a) by $\mu^{n+\frac{1}{2}}$ and integrate over the computational domain Ω. Assuming suitable boundary conditions, see Remark 3.7, we get the same equation plus one integral from the advective coupling term

$$\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} + \mathrm{ND}_{pot}^{n+1} + \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 + \int_\Omega \boldsymbol{u}^{n+1}\cdot\nabla\varphi^n\mu^{n+\frac{1}{2}} = 0\,. \tag{4.44}$$

Taking the inner product of (4.42b) and $\boldsymbol{u}^{n+1}$ and integrating over Ω yields

$$
\begin{aligned}
0 &= \int_\Omega \rho \frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t}\cdot\boldsymbol{u}^{n+1} + \int_\Omega \rho(\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+1}\cdot\boldsymbol{u}^{n+1} \\
&\quad - \int_\Omega \left(\nabla\cdot\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1}+(\nabla\boldsymbol{u}^{n+1})^T\right]\right\}\right)\cdot\boldsymbol{u}^{n+1} \\
&\quad + \int_\Omega \nabla p^{n+1}\cdot\boldsymbol{u}^{n+1} - \int_\Omega \mu^{n+\frac{1}{2}}\nabla\varphi^n\cdot\boldsymbol{u}^{n+1} \\
&= \int_\Omega \frac{\rho}{\Delta t}\left(|\boldsymbol{u}^{n+1}|^2 - \boldsymbol{u}^n\cdot\boldsymbol{u}^{n+1}\right) + \int_\Omega \rho(\boldsymbol{u}^n\cdot\nabla)\frac{1}{2}|\boldsymbol{u}^{n+1}|^2 \\
&\quad + \int_\Omega \left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1}+(\nabla\boldsymbol{u}^{n+1})^T\right]\right\} : \nabla\boldsymbol{u}^{n+1} - \int_{\partial\Omega}\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1}+(\nabla\boldsymbol{u}^{n+1})^T\right]\right\}\boldsymbol{u}^{n+1}\cdot\boldsymbol{n} \\
&\quad - \int_\Omega p^{n+1}(\nabla\cdot\boldsymbol{u}^{n+1}) + \int_{\partial\Omega} p^{n+1}\boldsymbol{u}^{n+1}\cdot\boldsymbol{n} - \int_\Omega \mu^{n+\frac{1}{2}}\nabla\varphi^n\cdot\boldsymbol{u}^{n+1} \\
&= \int_\Omega \frac{\rho}{2\,\Delta t}\left(|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n|^2 + |\boldsymbol{u}^{n+1}|^2 - |\boldsymbol{u}^n|^2\right) - \int_\Omega \rho(\nabla\cdot\boldsymbol{u}^n)\frac{1}{2}|\boldsymbol{u}^{n+1}|^2 + \int_{\partial\Omega}\rho(\boldsymbol{u}^n\cdot\boldsymbol{n})\frac{1}{2}|\boldsymbol{u}^{n+1}|^2 \\
&\quad + \int_\Omega \eta(\varphi^n)\left[|\nabla\boldsymbol{u}^{n+1}|^2 + \operatorname{tr}\left((\nabla\boldsymbol{u}^{n+1})^2\right)\right] - \int_{\partial\Omega}\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1}+(\nabla\boldsymbol{u}^{n+1})^T\right]\right\}\boldsymbol{u}^{n+1}\cdot\boldsymbol{n} \\
&\quad - \int_\Omega p^{n+1}(\nabla\cdot\boldsymbol{u}^{n+1}) + \int_{\partial\Omega} p^{n+1}\boldsymbol{u}^{n+1}\cdot\boldsymbol{n} - \int_\Omega \boldsymbol{u}^{n+1}\cdot\nabla\varphi^n\mu^{n+\frac{1}{2}}\,.
\end{aligned}
$$

Using the incompressibility and assuming suitable boundary conditions for $\boldsymbol{u}$, see Remark 3.7, we can eliminate a few integrals, including all boundary integrals. Further, since the discrete kinetic energy at time t_n reads

$$
E_{kin}(\boldsymbol{u}^n) = \int_\Omega \frac{1}{2}\rho|\boldsymbol{u}^n|^2\,,
$$

and since

$$
|\nabla\boldsymbol{u}^{n+1}|^2 + \operatorname{tr}\left((\nabla\boldsymbol{u}^{n+1})^2\right) = \frac{1}{2}\sum_{i,j=1}^{d}\left(\frac{\partial u_i^{n+1}}{\partial x_j}+\frac{\partial u_j^{n+1}}{\partial x_i}\right)^2 = \frac{1}{2}|\nabla\boldsymbol{u}^{n+1}+(\nabla\boldsymbol{u}^{n+1})^T|^2 = 2|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,,
$$

we can rewrite the above relation to

$$
\frac{E_{kin}(\boldsymbol{u}^{n+1})-E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \frac{\rho}{2\,\Delta t}\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|^2_{L^2(\Omega)} + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2 - \int_\Omega \boldsymbol{u}^{n+1}\cdot\nabla\varphi^n\mu^{n+\frac{1}{2}} = 0\,.
$$

Finally, summing this discrete energy law for the kinetic energy and the discrete energy law for the mixing energy (4.44), the integrals of the coupling terms cancel out and we have the desired energy law (4.43). □

Theorem 4.19. *Assuming suitable boundary conditions, see Remark 3.7, scheme (4.42) satisfies*

$$
\int_\Omega \varphi^{n+1} = \int_\Omega \varphi^0 \qquad \forall n\in\mathbb{N}\,.
$$

Thus, it is mass-conservative.

Proof. Multiplying equation (4.42a) by the smooth test function ψ, integrating over the computational domain Ω and applying integration by parts, we obtain

$$\begin{aligned}\int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t}\psi &= -\int_\Omega \boldsymbol{u}^{n+1}\cdot\nabla\varphi^n\psi + \int_\Omega \nabla\cdot\left(M(\varphi^n)\,\nabla\mu^{n+\frac{1}{2}}\right)\psi \\ &= \int_\Omega \nabla\cdot(\boldsymbol{u}^{n+1}\,\psi)\varphi^n - \int_{\partial\Omega}\boldsymbol{u}^{n+1}\varphi^n\psi\cdot\boldsymbol{n} - \int_\Omega M(\varphi^n)\,\nabla\mu^{n+1}\cdot\nabla\psi + \int_{\partial\Omega} M(\varphi^n)\frac{\partial\mu^{n+1}}{\partial\boldsymbol{n}}\psi\,.\end{aligned}$$

Then, setting $\psi = 1$, using the incompressibility and assuming suitable boundary conditions, see Remark 3.7, we have

$$\int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t} = 0 \qquad \Longleftrightarrow \qquad \int_\Omega \varphi^{n+1} = \int_\Omega \varphi^n\,.$$

Thus, the theorem follows by induction. □

4.4.2 Crank-Nicolson type scheme

Note that the norm

$$\frac{1}{2\,\Delta t}\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|^2_{L^2(\Omega)}$$

in the discrete energy law (4.43) can be eliminated analogously to the improved scheme for the Cahn-Hilliard equation (4.33). Thus, by using the midpoint approximation

$$\boldsymbol{u}^{n+\frac{1}{2}} = \frac{\boldsymbol{u}^{n+1}+\boldsymbol{u}^n}{2}\,,$$

we have

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla\varphi^n - \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right) = 0\,, \tag{4.45a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,$$

$$\rho\left(\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+\frac{1}{2}}\right) - \nabla\cdot\left[2\eta(\varphi^n)\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})\right] + \nabla p^{n+1} - \mu^{n+\frac{1}{2}}\nabla\varphi^n = \boldsymbol{0}\,, \tag{4.45b}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,. \tag{4.45c}$$

Theorem 4.20. *Let $f(\varphi^{n+1},\varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.45) is linear and, assuming suitable boundary conditions, see Remark 3.7, it is mass-conservative and satisfies the following discrete version of energy law (3.28)*

$$\begin{aligned}\frac{E_{total}(\varphi^{n+1},\boldsymbol{u}^{n+1}) - E_{total}(\varphi^n,\boldsymbol{u}^n)}{\Delta t} &= -\,\mathrm{ND}^{n+1}_{pot} - \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 \\ &\quad - \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2\,,\end{aligned} \tag{4.46}$$

where like above

$$\mathrm{ND}^{n+1}_{pot} = \int_\Omega f(\varphi^{n+1},\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t}\,.$$

Proof. It is clear that the proposed scheme is linear if $f(\varphi^{n+1}, \varphi^n)$ is linear. The mass conservation is calculated analogously to the proof of Theorem 4.19. Then, analogously to the proof of Theorem 4.12, we multiply (4.42a) by $\mu^{n+\frac{1}{2}}$ and integrate over the computational domain Ω. Assuming suitable boundary conditions, see Remark 3.7, we get the same equation plus one term from the coupling, reading

$$\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} + \mathrm{ND}_{pot}^{n+1} + \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 + \int_\Omega \boldsymbol{u}^{n+\frac{1}{2}} \cdot \nabla\varphi^n \mu^{n+\frac{1}{2}} = 0\,. \tag{4.47}$$

Further, analogously to the proof of Theorem 4.18, taking the inner product of (4.42b) and $\boldsymbol{u}^{n+\frac{1}{2}}$ and integrating over Ω yields

$$\begin{aligned}
0 = &\int_\Omega \rho \frac{\boldsymbol{u}^{n+1} - \boldsymbol{u}^n}{\Delta t} \cdot \boldsymbol{u}^{n+\frac{1}{2}} + \int_\Omega \rho(\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^{n+\frac{1}{2}} \cdot \boldsymbol{u}^{n+\frac{1}{2}} \\
&- \int_\Omega \left(\nabla \cdot \left\{\eta(\varphi^n)\left[\nabla \boldsymbol{u}^{n+\frac{1}{2}} + (\nabla \boldsymbol{u}^{n+\frac{1}{2}})^T\right]\right\}\right) \cdot \boldsymbol{u}^{n+\frac{1}{2}} \\
&+ \int_\Omega \nabla p^{n+1} \cdot \boldsymbol{u}^{n+\frac{1}{2}} - \int_\Omega \mu^{n+\frac{1}{2}} \nabla\varphi^n \cdot \boldsymbol{u}^{n+\frac{1}{2}} \\
= &\int_\Omega \frac{\rho}{2\Delta t}\left(|\boldsymbol{u}^{n+1}|^2 - |\boldsymbol{u}^n|^2\right) + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2 - \int_\Omega \boldsymbol{u}^{n+\frac{1}{2}} \cdot \nabla\varphi^n \mu^{n+\frac{1}{2}} \\
= &\frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2 - \int_\Omega \boldsymbol{u}^{n+\frac{1}{2}} \cdot \nabla\varphi^n \mu^{n+\frac{1}{2}}\,.
\end{aligned}$$

Summing the mixing and the kinetic energy law, the integrals of the coupling terms cancel out and we get the desired energy law (4.46). □

4.4.3 Second order scheme

Furthermore, scheme (4.45) can be modified to a second order in time two-step scheme by using the second order extrapolation

$$\omega^{n-\frac{1}{2}} = \frac{3\omega^n - \omega^{n-1}}{2}, \quad \omega \in \{\varphi, \boldsymbol{u}\},$$

for the explicit terms, yielding

$$\frac{\varphi^{n+1} - \varphi^n}{\Delta t} + \boldsymbol{u}^{n+\frac{1}{2}} \cdot \nabla\varphi^{n-\frac{1}{2}} - \nabla \cdot \left(M(\varphi^{n-\frac{1}{2}})\nabla\mu^{n+\frac{1}{2}}\right) = 0\,, \tag{4.48a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1}, \varphi^n)\,,$$

$$\rho\left(\frac{\boldsymbol{u}^{n+1} - \boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^{n-\frac{1}{2}} \cdot \nabla)\boldsymbol{u}^{n+\frac{1}{2}}\right) - \nabla \cdot \left[2\eta(\varphi^{n-\frac{1}{2}})\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})\right] + \nabla p^{n+1} - \mu^{n+\frac{1}{2}}\nabla\varphi^{n-\frac{1}{2}} = \boldsymbol{0}\,, \tag{4.48b}$$

$$\nabla \cdot \boldsymbol{u}^{n+1} = 0\,. \tag{4.48c}$$

Theorem 4.21. *Let $f(\varphi^{n+1}, \varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.48) is linear and, assuming suitable boundary*

conditions, see Remark 3.7, it is mass-conservative and satisfies the following discrete version of energy law (3.28)

$$\frac{E_{total}(\varphi^{n+1},\boldsymbol{u}^{n+1})-E_{total}(\varphi^{n},\boldsymbol{u}^{n})}{\Delta t} = -\mathrm{ND}_{pot}^{n+1} - \int_\Omega M(\varphi^{n-\frac{1}{2}})|\nabla\mu^{n+\frac{1}{2}}|^2 - \int_\Omega 2\eta(\varphi^{n-\frac{1}{2}})|\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2\,. \tag{4.49}$$

Proof. It is clear that the proposed scheme is linear if $f(\varphi^{n+1},\varphi^n)$ is linear. The mass conservation is calculated analogously to the proof of Theorem 4.19 and the energy law analogously to the proof of Theorem 4.20. □

4.4.4 Splitting schemes

Due to the increased computational effort of solving coupled systems, we want to decouple them if possible while conserving their thermodynamic consistency. Starting from scheme (4.42), we have to modify one of the coupling terms in order to decouple the calculation of (4.42a) and (4.42b). A suitable modification is given by substituting $\boldsymbol{u}^{n+1}$ in the advection term $\boldsymbol{u}^{n+1}\cdot\nabla\varphi$ of (4.42a) by

$$\boldsymbol{u}^* = \boldsymbol{u}^n + \frac{\Delta t}{\rho}\mu^{n+\frac{1}{2}}\nabla\varphi^n\,, \tag{4.50}$$

i.e., by its explicit discretization minus the φ-coupling term of the Navier-Stokes equation times the time step size. The resulting splitting scheme can be written

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^*\cdot\nabla\varphi^n - \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right) = 0\,, \tag{4.51a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,, \qquad \boldsymbol{u}^* = \boldsymbol{u}^n + \frac{\Delta t}{\rho}\mu^{n+\frac{1}{2}}\nabla\varphi^n\,,$$

$$\rho\left(\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^*}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+1}\right) - \nabla\cdot\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1} + (\nabla\boldsymbol{u}^{n+1})^T\right]\right\} + \nabla p^{n+1} = \boldsymbol{0}\,, \tag{4.51b}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,, \tag{4.51c}$$

where (4.51b) with (4.50) is identical to (4.42b).

Theorem 4.22. *Let $f(\varphi^{n+1},\varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.51) is linear and, assuming suitable boundary conditions, see Remark 3.7, it satisfies the following discrete version of energy law (3.28)*

$$\begin{aligned}\frac{E_{total}(\varphi^{n+1},\boldsymbol{u}^{n+1})-E_{total}(\varphi^{n},\boldsymbol{u}^{n})}{\Delta t} &= -\mathrm{ND}_{pot}^{n+1} - \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 \\ &\quad - \frac{\rho}{2\,\Delta t}\left(\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^*-\boldsymbol{u}^n\|^2_{L^2(\Omega)}\right) - \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,,\end{aligned} \tag{4.52}$$

where like above

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega f(\varphi^{n+1},\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t}\,.$$

Proof. It is clear that the proposed scheme is linear. Analogously to the proof of Theorem 4.12, we multiply (4.51a) by $\mu^{n+\frac{1}{2}}$ and integrate over the computational domain Ω. Assuming suitable boundary conditions, see Remark 3.7, we get the same energy law plus one term from the coupling, reading

$$\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} + \mathrm{ND}_{pot}^{n+1} + \int_\Omega M(\varphi^n)|\nabla\mu^{n+\frac{1}{2}}|^2 + \int_\Omega \boldsymbol{u}^* \cdot \nabla\varphi^n \, \mu^{n+\frac{1}{2}} = 0\,. \tag{4.53}$$

Then, analogously to the proof of Theorem 4.18, taking the inner product of (4.51b) and $\boldsymbol{u}^{n+1}$ and integrating over Ω yields

$$\begin{aligned} 0 &= \int_\Omega \rho \frac{\boldsymbol{u}^{n+1} - \boldsymbol{u}^*}{\Delta t} \cdot \boldsymbol{u}^{n+1} + \int_\Omega \rho(\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^{n+1} \cdot \boldsymbol{u}^{n+1} \\ &\quad - \int_\Omega \left(\nabla \cdot \left\{ \eta(\varphi^n) \left[\nabla \boldsymbol{u}^{n+1} + (\nabla \boldsymbol{u}^{n+1})^T \right] \right\} \right) \cdot \boldsymbol{u}^{n+1} + \int_\Omega \nabla p^{n+1} \cdot \boldsymbol{u}^{n+1} \\ &= \frac{\rho}{2\,\Delta t} \left(\|\boldsymbol{u}^{n+1}\|_{L^2(\Omega)}^2 - \|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|_{L^2(\Omega)}^2 \right) + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,. \end{aligned} \tag{4.54}$$

Taking the inner product of relation (4.50) and $\boldsymbol{u}^*$ and integrating over Ω yields

$$\begin{aligned} \|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 &= \int_\Omega \boldsymbol{u}^n \cdot \boldsymbol{u}^* + \int_\Omega \frac{\Delta t}{\rho} \mu^{n+\frac{1}{2}} \nabla\varphi^n \cdot \boldsymbol{u}^* \\ &= \int_\Omega \frac{1}{2} \left(-|\boldsymbol{u}^* - \boldsymbol{u}^n|^2 + |\boldsymbol{u}^*|^2 + |\boldsymbol{u}^n|^2 \right) + \int_\Omega \frac{\Delta t}{\rho} \boldsymbol{u}^* \cdot \nabla\varphi^n \, \mu^{n+\frac{1}{2}} \\ &= \frac{1}{2} \left(- \|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^n\|_{L^2(\Omega)}^2 \right) + \int_\Omega \frac{\Delta t}{\rho} \boldsymbol{u}^* \cdot \nabla\varphi^n \, \mu^{n+\frac{1}{2}}\,. \end{aligned}$$

Multiplying the above equation by two and subtracting $\|\boldsymbol{u}^*\|_{L^2(\Omega)}^2$ yields

$$\|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 = -\|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^n\|_{L^2(\Omega)}^2 + \int_\Omega \frac{2\Delta t}{\rho} \boldsymbol{u}^* \cdot \nabla\varphi^n \, \mu^{n+\frac{1}{2}}\,. \tag{4.55}$$

Inserting relation (4.55) in (4.54) reads

$$\begin{aligned} 0 &= \frac{\rho}{2\,\Delta t} \left(\|\boldsymbol{u}^{n+1}\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 - \|\boldsymbol{u}^n\|_{L^2(\Omega)}^2 - \int_\Omega \frac{2\Delta t}{\rho} \boldsymbol{u}^* \cdot \nabla\varphi^n \, \mu^{n+\frac{1}{2}} + \|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|_{L^2(\Omega)}^2 \right) \\ &\quad + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2 \\ &= \frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \frac{\rho}{2\,\Delta t} \left(\|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 \right) - \int_\Omega \boldsymbol{u}^* \cdot \nabla\varphi^n \, \mu^{n+\frac{1}{2}} \\ &\quad + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,. \end{aligned}$$

Summed up with the discrete mixing energy law (4.53), the integrals of the coupling terms cancel out and we obtain the desired energy law (4.52). □

Theorem 4.23. *Assuming suitable boundary conditions, see Remark 3.7, scheme (4.51) satisfies*

$$\int_\Omega \varphi^{n+1} = \int_\Omega \varphi^n \qquad \textit{if and only if} \qquad \int_\Omega \mu^{n+\frac{1}{2}} |\nabla\varphi^n|^2 = 0\,.$$

Therefore, it is in general not mass-conservative for φ.

Proof. Multiplying equation (4.51a) by the smooth test function ψ, integrating over the computational domain Ω and applying integration by parts, we obtain

$$\begin{aligned}\int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t}\psi &= -\int_\Omega \boldsymbol{u}^* \cdot \nabla\varphi^n\psi + \int_\Omega \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right)\psi \\ &= -\int_\Omega \boldsymbol{u}^n \cdot \nabla\varphi^n\psi - \int_\Omega \frac{\Delta t}{\rho}\mu^{n+\frac{1}{2}}\nabla\varphi^n\cdot\nabla\varphi^n\psi + \int_\Omega \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right)\psi \\ &= \int_\Omega \nabla\cdot(\boldsymbol{u}^n\psi)\,\varphi^n - \int_{\partial\Omega}\boldsymbol{u}^n\varphi^n\psi\cdot\boldsymbol{n} - \int_\Omega \frac{\Delta t}{\rho}\mu^{n+\frac{1}{2}}|\nabla\varphi^n|^2\psi \\ &\quad - \int_\Omega M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\cdot\nabla\psi + \int_{\partial\Omega} M(\varphi^n)\frac{\partial\mu^{n+\frac{1}{2}}}{\partial\boldsymbol{n}}\psi\,.\end{aligned}$$

Then, setting $\psi = 1$, using the incompressibility and assuming suitable boundary conditions, see Remark 3.7, we have

$$\int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t} = -\int_\Omega \frac{\Delta t}{\rho}\mu^{n+\frac{1}{2}}|\nabla\varphi^n|^2\,.$$

□

In order to construct a mass-conservative splitting scheme, we propose to rewrite the coupled scheme (4.42) at first. Using the incompressibility $\nabla\cdot\boldsymbol{u}^{n+1} = 0$, we can rewrite the advective Cahn-Hilliard equation (4.42a) into its conservative form analogously to Remark 3.15. For consistency, we also rewrite the φ-coupling term in the Navier-Stokes equation (4.42b) appropriately. The resulting scheme reads

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \nabla\cdot(\boldsymbol{u}^{n+1}\varphi^n) - \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right) = 0\,, \tag{4.56a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,$$

$$\rho\left(\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+1}\right) - \nabla\cdot\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1} + (\nabla\boldsymbol{u}^{n+1})^T\right]\right\} + \nabla p^{n+1} + \varphi^n\nabla\mu^{n+\frac{1}{2}} = \boldsymbol{0}\,, \tag{4.56b}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,. \tag{4.56c}$$

Again, we decouple the calculation of scheme (4.56) by substituting $\boldsymbol{u}^{n+1}$ in the advection term of (4.56a) by its explicit discretization minus the φ-coupling term of the Navier-Stokes equation times the time step size, now reading

$$\boldsymbol{u}^* = \boldsymbol{u}^n - \frac{\Delta t}{\rho}\varphi^n\nabla\mu^{n+\frac{1}{2}}\,. \tag{4.57}$$

The resulting splitting scheme reads

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \nabla\cdot(\boldsymbol{u}^*\varphi^n) - \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right) = 0\,, \tag{4.58a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,, \qquad \boldsymbol{u}^* = \boldsymbol{u}^n - \frac{\Delta t}{\rho}\varphi^n\nabla\mu^{n+\frac{1}{2}}\,,$$

$$\rho\left(\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^*}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+1}\right) - \nabla\cdot\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1} + (\nabla\boldsymbol{u}^{n+1})^T\right]\right\} + \nabla p^{n+1} = \boldsymbol{0}\,, \tag{4.58b}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,. \tag{4.58c}$$

Theorem 4.24. *Let $f(\varphi^{n+1}, \varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.58) is linear and, assuming suitable boundary conditions, see Remark 3.7, it satisfies the discrete energy law (4.52) with $\boldsymbol{u}^*$ given by relation (4.57).*

Proof. It is clear that the proposed scheme is linear. Analogously to the proof of Theorem 4.22, we multiply (4.58a) by $\mu^{n+\frac{1}{2}}$ and integrate over the computational domain Ω. Then, assuming suitable boundary conditions, see Remark 3.7, we get

$$\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} + \mathrm{ND}_{pot}^{n+1} + \int_\Omega M(\varphi^n)|\nabla \mu^{n+\frac{1}{2}}|^2 + \int_\Omega \nabla \cdot (\boldsymbol{u}^* \varphi^n)\, \mu^{n+\frac{1}{2}} = 0\,. \tag{4.59}$$

Taking the inner product of (4.58b) and $\boldsymbol{u}^{n+1}$ and integrating over Ω yields, analogously to the proof of Theorem 4.22,

$$0 = \frac{\rho}{2\,\Delta t}\Big(\|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} - \|\boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|^2_{L^2(\Omega)} \Big) + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,. \tag{4.60}$$

Taking the inner product of relation (4.57) and $\boldsymbol{u}^*$, integrating over Ω and applying integration by parts yields

$$\begin{aligned}
\|\boldsymbol{u}^*\|^2_{L^2(\Omega)} &= \int_\Omega \boldsymbol{u}^n \cdot \boldsymbol{u}^* - \int_\Omega \frac{\Delta t}{\rho} \varphi^n \nabla \mu^{n+\frac{1}{2}} \cdot \boldsymbol{u}^* \\
&= \int_\Omega \boldsymbol{u}^n \cdot \boldsymbol{u}^* + \int_\Omega \frac{\Delta t}{\rho} \nabla \cdot (\boldsymbol{u}^* \varphi^n)\, \mu^{n+\frac{1}{2}} - \int_{\partial\Omega} \frac{\Delta t}{\rho} \varphi^n \mu^{n+\frac{1}{2}} \boldsymbol{u}^* \cdot \boldsymbol{n}\,.
\end{aligned}$$

Assuming suitable boundary conditions, see Remark 3.7, the boundary integral vanishes and we get

$$\|\boldsymbol{u}^*\|^2_{L^2(\Omega)} = \frac{1}{2}\Big(- \|\boldsymbol{u}^* - \boldsymbol{u}^n\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^n\|^2_{L^2(\Omega)} \Big) + \int_\Omega \frac{\Delta t}{\rho} \nabla \cdot (\boldsymbol{u}^* \varphi^n)\, \mu^{n+\frac{1}{2}}\,.$$

Multiplying both sides of the equation by two and subtracting $\|\boldsymbol{u}^*\|^2_{L^2(\Omega)}$ yields

$$\|\boldsymbol{u}^*\|^2_{L^2(\Omega)} = -\|\boldsymbol{u}^* - \boldsymbol{u}^n\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^n\|^2_{L^2(\Omega)} + \int_\Omega \frac{2\Delta t}{\rho} \nabla \cdot (\boldsymbol{u}^* \varphi^n)\, \mu^{n+\frac{1}{2}}\,. \tag{4.61}$$

Inserting relation (4.61) in (4.60) yields

$$\begin{aligned}
0 = {} & \frac{\rho}{2\,\Delta t}\Big(\|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|^2_{L^2(\Omega)} - \|\boldsymbol{u}^n\|^2_{L^2(\Omega)} - \int_\Omega \frac{2\Delta t}{\rho} \nabla \cdot (\boldsymbol{u}^* \varphi^n)\, \mu^{n+\frac{1}{2}} + \|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|^2_{L^2(\Omega)} \Big) \\
& + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2 \\
= {} & \frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \frac{\rho}{2\,\Delta t}\Big(\|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|^2_{L^2(\Omega)} \Big) - \int_\Omega \nabla \cdot (\boldsymbol{u}^* \varphi^n)\, \mu^{n+\frac{1}{2}} \\
& + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,.
\end{aligned}$$

Summed with the discrete mixing energy law (4.59), the integrals of the coupling terms cancel out and we have the desired energy law (4.52). □

Theorem 4.25. *Assuming suitable boundary conditions, see Remark 3.7, scheme (4.58) satisfies*

$$\int_\Omega \varphi^{n+1} = \int_\Omega \varphi^0 \qquad \forall n \in \mathbb{N}\,.$$

Thus, it is mass-conservative.

Proof. Multiplying equation (4.58a) by the smooth test function ψ, integrating over the computational domain Ω and applying integration by parts reads

$$\begin{aligned}\int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t}\psi &= -\int_\Omega \nabla\cdot(\boldsymbol{u}^*\varphi^n)\psi + \int_\Omega \nabla\cdot\left(M(\varphi^n)\,\nabla\mu^{n+\frac{1}{2}}\right)\psi \\ &= \int_\Omega \boldsymbol{u}^*\varphi^n\cdot\nabla\psi - \int_{\partial\Omega}\boldsymbol{u}^*\varphi^n\psi\cdot\boldsymbol{n} - \int_\Omega M(\varphi^n)\,\nabla\mu^{n+\frac{1}{2}}\cdot\nabla\psi + \int_{\partial\Omega} M(\varphi^n)\frac{\partial\mu^{n+\frac{1}{2}}}{\partial\boldsymbol{n}}\psi\,.\end{aligned}$$

Then, setting $\psi = 1$ and assuming suitable boundary conditions, see Remark 3.7, we have

$$\int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t} = 0 \qquad \Longleftrightarrow \qquad \int_\Omega \varphi^{n+1} = \int_\Omega \varphi^n\,.$$

Thus, the theorem follows by induction. □

Theorem 4.26. *The splitting method used in scheme (4.58) has a truncation error of first order in time.*

Proof. Inserting relation (4.57) for $\boldsymbol{u}^*$ in the Cahn-Hilliard equation (4.58a) reads

$$\begin{aligned}0 &= \frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \nabla\cdot\left[\left(\boldsymbol{u}^n - \frac{\Delta t}{\rho}\varphi^n\nabla\mu^{n+\frac{1}{2}}\right)\varphi^n\right] - \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right) \\ &= \frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \nabla\cdot(\boldsymbol{u}^n\varphi^n) - \nabla\cdot\left(M(\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right) - \frac{\Delta t}{\rho}\nabla\cdot\left((\varphi^n)^2\nabla\mu^{n+\frac{1}{2}}\right)\,,\end{aligned}$$

where the last term is the splitting error, which holds

$$\Delta t\frac{1}{\rho}\nabla\cdot\left((\varphi^n)^2\nabla\mu^{n+\frac{1}{2}}\right) \in \mathcal{O}(\Delta t)\,.$$

□

4.4.5 Chorin's projection method

To further reduce the computational cost of our mass-conservative splitting scheme (4.58), we propose to use *Chorin's projection method* [17]. This algorithm enables a decoupling of the computation of the velocity and the pressure of the coupled Navier-Stokes equations (4.58b), (4.58c), by using the following pressure correction method

$$\rho\left(\frac{\boldsymbol{u}^\dagger-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^\dagger\right) - \nabla\cdot\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^\dagger + \left(\nabla\boldsymbol{u}^\dagger\right)^T\right]\right\} + \varphi^n\nabla\mu^{n+\frac{1}{2}} = \boldsymbol{0}\,, \tag{4.62a}$$

$$\boldsymbol{u}^\dagger = \boldsymbol{u}^{n+1} + \frac{\Delta t}{\rho}\nabla p^{n+1}\,, \tag{4.62b}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,. \tag{4.62c}$$

The above system is solved as follows:

1. Find $\boldsymbol{u}^\dagger$ by solving equation (4.62a).

2. Applying the divergence to (4.62b) yields

$$\nabla \cdot \boldsymbol{u}^\dagger = \nabla \cdot \boldsymbol{u}^{n+1} + \frac{\Delta t}{\rho} \Delta p^{n+1} .$$

   Consequently, due to the incompressibility condition (4.62c), we find p^{n+1} by solving the *Poisson equation*

$$\Delta p^{n+1} = \frac{\rho}{\Delta t} \nabla \cdot \boldsymbol{u}^\dagger .$$

3. Since $\boldsymbol{u}^\dagger$ and p^{n+1} are now known, we find $\boldsymbol{u}^{n+1}$ by solving equation (4.62c).

In summary, instead of solving a coupled system for $\boldsymbol{u}^{n+1}$ and p^{n+1}, we compute $\boldsymbol{u}^\dagger$, p^{n+1} and $\boldsymbol{u}^{n+1}$ one after another.

Theorem 4.27. *The Chorin projection method (4.62) has a truncation error of first order in time and does not influence the discrete energy law for a constant viscosity η. For a variable viscosity, the discrete energy law for the splitting scheme (4.58a), (4.62) reads*

$$\begin{aligned} \frac{E_{total}(\varphi^{n+1}, \boldsymbol{u}^{n+1}) - E_{total}(\varphi^n, \boldsymbol{u}^n)}{\Delta t} &= -\mathrm{ND}_{pot}^{n+1} - \mathrm{ND}_{chorin}^{n+1} \\ &\quad - \int_\Omega M(\varphi^n) |\nabla \mu^{n+\frac{1}{2}}|^2 - \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^\dagger)|^2 , \end{aligned} \tag{4.63}$$

where the numerical dissipation caused by splitting and projection reads

$$\begin{aligned} \mathrm{ND}_{chorin}^{n+1} &= \frac{\Delta t}{2\rho} \|\nabla p^{n+1}\|_{L^2(\Omega)}^2 + \frac{\rho}{2\,\Delta t} \left(\|\boldsymbol{u}^\dagger - \boldsymbol{u}^*\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 \right) , \\ u^* &= \boldsymbol{u}^n - \frac{\Delta t}{\rho} \varphi^n \nabla \mu^{n+\frac{1}{2}} . \end{aligned}$$

Proof. Inserting relation (4.62b) in equation (4.62a) reads

$$\begin{aligned} \mathbf{0} &= \rho \left(\frac{\boldsymbol{u}^{n+1} - \boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n \cdot \nabla) \left(\boldsymbol{u}^{n+1} + \frac{\Delta t}{\rho} \nabla p^{n+1} \right) \right) + \nabla p^{n+1} \\ &\quad - \nabla \cdot \left\{ \eta(\varphi^n) \left[\nabla \left(\boldsymbol{u}^{n+1} + \frac{\Delta t}{\rho} \nabla p^{n+1} \right) + \left(\nabla \left(\boldsymbol{u}^{n+1} + \frac{\Delta t}{\rho} \nabla p^{n+1} \right) \right)^T \right] \right\} + \varphi^n \nabla \mu^{n+\frac{1}{2}} \\ &= \rho \left(\frac{\boldsymbol{u}^{n+1} - \boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n \cdot \nabla) \boldsymbol{u}^{n+1} \right) - \nabla \cdot \left\{ \eta(\varphi^n) \left[\nabla \boldsymbol{u}^{n+1} + (\nabla \boldsymbol{u}^{n+1})^T \right] \right\} + \nabla p^{n+1} + \varphi^n \nabla \mu^{n+\frac{1}{2}} \\ &\quad + \Delta t (\boldsymbol{u}^n \cdot \nabla) \nabla p^{n+1} - \frac{\Delta t}{\rho} \nabla \cdot \left\{ \eta(\varphi^n) \left[\nabla^2 p^{n+1} + (\nabla^2 p^{n+1})^T \right] \right\} , \end{aligned}$$

where

$$\begin{aligned} &\Delta t (\boldsymbol{u}^n \cdot \nabla) \nabla p^{n+1} - \frac{\Delta t}{\rho} \nabla \cdot \left\{ \eta(\varphi^n) \left[\nabla^2 p^{n+1} + (\nabla^2 p^{n+1})^T \right] \right\} \\ &= \Delta t \left((\boldsymbol{u}^n \cdot \nabla) \nabla p^{n+1} - \frac{2}{\rho} \nabla \cdot \left[\eta(\varphi^n) \nabla^2 p^{n+1} \right] \right) \in \mathcal{O}(\Delta t) \end{aligned}$$

is the splitting error.

Assuming a constant viscosity η, we can use the following equality for the last term of the splitting error

$$\eta \nabla \cdot \nabla^2 p^{n+1} = \eta \nabla \cdot (\Delta p^{n+1} \boldsymbol{I}) .$$

Now, taking the inner product of the splitting error and $\boldsymbol{u}^{n+1}$ and integrating over Ω, we can calculate the influence of the splitting error on the discrete energy law. Assuming the same suitable boundary conditions as above, see Remark 3.7, and using the incompressibility, we have

$$\begin{aligned}
&\Delta t \int_\Omega (\boldsymbol{u}^n \cdot \nabla) \nabla p^{n+1} \cdot \boldsymbol{u}^{n+1} - \frac{2\Delta t \eta}{\rho} \int_\Omega \nabla \cdot (\Delta p^{n+1} \boldsymbol{I}) \cdot \boldsymbol{u}^{n+1} \\
&= -\Delta t \int_\Omega (\nabla \cdot \boldsymbol{u}^n \, \boldsymbol{u}^{n+1} + \boldsymbol{u}^n \nabla \cdot \boldsymbol{u}^{n+1}) \cdot \nabla p^{n+1} + \Delta t \int_{\partial\Omega} (\boldsymbol{u}^n \cdot \boldsymbol{n}) \nabla p^{n+1} \cdot \boldsymbol{u}^{n+1} \\
&\quad + \frac{2\Delta t \eta}{\rho} \int_\Omega \Delta p^{n+1} \boldsymbol{I} : \nabla \boldsymbol{u}^{n+1} - \frac{2\Delta t \eta}{\rho} \int_{\partial\Omega} \Delta p^{n+1} \, \boldsymbol{u}^{n+1} \cdot \boldsymbol{n} \\
&= \frac{2\Delta t \eta}{\rho} \int_\Omega \Delta p^{n+1} \nabla \cdot \boldsymbol{u}^{n+1} = 0 .
\end{aligned}$$

For a variable viscosity $\eta(\varphi^n)$, we substitute (4.57), i.e.

$$\boldsymbol{u}^* = \boldsymbol{u}^n - \frac{\Delta t}{\rho} \varphi^n \nabla \mu^{n+\frac{1}{2}} ,$$

in equation (4.62a), before we take the inner product with $\boldsymbol{u}^\dagger$ and integrate over Ω, yielding analogously to the derivation of energy law (4.52)

$$0 = \frac{\rho}{2\,\Delta t} \left(\|\boldsymbol{u}^\dagger\|^2_{L^2(\Omega)} - \|\boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^\dagger - \boldsymbol{u}^*\|^2_{L^2(\Omega)} \right) + \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^\dagger)|^2 . \tag{4.64}$$

Since

$$\begin{aligned}
\boldsymbol{u}^\dagger &= \boldsymbol{u}^{n+1} + \frac{\Delta t}{\rho} \nabla p^{n+1} , \\
\|\boldsymbol{u}^\dagger\|^2_{L^2(\Omega)} &= \|\boldsymbol{u}^{n+1} + \frac{\Delta t}{\rho} \nabla p^{n+1}\|^2_{L^2(\Omega)} \\
&= \|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} + \frac{(\Delta t)^2}{\rho^2} \|\nabla p^{n+1}\|^2_{L^2(\Omega)} + \frac{2\Delta t}{\rho} \int_\Omega \boldsymbol{u}^{n+1} \cdot \nabla p^{n+1} \\
&= \|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} + \frac{(\Delta t)^2}{\rho^2} \|\nabla p^{n+1}\|^2_{L^2(\Omega)} - \frac{2\Delta t}{\rho} \int_\Omega \nabla \cdot \boldsymbol{u}^{n+1} \, p^{n+1} + \frac{2\Delta t}{\rho} \int_{\partial\Omega} \boldsymbol{u}^{n+1} \cdot \boldsymbol{n} \, p^{n+1} \\
&= \|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} + \frac{(\Delta t)^2}{\rho^2} \|\nabla p^{n+1}\|^2_{L^2(\Omega)} ,
\end{aligned}$$

where the last equality holds true due to the incompressibility and suitable boundary conditions for $\boldsymbol{u}^{n+1}$, see Remark 3.7. Inserting the last relation into (4.64) reads

$$0 = \frac{\rho}{2\,\Delta t} \left(\|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} - \|\boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^\dagger - \boldsymbol{u}^*\|^2_{L^2(\Omega)} \right) + \frac{\Delta t}{2\rho} \|\nabla p^{n+1}\|^2_{L^2(\Omega)} + \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^\dagger)|^2 .$$

Now, analogously to the proof of Theorem 4.24, we insert relation (4.61) for $\|\boldsymbol{u}^*\|^2_{L^2(\Omega)}$, yielding the discrete kinetic energy law

$$0 = \frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \frac{\rho}{2\,\Delta t}\Big(\|\boldsymbol{u}^\dagger - \boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|^2_{L^2(\Omega)}\Big) - \int_\Omega \nabla\cdot(\boldsymbol{u}^*\varphi^n)\,\mu^{n+\frac{1}{2}} + \frac{\Delta t}{2\rho}\|\nabla p^{n+1}\|^2_{L^2(\Omega)} + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^\dagger)|^2\,.$$

Finally, adding the discrete mixing energy law (4.59), the integrals of the coupling terms cancel out and we have the desired energy law (4.63). □

Table 4.3: Summary of the schemes for the Cahn-Hilliard-Navier-Stokes model

	linear	en.-stable	mass-cons.	trunc. error	Δt
part. implicit (4.42)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
part. Crank-Nicolson (4.45)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
second order (4.48)	✓	✓	✓	$\mathcal{O}\left((\Delta t)^2\right)$	$\mathcal{O}(h^2)$
splitting scheme (4.51)	✓	✓	-	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
cons. splitting scheme (4.58)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
Chorin splitting (4.58a), (4.62)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$

Remark 4.28. *Summarized, see Table 4.3, the second order Crank-Nicolson type scheme (4.48) has the best properties on the face of it and we should note that again the two Crank-Nicolson type schemes are not only energy-stable but conserve the continuous energy laws up to the dissipation term* ND_{pot}*, where the latter results compulsorily from the linear approximation of the potential derivative* $f(\varphi)$*. Even though the mass-conservative (cons.) splitting scheme (4.58) has a larger truncation error and additional numerical dissipation from the splitting, the decoupled calculation also resulting from the splitting yields a major efficiency advantage and more flexibility regarding the spatial discretization of the advection due to its explicit discretization in time. Finally making the mass-conservative splitting scheme our method of choice for solving the Cahn-Hilliard-Navier-Stokes model numerically.*

4.5 Schemes for the simplified model for viscoelastic phase separation

We start this section proposing a one step numerical scheme for the simplified model (3.48), which is based on the Crank-Nicolson type scheme (4.33) for the Cahn-Hilliard equation. Note that we have originally introduced the following two schemes in *Strasser, Tierra, Dünweg and Lukáčová-Medviďová* [67].

4.5.1 Crank-Nicolson type scheme

We consider a uniform partition of the time interval $[0,T]$ with a constant time step Δt. Given (φ^n, q^n) from the previous time step, we compute (φ^{n+1}, q^{n+1}) such that

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \nabla\cdot\left\{\frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,, \tag{4.65a}$$

$$\begin{aligned}\mu^{n+\frac{1}{2}} &= -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,\\ \frac{q^{n+1}-q^n}{\Delta t} &+ \frac{1}{\tau_B(\varphi^n)}q^{n+\frac{1}{2}}\\ &+ G_B(\varphi^n)\nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,,\end{aligned} \tag{4.65b}$$

where

$$\varphi^{n+\frac{1}{2}} = \frac{\varphi^{n+1}+\varphi^n}{2} \qquad \text{and} \qquad q^{n+\frac{1}{2}} = \frac{q^{n+1}+q^n}{2}$$

are midpoint approximations, yielding a Crank-Nicolson type scheme.

Theorem 4.29. *Let $f(\varphi^{n+1},\varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.65) is linear and, assuming suitable boundary conditions, see Remark 3.11, it satisfies the following discrete version of energy law (3.50)*

$$\begin{aligned}\frac{E_{total}(\varphi^{n+1},q^{n+1}) - E_{total}(\varphi^n,q^n)}{\Delta t} &= -\,\mathrm{ND}_{pot}^{n+1} - \int_\Omega \frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2\\ &\quad - \int_\Omega \frac{1}{\zeta(\varphi^n)}\left|\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right|^2,\end{aligned} \tag{4.66}$$

where

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega f(\varphi^{n+1},\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t}\,.$$

Proof. It is clear that the proposed scheme is linear if $f(\varphi^{n+1},\varphi^n)$ is linear. The discrete mixing energy is calculated analogously to the discrete energy law (4.34), i.e., multiplying (4.65a) by $\mu^{n+\frac{1}{2}}$, integrating over Ω and applying integration by parts yields

$$\begin{aligned}0 = \frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} &+ \mathrm{ND}_{pot}^{n+1} - \int_{\partial\Omega}\frac{\varphi^{n+1}-\varphi^n}{\Delta t}\lambda\frac{\partial\varphi^{n+\frac{1}{2}}}{\partial\boldsymbol{n}}\\ &+ \int_\Omega \frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\cdot\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right]\\ &- \int_{\partial\Omega}\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\frac{\partial\mu^{n+\frac{1}{2}}}{\partial\boldsymbol{n}} - \frac{\partial(G_B(\varphi^n)\,q^{n+\frac{1}{2}})}{\partial\boldsymbol{n}}\right]\varphi^n(1-\varphi^n)\mu^{n+\frac{1}{2}}\,.\end{aligned}$$

Assuming suitable boundary conditions, see Remark 3.11, the two boundary integrals vanish, yielding the discrete energy law

$$\begin{aligned}\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} &= -\,\mathrm{ND}_{pot}^{n+1}\\ &\quad - \int_\Omega \frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\cdot\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}}\right].\end{aligned}$$

Multiplying (4.65b) by $q^{n+\frac{1}{2}}$, integrating over Ω and applying integration by parts reads

$$\begin{aligned}
0 &= \int_\Omega \frac{q^{n+1}-q^n}{\Delta t}\frac{q^{n+1}+q^n}{2} + \int_\Omega \frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2 \\
&\quad + \int_\Omega G_B(\varphi^n)\nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} q^{n+\frac{1}{2}} \\
&= \int_\Omega \frac{1}{2\Delta t}((q^{n+1})^2-(q^n)^2) + \int_\Omega \frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2 \\
&\quad - \int_\Omega \frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\cdot\nabla(G_B(\varphi^n)q^{n+\frac{1}{2}}) \\
&\quad + \int_{\partial\Omega} \frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\frac{\partial\mu^{n+\frac{1}{2}}}{\partial \boldsymbol{n}} - \frac{\partial(G_B(\varphi^n)\,q^{n+\frac{1}{2}})}{\partial \boldsymbol{n}}\right] G_B(\varphi^n)q^{n+\frac{1}{2}}\,.
\end{aligned}$$

Again, assuming suitable boundary conditions, see Remark 3.11, the boundary integral vanishes. Since the discrete bulk energy at time t_n reads

$$E_{bulk}(q^n) = \int_\Omega \frac{1}{2}(q^n)^2\,,$$

we have the following discrete energy law

$$\begin{aligned}
\frac{E_{bulk}(q^{n+1})-E_{bulk}(q^n)}{\Delta t} &= -\int_\Omega \frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2 \\
&\quad - \int_\Omega \frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\cdot\left[-\nabla(G_B(\varphi^n)q^{n+\frac{1}{2}})\right].
\end{aligned}$$

Summing up the two discrete energy laws yields the desired energy law (4.66). □

Theorem 4.30. *Assuming suitable boundary conditions, see Remark 3.11, scheme (4.65) satisfies*

$$\int_\Omega \varphi^{n+1} = \int_\Omega \varphi^0 \qquad \forall n\in\mathbb{N}\,.$$

Thus, it is mass-conservative.

Proof. Multiplying equation (4.65a) by the smooth test function ψ, integrating over the computational domain Ω and applying integration by parts yields

$$\begin{aligned}
\int_\Omega \frac{\varphi^{n+1}-\varphi^n}{\Delta t}\psi &= \int_\Omega \nabla\cdot\left\{\frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\}\psi \\
&= -\int_\Omega \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\cdot\nabla\psi \\
&\quad + \int_{\partial\Omega} \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\psi\cdot\boldsymbol{n} \\
&= -\int_\Omega \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\cdot\nabla\psi \\
&\quad + \int_{\partial\Omega} \frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\frac{\partial\mu^{n+\frac{1}{2}}}{\partial\boldsymbol{n}} - \frac{\partial(G_B(\varphi^n)\,q^{n+\frac{1}{2}})}{\partial\boldsymbol{n}}\right]\psi\,.
\end{aligned}$$

Then, setting $\psi = 1$ and assuming suitable boundary conditions, see Remark 3.11, we have

$$\int_\Omega \frac{\varphi^{n+1} - \varphi^n}{\Delta t} = 0 \qquad \Longleftrightarrow \qquad \int_\Omega \varphi^{n+1} = \int_\Omega \varphi^n \,.$$

Thus, it follows by induction that

$$\int_\Omega \varphi^{n+1} = \int_\Omega \varphi^0 \qquad \forall n \in \mathbb{N} \,.$$

□

The numerical scheme (4.65) is of first order in time, since some discretizations are evaluated explicitly at time t^n to avoid nonlinearities.

4.5.2 Second order scheme

Analogously to scheme (4.40), we propose a linear second order numerical scheme by using the second order Adams-Bashforth method to extrapolate the explicit terms φ^n by

$$\varphi^{n-\frac{1}{2}} = \frac{3\varphi^n - \varphi^{n-1}}{2} \,.$$

The resulting two-step numerical scheme reads

$$\frac{\varphi^{n+1} - \varphi^n}{\Delta t} - \nabla \cdot \left\{ \frac{\varphi^{n-\frac{1}{2}}(1-\varphi^{n-\frac{1}{2}})}{\zeta(\varphi^{n-\frac{1}{2}})} \left[\varphi^{n-\frac{1}{2}}(1-\varphi^{n-\frac{1}{2}}) \nabla \mu^{n+\frac{1}{2}} - \nabla (G_B(\varphi^{n-\frac{1}{2}})\, q^{n+\frac{1}{2}}) \right] \right\} = 0 \,, \tag{4.67a}$$

$$\begin{aligned} \mu^{n+\frac{1}{2}} &= -\lambda \Delta \varphi^{n+\frac{1}{2}} + f(\varphi^{n+1}, \varphi^n) \,, \\ \frac{q^{n+1} - q^n}{\Delta t} &+ \frac{1}{\tau_B(\varphi^{n-\frac{1}{2}})} q^{n+\frac{1}{2}} \\ &+ G_B(\varphi^{n-\frac{1}{2}}) \nabla \cdot \left\{ \frac{1}{\zeta(\varphi^{n-\frac{1}{2}})} \left[\varphi^{n-\frac{1}{2}}(1-\varphi^{n-\frac{1}{2}}) \nabla \mu^{n+\frac{1}{2}} - \nabla (G_B(\varphi^{n-\frac{1}{2}})\, q^{n+\frac{1}{2}}) \right] \right\} = 0 \,. \end{aligned} \tag{4.67b}$$

Theorem 4.31. *Let $f(\varphi^{n+1}, \varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$. Then, the resulting numerical scheme (4.67) is linear and, assuming suitable boundary conditions, see Remark 3.11, it is mass-conservative and satisfies the following discrete version of energy law (3.50)*

$$\begin{aligned} \frac{E_{total}(\varphi^{n+1}, q^{n+1}) - E_{total}(\varphi^n, q^n)}{\Delta t} &= -\,\mathrm{ND}_{pot}^{n+1} - \int_\Omega \frac{1}{\tau_B(\varphi^{n-\frac{1}{2}})} (q^{n+\frac{1}{2}})^2 \\ &\quad - \int_\Omega \frac{1}{\zeta(\varphi^{n-\frac{1}{2}})} \left| \varphi^{n-\frac{1}{2}}(1-\varphi^{n-\frac{1}{2}}) \nabla \mu^{n+\frac{1}{2}} - \nabla (G_B(\varphi^{n-\frac{1}{2}})\, q^{n+\frac{1}{2}}) \right|^2 . \end{aligned} \tag{4.68}$$

Proof. It is clear that the proposed scheme is linear if $f(\varphi^{n+1}, \varphi^n)$ is linear. The mass conservation is calculated analogously to the proof of Theorem 4.30 and the discrete energy law analogously to energy law (4.66). □

In order to compute the pair (φ^1, q^1) from (φ^0, q^0) a second order one-step nonlinear scheme can be considered. We overcome this by setting $\varphi^{-1} := \varphi^0$ and $q^{-1} := q^0$, i.e., we solve the first order scheme (4.65) in the first time step. As long as the initial data is sufficiently smooth, the influence is usually negligible for $T \gg 0$, see the experimental convergence presented in Chapter 5.

Note that due to the highly coupled nature of the two equations, we restrain from using a splitting scheme for the simplified model. But since both φ and q are scalar functions, unlike the velocity field $\boldsymbol{u}$, the computational effort of the coupled calculation is less severe.

Table 4.4: Summary of the schemes for the simplified model

	linear	energy-stable	mass-cons.	trunc. error	Δt
part. Crank-Nicolson (4.65)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
second order (4.67)	✓	✓	✓	$\mathcal{O}\left((\Delta t)^2\right)$	$\mathcal{O}(h^2)$

Remark 4.32. *Summarized, see Table 4.4, the second order Crank-Nicolson type scheme (4.67) has better properties and is therefore our method of choice for solving the simplified model numerically. Note that again both Crank-Nicolson type schemes are not only energy-stable but conserve the continuous energy laws up to the dissipation term* ND_{pot}*, which results from a linear potential approximation.*

4.6 Schemes for the full model for viscoelastic phase separation

Based on the methods introduced above, we present schemes for the full two-fluid model for viscoelastic phase separation (3.47) in the following. Note that most of these schemes have already been published in *Strasser, Tierra, Dünweg and Lukáčová-Medviďová* [67].

4.6.1 Mixed scheme

At first, we propose to combine the coupled scheme (4.42) for the Cahn-Hilliard-Navier-Stokes part with scheme (4.65) for the simplified model. In order to discretize the stress tensor evolution equation (3.47c), we primarily use the explicit Euler method here, since the equation contains only first order derivatives. Given $(\varphi^n, q^n, \boldsymbol{\sigma}^n, \boldsymbol{u}^n)$ from the previous time step, we

compute $(\varphi^{n+1}, q^{n+1}, \boldsymbol{\sigma}^{n+1}, \boldsymbol{u}^{n+1}, p^{n+1})$ such that

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^{n+1}\cdot\nabla\varphi^n - \nabla\cdot\left\{\frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,, \tag{4.69a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,$$

$$\frac{q^{n+1}-q^n}{\Delta t} + \boldsymbol{u}^n\cdot\nabla q^{n+\frac{1}{2}} + \frac{1}{\tau_B(\varphi^n)}q^{n+\frac{1}{2}} + G_B(\varphi^n)\nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,, \tag{4.69b}$$

$$\frac{\boldsymbol{\sigma}^{n+1}-\boldsymbol{\sigma}^n}{\Delta t} + (\boldsymbol{u}^{n+1}\cdot\nabla)\boldsymbol{\sigma}^n - (\nabla\boldsymbol{u}^{n+1})\boldsymbol{\sigma}^n - \boldsymbol{\sigma}^n\left(\nabla\boldsymbol{u}^{n+1}\right)^T + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}\boldsymbol{\sigma}^{n+\theta} - G_S(\varphi^{n+\frac{1}{2}})\left[\nabla\boldsymbol{u}^{n+1} + (\nabla\boldsymbol{u}^{n+1})^T\right] = \boldsymbol{0}\,, \tag{4.69c}$$

$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+1} - \nabla\cdot\left\{\eta(\varphi^n)\left[\nabla\boldsymbol{u}^{n+1} + (\nabla\boldsymbol{u}^{n+1})^T\right]\right\} + \nabla p^{n+1} - \mu^{n+\frac{1}{2}}\nabla\varphi^n - \nabla\cdot\boldsymbol{\sigma}^n = \boldsymbol{0}\,, \tag{4.69d}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,, \tag{4.69e}$$

where $\theta\in\{0,1\}$.

Theorem 4.33. *Let $f(\varphi^{n+1},\varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.69) is linear and, assuming suitable boundary conditions, see Remark 3.13, it is mass-conservative and satisfies the following discrete version of energy law (3.52)*

$$\begin{aligned}\frac{E_{total}(\varphi^{n+1},q^{n+1},\boldsymbol{\sigma}^{n+1},\boldsymbol{u}^{n+1}) - E_{total}(\varphi^n,q^n,\boldsymbol{\sigma}^n,\boldsymbol{u}^n)}{\Delta t} &= -\mathrm{ND}_{pot}^{n+1}\\ -\int_\Omega\frac{1}{\zeta(\varphi^n)}\left|\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right|^2 - \int_\Omega\frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2&\\ -\int_\Omega\frac{1}{2\,\tau_S(\varphi^{n+\frac{1}{2}})}\mathrm{tr}(\boldsymbol{\sigma}^{n+\theta}) - \frac{1}{2\,\Delta t}\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|^2_{L^2(\Omega)} - \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,,&\end{aligned} \tag{4.70}$$

where like above

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega f(\varphi^{n+1},\varphi^n)\frac{\varphi^{n+1}-\varphi^n}{\Delta t} - \int_\Omega\frac{F(\varphi^{n+1})-F(\varphi^n)}{\Delta t}\,.$$

Proof. Apart from the stress evolution equation (4.69c), it is clear that the proposed scheme is linear if $f(\varphi^{n+1},\varphi^n)$ is linear. The elastic stress tensor $\boldsymbol{\sigma}$ only occurs in its evolution equation (4.69c) and the Navier-Stokes equation (4.69d). Since it is discretized explicitly in the latter, the computation of (4.69c) is decoupled from the computation of the other (coupled) equations and performed afterwards, when the solution $(\varphi^{n+1}, q^{n+1}, \boldsymbol{u}^{n+1}, p^{n+1})$ is already known. Thus, equation (4.69c) is linear as well.

The mass conservation is calculated analogously to the proof of Theorem 4.19.

Then, analogously to the derivation of the discrete energy law (4.66) of the simplified model, (4.69a) is multiplied by $\mu^{n+\frac{1}{2}}$ and (4.69b) by $q^{n+\frac{1}{2}}$. Integrating both equations over Ω, using

partial integration, assuming suitable boundary conditions, see Remark 3.13, and summing up yields the same energy relation as above plus two terms from the previously missing advection

$$\begin{aligned}
&\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} + \frac{E_{bulk}(q^{n+1}) - E_{bulk}(q^n)}{\Delta t} + \mathrm{ND}_{pot}^{n+1} \\
&+ \int_\Omega \frac{1}{\tau_B(\varphi^n)} (q^{n+\frac{1}{2}})^2 + \int_\Omega \frac{1}{\zeta(\varphi^n)} \left| \varphi^n(1-\varphi^n) \nabla \mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\, q^{n+\frac{1}{2}}) \right|^2 \\
&+ \int_\Omega \boldsymbol{u}^{n+1} \cdot \nabla \varphi^n \mu^{n+\frac{1}{2}} + \int_\Omega \boldsymbol{u}^{n+1} \cdot \nabla q^{n+\frac{1}{2}}\, q^{n+\frac{1}{2}} = 0\,.
\end{aligned} \tag{4.71}$$

Here, the last term vanishes, since

$$\int_\Omega \boldsymbol{u}^{n+1} \cdot \nabla q^{n+\frac{1}{2}}\, q^{n+\frac{1}{2}} = \int_\Omega \boldsymbol{u}^{n+1} \cdot \frac{1}{2} \nabla (q^{n+\frac{1}{2}})^2 = -\int_\Omega \nabla \cdot \boldsymbol{u}^{n+1} \frac{1}{2} (q^{n+\frac{1}{2}})^2 + \int_{\partial\Omega} \boldsymbol{u}^{n+1} \frac{1}{2} (q^{n+\frac{1}{2}})^2 \cdot \boldsymbol{n} = 0\,,$$

using the incompressibility and assuming suitable boundary conditions, i.e., periodic or $\boldsymbol{u}^{n+1} \cdot \boldsymbol{n} = 0$ on $\partial\Omega$, see Remark 3.13.

Further, we take the Frobenius inner product of (4.69c) and $\frac{1}{2}\boldsymbol{I}$ and integrate over Ω, enabling analogous calculations as for the shear stress part of the continuous energy law (3.52), yielding

$$\frac{E_{el}(\boldsymbol{\sigma}^{n+1}) - E_{el}(\boldsymbol{\sigma}^n)}{\Delta t} - \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^{n+1} + \int_\Omega \frac{1}{2\,\tau_S(\varphi^{n+\frac{1}{2}})} \mathrm{tr}(\boldsymbol{\sigma}^{n+\theta}) = 0\,. \tag{4.72}$$

Taking the inner product of (4.69d) and $\boldsymbol{u}^{n+1}$ and integrating over Ω yields, analogously to the derivation of energy law (4.43),

$$\begin{aligned}
\frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \frac{1}{2\,\Delta t} \|\boldsymbol{u}^{n+1} - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 + \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2 \\
- \int_\Omega \boldsymbol{u}^{n+1} \cdot \nabla \varphi^n \mu^{n+\frac{1}{2}} - \int_\Omega (\nabla \cdot \boldsymbol{\sigma}^n) \cdot \boldsymbol{u}^{n+1} = 0\,,
\end{aligned}$$

where for the last integral it holds

$$-\int_\Omega (\nabla \cdot \boldsymbol{\sigma}^n) \cdot \boldsymbol{u}^{n+1} = \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^{n+1} - \int_{\partial\Omega} \boldsymbol{\sigma}^n \boldsymbol{n} \cdot \boldsymbol{u}^{n+1} = \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^{n+1}\,.$$

Thus, summing up the above three energy relations, all integrals which are related to coupling terms vanish and we get the desired energy law (4.70). □

4.6.2 Crank-Nicolson type scheme

It is possible to eliminate the term $\frac{1}{2\Delta t}\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^{n}\|^2_{L^2(\Omega)}$ from energy law (4.70), considering the following linear one-step scheme

$$\frac{\varphi^{n+1}-\varphi^{n}}{\Delta t}+\boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla\varphi^{n}-\nabla\cdot\left\{\frac{\varphi^{n}(1-\varphi^{n})}{\zeta(\varphi^{n})}\left[\varphi^{n}(1-\varphi^{n})\nabla\mu^{n+\frac{1}{2}}-\nabla(G_B(\varphi^{n})\,q^{n+\frac{1}{2}})\right]\right\}=0\,, \tag{4.73a}$$

$$\mu^{n+\frac{1}{2}}=-\lambda\Delta\varphi^{n+\frac{1}{2}}+f(\varphi^{n+1},\varphi^{n})\,,$$

$$\frac{q^{n+1}-q^{n}}{\Delta t}+\boldsymbol{u}^{n}\cdot\nabla q^{n+\frac{1}{2}}+\frac{1}{\tau_B(\varphi^{n})}q^{n+\frac{1}{2}} \tag{4.73b}$$

$$+G_B(\varphi^{n})\nabla\cdot\left\{\frac{1}{\zeta(\varphi^{n})}\left[\varphi^{n}(1-\varphi^{n})\nabla\mu^{n+\frac{1}{2}}-\nabla(G_B(\varphi^{n})\,q^{n+\frac{1}{2}})\right]\right\}=0\,,$$

$$\frac{\boldsymbol{\sigma}^{n+1}-\boldsymbol{\sigma}^{n}}{\Delta t}+(\boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla)\boldsymbol{\sigma}^{n}-(\nabla\boldsymbol{u}^{n+\frac{1}{2}})\boldsymbol{\sigma}^{n}-\boldsymbol{\sigma}^{n}\left(\nabla\boldsymbol{u}^{n+\frac{1}{2}}\right)^{T} \tag{4.73c}$$

$$+\frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}\boldsymbol{\sigma}^{n+\theta}-2G_S(\varphi^{n+\frac{1}{2}})\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})=\boldsymbol{0}\,,$$

$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^{n}}{\Delta t}+(\boldsymbol{u}^{n}\cdot\nabla)\boldsymbol{u}^{n+\frac{1}{2}}-\nabla\cdot\left\{2\eta(\varphi^{n})\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})\right\}+\nabla p^{n+1}-\mu^{n+\frac{1}{2}}\nabla\varphi^{n}-\nabla\cdot\boldsymbol{\sigma}^{n}=\boldsymbol{0}\,, \tag{4.73d}$$

$$\nabla\cdot\boldsymbol{u}^{n+1}=0\,, \tag{4.73e}$$

where $\theta\in\{0,1\}$.

Theorem 4.34. *Let $f(\varphi^{n+1},\varphi^{n})$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.73) is linear and, assuming suitable boundary conditions, see Remark 3.13, it is mass-conservative and satisfies the following discrete version of energy law (3.52)*

$$\begin{aligned}\frac{E_{total}(\varphi^{n+1},q^{n+1},\boldsymbol{\sigma}^{n+1},\boldsymbol{u}^{n+1})-E_{total}(\varphi^{n},q^{n},\boldsymbol{\sigma}^{n},\boldsymbol{u}^{n})}{\Delta t}&=-\mathrm{ND}^{n+1}_{pot}\\ -\int_\Omega\frac{1}{\zeta(\varphi^{n})}\left|\varphi^{n}(1-\varphi^{n})\nabla\mu^{n+\frac{1}{2}}-\nabla(G_B(\varphi^{n})\,q^{n+\frac{1}{2}})\right|^2-\int_\Omega\frac{1}{\tau_B(\varphi^{n})}(q^{n+\frac{1}{2}})^2&\\ -\int_\Omega\frac{1}{2\,\tau_S(\varphi^{n+\frac{1}{2}})}\mathrm{tr}(\boldsymbol{\sigma}^{n+\theta})-\int_\Omega 2\eta(\varphi^{n})|\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2\,,&\end{aligned} \tag{4.74}$$

which is, up to ND^{n+1}_{pot}, analogous to the continuous energy law.

Proof. Can be shown analogously to the proof of Theorem 4.33, except for the calculation of the kinetic energy law. For the latter, we calculate the inner product of (4.73d) and $\boldsymbol{u}^{n+\frac{1}{2}}$ analogously to the derivation of energy law (4.46). □

4.6.3 Second order scheme

Analogously to schemes (4.48) and (4.67), using the second order extrapolation

$$\omega^{n-\frac{1}{2}}=\frac{3\omega^{n}-\omega^{n-1}}{2},\quad \omega\in\{\varphi,\boldsymbol{u},\boldsymbol{\sigma}\},$$

for the explicit terms in scheme (4.73) yields the following linear and second order in time two-step numerical scheme

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla\varphi^{n-\frac{1}{2}} \tag{4.75a}$$
$$- \nabla\cdot\left\{\frac{\varphi^{n-\frac{1}{2}}(1-\varphi^{n-\frac{1}{2}})}{\zeta(\varphi^{n-\frac{1}{2}})}\left[\varphi^{n-\frac{1}{2}}(1-\varphi^{n-\frac{1}{2}})\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^{n-\frac{1}{2}})\,q^{n+\frac{1}{2}})\right]\right\} = 0\,,$$
$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,$$
$$\frac{q^{n+1}-q^n}{\Delta t} + \boldsymbol{u}^{n-\frac{1}{2}}\cdot\nabla q^{n+\frac{1}{2}} + \frac{1}{\tau_B(\varphi^{n-\frac{1}{2}})}q^{n+\frac{1}{2}} \tag{4.75b}$$
$$+ G_B(\varphi^{n-\frac{1}{2}})\nabla\cdot\left\{\frac{1}{\zeta(\varphi^{n-\frac{1}{2}})}\left[\varphi^{n-\frac{1}{2}}(1-\varphi^{n-\frac{1}{2}})\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^{n-\frac{1}{2}})\,q^{n+\frac{1}{2}})\right]\right\} = 0\,,$$
$$\frac{\boldsymbol{\sigma}^{n+1}-\boldsymbol{\sigma}^n}{\Delta t} + (\boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla)\boldsymbol{\sigma}^{n-\frac{1}{2}} - (\nabla\boldsymbol{u}^{n+\frac{1}{2}})\boldsymbol{\sigma}^{n-\frac{1}{2}} - \boldsymbol{\sigma}^{n-\frac{1}{2}}\left(\nabla\boldsymbol{u}^{n+\frac{1}{2}}\right)^T \tag{4.75c}$$
$$+ \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}\boldsymbol{\sigma}^{n-\frac{1}{2}+\theta} - G_S(\varphi^{n+\frac{1}{2}})2\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}}) = \boldsymbol{0}\,,$$
$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^{n-\frac{1}{2}}\cdot\nabla)\boldsymbol{u}^{n+\frac{1}{2}} - \nabla\cdot\left\{\eta(\varphi^{n-\frac{1}{2}})2\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})\right\} \tag{4.75d}$$
$$+ \nabla p^{n+\frac{1}{2}} - \mu^{n+\frac{1}{2}}\nabla\varphi^{n-\frac{1}{2}} - \nabla\cdot\boldsymbol{\sigma}^{n-\frac{1}{2}} = \boldsymbol{0}\,,$$
$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,, \tag{4.75e}$$

where $\theta\in\{0,1\}$.

Theorem 4.35. *Let $f(\varphi^{n+1},\varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.75) is linear and, assuming suitable boundary conditions, see Remark 3.13, it is mass-conservative and satisfies the following discrete version of energy law (3.52)*

$$\frac{E_{total}(\varphi^{n+1},q^{n+1},\boldsymbol{\sigma}^{n+1},\boldsymbol{u}^{n+1}) - E_{total}(\varphi^n,q^n,\boldsymbol{\sigma}^n,\boldsymbol{u}^n)}{\Delta t} = -\mathrm{ND}_{pot}^{n+1}$$
$$- \int_\Omega \frac{1}{\zeta(\varphi^{n-\frac{1}{2}})}\left|\varphi^{n-\frac{1}{2}}(1-\varphi^{n-\frac{1}{2}})\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^{n-\frac{1}{2}})\,q^{n+\frac{1}{2}})\right|^2 - \int_\Omega \frac{1}{\tau_B(\varphi^{n-\frac{1}{2}})}(q^{n+\frac{1}{2}})^2 \tag{4.76}$$
$$- \int_\Omega \frac{1}{2\,\tau_S(\varphi^{n+\frac{1}{2}})}\mathrm{tr}(\boldsymbol{\sigma}^{n-\frac{1}{2}+\theta}) - \int_\Omega 2\eta(\varphi^{n-\frac{1}{2}})|\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2\,,$$

which is, up to ND_{pot}^{n+1}, analogous to the continuous energy law.

Proof. Can be shown analogously to the proof of Theorem 4.34. □

4.6.4 Splitting scheme

In this subsection, we present yet another possibility to discretize system (3.56). In order to save computational cost, we split the computation into three different sub-steps. The first two steps are the interesting ones here, allowing us to decouple the calculation of the fluid part $(\boldsymbol{u},p)$ from the phase-field and bulk stress parts (φ,q) analogously to the mass-conservative

splitting scheme (4.58) for the Cahn-Hilliard-Navier-Stokes model, by modifying the advection term of the φ-equation.

Step 1. Find (φ^{n+1}, q^{n+1}) such that

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \nabla\cdot(\boldsymbol{u}^*\varphi^n) - \nabla\cdot\left\{\frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,, \tag{4.77a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,$$

$$\boldsymbol{u}^* = \boldsymbol{u}^n - \Delta t\,\varphi^n\nabla\mu^{n+\frac{1}{2}}\,, \tag{4.77b}$$

$$\frac{q^{n+1}-q^n}{\Delta t} + \nabla\cdot(\boldsymbol{u}^n q^{n+\frac{1}{2}}) + \frac{1}{\tau_B(\varphi^n)}q^{n+\frac{1}{2}} \tag{4.77c}$$
$$+ G_B(\varphi^n)\nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,.$$

Step 2. Find $(\boldsymbol{u}^{n+1}, p^{n+1})$ such that

$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^*}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+1} - \nabla\cdot\left\{\eta(\varphi^{n+\frac{1}{2}})\,2\boldsymbol{D}(\boldsymbol{u}^{n+1})\right\} + \nabla p^{n+1} - \nabla\cdot\boldsymbol{\sigma}^n = \boldsymbol{0}\,, \tag{4.78a}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,. \tag{4.78b}$$

Step 3. Find $\boldsymbol{\sigma}^{n+1}$ such that

$$\begin{aligned}\frac{\boldsymbol{\sigma}^{n+1}-\boldsymbol{\sigma}^n}{\Delta t} &+ (\boldsymbol{u}^{n+1}\cdot\nabla)\boldsymbol{\sigma}^n - (\nabla\boldsymbol{u}^{n+1})\boldsymbol{\sigma}^n - \boldsymbol{\sigma}^n\left(\nabla\boldsymbol{u}^{n+1}\right)^T\\ &+ \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}\boldsymbol{\sigma}^{n+\theta} - G_S(\varphi^{n+\frac{1}{2}})\,2\boldsymbol{D}(\boldsymbol{u}^{n+1}) = \boldsymbol{0}\,,\end{aligned} \tag{4.79}$$

where $\theta\in\{0,1\}$.

Theorem 4.36. *Let $f(\varphi^{n+1},\varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.77)-(4.79) is linear and, assuming suitable boundary conditions, see Remark 3.13, it is mass-conservative and satisfies the discrete energy law*

$$\begin{aligned}\frac{E_{total}(\varphi^{n+1},q^{n+1},\boldsymbol{\sigma}^{n+1},\boldsymbol{u}^{n+1}) - E_{total}(\varphi^n,q^n,\boldsymbol{\sigma}^n,\boldsymbol{u}^n)}{\Delta t} &= -\mathrm{ND}_{pot}^{n+1} - \mathrm{ND}_{split}^{n+1}\\ -\int_\Omega\frac{1}{\zeta(\varphi^n)}\left|\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right|^2 &- \int_\Omega\frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2\\ -\int_\Omega\frac{1}{2\,\tau_S(\varphi^{n+\frac{1}{2}})}\mathrm{tr}(\boldsymbol{\sigma}^{n+\theta}) &- \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,,\end{aligned} \tag{4.80}$$

where

$$\mathrm{ND}_{split}^{n+1} = \frac{1}{2\,\Delta t}\left(\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^*-\boldsymbol{u}^n\|^2_{L^2(\Omega)}\right).$$

Proof. Analogously to the proof of the discrete energy law (4.70), the proposed scheme is linear, and analogously to the proof of Theorem 4.25, it is mass-conservative. To calculate the discrete energy law, we multiply (4.77a) by $\mu^{n+\frac{1}{2}}$ and (4.77c) by $q^{n+\frac{1}{2}}$ and integrate over

Ω. Assuming suitable boundary conditions, see Remark 3.13, and summing up, we obtain an energy law analogous to (4.71), with only the last term being different, reading

$$\begin{aligned}&\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} + \frac{E_{bulk}(q^{n+1}) - E_{bulk}(q^n)}{\Delta t} + \mathrm{ND}_{pot}^{n+1} + \int_\Omega \frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2 \\ &+ \int_\Omega \frac{1}{\zeta(\varphi^n)} \left| \varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\, q^{n+\frac{1}{2}}) \right|^2 + \int_\Omega \nabla\cdot(\boldsymbol{u}^*\varphi^n)\,\mu^{n+\frac{1}{2}} = 0\,.\end{aligned} \tag{4.81}$$

Taking the Frobenius inner product of (4.79) and $\frac{1}{2}\boldsymbol{I}$ and integrating over Ω yields the identical elastic energy law (4.72) as above, i.e., we have

$$\frac{E_{el}(\boldsymbol{\sigma}^{n+1}) - E_{el}(\boldsymbol{\sigma}^n)}{\Delta t} - \int_\Omega \boldsymbol{\sigma}^n : \nabla\boldsymbol{u}^{n+1} + \int_\Omega \frac{1}{2\,\tau_S(\varphi^{n+\frac{1}{2}})}\mathrm{tr}(\boldsymbol{\sigma}^{n+\theta}) = 0\,. \tag{4.82}$$

Then, analogously to the proof of Theorem 4.22, taking the inner product of (4.78a) and $\boldsymbol{u}^{n+1}$ and integrating over Ω yields a similar equation as above, but now with an additional integral from the $\boldsymbol{\sigma}$-coupling term, reading

$$\begin{aligned}0 = \frac{1}{2\,\Delta t}\Big(&\|\boldsymbol{u}^{n+1}\|_{L^2(\Omega)}^2 - \|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|_{L^2(\Omega)}^2\Big) \\ &+ \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2 + \int_\Omega \boldsymbol{\sigma}^n : \nabla\boldsymbol{u}^{n+1}\,,\end{aligned} \tag{4.83}$$

where the last term is calculated identically as in the proof of the discrete energy law (4.70). Taking the inner product of relation (4.77b) and $\boldsymbol{u}^*$, integrating over Ω and applying integration by parts yields

$$\begin{aligned}\|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 &= \int_\Omega \boldsymbol{u}^n\cdot\boldsymbol{u}^* - \int_\Omega \Delta t\varphi^n\nabla\mu^{n+\frac{1}{2}}\cdot\boldsymbol{u}^* \\ &= \int_\Omega \boldsymbol{u}^n\cdot\boldsymbol{u}^* + \int_\Omega \Delta t\nabla\cdot(\boldsymbol{u}^*\varphi^n)\,\mu^{n+\frac{1}{2}} - \int_{\partial\Omega}\Delta t\varphi^n\mu^{n+\frac{1}{2}}\boldsymbol{u}^*\cdot\boldsymbol{n}\,.\end{aligned}$$

Assuming suitable boundary conditions, see Remark 3.13, the boundary integral vanishes and we get

$$\|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 = \frac{1}{2}\Big(-\|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^n\|_{L^2(\Omega)}^2\Big) + \int_\Omega \Delta t\,\nabla\cdot(\boldsymbol{u}^*\varphi^n)\,\mu^{n+\frac{1}{2}}\,.$$

Multiplying both sides of the equation by two and subtracting $\|\boldsymbol{u}^*\|_{L^2(\Omega)}^2$ yields

$$\|\boldsymbol{u}^*\|_{L^2(\Omega)}^2 = -\|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^n\|_{L^2(\Omega)}^2 + \int_\Omega 2\Delta t\,\nabla\cdot(\boldsymbol{u}^*\varphi^n)\,\mu^{n+\frac{1}{2}}\,. \tag{4.84}$$

Inserting relation (4.84) in (4.83) yields

$$\begin{aligned}0 &= \frac{1}{2\,\Delta t}\Big(\|\boldsymbol{u}^{n+1}\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 - \|\boldsymbol{u}^n\|_{L^2(\Omega)}^2 - \int_\Omega 2\Delta t\,\nabla\cdot(\boldsymbol{u}^*\varphi^n)\,\mu^{n+\frac{1}{2}} + \|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|_{L^2(\Omega)}^2\Big) \\ &\quad + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2 + \int_\Omega \boldsymbol{\sigma}^n : \nabla\boldsymbol{u}^{n+1} \\ &= \frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \frac{1}{2\,\Delta t}\Big(\|\boldsymbol{u}^{n+1} - \boldsymbol{u}^*\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2\Big) - \int_\Omega \nabla\cdot(\boldsymbol{u}^*\varphi^n)\,\mu^{n+\frac{1}{2}} \\ &\quad + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2 + \int_\Omega \boldsymbol{\sigma}^n : \nabla\boldsymbol{u}^{n+1}\,.\end{aligned}$$

Summed with the discrete energy laws (4.81) and (4.82), the integrals of the coupling terms cancel out and we obtain the desired energy law (4.80). □

4.6.5 Chorin's projection method

To further reduce the computational cost of our splitting scheme in **Step 2**, we propose to use Chorin's projection method analogously as in Subsection 4.4.5, now reading

$$\frac{\boldsymbol{u}^\dagger - \boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^\dagger - \nabla \cdot \left\{ \eta(\varphi^n) \left[\nabla \boldsymbol{u}^\dagger + \left(\nabla \boldsymbol{u}^\dagger\right)^T \right] \right\} + \varphi^n \nabla \mu^{n+\frac{1}{2}} - \nabla \cdot \boldsymbol{\sigma}^n = \boldsymbol{0}\,, \tag{4.85a}$$

$$\boldsymbol{u}^\dagger = \boldsymbol{u}^{n+1} + \Delta t\, \nabla p^{n+1}\,, \tag{4.85b}$$

$$\nabla \cdot \boldsymbol{u}^{n+1} = 0\,. \tag{4.85c}$$

The above system is solved as follows:

1. Find $\boldsymbol{u}^\dagger$ by solving equation (4.85a).

2. Applying the divergence to (4.85b) yields

$$\nabla \cdot \boldsymbol{u}^\dagger = \nabla \cdot \boldsymbol{u}^{n+1} + \Delta t\, \Delta p^{n+1}\,.$$

   Consequently, due to the incompressibility condition (4.85c), we find p^{n+1} by solving the Poisson equation

$$\Delta p^{n+1} = \frac{\nabla \cdot \boldsymbol{u}^\dagger}{\Delta t}\,.$$

3. Since $\boldsymbol{u}^\dagger$ and p^{n+1} are now known, we find $\boldsymbol{u}^{n+1}$ by solving equation (4.85b).

In summary, instead of solving a coupled system for $\boldsymbol{u}^{n+1}$ and p^{n+1} in (4.78), we compute $\boldsymbol{u}^\dagger$, p^{n+1} and $\boldsymbol{u}^{n+1}$ one after another.

Theorem 4.37. *The Chorin projection method* (4.85) *has a truncation error of first order in time and does not influence the discrete energy law for a constant viscosity η. For a variable viscosity, the discrete energy law for the splitting scheme* (4.77), (4.85), (4.79) *becomes*

$$\begin{aligned} \frac{E_{total}(\varphi^{n+1}, q^{n+1}, \boldsymbol{\sigma}^{n+1}, \boldsymbol{u}^{n+1}) - E_{total}(\varphi^n, q^n, \boldsymbol{\sigma}^n, \boldsymbol{u}^n)}{\Delta t} &= -\mathrm{ND}_{pot}^{n+1} - \mathrm{ND}_{chorin}^{n+1} \\ - \int_\Omega \frac{1}{\zeta(\varphi^n)} \left| \varphi^n(1-\varphi^n)\nabla \mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\, q^{n+\frac{1}{2}}) \right|^2 &- \int_\Omega \frac{1}{\tau_B(\varphi^n)} (q^{n+\frac{1}{2}})^2 \\ - \int_\Omega \frac{1}{2\,\tau_S(\varphi^{n+\frac{1}{2}})} \mathrm{tr}(\boldsymbol{\sigma}^{n+\theta}) - \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^\dagger)|^2\,, & \end{aligned} \tag{4.86}$$

where

$$\mathrm{ND}_{chorin}^{n+1} = \frac{\Delta t}{2} \|\nabla p^{n+1}\|_{L^2(\Omega)}^2 + \frac{1}{2\,\Delta t} \left(\|\boldsymbol{u}^\dagger - \boldsymbol{u}^*\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|_{L^2(\Omega)}^2 \right)\,.$$

Proof. Analogously to the proof of Theorem 4.27, inserting relation (4.85b) in equation (4.85a) yields the original Navier-Stokes part (4.78a) plus the following splitting error

$$\begin{aligned}&\Delta t(\boldsymbol{u}^n \cdot \nabla)\nabla p^{n+1} - \Delta t \nabla \cdot \left\{ \eta(\varphi^n) \left[\nabla^2 p^{n+1} + (\nabla^2 p^{n+1})^T \right] \right\} \\ &= \Delta t \left((\boldsymbol{u}^n \cdot \nabla)\nabla p^{n+1} - 2\nabla \cdot \left[\eta(\varphi^n) \nabla^2 p^{n+1} \right] \right) \in \mathcal{O}(\Delta t) .\end{aligned}$$

Assuming a constant viscosity η, we can use the following equality for the last term of the splitting error

$$\eta \nabla \cdot \nabla^2 p^{n+1} = \eta \nabla \cdot (\Delta p^{n+1} \boldsymbol{I}) .$$

Now, taking the inner product of the splitting error and $\boldsymbol{u}^{n+1}$ and integrating over Ω, we can calculate the influence of the splitting error on the discrete energy law. Assuming the same suitable boundary conditions as above, see Remark 3.13, and using the incompressibility yields

$$\begin{aligned}&\Delta t \int_\Omega (\boldsymbol{u}^n \cdot \nabla)\nabla p^{n+1} \cdot \boldsymbol{u}^{n+1} - 2\Delta t \eta \int_\Omega \nabla \cdot (\Delta p^{n+1} \boldsymbol{I}) \cdot \boldsymbol{u}^{n+1} \\ &= -\Delta t \int_\Omega (\nabla \cdot \boldsymbol{u}^n \, \boldsymbol{u}^{n+1} + \boldsymbol{u}^n \nabla \cdot \boldsymbol{u}^{n+1}) \cdot \nabla p^{n+1} + \Delta t \int_{\partial\Omega} (\boldsymbol{u}^n \cdot \boldsymbol{n}) \nabla p^{n+1} \cdot \boldsymbol{u}^{n+1} \\ &\quad + 2\Delta t \eta \int_\Omega \Delta p^{n+1} \boldsymbol{I} : \nabla \boldsymbol{u}^{n+1} - 2\Delta t \eta \int_{\partial\Omega} \Delta p^{n+1} \boldsymbol{u}^{n+1} \cdot \boldsymbol{n} \\ &= 2\Delta t \eta \int_\Omega \Delta p^{n+1} \nabla \cdot \boldsymbol{u}^{n+1} = 0 .\end{aligned}$$

For a variable viscosity $\eta(\varphi^n)$, we substitute (4.77b), reading $\boldsymbol{u}^* = \boldsymbol{u}^n - \Delta t \varphi^n \nabla \mu^{n+\frac{1}{2}}$, in equation (4.85a). Then, we take the inner product with $\boldsymbol{u}^\dagger$ and integrate over Ω, yielding analogously to the derivation of energy law (4.80)

$$\begin{aligned}0 = &\frac{1}{2\,\Delta t} \left(\|\boldsymbol{u}^\dagger\|^2_{L^2(\Omega)} - \|\boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^\dagger - \boldsymbol{u}^*\|^2_{L^2(\Omega)} \right) \\ &+ \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^\dagger)|^2 + \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^{n+1} .\end{aligned} \tag{4.87}$$

Since (4.62b) reads $\boldsymbol{u}^\dagger = \boldsymbol{u}^{n+1} + \Delta t \nabla p^{n+1}$, it holds

$$\begin{aligned}\|\boldsymbol{u}^\dagger\|^2_{L^2(\Omega)} &= \|\boldsymbol{u}^{n+1} + \Delta t \nabla p^{n+1}\|^2_{L^2(\Omega)} \\ &= \|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} + (\Delta t)^2 \|\nabla p^{n+1}\|^2_{L^2(\Omega)} + 2\Delta t \int_\Omega \boldsymbol{u}^{n+1} \cdot \nabla p^{n+1} \\ &= \|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} + (\Delta t)^2 \|\nabla p^{n+1}\|^2_{L^2(\Omega)} - 2\Delta t \int_\Omega \nabla \cdot \boldsymbol{u}^{n+1} \, p^{n+1} + 2\Delta t \int_{\partial\Omega} \boldsymbol{u}^{n+1} \cdot \boldsymbol{n} \, p^{n+1} \\ &= \|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} + (\Delta t)^2 \|\nabla p^{n+1}\|^2_{L^2(\Omega)} ,\end{aligned}$$

where the last equality holds true due to the incompressibility and suitable boundary conditions for $\boldsymbol{u}^{n+1}$, see Remark 3.13. Inserting the last relation into (4.87) reads

$$\begin{aligned}0 = &\frac{1}{2\,\Delta t} \left(\|\boldsymbol{u}^{n+1}\|^2_{L^2(\Omega)} - \|\boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^\dagger - \boldsymbol{u}^*\|^2_{L^2(\Omega)} \right) + \frac{\Delta t}{2} \|\nabla p^{n+1}\|^2_{L^2(\Omega)} \\ &+ \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^\dagger)|^2 + \int_\Omega \boldsymbol{\sigma}^n : \nabla \boldsymbol{u}^{n+1} .\end{aligned}$$

Now, we insert relation (4.84) for $\|\boldsymbol{u}^*\|^2_{L^2(\Omega)}$, yielding the discrete kinetic energy law

$$0 = \frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \frac{1}{2\,\Delta t}\Big(\|\boldsymbol{u}^\dagger - \boldsymbol{u}^*\|^2_{L^2(\Omega)} + \|\boldsymbol{u}^* - \boldsymbol{u}^n\|^2_{L^2(\Omega)}\Big) - \int_\Omega \nabla\cdot(\boldsymbol{u}^*\varphi^n)\,\mu^{n+\frac{1}{2}} + \frac{\Delta t}{2}\|\nabla p^{n+1}\|^2_{L^2(\Omega)} + \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^\dagger)|^2 + \int_\Omega \boldsymbol{\sigma}^n : \nabla\boldsymbol{u}^{n+1}\,.$$

Finally, adding the discrete energy laws (4.81) and (4.82), the integrals of the coupling terms cancel out and we have the desired energy law (4.86). □

Table 4.5: Summary of the schemes for the full model

	linear	en.-stable	mass-cons.	trunc. error	Δt
part. implicit (4.69)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
part. Crank-Nicolson (4.73)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
second order (4.75)	✓	✓	✓	$\mathcal{O}\left((\Delta t)^2\right)$	$\mathcal{O}(h^2)$
splitting scheme (4.77)-(4.79)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
Chorin spl. (4.77), (4.85), (4.79)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$

Remark 4.38. *Summarized, see Table 4.5, the second order Crank-Nicolson type scheme (4.75) has again the best properties on the face of it and we should note that again the two Crank-Nicolson type schemes are not only energy-stable but conserve the continuous energy laws up to the dissipation term* ND_{pot}*, where the latter results compulsorily from the linear approximation of the potential derivative* $f(\varphi)$*. Even though the splitting scheme (4.77), (4.85), (4.79), which uses the Chorin projection, has a larger truncation error and additional numerical dissipation from the splitting and the projection, the decoupled calculation also resulting from splitting and projection yields a major efficiency advantage and more flexibility regarding the spatial discretization of the advection due to its explicit discretization in time. Finally making the latter our method of choice for solving the full model for viscoelastic phase separation (3.47) numerically.*

4.6.6 Stiff stress tensor equation

Note that for the small shear rates $\boldsymbol{D}(\boldsymbol{u})$ and the Weissenberg numbers $\tau_S(\varphi)$ that typically arise in our numerical experiments in Chapter 5, the stiffness of the Oldroyd-B equation does not play a dominant role. For this reason, it is for the most part discretized explicitly in itself, i.e., in σ, in the schemes introduced above, or even completely explicit in itself for $\theta = 0$.

In case of a high Weissenberg number, it can be sufficient to choose $\theta = 1$, such that the relaxation term is computed implicitly. In extreme cases, the high Weissenberg problem can be treated by the addition of stress diffusion, see the diffusive stress equations (3.71) and (3.72), or by using additional techniques like a logarithmic transformation of the conformation tensor formulation. For more details on the latter, see, e. g., *Lukáčová-Medviďová, Notsu and She* [57].

Mixed scheme

For large shear rates $\boldsymbol{D}(\boldsymbol{u})$, an implicit approximation of the elastic shear stress $\boldsymbol{\sigma}$ is advisable. The proposed modification of scheme (4.73) is nonlinear and reads

$$\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla\varphi^n - \nabla\cdot\left\{\frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,, \tag{4.88a}$$

$$\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,$$

$$\frac{q^{n+1}-q^n}{\Delta t} + \boldsymbol{u}^n\cdot\nabla q^{n+\frac{1}{2}} + \frac{1}{\tau_B(\varphi^n)}q^{n+\frac{1}{2}} + G_B(\varphi^n)\nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,, \tag{4.88b}$$

$$\frac{\boldsymbol{\sigma}^{n+1}-\boldsymbol{\sigma}^n}{\Delta t} + (\boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla)\boldsymbol{\sigma}^{n+1} - (\nabla\boldsymbol{u}^{n+\frac{1}{2}})\boldsymbol{\sigma}^{n+1} - \boldsymbol{\sigma}^{n+1}\left(\nabla\boldsymbol{u}^{n+\frac{1}{2}}\right)^T + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}\boldsymbol{\sigma}^{n+1} + G_S(\varphi^{n+\frac{1}{2}})2\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}}) = \boldsymbol{0}\,, \tag{4.88c}$$

$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+\frac{1}{2}} - \nabla\cdot\left\{\eta(\varphi^n)\,2\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})\right\} + \nabla p^{n+\frac{1}{2}} - \mu^{n+\frac{1}{2}}\nabla\varphi^n - \nabla\cdot\boldsymbol{\sigma}^{n+1} = \boldsymbol{0}\,, \tag{4.88d}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,. \tag{4.88e}$$

Theorem 4.39. *Assuming suitable boundary conditions, scheme (4.88) satisfies the discrete energy law (4.74) with $\theta = 1$ and is mass-conservative.*

Proof. The energy law is derived analogously to Theorem 4.34 and the mass conservation is identical, since the φ-equation (4.88a) is identical to (4.73a). □

Note that we can linearize scheme (4.88) by, e. g., using a fixed-point iteration. Due to the highly coupled nature of scheme (4.88), a fixed-point iteration has high computational cost and it is preferable to split the scheme before its application, see the following two subsections.

Further, using the idea presented, e.g., in scheme (4.75) concerning the extrapolation of the explicit terms, while replacing $\boldsymbol{\sigma}^{n+1}$ by the Crank-Nicolson-type approximation $\boldsymbol{\sigma}^{n+\frac{1}{2}}$, one can obtain a nonlinear second order two-step scheme. Alternatively, the whole model can be solved with the Crank-Nicolson method, since scheme (4.88) is nonlinear anyway, yielding a nonlinear second order one-step scheme.

Splitting scheme

Since (4.88a) and (4.88b) are identical to (4.73a) and (4.73b) introduced above, the first step of an appropriate splitting scheme is identical to the first step of splitting scheme (4.77), (4.78), (4.79). The second and third step of the latter scheme are replaced by one combined step in this case, since the implicit discretization of both $\boldsymbol{\sigma}$ and $\boldsymbol{u}$ necessitates a coupled calculation.
Step 1. Find (φ^{n+1}, q^{n+1}) by solving system (4.77).

Step 2*. Find $(\boldsymbol{\sigma}^{n+1}, \boldsymbol{u}^{n+1}, p^{n+1})$ such that

$$\frac{\boldsymbol{\sigma}^{n+1}-\boldsymbol{\sigma}^{n}}{\Delta t} + (\boldsymbol{u}^{n+1}\cdot\nabla)\boldsymbol{\sigma}^{n+1} - (\nabla\boldsymbol{u}^{n+1})\boldsymbol{\sigma}^{n+1} - \boldsymbol{\sigma}^{n+1}\left(\nabla\boldsymbol{u}^{n+1}\right)^{T} + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}\boldsymbol{\sigma}^{n+1} + G_S(\varphi^{n+\frac{1}{2}})2\boldsymbol{D}(\boldsymbol{u}^{n+1}) = \boldsymbol{0}\,, \tag{4.89a}$$

$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^{n}}{\Delta t} + (\boldsymbol{u}^{n}\cdot\nabla)\boldsymbol{u}^{n+1} - \nabla\cdot\left\{\eta(\varphi^{n+\frac{1}{2}})\,2\boldsymbol{D}(\boldsymbol{u}^{n+1})\right\} + \nabla p^{n+1} + \varphi^{n}\nabla\mu^{n+\frac{1}{2}} - \nabla\cdot\boldsymbol{\sigma}^{n+1} = \boldsymbol{0}\,, \tag{4.89b}$$

$$\nabla\cdot\boldsymbol{u}^{n+1} = 0\,. \tag{4.89c}$$

Theorem 4.40. *Assuming suitable boundary conditions, see Remark 3.13, the numerical scheme (4.77), (4.89) is mass-conservative and satisfies the discrete energy law (4.80) with $\theta = 1$.*

Proof. Can be shown analogously to the proof of Theorem (4.36). □

Fixed-point method

It is possible to linearize and split **Step 2*** again by, e. g., using the following fixed-point iteration. Given $\boldsymbol{\sigma}^{n,0} = \boldsymbol{\sigma}^{n}$ and $\boldsymbol{u}^{n,0} = \boldsymbol{u}^{n}$ from the previous time step, we repeat **Step 2** and **3** for $l = 0, 1, ...$, until $||\omega^{n,l+1} - \omega^{n,l}|| \leq \delta||\omega^{n,l}||$, for $\omega \in \{\boldsymbol{\sigma}, \boldsymbol{u}, p\}$ and δ sufficiently small.
Step 2. Find $(\boldsymbol{u}^{n,l+1}, p^{n,l+1})$ such that

$$\frac{\boldsymbol{u}^{n,l+1}-\boldsymbol{u}^{n}}{\Delta t} + (\boldsymbol{u}^{n}\cdot\nabla)\boldsymbol{u}^{n,l+1} - \nabla\cdot\left\{\eta(\varphi^{n+\frac{1}{2}})\,2\boldsymbol{D}(\boldsymbol{u}^{n,l+1})\right\} + \nabla p^{n,l+1} + \varphi^{n}\nabla\mu^{n+\frac{1}{2}} - \nabla\cdot(\boldsymbol{\sigma}^{n,l}) = \boldsymbol{0}\,, \tag{4.90a}$$

$$\nabla\cdot\boldsymbol{u}^{n,l+1} = 0\,, \tag{4.90b}$$

where

$$\boldsymbol{D}(\boldsymbol{u}^{n,l+1}) = \frac{1}{2}\left[\nabla\boldsymbol{u}^{n,l+1} + (\nabla\boldsymbol{u}^{n,l+1})^{T}\right].$$

Step 3. Find $\boldsymbol{\sigma}^{n,l+1}$ such that

$$\frac{\boldsymbol{\sigma}^{n,l+1}-\boldsymbol{\sigma}^{n}}{\Delta t} + (\boldsymbol{u}^{n,l+1}\cdot\nabla)\boldsymbol{\sigma}^{n,l+1} - (\nabla\boldsymbol{u}^{n,l+1})\boldsymbol{\sigma}^{n,l+1} - \boldsymbol{\sigma}^{n,l+1}\left(\nabla\boldsymbol{u}^{n,l+1}\right)^{T} + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}\boldsymbol{\sigma}^{n,l+1} + G_S(\varphi^{n+\frac{1}{2}})2\boldsymbol{D}(\boldsymbol{u}^{n,l+1}) = \boldsymbol{0}\,. \tag{4.91}$$

Step 4. Update solution: $\boldsymbol{u}^{n+1} = \boldsymbol{u}^{n,l+1}$, $p^{n+1} = p^{n,l+1}$, $\boldsymbol{\sigma}^{n+1} = \boldsymbol{\sigma}^{n,l+1}$.

Note that we can also use Chorin's projection method from Subsection 4.6.5 in **Step 2** to further reduce the computational effort.

Table 4.6: Summary of the schemes for the full model with stiff stress tensor

	linear	en.-stable	mass-cons.	trunc. error	Δt
implicit $\boldsymbol{\sigma}$ (4.88)	-	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
splitting scheme (4.77), (4.89)	-	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
fixed-point (4.77), (4.90), (4.91)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$

4.7 Schemes for the full model in conformation tensor formulation

The full model in conformation tensor formulation (3.64) can be discretized analogously to scheme (4.73), reading

$$\begin{aligned}
&\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla\varphi^n && \text{(4.92a)}\\
&\quad - \nabla\cdot\left\{\frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,,\\
&\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,\\
&\frac{q^{n+1}-q^n}{\Delta t} + \boldsymbol{u}^n\cdot\nabla q^{n+\frac{1}{2}} + \frac{1}{\tau_B(\varphi^n)}q^{n+\frac{1}{2}} && \text{(4.92b)}\\
&\quad + G_B(\varphi^n)\nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,,\\
&\frac{\boldsymbol{C}^{n+1}-\boldsymbol{C}^n}{\Delta t} + (\boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla)\boldsymbol{C}^n - (\nabla\boldsymbol{u}^{n+\frac{1}{2}})\boldsymbol{C}^n - \boldsymbol{C}^n\left(\nabla\boldsymbol{u}^{n+\frac{1}{2}}\right)^T + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}(\boldsymbol{C}^n-\boldsymbol{I}) = \boldsymbol{0}\,, && \text{(4.92c)}\\
&\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+\frac{1}{2}} - \nabla\cdot\left\{\eta(\varphi^n)\,2\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})\right\} && \text{(4.92d)}\\
&\quad + \nabla p^{n+1} - \mu^{n+\frac{1}{2}}\nabla\varphi^n - \nabla\cdot[G_S(\boldsymbol{C}^n-\boldsymbol{I})] = \boldsymbol{0}\,,\\
&\nabla\cdot\boldsymbol{u}^{n+1} = 0\,, && \text{(4.92e)}
\end{aligned}$$

where

$$\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}}) = \frac{1}{2}\left[\nabla\boldsymbol{u}^{n+\frac{1}{2}} + (\nabla\boldsymbol{u}^{n+\frac{1}{2}})^T\right].$$

Theorem 4.41. *Let $f(\varphi^{n+1},\varphi^n)$ represent a suitably linearized approximation of $f(\varphi)$, see Table 4.1. Then, the resulting numerical scheme (4.92) is linear and, assuming suitable boundary conditions, see Remark 3.13, it is mass-conservative and satisfies the following discrete version of energy law (3.65)*

$$\begin{aligned}
&\frac{E_{total}(\varphi^{n+1},q^{n+1},\boldsymbol{C}^{n+1},\boldsymbol{u}^{n+1}) - E_{total}(\varphi^n,q^n,\boldsymbol{C}^n,\boldsymbol{u}^n)}{\Delta t} = -\mathrm{ND}_{pot}^{n+1} - \mathrm{ND}_{elastic}^{n+1}\\
&-\int_\Omega \frac{1}{\zeta(\varphi^n)}\left|\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right|^2 - \int_\Omega \frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2\\
&-\int_\Omega \frac{G_S}{2\tau_S(\varphi^{n+\frac{1}{2}})}\mathrm{tr}(\boldsymbol{C}^n + (\boldsymbol{C}^n)^{-1} - 2\boldsymbol{I}) - \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2\,,
\end{aligned} \tag{4.93}$$

where $\operatorname{tr}(\boldsymbol{C}^n + (\boldsymbol{C}^n)^{-1} - 2\boldsymbol{I}) \geq 0$ *due to estimate* (2.5c),

$$\mathrm{ND}_{pot}^{n+1} = \int_\Omega f(\varphi^{n+1}, \varphi^n) \frac{\varphi^{n+1} - \varphi^n}{\Delta t} - \int_\Omega \frac{F(\varphi^{n+1}) - F(\varphi^n)}{\Delta t}$$

and

$$\mathrm{ND}_{elastic}^{n+1} = \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}(\ln \boldsymbol{C}^{n+1}) - \operatorname{tr}(\ln \boldsymbol{C}^n)}{\Delta t} - \int_\Omega \frac{G_S}{2} \frac{\boldsymbol{C}^{n+1} - \boldsymbol{C}^n}{\Delta t} : (\boldsymbol{C}^n)^{-1} \leq 0\,.$$

Because of the latter, scheme (4.92) *is not necessarily energy-stable.*

Proof. The linearity is confirmed analogously to the proof of Theorem 4.33, and since (4.92a) and (4.92b) are identical to (4.69a) and (4.69b), the mass conservation is also directly transfered as well as the sum of the discrete energy laws for mixing and bulk energy, reading

$$\begin{aligned}
&\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} + \frac{E_{bulk}(q^{n+1}) - E_{bulk}(q^n)}{\Delta t} + \mathrm{ND}_{pot}^{n+1} \\
&+ \int_\Omega \frac{1}{\tau_B(\varphi^n)} (q^{n+\frac{1}{2}})^2 + \int_\Omega \frac{1}{\zeta(\varphi^n)} \left| \varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\, q^{n+\frac{1}{2}}) \right|^2 + \int_\Omega \boldsymbol{u}^{n+\frac{1}{2}} \cdot \nabla\varphi^n \mu^{n+\frac{1}{2}} = 0\,.
\end{aligned}$$

The discrete kinetic energy law is calculated analogously to the proof of Theorem 4.20, yielding

$$\frac{E_{kin}(\boldsymbol{u}^{n+1}) - E_{kin}(\boldsymbol{u}^n)}{\Delta t} + \int_\Omega 2\eta(\varphi^n) |\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2 - \int_\Omega \boldsymbol{u}^{n+\frac{1}{2}} \cdot \nabla\varphi^n \mu^{n+\frac{1}{2}} - \int_\Omega \nabla \cdot [G_S(\boldsymbol{C}^n - \boldsymbol{I})] \cdot u^{n+\frac{1}{2}} = 0\,,$$

where for the last term it holds

$$\begin{aligned}
-\int_\Omega \nabla \cdot [G_S(\boldsymbol{C}^n - \boldsymbol{I})] \cdot u^{n+\frac{1}{2}} &= -\int_\Omega G_S\, (\nabla \cdot \boldsymbol{C}^n) \cdot u^{n+\frac{1}{2}} \\
&= \int_\Omega G_S \boldsymbol{C}^n : \nabla \boldsymbol{u}^{n+\frac{1}{2}} - \int_{\partial\Omega} G_S\, (\boldsymbol{C}^n \boldsymbol{n}) \cdot u^{n+\frac{1}{2}} \\
&= \int_\Omega G_S \boldsymbol{C}^n : \nabla \boldsymbol{u}^{n+\frac{1}{2}}\,.
\end{aligned}$$

Taking the double dot product of (4.92c) and $\frac{G_S}{2}\left(\boldsymbol{I} - (\boldsymbol{C}^n)^{-1}\right)$, the discrete elastic energy law is, except for the first term, calculated analogously to the continuous case in the proof of Theorem 3.18, yielding

$$\begin{aligned}
0 = &\int_\Omega \frac{\boldsymbol{C}^{n+1} - \boldsymbol{C}^n}{\Delta t} : \frac{G_S}{2}\left(\boldsymbol{I} - (\boldsymbol{C}^n)^{-1}\right) \\
&- \int_\Omega G_S \boldsymbol{C}^n : \nabla \boldsymbol{u}^{n+\frac{1}{2}} + \int_\Omega \frac{G_S}{2\tau_S(\varphi)} \operatorname{tr}(\boldsymbol{C}^n - \boldsymbol{I}) + \int_\Omega \frac{G_S}{2\tau_S(\varphi)} \operatorname{tr}\left((\boldsymbol{C}^n)^{-1} - \boldsymbol{I}\right) \\
= &\int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}(\boldsymbol{C}^{n+1}) - \operatorname{tr}(\boldsymbol{C}^n)}{\Delta t} - \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}\left((\boldsymbol{C}^{n+1} - \boldsymbol{C}^n)(\boldsymbol{C}^n)^{-1}\right)}{\Delta t} \\
&- \int_\Omega G_S \boldsymbol{C}^n : \nabla \boldsymbol{u}^{n+\frac{1}{2}} + \int_\Omega \frac{G_S}{2\tau_S(\varphi)} \operatorname{tr}\left(\boldsymbol{C}^n + (\boldsymbol{C}^n)^{-1} - 2\boldsymbol{I}\right)\,.
\end{aligned} \tag{4.94}$$

Since the discrete elastic energy at time t_n reads

$$E_{el}(\boldsymbol{C}^n) = \int_\Omega \frac{G_S}{2} \operatorname{tr}(\boldsymbol{C}^n - \ln \boldsymbol{C}^n - \boldsymbol{I})\,,$$

we sum equation (4.94) with

$$0 = \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}(-\ln \boldsymbol{C}^{n+1} + \ln \boldsymbol{C}^n)}{\Delta t} + \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}(\ln \boldsymbol{C}^{n+1} - \ln \boldsymbol{C}^n)}{\Delta t},$$

yielding

$$\begin{aligned}
0 = &\int_\Omega \frac{G_S}{2} \frac{E_{el}(\boldsymbol{C}^{n+1}) - E_{el}(\boldsymbol{C}^n)}{\Delta t} + \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}(\ln \boldsymbol{C}^{n+1} - \ln \boldsymbol{C}^n)}{\Delta t} - \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}\left((\boldsymbol{C}^{n+1} - \boldsymbol{C}^n)(\boldsymbol{C}^n)^{-1}\right)}{\Delta t} \\
&- \int_\Omega G_S \boldsymbol{C}^n : \nabla \boldsymbol{u}^{n+\frac{1}{2}} + \int_\Omega \frac{G_S}{2\tau_S(\varphi)} \operatorname{tr}\left(\boldsymbol{C}^n + (\boldsymbol{C}^n)^{-1} - 2\boldsymbol{I}\right) \\
= &\int_\Omega \frac{G_S}{2} \frac{E_{el}(\boldsymbol{C}^{n+1}) - E_{el}(\boldsymbol{C}^n)}{\Delta t} + \mathrm{ND}_{elastic}^{n+1} \\
&- \int_\Omega G_S \boldsymbol{C}^n : \nabla \boldsymbol{u}^{n+\frac{1}{2}} + \int_\Omega \frac{G_S}{2\tau_S(\varphi)} \operatorname{tr}\left(\boldsymbol{C}^n + (\boldsymbol{C}^n)^{-1} - 2\boldsymbol{I}\right).
\end{aligned}$$

Summing up the above discrete energy laws yields energy law (4.93).

Due to estimate (2.5e), we have

$$\operatorname{tr}\left((\boldsymbol{C}^{n+1} - \boldsymbol{C}^n)(\boldsymbol{C}^n)^{-1}\right) \geq \operatorname{tr}\left(\ln \boldsymbol{C}^{n+1} - \ln \boldsymbol{C}^n\right),$$

and therefore also

$$\int_\Omega \operatorname{tr}\left((\boldsymbol{C}^{n+1} - \boldsymbol{C}^n)(\boldsymbol{C}^n)^{-1}\right) \geq \int_\Omega \operatorname{tr}\left(\ln \boldsymbol{C}^{n+1} - \ln \boldsymbol{C}^n\right).$$

Thus, it holds

$$\begin{aligned}
\mathrm{ND}_{elastic}^{n+1} &= \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}(\ln \boldsymbol{C}^{n+1} - \ln \boldsymbol{C}^n)}{\Delta t} - \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}\left((\boldsymbol{C}^{n+1} - \boldsymbol{C}^n)(\boldsymbol{C}^n)^{-1}\right)}{\Delta t} \\
&\leq \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}(\ln \boldsymbol{C}^{n+1} - \ln \boldsymbol{C}^n)}{\Delta t} - \int_\Omega \frac{G_S}{2} \frac{\operatorname{tr}(\ln \boldsymbol{C}^{n+1} - \ln \boldsymbol{C}^n)}{\Delta t} = 0.
\end{aligned}$$

□

Obviously, $\mathrm{ND}_{elastic}^{n+1}$ in energy law (4.93) has the wrong sign. It is possible to change the sign by calculating $\boldsymbol{C}$ implicitly and thus sacrificing the linearity while gaining robustness towards large shear rates $\boldsymbol{D}(\boldsymbol{u})$, see Subsection 4.6.6. The respective nonlinear one-step scheme is

analogous to scheme (4.88) and reads

$$\begin{aligned}
&\frac{\varphi^{n+1}-\varphi^n}{\Delta t} + \boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla\varphi^n && \text{(4.95a)}\\
&\qquad - \nabla\cdot\left\{\frac{\varphi^n(1-\varphi^n)}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,,\\
&\mu^{n+\frac{1}{2}} = -\lambda\Delta\varphi^{n+\frac{1}{2}} + f(\varphi^{n+1},\varphi^n)\,,\\
&\frac{q^{n+1}-q^n}{\Delta t} + \boldsymbol{u}^n\cdot\nabla q^{n+\frac{1}{2}} + \frac{1}{\tau_B(\varphi^n)}q^{n+\frac{1}{2}} && \text{(4.95b)}\\
&\qquad + G_B(\varphi^n)\nabla\cdot\left\{\frac{1}{\zeta(\varphi^n)}\left[\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right]\right\} = 0\,,\\
&\frac{\boldsymbol{C}^{n+1}-\boldsymbol{C}^n}{\Delta t} + (\boldsymbol{u}^{n+\frac{1}{2}}\cdot\nabla)\boldsymbol{C}^{n+1} && \text{(4.95c)}\\
&\qquad - (\nabla\boldsymbol{u}^{n+\frac{1}{2}})\boldsymbol{C}^{n+1} - \boldsymbol{C}^{n+1}\left(\nabla\boldsymbol{u}^{n+\frac{1}{2}}\right)^T + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}(\boldsymbol{C}^{n+1}-\boldsymbol{I}) = \boldsymbol{0}\,,\\
&\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t} + (\boldsymbol{u}^n\cdot\nabla)\boldsymbol{u}^{n+\frac{1}{2}} - \nabla\cdot\left\{\eta(\varphi^n)\,2\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})\right\} && \text{(4.95d)}\\
&\qquad + \nabla p^{n+1} - \mu^{n+\frac{1}{2}}\nabla\varphi^n - \nabla\cdot\left[G_S(\boldsymbol{C}^{n+1}-\boldsymbol{I})\right] = \boldsymbol{0}\,,\\
&\nabla\cdot\boldsymbol{u}^{n+1} = 0\,, && \text{(4.95e)}
\end{aligned}$$

Theorem 4.42. *Assuming suitable boundary conditions, see Remark 3.13, scheme (4.95) is mass-conservative and satisfies the following discrete version of energy law (3.65)*

$$\begin{aligned}
&\frac{E_{total}(\varphi^{n+1},q^{n+1},\boldsymbol{C}^{n+1},\boldsymbol{u}^{n+1}) - E_{total}(\varphi^n,q^n,\boldsymbol{C}^n,\boldsymbol{u}^n)}{\Delta t} = -\mathrm{ND}_{pot}^{n+1} - \mathrm{ND}_{elastic}^{n+1}\\
&-\int_\Omega \frac{1}{\zeta(\varphi^n)}\left|\varphi^n(1-\varphi^n)\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^n)\,q^{n+\frac{1}{2}})\right|^2 - \int_\Omega \frac{1}{\tau_B(\varphi^n)}(q^{n+\frac{1}{2}})^2 \qquad \text{(4.96)}\\
&-\int_\Omega \frac{G_S}{2\tau_S(\varphi^{n+\frac{1}{2}})}\mathrm{tr}(\boldsymbol{C}^{n+1} + (\boldsymbol{C}^{n+1})^{-1} - 2\boldsymbol{I}) - \int_\Omega 2\eta(\varphi^n)|\boldsymbol{D}(\boldsymbol{u}^{n+\frac{1}{2}})|^2\,,
\end{aligned}$$

where $\mathrm{tr}(\boldsymbol{C}^{n+1} + (\boldsymbol{C}^{n+1})^{-1} - 2\boldsymbol{I}) \geq 0$ *due to estimate (2.5c) and*

$$\mathrm{ND}_{elastic}^{n+1} = \int_\Omega \frac{G_S}{2}\frac{\mathrm{tr}(\ln\boldsymbol{C}^{n+1}) - \mathrm{tr}(\ln\boldsymbol{C}^n)}{\Delta t} - \int_\Omega \frac{G_S}{2}\frac{\boldsymbol{C}^{n+1}-\boldsymbol{C}^n}{\Delta t} : (\boldsymbol{C}^{n+1})^{-1} \geq 0\,.$$

Proof. The mass conservation and the discrete energy law are calculated analogously to the proof of Theorem 4.41, except that we take the double dot product of (4.95c) and $\frac{G_S}{2}\left(\boldsymbol{I} - (\boldsymbol{C}^{n+1})^{-1}\right)$. Consequently, since using estimate (2.5e), it holds

$$\mathrm{tr}\left((\boldsymbol{C}^{n+1}-\boldsymbol{C}^n)(\boldsymbol{C}^{n+1})^{-1}\right) \leq \mathrm{tr}\left(\ln\boldsymbol{C}^{n+1} - \ln\boldsymbol{C}^n\right)\,,$$

we have

$$\begin{aligned}
\mathrm{ND}_{elastic}^{n+1} &= \int_\Omega \frac{G_S}{2}\frac{\mathrm{tr}(\ln\boldsymbol{C}^{n+1} - \ln\boldsymbol{C}^n)}{\Delta t} - \int_\Omega \frac{G_S}{2}\frac{\mathrm{tr}\left((\boldsymbol{C}^{n+1}-\boldsymbol{C}^n)(\boldsymbol{C}^{n+1})^{-1}\right)}{\Delta t}\\
&\geq \int_\Omega \frac{G_S}{2}\frac{\mathrm{tr}(\ln\boldsymbol{C}^{n+1} - \ln\boldsymbol{C}^n)}{\Delta t} - \int_\Omega \frac{G_S}{2}\frac{\mathrm{tr}(\ln\boldsymbol{C}^{n+1} - \ln\boldsymbol{C}^n)}{\Delta t} = 0\,.
\end{aligned}$$

□

4.7.1 Splitting scheme

Since (4.95a) and (4.95b) are again identical to (4.73a) and (4.73b) introduced above, the first step of an appropriate splitting scheme is given by

Step 1. Solve (4.77) for (φ^{n+1}, q^{n+1}).

Since (4.95c) and (4.95d) are coupled due to the implicit calculation of both equations, the second step reads

Step 2*. Find $(\boldsymbol{C}^{n+1}, \boldsymbol{u}^{n+1}, p^{n+1})$, such that

$$\begin{aligned}
&\frac{\boldsymbol{C}^{n+1}-\boldsymbol{C}^{n}}{\Delta t} + (\boldsymbol{u}^{n+1}\cdot\nabla)\boldsymbol{C}^{n+1} && \text{(4.97a)}\\
&\qquad - (\nabla\boldsymbol{u}^{n+1})\boldsymbol{C}^{n+1} - \boldsymbol{C}^{n+1}\left(\nabla\boldsymbol{u}^{n+1}\right)^{T} + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}(\boldsymbol{C}^{n+1}-\boldsymbol{I}) = \boldsymbol{0}\,,\\
&\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^{*}}{\Delta t} + (\boldsymbol{u}^{n}\cdot\nabla)\boldsymbol{u}^{n+1} - \nabla\cdot\left\{\eta(\varphi^{n+\frac{1}{2}})\,2\boldsymbol{D}(\boldsymbol{u}^{n+1})\right\} && \text{(4.97b)}\\
&\qquad + \nabla p^{n+1} - \nabla\cdot\left[G_S(\boldsymbol{C}^{n+1}-\boldsymbol{I})\right] = \boldsymbol{0}\,,\\
&\nabla\cdot\boldsymbol{u}^{n+1} = 0\,, && \text{(4.97c)}
\end{aligned}$$

where

$$\boldsymbol{u}^{*} = \boldsymbol{u}^{n} - \Delta t\varphi^{n}\nabla\mu^{n+\frac{1}{2}}\,.$$

Theorem 4.43. *Assuming suitable boundary conditions, see Remark 3.13, the numerical scheme (4.77), (4.97) is mass-conservative and satisfies the discrete energy law*

$$\begin{aligned}
&\frac{E_{total}(\varphi^{n+1}, q^{n+1}, \boldsymbol{C}^{n+1}, \boldsymbol{u}^{n+1}) - E_{total}(\varphi^{n}, q^{n}, \boldsymbol{C}^{n}, \boldsymbol{u}^{n})}{\Delta t} = -\mathrm{ND}_{pot}^{n+1} - \mathrm{ND}_{elastic}^{n+1} - \mathrm{ND}_{split}^{n+1}\\
&-\int_\Omega \frac{1}{\zeta(\varphi^{n})}\left|\varphi^{n}(1-\varphi^{n})\nabla\mu^{n+\frac{1}{2}} - \nabla(G_B(\varphi^{n})\,q^{n+\frac{1}{2}})\right|^2 - \int_\Omega \frac{1}{\tau_B(\varphi^{n})}(q^{n+\frac{1}{2}})^2 \qquad \text{(4.98)}\\
&-\int_\Omega \frac{G_S}{2}\mathrm{tr}(\boldsymbol{C}^{n+1} + (\boldsymbol{C}^{n+1})^{-1} - 2\boldsymbol{I}) - \int_\Omega 2\eta(\varphi^{n+\frac{1}{2}})|\boldsymbol{D}(\boldsymbol{u}^{n+1})|^2\,,
\end{aligned}$$

where

$$\mathrm{ND}_{elastic}^{n+1} = \int_\Omega \frac{G_S}{2}\frac{\mathrm{tr}(\ln\boldsymbol{C}^{n+1}) - \mathrm{tr}(\ln\boldsymbol{C}^{n})}{\Delta t} - \int_\Omega \frac{G_S}{2}\frac{\boldsymbol{C}^{n+1}-\boldsymbol{C}^{n}}{\Delta t} : (\boldsymbol{C}^{n+1})^{-1} \geq 0$$

and

$$\mathrm{ND}_{split}^{n+1} = \frac{1}{2\,\Delta t}\left(\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^{*}\|_{L^2(\Omega)}^2 + \|\boldsymbol{u}^{*}-\boldsymbol{u}^{n}\|_{L^2(\Omega)}^2\right).$$

Proof. The mass conservation is identical to and the energy law is derived analogously to the proof of Theorem 4.36, except for the elastic energy, which is derived analogously to the proof of Theorem 4.42. □

4.7.2 Fixed-point method

It is possible to linearize and split **Step 2*** again by using a fixed-point iteration. Given $\boldsymbol{C}^{n,0} = \boldsymbol{C}^{n}$ and $\boldsymbol{u}^{n,0} = \boldsymbol{u}^{n}$ from the previous time step, we repeat **Step 2** and **3** for $l = 0, 1, ...,$

until $\|\omega^{n,l+1} - \omega^{n,l}\| \leq \delta \|\omega^{n,l}\|$ for $\omega \in \{\boldsymbol{C}, \boldsymbol{u}\}$ and δ sufficiently small.
Step 2 Find $(\boldsymbol{u}^{n,l+1}, p^{n,l+1})$ such that

$$\begin{aligned} \frac{\boldsymbol{u}^{n,l+1} - \boldsymbol{u}^*}{\Delta t} &+ (\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^{n,l+1} - \nabla \cdot \left\{ \eta(\varphi^{n+\frac{1}{2}})\, 2\boldsymbol{D}(\boldsymbol{u}^{n,l+1}) \right\} \\ &+ \nabla p^{n,l+1} - \nabla \cdot \left[G_S(\boldsymbol{C}^{n,l} - \boldsymbol{I}) \right] = \boldsymbol{0}\,, \qquad (4.99a) \\ \nabla \cdot \boldsymbol{u}^{n,l+1} &= 0\,, \qquad (4.99b) \end{aligned}$$

where

$$\boldsymbol{D}(\boldsymbol{u}^{n,l+1}) = \frac{1}{2}\left[\nabla \boldsymbol{u}^{n,l+1} + (\nabla \boldsymbol{u}^{n,l+1})^T \right].$$

Step 3 Find $\boldsymbol{C}^{n,l+1}$ such that

$$\begin{aligned} &\frac{\boldsymbol{C}^{n,l+1} - \boldsymbol{C}^n}{\Delta t} + (\boldsymbol{u}^{n,l+1} \cdot \nabla)\boldsymbol{C}^{n,l+1} \\ &\qquad - (\nabla \boldsymbol{u}^{n,l+1})\boldsymbol{C}^{n,l+1} - \boldsymbol{C}^{n,l+1} \left(\nabla \boldsymbol{u}^{n,l+1}\right)^T + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}(\boldsymbol{C}^{n,l+1} - \boldsymbol{I}) = \boldsymbol{0}\,. \end{aligned} \qquad (4.100)$$

Step 4 Update solution: $\boldsymbol{u}^{n+1} = \boldsymbol{u}^{n,l+1}$, $p^{n+1} = p^{n,l+1}$, $\boldsymbol{C}^{n+1} = \boldsymbol{C}^{n,l+1}$.

Note that we can also use Chorin's projection method from Subsection 4.6.5 in **Step 2** to further reduce the computational cost.

Table 4.7: Summary of the schemes for the full model in conformation tensor formulation

	linear	en.-stable	mass-cons.	trunc. error	Δt
explicit $\boldsymbol{C}$ (4.92)	✓	-	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
implicit $\boldsymbol{C}$ (4.95)	-	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
splitting scheme (4.77), (4.97)	-	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$
fixed-point (4.77), (4.99), (4.100)	✓	✓	✓	$\mathcal{O}(\Delta t)$	$\mathcal{O}(h^2)$

Let us point out that we have only investigated the thermodynamic consistency of semi-discrete schemes up to this stage. In order to investigate full discretizations, we will at first introduce problem-suited spatial discretizations in the following.

4.8 Spatial and full discretization

Our spatial discretization is either carried out solely by finite differences or by a combined finite volume - finite difference scheme on a staggered grid. Note that we only consider rectangular computational domains, which is typical in the literature studying viscoelastic phase separation and meets the situation of many molecular dynamics simulations as well as real world experiments. Due to this rather simple domain structure, a spatial discretization by finite differences is not only suitable but computationally very efficient. Since finite differences lack in capturing advection and convection accurately, we will additionally introduce upwind finite volume schemes for the discretization of the advection and convection terms. For an

accurate upwinding as well as for the above introduced pressure correction method, it is favorable to use a staggered grid, which is structured as follows.

The degrees of freedom for the components of the velocity vector $\boldsymbol{u}$ are the centers of the cell edges in two dimensions and of the cell faces in three dimensions. More detailed, in two space dimensions, the x-velocity component is given at the centers of vertical cell edges, while the y-velocity component is given at the centers of the horizontal cell edges. Consequently, the velocity components are piecewise linear in one direction and constant in the other. The other functions $(\varphi, q, \boldsymbol{\sigma}, p)$ are piecewise constant, since they are evaluated at the cell centers. This is analogous to the Marker and Cell (MAC) method by *Harlow and Welch* [42].

Again in two space dimensions, we consider a partition of our rectangular computational domain $\Omega \subset \mathbb{R}^2$ into a regular grid Ω_h with $N_x \times N_y$ grid cells and denote $\left[x_{i-1/2}, x_{i+1/2}\right] \times \left[y_{j-1/2}, y_{j+1/2}\right]$, $i \in \{1, 2, ..., N_x\}$, $j \in \{1, 2, ..., N_y\}$, as the i-th grid cell in x- and the j-th grid cell in y-direction. Hence, the height of each grid cell in x direction is given by $h_x = x_{i+1/2} - x_{i-1/2}$, which is equal for all i, and the height in y direction by $h_y = y_{j+1/2} - y_{j-1/2}$, which is equal for all j. Thus, each grid cell has the same size $h_x \times h_y$ such that the grid is regular.

The partition in three space dimensions is analogous, with z denoting the third dimension.

4.8.1 Finite difference derivation

Finite differences are approximate derivatives, which can be constructed by Taylor series expansions of a solution $\omega(x_i + kh) \in \mathbb{R}$ in one space dimension, where $k \in \mathbb{Q}$ and $h = x_{i+1/2} - x_{i-1/2}, i \in \{1, 2, ..., N\}$. Assuming $\omega \in C^4(\Omega, \mathbb{R})$, the Taylor formula yields

$$\omega(x_i + kh) = \omega(x_i) + kh\omega'(x_i) + \frac{(kh)^2}{2}\omega''(x_i) + \frac{(kh)^3}{6}\omega'''(x_i) + \mathcal{O}(h^4)\,. \tag{4.101}$$

Substituting the solution $\omega(x_i)$ by its approximation ω_i, $\omega(x_i + kh)$ by ω_{i+k} and the derivatives $\omega^{(l)}(x_i), l \in \mathbb{N}$, by $\partial^l \omega_i$, we have

$$\omega_{i+k} = \omega_i + kh\partial\omega_i + \frac{(kh)^2}{2}\partial^2\omega_i + \frac{(kh)^3}{6}\partial^3\omega_i + \mathcal{O}(h^4)\,, \tag{4.102}$$

which can be rearranged to

$$\partial\omega_i = \frac{\omega_{i+k} - \omega_i}{kh} + \frac{kh}{2}\partial^2\omega_i + \frac{(kh)^2}{6}\partial^3\omega_i + \mathcal{O}(h^3)\,. \tag{4.103}$$

Setting $k = 1$ yields

$$\partial^R\omega_i = \frac{\omega_{i+1} - \omega_i}{h} + \frac{h}{2}\partial^2\omega_i + \frac{h^2}{6}\partial^3\omega_i + \mathcal{O}(h^3) \tag{4.104}$$

$$\implies \partial^R\omega_i = \frac{\omega_{i+1} - \omega_i}{h} + \mathcal{O}(h)\,, \tag{4.105}$$

which is called the right difference and has obviously spatial error-order one.
Analogously, for $k = -1$, we get the left difference

$$\partial^L\omega_i = \frac{\omega_i - \omega_{i-1}}{h} - \frac{h}{2}\partial^2\omega_i + \frac{h^2}{6}\partial^3\omega_i + \mathcal{O}(h^3) \tag{4.106}$$

$$\implies \partial^L\omega_i = \frac{\omega_i - \omega_{i-1}}{h} + \mathcal{O}(h)\,, \tag{4.107}$$

which has error-order one as well. By summing up equations (4.104) and (4.106), the third Taylor link gets canceled out and the resulting central finite difference

$$\partial\omega_i = \frac{\omega_{i+1} - \omega_{i-1}}{2h} + \frac{h^2}{3}\partial^3\omega_i + \mathcal{O}(h^3) \tag{4.108}$$

$$\implies \quad \partial\omega_i = \frac{\omega_{i+1} - \omega_{i-1}}{2h} + \mathcal{O}(h^2) \tag{4.109}$$

is of second order. Analogously, for setting $k = \pm 1/2$ in equation (4.103) and by summing up both, we get the alternative central finite difference

$$\tilde{\partial}\omega_i = \frac{\omega_{i+1/2} - \omega_{i-1/2}}{h} + \mathcal{O}(h^2)\,. \tag{4.110}$$

To derive a finite difference approximation for the second derivative, we can sum up the discrete Taylor series expansion (4.102) for $k = \pm 1$, yielding

$$\omega_{i+1} + \omega_{i-1} = 2\omega_i + h^2\partial^2\omega_i + \mathcal{O}(h^4) \tag{4.111}$$

$$\iff \quad \partial^2\omega_i = \frac{\omega_{i+1} - 2\omega_i + \omega_{i-1}}{h^2} + \mathcal{O}(h^2)\,. \tag{4.112}$$

4.8.2 Two-dimensional finite differences

Since we are interested in higher dimensional numerical solutions and our models include divergences and gradients, which coincide in one space dimension, but are different differential operators in higher dimensions, it is reasonable to introduce finite differences for higher dimensions as well. Note that the proposed finite differences are in three dimensions analogous to the two-dimensional case. Therefore, only the latter will be introduced in the following.

We approximate the exact solution $\omega(x_i, y_j), \omega \in \{\varphi, q, \sigma_{k,l}, p\}, k, l = 1, 2$, by piecewise constant data $\omega_{i,j}$ in each grid cell.
The first order spatial derivative $\nabla\omega(x_i, y_j)$ is approximated using

$$\nabla_h\omega_{i,j} = \begin{pmatrix} \partial_x\omega_{i,j} \\ \partial_y\omega_{i,j} \end{pmatrix}, \tag{4.113}$$

where $\partial_x\omega_{i,j}$ and $\partial_y\omega_{i,j}$ are central finite differences, which are analogous to (4.109) and read

$$\partial_x\omega_{i,j} = \frac{\omega_{i+1,j} - \omega_{i-1,j}}{2h_x}\,, \qquad \partial_y\omega_{i,j} = \frac{\omega_{i,j+1} - \omega_{i,j-1}}{2h_y}\,. \tag{4.114}$$

This central differences can also be used to approximate the divergence of a vectorial function $\boldsymbol{u}(x_i, y_j) \approx \boldsymbol{u}_{ij} = (u_{ij}, v_{ij})^T$, i.e., $\nabla \cdot \boldsymbol{u}(x_i, y_j)$ is approximated by

$$\nabla_h \cdot \boldsymbol{u}_{ij} = \partial_x u_{ij} + \partial_y v_{ij}\,. \tag{4.115}$$

Further, the Laplacian $\Delta\omega(x_i, y_j)$ is approximated analogously to (4.112) by using the second order central finite difference

$$\Delta_h\omega_{i,j} = \frac{\omega_{i+1,j} - 2\omega_{i,j} + \omega_{i-1,j}}{h_x^2} + \frac{\omega_{i,j+1} - 2\omega_{i,j} + \omega_{i,j-1}}{h_y^2}\,. \tag{4.116}$$

Since this finite difference includes function values of one adjacent grid cell in each direction, it has a so-called 3×3 difference star.

The (φ, q)-equations include spatial derivatives of fourth order, which are – considering a variable mobility $M(\varphi)$ – given by $\nabla \cdot \{M(\varphi) \nabla [\Delta \varphi]\}$. Discretized by the above-mentioned finite differences, we have a 7×7 difference star, since each of the three discrete differences expands the difference star by one grid cell in each direction.

Considering a constant mobility, the fourth derivative holds

$$M \nabla \cdot \{\nabla [\Delta \varphi]\} = M \Delta [\Delta \varphi] = M \Delta^2 \varphi \,.$$

Thus, it can be calculated with a 5×5 difference star using solely finite difference (4.116), which causes less artificial diffusion.

In order to have a 5×5 difference star for any mobility, let us introduce the following finite differences for the gradient $\tilde{\nabla}_h$

$$\tilde{\partial}_x \omega_{ij} = \frac{\omega_{i+1/2,j} - \omega_{i-1/2,j}}{h_x} \,, \qquad \tilde{\partial}_y \omega_{ij} = \frac{\omega_{i,j+1/2} - \omega_{i,j-1/2}}{h_y} \,, \tag{4.117}$$

together with the following two shifted ones for the divergence $\tilde{\nabla}_h \cdot$

$$\tilde{\partial}_x \omega_{i-1/2,j} = \frac{\omega_{ij} - \omega_{i-1,j}}{h_x} \,, \qquad \tilde{\partial}_y \omega_{i,j-1/2} = \frac{\omega_{ij} - \omega_{i,j-1}}{h_y} \,, \tag{4.118}$$

yielding

$$\begin{aligned}
\tilde{\nabla}_h \cdot \tilde{\nabla}_h \omega_{ij} &= \tilde{\nabla}_h \cdot \begin{pmatrix} \tilde{\partial}_x \omega_{ij} \\ \tilde{\partial}_y \omega_{ij} \end{pmatrix} \\
&= \tilde{\nabla}_h \cdot \begin{pmatrix} \frac{\omega_{i+1/2,j} - \omega_{i-1/2,j}}{h_x} \\ \frac{\omega_{i,j+1/2} - \omega_{i,j-1/2}}{h_y} \end{pmatrix} \\
&= \tilde{\partial}_x \frac{\omega_{i+1/2,j} - \omega_{i-1/2,j}}{h_x} + \tilde{\partial}_y \frac{\omega_{i,j+1/2} - \omega_{i,j-1/2}}{h_y} \\
&= \frac{\frac{\omega_{i+1,j} - \omega_{ij}}{h_x} - \frac{\omega_{ij} - \omega_{i-1,j}}{h_x}}{h_x} + \frac{\frac{\omega_{i,j+1} - \omega_{ij}}{h_y} - \frac{\omega_{ij} - \omega_{i,j-1}}{h_y}}{h_y} \\
&= \frac{\omega_{i+1,j} - 2\omega_{i,j} + \omega_{i-1,j}}{h_x^2} + \frac{\omega_{i,j+1} - 2\omega_{i,j} + \omega_{i,j-1}}{h_y^2} = \Delta_h \omega_{ij} \,.
\end{aligned} \tag{4.119}$$

Thus, we have the same 3×3 difference star as the central finite difference (4.116) for the Laplacian, even though we apply two separate finite differences in a row.

The same discretization considering a variable mobility $M_{ij} := M(\varphi_{ij})$ reads

$$\begin{aligned}
\tilde{\nabla}_h \cdot \left[M_{ij}\tilde{\nabla}_h\omega_{ij}\right] &= \tilde{\partial}_x\left[M_{ij}\frac{\omega_{i+1/2,j}-\omega_{i-1/2,j}}{h_x}\right] + \tilde{\partial}_y\left[M_{ij}\frac{\omega_{i,j+1/2}-\omega_{i,j-1/2}}{h_y}\right] \\
&= \frac{M_{i+1/2,j}\frac{\omega_{i+1,j}-\omega_{ij}}{h_x} - M_{i-1/2,j}\frac{\omega_{ij}-\omega_{i-1,j}}{h_x}}{h_x} + \frac{M_{i,j+1/2}\frac{\omega_{i,j+1}-\omega_{ij}}{h_y} - M_{i,j-1/2}\frac{\omega_{ij}-\omega_{i,j-1}}{h_y}}{h_y} \\
&= \frac{M_{i+1/2,j}\omega_{i+1,j} - (M_{i+1/2,j}+M_{i-1/2,j})\omega_{i,j} + M_{i-1/2,j}\omega_{i-1,j}}{h_x^2} \\
&\quad + \frac{M_{i,j+1/2}\omega_{i,j+1} - (M_{i,j+1/2}+M_{i,j-1/2})\omega_{i,j} + M_{i,j-1/2}\omega_{i,j-1}}{h_y^2},
\end{aligned} \tag{4.120}$$

where

$$M_{i+1/2,j} = \frac{M_{i+1,j}+M_{ij}}{2} \qquad \text{and} \qquad M_{i,j+1/2} = \frac{M_{i,j+1}+M_{ij}}{2}.$$

To get rid of the $i, j\pm 1/2$ discretizations in the above calculations, we can replace the differences $\tilde{\partial}$ by the following left and right differences

$$\partial_x^L\omega_{ij} = \frac{\omega_{ij}-\omega_{i-1,j}}{h_x}, \qquad \partial_x^R\omega_{ij} = \frac{\omega_{i+1,j}-\omega_{ij}}{h_x}, \tag{4.121}$$

$$\partial_y^L\omega_{ij} = \frac{\omega_{ij}-\omega_{i,j-1}}{h_y}, \qquad \partial_y^R\omega_{ij} = \frac{\omega_{i,j+1}-\omega_{ij}}{h_y}, \tag{4.122}$$

which are originally only of first order. However, since they match the shifted differences (4.118), they become the same second order approximation as equation (4.119), if applied in the correct order, reading

$$\begin{aligned}
\nabla_h^L \cdot \nabla_h^R\omega_{ij} &= \nabla_h^L \cdot \begin{pmatrix} \partial_x^R\omega_{ij} \\ \partial_y^R\omega_{ij} \end{pmatrix} \\
&= \nabla_h^L \cdot \begin{pmatrix} \frac{\omega_{i+1,j}-\omega_{ij}}{h_x} \\ \frac{\omega_{i,j+1}-\omega_{ij}}{h_y} \end{pmatrix} \\
&= \partial_x^L\frac{\omega_{i+1,j}-\omega_{ij}}{h_x} + \partial_y^L\frac{\omega_{i,j+1}-\omega_{ij}}{h_y} \\
&= \frac{\frac{\omega_{i+1,j}-\omega_{ij}}{h_x} - \frac{\omega_{ij}-\omega_{i-1,j}}{h_x}}{h_x} + \frac{\frac{\omega_{i,j+1}-\omega_{ij}}{h_y} - \frac{\omega_{ij}-\omega_{i,j-1}}{h_y}}{h_y} \\
&= \frac{\omega_{i+1,j}-2\omega_{i,j}+\omega_{i-1,j}}{h_x^2} + \frac{\omega_{i,j+1}-2\omega_{i,j}+\omega_{i,j-1}}{h_y^2} = \Delta_h\omega_{ij}.
\end{aligned} \tag{4.123}$$

This extends to variable mobilities, if the mobility is shifted suitably as follows

$$
\begin{aligned}
\nabla_h^L \cdot \left[M_{ij}^R \nabla_h^R \omega_{ij} \right] &= \nabla_h^L \cdot \begin{pmatrix} \frac{M_{i+1,j}+M_{ij}}{2} \partial_x^R \omega_{ij} \\ \frac{M_{i,j+1}+M_{ij}}{2} \partial_y^R \omega_{ij} \end{pmatrix} \\
&= \nabla_h^L \cdot \begin{pmatrix} \frac{M_{i+1,j}+M_{ij}}{2} \frac{\omega_{i+1,j}-\omega_{ij}}{h_x} \\ \frac{M_{i,j+1}+M_{ij}}{2} \frac{\omega_{i,j+1}-\omega_{ij}}{h_y} \end{pmatrix} \\
&= \partial_x^L \left[\frac{M_{i+1,j}+M_{ij}}{2} \frac{\omega_{i+1,j}-\omega_{ij}}{h_x} \right] + \partial_y^L \left[\frac{M_{i,j+1}+M_{ij}}{2} \frac{\omega_{i,j+1}-\omega_{ij}}{h_y} \right] \\
&= \frac{\frac{M_{i+1,j}+M_{ij}}{2} \frac{\omega_{i+1,j}-\omega_{ij}}{h_x} - \frac{M_{ij}+M_{i-1,j}}{2} \frac{\omega_{ij}-\omega_{i-1,j}}{h_x}}{h_x} \\
&\quad + \frac{\frac{M_{i,j+1}+M_{ij}}{2} \frac{\omega_{i,j+1}-\omega_{ij}}{h_y} - \frac{M_{ij}+M_{i,j-1}}{2} \frac{\omega_{ij}-\omega_{i,j-1}}{h_y}}{h_y} \\
&= \frac{\frac{M_{i+1,j}+M_{ij}}{2} \omega_{i+1,j} - (\frac{M_{i+1,j}+M_{ij}}{2} + \frac{M_{ij}+M_{i-1,j}}{2}) \omega_{i,j} + \frac{M_{ij}+M_{i,j-1}}{2} \omega_{i-1,j}}{h_x^2} \\
&\quad + \frac{\frac{M_{i,j+1}+M_{ij}}{2} \omega_{i,j+1} - (\frac{M_{i,j+1}+M_{ij}}{2} + \frac{M_{ij}+M_{i,j-1}}{2}) \omega_{i,j} + \frac{M_{ij}+M_{i,j-1}}{2} \omega_{i,j-1}}{h_y^2} .
\end{aligned}
\tag{4.124}
$$

This is equivalent to equation (4.120) and its accuracy is therefore of second order in space as well.

Note that the shifted (or left and right) differences cannot only be used to reduce the numerical diffusion of higher order derivatives. They can also be used used to discretize derivatives in coupling terms on the staggered grid, since, e.g.

$$
\partial_x^L \omega_{ij} = \frac{\omega_{ij} - \omega_{i-1,j}}{h_x} = \tilde{\partial}_x \omega_{i-1/2,j} ,
$$

where $\omega_{i-1/2,j}$ is defined on the grid shifted in x-direction as is the velocity u. The same holds true vice versa for the differences $\tilde{\partial}_x u_{ij}$ and $\tilde{\partial}_y v_{ij}$ given by (4.117).

Remark 4.44. *Let us recall that we applied integration by parts in order to derive semi-discrete energy laws and mass conservation in the above subsections. The discrete counterpart to integration by parts is summation by parts. Therefore, fully discrete mass conservation and energy laws are only analogous to the semi-discrete laws, if the spatial discretization fulfills summation by parts rules which are analogous to the integration by parts rules used for the semi-discrete laws.*

Lemma 4.45. *(summation by parts for finite differences)*
Assuming suitable boundary conditions, the following summation by parts (SBP) formulas hold

for the above introduced finite differences

$$\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \boldsymbol{u}_{ij}^R \cdot \nabla_h^R \omega_{ij} = -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \nabla_h^L \cdot \left(\boldsymbol{u}_{ij}^R\right) \omega_{ij}\,, \tag{4.125a}$$

$$\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \nabla_h^L \cdot \boldsymbol{u}_{ij}^R = 0\,, \tag{4.125b}$$

$$\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{ij}^R |\nabla_h^R \omega_{ij}|^2 = -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \nabla_h^L \cdot \left(M_{ij}^R \nabla_h^R \omega_{ij}\right) \omega_{ij}\,, \tag{4.125c}$$

$$\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \nabla_h^L \cdot \left(M_{ij}^R \nabla_h^R \omega_{ij}\right) = 0\,. \tag{4.125d}$$

Note that these SBP formulas hold analogously for any space dimension. Further, note that these SBP formulas also hold analogously for central differences.

Proof. **Part 1.** Using index shifting gives

$$\begin{aligned}
\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \boldsymbol{u}_{ij}^R \cdot \nabla_h^R \omega_{ij} &= \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} u_{i+1/2,j} \left(\frac{\omega_{i+1,j} - \omega_{ij}}{h_x} + v_{i,j+1/2} \frac{\omega_{i,j+1} - \omega_{ij}}{h_y} \right) \\
&= \frac{1}{h_x} \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left[u_{i+1/2,j} \omega_{i+1,j} - u_{i+1/2,j} \omega_{ij} \right] + \frac{1}{h_y} \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left[v_{i,j+1/2} \omega_{i,j+1} - v_{i,j+1/2} \omega_{ij} \right] \\
&= \frac{1}{h_x} \left(\sum_{\substack{i=2,\dots,N_x+1\\ j=1,\dots,N_y}} u_{i-1/2,j} \omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} u_{i+1/2,j} \omega_{ij} \right) \\
&\quad + \frac{1}{h_y} \left(\sum_{\substack{i=1,\dots,N_x\\ j=2,\dots,N_y+1}} v_{i,j-1/2} \omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} v_{i,j+1/2} \omega_{ij} \right) \\
&= \frac{1}{h_x} \left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} u_{i-1/2,j} \omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} u_{i+1/2,j} \omega_{ij} + \sum_{j=1}^{N_y} \left[u_{N_x+1/2,j} \omega_{N_x+1,j} - u_{1/2,j} \omega_{1,j} \right] \right) \\
&\quad + \frac{1}{h_y} \left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} v_{i,j-1/2} \omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} v_{i,j+1/2} \omega_{ij} + \sum_{i=1}^{N_x} \left[v_{i,N_y+1/2} \omega_{i,N_y+1} - v_{i,1/2} \omega_{i,1} \right] \right).
\end{aligned}$$

Assuming periodic boundary conditions for $\boldsymbol{u}$ and ω, we have

$$u_{N_x+1/2,j} \omega_{N_x+1,j} = u_{1/2,j} \omega_{1,j}\,, \qquad v_{i,N_y+1/2} \omega_{i,N_y+1} = v_{i,1/2} \omega_{i,1}\,.$$

Assuming homogeneous Dirichlet boundary conditions for $\boldsymbol{u}$, we have

$$u_{N_x+1/2,j} = u_{1/2,j} = v_{i,N_y+1/2} = v_{i,1/2} = 0\,.$$

Thus, the boundary sums vanish for both boundary conditions and we have

$$\begin{aligned}\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \boldsymbol{u}_{ij}^R \cdot \nabla_h^R \omega_{ij} &= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(\frac{u_{i+1/2,j} - u_{i-1/2,j}}{h_x} + \frac{v_{i,j+1/2} - v_{i,j-1/2}}{h_y} \right) \omega_{ij} \\ &= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(\partial_x^L u_{i+1/2,j} + \partial_y^L v_{i,j+1/2} \right) \omega_{ij} \\ &= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \nabla_h^L \cdot \left(\boldsymbol{u}_{ij}^R \right) \omega_{ij} \,.\end{aligned}$$

Part 2. By setting $\omega_{ij} = 1$ in SBP formula (4.125a), we obtain

$$\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \nabla_h^L \cdot \left(\boldsymbol{u}_{ij}^R \right) 1 = -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \boldsymbol{u}_{ij}^R \cdot \nabla_h^R 1 = -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(u_{i+1/2,j} + v_{i,j+1/2} \right) 0 = 0 \,.$$

Part 3. This part directly follows by setting $\boldsymbol{u}_{ij}^R = M_{ij}^R \nabla_h^R \omega_{ij}$ in SBP formula (4.125a). The following computations are only meant to gain insight on the necessary boundary conditions. For $M_{i+1/2,j} = (M_{i+1,j} + M_{ij})/2$, $M_{i,j+1/2} = (M_{i,j+1} + M_{ij})/2$ and using index shifting, it holds

$$\begin{aligned}\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{ij}^R |\nabla_h^R \omega_{ij}|^2 &= \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(M_{i+1/2,j} \left(\frac{\omega_{i+1,j} - \omega_{ij}}{h_x} \right)^2 + M_{i,j+1/2} \left(\frac{\omega_{i,j+1} - \omega_{ij}}{h_y} \right)^2 \right) \\ &= \frac{1}{h_x^2} \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i+1/2,j} \left(\omega_{i+1,j}^2 - 2\omega_{i+1,j}\omega_{ij} + \omega_{ij}^2 \right) \\ &\quad + \frac{1}{h_y^2} \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i,j+1/2} \left(\omega_{i,j+1}^2 - 2\omega_{i,j+1}\omega_{ij} + \omega_{ij}^2 \right) \\ &= \frac{1}{h_x^2} \left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i+1/2,j} \left(\omega_{ij}^2 - \omega_{i+1,j}\omega_{ij} \right) + \sum_{\substack{i=2,\dots,N_x+1\\ j=1,\dots,N_y}} M_{i-1/2,j} \left(\omega_{ij}^2 - \omega_{ij}\omega_{i-1,j} \right) \right) \\ &\quad + \frac{1}{h_y^2} \left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i,j+1/2} \left(\omega_{ij}^2 - \omega_{i,j+1}\omega_{ij} \right) + \sum_{\substack{i=1,\dots,N_x\\ j=2,\dots,N_y+1}} M_{i,j-1/2} \left(\omega_{ij}^2 - \omega_{ij}\omega_{i,j-1} \right) \right),\end{aligned}$$

which is equivalent to

$$\begin{aligned}\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{ij}^R|\nabla_h^R\omega_{ij}|^2 = \frac{1}{h_x^2}\Bigg(&\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i+1/2,j}\left(\omega_{ij}^2-\omega_{i+1,j}\omega_{ij}\right)+\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i-1/2,j}\left(\omega_{ij}^2-\omega_{ij}\omega_{i-1,j}\right)\\ &+\sum_{j=1}^{N_y}\left[M_{N_x+1/2,j}\left(\omega_{N_x+1,j}^2-\omega_{N_x+1,j}\omega_{N_x,j}\right)-M_{1/2,j}\left(\omega_{1,j}^2-\omega_{1,j}\omega_{0,j}\right)\right]\Bigg)\\ +\frac{1}{h_y^2}\Bigg(&\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i,j+1/2}\left(\omega_{ij}^2-\omega_{i,j+1}\omega_{ij}\right)+\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i,j-1/2}\left(\omega_{ij}^2-\omega_{ij}\omega_{i,j-1}\right)\\ &+\sum_{i=1}^{N_x}\left[M_{i,N_y+1/2}\left(\omega_{i,N_y+1}^2-\omega_{i,N_y+1}\omega_{i,N_y}\right)-M_{i,1/2}\left(\omega_{i,1}^2-\omega_{i,1}\omega_{i,0}\right)\right]\Bigg).\end{aligned}$$

Assuming periodic boundary conditions for φ and ω, we have

$$\begin{aligned}M_{N_x+1/2,j}\left(\omega_{N_x+1,j}^2-\omega_{N_x+1,j}\omega_{N_x,j}\right)&=M_{1/2,j}\left(\omega_{1,j}^2-\omega_{1,j}\omega_{0,j}\right),\\ M_{i,N_y+1/2}\left(\omega_{i,N_y+1}^2-\omega_{i,N_y+1}\omega_{i,N_y}\right)&=M_{i,1/2}\left(\omega_{i,1}^2-\omega_{i,1}\omega_{i,0}\right).\end{aligned}$$

Assuming homogeneous Neumann boundary conditions for ω, we have

$$\begin{aligned}\omega_{N_x+1,j}^2-\omega_{N_x+1,j}\omega_{N_x,j}=\omega_{N_x,j}^2-\omega_{N_x,j}\omega_{N_x,j}=0,\qquad &\omega_{1,j}^2-\omega_{1,j}\omega_{0,j}=\omega_{1,j}^2-\omega_{1,j}\omega_{1,j}=0,\\ \omega_{i,N_y+1}^2-\omega_{i,N_y+1}\omega_{i,N_y}=\omega_{i,N_y}^2-\omega_{i,N_y}\omega_{i,N_y}=0,\qquad &\omega_{i,1}^2-\omega_{i,1}\omega_{i,0}=\omega_{i,1}^2-\omega_{i,1}\omega_{i,1}=0.\end{aligned}$$

Thus, the boundary sums vanish for both boundary conditions and we have

$$\begin{aligned}\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{ij}^R|\nabla_h^R\omega_{ij}|^2 &= \frac{1}{h_x^2}\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}}\left[M_{i+1/2,j}\left(\omega_{ij}-\omega_{i+1,j}\right)+M_{i-1/2,j}\left(\omega_{ij}-\omega_{i-1,j}\right)\right]\omega_{ij}\\ &\quad+\frac{1}{h_y^2}\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}}\left[M_{i,j+1/2}\left(\omega_{ij}-\omega_{i,j+1}\right)+M_{i,j-1/2}\left(\omega_{ij}-\omega_{i,j-1}\right)\right]\omega_{ij}\\ &=-\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}}\frac{M_{i+1/2,j}\omega_{i+1,j}-(M_{i+1/2,j}+M_{i-1/2,j})\omega_{ij}+M_{i-1/2,j}\omega_{i-1,j}}{h_x^2}\omega_{ij}\\ &\qquad+\frac{M_{i,j+1/2}\omega_{i,j+1}-(M_{i,j+1/2}+M_{i,j-1/2})\omega_{ij}+M_{i,j-1/2}\omega_{i,j-1}}{h_y^2}\omega_{ij}\\ &=-\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}}\nabla_h^L\cdot\left(M_{ij}^R\nabla_h^R\omega_{ij}\right)\omega_{ij},\end{aligned}$$

where the last equality holds true due to equation (4.124).

Part 4. This part directly follows by setting $\boldsymbol{u}_{ij}^R = M_{ij}^R \nabla_h^R \omega_{ij}$ in SBP formula (4.125b). Again, the following computations are only meant to gain insight on the necessary boundary conditions. Using equation (4.124) and index shifting, it holds

$$\begin{aligned}
\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \nabla_h^L \cdot \left(M_{ij}^R \nabla_h^R \omega_{ij}\right) &= \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \frac{M_{i+1/2,j}\omega_{i+1,j} - (M_{i+1/2,j} + M_{i-1/2,j})\omega_{ij} + M_{i-1/2,j}\omega_{i-1,j}}{h_x^2} \\
&\quad + \frac{M_{i,j+1/2}\omega_{i,j+1} - (M_{i,j+1/2} + M_{i,j-1/2})\omega_{ij} + M_{i,j-1/2}\omega_{i,j-1}}{h_y^2} \\
&= \frac{1}{h_x^2}\left(\sum_{\substack{i=2,\dots,N_x+1\\ j=1,\dots,N_y}} M_{i-1/2,j}\omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} (M_{i+1/2,j} + M_{i-1/2,j})\omega_{ij} + \sum_{\substack{i=0,\dots,N_x-1\\ j=1,\dots,N_y}} M_{i+1/2,j}\omega_{ij} \right) \\
&\quad + \frac{1}{h_y^2}\left(\sum_{\substack{i=1,\dots,N_x\\ j=2,\dots,N_y+1}} M_{i,j-1/2}\omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} (M_{i,j+1/2} + M_{i,j-1/2})\omega_{ij} + \sum_{\substack{i=1,\dots,N_x\\ j=0,\dots,N_y-1}} M_{i,j+1/2}\omega_{ij} \right) \\
&= \frac{1}{h_x^2}\Bigg(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i-1/2,j}\omega_{ij} - (M_{i+1/2,j} + M_{i-1/2,j})\omega_{ij} + M_{i+1/2,j}\omega_{ij} \\
&\qquad + \sum_{j=1}^{N_y} \left[M_{N_x+1/2,j}\omega_{N_x+1,j} - M_{1/2,j}\omega_{1,j} + M_{1/2,j}\omega_{0,j} - M_{N_x+1/2,j}\omega_{N_x,j} \right] \Bigg) \\
&\quad + \frac{1}{h_y^2}\Bigg(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} M_{i,j-1/2}\omega_{ij} - (M_{i,j+1/2} + M_{i,j-1/2})\omega_{ij} + M_{i,j+1/2}\omega_{ij} \\
&\qquad + \sum_{i=1}^{N_x} \left[M_{i,N_y+1/2}\omega_{i,N_y+1} - M_{i,1/2}\omega_{i,1} + M_{i,1/2}\omega_{i,0} - M_{i,N_y+1/2}\omega_{i,N_y} \right] \Bigg) \\
&= \frac{1}{h_x^2} \sum_{j=1}^{N_y} \left[M_{N_x+1/2,j}\omega_{N_x+1,j} - M_{1/2,j}\omega_{1,j} + M_{1/2,j}\omega_{0,j} - M_{N_x+1/2,j}\omega_{N_x,j} \right] \\
&\quad + \frac{1}{h_y^2} \sum_{i=1}^{N_x} \left[M_{i,N_y+1/2}\omega_{i,N_y+1} - M_{i,1/2}\omega_{i,1} + M_{i,1/2}\omega_{i,0} - M_{i,N_y+1/2}\omega_{i,N_y} \right] .
\end{aligned}$$

Assuming periodic boundary conditions for φ and ω, we have

$$\begin{aligned}
M_{N_x+1/2,j}\omega_{N_x+1,j} &= M_{1/2,j}\omega_{1,j} , & M_{1/2,j}\omega_{0,j} &= M_{N_x+1/2,j}\omega_{N_x,j} , \\
M_{i,N_y+1/2}\omega_{i,N_y+1} &= M_{i,1/2}\omega_{i,1} , & M_{i,1/2}\omega_{i,0} &= M_{i,N_y+1/2}\omega_{i,N_y} .
\end{aligned}$$

Assuming homogeneous Neumann boundary conditions for ω, we have

$$\begin{aligned}
M_{N_x+1/2,j}\omega_{N_x+1,j} &= M_{N_x+1/2,j}\omega_{N_x,j} , & M_{1/2,j}\omega_{1,j} &= M_{1/2,j}\omega_{0,j} , \\
M_{i,N_y+1/2}\omega_{i,N_y+1} &= M_{i,N_y+1/2}\omega_{i,N_y} , & M_{i,1/2}\omega_{i,1} &= M_{i,1/2}\omega_{i,0} .
\end{aligned}$$

Thus, the boundary sums vanish for both boundary conditions. □

4.8.3 Two-dimensional discretization of the Cahn-Hilliard equation

Using the above introduced left and right differences and the semi-discrete scheme (4.33) for the time discretization, the fully discretized Cahn-Hilliard equation in two space dimensions reads

$$\begin{aligned}\varphi_{ij}^{n+1} &= \varphi_{ij}^{n} + \Delta t \nabla_h^L \cdot \left(M(\varphi_{ij}^n)^R \nabla_h^R \mu_{ij}^{n+1/2} \right) \\ &= \varphi_{ij}^{n} + \Delta t \partial_x^L \left(\frac{M(\varphi_{i+1,j}^n) + M(\varphi_{ij}^n)}{2} \partial_x^R \mu_{ij}^{n+1/2} \right) \\ &\quad + \Delta t \partial_y^L \left(\frac{M(\varphi_{i,j+1}^n) + M(\varphi_{ij}^n)}{2} \partial_y^R \mu_{ij}^{n+1/2} \right),\end{aligned} \tag{4.126}$$

where $i = 1, ..., N_x$, $j = 1, ..., N_y$, and the discretization of the chemical potential $\mu_{ij}^{n+1/2}$, using the OD2 approximation (4.26) for the potential derivative, is given by

$$\mu_{ij}^{n+1/2} = -\lambda \Delta_h \frac{\varphi_{ij}^{n+1} + \varphi_{ij}^{n}}{2} + f(\varphi_{ij}^n) + \frac{\varphi_{ij}^{n+1} - \varphi_{ij}^{n}}{2} f'(\varphi_{ij}^n). \tag{4.127}$$

Theorem 4.46. *The fully discretized Cahn-Hilliard equation (4.126) is linear and, assuming suitable boundary conditions, see Remark 3.2, it is mass-conservative and fulfills the following discrete energy law*

$$\frac{E_{mix}(\varphi^{n+1}) - E_{mix}(\varphi^n)}{\Delta t} = - \sum_{\substack{i=1,...,N_x \\ j=1,...,N_y}} M(\varphi_{ij}^n)^R |\nabla^R \mu_{ij}^{n+\frac{1}{2}}|^2 - \mathrm{ND}_{pot}^{n+1}, \tag{4.128}$$

where the discrete energy reads

$$E_{mix}(\varphi^n) = \sum_{\substack{i=1,...,N_x \\ j=1,...,N_y}} \frac{\lambda}{2} |\nabla_h^R \varphi_{ij}^n|^2 + F(\varphi_{ij}^n),$$

and the numerical dissipation caused by the OD2 approximation reads

$$\mathrm{ND}_{pot}^{n+1} = -\frac{1}{6}(\Delta t)^2 \sum_{\substack{i=1,...,N_x \\ j=1,...,N_y}} f''(\zeta) \left(\partial_t \varphi_{ij}^{n+\frac{1}{2}} \right)^3 \in \mathcal{O}\left((\Delta t)^2 \right). \tag{4.129}$$

Note that this discrete energy law is analogous to energy law (4.34) of the semi-discrete scheme (4.33).

Proof. It is clear that the proposed scheme is linear.

Summing scheme (4.126) over all $i = 1, ..., N_x$ and $j = 1, ..., N_y$, and using the summation by parts formula (4.125d), we have

$$\sum_{\substack{i=1,...,N_x \\ j=1,...,N_y}} \varphi_{ij}^{n+1} = \sum_{\substack{i=1,...,N_x \\ j=1,...,N_y}} \varphi_{ij}^{n} + \Delta t \sum_{\substack{i=1,...,N_x \\ j=1,...,N_y}} \nabla_h^L \cdot \left(M(\varphi_{ij}^n)^R \nabla_h^R \mu_{ij}^{n+1/2} \right) = \sum_{\substack{i=1,...,N_x \\ j=1,...,N_y}} \varphi_{ij}^{n}.$$

Thus, the scheme is mass-conservative.

Multiplying scheme (4.126) by $\mu_{ij}^{n+1/2}$, summing over all $i = 1, ..., N_x$ and $j = 1, ..., N_y$, and using the summation by parts formulas (4.125a) and (4.125c), the discrete energy law (4.128) is calculated analogously to the semi-discrete energy law (4.34). The numerical dissipation (4.129) is calculated analogously to (4.27). □

Let us recall that the proposed linearized time discretizations of the Cahn-Hilliard equation have to be (partly) implicit to be provably energy-stable and that the implicit part includes spatial derivatives of φ. Further, finite differences always require information of neighboring cells. Summarized, we have to solve a coupled system of all N_x times N_y equations in every time step, e.g., (4.126) for $i = 1, ..., N_x$ and $j = 1, ..., N_y$. Therefore, a matrix formulation is favorable, which will be introduced in the following.

4.8.4 Finite difference matrices

The matrix notation of the one-dimensional (1D) central finite difference (4.109), reading

$$\partial\omega_i = \frac{\omega_{i+1} - \omega_{i-1}}{2h},$$

for $i = 1, 2, ..., N$ and homogeneous Dirichlet boundary conditions, i.e., $\omega_0 = 0 = \omega_{N+1}$, reads

$$\partial_N \omega^{1D} := \frac{1}{2h} \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 0 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_{N-1} \\ \omega_N \end{pmatrix} \in \mathbb{R}^N,$$

where ∂_N is a tridiagonal matrix of size $N \times N$. Further boundary conditions will be outlined at the end of this subsection.

To introduce matrix notations for higher dimensions, we need some preliminary work, since the discrete solution of ω is usually given by a tensor of the order of the dimension of the corresponding computational domain Ω_h, i.e., by a matrix in two dimensions and by a tensor of order three in three dimensions. In order to be able to operate with matrices on this higher dimensional solutions, it is convenient to vectorize them. We vectorize a $N_x \times N_y$-matrix column-wise to a vector of length $N_x N_y$ through lexicographic ordering, that is

$$\omega^{2D} := \mathrm{vec}\left(\begin{pmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1N_y} \\ \omega_{21} & \omega_{22} & & \vdots \\ \vdots & & \ddots & \\ \omega_{N_x 1} & \dots & & \omega_{N_x N_y} \end{pmatrix} \right) := \begin{pmatrix} \omega_{11} \\ \omega_{21} \\ \vdots \\ \omega_{N_x 1} \\ \omega_{12} \\ \vdots \\ \vdots \\ \omega_{N_x N_y} \end{pmatrix} \in \mathbb{R}^{N_x N_y}.$$

Analogously, vectorizing a (third order) $N_x \times N_y \times N_z$-tensor yields

$$\omega^{3D} := \begin{pmatrix} \omega_{\cdot\cdot 1} \\ \omega_{\cdot\cdot 2} \\ \vdots \\ \omega_{\cdot\cdot N_z} \end{pmatrix} \in \mathbb{R}^{N_x N_y N_z},$$

where $\omega_{\cdot\cdot k} := \omega_k^{2D}, k = 1, 2, ..., N_z$, i.e., ω^{3D} is a vector of length $N_x N_y N_z$.

Remark 4.47. *Note that we can always reverse the vectorization by matricization or tensorization.*

Having a vectorized discrete solution, we can now introduce matrix notations of finite differences for higher dimensions. The central difference matrices in two space dimensions (2D) are constructed by Kronecker products of the one-dimensional difference matrices with identity matrices, reading

$$\partial_{N_x^2} = \boldsymbol{I}_{N_y} \otimes \partial_{N_x} ,$$

where $\boldsymbol{I}_{N_y}$ is a $N_y \times N_y$ identity matrix, and

$$\partial_{N_y^2} = \partial_{N_y} \otimes \boldsymbol{I}_{N_x} .$$

These are both block matrices of size $(N_x N_y) \times (N_x N_y)$. For left and right differences in 2D, we need the following 1D matrices

$$\partial_N^L = \frac{1}{h} \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}, \qquad \partial_N^R = \frac{1}{h} \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & & -1 \end{pmatrix},$$

such that

$$\partial_{N_x^2}^L = \boldsymbol{I}_{N_y} \otimes \partial_{N_x}^L , \qquad \partial_{N_x^2}^R = \boldsymbol{I}_{N_y} \otimes \partial_{N_x}^R ,$$
$$\partial_{N_y^2}^L = \partial_{N_y}^L \otimes \boldsymbol{I}_{N_x} , \qquad \partial_{N_y^2}^R = \partial_{N_y}^R \otimes \boldsymbol{I}_{N_x} .$$

To complete the finite difference matrices for first order derivatives in 2D, we also adopt the discrete gradient, reading

$$\nabla_{N^2} \omega^{2D} = \begin{pmatrix} \partial_{N_x^2} \omega^{2D} \\ \partial_{N_y^2} \omega^{2D} \end{pmatrix}, \qquad \nabla_{N^2}^{L/R} \omega^{2D} = \begin{pmatrix} \partial_{N_x^2}^{L/R} \omega^{2D} \\ \partial_{N_y^2}^{L/R} \omega^{2D} \end{pmatrix},$$

and the discrete divergence, reading

$$\nabla_{N^2} \cdot \begin{pmatrix} \omega_1^{2D} \\ \omega_2^{2D} \end{pmatrix} = \partial_{N_x^2} \omega_1^{2D} + \partial_{N_y^2} \omega_2^{2D} ,$$
$$\nabla_{N^2}^{L/R} \cdot \begin{pmatrix} \omega_1^{2D} \\ \omega_2^{2D} \end{pmatrix} = \partial_{N_x^2}^{L/R} \omega_1^{2D} + \partial_{N_y^2}^{L/R} \omega_2^{2D} .$$

As mentioned before, the left and right differences cannot only be used to reduce the numerical diffusion of higher order derivatives, but also to discretize derivatives in coupling terms on staggered grids.

However, not all coupling terms consisting of functions discretized on different grids only contain derivatives. Hence, we also need matrices to shift approximations from one grid to the other, which is in 1D done by the left and right average matrices

$$A_N^L = \frac{1}{2} \begin{pmatrix} 1 & & & \\ 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \end{pmatrix}, \qquad A_N^R = \frac{1}{2} \begin{pmatrix} 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix},$$

and in 2D by

$$A^L_{N^2_x} = \boldsymbol{I}_{N_y} \otimes A^L_{N_x}\,, \qquad A^R_{N^2_x} = \boldsymbol{I}_{N_y} \otimes A^R_{N_x}\,,$$
$$A^L_{N^2_y} = A^L_{N_y} \otimes \boldsymbol{I}_{N_x}\,, \qquad A^R_{N^2_y} = A^R_{N_y} \otimes \boldsymbol{I}_{N_x}\,.$$

Finally, let us introduce the matrix formulation of the central finite difference for the Laplacian, which is in 1D given by

$$\Delta_N = \frac{1}{h^2}\begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{pmatrix},$$

and therefore in 2D by

$$\Delta_{N^2} = \boldsymbol{I}_{N_y} \otimes \Delta_{N_x} + \Delta_{N_y} \otimes \boldsymbol{I}_{N_x}\,.$$

Note that the extension of the above introduced differences and averages to three space dimensions (3D) is straight forward by taking another Kronecker product. Thus, central finite differences in 3D read

$$\partial_{N^3_x} = \boldsymbol{I}_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \partial_{N_x}\,,$$
$$\partial_{N^3_y} = \boldsymbol{I}_{N_z} \otimes \partial_{N_y} \otimes \boldsymbol{I}_{N_x}\,,$$
$$\partial_{N^3_z} = \partial_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \boldsymbol{I}_{N_x}\,.$$

Analogously to above, substituting the central difference matrices ∂_N by the left and right difference and average matrices yields the other necessary matrix operations. Hence, the left and right finite difference matrices in 3D read

$$\partial^{L/R}_{N^3_x} = \boldsymbol{I}_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \partial^{L/R}_{N_x}\,,$$
$$\partial^{L/R}_{N^3_y} = \boldsymbol{I}_{N_z} \otimes \partial^{L/R}_{N_y} \otimes \boldsymbol{I}_{N_x}\,,$$
$$\partial^{L/R}_{N^3_z} = \partial^{L/R}_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \boldsymbol{I}_{N_x}\,,$$

and the left and right average matrices

$$A^{L/R}_{N^3_x} = \boldsymbol{I}_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes A^{L/R}_{N_x}\,,$$
$$A^{L/R}_{N^3_y} = \boldsymbol{I}_{N_z} \otimes A^{L/R}_{N_y} \otimes \boldsymbol{I}_{N_x}\,,$$
$$A^{L/R}_{N^3_z} = A^{L/R}_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \boldsymbol{I}_{N_x}\,.$$

The discrete gradients in 3D are given by

$$\nabla_{N^3}\omega^{3D} = \begin{pmatrix} \partial_{N^3_x}\omega^{3D} \\ \partial_{N^3_y}\omega^{3D} \\ \partial_{N^3_z}\omega^{3D} \end{pmatrix}, \qquad \nabla^{L/R}_{N^3}\omega^{3D} = \begin{pmatrix} \partial^{L/R}_{N^3_x}\omega^{3D} \\ \partial^{L/R}_{N^3_y}\omega^{3D} \\ \partial^{L/R}_{N^3_z}\omega^{3D} \end{pmatrix},$$

and the divergences by

$$\nabla_{N^3} \cdot \begin{pmatrix} \omega_1^{3D} \\ \omega_2^{3D} \\ \omega_3^{3D} \end{pmatrix} = \partial_{N_x^3} \omega_1^{3D} + \partial_{N_y^3} \omega_2^{3D} + \partial_{N_z^3} \omega_3^{3D} ,$$

$$\nabla_{N^3}^{L/R} \cdot \begin{pmatrix} \omega_1^{3D} \\ \omega_2^{3D} \\ \omega_3^{3D} \end{pmatrix} = \partial_{N_x^3}^{L/R} \omega_1^{3D} + \partial_{N_y^3}^{L/R} \omega_2^{3D} + \partial_{N_z^3}^{L/R} \omega_3^{3D} .$$

And finally, the matrix formulation of the central finite difference for the Laplacian in 3D reads

$$\Delta_{N^3} = \boldsymbol{I}_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \Delta_{N_x} + \boldsymbol{I}_{N_z} \otimes \Delta_{N_y} \otimes \boldsymbol{I}_{N_x} + \Delta_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \boldsymbol{I}_{N_x} .$$

Boundary conditions

Note that all aforementioned 1D finite difference and average matrices are defined for homogeneous Dirichlet boundary conditions. For periodic boundary conditions, they read

$$\partial_N = \frac{1}{2h} \begin{pmatrix} 0 & 1 & & & -1 \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ 1 & & & -1 & 0 \end{pmatrix},$$

$$\partial_N^L = \frac{1}{h} \begin{pmatrix} 1 & & & & -1 \\ -1 & 1 & & & \\ & \ddots & \ddots & & \\ & & & -1 & 1 \end{pmatrix}, \qquad \partial_N^R = \frac{1}{h} \begin{pmatrix} -1 & 1 & & & \\ & \ddots & \ddots & & \\ & & & -1 & 1 \\ 1 & & & & -1 \end{pmatrix},$$

$$A_N^L = \frac{1}{2} \begin{pmatrix} 1 & & & & 1 \\ 1 & 1 & & & \\ & \ddots & \ddots & & \\ & & & 1 & 1 \end{pmatrix}, \qquad A_N^R = \frac{1}{2} \begin{pmatrix} 1 & 1 & & & \\ & \ddots & \ddots & & \\ & & & 1 & 1 \\ 1 & & & & 1 \end{pmatrix}$$

and

$$\Delta_N = \frac{1}{h^2} \begin{pmatrix} -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 1 & & & 1 & -2 \end{pmatrix}.$$

And for homogeneous Neumann boundary conditions, they read

$$\partial_N = \frac{1}{2h} \begin{pmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 1 \end{pmatrix},$$

$$\partial_N^L = \frac{1}{h}\begin{pmatrix} 0 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}, \qquad \partial_N^R = \frac{1}{h}\begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & & 0 \end{pmatrix},$$
$$A_N^L = \frac{1}{2}\begin{pmatrix} 2 & & & \\ 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \end{pmatrix}, \qquad A_N^R = \frac{1}{2}\begin{pmatrix} 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \\ & & & 2 \end{pmatrix}$$

and

$$\Delta_N = \frac{1}{h^2}\begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{pmatrix}.$$

4.8.5 Full discretization of the Cahn-Hilliard equation

We discretize all spatial derivatives except advection and convection terms by the above introduced finite difference matrices. Thus, we can now introduce full discretizations in matrix notation of the Cahn-Hilliard equation and of the simplified model. Note that since the discrete solutions are vectorized matrices and tensors, all functions of them are computed element-wise. Further, note that we diagonalize most functions to enable their inversion as well as their multiplication with subsequent vectors. The corresponding function diag reads

$$\operatorname{diag}\left(\begin{pmatrix} \omega_1 \\ \vdots \\ \omega_n \end{pmatrix}\right) := \begin{pmatrix} \omega_1 & & \\ & \ddots & \\ & & \omega_n \end{pmatrix}.$$

The following discretizations are generalized, such that they hold for any space dimension $d \in \{1, 2, 3\}$ and any suitable boundary conditions.

Based on the fully discretized 2D scheme (4.126), the linear equation system to calculate

the vectorized discrete solution φ^{n+1} reads

$$\begin{aligned}
\varphi^{n+1} &= \varphi^n + \Delta t \nabla^L_{N^d} \cdot \left(\operatorname{diag}(M(\varphi^n)^R)\nabla^R_{N^d}\mu^{n+1/2}\right) \\
&= \varphi^n + \Delta t \nabla^L_{N^d} \cdot \left(\operatorname{diag}(M(\varphi^n)^R)\nabla^R_{N^d}\left[-\lambda\Delta_{N^d}\varphi^{n+1/2} + f(\varphi^n) \right.\right. \\
&\qquad\qquad \left.\left. + \operatorname{diag}\left(f'(\varphi^n)\right)\frac{\varphi^{n+1}-\varphi^n}{2}\right]\right) \\
&= \varphi^n + \Delta t \sum_{a\in\alpha_d} \partial^L_a \left(\operatorname{diag}\left(A^R_a M(\varphi^n)\right)\partial^R_a \left[-\lambda\Delta_{N^d}\frac{\varphi^{n+1}+\varphi^n}{2} + f(\varphi^n) \right.\right. \\
&\qquad\qquad \left.\left. + \operatorname{diag}\left(f'(\varphi^n)\right)\frac{\varphi^{n+1}-\varphi^n}{2}\right]\right) \\
\iff \quad & \varphi^{n+1} - \Delta t \left(\sum_{a\in\alpha_d} \partial^L_a \operatorname{diag}\left(A^R_a M(\varphi^n)\right)\partial^R_a\right)\left[\frac{-\lambda}{2}\Delta_{N^d}\varphi^{n+1} + \frac{1}{2}\operatorname{diag}\left(f'(\varphi^n)\right)\varphi^{n+1}\right] \\
&= \varphi^n + \Delta t \left(\sum_{a\in\alpha_d} \partial^L_a \operatorname{diag}\left(A^R_a M(\varphi^n)\right)\partial^R_a\right)\left[\frac{-\lambda}{2}\Delta_{N^d}\varphi^n + f(\varphi^n) - \frac{1}{2}\operatorname{diag}\left(f'(\varphi^n)\right)\varphi^n\right] \\
\iff \quad & \left\{ \boldsymbol{I}_{N^d} - \Delta t \left(\sum_{a\in\alpha_d} \partial^L_a \operatorname{diag}\left(A^R_a M(\varphi^n)\right)\partial^R_a\right)\left[\frac{-\lambda}{2}\Delta_{N^d} + \frac{1}{2}\operatorname{diag}\left(f'(\varphi^n)\right)\right]\right\}\varphi^{n+1} \qquad (4.130) \\
&= \varphi^n + \Delta t \left(\sum_{a\in\alpha_d} \partial^L_a \operatorname{diag}\left(A^R_a M(\varphi^n)\right)\partial^R_a\right)\left[\frac{-\lambda}{2}\Delta_{N^d}\varphi^n + f(\varphi^n) - \frac{1}{2}\operatorname{diag}\left(f'(\varphi^n)\right)\varphi^n\right],
\end{aligned}$$

where the index set α_d is based on the respective dimension and reads

$$\alpha_1 = N, \quad \alpha_2 = \{N^2_x, N^2_y\}, \quad \text{or} \quad \alpha_3 = \{N^3_x, N^3_y, N^3_z\}.$$

Note that in order to have a second order in time discretization, we can compute the mobility function M of $\varphi^{n-1/2} = (3\varphi^n - \varphi^{n-1})/2$ like in semi-discretization (4.40).

Theorem 4.48. *The completely discretized Cahn-Hilliard scheme (4.130) is linear and, assuming suitable boundary conditions, see Remark 3.2, it is mass-conservative and fulfills the discrete energy law (4.128) in 2D (considering a non-vectorized discrete solution) and analogous versions in other dimensions.*

Proof. See the proof of Theorem 4.46, since the proposed scheme (4.130) is in 2D (up to the vectorized discrete solution) equivalent to (4.126). For other space dimensions, the proof is analogous, since the SBP formulas (4.125) hold for any space dimension. □

4.8.6 Full discretization of the simplified model

The system of linear equations for solving the simplified model is more complex, since we want to solve both equations in a coupled way. Let us at first introduce a full discretization of the semi-discrete scheme (4.65) for each grid cell in 2D, using the OD2 approximation (4.26).

It reads

$$\varphi_{ij}^{n+1} = \varphi_{ij}^{n} + \Delta t\, \nabla_h^L \cdot \left\{ \frac{(\varphi_{ij}^n(1-\varphi_{ij}^n))^R}{\zeta(\varphi_{ij}^n)^R} \left[(\varphi_{ij}^n(1-\varphi_{ij}^n))^R \nabla_h^R \mu_{ij}^{n+\frac{1}{2}} - \nabla_h^R (G_B(\varphi_{ij}^n)\, q_{ij}^{n+\frac{1}{2}}) \right] \right\}, \tag{4.131a}$$

$$\begin{aligned} \mu_{ij}^{n+1/2} &= -\lambda \Delta_h \frac{\varphi_{ij}^{n+1} + \varphi_{ij}^n}{2} + f(\varphi_{ij}^n) + \frac{\varphi_{ij}^{n+1} - \varphi_{ij}^n}{2} f'(\varphi_{ij}^n), \\ q_{ij}^{n+1} &= q_{ij}^n - \Delta t \frac{1}{\tau_B(\varphi_{ij}^n)} q_{ij}^{n+\frac{1}{2}} \\ &\quad - \Delta t\, G_B(\varphi_{ij}^n) \nabla_h^L \cdot \left\{ \frac{1}{\zeta(\varphi_{ij}^n)^R} \left[(\varphi_{ij}^n(1-\varphi_{ij}^n))^R \nabla_h^R \mu_{ij}^{n+\frac{1}{2}} - \nabla_h^R (G_B(\varphi_{ij}^n)\, q_{ij}^{n+\frac{1}{2}}) \right] \right\}. \end{aligned} \tag{4.131b}$$

Theorem 4.49. *The fully discretized simplified model (4.131) is linear and, assuming suitable boundary conditions, see Remark 3.11, it is mass-conservative and fulfills the following discrete energy law in 2D*

$$\begin{aligned} \frac{E_{total}(\varphi^{n+1}, q^{n+1}) - E_{total}(\varphi^n, q^n)}{\Delta t} &= -\mathrm{ND}_{pot}^{n+1} - \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \frac{1}{\tau_B(\varphi_{ij}^n)} (q_{ij}^{n+\frac{1}{2}})^2 \\ &\quad - \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \frac{1}{\zeta(\varphi_{ij}^n)^R} \left| (\varphi_{ij}^n(1-\varphi_{ij}^n))^R \nabla_h^R \mu_{ij}^{n+\frac{1}{2}} - \nabla_h^R (G_B(\varphi_{ij}^n)\, q_{ij}^{n+\frac{1}{2}}) \right|^2, \end{aligned} \tag{4.132}$$

where the discrete energy reads

$$E_{total}(\varphi^n, q^n) = \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \left(\frac{\lambda}{2} |\nabla_h^R \varphi_{ij}^n|^2 + F(\varphi_{ij}^n) + \frac{1}{2} q_{ij}^n \right)$$

and the numerical dissipation caused by the OD2 approximation reads

$$\mathrm{ND}_{pot}^{n+1} = -\frac{1}{6}(\Delta t)^2 \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} f''(\zeta) \left(\partial_t \varphi_{ij}^{n+\frac{1}{2}} \right)^3 \in \mathcal{O}\left((\Delta t)^2\right).$$

Note that this discrete energy law is analogous to energy law (4.66) of the semi-discrete scheme (4.65).

Proof. It is clear that the proposed scheme is linear. Summing equation (4.131a) over all $i = 1, ..., N_x$ and $j = 1, ..., N_y$ and using the summation by parts formula (4.125b), we have

$$\begin{aligned} \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \varphi_{ij}^{n+1} = \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \varphi_{ij}^n + \Delta t \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \nabla_h^L \cdot \left\{ \frac{(\varphi_{ij}^n(1-\varphi_{ij}^n))^R}{\zeta(\varphi_{ij}^n)^R} \left[(\varphi_{ij}^n(1-\varphi_{ij}^n))^R \nabla_h^R \mu_{ij}^{n+\frac{1}{2}} \right.\right. \\ \left.\left. - \nabla_h^R (G_B(\varphi_{ij}^n)\, q_{ij}^{n+\frac{1}{2}}) \right] \right\} = \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \varphi_{ij}^n. \end{aligned}$$

Thus, the scheme is mass-conservative.

Multiplying equation (4.131a) by $\mu_{ij}^{n+1/2}$ and (4.131b) by $q_{ij}^{n+1/2}$, summing over all $i = 1, ..., N_x$ and $j = 1, ..., N_y$, and using the summation by parts formula (4.125a), the discrete energy law (4.132) is calculated analogously to the semi-discrete energy law (4.66). The numerical dissipation caused by the OD2 approximation is calculated analogously to (4.27). □

Based on discretization (4.131), the system of linear equations to calculate the vectorized discrete solutions φ^{n+1} and q^{n+1} reads

$$\begin{pmatrix} \mathrm{LHS}_{11} & \mathrm{LHS}_{12} \\ \mathrm{LHS}_{21} & \mathrm{LHS}_{22} \end{pmatrix} \begin{pmatrix} \varphi^{n+1} \\ q^{n+1} \end{pmatrix} = \begin{pmatrix} \mathrm{RHS}_1 \\ \mathrm{RHS}_2 \end{pmatrix}, \tag{4.133a}$$

where LHS_{12} and LHS_{21} contain the implicit parts of the coupling terms and therefore necessitate a coupled calculation. LHS_{11} equals the curly bracket term of the left hand side of equation (4.130), reading

$$\mathrm{LHS}_{11} = \boldsymbol{I}_{N^d} - \Delta t \left(\sum_{a \in \alpha_d} \partial_a^L \operatorname{diag}\left(A_a^R M_1(\varphi^n)\right) \partial_a^R \right) \left[\frac{-\lambda}{2} \Delta_{N^d} + \frac{1}{2} \operatorname{diag}\left(f'(\varphi^n)\right) \right]. \tag{4.133b}$$

LHS_{12} is the implicit part of the q-coupling term in the Cahn-Hilliard equation, which reads

$$\mathrm{LHS}_{12} = \frac{\Delta t}{2} \left(\sum_{a \in \alpha_d} \partial_a^L \operatorname{diag}\left(A_a^R M_2(\varphi^n)\right) \partial_a^R \right) \operatorname{diag}(G_B(\varphi^n)). \tag{4.133c}$$

RHS_1 is equal to the right hand side of equation (4.130) plus the explicit part of the q-coupling term. Thus, it reads

$$\begin{aligned} \mathrm{RHS}_1 = \varphi^n + \Delta t \sum_{a \in \alpha_d} \partial_a^L \Bigg(& \operatorname{diag}\left(A_a^R M_1(\varphi^n)\right) \partial_a^R \left[\frac{-\lambda}{2} \Delta_{N^d} \varphi^n + f(\varphi^n) - \frac{1}{2} \operatorname{diag}\left(f'(\varphi^n)\right) \varphi^n \right] \\ & - \operatorname{diag}\left(A_a^R M_2(\varphi^n)\right) \partial_a^R \left[\frac{1}{2} \operatorname{diag}(G_B(\varphi^n)) q^n \right] \Bigg), \end{aligned} \tag{4.133d}$$

where

$$M_1(\varphi) = (\varphi(1-\varphi))^2 \zeta(\varphi)^{-1} \qquad \text{and} \qquad M_2(\varphi) = \varphi(1-\varphi)\zeta(\varphi)^{-1}.$$

RHS_2 is the right hand side of the q-equation and reads

$$\begin{aligned} \mathrm{RHS}_2 = {} & q^n - \frac{\Delta t}{2} \operatorname{diag}\left(\tau_B(\varphi^n)\right)^{-1} q^n \\ & + \Delta t \operatorname{diag}\left(G_B(\varphi^n)\right) \sum_{a \in \alpha_d} \partial_a^L \Bigg(\operatorname{diag}\left(A_a^R M_2(\varphi^n)\right) \partial_a^R \left[\frac{-\lambda}{2} \Delta_{N^d} \varphi^n + f(\varphi^n) - \frac{1}{2} \operatorname{diag}\left(f'(\varphi^n)\right) \varphi^n \right] \\ & - \operatorname{diag}\left(A_a^R \zeta(\varphi^n)\right)^{-1} \partial_a^R \left[\frac{1}{2} \operatorname{diag}(G_B(\varphi^n)) q^n \right] \Bigg). \end{aligned} \tag{4.133e}$$

LHS_{22} is the left hand side of the q-equation and reads

$$\begin{aligned} \mathrm{LHS}_{22} = \boldsymbol{I}_{N^d} + \frac{\Delta t}{2} \Bigg(& \operatorname{diag}\left(\tau_B(\varphi^n)\right)^{-1} \\ & + \operatorname{diag}\left(G_B(\varphi^n)\right) \left[\sum_{a \in \alpha_d} \partial_a^L \operatorname{diag}\left(A_a^R \zeta(\varphi^n)\right)^{-1} \partial_a^R \right] \operatorname{diag}\left(G_B(\varphi^n)\right) \Bigg). \end{aligned} \tag{4.133f}$$

And finally, LHS_{21} is the implicit part of the φ-coupling term in the q-equation, which reads

$$\text{LHS}_{21} = \Delta t \,\text{diag}\left(G_B(\varphi^n)\right)\left(\sum_{a\in\alpha_d} \partial_a^L \,\text{diag}\left(A_a^R M_2(\varphi^n)\right)\,\partial_a^R\right)\left[\frac{-\lambda}{2}\Delta_{N^d} + \frac{1}{2}\text{diag}\left(f'(\varphi^n)\right)\right]. \tag{4.133g}$$

Note that in order to have a second order in time discretization, we can compute the functions M_1, M_2, ζ, G_B and τ_B of $\varphi^{n-1/2} = (3\varphi^n - \varphi^{n-1})/2$ like in semi-discretization (4.67).

Theorem 4.50. *The fully discretized simplified model* (4.133) *is linear and, assuming suitable boundary conditions, see Remark 3.11, it is mass-conservative and fulfills the discrete energy law* (4.132) *in 2D (considering non-vectorized discrete solutions) and analogous versions in other dimensions.*

Proof. See the proof of Theorem 4.49, since the proposed scheme (4.133) is in 2D (up to the vectorized discrete solutions) equivalent to (4.126). For other space dimensions, the proof is analogous, since the SBP formulas (4.125) hold for any space dimension. □

Let us recall that the Cahn-Hilliard-Navier-Stokes model and the full model contain advection and convection terms. An inconsiderate application of finite differences to these terms usually yields numerically unstable results. Therefore, we will introduce finite volume methods in the following, which handle advection and convection equations with ease and are therefore excellently suited to be applied to the advection and convection terms of our respective models.

4.8.7 Finite volume derivation

In the following, we will introduce and discuss the applicability of several first and second order finite volume schemes. For the sake of clarity, we start from the advection equation in one space dimension (1D). This introduction is loosely following *LeVeque* [52, chapter 4].

We consider a nonlinear 1D advection equation in conservation form with density function $\rho := \rho(x,t)$ and velocity $u := u(x,t)$, i.e., the mass equation

$$\rho_t + (u\rho)_x = 0\,, \tag{4.134}$$

where $\rho, u : \Omega \times [0,\infty) \to \mathbb{R}$ are sufficiently smooth. We partition the computational domain $\Omega \subset \mathbb{R}$ into a uniform grid with N grid cells, also denoted as *finite volumes*, and indicate them by

$$C_i = (x_{i-1/2}, x_{i+1/2})\,.$$

Integrating (4.134) over a finite volume reads

$$\begin{aligned}
&\int_{C_i}\left[\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}(u(x,t)\rho(u,x))\right]dx = 0\\
\Leftrightarrow\quad &\int_{C_i}\frac{\partial}{\partial t}\rho(x,t)dx = -\int_{C_i}\frac{\partial}{\partial x}(u(x,t)\rho(u,x))dx\\
\Leftrightarrow\quad &\frac{d}{dt}\int_{C_i}\rho(x,t)dx = u(x_{i-1/2},t)\rho(x_{i-1/2},t) - u(x_{i+1/2},t)\rho(x_{i+1/2},t)\,.
\end{aligned} \tag{4.135}$$

Now, we integrate (4.135) over the time interval (t_n, t_{n+1}), yielding

$$\int_{t_n}^{t_{n+1}} \frac{d}{dt} \int_{C_i} \rho(x,t) = \int_{t_n}^{t_{n+1}} u(x_{i-1/2},t)\rho(x_{i-1/2},t)dt - \int_{t_n}^{t_{n+1}} u(x_{i+1/2},t)\rho(x_{i+1/2},t)dt\,.$$

Dividing by the size $h_x = x_{i+1/2} - x_{i-1/2}$ of a finite volume, we have

$$\begin{aligned}\frac{1}{h_x}\int_{C_i} \rho(x,t_{n+1})dx = \frac{1}{h_x}\int_{C_i} \rho(x,t_n)dx \\ - \frac{1}{h_x}\left[\int_{t_n}^{t_{n+1}} u(x_{i+1/2},t)\rho(x_{i+1/2},t)dt - \int_{t_n}^{t_{n+1}} u(x_{i-1/2},t)\rho(x_{i-1/2},t)dt\right].\end{aligned}$$

Substituting the density integrals by their numerical approximations

$$\rho_i^n \approx \frac{1}{h_x}\int_{C_i} \rho(x,t_n)dx\,,$$

and the advection integrals by the so-called numerical flux functions

$$F_{i-1/2}^n \approx \frac{1}{\Delta t_n}\int_{t_n}^{t_{n+1}} u(x_{i-1/2},t)\rho(x_{i-1/2},t))dt\,,$$

where $\Delta t_n = t_{n+1} - t_n$ is the size of the n-th time interval, we have the finite volume discretization of the mass equation in flux difference formulation

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x}(F_{i+1/2}^n - F_{i-1/2}^n)\,. \tag{4.136}$$

Note that there is a vast amount of numerical flux functions and each of them defines its own finite volume method. In the following, we introduce a few sensible choices and explain why some famous choices are unsuitable.

4.8.8 Upwinding

Upwinding is a standard finite volume method. It approximates transport based on the direction of flow. In the following, we will introduce it based on the 1D advection equation with constant velocity

$$\rho_t + u\rho_x = 0\,. \tag{4.137}$$

Since the velocity $u \in \mathbb{R}$ is either positive or negative, the density is transported in only one direction over time.

As pointed out before, the advection is approximated by a numerical flux function, which is defined on the cell borders. For the standard upwind method with a constant velocity $u \geq 0$, the flux at the left cell border reads

$$F_{i-1/2}^n = u\rho_{i-1}^n\,, \tag{4.138}$$

since the density is transported from the left cell to the right with velocity u. Thus, the full upwind method for the linear 1D advection equation with positive velocity reads

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x}(u\rho_i^n - u\rho_{i-1}^n)\,. \tag{4.139}$$

Analogously, for $u \leq 0$

$$F^n_{i-1/2} = u\rho^n_i \tag{4.140}$$

and

$$\rho^{n+1}_i = \rho^n_i - \frac{\Delta t}{h_x}(u\rho^n_{i+1} - u\rho^n_i)\,. \tag{4.141}$$

Now, let us consider a space dependent velocity $u \cong u(x)$, i.e., its sign is variable and the considered advection equation is identical to (4.134). We substitute the velocity by its numerical approximation $u_{i-1/2} \approx u(x_{i-1/2})$, i.e., on a staggered grid. Then, we can separate $u_{i-1/2}$ into positive and negative values by

$$u^+_{i-1/2} = \max(u_{i-1/2}, 0) = \frac{u_{i-1/2} + |u_{i-1/2}|}{2}\,,$$
$$u^-_{i-1/2} = \min(u_{i-1/2}, 0) = \frac{u_{i-1/2} - |u_{i-1/2}|}{2}\,,$$

such that the numerical flux function is the sum of flux (4.138) and flux (4.140) and reads

$$F^n_{i-1/2} = u^+_{i-1/2}\rho^n_{i-1} + u^-_{i-1/2}\rho^n_i\,. \tag{4.142}$$

This leads to the full upwind method for the linear 1D advection equation with variable velocity

$$\rho^{n+1}_i = \rho^n_i - \frac{\Delta t}{h_x}(u^+_{i+1/2}\rho^n_i + u^-_{i+1/2}\rho^n_{i+1} - u^+_{i-1/2}\rho^n_{i-1} - u^-_{i-1/2}\rho^n_i)\,. \tag{4.143}$$

For an additionally time dependent velocity $u \cong u(x,t)$ with approximations $u^n_i \approx u(x_i, t_n)$, the velocity is split in every time step like above by

$$(u^n_{i-1/2})^+ = \max(u^n_{i-1/2}, 0) = \frac{u^n_{i-1/2} + |u^n_{i-1/2}|}{2}\,, \tag{4.144}$$
$$(u^n_{i-1/2})^- = \min(u^n_{i-1/2}, 0) = \frac{u^n_{i-1/2} - |u^n_{i-1/2}|}{2}\,. \tag{4.145}$$

Defining $u^+_{i-1/2} := (u^n_{i-1/2})^+$ and $u^-_{i-1/2} := (u^n_{i-1/2})^-$, the numerical flux functions and the resulting upwind method are identical to equations (4.142) and (4.143) above.

Note that to the best of our knowledge, there is no efficient way to upwind using implicit discretizations of the velocity. An iteration in each time step like Newton or fixed-point would be required to get the necessary information about the direction of flow, which raises the computational cost disproportionately, unless the discretized advection term is nonlinear. Hence, it is usual to solely consider an explicit discretization of the velocity.

Nevertheless, the time discretization of the density function can be implicit, with the typical consequence of having to solve a linear equation system for all cells to calculate ρ^{n+1}. Let us recall that the usage of partly implicit flux functions is necessary to ensure the thermodynamic consistency of some proposed phase-field models with flow. For this purpose, let us introduce the θ-formulation of the flux function

$$F^{n+\theta}_{i-1/2} = u^+_{i-1/2}\rho^{n+\theta}_{i-1} + u^-_{i-1/2}\rho^{n+\theta}_i\,, \tag{4.146}$$

with $\theta \in [0,1]$ and

$$\rho^{n+\theta} = \theta\rho^{n+1} + (1-\theta)\rho^n \,.$$

Here, $u^+_{i-1/2} = (u^n_{i-1/2})^+$ and $u^-_{i-1/2} = (u^n_{i-1/2})^-$, as defined in equations (4.144) and (4.145). For $\theta = 0$, we have the explicit flux (4.142), while for example $\theta = 1/2$ yields a Crank-Nicolson type flux and $\theta = 1$ an implicit flux. The resulting θ-formulation of the upwind method for the advection equation reads

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x}(u^+_{i+1/2}\rho_i^{n+\theta} + u^-_{i+1/2}\rho_{i+1}^{n+\theta} - u^+_{i-1/2}\rho_{i-1}^{n+\theta} - u^-_{i-1/2}\rho_i^{n+\theta}) \,, \tag{4.147}$$

where $\theta \in [0,1]$. For the sake of simplicity, let us define $\rho_i := \rho_i^{n+\theta}$ and $F_{i-1/2} := F^{n+\theta}_{i-1/2}$ for the introduction of additional finite volume methods in the following.

Flux vector splitting

The beforehand introduced flux difference formulation for finite volume methods (4.136), reading

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x}(F_{i+1/2} - F_{i-1/2}) \,,$$

can be rewritten to the so-called flux vector splitting formulation

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x}(u^+\Delta\rho_{i-1/2} + u^-\Delta\rho_{i+1/2}) \,, \tag{4.148}$$

e.g., by defining

$$\begin{aligned} u^+\Delta\rho_{i-1/2} &:= F_i - F_{i-1/2} \,, \\ u^-\Delta\rho_{i+1/2} &:= F_{i+1/2} - F_i \,. \end{aligned}$$

Since the cell-centered numerical flux F_i is obviously canceled out by direct insertion of the above defined flux splitting in formulation (4.148), it can be chosen arbitrarily. Let us set it to an approximation to the real flux in the cell center by

$$F_i = (u^+_{i-1/2} + u^-_{i+1/2})\rho_i \,,$$

such that the flux splitting together with the upwind flux function (4.142) reads

$$\begin{aligned} u^+\Delta\rho_{i-1/2} &= (u^+_{i-1/2} + u^-_{i+1/2})\rho_i - u^+_{i-1/2}\rho_{i-1} - u^-_{i-1/2}\rho_i \,, \\ u^-\Delta\rho_{i+1/2} &= u^+_{i+1/2}\rho_i + u^-_{i+1/2}\rho_{i+1} - (u^+_{i-1/2} + u^-_{i+1/2})\rho_i \,, \end{aligned}$$

which can be rearranged to

$$\begin{aligned} u^+\Delta\rho_{i-1/2} &= u^+_{i-1/2}(\rho_i - \rho_{i-1}) + (u^-_{i+1/2} - u^-_{i-1/2})\rho_i \,, \\ u^-\Delta\rho_{i+1/2} &= u^-_{i+1/2}(\rho_{i+1} - \rho_i) + (u^+_{i+1/2} - u^+_{i-1/2})\rho_i \,. \end{aligned}$$

This leads to the finite volume scheme

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x}\Big(u_{i-1/2}^+(\rho_i - \rho_{i-1}) + u_{i+1/2}^-(\rho_{i+1} - \rho_i) + (u_{i+1/2}^- - u_{i-1/2}^-)\rho_i + (u_{i+1/2}^+ - u_{i-1/2}^+)\rho_i \Big) . \tag{4.149}$$

Note that this scheme can be rearranged to

$$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} + u_{i-1/2}^+ \frac{\rho_i - \rho_{i-1}}{h_x} + u_{i+1/2}^- \frac{\rho_{i+1} - \rho_i}{h_x} + \frac{u_{i+1/2} - u_{i-1/2}}{h_x}\rho_i = 0 ,$$

i.e., Scheme (4.149) is a direct discretization of the nonlinear 1D advection equation, since

$$\rho_t + (up)_x = 0 \quad \Longleftrightarrow \quad \rho_t + u\rho_x + u_x p = 0 .$$

Let us recall that all our mathematical models with flow are divergence free. Thus, in 1D, it holds

$$u_x = 0 ,$$

which reads discretized by the central difference (4.110)

$$\frac{u_{i+1/2} - u_{i-1/2}}{h_x} = 0 .$$

Obviously, in one space dimension, the divergence-freeness is equivalent to a constant velocity. But since the derivation of the flux vector splitting for higher dimensions, which we will use later on, is analogous and the equivalence is not given there, we will neglect it at this point. Applying the discrete divergence-freeness to finite volume scheme (4.149) eliminates the last two terms and we have

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x}\Big(u_{i-1/2}^+(\rho_i - \rho_{i-1}) + u_{i+1/2}^-(\rho_{i+1} - \rho_i) \Big) . \tag{4.150}$$

An advantage of this upwind scheme for divergence free flow is that we may exchange the left and right differences

$$\frac{\rho_i - \rho_{i-1}}{h_x} \quad \text{and} \quad \frac{\rho_{i+1} - \rho_i}{h_x}$$

by higher order derivatives, to receive higher order finite volume schemes.

4.8.9 Second order upwinding finite difference method

Since we use second order finite differences for the spatial discretization of our models except advection and convection, see Subsection 4.8.1, it can be favorable to use a second order finite volume method for the advection and convection terms as well.

We just introduced upwind method (4.150) with interchangeable spatial discretizations of the density function derivative. Since we are upwinding, it is sensible to use one-sided differences. For a one-sided finite difference to be of second order in space, at least a three point stencil is required. Hence, we need the discrete Taylor series expansion (4.103) for $k = \pm 2$, which are right and left differences of first order. If we subtract them from two times

equation (4.103) for $k = \pm 1$, i.e., the original right and left differences, we get the left and right second order differences

$$\rho_x(x_i, t_n) \approx \partial_x^L \rho_i = \frac{3\rho_i - 4\rho_{i-1} + \rho_{i-2}}{2h_x}, \tag{4.151}$$

$$\rho_x(x_i, t_n) \approx \partial_x^R \rho_i = \frac{-3\rho_i + 4\rho_{i+1} - \rho_{i+2}}{2h_x}. \tag{4.152}$$

Exchanging the left and right discretizations of ρ_x in upwind scheme (4.150) with their second order counterparts (4.151) and (4.152) leads to the following second order upwinding finite difference scheme

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{2h_x}\left(u_{i-1/2}^+(3\rho_i - 4\rho_{i-1} + \rho_{i-2}) - u_{i+1/2}^-(3\rho_i - 4\rho_{i+1} + \rho_{i+2})\right). \tag{4.153}$$

Note that this is not a classical finite volume scheme, since we cannot rewrite it to the flux difference formulation (4.136). Therefore, it is not necessarily conservative. We will confirm this experimentally in Section 5.1.6. Hence, let us introduce further second order methods in the following.

Lax-Wendroff method

To construct second order finite volume methods from scratch, we can use the Taylor series expansion in time of the exact solution. For the sake of simplicity, we use the constant-coefficient advection equation (4.137) once again. The Taylor series expansion in time of its solution reads

$$p(x_i, t_{n+1}) = p(x_i, t_n) + \Delta t \rho_t(x_i, t_n) + \frac{1}{2}(\Delta t)^2 \rho_{tt}(x_i, t_n) + \mathcal{O}((\Delta t)^3). \tag{4.154}$$

Rearranging the advection equation (4.137) reads

$$\begin{aligned} \rho_t &= -u\rho_x \\ \implies \rho_{tt} &= -u\rho_{xt} = u^2\rho_{xx}. \end{aligned}$$

Substituting both time derivatives in the Taylor expansion (4.154) yields

$$p(x_i, t_{n+1}) = p(x_i, t_n) - \Delta t u \rho_x(x_i, t_n) + \frac{1}{2}(\Delta t)^2 u^2 \rho_{xx}(x_i, t_n) + \mathcal{O}((\Delta t)^3).$$

Replacing $p(x_i, t_{n+1})$ and $p(x_i, t_n)$ by their numerical approximations and the spatial derivatives by central finite difference approximations yields the famous *Lax-Wendroff method*, reading

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{2h_x} u(\rho_{i+1}^n - \rho_{i-1}^n) + \frac{1}{2}\left(\frac{\Delta t}{h_x}\right)^2 u^2(\rho_{i-1}^n - 2\rho_i^n + \rho_{i+1}^n) + \mathcal{O}((\Delta t)^3). \tag{4.155}$$

This method can be rewritten to

$$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} + \frac{1}{2h_x} u(\rho_{i+1}^n - \rho_{i-1}^n) - \frac{\Delta t}{2(h_x)^2} u^2(\rho_{i-1}^n - 2\rho_i^n + \rho_{i+1}^n) + \mathcal{O}((\Delta t)^2) = 0, \tag{4.156}$$

and is therefore a second order in time approximation to the advection equation.

Further, since the inserted finite differences

$$\rho_x(x_i, t_n) \approx \frac{\rho_{i+1}^n - \rho_{i-1}^n}{2h_x},$$
$$\rho_{xx}(x_i, t_n) \approx \frac{\rho_{i+1}^n - 2\rho_i^n + \rho_{i-1}^n}{h_x^2}$$

are of second order in space, as shown in equations (4.109) and (4.112), the Lax-Wendroff method is not only of second order in time but also in space. It is easy to verify that the Lax-Wendroff method is a finite volume method, since we can construct the following numerical flux function

$$F_{i-1/2}^n = \frac{1}{2}u(\rho_{i-1}^n + \rho_i^n) - \frac{1}{2}\frac{\Delta t}{h_x}u^2(\rho_i^n - \rho_{i-1}^n), \tag{4.157}$$

which yields the Lax-Wendroff method when inserted into the flux difference formulation (4.136).

Note that having a variable velocity increases the complexity of this scheme substantially, if it is considered in the derivation. The derivation has to be done for the whole solution vector $\boldsymbol{\omega} \in \mathbb{R}^k$, where $k \in \mathbb{N}$ is the number of required solutions, and the partial differential equation system should be a gradient flow, having the general formulation (in 1D)

$$\boldsymbol{\omega}_t + f(\boldsymbol{\omega})_x = 0.$$

Then, we can substitute this system of equations for the advection equation in the above derivation, yielding a Lax-Wendroff method for the entire system. Note that neither the q nor the σ evolution equation are gradient flows, since they include relaxation terms. Therefore, we cannot derive the Lax-Wendroff method for our viscoelastic two-fluid models with flow.

A much simpler approach is to generalize the numerical flux function (4.157) derived above, by substituting the constant velocity u by $u_{i-1/2}^n$, yielding

$$F_{i-1/2}^n = \frac{1}{2}u_{i-1/2}^n(\rho_{i-1}^n + \rho_i^n) - \frac{1}{2}\frac{\Delta t}{h_x}\left(u_{i-1/2}^n\right)^2(\rho_i^n - \rho_{i-1}^n). \tag{4.158}$$

Let us recall that some of our semi-discretizations in time necessitate an implicit solution ρ in the advection term in order to be provably energy-stable. For this purpose, we could further generalize flux function (4.158) to a θ-formulation. But note that this contradicts above derivation of the Lax-Wendroff method, since this derivation is based on an explicit time discretization of the model equations.

Nevertheless, let us complete our introduction of finite volume methods by the following two common modifications of the Taylor formula approach used for the Lax-Wendroff method.

Beam-Warming method

The Lax-Wendroff method is a central three point method, since central differences are used. Another common approach is again motivated by upwinding and therefore uses one-sided second order differences. For $u \geq 0$, they are given by the left difference (4.151) and

$$\rho_{xx}(x_i, t_n) \approx \partial_{xx}^L p = \frac{\rho_i^n - 2\rho_i^n + \rho_{i-2}^n}{h_x^2}, \tag{4.159}$$

where the latter is the rewritten sum of the discrete Taylor series expansion (4.102) for $k = -1$ and $k = -2$. Replacing the central differences in the Lax-Wendroff method by the one-sided differences (4.151) and (4.159) yields the *Beam-Warming method* for $u \geq 0$

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{2h_x} u(3\rho_i^n - 4\rho_{i-1}^n + \rho_{i-2}^n) + \frac{1}{2}\left(\frac{\Delta t}{h_x}\right)^2 u^2(\rho_i^n - 2\rho_{i-1}^n + \rho_{i-2}^n), \tag{4.160}$$

which is also of second order in time and space.
Again, we can verify that it is a finite volume method by constructing the flux function

$$F_{i-1/2}^n = u\rho_{i-1}^n + \frac{1}{2}u\left(1 - \frac{\Delta t}{h_x}u\right)(\rho_{i-1}^n - \rho_{i-2}^n), \tag{4.161}$$

which yields the Beam-Warming method for $u \geq 0$ when inserted into the flux difference formulation (4.136).
Analogously, the Beam-Warming method for $u \leq 0$ reads

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{2h_x} u(-\rho_{i+2}^n + 4\rho_{i+1}^n - 3\rho_i^n) + \frac{1}{2}\left(\frac{\Delta t}{h_x}\right)^2 u^2(\rho_{i+2}^n - 2\rho_{i+1}^n + \rho_i^n), \tag{4.162}$$

having the related flux function

$$F_{i-1/2}^n = u\rho_i^n - \frac{1}{2}u\left(1 + \frac{\Delta t}{h_x}u\right)(\rho_{i+1}^n - \rho_i^n). \tag{4.163}$$

Finally, analogously to the Lax-Wendroff flux, one may substitute the constant velocity u by $u_{i-1/2}^{+/-}$ and sum both fluxes, yielding

$$\begin{aligned} F_{i-1/2}^n = {} & u_{i-1/2}^+ \rho_{i-1}^n + \frac{1}{2}u_{i-1/2}^+\left(1 - \frac{\Delta t}{h_x}u_{i-1/2}^+\right)(\rho_{i-1}^n - \rho_{i-2}^n) \\ & + u_{i-1/2}^- \rho_i^n - \frac{1}{2}u_{i-1/2}^-\left(1 + \frac{\Delta t}{h_x}u_{i-1/2}^-\right)(\rho_{i+1}^n - \rho_i^n). \end{aligned} \tag{4.164}$$

Note that each first term of the Beam-Warming fluxes (4.161) and (4.163) equals the classical upwind fluxes (4.138) and (4.140), respectively. Thus, the second term can be considered as a correction, which shifts the method to second order in time and space. This is also fulfilled by the Lax-Wendroff flux (4.157), if rewritten suitably. We will introduce one last method in the following, which is particularly based on this property.

Further, note that due to the analogous derivation of the Beam-Warming and the Lax-Wendroff method, using an explicit time discretization of the solution ρ, a θ-formulation is contradicting and therefore neither of both is well suited for our problems.

Slope-limiter methods

Since second order methods tend to cause oscillations if the solution exhibits extrema and discontinuities, let us introduce the *slope-limiter method*, which equals the classical upwind

method with case dependent corrections, which are based on the slopes of the solution. The slope-limiter method for $u \geq 0$ reads

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x} u(\rho_i^n - \rho_{i-1}^n) - \frac{1}{2}\frac{\Delta t}{h_x} u \left(1 - \frac{\Delta t}{h_x} u\right)(\alpha_i^n - \alpha_{i-1}^n). \tag{4.165}$$

There are several options for choosing the slopes, e.g.

$$\alpha_i^n = 0\,, \tag{4.166}$$

$$\text{the upwind-slope} \quad \alpha_i^n = \rho_i^n - \rho_{i-1}^n\,, \tag{4.167}$$

$$\text{the downwind-slope} \quad \alpha_i^n = \rho_{i+1}^n - \rho_i^n\,. \tag{4.168}$$

The slope-limiter method with slope (4.166) equals the upwind method. With up- or downwind-slope, it equals the Beam-Warming or Lax-Wendroff method, respectively, for $u \geq 0$, and the other way round for $u \leq 0$.

A reasonable combination of the above-mentioned slopes is called the *minmod method*, which is an adaptive slope-limiter method based on the *minmod slope*

$$\alpha_i^n = \text{minmod}(\rho_i^n - \rho_{i-1}^n, \rho_{i+1}^n - \rho_i^n)\,. \tag{4.169}$$

The minmod function of two arguments is defined as

$$\text{minmod}(a,b) = \begin{cases} a & \text{if } |a| < |b| \text{ and } ab > 0\,, \\ b & \text{if } |b| < |a| \text{ and } ab > 0\,, \\ 0 & \text{if } ab \leq 0\,. \end{cases}$$

Thus, for a and b having the same sign, the function gives the lower absolute value, else zero. Instead of using the same slope in every cell, the minmod method compares the two slopes 4.167 and 4.168 in each cell and chooses the absolutely smaller one. If the slopes have varying signs, ρ_i^n is a local minimum or maximum and slope (4.166) is chosen.

The numerical flux, which yields the slope-limiter method for $u \geq 0$, when inserted in the flux difference formulation (4.136), reads

$$F_{i-1/2}^n = u\rho_{i-1}^n + \frac{1}{2} u \left(1 - \frac{\Delta t}{h_x} u\right) \alpha_{i-1}^n\,. \tag{4.170}$$

Analogously, the slope limiter method for $u \leq 0$ reads

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{h_x} u(\rho_{i+1}^n - \rho_i^n) + \frac{1}{2}\frac{\Delta t}{h_x} u \left(1 + \frac{\Delta t}{h_x} u\right)(\alpha_{i+1}^n - \alpha_i^n) \tag{4.171}$$

and the corresponding numerical flux function

$$F_{i-1/2}^n = u\rho_i^n - \frac{1}{2} u \left(1 + \frac{\Delta t}{h_x} u\right) \alpha_i^n\,. \tag{4.172}$$

Substituting the constant velocity u by $u^{+/-}_{i-1/2}$ and summing both fluxes yields

$$\begin{aligned} F^n_{i-1/2} =& u^+_{i-1/2}\rho^n_{i-1} + \frac{1}{2}u^+_{i-1/2}\left(1 - \frac{\Delta t}{h_x}u^+_{i-1/2}\right)\alpha^n_{i-1} \\ &+ u^-_{i-1/2}\rho^n_i - \frac{1}{2}u^-_{i-1/2}\left(1 + \frac{\Delta t}{h_x}u^-_{i-1/2}\right)\alpha^n_i \,. \end{aligned} \tag{4.173}$$

Note that the necessary knowledge about the sign of the slopes requires an explicit time discretization of the solution ρ or a costly iterative scheme like fixed-point or Newton in every time step. Therefore, this method is even less suitable for our problems.

4.8.10 Finite volume methods for momentum equations

To derive finite volume - finite difference schemes for the Navier-Stokes equations, let us first investigate its greatly reduced form, the inviscid Burgers' equation. This equation only consists of a convective term aside from the time derivative, which is the only part of the Navier-Stokes equations that we want to discretize with finite volumes. In one space dimension, the inviscid Burgers' equation reads

$$u_t + uu_x = 0 \quad \Longleftrightarrow \quad u_t + \frac{1}{2}(u^2)_x = 0\,,$$

which is obviously equivalent in its non-conservative (quasi-linear) and its conservative form, no matter if it is divergence free or not. Since we use a staggered grid, the velocity is given on the cell borders, i.e., we approximate the solution $u^n_{i-1/2} \approx u(x_{i-1/2}, t_n)$. Thus, adapting the standard upwind method (4.143) is done by an index shift and reads

$$u^{n+1}_{i-1/2} = u^n_{i-1/2} - \frac{\Delta t}{h_x}\left(u^+_i u_{i-1/2} + u^-_i u_{i+1/2} - u^+_{i-1}u_{i-3/2} - u^-_{i-1}u_{i-1/2}\right)\,, \tag{4.174}$$

where

$$\begin{aligned} u^+_i &= \max\left(\frac{u^n_{i-1/2} + u^n_{i+1/2}}{2}, 0\right) = \frac{u^n_{i-1/2} + u^n_{i+1/2} + \left|u^n_{i-1/2} + u^n_{i+1/2}\right|}{4}\,, \\ u^-_i &= \min\left(\frac{u^n_{i-1/2} + u^n_{i+1/2}}{2}, 0\right) = \frac{u^n_{i-1/2} + u^n_{i+1/2} - \left|u^n_{i-1/2} + u^n_{i+1/2}\right|}{4}\,. \end{aligned}$$

Since the necessary averaging over two neighboring cells introduces numerical diffusion, this seems unfavorable.

The adaption of the flux vector splitting method (4.150) is different in this case of having the whole convection term discretized on a single grid, instead of two staggered ones. It is achieved by discretizing the non-conservative formulation, while still using left and right differences for the spatial derivative, yielding

$$u^{n+1}_{i+1/2} = u^n_{i+1/2} - \frac{\Delta t}{h_x}\left(u^+_{i+1/2}\left(u_{i+1/2} - u_{i-1/2}\right) + u^-_{i+1/2}\left(u_{i+3/2} - u_{i+1/2}\right)\right)\,. \tag{4.175}$$

Again, we may exchange the left and right first order differences with their second order counterparts to get the second order accurate scheme

$$\begin{aligned} u_{i+1/2}^{n+1} = u_{i+1/2}^{n} - \frac{\Delta t}{2h_x}\Big(& u_{i+1/2}^{+}\left(3u_{i+1/2} - 4u_{i+1/2} + u_{i+3/2}\right) \\ & - u_{i+1/2}^{-}\left(3u_{i+1/2} - 4u_{i+3/2} + u_{i+5/2}\right)\Big). \end{aligned} \tag{4.176}$$

Note that also in this advantageous case of working on a single grid, the second order version of the flux vector splitting method cannot be written in flux difference formulation and is therefore not a classical finite volume method.

4.8.11 Two-dimensional finite volume methods

Since we are particularly interested in numerical solutions in two and three space dimensions, we will at first adapt some suitable aforementioned finite volume methods to two space dimensions in the following. Most adaptions are straight forward. The flux difference upwind method for the two-dimensional (2D) advection equation reads

$$\begin{aligned} \rho_{ij}^{n+1} &= \rho_{ij}^{n} - \Delta t \tilde{\nabla}_h \cdot F_{ij} = \rho_{ij}^{n} - \Delta t\left(\tilde{\partial}_x F_{ij} + \tilde{\partial}_y F_{ij}\right) \\ &= \rho_{ij}^{n} - \frac{\Delta t}{h_x}\left(F_{i+1/2,j} - F_{i-1/2,j}\right) - \frac{\Delta t}{h_y}\left(F_{i,j+1/2} - F_{i,j-1/2}\right) \\ &= \rho_{ij}^{n} - \frac{\Delta t}{h_x}\left(u_{i+1/2,j}^{+}\rho_{ij} + u_{i+1/2,j}^{-}\rho_{i+1,j} - u_{i-1/2,j}^{+}\rho_{i-1,j} - u_{i-1/2,j}^{-}\rho_{ij}\right) \\ &\quad - \frac{\Delta t}{h_y}\left(v_{i,j+1/2}^{+}\rho_{ij} + v_{i,j+1/2}^{-}\rho_{i,j+1} - v_{i,j-1/2}^{+}\rho_{i,j-1} - v_{i,j-1/2}^{-}\rho_{ij}\right). \end{aligned} \tag{4.177}$$

The direct adaption of the flux vector splitting to two space dimensions is called donor cell upwind (DCU) method, since it neglects diagonal cells. It reads

$$\begin{aligned} \rho_{ij}^{n+1} = \rho_{ij}^{n} &- \frac{\Delta t}{h_x}\left(u_{i-1/2,j}^{+}\left(\rho_{ij} - \rho_{i-1,j}\right) + u_{i+1/2,j}^{-}\left(\rho_{i+1,j} - \rho_{ij}\right)\right) \\ &- \frac{\Delta t}{h_y}\left(v_{i,j-1/2}^{+}\left(\rho_{ij} - \rho_{i,j-1}\right) + v_{i,j+1/2}^{-}\left(\rho_{i,j+1} - \rho_{ij}\right)\right). \end{aligned} \tag{4.178}$$

The second order version of the DCU method is gained analogously to the 1D case, by replacing the left and right differences, yielding

$$\begin{aligned} \rho_{ij}^{n+1} = \rho_{ij}^{n} &- \frac{\Delta t}{2h_x}\left(u_{i-1/2,j}^{+}\left(3\rho_{ij} - 4\rho_{i-1,j} + \rho_{i-2,j}\right) - u_{i+1/2,j}^{-}\left(3\rho_{ij} - 4\rho_{i+1,j} + \rho_{i+2,j}\right)\right) \\ &- \frac{\Delta t}{2h_y}\left(v_{i,j-1/2}^{+}\left(3\rho_{ij} - 4\rho_{i,j-1} + \rho_{i,j-2}\right) - v_{i,j+1/2}^{-}\left(3\rho_{ij} - 4\rho_{i,j+1} + \rho_{i,j+2}\right)\right). \end{aligned} \tag{4.179}$$

Flux difference upwinding for the inviscid Burgers' equation in 2D reads

$$\begin{aligned}
u_{i-1/2,j}^{n+1} &= u_{i-1/2,j}^{n} - \Delta t\, \tilde{\nabla}\cdot_h F_{i-1/2,j} = u_{i-1/2,j}^{n} - \Delta t \left(\tilde{\partial}_x F_{i-1/2,j} + \tilde{\partial}_y F_{i-1/2,j}\right) \\
&= u_{i-1/2,j}^{n} - \frac{\Delta t}{h_x}\left(F_{ij} - F_{i-1,j}\right) - \frac{\Delta t}{h_y}\left(F_{i-1/2,j+1/2} - F_{i-1/2,j-1/2}\right) \\
&= u_{i-1/2,j}^{n} - \frac{\Delta t}{h_x}\Big(u_{ij}^{+}u_{i-1/2,j} + u_{ij}^{-}u_{i+1/2,j} - u_{i-1,j}^{+}u_{i-3/2,j} - u_{i-1,j}^{-}u_{i-1/2,j}\Big) \\
&\qquad - \frac{\Delta t}{h_y}\Big(v_{i-1/2,j+1/2}^{+}u_{i-1/2,j} + v_{i-1/2,j+1/2}^{-}u_{i-1/2,j+1} \\
&\qquad\qquad - v_{i-1/2,j-1/2}^{+}u_{i-1/2,j-1} - v_{i-1/2,j-1/2}^{-}u_{i-1/2,j}\Big),
\end{aligned} \tag{4.180}$$

$$\begin{aligned}
v_{i,j-1/2}^{n+1} &= v_{i,j-1/2}^{n} - \Delta t\, \tilde{\nabla}\cdot_h F_{i,j-1/2} = v_{i,j-1/2}^{n} - \Delta t \left(\tilde{\partial}_x F_{i,j-1/2} + \tilde{\partial}_y F_{i,j-1/2}\right) \\
&= v_{i,j-1/2}^{n} - \frac{\Delta t}{h_x}\left(F_{i+1/2,j-1/2} - F_{i-1/2,j-1/2}\right) - \frac{\Delta t}{h_y}\left(F_{ij} - F_{i,j-1}\right) \\
&= v_{i,j-1/2}^{n} - \frac{\Delta t}{h_x}\Big(u_{i+1/2,j-1/2}^{+}v_{i,j-1/2} + u_{i+1/2,j-1/2}^{-}v_{i+1,j-1/2} \\
&\qquad\qquad - u_{i-1/2,j-1/2}^{+}v_{i-1,j-1/2} - u_{i-1/2,j-1/2}^{-}v_{i,j-1/2}\Big) \\
&\qquad - \frac{\Delta t}{h_y}\Big(v_{ij}^{+}v_{i,j-1/2} + v_{ij}^{-}v_{i,j+1/2} - v_{i,j-1}^{+}v_{i,j-3/2} - v_{i,j-1}^{-}v_{i,j-1/2}\Big),
\end{aligned} \tag{4.181}$$

where for any index and $\omega \in \{u, v\}$

$$\begin{aligned}
\omega_{ij}^{+} &= \max\left(\omega_{ij}^{n}, 0\right) = \frac{\omega_{ij}^{n} + \left|\omega_{ij}^{n}\right|}{2}, \\
\omega_{ij}^{-} &= \min\left(\omega_{ij}^{n}, 0\right) = \frac{\omega_{ij}^{n} - \left|\omega_{ij}^{n}\right|}{2}.
\end{aligned}$$

Note that due to the structure of our staggered grid, we have to use the following interpolations

$$\begin{aligned}
u_{ij}^{n} &= \frac{u_{i-1/2,j}^{n} + u_{i+1/2,j}^{n}}{2}, & u_{i-1/2,j-1/2}^{n} &= \frac{u_{i-1/2,j-1}^{n} + u_{i-1/2,j}^{n}}{2}, \\
v_{ij}^{n} &= \frac{v_{i,j-1/2}^{n} + v_{i,j+1/2}^{n}}{2}, & v_{i-1/2,j-1/2}^{n} &= \frac{v_{i-1,j-1/2}^{n} + v_{i,j-1/2}^{n}}{2}.
\end{aligned}$$

The DCU method for the inviscid Burgers' equation in 2D reads

$$\begin{aligned}
u_{i-1/2,j}^{n+1} = u_{i-1/2,j}^{n} &- \frac{\Delta t}{h_x}\Big(u_{i-1/2,j}^{+}\left(u_{i-1/2,j} - u_{i-3/2,j}\right) + u_{i-1/2,j}^{-}\left(u_{i+1/2,j} - u_{i-1/2,j}\right)\Big) \\
&- \frac{\Delta t}{h_y}\Big(v_{i-1/2,j}^{+}\left(u_{i-1/2,j} - u_{i-1/2,j-1}\right) + v_{i-1/2,j}^{-}\left(u_{i-1/2,j+1} - u_{i-1/2,j}\right)\Big),
\end{aligned} \tag{4.182}$$

$$\begin{aligned}
v_{i,j-1/2}^{n+1} = v_{i,j-1/2}^{n} &- \frac{\Delta t}{h_x}\Big(u_{i,j-1/2}^{+}\left(v_{i,j-1/2} - v_{i-1,j-1/2}\right) + u_{i,j-1/2}^{-}\left(v_{i+1,j-1/2} - v_{i,j-1/2}\right)\Big) \\
&- \frac{\Delta t}{h_y}\Big(v_{i,j-1/2}^{+}\left(v_{i,j-1/2} - v_{i,j-3/2}\right) + v_{i,j-1/2}^{-}\left(v_{i,j+1/2} - v_{i,j-1/2}\right)\Big),
\end{aligned} \tag{4.183}$$

where $u^{\pm}_{i,j-1/2}$ and $v^{\pm}_{i-1/2,j}$ necessitate the following bilinear interpolations

$$u^n_{i,j-1/2} = \frac{1}{4}\left(u^n_{i+1/2,j-1} + u^n_{i+1/2,j} + u^n_{i-1/2,j-1} + u^n_{i-1/2,j}\right),$$
$$v^n_{i-1/2,j} = \frac{1}{4}\left(v^n_{i-1,j+1/2} + v^n_{i,j+1/2} + v^n_{i-1,j-1/2} + v^n_{i,j-1/2}\right).$$

Analogously to above, we can replace the first order left, right, top and bottom differences of $u_{i-1/2,j}$ and $v_{i,j-1/2}$ in the DCU method by second order one-sided differences, to get the following second order DCU scheme for the Burgers' equation in 2D

$$\begin{aligned} u^{n+1}_{i-1/2,j} = u^n_{i-1/2,j} &- \frac{\Delta t}{2h_x}\Big(u^+_{i-1/2,j}\left(3u_{i-1/2,j} - 4u_{i-3/2,j} + u_{i-5/2,j}\right) \\ &\quad - u^-_{i-1/2,j}\left(3u_{i-1/2,j} - 4u_{i+1/2,j} + u_{i+3/2,j}\right)\Big) \\ &- \frac{\Delta t}{2h_y}\Big(v^+_{i-1/2,j}\left(3u_{i-1/2,j} - 4u_{i-1/2,j-1} + u_{i-1/2,j-2}\right) \\ &\quad - v^-_{i-1/2,j}\left(3u_{i-1/2,j} - 4u_{i-1/2,j+1} + u_{i-1/2,j+2}\right)\Big), \end{aligned} \tag{4.184}$$

$$\begin{aligned} v^{n+1}_{i,j-1/2} = v^n_{i,j-1/2} &- \frac{\Delta t}{h_x}\Big(u^+_{i,j-1/2}\left(3v_{i,j-1/2} - 4v_{i-1,j-1/2} + v_{i-2,j-1/2}\right) \\ &\quad - u^-_{i,j-1/2}\left(3v_{i,j-1/2} - 4v_{i+1,j-1/2} + v_{i+2,j-1/2}\right)\Big) \\ &- \frac{\Delta t}{h_y}\Big(v^+_{i,j-1/2}\left(3v_{i,j-1/2} - 4v_{i,j-3/2} + v_{i,j-5/2}\right) \\ &\quad + v^-_{i,j-1/2}\left(3v_{i,j-1/2} - 4v_{i,j+1/2} + v_{i,j+3/2}\right)\Big). \end{aligned} \tag{4.185}$$

The extensions to three space dimensions are straight forward and will therefore only be given in matrix formulations, which we will introduce in the following subsection.

Lemma 4.51. *(summation by parts for flux difference finite volume methods)*
Assuming suitable boundary conditions and

$$\tilde{\nabla}_h \cdot F_{ij} = \tilde{\partial}_x F_{ij} + \tilde{\partial}_y F_{ij} = \frac{1}{h_x}\left(F_{i+1/2,j} - F_{i-1/2,j}\right) + \frac{1}{h_y}\left(F_{i,j+1/2} - F_{i,j-1/2}\right),$$

the following summation by parts formulas hold for the flux difference upwind method (4.177) for advection, and with half index shifts for methods (4.180) and (4.181) for convection

$$\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \tilde{\nabla}_h \cdot F_{ij} = 0, \tag{4.186a}$$

$$\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \tilde{\nabla}_h \cdot F_{ij}\omega_{ij} = -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F^R_{ij} \cdot \nabla^R_h \omega_{ij}, \tag{4.186b}$$

$$\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \tilde{\nabla}_h \cdot F(\boldsymbol{u}_{ij}, \omega_{ij})\omega_{ij} = \frac{1}{2}\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(h_x |u_{i+1/2,j}| (\partial^R_x \omega_{ij})^2 + h_y |v_{i,j+1/2}| (\partial^R_y \omega_{ij})^2\right) \geq 0. \tag{4.186c}$$

Note that above formulas hold analogously for any space dimension. Further, note that they also hold analogously for the first order DCU method (4.178), since it can be rewritten to flux difference method (4.177).

Proof. **Part 1.** Using index shifting gives

$$\begin{aligned}
\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \tilde{\nabla}_h \cdot F_{ij} &= \frac{1}{h_x} \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(F_{i+1/2,j} - F_{i-1/2,j}\right) + \frac{1}{h_y} \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(F_{i,j+1/2} - F_{i,j-1/2}\right) \\
&= \frac{1}{h_x}\left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{i+1/2,j} - \sum_{\substack{i=0,\dots,N_x-1\\ j=1,\dots,N_y}} F_{i+1/2,j} \right) + \frac{1}{h_y}\left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{i,j+1/2} - \sum_{\substack{i=1,\dots,N_x\\ j=0,\dots,N_y-1}} F_{i,j+1/2} \right) \\
&= \frac{1}{h_x} \sum_{j=1}^{N_y} \left(F_{N_x+1/2,j} - F_{1/2,j}\right) + \frac{1}{h_y} \sum_{i=1}^{N_x} \left(F_{i,N_y+1/2} - F_{i,1/2}\right) = 0\,,
\end{aligned}$$

where the boundary sums vanish in the last equality due to assuming suitable boundary conditions, e.g., periodic or homogeneous Dirichlet.
Part 2. Using index shifting leads to

$$\begin{aligned}
\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \tilde{\nabla}_h \cdot F_{ij}\omega_{ij} &= \frac{1}{h_x} \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(F_{i+1/2,j} - F_{i-1/2,j}\right) \omega_{ij} + \frac{1}{h_y} \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(F_{i,j+1/2} - F_{i,j-1/2}\right) \omega_{ij} \\
&= \frac{1}{h_x}\left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{i+1/2,j}\omega_{ij} - \sum_{\substack{i=0,\dots,N_x-1\\ j=1,\dots,N_y}} F_{i+1/2,j}\omega_{i+1,j} \right) \\
&\quad + \frac{1}{h_y}\left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{i,j+1/2}\omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=0,\dots,N_y-1}} F_{i,j+1/2}\omega_{i,j+1} \right).
\end{aligned}$$

Assuming the same suitable boundary conditions as above, e.g., periodic for F_{ij} and ω_{ij}, or homogeneous Dirichlet for F_{ij}, we have

$$\begin{aligned}
\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \tilde{\nabla}_h \cdot F_{ij}\omega_{ij} &= \frac{1}{h_x}\left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{i+1/2,j}\omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{i+1/2,j}\omega_{i+1,j} \right) \\
&\quad + \frac{1}{h_y}\left(\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{i,j+1/2}\omega_{ij} - \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{i,j+1/2}\omega_{i,j+1} \right) \\
&= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(F_{i+1/2,j}\frac{\omega_{i+1,j} - \omega_{ij}}{h_x} + F_{i,j+1/2}\frac{\omega_{i,j+1} - \omega_{ij}}{h_y} \right) \\
&= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F_{ij}^R \cdot \nabla_h^R \omega_{ij}\,.
\end{aligned}$$

Part 3. Setting $F_{ij} := F(\boldsymbol{u}_{ij}, \omega_{ij})$ in (4.186b) gives

$$\begin{aligned}
&\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \tilde{\nabla}_h \cdot F(\boldsymbol{u}_{ij}, \omega_{ij})\omega_{ij} = -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} F(\boldsymbol{u}_{ij}, \omega_{ij})^R \cdot \nabla_h^R \omega_{ij} \\
&= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(F(u_{i+1/2,j}, \omega_{i+1/2,j}) \frac{\omega_{i+1,j} - \omega_{ij}}{h_x} + F(v_{i,j+1/2}, \omega_{i,j+1/2}) \frac{\omega_{i,j+1} - \omega_{ij}}{h_y} \right) \\
&= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \Bigg(\left(u^+_{i+1/2,j}\omega_{ij} + u^-_{i+1/2,j}\omega_{i+1,j} \right) \frac{\omega_{i+1,j} - \omega_{ij}}{h_x} \\
&\qquad + \left(v^+_{i,j+1/2}\omega_{ij} + v^-_{i,j+1/2}\omega_{i,j+1} \right) \frac{\omega_{i,j+1} - \omega_{ij}}{h_y} \Bigg) \\
&= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \Bigg(\left(\frac{u_{i+1/2,j} + |u_{i+1/2,j}|}{2}\omega_{ij} + \frac{u_{i+1/2,j} - |u_{i+1/2,j}|}{2}\omega_{i+1,j} \right) \frac{\omega_{i+1,j} - \omega_{ij}}{h_x} \\
&\qquad + \left(\frac{v_{i,j+1/2} + |v_{i,j+1/2}|}{2}\omega_{ij} + \frac{v_{i,j+1/2} - |v_{i,j+1/2}|}{2}\omega_{i,j+1} \right) \frac{\omega_{i,j+1} - \omega_{ij}}{h_y} \Bigg) \\
&= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \Bigg(\left(u_{i+1/2,j} \left(\omega_{i+1,j} + \omega_{ij} \right) - |u_{i+1/2,j}| \left(\omega_{i+1,j} - \omega_{ij} \right) \right) \frac{\omega_{i+1,j} - \omega_{ij}}{2h_x} \\
&\qquad + \left(v_{i,j+1/2} \left(\omega_{i,j+1} + \omega_{ij} \right) - |v_{i,j+1/2}| \left(\omega_{i,j+1} - \omega_{ij} \right) \right) \frac{\omega_{i,j+1} - \omega_{ij}}{2h_y} \Bigg) \\
&= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \Bigg(\frac{1}{2h_x} \left(u_{i+1/2,j} \left(\omega^2_{i+1,j} - \omega^2_{ij} \right) - |u_{i+1/2,j}| \left(\omega_{i+1,j} - \omega_{ij} \right)^2 \right) \\
&\qquad + \frac{1}{2h_y} \left(v_{i,j+1/2} \left(\omega^2_{i,j+1} - \omega^2_{ij} \right) - |v_{i,j+1/2}| \left(\omega_{i,j+1} - \omega_{ij} \right)^2 \right) \Bigg).
\end{aligned}$$

Assuming suitable boundary conditions, e.g., periodic for $\boldsymbol{u}_{ij}$ and ω_{ij}, or homogeneous Dirichlet for $\boldsymbol{u}_{ij}$, and using index shifting, we have

$$\begin{aligned}
\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \tilde{\nabla}_h \cdot F(\boldsymbol{u}_{ij},\omega_{ij})\omega_{ij} &= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \Bigg(\frac{1}{2h_x}\left(u_{i+1/2,j}\omega_{i+1,j}^2 - u_{i+1/2,j}\omega_{ij}^2 - |u_{i+1/2,j}|\left(\omega_{i+1,j}-\omega_{ij}\right)^2\right) \\
&\qquad + \frac{1}{2h_y}\left(v_{i,j+1/2}\omega_{i,j+1}^2 - v_{i,j+1/2}\omega_{ij}^2 - |v_{i,j+1/2}|\left(\omega_{i,j+1}-\omega_{ij}\right)^2\right)\Bigg) \\
&= -\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \Bigg(\frac{1}{2h_x}\left(u_{i-1/2,j}\omega_{ij}^2 - u_{i+1/2,j}\omega_{ij}^2 - |u_{i+1/2,j}|\left(\omega_{i+1,j}-\omega_{ij}\right)^2\right) \\
&\qquad + \frac{1}{2h_y}\left(v_{i,j-1/2}\omega_{ij}^2 - v_{i,j+1/2}\omega_{ij}^2 - |v_{i,j+1/2}|\left(\omega_{i,j+1}-\omega_{ij}\right)^2\right)\Bigg) \\
&= \sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \Bigg(\frac{u_{i+1/2,j}-u_{i-1/2,j}}{2h_x}\omega_{ij}^2 + \frac{1}{2h_x}|u_{i+1/2,j}|\left(\omega_{i+1,j}-\omega_{ij}\right)^2 \\
&\qquad + \frac{v_{i,j+1/2}-v_{i,j-1/2}}{2h_y}\omega_{ij}^2 + \frac{1}{2h_y}|v_{i,j+1/2}|\left(\omega_{i,j+1}-\omega_{ij}\right)^2\Bigg) \\
&= \frac{1}{2}\sum_{\substack{i=1,\dots,N_x\\ j=1,\dots,N_y}} \left(\omega_{ij}^2\tilde{\nabla}_h\cdot\boldsymbol{u}_{ij} + h_x|u_{i+1/2,j}|(\partial_x^R\omega_{ij})^2 + h_y|v_{i,j+1/2}|(\partial_y^R\omega_{ij})^2\right) .
\end{aligned}$$

Recalling that the velocity field is divergence free, it holds $\tilde{\nabla}_h\cdot\boldsymbol{u}_{ij}=0$, such that above equation becomes (4.186c). □

4.8.12 Finite volume matrices

Let us recall that in order to have thermodynamically consistent semi-discretizations in time, we have discretized some advection and convection terms of our phase-field models with hydrodynamics implicitly. Combined with the fact that similarly to finite differences, finite volume methods always require information of neighboring cells, we have to solve a coupled system for all cells. For this purpose, we will introduce matrix formulations for the three aforementioned 2D finite volume methods in the following, using vectorized discrete solutions.

The matrix notation of the one-dimensional upwind flux (4.142), reading

$$F_{i-1/2}^n = u_{i-1/2}^+\rho_{i-1}^n + u_{i-1/2}^-\rho_i^n ,$$

for all $i = 1, ..., N$ and homogeneous Dirichlet boundary conditions, i.e., $\rho_0 = 0 = \rho_{N+1}$, reads

$$\begin{aligned}
F_-^n &= \begin{pmatrix} u^+_{1/2} & & & \\ & u^+_{3/2} & & \\ & & \ddots & \\ & & & u^+_{N-1/2} \end{pmatrix} \begin{pmatrix} 0 \\ \rho_1^n \\ \vdots \\ \rho_{N-1}^n \end{pmatrix} + \begin{pmatrix} u^-_{1/2} & & & \\ & u^-_{3/2} & & \\ & & \ddots & \\ & & & u^-_{N-1/2} \end{pmatrix} \begin{pmatrix} \rho_1^n \\ \rho_2^n \\ \vdots \\ \rho_N^n \end{pmatrix} \\
&= \operatorname{diag}\begin{pmatrix} u^+_{1/2} \\ u^+_{3/2} \\ \vdots \\ u^+_{N-1/2} \end{pmatrix} \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix} \begin{pmatrix} \rho_1^n \\ \rho_2^n \\ \vdots \\ \rho_N^n \end{pmatrix} + \operatorname{diag}\begin{pmatrix} u^-_{1/2} \\ u^-_{3/2} \\ \vdots \\ u^-_{N-1/2} \end{pmatrix} \begin{pmatrix} \rho_1^n \\ \rho_2^n \\ \vdots \\ \rho_N^n \end{pmatrix} \\
&= \operatorname{diag}(u^+) L_N \rho_{1D}^n + \operatorname{diag}(u^-) \rho_{1D}^n \\
&= \left(\operatorname{diag}(u^+) L_N + \operatorname{diag}(u^-)\right) \rho_{1D}^n ,
\end{aligned}$$

where L_N is the lower $N \times N$ shift matrix for homogeneous Dirichlet boundary conditions. The lower shift matrix for periodic boundary conditions reads

$$L_N = \begin{pmatrix} 0 & & & 1 \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix},$$

and for homogeneous Neumann boundary conditions

$$L_N = \begin{pmatrix} 1 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}.$$

Analogously, the matrix notation for the right upwind flux $F^n_{i+1/2}$ reads

$$F_+^n = \left(\operatorname{diag}(R_N u^+) + \operatorname{diag}(R_N u^-) R_N\right) \rho_{1D}^n ,$$

where R_N is the upper $N \times N$ shift matrix. For homogeneous Dirichlet boundary conditions, it reads

$$R_N = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix},$$

for periodic boundary conditions

$$R_N = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ 1 & & & 0 \end{pmatrix},$$

and for homogeneous Neumann boundary conditions

$$R_N = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 1 \end{pmatrix}.$$

Matrix formulations for higher dimensions necessitate higher order shift matrices and a vectorization of the discrete solutions, as introduced in Subsection 4.8.4 for finite difference matrices. Therefore, we denote

$$\begin{aligned} u^+ + u^- = u_{2D/3D} &:= \mathrm{vec}(u)\,, & v^+ + v^- = v_{2D/3D} &:= \mathrm{vec}(v)\,, \\ w^+ + w^- = w_{3D} &:= \mathrm{vec}(w)\,, & \rho_{2D/3D} &:= \mathrm{vec}(\rho)\,. \end{aligned}$$

The higher order shift matrices are again constructed by Kronecker products, reading

$$\begin{aligned} R_{N_x^2} &= \boldsymbol{I}_{N_y} \otimes R_{N_x}\,, & L_{N_x^2} &= \boldsymbol{I}_{N_y} \otimes L_{N_x}\,, \\ R_{N_y^2} &= R_{N_y} \otimes \boldsymbol{I}_{N_x}\,, & L_{N_y^2} &= L_{N_y} \otimes \boldsymbol{I}_{N_x} \end{aligned}$$

in two space dimensions, and in three dimensions

$$\begin{aligned} L_{N_x^3} &= \boldsymbol{I}_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes L_{N_x}\,, & R_{N_x^3} &= \boldsymbol{I}_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes R_{N_x}\,, \\ L_{N_y^3} &= \boldsymbol{I}_{N_z} \otimes L_{N_y} \otimes \boldsymbol{I}_{N_x}\,, & R_{N_y^3} &= \boldsymbol{I}_{N_z} \otimes R_{N_y} \otimes \boldsymbol{I}_{N_x}\,, \\ L_{N_z^3} &= L_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \boldsymbol{I}_{N_x}\,, & R_{N_z^3} &= R_{N_z} \otimes \boldsymbol{I}_{N_y} \otimes \boldsymbol{I}_{N_x}\,. \end{aligned}$$

The 2D flux difference upwind fluxes read

$$\begin{aligned} F^n_{x,-} &= \left(\mathrm{diag}(u^+)L_{N_x^2} + \mathrm{diag}(u^-)\right)\rho^n_{2D}\,, \\ F^n_{x,+} &= \left(\mathrm{diag}(R_{N_x^2}u^+) + \mathrm{diag}(R_{N_x^2}u^-)R_{N_x^2}\right)\rho^n_{2D}\,, \\ F^n_{y,-} &= \left(\mathrm{diag}(v^+)L_{N_y^2} + \mathrm{diag}(v^-)\right)\rho^n_{2D}\,, \\ F^n_{y,+} &= \left(\mathrm{diag}(R_{N_y^2}v^+) + \mathrm{diag}(R_{N_y^2}v^-)R_{N_y^2}\right)\rho^n_{2D}\,. \end{aligned}$$

For 3D flux difference upwind fluxes, we replace the indices N^2 and $2D$ above by N^3 and $3D$, and add the two fluxes

$$\begin{aligned} F^n_{z,-} &= \left(\mathrm{diag}(w^+)L_{N_z^3} + \mathrm{diag}(w^-)\right)\rho^n_{3D}\,, \\ F^n_{z,+} &= \left(\mathrm{diag}(R_{N_z^3}w^+) + \mathrm{diag}(R_{N_z^3}w^-)R_{N_z^3}\right)\rho^n_{3D}\,. \end{aligned}$$

Summarized, the flux difference upwind method in matrix formulation for the 2D advection equation reads

$$\begin{aligned} \rho^{n+1}_{2D} &= \rho^n_{2D} - \frac{\Delta t}{h_x}\left(F^n_{x,+} - F^n_{x,-}\right) - \frac{\Delta t}{h_y}\left(F^n_{y,+} - F^n_{y,-}\right) \\ &= \rho^n_{2D} - \frac{\Delta t}{h_x}\left(\mathrm{diag}(R_{N_x^2}u^+) + \mathrm{diag}(R_{N_x^2}u^-)R_{N_x^2} - \mathrm{diag}(u^+)L_{N_x^2} - \mathrm{diag}(u^-)\right)\rho^n_{2D} \\ &\quad - \frac{\Delta t}{h_y}\left(\mathrm{diag}(R_{N_y^2}v^+) + \mathrm{diag}(R_{N_y^2}v^-)R_{N_y^2} - \mathrm{diag}(v^+)L_{N_y^2} - \mathrm{diag}(v^-)\right)\rho^n_{2D}\,. \end{aligned}$$

Introducing the flux difference upwind matrix

$$\begin{aligned} H(u,v) = &\frac{1}{h_x}\left(\operatorname{diag}(R_{N_x^2}u^+) + \operatorname{diag}(R_{N_x^2}u^-)R_{N_x^2} - \operatorname{diag}(u^+)L_{N_x^2} - \operatorname{diag}(u^-)\right) \\ &+ \frac{1}{h_y}\left(\operatorname{diag}(R_{N_y^2}v^+) + \operatorname{diag}(R_{N_y^2}v^-)R_{N_y^2} - \operatorname{diag}(v^+)L_{N_y^2} - \operatorname{diag}(v^-)\right), \end{aligned} \tag{4.187}$$

we can rewrite the above matrix formulation to a general finite volume system for the 2D advection equation, reading

$$\rho_{2D}^{n+1} = \rho_{2D}^n - \Delta t\, H(u,v)\rho_{2D}^n .$$

This system yields a different finite volume method depending on the chosen flux matrix $H(u,v)$. Let us also introduce the respective θ-Formulation

$$\rho_{2D}^{n+1} = \rho_{2D}^n - \Delta t\, H(u,v)\rho_{2D}^{n+\theta}, \tag{4.188}$$

which is implicit for $\theta \in (0,1]$.

In three space dimensions, the finite volume system reads

$$\rho_{3D}^{n+1} = \rho_{3D}^n - \Delta t\, H(u,v,w)\rho_{3D}^{n+\theta}, \tag{4.189}$$

where the flux difference upwind matrix is given by

$$\begin{aligned} H(u,v,w) = &\frac{1}{h_x}\left(\operatorname{diag}(R_{N_x^3}u^+) + \operatorname{diag}(R_{N_x^3}u^-)R_{N_x^3} - \operatorname{diag}(u^+)L_{N_x^3} - \operatorname{diag}(u^-)\right) \\ &+ \frac{1}{h_y}\left(\operatorname{diag}(R_{N_y^3}v^+) + \operatorname{diag}(R_{N_y^3}v^-)R_{N_y^3} - \operatorname{diag}(v^+)L_{N_y^3} - \operatorname{diag}(v^-)\right) \\ &+ \frac{1}{h_z}\left(\operatorname{diag}(w^+)L_{N_z^3} + \operatorname{diag}(w^-) - \operatorname{diag}(R_{N_z^3}w^+) - \operatorname{diag}(R_{N_z^3}w^-)R_{N_z^3}\right). \end{aligned} \tag{4.190}$$

Remark 4.52. *Note that the first two lines of $H(u,v,w)$, given by (4.190), are similar to $H(u,v)$, given by (4.187), except that the latter contains 2D matrices. Thus, let us introduce $\tilde{H}(u,v)$, which is a modification of $H(u,v)$, where the 2D matrices are replaced by their 3D counterparts. Then, equation (4.190) is equal to*

$$H(u,v,w) = \tilde{H}(u,v) + \frac{1}{h_z}\left(\operatorname{diag}(w^+)L_{N_z^3} + \operatorname{diag}(w^-) - \operatorname{diag}(R_{N_z^3}w^+) - \operatorname{diag}(R_{N_z^3}w^-)R_{N_z^3}\right),$$

i.e., the expansion to 3D only requires the addition of the flux for the third dimension.

Next, let us also introduce flux matrices for the first and second order DCU method for two and three space dimensions. In 2D, the flux matrix for the DCU method reads

$$\begin{aligned} H(u,v) = &\frac{1}{h_x}\left(\operatorname{diag}(u^+)(\boldsymbol{I}_{N^2} - L_{N_x^2}) + \operatorname{diag}(R_{N_x^2}u^-)(R_{N_x^2} - \boldsymbol{I}_{N^2})\right) \\ &+ \frac{1}{h_y}\left(\operatorname{diag}(v^+)(\boldsymbol{I}_{N^2} - L_{N_y^2}) + \operatorname{diag}(R_{N_y^2}v^-)(R_{N_y^2} - \boldsymbol{I}_{N^2})\right). \end{aligned}$$

As shown in Remark 4.52, the expansion to 3D only requires the addition of the flux for the third dimension, i.e.

$$H(u,v,w) = \tilde{H}(u,v) + \frac{1}{h_z}\left(\operatorname{diag}(w^+)(\boldsymbol{I}_{N^3} - L_{N_z^3}) + \operatorname{diag}(R_{N_z^3}w^-)(R_{N_z^3} - \boldsymbol{I}_{N^3})\right).$$

The second order DCU method in 2D is given by

$$\begin{aligned} H(u,v) = {} & \frac{1}{2h_x}\left(\operatorname{diag}(u^+)(3\boldsymbol{I}_{N^2} - 4L_{N_x^2} + (L_{N_x^2})^2) - \operatorname{diag}(R_{N_x^2}u^-)(-3\boldsymbol{I}_{N^2} - 4R_{N_x^2} + (R_{N_x^2})^2)\right) \\ & + \frac{1}{2h_y}\left(\operatorname{diag}(v^+)(3\boldsymbol{I}_{N^2} - 4L_{N_y^2} + (L_{N_y^2})^2) - \operatorname{diag}(R_{N_y^2}v^-)(-3\boldsymbol{I}_{N^2} - 4R_{N_y^2} + (R_{N_y^2})^2)\right) \end{aligned}$$

and in 3D by

$$\begin{aligned} H(u,v,w) = {} & \tilde{H}(u,v) \\ & + \frac{1}{2h_z}\left(\operatorname{diag}(w^+)(3\boldsymbol{I}_{N^3} - 4L_{N_z^3} + (L_{N_z^3})^2) - \operatorname{diag}(R_{N_z^3}w^-)(-3\boldsymbol{I}_{N^3} - 4R_{N_z^3} + (R_{N_z^3})^2)\right). \end{aligned}$$

Neglecting the dependencies of the flux matrices and the dimension indices in equations (4.188) and (4.189), we have the following general matrix formulation for finite volume methods, which holds true for any space dimension

$$\rho^{n+1} = \rho^n - \Delta t\, H\rho^{n+\theta}. \tag{4.191}$$

To enable the full discretization in matrix formulation of the full model for viscoelastic phase separation (3.51), the last missing factor is a finite volume matrix formulation of the momentum equations. For this purpose, let us use the Burgers' equation like before. Its finite volume matrix formulation in two space dimensions reads

$$u^{n+1} = u^n - \Delta t\, H_u u^{n+\theta}, \tag{4.192a}$$
$$v^{n+1} = v^n - \Delta t\, H_v v^{n+\theta}, \tag{4.192b}$$

and in three space dimensions we simply add

$$w^{n+1} = w^n - \Delta t\, H_w w^{n+\theta}. \tag{4.192c}$$

The flux matrices vary for each flux direction and of course by dimension. The flux difference upwind matrices in 2D read

$$\begin{aligned} H_u = {} & \frac{1}{h_x}\left(\operatorname{diag}(A^R_{N_x^2}u^+) + \operatorname{diag}(A^R_{N_x^2}u^-)R_{N_x^2} - \operatorname{diag}(A^L_{N_x^2}u^+)L_{N_x^2} - \operatorname{diag}(A^L_{N_x^2}u^-)\right) \\ & + \frac{1}{h_y}\left(\operatorname{diag}(A^L_{N_x^2}R_{N_y^2}v^+) + \operatorname{diag}(A^L_{N_x^2}R_{N_y^2}v^-)R_{N_y^2} - \operatorname{diag}(A^L_{N_x^2}v^+)L_{N_y^2} - \operatorname{diag}(A^L_{N_x^2}v^-)\right), \\ H_v = {} & \frac{1}{h_x}\left(\operatorname{diag}(A^L_{N_y^2}R_{N_x^2}u^+) + \operatorname{diag}(A^L_{N_y^2}R_{N_x^2}u^-)R_{N_x^2} - \operatorname{diag}(A^L_{N_y^2}u^+)L_{N_x^2} - \operatorname{diag}(A^L_{N_y^2}u^-)\right) \\ & + \frac{1}{h_y}\left(\operatorname{diag}(A^R_{N_y^2}v^+) + \operatorname{diag}(A^R_{N_y^2}v^-)R_{N_y^2} - \operatorname{diag}(A^L_{N_y^2}v^+)L_{N_y^2} - \operatorname{diag}(A^L_{N_y^2}v^-)\right), \end{aligned}$$

and in 3D

$$H_u = \tilde{H}_u + \frac{1}{h_z}\Big(\operatorname{diag}(A^L_{N^3_x} R_{N^3_z} w^+) + \operatorname{diag}(A^L_{N^3_x} R_{N^3_z} w^-) R_{N^3_z} - \operatorname{diag}(A^L_{N^3_x} w^+) L_{N^3_z} - \operatorname{diag}(A^L_{N^3_x} w^-)\Big), \tag{4.193a}$$

$$H_v = \tilde{H}_v + \frac{1}{h_x}\Big(\operatorname{diag}(A^L_{N^3_y} R_{N^3_z} w^+) + \operatorname{diag}(A^L_{N^3_y} R_{N^3_z} w^-) R_{N^3_z} - \operatorname{diag}(A^L_{N^3_y} w^+) L_{N^3_z} - \operatorname{diag}(A^L_{N^3_y} w^-)\Big), \tag{4.193b}$$

$$\begin{aligned} H_w = {} & \frac{1}{h_x}\Big(\operatorname{diag}(A^L_{N^3_z} R_{N^3_x} u^+) + \operatorname{diag}(A^L_{N^3_z} R_{N^3_x} u^-) R_{N^3_x} - \operatorname{diag}(A^L_{N^3_z} u^+) L_{N^3_x} - \operatorname{diag}(A^L_{N^3_z} u^-)\Big) \\ & + \frac{1}{h_y}\Big(\operatorname{diag}(A^L_{N^3_z} R_{N^3_y} v^+) + \operatorname{diag}(A^L_{N^3_z} R_{N^3_y} v^-) R_{N^3_y} - \operatorname{diag}(A^L_{N^3_z} v^+) L_{N^3_y} - \operatorname{diag}(A^L_{N^3_z} v^-)\Big) \\ & + \frac{1}{h_z}\Big(\operatorname{diag}(A^R_{N^3_z} w^+) + \operatorname{diag}(A^R_{N^3_z} w^-) R_{N^3_z} - \operatorname{diag}(A^L_{N^3_z} w^+) L_{N^3_z} - \operatorname{diag}(A^L_{N^3_z} w^-)\Big). \end{aligned} \tag{4.193c}$$

The DCU flux matrices in 2D read

$$\begin{aligned} H_u = {} & \frac{1}{h_x}\Big(\operatorname{diag}(u^+)(\boldsymbol{I}_{N^2} - L_{N^2_x}) + \operatorname{diag}(u^-)(R_{N^2_x} - \boldsymbol{I}_{N^2})\Big) \\ & + \frac{1}{h_y}\Big(\operatorname{diag}(A^L_{N^2_x} A^R_{N^2_y} v^+)(\boldsymbol{I}_{N^2} - L_{N^2_y}) + \operatorname{diag}(A^L_{N^2_x} A^R_{N^2_y} v^-)(R_{N^2_y} - \boldsymbol{I}_{N^2})\Big), \\ H_v = {} & \frac{1}{h_x}\Big(\operatorname{diag}(A^L_{N^2_y} A^R_{N^2_x} u^+)(\boldsymbol{I}_{N^2} - L_{N^2_x}) + \operatorname{diag}(A^L_{N^2_y} A^R_{N^2_x} u^-)(R_{N^2_x} - \boldsymbol{I}_{N^2})\Big) \\ & + \frac{1}{h_y}\Big(\operatorname{diag}(v^+)(\boldsymbol{I}_{N^2} - L_{N^2_y}) + \operatorname{diag}(v^-)(R_{N^2_y} - \boldsymbol{I}_{N^2})\Big), \end{aligned}$$

and in 3D

$$\begin{aligned} H_u = {} & \tilde{H}_u + \frac{1}{h_z}\Big(\operatorname{diag}(A^L_{N^3_x} A^R_{N^3_z} w^+)(\boldsymbol{I}_{N^3} - L_{N^3_z}) + \operatorname{diag}(A^L_{N^3_x} A^R_{N^3_z} w^-)(R_{N^3_z} - \boldsymbol{I}_{N^3})\Big), \\ H_v = {} & \tilde{H}_v + \frac{1}{h_z}\Big(\operatorname{diag}(A^L_{N^3_y} A^R_{N^3_z} w^+)(\boldsymbol{I}_{N^3} - L_{N^3_z}) + \operatorname{diag}(A^L_{N^3_y} A^R_{N^3_z} w^-)(R_{N^3_z} - \boldsymbol{I}_{N^3})\Big), \\ H_w = {} & \frac{1}{h_x}\Big(\operatorname{diag}(A^L_{N^3_z} A^R_{N^3_x} u^+)(\boldsymbol{I}_{N^3} - L_{N^3_x}) + \operatorname{diag}(A^L_{N^3_z} A^R_{N^3_x} u^-)(R_{N^3_x} - \boldsymbol{I}_{N^3})\Big) \\ & + \frac{1}{h_y}\Big(\operatorname{diag}(A^L_{N^3_z} A^R_{N^3_y} v^+)(\boldsymbol{I}_{N^3} - L_{N^3_y}) + \operatorname{diag}(A^L_{N^3_z} A^R_{N^3_y} v^-)(R_{N^3_y} - \boldsymbol{I}_{N^3})\Big) \\ & + \frac{1}{h_z}\Big(\operatorname{diag}(w^+)(\boldsymbol{I}_{N^3} - L_{N^3_z}) + \operatorname{diag}(w^-)(R_{N^3_z} - \boldsymbol{I}_{N^3})\Big). \end{aligned}$$

And finally, the second order DCU flux matrices in 2D read

$$\begin{aligned} H_u = {} & \frac{1}{h_x}\Big(\operatorname{diag}(u^+)(3\boldsymbol{I}_{N^2} - 4L_{N^2_x} + (L_{N^2_x})^2) - \operatorname{diag}(u^-)(-3\boldsymbol{I}_{N^2} - 4R_{N^2_x} + (R_{N^2_x})^2)\Big) \\ & + \frac{1}{h_y}\Big(\operatorname{diag}(A^L_{N^2_x} A^R_{N^2_y} v^+)(3\boldsymbol{I}_{N^2} - 4L_{N^2_y} + (L_{N^2_y})^2) + \operatorname{diag}(A^L_{N^2_x} A^R_{N^2_y} v^-)(-3\boldsymbol{I}_{N^2} - 4R_{N^2_y} + (R_{N^2_y})^2)\Big), \\ H_v = {} & \frac{1}{h_x}\Big(\operatorname{diag}(A^L_{N^2_y} A^R_{N^2_x} u^+)(3\boldsymbol{I}_{N^2} - 4L_{N^2_x} + (L_{N^2_x})^2) - \operatorname{diag}(A^L_{N^2_y} A^R_{N^2_x} u^-)(-3\boldsymbol{I}_{N^2} - 4R_{N^2_x} + (R_{N^2_x})^2)\Big) \\ & + \frac{1}{h_y}\Big(\operatorname{diag}(v^+)(3\boldsymbol{I}_{N^2} - 4L_{N^2_y} + (L_{N^2_y})^2) + \operatorname{diag}(v^-)(-3\boldsymbol{I}_{N^2} - 4R_{N^2_y} + (R_{N^2_y})^2)\Big), \end{aligned}$$

and in 3D

$$
\begin{aligned}
H_u &= \tilde{H}_u + \frac{1}{h_z}\Big(\operatorname{diag}(A^L_{N_x^3} A^R_{N_z^3} w^+)(3\boldsymbol{I}_{N^3} - 4L_{N_z^3} + (L_{N_z^3})^2) \\
&\qquad + \operatorname{diag}(A^L_{N_x^3} A^R_{N_z^3} w^-)(-3\boldsymbol{I}_{N^3} - 4R_{N_z^3} + (R_{N_z^3})^2)\Big), \\
H_v &= \tilde{H}_v + \frac{1}{h_z}\Big(\operatorname{diag}(A^L_{N_y^3} A^R_{N_z^3} w^+)(3\boldsymbol{I}_{N^3} - 4L_{N_z^3} + (L_{N_z^3})^2) \\
&\qquad + \operatorname{diag}(A^L_{N_y^3} A^R_{N_z^3} w^-)(-3\boldsymbol{I}_{N^3} - 4R_{N_z^3} + (R_{N_z^3})^2)\Big), \\
H_w &= \frac{1}{h_x}\Big(\operatorname{diag}(A^L_{N_z^3} A^R_{N_x^3} u^+)(3\boldsymbol{I}_{N^3} - 4L_{N_x^3} + (L_{N_x^3})^2) + \operatorname{diag}(A^L_{N_z^3} A^R_{N_x^3} u^-)(-3\boldsymbol{I}_{N^3} - 4R_{N_x^3} + (R_{N_x^3})^2)\Big) \\
&\quad + \frac{1}{h_y}\Big(\operatorname{diag}(A^L_{N_z^3} A^R_{N_y^3} v^+)(3\boldsymbol{I}_{N^3} - 4L_{N_y^3} + (L_{N_y^3})^2) + \operatorname{diag}(A^L_{N_z^3} A^R_{N_y^3} v^-)(-3\boldsymbol{I}_{N^3} - 4R_{N_y^3} + (R_{N_y^3})^2)\Big) \\
&\quad + \frac{1}{h_z}\Big(\operatorname{diag}(w^+)(3\boldsymbol{I}_{N^3} - 4L_{N_z^3} + (L_{N_z^3})^2) + \operatorname{diag}(w^-)(-3\boldsymbol{I}_{N^3} - 4R_{N_z^3} + (R_{N_z^3})^2)\Big).
\end{aligned}
$$

4.8.13 Full discretization of the Cahn-Hilliard-Navier-Stokes model

The Cahn-Hilliard-Navier-Stokes model (3.27) is the most reduced version of the full model for viscoelastic phase separation (3.51), which still contains every kind of spatial derivative which we are facing numerically. It includes the fourth order derivative for the diffusive interface of the Cahn-Hilliard equation, as well as advection and convection induced by the Navier-Stokes equations. Hence, let us introduce a full discretization in two and three space dimensions for this reduced model, before we finally completely discretize the full model in the next subsection.

Let us at first introduce a full discretization of the semi-discrete splitting scheme (4.58a), (4.62) for each grid cell in 2D. We use the OD2 approximation (4.26) for the potential derivative, a constant viscosity η, and discretize the advection term $\nabla \cdot (\boldsymbol{u}^n \varphi^n)$ with upwind finite volume method (4.177) and the convection term $(\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^\dagger$ with (4.180) and (4.181). The resulting

discretization reads

$$\varphi_{ij}^{n+1} = \varphi_{ij}^{n} - \Delta t \tilde{\nabla}_h \cdot F(\boldsymbol{u}_{ij}^{n}, \varphi_{ij}^{n}) + \frac{(\Delta t)^2}{\rho} \nabla_h^L \cdot \left(\left(\varphi_{ij}^{n}\right)_R^2 \nabla_h^R \mu_{ij}^{n+\frac{1}{2}} \right) + \Delta t \nabla_h^L \cdot \left(M(\varphi_{ij}^{n})^R \nabla_h^R \mu_{ij}^{n+\frac{1}{2}} \right), \tag{4.194a}$$

$$\mu_{ij}^{n+1/2} = -\lambda \Delta_h \frac{\varphi_{ij}^{n+1} + \varphi_{ij}^{n}}{2} + f(\varphi_{ij}^{n}) + \frac{\varphi_{ij}^{n+1} - \varphi_{ij}^{n}}{2} f'(\varphi_{ij}^{n}),$$

$$u_{i-1/2,j}^{\dagger} = u_{i-1/2,j}^{n} - \Delta t \tilde{\nabla}_h \cdot F_u(\boldsymbol{u}_{i-1/2,j}^{n}, u_{i-1/2,j}^{\dagger}) + \frac{\Delta t}{\rho} \eta \Delta_h u_{i-1/2,j}^{\dagger} - \frac{\Delta t}{\rho} \varphi_{i-1/2,j}^{n} \partial_x^L \mu_{ij}^{n+\frac{1}{2}}, \tag{4.194b}$$

$$u_{i-1/2,j}^{\dagger} = u_{i-1/2,j}^{n+1} + \frac{\Delta t}{\rho} \partial_x^L p_{ij}^{n+1}, \tag{4.194c}$$

$$v_{i,j-1/2}^{\dagger} = v_{i,j-1/2}^{n} - \Delta t \tilde{\nabla}_h \cdot F_v(\boldsymbol{u}_{i,j-1/2}^{n}, v_{i,j-1/2}^{\dagger}) + \frac{\Delta t}{\rho} \eta \Delta_h v_{i,j-1/2}^{\dagger} - \frac{\Delta t}{\rho} \varphi_{i,j-1/2}^{n} \partial_y^L \mu_{ij}^{n+\frac{1}{2}}, \tag{4.194d}$$

$$v_{i,j-1/2}^{\dagger} = v_{i,j-1/2}^{n+1} + \frac{\Delta t}{\rho} \partial_y^L p_{ij}^{n+1}, \tag{4.194e}$$

$$\tilde{\nabla}_h \cdot \boldsymbol{u}_{ij}^{n+1} = 0. \tag{4.194f}$$

Theorem 4.53. *Scheme (4.194) is linear and, assuming suitable boundary conditions, see Remark 3.7, it is mass-conservative and satisfies the following discrete energy law*

$$\frac{E_{total}(\varphi^{n+1}, \boldsymbol{u}^{n+1}) - E_{total}(\varphi^{n}, \boldsymbol{u}^{n})}{\Delta t} = -\mathrm{ND}_{pot}^{n+1} - \mathrm{ND}_{split}^{n+1} - \mathrm{ND}_{FV}^{n+1} - \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} M(\varphi_{ij}^{n})^R |\nabla_h^R \mu_{ij}^{n+\frac{1}{2}}|^2 - \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \eta \left(|\nabla_h^R u_{i-1/2,j}^{n+1}|^2 + |\nabla_h^R v_{i,j-1/2}^{n+1}|^2 \right), \tag{4.195}$$

where

$$\mathrm{ND}_{pot}^{n+1} = -\frac{1}{6}(\Delta t)^2 \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} f''(\zeta) \left(\partial_t \varphi_{ij}^{n+\frac{1}{2}} \right)^3 \in \mathcal{O}\left((\Delta t)^2\right),$$

$$\mathrm{ND}_{split}^{n+1} = \frac{\rho}{2\,\Delta t} \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \Big((u_{i-1/2,j}^{n+1} - u_{i-1/2,j}^{*})^2 + (u_{i-1/2,j}^{*} - u_{i-1/2,j}^{n})^2 + (v_{i,j-1/2}^{n+1} - v_{i,j-1/2}^{*})^2 + (v_{i,j-1/2}^{*} - v_{i,j-1/2}^{n})^2 \Big) \geq 0,$$

$$u_{i-1/2,j}^{*} = u_{i-1/2,j}^{n} - \frac{\Delta t}{\rho} \varphi_{i-1/2,j}^{n} \partial_x^L \mu_{ij}^{n+\frac{1}{2}}, \qquad v_{i,j-1/2}^{*} = v_{i,j-1/2}^{n} - \frac{\Delta t}{\rho} \varphi_{i,j-1/2}^{n} \partial_y^L \mu_{ij}^{n+\frac{1}{2}},$$

$$\mathrm{ND}_{FV}^{n+1} = \frac{1}{2} \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \Big(h_x \left(|u_{i+1/2,j}^{n}| (\partial_x^R \varphi_{ij}^{n})^2 + |u_{ij}^{n}| (\partial_x^R u_{i-1/2,j}^{\dagger})^2 + |u_{i+1/2,j-1/2}^{n}| (\partial_x^R v_{i,j-1/2}^{\dagger})^2 \right) + h_y \left(|v_{i,j+1/2}^{n}| (\partial_y^R \varphi_{ij}^{n})^2 + |v_{i-1/2,j+1/2}^{n}| (\partial_y^R u_{i-1/2,j}^{\dagger})^2 + |v_{ij}^{n}| (\partial_y^R v_{i,j-1/2}^{\dagger})^2 \right) \Big) \geq 0.$$

This energy law is – up to the numerical diffusion caused by the finite volume method – analogous to its semi-discrete version (4.52).

Proof. It is clear that the proposed scheme is linear.

Summing equation (4.194a) over all $i = 1, ..., N_x$ and $j = 1, ..., N_y$, and using the summation by parts formulas (4.186a) and (4.125d), we have

$$\begin{aligned}\sum_{\substack{i=1,...,N_x\\ j=1,...,N_y}} \varphi_{ij}^{n+1} &= \sum_{\substack{i=1,...,N_x\\ j=1,...,N_y}} \varphi_{ij}^{n} - \Delta t \sum_{\substack{i=1,...,N_x\\ j=1,...,N_y}} \tilde{\nabla}_h \cdot F(\boldsymbol{u}_{ij}^n, \varphi_{ij}^n) + \Delta t \sum_{\substack{i=1,...,N_x\\ j=1,...,N_y}} \nabla_h^L \cdot \left(M(\varphi_{ij}^n)^R \nabla_h^R \mu_{ij}^{n+1/2} \right) \\ &= \sum_{\substack{i=1,...,N_x\\ j=1,...,N_y}} \varphi_{ij}^{n} .\end{aligned}$$

Thus, the scheme is mass-conservative.

To calculate the discrete energy law, we substitute $\boldsymbol{u}^\dagger$ defined by (4.194c) and (4.194e) in (4.194b) and (4.194d), respectively. Then, we multiply equation (4.194a) by $\mu_{ij}^{n+1/2}$, (4.194b) by $u_{i-1/2,j}^{n+1}$ and (4.194d) by $v_{i,j-1/2}^{n+1}$. Summing over all $i = 1, ..., N_x$ and $j = 1, ..., N_y$, and using the summation by parts formulas (4.125a), (4.125c) and (4.186c), the discrete energy law (4.195) is calculated analogously to energy law (4.52) of the semi-discrete splitting scheme. The additional terms in the Navier-Stokes equations from the Chorin projection vanish during the computations, analogously to the proof of Theorem 4.27, since we assume a constant viscosity. The additional numerical dissipation ND_{FV}^{n+1} is caused by the finite volume method, see (4.186c). The numerical dissipation ND_{pot}^{n+1} caused by the OD2 approximation is calculated analogously to (4.27). □

Let us recall that we have to compute some spatial derivatives implicitly in order to be energy-stable. Thus, we have to solve coupled systems for all grid cells, which will be introduced in the following.

We use the general matrix formulation (4.191) to rewrite the Cahn-Hilliard equation (4.194a), where analogously to above, we only discretize the advection part $\nabla\cdot(\boldsymbol{u}^n\varphi^n)$ with finite volumes and the other derivatives with finite differences. This yields

$$\begin{aligned}\varphi^{n+1} &= \varphi^n - \Delta t H \varphi^n + \frac{(\Delta t)^2}{\rho} \nabla_{N^d}^L \cdot \left(\mathrm{diag}\left(A^R \varphi^n\right)^2 \nabla_{N^d}^R \mu^{n+1/2} \right) + \Delta t \nabla_{N^d}^L \cdot \left(\mathrm{diag}(M(\varphi^n)^R) \nabla_{N^d}^R \mu^{n+1/2} \right) \\ \iff & \left\{ \boldsymbol{I}_{N^d} - \Delta t \sum_{a \in \alpha_d} \left(\partial_a^L \left[\frac{\Delta t}{\rho} \mathrm{diag}\left(A_a^R \varphi^n\right)^2 \right.\right.\right. \\ & \qquad \left.\left.\left. + \mathrm{diag}\left(A_a^R M(\varphi^n)\right) \right] \partial_a^R \right) \left[\frac{-\lambda}{2} \Delta_{N^d} + \frac{1}{2} \mathrm{diag}\left(f'(\varphi^n)\right) \right] \right\} \varphi^{n+1} \\ &= \varphi^n - \Delta t H \varphi^n \\ & \quad + \Delta t \sum_{a \in \alpha_d} \left(\partial_a^L \left[\frac{\Delta t}{\rho} \mathrm{diag}\left(A_a^R \varphi^n\right)^2 \right.\right. \\ & \qquad \left.\left. + \mathrm{diag}\left(A_a^R M(\varphi^n)\right) \right] \partial_a^R \right) \left[\frac{-\lambda}{2} \Delta_{N^d} \varphi^n + f(\varphi^n) - \frac{1}{2} \mathrm{diag}\left(f'(\varphi^n)\right) \varphi^n \right] ,\end{aligned} \tag{4.196}$$

where the index set α_d is based on the respective dimension and reads

$$\alpha_1 = N, \quad \alpha_2 = \{N_x^2, N_y^2\}, \quad \text{or} \quad \alpha_3 = \{N_x^3, N_y^3, N_z^3\}.$$

A full matrix discretization of the Chorin projection algorithm to solve the Navier-Stokes equations (4.194b)-(4.194f) numerically in 3D, using the general matrix formulation (4.192), reads as follows.

Step 1. For a variable viscosity η, we have to solve the coupled system

$$\begin{pmatrix} \mathrm{LHS}_{11} & \mathrm{LHS}_{12} & \mathrm{LHS}_{13} \\ \mathrm{LHS}_{21} & \mathrm{LHS}_{22} & \mathrm{LHS}_{23} \\ \mathrm{LHS}_{31} & \mathrm{LHS}_{32} & \mathrm{LHS}_{33} \end{pmatrix} \begin{pmatrix} u^\dagger \\ v^\dagger \\ w^\dagger \end{pmatrix} = \begin{pmatrix} \mathrm{RHS}_1 \\ \mathrm{RHS}_2 \\ \mathrm{RHS}_3 \end{pmatrix}, \tag{4.197a}$$

where

$$\mathrm{RHS}_1 = u^n - \frac{\Delta t}{\rho} \mathrm{diag}\left(A^L_{N_x^3} \varphi^n\right) \partial^L_{N_x^3} \mu^{n+1/2}, \tag{4.197b}$$

$$\mathrm{RHS}_2 = v^n - \frac{\Delta t}{\rho} \mathrm{diag}\left(A^L_{N_y^3} \varphi^n\right) \partial^L_{N_y^3} \mu^{n+1/2}, \tag{4.197c}$$

$$\mathrm{RHS}_3 = w^n - \frac{\Delta t}{\rho} \mathrm{diag}\left(A^L_{N_z^3} \varphi^n\right) \partial^L_{N_z^3} \mu^{n+1/2}, \tag{4.197d}$$

and

$$\mathrm{LHS}_{11} = \boldsymbol{I}_{N^3} + \Delta t H_u - \frac{\Delta t}{\rho} \left(\sum_{a \in \alpha_d} \partial_a \mathrm{diag}(A^L_{N_x^3} \eta(\varphi^n)) \partial_a + \partial_{N_x^3} \mathrm{diag}(A^L_{N_x^3} \eta(\varphi^n)) \partial_{N_x^3} \right), \tag{4.197e}$$

$$\mathrm{LHS}_{12} = -\frac{\Delta t}{\rho} \partial_{N_y^3} \mathrm{diag}(A^L_{N_x^3} \eta(\varphi^n)) \partial^L_{N_x^3} A^R_{N_y^3}, \tag{4.197f}$$

$$\mathrm{LHS}_{13} = -\frac{\Delta t}{\rho} \partial_{N_z^3} \mathrm{diag}(A^L_{N_x^3} \eta(\varphi^n)) \partial^L_{N_x^3} A^R_{N_z^3}, \tag{4.197g}$$

$$\mathrm{LHS}_{21} = -\frac{\Delta t}{\rho} \partial_{N_x^3} \mathrm{diag}(A^L_{N_y^3} \eta(\varphi^n)) \partial^L_{N_y^3} A^R_{N_x^3}, \tag{4.197h}$$

$$\mathrm{LHS}_{22} = \boldsymbol{I}_{N^3} + \Delta t H_v - \frac{\Delta t}{\rho} \left(\sum_{a \in \alpha_d} \partial_a \mathrm{diag}(A^L_{N_y^3} \eta(\varphi^n)) \partial_a + \partial_{N_y^3} \mathrm{diag}(A^L_{N_y^3} \eta(\varphi^n)) \partial_{N_y^3} \right), \tag{4.197i}$$

$$\mathrm{LHS}_{23} = -\frac{\Delta t}{\rho} \partial_{N_z^3} \mathrm{diag}(A^L_{N_y^3} \eta(\varphi^n)) \partial^L_{N_y^3} A^R_{N_z^3}, \tag{4.197j}$$

$$\mathrm{LHS}_{31} = -\frac{\Delta t}{\rho} \partial_{N_x^3} \mathrm{diag}(A^L_{N_z^3} \eta(\varphi^n)) \partial^L_{N_z^3} A^R_{N_x^3}, \tag{4.197k}$$

$$\mathrm{LHS}_{32} = -\frac{\Delta t}{\rho} \partial_{N_y^3} \mathrm{diag}(A^L_{N_z^3} \eta(\varphi^n)) \partial^L_{N_z^3} A^R_{N_y^3}, \tag{4.197l}$$

$$\mathrm{LHS}_{33} = \boldsymbol{I}_{N^3} + \Delta t H_w - \frac{\Delta t}{\rho} \left(\sum_{a \in \alpha_d} \partial_a \mathrm{diag}(A^L_{N_z^3} \eta(\varphi^n)) \partial_a + \partial_{N_z^3} \mathrm{diag}(A^L_{N_z^3} \eta(\varphi^n)) \partial_{N_z^3} \right). \tag{4.197m}$$

Step 2. Using the incompressibility, we find p^{n+1} by solving

$$\Delta_{N^3} p^{n+1} = \frac{\rho}{\Delta t} \nabla^R_{N^3} \cdot \boldsymbol{u}^\dagger = \frac{\rho}{\Delta t} \left(\partial^R_{N_x^3} u^\dagger + \partial^R_{N_y^3} v^\dagger + \partial^R_{N_z^3} w^\dagger \right). \tag{4.197n}$$

Step 3. Finally, we can calculate the discrete solutions at the new time t^{n+1} by

$$u^{n+1} = u^\dagger - \frac{\Delta t}{\rho}\partial^L_{N_x^3} p^{n+1}\,, \tag{4.197o}$$

$$v^{n+1} = v^\dagger - \frac{\Delta t}{\rho}\partial^L_{N_y^3} p^{n+1}\,, \tag{4.197p}$$

$$w^{n+1} = w^\dagger - \frac{\Delta t}{\rho}\partial^L_{N_z^3} p^{n+1}\,. \tag{4.197q}$$

Remark 4.54. *Note that we can reduce the spatial dimension of above algorithm to 2D by*

- *eliminating the third row of system (4.197a),*
- *eliminating the right column of the system block matrix,*
- *fixing $w = \mathbf{0}$, and*
- *substituting the index N^3 by N^2 everywhere.*

The 1D case is trivial, since the incompressibility is equivalent to a constant velocity in this situation.

Remark 4.55. *Let us recall from Theorem 3.5 that for a constant viscosity η, it holds*

$$\nabla \cdot (2\eta \boldsymbol{D}(\boldsymbol{u})) = \eta \nabla \cdot \left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right) = \eta \Delta \boldsymbol{u}\,.$$

Therefore, if the viscosity is constant, we can significantly simplify the block matrices (4.197e)-(4.197m) as follows

$$\mathrm{LHS}_{11} = \boldsymbol{I}_{N^3} + \Delta t H_u - \frac{\Delta t}{\rho}\eta \Delta_{N^3}\,, \tag{4.198a}$$

$$\mathrm{LHS}_{22} = \boldsymbol{I}_{N^3} + \Delta t H_v - \frac{\Delta t}{\rho}\eta \Delta_{N^3}\,, \tag{4.198b}$$

$$\mathrm{LHS}_{33} = \boldsymbol{I}_{N^3} + \Delta t H_w - \frac{\Delta t}{\rho}\eta \Delta_{N^3}\,, \tag{4.198c}$$

$$\mathrm{LHS}_{12} = \mathrm{LHS}_{13} = \mathrm{LHS}_{21} = \mathrm{LHS}_{23} = \mathrm{LHS}_{31} = \mathrm{LHS}_{32} = \mathbf{0}\,. \tag{4.198d}$$

This removes the coupling from the Navier-Stokes equations and therefore reduces the computational effort considerably. Instead of solving the coupled system (4.197a), we only have to solve the following three equations successively in step one of the algorithm

$$\mathrm{LHS}_{11} u^\dagger = \mathrm{RHS}_1\,, \tag{4.199a}$$

$$\mathrm{LHS}_{22} v^\dagger = \mathrm{RHS}_2\,, \tag{4.199b}$$

$$\mathrm{LHS}_{33} w^\dagger = \mathrm{RHS}_3\,. \tag{4.199c}$$

Theorem 4.56. *Let H, H_u, H_v, H_w represent the flux difference upwind matrices (4.190) and (4.193). Then, the resulting completely discretized numerical scheme (4.196), (4.197) is linear and, assuming suitable boundary conditions, see Remark 3.7, it is mass-conservative. Further, assuming a constant viscosity η, the scheme can be significantly simplified, see Remark 4.55. This simplified scheme satisfies the discrete energy law (4.195) in 2D (considering non-vectorized discrete solutions) and an analogous version in 3D.*

Proof. See the proof of Theorem 4.53, since the proposed scheme is in 2D (up to the vectorized discrete solutions) equivalent to (4.194). For three dimensions, the proof is analogous, since the SBP formulas (4.125) and (4.186) hold for any space dimension. □

4.8.14 Full discretization of the full model

Let us finally introduce a complete matrix discretization to solve the full model for viscoelastic phase separation (3.51) numerically. This discretization is based on the semi-discrete Chorin projection splitting scheme (4.77), (4.85), (4.79), using the OD2 approximation (4.26) for the potential derivative.

At first, we solve the linear equation system given by the full matrix discretization of the coupled semi-discrete (φ, q)-system (4.77). Note that this discretization is – up to the advection terms – analogous to the fully discretized simplified model (4.133). It reads

$$\begin{pmatrix} \mathrm{LHS}_{11} & \mathrm{LHS}_{12} \\ \mathrm{LHS}_{21} & \mathrm{LHS}_{22} \end{pmatrix} \begin{pmatrix} \varphi^{n+1} \\ q^{n+1} \end{pmatrix} = \begin{pmatrix} \mathrm{RHS}_1 \\ \mathrm{RHS}_2 \end{pmatrix}, \tag{4.200a}$$

where

$$\begin{aligned} \mathrm{LHS}_{11} = \boldsymbol{I}_{N^d} - \Delta t \sum_{a \in \alpha_d} \Bigg(\partial_a^L \Bigg[\frac{\Delta t}{\rho} \mathrm{diag}\left(A_a^R \varphi^n\right)^2 \\ + \mathrm{diag}\left(A_a^R M(\varphi^n)\right) \Bigg] \partial_a^R \Bigg) \left[\frac{-\lambda}{2} \Delta_{N^d} + \frac{1}{2} \mathrm{diag}\left(f'(\varphi^n)\right) \right], \end{aligned} \tag{4.200b}$$

$$\mathrm{LHS}_{12} = \frac{\Delta t}{2} \left(\sum_{a \in \alpha_d} \partial_a^L \mathrm{diag}\left(A_a^R M_2(\varphi^n)\right) \partial_a^R \right) \mathrm{diag}(G_B(\varphi^n)), \tag{4.200c}$$

$$\begin{aligned} \mathrm{LHS}_{21} = \Delta t \, \mathrm{diag}\left(G_B(\varphi^n)\right) \left(\sum_{a \in \alpha_d} \partial_a^L \mathrm{diag}\left(A_a^R M_2(\varphi^n)\right) \partial_a^R \right) \Bigg[\frac{-\lambda}{2} \Delta_{N^d} \\ + \frac{1}{2} \mathrm{diag}\left(f'(\varphi^n)\right) \Bigg], \end{aligned} \tag{4.200d}$$

$$\begin{aligned} \mathrm{LHS}_{22} = \boldsymbol{I}_{N^d} + \frac{\Delta t}{2} \Bigg(H + \mathrm{diag}\left(\tau_B(\varphi^n)\right)^{-1} \\ + \mathrm{diag}\left(G_B(\varphi^n)\right) \left(\sum_{a \in \alpha_d} \partial_a^L \mathrm{diag}\left(A_a^R \zeta(\varphi^n)\right)^{-1} \partial_a^R \right) \mathrm{diag}\left(G_B(\varphi^n)\right) \Bigg) \end{aligned} \tag{4.200e}$$

and

$$\begin{aligned}\mathrm{RHS}_1 = \varphi^n - \Delta t H\varphi^n \\ + \Delta t \sum_{a\in\alpha_d} \left(\partial_a^L \left[\frac{\Delta t}{\rho} \mathrm{diag}\left(A_a^R \varphi^n\right)^2 \right.\right. \\ \left.\left. + \mathrm{diag}\left(A_a^R M(\varphi^n)\right) \right] \partial_a^R \right) \left[\frac{-\lambda}{2}\Delta_{N^d}\varphi^n + f(\varphi^n) - \frac{1}{2}\mathrm{diag}\left(f'(\varphi^n)\right)\varphi^n \right] \\ - \Delta t \sum_{a\in\alpha_d} \partial_a^L \mathrm{diag}\left(A_a^R M_2(\varphi^n)\right) \partial_a^R \left[\frac{1}{2}\mathrm{diag}(G_B(\varphi^n))q^n \right], \end{aligned} \tag{4.200f}$$

$$\begin{aligned}\mathrm{RHS}_2 = q^n - \frac{\Delta t}{2} H q^n - \frac{\Delta t}{2}\mathrm{diag}\left(\tau_B(\varphi^n)\right)^{-1} q^n \\ + \Delta t\, \mathrm{diag}\left(G_B(\varphi^n)\right) \sum_{a\in\alpha_d} \partial_a^L \left(\mathrm{diag}\left(A_a^R M_2(\varphi^n)\right) \partial_a^R \left[\frac{-\lambda}{2}\Delta_{N^d}\varphi^n + f(\varphi^n) - \frac{1}{2}\mathrm{diag}\left(f'(\varphi^n)\right)\varphi^n \right] \right. \\ \left. - \mathrm{diag}\left(A_a^R \zeta(\varphi^n)\right)^{-1} \partial_a^R \left[\mathrm{diag}(G_B(\varphi^n)) \frac{1}{2} q^n \right] \right). \end{aligned} \tag{4.200g}$$

Next, we solve the completely discretized Chorin projection, which is analogous to algorithm (4.197) above, except for the following modifications to its right hand side in step one due to the additional $\boldsymbol{\sigma}$-coupling terms

$$\mathrm{RHS}_1 = u^n - \frac{\Delta t}{\rho}\mathrm{diag}\left(A^L_{N_x^3}\varphi^n\right)\partial^L_{N_x^3}\mu^{n+1/2} + \frac{\Delta t}{\rho}\left(\partial^L_{N_x^3}\sigma^n_{11} + A^L_{N_x^3}\partial_{N_y^3}\sigma^n_{12} + A^L_{N_x^3}\partial_{N_z^3}\sigma^n_{13}\right), \tag{4.201a}$$

$$\mathrm{RHS}_2 = v^n - \frac{\Delta t}{\rho}\mathrm{diag}\left(A^L_{N_y^3}\varphi^n\right)\partial^L_{N_y^3}\mu^{n+1/2} + \frac{\Delta t}{\rho}\left(A^L_{N_y^3}\partial_{N_x^3}\sigma^n_{21} + \partial^L_{N_y^3}\sigma^n_{22} + A^L_{N_y^3}\partial_{N_z^3}\sigma^n_{23}\right), \tag{4.201b}$$

$$\mathrm{RHS}_3 = w^n - \frac{\Delta t}{\rho}\mathrm{diag}\left(A^L_{N_z^3}\varphi^n\right)\partial^L_{N_z^3}\mu^{n+1/2} + \frac{\Delta t}{\rho}\left(A^L_{N_z^3}\partial_{N_x^3}\sigma^n_{31} + A^L_{N_z^3}\partial_{N_y^3}\sigma^n_{32} + \partial^L_{N_z^3}\sigma^n_{33}\right). \tag{4.201c}$$

Finally, we completely discretize the semi-discrete Oldroyd-B equations

$$\frac{\boldsymbol{\sigma}^{n+1} - \boldsymbol{\sigma}^n}{\Delta t} + (\boldsymbol{u}^{n+1}\cdot\nabla)\boldsymbol{\sigma}^n - (\nabla \boldsymbol{u}^{n+1})\boldsymbol{\sigma}^n - \boldsymbol{\sigma}^n\left(\nabla \boldsymbol{u}^{n+1}\right)^T + \frac{1}{\tau_S(\varphi^{n+\frac{1}{2}})}\boldsymbol{\sigma}^{n+\theta} - G_S(\varphi^{n+\frac{1}{2}})\, 2\boldsymbol{D}(\boldsymbol{u}^{n+1}) = \boldsymbol{0},$$

where $\theta \in \{0, 1\}$, reading

$$\begin{aligned}
&\mathrm{LHS}\sigma_{11}^{n+1} = \sigma_{11}^{n} - \Delta t H \sigma_{11}^{n} && (4.202a)\\
&+ 2\Delta t \Big[\mathrm{diag}\left(\partial^R_{N_x^3} u^{n+1}\right) \sigma_{11}^{n} + \mathrm{diag}\left(A^R_{N_x^3} \partial_{N_y^3} u^{n+1}\right) \sigma_{12}^{n} + \mathrm{diag}\left(A^R_{N_x^3} \partial_{N_z^3} u^{n+1}\right) \sigma_{13}^{n} \Big] \\
&- (1-\theta)\Delta t\, \mathrm{diag}\left(\tau_S(\varphi^{n+1/2})\right)^{-1} \sigma_{11}^{n} + \Delta t\, \mathrm{diag}\left(G_S(\varphi^{n+1/2})\right) 2\partial^R_{N_x^3} u^{n+1}\,, \\
&\mathrm{LHS}\sigma_{12}^{n+1} = \mathrm{LHS}\sigma_{21}^{n+1} = \sigma_{12}^{n} - \Delta t H \sigma_{12}^{n} && (4.202b)\\
&+ \Delta t \Big[\mathrm{diag}\left(\partial^R_{N_x^3} u^{n+1}\right) \sigma_{12}^{n} + \mathrm{diag}\left(A^R_{N_x^3} \partial_{N_y^3} u^{n+1}\right) \sigma_{22}^{n} + \mathrm{diag}\left(A^R_{N_x^3} \partial_{N_z^3} u^{n+1}\right) \sigma_{23}^{n} \\
&\quad + \mathrm{diag}\left(A^R_{N_y^3} \partial_{N_x^3} v^{n+1}\right) \sigma_{11}^{n} + \mathrm{diag}\left(\partial^R_{N_y^3} v^{n+1}\right) \sigma_{12}^{n} + \mathrm{diag}\left(A^R_{N_y^3} \partial_{N_z^3} v^{n+1}\right) \sigma_{13}^{n} \Big] \\
&- (1-\theta)\Delta t\, \mathrm{diag}\left(\tau_S(\varphi^{n+1/2})\right)^{-1} \sigma_{12}^{n} + \Delta t\, \mathrm{diag}\left(G_S(\varphi^{n+1/2})\right) \left(A^R_{N_x^3} \partial_{N_y^3} u^{n+1} + A^R_{N_y^3} \partial_{N_x^3} v^{n+1}\right), \\
&\mathrm{LHS}\sigma_{22}^{n+1} = \sigma_{22}^{n} - \Delta t H \sigma_{22}^{n} && (4.202c)\\
&+ 2\Delta t \Big[\mathrm{diag}\left(A^R_{N_y^3} \partial_{N_x^3} v^{n+1}\right) \sigma_{12}^{n} + \mathrm{diag}\left(\partial^R_{N_y^3} v^{n+1}\right) \sigma_{22}^{n} + \mathrm{diag}\left(A^R_{N_y^3} \partial_{N_z^3} v^{n+1}\right) \sigma_{23}^{n} \Big] \\
&- (1-\theta)\Delta t\, \mathrm{diag}\left(\tau_S(\varphi^{n+1/2})\right)^{-1} \sigma_{22}^{n} + \Delta t\, \mathrm{diag}\left(G_S(\varphi^{n+1/2})\right) 2\partial^R_{N_y^3} v^{n+1}\,, \\
&\mathrm{LHS}\sigma_{13}^{n+1} = \mathrm{LHS}\sigma_{31}^{n+1} = \sigma_{13}^{n} - \Delta t H \sigma_{13}^{n} && (4.202d)\\
&+ \Delta t \Big[\mathrm{diag}\left(\partial^R_{N_x^3} u^{n+1}\right) \sigma_{13}^{n} + \mathrm{diag}\left(A^R_{N_x^3} \partial_{N_y^3} u^{n+1}\right) \sigma_{23}^{n} + \mathrm{diag}\left(A^R_{N_x^3} \partial_{N_z^3} u^{n+1}\right) \sigma_{33}^{n} \\
&\quad + \mathrm{diag}\left(A^R_{N_z^3} \partial_{N_x^3} w^{n+1}\right) \sigma_{11}^{n} + \mathrm{diag}\left(A^R_{N_z^3} \partial_{N_y^3} w^{n+1}\right) \sigma_{12}^{n} + \mathrm{diag}\left(\partial^R_{N_z^3} w^{n+1}\right) \sigma_{13}^{n} \Big] \\
&- (1-\theta)\Delta t\, \mathrm{diag}\left(\tau_S(\varphi^{n+1/2})\right)^{-1} \sigma_{13}^{n} + \Delta t\, \mathrm{diag}\left(G_S(\varphi^{n+1/2})\right) \left(A^R_{N_x^3} \partial_{N_z^3} u^{n+1} + A^R_{N_z^3} \partial_{N_x^3} w^{n+1}\right), \\
&\mathrm{LHS}\sigma_{23}^{n+1} = \mathrm{LHS}\sigma_{32}^{n+1} = \sigma_{23}^{n} - \Delta t H \sigma_{23}^{n} && (4.202e)\\
&+ \Delta t \Big[\mathrm{diag}\left(A^R_{N_y^3} \partial_{N_x^3} v^{n+1}\right) \sigma_{13}^{n} + \mathrm{diag}\left(\partial^R_{N_y^3} v^{n+1}\right) \sigma_{23}^{n} + \mathrm{diag}\left(A^R_{N_y^3} \partial_{N_z^3} v^{n+1}\right) \sigma_{33}^{n} \\
&\quad + \mathrm{diag}\left(A^R_{N_z^3} \partial_{N_x^3} w^{n+1}\right) \sigma_{12}^{n} + \mathrm{diag}\left(A^R_{N_z^3} \partial_{N_y^3} w^{n+1}\right) \sigma_{22}^{n} + \mathrm{diag}\left(\partial^R_{N_z^3} w^{n+1}\right) \sigma_{23}^{n} \Big] \\
&- (1-\theta)\Delta t\, \mathrm{diag}\left(\tau_S(\varphi^{n+1/2})\right)^{-1} \sigma_{23}^{n} + \Delta t\, \mathrm{diag}\left(G_S(\varphi^{n+1/2})\right) \left(A^R_{N_y^3} \partial_{N_z^3} v^{n+1} + A^R_{N_z^3} \partial_{N_y^3} w^{n+1}\right), \\
&\mathrm{LHS}\sigma_{33}^{n+1} = \sigma_{33}^{n} - \Delta t H \sigma_{33}^{n} && (4.202f)\\
&+ 2\Delta t \Big[\mathrm{diag}\left(A^R_{N_z^3} \partial_{N_x^3} w^{n+1}\right) \sigma_{13}^{n} + \mathrm{diag}\left(A^R_{N_z^3} \partial_{N_y^3} w^{n+1}\right) \sigma_{23}^{n} + \mathrm{diag}\left(\partial^R_{N_z^3} w^{n+1}\right) \sigma_{33}^{n} \Big] \\
&- (1-\theta)\Delta t\, \mathrm{diag}\left(\tau_S(\varphi^{n+1/2})\right)^{-1} \sigma_{33}^{n} + \Delta t\, \mathrm{diag}\left(G_S(\varphi^{n+1/2})\right) 2\partial^R_{N_z^3} w^{n+1}\,,
\end{aligned}$$

where

$$\mathrm{LHS} = \left(\boldsymbol{I}_{N^3} + \theta \Delta t\, \mathrm{diag}\left(\tau_S(\varphi^{n+1/2})\right)^{-1} \right).$$

Thus, if $\theta \neq 0$, we have to solve a system of linear equations for each σ_{ij}, $i, j = 1, ..., d$, while, if $\theta = 0$, we can directly compute each.

Remark 4.57. *Note that we can reduce the spatial dimension of the above algorithm to 2D analogously to Remark 4.54, with additionally fixing*

$$\sigma_{13} = \sigma_{31} = \sigma_{23} = \sigma_{32} = \sigma_{33} = \mathbf{0}\,.$$

Further, note that for a constant viscosity η, we can decouple the calculation of the Navier-Stokes equations analogously to Remark 4.55.

Theorem 4.58. *Let H, H_u, H_v, H_w represent the flux difference upwind matrices (4.190) and (4.193). Then, the resulting completely discretized numerical scheme introduced above is linear and, assuming suitable boundary conditions, see Remark 3.13, it is mass-conservative. Further, assuming a constant viscosity η, the scheme can be significantly simplified, see Remark 4.55. This simplified scheme satisfies the following discrete energy law in 2D (considering non-vectorized discrete solutions) and an analogous version in 3D*

$$\begin{aligned}
&\frac{E_{total}(\varphi^{n+1}, q^{n+1}, \boldsymbol{\sigma}^{n+1}, \boldsymbol{u}^{n+1}) - E_{total}(\varphi^n, q^n, \boldsymbol{\sigma}^n, \boldsymbol{u}^n)}{\Delta t} = -\mathrm{ND}_{pot}^{n+1} - \mathrm{ND}_{split}^{n+1} - \mathrm{ND}_{FV}^{n+1} \\
&- \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \frac{1}{\tau_B(\varphi_{ij}^n)} (q_{ij}^{n+\frac{1}{2}})^2 - \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \frac{1}{\zeta(\varphi_{ij}^n)} \left| \varphi_{ij}^n (1-\varphi_{ij}^n) \nabla_h^R \mu_{ij}^{n+\frac{1}{2}} - \nabla_h^R (G_B(\varphi_{ij}^n)\, q_{ij}^{n+\frac{1}{2}}) \right|^2 \\
&- \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \frac{1}{\tau_S(\varphi_{ij}^{n+\frac{1}{2}})} \left((\sigma_{11}^{n+\theta})_{ij} + (\sigma_{22}^{n+\theta})_{ij} \right) - \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \eta \left(|\nabla_h^R u_{i-1/2,j}^{n+1}|^2 + |\nabla_h^R v_{i,j-1/2}^{n+1}|^2 \right),
\end{aligned} \tag{4.203}$$

where

$$\mathrm{ND}_{pot}^{n+1} = -\frac{1}{6}(\Delta t)^2 \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} f''(\zeta) \left(\partial_t \varphi_{ij}^{n+\frac{1}{2}} \right)^3 \in \mathcal{O}\left((\Delta t)^2\right),$$

$$\begin{aligned}
\mathrm{ND}_{split}^{n+1} = \frac{\rho}{2\,\Delta t} \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \Big(&(u_{i-1/2,j}^{n+1} - u_{i-1/2,j}^*)^2 + (u_{i-1/2,j}^* - u_{i-1/2,j}^n)^2 \\
&+ (v_{i,j-1/2}^{n+1} - v_{i,j-1/2}^*)^2 + (v_{i,j-1/2}^* - v_{i,j-1/2}^n)^2 \Big) \geq 0,
\end{aligned}$$

$$u_{i-1/2,j}^* = u_{i-1/2,j}^n - \frac{\Delta t}{\rho} \varphi_{i-1/2,j}^n \partial_x^L \mu_{ij}^{n+\frac{1}{2}}, \qquad v_{i,j-1/2}^* = v_{i,j-1/2}^n - \frac{\Delta t}{\rho} \varphi_{i,j-1/2}^n \partial_y^L \mu_{ij}^{n+\frac{1}{2}},$$

$$\begin{aligned}
\mathrm{ND}_{FV}^{n+1} = \frac{1}{2} \sum_{\substack{i=1,\dots,N_x \\ j=1,\dots,N_y}} \Bigg(&h_x \Big(|u_{i+1/2,j}^n| \left[(\partial_x^R \varphi_{ij}^n)^2 + (\partial_x^R q_{ij}^{n+1/2})^2 + (\partial_x^R (\sigma_{11}^{n+\theta})_{ij})^2 + (\partial_x^R (\sigma_{22}^{n+\theta})_{ij})^2 \right] \\
&+ |u_{ij}^n| (\partial_x^R u_{i-1/2,j}^\dagger)^2 + |u_{i+1/2,j-1/2}^n| (\partial_x^R v_{i,j-1/2}^\dagger)^2 \Big) \\
&+ h_y \Big(|v_{i,j+1/2}^n| \left[(\partial_y^R \varphi_{ij}^n)^2 + (\partial_y^R q_{ij}^{n+1/2})^2 + (\partial_y^R (\sigma_{11}^{n+\theta})_{ij})^2 + (\partial_y^R (\sigma_{22}^{n+\theta})_{ij})^2 \right] \\
&+ |v_{i-1/2,j+1/2}^n| (\partial_y^R u_{i-1/2,j}^\dagger)^2 + |v_{ij}^n| (\partial_y^R v_{i,j-1/2}^\dagger)^2 \Big) \Bigg) \geq 0.
\end{aligned}$$

This energy law is – up to the numerical diffusion caused by the finite volume method – analogous to its semi-discrete version (4.80).

Proof. It is clear that the proposed scheme is linear.

In order to calculate the discrete mass and energy law, we consider a cell-wise 2D version of the equations involved in the algorithm for non-vectorized solutions, analogously to discretizations (4.131) for the simplified model and (4.194) for the Cahn-Hilliard-Navier-Stokes model.

Then, using SBP formulas (4.125) and (4.186), the mass conservation is calculated analogously to the proof of Theorem 4.53 for the latter model.

Further, we multiply and sum up the cell-wise equations analogously to the calculations of energy laws (4.132) and (4.195), which are satisfied by the two above-mentioned models, while the cell-wise σ-equations are only summed up. Then, using the SBP formulas, the discrete energy law is calculated analogously to the semi-discrete energy law (4.80).

For three dimensions, the proof is analogous, since the SBP formulas (4.125) and (4.186) hold for any space dimension. □

4.9 Time step restrictions

Time-stepping is a crucial topic in numerics, since large time steps generally save precious computing time in numerical experiments. The key to a reasonable time step size is to ensure that it is as large as possible while ensuring, that the schemes are numerically stable and that no essential information is lost during the process. In this context, there are two main differentiations, fixed time step sizes and adaptive ones, where the latter may change in each time step. A general approach to adaptive time-stepping is the so called *CFL condition*, which is introduced in the following.

CFL condition

The CFL condition is named after *Courant, Friedrichs and Lewy*, see [20], and states that a numerical method can only be stable, if its numerical area of dependency includes the physical dependency area of the partial differential equation, at least in the limit $h \to 0$.

To check this condition, there is the so-called *C*ourant number ν. Its boundedness is a necessary condition for the stability of a numerical method.

Since implicit time discretizations have no time step restriction, we only have to pay attention to our explicitly discretized spatial derivatives. We discretize the advection and convection terms of our phase-field models with hydrodynamics at least partially explicitly. This terms are analogous to the advection term of the 2D advection equation

$$\rho_t + \nabla \cdot (\boldsymbol{u}\rho) = 0\,.$$

Applying the flux difference upwind method to above yields (4.177), reading

$$\begin{aligned}\rho_{ij}^{n+1} = \rho_{ij}^{n} &- \frac{\Delta t}{h_x}\left(u^{+}_{i+1/2,j}\rho_{ij} + u^{-}_{i+1/2,j}\rho_{i+1,j} - u^{+}_{i-1/2,j}\rho_{i-1,j} - u^{-}_{i-1/2,j}\rho_{ij}\right)\\ &- \frac{\Delta t}{h_y}\left(v^{+}_{i,j+1/2}\rho_{ij} + v^{-}_{i,j+1/2}\rho_{i,j+1} - v^{+}_{i,j-1/2}\rho_{i,j-1} - v^{-}_{i,j-1/2}\rho_{ij}\right)\,.\end{aligned}$$

The CFL condition for this discretization of the 2D advection equation is given by

$$\nu := \frac{|u|_\infty \Delta t}{h_x} + \frac{|v|_\infty \Delta t}{h_y} \leq \nu_{max} \,,$$

where $|u|_\infty = \max_{ij}(u_{i+1/2,j})$ and $|v|_\infty = \max_{ij}(v_{i,j+1/2})$. Further, $\nu_{max} = 1$, because the flux functions of the applied finite volume method consist of function values of adjacent cells only. The CFL condition leads to the following upper bound for the time step size

$$\frac{|u|_\infty \Delta t}{h_x} + \frac{|v|_\infty \Delta t}{h_y} \leq 1 \qquad \Leftrightarrow \qquad \Delta t \leq \frac{h_x h_y}{h_y |u|_\infty + h_x |v|_\infty} \,.$$

Note that in all our phase-field models with flow, the velocity field $\boldsymbol{u}$ is not only space but also time dependent. Consequently, the upper bound of the time step size changes in time, reading

$$\Delta t_{n+1} \leq \frac{h_x h_y}{h_y |u^n|_\infty + h_x |v^n|_\infty} \,.$$

Analogously, the upper bound of the time step size in 3D reads

$$\Delta t_{n+1} \leq \frac{h_x h_y h_z}{h_y h_z |u^n|_\infty + h_x h_z |v^n|_\infty + h_x h_y |w^n|_\infty} \,,$$

where $|u|_\infty = \max_{ijk}(u_{i+1/2,j,k})$, $|v|_\infty = \max_{ijk}(v_{i,j+1/2,k})$ and $|w|_\infty = \max_{ijk}(w_{i,j,k+1/2})$.

Note that we also have at least one explicitly calculated spatial second derivative in the Cahn-Hilliard equation due to the linearization of the nonlinear derivative $f(\varphi)$ of the double-well potential. This significantly complicates the construction of an accurate CFL condition, if not makes it impossible.

Further, note that even though an accurate CFL condition can yield an upper bound for the time step size, the latter may still need to be significantly smaller in order to have energy-stable simulations. In particular, using the OD2 approximation (4.26) for the potential derivative, we have a dissipation term of order $\mathcal{O}\left((\Delta t)^2\right)$. Thus, the impact of this dissipation term to the discrete energy law can be reduced by shrinking the time step size.

Summarized, we will use above CFL condition as an upper limit for our numerical experiments, but we will investigate further time step limitations of our numerical schemes experimentally, see Subsection 5.1.6.

Problem specific time-stepping

In general, long time dynamics of Cahn-Hilliard models tends to become slow, where dissipation terms like the above-mentioned one usually become small. Therefore, several methods to enlarge the time step size during this late time frame have been considered, see, e.g., *Cheng, Kurganov, Qu, Tang* [16], or *Guillén-González and Tierra* [39].

Since our focus is the full model for viscoelastic phase separation, which consists of far more than only the Cahn-Hilliard equation, a direct applicability of such methods is questionable.

Nevertheless, let us introduce one of the above-mentioned methods exemplarily. The adaptive time-stepping strategy for the Cahn-Hilliard equation of *Cheng, Kurganov et al.* [16] reads

$$\Delta t = \max\left(\Delta t_{min}, \frac{\Delta t_{max}}{\sqrt{1+\alpha|m'(t)|^2}}\right), \qquad \alpha = \text{const.}, \tag{4.204}$$

where

$$\begin{aligned} m(t) =& \sqrt{\frac{1}{|\Omega|}\int_\Omega (\varphi - \bar{\varphi})^2 \, d\mathbf{x}}, \\ \bar{\varphi} =& \frac{1}{|\Omega|}\int_\Omega \varphi d\mathbf{x}. \end{aligned}$$

The effectivity of this strategy has been proven in [16] for the Cahn-Hilliard equation using the Ginzburg-Landau potential and a non-degenerate mobility.

5 Numerical experiments

In this chapter, we demonstrate the behavior of the proposed fully discrete numerical schemes in two and three space dimensions. For this aim, we use our self-implemented universal MATLAB Code, which includes several vectorized full discretizations for the numerical solution of the full model as well as all sub-models, using sparse system matrices. Since MATLAB is a scripting language, one may question the efficiency of our implementation. Let us note in this context that MATLAB uses the to this date most advanced multithreaded linear algebra library *Intel Math Kernel Library* (MKL) for all relevant operations, including the linear equation system solvers. Thereby, we apply lower-upper (LU) decompositions to factorize small system matrices, whereas we use the biconjugate gradient stabilized (BiCGSTAB) method for large non-symmetric system matrices. In particular, we also use the preconditioned conjugate gradient (PCG) method for large symmetric system matrices. Such matrices appear in the full discretizations of the Cahn-Hilliard equation and the simplified model. Further, similar structures arise in the Poisson problem generated by the Chorin pressure correction algorithm.

For most numerical experiments, we consider two space dimensions, where we present the experimental order of convergence of important schemes introduced in Chapter 4 to demonstrate the accuracy of these schemes and we show some discretization peculiarities. Further, we perform and explain viscoelastic phase separation, an experimental sensitivity analysis in the form of parameter studies as well as comparisons of our macroscopic model simulations to mesoscopic coarse-grained molecular dynamics simulation results provided through the collaboration with our project partners from the Max Planck Institute for Polymer Research (MPIP).

We also consider three space dimensions, originally because the mesoscopic simulations include coarse-grained molecular dynamics of polymer chains, whose interaction is usually extremely limited in 2D, because they immediately entangle when they are close to each other.

Note that all beforehand introduced models are non-dimensionalized. Because of that, there is only one length scale in our problems, which is the domain size. Thus, all parameters as well as the time evolution scale with the domain size. Further, note that all computations are performed on an Intel Xeon E5-2650 v3 10-core CPU using MATLAB version 2017a.

5.1 Experimental convergence

In order to investigate the convergence of our schemes, we compute the *experimental order of convergence* (EOC) in time and/or space of some key discretizations as follows.

We compute several numerical solutions $\omega_{N,\Delta t}$ at the same final time t, using different uniform grids with $N \times N$ cells and different constant time step sizes Δt. The numerical solution computed on the finest resolution is used as the reference solution ω_{ref}. The other solutions have a consecutively doubled time step size (i.e., halved grid size in time) for the EOC in time and a consecutively halved grid size in each space dimension for the EOC in space. For a combined EOC in time and space, we halve both grid sizes consecutively. Next, we compute the L^1-error of the numerical solutions by

$$e(\omega_{N,\Delta t}) = \|\omega_{N,\Delta t} - \tilde{\omega}_{ref}\|_{L^1(\Omega)} ,$$

where $\tilde{\omega}_{ref}$ is mapped from its original grid to the respective $N \times N$ grid of $\omega_{N,\Delta t}$ in case of different spatial grids. This mapping is performed by averaging over the function values of the closest adjacent grid cells of the finer grid. In two space dimensions, this are four grid cells for all variables except the velocity field and two for each component of the velocity field, since the latter are defined on staggered grids. Note that for the vector- and matrix-valued functions $\boldsymbol{u}$ and σ, we calculate the error component-wise and sum up the errors of all components afterwards. Finally, the EOC in time is computed by

$$\mathrm{EOC}(\omega_{N,\Delta t}) = \log_2 \left(e(\omega_{N,2\Delta t}) / e(\omega_{N,\Delta t}) \right) ,$$

the EOC in space by

$$\mathrm{EOC}(\omega_{N,\Delta t}) = \log_2 \left(e(\omega_{N/2,\Delta t}) / e(\omega_{N,\Delta t}) \right) ,$$

and the EOC in time and space by

$$\mathrm{EOC}(\omega_{N,\Delta t}) = \log_2 \left(e(\omega_{N/2,2\Delta t}) / e(\omega_{N,\Delta t}) \right) .$$

We consider the following models and test cases.

5.1.1 Cahn-Hilliard equation (EOC in space and time)

We solve the full discretization (4.130) of the Cahn-Hilliard equation with the following functions and parameters:

- $f(\varphi)$ is the derivative of the Ginzburg-Landau potential (3.9),
- the interface constant reads $\lambda = 0.01$, and
- the constant mobility $M(\varphi) = 1$.

We use the following smooth initial conditions

$$\varphi^0 := \varphi(x, y, 0) = 0.05 \sin x \sin y + 0.001 , \qquad (x, y) \in [0, 2\pi] \times [0, 2\pi] ,$$

and compute the numerical solutions $\varphi_{N,\Delta t}$ up to the final time $t = 5$. Our reference solution reads $\varphi_{ref} = \varphi_{512,10^{-3}/4}$, i.e., it is calculated on a 512×512 grid with time step size $\Delta t = 10^{-3}/4$. The results displayed in Table 5.1 confirm our claimed second order of accuracy in time and space.

Table 5.1: **Cahn-Hilliard equation:** L^1-errors and experimental convergence rates in space and time at $t = 5$.

N	Δt	$e(\varphi_{N,\Delta t})$	$\mathrm{EOC}(\varphi_{N,\Delta t})$
64	$2 \cdot 10^{-3}$	0.26289	–
128	10^{-3}	0.064606	2.0247
256	$10^{-3}/2$	0.017824	1.8578

5.1.2 Cahn-Hilliard-Navier-Stokes model (EOC in space and time)

We solve the fully discretized algorithm (4.196), (4.197) for the Cahn-Hilliard-Navier-Stokes model with the modifications from Remark 4.54 for 2D and from Remark 4.55 for a constant viscosity, and the following functions and parameters:

- $f(\varphi)$ is the derivative of the Ginzburg-Landau potential (3.9),
- the interface constant reads $\lambda = 0.001$,
- the constant mobility $M(\varphi) = 1$, and
- the constant viscosity $\eta(\varphi) = 0.01$.

We use the following initial conditions

$$\begin{aligned} \varphi^0 &= 0.5 \sin(4\pi x) \sin(2\pi y)\,, \\ u^0 &= -0.25 \sin(\pi x)^2 \sin(2\pi y)\,, \\ v^0 &= 0.25 \sin(2\pi x) \sin(\pi y)^2\,, \qquad (x, y) \in [0,1] \times [0,1]\,, \end{aligned}$$

where (u^0, v^0) is divergence free. Then, we compute up to the final time $t = 1$ and use the numerical solution calculated on a staggered 512×512 grid with time step size $\Delta t = 10^{-3}/4$ as reference solution. The results displayed in Table 5.1 confirm that the scheme converges experimentally, but as expected only with around first order of accuracy in space and time.

Table 5.2: **Cahn-Hilliard-Navier-Stokes model:** L^1-errors and experimental convergence rates in space and time at $t = 1$.

N	Δt	$e(\varphi_{N,\Delta t})$	$\mathrm{EOC}(\varphi_{N,\Delta t})$	$e(\boldsymbol{u}_{N,\Delta t})$	$\mathrm{EOC}(\boldsymbol{u}_{N,\Delta t})$
32	$4 \cdot 10^{-3}$	$1.8236 \cdot 10^{-2}$	–	$4.3149 \cdot 10^{-3}$	–
64	$2 \cdot 10^{-3}$	$5.8273 \cdot 10^{-3}$	1.6459	$2.1858 \cdot 10^{-3}$	0.98117
128	10^{-3}	$2.4891 \cdot 10^{-3}$	1.2272	$9.8468 \cdot 10^{-4}$	1.1504
256	$10^{-3}/2$	$8.4118 \cdot 10^{-4}$	1.5651	$3.3807 \cdot 10^{-4}$	1.5423

5.1.3 Simplified model (EOC in space and time)

We solve the full discretization (4.133) of the simplified model with the following functions and parameters, where all functions of φ are calculated using the second order extrapolation $\varphi^{n-1/2}$:

- $f(\varphi)$ is the derivative of the Flory-Huggins potential (3.11) with $n_p = n_s = 1$ and $\chi = 2.8/1.1$,
- the interface constant reads $\lambda = 1$,
- the constant friction $\zeta(\varphi) = 0.1$,
- the relaxation time $\tau_B(\varphi) = \tau_B^0 \varphi^2$ with $\tau_B^0 = 10$, and
- the bulk relaxation modulus

$$G_B(\varphi) = G_B^0 \left[1 + \tanh\left(\frac{\cot(\pi\varphi^*) - \cot(\pi\varphi)}{\varepsilon} \right) \right] + G_B^1 ,$$

  where $G_B^0 = 0.5$, $G_B^1 = 0$, $\varphi^* = 0.4$ and $\varepsilon = 0.01$.

We use the following initial conditions

$$\varphi^0 = 0.4 + 0.05 \sin(8\pi x/128) \sin(4\pi y/128) , \qquad q^0 = 0 , \qquad (x, y) \in [0, 128] \times [0, 128] ,$$

compute up to the final time $t = 500$ and use the numerical solution calculated on a 1024×1024 grid with time step size $\Delta t = 0.0125$ as reference solution. The results displayed in Table 5.3 confirm our claimed second order of accuracy in time and space. Note that the L^1-errors are substantially larger than in Tables 5.1 and 5.2, since the error scales with domain size, which is significantly larger in this experiment.

Table 5.3: **Simplified model:** L^1-errors and experimental convergence rates in space and time at $t = 500$.

N	Δt	$e(\varphi_{N,\Delta t})$	$\mathrm{EOC}(\varphi_{N,\Delta t})$	$e(q_{N,\Delta t})$	$\mathrm{EOC}(q_{N,\Delta t})$
128	0.1	101.61	–	7.5542	–
256	0.05	25.553	1.9915	1.9439	1.9583
512	0.025	5.0666	2.3344	0.47251	2.0406

5.1.4 Simplified model (EOC in time)

Note that each time step size reduction also necessitates a grid cell size reduction for calculating the experimental convergence in space and time. Therefore, we can and will run out of random access memory (RAM) if we continue to decrease the step and cell sizes. By neglecting the spatial convergence, we are not limited by memory capacity anymore such that we can compute the experimental convergence using significantly more time step sizes and thereby get a more accurate order of accuracy in time.

Thus, let us also calculate the experimental convergence of the simplified model purely in time, using the following initial conditions

$$\varphi^0 = 0.4 + \delta(x, y) , \qquad q^0 = 0 , \qquad (x, y) \in [0, 128] \times [0, 128] ,$$

where $\delta(x, y)$ is a uniformly distributed random variable with range $[-0.001, 0.001]$, whose values are the same for each experiment. We compute up to the final time $t = 2000$ on a 128×128 grid and use the numerical solution calculated with time step size $\Delta t = 2^{-9}$ as reference solution. Everything else is identical to the EOC in space and time experiment 5.1.3 above. The more extensive results displayed in Table 5.3 again confirm our claimed second order of accuracy in time.

Table 5.4: **Simplified model:** L^1-errors and experimental convergence rates in time at $t = 2000$.

Δt	$e(\varphi_{\Delta t})$	$\mathrm{EOC}(\varphi_{\Delta t})$	$e(q_{\Delta t})$	$\mathrm{EOC}(q_{\Delta t})$
2^{-3}	167.12	–	2.6173	–
2^{-4}	44.203	1.9186	0.82712	1.6619
2^{-5}	15.989	1.4671	0.26185	1.6594
2^{-6}	5.0728	1.6562	0.087266	1.5852
2^{-7}	1.41	1.847	0.025212	1.7913
2^{-8}	0.30756	2.1968	0.006347	1.9899

Note that we have shown the experimental convergence in space and time of sub-models of the full model above. This sub-models already include all kinds of spatial discretizations which we use in the fully discretized full model. Therefore, and in order to limit our RAM usage, we only show the experimental convergence in time of the full model in what follows.

5.1.5 Full model (EOC in time)

We solve the fully discretized algorithm for the full model from Subsection 4.8.14 with the modifications from Remark 4.57 for 2D and a constant viscosity. Further, we use the functions and parameters from the simplified model above, complemented by the following:

- the shear relaxation time $\tau_S(\varphi) = \tau_S^0 \varphi^2$ with $\tau_S^0 = 5$,
- the shear relaxation modulus $G_S(\varphi) = G_S^0 \varphi^2$ with $G_S^0 = 0.5$, and
- the constant viscosity $\eta(\varphi) = 1$.

We consider the following initial conditions

$$\varphi^0 = 0.4 + \delta(x, y), \qquad q^0 = 0, \qquad \boldsymbol{\sigma}^0 = \mathbf{0}, \qquad \boldsymbol{u}^0 = \mathbf{0}, \qquad (x, y) \in [0, 128] \times [0, 128],$$

where $\delta(x, y)$ is again a uniformly distributed random variable with range $[-0.001, 0.001]$. We compute up to the final time $t = 1000$ on a staggered 128×128 grid and use the numerical solution calculated with time step size $\Delta t = 2^{-9}$ as reference solution. The results displayed in Table 5.5 confirm that the scheme converges experimentally, but as expected only with around first order of accuracy in time.

Table 5.5: **Full model:** L^1-errors and experimental convergence rates in time at $t = 1000$.

Δt	$e(\varphi_{\Delta t})$	EOC$(\varphi_{\Delta t})$	$e(q_{\Delta t})$	EOC$(q_{\Delta t})$	$e(\boldsymbol{\sigma}_{\Delta t})$	EOC$(\boldsymbol{\sigma}_{\Delta t})$	$e(\boldsymbol{u}_{\Delta t})$	EOC$(\boldsymbol{u}_{\Delta t})$
2^{-5}	1762.5	–	32.051	–	0.2284	–	1.3417	–
2^{-6}	624.96	1.4958	17.418	0.87978	0.10437	1.1299	0.65297	1.0389
2^{-7}	215.35	1.5371	8.2838	1.0722	0.035823	1.5428	0.1487	2.1346
2^{-8}	75.327	1.5154	3.4345	1.2702	0.014445	1.3104	0.05623	1.403

5.1.6 Discretization peculiarities

Expanding the space and time step choices from the experimental convergence experiments above to coarser grids demonstrates some peculiarities of our discretizations. In Figure 5.1, we see phase-fields, energy evolutions and L^1-errors at $t = 20$ of the Cahn-Hilliard experimental convergence experiment from Subsection 5.1.1, but purely in space, using different grid sizes and $\Delta t = 10^{-3}/4$. We can clearly observe that the phase-field and energy dynamics differ substantially for coarse grids. For grid sizes 16×16 and 32×32, the numerical solutions converge towards different steady states than for finer grids and have different energy evolutions. Consequently, the relative errors to the reference solution $\varphi_{ref} = \varphi_{512,10^{-3}/4}$ are very high. The reason being that the large grid cells are not able to capture the slim interface properly, which separates the two phases. But also the quadratic shape of the grid cells is a contributing factor using such coarse grids. Thus, we only use large grid sizes $N \geq 64$ for the EOC in space and time of the Cahn-Hilliard equation, see Table 5.1.

In Figure 5.2, we see phase-fields, energy evolutions and L^1-errors at $t = 2000$ of the simplified model experimental convergence in time experiment from Subsection 5.1.4, using different time step sizes and a 128×128 grid. Down to time step size $\Delta t = 1/8$, the numerical solutions of volume fraction φ at $t = 2000$ deviate noticeably, while further down to $\Delta t = 1/32$, deviations become much smaller. There are no further deviations visible for even smaller time step sizes. Therefore, we only use time step sizes $\Delta t \leq 1/8$ to calculate the EOC in time, see Table 5.4.

In Figure 5.3, we see phase-fields, energy evolutions and L^1-errors at $t = 1000$ of the full model experimental convergence in time experiment from Subsection 5.1.5, using different time step sizes and a 128×128 grid. Down to time step size $\Delta t = 1/32$, the numerical solutions of snapshots of volume fraction φ at $t = 2000$ deviate noticeably, while further down to $\Delta t = 1/64$, deviations become much smaller. There are no further deviations visible for even smaller time step sizes. The time evolution of the total energy also differs noticeably down to $\Delta t = 1/32$ and only minimally for even smaller time step sizes. Additionally, the error plot shows that the experimental convergence rate is low for time step sizes above $\Delta t = 1/32$. This implies that our full discretization is not well suited for such coarse time steps. Therefore, we only use time step sizes $\Delta t \leq 1/32$ to calculate the EOC in time, see Table 5.5.

Further, let us investigate the effects of differing scheme variants, which we introduced in Section 4.8. For this purpose, we compare phase-fields as well as errors of conservation of mass and momentum, and evolutions of total energies. Once, for our two different finite difference approaches, see Figures 5.4, 5.5 and 5.6, and once, for three different finite volume methods, see Figures 5.7, 5.8 and 5.9. These experiments are based on the full model experimental

convergence experiment from Subsection 5.1.5, using a 128×128 grid and time step size $\Delta t = 0.025$

In Figure 5.4, especially at time $t = 200$, we observe that the phase separation process is accelerated by the additional artificial diffusion introduced by central differences, compared to the left and right differences. The accelerated separation process is also confirmed by the earlier dissipation of the total energy when using central differences, see Figure 5.6. Since this effect is purely caused by the discretization and not by physics, it is not desirable. Thus, we use the left and right differences for our simulations in general. In Figure 5.5, we see a negligible deviation from the conservation of mass for both methods, given by the discrete version of

$$\int_\Omega \left(\varphi^n - \varphi^0\right) \qquad \forall n \in \mathbb{N}\,.$$

This verifies our proven conservation on the discrete level. The deviation from the conservation of momentum on the other hand is considerable in both cases. Even though one should keep in mind that the deviations, which are not normed, scale with domain size and that is $[0, 128]^2$. The deviation from the conservation of momentum is calculated by the discrete version of

$$\int_\Omega \left(u^n - u^0 + v^n - v^0\right) \qquad \forall n \in \mathbb{N}\,.$$

In Figure 5.7, we only compare the phase-fields at time $t = 1000$, since the three different finite volume methods cause no visible variation of the phase-field evolution in this experiment. This is primarily caused by the fact that the initial values for the velocity field are zero here. The error and total energy plots, see Figures 5.8 and 5.9, show no differences between the upwind and the DCU method, which is in line with their identical derivation. Both show the same negligible deviation from the conservation of mass and the same noticeable deviation from the conservation of momentum over time. However, while the second order DCU method causes only small variations in the momentum error and the evolution of the total energy, it introduces a significant mass conservation error. This is in line with our claim that this method is not necessarily conservative, see Subsection 4.8.9. For this reason, we use the first order upwind method for our simulations in general.

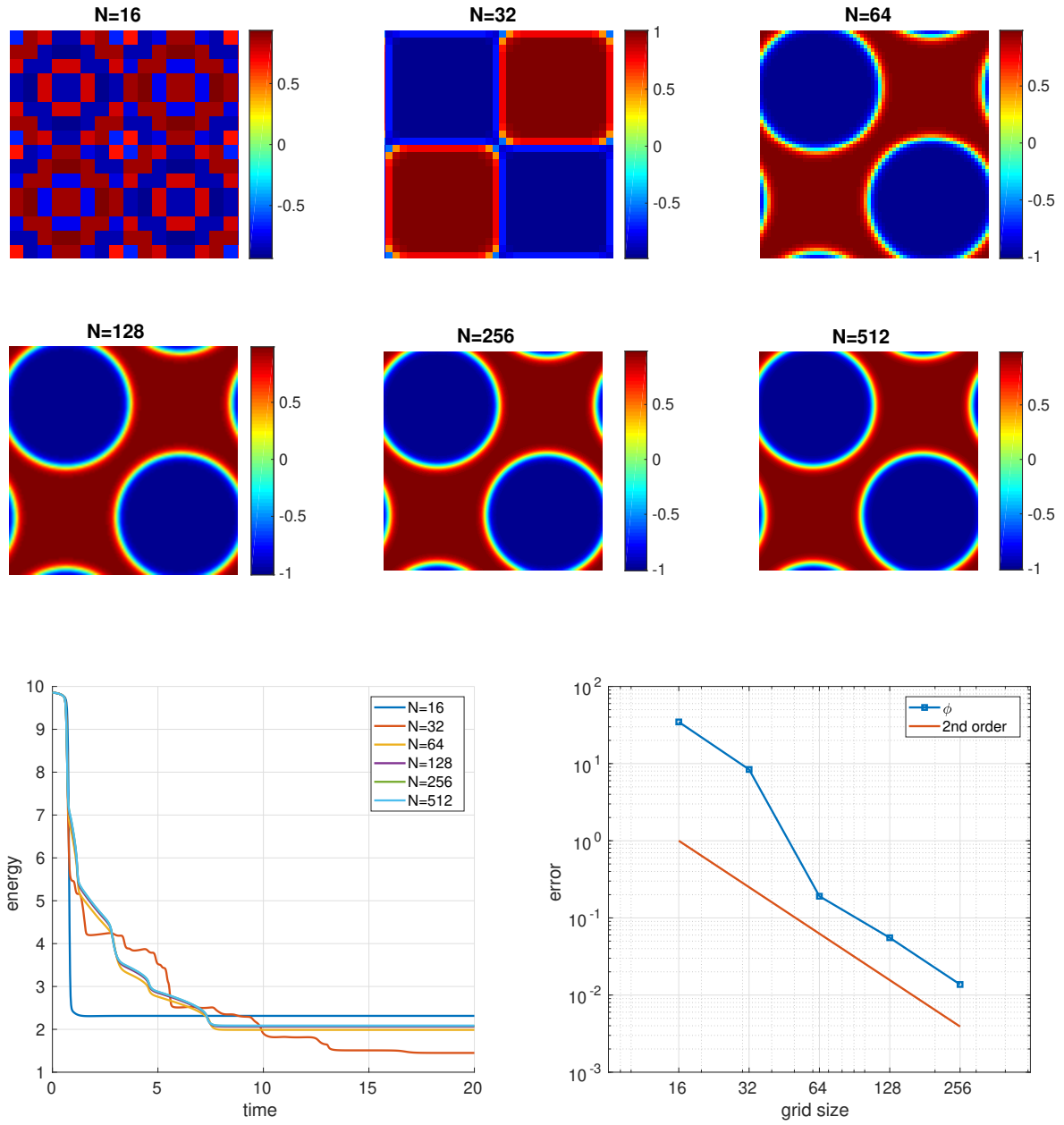
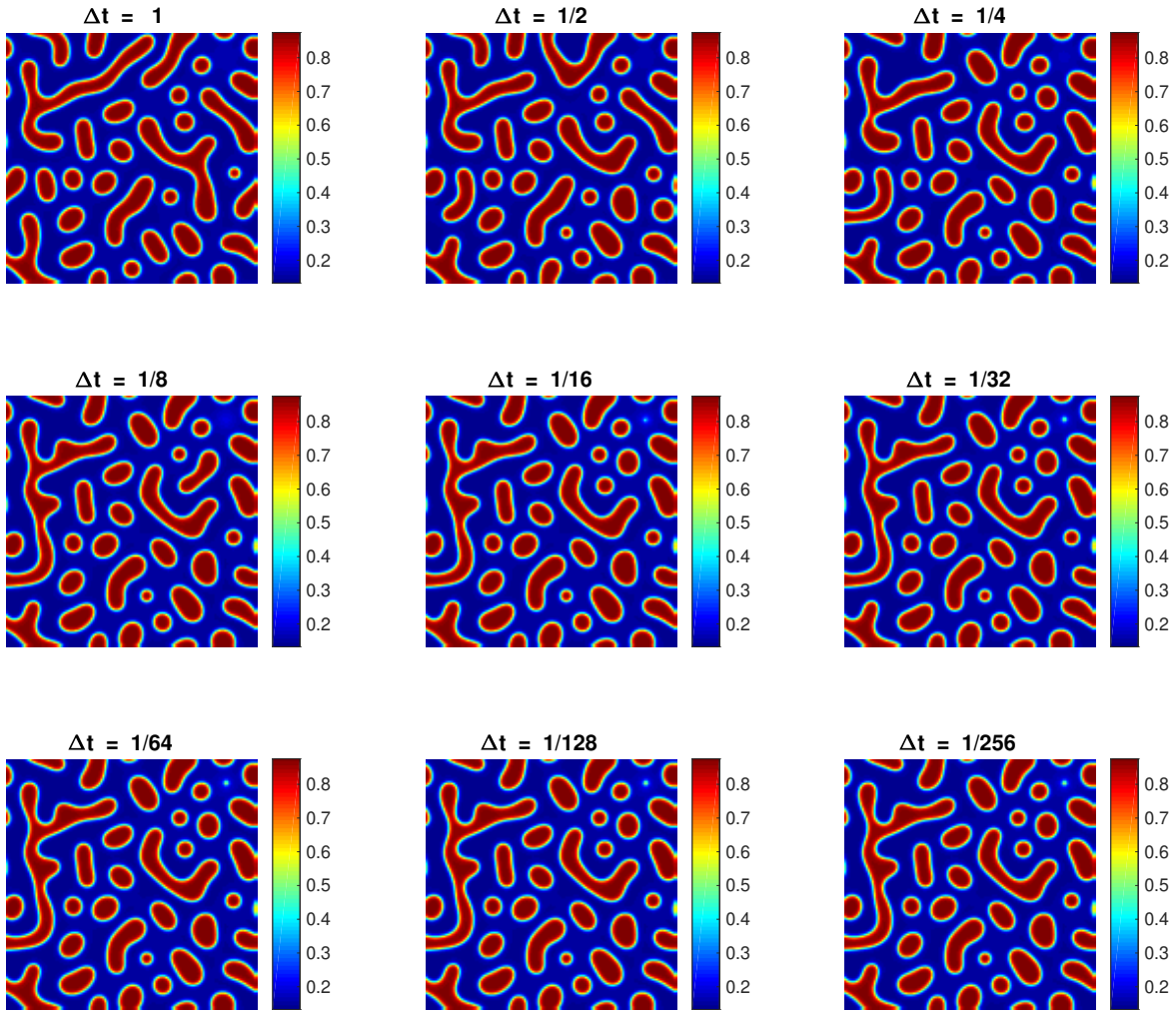


Figure 5.1: **Cahn-Hilliard equation** (EOC in space): snapshots of volume fraction φ at $t = 20$, evolutions of the total energies E_{total} and the L^1-errors, using different grid sizes $N \times N$ and $\Delta t = 10^{-3}/4$.



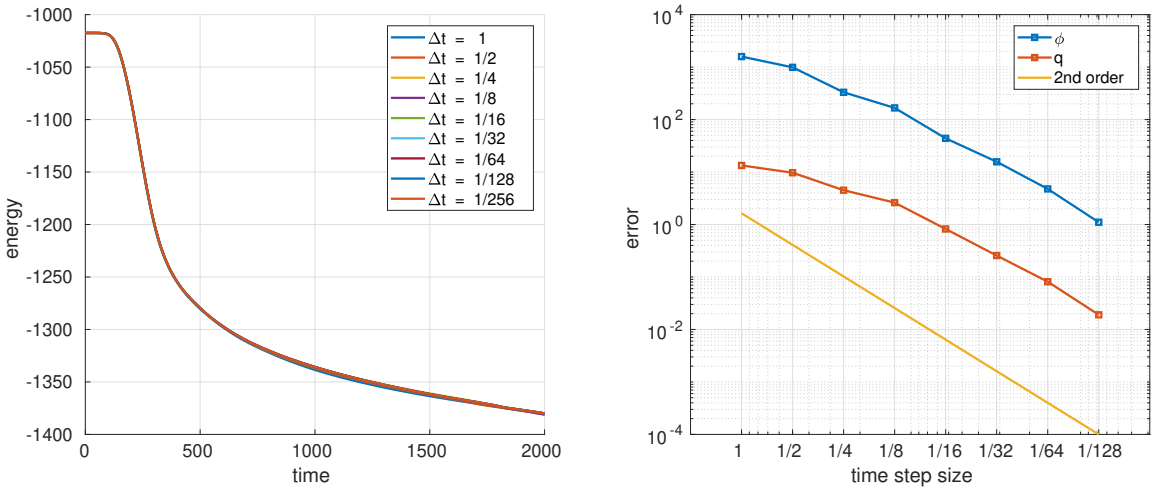
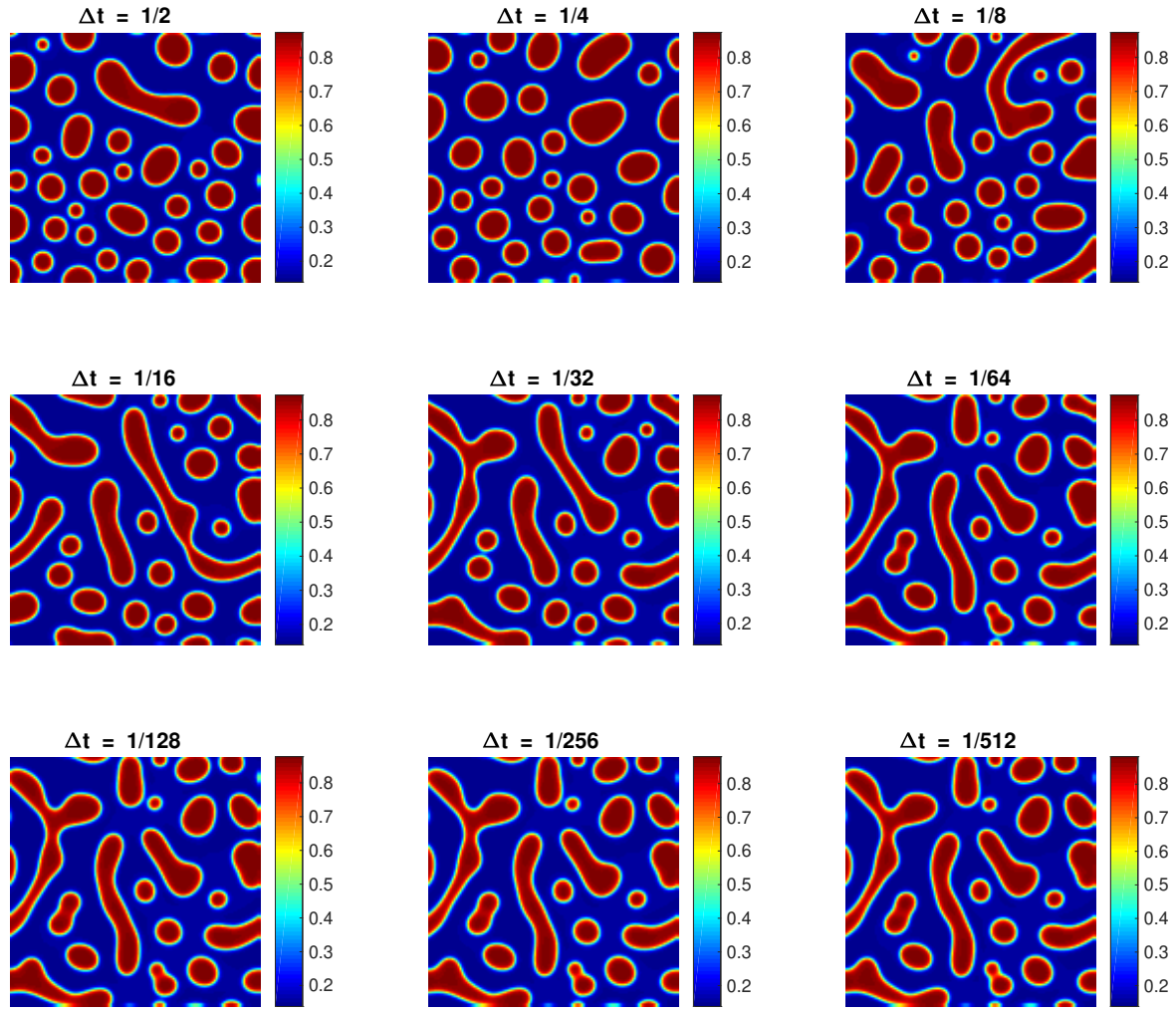


Figure 5.2: **Simplified model** (EOC in time): snapshots of volume fraction φ at $t = 2000$, evolutions of the total energies E_{total} and the L^1-errors, using different time step sizes Δt and a 128×128 grid.



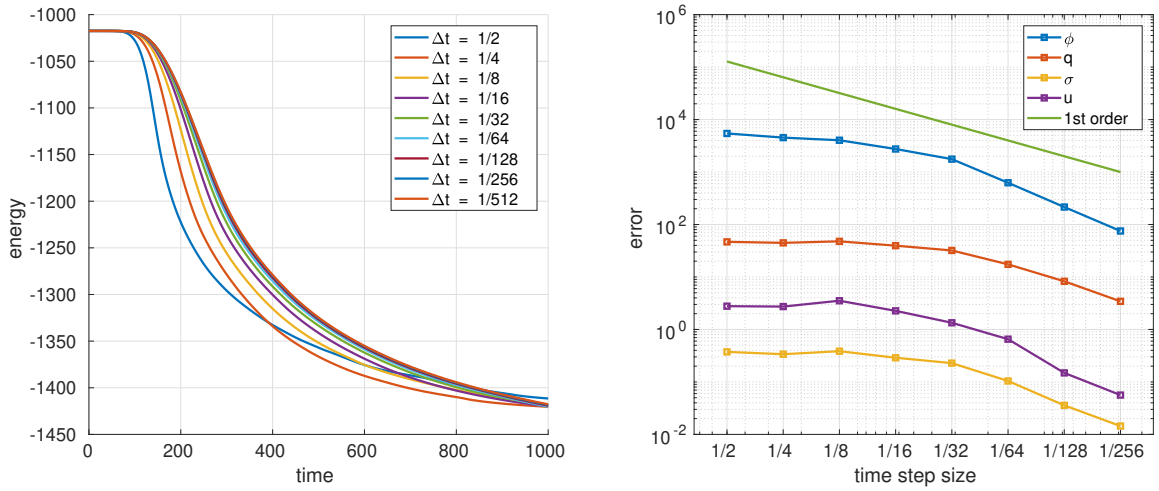


Figure 5.3: **Full model** (EOC in time): snapshots of volume fraction φ at $t = 1000$, evolutions of the total energies E_{total} and the L^1-errors, using different time step sizes Δt and a 128×128 grid.

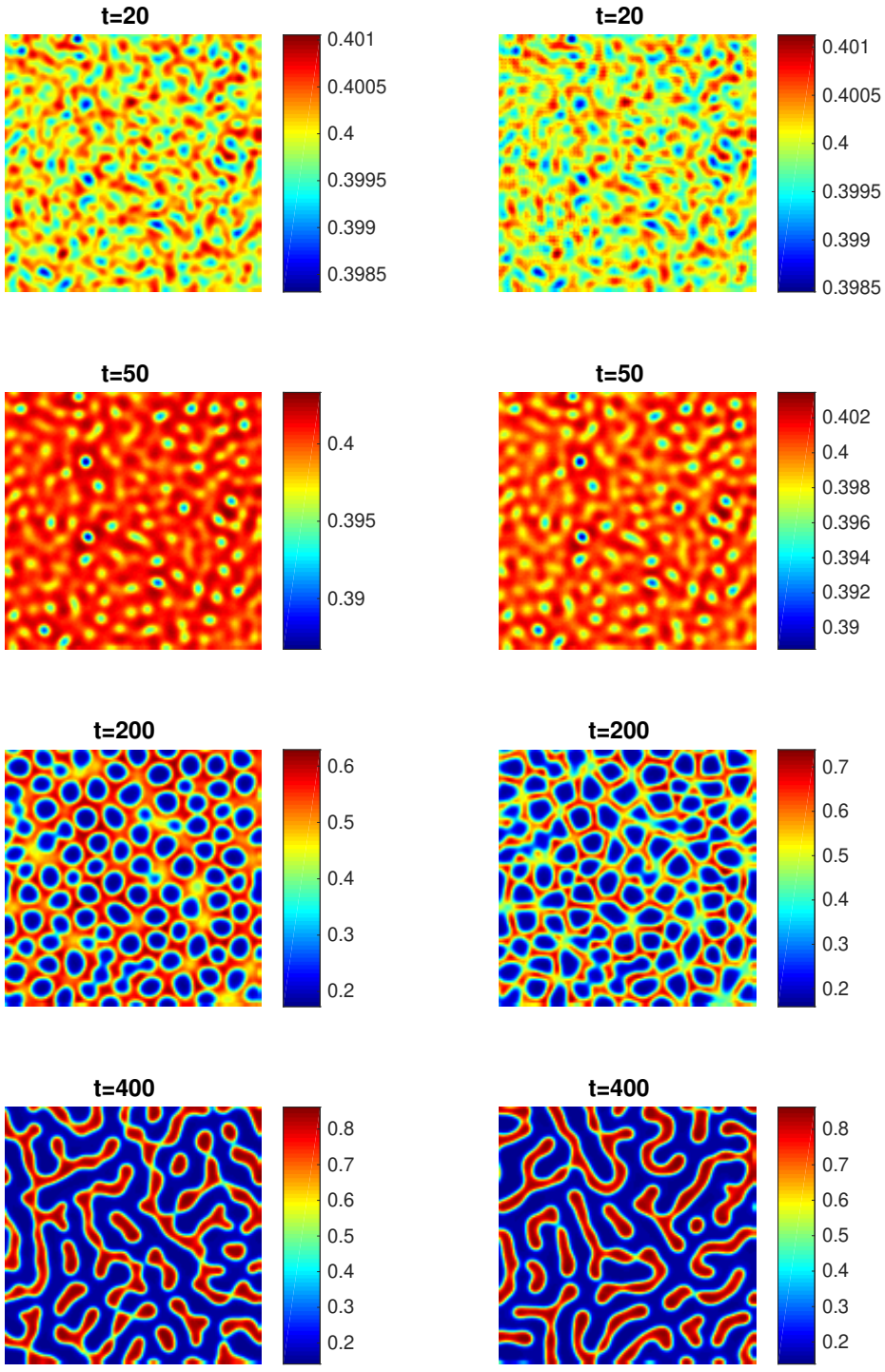


Figure 5.4: **Full model** 5.1.5: snapshots of volume fraction φ at different times t, using left and right finite differences (left) compared to using central differences (right).

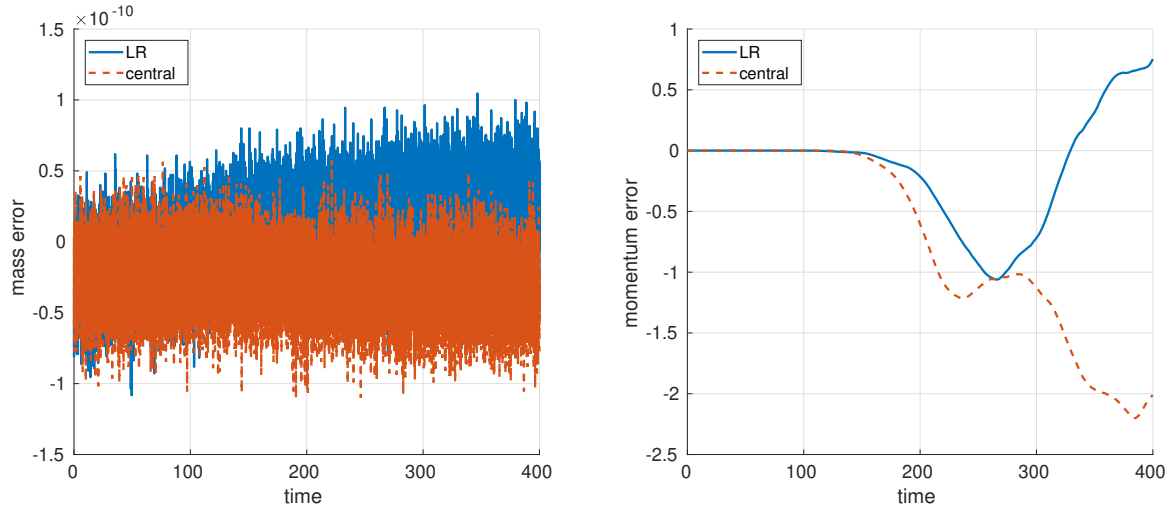


Figure 5.5: **Full model** 5.1.5: deviations from the conservation of mass and momentum, using left and right (LR) finite differences compared to using central differences.

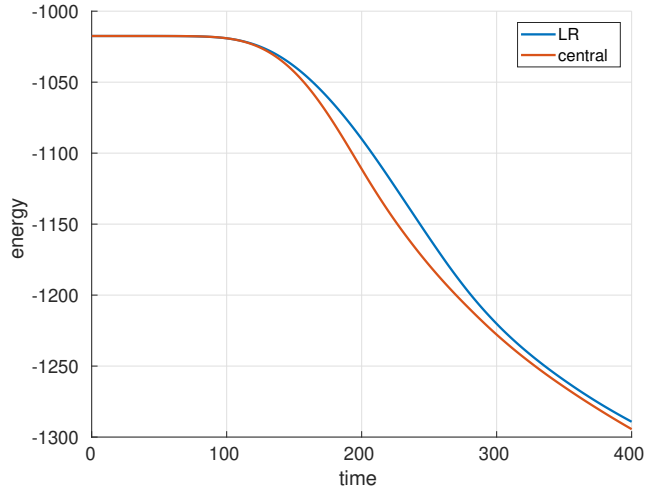


Figure 5.6: **Full model** 5.1.5: evolutions of the total energies E_{total}, using left and right (LR) finite differences compared to using central differences.

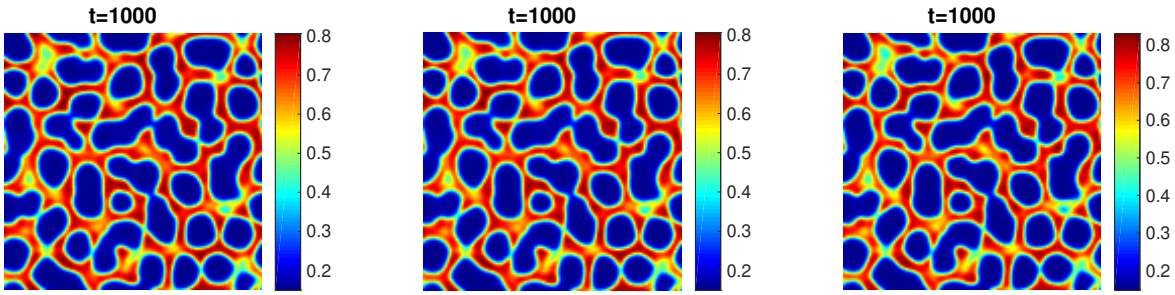


Figure 5.7: **Full model** 5.1.5: snapshots of volume fraction φ at time $t = 1000$, using flux difference upwinding (left), donor cell upwinding (center) or 2nd order DCU (right). Note that deviant from Experiment 5.1.5, $G_B^1 = 1$ instead of 0 here.

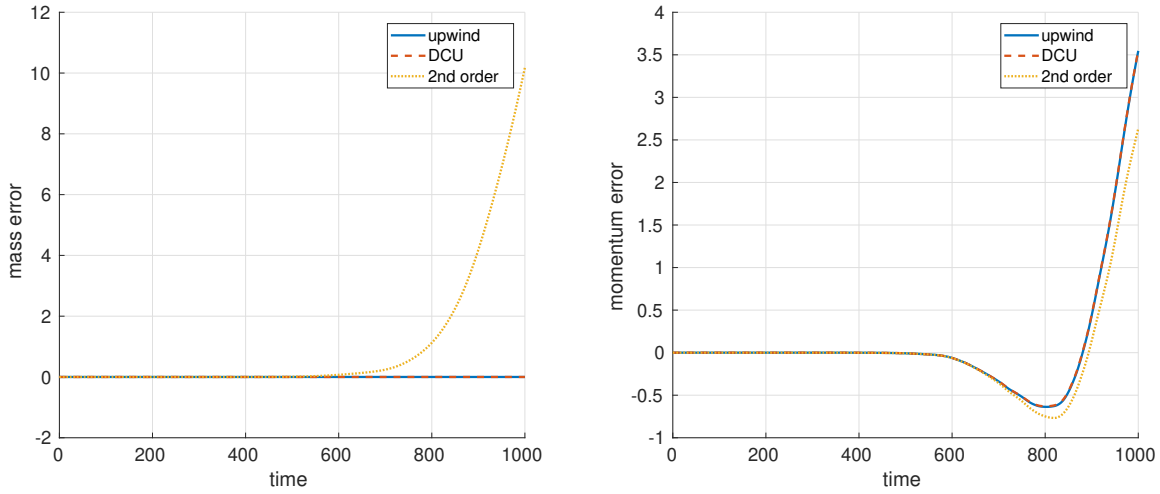


Figure 5.8: **Full model** 5.1.5: deviations from the conservation of mass and momentum, using flux difference upwinding (upwind), donor cell upwinding (DCU) or 2nd order DCU. Note that deviant from Experiment 5.1.5, $G_B^1 = 1$ instead of 0 here.

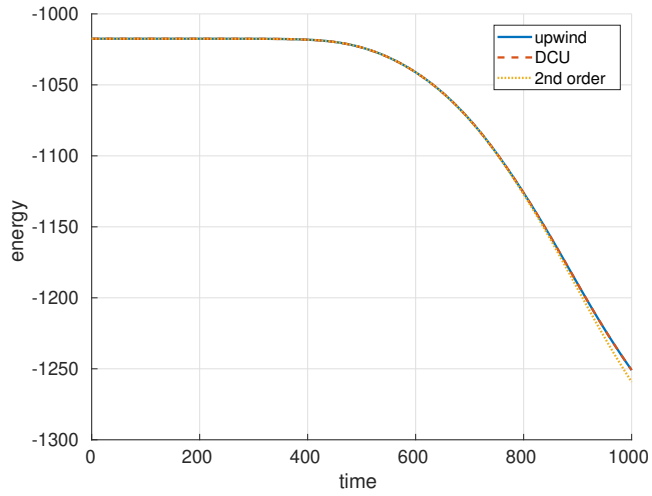


Figure 5.9: **Full model** 5.1.5: evolutions of the total energies E_{total}, using flux difference upwinding (upwind), donor cell upwinding (DCU) or 2nd order DCU. Note that deviant from Experiment 5.1.5, $G_B^1 = 1$ instead of 0 here.

5.2 Viscoelastic phase separation

The dynamics of viscoelastic phase separation can not be reproduced by the Cahn-Hilliard equation alone, see Figure 5.10. Here, we see the typical dynamics of the Cahn-Hilliard equation on the left side, which is phase separation by nucleation and coagulation of droplets. This phase separation process happens significantly faster than a viscoelastic phase separation process. Since a major share of the free energy originates from the interface between the phases and because this interface reduces during the phase separation process, the free energy of the Cahn-Hilliard equation decreases much earlier than the total energies of the simplified and the full model, see Figure 5.11.

In the right column of Figure 5.10 and in more detail in the long time experiment of the full model, see Figure 5.12, the whole viscoelastic phase separation process is exhibited as described by its essential features by *Tanaka* [68], based on observations from real lab experiments.

Starting from a homogeneous 40% polymer concentration with noise at $t = 0$, we see an aggregation of a minimally polymer-richer phase and solvent-rich droplets, which appear relatively even distributed in the computational domain. Since the volume fraction of the solvent-rich droplets deviates more from the initial polymer concentration than the polymer-richer phase, the volume of the latter increases initially until around $t = 60$. This is the so-called *frozen period*, since the variance of the volume fraction stays quite small during this initial process. This is emphasized by Figure 5.13, where we fixed the color bar to the entire solvent-polymer interval $[0, 1]$.

Then, the concentration of the polymer-rich phase continuously increases, see Figure 5.12, $t = 120$ and $t = 180$. In this process, the polymer-rich phase naturally shrinks in volume, since our closed system follows the law of conservation of mass. This *volume shrinking* of the polymer-rich phase results in the formation of a *network-like structure*.

Next, from around $t = 240$, this network-like structure starts breaking, whereby the solvent-rich droplets become a more and more coherent solvent-rich phase. Therefore, this process is called *phase inversion*.

Finally, the more and more separated polymer-rich structures slowly reduce their interface by reshaping to droplets and by coagulation. This is the typical long time dynamics of the Cahn-Hilliard equation.

In Figure 5.14, we see snapshots of the time evolution of the elastic bulk stress, which takes the same geometrical shape as the volume fraction φ. Since its initial value is zero and its evolution equation is relaxant, the absolute bulk stress can only increase due to the coupling to the other equations. In Figure 5.15, we visualize the time evolution of the velocity field by snapshots of its Euclidean norm. For this purpose, we have to map its two components to the cell centers, since they are originally on staggered grids. Note that we do not plot the shear stress, since a visualization of this tensor of rank two will not yield us much useful information.

Since the initial values for $q, \boldsymbol{\sigma}$ and $\boldsymbol{u}$ are zero, the respective energies, see Figure 5.16, can only increase in the beginning, which happens because of the coupling of the equations. But obviously, these energies increase significantly less than the mixing energy decreases. Therefore, the total energy evolves quite similarly to the mixing energy and is always non-increasing in time.

In Figure 5.17, we see a negligible deviation from the conservation of mass, given by the discrete version of

$$\int_\Omega \left(\varphi^n - \varphi^0\right) \qquad \forall n \in \mathbb{N}\,.$$

This verifies our proven conservation on the discrete level. The deviation from the conservation of momentum on the other hand is considerable, even though one should keep in mind that also the deviations scale with domain size and that is $[0, 128]^2$. It is calculated by the discrete version of

$$\int_\Omega \left(u^n - u^0 + v^n - v^0\right) \qquad \forall n \in \mathbb{N}\,.$$

The numerical dissipation of the potential ND_{pot}, see Figure 5.18, is always non-positive here and could therefore in principle violate the non-increasing nature of the total energy. But since it is significantly smaller than $(\Delta t)^2 = 0.025^2 = 6.25 \cdot 10^{-4}$ here, this is presumably only critical when using large time steps. The dissipation has eye-catching peaks here and there. These relate to major interface changes, e.g., during the merging of droplets.

The error of the discrete velocity divergence, see the right hand side of Figure 5.18, is negligible. It is calculated by

$$\|\nabla \cdot u^n\|_{L^\infty(\Omega)}\,.$$

In the central column of Figure 5.10 and in more detail in the long time experiment of the simplified model, see Figure 5.19, we can clearly observe that also this model captures the most important physical mechanisms of the viscoelastic phase separation process. Having all initial values set to zero, except for the volume fraction, only the long time dynamics differs on the face of it when comparing the numerical solutions of the simplified model to those of the full model, see Figure 5.10.

The snapshots of the time evolution of the elastic bulk stress in Figure 5.20 show the same geometrical shape as the snapshots of volume fraction φ, which is similar to the full model. Also the total energy evolves similarly, see Figure 5.21, where again the bulk energy increases slightly in the beginning due to the coupling of the equations, but significantly less than the mixing energy decreases. Thus, the total energy again evolves quite similarly to the mixing energy and is always non-increasing in time. The deviation from the conservation of mass, see Figure 5.22, is again negligible, and the numerical dissipation is again always non-positive but uncritical since it is significantly smaller than $(\Delta t)^2 = 0.1^2 = 10^{-2}$.

The main disadvantages of the simplified model compared to the full model are the reduced number of parameters to control the dynamics of the phase separation process as well as the missing velocity field. The latter is a crucial part of molecular dynamics simulations, to which we will compare our simulations later. Beforehand, we will have a look at the impact to the dynamics of most parameters and some initial values of the evolution equations in the following experimental sensitivity analysis.

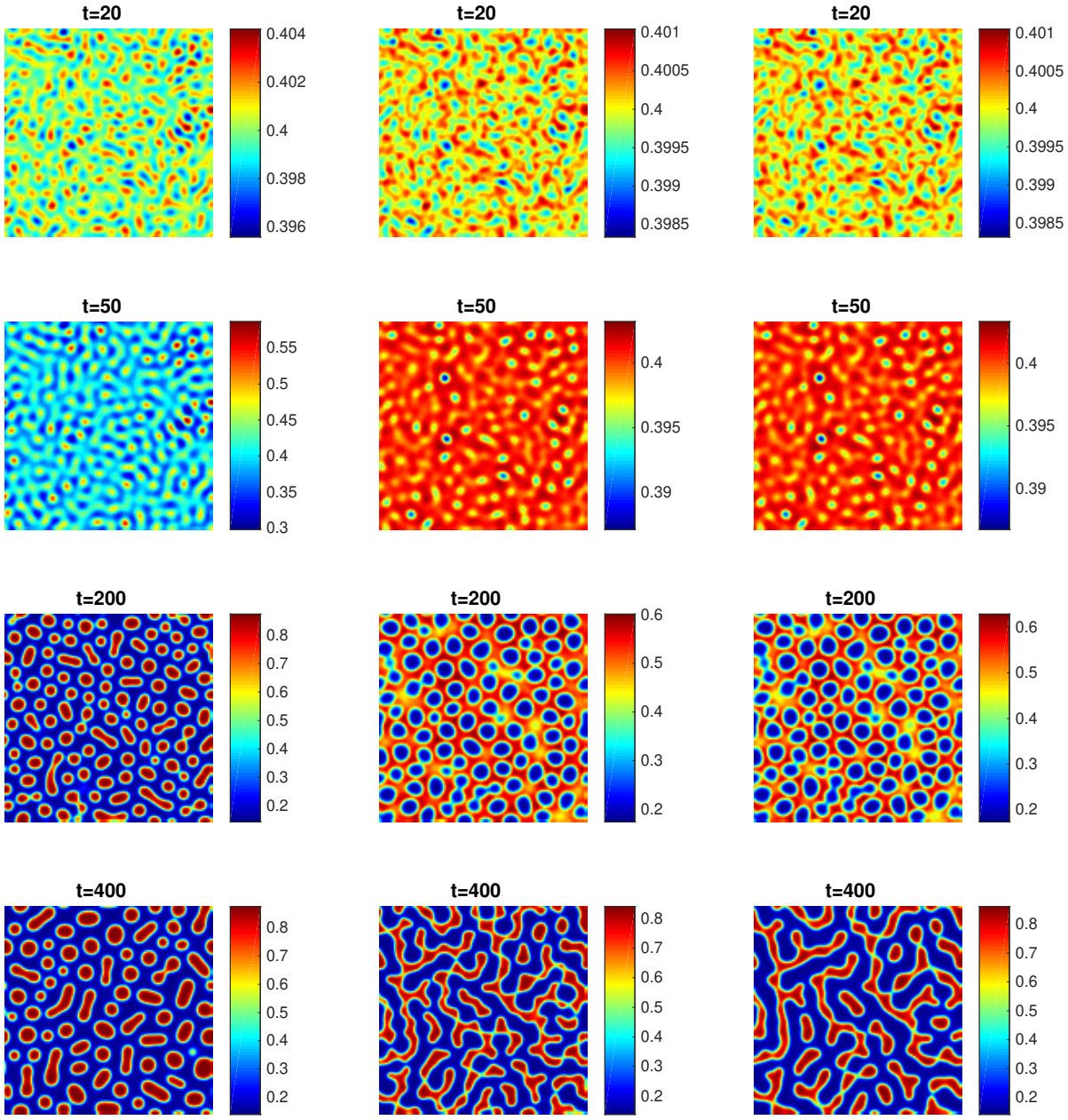


Figure 5.10: **Cahn-Hilliard equation** (left) vs. **simplified model** (center) vs. **full model** (right): snapshots of volume fraction φ at several times t, using the parameters and initial values of Experiment 5.1.5 and time step size $\Delta t = 0.025$.

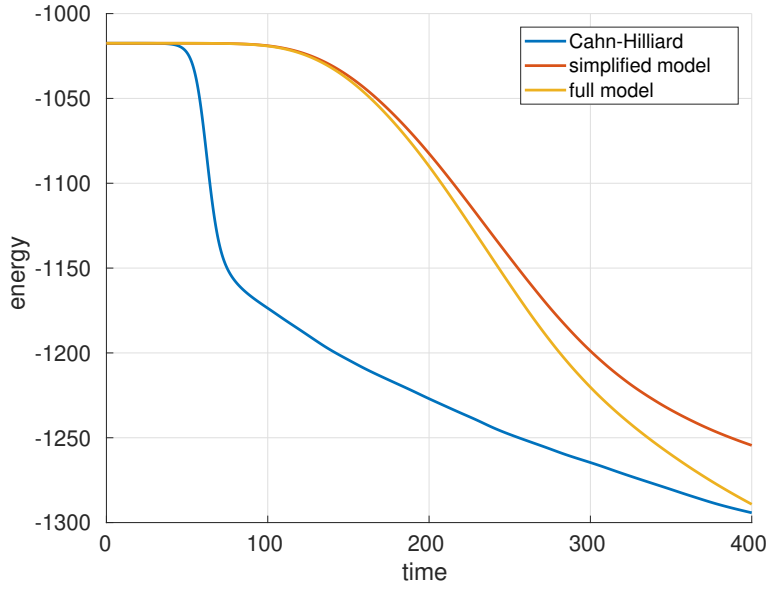


Figure 5.11: **Cahn-Hilliard equation** (left) vs. **simplified model** (center) vs. **full model** (right): evolutions of the total energies E_{total}, using the parameters and initial values of Experiment 5.1.5 and time step size $\Delta t = 0.025$.

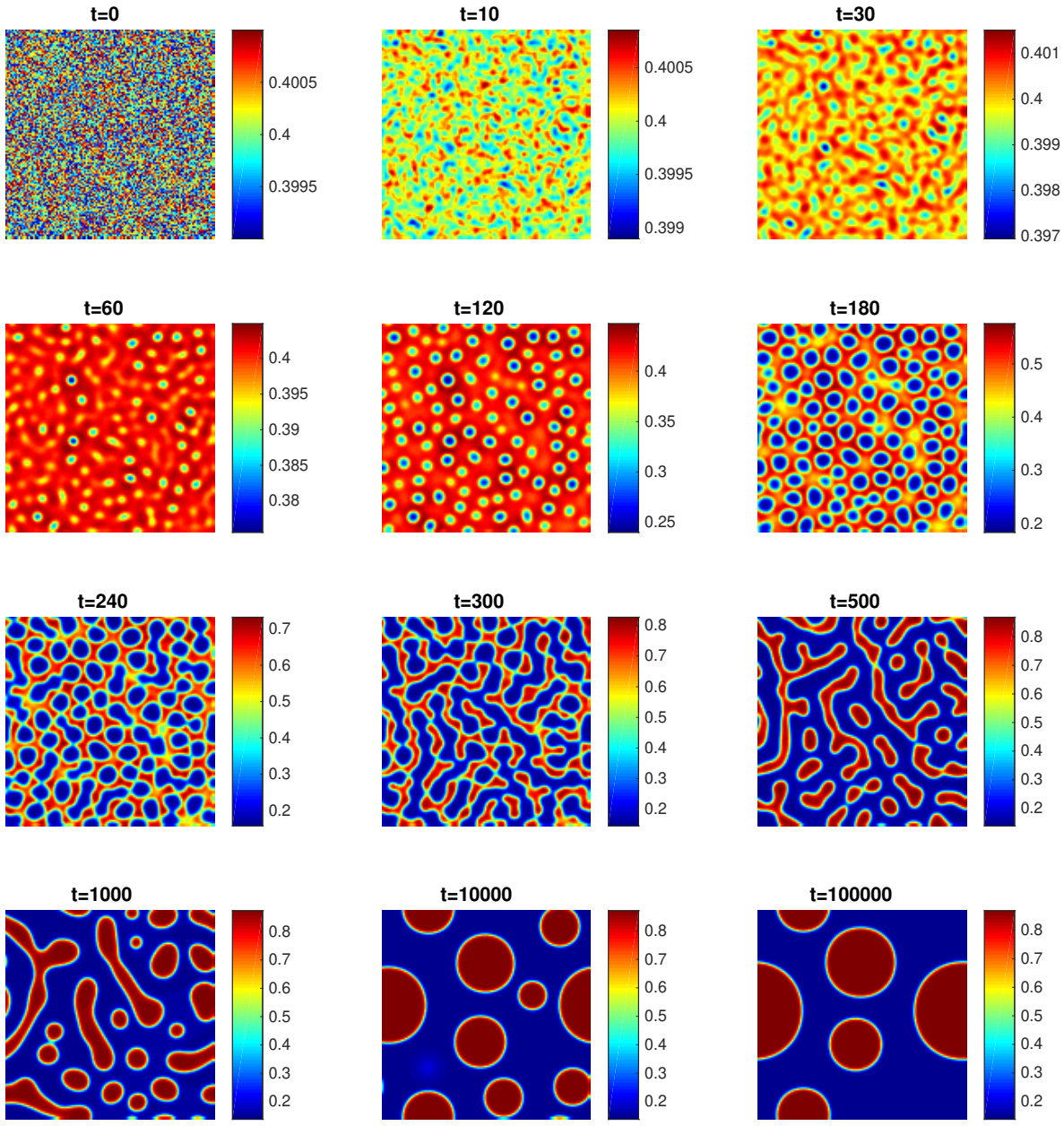


Figure 5.12: **Full model** 5.1.5: snapshots of volume fraction φ at different times t, using $\Delta t = 0.025$.

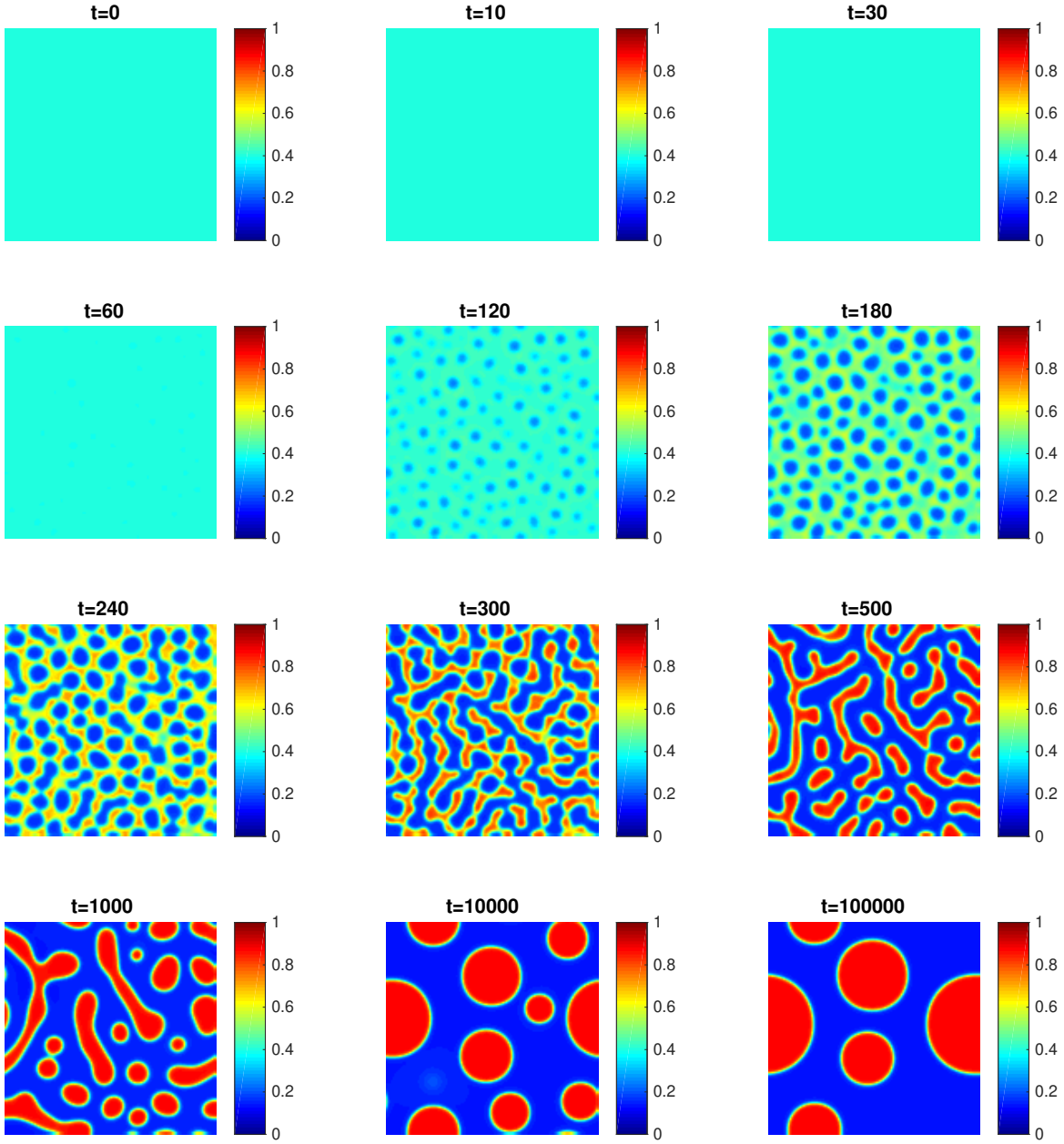


Figure 5.13: **Full model** 5.1.5: snapshots of volume fraction φ at different times t, using $\Delta t = 0.025$. Here, the color bar is fixed to the whole solvent-polymer interval $[0, 1]$.

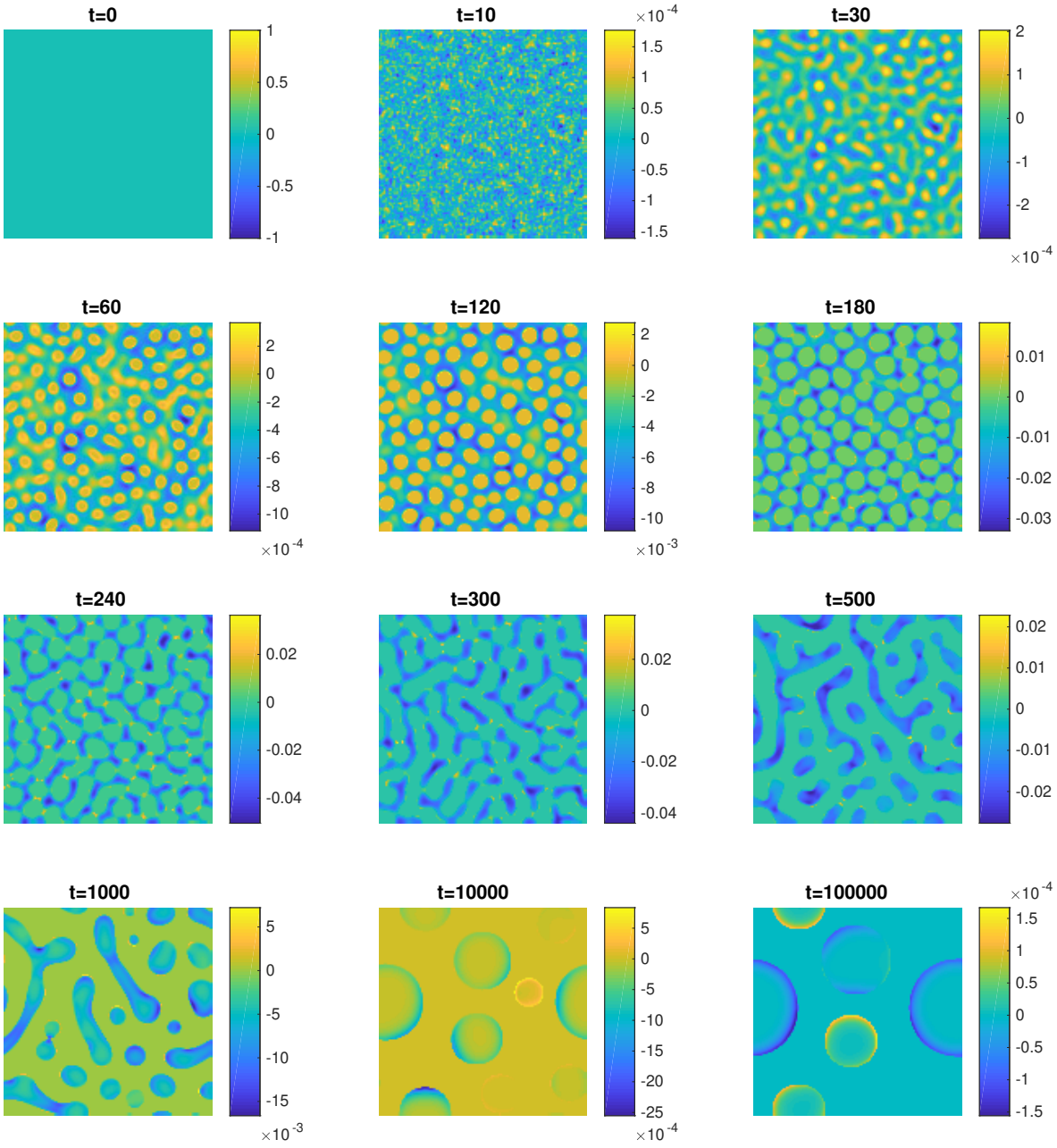


Figure 5.14: **Full model** 5.1.5: snapshots of bulk stress q at different times t, using $\Delta t = 0.025$.

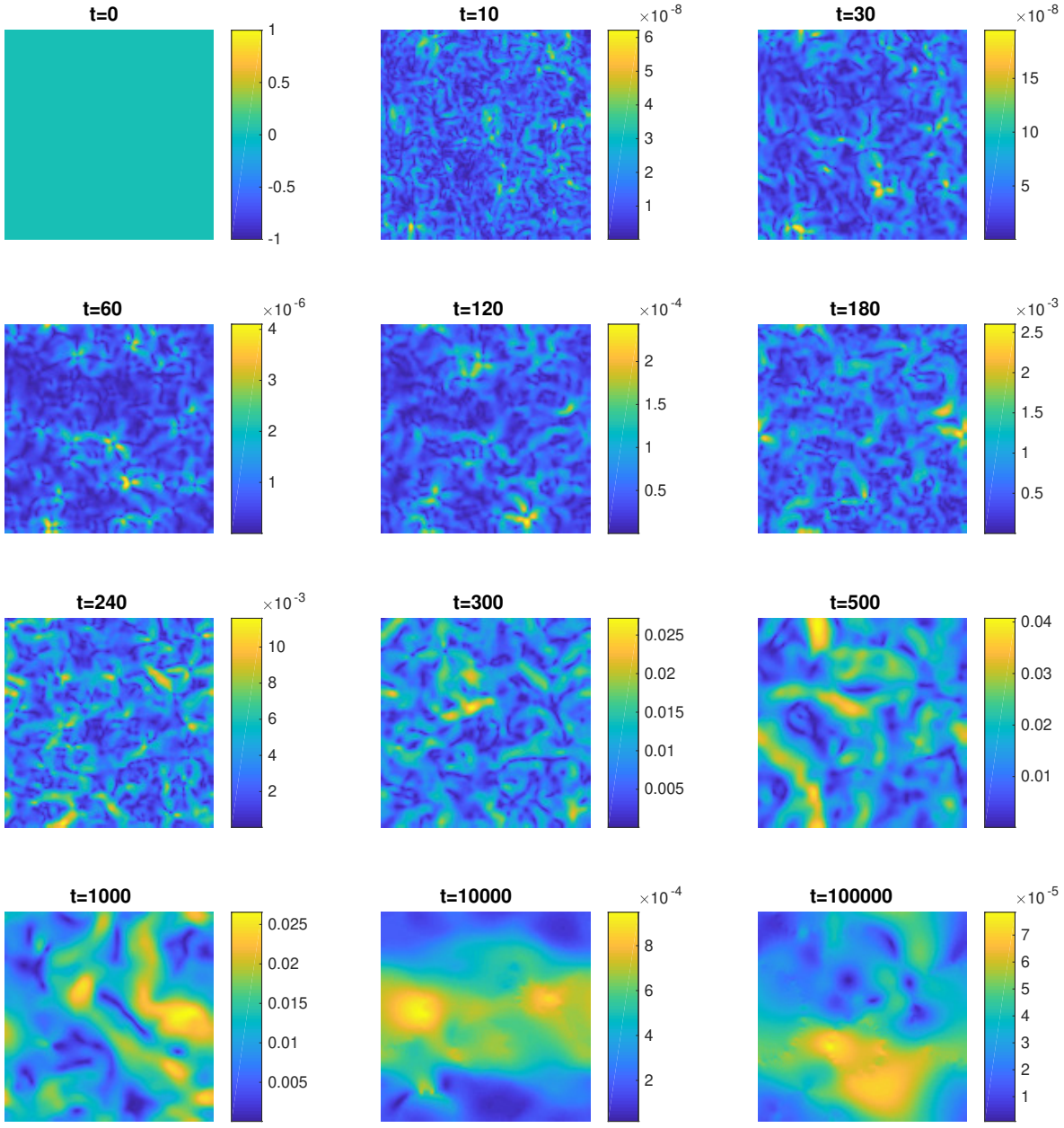


Figure 5.15: **Full model** 5.1.5: snapshots of the Euclidean norm of velocity field $\boldsymbol{u}$ at different times t, using $\Delta t = 0.025$.

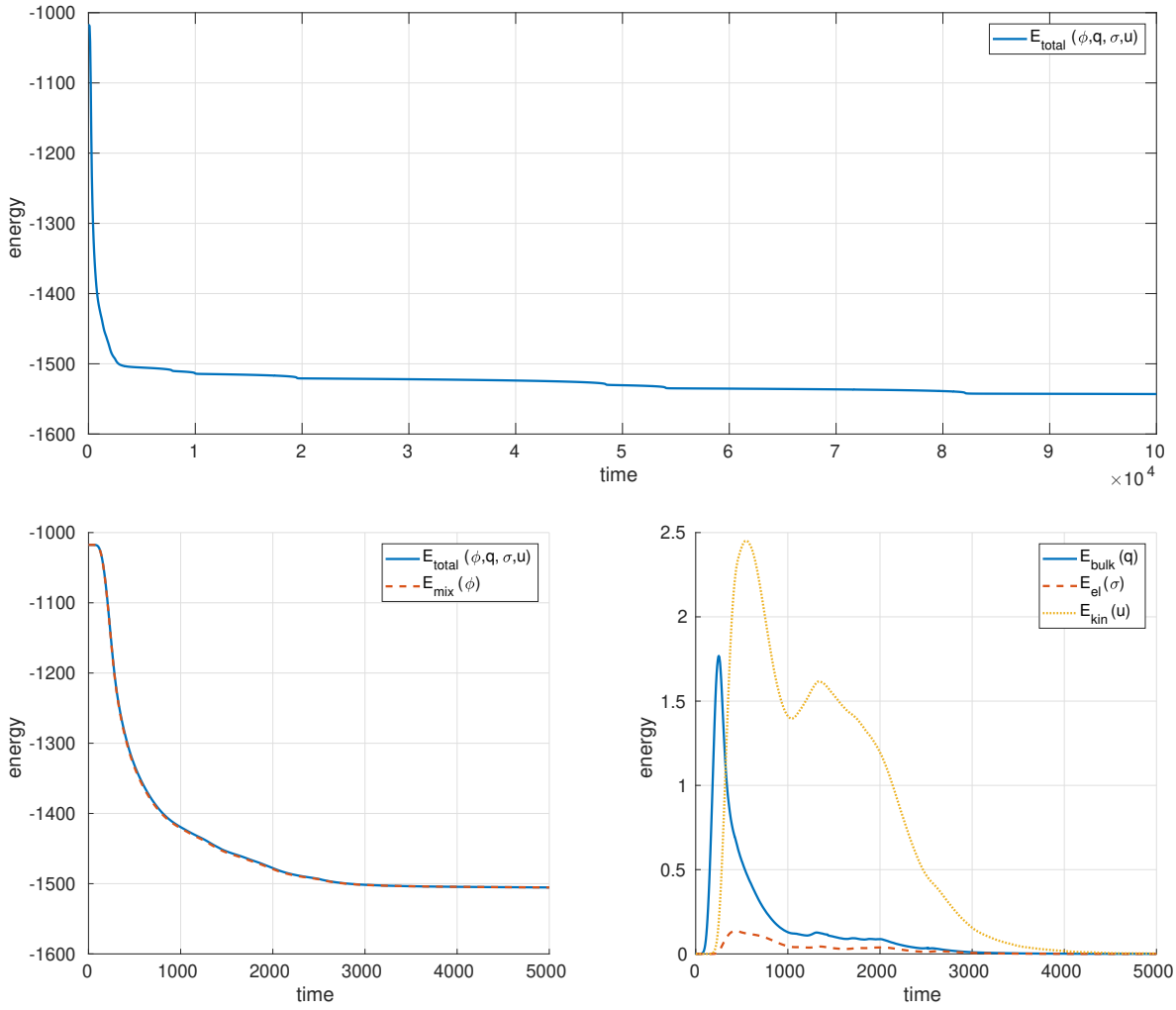


Figure 5.16: **Full model** 5.1.5: energy evolution, using $\Delta t = 0.025$. Plotted until $t = 10^5$ at the top and separately until $t = 5000$ below.

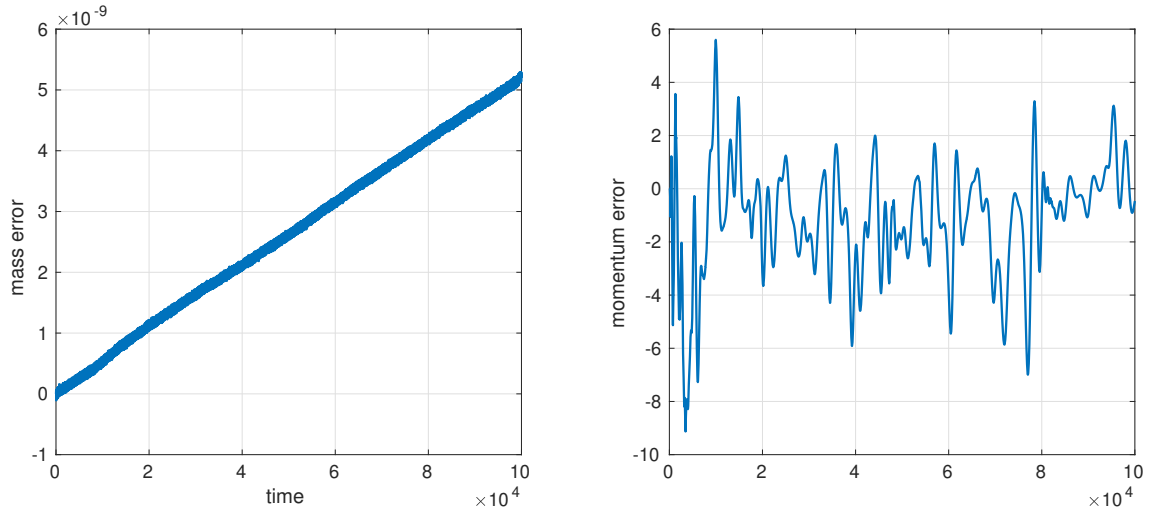


Figure 5.17: **Full model** 5.1.5: deviations from the conservation of mass and momentum, using $\Delta t = 0.025$.

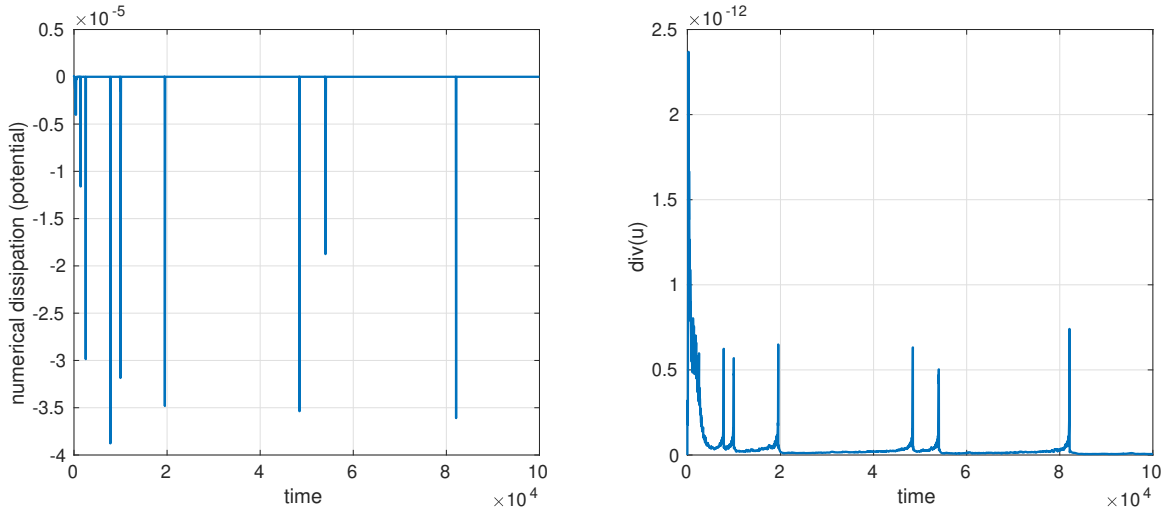


Figure 5.18: **Full model** 5.1.5: numerical dissipation ND_{pot} and the velocity divergence, using $\Delta t = 0.025$.

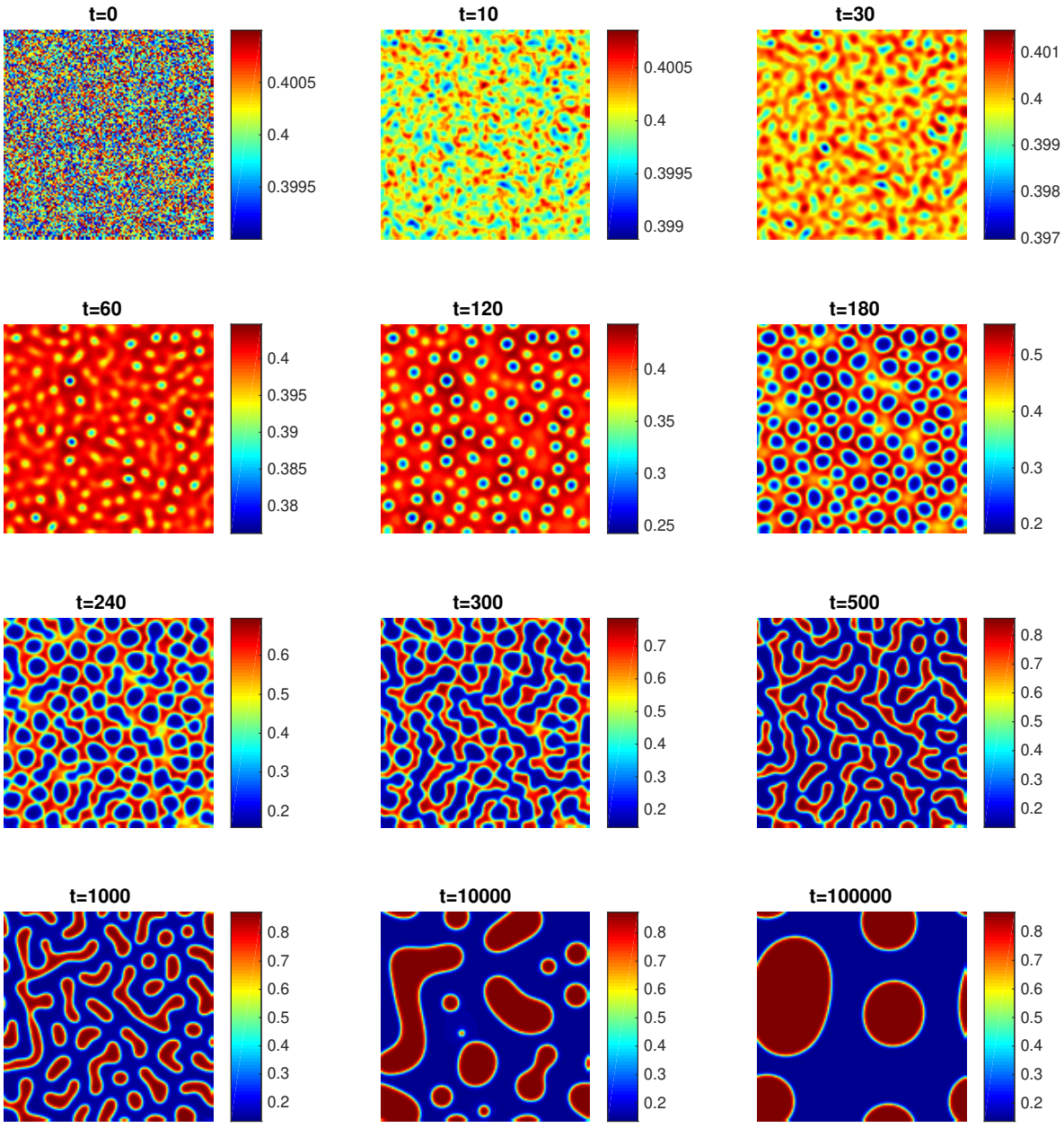


Figure 5.19: **Simplified model** 5.1.4: snapshots of volume fraction φ at different times t, using $\Delta t = 0.1$ and a 128×128 grid.

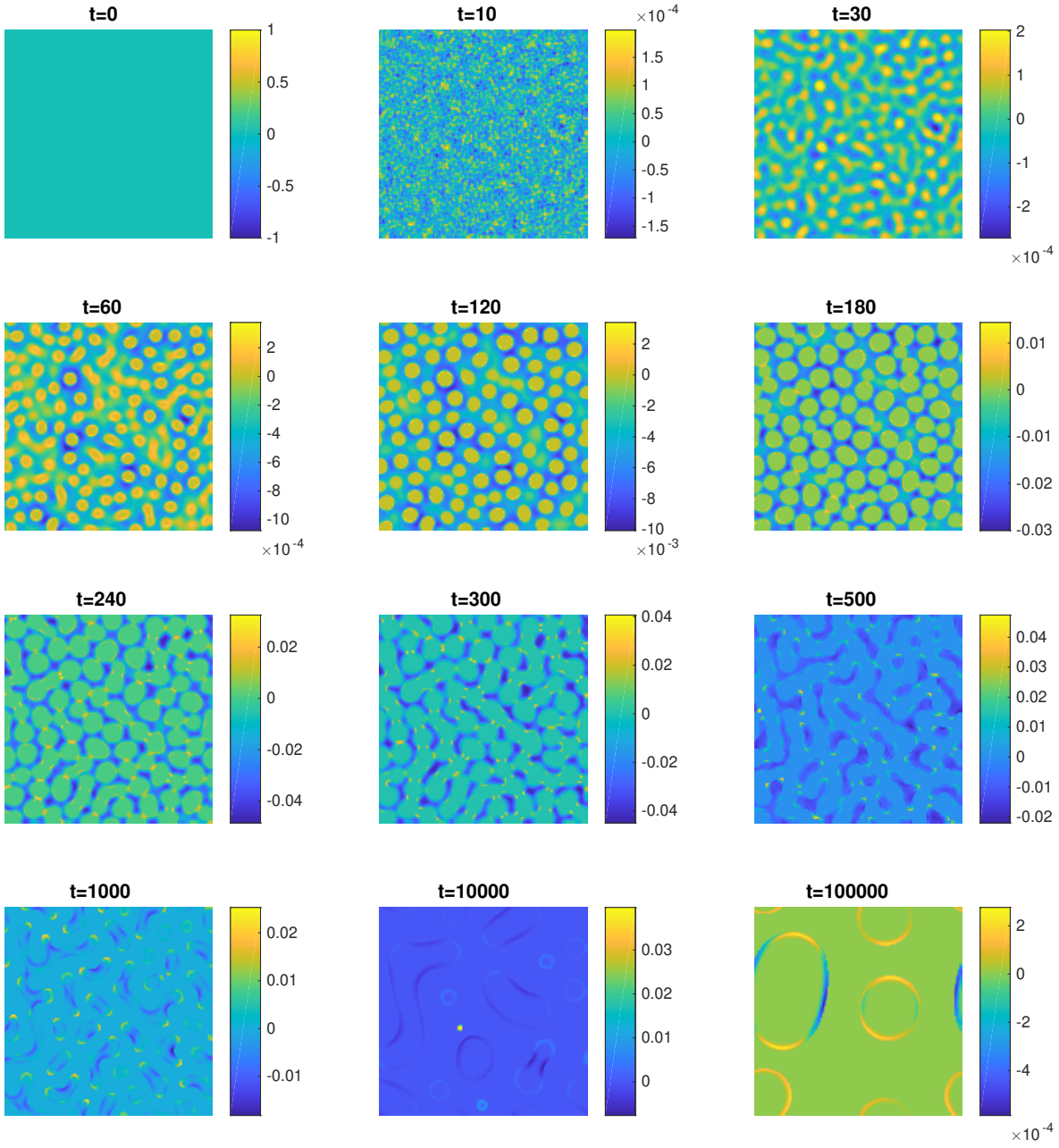


Figure 5.20: **Simplified model** 5.1.4: snapshots of bulk stress q at different times t, using $\Delta t = 0.1$ and a 128×128 grid.

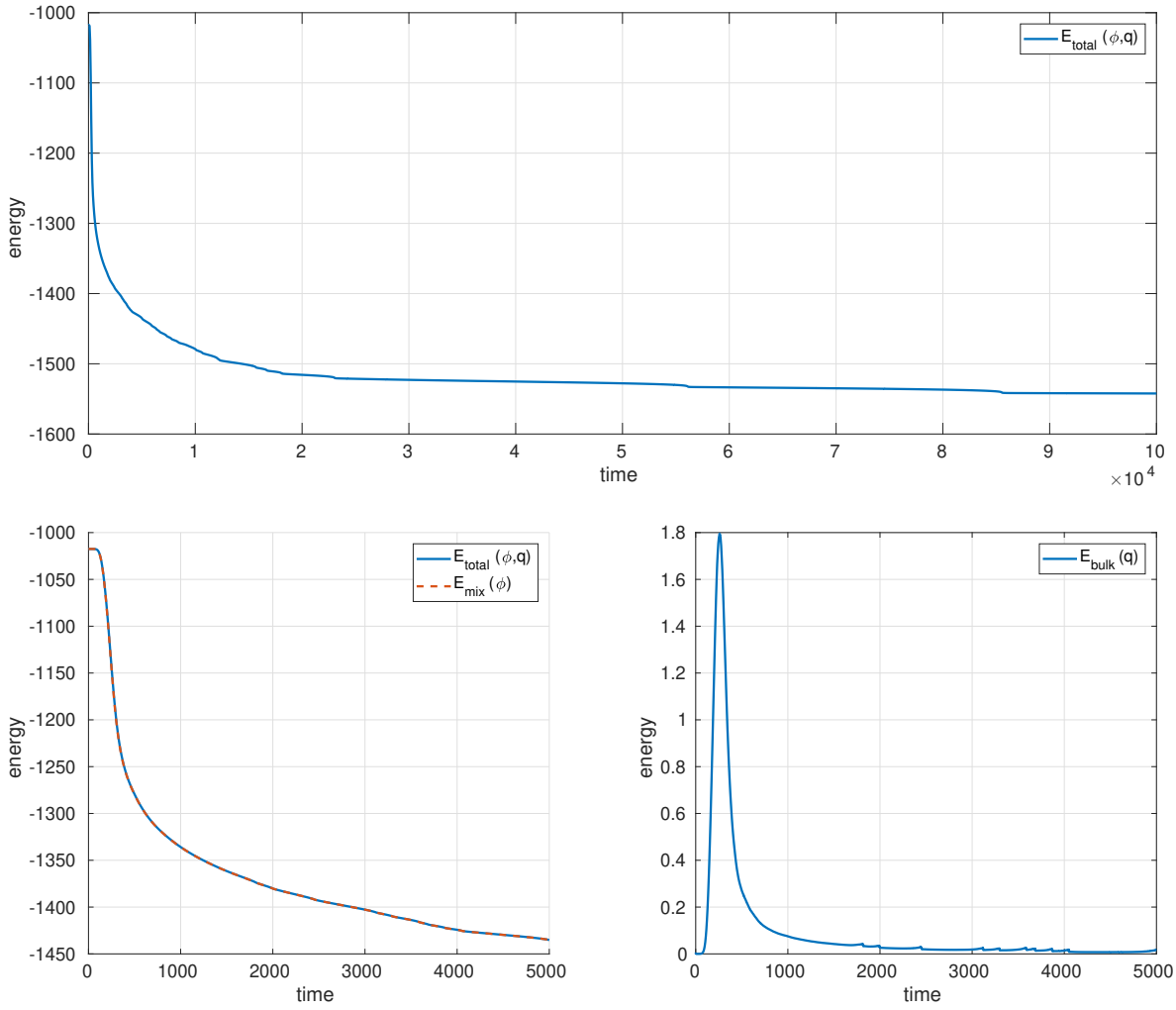


Figure 5.21: **Simplified model** 5.1.4: energy evolution, using $\Delta t = 0.1$ and a 128×128 grid. Plotted until $t = 10^5$ at the top and separately until $t = 5000$ below.

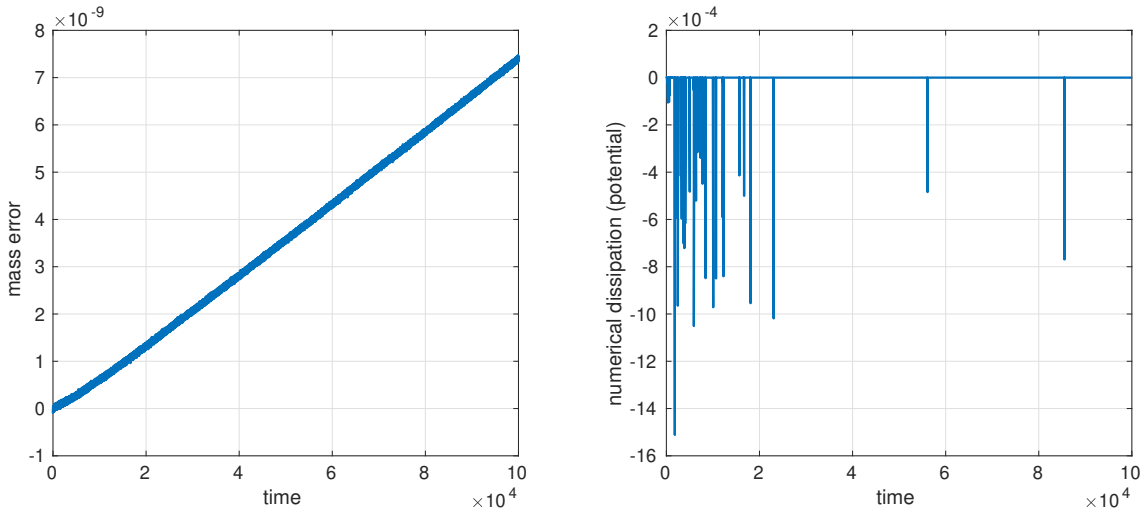


Figure 5.22: **Simplified model** 5.1.4: deviation from the conservation of mass and the numerical dissipation term ND_{pot}, using $\Delta t = 0.1$ and a 128×128 grid.

5.2.1 Experimental sensitivity analysis

In order to experimentally analyze the sensitivity of the full model and the simplified model with respect to their parameters and initial values, we compare simulations using an extensive collection of parameter sets, which are listed in Table 5.6. In this table, the first parameter set is our reference set, which we also used for the experimental convergence study in time of the full and the simplified model above. Deviations from this parameter set in the sets listed underneath are highlighted in bold. All simulations are computed on the two-dimensional domain $[0, 128]^2$, using a 128×128 grid, i.e., each grid cell has size 1×1, and using the time step size $\Delta t = 0.025$. Initial values except the volume fraction φ^0 are all set to zero and the degree of polymerization of the solvent remains $n_s = 1$. The initial volume fraction is defined by

$$\varphi^0 = \varphi_c^0 + \delta\Delta(x, y)\,, \qquad (x, y) \in [0, 128] \times [0, 128]\,,$$

where $\Delta(x, y)$ is a uniformly distributed random variable with range $[-1, 1]$, whose values are the same for each experiment.

Table 5.6: Parameter sets used for the experimental sensitivity analysis.

	φ_c^0	δ	ζ	λ	n_p	χ	τ_B^0	G_B^0	G_B^1	φ^*	τ_S^0	G_S^0	η	Figure
1	0.4	0.001	0.1	1	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.24
2	0.4	**0.01**	0.1	1	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.24
3	**0.35**	0.001	0.1	1	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.24
4	**0.45**	0.001	0.1	1	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.25
5	**0.5**	0.001	0.1	1	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.25
6	**0.6**	0.001	0.1	1	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.25
7	0.4	0.001	**1**	1	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.26
8	0.4	0.001	0.1	**2**	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.26
9	0.4	0.001	0.1	**4**	1	2.5455	10	0.5	0	φ^0	5	0.5	1	5.26
10	0.4	0.001	0.1	1	**3**	**1.8**	10	0.5	0	φ^0	5	0.5	1	5.27
11	0.4	0.001	0.1	1	**10**	**1.35**	10	0.5	0	φ^0	5	0.5	1	5.27
12	0.4	0.001	0.1	1	1	2.5455	**20**	0.5	0	φ^0	5	0.5	1	5.28
13	0.4	0.001	0.1	1	1	2.5455	**100**	0.5	0	φ^0	5	0.5	1	5.28
14	0.4	0.001	0.1	1	1	2.5455	10	0.5	**1**	φ^0	5	0.5	1	5.29
15	0.4	0.001	0.1	1	1	2.5455	10	**0**	**1**	φ^0	5	0.5	1	5.29
16	0.4	0.001	0.1	1	1	2.5455	10	0.5	0	**0.2**	5	0.5	1	5.29
17	0.4	0.001	0.1	1	1	2.5455	10	0.5	0	φ^0	**50**	0.5	1	5.33
18	0.4	0.001	0.1	1	1	2.5455	10	0.5	0	φ^0	5	**5**	1	5.33
19	0.4	0.001	0.1	1	1	2.5455	10	0.5	0	φ^0	5	0.5	**0.1**	5.33

Note that only the first one and the last three parameter sets from Table 5.6 are simulated using the full model. The other parameter sets are simulated using the simplified model.

Starting with Figure 5.24, where on the left hand side we have snapshots of a simulation using our reference parameter set 1, we see in the central column that the increased noise δ of the apart from that homogeneous initial volume fraction φ^0 in parameter set 2 slightly

accelerates the separation dynamics. This yields an energy evolution, which is slightly shifted to the left in Figure 5.23a. On the right hand side of Figure 5.24, one can observe that the lower initial polymer volume fraction in parameter set 3 not only slows down the initial dynamics but also accelerates the dynamics later on, because of having less network-like structure. Both is confirmed by the respective energy evolution in Figure 5.23a, where the major energy reduction happens later but in a shorter time interval. Note that the initial total energy is by definition lower for the lower initial volume fraction, since the symmetric Flory-Huggins potential we use here holds $F(0.35) < F(0.4)$.

In Figure 5.25, we see effects of the small to large increases of the initial polymer volume fraction in parameter sets 4, 5 and 6. While the initial dynamics accelerates through the initial volume fraction increases up to $\varphi_c^0 = 0.5$, it slightly decelerates for $\varphi_c^0 = 0.6$. The dynamics also changes substantially for the latter. No network-like structures arise and the phase inversion is missing completely in cause of having more polymer than solvent in the mixture. For the 50-50 mixture, it is unclear whether phase inversion occurs. For $\varphi^0 = 0.45$, the visible volume of the polymer-rich phase is at least slightly lower than that of the solvent-rich phase around $t = 1000$. Despite the substantial dynamics differences, the respective energy evolutions in Figure 5.23b are inconspicuous, differing primarily only initially while evolving quite similarly. Thus, similarities in the energy evolution do not necessarily point to similar dynamics.

Figure 5.26 shows the effects of the significantly increased friction ζ in parameter set 7 and of the increased interface sizes λ in parameter sets 8 and 9. Here, all modifications slow down the dynamics compared to the reference parameter set. This is confirmed by the respective energy evolutions in Figure 5.23c, where the major energy reductions occur later. The increased friction restricts the initial aggregation of a polymer-richer phase and of solvent-rich droplets, such that we see a different kind of frozen period and nearly no volume shrinking of the polymer-rich phase later on. For the increased interface sizes, the dynamics still follow the four essential features, but less solvent-rich droplets emerge and the formed network-like structures are significantly coarser. Summarized, the dynamics are very sensitive to both parameters.

In Figure 5.27, we compare the dynamics using the reference parameter set, having a symmetric double-well potential, to the dynamics using parameter sets 10 and 11, having asymmetric potentials. See Figure 5.30 for plots of the three different variants of the Flory-Huggins potential. Similarly to the initial 50-50 volume fraction, the dynamics using asymmetric potentials do not clearly show phase inversion. The total energies, see Figure 5.23d, differ substantially initially, because the potential is a major part of the mixing energy. Nevertheless, the evolutions of the total energies are quite similar.

In Figure 5.28, we compare the reference parameter set to parameters sets 12 and 13, in which we increase the bulk relaxation time from 1 to 20 and 100. The noticeably increased relaxation time clearly slows down the dynamics, which is confirmed by the respective energy evolutions in Figure 5.23e. For $\tau_B^0 = 100$, the dynamics become so slow that the simulation is still in the volume shrinking phase at $t = 1000$. Despite this, the dynamics evolve similarly, just on different time scales.

Our last comparison based on the simplified model, see Figure 5.29, demonstrates the influences of the bulk relaxation modulus changes in parameter sets 14, 15 and 16. Let us

recall the bulk relaxation modulus

$$G_B(\varphi) = G_B^0 \left[1 + \tanh\left(\frac{\cot(\pi\varphi^*) - \cot(\pi\varphi)}{\varepsilon} \right) \right] + G_B^1 ,$$

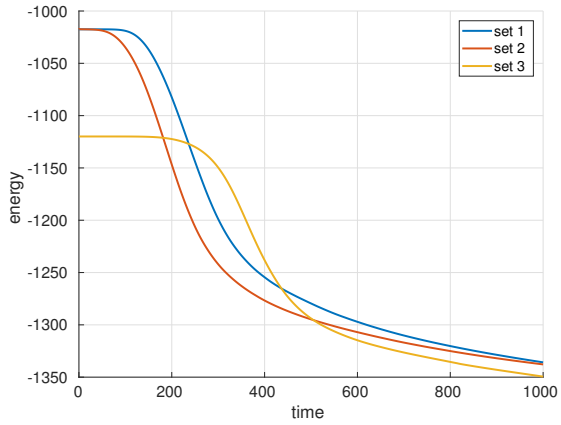
where in the reference parameter set $G_B^0 = 0.5$, $G_B^1 = 0$, $\varphi^* = 0.4$ and $\varepsilon = 0.01$. Thus, we have a smoothed step function, which is zero for volume fraction values below the critical concentration for polymers to crosslink φ^*, where the latter is equal to the conserved polymer mass φ_c^0. Parameter set 14 introduces an offset to the relaxation modulus. Therefore, we have bulk relaxation also below the critical concentration, but a twice as high one above it. Parameter set 15 removes the volume fraction dependency of the bulk relaxation modulus by setting it to a constant value of 1. While in parameter set 16, we reduce the critical concentration to 0.2, such that the bulk relaxation modulus only vanishes for volume fraction values below $\varphi = 0.2$, which are usually only reached in later stages of the phase separation. All these variants of the bulk relaxation modulus are visualized in Figure 5.31. On the left hand side of Figure 5.29, we see that the shifted bulk relaxation modulus slows down the dynamics noticeably compared to the reference parameter set. The simulation with a constant modulus evolves nearly identical to the one having a modulus with a lowered critical concentration. This is reasonable, since as mentioned above, volume fraction values below this concentration are reached pretty late in the separation process. Both simulations skip the frozen period and therefore also the volume shrinking of the polymer-rich phase. This leads to a significantly earlier and faster decreasing energy, see Figure 5.23f. Also, network-like structures are formed less pronounced than in simulations using the reference parameter set. Summarized, having an offset in the bulk relaxation modulus is applicable, while the other modifications of the bulk relaxation modulus are not suggested. Let us note in this context that *Zhou, Zhang and E* [74] use parameter set 14, i.e., an offset in the bulk relaxation modulus, for their simulations, while *Tanaka* [68] uses a bulk relaxation modulus without offset.

We finalize our parameter study by comparing three simulations using the full model in Figure 5.33. Here, we investigate the effects of changes to the shear relaxation time, shear relaxation modulus and viscosity, which do not appear in the simplified model due to the neglected hydrodynamics. We can see from the phase separation dynamics in Figure 5.33 as well as from the respective energy evolutions in Figure 5.32 that the increased shear relaxation time in parameter set 17 and the increased shear relaxation modulus in parameter set 18 have no perceptible effects. This is probably closely related to the fact that the dynamics of the simplified model are in general very similar to those of the full model, if the latter has zero initial velocity field and shear stress, as we have seen in Figure 5.10. Lowering the viscosity on the other hand yields coarser network-like structures. Thus, hydrodynamics can definitely have notable influences on the dynamics.

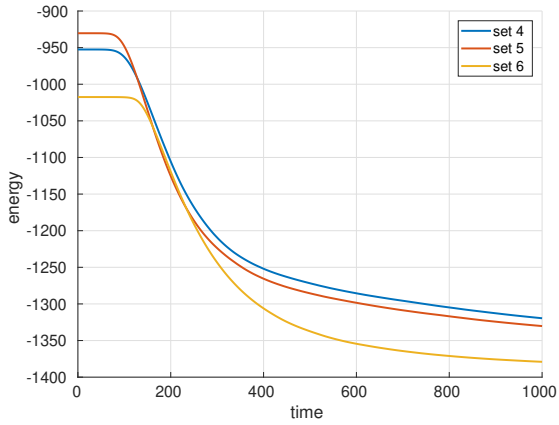
Additionally to the parameter sets from Table 5.6, let us have a look at the dynamics of viscoelastic phase separation in a vortex. For this purpose, we use the full model experiment from Subsection 5.1.5 (using parameter set 1 from Table 5.6), but with the following initial conditions for the velocity field

$$\begin{aligned} u^0 &= -\sin(\pi x/128)^2 \sin(2\pi y/128) , \\ v^0 &= \sin(2\pi x/128) \sin(\pi y/128)^2 , \qquad (x, y) \in [0, 128] \times [0, 128] , \end{aligned} \tag{5.1}$$

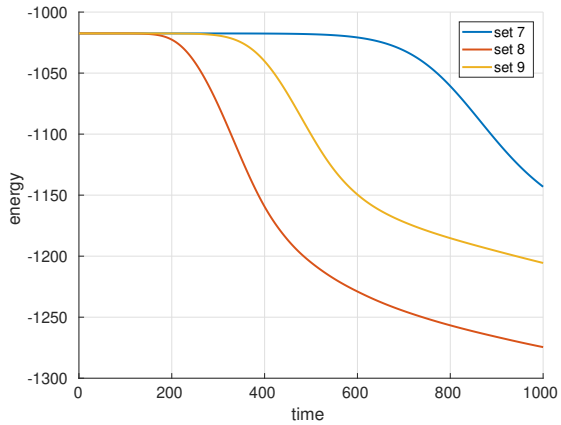
where (u^0, v^0) is divergence free. In Figure 5.34, we present detailed long time dynamics of the phase-field in a vortex. Since the velocity field is initially zero in the four corners of the computational domain and in its center, see Figure 5.36, the viscoelastic phase-field dynamics with its four essential features is still recognizable in these areas. In the vortex itself on the other hand, the dynamics clearly follow the stream. This yields less solvent droplet nucleation in the earlier stages, till around $t = 200$. Since the vortex slows down over time, more solvent-rich areas emerge around $t = 300$. Led by the remaining flow, the network-like structures forming around $t = 500$ have a dominant circular shape. Between $t = 700$ and $t = 1000$, the vortex vanishes such that the usual coagulation of polymer-rich structures and subsequently droplet forming begins. During all times, the elastic bulk stress q takes again the same geometric shape as the volume fraction, see Figure 5.35. The following other observables also behave similarly to the long time dynamics simulation of the full model with zero initial velocity field, which we described in the beginning of this section. The energy evolution is shown in Figure 5.37, the deviation of conservation of mass and momentum in Figure 5.38, and the numerical dissipation as well as the divergence of the velocity field in Figure 5.39.



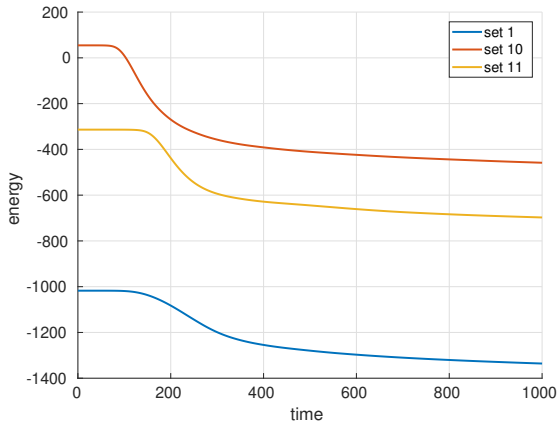
(a) See Figure 5.24



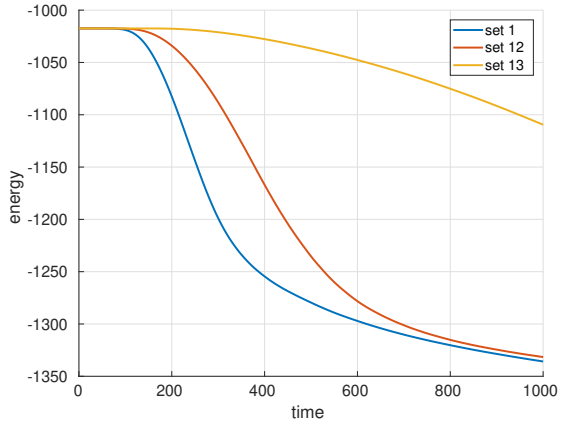
(b) See Figure 5.25



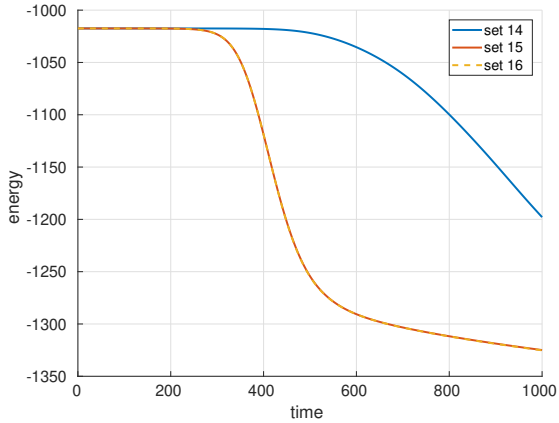
(c) See Figure 5.26



(d) See Figure 5.27



(e) See Figure 5.28



(f) See Figure 5.29

Figure 5.23: **Simplified model:** evolutions of the total energies E_{total}, using parameter sets **1-16** from Table 5.6. Corresponding phase-field evolutions are linked below each plot.

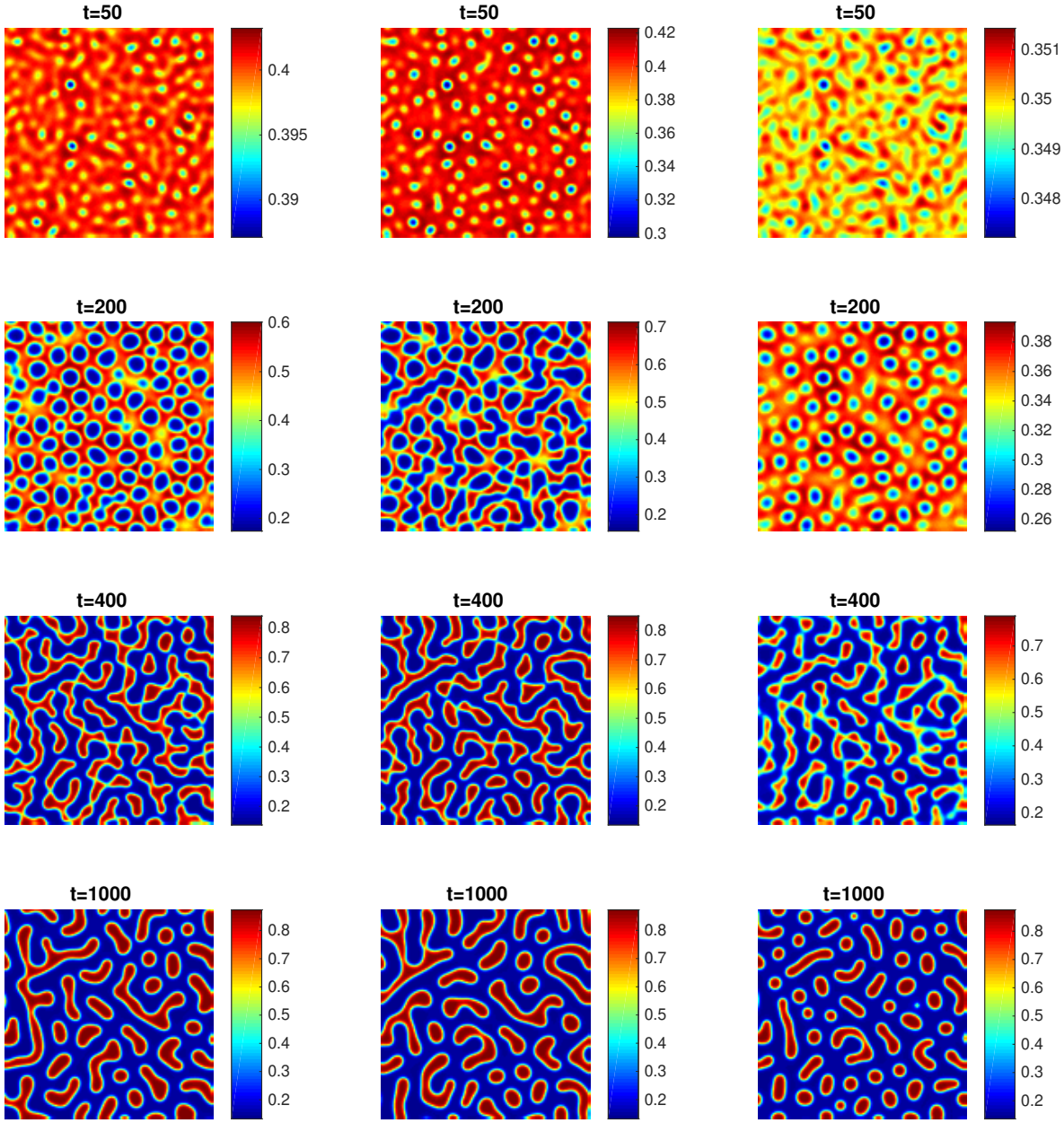


Figure 5.24: **Simplified model:** snapshots of volume fraction φ at different times t (from top to bottom), using the different parameter sets **1, 2** and **3** (from left to right) from Table 5.6.

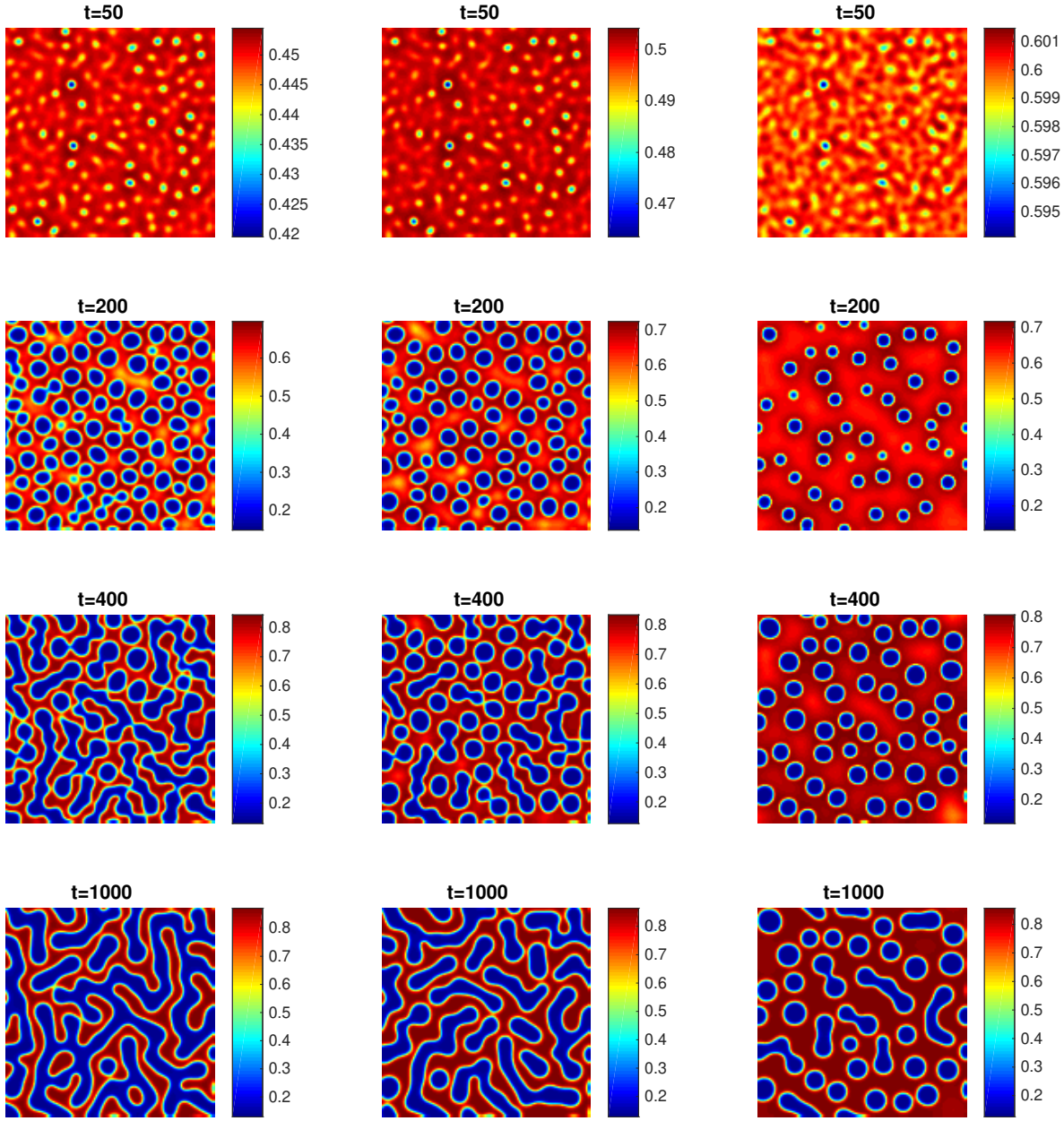


Figure 5.25: **Simplified model:** snapshots of volume fraction φ at different times t (from top to bottom), using the different parameter sets **4, 5** and **6** (from left to right) from Table 5.6.

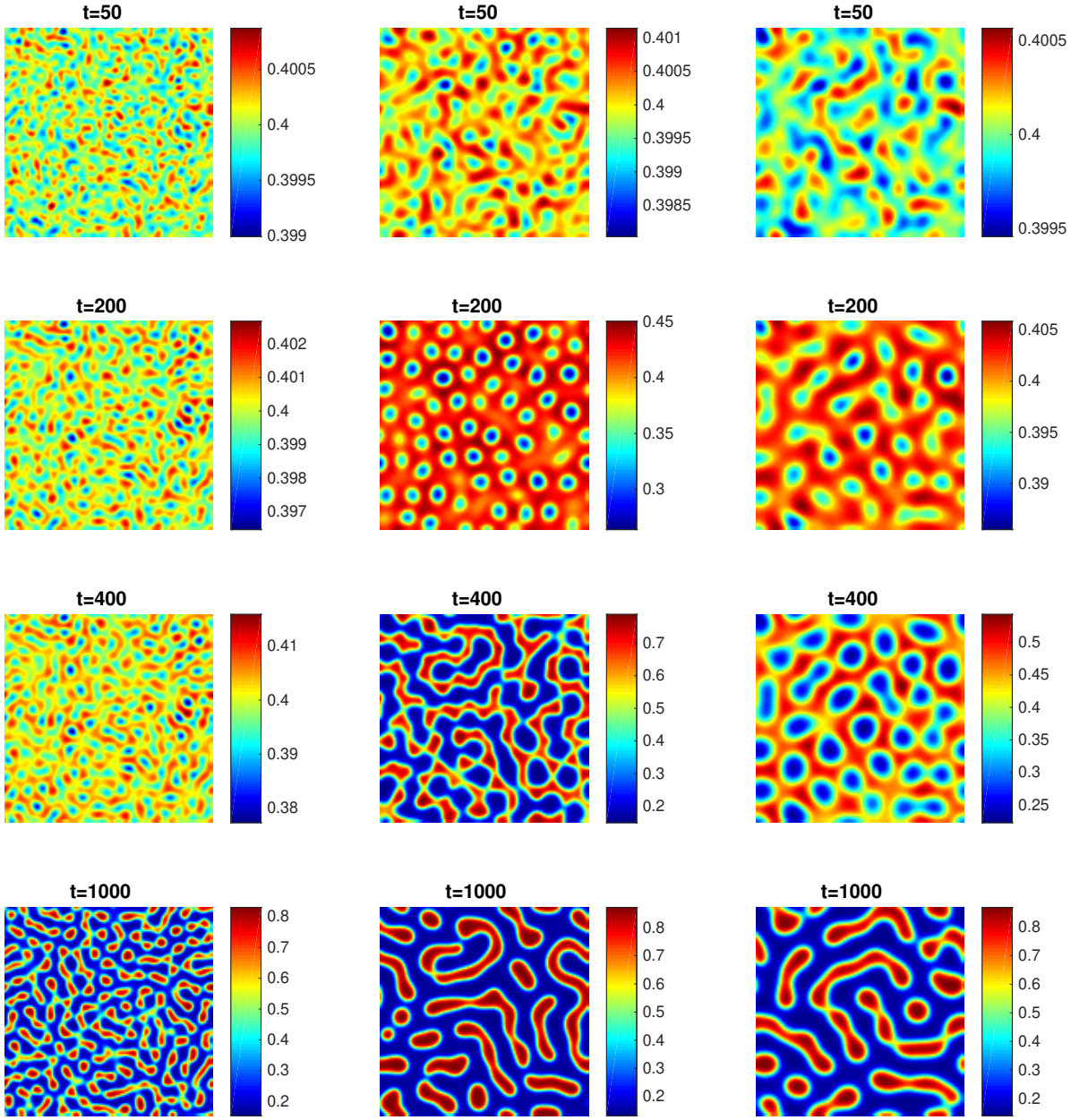


Figure 5.26: **Simplified model:** snapshots of volume fraction φ at different times t (from top to bottom), using the different parameter sets **7**, **8** and **9** (from left to right) from Table 5.6.

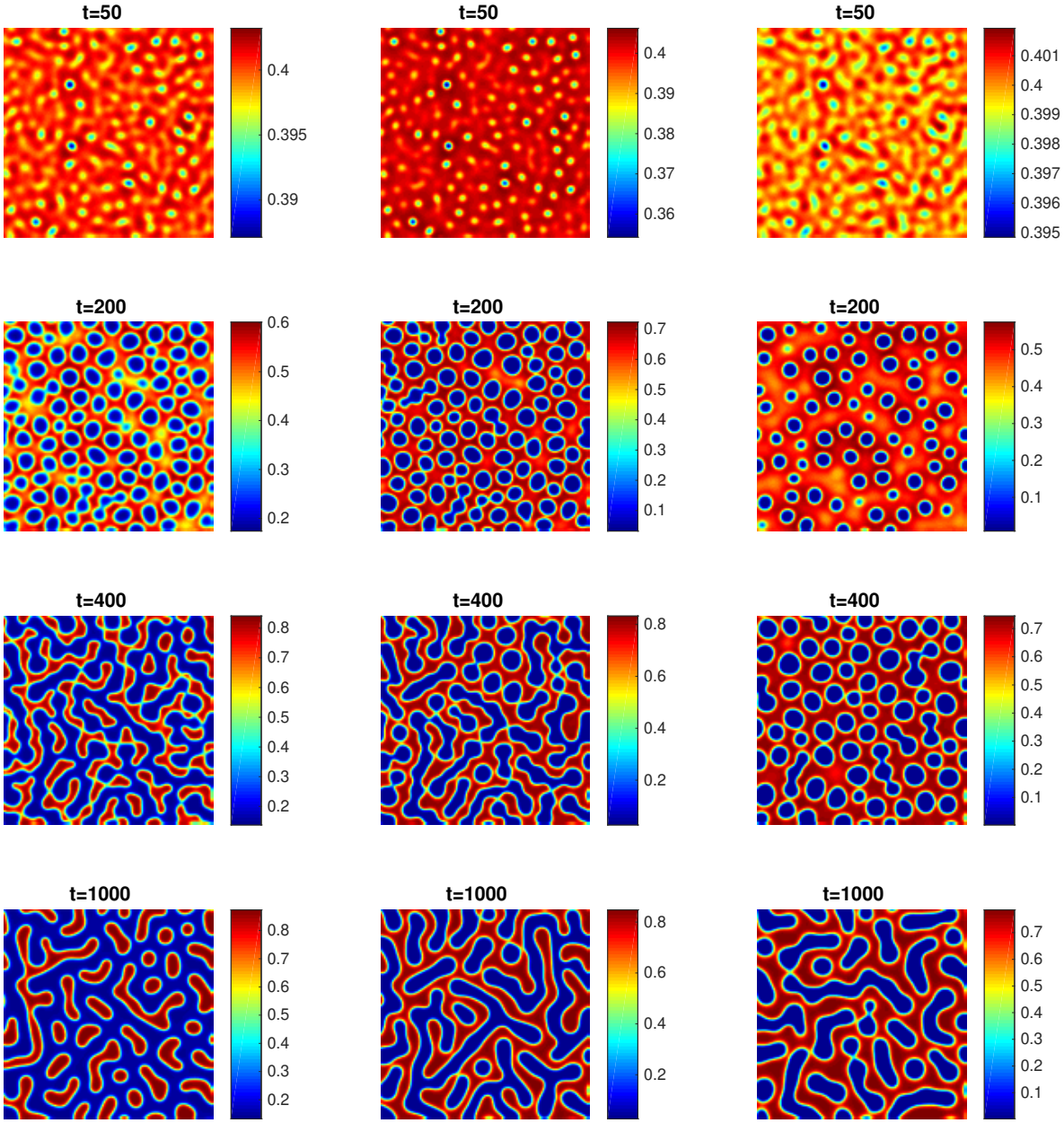


Figure 5.27: **Simplified model:** snapshots of volume fraction φ at different times t (from top to bottom), using the different parameter sets **1, 10** and **11** (from left to right) from Table 5.6.

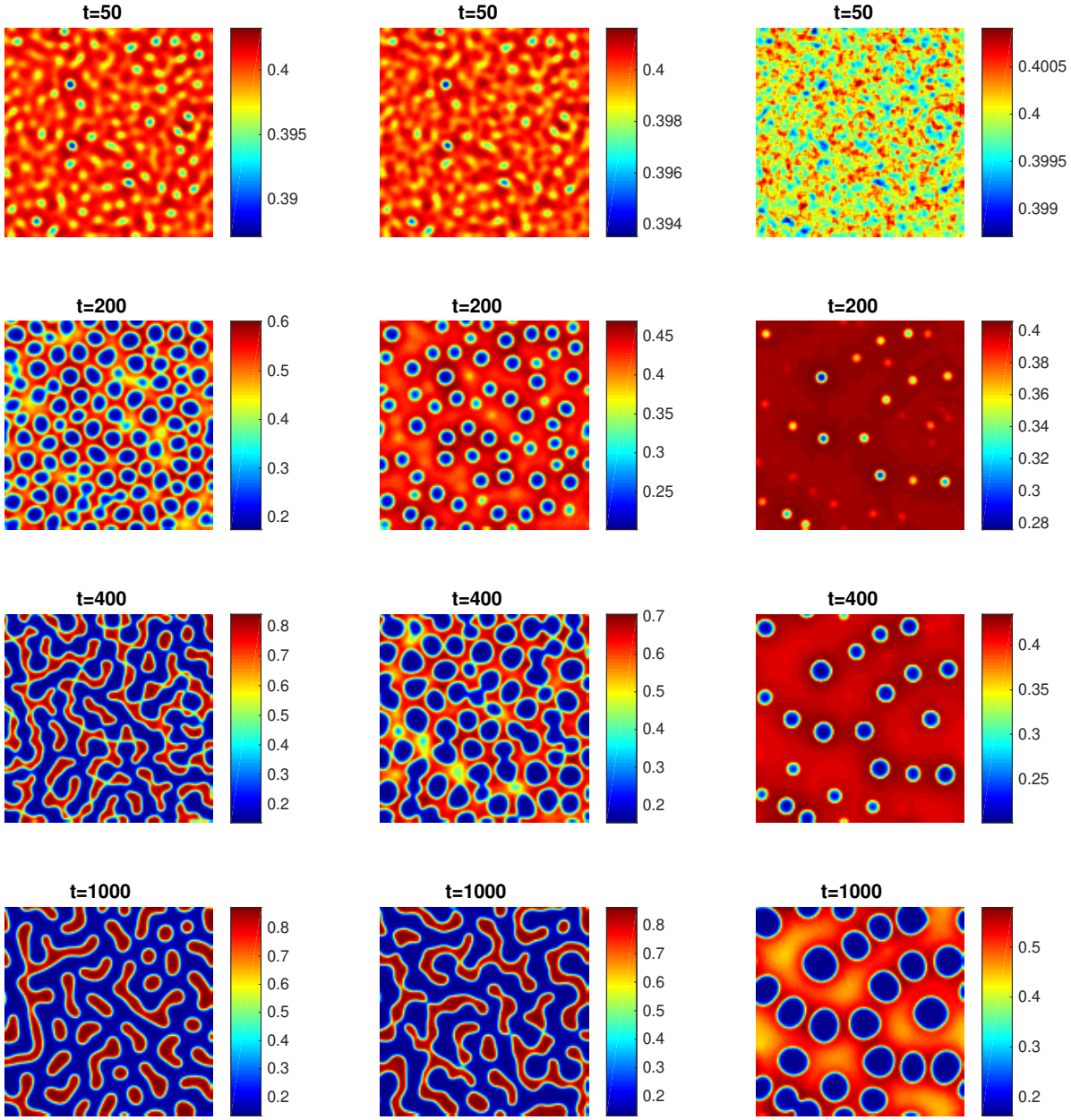


Figure 5.28: **Simplified model:** snapshots of volume fraction φ at different times t (from top to bottom), using the different parameter sets **1, 12** and **13** (from left to right) from Table 5.6.

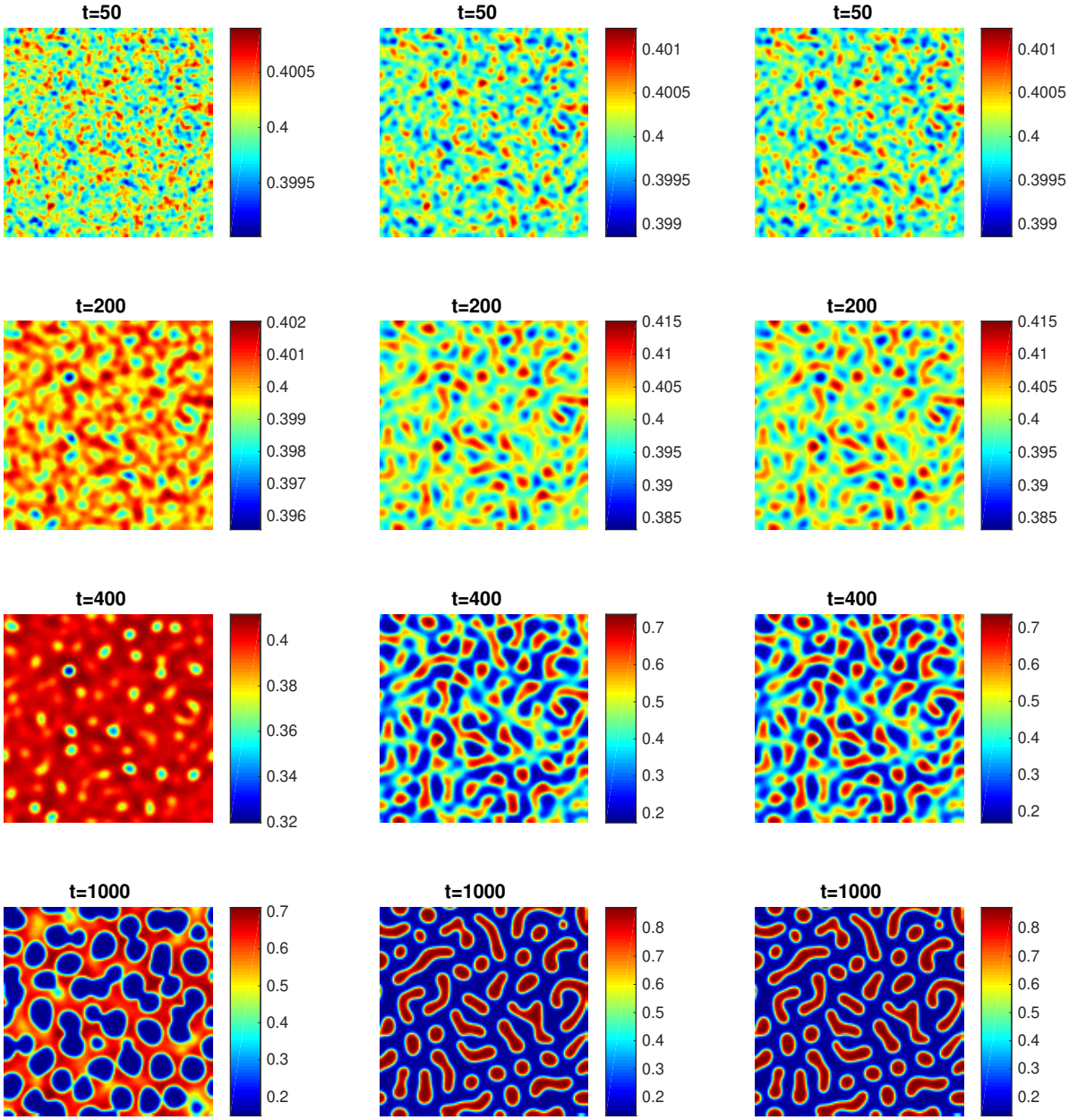


Figure 5.29: **Simplified model:** snapshots of volume fraction φ at different times t (from top to bottom), using the different parameter sets **14, 15** and **16** (from left to right) from Table 5.6.

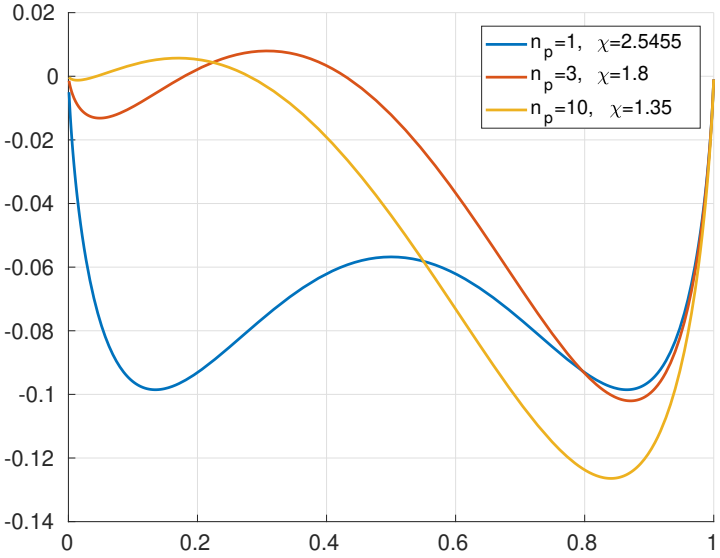


Figure 5.30: Flory-Huggins potential $F(\varphi)$ for different n_p and χ values. See Figure 5.27 for respective phase-field evolutions.

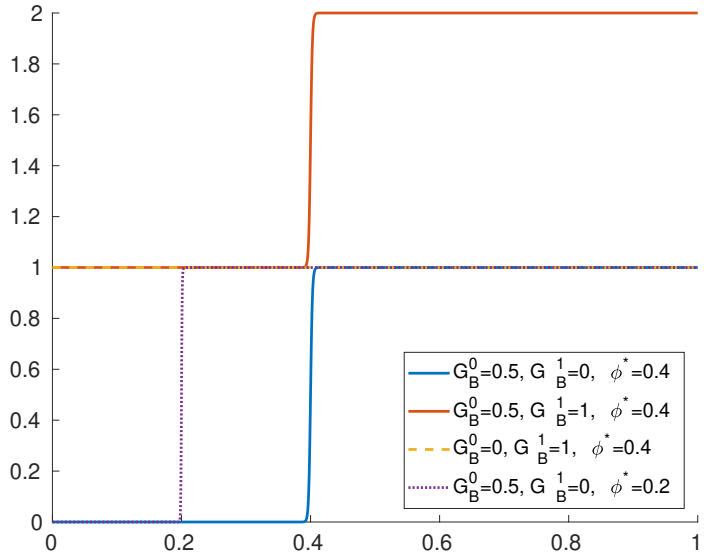


Figure 5.31: Bulk relaxation modulus $G_B(\varphi)$ for different G_B^0, G_B^1 and φ^* values. See Figure 5.29 for respective phase-field evolutions.

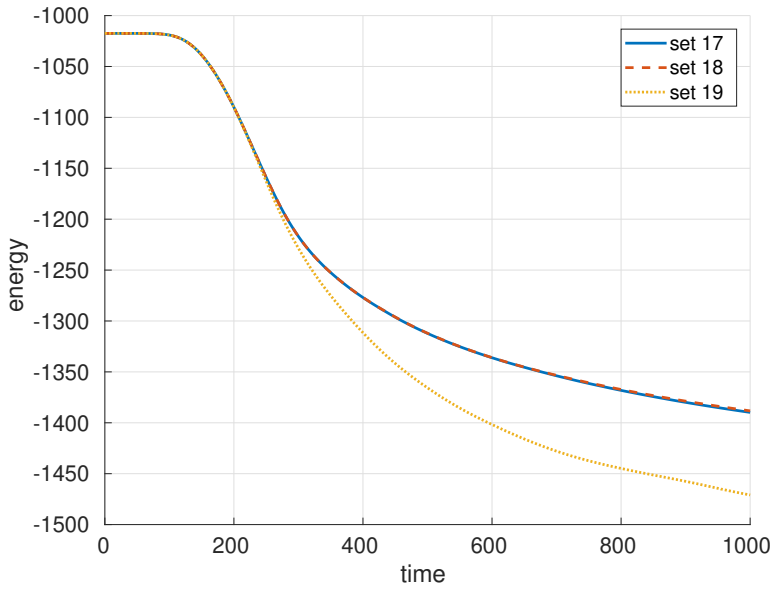


Figure 5.32: **Full model:** evolutions of the total energies E_{total}, using the different parameter sets **17, 18** and **19** from Table 5.6. See Figure 5.33 for corresponding phase-field evolutions.

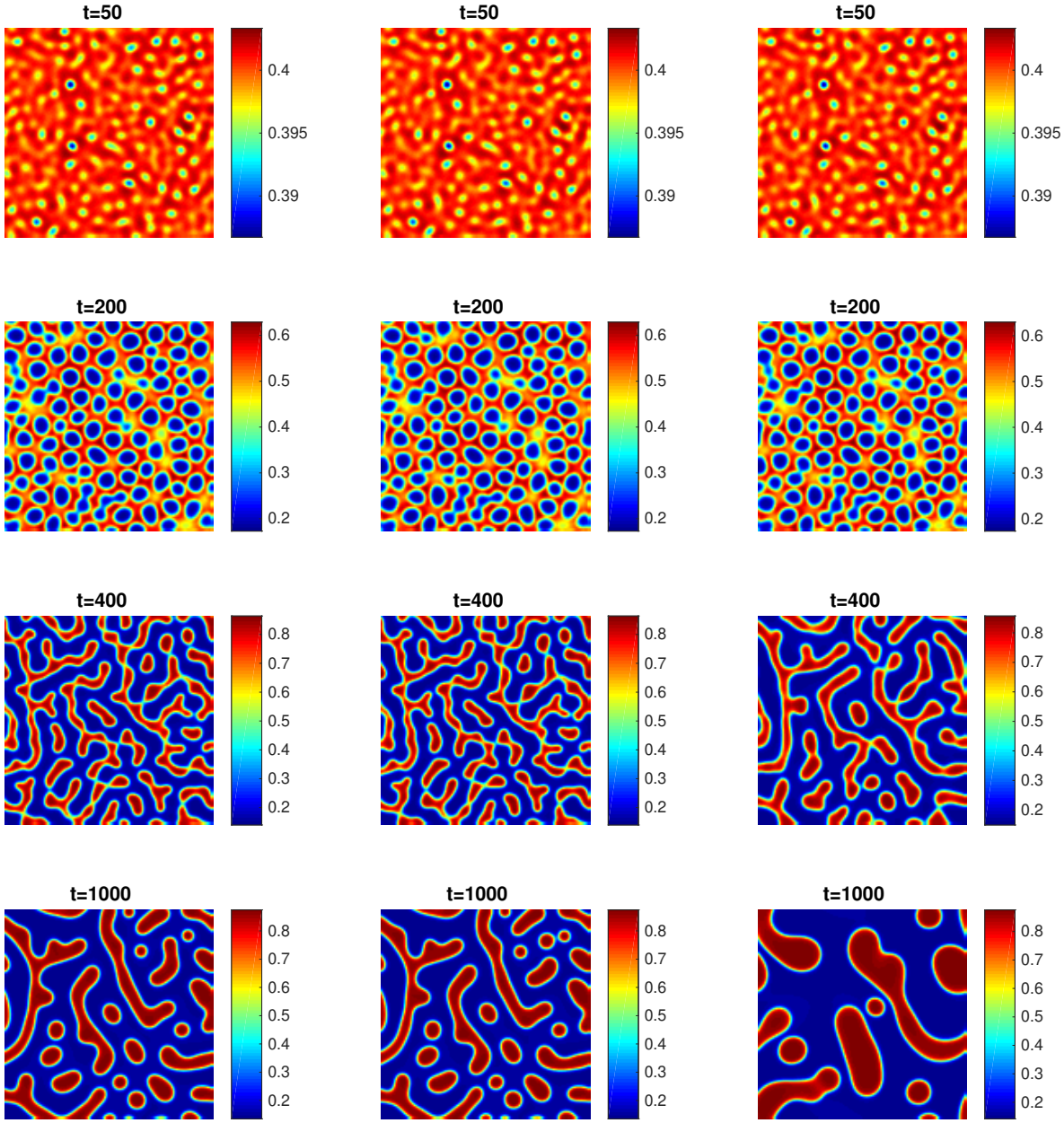


Figure 5.33: **Full model:** snapshots of volume fraction φ at different times t (from top to bottom), using the different parameter sets **17, 18** and **19** (from left to right) from Table 5.6.

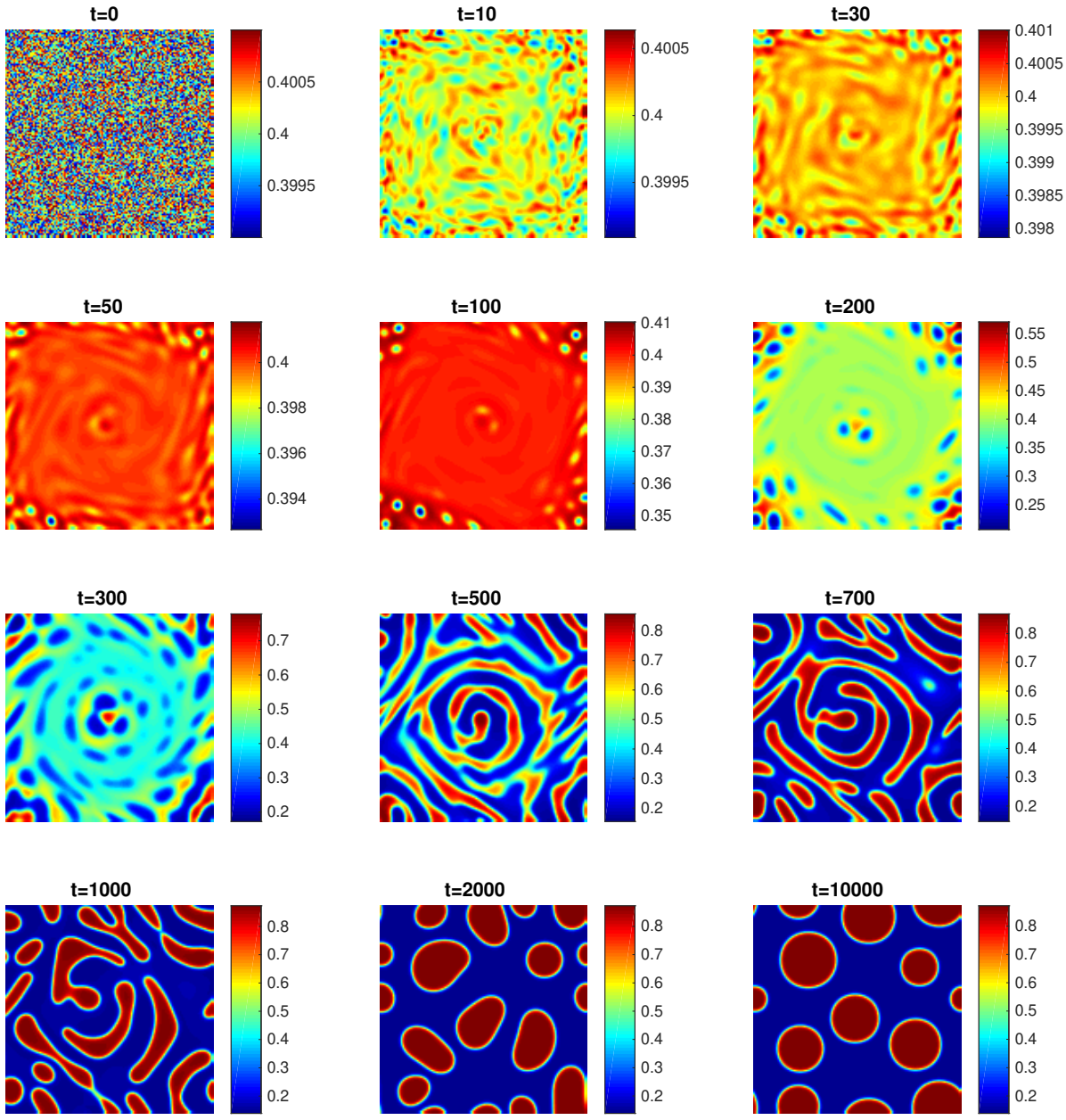


Figure 5.34: **Full model 5.1.5 with vortex** (5.1)**:** snapshots of volume fraction φ at different times t, using $\Delta t = 0.025$.

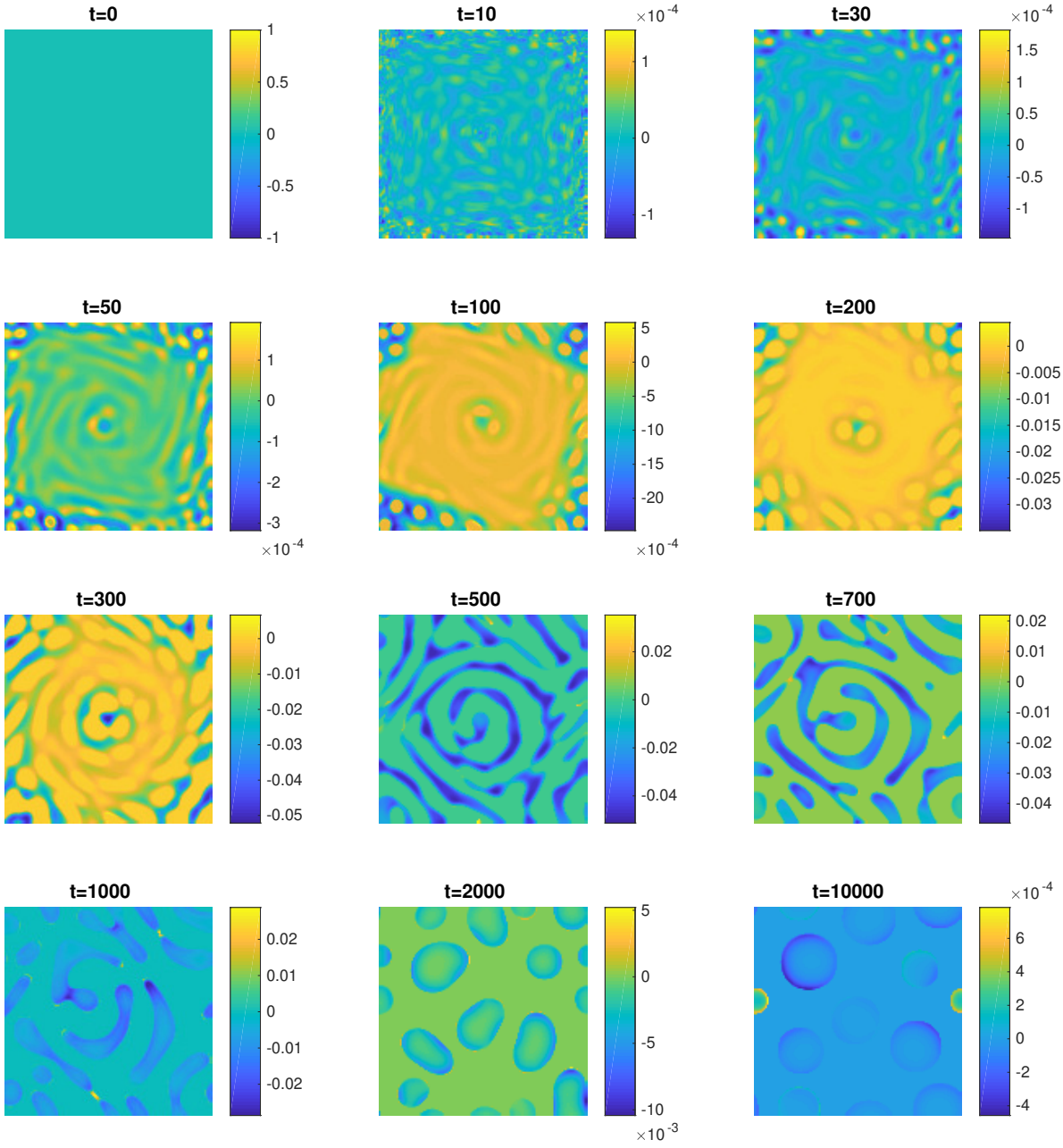


Figure 5.35: **Full model 5.1.5 with vortex** (5.1)**:** snapshots of bulk stress q at different times t, using $\Delta t = 0.025$.

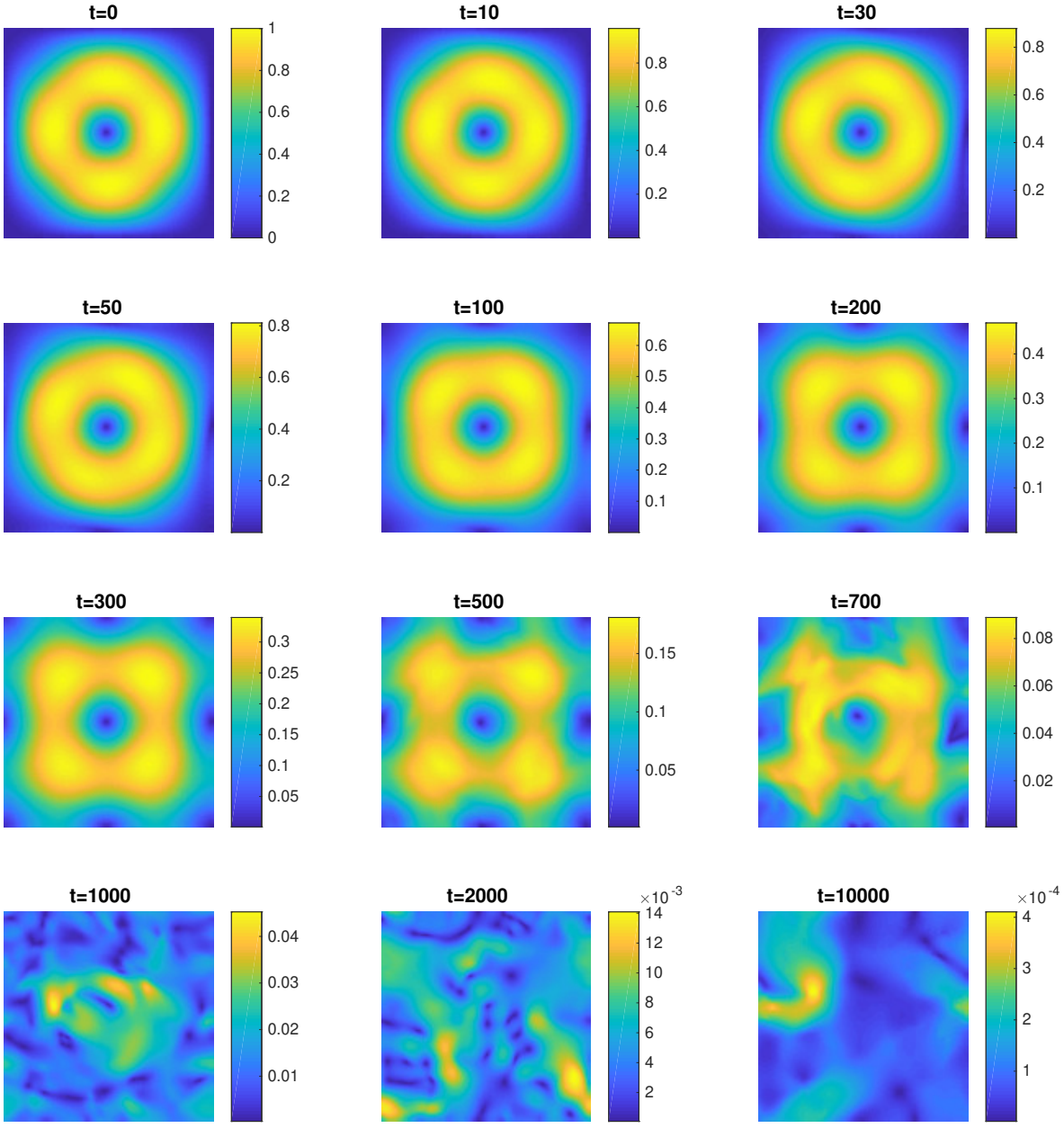


Figure 5.36: **Full model 5.1.5 with vortex (5.1):** snapshots of the Euclidean norm of velocity field $\boldsymbol{u}$ at different times t, using $\Delta t = 0.025$.

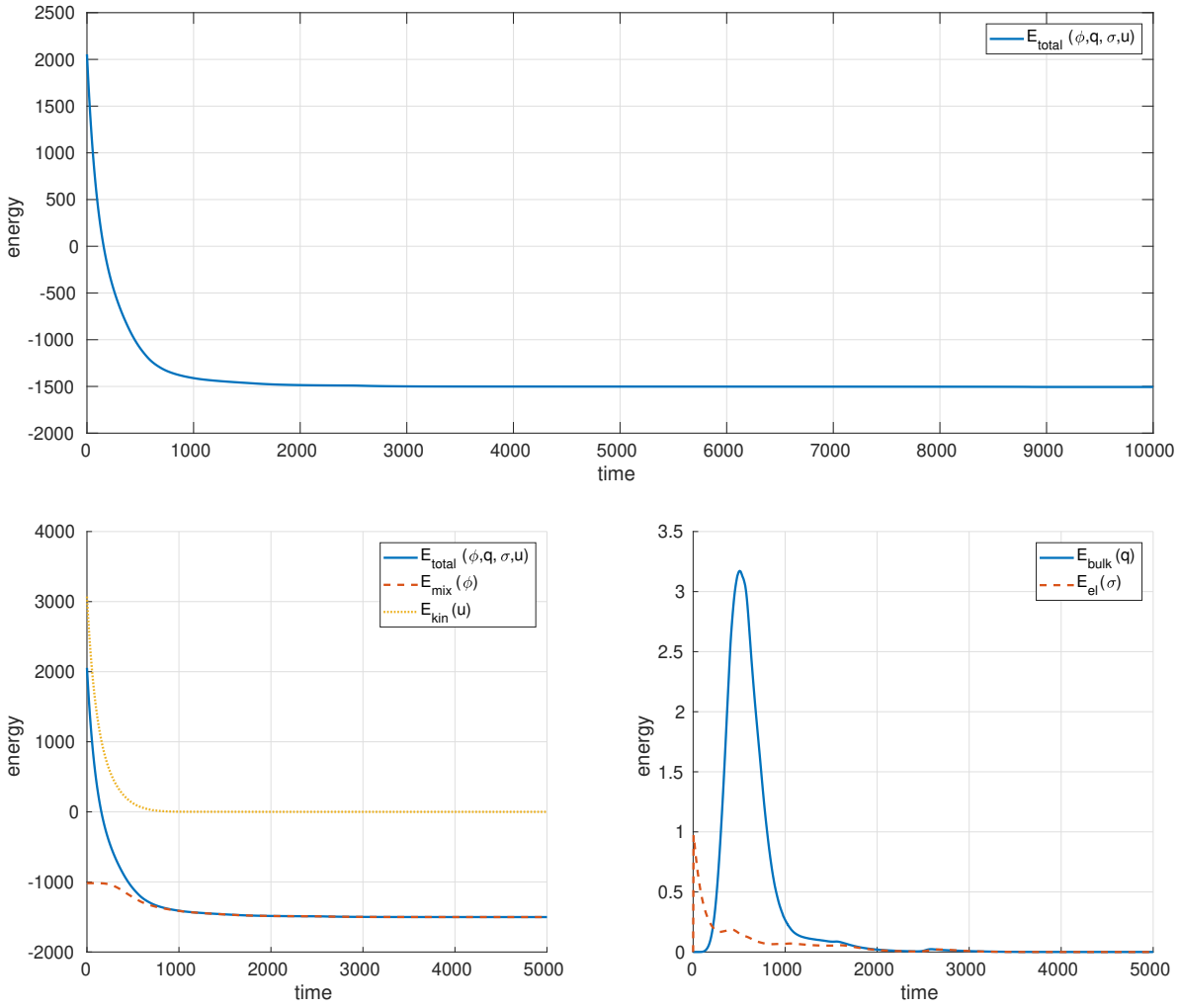


Figure 5.37: **Full model 5.1.5 with vortex** (5.1)**:** energy evolution, using $\Delta t = 0.025$. Plotted until $t = 10^5$ at the top and separately until $t = 5000$ below.

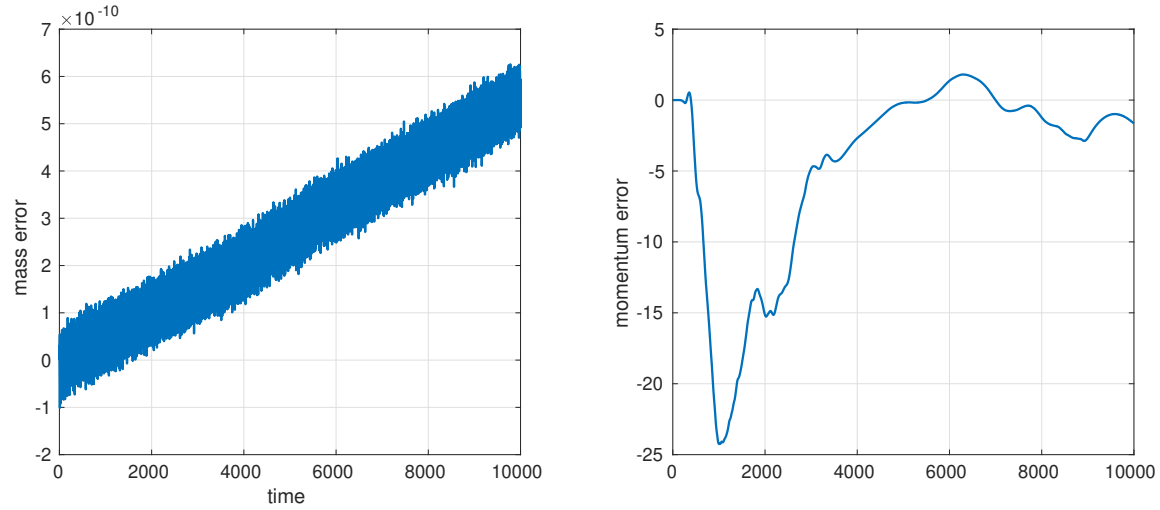


Figure 5.38: **Full model 5.1.5 with vortex (5.1):** deviations from the conservation of mass and momentum, using $\Delta t = 0.025$.

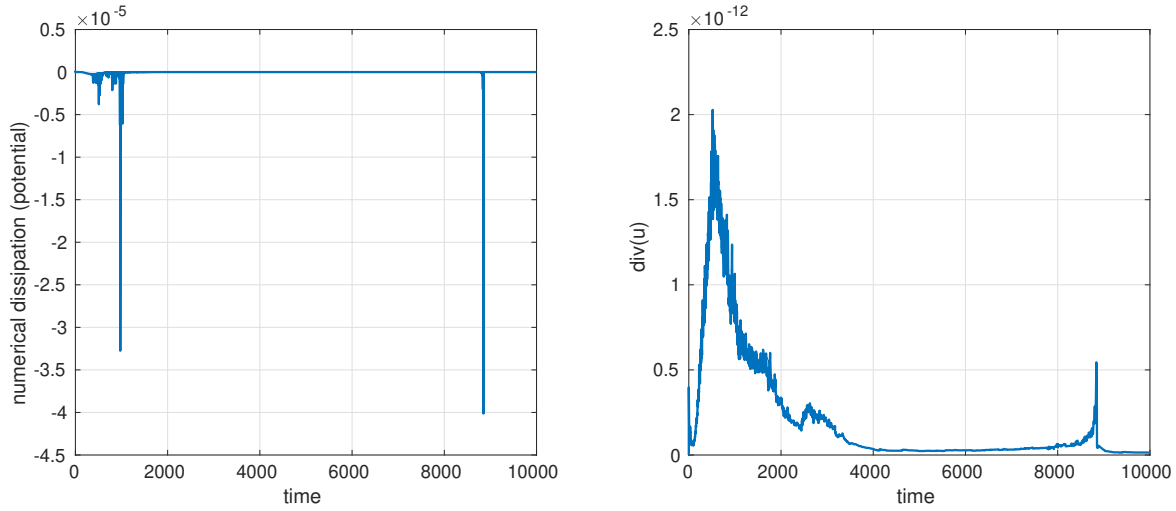


Figure 5.39: **Full model 5.1.5 with vortex (5.1):** numerical dissipation ND_{pot} and the velocity divergence, using $\Delta t = 0.025$.

5.3 Fluid dynamics versus molecular dynamics

As mentioned in the introduction, a crucial aim of this research is to investigate if the full model by *Zhou et al.* [74] is a fully satisfying solution of the underlying problem. For this reason, we also investigate how well this macroscopic system describes the physics, by linking and comparing numerically approximated solutions of it to computer experiments that are based upon a mesoscopic (coarse-grained molecular dynamics) model, which can be considered as physically sound beyond a reasonable doubt.

The starting point of this mesoscopic simulation model is a standard Kremer-Grest model [36], where each polymer chain is represented by sequences of beads. These beads interact via a bonded potential in order to ensure connectivity, and a non-bonded potential to model the excluded volume effect as well as the quality of the solvent. In a good solvent, i.e., a solvent where polymers and solvent mix homogeneously, the latter interaction is simply a purely repulsive Lennard-Jones potential. For details about these potentials, as well as used parameters, see *Grest and Kremer* [36]. The Lennard-Jones potential also defines the unit system of the simulation (each bead has unit mass). The effects of poor solvent quality, i.e., a solvent where polymers and solvent demix, are modeled by adding an attractive tail to the non-bonded pair interaction. For this tail, a suitably fitted cosine wave is taken. For details about the resulting potential, see *Soddemann, Dünweg and Kremer* [65]. This system is simulated by coarse-grained molecular dynamics (MD) and at the same time coupled to a standard D3Q19 Lattice Boltzmann (LB) model, i.e., a three-dimensional LB model that has nineteen degrees of freedom. The LB model represents the momentum transport through the solvent and has the thermodynamic properties of an ideal gas. The coupling of MD and LB is facilitated by a Stokes friction acting on each bead. The dissipative nature of this coupling ensures that it does not alter the thermodynamics of the polymer system. Both MD and LB are supplemented by a Langevin thermal noise such that the temperature is kept constant. For more technical details of this approach as well as the underlying theory, see, e.g., *Dünweg and Ladd* [22].

As mentioned above, the LB model is three-dimensional and likewise are the coarse-grained polymer chains. This is mandatory, since unlike in our phenomenological macroscopic model, a change of dimension substantially changes the (molecular) dynamics. This is easy to imagine, since, e.g., in two space dimensions, two long polymer chains in a bad solvent can hardly pass each other but immediately collide and therefore entangle if they come close. Thus, having a semidilute polymer configuration in 2D, modeled by MD, the demixing time scales are much shorter compared to 3D and the dynamics might differ substantially. Summarized, 2D MD simulations are unlike 3D ones not representative for real world experiments.

Primarily for this reason, we also derived 3D versions of our fully discrete numerical schemes. This facilitates us to map the initial data used for mesoscopic simulations to our staggered grid and use the resulting discrete initial values for our numerical experiments. In detail, we map the polymer bead positions to piecewise constant polymer volume fraction values for each grid cell, resulting in a tensor-valued initial polymer volume fraction φ^0 for the whole grid. Note that for this mapping, we also take into account the constant size of the polymer beads. We also map the velocity vectors of the polymer beads to our staggered grid, using the polymer bead positions for weighting, resulting in three tensor-valued initial

polymer velocities u_p^0, v_p^0, w_p^0. Let us recall that each velocity tensor is defined on a different grid, since we are using a staggered grid. Further, we map the solvent velocity field from the LB lattice to our staggered grid to get the three tensor-valued initial solvent velocities u_s^0, v_s^0, w_s^0. Then, we use the above constructed volume fraction to calculate a volume-averaged velocity field (element-wise) by $\omega_{\text{MD}}^0 = \varphi^0 \omega_p^0 + (1 - \varphi^0)\omega_s^0, \omega \in \{u, v, w\}$, where φ^0 is always mapped to the respective velocity grid. Finally, we remove the divergence from the velocity field, since the MD data is not divergence free. This is realized analogously to the pressure correction method by calculating

$$\begin{aligned} \Delta_{N^3} p &= \nabla_{N^3}^R \cdot \boldsymbol{u}_{\text{MD}}^0 \,, \\ \boldsymbol{u}^0 &= \boldsymbol{u}_{\text{MD}}^0 - \nabla_{N^3}^L p \,, \end{aligned}$$

where the latter is our initial divergence free and volume averaged velocity field.

A feature of many macroscopic models, and also of those that we use, as we will demonstrate in the next section, is that they are able to show similar dynamics in two and three space dimensions. Thus, it is not only reasonable but preferable to perform 2D simulations, since a lower spatial dimension naturally results in significantly less computational effort and is also visualized more clearly. This further widens the performance gap and therefore yields more reasons to use a macroscopic model over a mesoscopic one, if applicable.

But since we want to compare mesoscopic simulation results to macroscopic ones here, a workaround is necessary to use mesoscopic data as a starting point for a two-dimensional macroscopic simulation. One could, e.g., map the three-dimensional mesoscopic initial values to our two-dimensional staggered grid. We used this workaround back then, when only a macroscopic 2D simulation code was developed, see *Lukáčová-Medviďová, Dünweg, Strasser and Tretyakov* [56] for respective simulation results. Since this is rather inaccurate, our collaboration partners developed a variant of the MD LB model, which uses the third dimension to a minimum while still maintaining the typical 3D MD dynamics of the polymer solvent demixing, see *Spiller* [66] for details. In this pseudo 2D MD LB model, the third dimension is reduced to a few grid cells while most polymer beads are bonded to a plane with the thickness of one grid cell. Therefore, we can map the MD LB initial data of this plane to our two-dimensional staggered grid analogously to above and run our macroscopic 2D simulations using the resulting data as initial values.

Since two-dimensional phase-fields are perceived more clearly than three-dimensional visualizations, we will only use simulation results of the pseudo 2D MD LB method in the following for the comparison to our macroscopic simulations. This coarse-grained MD LB simulation results, see Figure 5.40, were thankfully obtained by *Dominic Spiller* and are also used in his thesis, see [66, Figure 5.9]. The system consists of 1024 polymer chains with 128 beads, each, and a $512 \times 512 \times 4$ Lattice Boltzman grid, which is at the same time the computational domain size ($[0, 512]^2 \times [0, 4]$).

In Figure 5.41, we map the initial polymer configuration and velocity field of the MD LB simulation as described above to two staggered grids of different sizes and perform simulations using the fully discretized algorithm for the full model from Subsection 4.8.14 with parameter set 1 from Table 5.6. In the top center of the figure, we map the MD LB initial data to a staggered 512×512 grid, since the Lattice Boltzman grid has the same dimensions. In the top

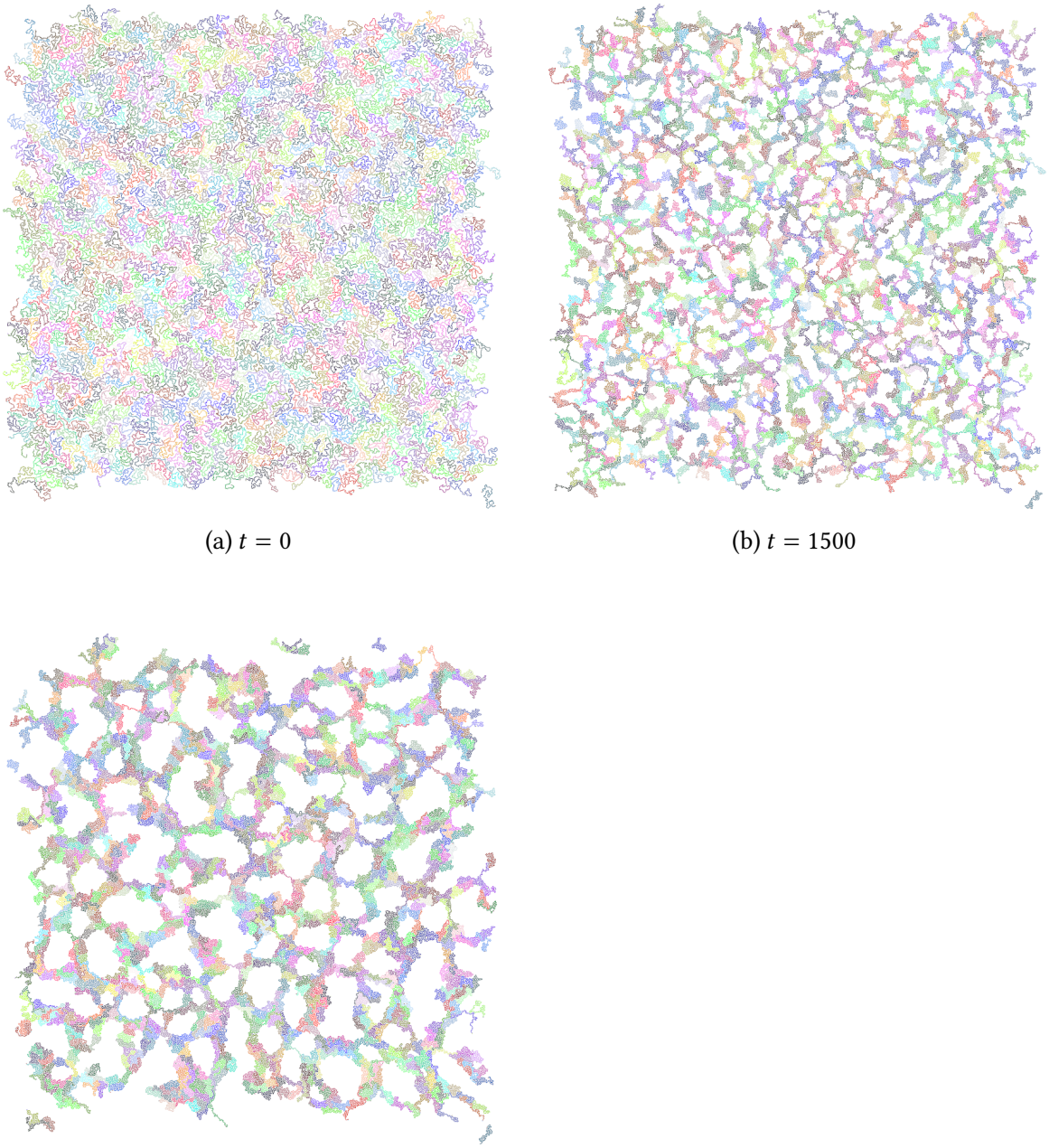
right, we map the MD LB initial data to a staggered 128×128 grid to see if a reduction to our usual simulation grid size, and therefore also a significant reduction in computational costs, leads to any major disadvantages. Even though the MD LB simulation provides another level of detail, we can see that not only are our simulations able to show similar dynamics but they even do so for the significantly reduced grid size. The snapshots of the polymer configurations gained from the MD LB simulation at times 1500 and 87500 fit at least by eye metric very well to the snapshots of the macroscopic volume fraction at times 10 and 80 for both the 512×512 and the 128×128 grid. Note that there is no frozen period and no real volume shrinking of the polymer-rich phase in the macroscopic simulations of this experiment. This is caused by the initially already high volume fraction differences between the polymer-richer and solvent-richer structures.

For a second comparison, see Figure 5.42, we map the same initial MD LB data as above to our staggered grids, but this time we use the initial polymer configuration just as noise for our initial volume fraction. More precisely, we use the initial volume fraction

$$\varphi^0 = 0.4 + 0.002 \left(\varphi^0_{\mathrm{MD}} - 0.4 \right) ,$$

where φ^0_{MD} is the mapped initial volume fraction of the above experiment. This yields much smoother initial data for the volume fraction. Note that there is no major difference between this initial volume fraction and the initial volume fraction with noise that we used in the last section. Thus, we see a frozen period as well as volume shrinking of the polymer-rich phase in this experiment, unlike in the one above. As a consequence, network-like structures are formed significantly later than in the experiment above, yielding a different time scale. Despite this, our simulations again show similar dynamics to the MD LB simulation for both the fine and the coarse grid.

Although the mapping considering only initial conditions is rather crude, it is nevertheless clear that the two systems evolve at least somewhat similarly. While the mass is already scaled between the two models by matching the polymer volume fractions, an energy scaling could also be established by matching the Flory parameters. Further, a detailed comparison of time scales has not yet been accomplished. In order to match the time scales, the effective viscosity of the macroscopic model has to be determined. This analysis is left for future work.



(a) $t = 0$

(b) $t = 1500$

(c) $t = 87500$

Figure 5.40: **Pseudo 2D MD LB simulation:** snapshots of the polymer configuration at different times t, using a $512 \times 512 \times 4$ LB grid. This simulation results were thankfully obtained by *Dominic Spiller* and are also used in his thesis, see [66, Figure 5.9].

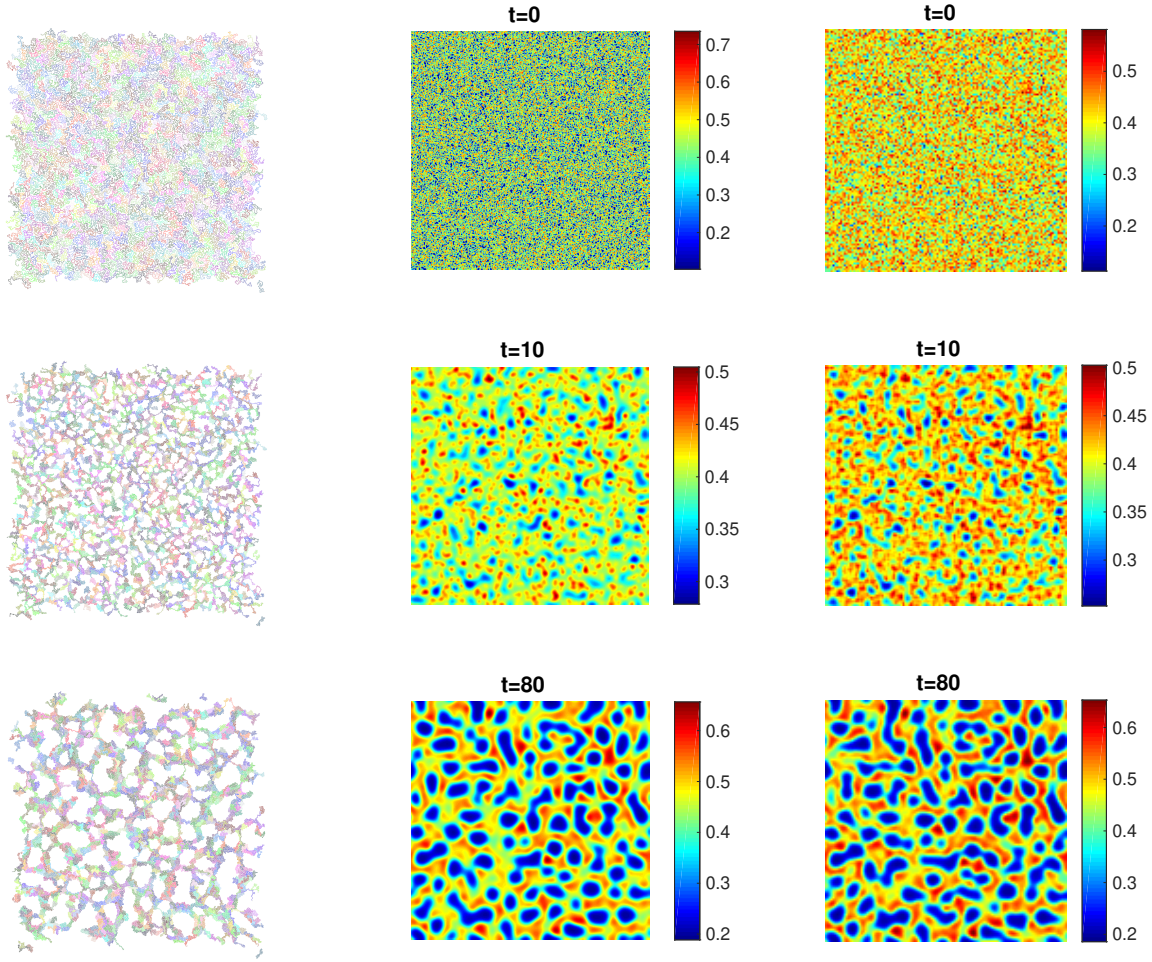


Figure 5.41: Snapshots of the polymer configuration from a MD LB simulation using a $512 \times 512 \times 4$ LB grid (left column, $t = 0, 1500, 87500$, see Figure 5.40) versus snapshots of volume fraction φ, simulated using our fully discretized algorithm for the full model from Subsection 4.8.14 with parameter set 1 from Table 5.6. Here, we mapped the initial MD LB data (top left) to a staggered 512×512 grid (top center) and a staggered 128×128 grid (top right) as initial φ and $\boldsymbol{u}$ values for the respective macroscopic simulation results below.

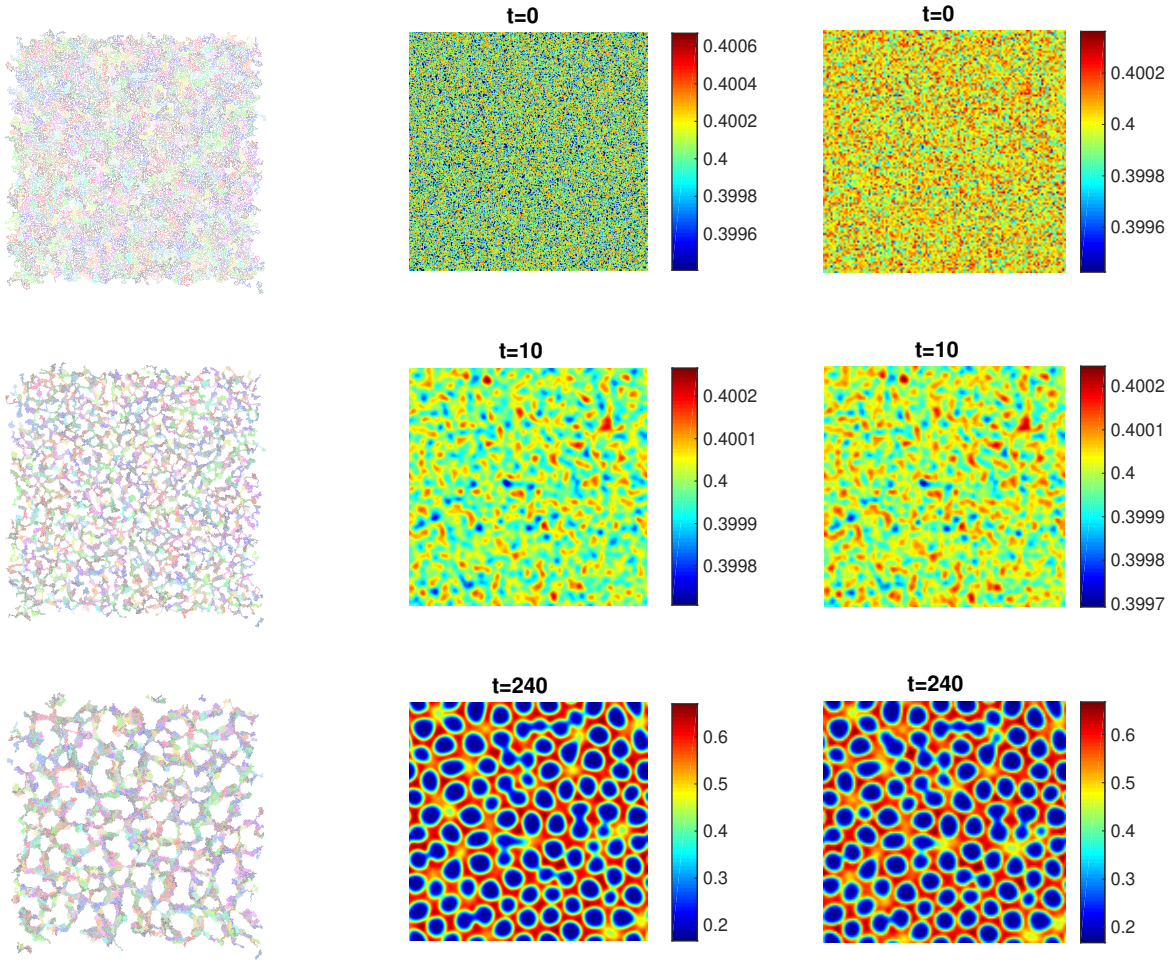


Figure 5.42: Snapshots of the polymer configuration from a MD LB simulation using a $512 \times 512 \times 4$ LB grid (left column, $t = 0, 1500, 87500$, see Figure 5.40) versus snapshots of volume fraction φ, simulated using the fully discretized algorithm for the full model from Subsection 4.8.14 with parameter set 1 from Table 5.6. Here, we mapped the initial MD LB data (top left) to a staggered 512×512 grid (top center) and a staggered 128×128 grid (top right) as initial $\boldsymbol{u}$ values and **as noise** for the initial φ values for the respective macroscopic simulation results below.

5.4 3D simulations

Due to the immense computational costs and random access memory demands of 3D simulations using small time steps and fine grids, we only consider the following model and test case to confirm experimental convergence in 3D. Thereafter, we consider a second model and test case to verify the applicability of our schemes to demonstrate viscoelastic phase separation in 3D.

5.4.1 Cahn-Hilliard equation in 3D (EOC in space and time)

We solve the full discretization (4.130) of the Cahn-Hilliard equation in three space dimensions with the following functions and parameters:

- $f(\varphi)$ is the derivative of the Ginzburg-Landau potential (3.9),
- the interface constant reads $\lambda = 0.01$, and
- the constant mobility $M(\varphi) = 1$.

We use the following smooth initial conditions

$$\varphi^0 := \varphi(x, y, z, 0) = 0.05 \sin x \sin y \sin z + 0.001\,, \qquad (x, y, z) \in [0, 2\pi]^3\,,$$

and compute the numerical solutions $\varphi_{N,\Delta t}$ up to the final time $t = 1$. Our reference solution reads $\varphi_{ref} = \varphi_{256,10^{-3}}$, i.e., it is calculated on a 256^3 grid with time step size $\Delta t = 10^{-3}$. The results displayed in Table 5.7 confirm our claimed second order of accuracy in time and space.

Table 5.7: **Cahn-Hilliard equation in 3D:** L^1-errors and experimental convergence rates in space and time at $t = 1$.

N	Δt	$e(\varphi_{N,\Delta t})$	$\mathrm{EOC}(\varphi_{N,\Delta t})$
32	$8 \cdot 10^{-3}$	37.413	–
64	$4 \cdot 10^{-3}$	6.1201	2.6119
128	$2 \cdot 10^{-3}$	1.2445	2.298

5.4.2 Full model in 3D

We solve the fully discretized algorithm for the full model from Subsection 4.8.14. Further, we use the following functions and parameters:

- $f(\varphi)$ is the derivative of the Flory-Huggins potential (3.11) with $n_p = n_s = 1$ and $\chi = 2.8/1.1$,
- the interface constant reads $\lambda = 1$,
- the constant friction $\zeta(\varphi) = 0.1$,
- the relaxation time $\tau_B(\varphi) = \tau_B^0 \varphi^2$ with $\tau_B^0 = 10$,
- the bulk relaxation modulus

$$G_B(\varphi) = G_B^0 \left[1 + \tanh\left(\frac{\cot(\pi\varphi^*) - \cot(\pi\varphi)}{\varepsilon}\right)\right] + G_B^1\,,$$

where $G_B^0 = 0.5$, $G_B^1 = 0$, $\varphi^* = 0.4$ and $\varepsilon = 0.01$,

- the shear relaxation time $\tau_S(\varphi) = \tau_S^0 \varphi^2$ with $\tau_S^0 = 5$,
- the shear relaxation modulus $G_S(\varphi) = G_S^0 \varphi^2$ with $G_S^0 = 0.5$, and
- the constant viscosity $\eta(\varphi) = 1$.

We consider the following initial conditions

$$\varphi^0 = 0.4 + \delta(x, y, z), \qquad q^0 = 0, \qquad \boldsymbol{\sigma}^0 = \mathbf{0}, \qquad \boldsymbol{u}^0 = \mathbf{0}, \qquad (x, y, z) \in [0, 128]^3,$$

where $\delta(x, y, z)$ is a uniformly distributed random variable with range $[-0.001, 0.001]$. We compute up to the final time $t = 1000$, using the time step size $\Delta t = 0.1$ and a 128^3 grid.

In Figure 5.43, we observe the phase separation process in 3D, visualized by 11 isosurfaces of values equally distributed between and including the minimum and maximum of the volume fraction, at different times t. The dynamics are as desired similar to those of the 2D simulations, including the four essential features: frozen period, volume shrinking of the polymer-rich phase, a resulting network-like structure and the phase inversion. And finally, also the typical long-time dynamics of the Cahn-Hilliard equation: droplet formation and coagulation.

Further, in Figure 5.44, we observe the same phase separation process in 3D, visualized by 11 isosurfaces of values equally distributed between and including the bounds of the whole solvent-polymer interval $[0, 1]$, at different times t.

In Figure 5.45, we see the energy evolution, which is again always decreasing for the mixing energy, and initially increasing for the bulk, elastic and kinetic energy. The latter is caused by the coupling of the equations and the initial stress as well as velocity values being zero. Since the decrease of the mixing energy is again dominant, the total energy is as desired always decreasing as well.

Further, we see the insignificant deviation of the conservation of mass and the again noticeable deviation of the conservation of momentum in Figure 5.46. Note that here in 3D, the deviation scales with the even larger domain size of 128^3. Thus, the domain size cleaned deviations are significantly smaller than the plotted values. And finally, we have the sufficiently small numerical dissipation and the negligible discrete divergence of the velocity field in Figure 5.47.

Summarized, the viscoelastic phase separation dynamics as well as all key observables behave utterly similar in two and three space dimensions.

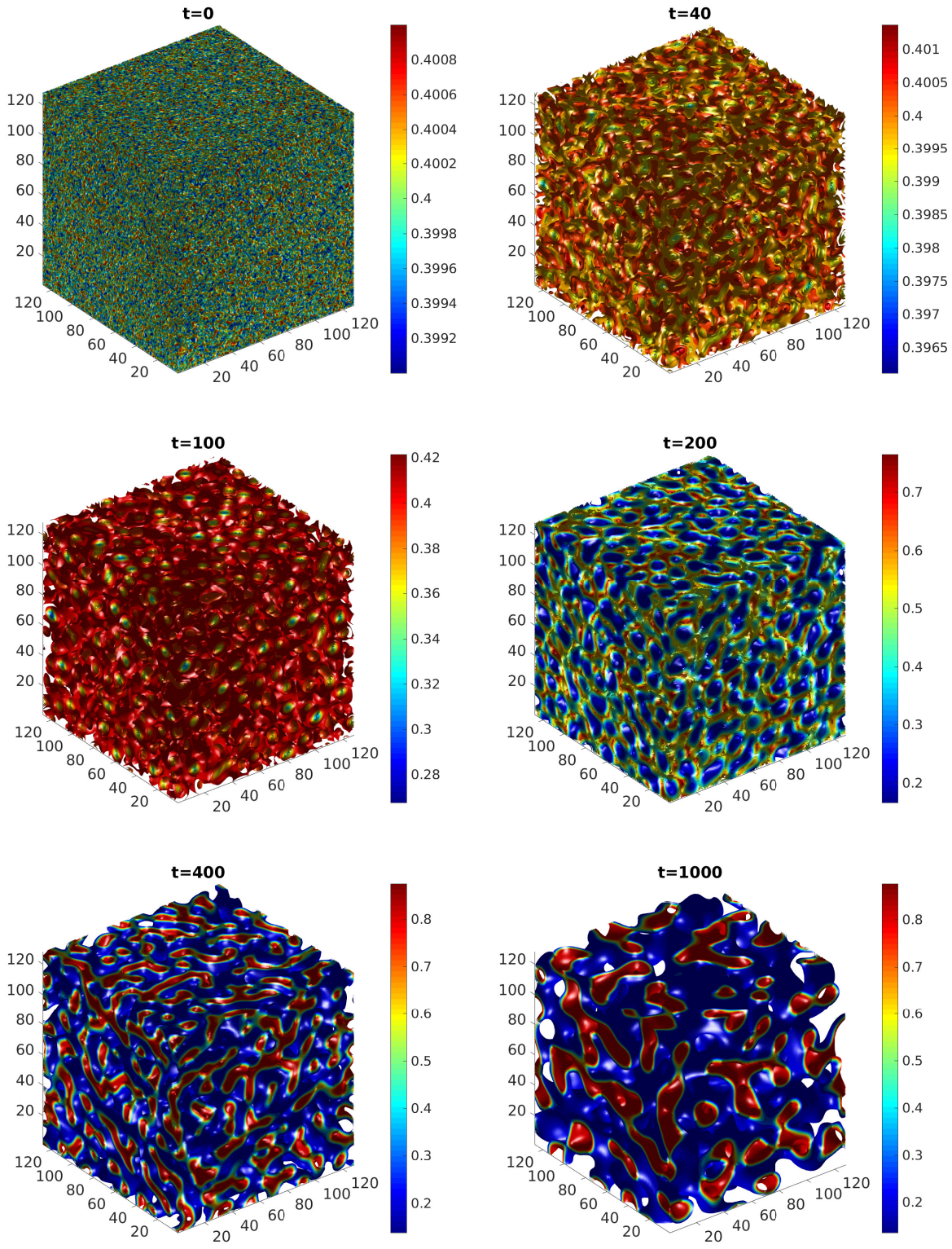


Figure 5.43: **Full model in 3D** 5.4.2: isosurfaces of volume fraction φ at different times t.

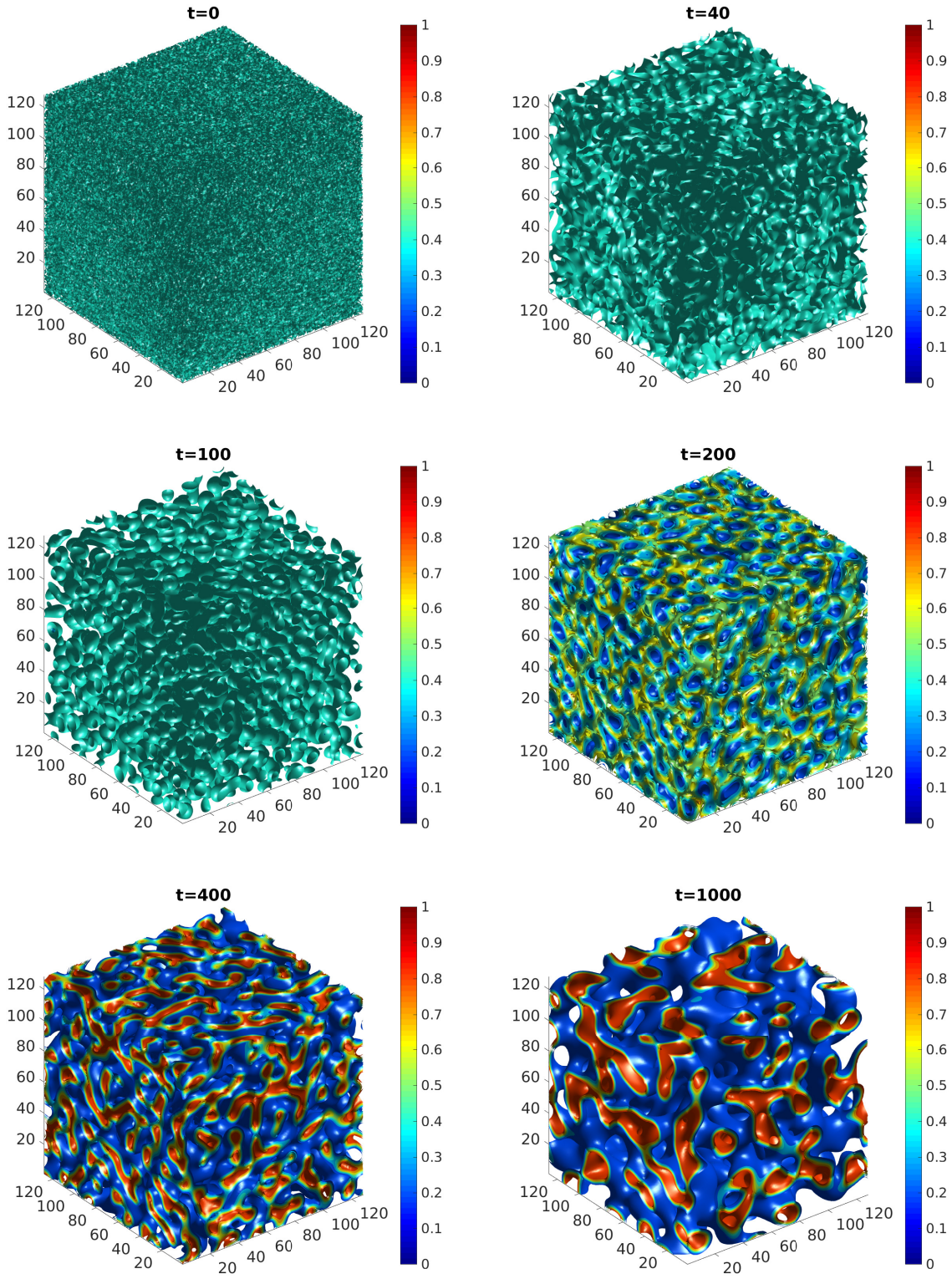


Figure 5.44: **Full model in 3D** 5.4.2: isosurfaces of volume fraction φ at different times t. Here, the color bar is fixed to the whole solvent-polymer interval $[0, 1]$.

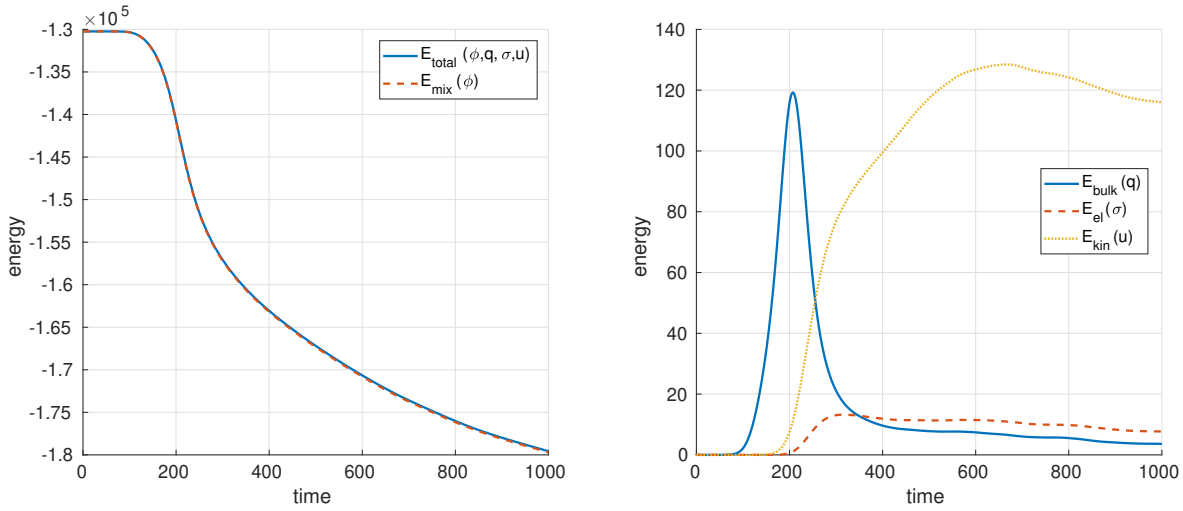


Figure 5.45: **Full model in 3D** 5.4.2: time evolution of the energy.

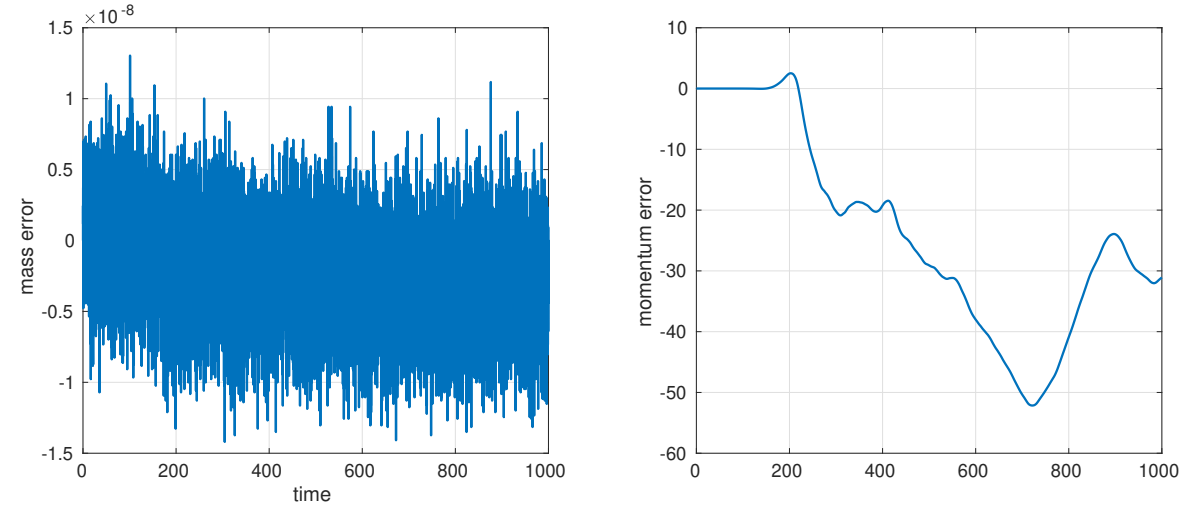


Figure 5.46: **Full model in 3D** 5.4.2: deviations from the conservation of mass and momentum.

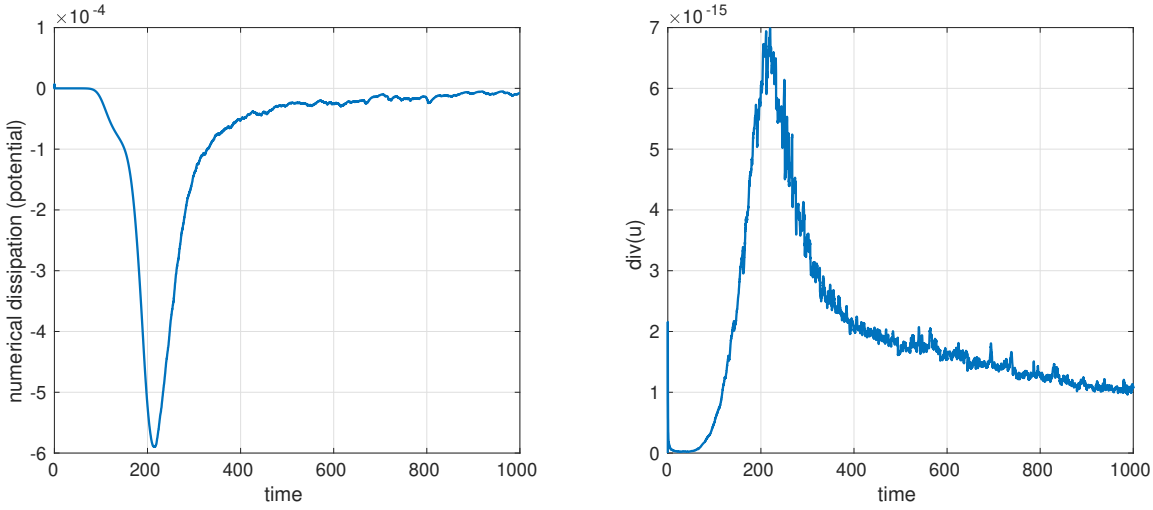


Figure 5.47: **Full model in 3D** 5.4.2: numerical dissipation ND_{pot} and the velocity divergence.

6 Conclusion

In this thesis, we have studied the phase separation of polymer-solvent mixtures by the Cahn-Hilliard equation, describing dynamics of a diffusive interface separating polymer and solvent phase, coupled to extended Oldroyd-B equations for the viscoelastic flow of the mixture. We also considered a significant simplification of this model, which is achieved by neglecting hydrodynamics. Both models have been proposed in [74], where the full model is derived through the variational principle as a minimizer of a total free energy. Consequently, this model is thermodynamically consistent because it satisfies the second law of thermodynamics through having a non-increasing free energy over time while being isothermal. The authors also present simulation results which confirm the capabilities of both models to reproduce all essential features of viscoelastic phase separation observed experimentally in the lab, see [68], reading: an initial frozen period, a volume shrinking of the polymer-rich phase, a resulting network-like structure and phase inversion. These simulation results are based on a space discretization by finite volumes and a time discretization by the explicit Euler method. However, we have shown in Theorem 4.4 that this explicit solver introduces positive terms to the time derivative of the free energy on a discrete level. Thus, one can not validate that the thermodynamic consistency of the model equations is conserved.

For this reason, we have developed and tested more suitable linear numerical schemes. We have demonstrated experimentally and proven theoretically up to the small numerical dissipation term of the double-well potential ND_{pot} that our schemes conserve the thermodynamic consistency of both models while still being utterly computationally efficient. Further, we have proven that our novel schemes also satisfy the conservation of mass of the original phase-field models.

Concerning two component phase-field models in general, we have studied several mathematical models in Chapter 3 and discussed their thermodynamic consistency to gain an elaborated background for the finally introduced full model for viscoelastic phase separation (3.47) by *Zhou et al.* [74] and its simplification (3.48). We have verified that the original viscoelastic phase-field models, which motivated the derivation of the full model (3.47), are not necessarily thermodynamically consistent. Further, we introduced a conformation tensor formulation of the full model to ensure that all terms in the time derivative of the free energy of the full model are non-increasing over time and that the full model is consequently indeed thermodynamically consistent.

In the subsequent Chapter 4, we initially derived problem-specific semi-discretizations in

time for sub-models of the full model and finally for the full model itself by, e.g., scheme (4.77), (4.85), (4.79). Here, we paid special attention to semi-discretizations of the double-well potential to ensure that our schemes are thermodynamically consistent. We focused on linear schemes and also introduced some splitting methods and Chorin's projection method in order to provide not only well-suited but also utterly computationally efficient results. Thereafter, we derived suitable efficient spatial discretizations in Section 4.8, based on finite differences and finite volumes for both two and three space dimensions. Here, the latter was primarily motivated by our aim of comparing our simulation results to usually three-dimensional molecular dynamics simulation results provided by our collaborators from the Max Planck Institute for Polymer Research. We then combined the finite difference discretizations in space with our earlier derived semi-discretizations in time to provide the full discretizations (4.130) of the Cahn-Hilliard equation and (4.133) of the simplified model, for which we claimed a second order of accuracy in both time and space. And finally, by the addition of finite volume methods for the spatial discretization of advection and convection terms, we provided the fully discretized algorithm (4.196), (4.197) for the Cahn-Hilliard-Navier-Stokes model, and in Subsection 4.8.14 the fully discretized algorithm for the full model for viscoelastic phase separation, for which we claimed a first order of accuracy.

At the very beginning of Chapter 5, we experimentally confirmed the previously claimed orders of accuracy of all four numerical schemes. Then, we presented the full viscoelastic phase separation process based on numerical experiments. Here, we also confirmed the reliability of our developed methods, that satisfy the conservation of mass and preserve the thermodynamic consistency of the underlying model equations by having a non-increasing free energy on the discrete level. For these experiments, we used not only our algorithm for the full model from Subsection 4.8.14, but also the full discretization (4.133) of the simplified model, confirming that also this model can describe the most important physical properties. Consequently, both can be used to efficiently simulate the complex dynamics of a phase separation process of a polymer-solvent mixture, including all key characteristics like the above-mentioned essential features observed experimentally in the lab. As part of our experiments, we conducted an in-depth parameter study to present and analyze the sensitivities of the underlying model equations to different parameter and initial value changes. Here, we observed that the simulation results vary noticeably for most proposed deviations from our reference parameter set. While most parameter deviations primarily only change the time scale of the phase separation process, some also influence the coarseness of the emerging network-like structures, e.g., the interface size, and a few can even change the whole phase separation dynamics, e.g., the friction and the bulk relaxation modulus. To elaborate the differences between the full and the simplified model, we also presented simulation results of viscoelastic phase separation in a vortex. We observed that the resulting dynamics can not be captured when neglecting hydrodynamics. Thanks to our collaborators from the Max Planck Institute, we also had the opportunity to compare our macroscopic simulation results to those of mesoscopic coarse-grained molecular dynamics simulations. Here, we mapped the initial data of one of their experiments, including the polymer configuration as well as the polymer and solvent velocity fields, to our discrete computational domain and compared the time evolutions of both models. The results, showing similar dynamics, are quite promising, especially when considering the rather crude mapping,

having neither energy nor parameter scaling. This again demonstrated the applicability and reliability of the proposed models and schemes. Finally, we presented some three-dimensional simulation results and observed that the viscoelastic phase separation dynamics are as desired similar to those of the two-dimensional simulation results, which verified that the proposed models and schemes perform as good in three space dimensions as they do in two.

Left for future work is a full error analysis of our proposed numerical schemes. Further future investigations should be aimed at extending the above-mentioned mapping by energy and parameter scaling, and at developing more refined and accurate macroscopic models as briefly touched at the end of Chapter 3, with the possible perspective of even constructing hybrid schemes for multiscale models of the viscoelastic phase separation process. Thus, the aim would be to combine the ideas of our proposed linear, thermodynamically consistent schemes for macroscopic models with the mesoscopic Lattice Boltzmann / coarse-grained molecular dynamics model for viscoelastic phase separation. It is to be believed that by such a hybrid multiscale simulation model the underlying physics will become more clear, which can not only provide a deeper insight but perhaps also the development of even more refined and accurate macroscopic models.

Bibliography

[1] H. Abels, D. Depner and H. Garcke. Existence of weak solutions for a diffuse interface model for two-phase flows of incompressible fluids with different densities. *J. Math. Fluid Mech.*, 15(3):453–480, 2013.

[2] H. Abels, D. Depner and H. Garcke. On an incompressible Navier–Stokes/Cahn–Hilliard system with degenerate mobility. *Ann. I. H. Poincare - An*, 30(6):1175–1190, 2013.

[3] H. Abels, H. Garcke and G. Grün. Thermodynamically consistent, frame indifferent diffuse interface models for incompressible two-phase flows with different densities. *Math. Models Methods Appl. Sci.*, 22, 2011.

[4] H. Abels, H. Garcke and G. Grün. Thermodynamically consistent diffuse interface models for incompressible two-phase flows with different densities. *Math. Models Methods Appl. Sci.*, 22, 2010.

[5] J. W. Barrett and S. Boyaval. Existence and approximation of a (regularized) Oldroyd-B model. *Math. Mod. Method. Appl. S.*, 21(09):1783–1837, 2011.

[6] F. Bashforth and J. C. Adams. *An attempt to test the theories of capillary action by comparing the theoretical and measured forms of drops of fluid. With an explanation of the method of integration employed in constructing the tables which give the theoretical forms of such drops.* Cambridge University Press, Cambridge, 1883.

[7] S. Boyaval, T. Lelièvre and C. Mangoubi. Free-energy-dissipative schemes for the Oldroyd-B model. *ESAIM: Math. Model. Numer. Anal.*, 43(3):523–561, 2009.

[8] F. Boyer. Mathematical study of multiphase flow under shear through order parameter formulation. *Asymptot. Anal.*, 20:175–212, 1999.

[9] F. Boyer. A theoretical and numerical model for the study of incompressible mixture flows. *Comput. Fluids*, 31:41–68, 2002.

[10] A. J. Bray. Theory of phase-ordering kinetics. *Adv. Phys.*, 51(2):481–587, 2002.

[11] A. Brunk, H. Egger, O. Habrich and M. Lukáčová-Medviďová. Relative energy estimates for the Cahn-Hilliard equation with concentration dependent mobility. *arXiv:2102.05704 [math.NA]*, 2021.

[12] A. Brunk and M. Lukáčová-Medviďová. Relative energy and weak-strong uniqueness of the two-phase viscoelastic phase separation model. *arXiv:2104.00589 [math.AP]*, 2021.

[13] A. Brunk and M. Lukáčová-Medviďová. Global existence of weak solutions to viscoelastic phase separation: Part I Regular case. *Nonlinearity*, 35(7):3417–3458, 2022.

[14] A. Brunk and M. Lukáčová-Medviďová. Global existence of weak solutions to viscoelastic phase separation: Part II Degenerate case. *Nonlinearity*, 35(7):3459–3486, 2022.

[15] J. W. Cahn and J. E. Hilliard. Free energy of a nonuniform system. I. Interfacial free energy. *J. Chem. Phys.*, 28:258–267, 1958.

[16] Y. Cheng, A. Kurganov, Z. Qu and T. Tang. Fast and stable explicit operator splitting methods for phase-field models. *J. Comput. Phys.*, 303:45–65, 2015.

[17] A. J. Chorin. A numerical method for solving incompressible viscous flow problems. *J. Comput. Phys.*, 2:12–26, 1967.

[18] L. Chupin. Existence result for a mixture of non Newtonian flows with stress diffusion using the Cahn-Hilliard formulation. *Discrete Contin. Dyn. Syst. Ser. B*, 3(1):45–68, 2003.

[19] P. Constantin and M. Kliegl. Note on Global Regularity for Two-Dimensional Oldroyd-B Fluids with Diffusive Stress. *Arch. Ration. Mech. Anal.*, 206(3):725–740, 2012.

[20] R. Courant, K. Friedrichs and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.*, 100:32–74, 1928.

[21] J. Crank and P. Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Adv. Comput. Math.*, 6:207–226, 1996.

[22] B. Dünweg and A. J. C. Ladd. Lattice Boltzmann simulations of Soft Matter systems. In C. Holm and K. Kremer, editors, *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, number 221 in Advances in Polymer Science, pages 89–166. Springer Berlin Heidelberg, 2009.

[23] C. M. Elliott and H. Garcke. On the Cahn-Hilliard equation with degenerate mobility. *SIAM J. Math. Anal.*, 27:404–423, 1996.

[24] C. M. Elliott and H. Garcke. Diffusional phase transitions in multicomponent systems with a concentration dependent mobility matrix. *Physica D*, 109:242–256, 1997.

[25] C. M. Elliott and Z. Songmu. On the Cahn-Hilliard equation. *Arch. Rat. Mech. Anal.*, 96:339–357, 1986.

[26] C. M. Elliott and H. Garcke. On the Cahn–Hilliard Equation with Degenerate Mobility. *SIAM J. Math. Anal.*, 27(2):404–423, 1996.

[27] D. J. Eyre. An unconditionally stable one-step scheme for gradient systems. *IEEE Trans. Image Process.*, 1997.

[28] R. Fattal and R. Kupferman. Time-dependent simulation of viscoelastic flows at high Weissenberg number using the log-conformation representation. *J. Non-Newton. Fluid*, 126(1):23–37, 2005.

[29] C. L. Fefferman. Existence and smoothness of the Navier-Stokes equation. In J. Carlson, A. Jaffe and A. Wiles, editors, *The Millennium Prize Problems*, pages 57–67. Clay Mathematics Institute, Cambridge, 2006.

[30] Weak-Strong Uniqueness in Fluid Dynamics. In C. Fefferman, J. C. Robinson and J. L. Rodrigo Diez, editors, *Partial differential equations in fluid mechanics*, 452, pages 289–326. Cambridge University Press, Cambridge, 2019.

[31] E. Fernández-Cara, F. M. Guillén-González and R. R. Ortega. Mathematical modeling and analysis of viscoelastic fluids of the Oldroyd kind. In *Handbook of numerical analysis (volume VIII)*, pages 543–660. Elsevier, 2002.

[32] P. J. Flory. Thermodynamics of high polymer solutions. *J. Chem. Phys*, 10:51–61, 1942.

[33] M. Giaquinta and S. Hildebrandt. *Calculus of Variations I*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2004.

[34] M.-H. Giga, A. Kirshtein and C. Liu. *Variational modeling and complex fluids*. In *Handbook of Mathematical Analysis in Mechanics of Viscous Fluids*. Y. Giga and A. Novotný, editors. Springer International Publishing, Cham, 2018, pages 73–113.

[35] V. Ginzburg, L. Landau, M. Leontovich and V. Fok. On the theory of superconductivity. *J. Exp. Theor. Phys.*, 20:1064, 1950.

[36] G. S. Grest and K. Kremer. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A*, 33(5):3628–3631, 1986.

[37] F. Guillén-González and G. Tierra. On linear schemes for a Cahn-Hilliard diffuse interface model. *J. Comput. Phys.*, 234:140–171, 2013.

[38] F. Guillén-González, M. Á. Rodríguez-Bellido and G. Tierra. Linear unconditional energy-stable splitting schemes for a phase-field model for nematic-isotropic flows with anchoring effects. *Int. J. Numer. Meth. Eng.*, 108(6):535–567, 2016.

[39] F. Guillén-González and G. Tierra. Second order schemes and time-step adaptivity for Allen–Cahn and Cahn–Hilliard models. *Comput. Math. Appl.*, 68(8):821–846, 2014.

[40] F. Guillén-González and G. Tierra. Splitting schemes for a Navier-Stokes-Cahn-Hilliard model for two fluids with different densities. *J. Comput. Math.*, 32(6):643–664, 2014.

[41] M. E. Gurtin. Generalized Ginzburg-Landau and Cahn-Hilliard equations based on a microforce balance. *Physica D*, 92(3):178 –192, 1996.

[42] F. H. Harlow and J. E. Welch. Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. *Phys. Fluids*, 8(12):2182–2189, 1965.

[43] E. Helfand and G. H. Fredrickson. Large fluctuations in polymer solutions under shear. *Phys. Rev. Lett.*, 62:2468–2471, 21, 1989.

[44] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Rev. Mod. Phys.*, 49(3):435–479, 1977.

[45] D. Hu and T. Lelièvre. New entropy estimates for the Oldroyd-B model and related models. *Commun. Math. Sci.*, 5(4):909–916, 2007.

[46] M. L. Huggins. Thermodynamics of high polymer solutions. *J. Chem. Phys*, 9:440, 1941.

[47] M. Hulsen. A sufficient condition for a positive definite configuration tensor in differential models. *J. Non-Newtonian Fluid Mech.*, 38:93–100, 1990.

[48] D. Kay and R. Welford. Efficient numerical solution of Cahn-Hilliard-Navier-Stokes fluids in 2D. *SIAM J. Sci. Comput.*, 29(6):2241–2257, 2007.

[49] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics, Second Edition: Volume 6 (Course of Theoretical Physics)*. Course of theoretical physics / by L. D. Landau and E. M. Lifshitz, Vol. 6. Butterworth-Heinemann, 2nd edition, 1987.

[50] R. Larson. Preface. In R. G. Larson, editor, *Constitutive Equations for Polymer Melts and Solutions*, Butterworths Series in Chemical Engineering, pages xv –xvi. Butterworth-Heinemann, 1988.

[51] C. Le Bris and T. Lelièvre. Micro-macro models for viscoelastic fluids: modelling, mathematics and numerics. *Sci. China Math.*, 55, 2011.

[52] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002.

[53] D. Lee, J.-Y. Huh, D. Jeong, J. Shin, A. Yun and J. Kim. Physical, mathematical, and numerical derivations of the Cahn-Hilliard equation. *Comp. Mater. Sci.*, 81:216–225, 2014.

[54] C. Liu, J. Shen and X. Yang. Decoupled energy stable schemes for a phase-field model of two-phase incompressible flows with variable density. *J. Sci. Comput.*, 62, 2015.

[55] J. Lowengrub and L. Truskinovsky. Quasi-incompressible Cahn-Hilliard fluids and topological transitions. *Proc. Math. Phys. Eng. Sci.*, 454(1978):2617–2654, 1998.

[56] M. Lukáčová-Medviďová, B. Dünweg, P. Strasser and N. Tretyakov. Energy-stable numerical schemes for multiscale simulations of polymer-solvent mixtures. In P. van Meurs, M. Kimura and H. Notsu, editors, *Mathematical Analysis of Continuum Mechanics and Industrial Applications II*. Volume 30, Mathematics for Industry, pages 153–165. Springer Singapore, 2018.

[57] M. Lukáčová-Medviďová, H. Notsu and B. She. Energy dissipative characteristic schemes for the diffusive Oldroyd-B viscoelastic fluid. *Int. J. Numer. Meth. Fl.*, 81(9):523–557, 2016.

[58] Masao Doi and Akira Onuki. Dynamic coupling between stress and composition in polymer solutions and blends. *J. Phys. II*, 2(8):1631–1656, 1992.

[59] *Mathematical Tools for the Study of the Incompressible Navier-Stokes Equations and Related Models*, volume 183 of *Applied Mathematical Sciences Ser.* Springer, New York, Nov. 2012.

[60] S. T. Milner. Dynamical theory of concentration fluctuations in polymer solutions under shear. *Phys. Rev. E*, 48:3674–3691, 5, 1993.

[61] H. Mizerová. *Analysis and numerical solution of the Peterlin viscoelastic model*. PhD thesis, University Mainz, 2015.

[62] A. Onuki. Shear-induced phase separation in polymer solutions. *J. Phys. Soc. Japan*, 59:3427–, 1990.

[63] A. Onuki. *Phase transition dynamics*. Cambridge University Press, 2002.

[64] J. Shen and X. Yang. A phase-field model and its numerical approximation for two-phase incompressible flows with different densities and viscosities. *SIAM J. Sci. Comput.*, 32:1159–1179, 2010.

[65] T. Soddemann, B. Dünweg and K. Kremer. A generic computer model for amphiphilic systems. *Eur. Phys. J. E*, 6(1):409–419, 2001.

[66] D. Spiller. *Spinodal Decomposition of Polymer-Solvent Systems: Theory and Simulation.* eng. PhD thesis, University Mainz, 2022.

[67] P. J. Strasser, G. Tierra, B. Dünweg and M. Lukáčová-Medviďová. Energy-stable linear schemes for polymer-solvent phase field models. *Comput. Math. Appl.*, 77(1):125–143, 2019.

[68] H. Tanaka. Viscoelastic phase separation. *J. Phys. Condens. Mat.*, 12(15):R207, 2000.

[69] H. Tanaka and T. Araki. Phase inversion during viscoelastic phase separation: roles of bulk and shear relaxation moduli. *Phys. Rev. Lett.*, 78:4966–4969, 26, 1997.

[70] B. Taylor. *Methodus incrementorum directa & inversa.* Impensis Gulielmi Innys, 1717.

[71] G. Tierra and F. Guillén-González. Numerical methods for solving the Cahn-Hilliard equation and its applicability to related energy-based models. *Arch. Comput. Method. E.*, 22(2):269–289, 2015.

[72] X. Wu, G. J. van Zwieten and K. G. van der Zee. Stabilized second-order convex splitting schemes for Cahn-Hilliard models with application to diffuse-interface tumor-growth models. *Int. J. Numer. Meth. Bio.*, 30(2):180–203, 2014.

[73] Y. Yan, W. Chen, C. Wang and S. M. Wise. A second-order energy stable bdf numerical scheme for the Cahn-Hilliard equation. *Commun. Comput. Phys.*, 23, 2018.

[74] D. Zhou, P. Zhang and W. E. Modified models of polymer phase separation. *Phys. Rev. E*, 73(6):061801, 2006.

Declaration

I hereby declare that I wrote the dissertation submitted without any unauthorized external assistance and used only sources acknowledged in this work. All textual passages which are appropriated verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice as formulated in the statutes of Johannes Gutenberg University Mainz to insure standard scientific practice.

Mainz, February 2023