SIAM J. SCI. COMPUT.
Vol. 42, No. 1, pp. B30–B56

A NOVEL ENERGY FACTORIZATION APPROACH FOR THE DIFFUSE-INTERFACE MODEL WITH PENG–ROBINSON EQUATION OF STATE*

JISHENG KOU†, SHUYU SUN‡, AND XIUHUA WANG§

Abstract. The Peng–Robinson equation of state (PR-EoS) has become one of the most extensively applied equations of state in chemical engineering and the petroleum industry due to its excellent accuracy in predicting the thermodynamic properties of a wide variety of materials, especially hydrocarbons. Although great effort has been made to construct efficient numerical methods for the diffuse interface models with PR-EoS, there is still not a linear numerical scheme that can be proved to preserve the original energy dissipation law. In order to pursue such a numerical scheme, we propose a novel energy factorization (EF) approach, which first factorizes an energy function into a product of several factors and then treats the factors using their properties to obtain the semi-implicit linear schemes. We apply the EF approach to deal with the Helmholtz free energy density determined by PR-EoS, and then propose a linear semi-implicit numerical scheme that inherits the original energy dissipation law. Moreover, the proposed scheme is proved to satisfy the maximum principle in both time semidiscrete form and cell-centered finite difference fully discrete form under certain conditions. Numerical results are presented to demonstrate the stability and efficiency of the proposed scheme.

Key words. diffuse interface model, Peng–Robinson equation of state, energy stability, maximum principle

AMS subject classifications. 65N30, 65N50, 49S05

DOI. 10.1137/19M1251230

1. Introduction.

The Peng–Robinson equation of state (PR-EoS) [34] has become one of the most popular and useful tools for describing the thermodynamic properties of fluids in both academic and industrial fields, especially chemical engineering and the petroleum industry [28, 46]. Compared to the well-known Van der Waals equation of state, PR-EoS can provide more reasonable accuracy in predicting the properties of a wide variety of materials, such as N_2, CO_2, and hydrocarbons. PR-EoS has been extensively applied for simulation of many important problems in petroleum and chemical engineering, for instance, phase equilibria calculations [10, 14, 15, 21, 29, 30, 31, 33, 40] and prediction of surface tension between gas and liquid [10, 18, 19, 32]. Modeling and simulation of compressible multicomponent two-phase flows with partial miscibility and realistic equations of state (e.g., PR-EoS) have been intensively studied in recent years [8, 19, 20, 22, 23, 25, 26, 27, 35, 36, 39, 40]. On the basis of the thermodynamic fundamental laws and realistic equations of state

*Submitted to the journal's Computational Methods in Science and Engineering section March 20, 2019; accepted for publication (in revised form) October 11, 2019; published electronically January 7, 2020.

https://doi.org/10.1137/19M1251230

Funding: This work was supported by the Scientific and Technical Research Project of Hubei Provincial Department of Education through grant D20192703.

†School of Civil Engineering, Shaoxing University, Shaoxing 312000, Zhejiang, China, and School of Mathematics and Statistics, Hubei Engineering University, Xiaogan 432000, Hubei, China (jishengkou@163.com).

‡Corresponding author. Computational Transport Phenomena Laboratory, Division of Physical Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia (shuyu.sun@kaust.edu.sa).

§School of Mathematics and Statistics, Hubei Engineering University, Xiaogan 432000, Hubei, China (wangxiuhua163email@163.com).

(e.g., PR-EoS), general diffuse interface models for compressible multicomponent two-phase flows have been proposed in [24, 39] for isothermal fluids and in [25, 26] for nonisothermal fluids.

This paper is primarily concerned with efficient numerical methods for an isothermal diffuse interface model with PR-EoS. There exist two primary challenging problems in designing numerical schemes for such model. One is that the Helmholtz free energy density determined by PR-EoS has complicated structures and strong nonlinearity. The other is that the model follows the energy dissipation law, and thus numerical schemes should be constructed to preserve this feature at the discrete level. In this paper, we will resolve the above challenging problems and develop a novel linear, energy stable numerical scheme.

We now provide an up-to-date review regarding the approaches in the literature employed to handle the bulk Helmholtz free energy density of PR-EoS and design energy stable numerical schemes. One approach is the convex splitting method [6, 7], which has been extensively used for various phase-field models [2, 37, 44]. For the diffuse interface models with PR-EoS, convex splitting schemes inheriting the discrete energy dissipation law have been developed in a series of recent works [8, 21, 24, 25, 35, 36]. When the convex splitting approach is applied to the PR-EoS-based Helmholtz free energy density, the ideal and repulsion terms are usually treated implicitly due to their convexity, while the attraction term with the concavity is treated explicitly. The convex splitting approach can produce unconditionally energy stable numerical schemes, but it results in nonlinear discrete equations that demand the complicated implement of efficient nonlinear iterative solvers and also are computationally expensive. Approximating chemical potential by a difference of Helmholtz free energy density, a fully implicit scheme was developed in [20]. This scheme is proved to be unconditionally energy stable, but it still suffers from the nonlinearity of the discrete equations.

The invariant energy quadratization (IEQ) approach [47, 48, 49] is a novel and efficient method developed in recent years and has been intensively applied for various phase-field models. The essential idea of IEQ is to transform the bulk free energy into a quadratic form through introducing a set of new variables. The new variables can be updated with time steps via semi-implicit linear schemes. A numerical scheme has been developed in [27] by applying the IEQ approach to PR-EoS. As a modification of the IEQ approach, the scalar auxiliary variable (SAV) approach has been proposed in [38], which introduces a scalar auxiliary variable instead of the space-dependent new variables in the IEQ approach. It leads to unconditionally stable numerical schemes, which only need to solve the linear equations with constant coefficients at each time step. Recently, in [23], the SAV approach was applied to construct unconditionally energy stable numerical schemes for the model proposed in [24]; moreover, a componentwise SAV approach has also been developed. The numerical schemes constructed by the IEQ and SAV approaches are linear, and consequently are easy to implement and very efficient in computation. However, while IEQ and SAV have become very useful and successful tools applied in a variety of phase-field models, the produced schemes use transformed free energies that are generally never equivalent to the original energies at the discrete level. Indeed, it has been indicated in [47] that the transformed free energies have errors of the order of time step size against the original energies. Consequently, the schemes constructed by IEQ and SAV may not inherit the original energy dissipation law in theory, although the dissipation of transformed energies can be proved. In numerical tests of [23], the original energy instability has been observed despite the transformed energies decrease with time steps.

To the best of our knowledge, for the diffuse interface model with PR-EoS, there is no linear semi-implicit numerical scheme inheriting the original energy dissipation law so far. In this paper, we will propose such a scheme.

In this paper, we will propose a novel energy factorization (EF) approach to construct an efficient numerical scheme for the diffuse interface model with PR-EoS. The key idea of EF is that we first factorize an energy function into a product of several factors and then handle them by the use of their properties to obtain the semi-implicit linear schemes. The EF approach will not introduce any new independent energy variable, and thus it can ensure the original energy dissipation law. Applying the EF approach to the model with PR-EoS, we will obtain a linear, efficient semi-implicit numerical scheme that inherits the original energy dissipation law.

We note that molar density is the primal variable in models with PR-EoS. Moreover, for a realistic substance, molar density shall have physical limits under given thermodynamical conditions, and as a result, the maximum principle is essential for numerical methods to ensure that numerical solutions are physically reasonable. The Flory–Huggins energy potential [3, 5, 11, 12, 16, 42, 45, 47] is the most useful logarithmic potential in the phase-field models, which also requires the maximum principle to avoid the singularity. For the Flory–Huggins potential, the fully implicit Euler scheme was analyzed in [5], and, more recently, the discrete maximum principle of the convex splitting scheme was rigorously proved in [3]. However, there are no results concerning the maximum principle of numerical schemes for models with PR-EoS so far. The cell-centered finite difference (CCFD) method can be equivalent to a special mixed finite element method with quadrature rules [1] and now has been widely applied for the phase-field models [4, 9, 13, 17, 43, 44, 50, 52]. For the first time, the proposed numerical scheme will be proved to preserve the maximum principle under certain conditions. The proof will be provided for both the semidiscrete time matching scheme and the CCFD fully discrete scheme. Error estimates of the discrete schemes will be analyzed as well.

The rest of this paper is organized as follows. In section 2, we provide a brief description of the diffuse interface model with PR-EoS. In section 3, we propose the energy factorization approach to deal with the bulk Helmholtz free energy density. In section 4, we present the semidiscrete time scheme and prove some theoretical results including the energy stability, maximum principle, and error estimate. The fully discrete scheme is developed and analyzed in section 5. In section 6, numerical results are presented to validate the proposed numerical scheme. Finally, some concluding remarks are given in section 7.

2. Model equations. We now give a brief description for the Helmholtz free energy density of a bulk fluid (denoted by f_b) determined by the PR-EoS [34]. We denote by c the molar density of a substance. For specified temperature T, f_b is a function of molar density c and can be expressed as a sum of three contributions,

$$f_b(c) = f_b^{\text{ideal}}(c) + f_b^{\text{repulsion}}(c) + f_b^{\text{attraction}}(c), \tag{2.1}$$

where

$$f_b^{\text{ideal}}(c) = c\vartheta_0 + cRT\ln(c), \tag{2.2}$$

$$f_b^{\text{repulsion}}(c) = -cRT\ln(1-\beta c), \tag{2.3}$$

$$f_b^{\text{attraction}}(c) = \frac{\alpha(T)c}{2\sqrt{2}\beta}\ln\left(\frac{1+(1-\sqrt{2})\beta c}{1+(1+\sqrt{2})\beta c}\right). \tag{2.4}$$

Here, R is the universal gas constant and ϑ_0 is an energy parameter that relies on the temperature and thermodynamical properties of a specific substance. The substance-specific parameters α and β can be determined from the critical properties and acentric factor of a specific substance,

$$\alpha = 0.45724\frac{R^2T_c^2}{P_c}\left[1+m(1-\sqrt{T_r})\right]^2, \quad \beta = 0.07780\frac{RT_c}{P_c}, \tag{2.5}$$

where $T_r = T/T_c$ is the reduced temperature, T_c and P_c stand for the critical temperature and critical pressure, respectively, and m is calculated from the acentric factor ω as follows:

$$m = 0.37464 + 1.54226\omega - 0.26992\omega^2, \quad \omega \le 0.49,$$

$$m = 0.379642 + 1.485030\omega - 0.164423\omega^2 + 0.016666\omega^3, \quad \omega > 0.49.$$

From a physical point of view, β is the effective volume of one mole of a substance. Let $v = \frac{1}{c}$ be the molar volume, i.e., the average volume occupied by one mole of a substance, which includes the effective volume β and the space between molecules. Therefore, we know that generally $\beta \ll v$ for gas and liquid. For an ideal and simple example, if the molecules can be approximated as spherical particles, then we have

$$\beta c = \frac{\beta}{v} \le \frac{\pi}{6}.$$

This fact suggests that βc has an upper bound for a specific substance under specified temperature. As a matter of fact, from the following form of PR-EOS [34],

$$P = \frac{cRT}{1-\beta c} - \frac{\alpha(T)c^2}{1+\beta c+\beta c(1-\beta c)} > \frac{cRT}{1-\beta c} - \alpha(T)c^2, \tag{2.6}$$

where P is the pressure, we can directly deduce

$$\beta c < 1 - \frac{cRT}{P+\alpha(T)c^2}.$$

To justify the boundedness of βc in realistic cases, we consider the species of n-butane. We can calculate $\beta = 7.2381 \times 10^{-5}$ m^3/mol using (2.5) and the physical data of n-butane. Molar densities of gas and liquid of n-butane at the temperature 330 K and pressure 106.39 bar are $c^G = 249.1123$ mol/m^3 and $c^L = 9526.8428$ mol/m^3, respectively. Then we have

$$\beta c^G = 0.0180, \quad \beta c^L = 0.6896.$$

On the basis of the above physical observation, we assume that molar density c is always bounded as

$$0 < c_m \le c \le c_M, \quad \beta c_M \le \epsilon_0 < 1, \tag{2.7}$$

where c_m, c_M, and ϵ_0 are determined by a specific substance under specified pressure and temperature. We remark that ϵ_0 usually does not take some very small value from the physical property as stated above.

Since the diffuse interfaces always exist between gas and liquid phases, in addition to the bulk free energy density, the gradient free energy density accounting for the effect of the interfaces is defined as

$$f_\nabla(c) = \frac{1}{2}\kappa|\nabla c|^2, \tag{2.8}$$

where $\kappa > 0$ is the influence parameter that can be calculated as follows:

$$\kappa = \alpha\beta^{2/3}\left[a_0(1-T_r)+a_1\right],$$

$$a_0 = -\frac{10^{-16}}{1.2326+1.3757\omega}, \qquad a_1 = \frac{10^{-16}}{0.9051+1.5410\omega}.$$

We denote by $f(c)$ the general Helmholtz free energy density:

$$f(c) = f_b(c) + f_\nabla(c). \tag{2.9}$$

The chemical potential is defined as the variational derivative of $f(c)$:

$$\mu(c) = \frac{\delta f(c)}{\delta c} = \mu_b(c) - \kappa\Delta c, \tag{2.10}$$

where $\mu_b(c) = f_b'(c)$ is the bulk chemical potential.

Let Ω be a connected and smooth space domain. We now state the model equation as follows [36]:

$$\frac{\partial c}{\partial t} - \kappa\Delta c + \mu_b(c) = \mu_e, \tag{2.11a}$$

$$\int_\Omega c d\mathbf{x} = c_t, \tag{2.11b}$$

$$\nabla c \cdot \boldsymbol{n}_{\partial\Omega} = 0, \quad \mathbf{x} \in \partial\Omega, \quad c(\mathbf{x},0) = c_0(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{2.11c}$$

where $\boldsymbol{n}_{\partial\Omega}$ denotes the normal unit outward vector to $\partial\Omega$, $c_t > 0$ is the total moles in Ω , and μ_e is a Lagrange multiplier incorporated to enforce total mole conservation. We note that μ_e is constant in space but could vary with time. Moreover, as time goes on, the spatial distribution of molar density c will approach an equilibrium state, and μ_e will become the chemical potential at the equilibrium state. So from a physical point of view, μ_e can be viewed as a general chemical potential. We also remark that μ_e will lead to a challenging issue for the maximum principle, which requires us to find some appropriate treatments to justify the maximum principle, so we will propose a reasonable condition given by (4.9) to resolve this issue, which is easily verified in computation. Here, we consider the homogeneous Neumann boundary condition, but the proposed numerical schemes and theoretical analysis can be directly extended to various boundary conditions, for instance, Dirichlet boundary conditions.

We note that the model (2.11) obeys the energy dissipation law

$$\frac{\partial}{\partial t}\int_\Omega f(c(\mathbf{x},t))d\mathbf{x} = -\int_\Omega \left(\frac{\partial c}{\partial t}\right)^2 d\mathbf{x}. \tag{2.12}$$

3. Energy factorization approach. In this section, we propose a novel energy factorization (EF) approach to deal with the bulk Helmholtz free energy density f_b. The basic idea of EF is to factorize an energy function or a term of the energy function into a product of several factors that can be separately treated by the use of their properties. Two different factorizations are proposed to deal with the ideal term and the repulsion term of the bulk Helmholtz free energy density, respectively.

At the time discrete level, we denote the time step size by τ and set $t_n = n\tau$. Furthermore, we use c^n to denote the approximation of molar density c at time t_n.

3.1. Factorization approach for the ideal term. The first EF approach is proposed to deal with the ideal term. We define the function $H(c) = c\ln(c)$, which can be factorized into the product of a linear function c and a logarithm function $\ln(c)$. Apparently, $\ln(c)$ is a concave function, and thus we have

$$\ln(c^{n+1}) \le \ln(c^n) + \frac{1}{c^n}\left(c^{n+1} - c^n\right). \tag{3.1}$$

For $c^n > 0$ and $c^{n+1} > 0$, using (3.1), we can deduce that

$$\begin{aligned} H(c^{n+1}) - H(c^n) &= c^{n+1}\ln\left(c^{n+1}\right) - c^n\ln(c^n) \\ &= \ln(c^n)\left(c^{n+1} - c^n\right) + c^{n+1}\left(\ln(c^{n+1}) - \ln(c^n)\right) \\ &\le \left(\ln\left(c^n\right) + \frac{c^{n+1}}{c^n}\right)\left(c^{n+1} - c^n\right). \end{aligned} \tag{3.2}$$

Then the ideal term $f_b^{\text{ideal}}(c)$ can be estimated as

$$f_b^{\text{ideal}}(c^{n+1}) - f_b^{\text{ideal}}(c^n) \le \vartheta_0\left(c^{n+1} - c^n\right) + RT\left(\ln\left(c^n\right) + \frac{c^{n+1}}{c^n}\right)\left(c^{n+1} - c^n\right), \tag{3.3}$$

which suggests that the ideal part of chemical potential at the $(n+1)$th time step be defined as

$$\mu_{\text{ideal}}^{n+1} = \vartheta_0 + RT\ln\left(c^n\right) + RT\frac{c^{n+1}}{c^n}. \tag{3.4}$$

Therefore, from (3.3), we have

$$f_b^{\text{ideal}}(c^{n+1}) - f_b^{\text{ideal}}(c^n) \le \mu_{\text{ideal}}^{n+1}\left(c^{n+1} - c^n\right). \tag{3.5}$$

We remark that the convex splitting approach for the ideal term $f_b^{\text{ideal}}(c)$ used in [36] leads to a highly nonlinear scheme, which demands the complicated implementations of efficient nonlinear iterative solvers and also is computationally expensive. In contrast, μ_{ideal}^{n+1} is semi-implicit and linear with respect to c^{n+1}, and thus it is easy to implement and can be solved at less computational cost. This approach can also be directly applied for the logarithmic Flory–Huggins potential [47, 51].

3.2. Factorization approach for the repulsion term. Since the function $-\ln(1-\beta c)$ is a convex function, we would obtain a nonlinear chemical potential when the first energy factorization approach dealing with the ideal term is employed for the repulsion term $f_b^{\text{repulsion}}(c)$. In order to pursue a linear scheme, we introduce the second energy factorization approach to deal with the repulsion term $f_b^{\text{repulsion}}(c)$.

We define the modified repulsion energy function as

$$\widehat{f}_b^{\text{repulsion}}(c) = \lambda c + \frac{1}{RT}f_b^{\text{repulsion}}(c) = \lambda c - c\ln(1-\beta c), \tag{3.6}$$

where λ is some positive constant relying on the specific substance. Apparently, $\widehat{f}_b^{\text{repulsion}}(c)$ is positive and bounded for molar density c satisfying (2.7). We further define the following intermediate energy function:

$$G(c) = \sqrt{\lambda c - c\ln(1-\beta c)}. \tag{3.7}$$

From this, we can factorize $\widehat{f}_b^{\text{repulsion}}(c)$ into the square of $G(c)$, i.e.,

$$\widehat{f}_b^{\text{repulsion}}(c) = G(c)^2. \tag{3.8}$$

For the function $G(c)$, we have the following key lemma regarding the choice of λ.

LEMMA 3.1. *Assume that molar density c satisfies* (2.7). *If λ is taken such that*

$$\lambda \geq \frac{\epsilon_0}{(1-\epsilon_0)^2} + \left(\frac{\epsilon_0^2}{(1-\epsilon_0)^4} - 2\ln(1-\epsilon_0)\frac{\epsilon_0}{(1-\epsilon_0)^2}\right)^{1/2}, \tag{3.9}$$

where ϵ_0 is given in (2.7), *then $G(c)$ is a concave function.*

Proof. We calculate the first and second derivatives of $G(c)$ as follows:

$$G'(c) = \frac{1}{2}G(c)^{-1}\left(\lambda - \ln(1-\beta c) + \frac{\beta c}{1-\beta c}\right), \tag{3.10}$$

$$\begin{aligned}
G''(c) &= -\frac{1}{4}G(c)^{-3}\left(\lambda - \ln(1-\beta c) + \frac{\beta c}{1-\beta c}\right)^2 \\
&\quad + \frac{1}{2}G(c)^{-1}\left(\frac{\beta}{1-\beta c} + \frac{\beta}{(1-\beta c)^2}\right) \\
&= -\frac{1}{4}G(c)^{-3}\left(\lambda^2 - 2\lambda\ln(1-\beta c) + (\ln(1-\beta c))^2 + \left(\frac{\beta c}{1-\beta c}\right)^2\right. \\
&\qquad\qquad \left. - 2\lambda\frac{\beta c}{(1-\beta c)^2} + 2\ln(1-\beta c)\frac{\beta c}{(1-\beta c)^2}\right) \\
&\leq -\frac{1}{4}G(c)^{-3}\left(\lambda^2 - 2\lambda\frac{\beta c}{(1-\beta c)^2} + 2\ln(1-\beta c)\frac{\beta c}{(1-\beta c)^2}\right) \\
&\leq -\frac{1}{4}G(c)^{-3}\left(\lambda^2 - 2\lambda\frac{\epsilon_0}{(1-\epsilon_0)^2} + 2\ln(1-\epsilon_0)\frac{\epsilon_0}{(1-\epsilon_0)^2}\right). \qquad (3.11)
\end{aligned}$$

Applying condition (3.9), we obtain $G''(c) \leq 0$, and thus $G(c)$ is concave. ∎

We can see from the proof of Lemma 3.1 that condition (3.9) can be further substantially relaxed. Moreover, λ usually does not demand a large value in practice since ϵ_0 generally is not a very small value; for instance, we calculate from (3.9) that $\lambda = 27.3656$ in numerical examples. We also note that λ is dimensionless.

The advantage of the factorization (3.8) is shown in the following lemma.

LEMMA 3.2. *Assume that molar density satisfies* (2.7) *and λ is taken such that* (3.9) *holds. Then we have*

$$G(c^{n+1})^2 - G(c^n)^2 \leq \left(G^{n+1} + G(c^n)\right)G'(c^n)\left(c^{n+1} - c^n\right), \tag{3.12}$$

where G^{n+1} is the linear approximation of $G(c^{n+1})$ as

$$G^{n+1} = G(c^n) + G'(c^n)\left(c^{n+1} - c^n\right). \tag{3.13}$$

Proof. The assumption implies the concavity of $G(c)$, so we have

$$G(c^{n+1}) \le G(c^n) + G'(\phi^n)\left(c^{n+1} - c^n\right) = G^{n+1}. \tag{3.14}$$

Since $G(c) \ge 0$, we get $G^{n+1} \ge 0$ from (3.14), and consequently

$$G(c^{n+1})^2 \le |G^{n+1}|^2. \tag{3.15}$$

From (3.13) and (3.15), we derive that

$$\begin{aligned} G(c^{n+1})^2 - G(c^n)^2 &\le |G^{n+1}|^2 - G(c^n)^2 \\ &= \left(G^{n+1} + G(c^n)\right)\left(G^{n+1} - G(c^n)\right) \\ &= \left(G^{n+1} + G(c^n)\right) G'(c^n)\left(c^{n+1} - c^n\right). \end{aligned} \tag{3.16}$$

This ends the proof. ∎

We are now ready to consider treatment of $f_b^{\text{repulsion}}(c)$ under condition (3.9). Using Lemma 3.2, we can estimate the energy difference between two time steps as

$$\begin{aligned} \widehat{f}_b^{\text{repulsion}}(c^{n+1}) - \widehat{f}_b^{\text{repulsion}}(c^n) &= G(c^{n+1})^2 - G(c^n)^2 \\ &\le \left(G^{n+1} + G(c^n)\right) G'(c^n)\left(c^{n+1} - c^n\right). \end{aligned} \tag{3.17}$$

It follows from (3.6) and (3.17) that

$$\begin{aligned} f_b^{\text{repulsion}}(c^{n+1}) - f_b^{\text{repulsion}}(c^n) &= RT\left(\widehat{f}_b^{\text{repulsion}}(c^{n+1}) - \widehat{f}_b^{\text{repulsion}}(c^n)\right) \\ &\quad - RT\lambda\left(c^{n+1} - c^n\right) \\ &\le RT\left(\left(G^{n+1} + G(c^n)\right) G'(c^n) - \lambda\right)\left(c^{n+1} - c^n\right). \end{aligned} \tag{3.18}$$

Thus, we can define the repulsion chemical potential at the $(n+1)$th time step as

$$\mu_{\text{repulsion}}^{n+1} = RTG'(c^n)\left(G^{n+1} + G(c^n)\right) - \lambda RT. \tag{3.19}$$

Substituting (3.13) into (3.19), we rewrite

$$\mu_{\text{repulsion}}^{n+1} = RTG'(c^n)\left(2G(c^n) + G'(c^n)\left(c^{n+1} - c^n\right)\right) - \lambda RT, \tag{3.20}$$

which is a linear function of c^{n+1}. We note that the convex splitting approach for the repulsion term $f_b^{\text{repulsion}}(c)$ [36] results in a nonlinear scheme as well as the ideal term.

The following lemma is a direct consequence of the above analysis and the definition of $\mu_{\text{repulsion}}^{n+1}$.

LEMMA 3.3. *Assume that molar density satisfies* (2.7) *and* λ *is taken such that* (3.9) *holds. Then we have*

$$f_b^{\text{repulsion}}(c^{n+1}) - f_b^{\text{repulsion}}(c^n) \le \mu_{\text{repulsion}}^{n+1}\left(c^{n+1} - c^n\right). \tag{3.21}$$

We remark that the proposed approach is different from the IEQ and SAV approaches. In the IEQ and SAV approaches, some new auxiliary energy variables are introduced, and from this the original energy is transformed to a quadratic form. Nevertheless, the transformed energy is generally not equivalent to the original energy at the time discrete level [47]. In the proposed approach, we just use $G(c)$ as an intermediate function of c, but never introducing any new independent variable. This is a key feature of the proposed approach that allows us to apply the concavity of $G(c)$ to obtain the linear numerical scheme inheriting the dissipation law of the original energy at the discrete level.

4. Semi-implicit time discrete scheme. In this section, we propose a semi-implicit time discrete scheme based on the results presented in section 3. The ideal and repulsion terms have been handled by the EF approach. Due to the concavity of the attraction term [36], we treat it explicitly and define the corresponding chemical potential term as

$$\mu_{\text{attraction}}(c^n) = \frac{\alpha(T)}{2\sqrt{2}\beta}\ln\left(\frac{1+(1-\sqrt{2})\beta c^n}{1+(1+\sqrt{2})\beta c^n}\right) - \frac{\alpha(T)c^n}{1+2\beta c^n-(\beta c^n)^2}. \tag{4.1}$$

Let τ be the time step size, and let c^0 be provided by the initial condition; we now state the semi-implicit linear time discrete scheme as follows:

$$\frac{c^{n+1}-c^n}{\tau} - \kappa\Delta c^{n+1} + \mu_{\text{ideal}}^{n+1} + \mu_{\text{repulsion}}^{n+1} + \mu_{\text{attraction}}(c^n) = \mu_e^{n+1}, \tag{4.2a}$$

$$\int_\Omega c^{n+1}d\mathbf{x} = c_t, \tag{4.2b}$$

$$\nabla c^{n+1}\cdot \boldsymbol{n}_{\partial\Omega} = 0, \quad \mathbf{x}\in\partial\Omega, \tag{4.2c}$$

where μ_{ideal}^{n+1} and $\mu_{\text{repulsion}}^{n+1}$ are defined in (3.4) and (3.20), respectively. For the convenience of theoretical analysis, we rewrite (4.2) in the following equivalent form:

$$\frac{1}{\tau}\left(c^{n+1}-c^n\right) - \kappa\Delta c^{n+1} + \nu(c^n)c^{n+1} = s_r(c^n) + \mu_e^{n+1}, \tag{4.3a}$$

$$\int_\Omega c^{n+1}d\mathbf{x} = c_t, \tag{4.3b}$$

$$\nabla c^{n+1}\cdot \boldsymbol{n}_{\partial\Omega} = 0 \ \text{ on } \partial\Omega, \tag{4.3c}$$

where $\nu(c)$ and $s_r(c)$ are the functions of c defined as

$$\nu(c) = RT\left(\frac{1}{c}+G'(c)^2\right), \tag{4.4}$$

$$s_r(c) = -\vartheta_0 - RT\ln(c) + RT\left(G'(c)^2c - 2G(c)G'(c)+\lambda\right) - \mu_{\text{attraction}}(c). \tag{4.5}$$

In what follows, we use traditional notation to denote the inner product of $L^2(\Omega)$ and $\left(L^2(\Omega)\right)^d$ by $(\cdot,\cdot)$ and the norm of $L^2(\Omega)$ and $\left(L^2(\Omega)\right)^d$ by $\|\cdot\|$, where d is the spatial dimension.

4.1. Well-posedness of the solution. We now show the existence and uniqueness of the solution of the semidiscrete scheme (4.3) as follows.

THEOREM 4.1. *Assume that c^n satisfies condition* (2.7) *and λ is taken to satisfy* (3.9). *There exists a unique c^{n+1} to solve* (4.3) *weakly in $H^1(\Omega)$.*

Proof. Suppose that c is a solution of the following homogeneous problem:

$$\frac{1}{\tau}c - \kappa\Delta c + \nu(c^n)c = \mu_e, \tag{4.6a}$$

$$\int_\Omega c d\mathbf{x} = 0, \tag{4.6b}$$

$$\nabla c \cdot \boldsymbol{n}_{\partial\Omega} = 0 \ \text{ on } \partial\Omega. \tag{4.6c}$$

By the Fredholm alternative theorem, it suffices to prove $c \equiv 0$. Multiplying (4.6a) by c and integrating it over Ω, we obtain

$$\frac{1}{\tau}\|c\|^2 + \kappa\|\nabla c\|^2 + \left(\nu(c^n), c^2\right) = (\mu_e, c)\,. \tag{4.7}$$

The right-hand side term vanishes due to (4.6b). Furthermore, $\nu(c^n) > 0$ holds for $c_m \le c^n \le c_M$. Equation (4.7) can be reduced to

$$\frac{1}{\tau}\|c\|^2 + \kappa\|\nabla c\|^2 \le 0. \tag{4.8}$$

This means that $c \equiv 0$ almost everywhere in Ω as well as $\mu_e = 0$ from (4.6a). ∎

4.2. Maximum principle. We now prove that the time scheme given in (4.2) follows the maximum principle, which can ensure that the schemes presented in section 3 are always well defined.

THEOREM 4.2. *Assume that c^n satisfies condition* (2.7) *and λ is taken to satisfy* (3.9). *If*

$$\max_{c_m \le c \le c_M} \left(c_m\nu(c) - s_r(c)\right) \le \mu_e^{n+1} \le \min_{c_m \le c \le c_M} \left(c_M\nu(c) - s_r(c)\right), \tag{4.9}$$

then we have $c_m \le c^{n+1} \le c_M$ almost everywhere.

Proof. Suppose that $c_m \le c^n \le c_M$ holds for $n \ge 0$. We first prove that $c^{n+1} \ge c_m$ almost everywhere. Let $c_-^{n+1} = \min(c^{n+1} - c_m, 0)$ and apparently $c_-^{n+1} \le 0$. Multiplying (4.3a) by c_-^{n+1} and then integrating it over Ω, we have

$$\begin{aligned}\frac{1}{\tau}\left(c^{n+1} - c^n, c_-^{n+1}\right) - \kappa\left(\Delta c^{n+1}, c_-^{n+1}\right) + \left(\nu(c^n)(c^{n+1} - c_m), c_-^{n+1}\right)\\ = \left(\mu_e^{n+1}, c_-^{n+1}\right) + \left(s_r(c^n) - c_m\nu(c^n), c_-^{n+1}\right).\end{aligned} \tag{4.10}$$

For the first term on the left-hand side of (4.10), thanks to $c^n \ge c_m$, we deduce

$$\left(c^{n+1} - c^n, c_-^{n+1}\right) = \|c_-^{n+1}\|^2 - \left(c^n - c_m, c_-^{n+1}\right) \ge \|c_-^{n+1}\|^2. \tag{4.11}$$

By using the boundary condition, the second term on the left-hand side of (4.10) becomes

$$-\kappa\left(\Delta c^{n+1}, c_-^{n+1}\right) = \kappa\|\nabla c_-^{n+1}\|^2. \tag{4.12}$$

We observe that $\nu(c)$ is a strictly monotonically decreasing positive function over the interval $[c_m, c_M]$. The third term on the left-hand side of (4.10) is bounded below:

$$\left(\nu(c^n)(c^{n+1} - c_m), c_-^{n+1}\right) = \left(\nu(c^n), |c_-^{n+1}|^2\right) \ge \nu(c_M)\|c_-^{n+1}\|^2. \tag{4.13}$$

Using condition (4.9) and taking into account $c_-^{n+1} \le 0$, we estimate the first term on the right-hand side of (4.10) as

$$\begin{aligned}\left(\mu_e^{n+1}, c_-^{n+1}\right) &\le \left(\max_{c_m \le c \le c_M} \left(c_m \nu(c) - s_r(c)\right), c_-^{n+1}\right) \\ &\le \left(c_m \nu(c^n) - s_r(c^n), c_-^{n+1}\right). \end{aligned} \tag{4.14}$$

Combining (4.11)–(4.14), we derive from (4.10) that

$$\frac{1}{\tau}\|c_-^{n+1}\|^2 + \kappa\|\nabla c_-^{n+1}\|^2 + \nu(c_M)\|c_-^{n+1}\|^2 \le 0. \tag{4.15}$$

It follows from (4.15) that $\|c_-^{n+1}\|^2 = 0$ and $\|\nabla c_-^{n+1}\|^2 = 0$. Consequently, $c^{n+1} \ge c_m$ almost everywhere.

We now turn to prove $c^{n+1} \le c_M$ almost everywhere. We define $c_+^{n+1} = \max(c^{n+1} - c_M, 0)$ and apparently $c_+^{n+1} \ge 0$. Similar to (4.10), we can get

$$\begin{aligned}&\frac{1}{\tau}\left(c^{n+1} - c^n, c_+^{n+1}\right) - \kappa\left(\Delta c^{n+1}, c_+^{n+1}\right) + \left(\nu(c^n)(c^{n+1} - c_M), c_+^{n+1}\right) \\ &= \left(\mu_e^{n+1}, c_+^{n+1}\right) + \left(s_r(c^n) - \nu(c^n)c_M, c_+^{n+1}\right). \end{aligned} \tag{4.16}$$

Taking into account $c^n \le c_M$, we deduce

$$\left(c^{n+1} - c^n, c_+^{n+1}\right) = \|c_+^{n+1}\|^2 - \left(c^n - c_M, c_+^{n+1}\right) \ge \|c_+^{n+1}\|^2. \tag{4.17}$$

Using condition (4.9) and $c_+^{n+1} \ge 0$, we derive

$$\begin{aligned}\left(\mu_e^{n+1}, c_+^{n+1}\right) &\le \left(\min_{c_m \le c \le c_M} \left(c_M \nu(c) - s_r(c)\right), c_+^{n+1}\right) \\ &\le \left(c_M \nu(c^n) - s_r(c^n), c_+^{n+1}\right). \end{aligned} \tag{4.18}$$

Using the routines similar to those in the derivations of (4.15), we can obtain

$$\frac{1}{\tau}\|c_+^{n+1}\|^2 + \kappa\|\nabla c_+^{n+1}\|^2 + \nu(c_M)\|c_+^{n+1}\|^2 \le 0. \tag{4.19}$$

This implies that $\|c_+^{n+1}\|^2 = 0$ and $\|\nabla c_+^{n+1}\|^2 = 0$, and thus $c^{n+1} \le c_M$ almost everywhere. ∎

We remark that it is reasonable to assume condition (4.9), as it is likely to be required in a typical physical setting. As a matter of fact, we can derive its a priori bounds by just assuming $c^{n+1} > 0$. Integrating (4.3a) over Ω and using the constraint (4.3b), we get

$$\mu_e^{n+1} = \frac{1}{|\Omega|}\int_\Omega \left(\nu(c^n)c^{n+1} - s_r(c^n)\right) d\mathbf{x}, \tag{4.20}$$

where $|\Omega|$ is the measure of Ω. Using the constraint (4.3b) again, we can estimate

$$\int_\Omega \nu(c^n)c^{n+1} d\mathbf{x} \le \max\left(\nu(c^n)\right)\int_\Omega c^{n+1} d\mathbf{x} \le \nu(c_m)c_t, \tag{4.21}$$

$$\int_\Omega \nu(c^n)c^{n+1} d\mathbf{x} \ge \min\left(\nu(c^n)\right)\int_\Omega c^{n+1} d\mathbf{x} \ge \nu(c_M)c_t. \tag{4.22}$$

Let us denote $s_r^m = \min_{c\in[c_m,c_M]}(s_r(c))$ and $s_r^M = \max_{c\in[c_m,c_M]}(s_r(c))$. Thus, μ_e^{n+1} is bounded as

$$\nu(c_M)\frac{c_t}{|\Omega|} - s_r^M \le \mu_e^{n+1} \le \nu(c_m)\frac{c_t}{|\Omega|} - s_r^m. \tag{4.23}$$

Numerical results will also be presented to verify (4.9).

4.3. Energy stability. We define the total free energy as

$$F(c^n) = (f_b(c^n), 1) + \frac{1}{2}\kappa\|\nabla c^n\|^2. \tag{4.24}$$

The following theorem shows that the proposed scheme inherits the dissipation law of the original energy at the discrete level.

THEOREM 4.3. *Assume that c^n satisfies* (2.7) *and λ is taken to satisfy* (3.9). *Under condition* (4.9), *for any time step size τ, we have*

$$F(c^{n+1}) \le F(c^n). \tag{4.25}$$

Proof. It follows from the concavity of $f_b^{\text{attraction}}(c)$ that

$$\left(f_b^{\text{attraction}}(c^{n+1}) - f_b^{\text{attraction}}(c^n), 1\right) \le \left(\mu_{\text{attraction}}(c^n), c^{n+1} - c^n\right). \tag{4.26}$$

Applying (3.5), (3.21), and (4.26), we deduce that

$$\left(f_b(c^{n+1}) - f_b(c^n), 1\right) \le \left(\mu_{\text{ideal}}^{n+1} + \mu_{\text{repulsion}}^{n+1} + \mu_{\text{attraction}}(c^n), c^{n+1} - c^n\right). \tag{4.27}$$

For the gradient contribution to the free energy, we can derive that

$$\begin{aligned}
\frac{1}{2}\left(\|\nabla c^{n+1}\|^2 - \|\nabla c^n\|^2\right) &= \frac{1}{2}\int_\Omega \left(|\nabla c^{n+1}|^2 - |\nabla c^n|^2\right) d\mathbf{x} \\
&= \int_\Omega \nabla c^{n+1}\cdot\nabla\left(c^{n+1} - c^n\right) d\mathbf{x} - \frac{1}{2}\int_\Omega |\nabla(c^{n+1} - c^n)|^2 d\mathbf{x} \\
&\le \int_\Omega \nabla c^{n+1}\cdot\nabla\left(c^{n+1} - c^n\right) d\mathbf{x} \\
&= -\int_\Omega \left(c^{n+1} - c^n\right)\Delta c^{n+1} d\mathbf{x}. \qquad (4.28)
\end{aligned}$$

Taking into account (4.2a) and the mass constraint (4.2b), we deduce from (4.27) and (4.28) that

$$\begin{aligned}
F(c^{n+1}) - F(c^n) &= \left(f_b(c^{n+1}) - f_b(c^n), 1\right) + \frac{1}{2}\kappa\left(\|\nabla c^{n+1}\|^2 - \|\nabla c^n\|^2\right) \\
&\le \left(\mu_{\text{ideal}}^{n+1} + \mu_{\text{repulsion}}^{n+1} + \mu_{\text{attraction}}(c^n) - \kappa\Delta c^{n+1}, c^{n+1} - c^n\right) \\
&= \left(\mu_e^{n+1} - \frac{c^{n+1} - c^n}{\tau}, c^{n+1} - c^n\right) \\
&= -\frac{1}{\tau}\|c^{n+1} - c^n\|^2. \qquad (4.29)
\end{aligned}$$

Thus, the energy inequality (4.25) is proved. ∎

4.4. Error estimate. We analyze the error estimate of the proposed semi-discrete scheme. Here, we denote $c(t_n) = c(\mathbf{x}, t_n)$ and $c^n = c^n(\mathbf{x})$ to simplify the notation. We define the time discrete error as

$$e^n = c(t_n) - c^n. \tag{4.30}$$

In what follows, we use C to represent a generic positive constant independent of the mesh size and time step seize, but probably having different values in different occurrences.

THEOREM 4.4. *Assume that the exact solution $c(\mathbf{x}, t)$ is sufficiently smooth and bounded in both time and space and the functions $\nu(c)$ and $s_r(c)$ are Lipschitz continuous for $c_m \le c \le c_M$. Under the same conditions as in Theorem* 4.3*, we have*

$$\|e^{n+1}\| \le C\tau, \quad \mu_e(t_{n+1}) - \mu_e^{n+1} \le C\tau. \tag{4.31}$$

Proof. It is deduced from (2.11) and (4.3) that

$$\frac{e^{n+1} - e^n}{\tau} - \kappa\Delta e^{n+1} + \nu(c^n)e^{n+1} = \mu_e(t_{n+1}) - \mu_e^{n+1} + \sum_{i=1}^{3} \Lambda_i, \tag{4.32a}$$

$$\int_\Omega e^{n+1} d\mathbf{x} = 0, \tag{4.32b}$$

where

$$\Lambda_1 = \frac{c(t_{n+1}) - c(t_n)}{\tau} - \frac{\partial c}{\partial t}(t_{n+1}), \tag{4.33}$$

$$\Lambda_2 = \nu(c^n)c(t_{n+1}) - \nu(c(t_{n+1}))c(t_{n+1}), \tag{4.34}$$

$$\Lambda_3 = s_r\left(c(t_{n+1})\right) - s_r(c^n). \tag{4.35}$$

Multiplying (4.32a) by e^{n+1} and then integrating it over Ω, we get

$$\begin{aligned} &\frac{1}{\tau}\left(e^{n+1} - e^n, e^{n+1}\right) - \kappa\left(\Delta e^{n+1}, e^{n+1}\right) + \left(\nu(c^n)e^{n+1}, e^{n+1}\right) \\ &\qquad = \left(\mu_e(t_{n+1}) - \mu_e^{n+1}, e^{n+1}\right) + \sum_{i=1}^{3}\left(\Lambda_i, e^{n+1}\right). \end{aligned} \tag{4.36}$$

The terms on the left-hand side of (4.36) can be estimated as

$$\left(e^{n+1} - e^n, e^{n+1}\right) = \frac{1}{2}\left(\|e^{n+1}\|^2 - \|e^n\|^2 + \|e^{n+1} - e^n\|^2\right) \ge \frac{1}{2}\left(\|e^{n+1}\|^2 - \|e^n\|^2\right), \tag{4.37}$$

$$-\kappa\left(\Delta e^{n+1}, e^{n+1}\right) = \kappa\|\nabla e^{n+1}\|^2, \tag{4.38}$$

$$\left(\nu(c^n)e^{n+1}, e^{n+1}\right) = \|\nu(c^n)^{1/2}e^{n+1}\|^2. \tag{4.39}$$

Equation (4.32b) leads to $\left(\mu_e(t_{n+1}) - \mu_e^{n+1}, e^{n+1}\right) = 0$. Since

$$\|\Lambda_1\|^2 \le \frac{\tau}{3}\int_{t_n}^{t_{n+1}} \left\|\frac{\partial^2 c}{\partial t^2}\right\|^2 dt, \tag{4.40}$$

$$\begin{aligned}\|\Lambda_2\|^2 &\le 2\|\left(\nu(c^n) - \nu(c(t_n))\right)c(t_{n+1})\|^2 + 2\|\left(\nu(c(t_{n+1})) - \nu(c(t_n))\right)c(t_{n+1})\|^2\\ &\le C\|e^n\|^2 + C\tau\int_{t_n}^{t_{n+1}} \left\|\frac{\partial c}{\partial t}\right\|^2 dt, \end{aligned}\tag{4.41}$$

$$\|\Lambda_3\|^2 \le C\|e^n\|^2 + C\tau\int_{t_n}^{t_{n+1}} \left\|\frac{\partial c}{\partial t}\right\|^2 dt, \tag{4.42}$$

we deduce that

$$\sum_{i=1}^{3}\left(\Lambda_i, e^{n+1}\right) \le C\left(\|e^n\|^2 + \|e^{n+1}\|^2 + \tau\right). \tag{4.43}$$

Substituting (4.37)–(4.43) into (4.36) and then summing it, we obtain

$$\frac{1}{2}\|e^{n+1}\|^2 + \tau\sum_{i=0}^{n}\kappa\|\nabla e^{i+1}\|^2 \le C\tau\sum_{i=0}^{n}\|e^{i+1}\|^2 + C\tau^2. \tag{4.44}$$

Thus, the first estimate of (4.31) is deduced from (4.44) using the discrete Gronwall inequality. At last, we deduce that

$$\begin{aligned}\mu_e(t_{n+1}) - \mu_e^{n+1} &= \frac{1}{|\Omega|}\int_\Omega \left(\nu(c(t_{n+1}))c(t_{n+1}) - s_r(c(t_{n+1})) - \nu(c^n)c^{n+1} + s_r(c^n)\right)d\mathbf{x}\\ &\le \frac{1}{|\Omega|^{1/2}}\|\nu(c(t_{n+1}))c(t_{n+1}) - s_r(c(t_{n+1})) - \nu(c^n)c^{n+1} + s_r(c^n)\|\\ &\le C\left(\|e^{n+1}\| + \|e^n\| + \tau\right) \le C\tau, \end{aligned}\tag{4.45}$$

where we have used the first estimate of (4.31) and C is independent of τ. ∎

5. Fully discrete scheme. In this section, we consider the fully discrete scheme. The cell-centered finite difference (CCFD) method [41] is employed as the spatial discretization method. We note that the CCFD method is equivalent to a mixed finite element method with quadrature rules [1]. Here, we present the numerical scheme in the two-dimensional case only, but it is straightforward to extend it to the three-dimensional case.

We consider a rectangular domain as $\Omega = [0, l_x] \times [0, l_y]$, where $l_x > 0$ and $l_y > 0$. For simplicity, a uniform mesh of Ω is used as $0 = x_0 < x_1 < \cdots < x_N = l_x$ and $0 = y_0 < y_1 < \cdots < y_M = l_y$, where N and M are integers. We also introduce the intermediate points $x_{i+\frac{1}{2}} = \frac{x_i + x_{i+1}}{2}$ and $y_{j+\frac{1}{2}} = \frac{y_j + y_{j+1}}{2}$. The mesh size is denoted by $h = x_{i+1} - x_i = y_{j+1} - y_j$.

5.1. Notations and fully discrete scheme. To formulate the fully discrete scheme, we define the following discrete function spaces:

$$\mathcal{V}_c = \left\{c : (x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}) \mapsto \mathbb{R}, \;\; 0 \le i \le N-1, \; 0 \le j \le M-1\right\},$$

$$\mathcal{V}_u = \big\{u : (x_i, y_{j+\frac{1}{2}}) \mapsto \mathbb{R}, \ \ 0 \le i \le N, \ 0 \le j \le M-1\big\},$$

$$\mathcal{V}_v = \big\{v : (x_{i+\frac{1}{2}}, y_j) \mapsto \mathbb{R}, \ \ 0 \le i \le N-1, \ 0 \le j \le M\big\}.$$

For components of discrete functions in the above spaces, we denote $c_{i+\frac{1}{2},j+\frac{1}{2}} = c(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}})$ for $c \in \mathcal{V}_c$, $u_{i,j+\frac{1}{2}} = u(x_i, y_{j+\frac{1}{2}})$ for $u \in \mathcal{V}_u$ and $v_{i+\frac{1}{2},j} = v(x_{i+\frac{1}{2}}, y_j)$ for $v \in \mathcal{V}_v$.

For $c \in \mathcal{V}_c$, we define the difference operators $\delta_x^c[c] \in \mathcal{V}_u$ and $\delta_y^c[c] \in \mathcal{V}_v$ as follows:

$$\delta_x^c[c]_{i,j+\frac{1}{2}} = \frac{c_{i+\frac{1}{2},j+\frac{1}{2}} - c_{i-\frac{1}{2},j+\frac{1}{2}}}{h}, \ \ 1 \le i \le N-1, \ 0 \le j \le M-1, \tag{5.1a}$$

$$\delta_y^c[c]_{i+\frac{1}{2},j} = \frac{c_{i+\frac{1}{2},j+\frac{1}{2}} - c_{i+\frac{1}{2},j-\frac{1}{2}}}{h}, \ \ 0 \le i \le N-1, \ 1 \le j \le M-1. \tag{5.1b}$$

On the boundary, applying the homogeneous Neumann boundary condition, we take the difference operators as

$$\delta_x^c[c]_{i,j+\frac{1}{2}} = 0, \ \ i \in \{0, N\}, \ 0 \le j \le M-1, \tag{5.2a}$$

$$\delta_y^c[c]_{i+\frac{1}{2},j} = 0, \ \ j \in \{0, M\}, \ 0 \le i \le N-1. \tag{5.2b}$$

We introduce the subsets of $\mathcal{V}_u$ and $\mathcal{V}_v$ involving the boundary condition as

$$\mathcal{V}_u^0 = \big\{u \in \mathcal{V}_u |\ u_{0,j+\frac{1}{2}} = u_{N,j+\frac{1}{2}} = 0, \ \ 0 \le j \le M-1\big\}, \tag{5.3}$$

$$\mathcal{V}_v^0 = \big\{v \in \mathcal{V}_v |\ v_{i+\frac{1}{2},0} = v_{i+\frac{1}{2},M} = 0, \ \ 0 \le i \le N-1\big\}. \tag{5.4}$$

Apparently, $\delta_x^c[c] \in \mathcal{V}_u^0$ and $\delta_y^c[c] \in \mathcal{V}_v^0$.

The difference operators for $u \in \mathcal{V}_u$ and $v \in \mathcal{V}_v$ are defined as

$$\delta_x^u[u]_{i+\frac{1}{2},j+\frac{1}{2}} = \frac{u_{i+1,j+\frac{1}{2}} - u_{i,j+\frac{1}{2}}}{h}, \tag{5.5}$$

$$\delta_y^v[v]_{i+\frac{1}{2},j+\frac{1}{2}} = \frac{v_{i+\frac{1}{2},j+1} - v_{i+\frac{1}{2},j}}{h}, \tag{5.6}$$

where $0 \le i \le N-1, \ 0 \le j \le M-1$.

We define the following discrete inner products:

$$\langle c, c' \rangle = h^2 \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} c_{i+\frac{1}{2},j+\frac{1}{2}} c'_{i+\frac{1}{2},j+\frac{1}{2}}, \quad c, c' \in \mathcal{V}_c,$$

$$\langle u, u' \rangle = h^2 \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} u_{i,j+\frac{1}{2}} u'_{i,j+\frac{1}{2}}, \quad u, u' \in \mathcal{V}_u^0,$$

$$\langle v, v' \rangle = h^2 \sum_{i=0}^{N-1} \sum_{j=1}^{M-1} v_{i+\frac{1}{2},j} v'_{i+\frac{1}{2},j}, \quad v, v' \in \mathcal{V}_v^0.$$

The discrete norms for $c \in \mathcal{V}_c$, $u \in \mathcal{V}_u^0$, and $v \in \mathcal{V}_v^0$ are denoted

$$\|c\| = \langle c, c\rangle^{1/2}, \quad \|u\| = \langle u, u\rangle^{1/2}, \quad \|v\| = \langle v, v\rangle^{1/2}.$$

The semi-implicit fully discrete scheme is stated as follows: Given $c^n \in \mathcal{V}_c$, find $c^{n+1} \in \mathcal{V}_c$ such that

$$\frac{c^{n+1} - n^n}{\tau} - \kappa\left(\delta_x^u[\delta_x^c[c^{n+1}]] + \delta_y^v[\delta_y^c[c^{n+1}]]\right) + \nu(c^n)c^{n+1} = s_r(c^n) + \mu_e^{n+1}, \tag{5.7a}$$

$$\langle c^{n+1}, 1\rangle = c_t, \tag{5.7b}$$

where the formulations of functions $\nu(c)$ and $s_r(c)$ are given in (4.4) and (4.5), respectively. We note that the boundary condition has already been considered in the definitions of the operators δ_x^c and δ_y^c in (5.1).

The following summation-by-parts formulas are derived by direct calculations [4, 23, 43, 44]:

$$\langle u, \delta_x^c[c]\rangle = -\langle \delta_x^u[u], c\rangle, \quad u \in \mathcal{V}_u^0,\ c \in \mathcal{V}_c, \tag{5.8}$$

$$\langle v, \delta_y^c[c]\rangle = -\langle \delta_y^v[v], c\rangle, \quad v \in \mathcal{V}_v^0,\ c \in \mathcal{V}_c. \tag{5.9}$$

5.2. Unique solvability. We first demonstrate the unique solvability of the linear system (5.7).

THEOREM 5.1. *Assume that $c^n \in \mathcal{V}_c$ satisfies* (2.7) *and λ is taken to satisfy* (3.9). *There exists a unique $c^{n+1} \in \mathcal{V}_c$ such that* (5.7) *holds.*

Proof. It suffices to prove that the following homogeneous problem has a unique zero solution in $\mathcal{V}_c$:

$$\frac{c}{\tau} - \kappa\left(\delta_x^u[\delta_x^c[c]] + \delta_y^v[\delta_y^c[c]]\right) + \nu(c^n)c = \mu_e, \tag{5.10a}$$

$$\langle c, 1\rangle = 0. \tag{5.10b}$$

Suppose that there exists a nonzero solution $c \in \mathcal{V}_c$ satisfying (5.10). We take the inner product of (5.10) with c,

$$\frac{1}{\tau}\|c\|^2 - \kappa\langle \delta_x^u[\delta_x^c[c]], c\rangle - \kappa\langle \delta_y^v[\delta_y^c[c]], c\rangle + \langle \nu(c^n)c, c\rangle = \langle \mu_e, c\rangle, \tag{5.11}$$

which, using (5.8), (5.9), and (5.10b), can be further reduced to

$$\frac{1}{\tau}\|c\|^2 + \kappa\left(\|\delta_x^c[c]\|^2 + \|\delta_y^c[c]\|^2\right) + \nu(c_M)\|c\|^2 \le 0. \tag{5.12}$$

Since $\nu(c_M) > 0$, we conclude from (5.12) that $c \equiv 0$. This yields a contradiction, and thus (5.7) is uniquely solvable. ∎

5.3. Discrete maximum principle. We are going to prove the discrete maximum principle of molar density. The following lemma is an essential ingredient of the proof.

LEMMA 5.1. *Let* $c^- = \min(c - c_m, 0)$ *and* $c^+ = \max(c - c_M, 0)$*, where* $c \in \mathcal{V}_c$ *and* $c_m < c_M$*. Then we have*

$$\langle \delta_x^c[c^-], \delta_x^c[c^-] \rangle \le -\langle \delta_x^u[\delta_x^c[c]], c^- \rangle, \tag{5.13}$$

$$\langle \delta_x^c[c^+], \delta_x^c[c^+] \rangle \le -\langle \delta_x^u[\delta_x^c[c]], c^+ \rangle, \tag{5.14}$$

$$\langle \delta_y^c[c^-], \delta_y^c[c^-] \rangle \le -\langle \delta_y^v[\delta_y^c[c]], c^- \rangle, \tag{5.15}$$

$$\langle \delta_y^c[c^+], \delta_y^c[c^+] \rangle \le -\langle \delta_y^v[\delta_y^c[c]], c^+ \rangle. \tag{5.16}$$

Proof. Let a and b be two real scalar numbers, and we further define $a^- = \min(a - c_m, 0)$ and $b^- = \min(b - c_m, 0)$. It is apparent that

$$(a - c_m)\, a^- = |a^-|^2, \quad (b - c_m)\, b^- = |b^-|^2. \tag{5.17}$$

Since $a - c_m \ge a^-$ and $b^- \le 0$, we get

$$(a - c_m)\, b^- \le a^- b^-. \tag{5.18}$$

Applying (5.8), (5.17), and (5.18), we can derive inequality (5.13) as follows:

$$\begin{aligned}
-\langle \delta_x^u[\delta_x^c[c]], c^- \rangle &= \langle \delta_x^c[c], \delta_x^c[c^-] \rangle \\
&= h^2 \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \delta_x^c[c]_{i,j+\frac{1}{2}} \delta_x^c[c^-]_{i,j+\frac{1}{2}} \\
&= \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \left(c_{i+\frac{1}{2},j+\frac{1}{2}} - c_{i-\frac{1}{2},j+\frac{1}{2}} \right) \left(c^-_{i+\frac{1}{2},j+\frac{1}{2}} - c^-_{i-\frac{1}{2},j+\frac{1}{2}} \right) \\
&\ge \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \left(|c^-_{i+\frac{1}{2},j+\frac{1}{2}}|^2 + |c^-_{i-\frac{1}{2},j+\frac{1}{2}}|^2 - 2c^-_{i+\frac{1}{2},j+\frac{1}{2}} c^-_{i-\frac{1}{2},j+\frac{1}{2}} \right) \\
&= \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \left(c^-_{i+\frac{1}{2},j+\frac{1}{2}} - c^-_{i-\frac{1}{2},j+\frac{1}{2}} \right)^2 \\
&= \langle \delta_x^c[c^-], \delta_x^c[c^-] \rangle. \qquad (5.19)
\end{aligned}$$

We now prove inequality (5.14). Let $a^+ = \max(a - c_M, 0)$ and $b^+ = \max(b - c_M, 0)$. Then we have

$$(a - c_M)\, a^+ = |a^+|^2, \quad (b - c_M)\, b^+ = |b^+|^2. \tag{5.20}$$

Since $a - c_M \le a^+$ and $b^+ \ge 0$, we get

$$(a - c_M)\, b^+ \le a^+ b^+. \tag{5.21}$$

Applying (5.8), (5.20), and (5.21), we deduce inequality (5.14) as

$$\begin{aligned}
-\langle \delta_x^u[\delta_x^c[c]], c^+\rangle &= \langle \delta_x^c[c], \delta_x^c[c^+]\rangle \\
&= h^2 \sum_{i=1}^{N-1}\sum_{j=0}^{M-1} \delta_x^c[c]_{i,j+\frac{1}{2}} \delta_x^c[c^+]_{i,j+\frac{1}{2}} \\
&= \sum_{i=1}^{N-1}\sum_{j=0}^{M-1} \left(c_{i+\frac{1}{2},j+\frac{1}{2}} - c_{i-\frac{1}{2},j+\frac{1}{2}}\right)\left(c^+_{i+\frac{1}{2},j+\frac{1}{2}} - c^+_{i-\frac{1}{2},j+\frac{1}{2}}\right) \\
&\geq \sum_{i=1}^{N-1}\sum_{j=0}^{M-1} \left(|c^+_{i+\frac{1}{2},j+\frac{1}{2}}|^2 + |c^+_{i-\frac{1}{2},j+\frac{1}{2}}|^2 - 2c^+_{i+\frac{1}{2},j+\frac{1}{2}} c^+_{i-\frac{1}{2},j+\frac{1}{2}}\right) \\
&= \langle \delta_x^c[c^+], \delta_x^c[c^+]\rangle. \qquad (5.22)
\end{aligned}$$

The remaining inequalities (5.15) and (5.16) can be proved by similar approaches. ∎

We are now ready to prove the maximum principle of the fully discrete scheme.

THEOREM 5.2. *Assume that $c^n \in \mathcal{V}_c$ satisfies* (2.7) *and λ is taken to satisfy* (3.9). *Under condition* (4.9), *we have $c_m \leq c^{n+1} \leq c_M$.*

Proof. It can be proved using routines similar to those in the proof of Theorem 4.2. The major difference is that we use Lemma 5.1 to treat the discrete form of the Laplace operator. So we just give a brief proof.

Assuming that $c_m \leq c^n \leq c_M$ for $n \geq 0$, we will prove that $c_m \leq c^{n+1} \leq c_M$. Let $c_-^{n+1} = \min(c^{n+1} - c_m, 0)$. It can be derived from (5.7a) that

$$\begin{aligned}
&\frac{1}{\tau}\langle c^{n+1} - c^n, c_-^{n+1}\rangle - \kappa\langle \delta_x^u[\delta_x^c[c^{n+1}]] + \delta_y^v[\delta_y^c[c^{n+1}]], c_-^{n+1}\rangle \\
&+ \langle \nu(c^n)(c^{n+1} - c_m), c_-^{n+1}\rangle = \langle \mu_e^{n+1}, c_-^{n+1}\rangle + \langle s_r(c^n) - c_m\nu(c^n), c_-^{n+1}\rangle. \qquad (5.23)
\end{aligned}$$

The second term on the left-hand side of (5.23) can be estimated using Lemma 5.1:

$$-\kappa\langle \delta_x^u[\delta_x^c[c^{n+1}]] + \delta_y^v[\delta_y^c[c^{n+1}]], c_-^{n+1}\rangle \geq \kappa\left(\|\delta_x^c[c_-^{n+1}]\|^2 + \|\delta_y^c[c_-^{n+1}]\|^2\right). \qquad (5.24)$$

Following routines similar to those used to deduce (4.15), we obtain

$$\frac{1}{\tau}\|c_-^{n+1}\|^2 + \kappa\left(\|\delta_x^c[c_-^{n+1}]\|^2 + \|\delta_y^c[c_-^{n+1}]\|^2\right) + \nu(c_M)\|c_-^{n+1}\|^2 \leq 0, \qquad (5.25)$$

which yields $c_-^{n+1} = 0$, and thus $c^{n+1} \geq c_m$. It is similar to prove $c^{n+1} \leq c_M$. ∎

5.4. Energy stability. We denote the discrete total free energy as

$$F_h(c^n) = \langle f_b(c^n), 1\rangle + \frac{1}{2}\kappa\left(\|\delta_x^c[c^n]\|^2 + \|\delta_y^c[c^n]\|^2\right). \qquad (5.26)$$

THEOREM 5.3. *Assume that c^n satisfies* (2.7) *and λ is taken to satisfy* (3.9). *Under condition* (4.9), *for any time step size τ, we have*

$$F_h(c^{n+1}) \leq F_h(c^n). \qquad (5.27)$$

Proof. Using (5.8), we can derive that

$$\begin{aligned}\frac{1}{2}\left(\|\delta_x^c[c^{n+1}]\|^2-\|\delta_x^c[c^n]\|^2\right)&=\frac{1}{2}\left(\langle\delta_x^c[c^{n+1}],\delta_x^c[c^{n+1}]\rangle-\langle\delta_x^c[c^n],\delta_x^c[c^n]\rangle\right)\\&=\langle\delta_x^c[c^{n+1}],\delta_x^c[c^{n+1}]-\delta_x^c[c^n]\rangle-\frac{1}{2}\|\delta_x^c[c^{n+1}]-\delta_x^c[c^n]\|^2\\&\le\langle\delta_x^c[c^{n+1}],\delta_x^c[c^{n+1}-c^n]\rangle\\&=-\langle\delta_x^u[\delta_x^c[c^{n+1}]],c^{n+1}-c^n\rangle.\end{aligned}\tag{5.28}$$

It is similar to deduce that

$$\frac{1}{2}\left(\|\delta_y^c[c^{n+1}]\|^2-\|\delta_y^c[c^n]\|^2\right)\le-\langle\delta_y^v[\delta_y^c[c^{n+1}]],c^{n+1}-c^n\rangle.\tag{5.29}$$

Using arguments similar to those of (4.27) and the estimates (5.28) and (5.29), we obtain

$$\begin{aligned}&\langle f_b(c^{n+1})-f_b(c^n),1\rangle+\frac{1}{2}\kappa\left(\|\delta_x^c[c^{n+1}]\|^2-\|\delta_x^c[c^n]\|^2\right)\\&+\frac{1}{2}\kappa\left(\|\delta_y^c[c^{n+1}]\|^2-\|\delta_y^c[c^n]\|^2\right)\le-\frac{1}{\tau}\|c^{n+1}-c^n\|^2,\end{aligned}\tag{5.30}$$

which yields (5.27). ∎

5.5. Error estimate. We carry out the convergence analysis for the proposed fully discrete scheme and present the error estimate results. We define the discrete error $e^n\in\mathcal{V}_c$, which has the following component form:

$$e^n_{i+\frac12,j+\frac12}=c(x_{i+\frac12},y_{j+\frac12},t_n)-c^n_{i+\frac12,j+\frac12}.\tag{5.31}$$

THEOREM 5.4. *Assume that the exact solution $c(\mathbf{x},t)$ is sufficiently smooth and bounded in both time and space and the functions $\nu(c)$ and $s_r(c)$ are Lipschitz continuous for $c_m\le c\le c_M$. Under the same conditions as in Theorem* 4.3*, we have*

$$\|e^{n+1}\|\le C\left(h+\tau\right),\quad|\mu_e(t_{n+1})-\mu_e^{n+1}|\le C\left(h+\tau\right),\tag{5.32}$$

where the values of C are independent of h and τ.

Proof. From (2.11) and (5.7), we can get

$$\begin{aligned}&\frac{e^{n+1}_{i+\frac12,j+\frac12}-e^n_{i+\frac12,j+\frac12}}{\tau}-\kappa\delta_x^u[\delta_x^c[e^{n+1}_{i+\frac12,j+\frac12}]]-\kappa\delta_y^v[\delta_y^c[e^{n+1}_{i+\frac12,j+\frac12}]]+\nu(c^n_{i+\frac12,j+\frac12})e^{n+1}_{i+\frac12,j+\frac12}\\&=\mu_e(t_{n+1})-\mu_e^{n+1}+\Lambda_{i+\frac12,j+\frac12}+\Upsilon_{i+\frac12,j+\frac12}+\Phi_{i+\frac12,j+\frac12}+\Psi_{i+\frac12,j+\frac12},\end{aligned}\tag{5.33a}$$

$$\langle e^{n+1},1\rangle=\sum_{i,j}\int_{\Omega_{i,j}}\Big(c(x_{i+\frac12},y_{j+\frac12},t_{n+1})-c(\mathbf{x},t_{n+1})\Big)\,d\mathbf{x},\tag{5.33b}$$

where $\Omega_{i,j}=[x_i,x_{i+1}]\times[y_j,y_{j+1}]$,

$$\Lambda^{n+1}_{i+\frac12,j+\frac12}=\frac{c(x_{i+\frac12},y_{j+\frac12},t_{n+1})-c(x_{i+\frac12},y_{j+\frac12},t_n)}{\tau}-\frac{\partial c}{\partial t}(x_{i+\frac12},y_{j+\frac12},t_{n+1}),\tag{5.34}$$

$$\Upsilon^{n+1}_{i+\frac12,j+\frac12}=\kappa\frac{\partial^2c}{\partial x^2}(x_{i+\frac12},y_{j+\frac12},t_{n+1})-\kappa\delta_x^u[\delta_x^c[c(x_{i+\frac12},y_{j+\frac12},t_{n+1})]],\tag{5.35}$$

$$\Phi^{n+1}_{i+\frac{1}{2},j+\frac{1}{2}} = \kappa\frac{\partial^2 c}{\partial y^2}(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1}) - \kappa\delta_y^v[\delta_y^c[c(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1})]], \tag{5.36}$$

$$\begin{aligned}\Psi^{n+1}_{i+\frac{1}{2},j+\frac{1}{2}} = &\left(\nu(c^n_{i+\frac{1}{2},j+\frac{1}{2}}) - \nu(c(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1}))\right) c(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1})\\ &+ s_r(c(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1})) - s_r(c^n_{i+\frac{1}{2},j+\frac{1}{2}}).\end{aligned} \tag{5.37}$$

Multiplying (5.33a) by $e^{n+1}_{i,j}$ and then taking the discrete inner product, we can deduce that

$$\begin{aligned}&\frac{1}{2\tau}\left(\|e^{n+1}\|^2 - \|e^n\|^2\right) + \kappa\|\delta_x^c[e^{n+1}]\|^2 + \kappa\|\delta_y^c[e^{n+1}]\|^2 + \|\nu(c^n)^{1/2}e^{n+1}\|^2\\ &\le \langle\mu_e(t_{n+1}) - \mu_e^{n+1}, e^{n+1}\rangle + \langle\Lambda^{n+1} + \Upsilon^{n+1} + \Phi^{n+1} + \Psi^{n+1}, e^{n+1}\rangle,\end{aligned} \tag{5.38}$$

the left-hand side of which is derived using approaches similar to those in Theorem 4.4 and the discrete properties. We now consider the terms on the right-hand side of (5.38). From (5.33b), we have the estimate

$$|\langle e^{n+1}, 1\rangle| \le \sum_{i,j}\left|\int_{\Omega_{i,j}} c(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1}) - c(\mathbf{x}, t_{n+1})d\mathbf{x}\right| \le Ch, \tag{5.39}$$

where C is independent of h, but depends on the exact solution. Similar to (4.45), using the discrete properties, we can deduce

$$\begin{aligned}|\mu_e(t_{n+1}) - \mu_e^{n+1}| &\le \frac{1}{|\Omega|}\sum_{i,j}\int_{\Omega_{i,j}}\Big(\left|\nu(c(\mathbf{x}, t_{n+1}))c(\mathbf{x}, t_{n+1}) - \nu(c^n_{i+\frac{1}{2},j+\frac{1}{2}})c^{n+1}_{i+\frac{1}{2},j+\frac{1}{2}}\right|\\ &\quad + \left|s_r(c(\mathbf{x}, t_{n+1})) - s_r(c^n_{i+\frac{1}{2},j+\frac{1}{2}})\right|\Big)d\mathbf{x}\\ &\le C\left(\|e^{n+1}\| + \|e^n\| + h + \tau\right).\end{aligned} \tag{5.40}$$

The estimates (5.39) and (5.40) yield

$$\langle\mu_e(t_{n+1}) - \mu_e^{n+1}, e^{n+1}\rangle \le C\left(\|e^{n+1}\|^2 + \|e^n\|^2 + h^2 + \tau^2\right). \tag{5.41}$$

Applying the Taylor expansion, we can obtain

$$c(x_{i+\frac{3}{2}}, y_{j+\frac{1}{2}}, t_{n+1}) = c_* + h\frac{\partial c_*}{\partial x} + \frac{1}{2}h^2\frac{\partial^2 c_*}{\partial x^2} + \frac{1}{6}h^3\frac{\partial^3 c_*}{\partial x^3} + \frac{1}{24}h^4\frac{\partial^4 \check{c}}{\partial x^4}, \tag{5.42}$$

$$c(x_{i-\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1}) = c_* - h\frac{\partial c_*}{\partial x} + \frac{1}{2}h^2\frac{\partial^2 c_*}{\partial x^2} - \frac{1}{6}h^3\frac{\partial^3 c_*}{\partial x^3} + \frac{1}{24}h^4\frac{\partial^4 \hat{c}}{\partial x^4}, \tag{5.43}$$

where $\frac{\partial^k c_*}{\partial x^k} = \frac{\partial^k c}{\partial x^k}(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1})$, $0 \le k \le 3$, and $\frac{\partial^4 \check{c}}{\partial x^4}$ and $\frac{\partial^4 \hat{c}}{\partial x^4}$ denote the derivatives at some points between $(x_{i-\frac{1}{2}}, y_{j+\frac{1}{2}}, t_{n+1})$ and $(x_{i+\frac{3}{2}}, y_{j+\frac{1}{2}}, t_{n+1})$. Equations (5.42) and (5.43) yield

$$\Upsilon^{n+1}_{i+\frac{1}{2},j+\frac{1}{2}} = \frac{1}{24}h^2\kappa\left(\frac{\partial^4 \hat{c}}{\partial x^4} + \frac{\partial^4 \check{c}}{\partial x^4}\right) \le \frac{1}{12}h^2\kappa\left\|\frac{\partial^4 c}{\partial x^4}\right\|_{L^\infty}, \quad 1 \le i \le N-2,\ 0 \le j \le M-1. \tag{5.44}$$

On the west boundary, i.e., $i = 0$, we have

$$0 = \frac{\partial c}{\partial x}(x_0, y_{j+\frac{1}{2}}, t_{n+1}) = \frac{\partial c_*}{\partial x} - \frac{1}{2}h\frac{\partial^2 c_*}{\partial x^2} + \frac{1}{8}h^2\frac{\partial^3 c_*}{\partial x^3} - \frac{1}{48}h^3\frac{\partial^4 \bar{c}}{\partial x^4}. \tag{5.45}$$

The Taylor expansion on the east boundary is similar, so we get

$$\Upsilon^{n+1}_{i+\frac{1}{2},j+\frac{1}{2}} \le \frac{7}{24}h\kappa \left\| \frac{\partial^3 c}{\partial x^3} \right\|_{L^\infty}, \quad i = 0, N-1,\ 0 \le j \le M-1. \tag{5.46}$$

For $\Phi^{n+1}_{i+\frac{1}{2},j+\frac{1}{2}}$, we can derive estimates similar to (5.44) and (5.46), and thus

$$\|\Upsilon^{n+1}\| \le Ch, \quad \|\Phi^{n+1}\| \le Ch. \tag{5.47}$$

Using approaches similar to those in Theorem 4.4, we deduce that

$$\|\Lambda\| \le C\tau, \quad \|\Psi\| \le C\left(\|e^n\| + \|e^{n+1}\| + \tau\right). \tag{5.48}$$

Combining the above estimates, we obtain

$$\frac{1}{2}\|e^{n+1}\|^2 + \tau\sum_{i=0}^{n}\kappa\left(\|\delta_x^c[e^{n+1}]\|^2 + \|\delta_y^c[e^{n+1}]\|^2\right) \le C\tau\sum_{i=0}^{n}\|e^{i+1}\|^2 + C\left(h^2+\tau^2\right). \tag{5.49}$$

Thus, the first inequality of (5.32) can be deduced from (5.49) using the discrete Gronwall inequality, and subsequently the second inequality can be obtained from the first inequality and (5.40). ∎

We remark that from the proof of Theorem 5.4, we can see that the spatial error can be improved to second order if the periodic boundary condition is applied.

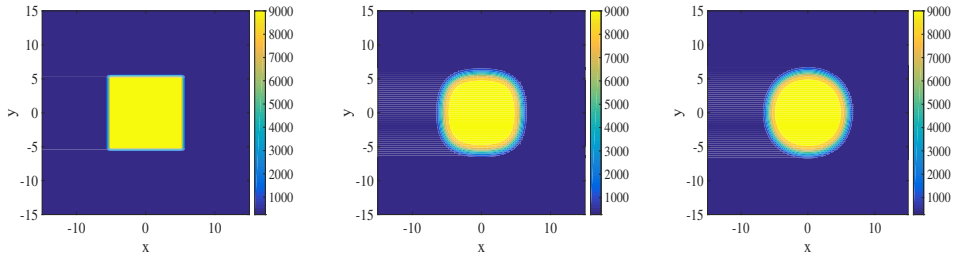
6. Numerical results. In this section, we present some numerical results to show the performance of the proposed method and verify the theoretical analysis. We consider a hydrocarbon substance, n-butane (nC_4). The related physical data is provided in Table 6.1. The temperature is fixed at 330 K. The value of ϑ_0 has no effects for the isothermal fluids, so we take $\vartheta_0 = 0$. In numerical tests, we use the gas molar density $c^G = 249.1123$ mol/m^3 and the liquid molar density $c^L = 9526.8428$ mol/m^3 to initialize the distribution of molar density. We note that mass transfer between two phases may take place in a dynamical process; that is, the gas phase may be condensed into the liquid phase, while the liquid phase may also be vaporized into the gas phase. As a result, we take $c_m = 0.9c^G$ and $c_M = 1.1c^L$ in (2.7) and calculate $\epsilon_0 = \beta c_M = 0.7585$. To satisfy condition (3.9), we use the value of ϵ_0 to calculate λ as follows:

$$\lambda = \frac{\epsilon_0}{(1-\epsilon_0)^2} + \left(\frac{\epsilon_0^2}{(1-\epsilon_0)^4} - 2\ln(1-\epsilon_0)\frac{\epsilon_0}{(1-\epsilon_0)^2}\right)^{1/2} = 27.3656. \tag{6.1}$$

In all of the numerical tests, the spatial domain is taken as $\Omega = [-L, L]^2$, where $L = 15$ nm, and a uniform rectangular mesh with 100×100 elements is employed. The proposed method admits the use of a very large time step size, so we take $\tau = 10^{10}$ s.

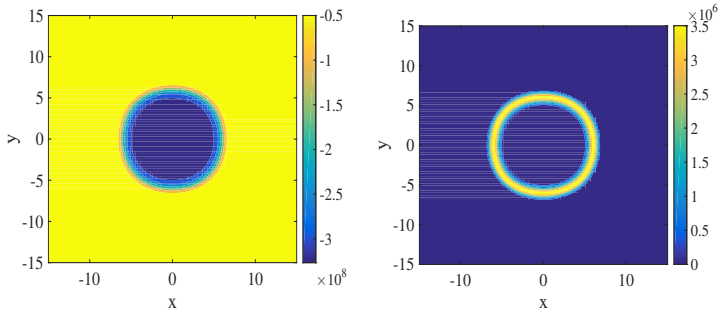
TABLE 6.1
Physical parameters of nC_4.

P_c(bar)	T_c(K)	ω
38.0	425.2	0.199



(a) Initial (b) $n = 100$ (c) $n = 200$

FIG. 6.1. *Example* 1: *Molar density distributions.*



(a) (b)

FIG. 6.2. *Example* 1: (a) *the bulk free energy density after* 200 *time steps;* (b) *the gradient free energy after* 200 *time steps.*

6.1. Example 1. In this example, we simulate the droplet shrinking problem for 200 time steps. In Figure 6.1, we show molar density distributions after different time steps, which demonstrate that the square droplet is changing to a circle shape. In Figure 6.2, we also illustrate the distributions of the bulk free energy density and gradient free energy density after 200 time steps.

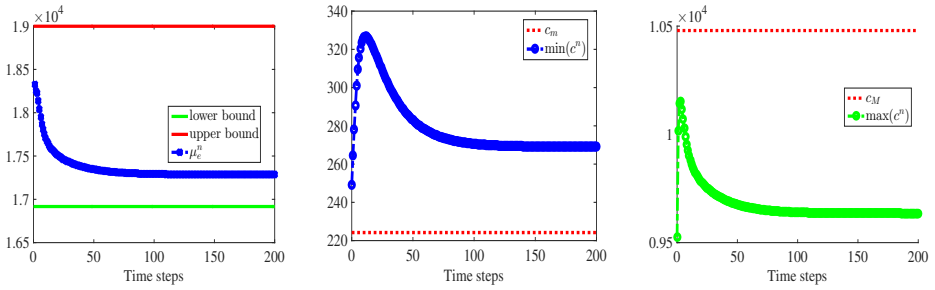
In Figure 6.3(a), the lower bound $\underline{\mu}$ and the upper bound $\bar{\mu}$ are calculated as

$$\underline{\mu} = \max_{c_m \le c \le c_M} \left(c_m \nu(c) - s_r(c)\right), \quad \bar{\mu} = \min_{c_m \le c \le c_M} \left(c_M \nu(c) - s_r(c)\right).$$

Figure 6.3(a) shows that μ_e^n always varies between $\underline{\mu}$ and $\bar{\mu}$, and thus condition (4.9) is invariably true. Figures 6.3(b) and 6.3(c) depict the minimum and maximum values of molar density c^n. Due to the effects of liquid vaporization and gas condensation, molar density changes a lot at the beginning period and then tends toward more steady values in the later period, but we observe that c^n always fluctuates between c_m and c_M; thus the maximum principle is verified.

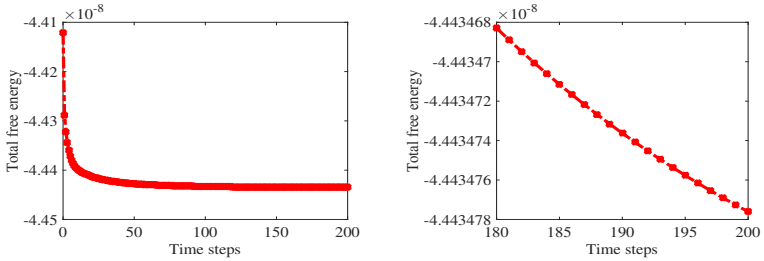
Figure 6.4(a) illustrates that the total energies are dissipated with time steps, while Figure 6.4(b) plots total energies at the last twenty time steps, which are still decreasing. Therefore, the proposed scheme can preserve the energy dissipation law.

6.2. Example 2. The initial molar density distribution of this example is illustrated in Figure 6.5(a). We simulate the dynamical process for 2000 time steps. Molar density distributions are illustrated in Figure 6.5, while the distributions of the



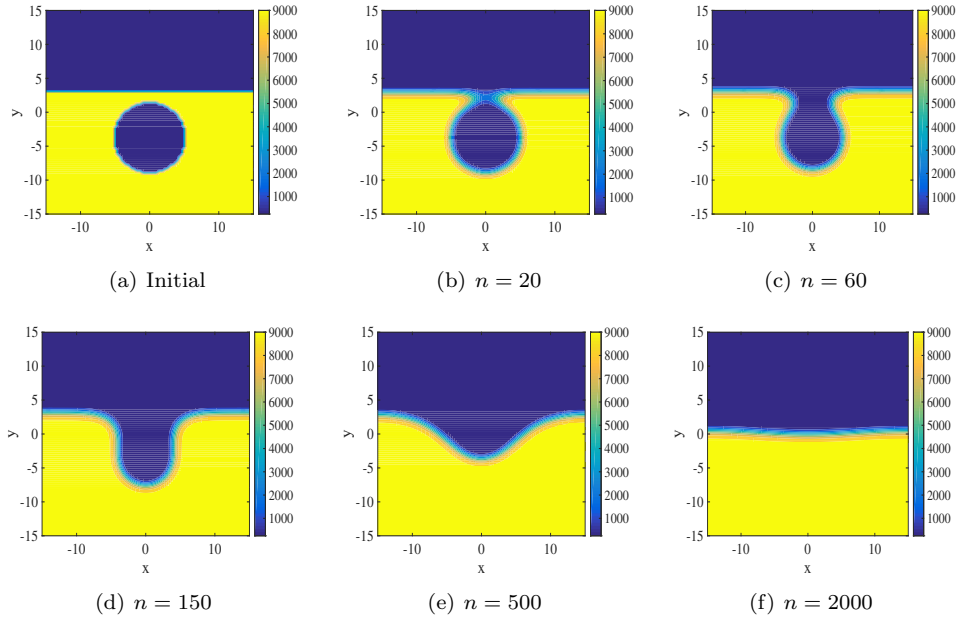
(a) Condition (4.9) (b) Minima of molar density (c) Maxima of molar density

FIG. 6.3. *Example* 1: *Verification of condition* (4.9) *and the maximum principle.*



(a) (b)

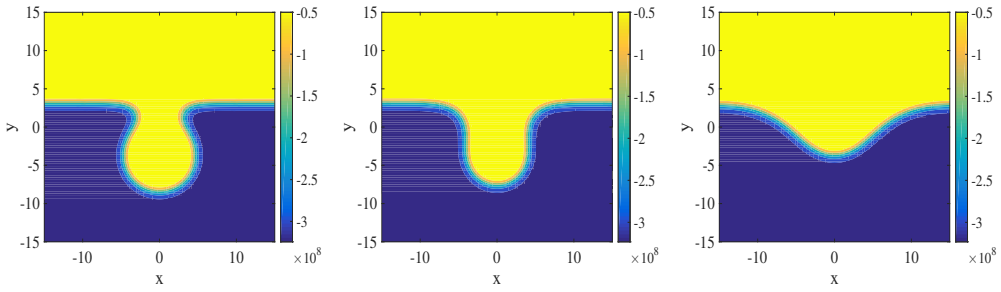
FIG. 6.4. *Example* 1: *Total energy profiles with time steps.*



(a) Initial (b) $n = 20$ (c) $n = 60$

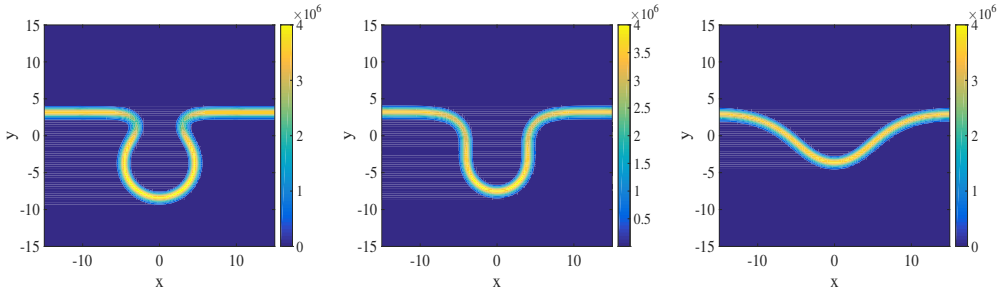
(d) $n = 150$ (e) $n = 500$ (f) $n = 2000$

FIG. 6.5. *Example* 2: *Molar density distributions.*



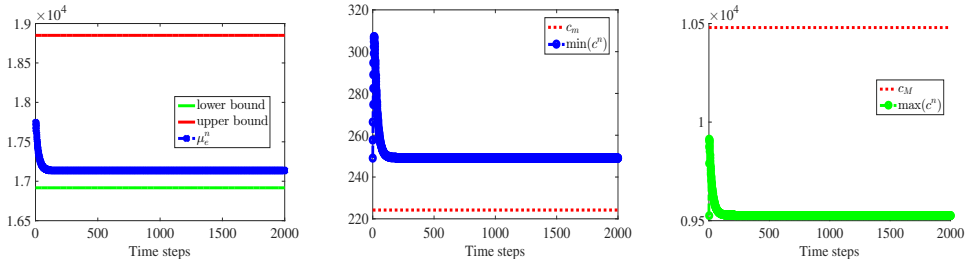
(a) $n = 60$ (b) $n = 150$ (c) $n = 500$

FIG. 6.6. *Example* 2: *The bulk free energy density distributions.*



(a) $n = 60$ (b) $n = 150$ (c) $n = 500$

FIG. 6.7. *Example* 2: *The gradient free energy density distributions.*



(a) Condition (4.9) (b) Minima of molar density (c) Maxima of molar density

FIG. 6.8. *Example* 2: *Verification of condition* (4.9) *and the maximum principle.*

free energy density and gradient free energy density are shown in Figures 6.6 and 6.7, respectively.

Figure 6.8(a) verifies the reasonability of condition (4.9) again, while Figures 6.8(b) and 6.8(c) validate the maximum principle of molar density. Moreover, Figure 6.9 shows that the total energies are always decreasing with time steps, even at the last twenty time steps.

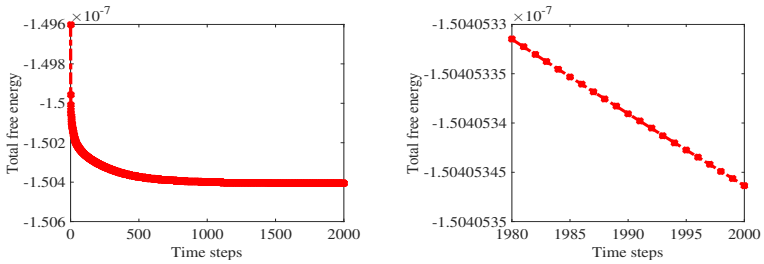


FIG. 6.9. *Example* 2: *Total energy profiles with time steps.*

7. Conclusions. A novel energy factorization (EF) approach has been proposed to construct the linear energy stable numerical scheme for the diffuse interface model with the Peng–Robinson equation of state, which is one of the most useful and prominent tools in chemical engineering and the petroleum industry. Compared with convex-splitting schemes, the semi-implicit numerical scheme constructed by the EF approach is linear and easy to implement. Compared with the IEQ/SAV schemes, the proposed method inherits the original energy dissipation law. Moreover, we prove that the maximum principle holds for both the time semidiscrete form and the cell-centered finite difference fully discrete form under certain conditions. Numerical results validate the stability and efficiency of the proposed scheme.

We remark that the proposed energy factorization approach in section 3 can be applied to Cahn–Hilliard-type equations [20], and numerical results (not presented here) demonstrate that the maximum principle and unconditional energy stability are still validated numerically. The maximum principle of the Allen–Cahn-type equation can be proved in proper conditions, but similar theoretical analysis may be not extended to the Cahn–Hilliard equation since the maximum principle of the Cahn–Hilliard equation is not available in theory. Certainly, we can prove unconditional energy stability of the proposed scheme under the boundedness condition of molar density, which is similar to the convex splitting approach [36].

In future work, we will study applications of the EF approach to multicomponent fluids and phase-field models. There have been quite a few works of second order, energy stable numerical schemes for the phase-field models, for example, [3, 11, 17, 47]. In other forthcoming work, we will study high-order time schemes based on the EF approach.

REFERENCES

[1] T. ARBOGAST, M. F. WHEELER, AND I. YOTOV, *Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences*, SIAM J. Numer. Anal., 34 (1997), pp. 828–852, https://doi.org/10.1137/S0036142994262585.

[2] A. BASKARAN, J. S. LOWENGRUB, C. WANG, AND S. M. WISE, *Convergence analysis of a second order convex splitting scheme for the modified phase field crystal equation*, SIAM J. Numer. Anal., 51 (2013), pp. 2851–2873, https://doi.org/10.1137/120880677.

[3] W. CHEN, C. WANG, X. WANG, AND S. M. WISE, *Positivity-preserving, energy stable numerical schemes for the Cahn-Hilliard equation with logarithmic potential*, J. Comput. Phys. X, 3 (2019), 100031.

[4] Y. CHEN AND J. SHEN, *Efficient, adaptive energy stable schemes for the incompressible Cahn-Hilliard Navier-Stokes phase-field models*, J. Comput. Phys., 308 (2016), pp. 40–56.

[5] M. I. M. Copetti and C. M. Elliott, *Numerical analysis of the Cahn-Hilliard equation with a logarithmic free energy*, Numer. Math., 63 (1992), pp. 39–65.
[6] C. M. Elliott and A. M. Stuart, *The global dynamics of discrete semilinear parabolic equations*, SIAM J. Numer. Anal., 30 (1993), pp. 1622–1663, https://doi.org/10.1137/0730084.
[7] D. J. Eyre, *Unconditionally gradient stable time marching the Cahn-Hilliard equation*, in Computational and Mathematical Models of Microstructural Evolution (San Francisco, CA, 1998), Mater. Res. Soc. Sympos. Proc. 529, MRS, Warrendale, PA, 1998, pp. 39–46.
[8] X. Fan, J. Kou, Z. Qiao, and S. Sun, *A componentwise convex splitting scheme for diffuse interface models with Van der Waals and Peng–Robinson equations of state*, SIAM J. Sci. Comput., 39 (2017), pp. B1–B28, https://doi.org/10.1137/16M1061552.
[9] D. Furihata, *A stable and conservative finite difference scheme for the Cahn-Hilliard equation*, Numer. Math., 87 (2001), pp. 675–699.
[10] A. Firoozabadi, *Thermodynamics of Hydrocarbon Reservoirs*, McGraw-Hill, New York, 1999.
[11] H. Gomez and T. J. R. Hughes, *Provably unconditionally stable, second-order time-accurate, mixed variational methods for phase-field models*, J. Comput. Phys., 230 (2011), pp. 5310–5327.
[12] H. Gomez, V. M. Calo, Y. Bazilevs, and T. J. R. Hughes, *Isogeometric analysis of the Cahn-Hilliard phase-field model*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 4333–4352.
[13] Z. Hu, S. M. Wise, C. Wang, and J. S. Lowengrub, *Stable and efficient finite-difference nonlinear-multigrid schemes for the phase field crystal equation*, J. Comput. Phys., 228 (2009), pp. 5323–5339.
[14] T. Jindrová and J. Mikyška, *Fast and robust algorithm for calculation of two-phase equilibria at given volume, temperature, and moles*, Fluid Phase Equilibria, 353 (2013), pp. 101–114.
[15] T. Jindrová and J. Mikyška, *General algorithm for multiphase equilibria calculation at given volume, temperature, and moles*, Fluid Phase Equilibria, 393 (2015), pp. 7–25.
[16] M. Kästner, P. Metsch, and R. de Borst, *Isogeometric analysis of the Cahn-Hilliard equation: A convergence study*, J. Comput. Phys., 305 (2016), pp. 360–371.
[17] N. Khiari, T. Achouri, M. Ben Mohamed, and K. Omrani, *Finite difference approximate solutions for the Cahn-Hilliard equation*, Numer. Methods Partial Differential Equations, 23 (2007), pp. 437–455.
[18] J. Kou, S. Sun, and X. Wang, *Efficient numerical methods for simulating surface tension of multi-component mixtures with the gradient theory of fluid interfaces*, Comput. Methods Appl. Mech. Engrg., 292 (2015), pp. 92–106.
[19] J. Kou and S. Sun, *Numerical methods for a multicomponent two-phase interface model with geometric mean influence parameters*, SIAM J. Sci. Comput., 37 (2015), pp. B543–B569, https://doi.org/10.1137/140969579.
[20] J. Kou and S. Sun, *Unconditionally stable methods for simulating multi-component two-phase interface models with Peng-Robinson equation of state and various boundary conditions*, J. Comput. Appl. Math., 291 (2016), pp. 158–182.
[21] J. Kou and S. Sun, *A stable algorithm for calculating phase equilibria with capillarity at specified moles, volume and temperature using a dynamic model*, Fluid Phase Equilibria, 456 (2018), pp. 7–24.
[22] J. Kou and S. Sun, *Multi-scale diffuse interface modeling of multi-component two-phase flow with partial miscibility*, J. Comput. Phys., 318 (2016), pp. 349–372.
[23] J. Kou, S. Sun, and X. Wang, *Linearly decoupled energy-stable numerical methods for multi-component two-phase compressible flow*, SIAM J. Numer. Anal., 56 (2018), pp. 3219–3248, https://doi.org/10.1137/17M1162287.
[24] J. Kou and S. Sun, *Thermodynamically consistent modeling and simulation of multi-component two-phase flow with partial miscibility*, Comput. Methods Appl. Mech. Engrg., 331 (2018), pp. 623–649.
[25] J. Kou and S. Sun, *Thermodynamically consistent simulation of nonisothermal diffuse-interface two-phase flow with Peng-Robinson equation of state*, J. Comput. Phys., 371 (2018), pp. 581–605.
[26] J. Kou and S. Sun, *Entropy stable modeling of non-isothermal multi-component diffuse-interface two-phase flows with realistic equations of state*, Comput. Methods Appl. Mech. Engrg., 341 (2018), pp. 221–248.
[27] H. Li, L. Ju, C. Zhang, and Q. Peng, *Unconditionally energy stable linear schemes for the diffuse interface model with Peng-Robinson equation of state*, J. Sci. Comput., 75 (2018), pp. 993–1015.
[28] J. S. Lopez-Echeverry, S. Reif-Acherman, and E. Araujo-Lopez, *Peng-Robinson equation of state: 40 years through cubics*, Fluid Phase Equilibria, 447 (2017), pp. 39–71.

[29] M. L. Michelsen, *State function based flash specifications*, Fluid Phase Equilibria, 158-160 (1999), pp. 617–626.
[30] J. Mikyška and A. Firoozabadi, *A new thermodynamic function for phase-splitting at constant temperature, moles, and volume*, AIChE J., 57 (2011), pp. 1897–1904.
[31] J. Mikyška, *A collection of analytical solutions for the flash equilibrium calculation problem*, Transp. Porous Media, 126 (2019), pp. 683–699, https://doi.org/10.1007/s11242-018-1160-9.
[32] C. Miqueu, B. Mendiboure, C. Graciaa, and J. Lachaise, *Modelling of the surface tension of binary and ternary mixtures with the gradient theory of fluid interfaces*, Fluid Phase Equilibria, 218 (2004), pp. 189–203.
[33] N. R. Nagarajan and A. S. Cullick, *New strategy for phase equilibrium and critical point calculations by thermodynamic energy analysis. Part* I. *Stability analysis and flash*, Fluid Phase Equilibria, 62 (1991), pp. 191–210.
[34] D. Peng and D. B. Robinson, *A new two-constant equation of state*, Indust. Engrg. Chem. Fund., 15 (1976), pp. 59–64.
[35] Q. Peng, *A convex-splitting scheme for a diffuse interface model with Peng-Robinson equation of state*, Adv. Appl. Math. Mech., 9 (2017), pp. 1162–1188.
[36] Z. Qiao and S. Sun, *Two-phase fluid simulation using a diffuse interface model with Peng–Robinson equation of state*, SIAM J. Sci. Comput., 36 (2014), pp. B708–B728, https://doi.org/10.1137/130933745.
[37] J. Shen and X. Yang, *Decoupled, energy stable schemes for phase-field models of two-phase incompressible flows*, SIAM J. Numer. Anal., 53 (2015), pp. 279–296, https://doi.org/10.1137/140971154.
[38] J. Shen, J. Xu, and J. Yang, *The scalar auxiliary variable (SAV) approach for gradient flows*, J. Comput. Phys., 353 (2018), pp. 407–416.
[39] S. Sun, *Darcy-scale phase equilibrium modeling with gravity and capillarity*, J. Comput. Phys., 399 (2019), 108908.
[40] S. Sun, *Energy stable simulation of two-phase equilibria with capillarity*, in Computational Science – ICCS 2019, Lecture Notes in Comput. Sci. 11539, J. Rodrigues et al., eds., Springer, Cham, 2019, pp. 538–550.
[41] G. Tryggvason, R. Scardovelli, and S. Zaleski, *Direct Numerical Simulations of Gas-Liquid Multiphase Flows*, Cambridge University Press, New York, 2011.
[42] G. N. Wells, E. Kuhl, and K. Garikipati, *A discontinuous Galerkin method for the Cahn-Hilliard equation*, J. Comput. Phys., 218 (2006), pp. 860–877.
[43] S. M. Wise, *Unconditionally stable finite difference, nonlinear multigrid simulation of the Cahn–Hilliard–Hele–Shaw system of equations*, J. Sci. Comput., 44 (2010), pp. 38–68.
[44] S. M. Wise, C. Wang, and J. S. Lowengrub, *An energy-stable and convergent finite-difference scheme for the phase field crystal equation*, SIAM J. Numer. Anal., 47 (2009), pp. 2269–2288, https://doi.org/10.1137/080738143.
[45] O. Wodo and B. Ganapathysubramanian, *Computationally efficient solution to the Cahn–Hilliard equation: Adaptive implicit time schemes, mesh sensitivity analysis and the* 3*D isoperimetric problem*, J. Comput. Phys., 230 (2011), pp. 6037–6060.
[46] H. Yang, S. Sun, Y. Li, and C. Yang, *A fully implicit constraint-preserving simulator for the black oil model of petroleum reservoirs*, J. Comput. Phys., 396 (2019), pp. 347–363.
[47] X. Yang and J. Zhao, *On linear and unconditionally energy stable algorithms for variable mobility Cahn-Hilliard type equation with logarithmic Flory-Huggins potential*, Commun. Comput. Phys., 25 (2019), pp. 703–728.
[48] X. Yang and L. Ju, *Efficient linear schemes with unconditional energy stability for the phase field elastic bending energy model*, Comput. Methods Appl. Mech. Engrg., 315 (2017), pp. 691–712.
[49] X. Yang, J. Zhao, and Q. Wang, *Numerical approximations for the molecular beam epitaxial growth model based on the invariant energy quadratization method*, J. Comput. Phys., 333 (2017), pp. 104–127.
[50] G. Zhu, J. Kou, B. Yao, Y. Wu, J. Yao, and S. Sun, *Thermodynamically consistent modelling of two-phase flows with moving contact line and soluble surfactants*, J. Fluid Mech., 879 (2019), pp. 327–359.
[51] G. Zhu, J. Kou, S. Sun, J. Yao, and A. Li, *Decoupled, energy stable schemes for a phase-field surfactant model*, Comput. Phys. Commun., 233 (2018), pp. 67–77.
[52] G. Zhu, H. Chen, J. Yao, and S. Sun, *Efficient energy-stable schemes for the hydrodynamics coupled phase-field model*, Appl. Math. Model., 70 (2019), pp. 82–108.