A numerical analysis focused comparison of several Finite Volume schemes for an Unipolar Degenerated Drift-Diffusion Model

Clément Cancès, Claire Chainais-Hillairet, Jürgen Fuhrmann, Benoît Gaudeul

A NUMERICAL ANALYSIS FOCUSED COMPARISON OF SEVERAL FINITE VOLUME SCHEMES FOR A UNIPOLAR DEGENERATE DRIFT-DIFFUSION MODEL

CLÉMENT CANCÈS, CLAIRE CHAINAIS-HILLAIRET, JÜRGEN FUHRMANN, AND BENOÎT GAUDEUL

ABSTRACT. In this paper, we consider a unipolar degenerate drift-diffusion system where the relation between the concentration of the charged species c and the chemical potential h is $h(c) = \log \frac{c}{1-c}$. We design four different finite volume schemes based on four different formulations of the fluxes. We provide a stability analysis and existence results for the four schemes. The convergence proof with respect to the discretization parameters is established for two of them. Numerical experiments illustrate the behaviour of the different schemes.

1. INTRODUCTION

1.1. **Motivation.** Unipolar drift-diffusion models describe the transport of a charged species in the presence of a fixed or moving countercharge. They consist of the coupling of a drift-diffusion equation on the density of the charged species c with a Poisson equation on the electric potential Φ. They can be written under a general form as

$$\begin{cases} \partial_t c + \operatorname{div}(\mathbf{J}) = 0, \quad \mathbf{J} = -\eta(c)\nabla(h(c) + \Phi), \\ -\lambda^2 \Delta \Phi = c + c^{\mathrm{dop}}, \end{cases}$$

where h is the chemical potential, η the mobility coefficient, λ the scaled Debye length coming from the nondimensionalisation of the physical model and c^{dop} describes the doping profile of the media.

Such models occur in many interesting application cases. Charge carriers in most classical semiconductors exhibit a relationship $c = \mathcal{F}(h)$, where $\mathcal{F}$ is the Fermi integral of index $\frac{1}{2}$ which can be approximated in the range $-\infty < h \lessapprox 1.3$ by the function $\mathcal{F}(h) = \frac{1}{\gamma + \exp(-h)}$ with $\gamma = 0.27$ [6]. For $\gamma = 1$, this relationship corresponds to the Fermi integral of index -1 and implies $h = \log \frac{c}{1-c}$. It is the limit for vanishing disorder of the Gauss-Fermi integral [42, 45] which is used to describe organic semiconductors [16]. A similar relationship is valid for the oxygen ion concentration in a solid oxide electrolyte [49] and a simple model of an ionic liquid [30].

While the relationship between chemical potential and concentration is sufficient to describe the thermodynamic equilibrium, the description of charge transport driven by the sum of the gradients of the chemical potential and the electrostatic potential Φ needs an additional specification of the mobility coefficient η. Setting this coefficient proportional to the concentration c is common in the case of semiconductors [48]. A similar ansatz describes the limit of large lattice mass density in solid oxide electrolytes. It also follows from a formal reduction of a generalized Nernst-Planck model [20, 19] to the case of a mixture of two charged species including an infinitely mobile and charged solvent – ionic liquids – as performed in [30]. We hint that more general and fully consistent models for both solid oxide electrolytes and ionic liquids consider mobility coefficients of the type $c(1-c)$ [49, 8, 37].

In this paper, we consider that the mobility coefficient is $\eta(c) = c$ and the chemical potential $h(c) = \log \frac{c}{1-c}$ (corresponding to $\mathcal{F}(h) = \frac{1}{1+\exp(-h)}$). Strong degeneration described by a bounded dependency of the concentration c on the chemical potential h leads to some structural mathematical challenges in the corresponding drift-diffusion models. These need to be addressed properly in numerical schemes. The consideration of this simplified model is a starting point for the study of generalized Nernst-Planck models for multiple ionic species in electroneutral solvents [20, 19, 29, 30]. Moreover, the design of discretization methods for the case where $\eta(c) = c(1-c)$ is also a possible topic of further investigation following the present paper.

1.2. **A simplified unipolar degenerate drift-diffusion model.** Let us now define the framework of the study. We consider the evolution of the concentration c of a charged species in a connected bounded open domain Ω of $\mathbb{R}^d$ ($d \leq 3$) with polyhedral and Lipschitz continuous boundary $\partial\Omega$ during a finite but arbitrary time $T > 0$. After nondimensionalisation with appropriate scaling, we regard the following system of partial differential equations (PDEs). The concentration c satisfies the conservation law

$$\partial_t c + \operatorname{div}(\mathbf{J}) = 0 \quad \text{in } (0,T)\times\Omega. \tag{1.1}$$ {eq:cons_l

The flux $\mathbf{J}$ is negatively proportional to the gradient of the electrochemical potential as expressed by the expression

$$\mathbf{J} = -c\nabla\left(h(c) + \Phi\right) \quad \text{in } (0,T)\times\Omega, \tag{1.2}$$ {cances fl

where $h(c) = \log\left(\frac{c}{1-c}\right)$ is the chemical potential. In what follows, we consider that the electrostatic potential Φ is related to space charge density thanks to the Poisson equation

$$-\Delta\Phi = c + c^{\mathrm{dop}} \quad \text{in } (0,T)\times\Omega, \tag{1.3}$$ {Poisson e

which means that the Debye length is set to 1. Extension to general Debye length is straightforward. The doping profile c^{dop} is assumed to be constant w.r.t. time and to be bounded, i.e., $c^{\mathrm{dop}} \in L^\infty(\Omega)$.

One interpretation of c is the concentration of majority carriers (holes) in a p-type organic semiconductor with constant in time doping. Another interpretation of c is the cation concentration in an ionic liquid following the formal approach introduced in [30].

The system is supplemented with the prescription of the initial concentration

$$c_{|t=0} = c^0 \in L^\infty(\Omega) \quad \text{with} \quad 0 \leq c^0 \leq 1 \quad \text{and} \quad 0 < \overline{c} = \fint_\Omega c^0 \mathrm{d}\boldsymbol{x} < 1, \tag{1.4}$$ {eq:c^0}

and of boundary conditions. The choice of the boundary conditions may depend on the targeted application: organic semiconductor or ionic liquid. For the analysis purpose, we consider boundary conditions which are well adapted to the ionic liquid model. Other boundary conditions will also be considered in the numerical simulations in Section 5. They are no-flux boundary conditions for the concentration:

$$\mathbf{J}\cdot\boldsymbol{n} = 0 \quad \text{on } (0,T)\times\partial\Omega. \tag{1.5}$$ {eq:no-flu

The Poisson equation (1.3) is supplemented by inhomogeneous Dirichlet boundary conditions on a part Γ_D of $\partial\Omega$ with positive measure, and by homogeneous Neumann boundary condition on the remaining part $\Gamma_N = \partial\Omega \setminus \Gamma_D$ of the boundary:

$$\Phi = \Phi^D \quad \text{on } (0,T)\times\Gamma_D, \qquad \nabla\Phi\cdot\boldsymbol{n} = 0 \quad \text{on } (0,T)\times\Gamma_N. \tag{1.6}$$ {eq:BC_Phi

Throughout the paper, we assume that Φ^D is defined on the whole domain Ω and does not depend on time, with $\Phi^D \in H^1(\Omega)\cap L^\infty(\Omega)$.

The goal of this paper is to study and compare several different Finite Volume schemes for the system (1.1)–(1.6). They are based on various reformulations of the flux $\mathbf{J}$. Indeed,

we may introduce either the so-called excess chemical potential [41] $\nu(c) = h(c) - \log(c) = -\log(1-c)$, or the activity and the inverse of the activity coefficient [44] respectively defined by $a(c) = e^{h(c)} = \frac{c}{1-c}$, and $\beta(c) = \frac{c}{a(c)} = 1-c$, or the diffusion enhancement [50] $r(c) = -\log(1-c)$ satisfying $r'(c) = ch'(c)$. Even though ν and r happen to be the same function for the nonlinearity considered in this paper, we keep different notations to emphasize their different physical meaning and expression in more complex systems. The different notations used throughout the paper are collected in Appendix D. Then the flux $\mathbf{J}$, initially defined by (1.2), satisfies

$$\mathbf{J} = -\nabla c - c\nabla\left(\Phi + \nu(c)\right), \tag{1.7}$$
$$= -\beta(c)(\nabla a(c) + a(c)\nabla\Phi), \tag{1.8}$$
$$= -r'(c)\nabla c - c\nabla\Phi. \tag{1.9}$$

{Sedan flu {Jurgen fl {Marianne

These formulations (1.2), (1.7), (1.8) and (1.9) lead to different schemes that we aim to compare from a numerical analysis point of view. We may notice that the flux $\mathbf{J}$ can also be expressed as

$$\mathbf{J} = -\nabla r(c) - c\nabla\Phi. \tag{1.10}$$

{Weak solu

This last formulation will be used to define the weak solution to (1.1)–(1.6).

Before going to the discretization of the problem, let us highlight the entropy structure of system (1.1)–(1.6), which plays a central role in what follows.

1.3. **Entropy structure and weak solutions.** The goal of this section is to shortly depict the gradient flow structure of the system (1.1)–(1.6). We stay here at a formal level and remain sloppy about regularity issues. The solutions (c, Φ) to (1.1)–(1.6) are supposed to be regular enough so that the following calculations are justified. Define the mixing entropy density

$$H(c) = c\log(c) + (1-c)\log(1-c),$$

which is an antiderivative of h, then the electrochemical energy is given by

$$E(c,\Phi) = \int_\Omega \left\{ H(c) + \frac{1}{2}|\nabla\Phi|^2 \right\} \mathrm{d}\boldsymbol{x} - \int_{\Gamma_D} \Phi^D \nabla\Phi\cdot\boldsymbol{n}\mathrm{d}\boldsymbol{\gamma}. \tag{1.11}$$

{eq:E}

The next proposition shows that the electrochemical energy is a Lyapunov functional. Moreover, the dissipation rate for the energy is explicitly given.

prop:E **Proposition 1.1.** *Let (c,Φ) be a smooth solution to (1.1)–(1.6), with c bounded away from 0 and 1, then*

$$\frac{\mathrm{d}}{\mathrm{d}t}E(c,\Phi) + \int_\Omega c\left|\nabla(h(c)+\Phi)\right|^2 \mathrm{d}\boldsymbol{x} = 0.$$

Proof. We notice first that since Φ^D does not depend on time,

$$\frac{\mathrm{d}}{\mathrm{d}t}E(c,\Phi) = \int_\Omega (h(c)\partial_t c + \nabla\Phi\cdot\partial_t\nabla\Phi)\mathrm{d}\boldsymbol{x} - \int_{\Gamma_D}\Phi^D\partial_t\nabla\Phi\cdot\boldsymbol{n}\mathrm{d}\boldsymbol{\gamma}.$$

Then we apply the Gauss theorem, and we use the Poisson equation (1.3) with a constant doping profile, to get

$$\frac{\mathrm{d}}{\mathrm{d}t}E(c,\Phi) = \int_\Omega (h(c)+\Phi)\partial_t c.$$

Multiplying the conservation law (1.1) by $h(c)+\Phi$ and integrating over the domain Ω yields

$$\int_\Omega \partial_t c(h(c)+\Phi) = -\int_\Omega c\left|\nabla(h(c)+\Phi)\right|^2\mathrm{d}\boldsymbol{x},$$

thanks to the no-flux boundary condition (1.5). This concludes the proof of Proposition 1.1. □

Let $c \in L^\infty(\Omega;[0,1])$. We denote by $\Phi[c]$ the unique solution to (1.3). One can easily check that the energy functional $c \mapsto E(c,\Phi[c])$ is bounded on $L^\infty(\Omega;[0,1])$. Indeed, H takes values in $[-\log 2, 0]$ and the bounds on the electrical energy can be obtained by multiplying the Poisson equation by $\Phi - \Phi^D$ and Φ and integrating over Ω. Therefore, $E(c(t),\Phi(t))$ is finite for all $t>0$, whence a $L^\infty((0,T);H^1(\Omega))$ estimate on Φ. We also deduce from Proposition 1.1 that the total energy dissipation is bounded, i.e.

$$\int_0^T \int_\Omega c\,|\nabla(h(c)+\Phi)|^2 \,\mathrm{d}\boldsymbol{x}\mathrm{d}t \le C \tag{1.12}$$ {eq:dissip

for some C uniform with respect to the final time horizon T. Using again that $0 \le c \le 1$, we deduce from (1.12) that

$$\int_0^T \int_\Omega |\nabla r(c)|^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t \le \int_0^T \int_\Omega c\,|\nabla h(c)|^2 \,\mathrm{d}\boldsymbol{x}\mathrm{d}t \le C. \tag{1.13}$$ {eq:L2H1_r

The aforementioned $L^\infty((0,T);H^1(\Omega))$ estimate on the potential Φ and Estimate (1.13) on $r(c)$ suggest a notion of weak solution which is based on the expression (1.10) of the flux $\mathbf{J}$. In what follows, we denote the vector spaces:

$$\mathcal{H}_{\Gamma^D} = \{f \in H^1(\Omega),\, f_{|\Gamma_D} = 0\} \qquad \text{and} \qquad Q_T = (0,T)\times\Omega.$$

ef:weaksol **Definition 1.** *A couple (c,Φ) is a weak solution of (1.1)–(1.6) if*

- $c \in L^\infty((Q_T;[0,1])$ *with* $r(c) \in L^2((0,T);H^1(\Omega))$, *and* $\Phi - \Phi^D \in L^\infty((0,T),\mathcal{H}_{\Gamma^D})$;
- *for all* $\varphi \in C_c^\infty([0,T)\times\overline{\Omega})$,

$$\iint_{Q_T} c\partial_t\varphi \mathrm{d}\boldsymbol{x}\mathrm{d}t + \int_\Omega c^0\varphi(0,\cdot)\mathrm{d}\boldsymbol{x} - \iint_{Q_T} (\nabla r(c) + c\nabla\Phi)\cdot\nabla\varphi \mathrm{d}\boldsymbol{x}\mathrm{d}t = 0; \tag{1.14}$$ {eq:weak_c

- *for all* $\psi \in \mathcal{H}_{\Gamma^D}$ *and almost all* $t \in (0,T)$,

$$\int_\Omega \nabla\Phi(t,\boldsymbol{x})\cdot\nabla\psi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \int_\Omega (c(t,\boldsymbol{x}) + c^{\mathrm{dop}}(\boldsymbol{x}))\psi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{1.15}$$ {eq:weak_P

The goal of this paper is to compare from a numerical analysis point of view several different numerical schemes to approximate the solutions to (1.1)–(1.6). We pay particular attention to the preservation at the discrete level of the key properties of the continuous model, in particular concerning the preservation of the physical bounds $0 \le c \le 1$ and the energy/energy dissipation relation highlighted in Proposition 1.1. The definition of the Finite Volume approximation is detailed in the next section.

Existence of weak solutions to (1.1)–(1.6) is a by-product of Theorem 2.2 which states the convergence of some finite volume approximations towards weak solutions. As far as we know, there is no uniqueness result covering the model in its full generality. It seems to us that the closest uniqueness result is due to Gajewski [34] in the framework of bipolar drift-diffusion system. This proof requires an $L^\infty(Q_T)$ bound on the chemical potential $h(c)$, which has not been yet established for our system.

2. Finite Volume approximations

This section is organized as follows. First, in Section 2.1, we state the requirements on the mesh and fix some notations. Then in Section 2.2, we describe the common basis for the different schemes to be studied in this paper. All the methods presented in this paper rely on so-called two-point flux approximations, but four different schemes are introduced in Section 2.3 based on the formulations (1.2), (1.7), (1.8) and (1.9) of the flux

J. Then in Section 2.4, we state our two main results. The first one, namely Theorem 2.1, focuses on the case of a fixed mesh. We are interested in the existence of a solution to the nonlinear system corresponding to the schemes, and the dissipation of the energy at the discrete level. More precisely, one establishes that all the studied schemes satisfy a discrete counterpart to Proposition 1.1. Our second main result, namely Theorem 2.2, is devoted to the convergence of the scheme as the time step and the mesh size tend to 0.

ssec:mesh

2.1. **Discretization of** $(0,T)\times\Omega$**.** In this paper, we perform a parallel study of four numerical schemes based on two-point flux approximation (TPFA) finite volume schemes. As explained in [21, 26], this approach appears to be very efficient as soon as the continuous problems to be solved numerically are isotropic and one has the freedom to choose a suitable mesh fulfilling the so-called orthogonality condition [40, 27]. We recall here the definition of such a mesh, which is illustrated in Figure 1.

def:mesh

Definition 2. *An admissible mesh of* Ω *is a triplet* $\left(\mathcal{T},\mathcal{E},(\boldsymbol{x}_K)_{K\in\mathcal{T}}\right)$ *such that the following conditions are fulfilled.*

(i) Each control volume (or cell) $K\in\mathcal{T}$ *is non-empty, open, polyhedral and convex. We assume that*

$$K\cap L=\emptyset \quad \text{if } K,L\in\mathcal{T} \text{ with } K\neq L, \qquad \text{while} \quad \bigcup_{K\in\mathcal{T}}\overline{K}=\overline{\Omega}.$$

(ii) Each face $\sigma\in\mathcal{E}$ *is closed and is contained in a hyperplane of* $\mathbb{R}^d$*, with positive* $(d-1)$*-dimensional Hausdorff (or Lebesgue) measure denoted by* $m_\sigma=\mathcal{H}^{d-1}(\sigma)>0$*. We assume that* $\mathcal{H}^{d-1}(\sigma\cap\sigma')=0$ *for* $\sigma,\sigma'\in\mathcal{E}$ *unless* $\sigma'=\sigma$*. For all* $K\in\mathcal{T}$*, we assume that there exists a subset* $\mathcal{E}_K$ *of* $\mathcal{E}$ *such that* $\partial K=\bigcup_{\sigma\in\mathcal{E}_K}\sigma$*. Moreover, we suppose that* $\bigcup_{K\in\mathcal{T}}\mathcal{E}_K=\mathcal{E}$*. Given two distinct control volumes* $K,L\in\mathcal{T}$*, the intersection* $\overline{K}\cap\overline{L}$ *either reduces to a single face* $\sigma\in\mathcal{E}$ *denoted by* $K|L$*, or its* $(d-1)$*-dimensional Hausdorff measure is* 0*.*

(iii) The cell centers $(\boldsymbol{x}_K)_{K\in\mathcal{T}}$ *belong to their cell:* $\boldsymbol{x}_K\in K$*, and are such that, if* $K,L\in\mathcal{T}$ *share a face* $K|L$*, then the vector* $\boldsymbol{x}_L-\boldsymbol{x}_K$ *is orthogonal to* $K|L$*.*

(iv) For the boundary faces $\sigma\subset\partial\Omega$*, we assume that either* $\sigma\subset\Gamma_D$ *or* $\sigma\subset\overline{\Gamma}_N$*. For* $\sigma\subset\partial\Omega$ *with* $\sigma\in\mathcal{E}_K$ *for some* $K\in\mathcal{T}$*, we assume additionally that there exists* $\boldsymbol{x}_\sigma\in\sigma$ *such that* $\boldsymbol{x}_\sigma-\boldsymbol{x}_K$ *is orthogonal to* σ*.*

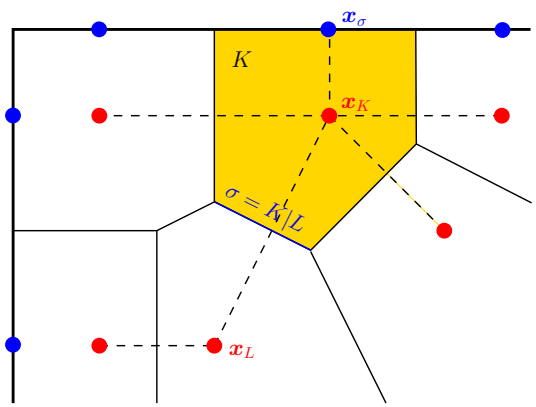

FIGURE 1. Illustration of an admissible mesh as in Definition 2.

fig:mesh

We denote by m_K the d-dimensional Lebesgue measure of the control volume K. The set of the faces is partitioned into two subsets: the set $\mathcal{E}_{\rm int}$ of the interior faces defined by $\mathcal{E}_{\rm int}=\{\sigma\in\mathcal{E}\mid\sigma=K|L \text{ for some } K,L\in\mathcal{T}\}$, and the set $\mathcal{E}_{\rm ext}$ of the exterior faces

defined by $\mathcal{E}_{\rm ext} = \{\sigma \in \mathcal{E} \mid \sigma \subset \partial\Omega\}$, which can also be partitioned into $\mathcal{E}^D = \{\sigma \subset \Gamma_D\}$ and $\mathcal{E}^N = \{\sigma \subset \overline{\Gamma}_N\}$. For a given control volume $K \in \mathcal{T}$, we also define $\mathcal{E}_{K,\rm int}$ the set of its faces which belong to $\mathcal{E}_{\rm int}$. For such a face $\sigma \in \mathcal{E}_{K,\rm int}$, we may write $\sigma = K|L$, meaning that $\sigma = \overline{K} \cap \overline{L}$.

Given $\sigma \in \mathcal{E}$, we let

$$d_\sigma = \begin{cases} |\boldsymbol{x}_K - \boldsymbol{x}_L| & \text{if } \sigma = K|L \in \mathcal{E}_{\rm int}, \\ |\boldsymbol{x}_K - \boldsymbol{x}_\sigma| & \text{if } \sigma \in \mathcal{E}_{\rm ext}, \end{cases} \qquad \text{and} \quad \tau_\sigma = \frac{m_\sigma}{d_\sigma}.$$

We finally introduce the size $h_\mathcal{T}$ and the regularity $\zeta_\mathcal{T}$ (which is assumed to be positive) of a discretization $(\mathcal{T}, \mathcal{E}, (\boldsymbol{x}_K)_{K\in\mathcal{T}})$ of Ω by setting

$$h_\mathcal{T} = \max_{K\in\mathcal{T}} \operatorname{diam}(K), \qquad \zeta_\mathcal{T} = \min_{K\in\mathcal{T}} \min_{\sigma\in\mathcal{E}_K} \frac{d(x_K, \sigma)}{d_\sigma}.$$

Concerning the time discretization of $(0, T)$, we consider an increasing finite family of times $0 = t_0 < t_1 < \ldots, < t_N = T$. We denote by $\Delta t_n = t_n - t_{n-1}$ for $1 \le n \le N$, by $\boldsymbol{\Delta t} = (\Delta t_n)_{1\le n\le N}$, and by $\overline{\boldsymbol{\Delta t}} = \max_{1\le n\le N} \Delta t_n$.

sec:scheme

2.2. **A common basis for the Finite Volume schemes.** All the numerical schemes studied in this paper are based on TPFA Finite Volumes. The initial data c_0 is discretized into $(c_K^0)_{K\in\mathcal{T}} \in \mathbb{R}^\mathcal{T}$ by setting

$$c_K^0 = \frac{1}{m_K} \int_K c^0(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \qquad \forall K \in \mathcal{T}, \tag{2.1}$$

{eq:cK0}

while the doping profile $c^{\rm dop}$ is discretized into $\left(c_K^{\rm dop}\right)_{K\in\mathcal{T}} \in \mathbb{R}^\mathcal{T}$ by

$$c_K^{\rm dop} = \frac{1}{m_K} \int_K c^{\rm dop}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \qquad \forall K \in \mathcal{T}. \tag{2.2}$$

{eq:cKdp}

Assume that $\boldsymbol{c}^{n-1} = \left(c_K^{n-1}\right)_{K\in\mathcal{T}}$ is given for some $n > 0$, then we have to define how to compute $(\boldsymbol{c}^n, \boldsymbol{\Phi}^n) = (c_K^n, \Phi_K^n)_{K\in\mathcal{T}}$.

First, we introduce some notations. For all $K \in \mathcal{T}$ and all $\sigma \in \mathcal{E}_K$, we define the mirror values $c_{K\sigma}^n$ and $\Phi_{K\sigma}^n$ of c_K^n and Φ_K^n respectively across σ by setting

$$c_{K\sigma}^n = \begin{cases} c_L^n & \text{if } \sigma = K|L \in \mathcal{E}_{\rm int}, \\ c_K^n & \text{if } \sigma \in \mathcal{E}_{\rm ext}, \end{cases} \quad \Phi_{K\sigma}^n = \begin{cases} \Phi_L^n & \text{if } \sigma = K|L \in \mathcal{E}_{\rm int}, \\ \Phi_K^n & \text{if } \sigma \in \mathcal{E}^N, \\ \Phi_\sigma^n = \frac{1}{m_\sigma}\int_\sigma \Phi^D \mathrm{d}\boldsymbol{\gamma} & \text{if } \sigma \in \mathcal{E}^D. \end{cases} \tag{2.3}$$

{eq:mirror

Given $\boldsymbol{u} = (u_K)_{K\in\mathcal{T}} \in \mathbb{R}^\mathcal{T}$, we define the oriented and absolute jumps of $\boldsymbol{u}$ across any edge by

$$D_{K\sigma}\boldsymbol{u} = u_{K\sigma} - u_K, \quad D_\sigma \boldsymbol{u} = |D_{K\sigma}\boldsymbol{u}|, \qquad \forall K \in \mathcal{T},\ \forall \sigma \in \mathcal{E}_K.$$

We consider a backward Euler scheme in time and a TPFA finite volume scheme in space. It is written as follows:

scheme

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D_{K\sigma}\boldsymbol{\Phi}^n = m_K \left(c_K^n + c_K^{\rm dop}\right), \qquad \forall K \in \mathcal{T}, \tag{2.4a}$$

{eq:scheme

$$m_K \frac{c_K^n - c_K^{n-1}}{\Delta t_n} + \sum_{\sigma\in\mathcal{E}_{K,\rm int}} F_{K\sigma}^n = 0, \qquad \forall K \in \mathcal{T}, \tag{2.4b}$$

{eq:scheme

where $F_{K\sigma}^n$ should be a conservative and consistent approximation of $\frac{1}{\Delta t_n}\int_{t_{n-1}}^{t_n}\int_\sigma \mathbf{J}\cdot\boldsymbol{n}_{K\sigma}$ ($\boldsymbol{n}_{K\sigma}$ denotes the normal to σ outward K). The explicit formulas relating the numerical fluxes $F_{K\sigma}^n$ to the primary unknowns are now the only remaining degree of freedom. Four possible choices are given in the next section.

sec:fluxes

2.3. **Numerical fluxes for the conservation of the chemical species.** To close the system (2.4a)–(2.4b), it remains to define the numerical fluxes $F^n_{K\sigma}$.

Due to the no-flux boundary condition we only have to define the inner fluxes. They are defined with a function $\mathcal{F}$ of the primary unknowns $(c^n_K, c^n_L, \Phi^n_K, \Phi^n_L)$:

$$F^n_{K\sigma} = \tau_\sigma \mathcal{F}(c^n_K, c^n_L, \Phi^n_K, \Phi^n_L), \quad \forall K \in \mathcal{T}, \forall \sigma = K|L. \tag{2.5}$$

{eq:fluxin

We discuss now four strategies that are based on the four expressions (1.2), (1.7), (1.8), and (1.9). They lead to different formulas for $\mathcal{F}$. Three of the discrete fluxes are extensions of the Scharfetter-Gummel scheme [47] and let the Bernoulli function $B(u) = \frac{u}{e^u - 1}$, with $B(0) = 1$, appear in their definition.

All the functions $\mathcal{F}$ defined below verify

$$\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) = -\mathcal{F}(c_L, c_K, \Phi_L, \Phi_K) \quad \forall (c_K, c_L, \Phi_K, \Phi_L) \in (0,1) \times (0,1) \times \mathbb{R} \times \mathbb{R},$$

so that the numerical fluxes are locally conservative, which means

$$F^n_{K\sigma} + F^n_{L\sigma} = 0 \quad \forall \sigma = K|L \in \mathcal{E}_{\rm int}. \tag{2.6}$$

{eq:cons_F

ec:centred

2.3.1. *The centred flux.* The so-called centred flux is derived from formula (1.2), which suggests the following definition of $\mathcal{F}$:

$$\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) = -\frac{c_K + c_L}{2} D_{K\sigma}\left(h(\boldsymbol{c}) + \boldsymbol{\Phi}\right). \tag{C}$$

{eq:centre

The associate flux can be seen as a particular case in the TPFA context of the fluxes introduced in [12, 10, 9, 13] in various multipoint flux approximations (MPFA) or finite element contexts. In opposition to the three next schemes, the centred scheme is not based on the Scharfetter-Gummel scheme. We can notice that even if the relation (1.10) between the flux and the concentration were be linear (i.e., if $h(c) = \log(c)$ so that $r(c) = c$), $\mathcal{F}$ would be nonlinear with respect to c_K and c_L, and also singular near 0.

ssec:Sedan

2.3.2. *The Sedan flux.* The second flux we introduce is named Sedan after the eponymous code SEDAN III [51][1]. Formula (1.7) for the flux $\mathbf{J}$ suggests to use a classical Scharfetter-Gummel scheme, but for a modified potential $\Phi + \nu(c)$ instead of only Φ, leading to the following definition of $\mathcal{F}$:

$$\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) = B\Big(D_{K\sigma}(\boldsymbol{\Phi} + \nu(\boldsymbol{c}))\Big)c_K - B\Big(-D_{K\sigma}(\boldsymbol{\Phi} + \nu(\boldsymbol{c}))\Big)c_L. \tag{S}$$

{eq:SEDAN_

rem:Sedan

Remark 2.1. We notice that the Sedan flux defined by (S) satisfies

$$\mathcal{F}(c_K, c_L, \Phi, \Phi) = r(c_K) - r(c_L), \quad \forall (c_K, c_L) \in (0,1) \times (0,1), \ \forall \Phi \in \mathbb{R}.$$

It means that when $\mathbf{J} = -\nabla r(c)$, we recover the classical two-point flux approximation:

$$F^n_{K\sigma} = \tau_\sigma (r(c^n_K) - r(c^n_L)), \quad \forall K \in \mathcal{T}, \forall \sigma = K|L.$$

[1] The online reference contains a link to a tar file http://www-tcad.stanford.edu/oldftp_sw/Sedan-III/relB.8830.tar.Z containing among others the FORTRAN source `diffg.f` which in the lines 53-56 contains

```
if(ferm) then
  dpsin=dpsin+dlog(gamn(ip1)/gamn(i))
  dpsip=dpsip-dlog(gamp(ip1)/gamp(i))
endif
```

Here, `dpsin` and `dpsip` are the arguments of the Bernoulli function and `ferm` is the switch for enabling the degenerate case (Fermi statistics). To our knowledge this is the earliest reference to this scheme.

c:activity

2.3.3. *The activity based flux.* The activity based flux we discuss now is a restriction to our simplified model of the flux introduced in [29, 33]. It relies on the expression (1.8) of the flux $\mathbf{J}$. Assume that $a(c)$ and $\beta(c)$ are independent one from another (even though this is of course not true), then the flux $\mathbf{J}$ is linear w.r.t. $a(c)$, while $\beta(c)$ is a multiplicative factor. This suggests choosing a particular average for $\beta(c)$ —here the arithmetic mean— and applying the Scharfetter-Gummel scheme to approximate $-\nabla a(c) - a(c)\nabla\Phi$. This yields

$$\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) = \frac{\beta(c_K) + \beta(c_L)}{2}\Big\{B(D_{K\sigma}\mathbf{\Phi})a(c_K) - B(-D_{K\sigma}\mathbf{\Phi})a(c_L)\Big\}. \qquad \text{(AB)}$$

{eq:activi

c:Marianne

2.3.4. *The Bessemoulin-Chatard flux.* The last numerical flux we consider here is named Bessemoulin-Chatard flux after the author's name of [3]. Formula (1.9) for the flux $\mathbf{J}$ suggests that, up to the introduction of a variable diffusion coefficient approximating the quantity $r'(c)$ per face, one can use the Scharfetter-Gummel scheme. Following [3], the approximation $\mathfrak{d}r(c_K, c_L)$ of $r'(c)$ is defined as

$$\mathfrak{d}r(c_K, c_L) = \begin{cases} \dfrac{h(c_K) - h(c_L)}{\log(c_K) - \log(c_L)} & \text{if } c_K \neq c_L, \\ r'(c_K) & \text{if } c_K = c_L. \end{cases}$$

This leads to the following definition of $\mathcal{F}$:

$$\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) = \mathfrak{d}r(c_K, c_L)\left\{B\left(\frac{D_{K\sigma}\mathbf{\Phi}}{\mathfrak{d}r(c_K, c_L)}\right)c_K - B\left(-\frac{D_{K\sigma}\mathbf{\Phi}}{\mathfrak{d}r(c_K, c_L)}\right)c_L\right\}. \qquad \text{(BC)}$$

{eq:Marian

ssec:main

2.4. **Main results and organisation of the paper.** We have introduced four schemes defined by (2.1)–(2.5), supplemented with one of the four definitions of $\mathcal{F}$: (C), (S), (AB), or (BC). Besides numerical comparisons between the different approaches —this will be the purpose of Section 5—, we aim at proposing shared pieces of numerical analysis for all the schemes.

All the four schemes proposed above yield a nonlinear system to be solved at each time step. The first theorem proven is this paper concerns the existence of discrete solutions for a given mesh, and the preservation of the physical bounds: boundedness of the concentration between 0 and 1, decay of the energy. The discrete energy functional $E_\mathcal{T}$ is the discrete counterpart of the continuous energy functional E, defined by:

$$E_\mathcal{T}(\boldsymbol{c}^n, \mathbf{\Phi}^n) = \sum_{K\in\mathcal{T}} m_K H(c_K^n) + \frac{1}{2}\sum_{\sigma\in\mathcal{E}} \tau_\sigma \left(D_\sigma \mathbf{\Phi}^n\right)^2 - \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}^D\cap\mathcal{E}_K} \tau_\sigma \Phi_\sigma^D D_{K\sigma}\mathbf{\Phi}^n. \qquad (2.7)$$

{eq:E_T}

As stated in Theorem 2.1 below, the nonlinear system corresponding to each scheme admits a solution which preserves the physical bounds on the concentrations and the decay of the energy. The proof of Theorem 2.1 will be the purpose of Section 3.

thm:main1

Theorem 2.1. *Let $(\mathcal{T}, \mathcal{E}, (\boldsymbol{x}_K)_{K\in\mathcal{T}})$ be an admissible mesh and let $\boldsymbol{c}^0$ be defined by (2.1). Then, for all $1 \leq n \leq N$, the nonlinear system of equations (2.3)–(2.5), supplemented either with (C), (S), (AB), or (BC), has a solution $(\boldsymbol{c}^n, \mathbf{\Phi}^n) \in [0,1]^\mathcal{T} \times \mathbb{R}^\mathcal{T}$. Moreover, the solution to the scheme satisfies, for all $1 \leq n \leq N$,*

$$E_\mathcal{T}(\boldsymbol{c}^n, \mathbf{\Phi}^n) \leq E_\mathcal{T}(\boldsymbol{c}^{n-1}, \mathbf{\Phi}^{n-1}) \text{ and } 0 < c_K^n < 1, \quad \forall K \in \mathcal{T}.$$

Knowing a discrete solution to the scheme, $(\boldsymbol{c}^n, \mathbf{\Phi}^n)_{1\leq n\leq N}$, we can define an approximate solution $(c_{\mathcal{T},\boldsymbol{\Delta t}}, \Phi_{\mathcal{T},\boldsymbol{\Delta t}})$. It is the piecewise constant function defined almost everywhere by

$$c_{\mathcal{T},\boldsymbol{\Delta t}}(t, \boldsymbol{x}) = c_K^n, \quad \Phi_{\mathcal{T},\boldsymbol{\Delta t}}(t, \boldsymbol{x}) = \Phi_K^n \quad \text{if } (t, \boldsymbol{x}) \in (t_{n-1}, t_n] \times K.$$

This definition will be developed in Section 4 and supplemented by other reconstruction operators.

Let $(\mathcal{T}_m, \mathcal{E}_m, (\boldsymbol{x}_K)_{K\in\mathcal{T}_m})_{m\geq 1}$ be a sequence of admissible meshes in the sense of Definition 2 such that $h_{\mathcal{T}_m}, \overline{\boldsymbol{\Delta t}}_m \underset{m\to\infty}{\longrightarrow} 0$ while the mesh regularity remains bounded, i.e., $\zeta_{\mathcal{T}_m} \geq \zeta^\star$ for some $\zeta^\star > 0$ not depending on m. A natural question is the convergence of the associated sequence of approximate solutions $(c_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}, \Phi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m})_{m\geq 1}$ towards a weak solution to the continuous problem. The convergence result is stated in Theorem 2.2, only for the centred scheme and the Sedan scheme.

thm:main2

Theorem 2.2. *For the centred scheme (inner fluxes defined by (2.5) and (C)) and the Sedan scheme (inner fluxes defined by (2.5) and (S)), a sequence of approximate solutions $(c_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}, \Phi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m})_{m\geq 1}$ satisfies, up to a subsequence,*

$$c_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} \underset{m\to\infty}{\longrightarrow} c \quad \text{a.e. in } Q_T, \qquad \Phi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} \underset{m\to\infty}{\longrightarrow} \Phi \quad \text{in } L^2(Q_T), \tag{2.8}$$

eq:conv

where (c, Φ) is a weak solution to (1.1)-(1.6) in the sense of Definition 1.

The above theorem deserves some comments. First, the convergence proof carried out in what follows does not encompass the activity based scheme and the Bessemoulin-Chatard scheme for reasons that will appear clearly in the proof later on. This does of course not mean that these schemes do not converge, but only that our analysis does not cover them. Second, the topologies for which the convergence is claimed in (2.8) is suboptimal when compared to the results we prove in Section 4. However, we choose to keep the statement as simple as possible. The interested reader can refer to Section 4 to get finer results, including the convergence of approximate gradients to be defined later on.

Section 5 is then devoted to the comparison of the numerical results produced by the different schemes.

3. Numerical analysis for fixed meshes

:existence

In this section, one aims to show that each scheme admits at least one solution and that the physical bounds are preserved by the schemes. Our approach is based on a topological degree argument [43, 18] to be detailed in Section 3.3. It relies on a priori estimates to be stated in Section 3.2. Let us start by some preliminary properties of the different functions $\mathcal{F}$, defined either by (C), (S), (AB), or (BC), and some consequences for the inner numerical fluxes $F^n_{K\sigma}$.

ec:fluxes2

3.1. Face concentration and face dissipation.

For each flux $\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L)$, we want to define a face concentration $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)$ satisfying

$$\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) = \mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)\,(h(c_K) + \Phi_K - h(c_L) - \Phi_L).$$

Lemma 3.1 states that the face concentration functional $\mathcal{C}$ can be continuously defined on $(0,1)\times(0,1)\times\mathbb{R}\times\mathbb{R}$ and that it verifies some bounds. Let us also note that it clearly satisfies $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = \mathcal{C}(c_L, c_K, \Phi_L, \Phi_K)$.

lem:avg

Lemma 3.1. *For a flux $\mathcal{F}$ defined either by (C), (S), (AB) or (BC), the corresponding face concentration functional defined by*

$$\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = \frac{\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L)}{h(c_K) + \Phi_K - h(c_L) - \Phi_L} \tag{3.1}$$

eq:EC

if $h(c_K) + \Phi_K - h(c_L) - \Phi_L \neq 0$ can be extended by continuity on $(0,1)\times(0,1)\times\mathbb{R}\times\mathbb{R}$. Moreover, if $\mathcal{F}$ is defined by (AB), we have for all $(c_K, c_L, \Phi_K, \Phi_L) \in (0,1)\times(0,1)\times\mathbb{R}\times\mathbb{R}$,

$$\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) \geq \frac{\min(c_K, c_L)}{2} > 0. \tag{3.2}$$

eq:avgA

In the other cases, $\mathcal{C}$ verifies a stronger result: for all $(c_K, c_L, \Phi_K, \Phi_L) \in (0,1) \times (0,1) \times \mathbb{R} \times \mathbb{R}$,

$$\min(c_K, c_L) \leq \mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) \leq \max(c_K, c_L). \tag{3.3}$$ {eq:avg}

Proof. We first remark that, for the centred flux (C),

$$\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = \frac{c_K + c_L}{2}.$$

Therefore, $\mathcal{C}$ is well defined in $(0,1) \times (0,1) \times \mathbb{R} \times \mathbb{R}$ and it satisfies the bounds (3.3).

The proof is more intricate for the Sedan flux (S) and the Bessemoulin-Chatard flux (BC). It relies on an elementary property of the Bernoulli function, which writes:

$$B(\log(a) - \log(b))a - B(\log(b) - \log(a))b = 0, \qquad \forall (a,b) \in (0,1)^2. \tag{3.4}$$ {eq:logBer

Let us consider first the Bessemoulin-Chatard flux (BC). Applying (3.4) with $a = c_K$ and $b = c_L$, we obtain, with $x = \log(c_K/c_L)$ and $y = (\Phi_L - \Phi_K)/\mathfrak{d}r(c_K, c_L)$,

$$\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) = \mathfrak{d}r(c_K, c_L)\Big((B(y) - B(x))c_K + (B(-y) - B(-x))c_L\Big).$$

But, we also notice that

$$\mathfrak{d}r(c_K, c_L)(x - y) = h(c_K) + \Phi_K - h(c_L) - \Phi_L,$$

so that (3.1) yields

$$\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = \frac{B(y) - B(x)}{x - y} c_K + \frac{B(-x) - B(-y)}{x - y} c_L \tag{3.5}$$ {eq:B_avg}

if $h(c_K) + \Phi_K - h(c_L) - \Phi_L \neq 0$, which means $x - y \neq 0$. First, we remark that this definition can be extended if $x - y \to 0$, so that $\mathcal{C}$ is defined in $(0,1) \times (0,1) \times \mathbb{R} \times \mathbb{R}$. Then, as the Bernoulli function is decreasing and satisfies $B(x) - B(-x) = -x$ for all $x \in \mathbb{R}$, which implies

$$\frac{B(y) - B(x)}{x - y} + \frac{B(-x) - B(-y)}{x - y} = 1,$$

we obtain that $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)$ is a convex combination of c_K and c_L. Therefore, (3.3) holds for the Bessemoulin-Chatard flux.

The proof is similar for the Sedan flux (S). Indeed, we still establish (3.5), but with $x = \log(c_K/c_L)$ and $y = \Phi_L + \nu(c_L) - \Phi_K - \nu(c_K)$, so that $x - y = h(c_K) + \Phi_K - h(c_L) - \Phi_L$. Here again, $\mathcal{C}$ is well defined in $(0,1) \times (0,1) \times \mathbb{R} \times \mathbb{R}$ and $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)$ is a convex combination of c_K and c_L, so that (3.3) holds for the Sedan flux.

The fact that (3.3) does not hold for the activity based flux (AB) is illustrated on Figure 2. Nevertheless, one can express the corresponding face concentration under the form

$$\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = \frac{\beta(c_K) + \beta(c_L)}{2} \times \left(\frac{B(y) - B(x)}{x - y} a(c_K) + \frac{B(-x) - B(-y)}{x - y} a(c_L) \right),$$

with $x = \log(a(c_K)) - \log(a(c_L))$ and $y = \Phi_L - \Phi_K$. Therefore, $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)$ is the product of the arithmetic mean of the positive quantities $\beta(c_K)$ and $\beta(c_L)$ with a convex combination of the positive quantities $a(c_K)$ and $a(c_L)$. As a is increasing, this convex combination is bounded by below by $a(\min(c_K, c_L))$. Using the identity $\beta(c)a(c) = c$, we get (3.2). □

Using this result we define one face concentration by internal face and by choice of flux:

$$\mathcal{C}_\sigma^n = \mathcal{C}\left(c_K^n, c_L^n, \Phi_K^n, \Phi_L^n\right) \quad \forall \sigma \in \mathcal{E}_{\text{int}}, \sigma = K|L. \tag{3.6}$$ {eq:Csig}

Each flux $F_{K\sigma}^n$ can be rewritten as

$$F_{K\sigma}^n = -\tau_\sigma \mathcal{C}_\sigma^n D_{K\sigma}(h(\boldsymbol{c}^n) + \boldsymbol{\Phi}^n), \quad \forall K \in \mathcal{T}, \forall \sigma = K|L. \tag{3.7}$$ {eq:FKsig_

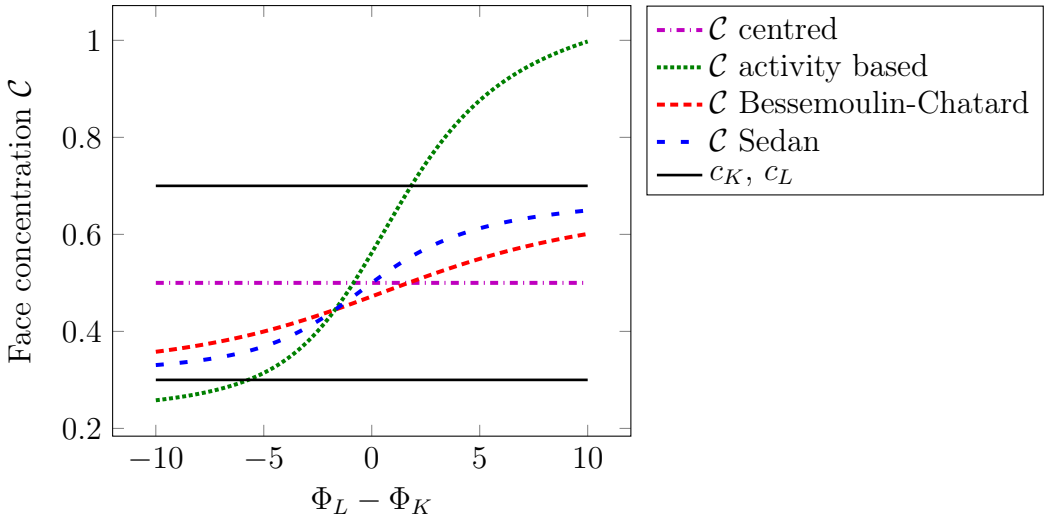

FIGURE 2. Evolution of the face concentration $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)$ as a function of the jump of the potential $\Phi_L - \Phi_K$ for the choice $c_K = 0.3$ and $c_L = 0.7$. fig:EC

We also introduce a face dissipation functional $\mathcal{D} : (0,1) \times (0,1) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, defined by

$$\mathcal{D}(c_K, c_L, \Phi_K, \Phi_L) = \mathcal{C}\left(c_K, c_L, \Phi_K, \Phi_L\right) \left|h(c_K) - h(c_L) + \Phi_K - \Phi_L\right|^2. \tag{3.8}$$ {eq:Dj}

We set, for each scheme:

$$\mathcal{D}_\sigma^n = \mathcal{D}(c_K^n, c_L^n, \Phi_K^n, \Phi_L^n), \quad \forall \sigma \in \mathcal{E}_{\rm int}, \sigma = K|L. \tag{3.9}$$ {eq:Dsig}

For $\delta \in (0,1)$ and $M \in \mathbb{R}$, we finally define two functions associated to $\mathcal{D}$, $\Psi_{\delta,M} : (0,1) \to \mathbb{R}$ and $\Upsilon_{\delta,M} : (0,1) \to \mathbb{R}$, by

$$\begin{aligned} \Psi_{\delta,M}(c_L) &= \inf\{\mathcal{D}(c_K, c_L, \Phi_K, \Phi_L); c_K \in [\delta, 1), (\Phi_K, \Phi_L) \in [-M, M]^2\}, \\ \Upsilon_{\delta,M}(c_L) &= \inf\{\mathcal{D}(c_K, c_L, \Phi_K, \Phi_L); c_K \in (0, 1-\delta], (\Phi_K, \Phi_L) \in [-M, M]^2\}. \end{aligned} \tag{3.10}$$ {eq:defPsi

Note that $\delta \mapsto \Psi_{\delta,M}(c_L)$ and $\delta \mapsto \Upsilon_{\delta,M}(c_L)$ are nondecreasing for all $M \in \mathbb{R}$ and all $c_L \in (0,1)$.

As a by-product of Lemma 3.1, we obtain that the face dissipation $\mathcal{D}$ is a nonnegative function as the product of nonnegative quantities. Lemma 3.2 is about the coercivity of the face dissipation functional. As its proof is technical, it is given in Appendix B.

lem:dissip **Lemma 3.2.** *The face dissipation functional defined by* (3.8) *and either* (C), (S), (AB) *or* (BC) *satisfies the following dissipation property: given* $\delta \in (0,1)$ *and* $M \in \mathbb{R}$, *for* Ψ *and* Υ *as defined in* (3.10)*:*

$$\lim_{c_L \to 0} \Psi_{\delta,M}(c_L) = +\infty,$$
$$\lim_{c_L \to 1} \Upsilon_{\delta,M}(c_L) = +\infty.$$

ec:apriori 3.2. **Uniform a priori estimates.** In all this section, we assume that $(\boldsymbol{c}^n, \boldsymbol{\Phi}^n)_{1 \le n \le N}$ is a solution to the scheme (2.3)–(2.5) with a numerical flux defined among (C) , (S), (AB), and (BC). We also assume that this solution verifies: $0 < c_K^n < 1$ for all $K \in \mathcal{T}$ and all $1 \le n \le N$. Then the goal of this section is to derive enough a priori estimates on $(\boldsymbol{c}^n, \boldsymbol{\Phi}^n)_{1 \le n \le N}$ in order to show the existence of a weak solution to the nonlinear system induced by the scheme.

The first lemma is the discrete counterpart of the global conservation of mass.

lem:mass **Lemma 3.3.** *One has*

$$\sum_{K\in\mathcal{T}} m_K c_K^n = \sum_{K\in\mathcal{T}} m_K c_K^{n-1} = \int_\Omega c^0 \mathrm{d}\boldsymbol{x}, \quad \forall 1 \le n \le N.$$

Proof. The first equality is obtained by summing (2.4b) over $K \in \mathcal{T}$ and by using the conservativity of the fluxes (2.6). A straightforward induction ensures the second equality thanks to (2.1). □

The second a priori estimate is related to energy dissipation and can be seen as a discrete counterpart of Proposition 1.1.

rop:E_disc **Proposition 3.1.** *Let $\mathcal{D}_\sigma^n$ be defined by (3.9), and $E_\mathcal{T}$ by (2.7). One has:*

$$\frac{E_\mathcal{T}(\boldsymbol{c}^n,\boldsymbol{\Phi}^n) - E_\mathcal{T}(\boldsymbol{c}^{n-1},\boldsymbol{\Phi}^{n-1})}{\Delta t_n} \le -\sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma \mathcal{D}_\sigma^n \le 0, \quad \forall 1 \le n \le N.$$

Proof. Due to the convexity of H and of $x \mapsto x^2/2$, we have:

$$E_\mathcal{T}(\boldsymbol{c}^n,\boldsymbol{\Phi}^n) - E_\mathcal{T}(\boldsymbol{c}^{n-1},\boldsymbol{\Phi}^{n-1}) \le \sum_{K\in\mathcal{T}} m_K (c_K^n - c_K^{n-1}) h(c_K^n) + \sum_{\sigma\in\mathcal{E}} \tau_\sigma D_\sigma \boldsymbol{\Phi}^n D_\sigma(\boldsymbol{\Phi}^n - \boldsymbol{\Phi}^{n-1}) - \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}^D\cap\mathcal{E}_K} \tau_\sigma \Phi_\sigma^D D_{K\sigma}(\boldsymbol{\Phi}^n - \boldsymbol{\Phi}^{n-1}).$$

A discrete integration by parts permits to rewrite the sum of the last two terms, which, combined to the scheme (2.4a), leads to

$$E_\mathcal{T}(\boldsymbol{c}^n,\boldsymbol{\Phi}^n) - E_\mathcal{T}(\boldsymbol{c}^{n-1},\boldsymbol{\Phi}^{n-1}) \le \sum_{K\in\mathcal{T}} m_K (c_K^n - c_K^{n-1})(h(c_K^n) + \Phi_K^n). \tag{3.11}$$ {E:step1}

Multiplying the equation (2.4b) by $h(c_K^n) + \Phi_K^n$ and summing over $K \in \mathcal{T}$, we obtain that

$$\sum_{K\in\mathcal{T}} m_K \frac{c_K^n - c_K^{n-1}}{\Delta t_n}(h(c_K^n) + \Phi_K^n) = -\sum_{K\in\mathcal{T}} (h(c_K^n) + \Phi_K^n) \sum_{\sigma\in\mathcal{E}_K} F_{K\sigma}^n = -\sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma \mathcal{C}_\sigma^n \left| D_\sigma\left(h(\boldsymbol{c}^n) + \boldsymbol{\Phi}^n\right)\right|^2 \tag{3.12}$$ {E:step2}

Combining (3.11) and (3.12) provides the desired estimate.

□

The third statement of this section is devoted to a uniform L^∞ estimate of $(\boldsymbol{\Phi}^n)_{1\le n\le N}$. It is a straightforward consequence of the slightly more general Proposition A.1 stated in appendix, together with the a priori bounds $0 < c_K^n < 1$ and $-\|c^{\rm dop}\|_\infty \le c_K^{\rm dop} \le \|c^{\rm dop}\|_\infty$.

em:LinfPhi **Lemma 3.4.** *There exists M_Φ depending only on Φ^D, $c^{\rm dop}$ and Ω such that*

$$|\Phi_K^n| \le M_\Phi, \qquad \forall K \in \mathcal{T},\ \forall 1 \le n \le N.$$

The next lemma concerns the discrete $L^\infty((0,T);H^1(\Omega))$ estimate on the electric potential and the control of the discrete dissipation.

:LinfH1Phi **Lemma 3.5.** *There exists C depending only on Φ^D, $c^{\rm dop}$, Ω and $\zeta_{\mathcal{T}}$, and C' depending also on c^0 such that:*

$$\sum_{\sigma\in\mathcal{E}} \tau_\sigma |D_\sigma \mathbf{\Phi}^n|^2 \le C, \qquad \forall 1 \le n \le N; \tag{3.13}$$ {est:H1dis

$$\left| \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}^D\cap\mathcal{E}_K} \tau_\sigma \Phi_\sigma^D D_{K\sigma}\mathbf{\Phi}^n \right| \le C, \qquad \forall 1 \le n \le N; \tag{3.14}$$ {est:Phi2}

$$\sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma \mathcal{D}_\sigma^n \le C'. \tag{3.15}$$ {est:diss}

Proof. As $\Phi^D \in L^\infty \cap H^1(\Omega)$, it is discretized into $\mathbf{\Phi}^D \in \mathbb{R}^{\mathcal{T}}$ by setting

$$\Phi_K^D = \frac{1}{m_K}\int_K \Phi^D {\rm d}\boldsymbol{x}, \quad \forall K\in\mathcal{T}, \quad \text{and } \Phi_\sigma^D = \frac{1}{m_\sigma}\int_\sigma \Phi^D {\rm d}\boldsymbol{\gamma}, \qquad \forall \sigma\in\mathcal{E}^D.$$

It satisfies $|\Phi_K^D| \le \|\Phi^D\|_\infty$ for all $K\in\mathcal{T}$. Multiplying (2.4a) by $\Phi_K^n - \Phi_K^D$ and summing over $K\in\mathcal{T}$ provides

$$\sum_{\sigma\in\mathcal{E}} \tau_\sigma D_{K\sigma}\mathbf{\Phi}^n D_{K\sigma}(\mathbf{\Phi}^n - \mathbf{\Phi}^D) = \sum_{K\in\mathcal{T}} m_K (c_K^n + c_K^{\rm dop})(\Phi_K^n - \Phi_K^D). \tag{3.16}$$ {eq:AB}

Using the elementary inequality $a(a-b) \ge \frac{a^2-b^2}{2}$, we get that

$$\sum_{\sigma\in\mathcal{E}} \tau_\sigma D_{K\sigma}\mathbf{\Phi}^n D_{K\sigma}(\mathbf{\Phi}^n - \mathbf{\Phi}^D) \ge \frac12 \sum_{\sigma\in\mathcal{E}} \tau_\sigma (D_\sigma \mathbf{\Phi}^n)^2 - \frac12 \sum_{\sigma\in\mathcal{E}} \tau_\sigma (D_\sigma \mathbf{\Phi}^D)^2.$$

Using the boundedness of c_K^n, $c_K^{\rm dop}$, Φ_K^D, and of Φ_K^n (cf. Lemma 3.4), we obtain that the right-hand side of (3.16) is bounded:

$$\sum_{K\in\mathcal{T}} m_K (c_K^n + c_K^{\rm dop})(\Phi_K^n - \Phi_K^D) \le C.$$

Following [27, Lemma 13.4], there exists C depending only on $\zeta_{\mathcal{T}}$ such that

$$\sum_{\sigma\in\mathcal{E}} \tau_\sigma (D_\sigma \mathbf{\Phi}^D)^2 \le C \|\Phi^D\|_{H^1(\Omega)}^2,$$

which concludes the proof of (3.13). Multiplying now the scheme (2.4a) by Φ_K^n and summing over $K\in\mathcal{T}$ leads to (3.14) by following the same kind of computations. Finally, these two inequality ensures that the functional $\boldsymbol{c}^n \mapsto E_{\mathcal{T}}(\boldsymbol{c}^n, \mathbf{\Phi}[\boldsymbol{c}^n])$ is bounded on $(0,1)^{\mathcal{T}}$. Therefore, Proposition 3.1 yields the control of the dissipation (3.15). □

As the last step before establishing the existence of a solution to the scheme, we show that the approximate concentrations $\boldsymbol{c}^n$ are bounded away from 0 and 1. Note that contrary to Lemmas 3.3 and 3.4 and to Proposition 3.1, the estimate of the following Lemma is not uniform with respect to mesh size and time step.

em:epsilon **Lemma 3.6.** *There exists $\epsilon > 0$ depending on $\mathcal{T}, \boldsymbol{\Delta t}$, Φ^D, $\bar c$, $c^{\rm dop}$ and Ω such that*

$$\epsilon < c_K^n < 1-\epsilon, \qquad \forall K\in\mathcal{T},\ \forall 1\le n\le N.$$

Proof. The proof follows the idea of [11, Lemma 3.10] (see also [12, Lemma 3.7]). Let us establish the lower bound only since the outline of the proof of the upper bound is similar.

Because of assumption (1.4) on the initial data and of the choice (2.1) for its discretization, one knows that

$$\frac{1}{m_\Omega}\sum_{K\in\mathcal{T}} m_K c_K^0 = \bar c \in (0,1).$$

Lemma 3.3 ensures the conservation of mass, so that

$$\frac{1}{m_\Omega}\sum_{K\in\mathcal{T}} m_K c_K^n = \overline{c} \in (0,1), \qquad \forall n \geq 1.$$

This implies that there exists $K_0 \in \mathcal{T}$ such that $c_{K_0}^n \geq \overline{c} > 0$. We set $\delta_0 = \overline{c}$.

Denote by $\boldsymbol{\Phi}[\boldsymbol{c}^n]$ the unique solution to the linear system (2.4a). The estimate (3.15) of Lemma 3.5 ensures that there exists $C_\mathcal{D}$ depending (among others) on Δt_n such that

$$\mathcal{D}_\mathcal{T}^n = \sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma \mathcal{D}_\sigma^n \leq C_\mathcal{D}, \qquad \forall 1 \leq n \leq N. \tag{3.17}$$ {eq:D_T2}

In particular, for all face $\sigma \in \mathcal{E}_{K_0}$, one gets that $\tau_\sigma \mathcal{D}_\sigma^n \leq C_\mathcal{D}$. Therefore, the concentration $c_{K_1}^n$ in any neighbouring cell K_1 of K_0 is bounded away from 0 by

$$\begin{aligned} c_{K_1}^n &\geq \inf\left\{ c_L \in (0,1) \ ; \ \Psi_{c_{K_0}^n, M_\Phi}(c_L) \leq C_\mathcal{D}/\tau_\sigma \right\} \\ &\geq \inf\left\{ c_L \in (0,1) \ ; \ \Psi_{\delta_0, M_\Phi}(c_L) \leq C_\mathcal{D}/\min_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma \right\} =: \delta_1 > 0 \end{aligned}$$

thanks to the monotonicity of $\delta \mapsto \Psi_{\delta,M}(c_L)$. Owing to Lemma 3.2, the above right-hand side is bounded away from 0 by some quantity that might also depend on $\mathcal{T}$ because of the presence of τ_σ. This lower bound can be set to δ_1, and we can then iterate the procedure to the neighbouring cells of K_1, and so on. Since the mesh is finite, only a finite number of iterations $I_\mathcal{T}$ is needed to cover all the cells, whence a uniform lower bound on c_K^n: $\epsilon = \min_{1\leq i\leq I_\mathcal{T}} \delta_i$, where

$$\delta_{i+1} = \inf\left\{ c_L \in (0,1) \ ; \ \Psi_{\delta_i, M_\Phi}(c_L) \leq C_\mathcal{D}/\min_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma \right\} > 0, \quad \delta_0 = \overline{c}.$$

□

:existence

3.3. **Existence of a solution to the schemes.** Based on the estimates derived in the previous section, we can establish the existence of at least one solution to each scheme. This completes the proof of Theorem 2.1.

:existence

Proposition 3.2. *Let $\boldsymbol{c}^0$ be defined by (2.1). Then, for all $1 \leq n \leq N$, the nonlinear system of equations (2.3)–(2.5), supplemented either with (C), (S), (AB), or (BC), has a solution $(\boldsymbol{c}^n, \boldsymbol{\Phi}^n) \in \mathbb{R}^\mathcal{T} \times \mathbb{R}^\mathcal{T}$.*

Proof. The proof is a proof by induction; it relies on a topological degree argument [43, 18] at each time step. The idea is to transform continuously our complex nonlinear system into a linear system while guaranteeing that a priori estimates controlling the solution remain valid all along the homotopy. We sketch the main ideas of the proof, making the homotopy (parametrized by $\lambda \in [0,3]$) explicit.

We denote by $\boldsymbol{c}^\star = \boldsymbol{c}^{n-1} \in (0,1)^\mathcal{T}$ the discrete concentration at the previous time step. We are interested in the existence of zeros for a functional

$$\mathcal{H} : \begin{cases} [0,3] \times (0,1)^\mathcal{T} \times \mathbb{R}^\mathcal{T} \to \mathbb{R}^\mathcal{T} \times \mathbb{R}^\mathcal{T} \\ (\lambda, \boldsymbol{c}, \boldsymbol{\Phi}) \mapsto \mathcal{H}(\lambda, \boldsymbol{c}, \boldsymbol{\Phi}) \end{cases}$$

that boils down to the scheme (2.4) when $\lambda = 3$. For sake of simplicity, instead of defining $\mathcal{H}$ for the different values of λ, we give a sense to the fact that $\boldsymbol{c}^{(\lambda)}, \boldsymbol{\Phi}^{(\lambda)}$ is solution to $\mathcal{H}(\lambda, \boldsymbol{c}^{(\lambda)}, \boldsymbol{\Phi}^{(\lambda)}) = 0$.

We start with $\lambda \in [0,1]$: $\boldsymbol{c}^{(\lambda)}$ is defined as the solution to the nonlinear system of equation

$$m_K \frac{c_K^{(\lambda)} - c_K^\star}{\Delta t_n} + (1-\lambda) \sum_{\sigma \in \mathcal{E}_{K,\mathrm{int}}} \tau_\sigma (c_K^{(\lambda)} - c_L^{(\lambda)}) + \lambda \sum_{\sigma \in \mathcal{E}_{K,\mathrm{int}}} \tau_\sigma \left(r(c_K^{(\lambda)}) - r(c_L^{(\lambda)}) \right) = 0, \quad (3.18)$$

{eq:lambda

while $\boldsymbol{\Phi}^{(\lambda)} = 0$. Let us remark that for $\lambda = 0$, it boils down to an invertible linear system of equations. Moreover, adapting the proof of Proposition 3.1 and using the property $(r(a) - r(b))(h(a) - h(b)) \geq (a-b)(h(a) - h(b))$ for all $(a,b) \in (0,1)^2$, we get:

$$E_{\mathcal{T}}(\boldsymbol{c}^{(\lambda)}, \mathbf{0}) - E_{\mathcal{T}}(\boldsymbol{c}^\star, \mathbf{0}) \leq -\Delta t_n \sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} \tau_\sigma (c_K^{(\lambda)} - c_L^{(\lambda)})(h(c_K^{(\lambda)}) - h(c_L^{(\lambda)})).$$

As the associated dissipation function defined by $\mathcal{D}(c_K, c_L) = (c_K - c_L)(h(c_K) - h(c_L))$ is clearly coercive in the sense of Lemma 3.2, we can deduce as in Lemma 3.6 the existence of $\epsilon_1 > 0$ such that $\epsilon_1 < c_K^{(\lambda)} < 1 - \epsilon_1$ for all $K \in \mathcal{T}$ and all $\lambda \in [0,1]$.

For $\lambda \in [1,2]$, one lets our system evolve from the monotone scheme corresponding to $\lambda = 1$ (which, due to Remark 2.1 corresponds to the Sedan scheme for the case without electrical potential) to the scheme with the expected numerical fluxes $F_{K\sigma}$. The electrical potential remains fixed to $\boldsymbol{\Phi}^{(\lambda)} = \mathbf{0}$, i.e.,

$$\begin{gathered} m_K \frac{c_K^{(\lambda)} - c_K^\star}{\Delta t_n} + (2-\lambda) \sum_{\sigma \in \mathcal{E}_{K,\mathrm{int}}} \tau_\sigma \left(r(c_K^{(\lambda)}) - r(c_L^{(\lambda)}) \right) + (\lambda - 1) \sum_{\sigma \in \mathcal{E}_{K,\mathrm{int}}} F_{K\sigma}^{(\lambda)} = 0, \\ F_{K\sigma}^{(\lambda)} = \tau_\sigma \mathcal{F}(c_K^{(\lambda)}, c_L^{(\lambda)}, 0, 0). \end{gathered} \quad (3.19)$$

{eq:lambda

with $\mathcal{F}$ defined either by (C), (S), (AB), or (BC). Thanks to Lemma 3.6, there exists $\epsilon_2 > 0$ such that $\epsilon_2 < c_K^{(\lambda)} < 1 - \epsilon_2$ for all $K \in \mathcal{T}$ and all $\lambda \in [1,2]$.

During the last step, $\lambda \in [2,3]$, we reactivate progressively the electrical potential while keeping equation (2.4b). Defining

$$\Phi_\sigma^{D,(\lambda)} = (\lambda - 2)\Phi_\sigma^D, \qquad \forall \sigma \in \mathcal{E}^D,$$

the solutions $(\boldsymbol{c}^{(\lambda)}, \boldsymbol{\Phi}^{(\lambda)})$ are defined, for all $\lambda \in [2,3]$ as the solution to the nonlinear system: for all $K \in \mathcal{T}$:

$$\begin{gathered} m_K \frac{c_K^{(\lambda)} - c_K^\star}{\Delta t_n} + \sum_{\sigma \in \mathcal{E}_{K,\mathrm{int}}} F_{K\sigma}^{(\lambda)} = 0, \text{ with } F_{K\sigma}^{(\lambda)} = \tau_\sigma \mathcal{F}\Big(c_K^{(\lambda)}, c_L^{(\lambda)}, (\lambda-2)\Phi_K^{(\lambda)}, (\lambda-2)\Phi_L^{(\lambda)} \Big), \\ - \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma D_{K\sigma} \boldsymbol{\Phi}^{(\lambda)} = (\lambda - 2) m_K (c_K^{(\lambda)} + c_K^{\mathrm{dop}}). \end{gathered}$$

Thanks to Proposition A.1, one has $\left| \Phi_K^{(\lambda)} \right| \leq M_\Phi$ for all $K \in \mathcal{T}$ and all $\lambda \in [2,3]$. Moreover, as in Lemma 3.6, we can establish the existence of $\epsilon_3 > 0$ such that $\epsilon_3 < c_K^{(\lambda)} < 1 - \epsilon_3$ for all $K \in \mathcal{T}$ and all $\lambda \in [2,3]$.

Finally, all along the homotopy parametrized by $\lambda \in [0,3]$, the solution $\left(\boldsymbol{c}^{(\lambda)}, \boldsymbol{\Phi}^{(\lambda)} \right)$ remains inside the compact subset $[\epsilon, 1-\epsilon]^{\mathcal{T}} \times [-M_\Phi - 1, M_\phi + 1]^{\mathcal{T}}$ with $\epsilon = \min(\epsilon_1, \epsilon_2, \epsilon_3)$. Thus, the topological degree corresponding to $\mathcal{H}(\lambda, \boldsymbol{c}, \boldsymbol{\Phi}) = \mathbf{0}$ and to the set $[\epsilon, 1-\epsilon]^{\mathcal{T}} \times [-M_\Phi - 1, M_\phi + 1]^{\mathcal{T}}$ is equal to one all along the homotopy and in particular for $\lambda = 3$. This ensures the existence of (at least) one solution to the scheme (2.4). □

onvergence

4. About the convergence towards a weak solution

The goal of this section is to prove Theorem 2.2, which states the convergence of the centred scheme (2.4), (C), and the Sedan scheme (2.4), (S), towards a weak solution to the continuous problem in the sense of Definition 1. Unfortunately, the proof we propose here neither applies to the activity based scheme (2.4), (AB), nor the Bessemoulin-Chatard scheme (2.4), (BC). This does not mean that these schemes do not converge. Indeed, numerical evidences provided in Section 5 seem to show that all the four schemes converge.

We consider here a sequence $\left(\mathcal{T}_m, \mathcal{E}_m, (\boldsymbol{x}_K)_{K\in\mathcal{T}_m}\right)_{m\geq 1}$ of admissible discretization with $h_{\mathcal{T}_m}, \overline{\boldsymbol{\Delta t}}_m$ tending to 0 as m tends to $+\infty$, while the regularity $\zeta_{\mathcal{T}_m}$ remains uniformly bounded from below by a positive constant $\zeta^\star$. Theorem 2.1 provides the existence of a family of discrete solutions $(\boldsymbol{c}_m, \boldsymbol{\Phi}_m)_m = \left((c_K^n)_{K\in\mathcal{T}_m, 1\leq n\leq N_m}, (\Phi_K^n)_{K\in\mathcal{T}_m, 1\leq n\leq N_m}\right)$. To prove Theorem 2.2, we first establish in Section 4.2 some compactness properties on the family of piecewise constant approximate solutions $(c_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}, \Phi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m})$ satisfied by the centred scheme and the Sedan scheme. Then we identify the limit as a weak solution in Section 4.3.

To enlighten the notations, we remove the subscript m as soon as it is not necessary for understanding.

econstruct

4.1. Reconstruction operators.

In order to carry out the analysis of convergence, we introduce some reconstruction operators following the methodology proposed in [24].

The operators $\pi_{\mathcal{T}} : \mathbb{R}^{\mathcal{T}} \to L^\infty(\Omega)$ and $\pi_{\mathcal{T},\boldsymbol{\Delta t}} : \mathbb{R}^{\mathcal{T}\times N} \to L^\infty((0,T)\times\Omega)$ are defined respectively by

$$\pi_{\mathcal{T}}\boldsymbol{u}(\boldsymbol{x}) = u_K \quad \text{if } \boldsymbol{x}\in K, \qquad \forall \boldsymbol{u} = (u_K)_{K\in\mathcal{T}},$$

and

$$\pi_{\mathcal{T},\boldsymbol{\Delta t}}\boldsymbol{u}(t,\boldsymbol{x}) = u_K^n \quad \text{if } (t,\boldsymbol{x})\in(t_{n-1},t_n]\times K, \qquad \forall \boldsymbol{u} = (u_K^n)_{K\in\mathcal{T}, 1\leq n\leq N}.$$

These operators allow passing from the discrete solution $(\boldsymbol{c}^n, \boldsymbol{\Phi}^n)_{1\leq n\leq N}$ to the approximate solution since

$$c_{\mathcal{T},\boldsymbol{\Delta t}} = \pi_{\mathcal{T},\boldsymbol{\Delta t}}\left(\boldsymbol{c}^n\right)_n, \qquad \Phi_{\mathcal{T},\boldsymbol{\Delta t}} = \pi_{\mathcal{T},\boldsymbol{\Delta t}}\left(\boldsymbol{\Phi}^n\right)_n.$$

To carry out the analysis, we further need to introduce approximate gradient reconstruction. Since the boundary conditions play a crucial role in the definition of the gradient, we need to enrich the discrete solution by face values $(c_\sigma^n)_{\sigma\in\mathcal{E}_{\rm ext}, 1\leq n\leq N}$ and $(\Phi_\sigma^n)_{\sigma\in\mathcal{E}_{\rm ext}, 1\leq n\leq N}$ defined by $c_\sigma^n = c_{K\sigma}^n$ and $\Phi_\sigma^n = \Phi_{K\sigma}^n$. With a slight abuse of notations, we still denote by $\boldsymbol{c}^n = ((c_K^n)_{K\in\mathcal{T}}, (c_\sigma^n)_{\sigma\in\mathcal{E}_{\rm ext}})$ and $\boldsymbol{\Phi}^n = ((\Phi_K^n)_{K\in\mathcal{T}}, (\Phi_\sigma^n)_{\sigma\in\mathcal{E}_{\rm ext}})$ the elements of $(0,1)^{\mathcal{T}\cup\mathcal{E}_{\rm ext}}$ and $\mathbb{R}^{\mathcal{T}\cup\mathcal{E}_{\rm ext}}$ containing both the cell values and the exterior faces values of the concentration and the potential respectively.

For $\sigma = K|L \in \mathcal{E}_{\rm int}$, we denote by Δ_σ the diamond cell corresponding to σ, that is the interior of the convex hull of $\sigma\cup\{\boldsymbol{x}_K, \boldsymbol{x}_L\}$. For $\sigma\in\mathcal{E}_{\rm ext}$, the diamond cell Δ_σ is defined as the interior of the convex hull of $\sigma\cup\{\boldsymbol{x}_K\}$. The approximate gradient $\nabla_{\mathcal{T}} : \mathbb{R}^{\mathcal{T}\cup\mathcal{E}_{\rm ext}} \to L^2(\Omega)^d$ we use in the analysis is merely weakly consistent (unless $d=1$) and takes its source in [15, 25]. It is piecewise constant on the diamond cells Δ_σ, and it is defined as follows:

$$\nabla_{\mathcal{T}}\boldsymbol{u}(\boldsymbol{x}) = -d\frac{D_{K\sigma}\boldsymbol{u}}{d_\sigma}\boldsymbol{n}_{K\sigma} \quad \text{if } \boldsymbol{x}\in\Delta_\sigma, \qquad \forall \boldsymbol{u}\in\mathbb{R}^{\mathcal{T}\cup\mathcal{E}_{\rm ext}}.$$

We also define $\nabla_{\mathcal{T},\boldsymbol{\Delta t}} : \mathbb{R}^{(\mathcal{T}\cup\mathcal{E}_{\rm ext})\times N} \to L^2(Q_T)^d$ by setting

$$\nabla_{\mathcal{T},\boldsymbol{\Delta t}}\boldsymbol{u}(t,\cdot) = \nabla_{\mathcal{T}}\boldsymbol{u}^n \quad \text{if } t\in(t_{n-1},t_n], \qquad \forall \boldsymbol{u} = (\boldsymbol{u}^n)_{1\leq n\leq N} \in \mathbb{R}^{(\mathcal{T}\cup\mathcal{E}_{\rm ext})\times N}.$$

Let us recall now some key properties to be used in the analysis. First, for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{\mathcal{T}\cup\mathcal{E}_{\rm ext}}$,

$$\sum_{\sigma\in\mathcal{E}} \tau_\sigma D_{K\sigma}\boldsymbol{u} D_{K\sigma}\boldsymbol{v} = \frac{1}{d}\int_\Omega \nabla_{\mathcal{T}}\boldsymbol{u}\cdot\nabla_{\mathcal{T}}\boldsymbol{v}\mathrm{d}\boldsymbol{x}.$$

This implies in particular that

$$\sum_{\sigma\in\mathcal{E}} \tau_\sigma |D_\sigma \boldsymbol{u}|^2 = \frac{1}{d}\int_\Omega |\nabla_{\mathcal{T}}\boldsymbol{u}|^2\mathrm{d}\boldsymbol{x}, \qquad \forall \boldsymbol{u}\in\mathbb{R}^{\mathcal{T}\cup\mathcal{E}_{\rm ext}}. \tag{4.1}$$ {eq:norm_L

ec:compact

4.2. **Compactness properties for the approximate concentration.** The goal here is to take advantage of the a priori estimates established in Section 3.2 to recover enough compactness for the sequences of approximate solutions.

em:L2H1_xi

Lemma 4.1. *Let $(\boldsymbol{c}_m, \boldsymbol{\Phi}_m)$ be the family of discrete solutions defined either by the centred scheme or by the Sedan scheme. There exists C depending only on Φ^D, Ω, $\zeta^\star$, c_0, $c^{\rm dop}$ and T, such that*

$$\iint_{Q_T} |\nabla_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} r(\boldsymbol{c}_m)|^2 + \left(\pi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} r(\boldsymbol{c}_m)\right)^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t \le C.$$

Proof. We get rid of the subscript m for the ease of reading. We will split the proof in two parts, first we focus on the proof of:

$$\iint_{Q_T} |\nabla_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c})|^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t \le C. \tag{4.2}$$ {eq:normL2

Thanks to (4.1), we have

$$\begin{aligned}\iint_{Q_T} |\nabla_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c})|^2 &= d\sum_{n=1}^N \Delta t_n \sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma |D_\sigma(r(\boldsymbol{c}^n))|^2,\\ &= d\sum_{n=1}^N \Delta t_n \sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma \left(\widetilde{\mathcal{C}}_\sigma^n\right)^2 |D_\sigma(h(\boldsymbol{c}^n))|^2,\end{aligned}$$

where we have defined the mean face concentrations $\left(\widetilde{\mathcal{C}}_\sigma^n\right)_{\sigma\in\mathcal{E}_{\rm int},1\le n\le N}$ by setting

$$\widetilde{\mathcal{C}}_\sigma^n = \frac{D_\sigma r(\boldsymbol{c}^n)}{D_\sigma h(\boldsymbol{c}^n)} \text{ if } c_K^n \neq c_L^n \quad \text{and} \quad \widetilde{\mathcal{C}}_\sigma^n = c_K^n \text{ otherwise}, \qquad \forall \sigma = K|L. \tag{4.3}$$ {eq:wtCsig

As noticed by (C.1), $\widetilde{\mathcal{C}}_\sigma^n$ is a mean value of c_K^n and c_L^n; so that $\widetilde{\mathcal{C}}_\sigma^n \in (0,1)$ for all $\sigma\in\mathcal{E}_{\rm int}$. Moreover, Lemma C.1 proved in Appendix C ensures that there exists $G>0$ such that

$$\frac{\widetilde{\mathcal{C}}_\sigma^n}{\mathcal{C}_\sigma^n} \le G, \qquad \forall \sigma\in\mathcal{E}_{\rm int},\ \forall n\in\{1,\dots,N\}. \tag{4.4}$$ {eq:alacon

Note that the above estimate may not hold for the Bessemoulin-Chatard scheme, as shown in Remark C.1. Then, thanks to the classical $(a+b)^2 \le 2a^2+2b^2$ inequality, we obtain

$$\begin{aligned}\iint_{Q_T} |\nabla_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c})|^2 \le& 2dG\sum_{n=1}^N \Delta t_n \sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma \mathcal{C}_\sigma^n |D_\sigma(h(\boldsymbol{c}^n)+\boldsymbol{\Phi}^n)|^2\\ &+ 2d\sum_{n=1}^N \Delta t_n \sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma |D_\sigma \boldsymbol{\Phi}^n|^2.\end{aligned}$$

Therefore, Lemma 3.5 yield the desired bound (4.2).

We now focus on the proof of:

$$\iint_{Q_T} \left(\pi_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c})\right)^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t \leq C. \tag{4.5}$$ {eq:normL2}

Noticing that for $c^* = \frac{1+\bar{c}}{2} > \bar{c}$:

$$r(c) \leq (r(c) - r(c^*))^+ + r(c^*),$$

we have, using $(a+b)^2 \leq 2(a^2+b^2)$:

$$\iint_{Q_T} |\pi_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c})|^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t \leq 2 \iint_{Q_T} |(\pi_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c}) - r(c^*))^+|^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t + 2r(c^*)^2 m(\Omega)T. \tag{4.6}$$ {eq:decomp}

Let $t \in [0,T]$ and $u = (\pi_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c}) - r(c^*))^+(t)$. We intend to show that we have a L^2 bound on u following ideas of [1, Appendix A.1]. As u is nonnegative, we have:

$$\int_\Omega |u-\bar{u}|^2 = \int_{u=0} \bar{u}^2 + \int_{\Omega\setminus\{u=0\}} |u-\bar{u}|^2 \geq m(\{u=0\})\bar{u}^2, \tag{4.7}$$ {eq:firstb}

where $\bar{u} = \fint_\Omega u$. Using Poincaré-Wirtinger inequality (see [4, Theorem 5] or [38, Theorem 2.1]), we have:

$$\int_\Omega |u-\bar{u}|^2 \leq \frac{C}{\zeta_{\mathcal{T}}} \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \tau_\sigma (D_\sigma u)^2. \tag{4.8}$$ {eq:Poinca}

If we had a lower bound on $m(\{u=0\})$, the equations (4.8) and (4.7) would yield an upper bound on $\bar{u}$. By definition of u and monotonicity of r, u is zero if and only if c is smaller than c^*. Using the monotonicity of integration and Lemma 3.3, we have:

$$c^*(m(\Omega) - m(\{u=0\})) = \int_{c>c^*} c^* \leq \int_\Omega \pi_{\mathcal{T},\boldsymbol{\Delta t}} \boldsymbol{c}(t) = m(\Omega)\bar{c}.$$

Hence, as $c^* = (1+\bar{c})/2$,

$$m(\Omega)\frac{1-\bar{c}}{1+\bar{c}} \leq m(\{u=0\}).$$

Finally, we have:

$$\int_\Omega u^2 \leq 2\left(\int_\Omega |u-\bar{u}|^2 + \bar{u}^2\right) \leq C \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \tau_\sigma (D_\sigma u)^2.$$

Using the definition of u, we have:

$$\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \tau_\sigma (D_\sigma u)^2 \leq \int_\Omega |\nabla_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c})(t)|^2.$$

Hence, integrating in time, and using (4.6):

$$\iint_{Q_T} |\pi_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c})|^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t \leq C \iint_{Q_T} |\nabla_{\mathcal{T},\boldsymbol{\Delta t}} r(\boldsymbol{c})(t)|^2 + C.$$

We then deduce (4.5) from (4.2). This concludes the proof of Lemma 4.1. □

op:compact **Proposition 4.1.** *Let $(\boldsymbol{c}_m, \boldsymbol{\Phi}_m)$ be the family of discrete solutions defined either by the centred scheme or by the Sedan scheme. In both cases, there exists $c \in L^\infty(Q_T;[0,1])$ with $r(c) \in L^2((0,T);H^1(\Omega))$ such that, up to a subsequence,*

$$\pi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} \boldsymbol{c}_m \underset{m\to\infty}{\longrightarrow} c \quad \textit{a.e. in } Q_T, \tag{4.9}$$ {eq:conv_a}

$$\nabla_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} r(\boldsymbol{c}_m) \underset{m\to\infty}{\longrightarrow} \nabla r(c) \quad \textit{weakly in } L^2(Q_T). \tag{4.10}$$ {eq:conv_g}

Remark 4.1. The limit c obtained in Proposition 4.1 could *a priori* depend on the chosen subsequence or be different for the centred scheme and the Sedan scheme. In Section 4.3, we will identify each limit as a weak solution to the initial problem.

Proof. Since $0 < \pi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}\boldsymbol{c}_m < 1$ for all $m \geq 1$, there exists $c \in L^\infty(Q_T;[0,1])$ such that $\pi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}\boldsymbol{c}_m$ tends to c in the $L^\infty(Q_T)$ weak-$\star$ sense. We still have to establish the almost everywhere convergence as well as the fact that $r(c)$ belongs to $L^2((0,T);H^1(\Omega))$. To this end, we make use of the black box [2, Theorem 3.9] which provides both the almost everywhere convergence and the identification of the limit of $\pi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} r(\boldsymbol{c}_m)$ as $r(c)$. We already have Lemma 4.1 at hand and c is bounded in L^∞, so that, owing to [2], it is sufficient to prove that there exists some C not depending on m such that, for all $\boldsymbol{\varphi}_m = (\varphi_K,\varphi_\sigma)_{K,\sigma} \in \mathbb{R}^{(\mathcal{T}_m\cup\mathcal{E}_{\mathrm{ext},m})\times N_m}$,

$$\left|\sum_n \Delta t_n \sum_{K\in\mathcal{T}_m} m_K \frac{c_K^n - c_K^{n-1}}{\Delta t_n}\varphi_K^n\right| \leq C\|\nabla_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}\boldsymbol{\varphi}_m\|_{L^\infty(Q_T)}.$$

We would then have, among other things, the desired convergence (4.9). Using (2.4b) and the writing (3.7) of the fluxes, we obtain that

$$\begin{aligned}\left|\sum_n \Delta t_n \sum_{K\in\mathcal{T}_m} m_K \frac{c_K^n - c_K^{n-1}}{\Delta t_n}\varphi_K^n\right| &= \left|\sum_n \Delta t_n \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \tau_\sigma \mathcal{C}_\sigma^n D_{K\sigma}(h(\boldsymbol{c}^n)+\boldsymbol{\Phi}^n) D_{K\sigma}\boldsymbol{\varphi}\right| \\ &\leq \left(\sum_n \Delta t_n \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \tau_\sigma \mathcal{C}_\sigma^n \left|D_\sigma(h(\boldsymbol{c}^n)+\boldsymbol{\Phi}^n)\right|^2\right)^{1/2} \\ &\quad\times\left(\sum_n \Delta t_n \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \tau_\sigma |D_\sigma\boldsymbol{\varphi}^n|^2\right)^{1/2} \\ &\leq C\|\nabla_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}\boldsymbol{\varphi}_m\|_{L^2(Q_T)} \leq C\|\nabla_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}\boldsymbol{\varphi}_m\|_{L^\infty(Q_T)},\end{aligned}$$

thanks to the boundedness of the dissipation in Lemma 3.5 which is a consequence of Proposition 3.1.

Since $\nabla_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} r(\boldsymbol{c}_m)$ is bounded in $L^2(Q_T)^d$, it converges weakly in $L^2(Q_T)^d$ towards some $\boldsymbol{U}$. The identification of $\boldsymbol{U}$ as $\nabla r(c)$ is classical (see for instance [15, Sec. 4], [25] or [22, Lemma 6.5]). □

We have two kinds of face values at hand: $\left(\mathcal{C}_\sigma^n\right)_{\sigma\in\mathcal{E}_{\mathrm{int}},1\leq n\leq N}$ and $\left(\widetilde{\mathcal{C}}_\sigma^n\right)_{\sigma\in\mathcal{E}_{\mathrm{int}},1\leq n\leq N}$ defined respectively by (3.6) and (4.3). Based on this, we can reconstruct two approximate concentration profiles $c_{\mathcal{E},\boldsymbol{\Delta t}}$ and $\widetilde{c}_{\mathcal{E},\boldsymbol{\Delta t}}$ that are piecewise constant on the diamond cells by setting

$$c_{\mathcal{E},\boldsymbol{\Delta t}}(t,\boldsymbol{x}) = \begin{cases}\mathcal{C}_\sigma^n & \text{if } (t,\boldsymbol{x})\in(t_{n-1},t_n]\times\Delta_\sigma, \quad \sigma\in\mathcal{E}_{\mathrm{int}},\\ c_K^n & \text{if } \boldsymbol{x}\in\Delta_\sigma, \quad \sigma\in\mathcal{E}_{\mathrm{ext}}\cap\mathcal{E}_K,\end{cases} \tag{4.11}$$

{eq:c_Ee}

and

$$\widetilde{c}_{\mathcal{E},\boldsymbol{\Delta t}}(t,\boldsymbol{x}) = \begin{cases}\widetilde{\mathcal{C}}_\sigma^n & \text{if } (t,\boldsymbol{x})\in(t_{n-1},t_n]\times\Delta_\sigma, \quad \sigma\in\mathcal{E}_{\mathrm{int}},\\ c_K^n & \text{if } \boldsymbol{x}\in\Delta_\sigma, \quad \sigma\in\mathcal{E}_{\mathrm{ext}}\cap\mathcal{E}_K.\end{cases} \tag{4.12}$$

{eq:wt_c_E

lem:c_Ee

Lemma 4.2. *For the centred scheme and the Sedan scheme, there holds*

$$c_{\mathcal{E}_m,\boldsymbol{\Delta t}_m} \underset{m\to\infty}{\longrightarrow} c \quad \text{in } L^p(Q_T) \text{ for all } p\in[1,\infty), \tag{4.13}$$

{eq:conv_c

$$\widetilde{c}_{\mathcal{E}_m,\boldsymbol{\Delta t}_m} \underset{m\to\infty}{\longrightarrow} c \quad \text{in } L^p(Q_T) \text{ for all } p\in[1,\infty), \tag{4.14}$$

{eq:conv_w

where c is as in Proposition 4.1.

Proof. We only prove (4.13) since the proof of (4.14) is similar. Here again, we get rid of m for clarity. Since $c_{\mathcal{T},\boldsymbol{\Delta t}}$ converges almost everywhere to c and remains bounded between 0 and 1, it converges in $L^p(Q_T)$. $c_{\mathcal{E},\boldsymbol{\Delta t}}$ is also uniformly bounded, hence it suffices to show that $\|c_{\mathcal{E},\boldsymbol{\Delta t}} - c_{\mathcal{T},\boldsymbol{\Delta t}}\|_{L^1(Q_T)}$ tends to 0. Denoting by $\Delta_{K\sigma}$ the half-diamond cell which is defined as the interior of the convex hull of $\sigma \cup \{\boldsymbol{x}_K\}$ for $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$, one has

$$\begin{aligned}\|c_{\mathcal{E},\boldsymbol{\Delta t}} - c_{\mathcal{T},\boldsymbol{\Delta t}}\|_{L^1(Q_T)} \leq & \sum_{n=1}^{N} \Delta t_n \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} m_{\Delta_{K\sigma}} |c_K^n - \mathcal{C}_\sigma^n| \\ \leq & \frac{h_\mathcal{T}}{d} \sum_{n=1}^{N} \Delta t_n \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} m_\sigma |c_K^n - \mathcal{C}_\sigma^n|,\end{aligned}$$

where we have used the geometric relation $m(\Delta_{K\sigma}) = \frac{1}{d} m_\sigma \mathrm{dist}(\boldsymbol{x}_K, \sigma) \leq \frac{h_\mathcal{T}}{d} m_\sigma$. For the internal faces, Lemma 3.1 (use (C.1) instead for $\widetilde{\mathcal{C}}_\sigma^n$) implies that

$$|c_K^n - \mathcal{C}_\sigma^n| + |c_L^n - \mathcal{C}_\sigma^n| = |c_K^n - c_L^n|, \qquad \forall \sigma = K|L.$$

Therefore, we obtain that

$$\begin{aligned}\|c_{\mathcal{E},\boldsymbol{\Delta t}} - c_{\mathcal{T},\boldsymbol{\Delta t}}\|_{L^1(Q_T)} \leq & \frac{h_\mathcal{T}}{d} \sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}_{\rm int}} m_\sigma D_\sigma \boldsymbol{c}^n \\ \leq & \frac{h_\mathcal{T}}{d} \left(\sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}_{\rm int}} m_\sigma d_\sigma \right)^{1/2} \left(\sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma |D_\sigma \boldsymbol{c}^n|^2 \right)^{1/2}.\end{aligned}$$

Since $|r(a) - r(b)| > |a - b|$ for all $a, b \in (0,1)$, we deduce from Lemma 4.1 that

$$\|c_{\mathcal{E},\boldsymbol{\Delta t}} - c_{\mathcal{T},\boldsymbol{\Delta t}}\|_{L^1(Q_T)} \leq C h_\mathcal{T}.$$

□

c:identify

4.3. **Convergence towards a weak solution.**

op:convPhi

Proposition 4.2. *Let c be as in Proposition 4.1 and let $\Phi \in L^\infty(Q_T) \cap L^\infty((0,T); H^1(\Omega))$ be the solution to the Poisson equation (1.3) with boundary conditions (1.6). Then, for the centred scheme and the Sedan scheme, there holds*

$$\pi_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} \boldsymbol{\Phi}_m \underset{m\to\infty}{\longrightarrow} \Phi \text{ in } L^2(Q_T) \text{ and in the } L^\infty(Q_T) \text{ weak-}\star \text{ sense}, \tag{4.15}$$

{eq:convPh

and

$$\nabla_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} \boldsymbol{\Phi}_m \underset{m\to\infty}{\longrightarrow} \nabla\Phi \quad \text{in the } L^\infty((0,T); L^2(\Omega)^d) \text{ weak-}\star \text{ sense}. \tag{4.16}$$

{eq:convPh

Proof. The existence of some $\Phi \in L^\infty(Q_T)$ such that (4.15) holds is a straightforward consequence of Lemma 3.4, whereas the existence of some $\boldsymbol{U} \in L^\infty((0,T); L^2(\Omega)^d)$ such that $\nabla_{\mathcal{T}_m,\boldsymbol{\Delta t}_m} \boldsymbol{\Phi}$ tends to $\boldsymbol{U}$ as m tends to ∞ follows from Lemma 3.5 together with (4.1). For the proof of the identification $\boldsymbol{U} = \nabla\Phi$, we refer to [15, 25, 22].

We show now that Φ satisfies the Poisson equation (1.3). Let $\psi \in C_c^\infty([0,T] \times \{\Omega \cup \Gamma^N\})$, then define $\psi_K^n = \psi(\boldsymbol{x}_K, t_n)$ and $\psi_\sigma^n = \psi(\boldsymbol{x}_\sigma, t_n)$ for $1 \leq n \leq N$, $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_{\rm ext}$. Following [23] (see [17] for a practical example), one can reconstruct a second approximate gradient operator $\widehat{\nabla}_\mathcal{T} : \mathbb{R}^\mathcal{T} \to L^\infty(\Omega)^d$ such that

$$\int_\Omega \nabla_\mathcal{T} \boldsymbol{u} \cdot \widehat{\nabla}_\mathcal{T} \boldsymbol{v} \mathrm{d}\boldsymbol{x} = \sum_{\sigma\in\mathcal{E}} \tau_\sigma D_{K\sigma} \boldsymbol{u} D_{K\sigma} \boldsymbol{v}, \qquad \forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^\mathcal{T},$$

and which is strongly consistent, i.e.,

$$\widehat{\nabla}_\mathcal{T} \boldsymbol{\psi}^n \underset{h_\mathcal{T}\to 0}{\longrightarrow} \nabla\psi(\cdot, t_n) \text{ uniformly in } \overline{\Omega}, \quad \forall n \in \{1, \ldots, N\}, \tag{4.17}$$

{eq:ov_gra

thanks to the smoothness of ψ. The scheme (2.4a) then reduces to

$$\int_\Omega \nabla_{\mathcal{T}} \boldsymbol{\Phi}^n \cdot \widehat{\nabla}_{\mathcal{T}} \boldsymbol{\psi}^n \mathrm{d}\boldsymbol{x} = \int_\Omega \pi_{\mathcal{T}}(\boldsymbol{c}^n + c^{\mathrm{dop}}) \pi_{\mathcal{T}} \boldsymbol{\psi}^n \mathrm{d}\boldsymbol{x}, \quad \forall n \in \{1, \dots, N\}, \ \forall \boldsymbol{\psi} \in \mathbb{R}^{(\mathcal{T} \cup \mathcal{E}_{\mathrm{ext}}) \times N}.$$

Integrating with respect to time over $(0,T)$ and passing to the limit $h_{\mathcal{T}}, \overline{\boldsymbol{\Delta t}} \to 0$ thanks to Proposition 4.1, (4.16) and (4.17) then yields

$$\iint_{Q_T} \nabla \Phi \cdot \nabla \psi \mathrm{d}\boldsymbol{x}\mathrm{d}t = \iint_{Q_T} (c + c^{\mathrm{dop}}) \psi \mathrm{d}\boldsymbol{x}\mathrm{d}t, \qquad \forall \psi \in C_c^\infty([0,T] \times \Omega \cup \Gamma^N).$$

In particular, (1.15) holds for almost every $t \in (0,T)$. Concerning the boundary conditions for Φ, the fact that $\Phi = \Phi^D$ on $(0,T) \times \Gamma^D$. can be proved for instance following the lines of [7, Section 4].

It remains to check that $\pi_{\mathcal{T}_m, \boldsymbol{\Delta t}_m} \boldsymbol{\Phi}_m$ strongly converges towards Φ in $L^2(Q_T)$. To this end, we make use of a discrete Aubin-Simon lemma [36] in the particular setting of [14, Lemma 9]. Since we have a discrete $L^\infty(H^1)$ estimate at hand thanks to Lemma 3.5, it suffices to show that there exists C not depending on m such that, for all $n \geq 1$ and all $\boldsymbol{\varphi} \in \mathbb{R}^{\mathcal{T}_m}$, we have

$$\sum_{K \in \mathcal{T}_m} m_K (\Phi_K^n - \Phi_K^{n-1}) \varphi_K = \int_\Omega \pi_{\mathcal{T}_m} \left(\boldsymbol{\Phi}^n - \boldsymbol{\Phi}^{n-1} \right) \pi_{\mathcal{T}_m} \boldsymbol{\varphi} \leq \Delta t_n C \| \pi_{\mathcal{T}_m} \boldsymbol{\varphi} \|_{L^2}. \tag{4.18}$$ {eq:L2Hm1}

By linearity of (2.4a) we have:

$$- \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma D_{K\sigma} \left(\boldsymbol{\Phi}^n - \boldsymbol{\Phi}^{n-1} \right) = m_K (c_K^n - c_K^{n-1}), \qquad \forall K \in \mathcal{T}_m.$$

Using (2.4b) and (3.7) there holds

$$\sum_{\sigma \in \mathcal{E}_K} \tau_\sigma D_{K\sigma} \left(\boldsymbol{\Phi}^n - \boldsymbol{\Phi}^{n-1} \right) = -\Delta t_n \sum_{\sigma \in \mathcal{E}_{K,\mathrm{int}}} \tau_\sigma \mathcal{C}_\sigma^n D_{K\sigma} (h(\boldsymbol{c}^n) + \boldsymbol{\Phi}^n), \qquad \forall K \in \mathcal{T}_m. \tag{4.19}$$ {eq:tsoint}

Let $\boldsymbol{\varphi} \in \mathbb{R}^{\mathcal{T}_m}$, then define $\boldsymbol{\psi} \in \mathbb{R}^{\mathcal{T}_m}$ as the solution of the linear system

$$- \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma D_{K\sigma} \boldsymbol{\psi} = m_K \varphi_K, \qquad \forall K \in \mathcal{T}_m, \tag{4.20}$$ {eq:def_ps

where we have set $\psi_{K\sigma} = \psi_L$ if $\sigma = K|L \in \mathcal{E}_{\mathrm{int}}$, and $\psi_{K\sigma} = 0$ if $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}$. Multiplying (4.19) by ψ_K, summing over $K \in \mathcal{T}_m$, performing discrete integration by parts and using (4.20) yields

$$\sum_{K \in \mathcal{T}_m} m_K (\Phi_K^n - \Phi_K^{n-1}) \varphi_K = \Delta t_n \sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} \tau_\sigma \mathcal{C}_\sigma^n D_{K\sigma} (h(\boldsymbol{c}^n) + \boldsymbol{\Phi}^n) D_{K\sigma} \boldsymbol{\psi}.$$

Using successively Cauchy Schwarz inequality and [27, Lemma 9.2] we get

$$\begin{aligned} \sum_{K \in \mathcal{T}_m} m_K (\Phi_K^n - \Phi_K^{n-1}) \varphi_K \leq & \Delta t_n \Big(\sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} \tau_\sigma \mathcal{C}_\sigma^n \big(D_\sigma (h(\boldsymbol{c}^n) + \boldsymbol{\Phi}^n) \big)^2 \Big)^{\frac{1}{2}} \Big(\sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} \tau_\sigma \big(D_{K\sigma} \boldsymbol{\psi} \big)^2 \Big)^{\frac{1}{2}} \\ \leq & C \Delta t_n \Big(\sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} \tau_\sigma \mathcal{C}_\sigma^n \big(D_\sigma (h(\boldsymbol{c}^n) + \boldsymbol{\Phi}^n) \big)^2 \Big)^{\frac{1}{2}} \| \pi_{\mathcal{T}_m} \boldsymbol{\varphi} \|_{L^2(\Omega)}. \end{aligned}$$

Then the control on the dissipation established in Proposition 3.1 allows to recover (4.18), hence the relative compactness in $L^2(Q_T)$ of $(\pi_{\mathcal{T}_m, \boldsymbol{\Delta t}_m} \boldsymbol{\Phi}_m)_m$. □

em:convPhi

Remark 4.2 (Enhanced convergence properties). The convergence described in Proposition 4.2 is suboptimal. One can establish the strong convergence of $\widehat{\nabla}_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}\boldsymbol{\Phi}_m$ towards $\nabla\Phi$, where the gradient reconstruction operator $\widehat{\nabla}_{\mathcal{T}_m,\boldsymbol{\Delta t}_m}$ is the extension to the time-space domain Q_T of the operator $\widehat{\nabla}_{\mathcal{T}_m}$ used in the proof of Proposition 4.2. We refer to [23] for details on these enhanced convergence properties.

rop:conv_c

Proposition 4.3. *Let c, Φ be as in Propositions 4.1 and 4.2, then the weak formulation (1.14) holds.*

Proof. Let $\varphi \in C_c^\infty([0,T)\times\overline{\Omega})$, then define $\varphi_K^n = \varphi(\boldsymbol{x}_K, t_n)$ for all $n \in \{0,\dots,N\}$ and $K \in \mathcal{T}$. Multiplying (2.4b) by $\Delta t_n \varphi_K^{n-1}$, then summing over $K \in \mathcal{T}$ and $n \in \{1,\dots,N\}$ and using expression (3.7) for the fluxes leads to

$$T_1 + T_2 + T_3 = 0, \tag{4.21}$$

{eq:T123}

where we have set

$$T_1 = \sum_{n=1}^{N}\sum_{K\in\mathcal{T}} m_K(c_K^n - c_K^{n-1})\varphi_K^{n-1},$$
$$T_2 = \sum_{n=1}^{N}\Delta t_n \sum_{\sigma\in\mathcal{E}} \tau_\sigma \mathcal{C}_\sigma^n D_{K\sigma}h(\boldsymbol{c}^n) D_{K\sigma}\boldsymbol{\varphi}^{n-1},$$
$$T_3 = \sum_{n=1}^{N}\Delta t_n \sum_{\sigma\in\mathcal{E}} \tau_\sigma \mathcal{C}_\sigma^n D_{K\sigma}\boldsymbol{\Phi}^n D_{K\sigma}\boldsymbol{\varphi}^{n-1}.$$

The term T_1 can be rewritten as

$$T_1 = \sum_{n=1}^{N}\Delta t_n \sum_{K\in\mathcal{T}} m_K c_K^n \frac{\varphi_K^{n-1} - \varphi_K^n}{\Delta t_n} - \sum_{K\in\mathcal{T}} m_K c_K^0 \varphi_K^0,$$

so that it follows from the convergence of $\pi_{\mathcal{T},\boldsymbol{\Delta t}}\boldsymbol{c}$ towards c and of $\pi_{\mathcal{T}}\boldsymbol{c}^0$ towards c^0 together with the regularity of φ that

$$T_1 \underset{m\to\infty}{\longrightarrow} -\iint_{Q_T} c\partial_t\varphi \mathrm{d}\boldsymbol{x}\mathrm{d}t - \int_\Omega c^0\varphi(0,\cdot)\mathrm{d}\boldsymbol{x}. \tag{4.22}$$

{eq:T1}

On the other hand, the term T_3 can be rewritten as

$$T_3 = \iint_{Q_T} c_{\mathcal{E},\boldsymbol{\Delta t}} \nabla_{\mathcal{T},\boldsymbol{\Delta t}}\boldsymbol{\Phi}\cdot\widehat{\nabla}_{\mathcal{T},\boldsymbol{\Delta t}}\boldsymbol{\varphi}\mathrm{d}\boldsymbol{x}\mathrm{d}t,$$

where $\widehat{\nabla}_{\mathcal{T},\boldsymbol{\Delta t}}$ is the strongly consistent gradient reconstruction operator introduced in the proof of Proposition 4.2 and in Remark 4.2. In particular, due to the smoothness of φ, $\widehat{\nabla}_{\mathcal{T},\boldsymbol{\Delta t}}\boldsymbol{\varphi}$ converges uniformly towards $\nabla\varphi$. Therefore, it follows from Lemma 4.2 and Proposition 4.2 that

$$T_3 \underset{m\to\infty}{\longrightarrow} \iint_{Q_T} c\nabla\Phi\cdot\nabla\varphi\mathrm{d}\boldsymbol{x}\mathrm{d}t. \tag{4.23}$$

{eq:T3}

Define the term

$$\widetilde{T}_2 = \sum_{n=1}^{N}\Delta t_n \sum_{\sigma\in\mathcal{E}} \tau_\sigma \widetilde{\mathcal{C}}_\sigma^n D_{K\sigma}h(\boldsymbol{c}^n) D_{K\sigma}\boldsymbol{\varphi}^{n-1} = \iint_{Q_T} \nabla_{\mathcal{T},\Delta t} r(\boldsymbol{c})\cdot\widehat{\nabla}_{\mathcal{T},\Delta t}\boldsymbol{\varphi}\mathrm{d}\boldsymbol{x}\mathrm{d}t,$$

then it follows from Proposition 4.1 that

$$\widetilde{T}_2 \underset{m\to\infty}{\longrightarrow} \iint_{Q_T} \nabla r(c)\cdot\nabla\varphi\mathrm{d}\boldsymbol{x}\mathrm{d}t.$$

Therefore, it only remains to show that $|T_2 - \widetilde{T}_2|$ tends to 0 to conclude the proof of Proposition 4.3. Thanks to the triangle and Cauchy-Schwarz inequalities, one has

$$|T_2 - \widetilde{T}_2| \leq \sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}} \tau_\sigma \left|\mathcal{C}_\sigma^n - \widetilde{\mathcal{C}}_\sigma^n\right| D_\sigma h(\boldsymbol{c}^n) D_\sigma \boldsymbol{\varphi}^{n-1}$$
$$\leq \left(\sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}} \tau_\sigma \mathcal{C}_\sigma^n |D_\sigma h(\boldsymbol{c}^n)|^2\right)^{1/2} \left(\sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}} \tau_\sigma \frac{(\mathcal{C}_\sigma^n - \widetilde{\mathcal{C}}_\sigma^n)^2}{\mathcal{C}_\sigma^n} |D_\sigma \boldsymbol{\varphi}^{n-1}|^2\right)^{1/2}.$$

The first term in the right-hand side is uniformly bounded thanks to Lemma 3.5, using the ideas of the proof of (4.2)). Thus our problem amounts to show that

$$\mathcal{R} := \sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}} \tau_\sigma \frac{(\mathcal{C}_\sigma^n - \widetilde{\mathcal{C}}_\sigma^n)^2}{\mathcal{C}_\sigma^n} |D_\sigma \boldsymbol{\varphi}^{n-1}|^2 \underset{m\to\infty}{\longrightarrow} 0. \tag{4.24}$$

{eq:Rto0}

Let us reformulate $\mathcal{R}$ as

$$\mathcal{R} := \sum_{n=1}^{N} \Delta t_n \sum_{\sigma\in\mathcal{E}} \tau_\sigma |\mathcal{C}_\sigma^n - \widetilde{\mathcal{C}}_\sigma^n| \left|1 - \frac{\widetilde{\mathcal{C}}_\sigma^n}{\mathcal{C}_\sigma^n}\right| |D_\sigma \boldsymbol{\varphi}^{n-1}|^2.$$

Thanks to (4.4), the quantity $\left|1 - \frac{\widetilde{\mathcal{C}}_\sigma^n}{\mathcal{C}_\sigma^n}\right|$ is uniformly bounded, whereas the regularity of φ implies that $D_\sigma \boldsymbol{\varphi}^{n-1} \leq \|\nabla\varphi\|_\infty d_\sigma$. Putting this in the above expression of $\mathcal{R}$, we obtain that

$$\mathcal{R} \leq C \|c_{\mathcal{E},\boldsymbol{\Delta t}} - \widetilde{c}_{\mathcal{E},\boldsymbol{\Delta t}}\|_{L^1(Q_T)} \underset{m\to\infty}{\longrightarrow} 0,$$

thanks to Lemma 4.2. □

Remark 4.3. The convergence proof of the scheme does not hold for the Bessemoulin-Chatard scheme because we lack compactness properties: Lemma C.1, yielding equation (4.4), is not satisfied for this scheme (see Remark C.1), which affects successively the proofs of Lemma 4.1, Proposition 4.1 and therefore Theorem 2.2.

The activity based scheme does not satisfy the bounds (3.3) of Lemma 3.1. This implies gaps in the proof of Lemma 4.2, and the convergence of the scheme stated in Theorem 2.2 is not established.

5. Numerical comparison of the schemes

c:numerics

The numerical examples [31] have been implemented in the Julia language [5] based on the package `VoronoiFVM.jl` [32] which realizes the implicit Euler Voronoi finite volume method for nonlinear diffusion-convection-reaction equations on simplicial grids. The resulting nonlinear systems of equations are solved using Newton's method with parameter embedding. An advantage of the implementation in Julia is the availability of `ForwardDiff.jl` [46], an automatic differentiation package. This package allows the assembly of analytical Jacobians based on a generic implementation of nonlinear parameter functions without the need to write source code for derivatives.

5.1. 1D time evolution and convergence test.

The first group of examples considers the problem as described by (1.1)-(1.3) in a one-dimensional domain with Dirichlet boundary conditions for Φ and homogeneous Neumann boundary conditions for c. We regard the time evolution from a zero potential Φ and constant concentrations c_0. In all examples, we assume a constant doping concentration $c^{\mathrm{dop}} = -\frac{1}{2}$. Calculations have been performed with subdivision of the domain $\Omega = (0, 50)$ into 100 control volumes. Time steps have been chosen in a geometric progression $t_i = t_1 * \delta^i$ with $\delta = 1.15$ and $t_1 = 10^{-4}$.

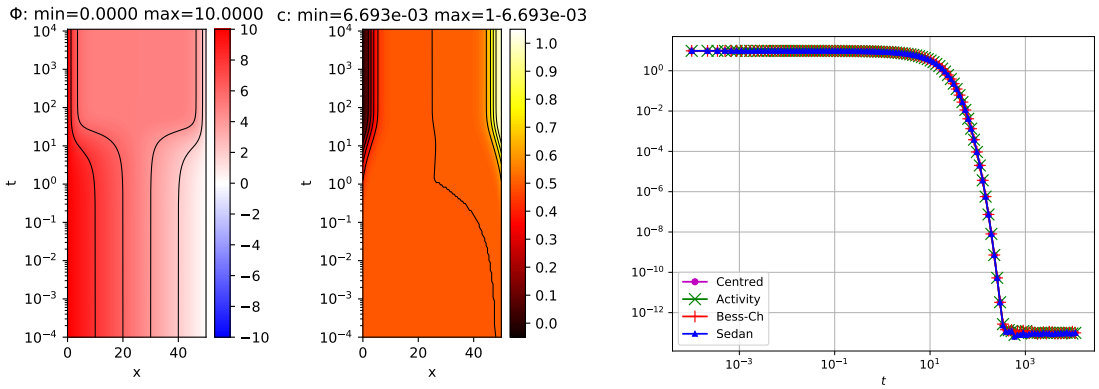

FIGURE 3. Left: time evolution of solution on domain $\Omega = (0, 50)$ with constant initial value $c = \frac{1}{2}$, Dirichlet boundary conditions $\Phi(0) = 10$, $\Phi(50) = 0$, $c^{\text{dop}} = -\frac{1}{2}$ and homogeneous Neumann boundary conditions for c. Right: Evolution of the relative free energy according to (1.11). fig:evoli

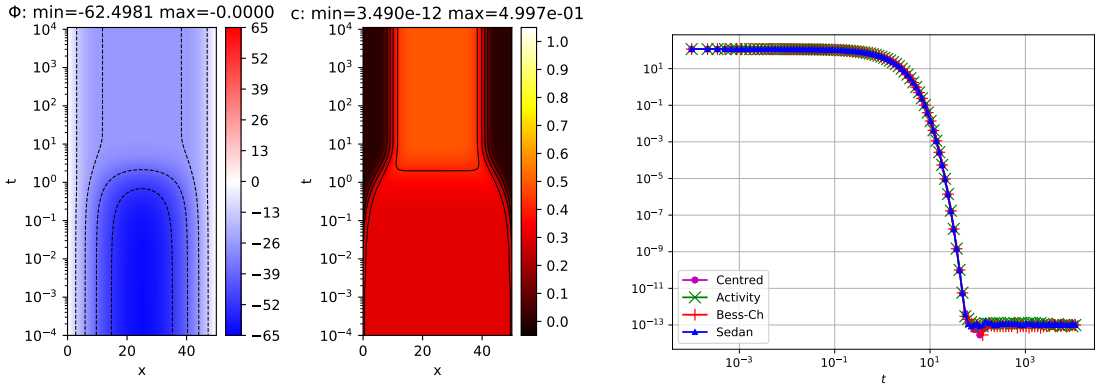

FIGURE 4. Left: time evolution of solution on domain $\Omega = (0, 50)$ with constant initial value $c = 0.3$, Dirichlet boundary conditions $\Phi(0) = 0$, $\Phi(50) = 0$, $c^{\text{dop}} = -\frac{1}{2}$ and homogeneous Neumann boundary conditions for c. Right: Evolution of the relative free energy according to (1.11). fig:evolii

In the first example (Fig. 3), $c_0 = 0.5$, and the initial amount of charge carriers exactly matches the amount of doping. With the start of time evolution, at $x = 0$ a potential of 10 is applied leading to a redistribution of the charge carrier concentration which for large t approaches a steady state with two space charge regions at the boundaries with opposite charge and an electroneutral region with $c = 0.5$ in the center of the domain. We remark that c stays in the range $(0, 1)$, and that the energy (1.11) decreases during time evolution for all four schemes discussed in this paper. We also remark that for zero applied potential, the constant values $\Phi = 0$ and $c = 0.5$ would comprise a solution for all $t > 0$.

Fig. 4 considers the case $c_0 = 0.3$. The available amount of charge carriers is not able to compensate for the amount of doping. At the end of the time evolution, the charge carriers are concentrated in the center of the domain, establishing an electroneutral region. At both boundaries, depletion boundary layers create equally charged space charge regions due to the lack of charge carriers able to compensate the doping.

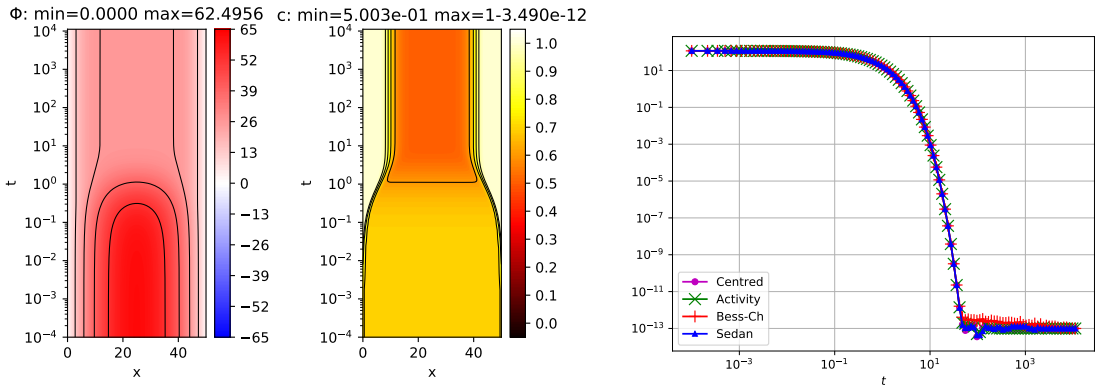

FIGURE 5. Left: time evolution of solution on domain $\Omega = (0, 50)$ with constant initial value $c = 0.7$, Dirichlet boundary conditions $\Phi(0) = 0$, $\Phi(50) = 0$, $c^{\text{dop}} = -\frac{1}{2}$ and homogeneous Neumann boundary conditions for c. Right: Evolution of the relative free energy according to (1.11). fig:evolii

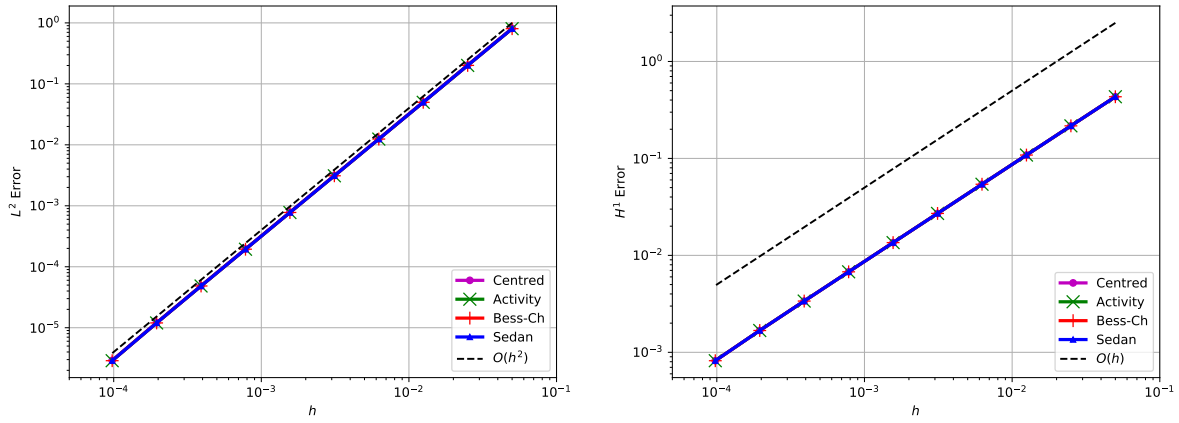

FIGURE 6. Convergence behaviour of the different schemes for the case depicted in Fig. 3: comparison of solutions at $t = 10$. Left: L^2-error, right: H^1 error. Correspondence to the equation in the paper: "centred": (C), "Sedan": (S), "Activity": (AB), "Bess-Ch": (BC). fig:scheme

Fig. 5 considers the case $c_0 = 0.7$ which in sense is symmetric to the previous one. There is again an electroneutral region in the center, and this time, "superfluous" charge carriers are forced to enrichment boundary layers.

Fig. 6 provides a comparison of the convergence behaviour for the test case discussed in Fig. 3. We compare the solutions at a moment of time where we observe a rather large descent of the relative free energy based on a reference solution obtained on a fine grid of 40960 nodes using scheme (S). No visible difference in the plot have been found when using one of the other schemes to obtain the reference solution.

We observe first order convergence in the H^1 norm and second order convergence in the L^2 norm. No significant difference between the results for the various schemes.

5.2. **1D stationary convergence test.** In order to reveal the behaviour of the various schemes under more extreme conditions, this convergence test outside of thermodynamic equilibrium includes regions of the solution with concentrations extremely close to 0 and

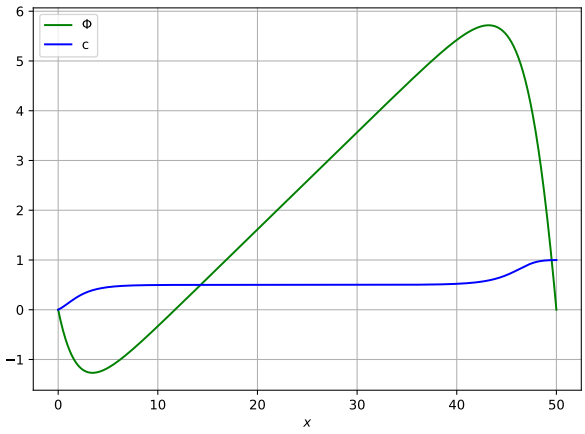

FIGURE 7. Stationary solution with Dirichlet boundary conditions for c and Φ. fig:refsol

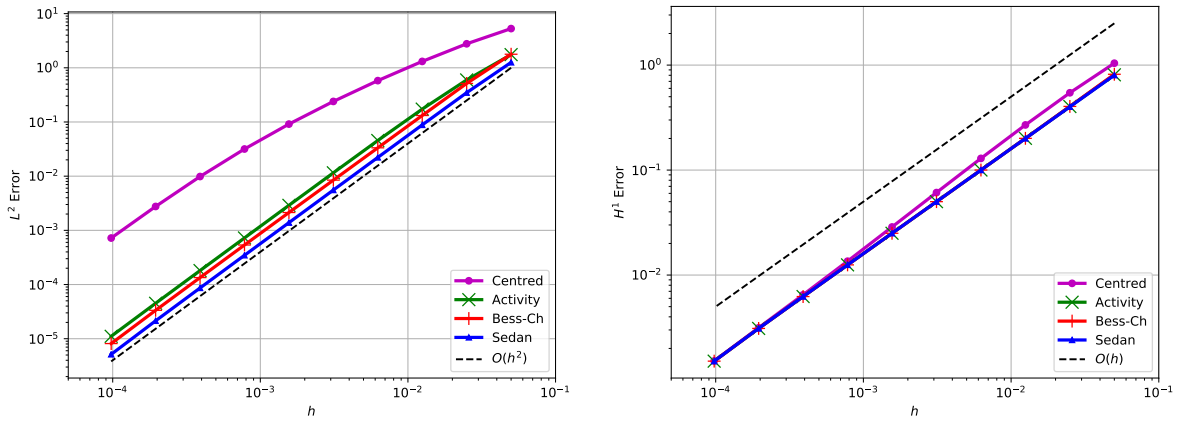

FIGURE 8. Convergence behaviour of the different schemes. Left: L^2-error, right: H^1 error. Correspondence to the equation in the paper: "centred": (C), "Sedan": (S), "Activity": (AB), "Bess-Ch": (BC). fig:scheme

1, respectively, enforced by inhomogeneous Dirichlet boundary conditions for the concentration, thus leaving the realm of the analysis in this paper. Once again, we assume $\Omega = (0, L)$ with $L = 50$, $c^{\mathrm{dop}} = -\frac{1}{2}$. We set boundary values $\Phi(0) = \Phi(L) = 0$ for the electrostatic potential, and $c(0) = 10^{-3}, c(L) = 1 - 10^{-3}$. We calculate a reference solution using the scheme (S) on a fine grid of 40960 nodes with grid spacing $h \simeq 1.22 \cdot 10^{-3}$, see Fig. 7. We use this solution as a surrogate for an analytical solution in a numerical investigation of the convergence rates of the different schemes. While no visible differences have been detected when using the schemes (AB) or (BC) for reference, for the slower converging scheme (C) as reference flux one would need a finer reference mesh to obtain similar results.

The result is shown in Fig. 8. We observe, that both in the H^1 and the L^2 norms, the schemes based on the modification of the Scharfetter-Gummel idea behave significantly better than the centred scheme. This is probably due to the Dirichlet boundary condition close to 0 where the function $c \mapsto h(c)$ appearing explicitly in the centred scheme is singular. Judging from the L^2 error plot in Fig. 8 (left), the scheme (S) converges better than all the others. Asymptotically, all schemes show the same standard behaviour: we

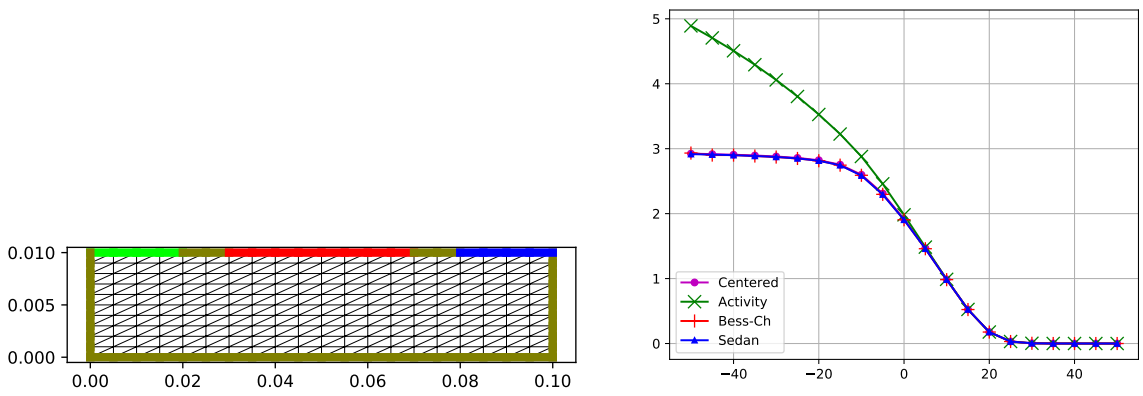

FIGURE 9. Discretization grid of refinement level $n_{ref} = 1$ (left) and corresponding I-V curves for different discretization schemes (right). fig:gridiv

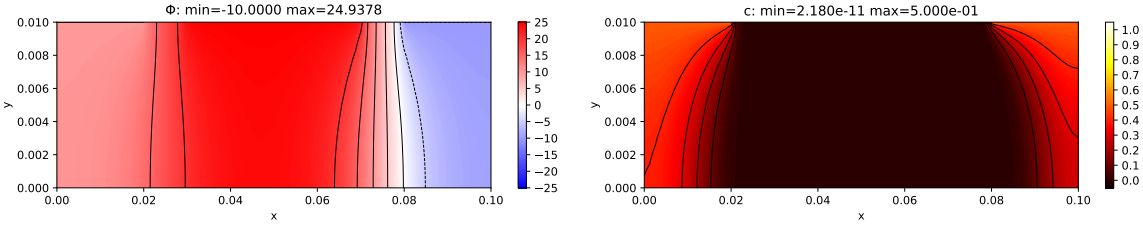

FIGURE 10. Electrostatic potential (left) and concentration (right) for closed gate ($U_{gate} = 50$). fig:closed

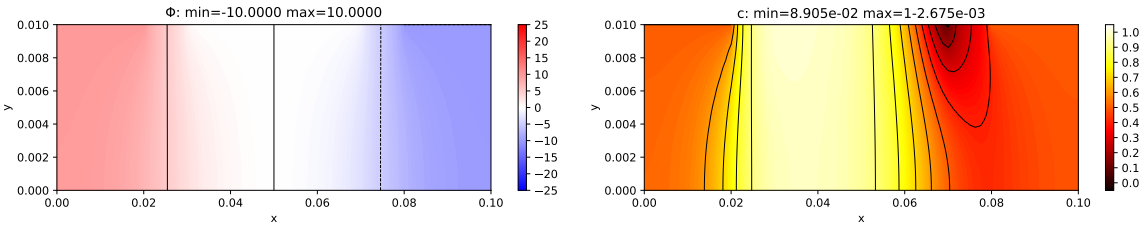

FIGURE 11. Electrostatic potential (left) and concentration (right) for $U_{gate} = 0$). fig:g0

observe second order convergence in the L^2 norm and first order convergence in the H^1-norm.

5.3. **2D Unipolar Field Effect Transistor.** As a second example, we consider a unipolar field effect transistor. The domain is $\Omega = (0, L) \times (0, H)$ with $L = 10^{-1}$, $H = 10^{-2}$.

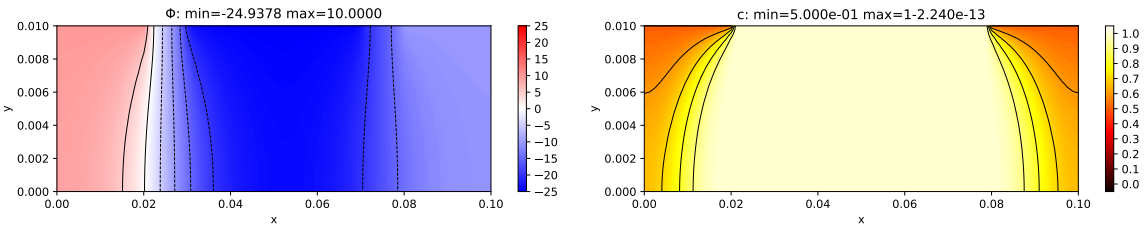

FIGURE 12. Electrostatic potential (left) and concentration (right) for open gate ($U_{gate} = -50$), with concentration in the channel reaching the saturation value 1. fig:open

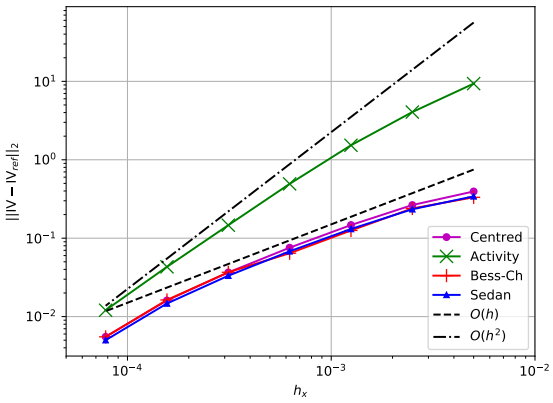

FIGURE 13. Convergence of the I-V curves calculated using the different discretization schemes. fig:erriv

We let $c^{\mathrm{dop}} = -\frac{1}{2}$, and set the following boundary conditions at the contacts:

$$\begin{pmatrix} \Phi \\ c \end{pmatrix} = \begin{pmatrix} -5 \\ \frac{1}{2} \end{pmatrix} \text{ at } \Gamma_{source} = (0, 0.2 \cdot L) \times H,$$

$$\begin{pmatrix} \Phi \\ c \end{pmatrix} = \begin{pmatrix} 5 \\ \frac{1}{2} \end{pmatrix} \text{ at } \Gamma_{drain} = (0.8 \cdot L, L) \times H,$$

$$\begin{pmatrix} \nabla\Phi \cdot \mathbf{n} \\ \mathbf{J} \cdot \mathbf{n} \end{pmatrix} = \begin{pmatrix} -\frac{1}{d}(\Phi - U_{gate}) \\ 0 \end{pmatrix} \text{ at } \Gamma_{gate} = (0.3 \cdot L, 0.7 \cdot L) \times H,$$

$$\begin{pmatrix} \nabla\Phi \cdot \mathbf{n} \\ \mathbf{J} \cdot \mathbf{n} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ at } \partial\Omega \setminus (\Gamma_{gate} \cup \Gamma_{source} \cup \Gamma_{drain}).$$

Here, $\Phi_{gate} \in (-50, 50)$ is the gate voltage, and $d = 0.1 \cdot H$ is the gate thickness. We introduce a slightly anisotropic rectangular grid $n_x \times n_y$ with $n_x = 10 \times 2^{n_{ref}}$ and $n_y = 5 \times 2^{n_{ref}}$, where n_{ref} is the refinement level. Each cell in the rectangular grid is subdivided into two triangles, see Fig. 9 (left). From the resulting triangle mesh, the Voronoi tessellation is obtained.

With fixed source and drain voltages, we vary the gate voltage U_{gate} from 50 to -50. At $U_{gate} = 50$, the positive applied potential pushes away the positively charged carriers from the channel – the region under the gate contact, see Fig. 10. The resulting lack of charge carriers results in a near zero current. With decreasing gate voltage, more and more charge carriers are allowed into the channel, leading to an increase in the current. When the gate voltage decreases further, charge carriers are attracted to the gate contact and fill up the channel. Due to the degeneration, their concentration cannot exceed 1.

As a result, we observe a saturation of the current close to some maximum value for gate voltages approaching -50, see Fig. 10.

All schemes under consideration except (AB) represent this saturation behaviour quite well already at rather coarse grids, see Fig. 9 (right). This appears to be in line with earlier investigations of the scheme based on activity averaging [28] which hint that its asymptotic behaviour for large electric fields is not satisfactory.

To get an idea about the convergence in this case, we produce a reference solution on a grid with 821121 nodes using the scheme (S) and compare the calculated I-V curves. The behaviour of the error in the I-V curves is shown in Fig. 13. While all four schemes exhibit convergence of order at least $O(h)$, the activity based scheme (AB) converges with a constant approximately one order of magnitude larger than the others.

6. Conclusion

Four finite volume numerical schemes for a degenerate unipolar drift-diffusion model have been studied both from a numerical and theoretical point of view. Three of them – the schemes (AB), (S) and (BC) – can be seen as generalizations of the classical Scharfetter-Gummel scheme [47] inspired by different ways to express the degeneracy of the carrier density in the continuous model. Existence of the discrete solution and monotone decrease of the relative free energy have been proven for all four of them. We were able to prove rigorously the convergence to a solution of the continuous problem for only two of the schemes, namely (S) and (C). However, numerical experiments suggest that all the four schemes converge and are of order two with respect to space, even though some particular test cases show limitations for the schemes (AB) and (C). Besides, the extension of the scheme (BC) to more complex physics involving several conservation laws is not straightforward. Moreover, a robust implementation of scheme (BC) requires additional efforts to handle the case of constant concentrations. The present study suggests a preference for scheme (S) in practical applications as long as the mobility is linear. In the case of nonlinear mobilities (like e.g. $c(1-c)$), the extension of the schemes (AB), (S) and (BC) is unclear and a scheme based on (C) seems to be a good option.

Acknowledgements. This work was partially supported by Labex CEMPI (ANR-11-LABX-0007-01). Claire Chainais-Hillairet was also supported by project MoHyCon (ANR-17-CE40-0027-01). Finally, the authors warmly thank the anonymous referees for their constructive feedback.

pp:LinfPhi

Appendix A. L^∞ bound on the TPFA FV approximate Poisson equation

It is well known that the solution to the Poisson equation

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u = u^D & \text{on } \Gamma^D, \\ \nabla u \cdot \boldsymbol{n} = 0 & \text{on } \Gamma^N, \end{cases} \tag{A.1}$$

{eq:Poisso

is bounded in $L^\infty(\Omega)$ provided $f \in L^\infty(\Omega)$ and $u^D \in L^\infty(\partial\Omega)$. The goal of this appendix is to get a discrete counterpart of this estimate for TPFA finite volume approximations of (A.1). The data u^D and f are discretized into

$$u_\sigma^D = \frac{1}{m_\sigma}\int_\sigma u^D(\boldsymbol{\gamma})\mathrm{d}\boldsymbol{\gamma}, \qquad f_K = \frac{1}{m_K}\int_K f\mathrm{d}\boldsymbol{x}, \qquad \sigma \in \mathcal{E}^D,\ K \in \mathcal{T}.$$

and the classical TPFA finite volume scheme is:

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D_{K\sigma} u = m_K f_K, \quad \forall K \in \mathcal{T}.$$

The associate linear system of equations can be written as

$$\mathbb{L}\boldsymbol{u} = \boldsymbol{b}, \tag{A.2}$$

{eq:Lu=b}

with $\boldsymbol{u} = (u_K, u_\sigma)_{K\in\mathcal{T},\sigma\in\mathcal{E}^D}$ (let us note that we keep the Dirichlet nodes in the set of unknowns), $\boldsymbol{b} = \left(f_K, u_\sigma^D\right)_{K\in\mathcal{T},\sigma\in\mathcal{E}^D}$ and $\mathbb{L} \in \mathbb{R}^{(\mathcal{T}\cup\mathcal{E}^D)\times(\mathcal{T}\cup\mathcal{E}^D)}$ is the sparse symmetric definite positive matrix defined by

$$\mathbb{L}_{\sigma,\sigma} = 1, \qquad \mathbb{L}_{\sigma,\ell} = 0 \text{ if } \ell \neq \sigma, \qquad \sigma \in \mathcal{E}^D,$$

$$\mathbb{L}_{K,K\sigma} = -\frac{\tau_\sigma}{m_K}, \qquad \mathbb{L}_{K,K} = \frac{1}{m_K}\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma, \qquad K \in \mathcal{T}.$$

In the above definition of $\mathbb{L}$, ℓ denotes an arbitrary index in $\mathcal{T}\cap\mathcal{E}^D$, whereas $K\sigma$ denotes the mirror index of K w.r.t. the faces $\sigma \in \mathcal{E}_K$, i.e., $K\sigma = L$ if $\sigma = K|L \in \mathcal{E}_K \cap \mathcal{E}_{\text{int}}$ and $K\sigma = \sigma$ if $\sigma \in \mathcal{E}_K \cap \mathcal{E}^D$.

The goal of this section is to derive an ℓ^∞ bound on the solution $\boldsymbol{u}$ to the linear system (A.2) which is uniform w.r.t. the mesh. This is the purpose of the following proposition.

nf_Poisson

Proposition A.1. *There exists C depending only on Ω such that*

$$|u_K| \leq C \max\left\{\|u^D\|_{L^\infty(\partial\Omega)}, \|f\|_{L^\infty(\Omega)}\right\}, \qquad \forall K \in \mathcal{T}.$$

Proof. The proof we propose here is an extension to the context of TPFA Finite Volumes of the proof of Hackbusch [39] for Finite Differences. An alternative proof of Proposition A.1 based on Stampacchia's truncation estimates is sketched in [35].

The definitions of u_σ^D and f_K ensure that

$$\|\boldsymbol{b}\|_\infty \leq \max\left\{\|u^D\|_{L^\infty(\partial\Omega)}, \|f\|_{L^\infty(\Omega)}\right\},$$

so that

$$\|\boldsymbol{u}\|_\infty \leq \|\mathbb{L}^{-1}\|_\infty\|\boldsymbol{b}\|_\infty \leq \|\mathbb{L}^{-1}\|_\infty \max\left\{\|u^D\|_{L^\infty(\partial\Omega)}, \|f\|_{L^\infty(\Omega)}\right\}.$$

Therefore, it only remains to check that $\|\mathbb{L}^{-1}\|_\infty \leq C$ for some C not depending on $\mathcal{T}$.

The matrix $\mathbb{L}$ is a M-matrix (see [39, Definition 4.8]). Therefore, owing to [39, Theorem 4.24], if we can exhibit some vector $\boldsymbol{w} \in \mathbb{R}^{\mathcal{T}\cup\mathcal{E}^D}$ such that $\mathbb{L}\boldsymbol{w} \geq \mathbf{1}$, then $\|\mathbb{L}^{-1}\|_\infty \leq \|\boldsymbol{w}\|_\infty$. Define the function $w : \overline{\Omega} \to \mathbb{R}$ by

$$w(\boldsymbol{x}) = 1 + \frac{1}{d}\left(\sup_{\boldsymbol{y}\in\Omega}|\boldsymbol{y}|^2 - |\boldsymbol{x}|^2\right) \geq 1, \qquad \boldsymbol{x} \in \overline{\Omega},$$

and the vector $\boldsymbol{w} = (w_K, w_\sigma)$ by $w_K = w(\boldsymbol{x}_K)$, $K \in \mathcal{T}$, and $w_\sigma = w(\boldsymbol{x}_\sigma)$, $\sigma \in \mathcal{E}^D$.

The estimate on the Dirichlet nodes is straightforward:

$$(\mathbb{L}\boldsymbol{w})_\sigma = w_\sigma \geq 1, \qquad \forall \sigma \in \mathcal{E}^D.$$

Now, let us focus on the inner nodes $K \in \mathcal{T}$. Since $\sum_{\ell \in \mathcal{T}\cup\mathcal{E}^D} \mathbb{L}_{K,\ell} = \sum_{\sigma\in\mathcal{E}_K} \mathbb{L}_{K,K\sigma} = 0$, one has

$$\begin{aligned}(\mathbb{L}\boldsymbol{w})_K =& \frac{1}{d}\sum_{\sigma\in\mathcal{E}_K} \mathbb{L}_{K,K\sigma}|\boldsymbol{x}_{K\sigma}|^2 = \frac{1}{dm_K}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\left(|\boldsymbol{x}_K|^2 - |\boldsymbol{x}_{K\sigma}|^2\right)\\ =& \frac{1}{dm_K}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\left(|\boldsymbol{x}_K - \boldsymbol{x}_{K\sigma}|^2 + 2\boldsymbol{x}_K\cdot(\boldsymbol{x}_K - \boldsymbol{x}_{K\sigma})\right)\\ =& \frac{1}{dm_K}\sum_{\sigma\in\mathcal{E}_K} m_\sigma d_\sigma + \frac{2}{dm_K}\boldsymbol{x}_K \cdot \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma(\boldsymbol{x}_K - \boldsymbol{x}_{K\sigma}).\end{aligned}$$

Because of the geometric relation $m_\sigma d_\sigma = dm_{\Delta_\sigma}$, and since $K \subset \bigcup_{\sigma\in\mathcal{E}_K}\Delta_\sigma$, there holds

$$\frac{1}{dm_K}\sum_{\sigma\in\mathcal{E}_K} m_\sigma d_\sigma = \frac{1}{m_K}\sum_{\sigma\in\mathcal{E}_K} m_{\Delta_\sigma} \geq 1.$$

On the other hand, the second term vanishes since

$$\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma(\boldsymbol{x}_K - \boldsymbol{x}_{K\sigma}) = -\sum_{\sigma\in\mathcal{E}_K} m_\sigma \boldsymbol{n}_{K\sigma} = \boldsymbol{0}.$$

Therefore, $(\mathbb{L}\boldsymbol{w})_K \geq 1$ for all $K \in \mathcal{T}$. As a consequence,

$$\|\mathbb{L}^{-1}\|_\infty \leq \|\boldsymbol{w}\|_\infty = 1 + \frac{1}{d}\sup_{\boldsymbol{y}\in\Omega}|\boldsymbol{y}|^2 \leq 1 + \frac{\mathrm{diam}(\Omega)^2}{4d}.$$

The last estimate comes from the fact that one can choose the origin for $\boldsymbol{y}$ arbitrarily. □

app:dissip

Appendix B. Proof of Lemma 3.2

Step 1. Let $\delta \in (0,1)$ and $M \in \mathbb{R}$. We start with the proof of

$$\lim_{c_L\to 1} \Upsilon_{\delta,M}(c_L) = +\infty, \tag{B.1}$$

{lim_Upsil

where

$$\Upsilon_{\delta,M}(c_L) = \inf\left\{\mathcal{D}(c_K, c_L, \Phi_K, \Phi_L); c_K \in (0, 1-\delta], (\Phi_K, \Phi_L) \in [-M, M]^2\right\}.$$

We recall that

$$\begin{aligned}\mathcal{D}(c_K, c_L, \Phi_K, \Phi_L) &= \mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)|h(c_K) + \Phi_K - h(c_L) - \Phi_L|^2\\ &= \mathcal{F}(c_K, c_L, \Phi_K, \Phi_L)(h(c_K) + \Phi_K - h(c_L) - \Phi_L),\end{aligned}$$

and we notice that the diffusion force blows up:

$$\lim_{c_L\to 1} \inf\left\{\left|h(c_K) - h(c_L) + \Phi_K - \Phi_L\right|; c_K \in (0, 1-\delta], (\Phi_K, \Phi_L) \in [-M, M]^2\right\} = +\infty. \tag{B.2}$$

{eq:diff-f

Therefore, we can get (B.1) by proving that either $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)$ or $\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L)$ stays bounded away from 0, uniformly in $c_K \in (0, 1-\delta]$, $(\Phi_K, \Phi_L) \in [-M, M]^2$, for $c_L \geq 1/2$.

For the centred flux, we have that, for all $(c_K, \Phi_K, \Phi_L) \in (0, 1-\delta] \times [-M, M]^2$, $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = (c_K + c_L)/2 \geq c_L/2$. This yields (B.1). For the three other schemes, we remark that, for any $\alpha \in (0, 1-\delta)$, we have

$$\Upsilon_{\delta,M}(c_L) = \min(\Upsilon^{\alpha,1}_{\delta,M}(c_L), \Upsilon^{\alpha,2}_{\delta,M}(c_L)),$$

where

$$\Upsilon^{\alpha,1}_{\delta,M}(c_L) = \inf\left\{\mathcal{D}(c_K,c_L,\Phi_K,\Phi_L); c_K \in (0,\alpha), (\Phi_K,\Phi_L)\in[-M,M]^2\right\},$$
$$\Upsilon^{\alpha,2}_{\delta,M}(c_L) = \inf\left\{\mathcal{D}(c_K,c_L,\Phi_K,\Phi_L); c_K \in [\alpha,1-\delta], (\Phi_K,\Phi_L)\in[-M,M]^2\right\}.$$

The Lemma 3.1 ensures that, independently of the choice of the numerical flux, we have at least $\mathcal{C}(c_K,c_L,\Phi_K,\Phi_L) \geq \min(c_K,c_L)/2$, so that $\mathcal{C}(c_K,c_L,\Phi_K,\Phi_L) \geq \alpha/2$ for all $(c_K,c_L,\Phi_K,\Phi_L) \in [\alpha,1-\delta]\times[1/2,1)\times[-M,M]^2$ if $\alpha \in (0,1-\delta)$.Therefore, for all $\alpha\in(0,1-\delta)$, we have

$$\lim_{c_L\to 1} \Upsilon^{\alpha,2}_{\delta,M}(c_L) = +\infty.$$

It remains to prove that for a given $\alpha\in(0,1-\delta)$ we also have

$$\lim_{c_L\to 1} \Upsilon^{\alpha,1}_{\delta,M}(c_L) = +\infty. \tag{B.3}$$ {lim_Upsil

Because of the monotonicity of $\delta \mapsto \Upsilon_{\delta,M}(c_L)$, we can restrict our attention to the case $\delta \leq 1/2$, so that we can seek for $\alpha \in (0,1/2]$.

For the Bessemoulin-Chatard flux, we have

$$\mathcal{F}(c_K,c_L,\Phi_K,\Phi_L) = \mathfrak{d}r(c_K,c_L)\left\{B\left(\frac{\Phi_L-\Phi_K}{\mathfrak{d}r(c_K,c_L)}\right)c_K - B\left(\frac{\Phi_K-\Phi_L}{\mathfrak{d}r(c_K,c_L)}\right)c_L\right\},$$

with $\mathfrak{d}r(c_K,c_L)\geq 1$. Using the monotonicity of the Bernoulli function and the bounds on Φ_K and Φ_L, we get:

$$B(2M) \leqslant B\left(\pm\frac{\Phi_L-\Phi_K}{\mathfrak{d}r(c_K,c_L)}\right) \leqslant B(-2M).$$

Hence, for $\alpha = \frac{B(2M)}{4B(-2M)}$:

$$\mathcal{F}(c_K,c_L,\Phi_K,\Phi_L) \leqslant \mathfrak{d}r(c_K,c_L)\left\{B(-2M)\alpha - B(2M)\frac{1}{2}\right\} \leqslant -\frac{B(2M)}{4}.$$

Then, thanks to (B.2), we deduce (B.3) for $\alpha = \frac{B(2M)}{4B(-2M)}$ and therefore (B.1).

For the Sedan flux, we use similarly the monotonicity of the function B and ν, so that

$$\mathcal{F}(c_K,c_L,\Phi_K,\Phi_L) \leq B\left(-2M+\nu(\frac{1}{2})-\nu(\alpha)\right)\alpha - B\left(2M-\nu(\frac{1}{2})+\nu(\alpha)\right)\frac{1}{2},$$
$$\forall c_K\in(0,\alpha), c_L\in(\frac{1}{2},1), (\Phi_K,\Phi_L)\in[-M,M]^2.$$

The right-hand side of the last inequality tends to $-B\left(2M-\nu(\frac{1}{2})\right)\frac{1}{2}$ when α tends to 0. The negativity of this limit means that for a given α small enough the flux remains bounded away from 0 so that we deduce (B.3) and therefore (B.1).

For the activity based flux, we also use the monotonicity of a and β, which yields

$$\mathcal{F}(c_K,c_L,\Phi_K,\Phi_L) \leq \frac{1}{4}\left(B(-2M)a(\alpha) - B(2M)a(\frac{1}{2})\right),$$
$$\forall c_K\in(0,\alpha), c_L\in(\frac{1}{2},1), (\Phi_K,\Phi_L)\in[-M,M]^2.$$

The right-hand side has a negative limit when α tends to 0. Thus, it remains bounded away from 0 for a given $\alpha < 1/2$, and we deduce (B.3) and therefore (B.1).

Step 2. We now focus on the proof of

$$\lim_{c_L\to 0} \Psi_{\delta,M}(c_L) = +\infty \tag{B.4}$$ {lim_Psi}

where

$$\Psi_{\delta,M}(c_L) = \inf\left\{\mathcal{D}(c_K, c_L, \Phi_K, \Phi_L); c_K \in [\delta, 1), (\Phi_K, \Phi_L) \in [-M, M]^2\right\}.$$

We use similar arguments than in Step 1. First, the diffusion force still blows up:

$$\lim_{c_L \to 0} \inf\left\{\left|h(c_K) - h(c_L) + \Phi_K - \Phi_L\right|; c_K \in [\delta, 1), (\Phi_K, \Phi_L) \in [-M, M]^2\right\} = +\infty. \quad \text{(B.5)}$$

{eq:diff-f

For the centred flux, we have: $\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = (c_K + c_L)/2 \geq \delta/2$ hence (B.4). For the other fluxes, we split using a parameter α again

$$\Psi_{\delta,M}(c_L) = \min(\Psi^{\alpha,1}_{\delta,M}(c_L), \Psi^{\alpha,2}_{\delta,M}(c_L)),$$

where

$$\Psi^{\alpha,1}_{\delta,M}(c_L) = \inf\left\{\mathcal{D}(c_K, c_L, \Phi_K, \Phi_L); c_K \in [\delta, \alpha], (\Phi_K, \Phi_L) \in [-M, M]^2\right\};$$
$$\Psi^{\alpha,2}_{\delta,M}(c_L) = \inf\left\{\mathcal{D}(c_K, c_L, \Phi_K, \Phi_L); c_K \in (\alpha, 1), (\Phi_K, \Phi_L) \in [-M, M]^2\right\}.$$

Using the symmetry of the flux $\mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) = -\mathcal{F}(c_L, c_K, \Phi_L, \Phi_K)$ and following the proof of (B.3), we get that for $\alpha = 1/2$,

$$\lim_{c_L \to 0} \Psi^{\alpha,2}_{\delta,M}(c_L) = +\infty$$

We now have to prove that, for $\alpha = 1/2$,

$$\lim_{c_L \to 0} \Psi^{\alpha,1}_{\delta,M}(c_L) = +\infty \quad \text{(B.6)}$$

{lim_Psi_a

To this end, we will show bounds on the flux. The set $[\delta, \alpha] \times [-M, M]^2$ is compact, and the flux functions are continuous. It is sufficient to show a positive lower bound for the limit at any $(c^*, \Phi^*, \Phi_*) \in [\delta, \alpha] \times [-M, M]^2$:

$$l^* = \lim_{(c_K, c_L, \Phi_K, \Phi_L) \to (c^*, 0, \Phi^*, \Phi_*)} \mathcal{F}(c_K, c_L, \Phi_K, \Phi_L) > 0.$$

For the Sedan scheme, we have:

$$l^* = B\left(\Phi_* - \Phi^* - \nu(c^*)\right) c^* \geq \delta B(2M).$$

For the Bessemoulin-Chatard scheme, we have: $\lim_{(c_K, c_L) \to (c^*, 0)} \mathfrak{d}r(c_K, c_L) = 1$, hence:

$$l^* = B(\Phi_* - \Phi^*)c^* \geq \delta B(2M).$$

For the activity based scheme we have:

$$l^* = \frac{\beta(c^*) + 1}{2} B(\Phi_* - \Phi^*)a(c^*) \geq \frac{a(\delta)}{2} B(2M).$$

As these limits are bounded away from zero we have (B.6) hence (B.4). This concludes the proof of Lemma 3.2.

Appendix C. Comparison of face concentration functionals

app:alacon

For each scheme, we have defined a face concentration functional $\mathcal{C} : (0,1) \times (0,1) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. We introduce a second face concentration functional $\widetilde{\mathcal{C}} : (0,1) \times (0,1) \to \mathbb{R}$, defined by

$$\widetilde{\mathcal{C}}(c_K, c_L) = \frac{r(c_K) - r(c_L)}{h(c_K) - h(c_L)} \text{ if } c_K \neq c_L \text{ and } c_K \text{ otherwise.}$$

As $r'(c) = ch'(c)$, it is clear that

$$\min(c_K, c_L) \leq \widetilde{\mathcal{C}}(c_K, c_L) \leq \max(c_K, c_L) \text{ for all } (c_K, c_L) \in (0,1) \times (0,1). \quad \text{(C.1)}$$

{eq:wt_avg

Lemma C.1 states a comparison between $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ for the centred and the Sedan schemes.

em:cctilde **Lemma C.1.** *For the centred scheme and the Sedan scheme, there exists $G > 0$, depending only on M, such that for all $(c_K, c_L, \Phi_K, \Phi_L) \in (0,1) \times (0,1) \times [-M, M] \times [-M, M]$,*

$$\frac{\widetilde{\mathcal{C}}(c_K, c_L)}{\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)} \leq G. \tag{C.2}$$

est:cctil

BCHtrouble *Remark* C.1. For the Bessemoulin-Chatard scheme, the bound (C.2) does not hold. Let us consider that $\Phi_K = \Phi_L = \Phi$, then with the notations $x = \log(c_K/c_L)$, and $y = \log(\frac{1-c_L}{1-c_K})$, we have:

$$\frac{\tilde{\mathcal{C}}(c_K, c_L)}{\mathcal{C}(c_K, c_L, \Phi, \Phi)} = \frac{xy}{(c_K - c_L)(x+y)}.$$

For $(c_K, c_L) \to (1, 0)$, x and y tends to $+\infty$, hence the blow up of the ratio.

Proof. The case of the centred scheme defined by (C) is the easiest one, since

$$\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = \frac{c_K + c_L}{2} \geq \frac{1}{2}\max(c_K, c_L),$$

so that (C.2) holds with $G = 2$ thanks to (C.1).

Let us now focus on the Sedan scheme defined by (S). We can introduce the function $\mathcal{G} : (0,1) \times (0,1) \to \mathbb{R}$ defined by

$$\mathcal{G}(c_K, c_L) = \frac{\widetilde{\mathcal{C}}(c_K, c_L)}{\min_{(\Phi_K, \Phi_L) \in [-M,M]^2} \mathcal{C}(c_K, c_L, \Phi_K, \Phi_L)}.$$

It is a continuous function which satisfies the symmetry property $\mathcal{G}(c_K, c_L) = \mathcal{G}(c_L, c_K)$ and the consistency $\mathcal{G}(c_K, c_K) = 1$.

Because of the symmetry and the consistency properties, we can assume without loss of generality that $c_K > c_L$. Using the average properties (3.3) and (C.1), one obtains that

$$\mathcal{G}(c_K, c_L) \leq \frac{c_K}{c_L} \leq \frac{1}{c_L}, \tag{C.3}$$

eq:angle

so that we only have to check that $\mathcal{G}(c_K, c_L)$ remains uniformly bounded as c_L tends to 0 to prove (C.2). To that extent, we compute explicitly the minimum of $\mathcal{C}$. We recall that we have, using (3.5):

$$\mathcal{C}(c_K, c_L, \Phi_K, \Phi_L) = t c_K + (1-t) c_L, \qquad t = \frac{B(y) - B(x)}{x - y},$$

where $y = \Phi_L + \nu(c_L) - \Phi_K - \nu(c_K)$ and $x = \log(c_K/c_L)$. Using the assumption $c_K > c_L$, $\mathcal{C}$ is minimal when t is minimal. As B is convex, and x fixed, this happens for y maximal, i.e.

$$y = 2M + \nu(c_L) - \nu(c_K).$$

Using this result, one can expand

$$\mathcal{G}(c_K, c_L) = \frac{(h(c_K) - h(c_L) - 2M)(r(c_K) - r(c_L))}{\big(B(2M + \nu(c_L) - \nu(c_K))c_K - B(-2M - \nu(c_L) + \nu(c_K))c_L\big)(h(c_K) - h(c_L))}$$

Therefore we study the limit of $\mathcal{G}$ when (c_K, c_L) tends to (1,0), (0,0) and $(c^*, 0)$ with $c^* \in (0,1)$.

We first consider the limit $(c_K, c_L) \to (1, 0)$. We have the following equivalences when $(c_K, c_L) \to (1, 0)$:

$$h(c_K) - h(c_L) \sim -\log(1 - c_K) - \log(c_L)$$
$$h(c_K) - h(c_L) - 2M \sim -\log(1 - c_K) - \log(c_L)$$
$$r(c_K) - r(c_L) \sim -\log(1 - c_K)$$
$$B(y)c_K - B(-y)c_L \sim -c_K \log(1 - c_K)$$

This yields:

$$\lim_{(c_K,c_L)\to(1,0)} \mathcal{G}(c_K, c_L, \Phi_K, \Phi_L) = 1. \tag{C.4}$$ {lim1}

With similar arguments, we compute the limit $(c_K, c_L) \to (c^*, 0)$ with $c^* \in (0, 1)$. We get:

$$\lim_{(c_K,c_L)\to(c^*,0)} \mathcal{G}(c_K, c_L) = \frac{r(c^*)}{B(y^*)c^*}, \tag{C.5}$$ {lim2}

with $y^* = 2M + \log(1 - c^*)$.

In the neighbourhood of $(0, 0)$, the behaviour is more complex, as the limit of $\log(c_K/c_L)$ is not defined and $\mathcal{G}$ does not have a limit. However, thanks to (C.3), $\mathcal{G}(c_K, c_L)$ stays bounded if c_K/c_L stays bounded while $(c_K, c_L) \to (0, 0)$. It remains to consider the case where $(c_K, c_L) \to (0, 0)$ while $c_K/c_L \to \infty$. In this case, we have:

$$\frac{(h(c_K) - h(c_L) - 2M)}{h(c_K) - h(c_L)} \to 1,$$
$$r(c_K) - r(c_L) \sim -c_L + c_K$$
$$B(y)c_K - B(-y)c_L \sim B(2M)c_K - B(-2M)c_L,$$

and

$$\lim_{\substack{(c_K,c_L)\to(0,0)\\ c_K/c_L\to\infty}} \mathcal{G}(c_K, c_L) = \frac{1}{B(2M)}.$$

We conclude that $\mathcal{G}(c_K, c_L)$ stays bounded when (c_K, c_L) is in the neighbourhood of $(0, 0)$. Combined with (C.3), (C.4) and (C.5), this concludes the proof of Lemma C.1. □

Appendix D. Some notations

:notations

In this Section, we recall the definition of some notations used along the paper. Table 1 gives the definition of the different quantities involved at the continuous level, while Table 2 gives the definition of the functions involved in the study of the numerical schemes.

$h(c)$	$\log \dfrac{c}{1-c}$	chemical potential
$\nu(c)$	$-\log(1-c)$	excess chemical potential
$r(c)$	$-\log(1-c)$	diffusion enhancement
$a(c)$	$\dfrac{c}{1-c}$	activity coefficient
$\beta(c)$	$1-c$	inverse activity coefficient
$H(c)$	$c\log(c)+(1-c)\log(1-c)$	entropy density
$E(c,\Phi)$	$\displaystyle\int_\Omega \left\{H(c)+\frac{1}{2}\lvert\nabla\Phi\rvert^2\right\}\mathrm{d}\boldsymbol{x}-\int_{\Gamma_D}\Phi^D\nabla\Phi\cdot\boldsymbol{n}\mathrm{d}\boldsymbol{\gamma}$	free energy

le:continu TABLE 1. Definition of the different functions involved in the continuous problem.

$B(x)$	$\dfrac{x}{e^x-1}$	Bernoulli function
$\mathfrak{d}r(c_K,c_L)$	$\begin{cases}\dfrac{h(c_K)-h(c_L)}{\log(c_K)-\log(c_L)} & \text{if } c_K\neq c_L,\\ r'(c_K) & \text{if } c_K=c_L.\end{cases}$	approximation of r' consistant with the thermal equilibrium
$\mathcal{F}(c_K,c_L,\Phi_K,\Phi_L)$	$\dfrac{F_{KL}}{\tau_{KL}}$	numerical flux intensity
$\mathcal{C}(c_K,c_L,\Phi_K,\Phi_L)$	$\dfrac{\mathcal{F}(c_K,c_L,\Phi_K,\Phi_L)}{h(c_K)+\Phi_K-h(c_L)-\Phi_L}$	face concentration
$\widetilde{\mathcal{C}}(c_K,c_L)$	$\begin{cases}\dfrac{r(c_K)-r(c_L)}{h(c_K)-h(c_L)} & \text{if } c_K\neq c_L\\ c_K & \text{otherwise}\end{cases}$	face concentration compliant with the weak solution
$\mathcal{D}(c_K,c_L,\Phi_K,\Phi_L)$	$\mathcal{F}(c_K,c_L,\Phi_K,\Phi_L)$ $\times\,(h(c_K)+\Phi_K-h(c_L)-\Phi_L)$	face entropy dissipation

ble:scheme TABLE 2. Definition of the different functions involved in the numerical schemes.

REFERENCES

[1] A. Ait Hammou Oulhaj, C. Cancès, and C. Chainais-Hillairet. Numerical analysis of a nonlinearly stable and positive Control Volume Finite Element scheme for Richards equation with anisotropy. ESAIM: Mathematical Modelling and Numerical Analysis, 52(4):1532–1567, 2018.

[2] B. Andreianov, C. Cancès, and A. Moussa. A nonlinear time compactness result and applications to discretization of degenerate parabolic–elliptic PDEs. J. Funct. Anal., 273(12):3633–3670, 2017.
[3] M. Bessemoulin-Chatard. A finite volume scheme for convection-diffusion equations with nonlinear diffusion derived from the Scharfetter-Gummel scheme. Numer. Math., 121(4):637–670, 2012.
[4] M. Bessemoulin-Chatard, C. Chainais-Hillairet, and F. Filbet. On discrete functional inequalities for some finite volume schemes. IMA J. Numer. Anal., 35:1125–1149, 2015.
[5] J. Bezanson, Al. Edelman, S. Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. SIAM review, 59(1):65–98, 2017.
[6] J.S. Blakemore. Approximations for fermi-dirac integrals, especially the function f12 (η) used to describe electron density in a semiconductor. Solid-State Electronics, 25(11):1067–1076, 1982.
[7] K. Brenner, C. Cancès, and D. Hilhorst. Finite volume approximation for an immiscible two-phase flow in porous media with discontinuous capillary pressure. Comput. Geosci., 17(3):573–597, 2013.
[8] F. Brochard, J. Jouffroy, and P. Levinson. Polymer-polymer diffusion in melts. Macromolecules, 16(10):1638–1641, 1983.
[9] C. Cancès. Energy stable numerical methods for porous media flow type problems. Oil & Gas Science and Technology-Rev. IFPEN, 73:1–18, 2018.
[10] C. Cancès, C. Chainais-Hillairet, and S. Krell. Numerical analysis of a nonlinear free-energy diminishing Discrete Duality Finite Volume scheme for convection diffusion equations. Computational Methods in Applied Mathematics, 2017. Special issue on "Advanced numerical methods: recent developments, analysis and application".
[11] C. Cancès and C. Guichard. Convergence of a nonlinear entropy diminishing Control Volume Finite Element scheme for solving anisotropic degenerate parabolic equations. Math. Comp., 85(298):549–580, 2016.
[12] C. Cancès and C. Guichard. Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure. Found. Comput. Math., 17(6):1525–1584, 2017.
[13] C. Cancès, F. Nabet, and M. Vohralík. Convergence and a posteriori error analysis for energy-stable finite element approximations of degenerate parabolic equations. HAL: hal-01894884, 2018.
[14] Clément Cancès, Claire Chainais-Hillairet, Anita Gerstenmayer, and Ansgar Jüngel. Finite-volume scheme for a degenerate cross-diffusion model motivated from ion transport. Numerical Methods for Partial Differential Equations, 35(2):545–575, 2019.
[15] C. Chainais-Hillairet, J.-G. Liu, and Y.-J. Peng. Finite volume scheme for multi-dimensional drift-diffusion equations and convergence analysis. ESAIM: M2AN, 37(2):319–338, 2003.
[16] R. Coehoorn, W.F. Pasveer, P.A. Bobbert, and M.A.J. Michels. Charge-carrier concentration dependence of the hopping mobility in organic materials with gaussian disorder. Physical Review B, 72(15):155206, 2005.
[17] Y. Coudière, J.-P. Vila, and P. Villedieu. Convergence rate of a finite volume scheme for a two dimensional convection-diffusion problem. ESAIM Mathematical Modelling and Numerical Analysis, 33:493–516, 1999.
[18] K. Deimling. Nonlinear functional analysis. Springer-Verlag, Berlin, 1985.
[19] W. Dreyer, C. Guhlke, and M. Landstorfer. A mixture theory of electrolytes containing solvation effects. Electrochemistry Communications, 43:75–78, 2014.
[20] W. Dreyer, C. Guhlke, and R. Müller. Overcoming the shortcomings of the nernst–planck model. Physical Chemistry Chemical Physics, 15(19):7075–7086, 2013.
[21] J. Droniou. Finite volume schemes for diffusion equations: introduction to and review of modern methods. Math. Models Methods Appl. Sci., 24(8):1575–1620, 2014.
[22] J. Droniou and R. Eymard. Study of the mixed finite volume method for stokes and navier-stokes equations. Numerical Methods for Partial Differential Equations, 25:137–171, 2009.
[23] J. Droniou and R. Eymard. The asymmetric gradient discretisation method. In C. Cancès and P. Omnes, editors, Finite volumes for complex applications VIII - methods and theoretical aspects, volume 199 of Springer Proc. Math. Stat., pages 311–319, Cham, 2017. Springer.
[24] J. Droniou, R. Eymard, T. Gallouët, C. Guichard, and R. Herbin. The Gradient Discretisation Method, volume 42 of Mathématiques et Applications. Springer International Publishing, 2018.
[25] R. Eymard and T. Gallouët. H-convergence and numerical schemes for elliptic problems. SIAM J. Numer. Anal., 41(2):539–562, 2003.
[26] R. Eymard, T. Gallouët, C. Guichard, R. Herbin, and R. Masson. TP or not TP, that is the question. Comput. Geosci., 18:285–296, 2014.

[27] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. Ciarlet, P. G. (ed.) et al., in Handbook of numerical analysis. North-Holland, Amsterdam, pp. 713–1020, 2000.
[28] P. Farrell, T. Koprucki, and J. Fuhrmann. Computational and analytical comparison of flux discretizations for the semiconductor device equations beyond Boltzmann statistics. J. Comput. Phys., 346:497–513, 2017.
[29] J. Fuhrmann. Comparison and numerical treatment of generalised Nernst–Planck models. Comput. Phys. Commun., 196:166–178, 2015.
[30] J. Fuhrmann. A numerical strategy for Nernst-Planck systems with solvation effect. Fuel Cells, 16, 12 2016.
[31] J. Fuhrmann. UnipolarDriftDiffusion.jl - Numerical examples for finite volume schemes for unipolar drift-diffusion problems, 2019. DOI:10.5281/zenodo.3351467.
[32] J. Fuhrmann. VoronoiFVM.jl: Solver for coupled nonlinear partial differential equations based on the voronoi finite volume method, 2019. DOI:10.5281/zenodo.3529808.
[33] J. Fuhrmann and C. Guhlke. A finite volume scheme for Nernst-Planck-Poisson systems with ion size and solvation effects. In C. Cancès & P. Omnes, editor, Finite Volumes for Complex Applications VIII - Hyperbolic, Elliptic and Parabolic Problems, volume 200 of Springer Proceedings in Mathematics & Statistics, pages 497–505, Lille, France, 2017. Springer International Publishing.
[34] H. Gajewski. On the uniqueness of solutions to the drift-diffusion model of semiconductor devices. Math. Models Methods Appl. Sci., 4(1):121–133, 1994.
[35] T. Gallouët. Nonlinear methods for linear equations. In Proceedings of Tamtam'07 conference held in Tipaza, 2007.
[36] T. Gallouët and J.-C. Latché. Compactness of discrete approximate solutions to parabolic PDEs—application to a turbulence model. Commun. Pure Appl. Anal., 11(6):2371–2391, 2012.
[37] N. Gavish and A. Yochelis. Theory of phase separation and polarization for pure ionic liquids. The journal of physical chemistry letters, 7(7):1121–1126, 2016.
[38] A. Glitzky and J. A. Griepentrog. Discrete Sobolev-Poincaré inequalities for Voronoi finite volume approximations. SIAM J. Numer. Anal., 48:372–391, 2010.
[39] W. Hackbusch. Elliptic Differential Equations: Theory and Numerical Treatment. Springer Series in Computational Mathematics 18. Springer-Verlag Berlin Heidelberg, 2nd edition, 2017.
[40] R. Herbin. An error estimate for a finite volume scheme for a diffusion–convection problem on a triangular mesh. Numer. Methods Partial Differential Equations, 11(2):165–173, 1995.
[41] Ernö Keszei. Chemical thermodynamics: an introduction. Springer Science & Business Media, 2013.
[42] T. Koprucki and K. Gärtner. Discretization scheme for drift-diffusion equations with strong diffusion enhancement. Optical and Quantum Electronics, 45(7):791–796, 2013.
[43] J. Leray and J. Schauder. Topologie et équations fonctionnelles. Ann. Sci. École Norm. Sup. (3), 51:45–78, 1934.
[44] A. D. McNaught and A. Wilkinson, editors. IUPAC Compendium of Chemical Terminology (the "Gold Book"). Blackwell Scientific, Oxford, 1997. Online version (2019-) created by S. J. Chalk. https://doi.org/10.1351/goldbook.
[45] G. Paasch and S. Scheinert. Charge carrier density of organics with gaussian density of states: analytical approximation for the gauss–fermi integral. Journal of Applied Physics, 107(10):104501, 2010.
[46] J. Revels, M. Lubin, and T. Papamarkou. Forward-mode automatic differentiation in julia. arXiv:1607.07892 [cs.MS], 2016.
[47] D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon read diode oscillator. Electron Devices, IEEE Transactions on, 16(1):64–77, 1969.
[48] S. Selberherr. Analysis and simulation of semiconductor devices. Springer, 2012.
[49] P. Vágner, C. Guhlke, V. Miloš, R. Müller, and J. Fuhrmann. A continuum model for yttria-stabilized zirconia incorporating triple phase boundary, lattice structure and immobile oxide ions. Journal of Solid State Electrochemistry, pages 1–20, 2019.
[50] SLM Van Mensfoort and Reinder Coehoorn. Effect of gaussian disorder on the voltage dependence of the current density in sandwich-type devices based on organic semiconductors. Physical Review B, 78(8):085207, 2008.
[51] Z. Yu and R. Dutton. SEDAN III. www-tcad.stanford.edu/tcad/programs/sedan3.html, 1988.

Clément Cancès (clement.cances@inria.fr): Inria, Univ. Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé, F-59000 Lille

Claire Chainais-Hillairet (claire.chainais@univ-lille.fr): Univ. Lille, CNRS, UMR 8524, Inria - Laboratoire Paul Painlevé, F-59000 Lille

Jürgen Fuhrmann (juergen.fuhrmann@wias-berlin.de): Weierstrass Institute (WIAS), Mohrenstr. 39, 10117 Berlin, Germany

Benoît Gaudeul (benoit.gaudeul@univ-lille.fr): Univ. Lille, CNRS, UMR 8524, Inria - Laboratoire Paul Painlevé, F-59000 Lille