Développement et analyse de schémas d'ordre élevé pour des modèles de convection-diffusion : étude du comportement en temps long

Julien Moatti

Université de Lille

Université de Lille

École doctorale **MADIS**

Unité de recherche **Centre Inria de l'Université de Lille**

Thèse présentée par **Julien Moatti**

Soutenue le **26 septembre 2023**

En vue de l'obtention du grade de docteur de l'Université de Lille

Discipline **Mathématiques**

Développement et analyse de schémas d'ordre élevé pour des modèles de convection-diffusion, étude du comportement en temps long

Thèse dirigée par Claire Chainais-Hillairet
Maxime Herda
Simon Lemaire

Composition du jury

Rapporteurs	Lourenço Beirão da Veiga	professeur à l'Università degli Studi di Milano-Bicocca	
	Jérôme Droniou	professeur à Monash University, Melbourne	
Examinateurs	Alexandre Ern	professeur à l'ENPC - Université Paris Est	président du jury
	Raphaèle Herbin	professeure à l'Université d'Aix-Marseille	
	Stella Krell	mcf hdr à l'Université Côte d'Azur	
	Ilaria Perugia	professeure à l'Universität Wien	
Invités	Maxime Herda	chargé de recherche chez Inria	
	Simon Lemaire	chargé de recherche chez Inria	
Directrice de thèse	Claire Chainais-Hillairet	professeure à l'Université de Lille	

Université de Lille

Université de Lille

École doctorale **MADIS**

Unité de recherche **Centre Inria de l'Université de Lille**

Thèse présentée par **Julien Moatti**

Soutenue le **26 septembre 2023**

En vue de l'obtention du grade de docteur de l'Université de Lille

Discipline **Mathématiques**

Développement et analyse de schémas d'ordre élevé pour des modèles de convection-diffusion, étude du comportement en temps long

Thèse dirigée par Claire Chainais-Hillairet
Maxime Herda
Simon Lemaire

Composition du jury

Rapporteurs	Lourenço Beirão da Veiga	professeur à l'Università degli Studi di Milano-Bicocca	
	Jérôme Droniou	professeur à Monash University, Melbourne	
Examinateurs	Alexandre Ern	professeur à l'ENPC - Université Paris Est	président du jury
	Raphaèle Herbin	professeure à l'Université d'Aix-Marseille	
	Stella Krell	mcf hdr à l'Université Côte d'Azur	
	Ilaria Perugia	professeure à l'Universität Wien	
Invités	Maxime Herda	chargé de recherche chez Inria	
	Simon Lemaire	chargé de recherche chez Inria	
Directrice de thèse	Claire Chainais-Hillairet	professeure à l'Université de Lille	

Université de Lille

Université de Lille

Doctoral School **MADIS**

University Department **Centre Inria de l'Université de Lille**

Thesis defended by **Julien Moatti**

Defended on **26th September, 2023**

In order to become Doctor from Université de Lille

Academic Field **Mathematics**

Development and numerical analysis of high-order schemes for convection-diffusion models, study of their long-time behaviour

Thesis supervised by Claire Chainais-Hillairet
Maxime Herda
Simon Lemaire

Committee members

Referees	Lourenço Beirão da Veiga	Professor at Università degli Studi di Milano-Bicocca	
	Jérôme Droniou	Professor at Monash University, Melbourne	
Examiners	Alexandre Ern	Professor at ENPC - Université Paris Est	Committee President
	Raphaèle Herbin	Professor at Université d'Aix-Marseille	
	Stella Krell	HDR Associate Professor at Université Côte d'Azur	
	Ilaria Perugia	Professor at Universität Wien	
Guests	Maxime Herda	Junior Researcher at Inria	
	Simon Lemaire	Junior Researcher at Inria	
Supervisor	Claire Chainais-Hillairet	Professor at Université de Lille	

Mots clés : analyse numérique, méthode d'entropie, méthodes numériques sur maillages généraux, méthodes numériques d'ordre élevé, semi-conducteurs, edp paraboliques

Keywords: numerical analysis, entropy method, numerical methods on general meshes, high-order numerical methods, semiconductors, parabolic pdes

Cette thèse a été préparée dans les laboratoires suivants.

Centre Inria de l'Université de Lille
Parc scientifique de la Haute-Borne
59650 Villeneuve d'Ascq
France

Laboratoire Paul Painlevé
Cité Scientifique
59655 Villeneuve-d'Ascq
France

Développement et analyse de schémas d'ordre élevé pour des modèles de convection-diffusion, étude du comportement en temps long

Résumé

Dans cette thèse, nous nous intéressons à l'approximation numérique de problèmes de convection-diffusion, potentiellement anisotropes, par des schémas d'ordre élevé sur maillages généraux. Notre objectif est de proposer des méthodes fiables, précises et efficaces : les solutions numériques doivent préserver les propriétés physiques des solutions continues (conservation de la masse, positivité des densités, comportement en temps long) tout en autorisant une large gamme de paramètres de discrétisation (pas de temps grands, maillages spatiaux généraux) et en optimisant la précision de calcul à coût donné. Les problèmes considérés sont des équations d'advection-diffusion ainsi que des systèmes couplés de dérive-diffusion qui modélisent les composants semi-conducteurs.

On se concentre d'abord sur une équation d'advection-diffusion seule, pour laquelle nous proposons et analysons trois méthodes d'ordre bas de type volumes finis hybrides (HFV). Cette comparaison met en avant la nécessité d'utiliser un schéma non-linéaire afin de préserver la positivité des solutions, tant d'un point de vue théorique que numérique. On s'intéresse alors à l'approximation d'un système de dérive-diffusion, constitué de deux équations d'advection-diffusion couplées avec une équation de Poisson. Pour ce problème, on introduit un schéma non-linéaire basé sur la méthode précédente, qui préserve les bornes des densités calculées (et en particulier leur positivité) et le comportement en temps long de la solution. Ce schéma HFV pour la dérive-diffusion est ensuite comparé numériquement avec un schéma présentant des propriétés similaires basé sur la méthode volumes finis en dualité discrète (DDFV).

Nous nous intéressons alors à des schémas d'ordre élevé (en espace). Ces schémas sont basés sur des méthodes de type hybride d'ordre élevé (HHO) qui peuvent être interprétées comme des extensions à l'ordre arbitraire des méthodes HFV. Nous introduisons deux méthodes pour les équations d'advection-diffusion linéaires. La première est linéaire, tandis que la seconde est non-linéaire, et permet de préserver la positivité. Pour ces deux schémas, nous prouvons l'existence de solutions discrètes et établissons des résultats de comportement en temps long. Nous confirmons également ces résultats numériquement, et mettons en avant la nécessité d'utiliser une méthode non-linéaire pour préserver la positivité. Par ailleurs, on observe que la méthode non-linéaire converge à l'ordre attendu. De plus, la montée en ordre permet un gain d'efficacité (précision de l'approximation pour un coût de calcul donné) conséquent par rapport aux méthodes d'ordre bas des premiers chapitres.

Ces travaux sont complétés par l'étude de problèmes de convection-diffusion avec convection très irrégulière, effectuée en collaboration avec des physiciens durant la thèse. Les recherches menées visent à comprendre comment concevoir des diodes électroluminescentes efficaces émettant dans l'ultraviolet profond, et soulèvent divers enjeux relatifs à la modélisation, l'analyse numérique et la simulation de ces problèmes.

Mots clés : analyse numérique, méthode d'entropie, méthodes numériques sur maillages généraux, méthodes numériques d'ordre élevé, semi-conducteurs, edp paraboliques

Centre Inria de l'Université de Lille – Parc scientifique de la Haute-Borne – 59650 Villeneuve d'Ascq – France
Laboratoire Paul Painlevé – Cité Scientifique – 59655 Villeneuve-d'Ascq – France

Development and numerical analysis of high-order schemes for convection-diffusion models, study of their long-time behaviour

Abstract

In this thesis, we are interested in the numerical approximation of anisotropic convection-diffusion problems using high-order schemes on general meshes. Our objective is to develop reliable, accurate, and efficient methods: the numerical solutions must preserve the physical properties of the continuous solutions (mass conservation, positivity of densities, long-time behaviour) while allowing for a wide range of discretisation parameters (large time steps, general spatial meshes) and optimising the accuracy at a given computational cost. The problems under study are advection-diffusion equations as well as coupled drift-diffusion systems that model semiconductor devices.

We first focus on a single advection-diffusion equation for which we propose and analyse three different hybrid finite volume (HFV) methods to approximate the solution. Their comparison highlights the necessity of using a nonlinear scheme to ensure the preservation of positivity, both from a theoretical and a numerical level. We then consider the numerical approximation of a drift-diffusion system, consisting of two coupled advection-diffusion equations with a Poisson equation. For this problem, we introduce a nonlinear scheme based on the previous method, which preserves the bounds on the computed densities (including their positivity) and the long-time behaviour of the solution. This HFV scheme for drift-diffusion is then numerically compared with another scheme with similar properties based on the discrete duality finite volume (DDFV) method.

We then focus on high-order (in space) schemes. These schemes are based on hybrid high-order (HHO) methods which can be interpreted as an arbitrary-order extension of HFV methods. We introduce two schemes for linear advection-diffusion: the first method is linear, while the second one is nonlinear and preserves the positivity of the solution. For both of these schemes, we prove the existence of discrete solutions and establish results about their long-time behaviour. We also confirm these results numerically and emphasise the need to use a nonlinear method to preserve positivity. We observe that the nonlinear method converges at the expected order. Furthermore, increasing the order leads to a significant gain in efficiency (accuracy of the approximation for a given computational cost) compared to the low-order methods discussed in the first chapters.

These works are complemented with the study of convection-diffusion problems with highly irregular convection, carried out in collaboration with physicists during the thesis. These investigations aim to understand how to design efficient deep ultraviolet-emitting electroluminescent diodes, and raise various issues related to the modeling, numerical analysis and simulation of these problems.

Keywords: numerical analysis, entropy method, numerical methods on general meshes, high-order numerical methods, semiconductors, parabolic pdes

Sommaire

Liste des tableaux

Table des figures

Avant-propos

Ce manuscrit présente les résultats obtenus durant la thèse réalisée au sein du centre Inria de l'Université de Lille et encadrée par Claire Chainais-Hillairet, Maxime Herda et Simon Lemaire. Ce document contient une introduction générale, qui décrit les modèles et méthodes numériques étudiés et dresse un panorama des contributions (en anglais). Cette introduction est suivie de quatre chapitres, correspondant chacun à des travaux publiés, et d'une conclusion dans laquelle nous exposons quelques perspectives. Le manuscrit est complété par quatre chapitres annexes, qui résultent de travaux réalisés en collaborations avec des mathématiciens et physiciens du Weierstrass Institute (Berlin, Allemagne) et du Tyndall National Institue (Cork, Irelande). L'Annexe A introduit les travaux et leur contextes physique et mathématique, tandis que les Annexes B-C-D rassemblent des travaux publiés ou soumis pour publication.

Liste des publications :

Chapitre 1 : *Long-time behaviour of hybrid finite volume schemes for advection-diffusion equations : linear and nonlinear approaches* ; en collaboration avec C. Chainais-Hillairet, M. Herda et Simon Lemaire ; paru [70] dans *Numerische Mathematik*, 151, 963–1016.

Chapitre 2 : *A structure preserving hybrid finite volume scheme for semiconductor models with magnetic field on general meshes* ; paru [213] dans *ESAIM : Mathematical Modelling and Numerical Analysis*, 57, 2557-2593.

Chapitre 3 : *Structure-preserving schemes for drift-diffusion systems on general meshes : DDFV vs HFV* ; en collaboration avec S. Krell ; accepté pour publication [187] comme *proceeding de la conférence Finite Volume for Complex Applications 10*.

Chapitre 4 : version étendue du proceeding *A skeletal high-order structure-preserving scheme for advection-diffusion equations* ; accepté pour publication [212] *à la conférence Finite Volume for Complex Applications 10*.

Annexe B : *Impact of random alloy fluctuations on the carrier distribution in multi-color (In,Ga)N/GaN quantum well systems* ; en collaboration avec M. O'Donovan, P. Farrell, T. Streckenbach, T. Koprucki et S. Schulz ; soumis pour publication [219].

Annexe C : *Importance of satisfying thermodynamic consistency in optoelectronic device simulations for high carrier densities* ; en collaboration avec P. Farrell, M. O'Donovan, S. Schulz et T. Koprucki ; paru [126] dans *Optical and Quantum Electronics*, 55, 978.

Annexe D : *Theoretical study of the impact of alloy disorder on carrier transport and recombination processes in deep UV (Al,Ga)N light emitters* ; en collaboration avec R. Finn, M. O'Donovan, P. Farrell, T. Streckenbach, T. Koprucki et S. Schulz ; paru [133] dans *Applied Physics Letters* 122 (24) : 241104.

Introduction générale

Sommaire du présent chapitre

Cette thèse est dédiée à l'approximation numérique de problèmes de convection-diffusion, éventuellement anisotropes, par des schémas d'ordre élevé sur maillage généraux. L'originalité principale de ces travaux est de traiter simultanément plusieurs enjeux : préservation de la structure, maillages généraux et ordre élevé. Bien que de nombreuses avancées aient été réalisées pour aborder chacun de ces enjeux individuellement, concevoir des méthodes qui présentent toutes ces caractéristiques simultanément est un travail difficile. Notre objectif est donc de proposer des méthodes fiables et efficaces : les solutions numériques doivent préserver les propriétés physiques des solutions continues tout en autorisant une large gamme de paramètres de discrétisation et en maintenant des coûts de calcul raisonnables.

Dans cette introduction générale, nous présentons dans un premier temps les modèles considérés dans ce manuscrit, en partant de différents contextes physiques, et en décrivant les propriétés (positivité et comportement en temps long) de leurs solutions. Cette présentation est l'occasion d'introduire la méthode d'entropie, outil fondamental des travaux de cette thèse. Nous nous intéressons ensuite à l'approximation numérique de ces modèles, en introduisant les grands concepts et enjeux de cette procédure, puis en proposant un état de l'art des différentes méthodes d'approximation pour les modèles de convection-diffusion. Enfin, nous dressons un

panorama (en anglais) des travaux théoriques (analyse numérique de schémas) et numériques (implémentation de schémas, investigation numérique de leurs propriétés et comparaison) réalisés au cours de cette thèse.

Phénomènes de convection-diffusion : modélisation, comportement et analyse

Les problèmes de convection-diffusion étudiés dans cette thèse constituent une brique de base pour décrire des phénomènes physiques et biologiques variés. Parmi les différents modèles complexes s'inscrivant dans le cadre discuté, on peut citer les modélisations d'écoulements en milieux poreux [17], l'étude de dispositif électronique comme les semi-conducteurs [258, 207, 208], les modélisations en biologie [135, 183, 179, 246], l'étude de la corrosion et de la dégradation de certains matériaux [3, 16, 53] ou encore des modèles de mécanique des fluides [206]. Tous ces phénomènes partagent une structure dissipative semblable. Ils échangent quelque chose avec leur environnement au cours du temps. Sous l'effet de ces échanges, le système va évoluer en vue d'atteindre un état plus stable. D'un point de vue mathématique, les modèles cités présentent tous un caractère diffusif [138, 130] et appartiennent à la grande classe des Équations aux Dérivées Partielles (EDP) paraboliques.

Modèles anisotropes d'advection-diffusion et de dérive-diffusion

Le premier modèle qui nous intéresse est une équation linéaire d'advection-diffusion, étudiée dans les Chapitres 1 et 4. Il s'agit d'un modèle dissipatif jouet, qui contient la structure de base des modèles dissipatifs. Il décrit l'évolution au cours du temps d'une quantité de matière dans un domaine borné ($\Omega \subset \mathbb{R}^d$, $d \in \{1,2,3\}$) qui est soumise à deux phénomènes : la diffusion d'une part (transfert naturel des zones de forte occupation vers celles faiblement occupées) et l'advection par un champ extérieur $V : \Omega \to \mathbb{R}^d$ (transport de la matière à cause d'une force externe). Le système mathématique obtenu pour l'inconnue notée $u : \mathbb{R}_+ \times \Omega \to \mathbb{R}$, qui représente la quantité de matière (ou sa concentration/densité), est le suivant :

$$\begin{cases} \partial_t u - \operatorname{div}(J(u)) = f & \text{dans } \mathbb{R}_+ \times \Omega, \\ J(u) = \Lambda(\nabla u + uV) & \text{dans } \mathbb{R}_+ \times \Omega, \\ u = u^D & \text{sur } \mathbb{R}_+ \times \Gamma^D, \\ J(u) \cdot n = g^N & \text{sur } \mathbb{R}_+ \times \Gamma^N, \\ u(0,\cdot) = u^{in} & \text{dans } \Omega. \end{cases} \tag{1}$$

Dans ce système, u^{in} est la donnée initiale, f est le terme source qui traduit l'apparition ou la disparition de matière dans le domaine. Les conditions de bord, de type Neumann sur Γ^N et Dirichlet sur Γ^D correspondent respectivement à une condition sur les flux au bord et une imposition des quantités de matière présentes. Le modèle est caractérisé par le flux d'advection-diffusion $J(u) = \Lambda(\nabla u + uV)$ qui décrit les échanges locaux de matière suivant la loi de Nernst–Planck [218, 227]. L'une des particularité de cette thèse est de s'intéresser à des modèles anisotropes : le flux dépend d'un tenseur $\Lambda : \Omega \to \mathbb{R}^{d\times d}$, qui traduit l'existence de directions privilégiées pour la diffusion, généralement à cause de la structure du milieu (comme c'est le cas dans les applications aux milieux poreux [17]) ou d'une force extérieure (comme dans les modèles de semi-conducteurs [143] présentés ci-dessous). Dans de nombreuses applications,

le champ d'advection dérive d'un potentiel $\phi : \Omega \to \mathbb{R}$: on peut écrire $V = \nabla\phi$, et

$$J(u) = \Lambda(\nabla u + u\nabla\phi).$$

Dans cette thèse, on s'intéresse également à des modèles de transport de charge dans des matériaux semi-conducteurs. Ces matériaux sont les constituants de base de nombreux dispositifs électroniques, notamment sous forme de diodes et de transistors qui constituent les circuits intégrés (et donc, par exemple, les processeurs et mémoires des ordinateurs ou des téléphones). Ils sont aussi utilisés dans des dispositifs opto-électroniques, comme les Diodes Électroluminescentes (DEL, light-emitting diode LED en anglais) ou les cellules photovoltaïques. Le modèle le plus simple pour les étudier est le système de dérive-diffusion introduit par Van Roosbroeck dans [258] en 1950. Il modélise l'évolution de charges électriques dans un matériau possédant une structure de réseau cristallin, qui constitue le semi-conducteur. L'idée fondamentale régissant ces structures est la suivante : les électrons (charges négatives) peuvent facilement se déplacer au sein du matériau, qui possède ainsi des propriétés intermédiaires entre un isolant électrique et un conducteur. Le transport des électrons laisse libres certains sites dans les bandes de valence du semi-conducteur : ce sont les trous d'électrons, aussi appelés simplement trous, qui correspondent à des charges positives. Au niveau mathématique, après adimensionnement, on obtient un système couplé d'inconnues N, P et ϕ vérifiant

$$\begin{cases} \partial_t N - \operatorname{div}(\nabla N - N\nabla\phi) = 0 & \text{dans } \mathbb{R}_+ \times \Omega, \\ \partial_t P - \operatorname{div}(\nabla P + P\nabla\phi) = 0 & \text{dans } \mathbb{R}_+ \times \Omega, \\ -\lambda^2 \operatorname{div}(\nabla\phi) = C + P - N & \text{dans } \mathbb{R}_+ \times \Omega, \\ N = N^D,\ P = P^D \text{ et } \phi = \phi^D & \text{sur } \mathbb{R}_+ \times \Gamma^D, \\ (\nabla N - N\nabla\phi)\cdot n = (\nabla P + P\nabla\phi)\cdot n = \nabla\phi \cdot n = 0 & \text{sur } \mathbb{R}_+ \times \Gamma^N, \\ N(0,\cdot) = N^{in} \text{ et } P(0,\cdot) = P^{in} & \text{dans } \Omega. \end{cases} \tag{2}$$

Dans ce système, $N : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ représente la densité de charges négatives (électrons), $P : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ la densité de charges positives (trous) et $\phi : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ est un potentiel électrostatique, induit par l'inhomogénéité spatiale de charge électrique dans le domaine. Les densités de charges suivent des lois de conservation décrites par des équations d'advection-convection, dont le champ $V_E = \pm\nabla\phi$ dérive du potentiel électrostatique et dépend de la charge de l'espèce considérée. Ce potentiel est lui-même solution d'une équation de Poisson, qui prend en compte la densité de charges totale, constituée de plusieurs contributions : les trous, les électrons et les impuretés (charges intrinsèques ajoutées volontairement dans le réseau cristallin).
Cet ajout de charge dans le matériau est nommé dopage et correspond à la fonction $C : \Omega \to \mathbb{R}$. Le dopage est un processus d'une importance capitale, car il permet de créer un dispositif aux propriétés voulues. Ainsi, le dopage C caractérise en partie le type de dispositif utilisé (diode, transistor, etc). Une illustration de géométrie typique de diode et son dopage associé est donnée en Figure 1, page 12.
Dans l'équation de Poisson qui régit le potentiel ϕ, la constante λ est nommée longueur de Debye (réduite). Elle est issue de l'adimensionnement du système, et dépend de certaines constantes physiques fondamentales (la conductivité diélectrique du milieu, la charge et la mobilité d'une particule élémentaire) et des caractéristiques du semi-conducteur étudié, dont la norme L^∞ du dopage et la taille du dispositif. Les valeurs typiques de λ varient de 10^{-3} à 10^{-5}.
Pour les modèles de semi-conducteurs, les conditions de bord sont des conditions mixtes Dirichlet-Neumann : les conditions de Dirichlet modélisent des contacts ohmiques (i.e. un contact avec un métal, conducteur, qui va imposer la quantité de charges au bord, ou avec

d'autres semi-conducteurs) et celles de Neumann correspondent à des parties isolées (sans contact avec l'extérieur, les charges ne peuvent ni rentrer ni sortir). .

Le modèle de Van Roosbroeck est un modèle simple, auquel ont peut ajouter des effets plus complexes. On peut notamment s'intéresser à l'effet d'un champ magnétique extérieur [143], à la prise en compte de diffusions non-linéaires [144, 257] ou encore prendre en compte les interactions entre les charges via des termes de réaction non-linéaires. La motivation principale de cette thèse est l'approximation du système ci-dessous, qui prend en compte ces différents effets :

$$\begin{cases} \partial_t N - \operatorname{div}(N\Lambda_N \nabla(h(N)-\phi)) = -R(N,P) & \text{dans } \mathbb{R}_+ \times \Omega \\ \partial_t P - \operatorname{div}(P\Lambda_P \nabla(h(P)+\phi)) = -R(N,P) & \text{dans } \mathbb{R}_+ \times \Omega \\ -\operatorname{div}(\Lambda_\phi \nabla\phi) = C + P - N & \text{dans } \mathbb{R}_+ \times \Omega \\ N = N^D,\ P = P^D \text{ et } \phi = \phi^D & \text{sur } \mathbb{R}_+ \times \Gamma^D \\ N\Lambda_N \nabla(h(N)-\phi)\cdot n = P\Lambda_P \nabla(h(P)+\phi)\cdot n = \Lambda_\phi \nabla\phi \cdot n = 0 & \text{sur } \mathbb{R}_+ \times \Gamma^N \\ N(0,\cdot) = N^{in} \text{ et } P(0,\cdot) = P^{in} & \text{dans } \Omega, \end{cases} \tag{3}$$

La nonlinéarité h dans les équations sur les charges est obtenue à partir de modèles cinétiques [172]. On peut alors décrire les flux sur les charges comme étant le produit de densités par des potentiels dits de quasi-Fermi w_N et w_P :

$$J_N = -N\nabla w_N \quad \text{et} \quad J_P = -P\nabla w_P.$$

Ces potentiels jouent un rôle fondamental dans le modèle, tant d'un point de vue physique que mathématique. Ils sont liés aux densités des espèces à travers les équations d'état

$$N = g(\phi - w_N) \quad \text{et} \quad P = g(-\phi + w_P), \tag{4}$$

où la fonction $g : \mathbb{R} \to \mathbb{R}_+$ est la fonction de distribution associée à la statistique. C'est une fonction strictement monotone, de bijection réciproque h. On peut donc exprimer les potentiels de quasi-Fermi en fonction des densités

$$w_N = h(N) - \phi \quad \text{et } w_P = h(P) + \phi,$$

et obtenir les équations de (3). Dans le cas du modèle de Van Roosbroeck, la statistique considérée est celle de Boltzmann, qui correspond à $g = \exp$ et $h = \log$; les termes de diffusion qui en découlent sont linéaires. Le type de statistique utilisée dans la modélisation dépend de la nature du semi-conducteur (organique ou non) et du cadre d'application [144, 257, 125]. Les expressions de g sont généralement complexes, mais il existe des approximations analytiques couramment utilisées [31, 223].
La présence des tenseurs d'anisotropie Λ_N et Λ_P est liée à l'existence d'un champ magnétique extérieur, par exemple dans le cadre d'étude de capteurs [234]. Ce champ va faire tourner les charges et ainsi modifier le comportement global du semi-conducteur. Un modèle pour de tels dispositifs a été introduit (dans le cas de la statistique de Boltzmann) par Gajewski et Gärtner dans [143], puis obtenu comme limite rigoureuse d'un modèle cinétique [23, 155]. Le champ magnétique extérieur est décrit par un champ de vecteur $B : \Omega \to \mathbb{R}^d$, qui modifie les flux d'électrons et de trous en rajoutant un terme dû à la contribution magnétique de la force de Lorentz. Des manipulations algébriques permettent ensuite d'exprimer directement les flux avec champ magnétique comme des produits de tenseurs Λ_N et Λ_P par les flux de charges sans champ magnétique. Pour un problème bidimensionnel, l'effet de rotation ne peut avoir lieu que dans le plan du semi-conducteur, et on peut se restreindre au cas $B = be_z$, où e_z est un vecteur

orthogonal au semi-conducteur et b est l'intensité du champ magnétique. Les tenseurs sont alors de la forme

$$\Lambda_N = \frac{1}{1+b^2}\begin{pmatrix} 1 & b \\ -b & 1 \end{pmatrix} \quad \text{et} \quad \Lambda_P = \frac{1}{1+b^2}\begin{pmatrix} 1 & -b \\ b & 1 \end{pmatrix}. \tag{5}$$

Pour un problème en trois dimensions de l'espace, en notant $B = (b_1, b_2, b_3)$, on a

$$\Lambda_N = \frac{1}{1+|B|^2}\begin{pmatrix} 1+b_1^2 & b_1b_2 - b_3 & b_1b_3 + b_2 \\ b_1b_2 + b_3 & 1+b_2^2 & b_2b_3 - b_1 \\ b_1b_3 - b_2 & b_2b_3 + b_1 & 1+b_3^2 \end{pmatrix}$$

et

$$\Lambda_P = \frac{1}{1+|B|^2}\begin{pmatrix} 1+b_1^2 & b_1b_2 + b_3 & b_1b_3 - b_2 \\ b_1b_2 - b_3 & 1+b_2^2 & b_2b_3 + b_1 \\ b_1b_3 + b_2 & b_2b_3 - b_1 & 1+b_3^2 \end{pmatrix}.$$

L'une des particularités de ces tenseurs est leur caractère non-symétrique.
Dans les équations de transport de charges, les termes $R(N,P)$ sont des termes de réaction, qui traduisent des interactions entre les charges et avec le milieu. En particulier, à cause de l'agitation thermique interne ou d'une stimulation externe (par exemple optique ou électrique) il existe un phénomène d'apparition de charges dans le dispositif, nommé génération : les électrons de valence sont arrachés pour devenir des électrons de conduction (libres). D'autre part, les électrons et les trous peuvent interagir entre eux via un phénomène nommé recombinaison : un électron de conduction peut devenir un électron de valence et neutraliser un trou. Ces termes ont une structure particulière, de la forme

$$R(N,P) = r(N,P)\left(\mathrm{e}^{h(N)+h(P)} - 1\right), \tag{6}$$

où r est une fonction positive. Ces effets de génération/recombinaison sont fondamentaux lorsque l'on s'intéresse à des dispositifs opto-électroniques, et la compétition entre eux a une influence sur le rendement des dispositifs. Ces questions font l'objet de l'Annexe D.

Positivité des solutions et conservation de la masse

L'équation linéaire (1) a été longuement étudiée et son analyse est maintenant classique [119, Section 7.1]. On sait en particulier montrer l'existence et l'unicité d'une solution globale, et évaluer sa régularité en fonctions des données.
Parmi les propriétés remarquables des solutions figure la positivité :

$$\text{si les données } u^{in}, f, g^N \text{ et } u^D \text{ sont positives, alors } u \geq 0 \text{ dans } \mathbb{R}_+ \times \Omega.$$

Il suffit d'imposer en plus que $\int_\Omega u^{in} > 0$ pour assurer la stricte positivité de la solution. Ces propriétés sont essentielles pour assurer une cohérence avec la physique du problème.
Une autre propriété remarquable concerne la préservation de la masse : si l'on impose des conditions de Neumann sur tout le bord du domaine ($\Gamma^N = \partial\Omega$), alors sous une hypothèse de compatibilité entre les données $\int_\Omega f + \int_{\partial\Omega} g^N = 0$, la masse de la solution est conservée au cours du temps :

$$\forall t > 0, \quad \int_\Omega u(t) = \int_\Omega u^{in}.$$

Concernant les systèmes de dérive-diffusion, l'existence et l'unicité de solutions globales ont été prouvées dans divers contextes, sous des hypothèses raisonnables sur les données (dont la

positivité des densités initiales et des densités sur les contacts ohmiques). On peut notamment citer les travaux fondateurs [214, 142, 208] pour le modèle de Van Roosbroeck (2), des études pour des statistiques non-Boltzmann [144, 149] et une analyse pour un modèle anisotrope [143]. L'étude du système (3) peut être abordée en suivant les arguments présentés dans ces travaux.
Comme pour le modèle d'advection-diffusion, les solutions des systèmes de dérive-diffusion satisfont une propriété de positivité : les densités de charges N et P sont strictement positives sur le domaine en tout temps. Pour certaines statistiques non-Boltzmann (typiquement une statistique de Blakemore qui correspond à une non-linéarité de la forme $h(s) = \log\left(\frac{s}{1-s}\right)$, on a également une borne supérieure a priori sur les densités donnée par $a = \sup_{\mathbb{R}} h$. Ainsi, en général

$$\forall t > 0,\ 0 < N(t) < a \ \text{ et } \ 0 < P(t) < a,$$

où $a \in]0, +\infty]$.

Comportement en temps long et méthode d'entropie

Comme annoncé précédemment, les phénomènes considérés sont caractérisés par une évolution vers des états stationnaires. L'objectif de cette section est de présenter la méthode d'analyse emblématique pour étudier ce phénomène.

Pour l'équation d'advection-diffusion (1), l'état stationnaire associé u^∞ est défini par

$$\begin{cases} -\operatorname{div}\left(\Lambda\left(\nabla u^\infty + u^\infty V\right)\right) = f & \text{dans } \Omega, \\ u^\infty = u^D & \text{sur } \Gamma^D, \\ \left(\Lambda\left(\nabla u^\infty + u^\infty V\right)\right)\cdot n = g^N & \text{sur } \Gamma^N. \end{cases} \tag{7}$$

Dans le cas où $\Gamma^D = \emptyset$, pour assurer une cohérence avec le problème évolutif, la masse de u^∞ doit aussi être égale à la masse initiale :

$$\int_\Omega u^\infty = \int_\Omega u^{in} = M.$$

L'état stationnaire est bien défini et unique [106, 114]. De plus, il a les mêmes propriétés de positivité que la solution transitoire.
Notons qu'il existe des états stationnaire particuliers, qui correspondent à des états physiques très stables, pour lesquels les flux sont nuls. Ces états stationnaires sont nommés équilibres thermiques. Lorsque le champ dérive d'un potentiel $V = \nabla\phi$, les équilibres sont associés à des potentiels de quasi-Fermi constants sur Ω. En pratique, pour l'équation (1) de tels état stationnaires ne peuvent exister que si les données sont compatibles avec l'équilibre. Typiquement, un terme source nul $f = 0$ et des conditions de Neumann homogènes sur tout le bord ($\Gamma^N = \partial\Omega$ avec $g^N = 0$) donnent un équilibre thermique. Lorsque ces conditions ne sont pas satisfaites, on obtient un état stationnaire dit hors-équilibre.

Pour le modèle d'advection-diffusion (1), la solution $u(t)$ converge exponentiellement vite, quand $t \to +\infty$, vers u^∞, solution de (7). Un tel résultat se généralise à d'autre modèles dissipatifs, dont les systèmes de dérive-diffusion (2)-(3). Pour démontrer un tel comportement, on utilise la méthode d'entropie. Celle-ci a initialement été introduite dans le cadre de la théorie cinétique des gaz par Boltzmann dans [34], à travers le théorème H. Comme présenté dans [8], cette méthode a par la suite été utilisée dans des cadres plus variés, notamment pour l'étude d'équations d'advection-diffusion [62, 254, 63, 33], ou plus généralement d'équations paraboliques [61]. On renvoie à l'ouvrage de Jüngel [174] pour une explication détaillée, et décrivons ci-dessous la

philosophie générale de la méthode.

(i) **Identification d'une entropie relative** : on définit une quantité, motivée par la physique, qui dépend de la solution et de l'état stationnaire, nommée entropie relative $\mathbb{E}$. Cette quantité est positive, et doit mesurer dans un certain sens l'écart entre u et u^∞. En particulier, l'entropie relative s'annule quand la solution transitoire coïncide avec l'état stationnaire.

(ii) **Obtention d'une relation d'entropie** : on dérive l'entropie relative par rapport au temps afin d'obtenir une relation sur l'entropie, parfois nommée relation d'entropie/dissipation d'entropie, ou simplement relation de dissipation d'entropie, de la forme :

$$\frac{d}{dt}\mathbb{E}(t) = -\mathbb{D}(t), \tag{8}$$

où $\mathbb{D}$ est une quantité positive nommée dissipation d'entropie ou information de Fischer. Cette relation implique la décroissance de l'entropie au cours du temps : l'entropie relative s'interprète alors comme une fonctionnelle de Lyapunov du problème.

(iii) **Obtention d'un contrôle de l'entropie par sa dissipation** : on compare l'entropie relative avec sa dissipation, pour obtenir une estimation du type

$$\forall t > 0,\ \nu\mathbb{E}(t) \leq \mathbb{D}(t), \tag{9}$$

où ν est une constante positive. Cette estimation s'obtient à l'aide d'inégalités fonctionnelles. On en déduit une décroissance exponentielle de l'entropie :

$$\forall t > 0,\ \mathbb{E}(t) \leq \mathrm{e}^{-\nu t}\,\mathbb{E}(0). \tag{10}$$

(iv) **Obtention d'un contrôle de la distance entre u et u^∞ par l'entropie** : pour conclure, on trouve une norme $\|\cdot\|$ sur un espace adapté tel que

$$\forall t > 0,\ \|u(t) - u^\infty\|^\gamma \leq C\mathbb{E}(t), \tag{11}$$

où γ et C sont des constantes strictement positives. On en déduit alors de (10) une convergence exponentiellement rapide de $\|u(t) - u^\infty\|$ vers 0 quand $t \to \infty$.

Pour l'équation (1), on peut par exemple partir de l'énergie quadratique pour définir une entropie relative

$$\mathbb{E}_2 = \frac{1}{2}\|u - u^\infty\|^2_{L^2(\Omega)},\ \text{de dissipation}\ \mathbb{D}_2 = \int_\Omega \Lambda\nabla(u - u^\infty)\cdot\nabla(u - u^\infty).$$

Une autre possibilité est d'utiliser l'entropie de Boltzmann (ou plus généralement des entropies de Tsallis [33])

$$\mathbb{E}_B = \int_\Omega u^\infty \Phi_1\left(\frac{u}{u^\infty}\right),\ \text{de dissipation}\ \mathbb{D}_B = \int_\Omega u(t)\Lambda\nabla(w - w^\infty)\cdot\nabla(w - w^\infty)$$

où $\Phi_1(s) = s\log(s) - s + 1$, $w = \log(u) + \phi$ et $w^\infty = \log(u^\infty) + \phi$. Cette entropie correspond à une quantité physique nommée énergie interne. Notons que la relation d'entropie (8) donne également des bornes sur l'entropie et sa dissipation. Il est souvent possible d'en déduire des estimations sur la solution elle-même. Cette stratégie est l'un des outils essentiels de cette thèse, permettant d'obtenir des estimations a priori sur des solutions discrètes.

Pour les modèles de semi-conducteur, on peut également étudier le comportement en temps long via une méthode d'entropie. La notion d'état stationnaire étant plus délicate à définir

(notamment de part leur non-unicité [215, 216]), l'analyse du comportement en temps long est restreinte à un cadre simple où l'état stationnaire est un équilibre thermique (N^e, P^e, ϕ^e). Comme pour l'équation d'advection-diffusion, celui-ci est caractérisé par des potentiels de quasi-Fermi constants :

$$h(N^e) - \phi^e = \alpha_N \quad \text{et} \quad h(P^e) + \phi^e = \alpha_P \text{ dans } \Omega,$$

où $(\alpha_N, \alpha_P) \in \mathbb{R}^2$. Cette relation impose naturellement une condition sur les données au bord :

$$h(N^D) - \phi^D = \alpha_N \text{ et } h(P^D) + \phi^D = \alpha_P \text{ sur } \Gamma^D. \tag{12}$$

En lien avec l'intuition physique, on impose également qu'il n'y ait plus de phénomène de recombinaison/génération lorsque le semi-conducteur est à l'équilibre, ce qui signifie que $R(N^e, P^e) = 0$. En revenant à la définition (6), ceci impose une autre condition de compatibilité dans le cas où $r \neq 0$:

$$\alpha_N + \alpha_P = 0. \tag{13}$$

Sous ces conditions, on obtient une relation simple entre les densités de charge et le potentiel à l'équilibre :

$$N^e = g(\alpha_N + \phi^e) \quad \text{et} \quad P^e = g(\alpha_P - \phi^e).$$

On en déduit que le potentiel à l'équilibre ϕ^e est une solution de l'équation de Poisson non-linéaire (nommée équation de Poisson-Boltzmann dans le cas de la statistique de Boltzmann $g = \exp$) suivante :

$$\begin{cases} -\operatorname{div}(\Lambda_\phi \nabla \phi^e) = C + g(\alpha_P - \phi^e) - g(\alpha_N + \phi^e) & \text{dans } \Omega, \\ \phi^e = \phi^D & \text{sur } \Gamma^D, \\ \Lambda_\phi \nabla \phi^e \cdot n = 0 & \text{sur } \Gamma^N. \end{cases} \tag{14}$$

Les propriétés de g permettent alors de montrer que (14) admet une unique solution ϕ^e. Ceci implique qu'il existe un unique équilibre thermique (N^e, P^e, ϕ^e). Sous ces hypothèses, l'analyse du comportement en temps long est alors menée pour différentes situations grâce à une méthode d'entropie [144, 143]. L'entropie relative utilisée pour le système (3) est définie par

$$\mathbb{E}(t) = \int_\Omega H(N) - H(N^e) - h(N^e)(N - N^e) + \int_\Omega H(P) - H(P^e) - h(P^e)(P - P^e) + \frac{1}{2} \int_\Omega \nabla(\phi - \phi^e) \cdot \Lambda_\phi \nabla(\phi - \phi^e),$$

où $H : x \to \int_1^x h(s)ds$. Les deux premiers termes sont des énergies internes (entropies de type Boltzmann sur les densités), et le dernier terme est une énergie électrique. On a alors la relation de dissipation d'entropie

$$\frac{d\mathbb{E}}{dt}(t) = -\mathbb{D}(t),$$

où le terme de dissipation d'entropie s'écrit

$$\mathbb{D}(t) = \int_\Omega N \nabla(h(N) - \phi) \cdot \Lambda_N \nabla(h(N) - \phi) + \int_\Omega P \nabla(h(P) + \phi) \cdot \Lambda_P \nabla(h(P) + \phi) + \int_\Omega R(N, P)\,(h(N) + h(P)).$$

Le dernier terme est lié aux processus de recombinaison/génération. Il est positif en vertu de la structure (6) des termes de réaction.

Approximation numérique des problèmes de convection-diffusion

Afin de comprendre les enjeux de l'approximation numérique, il est bon de rappeler que pour les modèles jouets (1), (2) et (3), les résultats théoriques ne fournissent pas d'expression explicite de la solution. C'est d'autant plus vrai pour des modèles plus réalistes utilisés dans l'industrie ou la recherche (tels que les modèles utilisés actuellement par des équipes de recherche en opto-électronique, c.f. Annexes B-C-D). Pour pouvoir exploiter pleinement les modèles mathématiques proposés, il faut donc être capable d'en tirer des informations quantitatives, en calculant une solution approchée qui pourra être exploitée au niveau industriel. A titre d'exemple, la conception de dispositifs modernes comme les LEDs émettant en UV profond (étudiées en Annexe D) nécessite beaucoup de tests par essai-erreur avant d'obtenir des résultats probants (par exemple, pour obtenir des LEDs efficaces en termes de consommation d'énergie). De tels tests ne peuvent se faire expérimentalement, à cause du nombre de configurations et du coût (économique et temporel) de chaque dispositif. La simulation numérique est alors la solution privilégiée pour réaliser ces tests à moindre coût, mais il faut être capable de garantir que la solution calculée est correcte.

Le travail réalisé dans cette thèse s'inscrit dans cette logique : nous proposons des méthodes numériques pour approcher les solutions des modèles (1), (2) et (3). Les méthodes introduites sont alors analysées théoriquement et étudiées au niveau numérique.

Dans cette section, nous présentons les enjeux majeurs de l'analyse numérique, et dressons un panorama des méthodes existantes pour approcher les problèmes étudiés dans cette thèse. L'accent est mis sur la question de la préservation des propriétés des solutions continues au niveau discret.

Enjeux de fiabilité de l'approximation

Pour les méthodes numériques qui nous intéressent, les solutions approchées sont obtenues via la discrétisation du domaine spatial en un maillage (partition en sous-domaines polyédriques nommés mailles ou volumes de contrôle). Dans l'idéal, on souhaite avoir une grande liberté lors de la conception du maillage, pour pouvoir par exemple avoir des mailles de formes différentes, ou des nœuds orphelins. Les questions relatives aux maillages sont cruciales dans certaines applications, comme en géosciences où le maillage est souvent imposé par la connaissance que l'on a du sol : il s'agit d'une donnée du problème. Un autre intérêt des maillages généraux est la facilité avec laquelle ils permettent de réaliser des raffinements locaux. Même pour des simulations de semi-conducteurs (associés à des dispositifs aux géométries assez simples, souvent cartésiennes), il peut être très intéressant d'avoir un maillage adapté localement (voir Figure 1 avec le cas d'une diode où les les zones d'intérêt sont raffinées) afin d'optimiser les coûts de calculs. Pour finir, les capacités de prise en compte d'un maillage général ou de la diffusion anisotrope sont fortement liées.

La qualité d'un schéma numérique est souvent évaluée par la précision du schéma (sur un maillage fixé). Cependant, un tel critère ne saurait être suffisant pour garantir une approximation fiable des phénomènes simulés. En effet, l'analyse théorique des modèles continus fournit des propriétés qualitatives importantes, qu'il convient de préserver au niveau discret. Les schémas qui suivent ce principe sont nommés de "schémas préservant la structure" (structure-preserving schemes en anglais). L'enjeu majeur de cette thèse est le développement de schémas préservant la structure d'entropie des problèmes de convection-diffusion, et ce sans condition sur le maillage

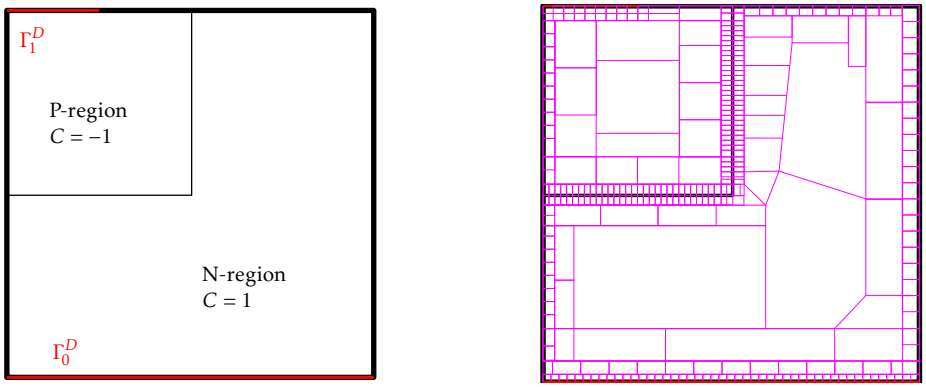


FIGURE 1 – Géométrie d'une diode et maillage adapté associé.

spatial ou le pas de temps utilisés. L'importance de la relation de dissipation d'entropie au niveau continu permet en particulier d'obtenir de nombreuses autres préservations, que nous détaillons ci-dessous.

Positivé des solutions discrètes. On peut considérer qu'obtenir une densité approchée négative est une aberration vis-à-vis de la physique. D'un point de vue pratique, l'exploitation d'une densité non positive est difficile, voire impossible. D'un point de vue théorique, la question de la préservation de la positivité peut aussi être très importante pour l'analyse. Une telle préservation s'obtient au moyen d'une entropie qui est affectée par le signe de la solution, comme l'entropie de Boltzmann.

Préservation des équilibres thermiques. La question des équilibres thermiques est elle aussi capitale, au vu du rôle joué par ces états stationnaires particuliers dans les modèles et leur analyse. Définir précisément une notion de schéma préservant les équilibres thermiques n'est pas aisé, et plusieurs définitions seront proposées et utilisées dans cette thèse, en fonction des méthodes considérées. Pour une équation d'advection-diffusion (1), une telle propriété correspond essentiellement au fait que l'état stationnaire discret pour des données compatibles avec l'équilibre est, en un certain sens, une interpolation de l'équilibre thermique continu. Pour les systèmes de dérive-diffusion, ceci correspond au fait que l'état stationnaire discret peut être interprété comme une approximation d'une solution de l'équation de Poisson-Boltzmann (14). En règle générale, l'annulation du terme de dissipation d'entropie discrète permet de caractériser les équilibres discrets.

Comportement en temps long des solutions discrètes. Pour les modèles étudiés ici, le comportement asymptotique $t \to \infty$ est très particulier : les solutions convergent exponentiellement vite vers des états stationnaires. Concevoir des schémas dont les solutions vérifient une propriété analogue est donc un gage conséquent de fiabilité. Pour certaines applications qui s'intéressent à des phénomènes évoluant sur des échelles de temps très longues, comme les phénomènes de corrosion [16, 53], la question du temps long est cruciale car elle assure la fiabilité des simulations. Un schéma avec un bon comportement en temps long permet également de calculer des états stationnaires à moindre coût. Si l'on considère l'exemple des semi-conducteurs, on s'intéresse généralement aux courbes courant-tension, qui concernent des états stationnaires. Pour simuler ces dispositifs, on a recours à des schémas (non-linéaires) qui nécessitent d'être initialisés par des données suffisamment proches de la solution. On a alors recours à des procédés de continuation. Une des possibilités consiste à utiliser cette initialisation comme une donnée initiale, et laisser évoluer

(en temps) le problème. Si la solution discrète converge exponentiellement vite vers un état stationnaire discret, on obtient une approximation de l'état stationnaire en très peu d'itérations en temps.

Enfin, pour les modèles où la masse est préservée (par exemple, advection-diffusion avec conditions de bord de Neumann homogènes), il est également important d'assurer cette caractéristique au niveau discret. Comme pour la positivité, cette conservation peut être utilisée à des fins d'analyse théorique des schémas.

Volumes finis à deux points : forces et faiblesses

Nous présentons maintenant une classe de schémas qui répond à presque tous les critères discutés à la section précédente, massivement utilisée pour simuler des modèles de semiconducteurs : les méthodes volumes finis à deux points. Ces méthodes sont basées sur une discrétisation du flux de matière. En définissant de manière adaptée les flux numériques, on obtient des schémas très robustes, puisqu'ils s'appuient sur des quantités qui portent un sens physique. Initialement introduites pour des problèmes hyperboliques [188, 150], l'utilisation de ces méthodes s'est ensuite répandue à des problèmes elliptiques et paraboliques [121].

L'idée des schémas volumes finis deux points (Two Points Flux Approximation, TPFA en anglais) pour les problèmes elliptiques est d'introduire, pour une discrétisation $\mathcal{D}$ du domaine donné, des inconnues discrètes

$$\underline{u}_{\mathcal{D}} = (u_K)_{K\in\mathcal{M}},$$

où $\mathcal{M}$ est le maillage et u_K l'inconnue de maille. Le problème continu $\operatorname{div}(J(u)) = f$ est alors discrétisé en intégrant l'équation sur chaque maille $K \in \mathcal{M}$, et en utilisant une formule de Green :

$$\forall K \in \mathcal{M}, \int_{\partial K} J(u)\cdot n = \int_K f.$$

Muni des inconnues de mailles, on approche le flux normal $J(u)\cdot n$ sur la face σ par un flux numérique $F_{K,\sigma}(\underline{u}_{\mathcal{D}})$. On obtient finalement le schéma

$$\forall K \in \mathcal{M}, \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(\underline{u}_{\mathcal{D}}) = \int_K f,$$

où $\mathcal{E}_K$ est l'ensemble des faces de la maille K. Dans le cas où le flux est diffusif ($J(u) = \Lambda\nabla u$), on utilise une discrétisation différences finies des gradients. Cette méthode est alors très simple à implémenter, et préserve les propriétés de la solution continue. Elle souffre néanmoins d'une limitation intrinsèque : pour être consistant avec les problèmes continus, il faut se placer sur un maillage dit admissible vérifiant une condition d'orthogonalité (pour le produit scalaire induit par le tenseur de diffusion Λ) [121, Définition 3.1]. En pratique, même pour des tenseurs anisotropes relativement simples, il est très compliqué de construire de tels maillages admissibles. Lorsque l'on considère des problèmes plus compliqués (diffusion anisotrope et hétérogène, plusieurs tenseurs de diffusion pour un système couplé), de tels maillages peuvent tout simplement ne pas exister. Ainsi, l'usage de ces schémas est essentiellement restreint à des cas isotropes, sur des maillages avec des géométries simples (et sans nœuds orphelins).

Les schémas TPFA sont l'objet d'une abondante littérature. Pour approcher l'équation d'advection-diffusion linéaire (1) (avec $\Lambda = I_d$), le flux de Scharfetter–Gummel [239, 160, 73] est une solution privilégiée : il permet de gérer des advections très fortes (ce flux peut être interprété comme une interpolation d'un flux parabolique et d'un flux hyperbolique, obtenue grâce à la fonction de Bernoulli $B(s) = \frac{s}{e^s-1}$), préserve la positivité et les équilibres thermiques au niveau

discret. Il présente également un bon comportement en temps long, tout comme certaines de ses généralisations [131, 69].
Concernant les modèles de dérive-diffusion isotropes, on peut également citer les travaux fondateurs de [239]. Suivant cette approche, de nombreuses recherches autour de schémas TPFA positifs préservant le comportement en temps longs ont été menées [68, 27, 28]. Notons également que la question de généraliser le flux de Scharfetter–Gummel à des statistiques non-Boltzmann a été un sujet actif [265, 120, 26, 140, 50]. Toutes ces méthodes préservent la positivité et l'équilibre thermique. De nombreux travaux comparatifs (numériques et théoriques) [125, 127, 50] ont également permis d'identifier les atouts et faiblesses de chacune de ces méthodes.

Malheureusement, tous les résultats énoncés plus haut ne sont valables que pour des problèmes isotropes sur maillages orthogonaux. Ainsi, il n'existe pas de schéma satisfaisant pour le modèle anisotrope (3). La seule tentative connue est le schéma TPFA de [143] pour la statistique de Boltzmann ($h = \log$), qui repose sur une perturbation des flux de Scharfetter–Gummel pour gérer l'anisotropie. Les auteurs parviennent à montrer que la positivité des solutions est préservée sous certaines conditions liant la géométrie du maillage avec l'intensité du champ magnétique. En pratique, ces conditions sont extrêmement restrictives et difficiles à vérifier : les résultats numériques présentés indiquent clairement une perte de la positivité au niveau discret.

Schémas pour problèmes anisotropes sur maillages généraux

Afin de gérer l'anisotropie et régler le défaut majeur des schémas TPFA, de nombreuses méthodes de type volumes finis ont été développées depuis les les vingt dernières années [157, 105, 40, 39, 110, 123, 124] pour discrétiser des problèmes de diffusion anisotrope. Par rapport aux schémas TPFA, ces méthodes reposent sur l'ajout d'inconnues auxiliaires qui permettent de gérer l'anisotropie de la diffusion et la généricité du maillage, via la reconstruction d'un opérateur gradient discret. Elles peuvent cependant s'exprimer grâce à des flux numériques, qui ne sont plus basés sur une formulation différences finies entre deux points. La plupart de ces méthodes peuvent s'inscrire dans le formalisme des "schémas gradients" [111]. Ces méthodes permettent d'approcher les problèmes anisotropes avec une grande robustesse, mais celle-ci s'obtient au prix d'une perte de la positivité des solutions numériques, comme obervé lors du benchmark de la conférence FVCA 5 [154].

L'utilisation de ces méthodes a ensuite été généralisée à des problèmes de convection-diffusion qui nous intéressent dans ce manuscrit. Pour le problème d'advection-diffusion linéaire (1), un schéma Hybride Mimétique Mixte (HMM) a été proposé et analysé [19]. Ce schéma, étudié dans le Chapitre 1, est conçu comme une généralisation des B-schémas TPFA [66]. Cependant, il ne préserve ni la positivité ni les équilibres thermiques. Concernant les modèles de semi-conducteurs isotropes, des schémas Volumes Finis en Dualité Discrète (Discrete Duality Finite Volume, DDFV en anglais) on été proposés [64] pour supporter les maillages généraux sur des domaines bi-dimensionnels. On peut aussi citer le schéma HMM [81] pour un modèle de plasma (qui est essentiellement un système de dérive diffusion en statistique de Boltzmann avec une anisotropie sur l'une des deux espèces). Ces schémas ne préservent pas la positivité des densités, ce qui restreint leur analyse.

Afin de palier le défaut de positivé des schémas précédemment cités, différentes stratégies ont été proposées. Elles reposent sur l'introduction d'une nonlinéarité dans le schéma.
On peut notamment citer les méthodes [245, 113, 32, 48] qui reposent sur l'ajout d'une nonlinéarité dans le flux numérique. Ces méthodes s'avèrent robustes vis-à-vis de l'anisotropie et des maillages pour des problèmes de diffusion linéaire, et l'existence et la consistance des flux sont prouvées [32]. Plus récemment, une méthode pour des systèmes de Poisson–Nernst–Planck (un

modèle pour décrire l'évolution d'ions dans une solution, semblable au système (2) mais avec des conditions de bord de type Neumann sur tout le bord du domaine) sur maillage généraux a été introduit et partiellement analysé [250] en suivant cette méthodologie.
Une autre solution a été proposée par Cancès et Guichard [55, 56] pour une classe d'équations non-linéaires contenant l'équation d'advection-diffusion linéaire : discrétiser le flux d'advection-diffusion

$$J(u) = \Lambda\,(\nabla u + u\nabla\phi)$$

sous la forme non-linéaire suivante

$$J(u) = u\Lambda\nabla(\log(u) + \phi).$$

Par rapport aux méthodes présentées plus haut, l'avantage de cette stratégie est qu'elle repose sur une non-linéarité qui a un sens au niveau continu. Celle-ci donne notamment au schéma une structure d'entropie (de Boltzmann), qui permet de proposer une analyse (existence de solutions positives et convergence). De plus, il est assez simple de généraliser ce schéma à des diffusions non-linéaires (correspondant à des modèles avec des statistiques non-Boltzmann). Suivant ce principe, un schéma non-linéaire Volumes Finis en Dualité Discrète a été dévéloppé et analysé [52, 51]. Grâce à des inégalités fonctionnelles de type Sobolev logarithmique, une analyse du comportement en temps long reposant sur l'entropie de Boltzmann est menée.

Méthodes d'ordre élevé

Nous nous intéressons maintenant aux méthodes d'ordre élevé. L'idée générale de ces méthodes est d'utiliser des inconnues discrètes polynomiales pour obtenir une précision accrue, pour peu que les solutions continues soient assez régulières. Ce gain de précision est généralement accompagné d'un gain d'efficacité : pour un coût de calcul donné, on pourra atteindre une meilleure précision avec une méthode d'ordre élevée qu'avec une méthode d'ordre bas.

Bien qu'il existe de nombreuses méthodes d'ordre élevé pour les problèmes diffusifs, dont les éléments finis $\mathbb{P}^k$, nous nous concentrons ici sur des méthodes conservatives, qui peuvent s'interpréter comme des généralisations des méthodes d'ordre bas présentées précédemment pour des problèmes de diffusion linéaires (pouvant être anisotropes et posés sur maillages généraux). Historiquement, les premières méthodes introduites sont les méthodes Galerkin Discontinues [13, 9, 10] (Discontinuous Galerkin, DG en anglais). On peut interpréter ces schémas comme une généralisation à l'ordre élevé des méthodes volumes finis TPFA : le schéma repose sur un équivalent discret du flux entre chaque maille. Cette caractéristique permet d'assurer le caractère conservatif de la méthode. Pour en améliorer l'efficacité, des versions hybrides ont été introduites [85, 87]. Ces méthodes de Galerkin Discontinue Hybrides (Hybridisable Discontinuous Galerkin, HDG en anglais) reposent sur l'ajout d'inconnues sur les faces afin de découpler les inconnues de mailles. Ceci permet de réaliser une élimination des inconnues de mailles (condensation statique) et ainsi obtenir un gain d'efficacité conséquent pour des ordres et dimensions élevés. Une approche similaire, en utilisant des stabilisations différentes, donne lieu aux méthodes Hybrides d'Ordre Elevé [100, 99, 97, 83] (Hybrid High Order, HHO en anglais). Elles permettent notamment d'obtenir un ordre de convergence optimal ($k+2$ en norme L^2 pour des inconnues de degré k), contrairement aux méthodes HDG originelles. Enfin, on peut citer les méthodes des Elements Virtuels [20, 21, 22, 7] (Virtual Element Method, VEM en anglais), dont l'analyse repose sur l'introduction d'espaces de fonctions non-polynomiales.

Les trois méthodes précédemment citées (HDG, VEM et HHO) sont fortement liées les unes aux autres, et partagent une caractéristique commune : des inconnues auxiliaires, placées sur le squelette du maillages (réunion des faces) sont utilisées pour réaliser une condensation statique,

qui permet d'éliminer les inconnues de maille. On parle parfois de méthodes squelettales. Les liens entre les méthodes HDG et HHO ont été étudiés et explicités dans [84], tandis que les similarités entre les méthodes VEM et HHO sont discutées dans [192].

L'utilisation de ces méthodes squelettales d'ordre élevé pour des problèmes de convection-diffusion est l'objet de travaux depuis plus d'une dizaine d'années. Concernant les équations d'advection-diffusion linéaires, de nombreuses méthodes HDG ont été développées [86, 78, 79, 139, 229]. On peut également citer la méthode HHO [98], qui est une extension à l'ordre élevé du schéma HMM [19]. Pour l'approximation de systèmes de dérive-diffusion, divers schémas DG et HDG ont aussi été proposés [200, 77]. Toutes ces méthodes, qui traitent la diffusion de manière linéaire, présentent un défaut de positivité.

A notre connaissance, la conception de schémas d'ordre élevé préservant la positivité est un travail encore assez peu exploré, qui est essentiellement restreint à des méthodes DG.
La première stratégie connue pour préserver la positivité repose sur l'utilisation de schémas DG dont la solution est positive en moyenne sur les mailles [198, 196]. Les inconnues peuvent prendre des valeurs négatives dans la maille, mais on s'assure que leurs moyennes sont positives en choisissant habilement des paramètres de stabilisation. Ce choix de stabilisation est délicat, et dépend de l'ordre de la méthode ainsi que de la dimension du problème. De plus, le fait que la solution puisse être ponctuellement négative limite pour le moment l'analyse des schémas. Cette méthode a aussi été appliquée à des systèmes de Poisson–Nernst–Planck [197].
Une autre approche est proposée dans [36], où un schéma DG pour une équation de Fisher–KPP (équation de la forme $\partial_t u - \triangle u = u(1-u)$, avec des conditions de Neumann homogènes) est introduit et analysé. Ce schéma suit les idées des schémas non-linéaires de [55, 56], et est basé sur la structure d'entropie du problème, utilisée avec un changement d'inconnue non-linéaire qui permet d'assurer la positivité de la solution. La positivité et la structure d'entropie discrète permettent alors de fournir une analyse étendue du schéma, incluant l'existence de solutions, le comportement en temps long et la convergence des solutions vers un problème semi-discrétisé en temps. L'analyse du schéma est réalisée pour un ordre et une dimension arbitraire. Cette analyse repose sur un terme de stabilisation bien choisi, qui fait intervenir les normes L^∞ des inconnues discrètes sur les faces du maillage. Les résultats numériques présentés pour un problème unidimensionnel indiquent bien une préservation de la positivité et un bon comportement en temps long, mais la question de l'ordre de convergence du schéma n'est pas discutée.

Méthodes étudiées dans cette thèse : HFV et HHO

Le positionnement initial de cette thèse est d'utiliser les stratégies introduites dans [55, 56, 52, 51] pour développer et analyser des méthodes pour le modèle anisotrope (3) permettant d'assurer la positivité des densités calculées. Parmi les méthodes volumes finis développées pour gérer l'anisotropie de problèmes linéaires, une classe de schémas, les méthodes HMM, se distingue des autres de par l'existence de généralisations d'ordre élevé. Puisque l'utilisation de méthodes d'ordre bas est un succès, la question de savoir s'il est possible de proposer et d'analyser des méthodes non-linéaires d'ordre élevé suivant les principes de [55, 56, 52, 51] se pose naturellement. Avant de pouvoir développer des méthodes d'ordre élevé, un travail préliminaire consiste à adapter les stratégies non-linéaires au cas des versions d'ordre bas des schémas. Les méthodes qui nous intéressent dans cette thèse sont dites hybrides, basées sur des inconnues dans les mailles et sur les faces.

Afin de concevoir des schémas préservant la structure, nous avons besoin d'utiliser des méthodes implicites (en temps) et généralement non-linéaires. Une des questions cruciales concernant leur analyse est l'existence de solutions. Comme pour l'étude des EDP il n'existe

pas de méthodologie générale pour ce genre de résultats. La technique utilisée dans cette thèse repose sur une méthode dite de continuité : l'idée est de comparer le schéma étudié avec un schéma plus simple, dont on sait qu'il admet des solutions (typiquement, un schéma linéaire). Cette comparaison se fait en reliant de manière continue les deux schémas. Il faut alors obtenir des estimations a priori sur les solutions, qui permettent d'en déduire l'existence de solution par des arguments de type degré topologique (voir par exemple [107] pour une introduction au degré ainsi que des applications à des résultats d'existence pour des problèmes continus et discrets). Ainsi, l'enjeu majeur de l'étude de schémas non-linéaires est l'obtention d'estimations a priori. Il est intéressant de noter que, contrairement au cas de schémas linéaires, la question de l'unicité des solutions ne découle en général pas des résultats d'existence. Cependant, cette problématique ne se limite pas à des enjeux théoriques : la qualité d'une méthode doit aussi être questionnée en pratique, à travers son implémentation dans un code de simulation. La question de l'existence est aussi délicate en pratique qu'en théorie : puisque calculer une solution revient à résoudre un système non-linéaire, comment peut-on la déterminer en pratique ? Dans notre cas, l'implémentation repose sur des méthodes de Newton, pour lesquelles nous n'avons pas de garantie théorique de convergence globale. Par conséquent, un enjeu réel est de vérifier que l'on arrive bien à calculer une solution. L'implémentation d'un schéma non-linéaire qui permet effectivement d'obtenir des solutions approchées est donc un résultat conséquent. Afin de s'assurer que le schéma (et son implémentation) est robuste, il est également nécessaire de réaliser des tests dans différentes situations (maillages, données physiques). Cette étape de test doit aussi permettre de confirmer que les propriétés théoriques du schéma sont présentes en pratique, notamment pour les schémas préservant la structure (positivité, comportement en temps long).
Notons finalement qu'il existe en général tout un panel de méthodes pour approcher un problème donné. Il est donc important de proposer une comparaison numérique, qui permet d'identifier les points forts des méthodes et d'en comprendre les cadres d'utilisation les plus appropriés. Cet aspect comparatif est assez présent dans cette thèse (Chapitres 1, 3 et 4). En particulier nous proposons lorsque c'est possible un travail de comparaison entre des méthodes linéaires et des méthodes non-linéaires. Il est important de noter que de tels travaux sont délicats, car les résultats peuvent dépendre de l'implémentation des schémas plus que des schémas eux même.

Les méthodes d'ordre bas étudiées dans cette thèse sont basées sur les Volumes Finis Hybrides [123] (Hybrid Finite Volume, HFV en anglais, qui sont un cas particulier du schéma SUSHI), qui coïncident avec les versions d'ordre bas des méthodes Hybrides d'Ordre Élevé [100] (Hybrid High Order, HHO en anglais). Notons que les méthodes HFV partagent des liens très forts avec les méthodes Différences Finies Mimétiques [40, 39] et les méthodes Volumes Finis Mixtes [110]. Dans [112], les auteurs unifient l'analyse et la description de ces trois méthodes, sous le nom de méthode Hybride Mimétiques Mixte (HMM).
Nous donnons ici une description succincte de la philosophie des schéma HFV. Pour une discrétisation $\mathcal{D}$ du domaine donné, on introduit des inconnues discrètes

$$\underline{u}_{\mathcal{D}} = ((u_K)_{K\in\mathcal{M}}, (u_\sigma)_{\sigma\in\mathcal{E}})$$

constituées d'inconnues de maille $u_K \in \mathbb{R}$ d'inconnues de face $u_\sigma \in \mathbb{R}$. L'approche de la méthode est locale : pour chaque maille K, on va définir un gradient discret sur K qui dépend des inconnues locales : l'inconnue de maille u_K et les inconnues de face de la maille $(u_\sigma)_{\sigma\in\mathcal{E}_K}$. Cette approche permet d'obtenir un gradient discret $\nabla_{\mathcal{D}}\underline{u}_{\mathcal{D}}$ sur Ω, qui est utilisé comme analogue discret du gradient continu. De manière équivalente, on peut définir des flux numériques $F_{K,\sigma}(\underline{u}_{\mathcal{D}})$ associés à ce gradient discret, et écrire la méthode comme une méthode volumes finis. Comme pour le gradient, les flux sont locaux, dans le sens où $F_{K,\sigma}(\underline{u}_{\mathcal{D}})$ dépend uniquement de

u_K et des inconnues de face locales $(u_\sigma)_{\sigma\in\mathcal{E}_K}$:

$$F_{K,\sigma}(\underline{u}_\mathcal{D}) = \sum_{\sigma'\in\mathcal{E}_K} A_K^{\sigma\sigma'}(u_K - u_{\sigma'}),$$

où les $A_K^{\sigma\sigma'}$ dépendent de la géométrie du maillage et du tenseur. L'une des forces majeures de cette approche locale consiste en la possibilité de réaliser une condensation statique lors de la résolution numérique des systèmes linéaires correspondant : l'inconnue de maille K peut être exprimée en fonction des inconnues de face associées. Il est alors possible d'éliminer localement les inconnues de maille, pour se ramener à un système linéaire sur les inconnues de face uniquement. Cette procédure offre un gain d'efficacité conséquent, en diminuant de fait le nombre d'inconnues de la méthode. Enfin, notons que les schémas HFV coïncident avec les schémas TPFA dans des cas isotropes sur des maillages fortement contraints, dits suradmissibles [123].

Comme pour les méthodes HFV, les schémas HHO sont basés sur des inconnues hybrides

$$\underline{u}_\mathcal{D} = ((u_K)_{K\in\mathcal{M}}, (u_\sigma)_{\sigma\in\mathcal{E}}),$$

mais ici les inconnues de maille et les inconnues de face sont polynomiales. Comme pour la méthode HFV, l'idée principale est de reconstruire un gradient sur chaque maille, en utilisant toutes les inconnues locales à disposition (inconnue de maille et inconnues de face). La définition de ce gradient suit une formule d'intégration par parties discrète, qui est testée contre des fonctions polynomiales. Ce gradient discret est consistant avec le gradient continu, mais il présente un défaut de stabilité. Pour obtenir un schéma stable, il faut également prendre en compte les sauts entre les inconnues de mailles et les inconnues de face, via un terme de stabilisation. Le choix de cette stabilisation est crucial, car il a une influence directe sur l'ordre de convergence. La stabilisation originelle de [100] permet, pour des inconnues de maille et de face de degré k d'obtenir un schéma d'ordre optimal $k+2$ en norme L^2. Avec ce choix, le schéma HHO pour $k=0$ coïncide avec la méthode HFV sous certaines hypothèses sur les données et le maillage [100]. Notons que la stabilisation initialement utilisée pour les méthodes HDG permet d'obtenir une convergence d'ordre $k+1$ en norme L^2 pour la même configuration.
La stabilisation utilisée au Chapitre 4 est différente de la stabilisation HHO originelle. Elle repose sur des inconnues de maille "enrichies" de degré $k+1$ et des inconnues de face de degré k. Avec de telles inconnues discrètes, la définition du gradient discret ne change pas, mais on peut utiliser une stabilisation dite de Lehrenfeld-Schöberl [190, 191], qui permet d'obtenir un schéma d'ordre $k+2$ en norme L^2. Dans le même temps, la condensation statique permet de garder un système global qui ne dépend que des inconnues de face, de degré k. De telles méthodes sont nommées méthodes HHO d'ordre mixte $(k,k+1)$ ou méthodes HDG+. L'intérêt de la stabilisation de Lehrenfeld-Schöberl est qu'elle est bien plus simple à exprimer, manipuler et implémenter que la stabilisation originelle de [100].

Overview of the works and results

In this section, we describe the main theoretical and numerical results presented in the manuscript. For each chapter, we also give some details about the codes implemented.

Long-time behaviour of Hybrid Finite Volume schemes for advection-diffusion (Chapter 1)

In the first chapter of this thesis, we focus on the numerical approximation of the linear advection-diffusion toy model

$$\partial_t u - \operatorname{div}(\Lambda(\nabla u + u\nabla\phi)) = f$$

with Hybrid Finite Volume (HFV) methods preserving entropy structures of the continuous problem. We focus on HFV methods because they are a first step towards HHO schemes.

We first introduce two linear methods (namely the one introduced in [19] and an exponential fitting scheme based on [43], relying on the change of unknown $\rho = u\,e^{\phi}$). Then, we develop a new nonlinear scheme based on the reformulation of the flux $u\Lambda\nabla(\log(u)+\phi)$ as in [56, 52] for Vertex Approximate Gradient (VAG) and Discrete Duality Finite Volume (DDFV) methods. The three HFV schemes have an entropy structure, related to quadratic entropies for the linear schemes and Boltzmann entropy for the nonlinear one. Thanks to these entropy dissipation estimates, we show several theoretical results.

The first theoretical result of this chapter is related to the existence of (positive) solutions to the nonlinear scheme, stated in Theorem 1. The associated proof differs from the existing ones of [56, 52, 51, 57], and is based on a discrete entropy method alongside with a Brouwer fixed-point argument. The method of proof is generic and can be easily adapted to any schemes with an entropy structure, see for instance [156, 1] and the other Chapters of the present manuscript. Other analysis results relate to the discrete long-time behaviour of the three schemes. These results are stated in Theorems 2, 3 and 4, and assert that the discrete solutions converge exponentially fast towards their associated steady-states as time tends to infinity. The proofs essentially rely on discrete functional inequalities. For both linear schemes, an analysis for data out of equilibrium is available.
As an intermediate step of the long-time behaviour analysis, we have established a Logarithmic-Sobolev inequality adapted to problems with Dirichlet boundary conditions in Appendices 1.A.2 and 1.A.3. Up to our knowledge, these results are new, both in the continuous and discrete frameworks.

We numerically validate our theoretical findings on a set of test-cases (see Figure 2) and, for completeness, we also compare the accuracy of the three schemes on stationary problems. These tests reveal that the two linear schemes can indeed compute non-positive solutions in practice. Moreover, all three schemes are accurate at order two in space for L^2 norm.

This work is an article published in Numerische Mathematik [70], in collaboration with Claire Chainais-Hillairet, Maxime Herda and Simon Lemaire.

About the codes

The HFV codes related to this first work are implemented in C++. The codes rely on an already existing mesh class for 2 dimensional meshes. The implementation is based on the "Finite Volume" flavour, using numerical fluxes. The main difficulty of these codes lies in the nonlinear schemes and the related Newton methods. Two codes were developed for the need of Chapter 1:

(i) **hfv_evol**: a code for linear anisotropic advection-diffusion equations with homogeneous Neumann boundary condition, zero source term and symmetric diffusion tensor. It includes

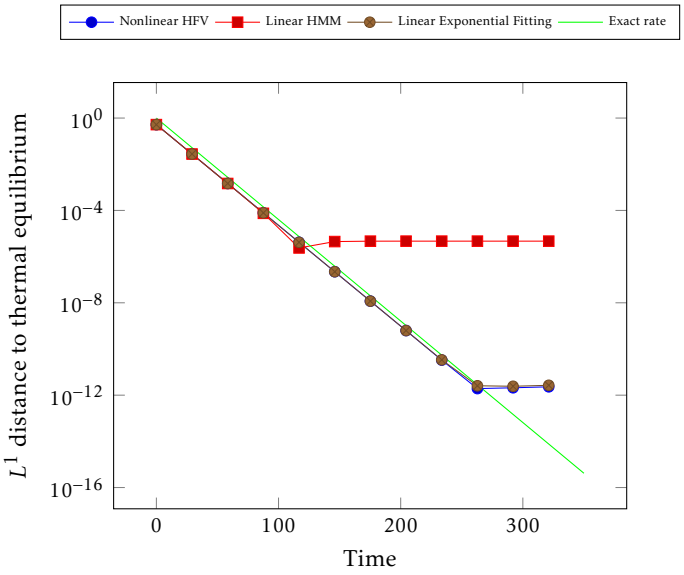


Figure 2 – Long-time behaviour of discrete solutions to HFV schemes on a Kershaw mesh.

the linear HMM scheme (1.26) of [19], the linear HFV exponential fitting scheme (1.35) and the nonlinear HFV scheme (1.45). The HMM scheme is implemented to handle general (not necessarily irrotational) advection fields. This code was used for the numerical results about positivity and long-time behaviour presented in Sections 1.5.2 and 1.5.3.

(ii) **hfv_sta**: a code for stationary linear anisotropic advection-diffusion equations with mixed Dirichlet-Neumann boundary condition (with a Dirichlet boundary of positive measure) with general boundary data, general source term and symmetric diffusion tensor. As for hfv_evol, this code include the three schemes presented in Chapter 1. For the nonlinear scheme, in order to ensure the convergence of the Newton method, a continuation procedure from the thermal equilibrium towards the real problem was implemented. This code was used for the accuracy study of Section 1.5.4.

Structure-preserving schemes for semiconductor models (Chapters 2 and 3)

In Chapters 2 and 3, we assert the relevance of the previously selected method (the nonlinear HFV scheme) by using it on a complex coupled system of PDEs (3) modelling semiconductor devices. The aim of this work is threefold:

— use the nonlinear HFV scheme on a nonlinear and anisotropic problem which cannot be approximated thanks to TPFA schemes;

— give us a better understanding of the theoretical features needed to approximate this kind of problems (with the aim of developing high-order schemes);

— check that such schemes actually work in practice for more complex problems than linear advection-diffusion equations.

In Chapter 2, we are interested in the numerical approximation of a drift-diffusion system which models semiconductor devices. The model under consideration is a generalisation of the one introduced in [143]. Its main feature is the fact that it takes into account an external magnetic field, which induces a rotation of the charge carriers (see Figure 3) and therefore

anisotropy on the convection-diffusion equations. From a mathematical point of view, such a system is modelled by the following system, where the unknowns N and P are the density of negative (electrons) and positive (holes) charges, and ϕ is the electrostatic potential induced by the spatial inhomogeneity of the electric charges:

$$\begin{cases} \partial_t N - \operatorname{div}(N\Lambda_N \nabla(h(N) - \phi)) = -R(N,P) \\ \partial_t P - \operatorname{div}(P\Lambda_P \nabla(h(P) + \phi)) = -R(N,P) \\ -\operatorname{div}(\Lambda_\phi \nabla\phi) = C + P - N. \end{cases}$$

In this system, Λ_N and Λ_P are (potentially) nonsymmetric anisotropic tensors related to the magnetic field, h is a nonlinearity in the diffusion (related to the so-called statistics used) and $R(N,P)$ is a reaction term. Because of the anisotropy induced by the magnetic field, it is not possible to design TPFA-admissible mesh for this problem. In [143] Gajewski and Gärtner propose a scheme based on a perturbation of the TPFA Scharfetter–Gummel scheme in order to handle this anisotropy for Boltzmann statistics ($h = \log$), but the scheme proved to compute negative densities on some situations. On the other hand, the positivity of the density is crucial both from a practical (since negative densities holds no physical meaning) and theoretical (since the analysis of the schemes relies on entropy structure, which do not have any meaning if the densities are negative) point of view. Thus, we use the nonlinear HFV method introduced in Chapter 1 in order to tackle both anisotropy and positivity issues, and introduce a structure-preserving scheme for this model.

With respect to the Chapter 1, the main novelty and difficulty of this work is the support of the coupling between unknowns. Another difference comes from the nonsymmetry of the tensors Λ_N and Λ_P, but it proves not to be a significant issue for the HFV method used. Last, since the diffusion considered for the equation on the charge carriers are nonlinear, one has to handle general nonlinearities and potentially two bounds on the densities. The analysis highlights the importance of the quasi-Fermi potentials $w_N = h(N) - \phi$ and $w_P = h(N) + \phi$, which are the key variables to handle the coupling. Thanks to this approach, we show that the scheme admits solution with relevant physical bounds on the densities (which are in particular positive), and preserve the long-time behaviour of the solution.

Our first result regarding the approximation of semiconductor models is Theorem 7, in which we state an existence result for the scheme. The computed densities are shown to be bounded in appropriate intervals (and in particular are positive) thanks to the use of a discrete entropy structure. We use the strategy employed in Chapter 1 together with a key correspondence discussed in Section 2.3.2 between discrete densities and discrete quasi-Fermi potentials through the nonlinear stationary scheme (2.42).
A second result of this work is Theorem 9, in which we establish the exponential decay of the discrete solutions towards the associated thermal equilibrium on a fixed mesh. The proof of this result relies on the use of a discrete Poincaré inequality on the discrete quasi-Fermi potentials.

From a numerical point of view, we give some evidence of our theoretical results, and investigate the relation between the cost of the scheme (number of Newton iterations) and the intensity of the magnetic field. Moreover, we assert the numerical robustness of the scheme with respect to the physical data, including the intensity of the magnetic field.

Following this work, we focus in Chapter 3 on the numerical comparison of two different structure-preserving methods for drift-diffusion systems, namely the HFV scheme introduced in Chapter 2 and a DDFV scheme based on the scheme for advection-diffusion [52, 51]. Up to our knowledge, such a comparative work between positivity preserving schemes on general meshes for coupled problems is new, but is fundamental in order to to understand the practical advantages and limitations of each method. For the sake of simplicity and conciseness, we

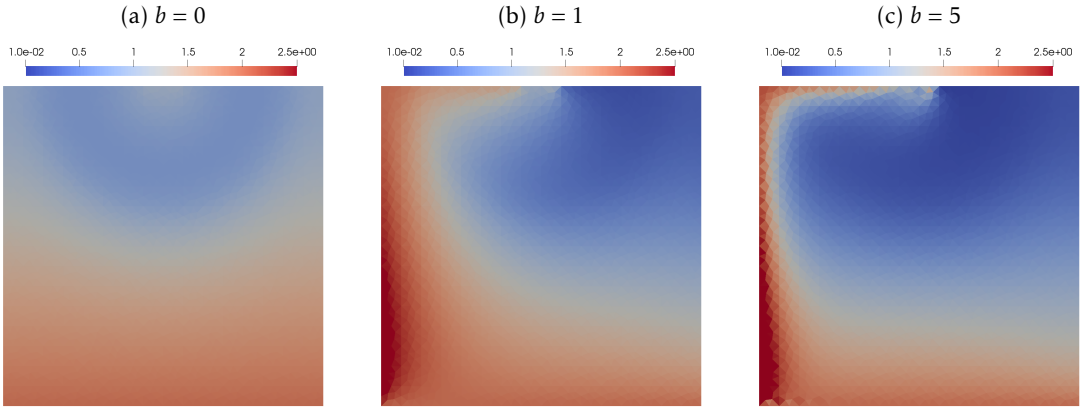


Figure 3 – Influence of the magnetic field intensity on the density of electrons of a non-equilibrium steady state.

consider a simplified model with Boltzmann statistics, without anisotropy neither reaction terms. We present both schemes in a similar way, in order to point out the numerous theoretical similarities between both frameworks. Therefore, the notations used in this Chapter (for the mesh and the HFV scheme) differ a little bit from these used in the rest of the manuscript. We then perform a numerical comparison about the long-time behaviour and the preservation of the positivity. Overall, both schemes present a similar behaviour, both from a quantitative and qualitative point of view.

These works are published in M2AN [213] (Chapter 2) and in the book of proceedings of the conference FVCA X [187] (Chapter 3, in collaboration with Stella Krell).

About the codes

The HFV code related to this part, **hfv_SC**, is implemented in C++ and follows the structure of **hfv_evol**. It is a code for transient anisotropic drift-diffusion system with mixed Dirichlet-Neumann boundary conditions. This code is developed to handle both Boltzmann and Blakemore statistics, realistic recombination/generation term (either Auger recombination, Shokley-Read-Hall term of a linear combination of both), anisotropic and non-symmetric tensors. With respect to the previous code, the main difficulty of the implementation lies in the coupling between the three different unknowns (positive and negative charges carrier density, electrostatic potential). Considering nonlinear diffusion (Blakemore statistics) for the densities also implies new challenges, including the fact that one has to handle two bounds on the densities (a lower one, which is 0 and upper one) in the Newton method. Another main difference lies in the stiff behaviour induced by the coupling (for small Debye lengths and strong magnetic fields), which requires a more refined choice of parameters (thresholds of truncation and stopping criteria). This code was used for the numerical results of Chapters 2 and 3.
Regarding the DDFV scheme studied in Chapter 3, it was implemented by Stella Krell.

High-order schemes for advection-diffusion (Chapter 4)

In the last chapter, we are interested in the discretisation of advection-diffusion equations with high-order methods in space. The main goal of this part is to develop and analyse a high-order (which means, at least of order 3 in L^2 norm in space) scheme which keeps the structure preserving features of the low-order schemes introduced in the previous chapters, and especially preserves the positivity. There is moreover a real need to investigate the numerical behaviour and the efficiency of such schemes for 2D problems. Thus, our objective is twofold:

— develop and analyse high-order structure-preserving schemes for diffusive problems;

— exhibit reasonable evidence that these high-order methods are appealing in terms of computational efficiency (compared to low-order structure preserving methods).

As in Chapter 1, this work focuses on a single linear advection-diffusion equation, but is of course intended to be used on more complex problems, including the anisotropic drift-diffusion system studied in Chapter 2.

We now provide a brief discussion about the locks and the key we had to identify in order to achieve the main goal of the thesis. The difficulties one needs to face in the design of a positivity-preserving high-order scheme inspired from the HFV scheme of Chapter 1 essentially lie in the following questions.

(i) **How can we ensure (both from a theoretical and a practical point of view) that the polynomial unknowns stay positive?**
Indeed, it is not easy a priori to infer the sign of a polynomial function (of arbitrary degree) from the degrees of freedom (values of the function at some points, or coefficients of the function in a given polynomial basis) which are the quantities available in practice in the code. Such a difficulty seems to be further reinforced by general meshes.

(ii) **How can we define high-order counterparts of the "reconstruction operators" used for the nonlinear HFV schemes?**
Indeed, the design and the analysis of the nonlinear HFV schemes rely on the definition of a local operator $r_K : \underline{V}_K \to \mathbb{R}$, which maps the local unknowns to a constant within the cell. A natural high-order generalisation of this operator (in order to get a consistent scheme at the expected order) should be an operator $r_K^k : \underline{V}_K^k \to \mathbb{P}^k(K)$, where $\underline{V}_K^k$ is the set of local (polynomial) unknowns, and $\mathbb{P}^k(K)$ is the space of polynomials functions on the cell K. However, using a simple mean between cell and face unknowns, as it was done for the low-order schemes, is not a relevant possibility anymore since the face unknowns are not defined on the cell.

(iii) **How can we generalise the method to obtain the L^∞ a priori estimates on the discrete solution?**
Indeed, the existence results for the low-order schemes rely on entropy methods: a control on the discrete entropy and its dissipation ensures that the discrete solution is bounded in L^∞ (with meshsize-dependant bounds). Up to now (and in similar works with different space discretisations [56, 52, 57]) the proofs of such results are based on a propagation of the information: one has a bound on a given cell unknown thanks to mass conservation, can therefore compare this unknown with the surrounding face unknowns and get estimates on these face unknowns. Then, proceeding in the same way, we compare the face unknowns with nearly cell unknowns, and get a bound for every unknown. For high-order schemes, a natural comparison can be performed on faces, between the trace of the cell unknown and the surrounding face unknowns. However, once we have a bound on a given face unknown, inferring a bound on the neighbouring cell unknown seems difficult, since we can only get information on the trace of this unknown.

Of course, these issues are somehow interconnected: the reconstruction operators should also be positive on the cells, and should be the key to get the algebraic structure allowing to get L^∞ bounds on the unknowns. The work presented in Chapter 4 proposes a first solution in order to answer the three previous points. We detail below the key ideas.

(1) **Potentials as polynomials and densities as exponentials of polynomials.** With regards to the analysis performed in Chapter 2, the key idea to solve the positivity issue (i) is to use the quasi-Fermi potential $w = \log(u) + \phi$ as the main unknown. Therefore, we want to approximate $\ell = \log(u)$ as a polynomial function (on each cell and face), and reconstruct a discrete density by taking the exponential of the discrete ℓ. The resulting discrete solution will thus be positive on the cells and faces, as the exponential of some real functions. Note that this process is easily generalisable to nonlinear diffusion since one just has to use the distribution function g of this statistics instead of the exponential.

(2) **Two different reconstructions for the consistent and stabilisation parts.** To address the reconstruction issue (ii), the solution chosen is to split the reconstruction operator into two different parts: a "consistent" reconstruction defined on the cells, which only takes into account the cell unknowns, and a "stabilising" reconstruction defined on the faces, which takes into account the face unknowns as well as the traces of the cell unknown. In order to simplify the design of the stabilising reconstruction, we choose to use a variant of the classical HHO scheme: instead of using cell and face unknowns of the same (maximal) degree k, we increase the maximal degree of the cell unknowns to $k+1$. Therefore, the scheme that we use is based on a mixed-order HHO$(k, k+1)$ scheme (see [83, Section 3.2.1], in HHO(l, p) l is the degree of face unknowns and p is the degree of cell unknwons), and on a stabilisation whose characteristics are derived from the so-called Lehrenfeld-Schöberl stabilisation [190, 191] in the context of HDG methods (which are sometimes called HDG+ schemes). This strategy allows for a scheme of order $k+2$ (in L^2 norm) while the systems to solve depend only on face unknowns, of degree k.

(3) **Over-dissipation thanks to over-stabilisation.** To tackle the issue (iii), we first have to understand how such results are achieved for low-order schemes. For all the nonlinear schemes presented and analysed in [56, 52, 57, 70], the proofs of L^∞ a priori estimates rely on the definition of the reconstruction operators, which ensure an over-dissipation: at the discrete level, one adds something to increase the dissipation (for the HFV scheme, we add the faces unknowns to get this over-dissipation, for the conforming $\mathbb{P}^1$ scheme of [57], the authors use the fact that the mean of a positive affine function on a simplex controls every local DOF of the unknown). In order to get analogous results in the high-order framework, we introduce a new term in the discretisation which induces such an over-dissipation. This term, which corresponds to a discretisation of $-\epsilon \operatorname{div}(\Lambda \nabla(\log(u) + \phi))$ can be seen as an over-stabilisation, where the ϵ depends on the mesh and tends to zero as the mesh is refined. Such a term provides sufficient control on the quasi-Fermi potential to obtain discrete L^∞ bounds from the entropy and its dissipation. Moreover, we scale this stabilisation parameter ϵ with respect to the size of the mesh so that it does not change the order of accuracy of the scheme. On the other hand, this stabilisation is designed to not disturb the mass preservation neither the thermodynamical consistency: the scheme has an entropy structure, and preserves the thermal equilibrium whether with or without this stabilisation. Up to now, we do not know if this over-stabilisation is really needed in the analysis, but numerical results indicates that the use of this term in the scheme does not have significant impact on its behaviour neither on its accuracy.

In order to demonstrate the interest of using high-order schemes, we introduce two arbitrary order skeletal schemes : a linear one (4.14), based on the exponential fitting strategy already

used in Chapter 1, and a nonlinear one (4.20), devised to preserved the positivity according to the previous discussion. Both schemes preserve the thermal equilibrium and enjoy discrete entropy structures. The exponential fitting schemes is essentially introduced for comparison purposes with respect to the nonlinear scheme.

In terms of analysis, our result is the existence of solutions (with positive densities) to the nonlinear scheme (4.20), stated in Theorem 10. This result is based on a discrete a priori estimate stated in Lemma 9, alongside with the entropy structure and the mass conservation property of the scheme. As previously discussed, the use of an over-dissipation term is the key ingredient to get these estimates.
We also state long-time behaviour results for both schemes in Propositions 14 and 16. Regarding the exponential fitting scheme, we obtained the exponential convergence of the solution towards the thermal equilibrium by the use an adapted discrete Poincaré inequality. For the nonlinear scheme, an analysis relying on discrete logarithmic Sobolev inequalities is not yet available, and the convergence of solution in time is based on a compactness argument alongside with the fact that the dissipation tends to vanish.

At the numerical level, we investigate the behaviour of both linear and nonlinear schemes, using the same test-cases as in Chapter 1, and compare some of the results with the low-order schemes of Chapter 1. In particular, we show that our theoretical results about long-time behaviour are satisfied for both scheme, and that the non-linear scheme preserves the positivity of the solution, while exhibiting an optimal order of accuracy (convergence of order $k+2$ for face unknowns of degree k). Moreover, this comparative work indicates two important outcomes. First, the use of the high-order version of the nonlinear scheme allows for significant gains in computational efficiency despite the fact that the time discretisation is merely of order one (see Figure 4). At the same time, when using high-order schemes, the use of nonlinear strategies seems to be needed to preserve the positivity of the discrete solution. Indeed, the solutions computed with the linear exponential fitting scheme demonstrates spurious oscillations (inducing some negative values) whose intensity increases with the order of approximation.

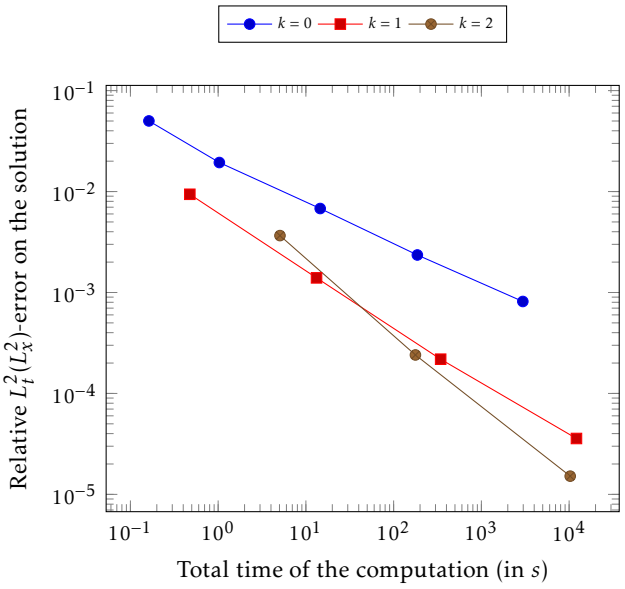


Figure 4 – Efficiency of non-linear HHO methods.

At the end, these promising results pave the way for the development and analysis of similar high-order and efficient schemes for complex dissipative systems.

The Chapter 4 is an extended version of the proceeding [212] for the conference FVCA X.

About the codes

The HHO codes related to Chapter 4 are implemented in C++. The codes rely on an existing code developed by Simon Lemaire for a linear Poisson equation, with isotropic diffusion and homogeneous Dirichlet boundary conditions on 2 dimensional meshes. This initial code relies on equal-order HHO method, and the consistent part used is the (continous) gradient of the potential reconstruction. Three codes were developed for the need of the last chapter:

(i) **hfv_fe**: a test code which was a preliminary work in order to understand how to write the finite volume scheme **hfv_evol** in a finite element flavour, with an implementation allowing a natural high-order extension. The nonlinear schemes implemented aim to approximate solution to transient linear advection-diffusion equations with homogeneous Neumann boundary conditions. The main attempt was to write separately the consistent term and the stabilisation term, in such a way that polynomial unknowns could be use instead of constant ones, while keeping an accuracy of order 2. The unknown used was either the density u or the potential $\ell = \log(u)$. The numerical results obtained with this code are not shown in this manuscript, but they indicated the importance of using an appropriate stabilisation term. These results also motivated the use of mixed-order HHO methods instead of equal-order one, in view of the complexity of both analysis and implementation of a relevant stabilisation term.

(ii) **hho_expfitt**: a code for linear anisotropic advection-diffusion equations with mixed Dirichlet-Neumann boundary conditions and general source term. The scheme implemented (4.14) is based on the exponential fitting strategy already used in Chapter 1, and is therefore a linear method. The main work regarding this code was to adapt the existing code to mixed-order method, implement general boundary conditions and discrete gradient operators. An other issue lies in the non-polynomial reconstruction of the discrete density (see Remark 19). There are two versions of this code: one for for stationary problems (with does not handle pure Neumann boundary conditions) and one for transient problems. This code was used for the numerical results of Chapter 4.

(iii) **hho_nl**: a code for transient anisotropic advection-diffusion equation with homogeneous Neumann boundary conditions. The scheme implemented is the nonlinear HHO scheme (4.20), main goal of this PhD. The main difficulties lies in the use of non-polynomial reconstructed densities and the nonlinearity of the scheme. Moreover, the use of high-order polynomials induces many numerical instabilities which hinders the convergence the Newton method. Various strategies had to be implemented in order to get a reliable solver, including truncation, filtering of oscillations and loop break for divergent Newton methods. The different tricks used are discussed in Section 4.4.1. The structure of this code is different from the previous ones, and essentially relies on two classes: one for representing the discrete unknowns (both potentials and densities), and an other one to handle the Newton method. This code was used for the numerical results of Chapter 4.

Semiconductors models with irregular convection fields (Appendices A-B-C-D)

The Appendices are dedicated to the numerical approximation of realistic modern semiconductor devices. This work began following Patricio Farrell's visit to Lille in October 2021, during which he presented to the Rapsodi team various projects and open questions concerning numerical approximations of real-life models of modern semiconductors. One of them was about semiconductor models which takes into account what is sometimes called random fluctuations in the alloys constituting the devices.

In a very simplified way, this consists in a drift-diffusion system with Boltzmann statistics (as the one studied in Chapter 3) where the convection fields are perturbed by some irregular quantities, called band-edge energies (BEEs). It reads

$$\begin{cases} \partial_t N - \operatorname{div}(\nabla N - N\nabla(\phi - E_c)) = 0 \\ \partial_t P - \operatorname{div}(\nabla P + P\nabla(\phi + E_v)) = 0 \\ -\lambda^2 \operatorname{div}(\nabla\phi) = C + P - N, \end{cases}$$

where E_c and E_v are the conduction and valence energies band, which are a characteristic of the material used in the semiconductor device. For classical models and devices, these BEEs are often considered to be constant functions, and therefore do not really appear in the equation since they only intervene trough their gradients. The main originality of the "random alloys fluctuations model" considered here lies in the fact that we do not consider constant BEEs anymore: these functions are now irregular ones. This irregularity comes from intern fluctuation inside the materials, and could even be described by a function which exhibits some jumps between each pair of atoms. This problem can then be reduced in a first time to the numerical approximation of solution to the advection-equation

$$\partial_t u - \operatorname{div}(\nabla u + u\nabla(\phi + E)) = 0,$$

where ϕ is a regular potential and E is an irregular band edge energy (which can be, to fix ideas, a piecewise constant function). The approach used by P. Farrell and his collaborators [221, 222] to get numerical simulations of such equations relied on a TPFA-Scharfetter–Gummel scheme, used as if the BEE E was regular, either with the discontinuous E or with a regularised version of E, obtained thanks to a Gaussian convolution. With this method, simulations provide numerical solutions in accordance with the physical intuition. On the other hand, this approach was lacking in a more robust analytical framework and some guarantees about the quality and the robustness of the approximation. Indeed, even at the continuous level, if E is a discontinuous function, its is not straightforward to define a natural notion of solution to such an equation (for example in $1D$, with a piecewise constant BEE, the gradient of E is a sum of Dirac masses and the corresponding advection field is extremely irregular). In the case one uses a regularisation on E, the question of knowing how to regularise in a good way to keep a similar behaviour to the non-regular model also needs to be answered.

Following the presentation of this problem, I suggested that one could use a strategy inspired from the exponential fitting procedure used in Chapter 1 in order to overcome the irregularity issue. Indeed, formally, using the change of variable $\rho = \frac{u}{\mathrm{e}^{-E}}$, we get the following equation on ρ

$$\mathrm{e}^{-E}\,\partial_t \rho - \operatorname{div}\left(\mathrm{e}^{-E}\,(\nabla\rho + \rho\nabla\phi)\right) = 0,$$

which does not take into account the undefined ∇E. Therefore, this formal computation suggests

to use the new "exponentially fitted" formulation to define a notion of solution to this irregular problem, and then define a solution of the irregular problem as $u = \rho e^{E}$. At the numerical level, this strategy implies to solve problems with a heterogeneous and discontinuous tensor e^{-E}. I then spent 10 weeks at the WIAS in Berlin during the summer of 2022 to work with Patricio Farrell on these issues. This stay led to some discussions and collaborations with physics researchers who worked on these models: Thomas Koprucki and Timo Streckenbach from the WIAS (Berlin, Germany), and Stefan Schulz, Michael O'Donovan and Robert Finn form the National Tyndall institute (Cork, Ireland). The content of the appendices presents the work resulting from this stay and these collaborations.

From a numerical point of view, this work is quite different from the previous ones, since we focus on TPFA finite volume methods. The main reason to this choice lies in the necessity to be able to use the methods developed in the existing codes currently used by the collaborators for the simulation of LEDs. However, we believe that the nonlinear hybrid methods studied during this PhD could be used in this framework. In fact, in the light of certain difficulties encountered to generate TPFA-admissible meshes for these problems in the situations studied in practice, the use of methods handling general meshes could be a solution in order to gain some computational efficiency, and therefore a first step towards the simulation of bigger devices. The same remark should apply to high-order schemes, in regards of the encouraging results of Chapter 4 about efficiency and computational costs. Thus, one can see this project as a preliminary work in order to apply state of the art numerical methods to current problems in physics and engineering.

In Appendix A, we are first interested in a the simple model of advection-diffusion with irregular advection field suggested by the physicists, and we propose a theoretical framework to analyse the equation, as well as different TPFA schemes to approximate the solution. Then, we discuss and summarise the works done with the physicists about the design of efficient LEDs.

For completeness, the full versions of the articles corresponding to these works are presented in the last appendices. In Appendix B, we compare the numerical results obtained thanks to our strategy with numerical results obtained by commercial packages and real experiments. In Appendix C, we investigate some issues related to the approximation of models with non-Boltzmann statistics. These issues are well-known in the mathematical community, but are rather poorly documented in physical communities. Especially, the main interest of this work is to show that unsuitable approximations lead to non-correct behaviours, which can even be seen on I-V curves. Last, in Appendix D we investigate numerically the way of designing efficient LEDs.

The Appendix B is an article submitted [219], in collaboration with Michael O'Donovan, Patricio Farrell, Timo Streckenbach, Thomas Koprucki and Stefan Schulz. Appendix C is an article pulished in Optical and Quantum Electronics [126], in collaboration with Patricio Farrell, Michael O'Donovan, Stefan Schulz and Thomas Koprucki. Last, the Appendix D is an article published in Applied Physics Letters [133], in collaboration with Robert Finn, Michael O'Donovan, Patricio Farrell, Timo Streckenbach, Thomas Koprucki and Stefan Schulz.

About the codes

The work of implementation related to this part differs from the previous ones, since it covers the modification of an existing code for realistic problems and the development of a demonstration code for toy problems.

In order to investigate the different possible discretisations of advection-diffusion with

irregular advection, a C++ code was implemented from scratch: **tpfa_qw**. It is developed to approximate the solutions of the model problem $-\mathrm{div}\,(u\nabla(h(u)+\phi+E)) = f$ in one space dimension. It can handle both Boltzmann and Blakemore statistics, generic electrostatic potential ϕ and piecewise constant band edge energy E. The main specificity of the code lies in the fine handling of the meshes and the discontinuities. Indeed, one of the strategies considered (the one before this work for the simulations) is a scheme where the quantities are not continuous on a given cell. With respect to the HFV codes this code marks the debut of the Object Oriented Programming in the implementation, and it introduces the general structure used to solve the nonlinear problem of Chapter 4. In particular, the code includes a class representing the discrete unknowns, and a class which handles the Newton method. This last class also allows one to test if the Jacobian is well computed.

On the other hand, an implementation work was performed on an existing code, **ddfermi** [104], which is designed to simulate drift-diffusion processes in semiconductor devices. This code handles 3D simulations, and various TPFA discretisations of the system, including the SEDAN flux [265] and the Bessemoulin-Chatard flux [26]. My main contribution was to add some implementations of numerical fluxes for non-Boltzmann statistics with varying band-edge energies. The simulations presented in Appendices B-C-D were obtained via **ddfermi**.

Chapter 1

Long-time behaviour of hybrid finite volume schemes for advection-diffusion equations: linear and nonlinear approaches

Outline of the current chapter

This chapter is a work in collaboration with Claire Chainais-Hillairet, Maxime Herda and Simon Lemaire published in Numerische Mathematik [70].

We are interested in the long-time behaviour of approximate solutions to anisotropic and heterogeneous linear advection-diffusion equations in the framework of hybrid finite volume (HFV) methods on general polygonal/polyhedral meshes. We consider two linear methods, as well as a new, nonlinear scheme, for which we prove the existence and the positivity of discrete solutions. We show that the discrete solutions to the three schemes converge exponentially fast in time towards the associated discrete steady-states. To illustrate our theoretical findings, we present some numerical simulations assessing long-time behaviour and positivity. We also compare the accuracy of the schemes on some numerical tests in the stationary case.

1.1 Introduction

We are interested in the numerical approximation of linear advection-diffusion equations on bounded domains. These equations constitute the main building block in the modelling of more complex problems stemming from physics (e.g., porous media flows [17], or corrosion models [16]), biology, or electronics (semi-conductor devices modelling [258]). Thus, designing reliable numerical schemes to approximate their solutions is a pre-requisite before discretising more complex models. Our aim here is the preservation of some key physical properties of these equations at the discrete level, on a large variety of meshes.

Let Ω be an open, bounded, connected polytopal subset of $\mathbb{R}^d$, $d \in \{2,3\}$, with boundary $\partial\Omega$ divided into two disjoint open subsets Γ^D and Γ^N, in such a way that $\partial\Omega = \overline{\Gamma^D} \cup \overline{\Gamma^N}$. We consider the following problem: Find $u : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ solution to

$$\begin{cases} \partial_t u - \operatorname{div}(\Lambda(\nabla u + u\nabla\phi)) = f & \text{in } \mathbb{R}_+ \times \Omega, \\ u = g^D & \text{on } \mathbb{R}_+ \times \Gamma^D, \\ \Lambda(\nabla u + u\nabla\phi)\cdot n = g^N & \text{on } \mathbb{R}_+ \times \Gamma^N, \\ u(0,\cdot) = u^{in} & \text{in } \Omega, \end{cases} \tag{1.1}$$

where n is the unit normal vector to $\partial\Omega$ pointing outward Ω, and the data satisfy:

— $\Lambda \in L^\infty(\Omega;\mathbb{R}^{d\times d})$ is a symmetric and uniformly elliptic diffusion tensor: there exist $\lambda_\flat, \lambda_\sharp$ with $0 < \lambda_\flat \le \lambda_\sharp < \infty$ such that, for a.e. x in Ω, $\xi\cdot\Lambda(x)\xi \ge \lambda_\flat|\xi|^2$ and $|\Lambda(x)\xi| \le \lambda_\sharp|\xi|$ for all $\xi \in \mathbb{R}^d$;

— $\phi \in C^1(\overline{\Omega})$ is a regular potential from which derives the advection field $V^\phi = -\Lambda\nabla\phi$, assumed to satisfy $V^\phi \in H(\mathrm{div};\Omega)$;

— $f \in L^2(\Omega)$ is a source term;

— $g^D \in H^{\frac{1}{2}}(\Gamma^D)$ is a Dirichlet datum, assumed to be the trace on Γ^D of $u^D \in H^1(\Omega)$ satisfying $\|u^D\|_{H^1(\Omega)} \leq C\|g^D\|_{H^{1/2}(\Gamma^D)}$ for a given $C > 0$;

— $g^N \in L^2(\Gamma^N)$ is a Neumann datum;

— $u^{in} \in L^2(\Omega)$ is an initial datum.

When $|\Gamma^D| = 0$, we assume that the compatibility condition $\int_\Omega f + \int_{\partial\Omega} g^N = 0$ holds true, and we denote by M the initial mass such that $M = \int_\Omega u^{in}$, which is known to be preserved along time: $\int_\Omega u(t) = M$ for almost every $t > 0$. For further use, we also let in that case $u^M = \frac{M}{|\Omega|} \in \mathbb{R}$, and we refer to this quantity as the mass lifting. Advection-diffusion models of the form (1.1) enjoy certain structural properties. First, when the data f, g^D, g^N, and u^{in} are positive, then the solution u is also positive. Second, the asymptotics $t \to \infty$, the so-called long-time behaviour of the solutions, is well understood (see [33, 62, 63, 254] for related models). Indeed, the solution u to (1.1) converges exponentially fast when $t \to \infty$ towards the steady-state u^∞, solution to the stationary problem

$$\begin{cases} -\mathrm{div}\,(\Lambda(\nabla u^\infty + u^\infty\nabla\phi)) = f & \text{in } \Omega, \\ u^\infty = g^D & \text{on } \Gamma^D, \\ \Lambda(\nabla u^\infty + u^\infty\nabla\phi)\cdot n = g^N & \text{on } \Gamma^N, \end{cases} \tag{1.2}$$

with additional constraint $\int_\Omega u^\infty = M$ when $|\Gamma^D| = 0$. The question of the long-time behaviour has been widely studied in the context of many-particle systems, for which the second law of thermodynamics ensures a relaxation of the transient phenomena towards an equilibrium. From a mathematical point of view, this evolution is strongly related to the dissipation of an entropy functional. Such a vision based on entropy dissipation has given birth to the so-called entropy method. As highlighted by Arnold et al. in [8], the successful use of the entropy method in kinetic theory paves the way to extended applications on various dissipative systems. We refer the reader to the book [174] of Jüngel for a presentation of some of these applications. In [33], Bodineau et al. proposed an entropy functional adapted to drift-diffusion equations with non-homogeneous Dirichlet boundary conditions. A direct adaptation of their method allows to conclude in the present case on the exponential convergence in time of the solution to Problem (1.1) towards the solution to Problem (1.2).

Under appropriate assumptions on the data (a sufficient condition, also valid for more general advection fields, is to assume that $\mathrm{div}\, V^\phi \geq 0$ a.e. in Ω and $V^\phi \cdot n \leq 0$ a.e. on Γ^N), the stationary Problem (1.2) is coercive and its well-posedness is straightforward. It turns out that, even if such assumptions on the data are not fulfilled, for an advection field of the form $V^\phi = -\Lambda\nabla\phi$, the problem is still coercive in the new unknown $\rho^\infty = u^\infty\, \mathrm{e}^\phi$, so that one can conclude on well-posedness by solving the problem in the new unknown. Concerning the evolution Problem (1.1), the same arguments show the existence and uniqueness of a global weak solution. For general advection fields (not necessarily deriving from a potential), we refer the reader to the results of Droniou [106] (for mixed Dirichlet-Neumann boundary conditions), and Droniou and Vázquez [114] (for pure Neumann boundary conditions) for detailed statements about well-posedness and regularity of the solutions.

When it comes to numerical approximation, the accuracy of the method is not the only important feature. In some applications (e.g., in subsurface modelling, where the mesh often results from seismic analysis), the mesh must be taken as a datum of the problem, and the

numerical method needs to be adapted so as to handle potentially fairly general meshes. In some other applications (e.g., power plant simulation), the preservation of the positivity of the solutions (or better, of the monotonicity properties of the equation) is an important quality criterion. In yet some other applications (e.g., nuclear waste repository management), finally, the reliability of the simulations in very large time proves to be crucial for sustainability purposes. The positivity and long-time behaviour of discrete solutions have been closely studied in the context of standard two-point flux approximation (TPFA) finite volume schemes, for isotropic diffusion (i.e., $\Lambda = \lambda I_d$ with $\lambda : \Omega \to \mathbb{R}_+^\star$) on orthogonal meshes. In [131], Filbet and Herda studied the long-time behaviour of a TPFA scheme for nonlinear boundary-driven Fokker–Planck equations, adapting to the discrete setting the arguments of [33]. In [69], Chainais-Hillairet and Herda proved on a variety of models that a whole family of TPFA schemes (the so-called B-schemes) preserves the exponential decay towards discrete steady-states. The results of [131] and [69] are valid for general advection fields, and a choice of data $|\Gamma^D| > 0$, $f = 0$, $g^D > 0$, $g^N = 0$, and $u^{in} \geq 0$. We also refer to [195, 73, 160, 46] for related schemes and similar issues. However, these TPFA schemes suffer from an intrinsic limitation: the mesh needs to be Λ-orthogonal, which, in practice, restricts their use to isotropic diffusion tensors and (standard) orthogonal meshes. In order to overcome this limitation, several linear finite volume methods using auxiliary unknowns have been designed (cf. [109] for a presentation of some of these schemes). As highlighted by Droniou in [109], these methods however suffer from a lack of monotonicity, and so do not preserve the positivity of discrete solutions. As a possible remedy, Cancès and Guichard introduced in [56] (see also the seminal paper [55]), for a class of models encompassing (1.1) for pure Neumann boundary conditions and a choice of data $f = 0$, $g^N = 0$, and $u^{in} \geq 0$ with $M > 0$, a nonlinear vertex approximate gradient (VAG) scheme, designed so as to preserve at the discrete level the positivity of the solutions and the entropy structure of the models, for arbitrary anisotropic diffusions and general meshes. Following the same ideas, Cancès et al. devised and analysed in [52] a (nonlinear) positivity-preserving discrete duality finite volume (DDFV) scheme, whose discrete entropy structure and long-time behaviour were fully studied in [51], based on the adaptation to the discrete setting of nonlinear functional inequalities. The DDFV scheme at hand is however limited to the two-dimensional case, and its adaptation to a three-dimensional framework seems difficult (cf. [109]). Let us also mention the work [241] (and the references therein), in which a general framework for the convergence analysis of positivity-preserving nonlinear cell-centred finite volume methods on general meshes is introduced. On another level, it is known that, given adequate assumptions hold on the data, the solutions to Problem (1.1) are regular in space (at least locally). This suggests that the use of high-order methods shall be an interesting track in order to increase the accuracy at fixed computational cost. Recently introduced by Di Pietro et al. in [100], hybrid high-order (HHO) methods can be seen as an arbitrary-order generalisation of hybrid finite volume (HFV) schemes, that were introduced by Eymard et al. in [123] as yet another way to overcome the limitations of TPFA schemes. HFV methods hinge on cell and face unknowns (whence the vocable hybrid), and as such benefit from a unified 2D/3D formulation. HFV methods have also been bridged to the larger family of hybrid mimetic mixed (HMM) methods in [112]. In view of the above elements, the study of HFV methods appears to be a natural first step in order to design structure-preserving high-order (HHO) schemes for Problem (1.1), that shall both increase the accuracy at fixed computational burden, and preserve the key properties (positivity and long-time behaviour) of the model at hand.

In this article, we study three different HFV schemes for Problem (1.1). The first one is the HFV variant of the HMM family of schemes introduced and analysed in the stationary setting by Beirão da Veiga et al. in [19, 108] (note that an arbitrary-order (HHO) generalisation of this scheme has been proposed in [98]). It is a linear scheme, based on a discretisation of the diffusive

and advective fluxes, that is well-posed under a coercivity condition. The second scheme is also a linear one. Its construction is based on exponential fitting, and takes inspiration from ideas in [43] (it also shares some features with the works [189, 205] and [131], which cover general advection fields). This scheme is unconditionally coercive. These two linear schemes are not expected to preserve positivity, which motivates the introduction of the third method. For pure Neumann boundary conditions, and a choice of data $f = 0$, $g^N = 0$, and $u^{in} \geq 0$ with $M > 0$ (see Appendix 1.B for the case of mixed Dirichlet-Neumann boundary conditions), we introduce a nonlinear HFV scheme, that is devised along the lines of the nonlinear VAG and DDFV schemes of [56] and [52, 51], so as to guarantee the positivity of discrete solutions. Our first result, stated in Theorem 1, is the existence of (positive) solutions to this nonlinear scheme. In a second time, we investigate the long-time behaviour of the three schemes at hand. We establish in Theorems 2, 3, and 4 the exponential decay in time of their discrete solutions towards the associated discrete steady-states. We numerically validate our theoretical findings on a set of test-cases and, for completeness, we also compare the accuracy of the three schemes on stationary problems.

The article is organised as follows. In Section 1.2, we introduce the HFV framework (mesh, discrete unknowns and discrete operators) on a steady variable diffusion problem. In Section 1.3, we introduce the three schemes for the transient advection-diffusion problem, and we discuss their well-posedness. In Section 1.4, we study the long-time behaviour of the three schemes, and prove exponential decay to equilibrium. In Section 1.5, we discuss the implementation of the schemes, and provide a numerical validation of our theoretical results, as well as a comparison of the stationary schemes in terms of accuracy. Appendices 1.A, 1.B, and 1.C finally collect some functional inequalities and the proofs of supplementary and auxiliary results.

1.2 Hybrid finite volume discretisation of a variable diffusion problem

The aim of this section is to recall the HFV framework on a steady variable diffusion problem, which corresponds to (1.2) without advection term ($V^\phi = 0$). For a detailed presentation of the method, we refer to [123].

1.2.1 Mesh and discrete unknowns

The definitions and notation we adopt for the discretisation are essentially the same as in [123]. A discretisation of the (open, bounded) polytopal set $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, is defined as a triplet $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$, where:

— $\mathcal{M}$ (the mesh) is a partition of Ω, i.e., a finite family of nonempty disjoint (open, connected) polytopal subsets K of Ω (the mesh cells) such that (i) for all $K \in \mathcal{M}$, $|K| > 0$, and (ii) $\overline{\Omega} = \bigcup_{K\in\mathcal{M}} \overline{K}$.

— $\mathcal{E}$ (the set of faces) is a partition of the mesh skeleton $\bigcup_{K\in\mathcal{M}} \partial K$, i.e., a finite family of nonempty disjoint (open, connected) subsets σ of $\overline{\Omega}$ (the mesh faces, or mesh edges if $d = 2$) such that (i) for all $\sigma \in \mathcal{E}$, $|\sigma| > 0$ and there exists $\mathcal{H}_\sigma$ affine hyperplane of $\mathbb{R}^d$ such that $\sigma \subset \mathcal{H}_\sigma$, and (ii) $\bigcup_{K\in\mathcal{M}} \partial K = \bigcup_{\sigma\in\mathcal{E}} \overline{\sigma}$. We assume that, for all $K \in \mathcal{M}$, there exists $\mathcal{E}_K \subset \mathcal{E}$ (the set of faces of K) such that $\partial K = \bigcup_{\sigma\in\mathcal{E}_K} \overline{\sigma}$. For $\sigma \in \mathcal{E}$, we let $\mathcal{M}_\sigma = \{K \in \mathcal{M} \mid \sigma \in \mathcal{E}_K\}$ be the set of cells whose σ is a face. Then, for all $\sigma \in \mathcal{E}$, either $\mathcal{M}_\sigma = \{K\}$ for a cell $K \in \mathcal{M}$, in which case σ is a boundary face ($\sigma \subset \partial\Omega$) and we note $\sigma \in \mathcal{E}_{ext}$, or $\mathcal{M}_\sigma = \{K, L\}$ for two cells $K, L \in \mathcal{M}$, in which case σ is an interface and we note $\sigma = K|L \in \mathcal{E}_{int}$.

— $\mathcal{P}$ (the set of cell centres) is a finite family $\{x_K\}_{K\in\mathcal{M}}$ of points of Ω such that, for all $K\in\mathcal{M}$, (i) $x_K\in K$, and (ii) K is star-shaped with respect to x_K. Moreover, we assume that the Euclidean (orthogonal) distance $d_{K,\sigma}$ between x_K and the affine hyperplane $\mathcal{H}_\sigma$ containing σ is positive (equivalently, the cell K is strictly star-shaped with respect to x_K).

For a given discretisation $\mathcal{D}$, we denote by $h_\mathcal{D}>0$ the size of the discretisation (the meshsize), defined by $h_\mathcal{D}=\sup_{K\in\mathcal{M}} h_K$ where, for all $K\in\mathcal{M}$, $h_K=\sup_{x,y\in\overline{K}}|x-y|$ is the diameter of the cell K. For all $\sigma\in\mathcal{E}$, we let $\overline{x}_\sigma\in\sigma$ be the barycentre of σ. Finally, for all $K\in\mathcal{M}$, and all $\sigma\in\mathcal{E}_K$, we let $n_{K,\sigma}\in\mathbb{R}^d$ be the unit normal vector to σ pointing outward K, and $P_{K,\sigma}$ be the (open) pyramid of base σ and apex x_K (notice that, when $d=2$, $P_{K,\sigma}$ is always a triangle). Since $|\sigma|$ and $d_{K,\sigma}$ are positive, we have $|P_{K,\sigma}|=\frac{|\sigma|d_{K,\sigma}}{d}>0$. We depict on Figure 1.1 an example of discretisation. Notice that the mesh cells are not assumed to be convex, neither x_K is assumed to be the barycentre of $K\in\mathcal{M}$. Notice that hanging nodes are seamlessly handled with our assumptions, so that meshes with non-conforming cells are allowed (see the orange cross in Figure 1.1; the cell K therein is treated as an hexagon). We consider the following measure of regularity for the discretisation (which is

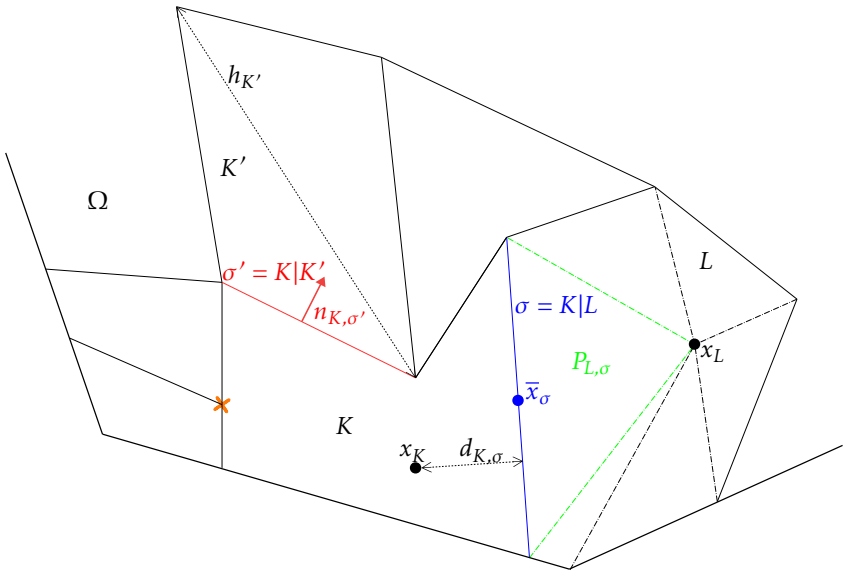


Figure 1.1 – Two-dimensional discretisation and corresponding notation.

slightly stronger than the ones advocated in [123, Eq. (4.1)] or in [111, Eq. (7.8)-(7.9)]):

$$\theta_\mathcal{D}=\max\left(\max_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}\frac{h_K}{d_{K,\sigma}},\max_{\sigma\in\mathcal{E},K\in\mathcal{M}_\sigma}\frac{h_K^{d-1}}{|\sigma|}\right). \tag{1.3}$$

Notice that $\theta_{\mathcal{D}} \geq 1$, and that for all $K \in \mathcal{M}$,

$$h_K^d \geq |K| = \sum_{\sigma \in \mathcal{E}_K} |P_{K,\sigma}| = \sum_{\sigma \in \mathcal{E}_K} \frac{|\sigma| d_{K,\sigma}}{d} \geq \sum_{\sigma \in \mathcal{E}_K} \frac{h_K^d}{d\theta_{\mathcal{D}}^2} = \frac{|\mathcal{E}_K|}{d\theta_{\mathcal{D}}^2} h_K^d.$$

Thus, the number of faces of any mesh cell is uniformly bounded:

$$\forall K \in \mathcal{M}, \qquad |\mathcal{E}_K| \leq d\theta_{\mathcal{D}}^2. \tag{1.4}$$

Also, it is an easy matter to verify that $\max_{\sigma = K|L \in \mathcal{E}_{int}} \max\left(\frac{d_{K,\sigma}}{d_{L,\sigma}}, \frac{d_{L,\sigma}}{d_{K,\sigma}}\right) \leq \theta_{\mathcal{D}}^{\frac{d}{d-1}}$. Given $\mathcal{F}$ a family of discretisations, we say that $\mathcal{F}$ is uniformly regular if there exists $\theta \geq 1$ such that for all $\mathcal{D} \in \mathcal{F}$, $\theta_{\mathcal{D}} \leq \theta$.

We now introduce the set of (hybrid, cell- and face-based) discrete unknowns:

$$\underline{V}_{\mathcal{D}} = \left\{ \underline{v}_{\mathcal{D}} = \left((v_K)_{K \in \mathcal{M}}, (v_\sigma)_{\sigma \in \mathcal{E}} \right) : v_K \in \mathbb{R}\ \forall K \in \mathcal{M}, v_\sigma \in \mathbb{R}\ \forall \sigma \in \mathcal{E} \right\}.$$

Given a mesh cell $K \in \mathcal{M}$, we let $\underline{V}_K = \mathbb{R} \times \mathbb{R}^{|\mathcal{E}_K|}$ be the restriction of $\underline{V}_{\mathcal{D}}$ to K, and $\underline{v}_K = \left(v_K, (v_\sigma)_{\sigma \in \mathcal{E}_K}\right) \in \underline{V}_K$ be the restriction of a generic element $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ to K. Also, for $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, we let $v_{\mathcal{M}} : \Omega \to \mathbb{R}$ and $v_{\mathcal{E}} : \bigcup_{K \in \mathcal{M}} \partial K \to \mathbb{R}$ be the piecewise constant functions such that

$$v_{\mathcal{M}|K} = v_K \text{ for all } K \in \mathcal{M}, \quad \text{and} \quad v_{\mathcal{E}|\sigma} = v_\sigma \text{ for all } \sigma \in \mathcal{E}.$$

In what follows, for any set $X \subset \overline{\Omega}$, we denote by $(\cdot,\cdot)_X$ the inner product in $L^2(X;\mathbb{R}^l)$, for $l \in \{1;d\}$. In particular, we have $(w_{\mathcal{M}}, v_{\mathcal{M}})_\Omega = \sum_{K \in \mathcal{M}} |K| w_K v_K$ and $(w_{\mathcal{E}}, v_{\mathcal{E}})_{\partial\Omega} = \sum_{\sigma \in \mathcal{E}_{ext}} |\sigma| w_\sigma v_\sigma$. For further use, we let $\underline{1}_{\mathcal{D}}$ denote the element of $\underline{V}_{\mathcal{D}}$ with all coordinates equal to 1. Also, given a function $f : \mathbb{R} \to \mathbb{R}$, and with a slight abuse in notation, we denote by $f(\underline{v}_{\mathcal{D}})$ the element of $\underline{V}_{\mathcal{D}}$ whose coordinates are the $(f(v_K))_{K \in \mathcal{M}}$ and the $(f(v_\sigma))_{\sigma \in \mathcal{E}}$. Finally, we let the product $\underline{w}_{\mathcal{D}} \times \underline{v}_{\mathcal{D}}$ denote the element of $\underline{V}_{\mathcal{D}}$ whose i-th coordinate is the product of the i-th coordinates of $\underline{w}_{\mathcal{D}}$ and $\underline{v}_{\mathcal{D}}$.

When considering mixed Dirichlet-Neumann boundary conditions, we assume that the discretisation $\mathcal{D}$ is compliant with the partition $\partial\Omega = \overline{\Gamma^D} \cup \overline{\Gamma^N}$ of the boundary of the domain, in the sense that the set $\mathcal{E}_{ext}$ can be split into two (necessarily disjoint) subsets $\mathcal{E}_{ext}^D = \left\{\sigma \in \mathcal{E}_{ext} \mid \sigma \subset \Gamma^D\right\}$ and $\mathcal{E}_{ext}^N = \left\{\sigma \in \mathcal{E}_{ext} \mid \sigma \subset \Gamma^N\right\}$ such that $\mathcal{E}_{ext} = \mathcal{E}_{ext}^D \cup \mathcal{E}_{ext}^N$. Notice that as soon as $|\Gamma^D| > 0$, $|\mathcal{E}_{ext}^D| \geq 1$. We define the following subspace of $\underline{V}_{\mathcal{D}}$, enforcing strongly a homogeneous Dirichlet boundary condition on Γ^D:

$$\underline{V}_{\mathcal{D},0}^D = \left\{ \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}} : v_\sigma = 0\ \forall \sigma \in \mathcal{E}_{ext}^D \right\}.$$

In view of the upcoming analysis, we define a discrete counterpart of the H^1 seminorm. Locally to any cell $K \in \mathcal{M}$, we let, for any $\underline{v}_K \in \underline{V}_K$, $|\underline{v}_K|_{1,K}^2 = \sum_{\sigma \in \mathcal{E}_K} \frac{|\sigma|}{d_{K,\sigma}} (v_K - v_\sigma)^2$. At the global level, for any $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, we let

$$|\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}} = \sqrt{\sum_{K \in \mathcal{M}} |\underline{v}_K|_{1,K}^2}.$$

Notice that $|\cdot|_{1,\mathcal{D}}$ does not define a norm on $\underline{V}_{\mathcal{D}}$, but if $|\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}} = 0$, then there is $c \in \mathbb{R}$ such that $\underline{v}_{\mathcal{D}} = c\,\underline{1}_{\mathcal{D}}$ ($\underline{v}_{\mathcal{D}}$ is constant). Thus, $|\cdot|_{1,\mathcal{D}}$ defines a norm on the space $\underline{V}_{\mathcal{D},0}^D$ as soon as $|\Gamma^D| > 0$, as

well as on the space of zero-mass vectors

$$\underline{V}_{\mathcal{D},0}^N = \left\{ \underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D} : \int_\Omega v_\mathcal{M} = 0 \right\}.$$

For further use, and to allow for a seamless treatment of pure Neumann boundary conditions, we introduce the notation $\underline{V}_{\mathcal{D},0}$, to denote either $\underline{V}_{\mathcal{D},0}^N$ whenever $|\Gamma^D| = 0$, or $\underline{V}_{\mathcal{D},0}^D$ otherwise.

1.2.2 Discrete problem

The HFV method hinges on the definition of a discrete gradient operator $\nabla_\mathcal{D}$, that maps any element $\underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}$ to a piecewise constant $\mathbb{R}^d$-valued function on the pyramidal submesh of $\mathcal{M}$ formed by all the $P_{K,\sigma}$'s, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$. More precisely, for all $K \in \mathcal{M}$, and all $\sigma \in \mathcal{E}_K$,

$$\nabla_\mathcal{D} \underline{v}_{\mathcal{D}|K} = \nabla_K \underline{v}_K \quad \text{with} \quad \nabla_K \underline{v}_{K|P_{K,\sigma}} = \nabla_{K,\sigma} \underline{v}_K = G_K \underline{v}_K + S_{K,\sigma} \underline{v}_K \in \mathbb{R}^d,$$

where $G_K \underline{v}_K$ is the consistent part of the gradient given by

$$G_K \underline{v}_K = \frac{1}{|K|} \sum_{\sigma' \in \mathcal{E}_K} |\sigma'|(v_{\sigma'} - v_K) n_{K,\sigma'} = \frac{1}{|K|} \sum_{\sigma' \in \mathcal{E}_K} |\sigma'| v_{\sigma'} n_{K,\sigma'},$$

and $S_{K,\sigma} \underline{v}_K$ is a stabilisation given, for some free parameter $\eta > 0$, by

$$S_{K,\sigma} \underline{v}_K = \frac{\eta}{d_{K,\sigma}} \Big(v_\sigma - v_K - G_K \underline{v}_K \cdot (\overline{x}_\sigma - x_K) \Big) n_{K,\sigma}. \tag{1.5}$$

Remark 1 (Choice of η)**.** *There are two specific values of the stabilisation parameter $\eta > 0$ for which one recovers known numerical schemes from the literature:*

(i) for $\eta = \sqrt{d}$, one recovers the original HFV scheme of [123], that coincides with the TPFA scheme on super-admissible meshes (see [123, Lemma 2.10]);

(ii) for $\eta = d$, one recovers the Discrete Geometric Approach (DGA) of [88], later bridged to the non-conforming finite element setting in [101].

The influence of the value of η on the numerical results has been investigated in [35] for anisotropic diffusion problems. It is shown that the above two values are appropriate choices (neither under- nor over-penalised).

Let us consider the stationary problem (1.2), without advection term ($V^\phi = 0$). Our aim is to write an HFV discretisation of this steady variable diffusion problem. Locally to any cell $K \in \mathcal{M}$, we introduce the discrete bilinear form $a_K^\Lambda : \underline{V}_K \times \underline{V}_K \to \mathbb{R}$ such that, for all $\underline{u}_K, \underline{v}_K \in \underline{V}_K$,

$$a_K^\Lambda(\underline{u}_K, \underline{v}_K) = \sum_{\sigma \in \mathcal{E}_K} |P_{K,\sigma}| \nabla_{K,\sigma} \underline{v}_K \cdot \Lambda_{K,\sigma} \nabla_{K,\sigma} \underline{u}_K = (\Lambda \nabla_K \underline{u}_K, \nabla_K \underline{v}_K)_K, \tag{1.6}$$

where we set $\Lambda_{K,\sigma} = \frac{1}{|P_{K,\sigma}|} \int_{P_{K,\sigma}} \Lambda$. At the global level, we let $a_\mathcal{D}^\Lambda : \underline{V}_\mathcal{D} \times \underline{V}_\mathcal{D} \to \mathbb{R}$ be the discrete bilinear form such that, for all $\underline{u}_\mathcal{D}, \underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}$,

$$a_\mathcal{D}^\Lambda(\underline{u}_\mathcal{D}, \underline{v}_\mathcal{D}) = \sum_{K \in \mathcal{M}} a_K^\Lambda(\underline{u}_K, \underline{v}_K) = (\Lambda \nabla_\mathcal{D} \underline{u}_\mathcal{D}, \nabla_\mathcal{D} \underline{v}_\mathcal{D})_\Omega.$$

The discrete HFV problem then reads: Find $\underline{u}_{\mathcal{D}}^z \in \underline{V}_{\mathcal{D},0}$ such that

$$a_{\mathcal{D}}^{\Lambda}(\underline{u}_{\mathcal{D}}^z, \underline{v}_{\mathcal{D}}) = (f, v_{\mathcal{M}})_{\Omega} + (g^N, v_{\mathcal{E}})_{\Gamma^N} - a_{\mathcal{D}}^{\Lambda}(\underline{u}_{\mathcal{D}}^l, \underline{v}_{\mathcal{D}}) \qquad \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, \tag{1.7}$$

where $\underline{u}_{\mathcal{D}}^l \in \underline{V}_{\mathcal{D}}$ is equal

(i) either, when $|\Gamma^D| > 0$, to the HFV interpolate $\underline{u}_{\mathcal{D}}^D$ of the known lifting u^D of the Dirichlet datum g^D (satisfying $|\underline{u}_{\mathcal{D}}^D|_{1,\mathcal{D}} \le C_{l,\Gamma^D} \|g^D\|_{H^{1/2}(\Gamma^D)}$, where $C_{l,\Gamma^D} > 0$ only depends on the discretisation $\mathcal{D}$ through $\theta_{\mathcal{D}}$),

(ii) or, when $|\Gamma^D| = 0$, to $\underline{u}_{\mathcal{D}}^M = u^M \underline{1}_{\mathcal{D}}$, where we recall that $u^M = \frac{M}{|\Omega|}$ is the mass lifting (remark that $a_{\mathcal{D}}^{\Lambda}(\underline{u}_{\mathcal{D}}^M, \underline{v}_{\mathcal{D}}) = 0$ for all $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$),

and the approximation of the solution to (1.2), denoted $\underline{u}_{\mathcal{D}}^{\infty} \in \underline{V}_{\mathcal{D}}$, is finally defined as

$$\underline{u}_{\mathcal{D}}^{\infty} = \underline{u}_{\mathcal{D}}^z + \underline{u}_{\mathcal{D}}^l. \tag{1.8}$$

Let us note that the superscript z stands for “zero", while l stands for “lifting".

Problem (1.7) defines a finite volume method, in the sense that it can be equivalently rewritten under a conservative form, with local mass balance, flux equilibration at interfaces, and boundary conditions. For all $K \in \mathcal{M}$, and all $\sigma \in \mathcal{E}_K$, the normal diffusive flux $-\int_{\sigma} \Lambda \nabla u^{\infty} \cdot n_{K,\sigma}$ is approximated by the following numerical flux:

$$F_{K,\sigma}^{\Lambda}(\underline{u}_K) = \sum_{\sigma' \in \mathcal{E}_K} A_K^{\sigma\sigma'}(u_K - u_{\sigma'}), \tag{1.9}$$

where the $A_K^{\sigma\sigma'}$ are defined by

$$A_K^{\sigma\sigma'} = \sum_{\sigma'' \in \mathcal{E}_K} |P_{K,\sigma''}| y_K^{\sigma''\sigma} \cdot \Lambda_{K,\sigma''} y_K^{\sigma''\sigma'}, \tag{1.10}$$

and the $y_K^{\sigma\sigma'} \in \mathbb{R}^d$ only depend on the geometry of the discretisation $\mathcal{D}$ (see, for example, [123, Eq. (2.22)] for an exact definition with $\eta = \sqrt{d}$). For all $K \in \mathcal{M}$, one can express the local discrete bilinear form a_K^{Λ} in terms of the local fluxes $\left(F_{K,\sigma}^{\Lambda}\right)_{\sigma \in \mathcal{E}_K}$: for all $\underline{u}_K, \underline{v}_K \in \underline{V}_K$,

$$a_K^{\Lambda}(\underline{u}_K, \underline{v}_K) = \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}^{\Lambda}(\underline{u}_K)(v_K - v_{\sigma}). \tag{1.11}$$

As for the VAG [56] and DDFV [52, 51] schemes, we can also express the local discrete bilinear form in a different way, which will be useful in the sequel:

$$a_K^{\Lambda}(\underline{u}_K, \underline{v}_K) = \delta_K \underline{v}_K \cdot \mathbb{\Lambda}_K \delta_K \underline{u}_K, \tag{1.12}$$

where, for all $\underline{v}_K \in \underline{V}_K$, $\delta_K \underline{v}_K \in \mathbb{R}^{|\mathcal{E}_K|}$ is defined by

$$\delta_K \underline{v}_K = (v_K - v_{\sigma})_{\sigma \in \mathcal{E}_K},$$

and $\mathbb{\Lambda}_K \in \mathbb{R}^{|\mathcal{E}_K| \times |\mathcal{E}_K|}$ is the symmetric (because Λ is) positive semi-definite matrix whose entries are the $A_K^{\sigma\sigma'}$, that can actually be proved to be nonsingular (cf. Lemma 4).

1.2.3 Well-posedness

As for the continuous case, the well-posedness of HFV methods for diffusion problems relies on a coercivity argument. Let $K \in \mathcal{M}$, and reason locally. By definition (1.6) of the local discrete bilinear form a_K^Λ, and from the bounds on the diffusion coefficient, we have $\lambda_\flat \|\nabla_K \underline{v}_K\|^2_{L^2(K;\mathbb{R}^d)} \le a_K^\Lambda(\underline{v}_K, \underline{v}_K) \le \lambda_\sharp \|\nabla_K \underline{v}_K\|^2_{L^2(K;\mathbb{R}^d)}$ for all $\underline{v}_K \in \underline{V}_K$. Furthermore, the following comparison result holds (cf. [111, Lemma 13.11, $p = 2$] and its proof): there exist $\alpha_\flat, \alpha_\sharp$ with $0 < \alpha_\flat \le \alpha_\sharp < \infty$, only depending on Ω, d, and $\theta_\mathcal{D}$ such that $\alpha_\flat |\underline{v}_K|^2_{1,K} \le \|\nabla_K \underline{v}_K\|^2_{L^2(K;\mathbb{R}^d)} \le \alpha_\sharp |\underline{v}_K|^2_{1,K}$ for all $\underline{v}_K \in \underline{V}_K$. Combining both estimates, we infer a local coercivity and boundedness result:

$$\forall \underline{v}_K \in \underline{V}_K, \qquad \lambda_\flat \alpha_\flat |\underline{v}_K|^2_{1,K} \le a_K^\Lambda(\underline{v}_K, \underline{v}_K) \le \lambda_\sharp \alpha_\sharp |\underline{v}_K|^2_{1,K}. \tag{1.13}$$

Summing over $K \in \mathcal{M}$, we get the following global estimates:

$$\forall \underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}, \qquad \lambda_\flat \alpha_\flat |\underline{v}_\mathcal{D}|^2_{1,\mathcal{D}} \le a_\mathcal{D}^\Lambda(\underline{v}_\mathcal{D}, \underline{v}_\mathcal{D}) \le \lambda_\sharp \alpha_\sharp |\underline{v}_\mathcal{D}|^2_{1,\mathcal{D}}. \tag{1.14}$$

The well-posedness of Problem (1.7)-(1.8) follows.

Proposition 1 (Well-posedness)**.** *There exists a unique solution $\underline{u}_\mathcal{D}^\infty \in \underline{V}_\mathcal{D}$ to Problem (1.7)-(1.8), which satisfies $|\underline{u}_\mathcal{D}^\infty|_{1,\mathcal{D}} \le C\left(\|f\|_{L^2(\Omega)} + \|g^N\|_{L^2(\Gamma^N)} + \|g^D\|_{H^{1/2}(\Gamma^D)}\right)$, for some $C > 0$ depending on the data, and on the discretisation $\mathcal{D}$ only through $\theta_\mathcal{D}$.*

Proof. The existence/uniqueness of $\underline{u}_\mathcal{D}^z \in \underline{V}_{\mathcal{D},0}$ solution to (1.7) (and in turn of $\underline{u}_\mathcal{D}^\infty = \underline{u}_\mathcal{D}^z + \underline{u}_\mathcal{D}^l$) is a direct consequence of the coercivity estimate (1.14), and of the fact that $|\cdot|_{1,\mathcal{D}}$ defines a norm on $\underline{V}_{\mathcal{D},0}$ (recall that $\underline{V}_{\mathcal{D},0}$ denotes either $\underline{V}_{\mathcal{D},0}^N$ when $|\Gamma^D| = 0$, or $\underline{V}_{\mathcal{D},0}^D$ otherwise). To prove the bound on $|\underline{u}_\mathcal{D}^\infty|_{1,\mathcal{D}}$, we use the triangle inequality:

$$|\underline{u}_\mathcal{D}^\infty|_{1,\mathcal{D}} \le |\underline{u}_\mathcal{D}^z|_{1,\mathcal{D}} + |\underline{u}_\mathcal{D}^l|_{1,\mathcal{D}}.$$

To estimate the first term, we test Problem (1.7) with $\underline{v}_\mathcal{D} = \underline{u}_\mathcal{D}^z \in \underline{V}_{\mathcal{D},0}$, we use (1.14), and we apply the Cauchy–Schwarz inequality. We get

$$\lambda_\flat \alpha_\flat |\underline{u}_\mathcal{D}^z|^2_{1,\mathcal{D}} \le \|f\|_{L^2(\Omega)} \|u_\mathcal{M}^z\|_{L^2(\Omega)} + \|g^N\|_{L^2(\Gamma^N)} \|u_\mathcal{E}^z\|_{L^2(\Gamma^N)} + \lambda_\sharp \alpha_\sharp |\underline{u}_\mathcal{D}^l|_{1,\mathcal{D}} |\underline{u}_\mathcal{D}^z|_{1,\mathcal{D}}.$$

Using a discrete Poincaré inequality, recalled in Proposition 7, and applied to $\underline{u}_\mathcal{D}^z \in \underline{V}_{\mathcal{D},0}$, as well as the discrete trace inequality of [111, Eq. (B.58), $p = 2$] combined with a discrete Poincaré inequality, we obtain

$$|\underline{u}_\mathcal{D}^z|_{1,\mathcal{D}} \le C_1 (\|f\|_{L^2(\Omega)} + \|g^N\|_{L^2(\Gamma^N)}) + C_2 |\underline{u}_\mathcal{D}^l|_{1,\mathcal{D}}.$$

It remains to estimate the norm of the lifting $|\underline{u}_\mathcal{D}^l|_{1,\mathcal{D}}$:

(i) if $|\Gamma^D| > 0$, one has $\underline{u}_\mathcal{D}^l = \underline{u}_\mathcal{D}^D$ (interpolate of the lifting u^D), therefore $|\underline{u}_\mathcal{D}^l|_{1,\mathcal{D}} \le C_{l,\Gamma^D} \|g^D\|_{H^{1/2}(\Gamma^D)}$;

(ii) if $|\Gamma^D| = 0$, since $\underline{u}_\mathcal{D}^l = \underline{u}_\mathcal{D}^M = u^M \underline{1}_\mathcal{D}$, the lifting is constant and $|\underline{u}_\mathcal{D}^l|_{1,\mathcal{D}} = 0$.

□

1.3 Definition of the schemes and well-posedness

In this section, we introduce and study the well-posedness of three HFV schemes, two linear ones and a nonlinear scheme, for the time-dependent advection-diffusion problem (1.1). For

the first two (linear) schemes, we introduce and study in the first place their steady versions on Problem (1.2). For the nonlinear scheme, by anticipation of the asymptotic analysis of Section 1.4.3, we restrict our study to the case where the (positive) solution to Problem (1.1) converges in long time towards the so-called thermal equilibrium (see (1.25)). However, as it will be verified numerically in Section 1.5.4 in the stationary setting, our scheme is applicable to more general data. We consider a fixed spatial discretisation $\mathcal{D}$ of Ω, which satisfies the conditions detailed in Section 1.2.1, and a fixed time step $\Delta t > 0$ for the time discretisation.

Remark 2 (Linear schemes and nonhomogeneous data)**.** *The linearity of Problem (1.1) implies that, (i) if $|\Gamma^D| > 0$, the shifted variable $u^z = u - u^D$ (recall that u^D is a known lifting of the Dirichlet datum g^D) satisfies an advection-diffusion equation with zero Dirichlet boundary condition on Γ^D, and (ii) otherwise, the shifted variable $u^z = u - u^M$ (recall that $u^M = \frac{M}{|\Omega|}$ is the mass lifting) satisfies a (compatible) pure Neumann advection-diffusion equation with zero-mass constraint. Thus, without loss of generality, we can restrict our study to the homogeneous case $g^D = 0$ or $M = 0$. For an example (in the steady, purely diffusive case) of how to handle at the discrete level nonhomogeneous data g^D or M, we refer the reader to Problem (1.7)-(1.8) and Proposition 1 above. Notice that such manipulations are possible for linear schemes only.*

1.3.1 Standard HFV scheme

We present here the HFV variant of the HMM family of schemes introduced in [19, 108].

Stationary problem

We consider Problem (1.2) with $g^D = 0$ when $|\Gamma^D| > 0$, or $M = \int_\Omega u^\infty = 0$ otherwise (cf. Remark 2). Locally to any cell $K \in \mathcal{M}$, we introduce the discrete bilinear form $a_K : \underline{V}_K \times \underline{V}_K \to \mathbb{R}$ such that, for all $\underline{u}_K, \underline{v}_K \in \underline{V}_K$,

$$a_K(\underline{u}_K, \underline{v}_K) = a_K^\Lambda(\underline{u}_K, \underline{v}_K) + a_K^\phi(\underline{u}_K, \underline{v}_K), \tag{1.15}$$

where the diffusive part a_K^Λ is defined by (1.6) (and rewrites as (1.11) in terms of the local diffusive fluxes $F_{K,\sigma}^\Lambda(\underline{u}_K)$ given by (1.9)), and the advective part a_K^ϕ is defined by

$$a_K^\phi(\underline{u}_K, \underline{v}_K) = \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}^\phi(\underline{u}_K)(v_K - v_\sigma), \tag{1.16}$$

with $F_{K,\sigma}^\phi(\underline{u}_K)$ an approximation of the normal advective flux $\int_\sigma u^\infty V^\phi \cdot n_{K,\sigma}$. In order to define the numerical advective fluxes, we need to introduce some data. We set $V_{K,\sigma}^\phi = \frac{1}{|\sigma|}\int_\sigma V^\phi \cdot n_{K,\sigma}$, and $\mu_\sigma = \min\Big(1, \min_{K\in\mathcal{M}_\sigma} \mathrm{Sp}(\Lambda_K)\Big) > 0$ where $\Lambda_K = \frac{1}{|K|}\int_K \Lambda$. We could as well use the finer local Péclet number introduced in [80], namely consider the value $\mu_\sigma = \min\Big(1, \min_{K\in\mathcal{M}_\sigma}(n_{K,\sigma}\cdot\Lambda_K n_{K,\sigma})\Big)$, but we choose here to stick to the formula advocated in [19, 108]. We also consider a Lipschitz continuous function $A : \mathbb{R} \to \mathbb{R}$, satisfying the following conditions:

$$\begin{aligned} A(0) &= 0, \\ \forall s \in \mathbb{R},\ A(-s) - A(s) &= s, \\ \forall s \in \mathbb{R},\ A(-s) + A(s) &\geq 0. \end{aligned} \tag{1.17}$$

Notice that $A = B - 1$, where B is the classical function used for the B-schemes introduced in [66]. In the B-schemes framework, advection and diffusion are simultaneously treated in the definition of the numerical flux. Here, as in [19], only the advective part is considered, whence the fact that $A = B - 1$. Standard choices of A functions include:

— the centred discretisation: $A : s \mapsto -\frac{s}{2}$;

— the upwind discretisation: $A : s \mapsto \max(-s, 0)$;

— the Scharfetter–Gummel discretisation: $A : s \mapsto \begin{cases} \frac{s}{e^s - 1} - 1 & \text{if } s \neq 0 \\ 0 & \text{if } s = 0 \end{cases}$.

We eventually define, for all $K \in \mathcal{M}$, and all $\sigma \in \mathcal{E}_K$, the numerical advective flux: for all $\underline{u}_K \in \underline{V}_K$,

$$F^{\phi}_{K,\sigma}(\underline{u}_K) = |\sigma| \frac{\mu_\sigma}{d_{K,\sigma}} \left(A\left(-\frac{d_{K,\sigma}}{\mu_\sigma} V^{\phi}_{K,\sigma} \right) u_K - A\left(\frac{d_{K,\sigma}}{\mu_\sigma} V^{\phi}_{K,\sigma} \right) u_\sigma \right). \tag{1.18}$$

Letting $a_{\mathcal{D}} : \underline{V}_{\mathcal{D}} \times \underline{V}_{\mathcal{D}} \to \mathbb{R}$ be the (global) discrete bilinear form such that, for all $\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$,

$$a_{\mathcal{D}}(\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = \sum_{K \in \mathcal{M}} a_K(\underline{u}_K, \underline{v}_K), \tag{1.19}$$

and recalling that $\underline{V}_{\mathcal{D},0}$ denotes either $\underline{V}^N_{\mathcal{D},0}$ whenever $|\Gamma^D| = 0$, or $\underline{V}^D_{\mathcal{D},0}$ otherwise, the discrete problem reads: Find $\underline{u}^{\infty}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}$ such that

$$a_{\mathcal{D}}(\underline{u}^{\infty}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = (f, v_{\mathcal{M}})_{\Omega} + (g^N, v_{\mathcal{E}})_{\Gamma^N} \qquad \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}. \tag{1.20}$$

Remark that, for pure Neumann boundary conditions with $M \neq 0$, as opposed to the purely diffusive case of Problem (1.7), $a_{\mathcal{D}}(\underline{u}^M_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \neq 0$ a priori for $\underline{v}_{\mathcal{D}} \in \underline{V}^N_{\mathcal{D},0}$. The well-posedness of (1.20) is discussed in the following proposition.

Proposition 2 (Well-posedness)**.** *Let A be a Lipschitz continuous function satisfying (1.17). If the advection field V^{ϕ} satisfies the two following conditions:*

$$(i)\ \textit{almost everywhere on } \Gamma^N,\ V^{\phi} \cdot n \leq 0, \tag{1.21a}$$

$$(ii)\ \textit{there exists } \beta < \frac{2\lambda_{\flat}\alpha_{\flat}}{C_P^2} \textit{such that, almost everywhere in } \Omega,\ \operatorname{div} V^{\phi} \geq -\beta, \tag{1.21b}$$

where $\lambda_{\flat}\alpha_{\flat}$ is the coercivity constant of (1.14), and C_P is either equal to C_{PW} if $|\Gamma^D| = 0$ or to C_{P,Γ^D} otherwise (where C_{PW}, C_{P,Γ^D} are the Poincaré constants of Proposition 7), then there exists $\kappa > 0$, only depending on Λ, β, Ω, d, Γ^D, and $\theta_{\mathcal{D}}$ such that

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, \qquad a_{\mathcal{D}}(\underline{v}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \geq \kappa |\underline{v}_{\mathcal{D}}|^2_{1,\mathcal{D}}. \tag{1.22}$$

Consequently, there exists a unique solution $\underline{u}^{\infty}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}$ to Problem (1.20). Moreover, one has

$$|\underline{u}^{\infty}_{\mathcal{D}}|_{1,\mathcal{D}} \leq C\Big(\|f\|_{L^2(\Omega)} + \|g^N\|_{L^2(\Gamma^N)} \Big),$$

for some $C > 0$ depending on the data, and on the discretisation $\mathcal{D}$ only through $\theta_{\mathcal{D}}$.

Proof. Let $K \in \mathcal{M}$, and $\sigma \in \mathcal{E}_K$. Let $s_{K,\sigma} = \frac{d_{K,\sigma}}{\mu_\sigma} V^{\phi}_{K,\sigma}$ and $\zeta_{K,\sigma} = A(-s_{K,\sigma}) + A(s_{K,\sigma})$. According to (1.17), $\zeta_{K,\sigma} \geq 0$, and we have $A(-s_{K,\sigma}) = \frac{s_{K,\sigma} + \zeta_{K,\sigma}}{2}$ and $-A(s_{K,\sigma}) = \frac{s_{K,\sigma} - \zeta_{K,\sigma}}{2}$. Consequently, for

all $\underline{v}_K \in \underline{V}_K$,

$$\Big(A(-s_{K,\sigma})v_K - A(s_{K,\sigma})v_\sigma\Big)(v_K - v_\sigma) = \frac{1}{2}s_{K,\sigma}(v_K^2 - v_\sigma^2) + \frac{1}{2}\zeta_{K,\sigma}(v_K - v_\sigma)^2 \geq \frac{1}{2}s_{K,\sigma}(v_K^2 - v_\sigma^2).$$

Recalling (1.16) and (1.18), we infer that, for all $\underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}$,

$$a_\mathcal{D}^\phi(\underline{v}_\mathcal{D}, \underline{v}_\mathcal{D}) = \sum_{K\in\mathcal{M}} a_K^\phi(\underline{v}_K, \underline{v}_K) \geq \frac{1}{2}\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} |\sigma| V_{K,\sigma}^\phi (v_K^2 - v_\sigma^2).$$

Since, for all $K \in \mathcal{M}$, $\sum_{\sigma\in\mathcal{E}_K} |\sigma| V_{K,\sigma}^\phi = \int_K \operatorname{div} V^\phi$ and, for all $\sigma \in \mathcal{E}_{int}$, $\sum_{K\in\mathcal{M}_\sigma} |\sigma| V_{K,\sigma}^\phi = 0$, we have, for all $\underline{v}_\mathcal{D} \in \underline{V}_{\mathcal{D},0}$,

$$a_\mathcal{D}^\phi(\underline{v}_\mathcal{D}, \underline{v}_\mathcal{D}) \geq \frac{1}{2}(\operatorname{div} V^\phi, v_\mathcal{M}^2)_\Omega - \frac{1}{2}(V^\phi \cdot n, v_\mathcal{E}^2)_{\Gamma^N}. \tag{1.23}$$

Combining (1.15) with (1.19), (1.21) with (1.23), and the coercivity result (1.14), we deduce that, for all $\underline{v}_\mathcal{D} \in \underline{V}_{\mathcal{D},0}$,

$$a_\mathcal{D}(\underline{v}_\mathcal{D}, \underline{v}_\mathcal{D}) \geq \lambda_\flat \alpha_\flat |\underline{v}_\mathcal{D}|_{1,\mathcal{D}}^2 - \frac{\beta}{2}\|v_\mathcal{M}\|_{L^2(\Omega)}^2. \tag{1.24}$$

Using a Poincaré inequality from Proposition 7, one has, for all $\underline{v}_\mathcal{D} \in \underline{V}_{\mathcal{D},0}$,

$$a_\mathcal{D}(\underline{v}_\mathcal{D}, \underline{v}_\mathcal{D}) \geq \left(\lambda_\flat \alpha_\flat - \frac{\beta C_P^2}{2}\right)|\underline{v}_\mathcal{D}|_{1,\mathcal{D}}^2,$$

therefore the estimate (1.22) holds for $\kappa = \lambda_\flat\alpha_\flat - \frac{\beta C_P^2}{2}$, which is positive according to (1.21b). The existence/uniqueness of $\underline{u}_\mathcal{D}^\infty \in \underline{V}_{\mathcal{D},0}$ solution to (1.20) is a direct consequence of the coercivity estimate (1.22), and of the fact that $|\cdot|_{1,\mathcal{D}}$ defines a norm on $\underline{V}_{\mathcal{D},0}$. The continuous dependency of $\underline{u}_\mathcal{D}^\infty$ with respect to the data can then be proved as in the proof of Proposition 1. □

Remark 3 (Assumptions on the advection field). *The well-posedness result of Proposition 2 does not use the fact that V^ϕ is related to the gradient of a potential, and thus extends to general advection fields. Even better, under a smallness assumption on the meshsize $h_\mathcal{D}$, it is actually possible to prove well-posedness for Problem (1.20) without assumptions (1.21) on the advection field. The starting point to prove so is a discrete Gårding inequality like (1.24), which can be easily obtained in full generality from (1.23) (in the case $|\Gamma^N| > 0$, it is obtained from the multiplicative discrete trace inequality of [111, Eq. (B.57), $p = 2$] and holds for $h_\mathcal{D}$ sufficiently small). The proof then proceeds by contradiction, assuming that a discrete inf-sup condition does not hold in the limit $h_\mathcal{D} \to 0$, and using a compactness argument (cf. [111, Lemmas B.27-B.33, $p = 2$]), together with the unconditional well-posedness of the continuous problem (1.2) (cf. [106]).*

Remark 4 (Choice of A). *The choice of the function A is of great importance. In particular, the Scharfetter–Gummel approximation is rather classical in various contexts. First introduced in [239] in the framework of TPFA schemes, this approximation of the flux ensures the preservation of the so-called thermal (or Gibbs) equilibrium at the discrete level, which has the form:*

$$u_{th}^\infty = \bar\rho\, \mathrm{e}^{-\phi} \tag{1.25}$$

where $\bar\rho \in \mathbb{R}_+$ is prescribed by the data. For instance, for pure Neumann boundary conditions, $f = 0$, and $g^N = 0$, we have $\bar\rho = M/\int_\Omega \mathrm{e}^{-\phi}$. The discrete solution obtained with the TPFA scheme is then the interpolate of u_{th}^∞. This property is no more true for the hybrid scheme and we observe numerically

that, for $\bar{\rho} \neq 0$, the discrete solution $\underline{u}_{\mathcal{D}}^{\infty} \in \underline{V}_{\mathcal{D}}$ is in general not the HFV interpolate of u_{th}^{∞}. However, as explained in [108, pp. 553-554], provided the parameters $(\mu_\sigma)_{\sigma\in\mathcal{E}}$ are well-chosen, the Scharfetter–Gummel flux ensures an automatic upwinding to the scheme in the advection-dominated regime (whereas it degenerates towards the centred scheme in the diffusion-dominated regime).

Evolution problem

We consider Problem (1.1) with $g^D = 0$ when $|\Gamma^D| > 0$, or $M = \int_\Omega u^{in} = 0$ otherwise (see Remark 2). We use a backward Euler discretisation in time, and the HFV discretisation introduced in Section 1.3.1 in space. The discrete problem reads: Find $\left(\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D},0}\right)_{n\geq 1}$ such that

$$\begin{cases} \dfrac{1}{\Delta t}(u_{\mathcal{M}}^n - u_{\mathcal{M}}^{n-1}, v_{\mathcal{M}})_\Omega + a_{\mathcal{D}}(\underline{u}_{\mathcal{D}}^n, \underline{v}_{\mathcal{D}}) = (f, v_{\mathcal{M}})_\Omega + (g^N, v_{\mathcal{E}})_{\Gamma^N} & \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, \quad (1.26a) \\ u_K^0 = \dfrac{1}{|K|}\displaystyle\int_K u^{in} & \forall K \in \mathcal{M}, \quad (1.26b) \end{cases}$$

where $a_{\mathcal{D}}$ is defined by (1.15) and (1.19). Since $a_{\mathcal{D}}$ is coercive, the bilinear form in (1.26a) is also coercive, so the scheme (1.26) is well-posed under the assumptions (1.21) on the advection field, as a straightforward consequence of Proposition 2.

Remark 5 (Pure Neumann case)**.** *When considering pure Neumann boundary conditions, and contrary to the stationary case, one can actually seek at each time step for a solution to Problem (1.26) in $\underline{V}_{\mathcal{D}}$, i.e., it is not necessary to seek for a solution in the constrained space $\underline{V}_{\mathcal{D},0}^N$. Indeed, testing (1.26a) by $\underline{1}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, and using that $\int_\Omega f + \int_{\partial\Omega} g^N = 0$ and $a_{\mathcal{D}}(\underline{u}_{\mathcal{D}}^n, \underline{1}_{\mathcal{D}}) = 0$, one can automatically infer that $\int_\Omega u_{\mathcal{M}}^n = \int_\Omega u_{\mathcal{M}}^{n-1}$ for all $n \geq 1$, that is, $\int_\Omega u_{\mathcal{M}}^n = \int_\Omega u_{\mathcal{M}}^0 = \int_\Omega u^{in} = M = 0$ for all $n \geq 1$, i.e., $\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D},0}^N$ for all $n \in \mathbb{N}^\star$.*

1.3.2 Exponential fitting HFV scheme

Following ideas in [43] (cf. also [189, 205] and [131] for general advection fields) in the context of finite element methods, we aim to design an unconditionally (i.e., without the need for assumptions (1.21) on the advection field V^ϕ) coercive scheme for the advection-diffusion problem in the HFV framework.

Stationary problem

We consider Problem (1.2). The strategy advocated in [43] is based on the following observation: at the continuous level, if u^∞ is a solution to (1.2), letting

$$\omega = \mathrm{e}^{-\phi}, \tag{1.27}$$

we can introduce the Slotboom change of variable $\rho^\infty = \frac{u^\infty}{\omega}$ (see [207, 208]). Then, noticing that $\nabla u^\infty + u^\infty \nabla\phi = \omega\nabla\rho^\infty - \rho^\infty\omega\nabla\phi + \rho^\infty\omega\nabla\phi = \omega\nabla\rho^\infty$, the Slotboom variable ρ^∞ equivalently solves the following pure diffusion problem:

$$\begin{cases} -\mathrm{div}\,(\omega\Lambda\nabla\rho^\infty) = f & \text{in } \Omega, \\ \rho^\infty = \omega^{-1} g^D & \text{on } \Gamma^D, \\ \omega\Lambda\nabla\rho^\infty \cdot n = g^N & \text{on } \Gamma^N, \end{cases} \tag{1.28}$$

with additional constraint $\int_\Omega \omega\rho^\infty = M$ when $|\Gamma^D| = 0$. Following Remark 2 (with ρ instead of u, ρ^D lifting of $\omega^{-1}g^D$ instead of u^D lifting of g^D, and $\rho^M = \frac{M}{\int_\Omega \omega} \in \mathbb{R}$ instead of u^M), we consider Problem (1.28) with $\rho^\infty = 0$ on Γ^D when $|\Gamma^D| > 0$, or $\int_\Omega \omega\rho^\infty = 0$ otherwise (which is equivalent to consider Problem (1.2) with $g^D = 0$ or $M = \int_\Omega u^\infty = 0$). Since ϕ is continuous on $\overline{\Omega}$, there exist $\omega_\flat, \omega_\sharp$ with $0 < \omega_\flat \le \omega_\sharp < \infty$, only depending on ϕ and Ω, such that $\omega_\flat \le \omega(x) \le \omega_\sharp$ for all $x \in \overline{\Omega}$. We then denote by $L^2_\omega(\Omega)$ the ω-weighted L^2 space on Ω.

At the discrete level, instead of discretising (1.2), we approximate the solution to (1.28). For any $K \in \mathcal{M}$, we let $a_K^\omega : \underline{V}_K \times \underline{V}_K \to \mathbb{R}$ be the discrete bilinear form such that, for all $\underline{\rho}_K, \underline{v}_K \in \underline{V}_K$,

$$a_K^\omega(\underline{\rho}_K, \underline{v}_K) = (\omega\Lambda\nabla_K\underline{\rho}_K, \nabla_K\underline{v}_K)_K, \tag{1.29}$$

and, classically, we let $a_\mathcal{D}^\omega : \underline{V}_\mathcal{D} \times \underline{V}_\mathcal{D} \to \mathbb{R}$ be the corresponding global discrete bilinear form obtained by sum of the local contributions. To account for the change of variable, we let $\underline{V}_{\mathcal{D},0}^\omega$ be the space $\underline{V}_{\mathcal{D},0}^D$ when $|\Gamma^D| > 0$, and the space $\left\{\underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D} : \int_\Omega \omega v_\mathcal{M} = 0\right\}$ otherwise. The discrete problem reads: Find $\underline{\rho}_\mathcal{D}^\infty \in \underline{V}_{\mathcal{D},0}^\omega$ such that

$$a_\mathcal{D}^\omega(\underline{\rho}_\mathcal{D}^\infty, \underline{v}_\mathcal{D}) = (f, v_\mathcal{M})_\Omega + (g^N, v_\mathcal{E})_{\Gamma^N} \qquad \forall \underline{v}_\mathcal{D} \in \underline{V}_{\mathcal{D},0}^\omega. \tag{1.30}$$

Remark that, for pure Neumann boundary conditions with $M \neq 0$, as for Problem (1.7), letting $\underline{\rho}_\mathcal{D}^M = \rho^M \underline{1}_\mathcal{D}$, $a_\mathcal{D}^\omega(\underline{\rho}_\mathcal{D}^M, \underline{v}_\mathcal{D}) = 0$ for all $\underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}$. Letting $\underline{\omega}_\mathcal{D} \in \underline{V}_\mathcal{D}$ be the HFV interpolate of ω, i.e.,

$$\omega_K = \frac{1}{|K|}\int_K \omega \quad \forall K \in \mathcal{M}, \qquad \omega_\sigma = \frac{1}{|\sigma|}\int_\sigma \omega \quad \forall \sigma \in \mathcal{E}, \tag{1.31}$$

the approximation of the solution to Problem (1.2) is finally defined as the product $\underline{u}_\mathcal{D}^\infty = \underline{\omega}_\mathcal{D} \times \underline{\rho}_\mathcal{D}^\infty$, that is

$$u_K^\infty = \omega_K\rho_K^\infty \quad \forall K \in \mathcal{M}, \qquad u_\sigma^\infty = \omega_\sigma\rho_\sigma^\infty \quad \forall \sigma \in \mathcal{E}. \tag{1.32}$$

Remark that $\underline{u}_\mathcal{D}^\infty \in \underline{V}_{\mathcal{D},0}$. Reasoning as in Section 1.2.3, one can easily prove that, for all $\underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}$,

$$a_\mathcal{D}^\omega(\underline{v}_\mathcal{D}, \underline{v}_\mathcal{D}) \ge \omega_\flat\lambda_\flat\alpha_\flat\,|\underline{v}_\mathcal{D}|_{1,\mathcal{D}}^2. \tag{1.33}$$

This estimate is instrumental to infer well-posedness for Problem (1.30)-(1.32).

Proposition 3 (Well-posedness). *There exists a unique solution $\underline{u}_\mathcal{D}^\infty \in \underline{V}_{\mathcal{D},0}$ to Problem (1.30)-(1.32), which satisfies $|\underline{u}_\mathcal{D}^\infty|_{1,\mathcal{D}} \le C\left(\|f\|_{L^2(\Omega)} + \|g^N\|_{L^2(\Gamma^N)}\right)$, for some $C > 0$ depending on the data, and on the discretisation $\mathcal{D}$ only through $\theta_\mathcal{D}$.*

Proof. The existence/uniqueness of $\underline{\rho}_\mathcal{D}^\infty \in \underline{V}_{\mathcal{D},0}^\omega$ solution to (1.30) (and in turn of $\underline{u}_\mathcal{D}^\infty \in \underline{V}_{\mathcal{D},0}$) is a direct consequence of the coercivity estimate (1.33), and of the fact that $|\cdot|_{1,\mathcal{D}}$ clearly defines a norm on $\underline{V}_{\mathcal{D},0}^\omega$. To prove the bound on $|\underline{u}_\mathcal{D}^\infty|_{1,\mathcal{D}}$, we use the fact that $\underline{u}_\mathcal{D}^\infty = \underline{\omega}_\mathcal{D} \times \underline{\rho}_\mathcal{D}^\infty$. For all $K \in \mathcal{M}$,

$$|\underline{u}_K^\infty|_{1,K}^2 = \sum_{\sigma\in\mathcal{E}_K}\frac{|\sigma|}{d_{K,\sigma}}(\omega_K\rho_K^\infty - \omega_\sigma\rho_\sigma^\infty)^2 \le 2\sum_{\sigma\in\mathcal{E}_K}\frac{|\sigma|}{d_{K,\sigma}}\left(\omega_\sigma^2(\rho_K^\infty - \rho_\sigma^\infty)^2 + (\rho_K^\infty)^2(\omega_K - \omega_\sigma)^2\right).$$

By definition (1.31) of $\underline{\omega}_\mathcal{D}$, and local stability of the HFV interpolant (cf. [111, Proposition B.7,

$p = 2$] and its proof), we infer

$$|\underline{u}_K^\infty|_{1,K}^2 \le 2\,\omega_\sharp^2\,|\underline{\rho}_K^\infty|_{1,K}^2 + 2\,(\rho_K^\infty)^2 C_{sta}^2 \|\nabla\omega\|_{L^2(K;\mathbb{R}^d)}^2,$$

with $C_{sta} > 0$ only depending on d and $\theta_\mathcal{D}$. Since $\nabla\omega = -\omega\nabla\phi$, $\phi \in C^1(\overline{\Omega})$, and $\omega > 0$, we have

$$\|\nabla\omega\|_{L^2(K;\mathbb{R}^d)}^2 \le \omega_\sharp \sup_{x\in\overline{\Omega}} |\nabla\phi(x)|^2 |K|\,\omega_K.$$

Summing over $K \in \mathcal{M}$ then yields

$$|\underline{u}_\mathcal{D}^\infty|_{1,\mathcal{D}}^2 \le 2\,\omega_\sharp^2\,|\underline{\rho}_\mathcal{D}^\infty|_{1,\mathcal{D}}^2 + 2\,C_{sta}^2\,\omega_\sharp \sup_{x\in\overline{\Omega}} |\nabla\phi(x)|^2 \|\rho_\mathcal{M}^\infty\|_{L^2_\omega(\Omega)}^2.$$

When $|\Gamma^D| > 0$, $\|\rho_\mathcal{M}^\infty\|_{L^2_\omega(\Omega)}^2 \le \omega_\sharp \|\rho_\mathcal{M}^\infty\|_{L^2(\Omega)}^2 \le \omega_\sharp\, C_{P,\Gamma^D}^2 |\underline{\rho}_\mathcal{D}^\infty|_{1,\mathcal{D}}^2$, where we have applied the discrete Poincaré inequality (1.83) to $\underline{\rho}_\mathcal{D}^\infty \in \underline{V}_{\mathcal{D},0}^D$. Otherwise, $\underline{\rho}_\mathcal{D}^\infty \in \underline{V}_{\mathcal{D},0}^\omega$ satisfies $\int_\Omega \omega\rho_\mathcal{M}^\infty = 0$, and one can use [51, Lemma 5.2] to infer that $\|\rho_\mathcal{M}^\infty\|_{L^2_\omega(\Omega)} \le 2\|\rho_\mathcal{M}^\infty - \frac{1}{|\Omega|}\int_\Omega \rho_\mathcal{M}^\infty\|_{L^2_\omega(\Omega)}$, and finally get that $\|\rho_\mathcal{M}^\infty\|_{L^2_\omega(\Omega)} \le 2\sqrt{\omega_\sharp}\, C_{PW} |\underline{\rho}_\mathcal{D}^\infty|_{1,\mathcal{D}}$ applying the discrete Poincaré inequality (1.82) to $\underline{\rho}_\mathcal{D}^\infty - \frac{1}{|\Omega|}\int_\Omega \rho_\mathcal{M}^\infty \underline{1}_\mathcal{D} \in \underline{V}_{\mathcal{D},0}^N$. In any case, we end up bounding $|\underline{u}_\mathcal{D}^\infty|_{1,\mathcal{D}}$ by $|\underline{\rho}_\mathcal{D}^\infty|_{1,\mathcal{D}}$, with multiplicative constant depending on the data, and on the discretisation $\mathcal{D}$ only through $\theta_\mathcal{D}$. The rest of the proof consists in bounding $|\underline{\rho}_\mathcal{D}^\infty|_{1,\mathcal{D}}$, and proceeds as in the proof of Proposition 1, using that $\|\rho_\mathcal{M}^\infty\|_{L^2(\Omega)} \le 2\sqrt{\frac{\omega_\sharp}{\omega_\flat}}\, C_{PW} |\underline{\rho}_\mathcal{D}^\infty|_{1,\mathcal{D}}$ when $|\Gamma^D| = 0$. □

In the sequel, the HFV scheme (1.30)-(1.32) will be referred to as "exponential fitting". Notice that no assumption on V^ϕ is needed to ensure its coercivity.

Remark 6 (Preservation of the thermal equilibrium)**.** *As for the exponential fitting scheme of [43], the method (1.30)-(1.32) preserves the thermal equilibrium, in the sense that $\underline{u}_\mathcal{D}^\infty = \bar{\rho}\,\underline{\omega}_\mathcal{D}$ when $u^\infty = u_{th}^\infty$ (see (1.25)). This property is analogous to what holds true for the TPFA Scharfetter–Gummel scheme.*

Evolution problem

We consider Problem (1.1). Following the previous strategy, letting $\rho = \frac{u}{\omega}$, one can show that ρ equivalently solves the following transient pure diffusion problem:

$$\begin{cases} \omega\partial_t\rho - \mathrm{div}(\omega\Lambda\nabla\rho) = f & \text{in } \mathbb{R}_+\times\Omega, \\ \rho = \omega^{-1} g^D & \text{on } \mathbb{R}_+\times\Gamma^D, \\ \omega\Lambda\nabla\rho\cdot n = g^N & \text{on } \mathbb{R}_+\times\Gamma^N, \\ \rho(0,\cdot) = \rho^{in} & \text{in } \Omega, \end{cases} \tag{1.34}$$

where $\rho^{in} = \frac{u^{in}}{\omega}$. Following Remark 2, we consider Problem (1.34) with $\rho = 0$ on Γ^D when $|\Gamma^D| > 0$, or $\int_\Omega \omega\rho^{in} = 0$ otherwise (which is equivalent to consider Problem (1.1) with $g^D = 0$ or $M = \int_\Omega u^{in} = 0$).

At the discrete level, instead of discretising (1.1), we approximate the solution to (1.34). We use a backward Euler discretisation in time, and the HFV discretisation introduced in

Section 1.3.2 in space. The discrete problem reads: Find $\left(\underline{\rho}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D},0}^{\omega}\right)_{n\geq 1}$ such that

$$\begin{cases} \dfrac{1}{\Delta t}(\rho_{\mathcal{M}}^n - \rho_{\mathcal{M}}^{n-1}, \omega_{\mathcal{M}} v_{\mathcal{M}})_{\Omega} + a_{\mathcal{D}}^{\omega}(\underline{\rho}_{\mathcal{D}}^n, \underline{v}_{\mathcal{D}}) = (f, v_{\mathcal{M}})_{\Omega} + (g^N, v_{\mathcal{E}})_{\Gamma^N} & \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}^{\omega}, \quad (1.35a) \\ \rho_K^0 = \dfrac{1}{\omega_K |K|} \displaystyle\int_K u^{in} & \forall K \in \mathcal{M}, \quad (1.35b) \end{cases}$$

where $a_{\mathcal{D}}^{\omega}$ is defined (locally) by (1.29), and the approximation $\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D},0}$ of the solution to Problem (1.1) is finally defined as $\underline{u}_{\mathcal{D}}^n = \underline{\omega}_{\mathcal{D}} \times \underline{\rho}_{\mathcal{D}}^n$, i.e., according to (1.32) (with superscript n instead of ∞). Once again, because of the coercivity of $a_{\mathcal{D}}^{\omega}$, the scheme (1.35) is unconditionally well-posed, as a straightforward consequence of Proposition 3.

Remark 7 (Pure Neumann case)**.** *When considering pure Neumann boundary conditions, as for the standard HFV scheme (see Remark 5), one can seek for a solution to Problem* (1.35) *in* $\underline{V}_{\mathcal{D}}$.

1.3.3 Nonlinear HFV scheme

We are interested in the evolution problem (1.1). We restrict our study to the pure Neumann case ($|\Gamma^D| = 0$), and to the choice of data $f = 0$, $g^N = 0$, and $u^{in} \geq 0$ with $M = \int_{\Omega} u^{in} > 0$ (see Appendix 1.B for the case of mixed Dirichlet-Neumann thermal equilibrium boundary conditions). Under these assumptions, it is known that the solution u to Problem (1.1) is strictly positive on $\mathbb{R}_+^{\star} \times \Omega$. Furthermore, in long time, $u(t)$ converges towards the thermal equilibrium. Indeed, we easily verify that the function $u^{\infty} = \rho^M \mathrm{e}^{-\phi}$, where we recall that $\rho^M = \frac{M}{\int_{\Omega} \mathrm{e}^{-\phi}} > 0$ (cf. Section 1.3.2), solves the steady problem (1.2) with same data. Since $u > 0$ on $\mathbb{R}_+^{\star} \times \Omega$, we can rewrite the flux $J = -\Lambda(\nabla u + u \nabla \phi)$ under the nonlinear form

$$J = -u\Lambda\nabla\Big(\log(u) + \phi\Big) = -u\Lambda\nabla\log\left(\frac{u}{u^{\infty}}\right).$$

At the continuous level, introducing this nonlinearity enables to highlight the following entropy/dissipation structure of the model at hand: testing the equation against $\log\left(\frac{u}{u^{\infty}}\right)$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(t) + \mathbb{D}(t) = 0, \tag{1.36}$$

where

$$\mathbb{E}(t) = \int_{\Omega} u^{\infty} \Phi_1\left(\frac{u(t)}{u^{\infty}}\right) \qquad \text{and} \qquad \mathbb{D}(t) = \int_{\Omega} u(t)\Lambda\nabla\log\left(\frac{u(t)}{u^{\infty}}\right) \cdot \nabla\log\left(\frac{u(t)}{u^{\infty}}\right), \tag{1.37}$$

with $\Phi_1(s) = s\log(s) - s + 1$ for all $s > 0$. Since $\Phi_1 \geq 0$, the relative entropy $\mathbb{E}(t)$ is a non-negative quantity (as well as the relative dissipation $\mathbb{D}(t)$). The entropy/dissipation structure (1.36)-(1.37) is instrumental to prove the exponential convergence in time of the solution $u(t)$ to Problem (1.1) towards the equilibrium u^{∞}. From the above nonlinear expression of the flux J, we build a nonlinear hybrid discretisation of the problem, leading to a scheme designed along the same principles as the nonlinear VAG and DDFV schemes of [56] and [52, 51]. This scheme is devised so as to ensure the positivity of discrete solutions, as well as to preserve at the discrete level the entropy/dissipation structure (and the long-time behaviour) of the model. The choice of designing a nonlinear HFV scheme is driven by the prospect of the design of hybrid high-order (HHO) schemes which could have similar features.

Definition of the scheme and key properties of discrete solutions

In the sequel, a vector of discrete unknowns $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ will be called positive if and only if, for all $K \in \mathcal{M}$ and all $\sigma \in \mathcal{E}$, $v_K > 0$ and $v_\sigma > 0$. Recall the definition (1.27) of $\omega = \mathrm{e}^{-\phi}$, as well as the definition (1.31) of the HFV interpolate $\underline{\omega}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ of ω. Remark that $\underline{\omega}_{\mathcal{D}}$ is positive. If $\underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ is positive, one can then define $\underline{w}_{\mathcal{D}}$ as the element of $\underline{V}_{\mathcal{D}}$ such that

$$w_K = \log\left(\frac{u_K}{\omega_K}\right) \quad \forall K \in \mathcal{M}, \qquad w_\sigma = \log\left(\frac{u_\sigma}{\omega_\sigma}\right) \quad \forall \sigma \in \mathcal{E}. \tag{1.38}$$

In what follows, to emphasise the dependency of $\underline{w}_{\mathcal{D}}$ upon $\underline{u}_{\mathcal{D}}$, we sometimes write $\underline{w}_{\mathcal{D}}(\underline{u}_{\mathcal{D}})$. Locally to any cell $K \in \mathcal{M}$, we define an approximation of

$$(u, v) \mapsto -\int_K J \cdot \nabla v = \int_K u \Lambda \nabla \log\left(\frac{u}{u^\infty}\right) \cdot \nabla v$$

under the form

$$T_K(\underline{u}_K, \underline{w}_K, \underline{v}_K) = \int_K r_K(\underline{u}_K) \Lambda \nabla_K \underline{w}_K \cdot \nabla_K \underline{v}_K,$$

for all $\underline{u}_K \in \underline{V}_K$ positive and all $\underline{v}_K \in \underline{V}_K$, where $r_K : \left(\underline{V}_K\right)_+^\star \to \mathbb{R}_+^\star$ is a local reconstruction operator. Since $r_K(\underline{u}_K)$ is a (positive) constant on K, we have

$$T_K(\underline{u}_K, \underline{w}_K, \underline{v}_K) = r_K(\underline{u}_K) a_K^\Lambda\left(\underline{w}_K, \underline{v}_K\right), \tag{1.39}$$

where a_K^Λ is defined by (1.6). Following (1.12), one can equivalently reformulate (1.39) using the local matrix $\mathbb{A}_K$ defined by (1.10):

$$T_K(\underline{u}_K, \underline{w}_K, \underline{v}_K) = r_K(\underline{u}_K) \delta_K \underline{v}_K \cdot \mathbb{A}_K \delta_K \underline{w}_K. \tag{1.40}$$

As already pointed out in the analysis of the nonlinear DDFV scheme of [52, 51], the definition of the local reconstruction operator is crucial to guarantee the existence of solutions and a good long-time behaviour to the scheme. The most natural choice in the HFV context would obviously be $r_K(\underline{u}_K) = u_K$, however it turns out that such a reconstruction embeds too few information on $\underline{u}_K$ to conclude, as already suggested in [49]. Therefore, we use a richer reconstruction, described below, which embeds information from both the local cell and face unknowns. For $\underline{u}_K \in \underline{V}_K$ positive, we let

$$r_K(\underline{u}_K) = f_{|\mathcal{E}_K|}\Big(\big(m(u_K, u_\sigma)\big)_{\sigma \in \mathcal{E}_K}\Big), \tag{1.41}$$

with $m : (\mathbb{R}_+^\star)^2 \to \mathbb{R}_+^\star$ and, for $k \geq 1$ integer, $f_k : (\mathbb{R}_+^\star)^k \to \mathbb{R}_+^\star$, such that

$$m \text{ is non-decreasing with respect to both its variables,} \tag{1.42a}$$

$$m(x,x) = x \text{ for all } x \in \mathbb{R}_+^\star \text{ and } m(y,x) = m(x,y) \text{ for all } (x,y) \in (\mathbb{R}_+^\star)^2, \tag{1.42b}$$

$$m(\lambda x, \lambda y) = \lambda m(x,y) \text{ for all } \lambda > 0 \text{ and all } (x,y) \in (\mathbb{R}_+^\star)^2, \tag{1.42c}$$

$$\frac{y-x}{\log(y) - \log(x)} \leq m(x,y) \leq \max(x,y) \text{ for all } (x,y) \in (\mathbb{R}_+^\star)^2,\ x \neq y, \tag{1.42d}$$

and

$$f_k(x_1,\dots,x_k)=\frac{1}{k}\sum_{i=1}^{k}x_i \qquad \text{or} \qquad f_k(x_1,\dots,x_k)=\max(x_1,\dots,x_k). \tag{1.43}$$

Note that, for all $(x,y)\in(\mathbb{R}_+^\star)^2$, one has

$$\frac{y-x}{\log(y)-\log(x)}\le\left(\frac{\sqrt{x}+\sqrt{y}}{2}\right)^2\le\frac{x+y}{2}\le\max(x,y),$$

and each expression of the previous sequence is a mean function m satisfying the properties (1.42). Heuristically, $r_K(\underline{u}_K)$ computes an average of the unknowns attached to the cell K, especially it contains information about all the local face unknowns. As far as the properties (1.42)-(1.43) are concerned, they will be instrumental to prove Lemma 2 and Proposition 6 below. As now standard, we finally let $T_\mathcal{D}$ be such that, for all $\underline{u}_\mathcal{D}\in\underline{V}_\mathcal{D}$ positive, and all $\underline{v}_\mathcal{D}\in\underline{V}_\mathcal{D}$,

$$T_\mathcal{D}(\underline{u}_\mathcal{D},\underline{w}_\mathcal{D},\underline{v}_\mathcal{D})=\sum_{K\in\mathcal{M}}T_K(\underline{u}_K,\underline{w}_K,\underline{v}_K), \tag{1.44}$$

where the local contributions T_K are defined by (1.39).

Using a backward Euler discretisation in time, and the HFV discretisation we have just introduced in space, our discrete problem reads: Find $\left(\underline{u}_\mathcal{D}^n\in\underline{V}_\mathcal{D}\right)_{n\ge1}$ such that

$$\begin{cases}\dfrac{1}{\Delta t}(u_\mathcal{M}^n-u_\mathcal{M}^{n-1},v_\mathcal{M})_\Omega+T_\mathcal{D}(\underline{u}_\mathcal{D}^n,\underline{w}_\mathcal{D}(\underline{u}_\mathcal{D}^n),\underline{v}_\mathcal{D})=0 & \forall\underline{v}_\mathcal{D}\in\underline{V}_\mathcal{D}, \quad (1.45a)\\ u_K^0=\dfrac{1}{|K|}\displaystyle\int_K u^{in} & \forall K\in\mathcal{M}. \quad (1.45b)\end{cases}$$

Notice that if $(\underline{u}_\mathcal{D}^n)_{n\ge1}$ solves Problem (1.45), then, necessarily, $\underline{u}_\mathcal{D}^n$ is positive for all $n\ge1$. Therefore, in the sequel, we will speak about the positive solutions to (1.45). Notice also that $u_\mathcal{M}^0$ may vanish in some cells of the mesh, since we only impose that $u^{in}\ge0$ (but $u_\mathcal{M}^0$ cannot be identically zero in Ω since $M>0$). Notice finally that $u_\mathcal{E}^0$ needs not be defined, as the scheme only uses $u_\mathcal{M}^0$.

Testing (1.45a) with $\underline{v}_\mathcal{D}=\underline{1}_\mathcal{D}$, and remarking that $T_\mathcal{D}(\underline{u}_\mathcal{D}^n,\underline{w}_\mathcal{D}(\underline{u}_\mathcal{D}^n),\underline{1}_\mathcal{D})=0$ for all $n\ge1$, we immediately infer the following discrete mass conservation property.

Proposition 4 (Mass conservation). *If $\left(\underline{u}_\mathcal{D}^n\in\underline{V}_\mathcal{D}\right)_{n\ge1}$ is a (positive) solution to (1.45), then*

$$\forall n\in\mathbb{N}^\star, \qquad \int_\Omega u_\mathcal{M}^n=\int_\Omega u_\mathcal{M}^0=\int_\Omega u^{in}=M.$$

Following Proposition 4, a discrete steady-state $\underline{u}_\mathcal{D}^\infty\in\underline{V}_\mathcal{D}$ of (1.45) shall satisfy

$$T_\mathcal{D}(\underline{u}_\mathcal{D}^\infty,\underline{w}_\mathcal{D}(\underline{u}_\mathcal{D}^\infty),\underline{v}_\mathcal{D})=0 \qquad \forall\underline{v}_\mathcal{D}\in\underline{V}_\mathcal{D}, \tag{1.46}$$

and $\int_\Omega u_\mathcal{M}^\infty=M$. Letting $\underline{w}_\mathcal{D}^\infty=\underline{w}_\mathcal{D}(\underline{u}_\mathcal{D}^\infty)$, and testing (1.46) with $\underline{v}_\mathcal{D}=\underline{w}_\mathcal{D}^\infty$, by (1.39) and (1.44), since $r_K(\underline{u}_K^\infty)>0$ for all $K\in\mathcal{M}$, we necessarily have $a_\mathcal{D}^\Lambda(\underline{w}_\mathcal{D}^\infty,\underline{w}_\mathcal{D}^\infty)=0$, which yields $|\underline{w}_\mathcal{D}^\infty|_{1,\mathcal{D}}=0$ by the coercivity property (1.14). Hence, $\underline{w}_\mathcal{D}^\infty=c\,\underline{1}_\mathcal{D}$ for some constant $c\in\mathbb{R}$, and since $\int_\Omega u_\mathcal{M}^\infty=M$, by (1.38), we necessarily have $\mathrm{e}^c=\rho^M$, that is $c=\log(\rho^M)$ and $\underline{w}_\mathcal{D}^\infty=\log(\rho^M)\underline{1}_\mathcal{D}$. As a consequence, again by (1.38), $\underline{u}_\mathcal{D}^\infty=\rho^M\underline{\omega}_\mathcal{D}$, i.e., $\underline{u}_\mathcal{D}^\infty$ is the HFV interpolate of u_{th}^∞. Thus, just like the exponential

fitting scheme (cf. Remark 6), the nonlinear scheme preserves the thermal equilibrium. We notice that $\underline{w}_{\mathcal{D}}$, first defined by (1.38), can actually be modified up to an additive constant without any impact on the scheme (1.45). Hence, we can redefine $\underline{w}_{\mathcal{D}}$ as

$$w_K = \log\left(\frac{u_K}{u_K^\infty}\right) \quad \forall K \in \mathcal{M}, \qquad w_\sigma = \log\left(\frac{u_\sigma}{u_\sigma^\infty}\right) \quad \forall \sigma \in \mathcal{E}. \tag{1.47}$$

Another important consequence of the fact that $\underline{u}_{\mathcal{D}}^\infty$ is the HFV interpolate of u_{th}^∞ is the following. Letting $u_\flat^\infty = \frac{M\omega_\flat}{|\Omega|\omega_\sharp} > 0$ and $u_\sharp^\infty = \frac{M\omega_\sharp}{|\Omega|\omega_\flat} > 0$ (recall that $\omega_\flat$ and $\omega_\sharp$ only depend on ϕ and Ω), we have $u_\flat^\infty \le u_{th}^\infty \le u_\sharp^\infty$, but we also have that

$$u_\flat^\infty \underline{1}_{\mathcal{D}} \le \underline{u}_{\mathcal{D}}^\infty \le u_\sharp^\infty \underline{1}_{\mathcal{D}}, \tag{1.48}$$

where the inequalities shall be understood coordinate-wise. In other words, the continuous bounds on the steady-state are transferred to the discrete level.

Given a (positive) solution $\left(\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}\right)_{n\ge1}$ to (1.45), we define the following discrete versions of the relative entropy and dissipation introduced in (1.37): for all $n \ge 1$,

$$\mathbb{E}^n = \int_\Omega u_{\mathcal{M}}^\infty \Phi_1\left(\frac{u_{\mathcal{M}}^n}{u_{\mathcal{M}}^\infty}\right) \qquad \text{and} \qquad \mathbb{D}^n = T_{\mathcal{D}}\left(\underline{u}_{\mathcal{D}}^n, \underline{w}_{\mathcal{D}}^n, \underline{w}_{\mathcal{D}}^n\right), \tag{1.49}$$

where we let $\underline{w}_{\mathcal{D}}^n = \underline{w}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}^n)$, and we recall that $\Phi_1(s) = s\log(s) - s + 1$ for all $s > 0$. Notice that $\mathbb{E}^n \ge 0$ for all $n \ge 1$ since $\Phi_1 \ge 0$. For further use, we extend the function Φ_1 by continuity to 0, letting $\Phi_1(0) = 1$. We can then define, in case there exists $K \in \mathcal{M}$ such that $u_K^0 = 0$, $\mathbb{E}^0 \ge 0$ according to (1.49). As far as $\mathbb{D}^n$ is concerned, by (1.39) and (1.44), for all $n \ge 1$, we have

$$\mathbb{D}^n = \sum_{K\in\mathcal{M}} r_K(\underline{u}_K^n) a_K^\Lambda(\underline{w}_K^n, \underline{w}_K^n) \ge 0.$$

We can now establish the following discrete counterpart of (1.36).

Proposition 5 (Entropy dissipation). *If $\left(\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}\right)_{n\ge1}$ is a (positive) solution to (1.45), then*

$$\forall n \in \mathbb{N}, \qquad \frac{\mathbb{E}^{n+1} - \mathbb{E}^n}{\Delta t} + \mathbb{D}^{n+1} \le 0. \tag{1.50}$$

Proof. Let $n \in \mathbb{N}$. By the expression (1.49) of the discrete relative entropy, and the convexity of Φ_1, we have

$$\mathbb{E}^{n+1} - \mathbb{E}^n \le \sum_{K\in\mathcal{M}} |K| u_K^\infty \Phi_1'\left(\frac{u_K^{n+1}}{u_K^\infty}\right)\left(\frac{u_K^{n+1} - u_K^n}{u_K^\infty}\right).$$

Thus, by (1.47), we get

$$\mathbb{E}^{n+1} - \mathbb{E}^n \le \int_\Omega \left(u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^n\right)\log\left(\frac{u_{\mathcal{M}}^{n+1}}{u_{\mathcal{M}}^\infty}\right) = (u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^n, w_{\mathcal{M}}^{n+1})_\Omega.$$

By (1.45a) and (1.49), we finally infer

$$\mathbb{E}^{n+1} - \mathbb{E}^n \le -\Delta t\, T_{\mathcal{D}}(\underline{u}_{\mathcal{D}}^{n+1}, \underline{w}_{\mathcal{D}}^{n+1}, \underline{w}_{\mathcal{D}}^{n+1}) = -\Delta t\, \mathbb{D}^{n+1},$$

which yields (1.50). □

We finally state the main result of Section 1.3.3, about the existence of (positive) solutions to the nonlinear scheme (1.45). The proof of this result is the subject of the next subsection.

Theorem 1 (Existence of positive solutions)**.** *Let $u^{in} \in L^2(\Omega)$ be a non-negative function such that $\int_\Omega u^{in} = M > 0$. There exists at least one positive solution $\left(\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}\right)_{n\geq 1}$ to the nonlinear scheme (1.45). Moreover, there exists $\varepsilon > 0$, depending on Λ, ϕ, u^{in}, M, Ω, d, Δt, and $\mathcal{D}$ such that*

$$\forall n \geq 1, \qquad u_K^n \geq \varepsilon \quad \forall K \in \mathcal{M} \quad \text{and} \quad u_\sigma^n \geq \varepsilon \quad \forall \sigma \in \mathcal{E}. \tag{1.51}$$

The uniform-in-time positivity result (1.51) on discrete solutions is the equivalent in the HFV context of [56, Lemma 3.7] and [52, Lemma 3.5] obtained, respectively, in the VAG and DDFV contexts.

Existence of discrete solutions

The existence of discrete solutions to the nonlinear scheme (1.45) is proved in two steps. First, we introduce a regularised scheme, for which we prove the existence of solutions by a fixed-point argument, inspired from the proof of existence in [25]. Then, we prove that sequences of regularised solutions satisfy uniform a priori bounds, which allows us to pass to the limit in the regularisation parameter. Notice that our proof of existence uses the same estimates, but follows a quite different path than the ones in the VAG [56] and DDFV [52] contexts, in which the proof is based on the topological degree, together with a monotonicity argument. Henceforth, we reason in the $\underline{w}_{\mathcal{D}}$ variable, and we recall that $\underline{u}_{\mathcal{D}} = \underline{u}_{\mathcal{D}}^\infty \times \exp(\underline{w}_{\mathcal{D}})$ according to (1.47). The advantage of doing so is that we can seek for solutions $\underline{w}_{\mathcal{D}}$ in the whole space $\underline{V}_{\mathcal{D}}$, with bijective correspondence with solutions $\underline{u}_{\mathcal{D}}$ that are automatically positive. Recalling the definition (1.49) of the discrete relative entropy and dissipation, and using (1.44) combined with (1.40), we let, for all $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$,

$$\mathbb{E}(w_{\mathcal{M}}) = \sum_{K\in\mathcal{M}} |K| u_K^\infty \Phi_1(\mathrm{e}^{w_K}), \qquad \mathbb{D}(\underline{w}_{\mathcal{D}}) = \sum_{K\in\mathcal{M}} r_K\Big(\underline{u}_K^\infty \times \exp(\underline{w}_K)\Big)\delta_K \underline{w}_K \cdot \mathbb{A}_K \delta_K \underline{w}_K, \tag{1.52}$$

in such a way that $\mathbb{E}^n = \mathbb{E}\Big(w_{\mathcal{M}}(u_{\mathcal{M}}^n)\Big)$ and $\mathbb{D}^n = \mathbb{D}\Big(\underline{w}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}^n)\Big)$ for all $n \geq 1$. Using the fact that $\Phi_1(0) = 1$, we extend the definition of $\mathbb{E}(w_{\mathcal{M}})$ to the case where some w_K's are equal to $-\infty$.

Before proceeding with the proof of Theorem 1, we state two preliminary lemmas. The first one, that can be found, e.g., in [119, Section 9.1], is a corollary of Brouwer's fixed-point theorem. This result is instrumental to show the existence of solutions to the regularised scheme.

Lemma 1. *Let $N \in \mathbb{N}^\star$, and let $P : \mathbb{R}^N \to \mathbb{R}^N$ be a continuous vector field. Assume that there is $r > 0$ such that*

$$P(x) \cdot x \geq 0 \qquad \text{if } |x| = r.$$

Then, there exists a point $x_0 \in \mathbb{R}^N$ such that $P(x_0) = 0$ and $|x_0| \leq r$.

The second lemma, whose proof is postponed until Appendix 1.C.1, establishes sufficient boundedness conditions on the discrete mass and (relative) dissipation so that a priori bounds hold for vectors of discrete unknowns. This result is instrumental to show that sequences of regularised solutions satisfy (uniform) a priori bounds.

Lemma 2. *Let $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, and assume that there exist $C_\sharp > 0$, and $M_\sharp \geq M_\flat > 0$ such that*

$$M_\flat \leq \sum_{K\in\mathcal{M}} |K| u_K^\infty \, \mathrm{e}^{w_K} \leq M_\sharp \qquad \text{and} \qquad \mathbb{D}(\underline{w}_{\mathcal{D}}) \leq C_\sharp. \tag{1.53}$$

Then, there exists $C > 0$, depending on Λ, $u_\flat^\infty$, $u_\sharp^\infty$, $M_\flat$, $M_\sharp$, $C_\sharp$, Ω, d, and $\mathcal{D}$ such that

$$|w_K| \leq C \quad \forall K \in \mathcal{M} \quad \text{and} \quad |w_\sigma| \leq C \quad \forall \sigma \in \mathcal{E}.$$

We can now proceed with the proof of Theorem 1. Let us first define the following inner product and corresponding norm on the space $\underline{V}_{\mathcal{D}}$: for all $\underline{z}_{\mathcal{D}}, \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$,

$$\langle \underline{z}_{\mathcal{D}}, \underline{v}_{\mathcal{D}} \rangle = \sum_{K\in\mathcal{M}} z_K v_K + \sum_{\sigma\in\mathcal{E}} z_\sigma v_\sigma \quad \text{and} \quad \|\underline{v}_{\mathcal{D}}\| = \sqrt{\langle \underline{v}_{\mathcal{D}}, \underline{v}_{\mathcal{D}} \rangle}.$$

Letting $N = |\mathcal{M}| + |\mathcal{E}|$, and identifying $\underline{V}_{\mathcal{D}}$ to $\mathbb{R}^N$, the inner product $\langle \cdot, \cdot \rangle$ is nothing but the standard inner product on $\mathbb{R}^N$. For all $K \in \mathcal{M}$, and all $\sigma \in \mathcal{E}_K$, we let, for $\underline{u}_K \in \underline{V}_K$ positive,

$$\mathcal{F}_{K,\sigma}^{\mathrm{nl}}(\underline{u}_K) = r_K(\underline{u}_K) \sum_{\sigma'\in\mathcal{E}_K} A_K^{\sigma\sigma'} \left(\log\left(\frac{u_K}{u_K^\infty}\right) - \log\left(\frac{u_{\sigma'}}{u_{\sigma'}^\infty}\right) \right), \tag{1.54}$$

where the $A_K^{\sigma\sigma'}$ are defined by (1.10). Combining (1.39) and (1.47) with (1.11) and (1.9), there holds that $T_K(\underline{u}_K, \underline{w}_K, \underline{v}_K) = \sum_{\sigma\in\mathcal{E}_K} \mathcal{F}_{K,\sigma}^{\mathrm{nl}}(\underline{u}_K)(v_K - v_\sigma)$ for all $\underline{v}_K \in \underline{V}_K$. In what follows, we let $n \in \mathbb{N}^\star$ and $u_{\mathcal{M}}^{n-1} \geq 0$ be given. We assume that $M^{n-1} = \int_\Omega u_{\mathcal{M}}^{n-1} > 0$ and that $u_{\mathcal{M}}^{n-1} > 0$ if $n > 1$. We also assume that $\mathbb{E}(w_{\mathcal{M}}^{n-1}) > 0$. If $\mathbb{E}(w_{\mathcal{M}}^{n-1}) = 0$ (which is equivalent to $u_{\mathcal{M}}^{n-1} = u_{\mathcal{M}}^\infty$), then necessarily, by (1.50), $\underline{u}_{\mathcal{D}}^n = \underline{u}_{\mathcal{D}}^\infty$ uniquely solves (1.45a). Letting, for any $\underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ positive, $\underline{\mathcal{G}}_{\mathcal{D}}^n(\underline{u}_{\mathcal{D}})$ be the element of $\underline{V}_{\mathcal{D}}$ such that

$$\mathcal{G}_K^n(\underline{u}_{\mathcal{D}}) = |K| \frac{u_K - u_K^{n-1}}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} \mathcal{F}_{K,\sigma}^{\mathrm{nl}}(\underline{u}_K) \qquad \forall K \in \mathcal{M}, \tag{1.55a}$$

$$\mathcal{G}_\sigma^n(\underline{u}_{\mathcal{D}}) = -\Big(\mathcal{F}_{K,\sigma}^{\mathrm{nl}}(\underline{u}_K) + \mathcal{F}_{L,\sigma}^{\mathrm{nl}}(\underline{u}_L) \Big) \qquad \forall \sigma = K \mid L \in \mathcal{E}_{int}, \tag{1.55b}$$

$$\mathcal{G}_\sigma^n(\underline{u}_{\mathcal{D}}) = -\mathcal{F}_{K,\sigma}^{\mathrm{nl}}(\underline{u}_K) \qquad \forall \sigma \in \mathcal{E}_{ext} \text{ with } \mathcal{M}_\sigma = \{K\}, \tag{1.55c}$$

we infer that, for all $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$,

$$\frac{1}{\Delta t} \left(u_{\mathcal{M}} - u_{\mathcal{M}}^{n-1}, v_{\mathcal{M}} \right)_\Omega + T_{\mathcal{D}}(\underline{u}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}), \underline{v}_{\mathcal{D}}) = \left\langle \underline{\mathcal{G}}_{\mathcal{D}}^n(\underline{u}_{\mathcal{D}}), \underline{v}_{\mathcal{D}} \right\rangle. \tag{1.56}$$

Hence, a positive vector $\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}$ is a solution to the nonlinear equation (1.45a) if and only if $\underline{\mathcal{G}}_{\mathcal{D}}^n(\underline{u}_{\mathcal{D}}^n) = \underline{0}_{\mathcal{D}}$. With this observation in hand, we now detail the two steps of the proof.

Step 1: Using the relation $\underline{u}_{\mathcal{D}} = \underline{u}_{\mathcal{D}}^\infty \times \exp(\underline{w}_{\mathcal{D}})$, we define the vector field $\underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu} : \underline{V}_{\mathcal{D}} \to \underline{V}_{\mathcal{D}}$ such that, for all $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$,

$$\underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}(\underline{w}_{\mathcal{D}}) = \underline{\mathcal{G}}_{\mathcal{D}}^n\Big(\underline{u}_{\mathcal{D}}^\infty \times \exp(\underline{w}_{\mathcal{D}}) \Big) + \mu \, \underline{w}_{\mathcal{D}}, \tag{1.57}$$

with $\underline{\mathcal{G}}_{\mathcal{D}}^n$ defined by (1.55) and $\mu \geq 0$. Notice that, unlike $\underline{\mathcal{G}}_{\mathcal{D}}^n$, the vector field $\underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}$ is continuous on the whole space $\underline{V}_{\mathcal{D}}$ for any $\mu \geq 0$. If $\underline{w}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}$ satisfies $\underline{\mathcal{P}}_{\mathcal{D}}^{n,0}(\underline{w}_{\mathcal{D}}^n) = \underline{0}_{\mathcal{D}}$, then letting $\underline{u}_{\mathcal{D}}^n =$

$\underline{u}_{\mathcal{D}}^{\infty} \times \exp(\underline{w}_{\mathcal{D}}^{n})$, we have $\underline{\mathcal{G}}_{\mathcal{D}}^{n}(\underline{u}_{\mathcal{D}}^{n}) = \underline{0}_{\mathcal{D}}$, therefore $\underline{u}_{\mathcal{D}}^{n}$ is a (positive) solution to (1.45a). For $\mu > 0$, the problem of finding $\underline{w}_{\mathcal{D}}^{n,\mu} \in \underline{V}_{\mathcal{D}}$ such that $\underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}(\underline{w}_{\mathcal{D}}^{n,\mu}) = \underline{0}_{\mathcal{D}}$ can thus be seen as a regularisation of the original problem. By (1.57) and (1.56), for all $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, we have

$$\begin{aligned}\left\langle \underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \right\rangle = \sum_{K\in\mathcal{M}} \frac{|K|}{\Delta t}\left(u_K^{\infty}\,\mathrm{e}^{w_K} - u_K^{n-1}\right) w_K \\ + \sum_{K\in\mathcal{M}} r_K\Big(\underline{u}_K^{\infty} \times \exp(\underline{w}_K)\Big)\delta_K \underline{w}_K \cdot \mathbb{A}_K \delta_K \underline{w}_K + \mu \|\underline{w}_{\mathcal{D}}\|^2.\end{aligned} \tag{1.58}$$

By (1.52), we recognise in the second term of the right-hand side the quantity $\mathbb{D}(\underline{w}_{\mathcal{D}}) \geq 0$. As far as the first term is concerned, for $n > 1$, by positivity of the $(u_K^{n-1})_{K\in\mathcal{M}}$, there exist real numbers $(w_K^{n-1})_{K\in\mathcal{M}}$ such that $u_K^{n-1} = u_K^{\infty}\,\mathrm{e}^{w_K^{n-1}}$ for all $K \in \mathcal{M}$, and since Φ_1 is convex,

$$\begin{aligned}\sum_{K\in\mathcal{M}} \frac{|K|}{\Delta t}\left(u_K^{\infty}\,\mathrm{e}^{w_K} - u_K^{n-1}\right) w_K &= \sum_{K\in\mathcal{M}} \frac{|K|}{\Delta t} u_K^{\infty} w_K \left(\mathrm{e}^{w_K} - \mathrm{e}^{w_K^{n-1}}\right) \\ &\geq \sum_{K\in\mathcal{M}} \frac{|K|}{\Delta t} u_K^{\infty}\left(\Phi_1\left(\mathrm{e}^{w_K}\right) - \Phi_1\left(\mathrm{e}^{w_K^{n-1}}\right)\right) = \frac{\mathbb{E}(w_{\mathcal{M}}) - \mathbb{E}(w_{\mathcal{M}}^{n-1})}{\Delta t},\end{aligned} \tag{1.59}$$

where we have used the definition (1.52) of $\mathbb{E}(w_{\mathcal{M}})$. For $n = 1$, now, it may happen that u_K^0 be zero for some $K \in \mathcal{M}$, and then w_K^0 such that $u_K^0 = u_K^{\infty}\,\mathrm{e}^{w_K^0}$ cannot be defined. However, letting in that case $w_K^0 = -\infty$, the inequality above still holds since $\Phi_1(0) = 1$ and $\Phi_1(\mathrm{e}^s) - 1 \leq s\,\mathrm{e}^s$ for all $s \in \mathbb{R}$. By non-negativity of $\mathbb{E}(w_{\mathcal{M}})$, we finally infer from (1.58) and (1.59) that

$$\left\langle \underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \right\rangle \geq \mu\|\underline{w}_{\mathcal{D}}\|^2 - \frac{\mathbb{E}(w_{\mathcal{M}}^{n-1})}{\Delta t},$$

so that, for $\mu > 0$, there holds $\left\langle \underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \right\rangle \geq 0$ if $\|\underline{w}_{\mathcal{D}}\| = \sqrt{\frac{\mathbb{E}(w_{\mathcal{M}}^{n-1})}{\mu\Delta t}} > 0$. By Lemma 1, we then conclude about the existence of solutions to the regularised scheme. There exists $\underline{w}_{\mathcal{D}}^{n,\mu} \in \underline{V}_{\mathcal{D}}$ such that

$$\underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}(\underline{w}_{\mathcal{D}}^{n,\mu}) = \underline{0}_{\mathcal{D}} \quad \text{and} \quad \|\underline{w}_{\mathcal{D}}^{n,\mu}\| \leq \sqrt{\frac{\mathbb{E}(w_{\mathcal{M}}^{n-1})}{\mu\Delta t}}. \tag{1.60}$$

Step 2: Since $\left\langle \underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}(\underline{w}_{\mathcal{D}}^{n,\mu}), \underline{w}_{\mathcal{D}}^{n,\mu} \right\rangle = 0$, by (1.58) and (1.59), we have

$$\frac{\mathbb{E}(w_{\mathcal{M}}^{n,\mu})}{\Delta t} + \mathbb{D}(\underline{w}_{\mathcal{D}}^{n,\mu}) + \mu\|\underline{w}_{\mathcal{D}}^{n,\mu}\|^2 \leq \frac{\mathbb{E}(w_{\mathcal{M}}^{n-1})}{\Delta t}.$$

The three terms on the left-hand side being non-negative, we infer that

$$\mathbb{D}(\underline{w}_{\mathcal{D}}^{n,\mu}) \leq C_\sharp, \tag{1.61}$$

with $C_\sharp = \frac{\mathbb{E}(w_{\mathcal{M}}^{n-1})}{\Delta t} > 0$. Moreover, since $\left\langle \underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}(\underline{w}_{\mathcal{D}}^{n,\mu}), \underline{1}_{\mathcal{D}} \right\rangle = 0$, by (1.57) and (1.56), we have

$$\sum_{K\in\mathcal{M}} |K| u_K^{\infty}\,\mathrm{e}^{w_K^{n,\mu}} - M^{n-1} = -\mu\Delta t \left\langle \underline{w}_{\mathcal{D}}^{n,\mu}, \underline{1}_{\mathcal{D}} \right\rangle.$$

Applying a Cauchy–Schwarz inequality, and recalling the bound (1.60), we obtain

$$\left|\sum_{K\in\mathcal{M}}|K|u_K^\infty \mathrm{e}^{w_K^{n,\mu}} - M^{n-1}\right| \le \mu\Delta t\|\underline{1}_{\mathcal{D}}\|\|\underline{w}_{\mathcal{D}}^{n,\mu}\| \le \sqrt{\mu}\sqrt{N\Delta t\,\mathbb{E}(w_{\mathcal{M}}^{n-1})},$$

so that, letting $\mu_0 = \frac{(M^{n-1})^2}{4N\Delta t\,\mathbb{E}(w_{\mathcal{M}}^{n-1})} > 0$, the following holds for all $0 < \mu \le \mu_0$:

$$0 < \frac{M^{n-1}}{2} = M_\flat \le \sum_{K\in\mathcal{M}}|K|u_K^\infty \mathrm{e}^{w_K^{n,\mu}} \le M_\sharp = \frac{3M^{n-1}}{2}. \tag{1.62}$$

By (1.62) and (1.61), we infer that $\underline{w}_{\mathcal{D}}^{n,\mu}$ satisfies (1.53) for μ sufficiently small with constants that are uniform in μ, so that by Lemma 2 the family $(\underline{w}_{\mathcal{D}}^{n,\mu})_{0<\mu\le\mu_0}$ is bounded uniformly in μ. As a consequence, by compactness, there is $\underline{w}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}$ such that, up to extraction, $\underline{w}_{\mathcal{D}}^{n,\mu}$ converges towards $\underline{w}_{\mathcal{D}}^n$ when μ tends to zero. Since $\underline{\mathcal{P}}_{\mathcal{D}}^{n,\mu}$ converges to $\underline{\mathcal{P}}_{\mathcal{D}}^{n,0}$ as μ tends to zero, we finally infer that $\underline{\mathcal{P}}_{\mathcal{D}}^{n,0}(\underline{w}_{\mathcal{D}}^n) = \underline{0}_{\mathcal{D}}$ (also, $\sum_{K\in\mathcal{M}}|K|u_K^\infty \mathrm{e}^{w_K^n} = M^{n-1}$).

Conclusion: Letting $\underline{u}_{\mathcal{D}}^n = \underline{u}_{\mathcal{D}}^\infty \times \exp(\underline{w}_{\mathcal{D}}^n)$, we have $\underline{\mathcal{G}}_{\mathcal{D}}^n(\underline{u}_{\mathcal{D}}^n) = \underline{0}_{\mathcal{D}}$, therefore $\underline{u}_{\mathcal{D}}^n$ is a (positive) solution to (1.45a). By Propositions 4 and 5, and since $\mathbb{D}\big(\underline{w}_{\mathcal{D}}^n\big) = \mathbb{D}^n$, $\mathbb{E}^n$ is non-negative, and $\mathbb{E}^n$ is non-increasing in n according to (1.50), we deduce that

$$\sum_{K\in\mathcal{M}}|K|u_K^\infty \mathrm{e}^{w_K^n} = M \quad \text{and} \quad \mathbb{D}(\underline{w}_{\mathcal{D}}^n) \le \frac{\mathbb{E}^0}{\Delta t}.$$

By Lemma 2 (recall also that $u_\flat^\infty, u_\sharp^\infty$ only depend on M, ϕ, and Ω), there exists $C > 0$, depending on Λ, ϕ, u^{in}, M, Ω, d, Δt, and $\mathcal{D}$, but not on $n \in \mathbb{N}^\star$, such that $-C\underline{1}_{\mathcal{D}} \le \underline{w}_{\mathcal{D}}^n \le C\underline{1}_{\mathcal{D}}$. By (1.48), we finally infer that $\underline{u}_{\mathcal{D}}^n \ge \varepsilon\underline{1}_{\mathcal{D}}$, with $\varepsilon = u_\flat^\infty \mathrm{e}^{-C} > 0$ still independent of $n \in \mathbb{N}^\star$. This concludes the proof of Theorem 1.

1.4 Long-time behaviour

In this section, we analyse the long-time behaviour of the three HFV schemes we have introduced in Section 1.3, thereby proving the main results of this paper.

Remark 8 (Linear schemes and nonhomogeneous data)**.** *In order to stay consistent with Section 1.3, we here below state our asymptotic results of Theorems 2 and 3, which respectively concern the (linear) standard and exponential fitting schemes, for discrete problems that feature homogeneous data (i.e., $g^D = 0$ when $|\Gamma^D| > 0$, or $M = 0$ otherwise). Nonetheless, Theorems 2 and 3 remain valid in the general case of nonhomogeneous data (we refer to Remark 2 for the straightforward adaptation of the schemes to this situation). Indeed, the proofs of the latter results only hinge on the fact that the difference between the discrete transient and steady-state solutions belongs to the homogeneous space $\underline{V}_{\mathcal{D},0}$ or $\underline{V}_{\mathcal{D},0}^\omega$, which is always true. This remark does not apply, however, to the nonlinear scheme.*

1.4.1 Asymptotic behaviour of the standard HFV scheme

We recall that u is the solution to Problem (1.1), and that u^∞ is the corresponding steady-state, solution to Problem (1.2), and we consider the following definition of the relative entropy

and dissipation:

$$\mathbb{E}(t) = \frac{1}{2}\|u(t) - u^\infty\|^2_{L^2(\Omega)}, \qquad \mathbb{D}(t) = \int_\Omega \Big(\Lambda\nabla(u(t)-u^\infty) - (u(t)-u^\infty)V^\phi\Big)\cdot\nabla(u(t)-u^\infty).$$

It can be easily verified that the following entropy/dissipation relation holds at the continuous level:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(t) + \mathbb{D}(t) = 0.$$

It is assumed that V^ϕ is such that $\mathbb{D}(t) \geq C\|\nabla(u(t)-u^\infty)\|^2_{L^2(\Omega;\mathbb{R}^d)}$ for some $C > 0$, so that $\mathbb{D}(t)$ indeed defines a dissipation.

At the discrete level, recalling that $\big(\underline{u}^n_\mathcal{D} \in \underline{V}_{\mathcal{D},0}\big)_{n\geq 1}$ is the solution to Problem (1.26), and that $\underline{u}^\infty_\mathcal{D} \in \underline{V}_{\mathcal{D},0}$ is the corresponding steady-state, solution to Problem (1.20), we consider the following equivalents of the relative entropy and dissipation: for all $n \in \mathbb{N}^\star$,

$$\mathbb{E}^n = \frac{1}{2}\left\|u^n_\mathcal{M} - u^\infty_\mathcal{M}\right\|^2_{L^2(\Omega)}, \qquad \mathbb{D}^n = a_\mathcal{D}\Big(\underline{u}^n_\mathcal{D} - \underline{u}^\infty_\mathcal{D}, \underline{u}^n_\mathcal{D} - \underline{u}^\infty_\mathcal{D}\Big),$$

where the discrete bilinear form $a_\mathcal{D}$ is defined by (1.15) and (1.19). The definition of the relative entropy is seamlessly extended to the case $n = 0$. Our main result on the standard HFV scheme is the following.

Theorem 2 (Asymptotic stability)**.** *Assume that the advection field V^ϕ satisfies the conditions (1.21) of Proposition 2, with constant β. Then, the following discrete entropy/dissipation relation holds true:*

$$\forall n \in \mathbb{N}, \qquad \frac{\mathbb{E}^{n+1} - \mathbb{E}^n}{\Delta t} + \mathbb{D}^{n+1} \leq 0. \tag{1.63}$$

Furthermore, the discrete entropy decays exponentially fast in time: there is $\nu = \frac{2\kappa}{C_P^2} > 0$, where κ is the constant of (1.22) (only depending on Λ, β, Ω, d, Γ^D, and $\theta_\mathcal{D}$), and C_P is either equal to C_{PW} if $|\Gamma^D| = 0$ or to C_{P,Γ^D} otherwise (where C_{PW}, C_{P,Γ^D} are the Poincaré constants of Proposition 7), such that

$$\forall n \in \mathbb{N}, \qquad \mathbb{E}^{n+1} \leq (1+\nu\Delta t)^{-1}\,\mathbb{E}^n. \tag{1.64}$$

Consequently, the discrete solution converges exponentially fast in time towards its associated discrete steady-state: for all $n \in \mathbb{N}^\star$,

$$\left\|u^n_\mathcal{M} - u^\infty_\mathcal{M}\right\|_{L^2(\Omega)} \leq (1+\nu\Delta t)^{-\frac{n}{2}}\left\|u^0_\mathcal{M} - u^\infty_\mathcal{M}\right\|_{L^2(\Omega)}. \tag{1.65}$$

Proof. Let $n \in \mathbb{N}$. One has

$$\mathbb{E}^{n+1} - \mathbb{E}^n = \sum_{K\in\mathcal{M}} \frac{|K|}{2}\Big(\big(u^{n+1}_K - u^\infty_K\big)^2 - \big(u^n_K - u^\infty_K\big)^2\Big).$$

Since $x \mapsto x^2$ is convex, for all $x, y \in \mathbb{R}$, we have $y^2 - x^2 \leq 2y(y-x)$, therefore

$$\mathbb{E}^{n+1} - \mathbb{E}^n \leq \sum_{K\in\mathcal{M}} |K|\Big(u^{n+1}_K - u^n_K\Big)\Big(u^{n+1}_K - u^\infty_K\Big) = \Big(u^{n+1}_\mathcal{M} - u^n_\mathcal{M}, u^{n+1}_\mathcal{M} - u^\infty_\mathcal{M}\Big)_\Omega. \tag{1.66}$$

Now, testing (1.26a) with $\underline{v}_{\mathcal{D}}^{n+1} = \underline{u}_{\mathcal{D}}^{n+1} - \underline{u}_{\mathcal{D}}^{\infty} \in \underline{V}_{\mathcal{D},0}$ yields

$$\left(u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^{n}, v_{\mathcal{M}}^{n+1}\right)_{\Omega} = -\Delta t a_{\mathcal{D}}\left(\underline{u}_{\mathcal{D}}^{n+1}, \underline{v}_{\mathcal{D}}^{n+1}\right) + \Delta t\left((f, v_{\mathcal{M}}^{n+1})_{\Omega} + (g^N, v_{\mathcal{E}}^{n+1})_{\Gamma^N}\right).$$

By definition (1.20) of the discrete steady-state $\underline{u}_{\mathcal{D}}^{\infty}$, we also have

$$(f, v_{\mathcal{M}}^{n+1})_{\Omega} + (g^N, v_{\mathcal{E}}^{n+1})_{\Gamma^N} = a_{\mathcal{D}}\left(\underline{u}_{\mathcal{D}}^{\infty}, \underline{v}_{\mathcal{D}}^{n+1}\right),$$

whence, by bilinearity of $a_{\mathcal{D}}$, we infer

$$\left(u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^{n}, v_{\mathcal{M}}^{n+1}\right)_{\Omega} = -\Delta t\, a_{\mathcal{D}}\left(\underline{u}_{\mathcal{D}}^{n+1} - \underline{u}_{\mathcal{D}}^{\infty}, \underline{v}_{\mathcal{D}}^{n+1}\right) = -\Delta t\, \mathbb{D}^{n+1}.$$

Combined to (1.66), this proves the entropy/dissipation relation (1.63). Now, since the advection field V^{ϕ} satisfies (1.21), we can invoke (1.22) from Proposition 2 to infer that

$$\mathbb{D}^{n+1} \geq \kappa |\underline{u}_{\mathcal{D}}^{n+1} - \underline{u}_{\mathcal{D}}^{\infty}|_{1,\mathcal{D}}^2,$$

where $\kappa > 0$ only depends on Λ ,β, Ω, d, Γ^D, and $\theta_{\mathcal{D}}$. Combining this estimate with a discrete Poincaré inequality from Proposition 7 (applied to $\underline{u}_{\mathcal{D}}^{n+1} - \underline{u}_{\mathcal{D}}^{\infty} \in \underline{V}_{\mathcal{D},0}$), and with the definition of the discrete (relative) entropy, yields

$$\mathbb{D}^{n+1} \geq \frac{\kappa}{C_P^2} \left\| u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^{\infty} \right\|_{L^2(\Omega)}^2 = \frac{2\kappa}{C_P^2} \mathbb{E}^{n+1},$$

where C_P is either equal to C_{PW} if $|\Gamma^D| = 0$ or to C_{P,Γ^D} otherwise. This last inequality, combined with the entropy/dissipation relation (1.63), implies the entropy decay (1.64). The inequality (1.65) is then a straightforward consequence of the definition of $\mathbb{E}^n$. □

The result of Theorem 2 does not use the fact that V^{ϕ} is related to the gradient of a potential, it thus extends to general advection fields.

1.4.2 Asymptotic behaviour of the exponential fitting scheme

We recall that $\rho = \frac{u}{\omega}$, with $\omega = \mathrm{e}^{-\phi}$, is the solution to Problem (1.34), and that ρ^{∞} is the corresponding steady-state, solution to Problem (1.28). We consider the following ω-weighted definitions of the relative entropy and dissipation:

$$\mathbb{E}_{\omega}(t) = \frac{1}{2} \left\| \rho(t) - \rho^{\infty} \right\|_{L^2_{\omega}(\Omega)}^2, \qquad \mathbb{D}_{\omega}(t) = \int_{\Omega} \omega \Lambda \nabla(\rho(t) - \rho^{\infty}) \cdot \nabla(\rho(t) - \rho^{\infty}).$$

It can be easily verified that the following entropy/dissipation relation holds at the continuous level:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}_{\omega}(t) + \mathbb{D}_{\omega}(t) = 0.$$

At the discrete level, let us recall that $\left(\underline{\rho}_{\mathcal{D}}^{n} \in \underline{V}_{\mathcal{D},0}^{\omega}\right)_{n \geq 1}$ is the solution to Problem (1.35). We then set $\underline{u}_{\mathcal{D}}^{n} = \underline{\omega}_{\mathcal{D}} \times \underline{\rho}_{\mathcal{D}}^{n}$ with $\underline{\omega}_{\mathcal{D}}$ defined by (1.31), in such a way that $\underline{u}_{\mathcal{D}}^{n} \in \underline{V}_{\mathcal{D},0}$. Similarly, $\underline{\rho}_{\mathcal{D}}^{\infty} \in \underline{V}_{\mathcal{D},0}^{\omega}$ is the corresponding steady-state, solution to Problem (1.30), and we set $\underline{u}_{\mathcal{D}}^{\infty} = \underline{\omega}_{\mathcal{D}} \times \underline{\rho}_{\mathcal{D}}^{\infty} \in \underline{V}_{\mathcal{D},0}$. We consider the following equivalents of the ω-weighted (relative) entropy and

dissipation: for all $n \in \mathbb{N}^\star$,

$$\mathbb{E}_\omega^n = \frac{1}{2}\left\|\rho_{\mathcal{M}}^n - \rho_{\mathcal{M}}^\infty\right\|_{L^2_\omega(\Omega)}^2, \qquad \mathbb{D}_\omega^n = a_{\mathcal{D}}^\omega\Big(\underline{\rho}_{\mathcal{D}}^n - \underline{\rho}_{\mathcal{D}}^\infty, \underline{\rho}_{\mathcal{D}}^n - \underline{\rho}_{\mathcal{D}}^\infty\Big),$$

where the discrete bilinear form $a_{\mathcal{D}}^\omega$ is defined (locally) by (1.29). The definition of the relative entropy is seamlessly extended to the case $n = 0$. Our main result on the exponential fitting HFV scheme is the following, whose proof is very similar to the one of Theorem 2.

Theorem 3 (Asymptotic stability)**.** *The following discrete entropy/dissipation relation holds true:*

$$\forall n \in \mathbb{N}, \qquad \frac{\mathbb{E}_\omega^{n+1} - \mathbb{E}_\omega^n}{\Delta t} + \mathbb{D}_\omega^{n+1} \le 0. \tag{1.67}$$

Furthermore, the discrete entropy decays exponentially fast in time: there is $\nu_\omega = \frac{2\omega_\flat \lambda_\flat \alpha_\flat}{\omega_\sharp C_P^2} > 0$*, where* $\lambda_\flat \alpha_\flat$ *is the coercivity constant of* (1.14) *(only depending on* Λ*,* Ω*,* d*, and* $\theta_{\mathcal{D}}$*), (we recall that) the bounds* $\omega_\flat, \omega_\sharp$ *only depend on* ϕ *and* Ω*, and* C_P *is either equal to* $2C_{PW}$ *if* $|\Gamma^D| = 0$ *or to* C_{P,Γ^D} *otherwise (where* C_{PW}*,* C_{P,Γ^D} *are the Poincaré constants of Proposition 7), such that*

$$\forall n \in \mathbb{N}, \qquad \mathbb{E}_\omega^{n+1} \le (1 + \nu_\omega \Delta t)^{-1}\, \mathbb{E}_\omega^n. \tag{1.68}$$

Consequently, the discrete solution converges exponentially fast in time towards its associated discrete steady-state: for all $n \in \mathbb{N}^\star$*,*

$$\|u_{\mathcal{M}}^n - u_{\mathcal{M}}^\infty\|_{L^2(\Omega)} \le \sqrt{\frac{\omega_\sharp}{\omega_\flat}}\,(1 + \nu_\omega \Delta t)^{-\frac{n}{2}}\, \|u_{\mathcal{M}}^0 - u_{\mathcal{M}}^\infty\|_{L^2(\Omega)}. \tag{1.69}$$

Proof. Let $n \in \mathbb{N}$. Reasoning as in the proof of Theorem 2, we infer

$$\mathbb{E}_\omega^{n+1} - \mathbb{E}_\omega^n \le -\Delta t\, a_{\mathcal{D}}^\omega\Big(\underline{\rho}_{\mathcal{D}}^{n+1} - \underline{\rho}_{\mathcal{D}}^\infty, \underline{\rho}_{\mathcal{D}}^{n+1} - \underline{\rho}_{\mathcal{D}}^\infty\Big) = -\Delta t\, \mathbb{D}_\omega^{n+1},$$

which proves (1.67). By the coercivity estimate (1.33), we have that $\mathbb{D}_\omega^{n+1} \ge \omega_\flat \lambda_\flat \alpha_\flat \big|\underline{\rho}_{\mathcal{D}}^{n+1} - \underline{\rho}_{\mathcal{D}}^\infty\big|_{1,\mathcal{D}}^2$, where $\alpha_\flat > 0$ from Section 1.2.3 only depends on Ω, d, and $\theta_{\mathcal{D}}$, and $\omega_\flat > 0$ only depends on ϕ and Ω. Reasoning as in the proof of Proposition 3, and using a discrete Poincaré inequality from Proposition 7 (combined with [51, Lemma 5.2] in the case $|\Gamma^D| = 0$), we also infer that

$$\|\rho_{\mathcal{M}}^{n+1} - \rho_{\mathcal{M}}^\infty\|_{L^2_\omega(\Omega)} \le \sqrt{\omega_\sharp}\, C_P |\underline{\rho}_{\mathcal{D}}^{n+1} - \underline{\rho}_{\mathcal{D}}^\infty|_{1,\mathcal{D}},$$

where $\omega_\sharp > 0$ only depends on ϕ and Ω, and C_P is either equal to $2C_{PW}$ if $|\Gamma^D| = 0$ or to C_{P,Γ^D} otherwise. Thus, we finally get that

$$\mathbb{D}_\omega^{n+1} \ge \frac{2\omega_\flat \lambda_\flat \alpha_\flat}{\omega_\sharp C_P^2}\, \mathbb{E}_\omega^{n+1}.$$

Combined to (1.67), this yields (1.68). Deriving (1.69) is then straightforward. □

1.4.3 Asymptotic behaviour of the nonlinear scheme

Recall that $u > 0$ is the solution to Problem (1.1) endowed with pure Neumann boundary conditions ($|\Gamma^D| = 0$), and data $f = 0$, $g^N = 0$, and $u^{in} \ge 0$ with $M = \int_\Omega u^{in} > 0$ (see Appendix 1.B

for the case of mixed Dirichlet-Neumann boundary conditions), and that $u^\infty > 0$, solution to Problem (1.2) with same data, is the thermal equilibrium u^∞_{th} given by (1.25) with $\bar\rho = \rho^M$. The analysis of the nonlinear scheme relies on the entropy/dissipation structure (1.36)-(1.37) introduced in Section 1.3.3. Notice that the relative dissipation (or relative Fisher information) of (1.37) can be equivalently rewritten

$$\mathbb{D}(t) = \int_\Omega u(t)\Lambda\nabla\log\left(\frac{u(t)}{u^\infty}\right)\cdot\nabla\log\left(\frac{u(t)}{u^\infty}\right) = 4\int_\Omega u^\infty\Lambda\nabla\sqrt{\frac{u(t)}{u^\infty}}\cdot\nabla\sqrt{\frac{u(t)}{u^\infty}}.$$

At the discrete level, recalling that $\left(\underline{u}^n_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}\right)_{n\geq 1}$ is a (positive) solution to Problem (1.45), and that $\underline{u}^\infty_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ is the corresponding steady-state, solution to Problem (1.46), that is equal to the HFV interpolate of u^∞_{th}, we consider the discrete entropy $\mathbb{E}^n$ and dissipation $\mathbb{D}^n$ defined by (1.49), and we define a discrete counterpart of the relative dissipation written in root-form: for all $n \geq 1$,

$$\hat{\mathbb{D}}^n = 4\sum_{K\in\mathcal{M}} u^\infty_{K,\flat}\int_K \Lambda\nabla_K\underline{\xi}^n_K\cdot\nabla_K\underline{\xi}^n_K = 4\sum_{K\in\mathcal{M}} u^\infty_{K,\flat}\delta_K\underline{\xi}^n_K\cdot\mathbb{A}_K\delta_K\underline{\xi}^n_K, \tag{1.70}$$

where, for all $K\in\mathcal{M}$, $u^\infty_{K,\flat} = \min\left(u^\infty_K, \min_{\sigma\in\mathcal{E}_K} u^\infty_\sigma\right)$ and the matrix $\mathbb{A}_K$ is defined by (1.10), and $\underline{\xi}^n_{\mathcal{D}}$ is the element of $\underline{V}_{\mathcal{D}}$ such that

$$\xi^n_K = \sqrt{\frac{u^n_K}{u^\infty_K}} \quad \forall K\in\mathcal{M}, \qquad \xi^n_\sigma = \sqrt{\frac{u^n_\sigma}{u^\infty_\sigma}} \quad \forall\sigma\in\mathcal{E}. \tag{1.71}$$

At the discrete level, and as opposed to the continuous level, the quantities $\mathbb{D}^n$ and $\hat{\mathbb{D}}^n$ are not equal, therefore we need to compare them. The definition of $u^\infty_{K,\flat}$ results from the following observation: according to the structures of $\mathbb{D}^n$ and $\hat{\mathbb{D}}^n$, locally, we expect to have to compare $u^\infty_{K,\flat}$ with $r_K(\underline{u}^n_K)$, which depends on u^n_K and on the $(u^n_\sigma)_{\sigma\in\mathcal{E}_K}$.

Proposition 6 (Fisher information). *There is $C_F > 0$, only depending on Λ, Ω, d, and $\theta_{\mathcal{D}}$ such that*

$$\forall n\geq 1, \qquad \hat{\mathbb{D}}^n \leq C_F\mathbb{D}^n. \tag{1.72}$$

Proof. Let $n\in\mathbb{N}^\star$, and $K\in\mathcal{M}$. By (1.96) from Lemma 4, we first have that

$$4u^\infty_{K,\flat}\delta_K\underline{\xi}^n_K\cdot\mathbb{A}_K\delta_K\underline{\xi}^n_K \leq 4u^\infty_{K,\flat}\delta_K\underline{\xi}^n_K\cdot\mathbb{B}_K\delta_K\underline{\xi}^n_K, \tag{1.73}$$

where the matrix $\mathbb{B}_K$ is the diagonal matrix defined by (1.95). Since for all $(x,y)\in\left(\mathbb{R}^\star_+\right)^2$, $\int_x^y \frac{\mathrm{d}z}{\sqrt{z}} = 2\left(\sqrt{y}-\sqrt{x}\right)$, the Cauchy–Schwarz inequality yields

$$4\left(\sqrt{y}-\sqrt{x}\right)^2 \leq (y-x)\left(\log(y)-\log(x)\right).$$

By the property (1.42d) of the function m, we then get that

$$\forall(x,y)\in\left(\mathbb{R}^\star_+\right)^2, \qquad 4\left(\sqrt{y}-\sqrt{x}\right)^2 \leq m(x,y)\left(\log(y)-\log(x)\right)^2. \tag{1.74}$$

Since $\mathbb{B}_K$ is diagonal, the combination of (1.73), (1.71), and (1.74), yields

$$4u_{K,\flat}^\infty \delta_K \underline{\xi}_K^n \cdot \mathbb{A}_K \delta_K \underline{\xi}_K^n \le \sum_{\sigma\in\mathcal{E}_K} 4u_{K,\flat}^\infty B_K^{\sigma\sigma} \left(\sqrt{\frac{u_K^n}{u_K^\infty}} - \sqrt{\frac{u_\sigma^n}{u_\sigma^\infty}}\right)^2$$
$$\le \sum_{\sigma\in\mathcal{E}_K} u_{K,\flat}^\infty B_K^{\sigma\sigma} m\left(\frac{u_K^n}{u_K^\infty}, \frac{u_\sigma^n}{u_\sigma^\infty}\right)\left(\log\left(\frac{u_K^n}{u_K^\infty}\right) - \log\left(\frac{u_\sigma^n}{u_\sigma^\infty}\right)\right)^2.$$

By definition of $u_{K,\flat}^\infty$, and monotonicity (1.42a) and homogeneity (1.42c) of m, we infer that, for all $\sigma \in \mathcal{E}_K$,

$$u_{K,\flat}^\infty m\left(\frac{u_K^n}{u_K^\infty}, \frac{u_\sigma^n}{u_\sigma^\infty}\right) \le u_{K,\flat}^\infty m\left(\frac{u_K^n}{u_{K,\flat}^\infty}, \frac{u_\sigma^n}{u_{K,\flat}^\infty}\right) = m\left(u_K^n, u_\sigma^n\right).$$

By definition (1.43) of $f_{|\mathcal{E}_K|}$, and the bound (1.4) on $|\mathcal{E}_K|$, we then have

$$\max_{\sigma\in\mathcal{E}_K} u_{K,\flat}^\infty m\left(\frac{u_K^n}{u_K^\infty}, \frac{u_\sigma^n}{u_\sigma^\infty}\right) \le \max_{\sigma\in\mathcal{E}_K} m\left(u_K^n, u_\sigma^n\right) \le |\mathcal{E}_K| r_K(\underline{u}_K^n) \le d\theta_\mathcal{D}^2\, r_K(\underline{u}_K^n).$$

We deduce that

$$4u_{K,\flat}^\infty \delta_K \underline{\xi}_K^n \cdot \mathbb{A}_K \delta_K \underline{\xi}_K^n \le d\theta_\mathcal{D}^2\, r_K(\underline{u}_K^n) \sum_{\sigma\in\mathcal{E}_K} B_K^{\sigma\sigma} \left(\log\left(\frac{u_K^n}{u_K^\infty}\right) - \log\left(\frac{u_\sigma^n}{u_\sigma^\infty}\right)\right)^2$$
$$= d\theta_\mathcal{D}^2\, r_K(\underline{u}_K^n)\, \delta_K \underline{w}_K^n \cdot \mathbb{B}_K \delta_K \underline{w}_K^n,$$

where $\underline{w}_K^n \in \underline{V}_K$ is such that $\underline{u}_K^n = \underline{u}_K^\infty \times \exp(\underline{w}_K^n)$. Using again (1.96) from Lemma 4, we finally infer that

$$4u_{K,\flat}^\infty \delta_K \underline{\xi}_K^n \cdot \mathbb{A}_K \delta_K \underline{\xi}_K^n \le d\theta_\mathcal{D}^2 C_B\, r_K(\underline{u}_K^n)\, \delta_K \underline{w}_K^n \cdot \mathbb{A}_K \delta_K \underline{w}_K^n,$$

with $C_B > 0$ only depending on Λ, Ω, d, and $\theta_\mathcal{D}$. Summing over $K \in \mathcal{M}$, and recalling the definitions (1.70) of $\hat{\mathbb{D}}^n$, and (1.49) of $\mathbb{D}^n$, eventually yields (1.72) with $C_F = d\theta_\mathcal{D}^2 C_B$. □

The long-time behaviour of the nonlinear HFV scheme is studied in the following result.

Theorem 4 (Asymptotic stability). *Recall the discrete entropy/dissipation relation of Proposition 5. The discrete entropy decays exponentially fast in time: there is* $\nu_{\mathrm{nl}} = \frac{4u_\flat^\infty \lambda_\flat \alpha_\flat}{C_F C_{LS,\infty}^2} > 0$, *depending on* Λ, ϕ, M, Ω, d, *and* $\theta_\mathcal{D}$ *such that*

$$\forall n \in \mathbb{N}, \qquad \mathbb{E}^{n+1} \le (1 + \nu_{\mathrm{nl}} \Delta t)^{-1} \mathbb{E}^n. \tag{1.75}$$

Consequently, the discrete solution converges exponentially fast in time towards its associated discrete steady-state: for all $n \in \mathbb{N}^\star$,

$$\|u_\mathcal{M}^n - u_\mathcal{M}^\infty\|_{L^1(\Omega)} \le \sqrt{2M\mathbb{E}^0}\,(1 + \nu_{\mathrm{nl}}\Delta t)^{-\frac{n}{2}}. \tag{1.76}$$

Proof. Let $n \in \mathbb{N}$. By definition (1.70) of $\hat{\mathbb{D}}^n$, and from the coercivity estimate (1.14), we first infer that

$$\hat{\mathbb{D}}^{n+1} = 4\sum_{K\in\mathcal{M}} u_{K,\flat}^\infty a_K^\Lambda\left(\underline{\xi}_K^{n+1}, \underline{\xi}_K^{n+1}\right) \ge 4u_\flat^\infty a_\mathcal{D}^\Lambda\left(\underline{\xi}_\mathcal{D}^{n+1}, \underline{\xi}_\mathcal{D}^{n+1}\right) \ge 4u_\flat^\infty \lambda_\flat \alpha_\flat \left|\underline{\xi}_\mathcal{D}^{n+1}\right|_{1,\mathcal{D}}^2,$$

which, combined with (1.72), implies that

$$\mathbb{D}^{n+1} \geq \frac{1}{C_F}\hat{\mathbb{D}}^{n+1} \geq \frac{4u_\flat^\infty \lambda_\flat \alpha_\flat}{C_F}\left|\underline{\xi}_{\mathcal{D}}^{n+1}\right|_{1,\mathcal{D}}^2.$$

In order to compare this quantity with the entropy, we use the discrete log-Sobolev inequality (1.86) from Proposition 9, applied to the couple $(\underline{u}_{\mathcal{D}}^{n+1}, \underline{u}_{\mathcal{D}}^{\infty})$ (which satisfies the mass condition owing to Proposition 4). We get

$$\mathbb{E}^{n+1} = \int_\Omega u_{\mathcal{M}}^\infty \Phi_1\left(\frac{u_{\mathcal{M}}^{n+1}}{u_{\mathcal{M}}^\infty}\right) \leq C_{LS,\infty}^2 \left|\underline{\xi}_{\mathcal{D}}^{n+1}\right|_{1,\mathcal{D}}^2,$$

which, combined with the previous estimate, yields

$$\mathbb{D}^{n+1} \geq \frac{4u_\flat^\infty \lambda_\flat \alpha_\flat}{C_F C_{LS,\infty}^2}\,\mathbb{E}^{n+1}.$$

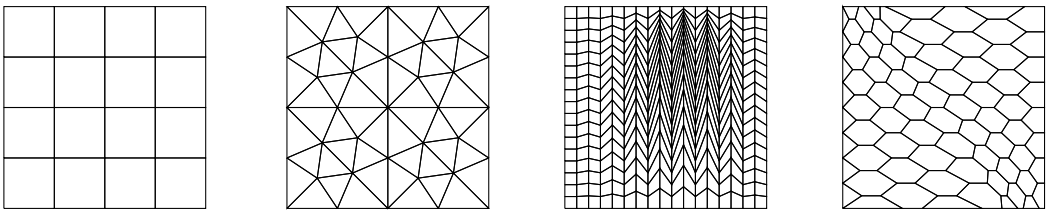
Combined with (1.50) from Proposition 5, this shows (1.75). The L^1-norm estimate (1.76) is then a direct consequence of (1.75) and of the Csiszár–Kullback lemma (cf., e.g., [51, Lemma 5.6]) applied to the probability measure $\mu(x)\,\mathrm{d}x = u_{\mathcal{M}}^\infty(x)\,\frac{\mathrm{d}x}{M}$ and to the function $g = \frac{u_{\mathcal{M}}^n}{u_{\mathcal{M}}^\infty}$ such that $\int_\Omega g\,\mathrm{d}\mu = 1$, which yields $\|u_{\mathcal{M}}^n - u_{\mathcal{M}}^\infty\|_{L^1(\Omega)} \leq \sqrt{2M\mathbb{E}^n}$ for all $n \geq 1$. □

Remark 9 (Norms and long-time behaviour)**.** *Notice that Theorem 4 states an exponential decay in L^1-norm, whereas Theorems 2 and 3 assert a convergence in L^2, and thus a convergence in L^p for any $p \in [1,2]$. This is reminiscent of the fact that the natural topologies for the linear and nonlinear problems differ.*

1.5 Numerical results

1.5.1 Implementation

In this section, we discuss some practical aspects concerning the implementation of the schemes described in this paper. In all the test-cases presented below, the two-dimensional domain is taken to be $\Omega = (0,1)^2$. The meshes used for the numerical tests, presented on Figure 1.2, are the classical Cartesian, triangular, and Kershaw meshes from the FVCA V benchmark (see [154]), as well as a tilted hexagonal-dominant mesh (cf. [101]). These meshes have convex



(a) Cartesian mesh (b) Triangular mesh (c) Kershaw mesh (d) Tilted hexagonal mesh

Figure 1.2 – **Implementation.** Coarsest meshes of each family used in the numerical tests.

cells, hence we always choose x_K to be the barycentre of $K \in \mathcal{M}$. In our implementation,

we compute the meshsize as $\tilde{h}_{\mathcal{D}} = \max_{K\in\mathcal{M}} \frac{|K|}{|\partial K|}$. Observe that $\tilde{h}_{\mathcal{D}}/h_{\mathcal{D}}$ is framed by constants only depending on the mesh regularity. Notice also that the Kershaw mesh family is not uniformly regular in the sense defined in Section 1.2.1. In practice, we use a fixed value $\sqrt{2} < \eta = 1.5 < 2$ of the stabilisation parameter (see Remark 1). In the sequel, we denote by HMM the classical HFV linear scheme for advection-diffusion, and we restrict our attention to the Scharfetter–Gummel discretisation of the flux (1.18), namely to the function $A(s) = \frac{s}{\mathrm{e}^s - 1} - 1$, extended by continuity to 0 at $s = 0$.

Linear systems and static condensation

The two linear (HMM and exponential fitting) schemes are implemented in the same way. To fix ideas, we consider the evolution problem with pure Neumann boundary conditions, of unknown solution $\underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$. We denote by $U_{\mathcal{M}} \in \mathbb{R}^{|\mathcal{M}|}$ and $U_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}$ the unknown vectors $(u_K)_{K\in\mathcal{M}}$ and $(u_\sigma)_{\sigma\in\mathcal{E}}$. The linear schemes result in the following block system:

$$\begin{pmatrix} \mathbb{M}_{\mathcal{M}} & \mathbb{M}_{\mathcal{M},\mathcal{E}} \\ \mathbb{M}_{\mathcal{E},\mathcal{M}} & \mathbb{M}_{\mathcal{E}} \end{pmatrix} \begin{pmatrix} U_{\mathcal{M}} \\ U_{\mathcal{E}} \end{pmatrix} = \begin{pmatrix} S_{\mathcal{M}} \\ S_{\mathcal{E}} \end{pmatrix}, \tag{1.77}$$

where $\mathbb{M}_{\mathcal{M}} \in \mathbb{R}^{|\mathcal{M}|\times|\mathcal{M}|}$, $\mathbb{M}_{\mathcal{M},\mathcal{E}} \in \mathbb{R}^{|\mathcal{M}|\times|\mathcal{E}|}$, $\mathbb{M}_{\mathcal{E},\mathcal{M}} \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{M}|}$, $\mathbb{M}_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{E}|}$, and $S_{\mathcal{M}} \in \mathbb{R}^{|\mathcal{M}|}$ and $S_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}$ stem from the loading term and the boundary conditions. By construction, the matrix $\mathbb{M}_{\mathcal{M}}$ is diagonal with non-zero diagonal entries, and can therefore be inverted at a very low computational cost. Thus, one can eliminate the cell unknowns, noticing that

$$U_{\mathcal{M}} = \mathbb{M}_{\mathcal{M}}^{-1}\left(S_{\mathcal{M}} - \mathbb{M}_{\mathcal{M},\mathcal{E}} U_{\mathcal{E}}\right). \tag{1.78}$$

Using this relation, one infers that $U_{\mathcal{E}}$ is the solution to the following linear system:

$$\left(\mathbb{M}_{\mathcal{E}} - \mathbb{M}_{\mathcal{E},\mathcal{M}}\mathbb{M}_{\mathcal{M}}^{-1}\mathbb{M}_{\mathcal{M},\mathcal{E}}\right) U_{\mathcal{E}} = S_{\mathcal{E}} - \mathbb{M}_{\mathcal{E},\mathcal{M}}\mathbb{M}_{\mathcal{M}}^{-1} S_{\mathcal{M}}, \tag{1.79}$$

where $\mathbb{M}_{\mathcal{D}} = \mathbb{M}_{\mathcal{E}} - \mathbb{M}_{\mathcal{E},\mathcal{M}}\mathbb{M}_{\mathcal{M}}^{-1}\mathbb{M}_{\mathcal{M},\mathcal{E}}$, the so-called Schur complement of the matrix $\mathbb{M}_{\mathcal{M}}$, is an invertible matrix of size $|\mathcal{E}| \times |\mathcal{E}|$. In practice, we solve the linear system (1.79) using an LU factorisation algorithm, and we use the solution $U_{\mathcal{E}}$ to reconstruct $U_{\mathcal{M}}$ from (1.78). This method, called *static condensation*, allows one to replace a system of size $|\mathcal{M}| + |\mathcal{E}|$ by a system of size $|\mathcal{E}|$ without additional fill-in. In the case of mixed Dirichlet-Neumann boundary conditions, the Dirichlet face unknowns are eliminated from the global linear system.

Exponential fitting scheme: choice of unknown and harmonic averaging

The exponential fitting scheme can be expressed in either the u or the $\rho = u\,\mathrm{e}^{\phi}$ variable. In the ρ variable, the resulting linear system is symmetric. One can then use, e.g., Cholesky factorisation or a conjugate gradient method. However, the formulation in ρ is ill-conditioned. In our numerical experiments, the ratio between the condition numbers of the linear systems in ρ and in u often exceeds 10^5. Because of this, we chose and we recommend to solve the linear system in the unknown u. Notice that solving the system in u is equivalent to right pre-condition the system in ρ with the inverse of the diagonal matrix with entries the coordinates of the interpolate of $\omega = \mathrm{e}^{-\phi}$.

In order to implement the exponential fitting scheme, one needs to evaluate averages of the diffusion tensor $\frac{1}{|P_{K,\sigma}|}\int_{P_{K,\sigma}} \omega\Lambda$. Observe that $\omega(x)/\omega(x_K)$ is of order $\mathrm{e}^{h_K\|\nabla\phi\|_{L^\infty(P_{K,\sigma})}}$ in $P_{K,\sigma}$. Therefore, with large advection fields, the diffusion problem (1.28) becomes strongly heterogeneous. It

is pointed out in [43] that an (empirical) solution to improve robustness to this heterogeneity is to use harmonic averages to approximate integrals of the diffusion tensor. In the numerical tests of the following subsections, we compare the "classical" exponential fitting scheme (for which the integral is approximated by a standard - second order - quadrature) with the "harmonic" one, in which case we choose to use the following approximation:

$$\frac{1}{|P_{K,\sigma}|}\int_{P_{K,\sigma}} \omega\Lambda \approx 3\left(\sum_{F\in\mathcal{E}_{P_{K,\sigma}}} \frac{1}{\omega(\overline{x}_F)}\right)^{-1}\Lambda(x_K),$$

where $\mathcal{E}_{P_{K,\sigma}}$ denotes the set of edges of the triangle $P_{K,\sigma}$ (recall that $d=2$ in our experiments), and $\overline{x}_F$ is the barycentre of $F\in\mathcal{E}_{P_{K,\sigma}}$.

Nonlinear scheme and Newton's method

The implementation of the nonlinear scheme relies on the following formulation: given $\underline{u}_{\mathcal{D}}^{n-1}\in\underline{V}_{\mathcal{D}}$ positive, we want to solve the nonlinear system $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}(\underline{u}_{\mathcal{D}}^{n})=\underline{0}_{\mathcal{D}}$, where $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}$ is defined as in (1.55) but with a time step δt instead of Δt. The resolution of this system relies on Newton's method.

First, one initialises the method with $\delta t=\Delta t$, and initial guess $\max(\underline{u}_{\mathcal{D}}^{n-1},\epsilon\underline{1}_{\mathcal{D}})\in\underline{V}_{\mathcal{D}}$ (where the maximum is taken coordinate by coordinate), in order to avoid potential problems due to the singularity of the log near 0. The successive linear systems to compute the residue have the same structure as (1.77). We thus perform static condensation at each Newton iteration. As a stopping criterion, we compare the l^∞ relative norm of the residue with a threshold tol. If the method does not converge after i_{max} iterations, we divide the time step by 2, and we restart the resolution. When the method converges, one can proceed with the approximation of $\underline{u}_{\mathcal{D}}^{n+1}$, with an initial time step of $\min(\Delta t, 2\delta t)$. In practice, we use $\epsilon=10^{-11}$, $i_{max}=50$, and $tol=10^{-11}$.

The implementation of the nonlinear scheme relies on the computation of $\log(\omega_K)$ and $\log(\omega_\sigma)$. Since we have chosen x_K to be the barycentre of K, we choose to approximate $\frac{1}{|K|}\int_K \mathrm{e}^{-\phi}$ by $\mathrm{e}^{-\phi(x_K)}$. Therefore, $\log(\omega_K)$ is computed as $\log(\mathrm{e}^{-\phi(x_K)})=-\phi(x_K)$. The same holds true for $\log(\omega_\sigma)$.

In the simulations shown below, we use arithmetic means for the functions m and $f_{|\mathcal{E}_K|}$ of the reconstruction r_K defined by (1.41)-(1.43). For all $K\in\mathcal{M}$, and all $\underline{u}_K\in\underline{V}_K$, we thus consider

$$r_K(\underline{u}_K)=\frac{1}{2}\left(u_K+\frac{1}{|\mathcal{E}_K|}\sum_{\sigma\in\mathcal{E}_K}u_\sigma\right).$$

This choice is close to the one advocated in [49, Eq. (58)]. For a discussion on other choices of reconstructions, we refer to [51, Section 6.2].

1.5.2 Long-time behaviour of discrete solutions

In this section, we present some numerical illustration of the long-time behaviour of discrete solutions. We focus on a test-case from [56, 52, 51]. We consider homogeneous pure Neumann boundary conditions ($\Gamma^D=\emptyset$ and $g^N=0$), and zero loading term ($f=0$). The advective potential and diffusion tensor are set to $\phi(x,y)=-x$ and $\Lambda=\begin{pmatrix}l_x & 0\\ 0 & 1\end{pmatrix}$ for $l_x>0$. The exact solution is given

by

$$u(t,x,y) = C_1\,\mathrm{e}^{-\alpha t+\frac{x}{2}}\,(2\pi\cos(\pi x)+\sin(\pi x)) + 2C_1\pi\,\mathrm{e}^{x-\frac{1}{2}},$$

where $C_1 > 0$ and $\alpha = l_x\left(\frac{1}{4}+\pi^2\right)$. Note that u^{in} vanishes on $\{x=1\}$, but for any $t>0$, $u(t,\cdot)>0$. The associated steady-state is

$$u^\infty(x,y) = 2C_1\pi\,\mathrm{e}^{x-\frac{1}{2}}.$$

Our experiments are performed using the following values:

$$l_x = 10^{-2} \qquad \text{and} \qquad C_1 = 10^{-1}.$$

We compute the solution on the time interval $[0,T_f]$, and we denote by $(\underline{u}_{\mathcal{D}}^n)_{1\le n\le N_f}$ the corresponding approximate solution. Note that the number of time steps N_f may differ between the linear and nonlinear schemes, because of the adaptive time step refinement procedure used for the nonlinear scheme.

We set $T_f = 350$ in order to see the complete evolution. Since the long-time behaviour of the schemes does not depend on the size of the discretisation, we can explore the evolution using a large time step $\Delta t = 10^{-1}$. We perform the numerical experiments on two Kershaw meshes (see Figure 1.2c) of sizes 0.02 and 0.006. In Figure 1.3, we depict, as a function of time, the L^1

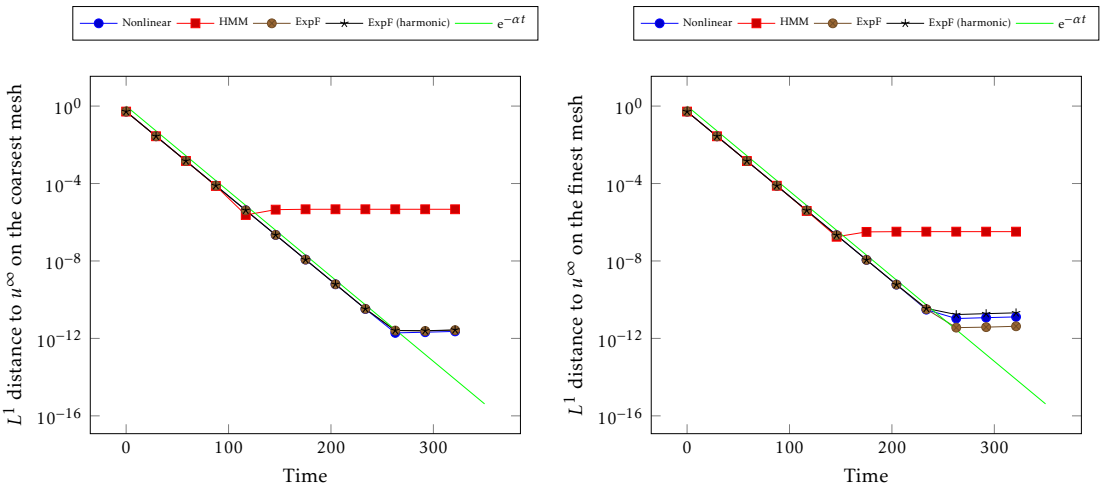


Figure 1.3 – **Long-time behaviour of discrete solutions.** Comparison of the long-time behaviour on Kershaw meshes for $T_f = 350$ and $\Delta t = 0.1$.

distance between $\underline{u}_{\mathcal{D}}^n$ and u^∞ (the exact steady-state) computed as

$$\sum_{K\in\mathcal{M}} |K|\|u_K^n - u^\infty(x_K)|.$$

We observe the exponential decay towards the steady-state, until some precision is reached. Note that for the HMM scheme, some saturation occurs at precision of magnitude 10^{-6} and 10^{-7}: the scheme does not preserve the thermal equilibrium (see Remark 4). This saturation corresponds to the accuracy of the stationary scheme (see Section 1.5.4), so the threshold is lower on the refined mesh. Note that one could also consider the error measure $\sum_{K\in\mathcal{M}} |K|\|u_K^n - u_K^\infty|$

between the discrete solution and the discrete steady-state: this quantity decays exponentially, with a lower saturation of magnitude 10^{-12}, corresponding to machine precision. The other (nonlinear and exponential fitting) schemes have the same decay rate, and the saturation occurs at machine precision. For the four schemes, the rates of convergence are similar to the real one α. In particular, the use of harmonic averages in the exponential fitting scheme does not have any impact on the long-time behaviour.

1.5.3 Positivity of discrete solutions

We are now interested in the positivity of the discrete solutions. We use the following test-case with anisotropic diffusion and homogeneous pure Neumann boundary conditions. We set $\Gamma^D = \emptyset$, $f = 0$, $g^N = 0$,

$$\phi(x,y) = -\Big((x-0.4)^2 + (y-0.6)^2\Big) \qquad \text{and} \qquad \Lambda = \begin{pmatrix} 0.8 & 0 \\ 0 & 1 \end{pmatrix}.$$

For the initial datum, we take

$$u^{in} = 10^{-3}\,\mathbb{1}_B + \mathbb{1}_{\Omega\setminus B},$$

where B is the Euclidean ball $\Big\{(x,y) \in \mathbb{R}^2 \mid (x-0.5)^2 + (y-0.5)^2 \leq 0.2^2\Big\}$. These data ensure that the solution u is positive on $\mathbb{R}_+ \times \Omega$. The experiment is performed on a tilted hexagonal-dominant mesh (see Figure 1.2d) of size $4.3 \cdot 10^{-3}$, made up of 4192 cells and 12512 edges. Since we deal with a diffusive phenomenon, the smallest values of u are expected for small time, hence we perform the simulation with a relatively small final time $T_f = 5.10^{-4}$, alongside with a time step of $\Delta t = 10^{-5}$.

The results are collected in Table 1.1. The cost is defined as the number of linear systems solved in order to compute the solution $(\underline{u}_{\mathcal{D}}^n)_{1\leq n\leq N_f}$, and the minimum values min_cells and min_edges are defined by

$$\min\Big\{u_K^n \mid 1 \leq n \leq N_f,\ K \in \mathcal{M}\Big\} \quad \text{and} \quad \min\Big\{u_\sigma^n \mid 1 \leq n \leq N_f,\ \sigma \in \mathcal{E}\Big\}.$$

The indicated number of negative unknowns is for the whole simulation. Here, $N_f = 50$ also for the nonlinear scheme (no sub-division of the time step was needed in Newton's method). As

	cost	min_cells	min_edges	# negative unknowns
Nonlinear	175	9.93e-04	7.36e-04	0
HMM	50	-5e-03	–7.74e-02	593
ExpF	50	-4.98e-03	-7.72e-02	590
ExpF (harmonic)	50	-4.98e-03	-7.74e-02	588

Table 1.1 – **Positivity of discrete solutions.** Numerical results for $T_f = 5.10^{-4}$ and $\Delta t = 10^{-5}$ on a tilted hexagonal-dominant mesh. At each time step, there are 4192 cell unknowns and 12512 edge unknowns.

expected, the nonlinear scheme has positive discrete solutions, whereas the linear ones exhibit a violation of positivity (the value of η can have some influence on positivity; see [122]). Note that the use of harmonic averages for the exponential fitting scheme has no impact on the undershoots.

We observe that the nonlinear scheme requires approximately 3.5 times more linear system inversions than the linear schemes. However, this value depends strongly on the final time

of simulation T_f. Indeed, the number of linear systems solved at step n decreases when n increases. The first time step costs 9 resolutions, but this number rapidly decreases as the solution approaches the steady-state (the second and the third time steps respectively cost 5 and 4 resolutions).

1.5.4 Accuracy of stationary solutions

In this section, we aim at comparing the accuracy of the different schemes for the stationary problem. To do so, we define the discrete L^2-norm and H^1-seminorm errors as (i) the L^2-norm of the difference $u_{\mathcal{M}} - \Pi_{\mathcal{M}}(u)$, and (ii) the $|\cdot|_{1,\mathcal{D}}$-seminorm of the difference $\underline{u}_{\mathcal{D}} - \underline{\Pi}_{\mathcal{D}}(u)$, where $\underline{u}_{\mathcal{D}}$ is the discrete solution, and $\underline{\Pi}_{\mathcal{D}}(u)$ is the HFV interpolate of the continuous solution u (computed as $\underline{\Pi}_{\mathcal{D}}(u) \approx ((u(x_K))_{K\in\mathcal{M}}, (u(\overline{x}_\sigma))_{\sigma\in\mathcal{E}})$). In what follows, we reason in relative errors.

The nonlinear scheme is extended to a more general setting, in order to take into account a loading term $f \geq 0$ and mixed Dirichlet-Neumann boundary conditions ($|\Gamma^D| > 0$) with $g^D > 0$ and $g^N \geq 0$. The scheme writes:

$$\text{Find } \underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}} \text{ positive such that } \underline{\mathcal{G}}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}) = \underline{0}_{\mathcal{D}}, \tag{1.80}$$

where $\underline{\mathcal{G}}_{\mathcal{D}} : (\underline{V}_{\mathcal{D}})_+^\star \to \underline{V}_{\mathcal{D}}$ is the vector field defined by

$$\mathcal{G}_K(\underline{u}_{\mathcal{D}}) = \sum_{\sigma\in\mathcal{E}_K} \mathcal{F}^{\mathrm{nl}}_{K,\sigma}(\underline{u}_K) - \int_K f \qquad \forall K \in \mathcal{M}, \tag{1.81a}$$

$$\mathcal{G}_\sigma(\underline{u}_{\mathcal{D}}) = -\left(\mathcal{F}^{\mathrm{nl}}_{K,\sigma}(\underline{u}_K) + \mathcal{F}^{\mathrm{nl}}_{L,\sigma}(\underline{u}_L)\right) \qquad \forall \sigma = K \mid L \in \mathcal{E}_{int}, \tag{1.81b}$$

$$\mathcal{G}_\sigma(\underline{u}_{\mathcal{D}}) = -\int_\sigma g^N - \mathcal{F}^{\mathrm{nl}}_{K,\sigma}(\underline{u}_K) \qquad \forall \sigma \in \mathcal{E}^N_{ext} \text{ with } \mathcal{M}_\sigma = \{K\}, \tag{1.81c}$$

$$\mathcal{G}_\sigma(\underline{u}_{\mathcal{D}}) = \frac{1}{|\sigma|}\int_\sigma g^D - u_\sigma \qquad \forall \sigma \in \mathcal{E}^D_{ext}, \tag{1.81d}$$

$$\mathcal{F}^{\mathrm{nl}}_{K,\sigma}(\underline{u}_K) = r^K(\underline{u}_K) \sum_{\sigma'\in\mathcal{E}_K} A_K^{\sigma\sigma'} (\log(u_K) + \phi(x_K) - \log(u_{\sigma'}) - \phi(\overline{x}_{\sigma'})) \qquad \forall K \in \mathcal{M}, \forall \sigma \in \mathcal{E}_K. \tag{1.81e}$$

The implementation of this scheme still relies on a Newton method similar to the one used for the evolution scheme. It is here initialised with $\underline{\Pi}_{\mathcal{D}}(\omega) = \underline{\omega}_{\mathcal{D}}$.

The first test-case we consider is the same as in [19]. It is an isotropic problem, with $\Lambda = I_2$, $\Gamma^N = \emptyset$, $\phi(x,y) = -(2x+3y)$, and exact solution

$$u(x,y) = \left(x - \mathrm{e}^{2(x-1)}\right)\left(y^2 - \mathrm{e}^{3(y-1)}\right),$$

the other data f and g^D being set accordingly. Note that for this test-case, the diffusion and advection terms are of the same order of magnitude. The numerical experiments are performed on the triangular mesh family (see Figure 1.2b). The convergence results are depicted in Figure 1.4. As expected, the two linear schemes are of order two in L^2-norm, and one in H^1-seminorm. The same holds for the nonlinear scheme, whose accuracy is rather the same as the classical HMM scheme, one order of magnitude better than the exponential fitting schemes. On this test-case, the use of harmonic averages for the exponential fitting scheme does not have a significant impact.

The second test-case is an advection-dominated problem, with anisotropic diffusion and

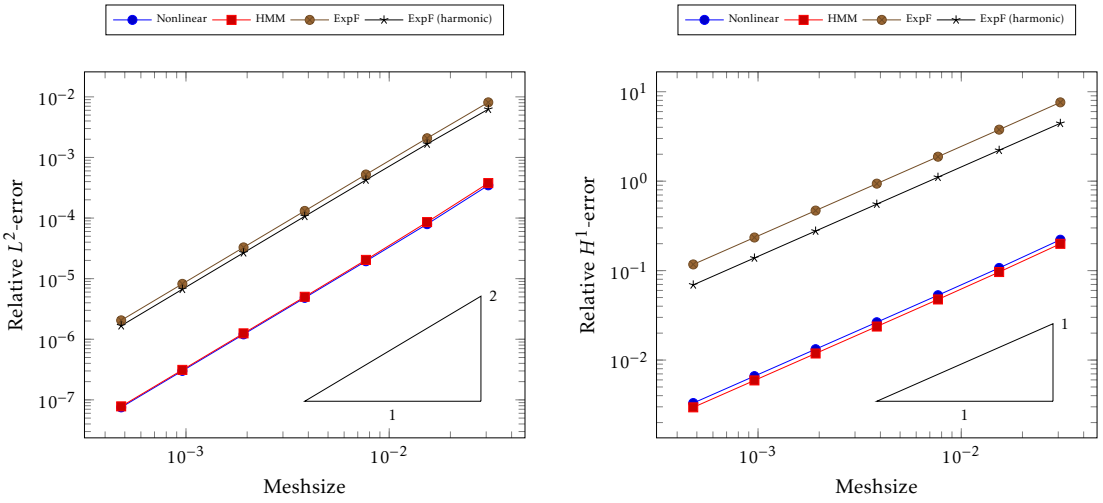


Figure 1.4 – **Accuracy of stationary solutions.** Relative errors in discrete L^2-norm and H^1-seminorm for the first test-case on triangular meshes.

mixed Dirichlet-Neumann boundary conditions. We set $\Gamma^D = (\{0\}\times(0,1))\cup(\{1\}\times(0,1))$, $\Gamma^N = ((0,1)\times\{0\})\cup((0,1)\times\{1\})$, $g^D = 1$, $g^N = 0$, and $f = 0$. The diffusion tensor and the potential are defined by the following expressions:

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & l_y \end{pmatrix} \qquad \text{and} \qquad \phi(x,y) = \log\left(\frac{1}{v} + x\right),$$

with $l_y, v > 0$. Note that the advection field $V^\phi = -\begin{pmatrix} \frac{v}{1+vx} \\ 0 \end{pmatrix}$ has a magnitude of order v when x is small. Thus, near the boundary $\{0\}\times[0,1]$, the problem is advection-dominated if v is large enough. Moreover, $\operatorname{div}(V^\phi) = \dfrac{v^2}{(1+vx)^2} > 0$ and $V^\phi \cdot n = 0$ on Γ^N, so the problem is coercive. The exact solution is given by

$$u(x,y) = \frac{v}{1+vx}\left(\frac{2vx}{2+v}\left(\frac{1}{v} + \frac{x}{2}\right) + \frac{1}{v}\right).$$

We perform our numerical experiments on the Cartesian mesh family (see Figure 1.2a), with

$$l_y = 100 \qquad \text{and} \qquad v = 200.$$

The results are depicted in Figure 1.5. They show that the HMM scheme suffers, most probably because of the fact that the advective term predominates over the diffusive term, at least in some part of the domain. The order of convergence of the HMM scheme is less than one in H^1-seminorm, and than two in L^2-norm. The other schemes converge with order one in H^1-seminorm, and two in L^2-norm. Moreover, on this test-case, their accuracy is better than that of the HMM scheme. Notice that $\omega = \mathrm{e}^{-\phi} = \frac{v}{1+vx}$ has small variations (i.e., not exponential) in the cells, even if v is large. Therefore, the diffusion tensor $\omega\Lambda$ of the problem in the ρ unknown for the exponential fitting schemes is not that heterogeneous (locally). It could explain the good performances of the exponential fitting schemes in this case. Moreover, on this test-case, using harmonic averages in the exponential fitting scheme gives a substantial gain in accuracy for both

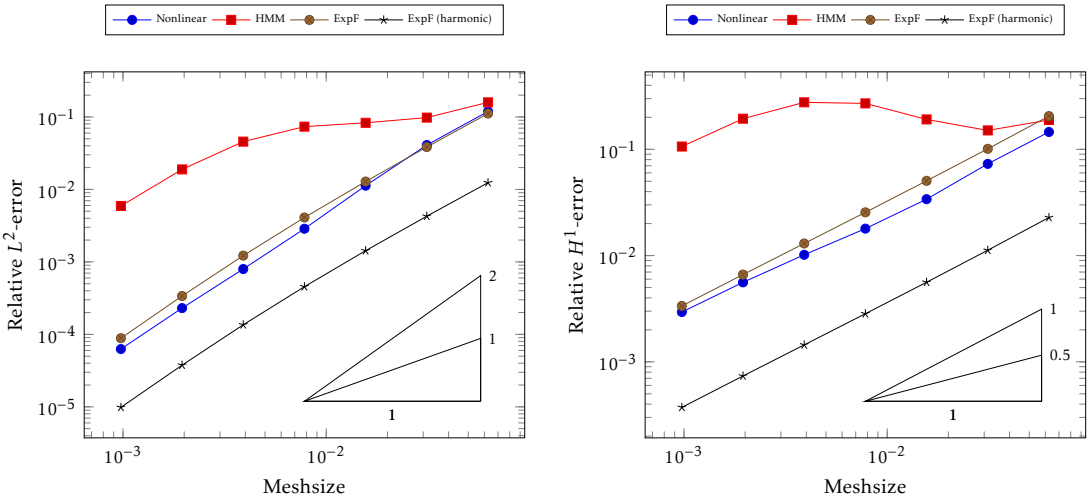


Figure 1.5 – **Accuracy of stationary solutions.** Relative errors in discrete L^2-norm and H^1-seminorm for the second test-case on Cartesian meshes.

the L^2 and H^1 relative errors, of magnitude 10^1.

1.6 Conclusion

In this paper, by means of discrete entropy methods, we have analysed the long-time behaviour of three hybrid finite volume schemes for linear advection-diffusion equations. We have proved that the solutions to all schemes converge exponentially fast in time towards the associated discrete steady-states. Two schemes among the three are new, that are the (linear) exponential fitting scheme (adapting known ideas to the HFV context) and the nonlinear scheme, for which we have proved the existence of solutions. All schemes can handle anisotropy and general meshes. The two linear schemes can deal with general data and mixed Dirichlet-Neumann boundary conditions, however they do not preserve the positivity of solutions. On the other hand, the nonlinear scheme preserves positivity and can be used in practice with general boundary conditions. However, at the moment, its asymptotic analysis is limited to systems that converge in time towards the thermal equilibrium, restricting the admissible data. We have finally validated our theoretical findings on different numerical tests, assessing long-time behaviour, positivity, and spatial accuracy of the schemes.

1.A Functional inequalities

1.A.1 Discrete Poincaré inequalities

We recall the following hybrid discrete Poincaré inequalities (cf. [111, Lemmas B.25 and B.32, $p = 2$]).

Proposition 7 (Discrete Poincaré inequalities)**.** *Let $\mathcal{D}$ be a given discretisation of Ω, with regularity parameter $\theta_{\mathcal{D}}$. There exists $C_{PW} > 0$, only depending on Ω, d, and $\theta_{\mathcal{D}}$ such that*

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}^N_{\mathcal{D},0}, \qquad \|v_{\mathcal{M}}\|_{L^2(\Omega)} \leq C_{PW} |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}. \tag{1.82}$$

Assume that $|\Gamma^D| > 0$. *Then, there exists* $C_{P,\Gamma^D} > 0$, *only depending on* Ω, d, Γ^D, *and* $\theta_{\mathcal{D}}$ *such that*

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}^D_{\mathcal{D},0}, \qquad \|v_{\mathcal{M}}\|_{L^2(\Omega)} \le C_{P,\Gamma^D} |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}. \tag{1.83}$$

1.A.2 Logarithmic Sobolev inequalities

In this section, we derive logarithmic Sobolev inequalities on a bounded domain, in the continuous setting. The intermediate results of Proposition 8 below will be useful in the discrete setting. In the following, μ is a probability measure on the bounded domain Ω, and the space $L^q_\mu(\Omega)$ denotes the space endowed with the norm $\|f\|^q_{L^q_\mu(\Omega)} = \int_\Omega |f|^q \mathrm{d}\mu$. We start with a preliminary lemma, which is an adaptation of part of the proof of [96, Theorem 6.1.22] (see also [153]). We recall that $\Phi_1(s) = s\log(s) - s + 1$.

Lemma 3. *For all* $t \in \mathbb{R}$ *and* $\psi \in L^2_\mu(\Omega)$ *such that* $\|\psi\|_{L^2_\mu(\Omega)} = 1$, *one has*

$$\int_\Omega \Phi_1\left((1+t\psi)^2\right)\mathrm{d}\mu \le t^2 \int_\Omega \psi^2 \log(\psi^2)\mathrm{d}\mu + (1+t^2)\log(1+t^2) + (1+|\langle\psi\rangle_\mu|)t^2,$$

where $\langle\psi\rangle_\mu = \int_\Omega \psi \,\mathrm{d}\mu$.

Proof. Let us define, for $\delta > 0$,

$$f_\delta(t) = \int_\Omega \Phi_1\left((1+t\psi)^2 + \delta\right)\mathrm{d}\mu - t^2 \int_\Omega \psi^2\log(\psi^2)\mathrm{d}\mu - (1+t^2)\log(1+t^2).$$

Differentiating f_δ yields

$$f_\delta'(t) = 2\int_\Omega (1+t\psi)\psi \log\left((1+t\psi)^2+\delta\right)\mathrm{d}\mu - 2t\int_\Omega \psi^2\log(\psi^2)\mathrm{d}\mu - 2t\log(1+t^2) - 2t.$$

In particular, $f_\delta'(0) = 2\log(1+\delta)\langle\psi\rangle_\mu$. Differentiating once more, and using that $\|\psi\|^2_{L^2_\mu(\Omega)} = 1$, we obtain

$$f_\delta''(t) = 2\int_\Omega \psi^2 \log\left(\frac{(1+t\psi)^2+\delta}{(1+t^2)\psi^2}\right)\mathrm{d}\mu + 4\int_\Omega \psi^2 \frac{(1+t\psi)^2}{\delta+(1+t\psi)^2}\mathrm{d}\mu - \frac{4t^2}{1+t^2} - 2.$$

Therefore, using that $\log(x) \le x-1$ in the first term, that $\frac{x}{\delta+x} \le 1$ in the second, together with the fact that μ is a probability measure and that $\|\psi\|^2_{L^2_\mu(\Omega)} = 1$, one gets

$$f_\delta''(t) \le \frac{2\delta}{1+t^2} + \frac{4t}{1+t^2}\langle\psi\rangle_\mu + 4 - \frac{4t^2}{1+t^2} - 2 \le 2\delta + 2|\langle\psi\rangle_\mu| + 2.$$

One concludes by integrating this inequality twice between 0 and t, using that $f_\delta(0) = \Phi_1(1+\delta)$, and letting $\delta \to 0$. □

Proposition 8. *For any* $\xi \in L^q_\mu(\Omega)$ *with* $q > 2$, *one has*

$$\int_\Omega \xi^2 \log\left(\frac{\xi^2}{\|\xi\|^2_{L^2_\mu(\Omega)}}\right)\mathrm{d}\mu \le \frac{q}{q-2}\|\xi - \langle\xi\rangle_\mu\|^2_{L^q_\mu(\Omega)} + \frac{q-4}{q-2}\|\xi-\langle\xi\rangle_\mu\|^2_{L^2_\mu(\Omega)}, \tag{1.84}$$

where $\langle \xi \rangle_\mu = \int_\Omega \xi \, \mathrm{d}\mu$. *Besides, one also has*

$$\int_\Omega \Phi_1(\xi^2)\,\mathrm{d}\mu \le \frac{q}{q-2}\|\xi-1\|^2_{L^q_\mu(\Omega)} + \frac{2q-6}{q-2}\|\xi-1\|^2_{L^2_\mu(\Omega)} + \Phi_1\left(1+\|\xi-1\|^2_{L^2_\mu(\Omega)}\right). \tag{1.85}$$

Proof. i) Assume that $\langle \xi \rangle_\mu \neq 0$, and take t and ψ such that $\xi = \langle \xi \rangle_\mu (1 + t\psi)$ and $\|\psi\|_{L^2_\mu(\Omega)} = 1$. In particular, $\langle \psi \rangle_\mu = 0$, and $1 + t^2 = \frac{\|\xi\|^2_{L^2_\mu(\Omega)}}{\langle \xi \rangle^2_\mu}$. Using Lemma 3, a somewhat tedious but straightforward computation yields

$$\int_\Omega \xi^2 \log\left(\frac{\xi^2}{\|\xi\|^2_{L^2_\mu(\Omega)}}\right)\mathrm{d}\mu \le \int_\Omega (\xi - \langle \xi \rangle_\mu)^2 \log\left(\frac{(\xi-\langle \xi \rangle_\mu)^2}{\|\xi - \langle \xi \rangle_\mu\|^2_{L^2_\mu(\Omega)}}\right)\mathrm{d}\mu + 2\|\xi - \langle \xi \rangle_\mu\|^2_{L^2_\mu(\Omega)}.$$

Observe that the last inequality also holds if $\langle \xi \rangle_\mu = 0$. Let $\phi = \xi - \langle \xi \rangle_\mu$. Then,

$$\int_\Omega \xi^2 \log\left(\frac{\xi^2}{\|\xi\|^2_{L^2_\mu(\Omega)}}\right)\mathrm{d}\mu \le \frac{2}{q-2}\|\phi\|^2_{L^2_\mu(\Omega)} \int_\Omega \frac{\phi^2}{\|\phi\|^2_{L^2_\mu(\Omega)}} \log\left(\frac{\phi^{q-2}}{\|\phi\|^{q-2}_{L^2_\mu(\Omega)}}\right)\mathrm{d}\mu + 2\|\phi\|^2_{L^2_\mu(\Omega)}.$$

Therefore, by Jensen's inequality for the probability measure $\frac{\phi^2}{\|\phi\|^2_{L^2_\mu(\Omega)}}\mathrm{d}\mu$ applied to the concave function log, one obtains

$$\begin{aligned}\int_\Omega \xi^2 \log\left(\frac{\xi^2}{\|\xi\|^2_{L^2_\mu(\Omega)}}\right)\mathrm{d}\mu &\le \frac{2}{q-2}\|\phi\|^2_{L^2_\mu(\Omega)} \log\left(\frac{\|\phi\|^q_{L^q_\mu(\Omega)}}{\|\phi\|^q_{L^2_\mu(\Omega)}}\right) + 2\|\phi\|^2_{L^2_\mu(\Omega)} \\ &= \frac{q}{q-2}\|\phi\|^2_{L^2_\mu(\Omega)} \log\left(\frac{\|\phi\|^2_{L^q_\mu(\Omega)}}{\|\phi\|^2_{L^2_\mu(\Omega)}}\right) + 2\|\phi\|^2_{L^2_\mu(\Omega)},\end{aligned}$$

and one concludes using that $\log(x) \le x - 1$.

ii) Take t and ψ such that $\xi = 1 + t\psi$ and $\|\psi\|_{L^2_\mu(\Omega)} = 1$. Remark that $t = \|\xi - 1\|_{L^2_\mu(\Omega)}$. Using that $|\langle \psi \rangle_\mu| \le \|\psi\|_{L^2_\mu(\Omega)} = 1$, Lemma 3 yields

$$\int_\Omega \Phi_1(\xi^2)\,\mathrm{d}\mu - \Phi_1\left(1+\|\xi-1\|^2_{L^2_\mu(\Omega)}\right) \le \int_\Omega (\xi-1)^2 \log\left(\frac{(\xi-1)^2}{\|\xi-1\|^2_{L^2_\mu(\Omega)}}\right)\mathrm{d}\mu + 3\|\xi-1\|^2_{L^2_\mu(\Omega)}.$$

Letting $\phi = \xi - 1$, the proof goes on as for i).

□

From there, logarithmic Sobolev inequalities are immediate consequences of Poincaré–Sobolev inequalities, of [51, Lemma 5.2], and of the fact that $\Phi_1(1+s) \le s\log(1+s)$.

Corollary 1 (Logarithmic Sobolev inequalities)**.** *Assume that* μ *has a density (still denoted by* μ*) with respect to the Lebesgue measure such that* $0 < \mu_\flat \le \mu(x) \le \mu_\sharp$ *for a.e.* $x \in \Omega$. *Then, for any*

$\xi \in H^1(\Omega)$, *one has*

$$\int_\Omega \xi^2 \log\left(\frac{\xi^2}{\|\xi\|^2_{L^2_\mu(\Omega)}}\right)\mathrm{d}\mu \le C(\Omega,d,\mu_\flat,\mu_\sharp)\|\nabla\xi\|^2_{L^2_\mu(\Omega;\mathbb{R}^d)}.$$

Besides, if $|\Gamma^D|>0$ *and* $\xi-1\in H^{1,D}_0(\Omega)=\{v\in H^1(\Omega)\mid v=0 \text{ on } \Gamma^D\}$, *then*

$$\int_\Omega \Phi_1(\xi^2)\,\mathrm{d}\mu \le C'(\Omega,d,\mu_\flat,\mu_\sharp)\left(1+\log\left(1+\|\xi-1\|^2_{L^2_\mu(\Omega)}\right)\right)\|\nabla\xi\|^2_{L^2_\mu(\Omega;\mathbb{R}^d)}.$$

1.A.3 Discrete logarithmic Sobolev inequalities

Similarly to what was done in [51], one can derive discrete logarithmic Sobolev inequalities adapted to the hybrid setting.

Proposition 9 (Discrete logarithmic Sobolev inequality, Neumann case)**.** *Let* $\mathcal{D}$ *be a given discretisation of* Ω, *with regularity parameter* $\theta_\mathcal{D}$. *Let* $\underline{v}_\mathcal{D},\underline{v}^\infty_\mathcal{D}\in\underline{V}_\mathcal{D}$ *be two positive vectors of unknowns such that*

$$\int_\Omega v_\mathcal{M}=\int_\Omega v^\infty_\mathcal{M}=M,$$

and set $v^\infty_{\mathcal{M},\sharp}=\sup_{K\in\mathcal{M}} v^\infty_K$. *Define* $\underline{\xi}_\mathcal{D}$ *as the element of* $\underline{V}_\mathcal{D}$ *such that*

$$\xi_K=\sqrt{\frac{v_K}{v^\infty_K}}\quad\forall K\in\mathcal{M},\qquad \xi_\sigma=\sqrt{\frac{v_\sigma}{v^\infty_\sigma}}\quad\forall\sigma\in\mathcal{E}.$$

Then, there exists $C_{LS,\infty}>0$, *only depending on* M, $v^\infty_{\mathcal{M},\sharp}$, Ω, d, *and* $\theta_\mathcal{D}$ *such that*

$$\int_\Omega v^\infty_\mathcal{M}\Phi_1\left(\xi^2_\mathcal{M}\right)\le C^2_{LS,\infty}\left|\underline{\xi}_\mathcal{D}\right|^2_{1,\mathcal{D}}. \tag{1.86}$$

Proof. By (1.84) and [51, Lemma 5.2] applied to the probability measure $\mu(x)\mathrm{d}x=v^\infty_\mathcal{M}(x)\frac{\mathrm{d}x}{M}$ and to the function $\xi_\mathcal{M}=\sqrt{\frac{v_\mathcal{M}}{v^\infty_\mathcal{M}}}$, we first infer that, for $q>2$,

$$\int_\Omega v_\mathcal{M}\log\left(\xi^2_\mathcal{M}\right)\le C(M,v^\infty_{\mathcal{M},\sharp},q)\left(\|\xi_\mathcal{M}-\overline{\xi_\mathcal{M}}\|^2_{L^q(\Omega)}+\|\xi_\mathcal{M}-\overline{\xi_\mathcal{M}}\|^2_{L^2(\Omega)}\right),$$

where we let $\overline{\xi_\mathcal{M}}=\frac{1}{|\Omega|}\int_\Omega\xi_\mathcal{M}\mathrm{d}x$. The conclusion then falls in two steps. On the one hand, since $\int_\Omega v_\mathcal{M}=\int_\Omega v^\infty_\mathcal{M}$, we remark that

$$\int_\Omega v^\infty_\mathcal{M}\Phi_1\left(\xi^2_\mathcal{M}\right)=\int_\Omega v_\mathcal{M}\log\left(\xi^2_\mathcal{M}\right).$$

On the other hand, we invoke (1.82) and the discrete Poincaré–Sobolev inequality of [111, Lemma B.25, $p=2$] for $2<q<\frac{2d}{d-2}$:

$$\forall\underline{w}_\mathcal{D}\in\underline{V}^N_{\mathcal{D},0},\qquad \|w_\mathcal{M}\|_{L^q(\Omega)}\le C_{PS}|\underline{w}_\mathcal{D}|_{1,\mathcal{D}},$$

where $C_{PS} > 0$ only depends on Ω, d, and $\theta_{\mathcal{D}}$, that we apply to $\underline{w}_{\mathcal{D}} = \underline{\xi}_{\mathcal{D}} - \overline{\xi_{\mathcal{M}}}\underline{1}_{\mathcal{D}} \in \underline{V}^N_{\mathcal{D},0}$. This proves (1.86).

□

Starting from (1.85), a similar proof yields the following result. The relevant discrete Poincaré–Sobolev inequality in this case is given in [111, Lemma B.32, $p = 2$].

Proposition 10 (Discrete logarithmic Sobolev inequality, Dirichlet case)**.** *Assume that $|\Gamma^D| > 0$. Let $\mathcal{D}$ be a given discretisation of Ω, with regularity parameter $\theta_{\mathcal{D}}$. Let $\underline{v}_{\mathcal{D}}, \underline{v}^\infty_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ be two positive vectors of unknowns such that*

$$\underline{v}_{\mathcal{D}} - \underline{v}^\infty_{\mathcal{D}} \in \underline{V}^D_{\mathcal{D},0},$$

and set $v^\infty_{\mathcal{M},\sharp} = \sup_{K\in\mathcal{M}} v^\infty_K$ and $M^\infty := \int_\Omega v^\infty_{\mathcal{M}}$. Define $\underline{\xi}_{\mathcal{D}}$ as the element of $\underline{V}_{\mathcal{D}}$ such that

$$\xi_K = \sqrt{\frac{v_K}{v^\infty_K}} \quad \forall K \in \mathcal{M}, \qquad \xi_\sigma = \sqrt{\frac{v_\sigma}{v^\infty_\sigma}} \quad \forall \sigma \in \mathcal{E}.$$

Then, letting $\mu = \frac{v^\infty_{\mathcal{M}}}{M^\infty}$, there exists $C_{LS,\Gamma^D,\infty} > 0$, only depending on M^∞, $v^\infty_{\mathcal{M},\sharp}$, Ω, d, Γ^D, and $\theta_{\mathcal{D}}$ such that

$$\int_\Omega v^\infty_{\mathcal{M}} \Phi_1(\xi^2_{\mathcal{M}}) \le C^2_{LS,\Gamma^D,\infty}\left(1 + \log\left(1 + \|\xi_{\mathcal{M}} - 1\|^2_{L^2_\mu(\Omega)}\right)\right)|\underline{\xi}_{\mathcal{D}}|^2_{1,\mathcal{D}}. \tag{1.87}$$

1.B Nonlinear scheme for mixed Dirichlet-Neumann boundary conditions

In this appendix, we introduce and analyse a version of the nonlinear scheme for the evolution problem (1.1) when $|\Gamma^D| > 0$. In order to perform the asymptotic analysis, we need to assume that the data are compatible with the thermal equilibrium:

$$f = 0, \quad g^N = 0, \quad \text{and there exists } \rho^D > 0 \text{ such that } g^D = \rho^D \mathrm{e}^{-\phi} = \rho^D \omega.$$

For such data, given $u^{in} \ge 0$, the solution u to (1.1) is positive for $t > 0$, and converges towards $u^\infty = \rho^D \mathrm{e}^{-\phi}$ when $t \to \infty$.

1.B.1 Scheme and well-posedness

Accordingly to this setting, we define $\underline{u}^\infty_{\mathcal{D}} = \rho^D \underline{\omega}_{\mathcal{D}}$. One has $u^\infty_\flat \underline{1}_{\mathcal{D}} \le \underline{u}^\infty_{\mathcal{D}} \le u^\infty_\sharp \underline{1}_{\mathcal{D}}$, where $0 < u^\infty_\flat \le u^\infty_\sharp$ only depend on ρ^D, ϕ, and Ω. Remind that, as in (1.47), given a positive $\underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, one defines $\underline{w}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D}}$ as

$$w_K = \log\left(\frac{u_K}{u^\infty_K}\right) \quad \forall K \in \mathcal{M}, \qquad w_\sigma = \log\left(\frac{u_\sigma}{u^\infty_\sigma}\right) \quad \forall \sigma \in \mathcal{E}.$$

For mixed boundary conditions, the discrete problem reads: Find $\left(\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}\right)_{n\geq 1}$ positive such that

$$\begin{cases} \dfrac{1}{\Delta t}(u_{\mathcal{M}}^n - u_{\mathcal{M}}^{n-1}, v_{\mathcal{M}})_\Omega + T_{\mathcal{D}}\left(\underline{u}_{\mathcal{D}}^n, \underline{w}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}^n), \underline{v}_{\mathcal{D}}\right) = 0 & \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}^D, \quad (1.88a) \\ \underline{w}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}^n) \in \underline{V}_{\mathcal{D},0}^D, & (1.88b) \\ u_K^0 = \dfrac{1}{|K|}\displaystyle\int_K u^{in} \qquad \forall K \in \mathcal{M}. & (1.88c) \end{cases}$$

Notice that, since for all $\sigma \in \mathcal{E}$, $w_\sigma^n = \log\left(\dfrac{u_\sigma^n}{u_\sigma^\infty}\right)$, the equation (1.88b) only means that, for all $\sigma \in \mathcal{E}_{ext}^D$, $u_\sigma^n = u_\sigma^\infty$, which enforces strongly the Dirichlet boundary condition on Γ^D. One can show the following existence result.

Theorem 5 (Existence of positive solutions and entropy/entropy dissipation relation)**.** *Let $u^{in} \in L^2(\Omega)$ be a non-negative function. There exists at least one positive solution $\left(\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}\right)_{n\geq 1}$ to the nonlinear scheme (1.88). It satisfies the following entropy/dissipation relation:*

$$\forall n \in \mathbb{N}, \qquad \frac{\mathbb{E}^{n+1} - \mathbb{E}^n}{\Delta t} + \mathbb{D}^{n+1} \leq 0, \tag{1.89}$$

where $\mathbb{E}^n$ and $\mathbb{D}^n$ are, respectively, the discrete relative entropy and dissipation defined in (1.49). Moreover, there exists $\varepsilon > 0$, depending on Λ, ϕ, u^{in}, ρ^D, Ω, d, Δt, and $\mathcal{D}$ such that, for all $n \geq 1$, $u_K^n \geq \varepsilon$ for all $K \in \mathcal{M}$ and $u_\sigma^n \geq \varepsilon$ for all $\sigma \in \mathcal{E}$.

The proof of this theorem relies on the same arguments as the one of Theorem 1 for (homogeneous) pure Neumann boundary conditions. The major difference lies in the counterpart of Lemma 2, which is no longer based on the positivity of the mass, but on the prescribed (zero) value on the Dirichlet faces.

1.B.2 Long-time behaviour

In the next theorem, we state the long-time behaviour of the discrete solutions to the nonlinear scheme (1.88).

Theorem 6 (Asymptotic stability)**.** *If $\left(\underline{u}_{\mathcal{D}}^n \in \underline{V}_{\mathcal{D}}\right)_{n\geq 1}$ is a (positive) solution to (1.88), then the discrete entropy decays exponentially fast in time: there is $\nu_{\mathrm{nl},\Gamma^D} > 0$, depending on Λ, ϕ, Γ^D, ρ^D, u^{in}, Ω, d, and $\theta_{\mathcal{D}}$ such that*

$$\forall n \in \mathbb{N}, \qquad \mathbb{E}^{n+1} \leq (1 + \nu_{\mathrm{nl},\Gamma^D}\,\Delta t)^{-1}\mathbb{E}^n. \tag{1.90}$$

Consequently, the discrete solution converges exponentially fast in time towards its associated discrete steady-state.

Proof. Let $n \in \mathbb{N}^\star$. As in Section 1.4.3, one has $\mathbb{D}^n \geq \dfrac{1}{C_F}\hat{\mathbb{D}}^n \geq \dfrac{4u_\flat^\infty \lambda_\flat \alpha_\flat}{C_F}\left|\underline{\xi}_{\mathcal{D}}^n\right|_{1,\mathcal{D}}^2$, where $C_F > 0$ depends on the data. Using the discrete log–Sobolev inequality (1.87) from Proposition 10, we

get

$$\mathbb{E}^n \le C_{LS,\Gamma^D,\infty}^2 \left(1+\log\left(1+\|\xi_{\mathcal{M}}^n-1\|_{L^2_{\mu^\infty}(\Omega)}^2\right)\right)|\underline{\xi}_{\mathcal{D}}^n|_{1,\mathcal{D}}^2$$
$$\le \frac{C_{LS,\Gamma^D,\infty}^2 C_F}{4u_\flat^\infty \lambda_\flat \alpha_\flat}\left(1+\log\left(1+\|\xi_{\mathcal{M}}^n-1\|_{L^2_{\mu^\infty}(\Omega)}^2\right)\right)\mathbb{D}^n. \quad (1.91)$$

Then, there is $C>0$ such that (recall that $\underline{\xi}_{\mathcal{D}}^n$ is positive)

$$\|\xi_{\mathcal{M}}^n-1\|_{L^2_{\mu^\infty}(\Omega)}^2 \le \|(\xi_{\mathcal{M}}^n)^2\|_{L^1_{\mu^\infty}(\Omega)}+1 \le \left\|\Phi_1\left((\xi_{\mathcal{M}}^n)^2\right)\right\|_{L^1_{\mu^\infty}(\Omega)}+C = (M^\infty)^{-1}\mathbb{E}^n+C,$$

where the last inequality is an application of the Fenchel–Young inequality $x \le \Phi_1(x)+\Phi_1^\star(1)$, where $\Phi_1^\star$ is the convex conjugate of Φ_1 and $x=(\xi_{\mathcal{M}}^n)^2$. But, since the entropy/dissipation relation (1.89) holds, the discrete entropy decays and $\mathbb{E}^n \le \mathbb{E}^0$. Therefore, one has

$$\|\xi_{\mathcal{M}}^n-1\|_{L^2_{\mu^\infty}(\Omega)}^2 \le (M^\infty)^{-1}\mathbb{E}^0+C.$$

Combining this estimate with (1.91), we deduce that there exists $\nu_{\mathrm{nl},\Gamma^D}>0$, depending on Λ, ϕ, Γ^D, ρ^D, u^{in}, Ω, d, and $\theta_{\mathcal{D}}$ such that $\mathbb{E}^n \le \nu_{\mathrm{nl},\Gamma^D}\mathbb{D}^n$. Then, using the entropy/dissipation relation (1.89), we get (1.90). □

1.C Proofs of technical results

1.C.1 Discrete boundedness by mass and dissipation

We prove Lemma 2 from Section 1.3.3. To ease the reading, we first recall the result.

Lemma 2. *Let $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, and assume that there exist $C_\sharp>0$, and $M_\sharp \ge M_\flat>0$ such that*

$$M_\flat \le \sum_{K\in\mathcal{M}} |K| u_K^\infty \mathrm{e}^{w_K} \le M_\sharp \qquad \text{and} \qquad \mathbb{D}(\underline{w}_{\mathcal{D}}) \le C_\sharp. \quad (1.53)$$

Then, there exists $C>0$, depending on Λ, $u_\flat^\infty$, $u_\sharp^\infty$, $M_\flat$, $M_\sharp$, $C_\sharp$, Ω, d, and $\mathcal{D}$ such that

$$|w_K| \le C \quad \forall K\in\mathcal{M} \quad \text{and} \quad |w_\sigma| \le C \quad \forall \sigma\in\mathcal{E}.$$

Proof. For $K\in\mathcal{M}$, using (1.12) and (1.13), we first infer that

$$\delta_K\underline{w}_K \cdot \mathbb{A}_K\delta_K\underline{w}_K = a_K^\Lambda(\underline{w}_K,\underline{w}_K) \ge \lambda_\flat\alpha_\flat|\underline{w}_K|_{1,K}^2 = \lambda_\flat\alpha_\flat \sum_{\sigma\in\mathcal{E}_K}\frac{|\sigma|}{d_{K,\sigma}}\left(w_K-w_\sigma\right)^2.$$

By definition (1.3) of the regularity parameter $\theta_{\mathcal{D}}$, we have that $\frac{|\sigma|}{d_{K,\sigma}} \ge \frac{h_K^{d-2}}{\theta_{\mathcal{D}}}$ for all $\sigma\in\mathcal{E}_K$, so that

$$\delta_K\underline{w}_K \cdot \mathbb{A}_K\delta_K\underline{w}_K \ge \frac{\lambda_\flat\alpha_\flat}{\theta_{\mathcal{D}}}h_K^{d-2}|\delta_K\underline{w}_K|^2. \quad (1.92)$$

By the expression (1.52) of $\mathbb{D}(\underline{w}_\mathcal{D})$, and the local lower bound (1.92), we thus get

$$\begin{aligned}\mathbb{D}(\underline{w}_\mathcal{D}) &= \sum_{K\in\mathcal{M}} r_K\Big(\underline{u}_K^\infty \times \exp(\underline{w}_K)\Big)\delta_K\underline{w}_K \cdot \mathbb{A}_K\delta_K\underline{w}_K \\ &\geq \frac{\lambda_\flat\alpha_\flat}{\theta_\mathcal{D}} \sum_{K\in\mathcal{M}} h_K^{d-2} r_K\Big(\underline{u}_K^\infty \times \exp(\underline{w}_K)\Big)|\delta_K\underline{w}_K|^2 \\ &= \frac{\lambda_\flat\alpha_\flat}{\theta_\mathcal{D}} \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} h_K^{d-2} r_K\Big(\underline{u}_K^\infty \times \exp(\underline{w}_K)\Big)(w_K - w_\sigma)^2.\end{aligned}$$

Let $K\in\mathcal{M}$ and $\sigma\in\mathcal{E}_K$ be fixed. Using, successively, the definition (1.41) of r_K combined with the definition (1.43) of $f_{|\mathcal{E}_K|}$, the combination of (1.48) with assumptions (1.42a) and (1.42c), and the assumptions (1.42b) and (1.42d) combined with the bound (1.4) on $|\mathcal{E}_K|$, we infer, for $w_\sigma \neq w_K$,

$$\begin{aligned}r_K\Big(\underline{u}_K^\infty \times \exp(\underline{w}_K)\Big)(w_K - w_\sigma)^2 &\geq \frac{1}{|\mathcal{E}_K|}\, m\Big(u_K^\infty \mathrm{e}^{w_K}, u_\sigma^\infty \mathrm{e}^{w_\sigma}\Big)(w_K - w_\sigma)^2 \\ &\geq \frac{u_\flat^\infty}{|\mathcal{E}_K|}\, m\Big(\mathrm{e}^{w_K}, \mathrm{e}^{w_\sigma}\Big)(w_K - w_\sigma)^2 \\ &\geq \frac{u_\flat^\infty}{d\theta_\mathcal{D}^2}\,(\mathrm{e}^{w_K} - \mathrm{e}^{w_\sigma})(w_K - w_\sigma) \geq 0,\end{aligned}$$

and we verify that this inequality still holds when $w_\sigma = w_K$. Since $\mathbb{D}(\underline{w}_\mathcal{D}) \leq C_\sharp$ by (1.53), for all $K\in\mathcal{M}$, and all $\sigma\in\mathcal{E}_K$, we have

$$0 \leq (\mathrm{e}^{w_K} - \mathrm{e}^{w_\sigma})(w_K - w_\sigma) \leq \zeta h_K^{2-d}, \tag{1.93}$$

with $\zeta = \frac{dC_\sharp\theta_\mathcal{D}^3}{\lambda_\flat\alpha_\flat u_\flat^\infty} > 0$ (recall that $\alpha_\flat$ depends on Ω, d, and $\theta_\mathcal{D}$). Besides, since $\sum_{K\in\mathcal{M}}|K|u_K^\infty \mathrm{e}^{w_K} \leq M_\sharp$ again by (1.53), we have $|K|u_K^\infty \mathrm{e}^{w_K} \leq M_\sharp$ for all $K\in\mathcal{M}$. Similarly, since $\sum_{K\in\mathcal{M}}|K|u_K^\infty \mathrm{e}^{w_K} \geq M_\flat$, there exists $K_0\in\mathcal{M}$ such that $|\Omega|u_{K_0}^\infty \mathrm{e}^{w_{K_0}} \geq M_\flat > 0$. Combining these bounds, we infer that there exists $K_0\in\mathcal{M}$ such that

$$\log\left(\frac{M_\flat}{|\Omega|u_\sharp^\infty}\right) \leq w_{K_0} \leq \log\left(\frac{M_\sharp}{|K_0|u_\flat^\infty}\right). \tag{1.94}$$

Now, let us show that we can similarly frame all the other components of $\underline{w}_\mathcal{D}$.

For $a, x\in\mathbb{R}$, let us define $E(a,x) = (\mathrm{e}^x - \mathrm{e}^a)(x-a) \geq 0$. Observe that $E(a, y+a)\mathrm{e}^{-a} = (\mathrm{e}^y - 1)y =: \xi(y)$ and that ξ is continuous, strictly decreasing for $y<0$, strictly increasing for $y>0$, $\xi(0)=0$, and $\xi(y)\to+\infty$ when $y\to\pm\infty$. Let $b, a_\sharp > 0$, and take $|a| \leq a_\sharp$. By the properties of ξ, if $E(a,x)\leq b$, then $|x| \leq \kappa_b(a_\sharp) := a_\sharp + \max\{|y| \text{ s.t. } \xi(y) = b\,\mathrm{e}^{a_\sharp}\}$. We can thus infer that if $(x_k)_{k=0,\dots,m}$ is a finite sequence of real numbers such that $E(x_k, x_{k+1}) \leq b$ and $|x_0| \leq a_\sharp$, then $|x_m| \leq \kappa_b^{(m)}(a_\sharp)$ where $\kappa_b^{(m)}$ is m compositions of κ_b. In particular, the bound only depends on $a_\sharp$, m and b.

Now we can conclude the proof. Because of the connectivity of the mesh, for any cell K (respectively, face σ) there is a finite sequence of components of $\underline{w}_\mathcal{D}$, denoted $(x_k)_{k=0,\dots,m}$, starting at $x_0 = w_{K_0}$ and finishing at $x_m = w_K$ (respectively, $x_m = w_\sigma$) such that, by (1.93), $E(x_k, x_{k+1}) \leq b := \zeta h_\mathcal{D}^{2-d}$. The inequality (1.94) yields the initial bound on $|x_0|$, and one concludes by the above argument.

□

1.C.2 A local comparison result

We prove a local comparison result between the matrices $\mathbb{A}_K$ and some (local) diagonal matrices. The proof relies on arguments that are similar to those advocated in [56] to analyse the VAG scheme.

Lemma 4. *For $K \in \mathcal{M}$, let $\mathbb{A}_K \in \mathbb{R}^{|\mathcal{E}_K|\times|\mathcal{E}_K|}$ be the matrix defined by (1.10). The matrices $\mathbb{A}_K$ are symmetric positive-definite, and there exists $C_A > 0$, only depending on Λ, Ω, d, and $\theta_{\mathcal{D}}$ such that*

$$\forall K \in \mathcal{M}, \quad \mathrm{Cond}_2(\mathbb{A}_K) \le C_A,$$

where $\mathrm{Cond}_2(\mathbb{A}_K) = \|\mathbb{A}_K^{-1}\|_2\|\mathbb{A}_K\|_2$ is the condition number of the matrix $\mathbb{A}_K$. Moreover, letting for $K \in \mathcal{M}$, $\mathbb{B}_K \in \mathbb{R}^{|\mathcal{E}_K|\times|\mathcal{E}_K|}$ be the diagonal matrix with entries

$$B_K^{\sigma\sigma} = \sum_{\sigma'\in\mathcal{E}_K} |A_K^{\sigma\sigma'}| \qquad \text{for all } \sigma \in \mathcal{E}_K, \tag{1.95}$$

there exists $C_B > 0$, only depending on Λ, Ω, d, and $\theta_{\mathcal{D}}$ such that

$$\forall K \in \mathcal{M}, \forall w \in \mathbb{R}^{|\mathcal{E}_K|}, \qquad w \cdot \mathbb{A}_K w \le w \cdot \mathbb{B}_K w \le C_B\, w \cdot \mathbb{A}_K w. \tag{1.96}$$

Proof. Let $K \in \mathcal{M}$ and $k = |\mathcal{E}_K|$. As a direct consequence of its definition (1.10), the matrix $\mathbb{A}_K \in \mathbb{R}^{k\times k}$ is symmetric and positive semi-definite. Now, let $w = (w_\sigma)_{\sigma\in\mathcal{E}_K} \in \mathbb{R}^k$, and define $\underline{v}_K \in \underline{V}_K$ such that

$$v_K = 0 \qquad \text{and} \qquad v_\sigma = -w_\sigma \text{ for all } \sigma \in \mathcal{E}_K.$$

Then, $\delta_K \underline{v}_K = (v_K - v_\sigma)_{\sigma\in\mathcal{E}_K} = w$. By (1.92), we immediately get that

$$w \cdot \mathbb{A}_K w \ge \frac{\lambda_\flat \alpha_\flat}{\theta_{\mathcal{D}}} h_K^{d-2}|w|^2,$$

which implies, since $w \in \mathbb{R}^k$ is arbitrary, that $\mathbb{A}_K$ is invertible, and gives us a lower bound on its smallest eigenvalue. By the same arguments advocated to prove (1.92), noticing that $\frac{|\sigma|}{d_{K,\sigma}} \le \theta_{\mathcal{D}} h_K^{d-2}$ for all $\sigma \in \mathcal{E}_K$, we infer that

$$w \cdot \mathbb{A}_K w \le \lambda_\sharp \alpha_\sharp \theta_{\mathcal{D}} h_K^{d-2}|w|^2.$$

We eventually get, using the estimates on the eigenvalues of $\mathbb{A}_K$, that

$$\mathrm{Cond}_2(\mathbb{A}_K) \le \frac{\lambda_\sharp \alpha_\sharp}{\lambda_\flat \alpha_\flat}\theta_{\mathcal{D}}^2 = C_A, \tag{1.97}$$

with $C_A > 0$ only depending on Λ, Ω, d, and $\theta_{\mathcal{D}}$. Now, by (1.95), since $\mathbb{A}_K$ is symmetric, we have

$$w \cdot \mathbb{B}_K w = \sum_{\sigma\in\mathcal{E}_K}\sum_{\sigma'\in\mathcal{E}_K} |A_K^{\sigma\sigma'}| w_\sigma^2 = \sum_{\sigma\in\mathcal{E}_K}\sum_{\sigma'\in\mathcal{E}_K} |A_K^{\sigma\sigma'}| w_{\sigma'}^2,$$

and we can use the half-sum to get

$$w \cdot \mathbb{B}_K w = \sum_{\sigma\in\mathcal{E}_K}\sum_{\sigma'\in\mathcal{E}_K} |A_K^{\sigma\sigma'}| \frac{w_\sigma^2 + w_{\sigma'}^2}{2}.$$

Using Young's inequality, we infer

$$w\cdot\mathbb{A}_K w=\sum_{\sigma\in\mathcal{E}_K}\sum_{\sigma'\in\mathcal{E}_K}A_K^{\sigma\sigma'}w_\sigma w_{\sigma'}\le\sum_{\sigma\in\mathcal{E}_K}\sum_{\sigma'\in\mathcal{E}_K}|A_K^{\sigma\sigma'}||w_\sigma||w_{\sigma'}|$$
$$\le\sum_{\sigma\in\mathcal{E}_K}\sum_{\sigma'\in\mathcal{E}_K}|A_K^{\sigma\sigma'}|\frac{w_\sigma^2+w_{\sigma'}^2}{2}=w\cdot\mathbb{B}_K w.$$

For the second inequality, by symmetry of $\mathbb{A}_K$, we have

$$w\cdot\mathbb{B}_K w=\sum_{\sigma\in\mathcal{E}_K}B_K^{\sigma\sigma}w_\sigma^2\le\max_{\sigma\in\mathcal{E}_K}(B_K^{\sigma\sigma})\sum_{\sigma\in\mathcal{E}_K}w_\sigma^2=\max_{\sigma\in\mathcal{E}_K}\left(\sum_{\sigma'\in\mathcal{E}_K}|A_K^{\sigma'\sigma}|\right)|w|^2=\|\mathbb{A}_K\|_1|w|^2.$$

The space $\mathbb{R}^{k\times k}$ being of finite dimension, the norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent, and there exists $\gamma_k>0$ such that $\|\cdot\|_1\le\gamma_k\|\cdot\|_2$. Moreover, since $\mathbb{A}_K$ is symmetric positive-definite, the following inequality holds:

$$w\cdot\mathbb{A}_K w\ge\frac{\|\mathbb{A}_K\|_2}{\mathrm{Cond}_2(\mathbb{A}_K)}|w|^2.$$

From the previous estimates and (1.97), we deduce that

$$w\cdot\mathbb{B}_K w\le\gamma_k\,\mathrm{Cond}_2(\mathbb{A}_K)\,w\cdot\mathbb{A}_K w\le\gamma_k\,C_A\,w\cdot\mathbb{A}_K w.$$

But, according to (1.4), we have $\max_{K\in\mathcal{M}}\gamma_k\le\max_{(d+1)\le l\le d\theta_{\mathcal{D}}^2}\gamma_l$, therefore

$$w\cdot\mathbb{B}_K w\le C_B\,w\cdot\mathbb{A}_K w,$$

where $C_B=C_A\max_{(d+1)\le l\le d\theta_{\mathcal{D}}^2}\gamma_l$ is a positive constant only depending on Λ, Ω, d, and $\theta_{\mathcal{D}}$. This completes the proof of the comparison result (1.96). □

Chapter 2

A structure preserving hybrid finite volume scheme for semiconductor models with magnetic field on general meshes

Outline of the current chapter

This chapter content is the article [213] published in ESAIM: Mathematical Modelling and Numerical Analysis.

We are interested in the discretisation of a drift-diffusion system in the framework of hybrid finite volume (HFV) methods on general polygonal/polyhedral meshes. The system under study is composed of two anisotropic and nonlinear convection-diffusion equations with nonsymmetric tensors, coupled with a Poisson equation and describes in particular semiconductor devices immersed in a magnetic field. We introduce a new scheme based on an entropy-dissipation relation and prove that the scheme admits solutions with values in admissible sets - especially, the computed densities remain positive. Moreover, we show that the discrete solutions to the scheme converge exponentially fast in time towards the associated discrete thermal equilibrium. Several numerical tests confirm our theoretical results. Up to our knowledge, this scheme is the first one able to discretise anisotropic drift-diffusion systems while preserving the bounds on the densities.

2.1 Introduction

We are interested in the numerical approximation of a generalised anisotropic Van Roosbroeck's drift-diffusion system, inspired by the model introduced in [143] by Gajewski and Gärtner to describe semiconductors devices immersed in exterior magnetic fields. Such a model differs from the classical drift-diffusion system from its anisotropic nature, induced by the magnetic field. From a numerical point of view, this anisotropy leads to several difficulties, including the preservation of bounds on the solution (see the numerical results of [143]). The main originality of our work is to address this difficulty using a nonlinear Hybrid Finite Volume method [123, 70], specifically designed to handle anisotropy while preserving the bounds. Hence, up to our knowledge, the scheme introduced here is the first bound-preserving scheme for anisotropic drift-diffusion systems. From a general perspective, the Van Roosbroeck's drift-diffusion system, initially introduced in [258], is one of the fundamental models for the description and the simulation of semiconductor devices. It is a macroscopic model taking into account the densities of the charge carriers (the electrons, negatively charged, and the holes, of positive charge) alongside with an electrostatic potential induced by the spatial inhomogeneity of the charges. Mathematically speaking, the system consists of two parabolic convection-diffusion equations coupled with a Poisson equation on the electrostatic potential. Among the other different generalisations of the initial model proposed in [258], one can mention the incorporation of nonlinear terms of recombination-generation (see for example [208]), as well as the use of non-Boltzmann statistics [144, 125] to describe some specific physical situations (high carrier densities) or devices (for example, organic semiconductors [257]), which leads to the modification of the diffusive term into a nonlinear one. Another generalisation of the system, as discussed above, is presented in [143]: the authors consider that the semiconductor is immersed into an exterior magnetic field. Such a situation leads to the introduction of anisotropy tensors (which are nonsymmetric) in the convection-diffusion equations, related to the magnetic field.

In the present work, we consider a general model that encompasses the different features mentioned above. More precisely, let Ω be an open, bounded, connected polytopal subsets of $\mathbb{R}^d$, with $d \in \{1, 2, 3\}$ whose boundary $\partial\Omega$ is divided into two disjoint subset $\partial\Omega = \overline{\Gamma^D} \cup \overline{\Gamma^N}$, with $|\Gamma^D| > 0$. We are interested in the following problem, where the unknowns N, P and ϕ are

functions from $\mathbb{R}_+ \times \Omega$ to $\mathbb{R}$:

$$\begin{cases} \partial_t N - \operatorname{div}(N\Lambda_N \nabla(h(N) - \phi)) = -R(N,P) & \text{in } \mathbb{R}_+ \times \Omega \\ \partial_t P - \operatorname{div}(P\Lambda_P \nabla(h(P) + \phi)) = -R(N,P) & \text{in } \mathbb{R}_+ \times \Omega \\ -\operatorname{div}(\Lambda_\phi \nabla\phi) = C + P - N & \text{in } \mathbb{R}_+ \times \Omega \\ N = N^D,\ P = P^D \text{ and } \phi = \phi^D & \text{on } \mathbb{R}_+ \times \Gamma^D \\ N\Lambda_N \nabla(h(N) - \phi) \cdot n = P\Lambda_P \nabla(h(P) + \phi) \cdot n = \Lambda_\phi \nabla\phi \cdot n = 0 & \text{on } \mathbb{R}_+ \times \Gamma^N \\ N(0,\cdot) = N^{in} \text{ and } P(0,\cdot) = P^{in} & \text{in } \Omega, \end{cases} \tag{2.1}$$

where n denotes the unit normal vector to $\partial\Omega$ pointing outward Ω.

Let us give some insight and explanations about this system, and make precise assumptions on the data. First, the unknowns N and P refer respectively to the densities of electrons and holes, and ϕ refers to the electrostatic potential. The densities N and P take values in the set of admissible densities $I_h =]0, a[$, where $a \in]0, +\infty]$ is the upper bound on the density. I_h is in fact the definition domain of the function $h : I_h \to \mathbb{R}$, which depends on the statistics used to describe the relation between the densities and the chemical potential. We assume that h is a $\mathcal{C}^1$ function, such that $h' > 0$ on I_h, with limits $\lim_0 h = -\infty$ and $\lim_a h = +\infty$. We denote by $g = h^{-1} : \mathbb{R} \to I_h$ the inverse function of h. Note that g is positive, and that $g' > 0$ on $\mathbb{R}$. Moreover, we assume that there exists $g_0 > 0$ such that for any real s, $g'(s) \le g_0 g(s)$. We refer the reader to [144, 125, 127] for more details about these statistics functions, but let us emphasise here three classical cases that fall under the scope of our assumptions:

(i) the Boltzmann statistics, for which $g = \exp$, $a = +\infty$ and $h = \log$;

(ii) the Blakemore statistics, for which $g(s) = (\gamma + \mathrm{e}^{-s})^{-1}$, $a = \frac{1}{\gamma}$ and $h(s) = \log(s/(1 - \gamma s))$, with $\gamma > 0$ (a relevant physical value is $\gamma = 0.27$);

(iii) the Fermi-Dirac statistics of order 1/2, for which $g(s) = \frac{2}{\sqrt{\pi}} \int_0^{+\infty} \frac{\sqrt{z}}{1+\mathrm{e}^{z-s}} dz$ and $a = +\infty$.

The statistics (ii) can be interpreted as an approximation of (iii) in low density situations (see the seminal paper of Blakemore [31]), and similarly (ii) can be approximated by (i) if the carrier densities are small enough. For the purpose of future analysis, we define $H : x \to \int_1^x h(s)ds$ and $G : x \mapsto \int_{-\infty}^x g(s)ds$. Notice that H and G are strictly convex, and that G is positive.
The function $C \in L^\infty(\Omega)$ is the doping profile of the semiconductor, which characterises the device under study. In practice, C is a discontinuous function.
The term $R(N,P)$ corresponds to the recombination-generation rate, which can be interpreted as a reaction term between electrons and holes. We assume that this term has the following form:

$$R(N,P) = r(N,P)\left(\mathrm{e}^{h(N)+h(P)} - 1\right), \tag{2.2}$$

where r is a continuous non-negative function, which can be nonconstant in space. Pertinent choices of r are for example (see [207, 208, 128]) the Auger recombination $r(N,P) = c_N N + c_P P$, the Shokley-Read-Hall (SRH) term $r(N,P) = (\tau_N N + \tau_P P + \tau_C)^{-1}$, or no recombination term $r = 0$.
The tensor $\Lambda_\phi \in L^\infty(\Omega, \mathbb{R}^{d\times d})$ is the (rescaled) permittivity of the medium. We assume that it is a symmetric and uniformly elliptic tensor, in the sense that there exists $\lambda_\sharp^\phi \ge \lambda_\flat^\phi > 0$ such that $\forall \xi \in \mathbb{R}^d$,

$$\lambda_\flat^\phi |\xi|^2 \le \xi \cdot \Lambda_\phi \xi \le \lambda_\sharp^\phi |\xi|^2,$$

In practical situations, the permittivity is often assumed isotropic, leading to a tensor of the

form $\Lambda_\phi = \lambda^2 \epsilon I_d$, where λ is the rescaled Debye length of the system, which accounts for the nondimensionalisation, and $\epsilon : \Omega \to R_+^*$ is a uniformly positive function corresponding to the dielectric permittivity of the material. Note that relevant values of the Debye length can be very small, inducing some stiff behaviours. Moreover, since the devices are often made of different materials, the function ϵ can be non-regular, and exhibits discontinuities at junctions between different materials.
The tensors Λ_N and Λ_P are diffusion tensors in $L^\infty(\Omega, \mathbb{R}^{d\times d})$, related to the exterior magnetic field. We refer to [143] for a detailed description of the semiconductor models with magnetic field. A typical example is the case $d = 2$, where the magnetic field $\vec{B}$ is orthogonal to the device: letting $b = |\vec{B}|$, the tensors write

$$\Lambda_N = \frac{\mu_N}{1+b^2}\begin{pmatrix} 1 & b \\ -b & 1 \end{pmatrix} \text{ and } \Lambda_P = \frac{\mu_P}{1+b^2}\begin{pmatrix} 1 & -b \\ b & 1 \end{pmatrix},$$

where $\mu_N : \Omega \to \mathbb{R}_+^\star$ and $\mu_P : \Omega \to \mathbb{R}_+^\star$ are the rescaled mobilities of the electrons and holes. For our analysis, we will consider general tensors, and only assume that the tensors are uniformly elliptic and bounded, in the sense that there exist $\lambda_\sharp \geq \lambda_\flat > 0$ such that for any $\xi \in \mathbb{R}^d$,

$$\lambda_\flat |\xi|^2 \leq \xi \cdot \Lambda_N \xi,\ \lambda_\flat |\xi|^2 \leq \xi \cdot \Lambda_P \xi \text{ and } |\Lambda_N \xi| \leq \lambda_\sharp |\xi|, |\Lambda_P \xi| \leq \lambda_\sharp |\xi|.$$

Note that these tensors are nonsymmetric in general (if $b \neq 0$).
Concerning the boundary data, we consider mixed Dirichlet-Neumann conditions. On Γ^D, which corresponds to the ohmic contacts, we impose that the values of the densities and the electrostatic potential are equal to N^D, P^D and ϕ^D which are the traces of H^1 functions, denoted the same way. On Γ^N, which corresponds to insulated contacts, we impose a null flux. Concerning the initial conditions, we impose the initial densities N^{in} and P^{in}, which are in $L^\infty(\Omega)$. We assume that the initial and boundary data are uniformly bounded in I_h: there exist $0 < m < M < a$ such that, almost everywhere on Ω,

$$m \leq N^{in}, P^{in}, N^D, P^D \leq M. \tag{2.3}$$

In the sequel, we will use the generic term of "data" or "physical data" to refer to the objects previously introduced, namely h, r, Λ_ϕ, Λ_N, Λ_P, C, N^D, P^D, ϕ^D, N^{in} and P^{in}.

The existence and uniqueness of the global solutions to the drift-diffusion system (2.1) have been originally investigated in the case of Boltzmann statistics ($h = \log$), see for example [214, 142, 208]. Concerning models with general statistics described above, we refer to [144, 149]. One of the main properties of the solutions is the fact that the carrier densities N and P take values in I_h: in the sequel we will say that the densities are I_h-valued. Another natural concern lies in the long-time behaviour of the system, that is to say the asymptotics of the solutions when $t \to +\infty$: under some assumptions on the data, the solutions are shown to converge exponentially fast towards some steady-state called thermal equilibrium. The thermal equilibrium is a specific steady-state (N^e, P^e, ϕ^e) for which the electrons and holes currents cancel. Under the hypothesis on the function h, according to [209], there is no vacuum in this state, which hence satisfies $\nabla(h(N^e) - \phi^e) = 0$ and $\nabla(h(P^e) + \phi^e) = 0$ on Ω. Therefore, there exist some constants α_N and α_P such that $h(N^e) - \phi^e = \alpha_N$ and $h(P^e) + \phi^e = \alpha_P$ in Ω (recall that Ω is connected). In particular, this implies the following compatibility condition on the boundary data:

$$h(N^D) - \phi^D = \alpha_N \text{ and } h(P^D) + \phi^D = \alpha_P \text{ on } \Gamma^D. \tag{2.4}$$

Moreover, at the thermal equilibrium, there is no more recombination nor generation, that is to

say $R(N^e, P^e) = 0$. This implies, according to the definition of R, that $h(N^e) + h(P^e) = 0$ if $r \neq 0$. Combined with the previous constraint, this imposes another compatibility condition on the constants if $r \neq 0$:

$$\alpha_N + \alpha_P = 0. \tag{2.5}$$

Under these conditions, we get that $N^e = g(\alpha_N + \phi^e)$ and $P^e = g(\alpha_P - \phi^e)$. Thus, the electrostatic potential at the equilibrium ϕ^e is a solution to the following nonlinear Poisson equation (called Poisson-Boltzmann in case of Boltzmann statistics, i.e. if $g = \exp$):

$$\begin{cases} -\operatorname{div}(\Lambda_\phi \nabla \phi^e) = C + g(\alpha_P - \phi^e) - g(\alpha_N + \phi^e) & \text{in } \Omega, \\ \phi^e = \phi^D & \text{on } \Gamma^D, \\ \Lambda_\phi \nabla \phi^e \cdot n = 0 & \text{on } \Gamma^N. \end{cases} \tag{2.6}$$

In the sequel, we will assume that the boundary data are compliant with the thermal equilibrium and satisfy the compatibility conditions (2.4) and (2.5) (if $r \neq 0$). Notice that (2.4) and (2.3) imply that ϕ^D is bounded on Γ^D. Hence, we assume that its lifting on Ω, also denoted by ϕ^D, is in $H^1(\Omega) \cap L^\infty(\Omega)$. The analysis of the long-time behaviour of the drift-diffusion system was carried out in [143] in the case of the Boltzmann statistics with a magnetic field, and in [144] for other statistics (without magnetic field). It relies on the use of the entropy method, which consists in using a Lyapunov functional -the entropy- decaying with time. Such a method was initially used in kinetic theory and then extended to the study of various dissipative systems, as explained in [8, 174]. Let us sketch here the main lines of this method: letting

$$\mathbb{E}(t) = \int_\Omega H(N) - H(N^e) - h(N^e)(N - N^e) + \int_\Omega H(P) - H(P^e) - h(P^e)(P - P^e) + \frac{1}{2} \int_\Omega \nabla(\phi - \phi^e) \cdot \Lambda_\phi \nabla(\phi - \phi^e),$$

$$\mathbb{D}(t) = \int_\Omega N \nabla(h(N) - \phi) \cdot \Lambda_N \nabla(h(N) - \phi) + \int_\Omega P \nabla(h(P) + \phi) \cdot \Lambda_P \nabla(h(P) + \phi) + \int_\Omega R(N, P)\,(h(N) + h(P)),$$

one has the following entropy-entropy dissipation relation:

$$\frac{d\mathbb{E}}{dt}(t) = -\mathbb{D}(t). \tag{2.7}$$

From the assumptions on R and h, one deduces that the entropy $\mathbb{E}$ and the dissipation $\mathbb{D}$ are non-negative quantities. Therefore, the entropy decreases, and this implies the convergence of $(N(t), P(t), \phi(t))$ towards (N^e, P^e, ϕ^e) when $t \to \infty$.

Note that the system (2.1) can also be interpreted as a Poisson–Nernst–Planck (PNP) system, which describes the evolution of charged particles, typically ions in a solution. Especially, the results presented in this article can be easily generalised to systems with more than two charge carriers.

From a numerical point of view, approximating the solutions to system (2.1) proves to be a challenging task due to various factors, including the nonlinear coupling between the equations, the discontinuity of the doping profile C, the stiffness induced by small values of the Debye

length λ, the anisotropy in the convection-diffusion equations or the non-symmetric nature of the tensors. Moreover, it is essential to design numerical schemes that preserve the qualitative physical properties of the continuous model. Here, the fact that the densities take values in I_h as well as the long-time behaviour are characteristic features of the continuous system that should be preserved at the discrete level. Another challenge lies in the variety of meshes the scheme can handle. Indeed, semiconductor devices are subject to boundary layers phenomena, which often require local mesh refinement to be performed: matching simplicial meshes are strongly constrained, and not very suitable for this purpose (see [207, Chapter 5.4]). Among the various numerical methods that have been proposed for solving drift-diffusion systems, the seminal work of Scharfetter and Gummel [239] in one dimension with Boltzmann statistics presents the keystone idea of numerous schemes: to use the exponential relation between chemical potentials and densities to ensure a preservation of the steady-state as well as a good discrete long-time behaviour. This idea was generalised for different statistics (of the form $h(s) = s^\alpha$, which do not match the assumptions presented above and allow vacuum) in [26] by Bessemoulin-Chatard. This scheme was proved to have a good long-time behaviour in [27, 28], following ideas from Chainais-Hilairet and Filbet introduced in [68] to analyse the long-time behaviour of an upwind TPFA scheme. The analysis strongly relies on the adaptation of the entropy method at the discrete level. The schemes discussed above are essentially part of the two-points flux approximation (TPFA) [121], which suffers from limitations: the mesh has to satisfy some orthogonality conditions, and the diffusion tensor has to be isotropic. In order to overcome these strong constraints, various finite volume methods have been introduced in the framework of the Poisson equation (see [109]), using auxiliary unknowns. Regrettably, these schemes do not enjoy monotonicity properties, and especially they can have negative solutions. In [55, 56], Cancès and Guichard proposed a solution to compensate this limitation, introducing nonlinear schemes based on the entropic structure of the problem. The specific question of approximating drift-diffusion systems on general meshes was already investigated in the past. One can cite the discrete duality finite volume (DDFV, see [157, 105] for a presentation and analysis of these schemes applied to Poisson equations) scheme of Chainais-Hillairet [64], which can handle general polygonal meshes with $d = 2$ in the framework of Boltzmann statistics. More recently, Su and Tang designed and analysed a scheme for Poisson–Nernst–Planck systems on general meshes in [250]. This scheme is based on two discretisation methods: a virtual element method [20] for the Poisson equation alongside with a positive nonlinear finite volume method [32, 48] for the convection-diffusion ones. The scheme is shown to admit solutions with positive densities and to have an entropic structure. However, it is restricted to the Boltzmann statistics with pure homogeneous Neumann boundary conditions and the Debye length is assumed to be 1. Apart from these finite volume schemes, many finite elements schemes were also designed for this type of drift-diffusion systems. Among them, one can cite the mixed and hybrid exponential fitting schemes [42, 43] of Brezzi, Marini and Pietra, which are proposed as finite elements generalisations of the Scharfetter–Gummel scheme [239], for linear diffusion on simplicial meshes. These schemes were then generalised to other statistics, especially in dimension 2, by Jüngel and Pietra in [175]. For other references about these schemes and their extensions, we refer the reader to [41]. Up to our knowledge, the only existing scheme which can handle the presence of a magnetic field in the model (anisotropic tensors Λ_N and Λ_P) is the one introduced and analysed by Gajewski and Gärtner in [143]. It can be seen as a modified Scharfetter–Gummel scheme, and is hence restricted to Boltzmann statistics ($h = \log$). Moreover, the scheme is restricted to triangular meshes, and bound to some strong constraints between the mesh geometry and the magnetic field intensity. Last, the scheme does not preserve the positivity of the densities in the presence of a strong magnetic field.

In this article, we are concerned with the design and the analysis of a numerical scheme

for (2.1) that preserves the continuous features of the system (long-time behaviour and I_h-valuation of the densities). One of our main concerns is the ability of the scheme to handle the large variety of possible data: statistics function h, recombination-generation term R, physical data (doping C, Debye length λ, initial and boundary data, magnetic field) and discretisation data (mesh, time step). Note that among the different motivations, for the support of general meshes is the desire to make local mesh adaptation much simpler in order to capture the boundary layers. Such an approach based on a local adaptation has already been investigated in 1D [58], but should become much more complicated or even impossible to use in higher dimension on constrained meshes. Hence, the scheme presented here could be a possible way to use local mesh refinement in 2D or 3D. To devise such a scheme, we use the hybrid finite volume (HFV) method, introduced in [123] by Eymard et al. in the framework of stationary diffusion problems. This method entails several interesting features: it can handle anisotropic and heterogeneous diffusion tensors alongside with very general polytopal meshes, and it benefits from a unified 1D/2D/3D formulation. The scheme introduced here and its analysis are based on the entropy structure (2.7) of the system, following the ideas of [29, 27] (in the framework of drift-diffusion systems with TPFA schemes) and [56, 52, 70] (in the framework of advection-diffusion equations, with positivity-preserving schemes supporting anisotropy and general meshes). In particular, the nonlinearity induced by the function h is discretised along the principles of the nonlinear HFV scheme introduced and analysed in [70] for advection-diffusion equations. Using this approach, we get a unified scheme for the system (2.1), robust with respect to the various data of the problem. The main result of this paper, stated in Theorem 7, is the existence of solutions (with I_h-valued discrete densities) to the nonlinear scheme. In Theorem 9, we establish the exponential decay of the discrete solutions towards the associated thermal equilibrium. These theoretical results are validated by various numerical simulations.

The article is organised as follows. In Section 2.2, we introduce the HFV framework (mesh, discrete unknowns, discrete operators), present the schemes for the equations (2.1) and a generalisation of (2.6) (namely, the semi-linear Poisson equation (2.30)), and state our main result. In Section 2.3, we show that the stationary scheme is well posed, and discuss an important consequence, namely the correspondence between discrete densities and discrete quasi-Fermi potentials. In Section 2.4, we analyse the scheme for the transient problem, showing that it has an entropy structure and admits solutions with I_h-valued densities whose long-time behaviour mimics that of the continuous solutions. In Section 2.5, we discuss the implementation of the schemes, and give some numerical evidences of our theoretical results. Finally, in Appendix 2.A, we state and prove a technical result that is instrumental to show the existence of solutions.

2.2 Discrete setting and schemes

The aim of this section is to recall the HFV framework for diffusive problems, and introduce the schemes for the discretisation of the drift-diffusion system (2.1) and the nonlinear Poisson equation (2.30). Here, for the purpose of analysis, we present the schemes from the finite element viewpoint. We refer the reader to Section 2.5.1 for a presentation of the schemes within the framework of finite volume methods (especially for the purpose of implementation).
For a detailed presentation of the HFV method in the framework of a steady variable diffusion problem with symmetric tensor, we refer the reader to [123].

2.2.1 Mesh, discrete unknowns and boundary data

The definitions and notation we adopt for the discretisation are essentially the same as in [123]. A discretisation of the (open, bounded, connected) polytopal set $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, is

defined as a triplet $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$, where:

— $\mathcal{M}$ (the mesh) is a partition of Ω, i.e., a finite family of nonempty disjoint (open, connected) polytopal subsets K of Ω (the mesh cells) such that (i) for all $K \in \mathcal{M}$, $|K| > 0$, and (ii) $\overline{\Omega} = \bigcup_{K\in\mathcal{M}} \overline{K}$.

— $\mathcal{E}$ (the set of faces) is a partition of the mesh skeleton $\bigcup_{K\in\mathcal{M}} \partial K$, i.e., a finite family of nonempty disjoint (open, connected) subsets σ of $\overline{\Omega}$ (the mesh faces, or mesh edges if $d = 2$) such that for all $\sigma \in \mathcal{E}$, $|\sigma| > 0$ and there exists $\mathcal{H}_\sigma$ affine hyperplane of $\mathbb{R}^d$ such that $\sigma \subset \mathcal{H}_\sigma$, and $\bigcup_{K\in\mathcal{M}} \partial K = \bigcup_{\sigma\in\mathcal{E}} \overline{\sigma}$. We assume that, for all $K \in \mathcal{M}$, there exists $\mathcal{E}_K \subset \mathcal{E}$ (the set of faces of K) such that $\partial K = \bigcup_{\sigma\in\mathcal{E}_K} \overline{\sigma}$. For $\sigma \in \mathcal{E}$, we let $\mathcal{M}_\sigma = \{K \in \mathcal{M} \mid \sigma \in \mathcal{E}_K\}$ be the set of cells whose σ is a face. Then, for all $\sigma \in \mathcal{E}$, either $\mathcal{M}_\sigma = \{K\}$ for a cell $K \in \mathcal{M}$, in which case σ is a boundary face ($\sigma \subset \partial\Omega$) and we note $\sigma \in \mathcal{E}_{ext}$, or $\mathcal{M}_\sigma = \{K, L\}$ for two cells $K, L \in \mathcal{M}$, in which case σ is an interface and we note $\sigma = K|L \in \mathcal{E}_{int}$.

— $\mathcal{P}$ (the set of cell centres) is a finite family $\{x_K\}_{K\in\mathcal{M}}$ of points of Ω such that, for all $K \in \mathcal{M}$, (i) $x_K \in K$, and (ii) K is star-shaped with respect to x_K. Moreover, we assume that the Euclidean (orthogonal) distance $d_{K,\sigma}$ between x_K and the affine hyperplane $\mathcal{H}_\sigma$ containing σ is positive (equivalently, the cell K is strictly star-shaped with respect to x_K).

For a given discretisation $\mathcal{D}$, we denote by $h_\mathcal{D} > 0$ the size of the discretisation (the meshsize), defined by $h_\mathcal{D} = \sup_{K\in\mathcal{M}} h_K$ where, for all $K \in \mathcal{M}$, $h_K = \sup_{x,y\in\overline{K}} |x-y|$ is the diameter of the cell K. For all $\sigma \in \mathcal{E}$, we let $\overline{x}_\sigma \in \sigma$ be the barycentre of σ. Finally, for all $K \in \mathcal{M}$, and all $\sigma \in \mathcal{E}_K$, we let $n_{K,\sigma} \in \mathbb{R}^d$ be the unit normal vector to σ pointing outward K, and $P_{K,\sigma}$ be the (open) pyramid of base σ and apex x_K (notice that, when $d = 2$, $P_{K,\sigma}$ is always a triangle). Since $|\sigma|$ and $d_{K,\sigma}$ are positive, we have $|P_{K,\sigma}| = \frac{|\sigma| d_{K,\sigma}}{d} > 0$. We depict on Figure 2.1 an example of discretisation. Notice that the mesh cells are not assumed to be convex, neither x_K is assumed to be the barycentre of $K \in \mathcal{M}$. Moreover, hanging nodes are seamlessly handled with our assumptions, so that meshes with non-conforming cells are allowed (see the orange cross in Figure 2.1; the cell K therein is treated as an hexagon).

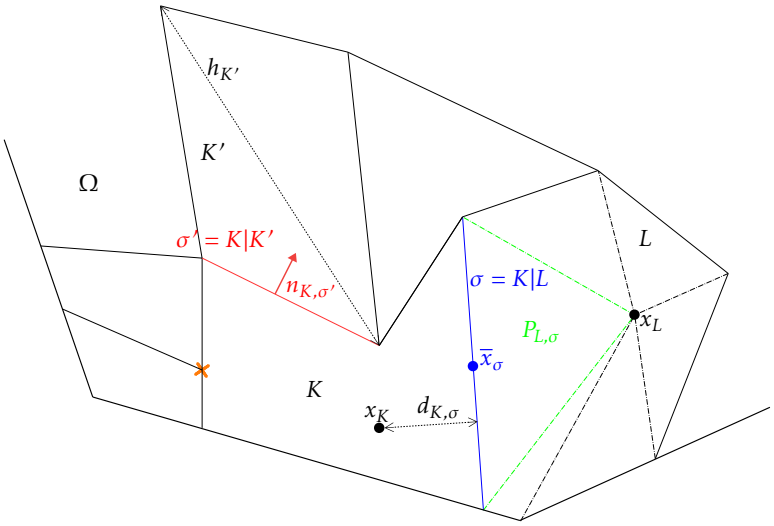


Figure 2.1 – Two-dimensional discretisation and corresponding notations.

We consider the following measure of regularity for the discretisation (which is the same as in [70, Eq. (2.1)]):

$$\theta_{\mathcal{D}} = \max\left(\max_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}\frac{h_K}{d_{K,\sigma}}, \max_{\sigma\in\mathcal{E},K\in\mathcal{M}_\sigma}\frac{h_K^{d-1}}{|\sigma|}\right). \tag{2.8}$$

We now introduce the set of (hybrid, cell- and face-based) discrete unknowns:

$$\underline{V}_{\mathcal{D}} = \left\{\underline{v}_{\mathcal{D}} = \left((v_K)_{K\in\mathcal{M}},(v_\sigma)_{\sigma\in\mathcal{E}}\right) \mid v_K\in\mathbb{R}\ \forall K\in\mathcal{M}, v_\sigma\in\mathbb{R}\ \forall\sigma\in\mathcal{E}\right\}.$$

Given a mesh cell $K\in\mathcal{M}$, we let $\underline{V}_K = \mathbb{R}\times\mathbb{R}^{|\mathcal{E}_K|}$ be the restriction of $\underline{V}_{\mathcal{D}}$ to K, and $\underline{v}_K = \left(v_K,(v_\sigma)_{\sigma\in\mathcal{E}_K}\right)\in\underline{V}_K$ be the restriction of a generic element $\underline{v}_{\mathcal{D}}\in\underline{V}_{\mathcal{D}}$ to K. Also, for $\underline{v}_{\mathcal{D}}\in\underline{V}_{\mathcal{D}}$, we let $v_{\mathcal{M}}:\Omega\to\mathbb{R}$ and $v_{\mathcal{E}}:\bigcup_{K\in\mathcal{M}}\partial K\to\mathbb{R}$ be the piecewise constant functions such that

$$v_{\mathcal{M}|K} = v_K \text{ for all } K\in\mathcal{M}, \quad \text{and} \quad v_{\mathcal{E}|\sigma} = v_\sigma \text{ for all } \sigma\in\mathcal{E}.$$

For further use, we let $\underline{1}_{\mathcal{D}}$ (respectively $\underline{0}_{\mathcal{D}}$) denote the element of $\underline{V}_{\mathcal{D}}$ with all coordinates equal to 1 (respectively 0). Also, given a function $f:\mathbb{R}\to\mathbb{R}$, and with a slight abuse in notation, we denote by $f(\underline{v}_{\mathcal{D}})$ the element of $\underline{V}_{\mathcal{D}}$ whose coordinates are the $(f(v_K))_{K\in\mathcal{M}}$ and the $(f(v_\sigma))_{\sigma\in\mathcal{E}}$. Given $\underline{u}_{\mathcal{D}}$ and $\underline{v}_{\mathcal{D}}$ two elements of $\underline{V}_{\mathcal{D}}$, we say that $\underline{u}_{\mathcal{D}}\le\underline{v}_{\mathcal{D}}$ (respectively $<,\ge,>$) if and only if for any $K\in\mathcal{M}$ and $\sigma\in\mathcal{E}$, $u_K\le v_K$ and $u_\sigma\le v_\sigma$ (respectively $<,\ge,>$). In particular, a vector of unknowns $\underline{v}_{\mathcal{D}}\in\underline{V}_{\mathcal{D}}$ is I_h-valued if and only if $\underline{0}_{\mathcal{D}}<\underline{v}_{\mathcal{D}}<a\underline{1}_{\mathcal{D}}$. Note that I_h-valued vectors have therefore positive coordinates.

We assume that the discretisation $\mathcal{D}$ is compliant with the partition $\partial\Omega = \overline{\Gamma^D}\cup\overline{\Gamma^N}$ of the boundary of the domain, in the sense that the set $\mathcal{E}_{ext}$ can be split into two disjoint subsets $\mathcal{E}_{ext}^D = \left\{\sigma\in\mathcal{E}_{ext}\mid\sigma\subset\Gamma^D\right\}$ and $\mathcal{E}_{ext}^N = \left\{\sigma\in\mathcal{E}_{ext}\mid\sigma\subset\Gamma^N\right\}$ such that $\mathcal{E}_{ext} = \mathcal{E}_{ext}^D\cup\mathcal{E}_{ext}^N$. Notice that, since $|\Gamma^D|>0$, one has $|\mathcal{E}_{ext}^D|\ge 1$. We define the following subspace of $\underline{V}_{\mathcal{D}}$, enforcing strongly a homogeneous Dirichlet boundary condition on Γ^D:

$$\underline{V}_{\mathcal{D},0} = \left\{\underline{v}_{\mathcal{D}}\in\underline{V}_{\mathcal{D}}\mid\forall\sigma\in\mathcal{E}_{ext}^D,\ v_\sigma = 0\right\}.$$

In view of the upcoming analysis, we define a discrete counterpart of the H^1 semi-norm. Locally to any cell $K\in\mathcal{M}$, we let, for any $\underline{v}_K\in\underline{V}_K$, $|\underline{v}_K|_{1,K}^2 = \sum_{\sigma\in\mathcal{E}_K}\frac{|\sigma|}{d_{K,\sigma}}(v_K-v_\sigma)^2$. At the global level, for any $\underline{v}_{\mathcal{D}}\in\underline{V}_{\mathcal{D}}$, we let

$$|\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}} = \sqrt{\sum_{K\in\mathcal{M}}|\underline{v}_K|_{1,K}^2}.$$

Notice that $|\cdot|_{1,\mathcal{D}}$ does not define a norm on $\underline{V}_{\mathcal{D}}$, but if $|\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}=0$, then there is $c\in\mathbb{R}$ such that $\underline{v}_{\mathcal{D}} = c\,\underline{1}_{\mathcal{D}}$ ($\underline{v}_{\mathcal{D}}$ is constant). Thus, $|\cdot|_{1,\mathcal{D}}$ defines a norm on the space $\underline{V}_{\mathcal{D},0}$.

We finally introduce discrete hybrid (Dirichlet) boundary data. The definitions below are motivated by the need to preserve the compatibility condition (2.4) at the discrete level. First, we define the discrete liftings of the densities $\underline{N}_{\mathcal{D}}^D$ and $\underline{P}_{\mathcal{D}}^D$ as the elements of $\underline{V}_{\mathcal{D}}$ such that for any $K\in\mathcal{M}$ and any $\sigma\in\mathcal{E}$,

$$N_\sigma^D = g\left(\frac{1}{|\sigma|}\int_\sigma h(N^D)\right) \text{ and } N_K^D = g\left(\frac{1}{|K|}\int_K h(N^D)\right), \tag{2.9}$$

$$P_\sigma^D = g\left(\frac{1}{|\sigma|}\int_\sigma h(P^D)\right) \text{ and } P_K^D = g\left(\frac{1}{|K|}\int_K h(P^D)\right). \tag{2.10}$$

We also define the hybrid interpolate $\underline{\phi}_{\mathcal{D}}^D \in \underline{V}_{\mathcal{D}}$ of the lifting ϕ^D, defined by $\phi_K^D = \frac{1}{|K|}\int_K \phi^D$ for any $K \in \mathcal{M}$ and $\phi_\sigma^D = \frac{1}{|\sigma|}\int_\sigma \phi^D$ for all $\sigma \in \mathcal{E}$. One has (see [111, Proposition B.7]) the following stability results:

$$|\underline{\phi}_{\mathcal{D}}^D|_{1,\mathcal{D}} \le C_{l,\Gamma^D}\|\phi^D\|_{H^{1/2}(\Gamma^D)} \text{ and } -\|\phi^D\|_{L^\infty(\Omega)}\underline{1}_{\mathcal{D}} \le \underline{\phi}_{\mathcal{D}}^D \le \|\phi^D\|_{L^\infty(\Omega)}\underline{1}_{\mathcal{D}}, \tag{2.11}$$

where C_{l,Γ^D} is a positive constant only depending on $\theta_{\mathcal{D}}$, Ω and Γ^D. By (2.4) and (2.9)-(2.10), the following compatibility condition, discrete counterpart of (2.4), holds:

$$\forall \sigma \in \mathcal{E}_{ext}^D,\ h(N_\sigma^D) - \phi_\sigma^D = \alpha_N \text{ and } h(P_\sigma^D) + \phi_\sigma^D = \alpha_P. \tag{2.12}$$

Remark 10 (Discrete liftings). *The definition of the discrete boundary densities (2.9) and (2.10) is driven by the desire to obtain the discrete compatibility condition (2.12). Such definition boils down to discretise quantities carrying a physical meaning, homogeneous to the quasi-Fermi potentials. Moreover, this definition ensures that $\underline{N}_{\mathcal{D}}^D$ and $\underline{P}_{\mathcal{D}}^D$ are I_h-valued. Notice that in practice, if N^D is regular enough, one can choose to approximate $\frac{1}{|\sigma|}\int_\sigma h(N^D)$ by $h(N^D(\overline{x}_\sigma))$ (respectively, $\frac{1}{|K|}\int_K h(N^D)$ by $h(N^D(x_K))$ if x_K is chosen to be the barycenter of K), and therefore find a classical expression for the discrete liftings, namely $N_\sigma^D = N^D(\overline{x}_\sigma)$ and $N_K^D = N^D(x_K)$. The same holds true for $\underline{P}_{\mathcal{D}}^D$.*

2.2.2 Foundations of the hybrid finite volume method

The HFV method hinges on the definition of a discrete gradient operator $\nabla_{\mathcal{D}}$, that maps any element $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ to a piecewise constant $\mathbb{R}^d$-valued function on the pyramidal submesh of $\mathcal{M}$ formed by all the $P_{K,\sigma}$'s, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$. More precisely, for all $K \in \mathcal{M}$, and all $\sigma \in \mathcal{E}_K$,

$$\nabla_{\mathcal{D}}\underline{v}_{\mathcal{D}|K} = \nabla_K \underline{v}_K \quad \text{with} \quad \nabla_K \underline{v}_{K|P_{K,\sigma}} = \nabla_{K,\sigma}\underline{v}_K = G_K\underline{v}_K + S_{K,\sigma}\underline{v}_K \in \mathbb{R}^d,$$

where $G_K\underline{v}_K$ is the consistent part of the gradient given by

$$G_K\underline{v}_K = \frac{1}{|K|}\sum_{\sigma' \in \mathcal{E}_K} |\sigma'|(v_{\sigma'} - v_K)n_{K,\sigma'} = \frac{1}{|K|}\sum_{\sigma' \in \mathcal{E}_K} |\sigma'|v_{\sigma'}n_{K,\sigma'},$$

and $S_{K,\sigma}\underline{v}_K$ is a stabilisation, given, for some parameter $\eta > 0$, by

$$S_{K,\sigma}\underline{v}_K = \frac{\eta}{d_{K,\sigma}}\Big(v_\sigma - v_K - G_K\underline{v}_K \cdot (\overline{x}_\sigma - x_K)\Big)n_{K,\sigma}. \tag{2.13}$$

Let us consider a generic tensor $\Lambda \in L^\infty(\Omega, \mathbb{R}^{d\times d})$, such that for any $\xi \in \mathbb{R}^d$, $\lambda_\flat|\xi|^2 \le \xi \cdot \Lambda\xi$ and $|\Lambda\xi| \le \lambda_\sharp|\xi|$. Locally to any cell $K \in \mathcal{M}$, we introduce the discrete bilinear form $a_K^\Lambda : \underline{V}_K \times \underline{V}_K \to \mathbb{R}$ such that, for all $\underline{u}_K, \underline{v}_K \in \underline{V}_K$,

$$a_K^\Lambda(\underline{u}_K, \underline{v}_K) = \sum_{\sigma \in \mathcal{E}_K} |P_{K,\sigma}|\nabla_{K,\sigma}\underline{v}_K \cdot \Lambda_{K,\sigma}\nabla_{K,\sigma}\underline{u}_K = \int_K \nabla_K\underline{v}_K \cdot \Lambda\nabla_K\underline{u}_K \tag{2.14}$$

where we set $\Lambda_{K,\sigma} = \frac{1}{|P_{K,\sigma}|}\int_{P_{K,\sigma}} \Lambda$. At the global level, we let $a_{\mathcal{D}}^{\Lambda} : \underline{V}_{\mathcal{D}} \times \underline{V}_{\mathcal{D}} \to \mathbb{R}$ be the discrete bilinear form such that, for all $\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$,

$$a_{\mathcal{D}}^{\Lambda}(\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = \sum_{K\in\mathcal{M}} a_K^{\Lambda}(\underline{u}_K, \underline{v}_K) = \int_\Omega \nabla_{\mathcal{D}}\underline{v}_{\mathcal{D}} \cdot \Lambda \nabla_{\mathcal{D}}\underline{u}_{\mathcal{D}}.$$

As for the continuous case, the well-posedness of HFV methods for diffusion problems relies on a coercivity argument. Let $K \in \mathcal{M}$, and reason locally. By definition (2.14) of the local discrete bilinear form a_K^{Λ}, and from the bounds on the diffusion coefficient, we have $\lambda_\flat \|\nabla_K \underline{v}_K\|^2_{L^2(K;\mathbb{R}^d)} \le a_K^{\Lambda}(\underline{v}_K, \underline{v}_K) \le \lambda_\sharp \|\nabla_K \underline{v}_K\|^2_{L^2(K;\mathbb{R}^d)}$ for all $\underline{v}_K \in \underline{V}_K$. Furthermore, the following comparison result holds (cf. [111, Lemma 13.11]): there exist $\alpha_\flat, \alpha_\sharp$ with $0 < \alpha_\flat \le \alpha_\sharp < \infty$, only depending on Ω, d, and $\theta_{\mathcal{D}}$ such that $\alpha_\flat |\underline{v}_K|^2_{1,K} \le \|\nabla_K \underline{v}_K\|^2_{L^2(K;\mathbb{R}^d)} \le \alpha_\sharp |\underline{v}_K|^2_{1,K}$ for all $\underline{v}_K \in \underline{V}_K$. Combining both estimates, we infer a local coercivity and boundedness result:

$$\forall \underline{v}_K \in \underline{V}_K, \qquad \lambda_\flat \alpha_\flat |\underline{v}_K|^2_{1,K} \le a_K^{\Lambda}(\underline{v}_K, \underline{v}_K) \le \lambda_\sharp \alpha_\sharp |\underline{v}_K|^2_{1,K}. \tag{2.15}$$

Summing over $K \in \mathcal{M}$, we get the following global estimates:

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}, \qquad \lambda_\flat \alpha_\flat |\underline{v}_{\mathcal{D}}|^2_{1,\mathcal{D}} \le a_{\mathcal{D}}^{\Lambda}(\underline{v}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \le \lambda_\sharp \alpha_\sharp |\underline{v}_{\mathcal{D}}|^2_{1,\mathcal{D}}. \tag{2.16}$$

Last, we recall the hybrid discrete Poincaré inequality (cf. [111, Lemma B.32, $p = 2$]): there exists $C_{P,\Gamma^D} > 0$, only depending on Ω, d, Γ^D, and $\theta_{\mathcal{D}}$ such that

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, \qquad \|v_{\mathcal{M}}\|_{L^2(\Omega)} \le C_{P,\Gamma^D} |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}. \tag{2.17}$$

In the sequel, we will omit the tensor in superscript and use the following notations instead:

$$a_{\mathcal{D}}^{\phi} = a_{\mathcal{D}}^{\Lambda_\phi}, \ a_{\mathcal{D}}^{P} = a_{\mathcal{D}}^{\Lambda_P} \text{ and } a_{\mathcal{D}}^{N} = a_{\mathcal{D}}^{\Lambda_N}. \tag{2.18}$$

One can note that, since Λ_ϕ is symmetric, one has

$$\forall (\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \in \underline{V}^2_{\mathcal{D},0}, \ |a_{\mathcal{D}}^{\phi}(u_{\mathcal{D}}, \underline{v}_{\mathcal{D}})| \le \alpha_\sharp \lambda_\sharp^{\phi} |\underline{u}_{\mathcal{D}}|_{1,\mathcal{D}} |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}. \tag{2.19}$$

For further use, we introduce the linear form $L_{\mathcal{D}}^{\phi} : \underline{V}_{\mathcal{D}} \to \mathbb{R}$ such that

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}, \ L_{\mathcal{D}}^{\phi}(\underline{v}_D) = \int_\Omega C v_{\mathcal{M}} - a_{\mathcal{D}}^{\phi}(\underline{\phi}_{\mathcal{D}}^D, \underline{v}_{\mathcal{D}}). \tag{2.20}$$

Note that, according to (2.19), the Poincaré inequality (2.17) and the estimates (2.11), one has

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, \ |L_{\mathcal{D}}^{\phi}(\underline{v}_D)| \le \|C\|_{L^2(\Omega)} \|v_{\mathcal{M}}\|_{L^2(\Omega)} + \alpha_\sharp \lambda_\sharp |\underline{\phi}_{\mathcal{D}}^D|_{1,\mathcal{D}} |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}} \le L_\sharp^{\phi} \ |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}, \tag{2.21}$$

where there exists $C_L > 0$, only depending on Ω, Γ^D, d, Λ_ϕ and $\theta_{\mathcal{D}}$ such that

$$L_\sharp^{\phi} \le C_L \Big(\|C\|_{L^2(\Omega)} + \|\phi^D\|_{H^{1/2}(\Gamma^D)} \Big).$$

2.2.3 Description of the schemes and main theorem

In this section, we introduce an HFV scheme for Problem (2.1). The scheme is designed so as to preserve the entropic structure presented in the introduction. In the sequel, we consider a fixed spatial discretisation $\mathcal{D}$ of Ω which satisfies the assumptions of Section 2.2.1, and a fixed time step $\Delta t > 0$ for the time dicretisation. For any $n \in \mathbb{N}$, we let $t^n = n\Delta t$.

Our HFV scheme relies on the hybrid nonlinear discretisation introduced in [70] for linear advection-diffusion equations, which was in turn inspired from the VAG and DDFV schemes of [56, 52]. We refer to [70] for a detailed description of the discretisation (with $h = \log$). The common idea of these schemes is to discretise the flux in a nonlinear way: formally, given a function u with values in I_h and a function ϕ, the problem $\operatorname{div}(-u\Lambda\nabla(h(u)+\phi)) = 0$ can be expressed in term of the flux

$$J = -u\Lambda\nabla(h(u)+\phi) = -u\Lambda\nabla w,$$

where $w = h(u)+\phi$ is a quasi-Fermi potential. Therefore, locally to any cell $K \in \mathcal{M}$, we define an approximation of

$$(u, w, v) \mapsto -\int_K J\cdot\nabla v = \int_K u\,\Lambda\nabla w\cdot\nabla v$$

under the form

$$T_K^\Lambda(\underline{u}_K, \underline{w}_K, \underline{v}_K) = \int_K r_K(\underline{u}_K)\,\Lambda\nabla_K\underline{w}_K\cdot\nabla_K\underline{v}_K$$

for all I_h-valued $\underline{u}_K \in \underline{V}_K$ and all $(\underline{w}_K, \underline{v}_K) \in \underline{V}_K^2$, where $r_K : \underline{V}_K \to I_h$, called local reconstruction operator, maps I_h-valued elements of $\underline{V}_K$ to I_h. Since $r_K(\underline{u}_K)$ is a (positive) constant on K, we have

$$T_K^\Lambda(\underline{u}_K, \underline{w}_K, \underline{v}_K) = r_K(\underline{u}_K)a_K^\Lambda\big(\underline{w}_K, \underline{v}_K\big), \tag{2.22}$$

where a_K^Λ is defined by (2.14). As already pointed out in previous works on similar nonlinear schemes [49, 52, 51, 70], the definition of the local reconstruction operator is crucial to guarantee the existence of solutions as well as the long-time behaviour of the discrete solutions. Especially, we use a reconstruction which embeds information from both the local cell and face unknowns. It is a natural generalisation of the one introduced in [70]. For $\underline{u}_K$ an I_h-valued element of $\underline{V}_K$, we let

$$r_K(\underline{u}_K) = \frac{1}{|\mathcal{E}_K|}\sum_{\sigma\in\mathcal{E}_K} m(u_K, u_\sigma) \tag{2.23}$$

where $m : I_h^2 \to I_h$ is a continuous (mean-)function satisfying

$$\forall (x,y)\in I_h^2,\ m_h(x,y) \le m(x,y) \le \max(x,y), \tag{2.24}$$

and the function m_h is defined by

$$m_h(x,y) = \frac{[xh(x)-H(x)]-[yh(y)-H(y)]}{h(x)-h(y)} \text{ if } x\neq y \text{ and } m_h(x,x) = x. \tag{2.25}$$

This function m_h was used in [131, Definition 3.3] to analyse an entropic scheme for nonlinear advection-diffusion equations. One can note that in the case of the Boltzmann statistics ($h = \log$), $m_h(x,y)$ is the classical logarithmic average of x and y, which coincides with the assumptions of [70]. Moreover, for any $(x,y)\in\mathbb{R}^2$, one has $m_h(g(x), g(y)) = \frac{xg(x)-H(g(x))-[yg(y)-H(g(y))]}{x-y}$. Since the

derivative of $x \mapsto xg(x) - H(g(x))$ is g, the following simple expression holds:

$$\forall (x,y) \in \mathbb{R}^2,\ m_h(g(x),g(y)) = \frac{G(x)-G(y)}{x-y}. \tag{2.26}$$

Note that, for all $(x,y) \in I_h^2$, one has $m_h(x,y) \le \frac{x+y}{2} \le \max(x,y)$, and each expression of the previous sequence is a mean function m satisfying (2.24). Heuristically, $r_K(\underline{u}_K)$ computes an average of the unknowns attached to the cell K, especially it contains information about all the local face unknowns. The property (2.24) will be instrumental to prove Lemma 7 and the existence of solutions to the scheme. Finally, we let $T_{\mathcal{D}}^{\Lambda}$ be such that, for all I_h-valued $\underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, and all $(\underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D}}^2$,

$$T_{\mathcal{D}}^{\Lambda}(\underline{u}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = \sum_{K\in\mathcal{M}} T_K^{\Lambda}(\underline{u}_K, \underline{w}_K, \underline{v}_K), \tag{2.27}$$

where the local contributions T_K^{Λ} are defined by (2.22). In the sequel, we will use the following notation

$$T_{\mathcal{D}}^N = T_{\mathcal{D}}^{\Lambda_N} \text{ and } T_{\mathcal{D}}^P = T_{\mathcal{D}}^{\Lambda_P} = (resp.,\ T_K^N = T_K^{\Lambda_N} \text{ and } T_K^P = T_K^{\Lambda_P}).$$

Using a backward Euler discretisation in time, and the nonlinear HFV discretisation previously described in space, we consider the following scheme for the drift-diffusion system (2.1): find $\left((\underline{N}_{\mathcal{D}}^n, \underline{P}_{\mathcal{D}}^n, \underline{\phi}_{\mathcal{D}}^n)\right)_{n\in\mathbb{N}} \in \underline{V}_{\mathcal{D}}^{3\,\mathbb{N}}$ such that

$$\begin{cases}
\underline{w}_{\mathcal{D}}^{N,n+1} = h\left(\underline{N}_{\mathcal{D}}^{n+1}\right) - \underline{\phi}_{\mathcal{D}}^{n+1} - \alpha_N \underline{1}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, & \text{(2.28a)}\\
\underline{w}_{\mathcal{D}}^{P,n+1} = h\left(\underline{P}_{\mathcal{D}}^{n+1}\right) + \underline{\phi}_{\mathcal{D}}^{n+1} - \alpha_P \underline{1}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, & \text{(2.28b)}\\
\underline{\psi}_{\mathcal{D}}^{n+1} = \underline{\phi}_{\mathcal{D}}^{n+1} - \underline{\phi}_{\mathcal{D}}^{D} \in \underline{V}_{\mathcal{D},0}, & \text{(2.28c)}\\
\displaystyle\int_\Omega \frac{N_{\mathcal{M}}^{n+1} - N_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}} + T_{\mathcal{D}}^N(\underline{N}_{\mathcal{D}}^{n+1}, \underline{w}_{\mathcal{D}}^{N,n+1}, \underline{v}_{\mathcal{D}}) = -\int_\Omega R(N_{\mathcal{M}}^{n+1}, P_{\mathcal{M}}^{n+1}) v_{\mathcal{M}} \quad \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, & \text{(2.28d)}\\
\displaystyle\int_\Omega \frac{P_{\mathcal{M}}^{n+1} - P_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}} + T_{\mathcal{D}}^P(\underline{P}_{\mathcal{D}}^{n+1}, \underline{w}_{\mathcal{D}}^{P,n+1}, \underline{v}_{\mathcal{D}}) = -\int_\Omega R(N_{\mathcal{M}}^{n+1}, P_{\mathcal{M}}^{n+1}) v_{\mathcal{M}} \quad \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, & \text{(2.28e)}\\
a_{\mathcal{D}}^{\phi}(\underline{\psi}_{\mathcal{D}}^{n+1}, \underline{v}_{\mathcal{D}}) - L_{\mathcal{D}}^{\phi}(\underline{v}_{\mathcal{D}}) = \displaystyle\int_\Omega \left(P_{\mathcal{M}}^{n+1} - N_{\mathcal{M}}^{n+1}\right) v_{\mathcal{M}} \quad \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}, & \text{(2.28f)}\\
N_K^0 = \dfrac{1}{|K|}\displaystyle\int_K N^{in} \text{ and } P_K^0 = \dfrac{1}{|K|}\int_K P^{in} \quad \forall K \in \mathcal{M}. & \text{(2.28g)}
\end{cases}$$

Notice that the equations (2.28a) and (2.28b) implicitly assume that, for any $n \ge 1$, the potentials $h\left(\underline{N}_{\mathcal{D}}^n\right)$ and $h\left(\underline{P}_{\mathcal{D}}^n\right)$ are defined, meaning that the discrete densities $\underline{N}_{\mathcal{D}}^n$ and $\underline{P}_{\mathcal{D}}^n$ are I_h-valued. Therefore, a solution to (2.28) has I_h-valued discrete densities by definition, which can be expressed as functions of the discrete quasi-Fermi potentials: for any $n \ge 1$,

$$\underline{N}_{\mathcal{D}}^n = g\left(\underline{w}_{\mathcal{D}}^{N,n} + \underline{\phi}_{\mathcal{D}}^n + \alpha_N \underline{1}_{\mathcal{D}}\right) \text{ and } \underline{P}_{\mathcal{D}}^n = g\left(\underline{w}_{\mathcal{D}}^{P,n} - \underline{\phi}_{\mathcal{D}}^n + \alpha_P \underline{1}_{\mathcal{D}}\right). \tag{2.29}$$

Remark 11 (Discrete potentials and boundary values)**.** *The definitions of the discrete quasi-Fermi potentials* (2.28a) *and* (2.28b) *are used to enforce the boundary conditions on the discrete unknowns. Indeed, since* $\underline{w}_{\mathcal{D}}^{N,n} \in \underline{V}_{\mathcal{D},0}$*, from* (2.29) *we get that for any* $\sigma \in \mathcal{E}_{ext}^D$*,* $N_\sigma^n = g\left(0 + \phi_\sigma^D + \alpha_N\right) = N_\sigma^D$

according to the compatibility condition (2.12). *A similar result holds for* $\underline{P}^n_{\mathcal{D}}$. *In fact, by definition of the trilinear forms, one can rewrite in a more natural way the equations* (2.28d) *and* (2.28e) *noticing that*

$$T^N_{\mathcal{D}}(\underline{N}^{n+1}_{\mathcal{D}}, \underline{w}^{N,n+1}_{\mathcal{D}}, \underline{v}^N_{\mathcal{D}}) = T^N_{\mathcal{D}}(\underline{N}^{n+1}_{\mathcal{D}}, h\left(\underline{N}^{n+1}_{\mathcal{D}}\right) - \underline{\phi}^{n+1}_{\mathcal{D}}, \underline{v}^N_{\mathcal{D}}), \text{ and}$$
$$T^P_{\mathcal{D}}(\underline{P}^{n+1}_{\mathcal{D}}, \underline{w}^{P,n+1}_{\mathcal{D}}, \underline{v}^P_{\mathcal{D}}) = T^P_{\mathcal{D}}(\underline{P}^{n+1}_{\mathcal{D}}, h\left(\underline{P}^{n+1}_{\mathcal{D}}\right) + \underline{\phi}^{n+1}_{\mathcal{D}}, \underline{v}^P_{\mathcal{D}}).$$

When considering general boundary conditions (i.e., if α_N *and* α_P *are no longer constant on* Γ^D, *and therefore with liftings no longer constant in* Ω), *the scheme will be the same, except for the equations* (2.28d) *and* (2.28e) *where arguments of the trilinear forms will be*

$$T^N_{\mathcal{D}}(\underline{N}^{n+1}_{\mathcal{D}}, \underline{w}^{N,n+1}_{\mathcal{D}} + \underline{\alpha_N}_{\mathcal{D}}, \underline{v}^N_{\mathcal{D}}) \text{ and } T^P_{\mathcal{D}}(\underline{P}^{n+1}_{\mathcal{D}}, \underline{w}^{P,n+1}_{\mathcal{D}} + \underline{\alpha_P}_{\mathcal{D}}, \underline{v}^P_{\mathcal{D}}).$$

The analysis of such a scheme is more sophisticated than the one presented here, and shall be the subject of a future work.

Plugging the expression of the discrete carrier densities (2.29) into Equation (2.28f), we get an elliptic semi-linear equation on $\underline{\psi}^{n+1}_{\mathcal{D}}$. This stationary equation gives a correspondence between densities and potentials at a given time, and will be of great importance in the analysis. At the continuous level, it corresponds to the following semi-linear Poisson equation:

$$\begin{cases} -\operatorname{div}\left(\Lambda_\phi \nabla\phi\right) = C + g(z^P - \phi) - g(z^N + \phi) & \text{in } \Omega, \\ \phi = \phi^D & \text{on } \Gamma^D, \\ \Lambda_\phi \nabla\phi \cdot n = 0 & \text{on } \Gamma^N, \end{cases} \tag{2.30}$$

where z^P and z^N are given functions in $L^\infty(\Omega)$.

Remark 12 (Electrostatic potential at thermal equilibrium). *If* z^P *and* z^N *are constant with respective values* α_P *and* α_N *in* Ω, *then the solution to* (2.30) *is the electrostatic potential at thermal equilibrium* ϕ^e *solution to* (2.6).

We introduce an HFV scheme to approximate (2.30): find $\underline{\phi}_{\mathcal{D}} = \underline{\psi}_{\mathcal{D}} + \underline{\phi}^D_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$ where $\underline{\psi}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}$ is such that

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0},\ a^\phi_{\mathcal{D}}(\underline{\psi}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = L^\phi_{\mathcal{D}}(\underline{v}_{\mathcal{D}}) + \int_\Omega \left[g(z^P_{\mathcal{M}} - \phi^D_{\mathcal{M}} - \psi_{\mathcal{M}}) - g(z^N_{\mathcal{M}} + \phi^D_{\mathcal{M}} + \psi_{\mathcal{M}})\right] v_{\mathcal{M}}, \tag{2.31}$$

where $z^P_{\mathcal{M}}$ and $z^N_{\mathcal{M}}$ are the piecewise constant functions such that, for any cell $K \in \mathcal{M}$, $z^P_{\mathcal{M}} = \frac{1}{|K|}\int_K z^P$ and $z^N_{\mathcal{M}} = \frac{1}{|K|}\int_K z^N$ on K. We will prove in Theorem 8 that this scheme is well-posed. Especially, in the light of Remark 12, one can define the discrete thermal equilibrium as a triplet $(\underline{N}^e_{\mathcal{D}}, \underline{P}^e_{\mathcal{D}}, \underline{\phi}^e_{\mathcal{D}}) \in \underline{V}^3_{\mathcal{D}}$ of discrete unknowns satisfying:

$$\begin{cases} \underline{\phi}^e_{\mathcal{D}} = \underline{\phi}^D_{\mathcal{D}} + \underline{\psi}^e_{\mathcal{D}}, \text{ where } \underline{\psi}^e_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0} \text{ is the solution to (2.31) with} (z^N, z^P) = (\alpha_N, \alpha_P), \\ \underline{N}^e_{\mathcal{D}} = g(\alpha_N \underline{1}_{\mathcal{D}} + \underline{\phi}^e_{\mathcal{D}}), \\ \underline{P}^e_{\mathcal{D}} = g(\alpha_P \underline{1}_{\mathcal{D}} - \underline{\phi}^e_{\mathcal{D}}). \end{cases} \tag{2.32}$$

Note that, by definition, the discrete densities at thermal equilibrium $\underline{N}^e_{\mathcal{D}}$ and $\underline{P}^e_{\mathcal{D}}$ are I_h-valued. We remind the reader that alternative formulations (in the spirit of finite volume methods) of the schemes (2.28) and (2.31) are presented in Section 2.5.1.

We are now in position to state the main result of this paper, whose proof is the subject of Section 2.4.

Theorem 7 (Existence of discrete I_h-valued solutions to (2.28))**.** *Let N^{in} and P^{in} two positive functions in $L^\infty(\Omega)$ satisfying the condition (2.3). There exists at least one solution $\left((\underline{N}_\mathcal{D}^n, \underline{P}_\mathcal{D}^n, \underline{\phi}_\mathcal{D}^n)\right)_{n\in\mathbb{N}} \in \underline{V}_\mathcal{D}^{3\,\mathbb{N}}$ to the coupled scheme (2.28). It satisfies the following entropy-dissipation relation:*

$$\forall n \in \mathbb{N}, \quad \frac{\mathbb{E}^{n+1} - \mathbb{E}^n}{\Delta t} + \mathbb{D}^{n+1} \leq 0, \tag{2.33}$$

where $\mathbb{E}^n$ and $\mathbb{D}^n$ are, respectively, the discrete relative entropy and dissipation at time $t^n = n\Delta t$, defined by

$$\begin{aligned}\mathbb{E}^n = &\int_\Omega H(N_\mathcal{M}^n) - H(N_\mathcal{M}^e) - h(N_\mathcal{M}^e)\left(N_\mathcal{M}^n - N_\mathcal{M}^e\right) \\ &+ \int_\Omega H(P_\mathcal{M}^n) - H(P_\mathcal{M}^e) - h(P_\mathcal{M}^e)\left(P_\mathcal{M}^n - P_\mathcal{M}^e\right) + \frac{1}{2} a_\mathcal{D}^\phi(\underline{\phi}_\mathcal{D}^n - \underline{\phi}_\mathcal{D}^e, \underline{\phi}_\mathcal{D}^n - \underline{\phi}_\mathcal{D}^e),\end{aligned} \tag{2.34}$$

$$\text{and } \mathbb{D}^n = T_\mathcal{D}^N(\underline{N}_\mathcal{D}^n, \underline{w}_\mathcal{D}^{N,n}, \underline{w}_\mathcal{D}^{N,n}) + T_\mathcal{D}^P(\underline{P}_\mathcal{D}^n, \underline{w}_\mathcal{D}^{P,n}, \underline{w}_\mathcal{D}^{P,n}) + \int_\Omega R(N_\mathcal{M}^n, P_\mathcal{M}^n)\left(h(N_\mathcal{M}^n) + h(P_\mathcal{M}^n)\right). \tag{2.35}$$

Moreover, there exist $0 < M_\flat \leq M_\sharp < a$, depending on the physical data, Δt, and $\mathcal{D}$ such that

$$\forall n \geq 1, \qquad M_\flat \underline{1}_\mathcal{D} \leq \underline{N}_\mathcal{D}^n \leq M_\sharp \underline{1}_\mathcal{D} \text{ and } M_\flat \underline{1}_\mathcal{D} \leq \underline{P}_\mathcal{D}^n \leq M_\sharp \underline{1}_\mathcal{D}. \tag{2.36}$$

Remark 13 (L^∞ bounds)**.** *Notice that the bounds $M_\flat$ and $M_\sharp$ on the discrete densities depend on the time step Δt and on the spatial discretisation $\mathcal{D}$ (and therefore, on the meshsize). Thus, our result is weaker than the one proved in [28, Theorem 1] in the framework of TPFA schemes. In fact, we believe that the discrete Moser iteration process used in [28] to get uniform bounds cannot be adapted in the hybrid framework, because the method is intrinsically non-monotone.*

2.3 Analysis of the stationary scheme

In this section, we analyse the scheme (2.31) for the approximation of the semi-linear elliptic equation (2.30). First, we show that the scheme is well-posed. Then, we use this result to show that there is a correspondence between the discrete densities and the discrete quasi-Fermi potentials, which will be very useful for the following.

2.3.1 Well-posedness

In the next theorem, we show that the scheme (2.31) admits a unique solution. The proof relies on an energy minimisation approach.

Theorem 8 (Well-posedness for (2.31))**.** *Let z^N and z^P be two functions in $L^\infty(\Omega)$. There exists a unique $\underline{\psi}_\mathcal{D} \in \underline{V}_{\mathcal{D},0}$ solution to (2.31), and there exists a positive constant $C_{Poisson}$, only depending on Λ_ϕ, Ω, Γ^D, d and $\theta_\mathcal{D}$ such that the following stability estimate holds:*

$$|\underline{\psi}_\mathcal{D}|_{1,\mathcal{D}}^2 \leq C_{Poisson}\left[G\left(\|z^P\|_{L^\infty(\Omega)} + \|z^N\|_{L^\infty(\Omega)} + \|\phi^D\|_{L^\infty(\Omega)}\right) + \|C\|_{L^2(\Omega)}^2 + \|\phi^D\|_{H^{1/2}(\Gamma^D)}^2\right]. \tag{2.37}$$

Moreover, the application $(z^N, z^P) \mapsto \underline{\psi}_{\mathcal{D}}$ *is continuous from* $L^\infty(\Omega)\times L^\infty(\Omega)$ *to* $\underline{V}_{\mathcal{D},0}$.

Proof. We define the discrete energy functional $J_{\mathcal{D}} : \underline{V}_{\mathcal{D},0} \to \mathbb{R}$ by

$$J_{\mathcal{D}}(\underline{u}_{\mathcal{D}}) = \frac{1}{2}a^{\phi}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}, \underline{u}_{\mathcal{D}}) - L^{\phi}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}) + \int_\Omega G(z^P_{\mathcal{M}} - \phi^D_{\mathcal{M}} - u_{\mathcal{M}}) + G(z^N_{\mathcal{M}} + \phi^D_{\mathcal{M}} + u_{\mathcal{M}}). \tag{2.38}$$

Since G is $\mathcal{C}^2$ on $\mathbb{R}$, $J_{\mathcal{D}}$ is $\mathcal{C}^2$ on $\underline{V}_{\mathcal{D},0}$, and one can compute its differential at a point $\underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}$:

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0},\ dJ_{\mathcal{D}}(\underline{u}_{\mathcal{D}})\cdot \underline{v}_{\mathcal{D}} = a^{\phi}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) - L^{\phi}_{\mathcal{D}}(\underline{v}_{\mathcal{D}}) - \int_\Omega \Big[g(z^P_{\mathcal{M}} - \phi^D_{\mathcal{M}} - u_{\mathcal{M}}) - g(z^N_{\mathcal{M}} + \phi^D_{\mathcal{M}} + u_{\mathcal{M}})\Big] v_{\mathcal{M}}.$$

Hence, a solution $\underline{\psi}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}$ to (2.31) is an element of $\underline{V}_{\mathcal{D},0}$ such that $dJ_{\mathcal{D}}(\underline{\psi}_{\mathcal{D}}) = 0$. Now, one can notice that $J_{\mathcal{D}}$ is strongly convex on $\underline{V}_{\mathcal{D},0}$. Indeed, since $a^{\phi}_{\mathcal{D}}$ is coercive and $L^{\phi}_{\mathcal{D}}$ is a continuous linear form, $\underline{u}_{\mathcal{D}} \mapsto \frac{1}{2}a^{\phi}_{\mathcal{D}}(\underline{u}_{\mathcal{D}}, \underline{u}_{\mathcal{D}}) - L^{\phi}_{\mathcal{D}}(\underline{u}_{\mathcal{D}})$ is strongly convex. Moreover, G is a convex function (since $G' = g$ is strictly increasing), so $\underline{u}_{\mathcal{D}} \mapsto \int_\Omega G(z^P_{\mathcal{M}} - \phi^D_{\mathcal{M}} - u_{\mathcal{M}}) + G(z^N_{\mathcal{M}} + \phi^D_{\mathcal{M}} + u_{\mathcal{M}})$ is also convex. Therefore, $J_{\mathcal{D}}$ has a unique minimum $\underline{\psi}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}$, which is a solution to (2.31). On the other hand, by convexity of $J_{\mathcal{D}}$, any critical point of $J_{\mathcal{D}}$ in $\underline{V}_{\mathcal{D},0}$ is a minimum, therefore the solution to (2.31) is unique. The continuity of $(z^N_{\mathcal{M}}, z^P_{\mathcal{M}}) \mapsto \underline{\psi}_{\mathcal{D}}$ follows from the regularity of $J_{\mathcal{D}}$ and the implicit function theorem. To obtain (2.37), one can use the monotonicity of G and (2.11) to get

$$\begin{aligned} J_{\mathcal{D}}(\underline{\psi}_{\mathcal{D}}) \le J_{\mathcal{D}}(\underline{0}_{\mathcal{D}}) &= \int_\Omega G(z^P_{\mathcal{M}} - \phi^D_{\mathcal{M}}) + G(z^N_{\mathcal{M}} + \phi^D_{\mathcal{M}}) \\ &\le \int_\Omega G\Big(\|z^P\|_{L^\infty(\Omega)} + \|\phi^D\|_{L^\infty(\Omega)}\Big) + G\Big(\|z^N\|_{L^\infty(\Omega)} + \|\phi^D\|_{L^\infty(\Omega)}\Big) \\ &\le 2|\Omega| G\Big(\|z^N\|_{L^\infty(\Omega)} + \|z^P\|_{L^\infty(\Omega)} + \|\phi^D\|_{L^\infty(\Omega)}\Big). \end{aligned}$$

Furthermore, using the fact that G is positive alongside with (2.16) and (2.21), and using Young's inequality on $L^{\phi}_{\sharp}|\underline{\psi}_{\mathcal{D}}|_{1,\mathcal{D}}$, one has the following lower bound:

$$J_{\mathcal{D}}(\underline{\psi}_{\mathcal{D}}) \ge \frac{1}{2}a^{\phi}_{\mathcal{D}}(\underline{\psi}_{\mathcal{D}}, \underline{\psi}_{\mathcal{D}}) - L^{\phi}_{\mathcal{D}}(\underline{\psi}_{\mathcal{D}}) \ge \frac{\alpha_{\flat}\lambda_{\flat}}{2}|\underline{\psi}_{\mathcal{D}}|^2_{1,\mathcal{D}} - L^{\phi}_{\sharp}|\underline{\psi}_{\mathcal{D}}|_{1,\mathcal{D}} \ge \frac{\alpha_{\flat}\lambda_{\flat}}{4}|\underline{\psi}_{\mathcal{D}}|^2_{1,\mathcal{D}} - \frac{1}{\alpha_{\flat}\lambda_{\flat}}L^{\phi\,2}_{\sharp}.$$

We conclude by combining the previous estimates and using the bound on $L^{\phi}_{\sharp}$. □

Notice that this result ensures the existence and uniqueness of the discrete thermal equilibrium $(\underline{N}^e_{\mathcal{D}}, \underline{P}^e_{\mathcal{D}}, \underline{\phi}^e_{\mathcal{D}})$ described in equation (2.32).

Remark 14 (Uniform bounds)**.** *The stability result* (2.37) *gives a uniform bound (with respect to the meshsize* $h_{\mathcal{D}}$*) on* $\underline{\psi}_{\mathcal{D}}$ *in energy norm* $|\cdot|_{1,\mathcal{D}}$*. There is however no clear* L^∞ *bound (uniform in* $\mathcal{D}$*) on the potential. It is rather different from some previous results in the TPFA framework (see [68]), where* L^∞ *bounds were obtained alongside with the* H^1 *estimate.*

2.3.2 Correspondence between discrete densities and discrete potentials

In this section, we discuss the link between discrete densities and discrete potentials.

For a given couple $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D}}^2$ of I_h-valued vectors, one associates a unique discrete electrostatic potential $\underline{\phi}_{\mathcal{D}} = \underline{\psi}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}^D \in \underline{V}_{\mathcal{D}}$ such that $\underline{\psi}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}$ is the solution to

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0},\ a_{\mathcal{D}}^{\phi}(\underline{\psi}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = L_{\mathcal{D}}^{\phi}(\underline{v}_{\mathcal{D}}) + \int_{\Omega} (P_{\mathcal{M}} - N_{\mathcal{M}}) v_{\mathcal{M}}. \tag{2.39}$$

Since $a_{\mathcal{D}}^{\phi}$ is coercive on $\underline{V}_{\mathcal{D},0}$ and $L_{\mathcal{D}}^{\phi}$ is a linear form, (2.39) is well-posed and $\underline{\psi}_{\mathcal{D}}$ is uniquely defined, so is $\underline{\phi}_{\mathcal{D}}$. Then, one associates to $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}})$ a unique couple of discrete quasi-Fermi potentials $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \in \underline{V}_{\mathcal{D}}^2$ defined by

$$\underline{w}_{\mathcal{D}}^N = h(\underline{N}_{\mathcal{D}}) - \underline{\phi}_{\mathcal{D}} - \alpha_N \underline{1}_{\mathcal{D}} \text{ and } \underline{w}_{\mathcal{D}}^P = h(\underline{P}_{\mathcal{D}}) + \underline{\phi}_{\mathcal{D}} - \alpha_P \underline{1}_{\mathcal{D}}. \tag{2.40}$$

We call these vectors the discrete electrostatic $(\underline{\phi}_{\mathcal{D}})$ and quasi-Fermi potentials $((\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P))$ associated to the discrete densities $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}})$.
Note that, as in Remark 11, the discrete quasi-Fermi potentials associated to $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}})$ are in $\underline{V}_{\mathcal{D},0}$ if and only if the densities satisfy the following discrete boundary condition:

$$\forall \sigma \in \mathcal{E}_{ext}^D,\ N_\sigma = N_\sigma^D \text{ and } P_\sigma = P_\sigma^D. \tag{2.41}$$

Conversely, for a given couple $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \in \underline{V}_{\mathcal{D}}^2$, one associates a unique electrostatic potential $\underline{\phi}_{\mathcal{D}} = \underline{\psi}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}^D \in \underline{V}_{\mathcal{D}}$ such that $\underline{\psi}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}$ is the solution to

$$\begin{aligned}\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0},\ a_{\mathcal{D}}^{\phi}(\underline{\psi}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = L_{\mathcal{D}}^{\phi}(\underline{v}_{\mathcal{D}}) + \\ \int_{\Omega} \Big(g(w_{\mathcal{M}}^P + \alpha_P - \phi_{\mathcal{M}}^D - \psi_{\mathcal{M}}) - g(w_{\mathcal{M}}^N + \alpha_N + \phi_{\mathcal{M}}^D + \psi_{\mathcal{M}}) \Big) v_{\mathcal{M}}.\end{aligned} \tag{2.42}$$

Indeed, since the functions $w_{\mathcal{M}}^P + \alpha_P$ and $w_{\mathcal{M}}^N + \alpha_N$ are in $L^\infty(\Omega)$, according to Theorem 8 applied to $(z^N, z^P) = (w_{\mathcal{M}}^P + \alpha_P, w_{\mathcal{M}}^N + \alpha_N)$, there exists a unique solution to problem (2.42). Then, one associates to $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P)$ a couple of I_h-valued discrete carrier densities $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D}}$ by

$$\underline{N}_D = g(\underline{w}_{\mathcal{D}}^N + \underline{\phi}_{\mathcal{D}} + \alpha_N \underline{1}_{\mathcal{D}}) \text{ and } \underline{P}_D = g(\underline{w}_{\mathcal{D}}^P - \underline{\phi}_{\mathcal{D}} + \alpha_P \underline{1}_{\mathcal{D}}).$$

Notice that with this process, according to (2.42) and the definition of $\underline{N}_{\mathcal{D}}$ and $\underline{P}_{\mathcal{D}}$, one has that

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0},\ a_{\mathcal{D}}^{\phi}(\underline{\psi}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = L_{\mathcal{D}}^{\phi}(\underline{v}_{\mathcal{D}}) + \int_{\Omega} (P_{\mathcal{M}} - N_{\mathcal{M}}) v_{\mathcal{M}}.$$

Therefore, it means that $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P)$ and $\underline{\phi}_{\mathcal{D}}$ are the discrete potentials associated to the discrete densities $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}})$ in the sense of (2.39)-(2.40).
Moreover, $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \in \underline{V}_{\mathcal{D},0}^2$ if and only if $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}})$ satisfy the discrete boundary condition (2.41).

Thus, there is a bijective correspondence between (I_h-valued) discrete densities and discrete quasi-Fermi potentials (in $\underline{V}_{\mathcal{D}}$).

Remark 15 (Thermal equilibrium and quasi-Fermi potentials). *From the definition (2.32) of the discrete thermal equilibrium, one can notice that $(\underline{N}_{\mathcal{D}}^e, \underline{P}_{\mathcal{D}}^e, \underline{\phi}_{\mathcal{D}}^e)$ are the discrete densities and electrostatic potential associated to the quasi-Fermi potentials at thermal equilibrium, which are the null vectors $\left(\underline{w}_{\mathcal{D}}^{N,e}, \underline{w}_{\mathcal{D}}^{P,e}\right) = (\underline{0}_{\mathcal{D}}, \underline{0}_{\mathcal{D}})$.*

2.4 Analysis of the transient scheme

In this section, we perform the analysis of the scheme (2.28) for the drift-diffusion system. The first two subsections are dedicated to the proof of Theorem 7. We also state and prove a qualitative property about the discrete solutions, namely the long-time behaviour in Theorem 9, in line with the work of [27].

2.4.1 Discrete entropy structure

The scheme (2.28) is designed so as to mimic the continuous entropic structure presented in introduction. In this part, we are interested in the entropic property of the scheme (2.28).

First, we define a discrete counterpart of the entropy and entropy dissipation. Given $(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D})$ a couple of I_h-valued vectors, one can define the relative discrete entropy as

$$\mathbb{E}(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D}) = \int_\Omega H(N_\mathcal{M}) - H(N^e_\mathcal{M}) - h(N^e_\mathcal{M})\left(N_\mathcal{M} - N^e_\mathcal{M}\right) + \int_\Omega H(P_\mathcal{M}) - H(P^e_\mathcal{M}) - h(P^e_\mathcal{M})\left(P_\mathcal{M} - P^e_\mathcal{M}\right) + \frac{1}{2} a^\phi_\mathcal{D}(\underline{\phi}_\mathcal{D} - \underline{\phi}^e_\mathcal{D}, \underline{\phi}_\mathcal{D} - \underline{\phi}^e_\mathcal{D}), \quad (2.43)$$

where $(\underline{N}^e_\mathcal{D}, \underline{P}^e_\mathcal{D}, \underline{\phi}^e_\mathcal{D})$ are defined in (2.32), and $\underline{\phi}_\mathcal{D}$ is the discrete electrostatic potential associated to $(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D})$ in the sense of Section 2.3.2. Notice that, by convexity of H, $H(N_\mathcal{M}) - H(N^e_\mathcal{M}) - h(N^e_\mathcal{M})\left(N_\mathcal{M} - N^e_\mathcal{M}\right)$ and the analogous term in $P_\mathcal{M}$ are non-negative. Therefore, the discrete entropy is non-negative. Similarly, one defines the discrete dissipation by

$$\mathbb{D}(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D}) = T^N_\mathcal{D}(\underline{N}_\mathcal{D}, \underline{w}^N_\mathcal{D}, \underline{w}^N_\mathcal{D}) + T^P_\mathcal{D}(\underline{P}_\mathcal{D}, \underline{w}^P_\mathcal{D}, \underline{w}^P_\mathcal{D}) + \int_\Omega R(N_\mathcal{M}, P_\mathcal{M})(h(N_\mathcal{M}) + h(P_\mathcal{M})). \quad (2.44)$$

By definition of $T^N_\mathcal{D}$ and $T^P_\mathcal{D}$, the first two terms are non-negative. One can also notice that the integrand of the third term writes

$$R(N_\mathcal{M}, P_\mathcal{M})(h(N_\mathcal{M}) + h(P_\mathcal{M})) = r(x, N_\mathcal{M}, P_\mathcal{M})(\exp(h(N_\mathcal{M}) + h(P_\mathcal{M})) - 1)(h(N_\mathcal{M}) + h(P_\mathcal{M})),$$

so since r is non-negative and $x(e^x - 1) \geq 0$ for any real x, the discrete dissipation is also non-negative. Note that, given a couple $(\underline{w}^N_\mathcal{D}, \underline{w}^P_\mathcal{D}) \in \underline{V}^2_\mathcal{D}$ of discrete quasi-Fermi potentials, one can also define the corresponding relative entropy and dissipation by $\mathbb{E}(\underline{w}^N_\mathcal{D}, \underline{w}^P_\mathcal{D}) = \mathbb{E}(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D})$ and $\mathbb{D}(\underline{w}^N_\mathcal{D}, \underline{w}^P_\mathcal{D}) = \mathbb{D}(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D})$, where $(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D})$ are the discrete densities associated to $(\underline{w}^N_\mathcal{D}, \underline{w}^P_\mathcal{D})$.

At the continuous level, the analysis of the drift-diffusion system relies on the entropic structure. At the discrete level, we use analogous results, based on the following lemma, which will be used in the proofs of Theorems 7 and 9.

Lemma 5 (Discrete differentiation of the entropy). *Let $\underline{N}^n_\mathcal{D}$, $\underline{P}^n_\mathcal{D}$, $\underline{N}_\mathcal{D}$ and $\underline{P}_\mathcal{D}$ be four I_h-valued elements of $\underline{V}_\mathcal{D}$, and $(\underline{w}^N_\mathcal{D}, \underline{w}^P_\mathcal{D}) \in \underline{V}^2_\mathcal{D}$ be the discrete quasi-Fermi potentials associated to the discrete densities $(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D})$. Then, the following inequality holds:*

$$\mathbb{E}(\underline{N}_\mathcal{D}, \underline{P}_\mathcal{D}) - \mathbb{E}(\underline{N}^n_\mathcal{D}, \underline{P}^n_\mathcal{D}) \leq \int_\Omega \left(N_\mathcal{M} - N^n_\mathcal{M}\right) w^N_\mathcal{M} + \int_\Omega \left(P_\mathcal{M} - P^n_\mathcal{M}\right) w^P_\mathcal{M}. \quad (2.45)$$

Proof. By definition of the discrete entropy, one has

$$\begin{aligned}\mathbb{E}(\underline{N}_{\mathcal{D}},\underline{P}_{\mathcal{D}})-\mathbb{E}(\underline{N}^n_{\mathcal{D}},\underline{P}^n_{\mathcal{D}}) &= \int_\Omega H(N_{\mathcal{M}})-H(N^n_{\mathcal{M}})-h(N^e_{\mathcal{M}})\left(N_{\mathcal{M}}-N^n_{\mathcal{M}}\right)\\ &\quad+\int_\Omega H(P_{\mathcal{M}})-H(P^n_{\mathcal{M}})-h(P^e_{\mathcal{M}})\left(P_{\mathcal{M}}-P^n_{\mathcal{M}}\right)\\ &\quad+\frac{1}{2}\left(a^\phi_{\mathcal{D}}(\underline{\phi}_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}},\underline{\phi}_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}})-a^\phi_{\mathcal{D}}(\underline{\phi}^n_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}},\underline{\phi}^n_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}})\right). \quad (2.46)\end{aligned}$$

Let us first consider the first term (in N) of (2.46): by convexity of H and definition of $\underline{N}^e_{\mathcal{D}}$, we get

$$\begin{aligned}\int_\Omega H(N_{\mathcal{M}})-H(N^n_{\mathcal{M}})-h(N^e_{\mathcal{M}})\left(N_{\mathcal{M}}-N^n_{\mathcal{M}}\right) &\le \int_\Omega (h(N_{\mathcal{M}})-h(N^e_{\mathcal{M}}))\left(N_{\mathcal{M}}-N^n_{\mathcal{M}}\right)\\ &= \int_\Omega (h(N_{\mathcal{M}})-\phi^e_{\mathcal{M}}-\alpha_N)\left(N_{\mathcal{M}}-N^n_{\mathcal{M}}\right).\end{aligned}$$

For the second term of (2.46), one has an analogous result, namely

$$\int_\Omega H(P_{\mathcal{M}})-H(P^n_{\mathcal{M}})-h(P^e_{\mathcal{M}})\left(P_{\mathcal{M}}-P^n_{\mathcal{M}}\right) \le \int_\Omega (h(P_{\mathcal{M}})+\phi^e_{\mathcal{M}}-\alpha_P)\left(P_{\mathcal{M}}-P^n_{\mathcal{M}}\right).$$

Now, notice that since $a^\phi_{\mathcal{D}}$ is a symmetric and positive semi-definite bilinear form on $\underline{V}_{\mathcal{D}}$, the quadratic form $\underline{u}_{\mathcal{D}}\mapsto a^\phi_{\mathcal{D}}(\underline{u}_{\mathcal{D}},\underline{u}_{\mathcal{D}})$ is convex, and therefore for any $(\underline{u}_{\mathcal{D}},\underline{v}_{\mathcal{D}})\in\underline{V}^2_{\mathcal{D}}$, one has

$$a^\phi_{\mathcal{D}}(\underline{u}_{\mathcal{D}},\underline{u}_{\mathcal{D}})-a^\phi_{\mathcal{D}}(\underline{v}_{\mathcal{D}},\underline{v}_{\mathcal{D}}) \le 2a^\phi_{\mathcal{D}}(\underline{u}_{\mathcal{D}}-\underline{v}_{\mathcal{D}},\underline{u}_{\mathcal{D}}).$$

Using this relation with $\underline{u}_{\mathcal{D}}=\underline{\phi}_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}}$ and $\underline{v}_{\mathcal{D}}=\underline{\phi}^n_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}}\in\underline{V}_{\mathcal{D}}$, we obtain the following estimate on the third term of (2.46):

$$\frac{1}{2}\left(a^\phi_{\mathcal{D}}(\underline{\phi}_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}},\underline{\phi}_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}})-a^\phi_{\mathcal{D}}(\underline{\phi}^n_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}},\underline{\phi}^n_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}})\right) \le a^\phi_{\mathcal{D}}(\underline{\phi}_{\mathcal{D}}-\underline{\phi}^n_{\mathcal{D}},\underline{\phi}_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}}).$$

We recall that the electrostatic potentials are defined by (2.39), and that $\underline{\phi}_{\mathcal{D}}-\underline{\phi}^n_{\mathcal{D}}=\underline{\psi}_{\mathcal{D}}-\underline{\psi}^n_{\mathcal{D}}$. Hence, letting $\underline{v}_{\mathcal{D}}=\underline{\phi}_{\mathcal{D}}-\underline{\phi}^e_{\mathcal{D}}\in\underline{V}_{\mathcal{D},0}$ and using the linearity of $a^\phi_{\mathcal{D}}$ with respect to the first variable, one gets

$$\begin{aligned}a^\phi_{\mathcal{D}}(\underline{\phi}_{\mathcal{D}}-\underline{\phi}^n_{\mathcal{D}},\underline{v}_{\mathcal{D}}) &= a^\phi_{\mathcal{D}}(\underline{\psi}_{\mathcal{D}},\underline{v}_{\mathcal{D}})-a^\phi_{\mathcal{D}}(\underline{\psi}^n_{\mathcal{D}},\underline{v}_{\mathcal{D}})\\ &= L^\phi_{\mathcal{D}}(\underline{v}_{\mathcal{D}})+\int_\Omega (P_{\mathcal{M}}-N_{\mathcal{M}})v_{\mathcal{M}}-L^\phi_{\mathcal{D}}(\underline{v}_{\mathcal{D}})-\int_\Omega (P^n_{\mathcal{M}}-N^n_{\mathcal{M}})v_{\mathcal{M}}\\ &= \int_\Omega (P_{\mathcal{M}}-P^n_{\mathcal{M}})v_{\mathcal{M}}-\int_\Omega (N_{\mathcal{M}}-N^n_{\mathcal{M}})v_{\mathcal{M}}.\end{aligned}$$

Combining the previous estimates, we obtain

$$\mathbb{E}(\underline{N}_{\mathcal{D}},\underline{P}_{\mathcal{D}})-\mathbb{E}(\underline{N}^n_{\mathcal{D}},\underline{P}^n_{\mathcal{D}})\leq\int_\Omega(h(N_{\mathcal{M}})-\phi^e_{\mathcal{M}}-\alpha_N-v_{\mathcal{M}})\left(N_{\mathcal{M}}-N^n_{\mathcal{M}}\right)\\+\int_\Omega(h(P_{\mathcal{M}})+\phi^e_{\mathcal{M}}-\alpha_P+v_{\mathcal{M}})\left(P_{\mathcal{M}}-P^n_{\mathcal{M}}\right).$$

But, by definition of $\underline{v}_{\mathcal{D}}$, we have $h(N_{\mathcal{M}})-\phi^e_{\mathcal{M}}-\alpha_N-v_{\mathcal{M}}=h(N_{\mathcal{M}})-\phi_{\mathcal{M}}-\alpha_N=w^N_{\mathcal{M}}$ and $h(P_{\mathcal{M}})+\phi^e_{\mathcal{M}}-\alpha_P+v_{\mathcal{M}}=w^P_{\mathcal{M}}$, therefore (2.45) holds. □

One can now state the following a priori result about the dissipation of entropy, which ensures that (2.33) holds.

Proposition 11 (Entropy dissipation). *Assume that* $\left((\underline{N}^n_{\mathcal{D}},\underline{P}^n_{\mathcal{D}},\underline{\phi}^n_{\mathcal{D}})\right)_{n\in\mathbb{N}}\in\underline{V}_{\mathcal{D}}^{3\,\mathbb{N}}$ *is a solution to* (2.28). *Then, the following entropy-entropy dissipation relation holds:*

$$\forall n\geq 0,\quad\frac{\mathbb{E}(\underline{N}^{n+1}_{\mathcal{D}},\underline{P}^{n+1}_{\mathcal{D}})-\mathbb{E}(\underline{N}^n_{\mathcal{D}},\underline{P}^n_{\mathcal{D}})}{\Delta t}\leq-\mathbb{D}(\underline{N}^{n+1}_{\mathcal{D}},\underline{P}^{n+1}_{\mathcal{D}}).$$

Proof. Since the discrete quasi-Fermi potentials $\underline{w}^{N,n+1}_{\mathcal{D}}$ and $\underline{w}^{P,n+1}_{\mathcal{D}}$ are in $\underline{V}_{\mathcal{D},0}$ by (2.28a) and (2.28b), one can use them as test functions in (2.28d) and (2.28e): summing the two identities, and noticing that $w^{N,n+1}_{\mathcal{M}}+w^{P,n+1}_{\mathcal{M}}=h(N^{n+1}_{\mathcal{M}})+h(P^{n+1}_{\mathcal{M}})$ because of (2.5), one gets that

$$\int_\Omega\frac{N^{n+1}_{\mathcal{M}}-N^n_{\mathcal{M}}}{\Delta t}w^{N,n+1}_{\mathcal{M}}+\int_\Omega\frac{P^{n+1}_{\mathcal{M}}-P^n_{\mathcal{M}}}{\Delta t}w^{P,n+1}_{\mathcal{M}}=-\mathbb{D}(\underline{N}^{n+1}_{\mathcal{D}},\underline{P}^{n+1}_{\mathcal{D}}).$$

One concludes using Lemma 5. □

From this entropic structure follows an important preservation property of the scheme (2.28).

Proposition 12 (Thermodynamic consistency). *Let* $(\underline{N}^\infty_{\mathcal{D}},\underline{P}^\infty_{\mathcal{D}},\underline{\phi}^\infty_{\mathcal{D}})\in\underline{V}^3_{\mathcal{D}}$ *be a discrete stationary state of the scheme* (2.28), *in the sense that there exist* $\left((\underline{N}^n_{\mathcal{D}},\underline{P}^n_{\mathcal{D}},\underline{\phi}^n_{\mathcal{D}})\right)_{n\in\mathbb{N}}\in\underline{V}_{\mathcal{D}}^{3\,\mathbb{N}}$ *a solution to* (2.28) *and* $n_\infty\in\mathbb{N}$ *such that,*

$$\forall n\geq n_\infty,\ (\underline{N}^n_{\mathcal{D}},\underline{P}^n_{\mathcal{D}},\underline{\phi}^n_{\mathcal{D}})=(\underline{N}^\infty_{\mathcal{D}},\underline{P}^\infty_{\mathcal{D}},\underline{\phi}^\infty_{\mathcal{D}}).$$

Then the discrete stationary state coincides with the discrete thermal equilibrium:

$$(\underline{N}^\infty_{\mathcal{D}},\underline{P}^\infty_{\mathcal{D}},\underline{\phi}^\infty_{\mathcal{D}})=(\underline{N}^e_{\mathcal{D}},\underline{P}^e_{\mathcal{D}},\underline{\phi}^e_{\mathcal{D}}).\tag{2.47}$$

Proof. Let $n\geq n_\infty$, we have

$$\mathbb{E}(\underline{N}^{n+1}_{\mathcal{D}},\underline{P}^{n+1}_{\mathcal{D}})-\mathbb{E}(\underline{N}^n_{\mathcal{D}},\underline{P}^n_{\mathcal{D}})=\mathbb{E}(\underline{N}^\infty_{\mathcal{D}},\underline{P}^\infty_{\mathcal{D}})-\mathbb{E}(\underline{N}^\infty_{\mathcal{D}},\underline{P}^\infty_{\mathcal{D}})=0.$$

According to Proposition 11, one has by positivity of the dissipation

$$\mathbb{D}(\underline{N}^\infty_{\mathcal{D}},\underline{P}^\infty_{\mathcal{D}})=\mathbb{D}(\underline{N}^{n+1}_{\mathcal{D}},\underline{P}^{n+1}_{\mathcal{D}})=0.$$

But, since all the terms of the dissipation are non-negative (see (2.44)), it means that

$$T^N_{\mathcal{D}}(\underline{N}^\infty_{\mathcal{D}},\underline{w}^{N,\infty}_{\mathcal{D}},\underline{w}^{N,\infty}_{\mathcal{D}})=T^P_{\mathcal{D}}(\underline{P}^\infty_{\mathcal{D}},\underline{w}^{P,\infty}_{\mathcal{D}},\underline{w}^{P,\infty}_{\mathcal{D}})=0,$$

where $(\underline{w}_{\mathcal{D}}^{N,\infty}, \underline{w}_{\mathcal{D}}^{P,\infty}) \in \underline{V}_{\mathcal{D},0}^2$ are the quasi-Fermi potentials associated to the discrete stationary densities $(\underline{N}_{\mathcal{D}}^{\infty}, \underline{P}_{\mathcal{D}}^{\infty})$. Now, let $K \in \mathcal{M}$. Since all the coordinates of $\underline{N}_K^{\infty}$ are positive, $r_K(\underline{N}_K^{\infty})$ is also positive, and by definition (2.22) of T_K^N, we get that

$$a_K^N(\underline{w}_K^{N,\infty}, \underline{w}_K^{N,\infty}) = 0.$$

Using the local coercivity estimates (2.15), we infer that $|\underline{w}_K^{N,\infty}|_{1,K}^2 = 0$. This holds for any cell $K \in \mathcal{M}$, so we have $|\underline{w}_{\mathcal{D}}^{N,\infty}|_{1,\mathcal{D}} = 0$, but $\underline{w}_{\mathcal{D}}^{N,\infty} \in \underline{V}_{\mathcal{D},0}$, therefore $\underline{w}_{\mathcal{D}}^{N,\infty} = \underline{0}_{\mathcal{D}}$. A similar result holds for $\underline{w}_{\mathcal{D}}^{P,\infty}$. Thus, one has $\underline{w}_{\mathcal{D}}^{N,\infty} = \underline{w}_{\mathcal{D}}^{P,\infty} = \underline{0}_{\mathcal{D}}$ and, by Remark 12, it means that $(\underline{w}_{\mathcal{D}}^{N,\infty}, \underline{w}_{\mathcal{D}}^{P,\infty})$ are equal to the quasi-Fermi potentials at thermal equilibrium. By correspondence between densities and potential discussed in Section 2.3.2, we get (2.47). □

Remark 16 (Preservation of the thermal equilibrium)**.** *The statement of Proposition 12 means that the scheme (2.28) preserves the thermal equilibrium: the only admissible discrete stationary state is the discrete thermal equilibrium. This remarkable feature is sometimes called thermodynamic consistency in the context of TPFA schemes (see [127]). In such a framework, this result is expressed in the following way: if the numerical fluxes of the convection-diffusion equations vanish, then the discrete quasi-Fermi potentials are constant (equal to zero). For our HFV scheme, this statement still holds, using the numerical fluxes (2.5.1) defined in Section 2.5.1. In fact, both statements in terms of fluxes and stationary states are equivalent. This notion of thermodynamic consistency can indeed be reformulated using the entropic property of the schemes: if the discrete dissipation associated to a discrete solution vanishes, then the solution is the discrete thermal equilibrium.*

2.4.2 Existence of solutions

The goal of this section is to prove the existence result and the estimates (2.36) of Theorem 7. The proof proposed here is an adaptation of the one introduced in [70] for similar schemes in the context of linear advection-diffusion equations. According to the discussion in Section 2.3.2, it is equivalent to seek discrete solutions in terms of densities or in terms of quasi-Fermi potentials. Therefore, we will seek discrete quasi-Fermi potentials in the whole space $\underline{V}_{\mathcal{D},0} \times \underline{V}_{\mathcal{D},0}$ instead of discrete densities in the set of I_h-valued elements of $\underline{V}_{\mathcal{D}} \times \underline{V}_{\mathcal{D}}$ (which is not a vector space).

In the sequel, we consider the vector space $\underline{V}_{\mathcal{D},0}^2$, and denote by $\underline{\mathbb{w}}_{\mathcal{D}} = (\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \in \underline{V}_{\mathcal{D},0}^2$ a typical element. We endow $\underline{V}_{\mathcal{D},0}^2$ with the following natural inner product and norm:

$$\langle \underline{\mathbb{w}}_{\mathcal{D}}, \underline{\mathbb{v}}_{\mathcal{D}} \rangle = \sum_{K\in\mathcal{M}} \left(w_K^N v_K^N + w_K^P v_K^P \right) + \sum_{\sigma \in \mathcal{E}_{int} \cup \mathcal{E}_{ext}^N} \left(w_\sigma^N v_\sigma^N + w_\sigma^P v_\sigma^P \right) \text{ and } \|\underline{\mathbb{w}}_{\mathcal{D}}\| = \sqrt{\langle \underline{\mathbb{w}}_{\mathcal{D}}, \underline{\mathbb{w}}_{\mathcal{D}} \rangle}.$$

One can also endow $\underline{V}_{\mathcal{D},0}^2$ with the l^∞ norm $\|.\|_\infty$ defined by

$$\|\underline{\mathbb{w}}_{\mathcal{D}}\|_\infty = \max_{K\in\mathcal{M},\ \sigma\in\mathcal{E}_{int}\cup\mathcal{E}_{ext}^N} \left(|w_K^N|, |w_K^P|, |w_\sigma^N|, |w_\sigma^P| \right).$$

Since $\underline{V}_{\mathcal{D},0}^2$ is a finite-dimensional vector space of dimension $2\left(|\mathcal{M}| + |\mathcal{E}_{int}| + |\mathcal{E}_{ext}^N|\right)$ there exists a constant $c_{\dim}$ only depending on $\dim(\underline{V}_{\mathcal{D},0})$ such that

$$\forall \underline{\mathbb{w}}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}^2,\ c_{\dim}\|\underline{\mathbb{w}}_{\mathcal{D}}\| \le \|\underline{\mathbb{w}}_{\mathcal{D}}\|_\infty. \tag{2.48}$$

We can now establish the existence result by induction on n. Let $\underline{N}_{\mathcal{D}}^n$ and $\underline{P}_{\mathcal{D}}^n$ be two I_h-valued vectors, we want to show that there exists a solution to (2.28a)-(2.28f). Given $\underline{\mathbb{w}}_{\mathcal{D}} =$

$(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \in \underline{V}_{\mathcal{D},0}^2$, (and $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}}, \underline{\psi}_{\mathcal{D}})$ the associated discrete densities and electrostatic potential), the application

$$\underline{\mathbb{v}}_{\mathcal{D}} \mapsto \int_\Omega \frac{N_{\mathcal{M}} - N_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}}^N + T_{\mathcal{D}}^N(\underline{N}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}^N, \underline{v}_{\mathcal{D}}^N) + \int_\Omega R(N_{\mathcal{M}}, P_{\mathcal{M}}) v_{\mathcal{M}}^N + \int_\Omega \frac{P_{\mathcal{M}} - P_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}}^P + T_{\mathcal{D}}^P(\underline{P}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}^P, \underline{v}_{\mathcal{D}}^P) + \int_\Omega R(N_{\mathcal{M}}, P_{\mathcal{M}}) v_{\mathcal{M}}^P$$

is a continuous linear form on $\underline{V}_{\mathcal{D},0}^2$. Hence, there exists a unique $\mathcal{G}(\underline{\mathbb{w}}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D},0}^2$ such that

$$\forall \underline{\mathbb{v}}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}^2,\ \langle \mathcal{G}(\underline{\mathbb{w}}_{\mathcal{D}}), \underline{\mathbb{v}}_{\mathcal{D}} \rangle = \int_\Omega \frac{N_{\mathcal{M}} - N_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}}^N + T_{\mathcal{D}}^N(\underline{N}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}^N, \underline{v}_{\mathcal{D}}^N) + \int_\Omega R(N_{\mathcal{M}}, P_{\mathcal{M}}) v_{\mathcal{M}}^N + \int_\Omega \frac{P_{\mathcal{M}} - P_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}}^P + T_{\mathcal{D}}^P(\underline{P}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}^P, \underline{v}_{\mathcal{D}}^P) + \int_\Omega R(N_{\mathcal{M}}, P_{\mathcal{M}}) v_{\mathcal{M}}^P, \quad (2.49)$$

and it is straightforward to see that $\mathcal{G}(\underline{\mathbb{w}}_{\mathcal{D}}) = (\underline{0}_{\mathcal{D}}, \underline{0}_{\mathcal{D}})$ if and only if $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}}, \underline{\phi}_{\mathcal{D}})$ is a solution to (2.28a)-(2.28f). Moreover, by the continuity result of Theorem 8, the vector field $\mathcal{G} : \underline{V}_{\mathcal{D},0}^2 \to \underline{V}_{\mathcal{D},0}^2$ is continuous. As in [70], our proof relies on two key results. The first one, that can be found, e.g., in [119, Section 9.1], is a corollary of Brouwer's fixed-point theorem.

Lemma 6. *Let k be a positive integer and let $P : \mathbb{R}^k \to \mathbb{R}^k$ be a continuous vector field. Assume that there is $r > 0$ such that*

$$P(x) \cdot x \geq 0 \qquad \text{if } |x| = r.$$

Then, there exists a point $x_0 \in \mathbb{R}^k$ such that $P(x_0) = 0$ and $|x_0| \leq r$.

The second lemma, whose proof is postponed until Appendix 2.A, establishes that bounds on the discrete entropy and dissipation terms imply bounds on the discrete quasi-Fermi potentials.

Lemma 7. *Let $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \in \underline{V}_{\mathcal{D},0}^2$, and assume that there exists $B_\sharp \geq 0$ such that*

$$\mathbb{E}(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \leq B_\sharp \qquad \text{and} \qquad \mathbb{D}(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \leq B_\sharp. \quad (2.50)$$

Then, there exists $C_\sharp > 0$, depending on the data, $B_\sharp$ and $\mathcal{D}$ such that

$$-C_\sharp \underline{1}_{\mathcal{D}} \leq \underline{w}_{\mathcal{D}}^N \leq C_\sharp \underline{1}_{\mathcal{D}},\ -C_\sharp \underline{1}_{\mathcal{D}} \leq \underline{w}_{\mathcal{D}}^P \leq C_\sharp \underline{1}_{\mathcal{D}} \text{ and } -C_\sharp \underline{1}_{\mathcal{D}} \leq \underline{\phi}_{\mathcal{D}} \leq C_\sharp \underline{1}_{\mathcal{D}},$$

where $\underline{\phi}_{\mathcal{D}}$ is the electrostatic potential associated to $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P)$.

We are now in position to prove the existence of solutions to the scheme and the estimates on these solutions.

Proof of existence - Theorem 7. In order to use the result of Lemma 6, notice that by definition of $\mathcal{G}$, and because $w_{\mathcal{M}}^N + w_{\mathcal{M}}^P = h(N_{\mathcal{M}}) + h(P_{\mathcal{M}})$, we have

$$\forall \underline{\mathbb{w}}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}^2,\ \langle \mathcal{G}(\underline{\mathbb{w}}_{\mathcal{D}}), \underline{\mathbb{w}}_{\mathcal{D}} \rangle = \int_\Omega \frac{N_{\mathcal{M}} - N_{\mathcal{M}}^n}{\Delta t} w_{\mathcal{M}}^N + \int_\Omega \frac{P_{\mathcal{M}} - P_{\mathcal{M}}^n}{\Delta t} w_{\mathcal{M}}^P + \mathbb{D}(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P).$$

Since $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P)$ are the quasi-Fermi potentials associated to $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}})$, the result of Lemma 5

ensures that

$$\forall \underline{\mathbb{w}}_{\mathcal{D}} \in \underline{V}^2_{\mathcal{D},0},\ \langle \mathcal{G}(\underline{\mathbb{w}}_{\mathcal{D}}), \underline{\mathbb{w}}_{\mathcal{D}} \rangle \geq \frac{\mathbb{E}(\underline{w}^N_{\mathcal{D}}, \underline{w}^P_{\mathcal{D}}) - \mathbb{E}(\underline{N}^n_{\mathcal{D}}, \underline{P}^n_{\mathcal{D}})}{\Delta t} + \mathbb{D}(\underline{w}^N_{\mathcal{D}}, \underline{w}^P_{\mathcal{D}}). \tag{2.51}$$

Let $B^n = \mathbb{E}(\underline{N}^n_{\mathcal{D}}, \underline{P}^n_{\mathcal{D}})$. According to Lemma 7, there exists $C^n > 0$ depending on the data, $\mathcal{D}$ and Δt such that

$$\text{if } \mathbb{E}(\underline{w}^N_{\mathcal{D}}, \underline{w}^P_{\mathcal{D}}) + \Delta t\, \mathbb{D}(\underline{w}^N_{\mathcal{D}}, \underline{w}^P_{\mathcal{D}}) \leq B^n, \text{ then } \|\underline{\mathbb{w}}_{\mathcal{D}}\|_\infty \leq C^n. \tag{2.52}$$

Now, letting $r^n = 2C^n/c_{\mathrm{dim}}$, one can notice that for any $\underline{\mathbb{w}}_{\mathcal{D}} \in \underline{V}^2_{\mathcal{D},0}$ such that $\|\underline{\mathbb{w}}_{\mathcal{D}}\| = r^n$, we have $C^n < 2C^n = c_{\mathrm{dim}}\|\underline{\mathbb{w}}_{\mathcal{D}}\| \leq \|\underline{\mathbb{w}}_{\mathcal{D}}\|_\infty$. Therefore, by (2.51) and contraposition of (2.52), if $\|\underline{\mathbb{w}}_{\mathcal{D}}\| = r^n$, one has

$$\Delta t \langle \mathcal{G}(\underline{\mathbb{w}}_{\mathcal{D}}), \underline{\mathbb{w}}_{\mathcal{D}} \rangle \geq \mathbb{E}(\underline{w}^N_{\mathcal{D}}, \underline{w}^P_{\mathcal{D}}) + \Delta t\, \mathbb{D}(\underline{w}^N_{\mathcal{D}}, \underline{w}^P_{\mathcal{D}}) - B^n > B^n - B^n \geq 0.$$

Thus, we can use the result of Lemma 6 applied to the vector field $\mathcal{G}$, and there exists at least one solution $(\underline{N}^{n+1}_{\mathcal{D}}, \underline{P}^{n+1}_{\mathcal{D}}, \underline{\phi}^{n+1}_{\mathcal{D}})$ to (2.28a)-(2.28f).

To prove the bounds (2.36), it suffices to note that the dissipation relation (2.33) implies that

$$\mathbb{E}^{n+1} \leq \mathbb{E}^n \leq \mathbb{E}^0 \text{ and } \mathbb{D}^{n+1} \leq \frac{\mathbb{E}^0}{\Delta t}.$$

Hence, by Lemma 7, there exists $C_\sharp$ depending on the data, $\mathcal{D}$, Δt and $\mathbb{E}^0$ (but not on n) such that

$$-C_\sharp \underline{1}_{\mathcal{D}} \leq \underline{w}^{N,n+1}_{\mathcal{D}} \leq C_\sharp \underline{1}_{\mathcal{D}},\ -C_\sharp \underline{1}_{\mathcal{D}} \leq \underline{w}^{P,n+1}_{\mathcal{D}} \leq C_\sharp \underline{1}_{\mathcal{D}} \text{ and } -C_\sharp \underline{1}_{\mathcal{D}} \leq \underline{\phi}^{n+1}_{\mathcal{D}} \leq C_\sharp \underline{1}_{\mathcal{D}}.$$

Therefore, by (2.29), $g(-2C_\sharp + \alpha_N)\underline{1}_{\mathcal{D}} \leq \underline{N}^{n+1}_{\mathcal{D}} \leq g(2C_\sharp + \alpha_N)\underline{1}_{\mathcal{D}}$, and an analogous estimate holds for $\underline{P}^{n+1}_{\mathcal{D}}$. □

2.4.3 Long-time behaviour

In this section, we analyse the long-time behaviour of the scheme (2.28). The analysis proposed here relies on an adaptation of the arguments presented in [27] in the framework of TPFA schemes.

Theorem 9 (Discrete long-time behaviour). *Let $\left((\underline{N}^n_{\mathcal{D}}, \underline{P}^n_{\mathcal{D}}, \underline{\phi}^n_{\mathcal{D}})\right)_{n\in\mathbb{N}} \in \underline{V}^{3\,\mathbb{N}}_{\mathcal{D}}$ be a solution to the coupled scheme (2.28). There exists $\nu_{\mathcal{D},\Delta t} > 0$ depending on the data, $\mathcal{D}$ and Δt such that*

$$\forall n \in \mathbb{N},\ \mathbb{E}^{n+1} \leq (1 + \nu_{\mathcal{D},\Delta t}\Delta t)^{-1}\mathbb{E}^n. \tag{2.53}$$

Moreover, the discrete solution converges geometrically fast towards the discrete thermal equilibrium: there exists a positive constant $c_{\mathcal{D},\Delta t}$ depending on the data, $\mathcal{D}$ and Δt such that

$$\forall n \in \mathbb{N},\ \|N^n_{\mathcal{M}} - N^e_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|P^n_{\mathcal{M}} - P^e_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|\phi^n_{\mathcal{M}} - \phi^e_{\mathcal{M}}\|^2_{L^2(\Omega)} \leq c_{\mathcal{D},\Delta t}\mathbb{E}^0(1 + \nu_{\mathcal{D},\Delta t}\Delta t)^{-n}. \tag{2.54}$$

Proof. In this proof, we will denote by c a generic constant depending on the data, $\mathcal{D}$ and Δt (but not on n). First, recall that the solutions satisfy the bounds (2.36). One can notice (using a Taylor expansion of H) that there exists two positive constants $\tilde{c_1}$ and $\tilde{c_2}$ such that for any $(x, y) \in [M_\flat, M_\sharp]$,

$$\tilde{c_1}(x-y)^2 \leq H(x) - H(y) - h(y)(x-y) \leq \tilde{c_2}(x-y)^2. \tag{2.55}$$

Therefore, letting $\hat{\mathbb{E}}^n = \|N^n_{\mathcal{M}} - N^e_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|P^n_{\mathcal{M}} - P^e_{\mathcal{M}}\|^2_{L^2(\Omega)} + \frac{1}{2}a^\phi_{\mathcal{D}}(\underline{\phi}_{\mathcal{D}} - \underline{\phi}^e_{\mathcal{D}}, \underline{\phi}_{\mathcal{D}} - \underline{\phi}^e_{\mathcal{D}})$ and using

(2.36) and (2.55), we deduce that there exist c_1 and c_2 positive such that

$$c_1\hat{\mathbb{E}}^n \le \mathbb{E}^n \le c_2\hat{\mathbb{E}}^n. \tag{2.56}$$

Since $\mathbb{E}(\underline{N}^e_{\mathcal{D}}, \underline{P}^e_{\mathcal{D}}) = 0$, by Lemma 5 (used with $(\underline{N}_{\mathcal{D}}, \underline{P}_{\mathcal{D}}) = (\underline{N}^n_{\mathcal{D}}, \underline{P}^n_{\mathcal{D}})$ and $(\underline{N}^n_{\mathcal{D}}, \underline{P}^n_{\mathcal{D}}) = (\underline{N}^e_{\mathcal{D}}, \underline{P}^e_{\mathcal{D}})$), one gets

$$c_1\hat{\mathbb{E}}^n \le \mathbb{E}^n \le \int_\Omega \left(N^n_{\mathcal{M}} - N^e_{\mathcal{M}}\right) w^{N,n}_{\mathcal{M}} + \int_\Omega \left(P^n_{\mathcal{M}} - P^e_{\mathcal{M}}\right) w^{P,n}_{\mathcal{M}}.$$

Then, we use Young inequality (with scaling c_1) to obtain

$$c_1\hat{\mathbb{E}}^n \le \frac{c_1}{2}\left(\|N^n_{\mathcal{M}} - N^e_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|P^n_{\mathcal{M}} - P^e_{\mathcal{M}}\|^2_{L^2(\Omega)}\right) + \frac{1}{2c_1}\left(\|w^{N,n}_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|w^{P,n}_{\mathcal{M}}\|^2_{L^2(\Omega)}\right).$$

Notice that $\|N^n_{\mathcal{M}} - N^e_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|P^n_{\mathcal{M}} - P^e_{\mathcal{M}}\|^2_{L^2(\Omega)} \le \hat{\mathbb{E}}^n$, therefore we have

$$\frac{c_1}{2}\hat{\mathbb{E}}^n \le \frac{1}{2c_1}\left(\|w^{N,n}_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|w^{P,n}_{\mathcal{M}}\|^2_{L^2(\Omega)}\right).$$

Combining this estimate with (2.56), we deduce that

$$\frac{c_1}{c_2}\mathbb{E}^n \le c_1\hat{\mathbb{E}}^n \le \frac{1}{c_1}\left(\|w^{N,n}_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|w^{P,n}_{\mathcal{M}}\|^2_{L^2(\Omega)}\right). \tag{2.57}$$

On the other hand, one has $\mathbb{D}^n \ge T^N_{\mathcal{D}}(\underline{N}^n_{\mathcal{D}}, \underline{w}^{N,n}_{\mathcal{D}}, \underline{w}^{N,n}_{\mathcal{D}}) + T^P_{\mathcal{D}}(\underline{P}^n_{\mathcal{D}}, \underline{w}^{P,n}_{\mathcal{D}}, \underline{w}^{P,n}_{\mathcal{D}})$, and, for any $K \in \mathcal{M}$, using (2.22) alongside with the positivity of $r_K(\underline{N}_K)$ and the local coercivity (2.15) of a^N_K, one gets that

$$T^N_K(\underline{N}^n_K, \underline{w}^{N,n}_K, \underline{w}^{N,n}_K) \ge r_K(\underline{N}^n_K)a^N_K(\underline{w}^{N,n}_K, \underline{w}^{N,n}_K) \ge \alpha_\flat\lambda_\flat r_K(\underline{N}^n_K)|\underline{w}^{N,n}_K|^2_{1,K}.$$

Moreover, by definition of r_K, properties on m and m_h and the bounds (2.36) on $\underline{N}^n_{\mathcal{D}}$, we deduce that

$$r_K(\underline{N}^n_K) = \sum_{\sigma\in\mathcal{E}_K}\frac{m(N^n_K, N^n_\sigma)}{|\mathcal{E}_K|} \ge \sum_{\sigma\in\mathcal{E}_K}\frac{m_h(N^n_K, N^n_\sigma)}{|\mathcal{E}_K|} \ge \sum_{\sigma\in\mathcal{E}_K}\frac{m_h(M_\flat, M_\flat)}{|\mathcal{E}_K|} \ge M_\flat.$$

The same holds for the term $T^P_{\mathcal{D}}(\underline{P}^n_{\mathcal{D}}, \underline{w}^{P,n}_{\mathcal{D}}, \underline{w}^{P,n}_{\mathcal{D}})$, therefore by the discrete Poincaré inequality (2.17) one has

$$\mathbb{D}^n \ge \alpha_\flat\lambda_\flat M_\flat\left(|\underline{w}^{N,n}_{\mathcal{D}}|^2_{1,\mathcal{D}} + |\underline{w}^{P,n}_{\mathcal{D}}|^2_{1,\mathcal{D}}\right) \ge \frac{\alpha_\flat\lambda_\flat M_\flat}{C^2_{P,\Gamma^D}}\left(\|w^{N,n}_{\mathcal{M}}\|^2_{L^2(\Omega)} + \|w^{P,n}_{\mathcal{M}}\|^2_{L^2(\Omega)}\right). \tag{2.58}$$

Combining (2.57) and (2.58), we get the following control on the entropy:

$$\nu_{\mathcal{D},\Delta t}\mathbb{E}^n \le \mathbb{D}^n \text{ with } \nu_{\mathcal{D},\Delta t} = \frac{\alpha_\flat\lambda_\flat M_\flat c_1^2}{c_2 C^2_{P,\Gamma^D}}.$$

From this inequality and the entropy dissipation relation (2.33), we deduce (2.53). To get (2.54), notice that $c_1\hat{\mathbb{E}}^n \le \mathbb{E}^n$, so using the discrete Poincaré inequality alongside with the coercivity of

a_D^ϕ, we get

$$\|N_\mathcal{M}^n - N_\mathcal{M}^e\|_{L^2(\Omega)}^2 + \|P_\mathcal{M}^n - P_\mathcal{M}^e\|_{L^2(\Omega)}^2 + \frac{\alpha_b \lambda_b^\phi}{2C_{P,\Gamma^D}^2}\|\phi_\mathcal{M}^n - \phi_\mathcal{M}^e\|_{L^2(\Omega)}^2 \le \hat{\mathbb{E}}^n \le \frac{1}{c_1}\mathbb{E}^n.$$

We conclude by using (2.53). □

Remark 17 (Non-uniformity with respect to the mesh)**.** *The result of Theorem 9 states a geometric convergence of the solutions to the scheme towards the discrete thermal equilibrium. However, the constants involved in (2.53) and (2.54) may depend strongly on the mesh (and in particular on h_D), which yelds a weaker result than the one of [27, 28] for TPFA schemes. Such a dependency is a reminiscence of the non-monotonicity of the HFV schemes (see Remark 13). In practice, the long-time behaviour of the scheme seems not to depend on the meshsize: see Section 2.5.3.*

2.5 Numerical results

In this section, we give some numerical evidences of the good behaviour of the scheme (2.28). We use test-cases inspired by the 2D PN-junction studied in [27], whose geometry is described in Figure 2.2. The domain Ω is the unit square $]0,1[^2$. For the boundary conditions, we split $\Gamma^D = \Gamma_0^D \cup \Gamma_1^D$ with $\Gamma_0^D = [0,1]\times\{0\}$ and $\Gamma_1^D = [0,0.25]\times\{1\}$. For $i \in \{0,1\}$, we let

$$N^D = N_i^D,\ P^D = P_i^D \text{ and } \phi^D = \frac{h(N_i^D) - h(P_i^D)}{2} \text{ on } \Gamma_i^D.$$

To be consistent with the compatibility condition (2.4) we assume that there exists a constant α_0 such that $h(N^D) + h(P^D) = \alpha_0$, therefore for given N^D and α_0 we set $P^D = g\left(\alpha_0 - h(N^D)\right)$ on Γ^D. Thus, we get $\alpha_N = \alpha_P = \frac{\alpha_0}{2}$. If $r \neq 0$, we finally impose that $\alpha_0 = 0$ to satisfy (2.5). We use the

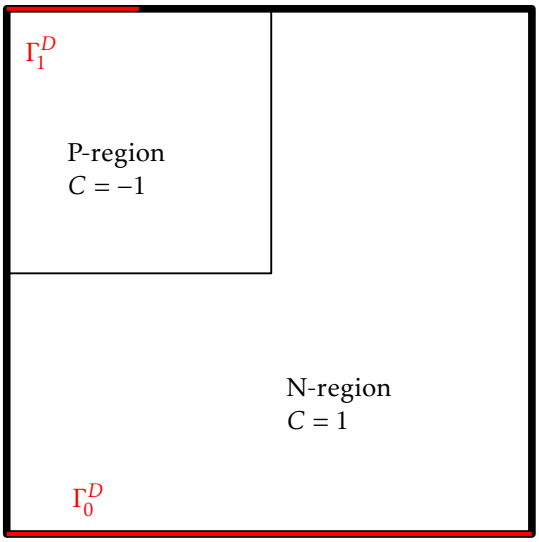


Figure 2.2 – The geometry of the PN diode.

following initial condition:

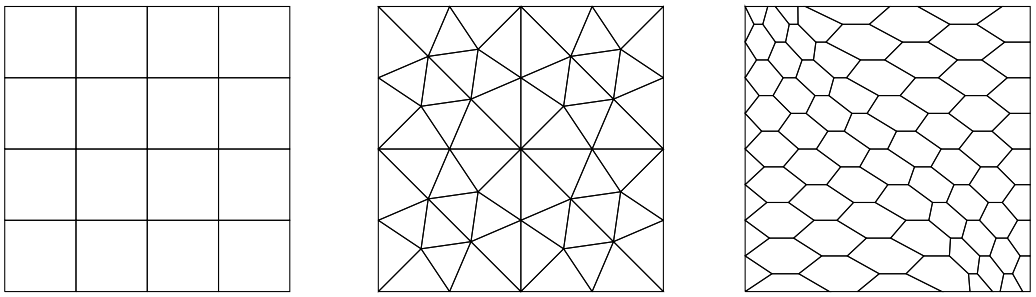
$$N_0(x,y) = N_1^D + (N_0^D - N_1^D)(1-\sqrt{y}) \text{ and } P_0(x,y) = P_1^D + (P_0^D - P_1^D)(1-\sqrt{y}).$$

Finally, we use a piecewise constant doping profile C, equal to -1 in the P-region and 1 in the N-region (see Figure 2.2). Concerning the tensors, we assume that the permittivity is isotropic, of the form $\Lambda_\phi = \lambda^2 I_2$, where $\lambda > 0$ is the rescaled Debye length. We also assume that the magnetic field is constant, of magnitude $b \geq 0$, and that the rescaled mobilities are equal to 1 ($\mu_N = \mu_P = 1$), therefore the tensors for the convection-diffusion equations are

$$\Lambda_N = \frac{1}{1+b^2}\begin{pmatrix} 1 & b \\ -b & 1 \end{pmatrix} \text{ and } \Lambda_P = \frac{1}{1+b^2}\begin{pmatrix} 1 & -b \\ b & 1 \end{pmatrix}. \tag{2.59}$$

2.5.1 Implementation

In this section, we discuss some practical details concerning the implementation of the schemes introduced in this paper. The mesh families used for the numerical tests are classical Cartesian and triangular families - see for example [154]- and a tilted hexagonal-dominant mesh family, depicted on Figure 2.3. These meshes have convex cells, hence we always choose x_K to be



(a) Cartesian mesh (b) Triangular mesh (c) Tilted hexagonal mesh

Figure 2.3 – **Implementation.** Coarsest meshes of each family used in the numerical tests.

the barycentre of K. Moreover, we use a fixed stabilisation parameter $\eta = 1.5$ (see (2.13)). In the simulations showed below, we use the arithmetic mean as an m function for the reconstruction defined in (2.23), therefore for $\underline{u}_\mathcal{D} \in \underline{V}_\mathcal{D}$ and $K \in \mathcal{M}$,

$$r_K(\underline{u}_K) = \frac{1}{2}\left(u_K + \frac{1}{|\mathcal{E}_K|}\sum_{\sigma\in\mathcal{E}_K} u_\sigma\right).$$

For a discussion on other choices of reconstructions we refer to [51, Section 6.2].

Finite volume formulation

The hybrid schemes described in this paper define finite volume methods, in the sense that they can be equivalently rewritten under a conservative forms, with local mass balances, flux equilibration at interfaces, and boundary conditions. For more details about this formulation in the framework of a linear Poisson equation, we refer the reader to [123]. Let us detail succinctly this formulation for our schemes. Let Λ be a generic uniformly elliptic tensor. For all $K \in \mathcal{M}$,

and all $\sigma \in \mathcal{E}_K$, in the framework of HFV methods, the normal diffusive flux $-\int_\sigma \Lambda \nabla u \cdot n_{K,\sigma}$ is approximated by the following numerical flux:

$$F^{\Lambda}_{K,\sigma}(\underline{u}_K) = \sum_{\sigma' \in \mathcal{E}_K} A^{\sigma\sigma'}_K (u_K - u_{\sigma'}), \tag{2.60}$$

where the $A^{\sigma\sigma'}_K$ are defined by

$$A^{\sigma\sigma'}_K = \sum_{\sigma'' \in \mathcal{E}_K} |P_{K,\sigma''}| y^{\sigma''\sigma}_K \cdot \Lambda_{K,\sigma''} y^{\sigma''\sigma'}_K, \tag{2.61}$$

and the $y^{\sigma\sigma'}_K \in \mathbb{R}^d$ only depend on the geometry of the discretisation $\mathcal{D}$ (see, for example, [123, Eq. (2.22)] for an exact definition with $\eta = \sqrt{d}$). For all $K \in \mathcal{M}$, one can express the local discrete bilinear form a^{Λ}_K in terms of the local fluxes $\left(F^{\Lambda}_{K,\sigma}\right)_{\sigma \in \mathcal{E}_K}$: for all $(\underline{u}_K, \underline{v}_K) \in \underline{V}^2_K$,

$$a^{\Lambda}_K(\underline{u}_K, \underline{v}_K) = \sum_{\sigma \in \mathcal{E}_K} F^{\Lambda}_{K,\sigma}(\underline{u}_K)(v_K - v_\sigma).$$

With these considerations, the scheme (2.31) for the Poisson equation (2.30) writes under the following form:

$$\begin{cases} \displaystyle\sum_{\sigma \in \mathcal{E}_K} F^{\Lambda_\phi}_{K,\sigma}(\underline{\phi}_K) = |K|\Big(C_K + g(z^P_K - \phi_K) - g(z^N_K + \phi_K)\Big) & \forall K \in \mathcal{M}, \quad (2.62a) \\ F^{\Lambda_\phi}_{K,\sigma}(\underline{\phi}_K) = -F^{\Lambda_\phi}_{L,\sigma}(\underline{\phi}_L) & \forall \sigma = K|L \in \mathcal{E}_{int}, \quad (2.62b) \\ \phi_\sigma = \phi^D_\sigma & \forall \sigma \in \mathcal{E}^D_{ext}, \quad (2.62c) \\ F^{\Lambda_\phi}_{K,\sigma}(\underline{\phi}_K) = 0 & \forall \sigma \in \mathcal{E}^N_{ext} \text{ with } \mathcal{M}_\sigma = \{K\}, \quad (2.62d) \end{cases}$$

where the fluxes are defined by (2.60), and $C_K = \frac{1}{|K|}\int_K C$. The first equation corresponds to local balance, the second imposes the local conservativity of the fluxes at interfaces and the last one enforces the boundary conditions.

Following an analogous approach, we define the nonlinear flux for the advection-diffusion: for any I_h-valued $\underline{u}_K \in \underline{V}_K$ and $\underline{\phi}_K \in \underline{V}_K$, we let

$$\mathcal{F}^{\Lambda}_{K,\sigma}(\underline{u}_K, \underline{\phi}_K) = r_K(\underline{u}_K) F^{\Lambda}_{K,\sigma}\left(h(\underline{u}_K) + \underline{\phi}_K\right). \tag{2.63}$$

Therefore, letting $\underline{w}_K = h(\underline{u}_K) + \underline{\phi}_K - \alpha \underline{1}_K$ (with $\alpha \in \mathbb{R}$), one can write the local trilinear form T^{Λ}_K in terms of the nonlinear flux: for any $\underline{v}_K \in \underline{V}_K$,

$$T^{\Lambda}_K(\underline{u}_K, \underline{w}_K, \underline{v}_K) = \sum_{\sigma \in \mathcal{E}_K} \mathcal{F}^{\Lambda}_{K,\sigma}(\underline{u}_K, \underline{\phi}_K)(v_K - v_\sigma).$$

The scheme (2.28) for the drift-diffusion system then writes under the following form:

$$\forall K \in \mathcal{M}, \qquad |K|\frac{N_K^{n+1}-N_K^n}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} \mathcal{F}_{K,\sigma}^{\Lambda_N}(\underline{N}_K^{n+1}, -\underline{\phi}_K^{n+1}) = -|K|R(N_K^{n+1}, P_K^{n+1}), \tag{2.64a}$$

$$\forall K \in \mathcal{M}, \qquad |K|\frac{P_K^{n+1}-P_K^n}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} \mathcal{F}_{K,\sigma}^{\Lambda_P}(\underline{P}_K^{n+1}, \underline{\phi}_K^{n+1}) = -|K|R(N_K^{n+1}, P_K^{n+1}), \tag{2.64b}$$

$$\forall K \in \mathcal{M}, \qquad \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{\Lambda_\phi}(\underline{\phi}_K^{n+1}) = |K|\Big(C_K + P_K^{n+1} - N_K^{n+1}\Big), \tag{2.64c}$$

$$\forall \sigma = K|L \in \mathcal{E}_{int}, \qquad \mathcal{F}_{K,\sigma}^{\Lambda_N}(\underline{N}_K^{n+1}, -\underline{\phi}_K^{n+1}) + \mathcal{F}_{L,\sigma}^{\Lambda_N}(\underline{N}_L^{n+1}, -\underline{\phi}_L^{n+1}) = 0, \tag{2.64d}$$

$$\forall \sigma = K|L \in \mathcal{E}_{int}, \qquad \mathcal{F}_{K,\sigma}^{\Lambda_P}(\underline{P}_K^{n+1}, \underline{\phi}_K^{n+1}) + \mathcal{F}_{L,\sigma}^{\Lambda_P}(\underline{P}_L^{n+1}, \underline{\phi}_L^{n+1}) = 0, \tag{2.64e}$$

$$\forall \sigma = K|L \in \mathcal{E}_{int}, \qquad F_{K,\sigma}^{\Lambda_\phi}(\underline{\phi}_K^{n+1}) + F_{L,\sigma}^{\Lambda_\phi}(\underline{\phi}_L^{n+1}) = 0, \tag{2.64f}$$

$$\forall \sigma \in \mathcal{E}_{ext}^D, \qquad N_\sigma^{n+1} = N_\sigma^D,\ P_\sigma^{n+1} = P_\sigma^D \text{ and } \phi_\sigma^{n+1} = \phi_\sigma^D, \tag{2.64g}$$

$$\forall \sigma \in \mathcal{E}_{ext}^N, \qquad \mathcal{F}_{K,\sigma}^{\Lambda_N}(\underline{N}_K^{n+1}, -\underline{\phi}_K^{n+1}) = \mathcal{F}_{K,\sigma}^{\Lambda_P}(\underline{P}_K^{n+1}, \underline{\phi}_K^{n+1}) = F_{K,\sigma}^{\Lambda_\phi}(\underline{\phi}_K^{n+1}) = 0, \tag{2.64h}$$

where the nonlinear fluxes are defined by (2.63), the initial data $(N_K^0, P_K^0)_{K\in\mathcal{M}}$ are defined as in (2.28g) and the cell K in (2.64h) is such that $\mathcal{M}_\sigma = \{K\}$.

Note that these formulations yield nonlinear systems. Thus, we can introduce natural functions $\mathcal{G}^{sta} : \underline{V}_\mathcal{D} \to \underline{V}_\mathcal{D}$ and $\mathcal{G}^{n,\Delta t} : \underline{V}_\mathcal{D}^3 \to \underline{V}_\mathcal{D}^3$ such that (2.62) rewrites $\mathcal{G}^{sta}(\underline{\phi}_\mathcal{D}) = \underline{0}_\mathcal{D}$ and (2.64) writes $\mathcal{G}^{n,\Delta t}(\underline{N}_\mathcal{D}^{n+1}, \underline{P}_\mathcal{D}^{n+1}, \underline{\phi}_\mathcal{D}^{n+1}) = (\underline{0}_\mathcal{D}, \underline{0}_\mathcal{D}, \underline{0}_\mathcal{D})$. The two functions are regular on their domains.

Newton's method and static condensation

The implementation of the nonlinear schemes relies on their finite volume formulation. To fix ideas, we consider the case of the transient scheme (2.64). Given $(\underline{N}_\mathcal{D}^n, \underline{P}_\mathcal{D}^n) \in \underline{V}_\mathcal{D}^2$ I_h-valued, we want to solve the nonlinear system $\mathcal{G}^{n,\delta t}(\underline{N}_\mathcal{D}^{n+1}, \underline{P}_\mathcal{D}^{n+1}, \underline{\phi}_\mathcal{D}^{n+1}) = (\underline{0}_\mathcal{D}, \underline{0}_\mathcal{D}, \underline{0}_\mathcal{D})$ (with a time step δt instead of Δt). The resolution of this system relies on a Newton method with time step adaptation.

First, one initialises the method with initial guess $(\underline{\tilde{N}}_\mathcal{D}^n, \underline{\tilde{P}}_\mathcal{D}^n) \in \underline{V}_\mathcal{D}^2$, where the coordinates of $\underline{\tilde{N}}_\mathcal{D}^n$ (respectively $\underline{\tilde{P}}_\mathcal{D}^n$) are the projections of the coordinates of $\underline{N}_\mathcal{D}^n$ (resp. $\underline{P}_\mathcal{D}^n$) on $[\epsilon, a-\epsilon]$ (if $a = +\infty$, we project on $[\epsilon, +\infty[$) in order to avoid potential problems due to the singularity of h near 0 and a.
The computation of the residues follows the process described below: let us denote by $R_\mathcal{M} \in \mathbb{R}^{3|\mathcal{M}|}$ and $R_\mathcal{E} \in \mathbb{R}^{3|\mathcal{E}|}$ the residue vectors $(r_K^N, r_K^P, r_K^\phi)_{K\in\mathcal{M}}$ and $(r_\sigma^N, r_\sigma^P, r_\sigma^\phi)_{\sigma\in\mathcal{E}}$. They are solution to the following linear block system:

$$\begin{pmatrix} \mathbb{M}_\mathcal{M} & \mathbb{M}_{\mathcal{M},\mathcal{E}} \\ \mathbb{M}_{\mathcal{E},\mathcal{M}} & \mathbb{M}_\mathcal{E} \end{pmatrix} \begin{pmatrix} R_\mathcal{M} \\ R_\mathcal{E} \end{pmatrix} = \begin{pmatrix} S_\mathcal{M} \\ S_\mathcal{E} \end{pmatrix}, \tag{2.65}$$

where $\mathbb{M}_\mathcal{M} \in \mathbb{R}^{3|\mathcal{M}|\times 3|\mathcal{M}|}$, $\mathbb{M}_{\mathcal{M},\mathcal{E}} \in \mathbb{R}^{3|\mathcal{M}|\times 3|\mathcal{E}|}$, $\mathbb{M}_{\mathcal{E},\mathcal{M}} \in \mathbb{R}^{3|\mathcal{E}|\times 3|\mathcal{M}|}$, $\mathbb{M}_\mathcal{E} \in \mathbb{R}^{3|\mathcal{E}|\times 3|\mathcal{E}|}$, and $S_\mathcal{M}$ and $S_\mathcal{E}$ are vectors of size $3|\mathcal{M}|$ and $3|\mathcal{E}|$ issued from the previous iteration. By construction, the matrix $\mathbb{M}_\mathcal{M}$ is block diagonal with 3×3 diagonal blocks, which are expected to be invertible. Therefore, this matrix can be inverted at a very low computational cost, inverting only small matrices. Thus, we

can eliminate the cell unknowns, noticing that

$$R_{\mathcal{M}} = \mathbb{M}_{\mathcal{M}}^{-1}\left(S_{\mathcal{M}} - \mathbb{M}_{\mathcal{M},\mathcal{E}} R_{\mathcal{E}}\right). \tag{2.66}$$

Using this relation, one shows that $R_{\mathcal{E}}$ is the solution to the following linear system:

$$\left(\mathbb{M}_{\mathcal{E}} - \mathbb{M}_{\mathcal{E},\mathcal{M}}\mathbb{M}_{\mathcal{M}}^{-1}\mathbb{M}_{\mathcal{M},\mathcal{E}}\right) R_{\mathcal{E}} = S_{\mathcal{E}} - \mathbb{M}_{\mathcal{E},\mathcal{M}}\mathbb{M}_{\mathcal{M}}^{-1} S_{\mathcal{M}}, \tag{2.67}$$

where $\mathbb{M}_{\mathcal{D}} = \mathbb{M}_{\mathcal{E}} - \mathbb{M}_{\mathcal{E},\mathcal{M}}\mathbb{M}_{\mathcal{M}}^{-1}\mathbb{M}_{\mathcal{M},\mathcal{E}}$, the Schur complement of the block $\mathbb{M}_{\mathcal{M}}$, is an invertible matrix of size $3|\mathcal{E}| \times 3|\mathcal{E}|$. In practice, we solve the linear system (2.67) using an LU factorization algorithm, and we use the solution $R_{\mathcal{E}}$ in order to compute $R_{\mathcal{M}}$ from (2.66). This technique, called static condensation, allows one to replace a system of size $3(|\mathcal{E}| + |\mathcal{M}|)$ by a system of size $3|\mathcal{E}|$ without any additional fill-in. As a stopping criterion for the Newton iterations, we compare the l^∞ relative norm of the residue with a threshold tol. If the method does not converge after i_{max} iterations, we divide the time step by 2 and restart the resolution. When the method converges, one can proceed to the approximation of $(\underline{N}_{\mathcal{D}}^{n+2}, \underline{P}_{\mathcal{D}}^{n+2}, \underline{\phi}_{\mathcal{D}}^{n+2})$, with an initial time step of $\min(\Delta t, 1.4 \times \delta t)$.
The initial time step (used to compute $(\underline{N}_{\mathcal{D}}^{1}, \underline{P}_{\mathcal{D}}^{1}, \underline{\phi}_{\mathcal{D}}^{1})$) is Δt. In practice, we use $\epsilon = 10^{-9}$, $i_{max} = 50$ and $tol = 10^{-10}$.

For the discrete thermal equilibrium, we solve the nonlinear system $\mathcal{G}^{sta}(\underline{\phi}_{\mathcal{D}}) = \underline{0}_{\mathcal{D}}$ (with $(z^P, z^N) = (\alpha_N, \alpha_P)$) using a similar approach, based on a Newton's method alongside with a continuation method. Moreover, note that the counterpart in this case of the matrix $\mathbb{M}_{\mathcal{M}}$ is diagonal with non-zero entries, so its inversion is straightforward.

2.5.2 Proof of concept

In this section, we are interested in qualitative and quantitative properties of the discrete solutions, and we present the profiles of some computed discrete solutions.

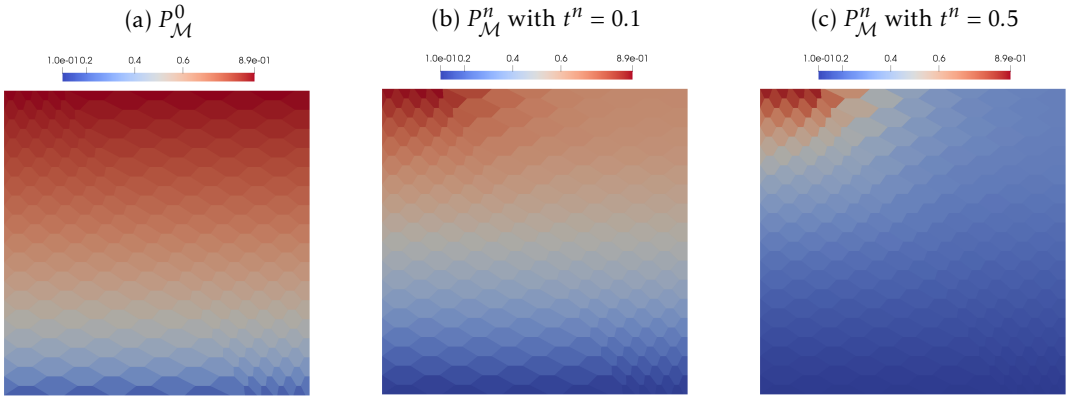


Figure 2.4 – **Test-case 1**. Evolution of the discrete density of holes

For the test-case 1, we use a set of data introduced in [68], with Boltzmann statistics ($h = \log$), no recombination ($r = 0$), no magnetic field ($b = 0$), $\lambda = 1$, boundary values $N_0^D = 0.9$, $N_1^D = 0.1$ and $\alpha_0 = \log\left(N_0^D \times N_1^D\right)$. We perform a simulation on a tilted hexagonal-dominant mesh constituted of 280 cells, with $\Delta t = 0.1$. On Figure 2.4, we show the profiles of $P_{\mathcal{M}}^n$ for different

values of n. The evolution is in accordance with results obtained in the TPFA context (see [68], and notice that we use a different initial condition). Note that the discrete density remains positive, as expected. In fact, even on coarser meshes, the positivity of the densities is preserved (both for the cell and edge values).

Now, we want to assert the robustness of the scheme with respect to the h function and the anisotropy. We consider the test-case 2, with Blakemore statistics ($h(s) = \log\left(\frac{s}{1-\gamma s}\right)$, $\gamma = 0.27$),

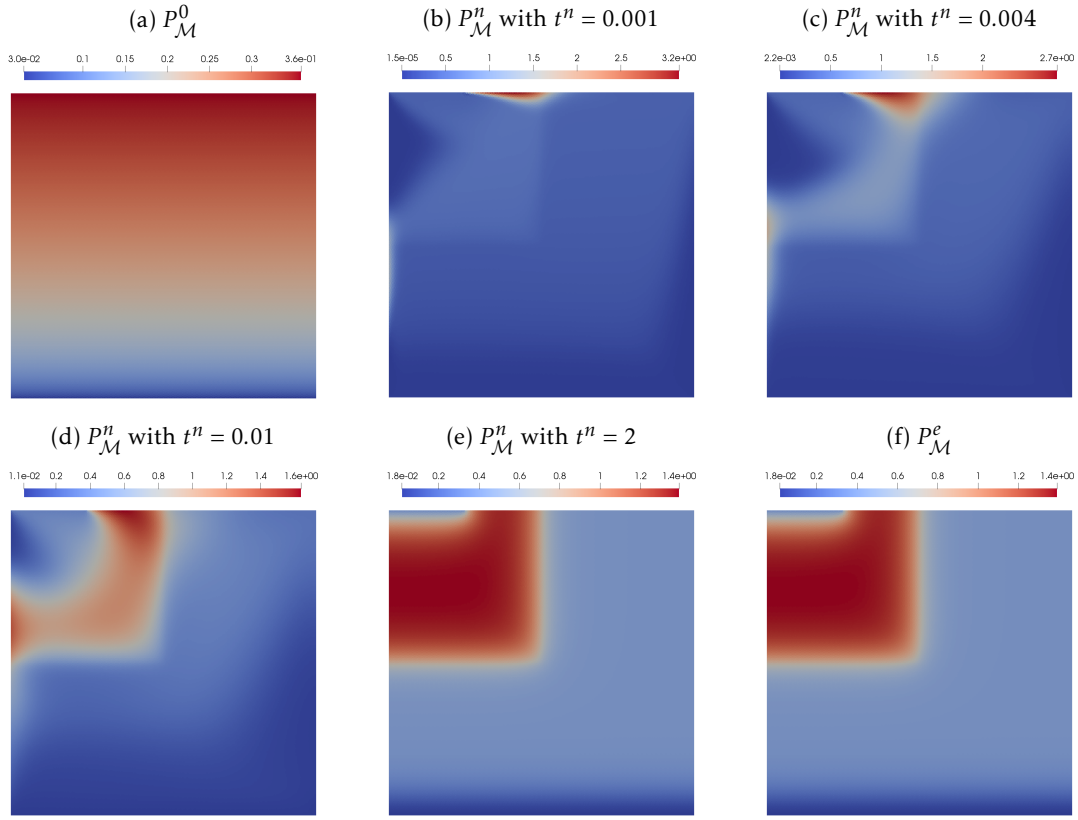


Figure 2.5 – **Test-case 2** ($b = 1$). Evolution of the discrete density of holes (note that the scale varies from a figure to the other)

a SRH recombination term ($r(N,P) = \frac{10}{1+N+P}$) and a strong magnetic field $b = 1$. We also use a (realistic) small Debye length $\lambda = 0.05$. In order to check the fact that the discrete densities are I_h-valued (here, $a = 1/\gamma \approx 3.7$), we consider boundary values close to the maximum admissible densities: $N_0^D = 3.5$ and $N_1^D = 1.5$. Since $r \neq 0$, we let $\alpha_0 = 0$. We perform a simulation on a refined triangular mesh (57 344 cells), with a time step $\Delta t = 0.1$. We show the profile of $P^n_{\mathcal{M}}$ for different values of n in Figure 2.5. Notice that this test-case is subject to boundary layers (essentially because λ is small and b is big, see the discussion below and Table 2.1), therefore the scheme performs numerous adaptations of the time step at the beginning of the simulation (the first admissible time step is $\delta t \approx 10^{-8}$). Moreover, one can see a rotation movement for the density of the holes, which is in accordance with the expected physical effect of the magnetic field.

To give more quantitative informations, we show in Figure 2.6 the evolution of the minimal

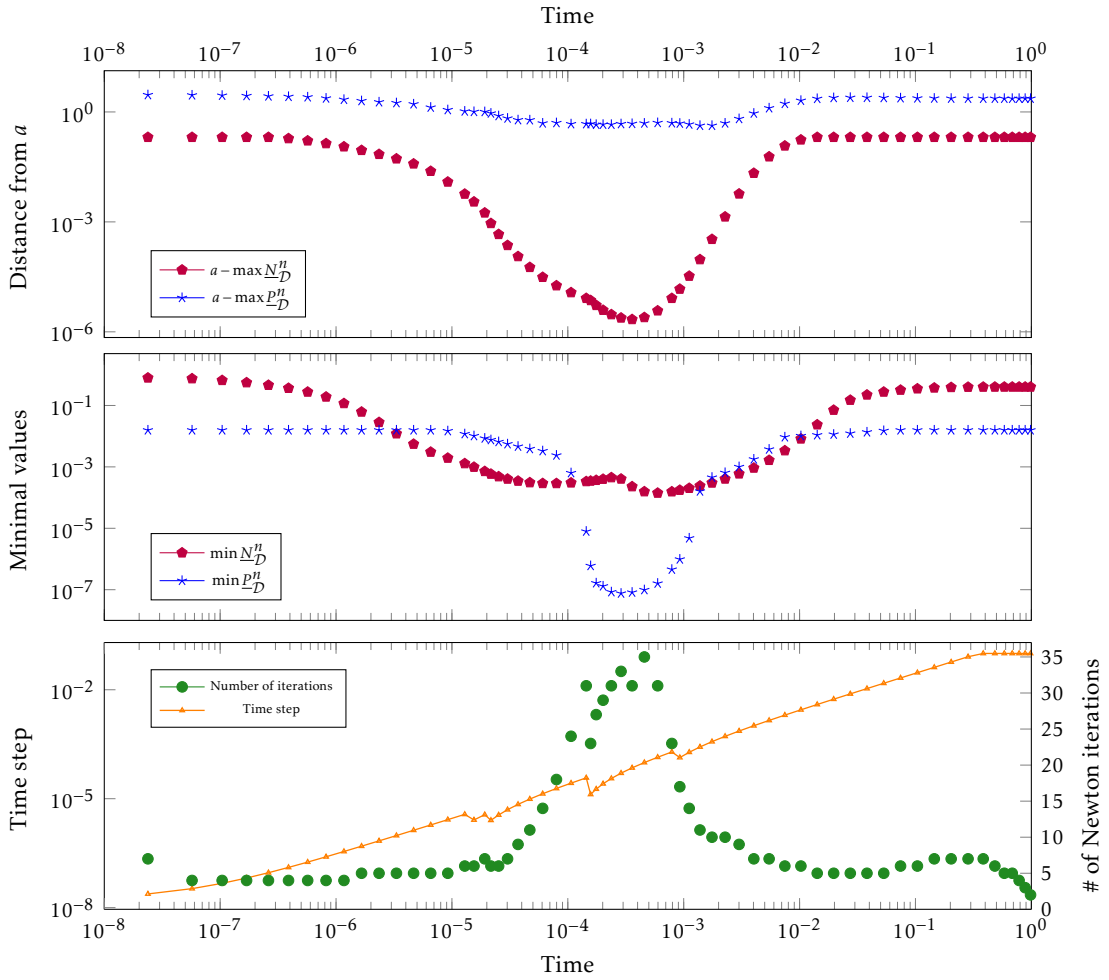


Figure 2.6 – **Test-case 2** ($b = 1$). Evolution of the discrete extremal values, time step and cost

and maximal values, along with the time step and the number of Newton's iterations needed to compute the solutions at a given time. First, note that we display each of the steps (from t^n to t^{n+1}) used to compute the solutions on the time interval $[0,1]$, there are 63 of them for this simulation. The extremal values account for the cell and edge unknowns: we let

$$\min \underline{N}_{\mathcal{D}}^n = \min\left(\min_{K\in\mathcal{M}} N_K^n, \min_{\sigma\in\mathcal{E}} N_\sigma^n\right) \text{ and } \max \underline{N}_{\mathcal{D}}^n = \max\left(\max_{K\in\mathcal{M}} N_K^n, \max_{\sigma\in\mathcal{E}} N_\sigma^n\right),$$

as well as analogous definitions for the holes density. In order to observe the extremal values on relevant scales, we look at the minimum and the distance between a (the maximal density allowed by the model) and the maximal densities computed, and print them on a log scale. As expected from Theorem 7, the values of the discrete densities stay in $I_h =]0, a[$. In particular, our scheme is not subject to the same lack of positivity as the scheme of [143] when the magnetic field is intense. Moreover, it is remarkable to note that the scheme seems very robust: the densities are close to the limit values, at a distance reaching 7.56e-8 in the most difficult situation. Note that

this value mean that in practice, we do not perform the projection step described in Section 2.5.1 to initialise the Newton's methods, since the projection threshold $\epsilon = 10^{-9}$ is smaller than 7.56e-8. Using larger values for ϵ could perhaps improve the convergence of Newton's methods, but the impact of such a modification has not been investigated here. As previously announced, one can see that the first effective time step needed to compute $(\underline{N}_{\mathcal{D}}^1, \underline{P}_{\mathcal{D}}^1, \underline{\phi}_{\mathcal{D}}^1)$ is relatively small: $0.1 \times 2^{-22} \approx 2.38\text{e-}8$. It means that the time step adaptation procedure needed to perform 22 time step reductions before managing to compute a solution. On the other hand, we can see that after this initial reduction, there are only five others time step reductions (one around 1.5e-5, one at 2e-5, two around 1.5e-4, and a last one at 9e-4). At the end of the simulation, around $t = 0.3$, the time step reach its maximal value $\Delta t = 0.1$ and there is no more time step adaptation. On the third graph, we show the number of Newton iterations needed to compute the discrete solution from one time to another. Note that we do not take into account the iterations used in non-convergent Newton's methods (i.e. methods that leads to a time step reduction). For all the time step reductions of this simulation, the Newton's methods do not converge because at least one of the computed discrete densities was not I_h-valued. On average, the convergent Newton's methods converge in 10.5 iterations. Finally, we can see on Figure 2.6 a clear correlation between the extremal values reached by the discrete densities and the number of Newton iterations needed to compute the solution. This can be explained by at least one fact: since h is singular in 0 and a, the values of its derivative near these limit values blow up. Therefore, the Jacobian matrix used in the Newton's method tends to be ill-conditioned, which induces numerical instabilities. We can give a last remark on this test-case: the most extreme values observed in Figure 2.6 are located near to the boundary and appear at the beginning of the simulation (see for example the minimal values on Figure 2.5b, on the left and right sides of the square). Hence, the difficulties are essentially due to the presence of the (strong) magnetic field, which induces rotation of the charges and creates boundary layers.

Magnetic field intensity	$b = 0$	$b = 0.5$	$b = 1$
$\min\left\{\min \underline{N}_{\mathcal{D}}^n \mid 0 \le t^n \le 1\right\}$	3.20e-1	5.36e-3	1.41e-4
$\min\left\{\min \underline{P}_{\mathcal{D}}^n \mid 0 \le t^n \le 1\right\}$	1.56e-2	2.53e-3	7.56e-8
$\min\left\{a - \max \underline{N}_{\mathcal{D}}^n \mid 0 \le t^n \le 1\right\}$	2.04e-1	4.17e-3	2.18e-6
$\min\left\{a - \max \underline{P}_{\mathcal{D}}^n \mid 0 \le t^n \le 1\right\}$	2.31	9.39e-1	4.23e-2
Number of steps	18	40	63
Minimal time step	3.13e-3	1.53e-6	2.38e-8
Number of initial time step reductions	5	16	22
Total number of time step reductions	5	16	27
Maximal number of Newton iterations	7	7	35
Average number of Newton iterations	5.05	5.45	10.54
Total cost	96	234	692

Table 2.1 – **Test-case 2**. Comparison of the extremal values and costs for different magnetic fields

To confirm this statement, we perform simulations with the same parameters (still on the time interval $[0,1]$) except that we consider situations with smaller magnetic field intensities ($b = 0$ and $b = 0.5$). The results are presented in Table 2.1. As expected, it seems that the extremal values are strongly related to the intensity of the magnetic field. It follows that the number of Newton iterations, and hence the global computational cost of the simulation increase with the intensity of the magnetic field. We can also notice that the time step reductions occur only at the

first time step for the moderate magnetic fields ($b = 0$ and $b = 0.5$): the computational difficulties lie in the boundary layers appearing at small times. As previously, note that the "number of iterations" mentioned on the table do not take into account the iterations of non-convergent Newton's methods. In the last line, we give the "Total cost" of the simulation, that is to say the number of linear systems solved during the simulation, including the iterations of non-convergent Newton's methods. Taking into account every iterations performed, the simulation with $b = 1$ is basically 7 times more time consuming than the one without magnetic field.

In order to assert the robustness of the method with respect to the mesh, we consider the test case 3, which is characterised by the same physical parameters as the previous one: $h(s) = \log\left(\frac{s}{1-\gamma s}\right)$, $\gamma = 0.27$), $r(N,P) = \frac{10}{1+N+P}$, $b = 1$, $\lambda = 0.05$, $N_0^D = 3.5$, $N_1^D = 1.5$ and $\alpha_0 = 0$. Contrarily to the previous test, we perform a simulation on a tilted hexagonal mesh, constituted of 4192 cells, with a time step $\Delta t = 0.1$. One has to notice that the spatial mesh used here is a "general polygonal" one, in the sense that it is not a admissible mesh for the TPFA scheme. Moreover, the geometry of this mesh is not well-suited with respect to the geometry of the device, since the junction of the PN-diode (which corresponds to the discontinuity of the doping profile C) crosses some cells. On Figure 2.7, we show the profile of the density of holes $P_{\mathcal{M}}$ computed

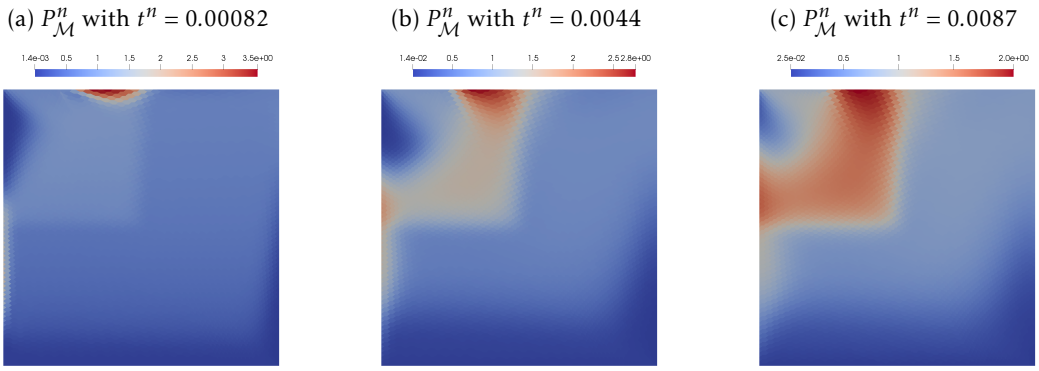


Figure 2.7 – **Test-case 3**. Evolution of the discrete density of holes, on a tilted hexagonal mesh.

at different times. These profiles are to be compared with these of the Figure 2.5: even if the tilted hexagonal does not fit the geometry of the problem perfectly and is much coarser than the triangular mesh used for the Figure 2.5, the profiles look quite the same.

We give quantitative information for this test case on Figure 2.8, which is the counterpart of Figure 2.6: we show the evolution of the bounds of the discrete densities, as well as the time step and the number of Newton's iterations used for a given time step. First, note that the first admissible time step ($\delta t \approx 10^{-6}$) is bigger than the one on the refined triangular mesh. Moreover, the extremal values are much farther from the bounds of I_h than on the triangular mesh. This behaviour can be explained by the fineness of the triangular mesh, which enable the scheme to capture in a very accurate way the boundary layers: the continuous solution take very small/big values near to the boundary of the domain, but if one average these values on relatively big cells (such as the cells of the tilted hexagonal mesh), then the mean values are not so extreme. We can also remark that the number of iterations needed to compute one time step peaks around $t \approx 3.10^{-4}$, as in Figure 2.6: it corresponds to the time when the densities are closest to the bounds of I_h. Such fact enforces the hypothesis formulated before: the values of the densities have an important impact on the Jacobian entries and hence on the convergence of the Newton's method.

Overall, the behaviour of the scheme does not depend on the geometry of the mesh, and the

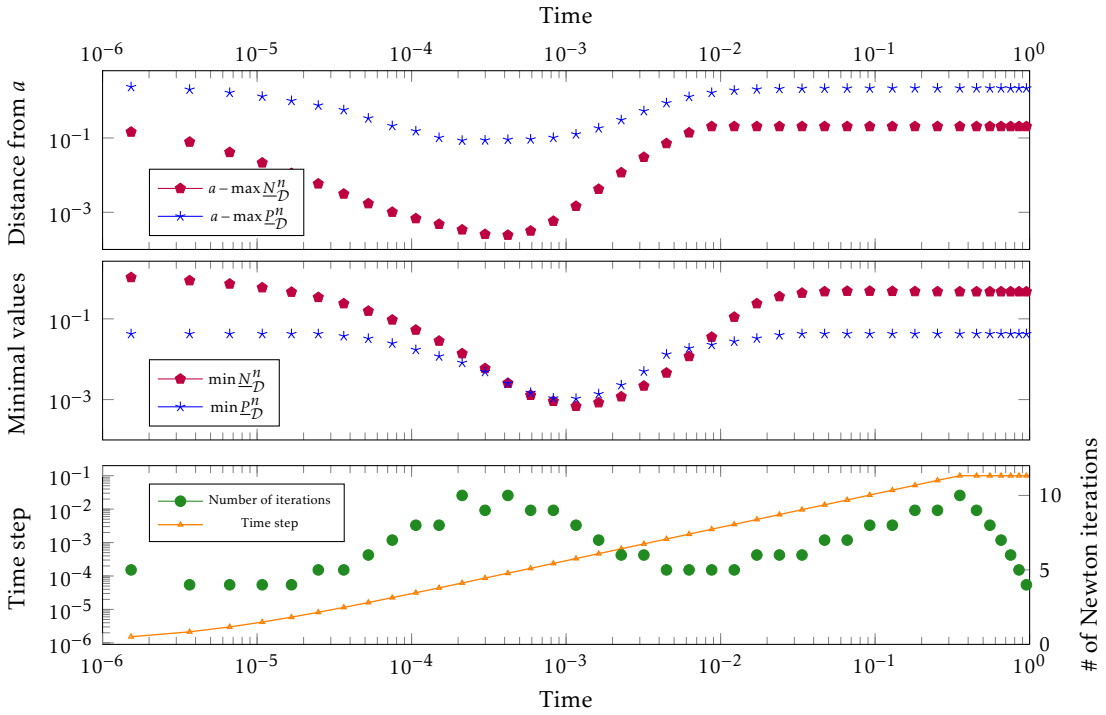


Figure 2.8 – **Test-case 3** ($b = 1$, tilted hexagonal mesh). Evolution of the discrete extremal values, time step and cost.

fact that the PN-junction crosses some cells does not have a noticeable impact.

2.5.3 Long-time behaviour of the discrete solutions

We are now interested in the long-time behaviour of the solutions computed with the scheme. For the test-case 4, we consider a test-case from [27] with Boltzmann statistics ($h = \log$), no recombination ($r = 0$), no magnetic field ($b = 0$), $\lambda = 1$ and boundary values $N_0^D = \mathrm{e}$, $N_1^D = 1$ and $\alpha_0 = 1$. We are interested in the the evolution of the discrete relative entropy and the L^2 distance to the equilibrium (namely, $\sqrt{\|N_{\mathcal{M}}^n - N_{\mathcal{M}}^e\|_{L^2(\Omega)}^2 + \|P_{\mathcal{M}}^n - P_{\mathcal{M}}^e\|_{L^2(\Omega)}^2 + \|\phi_{\mathcal{M}}^n - \phi_{\mathcal{M}}^e\|_{L^2(\Omega)}^2}$) with respect to time.

In Figure 2.9, we show the evolution of these quantities for simulations performed on the family of triangular meshes (the coarsest mesh has 56 cells, and a size h_0), with $\Delta t = 0.01$. First, one can note that the evolutions are exponentially fast, as expected from Theorem 9. One can also see a saturation phenomenon when the machine precision is reached. Moreover, as announced in Remark 17, the decay rate is not strongly impacted by the meshsize. Last but not least, the quantitative values of the decay rate are in accordance with those obtained in [27, Figure 2] with a Scharfetter–Gummel TPFA scheme.

In Figure 2.10, we show the evolution of the discrete relative entropy and the L^2 distance to the equilibrium for simulations computed with $\Delta t = 0.01$ on meshes with different geometry: a Cartesian one (64 cells, $h_{\mathcal{D}} = 3.12\ 10^{-2}$), a triangular one (56 cells, $h_{\mathcal{D}} = 3.07\ 10^{-2}$) and a tilted hexagonal one (76 cells, $h_{\mathcal{D}} = 3.42\ 10^{-2}$). Note that the meshes under consideration have a similar meshsize. As for the previous results, the expected exponential decay of entropy is

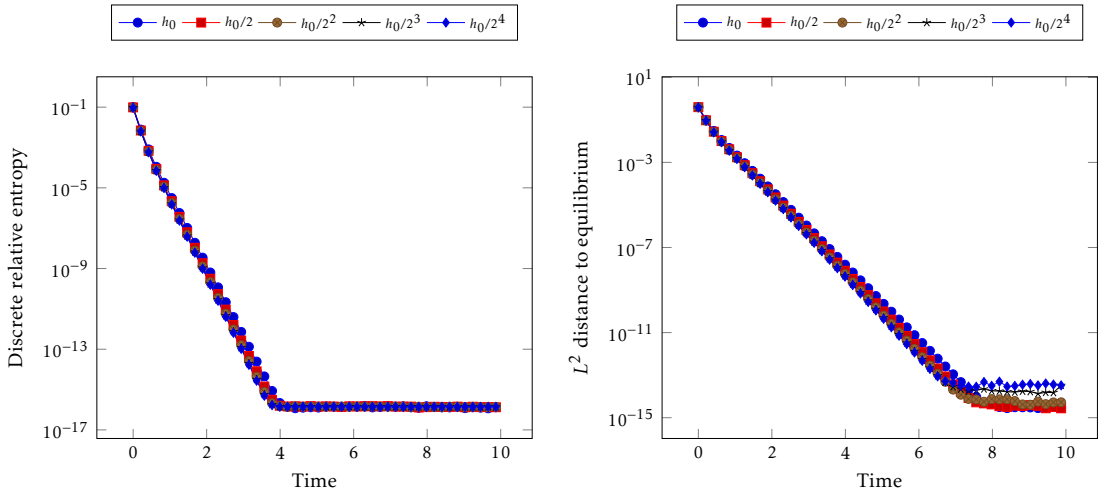


Figure 2.9 – **Test-case 4**. Influence of the meshsize: entropy and L^2 distance to equilibrium

observed. The decay rates computed are almost the same for the three mesh geometries, which reinforces the previous observations: in practice, the decay rate is not impacted by the mesh used for the simulation (either by its geometry or its size). It is also remarkable to note that on the

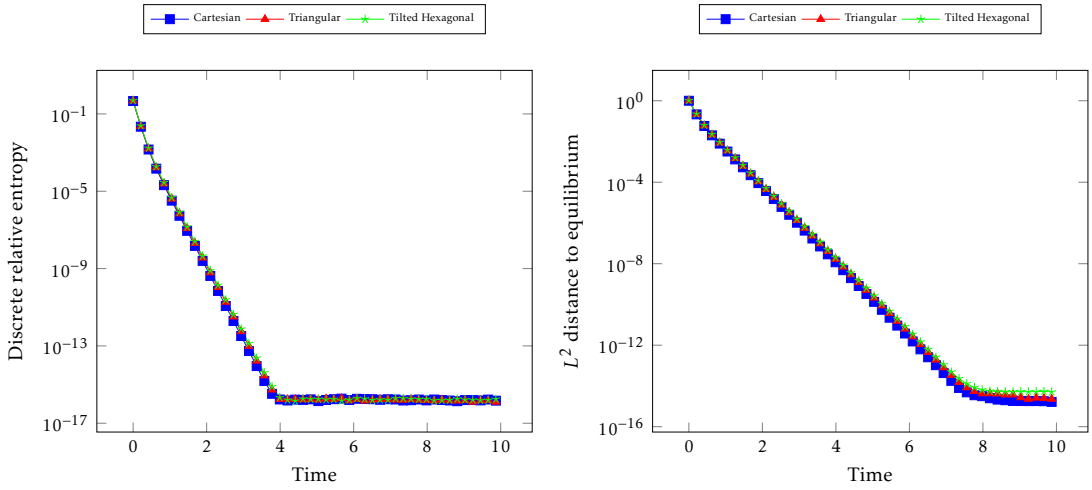
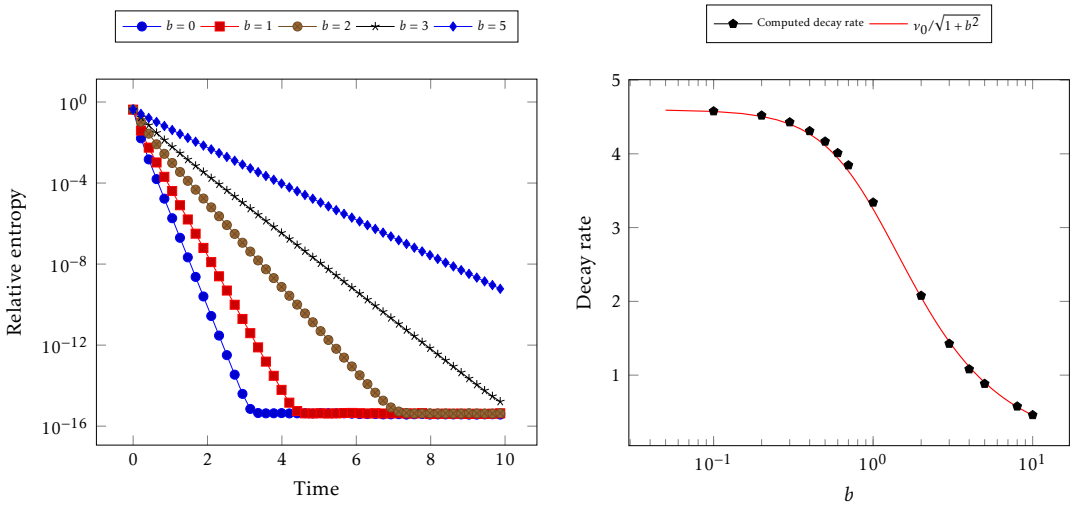


Figure 2.10 – **Test-case 5**. Influence of the mesh geometry on the long-time behaviour: entropy and L^2 distance to equilibrium.

tilted hexagonal mesh, which is not adapted to the geometry of the semiconductor device, the long-time behaviour is essentially similar to the one on meshes with adapted geometry. Again, this indicates the robustness of the scheme with respect to the mesh used.

Last, we investigate the influence of the magnetic field over the long-time behaviour. We consider the test-case 6, with Blakemore statistics ($h(s) = \log\left(\frac{s}{1-\gamma s}\right)$), no recombination ($r = 0$), $\lambda = 1$ and boundary values $N_0^D = \mathrm{e}$, $N_1^D = 1$ and $\alpha_0 = 1$. We perform our simulations on a Cartesian mesh, with a time step $\Delta t = 0.01$, with different values of b. The results are presented



(a) Evolution of the relative entropy

(b) Evolution of the decay rate with respect to b

Figure 2.11 – **Test-case 6**. Long-time behaviour for the Blakemore statistics and influence of the magnetic field

in Figure 2.11a. One can see the relative entropy decreases exponentially fast in time. Moreover, the presence of a magnetic field tends to attenuate the dissipative effects, and to slow down the evolution. This was expected from a physical point of view, since the magnetic field induces a rotation of the charge carriers which delays the natural relaxation towards equilibrium. On Figure 2.11b, we show the evolution of the entropy decay rate with respect to the intensity of the magnetic field b. We plot the values of the decay rate for different values of b, as well as a reference rate $\nu(b) = \frac{\nu_0}{\sqrt{1+b^2}}$, where ν_0 is the numerical value of the decay rate computed with no magnetic field ($b = 0$). The computed values seem to fit very well the reference rate. Note that the decay rate seems therefore to be proportional to $\frac{1}{\sqrt{1+b^2}}$, which happens to be the modulus of the eigenvalues of the diffusion tensors Λ_N and Λ_P.

2.6 Conclusion

In this article, we have designed and analysed a scheme for general anisotropic drift-diffusion systems on general polytopal meshes. The scheme is based on the Hybrid Finite Volume method. We have proved that the scheme has a discrete entropic structure, and used it to show the existence of solutions (with I_h-valued densities). As a by-product of this structure, we have also proved that the solutions to the scheme converge exponentially fast in time towards the associated discrete thermal equilibrium. The results are established for general statistics functions h, general diffusion tensors (potentially anisotropic and nonsymmetric) and general recombination terms, under the hypothesis that the boundary data are compatible with the thermal equilibrium (conditions (2.4) and (2.5)). Finally, we have validated our scheme on different numerical tests, highlighting the I_h-valuation of the discrete densities, the ability to withstand intense magnetic fields and the long-time behaviour. Numerical experiments (not presented in this paper) suggest that the scheme introduced here also works in situations where the compatibility condition with the thermal equilibrium does not hold. Hence, a future

direction would be to analyse the scheme in such a situation.

2.A Discrete boundedness by entropy and dissipation

In this appendix, we prove Lemma 7. To do so, we need a technical result stated below.

Lemma 8. *Let $E : (x,y) \mapsto (G(x)-G(y))(x-y)$, and $(x,y) \in \mathbb{R}^2$.*
If there exist two positive constants M and C_1 such that $|y| \le M$ and $0 \le E(x,y) \le C_1$, then, there exists a constant C_{M,C_1} only depending on M and C_1 such that $|x| \le C_{M,C_1}$.

Proof. First, we define $E_y : \delta \mapsto (G(y+\delta)-G(y))\delta$, therefore letting $\delta = x-y$, one has $E(x,y) = E_y(\delta) \le C_1$. The inequality $|x| \le M + |\delta|$ holds, so it suffices to get a bound on δ to conclude:

(i) if $\delta \ge 2M+1 > 0$, then since G is increasing, one has

$$G(y+\delta) \ge G(y+2M+1) \ge G(-M+2M+1) = G(M+1) \text{ and } G(y) \le G(M),$$

so $C_1 \ge E_y(\delta) \ge (G(M+1)-G(M))\,\delta$, and we get $2M+1 \le \delta \le C_1\,(G(M+1)-G(M))^{-1}$;

(ii) if $\delta \le -2M-1$, a similar computation shows that $-2M-1 \ge \delta \ge C_1\,(G(-M-1)-G(-M))^{-1}$.

Hence, one has $|\delta| \le \max\left(2M+1, \dfrac{C_1}{G(M+1)-G(M)}, \dfrac{C_1}{G(-M)-G(-M-1)}\right)$, which concludes the proof. □

We can now prove Lemma 7. The proof is similar to [70, Lemma 2], except for the use of the unknowns on $\mathcal{E}_{ext}^D$.

Lemma 7. *Let $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \in \underline{V}_{\mathcal{D},0}^2$, and assume that there exists $B_\sharp \ge 0$ such that*

$$\mathbb{E}(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \le B_\sharp \qquad \text{and} \qquad \mathbb{D}(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \le B_\sharp. \tag{2.50}$$

Then, there exists $C_\sharp > 0$, depending on the data, $B_\sharp$ and $\mathcal{D}$ such that

$$-C_\sharp \underline{1}_{\mathcal{D}} \le \underline{w}_{\mathcal{D}}^N \le C_\sharp \underline{1}_{\mathcal{D}},\ -C_\sharp \underline{1}_{\mathcal{D}} \le \underline{w}_{\mathcal{D}}^P \le C_\sharp \underline{1}_{\mathcal{D}} \text{ and } -C_\sharp \underline{1}_{\mathcal{D}} \le \underline{\phi}_{\mathcal{D}} \le C_\sharp \underline{1}_{\mathcal{D}},$$

where $\underline{\phi}_{\mathcal{D}}$ is the electrostatic potential associated to $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P)$.

Proof. The proof is divided into two steps. First, we prove bounds on the discrete electrostatic potential $\underline{\phi}_{\mathcal{D}}$ thanks to the bound on the entropy. Then, we use these bounds to estimate $(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P)$.

Since $\mathbb{E}(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \le B_\sharp$, by (2.16) one has $\dfrac{\alpha_\flat \lambda_\flat^\phi}{2} |\underline{\phi}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}}^e|_{1,\mathcal{D}}^2 \le \dfrac{1}{2} a_{\mathcal{D}}^\phi (\underline{\phi}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}}^e, \underline{\phi}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}}^e) \le B_\sharp$. Therefore, letting $c_1 = |\underline{\phi}_{\mathcal{D}}^e|_{1,\mathcal{D}} + \sqrt{\dfrac{2B_\sharp}{\alpha_\flat \lambda_\flat^\phi}}$, we get that $|\underline{\phi}_{\mathcal{D}}|_{1,\mathcal{D}} \le c_1$. By definition of $|.|_{1,\mathcal{D}}$, we deduce that, for any $K \in \mathcal{M}$ and any $\sigma \in \mathcal{E}_K$, $\dfrac{|\sigma|}{d_{K,\sigma}} (\phi_K - \phi_\sigma)^2 \le c_1^2$. Letting $c_2 = \max\limits_{K\in\mathcal{M}, \sigma\in\mathcal{E}_K} c_1 \sqrt{\dfrac{d_{K,\sigma}}{|\sigma|}}$, one has

$$\forall K \in \mathcal{M}, \forall \sigma \in \mathcal{E}_K, |\phi_K - \phi_\sigma| \le c_2.$$

On the other hand, there exists $\sigma_0 \in \mathcal{E}_{ext}^D$ such that $\phi_{\sigma_0} = \phi_{\sigma_0}^D$ and so, by (2.11), we have

$$|\phi_{\sigma_0}| \leq \|\phi^D\|_{L^\infty(\Omega)}.$$

Now, one can use the connectivity of the mesh: for any cell K (*resp.* face σ) there is a finite sequence of components of $\underline{\phi}_{\mathcal{D}}$, denoted $(x_k)_{0\leq k\leq l}$, starting at $x_0 = \phi_{\sigma_0}$ and finishing at $x_l = \phi_K$ (*resp.* $x_l = \phi_\sigma$) such that, for any k in $\{0,\dots l-1\}$, one has $|x_{k+1} - x_k| \leq c_2$. Therefore, one concludes that

$$|x_l| \leq lc_2 + |x_0| \leq 2|\mathcal{M}|c_2 + \|\phi^D\|_{L^\infty(\Omega)}.$$

Thus there exists $\phi_\sharp$ positive depending on the data and $\mathcal{D}$ such that

$$-\phi_\sharp \underline{1}_{\mathcal{D}} \leq \underline{\phi}_{\mathcal{D}} \leq \phi_\sharp \underline{1}_{\mathcal{D}}.$$

Now, note that since $\mathbb{D}(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^P) \leq B_\sharp$, one has $T_{\mathcal{D}}^N(\underline{N}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^N) \leq B_\sharp$ and $T_{\mathcal{D}}^P(\underline{P}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}^P, \underline{w}_{\mathcal{D}}^P) \leq B_\sharp$, where we recall that the discrete densities are defined by $\underline{N}_{\mathcal{D}} = g(\underline{w}_{\mathcal{D}}^N + \underline{\phi}_{\mathcal{D}} + \alpha_N \underline{1}_{\mathcal{D}})$ and $\underline{P}_{\mathcal{D}} = g(\underline{w}_{\mathcal{D}}^P - \underline{\phi}_{\mathcal{D}} + \alpha_P \underline{1}_{\mathcal{D}})$. For $K \in \mathcal{M}$, using (2.22) alongside with the positivity of $r_K(\underline{N}_K)$ and the local coercivity (2.15) of a_K^N, one gets that

$$B_\sharp \geq r_K(\underline{N}_K) a_K^N(\underline{w}_{\mathcal{D}}^N, \underline{w}_{\mathcal{D}}^N) \geq \alpha_\flat \lambda_\flat \sum_{\sigma\in\mathcal{E}_K} r_K(\underline{N}_K) \frac{|\sigma|}{d_{K,\sigma}} (w_K^N - w_\sigma^N)^2. \tag{2.68}$$

Given $\sigma \in \mathcal{E}_K$, by definition of r_K and $\underline{N}_K$, using (2.24) and the positivity of m, we obtain

$$r_K(\underline{N}_K) = \frac{1}{|\mathcal{E}_K|} \sum_{\sigma'\in\mathcal{E}_K} m(N_K, N_{\sigma'}) \geq \frac{1}{|\mathcal{E}_K|} m_h\Big(g(w_K^N + \phi_K + \alpha_N), g(w_\sigma^N + \phi_\sigma + \alpha_N)\Big).$$

Recall that formula (2.26) holds, so

$$r_K(\underline{N}_K) \geq \frac{1}{|\mathcal{E}_K|} \frac{G(w_K^N + \phi_K + \alpha_N) - G(w_\sigma^N + \phi_\sigma + \alpha_N)}{(w_K^N + \phi_K + \alpha_N) - (w_\sigma^N + \phi_\sigma + \alpha_N)}.$$

Moreover, since G is convex, $(x,y) \mapsto \frac{G(x)-G(y)}{x-y}$ is non-decreasing with respect to both its variables, therefore using the bound on $\underline{\phi}_{\mathcal{D}}$ proved previously we get

$$\begin{aligned} r_K(\underline{N}_K) &\geq \frac{1}{|\mathcal{E}_K|} \frac{G(w_K^N - \phi_\sharp + \alpha_N) - G(w_\sigma^N - \phi_\sharp + \alpha_N)}{(w_K^N - \phi_\sharp + \alpha_N) - (w_\sigma^N - \phi_\sharp + \alpha_N)} \\ &= \frac{1}{|\mathcal{E}_K|} \frac{G(w_K^N - \phi_\sharp + \alpha_N) - G(w_\sigma^N - \phi_\sharp + \alpha_N)}{w_K^N - w_\sigma^N}. \end{aligned}$$

Using (2.68), one gets that for any $K \in \mathcal{M}$ and for any $\sigma \in \mathcal{E}_K$,

$$0 \leq \Big(G(w_K^N - \phi_\sharp + \alpha_N) - G(w_\sigma^N - \phi_\sharp + \alpha_N)\Big)(w_K^N - w_\sigma^N) \leq \zeta$$

where $\zeta = \frac{B_\sharp}{\alpha_\flat \lambda_\flat} \max_{K\in\mathcal{M},\, \sigma\in\mathcal{E}_K} \frac{d_{K,\sigma}|\mathcal{E}_K|}{|\sigma|}$. Now, let $\underline{u}_{\mathcal{D}} = \underline{w}_{\mathcal{D}}^N + (-\phi_\sharp + \alpha_N)\underline{1}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$. It is clear that it

suffices to get bounds on $\underline{u}_{\mathcal{D}}$ to bound $\underline{w}_{\mathcal{D}}^N$. The previous relation writes

$$\forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}_K,\ 0 \le E(u_K, u_\sigma) \le \zeta, \tag{2.69}$$

where $E : \mathbb{R}^2 \to \mathbb{R}_+$ is the function defined in Lemma 8. Now, notice that since $\underline{w}_{\mathcal{D}}^N \in \underline{V}_{\mathcal{D},0}$ there exists $\sigma_0 \in \mathcal{E}_{ext}^D$ such that $u_{\sigma_0} = -\phi_\sharp + \alpha_N$. One can then use the connectivity of the mesh to conclude: for any cell K (*resp.* face σ) there is a finite sequence of components of $\underline{u}_{\mathcal{D}}$, denoted $(x_k)_{0\le k\le l}$, starting at $x_0 = u_{\sigma_0}$ and finishing at $x_l = u_K$ (*resp.* $x_l = u_\sigma$) such that for any k in $\{0,\dots,l-1\}$, one has $E(x_{k+1}, x_k) \le \zeta$. Therefore, by l successive applications of Lemma 8, we get the existence of some $c^N > 0$, depending on ζ, l, and $x_0 = -\phi_\sharp + \alpha_N$ such that $|x_l| \le c^N$. Thus, there exists $C_\sharp^N > 0$ depending on the data, $\mathcal{D}$ and $B_\sharp$ such that

$$\forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}, \quad |w_K^N| \le C_\sharp^N \text{ and } |w_\sigma^N| \le C_\sharp^N.$$

Using the same strategy, one gets bounds for $\underline{w}_{\mathcal{D}}^P$, which concludes the proof. □

Chapter 3

A comparison of structure-preserving schemes for drift-diffusion systems on general meshes: DDFV vs HFV

Outline of the current chapter

This chapter is a work in collaboration with Stella Krell, accepted as a proceeding of the FVCA X conference [187]. For the purpose of unified presentation and comparison, the notations used in this chapter are slightly different from those of the two previous chapters.

We made a comparison between a Discrete Duality Finite Volume (DDFV) scheme and a Hybrid Finite Volume (HFV) scheme for a drift-diffusion model with mixed boundary conditions on general meshes. Both schemes are based on a nonlinear discretisation of the convection-diffusion fluxes, which ensures the positivity of the discrete densities. We investigate the behaviours of the schemes on various numerical test cases.

3.1 Motivation

We are interested in the numerical discretization of a drift-diffusion model on general meshes. Let Ω be a polygonal connected open bounded subset of $\mathbb{R}^2$, whose boundary $\Gamma = \partial\Omega$ is divided into two parts $\Gamma = \Gamma^D \cup \Gamma^N$ with $m(\Gamma^D) > 0$. The problem writes:

$$\begin{cases} \partial_t N - \operatorname{div}(\nabla N - N\nabla\phi) = 0 & \text{in } \mathbb{R}_+ \times \Omega, \\ \partial_t P - \operatorname{div}(\nabla P + P\nabla\phi) = 0 & \text{in } \mathbb{R}_+ \times \Omega, \\ -\lambda^2 \operatorname{div}(\nabla\phi) = C + P - N & \text{in } \mathbb{R}_+ \times \Omega, \\ N = N^D,\ P = P^D \text{ and } \phi = \phi^D & \text{on } \mathbb{R}_+ \times \Gamma^D, \\ (\nabla N - N\nabla\phi)\cdot n = (\nabla P + P\nabla\phi)\cdot n = \nabla\phi \cdot n = 0 & \text{on } \mathbb{R}_+ \times \Gamma^N, \\ N(0,\cdot) = N^{in} \text{ and } P(0,\cdot) = P^{in} & \text{in } \Omega, \end{cases} \tag{3.1}$$

where n denotes the unit normal vector to $\partial\Omega$ pointing outward Ω. Regarding the data,

(i) the parameter $\lambda > 0$ is the rescaled Debye length of the system, which accounts for the nondimensionalisation (relevant values of this parameter can be very small, inducing some stiff behaviours),

(ii) the initial conditions N^{in} and P^{in} belong to $L^\infty(\Omega)$ and are positive,

(iii) the doping profile C is in $L^\infty(\Omega)$, and characterises the semiconductor device used.

In the following, we also assume that the boundary conditions are the trace of some H^1 function on Ω, such that the following relation holds:

$$\log(N^D) - \phi^D = \alpha_N \text{ and } \log(P^D) + \phi^D = \alpha_P \text{ on } \Gamma^D, \tag{3.2}$$

where α_N and α_P are two real constants. It follows that N^D and P^D are positive.

The solution to (3.1) enjoys some natural physical properties: the densities N and P are positive for all time, and the solution converges exponentially fast towards some thermal equilibrium (N^e, P^e, ϕ^e) as the time goes to infinity. This thermal equilibrium, which is a stationary solution to (3.1), is defined such that

$$N^e = \mathrm{e}^{\alpha_N + \phi^e} \text{ and } P^e = \mathrm{e}^{\alpha_P - \phi^e}, \tag{3.3}$$

where the electrostatic potential at the equilibrium ϕ^e is the solution to the Poisson-Boltzmann equation

$$\begin{cases} -\lambda^2 \operatorname{div}(\nabla\phi^e) = C + \exp(\alpha_P - \phi^e) - \exp(\alpha_N + \phi^e) & \text{in } \Omega, \\ \phi^e = \phi^D & \text{on } \Gamma^D, \\ \nabla\phi^e \cdot n = 0 & \text{on } \Gamma^N. \end{cases} \tag{3.4}$$

The relation (3.2) is a compatibility condition in order to ensure that the only stationary state of (3.1) is the thermal equilibrium defined by (3.4).

When designing numerical schemes for (3.1), it is crucial to ensure that the scheme preserves these properties at the discrete level. This structure preserving features are ensured by classical Two-Point Flux Approximation (TPFA) schemes on admissible orthogonal meshes (see [27]). Unfortunately, these schemes cannot be used on general meshes, and their use is essentially restricted to isotropic problems. On the other hand, it appears that the use of adaptative mesh refinement could be a relevant solution in order to handle the stiff behaviour of these problems, as presented in [58]. Therefore, a first step towards this kind of strategy lies in the development

of structure preserving scheme on general meshes. However, the main issue of schemes handling general meshes for diffusive problems is the lack of discrete positivity, both at the theoretical and practical level (see [109]). One of the possible solution in order tackle this problem is to use a nonlinear approach based on the entropy structure of the model. Such idea was originally introduced in [56] with a Vertex Approximate Gradient (VAG) scheme for advection-diffusion problems. Following the ideas of this seminal work, a nonlinear positivity preserving Discrete Duality Finite Volume (DDFV) scheme for linear advection-diffusion equations has been introduced in [52]. The good long-time behaviour of this scheme has then been established in [51]. In the spirit of these works, a nonlinear structure preserving Hybride Finite Volume (HFV) scheme was introduced and analysed in [213]: this scheme is proved to preserve the positivity of the densities as well as the long-time behaviour of the solutions. These three schemes are based on different spatial discretisation methods handling general meshes (VAG, DDFV and HFV), but they share the same spirit in terms of conception and analysis.

The aim of this paper is therefore to compare two of these methods for a realistic coupled problem, from a numerical point of view. To do so, we to introduce a nonlinear structure preserving DDFV scheme for (3.1) based on the scheme of [52] and we compare it numerically with the HFV scheme of [213].

3.2 Description of the schemes

The schemes used here are based on the same nonlinear strategy, introduced in [56], consisting in the reformulation of the convection-diffusion fluxes:

$$\nabla N - N\nabla\phi = N\nabla(\log(N) - \phi) \text{ and } \nabla P + P\nabla\phi = P\nabla(\log(P) + \phi).$$

The quantities $\log(N) - \phi$ and $\log(P) + \phi$ are often called quasi-Fermi potentials in the literature associated to semiconductor models, and have an important physical meaning. At the discrete level, both schemes rely on discrete gradients operators to approximate the continuous gradients. The major issue lies in the discretisation of the prefactors P and N, which will be handled by local reconstruction operators. This discretisation strategy is a way of ensuring, both at the theoretical and practical levels, the positivity of the discrete densities. We refer to [213, Theorem 1] (HFV scheme for drift-diffusion system) and [52, Theorem 2.1] (DDFV scheme for a single advection-diffusion equation) for proofs of this statement. We also refer the reader to these proofs for more insight about the reconstruction operators.
Both schemes are based on a backward Euler discretisation in time. To fix ideas, we will use a constant time step $\Delta t > 0$. For more precise descriptions and statements about the schemes and the meshes, we refer to [52] (DDFV) and [213] (HFV).

Remark 18 (Generalisation to anisotropic models)**.** *In this paper, we consider isotropic convection-diffusion equations for the charges carriers for the sake of brevity. One could add anisotropic diffusion tensors and consider the framework described in [213]. Thus, the presentation made here can easily be extended to models with external magnetic fields (see [143]).*

Both schemes rely on a spatial discretisation (or mesh) of the domain Ω. The (primal interior) mesh $\mathfrak{M}$ is a partition of Ω in polygonal control volumes (or cells). We let $\partial\mathfrak{M}$ be the set of boundary edges, seen either as degenerate control volumes (DDFV framework) or as edges (HFV framework). The primal mesh $\overline{\mathfrak{M}}$ is defined as the reunion of $\mathfrak{M}$ and $\partial\mathfrak{M}$. Given a cell $K \in \overline{\mathfrak{M}}$, we fix a point $x_K \in K$, called the center of K. For all neighboring primal cells K and L, we assume that $\partial K \cap \partial L$ is a segment, corresponding to an internal edge of the mesh $\mathfrak{M}$, denoted by $\sigma = K|L$ and we let $\mathcal{E}_{int}$ be the set of such edges. We denotes by $\mathcal{E} = \mathcal{E}_{int} \cup \partial\mathfrak{M}$ the set of all (internal and

exterior) edges of the mesh, and define $\mathcal{E}_K$ the set of edges of the cell $K \in \mathfrak{M}$. For any $K \in \mathfrak{M}$ and $\sigma \in \mathcal{E}_K$, we define $\mathbf{n}_{\sigma K}$ as the unit normal to σ outward K.
Given any measurable $X \subset \mathbb{R}^2$, we denote by m_X the measure of the object X.

3.2.1 The DDFV scheme

In order to define the DDFV scheme, we need to introduce two other meshes: the dual mesh denoted $\overline{\mathfrak{M}^*}$ and the diamond mesh denoted $\mathfrak{D}$ (see [52] for more details). The dual mesh $\overline{\mathfrak{M}^*}$ is also composed of interior dual mesh $\mathfrak{M}^*$ (corresponding of cells around vertex in Ω) and of boundary dual mesh $\partial\mathfrak{M}^*$ (corresponding of cells around vertex on $\partial\Omega$). For any vertex x_{K^*} of the primal mesh satisfying $x_{K^*} \in \Omega$, we define a polygonal control volume K^* by connecting all the centers of the primal cells sharing x_{K^*} as vertex. For any vertex $x_{K^*} \in \partial\Omega$, we define a polygonal control volume K^* by connecting the centers x_K of the interior primal cells and the midpoints of the boundary edges sharing x_{K^*} as vertex and x_{K^*}. We define the set $\mathcal{E}^*_{int}$ of internal edges of the dual mesh similarly as $\mathcal{E}_{int}$. We denote by $\mathbf{n}_{\sigma^* K^*}$ the unit normal to σ^* outward K^*. For each couple $(\sigma, \sigma^*) \in \mathcal{E} \times \mathcal{E}^*_{int}$ such that $\sigma = [x_{K^*}, x_{L^*}]$ and $\sigma^* = K^*|L^*$, we define the quadrilateral diamond $\mathcal{D}_{\sigma,\sigma^*}$ whose diagonals are σ and σ^* (if $\sigma \subset \partial\Omega$, it degenerates into a triangle). The set of the diamonds defines the diamond mesh $\mathfrak{D}$, which is a partition of Ω. Finally, the DDFV

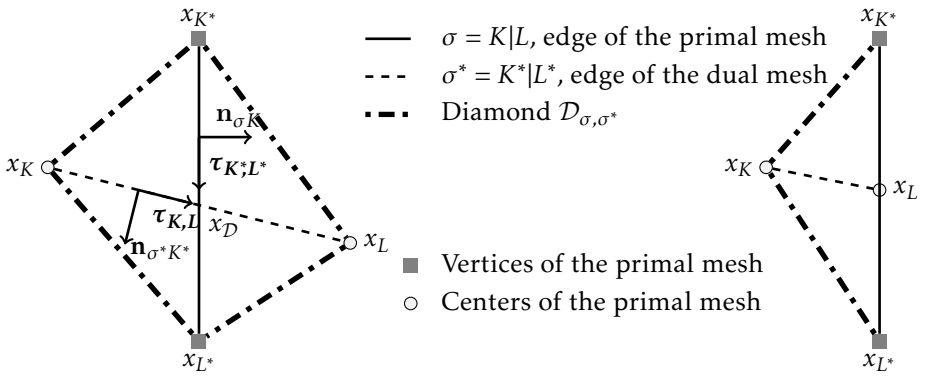


Figure 3.1 – Definition of the diamonds $\mathcal{D}_{\sigma,\sigma^*}$ and related notations.

mesh is made of $\mathcal{T} = (\overline{\mathfrak{M}}, \overline{\mathfrak{M}^*})$ and $\mathfrak{D}$.

We now introduce the space of scalar fields which are associated to each primal and dual cell $\mathbb{R}^{\mathcal{T}}$, and space of vector fields constant on the diamonds $\left(\mathbb{R}^2\right)^{\mathfrak{D}}$:

$$u_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}} \Longleftrightarrow u_{\mathcal{T}} = \left((u_K)_{K\in\overline{\mathfrak{M}}}, (u_{K^*})_{K^*\in\overline{\mathfrak{M}^*}}\right) \text{ and } \xi_{\mathfrak{D}} \in \left(\mathbb{R}^2\right)^{\mathfrak{D}} \Longleftrightarrow \xi_{\mathfrak{D}} = (\xi_{\mathcal{D}})_{\mathcal{D}\in\mathfrak{D}}.$$

To enforce Dirichlet boundary conditions, we introduce the set of Dirichlet boundary primal and dual cells: $\partial\mathfrak{M}_D = \{K \in \partial\mathfrak{M} : K \subset \Gamma_D\}$ and $\partial\mathfrak{M}^*_D = \{K^* \in \partial\mathfrak{M}^* : x_{K^*} \in \overline{\Gamma}_D\}$, and, for a given $v \in C(\Gamma^D)$, we define

$$\mathrm{E}_v^{\Gamma_D} = \{u_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}} \mid \forall K \in \partial\mathfrak{M}_D,\, u_K = v(x_K) \text{ and } \forall K^* \in \partial\mathfrak{M}^*_D,\, u_{K^*} = v(x_{K^*})\}.$$

These sets can be generalised to less regular boundary data v by substituting punctual evaluation

by averages. We also define discrete bilinear forms on $\mathbb{R}^{\mathcal{T}}$ and $\left(\mathbb{R}^2\right)^{\mathfrak{D}}$ by

$$\begin{aligned}
[\![v_{\mathcal{T}}, u_{\mathcal{T}}]\!]_{\mathcal{T}} &= \frac{1}{2}\left(\sum_{K\in\mathfrak{M}} \mathrm{m}_K u_K v_K + \sum_{K^*\in\overline{\mathfrak{M}^*}} \mathrm{m}_{K^*} u_{K^*} v_{K^*}\right), \quad \forall (u_{\mathcal{T}}, v_{\mathcal{T}}) \in \left(\mathbb{R}^{\mathcal{T}}\right)^2,\\
\left(\xi_{\mathfrak{D}}, \varphi_{\mathfrak{D}}\right)_{\mathfrak{D}} &= \sum_{\mathcal{D}\in\mathfrak{D}} \mathrm{m}_{\mathcal{D}}\, \xi_{\mathcal{D}} \cdot \varphi_{\mathcal{D}}, \quad \forall (\xi_{\mathfrak{D}}, \varphi_{\mathfrak{D}}) \in \left(\left(\mathbb{R}^2\right)^{\mathfrak{D}}\right)^2.
\end{aligned}$$

The DDFV method is based on the definition of a discrete gradient operator $\nabla^{\mathfrak{D}} : \mathbb{R}^{\mathcal{T}} \to \left(\mathbb{R}^2\right)^{\mathfrak{D}}$, defined by $\nabla^{\mathfrak{D}} u_{\mathcal{T}} = \left(\nabla^{\mathcal{D}} u_{\mathcal{T}}\right)_{\mathcal{D}\in\mathfrak{D}}$, where

$$\nabla^{\mathcal{D}} u_{\mathcal{T}} = \frac{1}{2m_{\mathcal{D}}} \left(\mathrm{m}_\sigma (u_L - u_K)\mathbf{n}_{\sigma K} + \mathrm{m}_{\sigma^*}(u_{L^*} - u_{K^*})\mathbf{n}_{\sigma^* K^*}\right) \quad \forall \mathcal{D} \in \mathfrak{D}. \tag{3.5}$$

Finally, we introduce a reconstruction operator on diamonds $r^{\mathfrak{D}}$. It is a mapping from $\mathbb{R}^{\mathcal{T}}$ to $\mathbb{R}^{\mathfrak{D}}$ defined for all $u_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$ by $r^{\mathfrak{D}} u_{\mathcal{T}} = \left(r^{\mathcal{D}} u_{\mathcal{T}}\right)_{\mathcal{D}\in\mathfrak{D}}$, where for $\mathcal{D} \in \mathfrak{D}$, whose vertices are x_K, x_L, x_{K^*}, x_{L^*}, $r^{\mathcal{D}} u_{\mathcal{T}} = \frac{1}{4}(u_K + u_L + u_{K^*} + u_{L^*})$. One can now introduce a DDFV discretisation of $(u, w, v) \mapsto \int_\Omega u \nabla w \cdot \nabla v$, defined by

$$T_{\mathfrak{D}} : (u_{\mathcal{T}}, w_{\mathcal{T}}, v_{\mathcal{T}}) \mapsto \sum_{\mathcal{D}\in\mathfrak{D}} \mathrm{m}_{\mathcal{D}} r^{\mathcal{D}} u_{\mathcal{T}}\, \nabla^{\mathcal{D}} w_{\mathcal{T}} \cdot \nabla^{\mathcal{D}} v_{\mathcal{T}}.$$

Now, we first discretise the data by taking the mean values of N^{in}, P^{in} and C on the primal and dual cells, which define $N^0_{\mathcal{T}}$, $P^0_{\mathcal{T}}$ and $C_{\mathcal{T}}$. Then, for all $n \geq 0$, we look for $(N^{n+1}_{\mathcal{T}}, P^{n+1}_{\mathcal{T}}, \phi^{n+1}_{\mathcal{T}}) \in \mathrm{E}^{\Gamma_D}_{N^D} \times \mathrm{E}^{\Gamma_D}_{P^D} \times \mathrm{E}^{\Gamma_D}_{\phi^D}$ solution to:

$$\left[\!\left[\frac{N^{n+1}_{\mathcal{T}} - N^n_{\mathcal{T}}}{\Delta t}, v_{\mathcal{T}}\right]\!\right]_{\mathcal{T}} + T_{\mathfrak{D}}(N^{n+1}_{\mathcal{T}}, \log(N^{n+1}_{\mathcal{T}}) - \phi^{n+1}_{\mathcal{T}}, v_{\mathcal{T}}) = 0 \quad \forall v_{\mathcal{T}} \in \mathrm{E}^{\Gamma_D}_0, \tag{3.6a}$$

$$\left[\!\left[\frac{P^{n+1}_{\mathcal{T}} - P^n_{\mathcal{T}}}{\Delta t}, v_{\mathcal{T}}\right]\!\right]_{\mathcal{T}} + T_{\mathfrak{D}}(P^{n+1}_{\mathcal{T}}, \log(P^{n+1}_{\mathcal{T}}) + \phi^{n+1}_{\mathcal{T}}, v_{\mathcal{T}}) = 0 \quad \forall v_{\mathcal{T}} \in \mathrm{E}^{\Gamma_D}_0, \tag{3.6b}$$

$$\lambda^2 \left(\nabla^{\mathcal{D}} \phi^{n+1}_{\mathcal{T}}, \nabla^{\mathcal{D}} v_{\mathcal{T}}\right)_{\mathfrak{D}} = \left[\!\left[C_{\mathcal{T}} + P^{n+1}_{\mathcal{T}} - N^{n+1}_{\mathcal{T}}, v_{\mathcal{T}}\right]\!\right]_{\mathcal{T}} \quad \forall v_{\mathcal{T}} \in \mathrm{E}^{\Gamma_D}_0. \tag{3.6c}$$

In (3.6a) and (3.6b), we use the notation $\log(u_{\mathcal{T}}) = \left((\log(u_K))_{K\in\overline{\mathfrak{M}}}, (\log(u_{K^*}))_{K^*\in\overline{\mathfrak{M}^*}}\right)$.

3.2.2 The HFV scheme

In order to define the HFV scheme, we need to introduce a pyramidal submesh. To do so, one has to assume that each cell $K \in \mathfrak{M}$ is star-shaped with respect to its center x_K (we recall that x_K is not necessarily the barycentre of K). We then define $P_{K,\sigma}$ as the pyramid (triangle) of base σ and apex x_K. Given any $\sigma \in \mathcal{E}$, we denote by $\overline{x}_\sigma$ the barycentre of σ, and by $d_{K,\sigma}$ the euclidean distance between σ and x_K. Finally, we define the hybrid discretisation (or mesh) as $\mathcal{D} = (\mathfrak{M}, \mathcal{E})$.

We now introduce the space of discrete (scalar) hybrid unknowns $\underline{V}_{\mathcal{D}}$:

$$\underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}} \Longleftrightarrow \underline{u}_{\mathcal{D}} = \left((u_K)_{K\in\mathfrak{M}}, (u_\sigma)_{\sigma\in\mathcal{E}}\right),$$

where the $u_K \in \mathbb{R}$ are the cell unknowns and the $u_\sigma \in \mathbb{R}$ are the edges unknowns (approximation

of the trace of the solutions on the edges). To enforce Dirichlet boundary conditions, for a given $v \in C(\Gamma^D)$, we define

$$\underline{V}_{\mathcal{D},v}^{\Gamma_D} = \{\underline{u}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}} \mid \forall \sigma \in \partial\mathfrak{M}_D,\, u_\sigma = v(\overline{x}_\sigma)\}.$$

As for the DDFV framework, we define a bilinear form on $\underline{V}_{\mathcal{D}}$, discrete counterpart of the inner product on $L^2(\Omega)$ as

$$[\![\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}]\!]_{\mathfrak{M}} = \sum_{K\in\mathfrak{M}} \mathrm{m}_K u_K v_K, \quad \forall (\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D}}^2.$$

Note that this discrete inner product only takes into account the values of the cell unknowns.

The HFV method is based on the definition of a discrete gradient operator $\nabla_{\mathcal{D}} : \underline{V}_{\mathcal{D}} \to (\mathbb{R}^2)^\Omega$ which maps discrete hybrid unknowns onto piecewise constant functions on the pyramidal submesh. More precisely, given $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, $K \in \mathfrak{M}$ and $\sigma \in \mathcal{E}_K$ we let

$$\nabla_{\mathcal{D}} \underline{v}_{\mathcal{D}|P_{K,\sigma}} = G_K \underline{v}_{\mathcal{D}} + S_{K,\sigma} \underline{v}_{\mathcal{D}},$$

where the consistent part of the gradient $G_K \underline{v}_{\mathcal{D}}$ and the stabilisation part $S_{K,\sigma}\underline{v}_{\mathcal{D}}$ are defined as

$$G_K \underline{v}_{\mathcal{D}} = \frac{1}{\mathrm{m}_K} \sum_{\sigma'\in\mathcal{E}_K} m_{\sigma'} v_{\sigma'} n_{K,\sigma'} \quad \text{and} \quad S_{K,\sigma}\underline{v}_{\mathcal{D}} = \frac{\eta}{d_{K,\sigma}}\Big(v_\sigma - v_K - G_K \underline{v}_K \cdot (\overline{x}_\sigma - x_K)\Big) n_{K,\sigma},$$

where η is a given positive stabilisation parameter. One can now define the discrete counterpart of $(u,v) \mapsto \int_\Omega \nabla u \cdot \nabla v$ as

$$a_{\mathcal{D}} : (\underline{u}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \mapsto \int_\Omega \nabla_{\mathcal{D}} \underline{u}_{\mathcal{D}} \cdot \nabla_{\mathcal{D}} \underline{v}_{\mathcal{D}}.$$

Finally, we introduce as previously local reconstruction operators on cells $r^K : \underline{V}_{\mathcal{D}} \to \mathbb{R}$, such that for any $u_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}$, $r^K(\underline{u}_{\mathcal{D}}) = \dfrac{1}{|\mathcal{E}_K|} \displaystyle\sum_{\sigma\in\mathcal{E}_K} \frac{u_K + u_\sigma}{2}$, where $|\mathcal{E}_K|$ is the cardinal of the finite set $\mathcal{E}_K$. One can now introduce a HFV discretisation of $(u,w,v) \mapsto \int_\Omega u \nabla w \cdot \nabla v$, defined by

$$T_{\mathcal{D}} : (\underline{u}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \mapsto \sum_{K\in\mathfrak{M}} r^K(\underline{u}_{\mathcal{D}}) \int_K \nabla_{\mathcal{D}} \underline{w}_{\mathcal{D}} \cdot \nabla_{\mathcal{D}} \underline{v}_{\mathcal{D}}.$$

We now discretise the data by taking the mean values of N^{in}, P^{in} and C on the cells and edges, which define $\underline{P}^0_{\mathcal{D}}$, $\underline{N}^0_{\mathcal{D}}$ and $\underline{C}_{\mathcal{D}}$. Then, for all $n \geq 0$, we look for $(\underline{N}^{n+1}_{\mathcal{D}}, \underline{P}^{n+1}_{\mathcal{D}}, \underline{\phi}^{n+1}_{\mathcal{D}}) \in \underline{V}^{\Gamma_D}_{\mathcal{D},N^D} \times \underline{V}^{\Gamma_D}_{\mathcal{D},P^D} \times \underline{V}^{\Gamma_D}_{\mathcal{D},\phi^D}$ solution to:

$$\left[\!\!\left[\frac{\underline{N}^{n+1}_{\mathcal{D}} - \underline{N}^n_{\mathcal{D}}}{\Delta t}, \underline{v}_{\mathcal{D}}\right]\!\!\right]_{\mathfrak{M}} + T_{\mathcal{D}}(\underline{N}^{n+1}_{\mathcal{D}}, \log(\underline{N}^{n+1}_{\mathcal{D}}) - \underline{\phi}^{n+1}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = 0 \quad \forall \underline{v}_{\mathcal{D}} \in \underline{V}^{\Gamma_D}_{\mathcal{D},0}, \tag{3.7a}$$

$$\left[\!\!\left[\frac{\underline{P}^{n+1}_{\mathcal{D}} - \underline{P}^n_{\mathcal{D}}}{\Delta t}, \underline{v}_{\mathcal{D}}\right]\!\!\right]_{\mathfrak{M}} + T_{\mathcal{D}}(\underline{P}^{n+1}_{\mathcal{D}}, \log(\underline{P}^{n+1}_{\mathcal{D}}) + \underline{\phi}^{n+1}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = 0 \quad \forall \underline{v}_{\mathcal{D}} \in \underline{V}^{\Gamma_D}_{\mathcal{D},0}, \tag{3.7b}$$

$$\lambda^2 a_{\mathcal{D}}\left(\underline{\phi}^{n+1}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}\right) = \left[\!\!\left[\underline{C}_{\mathcal{D}} + \underline{P}^{n+1}_{\mathcal{D}} - \underline{N}^{n+1}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}\right]\!\!\right]_{\mathfrak{M}} \quad \forall \underline{v}_{\mathcal{D}} \in \underline{V}^{\Gamma_D}_{\mathcal{D},0}. \tag{3.7c}$$

As previously, we use the notation $\log(\underline{u}_{\mathcal{D}}) = \Big((\log(u_K))_{K\in\mathfrak{M}}, (\log(u_\sigma))_{\sigma\in\mathcal{E}}\Big)$.

3.2.3 Some structural differences between schemes

As highlighted by the unified presentation above, both schemes are very similar and rely on the same features. Note that both local reconstruction operators $r^{\mathcal{D}}$ and r^{K} take into account all the local unknowns of the geometric entity considered (diamond or cells), this property is the key point of the analysis of this kind of schemes, see [52, 213].
However, the schemes exhibit differences, some of which are listed below:

— the discrete HFV gradient $\nabla_{\mathcal{D}}$ includes a stabilisation term for the sake of coercivity and the stabilisation parameter η has to be chosen a priori, whereas the DDFV one is simpler and do not need any choice of parameter;

— the DDFV unknowns are all "volumic", in the sense that there are associated to geometric entities with non-zero two-dimensional measures, whereas the faces unknowns of the HFV method have no mass and have no influence on the discrete time derivative terms $\left[\!\left[\underline{N}_{\mathcal{D}}^{n+1} - \underline{N}_{\mathcal{D}}^{n}, \underline{v}_{\mathcal{D}}\right]\!\right]_{\mathfrak{M}}$ and $\left[\!\left[\underline{P}_{\mathcal{D}}^{n+1} - \underline{P}_{\mathcal{D}}^{n}, \underline{v}_{\mathcal{D}}\right]\!\right]_{\mathfrak{M}}$;

— the cells unknowns of the HFV scheme can be eliminated before solving linear systems, using a static condensation procedure (see [213, Section 5.1.2.]), whereas one has to solve a system including all primal and dual unknowns for DDFV;

— the HFV scheme can be used in 3D without any modification (the edges become faces), whereas using a DDFV method in 3D requires more sophisticated changes (see [90]).

3.3 Numerical experiments

The two numerical schemes described here are nonlinear, hence their algebraic realisations boil down to the resolution of nonlinear systems of equations. To solve these systems, we use Newton method, with an adaptative time stepping strategies: if the Newton method does not converge, we try to compute the solution for a smaller time step $0.5 \times \Delta t$. If the method converges, we use a bigger time step $1.4 \times \Delta t$. We fix the initial time step as well as the maximal time step allowed, and denote these quantities by Δt_{ini} and Δt_{max}.
We use the same stopping criterion for both scheme, namely the size of the l^2 norm of the goal function (with the same threshold). Note that even if we use the same algebraic criterion, this does not mean in practice that the criterion has the same level of exigence for both schemes. In particular, since the HFV scheme has two different type of unknowns (cell and face), understanding how to create equivalent stopping criteria for the two schemes is unclear at this point. The results presented below indicate that, indeed, the criterion used leads to slightly different behaviours.

For the HFV scheme, at each system resolution, a static condensation is used to eliminate the cell unknowns (see [213, Section 5.1.2.]), and we use $\eta = 1.5$.

Note that we use N, P and ϕ as discrete unknowns in the schemes. One could also solve the system using the quasi-Fermi potentials and the electrostatic potential as discrete unknowns. Such a choice corresponds at the level of the implementation to a preconditioning. It could maybe leads to better performances in practice, but it seems difficult to find relevant criterion in order to know when switch between the two sets of unknowns.

3.3.1 Description of the test case

The test cases used below follow the framework used in [213] to describe a 2D PN-junction, whose geometry is described in Figure 3.2. The domain Ω is the unit square $]0,1[^2$. For the

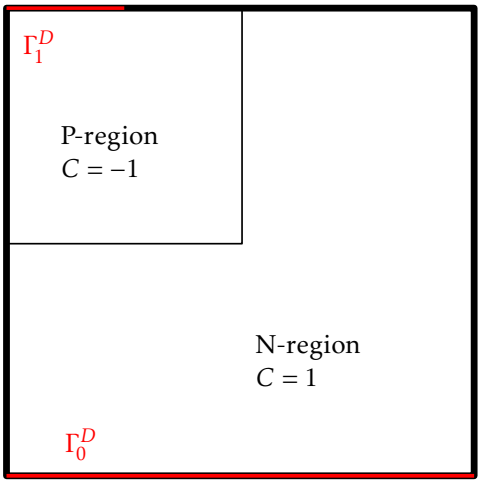


Figure 3.2 – PN diode geometry.

boundary conditions, we split $\Gamma^D = \Gamma_0^D \cup \Gamma_1^D$ with $\Gamma_0^D = [0,1]\times\{0\}$ and $\Gamma_1^D = [0,0.25]\times\{1\}$. For $i \in \{0,1\}$, we let

$$N^D = N_i^D,\ P^D = P_i^D \text{ and } \phi^D = \frac{h(N_i^D) - h(P_i^D)}{2} \text{ on } \Gamma_i^D.$$

To be consistent with the compatibility condition (3.2) we assume that there exists a constant α_0 such that $\log(N^D \times P^D) = \alpha_0$. Therefore for given N^D and α_0 we set

$$P^D = \frac{e^{\alpha_0}}{N^D} \text{ on } \Gamma^D.$$

Thus, one has $\alpha_N = \alpha_P = \frac{\alpha_0}{2}$. The doping profile C is piecewise constant, equal to -1 in the P-region and 1 in the N-region (see Figure 3.2). Last, we use the following smooth initial conditions:

$$N_0(x,y) = N_1^D + (N_0^D - N_1^D)(1-\sqrt{y}) \quad \text{and} \quad P_0(x,y) = P_1^D + (P_0^D - P_1^D)(1-\sqrt{y}).$$

3.3.2 Positivity

In this section, we compare the discrete positivity preservation of the schemes. The test used here corresponds to the following values:

$$\lambda = 0.05, \quad N_0^D = 0.1, \quad N_1^D = 1 \quad \text{and } \alpha_0 = -4.$$

We perform a test on a distorted quadrangle mesh (mesh_quad_6 of the FVCA 8 Benchmark), with $\Delta t_{ini} = 1.410^{-3}$ and $\Delta t_{max} = 0.1$. We show in Figure 3.3 the evolution of the minimal values of P and N, along with the time step and the number of Newton's iterations needed to compute the solutions at a given time for each time step. The minimal values are taken on every unknowns (primal and dual cells for the DDFV scheme, cells and faces for the HFV one). One can see that both schemes compute, as expected by the theoretical results, positive densities. The minimal values computed are of the same order for both schemes. Moreover, both computations proceed without the need of a time step reduction. Regarding the cost, it appears that the HFV scheme

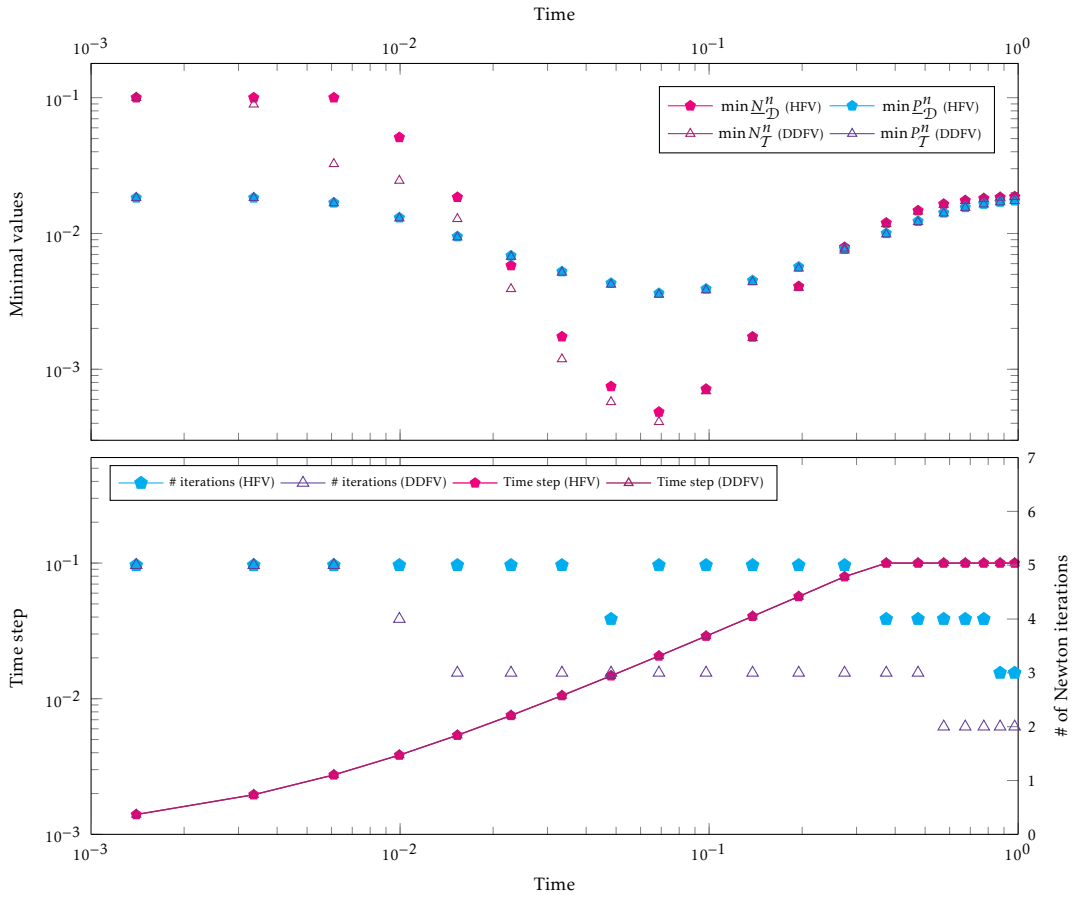


Figure 3.3 – **Positivity**. Evolution of the discrete minimal values, time step and cost

needs more Newton iterations than the DDFV one (90 vs 63). For both schemes, the number of iteration decay as the time increases, since the solutions converge exponentially fast towards the equilibrium.

3.3.3 Long-time behaviour

Here, we investigate the long-time behaviour of the schemes. At the continuous level, one usually quantify the distance between the solution (N, P, ϕ) and the equilibrium (N^e, P^e, ϕ^e) by looking at the relative entropy, defined as

$$\mathbb{E}(t) = \int_\Omega N^e H\left(\frac{N}{N^e}\right) + \int_\Omega P^e H\left(\frac{P}{P^e}\right) + \frac{\lambda^2}{2}\|\nabla(\phi - \phi^e)\|^2_{L^2(\Omega)},$$

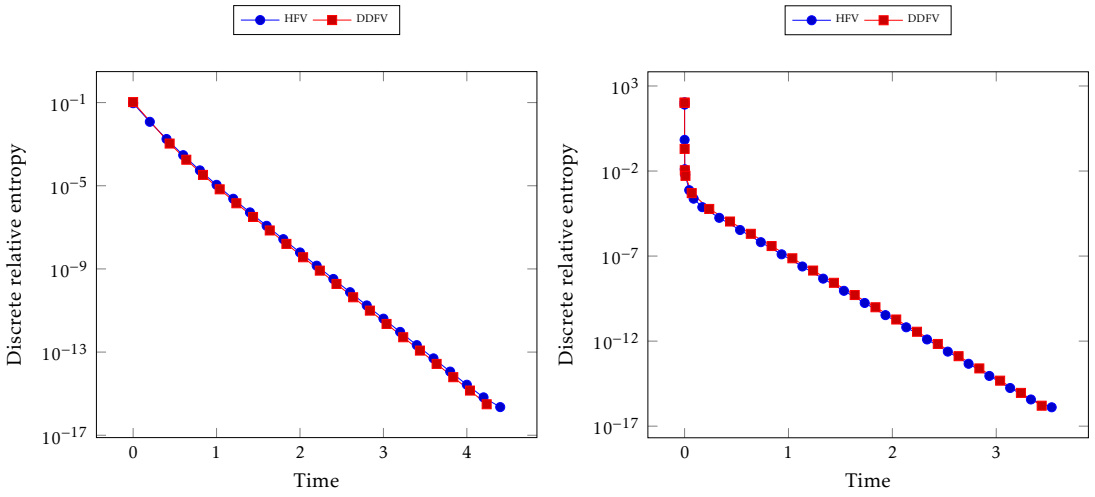
with $H : s \mapsto s\log(s) - s + 1$. One can check that (N, P, ϕ) coincides with the equilibrium if and only if the relative entropy cancels. In the following, we are interested in the evolution of the

discrete counterparts of this quantities, defined as

$$\mathbb{E}_{\mathcal{D}}^n = \left[\!\left[\underline{N}_{\mathcal{D}}^e H\left(\frac{\underline{N}_{\mathcal{D}}^n}{\underline{N}_{\mathcal{D}}^e}\right), \underline{1}_{\mathcal{D}} \right]\!\right]_{\mathfrak{M}} + \left[\!\left[\underline{P}_{\mathcal{D}}^e H\left(\frac{\underline{P}_{\mathcal{D}}^n}{\underline{P}_{\mathcal{D}}^e}\right), \underline{1}_{\mathcal{D}} \right]\!\right]_{\mathfrak{M}} + \frac{\lambda^2}{2} a_{\mathcal{D}}\left(\underline{\phi}_{\mathcal{D}}^n - \underline{\phi}_{\mathcal{D}}^e, \underline{\phi}_{\mathcal{D}}^n - \underline{\phi}_{\mathcal{D}}^e\right)$$

for the HFV scheme (where $\underline{1}_{\mathcal{D}}$ is the discrete elements whose coordinates are 1, and the product, quotient and functions are applied coordinate-wise) and similar definition for the DDFV scheme. Note that the HFV entropy does not take into account the edge unknowns of the discrete densities. To compute the discrete equilibrium, we use a nonlinear scheme for (3.4) and get $\underline{\phi}_{\mathcal{D}}^e$, then we defined the associated densities following the continuous relations $N^e = e^{\alpha_N + \phi^e}$ and $P^e = e^{\alpha_P - \phi^e}$ (defined as the discrete level coordinate by coordinate).

We consider a test case with physical data $N_0^D = \mathrm{e}$, $N_1^D = 1$ and $\alpha_0 = 0$. We also use two different values of the Debye length λ, respectively 1 and 0.01. We perform simulations on a triangular mesh, with a $\Delta t_{ini} = \Delta t_{max} = 0.1$. On Figure 3.4, we show the evolutions of the



(a) $\lambda = 1$ (b) $\lambda = 0.01$

Figure 3.4 – **Long-time behaviour.** Evolution of the discrete relative entropies.

discrete relative entropies along time, for the two values of λ and both schemes. As expected, the convergence towards the equilibrium is exponentially fast, as in the continuous framework. Moreover, it is remarkable to notice that the decay rates are almost the same for both schemes. Moreover, with the small Debye length (Figure 3.4b), both schemes are able to capture the behaviour with a very fast evolution far from the equilibrium, then slower once close to it.

3.4 Conclusion

We have compared numerically two schemes for drift-diffusion systems. Both schemes handle seamlessly general polygonal meshes, and are designed in order to preserve the positivity of the density as well as the entropy structure of the continuous problem. It appears that the two schemes present similar behaviours on practical test cases, both from a qualitative and quantitative viewpoint.

Chapter 4

High-order polytopal schemes for advection-diffusion equations: linear and nonlinear approaches

Outline of the current chapter

This chapter is an extended version of [212], which is an accepted proceeding of the FVCA X conference.

We are interested in the high-order approximation of anisotropic advection-diffusion equations on general meshes. We introduce two schemes, based on the Hybrid High-Order methodology. The first one is a linear scheme, while the second one is nonlinear. Both schemes admit solutions and possess a discrete entropy structure,

which ensures that the long-time behaviour of the discrete solutions mimics the continuous one. Moreover, the nonlinear scheme preserves the positivity of the solution at the discrete level. On the contrary, we present numerical evidence indicating that the linear scheme does not preserve positivity, independently of the order considered. Finally, we show on numerical experiments that the nonlinear scheme has optimal order of accuracy, and is less demanding, for a given accuracy, in terms of computational resources than a low-order Hybrid Finite Volume nonlinear method.

4.1 Motivations and context

We are interested in the discretisation of a linear advection-diffusion equation on general meshes with a high-order scheme. Our goal is to compare nonlinear structure-preserving high-order methods with similar low-order methods and linear high-order methods. More precisely, let Ω be an open, bounded, connected polytopal subset of $\mathbb{R}^d$, $d \in \{2,3\}$, with Lipschitz boundary. We consider the following linear advection-diffusion problem with homogeneous Neumann boundary conditions: find $u : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ solution to

$$\begin{cases} \partial_t u - \operatorname{div}(\Lambda(\nabla u + u\nabla\phi)) = 0 & \text{in } \mathbb{R}_+ \times \Omega, \\ \Lambda(\nabla u + u\nabla\phi)\cdot n = 0 & \text{on } \mathbb{R}_+ \times \partial\Omega, \\ u(0,\cdot) = u^{in} & \text{in } \Omega, \end{cases} \tag{4.1}$$

where n is the unit normal vector to $\partial\Omega$ pointing outward from Ω. We assume that the data satisfy:

(i) $\Lambda \in W^{1,\infty}(\Omega;\mathbb{R}^{d\times d})$ is a uniformly elliptic diffusion tensor: there exists $\lambda_\flat > 0$ such that, for a.e. x in Ω, $\Lambda(x)\xi\cdot\xi \geq \lambda_\flat|\xi|^2$ for all $\xi \in \mathbb{R}^d$;

(ii) $\phi \in C^1(\overline{\Omega})$ is a regular potential;

(iii) $u^{in} \in L^1(\Omega)$ is a non-negative initial datum, such that $\int_\Omega u^{in}\log\left(u^{in}\right) < \infty$.

The solutions to (4.1) enjoy some specific and well-known properties. First the mass of the solution is preserved along time, i.e. for almost every $t > 0$,

$$\int_\Omega u(t) = \int_\Omega u^{in} = M \tag{4.2}$$

where $M > 0$ is the initial mass. Second, the solution is positive:

$$\text{for } t > 0,\ u(t,\cdot) > 0 \text{ almost everywhere on } \Omega. \tag{4.3}$$

Last, the solution has a specific long-time behaviour: it converges exponentially fast when $t \to \infty$ towards the thermal equilibrium u^∞, solution to the stationary problem associated to (4.1), defined as

$$u^\infty = \frac{M}{\int_\Omega \mathrm{e}^{-\phi}}\,\mathrm{e}^{-\phi}. \tag{4.4}$$

In order to get a reliable numerical approximation of such problems, one has to preserve these structural properties at the discrete level. It is well-known that Two-Point Flux Approximation (TPFA) finite volume methods are structure-preserving (see [69] for the long-time behaviour), but these methods can only be used on meshes satisfying some orthogonality conditions (with

respect to the inner product induced by Λ), which essentially restricts the use of these methods to problems with isotropic tensors. On the other hand, a number of finite volume methods using auxiliary unknowns has been introduced for anisotropic problems on general meshes within the past twenty years. On can cite Discrete Duality Finite Volume (DDFV) methods, with additional unknowns on a dual mesh [157, 105], Vertex Approximate Gradient (VAG) methods, with auxiliary unknowns at the vertices of the mesh [124] or Hybrid Finite Volume (HFV) and Mimetic Finite Difference (MFD) methods, with auxiliary unknowns attached to the faces of the mesh [40, 123]. Such methods have proven to be relevant solutions to the anisotropy issue, but none of these linear methods preserves the positivity of the solutions (see [109]). A possible alternative was proposed in [56], with the introduction and analysis of a nonlinear positivity-preserving VAG scheme. The design and analysis of this scheme, as well as of its DDFV and HFV counterparts of [52, 70], relies on the entropy structure of (4.1): there exists some physically motivated quantity, called entropy, which decays along time. Reproducing this structure at the discrete level is a key point to get stability and convergence results.
One can also cite the methods introduced and analysed in [245, 113, 32], which rely on the introduction of nonlinearities within the schemes at the discrete level. These schemes exhibit a good robustness with respect to anisotropy and meshes for linear stationary diffusion. However they rely on nonlinearities which do not correspond to a particular equation at the continuous level. Therefore, it seems difficult to show that they enjoy discrete entropy structures (and associated long-time behaviour).

All the schemes discussed above are at most of order two in space in L^2-norm. A natural extension of these works is to develop similar methods with high-order accuracy. Such a task is not easy, because high-order polynomials are well-known to produce oscillations, which makes it difficult to ensure that the discrete solutions stay positive on the whole domain. There exist various positivity-preserving schemes in the litterature, based on Discontinuous Galerkin (DG) methods [13, 9, 10]. To our knowledge, the first high-order positive DG scheme for diffusive problems is the one introduced in [198] for linear advection-diffusion. The idea of this scheme is to choose some stabilisation parameters (depending on the degree of the polynomial unknowns and on the dimension of the domain) so that the positivity of the unknowns is ensured in average: the mean value of each cell unknown is positive, even though the (polynomial) unknowns can take negative values inside the cells. The main drawback of such a scheme lies in the difficulty to analyse it: since the solution can become negative on some area of the domain, it is not possible to adapt the analysis techniques used in the continuous framework, such as the use of the Boltzmann entropy.
Another positivity-preserving DG scheme was introduced in [36] for a Fisher–KPP equation $\partial_t u - \triangle u = u(1-u)$. This nonlinear scheme is developed so as to preserve the entropy structure of the continuous problem, and relies on a transformation of the equation based on the variable $\lambda = \log(u)$. The scheme provides positive densities, defined as $u = \mathrm{e}^{\lambda}$. Compared to the high-order DG schemes discussed above, the main improvement lies in the fact that the unknowns are positive everywhere. Such a feature allows the authors to fully analyse the scheme, including existence, long-time behaviour, and convergence towards a semi-discretised solution. The analysis is based on the properties of a well-chosen stabilisation function whose expression implies L^∞-norms of the polynomial unknowns along the faces of the mesh. One can also cite the space-time DG discretisation [37] for cross-diffusion systems.

The aim of this chapter is to introduce a high-order scheme for anisotropic problems on general meshes, preserving the three structural properties (4.2)-(4.3)-(4.4) discussed above. Since the HFV method can be seen as the low-order version of the Hybrid High-Order (HHO) scheme introduced in [100] for stationary diffusion, it is rather natural to try to adapt the nonlinear HFV scheme of [70] to the HHO framework. Our goal here is twofold. First, we want to design

a structure-preserving high-order scheme, and assert from a numerical point of view that the scheme is indeed of high order accuracy. On the other hand, we also want to compare this scheme with other approaches, based on two criteria:

(i) in terms of efficiency, with respect to the nonlinear low-order schemes of Chapter 1;

(ii) in terms of structure preservation, with respect to linear high-order schemes.

For the second criterion we mainly focus on the positivity of the discrete solutions. Indeed, with high-order schemes, one can expect to compute a numerical solution which is very close to the continuous one. Therefore, one could expect that the use of linear high-order schemes (with sufficiently high order) is already a solution in itself to preserve the positivity, at least in practice.

In order to perform this comparison, we introduce two HHO schemes. The first one is linear, and relies on the exponential fitting strategy [42] already used in Chapter 1. The key idea of this scheme is the change of unknown $\rho = u\,\mathrm{e}^{\phi}$, which allows one to reformulate (4.1) into an unconditionally coercive problem in ρ. As a by-product of this reformulation, we also get a scheme which preserves the thermal equilibrium. The second scheme is a nonlinear one, which can be seen as a high-order generalisation of the nonlinear HFV scheme introduced in Chapter 1. This scheme is designed so as to preserve the positivity of the discrete solutions. For the sake of simplicity, both schemes rely on a mixed-order HHO space: given an integer $k \geq 0$, the methods hinge on face unknowns of degree k and enriched cell unknowns of degree $k+1$. The main interest of such a discrepancy in the degree between face and cell unknowns is the simplification of the stabilisation term. In the meantime, such a choice preserves optimal accuracy (order $k+2$ in L^2-norm) and frugality (the linear system only takes into account the face unknowns, of degree k). This kind of method is sometimes referred to as HDG+ in the literature, and is associated to the so-called Lehrenfeld-Schöberl stabilisation [190, 191].

Our first theoretical results stated in Propositions 13 and 14 concern the well-posedness of the exponential fitting scheme and its long-time behaviour. Regarding the nonlinear scheme, we prove the existence of (positive) solutions in Theorem 10. Last, these discrete solutions are proved to converge (in time) towards an associated discrete thermal equilibrium in Proposition 16.

Note that one could also compare the schemes introduced in this chapter with the linear scheme for advection-diffusion of [98], based on separate discretisations of the advection and diffusion. This scheme is a high-order generalisation of the HMM scheme [19] which was considered in Chapter 1. As observed in [70], the low-order version of this scheme has two major drawbacks, namely the fact that it does not preserve the thermal equilibrium, and the need for some coercivity assumptions (which constrain the variety of potentials that can be considered for the analysis). We expect the high-order scheme to inherit these features. Moreover, in order to perform a fair comparison between the schemes, one should also consider a mixed-order version of the scheme of [98], but such a variation has, to our knowledge, not been studied yet. Thus, we will not consider this scheme here, but a numerical comparison including this scheme could be the subject of a future work .

The rest of this chapter is organised as follows. In Section 4.2, we first present the discrete framework, and describe the two schemes under consideration. Then, in Section 4.3 we discuss the main properties of both schemes, and provide some elements of analysis regarding the well-posedness and discrete long-time behaviours. Last, in Section 4.4 we discuss the implementation of the nonlinear scheme and we assert the behaviour of the schemes on some test-cases. These results show that the nonlinear high-order scheme offers a better efficiency in terms of computational cost than low-order schemes, while preserving the positivity of the solution, contrary to linear schemes.

4.2 Discrete setting and schemes

In this section, we introduce the discrete framework. Since our focus is mainly on the numerical aspects of the schemes, the description is concise. We refer the reader to [97, Section 1.1] for a more detailed description.
Here, as in the rest of this manuscript, we focus on space discretisation. Hence, both schemes are based on a backward Euler discretisation in time. For the sake of readability, we will consider a constant positive time step Δt. The time discretisation is then defined as $(t^n)_{n\in\mathbb{N}}$, where $t^n = n\Delta t$. Note that it is rather straightforward to generalise the discussion below to variable time steps.

4.2.1 Mesh

We define a discretisation of Ω as a couple $\mathcal{D} = (\mathcal{M}, \mathcal{E})$, where:

— the mesh $\mathcal{M}$ is a partition of Ω, i.e. $\mathcal{M}$ is a finite collection of disjoint, open, Lipschitz polytopes $K \subset \Omega$ with positive measure $|K| > 0$ (the cells) such that $\overline{\Omega} = \bigcup_{K\in\mathcal{M}} \overline{K}$;

— the set of faces $\mathcal{E}$ is a partition of the mesh skeleton $\bigcup_{K\in\mathcal{M}} \partial K$, i.e. $\mathcal{E}$ is a finite collection of disjoint, connected, relatively open subsets σ (the faces) of $\overline{\Omega}$ with positive measure $|\sigma| > 0$ such that $\bigcup_{K\in\mathcal{M}} \partial K = \bigcup_{\sigma\in\mathcal{E}} \overline{\sigma}$. It is assumed that, for all $\sigma \in \mathcal{E}$, σ is a Lipschitz polytopal subset of an hyperplane. We denote by $\mathcal{E}_K$ the set of faces of the cell K, and we let $n_{K,\sigma} \in \mathbb{R}^d$ be the (constant) unit normal vector to $\sigma \in \mathcal{E}_K$ pointing outward from K.

The diameter of a subset $X \subset \overline{\Omega}$ is denoted by

$$h_X = \sup\{|x - y| \,|\, (x, y) \in X^2\}.$$

We define the size of $\mathcal{D}$ as $h_\mathcal{D} = \sup\{h_K \mid K \in \mathcal{M}\}$.

When studying asymptotic behaviours with respect to the meshsize, one has to adopt a measure of regularity for (refined) families of discretisations. We classically follow [97, Definition 1.9], in which regularity for a refined family of discretisations is quantified by a uniform (with respect to the meshsize) parameter $0 < \theta < 1$, called regularity parameter. This parameter takes into account the ratios between the different characteristic scales of the mesh (typically, the size of the cells and the size of the faces).

4.2.2 Polynomials, discrete unknowns and discrete operators

In the following, k is a fixed non-negative integer. First, we introduce polynomial spaces on a subset $X \subset \overline{\Omega}$: $\mathbb{P}^k(X)$ and $\mathbb{P}^k(X)^d$ denote respectively the spaces of polynomial functions $X \to \mathbb{R}$ and polynomial vector fields $X \to \mathbb{R}^d$ of total degree at most k. If X is a geometric object of Hausdorff dimension $l \in \{1, 2, 3\}$, the dimension of the space $\mathbb{P}^k(X)$ is $\binom{l+k}{k}$. We also define the $L^2(X)$-orthogonal projector $\Pi_X^k : L^1(X) \to \mathbb{P}^k(X)$ such that, given any $v \in L^1(X)$, $\Pi_X^k(v)$ is the only element of $\mathbb{P}^k(X)$ which satisfies

$$\forall w \in \mathbb{P}^k(X), \int_X \Pi_X^k(v) w = \int_X v w.$$

Given any $K \in \mathcal{M}$, $\sigma \in \mathcal{E}_K$ and $v \in W^{1,1}(K)$, we also introduce the shortcut notation

$$\Pi_\sigma^k(v) = \Pi_\sigma^k\left(v_{|\sigma}\right).$$

We now introduce the set of discrete unknowns corresponding to the mixed-order HHO method [83, 97], with face unknowns of degree k and (enriched) cell unknowns of degree $k+1$:

$$\underline{V}_{\mathcal{D}}^k = \left\{ \underline{v}_{\mathcal{D}} = \left((v_K)_{K\in\mathcal{M}}, (v_\sigma)_{\sigma\in\mathcal{E}} \right) \middle| \begin{array}{ll} \forall K \in \mathcal{M}, & v_K \in \mathbb{P}^{k+1}(K) \\ \forall \sigma \in \mathcal{E}, & v_\sigma \in \mathbb{P}^k(\sigma) \end{array} \right\}.$$

Given a cell $K \in \mathcal{M}$, we let

$$\underline{V}_K^k = \mathbb{P}^{k+1}(K) \times \prod_{\sigma\in\mathcal{E}_K} \mathbb{P}^k(\sigma)$$

be the restriction of $\underline{V}_{\mathcal{D}}^k$ to K, and for any generic discrete unknown $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ we denote by $\underline{v}_K = \left(v_K, (v_\sigma)_{\sigma\in\mathcal{E}_K} \right) \in \underline{V}_K^k$ its local restriction to the cell K. Given any $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$, we associate two piecewise polynomial functions $v_{\mathcal{M}} : \Omega \to \mathbb{R}$ and $v_{\mathcal{E}} : \bigcup_{K\in\mathcal{M}} \partial K \to \mathbb{R}$ such that

$$v_{\mathcal{M}|K} = v_K \text{ for all } K \in \mathcal{M} \text{ and } v_{\mathcal{E}|\sigma} = v_\sigma \text{ for all } \sigma \in \mathcal{E}.$$

We also introduce $\underline{1}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ the discrete element such that $1_K = 1$ for any cell $K \in \mathcal{M}$ and $1_\sigma = 1$ for any face $\sigma \in \mathcal{E}$. Last, given a cell $K \in \mathcal{M}$, we define the local interpolator $\underline{I}_K^k : W^{1,1}(K) \to \underline{V}_K^k$ such that, for any $v \in W^{1,1}(K)$,

$$\underline{I}_K^k(v) = \left(\Pi_K^{k+1}(v), \left(\Pi_\sigma^k(v) \right)_{\sigma\in\mathcal{E}_K} \right).$$

Similarly, the global interpolator $\underline{I}_{\mathcal{D}}^k : W^{1,1}(\Omega) \to \underline{V}_{\mathcal{D}}^k$ is defined as

$$\underline{I}_{\mathcal{D}}^k(v) = \left(\left(\Pi_K^{k+1}(v) \right)_{K\in\mathcal{M}}, \left(\Pi_\sigma^k(v) \right)_{\sigma\in\mathcal{E}} \right) \ \forall v \in W^{1,1}(\Omega).$$

In order to analyse the schemes, one needs to introduce an H^1-like discrete semi-norm. To do so, given a cell $K \in \mathcal{M}$, we first introduce a local semi-norm $|\cdot|_{1,K}$ on $\underline{V}_K^k$, defined as

$$|\underline{v}_K|_{1,K}^2 = \|\nabla v_K\|_{L^2(K)^d}^2 + \sum_{\sigma\in\mathcal{E}_K} \frac{1}{h_\sigma} \|v_\sigma - v_K\|_{L^2(\sigma)}^2.$$

At the global level, we define the discrete H^1 semi-norm $|\cdot|_{1,\mathcal{D}}$ on $\underline{V}_{\mathcal{D}}^k$ as

$$|\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}} = \sqrt{\sum_{K\in\mathcal{M}} |\underline{v}_K|_{1,K}^2}. \tag{4.5}$$

Note that $|\cdot|_{1,\mathcal{D}}$ is not a norm on $\underline{V}_{\mathcal{D}}^k$, but any $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ satisfying $|\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}} = 0$ is a constant discrete element: there exists $c \in \mathbb{R}$ such that $\underline{v}_{\mathcal{D}} = c\,\underline{1}_{\mathcal{D}}$. In particular, this implies that $|\cdot|_{1,\mathcal{D}}$ is a norm on the space $\underline{V}_{\mathcal{D},0}^k$ of null-mass discrete unknowns, defined as

$$\underline{V}_{\mathcal{D},0}^k = \left\{ \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k \mid \int_\Omega v_{\mathcal{M}} = 0 \right\}.$$

The HHO schemes are based on a local approach, which ensures the local conservativity of the method. Given a cell $K \in \mathcal{M}$, we introduce local operators defined on $\underline{V}_K^k$, which are split into a consistent part and a stabilisation part. We first define a local discrete gradient operator

$G_K^k : \underline{V}_K^k \to \mathbb{P}^k(K)^d$ such that, for any $\underline{v}_K \in \underline{V}_K^k$, $G_K^k(\underline{v}_K)$ satisfies

$$\int_K G_K^k(\underline{v}_K)\cdot\tau = \int_K \nabla v_K\cdot\tau + \sum_{\sigma\in\mathcal{E}_K}\int_\sigma (v_\sigma - v_K)\tau\cdot n_{K,\sigma} \quad \forall \tau\in\mathbb{P}^k(K)^d. \tag{4.6}$$

This operator is a discrete counterpart of the continuous gradient and is consistent. It satisfies the following commutation property:

$$\forall v\in W^{1,1}(K),\ G_K^k\circ \underline{I}_K^k(v) = \Pi_K^k(\nabla v),$$

where Π_K^k acts component-wise. However, this operator suffers from a lack of coercivity. To ensure stability, we need to add a stabilisation, which controls the jumps between the cell unknown (or, to be more precise, the trace of this unknown on ∂K) and the face unknowns. To do so, given a face $\sigma\in\mathcal{E}_K$, we define the jump operator $J_{K,\sigma} : \underline{V}_K^k \to \mathbb{P}^k(\sigma)$ by

$$J_{K,\sigma}(\underline{v}_K) = \Pi_\sigma^k(v_K) - v_\sigma. \tag{4.7}$$

One can now introduce the classical (see [83, Section 3.2.1] and [97, Sections 4.2 and 5.1]) discrete counterpart of the local bilinear form $(w,v)\mapsto \int_K \Lambda\nabla w\cdot\nabla v$ as the bilinear form $a_K : \underline{V}_K^k\times\underline{V}_K^k\to\mathbb{R}$ such that

$$a_K : (\underline{w}_K,\underline{v}_K)\mapsto \int_K \Lambda G_K^k(\underline{w}_K)\cdot G_K^k(\underline{v}_K) + \eta_l\sum_{\sigma\in\mathcal{E}_K}\frac{\Lambda_{K\sigma}}{h_\sigma}\int_\sigma J_{K,\sigma}(\underline{w}_K)J_{K,\sigma}(\underline{v}_K), \tag{4.8}$$

where $\Lambda_{K\sigma} = \|\Lambda_{|K} n_{K,\sigma}\cdot n_{K,\sigma}\|_{L^\infty(\sigma)}$ and $\eta_l > 0$ is a stabilisation parameter (l stands for linear). The stabilisation used here, based on the jump operators $J_{K,\sigma}$, is the main specificity of the mixed-order HHO method with respect to the equal-order one [100]. Note that such a strategy used in mixed-form (formulation on the flux) is sometimes called HDG+ methods, and the corresponding stabilisation is called Lehrenfeld-Schöberl stabilisation, as it was first introduced in [190, 191]. Here, as in the rest of this manuscript, the HHO scheme is written on primal form.

One can now define a global bilinear form $a_\mathcal{D} : \underline{V}_\mathcal{D}^k\times\underline{V}_\mathcal{D}^k\to\mathbb{R}$, discretisation of the continuous one $(w,v)\mapsto\int_\Omega\Lambda\nabla w\cdot\nabla v$, by summing the local contributions:

$$a_\mathcal{D}(\underline{w}_\mathcal{D},\underline{v}_\mathcal{D}) = \sum_{K\in\mathcal{M}} a_K(\underline{w}_K,\underline{v}_K). \tag{4.9}$$

In terms of analysis, the definition of $a_\mathcal{D}$ implies (see for example [97, Section 5.1.6]) that the following stability estimate holds: there exists a positive constant $\alpha_\flat$, depending only on Ω, k and θ (the regularity of the mesh) such that

$$\forall \underline{v}_\mathcal{D}\in\underline{V}_\mathcal{D}^k,\ a_\mathcal{D}(\underline{v}_\mathcal{D},\underline{v}_\mathcal{D}) \geq \lambda_\flat\alpha_\flat|\underline{v}_\mathcal{D}|_{1,\mathcal{D}}^2. \tag{4.10}$$

In particular, since $|\cdot|_{1,\mathcal{D}}$ is a norm on $\underline{V}_{\mathcal{D},0}^k$, this estimate implies that $a_\mathcal{D}$ is a coercive bilinear form on $\underline{V}_{\mathcal{D},0}^k$.

In order to perform the analysis of the linear scheme, as in the continuous setting, we need a control of the L^2-norm of $v_\mathcal{M}$ by its energy-norm. This is the purpose of the discrete Poincaré-Wirtinger inequality: there exists a positive constant C_P, depending only on Ω, k and θ such that

$$\forall \underline{v}_\mathcal{D}\in\underline{V}_{\mathcal{D},0}^k,\ \|v_\mathcal{M}\|_{L^2(\Omega)} \leq C_P|\underline{v}_\mathcal{D}|_{1,\mathcal{D}}. \tag{4.11}$$

We refer to [97, Theorem 6.5, $p = q = 2$] for a demonstration of this result (in the equal-order case).

4.2.3 Exponential fitting scheme (linear)

The first scheme introduced herein is a linear one, which follows the exponential fitting strategy. Such an idea has, to our knowledge, initially been introduced at the numerical level in [43] in the context of mixed finite elements. In Chapter 1, we introduce and analyse an HFV scheme based on this strategy. In a nutshell, the exponential fitting approach hinges on the following reformulation of the continuous flux: letting $\omega = \mathrm{e}^{-\phi}$, we introduce the Slotboom variable $\rho = \dfrac{u}{\omega}$. Then, at least formally, one has the following relation:

$$\nabla u + u\nabla\phi = \omega\nabla\rho, \tag{4.12}$$

which allows one to transform an advection-diffusion equation into a purely diffusive problem. The main goal of this procedure is to ensure the coercivity of the new problem (since, by regularity of ϕ, ω is uniformly bounded away from zero). The exponential fitting scheme is based on this formulation. We discretise the Slotboom variable as a polynomial, that is to say we approximate ρ as a discrete unknown in $\underline{V}_{\mathcal{D}}^k$. Then, we reconstruct the discrete density by mimicking the continuous relation $u = \omega \times \rho$. Thus, a solution $(\underline{\rho}_{\mathcal{D}}^n)_{n\geq 1}$ to the scheme (4.14) described below corresponds to an approximation of the Slotboom variable associated to the density u.

More precisely, for a given discrete Slotboom variable $\underline{\rho}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$, we associate a discrete density

$$\underline{\underline{u}}_{\mathcal{D}} = \left((u_K^\omega)_{K\in\mathcal{M}}, (u_\sigma^\omega)_{\sigma\in\mathcal{E}}\right)$$

defined as a collection of (non-polynomial) regular functions:

$$\forall K \in \mathcal{M},\ u_K^\omega = \omega_{|K}\rho_K \quad \text{and} \quad \forall \sigma \in \mathcal{E},\ u_\sigma^\omega = \omega_{|\sigma}\rho_\sigma.$$

To this discrete density, we associate two piecewise continuous functions: $u_{\mathcal{M}}^\omega : \Omega \to \mathbb{R}$, the discrete density on the cells and $u_{\mathcal{E}}^\omega : \bigcup_{K\in\mathcal{M}} \partial K \to \mathbb{R}$, the discrete density on the faces. These reconstructed densities are defined as

$$u_{\mathcal{M}|K}^\omega = u_K^\omega \text{ for all } K \in \mathcal{M} \quad \text{and} \quad u_{\mathcal{E}|\sigma}^\omega = u_\sigma^\omega \text{ for all } \sigma \in \mathcal{E}. \tag{4.13}$$

Note that $\underline{\underline{u}}_{\mathcal{D}}$ is not a collection of polynomials in general (as highlighted by the use of a double underline).

Remark 19 (Non-polynomial reconstruction)**.** *Here, we choose to reconstruct the density as a non-polynomial function, since we locally multiply the polynomial unknown ρ_K (or ρ_σ for the faces) by the continuous (and, a priori non-polynomial) ω. One could also think of reconstructing a polynomial unknown, by multiplying (component by component) $\underline{\rho}_{\mathcal{D}}$ by $\underline{I}_{\mathcal{D}}^k(\omega)$. Such a polynomial reconstruction is the choice made for the HFV exponential fitting scheme of Chapter 1, in the sense that we use an interpolation of ω to define the discrete density. But for the low-order scheme of Chapter 1, both polynomials are constants, and therefore the reconstructed density is also constant. For the arbitrary-order scheme presented here, we believe that a polynomial reconstruction of the density is not so relevant. Indeed, the main issue of using $\underline{I}_{\mathcal{D}}^k(\omega)$ lies in the lack of positivity: there is no way of ensuring that the polynomials constituting $\underline{I}_{\mathcal{D}}^k(\omega)$ are positive on their definition domain. Therefore, the use of a*

polynomial reconstruction should lead to larger positivity issues. Such issues do not exist in Chapter 1, since the interpolation is positivity preserving.

In order to define a mixed-order HHO exponential fitting scheme, one needs to introduce a discrete counterpart of $(\rho, v) \mapsto \int_\Omega \omega \Lambda \nabla \rho \cdot \nabla v$. To do so, given $K \in \mathcal{M}$, leveraging the definition (4.8) of a_K, we define at the local level $a_K^\omega : \underline{V}_K^k \times \underline{V}_K^k \to \mathbb{R}$ as the bilinear form

$$a_K^\omega : (\underline{\rho}_K, \underline{v}_K) \mapsto \int_K \omega \Lambda G_K^k(\underline{\rho}_K) \cdot G_K^k(\underline{v}_K) + \eta_l \sum_{\sigma \in \mathcal{E}_K} \frac{\Lambda_{K\sigma}^\omega}{h_\sigma} \int_\sigma J_{K,\sigma}(\underline{\rho}_K) J_{K,\sigma}(\underline{v}_K),$$

where $\Lambda_{K\sigma}^\omega = \|\omega \Lambda_{|K} n_{K,\sigma} \cdot n_{K,\sigma}\|_{L^\infty(\sigma)}$. At the global level, as previously, we construct the bilinear form $a_\mathcal{D}^\omega : \underline{V}_\mathcal{D}^k \times \underline{V}_\mathcal{D}^k \to \mathbb{R}$ by summing the local contributions:

$$a_\mathcal{D}^\omega : (\underline{\rho}_\mathcal{D}, \underline{v}_\mathcal{D}) \mapsto \sum_{K \in \mathcal{M}} a_K^\omega(\underline{\rho}_K, \underline{v}_K).$$

We can now introduce the exponential fitting scheme for Problem (4.1): find $(\underline{\rho}_\mathcal{D}^n)_{n\geq 1} \in \left(\underline{V}_\mathcal{D}^k\right)^{\mathbb{N}^*}$ such that, for all $n \in \mathbb{N}$,

$$\begin{cases} \displaystyle\int_\Omega \frac{u_\mathcal{M}^{\omega,n+1} - u_\mathcal{M}^{\omega,n}}{\Delta t} v_\mathcal{M} = -a_\mathcal{D}^\omega(\underline{\rho}_\mathcal{D}^{n+1}, \underline{v}_\mathcal{D}) & \forall \underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}^k, \quad (4.14a) \\ u_{\mathcal{M}|K}^{\omega,n+1} = \omega_{|K} \rho_K^{n+1} & \forall K \in \mathcal{M}, \quad (4.14b) \\ u_{\mathcal{M}|K}^{\omega,0} = u_{|K}^{in} & \forall K \in \mathcal{M}. \quad (4.14c) \end{cases}$$

Given a solution $(\underline{\rho}_\mathcal{D}^n)_{n\geq 1}$ to (4.14), as described above, we associate a sequence of reconstructed densities $(\underline{\underline{u}}_\mathcal{D}^n)_{n\geq 1}$.

One can check that, for any $c \in \mathbb{R}$, the discrete element $\underline{\rho}_\mathcal{D} = c\underline{1}_\mathcal{D} \in \underline{V}_\mathcal{D}^k$ is a stationary solution to the exponential fitting scheme (4.14) (in Slotboom variable). Therefore, we define the discrete Slotboom variable at equilibrium $\underline{\rho}_\mathcal{D}^\infty \in \underline{V}_\mathcal{D}^k$ as

$$\underline{\rho}_\mathcal{D}^\infty = c_\omega^M \underline{1}_\mathcal{D}, \text{ where } c_\omega^M = \frac{M}{\int_\Omega \omega}.$$

It is naturally associated to the the discrete thermal equilibrium $\underline{\underline{u}}_\mathcal{D}^\infty = \left((u_K^{\omega,\infty})_{K\in\mathcal{M}}, (u_\sigma^{\omega,\infty})_{\sigma\in\mathcal{E}}\right)$. It is the only stationary solution to (4.14) associated to a density of mass M.

4.2.4 Nonlinear scheme

Following the ideas from [56, 70], our nonlinear scheme relies on a nonlinear reformulation of Problem (4.1). To do so, we introduce the logarithm potential $\ell = \log(u)$ and the quasi-Fermi potential $w = \ell + \phi$. At least formally, if u is positive, one has the following relation:

$$\nabla u + u \nabla \phi = u \nabla (\log(u) + \phi) = \mathrm{e}^\ell \nabla w. \quad (4.15)$$

We choose to discretise the potentials as polynomials, i.e. to approximate ℓ and w as discrete unknowns in $\underline{V}_\mathcal{D}^k$. Then, mimicking the relation $u = \mathrm{e}^\ell$, we reconstruct the density as the

exponential of polynomial functions, thus ensuring its positivity. Therefore, a solution $(\underline{\ell}_{\mathcal{D}}^n)_{n\geq 1}$ to the nonlinear scheme (4.20) described below corresponds to an approximation of the logarithms of the solution u (density).

More specifically, for a given discretisation $\underline{\ell}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ of the potential ℓ, one associates a discrete density

$$\underset{\sim}{u}_{\mathcal{D}} = ((u_K)_{K\in\mathcal{M}}, (u_\sigma)_{\sigma\in\mathcal{E}}),$$

defined as a collection of positive smooth functions

$$\forall K \in \mathcal{M},\ u_K = \exp(\ell_K) \quad \text{and} \quad \forall \sigma \in \mathcal{E},\ u_\sigma = \exp(\ell_\sigma).$$

As previously, we associate to this discrete density a couple of piecewise smooth functions $u_{\mathcal{M}} : \Omega \to \mathbb{R}$ (corresponding to the cell unknowns) and $u_{\mathcal{E}} : \bigcup_{K\in\mathcal{M}} \partial K \to \mathbb{R}$ (corresponding to the face unknowns), defined as

$$u_{\mathcal{M}} = \exp(\ell_{\mathcal{M}}) \text{ and } u_{\mathcal{E}} = \exp(\ell_{\mathcal{E}}). \tag{4.16}$$

Note that a discrete density $\underset{\sim}{u}_{\mathcal{D}}$ is not a collection of polynomials (which is highlighted by the use of the wave under u), but it enjoys positivity, both on the cells and faces, since it is defined as the exponential of real functions.

As previously, this scheme is based on local cell contributions, split into a consistent term and a stabilisation term. Given a cell $K \in \mathcal{M}$, we define a discretisation of $(\ell, w, v) \mapsto \int_K \mathrm{e}^{\ell}\, \Lambda \nabla w \cdot \nabla v$ as a sum of nonlinear consistent (4.17a) and stabilising (4.17b) contributions:

$$\mathcal{C}_K(\underline{\ell}_K, \underline{w}_K, \underline{v}_K) = \int_K \mathrm{e}^{\ell_K}\, \Lambda G_K^k(\underline{w}_K) \cdot G_K^k(\underline{v}_K), \tag{4.17a}$$

$$\mathcal{S}_K(\underline{\ell}_K, \underline{w}_K, \underline{v}_K) = \eta_{nl} \sum_{\sigma\in\mathcal{E}_K} \frac{\Lambda_{K\sigma}}{h_\sigma} \int_\sigma \frac{\mathrm{e}^{\Pi_\sigma^k(\ell_K)} + \mathrm{e}^{\ell_\sigma}}{2} J_{K,\sigma}(\underline{w}_K) J_{K,\sigma}(\underline{v}_K), \tag{4.17b}$$

where $\eta_{nl} > 0$ is a stabilisation parameter (nl stands for nonlinear). We can now define a local application $\mathcal{T}_K : \underline{V}_K^k \times \underline{V}_K^k \times \underline{V}_K^k \to \mathbb{R}$ by

$$\mathcal{T}_K(\underline{\ell}_K, \underline{w}_K, \underline{v}_K) = \mathcal{C}_K(\underline{\ell}_K, \underline{w}_K, \underline{v}_K) + \mathcal{S}_K(\underline{\ell}_K, \underline{w}_K, \underline{v}_K) + \varepsilon h_K^{k+2} a_K(\underline{w}_K, \underline{v}_K), \tag{4.18}$$

where ε is a non-negative parameter and a_K is the bilinear form defined in (4.8)-(4.9). At the global level, we define $\mathcal{T}_{\mathcal{D}} : \underline{V}_{\mathcal{D}}^k \times \underline{V}_{\mathcal{D}}^k \times \underline{V}_{\mathcal{D}}^k \to \mathbb{R}$ by summing the local contributions:

$$\mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = \sum_{K\in\mathcal{M}} \mathcal{T}_K(\underline{\ell}_K, \underline{w}_K, \underline{v}_K). \tag{4.19}$$

Remark 20 (Parameter ε)**.** *The operator $\mathcal{T}_{\mathcal{D}}$ is to be understood as a discretisation of $(\ell, w, v) \mapsto \int_\Omega (\mathrm{e}^{\ell} + \epsilon) \Lambda \nabla w \cdot \nabla v$, with $\epsilon \sim \varepsilon h_{\mathcal{D}}^{k+2}$ a small parameter. The ϵ perturbation is used in order to show the existence result of Proposition 10 and can be seen as a kind of stabilisation. The scaling factor h_K^{k+2} in (4.18) is used to get the expected order of convergence. Although the choice made here gives the right order of convergence, the use of larger parameters (such as h_K^{k+1}) has not yet been studied. In practice, numerical results for $\varepsilon = 1$ and $\varepsilon = 0$ are almost the same. The influence of this term will be further investigated in future works.*

We now let $\underline{\phi}_{\mathcal{D}} = \underline{I}_{\mathcal{D}}^k(\phi) \in \underline{V}_{\mathcal{D}}^k$ be the interpolate of ϕ. Now, using a backward Euler dis-

cretisation in time with time step $\Delta t > 0$, we introduce the following scheme for (4.1): find $(\underline{\ell}_{\mathcal{D}}^n)_{n\geq 1} \in \left(\underline{V}_{\mathcal{D}}^k\right)^{\mathbb{N}^*}$ such that, for all $n \geq 0$,

$$\begin{cases} \displaystyle\int_\Omega \frac{u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}} = -\mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{n+1}, \underline{\ell}_{\mathcal{D}}^{n+1} + \underline{\phi}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) & \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k, \quad (4.20a) \\ u_K^{n+1} = \mathrm{e}^{\ell_K^{n+1}} & \forall K \in \mathcal{M}, \quad (4.20b) \\ u_K^0 = u^{in}{}_{|K} & \forall K \in \mathcal{M}. \quad (4.20c) \end{cases}$$

Given a solution $(\underline{\ell}_{\mathcal{D}}^n)_{n\geq 1}$ to the scheme (4.20) (which are approximations of the logarithm of the density at times $t^n = n\Delta t$), as discussed above, we associate a sequence of positive discrete densities $\left(\underset{\sim}{u}_{\mathcal{D}}^n\right)_{n\geq 1}$.

Remark 21 (Initial condition)**.** *Note that the scheme (4.20) does not rely on discrete initial data, as described in (4.20c). This strategy allows one not to define $\underline{\ell}_{\mathcal{D}}^0$ (as the interpolate of $\ell^{in} = \log(u^{in})$), which is not possible in regions where u^{in} vanishes.*
The question of defining an initial discrete datum is a major difficulty when it comes to the numerical implementation and the initialisation of the Newton method (see Section 4.4.1).

We define the discrete thermal equilibrium as $\underset{\sim}{u}_{\mathcal{D}}^\infty = \left((u_K^\infty)_{K\in\mathcal{M}}, (u_\sigma^\infty)_{\sigma\in\mathcal{E}}\right)$ where

$$\forall K \in \mathcal{M},\ u_K^\infty = c_{nl}^M \mathrm{e}^{-\phi_K} \quad \text{and} \quad \forall \sigma \in \mathcal{E},\ u_\sigma^\infty = c_{nl}^M \mathrm{e}^{-\phi_\sigma},$$

with $c_{nl}^M = M/\int_\Omega \mathrm{e}^{-\phi_{\mathcal{M}}}$. One can show that $\underset{\sim}{u}_{\mathcal{D}}^\infty$ (and the associated logarithm potential at equilibrium $\underline{\ell}_{\mathcal{D}}^\infty = \log(c_{nl}^M)\underline{1}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$) is the only stationary solution to (4.20) with mass M.

4.3 Main features of the schemes

In this section, we present some theoretical results about the two schemes introduced above. We especially focus on the existence (and stability) of solutions, as well as on questions related to the long-time behaviour. The results presented below are generalisations of the analysis for low-order schemes presented in Chapter 1. In particular, the analysis strongly relies on the entropy structure of both schemes.

Remark 22 (Low-order versions of the schemes ($k = 0$))**.** *Note that the low-order versions of the two schemes introduced above do not coincide with the exponential fitting and nonlinear HFV schemes of Chapter 1. Indeed, even if the equal-order HHO method (face and cell unknowns of the same degree k) for $k = 0$ is essentially equivalent to the HFV method for diffusion problems [100], such a result does not hold for the mixed-order HHO method. In fact, the low-order versions of the schemes (4.14) and (4.20) have enriched cell unknowns, which are affine functions (and not constant). Therefore, the results presented here are new, even for $k = 0$.*

Before presenting individual results for each scheme, it should be noted that both schemes exhibits a similar important property: the preservation of the thermal equilibrium. Indeed, both scheme admit an unique stationary solution with mass M, namely $\underline{\rho}_{\mathcal{D}}^\infty = c_\omega^M \underline{1}_{\mathcal{D}}$ and $\underline{\ell}_{\mathcal{D}}^\infty = \log(c_{nl}^M)\underline{1}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}}$, which are associated to the following discrete cell densities:

$$u_{\mathcal{M}}^{\omega,\infty} = \frac{M}{\int_\Omega \mathrm{e}^{-\phi}} \mathrm{e}^{-\phi} \quad \text{and} \quad u_{\mathcal{M}}^\infty = \frac{M}{\int_\Omega \mathrm{e}^{-\phi_{\mathcal{M}}}} \mathrm{e}^{-\phi_{\mathcal{M}}}.$$

One can notice that $u_{\mathcal{M}}^{\omega,\infty}$ coincides with the continuous thermal equilibrium u^∞. This is a consequence of the choice of a non-polynomial reconstruction of the density in the exponential fitting scheme (see Remark 19). For the nonlinear scheme, we do not preserve the exact thermal equilibrium, since the discrete densities for this scheme are defined as exponentials of polynomial functions (so, contrarily to the scheme of Chapter 1, the discrete equilibrium is not $\underline{I}_{\mathcal{D}}^k(u^\infty)$. However, the discrete thermal equilibrium $u_{\mathcal{M}}^\infty$ is a reasonable discretisation of u^∞, in the sense that the logarithm potential of the discrete thermal equilibrium and the interpolate of the continuous equilibrium logarithm potential are equal up to an additive constant:

$$\forall K \in \mathcal{M},\ \log(u_K^\infty) - \Pi_K^{k+1}(\log(u^\infty)) = \log\left(\int_\Omega \mathrm{e}^{-\phi}\right) - \log\left(\int_\Omega \mathrm{e}^{-\phi_{\mathcal{M}}}\right).$$

In fact, at the continuous level, the thermodynamic consistency is sometimes expressed as follows: if the advection-diffusion flux $J = \nabla u + u\nabla\phi$ vanishes on Ω, then the quasi-Fermi potential $\log(u) + \phi$ is constant in Ω. Both schemes satisfy discrete counterparts of this principle: for the exponential fitting scheme, $\log(u_{\mathcal{M}}^\omega) + \phi$ and $\log(u_{\mathcal{E}}^\omega) + \phi_{|\partial\mathcal{M}}$ are constant; while for the for the nonlinear scheme, $\underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}$ is a constant discrete element (i.e. proportional to $\underline{1}_{\mathcal{D}}$).

4.3.1 Exponential fitting scheme

We present here the main properties of the exponential fitting scheme (4.14), and give detailed proofs of the results.

As a preliminary remark, note that since ϕ is C^1 on $\overline{\Omega}$, there exist positive constants $\omega_\flat$ and $\omega_\sharp$ such that

$$\forall x \in \Omega,\ \omega_\flat \le \omega(x) \le \omega_\sharp.$$

As a consequence, the diffusion tensor $\omega\Lambda$ is uniformly elliptic. This implies a stability estimate on the bilinear form $a_{\mathcal{D}}^\omega$ on $\underline{V}_{\mathcal{D}}^k$: there exists a positive constant $\alpha_\flat^\omega = \omega_\flat\lambda_\flat\alpha_\flat$ (which therefore depends only on Ω, k ϕ, Λ and the regularity of the mesh θ) such that

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k,\ a_{\mathcal{D}}^\omega(\underline{v}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \ge \alpha_\flat^\omega |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}^2. \tag{4.21}$$

We first state an existence result, which is mainly a consequence of the previous stability estimate.

Proposition 13 (Well-posedness of the exponential fitting scheme)**.** *The exponential fitting scheme* (4.14) *admits an unique solution* $(\underline{\rho}_{\mathcal{D}}^n)_{n\ge1}$. *Moreover, the corresponding discrete densities* $(\underline{\underline{u}}_{\mathcal{D}}^n)_{n\ge1}$ *have a mass equal to* M*:*

$$\forall n \ge 1,\ \int_\Omega u_{\mathcal{M}}^{\omega,n} = \int_\Omega u^{in} = M. \tag{4.22}$$

Proof. Let $n \ge 0$, and assume that $u_{\mathcal{M}}^{\omega,n}$ is defined. We want to show that equation (4.14b) admits a unique solution. To do so, we first define, for any $\underline{\rho}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$,

$$\|\underline{\rho}_{\mathcal{D}}\|_{\omega,\Delta t,\mathcal{D}} = \sqrt{\Delta t|\underline{\rho}_{\mathcal{D}}|_{1,\mathcal{D}}^2 + \int_\Omega \omega\rho_{\mathcal{M}}^2}.$$

Since $|\cdot|_{1,\mathcal{D}}$ is a semi-norm on $\underline{V}_{\mathcal{D}}^k$, it follows that $\|.\|_{\omega,\Delta t,\mathcal{D}}$ is an Euclidean norm on $\underline{V}_{\mathcal{D}}^k$. Thus,

by (4.21), the bilinear form $A : (\underline{\rho}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \mapsto \int_\Omega \omega \rho_{\mathcal{M}} v_{\mathcal{M}} + \Delta t a^\omega_{\mathcal{D}}(\underline{\rho}_{\mathcal{D}}, \underline{v}_{\mathcal{D}})$ satisfies the following coercivity estimate:

$$\forall \underline{\rho}_{\mathcal{D}} \in \underline{V}^k_{\mathcal{D}},\ A(\underline{\rho}_{\mathcal{D}}, \underline{\rho}_{\mathcal{D}}) \geq \min(1, \alpha^\omega_\flat) \|\underline{\rho}_{\mathcal{D}}\|^2_{\omega,\Delta t,\mathcal{D}}.$$

Therefore, by Lax-Milgram theorem, equation (4.14b) admits an unique solution $\underline{\rho}^{n+1}_{\mathcal{D}}$ in $\underline{V}^k_{\mathcal{D}}$. To get the mass preservation identity (4.22), we test (4.14a) by $\underline{1}_{\mathcal{D}}$ to get

$$\int_\Omega \frac{u^{\omega,n+1}_{\mathcal{M}} - u^{\omega,n}_{\mathcal{M}}}{\Delta t} = 0.$$

We conclude by noticing that $\int_\Omega u^{\omega,0}_{\mathcal{M}} = \int_\Omega u^{in}$ according to (4.14c). □

We now state our main result about the exponential fitting scheme, which ensures that the solution to (4.14) has similar long-time behaviour as the continuous solution. As usual with the entropy method, the main idea is to get a control of the entropy by its dissipation. Here, such an estimate is a consequence of the discrete Poincaré inequality (4.11).

Proposition 14 (Long-time behaviour of the exponential fitting scheme)**.** *Assume that $u^{in} \in L^2(\Omega)$, and let $(\underline{\rho}^n_{\mathcal{D}})_{n\geq 1}$ be the solution to the exponential fitting scheme (4.14). Then, the following discrete entropy relation holds:*

$$\forall n \in \mathbb{N}, \quad \frac{\mathbb{E}^{n+1}_\omega - \mathbb{E}^n_\omega}{\Delta t} \leq -\mathbb{D}^{n+1}_\omega, \tag{4.23}$$

where the discrete quadratic entropy is defined as

$$\mathbb{E}^n_\omega = \frac{1}{2}\int_\Omega \omega |\rho^n_{\mathcal{M}} - \rho^\infty_{\mathcal{M}}|^2 \text{ with } \rho^0_{\mathcal{M}} = \frac{u^{in}}{\omega} \in L^2(\Omega)$$

and the discrete dissipation is defined by

$$\mathbb{D}^n_\omega = a^\omega_{\mathcal{D}}(\underline{\rho}^n_{\mathcal{D}} - \underline{\rho}^\infty_{\mathcal{D}}, \underline{\rho}^n_{\mathcal{D}} - \underline{\rho}^\infty_{\mathcal{D}})\ \forall n \geq 1.$$

As a consequence, the discrete density converges exponentially fast in time towards the corresponding discrete density at equilibrium: there exist positive constants C_ω and ν_ω only depending on the data and on the mesh regularity parameter θ such that

$$\forall n \in \mathbb{N}, \|u^{\omega,n}_{\mathcal{M}} - u^{\omega,\infty}_{\mathcal{M}}\|_{L^2(\Omega)} \leq C_\omega (1 + \nu_\omega \Delta t)^{-\frac{n}{2}} \|u^{in} - u^{\omega,\infty}_{\mathcal{M}}\|_{L^2(\Omega)}. \tag{4.24}$$

Proof. Let $n \in \mathbb{N}$. By convexity of $x \mapsto x^2$ on $\mathbb{R}$, one has

$$\mathbb{E}^{n+1}_\omega - \mathbb{E}^n_\omega = \frac{1}{2}\int_\Omega \omega\left(|\rho^{n+1}_{\mathcal{M}} - \rho^\infty_{\mathcal{M}}|^2 - |\rho^n_{\mathcal{M}} - \rho^\infty_{\mathcal{M}}|^2\right) \leq \int_\Omega \omega\left(\rho^{n+1}_{\mathcal{M}} - \rho^n_{\mathcal{M}}\right)\left(\rho^{n+1}_{\mathcal{M}} - \rho^\infty_{\mathcal{M}}\right).$$

Therefore, testing (4.14a) against $\underline{\rho}^{n+1}_{\mathcal{D}} - \underline{\rho}^\infty_{\mathcal{D}} \in \underline{V}^k_{\mathcal{D}}$, we get

$$\frac{\mathbb{E}^{n+1}_\omega - \mathbb{E}^n_\omega}{\Delta t} \leq -a^\omega_{\mathcal{D}}(\underline{\rho}^{n+1}_{\mathcal{D}}, \underline{\rho}^{n+1}_{\mathcal{D}} - \underline{\rho}^\infty_{\mathcal{D}}).$$

Note that this estimate holds true also for $n = 0$ (using the definition of $\rho^0_{\mathcal{M}}$). On the other hand,

by definition of $\underline{\rho}_{\mathcal{D}}^{\infty}$ (proportional to $\underline{1}_{\mathcal{D}}$) alongside with the bilinearity of $a_{\mathcal{D}}^{\omega}$, one has

$$a_{\mathcal{D}}^{\omega}(\underline{\rho}_{\mathcal{D}}^{\infty}, \underline{\rho}_{\mathcal{D}}^{n+1} - \underline{\rho}_{\mathcal{D}}^{\infty}) = 0 \text{ and } a_{\mathcal{D}}^{\omega}(\underline{\rho}_{\mathcal{D}}^{n+1}, \underline{\rho}_{\mathcal{D}}^{n+1} - \underline{\rho}_{\mathcal{D}}^{\infty}) = \mathbb{D}_{\omega}^{n+1},$$

which yields the entropy relation (4.23). To get the exponential decay, one needs to compare $\mathbb{D}_{\omega}^{n+1}$ with $\mathbb{E}_{\omega}^{n+1}$. To do so, we let $\underline{v}_{\mathcal{D}} = \underline{\rho}_{\mathcal{D}}^{n+1} - \underline{\rho}_{\mathcal{D}}^{\infty} \in \underline{V}_{\mathcal{D}}^{k}$, and we define the probability measure $\mu = \frac{\omega}{M_\omega}\mathrm{d}x$ on Ω, where $M_\omega = \int_\Omega \omega$. We define $\mu v_{\mathcal{M}}$ as the mass of $v_{\mathcal{M}}$ for the measure μ, i.e.

$$\mu v_{\mathcal{M}} = \int_\Omega v_{\mathcal{M}} \mathrm{d}\mu.$$

The mass preservation identity (4.22) implies that

$$\mu v_{\mathcal{M}} = \int_\Omega v_{\mathcal{M}} \mathrm{d}\mu = \frac{1}{M_\omega}\int_\Omega \omega v_{\mathcal{M}} = \frac{1}{M_\omega}\int_\Omega \left(u_{\mathcal{M}}^{\omega,n+1} - u_{\mathcal{M}}^{\omega,\infty}\right) = \frac{M-M}{M_\omega} = 0.$$

Therefore, letting $\overline{v} = \frac{1}{|\Omega|}\int_\Omega v_{\mathcal{M}}$, we apply [51, Lemma 5.2, $q = 2$ and μ as defined above] to get

$$\int_\Omega \frac{\omega}{M_\omega} v_{\mathcal{M}}^2 = \int_\Omega (v_{\mathcal{M}} - \mu v_{\mathcal{M}})^2 \,\mathrm{d}\mu \le 4\int_\Omega (v_{\mathcal{M}} - \overline{v})^2 \,\mathrm{d}\mu.$$

Using the upper bound on ω and the definition of μ, we infer that

$$\int_\Omega \omega v_{\mathcal{M}}^2 \le 4\omega_\sharp \int_\Omega (v_{\mathcal{M}} - \bar{v})^2 = 4\omega_\sharp \|v_{\mathcal{M}} - \bar{v}\|_{L^2(\Omega)}^2.$$

By definition of $\bar{v}$, one has $\underline{v}_{\mathcal{D}} - \bar{v}\underline{1}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}^{k}$ so we can apply the discrete Poincaré inequality (4.11) to infer that

$$\|v_{\mathcal{M}} - \bar{v}\|_{L^2(\Omega)}^2 \le C_P^2 |\underline{v}_{\mathcal{D}} - \bar{v}\underline{1}_{\mathcal{D}}|_{1,\mathcal{D}}^2 = C_P^2 |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}^2.$$

Combining the two previous estimates, we get

$$\mathbb{E}_{\omega}^{n+1} \le 2\omega_\sharp C_P^2 |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}^2.$$

Now, one can use the stability estimate (4.21) on $a_{\mathcal{D}}^{\omega}$ to deduce that

$$\mathbb{D}_{\omega}^{n+1} = a_{\mathcal{D}}^{\omega}(\underline{v}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \ge \alpha_\flat^{\omega} |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}^2.$$

Therefore, by the last two estimates, one has the following relation between the entropy and its dissipation:

$$\mathbb{E}_{\omega}^{n+1} \le \frac{2\omega_\sharp C_P^2}{\alpha_\flat^{\omega}} \mathbb{D}_{\omega}^{n+1}.$$

Plugging this estimate into the entropy relation (4.23), we deduce that

$$(1 + \nu_\omega \Delta t)\mathbb{E}_{\omega}^{n+1} \le \mathbb{E}_{\omega}^{n}, \text{ with } \nu_\omega = \frac{\alpha_\flat^{\omega}}{2\omega_\sharp C_P^2}.$$

This implies the exponential decay of the entropy:

$$\forall n \ge 0,\ \mathbb{E}_{\omega}^{n} \le (1 + \nu_\omega \Delta t)^{-n} \mathbb{E}_{\omega}^{0}.$$

To conclude, we just use the bounds on ω to infer that

$$\omega_\flat \|u_{\mathcal{M}}^{\omega,n} - u_{\mathcal{M}}^{\omega,\infty}\|_{L^2(\Omega)}^2 \leq 2\mathbb{E}_\omega^n \leq \omega_\sharp \|u_{\mathcal{M}}^{\omega,n} - u_{\mathcal{M}}^{\omega,\infty}\|_{L^2(\Omega)}^2$$

which yields (4.24) with $C_\omega = \sqrt{\dfrac{\omega_\sharp}{\omega_\flat}}$. □

Remark 23 (Regularity of the initial datum and topology of the convergence)**.** *Notice that in Proposition 14 we made the extra assumption that $u^{in} \in L^2(\Omega)$. Indeed, the long-time analysis of the exponential fitting scheme relies on the decay of the quadratic entropy (on the Slotboom variable)*

$$\mathbb{E}_\omega(t) = \frac{1}{2}\int_\Omega \omega|\rho(t) - \rho^\infty|^2.$$

Therefore, even at the continuous level, one needs to have a finite initial quadratic entropy to perform the analysis, and assuming that the initial datum is in $L^2(\Omega)$ is a safe choice. At the end, the exponential fitting scheme gives a convergence (in time) result in the L^2 topology, as a reminiscence of its linearity. On the other hand, the nonlinear scheme will yield convergence in a weaker norm (typically L^1), but can be used to deal with less regular initial data, which are in $L\log(L)$ only.

4.3.2 Nonlinear scheme

We present here some results regarding the analysis of the nonlinear scheme (4.20). Since we deal with a nonlinear scheme, unlike the exponential fitting scheme, the question of the existence of solutions is the main difficulty here. As often for this type of scheme, we start by establishing some a priori estimates, and then give an existence result.
For the purpose of analysis, given $\underline{\ell}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ a discrete logarithm potential, we associate a discrete quasi-Fermi potential defined as

$$\underline{w}_{\mathcal{D}} = \underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}} - \log(c_{nl}^M)\underline{1}_{\mathcal{D}}.$$

By definition of c_{nl}^M, one has $w_{\mathcal{M}} = \log\left(\frac{u_{\mathcal{M}}}{u_{\mathcal{M}}^\infty}\right)$. Note that, on the other hand, for any $(\underline{\ell}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D}}^k \times \underline{V}_{\mathcal{D}}^k$, we have

$$\mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}, \underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}),$$

since $\underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}$ and $\underline{w}_{\mathcal{D}}$ only differ from an additive constant element . As in Chapter 2, the discrete quasi-Fermi potentials are the key variables in order to perform the existence analysis of the scheme.
As a last remark before analysing the scheme, note that by stability of the interpolator (see [97, Proposition 2.2]), there exists $\phi_\sharp > 0$ depending only on ϕ, Ω, k and θ such that

$$|\underline{\phi}_{\mathcal{D}}|_{1,\mathcal{D}} \leq \phi_\sharp. \tag{4.25}$$

We now state our fundamental a priori results. As for the exponential fitting scheme, the discrete entropy structure of the nonlinear scheme is mainly a consequence of the convexity of the entropy.

Proposition 15 (Fundamental a priori relations)**.** *Let $(\underline{\ell}_{\mathcal{D}}^n)_{n\geq 1}$ be a solution to the nonlinear scheme (4.20), and $\left(\underset{\sim}{u}_{\mathcal{D}}^n\right)_{n\geq 1}$ be the associated reconstructed discrete density. Then, the following a priori results hold:*

(i) the mass is preserved along time:

$$\forall n \in \mathbb{N}^*, \int_\Omega u_\mathcal{M}^n = \int_\Omega u^{in} = M, \tag{4.26}$$

(ii) a discrete entropy/dissipation relation holds true:

$$\forall n \in \mathbb{N}, \frac{\mathbb{E}^{n+1} - \mathbb{E}^n}{\Delta t} \leq -\mathbb{D}^{n+1}, \tag{4.27}$$

where the discrete entropy and dissipation are non-negative quantities defined by

$$\mathbb{E}^n = \int_\Omega u_\mathcal{M}^\infty \Phi_1\left(\frac{u_\mathcal{M}^n}{u_\mathcal{M}^\infty}\right) \geq 0 \quad \textit{and} \quad \mathbb{D}^n = \mathcal{T}_\mathcal{D}(\underline{\ell}_\mathcal{D}^n, \underline{w}_\mathcal{D}^n, \underline{w}_\mathcal{D}^n) \geq 0 \textit{ for } n \geq 1$$

with $\Phi_1 : s \mapsto s\log(s) - s + 1$ *(and* $\Phi_1(0) = 1$*).*

Proof. Let $n \geq 0$. Using $\underline{1}_\mathcal{D}$ as a test function in (4.20a), we get that the mass is preserved:

$$\int_\Omega u_\mathcal{M}^{n+1} = \int_\Omega u_\mathcal{M}^n.$$

Therefore, by (4.20c), we get the identity (4.26).
To get the entropy relation, we use the convexity of Φ_1 and notice that

$$\mathbb{E}^{n+1} - \mathbb{E}^n \leq \int_\Omega u_\mathcal{M}^\infty \Phi_1'\left(\frac{u_\mathcal{M}^{n+1}}{u_\mathcal{M}^\infty}\right)\frac{u_\mathcal{M}^{n+1} - u_\mathcal{M}^n}{u_\mathcal{M}^\infty}.$$

Then, since $w_\mathcal{M}^{n+1} = \log\left(\frac{u_\mathcal{M}^{n+1}}{u_\mathcal{M}^\infty}\right)$ and $\Phi_1' = \log$, one has

$$\mathbb{E}^{n+1} - \mathbb{E}^n \leq \int_\Omega w_\mathcal{M}^{n+1}\left(u_\mathcal{M}^{n+1} - u_\mathcal{M}^n\right). \tag{4.28}$$

On the other hand, testing (4.20a) with $\underline{w}_\mathcal{D}^{n+1} \in \underline{V}_\mathcal{D}^k$, we get by definition of the discrete dissipation

$$\int_\Omega w_\mathcal{M}^{n+1}\left(u_\mathcal{M}^{n+1} - u_\mathcal{M}^n\right) = -\Delta t \mathcal{T}_\mathcal{D}(\underline{\ell}_\mathcal{D}^{n+1}, \underline{\ell}_\mathcal{D}^{n+1} + \underline{\phi}_\mathcal{D}, \underline{w}_\mathcal{D}^{n+1}) = -\Delta t \mathcal{T}_\mathcal{D}(\underline{\ell}_\mathcal{D}^{n+1}, \underline{w}_\mathcal{D}^{n+1}, \underline{w}_\mathcal{D}^{n+1}),$$

which finally yields (4.27). □

Note that the previous results hold for any $\varepsilon \geq 0$. Following the ideas of [56, 70], the entropy/dissipation relation should allow one to show the exponential convergence of the discrete solutions towards the discrete thermal equilibrium and to get convergence results (when $(h_\mathcal{D}, \Delta t) \to (0,0)$). These aspects will be the topics of future works.

In the remainder of this section, we focus on the existence of solutions and on the long-time behaviour, for $\varepsilon > 0$. The two results stated rely on a discrete a priori estimate, which is a high-order counterpart of Lemma 2.
In order to perform the analysis, we first introduce an inner product $\langle \cdot, \cdot \rangle$ on $\underline{V}_\mathcal{D}^k$ such that

$$\forall (\underline{w}_\mathcal{D}, \underline{v}_\mathcal{D}) \in \underline{V}_\mathcal{D}^k \times \underline{V}_\mathcal{D}^k, \ \langle \underline{w}_\mathcal{D}, \underline{v}_\mathcal{D} \rangle = \int_\Omega w_\mathcal{M} v_\mathcal{M} + \sum_{\sigma \in \mathcal{E}} \int_\sigma w_\sigma v_\sigma.$$

We denote by $\|\cdot\|$ the corresponding Euclidean norm:

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k,\ \|\underline{v}_{\mathcal{D}}\| = \sqrt{\sum_{K\in\mathcal{M}} \|v_K\|_{L^2(K)}^2 + \sum_{\sigma\in\mathcal{E}} \|v_\sigma\|_{L^2(\sigma)}^2}.$$

We now state the fundamental discrete a priori result. Note that the proof proposed here adapts easily to non-Boltzmann statistics, since we do not use the morphism properties of the exponential function (but only the fact that it is non-decreasing).

Lemma 9 (Discrete boundedness by mass and energy semi-norm)**.** *Let $\underline{\ell}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ and assume that there exist $C_\sharp > 0$ and $M_\sharp \geq M_\flat > 0$ such that*

$$M_\flat \leq \int_\Omega \mathrm{e}^{\ell_{\mathcal{M}}} \leq M_\sharp \quad \text{and} \quad |\underline{\ell}_{\mathcal{D}}|_{1,\mathcal{D}}^2 \leq C_\sharp. \tag{4.29}$$

Then, there exists a positive constant C depending only on $M_\flat$, $M_\sharp$, $C_\sharp$, k, Ω and $\mathcal{D}$ such that

$$\|\underline{\ell}_{\mathcal{D}}\| \leq C.$$

Proof. In order to prove this lemma, we need two preliminary results.

First, we recall that in a finite-dimensional vector space, all the norms are equivalent. Therefore, there exist some positive constants c_∞ and c_2 (depending only on $\mathcal{D}$ and k) such that

$$\forall K \in \mathcal{M}, \forall v \in \mathbb{P}^{k+1}(K), \begin{cases} \|v\|_{L^\infty(K)} & \leq c_\infty \|v\|_{L^2(K)}, \\ \|v\|_{L^2(K)} & \leq c_2 \|v\|_{L^\infty(K)}, \\ \|v\|_{L^2(\sigma)} & \leq c_2 \|v\|_{L^\infty(\sigma)}\ \forall \sigma \in \mathcal{E}_K, \\ \|v\|_{L^\infty(\sigma)} & \leq c_\infty \|v\|_{L^2(\sigma)}\ \forall \sigma \in \mathcal{E}_K, \\ \|\nabla v\|_{L^\infty(K)} & \leq c_\infty \|\nabla v\|_{L^2(K)}, \end{cases}$$

and

$$\forall \sigma \in \mathcal{E}, \forall v \in \mathbb{P}^k(\sigma), \begin{cases} \|v\|_{L^\infty(\sigma)} & \leq c_\infty \|v\|_{L^2(\sigma)}, \\ \|v\|_{L^2(\sigma)} & \leq c_2 \|v\|_{L^\infty(\sigma)}. \end{cases}$$

Note that these inequalities essentially correspond to reverse Lebesgue and discrete trace inequalities, for which the precise dependency of the constant with respect to the considered cell is known [97, Section 1.2].

Second, let $K \in \mathcal{M}$. By the mean value theorem (see [118, Remark 2.12]), one shows that there exists some positive constant c_K depending on K (essentially its size and shape) such that

$$|\inf_K \ell_K - \sup_K \ell_K| \leq c_K \|\nabla \ell_K\|_{L^\infty(K)} \leq c_K c_\infty \|\nabla \ell_K\|_{L^2(K)} \leq c_K c_\infty \sqrt{C_\sharp}.$$

Therefore, defining the maximal oscillation over the cells as

$$o_{\mathcal{D}} = c_\infty \sqrt{C_\sharp} \max_{K\in\mathcal{M}} c_K,$$

we get that

$$\forall K \in \mathcal{M},\ |\inf_K \ell_K - \sup_K \ell_K| \leq o_{\mathcal{D}}. \tag{4.30}$$

We now proceed in three steps. First, we get an L^∞-control over one of the cell unknowns. Then, we propagate this bound to the neighbouring unknowns. Last, we use the connectivity of

the mesh to get an estimate on $\|\underline{\ell}_{\mathcal{D}}\|$.

Since $M_\flat \le \int_\Omega \mathrm{e}^{\ell_{\mathcal{M}}} = \sum_{K\in\mathcal{M}} \int_K \mathrm{e}^{\ell_K} \le M_\sharp$, there exists a cell $K_0 \in \mathcal{M}$ such that

$$\frac{|K_0|M_\flat}{|\Omega|} \le \int_{K_0} \mathrm{e}^{\ell_{K_0}} \le M_\sharp.$$

Let $\ell_\flat = \inf_{K_0} \ell_{K_0}$ and $\ell_\sharp = \sup_{K_0} \ell_{K_0}$ be the extremal values of ℓ_{K_0} on K_0. According to (4.30), one has

$$\ell_\sharp - o_{\mathcal{D}} \le \ell_\flat.$$

Thus, integrating this inequality we deduce that

$$\int_{K_0} \mathrm{e}^{\ell_\sharp - o_{\mathcal{D}}} \le \int_{K_0} \mathrm{e}^{\ell_{K_0}} \le M_\sharp \quad \text{so} \quad \ell_\sharp \le \log\left(\frac{M_\sharp}{|K_0|}\right) + o_{\mathcal{D}}.$$

For the other bound, one has

$$\frac{|K_0|M_\flat}{|\Omega|} \le \int_{K_0} \mathrm{e}^{\ell_{K_0}} \le \int_{K_0} \mathrm{e}^{\ell_\flat + o_{\mathcal{D}}} \quad \text{so} \quad \ell_\flat \ge \log\left(\frac{M_\flat}{|\Omega|}\right) - o_{\mathcal{D}}.$$

Letting $C_{K_0} = \max\left(\log\left(\frac{M_\sharp}{|K_0|}\right) + o_{\mathcal{D}}, o_{\mathcal{D}} - \log\left(\frac{M_\flat}{|\Omega|}\right)\right)$, we get an uniform control of the cell unknown ℓ_{K_0}:

$$\|\ell_{K_0}\|_{L^\infty(K_0)} \le C_{K_0}.$$

Now, we can estimate the neighbouring face unknowns. Let $\sigma \in \mathcal{E}_{K_0}$ be a fixed face. The bound on $|\underline{\ell}_{\mathcal{D}}|^2_{1,\mathcal{D}}$ implies that $\|\ell_{K_0} - \ell_\sigma\|^2_{L^2(\sigma)} \le h_\sigma C_\sharp$. Since ℓ_{K_0} is polynomial, one has

$$\|\ell_\sigma\|_{L^2(\sigma)} \le \|\ell_{K_0}\|_{L^2(\sigma)} + \sqrt{h_\sigma C_\sharp} \le c_2 \|\ell_{K_0}\|_{L^\infty(\sigma)} + \sqrt{h_\sigma C_\sharp} \le c_2 C_{K_0} + \sqrt{h_\sigma C_\sharp}.$$

Therefore, letting $C_\sigma = c_\infty\left(c_2 C_{K_0} + \sqrt{h_{\mathcal{D}} C_\sharp}\right)$ we get an uniform control of the face unknown:

$$\|\ell_\sigma\|_{L^\infty(\sigma)} \le C_\sigma.$$

On the other hand, let $L \in \mathcal{M}$ be a cell such that $\sigma \in \mathcal{E}_L$. As previously, one has $\|\ell_L - \ell_\sigma\|^2_{L^2(\sigma)} \le h_\sigma C_\sharp \le h_{\mathcal{D}} C_\sharp$, therefore

$$\|\ell_L\|_{L^\infty(\sigma)} \le c_\infty \|\ell_L\|_{L^2(\sigma)} \le c_\infty \left(\|\ell_\sigma\|_{L^2(\sigma)} + \sqrt{h_{\mathcal{D}} C_\sharp}\right) \le c_\infty \left(c_2 \|\ell_\sigma\|_{L^\infty(\sigma)} + \sqrt{h_{\mathcal{D}} C_\sharp}\right).$$

Thus, assuming that $\|\ell_\sigma\|_{L^\infty(\sigma)}$ is controlled, one gets an uniform estimate on the trace of ℓ_L on σ. But, using the control of the oscillations (4.30), we deduce that

$$\|\ell_L\|_{L^\infty(L)} \le \|\ell_L\|_{L^\infty(\sigma)} + o_{\mathcal{D}}.$$

All in all, we can get an uniform estimate on any cell or face unknown. Because of the connectivity of the mesh, for any cell K (respectively, face σ) there is a finite sequence of components of $\underline{\ell}_{\mathcal{D}}$, denoted $(\ell_k)_{k=0,\dots,m}$, starting at $\ell_0 = \ell_{K_0}$ and finishing at $\ell_m = \ell_K$ (respectively, $\ell_m = \ell_\sigma$) so that ℓ_K (resp. ℓ_σ) can be connected to ℓ_{K_0}. Given i, we denote by C_i the L^∞-norm of ℓ_i (either on a face or on a cell, depending on the nature of ℓ_i). According to the previous

estimates, if i is even (and corresponds to a cell unknown), one has

$$C_{i+1} \le c_\infty \left(c_2 C_i + \sqrt{h_\mathcal{D} C_\sharp} \right),$$

and else, if i is odd (and corresponds to a face unknown), we have

$$C_{i+1} \le c_\infty \left(c_2 C_i + \sqrt{h_\mathcal{D} C_\sharp} \right) + o_\mathcal{D}.$$

At the end, one can deduce that for any $0 \le i \le m-1$,

$$C_{i+1} \le o_\mathcal{D} + c_\infty \sqrt{h_\mathcal{D} C_\sharp} + c_\infty c_2 C_i.$$

Using a comparison with an arithmetico-geometric sequence, we infer that

$$C_m \le (c_\infty c_2)^m |C_0 - \gamma| + |\gamma| \quad \text{with } \gamma = \frac{o_\mathcal{D} + c_\infty \sqrt{h_\mathcal{D} C_\sharp}}{1 - c_\infty c_2}. \tag{4.31}$$

Therefore, C_m depends only on C_{K_0}, $o_\mathcal{D}$, $h_\mathcal{D}$, c_2, c_∞, $C_\sharp$ and m. Iterating this process for each unknown, since there is a finite number of cells and faces (at most $|\mathcal{M}| + |\mathcal{E}|$), we deduce that there exists a positive constant C depending on C_{K_0}, $o_\mathcal{D}$, $h_\mathcal{D}$, c_2, c_∞, $C_\sharp$ and $|\mathcal{M}| + |\mathcal{E}|$ such that

$$\forall K \in \mathcal{M},\ \|\ell_K\|_{L^\infty(K)} \le C \quad \text{and } \forall \sigma \in \mathcal{E},\ \|\ell_\sigma\|_{L^\infty(\sigma)} \le C.$$

To conclude and get a bound on $\|\underline{\ell}_\mathcal{D}\|$, we just use the fact that the local (cell and face) L^2-norms are dominated by the local L^∞-norms. □

Remark 24 (Non-uniform bound)**.** *Note that the estimate obtained in Lemma 9 involves quantities which strongly depend on the mesh, such as the meshsize $h_\mathcal{D}$ or the number of cells and faces. Therefore, this estimation is not uniform with respect to the mesh.*

We now state an existence result, which holds only for positive ε. The proof strongly relies on the methodology developed in Chapter 1 in order to show the existence result of Theorem 1 for the nonlinear HFV scheme.

Theorem 10 (Existence of solutions to the nonlinear scheme (4.20))**.** *Assume that the stabilisation parameter ε in (4.18) is positive. Then, there exists at least one solution $(\underline{\ell}_\mathcal{D}^n)_{n\ge1}$ to the scheme (4.20). The associated densities $\left(\underset{\sim}{u}_\mathcal{D}^n\right)_{n\ge1}$ are positive.*

Proof. The proof proceeds by induction. Let $n \in \mathbb{N}$, and assume that $u_\mathcal{M}^n$ is well defined, following the definition (4.20b) (if $n \ge 1$) or (4.20c) (if $n = 0$). We now prove the existence of a solution $\underline{\ell}_\mathcal{D}^{n+1} \in \underline{V}_\mathcal{D}^k$ to (4.20a). In order to simplify the computations, instead of looking for the logarithmic potential, we will seek for the discrete quasi-Fermi potential $\underline{w}_\mathcal{D}^{n+1} = \underline{\ell}_\mathcal{D}^{n+1} + \underline{\phi}_\mathcal{D} - \log(c_{nl}^M)\underline{1}_\mathcal{D}$.

To do so, we first notice that, given any $\underline{w}_\mathcal{D} \in \underline{V}_\mathcal{D}^k$ (in the following, we denote by $\underline{\ell}_\mathcal{D} = \underline{w}_\mathcal{D} - \underline{\phi}_\mathcal{D} + \log(c_{nl}^M)\underline{1}_\mathcal{D}$ the discrete logarithm potential and by $\underset{\sim}{u}_\mathcal{D}$ the associated density), the application

$$\underline{v}_\mathcal{D} \mapsto \int_\Omega \frac{u_\mathcal{M} - u_\mathcal{M}^n}{\Delta t} v_\mathcal{M} + \mathcal{T}_\mathcal{D}(\underline{\ell}_\mathcal{D}, \underline{w}_\mathcal{D}, \underline{v}_\mathcal{D})$$

is a continuous linear form on $\underline{V}_\mathcal{D}^k$. Therefore, by Riesz representation theorem, there exists a

unique $\underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D}}^k$ such that

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k,\ \left\langle \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}), \underline{v}_{\mathcal{D}} \right\rangle = \int_\Omega \frac{u_{\mathcal{M}} - u_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}).$$

On the other hand, since $\underline{V}_{\mathcal{D}}^k$ is a finite-dimensional vector space, one can show that $\underline{w}_{\mathcal{D}} \mapsto \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}})$ is a continuous mapping on $\underline{V}_{\mathcal{D}}^k$. Notice that, for any discrete quasi-Fermi potential $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ satisfying $\underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}) = \underline{0}_{\mathcal{D}}$, the corresponding discrete logarithm potential $\underline{\ell}_{\mathcal{D}} = \underline{w}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}} + \log(c_{nl}^{\mathcal{M}})\underline{1}_{\mathcal{D}}$ is a solution to (4.20a). Thus, our goal from now on is to prove that $\underline{\mathcal{G}}_{\mathcal{D}}$ vanishes on $\underline{V}_{\mathcal{D}}^k$.

We introduce a regularisation of $\underline{\mathcal{G}}_{\mathcal{D}}$: given any $\mu > 0$, we define an application $\underline{\mathcal{G}}_{\mathcal{D}}^\mu$ as

$$\underline{\mathcal{G}}_{\mathcal{D}}^\mu : \underline{w}_{\mathcal{D}} \mapsto \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}) + \mu \underline{w}_{\mathcal{D}}.$$

By definition of $\underline{\mathcal{G}}_{\mathcal{D}}$, one has

$$\begin{aligned}\left\langle \underline{\mathcal{G}}_{\mathcal{D}}^\mu(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \right\rangle &= \left\langle \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \right\rangle + \mu \|\underline{w}_{\mathcal{D}}\|^2 \\ &= \int_\Omega \frac{u_{\mathcal{M}} - u_{\mathcal{M}}^n}{\Delta t} w_{\mathcal{M}} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}) + \mu \|\underline{w}_{\mathcal{D}}\|^2.\end{aligned}$$

As it was already shown in the proof of Proposition 15 (equation (4.28)), by convexity of Φ_1 one has

$$\int_\Omega \frac{u_{\mathcal{M}} - u_{\mathcal{M}}^n}{\Delta t} w_{\mathcal{M}} \geq \frac{\mathbb{E}(\underline{w}_{\mathcal{D}}) - \mathbb{E}^n}{\Delta t},$$

where the relative entropies are defined by

$$\mathbb{E}(\underline{w}_{\mathcal{D}}) = \int_\Omega u_{\mathcal{M}}^\infty \Phi_1\left(\frac{u_{\mathcal{M}}}{u_{\mathcal{M}}^\infty}\right) \quad \text{and} \quad \mathbb{E}^n = \int_\Omega u_{\mathcal{M}}^\infty \Phi_1\left(\frac{u_{\mathcal{M}}^n}{u_{\mathcal{M}}^\infty}\right).$$

As already mentioned, since Φ_1 is a non-negative function, these two quantities are non-negative. Moreover, since $u_{\mathcal{M}}^n$ is piecewise C^∞ if $n \geq 1$ and $\int_\Omega u_{\mathcal{M}}^0 \log(u_{\mathcal{M}}^0) = \int_\Omega u^{in} \log(u^{in}) < \infty$, one has $\mathbb{E}^n < \infty$. Note that it may occur that $\mathbb{E}^n = 0$ (which is equivalent to $u_{\mathcal{M}}^n = u_{\mathcal{M}}^\infty$), in which case $\underline{\ell}_{\mathcal{D}} = \underline{\ell}_{\mathcal{D}}^\infty$ is the unique solution to (4.20a) by mass preservation and entropy relation (4.27). In the following, we therefore assume that $\mathbb{E}^n > 0$. The previous identities and the non-negativity of the dissipation and entropy imply that

$$\begin{aligned}\left\langle \underline{\mathcal{G}}_{\mathcal{D}}^\mu(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \right\rangle &\geq \frac{\mathbb{E}(\underline{w}_{\mathcal{D}}) - \mathbb{E}^n}{\Delta t} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}) + \mu \|\underline{w}_{\mathcal{D}}\|^2 \\ &\geq \mu \|\underline{w}_{\mathcal{D}}\|^2 - \frac{\mathbb{E}^n}{\Delta t}.\end{aligned} \tag{4.32}$$

Thus, for any $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ such that $\|\underline{w}_{\mathcal{D}}\| \geq \sqrt{\frac{\mathbb{E}^n}{\mu \Delta t}}$, one has $\left\langle \underline{\mathcal{G}}_{\mathcal{D}}^\mu(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \right\rangle \geq 0$. Therefore, according to Lemma 1 (see also [119, Section 9.1]), there exists a least one $\underline{w}_{\mathcal{D}}^\mu \in \underline{V}_{\mathcal{D}}^k$ such that

$$\underline{\mathcal{G}}_{\mathcal{D}}^\mu(\underline{w}_{\mathcal{D}}^\mu) = \underline{0}_{\mathcal{D}} \quad \text{and} \quad \|\underline{w}_{\mathcal{D}}^\mu\| \leq \sqrt{\frac{\mathbb{E}^n}{\mu \Delta t}}. \tag{4.33}$$

Now, using (4.32) with $\underline{w}_{\mathcal{D}}^{\mu}$, we get

$$\frac{\mathbb{E}(\underline{w}_{\mathcal{D}}^{\mu})}{\Delta t} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{\mu}, \underline{w}_{\mathcal{D}}^{\mu}, \underline{w}_{\mathcal{D}}^{\mu}) + \mu \|\underline{w}_{\mathcal{D}}^{\mu}\|^2 \leq \frac{\mathbb{E}^n}{\Delta t},$$

so $\mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{\mu}, \underline{w}_{\mathcal{D}}^{\mu}, \underline{w}_{\mathcal{D}}^{\mu}) \leq \frac{\mathbb{E}^n}{\Delta t}$. Therefore, recalling the definition (4.18)-(4.19) of $\mathcal{T}_{\mathcal{D}}$, alongside with the fact that $\varepsilon > 0$ and the global stability result (4.10) for $a_{\mathcal{D}}$, we then infer that

$$\varepsilon \alpha_{\flat} \lambda_{\flat} h_{\flat}^{k+2} |\underline{w}_{\mathcal{D}}^{\mu}|_{1,\mathcal{D}}^2 \leq \frac{\mathbb{E}^n}{\Delta t} \quad \text{with } h_{\flat} = \min_{K \in \mathcal{M}} h_K.$$

On the one hand, by definition of the discrete logarithm potential $\underline{\ell}_{\mathcal{D}}^{\mu}$ associated to $\underline{w}_{\mathcal{D}}^{\mu}$ and the estimate (4.25) on $|\underline{\phi}_{\mathcal{D}}|_{1,\mathcal{D}}$, it holds

$$|\underline{\ell}_{\mathcal{D}}^{\mu}|_{1,\mathcal{D}} = |\underline{w}_{\mathcal{D}}^{\mu} - \underline{\phi}_{\mathcal{D}} + \log(c_{nl}^{M})\underline{1}_{\mathcal{D}}|_{1,\mathcal{D}} \leq \sqrt{\frac{\mathbb{E}^n}{\varepsilon \alpha_{\flat} \lambda_{\flat} h_{\flat}^{k+2} \Delta t}} + \phi_{\sharp}. \tag{4.34}$$

On the other hand, by definition of $\underline{\mathcal{G}}_{\mathcal{D}}^{\mu}$, one first infers

$$\begin{aligned} 0 = \left\langle \underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(\underline{w}_{\mathcal{D}}^{\mu}), \underline{1}_{\mathcal{D}} \right\rangle &= \int_{\Omega} \frac{u_{\mathcal{M}}^{\mu} - u_{\mathcal{M}}^{n}}{\Delta t} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{\mu}, \underline{w}_{\mathcal{D}}^{\mu}, \underline{1}_{\mathcal{D}}) + \mu \left\langle \underline{w}_{\mathcal{D}}^{\mu}, \underline{1}_{\mathcal{D}} \right\rangle \\ &= \int_{\Omega} \frac{u_{\mathcal{M}}^{\mu} - u_{\mathcal{M}}^{n}}{\Delta t} + \mu \left\langle \underline{w}_{\mathcal{D}}^{\mu}, \underline{1}_{\mathcal{D}} \right\rangle. \end{aligned}$$

Second, recalling the estimate (4.33) on $\underline{w}_{\mathcal{D}}^{\mu}$ and using the Cauchy–Schwarz inequality, one gets

$$\left| \int_{\Omega} \left(u_{\mathcal{M}}^{\mu} - u_{\mathcal{M}}^{n} \right) \right| \leq \mu \Delta t \|\underline{w}_{\mathcal{D}}^{\mu}\| \|\underline{1}_{\mathcal{D}}\| \leq \sqrt{\mu} \sqrt{\Delta t \mathbb{E}^n \|\underline{1}_{\mathcal{D}}\|}.$$

Thus, letting $M^n = \int_{\Omega} u_{\mathcal{M}}^{n} > 0$ and $\mu_n = \frac{(M^n)^2}{4\Delta t \mathbb{E}^n \|\underline{1}_{\mathcal{D}}\|} > 0$, for all $0 < \mu < \mu_n$ one has

$$\frac{M^n}{2} \leq \int_{\Omega} \mathrm{e}^{\ell_{\mathcal{M}}^{\mu}} \leq \frac{3M^n}{2}. \tag{4.35}$$

Therefore, by estimates (4.34) and (4.35), one can apply Lemma 9 to $\underline{\ell}_{\mathcal{D}}^{\mu}$ with $M_{\flat} = \frac{M^n}{2}$, $M_{\sharp} = \frac{3M^n}{2}$ and $C_{\sharp} = \left(\sqrt{\frac{\mathbb{E}^n}{\varepsilon \alpha_{\flat} \lambda_{\flat} h_{\flat}^{k+2} \Delta t}} + \phi_{\sharp} \right)^2$ (note that these three constants do not depend on μ): there exists a constant $C > 0$ which does not depend on μ such that

$$\forall \mu \in]0, \mu_n[\,, \quad \|\underline{\ell}_{\mathcal{D}}^{\mu}\| \leq C.$$

Then, by compactness, there exists $\underline{\ell}_{\mathcal{D}}^{n+1} \in \underline{V}_{\mathcal{D}}^{k}$ such that, up to extraction, $\underline{\ell}_{\mathcal{D}}^{\mu} \to \underline{\ell}_{\mathcal{D}}^{n+1}$ when $\mu \to 0$. On the other hand, $\underline{\mathcal{G}}_{\mathcal{D}}^{\mu}$ tends to $\underline{\mathcal{G}}_{\mathcal{D}}$ as μ tends to 0. Therefore, letting $\underline{w}_{\mathcal{D}}^{n+1} = \underline{\ell}_{\mathcal{D}}^{n+1} + \underline{\phi}_{\mathcal{D}} - \log(c_{nl}^{M})\underline{1}_{\mathcal{D}}$, we have $\underline{0}_{\mathcal{D}} = \underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(\underline{w}_{\mathcal{D}}^{\mu}) \to \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}^{n+1})$ as $\mu \to 0$, which implies that

$$\underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}^{n+1}) = \underline{0}_{\mathcal{D}}.$$

It follows that $\underline{\ell}_{\mathcal{D}}^{n+1}$ is a solution to (4.20a). □

Remark 25 (Uniqueness of the solution). *As for the nonlinear HFV scheme of Chapter 1, the existence result only asserts that solutions exist. However, the question of uniqueness is still open currently. A possible approach to show such a result could be to consider the relative entropy of a solution with respect to another solution, and show that this quantity vanishes.*

Last, we state a result which indicates that the nonlinear scheme has a good long-time behaviour, in the sense that the solution converges in time towards the discrete thermal equilibrium. Note that this result is relatively weaker compared to the one of Proposition 14, since it does not give any indication about the speed of convergence.

Proposition 16 (Long-time behaviour of the nonlinear scheme). *Assume that the stabilisation parameter ε in (4.18) is positive, and let $(\underline{\ell}_{\mathcal{D}}^n)_{n\geq 1}$ be a solution to the scheme (4.20). Then, the solution converges towards the discrete logarithm potential at equilibrium as time tends to $+\infty$:*

$$\underline{\ell}_{\mathcal{D}}^n \xrightarrow[n\to\infty]{} \underline{\ell}_{\mathcal{D}}^\infty \text{ in } \underline{V}_{\mathcal{D}}^k. \tag{4.36}$$

Consequently, the corresponding discrete density $(u_{\mathcal{M}}^n)_{n\geq 1}$ converges to $u_{\mathcal{M}}^\infty$ in $L^\infty(\Omega)$.

Proof. First, note that by the entropy relation (4.27), one has

$$\sum_{n\geq 1} \mathbb{D}^n \leq \sum_{n\geq 1} \frac{\mathbb{E}^{n-1} - \mathbb{E}^n}{\Delta t} \leq \frac{\mathbb{E}^0}{\Delta t}.$$

Thus, according to the definition of the discrete dissipation $\mathbb{D}^n$, alongside with the definition (4.19) of $\mathcal{T}_{\mathcal{D}}$, the fact that $\varepsilon > 0$ and the global stability result (4.10) on $a_{\mathcal{D}}$, we infer that

$$\sum_{n\geq 1} |\underline{w}_{\mathcal{D}}^n|_{1,\mathcal{D}}^2 \leq \frac{\mathbb{E}^0}{\varepsilon h_{\flat}^{k+2} \alpha_{\flat} \lambda_{\flat} \Delta t},$$

where $h_{\flat} = \min_{K\in\mathcal{M}} h_K > 0$. This implies in particular that

$$\forall n \geq 1,\ |\underline{w}_{\mathcal{D}}^n|_{1,\mathcal{D}}^2 \leq \frac{\mathbb{E}^0}{\varepsilon h_{\flat}^{k+2} \alpha_{\flat} \lambda_{\flat} \Delta t} \quad \text{and} \quad |\underline{w}_{\mathcal{D}}^n|_{1,\mathcal{D}} \xrightarrow[n\to\infty]{} 0. \tag{4.37}$$

Let $n \geq 1$. By definition of $\underline{\ell}_{\mathcal{D}}^n$ and the bound on $\underline{\phi}_{\mathcal{D}}$, one has $|\underline{\ell}_{\mathcal{D}}^n|_{1,\mathcal{D}} \leq \phi_\sharp + \sqrt{\frac{\mathbb{E}^0}{\varepsilon h_{\flat}^{k+2} \alpha_{\flat} \lambda_{\flat} \Delta t}}$. On the other hand, by the mass preservation identity (4.26), we have $\int_\Omega \mathrm{e}^{\ell_{\mathcal{M}}^n} = M > 0$. Therefore, one can apply Lemma 9 to $\underline{\ell}_{\mathcal{D}}^n$: there exists a positive constant C (which depends on ε, $\mathcal{D}$, k, Δt and M, but not on n) such that

$$\forall n \geq 1,\ \|\underline{\ell}_{\mathcal{D}}^n\| \leq C.$$

Therefore, by compactness, there exists $\underline{\ell}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ such that, up to extraction,

$$\lim_{n\to\infty} \underline{\ell}_{\mathcal{D}}^n = \underline{\ell}_{\mathcal{D}} \text{ in } \underline{V}_{\mathcal{D}}^k.$$

On the other hand, by continuity of $|\cdot|_{1,\mathcal{D}}$ on $\underline{V}_{\mathcal{D}}^k$, by definition of $\underline{w}_{\mathcal{D}}^n$ and (4.37), we have

$$|\underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}|_{1,\mathcal{D}} = 0.$$

This means that there exists $a \in \mathbb{R}$ such that $\underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}} = a\underline{1}_{\mathcal{D}}$. Moreover, thanks to mass preservation, it holds

$$M = \int_\Omega \mathrm{e}^{\ell^n_{\mathcal{M}}} \xrightarrow[n\to\infty]{} \int_\Omega \mathrm{e}^{\ell_{\mathcal{M}}} \quad \text{so that} \quad \int_\Omega \mathrm{e}^{\ell_{\mathcal{M}}} = M.$$

Hence, one has $a = \log(c^M_{nl})$, which implies that $\underline{\ell}_{\mathcal{D}} = \log(c^M_{nl})\underline{1}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}} = \underline{\ell}^\infty_{\mathcal{D}}$. By uniqueness of the limit, we get the convergence of the whole sequence $(\underline{\ell}^n_{\mathcal{D}})_{n\geq 1}$ towards $\underline{\ell}^\infty_{\mathcal{D}}$. This implies in particular that $\ell^n_{\mathcal{M}} \xrightarrow[n\to\infty]{} \ell^\infty_{\mathcal{M}}$ in $L^\infty(\Omega)$ by norm equivalence in finite-dimensional vector spaces. Then, by the mean value theorem we deduce that

$$\|u^n_{\mathcal{M}} - u^\infty_{\mathcal{M}}\|_{L^\infty(\Omega)} = \|\mathrm{e}^{\ell^n_{\mathcal{M}}} - \mathrm{e}^{\ell^\infty_{\mathcal{M}}}\|_{L^\infty(\Omega)} \leq \mathrm{e}^{\max(\|\ell^n_{\mathcal{M}}\|_{L^\infty(\Omega)}, \|\ell^\infty_{\mathcal{M}}\|_{L^\infty(\Omega)})} \|\ell^n_{\mathcal{M}} - \ell^\infty_{\mathcal{M}}\|_{L^\infty(\Omega)},$$

which implies the convergence of the densities. □

Remark 26 (Estimation of the rate of convergence)**.** *The previous result states a convergence of the potential towards the corresponding equilibrium potential. It follows that $u^n_{\mathcal{M}} \xrightarrow[n\to\infty]{} u^\infty_{\mathcal{M}}$ in $L^\infty(\Omega)$. This result is relatively weaker than other long-time behaviour results presented in this manuscript. Indeed, there is no indication on the speed of convergence, even on a fixed mesh. However, note that the numerical results of Section 4.4.4 indicate that the convergence is as expected exponentially fast, and that the rate of convergence is similar to the continuous one. In order to state exponential convergence towards the equilibrium, one has to establish that the discrete dissipation controls the discrete entropy. Such an estimate (with a dependence on the meshsize) can be obtained for Dirichlet boundary conditions, using the same trick (comparison of the Boltzmann entropy with the H^1-norm of the quasi-Fermi potential, see (2.58)) as in Chapter 2. In order to get a uniform rate of convergence (with respect to the meshsize), one has to use further a discrete Logarithmic-Sobolev inequality. This strategy, already advocated for the low-order schemes of Chapter 1, is the subject of an ongoing work.*

4.4 Numerical results

In this section, we assert the properties of the schemes on some test-cases. We also aim at investigating the relevance of using (nonlinear) high-order schemes in terms of computational efficiency (accuracy for a given computational cost). The tests considered below are all set in the domain $\Omega =]0,1[^2$, and are the same tests as in Chapter 1, to which we refer for more detailed descriptions.

Given a (face) degree k, the exponential fitting scheme (4.14) will be referred to as expf_k, the nonlinear scheme (4.20) as nlhho_k, whereas the HFV scheme of Chapter 1 will be denoted nlhfv. Note that nlhho_0 hinges on affine cell unknowns, whereas the cell unknowns for nlhfv are constant: these two schemes hence do not coincide (see Remark 22), and nlhho_0 is expected to be more costly.

In the sequel, we use the following stabilisation parameters: $\eta_{nl} = \eta_l = 1$. Regarding the parameter ε we will use the value $\varepsilon = 1$. However, in some situations we will compare the two values $\varepsilon = 1$ and $\varepsilon = 0$. The nonlinear scheme with $\varepsilon = 0$ will then be denoted nlhho_k_0 for better readability.

The computations presented in this chapter were run on a laptop equipped with an Intel Core i7-9850H processor clocked at 2.60GHz and 32Gb of RAM.

4.4.1 Implementation

Here, we discuss some implementation aspects of the schemes. For a detailed introduction to the implementation of mixed-order HHO methods for standard linear diffusion problems (with a different consistent part), we refer the reader to [83, Chapter 8]. A different presentation (for equal-order methods) can be found in [97, Appendix B].

General setting

The HHO implementation here makes use of monomial basis functions for both the cell and face unknowns. Note that such a choice is known to introduce some numerical instabilities for large values of k. In the sequel, we will thus restrict the schemes to $k \leq 3$ in order to avoid as most as possible these instabilities. Keep in mind that the use of orthonormal basis functions could improve the results presented below (and especially the convergence of the Newton method used for the nonlinear scheme), and should be one of the next future improvements. The quadrature formulas used in our implementation are based on the Dunavant rules [115], which implies to subdivide the cells into triangles. For the tests below, we use mesh families with cells star-shaped with respect to their barycenters, therefore the subdivisions correspond to the pyramidal submesh introduced in the HFV context (see Chapter 1). To deal with integrals of non-polynomial functions (for both schemes), we use quadrature formulas of order $2k+5$. We performed a few tests (not shown here) with higher-order formulas and did not observe significant changes. Last but not least, the local computations are performed sequentially. One could expect a significant gain in terms of computational time parallelising these latter.

Exponential fitting scheme

For the linear exponential fitting scheme, the implementation follows the classical HHO strategy for linear diffusion problems. We directly solve for the discrete Slotboom variables $(\underline{\rho}_{\mathcal{D}}^n)_{n\geq 1}$. As standard for skeletal methods, we do not solve the full linear system but perform static condensation, which allows one to locally eliminate the cell unknowns (see [97, Appendix B.3.2]). Since the scheme relies on the same LHS matrix at each time step, we perform once and for all an LU decomposition of the matrix at the beginning of the computation. At each time step, the resolution is then inexpensive (the RHS has to be updated, but only through a matrix-vector product).

The implementation detailed here is relatively naive. In particular, we do not address the main questions which were highlighted in Chapter 1 about the (harmonic) averaging of ω (which is linked to the question of the quadrature formulas for the high-order scheme) and the preconditioning of the system (which was equivalent to choose to solve the system in the density variable u for the HFV scheme). These questions should be addressed in future works. However, in view of the results obtained in Chapter 1 for the HFV exponential fitting scheme, we believe these potential improvements will have no effect on the lack of positivity of this scheme.

Nonlinear scheme

The numerical scheme (4.20) requires to solve a nonlinear system of equations at each time step. For a given $n \in \mathbb{N}$, one wants to find $\underline{\ell}_{\mathcal{D}}^{n+1} \in \underline{V}_{\mathcal{D}}^k$ solution to the nonlinear scheme (4.20): this scheme can be written as the equation

$$\underline{\mathcal{G}}_{\mathcal{D}}^{n,\Delta t}(\underline{\ell}_{\mathcal{D}}^{n+1}) = \underline{0}_{\mathcal{D}},$$

where $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\Delta t} : \underline{V}_{\mathcal{D}}^k \to \underline{V}_{\mathcal{D}}^k$ is a regular (nonlinear) vector field. To find a zero of $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\Delta t}$, we use a Newton method.

In practice, the use of a naive method without any adaptation proved not to be enough to compute a solution in general. In order to get a robust implementation, which can handle various data and meshes, we use a few techniques. For further use, we denote by $\|\underline{\ell}_{\mathcal{D}}\|_{l^\infty}$ the l^∞-norm of the coefficients of $\underline{\ell}_{\mathcal{D}}$ in the polynomial bases (cells and faces) used to implement the scheme. This $\|\cdot\|_{l^\infty}$ is a norm on $\underline{V}_{\mathcal{D}}^k$ which is easily (and at very low cost) computable in practice. To fix the ideas and notation, the Newton method is defined as follows: given an initialisation $\underline{\ell}_{\mathcal{D},0} \in \underline{V}_{\mathcal{D}}^k$ and a time step $0 < \delta t \leq \Delta t$, one defines a sequence $\underline{\ell}_{\mathcal{D},i}$ of elements of $\underline{V}_{\mathcal{D}}^k$ such that

$$J_{\underline{\ell}_{\mathcal{D},i}}^{n,\delta t}(\underline{\ell}_{\mathcal{D},i+1} - \underline{\ell}_{\mathcal{D},i}) = -\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}\left(\underline{\ell}_{\mathcal{D},i}\right), \tag{4.38}$$

where $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}$ is the vector field associated to the nonlinear scheme (4.20) with time step δt instead of Δt, and $J_{\underline{\ell}_{\mathcal{D},i}}^{n,\delta t}$ is the differential (Jacobian in practice in the implementation) of $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}$ at the point $\underline{\ell}_{\mathcal{D},i}$. Note that in practice, we do not solve this linear system, but perform static condensation (see [97, Appendix B.3.2]) in order to eliminate (locally) the cell unknowns. The resulting linear system is called "condensed system" in what follows. We discuss below the main tricks deployed to reach robustness in the implementation of the Newton method.

Stopping criterion. In order to have information about the convergence of the Newton method, we consider two quantities: the relative norm of the residual r_{i+1} and the norm of the objective function g_i defined as

$$r_{i+1} = \frac{\|\underline{\ell}_{\mathcal{D},i+1} - \underline{\ell}_{\mathcal{D},i}\|_{l^\infty}}{\|\underline{\ell}_{\mathcal{D},i}\|_{l^\infty}} \quad \text{and} \quad g_i = \|\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}(\underline{\ell}_{\mathcal{D},i})\|_{l^\infty}.$$

We consider that the Newton method has converged when either

$$(r_{i+1} \leq 0.1 \times tol) \text{ or } (r_{i+1} \leq tol \text{ and } g_i \leq tol)$$

with $tol = 5.10^{-10}$, in which case we set $\underline{\ell}_{\mathcal{D}}^{n+1} = \underline{\ell}_{\mathcal{D},i+1}$. On the other hand, if this criterion is not met at $i = 50$, the method is considered as non-convergent (and we then proceed to a time step reduction, see below). In practice for the test of this chapter, we never reached $i = 50$, either because the method converged or because of a loop break, as described below.

Loop break for large $\underline{\ell}_{\mathcal{D}}$. The computation of $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}\left(\underline{\ell}_{\mathcal{D},i}\right)$ and $J_{\underline{\ell}_{\mathcal{D},i}}^{n,\delta t}$ implies punctual evaluations of $\mathrm{e}^{\ell_{K,i}}$ and $\mathrm{e}^{\ell_{\sigma,i}}$ (for K a cell and σ a face) in the quadrature formulas. Such computations can lead to severe numerical issues if the values at the quadrature nodes are too large. Therefore, we declare that $\underline{\ell}_{\mathcal{D},i}$ is too large for the computations if there exists a cell quadrature node $x_{K,q} \in \overline{K}$ or a face quadrature node $x_{\sigma,q} \in \overline{\sigma}$ such that

$$|\ell_{K,i}(x_{K,q})| \geq 100 \quad \text{or} \quad |\ell_{\sigma,i}(x_{\sigma,q})| \geq 100.$$

In such a case, the method is immediately considered as non-convergent, and we proceed with a time step reduction (see below). Note that the choice of the value 100 allows one to compute densities $\underline{u}_{\mathcal{D}}$ on a range from 10^{-43} to 10^{43}, and hence should not be a significant restriction in practice. In the numerical simulations presented in Sections 4.4.2-4.4.4, the use of this loop-breaking procedure is absolutely necessary in order to avoid some “explosion” of the method which leads to evaluations of too large quantities and crash of the code. Moreover, we also add a loop break if the linear solver does not

perform a successful resolution of the condensed linear system associated to (4.38), which corresponds to situations where either $J^{n,\delta t}_{\underline{\ell}_{\mathcal{D},i}}$ or its condensed counterpart are not invertible. Such situations occur in practice, essentially on very coarse meshes.

Adaptive time stepping. The previous strategies can lead to resolution failures for a given time step δt. If the Newton method does not converge, we try to compute the solution for a smaller time step $0.5 \times \delta t$. On the other hand, if the method converges we use for the subsequent time step the value $2 \times \delta t$. The maximal time step allowed is the initial time step, denoted by Δt. In practice, the scheme may perform numerous time step reductions at the beginning (early times) of the computation.

Initialisation by truncation and filtration. As for any Newton method, the question of the initialisation is fundamental in order to get a robust implementation. It appears that, for $n \geq 1$, the natural initialisation $\underline{\ell}_{\mathcal{D},0} = \underline{\ell}^n_{\mathcal{D}}$ is satisfactory when used with the adaptative time stepping strategy. However, for $n = 0$, such a choice is not possible, since $\underline{\ell}^0_{\mathcal{D}}$ does not exist in general if u^{in} vanishes locally or is too small (see Remark 21). A first way of tackling this problem (in the spirit of the techniques already used for the low-order schemes of the previous Chapters) is to define a truncated initial logarithm potential $\tilde{\ell}^{in}$ as

$$\tilde{\ell}^{in} = \log\Big(\max(u^{in}, 10^{-8})\Big),$$

and to initialise the Newton method with a polynomial interpolation $\underline{\tilde{\ell}}^0_{\mathcal{D}} \in \underline{V}^k_{\mathcal{D}}$ of $\tilde{\ell}^{in}$. In fact, such a strategy induces another issue: $\underline{\tilde{\ell}}^0_{\mathcal{D}}$ exhibits strong oscillations in the regions where the truncation is performed (this also holds true if u^{in} is not continuous on Ω near its discontinuities, as in Section 4.4.2). These oscillations usually make the method diverge, even with extremely small time step. Therefore, we eventually initialise the method with a "filtered" discrete unknown $\underline{\ell}^0_{\mathcal{D}} \in \underline{V}^k_{\mathcal{D}}$ with no oscillation, which corresponds to a projection of $\underline{\tilde{\ell}}^0_{\mathcal{D}}$ to piecewise constants (on cells and faces):

$$\forall K \in \mathcal{M},\ \ell^0_K = \Pi^0_K(\tilde{\ell}^0_K) = \frac{1}{|K|}\int_K \tilde{\ell}^0_K \text{ and } \forall \sigma \in \mathcal{E},\ \ell^0_\sigma = \Pi^0_\sigma(\tilde{\ell}^0_\sigma) = \frac{1}{|\sigma|}\int_\sigma \tilde{\ell}^0_\sigma.$$

In practice, using $\underline{\ell}_{\mathcal{D},0} = \underline{\ell}^0_{\mathcal{D}}$ as the first initialisation (when $n = 0$) yields convergent Newton methods for all tests presented below. In particular, the use of filtered initial discrete data seems crucial for high-order schemes ($k \geq 1$). For the lowest order version of the scheme ($k = 0$), the use of $\underline{\tilde{\ell}}^0_{\mathcal{D}}$ as the first initialisation often yields convergent Newton methods (up to time step reduction). Note that in order to define the initialisation on the faces, one has to assume more regularity on u^{in} than the initial hypothesis.

Of course, the chosen values for the stopping criterion and the thresholds are arbitrary and could be modified. However, the set of values advocated here makes the scheme robust enough so as to compute solutions for all the test-cases in this chapter.

Remark 27 (Potentials vs. densities)**.** *One of the main differences between the present scheme and the low-order ones from Chapter (1) lies in the fact that we use the potential ℓ as our unknown, whereas the density u was used in the low-order schemes. As a consequence, our stopping criterion only provides information on ℓ, while we are interested in the reconstructed density. Moreover, we only take into account the coefficients of the polynomials (through the norm $\|\cdot\|_{l^\infty}$), but such a norm does not give much information about the effective behaviour of the unknown. A more relevant stopping criterion*

could be to consider the residual in terms of densities

$$\frac{\|u_{\mathcal{M},i+1} - u_{\mathcal{M},i}\|_{L^2(\Omega)}}{\|u_{\mathcal{M},i}\|_{L^2(\Omega)}}$$

(and analogous definition for the face unknowns) in order to ensure a satisfying accuracy on u. However, the main issue of such a criterion is its evaluation cost. Thus, in this work we choose to use a purely algebraic stopping criterion on ℓ, whose cost is marginal. The testing of other stopping criteria will be the subject of future investigations.

Last, recall that for the HFV schemes, we used a loop-breaking strategy: when the computed density had non-positive values, we performed a time step reduction. Here, such a situation cannot occur since ℓ takes values in $\mathbb{R}$, but this apparent absence of negativity on the density is in fact pernicious. Indeed, situations where $\underline{\ell}_{\mathcal{D}}$ takes very negative values are the counterpart of a negative u for the low-order schemes, and lead to divergent Newton methods. The main difficulty here lies in the design of a relevant criterion in order to avoid this divergence.

4.4.2 Positivity

This first section is dedicated to assessing discrete positivity preservation. For the test considered here, we set the advection field and the diffusion tensor to

$$\phi(x,y) = -\Big((x-0.4)^2 + (y-0.6)^2\Big) \quad \text{and} \quad \Lambda = \begin{pmatrix} 0.8 & 0 \\ 0 & 1 \end{pmatrix}.$$

For the initial datum, we take

$$u^{in} = 10^{-3}\,\mathbb{1}_B + \mathbb{1}_{\Omega\setminus B},$$

where B is the Euclidean ball

$$\Big\{(x,y) \in \mathbb{R}^2 \mid (x-0.5)^2 + (y-0.5)^2 \leq 0.2^2\Big\}.$$

These data ensure that the solution u is positive on $\mathbb{R}_+ \times \Omega$. We perform the simulation on the time interval $[0, 5.10^{-4}]$ with $\Delta t = 10^{-5}$ on a refined tilted hexagonal-dominant mesh (4192 cells and 12512 edges). The computed discrete densities are denoted by $(\underset{\sim}{u}^n_{\mathcal{D}})_{1\leq n\leq N_f}$ and $(\underline{\underline{u}}^{\omega,n}_{\mathcal{D}})_{1\leq n\leq 50}$. Note that the situation $N_f > 50$ may occur if the nonlinear scheme has to perform time step adaptation. In this section, we also compare our results with the linear HMM scheme of [19] which was studied in Chapter 1.

In Table 4.1, we collect the minimal values reached by the discrete solutions. The values of "mincells" are defined by

$$\min\left\{\frac{1}{|K|}\int_K u^n_{\mathcal{M}} \mid K \in \mathcal{M}, 1 \leq n \leq N_f\right\} \quad \text{and} \quad \min\left\{\frac{1}{|K|}\int_K u^{\omega,n}_{\mathcal{M}} \mid K \in \mathcal{M}, 1 \leq n \leq 50\right\},$$

for, respectively, the nonlinear scheme and the exponential fitting scheme. The values of "mincellQN" are the minimal values taken by the densities at the cell quadrature nodes. Analogous definitions hold for the faces. The values of "#resol" correspond to the number of linear systems solved during the computation. Note that the size of these systems depends on the value of k, so it is not a relevant information to compare the cost of the schemes for different values of k. Last, the "computing time" is the total time needed to compute the discrete solution (it includes the precomputational steps, such as the computation of the matrices representing G^k_K). Recall

scheme	computing time (in s)	#resol	mincells	minfaces	mincellQN	minfaceQN
nlhfv	1.77e+01	175	9.93e-04	7.36e-04	9.93e-04	7.36e-04
nlhho_0	7.17e+01	224	1.00e-03	1.01-03	2.41e-06	1.01e-03
nlhho_1	4.13e+02	248	6.65e-04	2.05e-05	1.78e-04	3.57e-08
nlhho_2	1.45e+03	251	9.50e-04	5.99e-04	2.67e-07	1.06e-05
nlhho_3	3.87e+03	254	9.85e-04	8.58e-04	1.10e-05	1.79e-05
HMM	2.20e-01	50	-5.00e-03	-7.74e-02	-5.00e-03	-7.74e-02
expf_0	5.66e-01	50	1.02e-03	1.89e-03	-3.78e-01	1.89e-03
expf_1	2.23e+00	50	-1.29e-02	-2.40e-01	-4.91e-01	-3.71e-01
expf_2	6.34e+00	50	-6.14e-03	-1.02e-01	-5.08e-01	-5.35e-01
expf_3	1.53e+01	50	-3.24e-04	-1.02e-02	-5.52e-01	-4.05e-01

Table 4.1 – **Positivity of discrete solutions.**

that the HMM and the exponential fitting schemes are linear (with corresponding matrices not depending on time), whence their extremely low cost compared to the nonlinear schemes. However, when an LU decomposition is unaffordable and an iterative solver has to be used instead, nlhho_k is "only" five times more costly than expf_k.

The results of Table 4.1 first indicate that, as expected, all nonlinear schemes preserve the positivity of the discrete solutions. On the other hand, none of the linear schemes preserves positivity on the whole domain Ω. In fact, except for expf_0, all linear schemes studied here do not even preserve the average positivity on each cell, in the sense that there exists $K_0 \in \mathcal{M}$ and n_0 an integer such that

$$\int_{K_0} u_{\mathcal{M}}^{\omega,n_0} < 0 \quad (\text{ or } u_{K_0}^{n_0} < 0 \text{ for the HMM scheme}).$$

Moreover, it is interesting to note that the positivity violation peak (which can be approximated as |mincellQN| and |minfaceQN|) increases as k increases, while in average (values of mincells and minfaces) the lack of positivity becomes smaller as the order increases.

At this stage, it is worth pointing out the fact that quantifying the negativity of the solution is much more difficult for high-order schemes, since it is not possible to "count" the number of negative values (which corresponds to the degrees of freedom for low-order schemes). While the "mincellQN" value gives information about the minimum value reached on the whole domain, it does not give any indication about the measure of the set $\{x \in \overline{\Omega} \mid u_{\mathcal{M}}^{\omega,n}(x) < 0\}$ where the discrete cell unknown takes negative values. The same remark applies to "mincells". As an attempt to provide an idea of the size of this set, we display in Table 4.2 the number of cells with negative average over the whole simulation, defined as the cardinal of the set

$$\left\{(K,n) \in \mathcal{M} \times [\![1;50]\!] \mid \int_K u_{\mathcal{M}}^{\omega,n} < 0\right\}.$$

These data reveal that the higher the order, the smaller the size of the negative-values set (except for expf_0).

The previous observations seem to indicate a competition between two phenomena for the linear methods. As k increases, the accuracy is improved, and therefore the discrete solution becomes closer to the continuous one. Hence, in average, high-order schemes compute solutions with smaller area of negativity. However, high values of k induce larger oscillations for the

scheme	HMM	expf_0	expf_1	expf_2	expf_3
# cells with negative average	90	0	824	136	1

Table 4.2 – **Number of negative cell averages**

polynomial solution: the computed solution takes negative values on smaller sets but the undershoots become bigger as k increases. At the end, it seems that there is no hope to get a positive discrete solution on the whole domain Ω with a linear method on general meshes.

Remark 28 (An accuracy criterion taking into account positivity)**.** *The previous observations suggest that, for applications where preserving the positivity of the solution is an essential feature, the accuracy of the scheme should not simply be defined as an L^p distance between the discrete solution ($u_{\mathcal{M}}$) and the exact one (u). We believe that a relevant criterion in order to take into account both "classical accuracy" (distance between $u_{\mathcal{M}}$ and u) and positivity is to look at the relative Boltzmann entropy (or other kinds of "relative Φ-entropies" as defined in [33]) with respect to the exact solution, that is*

$$Err(u_{\mathcal{M}}) = \int_\Omega u\Phi_1\left(\frac{u_{\mathcal{M}}}{u}\right), \tag{4.39}$$

where $\Phi_1(s) = s\log(s) - s + 1$ for $s > 0$, $\Phi_1(0) = 1$ and $\Phi(s)$ with large values for $s < 0$. The interest of such a definition is twofold. First, the negativity of $u_{\mathcal{M}}$ is penalized. Second, if $u_{\mathcal{M}}$ is positive and $\int_\Omega u_{\mathcal{M}} = \int_\Omega u$ (which is the case in practice for problems with homogeneous Neumann boundary conditions), by Csiszár–Kullback inequality (see for example [51, Lemma 5.6]), one has

$$\|u_{\mathcal{M}} - u\|_{L^1(\Omega)} \leq \sqrt{\|u\|_{L^1(\Omega)} Err(u_{\mathcal{M}})}.$$

4.4.3 Convergence, accuracy and efficiency

Here, we are interested in the convergence as $(h_{\mathcal{D}}, \Delta t) \to (0,0)$ of the nonlinear schemes.

We consider a test-case with known exact solution. We set the advective potential and diffusion tensor to

$$\phi(x,y) = -x \quad \text{and} \quad \Lambda = \begin{pmatrix} l_x & 0 \\ 0 & 1 \end{pmatrix}$$

for $l_x > 0$. The exact solution is then given by

$$u(t,x,y) = C_1\, \mathrm{e}^{-\alpha t + \frac{x}{2}} \left(2\pi\cos(\pi x) + \sin(\pi x)\right) + 2C_1\pi\, \mathrm{e}^{x-\frac{1}{2}}, \tag{4.40}$$

where $C_1 > 0$ and $\alpha = l_x\left(\frac{1}{4} + \pi^2\right)$. Note that u^{in} vanishes on $\{x = 1\}$, but for any $t > 0$, $u(t,\cdot) > 0$. Here, our experiments are performed using $l_x = 1$ and $C_1 = 10^{-1}$.

To get information about the accuracy of the schemes, we compute the discrete solution on the time interval $[0, 0.1]$, and we denote by $(\underset{\sim}{u}^n_{\mathcal{D}})_{1\leq n\leq N_f}$ the corresponding discrete density. Then, we compute the relative $L^2_t(L^2_x)$ error on the solution and on the gradient of the solution, defined as

$$\frac{\sqrt{\sum_{n=1}^{N_f} \delta t^n \|u^n_{\mathcal{M}} - u(t^n,\cdot)\|^2_{L^2(\Omega)}}}{\|u\|_{L^2_t(L^2_x)}} \text{ and } \frac{\sqrt{\sum_{n=1}^{N_f} \delta t^n \|\mathbb{G}_{\mathcal{M}}(\underset{\sim}{u}^n_{\mathcal{D}}) - \nabla u(t^n,\cdot)\|^2_{L^2(\Omega)}}}{\|\nabla u\|_{L^2_t(L^2_x)}}$$

where $\delta t^n = t^n - t^{n-1}$ (note that $\sum_{1\leq n\leq N_f} \delta t^n = 0.1$) and $\mathbb{G}_{\mathcal{M}}(\underset{\sim}{u}^n_{\mathcal{D}})$ is the discrete gradient of the

densities. This discrete operator is defined by mimicking the continuous relation $\nabla u = \mathrm{e}^{\ell}\,\nabla\ell$ at the discrete level: $\mathbb{G}_{\mathcal{M}}(\underline{u}^n_{\mathcal{D}})$ is a piecewise continuous vector field satisfying, for any cell $K \in \mathcal{M}$,

$$\mathbb{G}_{\mathcal{M}}(\underline{u}_{\mathcal{D}})_{|K} = \mathrm{e}^{\ell_K}\, G^k_K(\underline{\ell}_K) \text{ in } K.$$

As already mentioned, L^2-norms (in space) are computed using quadrature formulas of order $2k+5$, and we use the approximation

$$\|u\|_{L^2_t(L^2_x)} \simeq \sqrt{\sum_{n=1}^{N_f} \delta t^n \|u(t^n,\cdot)\|^2_{L^2(\Omega)}}$$

for the exact solution norms (with an analogous definition for $\|\nabla u\|_{L^2_t(L^2_x)}$). Notice that, with the chosen definitions, we do not take into account the time $t = 0$. To plot the error graphs, we perform our simulations on a triangular mesh family $(\mathcal{D}_i)_{1\le i\le 5}$, such that $h_{\mathcal{D}_i}/h_{\mathcal{D}_{i+1}} = 2$. Since the time discretisation is of order one, on the i-th mesh of the family, we use an initial (and maximal) time step of

$$\Delta t_i = \frac{\Delta t_0}{2^{i(k+2)}}$$

where $\Delta t_0 = 0.05$. The goal of this choice is to fit the time step according to the expected order of convergence of the scheme (in space).

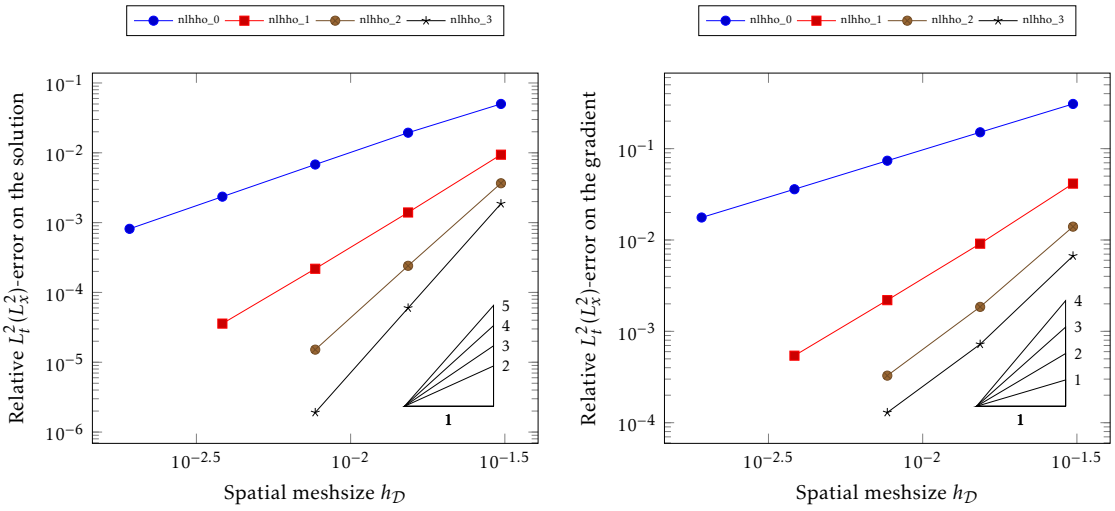


Figure 4.1 – **Accuracy of transient solutions.** Relative errors on triangular meshes.

In Figure 4.1, we plot the errors as functions of the meshsize $h_{\mathcal{D}}$ for different values of k. We see that, as one could expect, the scheme nlhho_k converges at order $k+1$ in energy-norm, and $k+2$ in L^2 norm of the density.

We now study efficiency, that is to say accuracy for a given computational cost. In Figure 4.2, we plot the errors as functions of the computing walltime. It is remarkable to see that, even with a low-order discretisation in time, significant efficiency gains can be obtained using a large value of k. The gain is expected to be even bigger after parallelising the local computations. Of course, the use of higher-order time-stepping methods should also lead to significant gains, and this

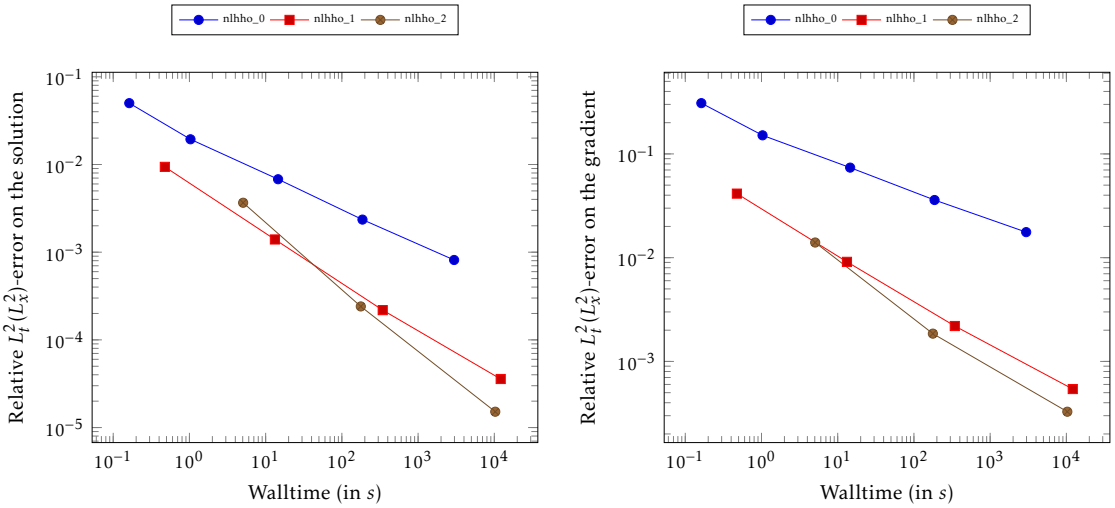


Figure 4.2 – **Accuracy vs. computational cost.** Relative errors on triangular meshes.

will be investigated in future works.

Remark 29 (High-order schemes in time and space)**.** *The extension of the nonlinear scheme* (4.20) *to arbitrary orders in time and space is a rather natural goal in order to achieve optimal efficiency. However, even with a discretisation of order* 2 *(like for example BDF2), there is currently no successful approach retaining the discrete entropy structure. Since this structure is the cornerstone of the analysis (including the existence of solutions), it is important to preserve it. Some numerical investigations on nonlinear entropic TPFA schemes for diffusive problems with order two time discretisation (BDF2) have been performed in [89, Chapter 3], and indicate that such a time discretisation could be a good candidate, even with respect to long-time behaviour (see [89, Chapter 3.4.4]).*

As a last accuracy test, we compare the nonlinear HHO scheme with the HFV scheme of Chapter 1. For the sake of completeness, we also include the scheme nlhho_2_0 (i.e. with $\varepsilon = 0$ in (4.18)) in our comparison. For the nlhfv scheme, we use the same definition of the errors as in Chapter 1. The results are presented on Figure 4.3. First, we can see that nlhho_2 and nlhho_2_0 have the same behaviour. Tests with other values of k, not shown here, indicate that the influence of ε (0 or 1) on the accuracy of the scheme is not noticeable. Regarding the efficiency, the interest of using high-order schemes seems clear (in particular when one is interested in the accuracy on the gradient). For the accuracy on the density, the trend is quite clear and shows that for (relative) errors smaller than 10^{-5}, the high-order schemes are more efficient. Note that the comparison presented here between the HFV and HHO schemes is not completely fair. Indeed, the HFV scheme has been implemented within an independent code, where a lot of optimisations (including precomputations and static condensation with explicit inverse matrices) are performed. Moreover, the question of the stopping criterion for the Newton method (see Remark 27) implies that it is difficult to compare "similar" versions.

Overall, the tests presented here confirm that the nonlinear high-order schemes allow for a consequent gain of efficiency compared with low-order schemes.

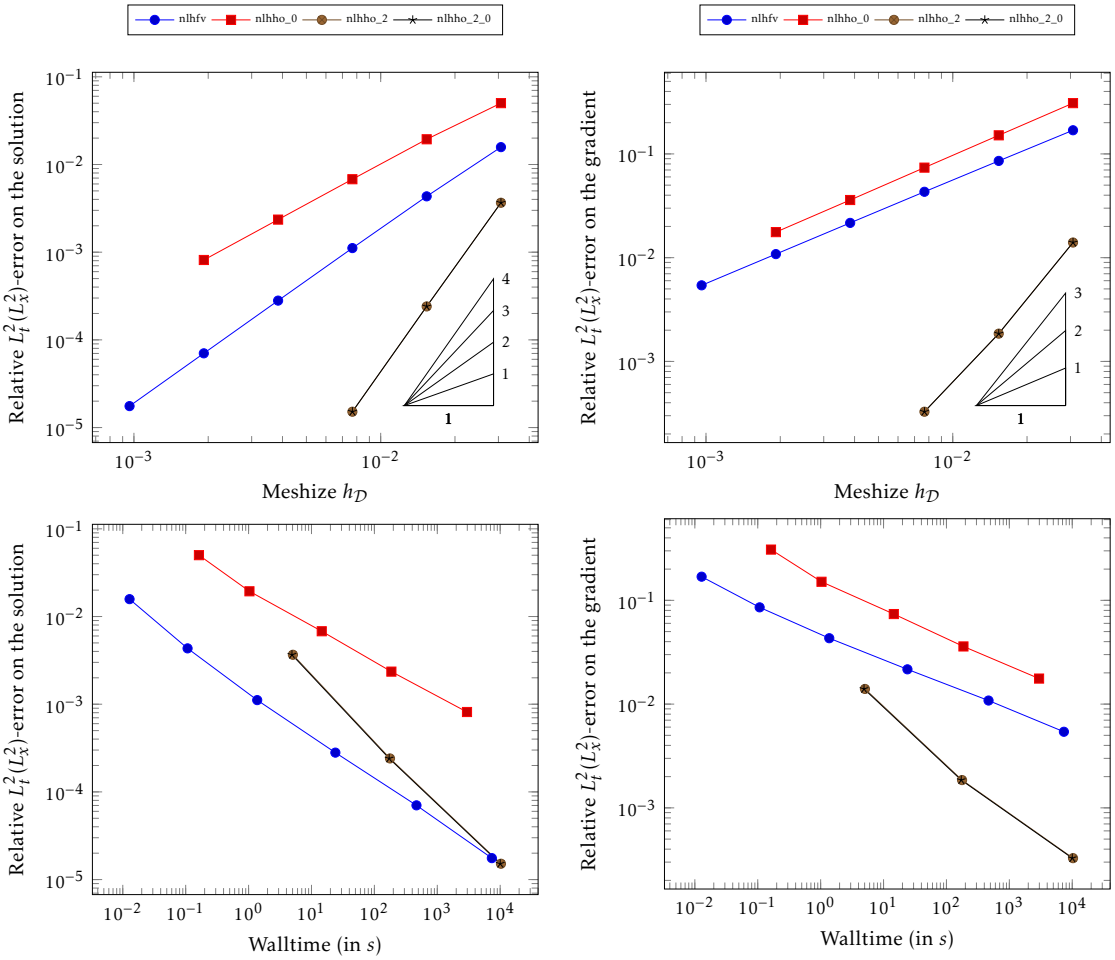


Figure 4.3 – **Accuracy: nlhfv vs. nlhho.** Relative errors on triangular meshes.

4.4.4 Discrete long-time behaviour

We are now interested in the long-time behaviour of discrete solutions. We use the same test-case as before, but with an anisotropic tensor: we set $l_x = 10^{-2}$. The corresponding steady-state is

$$u^\infty(x,y) = 2C_1\pi\, e^{x-\frac{1}{2}}.$$

We compute the solution on the time interval $[0,350]$, with $\Delta t = 10^{-1}$, on two Kershaw meshes of sizes 0.02 and 0.006. In Figure 4.4, we show the evolution along time of the L^1 distance between $\underset{\sim}{u}^n_{\mathcal{D}}$ and u^∞, computed as

$$\int_\Omega |u^n_{\mathcal{M}} - u^\infty|,$$

for the nlhho_k schemes, for $k \in \{0,1,2\}$. We observe the exponential convergence towards the thermal equilibrium for each value of k, until some precision is reached. For $k \geq 1$, the rates of convergence are similar to the exact one (α), and do not depend on the size of the mesh. For $k = 0$, the rate of convergence towards equilibrium differs a bit from α on the coarsest mesh,

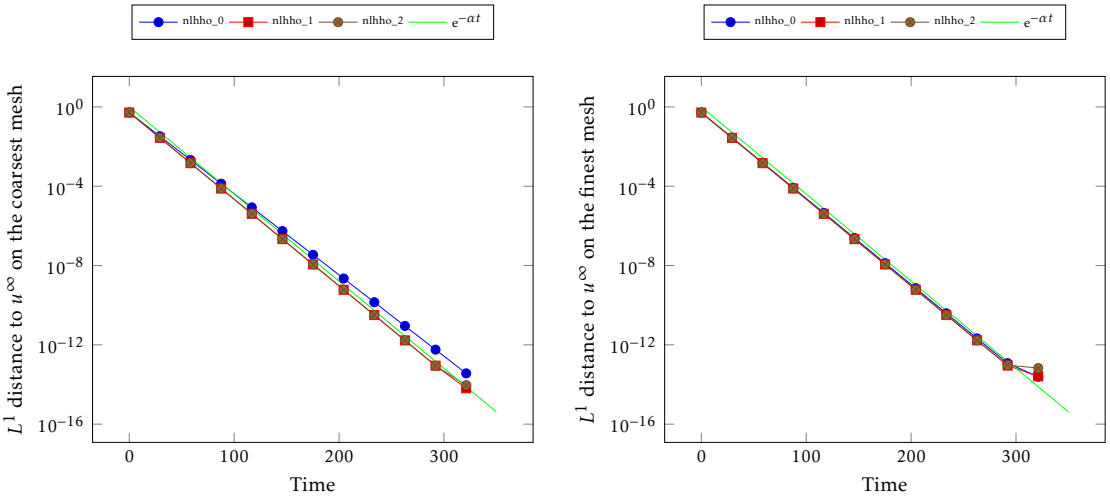


Figure 4.4 – **Long-time behaviour of discrete solutions.** Comparison of the long-time behaviour of the solutions to the nonlinear HHO schemes on Kershaw meshes for $T_f = 350$ and $\Delta t = 0.1$.

but theses two rates seem to coincide on sufficiently refined meshes. Note that for the test-case considered, $\phi \in \mathbb{P}^1(\Omega)$, therefore $\phi_{\mathcal{M}} = \phi$. It follows that $u_{\mathcal{M}}^\infty = u^\infty$, hence the very low saturation threshold.

For the exponential fitting scheme, the distance between $\underline{\underline{u}}_{\mathcal{D}}^n$ and u^∞ is computed as

$$\int_\Omega |u_{\mathcal{M}}^{\omega,n} - u^\infty|.$$

To assess the behaviour of the exponential fitting scheme, we last compare the results for three

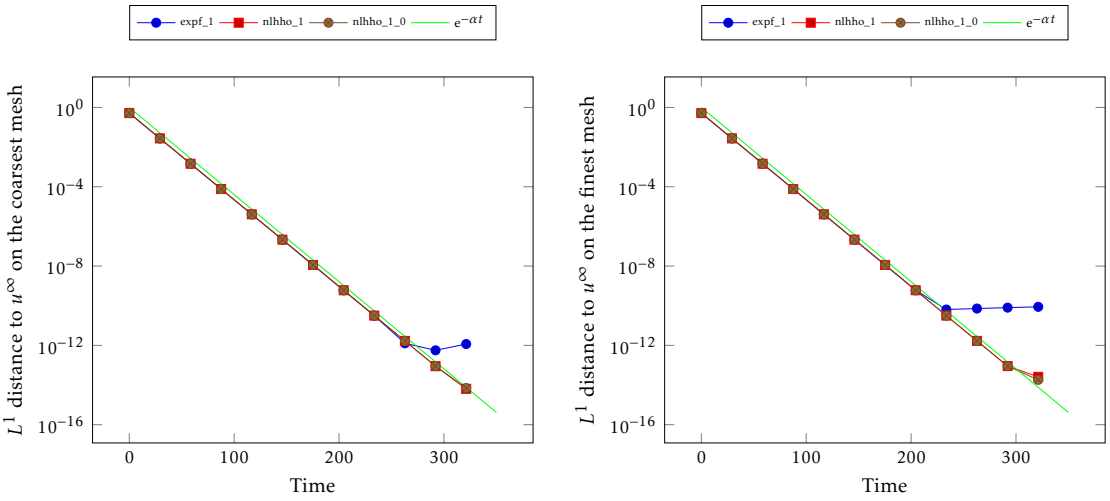


Figure 4.5 – **Long-time behaviour of discrete solutions.** Comparison of the long-time behaviour of the solutions to the nonlinear HHO schemes on Kershaw meshes for $T_f = 350$ and $\Delta t = 0.1$.

schemes with face unknowns of degree $k = 1$: nlhho_1, nlhho_1_0 and expf_1.
The results are presented in Figure 4.5. As previously, one can note that nlhho_1 and nlhho_1_0 exhibit an extremely similar behaviour. Regarding the exponential fitting scheme, as stated in Proposition 14, the discrete density converges exponentially fast towards the equilibrium. Moreover, as for the nonlinear schemes, the decay rate is similar to the continuous one.

4.5 Conclusion

In this chapter, we have introduced two arbitrary-order schemes for linear anisotropic advection-diffusion equations on general meshes. The first one is a linear scheme which follows the exponential fitting strategy, whereas the second is a nonlinear scheme which follows the low-order philosophy [56, 52, 70]. We have proved that both schemes admit solutions, have an entropy structure and preserve the long-time behaviour of the continuous problem. Moreover, the solutions to the nonlinear scheme are positive by construction. We have validated these theoretical results on numerical test-cases. We have also highlighted the lack of positivity of the linear method, which justifies the use of (more costly) nonlinear methods. In the meantime, the use of nonlinear schemes with polynomial unknowns of sufficient degree proves to improve the efficiency (accuracy at given computational cost) with respect to low-order nonlinear methods. These results confirm the benefits of using high-order nonlinear schemes in order to get reliable approximation of diffusive problems. Future directions of this work include a deeper analysis of the nonlinear scheme, as well a the development of similar schemes for more complex and nonlinear problems.

Conclusion and perspectives

Outline of the current chapter

The work presented in this thesis shows that high-order nonlinear structure-preserving methods for anisotropic diffusive problems on general meshes can be used successfully. The study performed in Chapter 1 shows that non-linear methods are needed in order to preserve positivity of solutions to anisotropic problems on general meshes, and introduces a nonlinear structure-preserving HFV method. This method is then used on complex and nonlinear anisotropic problems in Chapter 2 and 3, and allows one to approximate solutions to the drift-diffusion system (3) while preserving the structural bounds (implying in particular positivity) on the densities and the long-time behaviour. These encouraging results motivate the design of high-order schemes with similar structure-preserving features. Using the quasi-Fermi potentials as key variables, we successfully adapt the nonlinear HFV discretisation to an HHO one in Chapter 4. The scheme introduced exhibits an optimal order of accuracy, while preserving the positivity and the long-time behaviour of the solutions. The implementation proposed enjoys a robustness similar to the HFV one, and the use of sufficiently high-order schemes gives a consequent gain of computational efficiency with respect to the HFV scheme.

Thus, this work paves the way to various extensions and applications to other models. We discuss some of them below, regarding the analysis, the implementation and the applications.

Further analysis of the schemes

Regarding the analysis of the schemes, the most interesting challenges lie in the way of getting uniform (with respect to the meshsize) estimates needed in order to get long-time behaviour and convergence results.

The first natural question about the analysis of the HHO scheme is to understand how to get uniform (with respect to the meshsize) estimates. Indeed, such estimates are needed in order to prove both convergence of the scheme and uniform long-time behaviour. Performing an analysis of the long-time behaviour of the nonlinear HHO scheme similar to the one performed in Chapter 1 is a first important step and an ongoing work. The main issues is to get an high-order

counterpart of Proposition 6. The main difference when it comes to the high-order scheme is the difference between potential (polynomials) and densities: one has to deal with both in order to get the appropriate estimates. Another important task then is to adapt discrete functional inequalities of Section 1.A to the case of high-order unknowns.

The convergence analysis of the non-linear schemes for advection-diffusion equations is an important perspective of our work, for both HFV and HHO schemes. The analysis of the HFV scheme was not proved in this thesis, however the long-time behaviour analysis gives some uniform estimates on H^1 discrete semi-norms in Proposition 6. Such estimates, combined with the compactness toolbox [6] are the key ideas to show the convergence of similar DDFV and VAG schemes [56, 52]. The adaptation of these arguments to the HFV framework should be possible. For the high-order scheme, one needs to get uniform estimates (see above) and handle non-polynomial density. It is not known whether the toolbox [6] can be used directly.

The convergence analysis of schemes for drift-diffusion systems are more difficult. Indeed, the coupling between the unknowns makes it difficult to get uniform estimates on the relevant quantities. Up to our knowledge, the convergence analysis of existing schemes for drift-diffusion systems relies on L^∞ bounds on the densities and/or on the electrostatic potential. On the other hand, getting such estimates (uniform with respect to the meshsize) seems rather difficult for the methods considered here. A potential approach could be to adapt arguments used for the recent analysis of a continuous memristor model [170], which is essentially a drift-diffusion systems with three densities instead of two. The main interest of this analysis is the fact that it does not rely on L^∞ estimates on the solution (which are not available because there are three species instead of two) but rather on a pure entropy/energy method. We believe that such a method could be used at the discrete level with our schemes. Note that the analysis of long-time behaviour (with an uniform rate of decay with respect to the meshsize) of HFV schemes for the drift-diffusion system should also be performed using similar ideas.

Another important task is related to the analysis of the nonlinear schemes in situations out of equilibrium. Such analysis is performed for a nonlinear $\mathbb{P}^1$ finite element scheme [57] which enjoys the same structure as the nonlinear HFV scheme. The adaptation of this analysis to the nonlinear HFV scheme seems possible, but rather technical since one has to deal with non-conformal method. A related work could be the analysis of positivity-preserving schemes for stationary problems. The accuracy study of Chapter 1 suggests that the stationary version of the nonlinear HFV scheme is suited in order to approximate solutions to stationary problem, but up to now the analysis this scheme (or similar VAG or DDFV schemes) was not performed. A relevant analysis approach should be to use the relative entropy with respect to a lifting of the boundary condition in order to get entropy estimations. However, the main difficulty in order to analyse such a stationary scheme is related to the way of getting estimates on the entropy dissipation without the help of the time derivative term. Such estimates could be consequences of refined Log-Sobolev inequalities.

More efficient and robust implementations

In terms of implementation, we are working on a couple of ways to improve the numerical behaviour of the schemes.

A first interesting question about the numerical robustness is the implementation of the Newton method. Indeed, some recent work [38, 24] suggests that the use of parametrisation (or change of unknowns) in the Newton method could strongly improve its robustness. For the HFV schemes, such strategy would mean be to solve the problem using either densities or quasi-Fermi potentials as primal unknowns. Note that the use of quasi-Fermi potentials as

unknowns is the solution implemented in the TPFA code **ddfemi** [104]. On the other hand, for the HHO scheme, since we can only use the coefficients of the polynomials representing the quasi-Fermi potentials, it is not straightforward to switch between densities and potentials. Another possibility could be to consider more subtle resolution method, using for example line search techniques. Some numerical test performed during the phase development of **hho_nl** with relaxed Newton steps indicated that a better robustness can be achieved with such strategy, but at the cost of many iterations. Of course, such numerical work could also be associated to a theoretical study, in order to ensure the behaviour of the Newton method (invertible Jacobian matrices, global convergence of the methods).

Another important feature which should be investigated is the question of mesh adaptation. In [143], it is suggested that an equidistribution of the entropy dissipation over the domain (which means, the same dissipation on every cells) leads to a better repartition of the errors, and therefore to a better efficiency. We believe that refining the mesh in areas where the dissipation is too big can be a way of improving the convergence of the Newton methods. Such questions are the topic of an ongoing work on TPFA schemes with collaborators from WIAS and Tyndall National Institute for semiconductors models studied in Appendices: we aim at understanding how to use the entropy and its dissipation in order to get coarser mesh without loosing too much accuracy. Our ultimate goal is to simulate bigger devices than the ones studied in Appendix D. Of course, the use of HFV and HHO schemes should be a better solution when it comes to mesh adaptation, given the flexibility afforded by general meshes.

Last but not least, an important question which was not addressed during this PhD is the parallelisation of the computations for the hybrid schemes. Indeed, these methods are based on local contribution, and a lot of computations can be performed in parallel (static condensation, computation of every local quantity). Such implementation should allow one to get a way better efficiency for the scheme implemented, and could be part of the development of the ParaSkel++ [18] platform. Natural related development in this framework includes methods handling three dimensional domains and the use orthogonal polynomial basis instead of monomial ones.

Applications to other complex problems

In term of non-academic problems, my work essentially focused on semiconductor models, and the main goal was to approximate (3). In view of the results of Chapter 4, an extension of the work performed in Chapters 2 and 4 to develop and analyse an high-order scheme for semiconductor models is a short-term objective in order to get more efficient structure-preserving schemes for this problem.
Of course, the methodology developed in this thesis is intended to be used in a variety of different settings and models. Here, we discuss three situations of interest. As above, the schemes developed could be either HFV ones (as a preliminary work) or HHO ones.

A first natural application is related to porous media problems, for which there are challenges related to anisotropy. In particular, there exists some HMM, DDFV and HHO based schemes [65, 71, 72, 5] for Peaceman problem. These schemes are shown to be convergent, but there is no theoretical guaranties that the computed concentrations are non-negative. Therefore, the use of the non-linear methods of this PhD could be a solution in order to get a more reliable approximation of the solutions.

Another field of application could be the wide family of cross-diffusion systems. These are systems of multiples unknowns, in which the diffusion coefficients for each unknown depends on the other unknowns. As for the drift-diffusion systems (which can be seen as simple prototypes

of general cross-diffusion systems), the developpement and analysis of TPFA schemes for cross-diffusion is an active field of research [54, 147, 177]. On the other hand, there exists some anisotropic cross-diffusion models obtained as rigorous limits of stochastic models [230, 117]. For such models, the use of the previous TPFA schemes is not possible: the use of HFV and HHO schemes based on the work of this thesis should be a natural solution to get reliable approximation methods. Of course, the use of high-order schemes should also be an interesting solution to approximate such systems in an efficient way. Since these systems are generally analysed thanks to boundedness by entropy methods [47, 173], the methodology used to analyse our scheme should be adaptable without too much difficulty to cross-diffusion.

Another interesting question is the generalisation of the results presented in this thesis to convection fields, which are not the gradient of potentials. Indeed, in the study performed here, the only scheme which is able to handle such a general field is the HMM scheme [19] studied in Chapter 1. However, this scheme does not preserve positivity, neither thermal equilibrium. Moreover, its analysis relies on coercivity assumptions. From a more general point of view, the main difficulty of such a situation in order to develop structure-preserving methods is that the quasi-Fermi potentials, keystone of our structure preserving methods and the associated analysis, do not exist anymore. A first direction in order to solve this issue could be to use a Hodge decomposition of the field $V = \nabla\phi + V_0$, where V_0 is divergence-free, and treat the irrotational part in a non-linear way by using a partial quasi-Fermi potential $\log(u) + \phi$. The advective (and coercive) part induced by V_0 could then be discretised using the scheme of [19]. Computations at the continuous level indicate that this strategy -which is in a way similar to the one used to analyse non-coercive advection-diffusion problems [106]- allows one to get some estimates on the entropy dissipation. On the other hand, at the discrete level the way of getting this estimate is not straightforward and one has to understand how to perform the Hodge decomposition in an appropriate way. Note that this situation is not purely academic, and can be used in practical situations. In particular, there exists a series of work by Jochmann [163, 164, 165, 167, 166, 168] about semiconductor models in which the electrostatic filed does not solve a Poisson equation. Such a model correspond to a situation where the internal magnetic effects are not neglected anymore: the electrostatic field is not the gradient of a potential anymore, but is given by Maxwell equations. It could be interesting to develop positivity preserving high-order methods for such a system, using the methodology introduced in this thesis for the convection-diffusion equation alongside with some recent HHO methods for Maxwell equations [76]. Another interesting direction for such work could be to investigate the asymptotic preserving features of schemes with respect to the asymptotic analysis done by Jochman for the continuous model (long-time behaviour and singular limit when the magnetic susceptibility tends to zero).

Numerical comparisons with existing methods

As far as possible, we gave numerical comparison between different methods in order to identify and exhibit the strengths of each method introduced.

The results of Chapter 4 indicate that the high-order structure-preserving method developed in this thesis have a better efficiency that the nonlinear HFV method of Chapter 1. However, keeping in mind that the methods developed here address the need to handle anisotropy and substitute for TPFA schemes, a question of great importance is the comparison of the hybrid methods with the TPFA ones. Indeed, in situations where both classes of methods are available (isotropic problems on orthogonal meshes), the efficiency of the high-order schemes could be a strong argument in favour of HHO schemes. In order to understand the performances of each scheme, one need to perform tests on various nonlinear problems. In order to avoid the

limitations induced by the (first order accurate) time discretisation and give a fair comparison, the ideal situation should be to use stationary problems.

Another interesting work could be an extension of the results presented in Chapter 3. Indeed, we only focused on an isotropic drift-diffusion system, but one could ask if both HFV and DDFV schemes have the same behaviour for the anisotropic problem (3) with general statistics. Ongoing numerical investigations indicate that both schemes handle anisotropy for Boltzmann statistics. Moreover, the influence of the intensity of the magnetic field on the rate of exponential decay towards equilibrium (results in the same spirit as Figure 2.11b) is rather similar for both schemes. In view of the results of Chapter 4, it would be interesting to also compare the low-order schemes of Chapter 3 with a nonlinear structure-preserving HHO scheme for anisotropic drift-diffusion systems. Such comparative work could be a first step in order to investigate the behaviour of high-order structure-preserving schemes on complex problems and get reference results. On the other hand, such a work could also be the opportunity to investigate the question of efficiency (and the expected gain for the high-order methods) of the schemes on nonlinear problems.

High-order discretisations in time

A last important perspective of this work is related to the time-stepping method used. Indeed, all schemes presented in this manuscript are based on a backward Euler discretisation in time. Such a choice allows one to get discrete entropy structures relatively easily as a consequence of the convexity of the entropies. On the other hand, since every schemes are of order at least two in space, which leads to a rather large unbalanced space-time accuracy. This is even worse for high-order schemes: the use of high-order methods for time discretisation should provide a scheme higher efficiency.

The main difficulty when it comes to high-order methods in time is the way of getting the entropy relation, which is the base of all the nonlinear schemes developed here. A first step should be to test different numerical strategies in order to check both stability and efficiency of the schemes. Such a comparative work was performed for TPFA schemes in [89], and shows that BDF2 schemes provide good stability properties for entropic schemes. The analysis of such a scheme could rely on the methodology used in [146] to analyse an entropic scheme using BDF2. Another interresting direction could be the use of Runge-Kutta schemes following the work [176] on semi-discrete problems. A possible interesting solution could be the use of Strong Stability Preserving Methods, which are often used in the context of hyperbolic equations [151, 152]. Note that for the linear schemes used in this thesis, the use of a Crank–Nicolson discretisation in time leads to an exact entropy-dissipation relation at the discrete level (with an equality as in the continuous framework). However, getting a control of the entropy by the dissipation (which then depends on two times) in this situation seems rather difficult. Moreover, such a time discretisation is known to generate spurious oscillations which should degrade the positivity of solutions.

Appendix A

Semiconductor models with varying band edge energies: an overview

Outline of the current chapter

A.1 Motivations and context

In the Appendices, we focus on models with irregular convection fields. Such models are particularly useful in describing realistic devices, where the structure induces local fluctuations in internal energies. The prototype equation for modeling such a framework can be reduced to the linear problem:

$$\partial_t u - \operatorname{div}(\nabla u + u\nabla(\phi + E)) = 0, \tag{A.1}$$

where ϕ is a regular potential (solving a Poisson equation in the context of drift-diffusion systems) and E is a function known as the Band Edge Energy (BEE), which depends on the material (alloy) used in the semiconductor. In classical models (as in the drift-diffusions systems studied in Chapter 2 and 3), this energy is assumed to be constant. However, there are practical situations in which the BEEs are not constant and can even be irregular. Two cases of interest are:

— regions of the alloy with strong local fluctuations, referred to as quantum wells, where the BEEs exhibit extreme irregularity (essentially, a function taking different values at each atom);

— junctions of two layers made of different materials, known as heterojunctions, where the BEEs are piecewise constant.

Random alloy fluctuations significantly affect the electronic, optical, and transport properties of (In,Ga)N-based optoelectronic devices. To bridge the gap between macroscale drift-diffusion simulations and atomistic band-edge fluctuations, recently a multiscale framework was developed to integrate the macroscopic and microscopic worlds [221].

To understand how we can appropriately approximate the solutions of such problem, it is necessary to establish a framework for analysing the continuous equation (A.1). To the best of our knowledge, the most comparable situations in the literature are related to "distributional drifts" in stochastic problems [136, 137, 238, 161, 162]. In these works, the problems considered are driven by advection fields that lie in low-regularity Sobolev spaces, and the solutions are defined as limits of solutions to regularised problems. However, these advections are too regular for our study, as we aim to consider drifts that are singular measures, such as gradients of piecewise constant functions.
On the other hand, while the advection considered in (A.1) can be highly irregular, it has the specific structure of a gradient. By formally defining $\rho = \frac{u}{\omega}$, with $\omega = \mathrm{e}^{-\phi - E} \in L^\infty(\Omega)$, we can express an equivalent formulation of (A.1) in terms of ρ, namely

$$\omega \partial_t \rho - \operatorname{div}(\omega \nabla \rho) = 0. \tag{A.2}$$

Such an equation is classically well-posed, as the diffusion coefficient ω is bounded and uniformly positive on Ω. Hence, one can define the concept of solution to (A.1) by employing the exponential fitting formulation (A.2). Interestingly, this notion of solution can also be obtained as a limit of regularised solutions.
This approach can be generalised to problems with non-linear statistics (written here for the stationary problem to fix ideas)

$$-\operatorname{div}(u \nabla(h(u) + \phi + E)) = f,$$

where h is a function corresponding to the considered statistics (see, for example, the introduction of Chapter 2). By defining $\rho = \frac{\mathrm{e}^{h(u)}}{\omega}$, one formally obtains an exponential fitting formulation:

$$-\operatorname{div}\left(\frac{g(\log(\rho) - \phi - E)}{\rho} \nabla \rho\right) = f, \tag{A.3}$$

where g is a positive function, the inverse of h. In this formulation, similar to the previous discussion, the gradient of the irregular term E no longer appears.

At the numerical level, we propose a strategy based on the exponential fitting formulations and Two Point Flux Approximation (TPFA) finite volume methods. As for the HFV exponential fitting developed in Chapter 1, one of the main challenges lies in averaging the diffusion coefficient ω (or $g(\log(\rho) - \phi - E)$ for general statistics), which is irregular. It is worth noting that this issue has been addressed in the context of heterogeneous diffusion problems for TPFA schemes [121, Section 2.3.1]. Our numerical results (see Figure A.1) demonstrate the suitability of the exponential fitting for this type of problem, provided that an appropriate averaging technique is employed to handle the irregular BEE. In practice, it is necessary to use a harmonic mean for the averaging process.

The remainder of this appendix is organized as follows. First, we present the results of Appendices B-C-D, introducing the physical context of the problems and the issues addressed. Then, in Section A.5, we summarise various perspectives linked to this work, from both physical and mathematical points of view.

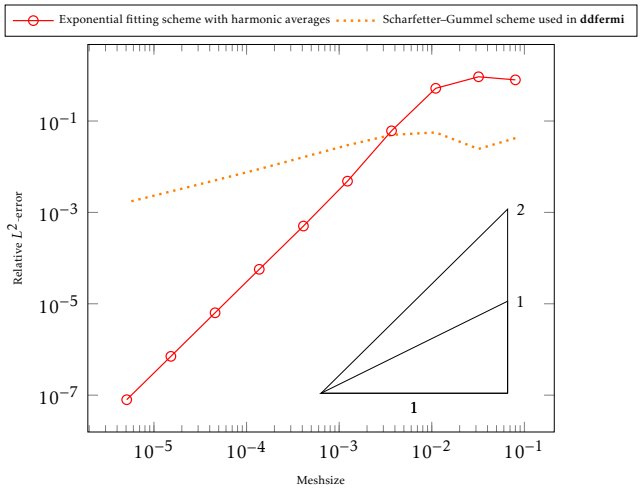


Figure A.1 – **Accuracy of schemes.** Boltzmann statistics and heterojunctions.

A.2 A comparison between numerical simulations and experimental results (Appendix B)

In Appendix B, we compare the numerical results obtained using our strategy with those obtained from commercial packages and experimental data.

Specifically, we consider an (In,Ga)N/GaN semiconductor system with quantum wells and introduce a model that incorporates random fluctuations in the alloy. This is achieved by computing appropriate band edge energies using the Localisation Landscape Theory (LLT) approach [11, 132, 224]. The LLT approach allows for the prediction of charge localisation regions and their corresponding energies. Our method differs from classical strategies implemented in commercial software, which employ a Schrödinger equation to incorporate quantum corrections. However, these Schrödinger-Poisson solvers have a significant computational cost, and currently, these commercial packages only support 1D simulations.

To assess the validity of our approach, we compare our simulations with results obtained using the commercial package next**nano** [30], which employs a full Schrödinger-Poisson equation, as well as experimental data. In particular, we investigate the impact of the well order on device behaviour. Our results demonstrate good agreement between our simulations and those presented in [30]. Moreover, considering the random fluctuations is necessary to accurately reproduce the observed trends in the experiment.

A.3 The question of thermodynamic consistency (Appendix C)

In Appendix C, we investigate certain issues related to the approximation of models with non-Boltzmann statistics. These discussions are motivated by observations resulting from the use of incorrect discretisation methods for solving problems with varying BEEs. However, they also apply to classical models with constant BEE.

These issues are well-known within the mathematical community and have led to numerous studies, ranging from the development of schemes handling general statistics to the comparison of such methods [265, 120, 26, 140, 50, 125, 127, 50]. However, in more applied scientific

communities, these questions are relatively under-documented.

The work presented in Appendix C aims to provide physics-based evidence that using inconsistent discretisations of the flux can significantly impact the results of simulations. The main novelty of this work lies in investigating the influence of discretisation on important physical behaviours, namely the recombination process and current-voltage (I-V) curves. The former is of major importance in LED applications, while the latter serves as a basis for engineers to model the device behaviour in an electrical circuit. The study conducted provides strong evidence that employing an inconsistent Scharfetter–Gummel flux in situations where non-Boltzmann statistics are used leads to non-physical results.

As a by-product of this study, we also introduce a new definition of thermodynamically consistent numerical flux (C.7). This notion is based on the relationship between the monotonicity of the quasi-Fermi potential and the sign of the flux. The introduced definition corresponds to a discrete version of this continuous feature. The main advantage of this notion is that, unlike the classical one [1], it holds meaning even outside of thermal equilibrium. It appears that this notion is linked to the methodology used in [50] to analyse TPFA schemes for Blakemore statistics: the numerical fluxes are expressed as the product of a positive interface term by the discrete derivative of the quasi-Fermi potentials.

A.4 Application to the design of efficient LEDs (Appendix D)

The last appendix focuses on a practical question: how to design efficient LEDs. We are specifically interested in LEDs operating in the deep-UV-C wavelength range (below 280 nm), which can be used for applications such as sterilisation or water purification filters. Aluminium gallium nitride ((Al,Ga)N) based LEDs are considered ideal candidates for such devices, as their emission wavelengths can be tuned across the entire UV spectrum without the need for toxic mercury. However, the use of (Al,Ga)N alloy is relatively new, and there is limited literature available compared to more classical (In,Ga)N-based optoelectronic systems. The main goal of our work is to understand the impact of random fluctuations in the alloy on LED efficiency.

To understand the behaviour and efficiency of LEDs, it is important to examine the recombination processes. In the situation under study, two main types are of interest:

— Radiative recombination, which involves electron transitions from the conduction band to the valence band accompanied by photon emission. This process is the primary source of light emission.

— Non-radiative Auger-Meitner recombination, which involves internal transfer of electrons (from the valence band to the conduction band) or of holes. This internal process can be viewed as an energy leak associated with the "emission" of phonons. Experimental results [148, 233] indicate that at high densities of charge carriers, this recombination process can physically degrade the device.

As a preliminary approach, we can consider that the ideal device will exhibit high radiative recombination and low non-radiative recombination. Our study focuses on these two types of recombination, comparing the device behaviour with and without random fluctuations.

Ultimately, our results suggest that random fluctuations increase both radiative and non-radiative recombination. To improve efficiency, careful device design is necessary to control the increase in non-radiative recombination. This can be achieved, for example, by using wider wells or increasing the number of wells in relevant areas.

1. If the numerical fluxes vanish, then the quasi-Fermi potentials are constant.

A.5 Some mathematical issues raised by these works

Regarding the work performed on the toy model, the numerical investigation suggests that it is possible to develop accurate schemes even for irregular BEEs. A useful first step before analysing schemes for full drift-diffusion systems could be to prove the convergence of the scheme for the toy problem equation. Convergence analysis for Boltzmann statistics is a classical approach using the compactness method since it only involves an equation with a heterogeneous diffusion coefficient. However, the case of non-Boltzmann statistics is more challenging. In fact, even for the continuous modified equation (A.3), we do not yet know if solutions exist. One possible first step could be to study the continuous equation and show, as in the linear case, that an equivalence holds between regularised solutions and exponentially fitted ones. Another possibility could be to directly prove the convergence of the numerical schemes using compactness. Such a result would imply, in particular, that a solution to the continuous problem exists. The techniques used in [50] could be relied upon for the analysis.

When it comes to simulating realistic devices, one of our medium-term objectives is to compare the numerical results obtained using our framework with experimental data that will soon be produced at the National Tyndall Institute. This comparative work will allow us to confirm the relevance of the approach used (LLT combined with the drift-diffusion system) in modelling complex devices. Once this confirmation has been obtained, it will be possible to apply our simulations to study optimal designs for LEDs.
The main hindrance to perform these comparisons is the cost of the simulations. Since we use an extremely refined mesh in the wells, our computations are not affordable for realistic-sized devices with a large number of wells. Furthermore, the adaptation between refined and coarse areas is challenging due to the orthogonality constraint on the (TPFA) admissible meshes. To address this issue, we are currently working on ways to locally coarsen the mesh while maintaining the same device behaviour. One of our goals is to find good heuristic criteria for mesh adaptation. A related question could be linked to homogenisation. Instead of using the exponential fitting approach to define a continuous solution, we could view the BEEs as oscillating functions and employ more classical homogenisation methods to obtain a well-posed limit equation. Such homogenised equation could be easier to discretise and would allow us to simulate larger devices with more quantum wells.

All the previous observations suggest that, according the promising results of Chapter 4, an interesting perspective could be to use HHO structure-preserving methods in order to simulate the devices under consideration. Regarding the irregularity of the BEEs, we believe that the non-linear strategies of the hybrid schemes developed in this thesis are adapted to handle it. Indeed, returning at the initial flux

$$J = u\nabla(\log(u) + \phi + E),$$

one can introduce a (partial) quasi-Fermi potential $w = \log(u) + E$, and expresses the flux as

$$J = \mathrm{e}^{w-E}\,\nabla(w + \phi).$$

Here, the irregular BEE E only appears in the diffusion coefficient. Hence, we can approximate the corresponding PDE using the same strategy as in Chapter 4, by discretising the potentials w and ϕ as polynomials. From a numerical perspective, the main difficulty caused by the irregularity of the BBE is the heterogeneity of the diffusion coefficient. However, this feature is not an issue in theory when it comes to the HHO method used (since it can handle heterogeneous and anisotropic diffusion). Of course, from a practical point of view, as usual with exponential fitting, the main challenge lies in approximating the integrals $\int_K \mathrm{e}^{-E}$ with irregular and oscillating

integrands.

Appendix B

Impact of random alloy fluctuations on the carrier distribution in multi-color (In,Ga)N/GaN quantum well systems

Outline of the current chapter

This chapter corresponds to a submitted work [219], in collaboration with Michael O'Donovan, Patricio Farrell, Timo Streckenbach, Thomas Koprucki and Stefan Schulz.

In this work, we study the impact that random alloy fluctuations have on the distribution of electrons and holes across the active region of a (In,Ga)N/GaN multi-quantum well based light emitting diode (LED). To do so, an atomistic tight-binding model is employed to account for alloy fluctuations on a microscopic level and the resulting tight-binding energy landscape forms input to a quantum corrected drift-diffusion model. Here, quantum corrections are introduced via localization landscape theory

and we show that when *neglecting alloy disorder* our established theoretical framework yields results very similar to commercial software packages that employ a self-consistent Schrödinger-Poisson-drift-diffusion solver; this provides validation of the developed quantum corrected transport model. Similar to experimental studies in the literature, we have focused on a multi-quantum well system where two of the three wells have the same In content while the third well differs in In content. By changing the order of wells in this 'multi-color' quantum well structure and looking at the relative radiative recombination rates of the different emitted wavelengths, we (i) gain insight into the distribution of carriers in such a system and (ii) can compare our findings to trends observed in experiment. Our results indicate that the distribution of carriers depends significantly on the treatment of the quantum well microstructure. For instance, when including random alloy fluctuations and quantum corrections in the simulations, the calculated trends in the relative radiative recombination rates as a function of the well ordering are consistent with previous experimental studies. However, the results from the widely employed virtual crystal approximation contradict the experimental data. Our calculations clearly demonstrate that when accounting for random alloy fluctuations in the simulations, no further ad-hoc modifications to the transport model are required, in contrast to previous studies neglecting alloy disorder. Overall, our work highlights the importance of a careful and detailed theoretical description of the carrier transport in an (In,Ga)N/GaN multi-quantum well system to ultimately guide the design of the active region of III-N-based LED structures.

B.1 Introduction

At the heart of modern light emitting diodes (LEDs) operating in the blue to violet spectral region are (In,Ga)N/GaN multi-quantum well (MQW) systems [158]. While the efficiency of these LEDs is and can be very high, further efficiency gains will still directly reduce the cost of operating such LEDs. Moreover, extending efficient operation of (In,Ga)N-based LEDs into the green to red spectral range is a topic of current research interest [201, 240, 12, 116, 159]. To achieve all this, understanding the fundamental electronic, optical and transport properties of (In,Ga)N-based MQW systems is of central importance to guide design of (In,Ga)N-based LEDs with new and improved capabilities. While experimental and theoretical studies on the electronic [59, 75, 252] and optical [169, 95] properties of such systems have already revealed that these properties are significantly impacted by alloy fluctuation induced carrier localization effects, the impact of alloy disorder on the carrier transport has only been targeted recently [220, 221, 222, 203, 44, 194]. Here, studies are ranging from fully atomistic quantum mechanical approaches [220] up to modified continuum-based models [221, 222, 203, 44, 194].

Recently, we have developed a three-dimensional (3-D) multiscale simulation framework that connects atomistic tight-binding theory with a modified, quantum corrected drift-diffusion (DD) solver. The framework has been employed to investigate uni-polar carrier transport and it was found that alloy fluctuations result in an increase in carrier transport of electrons (in an *n-i-n* system) [221], but decreases transport in the case of holes (in a *p-i-p* systems) [222]. In the present work we extend this scheme to investigate the active region of (In,Ga)N-based MQW LED structures (thus *p-i-n* systems) and study how carriers distribute across the active region. In general, understanding the carrier distribution can help to guide maximizing the efficiency in an LED, since ideally the carriers shall be distributed evenly across the entire MQW region so that all QWs will contribute to emission [266]. However, previous experimental studies on

carrier distribution in (In,Ga)N/GaN MQW systems have indicated that mainly the well closest to the *p*-doped contact side contributes to the light emission process [145, 92, 199, 266]. These samples were specifically designed to gain insight into the carrier distribution inside the active region of an LED.

Overall, this has been attributed to a sequential filling of the QWs, resulting in a high hole density only in the *p*-side QW. To establish accurate carrier transport models the trends found in the experimental studies of Refs. [145, 92, 199, 266] need to be captured. Previous theoretical studies have reproduced the experimentally observed behaviour, however this required (i) treating bound carriers in a quantum mechanical picture, (ii) softening of the QW barrier interface to account for tunneling effects, (iii) distinguishing between continuum and bound carriers in the carrier transport model (multi-population model), and (iv) allowing for scattering between the different populations [236]. But, the impact of alloy disorder is basically neglected in this advanced but also complex carrier transport model.

In this paper we show that when employing our quantum corrected 3-D simulation framework that accounts for random alloy fluctuations, the experimentally observed trends are captured, without introducing for instance a multi-population scheme. This highlights that our developed solver presents an ideal starting point for future device design studies.

To highlight clearly the impact that random alloy fluctuations have on the carrier distribution in the active region of an (In,Ga)N-based LED, we use as a reference point a virtual crystal approximation (VCA) which effectively can be described by a 1-D model. The benefit of this is twofold. Firstly, this enables us to compare directly the outcomes of our quantum corrected model with results from 1-D commercial software simulations; commercial software packages often employ a standard Schrödinger-Poisson-DD solver, which is numerically very costly and therefore unfeasible in large 3-D transport simulations. This motivates the need for an alternative implementation of quantum corrections. Secondly, and building on this benchmark, alloy fluctuations can be included in the calculations, revealing clearly their impact on the results. Our studies show, and when using the same input parameter set, only the model accounting for random alloy fluctuations produces trends that are consistent with the experimental data. The widely employed VCA yields results that are in contradiction with the experimental data, thus indicating that radiative recombination stems mainly from the well *furthest away* from the *p*-side. Overall, this highlights (i) that alloy fluctuations are essential to achieve an accurate description of the carrier transport and (ii) have to be taken into account when theoretically guiding the design of energy efficient III-N light emitters.

The paper is organized as follows: in Section B.2 we outline the model structure used for calculations and briefly summarise some of the literature experimental data from Ref. [145]. The theoretical framework which we use is summarized in Section B.3. Our results are discussed in Section B.4. Finally Section B.5 presents our conclusions.

B.2 Model MQW structures and literature experimental findings

To investigate the carrier distribution in (In,Ga)N/GaN MQW systems we proceed similar to experimental studies in the literature [145, 199] and target MQW systems where one of the wells in the MQW stack has a slightly higher In content compared to the remaining wells. In our case, we study MQW systems with three (In,Ga)N/GaN wells. Here two are $In_{0.1}Ga_{0.9}N$ ("shallow") wells and one is an $In_{0.125}Ga_{0.875}N$ ("deep") QW. These QWs are 3 nm wide and separated by 5 nm GaN barriers. The band edge profile of such a system along the transport (*c*-) direction, using a VCA, is shown in Fig. B.1 at a current density of 50 A/cm^2.

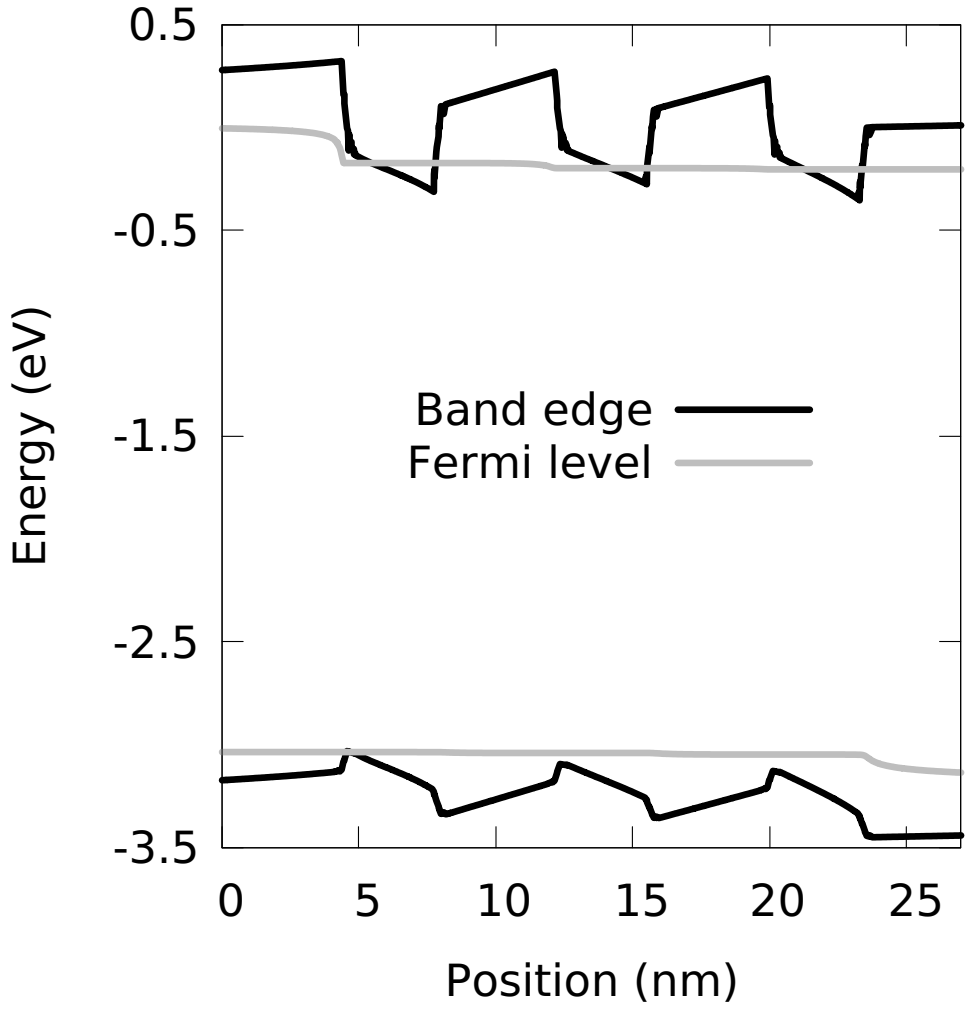


Figure B.1 – Conduction and valence band edges (black) along with the quasi-Fermi energies for electrons and holes (grey) in an (In,Ga)N/GaN multi-quantum well system described in virtual crystal approximation. The band edge profile and the quasi Fermi levels are shown at a current density of 50 A/cm^2. The leftmost (In,Ga)N quantum well contains 12.5% indium while the other two (In,Ga)N wells (centre and right) contain 10% indium.

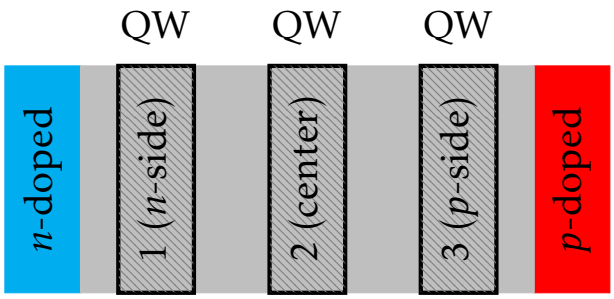


Figure B.2 – Schematic illustration of multi-quantum well system. The n-doped region is shown in cyan, the p-doped is in red and undoped regions are in grey. The quantum wells are numbered starting from the n-side.

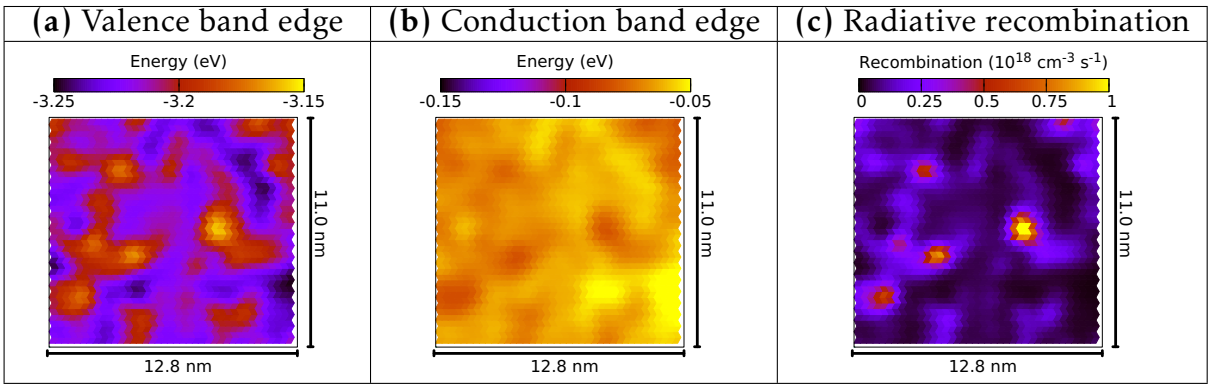


Figure B.3 – Profile of (a) valence band edge energy, (b) conduction band edge energy, and (c) radiative recombination rate in the growth plane (c-plane) of an $In_{0.1}Ga_{0.9}N$ quantum well; the current density is 50 A/cm^2 in all depicted figures. The slice displayed is the through the center well. The data are shown in all cases on a linear scale.

In the following, we investigate the carrier transport properties in two settings: (i) on an atomistic level accounting for random alloy fluctuations and (ii) in the frame of a VCA thus neglecting alloy fluctuations. In the latter VCA case (ii), at a given z-position (along the c-direction), there is no variation in material properties within the growth plane (c-plane). This assumption is also made in the widely used 1-D transport simulations on (In,Ga)N MQWs.

To study the carrier distribution in MQW systems using the simulation settings (i) and (ii), we follow again the experimental approach e.g. presented in Ref. [145] and the deep QW is moved from the n-side (position 1 (n-side) in Fig. B.2) to the p-side (position 3 (p-side) in Fig. B.2). In the case of the random alloy structures, the same microscopic configuration is kept for each well and only the ordering is changed.

For each of these systems the ratio of radiative recombination from the shallow wells to the deep well is calculated using:

$$\rho = \frac{\mathcal{R}^{RAD}_{\Omega_S}}{\mathcal{R}^{RAD}_{\Omega_D}} \tag{B.1}$$

where

$$\mathcal{R}^{RAD}_{\Omega_i} = \int_{\Omega_i} R^{RAD}(\mathbf{r}) dV \,, \tag{B.2}$$

is the total radiative recombination from the region Ω_i. Here, Ω_D is the region containing the deep QW, Ω_S is the region containing the shallow wells (as there are two shallow QWs this is the union of the two shallow QW regions). The radiative recombination rate at position $\mathbf{r}$, $R_{RAD}(\mathbf{r})$, is discussed in further detail in section B.3.2. Since we are studying a system with three QWs, an even distribution of carriers across the MQWs would result in a ratio of $\rho = 2$. Previous experimental work on a similar system by Galler *et. al.* [145] found that ρ *was small* (i.e. emission is dominated by the deep QW) *only when the deep well was closest to the p-doped side of the MQW system* (thus position 3 (p-side) in Fig.B.2). The authors conclude that holes are responsible for this behavior, and argue that they are mainly found in the p-side QW and not in wells further away from the p-side. As a consequence, the overall emission from the (In,Ga)N/GaN MQW system is dominated by the emission from this well closest to the p-doped region. In line with Ref. [145], we calculate ρ at a current density of 50 A/cm^2, which allows us to compare the here predicted trends with the trends found in the experimental studies. The theoretical framework employed to gain insight into ρ is discussed in the following section.

B.3 Theoretical framework

In this section, we introduce the underlying (microscopic) theory of our multiscale simulations. We start in Section B.3.1 with the electronic structure model, an atomistic tight-binding (TB) model, and discuss the drift-diffusion approach in Section B.3.2. Since all these ingredients have been discussed in detail in Refs. [221, 222], we here give only a brief summary.

B.3.1 Tight-binding energy landscape

In order to model the electronic structure of the above described (In,Ga)N MQW systems on an atomistic level, we employ a nearest-neighbour, sp^3 TB model [242]. Here, local strain and polarization effects are included using a valence force field model and local polarization theory, respectively [60].

To connect the TB model to a DD solver we proceed as follows. Firstly, the TB model is used to extract a potential energy landscape describing the MQW region of the device using an atomistic framework. To do so, at each atomic site in the three dimensional (3-D) supercell, a local TB Hamiltonian is constructed from the full TB Hamiltonian [75]. Subsequently, only the local TB Hamiltonian is diagonalized, yielding the conduction and valence band edge energy at each lattice site. These band edge energies include now already effects arising from alloy fluctuations and connected fluctuations in strain and built-in polarization field. The obtained 3-D confining energy landscape, after employing a Gaussian softening, forms the basis for our DD calculations. In previous studies we have investigated and discussed in detail the influence of the Gaussian softening on transport calculations for electrons and holes [221, 222]. Here, we choose a Gaussian broadening on the order of the GaN lattice constant, $\sigma_{c,v} = a^{\mathrm{GaN}} = 0.3189$ nm, in all calculations. This value is large enough to average over a number of neighboring sites, while also small enough to retain fluctuations in the energy landscape.

To obtain an accurate description of carrier transport in (In,Ga)N-based LEDs, the DD equations, which will be discussed below, are often coupled with solving the Schrödinger equation to account for quantum corrections. Such a Schrödinger-Poisson solver is widely available in commercial software packages. However, in these packages it is largely restricted

to 1-D simulations, since the extension to a 3-D system is computationally basically unfeasible. Instead of solving the large eigenvalue problem connected to evaluating the Schrödinger equation, we have implemented quantum corrections via localization landscape theory (LLT) [132]; this approach is numerically much more efficient and gives results similar to a full self-consistent Schrödinger-Poisson solver, as we will also discuss below. From LLT we extract an effective confining potential for the conduction and valence band edge starting from the TB energy landscape.

An example of the resulting quantum corrected energy landscape is given in Fig. B.3 (a) and (b). Here, in-plane band edge profiles for a single atomic plane through an $In_{0.1}Ga_{0.9}N$ QW, after LLT has been applied, are shown. As Fig. B.3 (a) reveals, the fluctuations in the valence band edge energy due to alloy fluctuations are of the order of 100 meV. In combination with the high effective hole mass, these fluctuations are large enough to give rise to strong carrier localization effects as seen in other studies already [260, 242, 102]. We therefore expect that, especially for holes, the inclusion of random alloy fluctuations in the simulation will impact the carrier distribution. Consequently recombination rates are also expected to be noticeably influenced.

The variation in the conduction band edge energy is significantly smaller (order of 30 meV), as can be seen in Figure B.3 (b). Since the effective electron mass is much lower in comparison with the holes, electron wave functions are less strongly perturbed by alloy fluctuations. The impact that these fluctuations in the band edge energies have on the radiative recombination is also seen in Fig. B.3 (c); the radiative recombination is calculated with `ddfermi` as will be described in section B.3.2. The correlation between the valence band edge maxima and regions of high radiative recombination can be clearly identified; similar spatial profiles can be seen for non-radiative (Auger) recombination (not shown).

In order to highlight the impact of random alloy fluctuations on carrier transport and the distribution of carriers across a MQW system, we compare our atomistic calculations with the outcome of a VCA. In the latter case a homogeneous effective crystal is constructed where material properties are chosen to be interpolated properties of the binaries InN and GaN within the QW region. Here a linear, composition weighted interpolation scheme is employed. A bandgap bowing of −2.0 eV is used, consistent with the underlying atomistic TB model [60]. The VCA description, without any Gaussian broadening, is similar to commercially available packages. However, and in contrast to commercial software packages, quantum corrections via LLT can also be taken into account in our VCA simulations, following the approach used for the random alloy case.

B.3.2 Device simulation

Having outlined above the generation of the energy landscape of the active region, e.g. the (In,Ga)N/GaN MQW system, a full device mesh, including the *n*- and *p*-doped regions, needs to be constructed on which the DD equations are solved. To achieve this we proceed as follows and divide the device mesh into two regions: an atomistic and a macroscopic one. The atomistic region is used to describe the MQW region and has as many grid points as atoms in the system. These points contain information about the conduction and valence band edge energies calculated from TB, as discussed above.

In order to capture the effects of carrier localization in the calculations, the in-plane dimensions of our 3-D simulation cell should be larger than the localization length of the holes, given that electrons are less strongly affected by alloy fluctuations [252]. In our atomistic calculations we use a system with in-plane dimensions of $12.8 \times 11.0\ \mathrm{nm}^2$. This is large enough to see the effects of hole localization as the in-plane hole localization length for $In_{0.1}Ga_{0.9}N$ QWs is of

the order of 1 nm [253]. The in-plane dimensions can be seen in Fig. B.3 (a) and (b) where the in-plane valence and conduction band edges of an $In_{0.1}Ga_{0.9}N$ QW are shown. In case of the VCA, given that there are no variations in material properties (band edge energies) within the growth plane (c-plane), a much smaller in-plane area is sufficient (1.3×1.1 nm^2), which reduces the numerical effort.

LLT is solved on this (finite element) mesh using a finite element method (FEM). The DD calculations are carried out employing a Voronoi finite volume method (FVM) [128]. Therefore, the generated FEM mesh must be transferred to an appropriate FVM mesh. Every point from the FEM mesh is included on the FVM mesh, as well as extra points required to produce a boundary-conforming Delaunay tetrahedral mesh; conduction and valence band data are then interpolated onto these additional nodes. This mesh is then embedded within a macroscopic device mesh which contains information about the p- and n-doped regions. In our atomistic transport studies here, we focus on systems without an (Al,Ga)N electron blocking layer (EBL). In principle, an atomistic mesh resolution would be required for the EBL, given the alloy fluctuations in (Al,Ga)N. However, and as we will discuss below, an (Al,Ga)N EBL is of secondary importance for the questions targeted in the present study. Therefore, outside active MQW region of the system, pure GaN is assumed. Thus, the conduction and valence band edge values are position independent (except for changes due an applied bias). The absence of strongly fluctuating band edges in the macroscopic mesh region allows us to use a sparse mesh and scale the simulation to a full device. The mesh is created using `TetGen` [247] and the interpolation is handled via `WIAS-pdelib` [141]. More details on the mesh generation can be found in Ref. [221].

Equipped with knowledge about the mesh generation, we turn now to the DD simulations. To do so we build on the van Roosbroeck system of equations [258]:

$$-\nabla \cdot (\varepsilon_s(\mathbf{r})\nabla\psi(\mathbf{r})) = q(p(\mathbf{r}) - n(\mathbf{r}) + C(\mathbf{r}))\,, \tag{B.3a}$$

$$\nabla \cdot \mathbf{j}_n(\mathbf{r}) = qR(\mathbf{r})\,, \tag{B.3b}$$

$$\nabla \cdot \mathbf{j}_p(\mathbf{r}) = -qR(\mathbf{r})\,, \tag{B.3c}$$

$$\mathbf{j}_n(\mathbf{r}) = -q\mu_n n(\mathbf{r})\nabla\varphi_n(\mathbf{r})\,, \tag{B.3d}$$

$$\mathbf{j}_p(\mathbf{r}) = -q\mu_p p(\mathbf{r})\nabla\varphi_p(\mathbf{r})\,. \tag{B.3e}$$

In the above equations, q is the elementary charge, $\varepsilon_s(\mathbf{r}) = \varepsilon_0\varepsilon_r(\mathbf{r})$ is the dielectric permittivity, $\psi(\mathbf{r})$ is the device electrostatic potential, $p(\mathbf{r})$ and $n(\mathbf{r})$ are the hole and electron densities, $C(\mathbf{r}) = N_D^+(\mathbf{r}) - N_A^+(\mathbf{r})$ is the net activated dopant density, $\mathbf{j}_n(\mathbf{r})$, $\mathbf{j}_p(\mathbf{r})$, $\varphi_n(\mathbf{r})$ and $\varphi_p(\mathbf{r})$ are the electron and hole current densities and the respective quasi-Fermi potentials. The total recombination rate is denoted by R.

The carrier densities are related to the band edge energies, $E_c^{\mathrm{dd}}(\mathbf{r})$ and $E_v^{\mathrm{dd}}(\mathbf{r})$, and quasi-Fermi potentials, $\varphi_n(\mathbf{r})$ and $\varphi_p(\mathbf{r})$, via the state equations [128]

$$n(\mathbf{r}) = N_c\mathcal{F}\left(\frac{q(\psi(\mathbf{r}) - \varphi_n(\mathbf{r})) - E_c^{dd}(\mathbf{r})}{k_BT}\right), \tag{B.4a}$$

$$p(\mathbf{r}) = N_v\mathcal{F}\left(\frac{E_v^{dd}(\mathbf{r}) - q(\psi(\mathbf{r}) - \varphi_p(\mathbf{r}))}{k_BT}\right). \tag{B.4b}$$

Here, N_c and N_v are the effective density of states for the conduction and valence band, respectively. For the distribution function, $\mathcal{F}$, we use Fermi-Dirac statistics, and k_B is the Boltzmann constant. A temperature of $T = 300$ K has been used in all calculations. It is to

note that the valence band, $E_v^{dd}(\mathbf{r})$, and conduction band edge energy, $E_c^{dd}(\mathbf{r})$, can be described either by a VCA, a VCA including quantum corrections via LLT, or an atomistic random alloy calculation including LLT-based quantum corrections.

The total recombination rate in Eqs. (B.3b) and (B.3c), R, is calculated using the ABC model [178, 226]. Here, R is the sum of (defect related) Shockley-Read-Hall, R^{SRH}, radiative, R^{RAD} and (non-radiative) Auger recombination rate, R^{AUG}. The SRH rate is obtained from:

$$R^{SRH}(\mathbf{r}) = \frac{r(n,p)}{\tau_p\Big(n(\mathbf{r}) + n_i(\mathbf{r})\Big) + \tau_n\Big(p(\mathbf{r}) + n_i(\mathbf{r})\Big)}\,, \tag{B.5}$$

the radiative part via

$$R^{RAD}(\mathbf{r}) = B_0 r(n,p)\,, \tag{B.6}$$

and the Auger rate is calculated as

$$R^{AUG}(\mathbf{r}) = (C_n n(\mathbf{r}) + C_p p(\mathbf{r}))r(n,p)\,. \tag{B.7}$$

In Eqs. (B.5) to (B.7)

$$r(n,p) = n(\mathbf{r})p(\mathbf{r}) - n_i^2(\mathbf{r})$$

and

$$n_i^2(\mathbf{r}) = n(\mathbf{r})p(\mathbf{r})\exp\left(\frac{q\varphi_n - q\varphi_p}{k_B T}\right).$$

The above equations require further input, namely the radiative recombination coefficient B_0, the Auger recombination coefficients C_p and C_n as well as the the SRH lifetimes τ_p and τ_n. All these parameters will in principle carry a composition dependence [210, 211, 181]. Furthermore, B_0, C_p and C_n will also be carrier density dependent [93, 94, 169]. We follow here the widely made assumption that these coefficients are constant across the InGaN MQW region [194, 203]. In the following we take a weighted average of parameters calculated in Ref. [210] for an electron and hole density of 3.8×10^{18} cm^{-3}, which is a good approximation for the average carrier densities in the QWs at a current density of 50 A/cm^2. As our active region consists of two $In_{0.1}Ga_{0.9}N$ QWs and one $In_{0.125}Ga_{0.875}N$ QW we evaluate the different recombination coefficients as follows:

$$R_i^{\text{eff}} = \frac{2 \times (R_i^{10\%}) + 0.5 \times (R_i^{15\%} + R_i^{10\%})}{3}\,. \tag{B.8}$$

Here, $R_i \in \{B_0, C_n, C_p\}$ are the radiative recombination, electron-electron-hole and hole-hole-electron Auger recombination coefficients, respectively. As there are no values for an $In_{0.125}Ga_{0.875}N$ QW in Ref. [210], a linear average of the coefficients in $In_{0.1}Ga_{0.9}N$ and $In_{0.15}Ga_{0.85}N$ wells has been used. A summary of the material parameters employed in all simulations is given in Table B.1.

The numerical approximation of the van Roosbroeck system is implemented (in 3-D) in `ddfermi` [104]. As already mentioned above, we employ the FVM and the current is discretized using the SEDAN (excess chemical potential) approach [265, 50, 2], which yields a thermodynamically consistent flux approximation in the sense of Ref. [128].

To simulate the devices under study, we also used the commercial software next**nano** [30], which relies on the simulation of a self-consistent Schrödinger-Poisson-DD system. In this work we use next**nano** to simulate the carrier transport in the above discussed MQW systems within a 1-D approximation. In next**nano** we utilize the same parameter set as in the `ddfermi` simulations. Therefore, the obtained results can be directly compared to our 3-D VCA model.

Table B.1 – Material parameters used in the different regions of the simulation supercell. Parameters denoted with † are taken from [194]; parameters denoted with ‡ are derived from [210] as described in the main text.

Parameter			Value in each region		
Name		**Units**	***p*-GaN**	***i*-InGaN**	***n*-GaN**
Doping		cm^{-3}	5×10^{18}	1×10^{16}	5×10^{18}
μ_h	†	cm^2/Vs	5	10	23
μ_e	†	cm^2/Vs	32	300	200
τ_p	†	s	10	1×10^{-7}	7×10^{-10}
τ_n	†	s	6×10^{-10}	1×10^{-7}	10
B_0	‡	cm^3/s	2.8×10^{-11}	2.8×10^{-11}	2.8×10^{-11}
C_p	‡	cm^6/s	5.7×10^{-30}	5.7×10^{-30}	5.7×10^{-30}
C_n	‡	cm^6/s	1×10^{-31}	1×10^{-31}	1×10^{-31}

When including quantum corrections in `ddfermi`, LLT is used. In next**nano** a self-consistent Schrödinger-Poisson-DD calculation is performed where a $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian is used to calculate eigenstates across the full simulation domain. Following the `ddfermi` set up, in next**nano** we employ also a 1-band model for the calculation of the electron and hole densities.

B.4 Results

In this section we present the results of our study on the carrier distribution in the above described (In,Ga)N/GaN MQW systems. To understand the impact of the alloy microstructure on the carrier distribution, in Section B.4.1 we start with standard 1-D calculations building on the commercial software package next**nano** [30]. We use this entirely continuum-based description of the QWs also to determine the impact (i) of an EBL and (ii) a self-consistent Schrodinger-Poisson-DD treatment on the transport properties. Moreover, and as already mentioned above, (ii) can also be used as a benchmark for our 3-D `ddfermi` solver. In Section B.4.2 we then proceed to investigate the influence of random alloy fluctuations on the carrier distribution in the (In,Ga)N/GaN MQW stack.

B.4.1 Continuum-based simulations of the carrier transport in (In,Ga)N-based LEDs

To examine the impact of random alloy fluctuations on the carrier distribution in an (In,Ga)N/GaN MQW stack, we start with a 'standard' 1-D simulation approach that is widely applied in the literature. In a first step we begin with next**nano** calculations and as outlined above, compare the results to our `ddfermi` data.

nextnano simulations

To study how the presence of an EBL affects the ratio of radiative recombination ρ, Eq. (B.1), the systems outlined in Section B.2 are simulated with and without a 20 nm $Al_{0.15}Ga_{0.85}N$ EBL using next**nano**. The EBL is separated from the *p*-side QW (position 3 (*p*-side) in Fig. B.2) by a 10 nm GaN barrier. Similar settings for an (Al,Ga)N EBL have been used in previous studies [194]. The next**nano** calculated ratio of radiative recombination ρ, when varying the position of the deep QW in the MQW stack, are depicted in Fig. B.4 (a). Turning first to the data without

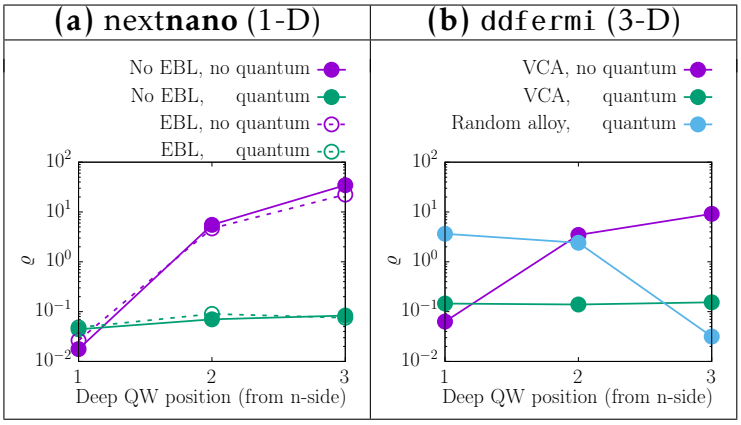


Figure B.4 – Ratio of radiative recombination ρ, Eq. (B.1), from the shallow wells ($In_{0.1}Ga_{0.9}N$) to recombination from the deep well ($In_{0.1}Ga_{0.9}N$) calculated as a function of the position of the deep well in the multi-quantum well stack. Here ρ is evaluated using (a) next**nano** excluding (purple) and including (green) quantum corrections via a self-consistent Schrödinger-Poisson-drift diffusion solver; results are shown when excluding (solid, filled circles) and including (dotted, open circles) an $Al_{0.15}Ga_{0.85}N$ blocking layer, and (b) `ddfermi` excluding (purple), including (green) quantum corrections via localization landscape theory (LLT) using a virtual crystal approximation (VCA) and a random alloy calculation including LLT-based quantum corrections (blue); these calculations neglect the AlGaN blocking layer.

quantum corrections, we find that in the case of the employed 1-D VCA-like continuum-based description, *ρ is small when the deep QW is at the n-side (position 1 (n-side) Fig. B.2) and larger when the deep well is at the p-side (position 3 (p-side) Fig. B.2)*. Thus, the 1-D model *predicts the opposite trend* when compared to experiment [145]. This trend is only slightly changed when including quantum corrections via a self-consistent Schrödinger-Poisson-DD model. In this case a much weaker dependence of the results on the position of the deep QW in the MQW stack is observed. However, even when including quantum corrections, the next**nano** results for ρ are not reflecting the experimentally observed behavior (see discussion above). Figure B.4 (a) reveals also that qualitatively the results do not depend on the presence of the EBL, indicating that for the structures considered, this feature of an LED is of secondary importance for the aims of this work.

`ddfermi` simulations

Since we are also able to use the atomistic framework in a VCA setting, we compare our `ddfermi` results, cf. Fig. B.4 (b) (purple), with those from next**nano**, cf. Fig. B.4 (a) (purple, solid). We focus on structures which neglect the EBL as we have found above that it does not impact results in a VCA. In both next**nano** and `ddfermi` a similar trend is found: the deep QW dominates recombination only when it is located at the *n*-side. This is illustrated further in Fig. B.5 (a), which displays the contribution (in percent) to the radiative recombination rate from each QW (colors) in the MQW stack. The data are shown as a function of position of the deep QW in the MQW system. This confirms that it is always the QW which is closest to the *n*-doped side (position 1) that dominates the recombination process; the *n*-side QW contributes $\approx 95\%$ when the deep QW is at position 1, $\approx 70\%$ when the deep QW is at position 2 or 3. Again, we stress that this is the *opposite* trend to the experimental findings in Ref. [145].

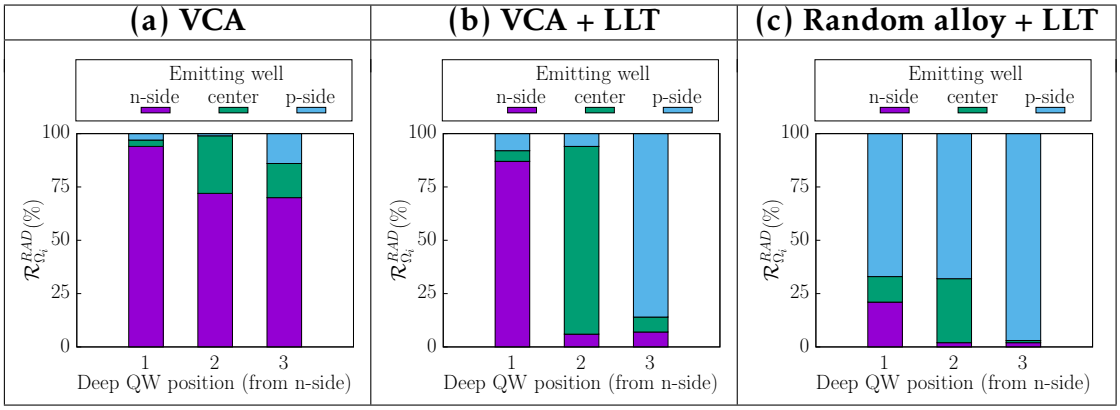


Figure B.5 – Contribution of each quantum well (n-side; centre; p-side) in the (In,Ga)N multi-quantum well system to the total radiative recombination $\mathcal{R}^{RAD}_{\Omega_i}$ for $i \in \{n\text{-side}, \text{center}, p\text{-side}\}$ as a percentage of the total radiative recombination from all 3 quantum wells for (a) virtual crystal approximation (VCA), (b) virtual crystal approximation with quantum corrections included via localization landscape theory (VCA + LLT) and (c) a random alloy calculation including localization landscape theory based quantum corrections (Random alloy + LLT). That data are shown as a function of the position of the deep quantum well (x-axis). Each bar contains the percentage recombination from the n-side quantum well (purple), the center quantum well (green) and the p-side quantum well (blue). Labelling is consistent with that introduced in Fig. B.2.

To shed more light on this result, the upper row in Figure B.6 depicts the average hole (black, solid), electron (black, dashed) and radiative recombination (red) rate along the c-axis when the deep QW ($In_{0.125}Ga_{0.875}N$ well) is (a) closest to the n-side (position 1), (b) in the centre of the MQW stack (position 2) and (c) closest to the p-side (position 3). Focusing on the VCA data, Figs. B.6 (i) (a-c), we see the cause of the dominant recombination from the n-side QW: the hole density is always high in this region, independent of which well is closest to the n-side. In particular, the p-side QW fails to capture holes effectively and consistently has the lowest hole density. We note that a similar behavior is also found in the next**nano** calculations discussed in Sec. B.4.1.

Given that our VCA `ddfermi` approach and next**nano** treat (In,Ga)N as a homogeneous alloy that can be described by averaged material parameters which do not vary throughout the wells (no alloy fluctuations included), it allows us also to compare the implemented methods for quantum corrections in DD simulations. Here, as discussed above, next**nano** builds on the widely used Schrödinger-Poisson-DD model while `ddfermi` utilizes the recently developed LLT method. It has been discussed and shown in the literature that the LLT method can produce results in good agreement with the solution of the Schrödinger equation in the case of a 1-D effective mass approximation [132, 74]. Looking at Fig. B.4 (a) (green, solid) and Fig. B.4 (b) (green) we see that the results from our in-house developed `ddfermi`-based 3-D model, which employs LLT (3-D, `ddfermi`), are very similar to the standard self-consistent 1-D Schrödinger-Poisson-DD calculation underlying next**nano**. This gives confidence that our LLT treatment is providing a comparable description of the quantum corrections in the system.

Overall, Fig. B.4 (a) reveals that when including quantum corrections in the VCA calculations, the position of the deep QW has little impact on the ratio of the relative radiative recombination, ρ. From Fig. B.5 (b) one can also gain more insight into this behavior and how quantum

corrections impact the carrier distribution in the MQW stack. In the *absence* of quantum corrections but utilizing a VCA, Fig. B.5 (a), the well closest to the *n*-side dominates the relative radiative recombination ratio ρ independent of the position of the deep well in MQW systems. When *including* quantum corrections this situation is now changed: the deep QW is now the dominant emitter independent of its position in the MQW stack.

This behavior becomes clear when analyzing the electron and hole densities as a function of the position of the deep well in the (In,Ga)N/GaN MQWs, as shown in Fig. B.6 (ii). Looking at the electron densities first, we find that electrons preferentially occupy the well closest to the *p*-side. This effect is enhanced when the deep QW is closest to the *p*-side (cf. Fig. B.6 (ii) (c)). In our previous study on uni-polar electron transport [221], we have already seen that including quantum corrections leads to a softening of the potential barrier at the QW barrier interfaces. This in turn can lead to an increased electron current at a fixed bias point, when compared to a VCA system without LLT treatment, and thus the electrons can more easily 'overshoot' the wells in the MQW system. As a consequence, a lower electron density in the well closest to the *n*-side is observed. Turning to the hole density, the situation is different. Here, we find that holes preferentially populate the well closest to the *n*-side. Only when the deep QW is closest to the *p*-side, the hole density in this well is noticeably increased. However, when comparing the distribution of holes in the MQW as a function of the position of deep well in absence (Fig. B.6 (i)) and presence (Fig. B.6 (ii)) of quantum corrections, the results are not very different. This indicates that quantum corrections, at least when employing a VCA, are of secondary importance for the hole distribution. This finding is consistent with our previous results on uni-polar hole transport [222], where we have discussed that due to the high effective hole mass and the small valence band offset, quantum corrections have a smaller impact on the hole transport when compared to electrons. As a consequence, the distribution of holes follows a similar pattern to that of the VCA where quantum corrections are neglected. Finally, when looking at the ratio of radiative recombination ρ, it is important to note that this quantity is not only determined by having both large electron and hole densities in the same well but also by their spatial overlap. As one can infer from Fig. B.6 (ii) (a-c), the largest radiative recombination rate is always observed in the deepest well. This indicates also that the spatial overlap of electron and hole densities largest in the deep QW regardless of its position across the MQW system. We stress again that even when including quantum corrections in the VCA calculations, the resulting trend in ρ is not reflecting the trend observed in experimental studies [145].

B.4.2 Impact of a random alloy fluctuations on the carrier transport in (In,Ga)N/GaN MQWs

In the last step, we move away from the VCA description of the system and include, in addition to quantum corrections, also random alloy fluctuations in the calculations. Figure B.4 (b) (blue) shows that, and this time in line with the experimental results by Galler *et. al* [145], the deep QW only contributes significantly to the radiative recombination when it is *closest to the p-side* (position 3). In fact, when including random alloy fluctuations in the calculations, the well closest to the *p*-side always has the largest contribution to total radiative recombination, as can be seen in Fig. B.5 (c).

To understand this behavior, Fig. B.6 (iii) depicts the electron and hole densities in the different wells as a function of the position of the deep well in the MQW systems. Looking at the electron density first, in comparison to the VCA calculations both including and excluding quantum corrections, random alloy fluctuations lead a reduction in electron density at the *n*-side QW. As discussed above and previously, quantum corrections can lead to increased electron transport, and including alloy fluctuations adds further to this effect due to the softening of the

barrier at the well interfaces [221]. As a consequence, the electrons can more easily 'overshoot' the wells in the MQWs, which can also be seen in the increased electron density beyond the *p*-side QW when alloy fluctuations and quantum corrections are included. However, in comparison to the VCA result including quantum corrections, the electron density in the *p*-side well is only slightly affected by alloy fluctuations.

In contrast, hole densities in the *p*-side QW are more dramatically changed by alloy fluctuations. As Figs. B.6 (ii) and B.6 (iii) show, in comparison to the VCA description, alloy fluctuations lead to an increase in the hole density in the *p-side QW* (position 3) even when the deep QW is closest to the *n*-side (position 1) or in the centre (position 2) of the MQW system. While the smoothing of the well barrier interface can increase hole transport, as in the case of electrons, there are now also alloy disorder induced localization effects to contend with. As discussed in our previous work, these localization effects are *detrimental* to hole transport [222] and result in an increased hole density in the *p*-side QW. As a consequence, the well closest to the *p*-side dominates radiative recombination

We note that there is still a reasonable hole density present in the *n*-side QW (Fig. B.6 (iii) (a-c)). In general, the distribution of carriers will also depend on the GaN barrier width and a 5 nm barrier is narrow enough to allow for some hole transport across the MQW [199]; a similar dependence of hole transmission on the barrier width has been seen in previous non-equilibrium Green's function studies [220]. Thus we expect that increasing the barrier width will mainly lead to a reduction of the hole density in the well furthest away from the *p*-side, but should to a lesser extent affect the hole density in the well closest to the *p*-side. Therefore, even for a larger barrier width than the here considered 5 nm, we expect that the recombination will still be dominated by the *p*-side QW.

We note that based on the VCA results we did not consider the EBL in the atomistic calculations. In general the EBL needs to be treated with an atomistic resolution. Previous studies of (Al,Ga)N barriers in uni-polar device settings have found that the impact of these barriers is lower than what is expected from a 1-D simulation for both electrons [45] and holes [231]. Thus given that our VCA calculations show that the presence of the EBL is of secondary importance for our study, we expect a similar conclusion when treating the EBL with alloy fluctuations. Therefore, it is unlikely that the EBL impacts the here presented result, however this question may be targeted in future studies.

B.5 Conclusions

In this work we apply a 3-D quantum corrected multiscale simulation framework to gain insight into the impact of random alloy fluctuations on the electron and hole distribution across the active region of an (In,Ga)N/GaN LED. To study the spatial distribution of carriers we have followed literature experimental studies [145] and analyzed the radiative recombination ratio in a multi-quantum well system, where one of the wells in the system has a higher indium content (deeper well) and its position is varied within the stack.

The here considered MQW systems are not only of interest for a comparison with experiment, they provide also the ideal opportunity to benchmark and validate results from our in-house developed 3-D multiscale simulation framework against commercially available software packages. To do so we treat the QWs in a virtual crystal approximation (VCA), to mimic the 1-D simulation widely used in the literature for (In,Ga)N QWs and implemented in the commercial software package next**nano**. In addition, this study allows us also to compare the different schemes to account for quantum corrections (localization landscape theory vs. Schrödinger-Poisson-DD simulations) in the simulations. Overall, this analysis showed very good agreement between

results obtained from our in-house software and next**nano** (without random alloy fluctuation).

Equipped with this benchmarked model, our analysis reveals that including (random) alloy fluctuations in the calculations is vital for reproducing trends seen in experiment. More specifically, when using the widely employed virtual crystal approximation (VCA), the hole density in the well closest to the *p*-doped region of the device is significantly reduced compared to our atomistic random alloy calculation. As a consequence, and in contrast to the experiment, in VCA the well closest to the *p*-side contributes very little to the radiative recombination process, an effect that can be reduced by accounting for quantum corrections. While this leads to enhanced radiative recombination from the well closest to the *p*-side, at least when this well is the deep well, it still does not reflect the trends observed in the experimental studies. However, when including random alloy fluctuations and quantum corrections in our 3-D simulation framework, these effects lead to an increase in the hole density in the well closest to the *p*-side. Consequently, this well dominates the radiative recombination process in line with the experimental data. We note that in addition to quantum corrections and alloy fluctuations no further ingredients are required (e.g. multi-population model) to explain the experimentally observed trends. Therefore, our calculations highlight that alloy fluctuations are a key ingredient in simulations guiding the design of III-N based devices. Thus, the here developed model presents an ideal starting point for future calculations.

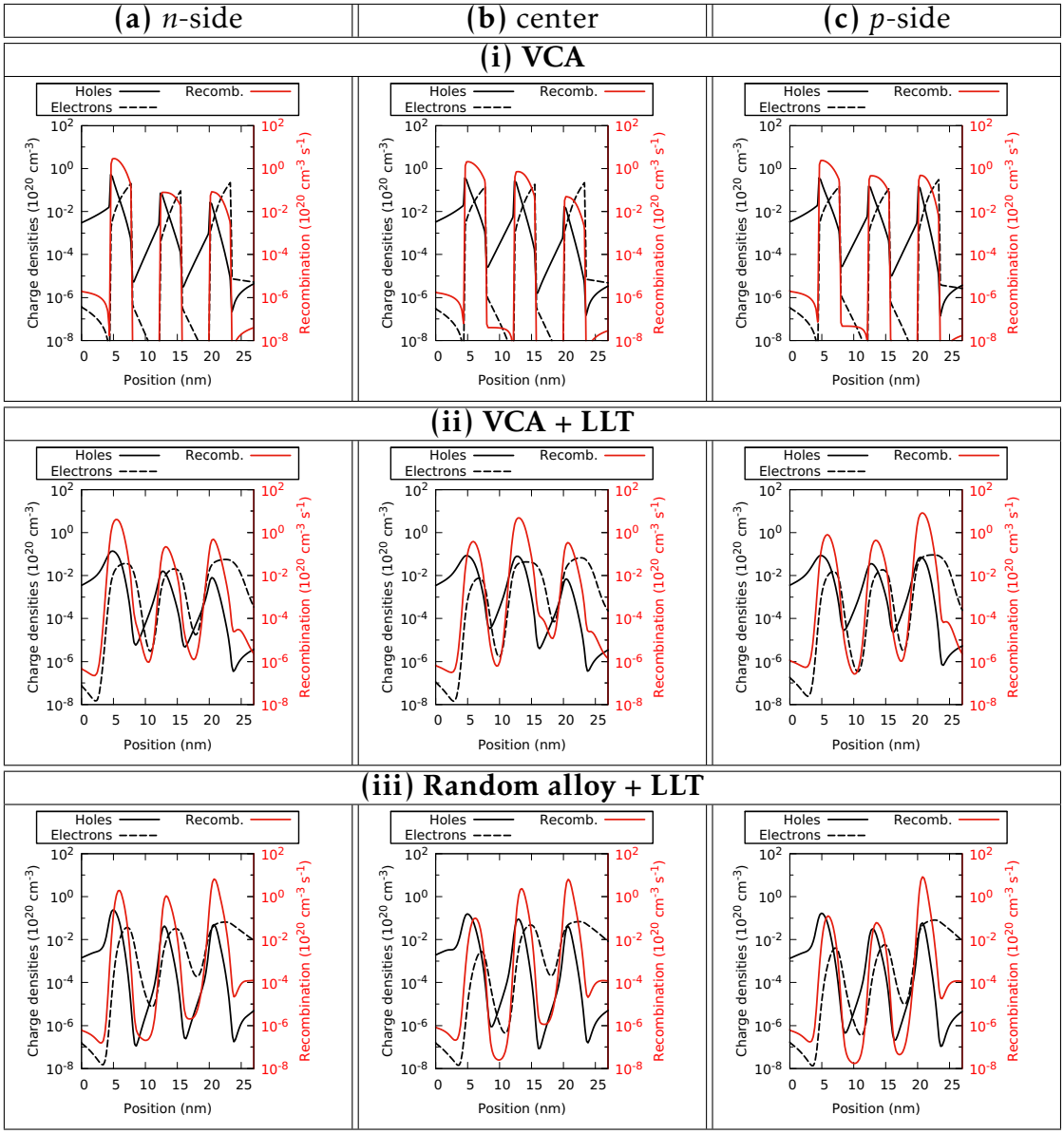


Figure B.6 – Hole density (black, solid), electron density (black, dashed), and radiative recombination rate (red, solid) averaged over each atomic plane along the transport direction. Results from calculations building on (i) a virtual crystal approximation (top), (ii) a virtual crystal including quantum corrections via localization landscape theory (LLT) (center) and a (iii) random alloy description including LLT-based quantum corrections (bottom); the deep well is located at (a) the *n*-side (left), (b) the center (middle) and (c) the *p*-side (right). The data are shown on a log scale.

Appendix C

Importance of satisfying thermodynamic consistency in optoelectronic device simulations for high carrier densities

Outline of the current chapter

This Appendix is a submitted [126] work, in collaboration with Patricio Farrell, Michael O'Donovan, Stefan Schulz and Thomas Koprucki.

We show the importance of using a thermodynamically consistent flux discretization when describing drift-diffusion processes within light emitting diode simulations. Using the classical Scharfetter-Gummel scheme with Fermi-Dirac statistics is an example of such an inconsistent scheme. In this case, for an (In,Ga)N multi quantum well device, the Fermi levels show an unphysical hump within the quantum well regions. This result originates from neglecting diffusion enhancement associated with Fermi-Dirac statistics in the numerical flux approximation. For a thermodynamically consistent scheme, such as the SEDAN scheme, the humps in the Fermi levels disappear. We show that thermodynamic inconsistency has

far reaching implications on the current-voltage curves and recombination rates.

C.1 Introduction

In recent years drift-diffusion simulations have offered a numerically attainable method for studying carrier dynamics in a wide variety of devices including light-emitting diodes (LEDs) [237, 225, 219], transistors [251, 91] and solar cells [232, 255]. The physical interpretation of the model is quite straightforward: carriers in a device tend to diffuse from regions of high carrier density to low, and have drift motion due to an applied force such as an electric field in the device. On the other hand, the numerical implementation of the model can carry pitfalls which lead to an incorrect description of the device behaviour as we will highlight below in detail.

The purpose of this publication is therefore to show the practical unphysical implications of disobeying thermodynamic consistency. In the stationary case, thermodynamic consistency for discretized drift-diffusion equations can be defined by the demand that the zero bias solution coincides with the thermodynamic equilibrium. In the transient case, it is closely related to the fact that, for boundary conditions compatible with the thermodynamic equilibrium, the solution converges to this equilibrium when time tends to infinity. It is already known that disobeying this property causes non-physical dissipation in the steady state, see [26]. However, the inconsistent discrete approximation of the numerical fluxes has also more direct consequences for the quasi Fermi potentials and is thus important for accurately describing the physics of a device in the frame of drift-diffusion simulations. If one inconsistently approximates the fluxes the quasi Fermi level will show a completely wrong behavior in for instance quantum well regions, which are at the heart of modern LEDs. This has a knock-on effect for the description of the carrier densities and thus also recombination and current-voltage (IV) curves, which we will illustrate with an (In,Ga)N quantum well structure, a material system of strong interest for energy efficient solid state lighting [158] and for which considerable effort has been undertaken to develop advanced carrier transport models [221].

For the Boltzmann distribution, the classical Scharfetter-Gummel scheme [239] presents such a thermodynamically consistent scheme. Strictly monotonically increasing non-Boltzmann distribution functions lead to *diffusion enhancement*. Various extensions of the Scharfetter-Gummel scheme have been suggested to account for this effect, see [228, 171, 249]. Unfortunately, these schemes are not thermodynamically consistent. In [184, 185] a thermodynamically consistent generalization for Blakemore statistics (which is itself a special case of [120]) is presented in the spirit of [239] by solving local Dirichlet problems. But this generalization requires solving local nonlinear equations during assembly and the iterative solution of the coupled system. It is therefore computationally prohibitively expensive. A computationally more affordable approach is presented in [67]. On the other hand, in [26] the author presents another extension of the Scharfetter-Gummel scheme using a proper average of the nonlinear diffusion guaranteeing thermodynamic consistency for a specific choice of the distribution function. An alternative interpretation of this approach based on averaging the diffusion enhancement for a very general class of statistical distribution functions was given in [186]. Finally, the SEDAN scheme [265, 50, 2] includes the nonlinearity in the drift instead of the diffusion part of the flux and thus also yields a thermodynamically consistent scheme.

The remainder of this paper is organized as follows: In Section C.2, we describe the bipolar drift-diffusion model for charge transport in semiconductors. Its finite volume discretization including the flux discretizations is described in Section C.3. The formal definition of discrete

thermodynamic consistency is also presented in this section. We compare thermodynamically consistent and inconsistent schemes in Section C.4 by studying the distribution of densities and quasi Fermi levels within an (In,Ga)N quantum well (QW) system, which is embedded in a *p-i-n* junction. Finally, we conclude in Section C.5.

C.2 Drift-diffusion equations and diffusion enhancement

We briefly introduce a model based on nonlinear partial differential equations which describes bipolar charge transport in a semiconductor. More details can be found in [128]. The dependence of the carrier densities n and p on the chemical potentials for electrons and holes η_n and η_p are described by a statistical distribution function $\mathcal{F}$ as well as conduction and valence band densities of states N_c and N_v via the state equations $n = N_c\mathcal{F}(\eta_n)$ and $p = N_v\mathcal{F}(\eta_p)$. Typical choices for the distribution function are $\mathcal{F}(\eta) = \exp(\eta)$, the so-called Boltzmann approximation, or $\mathcal{F}(\eta) = F_{1/2}(\eta) = \frac{2}{\sqrt{\pi}}\int_0^\infty \frac{E^{1/2}}{e^{E-\eta}+1}\,dE$, namely the Fermi-Dirac integral of order 1/2 describing degenerate semiconductors.

The chemical potentials are related to the quasi-Fermi potentials of electrons and holes φ_n and φ_p via

$$\eta_n = (q(\psi - \varphi_n) - E_c)/(k_B T) \text{ and } \eta_p = (q(\varphi_p - \psi) + E_v)/(k_B T).$$

Here q denotes the elementary charge, ψ the electrostatic potential, k_B the Boltzmann constant, T the temperature and E_c and E_v the conduction and valence band-edge energies. This model assumes that charge carriers behave as if they are in a bulk material, described by a 3-D density of states. In a slowly varying potential this is a valid description, however in a quantum well system the abrupt interface requires a more advanced treatment. To account for the quantum mechanical nature of electrons and holes one option is to solve the Schrödinger equation for the confining potential energy formed by $V_{c,v} = E_{c,v} - q\psi$. This is a numerically demanding approach, as the Schrödinger equation is an eigenvalue problem which would need to be solved self-consistently coupled to Poisson and drift-diffusion equations. While such a calculation is feasible in 1D, extending this to 2D or 3D structures is numerically prohibitive. Therefore, in recent years significant efforts have been undertaken to establish methods that account for quantum corrections but are computationally cheaper [129, 225]. One of these approaches is based on the so-called localization landscape theory (LLT), which allows to extract a (non-local) effective potential. It has been shown that LLT provides a good approximation of the single particle ground states, not only in square wells but also triangular wells which are relevant for systems with a polarization field – such as (In,Ga)N QWs [74]. The resulting (effective) confining potentials, $E_{c,v}^{\text{eff}}$, exhibit band edges that are softened and approximate the finite extent of carrier wavefunctions [1]. In the framework of LLT a linear system of equations is solved to determine the effective potential without introducing extra free parameters [132]. Our implementation of this method in conjunction with a drift-diffusion based carrier transport solver is discussed in more detail in [219] and [221]. In [219] we have compared results of our quantum corrected drift-diffusion model employing LLT with the results of a commercially available software package utilizing a 'standard' Schrödinger-Poisson solver. Our findings indicate that LLT can produce results in good agreement with the fully coupled Schrödinger-Poisson-drift-diffusion solver, highlighting that LLT captures quantum mechanical corrections sufficiently. In recent years this method has been used to study transport behaviour of numerous semi-conductor structures include LEDs [202, 193, 203, 236], blocking layers [231] and superlattices [256]; moreover, these

1. From here on within this work $E_{c,v}$ shall refer to the band edge values which are modified to account for the effective potential, $E_{c,v} \equiv E_{c,v}^{\text{eff}}$.

methods have been used successfully alongside experimental studies to gain insight into device behaviour [203, 236].

We model a bipolar semiconductor device as a domain $\Omega \subset \mathbb{R}^d$ where the carrier transport in a self-consistent electrical field is described by a system of partial differential equations. In the steady-state case this drift-diffusion system consists of Poisson's equation for ψ and continuity equations for electrons and holes:

$$-\nabla \cdot (\varepsilon_0 \varepsilon_r \nabla \psi) = q(C + p - n), \quad \vec{x} \in \Omega, \tag{C.1}$$

$$-\nabla \cdot \mathbf{j}_n = -qR, \qquad \nabla \cdot \mathbf{j}_p = -qR, \quad \vec{x} \in \Omega. \tag{C.2}$$

Here, ε_r is the relative permittivity, C is the net doping profile, and $R = R(n,p)$ describes carrier recombination. Electron and hole current densities can be expressed in terms of quasi-Fermi potentials by

$$\mathbf{j}_n = -q\mu_n n \nabla \varphi_n, \quad \mathbf{j}_p = -q\mu_p p \nabla \varphi_p, \tag{C.3}$$

or for any strictly monotonic Fermi-like distribution function $\mathcal{F}(\eta)$ in drift-diffusion form

$$\mathbf{j}_n = q\mu_n \left[U_T g\left(\frac{n}{N_c}\right) \nabla n - n \nabla \left(\psi - \frac{E_c}{q}\right)\right]$$

$$\text{and } \mathbf{j}_p = -q\mu_p \left[U_T g\left(\frac{p}{N_v}\right) \nabla p + p \nabla \left(\psi + \frac{E_v}{q}\right)\right], \tag{C.4}$$

where μ_n and μ_p denote the electron and hole mobilities, respectively, and $U_T = k_B T/q$ is the thermal voltage. The factor g can be defined in terms of densities, $g(x) = x(\mathcal{F}^{-1})'(x)$, for $x \in \mathbb{R}$. This factor is the so-called *diffusion enhancement* appearing as a density-dependent modification factor in the generalized Einstein relation, see [257], leading in general to a non-linear diffusion coefficient. For the Boltzmann distribution, $\mathcal{F}(\eta) = \exp(\eta)$, we have $g \equiv 1$ and the current expressions (C.4) reduce to the usual ones with linear diffusion.

C.3 Finite volume space discretization

We discretize the domain Ω using the Voronoï box based finite volume method introduced in [204], also known as "box method" due to [15]. It uses a simplical boundary conforming Delaunay grid ([248]) which allows to obtain control volumes surrounding each given collocation point $\mathbf{x}_K$ by joining the circumcenters of the simplices containing it, see [128] for details.

Let ∂K denote the boundary of the control volume K, and $|\xi|$ the measure of a geometrical object ξ. For each control volume K, we integrate the continuity equation (C.2) and apply Gauss's theorem to the integral of the flux divergence. Restricting our considerations to the electron transport equation, we obtain

$$\begin{aligned} 0 &= \int_{\partial K} \mathbf{j}_n \cdot \mathbf{n} ds - \int_K qR \, d\mathbf{x} = \sum_{L \text{ neighbor of } K} \int_{\partial K \cap \partial L} \mathbf{j}_n \cdot \mathbf{n}_{KL} ds - \int_K qR \, d\mathbf{x} \\ &\approx \sum_{L \text{ neighbor of } K} |\partial K \cap \partial L| j_{n,KL} - q|K| R(n_K, p_K), \end{aligned} \tag{C.5}$$

where $\mathbf{n}$ is the internal unit normal to ∂K and $\mathbf{n}_{KL}$ is the internal unit normal to the interface

$\partial K \cap \partial L$ for each neighbor L of K. The values n_K, p_K are the numerical approximations of the densities n and p at the collocation points x_K, and $j_{n,KL}$ are approximations of the normal currents through $\partial K \cap \partial L$. In the same manner the discretization of the Poisson equation can be obtained. A more detailed discussion of this method can be found in [128].

C.3.1 Discrete thermodynamic consistency

One property which holds on a continuous level to avoid unphysical state dissipation is the *preservation of thermodynamic equilibrium* [128]. Mathematically, this means that vanishing fluxes shall imply constant quasi Fermi potentials. The classical discrete counterpart of this property is formulated as below (see for example [186, 128]): a numerical flux $j = j_{KL}$ is said to be thermodynamically consistent if it satisfies an analogous discrete relation, i.e.

$$j = 0 \quad \text{implies} \quad \delta\varphi_{KL} = 0, \tag{C.6}$$

where $\delta\varphi_{KL} = (\varphi_L - \varphi_K)/U_T$. Similarly, we define $\delta\eta_{KL} = \eta_L - \eta_K$ and $\delta\psi_{KL} = (\psi_L - \psi_K)/U_T$ and $\delta E_{KL} = (E_{c,L} - E_{c,K})/(qU_T)$. We point out that the condition (C.6) holds in equilibrium. Here, we introduce a stronger notion of thermodynamic consistency, which holds outside of equilibrium, namely

$$j \le 0 \text{ implies } \delta\varphi_{KL} \ge 0 \qquad \text{and} \qquad j \ge 0 \text{ implies } \delta\varphi_{KL} \le 0. \tag{C.7}$$

An important property of defining thermodynamical consistency like above is that the sign of the numerical current is consistent with that of its continuous counterpart (C.3). Thermodynamic consistency is also important, when coupling the van Roosbroeck system to heat transport models [128]. We discuss now different numerical fluxes that may be used within a Voronoï finite volume framework.

C.3.2 The Scharfetter-Gummel scheme

First, we introduce the well known, classical Scharfetter-Gummel flux approximation [239] given by

$$j_{\mathrm{sg}} = j_0 \{B(\delta\psi_{KL} - \delta E_{KL})\mathcal{F}(\eta_L) - B(-\delta\psi_{KL} + \delta E_{KL})\mathcal{F}(\eta_K)\}, \tag{C.8}$$

where the constant j_0 is given by $j_0 = q\mu_n N_c \dfrac{U_T}{h_{KL}}$ for $h_{KL} = |\vec{x}_K - \vec{x}_L|$, and B is the Bernoulli function, $B(x) = \frac{x}{\exp(x)-1}$. It is important to point out that Scharfetter and Gummel introduced this numerical flux only in the Boltzmann regime, i.e. $\mathcal{F} = \exp$. In this case, the flux is thermodynamically consistent in the sense of (C.6). However, once we leave the Boltzmann regime, i.e. $\mathcal{F} \neq \exp$, and continue using (C.8) this numerical flux will no longer be thermodynamically consistent.

C.3.3 SEDAN scheme

Next, we present the SEDAN scheme, which yields a thermodynamically consistent approach also for state equations which do not necessarily rely on the Boltzmann approximation. The earliest reference we could find for such a excess chemical potential scheme is the source code of the SEDAN III simulator [265], which explains the reason we use this name. A numerical analysis focused comparison of this flux approximation is given in [50] and simulation results are presented in [2]. The scheme is motivated by rearranging the drift part to include the *excess*

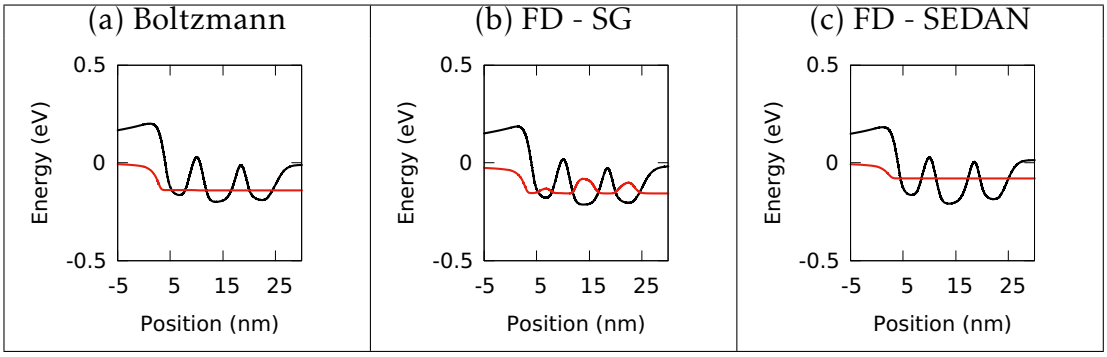


Figure C.1 – Conduction band edge (black) and quasi Fermi energy (red) at a bias of 3.3V (a) when using Boltzmann statistics, (b) when incorrectly using the Scharfetter-Gummel (SG) scheme with Fermi-Dirac (FD) statistics, and (c) when correctly using the SEDAN scheme with Fermi-Dirac statistics.

chemical potential, $\nu^{ex} = \ln \mathcal{F}(\eta) - \eta$, yielding

$$j_{\text{sedan}} = j_0 \{B(Q_{KL})\mathcal{F}(\eta_L) - B(-Q_{KL})\mathcal{F}(\eta_K)\} \tag{C.9}$$

with

$$Q_{KL} = \delta\psi_{KL} - \delta E_{KL} + \nu_L^{ex} - \nu_K^{ex} = \delta\psi_{KL} - \delta E_{KL} - \delta\eta_{KL} + \ln\frac{\mathcal{F}(\eta_L)}{\mathcal{F}(\eta_K)}. \tag{C.10}$$

Note that, using the definition of Q_{KL} and the fact that $e^x B(x) = B(-x)$, one can reformulate the SEDAN flux as

$$j_{\text{sedan}} = j_0 B(Q_{KL})\mathcal{F}(\eta_L)\left(1 - e^{\delta\varphi_{KL}}\right).$$

Therefore, it is easy to see that the SEDAN flux satisfies (C.7), since both $B(Q_{KL})$ and $\mathcal{F}(\eta_L)$ are positive: it is a thermodynamicaly consistent numerical flux.

Note that when applying Boltzmann statistics $\nu^{ex} = 0$ and the SEDAN flux becomes equivalent to the Scharfetter-Gummel expression. Therefore, in the next section, when displaying results using Boltzmann statistics, we only show results from one numerical scheme.

C.4 Simulations

To illustrate the importance of using thermodynamically consistent flux approximations, we study a simple (In,Ga)N multi QW (MQW) system. In particular, we consider three QWs and the same set of parameters as in [219]. For large negative values of η ($\eta \leq -2$, which correspond roughly to densities below 14% of the effective density of states N_c, thus a low carrier density regime in the conduction band of the wells) the Boltzmann approximation provides a good estimate of the Fermi-Dirac statistics. Therefore in certain cases Scharfetter-Gummel scheme can offer a good description of the drift-diffusion model (e.g. Ref. [222]).

On the other hand there are situations where Boltzmann statistics will not suffice and the importance of using a thermodynamically consistent scheme becomes apparent. Figure C.1 (a) displays the conduction band edge and quasi Fermi energies of the MQW system when treated using Boltzmann statistics at a bias of 3.3V. Applying the Scharfetter-Gummel scheme (C.8) to Fermi-Dirac statistics for a bias of 3.3V leads to humps in the quasi Fermi energy within each

QW, see Figure C.1 (b).

Recalling the condition (C.7), for thermodynamically consistent schemes, the discrete gradient of the discrete quasi Fermi level should indicates the direction of the discrete electron flow (as it is in the continuous case, according to (C.3)). However in Figure C.1 (b), one can see that, when the Scharfetter-Gummel scheme is applied to Fermi-Dirac statistics, resulting in a thermodynamically *inconsistent* scheme, the direction of electron flow is to the right outside the QW regions (e.g. between -5 nm and 3 nm) but to the left inside the QW regions (e.g. between 10 nm and 15 nm), which is not in accordance with (C.7). Moreover, as there is no generation of carriers in the system this change in direction of the electron flux is highly unphysical.

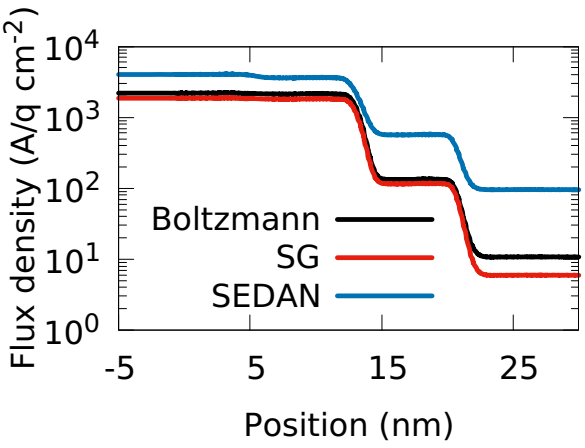


Figure C.2 – Numerical electron flux averaged over each atomic plane at a bias of 3.3V shown for Boltzmann statisitcs (black), Fermi-Dirac statistics incorrectly using the Scharfetter-Gummel (SG) flux discretization (red) and Fermi-Dirac statistics correctly using the SEDAN flux discretization (blue).

These humps in the quasi Fermi level are not present if one uses the Scharfetter-Gummel scheme with the Boltzmann approximation, which is a thermodynamically *consistent* scheme (Figure C.1 (a)). Similarly, using Fermi-Dirac statistics with the SEDAN scheme does not exhibit the unphysical humps in the Fermi level (Figure C.1 (c)).

Another perspective to interpret thermodynamic inconsistency is to note the incorrect interplay between quasi Fermi energies and local current fluxes. We see in Figure C.2 that the local numerical electron fluxes are positive and decrease monotonically across the QW regions. This is true for all three settings, that is, Boltzmann statistics in combination with the classical Scharfetter-Gummel scheme, as well as Fermi-Dirac statistics using both consistent and inconsistent numerical schemes. By our definition of a thermodynamically consistent scheme (C.7), a positive local numerical flux should imply a negative quasi Fermi potential discrete gradient. In fact, this is true for both consistent settings as one can see in Figure C.1 (a) and (c). However, for the inconsistent case the derivative of the quasi Fermi energies can become positive inside the wells, Fig. C.1 (b), also in disagreement of how the flux should behave at the continuous level (see the relation between the flux and the quasi Fermi potential (C.3)).

Previous studies of the numerical flux approximations have shown that a thermodynamically inconsistent scheme can result in the incorrect sign of the particle flux [186]. This is also reflected by the fact that (C.6) holds only for consistent schemes, which guarantee the physically correct sign of the current also far from equilibrium.

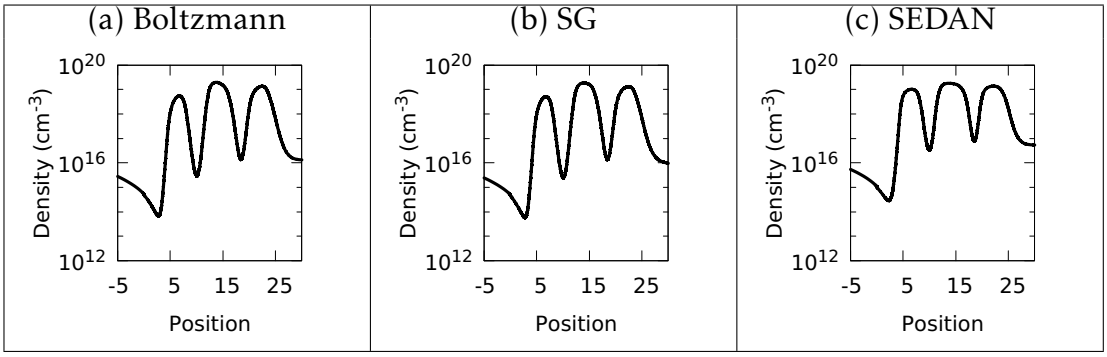


Figure C.3 – Electron density at a bias of 3.3V (a) when using Boltzmann statistics, (b) when incorrectly using the Scharfetter-Gummel (SG) scheme with Fermi-Dirac statistics, and (c) when correctly using the SEDAN scheme with Fermi-Dirac statistics.

The physical reason of why an inconsistent scheme produces humps within the QWs becomes apparent when looking at the corresponding densities at a bias of 3.3V. The Fermi-Dirac function grows like a polynomial while the Boltzmann approximation grows for large densities exponentially, see for example Figure 50.9 in [128]. This different behaviour leads to nonlinear diffusion, the so-called diffusion enhancement, for non-Boltzmann statistics, see (C.4), or [186]. The Scharfetter-Gummel scheme neglects the diffusion enhancement assuming only linear diffusion, which has a knock-on impact on the carrier density: the densities calculated using Boltzmann statistics (Fig. C.3 (a)) and using Fermi-Dirac statistics with the Scharfetter-Gummel scheme (Fig. C.3 (b)) are visibly indistinguishable. However, in order to produce the same density between a Boltzmann and Fermi-Dirac calculation, the quasi Fermi levels must differ. This results in the unusual behaviour of the quasi Fermi level exhibited in Figure C.1 (b). Comparing these densities with the correctly calculated density using Fermi-Dirac statistics in combination with the thermodynamically consistent SEDAN scheme (Figure C.3 (c)) we see that the choice of statistics function will impact carrier density in the well, and more strongly in the barrier regions at the here chosen example bias of 3.3V.

Because it influences the carrier density, thermodynamic inconsistency has direct implications for the computed recombination rates as well as the current-voltage (IV) curves. Next, we compare the recombination rates calculated using Boltzmann statistics with those calculated while incorrectly using the Scharfetter-Gummel scheme implementing Fermi-Dirac statistics.

This is highlighted in Figure C.4 (a), where the differences between Fermi-Dirac and Boltzmann-like behaviour are shown for the three recombination rates, calculated via

$$\Delta \log(\text{recomb.}) = \log(\mathrm{R}_i^{\mathrm{SEDAN}}) - \log(\mathrm{R}_i^{\mathrm{Boltzmann}}) \,,$$

where R_i^s is the recombination rate associated with the process i ($i \in$ {Shockley–Read–Hall, Radiative,Auger}) calculated with the scheme s ($s \in$ {Boltzmann, SEDAN}). From this figure it becomes clear that the Boltzmann behaviour overall underestimates the recombination across the multi QW region of the device. In particular, the Auger recombination is underestimated by up to two orders of magnitude at a bias of 3.3V. These differences increase as the bias is increased (not shown). This can have consequences for overall device behaviour such as the internal quantum efficiency and the IV curves. The latter are shown in Figure C.4 (b), where the decreased recombination current displayed in the Boltzmann and thermodynamically inconsistent Fermi-Dirac scheme leads to an underestimate of the current density by close to an

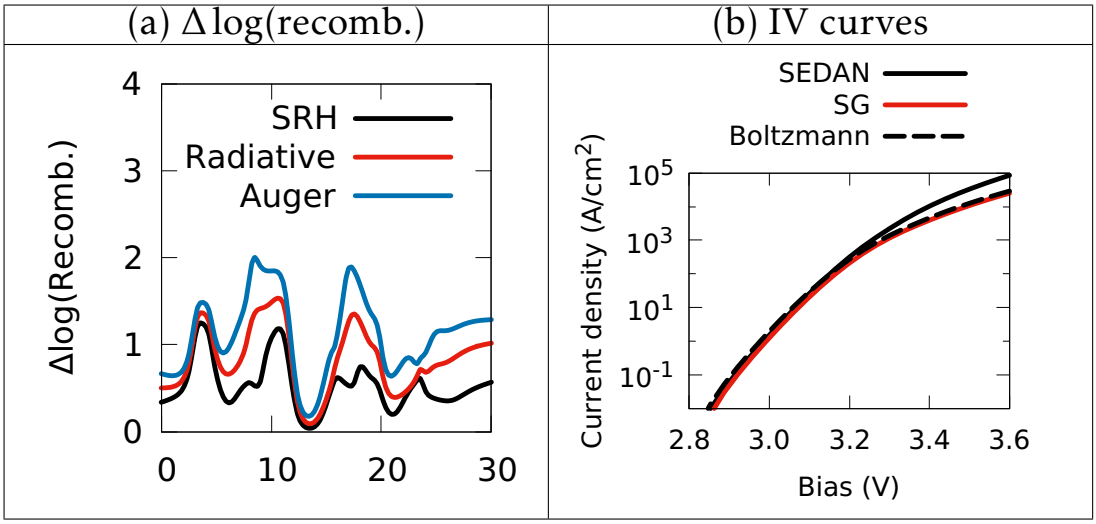


Figure C.4 – (a) Difference in magnitude of the Shockley-Read-Hall (SRH, black), radiative (red) and Auger (blue) recombination rates between Fermi-Dirac and Boltzmann statistics at a bias of 3.3V, calculated as described in the main text. (b) Current density-voltage (IV) curves using Boltzmann statistics (black, dashed), Fermi-Dirac statistics using the Scharfetter-Gummel scheme (SG, red) and Fermi-Dirac statistics using the SEDAN scheme (black, solid), shown on a log scale.

order of magnitude at 3.6V.

The results highlighted above indicate that Fermi statistics implemented using a thermodynamically inconsistent scheme will result in Boltzmann-like behaviour in LED simulations – at least in terms of carrier and current densities. If this is extended to *laser* simulations the consequences can be even more dramatic, as the gain calculation depends on the difference between the electron and hole quasi Fermi energies [14], expressed by the so-called Fermi voltage. In this case the unphysical humps seen in Figure C.1 (b) will lead to an incorrect prediction of the transparency density.

C.5 Conclusion

In this paper, we have shown the importance of using a thermodynamically consistent flux discretization when describing drift-diffusion processes within quantum well devices.

Using the classical Scharfetter-Gummel scheme with Fermi-Dirac statistics is an example of such an inconsistent scheme. Here we studied an (In,Ga)N multi quantum well structure as an example since it is a very important material system for optoelectronic devices. In this case, the Fermi levels show humps within the quantum wells resulting in an unphysical description of the direction of the current, e.g. assuming the usual continuous expression. This is explained by the omission of diffusion enhancement from the numerical current expression, that leads to a similar density distribution as using Boltzmann statistics. This has a knock-on effect for recombination and current-voltage behaviour, where using Fermi-Dirac statistics with a thermodynamically inconsistent scheme may incorrectly predict a Boltzmann-like behaviour.

Contrarily, for a thermodynamically consistent scheme, such as the SEDAN scheme, these unphysical humps in the Fermi levels disappear and accurate current curves and recombination

processes are predicted. Thus, thermodynamically consistent schemes are essential to address open questions, such as the efficiency drop in modern light emitting devices and to reliably guide their design.

Appendix D

Theoretical study of the impact of alloy disorder on carrier transport and recombination processes in deep UV (Al,Ga)N light emitters

This work is an article published in Applied Physics Letters [133], in collaboration with Robert Finn, Michael O'Donovan, Patricio Farrell, Timo Streckenbach, Thomas Koprucki and Stefan Schulz.

Aluminium gallium nitride ((Al,Ga)N) has gained significant attention in recent years due to its potential for highly efficient light emitters operating in the deep ultra-violet (UV) range (< 280 nm). However, given that current devices exhibit extremely low efficiencies, understanding the fundamental properties of (Al,Ga)N-based systems is of key importance. Here, using a multi-scale simulation framework, we study the impact of alloy disorder on carrier transport, radiative and non-radiative recombination processes in a *c*-plane $Al_{0.7}Ga_{0.3}N/Al_{0.8}Ga_{0.2}N$ quantum well embedded in a *p-n* junction. Our calculations reveal that alloy fluctuations can open "percolative" pathways that promote transport for the electrons and holes into the quantum well region. Such an effect is neglected in conventional, and widely used transport simulations. Moreover, we find that the resulting increased carrier density and alloy induced carrier localization effects significantly increase non-radiative Auger-Meitner recombination in comparison to the radiative process. Thus, to avoid such non-radiative process and potentially related material degradation, a careful design (wider well, multi quantum wells) of the active region is required to improve the efficiency of deep UV light emitters.

Light emitters operating in the ultraviolet (UV) spectral range have received significant attention in recent years due to their importance for a wide range of applications [182]. A region of particular interest is the deep-UV-C wavelength window (< 280 nm), which enables for instance water purification and sterilization. Aluminium gallium nitride ((Al,Ga)N) based

light emitting diodes (LEDs) are ideal candidates for such applications since their emission wavelengths can in principle be tuned across the entire UV spectrum and they do not require the use of toxic mercury. Unfortunately, the external quantum efficiency (EQE) of (Al,Ga)N-based LEDs is very low (≤1%) in the deep UV-C range and multiple factors contribute to this, e.g. low light extraction efficiencies and poor carrier injection, high defect densities and thus reduced radiative recombination in the active region [4, 182].

Understanding the fundamental material and device properties is of central importance in improving the efficiency of such devices. Theory and simulation plays an important role for guiding the device design. However, III-N systems and alloys, such as (In,Ga)N, (Al,In)N but also (Al,Ga)N, exhibit in general very different properties when compared to other III-V materials, e.g. (In,Ga)As. For instance, experimental and theoretical studies reveal strong alloy fluctuation induced carrier localization effects in III-N alloys and heterostructures [82, 243, 242, 103, 261, 134]. Accounting for such effects present a significant challenge for an accurate theoretical and numerical description, since a three-dimensional (3D) treatment of, for instance, a quantum well (QW), is required ideally on an atomistic level. Numerical challenges are further amplified when it comes to simulating LED structures, which require not only large 3D supercells but also self-consistent calculations. Significant progress has been made to address these questions [225, 103, 221]. But, in comparison to (In,Ga)N-based systems, understanding the fundamental impact of alloy fluctuations on electronic, optical and carrier transport properties in (Al,Ga)N-based QW systems is sparse.

Here, we target these questions for an $Al_{0.7}Ga_{0.3}N/Al_{0.8}Ga_{0.2}N$ QW embedded in *p*- and *n*-doped regions, thus a system relevant for deep UV (Al,Ga)N-based LEDs. The calculations build on our recently established quantum corrected multi-scale simulation framework that connects atomistic electronic structure theory with a drift-diffusion scheme to analyze carrier transport. To target recombination processes, the model is coupled with the widely used *ABC* model [226], where empirical *A*, *B* and *C* coefficients from the literature are employed. Our results indicate that alloy fluctuations can promote carrier transport through the intrinsic barriers by opening up percolative paths that are lower in energy. These pathways appear to improve the carrier injection into the QW active region – a feature that is neglected in 'conventional' calculations utilizing a virtual crystal approximation (VCA). Moreover, while we find that alloy disorder and carrier localization effects enhance radiative recombination, we find also that non-radiative Auger-Meitner recombination processes are strongly increased. The resulting increase in high energy carriers via non-radiative Auger-Meitner recombination may also lead to a degradation of the material as indicated in experimental studies [148]. As a consequence, our results suggest that alloy disorder can have a detrimental effect on the device efficiency. To reduce Auger-Meitner related processes, increasing the well width or employing multi QW (MQW) active regions to reduce the carrier density may be a way to improve quantum efficiencies in UV emitters, as the above discussed percolation paths may be exploited to distribute the carriers more evenly between MQWs.

In order to investigate the impact alloy disorder has on carrier transport and recombination processes in (Al,Ga)N-based LEDs, we describe the electronic structure of the active region by means of a empirical nearest neighbour sp^3 tight-binding (TB) model. We note that while we treat the TB matrix elements as free parameters to reproduce hybrid-functional DFT band structures, the TB matrix could also be constructed from maximally localized Wannier functions generated from DFT [259]. This would thus provide an avenue to extend our method towards a DFT-based TB Hamiltonian. In general, our employed TB framework treats strain and polarization field fluctuations due to random alloy fluctuations with an atomistic resolution. We have recently employed the TB model to study the influence of alloy fluctuations on the electronic and optical properties of (Al,Ga)N/AlN QWs. Our work revealed that random alloy fluctuations are sufficient

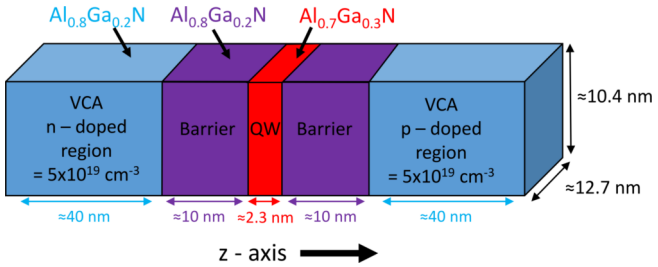


Figure D.1 – Schematic illustration of the (Al,Ga)N-based *p-i-n* system underlying the simulations. The intrinsic barriers plus quantum well region is denoted as the "atomistic" mesh in the main text, while the *n*- and *p*-regions are described by a sparser device mesh.

to lead to carrier localization effects in such systems. A detailed description of the model and electronic structure results is given in Ref. [134].

To connect our TB model to drift-diffusion (DD) simulations we have recently developed the following procedure [221]. As a first step a "local" TB Hamiltonian is constructed. By diagonalising this local Hamiltonian an energy landscape that accounts for alloy induced fluctuations in the conduction and valence band profile is obtained. This energy landscape with an atomistic resolution is mapped onto a finite element mesh using the software library `pdelib` [141]. The underlying finite element mesh has as many nodes as the atoms in the system. To target LED device simulations, the "atomistic" mesh is connected to a sparser device mesh representing *p*- and *n*-doped regions. Using localization landscape theory (LLT) [225] in conjunction with the software package `ddfermi`, [104] quantum corrected DD simulations are performed. More details about the carrier transport calculations, the mesh generation and numerical aspects can be found in Refs. [221] and [128].

The above outlined quantum corrected multi-scale simulation framework is here applied to a *p-i-n* system where the intrinsic (*i*) region contains a 2.3 nm wide $Al_{0.7}Ga_{0.3}N$ QW embedded into approximately 10 nm wide intrinsic $Al_{0.8}Ga_{0.2}N$ barriers, following the UV-C device design in Ref. [235]. The atomistic mesh region describing the active region of the device has the dimensions of roughly $17.4 \times 15.1 \times 22.2$ nm^3, corresponding to $\approx$ 560,000 atoms, and is coupled with a sparser device mesh by adding 40 nm of *n*- and *p*-doped regions, respectively. The doping density is 5×10^{19} cm^{-3} following the set-up described in Ref. [244]. A schematic illustration of the structure is given in Fig. D.1. To connect the conduction band and valence band edges between the atomistic and the sparser mesh, we use a blend of Gaussian softening and LLT to avoid discontinuities at such an interface. Details on the Gaussian softening and the effect of LLT on the energy landscape can be found elsewhere [221]. To test the impact of the in-plane dimensions of the simulation cell and thus also the influence of the alloy microstructure on the results, the calculations have been repeated for different cell sizes (not shown). These different calculations revealed in general the same trends as reported below, so that the above detailed simulation set up gives a representative picture of the impact of alloy disorder on the carrier transport in the discussed device.

To investigate the impact alloy fluctuations have on the carrier transport and the recombination process we proceed as follows: Our starting point is a fully atomistic description, meaning that alloy fluctuations in both the intrinsic (Al,Ga)N QW and (Al,Ga)N barriers are accounted for (cf. Fig. D.1). To study the impact of alloy disorder on the carrier injection, we investigate a second set-up in which alloy fluctuations in the (Al,Ga)N QW are treated on an atomistic level, however the intrinsic (Al,Ga)N barriers are described in a VCA. Finally, we employ a

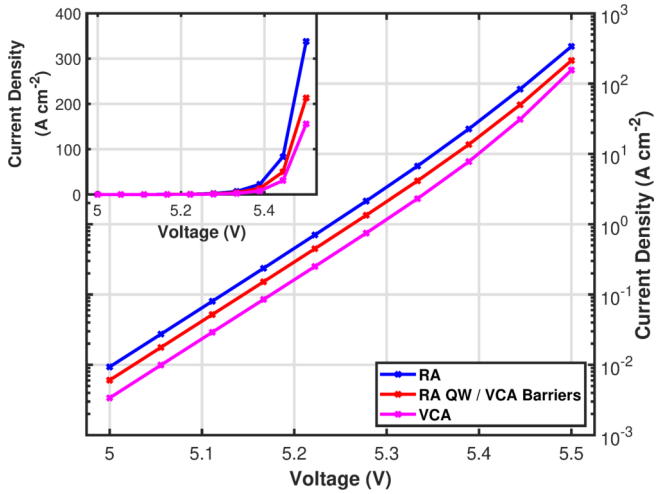


Figure D.2 – (a) Current-voltage curves, on a log-scale, for an $Al_{0.7}Ga_{0.3}N/Al_{0.8}Ga_{0.2}N$ quantum well embedded in a *p-n* junction. The data are shown for simulations that (i) account for random alloy (RA) fluctuations in the well and barriers (blue), (ii) account for RA fluctuations in the well but treat the barrier in a virtual crystal approximation (VCA) (red), and (iii) use a VCA in the entire *p-i-n* structure (pink). The inset shows the I-V curves on a linear scale.

VCA description of the entire QW plus barrier system. This last setting is basically equivalent to one-dimensional carrier transport simulations widely employed in commercially available software packages. In order to establish a comparable VCA description of the targeted system, without introducing additional free parameters (e.g. alloy dependence and or bowing parameters of material parameters), the atomistic energy landscape is averaged over the corresponding areas (e.g. well or barrier). To achieve this averaging procedure, the QW was treated initially with Dirichlet boundary conditions so that, except from alloy fluctuation induced variations, the conduction and valence band edges are flat. A VCA description of the barriers is then straightforward and achieved by averaging the atomistic energy landscape over the entire barrier volume. However, the QW VCA description is slightly more complicated as piezoelectric and spontaneous polarization induced built-in fields across the QW cause the band edges to slope. Thus, a volume average over the QW region is not possible. In this case, the atomistic energy landscape is averaged across each *c*-plane along the growth direction (*c*-axis). While this can lead to slight fluctuations in the energy landscape, the approach chosen here is better suited to compare the different calculations, since it avoids constructing a VCA that requires specific information on how to average e.g. material and TB parameters with composition, or to perform a separate VCA built-in field calculation.

Building on these different simulation set-ups, Fig. D.2 displays the calculated current-voltage (I-V) curves of the above described (Al,Ga)N *p-i-n* structure using (i) the fully atomistic approach (RA), (ii) accounting for alloy fluctuations in the well but not the barrier (RA QW/VCA Barriers) and (iii) a VCA of the entire system. One can infer from this figure that the fully atomistic treatment gives independent of the applied voltage the highest current density, while the full VCA description of the system (well and barriers) results in the lowest density. Figure D.2 also reveals that the current density, especially at lower bias values, is increased when accounting for alloy fluctuations in the barrier; this can be seen from the comparison of the I-V curves for the fully atomistic and 'partial' VCA (RA QW, VCA Barriers) treatment. Thus, our results indicate

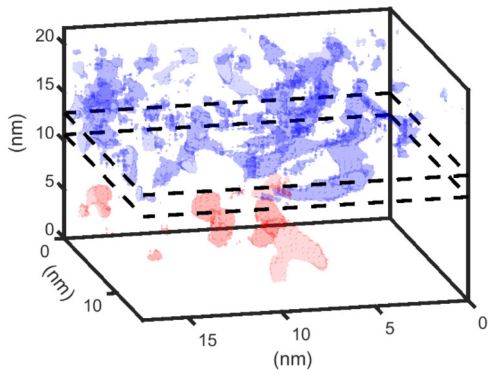


Figure D.3 – 3D isosurface plot of the current density in the intrinsic (Al,Ga)N quantum well and barrier regions at a bias of 5.5 V, using the fully atomistic description. The current density for the holes (blue) and electrons (red) are plotted at 700 A cm^{-2}, reflecting locally high current densities due to alloy fluctuations; dashed black lines indicate well boundaries.

that alloy fluctuations are beneficial for the carrier injection into the QW region.

To visualise this aspect further, Fig. D.3 depicts an isosurface plot of the current density in the intrinsic (Al,Ga)N QW and barrier regions, cf. Fig. D.1, at a fixed bias point of 5.5 V obtained from the fully atomistic description of the system. One can infer that alloy fluctuation induced band edge energy fluctuations lead to regions of higher and lower current densities. Thus, the fluctuations in the band edge energies enable percolation paths into the active region; such pathways are absent in a VCA description of the system. We note that this percolation path effect has also be observed in (In,Ga)N QWs and (Al,Ga)N barriers [263, 45, 231, 264]. We note that both electron and hole transport are affected by these percolation paths. However, due to the lower effective electron mass, the electron charge density is more smeared out, and compared to the holes, fewer regions reach the isosurface value at which the current density is plotted in Fig. D.3. As a consequence of these percolation paths one expects higher carrier densities in the active region in comparison to a simulation that neglects alloy disorder. Also, as we will see below, the percolation paths can lead to an increased hole density in the intrinsic barrier on the *n*-side of the studied system. Such a situation may be beneficial for inter-well transport in (Al,Ga)N multi-QW systems as recently reported in (In,Ga)N-based LEDs [203].

To study the impact of alloy fluctuations on the carrier transport and recombination in (Al,Ga)N-based LED structures in further detail, Fig. D.4 depicts electron and hole carrier densities above the turn-on voltage of the device, at a bias of 5.5 V. Figure D.4 (a) displays the *averaged* carrier density, over each *c*-plane, for the random alloy system along the *c*-axis (growth direction) together with the VCA data which neglects alloy disorder. When comparing the averaged random alloy data with the VCA results, it becomes clear that *(on average)* both electron and hole carrier densities are increased in the well when taking alloy fluctuations into account. Figure D.4 (a) indicates that the *average* hole charge density does not leak as much into the intrinsic (Al,Ga)N barrier material on the *n*-side of system (region below 10.7 nm), when compared to the hole density profile in VCA; one may attribute this to alloy induced hole localization effects which are not accounted for by the VCA description. The *average* electron density extends further into the (Al,Ga)N barrier region on the *p*-side (region above 13 nm) when compared to the VCA data; we relate this to the effect that alloy disorder and quantum corrections effectively reduce the quantum confinement between well and barrier and that the electrons are less strongly affected by alloy induced carrier localization effects as already seen

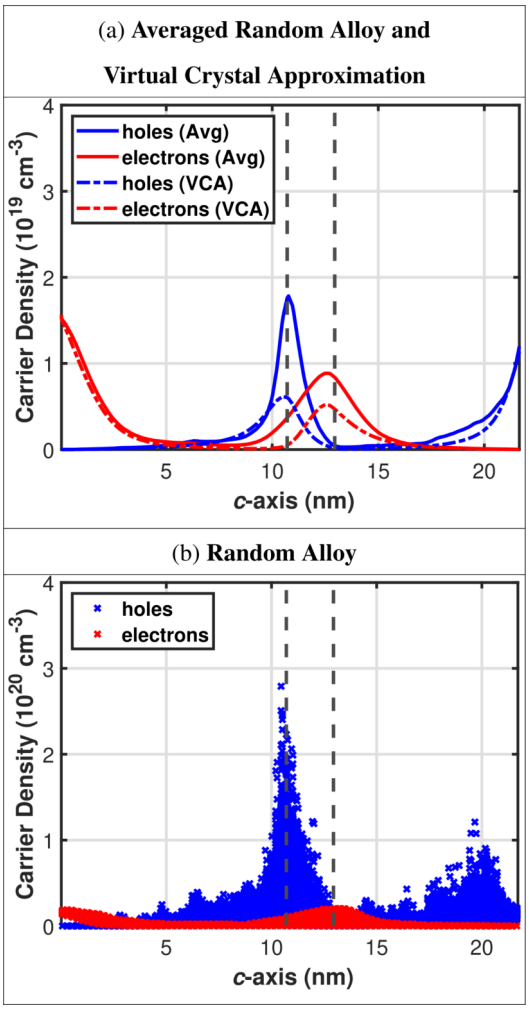


Figure D.4 – Electron (red) and hole (blue) density along the c-axis of a $Al_{0.7}Ga_{0.3}N/Al_{0.8}Ga_{0.2}N$ quantum well embedded in a p-i-n junction at a bias of 5.5 V; data are shown for the intrinsic regions of the device. (a) Averaged carrier distribution of the fully atomistic calculation over each c-plane along the c-axis (solid lines), and the carrier distribution for the virtual crystal approximation (dashed lines); (b) Scatter plot of the carrier distribution in the fully atomistic calculation. The vertical dashed black lines indicate well boundaries. Note (b) is an order of magnitude larger than (a) on the y-axis.

in (In,Ga)N-based systems [221] . However, it must be noted that the random alloy data in Fig. D.4 (a) are averaged over each c-plane. Thus, while on average the hole density may be smaller in e.g. the intrinsic barrier of the n-region when compared to VCA, local alloy regions may have high hole densities. To shed light onto this behavior, Fig. D.4 (b) displays a scatter plot of the carrier densities along the c-axis in the intrinsic (Al,Ga)N regions. The symbols give the carrier density at each lattice site within the respective c-planes. Figure D.4 (b) reveals strong fluctuations in the *local* densities which may be an order of magnitude higher than the *average* densities displayed in Fig. D.4 (a). The *local* hole carrier density is also noticeably higher

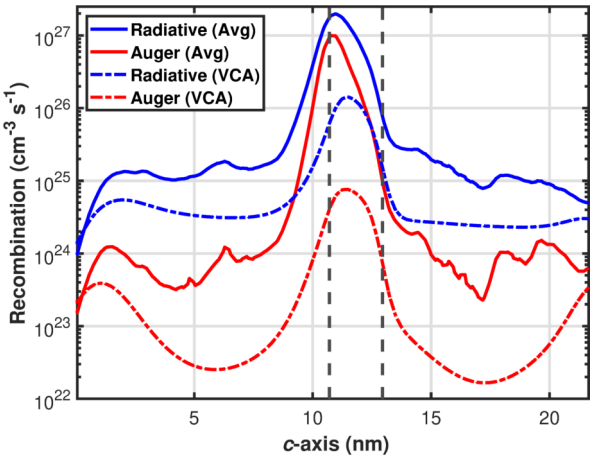


Figure D.5 – Radiative and Auger-Meitner non-radiative recombination rates along the *c*-axis of the $Al_{0.7}Ga_{0.3}N/Al_{0.8}Ga_{0.2}N$ *p-i-n* system at a bias of 5.5 V; the data are shown for the intrinsic regions of the device. Solid lines: results from the atomistic calculations taking alloy fluctuations into account (averaged over each *c*-plane). Dashed lines: data obtained from a virtual crystal approximation of the same system; vertical dashed black lines indicate well boundaries.

when compared to the electrons. In general, due to the higher effective mass, carrier localization for holes is more pronounced when compared to electrons [134]. Therefore, we conclude that percolative pathways, can promote easier transport of the carriers through the barriers. Thus, holes may transport to the *n*-side of the system more easily when compared to predictions from a VCA model. This aspect is of interest when designing multi-QW-based (Al,Ga)N LEDs, since it can help with a more equal distribution of holes between the wells, and thus be important for recombination processes.

Equipped with this information we turn our attention now to the recombination process in the system. To do so we employ the widely used *ABC* model. In the following we mainly focus our attention on the radiative and Auger-Meitner (non-radiative) processes; with increased current densities radiative and non-radiative Auger-Meitner processes become dominant over Shockley-Read-Hall (SRH) recombination as these contributions scale with carrier density n as $\propto n^2$ (radiative) and $\propto n^3$ (Auger-Meitner) compared to $\propto n$ (SRH). We note that defect related non-radiative processes, such as SRH and trap assisted tunnelling (TAT) processes can also affect the device performance of UV-C (Al,Ga)N-based LEDs significantly. While we take SRH into account, TAT processes have not been considered here but could be accounted for as a SRH-like process with different capture rates as detailed in Ref. [233]. Turning to radiative and Auger-Meitner (non-radiative) recombination rates, Fig. D.5 depicts these rates calculated (i) including random alloy fluctuations and thus carrier localization effects and (ii) using VCA for the entire system, consequently excluding alloy disorder. Figure D.5 displays the random alloy (RA) data averaged over each *c*-plane. We stress that the recombination rates are calculated using the *local* carrier density data, thus accounting for the effect that locally one may encounter both high electron and hole densities. The required radiative (B) and Auger-Meitner (C_{nnp} and C_{npp}) recombination coefficients are taken from Ref. [244]. We note that building on this often used carrier density dependent recombination model, the wave function character of electrons and holes is not taken into account. While the symmetry of the wave functions can affect radiative and non-radiative recombination rates, analyzing the spatial overlap of

electron and hole charge density provides first insight into the impact of alloy disorder on the recombination rates. Comparing the atomistic calculation with the VCA data it becomes clear that both radiative and non-radiative contributions are enhanced by alloy disorder since this can improve the carrier injection but also induce carrier localization in the wells. But, upon further inspection, Fig. D.5 reveals that the non-radiative rate increases more strongly when compared to the radiative. In the VCA the ratio of radiative to non-radiative peak values is $r^{\mathrm{VCA}} \approx 19$ while in the atomistic simulation it is $r^{\mathrm{RA}} \approx 2$ at a bias of 5.5 V. In general, and using the *ABC* model [226] here as a starting point, the radiative recombination contributes with Bn^2 while the Auger-Meitner non-radiative rate with Cn^3, where n is the carrier density. Thus, when alloy disorder increases the carrier density in local regions of the well, the Auger-Meitner rate is expected to increase more quickly than the radiative rate. Moreover, the Auger-Meitner rate is a three-particle process and has basically two contributions, namely an electron-electron-hole and a hole-hole-electron process [180]. Given that alloy induced carrier localization effects are significant for holes, locally the hole-hole wave function overlap is (strongly) increased. In turn this can result in increased Auger-Meitner rates [210]. Taking all this together, alloy disorder and related carrier localization effects are expected to lead to a significant increase of the non-radiative Auger-Meitner effect and thus have a detrimental effect on the performance of (deep) UV LEDs; a similar effect has been seen (In,Ga)N-base LEDs [264]. Moreover, strong Auger-Meitner non-radiative recombination may also lead to material degradation effects in deep-UV LEDs, as recently indicated in experiments [148, 233].

In summary, using a quantum corrected multi-scale simulation framework, we presented a detailed analysis of the impact of alloy fluctuations on carrier transport and recombination processes in an $Al_{0.7}Ga_{0.3}N/Al_{0.8}Ga_{0.2}N$ single quantum well system embedded in a *p*-*n* junction. Understanding the fundamental properties of such systems provides useful insight into the physics of deep UV light emitters. Our results show that alloy disorder can lead to improved carrier injection into the active region, i.e. quantum well, through percolative pathways. Also, these pathways can give rise to higher hole densities in the intrinsic region near the *n*-side of a device, an aspect of potential interest for tailoring inter-well transport in (Al,Ga)N-based multi quantum well systems. In general, these percolation effects are not captured by standard virtual crystal approximations employed in many commercially available carrier transport simulation packages. Moreover, we find an increase in the carrier density in the well which is accompanied by alloy induced carrier localization effects. In comparison to virtual crystal simulations, both radiative and non-radiative Auger-Meitner recombination processes are increased. We attribute this to alloy induced carrier localization effects, resulting in high local carrier densities which affect the Auger-Meitner effect more strongly than the radiative process. Moreover, the high energy of such Auger-Meitner generated carriers may lead to a degradation of the material and thus an enhancement of Shockley-Read Hall recombination. Overall, our results indicate that alloy disorder opens up percolative pathways improving carrier transport, but the enhanced Auger effect can have a detrimental effect on the efficiency of the device. While we have focused here on relatively high Al contents in the well and barrier, the above drawn conclusions are also expected to hold for systems with lower Al contents, given that our recent atomistic studies reveal that even at only 10% Al in (Al,Ga)N hole localization effects are observed. Thus, in general a careful design of the active region of an UV LED, e.g. by using wider wells as discussed in recent literature [217, 262] or multi quantum wells by exploiting percolation pawthways to improve the interwell transport, is necessary to reduce non-radiative recombination processes and prevent material damage.

Bibliography

[1] D. Abdel, C. Chainais-Hillairet, P. Farrell, and M. Herda, *Numerical analysis of a finite volume scheme for charge transport in perovskite solar cells*, IMA Journal of Numerical Analysis, (2023). in press.

[2] D. Abdel, P. Farrell, and J. Fuhrmann, *Assessing the quality of the excess chemical potential flux scheme for degenerate semiconductor device simulation*, Optical and Quantum Electronics, 53 (2021).

[3] T. Aiki and A. Muntean, *Existence and uniqueness of solutions to a mathematical model predicting service life of concrete structures*, Adv. Math. Sci. Appl., 19 (2009), pp. 109–129.

[4] H. Amano, R. Collazo, C. De Santi, S. Einfeldt, M. Funato, J. Glaab, S. Hagedorn, A. Hirano, H. Hirayama, R. Ishii, Y. Kashima, Y. Kawakami, R. Kirste, M. Kneissl, R. Martin, F. Mehnke, M. Meneghini, A. Ougazzaden, P. J. Parbrook, S. Rajan, P. Reddy, F. Römer, J. Ruschel, B. Sarkar, F. Scholz, L. J. Schowalter, P. Shields, Z. Sitar, L. Sulmoni, T. Wang, T. Wernicke, M. Weyers, B. Witzigmann, Y.-R. Wu, T. Wunderer, and Y. Zhang, *The 2020 UV emitter roadmap*, J. Phys. D: Appl. Phys, 53 (2020), p. 503001.

[5] D. Anderson and J. Droniou, *An arbitrary-order scheme on generic meshes for miscible displacements in porous media*, SIAM J. Sci. Comput., 40 (2018), pp. b1020–b1054.

[6] 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 (2017), pp. 3633–3670.

[7] P. F. Antonietti, L. Beirão da Veiga, and G. Manzini, eds., *The virtual element method and its applications*, vol. 31 of SEMA SIMAI Springer Ser., Cham: Birkhäuser, 2022.

[8] A. Arnold, J. A. Carrillo, L. Desvillettes, J. Dolbeault, A. Jüngel, C. Lederman, P. A. Markowich, G. Toscani, and C. Villani, *Entropies and equilibria of many-particle systems: An essay on recent research*, Monatsh. Math., 142 (2004), pp. 35–43.

[9] D. N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760.

[10] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., 39 (2002), pp. 1749–1779.

[11] D. N. Arnold, G. David, D. Jerison, S. Mayboroda, and M. Filoche, *Effective Confining Potential of Quantum States in Disordered Media*, Phys. Rev. Lett., 116 (2016), p. 056602.

[12] M. Auf der Maur, A. Pecchia, G. Penazzi, W. Rodrigues, and A. Di Carlo, *Efficiency Drop in Green* InGaN/GaN *Light Emitting Diodes: The Role of Random Alloy Fluctuations*, Phys. Rev. Lett., 116 (2016), p. 027401.

[13] I. Babuška, *The finite element method with penalty*, Math. Comput., 27 (1973), pp. 221–228.

[14] U. Bandelow, H. Gajewski, and R. Hünlich, *Fabry-perot lasers: Thermodynamics-based modeling*, Optoelectronic Devices: Advanced Simulation and Analysis, (2005), pp. 63–85.

[15] R. Bank and D. Rose, *Some error estimates for the box method*, SIAM J Numer Anal, 24 (1987), pp. 777–787.

[16] C. Bataillon, F. Bouchon, C. Chainais-Hillairet, C. Desgranges, E. Hoarau, F. Martin, S. Perrin, M. Tupin, and J. Talandier, *Corrosion modelling of iron based alloy in nuclear waste repository*, Electrochimica Acta, 55 (2010), pp. 4451–4467.

[17] J. Bear, *Dynamics of Fluids in Porous Media*, Courier Corporation, 1988.

[18] L. Beaude and S. Lemaire, *ParaSkel++: a C++ platform for the high-performance, arbitrary-order, 2/3D numerical approximation of PDEs on general polytopal meshes using skeletal Galerkin methods*, Aug. 2021.

[19] L. Beirão da Veiga, J. Droniou, and G. Manzini, *A unified approach for handling convection terms in finite volumes and mimetic discretization methods for elliptic problems*, IMA J. Numer. Anal., 31 (2011), pp. 1357–1401.

[20] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo, *Basic principles of virtual element methods*, Math. Models Methods Appl. Sci., 23 (2013), pp. 199–214.

[21] L. Beirão da Veiga, F. Brezzi, L. D. Marini, and A. Russo, *The Hitchhiker's guide to the virtual element method*, Math. Models Methods Appl. Sci., 24 (2014), pp. 1541–1573.

[22] L. Beirão da Veiga, F. Dassi, and A. Russo, *High-order virtual element method on polyhedral meshes*, Comput. Math. Appl., 74 (2017), pp. 1110–1122.

[23] N. Ben Abdallah and R. El Hajj, *Diffusion and guiding center approximation for particle transport in strong magnetic fields*, Kinet. Relat. Models, 1 (2008), pp. 331–354.

[24] I. Ben Gharbia, C. Cancès, T. Faney, M. Jonval, and Q. H. Tran, *Robust resolution of single-phase chemical equilibrium using parametrization and Cartesian representation techniques*. working paper or preprint, 2023.

[25] M. Bendahmane, Z. Khalil, and M. Saad, *Convergence of a finite volume scheme for gas-water flow in a multi-dimensional porous medium*, Math. Models Methods Appl. Sci., 24 (2014), pp. 145–185.

[26] M. Bessemoulin-Chatard, *A finite volume scheme for convection-diffusion equations with nonlinear diffusion derived from the Scharfetter-Gummel scheme*, Numer. Math., 121 (2012), pp. 637–670.

[27] M. Bessemoulin-Chatard and C. Chainais-Hillairet, *Exponential decay of a finite volume scheme to the thermal equilibrium for drift-diffusion systems*, J. Numer. Math., 25 (2017), pp. 147–168.

[28] ——, *Uniform-in-time bounds for approximate solutions of the drift-diffusion system*, Numer. Math., 141 (2019), pp. 881–916.

[29] M. Bessemoulin-Chatard, C. Chainais-Hillairet, and M.-H. Vignal, *Study of a finite volume scheme for the drift-diffusion system. Asymptotic behavior in the quasi-neutral limit*, SIAM J. Numer. Anal., 52 (2014), pp. 1666–1691.

[30] S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, and P. Vogl, *nextnano: General Purpose 3-D Simulations*, IEEE Transactions on Electron Devices, 54 (2007), pp. 2137–2142.

[31] J. Blakemore, *Approximations for Fermi-Dirac integrals, especially the function $\mathcal{F}_{1/2}(\eta)$ used to describe electron density in a semiconductor*, Solid-State Electronics, 25 (1982), pp. 1067–1076.

[32] X. Blanc and E. Labourasse, *A positive scheme for diffusion problems on deformed meshes*, ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik, 96 (2016), pp. 660–680.

[33] T. Bodineau, J. Lebowitz, C. Mouhot, and C. Villani, *Lyapunov functionals for boundary-driven nonlinear drift-diffusion equations*, Nonlinearity, 27 (2014), pp. 2111–2132.

[34] L. Boltzmann, *Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen*, k. und k. Hof- und Staatsdr., 1872.

[35] J. Bonelle, D. A. Di Pietro, and A. Ern, *Low-order reconstruction operators on polyhedral meshes: Application to Compatible Discrete Operator schemes*, Computer Aided Geometric Design, 35–36 (2015), pp. 27–41.

[36] F. Bonizzoni, M. Braukhoff, A. Jüngel, and I. Perugia, *A structure-preserving discontinuous Galerkin scheme for the Fisher-KPP equation*, Numer. Math., 146 (2020), pp. 119–157.

[37] M. Braukhoff, I. Perugia, and P. Stocker, *An entropy structure preserving space-time formulation for cross-diffusion systems: Analysis and galerkin discretization*, SIAM Journal on Numerical Analysis, 60 (2022), pp. 364–395.

[38] K. Brenner and C. Cancès, *Improving Newton's method performance by parametrization: the case of the Richards equation*, SIAM J. Numer. Anal., 55 (2017), pp. 1760–1785.

[39] F. Brezzi, K. Lipnikov, and M. Shashkov, *Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes*, SIAM J. Numer. Anal., 43 (2005), pp. 1872–1896.

[40] F. Brezzi, K. Lipnikov, and V. Simoncini, *A family of mimetic finite difference methods on polygonal and polyhedral meshes*, Math. Models Methods Appl. Sci., 15 (2005), pp. 1533–1551.

[41] F. Brezzi, L. D. Marini, S. Micheletti, P. Pietra, R. Sacco, and S. Wang, *Discretization of semiconductor device problems. I*, in Handbook of numerical analysis. Vol XIII. Special volume: Numerical methods in electromagnetics., Amsterdam: Elsevier/North Holland, 2005, pp. 317–441.

[42] F. Brezzi, L. D. Marini, and P. Pietra, *Numerical simulation of semiconductor devices*, Comput. Methods Appl. Mech. Eng., 75 (1989), pp. 493–514.

[43] ——, *Two-dimensional exponential fitting and applications to drift-diffusion models*, SIAM J. Numer. Anal., 26 (1989), pp. 1342–1355.

[44] D. Browne, B. Mazumder, Y.-R. Wu, and J. Speck, *Electron transport in unipolar InGaN/GaN multiple quantum well structures grown by NH_3 molecular beam epitaxy*, J. Appl. Phys., 117 (2015), p. 185703.

[45] D. A. Browne, M. N. Fireman, B. Mazumder, L. Y. Kuritzky, Y.-R. Wu, and J. S. Speck, *Vertical transport through AlGaN barriers in heterostructures grown by ammonia molecular beam epitaxy and metalorganic chemical vapor deposition*, Semiconductor Science and Technology, 32 (2017), p. 025010.

[46] C. Buet and S. Dellacherie, *On the Chang and Cooper scheme applied to a linear Fokker-Planck equation*, Commun. Math. Sci., 8 (2010), pp. 1079–1090.

[47] M. Burger, M. Di Francesco, J.-F. Pietschmann, and B. Schlake, *Nonlinear cross-diffusion with size exclusion*, SIAM J. Math. Anal., 42 (2010), pp. 2842–2871.

[48] J.-S. Camier and F. Hermeline, *A monotone nonlinear finite volume method for approximating diffusion operators on general meshes*, Int. J. Numer. Methods Eng., 107 (2016), pp. 496–519.

[49] C. Cancès, *Energy stable numerical methods for porous media flow type problems*, Oil & Gas Science and Technology – Rev. IFP Énergies nouvelles, 73 (2018).

[50] C. Cancès, C. Chainais-Hillairet, J. Fuhrmann, and B. Gaudeul, *A numerical-analysis-focused comparison of several finite volume schemes for a unipolar degenerate drift-diffusion model*, IMA J. Numer. Anal., 41 (2021), pp. 271–314.

[51] C. Cancès, C. Chainais-Hillairet, M. Herda, and S. Krell, *Large time behavior of nonlinear finite volume schemes for convection-diffusion equations*, SIAM J. Numer. Anal., 58 (2020), pp. 2544–2571.

[52] 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*, Comput. Methods Appl. Math., 18 (2018), pp. 407–432.

[53] C. Cancès, C. Chainais-Hillairet, B. Merlet, F. Raimondi, and J. Venel, *Mathematical analysis of a thermodynamically consistent reduced model for iron corrosion*, Z. Angew. Math. Phys., 74 (2023), p. 31. Id/No 96.

[54] C. Cancès and B. Gaudeul, *A convergent entropy diminishing finite volume scheme for a cross-diffusion system*, SIAM J. Numer. Anal., 58 (2020), pp. 2684–2710.

[55] 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 (2016), pp. 549–580.

[56] ——, *Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure*, Found. Comput. Math., 17 (2017), pp. 1525–1584.

[57] C. Cancès, F. Nabet, and M. Vohralík, *Convergence and a posteriori error analysis for energy-stable finite element approximations of degenerate parabolic equations*, Math. Comput., 90 (2021), pp. 517–563.

[58] X. Cao and H. Huang, *An adaptive conservative finite volume method for Poisson-Nernst-Planck equations on a moving mesh*, Commun. Comput. Phys., 26 (2019), pp. 389–412.

[59] M. A. Caro, S. Schulz, S. B. Healy, and E. P. O'Reilly, *Built-in field control in alloyed c-plane III-N quantum dots and wells*, J. Appl. Phys., 109 (2011), p. 084110.

[60] M. A. Caro, S. Schulz, and E. P. O'Reilly, *Theory of local electric polarization and its relation to internal strain: impact on the polarization potential and electronic properties of group-III nitrides*, Phys. Rev. B, 88 (2013), p. 214103.

[61] J. A. Carrillo, A. Jüngel, P. A. Markowich, G. Toscani, and A. Unterreiter, *Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities*, Monatsh. Math., 133 (2001), pp. 1–82.

[62] J. A. Carrillo and G. Toscani, *Exponential convergence toward equilibrium for homogeneous Fokker-Planck-type equations*, Math. Methods Appl. Sci., 21 (1998), pp. 1269–1286.

[63] ——, *Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity*, Indiana Univ. Math. J., 49 (2000), pp. 113–142.

[64] C. Chainais-Hillairet, *Discrete duality finite volume schemes for two-dimensional drift-diffusion and energy-transport models*, Int. J. Numer. Methods Fluids, 59 (2009), pp. 239–257.

[65] C. Chainais-Hillairet and J. Droniou, *Convergence analysis of a mixed finite volume scheme for an elliptic-parabolic system modeling miscible fluid flows in porous media*, SIAM J. Numer. Anal., 45 (2007), pp. 2228–2258.

[66] ——, *Finite-volume schemes for noncoercive elliptic problems with Neumann boundary conditions*, IMA J. Numer. Anal., 31 (2011), pp. 61–85.

[67] C. Chainais-Hillairet, R. Eymard, and J. Fuhrmann, *A monotone numerical flux for quasilinear convection diffusion equation*, arXiv preprint arXiv:2209.11657, (2022).

[68] C. Chainais-Hillairet and F. Filbet, *Asymptotic behaviour of a finite-volume scheme for the transient drift-diffusion model*, IMA J. Numer. Anal., 27 (2007), pp. 689–716.

[69] C. Chainais-Hillairet and M. Herda, *Large-time behaviour of a family of finite volume schemes for boundary-driven convection-diffusion equations*, IMA J. Numer. Anal., 40 (2020), pp. 2473–2504.

[70] C. Chainais-Hillairet, M. Herda, S. Lemaire, and J. Moatti, *Long-time behaviour of hybrid finite volume schemes for advection-diffusion equations: linear and nonlinear approaches*, Numer. Math., 151 (2022), pp. 963–1016.

[71] C. Chainais-Hillairet, S. Krell, and A. Mouton, *Study of discrete duality finite volume schemes for the Peaceman model*, SIAM J. Sci. Comput., 35 (2013), pp. a2928–a2952.

[72] ——, *Convergence analysis of a DDFV scheme for a system describing miscible fluid flows in porous media*, Numer. Methods Partial Differ. Equations, 31 (2015), pp. 723–760.

[73] J. Chang and G. Cooper, *A practical difference scheme for fokker-planck equations*, Journal of Computational Physics, 6 (1970), pp. 1–16.

[74] D. Chaudhuri, J. C. Kelleher, M. R. O'Brien, E. P. O'Reilly, and S. Schulz, *Electronic structure of semiconductor nanostructures: A modified localization landscape theory*, Physical Review B, 101 (2020), p. 035430.

[75] D. Chaudhuri, M. O'Donovan, T. Streckenbach, O. Marquardt, P. Farrell, S. K. Patra, T. Koprucki, and S. Schulz, *Multiscale simulations of the electronic structure of III-nitride quantum wells with varied indium content: Connecting atomistic and continuum-based models*, J. Appl. Phys., 129 (2021), p. 073104.

[76] F. Chave, D. A. Di Pietro, and S. Lemaire, *A discrete Weber inequality on three-dimensional hybrid spaces with application to the HHO approximation of magnetostatics*, Math. Models Methods Appl. Sci., 32 (2022), pp. 175–207.

[77] G. Chen, P. Monk, and Y. Zhang, *An HDG method for the time-dependent drift-diffusion model of semiconductor devices*, J. Sci. Comput., 80 (2019), pp. 420–443.

[78] Y. Chen and B. Cockburn, *Analysis of variable-degree HDG methods for convection-diffusion equations. I: General nonconforming meshes*, IMA J. Numer. Anal., 32 (2012), pp. 1267–1293.

[79] ——, *Analysis of variable-degree HDG methods for convection-diffusion equations. II: Semi-matching nonconforming meshes*, Math. Comput., 83 (2014), pp. 87–111.

[80] H. M. Cheng and J. t. T. Boonkkamp, *A generalised complete flux scheme for anisotropic advection-diffusion equations*, Adv. Comput. Math., 47 (2021).

[81] H. M. Cheng, J. ten thije Boonkkamp, J. Janssen, D. Mihailova, and J. van Dijk, *Combining the hybrid mimetic mixed method with the scharfetter-gummel scheme for magnetised transport in plasmas*, 2022.

[82] S. F. Chichibu, A. Uedono, T. Onuma, B. A. Haskell, A. Chakraborty, T. Koyama, P. T. Fini, S. Keller, S. P. DenBaars, J. S. Speck, U. K. Mishra, S. Nakamura, S. Yamaguchi, S. Kamiyama, H. Amano, I. Akasaki, J. Han, and T. Sota, *Origin of defect-insensitive emission probability in In-containing (Al,In,Ga)N alloy semiconductors*, Nature Mater., 5 (2006), p. 810.

[83] M. Cicuttin, A. Ern, and N. Pignet, *Hybrid high-order methods. A primer with applications to solid mechanics*, SpringerBriefs Math., Cham: Springer, 2021.

[84] B. Cockburn, D. A. Di Pietro, and A. Ern, *Bridging the hybrid high-order and hybridizable discontinuous Galerkin methods*, ESAIM, Math. Model. Numer. Anal., 50 (2016), pp. 635–650.

[85] B. Cockburn, B. Dong, and J. Guzmán, *A superconvergent LDG-hybridizable Galerkin method for second-order elliptic problems*, Math. Comput., 77 (2008), pp. 1887–1916.

[86] B. Cockburn, B. Dong, J. Guzmán, M. Restelli, and R. Sacco, *A hybridizable discontinuous Galerkin method for steady-state convection-diffusion-reaction problems*, SIAM J. Sci. Comput., 31 (2009), pp. 3827–3846.

[87] B. Cockburn, J. Gopalakrishnan, and R. Lazarov, *Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems*, SIAM J. Numer. Anal., 47 (2009), pp. 1319–1365.

[88] L. Codecasa, R. Specogna, and F. Trevisan, *A new set of basis functions for the discrete geometric approach*, J. Comput. Phys., 229 (2010), pp. 7401–7410.

[89] P. Colin, *Numerical analysis of drift-diffusion models: convergence and asymptotic behaviors*, PhD thesis, Université de Lille 1, June 2016.

[90] Y. Coudière and F. Hubert, *A 3d discrete duality finite volume method for nonlinear elliptic equations*, SIAM J. Sci. Comput., 33 (2011), pp. 1739–1764.

[91] M. Darwish and A. Gagliardi, *A drift-diffusion simulation model for organic field effect transistors: on the importance of the gaussian density of states and traps*, J. Phys. D: Appl. Phys., (2020), p. 105102.

[92] A. David, M. J. Grundmann, J. F. Kaeding, N. F. Gardner, T. G. Mihopoulos, and M. R. Krames, *Carrier distribution in (0001)InGaN/GaN multiple quantum well light-emitting diodes*, Applied Physics Letters, 92 (2008), p. 053502.

[93] A. David, N. G. Young, C. A. Hurni, and M. D. Craven, *Quantum Efficiency of III-Nitride Emitters: Evidence for Defect-Assisted Nonradiative Recombination and its Effect on the Green Gap*, Phys. Rev. Applied, 11 (2019), p. 031001.

[94] A. David, N. G. Young, C. Lund, and M. D. Craven, *Review—The Physics of Recombinations in III-Nitride Emitters*, ECS Journal of Solid State Science and Technology, 9 (2019), p. 016021.

[95] P. Dawson, S. Schulz, R. A. Oliver, M. J. Kappers, and C. J. Humphreys, *The nature of carrier localisation in polar and nonpolar InGaN/GaN quantum wells*, Journal of Applied Physics, 119 (2016), p. 181505.

[96] J.-D. Deuschel and D. W. Stroock, *Large deviations*, vol. 137 of Pure and Applied Mathematics, Academic Press, Inc., Boston, MA, 1989.

[97] D. A. Di Pietro and J. Droniou, *The hybrid high-order method for polytopal meshes. Design, analysis, and applications*, vol. 19 of MS&A, Model. Simul. Appl., Cham: Springer, 2020.

[98] D. A. Di Pietro, J. Droniou, and A. Ern, *A discontinuous-skeletal method for advection-diffusion-reaction on general meshes*, SIAM J. Numer. Anal., 53 (2015), pp. 2135–2157.

[99] D. A. Di Pietro and A. Ern, *A hybrid high-order locking-free method for linear elasticity on general meshes*, Comput. Methods Appl. Mech. Eng., 283 (2015), pp. 1–21.

[100] D. A. Di Pietro, A. Ern, and S. Lemaire, *An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators*, Comput. Methods Appl. Math., 14 (2014), pp. 461–472.

[101] D. A. Di Pietro and S. Lemaire, *An extension of the Crouzeix–Raviart space to general meshes with application to quasi-incompressible linear elasticity and Stokes flow*, Math. Comp., 84 (2015), pp. 1–31.

[102] A. Di Vito, A. Pecchia, A. Di Carlo, and M. Auf Der Maur, *Characterization of non-uniform InGaN alloys: spatial localization of carriers and optical properties*, Japanese Journal of Applied Physics, 58 (2019), p. SCCC03.

[103] ——, *Simulating random alloy effects in III-nitride light emitting diodes*, J. Appl. Phys., 128 (2020), p. 041102.

[104] D. H. Doan, P. Farrell, J. Fuhrmann, M. Kantner, T. Koprucki, and N. Rotundo, *ddfermi – a drift-diffusion simulation tool*, ddfermi – a drift-diffusion simulation tool, Weierstrass Institute (WIAS), doi: http://doi.org/10.20347/WIAS.SOFTWARE.DDFERMI, 2020.

[105] K. Domelevo and P. Omnes, *A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids*, ESAIM, Math. Model. Numer. Anal., 39 (2005), pp. 1203–1249.

[106] J. Droniou, *Non-coercive linear elliptic problems*, Potential Analysis, 17 (2002), pp. 181–203.

[107] ——, *Degrés topologiques et applications*, 2006.

[108] ——, *Remarks on discretizations of convection terms in Hybrid Mimetic Mixed methods*, Networks & Heterogeneous Media, 5 (2010), pp. 545–563.

[109] ——, *Finite volume schemes for diffusion equations: Introduction to and review of modern methods*, Math. Models Methods Appl. Sci., 24 (2014), pp. 1575–1619.

[110] J. Droniou and R. Eymard, *A mixed finite volume scheme for anisotropic diffusion problems on any grid*, Numer. Math., 105 (2006), pp. 35–71.

[111] J. Droniou, R. Eymard, T. Gallouët, C. Guichard, and R. Herbin, *The Gradient Discretisation Method*, vol. 82 of Mathématiques & Applications, Springer International Publishing, Cham, Switzerland, 2018.

[112] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin, *A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods*, Math. Models Methods Appl. Sci., 20 (2010), pp. 265–295.

[113] J. Droniou and C. le Potier, *Construction and convergence study of schemes preserving the elliptic local maximum principle*, SIAM J. Numer. Anal., 49 (2011), pp. 459–490.

[114] J. Droniou and J.-L. Vázquez, *Noncoercive convection-diffusion elliptic problems with Neumann boundary conditions*, Calc. Var., 34 (2009), pp. 413–434.

[115] D. A. Dunavant, *High degree efficient symmetrical Gaussian quadrature rules for the triangle*, Int. J. Numer. Methods Eng., 21 (1985), pp. 1129–1148.

[116] A. Dussaigne, F. Barbier, B. Damilano, S. Chenot, A. Grenier, A. M. Papon, B. Samuel, B. Ben Bakir, D. Vaufrey, J. C. Pillet, A. Gasse, O. Ledoux, M. Rozhavskaya, and D. Sotta, *Full InGaN red light emitting diodes*, Journal of Applied Physics, 128 (2020), p. 135704.

[117] C. Erignoux, *Hydrodynamic limit of boundary driven exclusion processes with nonreversible boundary dynamics*, J. Stat. Phys., 172 (2018), pp. 1327–1357.

[118] A. Ern and J.-L. Guermond, *Finite Elements I: Approximation and Interpolation*, Springer, Feb. 2021.

[119] L. C. Evans, *Partial Differential Equations: Second Edition*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, R.I., 2010.

[120] R. Eymard, J. Fuhrmann, and K. Gärtner, *A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local Dirichlet problems*, Numer. Math., 102 (2006), pp. 463–495.

[121] R. Eymard, T. Gallouët, and R. Herbin, *Finite volume methods*, in Techniques of Scientific Computing (Part 3), Handb. Numer. Anal., VII, North-Holland, Amsterdam, 2000, pp. 713–1020.

[122] ——, *Benchmark on anisotropic problems. SUSHI: a scheme using stabilization and hybrid interfaces for anisotropic heterogeneous diffusion problems*, in Finite Volumes for Complex Applications V - Problems & Perspectives, R. Eymard and J.-M. Hérard, eds., ISTE, London, 2008, pp. 801–814.

[123] ——, *Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes. SUSHI: a scheme using stabilization and hybrid interfaces*, IMA J. Numer. Anal., 30 (2010), pp. 1009–1043.

[124] R. Eymard, C. Guichard, and R. Herbin, *Small-stencil 3d schemes for diffusive flows in porous media*, ESAIM, Math. Model. Numer. Anal., 46 (2012), pp. 265–290.

[125] 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 (2017), pp. 497–513.

[126] P. Farrell, J. Moatti, M. O'Donovan, S. Schulz, and T. Koprucki, *Importance of satisfying thermodynamic consistency in optoelectronic device simulations for high carrier densities*, Optical and Quantum Electronics, 55 (2023), p. 978.

[127] P. Farrell, M. Patriarca, J. Fuhrmann, and T. Koprucki, *Comparison of thermodynamically consistent charge carrier flux discretizations for Fermi–Dirac and Gauss–Fermi statistics*, Optical and Quantum Electronics, 50 (2018), pp. 1–10.

[128] P. Farrell, N. Rotundo, D. H. Doan, M. Kantner, J. Fuhrmann, and T. Koprucki, *Mathematical Methods: Drift-Diffusion Models*, in Handbook of Optoelectronic Device Modeling and Simulation, J. Piprek, ed., vol. 2, CRC Press, 2017, ch. 50, pp. 733–771.

[129] D. Ferry, S. Ramey, L. Shifren, and R. Akis, *The effective potential in device modeling: The good, the bad and the ugly*, Journal of Computational Electronics, 1 (2002), pp. 59–65.

[130] D. A. Fick, *On liquid diffusion*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 10 (1855), pp. 30–39.

[131] F. Filbet and M. Herda, *A finite volume scheme for boundary-driven convection-diffusion equations with relative entropy structure*, Numer. Math., 137 (2017), pp. 535–577.

[132] M. Filoche, M. Piccardo, Y.-R. Wu, C.-K. Li, C. Weisbuch, and S. Mayboroda, *Localization landscape theory of disorder in semiconductors. I. Theory and modeling*, Physical Review B, 95 (2017), p. 144204.

[133] R. Finn, M. O'Donovan, P. Farrell, J. Moatti, T. Streckenbach, T. Koprucki, and S. Schulz, *Theoretical study of the impact of alloy disorder on carrier transport and recombination processes in deep UV (Al,Ga)N light emitters*, Applied Physics Letters, 122 (2023). 241104.

[134] R. Finn and S. Schulz, *Impact of random alloy fluctuations on the electronic and optical properties of (Al,Ga)N quantum wells: Insights from tight-binding calculations*, J. Chem. Phys., 157 (2022), p. 244705.

[135] R. A. Fisher, *The wave of advance of advantageous genes*, Annals of Eugenics, 7 (1937), pp. 355–369.

[136] F. Flandoli, F. Russo, and J. Wolf, *Some SDEs with distributional drift. I: General calculus*, Osaka J. Math., 40 (2003), pp. 493–542.

[137] ——, *Some SDEs with distributional drift. II: Lyons-Zheng structure, Itô's formula and semimartingale characterization*, Random Oper. Stoch. Equ., 12 (2004), pp. 145–184.

[138] J. Fourier, *Théorie analytique de la chaleur*, Manuscripta; History of science, 18th and 19th century, Chez Firmin Didot, père et fils, 1822.

[139] G. Fu, W. Qiu, and W. Zhang, *An analysis of HDG methods for convection-dominated diffusion problems*, ESAIM, Math. Model. Numer. Anal., 49 (2015), pp. 225–256.

[140] J. Fuhrmann, *Comparison and numerical treatment of generalised Nernst-Planck models*, Comput. Phys. Commun., 196 (2015), pp. 166–178.

[141] J. Fuhrmann, T. Streckenbach, et al., *pdelib: A finite volume and finite element toolbox for pdes. [software].*, pdelib: a finite volume and finite element toolbox for pdes. [software]. version: 2.4.20190405, Weierstrass Institute (WIAS), http://pdelib.org, 2019.

[142] H. Gajewski, *On existence, uniqueness and asymptotic behavior of solutions of the basic equations for carrier transport in semiconductors*, Z. Angew. Math. Mech., 65 (1985), pp. 101–108.

[143] H. Gajewski and K. Gärtner, *On the discretization of van Roosbroeck's equations with magnetic field*, Z. Angew. Math. Mech., 76 (1996), pp. 247–264.

[144] H. Gajewski and K. Gröger, *Semiconductor equations for variable mobilities based on Boltzmann statistics or Fermi-Dirac statistics*, Math. Nachr., 140 (1989), pp. 7–36.

[145] B. Galler, A. Laubsch, A. Wojcik, H. Lugauer, A. Gomez-Iglesias, M. Sabathil, and B. Hahn, *Investigation of the carrier distribution in InGaN-based multi-quantum-well structures*, physica status solidi c, 8 (2011), pp. 2372–2374.

[146] T. Gallouët, A. Natale, and G. Todeschi, *From geodesic extrapolation to a variational BDF2 scheme for Wasserstein gradient flows*. working paper or preprint, May 2023.

[147] B. Gaudeul and J. Fuhrmann, *Entropy and convergence analysis for two finite volume schemes for a Nernst-Planck-Poisson system with ion volume constraints*, Numer. Math., 151 (2022), pp. 99–149.

[148] J. Glad, J. Ruschel, N. L. Bloch, H. K. Cho, F. Mahnke, L. Sulmoni, M. Guttmann, T. Wernicke, M. Weyers, S. Einfeldt, and M. Kneissl, *Impact of operation parameters on the degradation of 233 nm AlGaN-based far UVC LEDs*, J. Appl. Phys., 131 (2022), p. 014501.

[149] A. Glitzky, M. Liero, and G. Nika, *An existence result for a class of electrothermal drift-diffusion models with Gauss-Fermi statistics for organic semiconductors*, Analysis and Applications, 19 (2021), pp. 275–304.

[150] S. K. Godunov and I. Bohachevsky, *Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics*, Matematičeskij sbornik, 47(89) (1959), pp. 271–306.

[151] S. Gottlieb, *On high order strong stability preserving Runge-Kutta and multi step time discretizations*, J. Sci. Comput., 25 (2005), pp. 105–128.

[152] S. Gottlieb, D. Ketcheson, and C.-W. Shu, *Strong stability preserving Runge-Kutta and multistep time discretizations.*, Hackensack, NJ: World Scientific, 2011.

[153] A. Guionnet and B. Zegarlinski, *Lectures on Logarithmic Sobolev Inequalities*, Séminaire de probabilités de Strasbourg, 36 (2002), pp. 1–134.

[154] R. Herbin and F. Hubert, *Benchmark on discretization schemes for anisotropic diffusion problems on general grids*, in Finite Volumes for Complex Applications V - Problems & Perspectives, R. Eymard and J.-M. Hérard, eds., ISTE, London, 2008, pp. 659–692.

[155] M. Herda, *On massless electron limit for a multispecies kinetic system with external magnetic field*, J. Differ. Equations, 260 (2016), pp. 7861–7891.

[156] M. Herda and A. Zurek, *Study of an entropy dissipating finite volume scheme for a nonlocal cross-diffusion system*, ESAIM: M2AN, 57 (2023), pp. 1589–1617.

[157] F. Hermeline, *A finite volume method for the approximation of diffusion operators on distorted meshes*, Journal of Computational Physics, 160 (2000), pp. 481–499.

[158] C. J. Humphreys, *Solid-State Lighting*, MRS Bulletin, 33 (2008), p. 459.

[159] J.-I. Hwang, R. Hashimoto, S. Saito, and S. Nunoue, *Development of InGaN-based red LED grown on (0001) polar surface*, Applied Physics Express, 7 (2014), p. 071003.

[160] A. M. Il'in, *Differencing scheme for a differential equation with a small parameter affecting the highest derivative*, Mathematical Notes of the Academy of Sciences of the USSR, 6 (1969), pp. 596–602.

[161] E. Issoglio, *Transport equations with fractal noise-existence, uniqueness and regularity of the solution*, Z. Anal. Anwend., 32 (2013), pp. 37–53.

[162] ——, *A non-linear parabolic PDE with a distributional coefficient and its applications to stochastic analysis*, J. Differ. Equations, 267 (2019), pp. 5976–6003.

[163] F. Jochmann, *Existence of weak solutions of the drift diffusion model coupled with Maxwell's equations*, J. Math. Anal. Appl., 204 (1996), pp. 655–676.

[164] ——, *Galerkin approximation of weak solutions of the drift diffusion model for semiconductors coupled with Maxwell's equations*, Math. Methods Appl. Sci., 19 (1996), pp. 1471–1488.

[165] ——, *A singular limit in the drift diffusion model for semiconductors coupled with Maxwell's equations*, Appl. Anal., 67 (1997).

[166] ——, *Convergence to stationary states of solutions of the transient drift diffusion equations for semiconductor devices with prescribed currents*, Asymptotic Anal., 18 (1998), pp. 67–109.

[167] ——, *Uniqueness and regularity for the two-dimensional drift-diffusion model for semiconductors coupled with Maxwell's equations*, J. Differ. Equations, 147 (1998), pp. 242–270.

[168] ——, *Some analytical results concerning the drift diffusion equations for semiconductor devices coupled with Maxwell's equations*, in Hyperbolic problems: Theory, numerics, applications. Proceedings of the 7th international conference, Zürich, Switzerland, February 1998. Vol. II, Basel: Birkhäuser, 1999, pp. 525–534.

[169] C. M. Jones, C.-H. Teng, Q. Yan, P.-C. Ku, and E. Kioupakis, *Impact of carrier localization on recombination in InGaN quantum wells and the efficiency of nitride light-emitting diodes: Insights from theory and numerical simulations*, Applied Physics Letters, 111 (2017), p. 113501.

[170] C. Jourdana, A. Jüngel, and N. Zamponi, *Three-species drift-diffusion models for memristors*. working paper or preprint, Apr. 2022.

[171] A. Jüngel, *Numerical approximation of a drift-diffusion model for semiconductors with nonlinear diffusion*, ZAMM, 75 (1995), pp. 783–799.

[172] ——, *Quasi-hydrodynamic semiconductor equations*, vol. 41 of Prog. Nonlinear Differ. Equ. Appl., Basel: Birkhäuser, 2001.

[173] ——, *The boundedness-by-entropy method for cross-diffusion systems*, Nonlinearity, 28 (2015), pp. 1963–2001.

[174] ——, *Entropy Methods for Diffusive Partial Differential Equations*, SpringerBriefs in Mathematics, Springer International Publishing, Cham, Switzerland, 2016.

[175] A. Jüngel and P. Pietra, *A discretization scheme for a quasi-hydrodynamic semiconductor model*, Math. Models Methods Appl. Sci., 7 (1997), pp. 935–955.

[176] A. Jüngel and S. Schuchnigg, *Entropy-dissipating semi-discrete Runge-Kutta schemes for nonlinear diffusion equations*, Commun. Math. Sci., 15 (2017), pp. 27–53.

[177] A. Jüngel and A. Zurek, *A convergent structure-preserving finite-volume scheme for the Shigesada-Kawasaki-Teramoto population system*, SIAM J. Numer. Anal., 59 (2021), pp. 2286–2309.

[178] S. Karpov, *ABC-model for interpretation of internal quantum efficiency and its droop in III-nitride LEDs: a review.*, Optical and Quantum Electronics, 47 (2015), pp. 1293–1303.

[179] E. F. Keller and L. A. Segel, *Initiation of slime mold aggregation viewed as an instability*, Journal of Theoretical Biology, 26 (1970), pp. 399–415.

[180] E. Kioupakis, D. Steiauf, P. Rinke, K. T. Delaney, and C. G. V. de Walle, *First-principles calculations of indirect Auger recombination in nitride semiconductors*, Phys. Rev. B, 92 (2015), p. 035207.

[181] E. Kioupakis, Q. Yan, D. Steiauf, and C. G. Van de Walle, *Temperature and carrier-density dependence of Auger and radiative recombination in nitride optoelectronic devices*, New Journal of Physics, 15 (2013), p. 125006.

[182] M. Kneissl, T.-Y. Seong, J. Han, and H. Amano, *The emergence and prospects of deep-ultraviolet light-emitting diode technologies*, Nat. Photonics, 13 (2019), p. 233.

[183] A. Kolmogorov, I. Petrovskii, and N. Piscounov, *A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem.*, Moscow University Bulletin of Mathematics, 1 (1937).

[184] T. Koprucki and K. Gärtner, *Discretization scheme for drift-diffusion equations with strong diffusion enhancement*, Opt Quant Electron, 45 (2013), pp. 791–796.

[185] ——, *Generalization of the scharfetter-gummel scheme*, in Numerical Simulation of Optoelectronic Devices (NUSOD), 2013 13th International Conference on, 2013, pp. 85–86.

[186] T. Koprucki, N. Rotundo, P. Farrell, D. Doan, and J. Fuhrmann, *On thermodynamic consistency of a scharfetter-gummel scheme based on a modified thermal voltage for drift-diffusion equations with diffusion enhancement*, Opt Quant Electron, 47 (2015), pp. 1327–1332.

[187] S. Krell and J. Moatti, *Structure-preserving schemes for drift-diffusion systems on general meshes: DDFV vs HFV*, in Finite Volumes for Complex Applications X - elliptic and parabolic problems, 2023. in press.

[188] P. D. Lax, *Weak solutions of nonlinear hyperbolic equations and their numerical computation*, Commun. Pure Appl. Math., 7 (1954), pp. 159–193.

[189] C. Le Bris, F. Legoll, and F. Madiot, *Stabilisation de problèmes non coercifs via une méthode numérique utilisant la mesure invariante (Stabilization of non-coercive problems using the invariant measure)*, Comptes Rendus Mathématique, 354 (2016), pp. 799–803.

[190] C. Lehrenfeld, *Hybrid Discontinuous Galerkin methods for solving incompressible flow problems*, PhD thesis, Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen, 05 2010.

[191] C. Lehrenfeld and J. Schöberl, *High order exactly divergence-free Hybrid Discontinuous Galerkin methods for unsteady incompressible flows*, Comput. Methods Appl. Mech. Eng., 307 (2016), pp. 339–361.

[192] S. Lemaire, *Bridging the hybrid high-order and virtual element methods*, IMA J. Numer. Anal., 41 (2021), pp. 549–593.

[193] G. Lheureux, C. Lynsky, Y. Wu, J. Speck, and C. Weisbuch, *A 3d simulation comparison of carrier transport in green and blue c-plane multi-quantum well nitride light emitting diodes*, J Appl Phys, 128 (2020), p. 235703.

[194] C.-K. Li, M. Piccardo, L.-S. Lu, S. Mayboroda, L. Martinelli, J. Peretti, J. S. Speck, C. Weisbuch, M. Filoche, and Y.-R. Wu, *Localization landscape theory of disorder in semiconductors. III. Application to carrier transport and recombination in light emitting diodes*, Phys. Rev. B, 95 (2017), p. 144206.

[195] L. Li and J.-G. Liu, *Large time behaviors of upwind schemes and B-schemes for Fokker-Planck equations on $\mathbb{R}$ by jump processes*, Mathematics of Computation, (2020).

[196] H. Liu and Z. Wang, *An entropy satisfying discontinuous Galerkin method for nonlinear Fokker-Planck equations*, J. Sci. Comput., 68 (2016), pp. 1217–1240.

[197] H. Liu, Z. Wang, P. Yin, and H. Yu, *Positivity-preserving third order DG schemes for Poisson-Nernst-Planck equations*, J. Comput. Phys., 452 (2022), p. 22. Id/No 110777.

[198] H. Liu and H. Yu, *Maximum-principle-satisfying third order discontinuous Galerkin schemes for Fokker-Planck equations*, SIAM J. Sci. Comput., 36 (2014), pp. a2296–a2325.

[199] J. P. Liu, J.-H. Ryou, R. D. Dupuis, J. Han, G. D. Shen, and H. B. Wang, *Barrier effect on hole transport and carrier distribution in InGaN/GaN multiple quantum well visible light-emitting diodes*, Applied Physics Letters, 93 (2008), p. 021102.

[200] Y. Liu and C.-W. Shu, *Analysis of the local discontinuous Galerkin method for the drift-diffusion model of semiconductor devices*, Sci. China, Math., 59 (2016), pp. 115–140.

[201] Q. Lv, J. Liu, C. Mo, J. Zhang, X. Wu, Q. Wu, and F. Jiang, *Realization of Highly Efficient InGaN Green LEDs with Sandwich-like Multiple Quantum Well Structure: Role of Enhanced Interwell Carrier Transport*, ACS Photonics, 6 (2019), pp. 130–138.

[202] C. Lynsky, A. Alhassan, G. Lheureux, B. Bonef, S. DenBaars, S. Nakamura, Y. Wu, C. Weisbuch, and J. Speck, *Barriers to carrier transport in multiple quantum well nitride-based c-plane green light emitting diodes*, Phys Rev Materials, 4 (2020), p. 054604.

[203] C. Lynsky, G. Lheureux, B. Bonef, K. S. Qwah, R. C. White, S. P. DenBaars, S. Nakamura, Y.-R. Wu, C. Weisbuch, and J. S. Speck, *Improved Vertical Carrier Transport for Green III-Nitride LEDs Using* (In, Ga)N *Alloy Quantum Barriers*, Phys. Rev. Applied, 17 (2022), p. 054048.

[204] R. Macneal, *An asymmetrical finite difference network*, Quart Math Appl, 11 (1953), pp. 295–310.

[205] F. Madiot, *Multiscale finite element methods for advection-diffusion problems*, PhD thesis, Université Paris-Est, December 2016.

[206] A. J. Majda and A. L. Bertozzi, *Vorticity and incompressible flow*, Camb. Texts Appl. Math., Cambridge: Cambridge University Press, 2002.

[207] P. A. Markowich, *The stationary semiconductor device equations*, Computational Microelectronics, Springer-Verlag, Vienna, 1986.

[208] P. A. Markowich, C. A. Ringhofer, and C. Schmeiser, *Semiconductor equations*, Wien: Springer-Verlag, 1990.

[209] P. A. Markowich and A. Unterreiter, *Vacuum solutions of a stationary drift-diffusion model*, Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, Ser. 4, 20 (1993), pp. 371–386.

[210] J. M. McMahon, E. Kioupakis, and S. Schulz, *Atomistic analysis of Auger recombination in c-plane (In,Ga)N/GaN quantum wells: Temperature-dependent competition between radiative and nonradiative recombination*, Physical Review B, 105 (2022), p. 195307.

[211] J. M. McMahon, D. S. P. Tanner, E. Kioupakis, and S. Schulz, *Atomistic analysis of radiative recombination rate, Stokes shift and density of states in c-plane InGaN/GaN qunatum wells*, Appl. Phys. Lett., 116 (2020), p. 181104.

[212] J. Moatti, *A skeletal high-order structure preserving scheme for advection-diffusion equations*, in Finite Volumes for Complex Applications X - elliptic and parabolic problems, 2023. in press.

[213] ——, *A structure preserving hybrid finite volume scheme for semiconductor models with magnetic field on general meshes*, ESAIM: M2AN, 57 (2023), pp. 2557–2593.

[214] M. S. Mock, *An initial value problem from semiconductor device theory*, SIAM J. Math. Anal., 5 (1974), pp. 597–612.

[215] ——, *An example of nonuniqueness of stationary solutions in semiconductor device models*, COMPEL-The international journal for computation and mathematics in electrical and electronic engineering, 1 (1982), pp. 165–174.

[216] ——, *Analysis of mathematical models of semiconductor devices*, vol. 3 of Adv. Numer. Comput. Ser., Boole Press, Dublin, 1983.

[217] G. Muziol, H. Turski, M. Siekacz, K. Szkudlarek, L. Janicki, M. Baranowski, S. Zolud, R. Kudrawiec, T. Suski, and C. Skierbiszewski, *Beyond Quantum Efficiency Limitations Originating from the Piezoelectric Polarization in Light-Emitting Devices*, ACS Photonics, 6 (2019), pp. 1963–1971.

[218] W. Nernst, *Die elektromotorische wirksamkeit der jonen*, Zeitschrift für physikalische Chemie, 4 (1889), pp. 129–181.

[219] M. O'Donovan, P. Farrell, J. Moatti, T. Streckenbach, T. Koprucki, and S. Schulz, *Impact of random alloy fluctuations on the carrier distribution in multi-color (In,Ga)N/GaN quantum well systems*. working paper or preprint, Sept. 2022.

[220] M. O'Donovan, M. Luisier, E. P. O'Reilly, and S. Schulz, *Impact of random alloy fluctuations on inter-well transport in InGaN/GaN multi-quantum well systems: an atomistic non-equilibrium Green's function study*, J. Phys.: Condens. Matter, 33 (2021), p. 045302.

[221] M. O'Donovan, D. Chaudhuri, T. Streckenbach, P. Farrell, S. Schulz, and T. Koprucki, *From atomistic tight-binding theory to macroscale drift–diffusion: Multiscale modeling and numerical simulation of uni-polar charge transport in (In,Ga)N devices with random fluctuations*, Journal of Applied Physics, 130 (2021), p. 065702.

[222] M. O'Donovan, P. Farrell, T. Streckenbach, T. Koprucki, and S. Schulz, *Multiscale simulations of uni-polar hole transport in (In,Ga)N quantum well systems*, Optical and Quantum Electronics, 54 (2022), p. 405.

[223] 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 (2010). 104501.

[224] P. Pelletier, D. Delande, V. Josse, A. Aspect, S. Mayboroda, D. N. Arnold, and M. Filoche, *Spectral functions and localization-landscape theory in speckle potentials*, Phys. Rev. A, 105 (2022), p. 023314.

[225] M. Piccardo, C.-K. Li, Y.-R. Wu, J. S. Speck, B. Bonef, R. M. Farrell, M. Filoche, L. Martinelli, J. Peretti, and C. Weisbuch, *Localization landscape theory of disorder in semiconductors. II. Urbach tails of disordered quantum well layers*, Phys. Rev. B, 95 (2017), p. 144205.

[226] J. Piprek, *Efficiency droop in nitride-based light-emitting diodes*, Phys. Status Solidi A, 207 (2010), pp. No. 10, 2217–2225.

[227] M. Planck, *Ueber die erregung von electricität und wärme in electrolyten*, Annalen der Physik, 275 (1890), pp. 161–186.

[228] O. Purbo, D. Cassidy, and S. Chisholm, *Numerical model for degenerate and heterostructure semiconductor devices*, J Appl Phys, 66 (1989), pp. 5078–5082.

[229] W. Qiu and K. Shi, *An HDG method for convection diffusion equation*, J. Sci. Comput., 66 (2016), pp. 346–357.

[230] J. Quastel, *Diffusion of color in the simple exclusion process*, Commun. Pure Appl. Math., 45 (1992), pp. 623–679.

[231] K. S. Qwah, M. Monavarian, G. Lheureux, J. Wang, Y.-R. Wu, and J. S. Speck, *Theoretical and experimental investigations of vertical hole transport through unipolar AlGaN structures: Impacts of random alloy disorder*, Applied Physics Letters, 117 (2020), p. 022107.

[232] X. Ren, Z. Wang, W. Sha, and W. Choy, *Exploring the way to approach the efficiency limit of perovskite solar cells by drift-diffusion model*, ACS Photonics, 4 (2017), pp. 934–942.

[233] N. Roccato, F. Piva, C. D. Santi, M. Buffolo, M. Fregolent, M. Pilati, N. Susilo, D. H. Vidal, A. Muhin, L. Sulmoni, T. Wernicke, M. Kneissl, G. Meneghesso, E. Zanoni, and M. Meneghini, *Modeling the electrical degradation of AlGaN-based UV-C LEDs by combined deep-level optical spectroscopy and TCAD simulations*, Appl. Phys. Lett., 122 (2023), p. 161105.

[234] R. Rodriguez-Torres, E. Gutierrez-Dominguez, R. Klima, and S. Selberherr, *Analysis of split-drain magfets*, IEEE Transactions on Electron Devices, 51 (2004), pp. 2237–2245.

[235] F. Römer, M. Guttmann, T. Wernicke, M. Kneissl, and B. Witzigmann, *Effect of inhomogeneous Broadening in Ultraviolet III-Nitride Light-Emitting Diodes*, Materials, 14 (2021), p. 7890.

[236] F. Römer and B. Witzigmann, *Luminescence distribution in the multi-quantum well region of iii-nitride light emitting diodes*, Proc. of SPIE, 10124 (2017), pp. 101240Y–1.

[237] ——, *Signature of the ideality factor in iii-nitride multi quantum well light emitting diodes*, Opt Quant Electron, 50 (2018), pp. 1–10.

[238] F. Russo and G. Trutnau, *Some parabolic PDEs whose drift is an irregular random noise in space*, Ann. Probab., 35 (2007), pp. 2213–2262.

[239] D. L. Scharfetter and H. K. Gummel, *Large-signal analysis of a silicon Read diode oscillator*, IEEE Transactions on Electron Devices, 16 (1969), pp. 64–77.

[240] D. Schiavon, M. Binder, M. Peter, B. Galler, P. Drechsel, and F. Scholz, *Wavelength-dependent determination of the recombination rate coefficients in single-quantum-well GaInN/GaN light emitting diodes*, physica status solidi (b), 250 (2013), pp. 283–290.

[241] M. Schneider, L. Agélas, G. Enchéry, and B. Flemisch, *Convergence of nonlinear finite volume schemes for heterogeneous anisotropic diffusion on general meshes*, J. Comput. Phys., 351 (2017), pp. 80–107.

[242] S. Schulz, M. A. Caro, C. Coughlan, and E. P. O'Reilly, *Atomistic analysis of the impact of alloy and well width fluctuations on the electronic and optical properties of InGaN/GaN quantum wells*, Phys. Rev. B, 91 (2015), p. 035439.

[243] S. Schulz, M. A. Caro, and E. P. O'Reilly, *Impact of cation-based localized electronic states on the conduction and valence band structure of $Al_{1-x}In_xN$ alloys*, Appl. Phys. Lett., 104 (2014), p. 172102.

[244] H.-T. Shen, Y.-C. Chang, and Y.-R. Wu, *Analysis of Light-Emission Polarization Ratio in Deep-Ultraviolet Light-Emitting Diodes by Considering Random Alloy Fluctuations with the 3D k.p Method*, Phys. Status Solidi, 16 (2022), p. 2100498.

[245] Z. Sheng, J. Yue, and G. Yuan, *Monotone finite volume schemes of nonequilibrium radiation diffusion equations on distorted meshes*, SIAM J. Sci. Comput., 31 (2009), pp. 2915–2934.

[246] N. Shigesada, K. Kawasaki, and E. Teramoto, *Spatial segregation of interacting species*, Journal of Theoretical Biology, 79 (1979), pp. 83–99.

[247] H. Si, *TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator*, ACM Transactions on Mathematical Software, 41 (2015), pp. 1–36.

[248] H. Si, K. Gärtner, and J. Fuhrmann, *Boundary conforming delaunay mesh generation*, Comput Math Math Phys, 50 (2010), pp. 38–53.

[249] S. Stodtmann, R. Lee, C. Weiler, and A. Badinski, *Numerical simulation of organic semiconductor devices with high carrier densities*, J Appl Phys, 112 (2012), p. 114909.

[250] S. Su and H. Tang, *A positivity-preserving and free energy dissipative hybrid scheme for the Poisson-Nernst-Planck equations on polygonal and polyhedral meshes*, Comput. Math. Appl., 108 (2022), pp. 33–48.

[251] M. Szymański, D. Tu, and R. Forchheimer, *2-d drift-diffusion simulation of organic electrochemical transistors*, IEEE Transactions on Electron Devices, 64 (2017), pp. 5114–5120.

[252] D. S. P. Tanner, M. A. Caro, E. P. O'Reilly, and S. Schulz, *Random alloy fluctuations and structural inhomogeneities in c-plane $In_xGa_{1-x}N$ quantum wells: theory of ground and excited electron and hole states*, RSC Adv., 6 (2016), pp. 64513–64530.

[253] D. S. P. Tanner, P. Dawson, M. J. Kappers, R. A. Oliver, and S. Schulz, *Polar InGaN/GaN quantum wells: Revisiting the impact of carrier localization on the "green gap" problem*, Phys. Rev. Applied, 13 (2020), p. 044068.

[254] G. Toscani, *Entropy production and the rate of convergence to equilibrium for the Fokker-Planck equation*, Quart. Appl. Math., 57 (1999), pp. 521–541.

[255] W. Tress, K. Leo, and M. Riede, *Optimum mobility, contact properties, and open-circuit voltage of organic solar cells: A drift-diffusion simulation study*, Phys Rev B, (2012), p. 155201.

[256] T.-Y. Tsai, K. Michalczewski, P. Martyniuk, C.-H. Wu, and Y. Wu, *Application of localization landscape theory and the k·p model for direct modeling of carrier transport in a type ii superlattice inas/inassb photoconductor system*, J Appl Phys, 127 (2020), p. 033104.

[257] S. L. M. van Mensfoort and R. Coehoorn, *Effect of gaussian disorder on the voltage dependence of the current density in sandwich-type devices based on organic semiconductors*, Phys. Rev. B, 78 (2008), p. 085207.

[258] W. Van Roosbroeck, *Theory of the flow of electrons and holes in germanium and other semiconductors*, The Bell System Technical Journal, 29 (1950), pp. 560–607.

[259] M. Wahn and J. Neugebauer, *Generalized Wannier functions: An efficient way to construct ab-initio tight-binding parameters for group-III nitrides*, phys. stat. sol. (b), 243 (2006), pp. No. 7, 1583–1587.

[260] D. Watson-Parris, M. J. Godfrey, P. Dawson, R. A. Oliver, M. J. Galtrey, M. J. Kappers, and C. J. Humphreys, *Carrier localization mechanisms in $In_xGa_{1-x}N/GaN$*, Phys. Rev. B, 83 (2011), p. 115321.

[261] C. Weisbuch, S. Nakamura, Y.-R. Wu, and J. S. Speck, *Disorder effects in nitride semiconductors: impact on fundamental and device properties*, Nanophotonics, 10(1) (2021), pp. 3–21.

[262] B. Witzigmann, F. Römer, M. Martens, C. Kuhn, T. Wernicke, and M. Kneissl, *Calculation of optical gain in AlGaN quantum wells for ultraviolet emission*, AIP Advances, 10 (2020), p. 095307.

[263] C.-K. Wu, C.-K. Li, and Y.-R. Wu, *Percolation transport study in nitride based LED by considering the random alloy fluctuation*, J. Comp. Elec., 14 (2015), pp. 416–424.

[264] T.-J. Yang, R. Shivaraman, J. S. Speck, and Y.-R. Wu, *The influence of random indium alloy fluctuations in indium gallium nitride quantum wells on the device behavior*, J. Appl. Phys., 116 (2014), p. 113104.

[265] Z. Yu and R. Dutton, *SEDAN III – A one-dimensional device simulator*. www-tcad.stanford.edu/tcad/programs/sedan3.html, 1988.

[266] J. H. Zhu, S. M. Zhang, H. Wang, D. G. Zhao, J. J. Zhu, Z. S. Liu, D. S. Jiang, Y. X. Qiu, and H. Yang, *The investigation on carrier distribution in InGaN/GaN multiple quantum well layers*, Journal of Applied Physics, 109 (2011), p. 093117.

Table des matières

Développement et analyse de schémas d'ordre élevé pour des modèles de convection-diffusion, étude du comportement en temps long

Résumé

Dans cette thèse, nous nous intéressons à l'approximation numérique de problèmes de convection-diffusion, potentiellement anisotropes, par des schémas d'ordre élevé sur maillages généraux. Notre objectif est de proposer des méthodes fiables, précises et efficaces : les solutions numériques doivent préserver les propriétés physiques des solutions continues (conservation de la masse, positivité des densités, comportement en temps long) tout en autorisant une large gamme de paramètres de discrétisation (pas de temps grands, maillages spatiaux généraux) et en optimisant la précision de calcul à coût donné. Les problèmes considérés sont des équations d'advection-diffusion ainsi que des systèmes couplés de dérive-diffusion qui modélisent les composants semi-conducteurs.

On se concentre d'abord sur une équation d'advection-diffusion seule, pour laquelle nous proposons et analysons trois méthodes d'ordre bas de type volumes finis hybrides (HFV). Cette comparaison met en avant la nécessité d'utiliser un schéma non-linéaire afin de préserver la positivité des solutions, tant d'un point de vue théorique que numérique. On s'intéresse alors à l'approximation d'un système de dérive-diffusion, constitué de deux équations d'advection-diffusion couplées avec une équation de Poisson. Pour ce problème, on introduit un schéma non-linéaire basé sur la méthode précédente, qui préserve les bornes des densités calculées (et en particulier leur positivité) et le comportement en temps long de la solution. Ce schéma HFV pour la dérive-diffusion est ensuite comparé numériquement avec un schéma présentant des propriétés similaires basé sur la méthode volumes finis en dualité discrète (DDFV).

Nous nous intéressons alors à des schémas d'ordre élevé (en espace). Ces schémas sont basés sur des méthodes de type hybride d'ordre élevé (HHO) qui peuvent être interprétées comme des extensions à l'ordre arbitraire des méthodes HFV. Nous introduisons deux méthodes pour les équations d'advection-diffusion linéaires. La première est linéaire, tandis que la seconde est non-linéaire, et permet de préserver la positivité. Pour ces deux schémas, nous prouvons l'existence de solutions discrètes et établissons des résultats de comportement en temps long. Nous confirmons également ces résultats numériquement, et mettons en avant la nécessité d'utiliser une méthode non-linéaire pour préserver la positivité. Par ailleurs, on observe que la méthode non-linéaire converge à l'ordre attendu. De plus, la montée en ordre permet un gain d'efficacité (précision de l'approximation pour un coût de calcul donné) conséquent par rapport aux méthodes d'ordre bas des premiers chapitres.

Ces travaux sont complétés par l'étude de problèmes de convection-diffusion avec convection très irrégulière, effectuée en collaboration avec des physiciens durant la thèse. Les recherches menées visent à comprendre comment concevoir des diodes électroluminescentes efficaces émettant dans l'ultraviolet profond, et soulèvent divers enjeux relatifs à la modélisation, l'analyse numérique et la simulation de ces problèmes.

Mots clés : analyse numérique, méthode d'entropie, méthodes numériques sur maillages généraux, méthodes numériques d'ordre élevé, semi-conducteurs, edp paraboliques

---

Centre Inria de l'Université de Lille – Parc scientifique de la Haute-Borne – 59650 Villeneuve d'Ascq – France
Laboratoire Paul Painlevé – Cité Scientifique – 59655 Villeneuve-d'Ascq – France

Development and numerical analysis of high-order schemes for convection-diffusion models, study of their long-time behaviour

Abstract

In this thesis, we are interested in the numerical approximation of anisotropic convection-diffusion problems using high-order schemes on general meshes. Our objective is to develop reliable, accurate, and efficient methods: the numerical solutions must preserve the physical properties of the continuous solutions (mass conservation, positivity of densities, long-time behaviour) while allowing for a wide range of discretisation parameters (large time steps, general spatial meshes) and optimising the accuracy at a given computational cost. The problems under study are advection-diffusion equations as well as coupled drift-diffusion systems that model semiconductor devices.

We first focus on a single advection-diffusion equation for which we propose and analyse three different hybrid finite volume (HFV) methods to approximate the solution. Their comparison highlights the necessity of using a nonlinear scheme to ensure the preservation of positivity, both from a theoretical and a numerical level. We then consider the numerical approximation of a drift-diffusion system, consisting of two coupled advection-diffusion equations with a Poisson equation. For this problem, we introduce a nonlinear scheme based on the previous method, which preserves the bounds on the computed densities (including their positivity) and the long-time behaviour of the solution. This HFV scheme for drift-diffusion is then numerically compared with another scheme with similar properties based on the discrete duality finite volume (DDFV) method.

We then focus on high-order (in space) schemes. These schemes are based on hybrid high-order (HHO) methods which can be interpreted as an arbitrary-order extension of HFV methods. We introduce two schemes for linear advection-diffusion: the first method is linear, while the second one is nonlinear and preserves the positivity of the solution. For both of these schemes, we prove the existence of discrete solutions and establish results about their long-time behaviour. We also confirm these results numerically and emphasise the need to use a nonlinear method to preserve positivity. We observe that the nonlinear method converges at the expected order. Furthermore, increasing the order leads to a significant gain in efficiency (accuracy of the approximation for a given computational cost) compared to the low-order methods discussed in the first chapters.

These works are complemented with the study of convection-diffusion problems with highly irregular convection, carried out in collaboration with physicists during the thesis. These investigations aim to understand how to design efficient deep ultraviolet-emitting electroluminescent diodes, and raise various issues related to the modeling, numerical analysis and simulation of these problems.

Keywords: numerical analysis, entropy method, numerical methods on general meshes, high-order numerical methods, semiconductors, parabolic pdes