SIAM J. NUMER. ANAL.
Vol. 59, No. 1, pp. 265–288

CONVERGENCE OF A SECOND-ORDER ENERGY-DECAYING METHOD FOR THE VISCOUS ROTATING SHALLOW WATER EQUATION*

GEORGIOS AKRIVIS†, BUYANG LI‡, AND JILU WANG§

Abstract. An implicit energy-decaying modified Crank–Nicolson time-stepping method is constructed for the viscous rotating shallow water equation on the plane. Existence, uniqueness, and convergence of semidiscrete solutions are proved by using Schaefer's fixed point theorem and H^2 estimates of the discretized hyperbolic–parabolic system. For practical computation, the semidiscrete method is further discretized in space, resulting in a fully discrete energy-decaying finite element scheme. A fixed-point iterative method is proposed for solving the nonlinear algebraic system. The numerical results show that the proposed method requires only a few iterations to achieve the desired accuracy, with second-order convergence in time, and preserves energy decay well.

Key words. viscous shallow water equation, energy decay, modified Crank–Nicolson, error estimate

AMS subject classifications. 65M12, 65M15, 35L60

DOI. 10.1137/20M1328051

1. Introduction.

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with smooth boundary $\partial\Omega$. We consider an initial and boundary value problem for the rotating viscous shallow water equation

$$\partial_t H = -\nabla \cdot (Hu) \qquad \text{in } \Omega \times (0,T], \tag{1.1}$$

$$\partial_t u = -\nabla\Big(\frac{1}{2}|u|^2 + g(H - H_b)\Big) - (\nabla \times u + f)\hat{k} \times u + \mathcal{G}(H,u) \quad \text{in } \Omega \times (0,T], \tag{1.2}$$

subject to the following initial and homogeneous Dirichlet boundary conditions:

$$u = 0 \qquad \text{on } \partial\Omega \times (0,T], \tag{1.3}$$

$$H|_{t=0} = H^0 \quad \text{and} \quad u|_{t=0} = u^0 \quad \text{in } \Omega, \tag{1.4}$$

where $H : \Omega \times [0,T] \to \mathbb{R}$ and $u = (u_1, u_2)^T : \Omega \times [0,T] \to \mathbb{R}^2$ denote the fluid thickness and velocity, respectively, and

$$\mathcal{G}(H,u) = \frac{\mu}{H}\nabla \cdot (H\nabla u) - c_f \frac{|u|u}{H} \tag{1.5}$$

*Received by the editors March 30, 2020; accepted for publication (in revised form) October 19, 2020; published electronically January 26, 2021.
https://doi.org/10.1137/20M1328051
Funding: The work of the authors was partially supported by the Hong Kong Polytechnic University project P0031035 ZZKQ and the National Natural Science Foundation of China grants U1930402 and 12071020.

†Department of Computer Science and Engineering, University of Ioannina, 451 10 Ioannina, Greece, and Institute of Applied and Computational Mathematics, FORTH, 700 13 Heraklion, Crete, Greece (akrivis@cse.uoi.gr).
‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong (buyang.li@polyu.edu.hk).
§Corresponding author. Beijing Computational Science Research Center, Beijing 100193, China (jiluwang@csrc.ac.cn).

consists of the viscous and friction forces, with $|u|$ denoting the magnitude of the velocity u.

In the two-dimensional plane, $\nabla \times u := \frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2}$ denotes the curl of the vector field u, and $\hat{k} \times \cdot$ denotes the rotation operator that rotates a vector field counterclockwise by the angle $\frac{\pi}{2}$, i.e.,

$$\hat{k} \times \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -v_2 \\ v_1 \end{pmatrix}.$$

The physical parameters and given functions in this model include

g: the gravity acceleration (positive constant),

f: the Coriolis term (function),

μ: viscosity of the fluid (positive constant),

c_f: Chezy coefficient for the bottom friction (positive constant),

H_b: the bathymetry (time-independent function).

The shallow water equations (1.1)–(1.2) describe the evolution of an incompressible fluid in response to gravitational and rotational accelerations for small enough ratio between the vertical and the horizontal scales. They are typically used to describe vertically averaged flows in three-dimensional domains in terms of horizontal velocity and depth variation. For smooth initial data such that $H^0 - H_b$ and u^0 are sufficiently small, it is known that the initial and boundary value problem (1.1)–(1.4) possesses a unique global smooth solution such that $H > 0$; cf. [27, 25, 26].

The numerical solution of the shallow water equation has wide applications in ocean modeling to study tidal fluctuations caused by earthquakes and storms and to study allowable discharge allocations by industries for water quality control. A nice introduction to the mathematical and computational modeling of ocean circulation, with a detailed derivation of the governing PDEs and an overview of early computational developments for such problems, is given in [14].

Many efforts have been devoted to developing efficient numerical methods and analyzing stability and convergence of numerical solutions for the shallow water equation. The energy boundedness of several first-order time-stepping methods for the viscous shallow water equation is proved in [1]. Convergence of numerical solutions to the viscous shallow water equation is established in [8] and [9] for a semidiscrete finite element method (FEM) and a fully discrete FEM, respectively, for a wave shallow water model proposed in [18]. The fully discrete FEM in [9] is linearly implicit and first-order in time and was shown to be convergent under a grid-ratio condition $\tau = O(h)$, which was used to prove the L^∞ boundedness of numerical solutions via an inverse inequality for finite element functions. Convergence of a fully discrete, first-order in time, nonlinearly implicit characteristic method for the shallow water equation is shown in [11], also under the grid-ratio condition $\tau = O(h)$ for the same reason. A leap-frog FEM is considered in [28] for the viscous shallow water equation on the unit sphere, and an error estimate is derived under the same grid-ratio condition $\tau = O(h)$. Exponential time differencing methods for the shallow water equations are constructed, discussed, and implemented in the recent paper [21]. A rigorous proof of convergence of the exponential time differencing method is still open.

Convergence of a Galerkin FEM with explicit Runge–Kutta methods in time is proved for the hyperbolic shallow water equation in one space dimension in [3], [4], and [15]. Optimal-order error estimates are established under the hyperbolic CFL condition $\tau = O(h)$. For an overview of numerical methods for the nonlinear hyperbolic shallow water equations and related models, with the main emphasis on the

spatial discretization and on practical issues, as well as for references to the original literature, we refer to the recent review paper [30].

Since the energy of the solutions to the viscous shallow water equation always decays in time, it is desirable to preserve this property in numerical solutions. Some high-order well-balanced and energy-conserving explicit methods have been constructed for the hyperbolic shallow water equation; for example, see [6, 7, 19, 31]. For the viscous shallow water equation, these explicit methods would preserve energy decay but require a CFL condition $\tau = O(h^2)$. As far as we know, no implicit methods have been reported to preserve energy decay for the viscous shallow water equation without requiring a CFL condition. Therefore, the construction of energy-decaying numerical methods (especially second-order methods without CFL conditions) for viscous shallow water equations is still challenging.

As far as we know, all existing error analyses of implicit and linearly implicit time-stepping methods for the viscous shallow water equation require the grid-ratio condition $\tau = O(h)$, which is natural for the hyperbolic shallow water equation but may not be necessary for the viscous model. Otherwise the semidiscretization in time (corresponding to the case $h \to 0$ in the full discretization) may not converge with optimal-order. The grid-ratio condition was used to prove the L^∞ boundedness of numerical solutions by an inverse inequality of finite element functions, which may be avoided if (i) convergence of the semidiscretization method in the H^2 norm with respect to the time stepsize can be established and (ii) the error splitting approach in [16, 17, 29] can be adopted for analysis of the fully discrete FEM. (This approach requires the temporal semidiscrete solutions to be bounded in the H^2 norm uniformly in temporal stepsizes.)

In this paper, we propose a second-order energy-decaying modified Crank–Nicolson method for the viscous problem (1.1)–(1.4) and establish (i), i.e., optimal-order convergence of the semidiscretization method in the H^2 norm with respect to the time stepsize. This would provide a foundation for error analysis of fully discrete FEMs using the approach mentioned in (ii) without a grid-ratio condition. The analysis of H^2 convergence of the proposed nonlinearly implicit temporal semidiscretization for the hyperbolic–parabolic system (1.1)–(1.4) is different from all existing work using the error splitting approach, e.g., [16, 17, 29], which all concern only nonlinear parabolic equations and linearly semi-implicit schemes. The derivation of the error estimates in this paper is based on the boundedness of the numerical solutions in $H^2(\Omega)$ (uniformly with respect to the stepsize τ), proved by Schaefer's fixed-point theorem, combined with discrete $L^\infty(0,T;H^2(\Omega))$ and $L^2(0,T;H^3(\Omega))$ estimates of the Crank–Nicolson scheme.

2. Energy decay and time discretization. In this section, we present a second-order implicit modified Crank–Nicolson method preserving the energy decay property of the viscous shallow water equation.

2.1. Energy decay property. Testing (1.1) by $\frac{1}{2}|u|^2 + g(H - H_b)$ and (1.2) by Hu, we obtain

$$\int_\Omega \partial_t H\Big(\frac{1}{2}|u|^2 + g(H - H_b)\Big)\mathrm{d}x = -\int_\Omega \nabla\cdot(Hu)\Big(\frac{1}{2}|u|^2 + g(H - H_b)\Big)\mathrm{d}x \tag{2.1}$$

and

$$\int_\Omega \partial_t u\cdot(Hu)\mathrm{d}x = -\int_\Omega \nabla\Big(\frac{1}{2}|u|^2+g(H-H_b)\Big)\cdot(Hu)\mathrm{d}x+\int_\Omega \mathcal{G}(H,u)\cdot(Hu)\mathrm{d}x, \tag{2.2}$$

respectively, where we used the orthogonality $(\hat{k} \times u) \cdot (Hu) = 0$ in the derivation of (2.2). By adding the relations (2.1) and (2.2), the first terms on their right-hand sides cancel, while the first terms on their left-hand sides can be combined using the product rule of differentiation. Therefore, we obtain

$$\frac{d}{dt} \int_\Omega \frac{1}{2} \Big(|u|^2 H + g(H - H_b)^2 \Big) \mathrm{d}x = \int_\Omega \mathcal{G}(H, u) \cdot (Hu) \mathrm{d}x. \tag{2.3}$$

We infer that the energy $E(u, H)$,

$$E(u, H) := \int_\Omega \frac{1}{2} \Big(|u|^2 H + g(H - H_b)^2 \Big) \mathrm{d}x, \tag{2.4}$$

satisfies the relation

$$\frac{d}{dt} E(u, H) = \int_\Omega \mathcal{G}(H, u) \cdot (Hu) \mathrm{d}x = -\mu \int_\Omega H |\nabla u|^2 \, \mathrm{d}x - c_f \int_\Omega |u|^3 \, \mathrm{d}x \leqslant 0 \tag{2.5}$$

if $H > 0$, where we have used the expression of $\mathcal{G}(H, u)$ in (1.5) and integration by parts. This shows that the energy is decaying.

2.2. A modified Crank–Nicolson method. Let $t_n := n\tau$, $n = 0, 1, \ldots, N$, be the nodes of a uniform partition of the time interval $[0, T]$ with stepsize $\tau = T/N$, and denote

$$\bar{\partial}_\tau v^n := (v^n - v^{n-1})/\tau \quad \text{and} \quad v^{n-\frac{1}{2}} := (v^n + v^{n-1})/2.$$

We consider the temporal discretization of the initial and boundary value problem (1.1)–(1.2) by the following implicit scheme: for given $H^{n-1} \in H^1(\Omega)$ and $u^{n-1} \in [H^2(\Omega) \cap H_0^1(\Omega)]^2$, find $H^n \in H^1(\Omega)$ and $u^n \in [H^2(\Omega) \cap H_0^1(\Omega)]^2$ satisfying the following equations:

$$\begin{cases} \bar{\partial}_\tau H^n = -\nabla \cdot \big(H^{n-\frac{1}{2}} u^{n-\frac{1}{2}} \big), \\ \bar{\partial}_\tau u^n = -\nabla \Big(\dfrac{1}{4} \big(|u^n|^2 + |u^{n-1}|^2 \big) + g(H^{n-\frac{1}{2}} - H_b) \Big) \\ \qquad\qquad - (\nabla \times u^{n-\frac{1}{2}} + f^{n-\frac{1}{2}}) \hat{k} \times u^{n-\frac{1}{2}} + \mathcal{G}(H^{n-\frac{1}{2}}, u^{n-\frac{1}{2}}), \quad n = 1, \ldots, N, \end{cases} \tag{2.6}$$

with starting values H^0 and u^0 being the given initial values in (1.4).

In the standard Crank–Nicolson method, the term $|u|^2$ in (1.2) would be discretized in the form $|u^{n-\frac{1}{2}}|^2$; instead, in the second equation of (2.6), we discretized it by $\big(|u^n|^2 + |u^{n-1}|^2 \big)/2$; this modification of the Crank–Nicolson method is familiar from [24] for the nonlinear Klein–Gordon equation and from [12] for the nonlinear Schrödinger equation; see also [2]. A different type of modified Crank–Nicolson method was also used for preserving the energy decay property of the Cahn–Hilliard equation; see [22].

2.3. Energy decay of discrete solutions.

THEOREM 2.1. *If* $(H^n, u^n) \in H^1(\Omega) \times [H^2(\Omega) \cap H_0^1(\Omega)]^2, n = 1, \ldots, N$, *is a solution of* (2.6) *satisfying*

$$H^n > 0, \quad n = 1, \ldots, N,$$

then the modified Crank–Nicolson scheme (2.6) *is energy-decaying, i.e.,*

$$E(u^n, H^n) \leqslant E(u^{n-1}, H^{n-1}), \quad n = 1, \ldots, N. \tag{2.7}$$

Proof. Testing the first equation in (2.6) by $\frac{1}{4}\big(|u^n|^2+|u^{n-1}|^2\big)+g(H^{n-\frac{1}{2}}-H_b)$ and the second by $H^{n-\frac{1}{2}}u^{n-\frac{1}{2}}$, adding the results, and noticing that, as in the continuous case, the first terms on the right-hand sides cancel, and the second term on the right-hand side of the second relation vanishes due to the orthogonality $(\hat{k}\times u^{n-\frac{1}{2}})\cdot(H^{n-\frac{1}{2}}u^{n-\frac{1}{2}})=0$, we obtain

$$\begin{aligned}\int_\Omega \Big[\bar\partial_\tau H^n\Big(\frac{1}{4}\big(|u^n|^2+|u^{n-1}|^2\big)+g(H^{n-\frac{1}{2}}-H_b)\Big)+\bar\partial_\tau u^n\cdot\big(H^{n-\frac{1}{2}}u^{n-\frac{1}{2}}\big)\Big]\mathrm{d}x\\ =\int_\Omega \mathcal{G}(H^{n-\frac{1}{2}},u^{n-\frac{1}{2}})\cdot\big(H^{n-\frac{1}{2}}u^{n-\frac{1}{2}}\big)\mathrm{d}x.\end{aligned}\tag{2.8}$$

Then, substituting the identities

$$\begin{aligned}\bar\partial_\tau u^n\cdot\big(H^{n-\frac{1}{2}}u^{n-\frac{1}{2}}\big)&=\frac{1}{2\tau}H^{n-\frac{1}{2}}\big(|u^n|^2-|u^{n-1}|^2\big)\\ &=\frac{1}{4\tau}\big(H^n|u^n|^2-H^{n-1}|u^{n-1}|^2\big)\\ &\quad+\frac{1}{4\tau}\big(H^{n-1}|u^n|^2-H^n|u^{n-1}|^2\big),\\ \bar\partial_\tau H^n\frac{1}{4}\big(|u^n|^2+|u^{n-1}|^2\big)&=\frac{1}{4\tau}\big(H^n|u^n|^2-H^{n-1}|u^{n-1}|^2\big)\\ &\quad-\frac{1}{4\tau}\big(H^{n-1}|u^n|^2-H^n|u^{n-1}|^2\big),\\ \bar\partial_\tau H^n g(H^{n-\frac{1}{2}}-H_b)&=\frac{g}{2\tau}\big[(H^n-H_b)^2-(H^{n-1}-H_b)^2\big]\end{aligned}$$

into (2.8), we obtain

$$\begin{aligned}\int_\Omega\frac{1}{2}\Big(H^n|u^n|^2+g\big(H^n-H_b\big)^2\Big)\mathrm{d}x&=\int_\Omega\frac{1}{2}\Big(H^{n-1}|u^{n-1}|^2+g\big(H^{n-1}-H_b\big)^2\Big)\mathrm{d}x\\ &\quad+\tau\int_\Omega\mathcal{G}(H^{n-\frac{1}{2}},u^{n-\frac{1}{2}})\cdot\big(H^{n-\frac{1}{2}}u^{n-\frac{1}{2}}\big)\mathrm{d}x.\end{aligned}$$

By using the expression (1.5) of the viscous and friction forces, we have

$$\int_\Omega\mathcal{G}(H^{n-\frac{1}{2}},u^{n-\frac{1}{2}})\cdot\big(H^{n-\frac{1}{2}}u^{n-\frac{1}{2}}\big)\mathrm{d}x=-\int_\Omega\big(\mu H^{n-\frac{1}{2}}|\nabla u^{n-\frac{1}{2}}|^2+c_f|u^{n-\frac{1}{2}}|^3\big)\,\mathrm{d}x\leqslant 0.$$

This implies the energy decay property (2.7). ∎

For practical computation, the semidiscrete scheme (2.6) can be further discretized in space by the FEM: find $(H_h^n,u_h^n)\in S_h\times\mathring{S}_h^2$, with $\mathring{S}_h\subset H_0^1(\Omega)$, satisfying the weak formulation

$$\begin{cases}(\bar\partial_\tau H_h^n,\phi_h)-\big(H_h^{n-\frac{1}{2}}u_h^{n-\frac{1}{2}},\nabla\phi_h\big)=0 & \forall\,\phi_h\in S_h,\\ (\bar\partial_\tau u_h^n,H_h^{n-\frac{1}{2}}v_h)+\big(\mu H_h^{n-\frac{1}{2}}\nabla u_h^{n-\frac{1}{2}},\nabla v_h\big)+\big(c_f|u_h^{n-\frac{1}{2}}|u_h^{n-\frac{1}{2}},v_h\big)\\ \quad=-\Big(\nabla P_h\Big[\frac{1}{4}\big(|u_h^n|^2+|u_h^{n-1}|^2\big)+g(H_h^{n-\frac{1}{2}}-H_b)\Big],H_h^{n-\frac{1}{2}}v_h\Big)\\ \quad\quad-\big((\nabla\times u_h^{n-\frac{1}{2}}+f^{n-\frac{1}{2}})\hat{k}\times u_h^{n-\frac{1}{2}},H_h^{n-\frac{1}{2}}v_h\big) & \forall\,v_h\in\mathring{S}_h^2\end{cases}\tag{2.9}$$

with starting values H_h^0 and u_h^0 being the Lagrange interpolants of H^0 and u^0, respectively. Here, S_h and $\mathring{S}_h^2$ are the scalar- and vector-valued finite element spaces, respectively, consisting of globally continuous piecewise linear polynomials. The L^2-projection P_h (onto the finite element space S_h) in (2.9) ensures that the energy decay property is unconditionally preserved also in the fully discrete case. The proof proceeds along the lines of the proof of Theorem 2.1, i.e., substituting

$$\phi_h = P_h\left[\frac{1}{4}\big(|u_h^n|^2 + |u_h^{n-1}|^2\big) + g(H_h^{n-\frac{1}{2}} - H_b)\right] \quad \text{and} \quad v_h = u_h^{n-\frac{1}{2}}$$

into (2.9) and summing up the two equations, we obtain

$$\begin{aligned}\int_\Omega \frac{1}{2}\Big(H_h^n|u_h^n|^2 + g\big(H_h^n - H_b\big)^2\Big)\mathrm{d}x &= \int_\Omega \frac{1}{2}\Big(H_h^{n-1}|u_h^{n-1}|^2 + g\big(H_h^{n-1} - H_b\big)^2\Big)\mathrm{d}x \\ &\quad - \int_\Omega \big(\mu H_h^{n-\frac{1}{2}}|\nabla u_h^{n-\frac{1}{2}}|^2 + c_f|u_h^{n-\frac{1}{2}}|^3\big)\,\mathrm{d}x \\ &\leqslant \int_\Omega \frac{1}{2}\Big(H_h^{n-1}|u_h^{n-1}|^2 + g\big(H_h^{n-1} - H_b\big)^2\Big)\mathrm{d}x,\end{aligned}$$

which holds whenever $H_h^{n-\frac{1}{2}} \geqslant 0$. This proves the energy decay property of the fully discrete FEM.

2.4. Existence, uniqueness, and convergence of discrete solutions. For simplicity, we denote by $(\cdot,\cdot)$ and $\|\cdot\|$ the inner products and norms on both $L^2(\Omega)$ and $\big(L^2(\Omega)\big)^2$, by $\|\cdot\|_{H^m}$ the norms on $H^m(\Omega)$ and on $\big(H^m(\Omega)\big)^2$, and similarly for norms on $L^p(\Omega)$-based Sobolev spaces.

We assume the following:

(A1) The domain Ω is sufficiently smooth, and the solution (H,u) of the initial and boundary value problem (1.1)–(1.4) is sufficiently smooth.

(A2) The solution of (1.1)–(1.4) satisfies

$$\inf_{(x,t)\in\Omega\times[0,T]} H(x,t) \geqslant H_{\min} \quad \text{for some positive constant } H_{\min}.$$

We denote by $B_{H,u}^n$ the set of pairs $(\widetilde{H},\widetilde{u}) \in H^2(\Omega)\times[H^3(\Omega)\cap H_0^1(\Omega)]^2$ such that

$$\|\widetilde{H} - H(t_n)\|_{H^2} + \|\widetilde{H} - H(t_n)\|_{L^\infty} \leqslant \frac{H_{\min}}{2} \quad \text{and} \quad \|\widetilde{u} - u(t_n)\|_{H^3} \leqslant 1, \tag{2.10}$$

where $H(t_n) := H(\cdot,t_n)$ and $u(t_n) := u(\cdot,t_n)$ are the exact solutions at the time level $t = t_n$. Thus $B_{H,u}^n$ is a neighborhood of the solution $(H(t_n),u(t_n))$ in the space $H^2(\Omega)\times[H^3(\Omega)\cap H_0^1(\Omega)]^2$.

The main theoretical result of this paper is the following.

THEOREM 2.2. *Under assumptions* (A1)–(A2)*, there exists a positive constant* τ_0 *such that, for* $\tau \leqslant \tau_0$*, the modified Crank–Nicolson method* (2.6) *has a unique solution* $(H^n,u^n) \in B_{H,u}^n$ *for* $n = 1,\dots,N$*. Moreover, the solution satisfies the following error estimate:*

$$\max_{1\leqslant n\leqslant N}\big(\|H(t_n) - H^n\|_{H^2} + \|u(t_n) - u^n\|_{H^2}\big) + \left(\tau\sum_{n=1}^N \|u(t_{n-\frac{1}{2}}) - u^{n-\frac{1}{2}}\|_{H^3}^2\right)^{\frac{1}{2}} \leqslant C\tau^2 \tag{2.11}$$

with a constant C *independent of* τ*.*

Remark 2.1. An immediate consequence of Theorem 2.2 is that the discrete solution satisfies, for $\tau \leqslant \tau_0$,

$$\|H^n\|_{H^2} + \|u^n\|_{H^3} + \|(u^n - u^{n-1})/\tau\|_{H^2} \leqslant C. \tag{2.12}$$

This result can be used for error analysis of fully discrete FEMs, which can be viewed as the spatial discretization of the semidiscrete problem (2.6), whose solution has regularity (2.12). By utilizing this regularity result (uniformly in the stepsize τ), one can expect that the error between semi and fully discrete solutions has the following bound:

$$\|H^n - H_h^n\| + \|u^n - u_h^n\| \leqslant Ch^2 \tag{2.13}$$

with a right-hand side independent of τ. Such a type of error estimate (independent of τ) has been proved in [17, 16, 29] for many nonlinear parabolic and wave equations. With such results as (2.13), convergence and boundedness of fully discrete numerical solutions in $L^\infty(0,T;L^\infty(\Omega))$ can be proved, for $\tau \leqslant \tau_0$ and $h \leqslant h_0$, by using (2.13) and an inverse inequality, without requiring any grid-ratio condition.

3. Proof of Theorem 2.2. In this section, we prove the existence and uniqueness of discrete solutions for sufficiently small time stepsize τ and establish a second-order error estimate.

3.1. H^2 and H^3 estimates for the Crank–Nicolson scheme. In this subsection, we present some H^2 and H^3 estimates of the Crank–Nicolson scheme for the heat equation, which will be used in our error estimation for the shallow water equation. Our main tool will be the following resolvent estimates.

LEMMA 3.1 (resolvent estimates). *The Dirichlet Laplacian operator $\Delta : H^2(\Omega) \cap H_0^1(\Omega) \to L^2(\Omega)$ satisfies the resolvent estimates*

$$\|\Delta(z-\Delta)^{-1}f\|_{L^2} \leqslant C\|f\|_{L^2} \quad \text{if } f \in L^2(\Omega),\ z \in \mathbb{C} \text{ and } \operatorname{Re} z \geqslant 0, \tag{3.1}$$

$$\|\Delta(z-\Delta)^{-1}f\|_{H^1} \leqslant C\|f\|_{H^1} \quad \text{if } f \in H^1(\Omega),\ z \in \mathbb{C} \text{ and } \operatorname{Re} z \geqslant 0 \tag{3.2}$$

with a constant C independent of z.

Proof. It is well known that the Dirichlet Laplacian $\Delta : H^2(\Omega) \cap H_0^1(\Omega) \to L^2(\Omega)$ generates a bounded analytic semigroup on $L^2(\Omega)$; see [20]. Equivalently, $z - \Delta$ is invertible for $z \in \mathbb{C}$ such that $\operatorname{Re} z \geqslant 0$ and the resolvent estimate (3.1) holds; see [5, Example 3.7.5 and Theorem 3.7.11].

For $f \in H_0^1(\Omega)$,

$$\begin{aligned}\|\Delta(z-\Delta)^{-1}f\|_{H^1} &= \|(-\Delta)^{\frac{1}{2}}(z-\Delta)^{-1}(-\Delta)^{\frac{1}{2}}f\|_{H^1}\\ &\leqslant \|\Delta(z-\Delta)^{-1}(-\Delta)^{\frac{1}{2}}f\|_{L^2}\\ &\leqslant C\|(-\Delta)^{\frac{1}{2}}f\|_{L^2} \leqslant C\|f\|_{H^1},\end{aligned}$$

where we have used (3.1) in the second to last inequality. This proves (3.2) for $f \in H_0^1(\Omega)$.

For $f \in H^1(\Omega)$, we choose a sequence of functions $f_n = \mathrm{e}^{n^{-1}\Delta} f \in H_0^1(\Omega)$, $n \in \mathbb{N}$. Then f_n is the solution of the heat equation at time $t = 1/n$ with initial value f, satisfying the following standard estimate:

$$f_n \to f \text{ in } L^2(\Omega) \text{ as } n \to \infty \quad \text{and} \quad \|f_n\|_{H^1} \leqslant C\|f\|_{H^1}. \tag{3.3}$$

In fact, we have $f_n = \mathrm{e}^{n^{-1}\Delta} f \to f$ in $L^2(\Omega)$ because the heat semigroup is strongly continuous (and analytic) on $L^2(\Omega)$; see [20, Theorem 2.4]. The estimate $\|f_n\|_{H^1} \leqslant C\|f\|_{H^1}$ can be proved as follows: let v be the solution of the heat equation

$$\partial_t v - \Delta v = 0 \quad \text{with initial condition } v(\cdot,0) = f \in H^1(\Omega),$$

so that $f_n = v(\cdot, 1/n)$. Then, testing this equation by $\partial_t v$ yields

$$\|\partial_t v\|_{L^2}^2 + \frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{1}{2}\|\nabla v\|_{L^2}^2\Big) = 0,$$

which implies $\|\nabla v(\cdot,t)\|_{L^2} \leqslant \|\nabla v(\cdot,0)\|_{L^2}$. This proves that $\|\nabla f_n\|_{L^2} \leqslant \|\nabla f\|_{L^2}$, and, together with the standard L^2 stability estimate $\|f_n\|_{L^2} \leqslant \|f\|_{L^2}$, leads to $\|f_n\|_{H^1} \leqslant \|f\|_{H^1}$.

Since (3.2) holds for $f_n \in H_0^1(\Omega)$, it follows that

$$\|\Delta(z-\Delta)^{-1} f_n\|_{H^1} \leqslant C\|f_n\|_{H^1} \leqslant C\|f\|_{H^1}. \tag{3.4}$$

This proves that $\Delta(z-\Delta)^{-1} f_n$ is bounded in $H^1(\Omega)$. On the one hand, there exists a subsequence $\Delta(z-\Delta)^{-1} f_{n_k}$ which converges weakly in $H^1(\Omega)$. On the other hand, $\Delta(z-\Delta)^{-1} f_{n_k}$ converges strongly in $L^2(\Omega)$, because $f_{n_k} \to f$ in $L^2(\Omega)$ and $\Delta(z-\Delta)^{-1}$ is a bounded linear operator on $L^2(\Omega)$. Thus $\Delta(z-\Delta)^{-1} f_{n_k}$ converges weakly in $H^1(\Omega)$ to $\Delta(z-\Delta)^{-1} f$. Then (3.4) implies (cf. [10, Theorem 5.12-2])

$$\|\Delta(z-\Delta)^{-1} f\|_{H^1} \leqslant \liminf_{n_k \to \infty} \|\Delta(z-\Delta)^{-1} f_{n_k}\|_{H^1} \leqslant C\|f\|_{H^1},$$

where we have used (3.4) in the last inequality. This proves the desired results. ∎

LEMMA 3.2. *For a given sequence $(f^n)_{n\in\mathbb{N}} \subset H^s(\Omega)$ with $s \in \{0,1\}$ and starting value $v^0 = 0$ in Ω, consider the sequence $(v^n)_{n\in\mathbb{N}}$ with $v^n \in H_0^1(\Omega)$ satisfying*

$$\bar\partial_\tau v^n - \Delta v^{n-\frac{1}{2}} = f^n \quad \text{in } \Omega, \quad n \in \mathbb{N}. \tag{3.5}$$

Then, there exists a positive constant C, independent of τ and of the sequence $(f^n)_{n\in\mathbb{N}}$, such that, for any $m \in \mathbb{N}$,

$$\max_{1\leqslant n\leqslant m} \|v^n\|_{H^{1+s}}^2 + \tau \sum_{n=1}^{m} \|v^{n-\frac{1}{2}}\|_{H^{2+s}}^2 \leqslant C\tau \sum_{n=1}^{m} \|f^n\|_{H^s}^2. \tag{3.6}$$

Proof. Without loss of generality, we assume that $f^n = 0$ for $n \geqslant m+1$. Otherwise we set $f^n = 0$ for $n \geqslant m+1$ without affecting the value of v^n for $n \leqslant m$. Then $f = (f^n)_{n=1}^\infty$ is an $L^2(\Omega)$-valued square summable sequence.

Let $v = (v^n)_{n=1}^\infty$ and $f = (f^n)_{n=1}^\infty$, and denote by $\mathcal{F}$ the Fourier $\mathbb{Z}$-transform, which transforms a square summable sequence $f = (f^n)_{n=1}^\infty$ to a function

$$\mathcal{F}f(\zeta) = \sum_{n=1}^{\infty} \zeta^n f^n$$

defined a.e. for ζ on the unit disk $\mathbb{D}$ on the complex plane.

Multiplying (3.5) by ζ^n and summing up the equations for $n = 1, 2, \ldots$, we obtain

$$\left(\frac{1-\zeta}{\tau} - \frac{1+\zeta}{2}\Delta\right)\mathcal{F}v(\zeta) = \mathcal{F}f(\zeta), \tag{3.7}$$

which furthermore implies

$$\frac{1+\zeta}{2}\Delta\mathcal{F}v(\zeta)=\Delta\left(\frac{2}{\tau}\frac{1-\zeta}{1+\zeta}-\Delta\right)^{-1}\mathcal{F}f(\zeta)$$

and therefore (taking the inverse transform of $\mathcal{F}$)

$$(\Delta v^{n-\frac{1}{2}})_{n=1}^{\infty}=\mathcal{F}^{-1}M(\zeta)\mathcal{F}\,(f^n)_{n=1}^{\infty}, \tag{3.8}$$

where $M(\zeta)=\Delta\big(\frac{2}{\tau}\frac{1-\zeta}{1+\zeta}-\Delta\big)^{-1}$.

Since $\mathrm{Re}(\frac{2}{\tau}\frac{1-\zeta}{1+\zeta})\geqslant 0$ for $\zeta\in\partial\mathbb{D}\backslash\{\pm1\}$, it follows that the operator $M(\zeta)$ is bounded in $L^2(\Omega)$ and in $H^1(\Omega)$, uniformly for $\zeta\in\partial\mathbb{D}\backslash\{\pm1\}$; see Lemma 3.1. By Parseval's identity, the boundedness of $M(\zeta)$ implies that the operator $\mathcal{F}^{-1}M(\zeta)\mathcal{F}$ is bounded on both $\ell^2(L^2(\Omega))$ and $\ell^2(H^1(\Omega))$, i.e.,

$$\tau\sum_{n=1}^{\infty}\|\Delta v^{n-\frac{1}{2}}\|_{L^2}^2\leqslant C\tau\sum_{n=1}^{m}\|f^n\|_{L^2}^2\quad\text{and}\quad\tau\sum_{n=1}^{\infty}\|\Delta v^{n-\frac{1}{2}}\|_{H^1}^2\leqslant C\tau\sum_{n=1}^{m}\|f^n\|_{H^1}^2.$$

Since $\|v^{n-\frac{1}{2}}\|_{H^{s+2}}\leqslant C\|\Delta v^{n-\frac{1}{2}}\|_{H^s}$ for $s=0,1$ (cf. [13, Theorem 5, Chapter 6]), these two inequalities imply

$$\tau\sum_{n=1}^{\infty}\|v^{n-\frac{1}{2}}\|_{H^2}^2\leqslant C\tau\sum_{n=1}^{m}\|f^n\|_{L^2}^2\quad\text{and}\quad\tau\sum_{n=1}^{\infty}\|v^{n-\frac{1}{2}}\|_{H^3}^2\leqslant C\tau\sum_{n=1}^{m}\|f^n\|_{H^1}^2.$$

This proves the estimate for the second term of (3.6).

Testing (3.5) by $-\Delta v^{n-\frac{1}{2}}$ immediately yields

$$\begin{aligned}\frac{\|\nabla v^n\|_{L^2}^2-\|\nabla v^{n-1}\|_{L^2}^2}{2\tau}+\|\Delta v^{n-\frac{1}{2}}\|_{L^2}^2&=(f^n,\Delta v^{n-\frac{1}{2}})\\&\leqslant\|f^n\|_{L^2}\|\Delta v^{n-\frac{1}{2}}\|_{L^2}\\&\leqslant\frac{1}{2}\|f^n\|_{L^2}^2+\frac{1}{2}\|\Delta v^{n-\frac{1}{2}}\|_{L^2}^2,\end{aligned}$$

which implies

$$\max_{1\leqslant n\leqslant m}\|v^n\|_{H^1}^2\leqslant C\tau\sum_{n=1}^{m}\|f^n\|_{L^2}^2. \tag{3.9}$$

This proves the estimate for the first term of (3.6) in the case $s=0$.

If $f^n\in H_0^1(\Omega)$, then $\Delta v^{n-\frac{1}{2}}\in H_0^1(\Omega)$, and testing (3.5) by $\Delta^2 v^{n-\frac{1}{2}}$ yields

$$\begin{aligned}\frac{\|\Delta v^n\|_{L^2}^2-\|\Delta v^{n-1}\|_{L^2}^2}{2\tau}+\|\nabla\Delta v^{n-\frac{1}{2}}\|_{L^2}^2&=(\nabla f^n,\nabla\Delta v^{n-\frac{1}{2}})\\&\leqslant\|f^n\|_{H^1}\|\nabla\Delta v^{n-\frac{1}{2}}\|_{L^2}\\&\leqslant\frac{1}{2}\|f^n\|_{H^1}^2+\frac{1}{2}\|\nabla\Delta v^{n-\frac{1}{2}}\|_{L^2}^2,\end{aligned}$$

which implies

$$\max_{1\leqslant n\leqslant m}\|\Delta v^n\|_{L^2}^2\leqslant C\tau\sum_{n=1}^{m}\|f^n\|_{H^1}^2;$$

therefore,

$$\max_{1\leqslant n\leqslant m}\|v^n\|_{H^2}^2\leqslant C\tau\sum_{n=1}^m\|f^n\|_{H^1}^2. \tag{3.10}$$

If $f^n\in H^1(\Omega)$, then we choose $f_j^n=e^{j^{-1}\Delta}f^n$. Similarly as (3.3), f_j^n has the following properties:

$$f_j^n\in H_0^1(\Omega),\quad f_j^n\to f^n\ \text{ in } L^2(\Omega)\quad\text{and}\quad \|f_j^n\|_{H^1}\leqslant C\|f^n\|_{H^1}\ \text{ as }\ j\to\infty.$$

Then, the corresponding v_j^n satisfies

$$\max_{1\leqslant n\leqslant m}\|v_j^n\|_{H^2}^2\leqslant C\tau\sum_{n=1}^m\|f_j^n\|_{H^1}^2\leqslant C\tau\sum_{n=1}^m\|f^n\|_{H^1}^2. \tag{3.11}$$

Since $f_j^n\to f^n$ in $L^2(\Omega)$ as $j\to\infty$, (3.9) implies that $v_j^n\to v^n$ in $H^1(\Omega)$. This together with (3.11) implies that v_j^n is bounded and convergent to v^n weakly in $H^2(\Omega)$. Letting $j\to\infty$, we obtain (3.10) for $f^n\in H^1(\Omega)$. This proves the estimate for the first term of (3.6) in the case $s=1$. ∎

3.2. Consistency of the method. We abbreviate $u(\cdot,t)$ by $u(t)$ and $H(\cdot,t)$ by $H(t)$. Furthermore, we denote

$$H_\star^n=H(t_n),\quad H_\star^{n-\frac12}=\frac12(H_\star^n+H_\star^{n-1}), \tag{3.12}$$

$$u_\star^n=u(t_n),\qquad u_\star^{n-\frac12}=\frac12(u_\star^n+u_\star^{n-1}), \tag{3.13}$$

and $t_{n-\frac12}:=(t_n+t_{n-1})/2$.

Let η_H^n and η_u^n be the consistency errors of the modified Crank–Nicolson method (2.6), defined by

(3.14)
$$\begin{cases}\bar\partial_\tau H_\star^n+\nabla\cdot\big(H_\star^{n-\frac12}u_\star^{n-\frac12}\big)=\eta_H^n,\\[2mm] \bar\partial_\tau u_\star^n+\nabla\Big(\dfrac14\big(|u_\star^n|^2+|u_\star^{n-1}|^2\big)+g(H_\star^{n-\frac12}-H_b)\Big)\\ \qquad+(\nabla\times u_\star^{n-\frac12}+f^{n-\frac12})\hat k\times u_\star^{n-\frac12}-\mathcal{G}(H_\star^{n-\frac12},u_\star^{n-\frac12})=\eta_u^n,\quad n=1,\dots,N.\end{cases}$$

It is straightforward to show that

$$\max_{1\leqslant n\leqslant N}\big(\|\eta_H^n\|_{H^2}+\|\eta_u^n\|_{H^2}\big)\leqslant C\tau^2, \tag{3.15}$$

provided that the solution (H,u) is sufficiently smooth.

3.3. Existence of discrete solutions. Let $e_u^n:=u_\star^n-u^n$ and $e_H^n:=H_\star^n-H^n$ denote the errors of the modified Crank–Nicolson method (2.6). Subtracting (2.6)

from the consistency equations (3.14), we obtain the error equations

(3.16)
$$\begin{cases} \bar{\partial}_\tau e_H^n + \nabla\cdot\big(e_H^{n-\frac12}u^{n-\frac12} + H_\star^{n-\frac12}e_u^{n-\frac12}\big) = \eta_H^n, \\ \bar{\partial}_\tau e_u^n - \dfrac{\mu}{H^{n-\frac12}}\nabla\cdot(H^{n-\frac12}\nabla e_u^{n-\frac12}) \\ \quad - \dfrac{\mu}{H^{n-\frac12}}\nabla\cdot(e_H^{n-\frac12}\nabla u_\star^{n-\frac12}) + \dfrac{\mu}{H^{n-\frac12}H_\star^{n-\frac12}}e_H^{n-\frac12}\nabla\cdot(H_\star^{n-\frac12}\nabla u_\star^{n-\frac12}) \\ \quad + c_f\left(\dfrac{|u_\star^{n-\frac12}|u_\star^{n-\frac12}}{H_\star^{n-\frac12}} - \dfrac{|u^{n-\frac12}|u^{n-\frac12}}{H^{n-\frac12}}\right) \\ \quad + \dfrac14\nabla\big(|u_\star^n|^2 - |u^n|^2 + |u_\star^{n-1}|^2 - |u^{n-1}|^2\big) + g\nabla e_H^{n-\frac12} \\ \quad + (\nabla\times u_\star^{n-\frac12})\hat{k}\times u_\star^{n-\frac12} - (\nabla\times u^{n-\frac12})\hat{k}\times u^{n-\frac12} + f^{n-\frac12}\hat{k}\times e_u^{n-\frac12} = \eta_u^n. \end{cases}$$

Remark 3.1. Obviously, if $(e_H^n, e_u^n)\in H^2(\Omega)\times[H^3(\Omega)\cap H_0^1(\Omega)]^2$ is a solution of (3.16) with $H^n = H_\star^n - e_H^n$ and $u^n = u_\star^n - e_u^n$, then (H^n, u^n) is a solution of (2.6).

To prove the existence of a solution (e_H^n, e_u^n) to (3.16) with $H^n = H_\star^n - e_H^n$ and $u^n = u_\star^n - e_u^n$, we first prove the existence of a solution for a regularized approximating problem (for which the proof of existence is easier, as explained in Remark 3.2). To this end, we let $E: L^1(\Omega)\to L^1(\mathbb{R}^2)$ be a linear extension operator that is bounded also from $W^{k,p}(\Omega)$ to $W^{k,p}(\mathbb{R}^2)$ for all $k\geqslant 0$ and $1\leqslant p\leqslant\infty$. Such an extension operator indeed exists; see [23, Theorem 5, p. 181]. Then, we let σ_ε be a standard smooth mollifier in $\mathbb{R}^2$ and define

$$\sigma_\varepsilon\star\varphi^{n-\frac12} := \sigma_\varepsilon\star E\varphi^{n-\frac12}.$$

The mollified function $\sigma_\varepsilon\star\varphi^{n-\frac12}$ is smooth and satisfies

(3.17)
$$\|\sigma_\varepsilon\star\varphi^{n-\frac12}\|_{H^m(\mathbb{R}^2)}\leqslant C_0\varepsilon^{-(m-k)}\|\varphi^{n-\frac12}\|_{H^k(\Omega)}\quad\forall\,\varphi^{n-\frac12}\in W^{k,p}(\Omega),\ \ 0\leqslant k\leqslant m.$$

We consider the following regularized problem an approximation to (3.16):

(3.18)
$$\begin{cases} \bar{\partial}_\tau e_H^n + \nabla\cdot\big(e_H^{n-\frac12}u^{n-\frac12} + H_\star^{n-\frac12}e_u^{n-\frac12}\big) = \eta_H^n, \\ \bar{\partial}_\tau e_u^n - \dfrac{\mu}{H_\varepsilon^{n-\frac12}}\nabla\cdot(H_\varepsilon^{n-\frac12}\nabla e_u^{n-\frac12}) \\ \quad - \dfrac{\mu}{H_\varepsilon^{n-\frac12}}\nabla\cdot(\sigma_\varepsilon\star e_H^{n-\frac12}\nabla u_\star^{n-\frac12}) + \dfrac{\mu}{H_\varepsilon^{n-\frac12}H_\star^{n-\frac12}}\sigma_\varepsilon\star e_H^{n-\frac12}\nabla\cdot(H_\star^{n-\frac12}\nabla u_\star^{n-\frac12}) \\ \quad + c_f\left(\dfrac{|u_\star^{n-\frac12}|u_\star^{n-\frac12}}{H_\star^{n-\frac12}} - \dfrac{|u^{n-\frac12}|u^{n-\frac12}}{H_\varepsilon^{n-\frac12}}\right) \\ \quad + \dfrac14\nabla\big((u_\star^n+u^n)\cdot(u_\star^n-u^n) + (u_\star^{n-1}+u^{n-1})\cdot(u_\star^{n-1}-u^{n-1})\big) + g\nabla(\sigma_\varepsilon\star e_H^{n-\frac12}) \\ \quad + \nabla\times(u_\star^{n-\frac12}-u^{n-\frac12})\hat{k}\times u^{n-\frac12} - (\nabla\times u_\star^{n-\frac12})\hat{k}\times(u_\star^{n-\frac12}-u^{n-\frac12}) \\ \quad + f^{n-\frac12}\hat{k}\times e_u^{n-\frac12} = \eta_u^n \end{cases}$$

with

(3.19)
$$H_\varepsilon^n = H_\star^n - \sigma_\varepsilon\star e_H^n \quad\text{and}\quad u^n = u_\star^n - e_u^n.$$

Remark 3.2. If (3.18) has a solution (e_H^n, e_u^n) bounded in $H^2(\Omega) \times [H^3(\Omega) \cap H_0^1(\Omega)]^2$ uniformly for $1 \leqslant n \leqslant N$ and $\varepsilon \in (0,1)$, then there exists a sequence $\varepsilon_k \to 0$ such that the corresponding sequence of solutions of (3.18) converges strongly in $H^1(\Omega) \times [H^2(\Omega) \cap H_0^1(\Omega)]^2$ and weakly in $H^2(\Omega) \times [H^3(\Omega) \cap H_0^1(\Omega)]^2$ to a solution of (3.16). The rigorous proof of this "pass to limit" is routine.

To prove the existence of solutions for the regularized problem (3.18), we use Schaefer's fixed point theorem (cf. [13, Chapter 9.2, Theorem 4]).

Schaefer's fixed point theorem: Let Y be a Banach space and let $M : Y \to Y$ be a continuous and compact map (possibly nonlinear). If the set

$$\big\{\phi \in Y : \ \phi = \theta M(\phi) \ \text{ for some } \ \theta \in [0,1]\big\} \tag{3.20}$$

is bounded in Y, then the map M has a fixed point.

Construction of the map M: Let $X = H^1(\Omega) \times \big(H^3(\Omega) \cap H_0^1(\Omega)\big)^2$ and consider the space $(X^N, \|\cdot\|_{\ell^\infty(X)})$ of sequences $(\phi^n, \varphi^n)_{n=1}^N$ with $(\phi^n, \varphi^n) \in X$, endowed with the following norm:

$$\|(\phi^n, \varphi^n)_{n=1}^N\|_{\ell^\infty(X)} := \max_{1 \leqslant n \leqslant N} \|(\phi^n, \varphi^n)\|_X.$$

For any sequence $(\phi^n, \varphi^n)_{n=1}^N \in X^N$ we define

$$\rho_{\phi,\varepsilon}^H := \min\left(\frac{H_{\min}}{2} \frac{1}{\max\limits_{1 \leqslant n \leqslant N} (\|\sigma_\varepsilon \star \phi^n\|_{H^2} + \|\sigma_\varepsilon \star \phi^n\|_{L^\infty})}, 1\right), \tag{3.21}$$

$$\rho_\varphi^u := \min\left(\frac{1}{\max\limits_{1 \leqslant n \leqslant N} (\|\varphi^n\|_{H^3} + \|\nabla \cdot \varphi^n\|_{L^\infty})}, 1\right), \tag{3.22}$$

$$H_{\phi,\varepsilon}^n = H_\star^n - \rho_{\phi,\varepsilon}^H \, \sigma_\varepsilon \star \phi^n \quad \text{and} \quad u_\varphi^n = u_\star^n - \rho_\varphi^u \varphi^n. \tag{3.23}$$

Then $\|\rho_{\phi,\varepsilon}^H \, \sigma_\varepsilon \star \phi^n\|_{L^\infty} \leqslant \frac{H_{\min}}{2}$ and therefore

$$H_{\phi,\varepsilon}^n \geqslant \frac{H_{\min}}{2}.$$

For any fixed $\varepsilon > 0$, the quantities $\rho_{\phi,\varepsilon}^H$ and ρ_φ^u depend continuously on $(\phi^n, \varphi^n)_{n=1}^N \in X^N$, and

$$\|\rho_{\phi,\varepsilon}^H \, \sigma_\varepsilon \star \phi^n\|_{H^2} \leqslant \frac{H_{\min}}{2} \quad \text{and} \quad \|\rho_\varphi^u \varphi^n\|_{H^3} \leqslant 1, \tag{3.24}$$

(3.25)
$$\|H_{\phi,\varepsilon}^n\|_{H^2} \leqslant \|H_\star^n\|_{H^2} + \frac{H_{\min}}{2} \ \text{ and } \ \|u_\varphi^n\|_{H^3} + \|\nabla \cdot u_\varphi^n\|_{L^\infty} \leqslant \|u_\star^n\|_{H^3} + \|\nabla \cdot u_\star^n\|_{L^\infty} + 1.$$

For any given $(\phi^n,\varphi^n)_{n=1}^N \in X^N$, we define $(e_H^n,e_u^n)_{n=1}^N \in X^N$ to be the solution of the following linear problem:

(3.26)

$$\begin{aligned}
&\bar\partial_\tau e_H^n + \nabla\cdot\big(e_H^{n-\frac12}u_\varphi^{n-\frac12} + H_\star^{n-\frac12}\varphi^{n-\frac12}\big) = \eta_H^n,\\
&\bar\partial_\tau e_u^n - \frac{\mu}{H_{\phi,\varepsilon}^{n-\frac12}}\nabla\cdot(H_{\phi,\varepsilon}^{n-\frac12}\nabla e_u^{n-\frac12})\\
&\quad - \frac{\mu}{H_{\phi,\varepsilon}^{n-\frac12}}\nabla\cdot(\sigma_\varepsilon\star\phi^{n-\frac12}\nabla u_\star^{n-\frac12}) + \frac{\mu}{H_{\phi,\varepsilon}^{n-\frac12}H_\star^{n-\frac12}}\sigma_\varepsilon\star\phi^{n-\frac12}\nabla\cdot(H_\star^{n-\frac12}\nabla u_\star^{n-\frac12})\\
&\quad + c_f\left(\frac{|u_\star^{n-\frac12}|u_\star^{n-\frac12}}{H_\star^{n-\frac12}} - \frac{|u_\varphi^{n-\frac12}|u_\varphi^{n-\frac12}}{H_{\phi,\varepsilon}^{n-\frac12}}\right)\\
&\quad + \frac14\nabla\big((u_\star^n+u_\varphi^n)\varphi^n + (u_\star^{n-1}+u_\varphi^{n-1})\varphi^{n-1}\big) + g\nabla\sigma_\varepsilon\star\phi^{n-\frac12}
\end{aligned}$$

(3.27)

$$+ (\nabla\times\varphi^{n-\frac12})\hat k\times u_\varphi^{n-\frac12} + (\nabla\times u_\star^{n-\frac12})\hat k\times\varphi^{n-\frac12} + f^{n-\frac12}\hat k\times\varphi^{n-\frac12} = \eta_u^n$$

with starting values $e_H^0 = e_u^0 = 0$. The map from $(\phi^n,\varphi^n)_{n=1}^N$ to $(e_H^n,e_u^n)_{n=1}^N$ is denoted by M.

LEMMA 3.3. *Let τ be sufficiently small (independent of ε),*

$$\tau \leqslant \frac{1}{\|u\|_{L^\infty(0,T;H^3)} + \|\nabla\cdot u\|_{L^\infty(0,T;L^\infty)} + 1}. \tag{3.28}$$

Then, for any given $\eta_H^n,\eta_u^n \in H^2(\Omega)$, given ε, and given $(\phi^n,\varphi^n)_{n=1}^N \in X^N$, the system (3.26)–(3.27) has a unique solution $(e_H^n,e_u^n) \in H^2(\Omega)\times[H^4(\Omega)\cap H_0^1(\Omega)]^2$, $n=1,\dots,N$. Moreover, the map $M: X^N \to X^N$ is well defined, continuous, and compact.

Proof. For given $e_H^{n-1} \in H^2(\Omega)$, (3.26) can be written as

$$\frac{2}{\tau}e_H^{n-\frac12} + u_\varphi^{n-\frac12}\cdot\nabla e_H^{n-\frac12} + (\nabla\cdot u_\varphi^{n-\frac12})e_H^{n-\frac12} = g_H^n \tag{3.29}$$

with

$$g_H^n = \eta_H^n - \nabla\cdot(H_\star^{n-\frac12}\varphi^{n-\frac12}) + \frac{2}{\tau}e_H^{n-1} \in H^2(\Omega).$$

The linear hyperbolic equation (3.29) has a unique solution $e_H^{n-\frac12} \in H^2(\Omega)$ and satisfies the following estimate (see the appendix):

$$\|e_H^{n-\frac12}\|_{H^2} \leqslant C\|g_H^n\|_{H^2}.$$

This implies that (3.26) has a unique solution $e_H^n \in H^2(\Omega)$, $n=1,\dots,N$.

Similarly, for given $e_u^{n-1} \in H^3(\Omega)\cap H_0^1(\Omega)$, (3.27) can be written as a linear elliptic equation

$$\frac{2}{\tau}e_u^{n-\frac12} - \frac{\mu}{H_{\phi,\varepsilon}^{n-\frac12}}\nabla\cdot(H_{\phi,\varepsilon}^{n-\frac12}\nabla e_u^{n-\frac12}) = -g_u^n \tag{3.30}$$

with $g_u^n \in H^2(\Omega)$ given by

$$
\begin{aligned}
g_u^n = &-\frac{\mu}{H_{\phi,\varepsilon}^{n-\frac12}}\nabla\cdot(\sigma_\varepsilon\star\phi^{n-\frac12}\nabla u_\star^{n-\frac12}) + \frac{\mu}{H_{\phi,\varepsilon}^{n-\frac12}H_\star^{n-\frac12}}\sigma_\varepsilon\star\phi^{n-\frac12}\nabla\cdot(H_\star^{n-\frac12}\nabla u_\star^{n-\frac12})\\
&+ c_f\Big(\frac{|u_\star^{n-\frac12}|u_\star^{n-\frac12}}{H_\star^{n-\frac12}} - \frac{|u_\varphi^{n-\frac12}|u_\varphi^{n-\frac12}}{H_{\phi,\varepsilon}^{n-\frac12}}\Big)\\
&+ \frac14\nabla\big((u_\star^n+u_\varphi^n)\varphi^n + (u_\star^{n-1}+u_\varphi^{n-1})\varphi^{n-1}\big) + g\nabla(\sigma_\varepsilon\star\phi^{n-\frac12})\\
&+ (\nabla\times\varphi^{n-\frac12})\hat{k}\times u_\varphi^{n-\frac12} + (\nabla\times u_\star^{n-\frac12})\hat{k}\times\varphi^{n-\frac12} + f^{n-\frac12}\hat{k}\times\varphi^{n-\frac12} - \eta_u^n + \frac{2}{\tau}e_u^{n-1}.
\end{aligned}
$$

It is well known that the elliptic equation (3.30) has a unique solution $e_u^{n-\frac12} \in [H^4(\Omega)\cap H_0^1(\Omega)]^2$, satisfying the following estimate:

$$\|e_u^{n-\frac12}\|_{H^4} \leqslant C\|g_u^n\|_{H^2}.$$

Therefore, the map $M : X^N \to X^N$ is well defined. Furthermore, if $(\phi^n,\varphi^n)_{n=1}^N$ is bounded in X^N, then $(e_H^n, e_u^n)_{n=1}^N$ is bounded in $(H^2(\Omega)\times[H^4(\Omega)\cap H_0^1(\Omega)]^2)^N$, which is compactly embedded into X^N. Thus the map $M : X^N \to X^N$ is compact.

The continuity of the map can be proved in the standard way, and the proof is omitted. ∎

Remark 3.3. For any fixed $\varepsilon > 0$, the mollified functions $H_{\phi,\varepsilon}^{n-\frac12}$ and $\sigma_\varepsilon\star\phi^{n-\frac12}$ are sufficiently smooth. As a result, the solution of (3.27) is in $[H^4(\Omega)\cap H_0^1(\Omega)]^2$, compactly embedded into $[H^3(\Omega)\cap H_0^1(\Omega)]^2$. Hence, the regularization using mollifiers guarantees that the map $M : X^N \to X^N$ is compact. Without the regularization, the map $M : X^N \to X^N$ is well defined and continuous, but it is difficult to prove its compactness.

In Lemma 3.3, we proved that the first condition of Schaefer's fixed point theorem is satisfied, i.e., that the proposed map $M : X^N \to X^N$ is well defined, continuous, and compact. In the following lemma, we will prove that the second condition of Schaefer's fixed point theorem is also satisfied, i.e., the set defined in (3.20) is bounded in $Y = X^N$.

LEMMA 3.4. *There exists a positive constant τ_0 such that the following result holds for $\tau \leqslant \tau_0$: if $(\phi^n,\varphi^n)_{n=1}^N$ satisfies*

$$(\phi^n,\varphi^n)_{n=1}^N = \theta M[(\phi^n,\varphi^n)_{n=1}^N] \quad \text{for some } \theta\in[0,1], \tag{3.31}$$

then $(\phi^n,\varphi^n)_{n=1}^N$ is bounded in $(H^2(\Omega)\times[H^3(\Omega)\cap H_0^1(\Omega)]^2)^N \hookrightarrow X^N$ uniformly for $\theta\in[0,1]$ and $\varepsilon\in(0,1)$. More precisely, $\rho_{\phi,\varepsilon}^H = \rho_\varphi^u = 1$ and

$$\max_{1\leqslant n\leqslant N}(\|\phi^n\|_{H^2} + \|\phi^n\|_{L^\infty}) \leqslant \frac{H_{\min}}{2} \quad \text{and} \quad \max_{1\leqslant n\leqslant N}\|\varphi^n\|_{H^3} \leqslant 1. \tag{3.32}$$

The proof of Lemma 3.4 is presented in the next two subsections together with error estimates for the discrete solutions.

Lemma 3.4 and Schaefer's fixed point theorem imply that the map M has at least one fixed point, which we denote by $(e_H^n, e_u^n)_{n=1}^N$. In the case $\rho_{\phi,\varepsilon}^H = \rho_\varphi^u = 1$, the fixed point $(e_H^n, e_u^n)_{n=1}^N$ of M satisfies

$$H_{\phi,\varepsilon}^n = H_\star^n - \sigma_\varepsilon\star e_H^n = H_\varepsilon^n \quad\text{and}\quad u_\varphi^n = u_\star^n - e_u^n = u^n,$$

where H_ε^n and u^n are defined in (3.19). Therefore, (3.26)–(3.27) reduces to (3.18). Hence, Lemma 3.4 implies the existence of a solution to the regularized problem (3.18).

Lemma 3.4 implies that the fixed point (e_H^n, e_u^n) is bounded in $(H^2(\Omega)\times[H^3(\Omega)\cap H_0^1(\Omega)]^2)^N$ uniformly for $\varepsilon \in (0,1)$. This implies that there exists a subsequence $\varepsilon_k \to 0$ such that the corresponding solutions of the regularized problem converge to a solution of (3.16) with

$$H^n = H_\star^n - e_H^n \quad \text{and} \quad u^n = u_\star^n - e_u^n.$$

This proves the existence of a solution for the proposed method (2.6), as explained in Remark 3.1, while (3.32) implies that the solution is in $B_{H,u}^n$, which is defined in (2.10).

Proof of Lemma 3.4. If $(\phi^n, \varphi^n)_{n=1}^N \in X^N$ satisfies (3.31), then $(e_H^n, e_u^n)_{n=1}^N = M[(\phi^n, \varphi^n)_{n=1}^N]$ is the solution of (3.26)–(3.27) and

$$\phi^n = \theta e_H^n \quad \text{and} \quad \varphi^n = \theta e_u^n. \tag{3.33}$$

In view of this, we assume that $(e_H^n, e_u^n)_{n=1}^N$ is the solution of (3.26)–(3.27) for some $\varepsilon \in [0,1]$ with (ϕ^n, φ^n) given by (3.33). Then, we estimate e_H^n and e_u^n separately in the next two subsections.

Remark 3.4. Although the proof of Lemma 3.4 only needs the case $\varepsilon \in (0,1)$, the estimates obtained in the next two subsections include the case $\varepsilon = 0$ (assuming that there exists a solution in this case).

3.4. Estimation of e_H^n. We rewrite (3.26) as

$$\bar\partial_\tau e_H^n + u_\varphi^{n-\frac12}\cdot\nabla e_H^{n-\frac12} = \eta_H^n + I_1^n + I_2^n + I_3^n \tag{3.34}$$

with

$$I_1^n = -e_H^{n-\frac12}\nabla\cdot u_\varphi^{n-\frac12}, \quad I_2^n = -\theta e_u^{n-\frac12}\cdot\nabla H_\star^{n-\frac12}, \quad I_3^n = -\theta H_\star^{n-\frac12}\nabla\cdot e_u^{n-\frac12}. \tag{3.35}$$

Let ∂_i be the partial differentiation operator with respect to x_i and let $\partial_{ij} = \partial_i\partial_j, i,j = 1,2$. Application of the differential operator $\partial_i\partial_j$ to (3.34) yields

$$\bar\partial_\tau \partial_{ij} e_H^n + u_\varphi^{n-\frac12}\cdot\nabla\partial_{ij} e_H^{n-\frac12} = \partial_{ij}\eta_H^n + \partial_{ij} I_1^n + \partial_{ij} I_2^n + \partial_{ij} I_3^n + I_4^n \tag{3.36}$$

with

$$I_4^n = -\partial_i u_\varphi^{n-\frac12}\cdot\nabla\partial_j e_H^{n-\frac12} - \partial_j u_\varphi^{n-\frac12}\cdot\nabla\partial_i e_H^{n-\frac12} - \partial_{ij} u_\varphi^{n-\frac12}\cdot\nabla e_H^{n-\frac12}. \tag{3.37}$$

Then, testing (3.36) by $2\partial_{ij} e_H^{n-\frac12}$, we obtain

$$\begin{aligned}\bar\partial_\tau\|\partial_{ij} e_H^n\|^2 &= (\nabla\cdot u_\varphi^{n-\frac12}, 2|\partial_{ij} e_H^{n-\frac12}|^2) + (\partial_{ij}\eta_H^n, 2\partial_{ij} e_H^{n-\frac12}) \\ &\quad + \sum_{\ell=1}^3 (\partial_{ij} I_\ell^n, 2\partial_{ij} e_H^{n-\frac12}) + (I_4^n, 2\partial_{ij} e_H^{n-\frac12}).\end{aligned} \tag{3.38}$$

By using the expressions in (3.35) and (3.37), we have

$$\begin{aligned}|(\partial_{ij}I_1^n, 2\partial_{ij}e_H^{n-\frac12})| &\leqslant |(\partial_{ij}e_H^{n-\frac12}\nabla\cdot u_\varphi^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})| + |(\partial_i e_H^{n-\frac12}\nabla\cdot\partial_j u_\varphi^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\quad + |(\partial_j e_H^{n-\frac12}\nabla\cdot\partial_i u_\varphi^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\quad + |(e_H^{n-\frac12}\nabla\cdot\partial_{ij} u_\varphi^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\leqslant C\|u_\varphi^{n-\frac12}\|_{H^3}\|e_H^{n-\frac12}\|_{H^2}^2,\end{aligned}$$

$$\begin{aligned}|(\partial_{ij}I_2^n, 2\partial_{ij}e_H^{n-\frac12})| &\leqslant |(\partial_{ij}e_u^{n-\frac12}\cdot\nabla H_\star^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})| + |(\partial_i e_u^{n-\frac12}\cdot\nabla\partial_j H_\star^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\quad + |(\partial_j e_u^{n-\frac12}\cdot\nabla\partial_i H_\star^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\quad + |(e_u^{n-\frac12}\cdot\nabla\partial_{ij} H_\star^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\leqslant C\|e_u^{n-\frac12}\|_{H^2}\|e_H^{n-\frac12}\|_{H^2},\end{aligned}$$

$$\begin{aligned}|(\partial_{ij}I_3^n, 2\partial_{ij}e_H^{n-\frac12})| &\leqslant |(\partial_{ij}H_\star^{n-\frac12}\nabla\cdot e_u^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})| + |(\partial_i H_\star^{n-\frac12}\nabla\cdot\partial_j e_u^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\quad + |(\partial_j H_\star^{n-\frac12}\nabla\cdot\partial_i e_u^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\quad + |(H_\star^{n-\frac12}\nabla\cdot\partial_{ij} e_u^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\leqslant C\|e_u^{n-\frac12}\|_{H^3}\|e_H^{n-\frac12}\|_{H^2},\end{aligned}$$

$$\begin{aligned}|(I_4^n, 2\partial_{ij}e_H^{n-\frac12})| &\leqslant |(\partial_i u_\varphi^{n-\frac12}\cdot\nabla\partial_j e_H^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\quad + |(\partial_j u_\varphi^{n-\frac12}\cdot\nabla\partial_i e_H^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12}) + |(\partial_{ij} u_\varphi^{n-\frac12}\cdot\nabla e_H^{n-\frac12}, 2\partial_{ij}e_H^{n-\frac12})|\\ &\leqslant C\|u_\varphi^{n-\frac12}\|_{H^3}\|e_H^{n-\frac12}\|_{H^2}^2.\end{aligned}$$

Substituting these estimates into (3.38) and using estimate (3.25), we obtain

$$\bar\partial_\tau\|\partial_{ij}e_H^n\|^2 \leqslant C\|e_H^{n-\frac12}\|_{H^2}^2 + C\|e_u^{n-\frac12}\|_{H^3}^2 + C\|\partial_{ij}\eta_H^n\|^2, \quad i,j=1,2.$$

Similar estimates can also be obtained for $\partial_j e_H^n$ and e_H^n,

$$\bar\partial_\tau\|\partial_j e_H^n\|^2 \leqslant C\|e_H^{n-\frac12}\|_{H^1}^2 + C\|e_u^{n-\frac12}\|_{H^3}^2 + C\|\partial_j\eta_H^n\|^2, \quad j=1,2,$$

and

$$\bar\partial_\tau\|e_H^n\|^2 \leqslant C\|e_H^{n-\frac12}\|^2 + C\|e_u^{n-\frac12}\|_{H^3}^2 + C\|\eta_H^n\|^2.$$

Summing up these estimates yields

$$\bar\partial_\tau\|e_H^n\|_{H^2}^2 \leqslant C\|e_H^{n-\frac12}\|_{H^2}^2 + C\|e_u^{n-\frac12}\|_{H^3}^2 + C\|\eta_H^n\|_{H^2}^2. \tag{3.39}$$

3.5. Estimation of e_u^n. We rewrite (3.27) as

$$\bar\partial_\tau e_u^n - \mu\Delta e_u^{n-\frac12} = \eta_u^n + \sum_{i=1}^8 J_i^n \tag{3.40}$$

with

$$
\begin{aligned}
J_1^n &= \frac{\mu}{H_{\phi,\varepsilon}^{n-\frac{1}{2}}}\nabla H_{\phi,\varepsilon}^{n-\frac{1}{2}}\cdot\nabla e_u^{n-\frac{1}{2}},\\
J_2^n &= \frac{\theta\mu}{H_{\phi,\varepsilon}^{n-\frac{1}{2}}}\nabla\cdot(\sigma_\varepsilon\star e_H^{n-\frac{1}{2}}\nabla u_\star^{n-\frac{1}{2}}),\\
J_3^n &= -\frac{\theta\mu}{H_{\phi,\varepsilon}^{n-\frac{1}{2}}H_\star^{n-\frac{1}{2}}}\sigma_\varepsilon\star e_H^{n-\frac{1}{2}}\nabla\cdot(H_\star^{n-\frac{1}{2}}\nabla u_\star^{n-\frac{1}{2}}),\\
J_4^n &= -c_f\left(\frac{|u_\star^{n-\frac{1}{2}}|u_\star^{n-\frac{1}{2}}}{H_\star^{n-\frac{1}{2}}}-\frac{|u_\varphi^{n-\frac{1}{2}}|u_\varphi^{n-\frac{1}{2}}}{H_{\phi,\varepsilon}^{n-\frac{1}{2}}}\right)\\
&= c_f\theta\frac{\rho_{\phi,\varepsilon}^H\sigma_\varepsilon\star e_H^{n-\frac{1}{2}}}{H_\star^{n-\frac{1}{2}}H_{\phi,\varepsilon}^{n-\frac{1}{2}}}|u_\star^{n-\frac{1}{2}}|u_\star^{n-\frac{1}{2}}-c_f\frac{|u_\star^{n-\frac{1}{2}}|u_\star^{n-\frac{1}{2}}-|u_\varphi^{n-\frac{1}{2}}|u_\varphi^{n-\frac{1}{2}}}{H_{\phi,\varepsilon}^{n-\frac{1}{2}}},\\
J_5^n &= -\frac{\theta}{4}\nabla\big((u_\star^n+u_\varphi^n)e_u^n+(u_\star^{n-1}+u_\varphi^{n-1})e_u^{n-1}\big),\\
J_6^n &= -\theta g\nabla(\sigma_\varepsilon\star e_H^{n-\frac{1}{2}}),\\
J_7^n &= -\theta(\nabla\times e_u^{n-\frac{1}{2}})\hat{k}\times u_\varphi^{n-\frac{1}{2}}-\theta(\nabla\times u_\star^{n-\frac{1}{2}})\hat{k}\times e_u^{n-\frac{1}{2}},\\
J_8^n &= -\theta f^{n-\frac{1}{2}}\hat{k}\times e_u^{n-\frac{1}{2}},
\end{aligned}
$$

where we have substituted $\phi^n=\theta e_H^n$ and $\varphi^n=\theta e_u^n$ into the expressions above. Notice that

$$
\sum_{n=1}^m\|\bar{\partial}_\tau e_u^n\|_{H^1}^2=\sum_{n=1}^m\|\mu\Delta e_u^{n-\frac{1}{2}}+\eta_u^n+\sum_{i=1}^7 J_i^n\|_{H^1}^2\leqslant C\sum_{n=1}^m\big(\|e_u^{n-\frac{1}{2}}\|_{H^3}^2+\|\eta_u^n+\sum_{i=1}^7 J_i^n\|_{H^1}^2\big).
$$

Therefore, applying Lemma 3.2 with $s=1$ to (3.40), we obtain

(3.41)
$$
\max_{1\leqslant n\leqslant m}\|e_u^n\|_{H^2}^2+\tau\sum_{n=1}^m\big(\|\bar{\partial}_\tau e_u^n\|_{H^1}^2+\|e_u^{n-\frac{1}{2}}\|_{H^3}^2\big)\leqslant C\tau\sum_{n=1}^m\Big(\|\eta_u^n\|_{H^1}^2+\sum_{i=1}^7\|J_i^n\|_{H^1}^2\Big).
$$

By using estimate (3.25), it is straightforward to verify that

$$
\begin{aligned}
\|J_1^n\|_{H^1}&\leqslant C(\|H_{\phi,\varepsilon}^{n-\frac{1}{2}}\|_{H^2}\|\nabla e_u^{n-\frac{1}{2}}\|_{L^\infty}+\|H_{\phi,\varepsilon}^{n-\frac{1}{2}}\|_{W^{1,4}}\|e_u^{n-\frac{1}{2}}\|_{W^{2,4}})\\
&\leqslant C\|e_u^{n-\frac{1}{2}}\|_{H^{\frac{5}{2}}}\qquad\qquad\text{(Sobolev embedding inequality)}\\
&\leqslant C_\delta\|e_u^{n-\frac{1}{2}}\|_{H^2}+\delta\|e_u^{n-\frac{1}{2}}\|_{H^3}\quad\text{(interpolation inequality)},
\end{aligned}
$$

where $\delta\in(0,1)$ can be arbitrarily small at the expense of enlarging the constant C_δ.

Similarly, we have

$$\begin{aligned}
&\|J_2^n\|_{H^1} \leqslant C\|e_H^{n-\frac{1}{2}}\|_{H^2},\\
&\|J_3^n\|_{H^1} \leqslant C\|e_H^{n-\frac{1}{2}}\|_{H^1},\\
&\|J_5^n\|_{H^1} \leqslant C\big(\|e_u^n\|_{H^2} + \|e_u^{n-1}\|_{H^2}\big),\\
&\|J_6^n\|_{H^1} \leqslant C\|e_H^{n-\frac{1}{2}}\|_{H^2},\\
&\|J_7^n\|_{H^1} \leqslant C\|e_u^{n-\frac{1}{2}}\|_{H^2},\\
&\|J_8^n\|_{H^1} \leqslant C\|e_u^{n-\frac{1}{2}}\|_{H^1}.
\end{aligned}$$

By the integral form of the mean value theorem,

$$\begin{aligned}
|u_\star^{n-\frac{1}{2}}|u_\star^{n-\frac{1}{2}} - |u_\varphi^{n-\frac{1}{2}}|u_\varphi^{n-\frac{1}{2}} &= (u_\star^{n-\frac{1}{2}} - u_\varphi^{n-\frac{1}{2}}) \cdot \int_0^1 2|(1-s)u_\star^{n-\frac{1}{2}} + su_\varphi^{n-\frac{1}{2}}|\mathrm{d}s\\
&= \rho_\varphi^u \varphi^{n-\frac{1}{2}} \int_0^1 2|(1-s)u_\star^{n-\frac{1}{2}} + su_\varphi^{n-\frac{1}{2}}|\mathrm{d}s\\
&= \rho_\varphi^u \theta e_u^{n-\frac{1}{2}} \int_0^1 2|(1-s)u_\star^{n-\frac{1}{2}} + su_\varphi^{n-\frac{1}{2}}|\mathrm{d}s,
\end{aligned}$$

which implies (together with $|\rho_\varphi^u| \leqslant 1$ and $\theta \leqslant 1$)

$$\|J_4^n\|_{H^1} \leqslant C(\|e_H^{n-\frac{1}{2}}\|_{H^1} + C\|e_u^{n-\frac{1}{2}}\|_{H^1}).$$

Substituting the estimates of $\|J_i^n\|_{H^1}$, $i = 1, \ldots, 8$, into (3.41), we obtain

$$\begin{aligned}
&\max_{1\leqslant n\leqslant m} \|e_u^n\|_{H^2}^2 + \tau \sum_{n=1}^m \big(\|\bar\partial_\tau e_u^n\|_{H^1}^2 + \|e_u^{n-\frac{1}{2}}\|_{H^3}^2\big)\\
&\leqslant C_\delta\, \tau \sum_{n=1}^m \big(\|\eta_u^n\|_{H^1}^2 + \|e_u^n\|_{H^2}^2 + \|e_u^{n-1}\|_{H^2}^2 + \|e_H^{n-\frac{1}{2}}\|_{H^2}^2\big) + \delta\tau \sum_{n=1}^m \|e_u^{n-\frac{1}{2}}\|_{H^3}^2.
\end{aligned} \tag{3.42}$$

Adding $\delta\times$(3.39) to (3.42), we have

$$\begin{aligned}
&\max_{1\leqslant n\leqslant m} \big(\delta\|e_H^n\|_{H^2}^2 + \|e_u^n\|_{H^2}^2\big) + \tau \sum_{n=1}^m \big(\|\bar\partial_\tau e_u^n\|_{H^1}^2 + \|e_u^{n-\frac{1}{2}}\|_{H^3}^2\big)\\
&\leqslant \delta\tau \sum_{n=1}^m \|e_u^{n-\frac{1}{2}}\|_{H^3}^2 + C_\delta\, \tau \sum_{n=1}^m \big(\|\eta_H^n\|_{H^2}^2 + \|\eta_u^n\|_{H^1}^2 + \|e_u^n\|_{H^2}^2 + \|e_u^{n-1}\|_{H^2}^2 + \|e_H^{n-\frac{1}{2}}\|_{H^2}^2\big).
\end{aligned}$$

Choosing here sufficiently small δ, the first term on the right-hand side of the above inequality can be absorbed by the left-hand side, and we infer that

$$\begin{aligned}
&\max_{1\leqslant n\leqslant m} \big(\|e_H^n\|_{H^2}^2 + \|e_u^n\|_{H^2}^2\big) + \tau \sum_{n=1}^m \big(\|\bar\partial_\tau e_u^n\|_{H^1}^2 + \|e_u^{n-\frac{1}{2}}\|_{H^3}^2\big)\\
&\leqslant C\tau \sum_{n=1}^m \big(\|\eta_H^n\|_{H^2}^2 + \|\eta_u^n\|_{H^1}^2 + \|e_u^n\|_{H^2}^2 + \|e_u^{n-1}\|_{H^2}^2 + \|e_H^{n-\frac{1}{2}}\|_{H^2}^2\big).
\end{aligned} \tag{3.43}$$

Then, by using the discrete Gronwall inequality, for sufficiently small τ we obtain

$$\begin{aligned}&\max_{1\leqslant n\leqslant N}\left(\|e_H^n\|_{H^2}^2+\|e_u^n\|_{H^2}^2\right)+\tau\sum_{n=1}^{N}\left(\|\bar\partial_\tau e_u^n\|_{H^1}^2+\|e_u^{n-\frac12}\|_{H^3}^2\right)\\&\leqslant C\tau\sum_{n=1}^{N}\left(\|\eta_H^n\|_{H^2}^2+\|\eta_u^n\|_{H^1}^2\right).\end{aligned}\tag{3.44}$$

This and the consistency estimate (3.15) imply

$$\max_{1\leqslant n\leqslant N}\left(\|e_H^n\|_{H^2}+\|e_u^n\|_{H^2}\right)\leqslant C\tau^2,\tag{3.45}$$

$$\tau\sum_{n=1}^{N}\|e_u^{n-\frac12}\|_{H^3}^2\leqslant C\tau^4.\tag{3.46}$$

In all the estimates above, the generic constant C is independent of ε.

Now, (3.46) implies $\|e_u^n\|_{H^3}\leqslant\|e_u^{n-1}\|_{H^3}+C\tau^{3/2}$ and, therefore,

$$\max_{1\leqslant n\leqslant N}\|e_u^n\|_{H^3}\leqslant C\tau^{\frac12}.\tag{3.47}$$

In view of (3.45) and (3.47), there exists a positive constant τ_0 such that, for $0<\tau\leqslant\tau_0$, we have

$$\max_{1\leqslant n\leqslant N}\|e_H^n\|_{H^2}\leqslant\frac{H_{\min}}{2}\quad\text{and}\quad\max_{1\leqslant n\leqslant N}\|e_u^n\|_{H^3}\leqslant1.\tag{3.48}$$

Thus,

$$\max_{1\leqslant n\leqslant N}\|\phi^n\|_{H^2}\leqslant\frac{H_{\min}}{2}\quad\text{and}\quad\max_{1\leqslant n\leqslant N}\|\varphi^n\|_{H^3}\leqslant1\tag{3.49}$$

and we infer that indeed

$$\rho_{\phi,\varepsilon}^H=\rho_\varphi^u=1.\tag{3.50}$$

In particular, (3.49) implies that $(\phi^n,\varphi^n)_{n=1}^N$ is bounded in $(H^2(\Omega)\times[H^3(\Omega)\cap H_0^1(\Omega)]^2)^N\hookrightarrow X^N$ uniformly for $\theta\in[0,1]$, $\varepsilon\in(0,1)$, and $0<\tau\leqslant\tau_0$. This proves Lemma 3.4.

3.6. Error estimate. The analysis following Lemma 3.4 proves the existence of a solution to (2.6) in $B_{H,u}^n$.

If $(H^n,u^n)\in B_{H,u}^n$ is a solution of (2.6), then the error (e_H^n,e_u^n) is a solution of (3.26)–(3.27) with (ϕ^n,φ^n) given by (3.33), with $\theta=1$ and $\varepsilon=0$. Then, the proof of Lemma 3.4 implies (3.45)–(3.46); see Remark 3.4 for the case $\varepsilon=0$. This proves the error estimate (2.11).

3.7. Uniqueness of discrete solutions. If there are two solutions of (2.6), say, $(H^n,u^n)\in B_{H,u}^n$ and $(\widetilde H^n,\tilde u^n)\in B_{H,u}^n$, then the error functions

$$e_H=\widetilde H^n-H^n\quad\text{and}\quad e_u=\tilde u^n-u^n$$

satisfy (3.16) with $H_\star^n$ and $u_\star^n$ replaced by $\widetilde H^n$ and $\tilde u^n$, respectively, and with $\eta_H^n=\eta_u^n=0$. Then, the error estimate (3.44) holds, which implies $e_H^n=e_u^n=0$. This proves the uniqueness of discrete solutions.

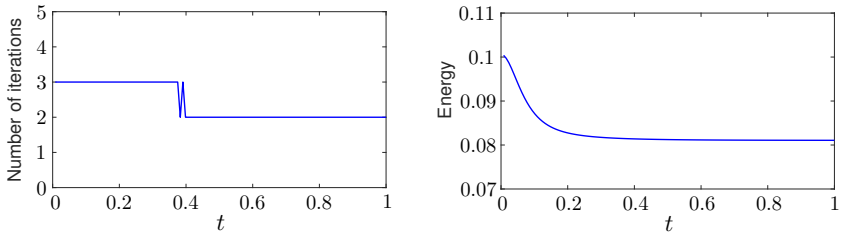

FIG. 4.1. *Number of iterations and energy at each time level.*

4. Numerical results. In this section, we present numerical results to support our theoretical analysis in Theorems 2.1 and 2.2.

We test energy decay and convergence rates of the proposed method by solving the initial and boundary value problem (1.1)–(1.4) in the domain $\Omega = [0,1] \times [0,1]$ with the following initial values:

$$H_0(x,y) = 1.0 + 0.2\sin(\pi x), \quad u_0(x,y) = (\sin^2(\pi x)y^2(1-y)^2, x^2(1-x)^2\sin^2(\pi y))^\top,$$

and with $g = 10$ and $H_b = \mu = c_f = 1$.

We solve the problem by using the proposed time discretization method with the FEM in (2.9), with a sufficiently small mesh size h such that the spatial discretization error is negligible in observing the temporal convergence rates. The nonlinear system (2.9) is solved by the following fixed-point iteration: choose $u^n_{h,0} = u^{n-1}_h$ and compute $(H^n_{h,\ell}, u^n_{h,\ell})$, $\ell = 1, 2, \ldots$, by

(4.1)
$$\begin{cases} (\bar\partial_\tau H^n_{h,\ell}, \phi_h) - \big(H^{n-\frac12}_{h,\ell} u^{n-\frac12}_{h,\ell-1}, \nabla\phi_h\big) = 0 & \forall\, \phi_h \in S_h, \\ (\bar\partial_\tau u^n_{h,\ell}, H^{n-\frac12}_{h,\ell} v_h) + \big(\mu H^{n-\frac12}_{h,\ell}\nabla u^{n-\frac12}_{h,\ell}, \nabla v_h\big) + \big(c_f|u^{n-\frac12}_{h,\ell-1}|u^{n-\frac12}_{h,\ell}, v_h\big) & \\ = -\Big(\nabla P_h\Big[\frac14\big(|u^n_{h,\ell-1}|^2 + |u^{n-1}_h|^2\big) + g(H^{n-\frac12}_{h,\ell} - H_b)\Big], H^{n-\frac12}_{h,\ell} v_h\Big) & \\ \quad - \big((\nabla\times u^{n-\frac12}_{h,\ell-1} + f^{n-\frac12})\hat k \times u^{n-\frac12}_{h,\ell}, H^{n-\frac12}_{h,\ell} v_h\big) & \forall\, v_h \in \mathring S^2_h. \end{cases}$$

For given $u^{n-\frac12}_{h,\ell-1}$, one can determine $H^n_{h,\ell}$ from the first equation of (4.1) and then compute $H^{n-\frac12}_{h,\ell}$. By using this computed $H^{n-\frac12}_{h,\ell}$, one can determine $u^n_{h,\ell}$ from the second equation of (4.1). The iteration is terminated when the following tolerance error is reached,

$$\|H^n_{h,\ell} - H^n_{h,\ell-1}\|_{L^\infty(\Omega)} < 10^{-7} \quad \text{and} \quad \|u^n_{h,\ell} - u^n_{h,\ell-1}\|_{L^\infty(\Omega)} < 10^{-7}, \tag{4.2}$$

which is much smaller than the temporal discretization errors observed in our numerical results.

The number of iterations at each time level, with $\tau = 1/128$, is presented in Figure 4.1 (left), which shows that the nonlinear system can be effectively solved with a few iterations to achieve the accuracy in (4.2). The energy of numerical solutions is presented in Figure 4.1 (right), which shows that the energy decays in time, consistent with the theoretical result of Theorem 2.1.

TABLE 4.1
Numerical results at $T = 1$.

	$\|H^N_{h,\tau} - H^N_{h,\tau/2}\|_{L^\infty(\Omega)}$	$\|u^N_{h,\tau} - u^N_{h,\tau/2}\|_{L^\infty(\Omega)}$
$\tau = 1/8$	6.706×10^{-4}	7.517×10^{-3}
$\tau = 1/16$	1.104×10^{-4}	1.805×10^{-3}
$\tau = 1/32$	2.073×10^{-5}	4.463×10^{-4}
$\tau = 1/64$	4.232×10^{-6}	1.176×10^{-4}
convergence rate	2.28	2.01

Since $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$ in the two-dimensional space, the error estimate in Theorem 2.2 implies that the proposed method has second-order convergence in $L^\infty(\Omega)$. Since the exact solution is unknown, we present the $L^\infty(\Omega)$-errors of numerical solutions in Table 4.1 based on the difference between two numerical solutions using consecutive stepsizes, with a sufficiently small mesh size h such that the spatial discretization error is negligible in observing the temporal convergence rates, which are computed by using the formula

$$\text{convergence rate} = \log_2\left(\frac{\|u^N_{h,2\tau} - u^N_{h,\tau}\|_{L^\infty(\Omega)}}{\|u^N_{h,\tau} - u^N_{h,\tau/2}\|_{L^\infty(\Omega)}}\right)$$

based on the three finest stepsizes. The numerical results indicate that the proposed method has second-order convergence in time, consistent with the theoretical result of Theorem 2.2.

Appendix: Well-posedness of the linear hyperbolic equation (3.29). First, we prove uniqueness of the solution of (3.29) for sufficiently small τ. If w and v are solutions of (3.29), then

$$\frac{2}{\tau}(w - v) + u_\varphi^{n-\frac{1}{2}} \cdot \nabla(w - v) + (\nabla \cdot u_\varphi^{n-\frac{1}{2}})(w - v) = 0. \tag{A.1}$$

Testing (A.1) by $w - v$ immediately yields

$$\frac{2}{\tau}\|w - v\|^2_{L^2} = -\frac{1}{2}(\nabla \cdot u_\varphi^{n-\frac{1}{2}}, |w - v|^2) \leqslant \frac{1}{2}\|\nabla \cdot u_\varphi^{n-\frac{1}{2}}\|_{L^\infty}\|w - v\|^2_{L^2}.$$

Hence, for

$$\tau \leqslant \frac{1}{\|\nabla \cdot u_\varphi^{n-\frac{1}{2}}\|_{L^\infty}}, \tag{A.2}$$

this estimate implies $w - v = 0$.

Second, we prove existence of a solution $e_H^{n-\frac{1}{2}} \in H^2(\Omega)$ to (3.29). To this end, we let g_H^n and $u_\varphi^{n-\frac{1}{2}}$ be extended to $H^2(\mathbb{R}^2)$ and $H^3(\mathbb{R}^2)$, respectively, both with compact supports in some bounded domain $\Omega' \supset \overline{\Omega}$, and consider the viscous approximating problem

$$\frac{2}{\tau}v_\delta + u_\varphi^{n-\frac{1}{2}} \cdot \nabla v_\delta + (\nabla \cdot u_\varphi^{n-\frac{1}{2}})v_\delta - \delta\Delta v_\delta = g_H^n \tag{A.3}$$

with a small parameter $\delta > 0$. It is well known that, for sufficiently small τ, satisfying (A.2), the elliptic equation (A.3) has a unique solution $v_\delta \in H^4(\mathbb{R}^2)$. Clearly, (3.25)

and (3.28) imply (A.2). Thus (A.3) has a solution $v_\delta \in H^4(\mathbb{R}^2)$ under condition (3.28).

In the following, we prove that v_δ converges strongly in $H^1(\Omega)$ and weakly in $H^2(\Omega)$ to a solution of (3.29) as $\delta \to 0$. In fact, differentiating (A.3) twice yields

$$\begin{aligned}
&\frac{2}{\tau}\partial_{ij}v_\delta + u_\varphi^{n-\frac{1}{2}} \cdot \nabla \partial_{ij} v_\delta + (\nabla \cdot u_\varphi^{n-\frac{1}{2}})\partial_{ij} v_\delta - \delta \Delta \partial_{ij} v_\delta \\
&= \partial_{ij} g_H^n - \partial_i u_\varphi^{n-\frac{1}{2}} \cdot \nabla \partial_j v_\delta - \partial_j u_\varphi^{n-\frac{1}{2}} \cdot \nabla \partial_i v_\delta - \partial_{ij} u_\varphi^{n-\frac{1}{2}} \cdot \nabla v_\delta \\
&\quad - (\nabla \cdot \partial_i u_\varphi^{n-\frac{1}{2}})\partial_j v_\delta - (\nabla \cdot \partial_j u_\varphi^{n-\frac{1}{2}})\partial_i v_\delta - (\nabla \cdot \partial_{ij} u_\varphi^{n-\frac{1}{2}}) v_\delta.
\end{aligned}$$

Then, testing this equation by $\partial_{ij} v_\delta$, we obtain

$$\begin{aligned}
\frac{2}{\tau}\|\partial_{ij} v_\delta\|_{L^2(\mathbb{R}^2)}^2 + \delta \|\nabla \partial_{ij} v_\delta\|_{L^2(\mathbb{R}^2)}^2 &\leqslant C\|g_H^n\|_{H^2}\|v_\delta\|_{H^2(\Omega')} + C\|u_\varphi^{n-\frac{1}{2}}\|_{H^3}\|v_\delta\|_{H^2(\Omega')}^2 \\
&\leqslant C\|g_H^n\|_{H^2}^2 + (C + C\|u_\varphi^{n-\frac{1}{2}}\|_{H^3})\|v_\delta\|_{H^2(\mathbb{R}^2)}^2.
\end{aligned}$$

Similarly, one can obtain

$$\frac{2}{\tau}\|v_\delta\|_{L^2(\mathbb{R}^2)}^2 + \delta \|\nabla v_\delta\|_{L^2(\mathbb{R}^2)}^2 \leqslant C\|g_H^n\|_{H^2}^2 + (C + C\|u_\varphi^{n-\frac{1}{2}}\|_{H^3})\|v_\delta\|_{H^2(\mathbb{R}^2)}^2$$

and

$$\frac{2}{\tau}\|\partial_j v_\delta\|_{L^2(\mathbb{R}^2)}^2 + \delta \|\nabla \partial_j v_\delta\|_{L^2(\mathbb{R}^2)}^2 \leqslant C\|g_H^n\|_{H^2}^2 + (C + C\|u_\varphi^{n-\frac{1}{2}}\|_{H^3})\|v_\delta\|_{H^2(\mathbb{R}^2)}^2.$$

These estimates imply

$$\frac{2}{\tau}\|v_\delta\|_{H^2(\mathbb{R}^2)}^2 + \delta \|\nabla \partial_{ij} v_\delta\|_{L^2(\mathbb{R}^2)}^2 \leqslant C\|g_H^n\|_{H^2}^2 + (C + C\|u_\varphi^{n-\frac{1}{2}}\|_{H^3})\|v_\delta\|_{H^2(\mathbb{R}^2)}^2. \tag{A.4}$$

When τ is sufficiently small, the last term on the right-hand side of (A.4) can be absorbed by the left-hand side. Then, we have

$$\|v_\delta\|_{H^2(\mathbb{R}^2)} \leqslant C\|g_H^n\|_{H^2}$$

with a constant C independent of δ. Hence, there exists a subsequence $\delta_k \to 0$ such that v_{δ_k} converges strongly in $H^1(\Omega)$ and weakly in $H^2(\Omega)$ to some function, which we denote by $e_H^n \in H^2(\Omega)$. Then, by letting $\delta = \delta_k \to 0$ in (A.3), we obtain that e_H^n is the solution of (3.29) on $\mathbb{R}^2$ (therefore it is also a solution on Ω). The uniqueness of such a solution has already been proved.

Acknowledgment. Georgios Akrivis gratefully acknowledges the hospitality at the Beijing Computational Science Research Center in the fall of 2019, where part of this work was conducted. We thank the anonymous referees for their helpful comments and suggestions.

REFERENCES

[1] V. Agoshkov, E. Ovchinnikov, A. Quarteroni, and F. Saleri, *Recent developments in the numerical simulation of shallow water equations.* II. *Temporal discretization*, Math. Models Methods Appl. Sci., 4 (1994), pp. 533–556.

[2] G. D. Akrivis, V. A. Dougalis, and O. A. Karakashian, *On fully discrete Galerkin methods of second-order temporal accuracy for the nonlinear Schrödinger equation*, Numer. Math., 59 (1991), pp. 31–53.
[3] D. C. Antonopoulos and V. A. Dougalis, *Galerkin-FEMs for the shallow water equations with characteristic boundary conditions*, IMA J. Numer. Anal., 37 (2017), pp. 266–295.
[4] D. C. Antonopoulos, V. A. Dougalis, and G. Kounadis, *On the standard Galerkin method with explicit* RK4 *time stepping for the shallow water equations*, IMA J. Numer. Anal., 40 (2020), pp. 2415–2449, https://doi.org/10.1093/imanum/drz033.
[5] W. Arendt, C. J. Batty, M. Hieber, and F. Neubrander, *Vector-valued Laplace Transforms and Cauchy Problems*, 2nd ed., Birkhäuser, Basel, 2011.
[6] P. Azerad, J.-L. Guermond, and B. Popov, *Well-balanced second-order approximation of the shallow water equation with continuous finite elements*, SIAM J. Numer. Anal., 55 (2017), pp. 3203–3224.
[7] M. J. Castro Díaz, C. Chalons, and T. Morales de Luna, *A fully well-balanced Lagrange-projection-type scheme for the shallow-water equations*, SIAM J. Numer. Anal., 56 (2018), pp. 3071–3098.
[8] S. Chippada, C. N. Dawson, M. L. Martínez-Canales, and M. F. Wheeler, *Finite element approximations to the system of shallow water equations, Part* I: *Continuous-time a priori error estimates*, SIAM J. Numer. Anal., 35 (1998), pp. 692–711.
[9] S. Chippada, C. N. Dawson, M. L. Martínez-Canales, and M. F. Wheeler, *Finite element approximations to the system of shallow water equations, Part* II: *Discrete-time a priori error estimates*, SIAM J. Numer. Anal., 36 (1998), pp. 226–250.
[10] P. G. Ciarlet, *Linear and Nonlinear Functional Analysis with Applications*, SIAM, Philadelphia, 2013.
[11] C. N. Dawson and M. L. Martínez-Canales, *A characteristic-Galerkin approximation to a system of shallow water equations*, Numer. Math., 86 (2000), pp. 239–256.
[12] M. Delfour, M. Fortin, and G. Payre, *Finite-difference solutions of a non-linear Schrödinger equation*, J. Comput. Phys., 44 (1981), pp. 277–288.
[13] L. C. Evans, *Partial Differential Equations*, 2nd ed., Grad. Stud. Math. 19, American Mathematical Society, Providence, RI, 2010.
[14] R. L. Higdon, *Numerical modelling of ocean circulation*, Acta Numer., 15 (2006), pp. 385–470.
[15] G. Kounadis and V. A. Dougalis, *Galerkin finite element methods for the shallow water equations over variable bottom*, J. Comput. Appl. Math., 373 (2020), 112315.
[16] B. Li and W. Sun, *Unconditional convergence and optimal error estimates of a Galerkin-mixed FEM for incompressible miscible flow in porous media*, SIAM J. Numer. Anal., 51 (2013), pp. 1959–1977.
[17] B. Li and W. Sun, *Error analysis of linearized semi-implicit Galerkin FEMs for nonlinear parabolic equations*, Int. J. Numer. Anal. Model., 10 (2013), pp. 622–633.
[18] D. R. Lynch and W. G. Gray, *A wave equation model for finite element tidal computations*, Comput. & Fluids, 7 (1979), pp. 207–228.
[19] S. Noelle, Y. Xing, and C.-W. Shu, *High-order well-balanced finite volume WENO schemes for shallow water equation with moving water*, J. Comput. Phys., 226 (2007), pp. 29–58.
[20] E. M. Ouhabaz, *Gaussian estimates and holomorphy of semigroups*, Proc. Amer. Math. Soc., 123 (1995), pp. 1465–1474.
[21] K. Pieper, K. C. Sockwell, and M. Gunzburger, *Exponential time differencing for mimetic multilayer ocean models*, J. Comput. Phys., 398 (2019), 108900.
[22] J. Shen and X. Yang, *Numerical approximations of Allen-Cahn and Cahn-Hilliard equations*, Discrete Contin. Dyn. Syst., 28 (2010), pp. 1669–1691.
[23] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, NJ, 1970.
[24] W. A. Strauss and L. Vazquez, *Numerical solution of a nonlinear Klein–Gordon equation*, J. Comput. Phys., 28 (1978), pp. 271–278.
[25] L. Sundbye, *Global existence for the Dirichlet problem for the viscous shallow water equations*, J. Math. Anal. Appl., 202 (1996), pp. 236–258.
[26] L. Sundbye, *Global existence for the Cauchy problem for the viscous shallow water equations*, Rocky Mountain J. Math., 28 (1998), pp. 1135–1152.
[27] B. A. Ton, *Existence and uniqueness of a classical solution of an initial-boundary value problem of the theory of shallow waters*, SIAM J. Math. Anal., 12 (1981), pp. 229–241.
[28] I. Tregubov and T. Tran, *A Galerkin method with spherical splines for the shallow water equations on a sphere: Error analysis*, Numer. Math., 129 (2015), pp. 783–814.

[29] J. Wang, Z. Si, and W. Sun, *A new error analysis of characteristics-mixed FEMs for miscible displacement in porous media*, SIAM J. Numer. Anal., 52 (2014), pp. 3000–3020.
[30] Y. Xing, *Numerical methods for the nonlinear shallow water equations*, in Handbook of Numerical Analysis, Vol. 18, R. Abgrall and C.-W. Shu, eds., Elsevier, New York, 2017, pp. 361–384.
[31] Y. Xing and C.-W. Shu, *High order finite difference WENO schemes with the exact conservation property for the shallow water equations*, J. Comput. Phys., 208 (2005), pp. 206–227.