



Unconditionally positivity preserving and energy dissipative schemes for Poisson–Nernst–Planck equations

Jie Shen¹ · Jie Xu²

Received: 24 June 2019 / Revised: 16 April 2021 / Accepted: 18 April 2021 / Published online: 28 May 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

We develop a set of numerical schemes for the Poisson–Nernst–Planck equations. We prove that our schemes are mass conservative, uniquely solvable and keep positivity unconditionally. Furthermore, the first-order scheme is proven to be unconditionally energy dissipative. These properties hold for various spatial discretizations. Numerical results are presented to validate these properties. Moreover, numerical results indicate that the second-order scheme is also energy dissipative, and both the first- and the second-order scheme preserves the maximum principle for cases where the equation satisfies the maximum principle.

Mathematics Subject Classification 65M12 · 35K61 · 35K55 · 65Z05 · 70F99

1 Introduction

The Poisson–Nernst–Planck (PNP) equations describe the dynamics of charged particles in an electric field that is also affected by these particles, and have been used to model physical systems involving motions of charged particles, including electrochemistry [4], semiconductors [19,40], and several biological phenomena [5,10,13]. When discussing the interplay of an electric field and a flow field, the PNP equation can also be coupled with the Navier–Stokes equation [41].

This work is supported in part by AFOSR FA9550-20-1-0309, NSF DMS-2012585, NSFC 11688101 and 12001524.

✉ Jie Shen
shen7@purdue.edu
Jie Xu
xujie@lsec.cc.ac.cn

¹ Department of Mathematics, Purdue University, West Lafayette IN 47907, USA

² LSEC & NCMIS, Institute of Computational Mathematics and Scientific/Engineering Computing (ICMSEC), Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences, Beijing 100190, China

A distinct feature of the PNP equations is that they are built as Wasserstein gradient flows [1]. Wasserstein gradient flows are usually used to describe the evolution of the concentration c which remains positive, given a positive initial condition. The dissipation operator in Wasserstein flow is nonlinear, given by $\nabla \cdot (c\nabla(\cdot))$, whose negativity also requires c to be positive. Meanwhile, in most cases the energy is well-defined with a lower bound only when c is positive, for which we mention a few models from different areas [2,7,12,23,29,35,38,49]. Numerically, it is thus crucial to construct schemes that preserve positivity and energy dissipation at the same time.

If one aims to design schemes featuring only unconditional positivity preservation, or only unconditional energy dissipation, there have been various approaches. However, this turns out not to be easy if both properties are desired. Typical techniques of designing energy dissipative time-discretized schemes for gradient flows, such as convex splitting [15,16,42], stabilization [47,54], auxiliary variable approaches [3,24] (including IEQ [50,53] and SAV [43,45,46]), usually have no definite effect on positivity preserving properties. One possible way to avoid dealing with positivity is to regularize the equation (see, for example [9,39,50]), so that the equation admits solutions that may be negative. However, regularization cannot preserve positivity in the strong sense. On the other hand, the tools for positivity preservation, such as limiter [51,52] or seeking discrete maximum principle [48], are often inconsistent with the structure needed for proving unconditional energy dissipation.

This difficulty is indeed reflected in the previous works on PNP equations. Various schemes have been applied [6,14,21,22,27,34,37]. Numerical analyses have also been carried out. Some schemes require CFL conditions for positivity preservation or energy dissipation [18,31,32,39]. Some schemes exhibit energy dissipation with a prerequisite on positivity that lacks rigorous proof [20,25,36]. Recently, a few schemes have been proposed based on a reformulation by rewriting the flux using a total differential [11,26,28,30] (see also [33] for applications beyond PNP equations). The advantage of the reformulation is that one can easily write down first-order-in-time finite difference schemes with positivity preserved unconditionally. As for the energy stability, a discrete energy law is proved in both [26] and [28], although the discrete energy in [26] is without a lower bound, and the existence of the numerical solution in [28] is proved with a restriction on the time step. On the other hand, a discrete energy law is proved in [30] for a linear scheme with an $O(1)$ bound on the time step.

In this paper, we shall construct schemes for PNP equations which unconditionally (independently of the time step and grid size) satisfy the four properties below: they are (i) mass conservative, (ii) uniquely solvable, (iii) positivity preserving, and (iv) energy dissipative. The key ingredient in constructing such schemes is to consider the PNP equation as a Wasserstein gradient flow, i.e., to interpret the dissipative term $\nabla \cdot (D\nabla c)$ as $\nabla \cdot (Dc\nabla \log c)$ from which an energy law in the Wasserstein metric can be derived. We discretize the PNP equations in the context of Wasserstein gradient flow, based on the form $\nabla \cdot (Dc\nabla \log c)$. The appearance of a logarithmic function in the schemes is essential to guarantee the concentration, which is also part of the diffusion coefficient, to be positive. This is attained by treating the coefficient c explicitly, and $\log c$ from the variational derivative of the energy implicitly. The key for achieving the nice properties stated above is that the schemes can be interpreted as minimization of a strictly convex functional, which implies the uniquely solvability, positivity and energy dissipation.

Such an idea has been successfully implemented for a Cahn–Hilliard equation with Flory–Huggins energy potential [8] (which is not a Wasserstein gradient flow), and for a class of Keller–Segel equations [44] (which is a Wasserstein gradient flow in a special case).

We start by constructing a first-order time discretization scheme and show that it satisfies the four properties (i)–(iv), and we believe that it is the only scheme which is both unconditionally positivity preserving and unconditionally energy dissipative. We then construct a second-order scheme, and show that it satisfies the properties (i)–(iii). We also discuss how to construct fully discrete schemes which can preserve the properties of the time discretization schemes. The schemes are suitable for both Galerkin-type (finite element, spectral, etc.) and finite difference discretizations. At each time step these schemes lead to a nonlinear system, but since the unique solution of this is the minimizer of a strictly convex functional, it can be solved efficiently by Newton’s iteration. We provide ample numerical results to show that both first- and second-order schemes satisfy the four properties. Moreover, in some special cases, where the solution of the PNP equation satisfies a maximum principle and electrostatic energy dissipation, both the first- and second-order scheme can also preserve the maximum principle and the electrostatic energy dissipation.

The rest of paper is organized as follows. In Sect. 2, we introduce the PNP equations and state some of their properties that we desire to inherit in numerical solutions. Then, we construct numerical schemes in Sect. 3 and prove that they satisfy the four properties stated earlier. We start by writing down the semi-discrete-in-time scheme, followed by careful discretization in space so that the properties of the time discretization schemes can be preserved under full discretization. Numerical results are presented in Sect. 4. Concluding remarks are given in the last section.

2 PNP equations

We first introduce the PNP equations in a general form, and then pay particular attention to a popular two-component system because it possesses extra properties.

2.1 General form

We consider a system with N species of charged particles driven by Brownian motion and the electric field in an open bounded domain $\Omega \subset \mathbb{R}^d$ ($d = 1, 2, 3$). The system is charged with a fixed constant density ρ_0 . To write down the dimensionless PNP equations governing the motion of this system, we introduce some dimensionless quantities (functions) below:

- $c_i(\mathbf{x})$ is the density of the i -th species;
- $\phi(\mathbf{x})$ is the internal electric potential contributed by the charged particles; $\phi_e(\mathbf{x})$ is a given external electric potential;
- The chemical potential w.r.t. c_i is $\mu_i = \log c_i + z_i(\phi + \phi_e)$;
- The constants $z_i, D_i > 0$ are the valence and the diffusion constant of the i -th species, and $\epsilon > 0$ is the permittivity.

Then, the PNP equations are written as

$$\frac{\partial c_i}{\partial t} = \nabla \cdot (D_i c_i \nabla \mu_i) = \nabla \cdot \left(D_i c_i \nabla (\log c_i + z_i (\phi + \phi_e)) \right) \quad \text{in } \Omega \quad (i = 1, \dots, N), \quad (2.1)$$

where the internal electric potential ϕ is determined by

$$-\nabla \cdot (\epsilon \nabla \phi) = \rho_0 + \sum_{i=1}^N z_i c_i \quad \text{in } \Omega. \quad (2.2)$$

Noticing that $\nabla c = c \nabla \log c$, we can rewrite (2.1) as

$$\frac{\partial c_i}{\partial t} = \nabla \cdot \left(D_i (\nabla c_i + z_i c_i \nabla (\phi + \phi_e)) \right) \quad \text{in } \Omega \quad (i = 1, \dots, N), \quad (2.3)$$

which is in the form most often used in the literature.

The boundary conditions are imposed on μ_i and ϕ . They can be either periodic on both μ_i and ϕ ; or, be of Neumann type on the flux to guarantee the mass conservation,

$$c_i \frac{\partial \mu_i}{\partial \mathbf{n}} = c_i \frac{\partial (\log c_i + z_i (\phi + \phi_e))}{\partial \mathbf{n}} = \frac{\partial c_i}{\partial \mathbf{n}} + z_i c_i \frac{\partial (\phi + \phi_e)}{\partial \mathbf{n}} = 0,$$

and either Dirichlet, Neumann, or Robin boundary conditions on ϕ ,

$$\phi = 0; \quad \text{or} \quad \frac{\partial \phi}{\partial \mathbf{n}} = 0; \quad \text{or} \quad \alpha \phi + \beta \frac{\partial \phi}{\partial \mathbf{n}} = 0, \quad \alpha, \beta > 0.$$

If using periodic or Neumann boundary conditions on ϕ , we require that

$$\rho_0 + \sum z_i \bar{c}_i = 0, \quad \int_{\Omega} \phi \, d\mathbf{x} = 0,$$

where \bar{c}_i is the average density of the i -th species.

Remark 2.1 We only consider periodic or homogeneous boundary conditions above on ϕ . For non-homogeneous boundary conditions such as $\phi|_{\partial\Omega} = g$, we can split ϕ as $\phi_1 + \phi_2$, with

$$\begin{aligned} -\nabla \cdot (\epsilon \nabla \phi_1) &= \rho_0 + \sum_{i=1}^N z_i c_i, & \phi|_{\partial\Omega} &= 0, \\ -\nabla \cdot (\epsilon \nabla \phi_2) &= 0, & \phi|_{\partial\Omega} &= g. \end{aligned}$$

Note that ϕ_2 does not depend on c_i . Thus, ϕ_2 actually acts as an external potential and could be added to ϕ_e . It is known that the profiles of c_i can sensitively depend on the boundary conditions [18]. In the above formulation, it actually implies that ϕ_2 , which goes in to the external potential, greatly affects the profile.

The total free energy of the system is given by

$$E(\{c_i\}, \phi) = \int_{\Omega} \sum_{i=1}^N c_i (\log c_i - 1) + \left(\rho_0 + \sum_{i=1}^N z_i c_i \right) \left(\frac{1}{2} \phi + \phi_e \right) \, dx. \tag{2.4}$$

Assuming $\|\phi_e\|_{L^\infty(\Omega)} \leq A$, we derive that the total energy is bounded from below. Indeed, we have

$$c_i \log c_i - c_i + z_i \phi_e c_i \geq c_i \log c_i - (|z_i|A + 1)c_i,$$

which is bounded from below. For the term with the internal potential ϕ , we derive by integration by parts that

$$\int \left(\rho_0 + \sum_{i=1}^N z_i c_i \right) \phi \, dx = \begin{cases} \int \epsilon |\nabla \phi|^2 \, dx, & \text{with periodic, Dirichlet or Neumann B.C.,} \\ \int \epsilon |\nabla \phi|^2 \, dx + \int_{\partial\Omega} \epsilon \frac{\alpha}{\beta} |\phi|^2 \, dS, & \text{with Robin B.C..} \end{cases}$$

The PNP equations (2.1)–(2.2) satisfy several important properties:

1. Mass conservation: Integrating (2.1) over Ω , we obtain immediately

$$\int_{\Omega} c_i(\cdot, t) \, dx = \int_{\Omega} c_i(\cdot, 0) \, dx.$$

2. Positivity: The well-posedness of (2.1)–(2.2) (cf. [41]) ensures that, if $c_i(\cdot, 0) > 0$, then we still have $c_i(\cdot, t) > 0$ for any $t > 0$.
3. Energy dissipation:

$$\frac{dE}{dt} = - \int \sum_{i=1}^N D_i c_i |\nabla \mu_i|^2 \, dx. \tag{2.5}$$

To derive the above energy dissipation, we need to observe that $\mu_i = \frac{\delta E}{\delta c_i}$. Actually, the variation $\delta\phi$ satisfies

$$-\nabla \cdot (\epsilon \nabla (\delta\phi)) = \sum_{i=1}^N z_i \delta c_i,$$

with the same boundary conditions as ϕ . Regardless of the type of the boundary conditions, we have

$$\int -\delta\phi \nabla \cdot (\epsilon \nabla \phi) \, dx = \int -\nabla \cdot (\epsilon \nabla (\delta\phi)) \phi \, dx.$$

So we have

$$\begin{aligned} \delta \int \left(\rho_0 + \sum_{i=1}^N z_i c_i \right) \phi \, dx &= \int \phi \sum_{i=1}^N z_i \delta c_i - \delta \phi \nabla \cdot (\epsilon \nabla \phi) \, dx \\ &= \int \phi \sum_{i=1}^N z_i \delta c_i - \nabla \cdot (\epsilon \nabla (\delta \phi)) \phi \, dx = 2 \int \phi \sum_{i=1}^N z_i \delta c_i \, dx \end{aligned}$$

Therefore, by multiplying the equation (2.1) with μ_i , taking the integral and summing up over i , we obtain (2.5).

2.2 A two-component system

We consider a two-component system ($N = 2$) which has attracted special attention in the literature. We set $z_1 = 1, z_2 = -1, \epsilon = 1$ and the external electric potential $\phi_e = 0$. Denote $p = c_1$ and $n = c_2$. Let the average density be $\bar{c}_1 = \bar{c}_2$ so that $\rho_0 = 0$. The PNP equation becomes

$$\frac{\partial p}{\partial t} = \nabla \cdot (D_1 p \nabla (\log p + \phi)), \tag{2.6}$$

$$\frac{\partial n}{\partial t} = \nabla \cdot (D_2 n \nabla (\log n - \phi)), \tag{2.7}$$

$$-\Delta \phi = p - n, \tag{2.8}$$

where p and n denote the concentration of positively and negatively charged particles, respectively, and ϕ is the electronic potential. This system is named after W. Nernst and M. Planck to describe the potential difference in a galvanic cell (e.g., rechargeable batteries, or biological cells), and has applications in many different fields including chemistry, biology, plasma physics, and modeling of semi-conductor devices.

The above system has two special properties stated below, which are satisfied only under the periodic or Neumann boundary conditions for p, n, ϕ . They do not necessarily hold for the general form of PNP equations.

1. The electrostatic energy $\|\nabla \phi\|^2/2$ is dissipative if $D_1 = D_2 = D$. Indeed, multiplying (2.6) and (2.7) with ϕ and calculating their difference, we obtain

$$\frac{d\|\nabla \phi\|^2/2}{dt} = -D \int [(p - n)^2 + (p + n)|\nabla \phi|^2] \, dx. \tag{2.9}$$

It needs to be pointed out that the electrostatic energy is part of the total energy (2.4), by noticing the derivation below (2.4).

2. The solutions p and n satisfy maximum principle. In [17], it is shown in a general setting that if initially we have $0 < \epsilon(0) \leq p, n \leq A(0)$, then at any time $T > 0$ there exist $A(T), \epsilon(T) > 0$ such that $\epsilon(T) \leq p, n \leq A(T)$. We give a short formal derivation for the maximum principle below. Multiplying (2.6) with p^{k-1} , we

obtain

$$\begin{aligned} \int \frac{1}{k} \cdot \frac{\partial p^k}{\partial t} \mathbf{dx} &= -D_1 \int (k-1)p^{k-2} |\nabla p|^2 + p \nabla(p^{k-1}) \cdot \nabla \phi \mathbf{dx} \\ &= -D_1 \int (k-1)p^{k-2} |\nabla p|^2 + \frac{k-1}{k} \nabla(p^k) \cdot \nabla \phi \mathbf{dx}. \end{aligned}$$

Since we adopt periodic or Neumann boundary conditions on ϕ , we deduce that

$$\int \frac{1}{k} \cdot \frac{\partial p^k}{\partial t} \mathbf{dx} = -D_1 \int (k-1)p^{k-2} |\nabla p|^2 + \frac{k-1}{k} p^k (p-n) \mathbf{dx}.$$

Similarly, multiplying (2.7) with n^{k-1} , we have

$$\int \frac{1}{k} \cdot \frac{\partial n^k}{\partial t} \mathbf{dx} = -D_2 \int (k-1)n^{k-2} |\nabla n|^2 - \frac{k-1}{k} n^k (p-n) \mathbf{dx}.$$

Taking the sum of the above two equations, and noting that $p, n > 0$, we obtain

$$\begin{aligned} \int \frac{\partial(p^k/D_1 + n^k/D_2)}{\partial t} \mathbf{dx} &= - \int k(k-1)(p^{k-2} |\nabla p|^2 + n^{k-2} |\nabla n|^2) \\ &\quad + (k-1)(p^k - n^k)(p-n) \mathbf{dx} \leq 0. \end{aligned} \tag{2.10}$$

So we have

$$\begin{aligned} \|p(t)\|_{L^k} &\leq \left(\|p(t)\|_{L^k}^k + \frac{D_1}{D_2} \|n(t)\|_{L^k}^k \right)^{1/k} \leq \left(\|p(0)\|_{L^k}^k + \frac{D_1}{D_2} \|n(0)\|_{L^k}^k \right)^{1/k} \\ &\leq \left(1 + \frac{D_1}{D_2} \right)^{1/k} \max\{\|p(0)\|_{L^\infty}, \|n(0)\|_{L^\infty}\}. \end{aligned}$$

Taking the limit $k \rightarrow +\infty$, we obtain

$$\max\{\|p(t)\|_{L^\infty}, \|n(t)\|_{L^\infty}\} \leq \max\{\|p(0)\|_{L^\infty}, \|n(0)\|_{L^\infty}\}.$$

Note that the inequality (2.10) also holds for $k < -1$, we then obtain by taking $k \rightarrow -\infty$ that

$$\max \left\{ \left\| \frac{1}{p(t)} \right\|_{L^\infty}, \left\| \frac{1}{n(t)} \right\|_{L^\infty} \right\} \leq \max \left\{ \left\| \frac{1}{p(0)} \right\|_{L^\infty}, \left\| \frac{1}{n(0)} \right\|_{L^\infty} \right\}.$$

Although we are not aiming to design numerical schemes guaranteeing these two properties theoretically, we are still interested in and will examine whether they can be kept in the numerical solutions.

3 Numerical scheme

We construct in this section numerical schemes for (2.1)–(2.2). We start from a first-order scheme and prove that it enjoys the four nice properties described in the introduction. We then construct a second-order scheme which enjoys the first three properties.

3.1 First-order scheme

We first write down the time-discretized scheme for (2.1)–(2.2):

$$\frac{c_i^{n+1} - c_i^n}{\delta t} = \nabla \cdot (D_i c_i^n \nabla \mu_i^{n+1}) = \nabla \cdot (D_i c_i^n \nabla (\log c_i^{n+1} + z_i (\phi^{n+1} + \phi_e))), \quad i = 1, \dots, N, \tag{3.1}$$

$$-\nabla \cdot (\epsilon \nabla \phi^{n+1}) = \rho_0 + \sum_{i=1}^N z_i c_i^{n+1}, \tag{3.2}$$

with the boundary conditions imposed on $\mu_i^{n+1} = \log c_i^{n+1} + z_i (\phi^{n+1} + \phi_e)$ and ϕ^{n+1} as in the PDE system (2.1)–(2.2).

Theorem 3.1 *Assume $c_i^n > 0$ for all i . For any solution to the scheme (3.1)–(3.2), we have*

1. *Mass conservation:*

$$\int c_i^{n+1} \, dx = \int c_i^n \, dx.$$

2. *Positivity preserving: $c_i^{n+1} > 0$ for all i .*

3. *Energy dissipation:*

$$E^{n+1} - E^n \leq -\delta t \int \sum_{i=1}^N D_i c_i^n \left| \nabla \mu_i^{n+1} \right|^2 \, dx, \quad n \geq 0, \tag{3.3}$$

where $E^k = \int_{\Omega} \sum_{i=1}^N c_i^k (\log c_i^k - 1) + \left(\rho_0 + \sum_{i=1}^N z_i c_i^k \right) \left(\frac{1}{2} \phi^k + \phi_e \right) \, dx.$

Proof We shall only prove the theorem for the Neumann boundary conditions on ϕ and μ_i . The results with other boundary conditions can be proved in the same way, as we will point out afterwards.

Taking the integral of (3.1) over Ω and using the Neumann boundary conditions on the chemical potential, we obtain the mass conservation.

The positivity follows from the appearance of $\log c_i^{n+1}$.

It remains to prove the energy dissipation. To this end, we take the inner product of the equation (3.1) with $\log c_i^{n+1} + z_i (\phi^{n+1} + \phi_e)$, summing up for $1 \leq i \leq N$, we

arrive at

$$\begin{aligned} & \sum_{i=1}^N (c_i^{n+1} - c_i^n, \log c_i^{n+1}) + \left(\rho_0 + \sum_{i=1}^N z_i c_i^{n+1} - \rho_0 - \sum_{i=1}^N z_i c_i^n, \phi^{n+1} + \phi_e \right) \\ &= \sum_{i=1}^N (c_i^{n+1} - c_i^n, \log c_i^{n+1}) + \left(\nabla \phi^{n+1} - \nabla \phi^n, \epsilon \nabla \phi^{n+1} \right) \\ & \quad + \left(\rho_0 + \sum_{i=1}^N z_i c_i^{n+1} - \rho_0 - \sum_{i=1}^N z_i c_i^n, \phi_e \right) \\ &= -\delta t \int \sum_{i=1}^N D_i c_i^n \left| \nabla (\log c_i^{n+1} + z_i (\phi^{n+1} + \phi_e)) \right|^2 dx. \end{aligned}$$

We note that by Taylor expansion we have

$$(a - b) \log a = (a \log a - a) - (b \log b - b) + \frac{(a - b)^2}{2\xi}, \quad \xi \in [\min\{a, b\}, \max\{a, b\}]. \tag{3.4}$$

We also have

$$(\nabla \phi^{n+1} - \nabla \phi^n) \cdot \nabla \phi^{n+1} = \frac{1}{2} (|\nabla \phi^{n+1}|^2 - |\nabla \phi^n|^2 + |\nabla \phi^{n+1} - \nabla \phi^n|^2). \tag{3.5}$$

With the above equalities, we immediately derive (3.3). □

It remains to examine whether there exists a solution for the scheme. Below, we give a *formal* derivation by formulating it as the minimizer of a strictly convex functional. Still, we examine the Neumann boundary conditions for ϕ and μ_i . Let us introduce linear operators \mathcal{L}_i^n , which are defined as follows: let $\mathcal{L}_i^n g = u$ if they satisfy the following elliptic equation with the Neumann boundary conditions,

$$-\nabla \cdot (c_i^n \nabla u) = g, \quad \int u dx = 0.$$

Also, we define \mathcal{L} as above where we replace c_i^n with ϵ . The linear operators \mathcal{L}_i^n and \mathcal{L} are symmetric and nonnegative in the sense $(u, \mathcal{L}u) \geq 0$. We consider the following functional

$$F[c_i^{n+1}] = \sum_{i=1}^N (c_i^{n+1} \log c_i^{n+1} - c_i^{n+1}, 1) + \frac{1}{2\delta t} \sum_{i=1}^N (c_i^{n+1} - c_i^n, \mathcal{L}_i^n (c_i^{n+1} - c_i^n))$$

$$\begin{aligned}
 &+ \frac{1}{2} \left(\rho_0 + \sum_{i=1}^N z_i c_i^{n+1}, \mathcal{L} \left(\rho_0 + \sum_{i=1}^N z_i c_i^{n+1} \right) \right) + \left(\rho_0 + \sum_{i=1}^N z_i c_i^{n+1}, \phi_e \right). \\
 & \hspace{15em} (3.6)
 \end{aligned}$$

The above functional is strictly convex, because $\int c_i^{n+1} \log c_i^{n+1} + (z_i \phi_e - 1) c_i^{n+1} \, d\mathbf{x}$ is strictly convex about c_i^{n+1} , and the remaining terms give a quadratic nonnegative functional. Its Euler–Lagrange equation under the constraints of mass is

$$\begin{aligned}
 &\frac{1}{\delta t} \mathcal{L}_i^n (c_i^{n+1} - c_i^n) + \log c_i^{n+1} + z_i \mathcal{L} \left(\rho_0 + \sum_{i=1}^N z_i c_i^{n+1} \right) + z_i \phi_e \\
 &= \frac{1}{\delta t} \mathcal{L}_i^n (c_i^{n+1} - c_i^n) + \log c_i^{n+1} + z_i (\phi^{n+1} + \phi_e) = \lambda_i, \\
 &\int (c_i^{n+1} - c_i^n) \, d\mathbf{x} = 0.
 \end{aligned}$$

where λ_i are the Lagrange multipliers for the mass conservation. It is easy to see that the above equations are equivalent to (3.1)–(3.2). The functional F has a unique minimizer. Moreover, the minimizer cannot have $c_i^{n+1}(\mathbf{x}) = 0$, because the derivative of the term $c_i^{n+1} \log c_i^{n+1} - c_i^{n+1}$ has the derivative $\log c_i^{n+1}$ that tends to $-\infty$ at zero. Hence, the unique minimizer must have $c_i^{n+1} > 0$ for all i , which is the unique solution to the Euler–Lagrange equation, hence to the scheme, because we can solve ϕ uniquely from (3.2).

The above formal derivation can be converted into a rigorous proof after we discretize in space. Before going on, let us explain the difference when using other boundary conditions, both for the theorem and for the formal derivation above. For the periodic boundary conditions, everything is exactly the same. When using Dirichlet or Robin boundary conditions, we do not need the average equals to zero when defining the operator \mathcal{L} (but still need for \mathcal{L}^n). For the energy dissipation for Robin boundary conditions, we need an extra term $\int_{\partial\Omega} \epsilon \alpha \beta |\phi|^2 \, dS$, which can be dealt with in the same way as $\int_{\Omega} \epsilon |\nabla \phi|^2 \, d\mathbf{x}$. Thus, we will still focus on the Neumann boundary conditions below.

We now discuss how to construct spatial discretizations which preserve the nice properties for the scheme (3.1)–(3.2). Note that in the proof of Theorem 3.1, we have used non-standard functions like $\log c_i^{n+1}$ as test function. Therefore, the proof can not be directly extended to a straightforward discretization in space since the discrete version of $\log c_i^{n+1}$ is usually not in the discrete test space. We need to carefully discretize the space to keep the properties stated in Theorem 3.1 in the discrete sense (cf. [32]).

Let us first discuss Galerkin type discretizations with finite-elements or spectral methods. Since there are differential operators with variable coefficients, we need to define a discrete inner product, i.e. numerical integration, on a finite set of points

$Z = \{z\}$:

$$[u, v] = \sum_{z \in Z} \beta_z u(z)v(z), \tag{3.7}$$

where we require that the weights $\beta_z > 0$. For finite element methods, the sum should be understood as $\sum_{K \subset \mathcal{T}} \sum_{z \in Z(K)}$ where \mathcal{T} is a given triangulation.

As we have mentioned, we still consider Neumann boundary conditions. Let $X_M \subset H^1(\Omega)$ be a finite dimensional approximation space. Assume that there is a unique function $\psi_z(\mathbf{x})$ in X_M satisfying $\psi_z(\mathbf{z}') = 1_{z=z'}$ for $z, z' \in Z$. Then, we can define $I_M : C(\Omega) \rightarrow X_M$ as the interpolation operator about the points in Z .

Our Galerkin method for the first-order scheme (3.1)–(3.2) is: to find $\{c_i^{n+1}\}$ and ϕ^{n+1} in X_M satisfying

$$\left[\frac{c_i^{n+1} - c_i^n}{\delta t}, v \right] = - \left[D_i c_i^n \nabla \left(I_M \left(\log c_i^{n+1} + z_i (\phi^{n+1} + \phi_e) \right) \right), \nabla v \right], \quad v \in X_M, \tag{3.8}$$

$$(\epsilon \nabla \phi^{n+1}, \nabla w) = \left[\rho_0 + \sum_{i=1}^N z_i c_i^{n+1}, w \right], \quad w \in X_M. \tag{3.9}$$

We emphasize that in the above, (\cdot, \cdot) represents the continuous L^2 inner product, while $[\cdot, \cdot]$ represents the discrete L^2 inner product defined in (3.7).

Theorem 3.2 *The fully discretized scheme (3.8)–(3.9) enjoys the following properties:*

1. *Mass conservation:*

$$[c_i^{n+1}, 1] = [c_i^n, 1].$$

2. *Unique solvability: the scheme (3.8)–(3.9) possesses a unique solution ($\{c_i^{n+1} \in X_M\}, \phi^{n+1} \in X_M$).*

3. *Positivity preserving: if $c_i^n(z) > 0$ for all i and $z \in Z$, we have $c_i^{n+1}(z) > 0$ for all i and $z \in Z$.*

4. *Energy dissipation:*

$$\tilde{E}^{n+1} - \tilde{E}^n \leq -\delta t \sum_{i=1}^N [D_i c_i^n \nabla \mu_i^{n+1}, \nabla \mu_i^{n+1}], \tag{3.10}$$

where $\mu_i^{n+1} = I_M(\log c_i^{n+1} + z_i(\phi^{n+1} + \phi_e))$ and the discrete energy is defined as

$$\tilde{E}^n = \sum_{i=1}^N [c_i^n \log c_i^n - c_i^n, 1] + \left[\rho_0 + \sum_{i=1}^N z_i c_i^n, \frac{1}{2} \phi^n + \phi_e \right]. \tag{3.11}$$

Proof The mass conservation is obtained by choosing $v = 1$.

Next, we look the unique solvability and positivity. Since we have $c_i^n(\mathbf{x}) = \sum_z c_i^n(\mathbf{z})\psi_z(\mathbf{x})$, let us denote the vector $(c_i^n(\mathbf{z}), \mathbf{z} \in Z)$ as \tilde{c}_i^n . Similarly we denote $(\phi^n(\mathbf{z}), \mathbf{z} \in Z)$ and $(\phi_e^n(\mathbf{z}), \mathbf{z} \in Z)$ by the vectors $\tilde{\phi}^n$ and $\tilde{\phi}_e$, respectively. We define the following stiffness and mass matrices:

$$A_i^n = [D_i c_i^n \nabla \psi_z, \nabla \psi_{z'}], \quad A = \epsilon(\nabla \psi_z, \nabla \psi_{z'}), \quad B = [\psi_z, \psi_{z'}].$$

It is clear that B is a diagonal matrix with positive elements since $\beta_z > 0$, and A is symmetric positive semi-definite. If $c_i^n(\mathbf{z}) > 0$ for $\mathbf{z} \in Z$, the matrices A_i^n are symmetric positive semi-definite. Furthermore, $A_i^n \tilde{x} = 0$, similarly $A \tilde{x} = 0$, if and only if all the components of \tilde{x} are equal. Therefore, A_i^n and A have one zero eigenvalue with all other eigenvalues being positive. Hence, the eigen-decomposition of A takes the form $A = T^t \Lambda T$ with $\Lambda = \text{diag}(0, \mu_2, \dots, \mu_M)$ and $\mu_j > 0$ for $j = 2, \dots, M$. We denote by A^* the pseudo-inverse given by $A^* = T^t \text{diag}(0, \mu_2^{-1}, \dots, \mu_M^{-1}) T$. Similarly we can define $(A_i^n)^*$ for $i = 1, \dots, N$. With the above notations, we can rewrite the scheme (3.8)–(3.9) in matrix form as follows:

$$\frac{1}{\delta t} B(\tilde{c}_i^{n+1} - \tilde{c}_i^n) = -A_i^n (\log \tilde{c}_i^{n+1} + z_i(\tilde{\phi}^{n+1} + \tilde{\phi}_e)), \tag{3.12}$$

$$A \tilde{\phi}^{n+1} = B \left(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1} \right). \tag{3.13}$$

Multiplying the above equations by pseudo-inverse $(A_i^n)^*$ and A^* , we find

$$\frac{1}{\delta t} (A_i^n)^* B(\tilde{c}_i^{n+1} - \tilde{c}_i^n) + \log \tilde{c}_i^{n+1} + z_i(\tilde{\phi}^{n+1} + \tilde{\phi}_e) = \lambda_i \mathbf{1}, \tag{3.14}$$

$$\tilde{\phi}^{n+1} = A^* B(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1}) + \lambda \mathbf{1}, \tag{3.15}$$

with $\mathbf{1}$ representing the all-one vector. Eliminating $\tilde{\phi}^{n+1}$ from the above, and then multiplying B to the first equation, we arrive at

$$\frac{1}{\delta t} B(A_i^n)^* B(\tilde{c}_i^{n+1} - \tilde{c}_i^n) + B \log \tilde{c}_i^{n+1} + z_i \left(BA^* B(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1}) + B \tilde{\phi}_e \right) = \lambda'_i B \mathbf{1},$$

along with the mass conservation $\mathbf{1}^t B(\tilde{c}_i^{n+1} - \tilde{c}_i^n) = 0$. Here, $\lambda'_i = \lambda_i - z_i \lambda$ acts as the Lagrange multiplier for the mass conservation. One can then easily check that the above is the Euler–Lagrange equation of the function

$$\tilde{F}[\tilde{c}_1^{n+1}, \dots, \tilde{c}_N^{n+1}]$$

$$\begin{aligned}
 &= \frac{1}{2\delta t} \sum_{i=1}^N (\tilde{c}_i^{n+1} - \tilde{c}_i^n)^t B(A_i^n)^* B(\tilde{c}_i^{n+1} - \tilde{c}_i^n) + \sum_{i=1}^N (\tilde{c}_i^{n+1})^t B(\log \tilde{c}_i^{n+1} - 1) \\
 &\quad + \frac{1}{2} \left(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1} \right)^t B A^* B \left(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1} \right) + \tilde{\phi}_e^t B \left(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1} \right).
 \end{aligned}$$

Since B is diagonal and positive definite, and $(A_i^n)^*$, A^* are symmetric and nonnegative, it is clear that the above function is strictly convex about \tilde{c}_i^{n+1} . Therefore, $\tilde{F}[\tilde{c}_1^{n+1}, \dots, \tilde{c}_N^{n+1}]$ has a unique minimizer in the region $c_i(z) \geq 0$, which is the domain of \tilde{F} by noticing that we have terms $c_i^{n+1}(z) \log c_i^{n+1}(z)$ in \tilde{F} . Below we eliminate the possibility of $\tilde{c}_i^{n+1}(z) = 0$. If this is done, the unique minimizer satisfies $\{\tilde{c}_i^{n+1} > 0\}_{i=1, \dots, N}$. With \tilde{c}_i^{n+1} , we can then determine a unique $\tilde{\phi}^{n+1}$ from (3.13).

Let us prove by contradiction. Without loss of generality, suppose the minimizer has $\tilde{c}_1^{n+1}(z) = 0$. Choose another z' such that $\tilde{c}_1^{n+1}(z') > 0$. Keep the other \tilde{c}_i^{n+1} , and substitute \tilde{c}_1^{n+1} by $\tilde{d}_1^{n+1} = \tilde{c}_1^{n+1} + \beta_{z'} \rho e_z - \beta_z \rho e_{z'}$, where we use e_z to denote the vector with the entry one for the z -component and zero entry for others. Next, we will show that when ρ is small enough, $\tilde{F}[\tilde{d}_1^{n+1}, \tilde{c}_i^{n+1}|_{i=2, \dots, n}] < \tilde{F}[\tilde{c}_i^{n+1}|_{i=1, \dots, n}]$. In the following, we denote two quantities in the inequality in short by $\tilde{F}[\tilde{d}_1^{n+1}]$ and $\tilde{F}[\tilde{c}_1^{n+1}]$.

Split \tilde{F} into two parts:

$$\tilde{F}_1 = \sum_{i=1}^N (\tilde{c}_i^{n+1})^t B(\log \tilde{c}_i^{n+1} - 1),$$

and $\tilde{F}_2 = \tilde{F} - \tilde{F}_1$. Note that \tilde{F}_2 is a quadratic function. Thus, there exists a constant $A_1 > 0$ such that for ρ small enough,

$$|\tilde{F}_2[\tilde{d}_1^{n+1}] - \tilde{F}_2[\tilde{c}_1^{n+1}]| < A_1 \rho.$$

Now we turn to \tilde{F}_1 . Let $a = \tilde{c}_1^{n+1}(z') > 0$. We can calculate that

$$\tilde{F}_1[\tilde{d}_1^{n+1}] - \tilde{F}_1[\tilde{c}_1^{n+1}] = \beta_z \beta_{z'} \rho \log(\beta_{z'} \rho) + \beta_{z'} ((a - \beta_z \rho) \log(a - \beta_z \rho) - a \log a).$$

Since $a > 0$, for ρ small enough, we have

$$|(a - \beta_z \rho) \log(a - \beta_z \rho) - a \log a| < A_2 \rho.$$

Thus, if we choose $\beta_z \beta_{z'} \log(\beta_{z'} \rho) < -A_1 - \beta_{z'} A_2$, we arrive at $\tilde{F}[\tilde{d}_1^{n+1}] < \tilde{F}[\tilde{c}_1^{n+1}]$, which is the contradiction we want.

It remains to prove the energy dissipation. To this end, we choose $v = \delta t \mu_i^{n+1} = \delta t I_M(\log c_i^{n+1} + z_i(\phi^{n+1} + \phi_e))$ in (3.8) and take the sum for $1 \leq i \leq N$, leading to

$$\begin{aligned} & -\delta t \sum_{i=1}^N [D_i c_i^n \nabla \mu_i^{n+1}, \nabla \mu_i^{n+1}] \\ & = \sum_{i=1}^N [c_i^{n+1} - c_i^n, I_M(\log c_i^{n+1} + z_i(\phi^{n+1} + \phi_e))] \\ & = \sum_{i=1}^N [c_i^{n+1} - c_i^n, \log c_i^{n+1}] + \left[\rho_0 + \sum_{i=1}^N z_i c_i^{n+1} - \rho_0 - \sum_{i=1}^N z_i c_i^n, \phi^{n+1} + \phi_e \right]. \end{aligned}$$

Then, by using (3.9) and (3.5), we have

$$\begin{aligned} & 2 \left[\rho_0 + \sum_{i=1}^N z_i c_i^{n+1} - \rho_0 - \sum_{i=1}^N z_i c_i^n, \phi^{n+1} \right] \\ & = 2(\nabla \phi^{n+1} - \nabla \phi^n, \epsilon \nabla \phi^{n+1}) \\ & = ((\nabla \phi^{n+1}, \epsilon \nabla \phi^{n+1}) - (\nabla \phi^n, \epsilon \nabla \phi^n) + (\nabla(\phi^{n+1} - \phi^n), \epsilon \nabla(\phi^{n+1} - \phi^n))) \\ & = \left[\rho_0 + \sum_{i=1}^N z_i c_i^{n+1}, \phi^{n+1} \right] - \left[\rho_0 + \sum_{i=1}^N z_i c_i^n, \phi^n \right] + (\nabla(\phi^{n+1} - \phi^n), \epsilon \nabla(\phi^{n+1} - \phi^n)). \end{aligned}$$

We can then obtain (3.10) by using (3.4). □

Remark 3.3 1. For Dirichlet boundary conditions on ϕ , we just need to change the function space for ϕ and w from X_M to X_{M0} requiring that the boundary value is zero. For Robin boundary conditions on ϕ , we just need to add the surface integral in (3.9).

2. The positivity is preserved on the discrete points $z \in Z$, while the function $\psi_z(x)$ may not be positive for any x .

Let us now briefly discuss how to construct finite difference schemes which preserve the properties of the time discretizations in the last section. An important aspect in finite difference schemes is to carefully implement the boundary conditions such that the summation by parts holds, which is crucial to guarantee the mass conservation (cf. [18] for comparison of non-conservative vs conservative discretization) and to derive the energy dissipation. This is not difficult on rectangular domains. We write down the 2D case, which is to be used in our numerical test, with the domain $[0, L]^2$ discretized at M^2 points $x_{j,k} = ((j - \frac{1}{2}) \delta x, (k - \frac{1}{2}) \delta x)$, $j, k = 1, \dots, M$ where $\delta x = L/M$. The scheme is written as

$$\frac{(c_i)_{j,k}^{n+1} - (c_i)_{j,k}^n}{\delta t} = \frac{D_i}{\delta x^2} \left[\frac{(c_i)_{j+1,k}^n + (c_i)_{j,k}^n}{2} ((\mu_i)_{j+1,k}^{n+1} - (\mu_i)_{j,k}^{n+1}) \right] \tag{3.16}$$

$$\begin{aligned}
 & - \frac{(c_i)_{j,k}^n + (c_i)_{j-1,k}^n}{2} \left((\mu_i)_{j,k}^{n+1} - (\mu_i)_{j-1,k}^{n+1} \right), \\
 & + \frac{(c_i)_{j,k+1}^n + (c_i)_{j,k}^n}{2} \left((\mu_i)_{j,k+1}^{n+1} - (\mu_i)_{j,k}^{n+1} \right) \\
 & - \frac{(c_i)_{j,k}^n + (c_i)_{j,k-1}^n}{2} \left((\mu_i)_{j,k}^{n+1} - (\mu_i)_{j,k-1}^{n+1} \right) \Big], \quad 1 \leq j, k \leq M, \quad 1 \leq i \leq N, \\
 & - \epsilon \frac{\phi_{j+1,k}^{n+1} + \phi_{j-1,k}^{n+1} + \phi_{j,k+1}^{n+1} + \phi_{j,k-1}^{n+1} - 4\phi_j^{n+1}}{h^2} = \rho_0 + \sum_{i=1}^N z_i (c_i)_{j,k}^{n+1}, \quad 1 \leq j, k \leq M,
 \end{aligned} \tag{3.17}$$

where $(\mu_i)_{j,k}^n = (\log c_i + z_i(\phi + \phi_e))_{j,k}^n$. To fix the idea, we still consider the Neumann boundary conditions. To have the summation by parts, we shall impose boundary terms like below,

$$\begin{aligned}
 \frac{(\mu_i)_{0,k}^{n+1} - (\mu_i)_{1,k}^{n+1}}{h} &= 0, \quad \frac{(\mu_i)_{M+1,k}^{n+1} - (\mu_i)_{M,k}^{n+1}}{h} = 0, \\
 \frac{\phi_{0,k}^{n+1} - \phi_{1,k}^{n+1}}{h} &= 0, \quad \frac{\phi_{M+1,k}^{n+1} - \phi_{M,k}^{n+1}}{h} = 0.
 \end{aligned} \tag{3.18}$$

The above boundary discretization is for $\partial u / \partial \mathbf{n} |_{\partial \Omega}$. The term $u |_{\partial \Omega}$ shall be discretized by $\frac{1}{2}(u_{0,k} + u_{1,k})$ for the summation by parts, if we consider Dirichlet or Robin boundary conditions on ϕ .

For the above scheme, we have

Theorem 3.4 *The finite difference scheme (3.16)–(3.18) enjoys the following properties:*

1. *Mass conservation:*

$$\delta x^2 \sum_{j,k=1}^M (c_i)_{j,k}^{n+1} = \delta x^2 \sum_{j,k=1}^M (c_i)_{j,k}^n, \quad 1 \leq i \leq N.$$

2. *Unique solvability: the scheme (3.16)–(3.18) possesses a unique solution $(\{(c_i)_{j,k}^{n+1}\}, \phi_{j,k}^{n+1})$.*

3. *Positivity preserving: if $(c_i)_{j,k}^n > 0$ for all i and (j, k) , we have $(c_i)_{j,k}^{n+1} > 0$ for all i and (j, k) .*

4. *Energy dissipation: we have*

$$\begin{aligned}
 \bar{E}^{n+1} - \bar{E}^n &\leq -\delta t \sum_{i=1}^N \frac{D_i}{\delta x^2} \sum_{\substack{1 \leq j \leq M-1 \\ 1 \leq k \leq M}} \frac{(c_i)_{j+1,k}^n + (c_i)_{j,k}^n}{2} \left((\mu_i)_{j+1,k}^{n+1} - (\mu_i)_{j,k}^{n+1} \right)^2 \\
 &+ \sum_{\substack{1 \leq j \leq M \\ 1 \leq k \leq M-1}} \frac{(c_i)_{j,k+1}^n + (c_i)_{j,k}^n}{2} \left((\mu_i)_{j,k+1}^{n+1} - (\mu_i)_{j,k}^{n+1} \right)^2, \tag{3.19}
 \end{aligned}$$

where the discrete energy is defined as

$$\tilde{E}^n = \sum_{i=1}^N \sum_{j,k=1}^M (c_i)_{j,k}^n (\log(c_i)_{j,k}^n - 1) + \sum_{j,k=1}^M \left[\rho_0 + \sum_{i=1}^N z_i (c_i)_{j,k}^n \right] \cdot \left[\frac{1}{2} \phi_{j,k}^n + (\phi_e)_{j,k} \right]. \tag{3.20}$$

Proof The mass conservation is obtained by taking the sum over $1 \leq j, k \leq M$ on (3.16) and using the boundary conditions of $(\mu_i)_{j,k}^{n+1}$ in (3.18).

The unique solvability and positivity can be proved similar to Theorem 3.2 by choosing the matrices as those given by finite difference discretization, which we write down and discuss below.

Now, the vector \tilde{c}_i^n is formed by $(c_i)_{j,k}^n$ for $1 \leq j, k \leq M$. To find the definition of the matrix A_i^n , we regard the right-hand side of (3.16) as $-(1/\delta x^2) A_i^n \tilde{\mu}_i^{n+1}$.

We say that (j, k) and (j', k') are adjacent if $|j - j'| + |k - k'| = 1$. To describe the matrix A_i^n , let us first fix a grid point (j, k) where $2 \leq j, k \leq M - 1$. In the row corresponding to the grid point (j, k) , one can check that the only off-diagonal nonzero entries correspond to the adjacent indices. Moreover, the diagonal entry,

$$D_i \left[2(c_i)_{j,k}^n + \frac{1}{2} \left((c_i)_{j+1,k}^n + (c_i)_{j-1,k}^n + (c_i)_{j,k+1}^n + (c_i)_{j,k-1}^n \right) \right],$$

is positive, while the off-diagonal nonzero ones are negative, with the sum of the these nonzero entries being zero. When j or k takes 1 or M , we take the boundary conditions into consideration and find that the above statement still holds. Another observation is that the matrix A_i^n is symmetric. As a result, the eigenvalues of A_i^n are nonnegative, thus positive semi-definite. Furthermore, the only zero eigenvector is the all-one vector. Actually, assume that $A_i^n \tilde{v} = 0$. Starting from a maximum entry of \tilde{v} which is denoted by v_0 , we deduce from the diagonal dominance that all the adjacent entries equal to v_0 . Since every grid point is accessible by a series of adjacent grid points, all the entries of \tilde{v} equal to v_0 .

Similarly, we can define A from the left-hand side of (3.17). The matrix B is given by $B = \delta x^2 I$ where I represents the identity matrix. In this way, we express the scheme (3.16)–(3.18) in the form of (3.12)–(3.13). From now on, we can follow the derivation in Theorem 3.2 to prove the unique solvability and positivity.

The energy dissipation is derived by multiplying (3.16) with $(\mu_i)_{j,k}^{n+1}$ and taking the sum over $1 \leq j, k \leq M$. On the right-hand side, the summation by parts is then done by noting the boundary conditions of $(\mu_i)_{j,k}^{n+1}$. On the left-hand side, we deal with the terms with $\phi_{j,k}^{n+1}$ in the same way as the last equation in the proof of Theorem 3.2, using (3.17). □

3.2 Second-order scheme

Apparently we can use second-order BDF scheme with Adams-Bashforth extrapolation to construct a second-order scheme. However, since the Adams-Bashforth

extrapolation can not preserve positivity, we need to modify it with

$$c_i^* = \begin{cases} 2c_i^n - c_i^{n-1}, & \text{if } c_i^n \geq c_i^{n-1}, \\ \frac{1}{2/c_i^{n-1} - 1/c_i^n}, & \text{if } c_i^n < c_i^{n-1}. \end{cases} \tag{3.21}$$

Note that both expressions above are second-order extrapolation, meanwhile $c_i^* > 0$ if $c_i^n, c_i^{n-1} > 0$. Then, a second order fully-discretized scheme can be written as follows: to find $\{c_i^{n+1}\}$ and ϕ^{n+1} in X_M satisfying

$$\left[\frac{3c_i^{n+1} - 4c_i^n + c_i^{n-1}}{2\delta t}, v \right] = - \left[D_i c_i^* \nabla \left(I_M \left(\log c_i^{n+1} + z_i (\phi^{n+1} + \phi_e) \right) \right), \nabla v \right], \quad v \in X_M, \tag{3.22}$$

$$(\epsilon \nabla \phi^{n+1}, \nabla w) = \left[\rho_0 + \sum_{i=1}^N z_i c_i^{n+1}, w \right], \quad w \in X_M. \tag{3.23}$$

To obtain c_i^1 , a natural choice is to use the first-order scheme given in the above, so that c_i^1 preserves the desired properties.

Similar to the first-order scheme, we have

Theorem 3.5 *Assume that c_i^1 is calculated from the first-order scheme. The fully discretized scheme (3.22)–(3.23) enjoys the following properties:*

1. *Mass conservation:*

$$[c_i^{n+1}, 1] = [c_i^n, 1].$$

2. *Unique solvability: the scheme (3.22)–(3.23) possesses a unique solution $(\{c_i^{n+1} \in X_M\}, \phi^{n+1} \in X_M)$.*

3. *Positivity preserving: $c_i^{n+1}(z) > 0$ for all i and $z \in Z$.*

Proof The proof follows the same route of Theorem 3.2. Now, the matrix A_i^n is defined as

$$A_i^n = [D_i c_i^* \nabla \psi_z, \nabla \psi_{z'}],$$

which is positive semi-definite by noticing $c_i^*(z) > 0$. The matrices A and B are the same as defined in Theorem 3.2. The convex function \tilde{F} is defined as

$$\begin{aligned} & \tilde{F}[\tilde{c}_1^{n+1}, \dots, \tilde{c}_N^{n+1}] \\ &= \frac{1}{12\delta t} \sum_{i=1}^N (3\tilde{c}_i^{n+1} - 4\tilde{c}_i^n + \tilde{c}_i^{n-1})^t B(A_i^n)^* B(3\tilde{c}_i^{n+1} - 4\tilde{c}_i^n + \tilde{c}_i^{n-1}) \\ & \quad + \sum_{i=1}^N (\tilde{c}_i^{n+1})^t B(\log \tilde{c}_i^{n+1} - 1) \end{aligned}$$

$$+ \frac{1}{2} \left(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1} \right)^t B A^* B \left(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1} \right) + \tilde{\phi}_e^t B \left(\rho_0 + \sum_{i=1}^N z_i \tilde{c}_i^{n+1} \right).$$

Using the same derivation of Theorem 3.2, we conclude the proof. □

Remark 3.6 Unfortunately, we are unable to prove the energy dissipation. The reason is that we do not have an analog of (3.4) to deal with the term $(3c_i^{n+1} - 4c_i^n + c_i^{n-1}, \log c_i^{n+1})$.

4 Numerical experiments

In this section, we present several numerical experiments to validate our theoretical results in the previous section. We first present two examples to examine accuracy and stability of our schemes. In these two examples, the equations are solved in $[0, 2\pi]^2$ with periodic boundary conditions and discretized by Fourier spectral method in space. We will verify the convergence order as well as the mass conservation, positivity preserving and energy dissipation. Then, we present two other examples with Dirichlet and Neumann boundary conditions, one for two species and one for three species, on the domain $[0, 1]^2$, discretized with the finite difference scheme (3.16)–(3.17).

Note that at each time step, the scheme is nonlinear, but it is shown that it possesses a unique solution which is the minimizer of a strictly convex function. Hence, it can be solved efficiently by Newton’s iteration method. For a given Newton’s direction, line search is incorporated to obtain a damped step length. We adopt a simple backtracking line search method, to half the step length until the residue of the nonlinear equations decreases, which requires the concentration to be positive since we have logarithm functions in the nonlinear equations. The linear system to obtain the Newton’s direction is solved using the preconditioned GMRES iteration. For Fourier spatial discretization, we utilize the preconditioner given by choosing $\{c_i\}, \phi$ as constant functions. For finite difference discretization, the preconditioner is constructed by incomplete LU factorization without filling. The tolerance for Newton’s iteration is 10^{-9} , and that for GMRES iteration is chosen as 10^{-6} . This approach proves to be quite efficient, as we will present below.

Example 1 (Accuracy test) Let $z_1 = 1, z_2 = -1, p = c_1, n = c_2, D_1 = D_2 = 1$, and $\rho_0 = 0$, like in Sect. 2.2. We set the external field $\phi_e = 0$ and $\epsilon = 1$. We use the first-order and second-order schemes with 64×64 Fourier spectral modes for spatial discretization. The initial value is chosen as

$$p(x, 0) = 1.1 + \sin x \cos y, \quad n(x, 0) = 1.1 - \sin x \cos y.$$

The reference solution is obtained by the second-order scheme with $\delta t = 10^{-4}$. The errors by the two schemes are plotted in Fig. 1, which clearly shows the expected first- and second-order accuracy.

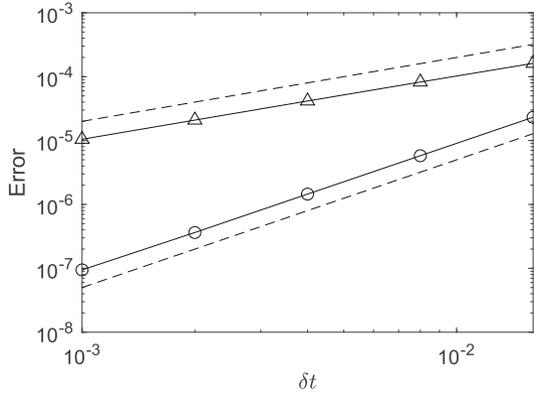


Fig. 1 (Example 1) Convergence rate of two schemes (triangle: first order; circle: second order). The error is calculated as $\sqrt{\|p^n(\cdot) - p(\cdot, t^n)\|^2 + \|n^n(\cdot) - n(\cdot, t^n)\|^2}$. The dashed lines represent the reference to first and second order convergence

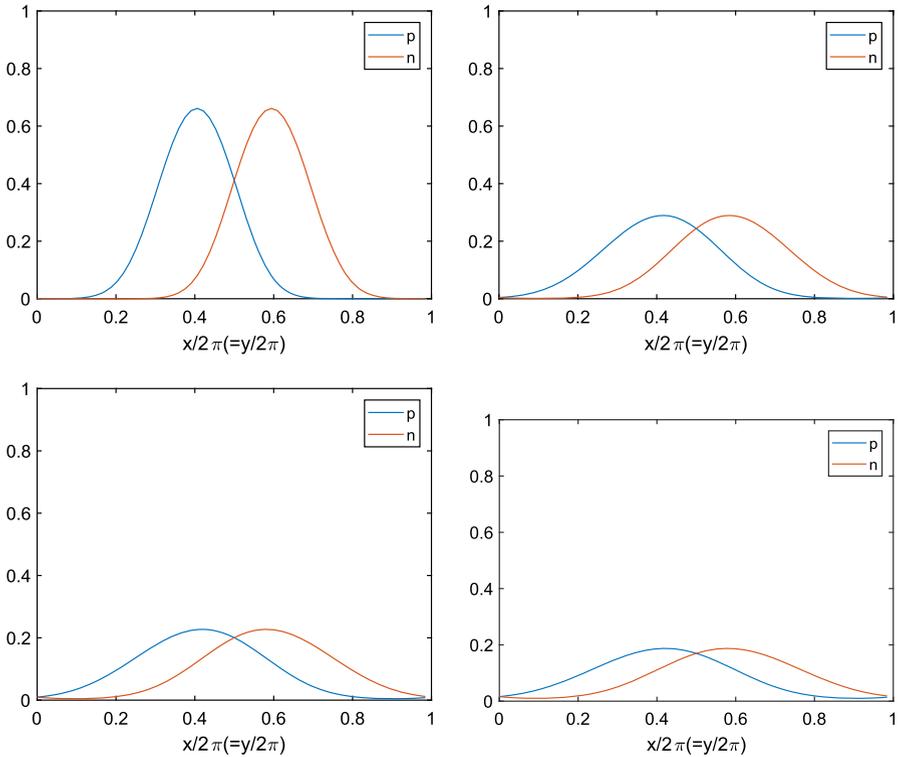


Fig. 2 (Example 2) Profiles of p and n on the line $x = y$, at $t = 0.2$ (upper-left), 0.6 (upper-right), 0.8 (lower-left), 1 (lower-right)

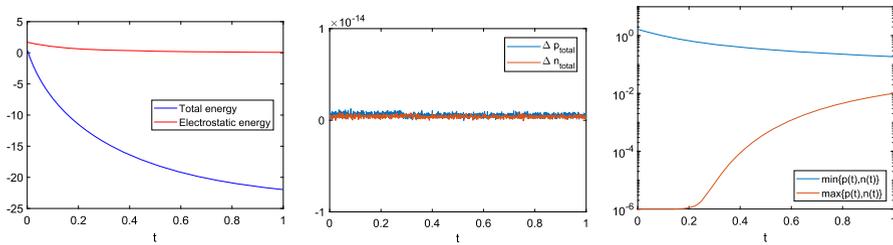


Fig. 3 (Example 2) Left: total energy density and electrostatic energy density. Middle: deviation of the average concentration to the initial. Right: Lower and upper bound

Example 2 (*Highly disparate initial value*) The domain, boundary conditions, ϕ_e, ϵ , and the spatial discretization are the same as Example 1. We choose the initial condition as follows,

$$p(x, y, 0) = 1 + 10^{-6} - \tanh\left(2((x - 0.8\pi)^2 + (y - 0.8\pi)^2 - (0.2\pi)^2)\right),$$

$$n(x, y, 0) = 1 + 10^{-6} - \tanh\left(2((x - 1.2\pi)^2 + (y - 1.2\pi)^2 - (0.2\pi)^2)\right),$$

so that $\min p(x, y, 0) = \min n(x, y, 0) \approx 10^{-6}$, $\max p(x, y, 0) = \max n(x, y, 0) \approx 1.65$. The initial condition indicates that the positive and negative charged particles accumulates in two regions centered at $(0.8\pi, 0.8\pi)$ and $(1.2\pi, 1.2\pi)$, respectively. By Sect. 2.2, the exact solution satisfies maximum principle and the dissipation of electrostatic potential.

We use the second-order scheme with the time step $\delta t = 10^{-3}$. To show the profiles of p and n , we plot them on the line $x = y$ at $t = 0.1, 0.2, 0.4, 1$ in Fig. 2. We also examine the energy dissipation of the total energy and the electrostatic energy in Fig. 3 (left), and find they indeed decrease as t grows. The change of average concentration is given in Fig. 3 (middle), where we find that the error is negligible. We also plot the lower and upper bounds of p and n about t in the right of Fig. 3, where we observe that the numerical results keep the maximum principle.

We also experiment with a larger time step $\delta t = 10^{-2}$, where the maximum principle and energy dissipation are still observed.

Efficiency of the scheme. Let us use the Example 2 to examine the efficiency. We plot the number of Newton iterations, and the maximum number of the GMRES iteration in each Newton step, for $\delta t = 10^{-3}$ and 10^{-2} . The number of the Newton iterations is slightly larger in the first few time steps, and for most time steps we only need 3–4 Newton iterations. For the larger time step, one intuitively expects that more Newton iterations are needed, but it turns out that we only need 1–2 more in this example.

Effect of boundary values. In the following two examples, we solve the PNP equations on $[0, 1]^2$. The Neumann boundary conditions are imposed on μ_i , while on ϕ the Dirichlet boundary conditions are imposed for the four solid line segments, $1/4 \leq x \leq 3/4, y = 0, 1$ and $1/4 \leq y \leq 3/4, x = 0, 1$, shown in Fig. 5. For other boundaries, the homogeneous Neumann boundary conditions are imposed. The

Fig. 4 (Example 2) Number of Newton iterations in each time step, for two δt

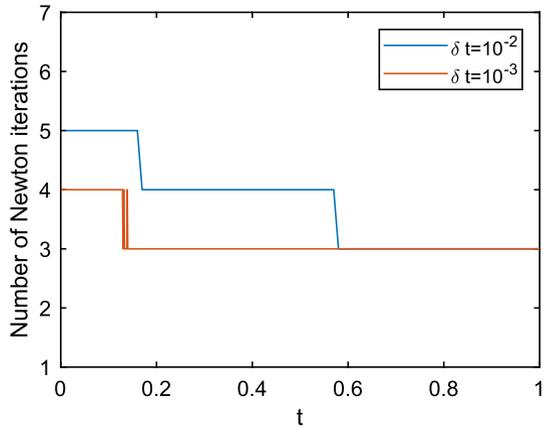
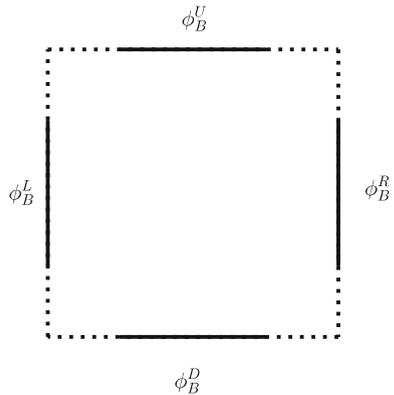


Fig. 5 Examples 3 & 4: illustration of boundary conditions on ϕ . On solid lines Dirichlet boundary conditions are imposed, while on dotted lines Neumann boundary conditions are imposed



external potential ϕ_e is obtained by solving $-\nabla \cdot (\epsilon \nabla \phi) = 0$ with the same types of boundary conditions in Fig. 5, but nonhomogeneous on the four solid line segments, specified by $\phi_B^L(y)$, $\phi_B^R(y)$, $\phi_B^D(x)$, $\phi_B^U(x)$. Recall that for ϕ we always assume homogeneous boundary conditions. So, it is equivalent to require that the total electric potential $\phi_{total} = \phi + \phi_e$ satisfies

$$-\nabla \cdot (\epsilon \nabla \phi_{total}) = \rho_0 + \sum_{i=1}^N z_i c_i,$$

$$\phi_{total}(0, y) = \phi_B^L(y), \quad \phi_{total}(1, y) = \phi_B^R(y), \quad \frac{1}{4} \leq y \leq \frac{3}{4}$$

$$\phi_{total}(x, 0) = \phi_B^D(x), \quad \phi_{total}(x, 1) = \phi_B^U(x), \quad \frac{1}{4} \leq x \leq \frac{3}{4}$$

$$\frac{\partial \phi_{total}}{\partial \mathbf{n}} = 0, \quad \text{elsewhere.}$$

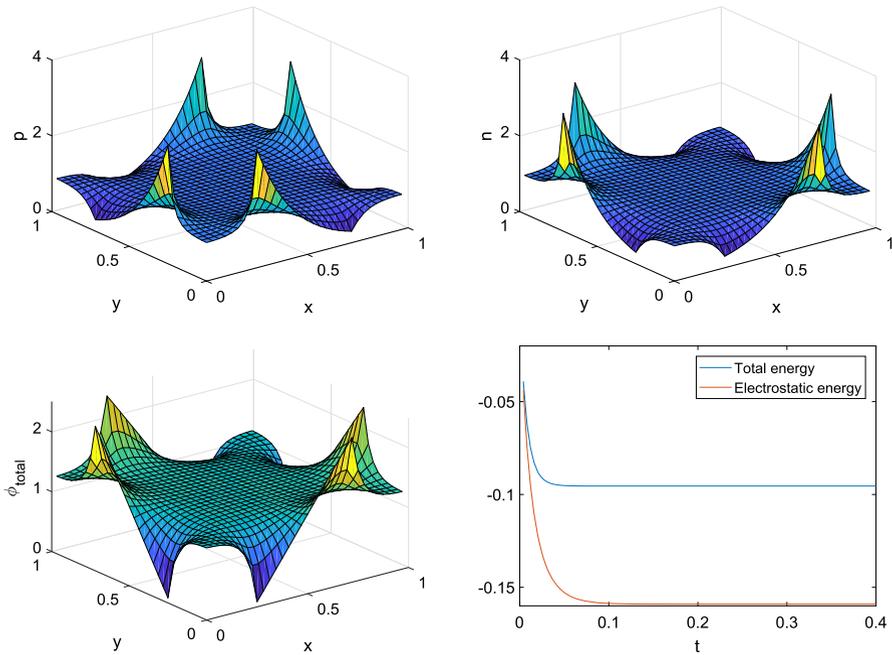


Fig. 6 (Example 3) Concentration, electric potential, and energy for $a = 2.5$

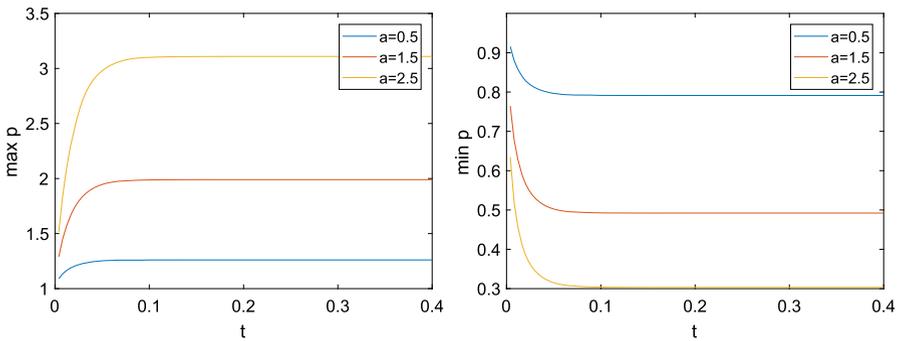


Fig. 7 (Example 3) Maximum and minimum concentration for different a

Example 3 (Two-component system with boundary potential) We let $z_1 = 1, z_2 = -1, p = c_1, n = c_2, D_1 = D_2 = 1,$ and $\epsilon = 0.01, \rho_0 = 0.$ The initial value is chosen as $p(x, y, 0) = n(x, y, 0) = 1.$ The boundary values are specified as follows,

$$\phi_B^L(y) = a \left(y - \frac{1}{4} \right), \phi_B^R(y) = a \left(\frac{3}{4} - y \right), \phi_B^D(x) = a \left(x - \frac{1}{4} \right), \phi_B^U(y) = a \left(\frac{3}{4} - x \right),$$

where a is a parameter to be varied. We discretize the space using finite difference method with 32×32 points, and solve the first-order scheme with the time step $\delta t = 4 \times 10^{-3}.$ The system reaches steady state after running 100 steps to $t = 0.4.$

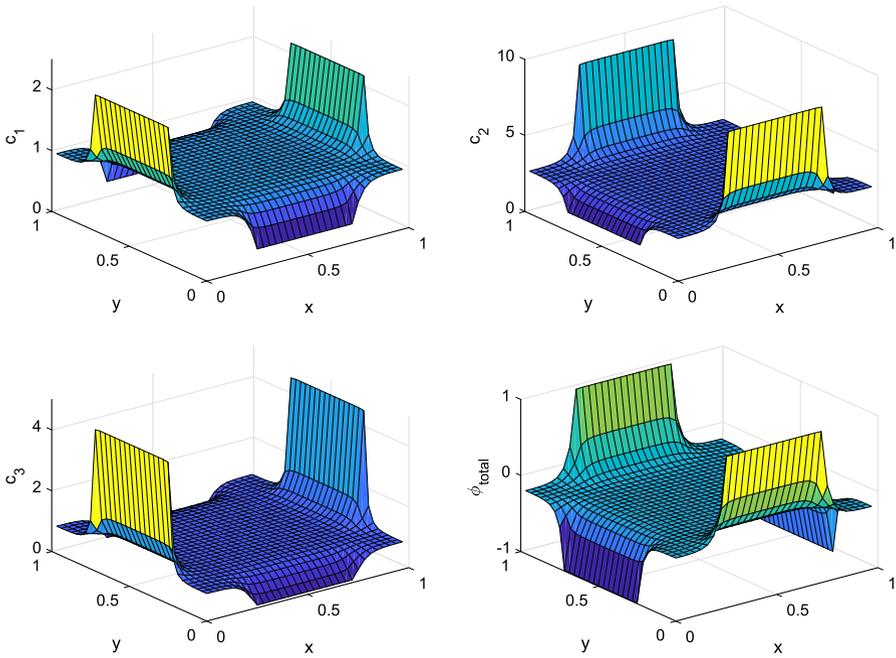


Fig. 8 (Example 4) Concentration and electric potential for $A = 1$

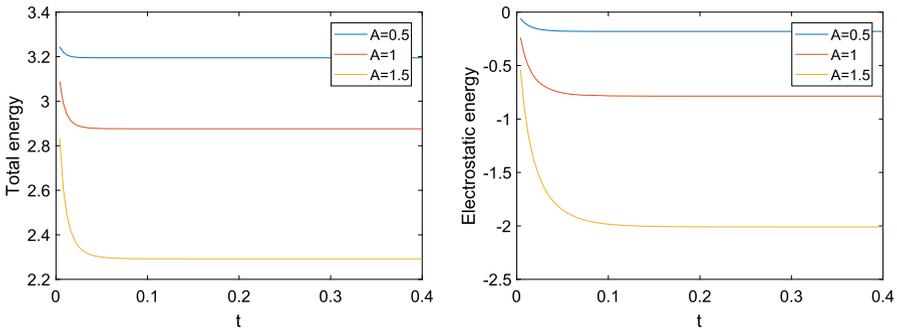


Fig. 9 (Example 4) Total energy and electrostatic energy for different boundary values

For $a = 2.5$, we plot p, n, ϕ_{total} in Fig. 6. They are mostly flat except near the boundary, with p peaking where ϕ_{total} reaches minimum on the boundary, n peaking where ϕ_{total} reaches maximum on the boundary. Actually, the profile of n is identical to the profile of p rotated by 90 degrees due to the symmetry of the boundary values on ϕ_{total} . The total energy and electrostatic energy are also plotted in Fig. 6, where the electrostatic energy here is defined by

$$\int \phi_{total}(p - n)dx dy.$$

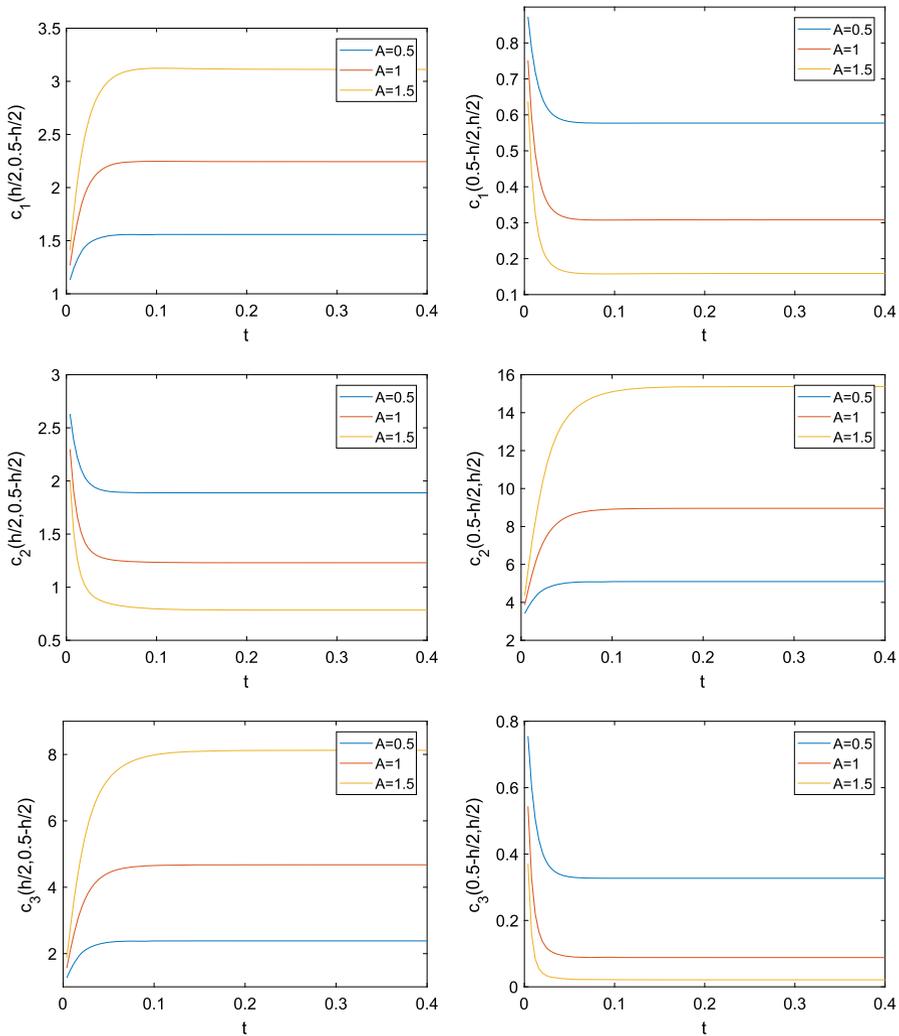


Fig. 10 (Example 4) The concentration evolution close to the left (left column) and the upper boundary (right column), for different boundary values. Here we plot the concentration at two grid points $(h/2, 0.5 - h/2)$ and $(0.5 - h/2, h/2)$, near the center of two boundaries, where $h = 1/32$

Both of them show dissipation, although for the latter it is not proved. We also examine how the maximum and minimum concentration evolve with different value of a . Because of the symmetry, we only plot p in Fig. 7.

Example 4 As the last example, we consider a three-component system. Choose $z_1 = 1, z_2 = -1, z_3 = 2$, and $D_1 = D_2 = D_3 = 1$. The other settings are identical to Example 3. The initial value is chosen as $c_1(x) = c_3(x) = 1$ and $c_2(x) = 3$ so that the system is electrically neutral. The boundary values are chosen as constants on each

line segments:

$$\phi_B^L = \phi_B^R = -A, \phi_B^D = \phi_B^U = A.$$

The spatial and time discretization are also identical to Example 3.

For $A = 1$, the concentration and total electric potential are plotted in Fig. 8. We also find that they are mostly flat except near the boundary. The two types of positive particles accumulate at the left and right boundaries, with c_3 larger, while the negative particles accumulate at the other two boundaries. We also compare the energy dissipation (Fig. 9) and the concentration near the boundaries (Fig. 10).

5 Concluding remarks

We proposed in this paper first- and second-order schemes for the PNP equations. We proved that both schemes are unconditionally mass conservative, uniquely solvable and positivity preserving; and that the first-order scheme is also unconditionally energy dissipative. While we can not prove the energy dissipation for the second-order scheme, our numerical result indicates that it is energy dissipative as well.

The schemes lead to nonlinear system at each time step but it possesses a unique solution which is the minimizer of a strictly convex functional. Hence, its solution can be efficiently obtained by using a Newton's iteration method. We presented ample numerical tests to verify the claimed properties for both first- and second-order schemes. In addition, in special cases where the PNP equation possesses maximum principle and electrostatic energy dissipation, our numerical results show that the schemes also satisfies them.

References

1. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Glows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich, 2nd edn. Birkhäuser Verlag, Basel (2008)
2. Archer, A.J., Rucklidge, A.M., Knobloch, E.: Quasicrystalline order and a crystal-liquid state in a soft-core fluid. *Phys. Rev. Lett.* **111**(16), 165501 (2013)
3. Badia, S., Guillén-González, F., Gutiérrez-Santacreu, J.V.: Finite element approximation of nematic liquid crystal flows using a saddle-point structure. *J. Comput. Phys.* **230**(4), 1686–1706 (2011)
4. Bazant, M.Z., Thornton, K., Ajdari, A.: Diffuse-charge dynamics in electrochemical systems. *Phys. Rev. E* **70**(2), 021506 (2004)
5. Biler, P., Hebisch, W., Nadzieja, T.: The Debye system: existence and large time behavior of solutions. *Nonlinear Anal.: Theory Methods Appl.* **23**(9), 1189–1209 (1994)
6. Cagni, E., Remondini, D., Mesirca, P., Castellani, G.C., Verondini, E., Bersani, F.: Effects of exogenous electromagnetic fields on a simplified ion channel model. *J. Biol. Phys.* **33**(3), 183–194 (2007)
7. Cahn, J.W., Elliott, C.M., Novick-Cohen, A.: The Cahn–Hilliard equation with a concentration dependent mobility: motion by minus the Laplacian of the mean curvature. *Eur. J. Appl. Math.* **7**(3), 287–302 (1996)
8. Chen, W., Wang, C., Wang, X., Wise, S.M.: Positivity-preserving, energy stable numerical schemes for the Cahn–Hilliard equation with logarithmic potential. *J. Comput. Phys.: X* **3**, 100031 (2019)
9. Copetti, M.I.M., Elliott, C.M.: Numerical analysis of the Cahn–Hilliard equation with a logarithmic free energy. *Numer. Math.* **63**(1), 39–65 (1992)

10. Corrias, L., Perthame, B., Zaag, H.: Global solutions of some chemotaxis and angiogenesis systems in high space dimensions. *Milan J. Math.* **72**(1), 1–28 (2004)
11. Ding, J., Wang, Z., Zhou, S.: Positivity preserving finite difference methods for Poisson–Nernst–Planck equations with steric interactions: application to slit-shaped nanopore conductance. *J. Comput. Phys.* **397**, 108864 (2019)
12. Doi, M., Edwards, S.F.: *The Theory of Polymer Dynamics*. Oxford University Press, Oxford (1988)
13. Eisenberg, B.: Ionic channels in biological membranes-electrostatic analysis of a natural nanotube. *Contemp. Phys.* **39**(6), 447–466 (1998)
14. Eisenberg, B., Hyon, Y., Liu, C.: A mathematical model for the hard sphere repulsion in ionic solutions. *Commun. Math. Sci.* **9**(2), 459–475 (2011)
15. Elliott, C.M., Stuart, A.M.: The global dynamics of discrete semilinear parabolic equations. *SIAM J. Numer. Anal.* **30**(6), 1622–1663 (1993)
16. Eyre, D.J.: Unconditionally gradient stable time marching the Cahn–Hilliard equation. In: *MRS Proceedings*, vol. 529, p. 39. Cambridge University Press (1998)
17. Fang, W.F., Itoi, K.: On the time-dependent drift-diffusion model for semiconductors. *J. Differ. Equ.* **117**(2), 245–280 (1995)
18. Flavell, A., Machen, M., Eisenberg, B., Kabre, J., Liu, C., Li, X.: A conservative finite difference scheme for Poisson–Nernst–Planck equations. *J. Comput. Electron.* **13**(1), 235–249 (2014)
19. Gajewski, H., Gröger, K.: On the basic equations for carrier transport in semiconductors. *J. Math. Anal. Appl.* **113**(1), 12–35 (1986)
20. Gao, H., He, D.: Linearized conservative finite element methods for the Nernst–Planck–Poisson equations. *J. Sci. Comput.* **72**(3), 1269–1289 (2017)
21. Gardner, C.L., Jones, J.R.: Electrodifffusion model simulation of the potassium channel. *J. Theor. Biol.* **291**, 10–13 (2011)
22. Gardner, C.L., Nonner, W., Eisenberg, R.S.: Electrodifffusion model simulation of ionic channels: 1D simulations. *J. Comput. Electron.* **3**(1), 25–31 (2004)
23. Gianazza, U., Savaré, G., Toscani, G.: The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.* **194**(1), 133–220 (2009)
24. Guillén-González, F., Tierra, G.: On linear schemes for a Cahn–Hilliard diffuse interface model. *J. Comput. Phys.* **234**, 140–171 (2013)
25. He, D., Pan, K.: An energy preserving finite difference scheme for the Poisson–Nernst–Planck system. *Appl. Math. Comput.* **287**, 214–223 (2016)
26. He, D., Pan, K., Yue, X.: A positivity preserving and free energy dissipative difference scheme for the Poisson–Nernst–Planck system. *J. Sci. Comput.* **81**(1), 436–458 (2019)
27. Horng, T.-L., Lin, T.-C., Liu, C., Eisenberg, B.: PNP equations with steric effects: a model of ion flow through channels. *J. Phys. Chem. B* **116**(37), 11422–11441 (2012)
28. Hu, J., Huang, X.: A fully discrete positivity-preserving and energy-dissipative finite difference scheme for Poisson–Nernst–Planck equations. *Numer. Math.* **145**, 1–39 (2020)
29. Keller, E.F., Segel, L.A.: Initiation of slime mold aggregation viewed as an instability. *J. Theor. Biol.* **26**(3), 399–415 (1970)
30. Liu, H., Maimaitiyiming, W.: Efficient, positive, and energy stable schemes for multi-D Poisson–Nernst–Planck Systems. *J. Sci. Comput.* **87**, 92 (2021)
31. Liu, H., Wang, Z.: A free energy satisfying finite difference method for Poisson–Nernst–Planck equations. *J. Comput. Phys.* **268**, 363–376 (2014)
32. Liu, H., Wang, Z.: A free energy satisfying discontinuous Galerkin method for one-dimensional Poisson–Nernst–Planck systems. *J. Comput. Phys.* **328**, 413–437 (2017)
33. Liu, J.-G., Wang, L., Zhou, Z.: Positivity-preserving and asymptotic preserving method for 2D Keller–Segal equations. *Math. Comput.* **87**(311), 1165–1189 (2018)
34. Lopreore, C.L., Bartol, T.M., Coggan, J.S., Keller, D.X., Sosinsky, G.E., Ellisman, M.H., Sejnowski, T.J.: Computational modeling of three-dimensional electrodiffusion in biological systems: application to the node of Ranvier. *Biophys. J.* **95**(6), 2624–2635 (2008)
35. Matthes, D., McCann, R.J., Savaré, G.: A family of nonlinear fourth order equations of gradient flow type. *Commun. Partial Differ. Equ.* **34**(11), 1352–1397 (2009)
36. Metti, M.S., Xu, J., Liu, C.: Energetically stable discretizations for charge transport and electrokinetic models. *J. Comput. Phys.* **306**, 1–18 (2016)
37. Nanninga, P.M.: A computational neuron model based on Poisson–Nernst–Planck theory. *ANZIAM J.* **50**, 46–59 (2008)

38. Otto, F.: Lubrication approximation with prescribed nonzero contact angle. *Commun. Partial Differ. Equ.* **23**(11–12), 2077–2164 (1998)
39. Prohl, A., Schmuck, M.: Convergent discretizations for the Nernst–Planck–Poisson system. *Numer. Math.* **111**(4), 591–630 (2009)
40. Ringhofer, C.A., Schmeiser, C., Markowich, P.A.: *Semiconductor Equations*. Springer, Berlin (1990)
41. Schmuck, M.: Analysis of the Navier–Stokes–Nernst–Planck–Poisson system. *Math. Models Methods Appl. Sci.* **19**(06), 993–1014 (2009)
42. Shen, J., Wang, C., Wang, X., Wise, S.M.: Second-order convex splitting schemes for gradient flows with Ehrlich–Schwoebel type energy: application to thin film epitaxy. *SIAM J. Numer. Anal.* **50**(1), 105–125 (2012)
43. Shen, J., Xu, J.: Convergence and error analysis for the scalar auxiliary variable (SAV) schemes to gradient flows. *SIAM J. Numer. Anal.* **56**(5), 2895–2912 (2018)
44. Shen, J., Xu, J.: Unconditionally bound preserving and energy dissipative schemes for a class of Keller–Segel equations. *SIAM J. Numer. Anal.* **58**(3), 1674–1695 (2020)
45. Shen, J., Xu, J., Yang, J.: The scalar auxiliary variable (SAV) approach for gradient flows. *J. Comput. Phys.* **353**, 407–416 (2018)
46. Shen, J., Xu, J., Yang, J.: A new class of efficient and robust energy stable schemes for gradient flows. *SIAM Rev.* **61**(3), 474–506 (2019)
47. Shen, J., Yang, X.: Numerical approximations of Allen–Cahn and Cahn–Hilliard equations. *Discrete Contin. Dyn. Syst.* **28**(4), 1669–1691 (2010)
48. Tang, T., Yang, J.: Implicit–explicit scheme for the Allen–Cahn equation preserves the maximum principle. *J. Comput. Math.* **34**(5), 451 (2016)
49. Jie, X., Zhang, P.: Onsager-theory-based dynamic model for nematic phases of bent-core molecules and star molecules. *J. Non-Newton. Fluid Mech.* **251**, 43–55 (2018)
50. Yang, X.: Linear, first and second-order, unconditionally energy stable numerical schemes for the phase field model of homopolymer blends. *J. Comput. Phys.* **327**, 294–316 (2016)
51. Zhang, X., Shu, C.-W.: On maximum-principle-satisfying high order schemes for scalar conservation laws. *J. Comput. Phys.* **229**(9), 3091–3120 (2010)
52. Zhang, X., Shu, C.-W.: On positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes. *J. Comput. Phys.* **229**(23), 8918–8934 (2010)
53. Zhao, J., Wang, Q., Yang, X.: Numerical approximations for a phase field dendritic crystal growth model based on the invariant energy quadratization approach. *Int. J. Numer. Methods Eng.* **110**(3), 279–300 (2016)
54. Zhu, J., Chen, L.Q., Shen, J., Tikare, V.: Coarsening kinetics from a variable mobility Cahn–Hilliard equation—application of semi-implicit Fourier spectral method. *Phys. Rev. E* **60**, 3564–3572 (1999)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.