Ondes internes en océanographie et cristaux photoniques : une approche mathématique

Vincent Duchêne

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité :

MATHÉMATIQUES

présentée par

Vincent Duchêne

pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet :

Ondes internes en océanographie et cristaux photoniques : une approche mathématique

Soutenue publiquement le 31 Mai 2011
devant le jury composé de

Didier BRESCH	Rapporteur
Stéphane LABBÉ	Rapporteur
David LANNES	Directeur de thèse
Jean-Michel ROQUEJOFFRE	Examinateur
Laure SAINT-RAYMOND	Examinateur
Nikolay TZVETKOV	Examinateur

Université Pierre et Marie Curie
Paris 6

Ecole Doctorale de Sciences
Mathématiques de Paris Centre

Département de Mathématiques
et Applications – UMR 8553
École Normale Supérieure, Paris

Remerciements

Je tiens à remercier en premier lieu David LANNES, mon directeur de thèse, pour avoir accepté de me guider au cours de ces trois dernières années, et de m'avoir fait découvrir un domaine et des sujets d'étude passionnants. Je n'ose imaginer ce que serait ce manuscrit sans ses conseils avisés, son jugement clairvoyant, sa disponibilité constante. David, je te remercie également pour ta patience, ton indulgence et ta décontraction. Je ne peux que témoigner de la chance que j'ai eue d'avoir un directeur de ce tonneau-là.

Stéphane LABBÉ et Didier BRESCH ont eu l'amabilité de rapporter cette thèse, et je les en remercie chaleureusement. Je remercie également Jean-Michel ROQUEJOFFRE, Laure SAINT-RAYMOND, Nikolay TZVETKOV. C'est un honneur pour moi qu'ils aient accepté de faire partie de mon jury de thèse.

Michael WEINSTEIN guided my very first steps as a researcher, and has never stopped helping me since. I enjoyed truly fruitful discussions, as he introduced me to very diverse and stimulating topics. I had a great pleasure working with him and Jeremy MARZUOLA, and I wish to thank them on this occasion.

J'ai croisé pendant ces trois ans de nombreux condisciples, qui furent d'agréable compagnie durant les séminaires et conférences. Je pense par exemple à Jean-Claude, Catherine, Florent, Samer, Robert, Chloé, Gael, Evelyne, Christophe, Yong, Maxime, *etc.* Leur contact amical n'est pas la moindre des raisons pour lesquelles j'aspire à continuer dans les métiers de la recherche. Qu'ils en soient remerciés.

Être accueilli au sein du DMA fut pour moi une chance inouïe. J'associe à ces remerciements toutes les personnes que j'y ai côtoyé. Je pense en particulier aux ex-habitants du R3, Daniel, Thibaut, Cyrille, Hua, Laure ; et aux actuels occupants du Bureau 15, Thomas, Cécile, Quentin, Vincent, Hugues, Vladimir. Leur présence fut agréable et précieuse, et je redoute le jour où je serai l'unique occupant d'un bureau.

J'ai particulièrement apprécié les pauses-déjeuner et cafés, à discuter de tout et de rien, mais toujours dans la bonne humeur, en compagnie de Thomas, Anne-Laure, Diogo, Colin, Jérémie, François, Cyril, Marie, Amandine, Sébastien, Gilles, Nicolas… Et que ferait tout ce beau monde sans Laurence, Bénédicte et Zaïna ?

Ma mémoire est capricieuse, et énumérer les personnes envers qui je suis redevable me semble un exercice aussi redoutable que de constituer une bibliographie — avec moins d'outils. Aussi, je tiens à remercier ceux que je n'ai pas manqué d'oublier, pour le flegme avec lequel ils excusent cette distraction.

Mes pensée se tournent aussi vers tous les amis qui, avec constance et détermination, ont su me détourner de mon travail pour les pires et les meilleures des raisons. Je leur dois — et en particulier à Alice, Aurel et Caro, BigB, Fabien, GB, Jean, Jim, Joseph, Laura, les Nours, Pierre, Spil — le courage en période de doute, et le recul nécessaire dans les phases de témérité.

Je voudrais également remercier chaudement l'armée de correcteurs courageux, et en premier lieu Tiny, Alice, Pierre et Jo, qui ont eu l'audace de se plonger dans un brouillon illisible, pour en extraire une version présentable.

Finalement, je remercie de tout cœur les membres de ma famille pour leur présence amicale, leurs encouragements, leur soutien sans faille.

Résumé

Ce mémoire est composé de deux parties indépendantes. Si les deux sujets traités sont de nature différente, l'étude asymptotique du problème est à chaque fois au centre de la réflexion.

Dans la première partie, nous nous intéressons au comportement d'un système composé de deux fluides non miscibles, soumis à la seule force de gravité. Un tel système est utilisé en *océanographie*, afin de modéliser une étendue d'eau de densité variable. On commence par écrire sous forme agréable les équations d'évolution gouvernant le système. Ensuite, on développe une panoplie de modèles asymptotiques, dans les régimes d'*eau peu profonde* – où l'on suppose que la profondeur des couches de fluide est petite devant la longueur d'onde caractéristique à l'interface – et d'*ondes longues* – où l'on ajoute une hypothèse de petitesse des déformations à la surface et à l'interface. Ces modèles sont rigoureusement justifiés, par des résultats de cohérence ou de convergence. Finalement, on s'intéresse particulièrement au phénomène d'*eaux mortes*, qui se manifeste par une forte résistance à l'avancement subie par un corps naviguant dans des eaux stratifiées. Cette résistance est une conséquence de l'énergie dépensée par le corps pour générer une onde à l'interface entre les deux couches de fluide. Là encore, des modèles asymptotiques sont construits, justifiés rigoureusement et simulés numériquement. Une telle étude nous permet de prédire dans quelles situations le phénomène d'eaux mortes se manifeste.

La deuxième partie est dédiée à l'étude de la propagation des ondes dans un milieu non-homogène. La motivation de cette étude se situe dans les *cristaux photoniques*, dont les propriétés de structure se répercutent sur la propagation des rayons lumineux. Plus particulièrement, nous nous intéressons à l'influence de la présence de *défauts* dans le matériau, modélisés par des singularités et/ou des discontinuités dans la microstructure. On montre que ces *interfaces* ont un effet prédominant sur les propriétés asymptotiques du matériau – notamment le coefficient de transmission – lorsque la longueur caractéristique de la microstructure tend vers zéro.

Mots clés : ondes de gravité, ondes internes, modèles asymptotiques, eaux mortes, opérateur de Schrödinger, théorie de diffusion, effets d'interface, homogénéisation.

Abstract

Two distinctive topics are investigated in this dissertation. However, we focus each time on the asymptotic study of the system.

The first part deals with the behavior of a flow, constituted of two immiscible, homogeneous fluids under the only influence of gravity. Such a system is widely used in *oceanography*, as a model for density-stratified fluids. First, we introduce the governing equations of our problem. Then, we construct several asymptotic models, in different regimes. The regimes at stake are the *shallow water* regime (where the depth of the fluids is assumed to be small when compared with the internal wavelength), and the *long wave* regime (with the additional smallness assumptions of small deformations at the surface and at the interface). Each of the models is rigorously justified, thanks to a consistency, or a convergence result. Finally, we deal with the so-called *dead water* phenomenon, which occurs when a ship sails in a stratified fluid, and experiences an important drag due to waves below the surface. Again, we construct, justify, and produce numerical simulations of asymptotic models for this problem. The provided analysis allows to predict the behavior of the flow, and in which situations the dead-water effect occurs.

The second part is dedicated to the wave propagation in inhomogeneous media, having in mind the applications of *photonic crystals*. The structure of these materials allow them to shape the flow of light. In particular, we study the effects of the presence of *defects* in the structure, modeled by singularities, or discontinuities in the microstructure. The effect of such *interfaces* is predominant in the asymptotic behavior of scattering quantities, such as the transmission coefficient, when the size of the microstructure vanishes.

Keywords : gravity waves, internal waves, asymptotic models, dead water, Schrödinger operator, scattering theory, interface effects, homogenization.

Table des matières

CHAPITRE 1

Introduction et vue d'ensemble

Tous les secrets de la nature gisent à découvert et frappent nos regards chaque jour sans que nous y fassions attention.

– André Gide, *Les nouvelles nourritures*

Ce mémoire est constitué de deux parties distinctes, pourtant assujetties à un objectif commun : celui de comprendre et d'analyser des problèmes issus de la Physique, *via* des outils Mathématiques, et en particulier des méthodes liées à l'étude des équations aux dérivées partielles. La première partie explore le comportement des ondes internes à l'interface entre deux fluides, dans un cadre océanographique ; la seconde partie traite de la propagation des ondes à travers des matériaux à structure non homogène, *i.e.* possédant des singularités, ou des microstructures.

Dans chacun de ces deux cas, les équations régissant le système sont complexes, et l'étude de leurs solutions difficile. Dès lors, notre approche consiste à considérer des solutions approchées du problème. À chaque fois, le cadre physique fournit un jeu de paramètres petits du système. En utilisant ces données, nous serons à même de construire des modèles et des solutions asymptotiques, qui seront bien plus agréables à analyser. Bien sûr, la difficulté consistera en la justification précise et rigoureuse de ces solutions, comme approximations du problème.

Au total, cinq articles (publiés ou soumis) composent le corps de ce mémoire, et figurent en des chapitres distincts. Nous avons choisi de les insérer dans leur forme originale, hormis quelques modifications mineures. Ce chapitre précis a pour but de motiver et lier entre eux ces travaux. Pour chacune des deux parties, nous présentons les problèmes qui y sont posés, l'état de l'art en la matière, les résultats obtenus ainsi que les méthodes utilisées.

Sommaire

Partie A. Ondes internes de gravité en océanographie

A.1. Modélisation des ondes internes. Dans cette partie, nous nous intéressons au problème de l'évolution de la surface et de l'interface entre deux couches de fluides supposés non miscibles et de densités différentes, sous l'influence de la gravité seule. De telles situations apparaissent notamment en météorologie (à l'interface océan-atmosphère), mais c'est dans le cadre de l'océanographie que nous présenterons le problème. En effet, la densité des eaux d'un lac, d'une mer ou d'un océan n'est en général pas constante. Dans bien des cas, on observe une séparation très marquée entre les eaux profondes, plus fraîches et plus salées, donc plus denses que les eaux de surface. Il est alors raisonnable de concevoir l'étendue d'eau comme composée de deux couches de fluides de densités différentes.

Si l'étude de l'évolution de la surface dans le cas à un fluide est ancienne (on peut remonter au XIX^{e} siècle avec Russel [**101**]), l'étude des ondes internes a réellement débuté à la fin des années 1960, après une conjonction intéressante d'avancées technologiques (amélioration des chaînes de thermistance, puis imagerie satellite avec le lancement de SEASAT) et de découvertes mathématiques (équations intégrables par la théorie du *scattering* inverse). Les observations et mesures font apparaître la présence de fortes déformations à l'interface, notamment sous la forme de paquets d'ondes internes, pouvant se propager sur des centaines de kilomètres et atteindre des amplitudes considérables.

On dissocie trois phases dans la « vie » de ces ondes, qui procèdent de mécanismes particuliers et demandent un traitement différent. D'abord la génération : comment se forment-elles et sous quelles conditions ? Parmi les causes, on peut distinguer : les dérèglements exceptionnels, qui sont par exemple à l'origine des tsunamis ; les perturbations plus régulières, comme les marées à l'origine des mascarets. Ensuite vient la propagation, où l'on se pose notamment la question de la stabilité de ces ondes et leur comportement face à des perturbations (courants, fonds marins, collisions, *etc.*). Finalement, ces ondes se dissipent, notamment par déferlement, dont on sait que le mélange induit est capital à l'équilibre sous-marin et au transport des sédiments. C'est la phase de propagation qui nous intéressera particulièrement (bien que le problème de la génération soit étudié au Chapitre 4).

Dans la section suivante, nous présentons les équations qui régissent le système. Cette modélisation elle-même suppose quelques hypothèses sur la nature du fluide considéré, que nous détaillons et justifions. À partir du système d'équations d'évolution obtenu, on construit à la section A.2 différents modèles asymptotiques, correspondant à différents régimes considérés (c'est-à-dire à différents ordres de grandeur des variables de notre problème). On peut distinguer deux grandes familles de modèles obtenus : les modèles fortement non-linéaires, s'appuyant uniquement sur une hypothèse d'eau peu profonde[1] et traités en particulier au Chapitre 2 ; les modèles faiblement non-linéaires, qui s'appuient sur une hypothèse supplémentaire de petitesse des amplitudes des ondes[2], étudiés en détail au Chapitre 3. On s'intéresse tout particulièrement dans la section A.3 aux ondes solitaires, comme solutions particulières d'un des modèles obtenus. À cette occasion, on présente de manière plus détaillée la théorie brièvement évoquée ci-dessus, ainsi que ses conséquences dans le cadre des ondes internes. Finalement, la section A.4 apporte un exemple d'application pratique de notre méthode et de nos résultats, avec un traitement nouveau du phénomène atypique d'*eaux mortes* (l'étude complète figure au Chapitre 4).

A.1.1. *Le système d'Euler complet.* Précisons le cadre physique dans lequel nous nous plaçons. Nous faisons effectivement plusieurs hypothèses simplificatrices mais raisonnables

[1]La faible profondeur est à comprendre par rapport à une longueur verticale caractéristique. Le régime peut donc être valide pour des profondeurs atteignant plusieurs kilomètres, comme on le verra par la suite.
[2]Encore une fois, la « petitesse » est à comprendre par comparaison, cette fois avec la profondeur des couches de fluide. Ainsi, elles admettent en fait des amplitudes relativement larges (plusieurs centaines de mètres).

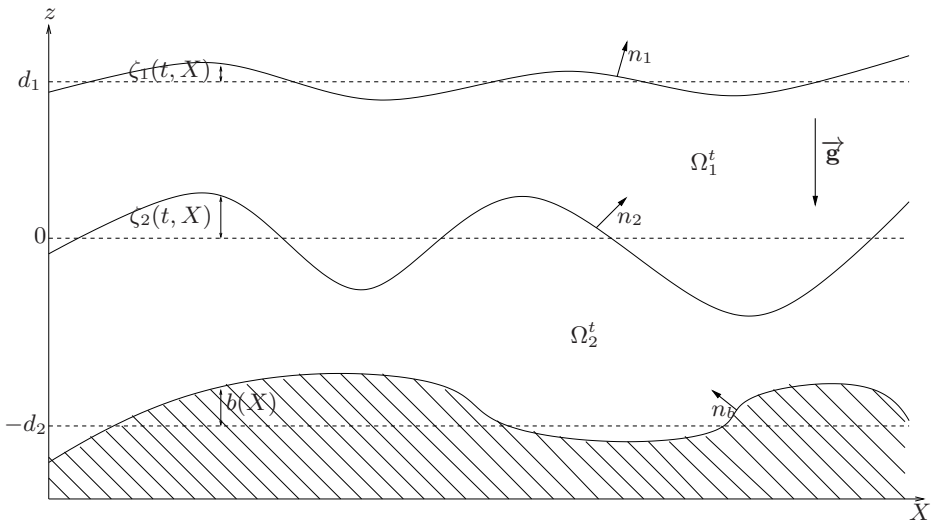


FIGURE 1. Domaine d'étude

dans le cadre océanographique (voir par exemple [**38**]) sur la nature du fluide, qui nous permettent d'écrire les équations de notre problème.

POSTULAT 1. *Le système est composé de deux couches de fluides non miscibles. Le domaine de chaque fluide est déterminé par le graphe de fonctions.*

On note Ω_i^t $(i = 1, 2)$ le domaine du fluide i au temps t (de manière générale, l'indice 1 se réfère au fluide du haut et l'indice 2 au fluide du bas ; voir Figure 1). Il existe alors $\zeta_i(t, X)$ $(i = 1, 2)$ et $b(X)$, tels que pour tous temps $t \geq 0$:

$$\begin{aligned}\Omega_1^t &= \{\ (X,z) \in \mathbb{R}^d \times \mathbb{R}, \quad \zeta_2(t,X) \ \leq\ z \ \leq\ d_1 + \zeta_1(t,X)\ \},\\ \Omega_2^t &= \{\ (X,z) \in \mathbb{R}^d \times \mathbb{R}, \quad -d_2 \ + b(X) \ \leq\ z \ \leq\ \zeta_2(t,X)\ \}.\end{aligned}$$

On note donc $z \in \mathbb{R}$ la variable verticale et $X \in \mathbb{R}^d$, où $d = 1$ ou 2 est la dimension de la variable horizontale.

POSTULAT 2. *Le domaine de chaque fluide reste strictement connexe.*

On se place ainsi loin du rivage, dans la zone de haut-fond. On supposera donc qu'il existe $h_{\min} > 0$, tel que pour tout $X \in \mathbb{R}^d$ et pour tout $t > 0$,

$$d_1 + \zeta_1(t,X) - \zeta_2(t,X) \geq h_{\min} > 0 \quad \text{et} \quad d_2 + \zeta_2(t,X) - b(X) \geq h_{\min} > 0. \tag{A.1}$$

POSTULAT 3. *Les fluides sont irrotationnels et incompressibles.*

L'irrotationalité se justifie dans la mesure où nous nous sommes éloignés des zones de *surf* (« déferlement ») et de *swash* (« jet de rive »), où les effets rotationnels ne sont plus négligeables. On peut alors introduire les potentiels de vitesse ϕ_i tels que

$$\mathbf{v}_i \ \equiv\ \nabla\phi_i \qquad \text{dans } \Omega_i^t.$$

L'incompressibilité de chaque fluide se traduit par une masse volumique ρ_i $(i = 1, 2)$ constante. L'équation de conservation de la masse devient donc

$$\operatorname{div}(\rho_i \mathbf{v}_i) \ =\ -\partial_t \rho_i \ =\ 0 \qquad \Longrightarrow \qquad \Delta\phi_i \ =\ 0 \quad \text{dans } \Omega_i^t. \tag{A.2}$$

POSTULAT 4. *Les fluides sont soumis uniquement à la force de gravité et sont idéaux : les effets de viscosité sont négligeables.*

Les fluides vérifient l'équation d'Euler, qui s'écrit en terme de potentiels de vitesse comme

$$\partial_t \phi_i + \frac{1}{2}|\nabla_{X,z}\phi_i|^2 = -\frac{P}{\rho_i} - gz \quad \text{dans } \Omega_i^t, \tag{A.3}$$

où g est l'accélération gravitationnelle et $P(X,z)$ la pression à l'intérieur du fluide.

POSTULAT 5. *Les fluides sont non miscibles ; les particules de chaque fluide ne traversent pas la surface, l'interface ni le fond.*

Par un simple argument géométrique, on peut se convaincre que cela se traduit par les équations cinématiques suivantes :

$$\partial_t \zeta_1 = \sqrt{1+|\nabla\zeta_1|^2}\partial_n\phi_1 \qquad \text{sur } \Gamma_1 \equiv \{z = d_1 + \zeta_1(t,X)\}, \tag{A.4}$$

$$\partial_t \zeta_2 = \sqrt{1+|\nabla\zeta_2|^2}\partial_n\phi_1 = \sqrt{1+|\nabla\zeta_2|^2}\partial_n\phi_2 \qquad \text{sur } \Gamma_2 \equiv \{z = \zeta_2(t,X)\}, \tag{A.5}$$

$$\partial_n \phi_2 = 0 \qquad \text{sur } \Gamma_b \equiv \{z = -d_2 + b(X)\}, \tag{A.6}$$

où ∂_n représente la dérivation par rapport au vecteur normal ascendant à la surface concernée :

$$\partial_n \equiv n \cdot \nabla_{X,z}, \quad \text{avec} \quad n \equiv \frac{1}{\sqrt{1+|\nabla\zeta|^2}}(-\nabla\zeta, 1)^T.$$

Finalement, notre système est fermé grâce à la dernière hypothèse :

POSTULAT 6. *La pression P est constante à la surface et continue à l'interface*[3].

Les équations (A.2)–(A.6) forment le système que nous allons étudier. En récapitulant, le système obtenu est le suivant :

DEFINITION A.1 (Système d'Euler complet). *On appelle système d'Euler complet l'ensemble des équations d'évolution suivantes :*

$$(\Sigma) \qquad \begin{cases} \Delta\phi_i = 0 & \text{dans } \Omega_i^t,\ i=1,2, \\ \partial_t\phi_i + \frac{1}{2}|\nabla_{X,z}\phi_i|^2 = -\frac{P}{\rho_i} - gz & \text{dans } \Omega_i^t,\ i=1,2, \\ \partial_t\zeta_1 = \sqrt{1+|\nabla\zeta_1|^2}\partial_n\phi_1 & \text{sur } \Gamma_1, \\ \partial_t\zeta_2 = \sqrt{1+|\nabla\zeta_2|^2}\partial_n\phi_1 = \sqrt{1+|\nabla\zeta_2|^2}\partial_n\phi_2 & \text{sur } \Gamma_2, \\ \partial_n\phi_2 = 0 & \text{sur } \Gamma_b, \\ P \text{ est constante à la surface et continue à l'interface.} \end{cases}$$

Les solutions de ce système seront pour nous des solutions *exactes* du problème. Pour autant, l'étude directe de ce problème est extrêmement compliquée, à la fois théoriquement et numériquement. Dans la section suivante, on commence par simplifier le système, en le transposant en des équations uniquement localisées à la surface et à l'interface. On obtient ainsi un gain dans le nombre d'inconnues du problème, ainsi que dans la dimension de l'espace considéré. Notre stratégie consiste ensuite à chercher des solutions *approchées* de ce système, en négligeant les termes dont l'influence est minime sur l'évolution de la solution. Ceci demande bien sûr des hypothèses de petitesse de certains paramètres du système – typiquement les *rapports* entre différentes longueurs – et requiert donc de travailler sur un système adimensionné. Les changements de variables qui nous permettent d'obtenir ce système sont présentés à la section A.1.3.

[3] Le problème de Cauchy associé au système d'Euler complet (Σ) à l'interface entre deux fluides est mal posé dans les espaces de Sobolev avec cette hypothèse (au moins lorsque $d = 1$), à cause d'instabilités de type Kelvin-Helmholtz. Dans [**70**], Lannes prouve qu'il suffit d'ajouter un terme de tension de surface pour que le problème devienne bien posé, sur un temps qui reste raisonnable même lorsque le coefficient de tension de surface est très petit (ce qui est le cas en général). L'idée derrière ce phénomène est que les instabilités de Kelvin-Helmholtz apparaissent à très haute fréquence, où les effets de la tension de surface seront importants. Au contraire, le comportement général du système que l'on veut approcher est localisé dans les basses fréquences ; il n'est donc que très peu modifié par le terme de tension de surface. Pour simplifier les notations, on décide de ne pas introduire ce terme. Le lecteur intéressé pourra se reporter au Chapitre 4, où le terme de tension de surface est inclus dans le système considéré.

A.1.2. *Réduction des équations et opérateurs de Dirichlet-Neumann.* Nous allons reformuler les équations (Σ), en utilisant une remarque basée sur les travaux de Zakharov **[122]**[4], que l'on peut résumer par :

« Connaître le système à la surface et à l'interface du fluide suffit à connaître le système dans l'ensemble du domaine. »

On va donc pouvoir reformuler le système d'Euler en équations d'évolution *localisées à la surface et à l'interface du fluide*, réduisant la dimension d'espace à d au lieu de $d+1$ et le nombre d'équations du problème à $2d+2$.

En effet, introduisons les traces des potentiels de vitesse à la surface et à l'interface :

$$\psi_1(t,X) \equiv \phi_1(t,X,d_1+\zeta_1(t,X)), \quad \text{et} \quad \psi_2(t,X) \equiv \phi_2(t,X,\zeta_2(t,X)).$$

Alors ϕ_2 est définie de manière unique comme la solution du problème de Laplace (A.2) dans le domaine inférieur, muni des conditions de bord de Neumann (A.6) sur Γ_b et de Dirichlet $\phi_2=\psi_2$ sur Γ_2. On obtient ensuite ϕ_1 comme l'unique solution du problème de Laplace dans le domaine supérieur, muni des conditions de Dirichlet $\phi_1=\psi_1$ sur Γ_1 et de la condition de Neumann (sous-tendue par (A.5)) $\partial_n\phi_2=\partial_n\phi_1$ sur Γ_2.

De manière plus précise, on va pouvoir définir les opérateurs suivants :

DEFINITION A.2 (Opérateurs de Dirichlet-Neumann). *Soient $\zeta_1,\ \zeta_2$ et $b\in W^{2,\infty}(\mathbb{R}^d)$, tels que $\Omega_1,\ \Omega_2$ vérifient* (A.1) *et soient $\nabla\psi_1,\nabla\psi_2\in H^{1/2}(\mathbb{R}^d)$. Alors les ϕ_i sont les uniques solutions dans $H^2(\Omega_i)$ $(i=1,2)$ des problèmes suivants :*

$$\text{(A.7)}\qquad \begin{cases} \Delta_{X,z}\phi_2=0 & \text{dans } \Omega_2,\\ \phi_2=\psi_2 & \text{sur } \Gamma_2,\\ \partial_n\phi_2=0 & \text{sur } \Gamma_b, \end{cases} \quad \text{et} \quad \begin{cases} \Delta_{X,z}\phi_1=0 & \text{dans } \Omega_1,\\ \phi_1=\psi_1 & \text{sur } \Gamma_1,\\ \partial_n\phi_1=\partial_n\phi_2 & \text{sur } \Gamma_2, \end{cases}$$

on définit $G_1[\zeta_1,\zeta_2,b](\psi_1,\psi_2),\ G_2[\zeta_2,b]\psi_2,\ H[\zeta_1,\zeta_2,b](\psi_1,\psi_2)\in H^{1/2}(\mathbb{R}^d)$ par

$$\begin{aligned} G_1[\zeta_1,\zeta_2,b](\psi_1,\psi_2) &\equiv \sqrt{1+|\nabla\zeta_1|^2}\,(\partial_n\phi_1)\Big|_{z=d_1+\zeta_1},\\ G_2[\zeta_2,b]\psi_2 &\equiv \sqrt{1+|\nabla\zeta_2|^2}\,(\partial_n\phi_2)\Big|_{z=\zeta_2},\\ H[\zeta_1,\zeta_2,b](\psi_1,\psi_2) &\equiv \nabla\big(\phi_1\Big|_{z=\zeta_2}\big). \end{aligned}$$

En utilisant les formules de dérivation de composées sur la trace du gradient des équations de Bernoulli (A.3) à la surface et à l'interface, il est aisé de déduire du système d'Euler (Σ) le système équivalent suivant :

$$(\Sigma')\qquad \begin{cases} \partial_t\zeta_1 - G_1[\zeta_1,\zeta_2,b](\psi_1,\psi_2) = 0,\\ \partial_t\zeta_2 - G_2[\zeta_2,b]\psi_2 = 0,\\ \partial_t\nabla\psi_1 + g\nabla\zeta_1 + \frac12\nabla(|\nabla\psi_1|^2) - \nabla\mathcal{N}_1 = 0,\\ \partial_t(\nabla\psi_2-\rho_1/\rho_2 H[\zeta_1,\zeta_2,b](\psi_1,\psi_2)) + g(1-\rho_1/\rho_2)\nabla\zeta_2\\ \qquad + \frac12\nabla(|\nabla\psi_2|^2-\rho_1/\rho_2|H[\zeta_1,\zeta_2,b](\psi_1,\psi_2)|^2) - \nabla\mathcal{N}_2 = 0, \end{cases}$$

où l'on note

$$\begin{aligned} \mathcal{N}_1 &\equiv \frac{\big(G_1[\zeta_1,\zeta_2,b](\psi_1,\psi_2)+\nabla\zeta_1\cdot\nabla\psi_1\big)^2}{2(1+|\nabla\zeta_1|^2)}, \quad \text{et}\\ \mathcal{N}_2 &\equiv \frac{\rho_2\big(G_2[\zeta_2,b]\psi_2+\nabla\zeta_2\cdot\nabla\psi_2\big)^2-\rho_1\big(G_2[\zeta_2,b]\psi_2+\nabla\zeta_2\cdot H[\zeta_1,\zeta_2,b](\psi_1,\psi_2)\big)^2}{2\rho_2(1+|\nabla\zeta_2|^2)}. \end{aligned}$$

[4]La formulation que l'on utilise, employant les opérateurs de Dirichlet-Neumann, est due à Craig et Sulem **[36]**. L'adaptation au problème à deux fluides est traité dans **[35, 40]**.

On a donc bien un système de $2d+2$ équations, selon les variables $(\zeta_1, \zeta_2, \psi_1, \psi_2)$. C'est ce système précis que nous allons étudier et dont nous allons construire des modèles approchés.

A.1.3. *Adimensionnement des équations.* Jusqu'ici, nous avons travaillé avec des grandeurs physiques et donc *dimensionnées*. Dans la suite de ce mémoire, nous attacherons une grande importance à l'ordre de grandeur relative des différents objets manipulés, choisissant de négliger ou au contraire de mettre en valeur tel ou tel terme. Ceci nous amène au questionnement suivant :

- Quels sont les paramètres du système ?
- Comment sont liés entre eux ces paramètres ?
- Quels ordres de grandeur pour ces paramètres indiquent les observations ?

Pour répondre à une partie de ces questions, nous nous appuyons sur l'étude du système linéarisé. En effet, nous sommes capables d'obtenir explicitement les solutions d'un tel système et de déduire ainsi de manière approchée le comportement des solutions du système complet, pour finalement choisir les changements de variables adaptés à notre problème. Afin de ne pas gêner la lecture, nous reportons ces calculs à la section C.2, page 48.

Commençons par définir les longueurs caractéristiques du système. On note a_1 l'amplitude maximale de la déformation de la surface, a_2 celle de l'interface et B celle du fond (bathymétrie). La longueur horizontale caractéristique est la longueur d'onde typique de l'interface, supposée identique selon les deux directions lorsque la dimension $d = 2$; elle est notée λ. La longueur verticale de référence est choisie comme étant la profondeur du fluide du haut d_1 et on rappelle que la profondeur de fluide du bas est notée d_2. L'analyse du système linéarisé nous offre une vitesse caractéristique des ondes c_0, définie par

$$c_0 \equiv \sqrt{g(\rho_2 - \rho_1)\frac{d_1 d_2}{\rho_2 d_1 + \rho_1 d_2}}.$$

On définit alors les variables adimensionnées[5]

$$\tilde{z} \equiv \frac{z}{d_1}, \quad \tilde{X} \equiv \frac{X}{\lambda}, \quad \tilde{t} \equiv \frac{t}{\lambda/c_0},$$

et les inconnues adimensionnées

$$\tilde{b}(\tilde{X}) \equiv \frac{b(X)}{B}, \quad \tilde{\zeta}_i(\tilde{X}) \equiv \frac{\zeta_i(X)}{a_i}, \quad \tilde{\psi}_i(\tilde{X}) \equiv \frac{d_1}{a_2 \lambda c_0}\psi_i(X).$$

Six paramètres indépendants du système apparaissent alors naturellement :

$$\gamma = \frac{\rho_1}{\rho_2}, \quad \delta \equiv \frac{d_1}{d_2}, \quad \mu \equiv \frac{d_1^2}{\lambda^2}, \quad \epsilon_1 \equiv \frac{a_1}{d_1}, \quad \epsilon_2 \equiv \frac{a_2}{d_1}, \quad \beta \equiv \frac{B}{d_1}.$$

On peut ainsi réécrire le système avec des variables, paramètres, inconnues sans dimension. En retirant tous les "*tildes*" par souci de lisibilité, on obtient les domaines de fluides adimensionnés :

$$\Omega_1[\epsilon_1\zeta_1, \epsilon_2\zeta_2] \equiv \{(X, z) \in \mathbb{R}^{d+1}, \epsilon_2\zeta_2(X) < z < 1 + \epsilon_1\zeta_1(X)\},$$

$$\Omega_2[\epsilon_2\zeta_2, \beta b] \equiv \{(X, z) \in \mathbb{R}^{d+1}, -\frac{1}{\delta} + \beta b(X) < z < \epsilon_2\zeta_2(X)\},$$

$$\Gamma_1[\epsilon_1\zeta_1] \equiv \{z = 1 + \epsilon_1\zeta_1\}, \quad \Gamma_2[\epsilon_2\zeta_2] \equiv \{z = \epsilon_2\zeta_2\}, \quad \Gamma_b[\beta b] \equiv \{z = -\frac{1}{\delta} + \beta b\},$$

[5]Le changement de variables présenté ici est légèrement différent de celui utilisé dans les chapitres 2 et 3 (voir la section 1.5 du Chapitre 2, page 60). On modifie les résultats obtenus en conséquence.

les problèmes de Laplace étirés

$$(A.8)\quad \left\{\begin{array}{lll} \left(\mu\Delta_X+\partial_z^2\right)\phi_2 &=& 0 \quad \text{dans } \Omega_2,\\ \phi_2 &=& \psi_2 \quad \text{sur } \Gamma_2,\\ \partial_n\phi_2 &=& 0 \quad \text{sur } \Gamma_b, \end{array}\right. , \qquad \left\{\begin{array}{lll} \left(\mu\Delta_X+\partial_z^2\right)\phi_1 &=& 0 \quad \text{dans } \Omega_1,\\ \phi_1 &=& \psi_1 \quad \text{sur } \Gamma_1,\\ \partial_n\phi_1 &=& \partial_n\phi_2 \quad \text{sur } \Gamma_2. \end{array}\right.$$

et les opérateurs de Dirichlet-Neumann

$$G_1(\psi_1,\psi_2)\equiv G_1^{\mu,\delta}[\epsilon_1\zeta_1,\epsilon_2\zeta_2,\beta b](\psi_1,\psi_2) \;\equiv\; -\mu\epsilon_1\nabla\zeta_1\cdot(\nabla\phi_1)\Big|_{z=1+\epsilon_1\zeta_1}+(\partial_z\phi_1)\Big|_{z=1+\epsilon_1\zeta_1},$$

$$G_2\psi_2\equiv G_2^{\mu,\delta}[\epsilon_2\zeta_2,\beta b]\psi_2 \;\equiv\; -\mu\epsilon_2\nabla\zeta_2\cdot(\nabla\phi_2)\Big|_{z=\epsilon_2\zeta_2}+(\partial_z\phi_2)\Big|_{z=\epsilon_2\zeta_2},$$

$$H(\psi_1,\psi_2)\equiv H^{\mu,\delta}[\epsilon_1\zeta_1,\epsilon_2\zeta_2,\beta b](\psi_1,\psi_2) \;\equiv\; \nabla\big(\phi_1\Big|_{z=\epsilon_2\zeta_2}\big),$$

et finalement le problème d'Euler adimensionné :

$$(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})\quad \left\{\begin{array}{l} \alpha\partial_t\zeta_1 - \dfrac{1}{\mu}G_1(\psi_1,\psi_2) = 0,\\[2ex] \partial_t\zeta_2 - \dfrac{1}{\mu}G_2\psi_2 = 0,\\[2ex] \partial_t\nabla\psi_1 + \alpha\dfrac{\gamma+\delta}{1-\gamma}\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla(|\nabla\psi_1|^2) = \mu\epsilon_2\nabla\mathcal{N}_1,\\[2ex] \partial_t(\nabla\psi_2-\gamma H(\psi_1,\psi_2)) + (\gamma+\delta)\nabla\zeta_2 + \dfrac{\epsilon_2}{2}\nabla(\delta^2|\nabla\psi_2|^2-\gamma|H(\psi_1,\psi_2)|^2)\\ \qquad\qquad = \mu\epsilon_2\nabla\mathcal{N}_2, \end{array}\right.$$

où l'on note $\alpha \equiv \epsilon_1/\epsilon_2$,

$$\mathcal{N}_1 \equiv \frac{\left(\frac{1}{\mu}G_1(\psi_1,\psi_2)+\epsilon_1\nabla\zeta_1\cdot\nabla\psi_1\right)^2}{2(1+\mu|\epsilon_1\nabla\zeta_1|^2)}, \qquad \text{et}$$

$$\mathcal{N}_2 \equiv \frac{\left(\frac{1}{\mu}G_2\psi_2+\epsilon_2\nabla\zeta_2\cdot\nabla\psi_2\right)^2-\gamma\left(\frac{1}{\mu}G_2\psi_2+\epsilon_2\nabla\zeta_2\cdot H(\psi_1,\psi_2)\right)^2}{2(1+\mu|\epsilon_2\nabla\zeta_2|^2)}.$$

Ce système est particulièrement compliqué à étudier directement ; aussi, très peu de résultats ont été obtenus. Comme nous l'avons dit dans la note 3, page 5, le problème de Cauchy à l'interface est mal posé dans les espaces de Sobolev (au moins pour la dimension $d=1$) **[109, 110, 73]**. Différents auteurs ont ensuite prouvé qu'ajouter un terme de tension de surface permet d'obtenir existence et unicité des solutions **[3, 4, 106, 28]**, mais le temps d'existence obtenu est très faible. Finalement, dans **[70]**, Lannes obtient un critère de stabilité permettant au problème d'être bien posé sur des temps raisonnables. Pour autant, ces études sont limitées au cas d'une interface entre deux fluides et à notre connaissance, aucun résultat n'existe dans le cadre du problème à deux fluides avec interface *et* surface libre. De plus, les problèmes de la description qualitative et quantitative des solutions demeurent largement hors de portée, et leur simulation numérique complexe et coûteuse. C'est ce qui nous conduit à recourir à des modèles, afin d'élaborer des solutions approchées qui soient à la fois proches des solutions exactes du problème et bien plus commodes à simuler et à analyser. Ces constructions s'appuient sur des hypothèses de petitesse sur certains paramètres du système, qui permettent de négliger les effets de certains termes, ayant en pratique très peu d'influence. On parle ainsi de modèles *asymptotiques*, puisqu'ils sont d'autant plus précis que les paramètres en question sont petits.

La section suivante s'attache à la construction et à la justification rigoureuse de ces modèles, pour des régimes particuliers.

A.2. Différents modèles asymptotiques. Dans cette section, nous construisons des modèles asymptotiques du système d'Euler complet (adimensionné) ($\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta}$). On se place donc dans un *régime* particulier, où l'un au moins des paramètres du système est supposé très petit. Présentons donc ici plus en détail les différents paramètres du système, leur signification et leur ordre de grandeur dans le cadre océanographique.

- μ mesure le caractère peu profond (*shallow*) du système. En effet, dans de nombreux cas, la profondeur des deux couches de fluide est petite devant la longueur d'onde caractéristique (voir Table 1 ci-dessous). Ainsi, dans la suite de notre étude, ce paramètre sera petit et constituera le socle de notre développement asymptotique :
$$\mu \ll 1$$
- ϵ_i $(i = 1, 2)$ sont des paramètres de non-linéarité[6]. Il sera fondamental de distinguer les cas où ces paramètres sont de l'ordre de l'unité (induisant des modèles dits *fortement non-linéaires*) et le cas où ces paramètres sont petits $\epsilon_i = \mathcal{O}(\mu)$ (induisant des modèles *faiblement non-linéaires*).
- α est défini par $\alpha \equiv \epsilon_1/\epsilon_2$. Il est donc dépendant des paramètres ci-dessus, mais joue également un rôle propre. On retrouve les modèles à toit solide en forçant $\alpha = 0$ et on supposera $\alpha = \mathcal{O}(\mu)$ dans un cadre particulier au Chapitre 4.
- δ est le ratio des hauteurs. Lors de l'adimensionnement, nous avons implicitement supposé que les deux couches sont de profondeur comparable ; ainsi, le choix de d_1 – et non d_2 – comme longueur verticale caractéristique est inoffensif. On choisira donc $\delta \in [\delta_{\min}, \delta_{\max}]$, fixé.
- γ est le ratio des densités (ou de manière équivalente des masses volumiques). Le fluide inférieur est supposé plus lourd que le fluide supérieur (le cas contraire entraîne des instabilités de Rayleigh-Taylor et le problème est mal posé) ; on fixe donc $\gamma \in (0, 1)$. Le cadre océanographique nous amène à traiter des valeurs proches de l'unité, si bien que l'on supposera parfois $1 - \gamma = \mathcal{O}(\mu)$, ce qui correspond à l'*approximation Boussinesq*.
- β est le paramètre de bathymétrie. L'influence de la topographie n'est pas le sujet majeur de ce mémoire (on pourra consulter **[26, 27, 54]** par exemple, pour son traitement dans le cas à un fluide) et nous nous limitons à $\beta \equiv 0$ aux Chapitres 3 et 4. Par souci de généralité, nous construisons les modèles pour $\beta \in [0, \beta_{\max}]$ dans le Chapitre 2.

Le choix même des changements de variables lors de l'adimensionnement, restreint implicitement le domaine de validité de notre analyse. En particulier, le choix de λ comme longueur caractéristique horizontale sur tout le domaine nous limite au cas de longueurs d'ondes internes et de surface de même ordre de grandeur. Enfin, puisqu'il n'y a pas de direction horizontale préférentielle, nous n'étudions pas les ondes transverses.

Nous présentons dans la Table 1 quelques paramètres caractéristiques des ondes qui ont pu être observées et mesurées (d'après **[97, 52]** et leurs références), justifiant ainsi le choix des régimes considérés.

	profondeurs	amplitudes	longueurs d'onde
bassins peu profonds	4–30m	3–10m	65–250m
lacs et mers de profondeur moyenne	30–500m	5–40m	200–1000m
mers profondes et eaux côtières	>500m	20–100m	200–3000m

TABLE 1. Paramètres typiques des observations.

[6]On retrouve le système linéarisé en posant $\epsilon_1 = \epsilon_2 = \beta = 0$.

Comme annoncé plus haut, on s'intéresse ici principalement à deux régimes particuliers[7], justifiés au vu des observations de la Table 1.

RÉGIME 1 (Eaux peu profondes).

$$\mu \ll 1.$$

Le régime d'eaux peu profondes (ou *shallow water*) a été largement étudié dans le cadre des ondes de gravité. Dans le cas à un fluide avec surface libre, on obtient à l'ordre $\mathcal{O}(\mu)$ les équations de Saint-Venant [**37**], et à l'ordre supérieur $\mathcal{O}(\mu^2)$ le système de Green-Naghdi [**48**], comportant des termes dispersifs. La justification de ces modèles a été achevée en toute généralité (*i.e.* avec bathymétrie et pour $d = 1, 2$) par Alvarez-Samaniego et Lannes [**1**]. En effet, ces auteurs ont obtenu l'existence et l'unicité des solutions de ces modèles, ainsi que leur convergence vers les solutions suffisamment régulières et bornées du système d'Euler complet.

RÉGIME 2 (Ondes longues).

$$\mu \sim \epsilon_2 \sim \epsilon_1 \ll 1.$$

Dans ce régime, on suppose *a priori* que les déformations de la surface et de l'interface resteront petites devant les profondeurs des deux couches de fluides. Ce régime a également été très largement étudié et mène à des modèles dits *faiblement non-linéaires*. Ces modèles sont intéressants dans la mesure où les effets non-linéaires et dispersifs sont de même taille et peuvent ainsi s'équilibrer. Les modèles dans le cas à un fluide sont extrêmement classiques – en particulier le système Boussinesq [**19, 20**] et les équations découplées de Korteweg-de Vries [**62**] – mais leur justification est plus récente (citons [**33, 104, 16, 26**]).

Pour ces deux régimes, nous n'avons volontairement présenté que les modèles correspondant au cas à une couche de fluide. Ces modèles constituent des étalons pour le cas bi-fluide et les systèmes que l'on obtiendra possèderont leur structure (le fait de retrouver les modèles à un fluide en passant formellement à la limite $\delta \to 1$, $\gamma \to 0$, est une bonne méthode pour vérifier la cohérence de nos modèles bi-fluides). Présentons maintenant un panorama non exhaustif des résultats obtenus dans le cas à deux fluides.

La panoplie des modèles existant est extrêmement large, notamment parce que les régimes considérés sont variés (différents ordres de grandeur des paramètres amènent bien sûr à différentes classes de modèles). Si ce constat est valable dans le cas d'un fluide, il l'est encore plus dans le cas de fluides stratifiés, puisque le nombre de paramètres est multiplié. En particulier, le ratio entre les profondeurs des différentes couches n'est pas anodin ; de nombreux modèles supposent que la couche inférieure (ou supérieure dans le cas de l'interface amosphère-océan) est « infinie ». Il serait fastidieux de présenter tous ces modèles[8], et nous nous limitons donc ici aux régimes qui nous intéressent : les eaux peu profondes et les ondes longues.

Les premiers travaux dans ce cadre utilisent l'hypothèse de toit rigide (et plat) : on pourra citer Miyata [**89, 90**], Mal′tseva [**81**], Matsuno [**82**] puis Camassa et Choi [**31**] et finalement [**17**] pour un traitement général et rigoureux. De tels modèles sont revus, simulés et comparés avec des expériences dans [**22**]. Les auteurs mettent l'accent en particulier sur les différences entre les modèles faiblement non-linéaires (*i.e.* dans le régime 2) et les modèles fortement non-linéaires (dans le régime 1). Dans le cadre de la surface libre, des modèles faiblement non-linéaires ont d'abord été introduits par Camassa et Choi [**30**]. Des

[7] Deux autres régimes, légèrement différents, sont présentés dans le cadre spécifique du phénomène d'eaux mortes dans la section A.4 (voir page 21) et au Chapitre 4.

[8] Le lecteur intéressé pourra consulter [**17**] par exemple, pour la construction et la justification de nombreux modèles, pour de nombreux régimes (dans le cas de deux fluides avec un toit rigide).

modèles légèrement différents, toujours de type Boussinesq, sont développés par Nguyen et Dias [**93**] et étudiés numériquement en détail. Des modèles fortement non-linéaires de type Green-Naghdi ont été obtenus par Barros, Gavrilyuk, Teshukov [**8, 7**]. En utilisant une approche différente, basée sur la formulation Hamiltonienne des équations d'Euler, Craig, Guyenne et Kalisch [**35**] retrouvent certaines de ces équations. Ces modèles sont comparés numériquement par Guyenne [**50**].

Malheureusement, ces travaux sont formels et pour la plupart restreints au cas unidimensionnel ($d = 1$). La méthode que l'on utilise a été introduite dans [**16, 14, 15**] et adaptée au cas de deux fluides (avec un toit solide) par Bona, Lannes et Saut [**17**]. On étend ainsi cette analyse au cas de deux fluides avec surface libre dans la section A.2.2 – et de manière plus détaillée au Chapitre 2.

Une classe de modèles présente un intérêt particulier, par la simplicité (relative) de leur étude : les modèles *unidirectionnels*. Ces modèles consistent à approcher le flot par une superposition d'ondes indépendantes, chaque composante étant déterminée par une équation scalaire. La plus célèbre de ces équations est bien sûr l'historique équation de Korteweg-de Vries (KdV). C'est l'équation la plus simple alliant un terme de non-linéarité et un terme de dispersion, dont les effets se compensent, permettant ainsi l'existence d'ondes solitaires parfaitement stables (on étudie ce phénomène en détail à la section A.3). Plusieurs modèles dérivent de l'équation KdV. On peut citer en particulier l'équation BBM [**10**], remplaçant le terme dispersif par un terme régularisant, et l'équation de Camassa-Holm [**23**] qui généralise ces deux derniers modèles. Dans le cas d'un fluide bi-couche, le coefficient devant le terme non-linéaire de l'équation KdV peut s'annuler pour un certain jeu de paramètres et divers modèles ont été présentés pour contrevenir à cet effet – notamment en ajoutant un terme de non-linéarité cubique [**39**]. Citons également l'équation de Kadomtsev-Petviashvili [**55**] (KP), généralisant l'équation KdV au cas de la dimension $d = 2$, admettant ainsi comme solution une ligne de solitons. Ces modèles et d'autres sont présentés par exemple dans [**98, 5, 53**].

Beaucoup de ces modèles sont obtenus de manière heuristique et leur justification rigoureuse est plus rare. Ainsi, ces modèles sont légitimés le plus souvent directement en comparaison avec des expériences ou des observations (citons en particulier [**98**] pour un panorama des études existantes). Il existe pourtant des exceptions et comme nous l'avons dit plus haut, l'approximation KdV dans le cas à un fluide est désormais parfaitement motivée (il en est de même pour l'équation KP, depuis les travaux de Lannes et Saut [**71**]). L'idée est de montrer que le flot se décompose en plusieurs ondes *indépendantes* (deux dans le cas d'un fluide ou bi-fluide avec toit rigide, quatre dans le cas de deux fluides avec surface libre), chacune d'entre elles vérifiant approximativement une équation unidimensionnelle (KdV dans notre cas). Cette décomposition peut être obtenue dans le cadre du régime 2 à partir du système de type Boussinesq ; c'est l'objet du Chapitre 3 et les modèles obtenus sous ce régime sont présentés section A.2.3.

Présentons désormais les idées générales de la méthode permettant d'obtenir et de justifier les modèles à deux couches avec surface libre.

A.2.1. *Développement des opérateurs.* La stratégie est la suivante. La première étape consiste à obtenir un développement asymptotique des opérateurs G_1, G_2 et H (voir page 8), par rapport au petit paramètre μ. En remplaçant ces opérateurs par les premiers ordres du développement dans le système d'Euler complet ($\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta}$), on obtient les différents modèles que nous présentons dans la suite. Les propriétés de cohérence (au sens présenté ci-dessous) de nos modèles découlent directement du développement asymptotique rigoureux présenté ici. Des propriétés (plus fortes) de convergence suivent par des méthodes d'énergie.

PROPOSITION A.3 (Développement des opérateurs). *Si l'hypothèse* (A.1) *est vérifiée et* $\nabla\psi_i,\ \zeta_i,\ b$ *(*$i=1,2$*) sont bornés dans des espaces de Sobolev suffisamment grands, alors on a les estimations suivantes :*

$$\big|G_2\psi_2 \ + \ \mu\nabla\cdot(h_2\nabla\psi_2) \ - \ \mu^2\nabla\cdot\mathcal{T}[h_2,\beta b]\nabla\psi_2\big|_{H^s} \ = \ \mathcal{O}(\mu^3), \tag{A.9}$$

$$\big|G_1(\psi_1,\psi_2)+\mu(\mathcal{A}_1+\mathcal{A}_2)-\mu^2\Big(\nabla\cdot\mathcal{T}_1+\nabla\cdot\mathcal{T}_2-\frac{1}{2}\nabla\cdot(h_1^2\nabla\mathcal{A}_2) -\nabla\cdot(h_1\epsilon_1\nabla\zeta_1\mathcal{A}_2)\Big)\big|_{H^s} \ = \ \mathcal{O}(\mu^3), \tag{A.10}$$

(A.11)

$$\big|H(\psi_1,\psi_2)-\nabla\psi_1-\mu\nabla\Big(h_1(\mathcal{A}_1+\mathcal{A}_2)-\frac{1}{2}h_1^2\Delta\psi_1-h_1\epsilon_1\nabla\zeta_1\cdot\nabla\psi_1\Big)\big|_{H^s} \ = \ \mathcal{O}(\mu^2),$$

où l'on a défini

$$\mathcal{A}_i \ \equiv \ \nabla\cdot(h_i\nabla\psi_i),\ (i=1,2), \quad \mathcal{T}_1 \ \equiv \ \mathcal{T}[h_1,\epsilon_2\zeta_2]\nabla\psi_1, \quad \mathcal{T}_2 \ \equiv \ \mathcal{T}[h_2,\beta b]\nabla\psi_2$$

avec $h_1\equiv 1+\epsilon_1\zeta_1-\epsilon_2\zeta_2$ *la profondeur de la couche supérieure,* $h_2\equiv\frac{1}{\delta}-\beta b+\epsilon_2\zeta_2$ *la profondeur de la couche inférieure et l'opérateur suivant :*

$$\mathcal{T}[h,b]V \ \equiv \ -\frac{1}{3}\nabla(h^3\nabla\cdot V)+\frac{1}{2}\big(\nabla(h^2\nabla b\cdot V)-h^2\nabla b\nabla\cdot V\big)+h\nabla b\nabla b\cdot V.$$

Idée de la preuve. La preuve de ce résultat suit la méthode utilisée dans **[17]** et amorcée dans **[16, 14, 15]**. L'idée est d'introduire en premier lieu un changement de variables, afin de transformer les problèmes de Laplace (A.8) en des problèmes sur des domaines plats. Le Laplacien est alors transformé en un opérateur variable (dépendant des variables ζ_1,ζ_2), mais on peut montrer que l'opérateur obtenu n'en est pas moins coercif. De plus, l'opérateur possède un développement asymptotique trivial en μ et il est aisé d'obtenir formellement un développement asymptotique des solutions ϕ_1 et ϕ_2 de (A.8), en résolvant à chaque ordre. On peut alors contrôler les normes de Sobolev de la différence entre la solution exacte et l'approximation à l'ordre voulu, en utilisant l'équation elliptique sur le domaine plat vérifiée par cette différence et la coercivité de l'opérateur. Ces estimations sont alors reportées aux opérateurs G_1, G_2, H et la proposition suit. On trouvera les détails de la preuve au Chapitre 2, section A (page 80). □

Comme nous l'avons dit, on justifie nos modèles en regard du système d'Euler complet $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ (voir page 8) par des résultats de cohérence ou de convergence, au sens suivant.

DEFINITION A.4 (Cohérence). *Le système d'Euler complet* $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ *est cohérent avec un système* (S) *de* $2d+2$ *équations, si toute solution régulière et bornée de* $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ *telle que* (A.1) *soit vérifiée, satisfait* (S) *à un petit reste près, appelé précision du modèle asymptotique. À travers ce mémoire, la précision sera entendue au sens des normes* $L^\infty H^s$*, i.e. le reste est bornée en norme* H^s*, uniformément par rapport au temps* t*.*

DEFINITION A.5 (Convergence). *Le système d'Euler complet* $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ *et un système de* $2d+2$ *équations bien posé* (S) *sont* convergents *à l'ordre* $\mathcal{O}(\varepsilon)$ *si toute solution* V *du système d'Euler complet, suffisamment régulière et bornée en norme* H^s*, est approchée par la solution* $\tilde{V}$ *du système* (S) *avec donnée initiale identique, avec l'estimation suivante :*

$$\big|V-\tilde{V}\big|_{L^\infty([0,T];H^s)} \ \leq \ C_0\ \varepsilon,$$

où C_0 *est indépendant de* ε*.*

La deuxième définition est évidemment bien plus forte que la première. Quand il est possible d'obtenir des résultats de convergence, la cohérence avec le système est en fait toujours une étape intermédiaire de la preuve dudit résultat.

A.2.2. *Les modèles d'eau peu profonde : $\mu \ll 1$.* On présente ici les deux modèles obtenus dans le régime d'eau peu profonde. Au premier ordre, on obtient un modèle équivalent aux équations *shallow water* ou Saint-Venant, dans le cas à un fluide ; le modèle obtenu à l'ordre suivant est de type Green-Naghdi.

Le modèle « shallow water/shallow water ». En utilisant le premier ordre des développements asymptotiques (A.9), (A.10) et (A.11) dans le système $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$, on obtient (en supprimant tous les termes d'ordre $\mathcal{O}(\mu)$) le système suivant :

$$(\mathcal{M}_{SW}) \qquad \begin{cases} \alpha\partial_t\zeta_1 + \nabla\cdot(h_1\nabla\psi_1) + \nabla\cdot(h_2\nabla\psi_2) = 0, \\ \partial_t\zeta_2 + \nabla\cdot(h_2\nabla\psi_2) = 0, \\ \partial_t\nabla\psi_1 + \alpha\dfrac{\gamma+\delta}{1-\gamma}\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_1|^2\right) = 0, \\ \partial_t\nabla\psi_2 + (\gamma+\delta)\nabla\zeta_2 + \alpha\gamma\dfrac{\gamma+\delta}{1-\gamma}\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_2|^2\right) = 0, \end{cases}$$

avec $h_1 = 1 + \epsilon_1\zeta_1 - \epsilon_2\zeta_2$ et $h_2 = \frac{1}{\delta} - \beta b + \epsilon_2\zeta_2$.

Ce système a été présenté dans **[35]** et sous une forme légèrement différente dans **[30]**. L'intérêt de notre stratégie est que le développement précis des opérateurs nous permet de contrôler rigoureusement la validité de l'approximation. En effet, on déduit de manière triviale, à partir de la Proposition A.3, le résultat de cohérence suivant :

Proposition A.6. *Le système d'Euler complet $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ est cohérent avec le modèle shallow-water/shallow-water $(\mathcal{M}_{SW})$, à la précision $\mathcal{O}(\mu)$.*

Le système possède une structure hyperbolique agréable. En utilisant la théorie des systèmes quasi-linéaires, on obtient une condition raisonnable sur la condition initiale, pour laquelle le système est Friedrichs-symétrisable. On peut alors en déduire une inégalité d'énergie, le caractère bien posé du système et finalement le résultat :

Proposition A.7. *Si $\beta \equiv 0$ et en se restreignant aux données initiales vérifiant la condition de symétrisabilité, le système d'Euler complet $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ et le modèle shallow-water/shallow-water $(\mathcal{M}_{SW})$ sont convergents à l'ordre $\mathcal{O}(\mu)$.*

Ces deux résultats sont détaillés au Chapitre 2, respectivement à la section 2.3.1 page 69 et à la section 3 page 74.

Le système de type Green-Naghdi. En utilisant désormais les termes d'ordre le plus élevé dans les développements des opérateurs (A.9), (A.10) et (A.11), on obtient le modèle suivant :

$(\mathcal{M}_{G-N})$

$$\begin{cases} \alpha\partial_t\zeta_1 + \mathcal{A}_1 + \mathcal{A}_2 \;=\; \mu\Big(\nabla\cdot\mathcal{T}_1 + \nabla\cdot\mathcal{T}_2 - \frac{1}{2}\nabla\cdot(h_1^2\nabla\mathcal{A}_2) - \nabla\cdot\big(h_1\epsilon_1\nabla\zeta_1\mathcal{A}_2\big)\Big), \\ \partial_t\zeta_2 + \mathcal{A}_2 \;=\; \mu\nabla\cdot\mathcal{T}_2, \\ \partial_t\nabla\psi_1 + \alpha\dfrac{\gamma+\delta}{1-\gamma}\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_1|^2\right) = \mu\epsilon_2\nabla\mathcal{N}_1, \\ \partial_t\nabla\psi_2 + (\gamma+\delta)\nabla\zeta_2 + \alpha\gamma\dfrac{\gamma+\delta}{1-\gamma}\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_2|^2\right) = \mu\Big(\gamma\partial_t\nabla\mathcal{H} + \gamma\epsilon_2\nabla(\nabla\psi_1\cdot\nabla\mathcal{H}) \\ \qquad\qquad +\epsilon_2\nabla\mathcal{N}_2 + \gamma\epsilon_2\nabla\mathcal{N}_1\Big), \end{cases}$$

où l'on a utilisé les notations suivantes :

$$\begin{aligned}
\mathcal{A}_1 &\equiv \nabla\cdot(h_1\nabla\psi_1), \qquad & \mathcal{A}_2 &\equiv \nabla\cdot(h_2\nabla\psi_2),\\
\mathcal{T}_1 &\equiv \mathcal{T}[h_1,\epsilon_2\zeta_2]\nabla\psi_1, \qquad & \mathcal{T}_2 &\equiv \mathcal{T}[h_2,\beta b]\nabla\psi_2,\\
\mathcal{H} &\equiv h_1(\nabla\cdot(h_1\nabla\psi_1)+\nabla\cdot(h_1\nabla\psi_2)-\tfrac{1}{2}h_1\Delta\psi_1-\epsilon_1\nabla\zeta_1\cdot\nabla\psi_1),\\
\mathcal{N}_1 &\equiv \tfrac{1}{2}(\epsilon_1\nabla\zeta_1\cdot\nabla\psi_1-\nabla\cdot(h_1\nabla\psi_1)-\nabla\cdot(h_2\nabla\psi_2))^2,\\
\mathcal{N}_2 &\equiv \tfrac{1}{2}(\epsilon_2\nabla\zeta_2\cdot\nabla\psi_2-\nabla\cdot(h_2\nabla\psi_2))^2-\tfrac{\gamma}{2}(\epsilon_2\nabla\zeta_2\cdot\nabla\psi_1-\nabla\cdot(h_2\nabla\psi_2))^2.
\end{aligned}$$

Comme précédemment, on obtient immédiatement de la Proposition A.3, le résultat de cohérence suivant (voir Chapitre 2, Section 2.3, page 68) :

PROPOSITION A.8. *Le système d'Euler complet* $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ *est cohérent avec le modèle de type Green-Naghdi* $(\mathcal{M}_{G-N})$*, à la précision* $\mathcal{O}(\mu^2)$*.*

A.2.3. *Le régime d'ondes longues :* $\mu\sim\epsilon_2\sim\epsilon_1\ll 1$. Deux types de modèles sont présentés dans le régime d'ondes longues : les modèles de type Boussinesq et l'approximation KdV découplée.

LES MODÈLES DE TYPE BOUSSINESQ À PARAMÈTRES. En supposant $\epsilon_2\sim\epsilon_1\sim\mu$, on peut déduire du modèle $(\mathcal{M}_{G-N})$ un ensemble de modèles, tous cohérents à l'ordre $\mathcal{O}(\mu^2)$ et dépendants d'un triplet de paramètres (a_1,a_2,b_1).

$$(\mathcal{M}^{a_1,a_2,b_1}_{Bouss})\begin{cases}
\alpha\partial_t\zeta_1+\nabla\cdot(h_1u_1)+\nabla\cdot(h_2u_2)+\mu\big(\beta_1\nabla\cdot\Delta u_1+\alpha_1\nabla\cdot\Delta u_2\big)=0,\\[2ex]
\partial_t\zeta_2+\nabla\cdot(h_2u_2)+\mu\frac{1+3a_2}{3\delta^3}\nabla\cdot\Delta u_2=0,\\[2ex]
(1+\mu b_1\Delta)\partial_t u_1+\mu\frac{a_1}{\delta}\Delta\partial_t u_2+\alpha\frac{\gamma+\delta}{1-\gamma}\nabla\zeta_1+\dfrac{\epsilon_2}{2}\nabla\left(|u_1|^2\right)=0,\\[2ex]
(1+\mu\alpha_2\Delta)\partial_t u_2-\mu\frac{\gamma}{2}\Delta\partial_t u_1+(\gamma+\delta)\nabla\zeta_2+\alpha\gamma\frac{\gamma+\delta}{1-\gamma}\nabla\zeta_1+\dfrac{\epsilon_2}{2}\nabla\left(|u_2|^2\right)=0,
\end{cases}$$

où $\alpha_1\equiv\frac{1+2a_1}{2\delta}+\frac{1+3a_2}{3\delta^3}$, $\alpha_2\equiv\frac{a_2}{\delta^2}-\frac{\gamma}{\delta}$, $\beta_1\equiv\frac{1+3b_1}{3}$ et u_i $(i=1,2)$ est défini par

$$u_1\equiv\nabla\psi_1-\mu b_1\Delta\nabla\psi_1-\mu\frac{a_1}{\delta}\Delta\nabla\psi_2,\qquad u_2\equiv\nabla\psi_2-\mu\frac{a_2}{\delta^2}\Delta\nabla\psi_2.$$

En choisissant $a_1=-\frac{1}{2}$, $a_2=-\frac{1}{3}$ et $b_1=-\frac{1}{3}$, on obtient le modèle présenté par Choi et Camassa dans **[30]** et où u_1 et u_2 représentent la vitesse horizontale du fluide moyennée verticalement sur la couche du fluide supérieur et inférieur, respectivement. Le choix des paramètres nous permet par exemple de s'assurer que le modèle est linéairement bien posé ou d'avoir une relation de dispersion correspondant à celle du système complet à l'ordre maximal (voir Chapitre 2, Section 2.3.2 page 70).

PROPOSITION A.9. *Le système d'Euler complet* $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ *est cohérent avec le modèle de type Boussinesq* $(\mathcal{M}^{a_1,a_2,b_1}_{Bouss})$*, à la précision* $\mathcal{O}(\mu^2)$*.*

LES MODÈLES DE TYPE BOUSSINESQ SYMÉTRIQUES. Pour simplifier les notations, on se place désormais dans le cas $d=1$ et on pose $\epsilon_1=\epsilon_2=\mu\equiv\varepsilon$. Le cas général peut être obtenu de la même manière.

Le système $(\mathcal{M}^{a_1,a_2,b_1}_{Bouss})$ s'écrit alors trivialement sous la forme suivante :

$$\partial_t U+\mathcal{A}_0\partial_x U+\varepsilon\big(\mathcal{A}(U)\partial_x U+\mathcal{A}_1\partial_x^3 U-\mathcal{A}_2\partial_x^2\partial_t U\big)=0, \tag{A.12}$$

avec $U=(\eta_1,\eta_2,v_1,v_2)$, $\mathcal{A}_0,\mathcal{A}_1,\mathcal{A}_2\in\mathcal{M}_4(\mathbb{R})$ et $\mathcal{A}(\cdot)$ une application linéaire de $\mathbb{R}^4$ à valeur dans $\mathcal{M}_4(\mathbb{R})$.

En multipliant par un opérateur bien choisi de la forme $S(U) = S_0 + \varepsilon S_1(U) - \varepsilon S_2 \partial_x^2$ et en supprimant les termes de reste d'ordre $\mathcal{O}(\varepsilon^2)$, on obtient le système parfaitement symétrique

$$(\mathcal{M}^{sym}_{Bouss}) \qquad \Big(S_0 + \varepsilon\big(S_1(U) - S_2\partial_x^2\big)\Big)\partial_t U + \Big(\Sigma_0 + \varepsilon\big(\Sigma_1(U) - \Sigma_2\partial_x^2\big)\Big)\partial_x U = 0,$$

avec les propriétés suivantes :

i. Les matrices S_0, Σ_0, S_2, $\Sigma_2 \in \mathcal{M}_4(\mathbb{R})$ sont symétriques.
ii. $S_1(\cdot)$ et $\Sigma_1(\cdot)$ sont des applications linéaires de $\mathbb{R}^4$ à valeur dans $\mathcal{M}_4(\mathbb{R})$ et pour tout $U \in \mathbb{R}^4$, $S_1(U)$ et $\Sigma_1(U)$ sont symétriques.
iii. S_0 et S_2 sont définies positives.

Par construction, le système obtenu est encore cohérent à la précision $\mathcal{O}(\varepsilon^2)$. De plus, les propriétés du système nous permettent d'utiliser des arguments énergétiques afin d'obtenir le caractère bien posé du système $(\mathcal{M}^{sym}_{Bouss})$ sur des temps de l'ordre $\mathcal{O}(1/\varepsilon)$, ainsi que le résultat suivant :

Proposition A.10. *Pour ε suffisamment petit, le système d'Euler complet $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ et le modèle Boussinesq symétrique $(\mathcal{M}^{sym}_{Bouss})$ sont convergents à l'ordre $\mathcal{O}(\varepsilon)$ sur des temps de l'ordre $\mathcal{O}(1/\varepsilon)$.*

Idée de la preuve. La preuve de ce résultats s'appuie sur les mêmes considérations énergétiques que dans le cas de la Proposition A.7, bien que des termes dispersifs doivent être gérés. L'énergie du système est désormais

$$E_s(U) \;\equiv\; 1/2(S_0\Lambda^s U, \Lambda^s U) + \varepsilon_2/2(S_1(U)\Lambda^s U, \Lambda^s U) + \varepsilon/2(S_2\Lambda^s\partial_x U, \Lambda^s\partial_x U).$$

Grâce aux propriétés de S_0 et S_2, on prouve que cette énergie est uniformément équivalente à la norme $\big|\cdot\big|_{H^{s+1}_\varepsilon}$, au sens où il existe $C_0 > 0$ tel que

$$\frac{1}{C_0}\left(\big|U\big|^2_{H^s} + \varepsilon\big|U\big|^2_{H^{s+1}}\right) \;\le\; E_s(U) \;\le\; C_0\left(\big|U\big|^2_{H^s} + \varepsilon\big|U\big|^2_{H^{s+1}}\right).$$

C'est cette quantité que l'on va contrôler et qui nous offrira le caractère bien posé, puis le résultat de convergence. On trouvera les détails de la preuve au Chapitre 3, page 93. □

L'approximation Korteweg-de Vries. L'approximation KdV consiste à montrer que dans la limite du régime d'ondes longues, la solution du système complet aura tendance à se décomposer en plusieurs fonctions *indépendantes* (quatre dans le cadre de deux fluides avec une surface libre, deux dans le cadre d'un fluide ou du toit solide), chacune étant solution d'une équation KdV. Si l'équation KdV a été constamment utilisée dans la littérature comme modèle pour la propagation des ondes de gravité, la justification rigoureuse de la décomposition est récente, même dans le cas à un fluide.

C'est à partir du système symétrique de type Boussinesq que l'on obtient cette décomposition. L'idée est de reprendre la méthode BKW (de Brillouin, Kramers, Wentzel) et l'adapter à notre problème. Ainsi, on suppose que la solution U de $(\mathcal{M}^{sym}_{Bouss})$ peut s'écrire sous la forme d'un développement asymptotique

$$U(t,x) \;\equiv\; U_0(\varepsilon t, t, x) \;+\; \varepsilon U_1(\varepsilon t, t, x) \;+\; \cdots$$

En plongeant cet *Ansatz* dans le système Boussinesq et en résolvant à chaque ordre, on obtient les équations régissant $U_0, U_1, \cdots$ En particulier, les propriétés de S_0 et Σ_0 permettent d'obtenir une base de $\mathbb{R}^4$ $(\mathbf{e}_1, \ldots, \mathbf{e}_4)$, qui diagonalise à la fois S_0 et Σ_0. On s'aperçoit alors que U_0 se décompose dans cette base en *quatre fonctions scalaires indépendantes, chacune solution d'une équation de Korteweg-de Vries.* Plus précisément, on obtient la décomposition suivante :

$$U_{\mathrm{KdV}}(t,x) \;\equiv\; U_0(\varepsilon t, t, x) \;\equiv\; \sum_{i=1}^{4} u_i(t,x)\ \mathbf{e}_i,$$

où u_i est la solution de l'équation

$$\partial_t u_i \ + \ c_i \partial_x u_i \ + \ \varepsilon \lambda_i u_i \partial_x u_i \ + \ \varepsilon \mu_i \partial_x^3 u_i \ = \ 0, \tag{A.13}$$

avec $u_i\,|_{t=0} = \mathbf{e}_i \cdot S_0 U\,|_{t=0}$ et c_i, λ_i, μ_i donnés page 97, remarque 3.5.

On obtient alors la proposition suivante :

PROPOSITION A.11. *Pour ε suffisamment petit, le système d'Euler complet $(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ et l'approximation KdV sont convergents à l'ordre $\mathcal{O}(\sqrt{\varepsilon})$ sur des temps de l'ordre $\mathcal{O}(1/\varepsilon)$.*

*De plus, si la donnée initiale est suffisamment localisée (*i.e. *décroissante à l'infini), alors l'approximation est d'ordre $\mathcal{O}(\varepsilon)$.*

L'énoncé exact de ce théorème se trouve au Chapitre 3, Section 3, page 95. Sa preuve est présentée à la Section 3.2 et fonctionne de la manière suivante. Tout d'abord, il est clair qu'il suffit d'obtenir les résultats de convergence entre l'approximation KdV et le modèle Boussinesq symétrique : ceux-ci seront propagés au système d'Euler complet grâce à la proposition A.10. On procède ensuite en deux temps. Il s'agit premièrement d'obtenir la convergence entre U la solution de $(\mathcal{M}^{sym}_{Bouss})$ et $U_0(\varepsilon t, t, x) \ + \ \varepsilon U_1(\varepsilon t, t, x)$ obtenue par le développement BKW. On utilise pour cela les méthodes d'énergie classiques, déjà mises en œuvre pour la Proposition A.10. Il s'agit ensuite de majorer convenablement U_1. Notamment, la fonction U_1 possède des termes d'interactions entre les différentes composantes u_i, qui sont totalement absentes dans U_0. C'est alors que l'on voit l'importance de la localisation de la donnée initiale. Les composantes u_i interagissent peu entre elles si elles sont localisées, car elles se déplacent à des vitesses différentes (tous les c_i sont différents) ; elles seront donc localisées après un temps fini en des lieux différents de l'espace. Cette différence dans le taux de convergence est bien réelle, comme on le voit à la figure 7, page 114.

A.3. L'approximation Korteweg-de Vries.

A.3.1. *L'équation KdV et les solitons.* La théorie des solitons trouve son origine en 1834 lorsque John Scott Russell, alors qu'il surveille le passage d'un bateau le long d'un canal, observe la propagation d'une onde de surface, qu'il rapporte en ces termes :

> I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped – not so the mass of water in the channel which it had put in motion ; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour, preserving its original figure some thirty feet long and a foot to a foot and a half in height. Its height gradually diminished, and after a chase of one or two miles I lost it in the windings of the channel. Such, in the month of August 1834, was my first chance interview with that singular and beautiful phenomenon which I have called the Wave of Translation.

Russel construit alors un bassin afin d'étudier le comportement de ces ondes particulières, que nous appellerons pour simplifier *solitons*. De ses expériences, il déduit plusieurs observations [**101**] :

1. Les solitons sont stables, au sens où ils se propagent sur de très longues distances, sans que leur forme ni leur amplitude ne soit affectée.

2. La vitesse des solitons est fonction (croissante) de leur amplitude.

3. La collision entre deux solitons ne vérifie qu'en apparence le principe de superposition : il se produit un déphasage.

On peut déduire plusieurs propriétés mathématiques de ces observations. D'abord, les points *2.* et *3.* témoignent de la non-linéarité du phénomène. La non-linéarité seule ne permet pas d'expliquer le point *1.* Ainsi que l'illustre l'équation de Burgers

$$\partial_t u \ + \ u \partial_x u \ = \ 0,$$

la présence d'un terme non-linéaire tend à contracter le profil des ondes et à favoriser la formation de chocs. La stabilité du soliton indique donc la présence d'une autre action, qui vient contrebalancer les effets de concentration de la non-linéarité : la dispersion.

La théorie mathématique de ces ondes commence en 1871 avec Boussinesq **[19, 20]**, puis Rayleigh **[100]** (1876) et Korteweg et de Vries **[62]** (1895). Ces derniers proposent une équation simple modélisant la propagation des solitons. Après un changement de variables bien choisi, cette équation s'écrit :

$$\partial_t u \ + \ u\partial_x u \ + \ \partial_x^3 u \ = \ 0. \tag{A.14}$$

Cette équation est la plus simple à équilibrer un terme non-linéaire, ainsi qu'un terme dispersif[9]. C'est cette propriété qui leur permet d'admettre une famille d'ondes solitaires, définies par

$$u(t,x) \ \equiv \ 3M \ \operatorname{sech}^2\left(\frac{M}{4}(x - x_0 - Mt)\right), \tag{A.15}$$

où sech est la sécante hyperbolique, et M et x_0 sont choisis arbitrairement. On retrouve donc les points *1.* et *2.* des observations de Russell. La troisième observation, concernant la collision entre deux solitons (et implicitement l'existence simultanée de plusieurs solitons), ne sera expliquée mathématiquement qu'à la fin des années 1960, quand plusieurs travaux relancent l'intérêt porté à cette équation.

En 1965, Zabusky et Kruskal **[121]** observent numériquement que des ondes solitaires de la forme (A.15) sont générées à partir de données initiales assez générales et demeurent intactes après collision. Des travaux théoriques **[46, 72, 64]** ainsi que des expériences **[87]** viennent confirmer ce résultat. On découvre ainsi que l'équation KdV est *intégrable*, c'est-à-dire qu'on peut se ramener par une transformation non-linéaire explicite (le *scattering inverse*) à une hiérarchie d'équations différentielles très simples. Les propriétés qui découlent de cette découverte sont notamment :

- Il existe une infinité de quantités indépendantes conservées par le système.
- Toute solution dont la donnée initiale est suffisamment régulière et localisée se décompose en un train d'ondes solitaires de la forme (A.15), accompagné d'une onde dispersée (dont l'amplitude maximale tend vers 0 en temps long, mais qui peut emporter une partie de l'énergie du système). Les solitons sont donc des attracteurs du flot Hamiltonien associé à (A.14).
- La collision entre deux solitons est parfaitement élastique (c'est-à-dire sans perte d'énergie) et formes, vitesses et amplitudes sont conservées. Le principe de superposition n'est pourtant pas respecté, puisqu'il se produit un décalage de phase.

Ces propriétés sont exceptionnelles. On voit qu'un soliton est une structure extrêmement stable qui émerge de manière naturelle et sera donc omniprésente dans tous les systèmes vérifiant une équation de la forme (A.14). Or, cette équation apparaît dans des domaines de la physique très variés, tels l'optique non-linéaire, la physique des plasmas, l'astrophysique et naturellement dans le cadre historique de la mécanique des fluides et des ondes de gravité.

Comme nous l'avons dit plus haut, la justification rigoureuse de (A.14) comme modèle pour le problème des ondes de surface est relativement récente **[33, 57, 103, 104, 16]** et apporte la touche finale à l'analyse, expliquant pourquoi les ondes solitaires sont si souvent observées.

A.3.2. *Le cas des ondes internes.* L'approximation KdV dans le cas des ondes internes a été obtenue formellement très tôt. Citons Keulegan **[60]** (1953) dans le cas d'un toit rigide et d'un ratio de densité proche de l'unité, Long **[77]** (1956) sans cette dernière hypothèse et Peters et Stoker dans le cas avec surface libre **[99]** (1960). Ces modèles ont attiré une

[9]Remarquons que cette équation est unidimensionnelle. Nous nous plaçons ici et pour le reste de la section dans le cadre d'une dimension spatiale $d = 1$.

attention croissante, depuis la conjonction des découvertes mathématiques décrites plus haut avec le développement de nouvelles techniques de mesures. D'abord, l'amélioration des chaînes de thermistances ont permis l'observation d'ondes solitaires internes dans les mers bordières et les eaux côtières des océans. L'inconvénient majeur de cette technique est qu'elle ne permet des mesures qu'à des localisations sporadiques et avec un très petit espacement des points de mesure. Il est donc impossible de « suivre » une onde solitaire. En 1978, le lancement du satellite SEASAT, utilisant un radar à ouverture de synthèse (SAR), a permis d'obtenir des photographies plus larges et plus générales de ces ondes. Il apparaît de manière très nette[10] des paquets d'ondes internes, qui sont générées de manière régulière au rythme des marées. Ces images sont convaincantes, mais ne permettent aucune mesure précise, si bien que l'existence ou non d'ondes internes solitaires de type soliton demeure un problème complexe (voir par exemple **[97, 52]** et les références incluses).

Dans ce cadre, la justification rigoureuse de l'approximation KdV (présentée en Proposition A.11) est une étape importante dans la modélisation du problème. Notamment, la décomposition *simultanée* des solutions du problème d'Euler permet de comparer directement les différents modèles (par exemple les modèles de type Boussinesq prenant en compte les interactions entre les composantes, ou des modèles fortement linéaires). En particulier, dans la section 3.4 du Chapitre 3 (page 104), on étudie la validité de l'hypothèse de toit rigide, c'est à dire pour quelles valeurs des coefficients γ et δ les approximations KdV pour les configurations avec surface libre ou toit rigide prédisent des comportements semblables. Présentons brièvement cette analyse.

On rappelle que la décomposition de l'approximation KdV, dans le cadre bi-fluide avec une surface libre, fait apparaître quatre composantes ($U_0 \equiv \sum_{i=1}^{4} u_i \mathbf{e}_i$). Quand on projette la décomposition sur les variables scalaires de déformation à la surface et à l'interface, on obtient :

- À l'interface : $\zeta_2 \equiv \sum_{(j,k)=(\pm,\pm)} \eta_{j,k}$, où $\eta_{\pm,\pm}$ vérifie l'équation

$$\partial_t \eta_{\pm,k} \pm c_k \partial_x \eta_{\pm,k} + \varepsilon \lambda_k^i \eta_{\pm,k} \partial_x \eta_{\pm,k} \pm \varepsilon \mu_k \partial_x^3 \eta_{\pm,k} = 0.$$

- À la surface : $\zeta_1 \equiv \sum_{(j,k)=(\pm,\pm)} \zeta_{j,k}$, où $\zeta_{\pm,\pm}$ vérifie l'équation

$$\partial_t \zeta_{\pm,k} \pm c_\pm \partial_x \zeta_{\pm,k} + \varepsilon \lambda_k^s \zeta_{\pm,k} \partial_x \zeta_{\pm,k} \pm \varepsilon \mu_k \partial_x^3 \zeta_{\pm,k} = 0.$$

Les coefficients sont les suivants :

$$c_\pm^2 = \frac{1+\delta \pm \sqrt{(1-\delta)^2+4\gamma\delta}}{2\delta}, \qquad \mu_\pm = \frac{c_\pm}{6} \frac{(1+\frac{3\gamma}{\delta}+\frac{1}{\delta^2})(c_\pm^2 - \frac{1-\gamma}{\delta+1}) - \frac{1}{\delta}c_\pm^2}{c_\pm^2 - 2\frac{1-\gamma}{\delta+1}},$$

$$\lambda_\pm^i = \frac{3c_\pm}{2} \frac{(2-\delta)c_\pm^2 + \delta - \frac{1}{\delta} - (1-\gamma)}{|(c_+^2 - c_-^2)(1-c_\pm^2)|}, \qquad \lambda_\pm^s = \frac{3c_\pm}{2} \frac{(2-\delta)c_\pm^2 + \delta - \frac{1}{\delta} - (1-\gamma)}{(c_+^2 - c_-^2)c_\pm^2}.$$

La décomposition si l'on suppose un toit plat et rigide ne fait apparaître que deux composantes. La variation à l'interface est ainsi obtenue par $\eta = \eta_+ + \eta_-$, où $\eta_\pm$ est la solution de l'équation KdV

$$\partial_t \eta_\pm \pm \sqrt{\frac{1-\gamma}{\gamma+\delta}} \partial_x \eta_\pm + \epsilon \frac{3c}{2} \frac{\delta^2-\gamma}{\gamma+\delta} \eta_\pm \partial_x \eta_\pm \pm \epsilon \frac{c}{6} \frac{(1+\gamma\delta)}{\delta(\gamma+\delta)} \partial_x^3 \eta_\pm = 0,$$

avec les données initiales déterminées par $\eta_\pm|_{t=0} = \frac{1}{2}\big(\zeta_2 \pm \frac{1}{\gamma+\delta}\, v\big)|_{t=0}$.

[10] Pour des images du satellite, ainsi que des données obtenues par des chaînes de thermistances, on pourra consulter **[52]** ou `http://www.internalwaveatlas.com/Atlas2_index.html` et les références incluses.

On retrouve ainsi la différence fondamentale entre les approximations KdV pour les configurations avec toit solide ou avec surface libre : dans le cas du toit solide, la solution prédite par l'approximation KdV se décompose en deux ondes, alors que la décomposition dans le cadre de la surface libre prédit quatre ondes indépendantes. Deux de ces ondes, correspondant aux coefficients de célérité $\pm c_+$, possèdent les mêmes propriétés que les ondes de surface et en particulier n'autorisent que des solitons d'élévation de la forme (A.15) avec $M > 0$ (ce fait est assuré par la positivité des coefficients c_+, λ_+, μ_+). On parle alors de *mode de surface* ou de *mode rapide*. Les deux autres ondes correspondent donc à l'inverse au *mode interne* ou au *mode lent*. Il existe pour ces ondes une valeur critique du paramètre de ratio des profondeurs δ, en fonction du ratio des densités γ, pour laquelle le coefficient de non-linéarité s'annule et aucun soliton n'est alors possible. Au-delà de cette valeur, les solitons seront de type élévation ; en deçà, ils seront de type dépression. Si cette caractéristique est commune au cas d'un toit rigide et au mode lent dans la configuration avec surface libre, il est intéressant de remarquer que les valeurs critiques δ_c dans les deux configurations peuvent être sensiblement différentes. En particulier, on a toujours $\delta_c > 1$ dans le cas de la surface libre et $\delta_c = \sqrt{\gamma} < 1$ dans le cas du toit solide. Il existe donc une zone, entre les deux courbes de la figure 2, où les deux approximations KdV amènent à des résultats largement différents, puisque les solitons admissibles ne sont pas du même type.

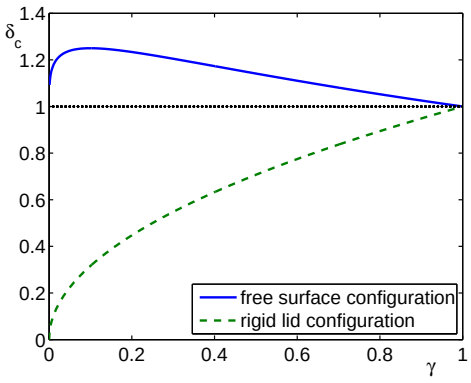


FIGURE 2. Le coefficient critique δ_c en fonction de γ, pour les configurations avec toit solide ou surface libre.

On peut utiliser les décompositions présentées dans la page précédente pour étudier plus en profondeur les amplitudes des déformations pour les différents modes. Tous ces résultats sont présentés au Chapitre 3, Section 3.4 (page 104) et résumés par la figure 3.

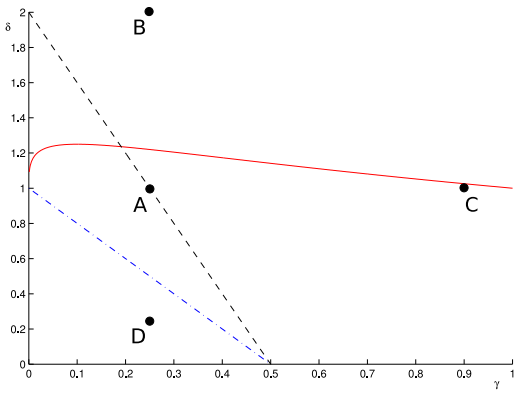


FIGURE 3. Propriétés de l'approximation KdV selon δ et γ.

Le trait plein représente le paramètre critique pour lequel le coefficient de non-linéarité s'annule.

Au dessus de la ligne pointillée ($\delta < 2(1-2\gamma)$), les déformations à l'interface sont plus importantes que celles à la surface dans le mode lent et inversement en dessous. Concernant le mode rapide, les déformations à la surface sont toujours plus importantes qu'à l'interface.

Le trait mixte concerne le cas d'une donnée initiale avec une surface plane et sans vitesse. Les ondes rapides (de surface) sont moins importantes que les ondes lentes (d'interface) au dessus de cette ligne et inversement en dessous ($\delta < 1 - 2\gamma$).

Il apparaît ainsi clairement que l'hypothèse du toit rigide n'est valable que lorsque γ et/ou δ sont grands. Lors de notre étude et pour la justification de nos modèles, on s'est restreint à des valeurs bornées $\delta \in [\delta_{\min}, \delta_{\max}]$ (le cas $\delta \to \infty$ correspond à une couche

de fluide inférieure très petite devant la couche supérieure et s'accorde plus au cadre de l'interface atmosphère-océan). La configuration de toit solide est donc valable uniquement dans le cas d'une très faible différence de densité entre les couches.

Les points **A**, **B**, **C** et **D** représentent les paramètres pour lesquels des simulations numériques ont été réalisées. On les trouvera au Chapitre 3, Section 4.2, page 112.

A.4. Le phénomène des eaux mortes. Dans cette section, on s'intéresse au phénomène d'*eaux mortes*. Ce phénomène est bien connu des marins et a été rapporté en premier lieu par l'explorateur Fridtjof Nansen, dans le rapport de ses expéditions polaires à bord du "Fram" [**92**]. Citons :

> This is a peculiar phenomenon, which occurs where a surface layer of fresh water rests upon the salt water of the sea. It manifests itself in the form of larger or smaller ripples or waves stretching across the wake, the one behind the other, arising sometimes as far forward as almost midships. When caught in dead water, "Fram" appeared to be held back, as if by some mysterious force, and she did not always answer the helm. In calm weather, with a light cargo, "Fram" was capable of 6 to 7 knots. When in dead water she was unable to make 1.5 knots. We made loops in our course, turned sometimes right around, tried all sorts of antics to get clear of it, but to very little purpose.

Il s'agit donc d'une force « mystérieuse » retenant le navire, alors qu'il navigue sur des flots stratifiés. Une caractéristique notable rapportée par Nansen est que, malgré la puissance de la force opposée (capable de diviser par cinq la vitesse du navire), la surface de l'eau reste dans l'ensemble plane. Très tôt, le problème fut soumis à Bjerknes et à son étudiant Ekman. Après l'élaboration d'expériences (voir [**43, 47**]), ce dernier confirma que cette force est due au caractère stratifié des flots et que l'explication de la résistance à l'avancement tient à ce qu'une partie de l'énergie dépensée par le navire est absorbée par la production d'une onde à l'interface entre les deux couches de fluide.

On se propose d'étudier plus en détail les mécanismes du phénomène et d'en déduire dans quelles situations un navire subira l'effet des eaux mortes. Ce travail a été motivé notamment par le récent article de Vasseur, Mercier et Dauxois [**114**], où les auteurs ont reproduit (entre autres) les expériences d'Ekman, et ont analysé en détail leur issue. On trouve relativement peu d'études mathématiques reliées à ce phénomène[11] ; elle sont toujours limitées au cadre de l'approximation linéaire (à l'exception notable de Wu et Chen [**117**] dont l'étude est exclusivement numérique et limitée au cas d'une seule couche de fluide). Nous nous proposons de construire et justifier rigoureusement des modèles non-linéaires, acceptant ainsi des ondes de forte amplitude, et qui reproduisent les caractéristiques principales du phénomène d'eaux mortes.

Notre modélisation est la suivante. Contrairement aux sections précédentes, la surface n'est plus libre, mais au contraire représentée par un toit solide traduisant la présence de la partie submergée du navire. Ainsi, on force

$$\zeta_1 \equiv \zeta_1(x - c_s t), \tag{A.16}$$

où c_s est la vitesse du navire, supposée constante[12]. On peut alors s'appuyer sur l'étude des sections précédentes, afin de décrire le système d'Euler complet lorsque la déformation de la surface est forcée comme en (A.16) et en déduire des modèles approchés. Une analyse de l'onde générée par le navire, ainsi que de la résistance à l'avancement qui l'accompagne, permettent ainsi de prédire pour quels paramètres du système le phénomène d'eaux mortes peut apparaître.

[11] Citons en premier lieu Lamb [**66**] et Sretenskii [**108**], puis les travaux plus récents de Miloh, Tulin et Zilman [**88**], suivis de Nguyen et Yeung [**94**]. Dans un cadre légèrement différent, on peut citer Motygin et Kuznetsov [**91**], Ten et Kashiwagi [**112**] et finalement Lu et Chen [**79**].

[12] Pour simplifier, on se place dans le cas unidimensionnel ($d = 1$) et avec fond plat ($\beta = 0$), bien que le cas général puisse être traité de manière identique.

A.4.1. *Le modèle d'Euler complet, les régimes considérés et la résistance à l'avancement.* Le modèle d'Euler complet, sous sa forme adimensionnée, est obtenu simplement à partir du système dans le cas de la surface libre ($(\Sigma^{\mu,\gamma,\delta}_{\epsilon_1,\epsilon_2,\beta})$ page 8), en posant ζ_1 comme dans (A.16). En translatant le référentiel afin que le navire y soit immobile, on obtient :

$$(\widetilde{\Sigma}) \quad \begin{cases} -\epsilon_1 \operatorname{Fr} \partial_x \zeta_1 \ - \ \dfrac{\epsilon_2}{\mu} G_1(\psi_1,\psi_2) \ = \ 0, \\ \\ (\partial_t - \operatorname{Fr}\partial_x)\zeta_2 \ - \ \dfrac{1}{\mu} G_2\psi_2 \ = \ 0, \\ \\ (\partial_t - \operatorname{Fr}\partial_x)\big(\partial_x\psi_2 - \gamma H(\psi_1,\psi_2)\big) + (\gamma+\delta)\partial_x\zeta_2 + \frac{\epsilon_2}{2}\partial_x\big(|\partial_x\psi_2|^2 - \gamma|H(\psi_1,\psi_2)|^2\big) \\ \qquad\qquad = \ \mu\epsilon_2\partial_x\mathcal{N}, \end{cases}$$

où les opérateurs G_1, G_2 et H sont définis comme précédemment (voir page 8) et

$$\mathcal{N} \ \equiv \ \frac{(\frac{1}{\mu}G_2\psi_2 + \epsilon_2\partial_x\zeta_2\partial_x\psi_2)^2 - \gamma(\frac{1}{\mu}G_2\psi_2 + \epsilon_2\partial_x\zeta_2 H(\psi_1,\psi_2))^2}{2(1+\mu|\epsilon_2\partial_x\zeta_2|^2)}.$$

On a introduit un nouveau paramètre appelé *nombre de Froude* : $\operatorname{Fr} \ \equiv \ c_s/c_0$, qui représente la vitesse relative du navire, comparée à la vitesse caractéristique de l'adimensionnement c_0 (voir Section A.1.3 page 7).

Puisque ζ_1 est un paramètre fixé de notre problème, la première équation du système $(\widetilde{\Sigma})$ est en fait une relation liant ψ_1 et ψ_2, si bien que l'on peut réduire le système à deux équations d'évolution, avec comme inconnues ζ_2 et v le *cisaillement* défini par

$$v \ \equiv \ \partial_x\Big(\big(\phi_2 - \gamma\phi_1\big)\Big|_{z=\epsilon_2\zeta_2}\Big) \ \equiv \ \partial_x\psi_2 - \gamma H(\psi_1,\psi_2). \tag{A.17}$$

Les modèles obtenus seront écrits explicitement sous cette forme.

On considère deux régimes particuliers, correspondant aux régimes présentés page 10, avec quelques hypothèses simplificatrices supplémentaires.

Régime 3 (Navire de petite taille).

$$\mu \ \ll \ 1 \ ; \qquad \alpha \equiv \frac{\epsilon_1}{\epsilon_2} \ = \ \mathcal{O}(\mu) \ , \ 1-\gamma \ = \ \mathcal{O}(\mu).$$

La première hypothèse est bien sûr celle d'une eau peu profonde, qui constitue le socle de notre étude asymptotique. On a ajouté deux hypothèses, peu restrictives, qui simplifient considérablement l'écriture du modèle obtenu. D'abord, on suppose que la partie submergée du navire est petite, comparativement à la profondeur du fluide et à la déformation à l'interface. Des simulations numériques montrent que les ondes générées par cette petite perturbation peuvent atteindre des tailles de l'ordre de l'unité, ce qui explique pourquoi la résistance produite peut être si forte. Ensuite, on suppose que les densités des deux fluides sont presque égales, ce qui est le cas en général. Les modèles obtenus sont présentés dans la section suivante.

Régime 4 (Faibles déformations).

$$\mu \ \ll \ 1 \ ; \qquad \epsilon_2 \ = \ \mathcal{O}(\mu) \ , \ \alpha \ = \ \mathcal{O}(\mu).$$

Dans ce régime, on suppose que l'amplitude des ondes internes générées par le navire resteront faibles, comparativement à la profondeur des deux couches de fluide. On suppose encore que la partie submergée du navire est petite devant ces ondes et donc que le navire subira une résistance relativement forte. Les modèles que l'on obtient se rapprochent donc des modèles d'ondes longues, à savoir de type Boussinesq ou Korteweg-de Vries (où la perturbation joue le rôle d'un terme de forçage). Ces modèles sont présentés et étudiés dans la section A.4.3.

Avant de présenter les modèles asymptotiques obtenus, étudions la force de résistance à l'avancement que subit le navire. Cette force est le reflet de l'énergie dépensée par le navire afin de déplacer le volume d'eau nécessaire à son avancement et est définie par

$$R_W \equiv \int_{\Gamma_{\text{ship}}} P\,(-\mathbf{e}_x \cdot \mathbf{n})\,\mathrm{d}S,$$

où Γ_{ship} est la partie submergée du navire, P la pression, $\mathbf{n}$ la normale unitaire extérieure à la coque et $\mathbf{e}_x$ le vecteur horizontal unitaire. Grâce à la loi de Bernoulli (A.3) satisfaite par le fluide (Postulat 4, page 5), on peut réécrire cette formule en utilisant uniquement les données de notre problème. Pour simplifier, nous écrivons ici directement les approximations au premier ordre obtenues dans chacun des régimes considérés (voir Chapitre 4, Section C, page 159). On note ces approximations C_W et on écrit *coefficients de résistance*, car ces coefficients sont obtenus après adimensionnement des variables.

CAS DU RÉGIME 3. En utilisant les hypothèses du régime 3, on obtient :

$$C_W = -\int_{\mathbb{R}} \left(\left(\zeta_2 + \frac{\epsilon_2}{2}\zeta_2^2\right) + \epsilon_2 \frac{\delta\ h_2 v}{1+\delta} + \epsilon_2 \left(\frac{\delta\ h_2 v}{1+\delta}\right)^2 \right) \frac{\mathrm{d}\zeta_1}{\mathrm{d}x}\ \mathrm{d}x + \mathcal{O}(\mu).$$

CAS DU RÉGIME 4. Les hypothèses du régime 4 amènent simplement à :

$$C_W = -\int_{\mathbb{R}} \zeta_2 \frac{\mathrm{d}\zeta_1}{\mathrm{d}x}\ \mathrm{d}x + \mathcal{O}(\mu).$$

On remarque que le coefficient de résistance sera faible si l'onde est faible ou localisée loin du navire (ce qui est évident), mais également si l'onde est symétrique et centrée sous le navire (une simple analyse de parité permet de s'en convaincre). Au contraire, le navire subira une résistance importante si une onde d'élévation est produite à la poupe du navire. C'est dans ce cas que le phénomène d'eaux mortes apparaîtra.

Pour résumer, le phénomène s'explique de la manière suivante : *un corps naviguant dans des eaux stratifiées peut générer une onde interne d'élévation dans sa traîne, qui provoque en retour une forte résistance à l'avancement.*

Voyons si les modèles asymptotiques obtenus dans nos régimes permettent de prédire ce phénomène.

A.4.2. *Le régime fortement linéaire.* Le modèle que l'on propose dans le cadre du régime 3 est le suivant :

$(\widetilde{\mathcal{M}})$

$$\begin{cases} (\partial_t - \mathrm{Fr}\,\partial_x)\zeta_2 \ +\ \partial_x\left(\dfrac{h_1 h_2}{h_1+\gamma h_2} w + h_2 \dfrac{\alpha\,\mathrm{Fr}}{h}\zeta_1\right) \ = 0, \\ \\ (\partial_t - \mathrm{Fr}\,\partial_x)\,(w - \mu\mathcal{S}_1 w) \ +\ (\gamma+\delta)\partial_x\zeta_2 \ +\ \epsilon_2\partial_x\left(\dfrac{1}{2}\dfrac{h_1^2-\gamma h_2^2}{(h_1+\gamma h_2)^2} w^2 + w\dfrac{\alpha\,\mathrm{Fr}}{h}\zeta_1\right) \\ \qquad\qquad + \mu\epsilon_2\partial_x\,(w\ \mathcal{S}_2 w) \ =\ 0, \end{cases}$$

où $h \equiv 1+\frac{1}{\delta}$, $h_1 \equiv 1+\epsilon_1\zeta_1-\epsilon_2\zeta_2$, $h_2 \equiv \frac{1}{\delta}+\epsilon_2\zeta_2$ et les opérateurs $\mathcal{S}_1$ et $\mathcal{S}_2$ sont définis par

$$\mathcal{S}_1 w \equiv \frac{1}{3}\left(\partial_x^2\big((h-h_2)h_2 w\big) - (\partial_x h_2)^2 w\right),$$
$$\mathcal{S}_2 w \equiv \frac{(h-h_2)h_2}{3h}\left((\partial_x^2 h_2)w + 2(\partial_x w)(\partial_x h_2)\right) + \frac{h-2h_2}{2h}(\partial_x h_2)^2 w.$$

Ce modèle est obtenu à partir de $(\widetilde{\Sigma})$, en y plongeant les développements asymptotiques des opérateurs G_1, G_2 et H obtenus dans la Proposition A.3, page 12. Les hypothèses du régime 3 permettent de simplifier de nombreux termes et notamment d'écrire simplement ψ_1 et ψ_2 en fonction du cisaillement défini en (A.17) :

$$h_2 v \;=\; -h\,\partial_x\psi_1 + \mathcal{O}(\mu), \qquad h_1 v \;=\; h\,\partial_x\psi_2 + \mathcal{O}(\mu).$$

La méthode présentée à la section A.2 permet alors d'obtenir la justification de ce modèle :

PROPOSITION A.12. *Le système d'Euler complet* $(\widetilde{\Sigma})$ *est cohérent avec le modèle* $(\widetilde{\mathcal{M}})$, *à la précision* $\mathcal{O}(\mu^2)$.

Des simulations numériques permettent alors de vérifier que le modèle prédit bien la génération d'une importante onde d'élévation, située juste derrière le navire. On présente l'une de ces simulations (voir Chapitre 4, Section 3.2, page 139 pour plus de détails) dans la figure 4. Le panneau de gauche représente la déformation à l'interface, en fonction des variables d'espace ($x \in [-20, 20]$) et du temps ($t \in [0, 15]$). L'évolution du coefficient de résistance en fonction du temps est donnée dans le panneau de droite et le dernier panneau représente l'état du système au temps final $t = 15$. Le phénomène d'eaux mortes est bien observé avec le jeu de paramètres choisis.

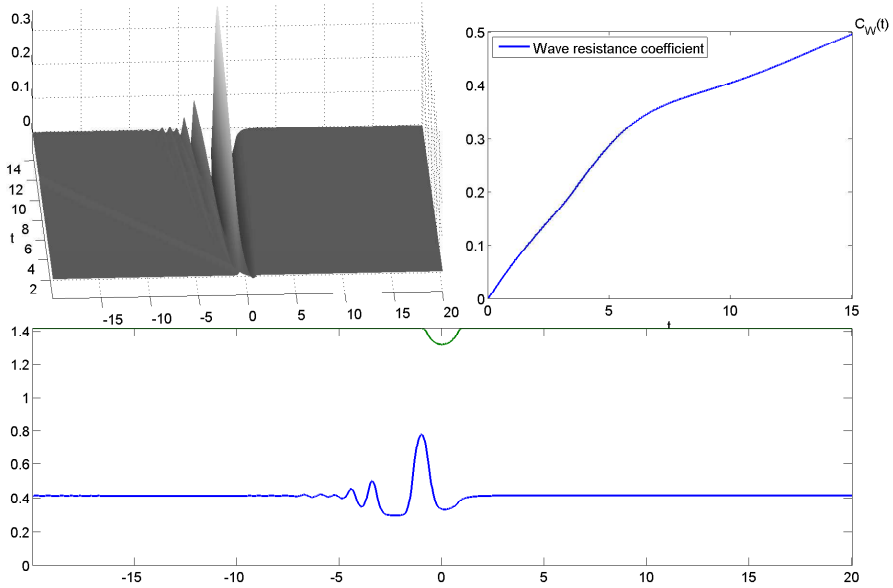


FIGURE 4. Prédiction du modèle $(\widetilde{\mathcal{M}})$, avec une donnée initiale nulle. $\epsilon_2 = 1$, $\alpha = \mu = 0.1$, $\gamma = 0.99$, $\mathrm{Fr} = 1$, $\delta = 12/5$.

A.4.3. *Le régime faiblement linéaire.* Le cadre du régime 4 permet de simplifier considérablement les modèles (en échange d'un domaine de validité plus restreint, puisque seules les déformations d'amplitude modérées sont acceptées). On aboutit alors à des modèles de type Boussinesq et Boussinesq symétrique comme à la section A.2.3 ; ce dernier modèle est justifié par un résultat de convergence à l'ordre $\mathcal{O}(\mu)$ sur des temps de l'ordre $\mathcal{O}(1/\mu)$. De ce modèle, on déduit grâce au développement BKW l'approximation KdV correspondante et c'est cette approximation que nous allons présenter plus en détail[13].

[13]Le lecteur intéressé pourra se reporter à la section 4 du Chapitre 4 (page 142) pour une présentation et une justification complètes de ces modèles.

Le modèle obtenu est le suivant. On va pouvoir approcher $V \equiv (\zeta_2, v)$ une solution du système d'Euler complet $(\widetilde{\Sigma})$ par la décomposition

$$V_{\mathrm{KdV}} \;\equiv\; \big(\, \eta_+ + \eta_- \;,\; (\gamma+\delta)(\eta_+ - \eta_-) \,\big),$$

où η_+ et η_- sont les solutions de l'équation de Korteweg-de Vries *avec un terme de forçage*

$$\text{(fKdV)}\;\; \partial_t \eta_\pm \,+\, (-\mathrm{Fr} \pm 1)\partial_x \eta_\pm \,\pm\, \epsilon_2 \frac{3}{2}\frac{\delta^2-\gamma}{\gamma+\delta}\eta_\pm \partial_x \eta_\pm \,\pm\, \mu \frac{1}{6}\frac{1+\gamma\delta}{\delta(\gamma+\delta)}\partial_x^3 \eta_\pm \;=\; -\alpha\,\mathrm{Fr}\,\gamma \frac{\mathrm{d}\zeta_1}{\mathrm{d}x},$$

avec la donnée initiale adaptée $\eta_\pm\,|_{t=0} = \frac{1}{2}\big(\zeta_2 \pm \frac{1}{\gamma+\delta}\, v\big)\,|_{t=0}$.

Ce modèle est justifié par la proposition suivante :

PROPOSITION A.13. *Pour μ suffisamment petit, le système d'Euler complet $(\widetilde{\Sigma})$ et l'approximation KdV sont convergents à l'ordre $\mathcal{O}(\sqrt{\mu})$ sur des temps de l'ordre $\mathcal{O}(1/\mu)$.*

*De plus, si la donnée initiale est suffisamment localisée (*i.e. *décroissante à l'infini), alors l'approximation est d'ordre $\mathcal{O}(\mu)$.*

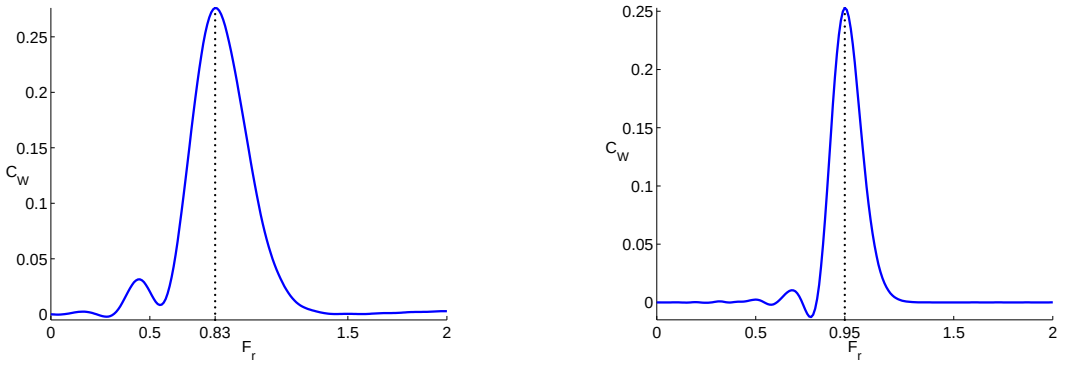
L'équation KdV avec un terme de forçage a été largement étudiée, notamment dans le cadre du problème à une couche (où un déplacement du fond ou d'une pression créé le terme de forçage). En particulier, on s'est intéressé au problème de génération d'ondes solitaires vers l'avant, qui a été découvert numériquement par Wu et Wu **[116]** et confirmé par des expériences par Lee **[74]** (citons **[119, 75, 107, 166]**). Il est apparu que le nombre de Froude Fr avait un rôle capital à jouer et que la génération d'ondes n'est importante que pour un petit intervalle de valeurs de Fr autour de l'unité.

On confirme ce phénomène par le résultat suivant (précisé Proposition 4.3, page 149) :

PROPOSITION A.14. *Si $\mathrm{Fr}-1 \neq 0$, alors il existe des constantes $\mu_0(\mathrm{Fr}-1)$, $C(\frac{1}{\mathrm{Fr}-1})$ et $T(\mathrm{Fr}-1)$ telles que pour $\mu < \mu_0$,*

$$\big|V_{KdV}\big|_{L^\infty([0,T/\mu];H^s)} \;\leq\; C\mu.$$

Ainsi, le phénomène d'eaux mortes sera toujours faible si le nombre de Froude est éloigné de sa valeur critique $\mathrm{Fr} = 1$. Les simulations numériques indiquent que le maximum est en fait atteint pour des valeurs légèrement sous-critiques (voir Figure 5 ci-dessous).



(a) $\alpha = \epsilon_2 = \mu = 0.1$, $\delta = 5/12$, $\gamma = 0.99$.

(b) $\alpha = \epsilon_2 = \mu = 0.1$, $\delta = 12/5$, $\gamma = 0.99$.

FIGURE 5. Coefficient de résistance C_W prédit par le modèle (fKdV) au temps $T = 10$, en fonction du nombre de Froude Fr.

A.4.4. *Conclusion et perspectives.* Les différents modèles que nous avons présenté, qu'ils soient dérivés dans le cadre fortement non-linéaire du Régime 4, ou dans le cadre faiblement non-linéaire du Régime 3, permettent de retrouver les principaux aspects du phénomène d'eaux mortes, tels que présentés dans [**114**]. En particulier, on a :

i. des ondes internes transverses sont produites à l'arrière d'un navire lorsqu'il navigue sur des eaux stratifiées ;
ii. le corps subit une résistance positive lorsqu'une onde interne d'élévation se situe en dessous de sa poupe ;
iii. le phénomène est important lorsque la vitesse du navire est proche d'une vitesse critique (la vitesse maximale des ondes internes) ;
iv. la résistance maximale est atteinte pour des valeurs légèrement sous-critique.

Nos modèles permettent également d'explorer les effets des différents paramètres du système sur le phénomène, et en particulier le rapport des profondeurs des deux couches de fluide. Il apparaît (voir Section 3.2) que le phénomène d'eaux mortes est plus important lorsque la couche supérieure est plus épaisse, ce qui est contraire à l'intuition puisque dans ce cas, l'onde d'élévation est éloignée de la coque du navire. Ceci est dû au fait que le coefficient de résistance ne dépend pas de la distance entre l'onde interne et le navire, mais plutôt de l'amplitude et de la déclivité de cette dernière.

Pourtant, une caractéristique importante du phénomène d'eaux mortes n'est pas prédite par nos modèles. En effet, les expériences (voir [**114, 160, 43**]) présentent un comportement périodique, que nos simulations ne permettent pas de retrouver. Ceci est une conséquence des hypothèses de notre cadre de travail : nous avons supposé que la vitesse du corps reste immuable, alors que les différentes expériences utilisent la contrainte plus naturelle d'une force constante appliquée au navire.

Afin de représenter grossièrement ce cadre, nous avons modifié nos schémas numériques, en ajustant à chaque pas de temps la vitesse du navire, tenant compte de la force de résistance qu'il subit. On reproduit dans la figure 6 le résultat d'une simulation obtenue dans cette configuration.On voit très nettement réapparaître le phénomène d'hystérésis décrit dans [**114**], avec une période de l'ordre de $\Delta T \approx 10$. Pourtant, le modèle utilisé est heuristique, et il serait intéressant de construire et justifier rigoureusement un système où la vitesse du corps ne serait pas supposée constante.

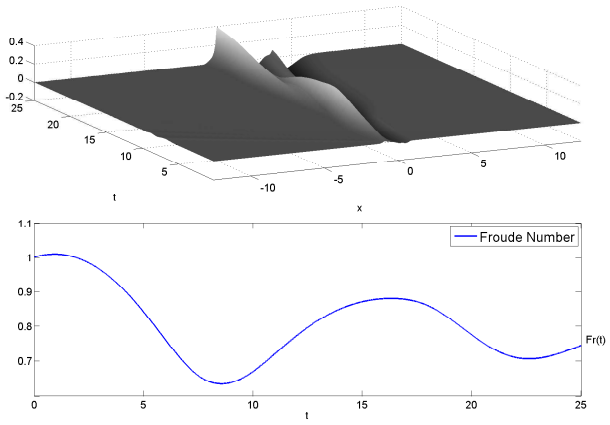


FIGURE 6. Comportement du flot, lorsque Fr n'est pas constant.

Par ailleurs, notre travail est limité au cas de la dimension horizontale $d = 1$. Étendre notre étude à la configuration $d = 2$ serait une importante contribution, et permettrait d'observer les ondes divergentes générées par le corps en marche, en même temps que les ondes transverses. On pourra consulter à ce sujet [**88, 120**].

Pour conclure, on présente dans le tableau 2, page suivante, un résumé des conclusions de notre analyse, fruits des résultats ci-dessus et de simulations numériques.

Nombre de Froude	Cas sous-critique : Fr < 1	Cas critique : Fr $= 1$		Cas surcritique : Fr > 1
Profondeur des deux couches		couche inférieure plus épaisse	couche supérieure plus épaisse	
Régime 3	Les ondes générées sont de faible amplitude, et les conclusions du Régime 4 sont valides.	Génération d'une onde d'élévation-dépression « molle » derrière le navire. Résistance modérée. Voir Figure 3, page 140	Génération d'une importante onde d'élévation « raide » derrière le navire. Forte résistance. Voir Figure 4, page 141	Les ondes générées sont de faible amplitude, et les conclusions du Régime 4 sont valides.
Régime 4	Génération en continu de vaguelettes à l'avant et à l'arrière du navire. Très faible résistance. Voir Figure 6, page 150	Génération périodique (avec une très grande période) d'ondes solitaires de dépression à l'avant du navire. Résistance oscillante, avec moyenne positive. Voir Figure 9, page 152	Génération périodique (avec une très grande période) d'ondes solitaires d'élévation à l'avant du navire. Résistance oscillante, avec moyenne nulle. Voir Figure 9, page 152	Génération de quelques vaguelettes à l'arrière du navire et convergence vers un état statique. Pas de résistance permanente. Voir Figure 7, page 151

TABLE 2. Prédictions des modèles, en fonction du nombre de Froude et du ratio des profondeurs.

Partie B. Propagation des ondes à travers un milieu non homogène

B.1. Présentation générale du problème. Dans cette partie, nous nous intéressons à la modélisation de la propagation d'une onde lumineuse à travers les cristaux photoniques. Ces cristaux photoniques possèdent une structure périodique dans une, deux ou trois dimensions. L'étude de ces cristaux a attiré récemment[14] une attention considérable au vu des applications potentielles nombreuses d'un tel matériau. En effet, les variations rapides de son indice de réfraction permettent de faire apparaître des bandes d'énergie interdites pour les photons, bloquant ainsi la propagation de la lumière dans certaines directions et pour certaines énergies. En cela, ces matériaux se comportent selon le même principe que les semi-conducteurs pour les électrons, qui sont au centre des composants électriques modernes (diode, transistor, *etc.*). Introduire des éléments photoniques dans les circuits actuels autoriserait donc à traiter toute l'information sous forme lumineuse, ce qui permet d'envisager la réalisation de dispositifs nanométriques. Le lecteur pourra se référer à [**154**] par exemple et à la bibliographie [**123**], pour une présentation détaillée des propriétés et applications de ces matériaux.

Pour toute utilisation pratique des cristaux photoniques, il est nécessaire de comprendre l'influence de la géométrie, la disposition et les échelles de leur structure, sur leurs propriétés. Une compréhension mathématique précise et efficace de la propagation de la lumière dans des milieux fortement hétérogènes est ainsi nécessaire à la conception et à l'analyse de ces nouveaux matériaux. Nos travaux portent sur l'influence de la présence de « défauts » dans la structure, dans un sens que l'on précisera par la suite.

Cette section particulière est dédiée à la présentation du problème que l'on étudie et des équations mathématiques qui le gouvernent. On commence par rappeler quelques éléments de la théorie des guides d'ondes (pour de plus amples détails, on pourra se rapporter à [**133, 173**] par exemple) et tout particulièrement la diffusion en milieu non homogène. On précise ensuite le type de structures considérées et les questions précises qui sont abordées dans les sections suivantes.

B.1.1. *Propagation de la lumière dans les guides d'ondes.* L'optique ondulatoire se propose d'étudier la propagation de la lumière en la considérant comme étant une onde électromagnétique, vérifiant donc les équations de Maxwell. Dans une région sans charge ni courant, ces équations s'écrivent simplement :

$$\nabla \cdot (\varepsilon \mathbf{E}) \;=\; 0, \quad \nabla \cdot \mathbf{B} \;=\; 0, \quad \nabla \wedge \mathbf{E} \;=\; -\partial_t \mathbf{B}, \quad \nabla \wedge (\mathbf{B}/\mu) \;=\; \varepsilon \partial_t \mathbf{E},$$

où $\mathbf{E}(t, X) \equiv (E_1, E_2, E_3)$ et $\mathbf{B}(t, X) \equiv (B_1, B_2, B_3)$ sont respectivement les champs électriques et magnétiques ; $\varepsilon(X)$ et $\mu(X)$ sont la permittivité et la perméabilité du milieu. On a utilisé les notations classiques ∂_t pour la dérivée temporelle et $\nabla\cdot$ et $\nabla\wedge$ sont des dérivées spatiales ($X = (x, y, z) \in \mathbb{R}^3$), respectivement la divergence et le rotationnel.

Afin de modéliser un guide d'onde, on suppose qu'il existe une direction d'espace invariante (disons z) et que le matériau est non-magnétique. Ainsi, on a $\varepsilon = \varepsilon(x, y)$ et $\mu \equiv \mu_0$. Dans ce cas, toute solution se décompose en une onde *transverse électrique*, dont la composante du champ électrique E_z est nulle et une onde *transverse magnétique*, dont la composante du champ magnétique B_z est nulle. On s'intéresse ici à la propagation de la composante transverse électrique $\mathbf{E}_\perp \equiv (E_x, E_y, 0)$. Plus particulièrement, on s'intéresse aux solutions harmoniques se propageant dans la direction z, c'est-à-dire

$$\mathbf{E}_\perp(t, X) \;\equiv\; \mathbf{E}_\beta(t, x, y) e^{i(\beta z - \omega t)}.$$

Il est aisé de vérifier que les composantes d'une telle solution vérifient l'équation

$$\nabla_\perp(\nabla_\perp \cdot \mathbf{E}_\beta) - \nabla_\perp^2 \mathbf{E}_\beta \;=\; \nabla \wedge (\nabla \wedge \mathbf{E}_\beta) \;=\; -\mu_0 \varepsilon \partial_t^2 \mathbf{E}_\beta \;=\; \omega^2 \mu_0 \varepsilon \mathbf{E}_\beta,$$

[14] La notion de cristal photonique date de 1987 [**179**] et le premier cristal photonique à trois dimensions a été réalisé en 1991 [**180**].

où l'on utilise le symbole $\nabla_\perp$ pour dénoter le gradient, la divergence et le Laplacien par rapport aux variables (x, y). Dans le cadre classique de l'approximation scalaire, le terme

$$\nabla_\perp(\nabla_\perp \cdot \mathbf{E}_\beta) \;=\; \nabla_\perp \left(\frac{\mathbf{E}_\beta \cdot \nabla_\perp \varepsilon}{\varepsilon} \right),$$

qui possède seul des interactions entre les différentes composantes vectorielles du champ électrique, est négligé. Il convient de noter que par la nature du matériau étudié, cette approximation ne sera valable que lorsque le champ $\mathbf{E}_\perp$ est faible à l'emplacement des interfaces de la structure (où $\nabla_\perp \varepsilon$ contribue fortement). Ceci n'est pas le cas en général, mais reste vérifié dans beaucoup de situations (voir **[143]**). On supposera que l'approximation est valable pour la suite de l'étude et ainsi que chaque composante du champ électrique transversal vérifie une équation scalaire découplée. De plus, on se limitera dans cette étude au cas du guide d'onde planaire (c'est à dire où y est également une direction d'espace invariante). Se limiter au cas d'une structure unidimensionnelle peut paraître une simplification grossière, dans le sens où les effets du matériau sur la propagation de la lumière ne s'appliqueront parfaitement qu'aux ondes à incidence normale (voir Figure 7 page 31). Pourtant, ce système simple illustre la majeure partie des phénomènes physiques mis en jeu dans les cristaux photoniques à deux ou trois dimensions et en particulier la structure en bande des énergies admises (voir par exemple **[151]**). En utilisant les notations suivantes :

$$k^2 \;=\; \mu_0\varepsilon_0 - \beta^2, \quad \text{et} \quad V(x) \;=\; \omega^2\mu_0(\varepsilon_0 - \varepsilon(x)),$$

où ε_0 est une permittivité de référence, le système est donc réduit à l'équation

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} \;+\; V(x) \right) \; u \;=\; k^2 \; u. \tag{B.18}$$

Cette équation est l'équation de Helmholtz unidimensionnelle en milieu hétérogène, où k est un paramètre réel fixé et le *potentiel* $V(x)$ porte les informations du milieu. L'équation (B.18) correspond également à l'équation de Schrödinger stationnaire (*i.e.* indépendante du temps). Cette similitude n'est pas surprenante et l'on peut étendre considérablement l'analogie entre électromagnétisme et mécanique quantique (voir **[154]**). Ainsi, notre étude s'applique également à l'opérateur de Schrödinger dans le cadre quantique.

B.1.2. *Coefficients de réflexion, transmission et ondes planes distordues.* Nous présentons ici quelques résultats classiques concernant l'opérateur de Schrödinger et tout particulièrement la théorie de diffusion (ou *scattering*). On trouvera les preuves de ce que l'on avance, ainsi qu'une étude approfondie, dans **[125, 135, 130]** par exemple.

CAS DU MILIEU HOMOGÈNE. Dans le cas d'un potentiel nul ($V(x) \;\equiv\; 0$), on obtient l'équation de Helmholtz homogène scalaire. La solution générale de cette équation s'écrit sous la forme

$$u(x) \;=\; A\; e^{ikx} \;+\; B\; e^{-ikx},$$

avec A et B des constantes complexes choisies librement. Les solutions particulières $e^{\pm ikx}$, appelées *ondes planes* (on choisira cette dénomination pour les seules ondes planes d'amplitude 1), forment donc une base de l'espace des solutions. On voit en particulier qu'on n'a pas unicité des solutions. Plus précisément, on a le résultat suivant :

PROPOSITION B.15. *Le spectre de l'opérateur autoadjoint positif (non borné) défini par* $H_0 \;:\; \left\{ \begin{array}{ccc} H^2 & \to & L^2 \\ u & \mapsto & -\frac{d^2}{dx^2}u \end{array} \right.$ *est purement continu et égal à l'intervalle* $\mathbb{R}^+$.

Il est donc impossible d'inverser l'opérateur $H_0 \;-\; k^2$ pour $k \in \mathbb{R}$. Il va alors falloir restreindre l'espace des solutions, c'est-à-dire imposer un comportement à l'infini pour

définir la « bonne » solution de l'équation de Helmholtz avec terme source

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} \; - \; k^2 \right) \; u \; = \; f. \tag{B.19}$$

Les calculs explicites du cas où k^2 n'appartient pas au spectre de H_0 (*i.e.* $k^2 \in \mathbb{C} \setminus \mathbb{R}^+$) nous poussent naturellement à définir la *fonction de Green* suivante :

$$G_0(x;k) \; \equiv \; \frac{ie^{ik|x|}}{2k}.$$

Cette fonction de Green vérifie au sens des distributions l'équation précédente, dans le cas où $f \equiv \delta$, la distribution de Dirac centrée en 0. De plus, cette fonction vérifie la *condition de radiation de Sommerfeld* ou condition d'onde sortante, définie ci-dessous.

Definition B.16 (Ondes sortantes). *On dit qu'une fonction $u(x)$ vérifie la* condition de radiation sortante *ou plus simplement qu'elle est* sortante *lorsque $|x| \to \infty$, si l'on a*

$$(\; \partial_x \; \mp \; ik \;) \, u_s \; \to \; 0, \quad \textit{lorsque} \quad x \to \pm\infty.$$

La fonction de Green permet donc d'obtenir la solution *sortante* de (B.19) par convolution :

$$u_f(x) \; \equiv \; G_0 \star f \; \equiv \int_{-\infty}^{+\infty} G_0(x-y;k) f(y) \; \mathrm{d}y.$$

Il convient de noter que le produit de convolution n'est pas défini pour n'importe quel terme source f. En particulier, puisque G_0 n'est pas intégrable, u_f n'est pas défini dans L^2 si $f \in L^2$. On voit ici apparaître l'importance du comportement à l'infini de nos fonctions. En particulier, si on suppose $f \in L^2_c$, c'est-à-dire de carré intégrable et à support compact, alors on a le résultat suivant :

Proposition B.17. *Pour $f \in L^2_c(\mathbb{R})$ et $k \in \mathbb{R}$, il existe une unique solution $u \in H^2_{loc}(\mathbb{R})$ vérifiant les conditions de radiation sortante.*

Cette solution est bien évidemment celle obtenue par convolution avec la fonction de Green. On a noté $H^2_{\mathrm{loc}}(\mathbb{R})$ l'espace des fonctions localement H^2.

On appelle alors *résolvante sortante* associée à l'opérateur $H_0 \; - \; k^2$ l'opérateur $R_0(k)$ défini par

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} \; - \; k^2 \right)^{-1} \; \equiv \; R_0(k) f \; \equiv \; \int_{-\infty}^{+\infty} G_0(x-y;k) f(y) \; \mathrm{d}y.$$

Cas du milieu hétérogène. On veut maintenant généraliser cette étude au cas de potentiels non-nuls (*i.e.* $V \; \neq \; 0$ dans (B.18)), qui est le cas qui nous intéresse. Pour simplifier, on se limite ici à des potentiels ne possédant pas de spectre discret. C'est le cas pour un large espace de potentiels suffisamment décroissants à l'infini. On a en particulier (voir [**135**])

Proposition B.18. *Si $V \in L^{1,2}$, c'est-à-dire vérifie*

$$\|V\|_{L^{1,2}} \; \equiv \; \int_{\mathbb{R}} (1+|s|)^2 \; |V(s)| \; ds \; < \; \infty,$$

alors le spectre continu de l'opérateur $H_0 \; + \; V \; \equiv \; -\frac{d^2}{dx^2} \; + \; V$ est égal à $\mathbb{R}^+$.

On peut définir l'équivalent des ondes planes dans le cas homogène.

DEFINITION B.19 (Ondes planes distordues). *On définit $e_{V\pm}(x;k)$, les ondes planes distordues associées au potentiel $V(x)$, comme les uniques solutions du problème* (B.18) *telles que*

$$e_{V\pm}(x;k) \;=\; e^{\pm ikx} \;+\; u_s(x), \tag{B.20}$$

où u_s est une onde sortante (on l'appelle le champ diffracté*).*

Il est intéressant de remarquer que lorsqu'on applique la résolvante R_0 à l'équation (B.18), on obtient que les *ondes planes distordues* vérifient l'équation suivante :

$$e_{V\pm}(x;k) \;=\; e^{\pm ikx} \;-\; R_0(k)\big(\, Ve_{V\pm}(\cdot;k)\,\big)\;. \tag{B.21}$$

Cette équation intégrale est appelée *équation de Lippmann-Schwinger* [**159**]. Elle peut être résolue grâce à l'alternative de Fredholm et garantit le caractère bien posé de la Définition B.19.

Ces ondes planes distordues vérifient le comportement à l'infini suivant :

$$\left\{\begin{array}{lll} e_{V+}(x;k) \;-\; \big(e^{ikx} + r_1(k)e^{-ikx}\big) \;\longrightarrow\; 0 & \text{quand } x \to -\infty, \\ e_{V+}(x;k) \;-\; t(k)e^{ikx} \;\longrightarrow\; 0 & \text{quand } x \to +\infty, \\ \\ e_{V-}(x;k) \;-\; t(k)e^{-ikx} \;\longrightarrow\; 0 & \text{quand } x \to -\infty, \\ e_{V-}(x;k) \;-\; \big(e^{-ikx} + r_2(k)e^{ikx}\big) \;\longrightarrow\; 0 & \text{quand } x \to +\infty. \end{array}\right. \tag{B.22}$$

On appelle $t(k)$ ainsi défini le *coefficient de transmission* associé au potentiel $V(x)$; $r_1(k)$ et $r_2(k)$ sont les *coefficients de réflexion*, définis pour $k \in \mathbb{R} \setminus \{0\}$.

On vérifie aisément que $t(k)$ peut se définir également *via* le Wronskien :

$$\mathrm{Wron}\,[e_{V+}(x;k), e_{V-}(x;k)] \;\equiv\; e_{V+}(\partial_x e_{V-}) \;-\; (\partial_x e_{V+})e_{V-} \;=\; -2ik\; t(k).$$

De simples calculs permettent alors de montrer que la résolvante de l'équation s'écrit sous la forme

$$\begin{aligned} \big(\; R_V(k)f \;\big)(x) \;&= \left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x) - k^2\right)^{-1} \; f(x) \;=\; \int_{-\infty}^{\infty} R_V(x,y;k) f(y)\; \mathrm{d}y \\ &\equiv\; \frac{1}{-2ik\; t(k)} \;\int_{-\infty}^{x} f(y)\; e_{V-}(y;k) e_{V+}(x;k)\; \mathrm{d}y \\ &\quad +\; \frac{1}{-2ik\; t(k)} \int_{x}^{+\infty} f(y)\; e_{V+}(y;k) e_{V-}(x;k)\; \mathrm{d}y. \end{aligned} \tag{B.23}$$

Ceci nous permet notamment d'obtenir le résultat suivant :

PROPOSITION B.1. *Soit $f \in L^1(\mathbb{R})$. Alors, pour $k \neq 0$, l'équation non homogène*

$$\left(-\frac{d^2}{dx^2} + V(x) - k^2\right)\; U \;=\; f \tag{B.24}$$

possède une unique solution sortante : $U = R_V(k)f$.

Idée de la preuve. L'existence est immédiatement offerte par la représentation intégrale en (B.23). L'unicité est obtenue en étudiant la différence de deux solutions et provient du fait qu'il n'existe pas de solution non nulle du problème $(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x) - k^2)u_s \;=\; 0$, avec $u_s(x)$ sortante lorsque $|x| \to \infty$, si $\Im(k^2) \;\geq\; 0$ (voir par exemple [**176**]). □

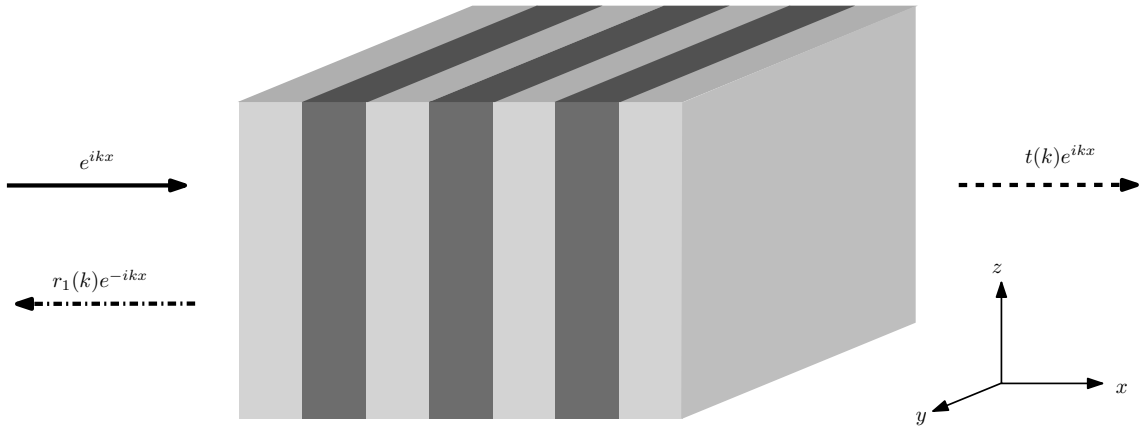


FIGURE 7. Diffusion à travers un matériau périodique unidimensionnel

B.1.3. *Présentation précise du problème.* Les coefficients de transmission et de réflexion sont au centre de notre étude, car ils représentent des données physiques naturelles du problème. En effet, on peut comprendre l'onde plane distordue $e_{V+}(x;k)$ comme la réponse du système à l'excitation par une onde monochromatique e^{ikx} venant de l'infini. Dans ce cas, le coefficient $|r_1(x;k)|^2$ représente l'énergie de l'onde réfléchie par le matériau et $|t(k)|^2$ l'énergie restante après avoir diffusé à travers le matériau[15] (voir Figure 7). C'est donc le contrôle de ces coefficients, en fonction de la forme du potentiel V, qui nous intéresse.

Ainsi, notre motivation première, qui peut se résumer en la question

« *Quelle est l'influence de la structure du matériau sur la propagation de la lumière ?* »

se traduit mathématiquement par

« *Quelle est la dépendance du coefficient de transmission en fonction du potentiel ?* »

La nature des matériaux étudiés nous pousse à envisager des potentiels pathologiques. Ces potentiels pourront comporter :

1. une composante singulière ; et/ou

2. une microstructure, pas nécessairement lisse.

La composante singulière est modélisée par la présence d'un nombre fini de distributions de Dirac. Ces distributions de Dirac forment des interfaces, à travers laquelle une partie de l'onde est transmise et l'autre partie réfléchie. Nous présentons dans la prochaine section quelques résultats montrant que la structure en bande des énergies admises apparaît dès la présence d'un petit nombre de distributions de Dirac équidistribuées. Le but de cette section est également décrire les résultats obtenus au Chapitre 5, généralisant les résultats connus pour des potentiels réguliers et d'étendre ainsi la théorie aux potentiels admettant des singularités.

La microstructure est modélisée par une fonction à très courte longueur caractéristique. Cette fonction sera périodique selon une variable rapide ou plus généralement localisée dans les hautes fréquences. On s'attache dans la section B.3 et au Chapitre 6 au développement asymptotique du coefficient de transmission lorsque le paramètre mesurant la longueur caractéristique de la microstructure tend vers 0.

Une attention particulière est apportée à la présence de « défauts » dans le système. Ces défauts sont modélisés par la présence de la composante singulière et par le caractère non lisse de la microstructure. On verra que ces défauts, même s'ils sont localisés et en nombre fini, ont une influence parfois plus importante que la microstructure elle-même.

[15]On a d'ailleurs toujours conservation de l'énergie : $|r_1(k)|^2 + |t(k)|^2 = 1$.

B.2. Potentiels admettant des singularités. Dans cette section, nous allons examiner l'effet de l'existence d'une composante singulière dans le potentiel. Nous allons donc considérer que le potentiel peut s'écrire sous la forme

$$V(x) \equiv V_{\text{sing}}(x) + V_{\text{reg}}(x), \tag{V1}$$

où l'on suppose vérifiées les hypothèses suivantes :

i. *Pour la composante singulière, V_{sing} :*

$$V_{\text{sing}}(x) = \sum_{j=0}^{N-1} c_j \, \delta(x - x_j), \quad \text{où} \quad c_j, x_j \in \mathbb{R}, \quad x_j < x_{j+1}.$$

ii. *Pour la composante régulière, V_{reg} :* $V_{\text{reg}} \in L^{1,3/2+}(\mathbb{R})$ avec

$$\|V\|_{L^{1,\rho}} \equiv \int_{\mathbb{R}} (1 + |s|)^{\rho} \, |V(s)| \, \mathrm{d}s < \infty.$$

On a donc une composante régulière suffisamment décroissante à l'infini et une composante singulière formée d'une somme finie de distributions de Dirac. La composante singulière nécessite une définition précise du sens dans lequel on entendra les solutions de l'équation de Helmholtz.

DEFINITION B.20. *On dit que u est une solution de l'équation* (B.18) *si u est C^2 par morceaux, vérifie* (B.18) *en dehors du support de V_{sing}, $\mathbb{R} \setminus \{x_0, \ldots, x_{N-1}\}$, ainsi que les* conditions de saut

$$\begin{cases} [u]_x = 0, \quad x \in \mathbb{R}, \\ \left[\dfrac{d}{dx} u\right]_x = 0 \ \textit{si} \ x \in \mathbb{R} \setminus \operatorname{supp} V_{sing}, \\ \left[\dfrac{d}{dx} u\right]_{x_j} = c_j u(x_j) \ \textit{où} \ x_j \in \operatorname{supp} V_{sing}, \end{cases} \tag{B.25}$$

où l'on a utilisé la notation suivante :

$$[u]_x \equiv u(x+) - u(x-) \equiv \lim_{\varepsilon \downarrow 0} u(x + \varepsilon) - u(x - \varepsilon).$$

On recherche ainsi des solutions fortes hors des singularités du potentiel et le recollement aux lieux des singularités x_j est dicté par des conditions de saut, obtenues naturellement en intégrant (B.18) au voisinage de x_j :

$$\lim_{\varepsilon \downarrow 0} \int_{x_j - \varepsilon}^{x_j + \varepsilon} \left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x) - k^2 \right) u = -\left[\frac{\mathrm{d}}{\mathrm{d}x} u\right]_{x_j} + c_j \, u(x_j) = 0.$$

B.2.1. *Potentiel uniquement composé de distributions de Dirac.* Étudions le cas où le potentiel V ne possède qu'une composante singulière ($V_{\text{reg}} \equiv 0$ dans la décomposition (V1)). Dans ce cas, l'étude de la section B.1.2 (avec un potentiel nul) nous indique qu'en dehors des singularités, il existe des constantes complexes A_j et B_j telles que

$$u(x) = A_j e^{ikx} + B_j e^{-ikx}, \quad \text{pour} \quad x \in (x_j, x_{j+1}).$$

Les solutions de (B.18) sont donc composées d'une superposition d'ondes planes, dont l'amplitude et la phase sont déterminées par les constantes A_j et B_j. Ces constantes sont reliées entre elles via les conditions de saut (B.25) et le problème est bien posé (existence et unicité de la solution) lorsque l'on impose deux de ces constantes.

L'onde plane distordue $e_{V_{\rm sing}+}$ associée au potentiel $V_{\rm sing}(x) = \sum_{j=0}^{N-1} c_j\, \delta(x-x_j)$ est ainsi donnée par

$$e_{V_{\rm sing}+}(x;k) = \begin{cases} e^{ikx} + B_0 e^{-ikx} \text{ for } x < x_0 \\ A_1 e^{ikx} + B_1 e^{-ikx} \text{ for } x_0 < x < x_1 \\ \vdots \\ A_N e^{ikx} \text{ for } x > x_{N-1} \end{cases}, \tag{B.26}$$

puisque l'on a fixé $A_0 \equiv 1$ et $B_N \equiv 0$ par convention. Les conditions de saut relient les constantes par les formules

$$\begin{aligned} e^{ikx_0} + B_0 e^{-ikx_0} &= A_1 e^{ikx_0} + B_1 e^{-ikx_0} \\ ik\left(A_1 e^{ikx_0} - B_1 e^{-ikx_0} - e^{ikx_0} + B_0 e^{-ikx_0}\right) &= c_0\left(A_1 e^{ikx_0} + B_1 e^{-ikx_0}\right) \\ &\vdots \\ A_{N-1} e^{ikx_{N-1}} + B_{N-1} e^{-ikx_{N-1}} &= A_N e^{ikx_{N-1}} \\ ik\left(A_N e^{ikx_{N-1}} - A_{N-1} e^{ikx_0} + B_{N-1} e^{-ikx_0}\right) &= c_{N-1}\left(A_N e^{ikx_{N-1}}\right). \end{aligned}$$

et l'on obtient immédiatement les coefficients de transmission $t(k) = A_N$ et de réflexion $r_1(k) = B_0$. On notera que le système ci-dessus est unitaire et que l'on a donc

$$|r_1(k)|^2 + |t(k)|^2 = |B_0|^2 + |A_N|^2 = 1.$$

Un exemple particulier vaut la peine d'être étudié : celui d'un nombre fini de distributions de Dirac de même hauteur, régulièrement espacées :

$$V = c\sum_{j=0}^{N-1} \delta(x-jl), \quad \text{avec } c \in \mathbb{R},\ N \in \mathbb{N}^*\ \text{ et }\ l > 0.$$

Dans ce cas, de simples calculs algébriques permettent d'obtenir une formule explicite pour le coefficient de transmission (voir Section C.4 page 51). On obtient

$$|t(k)| = \frac{1}{|P^N_{(1,1)}|} = \frac{1}{(1+(\frac{c}{k})^2 U_N(\xi)^2)^{1/2}}, \tag{B.27}$$

où U_N est le $(N-1)$-ième polynôme de Tchebychev de deuxième espèce (voir **[124]**) et $\xi \equiv \cos(kl) + \frac{c}{k}\sin(kl)$. On trouvera en Figure 8 (page suivante) les variations de ce coefficient, en fonction de k, pour un potentiel constitué de 3 et 20 distributions de Dirac.

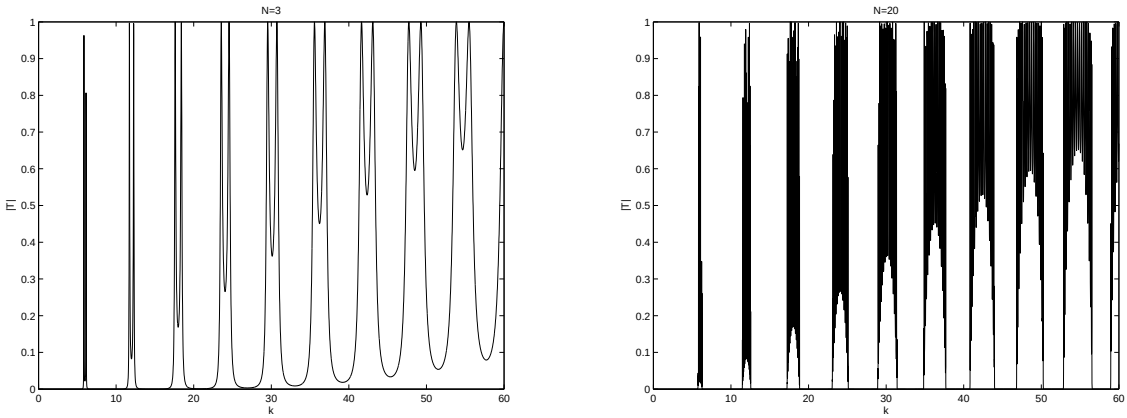
On voit très nettement apparaître une structure en bande : il existe des énergies pour lesquelles la transmission est quasi-parfaite ; l'onde entrante est pratiquement entièrement réfléchie ailleurs. Cette structure lacunaire est de plus en plus prononcée alors que N croît et annonce la structure caractéristique et recherchée du cas des potentiels périodiques (ou avec une microstructure). Chaque bande contient $N-1$ ondulations, que l'on voit apparaître dans la phase de Bloch $\gamma \equiv \arccos\xi$ (les ondulations correspondent à $\sin\gamma = 0$). Lorsque l'énergie augmente, la largeur des intervalles d'énergie où l'onde est majoritairement transmise devient plus large, la profondeur des lacunes plus faible ; l'onde est totalement transmise à haute énergie. On trouvera plus de détails dans **[146, 147, 156]**.

Ce dernier comportement à haute énergie se déduit aisément de notre formule : puisqu'il existe $C_0 > 0$ tel que pour tout $\xi \in \mathbb{R}$, $U_N(\xi) \le C_0$, il est immédiat de vérifier que l'on a

$$\Big|t(k)-1\Big|,\ \ \Big|\frac{\mathrm{d}}{\mathrm{d}k}t(k)\Big| \le \frac{C}{\langle k\rangle}, \quad \text{où } \langle k\rangle \equiv (1+k^2)^{1/2}.$$

De la même manière, le coefficient de réflexion vérifie

$$\Big|r_j(k)\Big|,\ \ \Big|\frac{\mathrm{d}}{\mathrm{d}k}r_j(k)\Big| \le \frac{C}{\langle k\rangle}.$$



(a) $N = 3$. (b) $N = 20$.

FIGURE 8. Coefficient de transmission, en fonction de k. Le potentiel est une somme de N distributions de Dirac, de hauteur $c = 80$, régulièrement espacées de $l = 0.5$.

Ces propriétés sont également vérifiées pour des potentiels réguliers (voir **[135]**) et s'avèrent fondamentales pour l'étude générale de l'opérateur de Schrödinger

$$H_V \equiv -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x),$$

et donc pour l'étude des solutions de l'équation (B.18), comme on le verra à la section suivante.

B.2.2. *Le cas général.* Dans cette section, nous étudions le cas plus général de la décomposition (V1). Notre but est de généraliser les résultats connus pour une classe de potentiels réguliers, au cas de potentiels admettant des singularités.

Pour cela, la notion d'*opérateurs d'ondes* est particulièrement indiquée. C'est en effet un outil précieux permettant de propager les propriétés de l'opérateur Hamiltonien « libre » $H_0 = -\frac{\mathrm{d}^2}{\mathrm{d}x^2}$, à l'opérateur à coefficient variable $H_V \equiv -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V$, sur son spectre continu.

DEFINITION B.21 (Opérateurs d'onde). *On appelle* opérateurs d'onde *associés aux hamiltoniens H_0 et H_V les opérateurs $W_\pm$ et $W_\pm^\star$ définis par*

$$W_\pm \equiv \operatorname*{s-lim}_{t\to\pm\infty} e^{itH_V} e^{-itH_0}, \tag{B.28}$$

$$W_\pm^\star \equiv \operatorname*{s-lim}_{t\to\pm\infty} e^{itH_0} e^{-itH_V} P_c, \tag{B.29}$$

où s-lim *correspond à la notion de limite forte et P_c est la projection sur le spectre continu de H_V.*

Ces opérateurs vérifient les propriétés suivantes :

$$W_\pm W_\pm^\star = P_c, \qquad W_\pm^\star W_\pm = \mathrm{Id}$$
$$f(H_V)P_c = W_\pm f(H_0) W_\pm^\star, \qquad f(H_0) = W_\pm^\star f(H_V) W_\pm,$$

pour toute fonction f Borélienne sur $\mathbb{R}$. Ainsi, un opérateur $f(H_V)P_c$ sera borné de $W^{k_1,p_1}(\mathbb{R}^d)$ sur $W^{k_2,p_2}(\mathbb{R}^d)$ si l'opérateur libre $f(H_0)$ l'est et si les opérateurs d'onde $W_\pm$ et $W_\pm^\star$ sont bornés, respectivement sur $W^{k_1,p_1}(\mathbb{R}^d)$ et $W^{k_2,p_2}(\mathbb{R}^d)$.

Obtenir des bornes sur les opérateurs d'onde est donc crucial. Ces bornes ont été obtenues par Yajima **[181]**, avec des hypothèses de régularité et de décroissance à l'infini

sur le potentiel $V(x)$ et pour des dimensions $d \geq 2$. Weder **[178]** a étendu ces résultats au cas de la dimension $d = 1$. Une analyse fine de sa preuve montre que le résultat peut être obtenu uniquement *via* les propriétés des ondes planes distordues ($e_{V\pm}$, voir Définition B.19) et du comportement des coefficients de réflexion et de transmission. Plus précisément, on prouve au Chapitre 5 le résultat suivant (Théorème 5.1, page 174) :

PROPOSITION B.22. *Soit $H_V = -\frac{d^2}{dx^2} + V(x)$ un opérateur auto-adjoint sur $L^2(\mathbb{R})$ tel que les propriétés suivantes soient vérifiées :*

i. *Les coefficients de transmission et de réflexion sont majorés par*

$$|r_1(k)| + |r_2(k)| + |t(k) - 1| \leq \frac{C}{\langle k \rangle}, \tag{B.30}$$

$$|\partial_k r_1(k)| + |\partial_k r_2(k)| + |\partial_k t(k)| = \mathcal{O}\left(\frac{1}{|k|}\right), \quad |k| \to \infty. \tag{B.31}$$

ii. *Les opérateurs $S_\pm$, définis par*

$$\left(S_\pm \Phi\right)(x) \equiv \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{1}{t(k)} e_\pm(x,k) - e^{ikx}\right) e^{-iky} \, dk \; \Phi(y) \, dy,$$

sont bornés dans $W^{1,p}(\mathbb{R}_+)$ et $W^{1,p}(\mathbb{R}_-)$ (respectivement pour S_+ et S_-), avec $1 < p < \infty$.

Alors $W_\pm$ et $W_\pm^\star$, définis à l'origine sur $W^{1,p} \cap L^2$, $1 \leq p \leq \infty$, s'étendent à des opérateurs $W^{1,p}$, $1 < p < \infty$. De plus, ces opérateurs sont bornés par C_p une constante :

$$\|W_\pm f\|_{W^{1,p}} \leq C_p \|f\|_{W^{1,p}}, \quad \|W_\pm^\star f\|_{W^{1,p}} \leq C_p \|f\|_{W^{1,p}},$$

pour tout $f \in W^{1,p} \cap L^2$, $1 < p < \infty$.

Idée de la preuve. La preuve repose sur une décomposition de l'opérateur en une composante à hautes fréquences et une composante à basses fréquences. Chacune de ces composantes pourra être bornée, grâce aux hypothèses du théorème.

En particulier, on utilise pour le terme à hautes fréquences l'inégalité de Young (voir par exemple **[138]**) sur L^p, $1 \leq p \leq \infty$:

$$C \equiv \sup_{x \in \mathbb{R}} \int_{\mathbb{R}} |S_\pm(x,y)| \, \mathrm{d}y + \sup_{y \in \mathbb{R}} \int_{\mathbb{R}} |S_\pm(x,y)| \, \mathrm{d}x < \infty$$
$$\implies \|S_\pm \phi\|_{L^p} \leq C \, \|\phi\|_{L^p}.$$

Le terme à basses fréquences fait naturellement apparaître la transformée de Hilbert

$$(\mathcal{H}\phi)(x) = \frac{1}{\pi} \, \text{P.V.} \int \frac{\phi(x-y)}{y} \mathrm{d}y \equiv \int_{-\infty}^{\infty} e^{ikx} \, (-i \, \mathrm{sgn}(k)) \, \hat{\phi}(k) \mathrm{d}k \tag{B.32}$$

Le caractère borné de la transformée de Hilbert est connu (voir **[174]**) ; celui du terme à basse fréquence est obtenu grâce aux estimations (B.30) et (B.31). □

Il suffit donc de vérifier les propriétés des ondes distordues de l'équation de Helmholtz (B.18), pour aboutir au caractère borné des opérateurs d'onde. Ceci est obtenu pour des potentiels vérifiant la décomposition (V1), principalement en utilisant la représentation intégrale qui découle de l'équation de Lippmann-Schwinger

$$e_{V+}(x;k) = t(k)e^{ikx} + \frac{1}{2ik} \int_0^\infty (e^{ik(y-x)} - e^{ik(x-y)}) V_{\mathrm{reg}}(y) e_{V+}(y;k) \, \mathrm{d}y$$
$$+ \frac{1}{2ik} \sum_{l=0}^{N-1} c_l (e^{ik(x_l - x)} - e^{ik(x-x_l)}) e_{V+}(x_l,k) \mathbf{1}(x_l - x),$$

où **1** est la fonction de Heaviside définie par

$$\mathbf{1}(x) = \begin{cases} 1\ , \text{ si } x > 0, \\ \frac{1}{2}\ , \text{ si } x = 0, \\ 0\ , \text{ si } x < 0, \end{cases}$$

On vérifie aisément que la fonction, définie ainsi, sera bien solution de (B.18) avec les conditions à l'infini (B.22) caractérisant $e_{V+}(x;k)$. On peut alors déduire d'une analyse des itérées de l'intégrale de Volterra les hypothèses de la proposition B.22. On obtient alors le théorème principal :

PROPOSITION B.23. *Soit $V(x)$, un potentiel satisfaisant* (V1) *(voir page 32). Alors $W_\pm$ et $W_\pm^\star$, définis par* (B.28) *et* (B.29) *sur $W^{1,p} \cap L^2$, $1 \le p \le \infty$, s'étendent en des opérateurs bornés sur $W^{1,p}$, $1 < p < \infty$:*

$$\|W_\pm f\|_{W^{1,p}(\mathbb{R})} \le C_p \|f\|_{W^{1,p}(\mathbb{R})},\ \|W_\pm^\star f\|_{W^{1,p}(\mathbb{R})} \le C_p \|f\|_{W^{1,p}(\mathbb{R})},\ f \in W^{1,p}(\mathbb{R}),\ 1 < p < \infty.$$

Les applications de ce résultat sont nombreuses. Il a par exemple été implicitement utilisé dans **[152]** et **[145]**. On peut en particulier étendre de nombreux résultats théoriques au cas de potentiels singuliers. Par exemple, on étend les estimations dispersives et de Strichartz H^1 au cas de potentiels singuliers et on peut en déduire le caractère bien posé dans H^1 de l'équation de Schrödinger non linéaire

$$\text{(B.33)} \qquad \begin{cases} i\partial_t u + H_V u - |u|^{2\sigma} u \ = \ 0, \\ \quad u(x,0) = u_0(x) \in H^1, \end{cases}$$

où $H_V \ \equiv \ -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x)$ admet un nombre fini de singularités.

Notre but n'est pas ici de détailler ces résultats (on pourra consulter la section 8 du Chapitre 5, page 182). Notre motivation principale est d'étudier la diffusion de potentiels admettant des singularités *et possédant une microstructure*. C'est ce que nous développons dans la section suivante et au chapitre 6 ; les résultats obtenus en Section B.3.2 requièrent des bornes sur $(m^2 + H_V)^{-1} P_c(H_V)(m^2 - \partial_x^2)$, qui découlent de la proposition B.23.

B.3. Potentiels possédant une microstructure. Dans cette partie, nous ajoutons un terme de microstructure au potentiel étudié. On s'intéresse donc aux potentiels admettant la décomposition suivante :

$$\text{(V2)} \qquad \begin{aligned} V^\epsilon(x) \ &\equiv \ V_0(x) \ + \ q_\epsilon(x) \\ &\equiv \ V_{\text{sing}}(x) \ + \ V_{\text{reg}}(x) + \ q_\epsilon(x), \end{aligned}$$

où les composantes vérifient :

i. *Composante singulière, V_{sing} :*

$$V_{\text{sing}}(x) \ = \ \sum_{j=0}^{N-1} c_j\ \delta(x - x_j), \quad \text{où} \quad c_j, x_j \in \mathbb{R}, \ \ x_j < x_{j+1}.$$

ii. *Composante régulière, V_{reg} :* $V_{\text{reg}} \in L^{1,2}(\mathbb{R})$ avec

$$\|V\|_{L^{1,2}} \ \equiv \ \int_{\mathbb{R}} (1 + |s|)^2\ |V(s)|\ \mathrm{d}s \ < \ \infty.$$

iii. *Composante oscillante, $q_\epsilon(x) = q\left(x, \frac{x}{\epsilon}\right)$* où $(x,y) \mapsto q(x,y)$ vérifie
   (a) q est $1-$ périodique, *i.e.* pour tout $x \in \mathbb{R}$, $\ q(x, y+1) = q(x,y)$,
   (b) q est de moyenne nulle par rapport à y, *i.e.* pour tout $x \in \mathbb{R}$,

$$\int_0^1 q(x,y)\ \mathrm{d}y = 0,$$

(c) Il existe une subdivision de $\mathbb{R}$

$$-\infty = a_0 < a_1 < a_2 < \cdots < a_M < a_{M+1} = +\infty$$

telle que l'on ait

$$\sum_{j=0}^{M+1} \int_0^1 \|q(\cdot, y)\|^2_{C^3(a_j, a_{j+1})} \, \mathrm{d}y \ < \ \infty.$$

iv. On travaillera avec le développement de Fourier de $q(x, y)$, qui s'écrit

$$q(x,y) \ = \ \sum_{j \neq 0} \ q_j(x) \ e^{2\pi i j y}, \quad q_j(x) \equiv \int_0^1 e^{-2\pi i j y} q(x,y) \ \mathrm{d}y \tag{B.34}$$

et on suppose les caractères bornés et décroissants à l'infini :

$$\int_{\mathbb{R}} \int_0^1 |q(x,y)|^2 \ \mathrm{d}y \ \mathrm{d}x \ = \ \sum_{|j| \geq 1} \int_{\mathbb{R}} |q_j(x)|^2 < \infty,$$

$$\int_0^1 |q(x,y)|^2 \ \mathrm{d}y \ = \ \sum_{|j| \geq 1} |q_j(x)|^2 \ \to 0, \quad |x| \to \infty \ .$$

On voit que l'on autorise en plus d'une composante singulière, des discontinuités dans la composante rapidement oscillante qui traduisent des défauts dans la microstructure. La question que l'on se pose est alors la suivante :

Comment les défauts du milieu influencent les propriétés de diffusion, tels les coefficients de réflexion et transmission ?

Plus précisément, on va voir que l'on peut obtenir un développement asymptotique du coefficient de transmission $t^\epsilon(k)$ en fonction de ϵ. Nous allons comparer les termes du développement asymptotique en présence de défauts, avec ceux dans le cas lisse. Le résultat sera obtenu de deux manières différentes, qui ont chacune leur intérêt propre. La première, dans la section suivante, utilise les techniques de l'homogénéisation. Cette méthode est constructive et permet de comprendre l'origine des termes générés par la présence des défauts. La deuxième méthode est abordée en Section B.3.2 et permet de justifier rigoureusement les résultats obtenus par la méthode d'homogénéisation. De plus, cette dernière méthode fonctionne dans un cadre plus général, dans le sens où la composante rapidement oscillante du potentiel peut ne pas être périodique, mais localisée en hautes fréquences, dans un sens que l'on précisera alors. Deux exemples précis seront finalement étudiés à la section B.3.3.

B.3.1. *Théorie de l'homogénéisation.* La méthode d'homogénéisation est particulièrement indiquée pour l'étude asymptotique de matériaux périodiques comportant plusieurs échelles (voir **[129, 126]** par exemple). Dans ce cadre, la méthode consiste à chercher une solution de l'équation

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V_0(x) + q\left(x, \frac{x}{\epsilon}\right) - k^2 \right) e_{V^\epsilon +}(x;k) = 0,$$

sous la forme d'une fonction à deux échelles, $e_{V^\epsilon +}(x;k) = U^\epsilon(x, \frac{x}{\epsilon})$, où x et y sont des variables *indépendantes*. Naturellement, dériver par rapport à la variable rapide fait apparaître un coefficient $1/\epsilon$ et l'on cherche donc $U^\epsilon(x,y)$ comme solution de

$$\left(-\left(\frac{\partial}{\partial x} + \frac{1}{\epsilon} \frac{\partial}{\partial y} \right)^2 + V_0(x) + q(x,y) - k^2 \right) U^\epsilon(x,y) \ = \ 0. \tag{B.35}$$

On postule de plus un développement asymptotique de $U^\epsilon(x,y)$ en puissance de ϵ :

$$U^\epsilon(x,y) \;=\; \sum_{j=0}^{\infty} \epsilon^j \; U_j(x,y), \tag{B.36}$$

vérifiant les conditions de périodicité, de radiation et de saut suivantes (on rappelle que les conditions de saut sont déterminées par la présence – ou non – de singularités, *via* (B.25)) :

$$\begin{aligned} & U_j(x,y+1) \;=\; U_j(x,y), \quad j \geq 0, \\ & U_0(x,y) - e^{ikx}, \;\; U_j(x,y) \;\; j \geq 1 \quad \text{sortante pour} \;\; |x| \to \infty, \\ & U_j(x,y)|_{y=x/\epsilon} \;\; \text{vérifie les conditions de saut} \;\; (\text{B.25}). \end{aligned} \tag{B.37}$$

Les termes $U_j(x,y)$ sont alors obtenus en introduisant l'*Ansatz* (B.36) dans l'équation (B.35) et en résolvant à chaque ordre. On obtient ainsi la hiérarchie d'équations suivante :

$$\begin{aligned} \mathcal{O}(\epsilon^{-2}) \qquad & r_{-2} = -\partial_y^2 U_0 \;=\; 0, \\ \mathcal{O}(\epsilon^{-1}) \qquad & r_{-1} = -\partial_y^2 U_1 - 2\partial_y\partial_x U_0 \;=\; 0, \\ \mathcal{O}(\epsilon^j),\; j \geq 0 \qquad & r_j = -\partial_y^2 U_{j+2} \;-\; 2\partial_y\partial_x U_{j+1} \;-\; \partial_x^2 U_j + (V_0+q)U_j - k^2 U_j \;=\; 0. \end{aligned}$$

En résolvant simultanément les équations $r_j = 0$ pour $j = -2, \cdots, 3$, on détermine les fonctions U_0, U_1, U_2 et U_3. On obtient ainsi formellement une approximation de U^ϵ à l'ordre $\mathcal{O}(\epsilon^3)$. Les calculs apparaîssent au Chapitre 6 (section 4 page 197) et nous récapitulons ici le résultat obtenu.

DÉVELOPPEMENT « NAÏF ». La solution $e_{V^\epsilon +}(x;k) = U^\epsilon(x, \frac{x}{\epsilon})$ vérifie le développement suivant :

$$U^\epsilon(x,y) \;=\; U_0(x;k) \;+\; \epsilon^2 \left(U_2^{(p)}(x,y) + U_2^{(h)}(x) \right) \;+\; \epsilon^3 \left(U_3^{(p)}(x,y) + U_3^{(h)}(x) \right) \;+\; \ldots, \tag{B.39}$$

avec les termes définis par :

- $U_0(x) \;\equiv\; e_{V_0+}(x;k)$ la solution homogénéisée, vérifiant

$$-\frac{\mathrm{d}^2}{\mathrm{d}x^2} U_0(x) + V_0(x) U_0(x) - k^2 U_0(x) \;=\; 0$$

  et telle que $U_0 - e^{ikx}$ vérifie la condition de radiation sortante.
- $U_2^{(p)}(x,y)$ et $U_3^{(p)}(x,y)$ sont obtenues grâce au développement en série de Fourier de $y \mapsto q(x,y)$ donné en (B.34). On a précisément les formules suivantes

$$\begin{aligned} U_2^{(p)}(x,y) \;&=\; -\frac{e_{V_0+}(x;k)}{4\pi^2} \sum_{|j|\geq 1} \frac{q_j(x)}{j^2} e^{2i\pi jy}, \\ U_3^{(p)}(x,y) \;&=\; -\frac{i}{4\pi^3} \sum_{|j|\geq 1} \frac{\partial_x (e_{V_0+}(x;k) q_j(x))}{j^3} e^{2i\pi jy}. \end{aligned}$$

- $U_2^{(h)}(x)$ et $U_3^{(h)}(x)$ sont les solutions sortantes des équations

$$\begin{aligned} -\frac{\mathrm{d}^2}{\mathrm{d}x^2} U_2^{(h)}(x) + V_0(x) U_2^{(h)}(x) - k^2 U_2^{(h)}(x) &= -\int_0^1 q(x,y) U_2^{(p)}(x,y) \; \mathrm{d}y, \\ -\frac{\mathrm{d}^2}{\mathrm{d}x^2} U_3^{(h)}(x) + V_0(x) U_3^{(h)}(x) - k^2 U_3^{(h)}(x) &= -\int_0^1 U_3^{(p)}(x,y) q(x,y) \; \mathrm{d}y. \end{aligned}$$

Nous avons utilisé le terme de développement « naïf » , car aucune précaution particulière n'a été prise concernant la présence de singularités et de discontinuités dans le terme rapidement oscillant. De plus, le cahier des charges n'est pas totalement rempli, puisque les conditions de saut dictées par l'équation n'ont pas été vérifiées.

En effet, il est aisé de se rendre compte que si les termes homogènes (U_0, $U_2^{(h)}$, $U_3^{(h)}$) vérifient naturellement les conditions de saut grâce aux équations qu'ils vérifient, les termes $U_2^{(p)}$ et $U_3^{(p)}$ n'ont aucune raison de vérifier ces conditions *et ne les vérifient pas en général*. On vérifie en fait facilement que *le développement* (B.39) *est valable pour le cas de potentiels réguliers avec microstructure lisse*. Dans le cas contraire, il est nécessaire d'ajouter des correcteurs, qui sont choisis de la forme suivante :

Definition B.24 (Correcteur d'interface). *La fonction $\mathcal{U}(x)$ est un correcteur à l'interface $x = a$, si on peut l'écrire sous la forme*

$$\mathcal{U}(x) \;=\; \begin{cases} \alpha\; e_{V_0-}(x;k) & \text{si } \; x < a, \\ \beta\; e_{V_0+}(x;k) & \text{si } \; x > a, \end{cases} \tag{B.40}$$

avec α et β des constantes à déterminer.

Ces correcteurs sont C^2 par morceaux et vérifient

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V_0(x) - k^2\right)\mathcal{U} \;=\; 0\;, \qquad \text{pour} \quad x < a \text{ et } x > a,$$

ainsi que la condition de radiation sortante. Les coefficients α et β peuvent alors être manipulés afin d'obtenir les conditions de saut désirées à l'interface $x = a$.

Le principe est donc le suivant : à chaque défaut dans la vérification des conditions de saut (il y en a un nombre fini par construction), on ajoute un correcteur d'interface, avec des paramètres bien choisis, de telle sorte que les conditions de saut soient rétablies.

Les correcteurs d'interface sont ajustés dans la section 4.1 du Chapitre 6 (page 197), et l'on présente ci-dessous le développement final obtenu.

Développement corrigé. Après avoir ajouté les correcteurs d'interface adaptés, on aboutit au développement suivant :

$$U^\epsilon(x,y) \;=\; U_0(x;k) \;+\; \epsilon\,\mathcal{U}_1^\epsilon(x) \;+\; \epsilon^2\left(\mathcal{U}_2^\epsilon(x) \;+\; U_2^{(p)}(x,y) + U_2^{(h)}(x)\right) \;+\; \ldots, \tag{B.41}$$

où $U_0(x)$, $U_2^{(p)}(x,y)$, $U_2^{(h)}(x)$ sont inchangés par rapport au développement « naïf » et

- $\mathcal{U}_1^\epsilon$ est déterminée par les violations des conditions de saut de $U_2^{(p)}(x,y)$, qui interviennent aux points de discontinuité de $q(x,y)$.
- $\mathcal{U}_2^\epsilon$ est déterminée par les violations des conditions de saut de $U_3^{(p)}(x,y)$, qui interviennent
  - (a) aux points a_j $(j = 1,\ldots,M)$ pour lesquels $[q_j]_a \neq 0$ ou $[\partial_x q_j]_a \neq 0$;
  - (b) au support des singularités : $\operatorname{supp} V_{\text{sing}} = \{x_0,\ldots,x_{N-1}\}$.

On voit ainsi que les termes correcteurs apparaissent à un ordre précédant les termes qu'ils corrigent. Ceci s'explique par le fait que les coefficients α et β des termes correcteurs font intervenir les dérivées par rapport à la variable rapide. Pousser le développement un cran plus loin requiert donc de calculer $U_4 \;=\; U_4^{(p)}(x,y) \;+\; U_4^{(h)}(x)$ et les termes $U_3^{(p)}(x,y)$, $U_3^{(h)}(x)$ ne nous sont utiles que pour déterminer $\mathcal{U}_2^\epsilon(x)$.

On a ainsi deux conséquences immédiates :

i. les effets de la présence de discontinuités dans la microstructure interviennent à un ordre précédant les effets de la présence de singularités ;
ii. la différence entre l'onde distordue $e_{V+}(x;k)$ et l'onde homogénéisée $e_{V_0+}(x;k)$ est due au premier ordre à la présence de discontinuités dans la microstructure.

APPLICATION AU COEFFICIENT DE TRANSMISSION, $t^\epsilon(k)$. Connaissant le développement de $e_{V^\epsilon+}(x,y) = U^\epsilon(x, x/\epsilon)$, il est facile de déduire celui du coefficient de transmission $t^\epsilon(k)$. Ces calculs sont présentés dans la section 4.2 du Chapitre 6 (page 204). On obtient

$$t^\epsilon(k) \;=\; t_0^{hom}(k) \;+\; \epsilon\, t_1^\epsilon(k) \;+\; \epsilon^2 \,\Big(\, t_2^{hom}(k) \;+\; t_2^\epsilon(k) \,\Big) \;+\; \ldots, \tag{B.42}$$

où $t_0^{hom}(k)$ est le coefficient de transmission associé au potentiel V_0 et

$$t_1^\epsilon(k) = \frac{1}{4k\pi} \sum_{j=1}^{M} e_{V_0+}(a_j;k) e_{V_0-}(a_j;k) \sum_{|l|\geq 1} [q_l]_{a_j} \, \frac{e^{2i\pi l \frac{a_j}{\epsilon}}}{l},$$

$$t_2^{hom}(k) = \frac{i}{8k\pi^2} \sum_{|j|\geq 1} j^{-2} \int_{\mathbb{R}} |q_j(z)|^2 \; e_{V_0-}(z;k) e_{V_0+}(z;k) \mathrm{d}z,$$

$$t_2^\epsilon(k) = \; \frac{i}{8k\pi^2} \sum_{j=1}^{M} \sum_{|l|\geq 1} [\; \partial_x \left(e_{V_0+}(x;k) \; e_{V_0-}(x;k) q_l(x) \right) \;]_{a_j} \, \frac{e^{2i\pi l \frac{a_j}{\epsilon}}}{l^2}.$$

Les contributions de chacun des termes de (B.41) sur ce développement sont les suivantes :

(a) t_j^{hom}, $(j = 0, 2)$ sont les contributions des termes homogènes du développement « naïf » $U_0(x)$ et $U_2^{(h)}(x)$ et sont valides pour les potentiels réguliers ;

(b) t_1^ϵ provient de $\mathcal{U}_1^\epsilon$ et est dû aux discontinuités de $x \mapsto q(x,\cdot)$;

(c) t_2^ϵ provient de $\mathcal{U}_2^\epsilon$ et est dû à la fois aux discontinuités de $x \mapsto q(x,\cdot)$ ou $x \mapsto \partial_x q(x,\cdot)$ et à la présence de singularités dans le potentiel V_{sing}.

Les coefficients t_1^ϵ and t_2^ϵ sont uniformément bornés pour ϵ petit, mais font apparaître des oscillations rapides, de telle manière qu'on a en général[16] le comportement suivant :

- Si q_ϵ est discontinue, alors

  $t^\epsilon - t_0^{hom} = \mathcal{O}(\epsilon), \quad \epsilon \to 0, \quad$ mais $\quad \epsilon^{-1} \Big(\, t^\epsilon - t_0^{hom} \, \Big) \quad$ ne possède pas de limite.

- Si q_ϵ est lisse, mais V_0 possède une composante singulière (distributions de Dirac), alors

  $t^\epsilon - t_0^{hom} = \mathcal{O}(\epsilon^2), \quad \epsilon \to 0, \quad$ mais $\quad \epsilon^{-2} \Big(\, t^\epsilon - t_0^{hom} \, \Big) \quad$ ne possède pas de limite.

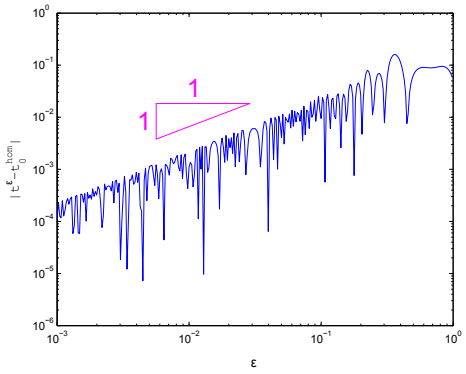
- Si q_ϵ est lisse et V_0 ne possède pas de composante singulière, alors

  $t^\epsilon - t_0^{hom} = \mathcal{O}(\epsilon^2), \quad \epsilon \to 0, \quad$ et $\quad \lim_{\epsilon \downarrow 0} \epsilon^{-2} \Big(\, t^\epsilon - t_0^{hom} \, \Big) \quad$ est bien définie.

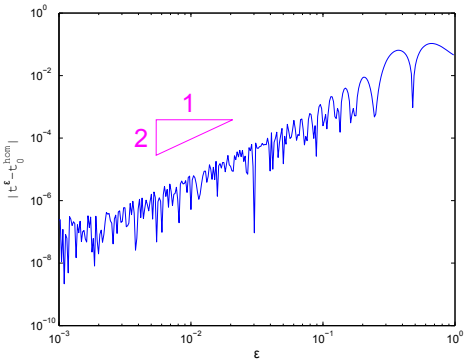
Les trois graphes de la figure 9, page suivante, illustrent ce comportement. Le détail des potentiels utilisés pour obtenir ces graphes, ainsi que la méthode numérique utilisée, sont indiqués à la section A du Chapitre 6, page 218.

REMARQUE B.25. *Il est important de noter que les développements présentés dans cette section ne sont pas obtenus de manière rigoureuse. Ils relèvent d'une construction naturelle, mais faisant intervenir des dérivées $e_{V_0+}(x;k)$ et $q_j(x)$ à un ordre arbitraire, malgré les potentielles discontinuités de ces fonctions. Les développements sont donc à comprendre au sens des distributions. Ces développements sont prouvés rigoureusement par une méthode totalement différente, dans la section suivante.*

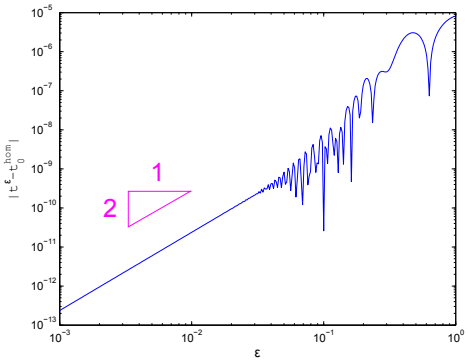
[16] Il existe des distributions particulières pour lesquelles les contributions des termes correcteurs s'annulent. Ce point est abordé dans la section B.3.3, page 44.



(a) q est discontinue et V_0 est une somme de distributions de Dirac



(b) q est lisse et V_0 est une somme de distributions de Dirac



(c) $V^\epsilon = V_0$ est une *approximation lisse* d'une somme de distributions de Dirac

FIGURE 9. Comportement de $\log \left|t^\epsilon - t_0^{hom}\right|$ versus $\log \epsilon^{-1}$ pour plusieurs formes de potentiels.

B.3.2. *Démonstration rigoureuse.* Le but de cette section est d'offrir une démonstration rigoureuse aux développements présentés dans la section précédente. Ces démonstrations utilisent une classe de potentiels légèrement différente de celle présentée en (V2) : on demande plus de décroissance à l'infini et la microstructure n'est plus nécessairement représentée par un potentiel périodique, mais simplement par une fonction localisée dans les hautes fréquences. De manière détaillée, on suppose

$$\text{(V3)} \qquad \begin{aligned} V(x) &\equiv V_0(x) + Q(x) \\ &\equiv V_{\text{sing}}(x) + V_{\text{reg}}(x) + Q(x), \end{aligned}$$

vérifiant les propriétés suivantes :

i. *Composante singulière, V_{sing} :*

$$V_{\text{sing}}(x) = \sum_{j=0}^{N-1} c_j\ \delta(x - x_j), \quad \text{où} \quad c_j, x_j \in \mathbb{R}, \quad x_j < x_{j+1}.$$

ii. *Composante régulière, V_{reg} :* $V_{\text{reg}} \in L^{1,8}(\mathbb{R})$ avec

$$\|V\|_{L^{1,8}} \equiv \int_{\mathbb{R}} (1 + |s|)^8\ |V(s)|\ \mathrm{d}s < \infty.$$

iii. *Composante à haute fréquence, $Q(x)$:* Q est bornée pour la norme

$$|||Q||| \equiv \left\| \langle D_0\rangle^{-1} \ \chi^{-1} Q \ \chi^{-1} \langle D_0\rangle^{-1} \right\|_{L^2\to L^2}, \tag{B.43}$$

où l'on définit $\langle D_0\rangle^s$ grâce à la transformée de Fourier par

$$\langle D_0\rangle^s g \equiv (I-\Delta)^{s/2} g \equiv \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{ix\xi}\ (1+\xi^2)^{s/2}\ \hat{g}(\xi)\ \mathrm{d}\xi. \tag{B.44}$$

et χ la fonction poids

$$\chi(x) = \langle x\rangle^{-\sigma} \ = \ \left(1+x^2\right)^{-\frac{\sigma}{2}}, \ \text{ avec } \sigma > 4. \tag{B.45}$$

Ainsi, la composante Q peut être grande en norme L^∞, mais elle est suffisamment décroissante à l'infini, et supportée à hautes fréquences : on s'intéresse au comportement du système pour $|||Q|||$ petit. Le cas $Q(x) = q_\epsilon(x) = q(x, x/\epsilon)$, $\epsilon \to 0$ est un cas particulier, comme on le verra par la suite.

La méthode est alors la suivante. L'onde distordue $e_{V+}(x;k)$, définie par (B.18) et par (B.22) est une perturbation de l'onde homogénéisée $e_{V_0+}(x;k)$ et l'on définit u_s par

$$e_{V+}(x;k) \ = \ e_{V_0+}(x;k) \ + \ u_s(x;k).$$

La fonction u_s vérifie la condition de radiation sortante lorsque $x \to \pm\infty$ et est solution de

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + (V_0+Q)(x) - k^2\right) u_s(x;k) \ = \ -Q e_{V_0+}(x;k), \tag{B.46}$$

On applique alors la résolvante $R_{V_0}(k)$, obtenue et étudiée en Section B.2, à cette équation. En réarrangeant les termes, on obtient alors l'équation de Lippmann-Schwinger

$$u_s \ = \ -\left(I + R_{V_0}(k)Q\right)^{-1} \ R_{V_0}(k)\ Q\ e_{V_0}(\cdot;k).$$

On a ainsi *formellement* le développement de Neumann suivant :

$$\begin{aligned} e_{V+}(x;k) \ &= \ e_{V_0+}(x;k) - R_{V_0}(k)(Qe_{V_0+}(x;k)) + R_{V_0}(k)QR_{V_0}(k)(Qe_{V_0+}(x;k)) + \dots \\ &= \ e_{V_0+}(x;k) \ + \ \sum_{m=0}^{\infty} \left[(-R_{V_0}(k)\ Q)^m \ e_{V_0+}\right](x;k). \end{aligned}$$

On définit ensuite l'onde plane pondérée en espace et en fréquence $E_{V+}(x;k)$ par

$$E_{V+}(x;k) \ \equiv \ (\ \langle D_0\rangle \chi e_{V+}\ )(x;k),$$

où $\langle D_0\rangle$ et χ sont définis par (B.44) et (B.45). Définissons alors les opérateurs

$$T_{R_{V_0}}(k) \ \equiv \ \langle D_0\rangle \chi R_{V_0}(k) \chi \langle D_0\rangle, \qquad \text{et} \qquad T_Q \ \equiv \ \langle D_0\rangle^{-1} \chi^{-1} Q \chi^{-1} \langle D_0\rangle^{-1},$$

et $E_{V+}(x;k)$ vérifie

$$\left(I + T_{R_{V_0}}\ T_Q\right)\ (\ E_{V+}(\cdot;k) - E_{V_0+}(\cdot;k)\ ) \ = \ -\langle D_0\rangle \chi R_{V_0}(k)\ Q e_{V_0+}(x;k). \tag{B.47}$$

Ainsi, si $\|T_Q\|_{L^2\to L^2} < \|T_{R_{V_0}}\|_{L^2\to L^2}$, le développement de Neumann suivant est vérifié :

$$\begin{aligned} E_{V+}(\cdot;k) \ &= \ E_{V_0+}(\cdot;k) \ - \ \left(I + T_{R_{V_0}}\ T_Q\right)^{-1} \ \langle D_0\rangle \chi R_{V_0}(k)\ Q e_{V_0+}(x;k) \\ &= \ E_{V_0+}(x;k) \ - \ \langle D_0\rangle \chi \ R_{V_0}(k)\ Q\ e_{V_0+}(x;k) \\ &\quad + \ T_{R_{V_0}}(k)\ T_Q \ \langle D_0\rangle \chi \ R_{V_0}(k)\ Q\ e_{V_0+}(x;k) - \dots . \end{aligned} \tag{B.48}$$

On a alors pour tout $M \geq 1$:

$$\left\| \langle D_0\rangle \chi \ \left(u_s(\cdot;k) \ + \ \sum_{m=0}^{M} \left[(-R_{V_0}(k)Q)^m\, e_{V_0+}\right](\cdot;k) \right) \right\|_{L^2(\mathbb{R})} \leq \ C|||Q|||^{M+1} \tag{B.49}$$

avec $u_s(x;k) \ \equiv \ e_{V_0+Q,+}(x;k) \ - \ e_{V_0+}(x;k)$.

Incidemment, la norme $\|T_Q\|_{L^2\to L^2}$ est exactement la norme $|||Q|||$; le caractère borné de T_Q est donc une hypothèse. Le caractère borné de $T_{R_{V_0}}$ est le résultat principal et sa preuve est technique. On y utilise de manière centrale le résultat obtenu en Section B.2 sur l'opérateur des ondes. Ce résultat nous permet d'utiliser les ondes planes distordues $e_{V_0+}(x;k)$ comme base d'un opérateur de Fourier généralisé et d'étudier de manière fine le comportement à basses et hautes fréquence de l'opérateur $T_{R_{V_0}}$. La preuve complète est donnée au Chapitre 6, section C, page 219.

À la fin de la journée, on obtient le résultat suivant :

PROPOSITION B.26. *Pour $k \neq 0$, il existe $\tau_0 = \tau_0(k)$, tel que si $|||Q||| < \tau_0(k)$, alors l'onde plane distordue $e_{(V_0+Q)+}(x;k)$ vérifie pour $M \geq 1$:*

$$\left\| \langle D_0\rangle \, \chi \, \left(u_s(\cdot;k) \; + \; \sum_{m=0}^{M} \left[(-R_{V_0}(k)Q)^m \, e_{V_0+} \right](\cdot;k) \right) \right\|_{L^2(\mathbb{R})} \leq \; C|||Q|||^{M+1}$$

avec $u_s(x;k) \; \equiv \; e_{(V_0+Q)+}(x;k) \; - \; e_{V_0+}(x;k)$.

Cette proposition s'applique notamment à une large classe de potentiels de la forme (V2) (*i.e.* avec une microstructure périodique, voir page 36), grâce au résultat suivant :

PROPOSITION B.27. *S'il existe $\epsilon > 0$ tel que $\left\|T_{q_\epsilon}\right\|_{L^2\to L^2} \; < \; \infty$, alors on a pour ϵ petit,*

$$\left\|T_{q_\epsilon}\right\|_{L^2\to L^2} = \mathcal{O}(\epsilon).$$

APPLICATION AU COEFFICIENT DE TRANSMISSION $t(k) \; = \; t[k;Q]$.

En utilisant le développement de la proposition B.26 et en utilisant la relation entre le coefficient de transmission et l'onde plane distordue donnée par (B.22), on aboutit au développement suivant :

$$t[Q] \; = \; t_0^{hom} \; + \; t_1[Q] \; + \; t_2[Q,Q] \; + \; t_3[Q,Q,Q] \; + \; \ldots, \tag{B.50}$$

où $t_j[Q,Q,\ldots,Q]$ est $j-$ linéaire en Q et provient de la contribution du $j-$ème terme $(-R_{V_0}(k)\; Q)^j \; e_{V_0+}(x;k)$. Précisons les premiers termes de ce développement :

$$t_1[Q] \; \equiv \; \frac{1}{2ik} \int_{-\infty}^{\infty} Q(\zeta) \; e_{V_0+}(\zeta;k) e_{V_0-}(\zeta;k) \; \mathrm{d}\zeta.$$

$$\begin{aligned} t_2[Q,Q] \; &\equiv \; \frac{1}{2ik} \int_{-\infty}^{\infty} Q(\zeta) R_{V_0}(k) (Q(\zeta) \; e_{V_0+}(\zeta;k)) e_{V_0-}(\zeta;k) \; \mathrm{d}\zeta \\ &= \; \frac{1}{2ik} \frac{1}{-2ik \; t_0^{hom}} \int_{-\infty}^{+\infty} Q(\zeta) e_{V_0-}(\zeta;k) \, (\; I_l(\zeta) + I_r(\zeta) \;) \; \mathrm{d}\zeta, \quad \text{avec} \\ &\qquad I_l(\zeta) \; = \; \textstyle\int_{-\infty}^{\zeta} Q(z) \; e_{V_0+}(z;k) \; e_{V_0-}(z;k) \; e_{V_0+}(\zeta;k) \; \mathrm{d}z, \\ &\qquad I_r(\zeta) \; = \; \textstyle\int_{\zeta}^{+\infty} Q(z) \; e_{V_0+}(z;k) \; e_{V_0+}(z;k) \; e_{V_0-}(\zeta;k) \; \mathrm{d}z. \end{aligned}$$

Il reste désormais à estimer le terme d'erreur, défini par

$$t_{rem}(k;Q) \; \equiv \; t(k;Q) \; - \; t_0^{hom}(k) \; - \; t_1[Q] \; - \; t_2[Q,Q]. \tag{B.51}$$

On a le résultat suivant :

PROPOSITION B.28. *Soit $t_{rem}(k;Q)$ défini par* (B.51). *Alors,*

i. *si V est à support compact, alors $t_{rem}(k) \; = \; \mathcal{O}(|||Q|||^3)$;*
ii. *si V est exponentiellement décroissante, alors $t_{rem}(k) \; = \; \mathcal{O}(|||Q|||^{3-})$;*
iii. *si $\langle x\rangle^{\rho+1} V_0(x) \in L^1(\mathbb{R})$ et $\langle x\rangle^{\rho} Q(x) \in L^2(\mathbb{R})$, $\rho > 8$, alors il existe $2 < \beta < 3$ tel que $t_{rem}(k) \; = \; \mathcal{O}(|||Q|||^{\beta})$.*

Le développement ainsi obtenu est en fait rigoureusement identique au développement (B.42) de la section précédente, dans le cas où $Q \equiv q_\epsilon(x) = q(x, x/\epsilon)$. Après quelques calculs, on s'aperçoit que $t_1[q_\epsilon]$ fournit les termes correcteurs $t_1^\epsilon(k) + \epsilon t_2^\epsilon(k)$, *via* les termes de bords après intégrations par parties ; $t_2[q_\epsilon, q_\epsilon]$ contribue au premier ordre à $t_2^{hom}(k)$. On a ainsi prouvé

PROPOSITION B.29. *S'il existe $\epsilon > 0$ tel que $\left\|T_{q_\epsilon}\right\|_{L^2\to L^2} = |||q(x, x/\epsilon))|||_{L^2\to L^2} < \infty$, alors on a pour ϵ petit,*

$$t^\epsilon(k) \equiv t_0^{hom}(k) + \epsilon t_1^\epsilon(k) + \epsilon^2 \left(t_2^{hom}(k) + t_2^\epsilon(k) \right) + t_{rem}^\epsilon(k), \tag{B.52}$$

avec

i. $t_{rem}^\epsilon(k) = \mathcal{O}(\epsilon^3)$ *si V est à support compact ;*
ii. $t_{rem}^\epsilon(k) = \mathcal{O}(\epsilon^{3-})$ *si V est exponentiellement décroissante ;*
iii. $t_{rem}^\epsilon(k) = \mathcal{O}(\epsilon^\beta)$ *avec $2 < \beta < 3$ si $\langle x\rangle^\rho V_0 \in L^1$, pour $\rho > 9$.*

B.3.3. *Deux exemples de microstructures avec discontinuités.* Il est intéressant de remarquer que mis sous la forme (B.50), le développement met en valeur un comportement particulier si le potentiel admet quelques symétries. Par exemple, supposons V_0 paire et q_ϵ « séparable » , dans le sens où

$$V_0(x) = V_0(-x), \quad q_\epsilon(x) = q_0(x)\, q_{\rm per}\left(\frac{x}{\epsilon}\right), \quad q_{\rm per} \text{ est } 1-\text{périodique}.$$

On peut alors facilement vérifier que $e_{V_0+}(\cdot;k)e_{V_0-}(\cdot;k)$ est paire. Ainsi, si q_0 et $q_{\rm per}$ sont de parité opposée, alors $x \mapsto e_{V_0+}(x;k)\, e_{V_0-}(x;k)\, q_\epsilon(x)$ est impaire et $t_1[q_\epsilon][k] \equiv 0$ quelque soit $\epsilon > 0$. Le premier ordre de correction à $t_0^{hom}(k)$ pour une telle classe de potentiels est $t_2[Q, Q]$, qui est d'ordre $\mathcal{O}(\epsilon^2)$, *même si q_ϵ est discontinue.* On a dans ce cas

$$\lim_{\epsilon\downarrow 0} \epsilon^{-2} \left(t^\epsilon(k) - t_0^{hom}(k) \right) = t_2^{hom}(k).$$

Étudions maintenant deux exemples particuliers.

i. Considérons une famille de structures, qui soient la troncature d'un potentiel lisse, telle que la troncature puisse générer des discontinuités. Pour préciser les idées, définissons

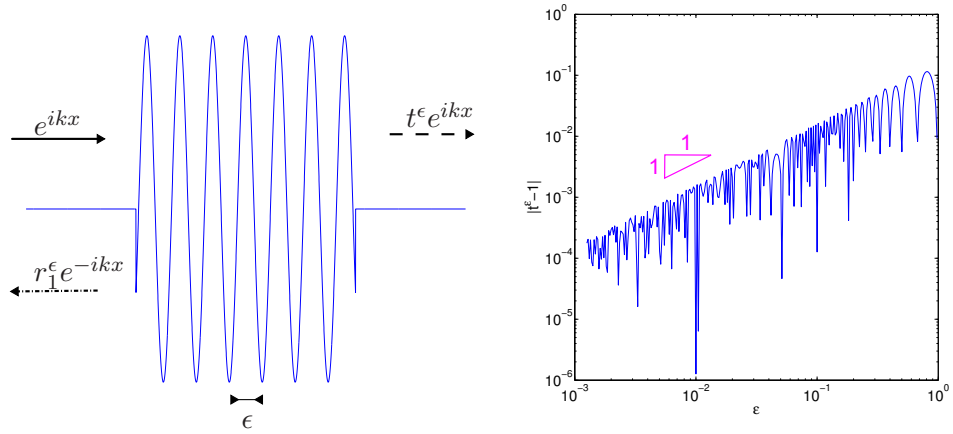
$$V_1^\epsilon(x;\theta) = \cos\left(\frac{2\pi x}{\epsilon} + \theta\right) \mathbf{1}_{[-1,1]}(x). \tag{B.53}$$

On est clairement dans le cas décrit ci-dessus, avec $V_0 \equiv 0$ (et donc $e_{V_0\pm}(x;k) = e^{\pm ikx}$ et $t_0^{hom} = 1$). Plus précisément, on prouve facilement que

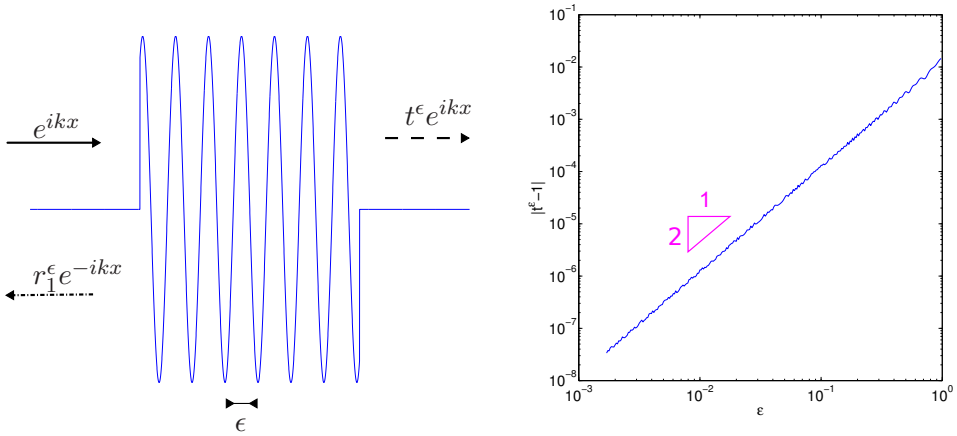
$$\begin{aligned} t_1^\epsilon(k;\theta) &\equiv \frac{1}{4k\pi} \sum_{j=1}^{M} e_{V_0+}(a_j;k)e_{V_0-}(a_j;k) \sum_{|l|\geq 1} [q_l]_{a_j} \frac{e^{2i\pi l \frac{a_j}{\epsilon}}}{l} \\ &= \frac{-i}{2k\pi} \cos(\theta) \sin\left(\frac{2\pi}{\epsilon}\right). \end{aligned}$$

En général, $t_1^\epsilon(k) \neq 0$ mais pour $\theta = \frac{\pi}{2} + m\pi$, $m \in \mathbb{Z}$, $q_\epsilon(\cdot;\theta)$ est paire et pour tout $k \in \mathbb{R}$ et $\epsilon > 0$, on a $t_1[k; q_\epsilon] = 0$.

Ces résultats sont résumés dans la figure 10 page suivante.



(a) $\theta = 0$.



(b) $\theta = \pi/2$.

FIGURE 10. Propagation des ondes à travers les matériaux de type i.

ii. Notre deuxième exemple provient d'une structure constante par morceaux, tronquée de manière lisse. On définit

$$V_2^\epsilon(x;\theta) \equiv h_{\text{per}}\left(\frac{x}{\epsilon}+\theta\right) e^{-\frac{x^2}{(x-1)(x+1)}} \mathbf{1}_{[-1,1]}(x), \tag{B.54}$$

où $h_{\text{per}}(y)$ est 1-périodique et vérifie $h(y) = -1$ pour $y \in (-1/2, 1/2]$ et $h(y) = 1$ pour $y \in (1/2, 3/2]$.

Puisque la composante à variable lente de $q_\epsilon(x)$ est lisse et que V_0 ne possède aucune singularité, on a

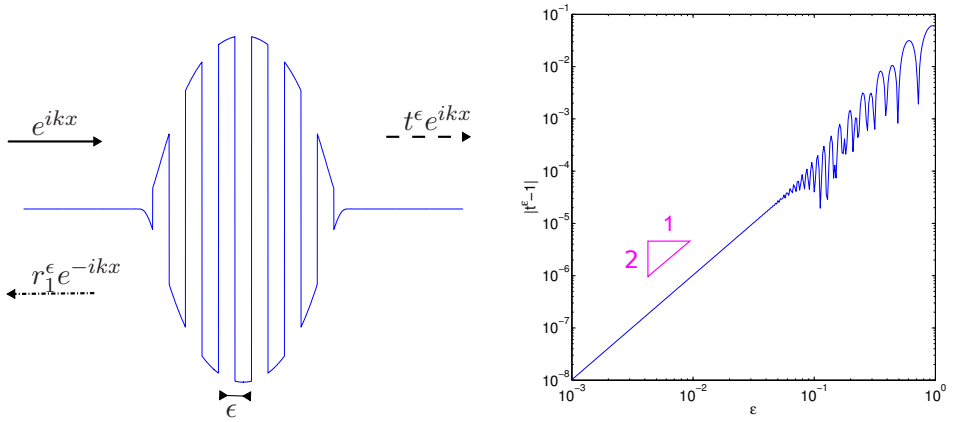
$$t^\epsilon - t_0^{hom} = \mathcal{O}(\epsilon^2), \quad \epsilon \to 0, \quad \text{et} \quad \lim_{\epsilon \downarrow 0} \epsilon^{-2}\left(t^\epsilon - t_0^{hom} \right) = t_2^{hom} \text{ est bien définie,}$$

et ceci même si $q_\epsilon(x)$ possède des discontinuités internes.

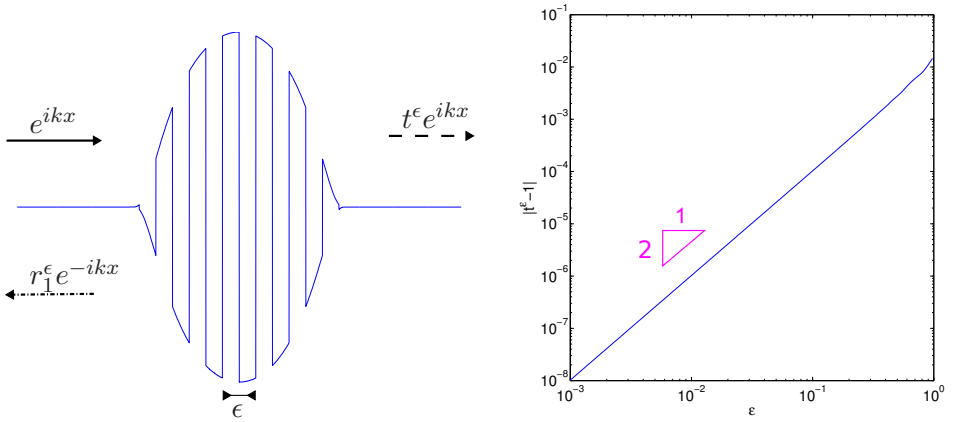
Pourtant, comme dans le cas de potentiels de type i., il existe une différence de comportement selon la parité du potentiel. On a :

- si $\theta \in \frac{1}{4} + \frac{1}{2}\mathbb{N}$ (*i.e.* q_ϵ est impaire), alors $t_1[k; q_\epsilon] = 0$;
- dans tous les cas, $t_1[k; q_\epsilon] = \mathcal{O}(\epsilon^3)$ $(\epsilon \to 0)$.

La réponse du système est effectivement dissemblable selon la parité du potentiel, comme on le voit dans la figure 11 page suivante.



(a) $\theta = 0$.



(b) $\theta = 1/4$.

FIGURE 11. Propagation des ondes à travers les matériaux de type ii.

Notations et autres supports

C.1. Notations utilisées. Les notations suivantes sont utilisées le long de l'introduction et sont respectées autant que possible tout au long du mémoire. Dans le cas contraire, les notations sont rappelées dans chaque Chapitre ou précisées lors de leur première utilisation, évitant – on l'espère – toute confusion.

Symboles.

- La notation $C_0(\lambda_1, \lambda_2, \ldots)$ décrit une constante positive, dépendant uniquement des paramètres $\lambda_1, \lambda_2, \ldots$ et dont la dépendance en λ_j est supposée croissante.
- On note $A \equiv B$ et parfois $A := B$ pour signifier que A est égal à B par définition.
- Les parties réelle et imaginaire d'un nombre complexe z sont respectivement notées $\Re(z)$ et $\Im(z)$.
- On utilise la notation des crochets japonais, à savoir
$$\langle x \rangle \ \equiv \ (1 + x^2)^{1/2}.$$
- Pour toute distribution tempérée u, on note $\widehat{u}$ (ou $\mathcal{F}u$) sa transformée de Fourier. On utilise la notation de Fourier standard $f(D)u$, définie en terme de transformée de Fourier par
$$\widehat{f(D)u}(\zeta) \ \equiv \ f(\zeta)\widehat{u}(\zeta).$$
- Λ est l'opérateur défini par $\Lambda = (1 + |D|^2)^{1/2}$.
- δ peut représenter selon le contexte la distribution de Dirac centrée en 0, ou une simple constante.

Dérivées partielles. Soit $0 < T \leq \infty$ et $f(t, x_1, \ldots, x_d)$ une fonction définie sur $[0, T] \times \mathbb{R}^d$. On note

- $\partial_t f$ la dérivée partielle par rapport au temps t.
- $\partial_{x_i} f$ la dérivée partielle par rapport à la variable d'espace x_i.
- Si f ne dépend que de la variable x_i, alors on notera
$$\frac{\mathrm{d}}{\mathrm{d}x_i} f \ \equiv \ \partial_{x_i} f.$$
- ∇f le gradient défini pour $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ par
$$\nabla f \ \equiv \ (\partial_{x_i})_{i=1,\ldots,d}$$
- $\operatorname{div} f$ la divergence définie pour $f \equiv (f_1, \ldots, f_d) : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ par
$$\operatorname{div} f \ \equiv \ \sum_{i=1}^{d} \partial_{x_i} f_i \ \equiv \nabla \cdot f$$
- Δf le Laplacien défini pour $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ par
$$\Delta f \ \equiv \ \sum_{i=1}^{d} \partial_{x_i}^2 f$$
et pour $f \equiv (f_1, \ldots, f_n) : [0, T] \times \mathbb{R}^d \to \mathbb{R}^n$ par
$$\Delta f \ \equiv \ (\Delta f_1, \ldots, \Delta f_n).$$

ESPACES FONCTIONNELS. Soit Ω un ouvert de $\mathbb{R}^d$, $d \geq 1$.

- On note $L^\infty(\Omega)$ l'espace de Lebesgue des fonctions mesurables $F : \Omega \to \mathbb{R}^n$ presque partout bornées par une constante finie. Cet espace est muni de la norme :
$$\left|F\right|_\infty \equiv \inf\{ M, \ |F| \leq M \ \text{ presque partout } \}.$$
- Pour $1 \leq p < \infty$, on note $L^p(\Omega)$ l'espace de Lebesgue des fonctions mesurables $F : \Omega \to \mathbb{R}^n$ telles que $|f|^p$ soit intégrable :
$$\left|F\right|_p \equiv \left(\int_{\mathbb{R}} |F(x)|^p dx\right)^{1/p} < \infty.$$
- Pour $k \in \mathbb{N}$, on note $W^{k,p}(\Omega)$ l'espace de Sobolev muni de la norme :
$$\left|F\right|_{W^{k,p}} \equiv \left|F\right|_p + \sum_{|\alpha|=k} \left|\nabla^\alpha F\right|_p.$$
- Les espaces de Sobolev $H^s(\Omega)$ (ou simplement H^s si le domaine est clair d'après le contexte) sont définis par la norme
$$\left|F\right|_{H^s} \equiv \left|\Lambda^s F\right|_2 \equiv \left|(1 + |\cdot|^2)^{s/2} \widehat{F}\right|_2.$$
- Soit $0 < T \leq \infty$ et $f(t,x)$ une fonction définie sur $[0,T] \times \mathbb{R}^d$. On a $f \in L^\infty([0,T]; H^s)$ si f est bornée dans $H^s(\mathbb{R}^d)$, uniformément par rapport à $t \in [0,T]$.
- Soit $0 < T \leq \infty$ et $f(t,x)$ une fonction définie sur $[0,T] \times \mathbb{R}^d$. On note alors $f \in W^{1,\infty}([0,T]; H^s)$ si $f \in L^\infty([0,T]; H^s)$ et $\partial_t f \in L^\infty([0,T]; H^s)$.
- On utilise la notation H^s_{loc} pour les fonctions localement dans H^s, c'est-à-dire vérifiant pour tout compact $\Omega \subset \mathbb{R}$, $u\,|_\Omega \in H^s$.
- Pour $\rho \in \mathbb{R}$, une fonction de Lebesgue $f \in L^{1,\rho}$ si elle est bornée pour la norme
$$\|f\|_{L^{1,\rho}} \equiv \int_{\mathbb{R}} (1 + |s|)^\rho \ |V(s)| \ \mathrm{d}s < \infty.$$

C.2. Étude du linéarisé du système d'Euler complet. Dans cette section, on étudie le système d'Euler complet (Σ') (voir page 6), linéarisé autour de la solution au repos. Un des intérêts de cette étude est d'obtenir un comportement approché des solutions de (Σ') pour de faibles perturbations, afin de choisir un adimensionnement adapté dans la section (A.1.3). Le système linéarisé est le suivant :

$$\text{(C.55)} \qquad \begin{cases} \partial_t \zeta_1 - G_1[0,0,0](\psi_1, \psi_2) = 0, \\ \partial_t \zeta_2 - G_2[0,0]\psi_2 = 0, \\ \partial_t \nabla\psi_1 + g\nabla\zeta_1 = 0, \\ \partial_t(\nabla\psi_2 - \rho_1/\rho_2 H[0,0,0](\psi_1,\psi_2)) + g(1-\rho_1/\rho_2)\nabla\zeta_2 = 0. \end{cases}$$

Or, on peut obtenir une expression explicite des opérateurs linéarisés $G_1[0,0,0]$, $G_2[0,0]$ et $H[0,0,0]$. En effet, appliquons l'opérateur de Fourier sur les variables horizontales des équations de Laplace (A.7). On obtient alors que $\widehat{\phi_2}$ et $\widehat{\phi_1}$ sont solutions de l'équation différentielle ordinaire suivante :
$$-|\xi|^2 y(\xi) + y''(\xi) = 0.$$

En utilisant les conditions aux bords, on en déduit :

$$\phi_2(X,z) = \cosh(|D|z)\psi_2(X) + \tanh(d_2|D|)\sinh(|D|z)\psi_2(X).$$

De la même manière, il suit

$$\begin{aligned}\phi_1(X,z) &= \cosh(|D|z)\big(\cosh^{-1}(d_1|D|)\psi_1(X) - \tanh(d_2|D|)\tanh(d_1|D|)\psi_2(X)\big)\\ &\quad + \tanh(d_2|D|)\sinh(|D|z)\psi_2(X).\end{aligned}$$

On en déduit finalement les expressions des opérateurs

$$\begin{aligned}G_1[0,0,0](\psi_1,\psi_2) &= \frac{|D|}{\cosh(d_1|D|)}\left(\sinh(d_1|D|)\psi_1 + \tanh(d_2|D|)\psi_2\right),\\ G_2[0]\psi_2 &= |D|\tanh(d_2|D|)\psi_2,\\ H[0,0,0](\psi_1,\psi_2) &= \frac{1}{\cosh(d_1|D|)}\nabla\psi_1(X) - \tanh(d_2|D|)\tanh(d_1|D|)\nabla\psi_2(X).\end{aligned}$$

Supposons maintenant que la surface est solide et plate[17] (*i.e.* $\zeta_1 \equiv 0$) et plongeons dans (C.55) l'expression des opérateurs. Il apparaît alors

$$\begin{aligned}&G_1[0,0,0](\psi_1,\psi_2) \equiv 0 \Longrightarrow \sinh(d_1|D|)\psi_1 = -\tanh(d_2|D|)\psi_2, \quad \text{et par suite}\\ &H[0,0,0](\psi_1,\psi_2) = -\frac{\tanh(d_2|D|)}{\tanh(d_1|D|)}\nabla\psi_2(X).\end{aligned}$$

Le système devient donc simplement

$$\left\{\begin{array}{l}\partial_t\zeta_2 - |D|\tanh(d_2|D|)\psi_2 = 0,\\ \partial_t\left(1+\rho_1/\rho_2\frac{\tanh(d_2|D|)}{\tanh(d_1|D|)}\right)\nabla\psi_2 + g(1-\rho_1/\rho_2)\nabla\zeta_2 = 0.\end{array}\right. \tag{C.56}$$

La *relation de dispersion* de ce système est immédiate. En effet, les fréquences d'onde $\omega(k)$, correspondant aux solutions d'onde plane $e^{ik\cdot X - i\omega t}$, sont solutions de l'équation

$$\omega^2\left(\frac{\rho_1}{\tanh(|k|d_1)} + \frac{\rho_2}{\tanh(|k|d_2)}\right) - g(\rho_2-\rho_1)|k| = 0 \tag{C.57}$$

On en déduit la vitesse caractéristique de propagation des ondes internes, par

$$c_0 \equiv \lim_{|k|\to 0}\left|\frac{\omega}{k}\right| = \sqrt{g(\rho_2-\rho_1)\frac{d_1 d_2}{\rho_2 d_1 + \rho_1 d_2}}.$$

Notons enfin que (C.56) offre les relations suivantes[18] :

$$\begin{aligned}|\psi_2| &\underset{|k|\to 0}{\sim} \frac{\omega}{k^2 d_2}|\zeta_2| = \frac{c_0}{k\ d_2}|\zeta_2|,\\ |\psi_1| &\underset{|k|\to 0}{\sim} \frac{d_2}{d_1}|\psi_2| = \frac{c_0}{k\ d_1}|\zeta_2|.\end{aligned}$$

[17] Cette hypothèse, souvent utilisée pour simplifier le système, est étudiée en section A.3.2 page 17. Elle est valable notamment lorsque $\rho_1/\rho_2 \sim 1$, ce qui est le cas dans la plupart des applications océanographiques. Le cas de la surface libre se traite identiquement et entraîne deux modes, avec deux vitesses caractéristiques.

[18] Nous nous intéressons au comportement des solutions lorsque $k \to 0$, car cela correspond aux régimes d'eaux peu profondes auxquels nous nous attacherons dans la suite, voir Section A.2 page 9.

C.3. Schémas numériques des simulations de la partie A. Dans cette section, on présente la méthode numérique utilisée pour les différentes simulations présentées dans la section A et plus généralement aux chapitres 3 et 4. Ces simulations utilisent un schéma aux différences finies de type Crank-Nicholson ; les termes non-linéaires sont traités grâce à un demi-pas prédictif, selon une méthode présentée par Besse et Bruneau [**12, 13**]. De tels schémas sont utilisés dans le cadre des ondes de surface dans [**27, 42**] par exemple. Nous donnons dans la suite les idées générales de la méthode et le schéma explicite dans le cas simple de l'équation Korteweg-de Vries. Le cas des équations couplées est plus lourd à présenter et on les omet ici afin de faciliter la lecture ; le lecteur intéressé pourra se reporter aux sections 4.1 (Chapitre 3, page 110) et D (Chapitre 4, page 162) pour plus de détails.

DISCRÉTISATION EN TEMPS. Notons Δt le pas de temps du schéma. La fonction $u(t,x)$ est approchée au temps $t = n\Delta t$ par la fonction $u^n(x)$: $u(n\Delta t, x) \equiv u^n(x)$. Le schéma itératif permettant de définir u^n est alors obtenu à partir des équations vérifiées par u et par les approximations suivantes : la dérivée en temps discrète est définie par

$$\partial_t u \approx \frac{u^{n+1} - u^n}{\Delta t},$$

et pour F une application linéaire en u, on pose

$$F(u) \approx \frac{F(u^{n+1}) + F(u^n)}{2}.$$

Comme nous l'avons dit plus haut, les non-linéarités sont traitées à l'aide d'un demi-pas prédictif, défini à chaque itération par

$$\frac{u^{n+\frac{1}{2}} + u^{n-\frac{1}{2}}}{2} \;=\; u^n.$$

Ainsi, la discrétisation d'un terme bilinéaire $F(u,v)$ sera une combinaison linéaire des deux discrétisations possibles :

$$F(u,v) \approx \alpha F\left(u^{n+\frac{1}{2}}, \frac{v^{n+1} + v^n}{2}\right) + (1-\alpha) F\left(\frac{u^{n+1} + u^n}{2}, v^{n+\frac{1}{2}}\right).$$

Le paramètre α nous offre une liberté ; il pourra être choisi afin que des quantités naturelles du système soient conservées.

DISCRÉTISATION EN ESPACE. En notant Δx le pas d'espace du schéma, la discrétisation spatiale utilise les différences centrées. En particulier, on utilise

$$u(x,t) \approx M(\beta)u \text{ où } (M(\beta)u^n)_i = (1-\beta)u_i^n + \frac{\beta}{2}(u_{i+1}^n + u_{i-1}^n),$$

et le paramètre β peut encore être choisi librement. Les dérivées spatiales discrètes sont données par

$$\partial_x u \approx \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} \equiv (D_1 u^n)_i, \quad \partial_x^2 u \approx \frac{u_{i+1}^n - 2u_i^2 + u_{i-1}^n}{\Delta x^2} \equiv (D_2 u^n)_i, \quad \ldots$$

On utilise des conditions périodiques au bord.

Finalement, en fixant $u_i^0 \equiv u(i\Delta x, 0)$ et en calculant le premier demi-pas $u^{1/2}$ à l'aide d'un simple schéma explicite, on obtient le schéma complet. Les schémas obtenus de cette manière sont formellement d'ordre $\mathcal{O}(\Delta x^2 + \Delta t^2)$, ce qui est confirmé par nos simulations, et s'avèrent être inconditionellement stable. On trouvera en Table 2, page 112, quelques données sur la taille des erreurs engendrées par nos schémas.

L'équation de Korteweg-de Vries. Donnons un exemple précis : le cas de l'équation de Korteweg-de Vries. On veut résoudre numériquement l'équation

$$\partial_t u \ + \ c\partial_x u \ + \ \lambda u\partial_x u \ + \ \nu\partial_x^3 u \ = \ 0, \tag{C.58}$$

dont on sait qu'elle préserve l'énergie $E \equiv \left|u\right|_{L^2}^2$. Israwi et Duruflé [**42**, Théorème 2] ont présenté le schéma suivant, qui découle de la méthode présentée ci-dessus :

$$\frac{u_i^{n+1}-u_i^n}{\Delta t} + c\left(D_1\frac{u^{n+1}+u^n}{2}\right)_i + \nu\left(D_3\frac{u^{n+1}+u^n}{2}\right)_i \tag{C.59}$$
$$+\frac{\lambda}{3}\left(\left(M(1)\frac{u_i^{n+1}+u_i^n}{2}\right)_i\left(D_1u^{n+\frac{1}{2}}\right)_i + 2\left(M(1/2)u^{n+\frac{1}{2}}\right)_i\left(D_1\frac{u^{n+1}+u^n}{2}\right)_i\right) = 0.$$

Ce schéma a la propriété agréable de conserver l'énergie discrète, garantissant sa stabilité.

Proposition C.2. *Le schéma* (C.59) *conserve la norme discrète* l^2, i.e. *pour tout* $n \in \mathbb{N}$

$$\left|u^n\right|_{l^2}^2 \ \equiv \ \sum_i |u_i^n|^2 \ = \ \sum_i |u_i^0|^2.$$

C.4. Coefficient de transmission dans le cas d'un potentiel singulier uniforme. On s'intéresse ici à l'équation (B.18) (voir page 28), dans le cas d'un potentiel uniquement composé d'une somme finie de distributions de Dirac, de même hauteur et régulièrement espacées. On suppose donc

$$V \ = \ c\sum_{j=0}^{N-1}\delta(x-jl), \quad \text{avec } c\in\mathbb{R},\ N\in\mathbb{N}^* \ \text{ et } \ l>0.$$

D'après la section B.2.1, on sait que l'onde plane distordue vérifie

$$e_{V_{\text{sing}}+}(x;k) = \begin{cases} e^{ikx}+B_0e^{-ikx} & \text{pour } x<x_0 \\ A_1e^{ikx}+B_1e^{-ikx} & \text{pour } x_0<x<x_1 \\ \vdots & \\ A_Ne^{ikx} & \text{pour } x>x_{N-1} \end{cases}$$

et les conditions de saut ci-dessus s'écrivent dans notre cas

$$\begin{pmatrix} A_je^{-ikjl} \\ B_je^{ikjl}\end{pmatrix} = \begin{pmatrix} 1+i\beta & i\beta e^{-2ikjl} \\ -i\beta e^{2ikjl} & 1-i\beta\end{pmatrix}\begin{pmatrix} A_{j+1}e^{-ik(j+1)l} \\ B_{j+1}e^{ik(j+1)l}\end{pmatrix}, \text{ pour } j=0,\dots,N-1,$$

où l'on note $\beta=\frac{c}{k}$. On a donc

$$\begin{pmatrix} A_0 \\ B_0\end{pmatrix} = P^N\begin{pmatrix} A_N \\ B_N\end{pmatrix},$$

avec P la *matrice de transmission* définie par

$$P \ \equiv \ \begin{pmatrix} (1+i\beta)e^{-ikl} & i\beta e^{ikl} \\ -i\beta e^{-ikl} & (1-i\beta)e^{ikl}\end{pmatrix}\begin{pmatrix} A_{j+1} \\ B_{j+1}\end{pmatrix}.$$

Le théorème de Cayley-Hamilton, en utilisant le fait que P soit unitaire, nous donne la relation suivante

$$P^2 \ - \ 2\xi P \ + \ I \ = \ 0. \tag{C.60}$$

avec $\xi=\frac{\mathrm{tr}(P)}{2}=\cos(kl)+\beta\sin(kl)$.

Ainsi, P^N est un polynôme en P, de degré au plus 1. Il existe donc U_N et V_N tels que

$$P^N = U_N(\xi)P + V_N(\xi)I. \tag{C.61}$$

De (C.60) et (C.61), on déduit

$$P^{N+1} = U_{N+1}(\xi)P + V_{N+1}(\xi) = (V_N(\xi)+2\xi U_N(\xi))P - U_N(\xi)I,$$

et U_N et V_N sont donc définis par la relation de récurrence suivante :

- $U_0(\xi) = 0, \qquad U_1(\xi) = 1\,;$
- $U_{N+1}(\xi) = 2\xi U_N(\xi) - U_{N-1}(\xi)\,;$
- $V_{N+1}(\xi) = -U_N(\xi).$

On reconnait ici la relation de récurrence des polynômes de Tchebychev. En particulier, si $|\xi| < 1$, on a

$$U_N(\xi) = \frac{\sin(N\gamma)}{\sin\gamma}, \text{ où } \gamma = \cos^{-1}(\xi). \tag{C.62}$$

On peut ensuite utiliser les calculs suivants :

$$\begin{aligned}
|P^N_{(1,1)}|^2 =& |U_N(\xi)P_{(1,1)} - U_{N-1}(\xi)|^2 \\
=& |U_N(\xi)(1+i\beta)e^{-ikl} - U_{N-1}(\xi)|^2 \\
=& U_N(\xi)^2|1+i\beta|^2 + U_{N-1}(\xi)^2 - ((1+i\beta)e^{-ikl} + \overline{(1+i\beta)e^{-ikl}})U_N(\xi)U_{N-1}(\xi) \\
=& U_N(\xi)^2 + U_{N-1}(\xi)^2 - 2\xi U_N(\xi)U_{N-1}(\xi) + \beta^2 U_N(\xi)^2 \\
=& W_N(\xi) + \beta^2 U_N(\xi)^2,
\end{aligned}$$

où l'on a utilisé

$$W_N(\xi) \equiv U_N(\xi)^2 + U_{N-1}(\xi)^2 - 2\xi U_N(\xi)U_{N-1}(\xi).$$

De plus, on vérifie aisément

$$\begin{aligned}
W_N(\xi) - W_{N-1}(\xi) =& U_N(\xi)^2 - U_{N-2}(\xi)^2 - 2\xi U_{N-1}(\xi)(U_N(\xi) - U_{N-2}(\xi)) \\
=& U_N(\xi)^2 - U_{N-2}(\xi)^2 - 2\xi(U_N(\xi) + U_{N-2}(\xi))(U_N(\xi) - U_{N-2}(\xi)) \\
=& 0.
\end{aligned}$$

Ainsi, $W_N(\xi) = W_1(\xi) = 1$ et donc $|P^N_{(1,1)}|^2 = 1 + \beta^2 U_N(\xi)^2$. Finalement, on aboutit à

$$|t(k)| = \frac{1}{|P^N_{(1,1)}|} = \frac{1}{(1 + (\frac{c}{k})^2 U_N(\xi)^2)^{1/2}}. \tag{C.63}$$

On a donc bien obtenu une formule explicite du coefficient de transmission. On trouvera en Figure 8, page 34, les variations de ce coefficient, en fonction de k, pour un potentiel contenant respectivement 3 et 20 distributions de Dirac.

Partie A. Ondes internes de gravité en océanographie

CHAPITRE 2

Asymptotic shallow water models for internal waves in a two-fluid system with a free surface

Article publié dans *SIAM Journal on Mathematical Analysis* **42**, 5 (2010), pages 2229–2260.

Sommaire

Abstract

In this chapter, we derive asymptotic models for the propagation of two- and three-dimensional gravity waves at the free surface and the interface between two layers of immiscible fluids of different densities over an uneven bottom. We assume the thickness of the upper and lower fluids to be of comparable size and small compared to the characteristic wavelength of the system (shallow water regimes). Following a method introduced by Bona, Lannes, and Saut [*J. Math. Pures Appl.* (9), 89 (2008), pp. 538–566] based on the expansion of the involved Dirichlet-to-Neumann operators, we are able to give a rigorous justification of classical models for weakly and strongly nonlinear waves, as well as interesting new ones. In particular, we derive linearly well-posed systems in the so-called Boussinesq/Boussinesq regime. Furthermore, we establish the consistency of the full Euler system with these models and deduce the convergence of the solutions.

1. Introduction

1.1. General settings. This chapter deals with weakly and strongly nonlinear internal waves in a two-fluid system. We consider the case of uneven bottom topography and free surface, although the rigid-lid assumption is mentioned. The idealized system studied here consists of two layers of immiscible, homogeneous, ideal, incompressible, and irrotational fluids under only the influence of gravity.

The mathematical theory of internal waves, following the theory of free-surface water waves, has attracted lots of interest over the past decades. We let the reader refer to the survey article of Helfrich and Melville [**52**] for a good overview of the ins and outs on this problem. The governing equations, which we call the full Euler system, are fully nonlinear, and their direct study and computation remain a real obstacle. In particular, the well-posedness of the equations in Sobolev space is challenging, as discussed in Remark 1.1. An alternative way is to look for approximations through the use of asymptotic models. Such models can be derived from the full Euler system by introducing natural dimensionless parameters of the system and by setting some smallness hypotheses on these parameters (thus reducing the framework to more specific physical regimes).

Many models for a two-fluid system have already been derived and studied. Systems under the rigid-lid assumption have first been investigated (see [**89**] or [**81**], for example). Weakly nonlinear models in the free-surface case have been presented by Choi and Camassa [**30**]. Nguyen and Dias [**93**] presented a great deal of numeric simulations for such Boussinesq-type systems. Strongly nonlinear regimes have been derived by Matsuno [**82**], Choi and Camassa [**31**], and Barros, Gavrilyuk, and Teshukov [**8**], generalizing the classical Green–Naghdi model (see [**48**]). A different approach has been carried out by Craig, Guyenne, and Kalisch [**35**], using the Hamiltonian formulation of the Euler equations. Most of these works are formal and are restricted to two-dimensional flows or to the flat-bottom case. Finally, we refer to the work of Bona, Lannes, and Saut [**17**] who, following a strategy initiated in [**16, 14, 15**], rigorously derived a large class of models in different regimes under the rigid-lid assumption. This chapter is concerned with the more complex case where the rigid-lid assumption is removed and replaced by a free surface.

The strategy consists of rewriting the full system as a system of four evolution equations located on the surface and the interface between the two fluids (as opposed to two equations in the rigid-lid case). The reformulation introduces a Dirichlet-to-Neumann operator $G[\zeta]$ and an interface operator $H[\zeta]$, defined precisely below. The computation of asymptotic expansions of these operators leads to the models presented here. We focus here on shallow water regimes, allowing strongly nonlinear waves.

Our analysis gives a rigorous derivation of most of the models existing in the literature and also interesting new ones. In particular, we derive a set of models in the Boussinesq/Boussinesq regime, with coefficients that can be chosen so that the system is linearly well-posed. We prove that the full Euler system is consistent with each of our models, which roughly means that any solution of the full system solves the asymptotic systems up to a small error. Then in the case of the shallow water/shallow water model, using energy methods together with consistency, we also prove that the solutions of our models converge toward the solutions of the full Euler system, assuming that such solutions exist.

The chapter is organized as follows. Section 1 is devoted to the reformulation of the full system, from the Euler equation to the "Zakharov formulation," written in dimensionless form. In section 1.6, we focus on the linearized system, and its dispersion relations are derived. From the asymptotic expansion of the operators $G[\zeta]$ and $H[\zeta]$ presented in section 2, the asymptotic models under different regimes are rigorously obtained and are presented in section 2.3. The consistency of the full Euler system with each of our models is proved. Then section 3 gives convergence results: we show that the solutions of the full Euler system tend to associated solutions of one of our models in the shallow water limit.

Finally the links with different models already existing in the literature are presented in section 4 for rigid-lid models [**17**] and layer-mean equations [**31, 30**]. The proof of Proposition 2.5 is given in Appendix A.

1.2. Notation. We use the Cartesian coordinates (X, z), where z is the vertical variable and X is the d-dimensional horizontal variable: $X = x$ when $d = 1$ and $X = (x, y)$ when $d = 2$.

The symbols ∇ and Δ denote the gradient and Laplace operators in the horizontal variables, respectively, whereas $\nabla_{X,z}$ and $\Delta_{X,z}$ are their $(d+1)$-variable version. For $\mu > 0$, we also define the scaled version of the gradient and Laplace operators, namely, $\nabla^{\mu}_{X,z} := (\sqrt{\mu}\nabla^T, \partial_z)^T$ and $\Delta^{\mu}_{X,z} := \mu\Delta + \partial_z^2$, respectively.

Given a surface $\Gamma := \{(X, z), z = \zeta(X)\}$, we denote by ∂_n the upward normal derivative at Γ:

$$\partial_n := n \cdot \nabla_{X,z} \text{ with } n := \frac{1}{\sqrt{1 + |\nabla\zeta|^2}}(-\nabla\zeta, 1)^T.$$

If we consider an elliptic operator $\mathbf{P} = \nabla_{X,z} \cdot P\nabla_{X,z}$, then the conormal derivative associated to $\mathbf{P}$ is

$$\partial_n^P := n \cdot P\nabla_{X,z},$$

which we simply denote ∂_n when there is no risk of confusion.

For any tempered distribution u, we denote by $\widehat{u}$ its Fourier transform. We use the standard Fourier multiplier notation $f(D)u$, defined in terms of Fourier transforms by

$$\widehat{f(D)u} := f\widehat{u}.$$

The operator $\Lambda = (1 - \Delta)^{1/2}$ is equivalently defined using the Fourier multiplier notation to be $\Lambda = (1 + |D|^2)^{1/2}$.

We denote by $H^s(\mathbb{R}^d)$ (or simply H^s if the underlying domain is clear from the context) the L^2-based Sobolev spaces. Their norm is written $\left|\cdot\right|_{H^s}$ and simply $\left|\cdot\right|_2$ for the L^2 norm.

Then, for $0 < T \leq \infty$, $q \in \mathbb{N}$, $W^{q,\infty}([0,T]; H^s(\mathbb{R}^d))$ (or simply $W^{q,\infty}H^s$ and $L^\infty H^s$ when $q = 0$) denotes the space of the functions $f(t, X)$ defined on $[0, T] \times \mathbb{R}^d$, whose derivatives up to the order q in t are bounded in $H^s(\mathbb{R}^d)$ uniformly with respect to $t \in [0, T)$. Their norm is written $\left|\cdot\right|_{W^{q,\infty}H^s}$.

Since it often appears, it is convenient to introduce for s and $T > 0$ the space $\mathbf{H}^s([0,T])$, made up of the quadruplets $(\zeta_1, \zeta_2, u_1, u_2)$ such that their components satisfy $u_1,\ u_2 \in W^{1,\infty}([0,T]; H^{s+d+7/2}(\mathbb{R}^d))^d$, $\zeta_1 \in W^{1,\infty}([0,T]; H^{s+d+3/2}(\mathbb{R}^d))$, and finally $\zeta_2 \in W^{1,\infty}([0,T]; H^{s+d+5/2}(\mathbb{R}^d))$. Their norm is written $\left|\cdot\right|_{\mathbf{H}^s}$.

Finally we denote by $\mathcal{S}^+$ the planar strip $\mathbb{R}^d \times (0, 1)$ and by $\mathcal{S}^-$ the planar strip $\mathbb{R}^d \times (-1, 0)$. We use the notation $\left\|\cdot\right\|_{H^s}$ for the usual norm of $H^s(\mathcal{S}^\pm)$ and simply $\left\|\cdot\right\|_2$ for the $L^2(\mathcal{S}^\pm)$ norm. We also for $s \in \mathbb{R}$ and $k \in \mathbb{N}$ introduce the spaces

$$H^{s,k}(\mathcal{S}^\pm) = \{f \in \mathcal{D}'(\overline{\mathcal{S}^\pm}) : \left\|f\right\|_{H^{s,k}} < \infty\},$$

where $\left\|f\right\|_{H^{s,k}} = \sum_{j=0}^k \left\|\Lambda^{s-j}\partial_z^j f\right\|_2$.

1.3. The basic equations. We assume that each fluid is irrotational and incompressible so that we can introduce velocity potentials ϕ_i $(i = 1, 2)$, respectively, associated to the upper and lower fluid layers. The velocity potentials satisfy

$$\Delta_{X,z}\phi_i = 0 \text{ in } \Omega_t^i, \tag{1.1}$$

where Ω_t^i denotes the domain of the fluid i at time t (see Figure 1). Moreover, we assume the fluids to satisfy the Euler equation and their respective density ρ_i is constant so that the velocity potentials satisfy the Bernoulli equation

$$\partial_t\phi_i + \frac{1}{2}|\nabla_{X,z}\phi_i|^2 = -\frac{P}{\rho_i} - gz \text{ in } \Omega_t^i, \tag{1.2}$$

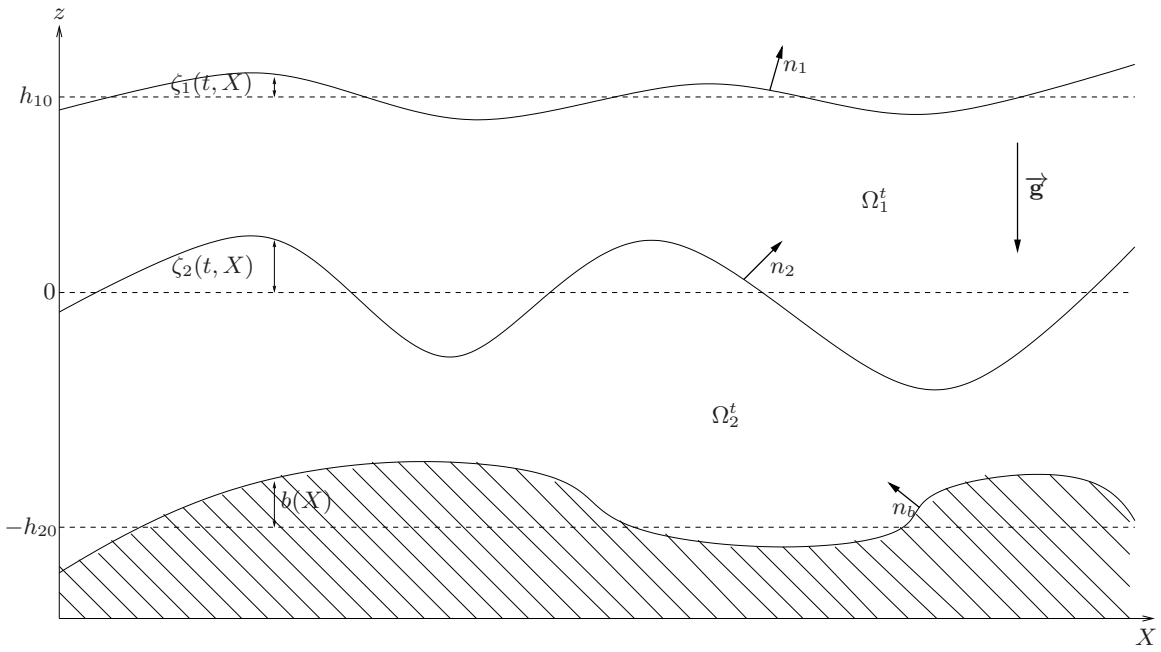


Figure 1. Sketch of the domain.

where g denotes the acceleration of gravity and P is the pressure inside the fluid. The kinematic boundary condition at the known, constant with respect to time, bottom topography $\Gamma_b := \{z = -h_{20} + b(X)\}$ is given by

$$\partial_n \phi_2 = 0 \text{ on } \Gamma_b. \tag{1.3}$$

It is presumed that the surface and the interface are given as the graph of functions (respectively, $\zeta_1(t,X)$ and $\zeta_2(t,X)$) which express the deviation from their rest position (respectively, (X,h_{10}) and $(X,0)$) at the spatial coordinate X and at time t. The assumption that no fluid particle crosses the surface or the interface gives the following kinematic boundary conditions:

$$\begin{array}{lll} \partial_t \zeta_1 &= \sqrt{1+|\nabla\zeta_1|^2}\partial_n\phi_1 & \text{on } \Gamma_1 := \{z = h_{10} + \zeta_1(t,X)\}, \\ \partial_t \zeta_2 &= \sqrt{1+|\nabla\zeta_2|^2}\partial_n\phi_1 = \sqrt{1+|\nabla\zeta_2|^2}\partial_n\phi_2 & \text{on } \Gamma_2 := \{z = \zeta_2(t,X)\}. \end{array} \tag{1.4}$$

Finally we close the set of equations by assuming that

$$P \text{ is constant at the surface and continuous at the interface.} \tag{1.5}$$

Remark 1.1. Unlike the water wave problem (air-water interface), the Cauchy problem associated with waves at the interface of two fluids of positive different densities is known to be ill-posed in Sobolev spaces in the absence of surface tension, as Kelvin–Helmholtz instabilities appear. However, when adding a small amount of surface tension, Lannes **[70]** proved that, thanks to a stability criterion, the problem becomes well-posed with a time of existence that does not vanish as the surface tension goes to zero and thus is consistent with the observations. The Kelvin–Helmholtz instabilities appear for high frequencies, where the regularization effect of the surface tension is relevant, while the main profile of the wave that we want to capture is located in lower frequencies and is unaffected by surface tension. Therefore, adding a small amount of surface tension at the interface[1] in the Euler system guarantees the well-posedness of the system and does not change our asymptotic models. For simplicity, we decide to omit this surface tension term.

[1]The study of Lannes focuses on the two-layer fluid system with a rigid lid. However, we believe that the theory in the free-surface case does not differ much from the one in the rigid-lid configuration.

1.4. Reduction of the equations. In **[122]**, Zakharov remarked that the surface wave system can be fully deduced from the knowledge of the surface elevation and the trace of the velocity potential at the surface. We extend it here for two fluids in the free-surface case. Indeed, if we introduce the traces

$$\psi_1(t,X) := \phi_1(t,X,h_{10}+\zeta_1(t,X)) \text{ and } \psi_2(t,X) := \phi_2(t,X,\zeta_2(t,X)),$$

then ϕ_2 is uniquely given as the solution of Laplace's equation (1.1) in the lower fluid domain, with the Neumann condition (1.3) on Γ_b and the Dirichlet condition $\phi_2 = \psi_2$ on Γ_2. Then ϕ_1 is obtained as the solution of Laplace's equation on the upper fluid domain, with the Neumann condition given by (1.4) as $\partial_n\phi_2 = \partial_n\phi_1$ on Γ_2 and the Dirichlet condition as $\phi_1 = \psi_1$ on Γ_1.

Following the formalism introduced by Craig, Sulem, and Sulem in **[36]** (see **[11, 34]** for the bi-fluidic rigid-lid case), we first define the Dirichlet–Neumann operators:

$$\begin{aligned} G_1[\zeta_1,\zeta_2,b](\psi_1,\psi_2) &:= \sqrt{1+|\nabla\zeta_1|^2}\partial_n\phi_{1|z=h_{10}+\zeta_1}, \\ G_2[\zeta_2,b]\psi_2 &:= \sqrt{1+|\nabla\zeta_2|^2}\partial_n\phi_{2|z=\zeta_2}. \end{aligned}$$

We also define the following operator:

$$H[\zeta_1,\zeta_2,b](\psi_1,\psi_2) := \nabla\big(\phi_{1|z=\zeta_2}\big).$$

Using the chain rule and the last definitions in the relation (1.2) evaluated at the surface, we obtain

$$\partial_t\psi_1 + g(h_{10}+\zeta_1) + \frac{1}{2}|\nabla\psi_1|^2 - \frac{(G_1[\zeta_1,\zeta_2,b](\psi_1,\psi_2)+\nabla\zeta_1\cdot\nabla\psi_1)^2}{2(1+|\nabla\zeta_1|^2)} = -\frac{P_1}{\rho_1}, \tag{1.6}$$

where P_1 is the constant pressure at the surface. Using again the Bernoulli equation for the upper and the lower fluids evaluated at the interface, we have

$$\begin{aligned} &\partial_t(\phi_{1|z=\zeta_2}) + g\zeta_2 + \frac{1}{2}|H[\zeta_1,\zeta_2,b](\psi_1,\psi_2)|^2 \\ &\qquad - \frac{(G_2[\zeta_2,b]\psi_2+\nabla\zeta_2\cdot H[\zeta_1,\zeta_2,b](\psi_1,\psi_2))^2}{2(1+|\nabla\zeta_2|^2)} = -\frac{P_2}{\rho_1}, \end{aligned} \tag{1.7}$$

$$\partial_t\psi_2 + g\zeta_2 + \frac{1}{2}|\nabla\psi_2|^2 - \frac{(G_2[\zeta_2,b]\psi_2+\nabla\zeta_2\cdot\nabla\psi_2)^2}{2(1+|\nabla\zeta_2|^2)} = -\frac{P_2}{\rho_2}, \tag{1.8}$$

where P_2 is the pressure at the interface, identical in (1.7) and (1.8), thanks to the continuity assumption in (1.5).

Finally, using (1.4), the gradient of the equality (1.6), and a straightforward combination of (1.7) and (1.8), we obtain the system of equations

$$\begin{cases} \partial_t\zeta_1 - G_1[\zeta_1,\zeta_2,b](\psi_1,\psi_2) = 0, \\ \partial_t\zeta_2 - G_2[\zeta_2,b]\psi_2 = 0, \\ \partial_t\nabla\psi_1 + g\nabla\zeta_1 + \frac{1}{2}\nabla(|\nabla\psi_1|^2) - \nabla\mathcal{N}_1(\zeta_1,\zeta_2,b,\psi_1,\psi_2) = 0, \\ \partial_t(\nabla\psi_2 - \gamma H[\zeta_1,\zeta_2,b](\psi_1,\psi_2)) + g(1-\gamma)\nabla\zeta_2 \\ \quad + \frac{1}{2}\nabla(|\nabla\psi_2|^2 - \gamma|H[\zeta_1,\zeta_2,b](\psi_1,\psi_2)|^2) - \nabla\mathcal{N}_2(\zeta_1,\zeta_2,b,\psi_1,\psi_2) = 0, \end{cases} \tag{1.9}$$

where $\gamma = \frac{\rho_1}{\rho_2}$ and

$$\begin{aligned} \mathcal{N}_1(\zeta_1,\zeta_2,b,\psi_1,\psi_2) &= \tfrac{(G_1[\zeta_1,\zeta_2,b](\psi_1,\psi_2)+\nabla\zeta_1\cdot\nabla\psi_1)^2}{2(1+|\nabla\zeta_1|^2)}, \\ \mathcal{N}_2(\zeta_1,\zeta_2,b,\psi_1,\psi_2) &= \tfrac{(G_2[\zeta_2,b]\psi_2+\nabla\zeta_2\cdot\nabla\psi_2)^2-\gamma(G_2[\zeta_2,b]\psi_2+\nabla\zeta_2\cdot H[\zeta_1,\zeta_2,b](\psi_1,\psi_2))^2}{2(1+|\nabla\zeta_2|^2)}. \end{aligned}$$

This is the system of equations that we use in order to derive asymptotic models.

1.5. Nondimensionalization of the equations. In this subsection, we rewrite the system (1.9) in dimensionless variables, introducing dimensionless parameters which are crucial to study the asymptotic dynamics. We denote by a_1 the typical amplitude of the surface deformation and by a_2 that of the interface. λ is a characteristic horizontal length, say, the wavelength of the interface. Finally B is the order of bottom topography variation.

We define the dimensionless variables

$$\tilde{X} := \frac{X}{\lambda}, \quad \tilde{z} := \frac{z}{h_{10}}, \quad \tilde{t} := \frac{t}{\lambda/\sqrt{gh_{10}}}, \quad \tilde{b}(\tilde{X}) := \frac{b(X)}{B}$$

and the dimensionless unknowns

$$\tilde{\zeta}_i(\tilde{X}) := \frac{\zeta_i(X)}{a_i}, \quad \tilde{\psi}_i(\tilde{X}) := \frac{\psi_i(X)}{a_2\lambda\sqrt{g/h_{10}}}.$$

Five independent parameters of the system are thus added to $\gamma = \frac{\rho_1}{\rho_2}$:

$$\epsilon_1 := \frac{a_1}{h_{10}}, \quad \epsilon_2 := \frac{a_2}{h_{10}}, \quad \mu := \frac{h_{10}^2}{\lambda^2}, \quad \delta := \frac{h_{10}}{h_{20}}, \quad \beta := \frac{B}{h_{10}}.$$

So, ϵ_1 and ϵ_2 are the nonlinearity parameters, and μ is the shallowness parameter. We also define the convenient notation

$$\alpha := \frac{a_1}{a_2} = \frac{\epsilon_1}{\epsilon_2}.$$

Remark 1.2. The scaling for nondimensionalization has been chosen considering the solutions of the linearized problem that can be computed with the physical variables using the method of section 1.6 (see **[105]**, for example). Using such a scaling, we implicitly assume that the two layers are of similar depth (i.e., $\delta \sim 1$). Therefore, the choice of h_{10} (and not h_{20}) as the reference vertical length and $\sqrt{gh_{10}}$ as the reference velocity are harmless. We refer, for example, to **[65, 61]** for the investigation of different situations such as the deep-water regime or the finite-depth regime.

In the same way, we decide to use the same scaling on ψ_1 and ψ_2 in order to simplify Definitions 1.3 and 1.4 and especially to keep the relation $\partial_n\phi_1 = \partial_n\phi_2$ on the interface. We choose a_2 instead of a_1 so that the expansions of section 2.2 hold for α tending to zero. In that way, we are able to retrieve the shallow water/shallow water with rigid-lid model in section 4.1.

Finally, as we choose a unique characteristic horizontal length, we focus only on the case where the internal and surface waves have length scales of the same order and hence do not consider phenomenons such as the resonant interaction between a long internal wave and short surface waves, as studied, for example, in **[44]**. Moreover, in the case of three-dimensional waves ($d = 2$), a unique characteristic horizontal length means that there is no preferential horizontal direction and we therefore do not study transverse waves.

We now rewrite the system in terms of dimensionless variables. First, we have to define the dimensionless operators associated to the dimensionless fluid domains:

$$\begin{aligned}\Omega_1 &:= \{(X,z) \in \mathbb{R}^{d+1}, \epsilon_2\zeta_2(X) < z < 1 + \epsilon_1\zeta_1(X)\},\\ \Omega_2 &:= \{(X,z) \in \mathbb{R}^{d+1}, -\frac{1}{\delta} + \beta b(X) < z < \epsilon_2\zeta_2(X)\}.\end{aligned}$$

In the following, we always assume that the domains remain strictly connected, so there is a positive value h such that, for all $X \in \mathbb{R}^d$,

$$1 + \epsilon_1\zeta_1(X) - \epsilon_2\zeta_2(X) \geq h > 0 \qquad \text{and} \qquad \frac{1}{\delta} + \epsilon_2\zeta_2(X) - \beta b(X) \geq h > 0. \tag{1.10}$$

DEFINITION 1.3. *Let ζ_2 and $b \in W^{1,\infty}(\mathbb{R}^d)$ such that Ω_2 satisfies (1.10), and suppose that $\nabla\psi_2 \in H^{1/2}(\mathbb{R}^d)$. Then, with ϕ_2 the unique solution in $H^2(\Omega_2)$ of the boundary value problem*

$$\left\{\begin{array}{ll} \Delta^{\mu}_{X,z}\phi_2 = 0 & \text{in } \Omega_2, \\ \phi_2 = \psi_2 & \text{on } \Gamma_2 := \{z = \epsilon_2\zeta_2\}, \\ \partial_n\phi_2 = 0 & \text{on } \Gamma_b := \{z = -\frac{1}{\delta} + \beta b\}, \end{array}\right. \tag{1.11}$$

we define $G_2^{\mu,\delta}[\epsilon_2\zeta_2, \beta b]\psi_2 \in H^{1/2}(\mathbb{R}^d)$ by

$$G_2^{\mu,\delta}[\epsilon_2\zeta_2, \beta b]\psi_2 := -\mu\epsilon_2\nabla\zeta_2 \cdot \nabla\phi_{2|z=\epsilon_2\zeta_2} + \partial_z\phi_{2|z=\epsilon_2\zeta_2}.$$

DEFINITION 1.4. *Let now ζ_1, ζ_2, and $b \in W^{1,\infty}(\mathbb{R}^d)$ be such that Ω_1 and Ω_2 satisfy (1.10), and suppose $\nabla\psi_1$, $\nabla\psi_2 \in H^{1/2}(\mathbb{R}^d)$. Let ϕ_1 be the unique solution in $H^2(\Omega_2)$ of the boundary value problem*

$$\left\{\begin{array}{ll} \Delta^{\mu}_{X,z}\phi_1 = 0 & \text{in } \Omega_1, \\ \phi_1 = \psi_1 & \text{on } \Gamma_1 := \{z = 1 + \epsilon_1\zeta_1\}, \\ \partial_n\phi_1 = \frac{1}{\sqrt{1+\mu\epsilon_2^2|\nabla\zeta_2|^2}} G_2^{\mu,\delta}[\epsilon_2\zeta_2, \beta b]\psi_2 & \text{on } \Gamma_2. \end{array}\right. \tag{1.12}$$

Then we define $G_1^{\mu,\delta}[\epsilon_1\zeta_1, \epsilon_2\zeta_2, \beta b](\psi_1, \psi_2) \in H^{1/2}(\mathbb{R}^d)$ by

$$G_1^{\mu,\delta}[\epsilon_1\zeta_1, \epsilon_2\zeta_2, \beta b](\psi_1, \psi_2) := -\mu\epsilon_1\nabla\zeta_1 \cdot \nabla\phi_{1|z=1+\epsilon_1\zeta_1} + \partial_z\phi_{1|z=1+\epsilon_1\zeta_1}$$

and $H^{\mu,\delta}[\epsilon_1\zeta_1, \epsilon_2\zeta_2, \beta b](\psi_1, \psi_2) \in H^{1/2}(\mathbb{R}^d)$ by

$$H^{\mu,\delta}[\epsilon_1\zeta_1, \epsilon_2\zeta_2, \beta b](\psi_1, \psi_2) = \nabla\big(\phi_{1|z=\epsilon_2\zeta_2}\big).$$

In the following, when there is no possibility of mistake, we simply write

$$\begin{aligned} G_2\psi_2 &:= G_2^{\mu,\delta}[\epsilon_2\zeta_2, \beta b]\psi_2, \\ G_1(\psi_1, \psi_2) &:= G_1^{\mu,\delta}[\epsilon_1\zeta_1, \epsilon_2\zeta_2, \beta b](\psi_1, \psi_2), \\ H(\psi_1, \psi_2) &:= H^{\mu,\delta}[\epsilon_1\zeta_1, \epsilon_2\zeta_2, \beta b](\psi_1, \psi_2). \end{aligned}$$

REMARK 1.5. The existence and uniqueness of such solutions ϕ_2 and ϕ_1 are given by Proposition 2.1.

Using these last definitions, it is straightforward to check that the system (1.9) becomes in dimensionless variables (where we omit the tildes for the sake of clarity)

$$\left\{\begin{array}{l} \alpha\partial_t\zeta_1 - \frac{1}{\mu}G_1(\psi_1, \psi_2) = 0, \\ \partial_t\zeta_2 - \frac{1}{\mu}G_2\psi_2 = 0, \\ \partial_t\nabla\psi_1 + \alpha\nabla\zeta_1 + \frac{\epsilon_2}{2}\nabla(|\nabla\psi_1|^2) = \mu\epsilon_2\nabla\mathcal{N}_1, \\ \partial_t(\nabla\psi_2 - \gamma H(\psi_1, \psi_2)) + (1-\gamma)\nabla\zeta_2 + \frac{\epsilon_2}{2}\nabla(|\nabla\psi_2|^2 - \gamma|H(\psi_1, \psi_2)|^2) = \mu\epsilon_2\nabla\mathcal{N}_2, \end{array}\right. \tag{1.13}$$

where

$$\begin{aligned} \mathcal{N}_1 &:= \frac{\Big(\frac{1}{\mu}G_1(\psi_1, \psi_2) + \epsilon_1\nabla\zeta_1 \cdot \nabla\psi_1\Big)^2}{2(1 + \mu|\epsilon_1\nabla\zeta_1|^2)}, \\ \mathcal{N}_2 &:= \frac{\Big(\frac{1}{\mu}G_2\psi_2 + \epsilon_2\nabla\zeta_2 \cdot \nabla\psi_2\Big)^2 - \gamma\Big(\frac{1}{\mu}G_2\psi_2 + \epsilon_2\nabla\zeta_2 \cdot H(\psi_1, \psi_2)\Big)^2}{2(1 + \mu|\epsilon_2\nabla\zeta_2|^2)}. \end{aligned}$$

We derive the asymptotic models from this system of nondimensionalized equations, corresponding to different sizes for the dimensionless parameters.

1.6. The linearized equation. Linearizing the system (1.13) around the rest state, we obtain

$$\begin{cases} \alpha\partial_t\zeta_1 - \frac{1}{\mu}G_1^{\mu,\delta}[0,0,0](\psi_1,\psi_2) = 0, \\ \partial_t\zeta_2 - \frac{1}{\mu}G_2^{\mu,\delta}[0,0]\psi_2 = 0, \\ \partial_t\nabla\psi_1 + \alpha\nabla\zeta_1 = 0, \\ \partial_t(\nabla\psi_2 - \gamma H^{\mu,\delta}[0,0,0](\psi_1,\psi_2)) + (1-\gamma)\nabla\zeta_2 = 0. \end{cases} \tag{1.14}$$

Now when the surface, the interface, and the bottom are flat, we have explicit expressions for the operators G_1, G_2, and H. Indeed, taking the horizontal Fourier transform of the Laplace equations in (1.11) and (1.12), we obtain that $\widehat{\phi_2}$ and $\widehat{\phi_1}$ are solutions of the following ordinary differential equations:

$$-\mu|D|^2 y + y'' = 0.$$

Then, using the boundary conditions, we deduce

$$\phi_2(X,z) = \cosh(\sqrt{\mu}|D|z)\psi_2(X) + \tanh\left(\frac{\sqrt{\mu}}{\delta}|D|\right)\sinh(\sqrt{\mu}|D|z)\psi_2(X)$$

so that we have

$$G_2^{\mu,\delta}[0,0]\psi_2 = \sqrt{\mu}|D|\tanh\left(\frac{\sqrt{\mu}}{\delta}|D|\right)\psi_2.$$

Then we obtain

$$\begin{aligned}\phi_1(X,z) = \cosh(\sqrt{\mu}|D|z)&\left(\frac{1}{\cosh(\sqrt{\mu}|D|)}\psi_1(X) - \tanh\left(\frac{\sqrt{\mu}}{\delta}|D|\right)\tanh(\sqrt{\mu}|D|)\psi_2(X)\right) \\ &+ \tanh\left(\frac{\sqrt{\mu}}{\delta}|D|\right)\sinh(\sqrt{\mu}|D|z)\psi_2(X)\end{aligned}$$

so that we have

$$G_1^{\mu,\delta}[0,0,0](\psi_1,\psi_2) = \frac{\sqrt{\mu}|D|}{\cosh(\sqrt{\mu}|D|)}\left(\sinh(\sqrt{\mu}|D|)\psi_1 + \tanh\left(\frac{\sqrt{\mu}}{\delta}\right)\psi_2\right)$$

and finally

$$H^{\mu,\delta}[0,0,0](\psi_1,\psi_2) = \frac{1}{\cosh(\sqrt{\mu}|D|)}\nabla\psi_1(X) - \tanh\left(\frac{\sqrt{\mu}}{\delta}|D|\right)\tanh(\sqrt{\mu}|D|)\nabla\psi_2(X).$$

Using these expressions in the system (1.14), we can easily calculate the dispersion relations. Indeed, the wave frequencies $\omega_\pm^2(k)$, corresponding to plane-wave solutions $e^{ik\cdot X - i\omega_\pm t}$, are the solutions of the quadratic equation

$$\omega^4 - \frac{|k|}{\sqrt{\mu}}\frac{\tanh(\sqrt{\mu}|k|) + \tanh\left(\frac{\sqrt{\mu}}{\delta}|k|\right)}{1+\gamma\tanh(\sqrt{\mu}|k|)\tanh\left(\frac{\sqrt{\mu}}{\delta}|k|\right)}\omega^2 + \frac{|k|^2}{\mu}\frac{(1-\gamma)\tanh(\sqrt{\mu}|k|)\tanh\left(\frac{\sqrt{\mu}}{\delta}|k|\right)}{1+\gamma\tanh(\sqrt{\mu}|k|)\tanh\left(\frac{\sqrt{\mu}}{\delta}|k|\right)} = 0. \tag{1.15}$$

This equation has two strictly positive solutions (and their opposites) if and only if $\gamma < 1$, corresponding to the case wherein the lower fluid is heavier than the upper one. This expression also appears in **[35]** and **[99]**. Figure 2 shows the evolution of the wave frequencies $\omega_\pm$, $-\omega_\pm$ as functions of the wave number k. We chose the parameters $\mu = 0.1$, $\delta = 1/3$, $\gamma = 2/3$.

Remark 1.6. We remark that setting $\gamma = 0$ and $\delta = 1$, we recover the classical dispersion relation for the one-fluid system

$$\omega^2 = \frac{|k|}{\sqrt{\mu}}\tanh(\sqrt{\mu}|k|).$$

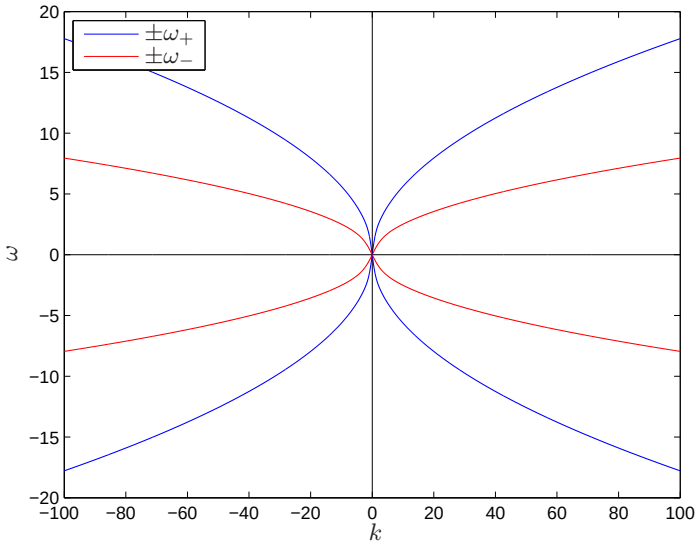


FIGURE 2. Full system dispersion.

2. Asymptotic models

We derive asymptotic models for the system by obtaining explicit expansions of the operators. Following the method of **[17]**, it is convenient to first reduce the problems (1.11) and (1.12) to elliptic equations on a flat strip.

2.1. Flattening of the domain. We define the mappings

$$R_1 := \begin{array}{ccc} \Omega_1 & \to & \mathcal{S}^+ \\ (X,z) & \mapsto & (X, r_1(X,z)) \end{array} \quad \text{with } r_1(X,z) := \frac{z-\epsilon_2\zeta_2(X)}{1+\epsilon_1\zeta_1(X)-\epsilon_2\zeta_2(X)},$$

$$R_2 := \begin{array}{ccc} \Omega_2 & \to & \mathcal{S}^- \\ (X,z) & \mapsto & (X, r_2(X,z)) \end{array} \quad \text{with } r_2(X,z) := \frac{z-\epsilon_2\zeta_2(X)}{1/\delta-\beta b(X)+\epsilon_2\zeta_2(X)},$$

and we denote their inverse

$$S_1 := \begin{array}{ccc} \mathcal{S}^+ & \to & \Omega_1 \\ (X,z) & \mapsto & (X, s_1(X,z)) \end{array} \quad \text{and} \quad S_2 := \begin{array}{ccc} \mathcal{S}^- & \to & \Omega_2 \\ (X,z) & \mapsto & (X, s_2(X,z)) \end{array}$$

with

$$\begin{aligned} s_1(X,z) &:= (1+\epsilon_1\zeta_1(X)-\epsilon_2\zeta_2(X))z+\epsilon_2\zeta_2(X), \\ s_2(X,z) &:= (1/\delta-\beta b(X)+\epsilon_2\zeta_2(X))z+\epsilon_2\zeta_2(X). \end{aligned}$$

Introducing the $(d+1)\times(d+1)$ matrices

$$\begin{aligned} P_i &:= \frac{1}{\partial_z s_i}\begin{pmatrix} \partial_z s_i I_d & 0_{d,1} \\ -\nabla_X s_i{}^T & 1 \end{pmatrix}\begin{pmatrix} \mu I_d & 0_{d,1} \\ 0_{1,d} & 1 \end{pmatrix}\begin{pmatrix} \partial_z s_i I_d & -\nabla_X s_i \\ 0_{1,d} & 1 \end{pmatrix} \\ &= \begin{pmatrix} \mu\partial_z s_i I_d & -\mu\nabla_X s_i \\ -\mu\nabla_X s_i{}^T & \frac{1+\mu|\nabla_X s_i|^2}{\partial_z s_i} \end{pmatrix}, \end{aligned} \tag{2.1}$$

where $0_{m,n}$ is the $m\times n$ zero matrix and I_d the $d\times d$ identity matrix, we can transform the Laplace equations (1.11) and (1.12) into elliptic boundary value problems on flat strips.

PROPOSITION 2.1. *Let ζ_1, ζ_2, and $b\in W^{1,\infty}(\mathbb{R}^d)$ such that Ω_1 and Ω_2 satisfy* (1.10)*, and suppose $\nabla\psi_1$, $\nabla\psi_2\in H^{1/2}(\mathbb{R}^d)$. Then there exists a unique solution $\phi_1\in H^2(\mathcal{S}^+)$ and*

$\phi_2 \in H^2(\mathcal{S}^-)$ to the boundary value problems

$$\left\{\begin{array}{ll} \nabla_{X,z}\cdot P_2\nabla_{X,z}\phi_2 = 0 & \text{in } \mathcal{S}^-,\\ \phi_2 = \psi_2 & \text{on } \{z=0\},\\ \partial_n\phi_2 = 0 & \text{on } \{z=-1\}\end{array}\right. \tag{2.2}$$

and

$$\left\{\begin{array}{ll} \nabla_{X,z}\cdot P_1\nabla_{X,z}\phi_1 = 0 & \text{in } \mathcal{S}^+,\\ \phi_1 = \psi_1 & \text{on } \{z=1\},\\ \partial_n\phi_1 = \partial_n\phi_2 & \text{on } \{z=0\},\end{array}\right. \tag{2.3}$$

where $\partial_n\phi$ stands for the upward conormal derivative associated to the elliptic operator involved:

$$\partial_n\phi := e_{d+1}\cdot P\nabla_{X,z}\phi.$$

Moreover, $\tilde{\phi}_i := (X,z)\mapsto \phi_i(X,r_i(X,z))$ $(i=1,2)$, respectively, solve the problems (1.12) *and* (1.11)*. Thus, the operators G_1, G_2, and H can equivalently be defined with*

$$\begin{aligned} G_2\psi_2 &= e_{d+1}\cdot P_2\nabla_{X,z}\phi_{2|z=0},\\ G_1(\psi_1,\psi_2) &= e_{d+1}\cdot P_1\nabla_{X,z}\phi_{1|z=1},\\ H(\psi_1,\psi_2) &= \nabla\big(\phi_{1|z=0}\big).\end{aligned}$$

Proof. The reduction of the problems (1.12) and (1.11) on the flat strip can be found on **[68]** (Proposition 2.7). The coercivity condition is satisfied thanks to (1.10) and the assumptions on ζ_1, ζ_2 (see Proposition 2.3 of **[71]**):

$$\exists k>0\ \forall\Theta\in\mathbb{R}^{d+1}, \Theta\cdot P_i\Theta \geq \frac{1}{k}\big|\Theta\big|^2. \tag{2.4}$$

Thus, we just prove here the existence and uniqueness of the H^2-solutions ϕ_i $(i=1,2)$. Since, for all $g\in H^{-1/2}(\mathbb{R}^d), h\in H^{1/2}(\mathbb{R}^d)$, we can easily construct a function $w\in H^1(\mathcal{S}^+)$ such that $w_{|z=1}=h$ and $\partial_n w_{|z=0}=g$, (2.3) and (2.2) clearly reduce to the problem

$$\left\{\begin{array}{ll} \nabla_{X,z}\cdot P\nabla_{X,z}\phi_1 = f & \text{in } \mathcal{S}^+,\\ \phi_1 = 0 & \text{on } \Gamma_1 := \mathbb{R}^d\times\{1\},\\ \partial_n\phi_1 = 0 & \text{on } \Gamma_2 := \mathbb{R}^d\times\{0\},\end{array}\right. \tag{2.5}$$

where $f\in H^{-1}(\mathcal{S}^+)$ and P satisfies (2.4).

As a first step, we introduce the variational formulation of this problem. Let us define the functional space

$$V := \{v\in H^1(\mathcal{S}^+), \gamma_0(v)=0 \text{ on } \mathbb{R}^d\}$$

with $\gamma_0 : H^1(\mathcal{S}^+)\to H^{1/2}(\mathbb{R}^d)$ the trace operator on Γ_1. Since γ_0 is continuous, V, equipped with the scalar product of $H^1(\mathcal{S}^+)$ and the corresponding norm, is a closed subspace of $H^1(\mathcal{S}^+)$ and hence a Hilbert space. A solution of the variational problem related to (2.5) is then a function $u\in V$ such that

$$\forall v\in V, \int_{\mathcal{S}^+} P\nabla u\cdot\nabla v = -\int_{\mathcal{S}^+} fv. \tag{2.6}$$

Since $\mathcal{V}=\{v\in\mathcal{D}(\bar{\mathcal{S}^+}), v=0 \text{ on } \Gamma_1\}$ is dense in V, a solution of the variational problem (2.6) is a weak solution of the problem (2.5).

Now we can check that $a(u,v) := \int_{\mathcal{S}^+} P\nabla u\cdot\nabla v$ is a continuous bilinear form. The coercivity of a is given by (2.4) and a generalized Poincaré inequality (see **[6]**, Theorem 5.4.3). Finally, since $b : v\in V\mapsto -\int_{\mathcal{S}^+} fv$ is clearly continuous, the Lax–Milgram theorem gives the existence and uniqueness of a solution $u\in V$ of (2.6) and thus a weak solution of (2.5). Moreover, we have

$$\big\|u\big\|_{H^1} \leq C\big\|f\big\|_{H^{-1}}.$$

The last step consists in proving that the solution u lives in $H^2(\mathcal{S}^+)$ if we assume that $f \in L^2$. We introduce for $h > 0$,

$$u_h := (x,y,z) \mapsto \frac{\tau_h u(x,y,z) - u(x,y,z)}{h} = \frac{u(x+h,y,z) - u(x,y,z)}{h}.$$

Then u_h is the solution (2.5) with $f_h = \frac{\tau_h f - f}{h}$ and $g_h = \frac{\tau_h g - g}{h}$ so that

$$\big\|u_h\big\|_{H^1} \leq C\big\|f_h\big\|_{H^{-1}}.$$

Then we remark that, for any $v \in H^1(\mathcal{S}^+)$, $v_h(x,y,z) = \frac{1}{h}\int_x^{x+h} \partial_x v(t,y,z)dt$ so that

$$\big\|v_h\big\|_{L^2} \leq \frac{1}{h}\int_0^h \big\|\partial_x v\big\|_{L^2} dt \leq \big\|v\big\|_{H^1}.$$

Thus, we have, thanks to the duality between H^1 and H^1_0,

$$\big\|f_h\big\|_{H^{-1}} \leq \sup_{v\in H^1_0} \frac{|(f_h,v)|}{\big\|v\big\|_{H^1}} \leq \sup_{v\in H^1_0} \frac{\big\|f\big\|_{L^2}\big\|v_h\big\|_{L^2}}{\big\|v\big\|_{H^1}} \leq \big\|f\big\|_{L^2}.$$

We finally have

$$\big\|u_h\big\|_{H^1} \leq C\big\|f\big\|_{L^2}.$$

Since V is a Hilbert space, we deduce that there exists $w \in V$ and a subsequence (u_{h_k}) such that u_{h_k} weakly converges toward w. Moreover, we know that u_{h_k} converges toward $\partial_x u$ in $\mathcal{D}'(\mathcal{S}^+)$, so we deduce $\partial_x u \in V \subset H^1$.

We prove in the same way that $\partial_y u \in H^1$ so that $\Delta_X u \in L^2$. Finally, thanks to (2.4), we have

$$|\partial_z^2 u| \leq |\Delta_X u| + k|\nabla_{X,z} \cdot P\nabla_{X,z} u| = |\Delta_X u| + k|f|$$

so that $u \in H^2(\mathcal{S}^+)$, and the proposition is proved. □

2.2. Asymptotic expansion of the operators. We are looking for shallow water models ($\mu \ll 1$) and therefore need to obtain an expansion of the operators in terms of μ. The method is the following. We first exhibit the expansion of the matrix P_i in terms of μ. Then we look for approximate solutions ϕ_i^{app} $(i = 1,2)$ under the form

$$\phi_i^{app} = \phi_i^0 + \mu\phi_i^1 + \mu^2\phi_i^2.$$

Plugging this ansatz into (2.2) and (2.3) and solving at each order of μ gives the ϕ_i^j, from which we can deduce the expansion of the operators by computing the normal derivative of ϕ_i^{app}.

Since (2.2) is exactly the same problem as involved (in the case of the water wave) in [**71**], we can directly apply Proposition 3.8 to the lower fluid.

PROPOSITION 2.2. *Let $t_0 > d/2$ and $s \geq t_0 + 1/2$, $\nabla\psi_2 \in H^{s+\frac{11}{2}}(\mathbb{R}^2)$, $\zeta_2 \in H^{s+\frac{9}{2}}(\mathbb{R}^2)$, and $b \in H^{s+\frac{11}{2}}(\mathbb{R}^2)$ such that (1.10) is satisfied. Then we have*

$$\big|G_2\psi_2 + \mu\nabla\cdot(h_2\nabla\psi_2)\big|_{H^s} \leq \mu^2 C_0, \tag{2.7}$$

$$\big|G_2\psi_2 + \mu\nabla\cdot(h_2\nabla\psi_2) - \mu^2\nabla\cdot\mathcal{T}[h_2,\beta b]\nabla\psi_2\big|_{H^s} \leq \mu^3 C_1, \tag{2.8}$$

with $C_j = C(\frac{1}{h}, \beta\big|b\big|_{H^{s+7/2+2j}}, \epsilon_2\big|\zeta_2\big|_{H^{s+5/2+2j}}, \big|\nabla\psi_2\big|_{H^{s+7/2+2j}})$, and where we denote by $h_2 := \frac{1}{\delta} - \beta b + \epsilon_2\zeta_2$ the thickness of the lower layer, and

$$\mathcal{T}[h,b]V := -\frac{1}{3}\nabla(h^3\nabla\cdot V) + \frac{1}{2}\big(\nabla(h^2\nabla b\cdot V) - h^2\nabla b\nabla\cdot V\big) + h\nabla b\nabla b\cdot V.$$

Remark 2.3. To obtain the estimate (2.7), we use the approximate solution

$$\phi_2^{app,1} = \psi_2 - \mu h_2 \left(h_2 \left(\frac{z^2}{2} + z \right) \Delta \psi_2 - z\beta \nabla b \cdot \nabla \psi_2 \right).$$

We need a higher order approximation to obtain (2.8), namely, $\phi_2^{app,2} = \phi_2^{app,1} + \mu^2 \phi_2^2$, where ϕ_2^2 can be obtained using the same method as in the following study. Proposition 2.2 is then obtained following the path of Appendix A for the lower fluid (see **[26]** for a rigorous proof).

The study of the upper fluid is different from the one of the lower fluid since we have now a nonhomogeneous Neumann condition on the interface. In order to manage this, we first decompose $\phi_1 := \check{\phi}_1 + \bar{\phi}_1$, where $\check{\phi}_1$ is the unique solution of

$$\left\{ \begin{array}{ll} \nabla_{X,z} \cdot P_1 \nabla_{X,z} \check{\phi}_1 = 0 & \text{in } \mathcal{S}^+, \\ \check{\phi}_1 = \psi_1 & \text{on } \{z = 1\}, \\ \partial_n \check{\phi}_1 = 0 & \text{on } \{z = 0\} \end{array} \right. \tag{2.9}$$

and $\bar{\phi}_1$ is the unique solution of

$$\left\{ \begin{array}{ll} \nabla_{X,z} \cdot P_1 \nabla_{X,z} \bar{\phi}_1 = 0 & \text{in } \mathcal{S}^+, \\ \bar{\phi}_1 = 0 & \text{on } \{z = 1\}, \\ \partial_n \bar{\phi}_1 = G_2 \psi_2 & \text{on } \{z = 0\}. \end{array} \right. \tag{2.10}$$

Again the system satisfied by $\check{\phi}_1$ reduces to the water-wave problem (where the topography of the bottom would be given by $\epsilon_2 \zeta_2$), so we introduce as in Remark 2.3 the approximate solutions

$$\begin{array}{rcl} \check{\phi}_1^{app,1} & := & \psi_1 - \mu h_1 \left(h_1 \left(\dfrac{(z-1)^2}{2} + (z-1) \right) \Delta \psi_1 - (z-1) \epsilon_2 \nabla \zeta_2 \cdot \nabla \psi_1 \right), \\ \check{\phi}_1^{app,2} & := & \check{\phi}_1^{app,1} + \mu^2 \check{\phi}_1^2. \end{array}$$

It follows that $\check{G}_1 \psi_1$ the contribution on the Dirichlet–Neumann operator from $\check{\phi}_1$ can be expanded as in the following proposition.

Proposition 2.4. *Let $t_0 > d/2$ and $s \geq t_0 + 1/2$, $\nabla \psi_1, \zeta_2 \in H^{s+11/2}(\mathbb{R}^2)$, $\zeta_1 \in H^{s+9/2}(\mathbb{R}^2)$ such that* (1.10) *is satisfied. Then we have*

$$\left| \check{G}_1 \psi_1 + \mu \nabla \cdot (h_1 \nabla \psi_1) \right|_{H^s} \leq \mu^2 C_0, \tag{2.11}$$

$$\left| \check{G}_1 \psi_1 + \mu \nabla \cdot (h_1 \nabla \psi_1) - \mu^2 \nabla \cdot \mathcal{T}[h_1, \epsilon_2 \zeta_2] \nabla \psi_1 \right|_{H^s} \leq \mu^3 C_1 \tag{2.12}$$

with $C_j = C(\frac{1}{h}, \epsilon_2 |\zeta_2|_{H^{s+7/2+2j}}, \epsilon_1 |\zeta_1|_{H^{s+5/2+2j}}, |\nabla \psi_1|_{H^{s+7/2+2j}})$, and where we denote by $h_1 := 1 + \epsilon_1 \zeta_1 - \epsilon_2 \zeta_2$ the thickness of the upper layer, and $\mathcal{T}[h, b]V$ is defined as in Proposition 2.2.

The last step consists in computing the contribution on the Dirichlet–Neumann operator from $\bar{\phi}_1$. We first define $\bar{\phi}_1^{app} = \phi^0 + \mu \phi^1 + \mu^2 \phi^2$. It is straightforward that

$$P_1 = P^0 + \mu P^1 \text{ with } P^0 := \begin{pmatrix} 0_{d,d} & 0_{d,1} \\ 0_{1,d} & \frac{1}{h_1} \end{pmatrix} \text{ and } P^1 := \begin{pmatrix} h_1 I_d & -\nabla_X s_1 \\ -\nabla_X s_1{}^T & \frac{|\nabla_X s_1|^2}{h_1} \end{pmatrix},$$

where we have used the notations $0_{m,n}$ for the $m \times n$ zero matrix and I_d for the $d \times d$ identity matrix. Plugging these expansions into (2.10), using Proposition 2.2, and solving at each order, we get the following:

At order $O(1)$:

$$\left\{ \begin{array}{ll} \frac{1}{h_1} \partial_z^2 \phi^0 = 0 & \text{in } \mathcal{S}^+, \\ \phi^0 = 0 & \text{on } \{z = 1\}, \\ \frac{1}{h_1} \partial_z \phi^0 = 0 & \text{on } \{z = 0\} \end{array} \right.$$

so that we have

$$\phi^0 = 0. \tag{2.13}$$

At order $O(\mu)$:

$$\begin{cases} \frac{1}{h_1}\partial_z^2\phi^1 = -\nabla_{X,z}\cdot P^1\nabla_{X,z}\phi^0 = 0 & \text{in } \mathcal{S}^+, \\ \phi^1 = 0 & \text{on } \{z=1\}, \\ \frac{1}{h_1}\partial_z\phi^1 = -e_{d+1}\cdot P^1\nabla_{X,z}\phi^0 - \nabla\cdot(h_2\nabla\psi_2) & \text{on } \{z=0\}, \end{cases}$$

which gives immediately

$$\phi^1 = -h_1\nabla\cdot(h_2\nabla\psi_2)(z-1). \tag{2.14}$$

At order $O(\mu^2)$:

$$\begin{cases} \frac{1}{h_1}\partial_z^2\phi^2 = h_1\big((z-1)h_1\nabla\cdot\nabla\mathcal{A}_2 - 2\epsilon_1\nabla\zeta_1\cdot\nabla\mathcal{A}_2 - \epsilon_1\Delta\zeta_1\mathcal{A}_2\big) & \text{in } \mathcal{S}^+, \\ \phi^2 = 0 & \text{on } \{z=1\}, \\ \frac{1}{h_1}\partial_z\phi^2 = \nabla\cdot\mathcal{T}[h_2,\beta b]\nabla\psi_2 + \epsilon_2\nabla\zeta_2\cdot(h_1\nabla\mathcal{A}_2 + \epsilon_1\nabla\zeta_1\mathcal{A}_2) & \text{on } \{z=0\} \end{cases}$$

with the notation $\mathcal{A}_2 := \nabla\cdot(h_2\nabla\psi_2)$. This leads to the solution

$$\begin{aligned} \phi^2 = \quad & h_1\bigg((h_1^2\nabla\cdot\nabla\mathcal{A}_2)\left(\frac{z^3}{6} - \frac{z^2}{2} + \frac{1}{3}\right) - h_1(2\epsilon_1\nabla\zeta_1\cdot\nabla\mathcal{A}_2 + \epsilon_1\Delta\zeta_1\mathcal{A}_2)\left(\frac{z^2}{2} - \frac{1}{2}\right) \\ & + (\nabla\cdot\mathcal{T}[h_2,\beta b]\nabla\psi_2 + \epsilon_2\nabla\zeta_2\cdot(h_1\nabla\mathcal{A}_2 + \epsilon_1\nabla\zeta_1\mathcal{A}_2))(z-1)\bigg). \end{aligned} \tag{2.15}$$

This formal derivation of $\bar{\phi}_1^{app}$ allows us to obtain the expansion of $\bar{G}_1\psi_2$, the contribution on the Dirichlet–Neumann operator from $\bar{\phi}_1$. Formally we have

$$\bar{G}_1\psi_2 \approx -\mu\mathcal{A}_2 + \mu^2\Big(\nabla\cdot\mathcal{T}[h_2,\beta b]\nabla\psi_2 - \frac{1}{2}\nabla\cdot(h_1^2\nabla\mathcal{A}_2) - \nabla\cdot(h_1\epsilon_1\nabla\zeta_1\mathcal{A}_2)\Big). \tag{2.16}$$

Summing this expansion with the one of Proposition 2.4 gives immediately the expansion of the full operator $G_1(\psi_1,\psi_2)$. The following proposition gives a rigorous statement of this fact; its proof is postponed to Appendix A.

PROPOSITION 2.5. *Let $t_0 > d/2$ and $s \geq t_0 + 1/2$, $\nabla\psi_1$, $\nabla\psi_2 \in H^{s+11/2}(\mathbb{R}^2)$, $\zeta_1 \in H^{s+7/2}(\mathbb{R}^2)$, $\zeta_2 \in H^{s+9/2}(\mathbb{R}^2)$, and $b \in H^{s+11/2}(\mathbb{R}^2)$ such that (1.10) is satisfied. Then we have*

$$\big|G_1(\psi_1,\psi_2) + \mu(\mathcal{A}_1+\mathcal{A}_2)\big|_{H^s} \leq \mu^2 C_0, \tag{2.17}$$

$$\begin{aligned} \big|G_1(\psi_1,\psi_2) + \mu(\mathcal{A}_1+\mathcal{A}_2) - \mu^2\Big(\nabla\cdot\mathcal{T}_1 + \nabla\cdot\mathcal{T}_2 - \frac{1}{2}\nabla\cdot(h_1^2\nabla\mathcal{A}_2) \\ - \nabla\cdot(h_1\epsilon_1\nabla\zeta_1\mathcal{A}_2)\Big)\big|_{H^s} \leq \mu^3 C_1 \end{aligned} \tag{2.18}$$

with the constants

$$\begin{aligned} C_0 &= C\left(\tfrac{1}{h}, \beta|b|_{H^{s+7/2}}, \epsilon_2|\zeta_2|_{H^{s+5/2}}, \epsilon_1|\zeta_1|_{H^{s+3/2}}, |\nabla\psi_1|_{H^{s+7/2}}, |\nabla\psi_2|_{H^{s+7/2}}\right), \\ C_1 &= C\left(\tfrac{1}{h}, \beta|b|_{H^{s+11/2}}, \epsilon_2|\zeta_2|_{H^{s+9/2}}, \epsilon_1|\zeta_1|_{H^{s+7/2}}, |\nabla\psi_1|_{H^{s+11/2}}, |\nabla\psi_2|_{H^{s+11/2}}\right), \end{aligned}$$

and the notations

$$\begin{aligned} \mathcal{A}_1 &:= \nabla\cdot(h_1\nabla\psi_1), \quad & \mathcal{A}_2 &:= \nabla\cdot(h_2\nabla\psi_2), \\ \mathcal{T}_1 &:= \mathcal{T}[h_1,\epsilon_2\zeta_2]\nabla\psi_1, \quad & \mathcal{T}_2 &:= \mathcal{T}[h_2,\beta b]\nabla\psi_2. \end{aligned}$$

REMARK 2.6. As in Remark 2.3, the proof of the estimate (2.17) requires the approximate solution $\phi_1^{app,1}$ with

$$\phi_1^{app,1} := \check{\phi}_1^{app,1} + \phi^0 + \mu\phi^1,$$

and the second estimate (2.18) uses

$$\phi_1^{app,2} := \check{\phi}_1^{app,2} + \phi^0 + \mu\phi^1 + \mu^2\phi^2.$$

In Appendix A (Steps 4 and 5), we give estimates on $\phi_1 - \phi_1^{app}$, obtained thanks to the trace theorem and an elliptic estimate on the boundary value problem solved by $\phi_1 - \phi_1^{app}$. This leads to the desired inequalities since

$$G_1(\psi_1,\psi_2) - \partial_n \phi_1^{app}{}_{|z=1} = \partial_n(\phi_1 - \phi_1^{app})_{|z=1}.$$

The last expansion to obtain is the one of $H(\psi_1,\psi_2)$, which is given by the following.

Proposition 2.7. *Let $t_0 > d/2$ and $s \geq t_0 + 1/2$, $\nabla\psi_1$, $\nabla\psi_2 \in H^{s+11/2}(\mathbb{R}^2)$, $\zeta_1 \in H^{s+7/2}(\mathbb{R}^2)$, $\zeta_2 \in H^{s+9/2}(\mathbb{R}^2)$, and $b \in H^{s+11/2}(\mathbb{R}^2)$ such that* (1.10) *is satisfied. Then we have*

$$\big|H(\psi_1,\psi_2) - \nabla\psi_1\big|_{H^s} \leq \mu C_0, \tag{2.19}$$

$$\big|H(\psi_1,\psi_2) - \nabla\psi_1 - \mu\nabla\Big(h_1(\mathcal{A}_1+\mathcal{A}_2) - \frac{1}{2}h_1^2\Delta\psi_1 - h_1\epsilon_1\nabla\zeta_1\cdot\nabla\psi_1\Big)\big|_{H^s} \leq \mu^2 C_1 \tag{2.20}$$

with

$$C_0 = C\left(\tfrac{1}{h}, \beta\big|b\big|_{H^{s+7/2}}, \epsilon_2\big|\zeta_2\big|_{H^{s+5/2}}, \epsilon_1\big|\zeta_1\big|_{H^{s+3/2}}, \big|\nabla\psi_1\big|_{H^{s+7/2}}, \big|\nabla\psi_2\big|_{H^{s+7/2}}\right),$$
$$C_1 = C\left(\tfrac{1}{h}, \beta\big|b\big|_{H^{s+11/2}}, \epsilon_2\big|\zeta_2\big|_{H^{s+9/2}}, \epsilon_1\big|\zeta_1\big|_{H^{s+7/2}}, \big|\nabla\psi_1\big|_{H^{s+11/2}}, \big|\nabla\psi_2\big|_{H^{s+11/2}}\right),$$

and using the notations of Proposition 2.5.

Proof. The proof uses the estimates (A.7) and (A.9) on $u := \phi_1 - \phi_1^{app,1}$. Indeed, we have to give an estimate for $\big|\nabla u_{|z=0}\big|_{H^s}$, and a trace theorem (see Métivier **[83]** pp. 23–27) gives, for all $s \geq 0$,

$$\big|\nabla u_{|z=0}\big|_{H^s} \leq \mathrm{Cst}(\big\|\Lambda^{s+1/2}\nabla u\big\|_{L^2} + \big\|\Lambda^{s-1/2}\partial_z\nabla u\big\|)_{L^2} \leq \frac{\mathrm{Cst}}{\sqrt{\mu}}\big\|\Lambda^{s+1/2}\nabla^{\mu}_{X,z}u\big\|_{L^2}.$$

Then the estimate (A.7) allows us to conclude:

$$\big|\nabla u_{|z=0}\big|_{H^s} \leq \tfrac{C_{s,t_0}}{\sqrt{\mu}}\Big(\tfrac{1}{h}, \epsilon_1\big|\zeta_1\big|_{H^{\max\{t_0+2,s+3/2\}}}, \epsilon_2\big|\zeta_2\big|_{H^{\max\{t_0+2,s+3/2\}}})\big)(\mu^2\big\|\mathbf{h}\big\|_{H^{s+1/2}} + \tfrac{1+\sqrt{\mu}}{\sqrt{\mu}}\big|V\big|_{H^{s+1}}\Big).$$

The first estimate (2.19) follows from this relation, together with the estimates (A.2) and (A.4).

As for the Proposition 2.5, the second estimate (2.20) requires the use of the higher order approximate solution $\tilde{u} := \phi_1 - \phi_1^{app,2}$, and the result is obtained in the same way. □

Remark 2.8. Using the same approximate solution as for the expansion of $G_1(\psi_1,\psi_2)$, we obtain an estimate one order less precise in μ than in (2.17) and (2.18). This loss of precision is not seen at the formal level and comes from the $\frac{1}{\sqrt{\mu}}$ term due to the horizontal scaling, which is necessary in order to have a uniformly elliptic operator.

2.3. Asymptotic models. The expansions of the operators we obtained allow us to derive asymptotic models from (1.13). The frame of this study is limited to shallow water/shallow water regimes, that is to say, long waves and layers of similar depth ($\mu \ll 1$, and $\delta \sim 1$). However, the method could be extended to many different regimes, as it has been done in **[17]** with the rigid-lid assumption. As we see in section 4, we recover most of the models which have been introduced in the literature, as well as interesting new ones (the Boussinesq/Boussinesq model with coefficients (2.27) and the higher order system (2.29)). Furthermore, we show rigorously that (1.13) is consistent with all of these models in the following sense (see **[16]**).

DEFINITION 2.9. *The internal-wave system* (1.13) *is consistent with a system* S *of* $2d+2$ *equations, if any sufficiently smooth solution of* (1.13) *such that* (1.10) *is satisfied solves* S *up to a small residual called the precision of the asymptotic model. Throughout this chapter, the precision is given in the sense of* $L^\infty H^s$ *norms, which means that the* H^s *norm of the residual is uniformly bounded with respect to* t *where the solution is defined.*

REMARK 2.10. The consistency does not require the well-posedness of (1.13) and concerns only the properties of smooth solutions of the system. However, if we assume the existence of such functions, we can prove that they are approximated by the solutions of consistent systems, as we see in section 3.

2.3.1. *The shallow water/shallow water regime:* $\mu \ll 1$. We assume here that both layers are in the shallow water regime ($\mu \ll 1$), whereas strong nonlinearity is allowed ($\epsilon_1, \epsilon_2 = O(1)$). We use the first order expansions (2.7), (2.17), and (2.19), and we plug them into (1.13). We obtain, discarding the $O(\mu)$ terms, the following system:

$$\left\{\begin{array}{l} \alpha\partial_t\zeta_1 + \nabla\cdot(h_1\nabla\psi_1) + \nabla\cdot(h_2\nabla\psi_2) = 0,\\ \partial_t\zeta_2 + \nabla\cdot(h_2\nabla\psi_2) = 0,\\ \partial_t\nabla\psi_1 + \alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_1|^2\right) = 0,\\ \partial_t\nabla\psi_2 + (1-\gamma)\nabla\zeta_2 + \gamma\alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_2|^2\right) = 0, \end{array}\right. \tag{2.21}$$

where $h_1 = 1 + \epsilon_1\zeta_1 - \epsilon_2\zeta_2$ and $h_2 = \frac{1}{\delta} - \beta b + \epsilon_2\zeta_2$.

REMARK 2.11. This system has already been introduced in the flat bottom case in **[35]** and equivalently, although under a different form, in **[30]**. We say more about this in section 4.2.

PROPOSITION 2.12. *The full system* (1.13) *is consistent with* (2.21)*, at the precision* μC_0*, with*

$$C_0 = C\Big(\tfrac{1}{h}, \beta\big|b\big|_{W^{1,\infty}H^{s+7/2}}, \epsilon_2\big|\zeta_2\big|_{W^{1,\infty}H^{s+5/2}}, \epsilon_1\big|\zeta_1\big|_{W^{1,\infty}H^{s+3/2}}, \big|\nabla\psi_1\big|_{W^{1,\infty}H^{s+7/2}}, \big|\nabla\psi_2\big|_{W^{1,\infty}H^{s+7/2}}\Big).$$

Proof. Let $t_0 > d/2$ and $s \geq t_0 + 1/2$. Let $U := (\zeta_1, \zeta_2, \nabla\psi_1, \nabla\psi_2)$ be a solution of (1.13) such that (1.10) is satisfied and $U \in \mathbf{H}^s$. It is straightforward to check that we have

$$\left\{\begin{array}{l} \alpha\partial_t\zeta_1 + \nabla\cdot(h_1\nabla\psi_1) + \nabla\cdot(h_2\nabla\psi_2) = \nabla\cdot(h_1\nabla\psi_1) + \nabla\cdot(h_2\nabla\psi_2) + \frac{1}{\mu}G_1(\psi_1,\psi_2),\\ \partial_t\zeta_2 + \nabla\cdot(h_2\nabla\psi_2) = \nabla\cdot(h_2\nabla\psi_2) + \frac{1}{\mu}G_2\psi_2,\\ \partial_t\nabla\psi_1 + \alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_1|^2\right) = \mu\epsilon_2\nabla\mathcal{N}_1,\\ \partial_t\nabla\psi_2 + (1-\gamma)\nabla\zeta_2 + \gamma\alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_2|^2\right) = \gamma\partial_t(H(\psi_1,\psi_2) - \nabla\psi_1)\\ \qquad\qquad + \frac{\epsilon_2}{2}\gamma\nabla(|H(\psi_1,\psi_2)|^2 - |\nabla\psi_1|^2) + \mu\epsilon_2\nabla\mathcal{N}_2 + \mu\gamma\epsilon_2\nabla\mathcal{N}_1. \end{array}\right. \tag{2.22}$$

Except for $\partial_t(H(\psi_1,\psi_2) - \nabla\psi_1)$, the right-hand side is immediately bounded by μC_0, thanks to the estimates (2.7), (2.17), and (2.19). The estimate on the derivative is obtained as in the following.

We use the study of Appendix A; we derive (A.1) with respect to t on both sides and get

$$\left\{\begin{array}{ll} \nabla^\mu_{X,z}\cdot P^\mu\nabla^\mu_{X,z}(\partial_t u) = \mu^2\ \nabla^\mu_{X,z}\cdot\partial_t\mathbf{h} - \nabla^\mu_{X,z}\cdot\partial_t(P^\mu)\nabla^\mu_{X,z}u & \text{in } \mathcal{S}^+,\\ \partial_t u = 0 & \text{on } \{z=1\},\\ \partial_n(\partial_t u) = \nabla\cdot\partial_t V + \mu^2 e_{d+1}\cdot\partial_t\mathbf{h} - e_{d+1}\cdot\partial_t(P^\mu)\nabla^\mu_{X,z}u & \text{on } \{z=0\}. \end{array}\right. \tag{2.23}$$

We now need estimates on the right-hand side of the system. Directly from the definition of $\mathbf{h}$, we have

$$\big\|\partial_t \mathbf{h}\big\|_{H^{s+3/2,1}} \leq C_0. \tag{2.24}$$

Thanks to Step 4 of section A.2, we have

$$\big\|\partial_t (P^\mu)\nabla^\mu_{X,z} u\big\|_{H^{s+3/2,1}} \leq C_0.$$

Finally we can obtain the estimate on $\partial_t V$, using the same method as here on the lower layer:

$$\big|\partial_t V\big|_{H^s} \leq \mu^2 C\left(\frac{1}{h}, \beta|b|_{W^{1,\infty}H^{s+5/2}}, \big|(\epsilon_1\zeta_1, \epsilon_2\zeta_2, \nabla\psi_1, \nabla\psi_2)\big|_{\mathbf{H}^{s-1}}\right).$$

Then we use the study of Appendix A and obtain the estimates of Steps 4 and 5 for $\partial_t u$, and we use them as in Proposition 2.7 in order to obtain the desired inequality:

$$\big|\partial_t(H(\psi_1,\psi_2) - \nabla\psi_1)\big|_{H^s} = \big|\nabla\partial_t u\big|_{H^s} \leq \mu C_0. \qquad \square$$

Conservation laws. The first two equations of (2.21) reveal the conservation of mass since a straightforward linear combination gives

$$\begin{cases} \partial_t h_1 + \epsilon_2 \nabla\cdot(h_1\nabla\psi_1) = 0, \\ \partial_t h_2 + \epsilon_2 \nabla\cdot(h_2\nabla\psi_2) = 0. \end{cases} \tag{2.25}$$

We can play with the system to obtain other conservation laws. The conservations of total momentum and energy are given by

$$\begin{aligned} \partial_t(\gamma h_1 u_1 + h_2 u_2) + \nabla p + (\gamma h_1 + h_2)\beta\nabla b + \nabla\cdot(\gamma h_1 u_1\otimes u_1 + h_2 u_2\otimes u_2) &= 0, \\ \partial_t\Big(\frac{1}{2}\big(\gamma h_1|u_1|^2 + h_2|u_2|^2\big) + p\Big) + \frac{1}{2}\nabla\cdot(\gamma h_1|u_1|^2 u_1 + h_2|u_2|^2 u_2) & \\ +\nabla\cdot(\gamma h_1^2 u_1 + h_2^2 u_2 + \gamma h_1 h_2(u_1+u_2)) + (\gamma h_1 u_1 + h_2 u_2)\beta\nabla b &= 0, \end{aligned}$$

with the notations $h_1 = 1 + \epsilon_1\zeta_1 - \epsilon_2\zeta_2$, $h_2 = \frac{1}{\delta} - \beta b + \epsilon_2\zeta_2$, $u_i = \epsilon_2\nabla\psi_i$ $(i = 1, 2)$, and the "pressure" $p := \frac{1}{2}\gamma h_1^2 + \frac{1}{2}h_2^2 + \gamma h_1 h_2$.

Dispersion relations. When we calculate the linearized dispersion relations as in section 1.6, we obtain that $\omega_\pm^2(k)$ satisfy

$$\omega_\pm^2(k) = \frac{1+\delta \pm \sqrt{(1-\delta)^2 + 4\gamma\delta}}{2\delta}|k|^2.$$

This dispersion relation is not the same as the one of the full system (it corresponds to the first order of the expansion in μ of the solutions of (1.15)), but we still have the condition $\gamma < 1$ for the system to be linearly well-posed. Figure 3 presents shallow water/shallow water model dispersion, compared with the dispersion of the full system, with the parameters $\mu = 0.1$, $\delta = 1/3$, $\gamma = 2/3$.

2.3.2. *The Boussinesq/Boussinesq regime:* $\mu \sim \epsilon_2 \sim \epsilon_1 \ll 1$. In this regime, the shallowness and the nonlinearity are supposed to be small and of the same size. This time, we use the second order of the expansions and obtain

$$\begin{cases} \alpha\partial_t\zeta_1 + \nabla\cdot(h_1\nabla\psi_1) + \nabla\cdot(h_2\nabla\psi_2) = \mu\big(\frac{-1}{3}\Delta\nabla\cdot\nabla\psi_1 + \nabla\cdot\mathcal{T}[h_2,\beta b]\nabla\psi_2 \\ \qquad\qquad -\frac{1}{2\delta}\Delta\nabla\cdot\nabla\psi_2\big), \\ \partial_t\zeta_2 + \nabla\cdot(h_2\nabla\psi_2) = \mu\left(\nabla\cdot\mathcal{T}[h_2,\beta b]\nabla\psi_2\right), \\ \partial_t\nabla\psi_1 + \alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_1|^2\right) = 0, \\ \partial_t\nabla\psi_2 + (1-\gamma)\nabla\zeta_2 + \alpha\gamma\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_2|^2\right) = \mu\gamma\partial_t\big(\frac{1}{\delta}\nabla\Delta\psi_2 + \frac{1}{2}\nabla\Delta\psi_1\big) \end{cases} \tag{2.26}$$

with $\mathcal{T}[h,b]V$ defined as in Proposition 2.2.

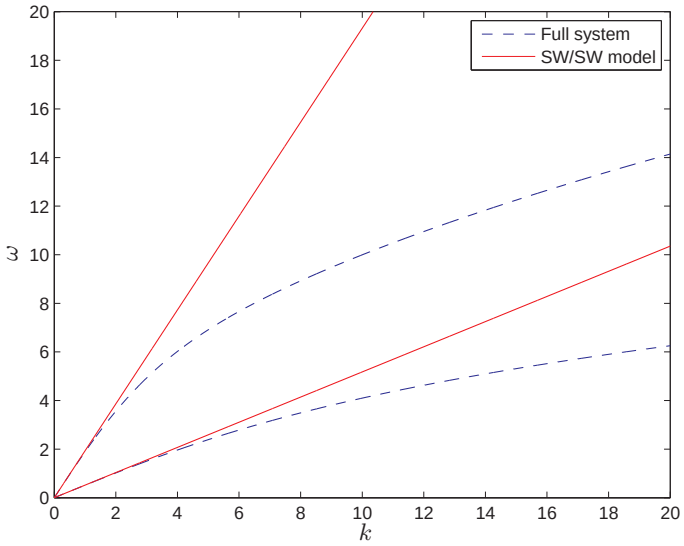


FIGURE 3. The shallow water/shallow water model dispersion.

REMARK 2.13. If the bottom is flat, then $\mathcal{T}[h_2, \beta b]\nabla\psi_2$ is simply $\frac{-1}{3\delta^3}\nabla\Delta\psi_2$.

MODEL WITH IMPROVED FREQUENCY DISPERSION. This model is linearly ill-posed. Fortunately, following **[16, 14]**, we can easily derive asymptotically equivalent models with coefficients which can be chosen so that the system is well-posed. For simplicity, we assume now to be in the case of a flat bottom (see **[26]** for the varying bottom case).

We rewrite the system (2.26) with new variables: $u_i := \nabla\big(\phi_i(\cdot, z_i)\big)$ $(i = 1, 2)$. From the calculations of section 2.2, we obtain

$$\begin{aligned}
\phi_1^{app,1}(\cdot, z) &= \psi_1 - \mu\left(\frac{(z-1)^2}{2} + (z-1)\right)\Delta\psi_1 - \mu\frac{1}{\delta}(z-1)\Delta\psi_2,\\
\phi_2^{app,1}(\cdot, z) &= \psi_2 - \mu\frac{1}{\delta^2}\left(\frac{z^2}{2} + z\right)\Delta\psi_2.
\end{aligned}$$

We then define u_1 and u_2 as in the following:

$$\begin{aligned}
u_1 &:= \nabla\big(\phi_1^{app,1}(\cdot, z_1)\big) = \nabla\psi_1 - \mu b_1\Delta\nabla\psi_1 - \mu\frac{1}{\delta}a_1\Delta\nabla\psi_2,\\
u_2 &:= \nabla\big(\phi_2^{app,1}(\cdot, z_2)\big) = \nabla\psi_2 - \mu\frac{1}{\delta^2}a_2\Delta\nabla\psi_2,
\end{aligned}$$

with $z_1 \in (0, 1)$ for the upper fluid, $z_2 \in (-1, 0)$ for the lower fluid, and the coefficients

$$a_1 := z_1 - 1 \in [-1, 0], \qquad a_2 := \frac{z_2^2}{2} + z_2 \in [-1/2, 0], \qquad b_1 := \frac{a_1^2}{2} + a_1 \in [-1/2, 0].$$

We plug this into (2.26) and obtain

$$
\left\{
\begin{array}{l}
\alpha\partial_t\zeta_1 + \nabla\cdot(h_1u_1) + \nabla\cdot(h_2u_2) + \mu\Big(\frac{1+3b_1}{3}\nabla\cdot\Delta u_1 \\
\qquad\qquad + \big(\frac{1+2a_1}{2\delta} + \frac{1+3a_2}{3\delta^3}\big)\nabla\cdot\Delta u_2\Big) = 0, \\
\partial_t\zeta_2 + \nabla\cdot(h_2u_2) + \mu\frac{1+3a_2}{3\delta^3}\nabla\cdot\Delta u_2 = 0, \\
(1+\mu b_1\Delta)\partial_t u_1 + \mu\frac{a_1}{\delta}\Delta\partial_t u_2 + \alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\big(|u_1|^2\big) = 0, \\
(1+\mu(\frac{a_2}{\delta^2} - \frac{\gamma}{\delta})\Delta)\partial_t u_2 - \mu\frac{\gamma}{2}\Delta\partial_t u_1 + (1-\gamma)\nabla\zeta_2 + \alpha\gamma\nabla\zeta_1 \\
\qquad\qquad + \dfrac{\epsilon_2}{2}\nabla\big(|u_2|^2\big) = 0.
\end{array}
\right. \tag{2.27}
$$

Remark 2.14. If we choose $a_1 = -\frac{1}{2}$, $a_2 = -\frac{1}{3}$, and $b_1 = -\frac{1}{3}$, we obtain the classical "layer-mean" model (4.12), introduced by Choi and Camassa in [**30**]. As we see below, this system is linearly ill-posed. One of the interests of (2.27) is to offer a large class of equivalent models with parameters which can be chosen so that the system is linearly well-posed.

Proposition 2.15. *The full system* (1.13) *is consistent with* (2.27)*, at the precision* $\mu^2 C_1$*, with*

$$
C_1 = C\bigg(\tfrac{1}{h}, \beta\big|b\big|_{W^{1,\infty}H^{s+11/2}}, \epsilon_2\big|\zeta_2\big|_{W^{1,\infty}H^{s+9/2}}, \epsilon_1\big|\zeta_1\big|_{W^{1,\infty}H^{s+7/2}}, \big|\nabla\psi_1\big|_{W^{1,\infty}H^{s+11/2}}, \\
\big|\nabla\psi_2\big|_{W^{1,\infty}H^{s+11/2}}\bigg).
$$

Proof. Let $t_0 > d/2$ and $s \geq t_0 + 1/2$. Let $U := (\zeta_1, \zeta_2, \nabla\psi_1, \nabla\psi_2)$ be a solution of (1.13) such that (1.10) is satisfied and $U \in \mathbf{H}^{s+2}$.

We first give the proof for $a_1 = b_1 = a_2 = 0$ corresponding to the original system (2.26). We just have to plug U in (2.26), as in the proof of Proposition 2.12. Since $\epsilon_2 \sim \mu$, we have $\big|\mu\epsilon_2\nabla\mathcal{N}_1\big|_{H^s} + \big|\mu\epsilon_2\nabla\mathcal{N}_2\big|_{H^s} \leq \mu^2 C_1$. The other residuals are bounded by $\mu^2 C_1$, thanks to the estimates (2.8), (2.18), and (2.20) with $\epsilon_2 \ll 1$ and the equivalent estimates on the derivatives which are obtained as in the proof of Proposition 2.12.

The general case is obtained when we substitute $\nabla\psi_1 - \mu b_1\Delta\nabla\psi_1 - \mu\frac{1}{\delta}a_1\Delta\nabla\psi_2$ for u_1 and $\nabla\psi_2 - \mu\frac{1}{\delta^2}a_2\Delta\nabla\psi_2$ for u_2 in (2.27). We obtain (2.26) up to additional terms that are clearly bounded by $\mu^2 C_1$. □

Dispersion relations. As we have said previously, the coefficients can be chosen so that the system (2.27) is linearly well-posed. Indeed, it is straightforward to check from the linearized system that $\omega_\pm^2(k)$, corresponding to plane-wave solutions $e^{ik\cdot X - i\omega_\pm t}$, must be the solutions of the equation

$$
\omega^4 - A(\mu|k|^2)|k|^2\omega^2 + B(\mu|k|^2)|k|^4 = 0 \tag{2.28}
$$

with

$$
A(Y) := \frac{(1-\beta_1 Y)\Big(1+\frac{\gamma\delta(a_1+1)-a_2}{\delta^2}Y\Big)+\gamma\big(\frac{1}{\delta}-(\alpha_1+\alpha_2)Y\big)\big(1-\big(b_1+\frac{1}{2}\big)Y\big)+(1-\gamma)\big(\frac{1}{\delta}-\alpha_2 Y\big)(1-b_1 Y)}{(1-b_1Y)\Big(1-\frac{a_2-\gamma\delta}{\delta^2}Y\Big)+\frac{\gamma}{2\delta}a_1Y^2},
$$

$$
B(Y) := (1-\gamma)\frac{\big(\frac{1}{\delta}-\alpha_2 Y\big)(1-\beta_1 Y)}{(1-b_1Y)\Big(1-\frac{a_2-\gamma\delta}{\delta^2}Y\Big)+\frac{\gamma}{2\delta}a_1Y^2}
$$

and the notations

$$
\alpha_1 := \frac{1+2a_1}{2\delta}, \qquad \alpha_2 := \frac{1+3a_2}{3\delta^3}, \qquad \beta_1 := \frac{1+3b_1}{3}.
$$

In order to have two positive solutions of (2.28), the coefficients have to satisfy $a_2 \leq -1/3$ and $b_1 \leq -1/2$. We see that the original system (2.26) and the classical layer-mean

model (4.12) are ill-posed. However, there exist sets of parameters a_1, a_2, b_1 such that the generalized system is well-posed. Moreover, we can choose the coefficients such that the dispersions meet with the ones of the full system at the order 3 in $\mu|k|^2$. We present in Figure 4 the difference between the dispersion of the full system and the one of the Boussinesq/Boussinesq model for three sets of parameters: $a_1 = b_1 = a_2 = 0$ corresponding to the original system (2.26); $a_1 = -\frac{1}{2}$, $a_2 = -\frac{1}{3}$, and $b_1 = -\frac{1}{3}$ corresponding to the layer-mean system (4.12); and finally $a_1 \approx 0.4714$, $a_2 \approx -0.3942$, and $b_1 = -1$ corresponding to optimized parameters in (2.27). Moreover, we chose $\mu = 0.1$, $\delta = 1/3$, and $\gamma = 2/3$. Note that except for the last set of parameters, the system is linearly ill-posed so that the computation breaks for high wave numbers.

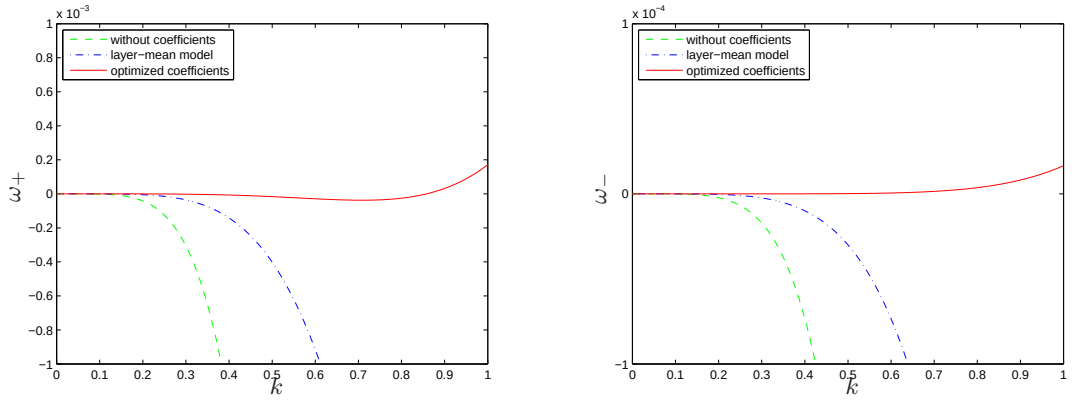


FIGURE 4. The Boussinesq/Boussinesq models dispersion error.

2.3.3. *The higher order system.* We are now back in the strong linearity regime allowing large amplitude (ϵ_1, $\epsilon_2 = O(1)$). But now we use the higher order expansions (2.8), (2.18), and (2.20) and thus obtain the strongly nonlinear model

(2.29)
$$\begin{cases} \alpha\partial_t\zeta_1 + \mathcal{A}_1 + \mathcal{A}_2 = \mu\Big(\nabla\cdot\mathcal{T}_1 + \nabla\cdot\mathcal{T}_2 - \frac{1}{2}\nabla\cdot(h_1^2\nabla\mathcal{A}_2) - \nabla\cdot\big(h_1\epsilon_1\nabla\zeta_1\mathcal{A}_2\big)\Big), \\ \partial_t\zeta_2 + \mathcal{A}_2 = \mu\nabla\cdot\mathcal{T}_2, \\ \partial_t\nabla\psi_1 + \alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_1|^2\right) = \mu\epsilon_2\nabla\mathcal{N}_1, \\ \partial_t\nabla\psi_2 + (1-\gamma)\nabla\zeta_2 + \gamma\alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_2|^2\right) = \mu\Big(\gamma\partial_t\nabla\mathcal{H} + \gamma\epsilon_2\nabla(\nabla\psi_1\cdot\nabla\mathcal{H}) \\ \qquad\qquad +\epsilon_2\nabla\mathcal{N}_2 + \gamma\epsilon_2\nabla\mathcal{N}_1\Big), \end{cases}$$

where we have used the following notations:

$$\begin{aligned} &\mathcal{A}_1 := \nabla\cdot(h_1\nabla\psi_1), \qquad \mathcal{A}_2 := \nabla\cdot(h_2\nabla\psi_2), \\ &\mathcal{T}_1 := \mathcal{T}[h_1, \epsilon_2\zeta_2]\nabla\psi_1, \quad \mathcal{T}_2 := \mathcal{T}[h_2, \beta b]\nabla\psi_2, \\ &\mathcal{H} := h_1(\nabla\cdot(h_1\nabla\psi_1) + \nabla\cdot(h_2\nabla\psi_2) - \tfrac{1}{2}h_1\Delta\psi_1 - \epsilon_1\nabla\zeta_1\cdot\nabla\psi_1), \\ &\mathcal{N}_1 := \tfrac{(\epsilon_1\nabla\zeta_1\cdot\nabla\psi_1 - \nabla\cdot(h_1\nabla\psi_1) - \nabla\cdot(h_2\nabla\psi_2))^2}{2}, \\ &\mathcal{N}_2 := \tfrac{(\epsilon_2\nabla\zeta_2\cdot\nabla\psi_2 - \nabla\cdot(h_2\nabla\psi_2))^2 - \gamma(\epsilon_2\nabla\zeta_2\cdot\nabla\psi_1 - \nabla\cdot(h_2\nabla\psi_2))^2}{2}. \end{aligned}$$

PROPOSITION 2.16. *The full system* (1.13) *is consistent with* (2.29)*, at the precision* $\mu^2 C_1$*, with*

$$C_1 = C\left(\tfrac{1}{h}, \beta\big|b\big|_{W^{1,\infty}H^{s+11/2}}, \epsilon_2\big|\zeta_2\big|_{W^{1,\infty}H^{s+9/2}}, \epsilon_1\big|\zeta_1\big|_{W^{1,\infty}H^{s+7/2}}, \big|\nabla\psi_1\big|_{W^{1,\infty}H^{s+11/2}}, \big|\nabla\psi_2\big|_{W^{1,\infty}H^{s+11/2}}\right).$$

Proof. Let $t_0 > d/2$ and $s \geq t_0 + 1/2$. Let $U := (\zeta_1, \zeta_2, \nabla\psi_1, \nabla\psi_2)$ be a solution of (1.13) such that (1.10) is satisfied and $U \in \mathbf{H}^{s+2}$. We plug U into (2.29), and, thanks to the estimates (2.8), (2.18), and (2.20) and that the equivalent estimates on the derivatives are obtained as in the proof of Proposition 2.12, we can check that the residuals are bounded by $\mu^2 C_1$. □

Dispersion relations. The linearized system is the same as the one of (2.26). So the system is linearly ill-posed, and we should derive models with parameters to obtain well-posed systems.

3. Convergence results

We show here how to use the consistency results obtained in section 2.3 to prove convergence results, stating that solutions of (1.13), if they exist, remain close to the solutions of the asymptotic models that are symmetrizable hyperbolic systems.

Remark 3.1. It is not clear that each of our models can be written as a symmetrizable hyperbolic system. That is why we focus here on the shallow water/shallow water model (2.21) in the flat-bottom case ($\beta = 0$). We set $d = 2$, and the case $d = 1$ follows immediately. The case of the Boussinesq/Boussinesq models will be discussed in a later work.

The analysis is based on classical results for quasilinear systems, which can be found, for example, in **[84]** and **[155]** and that we recall here.

Lemma 3.2. *Let $s > \frac{d}{2} + 1$ and $T > 0$. We assume that A_j are smooth functions of $u \in \mathbb{R}^n$ such that the system*

$$\partial_t U + \sum_{j=1}^{d} A_j(U)\partial_x U = F(t, x, U) \tag{3.1}$$

is Friedrichs-symmetrizable. Moreover, we assume that $u \mapsto F(t, x, u)$ is a smooth function of $u \in \mathbb{R}^n$ and that $F(t, x, u)$ is bounded in H^s uniformly with respect to $t \in [0, T]$. Then, for $g \in H^s(\mathbb{R}^d)$, taking values in $\mathbb{R}^n$, there exist $0 < T' \leq T$ and a unique $U \in C^0([0, T'); H^s(\mathbb{R}^d))^n$ such that U satisfies (3.1) and $U(t = 0) = g$. Moreover, U belongs to $U \in C^0([0, T'); H^s)^n \cap C^1([0, T'); H^{s-1})^n$, and if U satisfies

$$\big|U\big|_{W^{1,\infty}([0,T]\times\mathbb{R}^d)} \leq M,$$

for $M > 0$, then there are constants $C(M)$ and $K(M)$ such that

$$\big|U(t)\big|_{H^s} \leq Ce^{Kt}\big|g\big|_{H^s} + C\int_0^t e^{K(t-s)}\big|f(s)\big|_{H^s} ds$$

with $f(t, x) = F(t, x, g)$.

First we remark that the shallow water/shallow water model (2.21) in the flat-bottom case ($\beta = 0$) can be written as a quasilinear system

$$\partial_t U + A_1(U)\partial_x U + A_2(U)\partial_y U = 0 \tag{3.2}$$

with the notation

$$\begin{aligned} U &:= (h_1, h_2, u_{1x}, u_{1y}, u_{2x}, u_{2y}) \\ &= (1 + \epsilon_1\zeta_1 - \epsilon_2\zeta_2, \frac{1}{\delta} + \epsilon_2\zeta_2, \epsilon_2\partial_x\psi_1, \epsilon_2\partial_y\psi_1, \epsilon_2\partial_x\psi_2, \epsilon_2\partial_y\psi_2) \end{aligned}$$

and the matrices

$$A_1(U) := \begin{pmatrix} u_{1x} & 0 & h_1 & 0 & 0 & 0 \\ 0 & u_{2x} & 0 & 0 & h_2 & 0 \\ 1 & 1 & u_{1x} & u_{1y} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma & 1 & 0 & 0 & u_{2x} & u_{2y} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_2(U) := \begin{pmatrix} u_{1y} & 0 & 0 & h_1 & 0 & 0 \\ 0 & u_{2y} & 0 & 0 & 0 & h_2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & u_{1x} & u_{1y} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma & 1 & 0 & 0 & u_{2x} & u_{2y} \end{pmatrix}.$$

We prove now that the Cauchy problem associated with (3.2) is well-posed under some assumptions on the initial data since the quasilinear system is Friedrichs-symmetrizable.

PROPOSITION 3.3. *Let $s > \frac{d}{2} + 1$. Let $U_0 \in H^s(\mathbb{R}^d)^6$ such that there exists $h > 0$ such that for all X in $\mathbb{R}^d$, $U_0(X)$ satisfies the assumptions*

$$h_1, h_2 > h,\ |u_{1x}^2 + u_{1y}^2|, |u_{2x}^2 + u_{2y}^2| < h, \quad \text{and } (h_1 - u_{1x}^2 - u_{1y}^2)(h_2 - u_{2x}^2 - u_{2y}^2) > \gamma h_1 h_2. \tag{3.3}$$

Then there exists $T' > 0$ and a unique $U \in C^0([0, T'); H^s(\mathbb{R}^d))^6$ such that U satisfies (3.2) and $U(t = 0) = U_0$.

PROOF. We introduce the following matrix S, namely,

$$S(U) := \begin{pmatrix} \gamma & \gamma & \gamma u_{1x} & \gamma u_{1y} & 0 & 0 \\ \gamma & 1 & 0 & 0 & u_{2x} & u_{2y} \\ \gamma u_{1x} & 0 & \gamma h_1 & 0 & 0 & 0 \\ \gamma u_{1y} & 0 & 0 & \gamma h_1 & 0 & 0 \\ 0 & u_{2x} & 0 & 0 & h_2 & 0 \\ 0 & u_{2y} & 0 & 0 & 0 & h_2 \end{pmatrix}.$$

It is straightforward to check that $S(U)$ and $S(U)A(U, \xi)$ are self-adjoint with

$$A(U, \xi) := \xi_1 A_1(U) + \xi_2 A_2(U).$$

Then, using the Gauss reduction algorithm, we can check that $S(U)$ is definite positive if U satisfies (3.3). These requirements are satisfied at time $t = 0$ by U_0, and we define T as the maximum time such that they remain satisfied for all $t < T$. We know that $T > 0$, thanks to a continuity argument. Then since we have proved that S is a symmetrizer of (3.2), Lemma 3.2 gives $0 < T' \leq T$ such that U is uniquely defined on $[0, T')$. □

The last step consists of proving that the solutions of (3.2) approximate the solutions of the full system (1.13), assuming that the latter exist. This is obtained thanks to the energy estimate of Lemma 3.2.

Proposition 3.4. *We fix $\gamma \in (0,1)$ and $\delta \in (0,+\infty)$. For $t_0 > d/2$ and $s \geq t_0+1/2$, let $U \in C^1([0;T];H^s)^6 \cap C^0([0;T];H^{s+1})^6$ be a solution of (1.13) such that (1.10) is satisfied and U is bounded in $\mathbf{H}^s([0,T])$, uniformly with respect to ϵ_1, $\epsilon_2 \in [0,1)$, and $\mu \in (0,\mu^{max}]$. We denote by $\tilde{U} := (\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{u}_1, \tilde{u}_2)$ the solution of (2.21) with the same initial values that we assume to satisfy (3.3). Then we have*

$$\left|U - \tilde{U}\right|_{H^s} \leq \mu C_0$$

with $C_0 = C(\frac{1}{h}, \gamma, \delta, \mu^{max}, \left|U\right|_{\mathbf{H}^s}, T)$.

Proof. Thanks to the consistency result (Proposition 2.12), we know that U satisfies (3.1) with $F(t,x,U) = f(t,x)$ and

$$\left|f\right|_{H^s} \leq \mu C_0$$

with $C_0 = C(\frac{1}{h}, \left|U\right|_{\mathbf{H}^s})$. Then the difference between the two solutions $R^\mu := U - \tilde{U}$ satisfies (3.1) with the same f and

$$F(t,x,R^\mu) := f(t,x) - A_1(R^\mu)\partial_x\tilde{U} - A_2(R^\mu)\partial_y\tilde{U}.$$

Taking a smaller T if necessary, we have

$$\left|U\right|_{(W^{1,\infty}([0,T]\times\mathbb{R}^d))^6} + \left|\tilde{U}\right|_{(W^{1,\infty}([0,T]\times\mathbb{R}^d))^6} \leq M,$$

where M is independent of ϵ_1, ϵ_2 and μ. Thus we can apply Lemma 3.2, and we have

$$\left|R^\mu(t)\right|_{H^s} \leq C\int_0^t e^{K(t-s)}\left|f\right|_{H^s} ds \leq \mu C\left(\frac{1}{h}, \gamma, \delta, \mu^{max}, \left|U\right|_{\mathbf{H}^s}, T\right). \qquad \square$$

4. Links to other models

4.1. Rigid lid in the shallow water/shallow water case. In [**17**], Bona, Lannes, and Saut presented a model for internal waves in the shallow water regime with the rigid-lid assumption. They showed that a nonlocal operator has to appear for $d = 2$ (see observations in [**51**]). This operator cannot be seen in our model (2.21) so that it is a purely two-dimensional, rigid-lid effect. However, we show in the following how to make it appear from (2.21).

Indeed, the rigid-lid assumption means that $\epsilon_1 = 0$ when ϵ_2 remains > 0 so that $\alpha = 0$. The system (2.21) becomes

$$\left\{\begin{array}{l} \nabla\cdot(h_1\nabla\psi_1) + \nabla\cdot(h_2\nabla\psi_2) = 0, \\ \partial_t\zeta_2 + \nabla\cdot(h_2\nabla\psi_2) = 0, \\ \partial_t\nabla\psi_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_1|^2\right) = 0, \\ \partial_t\nabla\psi_2 + (1-\gamma)\nabla\zeta_2 + \dfrac{\epsilon_2}{2}\nabla\left(|\nabla\psi_2|^2\right) = 0, \end{array}\right. \tag{4.1}$$

where $h_1 = 1 - \epsilon_2\zeta_2$ and $h_2 = \frac{1}{\delta} - \beta b + \epsilon_2\zeta_2$.

For simplicity, we restrict ourself to the case of a flat bottom ($\beta = 0$), but we could do the same calculations with $\beta > 0$. We first define the shear velocity

$$v := \nabla\psi_2 - \gamma\nabla\psi_1.$$

From the first line, we deduce

$$\nabla\cdot(h_2 v) = -\nabla\cdot((h_1+\gamma h_2)\nabla\psi_1) = -\frac{\gamma+\delta}{\delta}\nabla\cdot\left(\left(1 + \frac{\gamma-1}{\gamma+\delta}\delta\epsilon_2\zeta_2\right)\nabla\psi_1\right).$$

Then we define the nonlocal operator $\mathfrak{Q}$ as follows.

DEFINITION 4.1. *Assuming that $\zeta \in L^\infty(\mathbb{R}^d)$, we define the mapping*

$$\mathfrak{Q}[\zeta] := \begin{array}{ccc} L^2(\mathbb{R}^d)^d & \to & L^2(\mathbb{R}^d)^d, \\ W & \mapsto & V, \end{array}$$

where V is the unique gradient vector in $L^2(\mathbb{R}^d)^d$, solution of

$$\nabla \cdot ((1+\zeta)V) = \nabla \cdot W.$$

So from the definition, we have

$$\nabla\psi_1 = \mathfrak{Q}[\frac{\gamma-1}{\gamma+\delta}\delta\epsilon_2\zeta_2]\left(-\frac{\delta}{\gamma+\delta}h_2 v\right).$$

We plug this expression into (4.1) and obtain immediately

$$\left\{\begin{array}{l} \partial_t\zeta_2 + \frac{\delta}{\gamma+\delta}\nabla\cdot\big(h_1\mathfrak{Q}[\frac{\gamma-1}{\gamma+\delta}\delta\epsilon_2\zeta_2](h_2 v)\big) = 0, \\ \partial_t v + (1-\gamma)\nabla\zeta_2 + \dfrac{\epsilon_2}{2}\nabla\Big(|v - \frac{\gamma\delta}{\gamma+\delta}\mathfrak{Q}[\frac{\gamma-1}{\gamma+\delta}\delta\epsilon_2\zeta_2](h_2 v)|^2 \\ \qquad\qquad -\frac{\gamma\delta^2}{(\gamma+\delta)^2}|\mathfrak{Q}[\frac{\gamma-1}{\gamma+\delta}\delta\epsilon_2\zeta_2](h_2 v)|^2\Big) = 0, \end{array}\right. \tag{4.2}$$

where $h_1 = 1-\epsilon_2\zeta_2$ and $h_2 = \frac{1}{\delta}+\epsilon_2\zeta_2$. This is exactly the system derived in [**17**].

Using the same method, we could derive rigid-lid models from (2.27) and (2.29). The rigid-lid model in the Boussinesq regime has already been exhibited in [**17**], and a fully nonlinear model is presented in [**31**].

4.2. The layer-mean equations. In the literature, the water-wave system is often given by layer-mean equations (see, for example, [**118**]), using as unknowns the depth-mean velocity across the layers:

$$\begin{array}{lll} \overline{u}_1(X) & := \frac{1}{h_1}\int_{\epsilon_2\zeta_2}^{1+\epsilon_1\zeta_1}\nabla\phi_1(X, r_1(X,z))dz & \text{with } h_1 := 1+\epsilon_1\zeta_1-\epsilon_2\zeta_2, \\ \overline{u}_2(X) & := \frac{1}{h_2}\int_{-1/\delta+\beta b}^{\epsilon_2\zeta_2}\nabla\phi_2(X, r_2(X,z))dz & \text{with } h_2 := \dfrac{1}{\delta}-\beta b+\epsilon_2\zeta_2. \end{array}$$

The systems under this form (obtained, for example, in [**30**] and [**8**]) are equivalent to the system we derived since we can approximate $\overline{u}_1$ and $\overline{u}_2$ thanks to our previous unknowns ψ_1 and ψ_2 (as we see in the following proposition) and conversely. Thus, our study gives a rigorous justification of these models, and we are able to offer consistency results.

PROPOSITION 4.2. *Let $t_0 > d/2$ and $s \geq t_0 + 1/2$, $\nabla\psi_1$, $\nabla\psi_2 \in H^{s+11/2}(\mathbb{R}^2)$, $\zeta_1 \in H^{s+7/2}(\mathbb{R}^2)$, $\zeta_2 \in H^{s+9/2}(\mathbb{R}^2)$, and $b \in H^{s+11/2}(\mathbb{R}^2)$ such that (1.10) is satisfied. Then we have*

$$\big|\overline{u}_1 - \nabla\psi_1\big|_{H^{s+1}} \leq \mu C_0, \tag{4.3}$$

$$\big|\overline{u}_2 - \nabla\psi_2\big|_{H^{s+1}} \leq \mu C_0, \tag{4.4}$$

$$\big|\overline{u}_1 - \nabla\psi_1 - \mu\mathcal{D}_1(\nabla\psi_1, \nabla\psi_2)\big|_{H^{s+1}} \leq \mu^2 C_1, \tag{4.5}$$

$$\big|\overline{u}_2 - \nabla\psi_2 - \mu\mathcal{D}_2\nabla\psi_2\big|_{H^{s+1}} \leq \mu^2 C_1, \tag{4.6}$$

with

$$C_j = C\left(\tfrac{1}{h}, \beta\big|b\big|_{H^{s+7/2+2j}}, \epsilon_2\big|\zeta_2\big|_{H^{s+5/2+2j}}, \epsilon_1\big|\zeta_1\big|_{H^{s+3/2+2j}}, \big|\nabla\psi_1\big|_{H^{s+7/2+2j}}, \big|\nabla\psi_2\big|_{H^{s+7/2+2j}}\right)$$

and where $\mathcal{D}_1$ and $\mathcal{D}_2$ are defined by

$$\begin{array}{rl} \mathcal{D}_1(\nabla\psi_1, \nabla\psi_2) & = -\frac{1}{h_1}\Big(\mathcal{T}_1 - \frac{1}{2}(h_1^2\nabla\mathcal{A}_2) - (h_1\epsilon_1\nabla\zeta_1\mathcal{A}_2)\Big), \\ \mathcal{D}_2\nabla\psi_2 & = -\frac{1}{h_2}\mathcal{T}_2 \end{array}$$

with the notations of Proposition 2.5.

Proof. Using the Green formula with ϕ_1 the solution of (1.12) and a test function $\tilde{\varphi} := (X, z) \mapsto \varphi(X)$, we have

$$\begin{aligned}\int_{\Omega_1} \tilde{\varphi}\Delta^{\mu}_{X,z}\phi_1 dXdz &= -\int_{\Omega_1} \nabla^{\mu}_{X,z}\phi_1 \cdot \nabla^{\mu}_{X,z}\tilde{\varphi} dXdz + \int_{\Gamma_1} \varphi\partial_{n_1}\phi_1 dn_1 + \int_{\Gamma_2} \varphi\partial_{n_2}\phi_1 dn_2 \\ &= -\mu\int_{\mathbb{R}^d} \nabla\varphi \int_{\epsilon_2\zeta_2}^{1+\epsilon_1\zeta_1} \nabla\phi_1 dzdX + \int_{\mathbb{R}^d} \varphi\left(G_1(\psi_1,\psi_2) - G_2\psi_2\right) dX.\end{aligned}$$

Thus we deduce

$$\nabla \cdot (h_1\overline{u}_1) = \frac{-1}{\mu}(G_1(\psi_1,\psi_2) - G_2\psi_2). \tag{4.7}$$

Identically we have

$$\nabla \cdot (h_2\overline{u}_2) = \frac{-1}{\mu}G_2\psi_2. \tag{4.8}$$

We now prove the estimate (4.3), and the others are obtained in the same way.

Using Propositions 2.2 and 2.5 together with (4.7), and since (1.10) is satisfied, we have immediately

$$\left|\nabla \cdot (\overline{u}_1 - \nabla\psi_1)\right|_{H^s} \leq \mu C_0$$

so that we have only to obtain an L^2-estimate on $\overline{u}_1 - \nabla\psi_1$. Using the definition of $\overline{u}_1$ and the mappings defined on section 2, we obtain

$$\overline{u}_1 - \nabla\psi_1 = \int_0^1 \nabla(\tilde{\phi}_1 - \psi_1) + \nabla s_1\partial_{\tilde{z}}\tilde{\phi}_1 d\tilde{z}$$

with $\phi_1 : (X, \tilde{z}) \in \mathcal{S}^+ \mapsto \tilde{\phi}_1(X, s_1(X, \tilde{z}))$. We deduce

$$\left|\overline{u}_1 - \nabla\psi_1\right|_2 \leq C(\epsilon_1\left|\zeta_1\right|_{W^{1,\infty}}, \epsilon_2\left|\zeta_2\right|_{W^{1,\infty}})\left\|\nabla_{X,\tilde{z}}(\tilde{\phi}_1 - \psi_1)\right\|_2.$$

The estimate follows now from Step 3 of section A.2, together with the estimates (A.2) and (A.4). □

4.2.1. *The shallow water/shallow water regime:* $\mu \ll 1$, $\epsilon = O(1)$. We use (4.3) and (4.4) in the system (2.21), and with a straightforward linear combination, we obtain

$$\begin{cases} \partial_t h_1 + \epsilon_2\nabla \cdot (h_1\overline{u}_1) = 0, \\ \partial_t h_2 + \epsilon_2\nabla \cdot (h_2\overline{u}_2) = 0, \\ \partial_t\overline{u}_1 + \nabla h_2 + \beta\nabla b + \nabla h_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\overline{u}_1|^2\right) = 0, \\ \partial_t\overline{u}_2 + \nabla h_2 + \beta\nabla b + \gamma\nabla h_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\overline{u}_2|^2\right) = 0. \end{cases} \tag{4.9}$$

Proposition 4.3. *The full system* (1.13) *is consistent with* (4.9)*, at the precision* μC_0*, with*

$$C_0 = C\left(\tfrac{1}{h}, \beta\left|b\right|_{W^{1,\infty}H^{s+7/2}}, \epsilon_2\left|\zeta_2\right|_{W^{1,\infty}H^{s+5/2}}, \epsilon_1\left|\zeta_1\right|_{W^{1,\infty}H^{s+3/2}}, \left|\nabla\psi_1\right|_{W^{1,\infty}H^{s+7/2}}, \left|\nabla\psi_2\right|_{W^{1,\infty}H^{s+7/2}}\right).$$

Proof. We know from Proposition 2.12 that (1.13) is consistent with (2.21), at the precision μC_0. From (4.3) and (4.4), we deduce that $(\zeta_1, \zeta_2, \overline{u}_1, \overline{u}_2)$ satisfies (4.9) up to a residual of the same order. □

Remark 4.4. Note that the first two equations of (4.9) are equalities (where the last two equations are first order approximations in μ), as we can see from (1.13), (4.7), and (4.8). They reveal the conservation of mass. Conservation of momentum and energy are the one obtained in section 2.3.1 when we substitute $\overline{u}_i$ for $\nabla\psi_i$ $(i = 1, 2)$. These conservation laws and the one of higher order systems have already been introduced in the flat-bottom case in [**8**].

4.2.2. *The Boussinesq/Boussinesq regime:* $\mu \sim \epsilon_2 \sim \epsilon_1 \ll 1$. We now restrict ourself to the flat-bottom case since it considerably simplifies the notations, but the following could be derived with $\beta \neq 0$ without any difficulty. The estimates (4.5) and (4.6) with $\epsilon_2 \sim \mu$ and $\beta = 0$ give the following formal relations:

$$\overline{u}_1 \approx \nabla\psi_1 + \mu\left(\frac{1}{3}\nabla\Delta\psi_1 + \frac{1}{2\delta}\nabla\Delta\psi_2\right), \tag{4.10}$$

$$\overline{u}_2 \approx \nabla\psi_2 + \mu\frac{1}{3\delta^2}\nabla\Delta\psi_2. \tag{4.11}$$

Plugging this into (2.26), we obtain the system

$$\begin{cases} \partial_t h_1 + \epsilon_2\nabla\cdot(h_1\overline{u}_1) = 0, \\ \partial_t h_2 + \epsilon_2\nabla\cdot(h_2\overline{u}_2) = 0, \\ \partial_t\overline{u}_1 + \alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\overline{u}_1|^2\right) = \mu\partial_t\left(\frac{1}{3}\Delta\overline{u}_1 + \frac{1}{2\delta}\Delta\overline{u}_2\right), \\ \partial_t\overline{u}_2 + (1-\gamma)\nabla\zeta_2 + \alpha\gamma\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\overline{u}_2|^2\right) = \mu\partial_t\left(\left(\frac{1}{3\delta^2} + \frac{\gamma}{\delta}\right)\Delta\overline{u}_2 + \frac{\gamma}{2}\Delta\overline{u}_1\right). \end{cases} \tag{4.12}$$

Remark 4.5. This set of equations had been revealed in **[30]**. It corresponds to (2.27) with the choice of parameters: $a_1 = -\frac{1}{2}$, $a_2 = -\frac{1}{3}$, $b_1 = -\frac{1}{3}$. This particular choice of parameters leads to a linearly ill-posed system. That is why it is interesting to obtain, as in section 2.3.2, a larger class of models allowing linearly well-posed systems.

Since this system is a particular case of the Boussinesq/Boussinesq model (2.27), we can apply Proposition 2.15.

Proposition 4.6. *The full system* (1.13) *is consistent with* (4.12)*, at the precision* $\mu^2 C_1$*, with*

$$C_1 = C\left(\tfrac{1}{h}, \beta\left|b\right|_{W^{1,\infty}H^{s+11/2}}, \epsilon_2\left|\zeta_2\right|_{W^{1,\infty}H^{s+9/2}}, \epsilon_1\left|\zeta_1\right|_{W^{1,\infty}H^{s+7/2}}, \left|\nabla\psi_1\right|_{W^{1,\infty}H^{s+11/2}}, \left|\nabla\psi_2\right|_{W^{1,\infty}H^{s+11/2}}\right).$$

4.2.3. *The higher order system.* We now do the same study without assuming any smallness on ϵ_1, ϵ_2. We plug (4.5) and (4.6) into (2.29), and we obtain

$$\begin{cases} \partial_t h_1 + \epsilon_2\nabla\cdot(h_1\overline{u}_1) = 0, \\ \partial_t h_2 + \epsilon_2\nabla\cdot(h_2\overline{u}_2) = 0, \\ \partial_t\overline{u}_1 + \alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\overline{u}_1|^2\right) = \mu\epsilon_2\nabla\mathcal{N}_1 + \mu\epsilon_2\nabla(\overline{u}_1\cdot\mathcal{D}_1) + \mu\partial_t\mathcal{D}_1, \\ \partial_t\overline{u}_2 + (1-\gamma)\nabla\zeta_2 + \gamma\alpha\nabla\zeta_1 + \dfrac{\epsilon_2}{2}\nabla\left(|\overline{u}_2|^2\right) = \mu\Big(\partial_t(\gamma\nabla\mathcal{H} + \mathcal{D}_2) + \epsilon_2\nabla(\gamma\overline{u}_1\cdot\nabla\mathcal{H} \\ \qquad\qquad + \overline{u}_2\cdot\mathcal{D}_2 + \mathcal{N}_2 + \gamma\mathcal{N}_1)\Big) \end{cases} \tag{4.13}$$

with the notations of Proposition 4.2 and (2.29) and when we substitute $\overline{u}_i$ for $\nabla\psi_i$ ($i = 1, 2$).

Proposition 4.7. *The full system* (1.13) *is consistent with* (4.13)*, at the precision* $\mu^2 C_1$*, with*

$$C_1 = C\left(\tfrac{1}{h}, \beta\left|b\right|_{W^{1,\infty}H^{s+11/2}}, \epsilon_2\left|\zeta_2\right|_{W^{1,\infty}H^{s+9/2}}, \epsilon_1\left|\zeta_1\right|_{W^{1,\infty}H^{s+7/2}}, \left|\nabla\psi_1\right|_{W^{1,\infty}H^{s+11/2}}, \left|\nabla\psi_2\right|_{W^{1,\infty}H^{s+11/2}}\right).$$

Proof. Let $t_0 > d/2$ and $s \geq t_0 + 1/2$. Let $(\zeta_1, \zeta_2, \nabla\psi_1, \nabla\psi_2)$ be a sufficiently smooth solution of (1.13) such that (1.10) is satisfied. We know from Proposition 2.16 that $(\zeta_1, \zeta_2, \nabla\psi_1, \nabla\psi_2)$ satisfies (2.29) up to a residual bounded by $\mu^2 C_1$. Then the estimates (4.5) and (4.6) give that $(\zeta_1, \zeta_2, \overline{u}_1, \overline{u}_2)$ satisfies (4.13) up to a residual of the same order. □

A. Proof of Proposition 2.5

Our proof contains three parts. First we introduce u, the correction to the expansion of ϕ_1 formally obtained in section 2.2, and we present the system solved by u. Then we use the elliptic form of the operator to obtain H^s estimates on u. Finally we use these estimates to prove the desired inequalities.

A.1. System solved by u. We first define the second order correction to the formal expansion

$$u := \phi_1 - \psi_1 + \mu \underbrace{h_1(z-1)\Big(h_1\big(\frac{z+1}{2}\big)\Delta\psi_1 - \epsilon_2\nabla\zeta_2\cdot\nabla\psi_1 + \nabla\cdot(h_2\nabla\psi_2)\Big)}_{:=\phi^1}.$$

From the computation carried out in section 2.2, we know that u satisfies the following equalities:

$$\begin{aligned}
\nabla_{X,z}\cdot P_1\nabla_{X,z}u &= \mu^2\nabla\cdot P^1\nabla\phi^1 && \text{in } \mathcal{S}^+,\\
u &= 0 && \text{on } \{z=1\},\\
\partial_n u &= G_2\psi_2 + \mu\nabla\cdot(h_2\nabla\psi_2) + \mu^2(\partial_n^{P^1}\phi^1) && \text{on } \{z=0\}.
\end{aligned}$$

Moreover, we notice that (4.8) gives $G_2\psi_2 + \mu\nabla\cdot(h_2\nabla\psi_2) = \nabla\cdot V$ with

$$V = \mu h_2(\nabla\psi_2 - \overline{u}_2).$$

Thus, using the definition of P_1 in (2.1), we finally have the system

$$\left\{\begin{array}{ll}
\nabla^\mu_{X,z}\cdot P^\mu\nabla^\mu_{X,z}u = \mu^2\ \nabla^\mu_{X,z}\cdot\mathbf{h} & \text{in } \mathcal{S}^+,\\
u = 0 & \text{on } \{z=1\},\\
\partial_n u = \nabla\cdot V + \mu^2 e_{d+1}\cdot\mathbf{h} & \text{on } \{z=0\},
\end{array}\right. \tag{A.1}$$

where we have introduced the notation $\nabla^\mu_{X,z} := (\sqrt{\mu}\nabla^T, \partial_z)^T$, $\mathbf{h} := P^\mu\nabla^\mu_{X,z}\phi^1$, and

$$P^\mu := \begin{pmatrix} h_1 I_d & -\sqrt{\mu}\nabla s_1 \\ -\sqrt{\mu}\nabla {s_1}^T & \frac{1+\mu|\nabla s_1|^2}{h_1}\end{pmatrix}.$$

We now give the useful estimates of the right-hand side of the system. It is straightforward to check that

$$\big\|\mathbf{h}\big\|_{H^{s+1/2,1}} \le C\left(\frac{1}{h}, \epsilon_1\big|\zeta_1\big|_{H^{s+3/2}}, \epsilon_2\big|\zeta_2\big|_{H^{s+5/2}}, \beta\big|b\big|_{H^{s+5/2}}, \big|\nabla\psi_1\big|_{H^{s+7/2}}, \big|\nabla\psi_2\big|_{H^{s+7/2}}\right). \tag{A.2}$$

Then, since we have $G_2\psi_2 + \mu\nabla\cdot(h_2\nabla\psi_2) = \nabla\cdot V$, Proposition 2.2 immediately gives

$$\big|\nabla\cdot V\big|_{H^s} \le \mu^2 C\left(\frac{1}{h}, \beta\big|b\big|_{H^{s+7/2}}, \epsilon_2\big|\zeta_2\big|_{H^{s+5/2}}, \big|\nabla\psi_2\big|_{H^{s+7/2}}\right). \tag{A.3}$$

We now seek a L^2-estimate of $V = \mu h_2(\nabla\psi_2 - \overline{u}_2)$. Using the definition of $\overline{u}_2$ and the mappings defined on section 2, we obtain easily that

$$\overline{u}_2 - \nabla\psi_2 = \int_{-1}^{0} \nabla(\tilde{\phi}_2 - \psi_2) + \nabla s_2\partial_{\tilde{z}}\tilde{\phi}_2 d\tilde{z}$$

with $\phi_2 : (X,\tilde{z}) \in \mathcal{S}^- \mapsto \tilde{\phi}_2(X, s_2(X,\tilde{z}))$. Then the method of our proof adapted for the lower fluid (this is done, for example, in **[26]**) leads at Step 3 to a L^2-estimate on $\nabla^\mu_{X,z}(\tilde{\phi}_2 - \psi_2)$. We then plug this estimate into the previous equality, deduce the desired estimate on $\big|V\big|_{L^2}$, and finally get with (A.3)

$$\big|V\big|_{H^s} \le \mu^2 C\left(\frac{1}{h}, \beta\big|b\big|_{H^{s+5/2}}, \epsilon_2\big|\zeta_2\big|_{H^{s+3/2}}, \big|\nabla\psi_2\big|_{H^{s+5/2}}\right). \tag{A.4}$$

A.2. $H^{s,1}$-estimate ($s \geq 0$) on u. We follow the sketch of the proof of Proposition 3 in [**17**], which contains five steps.

STEP 1. *Coercivity of the operator.* Since ζ_1, $\zeta_2 \in W^{1,\infty}$ and satisfy (1.10), we can check (see Proposition 2.3 of [**71**]) that, for any $\Theta \in \mathbb{R}^{d+1}$,

$$\Theta \cdot P^\mu \Theta \geq \frac{1}{k}\big|\Theta\big|^2$$

with $k = \big\|h_1\big\|_\infty + \frac{1}{h}(1 + \mu\big\|\nabla s_1\big\|_\infty^2)$. The operator is uniformly coercive in μ.

STEP 2. *Existence and uniqueness of the solution.* The result is given by the coercivity of the operator. From the assumptions on ζ_1, ζ_2, b, ψ_1, and ψ_2, we know that $\mathbf{h} \in H^{s+1/2,1}(\mathcal{S}^+)^{d+1}$ and $V \in H^{s+1}(\mathbb{R}^d)$. For $s \geq 1/2$, the proof of Proposition 2.1 works for the system (A.1) so that we know that there exists a unique solution in $H^2(\mathcal{S}^+)$. We now prove by induction that, for $k \in \mathbb{N}$,

$$\mathbf{h} \in H^{k+1} \text{ and } V \in H^k \Longrightarrow u \in H^{k+2}. \tag{A.5}$$

We assume that $\mathbf{h} \in H^{k+2}$ and $V \in H^{k+1}$. We thus know that $u \in H^{k+2}$ so that $v := \Lambda u \in H^{k+1} \subset H^1$. Hence, v is the classical solution of

$$\left\{\begin{array}{ll} \nabla^\mu_{X,z} \cdot P^\mu \nabla^\mu_{X,z} v = \mu^2 \ \nabla_{X,z} \cdot \tilde{\mathbf{h}} & \text{in } \mathcal{S}^+, \\ v = 0 & \text{on } \{z = 1\}, \\ \partial_n v = \nabla \cdot \Lambda V + \mu^2 e_{d+1} \cdot \partial_x \tilde{\mathbf{h}} & \text{on } \{z = 0\} \end{array}\right. \tag{A.6}$$

with $\mu^2\tilde{\mathbf{h}} = \mu^2 \Lambda \mathbf{h} + [\Lambda, P^\mu]\nabla^\mu_{X,z} u$. Thanks to Theorem 6 of [**69**], for $t_0 > \frac{d}{2}$, we have

$$\big\|[\Lambda, P^\mu]\nabla^\mu_{X,z} u\big\|_2 \leq C_{t_0}\big\|\nabla P^\mu\big\|_{H^{t_0}}\big\|\nabla^\mu_{X,z} u\big\|_2$$

so that $\tilde{\mathbf{h}} \in H^{k+1}$ and $\Lambda V \in H^k$. The inductive hypotheses are satisfied so that we know that $v \in H^{k+2}$. Finally we use the coercivity of the operator (Step 1) with the n^{th} derivative of (A.1) and obtain

$$\big\|\partial_z^2 \partial^n u\big\|_2 \leq k\big\|\nabla^\mu_{X,z} \cdot P^\mu \nabla^\mu_{X,z} \partial^n u\big\|_2 + \big\|\Delta_X \partial^n u\big\|_2.$$

It follows that $u \in H^{k+3}$, and (A.5) is proved. The interpolation theory leads to the final result: for $s \geq 1/2$, there exists a unique solution $u \in H^{s+3/2}$ of (A.1).

STEP 3. *L^2-estimate on $\nabla^\mu_{X,z} u$.* We multiply (A.1) by u, integrate by parts on both sides, and use the boundary conditions to finally obtain

$$\int_{\mathcal{S}} \nabla^\mu_{X,z} u \cdot P^\mu \nabla^\mu_{X,z} u = \mu^2 \int_{\mathcal{S}} u \nabla_{X,z} \cdot \mathbf{h} + \int_{\{z=0\}} \nabla u \cdot V.$$

From the coercivity, the Cauchy-Schwarz inequality and a Poincaré inequality, we deduce

$$\big\|\nabla^\mu_{X,z} u\big\|_2^2 \leq k(\mu^2\big\|\nabla_{X,z} \cdot \mathbf{h}\big\|_2\big\|\partial_z u\big\|_2 + \big|V\big|_{H^{1/2}}\big|\nabla u_{|z=0}\big|_{H^{-1/2}}).$$

Then a trace theorem (see Métivier [**83**], pp. 23–27) gives

$$\begin{aligned} \big|\nabla u_{|z=0}\big|_{H^{-1/2}} &\leq \mathrm{Cst}(\big\|\nabla u\big\|_2 + \big\|\Lambda^{-1}\partial_z \nabla u\big\|_2) \\ &\leq \mathrm{Cst}(\frac{1}{\sqrt{\mu}} + 1)\big\|\nabla^\mu_{X,z} u\big\|_2. \end{aligned}$$

This finally gives the estimate

$$\big\|\nabla^\mu_{X,z} u\big\|_2 \leq C\left(\frac{1}{h}, \epsilon_1\big|\zeta_1\big|_{W^{1,\infty}}, \epsilon_2\big|\zeta_2\big|_{W^{1,\infty}}\right)\left(\mu^2\big\|\nabla_{X,z} \cdot \mathbf{h}\big\|_2 + \frac{1+\sqrt{\mu}}{\sqrt{\mu}}\big|V\big|_{H^{1/2}}\right). \tag{A.7}$$

Step 4. *L^2-estimate on $\Lambda^s\nabla^\mu_{X,z}u$ ($s\geq 0$).* We define $v=\Lambda^s u$. Multiplying (A.1) by Λ^s on both sides, we obtain

$$\text{(A.8)}\qquad \begin{cases} \nabla^\mu_{X,z}\cdot P^\mu\nabla^\mu_{X,z}v=\mu^2\ \nabla_{X,z}\cdot\tilde{\mathbf{h}} & \text{in } \mathbb{R}^d\times(0,1),\\ v=0 & \text{on } \{z=1\},\\ \partial_n v=\nabla\cdot\Lambda^s V+\mu^2 e_{d+1}\cdot\tilde{\mathbf{h}} & \text{on } \{z=0\}\end{cases}$$

with $\mu^2\tilde{\mathbf{h}}=\mu^2\Lambda^s\mathbf{h}-[\Lambda^s,P^\mu]\nabla^\mu_{X,z}u$. We can use Step 3 with v and obtain

$$\left\|\nabla^\mu_{X,z}v\right\|_2\leq C\left(\tfrac{1}{h},\epsilon_1\left|\zeta_1\right|_{W^{1,\infty}},\epsilon_2\left|\zeta_2\right|_{W^{1,\infty}}\right)\Big(\mu^2\left\|\Lambda^s\nabla_{X,z}\cdot\mathbf{h}\right\|_2+\left\|[\Lambda^s,P^\mu]\nabla^\mu_{X,z}u\right\|_2 \\ +\tfrac{1+\sqrt{\mu}}{\sqrt{\mu}}\left|V\right|_{H^{s+1/2}}\Big).$$

We obtain the commutator estimate thanks to Theorem 6 of **[69]**: for $s>-\frac{d}{2}$ and $t_0>\frac{d}{2}$, we have

$$\left\|[\Lambda^s,f]g\right\|_2\leq C_{s,t_0}\left\|\nabla f\right\|_{H^{\max\{t_0,s-1\}}}\left\|g\right\|_{H^{s-1}}.$$

In our case, it gives

$$\left\|[\Lambda^s,P^\mu]\nabla^\mu_{X,z}u\right\|_2\leq C_{s,t_0}\left(\frac{1}{h},\epsilon_1\left|\zeta_1\right|_{H^{\max\{t_0+2,s+1\}}},\epsilon_2\left|\zeta_2\right|_{H^{\max\{t_0+2,s+1\}}}\right)\left\|\Lambda^{s-1}\nabla^\mu_{X,z}u\right\|_2.$$

We finally get an estimate on $\left\|\Lambda^s\nabla^\mu_{X,z}u\right\|_2$ in terms of $\left\|\Lambda^{s-1}\nabla^\mu_{X,z}u\right\|_2$. Step 3 is the case when $s=0$. By induction and by interpolation when $s\in(0,1)$, we obtain the following relation for all $s\geq 0$:

$$\text{(A.9)}\qquad \left\|\Lambda^s\nabla^\mu_{X,z}u\right\|_2\leq C_{s,t_0}\left(\tfrac{1}{h},\epsilon_1\left|\zeta_1\right|_{H^{\max\{t_0+2,s+1\}}},\epsilon_2\left|\zeta_2\right|_{H^{\max\{t_0+2,s+1\}}}\right)\left(\mu^2\left\|\Lambda^s\nabla_{X,z}\cdot\mathbf{h}\right\|_2\right. \\ \left.+\tfrac{1+\sqrt{\mu}}{\sqrt{\mu}}\left|V\right|_{H^{s+1/2}}\right).$$

Step 5. *L^2-estimate ($s\geq 0$) on $\Lambda^s\partial_z\nabla^\mu_{X,z}u$.* (A.1) gives the formula

$$\tfrac{1+\mu\left|\nabla s_1\right|^2}{h_1}\partial_z^2u=\mu^2\nabla^\mu_{X,z}\cdot\mathbf{h}-\mu\nabla\cdot(h_1\nabla u-\nabla s_1\partial_z u)+\mu\partial_z(\nabla s_1\cdot\nabla u) \\ -\partial_z\left(\tfrac{1+\mu\left|\nabla s_1\right|^2}{h_1}\right)\partial_z u,$$

from which we deduce

$$\left\|\Lambda^s\partial_z^2u\right\|_2\leq C\left(\frac{1}{h},\epsilon_1\left|\zeta_1\right|_{W^{1,\infty}},\epsilon_2\left|\zeta_2\right|_{W^{1,\infty}}\right)\left(\mu^2\left\|\Lambda^s\nabla^\mu_{X,z}\cdot\mathbf{h}\right\|_2+\sqrt{\mu}\left\|\Lambda^{s+1}\nabla^\mu_{X,z}u\right\|_2\right).$$

Thus, we have the estimate

$$\left\|\Lambda^s\partial_z\nabla^\mu_{X,z}u\right\|_2\leq C\left(\frac{1}{h},\epsilon_1\left|\zeta_1\right|_{W^{1,\infty}},\epsilon_2\left|\zeta_2\right|_{W^{1,\infty}}\right)\left(\mu^2\left\|\Lambda^s\nabla^\mu_{X,z}\cdot\mathbf{h}\right\|_2+\sqrt{\mu}\left\|\Lambda^{s+1}\nabla^\mu_{X,z}u\right\|_2\right),$$

and Step 4 allows us to conclude

$$\left\|\Lambda^s\partial_z\nabla^\mu_{X,z}u\right\|_2\leq C_{s,t_0}\left(\tfrac{1}{h},\epsilon_1\left|\zeta_1\right|_{H^{\max\{t_0+2,s+2\}}},\epsilon_2\left|\zeta_2\right|_{H^{\max\{t_0+2,s+2\}}}\right)\left(\mu^2\left\|\mathbf{h}\right\|_{H^{s+1,1}}\right. \\ \left.+\left|V\right|_{H^{s+3/2}}\right).$$

A.3. Proof of the inequalities. To obtain the first estimate, we remark that

$$G_1(\psi_1,\psi_2)+\mu(\mathcal{A}_1+\mathcal{A}_2)=\partial_n u_{|z=1}-\mu^2u_0$$

with $u_0:=|\epsilon_1\nabla\zeta_1|^2(h_1\Delta\psi_1-\epsilon_2\zeta_2\cdot\nabla\psi_1+\nabla\cdot(h_2\nabla\psi_2))$. It is straightforward to check that

$$\left|u_0\right|_{H^s}\leq C\left(\frac{1}{h},\beta\left|b\right|_{H^{s+1}},\epsilon_2\left|\zeta_2\right|_{H^{s+1}},\epsilon_1\left|\zeta_1\right|_{H^{s+1}},\left|\nabla\psi_1\right|_{H^{s+2}},\left|\nabla\psi_2\right|_{H^{s+2}}\right)$$

so that we just have to bound $\left|\partial_n u_{|z=1}\right|_{H^s}$. We now use the trace theorem to get

$$\begin{aligned}\left|\partial_n u_{|z=1}\right|_{H^s} &\leq C\left(\frac{1}{h}, \epsilon_1\left|\zeta_1\right|_{W^{1,\infty}}, \epsilon_2\left|\zeta_2\right|_{W^{1,\infty}}\right)\left(\mu\left|\nabla u_{|z=1}\right|_{H^s} + \left|\partial_z u_{|z=1}\right|_{H^s}\right)\\ &\leq C\left(\frac{1}{h}, \epsilon_1\left|\zeta_1\right|_{W^{1,\infty}}, \epsilon_2\left|\zeta_2\right|_{W^{1,\infty}}\right)\left(\sqrt{\mu}\left\|\nabla^\mu_{X,z} u\right\|_{H^{s+1/2,0}}\right.\\ &\qquad \left. + \left\|\partial_z \nabla^\mu_{X,z} u\right\|_{H^{s-1/2,0}}\right). \qquad \text{(A.10)}\end{aligned}$$

The estimates obtained in Steps 4 and 5, together with (A.2) and (A.4), give immediately the desired result.

To obtain the second estimate, we have to carry on the proof with the higher order approximate solution obtained in section 2.2,

$$\tilde{u} := \phi_1 - \phi_1^{app,2},$$

and we would obtain the estimates exactly as above. We omit this technical step.

CHAPITRE 3

Boussinesq/Boussinesq systems for internal waves with a free surface, and the KdV approximation

À paraître dans Mathematical Modelling and Numerical Analysis (M2AN).

Sommaire

Abstract

We study here some asymptotic models for the propagation of internal and surface waves in a two-fluid system. We focus on the so-called long wave regime for one-dimensional waves, and consider the case of a flat bottom. Following the method presented in [**16**] for the one-layer case, we introduce a new family of symmetric hyperbolic models, that are equivalent to the classical Boussinesq/Boussinesq system displayed in [**30**] and rigorously justified in Chapter 2. We study the well-posedness of such systems, and the asymptotic convergence of their solutions towards solutions of the full Euler system. Then, we provide a rigorous justification of the so-called KdV approximation, stating that any bounded solution of the full Euler system can be decomposed into four propagating waves, each of them being well approximated by the solutions of uncoupled Korteweg-de Vries equations. Our method also applies for models with the rigid lid assumption, using the Boussinesq/Boussinesq models introduced in [**17**]. Our explicit and simultaneous decomposition allows to study in details the behavior of the flow depending on the depth and density ratios, for both the rigid lid and free surface configurations. In particular, we consider the influence of the rigid lid assumption on the evolution of the interface, and

specify its domain of validity. Finally, solutions of the Boussinesq/Boussinesq systems and the KdV approximation are numerically computed, using a Crank-Nicholson scheme with a predictive step inspired from **[12, 13]**.

1. Introduction

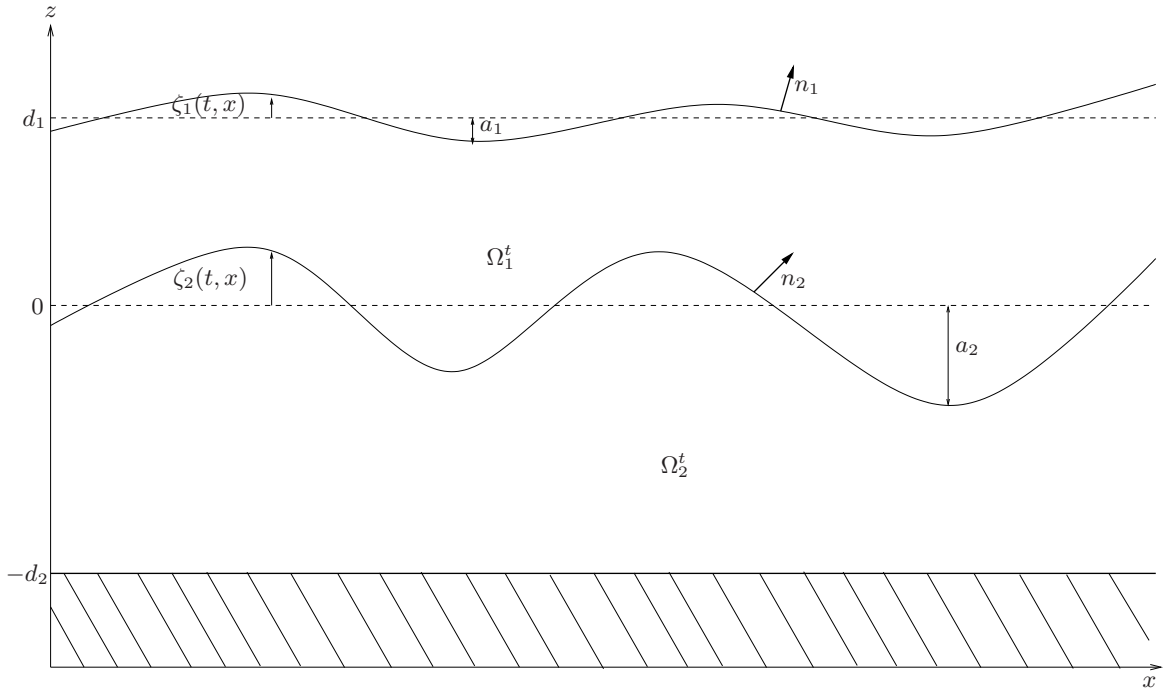


Figure 1. Sketch of the domain.

1.1. Motivation of the problem. This chapter deals with different asymptotic models for the propagation of weakly nonlinear internal waves in a two-fluid system. The system we study consists in two layers of immiscible, homogeneous, ideal, incompressible and irrotationnal fluids under the only influence of gravity. Since we are interested in KdV equations, which are unidirectional, we focus on the one-dimensional case, and the bottom is assumed to be flat (see Figure 1).

Let us denote by ρ_1 the density of the upper fluid and d_1 its depth, ρ_2 the density of the lower fluid, and d_2 its depth, a_1 the typical amplitude of the deformation of the surface and a_2 the one of the interface, and finally λ is a characteristic horizontal length, say the typical wavelength of the interface. Then the regime of the system is described by the following dimensionless parameters:

$$\epsilon_1 = \frac{a_1}{d_1}, \epsilon_2 = \frac{a_2}{d_1} \in (0,1),\ \mu = \frac{d_1^2}{\lambda^2} \in (0,+\infty),\ \delta = \frac{d_1}{d_2} \in (0,+\infty),\ \gamma = \frac{\rho_1}{\rho_2} \in (0,1).$$

The governing equations of such a system, that we call "full Euler", have been obtained in **[40]**; we briefly recall the system in Section 1.3 below. This system is strongly nonlinear, and remains complex, for its direct study as well as for numerical computations. That is why the derivation of approximate asymptotic models has attracted lots of interests in the past decades. We focus here on the so-called *long wave regime*, where the dimensionless parameters ϵ_1, ϵ_2, and μ are small and of the same order:

$$\epsilon_1 \ \sim\ \epsilon_2 \ \sim\ \mu \ \ll\ 1.$$

The long wave regime for the one-fluid system (water wave) has been considerably studied, and has led to many approximate equations. Among them, of particular interest

are the Boussinesq systems (from the work of Boussinesq [**19, 20**]), and the KdV approximation (from Korteweg and de Vries [**62**]). The latter model states that any solution of the one-layer water wave problem in the long-wave limit splits up into two counter-propagating waves, each of them evolving independently as a solution of a KdV equation. A justification of such models has been investigated among others by Craig [**33**], Schneider-Wayne [**104**] and Bona-Colin-Lannes [**16**]. The study of internal waves followed quickly. When the surface is assumed to be fixed as a rigid lid, the KdV equations related to such a system have been formally obtained by Keulegan [**60**] and Long [**77**]. The related Boussinesq-type systems have been justified (among many other asymptotic models) by Bona-Lannes-Saut [**17**]. When the surface is not rigid and allowed to move as a free surface, it is known that there exist two different modes of wave motion, corresponding to different linear phase speeds (see Kakutani-Yamasaki [**56**], Leone-Segur-Hammack [**76**], Michallet-Barthélemy [**86**] and Craig-Guyenne-Kalisch [**35**] for example). Accordingly, the KdV approximation states that any deformation of the surface and/or the interface will split up into four waves, each of them being lead by KdV equations.

However, to our knowledge, the four KdV equations related to the problem have never been revealed *simultaneously*, neither rigorously justified with a convergence theorem. Such a precise and complete decomposition allows to directly compare the different models. In particular, as an application of our results, we present an in-depth study of the influence of the rigid lid assumption on the evolution of the interface, and therefore assert the domain of validity of such a hypothesis.

An intermediate result for the construction of the KdV approximation is the full justification (in the sense of well-posedness, and convergence of the solutions towards solutions of the full Euler system) of symmetric coupled models, that are equivalent to the Boussinesq-type models derived by Camassa-Choi [**30**], Craig-Guyenne-Kalisch [**35**] and Duchêne [**40**]. The latter systems are justified only by a consistency result, and the stronger properties of the symmetric models make them interesting by their own.

The construction and rigorous justification of the symmetric Boussinesq-type models and the KdV approximation are the main motivation of this article.

1.2. Main results and outline of the chapter. Our study applies the methods of [**16**] to the case of internal waves, and accordingly uses as direct supports the full Euler system and the Boussinesq/Boussinesq model obtained in [**40**]. The derivation of the full Euler system as governing equations of our problem is quickly explained in Section 1.3. The full Euler system is consistent with the Boussinesq/Boussinesq model at order $\mathcal{O}(\mu^2)$, provided that $\epsilon_1,\ \epsilon_2 = \mathcal{O}(\mu)$; this result is recalled and precisely stated in Proposition 2.1, below.

The first step of our study lies in the construction of symmetric systems, obtained from the original Boussinesq/Boussinesq model by using a first order symmetrizer, and withdrawing the $\mathcal{O}(\epsilon^2)$ terms. Section 2 is dedicated to the construction, analysis and justification of such models. The systems we obtain are of the form

$$\Big(S_0 + \epsilon\big(S_1(U) - S_2\partial_x^2\big)\Big)\partial_t U + \Big(\Sigma_0 + \epsilon\big(\Sigma_1(U) - \Sigma_2\partial_x^2\big)\Big)\partial_x U = 0,$$

with $U(t,x) \in \mathbb{R}^4$, symmetric matrices S_0, Σ_0, S_2, $\Sigma_2 \in \mathcal{M}_4(\mathbb{R})$ (S_0 and S_2 being definite positive), and linear mappings $S_1(\cdot)$ and $\Sigma_1(\cdot)$ with values in real symmetric matrices. The well-posedness of such systems over times of order $\mathcal{O}(1/\epsilon)$ is stated in Proposition 2.6. The convergence of their solutions towards solutions of the full Euler system is then proved to be of order $\mathcal{O}(\epsilon)$, for times of order $\mathcal{O}(1/\epsilon)$, in Proposition 2.8.

From these models, using the classical WKB method, we prove in Section 3 that a rougher approximation consists in *four uncoupled KdV equations*, that is to say that any

bounded solution of the full Euler system in the long-wave limit splits up into four propagating waves, two of them moving to the right at different speed, and the other two moving to the left, each of them approximated by independent solutions of KdV equations. Our main result is Theorem 3.1, which states explicitly the decomposition, and precise the convergence rate between bounded solutions of the full Euler system and the solutions of the KdV approximation. The case of the rigid-lid configuration can be treated in the same way, and is quickly tackled in Section 3.3.

The complete, simultaneous decomposition of the flow, and its rigorous justification, is a key point for the comparison with different models. In particular, we present in Section 3.4 an in-depth study of the behavior of the flow predicted by the KdV approximation, depending of the density ratio γ and the depth ratio δ, for both the rigid lid and free surface configurations. As a result, we show that the rigid lid hypothesis is valid only for small density differences between the two fluids. To our knowledge, this fact has never been established, thought it has been addressed for example in **[60, 86]**.

Finally, the Boussinesq/Boussinesq models and the KdV approximation are numerically computed and compared for different initial data and parameters in Section 4. The numerical schemes we use are based on Crank-Nicholson methods, with a predictive step in order to deal with the nonlinearities. Such a scheme has been introduced by Besse and Bruneau in **[12, 13]**; it is formally of order 2 in space and time, and appears to be unconditionally stable. The precise schemes in our framework are presented in Section 4.1.

1.3. The full Euler system. Let us recall here briefly the governing equations of our system (see **[40]** for more details). The velocity fields of an irrotational flow can be expressed as gradients of potentials (that we call ϕ_1 for the upper fluid and ϕ_2 for the lower fluid). The conservation of mass, together with the incompressibility of the fluids, yields Laplace equation for the potentials. The momentum equations of the Euler equations can then be integrated, which yields the Bernoulli equation in terms of potentials. The system is finally closed by kinematic boundary conditions (stating that no particles of fluid cross the bottom, the surface or the interface) and pressure laws (the pressure is assumed to be constant at the surface, and continuous at the interface, see Remark 1.2 below). Thanks to an appropriate scaling, the two-layer full Euler system can be written in dimensionless form.

The key point is then to remark that the system can be reduced into four evolution equations coupling Zakharov's canonical variables **[122]**, specifically the deviation of the free surface and interface from their rest position (respectively ζ_1, ζ_2), and the trace of the velocity potentials of the upper fluid evaluated at the surface (namely ψ_1) and of the lower fluid evaluated at the interface (namely ψ_2). The system is then given by

$$
\left\{
\begin{array}{l}
\alpha\partial_t\zeta_1 - \frac{1}{\mu}G_1(\psi_1,\psi_2) = 0,\\
\partial_t\zeta_2 - \frac{1}{\mu}G_2\psi_2 = 0,\\
\partial_t\partial_x\psi_1 + \alpha\partial_x\zeta_1 + \frac{\epsilon_2}{2}\partial_x(|\partial_x\psi_1|^2) - \mu\epsilon_2\partial_x\mathcal{N}_1 = 0,\\
\partial_t(\partial_x\psi_2 - \gamma H(\psi_1,\psi_2)) + (1-\gamma)\partial_x\zeta_2 + \frac{\epsilon_2}{2}\partial_x(|\partial_x\psi_2|^2 - \gamma|H(\psi_1,\psi_2)|^2)\\
\qquad\qquad -\mu\epsilon_2\partial_x\mathcal{N}_2 = 0,
\end{array}
\right. \tag{1.1}
$$

with $\alpha = \frac{\epsilon_1}{\epsilon_2}$ and where $\mathcal{N}_1$ and $\mathcal{N}_2$ are given by the formulae

$$
\mathcal{N}_1 \equiv \frac{(\frac{1}{\mu}G_1(\psi_1,\psi_2) + \epsilon_1\partial_x\zeta_1\partial_x\psi_1)^2}{2(1+\mu|\epsilon_1\partial_x\zeta_1|^2)},
$$

$$
\mathcal{N}_2 \equiv \frac{(\frac{1}{\mu}G_2\psi_2 + \epsilon_2\partial_x\zeta_2\partial_x\psi_2)^2 - \gamma(\frac{1}{\mu}G_2\psi_2 + \epsilon_2\partial_x\zeta_2 H(\psi_1,\psi_2))^2}{2(1+\mu|\epsilon_2\partial_x\zeta_2|^2)},
$$

and with G_1 and G_2 the Dirichlet-Neumann operators and H the interface operator, defined as follows.

DEFINITION 1.1. *For ζ_1, $\zeta_2 \in W^{2,\infty}(\mathbb{R})$ and $\partial_x\psi_1$, $\partial_x\psi_2 \in H^{1/2}(\mathbb{R})$, the operators G_1, G_2 and H are defined by*

$$G_2[\epsilon_2\zeta_2]\psi_2 \equiv (\partial_z\phi_2 - \mu\epsilon_2\partial_x\zeta_2\partial_x\phi_2)\Big|_{z=\epsilon_2\zeta_2} .$$
$$G_1[\epsilon_1\zeta_1, \epsilon_2\zeta_2](\psi_1, \psi_2) \equiv (\partial_z\phi_1 - \mu\epsilon_1\partial_x\zeta_1\partial_x\phi_1)\Big|_{z=1+\epsilon_1\zeta_1} ,$$
$$H[\epsilon_1\zeta_1, \epsilon_2\zeta_2](\psi_1, \psi_2) \equiv \partial_x\big(\phi_1\Big|_{z=\epsilon_2\zeta_2}\big),$$

with ϕ_1 and ϕ_2 the unique solutions of the boundary problems

(1.2)

$$\begin{cases} \Delta^\mu_{x,z}\phi_2 = 0 & \text{in } \Omega_2^t, \\ \phi_2 = \psi_2 & \text{on } \{z = \epsilon_2\zeta_2\}, \\ \partial_z\phi_2 = 0 & \text{on } \{z = -\frac{1}{\delta}\}, \end{cases} \quad \text{and} \quad \begin{cases} \Delta^\mu_{x,z}\phi_1 = 0 & \text{in } \Omega_1^t, \\ \phi_1 = \psi_1 & \text{on } \{z = 1 + \epsilon_1\zeta_1\}, \\ \partial_n\phi_1 = \frac{G_2[\epsilon_2\zeta_2]\psi_2}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}} & \text{on } \{z = \epsilon_2\zeta_2\}. \end{cases}$$

Here, we denote respectively by $\Omega_1^t = \{(x,z) \in \mathbb{R}^2, \epsilon_2\zeta_2(t,x) < z < 1 + \epsilon_1\zeta_1(t,x)\}$ and by $\Omega_2^t = \{(x,z) \in \mathbb{R}^2, -\frac{1}{\delta} < z < \epsilon_2\zeta_2(t,x)\}$ the domains of the fluids, by $\Delta^\mu_{x,z} \equiv \mu\partial_x^2 + \partial_z^2$ the scaled Laplace operator, and by ∂_n the upward conormal derivative:

$$(\partial_n\phi)\Big|_{z=\epsilon_2\zeta_2} = \frac{1}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}}(\partial_z\phi - \mu\epsilon_2\partial_x\zeta_2\partial_x\phi)\Big|_{z=\epsilon_2\zeta_2} .$$

The domains of the two fluids are assumed to remain strictly connected, *i.e.* there exists $h_{\min} > 0$ such that

(1.3)

$$\forall x \in \mathbb{R}, \quad h_1(x) \equiv 1 + \epsilon_1\eta_1(x) \ \geq \ h_{\min} \ > \ 0 \quad \text{and} \quad h_2(x) \equiv \frac{1}{\delta} + \epsilon_2\eta_2(x) \ \geq \ h_{\min} \ > \ 0.$$

This assumption is necessary in order to obtain the consistency of the full Euler system (1.1) with the Boussinesq/Boussinesq model (2.1), as seen in [**40**], and recalled in Proposition 2.1 below. We do not always precise this assumption thereafter.

REMARK 1.2. *Even if the Cauchy problem associated to the Euler system at the interface of two fluids of different positive densities is known to be ill-posed in Sobolev spaces in the absence of surface tension (as Kelvin-Helmholtz instabilities appear), Lannes* [**70**] *proved thanks to a stability criterion that adding a small amount of surface tension guarantees the well-posedness of such a problem, with a time of existence that does not vanish as the surface tension goes to zero, and thus is consistent with the observations. The stability criterion states that the Kelvin-Helmholtz instabilities appear for high frequencies, where the regularization effect of the surface tension is relevant, while the main profile of the wave that we want to capture is located in lower frequencies, and is unaffected by surface tension. Consequently, we decide to neglect the surface tension term, as its effect does not appear in our asymptotic models.*

Furthermore, we know from Theorem 5.8 of [**70**] *that in the long wave regime, there exists uniformly bounded solutions of the full Euler system for times of order $\mathcal{O}(1/\mu)$ (again with a small amount of surface tension) in the rigid lid case. With this result in mind, we assume in the following that smooth, uniformly bounded families of solutions to* (1.1)*, whose existence is assumed in Proposition 2.8 and in Theorem 3.1, indeed exist.*

2. Derivation and analysis of Boussinesq/Boussinesq models

As said in the introduction, the starting point of our study is the Boussinesq/Boussinesq model, obtained from (1.1) thanks to an asymptotic expansion of the operators G_1, G_2 and H (see [**40**]). In order to simplify the notations, we assume that the small parameters of

the long wave regime are equal (the general case can easily be obtained by modifying some constants) and set

$$\epsilon_1 \;=\; \epsilon_2 \;=\; \mu \;\equiv\; \epsilon \;\ll\; 1.$$

The Boussinesq/Boussinesq system, in the one-dimensional case, can then be written using the convenient set of variables $(\eta_1, \eta_2, u_1, u_2) \;\equiv\; (\zeta_1 - \zeta_2, \zeta_2, \partial_x\psi_1, \partial_x\psi_2)$ as

$$\left\{\begin{array}{l} \partial_t\eta_1 + \partial_x(h_1u_1) = 0, \\ \partial_t\eta_2 + \partial_x(h_2u_2) = 0, \\ \partial_t u_1 + \partial_x(\eta_1+\eta_2) + \epsilon\big(u_1\partial_x u_1 - \dfrac{1}{3}\partial_x^2\partial_t u_1 - \dfrac{1}{2\delta}\partial_x^2\partial_t u_2\big) = 0, \\ \partial_t u_2 + \partial_x(\gamma\eta_1+\eta_2) + \epsilon\big(u_2\partial_x u_2 - \dfrac{1+3\gamma\delta}{3\delta^2}\partial_x^2\partial_t u_2 - \dfrac{\gamma}{2}\partial_x^2\partial_t u_1\big) = 0, \end{array}\right. \tag{2.1}$$

where $h_1 \;\equiv\; 1+\epsilon\eta_1$ and $h_2 \;\equiv \frac{1}{\delta} + \epsilon\eta_2$ are the respective nondimensionalized depths of the upper and lower layer. Let us recall the following consistency result, that have been obtained in [**40**, Proposition 2.15]

Proposition 2.1. *Let $U \;\equiv\; (\eta_1, \eta_2, u_1, u_2)$ be a strong solution of system (1.1), bounded in $L^{1,\infty}_t([0,T]; H^{s+t_0})$ with $s > 1$ and $t_0 \geq 9/2$, and such that (1.3) is satisfied. Then U satisfies (2.1) up to a residual R bounded by*

$$\big|R\big|_{L^\infty H^s} \;\leq\; \epsilon^2 C_0\Big(\frac{1}{h_{\min}}, \big|U\big|_{L^{1,\infty}_t H^{s+t_0}}\Big).$$

Remark 2.2. *Here, and in the following, we denote by $C_0(\lambda_1, \lambda_2, \ldots)$ any positive constant, depending on the parameters $\lambda_1, \lambda_2, \ldots$, and whose dependence on λ_j is assumed to be nondecreasing. Moreover, for $0 < T \leq \infty$ and $f(t,x)$ a function defined on $[0,T]\times\mathbb{R}$, we write $f \in L^\infty([0,T]; H^s)$ if f is uniformly (with respect to $t \in [0,T]$) bounded in $H^s = H^s(\mathbb{R})$ the L^2-based Sobolev space. Finally, one has $f \in L^{1,\infty}_t([0,T]; H^s)$ if $f \in L^\infty([0,T]; H^s)$ and $\partial_t f \in L^\infty([0,T]; H^{s-1})$. Their respective norm is denoted $\big|\cdot\big|_{L^\infty H^s}$ and $\big|\cdot\big|_{L^{1,\infty}_t H^s}$.*

2.1. A new family of symmetric models. In order to prove the well-posedness of hyperbolic systems as (2.1), one can use energy methods, that require symmetries of the system. Although our system is not entirely symmetrizable, we show in this section that it is equivalent (in the sense of consistency) at order $\mathcal{O}(\epsilon^2)$ to a system of the form

$$\Big(S_0 + \epsilon\big(S_1(U) - S_2\partial_x^2\big)\Big)\partial_t U + \Big(\Sigma_0 + \epsilon\big(\Sigma_1(U) - \Sigma_2\partial_x^2\big)\Big)\partial_x U = 0, \tag{2.2}$$

which satisfies the following crucial properties:

Assumption 2.3. i. *The matrices S_0, Σ_0, S_2, $\Sigma_2 \in \mathcal{M}_4(\mathbb{R})$ are symmetric.*
ii. *$S_1(\cdot)$ and $\Sigma_1(\cdot)$ are linear mappings with values in $\mathcal{M}_4(\mathbb{R})$, and for all $U \in \mathbb{R}^4$, the matrices $S_1(U)$ and $\Sigma_1(U)$ are symmetric.*
iii. *S_0 and S_2 are definite positive.*

Remark 2.4. *We sometimes write the system (2.2) under the form*

$$P_\epsilon(U, \partial_x)\partial_t U + Q_\epsilon(U, \partial_x)\partial_x U = 0, \tag{2.3}$$

with $P_\epsilon(U,\partial) = S_0 + \epsilon\big(S_1(U) - S_2\partial^2\big)$ and $Q_\epsilon(U,\partial) = \Sigma_0 + \epsilon\big(\Sigma_1(U) - \Sigma_2\partial^2\big)$.

First of all, let us point out that system (2.1) is only one among many other Boussinesq-type systems. In [**14**], Bona, Chen and Saut studied in the one-layer case a three-parameter family of Boussinesq systems, which are all approximations to the full Euler equations at the same order (in the sense of consistency). The same structure applies also in the two-layer case, and we describe it quickly below. We then exhibit first order symmetrizers *adapted to each of the Boussinesq-type systems*, and leading to systems of the form (2.2).

All of these models are equivalent in the sense that the original Boussinesq/Boussinesq system (2.1) is consistent at order $\mathcal{O}(\epsilon^2)$ with any of the systems presented in this section, as stated in Proposition 2.5.

As a first step, one can use the following change of variables:

$$v_2 \equiv (1-\epsilon a_2\partial_x^2)^{-1}u_2, \quad \text{and} \quad v_1 \equiv (1-\epsilon b_1\partial_x^2)^{-1}(u_1+\epsilon a_1\partial_x^2 v_2), \tag{2.4}$$

with $a_1 \in \mathbb{R}$ and $a_2, b_1 \geq 0$, and one recovers the three-parameter family of Boussinesq/Boussinesq systems introduced in [**40**].

Then, as it has been achieved in [**14**], one can also use the classical BBM-trick [**10**], and inherit new choices as parameters. This trick is based on the following calculations: since we have from (2.1) at first order

$$\begin{array}{ll} \partial_t\eta_1 = -\partial_x v_1 + \mathcal{O}(\epsilon), & \partial_t\eta_2 = -\frac{1}{\delta}\partial_x v_2 + \mathcal{O}(\epsilon), \\ \partial_t v_1 = -\partial_x(\eta_1+\eta_2) + \mathcal{O}(\epsilon), & \partial_t v_2 = -\partial_x(\gamma\eta_1+\eta_2) + \mathcal{O}(\epsilon), \end{array} \tag{2.5}$$

we can deduce the following, with the parameters $\lambda_i \in [0,1]$ $(i = 1\dots 4)$,

$$\begin{aligned} \partial_x^3 v_1 &= \lambda_1\partial_x^3 v_1 - (1-\lambda_1)\partial_x^2\partial_t\eta_1 + \mathcal{O}(\epsilon), \\ \partial_x^3 v_2 &= \lambda_2\partial_x^3 v_2 - \delta(1-\lambda_2)\partial_x^2\partial_t\eta_2 + \mathcal{O}(\epsilon), \\ \partial_x^2\partial_t v_1 &= (1-\lambda_3)\partial_x^2\partial_t v_1 - \lambda_3\partial_x^3(\eta_1+\eta_2) + \mathcal{O}(\epsilon), \\ \partial_x^2\partial_t v_2 &= (1-\lambda_4)\partial_x^2\partial_t v_2 - \lambda_4\partial_x^3(\gamma\eta_1+\eta_2) + \mathcal{O}(\epsilon). \end{aligned}$$

In the end, one obtains the following system, formally equivalent to (2.1) at order $\mathcal{O}(\epsilon^2)$:

$$\partial_t U + \mathcal{A}_0\partial_x U + \epsilon\big(\mathcal{A}(U)\partial_x U - \mathcal{A}_1\partial_x^3 U - \mathcal{A}_2\partial_x^2\partial_t U\big) = 0, \tag{2.6}$$

denoting $U = (\eta_1, \eta_2, v_1, v_2)$ and

$$\mathcal{A}_0 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/\delta \\ 1 & 1 & 0 & 0 \\ \gamma & 1 & 0 & 0 \end{pmatrix}, \quad \mathcal{A}(U) = \begin{pmatrix} v_1 & 0 & \eta_1 & 0 \\ 0 & v_2 & 0 & \eta_2 \\ 0 & 0 & v_1 & 0 \\ 0 & 0 & 0 & v_2 \end{pmatrix},$$

$$\mathcal{A}_1 = \begin{pmatrix} 0 & 0 & -\lambda_1\beta_1 & -\lambda_2\alpha_1 \\ 0 & 0 & 0 & -\lambda_2\frac{\alpha_2}{\delta} \\ -\lambda_3 b_1 - \lambda_4\gamma a_1 & -\lambda_3 b_1 - \lambda_4 a_1 & 0 & 0 \\ -\lambda_4\gamma\beta_2 - \lambda_3\frac{\gamma}{2} & -\lambda_4\beta_2 - \lambda_3\frac{\gamma}{2} & 0 & 0 \end{pmatrix},$$

$$\mathcal{A}_2 = \begin{pmatrix} (1-\lambda_1)\beta_1 & \delta(1-\lambda_2)\alpha_1 & 0 & 0 \\ 0 & (1-\lambda_2)\alpha_2 & 0 & 0 \\ 0 & 0 & (1-\lambda_3)b_1 & (1-\lambda_4)a_1 \\ 0 & 0 & (1-\lambda_3)\frac{\gamma}{2} & (1-\lambda_4)\beta_2 \end{pmatrix},$$

with the parameters $\beta_1 = \frac{1}{3} - b_1$, $\alpha_1 = \frac{1}{2\delta} - a_1$, $\alpha_2 = \frac{1}{3\delta^2} - a_2$ and $\beta_2 = a_2 + \frac{\gamma}{\delta}$, so that the system depends on the real parameters a_1, a_2, b_1, and λ_i $(i = 1\dots 4)$, that can be chosen freely.

In order to derive a system of the form (2.2), we exhibit a good symmetrizer of (2.6), that is to say

$$S(U) \equiv S_0 + \epsilon S_1(U) - \epsilon\tilde{S}_2\partial_x^2,$$

such that when we multiply (2.6) on the left by $S(U)$, and withdrawing the $\mathcal{O}(\epsilon^2)$ terms, we obtain a system (2.2) satisfying Assumption 2.3.

The symmetrization of the one-layer shallow water system is well known, and consists in multiplying the velocity equation by the water depth. An adaptation of this to our

two-layer model leads to

$$S_0 \equiv \begin{pmatrix} \gamma & \gamma & 0 & 0 \\ \gamma & 1 & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1/\delta \end{pmatrix} \quad \text{and} \quad S_1(U) \equiv \begin{pmatrix} 0 & 0 & \gamma v_1 & 0 \\ 0 & 0 & 0 & v_2 \\ \gamma v_1 & 0 & \gamma\eta_1 & 0 \\ 0 & v_2 & 0 & \eta_2 \end{pmatrix}, \tag{2.7}$$

so that S_0 and $\Sigma_0 \equiv S_0\mathcal{A}_0$ are symmetric, and for all $U \in \mathbb{R}^4$, the matrices $S_1(U)$ and $\Sigma_1(U) \equiv S_1(U)\mathcal{A}_0 + S_0\mathcal{A}(U)$ are symmetric. Moreover, S_0 is definite positive for $\delta > 0$ and $\gamma \in (0,1)$: its eigenvalues are

$$\gamma,\ 1/\delta,\ \text{and}\ \ 1/2(1+\gamma \pm \sqrt{(1-\gamma)^2+4\gamma^2}).$$

Then, one can check that when we set

$$\tilde{S}_2 \equiv \begin{pmatrix} a+(\delta-1+\gamma)b & 0 & 0 & 0 \\ a+\delta b & 0 & 0 & 0 \\ 0 & 0 & a+(\delta-1)b+\gamma(b_1\lambda_3-\lambda_1\beta_1) & b+\gamma\lambda_4 a_1 \\ 0 & 0 & \gamma(\frac{\lambda_3}{2\delta}-\lambda_2\alpha_1) & \frac{\lambda_4\beta_2-\lambda_2\alpha_2}{\delta} \end{pmatrix}$$

with $a = \gamma((1-\lambda_2)\alpha_2 - (1-\lambda_1)\beta_1)$ and $b = \gamma(1-\lambda_2)\alpha_1$, then $S_2 \equiv S_0\mathcal{A}_2 + \tilde{S}_2$ and $\Sigma_2 \equiv S_0\mathcal{A}_1 + \tilde{S}_2\mathcal{A}_0$ are symmetric. Then for any $K \in \mathbb{R}$, we can substitute $\tilde{S}_2 + KS_0$ for $\tilde{S}_2$, and one has again S_2 and Σ_2 are symmetric. Moreover, since S_0 is definite positive, one can choose K big enough for S_2 to be definite positive.

Therefore, when we multiply (2.6) by $S(U) = S_0 + \epsilon S_1(U) - \epsilon(\tilde{S}_2 + KS_0)\partial_x^2$, and withdrawing the $\mathcal{O}(\epsilon^2)$ terms, we obtain the perfectly symmetric system (2.2).

Using the above calculations, it is now straightforward to obtain the following consistency result:

Proposition 2.5. *Let $U = (\eta_1, \eta_2, u_1, u_2)$ be a strong solution of system (2.1) such that $V = (\eta_1, \eta_2, v_1, v_2)$, given by the change of variables (2.4), is uniformly bounded in $L_t^{1,\infty}([0,T]; H^{s+5})$ with $s > 1/2$. Then V satisfies (2.6) and (2.2) up to a residuals bounded by $\epsilon^2 C_0$ in the sense of $L^\infty H^s$ norm, with*

$$C_0 = C_0\big(a_1, a_2, b_1, K, \delta + \frac{1}{\delta}\big)\big|V\big|_{L_t^{1,\infty}([0,T];H^{s+5})}.$$

Proof. The first step in order to obtain (2.6) from (2.1) is to use the change of variables (2.4). Thus, when replacing in u_1 by $v_1 - \epsilon b_1\partial_x^2 v_1 - \epsilon a_1\partial_x^2 v_2$ and u_2 by $v_2 - \epsilon a_2\partial_x^2 v_2$, we obtain straightforwardly that V satisfies (2.6) in the case $1-\lambda_1 = 1-\lambda_2 = \lambda_3 = \lambda_4 = 0$, up to terms of the form

$$\epsilon^2\partial_x^4\partial_t f, \quad \epsilon^2\partial_x(f\partial_x^2 g) \quad \text{and} \quad \epsilon^3\partial_x\big((\partial_x^2 f)^2\big),$$

where f and g are components of V. Using the fact that V is bounded in $L_t^{1,\infty}([0,T); H^{s+5})$, and $H^s(\mathbb{R})$ is an algebra for $s > 1/2$, the remaining terms are clearly uniformly bounded by $\epsilon^2 C_0(a_1, a_2, b_1, K, \delta + \frac{1}{\delta},)\big|V\big|_{L_t^{1,\infty}H^{s+5}}$.

In the same way, when we substitute the relations of the BBM trick (2.5) into the third-derivative terms of the equations, we obtain (2.6) up to extra terms bounded by $\epsilon^2 C_0(a_1, a_2, b_1, \frac{1}{\delta})\big|V\big|_{L_t^{1,\infty}H^{s+5}}$.

Finally, in order to obtain (2.2), we multiply (2.6) by $S_0 + \epsilon S_1(V) - \epsilon(\tilde{S}_2 + KS_0)\partial_x^2$, and withdraw the terms

$$\epsilon^2\big(S_1(V) - (\tilde{S}_2 + KS_0)\partial_x^2\big)\big(\mathcal{A}(V)\partial_x V - \mathcal{A}_1\partial_x^3 V - \mathcal{A}_2\partial_t\partial_x^2 V\big).$$

Each of these terms are clearly bounded by $\epsilon^2 C_0$, since $H^s(\mathbb{R})$ is an algebra for $s > 1/2$, and

$$\big|A(V)\partial_x V\big|_{H^s} \le C_0\big|V\big|_{L^\infty H^{s+3}},\ \ \big|S_1(V)\big|_{L^\infty} \le C_0\big|V\big|_{L^\infty H^s},$$

$$\text{and} \quad \big|\mathcal{A}_1\partial_x^3 V\big|_{H^s} + \big|\mathcal{A}_2\partial_t\partial_x^2 V\big|_{H^s} \leq C_0\big|V\big|_{L_t^{1,\infty}H^{s+1}}$$

Therefore, the system (2.1) is consistent with the system (2.2) at the precision $\epsilon^2 C_0$. □

2.2. Well-posedness and convergence results. The system (2.2) is a symmetric hyperbolic system, and can be studied using classical energy methods. We first prove in Proposition 2.6 that such a system is well-posed over times of order $\mathcal{O}(1/\epsilon)$. The proof uses in particular an a priori estimate of the solution, in an adapted norm:

$$\big|U\big|^2_{H^{s+1}_\epsilon} \;\equiv\; \big|U\big|^2_{H^s} \;+\; \epsilon\big|U\big|^2_{H^{s+1}}.$$

Then, using a consistency result with energy estimates, we show that the solutions of our models converge towards bounded solutions of the full Euler system, assuming that such solutions exist (see Remark 1.2).

Here and thereafter, we fix $0 < \gamma_{\min} \leq \gamma \leq \gamma_{\max} < 1$ and $0 < \delta_{\min} \leq \delta \leq \delta_{\max} < +\infty$. The limit cases ($\delta \to 0, \infty$ and $\gamma \to 0, 1$) demand other scalings in the nondimensionalization than the ones presented in **[40]** (see Section A of **[70]** for example) and correspond to different regimes, such as the deep-water theory (from Benjamin **[9]** and Ono **[95]**), and lead to different models (see for example **[35, 17]** in the rigid lid configuration, and **[105, 61, 98]** in the free surface case). In all of these cases, the calculations of our justification break: the dependence of the constants $C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}) = C_0(\|S_0\|, \|S_0^{-1}\|)$ in the following theorems prevents the parameters to approach the limits.

In the same way, the following constants C_0 also depend on the set of parameters $(a_1, a_2, b_1, K, \lambda_i)$ $(i = 1 \cdots 4)$. We decide to fix these to constants once for all, and do not write explicitly this dependency, in order to simplify the notations.

We now state the well-posedness of our symmetric Boussinesq/Boussinesq model.

PROPOSITION 2.6. *Let $U^0 \in H^{s+1}$, with $s > 3/2$. Then there exists a constant $C_0 = C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}) > 0$ such that for $\epsilon \leq \epsilon_0 = (C_0\big|U^0\big|_{H^{s+1}_\epsilon})^{-1}$, there exists a time $T > 0$ independent of ϵ, and a unique solution $U \in C^0([0, T/\epsilon); H^{s+1}_\epsilon) \cap C^1([0, T/\epsilon); H^s_\epsilon)$ of the Cauchy problem (2.2) with $U\,|_{t=0} = U^0$.*

Moreover, one has the following estimate for $t \in [0, T/\epsilon]$:

$$\big|U\big|_{L^\infty([0,t];H^{s+1}_\epsilon)} + \big|\partial_t U\big|_{L^\infty([0,t];H^s_\epsilon)} \leq C_0 \frac{\big|U^0\big|_{H^{s+1}_\epsilon}}{1 - C_0\big|U^0\big|_{H^{s+1}_\epsilon}\epsilon t}. \tag{2.8}$$

We postpone the somewhat technical proof to Appendix A. The key ingredients of the proof are quickly presented in the following remark.

REMARK 2.7. *The condition $s > 3/2$ is necessary in order to obtain estimate (2.8), and thus the well-posedness over times of order $\mathcal{O}(1/\epsilon)$. One could obtain, using the exact same method as in the proof, the same result over times of order $\mathcal{O}(1)$, with the sharper assumption $s > 1/2$ (which is the standard regularity for one-dimensional hyperbolic systems).*

The smallness condition ($\epsilon\big|U^0\big|_{H^{s+1}_\epsilon} \leq {C_0}^{-1}$) is also necessary, for the nonlinear terms in $S_1(U)$ to remain negligible when compared with S_0. Indeed, under this condition, the energy of the system (2.2), defined by

$$E_s(U) \;\equiv\; 1/2(S_0\Lambda^s U, \Lambda^s U) \;+\; \epsilon/2(S_1(U)\Lambda^s U, \Lambda^s U) \;+\; \epsilon/2(S_2\Lambda^s\partial_x U, \Lambda^s\partial_x U),$$

is uniformly equivalent to the $\big|\cdot\big|_{H^{s+1}_\epsilon}$ norm, that is to say there exists $\alpha > 0$ such that

$$\frac{1}{\alpha}\big|U\big|^2_{H^{s+1}_\epsilon} \;\leq\; E_s(U) \;\leq\; \alpha\big|U\big|^2_{H^{s+1}_\epsilon}.$$

Moreover, the smallness condition on $\epsilon\big|U^0\big|_{H^{s+1}_\epsilon}$ *allows us to prove that the operator*

$$P_\epsilon(U,\partial) \;=\; S_0+\epsilon\big(S_1(U)-S_2\partial^2\big) : H^{s+1}\to H^{s-1}$$

is one-to-one and onto, and that $P_\epsilon(U,\partial)^{-1}Q_\epsilon(U,\partial)$ *is uniformly bounded* $H^s_\epsilon\to H^s_\epsilon$*. This leads to*

$$\big|\partial_t U\big|_{H^s_\epsilon}=\big|P_\epsilon(U,\partial)^{-1}Q_\epsilon(U,\partial)\partial_x U\big|_{H^s_\epsilon}\le C_0\big|U\big|_{H^{s+1}_\epsilon}.$$

Both of these properties are crucial in order to prove estimate (2.8)*.*

The existence and uniqueness of the solution of the Cauchy problem follow from the a priori estimate, using classical methods.

Using the previous Proposition, and the consistency of the full Euler system (1.1) with our symmetric Boussinesq/Boussinesq model (2.2), one can now easily deduce the following convergence Proposition:

Proposition 2.8. *Let* $s>3/2$*,* $\epsilon>0$ *and* $U=(\zeta_1,\zeta_2,\psi_1,\psi_2)$ *be a solution of the full Euler system* (1.1) *such that* $V=(\eta_1,\eta_2,v_1,v_2)\in C^0([0;T/\epsilon);H^{s+1}_\epsilon)\cap C^1([0;T/\epsilon);H^s_\epsilon)$ *defined by*

$$V\equiv(\zeta_1-\zeta_2,\zeta_2,(1-\epsilon b_1\partial_x^2)^{-1}(\partial_x\psi_1+\epsilon a_1\partial_x^2 v_2),(1-\epsilon a_2\partial_x^2)^{-1}\partial_x\psi_2)$$

is uniformly bounded in $L^{1,\infty}_t([0,T/\epsilon];H^{s+5})$ *and* (1.3) *is satisfied. We assume that* ϵ *satisfies the smallness condition of Proposition 2.6, and denote by* V_B *the solution of the symmetric Boussinesq/Boussinesq system* (2.2)*, with the same initial value* $V_B\,|_{t=0}=V\,|_{t=0}=V^0$ *and the same domain of existence. Then one has for all* $t\in[0,T/\epsilon]$*,*

$$\big|V-V_B\big|_{L^\infty([0,t];H^{s+1}_\epsilon)}\le\epsilon^2 tC_0,$$

with $C_0=C_0(\frac{1}{h_{\min}},\frac{1}{\gamma(1-\gamma)},\delta+\frac{1}{\delta},\big|V\big|_{L^{1,\infty}_t H^{s+5}},T)$*. In particular, one has*

$$\big|V-V_B\big|_{L^\infty([0,T/\epsilon];H^{s+1}_\epsilon)}\le\epsilon C_0,$$

with C_0 *independent of* ϵ*.*

Proof. Let us first point out that the full Euler system (1.1) is consistent with (2.2) at the precision ϵ^2C_0, that is to say that for any solution U of the full Euler system such that $V\in C^0([0;T/\epsilon);H^{s+1})\cap C^1([0;T/\epsilon);H^s)$ is uniformly bounded in $L^{1,\infty}_t H^{s+5}([0,T/\epsilon])$, then V satisfies (2.2) up to a residual bounded by ϵ^2C_0 in the sense of $L^\infty H^s$ norm, with $C_0=C_0(\frac{1}{h_{\min}},\delta+\frac{1}{\delta},\big|V\big|_{L^{1,\infty}_t H^{s+5}})$. Proposition 2.1 states that the full Euler system is consistent with (2.1) at the precision ϵ^2C_0, and Proposition 2.5 achieves the result.

Therefore, we know that V satisfies (2.2) up to $\epsilon^2 f$, with $f\in L^\infty([0,T/\epsilon];H^s)$, so that $R_s\equiv\Lambda^sV-\Lambda^sV_B$, Λ^s being the Fourier multiplier defined by $\widehat{\Lambda^s u}(\zeta)\equiv(1+|\zeta|^2)^{s/2}\widehat{u}(\zeta)$, satisfies the system

$$\begin{aligned}\big(S_0-\epsilon S_2\partial_x^2\big)\partial_t R_s+\epsilon\Lambda^s(S_1(V)\partial_t R_0+S_1(R_0)\partial_t V_B)+\big(\Sigma_0-\epsilon\Sigma_2\partial_x^2\big)\partial_x R_s\\ +\epsilon\Lambda^s(\Sigma_1(V)\partial_x R_0+\Sigma_1(R_0)\partial_x V_B)=\epsilon^2\Lambda^s f,\end{aligned}\tag{2.9}$$

with $\big|f\big|_{L^\infty([0,T/\epsilon];H^s)}\le C_0(\frac{1}{h_{\min}},\delta+\frac{1}{\delta},\big|V\big|_{L^{1,\infty}_t H^{s+5}})$.

We can then carry on the calculations of Section A.3, with the extra term $\epsilon^2\Lambda^s f$. We obtain that there exists $C_0(\frac{1}{\gamma(1-\gamma)},\delta+\frac{1}{\delta})$ such that as long as

$$\epsilon\big|R_0\big|_{H^{s+1}_\epsilon}\le 1/C_0,\tag{2.10}$$

one has the estimate

$$\frac{d}{dt}E(R_s)\;\le\;\epsilon C_0(\big|V_B\big|_{H^s}+\big|V\big|_{H^s})\big|R_0\big|^2_{H^s}\;+\;\epsilon^2(\Lambda^s f,\Lambda^s R_s),\tag{2.11}$$

with the energy $E(R_s)$ defined by

$$E(R_s) \equiv 1/2(S_0 R_s, R_s) + \epsilon/2(S_1(R_0)R_s, R_s) + \epsilon/2(S_2\partial_x R_s, \partial_x R_s).$$

Now, since S_0 and S_2 are definite positive, the condition (2.10) implies in particular

$$\frac{1}{C_0}\big|R_0\big|^2_{H^{s+1}_\epsilon} \leq E(R_s) \leq C_0\big|R_0\big|^2_{H^{s+1}_\epsilon}. \tag{2.12}$$

Thus, under this condition, and since V_B is uniformly bounded in $L^\infty([0,T/\epsilon];H^s)$ with respect to ϵ (from Proposition 2.6), one has

$$\frac{d}{dt}E(R_s) \leq \epsilon C_0 E(R_s) + \epsilon^2 C_0\big|f\big|_{H^s} E(R_s)^{1/2},$$

and Gronwall-Bihari's Lemma leads to

$$E(R_s)^{1/2} \leq C_0\epsilon\big|f\big|_{H^s}(e^{\epsilon C_0 t} - 1).$$

Finally, since $R_0\,|_{t=0} = 0$, and thanks to a continuity argument, there exists a time $T(C_0, \big|f\big|_{H^s}) > 0$ such that (2.10) holds for $0 \leq t \leq T/\epsilon$, and the estimate of the Proposition follows:

$$\big|R_0\big|_{H^{s+1}_\epsilon} \leq C_0 E(R_s)^{1/2} \leq C_0\epsilon^2\big|f\big|_{H^s} t.$$

□

3. The KdV approximation

In this section, we offer a rigorous justification of the so-called KdV approximation, from a class of symmetric systems that contains the symmetric Boussinesq/Boussinesq system (2.2). The KdV approximation consists in a decomposition of the flow into four parts, each of the components being lead by a Korteweg-de Vries equation. The construction of the KdV approximation is precisely explained in Section 3.1. Then, in Section 3.2, we obtain the convergence rate between the solutions of the coupled systems and the solutions defined by the KdV approximation. As a consequence, when we combine this result with the convergence Proposition 2.8, it follows immediately that any strong solution of the full Euler system existing over times $\mathcal{O}(1/\epsilon)$ and bounded in a sufficiently high Sobolev norm, is well approximated by the KdV approximation. More precisely, we state the following:

THEOREM 3.1. *Let $s > 3/2$ and $U = (\zeta_1, \zeta_2, \psi_1, \psi_2)$ be a solution of the full Euler system (1.1) such that $V = (\eta_1, \eta_2, v_1, v_2) \in C^0([0;T/\epsilon);H^{s+1}_\epsilon) \cap C^1([0;T/\epsilon);H^s_\epsilon)$ defined by*

$$V \equiv (\ \zeta_1 - \zeta_2\ ,\ \zeta_2\ ,\ (1-\epsilon b_1\partial_x^2)^{-1}(\partial_x\psi_1 + \epsilon a_1\partial_x^2 v_2)\ ,\ (1-\epsilon a_2\partial_x^2)^{-1}\partial_x\psi_2\ )$$

is uniformly bounded in $L^{1,\infty}_t([0,T/\epsilon];H^{s+5})$ and (1.3) is satisfied. Then there exists[1] *$(\boldsymbol{e}_1,\dots,\boldsymbol{e}_4)$ a basis of $\mathbb{R}^4$, and coefficients c_i, λ_i, μ_i ($i = 1\cdots 4$) such that, denoting by u_i the solution of the KdV equation*

$$\partial_t u_i + c_i\partial_x u_i + \epsilon\lambda_i u_i\partial_x u_i + \epsilon\mu_i\partial_x^3 u_i = 0 \tag{3.1}$$

with $u_i\,|_{t=0} = \boldsymbol{e}_i \cdot S_0 V\,|_{t=0}$ (S_0 defined in (2.7)), one has for all $t \in [0,T/\epsilon]$,

$$\big|V - \sum_{i=1}^4 u_i\boldsymbol{e}_i\big|_{L^\infty([0,t];H^{s+1}_\epsilon)} \leq \epsilon\sqrt{t}C_0,$$

with $C_0 = C_0(\frac{1}{h_{\min}}, \frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|V\big|_{L^{1,\infty}_t H^{s+5}})$.

[1] one has explicit expressions for the basis $(\mathbf{e}_1,\dots,\mathbf{e}_4)$ and the coefficients c_i, λ_i, μ_i, that we display in Remark 3.5 page 97.

Moreover, if V satisfies $(1+x^2)V\,|_{t=0} \in H^{s+4}$, then one has the better estimate

$$\big|V - \sum_{i=1}^{4} u_i \boldsymbol{e}_i\big|_{L^\infty([0,T/\epsilon];H^{s+1}_\epsilon)} \leq \epsilon C_0',$$

with $C_0' = C_0(\frac{1}{h_{\min}}, \frac{1}{\gamma(1-\gamma)}, \delta+\frac{1}{\delta}, \big|V\big|_{L^{1,\infty}_t H^{s+5}}, \big|(1+x^2)V\,|_{t=0}\,\big|_{H^{s+4}})$.

Proof. The proof proceeds from different results of the chapter; the completion is as follows. In Proposition 3.12, we prove the convergence between the solutions of systems of the form (3.2) (thus containing symmetric Boussinesq/Boussinesq systems (2.2)) and the approximate solution $U_{\rm app} = \sum_{i=1}^4 u_i \mathbf{e}_i + \epsilon U_1$, defined in Definition 3.6. The residual U_1 is then estimated in Proposition 3.9. Finally, since we have from Proposition 2.8 the convergence between the solutions of the full Euler system (1.1) and the solutions of the symmetric Boussinesq/Boussinesq system (2.2) with a better rate, the Theorem follows; see also Remark 3.13 below. □

Remark 3.2. *The difference on the convergence rate for different sets of initial values is not simply a technical issue. Indeed, one can see in Figure 7, page 114, that if the condition of sufficient decreasing in space is not satisfied, the convergence will be worse than $\mathcal{O}(\epsilon)$. Requiring the initial data (and thus the solutions of the KdV equations) to lie in weighted Sobolev spaces guarantees that the nonlinear interaction between the four traveling waves can be neglected. As a matter of fact, this condition on the sufficient decay in space of the initial data appears also naturally for the KdV approximation of the one-layer problem, as we see in* **[104, 16]**.

3.1. Formal derivation. The class of system that we now study is the following:

$$\Big(S_0 + \epsilon\big(S_1(U) - S_2\partial_x^2\big)\Big)\partial_t U + \Big(\Sigma_0 + \epsilon\big(\Sigma_1(U) - \Sigma_2\partial_x^2\big)\Big)\partial_x U = 0, \tag{3.2}$$

with the following hypothesis:

Assumption 3.3. i. *The matrices S_0, Σ_0, S_2, $\Sigma_2 \in \mathcal{M}_4(\mathbb{R})$ are symmetric.*
ii. *$S_1(\cdot)$ and $\Sigma_1(\cdot)$ are linear mappings with values in $\mathcal{M}_4(\mathbb{R})$, and for all $U \in \mathbb{R}^4$, $S_1(U)$ and $\Sigma_1(U)$ are symmetric.*
iii. *S_0 is definite positive, and $S_0^{-1}\Sigma_0$ has four different non zero eigenvalues c_i ($i = 1\dots 4$).*

Remark 3.4. *The symmetric Boussinesq/Boussinesq systems* (2.2) *derived in Section 2.1 immediately satisfy Assumption 3.3, with*

$$c_i = \pm\sqrt{\frac{1+\delta \pm \sqrt{(1-\delta)^2+4\gamma\delta}}{2\delta}}.$$

Following the classical WKB method, we look for an approximate solution of the Cauchy problem (3.2) with initial data U^0 under the form

$$U_{\rm app}(t,x) = U_0(\epsilon t, t, x) + \epsilon U_1(\epsilon t, t, x),$$

with the profiles $U_0(\tau,t,x)$ and $\epsilon U_1(\tau,t,x)$ satisfying $U_0\,|_{t=\tau=0} = U^0$ et $U_1\,|_{t=\tau=0} = 0$.

We plug the Ansatz into (3.2), and obtain

$$\begin{aligned}(S_0\partial_t + \Sigma_0\partial_x)U_0 + \epsilon S_0\partial_\tau U_0 + \epsilon\big(S_1(U_0)\partial_t U_0 + \Sigma_1(U_0)\partial_x U_0 - S_2\partial_x^2\partial_t U_0 - \Sigma_2\partial_x^3 U_0\big)\\ +\epsilon(S_0\partial_t + \Sigma_0\partial_x)U_1 + \epsilon^2 R = 0.\end{aligned} \tag{3.3}$$

We now deduce the equations satisfied by $U_0(\tau,t,x)$ and $U_1(\tau,t,x)$, solving (3.3) at each order.

At order $\mathcal{O}(1)$: We solve

$$(S_0\partial_t + \Sigma_0\partial_x)U_0 = 0. \tag{3.4}$$

We assumed that S_0 is symmetric definite positive, and hence induce a scalar product:

$$\langle u, v\rangle \equiv u \cdot S_0 v = u^T S_0 v.$$

Since $S_0^{-1}\Sigma_0$ is real and symmetric for the scalar product $\langle\cdot,\cdot\rangle$, it is diagonalizable in an orthonormal basis. We denote by $\mathbf{e}_i$ ($i = 1\dots 4$) the basis vectors, which are the unitary eigenvectors of $S_0^{-1}\Sigma_0$. By definition, they satisfy for $1 \le i, j \le 4$,

$$\langle \mathbf{e}_i, S_0^{-1}\Sigma_0\mathbf{e}_j\rangle \;\equiv\; \mathbf{e}_i\cdot\Sigma_0\mathbf{e}_j \;=\; c_i\delta_{i,j}, \quad \text{and} \quad \langle \mathbf{e}_i, \mathbf{e}_j\rangle \;\equiv\; \mathbf{e}_i\cdot S_0\mathbf{e}_j \;=\; \delta_{i,j},$$

with $\delta_{i,j}$ the classical Kronecker delta symbol. Therefore, when we define $u_i \equiv \mathbf{e}_i \cdot S_0 U_0$ (and hence $U_0 = \sum_{i=1}^{4} u_i\mathbf{e}_i$), multiplying (3.4) on the left by $\mathbf{e}_i$, we obtain

$$(\partial_t + c_i\partial_x)u_i \;=\; 0$$

for all $i = 1, \dots, 4$. Finally, since u_i satisfies a transport equation, we use the notation

$$u_i(\tau, t, x) \;=\; u_i(\tau, x - c_i t) \;=\; u_i(\tau, x_i), \tag{3.5}$$

with initial data $u_i(0, x_i) = \mathbf{e}_i \cdot S_0 U^0(x_i)$.

At order $\mathcal{O}(\epsilon)$: We solve

$$S_0\partial_\tau U_0 + \Sigma_1(U_0)\partial_x U_0 + S_1(U_0)\partial_t U_0 - \Sigma_2\partial_x^3 U_0 - S_2\partial_x^2\partial_t U_0 + (S_0\partial_t + \Sigma_0\partial_x)U_1 = 0, \tag{3.6}$$

that we can split[2] in

$$\partial_\tau u_i \;+\; \lambda_i u_i \partial_{x_i} u_i \;+\; \mu_i \partial^3_{x_i} u_i \;=\; 0, \tag{3.7}$$

with $\lambda_i \equiv \mathbf{e}_i \cdot \big(\Sigma_1(\mathbf{e}_i) - c_i S_1(\mathbf{e}_i)\big)\mathbf{e}_i$ and $\mu_i \equiv \mathbf{e}_i \cdot \big(-\Sigma_2 + c_i S_2\big)\mathbf{e}_i$; and in the other hand,

$$(\partial_t + c_i\partial_x)\langle \mathbf{e}_i, U_1\rangle + \sum_{(j,k)\neq(i,i)} \alpha_{ijk} u_k(\tau, x - c_k t)\partial_x u_j(\tau, x - c_j t) = \sum_{j\neq i} \beta_{ij}\partial_x^3 u_j(\tau, x - c_j t), \tag{3.8}$$

with $\alpha_{ijk} \equiv \mathbf{e}_i \cdot (\Sigma_1(\mathbf{e}_k) - c_j S_1(\mathbf{e}_k))\mathbf{e}_j$ and $\beta_{ij} \equiv \mathbf{e}_i \cdot (\Sigma_2 - c_j S_2)\mathbf{e}_j$.

It is clear that u_i satisfies (3.5) and (3.7), if and only if $u_i(\epsilon t, t, x)$ satisfies the Korteweg-de Vries equation of Theorem 3.1:

$$\partial_t u_i \;+\; c_i\partial_x u_i \;+\; \epsilon\lambda_i u_i \partial_x u_i \;+\; \epsilon\mu_i\partial_x^3 u_i \;=\; 0. \tag{3.9}$$

REMARK 3.5. *In the specific case of symmetric Boussinesq/Boussinesq systems* (2.2)*, one obtains the following values for the coefficients:*

$$(c_i)_{i\in\{1\dots4\}} = (c_+, -c_+, c_-, -c_-), \qquad \text{with } c_\pm^2 = \frac{1+\delta \pm \sqrt{(1-\delta)^2 + 4\gamma\delta}}{2\delta},$$

$$(\lambda_i)_{i\in\{1\dots4\}} = (\lambda_+, \lambda_+, \lambda_-, \lambda_-), \qquad \text{with } \lambda_\pm = \frac{3}{2}\frac{(2-\delta)c_\pm^2 + \delta - \frac{1}{\delta} - (1-\gamma)}{\Theta_\pm(c_+^2 - c_-^2)},$$

$$(\mu_i)_{i\in\{1\dots4\}} = (\mu_+, -\mu_+, \mu_-, -\mu_-), \qquad \text{with } \mu_\pm = \frac{c_\pm}{6}\frac{(1 + \frac{3\gamma}{\delta} + \frac{1}{\delta^2})(c_\pm^2 - \frac{1-\gamma}{\delta+1}) - \frac{1}{\delta}c_\pm^2}{c_\pm^2 - 2\frac{1-\gamma}{\delta+1}},$$

[2]This splitting is in fact necessary. Indeed, the multiscale WKB expansion can be justified for times of order $\mathcal{O}(1/\epsilon)$ only if the growth of the corrector term U_1 is sublinear. As we see in Proposition 3.9, thanks to Lemma 3.8, the particular form of (3.8) allows to obtain a square-root growth, and even better if U^0 is sufficiently decreasing in space.

with $\Theta_\pm \equiv \sqrt{2\delta(c_+^2 - c_-^2)|c_\pm^2 - 1|}$. *The unitary eigenvectors of* $S_0^{-1}\Sigma_0$ *are given by*

$$\boldsymbol{e}_1 = \frac{1}{\Theta_+}\begin{pmatrix} \frac{1}{c_+} \\ c_+ - \frac{1}{c_+} \\ 1 \\ \delta c_+^2 - \delta \end{pmatrix}, \qquad \boldsymbol{e}_2 = \frac{1}{\Theta_+}\begin{pmatrix} -\frac{1}{c_+} \\ \frac{1}{c_+} - c_+ \\ 1 \\ \delta c_+^2 - \delta \end{pmatrix},$$

$$\boldsymbol{e}_3 = \frac{1}{\Theta_-}\begin{pmatrix} \frac{-1}{c_-} \\ \frac{1}{c_-} - c_- \\ -1 \\ \delta - \delta c_-^2 \end{pmatrix}, \qquad \boldsymbol{e}_4 = \frac{1}{\Theta_-}\begin{pmatrix} \frac{1}{c_-} \\ c_- - \frac{1}{c_-} \\ -1 \\ \delta - \delta c_-^2 \end{pmatrix}.$$

Let us first remark that these coefficients only depend on γ *and* δ, *so that neither the change of variables* (2.4) *nor the BBM-trick* (2.5) *affect the coefficients of the KdV approximation.*

It is also worth pointing out that the dispersion coefficients $\mu_\pm$ *cannot be zeros for* $\gamma \in (0,1)$ *and* $\delta > 0$. *Therefore, the KdV approximation cannot degenerate into Burgers-type equations.*

The coefficients of the KdV equation as an approximation to describe solitary internal waves in the free surface case have already been introduced in [**56**], and then in [**76, 82, 86**], and correspond to the ones we present here. In a slightly different regime, Craig, Guyenne and Kalisch [**35**] obtained a model that consists in four independent propagating waves, two of them satisfying the KdV equations of our slow mode waves (the other two being solutions of the first order transport equation $\partial_t u \pm c_+ \partial_x u = 0$). We state here that in the long wave regime, *any* bounded solution of the full Euler system can be decomposed into four propagating waves, each of them being well approximated by independent solutions of four KdV equations. This simultaneous decomposition of the flow is new, and allows to compare with other models, such as Boussinesq-type models.

3.2. Rigorous demonstration. The strategy is the following. Using the previous calculations, we define the approximate solution:

Definition 3.6. *Let* $U^0 \in H^{s+2}(\mathbb{R})$, *with* $s > 1/2$. *We call* approximate solution *of the system* (3.2) *any function* $U_{app}(t,x) \in C^0([0,T/\epsilon); H^s)^4$ *such that*

$$U_{app}(t,x) = \sum_{i=1}^{4} u_i(t,x)\boldsymbol{e}_i + \epsilon U_1(\epsilon t, t, x), \tag{3.10}$$

where $(u_i)_{i=1\dots4}$ *is the solution of the four uncoupled KdV equations* (3.9), *with* $u_i\,|_{t=0} = \boldsymbol{e}_i \cdot S_0 U^0$, *and the correcting term* U_1 *is a solution of* (3.8) *with* $U_1\,|_{\tau=t=0} = 0$.

We first prove that such solutions exist in the strong sense for sufficiently smooth initial data, over times of order $\mathcal{O}(1/\epsilon)$. Then we use estimates on U_0 and U_1, to obtain a consistency result. This result allows us to show that $U_{\rm app}(t,x)$ indeed approximates the solution of (3.2) with the same initial data, at least at order $\mathcal{O}(\epsilon^{3/2}t)$.

Proposition 3.7. *Let* $U^0 \in H^{s+2}(\mathbb{R})$, *with* $s > 1/2$. *Then, one has:*

i. *For all* $i = 1\dots4$, *there exists a unique strong solution of the Cauchy problem* (3.7), *with initial data* $u_i\,|_{t=0} = \boldsymbol{e}_i \cdot S_0 U^0$. *Moreover, there exists* $T > 0$, *such that one has the estimate*

$$\sum_{i=1}^{4} \left|u_i\right|_{L^\infty([0,T];H^{s+2})} \leq C_0\left(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}\right)\left|U^0\right|_{H^{s+2}}.$$

ii. *There exists a function* $U_1 \in L^\infty([0,T]\times\mathbb{R};H^s)$*, strong solution of* (3.8)*, with* $U_1\,|_{\tau=t=0}=0$.

PROOF. i. The existence and uniqueness of the solutions of (3.7) is classical: see **[59]** for the local well-posedness of the KdV equation, and **[32]** for the global well-posedness. One obtains the estimate as usual: as we multiply the equation (3.7) by $\Lambda^{2k}u_i$ (with $3/2<k\leq s+2$) and integrate with respect to the space variable, one obtains

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}}(\Lambda^k u_i)^2\mathrm{d}x=\left|\lambda_i\int_{\mathbb{R}}\Lambda^k(u_i\partial_x u_i)\Lambda^k u_i\mathrm{d}x\right|.$$

Thanks to the Kato-Ponce Lemma, we manage to estimate the right-hand side as follows:

$$\left|\int_{\mathbb{R}}\Lambda^k(u_i\partial_x u_i)\Lambda^k u_i\mathrm{d}x\right|\leq\left|\frac{1}{2}\int_{\mathbb{R}}\partial_x u_i(\Lambda^k u_i)^2\right|+\left|\int_{\mathbb{R}}[\Lambda^k,u_i]\partial_x u_i(\Lambda^k u_i)\mathrm{d}x\right|\leq C_0\big|u_i\big|_{H^k}^3,$$

and we conclude by applying Gronwall-Bihari's lemma, that reads

$$\big|u_i\big|_{H^k}\leq C_0\frac{\big|u_i\,|_{t=0}\,\big|_{H^k}}{1-C_0t\big|u_i\,|_{t=0}\,\big|_{H^k}},$$

so that the estimate of the proof follows for T sufficiently small, since

$$\big|u_i\,|_{t=0}\,\big|_{H^{s+2}}=\big|\mathbf{e}_i\cdot S_0U^0\big|_{H^{s+2}}\leq C_0\left(\frac{1}{\gamma(1-\gamma)},\delta+\frac{1}{\delta}\right)\big|U^0\big|_{H^{s+2}}.$$

Let us recall the notation (3.5): we have proved that $u_i(\tau,t,x)\in L^\infty([0,T]\times\mathbb{R};H^{s+2})$, with

$$\sum_{i=1}^4\big|u_i\big|_{L^\infty([0,T]\times\mathbb{R};H^{s+2})}\leq C_0\left(\frac{1}{\gamma(1-\gamma)},\delta+\frac{1}{\delta}\right)\big|U^0\big|_{H^{s+2}}.$$

ii. We can then exhibit U_1: let us write (3.8) under the form

$$(\partial_t+c_i\partial_x)\langle\mathbf{e}_i,U_1\rangle=\sum_{(j,k)\neq(i,i)}f_{ijk}(\tau,t,x)+\sum_{j\neq i}\partial_x g_{ij}(\tau,x-c_jt).$$

From the above estimate on u_i, one has $f_{ijk}\in L^\infty([0,T]\times\mathbb{R};H^{s+1})$, and $g_{ij}\in L^\infty([0,T];H^s)$. Hence, for $s>1/2$, one can set

$$\begin{array}{rl}\langle\mathbf{e}_i,U_1\rangle(\tau,t,x)=&\sum_{(j,k)\neq(i,i)}\int_0^t f_{ijk}(\tau,s,x+c_i(s-t))\mathrm{d}s\\&+\sum_{j\neq i}\frac{1}{c_i-c_j}\big(g_{ij}(\tau,x-c_jt)-g_{ij}(\tau,x-c_it)\big).\end{array}$$

and U_1 satisfies the hypotheses of the Proposition: $U_1(\tau,t,x)\in L^\infty([0,T]\times\mathbb{R};H^s)$ and $U_1\,|_{\tau=t=0}=0$.

□

We now prove that U_1, which is the corrector term defined by (3.8), and that contains all the coupling effects between the different components, obeys to a sublinear secular growth. The key point is given by the following Lemma, that proceeds from Propositions 3.2 and 3.5 of **[67]**:

LEMMA 3.8. *Let* u *be the solution of*

$$\left\{\begin{array}{l}(\partial_t+c\partial_x)u=g(v_1,v_2)\\ u\,|_{t=0}=0\end{array}\right.\quad\text{with}\quad\forall i\in\{1,2\},\quad\left\{\begin{array}{l}(\partial_t+c_i\partial_x)v_i=0\\ v_i\,|_{t=0}=v_i^0\end{array}\right.\tag{3.11}$$

with $c_1\neq c_2$, v_1^0, $v_2^0\in H^s(\mathbb{R})$, $s>1/2$, *and* g *is a bilinear mapping defined on* $\mathbb{R}^2$ *and with values in* $\mathbb{R}$. *Then one has the following estimates:*

i. *If* $c=c_1$ *, then* $\lim_{t\to\infty}\frac{1}{\sqrt{t}}\big|u(t,\cdot)\big|_{H^s(\mathbb{R})}=0$.

ii. *If* $c\neq c_1\neq c_2$*, then* $\frac{1}{\sqrt{t}}\big|u(t,\cdot)\big|_{H^s(\mathbb{R})}=\mathcal{O}(1)$.

Moreover, if there exists $\alpha > 1/2$ such that $v_1^0(1+x^2)^\alpha$, and $v_2^0(1+x^2)^\alpha \in H^s(\mathbb{R})$, then one has the better estimate

$$\big|u\big|_{L^\infty H^s(\mathbb{R})} \ \leq \ C_0\big|v_1^0(1+x^2)^\alpha\big|_{H^s(\mathbb{R})}\big|v_2^0(1+x^2)^\alpha\big|_{H^s(\mathbb{R})},$$

with $C_0 = C_0(c, c_1, c_2)$.

We are now able to give the following crucial estimate on U_1:

Proposition 3.9. *Let $s > 1/2$ and $U^0 \in H^{s+2}$. Then with $U_1 \in L^\infty([0,T]\times\mathbb{R}; H^s)$ a strong solution of (3.8) with $U_1\,|_{\tau=t=0} = 0$, one has the estimate:*

$$\big|U_1\big|_{L^\infty([0,T]\times[0,t];H^s)} \leq C_0\sqrt{t},$$

with $C_0 = C_0\big(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+2}}\big)$.

Moreover, if U^0 satisfies $U^0(1+x^2) \in H^{s+1}(\mathbb{R})$, then one has the uniform estimate

$$\big|U_1\big|_{L^\infty([0,T]\times\mathbb{R};H^s)} \leq C_0\big|U^0(1+x^2)\big|^2_{H^{s+1}},$$

with $C_0 = C_0\big(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+2}}\big)$.

Proof. Let us decompose U_1 as a sum of functions as in Proposition 3.7:

$$\begin{aligned}\langle \mathbf{e}_i, U_1\rangle(\tau,t,x) &= \sum_{(j,k)\neq(i,i)} \int_0^t f_{ijk}(\tau, s, x + c_i(s-t))\mathrm{d}s \\ &\quad + \sum_{j\neq i} \frac{1}{c_i - c_j}\big(g_{ij}(\tau, x - c_j t) - (g_{ij}(\tau, x - c_i t)\big) \\ &= \sum_{(j,k)\neq(i,i)} U^{ijk} + \sum_{j\neq i} V^{ij}.\end{aligned}$$

with the functions f_{ijk} and g_{ij} coming from (3.8) written in a simplified form:

$$(\partial_t + c_i\partial_x)\langle \mathbf{e}_i, U_1\rangle = \sum_{(j,k)\neq(i,i)} f_{ijk}(\tau,t,x) \ + \ \sum_{j\neq i} \partial_x g_{ij}(\tau, x - c_j t).$$

From Proposition 3.7, we know that the following bounds hold:

$$\forall \tau \in [0,T], \quad \sum_{(j,k)\neq(i,i)} \big|f_{ijk}\big|_{L^\infty([0,T]\times\mathbb{R};H^{s+1})} \ + \ \sum_{j\neq i} \big|g_{ij}\big|_{L^\infty([0,T];H^s)} \leq C_0\big|U^0\big|_{H^{s+2}},$$

with $C_0 = C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta})$. Therefore, one has

$$\forall j \neq i, \qquad \big|V^{ij}\big|_{L^\infty([0,T]\times\mathbb{R};H^s)} \ \leq \ C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta})\big|U^0\big|_{H^{s+2}}.$$

Moreover, again for $j \neq i$, we remark that f_{ijj} can be written as

$$f_{ijj}(\tau,t,x) \ \equiv \ \alpha_{ijj}u_j(\tau, x - c_j t)\partial_x u_j(\tau, x - c_j t) \ \equiv \ \partial_x h_{ij}(\tau, x - c_j t),$$

so that U^{ijj} has the same form as V^{ij}, and can be treated in the same way. And since $f_{ijj} \in L^\infty([0,T]\times\mathbb{R}; H^{s+1})$, U^{ijj} is uniformly bounded in H^s. The last terms that have to be bounded are U^{ijk}, for $(j,k) \neq (i,i)$ with $j \neq k$, U^{ijk}. One can easily check that U^{ijk} satisfies the hypothesis of Lemma 3.8, with $f_{ijk} = g(u_j, \partial_x u_k)$, for any $\tau \in [0,T]$. We then immediately deduce

$$\big|U_1\big|_{L^\infty([0,T]\times[0,t];H^s)} \leq \sqrt{t}C_0\Big(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+2}}\Big).$$

As for the second estimate of the proposition, let us first remark that the estimates of V^{ij} and U^{ijj} are time-independant, and in agreement with the improved estimate. Therefore, the only remaining terms we have to control are U^{ijk} with $j \neq k$. Of course, we

will use the second case of Lemma 3.8, but we have to check first that for every $\tau \in [0,T]$, the initial data $u_j(\tau,0,x)$ and $\partial_x u_k(\tau,0,x)$ are localized in space, that is

$$\forall \tau \in [0,T], \qquad \big|(1+x^2)u_j(\tau,0,x)\big|_{H^s} \ + \ \big|(1+x^2)\partial_x u_k(\tau,0,x)\big|_{H^s} \ < \ \infty.$$

This property is true at $\tau = 0$ (by hypothesis of the proposition), and is propagated to $\tau > 0$, using the fact that $u_i(\tau, x_i)$ satisfies the KdV equation (3.7). This propagation of the localization in space has been proved by Schneider and Wayne in [**104**, Lemma 6.4]. We do not recall the proof here, and use directly the statement:

LEMMA 3.10. *If $(1+x^2)U^0\,|_{t=0} \in H^{s+1}$, then there exists $C_1, \tilde{C}_1 > 0$ such that*

$$\big|(1+x^2)u_j(\tau,0,x)\big|_{H^{s+1}} \ \leq \ C_1\big|(1+x^2)u_j\,|_{\tau=t=0}\,\big|_{H^{s+1}} \ \leq \ \tilde{C}_1\big|(1+x^2)U^0\,|_{t=0}\,\big|_{H^{s+1}}.$$

This Lemma, together with the second estimate of Lemma 3.8, allows to control U^{ijk}, uniformly in time. Every term of the decomposition of U_1 has been controlled, and one has the following estimate:

$$\big|U_1\big|_{L^\infty([0,T]\times\mathbb{R};H^s)} \leq \ C_0\Big(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+2}}\Big)\big|(1+x^2)U^0\big|^2_{H^{s+1}}. \tag{3.12}$$

This concludes the proof. □

The next step consists in proving the consistency of our approximation with the symmetric system (3.2).

PROPOSITION 3.11. *If $U^0 \in H^{s+5}$ with $s > 1/2$, then $U_{app}(t,x)$ defined in Definition 3.6 satisfies the symmetric system (3.2) up to a residual of order $\mathcal{O}(\epsilon^{3/2})$ in $L^\infty([0,T/\epsilon];H^s)$.*

Moreover, U_0 satisfies $U^0(1+x^2) \in H^{s+4}(\mathbb{R})$, then the residual is uniformly bounded in $L^\infty([0,T/\epsilon];H^s)$ by $\epsilon^2 C_0$, with $C_0 = C_0\left(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|(1+x^2)U^0\big|_{H^{s+4}}, T\right)$.

PROOF. Plugging $U_{\rm app}(t,x)$ into (3.3), we see from the calculations of Section 3.1 that the only remaining term we have to control is $\epsilon^2 R(\epsilon t, t, x)$, with

$$\begin{aligned} R \equiv \partial_\tau U_1 + \Sigma_1(U_0)\partial_x U_1 + \Sigma_1(U_1)\partial_x U_0 + S_1(U_0)\partial_t U_1 + S_1(U_1)\partial_t U_0 \\ -\Sigma_2\partial_x^3 U_1 - S_2\partial_x^2\partial_t U_1 + \epsilon\Sigma_1(U_1)\partial_x U_1 + \epsilon S_1(U_1)\partial_x U_1, \end{aligned}$$

where $U_0(\epsilon t, t, x) = \sum_{i=1}^4 u_i(t,x)\mathbf{e}_i$.

Each term of the right hand side is suitably bounded in the Sobolev H^s-norm, as we show in the following. Indeed, from Proposition 3.9, one has immediately

$$\big|\Sigma_2\partial_x^3 U_1(\epsilon t, t, \cdot)\big|_{H^s} \leq C_0\big|U_1(\epsilon t, t, \cdot)\big|_{H^{s+3}} \leq \sqrt{t}C_0\left(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+5}}, T\right).$$

Then, from (3.8), we deduce

$$\langle \mathbf{e}_i, \partial_t U_1\rangle = -c_i\langle \mathbf{e}_i, \partial_x U_1\rangle + f_i$$

with $f_i \in L^\infty([0,T]\times\mathbb{R}; H^{s+2})$, and $\big|f_i\big|_{H^{s+2}} \leq C_0\big|U^0\big|_{H^{s+5}}$, so that one has identically

$$\big|S_2\partial_x^2\partial_t U_1(\epsilon t, t, \cdot)\big|_{H^s} \leq C_0\big|\partial_t U_1(\epsilon t, t, \cdot)\big|_{H^{s+2}} \leq \sqrt{t}C_0\left(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+5}}\right).$$

One obtains in the same way the desired estimates for $\Sigma_1(U_0)\partial_x U_1$, $\Sigma_1(U_1)\partial_x U_0$, $S_1(U_0)\partial_t U_1$, $S_1(U_1)\partial_t U_0$, $\Sigma_1(U_1)\partial_x U_1$ and $S_1(U_1)\partial_x U_1$.

Finally, in order to estimate $\partial_\tau U_1$, we differentiate (3.8) with respect to τ. Since u_i satisfies (3.9), one has $\partial_\tau u_i \in L^\infty([0,T); H^{s+2})$. We are on the frame of the Lemma 3.8, so that we can obtain as in Proposition 3.9 that $\partial_\tau U_1 \in L^\infty([0,T]\times[0,t]; H^s)$, with

$$\big|\partial_\tau U_1(\epsilon t, t, \cdot)\big|_{H^s} \leq C_0\sqrt{t}\big|\partial_t U^0\big|_{H^{s+2}} \leq C_0\sqrt{t}\big|U^0\big|_{H^{s+5}}.$$

Hence, $R(\epsilon t, t, \cdot) \in L^\infty([0,T/\epsilon]; H^s)$, and

$$\big|R\big|_{H^s} \leq C_0\sqrt{T/\epsilon}\big|U^0\big|_{H^{s+5}},$$

which concludes the first part of the proof.

The second part follows in the exact same way, using the second estimate of Proposition 3.9. □

Finally, thanks to the consistency result and the estimate on U_1, we are able to set the following convergence Proposition:

Proposition 3.12. *Let $U^0 \in H^{s+5}$, $s > 1/2$, $U_B \in L^\infty([0,T/\epsilon];H^{s+5})$ be a family of solutions of (3.2) with $U_B\,|_{t=0} = U^0$ and U_{app} be defined by Definition 3.6, with the same initial value. Then one has*

$$\big|U_{app} - U_B\big|_{L^\infty([0,t];H^{s+1}_\epsilon)} \leq C_0\epsilon^{3/2}t,$$

with $C_0 = C_0\big(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+5}}, T\big)$.

Moreover, U^0 satisfies $U^0(1+x^2) \in H^{s+4}(\mathbb{R})$, then one has the better estimate

$$\big|U_{app} - U_B\big|_{L^\infty([0,t];H^{s+1}_\epsilon)} \leq C_0\epsilon^{2}t,$$

with $C_0 = C_0\big(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0(1+x^2)\big|_{H^{s+4}}, T\big)$.

Proof. Let us set $R^\epsilon \equiv U_{\rm app} - U_B$. Thanks to Proposition 3.11, we know that

$$\big(S_0 - \epsilon S_2\partial_x^2 + \epsilon S_1(U_{\rm app})\big)\partial_t R^\epsilon + \big(\Sigma_0 - \epsilon\Sigma_2\partial_x^2 + \epsilon\Sigma_1(U_{\rm app})\big)\partial_x R^\epsilon = \epsilon^{3/2}f + \epsilon\mathcal{A} + \epsilon\mathcal{B}, \tag{3.13}$$

with $\mathcal{A} = \partial_t S_1(U_{\rm app})R^\epsilon - S_1(R^\epsilon)\partial_t U_B$, $\mathcal{B} = \partial_x\Sigma_1(U_{\rm app})R^\epsilon - \Sigma_1(R^\epsilon)\partial_x U_B$ and a function $f \in L^\infty H^s$.

Then, we can follow the same path as for the proof of Proposition 2.6, in Section A.3 (see also the proof of Proposition 2.8). We define the energy as

$$E_s(R^\epsilon) \equiv \frac{1}{2}(S_0\Lambda^s R^\epsilon, \Lambda^s R^\epsilon) + \frac{\epsilon}{2}(S_2\partial_x\Lambda^s R^\epsilon, \partial_x\Lambda^s R^\epsilon) + \frac{\epsilon}{2}(S_1(U_{\rm app})\Lambda^s R^\epsilon, \Lambda^s R^\epsilon),$$

and the exact same calculations lead to the following inequality:

$$\frac{d}{dt}E_s(R^\epsilon) \leq C_0\epsilon E_s(R^\epsilon) + C_0\epsilon^{3/2}(E_s(R^\epsilon))^{1/2},$$

with $C_0 = C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+5}})$.

From Gronwall-Bihari's theorem, we get $E_s(R^\epsilon) \leq C_0\epsilon^{1/2}(e^{C_0\epsilon t} - 1)$, and finally for $\epsilon t \leq T$,

$$\big|U_{\rm app} - U_B\big|_{H^{s+1}_\epsilon} \leq C_0 E_s(R^\epsilon) \leq C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+5}}, T)\epsilon^{3/2}t.$$

The second part of the proof follows in the same way, using the consistency at order $\mathcal{O}(\epsilon^2)$ of Proposition 3.11. □

Remark 3.13. *From this Proposition, one can immediately deduce the convergence rate between the solution of the symmetric system (3.2), and the KdV approximation (3.9). Indeed, since we know from Proposition 3.9 the growth of the correcting term U_1, and since we restrict ourselves to times $0 \leq t \leq T/\epsilon$, the convergence rate is of order $\mathcal{O}(\epsilon\sqrt{t})$ in general, and of order $\mathcal{O}(\epsilon)$ if the initial data is sufficiently decreasing in space.*

In the same way, we obtain the convergence rate between bounded solutions of the full Euler system (1.1), and the KdV approximation (3.9), using Proposition 2.8. This result is stated rigorously in Theorem 3.1. One sees that even in the case where the initial data is rapidly decreasing in space, the convergence rate of the symmetric Boussinesq/Boussinesq model (2.2) (namely $\mathcal{O}(\epsilon^2 t)$) is better than the one of the KdV approximation (namely $\mathcal{O}(\epsilon)$). This is due to the interaction between the traveling waves of different wave modes, that is captured by the Boussinesq/Boussinesq system, and not by the uncoupled KdV approximation, and which is of order $\mathcal{O}(\epsilon)$ for times of order $\mathcal{O}(1)$. The decreasing in space of the initial data allows this error to remain of order $\mathcal{O}(\epsilon)$ for times of order $\mathcal{O}(1/\epsilon)$.

Numerical simulations for both the Boussinesq/Boussinesq models and the KdV approximation are computed in Section 4.2. In particular, the relationship between the convergence rate and the decreasing in space of the initial data is discussed and enhanced in Figure 7.

3.3. The models under the rigid lid assumption. In this section, we formally recover models existing in the literature in the rigid lid configuration. Starting from our Boussinesq/Boussinesq model (2.1), we recover the three-parameter family of rigid lid Boussinesq/Boussinesq systems presented in [**17**]. One can then apply the method presented in the previous section, in order to obtain the KdV approximation in this case.

The rigid lid models use the variables (ζ, v), where ζ is the interface deviation ($-\eta_1 = \eta_2 \equiv \zeta$), and v is the shear velocity defined by

$$v \equiv (\partial_x \phi_2 - \gamma \partial_x \phi_1)\big|_{z=\epsilon\zeta} .$$

Using the calculations in [**40**], one has

$$v = u_2 - \gamma u_1 - \epsilon \left(\frac{\gamma}{6} \partial_x^2 u_1 + \left(\frac{1}{3\delta^2} + \frac{\gamma}{2\delta} \right) \partial_x^2 u_2 \right) + \mathcal{O}(\epsilon^2).$$

Then, adding the first two equations of (2.1) leads to $\partial_x(h_1 u_1 + h_2 u_2) = 0$, so that one has $u_1 + \frac{1}{\delta} u_2 = \mathcal{O}(\epsilon)$, and

$$v = \frac{\delta + \gamma}{\delta} u_2 + \mathcal{O}(\epsilon) = -(\delta + \gamma) u_1 + \mathcal{O}(\epsilon).$$

Therefore, using a straightforward combination the equations of (2.1), one checks that the system becomes

$$\left\{ \begin{array}{l} \partial_t \zeta + \frac{1}{\delta+\gamma} \partial_x v + \epsilon \frac{\delta^2 - \gamma}{(\gamma+\delta)^2} \partial_x (\zeta v) + \epsilon \frac{1+\gamma\delta}{3\delta(\gamma+\delta)^2} \partial_x^3 v = \mathcal{O}(\epsilon^2), \\ \partial_t v + (1-\gamma) \partial_x \zeta + \epsilon \frac{\delta^2 - \gamma}{(\delta+\gamma)^2} v \partial_x v = \mathcal{O}(\epsilon^2). \end{array} \right. \tag{3.14}$$

Finally, using BBM-tricks as in (2.5), and the change of variable $v_\beta = (1 - \epsilon\beta\partial_x^2)^{-1} v$ (with $\beta \geq 0$), one obtains eventually the three-parameter family of rigid lid Boussinesq/Boussinesq systems presented in [**17**]:

$$\left\{ \begin{array}{l} (1 - \epsilon b \partial_x^2) \partial_t \zeta + \frac{1}{\delta+\gamma} \partial_x v_\beta + \epsilon \frac{\delta^2 - \gamma}{(\gamma+\delta)^2} \partial_x (\zeta v_\beta) + \epsilon a \partial_x^3 v_\beta = \mathcal{O}(\epsilon^2), \\ (1 - \epsilon d \partial_x^2) \partial_t v_\beta + (1-\gamma) \partial_x \zeta + \epsilon \frac{\delta^2 - \gamma}{(\delta+\gamma)^2} v_\beta \partial_x v_\beta + \epsilon c (1-\gamma) \partial_x^3 \zeta = \mathcal{O}(\epsilon^2), \end{array} \right. \tag{3.15}$$

with a, b, c and d set (with $\theta_1 \geq 0$, $\theta_2 \leq 1$, $\beta \geq 0$) as

$$(\gamma+\delta) a = (1-\theta_1) \frac{1+\gamma\delta}{3\delta(\gamma+\delta)} - \beta, \quad b = \theta_1 \frac{1+\gamma\delta}{3\delta(\gamma+\delta)}, \quad c = \beta\theta_2, \quad d = \beta(1-\theta_2).$$

From this system, one can easily follow the path of Section 3.1, and deduce the KdV approximation related to system (3.15). One would obtain a similar result as in Theorem 3.1. Eventually, the KdV approximation consists in decomposing the approximate solution $(\zeta_{\mathrm{KdV}}, v_{\mathrm{KdV}})$ as

$$(\zeta_{\mathrm{KdV}}, v_{\mathrm{KdV}}) = u_+ \mathbf{e}_+ + u_- \mathbf{e}_-,$$

with u_+ and u_- two solutions of independent Korteweg-de Vries equations, namely

$$\partial_t u_\pm \ \pm \ c \partial_x u_\pm \ + \ \epsilon \lambda_\pm u_\pm \partial_x u_\pm \ \pm \ \epsilon \mu \partial_x^3 u_\pm = 0, \tag{3.16}$$

with the following vectors and coefficients:

$$\mathbf{e}_\pm = \frac{1}{\sqrt{2}} \begin{pmatrix} \pm \frac{1}{\sqrt{1-\gamma}} \\ \sqrt{\gamma+\delta} \end{pmatrix}, \ c = \sqrt{\frac{1-\gamma}{\gamma+\delta}}, \ \lambda = \frac{1}{\sqrt{2(1-\gamma)}} \frac{3c}{2} \frac{\delta^2 - \gamma}{\gamma+\delta}, \ \mu = \frac{c}{6} \frac{(1+\gamma\delta)}{\delta(\gamma+\delta)}.$$

In that way, when looking at the decomposition of deformation of the interface, the KdV approximation leads to two counter-propagating waves, that is to say that one can write $\eta = \eta_+ + \eta_-$, with $\eta_\pm$ solution of

$$\partial_t \eta_\pm \ \pm \ \sqrt{\frac{1-\gamma}{\gamma+\delta}}\partial_x \eta_\pm \ + \ \epsilon \frac{3c}{2}\frac{\delta^2-\gamma}{\gamma+\delta}\eta_\pm \partial_x \eta_\pm \ \pm \ \epsilon \frac{c}{6}\frac{(1+\gamma\delta)}{\delta(\gamma+\delta)}\partial_x^3 \eta_\pm \ = \ 0.$$

We recover the classical KdV equations in the rigid lid configuration (see for example **[39, 61, 85, 35]**). Following Section 3.2, one would obtain in the same way a rigorous justification for the KdV approximation, under the rigid lid assumption.

One sees that whereas the KdV approximation in the rigid lid case leads to a decomposition into two waves, the free surface configuration predicts the decomposition into four waves, each of them solution of a KdV equation with different velocities. This striking fact leads to think that the rigid lid assumption may induce a significant alteration of the behavior of the solutions, and thus cannot be considered as a harmless statement in all configurations. This remark has already been addressed in **[60, 86]**, but to our knowledge, in the absence of the exhaustive decomposition given in Theorem 3.1, the analysis of the difference between the two configurations has never been extensively discussed.

The following Section is devoted to a detailed study of the differences between the rigid lid and free surface configurations, depending on the values of the density ratio γ and the depth ratio δ.

3.4. Discussion. Let us recall here that the KdV approximation in the free surface configuration consists in decomposing $U = (\eta_1, \eta_2, u_1, u_2)$ as $U \sim \sum_{i=1}^4 u_i \mathbf{e}_i$, with u_i satisfying the KdV equation

$$\partial_t u_i \ + \ c_i \partial_x u_i \ + \ \epsilon \lambda u_i \partial_x u_i \ + \ \epsilon \mu_i \partial_x^3 u_i \ = \ 0.$$

The coefficients c_i, λ_i and μ_i, as well as the vectors $\mathbf{e}_i$ are given in Remark 3.5, page 97. This leads to the following decomposition for the respective deformations of the interface and the surface:

i. For the interface: $\eta_2 = \displaystyle\sum_{i=1}^4 u_i \mathbf{e}_{i,2} \equiv \sum_{(j,k)=(\pm,\pm)} \eta_{j,k}$, and $\eta_{\pm,\pm}$ satisfies the KdV equation

$$\partial_t \eta_{\pm,k} \ \pm \ c_k \partial_x \eta_{\pm,k} \ + \ \epsilon \lambda_k^i \eta_{\pm,k} \partial_x \eta_{\pm,k} \ \pm \ \epsilon \mu_k \partial_x^3 \eta_{\pm,k} \ = \ 0,$$

where $\mathbf{e}_{i,j}$ denotes the j^{th} component of the vector $\mathbf{e}_i$ and with the following coefficients[3]:

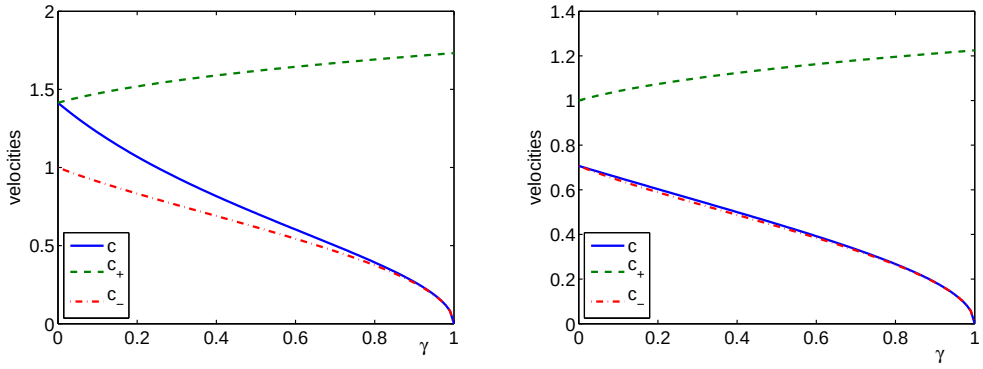
$$c_\pm^2 = \frac{1+\delta \pm \sqrt{(1-\delta)^2+4\gamma\delta}}{2\delta}, \quad \mu_\pm = \frac{c\pm}{6}\frac{(1+\frac{3\gamma}{\delta}+\frac{1}{\delta^2})(c_\pm^2 - \frac{1-\gamma}{\delta+1}) - \frac{1}{\delta}c_\pm^2}{c_\pm^2 - 2\frac{1-\gamma}{\delta+1}},$$
$$\lambda_\pm^i = \frac{3c_\pm}{2}\frac{(2-\delta)c_\pm^2 + \delta - \frac{1}{\delta} - (1-\gamma)}{|(c_+^2 - c_-^2)(1-c_\pm^2)|}.$$

ii. For the surface: $\zeta_1 = \eta_1 + \eta_2 = \displaystyle\sum_{i=1}^4 u_i(\mathbf{e}_{i,1} + \mathbf{e}_{i,2}) \equiv \sum_{(j,k)=(\pm,\pm)} \zeta_{j,k}$, and $\zeta_{\pm,\pm}$ satisfies the KdV equation

$$\partial_t \zeta_{\pm,k} \ \pm \ c_\pm \partial_x \zeta_{\pm,k} \ + \ \epsilon \lambda_k^s \zeta_{\pm,k} \partial_x \zeta_{\pm,k} \ \pm \ \epsilon \mu_k \partial_x^3 \zeta_{\pm,k} \ = \ 0,$$

with the same values as previously for $c_\pm$ and $\mu_\pm$, and

$$\lambda_\pm^s = \frac{3c_\pm}{2}\frac{(2-\delta)c_\pm^2 + \delta - \frac{1}{\delta} - (1-\gamma)}{(c_+^2 - c_-^2)c_\pm^2}.$$



(a) $\delta = 1/2$ (b) $\delta = 2$

FIGURE 2. The different velocities, in the rigid lid configuration and free surface case, for $\gamma \in (0,1)$ and (a) $\delta = 1/2$, (b) $\delta = 2$.

Therefore, one sees that for both the surface and the interface elevations, the KdV approximation predicts the evolution of four different waves, two of them corresponding to the velocities $\pm c_+$ (we call them fast mode waves), and the other two corresponding to the velocities $\pm c_-$ (we call them slow mode waves), with $c_+ > c_- > 0$. The fact that such different modes exist is characteristic of the free surface configuration, as only two counter-propagating waves appear in the rigid lid case. Moreover, as we see in Figure 2, the two velocities corresponding to the free surface configuration can be very different from the velocity in the rigid lid case (namely $c = \sqrt{\frac{1-\gamma}{\gamma+\delta}}$), depending on the values of δ and γ. In these cases, one expect the solutions in the two configurations to behave very differently.

The aim of this section is to study more in depth the behavior of the KdV approximation in the two different configurations, with respect to the parameters γ and δ. The first part is devoted to the study of solitary waves, as known solutions of the KdV equations. Then, we study the case where the initial data satisfy the rigid lid hypothesis, and explore the evolution the surface in that case. The results we obtain are summarized in Section 3.4.3; we let the reader refer to Figures 8–15 for a numerical illustration of our statements.

REMARK 3.14. *In the following study, and especially in Table 1, we allow ourselves to look at the behavior of the system in the limit cases of the parameters ($\delta \to 0, \infty$ and $\gamma \to 0, 1$), despite the fact that the rigorous justification of the KdV approximation, as well as the Boussinesq/Boussinesq models, break in these limits. However, the KdV approximation when $\gamma \to 1$ has been widely used in the literature (see for example* **[49, 52, 102, 105]***), and such limits offer striking illustrations of our discussion.*

3.4.1. *Solitary waves.* It is well known that the solitary wave solutions of the generic KdV equation

$$\partial_t u \ + \ c\partial_x u \ + \ \epsilon\lambda u\partial_x u \ + \ \epsilon\mu\partial_x^3 u \ = \ 0$$

can be expressed as follows:

$$u(t,x) = \frac{M}{\cosh(k(x - x_0 - c't))^2}, \tag{3.17}$$

with $c' = c + \epsilon\frac{\lambda M}{3}$, $k = \sqrt{\frac{\lambda M}{12\mu}}$, and M and x_0 arbitrary.

We discuss in the following the polarity, magnitude and thickness of such waves, as the parameters γ and δ specify the coefficients of the KdV approximation.

[3]These are the coefficients that are displayed in **[56, 76, 86]**.

Polarity. It is obvious that for $k = \sqrt{\frac{\lambda M}{12\mu}}$ to be real valued, the sign of the ratio λ/μ determines the sign of the acceptable values of M. Therefore, we are able to predict, depending on the parameters γ and δ, the polarity of the solitary waves predicted by the KdV approximation (elevation or depression). We give here the result, first for the free surface case, and then in the rigid lid configuration.

First, one can check that for every value of $\delta > 0$ and $\gamma \in (0,1)$, the three coefficients λ_+^i, λ_+^s and $\mu_+ > 0$ are positive. Hence, the fast mode solitary waves will always be of elevation/elevation type (both surface wave and interface wave are convex upward). The qualitative nature of the fast mode is thus similar to that of the one-layer water-wave problem, which is always of elevation type. In particular, when we set $\gamma \to 0^+$ and $\delta \to 1^-$, one recovers the classical KdV equation for a single layer at the interface ($c_+ \to 1$, $\lambda_+^i \to 3/2$, $\mu_+ \to 1/6$), and when we set $\delta \to \infty$, one recovers the classical KdV equation for a single layer at the surface ($c_+ \to 1$, $\lambda_+^s \to 3/2$, $\mu_+ \to 1/6$).

The behavior of the slow mode is more peculiar, as the nonlinear coefficients $\lambda_-^{s,i}$ can have both signs, depending on the size of the thickness ratio δ. Indeed, as it has been pointed out in **[115, 56]** and then in **[86, 98]**, for every $\gamma \in (0,1)$, there exists a critical ratio $\delta_c(\gamma)$ such that if $\delta > \delta_c(\gamma)$, then $\lambda_-^i > 0$ and $\lambda_-^s < 0$ (and conversely if $\delta < \delta_c$). Since one has $\mu_- > 0$ for every value of $\delta > 0$ and $\gamma \in (0,1)$, we know that the slow mode solitary wave will be of elevation/depression type (surface wave convex upward, and interface wave concave) if $\delta < \delta_c$, and of depression/elevation type if $\delta > \delta_c$. There is no solitary waves in the case $\delta = \delta_c$.

More precisely, the critical ratio is the unique real solution of the equation

$$X^3 + (\gamma^2 + 3\gamma - 3)X^2 + (3 - 4\gamma)X - 1 = 0,$$

and takes values in $\delta_c \in (1, 5/4]$ for $\gamma \in (0,1)$ (see Figure 3).

In that way, the behavior of the slow mode waves resembles that of the interface waves with a rigid lid. Indeed, such a critical ratio appears straightforwardly in the rigid lid configuration: if $\delta < \delta_c' \equiv \sqrt{\gamma}$, then the interface wave is of depression type, and if $\delta > \delta_c'$, then the interface wave is of elevation type. This critical ratio is well known in the literature, and has led to many extended models, where a cubic nonlinear term becomes the major source of nonlinearity for $\delta \sim \delta_c$ (see for example **[39, 61, 45, 50, 52]**).

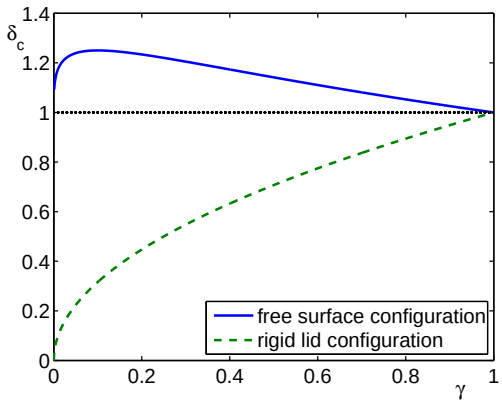


Figure 3. Dependence of the critical ratio δ_c on the density ratio γ.

It is interesting to see that even if the respective polarity of the interface slow mode waves in the free surface configuration, and the interface waves in the rigid lid configuration follow qualitatively the same behavior, the value of the critical ratio is notably different. Indeed δ_c is always located above 1 in the free surface case, on the contrary to the rigid lid case. Moreover, as we can see in Figure 3, they have considerably different values, except when $\gamma \sim 1$. In the area between the two curves, the polarities of the interface waves in the free surface and in the rigid lid configurations are reversed; the two models therefore lead to considerably different results.

Magnitude of the deformations. Depending of the parameters γ and δ, we are able to compare the magnitudes of the respective amplitudes of the surface and the interface waves. Indeed, since $\eta_2 = \sum_{i=1}^4 u_i \mathbf{e}_{i,2} = \sum \eta_{\pm,\pm}$, and $\zeta_1 = \eta_1 + \eta_2 = \sum_{i=1}^4 u_i(\mathbf{e}_{i,1} + \mathbf{e}_{i,2}) = \sum \zeta_{\pm,\pm}$, one sees immediately that the surface and interface deformations, for each mode, are proportional, and satisfy

$$\frac{\eta_{\pm,\pm}}{\zeta_{\pm,\pm}} = \frac{\mathbf{e}_{i,2}}{\mathbf{e}_{i,1} + \mathbf{e}_{i,2}} = \frac{c_\pm^2 - 1}{c_\pm^2}.$$

In that way, one deduces that for the fast mode, the surface deformation is always bigger than the interface deformation, and the ratio tends to zero when $\gamma \to 0$ with $\delta \geq 1$, or when $\delta \to \infty$.

Meanwhile, as remarked in [**56**], the amplitude of the surface deformation is bigger than the one of the interface for the slow mode if $0 < \delta \leq 2(1 - 2\gamma)$, and conversely if $\delta > 2(1 - 2\gamma)$. Moreover, the ratio tends to zero when $\gamma \to 0$ with $\delta \leq 1$, and tends to ∞ when $\gamma \to 1$ or $\delta \to \infty$.

Thickness. In addition to forcing the polarity of the solitary wave, the ratio $\frac{\lambda M}{12\mu}$ is also related to the thickness, or the wavelength of the wave. Indeed, defining the thickness of a wave as in [**86**] by

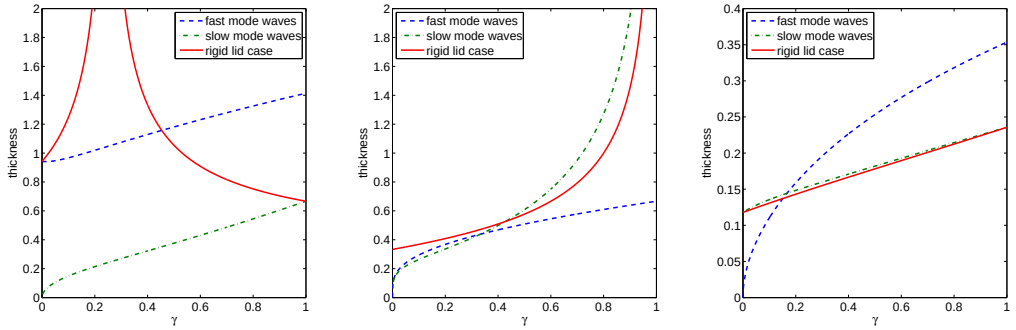
$$l(u) \equiv \frac{1}{2M} \int_{-\infty}^{+\infty} u(x) \mathrm{d}x,$$

one obtains for the function (3.17): $l(u) = \frac{1}{k} = \sqrt{\frac{12\mu}{\lambda M}}$.

As an immediate result, and since we know the ratio between the magnitude of the deformations at the surface and the interface, the thickness of the deformations are identical at the surface and at the interface:

$$\frac{l_\pm^s}{l_\pm^i} = \sqrt{\frac{\lambda_\pm^i}{\lambda_\pm^s} \left| \frac{c_\pm^2 - 1}{c_\pm^2} \right|} = 1.$$

One can now compare the thickness of the different wave modes, and the ones in the rigid lid configuration, for waves of same heights. One computes in Figure 4 the ratio $\sqrt{\frac{\mu_+}{\lambda_+^i}}$, $\sqrt{\frac{\mu_-}{\lambda_-^i}}$ and $\sqrt{\frac{\mu}{\lambda^i}}$ for $\gamma \in (0,1)$, and $\delta = 1/2, 1, 2$.



(a) $\delta = 1/2$ (b) $\delta = 1$ (c) $\delta = 2$

FIGURE 4. The different thicknesses, in the rigid lid and free surface configuration, for the parameters $\gamma \in (0,1)$ and (a) $\delta = 1/2$, (b) $\delta = 1$, (c) $\delta = 2$.

Of course, the thickness of the solitary wave tends to infinity when the depth ratio approaches its critical ratio $\delta = \delta_c$, so that the waves predicted in the rigid lid and in the

free surface configurations are excessively unlike. Additionally, one sees that when δ is small, then the thickness of the internal waves predicted in the free surface configuration is largely different from both the slow mode and fast mode thicknesses, except in the limit cases $\gamma \to 0$ and $\gamma \to 1$.

When $\gamma \ll 1$ and $\delta \leq 1$, one has at the same time the similitude of the thickness and the velocities of the waves in the rigid lid configuration, and the fast mode waves in the free surface configuration. However, the amplitude of the surface deformation is bigger than the one of the interface for the fast mode waves, so that the rigid lid is not valid hypothesis. This resemblance has to do with the fact that both the fast mode waves and the waves in the rigid lid configuration converge to waves of the one-layer problem, when $\gamma \to 0$ and $\delta \to 1^-$. Conversely, when $\gamma \sim 1$ and when $\delta \gg 1$, the characteristics of the waves in the rigid lid configuration resemble the slow mode waves ones. We explore thereafter the validity of this hypothesis for different situations.

3.4.2. *Evolution of the rigid lid hypothesis.* Now, we restrict ourselves to initial data that are compatible for both the free surface and rigid lid configurations, and compare the different evolutions of the solutions. Following Section 3.3, if the initial data with the rigid lid assumption is (η^0, v^0), then the corresponding initial data in the free surface case (in the limit $\epsilon \to 0$) is $U^0 = (-\eta^0, \eta^0, \frac{-1}{\gamma+\delta}v^0, \frac{\delta}{\gamma+\delta}v^0))$. Therefore, one has initially $u_i\,|_{t=0} = \mathbf{e}_i \cdot S_0 U^0$, so

$$\sum_{(j,k)=(\pm,\pm)} \eta_{j,k}\,|_{t=0} = \sum_{i=1}^{4} \mathbf{e}_{i,2} u_i\,|_{t=0} = \mathbf{e}_{i,2}\left((1-\gamma)\mathbf{e}_{i,2}\eta^0 + \frac{(\mathbf{e}_{i,3} - \gamma \mathbf{e}_{i,4})v^0}{\gamma+\delta}\right),$$

$$\sum_{(j,k)=(\pm,\pm)} \zeta_{j,k}\,|_{t=0} = (\mathbf{e}_{i,1} + \mathbf{e}_{i,2})\left((1-\gamma)\mathbf{e}_{i,2}\eta^0 + \frac{(\mathbf{e}_{i,3} - \gamma \mathbf{e}_{i,4})v^0}{\gamma+\delta}\right).$$

Using the values given in Remark 3.5 page 97, this reads

$$\eta^0_{j,k} = \eta_{j,k}\,|_{t=0} = \frac{k}{2\delta(c_+^2 - c_-^2)}\left((1-\gamma)\frac{c_k^2-1}{c_k^2}\eta^0 + j\frac{1+\gamma\delta-\gamma\delta c_k^2}{(\gamma+\delta)c_k}v^0\right),$$

$$\zeta^0_{j,k} = \zeta_{j,k}\,|_{t=0} = \frac{k}{2\delta(c_+^2 - c_-^2)}\left((1-\gamma)\eta^0 + j\frac{c_k^2}{c_k^2-1}\frac{1+\gamma\delta-\gamma\delta c_k^2}{(\gamma+\delta)c_k}v^0\right).$$

Since the KdV equation preserves mass, knowing the size of the initial data allows to predict the significance of the waves. In particular, the rigid lid hypothesis will be valid for long times only if $|\zeta^0_{\pm,\pm}| \ll |\eta^0_{\pm,\pm}|$. We give in Table 1 the different behavior of these variables, in the limits $\gamma \to 1$, $\gamma \to 0$, $\delta \to \infty$ and $\delta \to 0$. As we can see, the rigid lid hypothesis will be valid for long times when $\gamma \sim 1$, or when $\delta \gg 1$. For each of these cases, one sees that the main deformation comes from the slow mode waves, which correspond to the waves predicted by the models in the rigid lid configuration.

As a specific example, when the initial data has zero velocities (that is to say $v^0 = 0$), then we are able to compare straightforwardly the different magnitudes of the four waves. Indeed, one deduces from the previous calculations that when $v^0 = 0$, one has

$$\eta^0_{j,k} = \frac{k(1-\gamma)}{2\delta(c_+^2 - c_-^2)}\frac{c_k^2-1}{c_k^2}\eta^0, \quad \text{and} \quad \zeta^0_{j,k} = \frac{k(1-\gamma)}{2\delta(c_+^2 - c_-^2)}\eta^0.$$

Consequently, the four different waves have the same weight at the surface. The situation is more sophisticated at the interface, and one has eventually

$$\begin{array}{rcll} |\eta_{\pm,+}|_{L^2} & \geq & |\eta_{\pm,-}|_{L^2} & \text{if } \delta \leq 1-2\gamma, \\ |\eta_{\pm,+}|_{L^2} & < & |\eta_{\pm,-}|_{L^2} & \text{if } \delta > 1-2\gamma, \end{array}$$

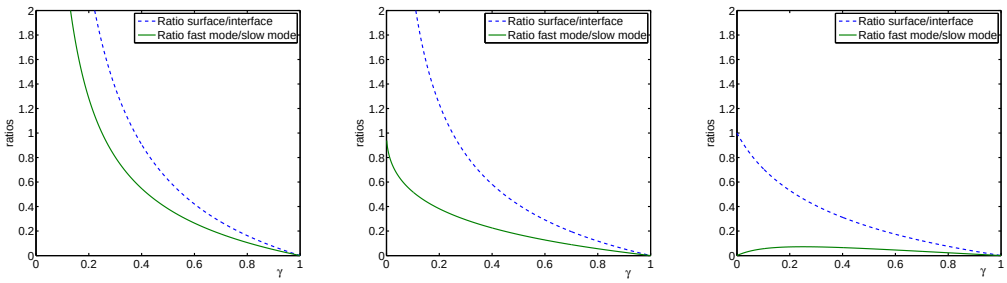
	$\gamma \to 1$	$\gamma \to 0,\ \delta > 1$	$\gamma \to 0,\ \delta < 1$
$\eta^0_{\pm,+}$	0	$\dfrac{\pm\nu}{2\sqrt{\delta}} v^0$	$\frac{1}{2}\left(\eta^0 \pm \nu v^0\right)$
$\eta^0_{\pm,-}$	$\frac{1}{2}\eta^0 \pm \varsigma v^0$	$\frac{1}{2}\left(\eta^0 \mp \nu v^0\right)$	$\mp\frac{1}{2}\nu v^0$
$\zeta^0_{\pm,+}$	0	$\dfrac{1}{2(\delta-1)}\left(\eta^0 \pm \tau v^0\right)$	$\dfrac{1}{2(1-\delta)}\left(\eta^0 \pm \nu v^0\right)$
$\zeta^0_{\pm,-}$	0	$\dfrac{1}{2(\delta-1)}\left(\eta^0 \mp \nu v^0\right)$	$\dfrac{1}{2(1-\delta)}\left(-\eta^0 \pm \tau v^0\right)$

with $\varsigma \underset{\gamma\to 1}{\sim} \frac{1}{2\sqrt{(\delta+1)(1-\gamma)}}$, $\tau \underset{\gamma\to 0}{\sim} \frac{\delta-1}{\delta\gamma}$ and $\nu = \frac{1}{\sqrt{\delta}(\delta-1)}$.

	$\delta \to \infty$	$\delta \to 0$
$\eta^0_{\pm,+}$	0	$\frac{1}{2}(1-\gamma)\eta^0$
$\eta^0_{\pm,-}$	$\frac{1}{2}\eta^0$	$\frac{1}{2}\left(\gamma\eta^0 \mp \frac{1}{\gamma\sqrt{1-\gamma}} v^0\right)$
$\zeta^0_{\pm,+}$	0	$\frac{1}{2}(1-\gamma)\eta^0$
$\zeta^0_{\pm,-}$	0	$\dfrac{-1}{2}\left((1-\gamma)\eta^0 \mp \frac{\sqrt{1-\gamma}}{\gamma^2} v^0\right)$

TABLE 1. Initial magnitudes of the different waves, at the surface and at the interface, in the limit cases.

so that the fast mode wave is more significant than the slow mode wave when $\delta \leq 1 - 2\gamma$, and conversely otherwise. In the limits $\delta \to \infty$ and $\gamma \to 1$, the magnitudes of the fast mode waves tend to 0, so that the energy is only shared by the slow mode waves. Meanwhile, in the limit $\delta \to 0$, the significance of the fast mode and the slow modes respectively tend to $\frac{1-\gamma}{2}|\eta^0|_{L^2}$ and $\frac{\gamma}{2}|\eta^0|_{L^2}$. Finally, in the limit $\gamma \to 0$, then the magnitudes of the fast mode waves tend to 0 when $\delta > 1$, and in the contrary carry all the energy when $\delta < 1$.



(a) $\delta = 1/2$ (b) $\delta = 1$ (c) $\delta = 2$

FIGURE 5. Magnitude of the deformations: the fast mode internal waves and the surface waves, when compared with the slow mode internal waves. Ratios for $\gamma \in (0,1)$ and (a) $\delta = 1/2$, (b) $\delta = 1$, (c) $\delta = 2$.

We plot in Figure 5 the different ratio of magnitudes, with the slow mode internal waves chosen as reference; that is to say

$$\frac{|\zeta_{\pm,+}|_{L^2}}{|\eta_{\pm,-}|_{L^2}} = \frac{|\zeta_{\pm,-}|_{L^2}}{|\eta_{\pm,-}|_{L^2}} \quad \text{and} \quad \frac{|\eta_{\pm,+}|_{L^2}}{|\eta_{\pm,-}|_{L^2}} \ .$$

The rigid lid hypothesis is valid for small values of these ratios. Again, one sees that it occurs only when $\gamma \sim 1$, or when δ is big. This fact has already been addressed, for example in **[60, 86]**, but its precise confirmation has never been exposed, to our knowledge. For

example, if $\delta = 1/2$ and $\gamma = 0.8$, one has

$$\frac{|\zeta_{\pm,+}|_{L^2}}{|\eta_{\pm,-}|_{L^2}} = \frac{|\zeta_{\pm,-}|_{L^2}}{|\eta_{\pm,-}|_{L^2}} \sim \frac{1}{6} \quad \text{and} \quad \frac{|\eta_{\pm,+}|_{L^2}}{|\eta_{\pm,-}|_{L^2}} \sim \frac{1}{10} \ ,$$

so that the rigid lid assumption is inaccurate.

3.4.3. *Summary.* Let us summarize in Figure 6 the different results we obtained concerning the dependence of the behavior of the KdV approximation, depending on the parameters of the problem.

When δ is above the plain curve ($\delta > \delta_c$), then we know that the slow mode solitary waves will be of elevation type at the interface and of depression type at the surface, and conversely if $\delta < \delta_c$. The fast mode solitary waves are always of elevation type at the surface and at the interface.

Above the dashed line, the interface deformation is more important than the surface one for the slow mode solitary waves, and conversely if $\delta < 2(1-2\gamma)$. As for the fast mode waves, the surface elevation is always bigger than the interface elevation.

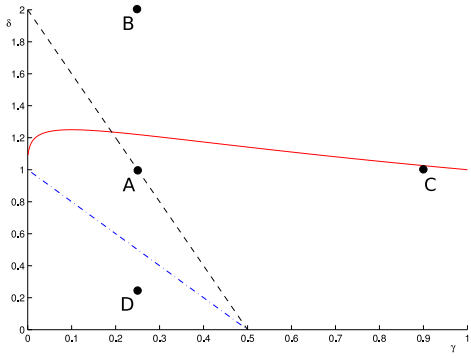


Figure 6. Properties of the KdV approximation, depending on the depths ratio δ and the density ratio γ.

The dash-dotted line concerns waves created by an initial data with zero velocities, and with a flat surface. In that case, the fast mode waves will be smaller than the slow mode waves above the line $\delta = 1 - 2\gamma$, and conversely below.

We see that for big values of γ and/or δ, the KdV approximation of the two-layer problem with a free surface gives a solution that resembles the interface problem with a rigid lid: the fast mode waves are smaller than the slow mode waves, and the magnitude of the deformation of the surface is of less importance than the deformation of the interface. On the contrary, when γ and δ are small, then the solutions of our problem, when considered at the interface, are comparable to solutions of the one-layer water wave models.

The dots on Figure 6 represent the pair of parameters (γ, δ) for which numerical simulations have been computed in Section 4.2. We have first computed the solutions of the symmetric Boussinesq/Boussinesq model and the KdV approximation, for both for the case of solitary waves, and zero velocities-flat surface initial data. The results are plotted for parameters corresponding to **A** ($\gamma = 1/4, \delta = 1$) in Figures 8 and 9, and to **B** ($\gamma = 1/4, \delta = 2$) in Figures 10 and 11. Then, we have compared the KdV approximation in the rigid lid and free surface configurations. Figure 12, 14, 13 and 15 corresponds respectively to the points **A**, **B**, **C** ($\gamma = 9/10, \delta = 1$) and **D** ($\gamma = \delta = 1/4$).

4. Numerical comparison

4.1. The numerical schemes. This section is devoted to the numerical comparison between the different models displayed in this article, namely the symmetric Boussinesq/Boussinesq model (2.2), and the KdV equations (3.9). We first provide a numerical scheme for generic KdV equations that can easily be adjusted for the uncoupled KdV approximations (3.9) and (3.16), and its adaptation to the Boussinesq/Boussinesq system (2.2).

Each time, we use a Crank Nicholson scheme and replace the costly numerical treatment of the nonlinear term by a predictive step. This method has been introduced by Besse and Bruneau in [**13**], justified in [**12**], and used in the water wave framework by Chazel in [**27**], and more recently by Duruflé and Israwi in [**42**]. The method is formally of order two in space and time, which is confirmed by the simulations, and appears to be unconditionally stable.

4.1.1. *The KdV equation.* We present here the numerical scheme for the generic KdV equation:

$$\partial_t u + c\partial_x u + \lambda u \partial_x u + \mu \partial_x^3 u. \tag{4.1}$$

First, we use the following semi-discretized in time equation:

$$\begin{aligned}\frac{u^{n+1}-u^n}{dt} + c\partial_x\left(\frac{u^{n+1}+u^n}{2}\right) + \lambda\Big(\alpha u^{n+1/2}\partial_x\left(\tfrac{u^{n+1}+u^n}{2}\right)\\ +(1-\alpha)\partial_x u^{n+1/2}\left(\tfrac{u^{n+1}+u^n}{2}\right)\Big) + \mu\partial_x^3\left(\tfrac{u^{n+1}+u^n}{2}\right) = 0,\end{aligned}$$

with $\alpha \in [0,1]$, and where $u^{n+1/2}$ is a predictive term defined by

$$u^n = \frac{u^{n+1/2}+u^{n-1/2}}{2}. \tag{4.2}$$

The scheme takes advantage of the two possible discretizations of the nonlinear term $u\partial_x u$, that is to say $u^{n+1/2}\partial_x\left(\frac{u^{n+1}+u^n}{2}\right)$ and $\frac{u^{n+1}+u^n}{2}\partial_x\left(u^{n+1/2}\right)$, by introducing a parameter $\alpha \in [0,1]$ and taking a convex combination of these possibilities.

It is easy to check that, in order to preserve the semi-discrete L^2-norm, one has to choose $\alpha = 2/3$. As for the spatial discretization, we use the Crank-Nicholson scheme, adjusted so that the discrete L^2-norm is preserved. This leads to the final discretization:

$$\begin{aligned}\frac{u_i^{n+1}-u_i^n}{dt} + c\left(D_1\frac{u^{n+1}+u^n}{2}\right)_i + \frac{\lambda}{3}\Bigg(\left(\frac{u_{i+1}^{n+1}+u_{i+1}^n+u_{i-1}^{n+1}+u_{i-1}^n}{4}\right)\left(D_1 u^{n+1/2}\right)_i\\ + \left(u_i^{n+1/2}+\frac{u_{i+1}^{n+1/2}+u_{i-1}^{n+1/2}}{2}\right)\left(D_1\frac{u^{n+1}+u^n}{2}\right)_i\Bigg) + \mu\left(D_3\frac{u^{n+1}+u^n}{2}\right)_i = 0,\end{aligned} \tag{4.3}$$

with D_1 and D_3 the classical centered discretizations of the derivatives ∂_x and ∂_x^3 with periodic boundary conditions. The scheme is given at each step by (4.3) and (4.2), with a simple explicit scheme for the first half step. One can then check (see [**42**, Theorem 2]):

$$\forall\, n \in \mathbb{N}, \quad \sum_i |u_i^n|^2 \;=\; \sum_i |u_i^0|^2.$$

4.1.2. *The Boussinesq/Boussinesq system.* We use the same ideas as in the previous section for the discretization of the Boussinesq/Boussinesq system. Even if the L^2 norm is not preserved by (2.2), and no simple quantity either, we decide to use the same parameter $\alpha = 2/3$ as in the previous section. This leads to the following discretization of the spatial nonlinear term:

$$\Sigma_1(U)\partial_x U \sim \frac{2}{3}\Sigma_1(U^{n+1/2})\partial_x\left(\frac{U^{n+1}+U^n}{2}\right) + \frac{1}{3}\Sigma_1\left(\frac{U^{n+1}+U^n}{2}\right)\partial_x U^{n+1/2}.$$

As for the nonlinear term in time, we simply use

$$S_1(U)\partial_t U \sim S_1(U^{n+1/2})\partial_t\left(\frac{U^{n+1}+U^n}{2}\right).$$

We also need to construct the linear mappings $\tilde{\Sigma}_1$ with values on $\mathcal{M}_4(\mathbb{R})$, such that

$$\forall U, V \in \mathbb{R}^4, \quad \Sigma_1(U)V \;=\; \tilde{\Sigma}_1(V)U.$$

This finally leads to the following scheme:

(4.4)
$$\Big(S_0+\epsilon S_1(U_i^{n+1/2})-\epsilon S_2D_2\Big)\tfrac{U_i^{n+1}-U_i^n}{dt}+\Big(\Sigma_0D_1\tfrac{U^{n+1}+U^n}{2}\Big)_i-\epsilon\Big(\Sigma_2D_3\tfrac{U^{n+1}+U^n}{2}\Big)_i$$
$$+\tfrac{\epsilon}{3}\left(\Sigma_1\left(U_i^{n+1/2}+\tfrac{U_{i+1}^{n+1/2}+U_{i-1}^{n+1/2}}{2}\right)\Big(D_1\tfrac{U^{n+1}+U^n}{2}\Big)_i\right.$$
$$\left.+\Big(\tilde\Sigma_1\big(D_1U^{n+1/2}\big)\tfrac{U^{n+1}+U^n}{2}\Big)_i\right)=0,$$

where D_1, D_2 and D_3 are block-diagonal, with the classical centered discretizations of the derivatives ∂_x, ∂_x^2 and ∂_x^3 (with periodic boundary conditions) as diagonal blocks.

The scheme is given at each step by (4.4) and (4.2), with a simple explicit scheme for the first half step.

4.1.3. *Validation of the numerical method.* As said previously, the method is formally of order 2 in space and time, which is confirmed by the simulations, and appears to be unconditionally stable. In order to validate the schemes, we use the known solitary wave solutions of (4.1), expressed as follows:

$$u(t,x)=\frac{M}{\cosh^2(k(x-c't))}, \tag{4.5}$$

with $c'=c+\frac{\lambda M}{3}$, $k=\sqrt{\frac{\lambda M}{12\mu}}$, and M arbitrary.

Therefore, we are able to construct an initial data that will lead the solutions of the KdV approximation defined in Theorem 3.1 to be steady traveling waves. Indeed, if we set $\gamma\in(0,1)$ and $\delta>0$, the coefficients of the KdV equations solved by the KdV approximation are explicit and given in Remark 3.5 page 97. Consequently, we choose $M_1,\dots,M_4\in\mathbb{R}$, and set $U^0=\sum_{i=1}^4 u_i(x,0)\mathbf{e}_i$, with u_i of the form given by (4.5), with the given parameters, and the uncoupled KdV approximation will be given by $U(t,x)=\sum_{i=1}^4 u_i(\epsilon t,x-c_it)\mathbf{e}_i$.

Unfortunately, we cannot exhibit such an exact solution for the symmetric Boussinesq/Boussinesq system (2.2). In order to validate the scheme (4.4), we plug the solution of the uncoupled KdV approximation in (2.2), and obtain a forcing term $F(t,x)$, so that $U(t,x)$ is solution of the modified system

$$\Big(S_0+\epsilon\big(S_1(U)-S_2\partial_x^2\big)\Big)\partial_tU+\Big(\Sigma_0+\epsilon\big(\Sigma_1(U)-\Sigma_2\partial_x^2\big)\Big)\partial_xU=F.$$

It is trivial then to modify the scheme (4.4) by adding a forcing term $F_i^n=F(t^n,x_i)$.

We present in Table 2 the results that we obtain using the scheme (4.3) and (4.4), for several values of spatial and time discretization steps dx, dt, and different values of ϵ, and for times $T=1/\epsilon$. The relative errors are computed in the discrete L^2 norm. These results allow to validate the schemes proposed.

dx	dt	L	T	ϵ	KdV scheme	Boussinesq scheme
0.01	0.01	120	5	0.2	$9.6317.10^{-5}$	$9.8719.10^{-5}$
0.02	0.02	120	10	0.1	$7.7094.10^{-4}$	$7.9861.10^{-4}$
0.05	0.05	120	20	0.05	$9.5663.10^{-3}$	$9.8587.10^{-3}$

Table 2. Numerical errors of the KdV and the Boussinesq schemes.

In the following simulations, we always choose d$x=dt=0.01$.

4.2. Numerical results. We compute our schemes for two different different forms of initial data. The first leads to solitary waves as exact solutions of the KdV approximation (thus of the form (3.17)), and is therefore exponentially decreasing in space. The other initial data consist in deformations of the interface of the form $\frac{M}{\sqrt{1+kx^2}}$ (therefore, they do not satisfy the spatial rapidly decreasing assumption for the better convergence rate in

Proposition 3.12), as the surface is flat, and initial velocities are zeros. All the forthcoming results are expressed in non-dimensionalized variables.

We first look at the behavior of the relative difference between the solutions of the different models for fixed time, and for different values of ϵ. In Table 3, we provide the difference between the original Boussinesq/Boussinesq model (2.1) and our symmetric Boussinesq/Boussinesq system (2.2), at time $T = 1$ and for different values of ϵ. The same results are given for the comparison between the symmetric Boussinesq/Boussinesq system and the KdV approximation in Table 4. As predicted, when compared with the solution of the symmetric Boussinesq/Boussinesq model, the deviation of the original Boussinesq/Boussinesq system at fixed time is of order $\mathcal{O}(\epsilon^2)$, whereas the error of the KdV approximation is of order $\mathcal{O}(\epsilon)$. This supports our choice of the symmetric Boussinesq/Boussinesq system (2.2), as an equivalent model for system (2.1).

$\mathrm{d}x$	dt	L	T	ϵ	relative error of the solutions with the initial value: $U^0 = \frac{M}{\cosh(kx)^2}$	$U^0 = \frac{M}{\sqrt{1+(kx)^2}}$
0.01	0.01	120	1	0.1	$1.4859.10^{-3}$	$2.4785.10^{-3}$
0.01	0.01	120	1	0.05	$5.2714.10^{-4}$	$1.0882.10^{-3}$
0.01	0.01	120	1	0.01	$3.1656.10^{-5}$	$7.1958.10^{-5}$

TABLE 3. Relative error between the solutions of the symmetric Boussinesq/Boussinesq system and the original Boussinesq/Boussinesq model.

$\mathrm{d}x$	dt	L	T	ϵ	relative error of the solutions with the initial value: $U^0 = \frac{M}{\cosh(kx)^2}$	$U^0 = \frac{M}{\sqrt{1+(kx)^2}}$
0.01	0.01	120	1	0.1	$4.0805.10^{-3}$	$5.9295.10^{-3}$
0.01	0.01	120	1	0.05	$2.2785.10^{-3}$	$2.5506.10^{-3}$
0.01	0.01	120	1	0.01	$5.2496.10^{-4}$	$3.5434.10^{-4}$

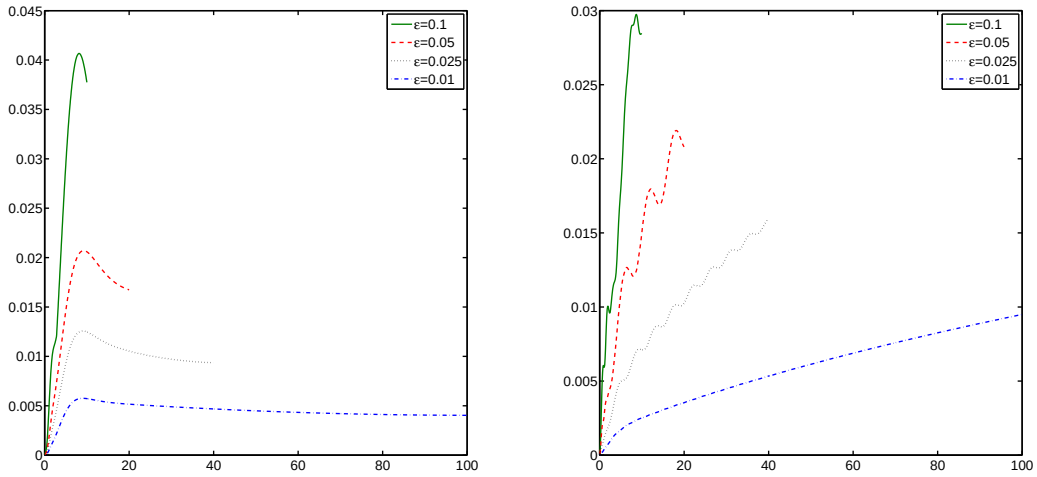
TABLE 4. Relative error between the solution of the symmetric Boussinesq/Boussinesq system and the solution of the KdV approximation.

Then, we give a numerical confirmation of the convergence rate obtained in Proposition 3.12. It is stated that for any initial data, the difference between the solution of the Boussinesq/Boussinesq system and the KdV approximation is bounded by $\mathcal{O}(\epsilon)$ over times of order $\mathcal{O}(1/\epsilon)$ if the initial data is sufficiently decreasing in space, and $\mathcal{O}(\epsilon\sqrt{t})$ otherwise.

We plot in Figure 7 the difference (in the discrete L^2 norm) of two solutions, obtained respectively by the KdV scheme and the Boussinesq scheme. The first plot concerns solitary waves of the form (3.17), thus exponentially decreasing in space, and the second is given by a zero velocities-flat surface initial data, with the interface of the form $\frac{M}{\sqrt{1+(kx)^2}}$.

For both configurations, we set $\delta = 1$, $\gamma = 1/4$, and simulate at the different values $\epsilon = 0.1, 0.05, 0.025, 0.01$, throughout time $T = 1/\epsilon$. One sees that the results match the theory, and that the decreasing in space at infinity is indeed of great concern.

The Figures 8 and 9 provide the snapshots of the previous simulations for $\epsilon = 0.1$, and at times $t = 20$ and $t = 40$. The results of the two schemes are plotted in the same figure, as well as the initial data and the (ten times) emphasized error at the interface, for readability. We only show the right side of the simulations, as the left side is obtained symmetrically.



(a) solitary waves

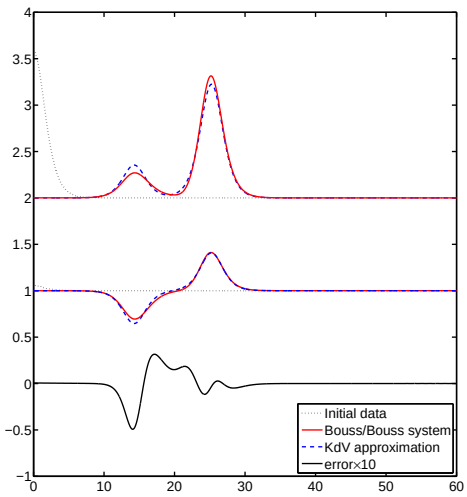
(b) zero velocities-flat surface initial value

Figure 7. Relative error between the KdV approximation and the symmetric Boussinesq/Boussinesq system: (a) for solitary waves, exponentially decreasing in space (b) for zero velocities-flat surface initial data, slowly decreasing in space.

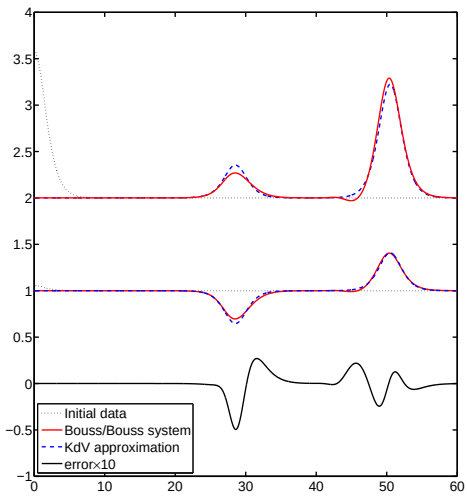
We see that in the case of a solitary wave, the Boussinesq/Boussinesq system produces an almost perfect soliton, that is predicted by the KdV approximation (Figure 8). The difference between the two solutions is located in space. However, when the initial data is not sufficiently decreasing at infinity, there is a big qualitative difference between the solution of the Boussinesq model, and the KdV approximation (Figure 9), since their difference is not exclusively located in the area of the solitary waves. Therefore, apart from a good initial localization in space, there is a significant interaction between the traveling waves of different wave modes, that is captured by the Boussinesq/Boussinesq system, and not by the KdV approximation.

The same simulations are produced for different parameters, *i.e.* $\delta = 2$ and $\gamma = 1/4$, and are displayed in Figures 10 and 11. We see that the exact same phenomenon appears.

Finally, we present from Figures 12 to 15 numerical simulations of the KdV approximation, in both the free surface and rigid lid configurations. The simulation of the Boussinesq/Boussinesq models leads to almost identical results, so that we do not plot them for readability. Again, we set $\epsilon = 0.1$, and the different times of the snapshots are $t = 20$ and $t = 40$. In Figures 12 and 13, we choose the parameters $\delta = 1$, and respectively $\gamma = 1/4$ and $\gamma = 9/10$. In Figures 14 and 15, we choose the parameters $\gamma = 1/4$, and respectively $\delta = 2$ and $\delta = 1/4$. Each time, the initial data consists in a bell curve in the interface, with no velocities and with a flat top. One clearly sees that, as discussed in Section 3.4, the rigid lid assumption is satisfactory as an approximation of the free surface, only in the case of small difference of densities ($\gamma \sim 1$), or when the depth ratio becomes large.

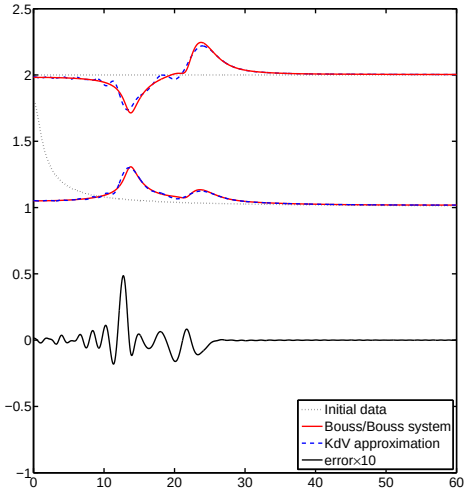


(a) $t = 20$. $\epsilon = 0.1, \delta = 1, \gamma = 1/4$.

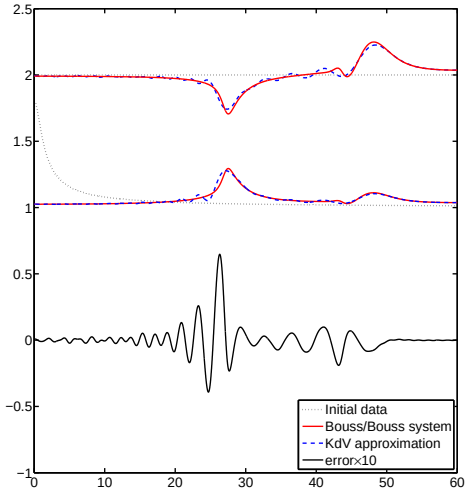


(b) $t = 40$. $\epsilon = 0.1, \delta = 1, \gamma = 1/4$.

FIGURE 8. Solitary wave solution of the KdV approximation, and the symmetric Boussinesq/Boussinesq model, at times (a) t=20, (b) t=40.

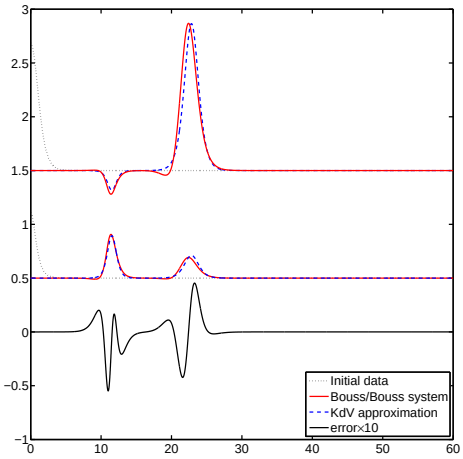


(a) $t = 20$. $\epsilon = 0.1, \delta = 1, \gamma = 1/4$.

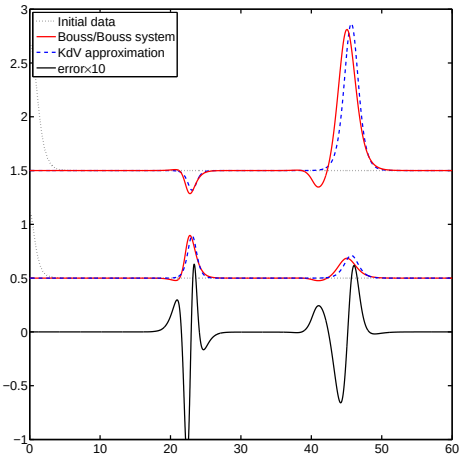


(b) $t = 40$. $\epsilon = 0.1, \delta = 1, \gamma = 1/4$.

FIGURE 9. Solution of the KdV approximation, and symmetric Boussinesq/Boussinesq system for a zero velocities-flat surface initial value, at times (a) t=20, (b) t=40.

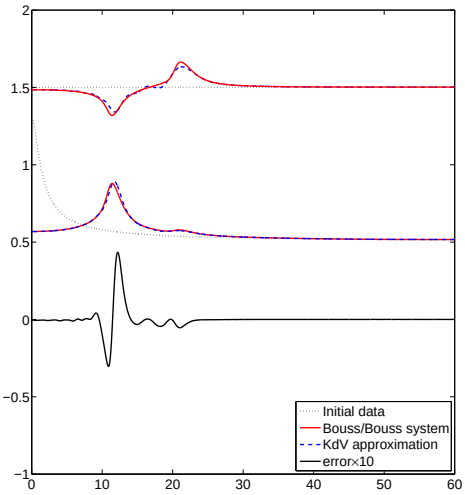


(a) $t = 20.\ \epsilon = 0.1, \delta = 2, \gamma = 1/4.$

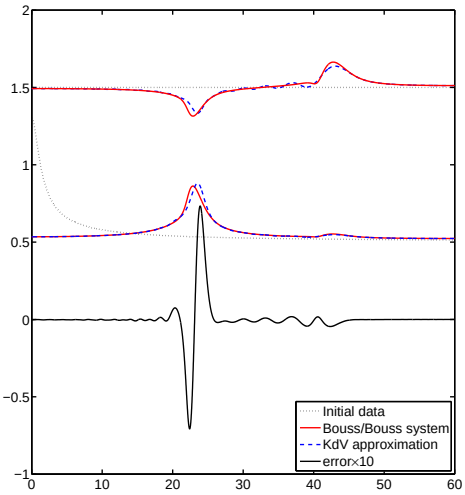


(b) $t = 40.\ \epsilon = 0.1, \delta = 2, \gamma = 1/4.$

Figure 10. Solitary wave solution of the KdV approximation, and symmetric Boussinesq/Boussinesq model, at times (a) t=20, (b) t=40.

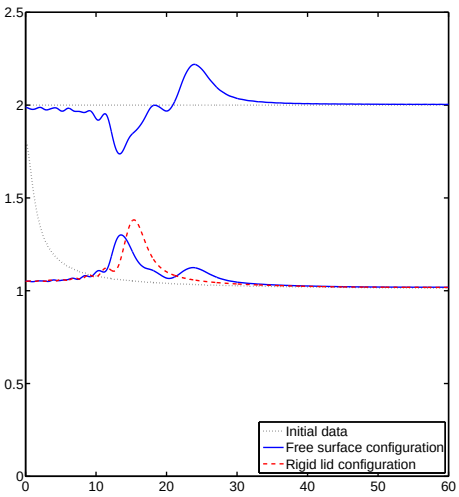


(a) $t = 20.\ \epsilon = 0.1, \delta = 2, \gamma = 1/4.$

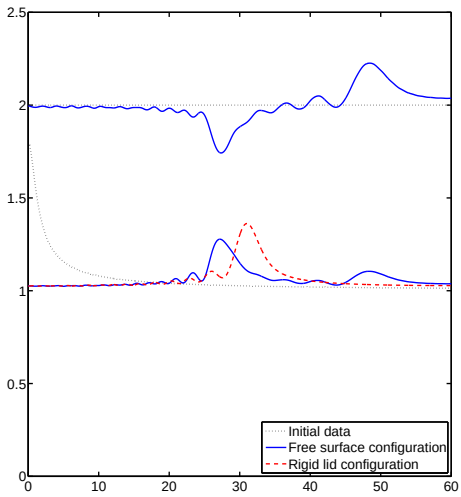


(b) $t = 40.\ \epsilon = 0.1, \delta = 2, \gamma = 1/4.$

Figure 11. Solution of the KdV approximation, and symmetric Boussinesq/Boussinesq system for a zero velocities-flat surface initial value, at times (a) t=20, (b) t=40.

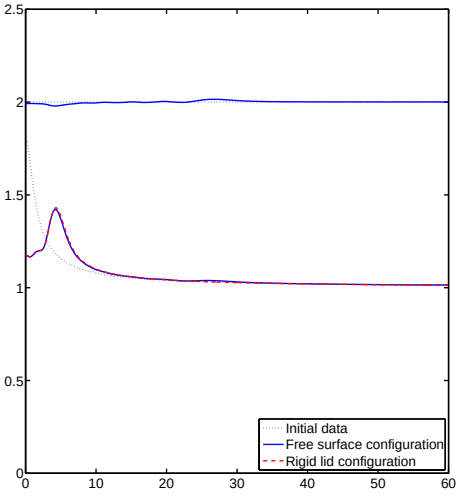


(a) $t = 20$. $\epsilon = 0.1, \delta = 1, \gamma = 1/4$.

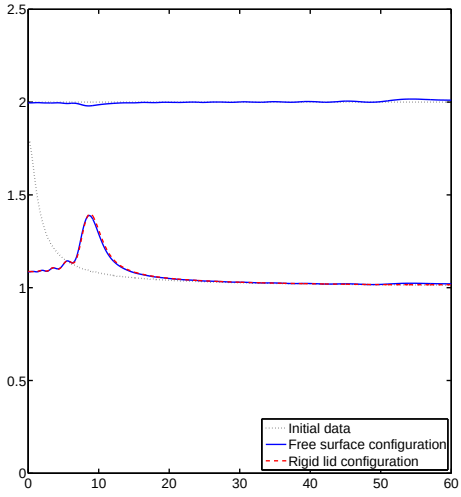


(b) $t = 40$. $\epsilon = 0.1, \delta = 1, \gamma = 1/4$.

FIGURE 12. Solution of the KdV approximations, in both rigid lid and free surface configurations, with two fluids of highly different densities, at times (a) t=20, (b) t=40.

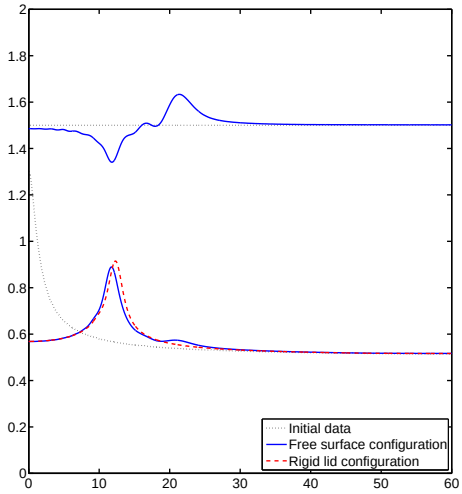


(a) $t = 20$. $\epsilon = 0.1, \delta = 1, \gamma = 9/10$.

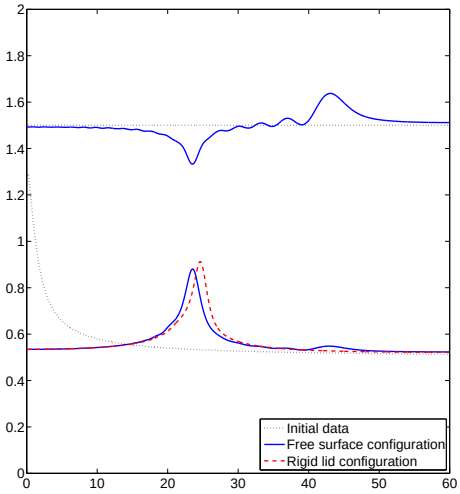


(b) $t = 40$. $\epsilon = 0.1, \delta = 1, \gamma = 9/10$.

FIGURE 13. Solution of the KdV approximations, in both rigid lid and free surface configurations, with two fluids of near equal densities, at times (a) t=20, (b) t=40.

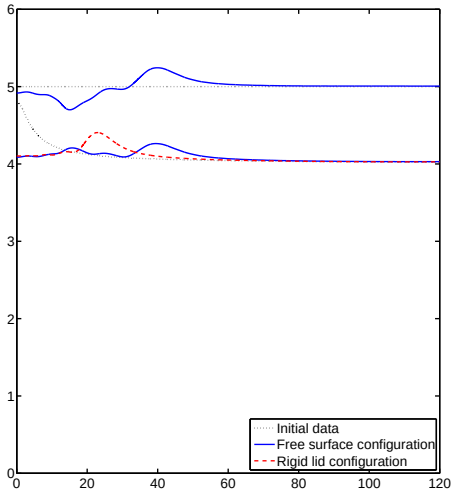


(a) $t = 20.\ \epsilon = 0.1, \delta = 2, \gamma = 1/4.$

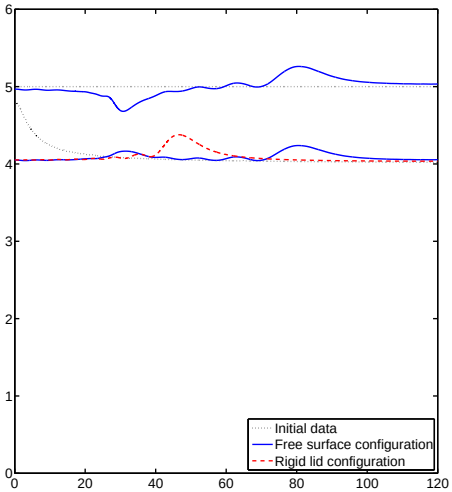


(b) $t = 40.\ \epsilon = 0.1, \delta = 2, \gamma = 1/4.$

FIGURE 14. Solution of the KdV approximations, in both rigid lid and free surface configurations, with the upper fluid of greater depth, at times (a) t=20, (b) t=40.



(a) $t = 20.\ \epsilon = 0.1, \delta = 1/4, \gamma = 1/4.$



(b) $t = 40.\ \epsilon = 0.1, \delta = 1/4, \gamma = 1/4.$

FIGURE 15. Solution of the KdV approximations, in both rigid lid and free surface configurations, with the lower fluid of greater depth, at times (a) t=20, (b) t=40.

A. Proof of Proposition 2.6

The proof is made of three steps. First, we introduce an energy of the system, and obtain an a priori estimate on this energy. Then, using this estimate and regularization operators, we prove the existence of a solution of our problem. Finally, using the energy estimate on the difference of two solutions, we get the uniqueness of the solution.

A.1. Energy estimate. In the following, we denote by U a solution of (2.2) on $[0,T]$, with $U \in L^\infty([0,T];H^{s+1})^4$. When we multiply (2.2) by Λ^s, the system becomes

$$\Big(S_0+\epsilon S_1(U)-\epsilon S_2\partial_x^2\Big)\partial_t\Lambda^s U+\epsilon[\Lambda^s,S_1(U)]\partial_t U+\Big(\Sigma_0+\epsilon\Sigma_1(U)-\epsilon\Sigma_2\partial_x^2\Big)\partial_x\Lambda^s U \\ +\epsilon[\Lambda^s,\Sigma_1(U)]\partial_x U=0. \tag{A.1}$$

We then introduce the energy associated to the system

Definition A.1. *Let $U\in H^{s+1}(\mathbb{R}^4)$. We define the energy of the function U associated to the system* (2.2) *as*

$$E_s(U)\equiv 1/2(S_0\Lambda^s U,\Lambda^s U)+\epsilon/2(S_1(U)\Lambda^s U,\Lambda^s U)+\epsilon/2(S_2\Lambda^s\partial_x U,\Lambda^s\partial_x U).$$

If there is no risk of confusion, we simply write E_s.

From Assumption 2.3, S_0 and S_2 are definite positive (with eigenvalues depending on γ and δ). Moreover, since $s>1/2$, one has by Sobolev embeddings $\big|S_1(U)\big|_{L^\infty}\le C_0\big|U\big|_{H^s}$. Hence, there exists $\alpha=C_0(\frac{1}{\gamma(1-\gamma)},\delta+\frac{1}{\delta})$ such that if $\epsilon\big|U\big|_{H^s}<\frac{1}{C_0(\alpha)}$, then

$$\frac{1}{\alpha}\big|U\big|^2_{H^{s+1}_\epsilon}\le E_s(U)\le\alpha\big|U\big|^2_{H^{s+1}_\epsilon}. \tag{A.2}$$

Let us multiply (A.1) on the right by $\Lambda^s U$ and integrate. One obtains

$$\frac{d}{dt}E_s=\epsilon/2(S_1(\partial_t U)\Lambda^s U,\Lambda^s U)-\epsilon([\Lambda^s,S_1(U)]\partial_t U,\Lambda^s U)+\epsilon/2((\Sigma_1(\partial_x U)\Lambda^s U),\Lambda^s U) \\ -\epsilon([\Lambda^s,\Sigma_1(U)]\partial_x U,\Lambda^s U). \tag{A.3}$$

Now, one has thanks to Cauchy-Schwarz inequality and Sobolev embeddings

$$|(\Sigma_1(\partial_x U)\Lambda^s U,\Lambda^s U)|\ \le\ \big|\Sigma_1(\partial_x U)\big|_{L^\infty}\big|\Lambda^s U\big|^2_{L^2}\ \le\ C_0\big|U\big|^3_{H^s}. \tag{A.4}$$

We then use classical Kato-Ponce commutator estimate **[58]**:

Lemma A.2. (Kato-Ponce) *For $s\ge 0$, if $f\in H^s$ and $g\in H^{s-1}$, then one has the estimate*

$$\big|[\Lambda^s,f]g\big|_{L^2}\le C_0\big|f\big|_{H^s}\big|g\big|_{L^\infty}+C_0\big|\partial_x f\big|_{L^\infty}\big|g\big|_{H^{s-1}}.$$

In our case, it leads to the commutator estimate

$$\big|([\Lambda^s,\Sigma_1(U)]\partial_x U,\Lambda^s U)\big|\le\big|([\Lambda^s,\Sigma_1(U)]\partial_x U\big|_2\big|\Lambda^s U\big|_2\le C_0\big|U\big|^3_{H^s}. \tag{A.5}$$

In order to deal with the nonlinear terms with a time-derivative ∂_t, we will need the following Lemma, using the elliptic form of the operator P_ϵ defined in (2.3):

Lemma A.3. *Let $P_\epsilon(U,\partial)$ and $Q_\epsilon(U,\partial)$ the differential operators defined in* (2.3) *and $s>1/2$. Then there exists $C_0=C_0(\frac{1}{\gamma(1-\gamma)},\delta+\frac{1}{\delta})$ such that if $\epsilon\big|U\big|_{H^s}<1/C_0$, one has the following properties:*

i. *$P_\epsilon(U,\partial):H^1\to H^{-1}$ is one-to-one, and onto.*
ii. *For $k\in[0,s]$, if one has $P_\epsilon(U,\partial)V\in H^{k-1}$, then $V\in H^{k+1}$.*
iii. *For $1/2<k\le s$ and $V\in H^{k+1}$, one has*

$$1/C_0\big|V\big|_{H^{k+1}_\epsilon}\le(P_\epsilon(U,\partial)V,V)\le C_0\big|V\big|_{H^{k+1}_\epsilon}.$$

iv. *For $1/2 < k \leq s$, the operator $P_\epsilon(U,\partial_x)^{-1}Q_\epsilon(U,\partial_x)$ is bounded $H^{k+1}_\epsilon \to H^{k+1}_\epsilon$, uniformly with respect to ϵ.*

Proof. i. Using the fact that S_0 and S_2 are symmetric definite positive (with eigenvalues depending on γ and δ) in the formulation of P_ϵ, it is obvious to see that one can choose $C_0 = C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}) > 0$ such that for $\epsilon\big|U\big|_{L^\infty} < 1/C_0$ and any $V \in H^1$,

$$(P_\epsilon(U,\partial)V, V) = (S_0V, V) + \epsilon(S_1(U)V, V) + \epsilon(S_2\partial_x V, \partial_x V) \geq \frac{\epsilon}{C_0}\big|V\big|^2_{H^1}.$$

In the same way, the bilinear form $a(V,W) = (P_\epsilon(U,\partial)V, W)$ is clearly continuous on $H^1 \times H^1$:

$$(P_\epsilon(U,\partial)V, V) = (S_0V, V) + \epsilon(S_1(U)V, V) + \epsilon(S_2\partial_x V, \partial_x V) \leq C_0\big|V\big|^2_{H^1}.$$

Using Lax-Milgram lemma, we obtain that for all $F \in H^{-1}$, there exists a unique $V \in H^1$ such that for all $W \in H^1$, $a(V,W) = (F,W)$, and hence there exists a unique variational solution of

$$P_\epsilon(U,\partial)V = F.$$

ii. We will prove the second point by induction. The result is known for $k = 0$ from the previous point. Then, we remark that if $P_\epsilon(U,\partial)V = W$ with $W \in L^2$ (and $V \in H^1$), then

$$\epsilon S_2\partial_x^2 V = S_0V + \epsilon S_1(U)V - W.$$

Therefore, since S_2 is invertible, we get $\partial_x^2 V \in L^2$, and thus $V \in H^2$. The second point is therefore proved for $k = 1$, and the intermediate values $k \in [0,1]$ follow by interpolation.

Let now assume that the result is known for a $k \in \mathbb{N}^*$ and let $P_\epsilon(U,\partial)V \in H^k \subset H^{k-1}$. From the induction hypothesis, one has $V \in H^{k+1}$, and moreover

$$\epsilon S_2\partial_x^2\Lambda^k V = S_0\Lambda^k V + \epsilon\Lambda^k(S_1(U)V) - \Lambda^k P_\epsilon(U,\partial)V.$$

We know that $\Lambda^k V \in L^2$ and $\Lambda^k P_\epsilon(U,\partial)V \in L^2$. Moreover, since $U \in H^s$ and $s \geq k > 1/2$, we know from the Sobolev embedding that $\Lambda^k(S_1(U)V) \in L^2$ and $\big|\Lambda^k(S_1(U)V)\big|_2 \leq C_0\big|U\big|_{H^k}\big|V\big|_{H^k}$. Therefore, $\partial_x^2\Lambda^k V \in L^2$, and as we know from the induction hypothesis that $V \in H^{k+1}$, one has indeed $V \in H^{k+2}$. The second point is thus proved for all $k \in \mathbb{N} \cap [0,s]$, and we obtain the result for $0 \leq k \leq s$ in the same way, starting the induction from $k - \lfloor k \rfloor \in [0,1)$.

iii. Since S_0 and S_2 are symmetric definite positive, one has for $\epsilon\big|U\big|_{H^s} < 1/C_0$ with $C_0 = C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta})$, and for any $V \in H^{k+1}$,

$$\begin{aligned}(\Lambda^k(P_\epsilon(U,\partial)V), \Lambda^k V) &= (S_0\Lambda^k V, \Lambda^k V) + \epsilon(\Lambda^k(S_1(U)V), \Lambda^k V) + \epsilon(S_2\partial_x\Lambda^k V, \partial_x\Lambda^k V)\\ &\geq \frac{1}{C_0}\big|V\big|^2_{H^k} + \frac{1}{C_0}\epsilon\big|V\big|^2_{H^{k+1}}.\end{aligned}$$

The second inequality is straightforward.

iv. Let $W \in H^{k+1}$. It is obvious that, since $s \geq k > 1/2$, one has

$$Q_\epsilon(U,\partial_x)W = \big(\Sigma_0 + \epsilon\Sigma_1(U)\big)W - \epsilon\Sigma_2\partial_x^2 W \in H^{k-1}.$$

Hence, we know from the previous points that $V \equiv P_\epsilon(U,\partial_x)^{-1}Q_\epsilon(U,\partial_x)W \in H^{k+1}$, and

$$\frac{1}{C_0}\big|V\big|^2_{H^{k+1}_\epsilon} \leq (\Lambda^k(P_\epsilon(U,\partial)V), \Lambda^k V) \leq C_0\big|V\big|^2_{H^{k+1}_\epsilon}.$$

Therefore, one has:

$$\begin{aligned}\big|V\big|^2_{H^{k+1}_\epsilon} &\leq C_0(\Lambda^k(Q_\epsilon(U,\partial)W),\Lambda^k V)\\ &\leq C_0(\Lambda^k(\Sigma_0+\epsilon\Sigma_1(U))W),\Lambda^k V)+C_0(\partial_x\Lambda^k\Sigma_2 W),\partial_x\Lambda^k V)\\ &\leq C_0(\gamma,\frac{1}{\delta},\big|U\big|_{H^k})\big(\big|W\big|_{H^k}\big|V\big|_{H^k}+\epsilon\big|W\big|_{H^{k+1}}\big|V\big|_{H^{k+1}}\big).\end{aligned}$$

Finally, one has

$$\begin{aligned}\big|V\big|^2_{H^{k+1}_\epsilon} \leq&\frac{1}{2}\big(C_0\big|W\big|_{H^k}-\big|V\big|_{H^k}\big)^2+\frac{\epsilon}{2}\big(C_0\big|W\big|_{H^{k+1}}-\big|V\big|_{H^{k+1}}\big)^2\\ &+\frac{1}{2}\big|V\big|^2_{H^{k+1}_\epsilon}+\frac{1}{2}C_0^2\big|W\big|^2_{H^{k+1}_\epsilon}\\ \leq&\frac{1}{2}\big|V\big|^2_{H^{k+1}_\epsilon}+\frac{1}{2}C_0^2\big|W\big|^2_{H^{k+1}_\epsilon}.\end{aligned}$$

The operator $P_\epsilon(U,\partial_x)^{-1}Q_\epsilon(U,\partial_x)$ is thus bounded $H^{k+1}_\epsilon\to H^{k+1}_\epsilon$ by C_0^2, which ends the proof of the Lemma.

□

Using this Lemma, we immediately obtain that for $U\in H^{s+1}$ satisfying (2.3) and the hypothesis of the Lemma, one has $\partial_t U\in H^s$, and

$$\big|\partial_t U\big|_{H^s_\epsilon}=\big|P_\epsilon(U,\partial_x)^{-1}Q_\epsilon(U,\partial_x)\partial_x U\big|_{H^s_\epsilon}\leq C_0\big|\partial_x U\big|_{H^s_\epsilon}\leq C_0\big|U\big|_{H^{s+1}_\epsilon}. \tag{A.6}$$

Therefore, we can use the same calculations as in (A.4) and (A.5), and obtain

$$|(S_1(\partial_t U)\Lambda^s U,\Lambda^s U)|\leq\big|S_1(\partial_t U)\big|_{L^\infty}\big|\Lambda^s U\big|^2_{L^2}\leq C_0\big|U\big|^2_{H^s}\big|U\big|_{H^{s+1}_\epsilon} \tag{A.7}$$

(since $s-1>1/2$ such that $\big|\partial_t U\big|_{L^\infty}\leq C_0\big|\partial_t U\big|_{H^{s-1}}\leq C_0\big|U\big|_{H^{s+1}_\epsilon}$), and with Kato-Ponce theorem,

$$\begin{aligned}|([\Lambda^s,S_1(U)]\partial_t U,\Lambda^s U)|&\leq\big|([\Lambda^s,S_1(U)]\partial_t U\big|_{L^2}\big|\Lambda^s U\big|_{L^2}\\ &\leq C_0\big|U\big|^2_{H^s}\big|\partial_t U\big|_{H^{s-1}}\leq C_0\big|U\big|^2_{H^s}\big|U\big|_{H^{s+1}_\epsilon}.\end{aligned} \tag{A.8}$$

Finally, one deduces from (A.2), (A.3) and the estimates (A.4)-(A.8):

$$\frac{d}{dt}E_s\leq\epsilon C_0\ \big|U\big|^2_{H^s}\big|U\big|_{H^{s+1}_\epsilon}\leq\epsilon C_0E_s^{3/2},$$

with $C_0=C_0(\frac{1}{\gamma(1-\gamma)},\delta+\frac{1}{\delta})$, and providing the fact that $\epsilon\big|U\big|_{L^\infty H^s}<1/C_0$.

From Gronwall-Bihari's inequality and (A.2), it follows

$$\big|U\big|_{H^{s+1}_\epsilon}\leq C_0E_s^{1/2}\leq C_0\frac{E_s^{1/2}\big|_{t=0}}{1-C_0\epsilon tE_s^{1/2}\big|_{t=0}}\leq C_0\frac{\big|U^0\big|_{H^{s+1}_\epsilon}}{1-C_0\epsilon\big|U^0\big|_{H^{s+1}_\epsilon}t}. \tag{A.9}$$

Thus, one sees that there exists C_0 such that if $\epsilon\big|U_0\big|_{H^{s+1}_\epsilon}<1/C_0$, then one can choose $T=T(C_0)$ such that the smallness assumption $\epsilon\big|U\big|_{H^s}<1/C_0$ remains valid for any $t\in[0,T/\epsilon]$. Hence, in the frame of the proposition, one has the following estimate:

$$\big|U\big|_{L^\infty([0,T/\epsilon];H^{s+1}_\epsilon)}\leq C_0\frac{\big|U^0\big|_{H^{s+1}_\epsilon}}{1-C_0\epsilon\big|U^0\big|_{H^{s+1}_\epsilon}t}, \tag{A.10}$$

with $C_0=C_0(\frac{1}{\gamma(1-\gamma)},\delta+\frac{1}{\delta})$ and $T>0$, independent of ϵ.

Let us note that by Lemma A.3, we know that we can also control the time derivative $\big|\partial_t U\big|_{L^\infty([0,T/\epsilon];H^s_\epsilon)}$ with the same bound.

A.2. Existence of a solution. We can deduce from the energy estimate (A.10) the existence of a maximal solution $U \in L^{1,\infty}_t([0,T/\epsilon];H^{s+1})$ of (2.2) for any initial data $U^0 \in H^{s+1}$. We follow the classical Friedrichs proof, using the regularization operators defined thanks to the Fourier transform as below:

$$\forall v \in L^2, \quad \forall \xi \in \mathbb{R}, \quad \widehat{J_\nu v(\xi)} \equiv \varphi(\nu\xi)\widehat{v}(\xi), \tag{A.11}$$

with φ a smooth numeric function with compact support, such that $\varphi(0) = 1$. These operators have the following classical properties

Lemma A.4. i. *J_ν is bounded $H^s \to H^s$: there exists $C_0(s,\nu)$ such that for any $v \in H^s$,*

$$\big|J_\nu v\big|_{H^s} \leq C_0(s,\nu)\big|v\big|_{H^s} \tag{A.12}$$

ii. *J_ν commutes with Λ^s, and is a self-adjoint operator*
iii. *There exists C_0 independent of ν such that*

$$\big|J_\nu v\big|_2 \leq C_0\big|v\big|_2 \tag{A.13}$$

We obtain then a solution of (2.3) as the limit of U_ν the solutions of

$$\partial_t U_\nu + J_\nu P_\epsilon(J_\nu U_\nu, \partial_x)^{-1} Q_\epsilon(J_\nu U_\nu, \partial_x) J_\nu \partial_x U_\nu = 0, \tag{A.14}$$

with $U_\nu\,|_{t=0} = U^0$.

From Lemma A.4, one has that (A.14) is an ordinary differential equation on the Banach space H^s. Thus, thanks to Cauchy-Lipschitz Theorem, we know that there exists a unique maximal solution $U_\nu \in C([0,T_\nu),H^s)$.

Using the calculations of Section A.1 and the properties of J_ν, one obtains the energy estimate

$$\big|U_\nu\big|_{H^{s+1}_\epsilon} \leq C_0 \frac{\big|U^0\big|_{H^{s+1}_\epsilon}}{1 - C_0\epsilon\big|U^0\big|_{H^{s+1}_\epsilon} t},$$

with C_0 independent of ν. Moreover, U_ν satisfies (A.6):

$$\big|\partial_t U_\nu\big|_{H^{s-1}} \leq C_0\big|\partial_x U\big|_{H^{s-1}} \leq C_0\big|U\big|_{H^s}. \tag{A.15}$$

Thus one can find $T > 0$ independent of ν such that the solution U_ν does not blow up for $t \in [0,T/\epsilon]$. In particular, one has $T_\nu > T/\epsilon > 0$.

Now, since $(U_\nu)_{\nu>0}$ is uniformly bounded in $L^{1,\infty}_t([0,T/\epsilon],H^{s+1}_\epsilon)$, one can extract a subsequence that weakly converges towards a function $U \in L^{1,\infty}_t([0,T/\epsilon],H^{s+1}_\epsilon)$. We want to use Ascoli theorem, but we need the injection $H^s(\mathbb{R}) \subset H^{s-1}(\mathbb{R})$ to be compact, which is not true since $\mathbb{R}$ is unbounded. However, one can easily circumvent this problem, using weighted Sobolev spaces for example. Finally, one obtains that there exists $U \in C^0([0,T/\epsilon),H^{s-1})$ such that U_ν converges strongly towards U as the subsequence ν tends to 0. Also, by interpolation inequalities, U_ν converges strongly towards U in $C^0([0,T/\epsilon),H^{s-\alpha})$, for $0 < \alpha \leq 1$. Then, since one can find α such that $H^{s-\alpha}$ injects continuously in $C^1(\mathbb{R})$, one proves that $Q_\epsilon(J_\nu U_\nu,\partial_x)J_\nu\partial_x U_\nu$ converges to $Q_\epsilon(U,\partial_x)\partial_x U$ and $J_\nu P_\epsilon(J_\nu U_\nu,\partial_x)\partial_t U$ converges to $P_\epsilon(U_\nu,\partial_x)\partial_t U$ as $\nu \to 0$. Hence, U is indeed a solution of (2.3).

From Section A.1, we know that $U \in L^{1,\infty}_t([0,T/\epsilon];H^{s+1}_\epsilon)$, and one can prove (see **[111]** XVI.1.4 for example) that $U \in C^0([0,T/\epsilon);H^{s+1}) \cap C^1([0,T/\epsilon);H^s)$.

A.3. Uniqueness. Let $U_1, U_2 \in C^0([0,T/\epsilon);H^{s+1})\cap C^1([0,T/\epsilon);H^s)$ be two solutions of the Cauchy problem (2.2) with initial data $U_1\,|_{t=0} = U_2\,|_{t=0} = U^0$. One can immediately check that $R \equiv U_1 - U_2$ satisfies

$$\begin{aligned}\Big(S_0 + \epsilon S_1(U_1) - \epsilon S_2\partial_x^2\Big)\partial_t\Lambda^s R + \Big(\Sigma_0 + \epsilon\Sigma_1(U_1) - \epsilon\Sigma_2\partial_x^2\Big)\partial_x\Lambda^s R \\ +\epsilon[\Lambda^s, S_1(U_1)]\partial_t R + \epsilon[\Lambda^s,\Sigma_1(U_1)]\partial_x R = \epsilon F,\end{aligned} \tag{A.16}$$

with $F = -\Lambda^s\Big(S_1(R)\partial_t U_2 + \Sigma_1(R)\partial_x U_2\Big)$. Then, we can carry out the same calculations as in Section A.1 on R, and obtain the equivalent energy estimate

$$\frac{d}{dt}E_s(R) \le \epsilon C_0(\big|U_1\big|_{H^s} + \big|U_2\big|_{H^s})E_s,$$

with $C_0 = C_0(\frac{1}{\gamma(1-\gamma)}, \delta + \frac{1}{\delta}, \big|U^0\big|_{H^{s+1}_\epsilon})$.

From Gronwall-Bihari's inequality and the estimate (A.10) on U_1 and U_2, and since $E_s(R)\,|_{t=0} = 0$, one has immediately $E_s(R) = 0$ on $[0,T/\epsilon]$, and finally $U_1 = U_2$.

CHAPITRE 4

Asymptotic models for the generation of internal waves by a moving ship, and the dead-water phenomenon

Article soumis (preprint arXiv :1012.5892).

Sommaire

Abstract

This chapter deals with the dead-water phenomenon, which occurs when a ship sails in a stratified fluid, and experiences an important drag due to waves below the surface. More generally, we study the generation of internal waves by a disturbance moving at constant speed on top of two layers of fluids of different densities. Starting from the full Euler equations, we present several nonlinear asymptotic models, in the long wave regime. These models are rigorously justified by consistency or convergence results. A careful theoretical and numerical analysis is then provided, in order to predict the behavior of the flow and in which situations the dead-water effect appears.

1. Introduction

The so-called "dead-water" phenomenon has been first reported by Fridtjof Nansen **[92]**, describing a severe (and then inexplicable) decrease of velocity encountered by a ship, sailing on calm seas. Bjerknes, and then Ekman **[43]** soon explained that this phenomenon occurs as the ship, on top of density-stratified fluids (due to variations in temperature

or salinity), concedes a large amount of energy by generating internal waves. Our work is motivated by the recent paper of Vasseur, Mercier and Dauxois [**114**], who performed experiments that revealed new insights on the phenomenon.

Relatively little consideration has been given to this problem, after the early works of Lamb [**66**] and Sretenskii [**108**]. Miloh, Tulin and Zilman in [**88**] produce a model and numerical simulations for the case of a semi-submersible moving steadily on the free surface. The authors assume that the density difference between the two layers is small, an assumption that is removed by Nguyen and Yeung [**94**]. Motygin and Kuznetsov [**91**] offer a rigorous mathematical treatment of the issue when the body is totally submerged in one of the two layers, and Ten and Kashiwagi [**112**] present a comparison between theory and experiments, in the case of a body intersecting the interface as well as the surface. Finally, we would like to cite Lu and Chen [**79**] for a more general treatment of the problem. All of these works use linearized boundary conditions, that rely on the assumption that the amplitudes of the generated waves are small. The linear theory is convenient as it allows to obtain the flow field by a simple superposition of Green functions, replacing the moving body by a sum of singularities. The integral representation of the flow, as well as the wave resistance experienced by the body, are therefore found explicitly. However, the smallness assumption on the wave amplitudes, that is necessary to the linear theory, is quite restrictive, and the experiences in [**114**] clearly exhibit nonlinear features. Our aim is to produce simple nonlinear models, that are able to predict the apparition and the magnitude of the dead-water effect, and recover duly most of its key aspects.

In this paper, we introduce several different asymptotic models, corresponding to two different regimes (size of the parameters of the system). Each time, we assume that the depth of the fluid layers are small when compared with the wave lengths (shallow water). In the one-layer case, the equations obtained with this single assumption are the shallow-water or Saint Venant equations (at first order), or the Green-Naghdi extension (one order further, therefore involving nonlinear dispersion terms). See Wu, Chen [**117**] and references therein for a numerical treatment of the waves generated by a moving ship in the one layer case, using the Green-Naghdi equations. The two regimes we study carry additional smallness assumptions, that allow to substantially simplify the models. The first regime considers the case of small surface deformations, and small differences between the densities of the two layers. Such additional assumptions are very natural in the oceanographic framework, and have been frequently used in the literature (see [**60, 78, 163**] for example). In the second regime, we assume that the magnitude of the produced internal waves, when compared with the depth of the two layers, are small and of the same order of magnitude as the shallowness parameter. This regime, known as the Boussinesq regime, is particularly interesting as it allows models with competing dispersion and nonlinearity. Along with the coupled Boussinesq-type model, we introduce the KdV approximation, which consists in a decomposition of the flow into two waves, each one being a solution of an independent forced Korteweg-de Vries equation (fKdV). The fKdV equation has been extensively studied in the framework of the one-layer water wave problem (where a moving topography, or pressure, is the forcing term that generates waves); see [**119, 80, 75**], for example.

The system we study consists in two layers of immiscible, homogeneous, ideal, incompressible fluids only under the influence of gravity. The bottom is assumed to be flat, and we use the rigid-lid approximation at the surface. However, the surface is *not flat*, but a given function, moving at constant speed, that reflects the presence of the ship. Moreover, as we are interested in unidirectional equations (the fKdV equations), we focus on the two-dimensional case, *i.e.* the horizontal dimension $d = 1$. However, our method could easily be extended to the three-dimensional configuration. Starting from the governing equations of our problem, the so-called full Euler system, and armed with the analysis of [**40, 41**] in the free surface case, we are able to deduce asymptotic models for each of the regimes

presented above. Each of the models presented here are justified by a consistency result, or a convergence theorem (in a sense precisely given in Section 2.3 page 134). We compute numerically the fully nonlinear system of the first regime, as well as the KdV approximation, which allows us to investigate the effect of different parameters of the system, such as the velocity of the boat, or the depth ratio of the two layers. The wave resistance encountered by the ship is also computed, so that we are able to predict in which situations the dead-water effect shows up.

Our models allow to recover most of the known aspects of the dead-water phenomenon (see **[114, 88, 94]**), especially:

i. transverse internal waves are generated at the rear of a body while moving at the surface;
ii. the body suffers from a positive drag when an internal elevation wave is located at its stern;
iii. this effect can be strong (as the generated wave reach large amplitudes) near critical Froude numbers, that is when the velocity of the body approaches maximum internal wave speed. The effect is always small away from the critical value;
iv. the maximum peak of the drag is reached at slightly subcritical values.

We would like to emphasize that these considerations are not new. In particular, the fact that the drag experienced by the ship is strong only when its velocity approaches the maximum internal wave speed has been observed through experiments **[43, 114]**, and the maximum peak of the drag being reached at slightly subcritical values is predicted by linear models **[88, 94]**. The main contribution of our work is to offer simple adapted models, that are rigorously justified for a wide range of parameters (in particular, contrarily to linear models, our models allow large amplitude waves, in agreement with experiments).

Our models allow to easily investigate the dead-water phenomenon for several values of the parameter. For example, we numerically compute the system for several values of the depth ratio between the two layers. It appears that, contrarily to the intuition, the dead-water phenomenon is stronger when the upper layer is thicker than the lower one. This observation is new, as far as we know. This curiosity is due to the fact that the magnitude of the drag suffered by the body does not depend on the distance between the body and the generated internal wave, but rather on the amplitude and sharpness of the latter. A thicker upper layer allows the generation of internal elevation waves reaching very large amplitudes, when compared with the size of the body, which strengthen the dead-water phenomenon.

The behavior of the system, depending on the depth ratio of the two layers of fluid and the normalized velocity of the body, is summarized in Table 1, below.

However, one peculiar phenomenon, that is described in details in **[114]**, is not recovered by our models. Indeed, the dead-water effects exhibits a somewhat periodic behavior, where during each period, a wave is generated, slows down the ship, and then breaks. This discrepancy between our simulations and the experiments is due to the fact that our models are based on the assumption of a constant velocity for the traveling body, while their experiments are conducted with a constant force brought to the body **[113, 114]**. The latter setting is of course more natural, but the constant velocity hypothesis is crucial in our analysis (and is, in fact, consistently assumed within existing models in the literature **[43, 88, 120]**). We perform a numeric experiment which roughly matches this setting (the velocity is then adjusted at each time step as a function of the drag suffered by the ship), and recover the oscillating behavior. This point is discussed in more details on page 141.

Velocity of the ship	subcritical case	critical case		supercritical case
Depth ratio between the layers		thicker lower layer	thicker upper layer	
Regime 1	The generated waves are small, and conclusions of Regime 2 apply.	Generation of dull elevation-depression wave below the body. Moderate wave resistance. See Figure 3, page 140	Generation of a sharp elevation wave below the body. Strong wave resistance. See Figure 4, page 141	The generated waves are small, and conclusions of Regime 2 apply.
Regime 2	Continuous generation of small up-stream propagating and down-stream propagating waves. Very weak wave resistance. See Figure 6, page 150	Periodic generation of up-stream propagating depression waves with very large time-period. Oscillating wave resistance, with positive mean. See Figure 9, page 152	Periodic generation of up-stream propagating elevation waves with very large time-period. Oscillating wave resistance, with mean zero. See Figure 9, page 152	Generation of small down-stream propagating waves. Convergence towards a steady state. No lasting wave resistance. See Figure 7, page 151

TABLE 1. Behavior of the flow in the two studied regimes, depending on the velocity of the body and the depth ratio of the fluids

OUTLINE OF THE PAPER In Section 2.1, we introduce the governing equations of our problem: the so-called full Euler system. A specific care is given to an adapted rescaling, that leads to the dimensionless version of the full Euler system (2.4). The models presented here are asymptotic approximations of this system. In Section 2.3, we present precisely the different means of justification of our models, as well as the regimes at stake. The main tool that we use in order to construct our models, as well as a broad, strongly nonlinear model (2.7) are presented in Section 2.4.

The simpler models we study are deduced from (2.7), using the additional assumptions of the regimes considered. Strongly nonlinear models (3.1) and (3.3) are introduced in Section 3, and justified with consistency results. Numerical simulations are then displayed and discussed in Section 3.2. The weakly nonlinear models, *i.e.* the Boussinesq-type system (4.2) (and its symmetrized version (4.3)), and the KdV approximation (4.12), are presented in Sections 4.1 and 4.2, respectively. The convergence of their solutions towards the solutions of the full Euler system are given in Propositions 4.1 and 4.2, respectively. An in-depth analysis of the forced Korteweg-de Vries equation, and its consequences to the dead-water effect, is then displayed in Section 4.3.

Finally, in Section 5, we conclude with an overview of the results presented here, together with suggestions for future work.

Calculations that lead to (2.7) are postponed until Appendix A, and Appendix B contains the proof of Proposition 4.2. Finally, Appendix C is devoted to the analysis of the wave resistance encountered by the ship, and the numerical schemes used in the simulations are presented and justified in Appendix D.

NOTATIONS We denote by $C(\lambda_1, \lambda_2, \ldots)$ any positive constant, depending on the parameters $\lambda_1, \lambda_2, \ldots$, and whose dependence on λ_j is always assumed to be nondecreasing.

Let $f(x_1, \ldots, x_d)$ be a function defined on $\mathbb{R}^d$. We denote by $\partial_{x_i} f$ the derivative with respect to the variable x_i. If f depends only on x_i, then we use the notation

$$\frac{\mathrm{d}}{\mathrm{d}x_i} f \;\equiv\; \partial_{x_i} f.$$

We denote by $L^2 = L^2(\mathbb{R})$ the Hilbert space of Lebesgue measurable functions $F : \mathbb{R} \to \mathbb{R}^n$ with the standard norm $\big|F\big|_2 = \big(\int_{\mathbb{R}} |F(x)|^2 \, \mathrm{d}x \big)^{1/2} < \infty$. Its inner product is denoted by $\big(F_1 \, , \, F_2 \big) = \int_{\mathbb{R}} F_1 \cdot F_2$.

Then, we denote by $H^s = H^s(\mathbb{R})$ the L^2-based Sobolev spaces. Their norm is written $\big| \cdot \big|_{H^s} = \big| \Lambda^s \cdot \big|_2$, with the pseudo-differential operator $\Lambda \equiv (1 - \partial_x^2)^{1/2}$. It is also convenient to introduce the following norm on the Sobolev space H^{s+1} (for $\varepsilon > 0$ a small parameter):

$$\big|U\big|^2_{H^{s+1}_\varepsilon} = \big|U\big|^2_{H^s} + \varepsilon \big|U\big|^2_{H^{s+1}}.$$

Let $0 < T \leq \infty$ and $f(t,x)$ be a function defined on $[0,T] \times \mathbb{R}$. Then one has $f \in L^\infty([0,T]; H^s)$ if f is bounded in H^s, uniformly with respect to $t \in [0,T]$. Moreover, $f \in W^{1,\infty}([0,T]; H^s)$ if f, $\partial_t f \in L^\infty([0,T]; H^s)$. Their respective norm is written $\big| \cdot \big|_{L^\infty H^s}$ and $\big| \cdot \big|_{W^{1,\infty} H^s}$.

2. Construction of asymptotic models

Our framework, the formulation of the system using Dirichlet-Neumann operators, as well as the techniques used for constructing the asymptotic models, are not new. We follow here the strategy initiated in **[14, 15, 16]** in the one-layer (water wave) configuration, and extended to the bi-fluidic case in **[17, 40, 41]**. The major difference here is that the surface is not free, but rather a non-flat rigid lid, moving at constant velocity.

We present in the following, as briefly as possible, the governing equations of our problem. The full details of the construction can be found in the references above. Our models are then derived starting from the non-dimensionalized version of these equations (the full Euler system (2.4); see Section 2.2). The idea is to use smallness assumptions on some parameters, that characterize natural quantities of the system, to construct the asymptotic models. Together with the means to justify our models, the regimes under study are precisely described in Section 2.3. Finally, the main tool that we use in order to construct our models, that is the asymptotic expansion of the Dirichlet-Neumann operators in terms of small shallowness parameter μ, is presented in Section 2.4. A broad, nonlinear model (2.7) is then deduced, and the precise systems we study are derived from it, using the additional assumptions of Regime 1 and Regime 2, respectively in Section 3 and Section 4.

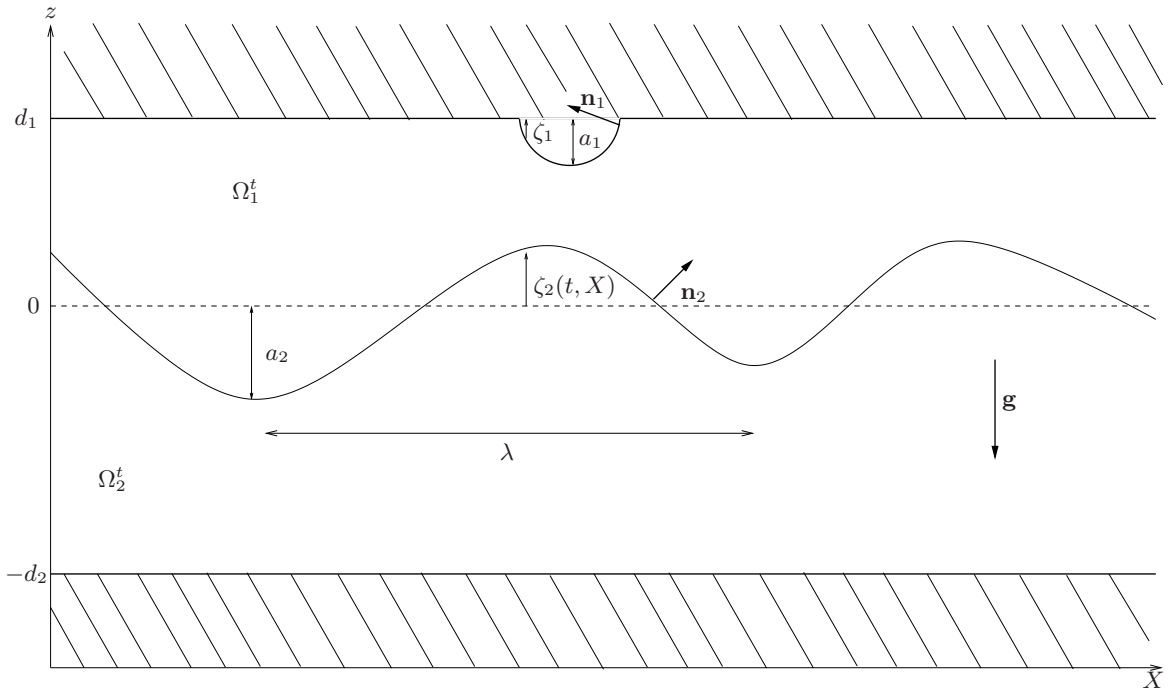


Figure 1. Sketch of the domain

2.1. The basic equations. The system we study consists in two layers of immiscible, homogeneous, ideal, incompressible fluids only under the influence of gravity (see Figure 1). Since we are interested in unidirectional models (the KdV approximation), we focus on the two-dimensional case, *i.e.* the horizontal dimension $d = 1$. However, most of this study could easily be extended to the case $d = 2$, using the techniques presented here.

We denote by ρ_1 and ρ_2 the density of, respectively, the upper and the lower fluid. Since we assume that each fluid is incompressible, ρ_1 and ρ_2 are constant and the velocity potentials ϕ_i $(i = 1, 2)$, respectively associated to the upper and lower fluid layers, satisfy the Laplace equation

$$\partial_x^2\phi_i \ + \ \partial_z^2\phi_i \ = \ 0.$$

Moreover, it is presumed that the surface and the interface are given as the graph of functions (respectively, $\zeta_1(t, x)$ and $\zeta_2(t, x)$) which express the deviation from their rest position (respectively, (x, d_1) and $(x, 0)$) at the spatial coordinate x and at time t. The bottom is assumed to be flat, and the surface is flat away from the location of the ship moving at constant speed c_s, so that ζ_1 matches the submerged part of the ship and is given by

$$\zeta_1(t, x) \ = \ \zeta_1(x - c_s t).$$

Therefore, at each time $t \geq 0$, the domains of the upper and lower fluid (denoted, respectively, Ω_1^t and Ω_2^t), are given by

$$\begin{aligned}\Omega_1^t &= \{\ (x,z) \in \mathbb{R}^d \times \mathbb{R}, \quad \zeta_2(t,x) \ \leq\ z \ \leq\ d_1 + \zeta_1(x - c_s t)\ \},\\ \Omega_2^t &= \{\ (x,z) \in \mathbb{R}^d \times \mathbb{R}, \quad -d_2 \ \leq\ z \ \leq\ \zeta_2(t,x)\ \}.\end{aligned}$$

The fluids being ideal, they satisfy the Euler equation (or Bernoulli equation when written in terms of the velocity potentials), and kinematic boundary conditions are given through the assumption that no fluid particle crosses the surface, the bottom or the interface. Finally, the set of equations is closed by the continuity of the stress tensor at the interface, which takes into account the effects of surface tension.

Altogether, the governing equations of our problem are given by the following

(2.1)
$$\begin{cases} \partial_x^2\phi_i \ +\ \partial_z^2\phi_i \ =\ 0 & \text{in } \Omega_i^t,\ i=1,2,\\ \partial_t\phi_i + \frac{1}{2}|\nabla_{x,z}\phi_i|^2 = -\frac{P}{\rho_i} - gz & \text{in } \Omega_i^t,\ i=1,2,\\ \partial_t\zeta_1 \ =\ \sqrt{1+|\partial_x\zeta_1|^2}\partial_{n_1}\phi_1 & \text{on } \Gamma_1 \equiv \{(x,z), z = d_1+\zeta_1(t,x)\},\\ \partial_t\zeta_2 = \sqrt{1+|\partial_x\zeta_2|^2}\partial_{n_2}\phi_1 = \sqrt{1+|\partial_x\zeta_2|^2}\partial_{n_2}\phi_2 & \text{on } \Gamma_2 \equiv \{(x,z), z = \zeta_2(t,x)\},\\ \partial_z\phi_2 \ =\ 0 & \text{on } \Gamma_b \equiv \{(x,z), z = -d_2\},\\ [\![P]\!] \ =\ \sigma\ \partial_x\left(\dfrac{\partial_x\zeta_2}{\sqrt{1+|\partial_x\zeta_2|^2}}\right) & \text{on } \Gamma_2, \end{cases}$$

where ∂_{n_i} is the upward normal derivative at Γ_i:

$$\partial_{n_i} \ \equiv\ n_i \cdot \nabla_{x,z}, \qquad \text{with}\ \ n_i \ \equiv\ \frac{1}{\sqrt{1+|\partial_x\zeta_i|^2}}(-\partial_x\zeta_i, 1)^T;$$

we denote $\sigma > 0$ the surface tension coefficient and $[\![P]\!]$ the jump of the pressure at the interface:

$$[\![P(t,x)]\!] \ \equiv\ \lim_{\varepsilon\to 0}\Big(\ P(t,x,\zeta_2(t,x)+\varepsilon)\ -\ P(t,x,\zeta_2(t,x)-\varepsilon)\ \Big).$$

This system can be reduced into evolution equations located at the surface and at the interface. Indeed, when we define the trace of the potentials at the surface and at the interface

$$\psi_1(t,x) \ \equiv\ \phi_1(t,x,d_1+\zeta_1(t,x)), \qquad \text{and} \qquad \psi_2(t,x) \ \equiv\ \phi_2(t,x,\zeta_2(t,x)),$$

then ϕ_1 and ϕ_2 are uniquely given by the Laplace equation, and the (Dirichlet or Neumann) boundary conditions (see Figure 2):

(2.2)
$$\begin{cases} (\ \partial_x^2 \ +\ \partial_z^2\ )\ \phi_2 = 0 & \text{in } \Omega_2,\\ \phi_2 = \psi_2 & \text{on } \Gamma_2,\\ \partial_z\phi_2 = 0 & \text{on } \Gamma_b, \end{cases} \quad \text{and} \quad \begin{cases} (\ \partial_x^2 \ +\ \partial_z^2\ )\ \phi_1 = 0 & \text{in } \Omega_1,\\ \phi_1 = \psi_1 & \text{on } \Gamma_1,\\ \partial_{n_2}\phi_1 = \partial_{n_2}\phi_2 & \text{on } \Gamma_2. \end{cases}$$

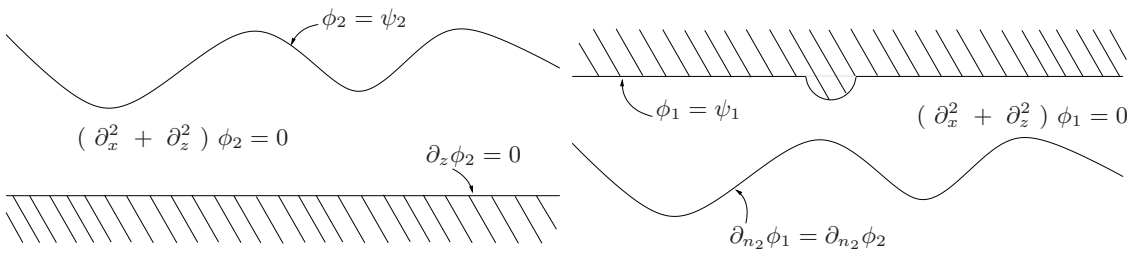


FIGURE 2. Laplace problems in the two domains.

Therefore, the following operators are well-defined:

$$\begin{aligned}
G_1[\zeta_1,\zeta_2](\psi_1,\psi_2) &\equiv \sqrt{1+|\partial_x\zeta_1|^2}(\partial_{n_1}\phi_1)\Big|_{z=1+\zeta_1}\,,\\
G_2[\zeta_2]\psi_2 &\equiv \sqrt{1+|\partial_x\zeta_2|^2}(\partial_{n_2}\phi_2)\Big|_{z=\zeta_2}\,,\\
H[\zeta_1,\zeta_2](\psi_1,\psi_2) &\equiv \partial_x\big(\phi_1\Big|_{z=\zeta_2}\big).
\end{aligned}$$

Then, the chain rule and straightforward combinations of the equations of (2.1) lead to the following equivalent system:

(2.3)
$$\begin{cases}
\partial_t\zeta_1 - G_1[\zeta_1,\zeta_2](\psi_1,\psi_2) = 0,\\
\partial_t\zeta_2 - G_2[\zeta_2]\psi_2 = 0,\\
\partial_t\Big(\rho_2\partial_x\psi_2-\rho_1H[\zeta_1,\zeta_2](\psi_1,\psi_2)\Big) + g(\rho_2-\rho_1)\partial_x\zeta_2\\
\qquad + \dfrac{1}{2}\partial_x\Big(\rho_2|\partial_x\psi_2|^2-\rho_1|H[\zeta_1,\zeta_2](\psi_1,\psi_2)|^2\Big) = \partial_x\mathcal{N} + \sigma\,\partial_x^2\Big(\dfrac{\partial_x\zeta_2}{\sqrt{1+|\partial_x\zeta_2|^2}}\Big),
\end{cases}$$

with

$$\mathcal{N} = \frac{\rho_2\big(G_2[\zeta_2]\psi_2+(\partial_x\zeta_2)(\partial_x\psi_2)\big)^2 - \rho_1\big(G_2[\zeta_2]\psi_2+(\partial_x\zeta_2)H[\zeta_1,\zeta_2](\psi_1,\psi_2)\big)^2}{2(1+|\partial_x\zeta_2|^2)}.$$

2.2. Nondimensionalization of the system. The next step consists in nondimensionalizing the system. The study of the linearized system (see **[70]**, for example), which can be solved explicitly, leads to a well-adapted rescaling. Moreover, it is convenient to write the equations in the frame of reference of the ship.

Let a_1 and a_2 be the maximum amplitude of the deformation of, respectively, the surface and the interface. We denote by λ a characteristic horizontal length, say the wavelength of the interface. Then the typical velocity of small propagating internal waves (or wave celerity) is given by

$$c_0 = \sqrt{g\frac{(\rho_2-\rho_1)d_1d_2}{\rho_2d_1+\rho_1d_2}}.$$

Consequently, we introduce the dimensionless variables

$$\tilde z \equiv \frac{z}{d_1}, \qquad \tilde x \equiv \frac{x+c_st}{\lambda}, \qquad \tilde t \equiv \frac{c_0}{\lambda}t,$$

the dimensionless unknowns

$$\tilde\zeta_i(\tilde x) \equiv \frac{\zeta_i(x)}{a_i}, \qquad \tilde\psi_i(\tilde x) \equiv \frac{d_1}{a_2\lambda c_0}\psi_i(x),$$

and the seven independent dimensionless parameters

$$\gamma = \frac{\rho_1}{\rho_2},\quad \epsilon_1 \equiv \frac{a_1}{d_1},\quad \epsilon_2 \equiv \frac{a_2}{d_1},\quad \mu \equiv \frac{d_1^2}{\lambda^2},\qquad \delta \equiv \frac{d_1}{d_2},\quad \mathrm{Fr} \equiv \frac{c_s}{c_0},\quad \mathrm{Bo} \equiv \frac{c_0^2\lambda^2\rho_2}{d_1\sigma}.$$

With this rescaling, the full Euler system (2.1) becomes (we have withdrawn the tildes for the sake of readability)

(2.4)
$$\left\{\begin{array}{l} -\epsilon_1 \operatorname{Fr} \partial_x\zeta_1 \;-\; \dfrac{\epsilon_2}{\mu} G_1(\psi_1,\psi_2) \;=\; 0, \\[2ex] (\partial_t - \operatorname{Fr}\partial_x)\zeta_2 \;-\; \dfrac{1}{\mu} G_2\psi_2 \;=\; 0, \\[2ex] (\partial_t - \operatorname{Fr}\partial_x)\Big(\partial_x\psi_2 - \gamma H(\psi_1,\psi_2)\Big) \;+\; (\gamma+\delta)\partial_x\zeta_2 \\ \qquad + \dfrac{\epsilon_2}{2}\partial_x\Big(|\partial_x\psi_2|^2 - \gamma|H(\psi_1,\psi_2)|^2\Big) \;=\; \mu\epsilon_2\partial_x\mathcal{N} \;+\; \dfrac{1}{\operatorname{Bo}}\partial_x^2\Big(\dfrac{\partial_x\zeta_2}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}}\Big), \end{array}\right.$$

with

$$\mathcal{N} \;\equiv\; \frac{\big(\frac{1}{\mu}G_2\psi_2 + \epsilon_2(\partial_x\zeta_2)(\partial_x\psi_2)\big)^2 \;-\; \gamma\big(\frac{1}{\mu}G_2\psi_2 + \epsilon_2(\partial_x\zeta_2)H(\psi_1,\psi_2)\big)^2}{2(1+\mu|\epsilon_2\partial_x\zeta_2|^2)},$$

and the dimensionless Dirichlet-to-Neumann operators defined by

$$G_1(\psi_1,\psi_2) \;\equiv\; G_1^{\mu,\delta}[\epsilon_1\zeta_1,\epsilon_2\zeta_2](\psi_1,\psi_2) \;\equiv\; -\mu\epsilon_1\big(\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big)(\partial_x\phi_1)\Big|_{z=1+\epsilon_1\zeta_1} + (\partial_z\phi_1)\Big|_{z=1+\epsilon_1\zeta_1},$$

$$G_2\psi_2 \;\equiv\; G_2^{\mu,\delta}[\epsilon_2\zeta_2]\psi_2 \equiv -\mu\epsilon_2(\partial_x\zeta_2)(\partial_x\phi_2)\Big|_{z=\epsilon_2\zeta_2} + (\partial_z\phi_2)\Big|_{z=\epsilon_2\zeta_2},$$

$$H(\psi_1,\psi_2) \;\equiv\; H^{\mu,\delta}[\epsilon_1\zeta_1,\epsilon_2\zeta_2](\psi_1,\psi_2) \;\equiv\; (\partial_x\phi_1)\Big|_{z=\epsilon_2\zeta_2} + \epsilon_2(\partial_x\zeta_2)(\partial_z\phi_1)\Big|_{z=\epsilon_2\zeta_2},$$

where ϕ_1 and ϕ_2 are the solutions of the Laplace problems

$$\left\{\begin{array}{ll} \big(\,\mu\partial_x^2 \;+\; \partial_z^2\,\big)\;\phi_2 = 0 & \text{in } \Omega_2 \equiv \{(x,z)\in\mathbb{R}^2, -\frac{1}{\delta} < z < \epsilon_2\zeta_2(x)\}, \\ \phi_2 = \psi_2 & \text{on } \Gamma_2 \equiv \{z = \epsilon_2\zeta_2\}, \\ \partial_z\phi_2 = 0 & \text{on } \Gamma_b \equiv \{z = -\frac{1}{\delta}\}, \end{array}\right.$$

$$\left\{\begin{array}{ll} \big(\,\mu\partial_x^2 \;+\; \partial_z^2\,\big)\;\phi_1 = 0 & \text{in } \Omega_1 \equiv \{(x,z)\in\mathbb{R}^2, \epsilon_2\zeta_2(x) < z < 1+\epsilon_1\zeta_1(x)\}, \\ \phi_1 = \psi_1 & \text{on } \Gamma_1 \equiv \{z = 1+\epsilon_1\zeta_1\}, \\ \partial_{n_2}\phi_1 = G_2\psi_2 & \text{on } \Gamma_2. \end{array}\right.$$

Remark 2.1. *Let us keep in mind that in our case, the function ζ_1 is not an unknown, but some fixed data of the problem:*

$$\zeta_1(t,x) \;=\; \zeta_1(x),$$

where ζ_1 is the submerged part of the ship (independent of time thanks to the change in the frame of reference). In that way, the first line of (2.4) *is a relation connecting ψ_1 with ψ_2, and* (2.4) *can be reduced into a system of two equations with two unknowns. This reduction is computed explicitly in the following asymptotic models.*

Remark 2.2. *The Cauchy problem associated to the full Euler system at the interface of two fluids is known to be ill-posed in Sobolev spaces in the absence of surface tension, as Kelvin-Helmholtz instabilities appear. However, in* **[70]**, *Lannes proved that adding a small amount of surface tension guarantees the well-posedness of the Cauchy problem (with a flat rigid lid), with a time of existence that may remain quite large even with a small surface tension coefficient, and thus is consistent with the observations. The idea behind this result is that the Kelvin-Helmholtz instabilities appear for high frequencies, where the regularization effect of the surface tension is relevant, even if* Bo *the Bond number measuring the ratio of gravity forces over capillary forces is very large.*[1] *On the contrary,*

[1]As an example, let us consider the values of the experiments of Vasseur, Mercier and Dauxois **[114]**. One has $d_1 = 0.05\text{m}$, $d_2 = 0.12\text{m}$, $\rho_1 = 1\,000.5\text{kg.m}^{-3}$, $\rho_2 = 1\,022.7\text{kg.m}^{-3}$. Taking advantage of the analysis of Lannes **[70]**, we use $\sigma = 0.095\text{N.m}^{-1}$ as a typical value for the interfacial tension coefficient. As a

the main profile of the wave that we want to capture is located at lower frequencies, and will therefore be unaffected by surface tension. Driven by this analysis, we decide to omit the surface tension term in the models of the Regime 2 presented below, as they do not give rise to Kelvin-Helmholtz instabilities. On the contrary, we keep the surface tension term in the fully nonlinear models of Regime 1. As a matter of fact the regularization effect of the surface tension in this case plays a critical role in our numerical simulations, as it stabilizes the scheme, even if Bo^{-1} *is one order smaller than the coefficients in front of all other terms.*

The following terminology is used throughout the paper.

Definition 2.3 (Adapted solutions). *We call* adapted *solution of* (2.4) *any strong solution* $(\zeta_1, \zeta_2, \psi_1, \psi_2)$*, bounded in* $W^{1,\infty}([0,T]; H^{s+t_0})$ *with* $T > 0$*,* $s > 1$ *and* t_0 *big enough, and such that* $\zeta_1(t,x) = \zeta_1(x)$ *(in the frame of reference of the ship) and the domains of the fluids remain strictly connected,* i.e. *there exists* $h_{\min} > 0$ *such that*

$$1 + \epsilon_1\zeta_1(x) - \epsilon_2\zeta_2(t,x) \geq h_{\min} > 0 \quad \textit{and} \quad \frac{1}{\delta} + \epsilon_2\zeta_2(t,x) \geq h_{\min} > 0. \tag{2.5}$$

From the discussion above, it is legitimate to assume that such a smooth, uniformly bounded family of solutions of (2.4) indeed exists.

2.3. Description of the results, and the regimes under study. The models displayed in the following sections are all justified with a consistency result; see definition below. When it is possible, that is when energy estimates are attainable, we provide a stronger justification with a convergence rate (the convergence theorems are precisely disclosed in Proposition 4.1 and 4.2).

Definition 2.4 (Consistency). *The full Euler system* (2.4) *is* consistent *with a system of two equations* (Σ) *on* $[0,T]$ *if any* adapted solution U *defines,* via *the changes of variables explained throughout the paper, a pair of functions satisfying* (Σ) *up to a small residual, called the* precision *of the model. The order of the precision will be* $\mathcal{O}(\varepsilon)$*, if there exists* $s_0, t_0 \geq 0$ *such that if* $U \in W^{1,\infty}([0,T]; H^{s+t_0})$ *with* $s > s_0$*, then the* $L^\infty([0,T]; H^s)$ *norm of the residual is bounded by* $C_0\ \varepsilon$*, with* C_0 *a constant independent of* ε*.*

Definition 2.5 (Convergence). *The full Euler system* (2.4) *and a well-posed system* (Σ) *of two equations are* convergent *at order* $\mathcal{O}(\varepsilon)$ *on* $[0,T]$ *if any adapted solution with small initial data* U *defines,* via *the changes of variables explained throughout the paper, a pair of functions* V *such that* $\tilde{V}$ *the solution of* (Σ) *with same initial data satisfies*

$$\left|V - \tilde{V}\right|_{L^\infty([0,T];H^s)} \ \leq\ C_0\ \varepsilon,$$

with C_0 *independent of* ε*.*

The small parameter ε in these definitions is a function of some of the dimensionless parameters of the system that are assumed to be small. The regimes that we study throughout the paper have been briefly presented in the introduction; let us describe them here in more details. When nothing is specifically said in the regimes below, we assume the parameters to be fixed as

$$\gamma \in (0,1), \quad \epsilon_1, \epsilon_2 \in (0,1), \quad 0 < \delta_{\min} \leq \delta \leq \delta_{\max}, \quad \mathrm{Fr} \in [0, \mathrm{Fr}_{\max}], \quad 0 < \mathrm{Bo}_{\min} \leq \mathrm{Bo}\,.$$

In particular, the two layers are assumed to be of finite, comparable depth.

conclusion, one has $\mathrm{Bo} = \frac{c_0^2\rho_2 d_1}{\sigma}\frac{1}{\mu} \approx \frac{40}{\mu}$. The coefficient in front of the surface tension term is therefore much smaller than the coefficients in front of any other term.

The mutual and crucial assumption between the two regimes is that the depth of the two layers of fluids is small when compared with the internal wavelength. This is commonly called *shallow water regime* and is simply given by (with our notation)

$$\mu \ll 1.$$

The shallow water regime has been widely used in the framework of gravity waves. In the one layer, free surface case, it leads at order $\mathcal{O}(\mu)$ to the shallow water equations [**37**], and at order $\mathcal{O}(\mu^2)$ to the Green-Naghdi equations[**48**]. The analysis has been extended to the case of two layers with a free surface in [**40**], and the models presented here can be recovered from models in this situation, when fixing the surface as some data of the problem.

Regime 1 (Small sized boats).

$$\mu \ll 1 \ ; \qquad \alpha \equiv \frac{\epsilon_1}{\epsilon_2} \ = \ \mathcal{O}(\mu) \ , \ 1-\gamma \ = \ \mathcal{O}(\mu).$$

The two additional smallness assumptions of this regime are very natural. First, we assume that the depth of the submerged part of the ship is small when compared with the depth of the fluid, and the attainable size of the deformation. The numerical simulations of our model show that even with this assumption, the waves generated by the small disturbance can be very large (of the order of the depth of the two layers, and therefore very large when compared with the variation of the surface induced by the ship). This explains why the dead-water phenomenon can be so powerful. What is more, we assume that the densities of the two fluids are almost equal. This is known as the Boussinesq approximation, and is valid most of the time (the value of $1-\gamma$ reported in Celtic Sea in [**96**] is $\approx 10^{-3}$, and the value in the experiments of [**114**] is $\approx 10^{-2}$).

Regime 2 (Small wave amplitudes).

$$\mu \ll 1 \ ; \qquad \epsilon_2 \ = \ \mathcal{O}(\mu) \ , \ \alpha \equiv \frac{\epsilon_1}{\epsilon_2} \ = \ \mathcal{O}(\mu).$$

In this regime, we assume that the internal wave generated by the disturbance will remain small when compared with the depth of the two fluids. As previously, we assume that the waves generated by the ship are large when compared with the depth of the disturbance, so that the ship suffers from a relatively large wave resistance. In that way, the models we obtain involve only weak nonlinearities, and will be much easier to study. In particular, we are able to obtain well-posedness and convergence results for both of the models presented here. The counterpart is that these models remain valid only for relatively small waves. Again, this regime has been widely used in the literature. The one layer, free surface equivalent systems of the models presented here are the classical Boussinesq system [**19, 20**] and the KdV approximation [**62**]. These models have been rigorously justified in [**104, 16, 1**], and the extension to the case of two layers, with a free surface, can be found in [**41**].

2.4. Expansion of the Dirichlet-Neumann operators. The key assumption in our models, which is shared by both regimes 1 and 2, is the so-called shallow water assumption

$$\mu \ll 1,$$

which states that the depth of the two layers of fluids are small when compared with the characteristic wavelength of the system. The main ingredient in the construction of asymptotic models in the shallow water regime, then lies in the construction and the justification of the following expansion of the operators G_1, G_2 and H:

Lemma 2.6. *Let $\zeta_1, \zeta_2, \psi_1, \psi_2$, such that $(\zeta_1, \zeta_2, \partial_x\psi_1, \partial_x\psi_2) \in W^{1,\infty}([0,T]; H^{s+t_0}(\mathbb{R}))$, with $s > 1$ and $t_0 \geq 9/2$, and such that (2.5) is satisfied. Then one has*

$$\begin{aligned}
\Big| G_1(\psi_1,\psi_2) + \mu\partial_x(h_1\partial_x\psi_1 + h_2\partial_x\psi_2) - \mu^2\partial_x\Big(\mathcal{T}[h_1,h_2]\partial_x\psi_1 + \mathcal{T}[h_2,0]\partial_x\psi_2 \\
-\frac{1}{2}(h_1^2\partial_x^2(h_2\partial_x\psi_2)) - (h_1\epsilon_1\frac{d}{dx}\zeta_1\partial_x(h_2\partial_x\psi_2))\Big)\Big|_{L^\infty H^s} &\leq \mu^3 C_0, \\
\Big| G_2\psi_2 + \mu\partial_x(h_2\partial_x\psi_2) - \mu^2\partial_x\mathcal{T}[h_2,0]\partial_x\psi_2\Big|_{L^\infty H^s} &\leq \mu^3 C_0, \\
\Big| H(\psi_1,\psi_2) - \partial_x\psi_1 - \mu\partial_x\Big(h_1(\partial_x(h_1\partial_x\psi_1) + \partial_x(h_2\partial_x\psi_2)) \\
-\frac{1}{2}h_1^2\partial_x^2\psi_1 - h_1\epsilon_1(\partial_x\psi_1)\frac{d}{dx}\zeta_1(x)\Big)\Big|_{W^{1,\infty}H^s} &\leq \mu^2 C_0,
\end{aligned}$$

with $C_0 = C_0\left(\frac{1}{h_{\min}}, |U|_{W^{1,\infty}H^{s+t_0}}\right)$, and the operator $\mathcal{T}$ defined by

$$\mathcal{T}[h,b]V \equiv -\frac{1}{3}\partial_x(h^3\partial_x V) + \frac{1}{2}\big(\partial_x(h^2(\partial_x b)V)) - h^2(\partial_x b)(\partial_x V)\big) + h(\partial_x b)^2 V.$$

These estimates have been proved for $L^\infty([0,T]; H^s)$ norms in Propositions 2.2, 2.5 and 2.7 of [**40**]. The proof can easily be adapted to work with the time derivative of the functions, following the proof of [**40**, Proposition 2.12].

The idea is then simply to plug these expansions into the full Euler system (2.4), and drop all $\mathcal{O}(\mu^2)$ terms. Then, using the fact that ζ_1 is a forced parameter of our problem, the system reduces to two evolution equations for (ζ_2, v), with v the *shear velocity* defined by

$$v \equiv \partial_x\Big(\big(\phi_2 - \gamma\phi_1\big)\Big|_{z=\epsilon_2\zeta_2}\Big) = \partial_x\psi_2 - \gamma H(\psi_1,\psi_2). \tag{2.6}$$

These calculations are postponed to Appendix A for the sake of readability, and we directly present here the system thus obtained:

$$\left\{\begin{array}{l}
(\partial_t - \mathrm{Fr}\,\partial_x)\zeta_2 + \partial_x\left(\dfrac{h_2}{h_1+\gamma h_2}(h_1 v + \gamma\alpha\,\mathrm{Fr}\,\zeta)\right) + \mu\partial_x\big(\mathcal{L}(v_1,v_2)\big) = 0, \\[2ex]
(\partial_t - \mathrm{Fr}\,\partial_x)v + (\gamma+\delta)\partial_x\zeta_2 + \dfrac{\epsilon_2}{2}\partial_x\left(\dfrac{|h_1 v + \gamma\alpha\,\mathrm{Fr}\,\zeta|^2 - \gamma|h_2 v - \alpha\,\mathrm{Fr}\,\zeta|^2}{(h_1+\gamma h_2)^2}\right) \\[2ex]
\qquad\qquad +\mu\epsilon_2\partial_x\big(\mathcal{Q}[v_1,v_2]\big) = \dfrac{1}{\mathrm{Bo}}\partial_x^2\left(\dfrac{\partial_x\zeta_2}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}}\right),
\end{array}\right. \tag{2.7}$$

where $\mathcal{L}$ and $\mathcal{Q}$ are respectively linear and quadratic in (v_1, v_2), the latter being the approximation at order $\mathcal{O}(\mu)$ of $(\partial_x\psi_1, \partial_x\psi_2)$ given by

$$v_1 \equiv -\frac{h_2 v - \alpha\,\mathrm{Fr}\,\zeta_1}{h_1+\gamma h_2}, \quad \text{and} \quad v_2 \equiv \frac{h_1 v + \gamma\alpha\,\mathrm{Fr}\,\zeta_1}{h_1+\gamma h_2}. \tag{2.8}$$

The operators $\mathcal{L}$ and $\mathcal{Q}$ are precisely disclosed in Appendix A. The full Euler system (2.4) is consistent with this system at order $\mathcal{O}(\mu^2)$ on $[0,T]$, $T > 0$, as stated in Proposition A.1, page 155.

The simplified models we study are deduced from system (2.7), using the additional assumptions of Regime 1 and Regime 2. The full Euler system (2.4) is still consistent at order $\mathcal{O}(\mu^2)$ with the models thus obtained, that is (see below) the strongly nonlinear systems (3.1) and (3.3), the Boussinesq-type system (4.2) and its symmetrized version (4.3). The KdV approximation (4.12) is then deduced from the symmetric Boussinesq-type system.

3. Strongly nonlinear models

In this section, we introduce different strongly nonlinear asymptotic models for the full Euler system (2.4). These models are obtained starting from system (2.7), using the additional properties of Regime 1. The weakly nonlinear models (Regime 2) are presented in following Section 4. Indeed, the system (2.7) only presumes some smallness on the shallowness parameter μ, that is to say the depth of the fluids are small when compared with the characteristic wavelength of the system. In the framework of Regime 1, we assume that

$$\mu \ll 1 ; \qquad \alpha \equiv \frac{\epsilon_1}{\epsilon_2} = \mathcal{O}(\mu) , \ 1-\gamma = \mathcal{O}(\mu),$$

where we recall that γ is the density ratio and α the ratio between the amplitudes of the deformations at the surface and at the interface. As it has been said, page 135, these assumptions are very natural in the framework of our study. The first one supposes that the deformation induced by the presence of the ship at the surface is small when compared with the depth of the two layers of fluid, and small when compared with the attainable size of the wave generated at the interface. The second one is the classical Boussinesq approximation. The simplified systems we obtain remain consistent at order $\mathcal{O}(\mu^2)$, as stated in Proposition 3.1. In Section 3.2, we present and discuss numerical computations from the constructed models, that reproduce the key aspects of the dead-water phenomenon.

3.1. The fully nonlinear model in Regime 1. The first obvious observation is that the total depth of the fluid is approximatively constant in Regime 1, *i.e.* equal to $h \equiv 1+\frac{1}{\delta}$ at order $\mathcal{O}(\mu)$. What is more, one has

$$h_1+\gamma h_2 = h + \mathcal{O}(\mu) \equiv 1+\frac{1}{\delta} + \mathcal{O}(\mu).$$

Therefore, the approximations of $(\partial_x\psi_1, \partial_x\psi_2)$ at order $\mathcal{O}(\mu)$ given in (2.8) are now simply

$$h_2 v = -h\ v_1 + \mathcal{O}(\mu), \qquad h_1 v = h\ v_2 + \mathcal{O}(\mu).$$

It follows then some substantial simplifications, and one obtains in the end the system

$$\left\{\begin{array}{l} (\partial_t - \mathrm{Fr}\,\partial_x)\zeta_2 + \partial_x\left(\dfrac{h_1h_2}{h_1+\gamma h_2}v + h_2\dfrac{\alpha\,\mathrm{Fr}}{h}\zeta_1\right) + \mu\partial_x\big(\mathcal{P}_1 v\big) = 0, \\[2ex] (\partial_t - \mathrm{Fr}\,\partial_x)v + (\gamma+\delta)\partial_x\zeta_2 + \epsilon_2\partial_x\left(\dfrac{1}{2}\dfrac{|h_1v|^2-\gamma|h_2v|^2}{(h_1+\gamma h_2)^2} + \dfrac{\alpha\,\mathrm{Fr}}{h}\zeta_1 v\right) \\[2ex] \qquad\qquad + \mu\epsilon_2\partial_x\big(\mathcal{P}_2[v]\big) = \dfrac{1}{\mathrm{Bo}}\partial_x^2\left(\dfrac{\partial_x\zeta_2}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}}\right), \end{array}\right. \tag{3.1}$$

where $h \equiv 1+\frac{1}{\delta}$ and the operators $\mathcal{P}_1$ and $\mathcal{P}_2$ are defined by

$$\begin{aligned} h^2\ \mathcal{P}_1 v &= \frac{1}{3}\left(h_2\partial_x(h_1^3\partial_x(h_2v)) + h_1\partial_x(h_2^3\partial_x(h_1v))\right), \\ h^2\ \mathcal{P}_2[v] &= \frac{v}{3}\partial_x\left[h_1^3\partial_x(h_2v) - h_2^3\partial_x(h_1v)\right] + \frac{1}{2}\left((h_1\partial_x(h_2v))^2 - (h_2\partial_x(h_1v))^2\right). \end{aligned}$$

Linearizing the system (3.1) around the rest state, leads to

$$\left\{\begin{array}{l} (\partial_t - \mathrm{Fr}\,\partial_x)\zeta_2 + \dfrac{1}{\gamma+\delta}\partial_x v + \mu\dfrac{1}{3\delta(1+\delta)}\partial_x^3 v = \dfrac{\alpha\,\mathrm{Fr}}{1+\delta}\partial_x\zeta_1, \\[2ex] (\partial_t - \mathrm{Fr}\,\partial_x)v + (\gamma+\delta)\partial_x\zeta_2 = \dfrac{1}{\mathrm{Bo}}\partial_x^3\zeta_2, \end{array}\right.$$

from which we can easily deduce dispersion relation of the system, without forcing. Indeed, setting $\alpha \equiv 0$, the wave frequency $\omega(k)$, corresponding to plane-wave solutions $e^{ik\cdot X - i\omega t}$, is solution of the quadratic equation

$$(\mathrm{Fr}\, k \ + \ \omega)^2 \ = \ k^2 \Big(\ 1 \ + \ \frac{k^2}{\mathrm{Bo}(\gamma+\delta)} \ \Big)\Big(\ 1 \ - \ \mu k^2 \frac{\delta+\gamma}{3\delta(1+\delta)} \ \Big). \tag{3.2}$$

Therefore, for high wave numbers, k, there is no real-valued solution $\omega(k)$ of (3.2), which means that the system (3.1) is linearly ill-posed, and leads to instabilities. In order to deal with this issue, we use a nonlinear change of variables on the shear velocity v, that leads to a formally equivalent system. This idea is not new (see **[14, 93, 29, 40]** for example), and usually relies on a "shear velocity" constructed from horizontal velocities *at any depth* in the upper and lower layers. The choice of the depth allows then to control properties of the linear relation dispersion. Here, we define w with

$$\frac{h_1 h_2}{h_1+\gamma h_2} w \ \equiv \ \frac{h_1 h_2}{h_1+\gamma h_2} v \ + \ \mu \mathcal{P}_1 v \implies v \ = \ w \ - \ \mu \ \frac{h \mathcal{P}_1 w}{(h-h_2)h_2} \ + \mathcal{O}(\mu^2).$$

This leads immediately to the following system, equivalent at order $\mathcal{O}(\mu^2)$:

$$\left\{ \begin{array}{l} (\partial_t - \mathrm{Fr}\,\partial_x)\zeta_2 \ + \ \partial_x \left(\dfrac{h_1 h_2}{h_1+\gamma h_2} w + h_2 \dfrac{\alpha\,\mathrm{Fr}}{h} \zeta_1 \right) \ = 0, \\ \\ (\partial_t - \mathrm{Fr}\,\partial_x)\left(w - \mu \mathcal{S}_1 w\right) \ + \ (\gamma+\delta)\partial_x \zeta_2 \ + \ \epsilon_2 \partial_x \left(\dfrac{1}{2} \dfrac{h_1^2 - \gamma h_2^2}{(h_1+\gamma h_2)^2} w^2 + w \dfrac{\alpha\,\mathrm{Fr}}{h} \zeta_1 \right) \\ \\ \qquad\qquad + \ \mu \epsilon_2 \partial_x \left(w \ \mathcal{S}_2 w \right) \ = \ \dfrac{1}{\mathrm{Bo}} \partial_x^2 \left(\dfrac{\partial_x \zeta_2}{\sqrt{1+\mu\epsilon_2^2 |\partial_x \zeta_2|^2}} \right), \end{array} \right. \tag{3.3}$$

with $h \ \equiv \ 1 + \frac{1}{\delta}$, $h_1 \ \equiv \ 1 + \epsilon_1 \zeta_1 - \epsilon_2 \zeta_2$, $h_2 \ \equiv \ \frac{1}{\delta} + \epsilon_2 \zeta_2$, and the operators

$$\begin{aligned} \mathcal{S}_1 w \equiv & \ \frac{1}{3} \left(\partial_x^2 \big(h_1 h_2 w \big) - (\partial_x h_2)^2 w \right), \\ \mathcal{S}_2 w \equiv & \ \frac{h_1 h_2}{3h} \left((\partial_x^2 h_2) w + 2 (\partial_x w)(\partial_x h_2) \right) + \frac{h_1 - h_2}{2h} (\partial_x h_2)^2 w. \end{aligned}$$

Under this form, our system corresponds to the Green-Naghdi model presented in the one-layer case in **[1, 2]**, and proved to be well-posed and convergent, in the sense that their solutions provide an approximation at order $\mathcal{O}(\mu^2)$ in $L^\infty([0,T];H^s)$ to the solutions of the one layer water wave equations. When dropping all $\mathcal{O}(\mu)$ terms, one obtains a shallow water model, similar to the ones derived in **[31]** and **[35]** in the flat rigid lid case (the forcing terms induced by the presence of the body do not appear). Such a system has been studied in details and analyzed as a hyperbolic system (leading to well-posedness results under reasonable assumptions on the initial data) in **[163, 51, 21]**.

One can check that the issue of linear ill-posedness is now solved. Indeed, linearizing the system (3.3) around the rest state, leads to

$$\left\{ \begin{array}{l} (\partial_t - \mathrm{Fr}\,\partial_x)\zeta_2 \ + \ \dfrac{1}{\gamma+\delta} \partial_x w \ = \ \dfrac{\alpha\,\mathrm{Fr}}{1+\delta} \partial_x \zeta_1, \\ (\partial_t - \mathrm{Fr}\,\partial_x)(1 - \dfrac{\mu}{3\delta}\partial_x^2) w \ + \ (\gamma+\delta)\partial_x \zeta_2 \ = \ \dfrac{1}{\mathrm{Bo}} \partial_x^3 \zeta_2. \end{array} \right.$$

The dispersion relation of this linear system, when setting $\alpha \equiv 0$, is

$$(\mathrm{Fr}\, k \ + \ \omega)^2 \Big(1 \ + \ \frac{\mu}{3\delta} k^2 \Big) \ = \ k^2 \Big(1 \ + \ \frac{1}{(\gamma+\delta)\,\mathrm{Bo}} k^2 \Big), \tag{3.4}$$

which leads to real-valued waves frequencies ω, for any values of $k \in \mathbb{R}$.

Let us remark that under the assumptions of Regime 1, the relations (3.2) and (3.4) are asymptotically equivalent at order $\mathcal{O}(\mu^2 k^4)$, so that the effect of the nonlinear change of variable only affects high frequencies.

We now state that the two systems (3.1) and (3.3) are equivalently justified as models for the full Euler system, with the following consistency result.

PROPOSITION 3.1. *Assuming that $\alpha = \mathcal{O}(\mu)$ and $1-\gamma = \mathcal{O}(\mu)$, the full Euler system* (2.4) *is consistent with the models* (3.1) *and* (3.3)*, both at precision $\mathcal{O}(\mu^2)$ on $[0,T]$, with $T>0$.*

PROOF. Let $U \equiv (\zeta_1, \zeta_2, \psi_1, \psi_2)$ be a solution of (2.4), bounded in $W^{1,\infty}([0,T]; H^{s+t_0})$ with $s>1$ and $t_0 \geq 9/2$, and such that (2.5) is satisfied with $\zeta_1(t,x) \equiv \zeta_1(x)$. The consistency result of Proposition A.1 states that (ζ_2, v), with $v \equiv \partial_x\psi_2 - \gamma H(\psi_1,\psi_2)$, satisfies (2.7), up to $R_1 = (r_1, r_2)^T \in L^\infty([0,T]; H^s)^2$, with (for $i=1,2$)

$$\left|r_i\right|_{L^\infty H^s} \leq \mu^2\, C_0\left(\frac{1}{h_{\min}}, \left|U\right|_{W^{1,\infty}H^{s+t_0}}\right).$$

Since $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{T}$ and $\partial_x\mathcal{H}$ (defined in Appendix A) involve two spatial derivatives, and thanks to straightforward calculations using the smallness assumptions of regime 1, one has for $i=1,2$

$$\left|\mathcal{L}_i - \mathcal{P}_i\right|_{L^\infty([0,t);H^s)} \leq \mu^2\, C_0\left(\frac{1}{h_{\min}}, \left|(h_1,h_2,v)\right|_{L^\infty H^{s+2}}\right) \leq \mu^2\, C_0\left(\frac{1}{h_{\min}}, \left|U\right|_{W^{1,\infty}H^{s+t_0}}\right).$$

The consistency of (2.4) with (3.1) is therefore proved.

Now, we set $w \equiv v + \mu\frac{h_1+\gamma h_2}{h_1 h_2}\mathcal{P}_1[h_2]v$, and one has immediately, using the fact that $H^s(\mathbb{R})$ is an algebra for $s>1/2$,

$$\begin{aligned}\left|v - w + \mu\,\frac{h\mathcal{P}_1[h_2]w}{(h-h_2)h_2}\right|_{W^{1,\infty}H^s} &\leq \mu^2\, C_0\left(\frac{1}{h_{\min}}, \left|(h_1,h_2,v)\right|_{W^{1,\infty}H^{s+2}}\right)\\ &\leq \mu^2\, C_0\left(\frac{1}{h_{\min}}, \left|U\right|_{W^{1,\infty}H^{s+t_0}}\right).\end{aligned}$$

It is then straightforward to deduce from the previous consistency result, the consistency of (2.4) with (3.3). □

3.2. Numerical simulations. In figures 3 and 4, we plot the behavior of the flow predicted by the model (3.3), for different values of the parameters. Each figure contains three panels. The left panel represents the interface deformation, depending on space ($x \in [-20,20]$) and time ($t \in [0,15]$) variables. The right panel contains the time evolution of the wave resistance coefficient, computed thanks to formula (C.4). Finally, we plot in the bottom panel the situation of the system (*i.e.* the surface and interface deformations) at final time $t=15$.

In these simulations, and throughout the paper, we use zero initial data conditions, and a function $\zeta_1(x)$ defined by

$$\zeta_1(x) \equiv \begin{cases} -\exp(-\frac{x^2}{(1-x)(x+1)}) & \text{if } x\in(-1,1),\\ 0 & \text{otherwise.}\end{cases}$$

The parameters of the system ϵ_2, μ, $\alpha \equiv \epsilon_1/\epsilon_2$, δ, γ, Fr and Bo, are specified below each of the figures. The schemes used are described and justified in Appendix D.

Here, we decide to study the effect of the depth ratio coefficient δ. We choose two different values for the depth ratio: $\delta = 5/12$ (corresponding to the experiments of [**114**]), or $\delta = 12/5$ for a thicker upper layer. We plot in figures 3 and 4 the outcome of our

scheme (D.4), setting the Froude number as Fr = 1. Away from this critical value, as predicted by experiments [**43, 114**] and confirmed by our numerical simulations, the amplitude of the generated waves (and therefore the magnitude of the wave resistance coefficient) are significantly smaller. In that way, the outcome is similar to the one predicted by the weakly nonlinear models, and we do not present these results for the sake of brevity. We let the reader refer to Section 4.3.1 for a discussion on the case of subcritical and supercritical Froude values, in Regime 2.

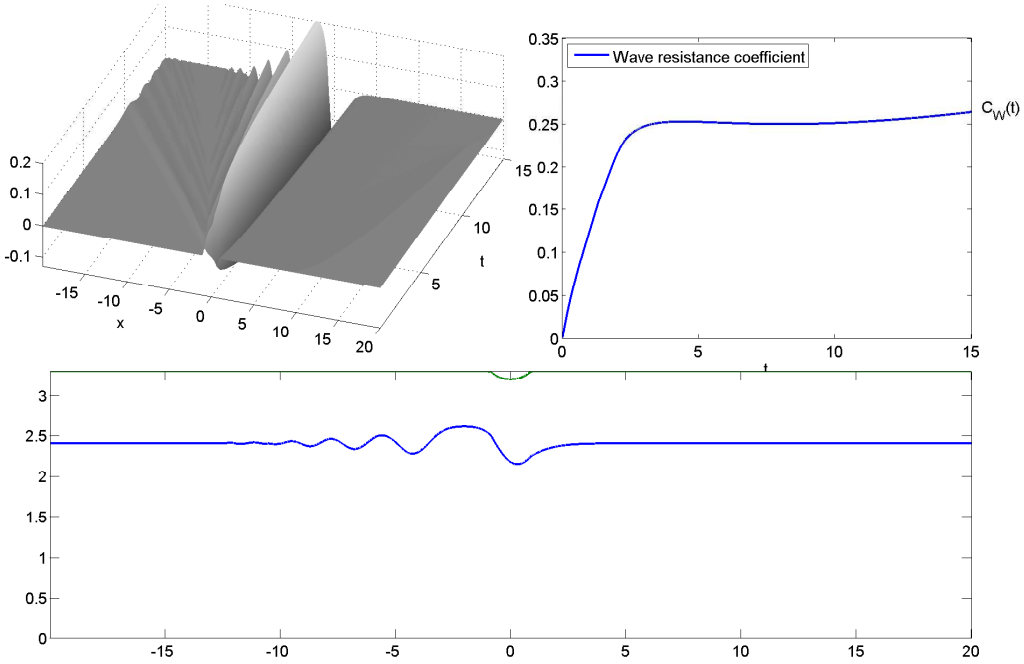


Figure 3. Flow predicted by (3.3), with steady initial data and a thicker lower layer.
$\epsilon_2 = 1$, $\alpha = \mu = 0.1$, $\gamma = 0.99$, Fr = 1, Bo = 100, $\delta = 5/12$.

Let us express a few remarks on these results. There are some similarities for each of the situations. First, one sees that there is no deformation on the right-hand side of the plots, which corresponds to the up-stream propagating part. This is due to the fact that we have set Fr = 1, which corresponds to the maximum gravitational wave velocity in the flat rigid lid case. On the contrary, on the left-hand side (down-stream propagating part), one remarks a small elevation wave, followed by even smaller perturbations, progressing with velocity $c_- = -2$ (in the frame of the ship). This corresponds to the η_- part of the KdV approximation decomposition, and is studied with more details in Section 4.3. Finally, the most important part takes place just behind the location of the body. Each time, an important wave of elevation is generated just behind the body, producing a severe wave resistance (see Remark C.1 page 161). This wave comes with a tail of smaller waves, that are located away from the boat, and therefore do not produce any drag.

However, one can see that the shape and time-behavior of this elevation wave is quite different, depending on the value of the depth ratio δ. When the upper fluid domain is thinner than the lower fluid's, the generated wave is flattening as it is growing up. Its height is relatively low, but the deformation carries on with a depression wave, located just below the body. On the contrary, when the upper fluid domain is the thicker one, then the generated wave remains sharp and is continuously growing. It is separated with its tail, while all of the energy produced by the ship contributes to the elevation of the wave. The wave produced wave resistance is greater in this case, and the dead-water effect appears

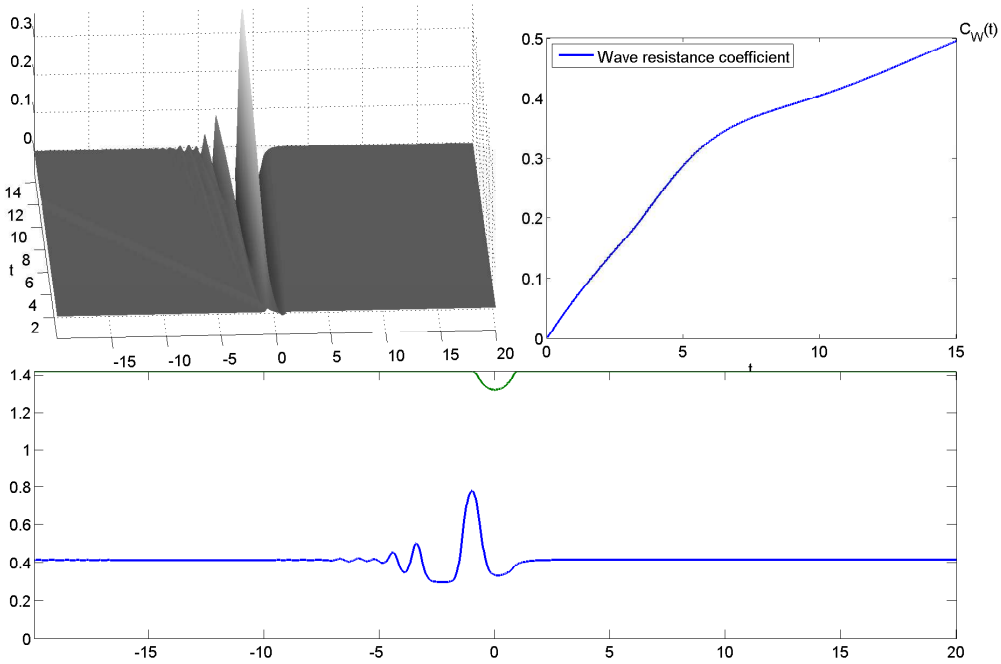


FIGURE 4. Flow predicted by (3.3), with steady initial data and a thicker upper layer.
$\epsilon_2 = 1$, $\alpha = \mu = 0.1$, $\gamma = 0.99$, Fr $= 1$, Bo $= 100$, $\delta = 12/5$.

therefore much stronger. The fact that the wave resistance is stronger when the upper fluid domain is thicker is counterintuitive, as the distance between the ship and the interface is larger. However, as we can see in Appendix (C), the wave resistance coefficient does not depend on the distance between the two interfaces $(\zeta_1 - \zeta_2)$, but rather on the amplitude of the generated wave. It is therefore not surprising that, when the upper layer is thicker, the generated elevation wave is able to reach larger amplitudes, and the drag suffered by the body is stronger.

The experiments conducted in **[114]** (this has also been reported in **[160]** for example) exhibit an interesting behavior, that cannot be seen in our simulations in figures 3 and 4. Indeed, with their setting, the dead-water phenomenon appears to be periodic in some sense, as the generated wave increases its amplitude, slows down the boat, breaks, and the process repeats. As a matter of fact, even when running simulations for much longer (around ten times) time intervals, the solution of (3.3) does not display such a phenomenon. It is interesting to see that this is a major difference with the solutions of the KdV approximation, which generate periodically up-stream propagating waves, inducing an oscillation in the related wave resistance (see Figure 9 page 152). However, even with this model, the time-period of generation of these solitons ($\Delta T \approx 100$ in our situation) is too large to explain positively the phenomenon.

This discrepancy may be due to our assumption of a constant velocity for the body, as the setting in **[114]** constrains a constant force to move the boat. As a numerical experiment, we performed simulations with a Froude number Fr adjusted at each step:

$$\mathrm{Fr}((n+1)\Delta t) \equiv \mathrm{Fr}(n\Delta t) - \Delta t \mathrm{Cstt}_1(C_W(n\Delta t) - \mathrm{Cstt}_2).$$

This roughly corresponds to a case where the acceleration of the ship is given by a constant force imposed to the ship, minus the resistance suffered by the body.[2] We present in

[2]Of course, this modification is careless, and the produced scheme reflects only roughly what would be a model where a constant force is imposed to the body. To construct and rigorously justify a constant-force

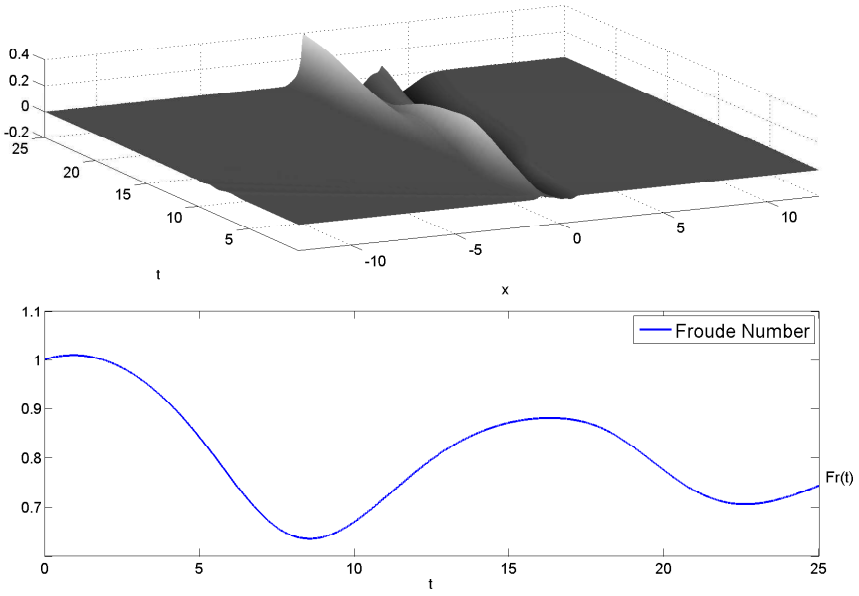


Figure 5. Flow predicted by (3.3), when the velocity of the body Fr is not constant.
$\epsilon_2 = 1$, $\delta = 1$, $\alpha = \mu = 0.1$, $\gamma = 0.99$, $\mathrm{Bo} = 100$.

Figure 5 the result of this computation. A periodicity can clearly be seen in the wave resistance (with a time period of order $\Delta T \approx 10$). One can explain the phenomenon as follows. The velocity of the ship, suffering from the drag generated by the wave resistance, decreases down to a value where the generated wave resistance is very small. Released from its drag, the ship speeds up, and the phenomenon repeats periodically. This corresponds to the observations of [**114**], and is not recovered by constant velocity models (see [**88, 91, 120**]).

4. Weakly nonlinear models

The aim of this section is to introduce simple weakly nonlinear models for our problem, using the smallness assumptions of Regime 2:

$$\mu \ll 1 \; ; \qquad \epsilon_2 \; = \; \mathcal{O}(\mu), \quad \alpha \; = \; \mathcal{O}(\mu).$$

We recall that μ stands for the shallowness parameter, that is the ratio between the depth of the upper fluid layer and the internal wavelength. The parameter ϵ_2 measures the magnitude of the deformation at the interface when compared with the upper fluid layer, and α is the ratio between the amplitudes of the deformations at the surface and at the interface. As discussed in Remark 2.2, the effect of the surface tension term will be relevant only if Kelvin-Helmholtz instabilities, located at high frequencies, appear. The models we obtain in this section do not give rise to these instabilities, and we therefore neglect the surface tension term, and set $\mathrm{Bo}^{-1} \equiv 0$ ($\mathrm{Bo}^{-1} = \mathcal{O}(\mu^2)$ would in fact suffice).

Starting with the strongly nonlinear model (2.7), we deduce easily a Boussinesq-type model and its symmetrized version. Then, using a classical WKB (Wentzel-Kramers-Brillouin) expansion, we obtain a rougher approximation, that consists in two *uncoupled Korteweg-de Vries equation, with a forcing term.* These two equations are studied in details, and numerically computed, in Section 4.3. The symmetric Boussinesq-type model, as well

model would be much more difficult, as we explicitly use the constant-velocity hypothesis, as early as in (A.4).

as the KdV approximation, are justified thanks to convergence results (see Propositions 4.1 and 4.2, respectively).

4.1. The Boussinesq-type models. Let us drop $\mathcal{O}(\mu^2)$ terms in system (2.7), using the assumptions of Regime 2. One obtains straightforwardly the following system

(4.1)
$$\begin{cases} (\partial_t - \mathrm{Fr}\,\partial_x)\zeta_2 + \dfrac{1}{\delta+\gamma}\partial_x v + \epsilon_2 \dfrac{\delta^2-\gamma}{(\gamma+\delta)^2}\partial_x(\zeta_2 v) + \mu \dfrac{1+\gamma\delta}{3\delta(\delta+\gamma)^2}\partial_x^3 v = -\alpha \dfrac{\mathrm{Fr}\,\gamma}{\delta+\gamma}\dfrac{\mathrm{d}}{\mathrm{d}x}\zeta_1, \\ \\ (\partial_t - \mathrm{Fr}\,\partial_x)v \;+\; (\gamma+\delta)\partial_x\zeta_2 \;+\; \dfrac{\epsilon_2}{2}\dfrac{\delta^2-\gamma}{(\gamma+\delta)^2}\partial_x\left(|v|^2\right) \;=\; 0. \end{cases}$$

This Boussinesq-type system can be written in a compact form as

$$\partial_t U \;+\; A_0\partial_x U \;+\; \epsilon_2 A_1(U)\partial_x U \;-\; \mu A_2 \partial_x^3 U \;=\; \alpha b_0(x), \tag{4.2}$$

with $U \;=\; (\zeta_2, v)^T$, $b_0 \;=\; -\,\mathrm{Fr}\,\frac{\gamma}{\delta+\gamma}(\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1, 0)^T$, and

$$A_0 = \begin{pmatrix} -\,\mathrm{Fr} & \frac{1}{\delta+\gamma} \\ \delta+\gamma & -\,\mathrm{Fr} \end{pmatrix}, \quad A_1(U) = \frac{\delta^2-\gamma}{(\gamma+\delta)^2}\begin{pmatrix} v & \zeta_2 \\ 0 & v \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & -\frac{1+\gamma\delta}{3\delta(\delta+\gamma)^2} \\ 0 & 0 \end{pmatrix}.$$

Following the classical theory of hyperbolic quasilinear equations, our aim is now to obtain an appropriate symmetrizer of this system. Let us define

$$\begin{aligned} S(U) \;&\equiv\; S_0 \;+\; \epsilon_2 S_1(U) \;-\; \mu S_2\partial_x^2 \\ &=\; \begin{pmatrix} \delta+\gamma & 0 \\ 0 & \frac{1}{\delta+\gamma} \end{pmatrix} + \epsilon_2\, \frac{\delta^2-\gamma}{(\gamma+\delta)^2}\begin{pmatrix} 0 & -v \\ -v & \zeta_2 \end{pmatrix} - \mu\, \begin{pmatrix} 0 & 0 \\ 0 & -\frac{1+\gamma\delta}{3\delta(\delta+\gamma)^2} \end{pmatrix}\, \partial_x^2\,. \end{aligned}$$

Multiplying (4.2) on the left by $S(U) - \mu K S_0\partial_x^2$, and dropping the $\mathcal{O}(\mu^2)$ terms, leads to the following equivalent system

(4.3)
$$\left(\; S_0 \;+\; \epsilon_2 S_1(U) \;-\; \mu(S_2 + K S_0)\partial_x^2 \;\right)\partial_t U \;+\; \left(\; \Sigma_0 \;+\; \epsilon_2\Sigma_1(U) \;-\; \mu\Sigma_2 \;\right)\partial_x U \;=\; \alpha b(x),$$

with $b \;=\; -\,\mathrm{Fr}\,\gamma(\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1, 0)^T$, and

$$\begin{aligned} \Sigma_0 \;&=\; \begin{pmatrix} -\,\mathrm{Fr}(\gamma+\delta) & 1 \\ 1 & \frac{-\,\mathrm{Fr}}{\gamma+\delta} \end{pmatrix}, \qquad \Sigma_1(U) \;=\; \frac{\delta^2-\gamma}{\gamma+\delta}\begin{pmatrix} 0 & \zeta_2 + \frac{\mathrm{Fr}\;v}{\gamma+\delta} \\ \zeta_2 + \frac{\mathrm{Fr}\;v}{\gamma+\delta} & \frac{-\,\mathrm{Fr}\;\zeta_2}{\gamma+\delta} \end{pmatrix}, \\ \Sigma_2 \;&=\; -\frac{1+\gamma\delta}{3\delta(\delta+\gamma)}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \;+\; K\;\Sigma_0. \end{aligned}$$

As we can see, the system (4.3) is perfectly symmetric, and K can be chosen so that S_0 and $S_2 + KS_0$ are definite positive. With these properties, we are able to use energy methods in order to prove that (4.3) is well-posed, and convergent with the full Euler system (2.4), at order $\mathcal{O}(\mu)$ up to times of order $\mathcal{O}(1/\mu)$. More precisely, one has

Proposition 4.1. *Let $s > 3/2$, $t_0 \geq 9/2$ and $U = (\zeta_1, \zeta_2, \psi_1, \psi_2)$ be an* adapted *solution of the full Euler system* (2.4)*, bounded in $W^{1,\infty}([0, T/\mu); H^{s+t_0})$. We define then $V \equiv (\zeta_2, v)$ by*

$$v \equiv\; \partial_x\left(\phi_2\Big|_{z=\epsilon_2\zeta_2} - \gamma\phi_1\Big|_{z=\epsilon_2\zeta_2}\right) \;\equiv\; \partial_x\psi_2 - \gamma H(\psi_1, \psi_2).$$

Moreover, let us assume that there exists a constant C_0 such that $\epsilon_2 \leq C_0\mu$ and $\alpha \leq C_0\mu$. Then there exists a constant $C_1 = C(\frac{1}{\gamma+\delta}, \gamma+\delta, C_0) > 0$ such that if $\epsilon_2 \big| V\,|_{t=0}\,\big|_{H^{s+1}_\mu} \;\leq\; \frac{1}{C_1}$, then there exists a constant $T(C_1) \;>\; 0$, independent of μ, and a unique solution of (4.3)

$V_B \in C^0([0,T/\mu);H^{s+1}_\mu)\cap C^1([0,T/\mu);H^s_\mu)$, bounded in $W^{1,\infty}([0,T/\mu);H^s_\mu)$, with initial data $V_B\,|_{t=0} = V\,|_{t=0}$. Moreover, one has for all $t\in[0,T/\mu)$,

$$\big|V - V_B\big|_{L^\infty([0,t];H^s)} \;\leq\; \mu^2 t C,$$

with $C = C(\frac{1}{h_{\min}}, \frac{1}{\gamma+\delta}, \gamma+\delta, \big|V\big|_{W^{1,\infty}H^{s+t_0}}, T)$.

Before starting with the proof, let us remark that the proposition is not empty only if there actually exists a family of solutions of (2.4), which is smooth and bounded in $W^{1,\infty}([0,T/\mu);H^{s+t_0})$. As discussed in Remark 2.2 page 133, this requires adding a surface tension term. However, this surface tension term is very small in practical cases, so that we can assume $\frac{1}{\mathrm{Bo}} = \mathcal{O}(\mu^2)$, and the result is obtained as in the proof presented below.

Proof. *Step 1: Well-posedness.* In order to prove the well-posedness of the symmetric system (4.3), we use techniques of [**41**]. It is proved in Proposition 2.4 therein that systems of the form (4.3) (with four equations instead of two and without the right-hand size), satisfying

i. The matrices S_0, Σ_0, S_2, Σ_2 are symmetric,
ii. $S_1(\cdot)$ and $\Sigma_1(\cdot)$ are linear mappings, and for all $U\in\mathbb{R}^4$, $S_1(U)$ and $\Sigma_1(U)$ are symmetric,
iii. S_0 and S_2 are definite positive,

are well-posed and satisfy an energy estimate. The proof is easily adapted for the case of a non-zero right-hand side, and we briefly give the arguments here. The key point of the proof relies on a differential inequality, satisfied by the following energy (with $\Lambda^s \equiv (1-\partial_x^2)^{s/2}$)

$$E_s(U) \;\equiv\; 1/2(S_0\Lambda^s U,\Lambda^s U) + \epsilon_2/2(S_1(U)\Lambda^s U,\Lambda^s U) + \mu/2(S_2\Lambda^s\partial_x U,\Lambda^s\partial_x U).$$

This energy is easily proved, thanks to the positiveness of S_0 and S_2 and using the smallness assumptions of Regime 2, to be equivalent to the $\big|\cdot\big|_{H^{s+1}_\mu}$ norm, that is to say there exists $C_0>0$ such that

$$\frac{1}{C_0}\left(\big|U\big|^2_{H^s} + \mu\big|U\big|^2_{H^{s+1}}\right) \;\leq\; E_s(U) \;\leq\; C_0\left(\big|U\big|^2_{H^s} + \mu\big|U\big|^2_{H^{s+1}}\right).$$

Then, we prove that there exists $C_1 = C_1(\frac{1}{\gamma+\delta},\gamma+\delta)>0$ such that if $\epsilon_2\big|U\big|_{H^{s+1}_\mu} \leq \frac{1}{C_1}$, then the operator defined by $P(U,\partial_x) = S_0 + \epsilon_2 S_1(U) - \mu S_2\partial_x^2 : H^{s+1}\to H^{s-1}$ is one-to-one and onto, and that $P(U,\partial_x)^{-1}\big(\Sigma_0+\epsilon_2\Sigma_1(U)-\mu\Sigma_2\partial_x^2\big)$ is uniformly bounded $H^s_\epsilon\to H^s_\epsilon$, so that any solution V of (4.3) will satisfy the a priori estimate

$$\big|\partial_t V\big|_{H^s_\epsilon} = \Big|P(V,\partial_x)^{-1}\Big(\big(\Sigma_0+\epsilon_2\Sigma_1(V)-\mu\Sigma_2\partial_x^2\big)\partial_x V - \alpha b\Big)\Big|_{H^s_\epsilon} \leq C_2(\big|V\big|_{H^{s+1}_\epsilon} + \alpha\big|b\big|_{H^s}),$$

with C_2 independent of ϵ_2 and μ, as long as $\epsilon_2\big|V\big|_{H^{s+1}_\mu} \leq \frac{1}{C_1}$. It follows that $E_s(V)$ satisfies

$$\begin{aligned}
\frac{d}{dt}E_s(V) &= \epsilon_2/2(S_1(\partial_t V)\Lambda^s V,\Lambda^s V) - \epsilon_2([\Lambda^s,S_1(V)]\partial_t V,\Lambda^s V) + \alpha(\Lambda^s b,\Lambda^s V)\\
&\quad + \epsilon_2/2((\Sigma_1(\partial_x V)\Lambda^s V),\Lambda^s V) - \epsilon_2([\Lambda^s,\Sigma_1(V)]\partial_x V,\Lambda^s V)\\
&\leq C_3\left(\mu\big|V\big|^2_{H^s}\big(\big|V\big|_{H^{s+1}_\mu} + \alpha\big|b\big|_{H^s}\big) + \alpha\big|b\big|_{H^s}\big|V\big|_{H^s}\right)\\
&\leq C_4\left(\,\epsilon_2 E_s(V)^{3/2} + \alpha\big|b\big|_{H^s}E_s(V)^{1/2}\right).
\end{aligned}$$

The last inequalities come from Cauchy-Schwarz inequality, Sobolev embeddings, Kato-Ponce commutator estimates and the above a priori estimate (see Appendix A of [**41**]). The Gronwall-Bihari's Lemma allows to conclude that as long as $\epsilon_2\big|V\big|_{H^{s+1}_\mu} \leq \frac{1}{C_0}$, one has

$$\big|V\big|_{H^{s+1}_\mu} \;\leq\; C_0\; E(R_s)^{1/2} \leq C_5\alpha\big|f\big|_{H^s}\; t. \tag{4.4}$$

Using the assumptions of Regime 2 ($\alpha = \mathcal{O}(\mu), \epsilon_2 = \mathcal{O}(\mu)$), one can then follow the classical Friedrichs proof, and obtain the existence of $T(\gamma+\delta, \frac{1}{\gamma+\delta}, (\epsilon_2 \big|V^0\big|_{H^{s+1}_\mu})^{-1}) > 0$ and a solution V of (4.3), defined over times $[0, T/\mu)$, such that $V \in C^0([0,T/\mu); H^{s+1}) \cap C^1([0,T/\mu); H^s)$.

Finally, the uniqueness of the solution is obtained in the same way, applying the energy estimate to the difference of two solutions. Indeed, let $V_1, V_2 \in C^0([0,T/\epsilon); H^{s+1}) \cap C^1([0,T/\epsilon); H^s)$ be two solutions of the Cauchy problem (4.3) with same initial value $V_1\,|_{t=0} = V_2\,|_{t=0} = V^0$. The functions V_1 and V_2 are uniformly bounded thanks to (4.4). One can immediately check that $R \equiv V_1 - V_2$ satisfies

$$\begin{aligned}\Big(S_0 + \epsilon_2 S_1(V_1) - \mu S_2 \partial_x^2\Big)\partial_t \Lambda^s R + \Big(\Sigma_0 + \epsilon_2 \Sigma_1(V_1) - \mu \Sigma_2 \partial_x^2\Big)\partial_x \Lambda^s R \\ + \epsilon_2 [\Lambda^s, S_1(V_1)]\partial_t R + \epsilon_2 [\Lambda^s, \Sigma_1(V_1)]\partial_x R = \epsilon_2 F,\end{aligned} \tag{4.5}$$

with $F = -\Lambda^s\Big(S_1(R)\partial_t V_2 + \Sigma_1(R)\partial_x U_2\Big)$. Then, we can carry out the above method on R with a modified right-hand side, and obtain the equivalent energy estimate

$$\frac{d}{dt} E_s(R) \leq \epsilon_2 C_6 (\big|V_1\big|_{H^s} + \big|V_2\big|_{H^s}) E_s,$$

with $C_6 = C(\frac{1}{\gamma+\delta}, \delta+\gamma, \big|U^0\big|_{H^{s+1}_\epsilon})$. From Gronwall-Bihari's inequality, using the uniform boundedness of V_1 and V_2 on $[0,T/\epsilon)$, and since $E_s(R)\,|_{t=0} = 0$, one has immediately $E_s(R) = 0$ on $[0,T/\epsilon)$, and finally $V_1 = V_2$.

Step 2: Consistency. We prove that, assuming that $\alpha = \mathcal{O}(\mu)$ and $\epsilon_2 = \mathcal{O}(\mu)$, the full Euler system (2.4) is consistent with the models (4.2) and (4.3), both at precision $\mathcal{O}(\mu^2)$ on $[0,T]$, $T > 0$.

Let $U \equiv (\zeta_1, \zeta_2, \psi_1, \psi_2)$ be a strong solution of (2.4), bounded in $W^{1,\infty}([0,T]; H^{s+t_0})$ with $s > 1$ and $t_0 \geq 9/2$, and such that (2.5) is satisfied with $\zeta_1(t,x) \equiv \zeta_1(x)$. The consistency result of Proposition A.1 states that (ζ_2, v), with $v \equiv \partial_x \psi_2 - \gamma H(\psi_1, \psi_2)$, satisfies (2.7), up to $R_1 = (r_1, r_2)^T \in L^\infty([0,T]; H^s)^2$, satisfying (for $i = 1, 2$)

$$\big|r_i\big|_{L^\infty H^s} \leq \mu^2\, C_0\left(\frac{1}{h_{\min}}, \big|U\big|_{W^{1,\infty} H^{s+t_0}}\right).$$

It is the straightforward, as in the proof of Proposition 3.1, to show that the terms dropped in (4.1), and later in (4.3), are all bounded by $\mu^2\, C_0\left(\frac{1}{h_{\min}}, \big|U\big|_{W^{1,\infty}H^{s+4}}\right)$ in $L^\infty([0,T]; H^s)$ norm. Finally, (ζ_2, v) satisfies (4.2) and (4.3) with modified right-hand sides, *i.e.* , respectively,

$$\alpha \tilde{b_0} = \alpha b_0 + \mu^2 f_0, \quad \text{and} \quad \alpha \tilde{b} = \alpha b + \mu^2 f,$$

with f_0, f uniformly bounded in $L^\infty([0,T]; H^s)$.

Step 3: Convergence. The convergence estimates of the proposition follow easily from the calculations of *Step 1*, together with *Step 2*. Indeed, thanks to the consistency result, $V - V_B$ satisfy (4.5), with modified right-hand side

$$\tilde{F} \equiv F + \mu^2 f, \qquad f \in L^\infty([0,T]; H^s).$$

It follows that $E_s(V - V_B)$ satisfies

$$\frac{d}{dt} E_s(V - V_B) \leq \epsilon_2 C_6 E_s(V - V_B) + \mu^2 C_7 \big|f\big|_{H^s} E_s(V - V_B)^{1/2},$$

and the Gronwall-Bihari's Lemma leads to

$$\big|V - V_B\big|_{L^\infty([0,T/\mu); H^{s+1}_\mu)} \leq C_0 E_s(V - V_B)^{1/2} \leq \frac{C_7 \mu^2}{C_6 \epsilon_2} \big|f\big|_{H^s} (e^{C_6 \epsilon_2 t} - 1).$$

The convergence estimate of the proposition follow from $\epsilon_2 = \mathcal{O}(\mu)$. □

4.2. The Korteweg-de Vries approximation. In this section, we use a WKB approximation, in order to deduce from the symmetric Boussinesq-type system (4.3), an approximated model that consists in two uncoupled forced Korteweg-de Vries (fKdV) equations. This method has been used, for example, in [**41, 71**], and is briefly presented in the following.

The idea is to look for an approximate solution of the Cauchy problem (4.3) with initial data U^0, under the form

$$U_{\rm app}(t,x) = U_0(\mu t, t, x) + \mu U_1(\mu t, t, x),$$

with the profiles $U_0(\tau,t,x)$ and $\mu U_1(\tau,t,x)$ satisfying $U_0\,|_{t=\tau=0} = U^0$ and $U_1\,|_{t=\tau=0} = 0$.

Plugging the *Ansatz* into (4.3) leads to the following equation

$$\begin{aligned}(S_0\partial_t + \Sigma_0\partial_x)U_0 + \mu S_0\partial_\tau U_0 + \epsilon_2\big(S_1(U_0)\partial_t U_0 + \Sigma_1(U_0)\partial_x U_0\big) - \mu\big(S_2\partial_x^2\partial_t U_0 + \Sigma_2\partial_x^3 U_0\big)\\ +\mu(S_0\partial_t + \Sigma_0\partial_x)U_1 + \mu^2 R = 0.\end{aligned} \tag{4.6}$$

We now deduce $U_0(\tau,t,x)$ and $U_1(\tau,t,x)$, by solving (4.6) at each order.

At order $\mathcal{O}(1)$ We solve

$$(\ S_0\partial_t\ +\ \Sigma_0\partial_x\ )U_0\ =\ 0. \tag{4.7}$$

Let us define $\mathbf{e}_\pm \equiv \frac{1}{\sqrt{2}}(\pm\frac{1}{\sqrt{\gamma+\delta}}, \sqrt{\gamma+\delta})^T$. One can check that the basis satisfies

$$\mathbf{e}_i\cdot\Sigma_0\mathbf{e}_j = c_i\delta_{i,j}, \quad \text{and} \quad \mathbf{e}_i\cdot S_0\mathbf{e}_j = \delta_{i,j},$$

with $\delta_{i,j}$ the classical Kronecker delta symbol, and $c_\pm = \pm 1 - \mathrm{Fr}$.

Therefore, when we define $u_\pm \equiv \mathbf{e}_\pm\cdot S_0U_0$ (and hence $U_0 = u_+\mathbf{e}_+ + u_-\mathbf{e}_-$), the scalar product of (4.7) with $\mathbf{e}_\pm$ leads to

$$(\ \partial_t\ +\ c_\pm\partial_x\ )u_\pm\ =\ 0.$$

Finally, since $u_\pm$ satisfies a scalar transport equation, we use the notation

$$u_\pm(\tau,t,x)\ =\ u_\pm(\tau,x-c_\pm t)\ =\ u_\pm(\tau,x_\pm), \tag{4.8}$$

with initial data $u_\pm(0,x_\pm)\ =\ \mathbf{e}_\pm\cdot S_0U^0(x_\pm)$.

At order $\mathcal{O}(\mu)$ We solve

$$S_0\partial_\tau U_0 + \frac{\epsilon_2}{\mu}\left(\Sigma_1(U_0)\partial_x U_0 + S_1(U_0)\partial_t U_0\right) - \Sigma_2\partial_x^3 U_0 - S_2\partial_x^2\partial_t U_0 + (S_0\partial_t + \Sigma_0\partial_x)U_1 = \frac{\alpha}{\mu}b(x), \tag{4.9}$$

that we can split in

$$\partial_\tau u_\pm\ +\ \lambda_\pm u_\pm\partial_{x_\pm}u_\pm\ +\ \nu_\pm\partial^3_{x_\pm}u_\pm\ =\ \beta_\pm(x), \tag{4.10}$$

with $\lambda_\pm \equiv \frac{\epsilon_2}{\mu}\mathbf{e}_\pm\cdot(\Sigma_1 - c_\pm S_1)(\mathbf{e}_\pm))\mathbf{e}_\pm$, $\nu_\pm \equiv \mathbf{e}_\pm\cdot(-\Sigma_2 + c_\pm S_2)\mathbf{e}_\pm$ and $\beta_\pm \equiv \frac{\alpha}{\mu}\mathbf{e}_\pm\cdot b$; and on the other hand,

$$(\partial_t + c_i\partial_x)\mathbf{e}_i\cdot S_0U_1 + \sum_{(j,k)\neq(i,i)}\alpha_{ijk}u_k(\tau,x-c_kt)\partial_x u_j(\tau,x-c_jt) = \sum_{j\neq i}\beta_{ij}\partial_x^3 u_j(\tau,x-c_jt), \tag{4.11}$$

with $\alpha_{ijk} \equiv \mathbf{e}_i\cdot(\Sigma_1(\mathbf{e}_k) - c_jS_1(\mathbf{e}_k))\mathbf{e}_j$ and $\beta_{ij} \equiv \mathbf{e}_i\cdot(\Sigma_2 - c_jS_2)\mathbf{e}_j$.

It is clear that u_i satisfies (4.8) and (4.10), if and only if $u_i(\epsilon t,t,x)$ satisfies the Korteweg-de Vries equation:

$$\partial_t u_\pm\ +\ c_\pm\partial_x u_\pm\ +\ \mu\left(\ \lambda_\pm u_\pm\partial_x u_\pm\ +\ \nu_\pm\partial_x^3 u_\pm\ \right)\ =\ \mu\beta_\pm(x).$$

Finally, simple calculations show that in our case, we can decompose

$$\zeta_2 \;\equiv\; \eta_+ \;+\; \eta_-, \quad \text{with} \;\; \eta_\pm \;\equiv\; \pm\frac{1}{\sqrt{2(\gamma+\delta)}}u_\pm$$

satisfying precisely the following KdV equation

$$\partial_t\eta_\pm \;+\; (-\operatorname{Fr}\pm 1)\partial_x\eta_\pm \;\pm\; \epsilon_2\frac{3}{2}\frac{\delta^2-\gamma}{\gamma+\delta}\eta_\pm\partial_x\eta_\pm \;\pm\; \mu\frac{1}{6}\frac{1+\gamma\delta}{\delta(\gamma+\delta)}\partial_x^3\eta_\pm \;=\; -\alpha\operatorname{Fr}\gamma\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1.$$

Unsurprisingly, we recover the KdV approximation with a flat rigid lid, when $\alpha = 0$ (see [**41**] and references therein).

Using these forced Korteweg-de Vries equations as an approximation of the full problem is justified up to times of order $\mathcal{O}(1/\mu)$ by the following proposition.

PROPOSITION 4.2. *Let $s > 1/2$, $t_0 \geq 5 + 5/2$ and $U = (\zeta_1, \zeta_2, \psi_1, \psi_2)$ be an* adapted *solution of the full Euler system* (2.4)*, bounded in $W^{1,\infty}([0, T/\mu); H^{s+t_0})$. We define $V \equiv (\zeta_2, v)$ by*

$$v \;\equiv\; \partial_x\left(\phi_2\Big|_{z=\epsilon_2\zeta_2} - \gamma\phi_1\Big|_{z=\epsilon_2\zeta_2}\right) \;\equiv\; \partial_x\psi_2 - \gamma H(\psi_1,\psi_2).$$

Then there exists η_+ and η_-, the two solutions of the following forced Korteweg-de Vries equation

(4.12)

$$\partial_t\eta_\pm \;+\; (-\operatorname{Fr}\pm 1)\partial_x\eta_\pm \;\pm\; \epsilon_2\frac{3}{2}\frac{\delta^2-\gamma}{\gamma+\delta}\eta_\pm\partial_x\eta_\pm \;\pm\; \mu\frac{1}{6}\frac{1+\gamma\delta}{\delta(\gamma+\delta)}\partial_x^3\eta_\pm \;=\; -\alpha\operatorname{Fr}\gamma\frac{d}{dx}\zeta_1(x),$$

with $\eta_\pm\big|_{t=0} = \frac{1}{2}\big(\zeta_2 \pm \frac{1}{\gamma+\delta}v\big)\big|_{t=0}$. Moreover, if there exists a constant C_0 such that $\epsilon_2 \leq C_0\mu$ and $\alpha \leq C_0\mu$, then one has for all $t \in [0, T/\mu)$,

$$\big|\zeta_2 - (\eta_+ + \eta_-)\big|_{L^\infty([0,t];H^s} \;+\; \big|v - (\gamma+\delta)(\eta_+ - \eta_-)\big|_{L^\infty([0,t];H^s)} \;\leq\; \mu\sqrt{t}C,$$

with $C = C(\frac{1}{h_{\min}}, \frac{1}{\gamma+\delta}, \gamma+\delta, \big|U\big|_{W^{1,\infty}H^{s+t_0}}, C_0)$.

Additionally, if V satisfies $(1+x^2)V\big|_{t=0} \in H^{s+5}$, then one has the improved estimate

$$\big|\zeta_2 - (\eta_+ + \eta_-)\big|_{L^\infty([0,t];H^s} \;+\; \big|v - (\gamma+\delta)(\eta_+ - \eta_-)\big|_{L^\infty([0,t];H^s)} \;\leq\; \epsilon C',$$

with $C' = C(\frac{1}{h_{\min}}, \frac{1}{\gamma+\delta}, \gamma+\delta, \big|U\big|_{W^{1,\infty}H^{s+t_0}}, \big|(1+x^2)V\big|_{t=0}\big|_{H^{s+5}})$.

The proposition is obtained by a simple adaptation of the techniques presented in [**41**], with additional forcing terms in the Korteweg-de Vries equations. The proof is given in Appendix B, for the sake of completeness.

4.3. Analysis of the forced Korteweg-de Vries equation. The forced Korteweg-de Vries equation

$$\partial_t u \;+\; c\partial_x u \;+\; \varepsilon\lambda u\partial_x u \;+\; \varepsilon\nu\partial_x^3 u \;=\; \varepsilon\frac{\mathrm{d}}{\mathrm{d}x}f(x) \tag{4.13}$$

has attracted a lot of interests, especially in the framework of the one layer water wave problem (where a moving topography, or pressure, is the forcing term that generates waves). Of particular interest is the problem of the generation of solitons, that have first been numerically discovered by Wu and Wu [**116**], and validated with experiments by Lee [**74**]. Using the Boussinesq-type system or the KdV approximation, they found that starting with a zero initial data, the solution can generate periodically an infinite number of solitary waves. Numerous work have have then tackled this problem, including [**119, 75, 107, 166**]. It appeared that the Froude number (which is given by $1 - c$ in (4.13)) is playing a predominant role in this phenomenon, as the generation of solitons only occurs for a narrow band of its values. One could roughly summarize the observations by the existence of $F_c > 1$ such that

i. if $\mathrm{Fr} > F_c$, then the flow approaches a steady state, symmetric and localized at the site of forcing;
ii. for $\mathrm{Fr} < F_c$, solitons are periodically generated at the site of forcing and radiated up-stream;
iii. the amplitude of the generated solitons goes to zero as $\mathrm{Fr} \to -\infty$.

The existence of steady solitary waves of (4.13), and their stability is therefore essential. This issue has been studied for specific forcing terms by Camassa and Wu in **[25, 24]**, and for general forcing terms by Choi, Lin, Sun and Whang in **[132, 131]**. They prove that for Fr sufficiently large, there exists a unique small steady solution. Moreover, this solution is proved to be symmetric and localized at the site of forcing, and stable in $H^1(\mathbb{R})$ (in the Lyapunov sense). The behavior when the Froude number Fr approaches unity is more peculiar and depends on the sign of the forcing term. Subcritical Froude numbers are not studied.

In the framework of our analysis, the values of the coefficients λ and ν depend on the parameters γ and δ, and their order of magnitude depends on ϵ_2 and μ (the magnitude of the forcing term depends mostly on α). We use the factor ε to keep in mind that all of these coefficients are assumed to be small. More precisely, we recall

$$c_\pm = \pm 1 - \mathrm{Fr}, \quad \varepsilon\lambda_\pm = \pm\epsilon_2 \frac{3}{2}\frac{\delta^2-\gamma}{\gamma+\delta}, \quad \varepsilon\nu_\pm = \pm\mu\frac{1}{6}\frac{1+\gamma\delta}{\delta(\gamma+\delta)}, \quad \varepsilon f(x) = -\,\mathrm{Fr}\,\alpha\gamma\zeta_1(x).$$

When $\delta^2 - \gamma \neq 0$, a simple change of parameters allows to recover values of the one layer problem, and one obtains straightforwardly similar results. Our approach is quite different. In the following section, we prove that away from the critical speed ($c = 0$), the solution of (4.13) with small initial data will remain small (in $H^s(\mathbb{R})$ norm) for long times, using smallness assumptions on the forcing terms and the coefficients $\varepsilon\lambda$ and $\varepsilon\nu$. Numerically, it appears that the solution converges locally towards a negative steady state in the supercritical case, and that small solitons are continuously generated otherwise. Then, in Section 4.3.2, we study numerically the generation of up-stream propagating solitons for Froude numbers around the critical value.

The consequences of this study to the dead-water effect are the following:

i. away from the critical Froude number, the drag suffered by the ship is always small;
ii. the peak of wave resistance occurs for Froude numbers just below the critical value;
iii. the time-period of the generation of solitons predicted by the KdV approximation is of the same order as the time-scale of relevance to the model, so that in cannot clearly be held responsible for the periodic aspect described among others in **[114]**.

4.3.1. *Non-critical Froude numbers.* In the following proposition, we obtain an improved growth rate for the solution of the forced Korteweg-de Vries equation (4.13), if the velocity coefficient c is away from zero, when assuming smallness on the nonlinearity and dispersion coefficients λ and ν, and on the initial data and the forcing term. This phenomenon is easily explained, when looking at the linear transport equation related to (4.13):

$$\partial_t v \; + \; c\partial_x v \; = \; \varepsilon \frac{\mathrm{d}}{\mathrm{d}x} f(x). \tag{4.14}$$

The Cauchy problem, with initial data $v\,|_{t=0} = \varepsilon u_0$, is solved by

$$v \; \equiv \; \varepsilon \; u_0(x-ct) \; + \; \frac{\varepsilon}{c}\Big(\; f(x) - f(x-ct) \; \Big).$$

The solution is therefore bounded for all times as soon as $c \neq 0$, and small if the initial data and the forcing terms are small. We will obtain improved bound estimates on u the solution of the forced KdV equation, by estimating the difference $\big|u - v\big|_{H^s}$.

PROPOSITION 4.3. *Let $s > 3/2$ and u be the solution of*

$$\partial_t u \ + \ c\partial_x u \ + \ \varepsilon\lambda u\partial_x u \ + \ \varepsilon\nu\partial_x^3 u \ = \ \varepsilon\frac{d}{dx}f(x),$$

such that $u\,|_{t=0} = \varepsilon u_0$. Let us assume that there exists $M > 0$ such that

$$\left|\frac{1}{c}\right|, \ |\lambda|, \ |\nu|, \ \big|u_0\big|_{H^{s+3}}, \ \big|f\big|_{H^{s+3}} \ \leq M.$$

Then there exists $T(M,s) > 0$ and $C = C(M,s) > 0$, such that the function u is bounded on $[0, T/\varepsilon]$ by

$$\big|u(t)\big|_{L^\infty([0,T/\varepsilon];H^s)} \ \leq \ C\varepsilon.$$

PROOF. Thanks to **[18]**, we know that the function u exists and is unique. We define as above

$$v \ \equiv \ \varepsilon \, u_0(x-ct) \ + \ \frac{\varepsilon}{c}\Big(\ f(x) - f(x-ct) \ \Big),$$

the solution of the transport equation

$$\partial_t v \ + \ c\partial_x v \ = \ \varepsilon\frac{\mathrm{d}}{\mathrm{d}x}f(x),$$

with same initial data $v\,|_{t=0} = \varepsilon u_0$. Since the function v is uniformly bounded by

$$\big|v(t)\big|_{L^\infty(\mathbb{R};H^s)} \ \leq \ C\varepsilon,$$

the estimate of the proposition follows from the same estimate on the difference $r \ \equiv \ u - v$. It is easy to check that r satisfies

$$\partial_t r \ + \ c\partial_x r \ + \ \varepsilon\lambda(r+v)\partial_x(r+v) \ + \ \varepsilon\nu\partial_x^3(r+v) = \ 0, \tag{4.15}$$

and $r\,|_{t=0} = 0$. When multiplying (4.15) by $\Lambda^{2s}r$, and integrating, one obtains

$$\frac{1}{2}\frac{d}{dt}(\Lambda^s r, \Lambda^s r) + c(\Lambda^s\partial_x r, \Lambda^s r) + \varepsilon\lambda(\Lambda^s\left((r+v)\partial_x(r+v)\right), \Lambda^s r) - \varepsilon\nu(\Lambda^s\partial_x^2(r+v), \Lambda^s\partial_x r) = 0,$$

denoting $(f,g) \ \equiv \int_{\mathbb{R}} fg$ the L^2-based inner product. It follows then

$$\frac{1}{2}\frac{d}{dt}\left(|r|_{H^s}^2\right) \ = \ -\varepsilon\lambda(\Lambda^s\left((r+v)\partial_x(r+v)\right), \Lambda^s r) \ - \ \varepsilon\nu(\Lambda^s\partial_x^3 v, \Lambda^s r).$$

Some of the terms of the right-hand side are straightforwardly estimated using Cauchy-Schwarz inequality, and the algebraic properties of $H^s(\mathbb{R})$, $s > 1/2$:

$$\begin{aligned}
\big|(\Lambda^s\partial_x^3 v, \Lambda^s r)\big| &\leq \ \big|v\big|_{H^{s+3}}\big|r\big|_{H^s}, \\
\big|(\Lambda^s(v\partial_x v), \Lambda^s r)\big| &\leq \ C_0(s)\big|v\big|_{H^{s+1}}\big|v\big|_{H^s}\big|r\big|_{H^s}, \\
\big|(\Lambda^s(r\partial_x v), \Lambda^s r)\big| &\leq \ C_0(s)\big|v\big|_{H^{s+1}}\big|r\big|_{H^s}^2.
\end{aligned}$$

As for the remaining terms, we integrate by part and obtain

$$\begin{aligned}
(\Lambda^s(r\partial_x r), \Lambda^s r) &= \ -\frac{1}{2}(\partial_x r\Lambda^s r, \Lambda^s r) + ([\Lambda^s, r]\partial_x r, \Lambda^s r), \\
(\Lambda^s(v\partial_x r), \Lambda^s r) &= \ -\frac{1}{2}(\partial_x v\Lambda^s r, \Lambda^s r) + ([\Lambda^s, v]\partial_x r, \Lambda^s r),
\end{aligned}$$

with $[T,f]$ the commutator defined by $[T,f]g \equiv T(fg) - fT(g)$. Since for $f \in H^s$, $s > 3/2$, one has $\big|\partial_x f\big|_{L^\infty} \leq \big|f\big|_{H^s}$, one can use the classical Kato-Ponce Lemma **[58]** to deduce the following estimates

$$\begin{aligned}
\big|(\Lambda^s(v\partial_x r), \Lambda^s r)\big| &\leq \ C_0(s)\big|v\big|_{H^s}\big|r\big|_{H^s}^2, \\
\big|(\Lambda^s(r\partial_x r), \Lambda^s r)\big| &\leq \ C_0(s)\big|r\big|_{H^s}^3.
\end{aligned}$$

Altogether, it follows that $|r|_{H^s}$ satisfies the differential inequality

$$\frac{1}{2}\frac{d}{dt}\left(|r|_{H^s}^2\right) \;\leq\; C_0(s)\varepsilon\lambda\left(|v|_{H^{s+1}}^2|r|_{H^s} + |v|_{H^{s+1}}|r|_{H^s}^2 + |r|_{H^s}^3\right) + \varepsilon\nu|v|_{H^{s+3}}|r|_{H^s},$$

with $C_0(s)$ a constant depending only on the parameter $s > 3/2$. Now, using the assumptions of the proposition, it is straightforward to see that one can rewrite

$$\frac{d}{dt}|r|_{H^s} \;\leq\; C_0\varepsilon^2 \;+\; C_0(s)\varepsilon\left(\varepsilon^2 + \varepsilon|r|_{H^s} + |r|_{H^s}^2\right).$$

Now, let us define $T^\star$ the maximum time such that $|r|_{H^s} \leq \varepsilon$ on $[0, T^\star]$ (this is true at $t = 0$, so that $T^\star > 0$ by a continuity argument). Then we deduce from the above inequality, and Gronwall's Lemma, that for all $t \in [0, T^\star]$,

$$|r|_{H^s} \;\leq\; C(s)\varepsilon^2 t,$$

with a constant $C(s)$ depending only on the parameter $s > 3/2$. From this very a priori estimate, we deduce that $T^\star \geq (C(s)\varepsilon)^{-1}$, and the proposition is proved. □

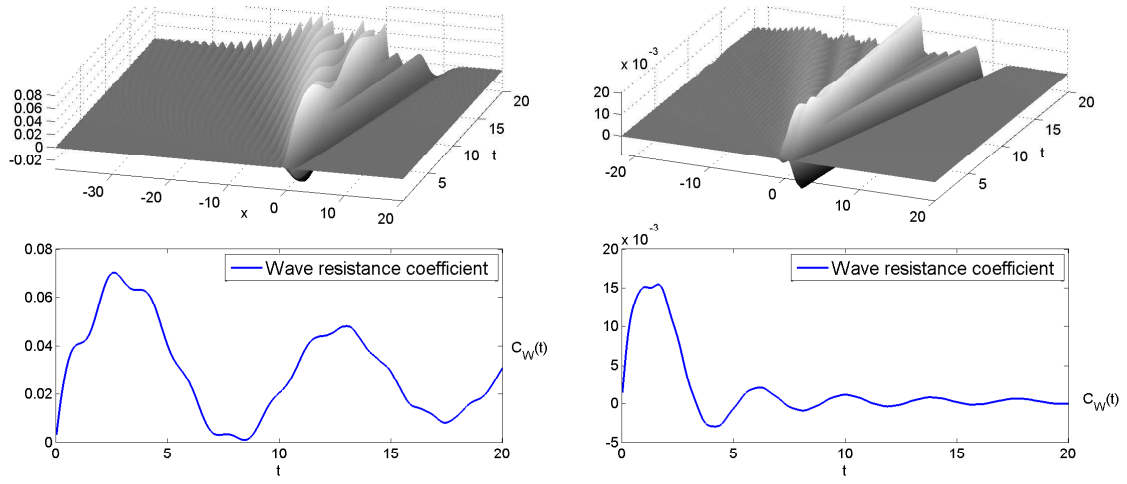
Remark 4.4. *One of the immediate consequences of Proposition 4.3 to our problem is that, in the decomposition of the KdV approximation (Proposition 4.2), the function η_- remains small on the time interval where the KdV approximation is a relevant model. Indeed, as $c_- \equiv -1 - \mathrm{Fr} < -1$, then if $\eta_-|_{t=0} = 0$ (small in H^{s+3} would suffice), then for $t \in [0, T/\mu)$, one has*

$$|\eta_-(t)|_{H^s} \;\leq\; C\Big(\frac{1}{\gamma+\delta}, \gamma+\delta, |\zeta_1|_{H^{s+3}}\Big)\mu.$$

Therefore, we focus in the following on η_+, the solution of

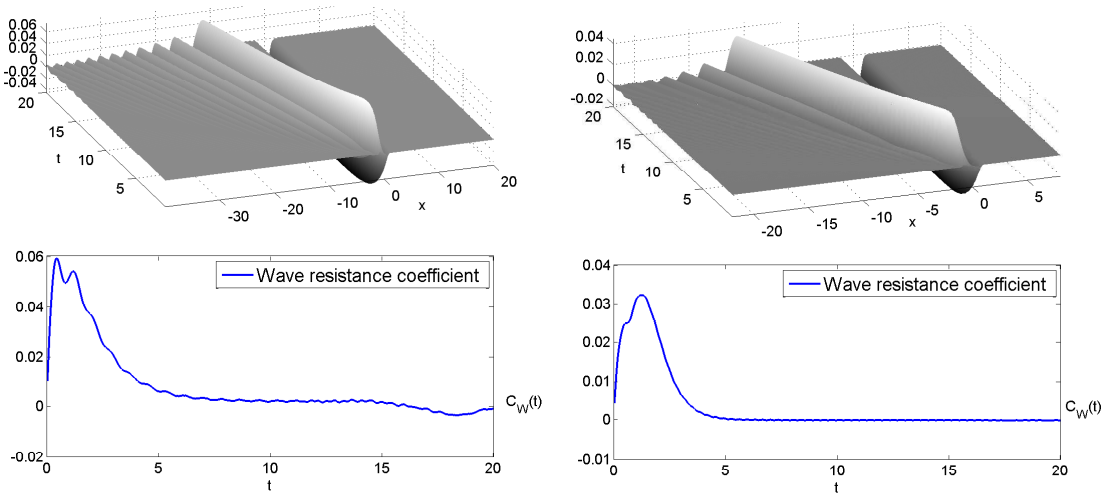
$$\partial_t\eta_+ \;+\; (1-\mathrm{Fr})\partial_x\eta_+ \;+\; \epsilon_2\frac{3}{2}\frac{\delta^2-\gamma}{\gamma+\delta}\eta_+\partial_x\eta_+ \;+\; \mu\frac{1}{6}\frac{1+\gamma\delta}{\delta(\gamma+\delta)}\partial_x^3\eta_+ \;=\; -\alpha\,\mathrm{Fr}\,\gamma\frac{d}{dx}\zeta_1. \tag{4.16}$$

In Figures 6 and 7, we compute the solution of (4.16), with zero initial data, in the supercritical ($\mathrm{Fr} = 1.5$) and subcritical ($\mathrm{Fr} = 0.5$) cases. For each of the figures, we plot the flow, depending on space and time variables, as well as the evolution of the related wave resistance coefficient C_W, calculated with formula (C.5) page 161. We use two values for the depth ratio: $\delta \in \{5/12, 12/5\}$.



(a) $\delta = 5/12$. (b) $\delta = 12/5$.

Figure 6. Subcritical flow: $\mathrm{Fr} = 0.5$. $\alpha = \epsilon_2 = \mu = 0.1$, $\gamma = 0.99$.



(a) $\delta = 5/12$. (b) $\delta = 12/5$.

FIGURE 7. Supercritical flow: Fr $= 1.5$. $\alpha = \epsilon_2 = \mu = 0.1$, $\gamma = 0.99$.

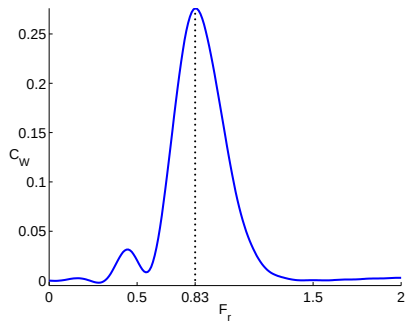
As we are away from the critical Froude number, the amplitudes of the generated waves, as well as the magnitude of the wave resistance coefficient, are small (at most of the order $\mathcal{O}(\mu)$). One remarks that the behavior of the flow is roughly independent of the value of the depth ratio δ. This is easily explained, as the quadratic nonlinearities do not play an important role when the deformation is small (as they are of order $\mathcal{O}(\mu^3)$). Therefore, the variations of δ are mostly seen through a variation of the size of the dispersion coefficient ν, which do not change the behavior pattern of the solution.

On the contrary, the Froude number parameter has an essential role on the behavior of the solution. One sees that for subcritical flows (Figure 6), up-stream propagating solitary waves are continuously generated, as a widening oscillatory tail is generated at the down-stream propagating area. The former will generate most of the wave resistance, as they overtake the body. However, the drag remains small, and possibly tends to zero in the long time limit. The behavior of supercritical flows (Figure 7) is quite different. A down-stream propagating solitary wave, with a small oscillatory tail, is generated and travels at constant velocity $c_0 \approx -1/2$. It remains a symmetric, depression wave at the location of the body, that does not generate any wave resistance. Unsurprisingly, this corresponds to the observations of the one-layer theory in the supercritical case.

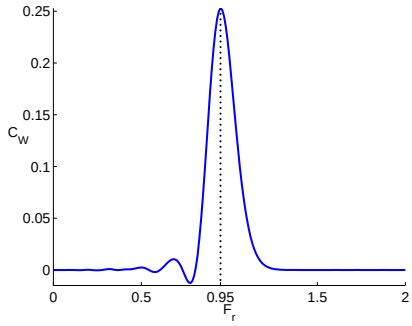
4.3.2. *Critical Froude numbers.* As we have seen in the previous section, the solution of the fKdV equation, and therefore the wave resistance coefficient (C.5), are small when the Froude number is away from its critical number Fr $= 1$. We present in Figure 8 the behavior of the wave resistance at time $t = 10$, depending on the Froude number, for different values of the depth ratio δ, and different values of the shallowness parameter μ.

It appears that the maximum of the critical wave is obtained below the critical speed, at approximate value Fr $\approx 1 - \text{Cstt } \lambda_+ = 1 - \text{Cstt } \mu\frac{1+\gamma\delta}{\gamma+\delta}$. The fact that the maximum peak is slightly subcritical has been obtained using the linear theory in [**88**], and explained as a result of dispersion effects (without the nonlinear and dispersion terms, the peak is infinite and obtained at Fr $= 1$). This effect is therefore preserved in the nonlinear theory.

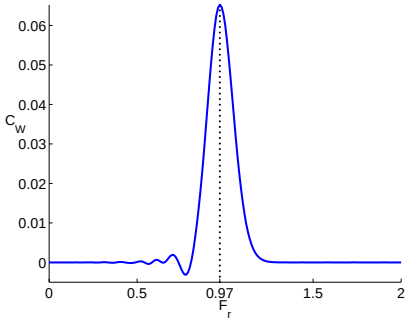
In Figure 9, we compute the solution of (4.16), with zero initial data, in the critical case (Fr $= 1$). Again, we plot the flow, depending on space and time variables, as well as the wave resistance coefficient, and we set $\delta \in \{5/12, 12/5\}$. It is known, since the work of Wu [**119**], that the forced KdV equation can generate up-stream propagating solitary



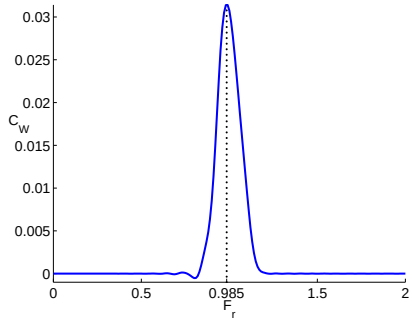
(a) $\alpha = \epsilon_2 = \mu = 0.1$, $\delta = 5/12$, $\gamma = 0.99$.



(b) $\alpha = \epsilon_2 = \mu = 0.1$, $\delta = 12/5$, $\gamma = 0.99$.



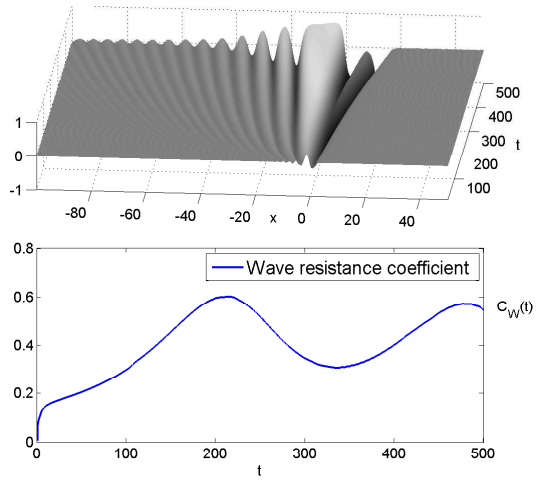
(c) $\alpha = \epsilon_2 = \mu = 0.01$, $\delta = 5/12$, $\gamma = 0.99$.



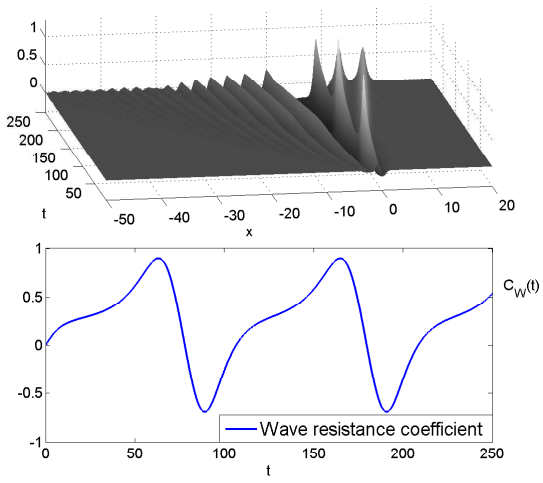
(d) $\alpha = \epsilon_2 = \mu = 0.01$, $\delta = 12/5$, $\gamma = 0.99$.

Figure 8. Dependence of the wave resistance coefficient on the Froude number, at time $T = 10$.

waves. Our simulations exhibit that behavior, and show that the depth ratio δ now plays an essential role, as it determines the sign of the coefficient of nonlinearity. Indeed, even if up-stream propagating waves are generated for both of the values of δ, these waves are of elevation if $\delta^2 > \gamma$, and of depression in the opposite case.



(a) $\delta = 5/12$.



(b) $\delta = 12/5$.

Figure 9. Critical flow: Fr $= 1$. $\alpha = \epsilon_2 = \mu = 0.1$, $\gamma = 0.99$.

The continuous generation of up-stream propagating waves induce a periodic aspect to the wave resistance, that recalls the hysteretic behavior of the dead-water phenomenon recorded in [**114**]. However, we do not believe that these waves are the cause of the periodic behavior recorded in [**114**], based on the two following facts.

i. The time period of the generation of waves ΔT is very large. A simple scaling argument shows that $\Delta T \propto 1/\mu$ if $c = 0$ and $\lambda, \alpha, \nu \propto \mu$, and the proportionality constant appears to be big (see also scaling arguments in [**119**]). By Proposition 4.2, the KdV approximation is a valid model only up to times of order $\mathcal{O}(1/\mu)$, so that nothing shows that the solutions of the full Euler equations (2.4) follow this behavior. It is possible that the time-range of validity of the model is shorter than one time period of generation of up-stream propagating waves.
ii. As a matter of fact, the solutions of the strongly nonlinear model (3.3) do not exhibit such a phenomenon, as the generated internal wave stays in the trail of the body, and never overtakes the ship, even for long times.

In Figure 5 page 142, we present the results of a simulation where the velocity of the ship is affected by the generated wave resistance. A similar periodic behavior is exhibited, with a more reasonable time period ($\Delta T \approx 10$). What is more, in this case, the up-stream propagating solitons vanish as soon as they overtake and slow down the body. This phenomenon corresponds to the observations of [**114**], but cannot be seen in Figure 9. This is why we believe that constant-velocity models do not accurately predict the hysteretic aspect of the dead-water phenomenon, and that only a constant-force model would be able to recover this behavior (see also discussion in [**114**], section 1.3).

5. Overview of results and discussion

We have presented several nonlinear models describing the behavior of a system composed of two homogeneous fluids, when a body moves at the surface with constant velocity. Our asymptotic models are obtained through smallness assumptions on parameters of the system (Regime 1 and 2, see page 135). Contrarily to (linear) models existing in the literature, our models are valid for large (Regime 1) and moderate (Regime 2) amplitude waves. They are justified by consistency results (Proposition 3.1) or convergence results (Propositions 4.1 and 4.2).

These models duly recover the key features of the dead-water phenomenon, described in details in [**114**], namely:

i. transverse internal waves are generated at the rear of a body while moving at the surface;
ii. the body suffers from a positive drag when an internal elevation wave is located at its stern;
iii. this effect is strong only near critical Froude numbers;
iv. the maximum peak of the drag is reached at slightly subcritical values.

Moreover, we numerically computed our models using several ratios between the depth of the two layers of fluid. It appears that when the upper layer is thicker than the lower layer, the generated wave is able to reach larger amplitudes, and the wave resistance suffered by the body is stronger than in the opposite case (see figures 3, 4 and 9). This aspect have never been expressed before, as far as we know.

One important characteristic of the dead-water phenomenon, investigated in [**114**], is the hysteretic behavior of the system, and is not predicted by our models. This is due to the fact that our analysis assumes the velocity of the body to be constant, whereas a constant force was imposed to the body in their experimental setting. It would be of great interest to construct and justify models in the constant-force setting, in order to recover the complex behavior of the system in that configuration.

Finally, our work is limited to the two-dimensional configuration, that is horizontal dimension $d = 1$. Extending the study to the three-dimensional configuration would allow to observe divergent waves generated by the moving body, as well as the transverse ones [**88, 120**].

A. Derivation of the Green-Naghdi type model

In this section, we derive the strongly nonlinear model (2.7). This model only requires the shallow water assumption

$$\mu \ll 1,$$

and is a convenient intermediate step to construct the models of Regimes 1 and 2 used throughout the paper. The full Euler system is proved to be consistent with our model at order $\mathcal{O}(\mu^2)$, in Proposition A.1 below.

Let us first plug the expansions of Lemma 2.6 into (2.4), and drop $\mathcal{O}(\mu^2)$ terms. One can easily check that we obtain the approximate system

$$\text{(A.1)}\quad \begin{cases} -\alpha\,\mathrm{Fr}\,\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1 + \partial_x(h_1\partial_x\psi_1) + \partial_x(h_2\partial_x\psi_2) \;=\; \mu\partial_x\Big(\mathcal{R}_1(\partial_x\psi_1,\partial_x\psi_2)\Big), \\ (\partial_t - \mathrm{Fr}\,\partial_x)\zeta_2 + \partial_x(h_2\partial_x\psi_2) \;=\; \mu\partial_x\mathcal{T}[h_2,0]\partial_x\psi_2, \\ (\partial_t - \mathrm{Fr}\,\partial_x)\left(\partial_x\psi_2 - \gamma\partial_x\psi_1\right) + (\gamma+\delta)\partial_x\zeta_2 + \frac{\epsilon_2}{2}\partial_x\left(|\partial_x\psi_2|^2 - \gamma|\partial_x\psi_1|^2\right) \\ \qquad\qquad = \;\mu\Big(\gamma(\partial_t - \mathrm{Fr}\,\partial_x)\partial_x\mathcal{H}(\partial_x\psi_1,\partial_x\psi_2) + \epsilon_2\partial_x\mathcal{R}_2[\partial_x\psi_1,\partial_x\psi_2]\Big) \\ \qquad\qquad\qquad + \frac{1}{\mathrm{Bo}}\partial_x^2\left(\frac{\partial_x\zeta_2}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}}\right), \end{cases}$$

where we have used the following notations:

$$\begin{aligned} \mathcal{T}[h,b]V &\equiv -\frac{1}{3}\partial_x(h^3\partial_x V) + \frac{1}{2}\big(\partial_x(h^2(\partial_x b)V)) - h^2(\partial_x b)(\partial_x V)\big) + h(\partial_x b)^2 V, \\ \mathcal{N}[V_1,V_2] &\equiv \frac{((\partial_x h)V_2 - \partial_x(hV_2))^2 - \gamma((\partial_x h)V_1 - \partial_x(hV_2))^2}{2}, \\ \mathcal{H}(V_1,V_2) &\equiv h_1\Big(\partial_x(h_1V_1) + \partial_x(h_2V_2) - \frac{1}{2}h_1\partial_x V_1 - \partial_x(h_1+h_2)V_1\Big), \\ \mathcal{R}_1(V_1,V_2) &\equiv \mathcal{T}[h_1,h_2]V_1 + \mathcal{T}[h_2,0]V_2 - \frac{1}{2}\partial_x(h_1^2\partial_x(h_2V_2)) - h_1\partial_x h_2\partial_x(h_2V_2)), \\ \mathcal{R}_2[V_1,V_2] &\equiv \gamma V_1\partial_x\big(\mathcal{H}(V_1,V_2)\big) + \mathcal{N}[V_1,V_2]. \end{aligned}$$

Now, thanks to the fact that ζ_1 is a forced parameter of our problem, the system reduces to two evolution equations for (ζ_2, v), with v the *shear velocity* defined by

$$v \;\equiv\; \partial_x\left(\phi_2\Big|_{z=\epsilon_2\zeta_2} - \gamma\phi_1\Big|_{z=\epsilon_2\zeta_2}\right) \;=\; \partial_x\psi_2 - \gamma H(\psi_1,\psi_2).$$

From the last estimate of Lemma 2.6, one has immediately

$$\text{(A.2)}\qquad v \;=\; \partial_x\psi_2 - \gamma\partial_x\psi_1 - \mu\gamma\partial_x\mathcal{H}(\partial_x\psi_1,\partial_x\psi_2) + \mathcal{O}(\mu^2).$$

Now, one deduces from the first equation of (A.1), and (A.2), the following relations

$$\begin{aligned} h_1\partial_x\psi_1 \;+\; h_2\partial_x\psi_2 \;&=\; \alpha\,\mathrm{Fr}\,\zeta_1 \;+\; \mu\,\mathcal{R}_1(\partial_x\psi_1,\partial_x\psi_2) \;+\mathcal{O}(\mu^2), \\ v \;&=\; \partial_x\psi_2 - \gamma\partial_x\psi_1 \;-\; \mu\,\gamma\partial_x\mathcal{H}(\partial_x\psi_1,\partial_x\psi_2) \;+\mathcal{O}(\mu^2). \end{aligned} \qquad \text{(A.3)}$$

It follows that any linear operator defined above can be approximated as in the following example

$$\mathcal{H}(\partial_x\psi_1,\partial_x\psi_2) \;=\; \mathcal{H}\left(-\frac{h_2v - \alpha\,\mathrm{Fr}\,\zeta_1}{h_1+\gamma h_2}, \frac{h_1v + \gamma\alpha\,\mathrm{Fr}\,\zeta_1}{h_1+\gamma h_2}\right) + \mathcal{O}(\mu).$$

For the sake of readability, we do not precise the arguments in the following, and simply write

$$\mathcal{H} \equiv \mathcal{H}\left(-\frac{h_2 v - \alpha \operatorname{Fr} \zeta_1}{h_1+\gamma h_2}, \frac{h_1 v + \gamma\alpha \operatorname{Fr} \zeta_1}{h_1+\gamma h_2}\right),\ \mathcal{R}_1 \equiv \mathcal{R}_1\left(-\frac{h_2 v - \alpha \operatorname{Fr} \zeta_1}{h_1+\gamma h_2}, \frac{h_1 v + \gamma\alpha \operatorname{Fr} \zeta_1}{h_1+\gamma h_2}\right).$$

Using (A.3), one can approximate $\partial_x\psi_1$ and $\partial_x\psi_2$ at order $\mathcal{O}(\mu^2)$ with

$$\begin{aligned}(h_1+\gamma h_2)\partial_x\psi_1 &= -h_2 v + \alpha \operatorname{Fr}\zeta + \mu\left(\mathcal{R}_1 - \gamma h_2 \partial_x \mathcal{H}\right) + \mathcal{O}(\mu^2),\\ (h_1+\gamma h_2)\partial_x\psi_2 &= h_1 v + \gamma\alpha \operatorname{Fr}\zeta + \mu\gamma\left(\mathcal{R}_1 + h_1\partial_x\mathcal{H}\right) + \mathcal{O}(\mu^2).\end{aligned} \tag{A.4}$$

Using these formulae, the system (A.1) becomes (dropping $\mathcal{O}(\mu^2)$ terms)

$$\begin{cases} (\partial_t - \operatorname{Fr}\partial_x)\zeta_2 + \partial_x\left(\dfrac{h_2}{h_1+\gamma h_2}(h_1 v + \gamma\alpha\operatorname{Fr}\zeta)\right) + \mu\partial_x\mathcal{L} = 0,\\[2ex] (\partial_t - \operatorname{Fr}\partial_x)v + (\gamma+\delta)\partial_x\zeta_2 + \dfrac{\epsilon_2}{2}\partial_x\left(\dfrac{|h_1 v + \gamma\alpha\operatorname{Fr}\zeta|^2 - \gamma|h_2 v - \alpha\operatorname{Fr}\zeta|^2}{(h_1+\gamma h_2)^2}\right) + \mu\epsilon_2\partial_x\mathcal{Q}\\ \qquad\qquad = \dfrac{1}{\mathrm{Bo}}\partial_x^2\left(\dfrac{\partial_x\zeta_2}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}}\right),\end{cases} \tag{A.5}$$

where $\mathcal{L}$ and $\mathcal{Q}$ are defined by

$$\begin{aligned}\mathcal{L} &= \gamma\frac{h_2}{h_1+\gamma h_2}\left(\mathcal{R}_1 + h_1\partial_x\mathcal{H}\right) - \mathcal{T}[h_2,0]\left(\frac{h_1 v + \gamma\alpha\operatorname{Fr}\zeta_1}{h_1+\gamma h_2}\right),\\ \mathcal{Q} &= \gamma\frac{(h_1+h_2)v - (1-\gamma)\alpha\operatorname{Fr}\zeta_1}{(h_1+\gamma h_2)^2}\mathcal{R}_1 - \mathcal{R}_2\left[-\frac{h_2 v - \alpha\operatorname{Fr}\zeta_1}{h_1+\gamma h_2}, \frac{h_1 v + \gamma\alpha\operatorname{Fr}\zeta_1}{h_1+\gamma h_2}\right]\\ &\qquad + \gamma\frac{((h_1^2-\gamma h_2^2)v + \gamma(h_1+h_2)\alpha\operatorname{Fr}\zeta_1}{(h_1+\gamma h_2)^2}\partial_x\mathcal{H}.\end{aligned}$$

It is now simple to check that the system (A.5) is exactly the system (2.7).

This model is justified by the following proposition:

Proposition A.1. *The full Euler system* (2.4) *is consistent with* (A.5)*, at precision* $\mathcal{O}(\mu^2)$ *on* $[0,T]$*, with* $T>0$*.*

Proof. Let $U \equiv (\zeta_1,\zeta_2,\psi_1,\psi_2)$ be a solution of (2.4), bounded in $W^{1,\infty}([0,T];H^{s+t_0})$ with $s>1$ and $t_0 \geq 9/2+5$, and such that (2.5) is satisfied with $\zeta_1(t,x) \equiv \zeta_1(x)$. Using Lemma 2.6, it is clear that $(\zeta_1,\zeta_2,\partial_x\psi_1,\partial_x\psi_2)$ satisfies (A.1), up to $R_1 = (r_1,r_2,r_3,r_4)^T \in L^\infty([0,T];H^{s+5})^4$, satisfying (for $i=1\dots4$)

$$\left|r_i\right|_{W^{1,\infty}H^{s+5}} \leq \mu^2\, C_0\left(\frac{1}{h_{\min}}, \left|U\right|_{W^{1,\infty}H^{s+t_0}}\right).$$

Now, we set $v \equiv \partial_x\psi_2 - \gamma H(\psi_1,\psi_2)$, and v satisfies (A.2), up to $R_2 \in W^{1,\infty}([0,T];H^{s+5})$, with (again, thanks to Lemma 2.6)

$$\left|R_2\right|_{W^{1,\infty}H^{s+5}} \leq \mu^2\, C_0\left(\frac{1}{h_{\min}}, \left|U\right|_{W^{1,\infty}H^{s+t_0}}\right).$$

Therefore, since $\mathcal{R}_1$ and $\partial_x\mathcal{H}$ involve two spatial derivatives, one has

$$\begin{cases} (h_1+\gamma h_2)\partial_x\psi_1 &= -h_2 v + \alpha\operatorname{Fr}\zeta + \mu R_3,\\ (h_1+\gamma h_2)\partial_x\psi_2 &= h_1 v + \gamma\alpha\operatorname{Fr}\zeta + \mu R_4,\end{cases}$$

with $R_3,\ R_4 \in L^\infty([0,T];H^{s+3})$. Consequently, using the fact that $H^s(\mathbb{R})$ is an algebra for $s>1/2$ and that (2.5) is satisfied, we deduce that (A.4) is satisfied, up to $R_5 \in$

$L^\infty([0,T];H^{s+1})$, with

$$\big|R_5\big|_{L^\infty H^{s+1}} \ \leq \ \mu^2 \ \big(\big|R_3\big|_{L^\infty H^{s+3}} + \big|R_4\big|_{L^\infty H^{s+3}}\big) \ C_0\left(\frac{1}{h_{\min}}, \big|U\big|_{W^{1,\infty}H^{s+t_0}}\right).$$

Finally, when plugging (v,ζ_2) into (A.1), the residuals are clearly bounded by $\mu^2 C_0$, uniformly in $L^\infty([0,T];H^s)$, and the proposition is proved. □

B. Proof of Proposition 4.2

The sketch of the proof is the following. We will first prove the convergence between

$$U_{\rm app}(t,x) = U_0(\mu t, t, x) + \mu U_1(\mu t, t, x),$$

defined by (4.8)–(4.11), and the solutions of the symmetric Boussinesq-type system (4.3). This is obtained thanks to a consistency result, together with energy estimates. The convergence towards solutions of the full Euler system (2.4) follows then immediately from Proposition 4.1. Finally, the proposition is completed when remarking that the corrector term U_1 obeys to a sublinear growth. This result is a consequence of the following Lemma, that proceeds from Propositions 3.2 and 3.5 of [**67**]:

Lemma B.1. *Let u be the solution of*

$$\text{(B.1)} \qquad \left\{\begin{array}{l}(\partial_t + c\partial_x)u = g(v_1,v_2),\\ u\,|_{t=0} = 0,\end{array}\right. \quad \textit{with} \ \ \forall i\in\{1,2\}, \quad \left\{\begin{array}{l}(\partial_t + c_i\partial_x)v_i = 0,\\ v_i\,|_{t=0} = v_i^0,\end{array}\right.$$

with $c_1 \neq c_2$, v_1^0, $v_2^0 \in H^s(\mathbb{R})$, $s>1/2$, and g is a bilinear mapping defined on $\mathbb{R}^2$ and with values in $\mathbb{R}$. Then one has the following estimates:

i. *If $c = c_1$, then* $\lim\limits_{t\to\infty} \frac{1}{\sqrt{t}}\big|u(t,\cdot)\big|_{H^s(\mathbb{R})} = 0.$

ii. *If $c \neq c_1 \neq c_2$, then* $\frac{1}{\sqrt{t}}\big|u(t,\cdot)\big|_{H^s(\mathbb{R})} = \mathcal{O}(1).$

Moreover, if there exists $a>1/2$ such that $v_1^0(1+x^2)^a$, and $v_2^0(1+x^2)^a \in H^s(\mathbb{R})$, then one has the better estimate

$$\big|u\big|_{L^\infty H^s(\mathbb{R})} \ \leq \ C_0\big|v_1^0(1+x^2)^a\big|_{H^s(\mathbb{R})}\big|v_2^0(1+x^2)^a\big|_{H^s(\mathbb{R})},$$

with $C_0 = C(c,c_1,c_2)$.

We can now proceed with the proof.

Step 1: Well-posedness of $U_{app}(t,x)$. The global well-posedness of the forced Korteweg-de Vries equation is given by Bona and Zhang in [**18**].

The proof relies on an *a priori* estimate, that we recall here, as it will be useful for the following arguments. Let u be a solution of

$$\partial_t u \ + \ c\partial_x u \ + \ \lambda u\partial_x u \ + \ \nu\partial_x^3 u \ = \ b(x).$$

As we multiply the equation by $\Lambda^{2s'}u$ (with $s'>3/2$), and integrate with respect to the space variable, one obtains

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}}(\Lambda^{s'}u)^2\ {\rm d}x = \left|\lambda\int_{\mathbb{R}}\Lambda^{s'}(u\partial_x u)\Lambda^{s'}u\ {\rm d}x\right| + \left|\int_{\mathbb{R}}(\Lambda^{s'}b)(\Lambda^{s'}u)\ {\rm d}x\right|.$$

Thanks to the Kato-Ponce Lemma, we estimate the right-hand side as follows:

$$\left|\int_{\mathbb{R}}\Lambda^{s'}(u\partial_x u)\Lambda^{s'}u\ {\rm d}x\right| \leq \left|\frac{1}{2}\int_{\mathbb{R}}\partial_x u(\Lambda^{s'}u)^2\right| + \left|\int_{\mathbb{R}}[\Lambda^{s'},u]\partial_x u(\Lambda^{s'}u)\ {\rm d}x\right| \leq C_s\big|u\big|_{H^{s'}}^2\big|\partial_x u\big|_{L^\infty},$$

and the Cauchy-Schwarz inequality leads to

$$\frac{1}{2}\frac{d}{dt}\big|u\big|_{H^{s'}}^2 \ \leq \ C_0\lambda\big|u\big|_{H^{s'}}^2\big|\partial_x u\big|_{L^\infty} \ + \ \big|u\big|_{H^{s'}}\big|b\big|_{H^{s'}}.$$

It follows

$$\sup_{t\in[0,T]} \big|u\big|_{H^{s'}} \leq \exp\left(\lambda\int_0^T \big|\partial_x u(s,\cdot)\big|_{L^\infty}\,\mathrm{d}s\right)\big(\big|u\,|_{t=0}\,\big|_{H^{s'}} + T\big|b\big|_{H^{s'}}\big)\,.$$

In the particular case of (4.12), since $\alpha = \mathcal{O}(\mu)$, $\epsilon_2 = \mathcal{O}(\mu)$ and $\eta_\pm\,|_{t=0} = \frac{1}{2}\big(\zeta_2 \pm \frac{1}{\gamma+\delta}v\big)\,|_{t=0}$, it is straightforward that there exists $T = T(\gamma+\delta, \frac{1}{\gamma+\delta}) > 0$ such that for all $s' > 3/2$,

$$\sup_{t\in[0,T/\mu]} \big|u\big|_{H^{s'}} \leq C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s'}}, \big|V\,|_{t=0}\,\big|_{H^{s'}}, T\big). \tag{B.2}$$

We can then exhibit U_1. Indeed, (4.11) can be written in a simplified form as

$$(\partial_t + c_i\partial_x)\mathbf{e}_i\cdot S_0U_1 = \sum_{(j,k)\neq(i,i)} f_{ijk}(\tau,t,x) + \sum_{j\neq i}\partial_x g_{ij}(\tau,t,x), \tag{B.3}$$

with

$$f_{ijk}(\tau,t,x) \equiv \alpha_{ijk}u_j(\tau,x-c_jt)\partial_x u_k(\tau,x-c_kt) \quad \text{and} \quad g_{ij}(\tau,t,x) \equiv \beta_{ij}\partial_x^2 u_j(\tau,x-c_jt).$$

Thanks to estimate (B.2), one has for $s' > 1/2$,

$$\big|f_{ijk}\big|_{L^\infty([0,T]\times[0,T/\epsilon);H^{s'+1})} + \big|g_{ij}\big|_{L^\infty([0,T];H^{s'})} \leq C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s'+2}}, \big|V\,|_{t=0}\,\big|_{H^{s'+2}}, T\big).$$

with $C_0 = C_0(\frac{1}{\gamma+\delta}, \gamma+\delta)$. Hence, for $s' > 1/2$, one can set

$$\begin{aligned}\mathbf{e}_\pm\cdot S_0U_1(\tau,t,x) &= \sum_{(j,k)\neq(i,i)}\int_0^t f_{ijk}(\tau,s,x+c_i(s-t))\,\mathrm{d}s \\ &\quad + \sum_{j\neq i}\frac{1}{c_i-c_j}\big(g_{ij}(\tau,x-c_jt) - g_{ij}(\tau,x-c_it)\big) \\ &\equiv \sum_{(j,k)\neq(i,i)} U^{ijk} + \sum_{j\neq i} V^{ij}. \end{aligned}\tag{B.4}$$

One checks immediately that $U_1 \in L^\infty([0,T]\times[0,T/\mu);H^{s'})$ and $U_1\,|_{\tau=t=0} = 0$.

Step 2: Estimate on U_1. Let us estimate each term of the decomposition (B.4). Thanks to the uniform estimates of Step *1.* above, one has $g_{ij} \in L^\infty([0,T];H^{s'})$, so that it follows immediately

$$\forall j\neq i, \qquad \big|V^{ij}\big|_{L^\infty(\mathbb{R}^+\times[0,T];H^{s'})} \leq C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s'+2}}, \big|V\,|_{t=0}\,\big|_{H^{s'+2}}, T\big).$$

Moreover, for $j\neq i$, we remark that f_{ijj} can be written as

$$f_{ijj}(\tau,t,x) \equiv \alpha_{ijj}u_j(\tau,x-c_jt)\partial_x u_j(\tau,x-c_jt) \equiv \partial_x h_{ij}(\tau,x-c_jt),$$

so that U^{ijj} has actually the same form as V^{ij}, and can be treated in the same way. Since $f_{ijj}\in L^\infty([0,T]\times[0,T/\epsilon);H^{s'+1})$, U^{ijj} is bounded in $H^{s'}$ by $C_0(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s'}}, \big|V\,|_{t=0}\,\big|_{H^{s'}}, T)$, for all $j\neq i$.

Finally, for all $(j,k)\neq(i,i)$ with $j\neq k$, U^{ijk} satisfies the hypothesis of Lemma B.1, with $f_{ijk} = g(u_j,\partial_x u_k)$. Therefore, we deduce

$$\big|U_1\big|_{L^\infty(\mathbb{R}^+\times[0,T];H^{s'})} \leq \sqrt{t}C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s'+2}}, \big|V\,|_{t=0}\,\big|_{H^{s'+2}}, T\big). \tag{B.5}$$

As for the second estimate of the proposition, let us first remark that the estimates of V^{ij} and U^{ijj} are time-independent, and in agreement with the improved estimate. Therefore, the only remaining terms we have to control are U^{ijk} with $j\neq k$. Of course, we will

use the second case of Lemma B.1, but we have to check first that for every $\tau \in \mathbb{R}^+$, the initial data $u_j(\tau,0,x)$ and $\partial_x u_k(\tau,0,x)$ are localized in space, that is

$$\forall \tau \in \mathbb{R}^+, \qquad \big|(1+x^2)u_\pm(\tau,0,x)\big|_{H^{s'}} \ + \ \big|(1+x^2)\partial_x u_\pm(\tau,0,x)\big|_{H^{s'}} \ < \ \infty.$$

This property is true at $\tau = 0$ (by hypothesis of the proposition), and is propagated to $\tau > 0$, using the fact that $u_\pm(\tau, x_\pm)$ satisfies the KdV equation (4.10). This propagation of the localization in space has been proved for by Schneider and Wayne in [**104**, Lemma 6.4]. We do not recall the proof here, and make direct use of the statement:

Lemma B.2. *If $(1+x^2)V\,|_{t=0} \in H^{s'+2}$, then there exists $C_1, \tilde{C}_1 > 0$ such that*

$$\big|(1+x^2)u_\pm(\tau,0,x)\big|_{H^{s'+1}} \ \le \ C_1\big|(1+x^2)u_\pm\,|_{\tau=t=0}\,\big|_{H^{s'+2}} \ \le \ \tilde{C}_1\big|(1+x^2)V\,|_{t=0}\,\big|_{H^{s'+2}}.$$

This Lemma, together with the second estimate of Lemma B.1, allows to control uniformly $U^{ijk} \in L^\infty(\mathbb{R}^+ \times [0,T]; H^{s'})$. Every term of the decomposition (B.4) has been controlled, and one has the following estimate:

$$\big|U_1\big|_{L^\infty(\mathbb{R}^+\times[0,T];H^{s'})} \le \ C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s'+2}}, \big|V\,|_{t=0}\,\big|_{H^{s'+2}}, \big|(1+x^2)V\,|_{t=0}\,\big|_{H^{s'+2}}, T\big). \tag{B.6}$$

Step 3: Consistency results. We first prove the the consistency of the Boussinesq-type system (4.3) with our approximation. The precision is $\mathcal{O}(\mu^{3/2})$ in the general case, and $\mathcal{O}(\mu^2)$ in the second case. Here and in the following, we set $\alpha = \epsilon_2 = \mu$ in order to simplify the notations. The general case $\alpha = \mathcal{O}(\mu)$, $\epsilon_2 = \mathcal{O}(\mu)$ is obtained with slight obvious modifications.

Plugging $U_{\mathrm{app}}(t,x)$ into (4.6), we see from (4.7)-(4.11) that the only remaining term we have to control is $\mu^2 R(\mu t, t, x)$, with

$$\begin{aligned} R \ \equiv \ & \partial_\tau U_1 + \Sigma_1(U_0)\partial_x U_1 + \Sigma_1(U_1)\partial_x U_0 + S_1(U_0)\partial_t U_1 + S_1(U_1)\partial_t U_0 \\ & -\Sigma_2\partial_x^3 U_1 - S_2\partial_x^2\partial_t U_1 + \mu\Sigma_1(U_1)\partial_x U_1 + \mu S_1(U_1)\partial_x U_1, \end{aligned}$$

where $U_0(\mu t, t, x) \ = \ u_+(\mu t, x - c_+ t)\mathbf{e}_i \ + \ u_-(\mu t, x - c_- t)\mathbf{e}_-$. We bound each term of the right-hand side in the Sobolev H^s-norm, with $s > 1/2$ as in the proposition.

The estimate (B.5), with $s' = s + 3$, leads to

$$\big|\Sigma_2\partial_x^3 U_1\big|_{H^s} \le \sqrt{t}\ C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s+5}}, \big|V\,|_{t=0}\,\big|_{H^{s+5}}, T\big).$$

Then, from (4.11), one has

$$\mathbf{e}_i \cdot \partial_t S_0 U_1 = -c_i \mathbf{e}_i \cdot \partial_x S_0 U_1 + f_i$$

with $f_i \in L^\infty([0,T]\times[0,T/\mu); H^{s+2})$, and $\big|f_i\big|_{H^{s+2}} \le C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s+5}}, \big|V\,|_{t=0}\,\big|_{H^{s+5}}, T\big)$, so that

$$\big|S_2\partial_x^2\partial_t U_1\big|_{H^s} \le \sqrt{t}\ C_0\big|\partial_t U_1\big|_{H^{s+2}} \le \sqrt{T/\mu}C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s+5}}, \big|V\,|_{t=0}\,\big|_{H^{s+5}}, T\big).$$

One obtains in the same way the desired estimates for the following terms : $\Sigma_1(U_0)\partial_x U_1$, $\Sigma_1(U_1)\partial_x U_0$, $S_1(U_0)\partial_t U_1$, $S_1(U_1)\partial_t U_0$, $\Sigma_1(U_1)\partial_x U_1$ and $S_1(U_1)\partial_x U_1$.

Finally, in order to estimate $\partial_\tau U_1$, we differentiate (4.11) with respect to τ. Since u_i satisfies (4.10), one has $\partial_\tau u_i \in L^\infty([0,T]; H^{s+2})$. We are on the framework of Lemma B.1, so that we can obtain, as for (B.5), that $\partial_\tau U_1 \in L^\infty([0,T]; H^s)$, with

$$\big|\partial_\tau U_1\big|_{H^s} \le \sqrt{t}\ C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s+5}}, \big|V\,|_{t=0}\,\big|_{H^{s+5}}, T\big).$$

Hence, the residual $\mu^2 R$ is uniformly bounded in $L^\infty([0,T/\mu); H^s)$, and

$$\big|R\big|_{H^s} \le C_0\mu^{-1/2}C_0\big(\big|\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\big|_{H^{s+5}}, \big|V\,|_{t=0}\,\big|_{H^{s+5}}, T\big),$$

which gives the consistency at order $\mathcal{O}(\mu^{3/2})$.

The consistency at order $\mathcal{O}(\mu^2)$, for initial data sufficiently decreasing in space, is obtained in the same way, using the second estimate (B.6).

Step 4: Convergence results. The convergence is deduced from the consistency, as in the proof of Proposition 4.1. Indeed, when setting $R^\mu \equiv U_{\rm app} - V \equiv U_0(\mu t, t, x) + \mu U_1(\mu t, t, x) - V(t, x)$, one can check that R^μ satisfies the following equation:

$$\big(S_0 - \mu S_2\partial_x^2 + \mu S_1(U_{\rm app})\big)\partial_t R^\mu + \big(\Sigma_0 - \mu\Sigma_2\partial_x^2 + \mu\Sigma_1(U_{\rm app})\big)\partial_x R^\mu = \mu^{3/2} f + \mu\mathcal{A} + \mu\mathcal{B}, \tag{B.7}$$

with $\mathcal{A} = \partial_t S_1(U_{\rm app})R^\mu - S_1(R^\mu)\partial_t V$, $\mathcal{B} = \partial_x\Sigma_1(U_{\rm app})R^\mu - \Sigma_1(R^\mu)\partial_x V$ and the function f proved to be uniformly bounded in H^s, by the consistency result. We define the energy as

$$E_s(R^\mu) \equiv \frac{1}{2}(S_0\Lambda^s R^\mu, \Lambda^s R^\mu) + \frac{\mu}{2}(S_2\partial_x\Lambda^s R^\mu, \partial_x\Lambda^s R^\mu) + \frac{\mu}{2}(S_1(U_{\rm app})\Lambda^s R^\mu, \Lambda^s R^\mu),$$

as in the proof of Proposition 4.1. The same calculations lead to the following differential inequality:

$$\frac{d}{dt}E_s(R^\mu) \le C_0\mu E_s(R^\mu) + C_0\mu^{3/2}(E_s(R^\mu))^2,$$

with $C_0 = C_0(\big|\frac{\rm d}{{\rm d}x}\zeta_1\big|_{H^{s+5}}, \big|V\,|_{t=0}\,\big|_{H^{s+5}}, T)$. Then, Gronwall-Bihari's theorem allows to obtain $(E_s(R^\mu))^{1/2} \le C_0\mu^{1/2}(e^{C_0\mu t} - 1)$, and finally for $\mu t \le T(\frac{1}{\gamma+\delta}, \gamma + \delta)$,

$$\big|U_{\rm app} - V\big|_{H^s} \le C_0(E_s(R^\mu))^{1/2} \le \mu^{3/2} t\; C_0(\big|\frac{\rm d}{{\rm d}x}\zeta_1\big|_{H^{s+5}}, \big|V\,|_{t=0}\,\big|_{H^{s+5}}, T).$$

The first estimate of the proposition is now a direct consequence of (B.5), and

$$\big|\frac{\rm d}{{\rm d}x}\zeta_1(x)\big|_{H^{s+5}} + \big|V(t,x)\big|_{H^{s+5}} \le \big|U(t,x)\big|_{H^{s+t_0}},$$

thanks to an appropriate estimate of the operator $H(\psi_1, \psi_2)$; see [**40**, Proposition 2.7].

The second estimate of the proposition follows in the same way, using (B.6) and the consistency at order $\mathcal{O}(\mu^2)$.

C. Wave resistance and the dead-water phenomenon

Following Lamb [**66**] and Kostyukov [**63**], we assume that the drag experienced by ships is mostly due to the wave (making) resistance. This section is devoted to the analysis of this wave resistance. We first work with the variables *with dimension*, and deduce an explicit formula for the wave resistance R_W, depending on the flow, given as a solution of the full Euler equation (2.1). Accordingly with the nondimensionalization performed in Section 2.2, we introduce the dimensionless version of the wave resistance, that we call *wave resistance coefficient* and denote C_W. Finally, we derive simple approximations, for the two regimes considered throughout the paper.

The wave resistance acknowledges the energy required from the body to push the water away, and is defined by (see [**91, 88, 117**] and references therein)

$$R_W \;\equiv\; \int_{\Gamma_{\rm ship}} P\;(-\mathbf{e}_x\cdot\mathbf{n})\;{\rm d}S,$$

where $\Gamma_{\rm ship}$ is the exterior domain of the ship, P is the pressure, $\mathbf{e}_x$ is the horizontal unit vector and $\mathbf{n}$ the normal unit vector exterior to the ship. The pressure being constant ($\equiv P_\infty$) on the non-submerged part of the ship, one has

$$\begin{aligned} R_W &= \int_{\Gamma_{\rm ship}} (P - P_\infty)(-\mathbf{e}_x\cdot\mathbf{n})\;{\rm d}S \;=\; \int_{\mathbb{R}} \big(P\,\big|_{d_1+\zeta_1} - P_\infty\big)(-\partial_x\zeta_1)\;{\rm d}x \\ &= -\int_{\mathbb{R}} P\,\big|_{d_1+\zeta_1}\;\partial_x\zeta_1\;{\rm d}x. \end{aligned}$$

As a solution of the Bernoulli equation in (2.1), the pressure P is given in the upper domain by

$$\frac{P(x,z)}{\rho_1} \;=\; -\partial_t\phi_1(x,z) \;-\; \frac{1}{2}|\nabla_{x,z}\phi_1(x,z)|^2 \;-\; gz,$$

so that the wave resistance satisfies

$$\begin{aligned} R_W =\; & \rho_1 \int_{\mathbb{R}} g(d_1+\zeta_1)\partial_x\zeta_1 - \left(\partial_x\partial_t\phi_1 + \frac{1}{2}\partial_x\left(\partial_x\phi_1\right)^2 + \frac{1}{2}\partial_x\left(\partial_z\phi_1\right)^2\right)\Big|_{z=d_1+\zeta_1} \zeta_1 \,\mathrm{d}x. \\ \equiv\; & \int_{\mathbb{R}} F_W[\zeta_2,\phi_1]\,(d_1+\zeta_1)\,\zeta_1 \,\mathrm{d}x. \end{aligned}$$

Let us now construct the dimensionless version of this formula. Using the same change of variables as in Section 2.2, it is straightforward to obtain the dimensionless wave resistance (that we call *wave resistance coefficient*, and denote C_W):

$$\text{(C.1)} \qquad \begin{aligned} C_W \;\equiv\; \frac{\rho_1\lambda}{a_2c_0^2}\,R_W \;=\; & \int_{\mathbb{R}} \tilde{F}_W[\tilde{\phi}_1](\tilde{x})\tilde{\zeta}_1(\tilde{x})\,\mathrm{d}\tilde{x}, \quad \text{with} \\ -\tilde{F}_W[\tilde{\phi}_1] \;\equiv\; & (\partial_{\tilde{t}} - \mathrm{Fr}\,\partial_{\tilde{x}})\partial_{\tilde{x}}\tilde{\phi}_1 + \frac{\epsilon_2}{2}\partial_{\tilde{x}}\left(\left(\partial_{\tilde{x}}\tilde{\phi}_1\right)^2 + \frac{\epsilon_2}{2\mu}\left(\partial_{\tilde{z}}\tilde{\phi}_1\right)^2\right). \end{aligned}$$

Again, we omit the tildes in the following, for readability. Let us remark that by definition of the Dirichlet-Neumann operator G_1, and using (2.4), one has

$$\text{(C.2)} \quad -F_W[\phi_1] \;=\; (\partial_t - \mathrm{Fr}\,\partial_x)\partial_x\phi_1 + \frac{\epsilon_2}{2}\partial_x\left(\left(\partial_x\phi_1\right)^2 + \left(-\alpha\,\mathrm{Fr}\,\partial_x\zeta_1 + \epsilon_1\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\partial_x\phi_1\right)^2\right).$$

For practical purposes, we use approximations to compute the wave resistance coefficient, corresponding to the leading order term in the asymptotic expansion of C_W, for the different regimes at stake. These formulae use the variables of Sections 3 and 4, *i.e.* the surface and interface deviations ζ_1 and ζ_2 and the *shear velocity*, defined by

$$v \;\equiv\; \partial_x\left(\phi_2\Big|_{z=\epsilon_2\zeta_2} - \gamma\phi_1\Big|_{z=\epsilon_2\zeta_2}\right).$$

First of all, we use the following estimate, justified in [**40**]:

$$\phi_1(x,z) \;=\; \psi_1(x) \;+\; \mathcal{O}(\mu).$$

When combined with the first equation of (A.4):

$$\partial_x\psi_1 \;=\; \frac{-h_2v + \alpha\,\mathrm{Fr}\,\zeta_1}{h_1+\gamma h_2} \;+\mathcal{O}(\mu), \quad \text{with} \quad h_1 \equiv 1+\epsilon_1\zeta_1, \quad \text{and} \quad h_2 \equiv \frac{1}{\delta}+\epsilon_2\zeta_2,$$

one has immediately that (C.2) simplifies into

$$\begin{aligned} -F_W[\phi_1] = \; & (\partial_t - \mathrm{Fr}\,\partial_x)\left(\frac{-h_2v+\alpha\,\mathrm{Fr}\,\zeta_1}{h_1+\gamma h_2}\right) \\ & + \frac{\epsilon_2}{2}\partial_x\left(\frac{(-h_2v+\alpha\,\mathrm{Fr}\,\zeta_1)^2 + (\alpha\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1)^2\left((h_1+\gamma h_2)\,\mathrm{Fr} + \epsilon_2(h_2v-\alpha\,\mathrm{Fr}\,\zeta_1)\right)^2}{(h_1+\gamma h_2)^2}\right). \end{aligned} \qquad \text{(C.3)}$$

Now, $\partial_t h_2$, and $\partial_t v$ can be written using only ζ_1, ζ_2, v, and their spatial derivatives, since they satisfy system (A.5) up to $\mathcal{O}(\mu^2)$. Therefore, the leading order term of the wave resistance coefficient C_W can be deduced from the knowledge of the solution (ζ_2, v). We do not write explicitly this expression, as the models used in our simulations benefit from extra smallness assumptions, and simpler formulae are deduced in these cases.

Case of Regime 1

$$\mu \ll 1 ; \qquad \alpha \equiv \frac{\epsilon_1}{\epsilon_2} = \mathcal{O}(\mu) , \ 1-\gamma = \mathcal{O}(\mu).$$

The first immediate simplification in this regime is

$$h_1 + \gamma h_2 = h_1 + h_2 + \mathcal{O}(\mu) = 1 + \frac{1}{\delta} + \mathcal{O}(\mu).$$

Then, we use that (ζ_2, v) satisfies system (3.1) up to $\mathcal{O}(\mu^2)$ to deduce[3] (with $\bar{h}_1 \equiv 1+\frac{1}{\delta}-h_2$)

$$\begin{aligned}\partial_t(h_2 v) = & \ \mathrm{Fr}\,\partial_x(h_2 v) + (1+\delta)h_2\partial_x\zeta_2 + \epsilon_2\frac{\delta}{1+\delta}\left(v\partial_x\left(\bar{h}_1 h_2 v\right) + \frac{h_2}{2}\partial_x\left(\bar{h}_1 - h_2\right)v^2\right)\right) \\ = & \ \mathrm{Fr}\,\partial_x(h_2 v) + (1+\delta)h_2\partial_x h_2 + \epsilon_2\partial_x(h_2 v) + \frac{3\epsilon_2}{2}\frac{\delta}{1+\delta}\partial_x(h_2^2 v^2) + \mathcal{O}(\mu).\end{aligned}$$

Finally, using these approximations into (C.1) and (C.3), one has

$$\begin{aligned}C_W = & \ \frac{\delta}{1+\delta}\int_{\mathbb{R}}\left((1+\delta)h_2\partial_x\zeta_2 + \epsilon_2\partial_x(h_2 v) + \epsilon_2\frac{\delta}{1+\delta}\partial_x(h_2^2 v^2)\right)\zeta_1(x)\ \mathrm{d}x + \mathcal{O}(\mu^2) \\ = & \ -\int_{\mathbb{R}}\left(\left(\zeta_2 + \frac{\epsilon_2}{2}\zeta_2^2\right) + \epsilon_2\frac{\delta\ h_2 v}{1+\delta} + \epsilon_2\left(\frac{\delta\ h_2 v}{1+\delta}\right)^2\right)\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1(x)\ \mathrm{d}x + \mathcal{O}(\mu). \qquad \text{(C.4)}\end{aligned}$$

This is the formula used in Figures 3–5, in Section 3.2.

Case of Regime 2

$$\mu \ll 1 ; \qquad \epsilon_2 = \mathcal{O}(\mu) , \ \alpha = \mathcal{O}(\mu).$$

Most of the terms of (C.3) are now of size $\mathcal{O}(\mu)$. The first order of system (3.3) leads immediately to

$$\partial_t(h_2 v) - \mathrm{Fr}\,\partial_x(h_2 v) = -\frac{\gamma+\delta}{\delta}\partial_x\zeta_2 + \mathcal{O}(\mu),$$

so that we obtain simply

$$C_W = -\int_{\mathbb{R}}\zeta_2\frac{\mathrm{d}}{\mathrm{d}x}\zeta_1\ \mathrm{d}x + \mathcal{O}(\mu). \qquad \text{(C.5)}$$

This formula is used in Figures 6–9, in Section 4.3.

Remark C.1. *As we can see in* (C.4) *and* (C.5)*, the wave resistance coefficient will be small in the two following cases*

i. *The internal wave is small, or is not located below the ship,*
ii. *The internal wave is symmetric and centered at the location of the ship.*

The first case is obvious, and the last case reflects the fact that the integrands of (C.4) *and* (C.5) *are odd if* ζ_2*,* v *and* ζ_1 *are even.*

On the contrary, the ship will encounter a strong positive wave resistance if the functions ζ_2 *and* $h_2 v$ *are* decreasing at the location of the body. *This is the case when an internal wave of elevation is located just behind the ship. The dead-water effect is then explained in that way: the ship, traveling in a density-stratified water, generates internal waves of elevation in its trail, and therefore suffers from a severe wave resistance.*

[3] as we use $\mathrm{Bo}^{-1} = \mu^2$ in our simulations, we do not take here into account the surface tension term.

D. The Numerical schemes

We present in this section the numerical method used to obtains the figures of this study. For all of the simulations, we use a scheme based one the Crank-Nicholson method, and take care of the nonlinearities using a predictive step. This method has been introduced in **[13, 12]**, and used in the water wave framework in **[27, 42, 41]**. We give the general directions of such a method in the following, and then present the exact schemes we used for each of the models.

Time discretization. Denoting Δt the time step of the scheme, we approximate a function $u(x,t)$ at time $t = n\Delta t$ by $u(x, n\Delta t) \equiv u^n(x)$. Then, we approximate the time derivative by

$$\partial_t u \approx \frac{u^{n+1} - u^n}{\Delta t}$$

and any linear function of u by

$$F(u) \;\approx\; \frac{F(u^{n+1}) + F(u^n)}{2} \;=\; F\Big(\frac{u^{n+1} + u^n}{2}\Big).$$

Now we deal with the nonlinearities using a predictive step, $u^{n+\frac{1}{2}}$, defined by

$$\frac{u^{n+\frac{1}{2}} + u^{n-\frac{1}{2}}}{2} = u^n.$$

The first half-step, $u^{\frac{1}{2}}$, is computed through a simple explicit scheme.

With the help of the predictive step, the discretization of a quadratic nonlinearity $F(u,v)$ (F being linear with respect to both variables) is then a linear combination of the two possible discretizations, namely

$$F(u,v) \approx \alpha F\left(u^{n+\frac{1}{2}}, \frac{v^{n+1} + v^n}{2}\right) + (1-\alpha) F\left(\frac{u^{n+1} + u^n}{2}, v^{n+\frac{1}{2}}\right).$$

The parameter α can be chosen so that natural quantities are conserved.

Space discretization. With Δx the space step of the scheme, the functions are discretized spatially with a central difference. It is also useful to introduce a Lax-type averaging, by

$$u(x,t) \approx M(\beta)u \text{ with } (M(\beta)u^n)_i = (1-\beta)u_i^n + \frac{\beta}{2}(u_{i+1}^n + u_{i-1}^n).$$

The spatial derivatives are given by

$$\partial_x u \approx \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} \equiv (D_1 u^n)_i, \quad \partial_x^2 u \approx \frac{u_{i+1}^n - 2u_i^2 + u_{i-1}^n}{\Delta x^2} \equiv (D_2 u^n)_i, \quad \dots$$

We use periodic boundary conditions.

As we see in the following section, the parameters α and β can be set so that natural quantities are conserved. In the case of a Korteweg-de Vries equation, when choosing carefully the parameters, one can obtain a numerical scheme that preserves the discrete l^2-norm, in agreement with the solutions of the KdV equation having constant L^2-norms. As for the more complicated case of system (3.3), we set the parameters by analogy with the simple Korteweg-de-Vries equation.

D.1. The forced Korteweg-de-Vries equation. We want to solve numerically the following generic forced KdV equation

$$\partial_t u + c\partial_x u + \lambda u \partial_x u + \nu \partial_x^3 u = f(x). \tag{D.1}$$

When there is no forcing term ($f \equiv 0$), it is known that the KdV equation preserves the energy $E \equiv \left|u\right|^2_{L^2}$. The following scheme has been presented and studied (without the forcing term) in [**42**], and has the property to preserve the discrete energy when $f \equiv 0$.

$$\begin{aligned}&\frac{u_i^{n+1}-u_i^n}{\Delta t}+c\left(D_1\frac{u^{n+1}+u^n}{2}\right)_i+\frac{\lambda}{3}\left(\left(M(1)\frac{u_i^{n+1}+u_i^n}{2}\right)_i\left(D_1u^{n+\frac12}\right)_i\right.\\ &\left.+2\left(M(1/2)u^{n+\frac12}\right)_i\left(D_1\frac{u^{n+1}+u^n}{2}\right)_i\right)+\nu\left(D_3\frac{u^{n+1}+u^n}{2}\right)_i=f(i\Delta x).\end{aligned}\tag{D.2}$$

PROPOSITION D.1. *If $f \equiv 0$, then the scheme* (D.2) *preserves the discrete l^2-norm:*

$$\forall\ n\in\mathbb{N},\quad \left|u^n\right|^2_{l^2}\ \equiv\ \sum_i|u_i^n|^2\ =\ \sum_i|u_i^0|^2.$$

The proof is given in [**42**, Theorem 2].

D.2. The fully nonlinear model of Regime 1. The system we deal with now is (3.3), that we can write under the compact form

$$(\partial_t-\mathrm{Fr}\,\partial_x)\left(U-\mu\mathcal{R}_1[U]\right)+\partial_x\left((\mathcal{A}[U]+\epsilon_1\mathcal{B}(x))U\right)+\mu\epsilon_2\partial_x(\mathcal{R}_2[U]\cdot U)=b(x)+\frac{1}{\mathrm{Bo}}\partial_x^2\left(\mathcal{T}[U]\right),\tag{D.3}$$

with $U\ \equiv\ (\zeta_2,w)^T$, $b(x)\ \equiv\ -\alpha\,\mathrm{Fr}\,/(1+\delta)\ (\zeta_1(x),0)^T$, and

$$\mathcal{A}[U]\ \equiv\ \begin{pmatrix}0 & \frac{h_1h_2}{h_1+\gamma h_2}\\ \gamma+\delta & \epsilon_2\frac{h_1^2-\gamma h_2^2}{(h_1+\gamma h_2)^2}w\end{pmatrix},\qquad \mathcal{B}(x)\ \equiv\ \frac{\mathrm{Fr}}{h}\begin{pmatrix}\zeta_1(x) & 0\\ 0 & \zeta_1(x)\end{pmatrix},$$

$$\mathcal{R}_i[U]\ \equiv\ \begin{pmatrix}0\\ \mathcal{S}_i[h_2]w\end{pmatrix},\qquad \mathcal{T}[U]\ \equiv\ \begin{pmatrix}0\\ \frac{\partial_x\zeta_2}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}}\end{pmatrix}.$$

It is convenient to denote, with $h_1\equiv 1+\epsilon_1\zeta_1-\epsilon_2\zeta_2$ and $h_2\equiv\frac{1}{\delta}+\epsilon_2\zeta_2$,

$$f[\zeta_2]\ \equiv\ \frac{h_1h_2}{h_1+\gamma h_2},\qquad g[\zeta_2]\ \equiv\ \epsilon_2\frac{h_1^2-\gamma h_2^2}{(h_1+\gamma h_2)^2},\quad\text{and}\quad t[\zeta_2]\ \equiv\ \frac{\partial_x\zeta_2}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta_2|^2}}.$$

Advised by the above work on the KdV equation, we use the following time discretization:

$$f[\zeta]w\approx f[\zeta^{n+\frac12}]\frac{w^n+w^{n+1}}{2},\qquad g[\zeta]w^2\approx g[\zeta^{n+\frac12}]w^{n+\frac12}\frac{w^n+w^{n+1}}{2},$$

$$t[\zeta]\approx\frac{1}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta^{n+\frac12}|^2}}\partial_x\left(\frac{\zeta^{n+1}+\zeta^n}{2}\right),$$

$$\partial_t(\mathcal{S}_1[h]w)\approx\mathcal{S}_1[h^{n+\frac12}]\frac{w^{n+1}-w^n}{\Delta t}+\tilde{\mathcal{S}}_1[h^{n+\frac12},w^{n+\frac12}]\frac{\zeta^{n+1}-\zeta^n}{\Delta t},$$
$$\text{with}\quad \tilde{\mathcal{S}}_1[h,w]\zeta\ \equiv\ \frac{\epsilon_2}{3}\partial_x^2(\zeta(1+1/\delta-2h)w)-2\epsilon_2(\partial_x\zeta)(\partial_xh)w,$$

$$w\ \mathcal{S}_2[H]w\approx\frac13w^{n+\frac12}\mathcal{S}_2[H^{n+\frac12}]\left(\frac{w^n+w^{n+1}}{2}\right)+\frac23\frac{w^n+w^{n+1}}{2}\mathcal{S}_2[H^{n+\frac12}]w^{n+\frac12}.$$

Finally, after the space discretization, this leads to the following scheme (with $h^n \equiv \frac{1}{\delta} + \epsilon_2\zeta^n$)

(D.4)

$$\frac{\zeta_i^{n+1}-\zeta_i^n}{\Delta t} - \mathrm{Fr}\left(D_1\frac{\zeta^{n+1}+\zeta^n}{2}\right)_i + \alpha\frac{\mathrm{Fr}\,\delta}{1+\delta}\left(D_1\left(\left(\frac{1}{\delta}+\epsilon_2\frac{\zeta^{n+1}+\zeta^n}{2}\right)\zeta_1(i\Delta x)\right)\right)_i + \frac{1}{3}D_1\left(M(1)f[h^{n+\frac{1}{2}}]\frac{w^n+w^{n+1}}{2} + f[h^{n+\frac{1}{2}}]M(1/2)(w^n+w^{n+1})\right)_i = 0,$$

(D.5)

$$\begin{aligned}&\frac{(I-\mu\mathcal{S}_1[h^{n+\frac{1}{2}}])(w^{n+1}-w^n)-\mu\tilde{\mathcal{S}}_1[h^{n+\frac{1}{2}},w^{n+\frac{1}{2}}](\zeta^{n+1}-\zeta^n)}{\Delta t} + (\gamma+\delta)\left(D_1\frac{\zeta^n+\zeta^{n+1}}{2}\right)_i\\ &- \mathrm{Fr}\left(D_1(I-\mu\mathcal{S}_1[h^{n+\frac{1}{2}}])\frac{w^{n+1}+w^n}{2}\right)_i + \alpha\frac{\mathrm{Fr}\,\delta}{1+\delta}\left(D_1\left(\frac{w^{n+1}+w^n}{2}\zeta_1(i\Delta x)\right)\right)_i\\ &+ \frac{\epsilon_2}{3}D_1\left(M(1)g[h^{n+\frac{1}{2}}]w^{n+\frac{1}{2}}\frac{w^n+w^{n+1}}{2} + g[h^{n+\frac{1}{2}}]w^{n+\frac{1}{2}}M(1/2)(w^n+w^{n+1})\right)_i\\ &+ \frac{\mu\epsilon_2}{3}D_1\left(w^{n+\frac{1}{2}}\mathcal{S}_2[H^{n+\frac{1}{2}}]\left(\frac{w^n+w^{n+1}}{2}\right) + (w^n+w^{n+1})\mathcal{S}_2[H^{n+\frac{1}{2}}]w^{n+\frac{1}{2}}\right)_i\\ &= \frac{1}{\mathrm{Bo}}\left(D_2\frac{1}{\sqrt{1+\mu\epsilon_2^2|\partial_x\zeta^{n+\frac{1}{2}}|^2}}D_1\left(\frac{\zeta^{n+1}+\zeta^n}{2}\right)\right)_i.\end{aligned}$$

D.3. Validation of the method. In order to validate the proposed schemes, we use known explicit solutions of the forced Korteweg-de Vries equation (4.13). The first test function we use is the travelling wave

$$u_0 \equiv -\operatorname{sech}^2(k(x-c_{sw}t)), \tag{D.6}$$

with $c_{sw} = c + 4\nu k^2,\ k = \sqrt{-\lambda/(12\nu)}$, corresponding to the classical case where there is no forcing.

The second test function is the steady solution

$$u_1 \equiv \pm\operatorname{sech}^2(kx), \tag{D.7}$$

with $k^2 = \frac{\lambda}{12\nu}$ and the corresponding forcing

$$f(x) = \left(c+\frac{\lambda}{3}\right)\operatorname{sech}^2(kx).$$

These two functions are exact solutions of the forced KdV equations

$$\partial_t u + c\partial_x u + \lambda u\partial_x u + \nu\partial_x^3 u = \frac{\mathrm{d}}{\mathrm{d}x}f(x).$$

In the absence of non-trivial solutions of system (3.3), we use the same functions as reference. After an appropriate change of variables, u_0 and u_1 will satisfy system (3.3) up to small terms. Therefore, adding the corresponding forcing term to the equations, u_0 and u_1 are exact solutions of the modified system, and we are able to compare the results of our numerical schemes with the theoretical solution.

These comparisons are presented in Tables 2 and 3. The error are givens in term of the normalized l^2 norm. The results exhibit a convergence behavior of order $\mathcal{O}((\Delta x)^2+(\Delta t)^2)$.

The test functions (D.6) and (D.7) are very smooth, and our problem can lead to rapid variation in the solution, as we can see especially in figures 6 and 7. In order to check the accuracy of our method in such cases, we compute the KdV scheme with different space and time steps, and compare the outcome with the result of a more refined grid.

$\Delta x = \Delta t$	L	T	KdV scheme	fully nonlinear scheme
0.1	20	10	$8.1530\ 10^{-4}$	$4.9498\ 10^{-4}$
0.05	20	10	$2.0393\ 10^{-4}$	$1.5154\ 10^{-4}$
0.01	20	10	$8.1604\ 10^{-6}$	$1.8363\ 10^{-5}$

TABLE 2. Numerical errors of the KdV and Green-Naghdi schemes for the initial data (D.6). We choose the parameters $\mu = \epsilon_2 = \alpha = 0.1$, $\gamma = 0.9$, $\delta = 5/12$, Fr $= 1.1$, Bo $= 100$.

$\Delta x = \Delta t$	L	T	KdV scheme	fully nonlinear scheme
0.1	20	10	$4.7397\ 10^{-4}$	$1.7844\ 10^{-4}$
0.05	20	10	$1.1850\ 10^{-4}$	$4.4557\ 10^{-5}$
0.01	20	10	$4.7409\ 10^{-6}$	$8.1604\ 10^{-6}$

TABLE 3. Numerical errors of the KdV and Green-Naghdi schemes for the initial data (D.7). We choose the parameters $\mu = \epsilon_2 = \alpha = 0.1$, $\gamma = 0.9$, $\delta = 5/12$, Fr $= 1.1$, Bo $= 100$.

More precisely, we compute the scheme for $\Delta x = \Delta t = 0.1,\ 0.05,\ 0.01$, and measure the difference with a reference solution obtained with $\Delta x = \Delta t = 0.001$. The error is then discrete l^2 norm of the difference, normalized with the l^2 norm of the reference solution.

These comparisons, using the settings of figures 6 and 7, are presented in Table 4. The normalized error can be relatively large; this is due to the fact that the reference solution is very small (of order 10^{-2}). However, one clearly sees that the schemes are stable, and converge fast when space and time steps become small. Moreover, the precision for $\Delta x = \Delta t = 0.01$ is sufficient to validate our results.

$\Delta x = \Delta t$	Figure 6 (a)	Figure 6 (b)	Figure 7 (a)	Figure 7 (b)
0.1	$7.2320\ 10^{-1}$	1.2158	$2.4538\ 10^{-1}$	$2.0070\ 10^{-1}$
0.05	$1.9339\ 10^{-1}$	$5.1843\ 10^{-1}$	$9.6817\ 10^{-2}$	$1.0586\ 10^{-1}$
0.01	$9.3586\ 10^{-3}$	$5.2038\ 10^{-2}$	$1.9121\ 10^{-2}$	$6.0949\ 10^{-2}$

TABLE 4. Convergence of the KdV scheme, with the settings of figures 6 and 7.

Partie B. Propogation des ondes à travers un milieu non homogène

CHAPITRE 5

Wave operator bounds for 1-dimensional Schrödinger operators with singular potentials and applications

Article publié dans *Journal of Mathematical Physics* **52** (2011)
en collaboration avec Jeremy L. Marzuola et Michael I. Weinstein.

Sommaire

Abstract

Boundedness of wave operators for Schrödinger operators in one space dimension for a class of singular potentials, admitting finitely many Dirac delta distributions, is proved. Applications are presented to, for example, dispersive estimates and commutator bounds.

1. Introduction

Wave operators provide a means for converting operator bounds for a "free" dynamics generated by a constant coefficient Hamiltonian, $H_0 = -\Delta$ to analogous operator bounds about "interacting" dynamics associated with a variable coefficient Hamiltonian, $H = -\Delta + V$, on its continuous spectral subspace. Indeed let $W_\pm$ and $W_\pm^*$ denote wave operators associated with the free and interacting Hamiltonians H_0 and H (defined by (2.1) and (2.2)). Then we have

$$W_\pm W_\pm^* = P_c, \quad W_\pm^* W_\pm = Id \tag{1.1}$$

$$f(H)P_c = W_\pm f(H_0)W_\pm^*,\ f(H_0) = W_\pm^* f(H) W_\pm, \quad f \text{ Borel on } \mathbb{R}\ . \tag{1.2}$$

It follows that bounds on $f(H)P_c$ acting between $W^{k_1,p_1}(\mathbb{R}^d)$ and $W^{k_2,p_2}(\mathbb{R}^d)$ can be derived from bounds on $f(H_0)$ between these spaces if the wave operators $W_\pm$ are bounded between $W^{k_1,p_1}(\mathbb{R}^d)$ and $W^{k_2,p_2}(\mathbb{R}^d)$ for $k_j \geq 0$ and $p \geq 1$. Here, $W^{k,p}(\mathbb{R}^d),\ k \geq 1,\ p \geq 1$ denotes the Sobolev space of functions having derivatives up to order k in $L^p(\mathbb{R}^d)$.

Boundedness of wave operators in $W^{k,p}(\mathbb{R}^d)$, under smoothness and decay assumptions on $V(x)$ was proved by Yajima [**181**] in dimensions $d \geq 2$. Weder [**178**] proved boundedness in dimension one; see also the article of D'Ancona and Fanelli [**134**]. In [**178**] it is assumed that $V \in L^1_\gamma(\mathbb{R})$, the space of all complex-valued measurable functions ϕ defined on $\mathbb{R}$ such that

$$\|\phi\|_{L^1_\gamma} = \int |\phi(x)|(1+|x|)^\gamma \mathrm{d}x < \infty. \tag{1.3}$$

For V in a class of generic potentials, the assumption is $\gamma > 3/2$, and otherwise it is assumed $\gamma > 5/2$. Wave operator bounds can be used to establish dispersive estimates, namely

$$\begin{aligned} \left\|e^{-iHt}P_c(H)f\right\|_{L^p(\mathbb{R}^d)} &= \left\|W_\pm e^{-iH_0t}W_\pm^* f\right\|_{L^p(\mathbb{R}^d)} \\ &\leq C\ |t|^{-\frac{d}{2}-\frac{d}{p}}\ \|f\|_{L^q(\mathbb{R}^d)}, \qquad p^{-1}+q^{-1}=1,\ \ p \geq 1. \end{aligned}$$

Applications of wave operator bounds for singular potentials appear in [**162, 137, 152**]. Schrödinger operators with singular potentials arise in several mathematical models, which have recently been extensively investigated. For example, see [**145, 152, 148, 149, 140, 158, 139, 162**], where Dirac delta function potentials are considered. Boundedness of wave operators in $W^{1,2}(\mathbb{R})$ for singular potentials is used implicitly in references [**152**] and [**145**], but this property appears not to have been addressed previously. This gap in the literature is addressed in the present work. Another motivation for the present work is the study of scattering for highly oscillatory potentials, containing local singularities, in the homogenization limit [**137**]. In this work, bounds on $(m^2+H)^{-1}P_c(H)(m^2-\partial_x^2)$, where $H=-\partial_x^2+V(x)$ is a Schrödinger operator with a singular (distribution) part to the potential $V(x)$, are required; see section 8.

This article is devoted to an extension of the one-dimensional results [**178**] to the case of singular potentials. Specifically, our results apply to Hamiltonians of the form

$$H \ = \ -\partial_x^2 \ + \ V(x),$$

where $V(x)$ satisfies:

Hypotheses (V)

$$V(x) \ = \ V_{sing}(x) \ + \ V_{reg}(x), \tag{1.4}$$

$$V_{sing}(x) \ = \ \sum_{j=0}^{N-1} q_j\ \delta(x-y_j), \quad q_j, y_j \in \mathbb{R}, \quad y_j < y_{j+1}, \quad q_j \neq 0, \tag{1.5}$$

$$\|V_{reg}\|_{L^1_{\frac{3}{2}+}(\mathbb{R})} \ \equiv \ \int_{\mathbb{R}} (1+|s|)^{\frac{3}{2}+}|V_{reg}(s)|\ ds \ < \ \infty. \tag{1.6}$$

The chapter is structured as follows. In section 2 we state our main result, Theorem 2.1, concerning boundedness of wave operators. In section 3 the strategy of proof is outlined. Section 4 summarizes facts about Jost solutions, distorted plane waves, reflection and transmission coefficients, *etc.* In section 5 we state a general result, Theorem 5.1, from which Theorem 2.1 follows. The proof of Theorem 5.1 is given in section 6, and the completion of Theorem 2.1 in section 7. Finally, in section 8 we present examples (multi-delta function potentials) and applications to dispersive estimates, commutator bounds and well posedness.

2. Main results

We first define and review properties of the wave operators. For basic results on wave operators see, for example, **[125, 168, 170]**.

Introduce the self-adjoint operators $H_0 = -\Delta$ and $H = -\Delta + V$. Here, V is a real-valued potential, satisfying assumptions given below; see Section 5. Let $P_c = P_c(H)$ denote the continuous spectral projection associated with H. The wave operators, $W_\pm$ and their adjoints $W_\pm^*$ are defined by

$$W_\pm \equiv s - \lim_{t\to\pm\infty} e^{itH}e^{-itH_0} \tag{2.1}$$

$$W_\pm^* \equiv s - \lim_{t\to\pm\infty} e^{itH_0}e^{-itH}P_c. \tag{2.2}$$

The wave operators satisfy the properties (1.1) and (1.2). The notion of wave operators is intimately related to the idea of distorted Fourier bases, which are discussed in detail in **[125]**, **[150]**, **[167]**. In one dimension, this is directly related to the Jost solutions, studied for general self-adjoint Schrödinger operators in **[167, 135]** and for a certain class of non-self-adjoint operators in **[157]**.

Theorem 5.1 of section 5, combined with the calculations of section 7, implies the following:

THEOREM 2.1. *Consider the Schrödinger operator with a potential, $V(x)$, satisfying* **Hypotheses (V)**. *Then $W_\pm$ and $W_\pm^*$ originally defined on $W^{1,p}\cap L^2$, $1\le p\le\infty$, have extensions to bounded operators on $W^{1,p}$, $1<p<\infty$. Moreover, there are constants C_p such that:*

$$\begin{aligned}\|W_\pm f\|_{W^{1,p}(\mathbb{R})} &\le C_p\|f\|_{W^{1,p}(\mathbb{R})},\\ \|W_\pm^* f\|_{W^{1,p}(\mathbb{R})} &\le C_p\|f\|_{W^{1,p}(\mathbb{R})}, \qquad f\in W^{1,p}(\mathbb{R}),\ 1<p<\infty.\end{aligned}$$

REMARK 2.2. *In general, the wave operators are not bounded on L^1. The constraint $p>1$ is due to the Hilbert transform, $\mathcal{H}$ not being bounded on L^1; see* **[178]**.

3. Strategy of Proof

We use the approach for wave operators on $\mathbb{R}$ initiated by Weder in **[178]**. The heart of the matter concerns the detailed low and high frequency behavior of Jost solutions, worked out by Deift and Trubowitz **[135]**, or a consequence of their methods. The idea is to split the wave operators into high and low frequency components:

$$W_\pm = W_{\pm,high} + W_{\pm,low}.$$

For the high frequency component we prove for $\phi\in\mathcal{S}$,

$$W_{\pm,high}\phi = \sum_j S_{A_j}\phi, \quad \text{where} \quad S_A\phi \equiv \int_{-\infty}^{\infty} A(x,y)\phi(y)\mathrm{d}y.$$

For each $A = A_j$, we use the criterion (Young's inequality **[138]**) for L^p, $1\le p\le\infty$ boundedness:

$$\begin{aligned}C_A \equiv \sup_{x\in\mathbb{R}}\int_{\mathbb{R}}|A(x,y)|\,\mathrm{d}y + \sup_{y\in\mathbb{R}}\int_{\mathbb{R}}|A(x,y)|\,\mathrm{d}x < \infty\\ \Longrightarrow \|S_A\phi\|_{L^p} \le C_A\,\|\phi\|_{L^p}.\end{aligned}$$

to prove

$$\|W_{\pm,high}\phi\|_{W^{1,p}} \le C_p\,\|\phi\|_{W^{1,p}}, \quad 1<p<\infty, \quad k\ge 0. \tag{3.1}$$

For the low frequency components, we have

$$W_{\pm,low} \sim \mathcal{H} \ + \ \sum_j S_{A_j},$$

where S_{A_j} is as above and $\mathcal{H}$ denotes the Hilbert Transform

$$(\mathcal{H}\phi)(x) \ = \ \frac{1}{\pi} \text{ P.V. } \int \frac{\phi(x-y)}{y} \mathrm{d}y \ = \ \int_{-\infty}^{\infty} e^{ikx} \ (-i \ \mathrm{sgn}(k)) \ \hat{\phi}(k) \mathrm{d}k \tag{3.2}$$

Here, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier Transform on $\mathbb{R}$ and its inverse, defined by

$$\hat{\phi}(k) \ \equiv \ \mathcal{F}\phi(k) = \frac{1}{2\pi} \int e^{-ikx} \phi(x) \mathrm{d}x, \quad \check{\Phi}(x) \ \equiv \ \mathcal{F}^{-1}\Phi(x) = \int e^{ikx} \Phi(k) \mathrm{d}k. \tag{3.3}$$

Thus, for low frequencies, boundedness

$$\|W_{\pm,low}\phi\|_{W^{1,p}} \ \leq \ C_p \ \|\phi\|_{W^{1,p}}, \ \ 1 < p < \infty. \tag{3.4}$$

reduces to the boundedness properties of the Hilbert transform [**174**]:

Theorem 3.1. *$\mathcal{H} : W^{s,p} \to W^{s,p}$, for $1 < p < \infty$ and $s \geq 0$, with $\|\mathcal{H}\phi\|_{W^{s,p}(\mathbb{R})} \ \leq K_p \ \|\phi\|_{W^{s,p}(\mathbb{R})}$.*

Estimates (3.1) and (3.4) then imply the theorem. The proof of (3.1) and (3.4) is given in section 6. We now develop some background for implementing the strategy.

4. Background spectral theory of $H = -\partial_x^2 + V$

4.1. Distorted plane waves, $e_\pm(x;k)$. Consider the operator $H = -\partial_x^2 + V(x)$, defined as a self-adjoint operator on $L^2(\mathbb{R})$. Denote by P_d and P_c the discrete and continuous spectrum projections. P_d and P_c are orthogonal projections with $P_c = Id - P_d$.

Denote by R_0 the outgoing "free" resolvent operator $R_0(k) = (-\partial_x^2 - k^2)^{-1}$ with kernel

$$R_0(k)(x,y) = -(2ik)^{-1} \exp(ik|x-y|)$$

and finally introduce the *distorted plane waves*, $e_\pm(x;k)$:

Definition 4.1. *The functions $u = e_\pm(x;k)$ are the unique solutions to $(H - k^2)u = 0$ satisfying*

$$e_\pm(x;k) \ = \ e^{\pm ikx} \ + \ \mathrm{outgoing}(x), \tag{4.1}$$

where a function U is said to be outgoing as $|x| \to \infty$ if

$$(\ \partial_x \ \mp \ ik \ )U \to \ 0, \ x \to \pm\infty.$$

Thus, $e_\pm(x;k)$ is given by the integral equation:

$$e_\pm(x;k) = e^{\pm ikx} - R_0(k) V e_\pm(x;k) \tag{4.2}$$

or equivalently

$$e_\pm(x;k) = e^{\pm ikx} - R_V(k) V e^{\pm ikx}. \tag{4.3}$$

The continuous spectral projection, P_c, is given by

$$P_c f(x) = \frac{1}{2\pi} \int \int_0^\infty \left(e_+(x,k) \ \overline{e_+(y,k)} + e_-(x,k) \ \overline{e_-(y,k)} \right) f(y) \mathrm{d}k \mathrm{d}y. \tag{4.4}$$

see, for example, [**176**].

We write

$$P_c f \ \equiv \ F_+^* \ F_+ \ f, \ \text{ where it follows from (4.4) that}$$

$$F_+ f \ \equiv \ \int_{\mathbb{R}} \overline{\Psi_+(y,k)} \ f(y) \ \mathrm{d}y, \quad F_+^* f \ \equiv \ \int_{\mathbb{R}} \Psi_+(y,k) \ f(k) \ \mathrm{d}k \ \text{ and} \tag{4.5}$$

$$\Psi_+(y,k) = \frac{1}{\sqrt{2\pi}} \begin{cases} e_+(x;k) & k \geq 0, \\ e_-(x;-k) & k < 0. \end{cases} \tag{4.6}$$

We also define $\Psi_-(x,k) = \overline{\Psi_+(x,-k)}$. Recall $W_\pm = F_\pm^* \mathcal{F}$ **[170]**.

4.2. Jost solutions. To make direct use of the arguments in **[178]** and **[135]**, we express the results of the preceding subsection in terms of *Jost solutions*, commonly introduced for one-dimensional Schrödinger operators.

Given the Schrödinger equation

$$-\frac{\mathrm{d}^2}{\mathrm{d}x^2}u + Vu = k^2 u, \ k \in \mathbb{C}, \tag{4.7}$$

we define the Jost solutions, $f_j(x,k)$, $j = 1,2$, $\mathrm{Im}k \geq 0$, to be the unique solutions of (4.7) satisfying the conditions:

$$\begin{aligned} f_1(x,k) \ &- \ e^{ikx} \to 0, \quad x \to \infty, \ \text{ and} \\ f_2(x,k) \ &- \ e^{-ikx} \to 0, \quad x \to -\infty. \end{aligned} \tag{4.8}$$

The Jost solutions are linearly independent solutions of (4.7) for $k \neq 0$. Therefore, there are unique functions $T(k)$, $R_j(k)$, $j = 1,2$ such that for $k \in \mathbb{R} \setminus 0$

$$f_2(x,k) = \frac{R_1(k)}{T(k)} f_1(x,k) + \frac{1}{T(k)} f_1(x,-k), \tag{4.9}$$

$$f_1(x,k) = \frac{R_2(k)}{T(k)} f_2(x,k) + \frac{1}{T(k)} f_2(x,-k) \tag{4.10}$$

For a potential, V, with compact support within $(-r,r)$, $R_j(k)$ and $T(k)$ are defined via the solutions:

$$e_+(x;k) = t(k) f_1(x;k) = \begin{cases} e^{ikx} + R_2(k)e^{-ikx}, & x < -r, \\ T(k)e^{ikx}, & x > r \end{cases} \tag{4.11}$$

$$e_-(x;k) = t(k) f_2(x;k) = \begin{cases} T(k)e^{-ikx}, & x < -r, \\ e^{-ikx} + R_1(k)e^{ikx}, & x > r. \end{cases} \tag{4.12}$$

Generically,

$$T(k) = \alpha k + o(k), \quad 1 + R_j(k) = \alpha_j k + o(k), \quad j = 1,2, \quad k \to 0. \tag{4.13}$$

$T(k)$ is called the transmission coefficient associated with H. $R_1(k)$ is the right to left reflection coefficient, and $R_2(k)$ the left to right reflection coefficient.

Finally, it is convenient to denote by $m_j(x,k)$, $j = 1,2$

$$m_1(x,k) = e^{-ikx} f_1(x,k), \qquad \text{and} \qquad m_2(x,k) = e^{ikx} f_2(x,k). \tag{4.14}$$

It follows from (4.1), (4.8) and (4.9) that

$$\Psi_+(x,k) = \frac{1}{\sqrt{2\pi}} \begin{cases} T(k)\, e^{ikx}\, m_1(x,k) & k \geq 0, \\ T(-k)\, e^{ikx}\, m_2(x,-k) & k < 0, \end{cases} \tag{4.15}$$

where $m_1(x,k) - 1 \to 0$ as $x \to \infty$ and $m_2(x,k) - 1 \to 0$ as $x \to -\infty$. The detailed smoothness and decay properties, in x and k, of $m_j(x;k) - 1$ are required in estimates. These are given in section 7.

5. Statement of the Central Theorem

Our central result, from which Theorem 2.1 follows, is:

Theorem 5.1. *Let $H = -\partial_x^2 + V(x)$ be self-adjoint on $L^2(\mathbb{R})$ for which the transmission and reflection coefficients (see (4.9)) satisfy the bounds:*

$$|R_1(k)| \ + \ |R_2(k)| \ + \ |T(k)-1| \quad \leq \quad \frac{C}{\langle k \rangle}, \tag{5.1}$$

$$|\partial_k R_1(k)| \ + \ |\partial_k R_2(k)| \ + \ |\partial_k T(k)| \quad = \quad \mathcal{O}\left(\frac{1}{|k|}\right), \ \ |k| \to \infty. \tag{5.2}$$

Let S_1 and S_2 be defined by

$$(S_j \Phi)(x) \ \equiv \ \int_{\mathbb{R}} R_j(x,y) \ \Phi(y) \ dy, \quad \text{where} \tag{5.3}$$

$$R_j(x,y) \ \equiv \ \int_{\mathbb{R}} e^{ikx} \left(m_j(x,k) - 1\right) e^{-iky} \ dk. \tag{5.4}$$

and assume, for $1 < p < \infty$, that S_1 is bounded on $W^{1,p}(\mathbb{R}_+)$ and S_2 is bounded on $W^{1,p}(\mathbb{R}_-)$.

Then $W_\pm$ and $W_\pm^*$ originally defined on $W^{1,p} \cap L^2$, $1 \leq p \leq \infty$, extend to bounded operators on $W^{1,p}$, $1 < p < \infty$. Furthermore, there are constants C_p such that:

$$\|W_\pm f\|_{W^{1,p}} \leq C_p \|f\|_{W^{1,p}}, \ \|W_\pm^* f\|_{W^{1,p}} \leq C_p \|f\|_{W^{1,p}}, \ f \in W^{1,p} \cap L^2, \ 1 < p < \infty. \tag{5.5}$$

Remark 5.2. *Deift and Trubowitz* **[135]** *establish the hypotheses of the Theorem for any potential $V(x)$, for which $(1+|x|)^{\frac{3}{2}+} |V(x)| \in L^1(\mathbb{R})$ (see section 7). We show in section 7.2 that their proof also applies to a potential of the type in Hypothesis* **(V)**, *$V = V_{sing} + V_{reg}$, where V_{sing} has a finite set of Dirac masses within an interval $(-A, A)$, and such that $(1+|x|)^{\frac{3}{2}+} |V_{reg}(x)| \in L^1(\mathbb{R})$.*

Remark 5.3. *In fact, less restrictive bounds on V_{reg} as developed in* **[134]** *would suffice. However, for simplicity we will follow the work of* **[178]** *as it makes some computations more explicit.*

6. Proof of Central Theorem 5.1

We follow the strategy described in section 3. Theorem 2.1 will follow from 5.1 by verifying the hypotheses of Theorem 5.1 for $V = V_{sing} + V_{reg}$. This verification is computed in section 7.

Let $\chi(x \geq 1) \in C^\infty(\mathbb{R})$ denote non-decreasing cut-off functions such that

$$\chi(x \geq 1) = \begin{cases} 0 & x \leq \frac{1}{2}, \\ 1 & x \geq 1 \ . \end{cases} \tag{6.1}$$

To localize in frequency space, introduce $\psi(|k| \leq k_0) \in C_0^\infty(\mathbb{R})$ be a compactly supported cut-off function, depending on a parameter, k_0, to be chosen, such that

$$\psi(|k| \leq k_0) = \begin{cases} 1 & |k| \leq \ k_0, \\ 0 & |k| \geq \ 2k_0 \ . \end{cases} \tag{6.2}$$

We decompose any $\phi \in L^2(\mathbb{R})$ into its low and high frequency parts:

$$\phi(x) \ = \ \phi_{low}(x) \ + \ \phi_{high}(x), \qquad \text{where using} \quad D \equiv -i\partial_x, \tag{6.3}$$

$$\phi_{low}(x) \ \equiv \ \psi(|D| \leq k_0)\phi(x) \ \equiv \ \int_{\mathbb{R}} e^{ikx} \psi(|k| \leq k_0) \, \hat{\phi}(k) \ \mathrm{d}k, \tag{6.4}$$

$$\phi_{high}(x) \ \equiv \ (\ 1 \ - \ \psi(|D| \leq k_0)) \ \phi(x) \ \equiv \ \int_{\mathbb{R}} e^{ikx} \ (\ 1 - \psi(|k| \leq k_0) \ ) \, \hat{\phi}(k) \ \mathrm{d}k \ . \tag{6.5}$$

6.1. Bounds on $W_+\phi_{low}$. For $x \geq 0$, we can express $W_+\phi_{low}(x)$, in terms of $m_1(x,k)$, and for $x \leq 0$, we can express $W_+\phi_{low}(x)$, in terms of $m_2(x,k)$. Since the cases $x \geq 0$ and $x \leq 0$ are very similar, we only carry this calculation out in detail for $x \geq 0$. We have, using the notation $Pf(x) = f(-x)$,

$$
\begin{aligned}
W_+\phi_{low} &= F_+^* \mathcal{F}\ \psi(|D| \leq k_0)\phi \\
&= \int_0^\infty e^{ikx}\ T(k)\ m_1(x,k)\ \psi\left(|k| \leq k_0\right)\ \hat{\phi}(k)\ \mathrm{d}k \\
&\quad + \int_{-\infty}^0 e^{ikx}\ T(-k)\ m_2(x,-k)\ \psi\left(|k| \leq k_0\right)\ \hat{\phi}(k)\ \mathrm{d}k \\
&= \int_0^\infty e^{ikx}\ T(k)\ m_1(x,k)\ \psi\left(|k| \leq k_0\right)\ \hat{\phi}(k)\ \mathrm{d}k \\
&\quad + \int_{-\infty}^0 e^{ikx}\ [R_1(-k)e^{-2ikx}\ m_1(x,-k) + m_1(x,k)]\ \psi\left(|k| \leq k_0\right) \hat{\phi}(k)\ \mathrm{d}k \\
&= \int_0^\infty e^{ikx}\ m_1(x,k)\ [T(k) + R_1(k)P]\ \psi\left(|k| \leq k_0\right)\ \hat{\phi}(k)\ \mathrm{d}k + \int_{-\infty}^0 e^{ikx}\ m_1(x,k)\ \hat{\phi}(k)\ \mathrm{d}k,
\end{aligned}
$$

where we have applied (4.5) and (4.15).

We continue by using that $\int_0^\infty\ [\dots]\ \mathrm{d}k = \frac{1}{2}\int_{-\infty}^\infty\ (1 + \mathrm{sgn}(k))\ [\dots]\ \mathrm{d}k$, we have

$$
\begin{aligned}
W_+\phi_{low} =& \frac{1}{2}\int_{-\infty}^\infty (1+\mathrm{sgn}(k))\ e^{ikx}(m_1(x,k)-1)T(k)\ \psi\left(|k| \leq k_0\right) \hat{\phi}(k)\mathrm{d}k \\
&+ \frac{1}{2}\int_{-\infty}^\infty (1+\mathrm{sgn}(k))\ e^{ikx}(m_1(x,k)-1)R_1(k)P\ \psi\left(|k| \leq k_0\right) \hat{\phi}(k)\mathrm{d}k \\
&+ \frac{1}{2}\int_{-\infty}^\infty (1-\mathrm{sgn}(k))\ e^{ikx}(m_1(x,k)-1)\ \psi\left(|k| \leq k_0\right) \hat{\phi}(k)\mathrm{d}k \\
&+ \frac{1}{2}\int_{-\infty}^\infty (1+\mathrm{sgn}(k))\ e^{ikx}T(k)\ \psi\left(|k| \leq k_0\right) \hat{\phi}(k)\mathrm{d}k \\
&+ \frac{1}{2}\int_{-\infty}^\infty (1+\mathrm{sgn}(k))\ e^{ikx}R_1(k)P\ \psi\left(|k| \leq k_0\right) \hat{\phi}(k)\mathrm{d}k \\
&+ \frac{1}{2}\int_{-\infty}^\infty (1-\mathrm{sgn}(k))\ e^{ikx}\ \psi\left(|k| \leq k_0\right) \hat{\phi}(k)\mathrm{d}k, \quad x \geq 0.
\end{aligned} \tag{6.6}
$$

For $x \leq 0$ an analogous representation holds with $m_1(x,k)$ replaced by $m_2(x,k)$.

We now show that $W_{+,1,low}$ is a bounded operator on $W^{1,p}(\mathbb{R}_+)$. Each term in the first three lines of (6.6) is of the form:

$$
\phi\ \mapsto\ S_1\ \circ\ (I \pm i\ \mathcal{H})\ \circ\ \Psi(D)\ \phi\ , \tag{6.7}
$$

and each term in the last three lines is of the form

$$
\phi\ \mapsto\ (I \pm i\ \mathcal{H})\ \circ\ \Psi(D)\ \phi\ , \tag{6.8}
$$

where S_1, is defined in (5.3)-(5.4), $\mathcal{H}$ denotes the Hilbert transform (3.2), and

$$
\begin{aligned}
\Psi(D) &= \mathcal{F}^{-1}\ \hat{\Psi}(k)\ \mathcal{F} \quad \text{and} \\
\hat{\Psi}(k) &= T(k)\ \psi(|k| \leq k_0) \text{ or } R_1(k)\ P\ \psi(|k| \leq k_0) \text{ or } \psi(|k| \leq k_0).
\end{aligned}
$$

(For $x \leq 0$, the argument is parallel with S_1 replaced by S_2.)

By hypotheses on $T(k)$ and $R(k)$, $\hat{\Psi}(k)$ is a multiplier on $W^{1,p}(\mathbb{R})$ for $1 < p < \infty$ **[174]**. The Hilbert transform is bounded (Theorem 3.1), so that the boundedness of the

operators in (6.7) and (6.8) on $W^{1,p}$ for $1 < p < \infty$ follows from boundedness of S_j, that holds by hypothesis. Therefore, one has

$$\|W_+\phi_{low}\|_{W^{1,p}(\mathbb{R})} \;\leq\; C\;\|\phi\|_{W^{1,p}(\mathbb{R})}\;, \tag{6.9}$$

and this completes the low frequency analysis.

6.2. High Frequencies. We have, using (4.9) and the notation $Pf(x) = f(-x)$,

$$\begin{aligned}
W_+\phi_{high} =& F_+^*\mathcal{F}\left(1-\psi(|D|\leq k_0)\right)\phi\\
=&\int_0^\infty T(k)e^{ikx}m_1(x,k)(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
&+\int_{-\infty}^0 T(-k)e^{ikx}m_2(x,-k)(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
=&\int_0^\infty T(k)e^{ikx}m_1(x,k)(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
&+\int_{-\infty}^0 e^{ikx}[R_1(-k)e^{-2ikx}m_1(x,-k)+m_1(x,k)](1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
=&\int_0^\infty e^{ikx}m_1(x,k)[T(k)+R_1(k)P](1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
&+\int_{-\infty}^0 e^{ikx}m_1(x,k)\hat{\phi}(k)\mathrm{d}k.
\end{aligned}$$

For $x \geq 0$ we rewrite this expression as

$$\begin{aligned}
W_+\phi_{high} \;=\;& \frac{1}{2}\int_{-\infty}^\infty e^{ikx}\;(1+\mathrm{sgn}(k))\;(m_1(x,k)-1)T(k)(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
&+\frac{1}{2}\int_{-\infty}^\infty e^{ikx}\;(1+\mathrm{sgn}(k))(m_1(x,k)-1)R_1(k)P(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
&+\frac{1}{2}\int_{-\infty}^\infty e^{ikx}\;(1-\mathrm{sgn}(k))(m_1(x,k)-1)(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
&+\frac{1}{2}\int_{-\infty}^\infty e^{ikx}\;(1+\mathrm{sgn}(k))T(k)(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
&+\frac{1}{2}\int_{-\infty}^\infty e^{ikx}\;(1+\mathrm{sgn}(k))R_1(k)P(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k\\
&+\frac{1}{2}\int_{-\infty}^\infty e^{ikx}\;(1-\mathrm{sgn}(k))(1-\psi\left(|k|\leq k_0\right))\hat{\phi}(k)\mathrm{d}k,\quad x\geq 0.
\end{aligned}$$

An analogous expression, with $m_1(x,k)$ replaced by $m_2(x,k)$, is used for $x \leq 0$. We proceed now to show that each term is bounded on $W^{1,p}(\mathbb{R}_+)$, $p \geq 1$.

Each summand in this decomposition of $W_+\phi_{high}$ is of the form:

$$\phi \mapsto S_j \circ \rho(D)\;\phi, \qquad \text{or} \qquad \phi \mapsto \rho(D)\phi. \tag{6.10}$$

where $\rho(D) = \mathcal{F}^{-1}\hat{\rho}(k)\mathcal{F}$. Here, S_j, $j = 1, 2$, defined in (5.3) and (5.4), is bounded on $W^{1,p}(\mathbb{R}_+)$ for $1 < p < \infty$ by hypothesis. Moreover, $\rho(k)$ is a multiplier on $W^{1,p}(\mathbb{R})$ for $1 < p < \infty$ due to hypotheses on $R(k), T(k) - 1, \partial_k R(k)$ and $\partial_k T(k)$, and the fact that $1 - \psi(|k| \leq k_0)$ is smooth, asymptotically constant as $k \to \infty$ and vanishing in a neighborhood of 0. It follows that

$$\|W_+\phi_{high}\|_{W^{1,p}(\mathbb{R}_+)} \leq C\|V\|_{L^1_{\frac{3}{2}^+}(\mathbb{R})}\;\|\phi\|_{W^{1,p}(\mathbb{R}_+)}. \tag{6.11}$$

An estimate analogous to (6.11), similarly proved using a representation of $W_+\phi_{high}(x)$ for $x \leq 0$, in terms of S_2, also holds. Thus,

$$\|W_+\phi_{high}\|_{W^{1,p}(\mathbb{R})} \leq C\|V\|_{L^1_{\frac{3}{2}+}(\mathbb{R})} \ \|\phi\|_{W^{1,p}(\mathbb{R})} \ . \tag{6.12}$$

The decomposition (6.3) and the bounds (6.9) and (6.12) imply the result. This completes the proof of the central result, Theorem 5.1.

7. Completion of the proof of Theorem 2.1

This section is devoted to the completion of the proof of Theorem 2.1, as a consequence of Theorem 5.1. The hypotheses of Theorem 5.1 are satisfied for potentials $V \in L^1_{\frac{3}{2}+}(\mathbb{R})$, by results in [**135**]. We briefly recall the argument below, and then we generalize it to potentials of the form (1.4), $V = V_{sing} + V_{reg}$, in section 7.2.

7.1. The case of regular potentials. From the relation $m_1(x,k) = e^{-ikx} \ f_1(x,k)$, with $k \in \mathbb{C}$, we have that $m_1(x,k)$ is the unique solution of

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} m_1 + 2ik \ \frac{\mathrm{d}}{\mathrm{d}x} m_1 = V \ m_1, \text{ and } \ m_1(x;k) \to 1, \text{ as } x \to \infty \ . \tag{7.1}$$

Consequently, we have

$$m_1(x,k) = 1 + \int_x^\infty D_k(y-x)V(y)m_1(y,k)\mathrm{d}y, \quad \text{where } \ D_k(x) \equiv \int_0^x e^{2iky}\mathrm{d}y. \tag{7.2}$$

Indeed, for $V \in L^1_{\frac{3}{2}+}(\mathbb{R})$, the iterates of the Volterra integral are bounded by $\frac{\gamma(x)^n}{n!}$, with

$$\gamma(x) \equiv \int_x^\infty (t-x)|V(t)| \ \mathrm{d}t.$$

Summing on n, we find that the majoring series converges and $m_1(x,k)$ satisfies the bound

$$|m_1(x,k)| \ \leq \ e^{\gamma(x)}\gamma(x).$$

By a careful analysis for $x \to -\infty$, one has the improved estimate

$$|m_1(x,k)| \ \leq \ C(1+\max(-x,0)) \int_x^\infty (1+|t|)|V(t)| \ \mathrm{d}t. \tag{7.3}$$

As a consequence, the function $m_1(x,k) - 1$ is in the Hardy space, and therefore there exists $B_1 \in L^2(\mathbb{R} \times \mathbb{R}_+)$ such that

$$m_1(x,k) = 1 + \int_0^\infty B_1(x,y)e^{2iky}\mathrm{d}y. \tag{7.4}$$

Now, the function $B_1(x,y)$ is equivalently defined with

$$B_1(x,y) \ \equiv \ \int_{x+y}^\infty V(t) \ \mathrm{d}t \ + \ \int_0^y \int_{x+y-z}^\infty V(t)B_1(t,z) \ \mathrm{d}t \ \mathrm{d}z \tag{7.5}$$

$$= \sum_{n=0}^\infty K_n(x,y), \tag{7.6}$$

where K_n is defined by induction with

$$K_0(x,y) = \int_{x+y}^\infty V(t)\mathrm{d}t \ , \qquad K_{n+1}(x,y) = \int_0^y \int_{x+y-z}^\infty V(t)K_n(t,z)\mathrm{d}t\mathrm{d}z.$$

It is then easy to prove by induction that

$$|K_n(x,y)| \ \leq \ \frac{\gamma^n(x)}{n!}\eta(x+y), \quad \text{ with } \eta(x) \ \equiv \ \int_x^\infty |V(t)| \ \mathrm{d}t.$$

This allows us to confirm that the sum in (7.6) is well-defined and satisfies (7.5), plus the estimates

$$|B_1(x,y)| \le e^{\gamma(x)}\eta(x+y), \quad \|B_1(x,\cdot)\|_{L^1} \le e^{\gamma(x)}\gamma(x). \tag{7.7}$$

From (7.7), taking the $x-$derivative of (7.5), we have

$$|\partial_x B_1(x,y)| \le\ C\ e^{\gamma_1(x)}\ \left(\ V(x+y)\ +\ \int_{x+y}^{\infty}|V(t)|\mathrm{d}t\ \right),\ x\in\mathbb{R},\ y>0. \tag{7.8}$$

The construction above, and (7.4) with the estimates (7.7) and (7.8), are sufficient to prove

Lemma 7.1. *S_1 is bounded on $W^{1,p}(\mathbb{R}_+)$ and S_2 is bounded on $W^{1,p}(\mathbb{R}_-)$, $1<p<\infty$.*

Proof. We focus on the bound for S_1 on $W^{1,p}(\mathbb{R}_+)$. The bound for S_2 on $W^{1,p}(\mathbb{R}_-)$ is similar. To prove boundedness of the operatorS_1 and ∂S_1 on L^p, we use that the operator

$$S_R\Phi(x)\ =\ \int_{\mathbb{R}} R(x,y)\ \Phi(y)\ \mathrm{d}y,$$

is bounded on L^p with estimate

$$\|S_R\Phi\ \|_{L^p}\ \le\ C_R\ \|\Phi\|_{L^p}\,,\quad 1\le p\le\infty \tag{7.9}$$

if

$$C_R\equiv\sup_{x\ge0}\int_{\mathbb{R}}|R(x,y)|\ \mathrm{d}y\ +\ \sup_{y\ge0}\int_{\mathbb{R}}|R(x,y)|\ \mathrm{d}x\ <\ \infty\ . \tag{7.10}$$

Using the representation formula (7.4) we have

$$R_j(x,y)\ \equiv\ \int_{\mathbb{R}}e^{ik(x-y)}\int_0^{\infty}e^{2ikz}\ B_1(x,z)\ \mathrm{d}k\ \mathrm{d}z\ =\ B_1\left(x,\frac{y-x}{2}\right).$$

Thus, the operator S_1 simplifies to

$$\left(S_1\Phi\right)(x)=\ \int_x^{\infty}B_1\left(x,\frac{y-x}{2}\right)\Phi(y)\mathrm{d}y\ =\ \int_0^{\infty}B_1\left(x,\frac{\zeta}{2}\right)\ \Phi(\zeta-x)\ d\zeta,\quad x\ge0\ .$$

Since we must estimate S_1 on $W^{1,p}$ we also compute

$$\begin{aligned}\partial_x\left(S_1\Phi\right)(x)=&\quad\int_0^{\infty}B_1\left(x,\frac{\zeta}{2}\right)\ (-\partial_\zeta)\Phi(\zeta-x)\ d\zeta\ +\ \int_0^{\infty}\partial_xB_1\left(x,\frac{\zeta}{2}\right)\ \Phi(\zeta-x)\ d\zeta\\ =&\ \int_x^{\infty}B_1\left(x,\frac{y-x}{2}\right)\ (-\partial_y)\Phi(y)\ \mathrm{d}y\ +\ \int_x^{\infty}\partial_xB_1\left(x,\frac{y-x}{2}\right)\ \Phi(y)\ \mathrm{d}y,\quad .\end{aligned}$$

Note that by (7.7) and (7.8) we have for large enough x that

$$|B_1(x,z)|\lesssim\int_{x+z}^{\infty}|V(s)|\mathrm{d}s\ \text{ and }\ |\partial_xB_1(x,z)|\lesssim|V(x)|\ +\ \int_{x+z}^{\infty}|V(s)|\mathrm{d}s. \tag{7.11}$$

Therefore,

$$\begin{aligned}&\sup_{x\ge0}\int\mathbf{1}_{y\ge x}\left|B_1\left(x,\frac{y-x}{2}\right)\right|\ \mathrm{d}y\ +\ \sup_{y\ge0}\int\mathbf{1}_{y\ge x}\left|B_1\left(x,\frac{y-x}{2}\right)\right|\ \mathrm{d}x\\ &\quad\le\ 2\sup_{x\ge0}\int_0^{\infty}\int_{\frac{x+y}{2}}^{\infty}\ |V(s)|\mathrm{d}s\ \mathrm{d}y\\ &\quad\le\ 2\int_0^{\infty}\ \left(1+\frac{x+y}{2}\right)^{-\frac{3}{2}-}\ \int_{\frac{x+y}{2}}^{\infty}(1+s)^{\frac{3}{2}+}|V(s)|\mathrm{d}s\\ &\quad\le const\times\|V\|_{L^1_{\frac{3}{2}+}(\mathbb{R})}.\end{aligned}$$

A similar bound applies to the kernel $\mathbf{1}_{x\geq y}\partial_x B_1\left(x, \frac{y-x}{2}\right)$. Thus, we have

$$\|S_1\Phi\|_{W^{1,p}(\mathbb{R}_+)} \equiv \|S_1\Phi\|_{L^p(\mathbb{R}_+)} + \|\partial_x\left(S_1\Phi\right)\|_{L^p(\mathbb{R}_+)} \leq\ C\|V\|_{L^1_{\frac{3}{2}+}(\mathbb{R})}\ \|\Phi\|_{W^{1,p}(\mathbb{R}_+)}\ .$$

Applying similar arguments with S_1 replaced by S_2 for $x \leq 0$ yields boundedness of S_2 on $W^{1,p}(\mathbb{R}^-)$. □

LEMMA 7.2.

$$|R_j(k)|,\ |T(k)-1| \leq \frac{C}{\langle k\rangle} \quad \forall k \in \mathbb{R}, \qquad |\partial_k T(k)|,\ |\partial_k R_1(k)|,\ |\partial_k R_2(k)| \leq \frac{C}{|k|} \quad \text{as } |k| \to \infty.$$

Proof of Lemma 7.2 for $V = V_{reg}$:
We follow again the method of **[135]**. From (7.2), one has

$$\begin{aligned} m_1(x,k) &= 1 + \frac{1}{2ik}\int_0^\infty (e^{2ik(y-x)} - 1)V(y)m_1(y,k)\ \mathrm{d}y \\ &=\ e^{-2ikx}\left(\frac{1}{2ik}\int_{-\infty}^{+\infty} e^{2iky}m_1(y,k)V(y)\ \mathrm{d}y\right) \\ &\quad +\ \left(1 - \frac{1}{2ik}\int_{-\infty}^{+\infty} m_1(y,k)V(y)\ \mathrm{d}y\right) + o(1), \quad x \to -\infty. \end{aligned}$$

Moreover, one has from (4.8) and (4.9)

$$m_1(x,k)\ =\ e^{-2ikx}\frac{R_2(k)}{T(k)}\ +\ \frac{1}{T(k)}\ +\ o(1), \quad x \to -\infty.$$

This, and the same study on $m_2(x,k)$, leads to the following integral representations:

$$\frac{1}{T(k)}\ =\ 1 - \frac{1}{2ik}\int_{-\infty}^{+\infty} m_1(y,k)V(y)\ \mathrm{d}y, \tag{7.12}$$

$$\frac{R_2(k)}{T(k)}\ =\ \frac{1}{2ik}\int_{-\infty}^{+\infty} e^{2iky}m_1(y,k)V(y)\ \mathrm{d}y, \tag{7.13}$$

$$\frac{R_1(k)}{T(k)}\ =\ \frac{1}{2ik}\int_{-\infty}^{+\infty} e^{2iky}m_2(y,k)V(y)\ \mathrm{d}y. \tag{7.14}$$

These integral representations, together with the $L^1_{\frac{3}{2}+}$ decay assumption on the potential and the uniform bounds (7.3), lead immediately to

$$|R_j(k)|,\ |T(k)-1| \leq \frac{C}{\langle k\rangle}.$$

Now, differentiating (7.4) with respect to k leads to the uniform estimate

$$|\ \partial_k m_1(x,k)\ |\ \leq\ C\langle k\rangle\langle x\rangle,$$

so that (7.12) yields

$$|\partial_k T(k)| \leq \frac{C}{|k|}, \qquad \text{as} \qquad |k| \to \infty.$$

The equivalent estimates for $R_1(k), R_2(k)$ follow similarly from (7.13) and (7.14), and Lemma 7.2 holds.

7.2. The case of potentials with a singular component. In this section we prove that one can generalize the construction above for generalized potentials, satisfying Hypothesis **(V)**, with equivalent estimates, so that Lemma 7.1 and 7.2 hold. As a consequence, the hypotheses of Theorem 5.1 are satisfied, and Theorem 2.1 is proved.
*Proof of Lemma 7.1 for V satisfying Hypotheses (**V**):*
We prove the desired estimates for $m_1(x,k)$, $x \geq 0$, and similar results apply to $m_2(x,k)$, $x \leq 0$. Let us define the function $B_1(x,y)$ with

$$B_1(x,y) \equiv \int_{x+y}^{\infty} V_{reg}(t)\ \mathrm{d}t \ + \ \sum_{l=0}^{N-1} c_l \mathbf{1}(x_l - (x+y)) \ + \ \int_0^y \int_{x+y-z}^{\infty} V_{reg}(t) B_1(t,z)\ \mathrm{d}t\ \mathrm{d}z$$

$$+ \ \int_0^y \sum_{l=0}^{N-1} c_l B_1(x_l,z) \mathbf{1}(x_l - (x+y-z))\ \mathrm{d}z \tag{7.15}$$

$$= \sum_{n=0}^{\infty} K_n(x,y), \tag{7.16}$$

with $\mathbf{1}$ the classical symmetric Heaviside function defined such that

$$\mathbf{1}(x) = \begin{cases} 1\ , & x > 0, \\ \frac{1}{2}\ , & x = 0, \\ 0\ , & x < 0, \end{cases}$$

and K_n defined by induction, with

$$K_0(x,y) = \int_{x+y}^{\infty} V_{reg}(t)\mathrm{d}t \ + \ \sum_{l=0}^{N-1} c_l \mathbf{1}(x_l - (x+y)),$$

$$K_{n+1}(x,y) = \int_0^y \int_{x+y-z}^{\infty} V_{reg}(t) K_n(t,z)\mathrm{d}t\mathrm{d}z + \ \sum_{l=0}^{N-1} c_l \int_0^y K_n(x_l,z)\mathbf{1}(x_l - (x+y-z))\ \mathrm{d}z.$$

Following the proof of Lemma 3, **[135]**, it is easy to show by induction the following pointwise bound

$$|K_n(x,y)| \ \leq \ \frac{\gamma_1^n(x)}{n!}\eta_1(x+y),$$

with γ_1 and η_1 defined as

$$\gamma_1(x) \equiv \int_x^{\infty} (t-x)|V_{reg}(t)|\ \mathrm{d}t \ + \ \sum_{l=0}^{N-1} |c_l|(x_l - x)\mathbf{1}(x_l - x),$$

$$\eta_1(x) \equiv \int_x^{\infty} |V_{reg}(t)|\ \mathrm{d}t \ + \ \sum_{l=0}^{N-1} |c_l|\mathbf{1}(x_l - x).$$

This allows us to confirm that the sum in (7.16) is well-defined and satisfies (7.15), plus the estimates

$$|B_1(x,y)| \leq e^{\gamma_1(x)}\eta_1(x+y), \quad \|B_1(x,\cdot)\|_{L^1} \leq e^{\gamma(x)}\gamma(x) \tag{7.17}$$

and, differentiating with respect to x,

$$|\partial_x B_1(x,y)| \leq \ C\ e^{\gamma_1(x)} \ \left(\ V(x+y) \ + \ \int_{x+y}^{\infty} |V(t)|\mathrm{d}t \ \right), \ x \in \mathbb{R}, \ y > 0. \tag{7.18}$$

Abusing notation (see justification below), we define

$$m_1(x,k) \;\equiv\; 1 \;+\; \int_0^\infty B_1(x,y)e^{2iky}\,\mathrm{d}y, \tag{7.19}$$

it is easy to deduce from (7.15) that

$$\begin{aligned}
\partial_x m_1(x,k) &= \int_0^\infty \left(\partial_x B_1(x,y) - \partial_y B_1(x,y)\right) e^{2iky} \;+\; \partial_y B_1(x,y)e^{2iky}\,\mathrm{d}y \\
&= -\int_0^\infty \int_x^\infty V_{reg}(t)B_1(t,y)\,\mathrm{d}t\; e^{2iky} \;-\; \sum_{l=0}^{N-1} c_l B_1(x_l,y)\mathbf{1}(x_l-x)e^{2iky}\,\mathrm{d}y \\
&\quad -\; \int_0^\infty 2ikB_1(x,y)e^{2iky}\,\mathrm{d}y \;-\; B_1(x,0). \qquad (7.20)\\
\partial_x^2 m_1(x,k) &= \int_0^\infty V_{reg}(x)B_1(x,y)e^{2iky}\,\mathrm{d}y \;+\; \int_0^\infty \sum_{l=0}^{N-1} c_l B_1(x_l,y)\delta(x_l-x)e^{2iky}\,\mathrm{d}y \\
&\quad -\; \int_0^\infty 2ik\partial_x B_1(x,y)e^{2iky}\,\mathrm{d}y \;+\; V_{reg}(x) \;+\; \sum_{l=0}^{N-1} c_l\delta(x_l-x). \qquad (7.21)
\end{aligned}$$

Therefore, $m_1(x,k)$ is the unique function satisfying

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}m_1 + 2ik\,\frac{\mathrm{d}}{\mathrm{d}x}m_1 = \sum_l c_l\delta(x-x_l) + V_{reg}\; m_1,\;\; k\in\mathbb{C},$$

with $m_1(x;k)\to 1$ as $x\to\infty$. Equation (7.19) is thus justified.

Finally, Lemma 7.1 follows from (7.19), with the estimates (7.17) and (7.18).

Proof of Lemma 7.2. We follow again the method of **[135]**. The generalization of (7.2) to potentials satisfying Hypotheses **(V)** is

$$\begin{aligned}
m_1(x,k) &= 1 + \frac{1}{2ik}\int_0^\infty (e^{2ik(y-x)}-1)V_{reg}(y)m_1(y,k)\,\mathrm{d}y \\
&\quad + \frac{1}{2ik}\sum_{l=0}^{N-1} c_l(e^{2ik(x_l-x)}-1)m_1(x_l,k)\mathbf{1}(x_l-x) \\
&= \frac{e^{-2ikx}}{2ik}\left(\int_{-\infty}^{+\infty} e^{2iky}m_1(y,k)V_{reg}(y)\,\mathrm{d}y \;+\; \sum_{l=0}^{N-1} c_l e^{2ikx_l}m_1(x_l,k)\right) \\
&\quad +\; 1 - \frac{1}{2ik}\left(\int_{-\infty}^{+\infty} m_1(y,k)V_{reg}(y)\,\mathrm{d}y + \sum_{l=0}^{N-1} c_l m_1(x_l,k)\right) + o(1)(x\to-\infty).
\end{aligned}$$

Moreover, one has from (4.8) and (4.9)

$$m_1(x,k) \;=\; e^{-2ikx}\frac{R_2(k)}{T(k)} \;+\; \frac{1}{T(k)} \;+\; o(1)(x\to-\infty).$$

This, and the same study on $m_2(x,k)$, leads to the following integral representations:

$$\frac{1}{T(k)} \;=\; 1 - \frac{1}{2ik}\int_{-\infty}^{+\infty} m_1(y,k)V_{reg}(y)\,\mathrm{d}y \;-\; \frac{1}{2ik}\sum_{l=0}^{N-1} c_l m_1(x_l,k), \tag{7.22}$$

$$\frac{R_2(k)}{T(k)} \;=\; \frac{1}{2ik}\int_{-\infty}^{+\infty} e^{2iky}m_1(y,k)V_{reg}(y)\,\mathrm{d}y \;+\; \frac{1}{2ik}\sum_{l=0}^{N-1} c_l e^{2ikx_l}m_1(x_l,k), \tag{7.23}$$

$$\frac{R_1(k)}{T(k)} \;=\; \frac{1}{2ik}\int_{-\infty}^{+\infty} e^{2iky} m_2(y,k) V_{reg}(y)\,\mathrm{d}y \;+\; \frac{1}{2ik}\sum_{l=0}^{N-1} c_l e^{2ikx_l} m_2(x_l,k). \tag{7.24}$$

The identity (7.19), with the estimates (7.17), guarantees the uniform bounds

$$|m_1(x,k)| \le C\langle x\rangle, \qquad |\partial_k m_1(x,k)| \;\le\; C\langle k\rangle\langle x\rangle.$$

Therefore the $L^1_{\frac{3}{2}+}$ decay assumption on the potential V_{reg} leads immediately to

$$|R_j(k)|,\ |T(k)-1| \le \frac{C}{\langle k\rangle}.$$

Now, differentiating (7.22) with respect to k yields

$$|\partial_k T(k)| \le \frac{C}{|k|}, \qquad \text{as} \qquad |k|\to\infty.$$

The equivalent estimates for $R_1(k), R_2(k)$ follow similarly from (7.23) and (7.24), and Lemma 7.2 holds.

8. Examples and Applications

8.1. $V(x)$ = a sum of Dirac delta masses. In this section we verify directly the hypotheses of Theorem 5.1 for the case of a potential, which is the sum of Dirac delta functions, thereby establishing the applicability of our main results to this case.

We follow the analysis from **[148]** and **[176]**, see also **[146]**, **[147]** for specific examples. Seek solutions of the form

$$\left(H_{\vec{q},\vec{y}} - \frac{1}{2}k^2 \right) e_\pm(x,k) = 0, \tag{8.1}$$

where $H_{\vec{q},\vec{y}} = \sum_{j=0}^{N-1} q_j \delta(x-y_j)$ when $\vec{q}=(q_0,\cdots,q_{N-1})$, $\vec{y}=(y_0,\cdots,y_{N-1})$, and where $e_\pm(x,k)$ represent the distorted Fourier basis functions as defined (4.1). Thus,

$$e_+(x,k) = \begin{cases} e^{ikx} + B_0 e^{-ikx} \text{ for } x < y_0, \\ A_1 e^{ikx} + B_1 e^{-ikx} \text{ for } y_0 < x < y_1, \\ \vdots \\ A_N e^{ikx} \text{ for } x > y_{N-1}, \end{cases} \tag{8.2}$$

where we have taken $A_0 = 1$ and $B_N = 0$. With this choice of notation, we have, referring to (4.11) and (4.12), $A_N = T$ the transmission coefficient and $B_0 = R_1$ the reflection coefficient for the "incoming" plane wave e^{ikx} from $-\infty$. Then, we have the following system of equations implied by continuity and jump conditions at the points $\{y_j\}$ for $j = 0,\dots,N-1$:

$$\begin{aligned} e^{iky_0} + B_0 e^{-iky_0} &= A_1 e^{ikx_0} + B_1 e^{-iky_0} \\ ik\left[A_1 e^{iky_0} - B_1 e^{-iky_0} - e^{iky_0} + B_0 e^{-iky_0}\right] &= 2q_0\left[A_1 e^{iky_0} + B_1 e^{-iky_0}\right] \\ &\vdots \\ A_{N-1} e^{iky_{N-1}} + B_{N-1} e^{-iky_{N-1}} &= A_N e^{iky_{N-1}} \\ ik\left[A_N e^{iky_{N-1}} - A_{N-1} e^{iky_0} + B_{N-1} e^{-iky_0}\right] &= 2q_{N-1}\left[A_N e^{iky_{N-1}}\right]. \end{aligned}$$

Note, the above system guarantees unitarity, or that

$$|B_0|^2 + |A_N|^2 = 1. \tag{8.3}$$

We can define similarly

$$e_-(x,k) = \begin{cases} D_0 e^{-ikx} \text{ for } x < y_0, \\ C_1 e^{ikx} + D_1 e^{-ikx} \text{ for } y_0 < x < y_1, \\ \vdots \\ C_N e^{ikx} + e^{-ikx} \text{ for } x > y_{N-1}, \end{cases} \tag{8.4}$$

where now the incoming wave is e^{-ikx} from ∞ and the scattering matrix is determined by the transmission coefficients $D_0 = T$ and the reflection coefficient $C_N = R_2$ for the "incoming" plane wave e^{-ikx} from ∞.

8.1.1. *Bounds on m_1, m_2:* In addition, for general singular potentials with compact support, we have

$$m_1(x,k) = e^{-ikx} f_1(x,k) = \begin{cases} e^{-ikx} \frac{e_+(x,k)}{T(k)} \text{ for } x < y_{N-1}, \\ 1, \text{ for } x > y_{N-1}, \end{cases}$$

$$m_2(x,k) = e^{ikx} f_2(x,k) = \begin{cases} e^{ikx} \frac{e_-(x,k)}{T(k)} \text{ for } x > y_0, \\ 1, \text{ for } x < y_0. \end{cases}$$

Hence, there exists constants $C^1_\alpha(y_{N-1})$ and $C^2_\alpha(y_0)$ such that

$$|\partial_k^\alpha m_1(x,k)| \le C^1_\alpha(y_{N-1}) \text{ for } y_{N-1} > x \ge 0, \tag{8.5}$$

$$|\partial_k^\alpha m_2(x,k)| \le C^2_\alpha(y_0) \text{ for } y_0 < x \le 0. \tag{8.6}$$

As a result, we see that an arbitrary collection of δ functions satisfies the required estimate for the proof of Lemma 7.1.

We conclude this subsection with explicit computations of the transmission and reflection coefficients for single and double δ well potentials:

8.1.2. **Single δ potential ($H_q = -q\delta(x)$)***:* Setting up the appropriate equations, we have

$$R_1 = r_q = \frac{q}{ik - q}, \tag{8.7}$$

$$T = t_q = \frac{ik}{ik - q}, \tag{8.8}$$

where r_q, t_q are the reflection and transmission coefficients for H_q respectively. We must show the bounds from (5.1) hold, however such bounds follow clearly for (8.8), (8.7).

8.1.3. **Double δ potential ($H_{q,L} = -q(\delta(x+L) + \delta(x-L))$)***:* Setting up the appropriate equations, we have

$$R_1 = r_{q,L} = \left(\frac{q(ik-q)e^{2ikL} + q(ik+q)e^{-2ikL}}{q^2 e^{2ikL} - (ik+q)^2 e^{-2ikL}} \right) e^{-2ikL}, \tag{8.9}$$

$$T = t_{q,L} = \left(\frac{k^2}{q^2 e^{2ikL} - (ik+q)^2 e^{-2ikL}} \right) e^{-2ikL}, \tag{8.10}$$

where $r_{q,L}$, $t_{q,L}$ are the reflection and transmission coefficients for $H_{q,L}$ respectively.

Again, we must verify Lemma (7.2), hence we must prove for instance

$$|\partial_k t_{q,L}(k)| \le C(1 + |k|)^{-1},$$

provided $qL \neq 1/2$. Indeed, we have

$$\partial_k t_{q,L}(k) = \frac{2k(k^2 - 2ikq + q^2(e^{4ikL} - 1)) - 2ik^2(2Lq^2 e^{4ikL} - (ik+q))}{(k^2 - 2ikq + q^2(e^{4ikL} - 1))^2},$$

which satisfies

$$|\partial_k t_{q,L}(k)| \sim \mathcal{O}(|k|^{-1})$$

as $k \to \infty$ and

$$|\partial_k t_{q,L}(k)| \sim \mathcal{O}\left(\frac{1}{4q^2L - 2q}\right)$$

as $k \to 0$. A similar computation holds for $r_{q,L}$.

8.2. Commutator / Resolvent type bounds. In **[137]**, where homogenization of high contrast oscillatory structures with defects is studied, bounds on $(H_0+1)^{-1}(H_{\vec{q},\vec{y}}+1)$ are required to estimate a Lipmann Schwinger equation. We have, by our main theorem that

$$(H_0+1)^{-1}(H_{\vec{q},\vec{y}}+1)P_c = (H_0+1)^{-1}W_+(H_0+1)W_+^* : L^2 \to L^2.$$

8.3. Dispersive and Strichartz estimates in H^1 for δ-Schrödinger. We may represent

$$e^{itH}P_c f = \frac{1}{2\pi}\int\int_0^\infty e^{-\frac{itk^2}{2}}\left(e_+(x,k)\,\overline{e_+(x,k)} + e_-(x,k)\,\overline{e_-(x,k)}\right) f(y)\mathrm{d}k\mathrm{d}y. \tag{8.11}$$

From here, we may use direct computations to arrive at Strichartz estimates and apply Weder's results on wave operators since the potentials are all in L^1 with compact support.

Using the properties of wave operators, we have

$$\|e^{iHt}P_c f\|_{L^p} = \|W_\pm e^{itH_0}W_\pm^* f\|_{L^p} \tag{8.12}$$

and using standard dispersive estimates for the linear Schrödinger operator (see for instance **[175]** for a concise overview), we arrive at

$$\|e^{iHt}P_c f\|_{L^p} \leq C_p t^{-(\frac{1}{2}-\frac{1}{p})}\|f\|_{W^{1,p}}. \tag{8.13}$$

Define a Strichartz pair (q,r) to be admissible if

$$\frac{2}{q} = \frac{1}{2} - \frac{1}{r} \tag{8.14}$$

with $2 \leq r < \infty$. Then, we arrive at the celebrated Strichartz estimates

$$\|e^{iHt}P_c u_0\|_{L^qW^{1,r}} \lesssim \|u_0\|_{W^{1,2}} \tag{8.15}$$

and

$$\left\|\int_0^t e^{iH(t-s)}P_c f\right\|_{L^qW^{1,r}} \lesssim \|f(x,t)\|_{L_t^{\tilde{q}}W_x^{1,\tilde{r}}} \tag{8.16}$$

using duality techniques and once again the boundedness of the wave operators.

As a side note, using positive commutators and well crafted local smoothing spaces, from **[161]** we have the Strichartz estimate

$$\left\|\int_0^t e^{iH(t-s)}P_c f\right\|_{L^\infty L^2} \lesssim \|f(x,t)\|_{L_t^{\tilde{p}}L_x^{\tilde{q}}}. \tag{8.17}$$

Now, by boundedness of wave operators on $W^{1,p}$ spaces for singular potentials as proved in Theorem 5.1, we have the following useful relation

$$\left\|\int_0^t e^{iH(t-s)}P_c f\right\|_{L^\infty H^1} \lesssim \|f(x,t)\|_{L_t^{\tilde{p}}W_x^{1,\tilde{q}}}, \tag{8.18}$$

where $(\tilde{p},\tilde{q})$ is a dual Strichartz pair without first going through the dispersive estimates.

8.4. Local Well-Posedness in H^1 for δ-NLS. Consider the nonlinear Schrödinger / Gross-Pitaevskii, with a potential consisting of a finite set of Dirac delta functions:

$$\begin{cases} i\partial_t u + H_{\vec{q},\vec{y}} u - |u|^{2\sigma} u = 0, \\ u(x,0) = u_0(x) \in H^1, \end{cases}$$

for $0 < \sigma < \infty$. We seek a solution in the following sense:

$$u = \Lambda[u],$$

where

$$\Lambda[u](t) = e^{-iH_{\vec{q},\vec{y}}t} u_0 - i \int_0^t e^{-iH_{\vec{q},\vec{y}}(t-s)} |u|^{2\sigma} u(s) ds. \tag{8.19}$$

We claim that local well-posedness can be established via the contraction mapping principle in the space $C^0([0,T); H^1(\mathbb{R}))$ for T sufficiently small. To prove the necessary boundedness and contraction estimates, it is natural to apply the operator $(I + H_{\vec{q},\vec{y}})^{\frac{1}{2}} P_c$, which commutes with the group $e^{-iH_{\vec{q},\vec{y}}t}$ to (8.19). Then, estimates follow in a straightforward way, using that $H^1(\mathbb{R})$ is an algebra, provided the space

$$\mathcal{H}^1(\mathbb{R}) = \left\{ f : (I + H_{\vec{q},\vec{y}})^{\frac{1}{2}} P_c f \in L^2(\mathbb{R}) \right\} \tag{8.20}$$

is equivalent to the classical Sobolev space H^1. This follows from the relations

$$(I+H)^{\frac{1}{2}} P_c = W (I - \partial_x^2)^{\frac{1}{2}} W^*, \quad W^* (I+H)^{\frac{1}{2}} W = (I - \partial_x^2)^{\frac{1}{2}}$$

and our results on the boundedness of wave operators associated with $H_{\vec{q},\vec{y}}$ on H^1.

8.5. Long time dynamics for NLS with a double δ well potential. In **[162]**, the long time dynamics of solutions to the nonlinear Schrödinger / Gross-Pitaevskii equation

$$i\partial_t u = (-\Delta + V(x))u \; + \; gK\left[|u|^2\right] u, \tag{8.21}$$

where V is a symmetric, double well potential, are studied. In particular, under appropriate spectral assumptions on the operator $H = -\partial_x^2 + V(x)$, in a neighborhood of a symmetry breaking bifurcation point, there are different classes of oscillating solutions (8.21) which *shadow* periodic orbits of a finite dimensional reduction on very long, but finite, time scales. These solutions correspond to states with mass concentrations oscillating between the two wells of a symmetric potential well. The proof requires dispersive / Strichartz type estimates. The results of this chapter imply that the results of **[162]** extend to (8.21) for the case of singular potentials, such as

$$V(x) = -q[\delta(x-L) + \delta(x+L)].$$

CHAPITRE 6

Scattering, homogenization and edge effects for oscillatory potentials with strong singularities

À paraître dans Multiscale Model. Simul., en collaboration avec Michael I. Weinstein.

Sommaire

Abstract

We study one-dimensional scattering for a decaying potential with rapid periodic oscillations and strong localized singularities. In particular, we consider the Schrödinger equation

$$H_\epsilon\,\psi \equiv \left(\, -\partial_x^2 + V_0(x) + q\,(x, x/\epsilon)\,\right)\psi = k^2\psi$$

for $k \in \mathbb{R}$ and $\epsilon \ll 1$. Here, $q(\cdot, y+1) = q(\cdot, y)$, has mean zero and $|V_0(x) + q(x,\cdot)| \to 0$ as $|x| \to \infty$. The distorted plane waves of H_ϵ are solutions of the form: $e_{V^\epsilon\pm}(x;k) = e^{\pm ikx} + u^s_\pm(x;k)$, $u^s_\pm$ outgoing as $|x| \to \infty$. We derive their ϵ small asymptotic behavior, from which the asymptotic behavior of scattering quantities such as the transmission coefficient, $t^\epsilon(k)$, follow.

Let $t_0^{hom}(k)$ denote the homogenized transmission coefficient associated with the average potential V_0. If the potential is smooth, then classical homogenization theory gives asymptotic expansions of, for example, distorted plane waves, and transmission and reflection coefficients. Singularities of V_0 or discontinuities of q_ϵ are "interfaces" across which a solution must satisfy interface conditions (continuity or jump conditions). To satisfy these conditions it is necessary to introduce *interface correctors*, which are highly oscillatory in ϵ.

Our theory admits potentials which have discontinuities in the microstructure, $q_\epsilon(x)$ as well as strong singularities in the background potential, $V_0(x)$. A consequence of our main results is that $t^\epsilon(k) - t_0^{hom}(k)$, the error in the homogenized transmission coefficient is (i) $\mathcal{O}(\epsilon^2)$ if q_ϵ is continuous and (ii) $\mathcal{O}(\epsilon)$ if q_ϵ has

discontinuities. Moreover, in the discontinuous case the correctors are highly oscillatory in ϵ, *i.e.* $\sim \exp(2\pi i \frac{\nu}{\epsilon})$, for $\epsilon \ll 1$. Thus a first order corrector is not well-defined since $\epsilon^{-1}\left(t^\epsilon(k) - t_0^{hom}(k)\right)$ does not have a limit as $\epsilon \to 0$. This expression may have limits which depend on the particular sequence through which ϵ tends to zero. The analysis is based on a (pre-conditioned) Lippman-Schwinger equation, introduced in [**144**].

1. Introduction

An important method for computing the effective properties of highly oscillatory media is the method of homogenization. The goal of homogenization is to approximate a highly oscillatory medium, described by a differential equation with oscillatory coefficients, by an approximate and homogeneous medium, described by a "homogenized" differential equation with constant or slowly varying coefficients. In its regime of validity, the homogenized differential equation (i) predicts effective properties which are approximately those of the heterogeneous medium and (ii) is, by comparison with the full problem, much simpler to study either analytically or by numerical simulation.

While the homogenized limit can often be obtained by a formal multiple scale expansion or by variational methods [**129, 153, 127, 177**], these expansions are typically valid in the *bulk medium*, away from boundaries, discontinuities or more singular sets of coefficients. Indeed, solutions to elliptic operators with oscillatory coefficients on bounded domains have been shown to require boundary layer correctors, which are sensitive to the manner in which the microstructure meets a boundary [**169, 164, 128, 141, 142**] or interface [**172**]. Furthermore, the importance of correctors to homogenization due to interface effects, boundary layers *etc.* is explored analytically and computationally, in the context of accurate estimation of scattering resonances in [**143, 144**].

In this article we study the scattering problem for the one-dimensional time-independent Schrödinger equation

$$\left(H_\epsilon - k^2 \right) \psi \equiv \left(-\frac{d^2}{dx^2} + V^\epsilon(x) - k^2 \right) \psi(x) = 0. \tag{1.1}$$

The potential, $V^\epsilon(x) = V_0(x) + q\left(x, x/\epsilon\right)$, is the sum of a slowly varying part with smooth and singular components, $V_0 = V_{reg} + V_{sing}$, and a rapidly oscillatory part, $q_\epsilon(x) = q\left(x, x/\epsilon\right)$, $\epsilon \ll 1$. $V^\epsilon(x)$ is assumed to decay to zero as x tends to infinity. We also assume $V^\epsilon(x) \geq 0$, a simple way to restrict to the case where H_ϵ has no discrete eigenvalues (bound states) and has only continuous spectrum (extended / radiation states). The wave number, k, is fixed and we study the $\epsilon-$ small behavior.

Many physically important scattering properties are not captured by leading order homogenization. Line-widths and imaginary parts of scattering resonances are key to quantifying the lifetimes of metastable states in quantum systems, or in electro-magnetics, the leakage rates of energy from photonic structures; see [**143, 144**] and references therein. In [**143, 144**] it was shown that inclusion of even the first non-trivial correction due to microstructure can yield large improvements in the approximation of such scattering quantities. Since, as we shall see, defects and singularities can be responsible for the *dominant* correctors and these contributions are not captured in smooth homogenization setting, we therefore seek a better understanding of homogenization for wave / scattering problems in their presence. In this paper we ask:

How are scattering properties, such as transmission and reflection coefficients, $t_\epsilon(k)$ and $r_\epsilon(k)$, influenced by interfaces, defects and singularities?

The heart of the matter is an asymptotic study of the *distorted plane waves*, solutions of $(H_\epsilon - k^2)\psi = 0$ of the form:

$$e_{V^\epsilon \pm}(x;k) = e^{\pm ikx} + u^s_\pm(x;k), \quad u^s_\pm \text{ outgoing as } |x| \to \infty, \quad \text{for } \epsilon \text{ small.}$$

Consequences of our analysis include the following:

i. Theorem 5.1 provides a convergent expansion of the distorted plane waves of $H_Q = -\partial_x^2 + V_0 + Q$, which is valid for a large class of perturbing potentials, Q, which may be pointwise large, but highly oscillatory (supported at high frequencies although not necessarily periodic). Theorem 5.6 is the corresponding expansion for the transmission coefficient $t[k;Q]$. By Proposition 5.4 we can apply Theorems 5.1 and 5.6 to $Q(x) = q_\epsilon(x) = q\,(x, x/\epsilon)$, where $q(x,y)$ is $1-$ periodic in y, decaying as $|x| \to \infty$, and satisfies Hypotheses **(V)**.

ii. Theorem 2.2 implies that:
(i) $t^\epsilon(k) - t_0^{hom}(k) = \mathcal{O}(\epsilon^2)$ if q_ϵ is continuous and
(ii) $t^\epsilon(k) - t_0^{hom}(k) = \mathcal{O}(\epsilon)$ if q_ϵ has discontinuities.
For q_ϵ discontinuous *interface correctors*, which are highly oscillatory in ϵ, enter the expansion; see the discussion in section 4 concerning failure and restoration of interface conditions at singularities of V_0 or discontinuities of q_ϵ. These correctors are related to the asymptotics of boundary layers arising in work on homogenization of divergence form operators on bounded domains [**169, 164, 128, 141, 142**]. Since these correctors involve ϵ dependence of the form: $\sim \exp(2\pi i \frac{\nu}{\epsilon})$, $\epsilon \ll 1$, $0 \neq \nu \in \mathbb{R}$, the expression $\epsilon^{-1}\left(t^\epsilon(k) - t_0^{hom}(k)\right)$ does not have a limit as $\epsilon \to 0$, and a correction to the value of $t_0^{hom}(k)$ is not well-defined. However, there can be limits which depend on the particular sequences through which ϵ tends to zero. See the more detailed discussion after the statement of Theorem 2.3.

Outline of paper: In section 2 we state detailed hypotheses and our main theorems on transmission coefficients, Theorems 2.2 and 2.3, which depend on our analysis of distorted plane waves (Theorem 5.1). We also present the results of numerical simulations designed to illustrate the relationship between regularity of the potential, V^ϵ, and ϵ small asymptotics of the transmission coefficient, stated in Theorem 2.2. In section 3 we present the technical background on one-dimensional scattering theory. In section 4 we derive, by including *interface correctors* to an expansion derived by the classical method of multiple scales, an expansion of the distorted plane waves and of the transmission coefficient valid to all orders in the small parameter ϵ . Section 5 contains rigorous proofs of the expansion of the distorted plane waves (Theorem 5.1) and transmission coefficients (Theorem 5.1 and 5.6) with error bounds. The proof is based on the reformulation of the scattering problem as a pre-conditioned Lippman-Schwinger equation, an approach introduced in [**144**]. Appendix A contains a brief discussion of the numerical methods used in the simulations. Appendix C contains the technical proof of operator bounds which are central to the proofs in section 5.

Acknowledgements: The authors wish to thank R.V. Kohn and J. Marzuola for fruitful discussions. VD was supported, in part, by Agence Nationale de la Recherche Grant ANR-08-BLAN-0301-01. MIW was supported in part by NSF grant DMS-07-07850 and DMS-10-08855. MIW would also like to acknowledge the hospitality of the Courant Institute of Mathematical Sciences, where he was on sabbatical during the preparation of this article. VD would like to thank the Department of Applied Physics and Applied Mathematics (APAM) at Columbia University for its hospitality during the Spring of 2008 when this work was initiated.

2. Main results and Discussion

We begin with the key hypotheses. Hypotheses **(V)** make precise the decomposition of the potential, V, into regular, singular and oscillatory parts. Hypothesis **(G)** specifies, for the cases of *generic* and *non-generic* potentials, V_0, the admissible values of the wave number, k. We then state and discuss our main results concerning the transmission coefficients, in the small ϵ limit.

Hypotheses (V)

$$V^\epsilon(x) \equiv V_0(x) + q_\epsilon(x), \quad \text{real}-\text{valued} \tag{2.1}$$
$$\equiv V_{sing}(x) + V_{reg}(x) + q_\epsilon(x),$$
$$q_\epsilon(x) \equiv q\left(x, \frac{x}{\epsilon}\right), \quad V^\epsilon(x) \geq 0, \tag{2.2}$$

where

i. *Singular part of V^ϵ, V_{sing}:*

$$V_{sing}(x) = \sum_{j=0}^{N-1} c_j\, \delta(x - x_j), \quad \text{where } c_j, x_j \in \mathbb{R}, \quad x_j < x_{j+1}. \tag{2.3}$$

ii. *Regular part of V^ϵ:* $V_{reg} \in L^{1,2}(\mathbb{R})$ with

$$\|V\|_{L^{1,2}} \equiv \int_{\mathbb{R}} (1+|s|)^2\, |V(s)|\, \mathrm{d}s < \infty. \tag{2.4}$$

iii. *Rapidly varying part of V^ϵ, $q_\epsilon(x) = q\left(x, \frac{x}{\epsilon}\right)$:* The mapping $(x,y) \mapsto q(x,y)$ is
(a) $1-$ periodic, *i.e.* for each $x \in \mathbb{R}$, $q(x, y+1) = q(y)$,
(b) mean zero with respect to y, *i.e.* for each $x \in \mathbb{R}$,

$$\int_0^1 q(x,y)\, \mathrm{d}y = 0, \tag{2.5}$$

(c) $q \in pC^3_x L^2_{y,per}$, the set of functions $q : \mathbb{R} \times S^1 \to \mathbb{R}$, such that there exists a finite partition of $\mathbb{R}$

$$-\infty = a_0 < a_1 < a_2 < \cdots < a_M < a_{M+1} = +\infty$$

with

$$\sum_{j=0}^{M+1} \int_0^1 \|q(\cdot, y)\|^2_{C^3(a_j, a_{j+1})}\, \mathrm{d}y < \infty \tag{2.6}$$

iv. We shall work with the Fourier expansion of $q(x,y)$, written as

$$q(x,y) = \sum_{j \neq 0} q_j(x)\, e^{2\pi i j y}, \quad q_j(x) \equiv \int_0^1 e^{-2\pi i j y} q(x,y)\, \mathrm{d}y \tag{2.7}$$

and assume

$$\int_{\mathbb{R}} \int_0^1 |q(x,y)|^2\, \mathrm{d}y\, \mathrm{d}x = \sum_{|j| \geq 1} \int_{\mathbb{R}} |q_j(x)|^2 < \infty, \tag{2.8}$$

$$\int_0^1 |q(x,y)|^2\, \mathrm{d}y = \sum_{|j| \geq 1} |q_j(x)|^2 \to 0, \quad |x| \to \infty\,. \tag{2.9}$$

v. Proposition 5.4, which is a step in proving Theorem 2.2, requires more decay at infinity for q_ϵ: there exists $\rho > 8$ such that

$$(1+|\cdot|^2)^{\rho/2} q_j \in L^2,\ |j| \geq 1, \quad \text{and} \quad \sum_{|j|\geq 1} \left\| (1+|\cdot|^2)^{\rho/2} q_j \right\|_{L^2} < \infty, \tag{2.10}$$

$$\frac{d}{\mathrm{d}x}\left((1+|x|^2)^{\rho/2} q_j(x)\right) \in L^2 \text{ and } \sup_{|j|\geq 1} \left\| \frac{d}{\mathrm{d}x}\left((1+|x|^2)^{\rho/2} q_j(x)\right) \right\|_{L^2_x} < \infty. \tag{2.11}$$

Hypothesis (G) If V_0 is generic (see Definition 3.6), then the wave number, $k \in K$, an arbitrary compact subset of $\mathbb{R}$. If V_0 is not generic, then the compact set K must be such that $0 \notin K$.

REMARK 2.1. *If V_0 is not generic (as for example $V_0 \equiv 0$), then the expansions we present in Theorem 2.2 and Theorem 2.3, are not uniform in a neighborhood of $k = 0$. This will be the subject of a future paper.*

The aim of this article is to understand the scattering properties for this class of potentials. In particular, we are interested in the influence of combined microstructure (q_ϵ) and singularities (V_{sing}) on the reflection and transmission coefficients and distorted plane waves (see below). Formal application of classical homogenization theory (see for example [**129**]) suggests that the leading order (in $\epsilon \to 0$) scattering behavior is governed by the averaged (homogenized) operator $-\partial_x^2 + V_0(x)$; see (2.1). For example, if $V^\epsilon(x)$ is smooth (in particular, $V_{sing} \equiv 0$), then the transmission coefficient satisfies the expansion

$$t^\epsilon(k) \quad \sim \quad t_0^{hom}(k) + \epsilon t_1^{hom}(k) \ + \epsilon^2 t_2^{hom}(k) \ + \ \ldots \tag{2.12}$$

where t_j^{hom} are computed from the formal 2-scale homogenization expansion. In particular, t_0^{hom} is the transmission coefficient associated with the averaged potential $V_0(x)$, However, homogenization is a theory valid only in the *bulk*, away from boundaries or non-smooth points of coefficients. For our class of potentials, this expansion must be corrected.

Our main result is the small ϵ characterization of the distorted plane waves presented in Theorem 5.1. A key consequence of our analysis is the following:

THEOREM 2.2. *Let $V^\epsilon(x) = V_0(x) + q_\epsilon(x)$ with V_0 and $q_\epsilon(x) = q(x, x/\epsilon)$ satisfying Hypotheses **(V)**, and $k \in K$ a compact subset of $\mathbb{R}$ satisfying Hypothesis **(G)**. Denote by $e_{V_0\pm}(x;k)$ the distorted plane waves associated with the unperturbed operator $-\partial_x^2 + V_0(x)$; see section 3.*

Then, there exists $\epsilon_0 = \epsilon_0(K)$, such that for $0 < \epsilon < \epsilon_0$, the transmission coefficient $t^\epsilon = t^\epsilon(k)$ (see (3.7)) associated with $V^\epsilon(x)$ satisfies the following expansion uniformly in $k \in K$:

$$t^\epsilon(k) \ = \ t_0^{hom}(k) \ + \ \epsilon\, t_1^\epsilon(k) \ + \ \epsilon^2 \ \left(\ t_2^{hom}(k) \ + \ t_2^\epsilon(k) \ \right) \ + \ t_{rem}^\epsilon(k), \tag{2.13}$$

where $t_0^{hom}(k)$ denotes the transmission coefficient, associated with the average (homogenized) potential V_0 and

$$t_1^\epsilon(k) = \frac{1}{4k\pi} \sum_{j=1}^{M} e_{V_0+}(a_j;k) e_{V_0-}(a_j;k) \sum_{|l|\geq 1} [q_l]_{a_j} \frac{e^{2i\pi l \frac{a_j}{\epsilon}}}{l}, \tag{2.14}$$

$$t_2^{hom}(k) = \frac{i}{8k\pi^2} \sum_{|j|\geq 1} j^{-2} \int_{\mathbb{R}} |q_j(z)|^2 \ e_{V_0-}(z;k) e_{V_0+}(z;k) dz, \tag{2.15}$$

$$t_2^\epsilon(k) = \frac{i}{8k\pi^2}\sum_{j=1}^{M}\sum_{|l|\geq 1}\left[\ \partial_x\left(e_{V_0+}(x;k)\ e_{V_0-}(x;k)q_l(x)\right)\ \right]_{a_j}\frac{e^{2i\pi l\frac{a_j}{\epsilon}}}{l^2}, \tag{2.16}$$

$$t_{rem}^\epsilon(k) = o(\epsilon^{2+}), \text{ more precisely quantified in Proposition 5.7.} \tag{2.17}$$

(a) t_j^{hom}, $j = 0, 2\ldots$, *denote the expansion coefficients for the transmission coefficient obtained from the two-scale (bulk) homogenization expansion, valid for smooth potentials.*
(b) t_1^ϵ *arises due to discontinuities in* $x \mapsto q(x,\cdot)$, *and*
(c) t_2^ϵ *arises due to both the singular part of the potential,* V_{sing}, *and discontinuities in* $x \mapsto q(x,\cdot)$ *or* $x \mapsto \partial_x q(x,\cdot)$.

t_1^ϵ *and* t_2^ϵ *are uniformly bounded, for* ϵ *small. However each is a sum over rapidly oscillating (as* $\epsilon \to 0$*) terms of the form* $\exp(i\frac{\nu}{\epsilon})$, *corresponding to discontinuity points of* q_ϵ, *respectively points in the support of* V_{sing}.

Theorem 2.2 is a consequence of the more general Theorem 2.3, stated below, which follows from the asymptotic study of the convergent expansion of the distorted plane waves, presented in Theorem 5.1. The proof of Theorem 5.1 is based on construction and asymptotic study of the scattering problem via a *pre-conditioned* Lippman-Schwinger equation. This approach is quite general and applies to the perturbation theory of Schrödinger operators of the form

$$H = -\partial_x^2 + V_0(x)\ +\ Q(x),$$

where Q is small in the sense that $|||Q||| \sim \left\|(I-\Delta)^{-\frac{1}{2}}Q(I-\Delta)^{-\frac{1}{2}}\right\|_{L^2\to L^2}$ is small. This formulation was introduced in **[144]** to study the perturbation of scattering resonances due to high contrast microstructure perturbations of a potential. If Q is a "microstructure", roughly meaning that it is supported at high frequencies, then $|||Q|||$ is small. Here, we apply this method and obtain a convergent expansion of $Q \mapsto e_{V_0+Q}(x,k)$ for fixed k and $|||Q|||$ sufficiently small. The expansion of the transmission coefficient, $Q \mapsto t_{V_0+Q}(k)$, is a direct consequence of:

Theorem 2.3. *Let* $V(x) = V_0(x) + Q(x)$ *with* V_0 *satisfying Hypotheses* ***(V)*** *and* $(1+|x|^2)^{\rho/2}Q \in L^2$, *for* $\rho > 8$. *We use the following norm on* Q, *see section 5.2:*

$$|||Q||| \ \equiv\ \left\|\langle D_0\rangle^{-1}\ (1+|x|^2)^{\rho/4}\ Q\ (1+|x|^2)^{\rho/4}\ \langle D_0\rangle^{-1}\right\|_{L^2\to L^2}.$$

Set $k \in K$ *a compact subset of* $\mathbb{R}$ *satisfying Hypothesis* ***(G)***, *and denote by* $e_{V_0\pm}(x;k)$ *the distorted plane waves associated with the unperturbed operator* $-\partial_x^2 + V_0(x)$; *see section 3. Denote by* $t = t(k,Q) = t(k)$ *the transmission coefficient (see* (3.7)*) associated with* $V(x)$. *There exists* $\tau_0 = \tau_0(K)$ *such that for* $0 < |||Q||| < \tau_0(K)$, *we have the following expansion which holds uniformly in* $k \in K$*:*

$$t(k,Q) \ =\ t_0^{hom}(k)\ +\ t_1[Q]\ +\ t_2[Q,Q]\ +\ t_{rem}(k), \tag{2.18}$$

with $t_0^{hom}(k)$ *the transmission coefficient, associated with the average (homogenized) potential* V_0, *and the following:*

$$t_1[Q] \ =\ \frac{1}{2ik}\int_{-\infty}^{\infty} Q(\zeta)\ e_{V_0+}(\zeta;k)\ e_{V_0-}(\zeta;k)\ d\zeta, \tag{2.19}$$

$$t_2[Q,Q] \ =\ \frac{1}{2ik}\int_{-\infty}^{\infty} Q\ R_{V_0}(k)(Q(\zeta)\ e_{V_0+}(\zeta;k))\ e_{V_0-}(\zeta;k)\ d\zeta, \tag{2.20}$$

$$t_{rem}(k) \ =\mathcal{O}\left(|||Q|||^{2+}\right) \quad \text{and more precisely estimated in Theorem 5.6.} \tag{2.21}$$

Here, $R_{V_0}(k)$, $e_{V_0+}(x;k)$ *and* $e_{V_0-}(x;k)$ *being defined in section 3.*

REMARK 2.4. **Symmetry considerations:** *There is a class of potentials, q_ϵ, whose members are discontinuous, and yet the (oscillatory in ϵ) correctors, t_j^ϵ, $j \geq 1$ vanish. In subsection 2.1 we explore families of such structures. Indeed, let us apply Theorem 2.3 with $V \equiv V_\epsilon$ satisfies Hypotheses* **(V)** *as well as the additional properties: V_0 even and q_ϵ "separable":*

$$V_0(x) = V_0(-x), \qquad q_\epsilon(x) \;=\; q_0(x)\; q_{per}\left(\frac{x}{\epsilon}\right).$$

One can easily see that V_0 is even implies that $e_{V_0+}(\cdot;k)e_{V_0-}(\cdot;k)$ is even. Therefore, if q_0 and q_{per} are of opposite parity, then $x \mapsto e_{V_0+}(x;k)\; e_{V_0-}(x;k)\; q_\epsilon(x)$ is odd and therefore $t_1[q_\epsilon](k) \equiv 0$ for any $\epsilon > 0$. It follows that for such potentials, and even if q_ϵ is discontinuous, *the leading order correction to $t_0^{hom}(k)$ is $t_2[Q,Q]$ which is of order $\mathcal{O}(\epsilon^2)$ (see section 5.4). Moreover, in this special case the second order corrector is well defined:*

$$\lim_{\epsilon \downarrow 0}\; \epsilon^{-2}\left(\; t^\epsilon(k) - t_0^{hom}(k)\;\right) = t_2^{hom}(k).$$

The three subplots of figure 1 illustrate the results of Theorem 2.2 on the behavior of $t^\epsilon - t_0$ for several contrasting choices of potential $V^\epsilon = V_0 + q_\epsilon$, where V_0 is a finite sum of Dirac delta functions, at equally spaced points.[1]

- The left panel of figure 1 corresponds to the case where q_ϵ is discontinuous. It shows that

  $$t^\epsilon - t_0^{hom} = \mathcal{O}(\epsilon), \quad \epsilon \to 0, \quad \text{and yet} \quad \epsilon^{-1}\left(\; t^\epsilon - t_0^{hom}\;\right) \quad \text{does not have a limit.}$$

- The center panel of figure 1 corresponds to the case where q_ϵ is a smooth function, and V_0 is a Dirac delta function. Here,

  $$t^\epsilon - t_0^{hom} = \mathcal{O}(\epsilon^2), \quad \epsilon \to 0, \quad \text{and yet} \quad \epsilon^{-2}\left(\; t^\epsilon - t_0^{hom}\;\right) \quad \text{does not have a limit.}$$

- The right panel of figure 1 corresponds to the case where q_ϵ is a smooth function, and V_0 is a *smoothed out* Dirac delta function. Here we find

  $$t^\epsilon - t_0^{hom} = \mathcal{O}(\epsilon^2), \quad \epsilon \to 0, \quad \text{and} \quad \lim_{\epsilon \downarrow 0} \epsilon^{-2}\left(\; t^\epsilon - t_0^{hom}\;\right) \quad \text{is well} - \text{defined.}$$

This phenomenon of indeterminacy of higher order correctors, due to boundary layer effects is discussed, in the context of a Dirichlet spectral problem **[169, 164]**.

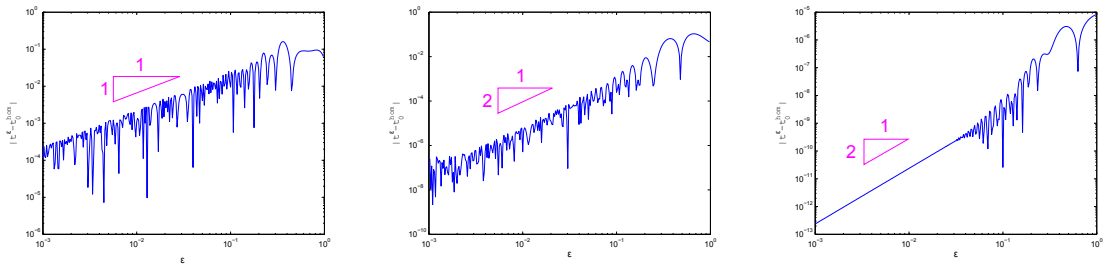


FIGURE 1. Illustration of Theorem 2.2 via plot of $\log\left|t^\epsilon - t_0^{hom}\right|$ versus $\log \epsilon^{-1}$ for the case of q discontinuous and V_0 a sum of Dirac delta functions (left panel, average slope 1), q smooth and V_0 a sum of Dirac delta functions (center panel, average slope 2). The right panel (slope 2) is for the case where $V^\epsilon = V_0$ is a *smooth approximation* of a finite sum of Dirac delta-functions.

[1]The precise functions and parameters used to obtain the plots displayed in figures 1 and 2 are given in Appendix A, page 218.

The transition between the cases of a regular potential and a potential containing singularities is illustrated in Figure 2. The three panels show the behavior of $t^\epsilon - t_0$ with respect to ϵ, where the potential $V^\epsilon = V_0 + q_\epsilon$ satisfies q_ϵ is smooth and V_0 is a sum of smoothed out Dirac delta functions. From right to left, V_0 is an improving approximation of Dirac delta functions.

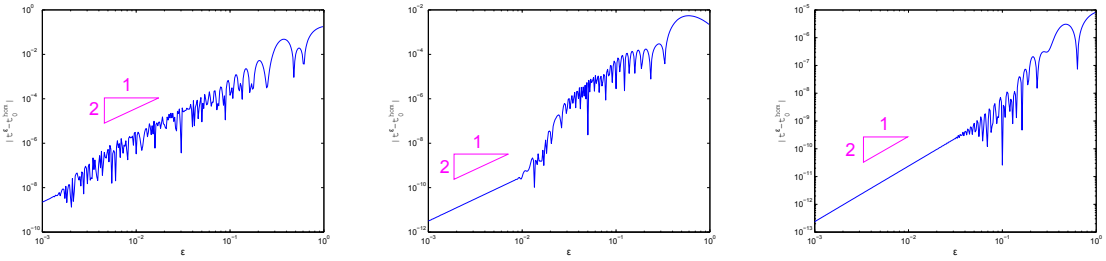


FIGURE 2. Plot of $\log\left|t^\epsilon - t_0^{hom}\right|$ versus $\log\epsilon^{-1}$ for the case of q smooth and V_0 a sum of three approximate Dirac delta functions $\delta_\rho(x) \equiv \frac{1}{\rho\sqrt{\pi}}e^{-x^2/\rho^2}$, with $\rho = 0.001, 0.01, 0.1$.

2.1. Some specific structures. We now study in detail two natural and illustrative classes of potentials:

i. We first consider a one-parameter family of structures, which are truncations of a smooth potential, where for certain parameter ranges the manner of truncation causes a discontinuity. The latter corresponds to cleaving a periodic structure in a manner not commensurate with the background medium:

$$V_1^\epsilon(x;\theta) = \cos\left(\frac{2\pi x}{\epsilon} + \theta\right)\mathbf{1}_{[-1,1]}(x). \tag{2.22}$$

We are obviously in the case related in Remark 2.4, with $V_0 \equiv 0$ (so that $e_{V\pm}(x;k) = e^{\pm ikx}$ and $t_0^{hom} = 1$). More precisely, it is easy to show that

$$\begin{aligned} t_1^\epsilon(k;\theta) &\equiv \frac{1}{4k\pi}\sum_{j=1}^{M} e_{V_0+}(a_j;k)e_{V_0-}(a_j;k)\sum_{|l|\geq 1}[q_l]_{a_j}\frac{e^{2i\pi l\frac{a_j}{\epsilon}}}{l} \\ &= \frac{-i}{2k\pi}\cos(\theta)\sin\left(\frac{2\pi}{\epsilon}\right). \end{aligned}$$

In general, $t_1^\epsilon(k) \neq 0$ but for $\theta = \frac{\pi}{2} + m\pi$, $m \in \mathbb{Z}$, $q_\epsilon(\cdot;\theta)$ is even and therefore for all $k \in \mathbb{R}$ and $\epsilon > 0$, we have $t_1[q_\epsilon](k) = 0$.

ii. Our second example is a piecewise constant and discontinuous structure, which is smoothly truncated:

$$V_2^\epsilon(x;\theta) \equiv h_{per}\left(\frac{x}{\epsilon} + \theta\right) e^{-\frac{x^2}{(x-1)(x+1)}}\mathbf{1}_{[-1,1]}(x), \tag{2.23}$$

with $h_{per}(y)$ the 1-periodic function such that $h(y) = -1$ for $y \in (-1/2, 1/2]$, and $h(y) = 1$ for $y \in (1/2, 3/2]$.

Since the slow-varying part of $q_\epsilon(x)$ is smooth, and V_0 has no singularity, Theorem 2.2 predicts that

$$t^\epsilon - t_0^{hom} = \mathcal{O}(\epsilon^2), \quad \epsilon \to 0, \quad \text{and} \quad \lim_{\epsilon\downarrow 0}\epsilon^{-2}\left(t^\epsilon - t_0^{hom}\right) = t_2^{hom} \text{ is well-defined},$$

even though the function $q_\epsilon(x)$ has internal discontinuities. In Figure 3, we plot $\log\left|t^\epsilon - t_0^{hom}\right|$ versus $\log\epsilon^{-1}$ for the two potentials V_1^ϵ and V_2^ϵ, setting $k = 1$, and $\theta = 0$.

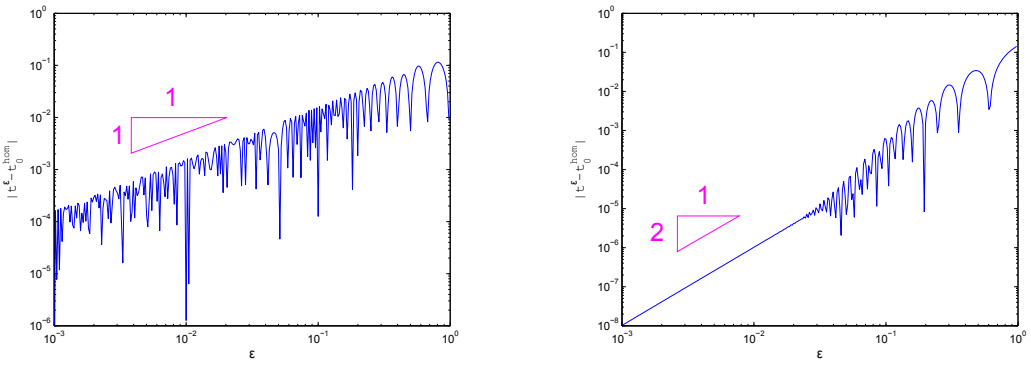


FIGURE 3. Plot of $\log \left| t^\epsilon - t_0^{hom} \right|$ versus $\log \epsilon^{-1}$ for the case of the potentials $V_1^\epsilon(x;\theta)$ in (2.22) (left panel, slope 1), and $V_2^\epsilon(x;\theta)$ in (2.23) (right panel, slope 2). One has $k = 1$, and $\theta = 0$.

3. Background on one-dimensional scattering theory

For simplicity, we consider potentials, W, which have no localized eigenstates, *i.e.* the spectrum of $-\partial_x^2 + W(x)$ is continuous. We further assume that W has the form

$$\begin{aligned} W &= W_{reg} + W_{sing}, \quad \text{with} \\ W_{reg} &\in L^{1,3/2+}(\mathbb{R}), \\ W_{sing} &= \sum_{j=0}^{N-1} c_j \, \delta(x - x_j), \quad \text{where} \quad c_j, x_j \in \mathbb{R}, \quad x_j < x_{j+1}. \end{aligned}$$

We now introduce an appropriate notion of solution to the Schrödinger equation

$$\left(H_W - k^2 \right) u \equiv \left(-\frac{d^2}{dx^2} + W(x) - k^2 \right) u = 0. \tag{3.1}$$

Let $[U]_\xi$ denote the jump in U at the point ξ, *i.e.*

$$[U]_\xi = \lim_{x \to \xi^+} U(x) - \lim_{x \to \xi^-} U(x). \tag{3.2}$$

DEFINITION 3.1. *We say that u is a solution of time-independent Schrödinger equation* (3.1) *if u is piecewise C^2, and satisfies the* (3.1) *on $\mathbb{R} \backslash \operatorname{supp} W_{sing} = \mathbb{R} \backslash \{x_0, \dots, x_{N-1}\}$ as well as the jump conditions*

$$\begin{cases} [u]_x = 0, \quad x \in \mathbb{R}, \\ \left[\frac{d}{dx} u \right]_x = 0 \text{ if } x \in \mathbb{R} \setminus \operatorname{supp} W_{sing}, \\ \left[\frac{d}{dx} u \right]_{x_j} = c_j u(x_j) \text{ where } x_j \in \operatorname{supp} W_{sing}. \end{cases} \tag{3.3}$$

Of special interest are the Jost solutions, defined below.

DEFINITION 3.2. *The Jost solutions $f_\pm(x;k) \equiv m_\pm(x;k) e^{\pm ikx}$ are the unique solutions of* (3.1)*, such that*

$$\lim_{x \to \pm\infty} m_\pm(x;k) = 1.$$

This definition is valid, as we see in Appendix B. We shall use some smoothness and decay properties of these solutions, that are also postponed to Appendix B, for the sake of readability.

With the help of the Jost solutions, we are able to define scattering quantities, as the *transmission and reflection coefficients*, and the *distorted plane waves*.

Since $f_\pm(x;k)$ and $f_\pm(x;-k)$ are solutions of (3.1), and are independent for $k \neq 0$, there exists unique functions $t_\pm(k)$ and $r_\pm(k)$, such that

$$\begin{aligned} f_-(x,k) &= \frac{r_+(k)}{t_+(k)}\, f_+(x,k) + \frac{1}{t_+(k)}\, f_+(x,-k), \\ f_+(x,k) &= \frac{r_-(k)}{t_-(k)}\, f_-(x,k) + \frac{1}{t_-(k)}\, f_-(x,-k). \end{aligned}$$

It is then easy to check that $t_+(k) = t_-(k) \equiv t(k)$, and that $t(k)$ and $r_\pm(k)$ are continous at $k = 0$. The distorted plane waves $e_{W\pm}(x;k)$ are then defined by:

Definition 3.3. *Given a potential $W(x)$, we define $e_{W\pm}(x;k)$, the distorted plane waves associated with H_W by*

$$e_+(x;k) \equiv t(k) f_+(x;k) \equiv t(k) m_+(x;k) e^{ikx} \tag{3.4}$$

$$e_-(x;k) \equiv t(k) f_-(x;k) \equiv t(k) m_-(x;k) e^{-ikx}. \tag{3.5}$$

The distorted plane waves $e_{W\pm}(x;k)$ play the role for H_W that the plane waves $e^{\pm ikx}$ play for $H_0 = -\partial_x^2$, as we see below. Let us first introduce the notion of *outgoing radiation* as $|x| \to \infty$.

Definition 3.4. *$U(x)$ is said to satisfy an outgoing radiation condition or to be outgoing as $|x| \to \infty$ if*

$$(\partial_x \mp ik) U \to 0, \quad \text{as } x \to \pm\infty.$$

Proposition 3.5. *Given a potential $W(x)$, the distorted plane waves $e_{W\pm}(x;k)$ are the unique solutions of (3.1) satisfying*

$$e_{W\pm}(x;k) = e^{\pm ikx} + \text{outgoing}(x). \tag{3.6}$$

More precisely, they satisfy the following asymptotic relations [**135**]:

$$\begin{cases} e_{W+}(x;k) - \left(e^{ikx} + r_+(k)e^{-ikx}\right) \longrightarrow 0 & \text{as } x \to -\infty, \\ e_{W+}(x;k) - t(k)e^{ikx} \longrightarrow 0 & \text{as } x \to +\infty, \\ e_{W-}(x;k) - t(k)e^{-ikx} \longrightarrow 0 & \text{as } x \to -\infty, \\ e_{W-}(x;k) - \left(e^{-ikx} + r_-(k)e^{ikx}\right) \longrightarrow 0 & \text{as } x \to +\infty. \end{cases} \tag{3.7}$$

A consequence of the relations (3.7) is the Wronskian identity:

$$Wron\left(e_{W+}(\cdot;k), e_{W-}(\cdot;k)\right) \equiv e_{W+}\partial_x e_{W-} - \partial_x e_{W+} e_{W-} = -2ik\, t(k). \tag{3.8}$$

In terms of the Jost solutions:

$$Wron\left(f_+(\cdot;k), f_-(\cdot;k)\right) = -\frac{2ik}{t(k)}, \quad k \neq 0. \tag{3.9}$$

By analyticity in W, potentials for which $Wron\left(f_+(\cdot;k), f_-(\cdot;k)\right)|_{k=0} = 0$ are isolated in the space of potentials.

Definition 3.6. *A potential W is said to be generic if*

$$Wron\left(f_+(x;0), f_-(x;0)\right) = Wron\left(m_+(x;0), m_-(x;0)\right) \neq 0.$$

Otherwise, the operator H_W is said to have a zero-energy resonance, i.e. *$H_W u = 0$ has a non-trivial solution that is bounded both as $x \to \infty$ and as $x \to -\infty$.*

Note that the potential $W(x) \equiv 0$ is not generic since $m_+(x;k) \equiv m_-(x;k) \equiv 1$. If W is generic, we have [**135, 165, 178**]

$$t(k) = -\frac{2ik}{Wron\left(f_+(\cdot;0), f_-(x;0)\right)} + o(k) = \mathcal{O}(k), \quad |k| \to 0. \tag{3.10}$$

In particular, $t(0) = 0$ and $r_\pm(0) = -1$.

A simple calculation then yields the following expressions for the outgoing Green's function (resolvent kernel) and the outgoing resolvent, $R_W(k),\ k \neq 0$:

$$R_W(x,y;k) \;=\; \begin{cases} \frac{1}{-2ik\ t(k)}\ e_{W-}(y;k)\ e_{W+}(x;k), & y < x, \\ \\ \frac{1}{-2ik\ t(k)}\ e_{W-}(x;k)\ e_{W+}(y;k), & y > x, \end{cases} \tag{3.11}$$

$$R_W(k)F(x) \;=\left(-\frac{d^2}{dx^2} + W(x) - k^2\right)^{-1} F(x) \;=\; \int_{-\infty}^{\infty} R_W(x,\zeta;k)F(\zeta)\mathrm{d}\zeta. \tag{3.12}$$

Remark 3.7. *Note that these expressions, originally defined for $k \neq 0$, are easily extended to the point $k = 0$, for generic potentials. Indeed, one has by Definition 3.6:*

$$\frac{1}{-2ik\ t(k)}\ e_{W-}(y;k)\ e_{W+}(x;k) \;=\; \frac{f_-(y;k)\ f_+(x;k)}{Wron\left(f_+(\cdot;k), f_-(\cdot;k)\right)}.$$

In the generic case, this expression has a limit when $k \to 0$ by (3.9) *and* (3.10)*. In the following, we work with the distorted plane waves, which sometimes lead to expressions which are only defined for $k \neq 0$. By the above considerations, it is easy to check that in the case of a generic potential, these expressions have a well-defined finite limit when $k \to 0$.*

In particular, we have the following

Proposition 3.8. *Let $F \in L^1(\mathbb{R})$. Assume $W(x)$ satisfying Hypotheses **(V)** and $k \in K$ satisfying Hypothesis **(G)**. Then the inhomogeneous equation*

$$\left(-\frac{d^2}{dx^2} + W(x) - k^2\right)\ U \;=\; F \tag{3.13}$$

has the unique outgoing solution $U = R_W(k)F$. Moreover, $\|U\|_{L^\infty} \leq C\ \|F\|_{L^1}$, with a constant, $C(K)$.

Proof. Existence follows from the explicit integral representation (3.12). Note that if W is generic, then $R_W(k)F$ is defined for any $k \in \mathbb{R}$, whereas in the non-generic case, $Wron\left(f_+(x;k), f_-(x;k)\right) \longrightarrow 0$ $(k \to 0)$ and $f_\pm(x,k)$ does not tend to zero as $k \to 0$ **[135]**, so that $R_W(k)F$ has a simple pole at $k = 0$.

To prove uniqueness, note that if the difference, $d(x)$, of two solutions is non-zero, then $d(x)$ is a non-trivial solution of the scattering resonance problem, that is $(H_W - k^2)d = 0$, $d(x)$ outgoing at $|x| \to \infty$ with scattering resonance energy $k^2 \in \mathbb{R}$. However, the scattering resonance energies must satisfy $\Im(k^2) < 0$; see, for example, **[176]**. Therefore, $d(x) \equiv 0$. This completes the proof. □

4. Homogenization / Multiple Scale Perturbation Expansion

4.1. Multiple scale expansion. In this section, our goal is to formally obtain the expansion displayed in Theorem 2.2, using a systematic two-scale / homogenization perturbation scheme. A proof (and derivation by other means) of this expansion is presented in section 5.

We seek a solution of the equation

$$\left(-\frac{d^2}{dx^2} + V_0(x) + q\left(x, \frac{x}{\epsilon}\right) - k^2\right) e_{V^\epsilon+}(x;k) = 0, \tag{4.1}$$

in the form of a two-scale function, $e_{V^\epsilon+}(x;k) = U^\epsilon(x, \frac{x}{\epsilon})$ which satisfies the jump conditions (3.3) and the outgoing radiation condition of Definition 3.6. Treating x and y as independent variables, we find that $U^\epsilon(x,y)$ is a solution of

$$\left(-\left(\frac{\partial}{\partial x}+\frac{1}{\epsilon}\frac{\partial}{\partial y}\right)^2 + V_0(x) + q(x,y) - k^2\right) U^\epsilon(x,y) = 0. \tag{4.2}$$

We then formally expand $U^\epsilon(x,y)$ as

$$U^\epsilon(x,y) = \sum_{j=0}^{\infty} \epsilon^j\ U_j(x,y), \tag{4.3}$$

and require that

$$\begin{aligned} &U_j(x,y+1) \ = \ U_j(x,y), \quad j \geq 0, \\ &U_0(x,y) - e^{ikx}, \quad U_j(x,y) \quad j \geq 1 \ \text{ outgoing as } |x| \to \infty, \\ &U_j(x,y)|_{y=x/\epsilon} \ \text{ satisfies jump conditions (3.3).} \end{aligned} \tag{4.4}$$

The problem is solved by substituting the expansion (4.3) into (4.2) and imposing the equation, jump conditions and radiation condition at each order in ϵ. The differential equation becomes

$$\left(-\left(\frac{\partial}{\partial x}+\frac{1}{\epsilon}\frac{\partial}{\partial y}\right)^2 + V_0(x) + q(x,y) - k^2\right)\ U^\epsilon\left(x, \frac{x}{\epsilon}\right) = \sum_{j=-2}^{\infty}\ \epsilon^j\ r_j \ = \ 0, \tag{4.5}$$

implying the following hierarchy of equations at each order in ϵ

$$\mathcal{O}(\epsilon^{-2}) \qquad r_{-2} = -\partial_y^2 U_0 \ = \ 0, \tag{4.6a}$$

$$\mathcal{O}(\epsilon^{-1}) \qquad r_{-1} = -\partial_y^2 U_1 - 2\partial_y\partial_x U_0 \ = \ 0, \tag{4.6b}$$

$$\mathcal{O}(\epsilon^{0}) \qquad r_{0} = -\partial_y^2 U_2 - 2\partial_y\partial_x U_1 - \partial_x^2 U_0 + (V_0+q)U_0 - k^2 U_0 = 0, \tag{4.6c}$$

$$\mathcal{O}(\epsilon^{1}) \qquad r_{1} = -\partial_y^2 U_3 - 2\partial_y\partial_x U_2 - \partial_x^2 U_1 + (V_0+q)U_1 - k^2 U_1 \ = \ 0, \tag{4.6d}$$

$$\mathcal{O}(\epsilon^{2}) \qquad r_{2} = -\partial_y^2 U_4 - 2\partial_y\partial_x U_3 - \partial_x^2 U_2 + (V_0+q)U_2 - k^2 U_2 \ = \ 0, \tag{4.6e}$$

$$\mathcal{O}(\epsilon^{3}) \qquad r_{3} = -\partial_y^2 U_5 - 2\partial_y\partial_x U_4 - \partial_x^2 U_3 + (V_0+q)U_3 - k^2 U_3 \ = \ 0, \tag{4.6f}$$

$$\cdots \quad \cdots$$

$$\mathcal{O}(\epsilon^{j}) \qquad r_{j} = r_j\left[U_{j+2}, U_{j+1}, U_j\right] \ = \ 0\ . \tag{4.6g}$$

For example, to construct an approximate solution of (4.2) satisfying (4.4) up to the order 3, we solve simultaneously the equations $r_j = 0$ for $j = -2, \cdots, 3$. This will determine the functions U_0, U_1, U_2 and U_3 which make U^ϵ an approximate solution through order $\mathcal{O}(\epsilon^3)$. Since $e_{V^\epsilon+}(x;k) - e^{ikx}$ is to be outgoing, we require $U_0 - e^{ikx}$ and each U_i $(i = 1, \ldots, 3)$ to satisfy the outgoing condition. We now proceed with the implementation.

Caveat lector! *The formal expansion presented in the remainder of this section yields terms involving spatial derivatives of $e_{V_0+}(x;k)$ and $q_j(x)$ of arbitrarily high order. Now $\partial_x e_{V_0+}(x;k)$ has jump discontinuities on* $\operatorname{supp} V_{sing}$ *and $q_j(x)$ has jump discontinuities. Hence, the expansion must viewed in a distributional sense, e.g. involving terms, such as $\partial_x^\alpha \delta(x - x_j)$ etc. Furthermore, when we impose the jump conditions* (3.3) *to the expansion, order by order in ϵ, we shall throughout assign $[\partial_x^\alpha \delta(x - x_j)]_{x=x_j} = 0$. Although seemingly risky, in section 5 we give a complete rigorous proof of the expansion with error bounds.*

Beginning at $\mathcal{O}(\epsilon^{-2})$, one has from (4.6a)

$$r_{-2} = 0 \Longrightarrow \partial_y^2 U_0 = 0 \Longrightarrow U_0(x,y) = U_0(x). \tag{4.7}$$

Consequently, one has from (4.6b)

$$r_{-1} = 0 \Longrightarrow \partial_y^2 U_1 = -2\partial_y \partial_x U_0 = 0 \Longrightarrow U_1(x,y) = U_1(x). \tag{4.8}$$

Recall that $y \mapsto q(x,y)$ is 1-periodic and $\int_0^1 q(x,y)\ \mathrm{d}y = 0$. Integration of the equation (4.6c) with respect to y yields:

$$\int_0^1 r_0(x,y)\ \mathrm{d}y = 0 \Longrightarrow -\frac{d^2}{dx^2}U_0(x) + V_0(x)U_0(x) - k^2 U_0(x) = 0. \tag{4.9}$$

Furthermore, since $U_0 - e^{ikx}$ is outgoing, one has by Proposition 3.5

$$U_0(x) \equiv e_{V_0+}(x;k). \tag{4.10}$$

By (4.9) and (4.6c) leads to

$$r_0 = 0 \iff -\partial_y^2 U_2(x,y) + q(x,y)e_{V_0+}(x) = 0. \tag{4.11}$$

Thus, we decompose U_2 as:

$$U_2 = U_2^{(h)}(x) + U_2^{(p)}(x,y),$$

with $U_2^{(p)}(x,y)$ a particular solution, and $U_2^{(h)}(x)$ an homogeneous solution to be determined.

Again, since $y \mapsto q(x,y)$ is 1-periodic and $\int_0^1 q(x,y)\ \mathrm{d}y = 0$, when by (4.6d),

$$\int_0^1 r_1(x,y)\ \mathrm{d}y = 0 \Longrightarrow -\frac{d^2}{dx^2}U_1(x) + V_0(x)U_1(x) - k^2 U_1(x) = 0. \tag{4.12}$$

Since U_1 is outgoing, we claim

$$U_1 \equiv 0. \tag{4.13}$$

Indeed, in this case k^2 is a scattering resonance energy and U_1 its corresponding mode. Scattering resonances necessarily satisfy $\Im k^2 < 0$ **[176]**. However, $k^2 \in \mathbb{R}$ and hence $U_1 \equiv 0$.

Consequently,

$$r_1 = 0 \Longrightarrow -\partial_y^2 U_3(x,y) - 2\partial_y \partial_x U_2 = 0. \tag{4.14}$$

In the same way as for U_2, we decompose U_3 as

$$U_3 = U_3^{(h)}(x) + U_3^{(p)}(x,y),$$

with $U_3^{(p)}(x,y)$ a particular solution, and $U_3^{(h)}(x)$ an homogeneous solution to be determined.

Integration of the equations (4.6e) and (4.6f) with respect to y, respectively, yields:

$$-\frac{d^2}{dx^2}U_2^{(h)}(x) + V_0(x)U_2^{(h)}(x) + \int_0^1 q(x,y)U_2^{(p)}(x,y)\mathrm{d}y - k^2 U_2^{(h)}(x) = 0. \tag{4.15}$$

$$-\frac{d^2}{dx^2}U_3^{(h)}(x) + V_0(x)U_3^{(h)}(x) - k^2 U_3^{(h)}(x) = -\int_0^1 U_3^{(p)}(x,y)q(x,y)\ \mathrm{d}y. \tag{4.16}$$

We now solve (4.11), (4.14), (4.15) and (4.16) to obtain a unique (approximate) solution satisfying both outgoing and jump conditions, as we see in the following. First, we use the decomposition in Fourier series of $q(x,y)$ in y :

$$q(x,y) = \sum_{j\neq 0} q_j(x) e^{2i\pi jy}.$$

Consequently, equation (4.11) leads immediately to

$$U_2^{(p)}(x,y) = -\frac{e_{V_0+}(x;k)}{4\pi^2}\sum_{|j|\geq 1}\frac{q_j(x)}{j^2}e^{2i\pi jy}. \tag{4.17}$$

From (4.14), one deduces

$$\partial_y^2 U_3(x,y) = \frac{i}{\pi}\sum_{|j|\geq 1}\frac{\partial_x(e_{V_0+}(x;k)q_j(x))}{j}e^{2i\pi jy}.$$

A particular solution $U_3^{(p)}(x,y)$ is therefore given by

$$U_3^{(p)}(x,y) = -\frac{i}{4\pi^3}\sum_{|j|\geq 1}\frac{\partial_x(e_{V_0+}(x;k)q_j(x))}{j^3}e^{2i\pi jy}. \tag{4.18}$$

Then, using the Fourier series of q and $U_2^{(p)}$, we obtain the following equations from (4.15) and (4.16):

$$-\frac{d^2}{dx^2}U_2^{(h)}(x) + V_0(x)U_2^{(h)}(x) - k^2U_2^{(h)}(x) = \frac{e_{V_0+}(x;k)}{4\pi^2}\sum_{|j|\geq 1}\frac{|q_j(x)|^2}{j^2}, \quad \text{and} \tag{4.19}$$

$$\begin{aligned}&-\frac{d^2}{dx^2}U_3^{(h)}(x) + V_0(x)U_3^{(h)}(x) - k^2U_3^{(h)}(x)\\ &\quad = -\int_0^1 U_3^{(p)}(x,y)q(x,y)\,\mathrm{d}y \;=\; \frac{i}{4\pi^3}\sum_{|j|\geq 1}\frac{\partial_x(e_{V_0+}(x;k)q_j(x))q_{-j}(x)}{j^3}.\end{aligned} \tag{4.20}$$

By Proposition 3.8, equations (4.19) and (4.20) have unique outgoing solutions. We refer to the expansion of U^ϵ obtained in this way as the

Bulk (homogenization) expansion:

$$U^\epsilon(x,y) \;=\; e_{V_0+}(x;k) \;+\; \epsilon^2\left(U_2^{(p)}(x,y) + U_2^{(h)}(x)\right) \;+\; \epsilon^3\left(U_3^{(p)}(x,y) + U_3^{(h)}(x)\right) \;+\; \dots \tag{4.21}$$

It consists of a leading order average term (homogenization) plus correctors at each order in ϵ due to microstructure.

Failure of Jump conditions at interfaces:

Recall that we seek a solution which satisfies the jump conditions (3.3) on $U^\epsilon(x,y)$ for all $(x,y) = (x, x/\epsilon)$ at each order in ϵ. The leading order term, e_{V_0+} satisfies all jump conditions. Now consider the terms $U_j^{(p)}(x,y) + U_j^{(h)}(x)$, arising at order $\mathcal{O}(\epsilon^j)$. By construction, $U_j^{(h)}$ satisfies (3.3). However $U_j^{(p)}(x, x/\epsilon)$ does not. Indeed, for the cases $j = 2, 3$, referring to expressions (4.17) and (4.18) we observe violation of (3.3) in $U_j^{(p)}(x, x/\epsilon)$ at discontinuities of $q_j(x)$ and $e_{V_0+}(x;k)$, and their derivatives.

More precisely, the jump conditions for $U_2^{(p)}$ fail at a_l $(l = 1, \dots, M)$ each point of discontinuity of $q(x, x/\epsilon)$, since one has

$$\left[U_2^{(p)}(x, \frac{x}{\epsilon})\right]_a = F_{2,a}^\epsilon, \tag{4.22}$$

$$\left[\frac{d}{\mathrm{d}x}U_2^{(p)}(x, \frac{x}{\epsilon})\right]_a = \frac{1}{\epsilon}G_{2,a}^\epsilon \;+\; H_{2,a}^\epsilon. \tag{4.23}$$

with

$$
\begin{aligned}
F^\epsilon_{2,a} &\equiv \frac{-1}{4\pi^2}\sum_{|j|\geq 1} e_{V_0+}(a;k)\,[q_j]_a\,\frac{e^{2i\pi ja/\epsilon}}{j^2},\\
G^\epsilon_{2,a} &\equiv \frac{-i}{2\pi}\sum_{|j|\geq 1} e_{V_0+}(a;k)\,[q_j]_a\,\frac{e^{2i\pi ja/\epsilon}}{j},\\
H^\epsilon_{2,a} &\equiv \frac{-1}{4\pi^2}\sum_{|j|\geq 1} [\partial_x(e_{V_0+}(x;k)q_j(x))]_a\,\frac{e^{2i\pi ja/\epsilon}}{j^2}.
\end{aligned}
\tag{4.24}
$$

In the same way, the jump conditions for $U_3^{(p)}$ fail at points of discontinuity of the functions $q(x,x/\epsilon)$ and $\partial_x q(x,x/\epsilon)$, and for $x\in\{x_0,\cdots,x_{N-1}\}$ the support of V_{sing} (recall: $V_{sing}=\sum_{j=0}^{N-1} c_j\ \delta(x-x_j)$):

$$
\left[U_3^{(p)}(x,\frac{x}{\epsilon})\right]_a = F^\epsilon_{3,a},
\tag{4.25}
$$

$$
\left[\frac{d}{\mathrm{d}x}U_3^{(p)}(x,\frac{x}{\epsilon})\right]_a = \frac{1}{\epsilon}G^\epsilon_{3,a}\ +\ H^\epsilon_{3,a},
\tag{4.26}
$$

with $F^\epsilon_{3,a}$ and $H^\epsilon_{3,a}$ bounded highly oscillating functions and

$$
G^\epsilon_{3,a} \equiv \frac{1}{2\pi^2}\sum_{|j|\geq 1} [\partial_x(e_{V_0+}(x;k)q_j(x))]_a\,\frac{e^{2i\pi ja/\epsilon}}{j^2}.
\tag{4.27}
$$

$F^\epsilon_{3,a}$ and $H^\epsilon_{3,a}$ can be made explicit, but we omit these expressions as they contribute only at $\mathcal{O}(\epsilon^3)$.

Restoring the Jump Conditions at interfaces:

In order to restore the jump conditions (3.3), we must add to the expansion, at each point where the jump conditions are not satisfied, an appropriate *corrector*. These correctors each solve a non-homogeneous equation, driven by the jumps in the bulk expansion (4.21).

To see this, first note that $\left[\ \frac{d}{\mathrm{d}x}U_j^{(p)}\ \right]_a = \mathcal{O}(\epsilon^{-1}),\ j=2,3$. Since $U_j^{(p)}$ contributes at order ϵ^j, this suggests adding a corrector at order ϵ^{j-1}. Thus, we introduce the

Bulk expansion with corrector terms:

$$
\begin{aligned}
U^\epsilon(x,y) \ =\ e_{V_0+}(x;k)\ +\ \epsilon\,\mathcal{U}_1^\epsilon(x)\ +\ \epsilon^2\Big(\ U_2^{(p)}(x,y)+U_2^{(h)}(x)+\mathcal{U}_2^\epsilon(x)\Big)\\
+\ \epsilon^3\Big(\ U_3^{(p)}(x,y)+U_3^{(h)}(x)+\mathcal{U}_3^\epsilon(x)\Big)\ +\ \ldots
\end{aligned}
\tag{4.28}
$$

The interface correctors $\mathcal{U}_j^\epsilon(x)$ are to be determined so that, at each order in ϵ, the expansion (4.28) *satisfies the jump conditions* (3.3)*, the differential equation* (4.1) *and outgoing radiation condition.*

We construct $\mathcal{U}_j^\epsilon(x),\ j=1,2$ below. The general construction uses the following

LEMMA 4.1. *Let $F_1,F_2\in\mathbb{R}$ and $V_0=V_{sing}+V_{reg}$ as in* (2.1)*. Then there exists $\mathcal{U}(x)$, an outgoing piecewise C^2 solution of:*

$$
\left(-\frac{d^2}{dx^2}+V_0(x)-k^2\right)\mathcal{U}=0,\ \textit{for } x<a \textit{ and } x>a,
\tag{4.29}
$$

which also satisfies the following jump conditions at the point $x=a$:

$$
\begin{aligned}
[\mathcal{U}(x)]_a \ &=\ F_1,\\
\left[\frac{d}{dx}\mathcal{U}(x)\ \right]_a\ -\ c\,\mathcal{U}(a-)\ &=\ F_2.
\end{aligned}
$$

Here, $\mathcal{U}(a-) = \lim_{x\uparrow a}\mathcal{U}(x)$*, and the constant*

$$c = \begin{cases} 0 & \text{if } a \notin \operatorname{supp} V_{sing}, \\ c_{j_0} & \text{if } a = x_{j_0} \in \operatorname{supp} V_{sing}\ ; \end{cases} \tag{4.30}$$

recall $V_{sing}(x) = \sum_{j=0}^{N-1} c_j\ \delta(x - x_j)$.

$\mathcal{U}(x)$ *has the form*

$$\mathcal{U}(x) = \begin{cases} \alpha e_{V_0-}(x;k) & \text{if } x < a, \\ \beta e_{V_0+}(x;k) & \text{if } x > a, \end{cases} \tag{4.31}$$

for appropriate choice of α *and* β*, namely*

(4.32)
$$\alpha = \frac{F_2 e_{V_0+}(a;k)\ -\ F_1\partial_x e_{V_0+}(a+;k)}{2ik\ t_0^{hom}(k)} \quad \text{and} \quad \beta = \frac{F_2 e_{V_0-}(a;k)\ -\ F_1\partial_x e_{V_0-}(a+;k)}{2ik\ t_0^{hom}(k)}.$$

Before giving the proof, we explain why choosing $\mathcal{U}_j^\epsilon$ as in Lemma 4.1 does not change the bulk expansion (4.21) constructed above. Therefore, our approach which first computes the bulk-expansion and then the correctors is consistent.

As pointed out the expressions in the bulk expansion (4.21)

$$U_j(x,y)\ =\ U_j^{bulk}(x,y)\ \equiv\ U_j^{(p)}(x,x/\epsilon) + U_j^{(h)}(x) \tag{4.33}$$

do not satisfy jump conditions (3.3). Suppose now that we replace the functions $U_j(x,y) = U_j^{bulk}(x,y)$ by $U_j(x,y) = U_j^{bulk}(x,y) + \mathcal{U}_a^\epsilon(x)$ and we seek $\mathcal{U}_a^\epsilon(x)$ so as to ensure jump conditions (3.3). (Assume only one corrector is required). Note that since $\mathcal{U}_a^\epsilon(x)$ lies in the kernel of ∂_y, adding such a term has no effect on the equations determining $U_j^{(p)}(x,y)$. Further, we want to preserve the form of $U_j^{(h)}(x)$, which has previously been constructed. Thus,

$$\begin{aligned} r_j\left[U_{j+2}, U_{j+1}, U_j^{bulk} + \mathcal{U}_a^\epsilon\right]\ &= r_j\left[U_{j+2}, U_{j+1}, U_j^{bulk}\right] \\ &\quad + \left(-\partial_x^2 + V_0(x) - k^2\right)\mathcal{U}_a^\epsilon(x)\ +\ q(x,y)\mathcal{U}_a^\epsilon(x). \end{aligned} \tag{4.34}$$

The equation for $U_j^{(h)}(x)$ is obtained by averaging (4.34) with respect to y. Since $q(x,y)$ has mean zero with respect to y, this gives

$$\int_0^1 r_j\left[U_{j+2}, U_{j+1}, U_j^{bulk}\right](x,y)\ \mathrm{d}y\ +\ \left(-\partial_x^2 + V_0(x) - k^2\right)\mathcal{U}_a^\epsilon(x)\ =\ 0. \tag{4.35}$$

Thus, if we choose $\mathcal{U}_a^\epsilon(x)$ to satisfy (4.29), then the second term in (4.35) vanishes and the equation for $U_j^{(h)}(x)$ is preserved. Therefore, if Lemma 4.1 is used to determine the jump-driven correctors at each order in ϵ, then the corrected bulk expansion (4.28) is the solution we seek.

Proof of Lemma 4.1: The piecewise form of $\mathcal{U}$ (4.31) satisfies the outgoing radiation condition, by construction. The constants α and β are determined by the jump conditions.

Using the fact that $e_{V_0+}(x;k)$ and $e_{V_0-}(x;k)$ satisfy the jump conditions (3.3), one has

$$\begin{cases} & [\mathcal{U}(x)]_a & = \beta e_{V_0+}(a;k) - \alpha e_{V_0-}(a;k), \\ \left[\frac{d}{dx}\mathcal{U}(x)\right]_a & - c\mathcal{U}(a-) & = \beta\partial_x e_{V_0+}(a+;k) - \alpha\partial_x e_{V_0-}(a-;k) - c\alpha\ e_{V_0-}(a-;k) \\ & & = \beta\partial_x e_{V_0+}(a+;k) - \alpha\partial_x e_{V_0-}(a+;k). \end{cases}$$

Solving this inhomogeneous system, using the value of the Wronskian, given in (3.8), leads immediately to (4.32). This completes the proof of Lemma 4.1. □

We now proceed to apply Lemma 4.1 to determine the correctors associated with $U_2^{(p)}$ and $U_3^{(p)}$. Using (4.22)-(4.23) and (4.25)-(4.26), the jump conditions (3.3) applied to $\epsilon\mathcal{U}_1^\epsilon + \epsilon^2\mathcal{U}_2^\epsilon + \epsilon^2 U_2^{(p)} + \epsilon^3 U_3^{(p)}$ read:

$$\epsilon\,[\,\mathcal{U}_1^\epsilon\,]_a + \epsilon^2\big(\,F_{2,a}^\epsilon + [\,\mathcal{U}_2^\epsilon\,]_a\,\big) = \mathcal{O}(\epsilon^3), \tag{4.36}$$

$$\epsilon\Big(G_{2,a}^\epsilon + \Big[\,\frac{d}{\mathrm{d}x}\mathcal{U}_1^\epsilon\,\Big]_a - c\mathcal{U}_1^\epsilon(a-)\Big) + \epsilon^2\Big(H_{2,a}^\epsilon - c\,U_2^{(p)}(a-) + G_{3,a}^\epsilon + \Big[\,\frac{d}{\mathrm{d}x}\mathcal{U}_2^\epsilon\,\Big]_a - c\,\mathcal{U}_1^\epsilon(a-)\Big) = \mathcal{O}(\epsilon^3). \tag{4.37}$$

Equations (4.36) and (4.37) imply jump conditions at order ϵ and order ϵ^2. Therefore, we construct $\mathcal{U}_{j,a}^\epsilon$, $j = 1, 2$ solving the two inhomogeneous problems at each point, a, of non-smoothness.

System for corrector $\mathcal{U}_{1,a}^\epsilon$:

$$\left(-\frac{d^2}{dx^2} + V_0(x) - k^2\right)\mathcal{U}_{1,a}^\epsilon = 0, \quad x \neq a, \tag{4.38}$$

$$\big[\,\mathcal{U}_{1,a}^\epsilon\,\big]_a = 0, \quad \left[\frac{d}{\mathrm{d}x}\mathcal{U}_{1,a}^\epsilon\right]_a - c\,\mathcal{U}_{1,a}^\epsilon(a-) = -G_{2,a}^\epsilon. \tag{4.39}$$

System for corrector $\mathcal{U}_{2,a}^\epsilon$:

$$\left(-\frac{d^2}{dx^2} + V_0(x) - k^2\right)\mathcal{U}_{2,a}^\epsilon = 0, \quad x \neq a, \tag{4.40}$$

$$\big[\,\mathcal{U}_{2,a}^\epsilon\,\big]_a = -F_{2,a}^\epsilon, \quad \left[\frac{d}{\mathrm{d}x}\mathcal{U}_{2,a}^\epsilon(x)\right]_a - c\,\mathcal{U}_{2,a}^\epsilon(a-) = -H_{2,a}^\epsilon - G_{3,a}^\epsilon + c\,U_2^{(p)}(a-). \tag{4.41}$$

Lemma 4.1, applied to (4.38)-(4.39) and (4.40)-(4.41) defines the unique correctors $\mathcal{U}_{1,a}^\epsilon$ and $\mathcal{U}_{2,a}^\epsilon$: $\mathcal{U}_{1,a}^\epsilon$ is given by (4.31), *i.e.*

$$\mathcal{U}_{1,a}^\epsilon(x) = \begin{cases} \alpha_{1,a}^\epsilon\, e_{V_0-}(x;k) & \text{if } x < a, \\ \beta_{1,a}^\epsilon\, e_{V_0+}(x;k) & \text{if } x > a, \end{cases} \tag{4.42}$$

with $\alpha_{1,a}^\epsilon$ and $\beta_{1,a}^\epsilon$ given by

$$\alpha_{1,a}^\epsilon = -\frac{G_{2,a}^\epsilon}{2ikt_0^{hom}(k)}\, e_{V_0+}(a;k), \quad \beta_{1,a}^\epsilon = -\frac{G_{2,a}^\epsilon}{2ikt_0^{hom}(k)}\, e_{V_0-}(a;k), \tag{4.43}$$

where $G_{2,a}^\epsilon$ is given in (4.24). Then, $\mathcal{U}_{2,a}^\epsilon$ is given by (4.31) with $\alpha_{2,a}^\epsilon$ and $\beta_{2,a}^\epsilon$ α and β given by

$$\alpha_{2,a}^\epsilon = \frac{1}{2ikt_0^{hom}(k)}\left(\Big(-H_{2,a}^\epsilon - G_{3,a}^\epsilon + c\,U_2^{(p)}(a-)\Big)\,e_{V_0+}(a;k) + F_{2,a}^\epsilon \partial_x e_{V_0+}(a+;k)\right),$$

$$\beta_{2,a}^\epsilon = \frac{1}{2ikt_0^{hom}(k)}\left(\Big(-H_{2,a}^\epsilon - G_{3,a}^\epsilon + c\,U_2^{(p)}(a-)\Big)\,e_{V_0-}(a;k) + F_{2,a}^\epsilon \partial_x e_{V_0-}(a+;k)\right), \tag{4.44}$$

where $H_{2,a}^\epsilon$, $F_{2,a}^\epsilon$ and $G_{3,a}^\epsilon$ are given in (4.24) and (4.27).

Therefore at $\mathcal{O}(\epsilon)$, we define the corrector $\mathcal{U}_1^\epsilon$ as

$$\mathcal{U}_1^\epsilon = \sum_{j=1}^{M} \mathcal{U}_{1,a_j}^\epsilon \tag{4.45}$$

where a_j, $j = 1, \ldots, M$ denote the points of discontinuity of $q\,(x, x/\epsilon)$.

At order $\mathcal{O}(\epsilon^2)$, we have a violation of the jump conditions (3.3) due to

(i) points of "discontinuity" of $q_j(x)$, *i.e.* a_j $(j = 1, \dots, M)$ for which $[q_j]_a \neq 0$ or $[\partial_x q_j]_a \neq 0$, and

(ii) the singular set $\operatorname{supp} V_{sing} = \{x_0, \dots, x_{N-1}\}$.

Thus we construct, $\mathcal{U}^\epsilon_{2,a}$, for all a in the set, Ω, of non-smooth points of $V^\epsilon(x)$:

$$\Omega = \{x_0, \dots, x_{N-1}\} \cup \{-\infty = a_0, a_1, \dots, a_M = \infty\} \tag{4.46}$$

and define the corrector $\mathcal{U}^\epsilon_2$ by:

$$\mathcal{U}^\epsilon_2 = \sum_{a\in\Omega} \mathcal{U}^\epsilon_{2,a}. \tag{4.47}$$

We summarize the preceding calculation in the following

Proposition 4.2.

$$\begin{aligned} e_{V^\epsilon+}(x;k) \; &= U^\epsilon(x, x/\epsilon) = e_{V_0+}(x;k) \; + \; \epsilon\, \mathcal{U}^\epsilon_1(x) \\ &+ \; \epsilon^2 \; \Big(\; U^{(h)}_2(x) \; + \; U^{(p)}_2(x, x/\epsilon) \; + \; \mathcal{U}^\epsilon_2(x) \; \Big) + \mathcal{O}(\epsilon^3) \end{aligned} \tag{4.48}$$

gives a formal construction of the distorted plane wave $e_{V^\epsilon+}(x;k)$, through $\mathcal{O}(\epsilon^2)$ with error of size $\mathcal{O}(\epsilon^3)$. The correctors $\mathcal{U}^\epsilon_1(x)$ and $\mathcal{U}^\epsilon_2(x)$ are given by (4.38)-(4.39) *and* (4.40)-(4.41).

Finally, $U^{(p)}_j(x,y)$ and $U^{(h)}_j(x)$ are given by

$$U^{(p)}_2(x,y) = -\frac{e_{V_0+}(x;k)}{4\pi^2} \sum_{|j|\geq 1} q_j(x) \frac{e^{2i\pi jy}}{j^2},$$

$$\left(-\frac{d^2}{dx^2} + V_0(x) - k^2\right) U^{(h)}_2(x) = \frac{e_{V_0+}(x;k)}{4\pi^2} \sum_{|j|\geq 1} \frac{|q_j(x)|^2}{j^2}, \quad U^{(h)}_2 \text{ outgoing},$$

$$U^{(p)}_3(x,y) = -\frac{i}{4\pi^3} \sum_{|j|\geq 1} \partial_x(e_{V_0+}(x;k) q_j(x)) \frac{e^{2i\pi jy}}{j^3},$$

$$\left(-\frac{d^2}{dx^2} + V_0(x) - k^2\right) U^{(h)}_3(x) = i \sum_{|j|\geq 1} \frac{\partial_x(e_{V_0+}(x;k) q_j(x)) q_{-j}(x)}{4\pi^3\, j^3}, \; U^{(h)}_3 \text{ outgoing}.$$

4.2. Expansion of the transmission coefficient, $t^\epsilon(k)$. The results of the previous section can now be used to derive expansion (2.13) for the transmission coefficient, $t^\epsilon(k)$ associated with the potential $V^\epsilon(x)$. $t^\epsilon(k)$, through order ϵ^2 is derived by isolating appropriate terms in the expansion (4.48). The sense in which the remainder is small is proved, by entirely different means, in section 5.

$\mathcal{O}(\epsilon^0)$: The only term at order one is e_{V_0+}, which gives the leading order transmission coefficient, $t^{hom}_0(k)$, corresponding to the average potential V_0.

$\mathcal{O}(\epsilon^1)$: At order ϵ, we seek the contribution to $t_\epsilon(k)$ from $\mathcal{U}^\epsilon_1$. From (4.31) we have, since $e_{V_0+}(x;k) \sim t^{hom}_0(k)\, e^{ikx}$ as $x \to +\infty$, that the contribution of $\mathcal{U}^\epsilon_{1,a}$ to the transmission coefficient is given by

$$t^\epsilon_{1,a} = \; \beta^\epsilon_{1,a}\, t^{hom}_0(k) \; = \; \frac{1}{4\pi k} e_{V_0+}(a;k) e_{V_0-}(a;k) \sum_{|j|\geq 1} [q_j]_a \, \frac{e^{2i\pi ja/\epsilon}}{j}.$$

Finally, summing over all contributions from points of discontinuity of q_j, one obtains the complete first order contribution from $\mathcal{U}^\epsilon_1$:

$$t^\epsilon_1 = \sum_j t^\epsilon_{1,a_j} \; = \; \sum_{j=1}^M \frac{e_{V_0+}(a_j;k) e_{V_0-}(a_j;k)}{4\pi k} \sum_{|l|\geq 1} [q_l]_{a_j} \, \frac{e^{2i\pi l \frac{a_j}{\epsilon}}}{l}. \tag{4.49}$$

$\mathcal{O}(\epsilon^2)$: (a) *No contribution to $t^\epsilon(k)$ from $\epsilon^2 U_2^{(p)}$:* We estimate $U_2^{(p)}$ pointwise.

$$|U_2^{(p)}\left(x,\frac{x}{\epsilon}\right)| \leq \frac{1}{4\pi^2}|e_{V_0+}(x;k)| \sum_{|j|\geq 1}\frac{|q_j(x)|}{j^2} \leq C\,\Big(\sum_{|j|\geq 1}|q_j(x)|^2\Big)^{1/2} \;\to\; 0,\ |x|\to\infty.$$

Here we have used the uniform bound (B.3) on e_{V_0+} for $x \geq 0$ and the hypothesis (2.9). Since, $U_2^{(p)}(x,\frac{x}{\epsilon}) \to 0$ as $x \to \infty$, it does not contribute to the transmission coefficient.

(b) *Contribution of $\epsilon^2 U_2^{(h)}(x)$ to $t^\epsilon(k)$:*

From (4.19), one has

$$\left(-\frac{d^2}{dx^2}+V_0(x)-k^2\right)U_2^{(h)}(x) = \frac{e_{V_0+}(x;k)}{4\pi^2}\sum_{|j|\geq 1}\frac{|q_j(x)|^2}{j^2}.$$

Using expression (3.12) for the outgoing resolvent we have

$$\begin{aligned}U_2^{(h)}(x) =& R_{V_0}(k)\left(\frac{e_{V_0+}(\cdot;k)}{4\pi^2}\sum_{|j|\geq 1}\frac{|q_j(\cdot)|^2}{j^2}\right)\\ =&\frac{-1}{2ik\ t_0^{hom}(k)}\int_{-\infty}^{x}\frac{e_{V_0+}(\zeta;k)}{4\pi^2}\sum_{|j|\geq 1}\frac{|q_j(\zeta)|^2}{j^2}e_{V_0-}(\zeta;k)e_{V_0+}(x;k)\ \mathrm{d}\zeta\\ &+\frac{-1}{2ik\ t_0^{hom}(k)}\int_{x}^{+\infty}\frac{e_{V_0+}(\zeta;k)}{4\pi^2}\sum_{|j|\geq 1}\frac{|q_j(\zeta)|^2}{j^2}e_{V_0+}(\zeta;k)e_{V_0-}(x;k)\ \mathrm{d}\zeta.\end{aligned}$$

Therefore, since $q_j \in L^2$ for all $j \in \mathbb{Z}$, and $e_{V_0\pm} \in L^\infty$, one has when $x \to \infty$:

$$\lim_{x\to\infty}U_2^{(h)}(x) - \left(\frac{-e_{V_0+}(x;k)}{8ik\pi^2\ t_0^{hom}(k)}\int_{\mathbb{R}}\sum_{|j|\geq 1}\frac{|q_j(\zeta)|^2}{j^2}e_{V_0-}(\zeta;k)e_{V_0+}(\zeta;k)\mathrm{d}\zeta\right)=0.$$

It follows that the contribution of $U_2^h(x)$ to the transmission coefficient is

$$t_2^{hom}(k) \equiv -\frac{1}{8ik\pi^2}\int_{\mathbb{R}}\sum_{|j|\geq 1}\frac{|q_j(\zeta)|^2}{j^2}e_{V_0-}(\zeta;k)e_{V_0+}(\zeta;k)\mathrm{d}\zeta. \tag{4.50}$$

(c) *Contribution of $\epsilon^2\mathcal{U}_2^\epsilon$ to $t^\epsilon(k)$:*

We study $\mathcal{U}_2^\epsilon$ as above. From (4.31) we have, since $e_{V_0+}(x;k) \sim t_0^{hom}(k)\ e^{ikx}$ as $x \to +\infty$, that the contribution of $\mathcal{U}_{2,a}^\epsilon$ to the transmission coefficient is given by $t_2^\epsilon \equiv t_0^{hom}(k)\ \beta_{2,a}^\epsilon$.

From (4.44), (4.24) and (4.26) we have

$$\begin{aligned}t_{2,a}^\epsilon(k) \;&=\; \beta_{2,a}^\epsilon\ t_0^{hom}(k)\\ &= \frac{-e_{V_0-}(a;k)}{8\pi^2 ik}\ \sum_{|j|\geq 1}\left(c\ e_{V_0+}(a;k)q_j(a-)+[\partial_x(e_{V_0+}(x;k)q_j(x))]_a\ \right)\frac{e^{2i\pi ja/\epsilon}}{j^2}\\ &\quad-\ \frac{1}{8\pi^2 ik}\ \sum_{|j|\geq 1}e_{V_0+}(a;k)\,[q_j]_a\,\partial_x e_{V_0-}(a+;k)\frac{e^{2i\pi ja/\epsilon}}{j^2}.\end{aligned}$$

Using the easily verified identity

$$[\ q_j(x)\partial_x e_{V_0-}(x;k)\ ]_a \;=\partial_x e_{V_0-}(a+;k)\,[\ q_j(x)\ ]_a \;+\; c\ e_{V_0-}(a;k)\ q_j(a-), \tag{4.51}$$

we obtain

$$\begin{aligned} t^\epsilon_{2,a}(k) &= \frac{-1}{8\pi^2 ik}\sum_{|j|\geq 1}\Big(\left[\partial_x(e_{V_0+}(x;k)q_j(x))\right]_a e_{V_0-}(a;k) \\ &\quad + \left[q_j(x)\partial_x e_{V_0-}(x;k)\right]_a e_{V_0+}(a;k)\Big)\frac{e^{2i\pi ja/\epsilon}}{j^2} \\ &= \frac{i}{8\pi^2 k}\sum_{|j|\geq 1}\left[\partial_x(e_{V_0-}(x;k)e_{V_0+}(x;k)q_j(x))\right]_a \frac{e^{2i\pi ja/\epsilon}}{j^2}. \end{aligned}$$

Finally, summing over all contributions of all singular and / or discontinuity points of V^ϵ, we obtain the simple expression:

$$t^\epsilon_2(k) = \sum_{a\in\Omega} t^\epsilon_{2,a}(k) = \frac{i}{8\pi^2 k}\sum_{a\in\Omega}\sum_{|l|\geq 1}\left[\partial_x(e_{V_0-}(x;k)e_{V_0+}(x;k)q_l(x))\right]_a \frac{e^{2i\pi la/\epsilon}}{l^2}. \tag{4.52}$$

$\mathcal{O}(\epsilon^3)$: By similar considerations to the above discussion of $U_2^{(h)}$ and $U_2^{(p)}$, the terms $U_3^\epsilon = U_3^{(h)} + U_3^{(p)}$ in the expansion of $e_{V_+^\epsilon}$ give a correction to $t^\epsilon(k)$ of order ϵ^3, and is therefore subsumed by the error term in the expansion (2.13).

In summary, we have an expansion of $t^\epsilon(k)$, agreeing with the expansion (2.13) in Theorem 2.2:

Proposition 4.3. **Formal corrected homogenization expansion:**

$$t^\epsilon(k) = t_0^{hom}(k) \ + \ \epsilon\, t_1^\epsilon(k) + \epsilon^2 \ \Big(\ t_2^{hom}(k) + t_2^\epsilon(k) \ \Big) \ + \ \mathcal{O}(\epsilon^3), \tag{4.53}$$

where the leading order term, $t_0^{hom}(k)$, is the transmission coefficient associated with the homogenized (average with respect to the fast scale) potential V_0, $t_2^{hom}(k)$ is a classical homogenization theory corrector given by (4.50)*, and t_j^ϵ, $j = 1, 2$ are interface correctors given by* (4.49) *and* (4.52)*.*

Note that if V_0 is generic, then since using that $t_{V_0}(k)$ and $e_{V_0\pm}(x,k)$ are $\mathcal{O}(k)$ as $k \to 0$, we see the expansion is formally valid for any $k \in \mathbb{R}$. However, if V_0 is not generic then we must exclude $k = 0$; see the discussion of Remark 3.7.

5. Rigorous analysis of the scattering problem

In the preceding section, we applied the classical method of multiple scales to derive a formal expansion for the distorted plane wave $e_{V^\epsilon+}(x;k)$ and transmission coefficient $t^\epsilon(k)$; see section 3. For sufficiently smooth potentials, this expansion satisfies, at each order in ϵ, all necessary continuity conditions as well as the radiation condition at infinity; see Definitions 3.1 and 3.4.

We found, however, that if the potential is non-smooth this expansion, while valid *in the bulk*, violates continuity conditions at (i) discontinuities, and (ii) at strong singularities of the background, unperturbed potential, $V_0(x) = V_{reg}(x) + V_{sing}(x)$. We found, in Section 4 that we can, "by hand", construct interface correctors for each point of non-smoothness, thereby giving a corrected expansion (bulk expansion plus interface correctors) which is a valid solution to any finite order in ϵ. The expansion of Proposition 4.3 is explicit through order ϵ^2 with order ϵ^3 correctors.

Question: Does the procedure of section 4 yield a valid expansion with an error terms satisfying an appropriate higher order error bound?

It turns out that the formal expansion is correct with an appropriate error estimate. However, we obtain this result, not by expansion in scalar ϵ but rather in the the function

$q_\epsilon(x)$, with respect to which there is an analytic perturbation theory in an appropriate function space . Smallness required for control of the perturbation expansion derives from $q_\epsilon(x, x/\epsilon)$ being supported at high frequencies if ϵ is small. The principle terms, displayed in the expansion of Proposition 4.3 (and indeed the terms at any finite order in the small parameter, ϵ), are obtained via small ϵ asymptotics of the leading order terms in the q_ϵ expansion. The approach we use was introduced by Golowich and Weinstein in [**144**].

5.1. Formulation of the problem. We consider the general one-dimensional scattering problem

$$\left(-\frac{d^2}{dx^2} + V_0(x) + Q(x) - k^2\right) e_{V+} = 0$$

$$e_{V_0+Q,+}(x;k) \; - \; e^{ikx} \; \to 0, \quad x \to -\infty, \tag{5.1}$$

where $V_0(x)$ as hypothesized in section 1 and Q is a spatially localized perturbing potential, which we think of as being spectrally supported at high frequencies. Q may be large in L^∞. As a model, we have in mind $Q(x) = q_\epsilon(x) = q(x, x/\epsilon)$, with ϵ small.

We introduce the scattered field, u_s, via

$$e_{V+}(x;k) = e_{V_0+}(x;k) + u_s(x;k) \tag{5.2}$$

where u_s is outgoing as $x \to \pm\infty$. Therefore, u_s is the solution of

$$\left(-\frac{d^2}{dx^2} + (V_0 + Q)(x) - k^2\right) u_s(x;k) = -Q e_{V_0+}(x;k), \tag{5.3}$$

with outgoing conditions : $(\partial_x \mp ik) u_s \to 0, \; x \to \pm\infty$.

Applying the outgoing resolvent, $R_{V_0}(k)$, to (5.3) and rearranging terms we obtain the Lippman-Schwinger equation

$$\begin{aligned} u_s \; &= \; -(I + R_{V_0}(k)Q)^{-1} \; R_{V_0}(k) \; Q \; e_{V_0+}(\cdot;k) \quad \Longrightarrow \\ e_{V+}(x;k) = \; &e_{V_0+}(x;k) \; - (I + R_{V_0}(k)Q)^{-1} \; R_{V_0}(k) \; Q \; e_{V_0+}(\cdot;k). \end{aligned} \tag{5.4}$$

Consider now the *formal* Neumann expansion, obtained from (5.4).

$$\begin{aligned} e_{V+}(x;k) \; &= e_{V_0+}(x;k) - R_{V_0}(k)(Q e_{V_0+}(x;k)) + R_{V_0}(k)QR_{V_0}(k)(Q e_{V_0+}(x;k)) + \dots \\ &= e_{V_0+}(x;k) \; + \; \sum_{m=0}^{\infty} \left[(-R_{V_0}(k) \; Q)^m \; e_{V_0+}\right](x;k). \end{aligned} \tag{5.5}$$

In this section we show for a class of Q, which include high-contrast (pointwise large) microstructure (highly oscillatory) potentials that the expansion (5.5) converges in an appropriate sense and that any truncation satisfies an error bound.

5.2. Reformulation of the Lippman-Schwinger equation and the norm $|||Q|||$. We seek a reformulation of the Lippman-Schwinger equation (5.4) in which it is explicitly clear that if Q is highly oscillatory, then the terms of the Neumann series are successively smaller. Introduce, via the Fourier transform, the operator $\langle D_0 \rangle^s$

$$\langle D_0 \rangle^s \, g \equiv (I - \Delta)^{s/2} g \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ix\xi} \; (1 + \xi^2)^{s/2} \; \hat{g}(\xi) \mathrm{d}\xi. \tag{5.6}$$

and the localized function χ

$$\chi(x) = \langle x \rangle^{-\sigma} \; = \; \left(1 + x^2\right)^{-\frac{\sigma}{2}}, \quad \sigma > 4. \tag{5.7}$$

Now introduce the *spatially and frequency weighted distorted plane wave*, $E_{V+}(x;k)$, given by:

$$E_{V+}(x;k) \; \equiv \; (\; \langle D_0 \rangle \, \chi e_{V+} \;) \, (x;k). \tag{5.8}$$

With the operator definitions

$$T_{R_{V_0}}(k) \equiv \langle D_0\rangle\, \chi R_{V_0}(k)\chi\, \langle D_0\rangle\,, \tag{5.9}$$

$$T_Q \equiv \langle D_0\rangle^{-1}\, \chi^{-1} Q \chi^{-1}\, \langle D_0\rangle^{-1} \;=\; \langle D_0\rangle^{-1}\, \langle x\rangle^{\sigma}\cdot Q\cdot \langle x\rangle^{\sigma}\, \langle D_0\rangle^{-1}\;, \tag{5.10}$$

$E_{V+}(x;k)$ can be seen to satisfy

$$\Big(I + T_{R_{V_0}}\; T_Q\Big)\;\big(\; E_{V+}(\cdot;k) - E_{V_0+}(\cdot;k)\;\big) = -\,\langle D_0\rangle\, \chi R_{V_0}(k)\; Q\; e_{V_0+}(x;k). \tag{5.11}$$

Here's the motivation for our strategy. Note that T_Q has the operator $\langle D_0\rangle^{-1}$ as both a pre- and post- multiplier. This has the effect of a high frequency cutoff. Therefore, for highly oscillatory Q, T_Q is expected to be of small operator norm. If the norm of $T_{R_{V_0}}\circ T_Q$ is small then $I + T_{R_{V_0}}\circ T_Q$ is invertible and we have the *preconditioned* Lippman-Schwinger equation

$$E_{V+}(\cdot;k) \;=\; E_{V_0+}(\cdot;k) \;-\; \Big(I + T_{R_{V_0}}\; T_Q\Big)^{-1}\; \langle D_0\rangle\, \chi R_{V_0}(k)\; Q e_{V_0+}(x;k). \tag{5.12}$$

We proceed now to construct a norm, $|||Q|||$, such that if $|||Q|||$ is small then $T_{R_{V_0}}\; T_Q$ is bounded and of small norm as an operator norm from L^2 to L^2.

The norm we choose for the perturbing potential is defined as follows:

$$|||Q||| \;\equiv\; \|T_Q\|_{L^2\to L^2} \;=\; \Big\|\langle D_0\rangle^{-1}\; \langle x\rangle^{\sigma} Q\; \langle x\rangle^{\sigma}\, \langle D_0\rangle^{-1}\Big\|_{L^2\to L^2}\,, \quad \sigma > 4. \tag{5.13}$$

The next result establishes the expansion of the distorted plane waves $e_{V+}(x;k)$ in a $H^1(\mathbb{R};\chi(x)\mathrm{d}x)$ and therefore, by the Sobolev inequality, a $L^\infty(\mathbb{R};\chi(x)\mathrm{d}x)$ convergent expansion for $|||Q|||$ sufficiently small.

Theorem 5.1. *Let V satisfy Hypotheses **(V)**, and $k \in K$ a compact subset of $\mathbb{R}$, satisfying Hypothesis **(G)**. Define*

$$\tau_0(K) \;\equiv\; \frac{1}{\max_{k\in K}\Big\|T_{R_{V_0}}(k)\Big\|_{L^2\to L^2}} \;>\; 0.$$

If $|||Q||| < \tau_0(K)$, then for all $k \in K$:

- *The preconditioned Lippman Schwinger equation (5.12) has a unique spatially and spectrally weighted distorted plane solution, $E_{V+}(x;k)$.*
- *This solution can be expressed as a series, which converges in $L^2(\mathbb{R})$, uniformly in $k \in K$:*

$$\begin{aligned} E_{V+}(x;k) \;&=\; E_{V_0+}(x;k) \;+\; \sum_{m=0}^{\infty}\; \Big(-T_{R_{V_0}}(k)\; T_Q\Big)^m\, [\langle D_0\rangle\, \chi R_{V_0}(k)\; Q\; e_{V_0+}]\,(x;k) \\ &=\; E_{V_0+}(x;k) \;-\; \langle D_0\rangle\, \chi\; R_{V_0}(k)\; Q\; e_{V_0+}(x;k) \\ &\qquad +\; T_{R_{V_0}}(k)\; T_Q\;\; \langle D_0\rangle\, \chi\; R_{V_0}(k)\; Q\; e_{V_0+}(x;k) - \dots \end{aligned}$$

- *It follows that the distorted plane wave, $e_{V_0+Q,+}(x;k)$ satisfies the approximation for any $M \geq 1$*

$$\left\|\langle D_0\rangle\, \chi\; \left(u_s(\cdot;k) \;+\; \sum_{m=0}^{M}\, [(-R_{V_0}(k)Q)^m\, e_{V_0+}]\,(\cdot;k)\;\right)\right\|_{L^2(\mathbb{R})} \;\leq\; C|||Q|||^{M+1} \tag{5.14}$$

with $u_s(x;k) \;\equiv\; e_{V_0+Q,+}(x;k) \;-\; e_{V_0+}(x;k)$.

Remark 5.2. *In the proof of Theorem 5.1, the distinction between generic and non-generic cases arises through the properties of the unperturbed resolvent, $R_{V_0}(k)$, as $k \to 0$; see Proposition 3.8.*

In the following, we prove that both $T_{R_{V_0}}$ and T_Q are well-defined operators, bounded in L^2. Then, Theorem 5.1 follows immediately if Q satisfies the smallness condition

$$\left\|T_Q\right\|_{L^2\to L^2} \;<\; \min_{k\in K}\left(\left\|T_{R_{V_0}}(k)\right\|_{L^2\to L^2}\right)^{-1} \;\equiv\; \tau_0(k). \tag{5.15}$$

PROPOSITION 5.3. *Let $\langle x\rangle^{2\sigma}Q(x)\in L^2(\mathbb{R})$. Then T_Q, as defined in* (5.10)*, is a Hilbert-Schmidt operator and is therefore compact.*

PROPOSITION 5.4. *Let $q_\epsilon(x)$ satisfy the conditions in Hypotheses* **(V)**. *Then, for ϵ small,*

$$\left\|T_{q_\epsilon}\right\|_{L^2\to L^2} = \mathcal{O}(\epsilon). \tag{5.16}$$

PROPOSITION 5.5. *$T_{R_{V_0}}(k)$ is a bounded operator from L^2 to L^2.*

Propositions 5.3 and 5.4 are proved below. The proof of Proposition 5.5 is somewhat more technical proof and is found in Appendix C.

We now prove Proposition 5.3. The proof of Proposition 5.4 begins on page 210.

PROOF. **Proof of Proposition 5.3:** We begin by introducing the notation

$$Q^\sharp = \chi^{-1}Q\chi^{-1}. \tag{5.17}$$

Then, one uses the following calculation:

$$\begin{aligned}
\langle D_0\rangle^{-1}Q^\sharp\langle D_0\rangle^{-1}f(x) &= \langle D_0\rangle^{-1}Q^\sharp\langle D_0\rangle^{-1}\left(\frac{1}{2\pi}\int_\xi e^{ix\xi}\hat f(\xi)\,\mathrm{d}\xi\right)\\
&= \int_\xi \frac{\hat f(\xi)}{2\pi}\langle D_0\rangle^{-1}Q^\sharp\langle D_0\rangle^{-1}e^{ix\xi}\mathrm{d}\xi\\
&= \int_\xi \frac{\hat f(\xi)}{4\pi^2}\langle D_0\rangle^{-1}Q^\sharp e^{ix\xi}(1+\xi^2)^{-1/2}\mathrm{d}\xi\\
&= \int_\xi \frac{\hat f(\xi)}{4\pi^2}(1+\xi^2)^{-1/2}\frac{1}{2\pi}\int_\eta e^{ix\eta}(1+\eta^2)^{-1/2}\widehat{Q^\sharp e^{i\eta\xi}}\,\mathrm{d}\eta\,\mathrm{d}\xi\\
&= \int_\xi \frac{1}{8\pi^3}\left(\int_\zeta e^{-iy\xi}f(\zeta)\,\mathrm{d}\zeta\right)\int_\eta(\langle\xi\rangle\langle\eta\rangle)^{-1}e^{ix\eta}\widehat{Q^\sharp}(\xi-\eta)\,\mathrm{d}\eta\,\mathrm{d}\xi\\
&= \int_\zeta f(\zeta)K(x,\zeta)\,\mathrm{d}\zeta.
\end{aligned}$$

with the kernel

$$K(x,\zeta)\equiv\frac{1}{8\pi^3}\int_\xi\int_\eta(1+\xi^2)^{-1/2}(1+\eta^2)^{-1/2}e^{i(x\eta-\zeta\xi)}\widehat{Q^\sharp}(\xi-\eta)\,\mathrm{d}\eta\,\mathrm{d}\xi \tag{5.18}$$

We want to prove that $\iint|K(x,\zeta)|^2\,\mathrm{d}x\,\mathrm{d}\zeta<+\infty$, *i.e.* $K\in L^2(\mathbb{R}^2)$. One has

$$\begin{aligned}
\widehat K(s,z) =& \iint_{\mathbb{R}^2}K(x,\zeta)e^{-ixs}e^{-i\zeta z}\,\mathrm{d}x\,\mathrm{d}\zeta\\
=&\frac{1}{8\pi^3}\iint_{x,\zeta}\iint_{\eta,\xi}\langle\xi\rangle^{-1}\,\langle\eta\rangle^{-1}e^{ix(\eta-s)}e^{-i\zeta(\xi+z)}\widehat{Q^\sharp}(\xi-\eta)\,\mathrm{d}x\,\mathrm{d}\zeta\,\mathrm{d}\eta\,\mathrm{d}\xi\\
=&\frac{1}{8\pi^3}\frac{\widehat{Q^\sharp}(-s-z)}{(1+s^2)^{1/2}(1+z^2)^{1/2}}.
\end{aligned}$$

Therefore, we deduce

$$\iint_{x,\zeta} |K(x,\zeta)|^2 \, \mathrm{d}x \, \mathrm{d}\zeta = \iint_{s,z} |\widehat{K}(s,z)|^2 \, \mathrm{d}s \, \mathrm{d}z = \frac{1}{8\pi^3} \int_s \frac{1}{1+s^2} \int_z \frac{|\widehat{Q^\sharp}(s+z)|^2}{1+z^2} \, \mathrm{d}z \, \mathrm{d}s.$$

Since $Q^\sharp \in L^2(\mathbb{R})$, one has immediately $\iint_{x,\zeta} |K(x,\zeta)|^2 \, \mathrm{d}x \, \mathrm{d}\zeta < \infty$, and

$$\big\|K(x,\zeta)\big\|_{L^2(\mathbb{R}^2)} \leq C \big\|Q^\sharp\big\|_{L^2(\mathbb{R})}.$$

Therefore T_Q is a Hilbert-Schmidt integral operator, and is therefore bounded, with

$$\big\|T_Q\big\|_{L^2 \to L^2} \leq C \big\|Q^\sharp\big\|_{L^2(\mathbb{R})}.$$

This completes the proof of Proposition 5.3. □

Proof. **Proof of Proposition 5.4:**

Consider T_Q, where $Q = q_\epsilon(x) = q\left(x, \frac{x}{\epsilon}\right)$ as in (2.2). From the proof below, one has $T_{q_\epsilon} f(x) = \int_\zeta f(\zeta) K_\epsilon(x,\zeta) \, \mathrm{d}\zeta$, with the kernel $K_\epsilon(x,\zeta)$ satisfying

$$\widehat{K_\epsilon}(s,z) = \frac{1}{8\pi^3} \frac{\widehat{q_\epsilon^\sharp}(-s-z)}{(1+s^2)^{1/2}(1+z^2)^{1/2}}.$$

Using the decomposition in Fourier series of $q_\epsilon(x) = q\left(x, \frac{x}{\epsilon}\right)$, one has

$$q_\epsilon^\sharp(x) \equiv \chi^{-1} q_\epsilon \chi^{-1}(x) = \sum_{|j|\geq 1} q_j^\sharp(x) e^{2i\pi j(x/\epsilon)},$$

and therefore

$$\widehat{q_\epsilon^\sharp}(\xi) = \sum_{|j|\geq 1} \int_x q_j^\sharp(x) e^{2i\pi j(x/\epsilon)} e^{-ix\xi} \mathrm{d}x = \sum_{|j|\geq 1} \widehat{q_j^\sharp}\left(\frac{2\pi j}{\epsilon} - \xi\right).$$

One deduces then

$$\begin{aligned}
\iint_{s,z} |\widehat{K_\epsilon}(s,z)|^2 \, \mathrm{d}s \, \mathrm{d}z &= \frac{1}{8\pi^3} \sum_{|j|\geq 1} \iint_{s,z} \frac{|\widehat{q_j^\sharp}(s+z+(2\pi j/\epsilon))|^2}{(1+s^2)(1+z^2)} \, \mathrm{d}s \, \mathrm{d}z \\
&= \frac{1}{8\pi^3} \sum_{|j|\geq 1} \int_{\mathbb{R}} \mathrm{d}z \int_{\mathbb{R}} \mathrm{d}\eta \frac{|\widehat{q_j^\sharp}(\eta + 2\pi j/\epsilon))|^2}{(1+(\eta-z)^2)(1+z^2)} \\
&= \sum_{|j|\geq 1} \int_{\mathbb{R}} \mathrm{d}z \int_{|\eta|\geq \frac{\pi j}{\epsilon}} \mathrm{d}\eta \frac{|\widehat{q_j^\sharp}(\eta + (2\pi j/\epsilon))|^2}{(1+(\eta-z)^2)(1+z^2)} \\
&\quad + \sum_{|j|\geq 1} \int_{\mathbb{R}} \mathrm{d}z \int_{|\eta|\leq \frac{\pi j}{\epsilon}} \mathrm{d}\eta \frac{|\widehat{q_j^\sharp}(\eta + (2\pi j/\epsilon))|^2}{(1+(\eta-z)^2)(1+z^2)} \equiv \; I_1 \; + \; I_2.
\end{aligned}$$

Estimation of I_1:

$$\begin{aligned}
\int_{\mathbb{R}} \mathrm{d}z \int_{|\eta|\geq \frac{\pi j}{\epsilon}} \frac{|\widehat{q_j^\sharp}(\eta + \frac{2\pi j}{\epsilon})|^2}{(1+(\eta-z)^2)(1+z^2)} \, \mathrm{d}\eta \; &= \; \left(\int_{|z|\geq \frac{\pi j}{2\epsilon}} \mathrm{d}z + \int_{|z|\leq \frac{\pi j}{2\epsilon}} \mathrm{d}z \right) \int_{|\eta|\geq \frac{\pi j}{\epsilon}} \mathrm{d}\eta \\
&= \; I_{1,A} \; + \; I_{1,B}, \quad \text{with} \\
I_{1,A} = \int_{|z|\geq \frac{\pi j}{2\epsilon}} \mathrm{d}z \int_{|\eta|\geq \frac{\pi j}{\epsilon}} \mathrm{d}\eta &\leq C \frac{\epsilon^2}{j^2} \int_{|\eta|\geq \frac{\pi j}{\epsilon}} \left| \widehat{q_j^\sharp}\left(\eta + \frac{2\pi j}{\epsilon}\right) \right|^2 \int_{|z|\geq \frac{\pi j}{2\epsilon}} \frac{1}{1+(\eta-z)^2} \mathrm{d}z \mathrm{d}\eta
\end{aligned}$$

$$\leq C' \frac{\epsilon^2}{j^2}\|q_j^\sharp\|_{L^2}^2; \quad \text{and as } |\eta - z| \geq \frac{\pi j}{2\epsilon} \text{ for } |z| \leq \frac{\pi j}{2\epsilon},\ |\eta| \geq \frac{\pi j}{\epsilon},$$

$$I_{1,B} = \int_{|z|\leq \frac{\pi j}{2\epsilon}} \mathrm{d}z \int_{|\eta|\geq \frac{\pi j}{\epsilon}} \mathrm{d}\eta \leq C\frac{\epsilon^2}{j^2} \int_{|z|\leq \frac{\pi j}{2\epsilon}} \frac{1}{1+z^2}\,\mathrm{d}z \int_{|\eta|\geq \frac{\pi j}{\epsilon}} \left|\widehat{q_j^\sharp}\left(\eta + \frac{2\pi j}{\epsilon}\right)\right|^2 \mathrm{d}\eta$$
$$\leq C' \frac{\epsilon^2}{j^2}\|q_j^\sharp\|_{L^2}^2.$$

Now, summing on j, one obtains $I_1(\epsilon) = \mathcal{O}(\epsilon^2)$.

Estimation of I_2: We first show that if we assume only that $\sum_{|j|\geq 1}\|q_j^\sharp\|_{L^2}^2 < \infty$, then $I_2(\epsilon) \to 0$ as $\epsilon \to 0$, and therefore $\|T_{q_\epsilon}\|_{L^2\to L^2}^2 = o(1) + \mathcal{O}(\epsilon^2) = o(1)$ as $\epsilon \to$ with no specified rate.

We then show that if q_ϵ is as in hypotheses **(V)** then $\|T_{q_\epsilon}\|_{L^2\to L^2} = \mathcal{O}(\epsilon)$ as $\epsilon \to 0$.

Assume $\sum_{|j|\geq 1}\|q_j^\sharp\|_{L^2}^2 < \infty$. Then,

$$\begin{aligned} I_2 &\equiv \sum_{|j|\geq 1}\int_{\mathbb{R}} \mathrm{d}z \int_{|\eta|\leq \frac{\pi j}{\epsilon}} \mathrm{d}\eta \frac{|\widehat{q_j^\sharp}(\eta + (2\pi j/\epsilon))|^2}{(1+(\eta-z)^2)(1+z^2)} \\ &\leq \int_{\mathbb{R}} \frac{1}{(1+z^2)}\,\mathrm{d}z \sum_{|j|\geq 1}\int_{|\eta|\leq \frac{\pi j}{\epsilon}} |\widehat{q_j^\sharp}(\eta + (2\pi j/\epsilon))|^2\,\mathrm{d}\eta \\ &\leq C\sum_{|j|\geq 1}\int_{\frac{\pi j}{\epsilon}\leq \tau \leq \frac{3\pi j}{\epsilon}} |\widehat{q_j^\sharp}(\tau)|^2\,d\tau. \end{aligned}$$

Note that $\sum_{|j|\geq 1}\int_{\mathbb{R}}|\widehat{q_j^\sharp}(\tau)|^2\,d\tau = \sum_{|j|\geq 1}\|q_j^\sharp\|_{L^2}^2 < \infty$, implying $I_2 = o(1)$ as $\epsilon \to 0$.

We now turn to the case where q_ϵ satisfies the condition in Hypotheses **(V)** in order to establish that $\|T_{q_\epsilon}\|_{L^2\to L^2} = \mathcal{O}(\epsilon)$ as $\epsilon \to 0$. The estimate for $I_1(\epsilon)$ is as above: $I_1(\epsilon) = \mathcal{O}(\epsilon^2)$.

Now, we estimate $I_2(\epsilon)$ using the fact that since $q_j^\sharp \in L^2$ and $(q_j^\sharp)' \in L^2$:

$$\begin{aligned} \left|\widehat{q_j^\sharp}(\tau)\right| &= \left|\sum_{l=0}^{M}\int_{a_l}^{a_{l+1}} q_j^\sharp(x)e^{-i\tau x}\,\mathrm{d}x\right| \\ &= \left|\frac{1}{i\tau}\sum_{l=0}^{M}\left(\int_{a_l}^{a_{l+1}} (q_j^\sharp)'(x)e^{-i\tau x}\,\mathrm{d}x - q_j^\sharp(a_{l+1}^-)e^{-i\tau a_{l+1}} + q_j^\sharp(a_l+)e^{-i\tau a_l}\right)\right| \\ &\leq C\frac{1}{\tau}\left(\left\|(q_j^\sharp)'\right\|_{L^2} + \sum_{l=1}^{M}\left[q_j^\sharp(x)\right]_{a_j}\right) = \mathcal{O}\left(\frac{1}{\tau}\right). \end{aligned}$$

Therefore, one has

$$\begin{aligned} I_2 &\equiv \sum_{|j|\geq 1}\int_{\mathbb{R}} \mathrm{d}z \int_{|\eta|\leq \frac{\pi j}{\epsilon}} \mathrm{d}\eta \frac{|\widehat{q_j^\sharp}(\eta + (2\pi j/\epsilon))|^2}{(1+(\eta-z)^2)(1+z^2)} \\ &\leq C\int_{\mathbb{R}} \frac{1}{(1+z^2)}\,\mathrm{d}z \sum_{|j|\geq 1}\int_{|\eta|\leq \frac{\pi j}{\epsilon}} \left(\frac{\epsilon}{\pi j}\right)^2 \frac{1}{1+(\eta-z)^2}\mathrm{d}\eta \leq C'\epsilon^2\sum_{|j|\geq 1}\frac{1}{j^2}. \end{aligned}$$

One deduces finally that $\left\|T_{q_\epsilon}\right\|_{L^2\to L^2} = \left\|K_\epsilon\right\|_{L^2(\mathbb{R}^2)} = I_1 + I_2 = \mathcal{O}(\epsilon)$. This completes the proof of Proposition 5.4. □

5.3. Application to the transmission coefficient, $t(k) = t[k;Q]$. This section is devoted to the proof of Theorem 2.3. The heart of the matter is to view $t(k)$ as a functional of the perturbing microstructure potential, $Q(x)$

$$Q \;\mapsto\; t[Q] \tag{5.19}$$

and to use the Lippman-Schwinger expansion of Theorem 5.1 to expand $t[Q]$ for small $|||Q|||$:

$$t[Q] \;=\; t_0^{hom} \;+\; t_1[Q] \;+\; t_2[Q,Q] \;+\; t_3[Q,Q,Q] \;+\; \ldots, \tag{5.20}$$

where $t_j[Q,Q,\ldots,Q]$ is $j-$ linear in Q. The transmission coefficient expansion of Theorem 2.3 is recovered from the small $|||Q|||$ asymptotics of the first several terms of the expansion of $t[q_\epsilon]$. Finally, the error terms are estimated.

Recall that from (3.7) the transmission coefficient, $t_W(k)$, associated with the distorted plane wave $e_{W+}(x;k)$, is given by

$$t_W(k) \;=\; \lim_{x\to+\infty}\; e^{-ikx}\; e_{W+}(x;k).$$

We denote the transmission coefficients of $e_{V_0+}(x;k)$ and $e_{V_0+Q,+}(x;k)$, respectively,

$$\begin{aligned} t_{V_0}(k) \;&\equiv\; t_0(k) \;\equiv\; t_0^{hom}(k), \\ t_V(k) \;\equiv\; t(k) \;&=\; \lim_{x\to+\infty}\; e^{-ikx}\; e_{V_0+Q,+}(x;k) \;=\; t_0^{hom}(k) \;+\; \lim_{x\to+\infty} e^{-ikx}\; u_s(x;k). \end{aligned}$$

To obtain the desired leading order expansion of $t(k)$ of Theorem 2.3 we now derive the small $|||Q|||$ asymptotics of the *linear and quadratic terms in* Q of (5.14).

Calculation of $t_1[Q]$:

One has from (3.11) that

$$\begin{aligned} -R_{V_0}(k)\; Q\; e_{V_0+}(x;k) = \;& \int_{-\infty}^{x} Q(\zeta)\; e_{V_0+}(\zeta;k) e_{V_0-}(\zeta;k)\; \mathrm{d}\zeta\; \frac{e_{V_0+}(x;k)}{2ik\; t_0^{hom}} \\ & + \int_{x}^{+\infty} Q(\zeta)\; e_{V_0+}(\zeta;k) e_{V_0+}(\zeta;k)\; \mathrm{d}\zeta\; \frac{e_{V_0-}(x;k)}{2ik\; t_0^{hom}} \\ \sim\;& t_1[Q]\; e^{ikx}, \quad \text{as } x\to\infty, \end{aligned}$$

where

$$t_1[Q] \;\equiv\; \frac{1}{2ik}\int_{-\infty}^{\infty} Q(\zeta)\; e_{V_0+}(\zeta;k) e_{V_0-}(\zeta;k)\; \mathrm{d}\zeta. \tag{5.21}$$

Calculation of $t_2[Q,Q]$:

One has from (3.11) that

$$\begin{aligned} & R_{V_0}(k)\; Q\; R_{V_0}(k)\; Q(\zeta)\; e_{V_0+}(x;k) \\ & = \int_{-\infty}^{x} Q(\zeta) R_{V_0}(k)(Q(\zeta)\; e_{V_0+}(\zeta;k)) e_{V_0-}(\zeta;k)\; \mathrm{d}\zeta\; \frac{e_{V_0+}(x;k)}{2ik\; t_0^{hom}} \\ & \quad + \int_{x}^{+\infty} Q(\zeta) R_{V_0}(k)(Q(\zeta)\; e_{V_0+}(\zeta;k)) e_{V_0+}(\zeta;k)\; \mathrm{d}\zeta\; \frac{e_{V_0-}(x;k)}{2ik\; t_0^{hom}} \\ & \sim\; t_2[Q,Q]\; e^{ikx}, \end{aligned}$$

where

$$\begin{aligned} t_2[Q,Q] &\equiv \frac{1}{2ik}\int_{-\infty}^{\infty} Q(\zeta) R_{V_0}(k)(Q(\zeta)\; e_{V_0+}(\zeta;k)) e_{V_0-}(\zeta;k)\; \mathrm{d}\zeta \\ &= \frac{1}{2ik}\frac{1}{-2ik\; t_0^{hom}}\int_{-\infty}^{+\infty} Q(\zeta) e_{V_0-}(\zeta;k)\,(\; I_l(\zeta) + I_r(\zeta)\;)\; \mathrm{d}\zeta, \quad \text{with} \end{aligned} \tag{5.22}$$

$$I_l(\zeta) \;=\; \int_{-\infty}^{\zeta} Q(z)\; e_{V_0+}(z;k)\; e_{V_0-}(z;k)\; e_{V_0+}(\zeta;k)\; \mathrm{d}z,$$
$$I_r(\zeta) \;=\; \int_{\zeta}^{+\infty} Q(z)\; e_{V_0+}(z;k)\; e_{V_0+}(z;k)\; e_{V_0-}(\zeta;k)\; \mathrm{d}z.$$

Estimation of the error terms:

The final step for the proof of Theorem 2.3 consists in a bound on the contribution to the transmission coefficient from the remainder term in expansion (5.20). This is given by the following Theorem:

THEOREM 5.6. *Let K denote a compact subset of $\mathbb{R}$, satisfying Hypothesis **(G)**. Introduce for $k \in K$*

$$t_{rem}(k;Q) \;\equiv\; t(k;Q) \;-\; t_0^{hom}(k) \;-\; t_1[Q] \;-\; t_2[Q,Q]. \tag{5.23}$$

Then we have, uniformly in $k \in K$:

i. *If V has compact support, then $t_{rem}(k) \;=\; \mathcal{O}(|||Q|||^3)$.*
ii. *If V is exponentially decreasing, then $t_{rem}(k) \;=\; \mathcal{O}(|||Q|||^{3-})$.*
iii. *If $\langle x\rangle^{\rho+1} V_0(x) \in L^1(\mathbb{R})$ and $\langle x\rangle^{\rho} Q(x) \in L^2(\mathbb{R})$, $\rho > 8$, then there exists $2 < \beta < 3$ such that $t_{rem}(k) \;=\; \mathcal{O}(|||Q|||^{\beta})$.*

PROOF. It is convenient to first introduce

$$\begin{aligned} f_{rem} \;&\equiv\; -\Big(I + T_{R_{V_0}} T_Q\Big)^{-1} (T_{R_{V_0}} T_Q)^3\, \langle D_0\rangle\, \chi e_{V_0+}(x;k) \qquad (5.24)\\ &\equiv\; \langle D_0\rangle\, \chi u_s \;+\; \langle D_0\rangle\, \chi R_{V_0}(k)\; Q(x)\; e_{V_0+}(x;k)\\ &\quad -\; \langle D_0\rangle\, \chi R_{V_0}(k)\; Q\; R_{V_0}(k)\; Q(x)\; e_{V_0+}(x;k). \qquad (5.25)\end{aligned}$$

Using (B.4), one deduces that $\langle D_0\rangle\, \chi e_{V_0+}(x;k) \in L^2_x$, with

$$\begin{aligned} \|\langle D_0\rangle\, \chi e_{V_0+}\|_{L^2} &= \big\|\langle\eta\rangle \widehat{\chi e_{V_0+}}(\eta;k)\big\|_{L^2_\eta}\\ &\le \|\chi(x) e_{V_0+}(x;k)\|_{L^2_x} \;+\; \|\partial_x\left(\chi(x) e_{V_0+}(x;k)\right)\|_{L^2_x} \;\le\; \|\chi(x)\langle x\rangle\|_{L^2_x}\,. \end{aligned}$$

Therefore, thanks to Propositions 5.3 and 5.5, and using (5.24), one has for $|||Q|||$ small enough,

$$f_{rem} \in L^2 \quad \text{and} \quad \|f_{rem}\|_{L^2} \;\le\; C_\chi\; |||Q|||^3. \tag{5.26}$$

The following pointwise bound can be also be deduced

$$\Big|\langle D_0\rangle^{-1} f_{rem}\Big| \;\le \left|\int_\eta \langle\eta\rangle^{-1} \widehat{f_{rem}}(\eta) e^{i\eta x}\right| \;\le \|\langle\eta\rangle^{-1}\|_{L^2_\eta}\; \|f_{rem}\|_{L^2}\,,$$

which implies

$$\Big|\chi^{-1}(x)\, \langle D_0\rangle^{-1} f_{rem}(x)\Big| \le C_\chi\; \chi^{-1}(x)\; |||Q|||^3. \tag{5.27}$$

From (5.25) we have that $t_{rem}(k)$ is the complex number for which

$$\lim_{x\to\infty} \Big(\; \chi^{-1}(x)\, \langle D_0\rangle^{-1} f_{rem} \;-\; t_{rem}(k) e^{ikx} \;\Big) \;=\; 0.$$

We now use the decay properties of the potential V to estimate the magnitude of $t_{rem}(k)$ for $|||Q|||$ small.

Case 1: V has compact support

Assume $\operatorname{supp} V \subset [-M, M],\ M > 0$. Using the explicit representation of R_{V_0}, (3.12), for $x > M$ we have:

$$\begin{aligned} -R_{V_0}(k) Q\; e_{V_0+}(x;k) \;&=\; \frac{1}{2ik\; t_0^{hom}} \int_{-\infty}^{x} Q(\zeta)\; e_{V_0+}(\zeta;k)\; e_{V_0-}(\zeta;k) e_{V_0+}(x;k)\; \mathrm{d}\zeta\\ &\quad +\; \frac{1}{2ik\; t_0^{hom}} \int_{x}^{+\infty} Q(\zeta)\; e_{V_0+}(\zeta;k)\; e_{V_0+}(\zeta;k) e_{V_0-}(x;k)\; \mathrm{d}\zeta \end{aligned}$$

$$= \frac{1}{2ik\ t_0^{hom}} \int_{-\infty}^{+\infty} Q(\zeta)\ e_{V_0+}(\zeta;k)\ e_{V_0-}(\zeta;k) e_{V_0+}(x;k)\ \mathrm{d}\zeta$$
$$= \frac{t_1[Q]}{t_0^{hom}} e_{V_0+}(x;k) \ = \ t_1[Q]\ e^{ikx}.$$

Similarly, for the quadratic in Q-term we have

$$R_{V_0}(k)\ Q\ R_{V_0}(k)\ Q\ e_{V_0+}(x;k) \ = \ t_2[Q,Q]\ e^{ikx}.$$

Therefore,

$$u_s \ = \ t_1[Q]\ e^{ikx} \ + \ t_2[Q,Q]\ e^{ikx} \ + \ t_{rem}\ e^{ikx},$$

where for $x > M$

$$\chi^{-1}\ \langle D_0\rangle^{-1}\ f_{rem}(x) \ = \ t_{rem}(k) e^{ikx}.$$

Therefore, using the pointwise bound (5.27) we have

$$|t_{rem}(k)| \ \leq \ C_\chi \chi^{-1}(M)\ |||Q|||^3 \ = \ \mathcal{O}(|||Q|||^3).$$

Case 2: V is exponentially decreasing
Assume $|V_0(x)| + |Q(x)| \leq Ce^{-\alpha|x|}$ for some $C,\ \alpha > 0$ and $x > M$. As in the first case, the formula for the resolvent (3.12) leads to

$$- R_{V_0}(k)Q\ e_{V_0+}(x;k) \ = \ t_1[Q]\ e^{ikx} \ + \ \frac{t_1[Q]}{t_0^{hom}} \Big(e_{V_0+}(x;k) - t_0^{hom} e^{ikx} \Big)$$
$$+ \frac{1}{2ik\ t_0^{hom}} \int_x^{+\infty} Q(\zeta) e_{V_0+}(\zeta;k) \Big(e_{V_0+}(\zeta;k) e_{V_0-}(x;k) - e_{V_0+}(x;k) e_{V_0-}(\zeta;k) \Big)\ \mathrm{d}\zeta.$$

Using (B.4), one can easily bound for $x > M$

$$\left| \frac{1}{2ik\ t_0^{hom}} \int_x^{+\infty} Q(\zeta)\ e_{V_0+}(\zeta;k)\ (e_{V_0+}(\zeta;k) e_{V_0-}(x;k) - e_{V_0+}(x;k) e_{V_0-}(\zeta;k))\ \mathrm{d}\zeta \right|$$
$$\leq C \int_x^{+\infty} Ce^{-\alpha|\zeta|} \langle\zeta\rangle\ \mathrm{d}\zeta \ \leq \ C' e^{-\alpha/2|x|}.$$

Now, we use the estimate (B.3)

$$|m_+(x;k) - 1| \leq \frac{1 + \max(-x,0)}{1+|k|} \int_x^\infty (1+|s|)|V_0(s)|\mathrm{d}s,$$

so that $|e_{V_0+}(x;k) - t_0^{hom} e^{ikx}| \leq Ce^{-\alpha/2x}$ for $x > M$. Finally, one obtains

$$\left| R_{V_0}(k) Q\ e_{V_0+}(x;k) \ - t_1[Q]\ e^{ikx} \right| \ \leq \ Ce^{-\alpha/2x}.$$

A similar estimate holds for $u_s \ = \ e_{V+}(x;k) - e_{V_0+}(x;k)$, and for the quadratic term $R_{V_0}(k)\ Q\ R_{V_0}(k)\ Q\ e_{V_0+}(x;k)$. Therefore, for $x > M$ we have

$$\left| \chi^{-1}\ \langle D_0\rangle^{-1}\ f_{rem}(x) \ - \ t_{rem}(k) e^{ikx} \right| \leq \ Ce^{-\alpha/2x}.$$

Again the pointwise bound (5.27) implies, for $x > M$, that

$$|t_{rem}(k)| \ \leq \ C_\chi \chi^{-1}(x) |||Q|||^3 \ + \ Ce^{-\alpha/2x}.$$

Finally, choosing $x = -\frac{6}{\alpha} \ln |||Q|||$, one has for $|||Q|||$ small enough,

$$|t_{rem}(k)| \leq \ C|||Q|||^3 (1 + \langle \ln |||Q||| \rangle) \leq C|||Q|||^{3-}.$$

Case 3: $\langle x\rangle^{\rho+1} V_0(x) \in L^1(\mathbb{R})$ and $\langle x\rangle^\rho Q(x) \in L^2(\mathbb{R})$, with $\rho > 8$
We use again the formula of the resolvent (3.12):

$$- R_{V_0}(k)Q\ e_{V_0+}(x;k) \ = \ t_1[Q]\ e^{ikx} \ + \ \frac{t_1[Q]}{t_0^{hom}} \Big(e_{V_0+}(x;k) - t_0^{hom} e^{ikx} \Big)$$

$$+\frac{1}{2ik\ t_0^{hom}}\int_x^{+\infty} Q(\zeta)e_{V_0+}(\zeta;k)\Big(e_{V_0+}(\zeta;k)e_{V_0-}(x;k)-e_{V_0+}(x;k)e_{V_0-}(\zeta;k)\Big)\ \mathrm{d}\zeta.$$

Using the estimate (B.3) leads to

$$|e_{V_0+}(x;k)-t_0^{hom}(k)e^{ikx}|\le C\int_x^{\infty}\frac{1}{(1+|s|)^{\rho}}(1+|s|)^{\rho+1}|V_0(s)|\ \mathrm{d}s\le C\frac{1}{\langle x\rangle^{\rho}}\|V_0\|_{L^{1,\rho+1}}.$$

Therefore, one has

$$\left|\frac{1}{2ik\ t_0^{hom}}\int_x^{+\infty} Q(\zeta)\ e_{V_0+}(\zeta;k)\ (e_{V_0+}(\zeta;k)e_{V_0-}(x;k)-e_{V_0+}(x;k)e_{V_0-}(\zeta;k))\ \mathrm{d}\zeta\right|$$
$$\le C\int_x^{+\infty}|Q(\zeta)|\frac{1}{\langle\zeta\rangle^{\rho}}\left(\frac{\langle x\rangle}{\langle\zeta\rangle^{\rho}}+\frac{\langle\zeta\rangle}{\langle x\rangle^{\rho}}\right)\ \mathrm{d}\zeta\le\frac{C}{\langle x\rangle^{\rho}}\left\|\frac{1}{\langle\zeta\rangle^{\rho-1}}\right\|_{L^2_\zeta}\|Q(\zeta)\|_{L^2_\zeta},$$

from which we deduce

$$\left|R_{V_0}(k)Q\ e_{V_0+}(x;k)\ -t_1[Q]\ e^{ikx}\right|\le\frac{C}{\langle x\rangle^{\rho}}\left(\|V_0\|_{L^{1,\rho+1}}+\left\|\frac{1}{\langle\zeta\rangle^{\rho-1}}\right\|_{L^2_\zeta}\|Q\|_{L^2}\right).$$

Similar estimates hold for $u_s = e_{V+}(x;k)-e_{V_0+}(x;k)$, and for the quadratic term $R_{V_0}(k)\ Q\ R_{V_0}(k)\ Q\ e_{V_0+}(x;k)$. Therefore, for $x>M$, one has

$$\left|\chi^{-1}\langle D_0\rangle^{-1}f_{rem}(x)\ -\ t_{rem}(k)e^{ikx}\right|\le\ \frac{C}{\langle x\rangle^{\rho}}\|V_0\|_{L^{1,\rho+1}}.$$

Since $\chi(x)=\langle x\rangle^{-\alpha}$ with $\alpha>4$, the pointwise bound (5.27) yields

$$t_{rem}(k)\ \le\ C_\chi\langle x\rangle^{\alpha}|||Q|||^3\ +\ C\frac{1}{\langle x\rangle^{\rho}},$$

so that choosing $x=|||Q|||^{-3/(\rho+\alpha)}$, which tends to infinity as $|||Q|||$ tends to 0, one has

$$|t_{rem}(k)|\ \le\ C\ |||Q|||^{\frac{3\rho}{\rho+\alpha}}.$$

It follows that with $\alpha>4$ and $\rho>2\alpha$, one has

$$|t_{rem}(k)|\ =\ \mathcal{O}(|||Q|||^{\beta}),\ \ 2<\beta\equiv\frac{3\rho}{\rho+\alpha}.$$

This completes the proof. □

5.4. Completion of the proof of Theorem 2.2. In this section, we show how to derive the corrected multi-scale / homogenization expansion of section 4 from the rigorous results of the previous section with a potential $V=V_0+q_\epsilon$ satisfying Hypotheses **(V)**, and using Proposition 5.4. Theorem 2.2 follows then as a direct consequence.

The small ϵ asymptotics of $t_1[q_\epsilon]$:

We use the decomposition of q_ϵ in Fourier series in y

$$q_\epsilon(x)=q\left(x,\frac{x}{\epsilon}\right)=\sum_{|j|\ge1}q_j(x)e^{2i\pi j(x/\epsilon)},$$

that we plug into $t_1[q_\epsilon]$, given in (5.21):

$$t_1[q_\epsilon]\ =\ \frac{1}{2ik}\sum_{|j|\ge1}t_1[q_\epsilon]^j,\ \text{with}\ t_1[q_\epsilon]^j=\int_{-\infty}^{+\infty}q_j(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k)e^{2i\pi j(\zeta/\epsilon)}\ \mathrm{d}\zeta.$$

We assume that q_j is piecewise C^3, so that there exists $-\infty=a_0<a_1<\cdots<a_M<a_{M+1}=\infty$, such that $q_j\in C^3(a_l,a_{l+1})$. Then, one has

$$t_1^{j,l}[q_\epsilon]=\frac{1}{2ik}\int_{a_l}^{a_{l+1}}q_j(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k)e^{2i\pi j(\zeta/\epsilon)}\ \mathrm{d}\zeta$$

$$= \frac{-1}{2ik}\int_{a_l}^{a_{l+1}} \partial_\zeta(q_j(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k))\frac{\epsilon}{2i\pi j}e^{2i\pi j(\zeta/\epsilon)}\,\mathrm{d}\zeta + b_1^{j,l}$$
$$= \frac{1}{2ik}\int_{a_l}^{a_{l+1}} \partial_\zeta^2(q_j(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k))\left(\frac{\epsilon}{2i\pi j}\right)^2 e^{2i\pi j(\zeta/\epsilon)}\,\mathrm{d}\zeta + b_1^{j,l} + b_2^{j,l},$$

with the following boundary terms

$$\begin{aligned} b_1^{j,l} &= \tfrac{-\epsilon}{4k\ \pi j}\big(q_j(a_{l+1}^-)e_{V_0+}(a_{l+1};k)e_{V_0-}(a_{l+1};k))e^{2i\pi j(a_{l+1}/\epsilon)} \\ &\quad -q_j(a_l^+)e_{V_0+}(a_l;k)e_{V_0-}(a_l;k))e^{2i\pi j(a_l/\epsilon)}\big), \\ b_2^{j,l} &= \tfrac{-i\ \epsilon^2}{8k\ \pi^2 j^2}\Big(\partial_\zeta(q_j(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k))|_{\zeta=a_{l+1}^-}\ e^{2i\pi j(a_{l+1}/\epsilon)} \\ &\quad -\partial_\zeta(q_j(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k))|_{\zeta=a_l^+}\ e^{2i\pi j(a_l/\epsilon)}\Big). \end{aligned}$$

Now, one has

$$\begin{aligned} \partial_x^2\big(q_j(x)e_{V_0+}(x;k)e_{V_0-}(x;k)\big) &= \frac{d^2q_j}{\mathrm{d}x^2}(x)e_{V_0+}(x;k)e_{V_0-}(x;k) \\ + 2\frac{dq_j}{\mathrm{d}x}(x)\partial_x(e_{V_0+}(x;k)e_{V_0-}(x;k)) &+ 2q_j(x)\partial_x e_{V_0+}(x;k)\partial_x e_{V_0-}(x;k) \\ &+ q_j(x)\big((\partial_x^2 e_{V_0+}(x;k))e_{V_0-}(x;k) + e_{V_0+}(x;k)\partial_x^2 e_{V_0-}(x;k)\big). \end{aligned}$$

The first three terms are piecewise-C^1, so that oscillatory integrals predict that

$$\begin{aligned} \int_{a_l}^{a_{l+1}} \Big(&\frac{d^2q_j}{\mathrm{d}\zeta^2}(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k) + \frac{dq_j}{\mathrm{d}\zeta}(\zeta)\partial_\zeta(e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k)) \\ &+2q_j(\zeta)\partial_\zeta e_{V_0+}(\zeta;k)\partial_\zeta e_{V_0-}(\zeta;k)\Big)e^{2i\pi j(\zeta/\epsilon)}\,\mathrm{d}\zeta = \mathcal{O}(\epsilon). \end{aligned} \tag{5.28}$$

For the fourth term, we use the fact that e_{V_0+} and e_{V_0-} satisfy $\big(-\frac{d^2}{dx^2}+V_0-k^2\big)u=0$, so that one has, with $\Omega_j = \{x_0,\dots,x_{N-1}\}\cap(a_j,a_{j+1})$,

$$\begin{aligned} t_1^{j,l}[q_\epsilon] &= \frac{i\ \epsilon^2}{8k\pi^2 j^2}\sum_{x_i\in\Omega_j} 2c_i q_j(x_i)e_{V_0+}(x_i;k)e_{V_0-}(x_i;k)e^{\frac{2i\pi j\ x_i}{\epsilon}} + b_1^{j,l} + b_2^{j,l} + \mathcal{O}\left(\epsilon^3/j^2\right) \\ &= \frac{i\ \epsilon^2}{8k\ \pi^2 j^2}\sum_{x_i\in\Omega_j}\left[\partial_x(q_j(x)e_{V_0+}(x;k)e_{V_0-}(x;k))\right]_{a_j} e^{\frac{2i\pi j\ x_i}{\epsilon}} + b_1^{j,l} + \mathcal{O}\left(\epsilon^3/j^2\right). \end{aligned}$$

Finally, we have $t_1[q_\epsilon] = \sum_{l=0}^{M-1}\ \sum_{|j|\geq 1}\ t_1^{j,l}[q_\epsilon] + \mathcal{O}(\epsilon^3)$, and one recovers immediately terms of the expansion of Theorem 2.2:

$$\sum_{l=0}^{M-1}\sum_{|j|\geq 1} b_1^{j,l} = \epsilon t_1^\epsilon \quad \text{and}$$
$$\sum_{l=0}^{M-1}\sum_{|j|\geq 1}\frac{i\ \epsilon^2}{8k\ \pi^2 j^2}\sum_{x_i\in\Omega_j}\Big[\partial_x(q_j(x)e_{V_0+}(x;k)e_{V_0-}(x;k))\Big]_{a_j} e^{2i\pi j(x_i/\epsilon)} = \epsilon^2 t_2^\epsilon + \mathcal{O}(\epsilon^3),$$

so that $t_1[q_\epsilon] = \epsilon t_1^\epsilon(k) + \epsilon^2 t_2^\epsilon(k) + \mathcal{O}(\epsilon^3)$.

The small ϵ asymptotics of $t_2[q_\epsilon, q_\epsilon]$:

Let us assume that ζ is fixed outside $\operatorname{supp} V_{sing}$, and outside the discontinuities of $q_j, \partial_x q_j$ (this particular case arises for a finite number of values of ζ, and therefore brings no contribution to the transmission coefficient, when integrated). Then integrating by part leads to the following expansion for ϵ small:

$$I_l^j(\zeta) \equiv -e_{V_0+}(\zeta;k)\int_{-\infty}^{\zeta}\partial_z\big(q_j(z)e_{V_0+}(z;k)e_{V_0-}(z;k)\big)\ \frac{\epsilon}{2i\pi j}e^{2i\pi j(z/\epsilon)}\,\mathrm{d}z$$

$$= \frac{\epsilon^2}{4\pi^2 j^2}\, e_{V_0+}(\zeta;k)\, \Big(-\int_{-\infty}^{\zeta} \partial_z^2\big(q_j(z)e_{V_0+}(z;k)e_{V_0-}(z;k)\big)\, e^{2i\pi j(z/\epsilon)}\, \mathrm{d}z$$
$$+ \Big[\partial_z\big(q_j(z)e_{V_0+}(z;k)e_{V_0-}(z;k)\big)\, e^{2i\pi j(./\epsilon)}\Big]_{-\infty}^{\zeta}\Big).$$

The first term, treated as previously and using the fact that the functions q_j, e_{V_0+}, and e_{V_0-} are piecewise-C^3, brings a contribution of order $\mathcal{O}(\epsilon^3)$.

Now, using the same analysis on $I_r^j(\zeta)$ and the Wronskian identity (3.8), one obtains the following expansion for the $-\mathrm{d}\zeta$ integrand of (5.22):

$$\begin{aligned} I_l^j(\zeta) + I_r^j(\zeta) &= \Big(e_{V_0+}(\zeta;k)\partial_\zeta(q_j(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k)) \\ &\quad - e_{V_0-}(\zeta;k)\partial_\zeta(q_j(\zeta)e_{V_0+}(\zeta;k)e_{V_0-}(\zeta;k))\Big) + \mathcal{O}(\epsilon^3) \\ &= -2ik\; t_0^{hom}\; q_j(\zeta)\; e_{V_0+}(\zeta;k)\; e^{2i\pi j(\zeta/\epsilon)} + \mathcal{O}(\epsilon^3). \end{aligned}$$

Therefore, one has

$$\begin{aligned} t_2[q_\epsilon, q_\epsilon] &= \frac{1}{2ik}\frac{1}{-2ik\; t_0^{hom}} \sum_{|j|\geq 1} \frac{\epsilon^2}{4\pi^2 j^2} \int_{-\infty}^{+\infty} \sum_{|m|\geq 1} q_m(\zeta) e^{2i\pi m(\zeta/\epsilon)} e_{V_0-}(\zeta;k) \\ &\quad \Big(-2ik\; t_0^{hom} q_j(\zeta) e_{V_0+}(\zeta;k) e^{2i\pi j(\zeta/\epsilon)}\Big)\; \mathrm{d}\zeta + \mathcal{O}(\epsilon^3) \\ &= \epsilon^2 \frac{i}{8k\pi^2} \int_{-\infty}^{+\infty} \sum_{|j|\geq 1} \frac{q_{-j}(\zeta)q_j(\zeta)}{j^2} e_{V_0-}(\zeta;k) e_{V_0+}(\zeta;k)\; \mathrm{d}\zeta + \mathcal{O}(\epsilon^3). \end{aligned} \tag{5.29}$$

One recovers finally: $t_2[q_\epsilon, q_\epsilon] = \epsilon^2 t_2^{hom} + \mathcal{O}(\epsilon^3)$.

Estimate of t_{rem}^ϵ:

Using Proposition 5.4 with Theorem 5.6 yields:

PROPOSITION 5.7. *Let K denote a compact subset of $\mathbb{R}$, satisfying Hypothesis **(G)**. Introduce for $k \in K$*

$$t_{rem}^\epsilon(k) \equiv t^\epsilon(k) - t_0^{hom}(k) - \epsilon t_1^\epsilon(k) - \epsilon^2 \Big(t_2^{hom}(k) + t_2^\epsilon(k) \Big). \tag{5.30}$$

Then we have

i. *If V has compact support, then $t_{rem}^\epsilon(k) = \mathcal{O}(\epsilon^3)$.*
ii. *If V is exponentially decreasing, then $t_{rem}^\epsilon(k) = \mathcal{O}(\epsilon^{3-})$.*
iii. *If $\langle x\rangle^\rho V_0 \in L^1$, $\rho > 9$, then there exists $2 < \beta < 3$ such that $t_{rem}^\epsilon(k) = \mathcal{O}(\epsilon^\beta)$.*

The proof of Theorem 2.2 is now complete.

A. The numerical computations

In this section we outline the numerical method we used to obtain results displayed in figures 1 and 2.

We approach the computation of $t(k)$, the transmission coefficient associated with the potential $V(x)$, by numerical approximation of the function

$$u(x;k) \equiv \frac{1}{t(k)} e_{V-}(x;k),$$

where $e_{V-}(x;k)$ denotes the distorted plane wave generate by an incoming wave from positive infinity; see (3.7). We rewrite the equation

$$\Big(-\frac{d^2}{dx^2} + V(x) - k^2\Big)\, u(x;k) = 0,$$

equivalently in terms of the variable $U(x;k) \equiv (u(x;k), \partial_x u(x;k))^T$ as the first order system

$$\frac{d}{\mathrm{d}x}U \;=\; \begin{pmatrix} 0 & 1 \\ V(x)-k^2 & 0 \end{pmatrix} U. \tag{A.1}$$

Note that if V is assumed to have compact support ($\operatorname{supp} V \subset [-M, M]$ with $M > 0$), then

$$U(x;k) \;\equiv\; \begin{pmatrix} e^{-ikx} \\ -ike^{-ikx} \end{pmatrix} \quad \text{for } x < -M, \tag{A.2}$$

$$U(x;k) \;\equiv\; \begin{pmatrix} \frac{r_r(k)}{t(k)}e^{ikx} + \frac{1}{t(k)}e^{-ikx} \\ \frac{ik\ r_r(k)}{t(k)}e^{ikx} - \frac{ik}{t(k)}e^{-ikx} \end{pmatrix} \quad \text{for } x > M. \tag{A.3}$$

Starting with the initial data given by (A.2), we numerically solve the system of first order ODEs defined by (A.1) up to $x > M$, and (A.3) allows to recover the desired value of $t(k)$. At the location of the singularities $x = x_j$, the jump conditions (3.3) allow to obtain $U(x+;k)$ from $U(x-;k)$ via a transfer matrix. Between the singularities, one approximatively solves (A.1) using for example Runge-Kutta formulae. We used the Matlab solver `ode45`; see [**171**] for more information about the Matlab ODE Suite.

We conclude this section by stating the precise functions and parameters used to obtain the plots displayed in figures 1 and 2.

For the case when V_0 has singularities, as in the left and center panels of figure 1, we set

$$V_0 = V_{sing}(x) \equiv 40\left(\delta(x) + \delta(x-0.5) + \delta(x-1)\right).$$

Otherwise, we set

$$V_0 = V_{reg}(x) \equiv 40\left(\delta_\rho(x) + \delta_\rho(x-0.5) + \delta_\rho(x-1)\right),$$

with $\delta_\rho(x) \equiv \frac{1}{\rho\sqrt{\pi}}\mathrm{e}^{-x^2/\rho^2}$ the smoothed out approximation. One has $\rho = 0.1$ for the right panels of figures 1 and 2, and respectively $\rho = 0.01$ and $\rho = 0.001$ for the center and left panels of figure 2.

We set $q_\epsilon(x) \;=\; f(x)\sin(2\pi x/\epsilon)$, with $f(x) \equiv 0$ for $x \in \mathbb{R} \setminus [-2/3; 2/3]$, and elsewhere

$$\begin{cases} f(x) = 40 & \text{in the discontinuous case (left panel of figure 1), or} \\ f(x) = 40e^{-\frac{x^2}{(x-2/3)(x+2/3)}} & \text{in the smooth cases (all other panels).} \end{cases}$$

Finally, we set $k = 5.5$, since it corresponds to a case where $t_0^{hom}(k)$ approaches unity when $V_0 = V_{sing}$.

B. The Jost solutions

In this section, we provide a construction of the Jost solutions and a rigorous derivation of their properties, including bounds that are used in the proof of Proposition 5.5, Appendix C. We recall that by Definition 3.2, the Jost solutions are the unique solutions $f_\pm(x;k)$ of

$$\left(\ H_W \;-\; k^2\ \right) u \;\equiv\; \left(-\frac{d^2}{dx^2} + W(x) - k^2\right) u = 0. \tag{B.1}$$

such that $f_\pm(x;k) = e^{\pm ikx}\ m_\pm(x;k)$ and

$$\lim_{x\to\pm\infty} m_\pm(x;k) \;=\; 1.$$

The existence of Jost solutions for regular potentials $W \in L^{1,3/2+}(\mathbb{R})$ is established in [**135**]. The generalization to potentials allowing a singular component

$$W = W_{reg} + W_{sing}, \quad \text{with}$$

$$W_{reg} \in L^{1,3/2+}(\mathbb{R}),$$
$$W_{sing} = \sum_{j=0}^{N-1} c_j \, \delta(x - x_j), \quad \text{where} \quad c_j, x_j \in \mathbb{R}, \quad x_j < x_{j+1}.$$

can be found in [**136**].

As an intermediate step of the proof, one introduces an equivalent definition of the Jost solution, as solutions of integral equations. In the case where W is regular, one has

$$m_+(x;k) = 1 + \int_x^\infty D_k(\zeta - x) W(\zeta) m_+(\zeta;k) \mathrm{d}\zeta, \tag{B.2}$$
$$m_-(x;k) = 1 + \int_{-\infty}^x D_k(x - \zeta) W(\zeta) m_-(\zeta;k) \mathrm{d}\zeta, \qquad D_k(x) = \int_0^x e^{2ik\zeta} \mathrm{d}\zeta.$$

If W has regular and singular components, we work with a variant of equations (B.2):

$$m_+(x;k) = 1 + \int_x^\infty D_k(\zeta - x) W(\zeta) m_+(\zeta;k) \mathrm{d}\zeta \; + \sum_{x_j > x} D_k(x_j - x) c_j m_+(x_j;k),$$
$$m_-(x;k) = 1 + \int_{-\infty}^x D_k(x - \zeta) W(\zeta) m_-(\zeta;k) \mathrm{d}\zeta \; + \sum_{x_j < x} D_k(x_j - x) c_j m_+(x_j;k).$$

From these integral equations, one deduces

$$|m_+(x;k) - 1| \leq \frac{1 + \max(-x, 0)}{1 + |k|} \int_x^\infty (1 + |s|) |W(s)| \mathrm{d}s,$$
$$|m_-(x;k) - 1| \leq \frac{1 + \max(-x, 0)}{1 + |k|} \int_{-\infty}^{-x} (1 + |s|) |W(s)| \mathrm{d}s. \tag{B.3}$$

Then, since m_+ satisfies

$$\partial_x m_+(x;k) = \int_x^\infty e^{2ik(t-x)} W(t) m_+(t;k), \text{ and}$$
$$\partial_k m_+(x;k) = \int_x^\infty D_k(t - x) W(t) \partial_k m_+(t;k) \; + \int_x^\infty \partial_k D_k(t - x) W(t) m_+(t;k),$$

one obtains easily the following uniform bounds

$$|m_+(x;k)| \leq C\langle x \rangle, \qquad |\partial_x m_+(x;k)| \leq C,$$
$$|\partial_k m_+(x;k)| \leq C\langle x \rangle^2, \qquad |\partial_x \partial_k m_+(x;k)| \leq C\langle x \rangle, \tag{B.4}$$

where C is independent of k. The same bounds clearly hold for $m_-(x;k)$.

C. Proof of Proposition 5.5

This Section is dedicated to the proof of Proposition 5.5, namely

$$T_{R_{V_0}}(k) \equiv \langle D_0 \rangle \, \chi R_{V_0}(k) \chi \, \langle D_0 \rangle \text{ is a bounded operator from } L^2 \text{ to } L^2.$$

This result has been proved by in [**144**], for $V_0 \equiv 0$, and spatial dimensions $n = 1, 2, 3$. We generalize this result in the one dimensional case for $V_0 = V_{reg} + V_{sing}$ as in (2.1), so that singularities in the potential are allowed.

Our proof requires the use of the generalized Fourier transform, described in terms of the distorted plane waves. We introduce

$$\Psi(x;\zeta) = \frac{1}{\sqrt{2\pi}} \begin{cases} e_{V_0+}(x;\zeta) & \zeta \geq 0, \\ e_{V_0-}(x;-\zeta) & \zeta < 0, \end{cases} \equiv \frac{1}{\sqrt{2\pi}} \begin{cases} t(\zeta) m_+(x;\zeta) e^{ix\zeta} & \zeta \geq 0, \\ t(-\zeta) m_-(x;-\zeta) e^{ix\zeta} & \zeta < 0, \end{cases}$$

with $m_+(x;\zeta) \to 0$ as $x \to \infty$ and $m_-(x;\zeta) \to 0$ as $x \to -\infty$.

Then $\mathcal{F}$ and $\mathcal{F}^*$ the distorted Fourier transform and its adjoint are defined by

$$\begin{array}{rcll} \mathcal{F} & : & L^2 & \to L^2 \\ & & \phi & \mapsto \mathcal{F}[\phi](\xi) \equiv \int_{-\infty}^{+\infty} \phi(x)\overline{\Psi(x,\xi)}\ \mathrm{d}x, \\ \mathcal{F}^* & : & L^2 & \to L^2 \\ & & \Phi & \mapsto \int_{-\infty}^{+\infty} \Phi(\xi)\Psi(x,\xi)\ \mathrm{d}\xi. \end{array}$$

One has the following property:

$$P_c\phi = \mathcal{F}^*\mathcal{F}\phi,$$

where P_c denotes the spectral projection onto the continuous spectral subspace associated with the operator

$$H \equiv -\partial_x^2 + V_0. \tag{C.1}$$

To construct a smoothing operator which commutes with functions of H, it is convenient to introduce, using the distorted plane wave spectral representation of H:

$$\langle D_{V_0}\rangle\ f\ =\ (I-\Delta+V_0)^{1/2}\ f\ =\ \int_{\mathbb{R}} \langle\eta\rangle \mathcal{F}[f](\eta)\ \Psi(x;\eta)\ \mathrm{d}\eta \tag{C.2}$$

Therefore, one has

$$T_{R_{V_0}} = \langle D_0\rangle \langle D_{V_0}\rangle^{-1} \langle D_{V_0}\rangle \chi R_{V_0}(k)\chi \langle D_{V_0}\rangle \langle D_{V_0}\rangle^{-1} \langle D_0\rangle \tag{C.3}$$

$$\equiv \langle D_0\rangle \langle D_{V_0}\rangle^{-1} \circ \tilde{T}_{R_{V_0}} \circ \langle D_{V_0}\rangle^{-1} \langle D_0\rangle . \tag{C.4}$$

There are thus three terms to estimate. In order to deal with $\langle D_0\rangle \langle D_{V_0}\rangle^{-1}$ and $\langle D_{V_0}\rangle^{-1} \langle D_0\rangle$, we introduce the classical wave operator, W and its adjoint W^*, defined by

$$W \equiv s - \lim_{t\to\infty} e^{itH}e^{-itH_0}, \tag{C.5}$$

$$W^* \equiv s - \lim_{t\to\infty} e^{itH_0}e^{-itH}P_c, \tag{C.6}$$

with $H \equiv -\partial_x^2 + V_0$ and $H_0 \equiv -\partial_x^2$. The wave operators have the property to intertwine between the continuous part of H and H_0, so that for any Borel function f:

$$f(H)P_c = W f(H_0) W^*.$$

Especially, one has $\langle D_{V_0}\rangle = W \langle D_0\rangle W^*$, so that

$$\langle D_0\rangle \langle D_{V_0}\rangle^{-1} = \langle D_0\rangle W \langle D_0\rangle^{-1} W^*. \tag{C.7}$$

Let us state the following result, that has been introduced in **[178]** and extended in **[136]** to potentials $V_0 = V_{reg} + V_{sing}$ as in (2.1), thus allowing Dirac delta functions:

Lemma C.1. *W and W^* have extensions to bounded operators on H^k, for $k = -1, 0, 1$.*

Using this last result and the known fact that $\langle D_0\rangle^s$ is bounded from H^k to H^{k-s}, we obtain directly from (C.7) that

$$\langle D_0\rangle \langle D_{V_0}\rangle^{-1} \text{ is bounded from } L^2 \text{ to } L^2.$$

Similarly,

$$\langle D_{V_0}\rangle^{-1} \langle D_0\rangle \text{ is bounded from } L^2 \text{ to } L^2.$$

In order to deal with the last term of (C.4), we decompose $\tilde{T}_{R_{V_0}}$ as a sum of four operators, commuting $\langle D_0\rangle$.

$$\begin{aligned} \tilde{T}_{R_{V_0}} &\equiv \langle D_{V_0}\rangle \chi R_{V_0}(k)\chi \langle D_{V_0}\rangle \\ &= \left(\chi \langle D_{V_0}\rangle + [\langle D_{V_0}\rangle, \chi]\right) R_{V_0}(k) \left(\langle D_{V_0}\rangle \chi + [\chi, \langle D_{V_0}\rangle]\right) \end{aligned}$$

$$
\begin{aligned}
=&\chi \langle D_{V_0}\rangle R_{V_0}(k) \langle D_{V_0}\rangle \chi + ([\langle D_{V_0}\rangle, \chi])(R_{V_0}(k) \langle D_{V_0}\rangle \chi) \\
&+ (\chi \langle D_{V_0}\rangle R_{V_0}(k))([\chi, \langle D_{V_0}\rangle]) + ([\langle D_{V_0}\rangle, \chi])(R_{V_0}(k))([\chi, \langle D_{V_0}\rangle]) \\
=&A_I + A_{II}^{(a)} + A_{II}^{(b)} + A_{III}.
\end{aligned}
$$

Each of these terms is proved to be bounded from L^2 to L^2. We treat each term separately in Propositions C.2, C.4, C.5, and C.6.

PROPOSITION C.2. *$A_I \equiv \chi \langle D_{V_0}\rangle R_{V_0}(k) \langle D_{V_0}\rangle \chi$ is bounded $L^2 \to L^2$, i.e.*

$$
\| \chi \langle D_{V_0}\rangle R_{V_0}(k) \langle D_{V_0}\rangle \chi \, g \|_{L^2} \leq C \|g\|_{L^2}, \quad g \in L^2(\mathbb{R}) \tag{C.8}
$$

First we commute $\langle D_{V_0}\rangle$ and R_{V_0}. It is obvious that R_0 and $\langle D_0\rangle$ commute, so that using the wave operators introduced above (so that $\langle D_{V_0}\rangle = W \langle D_0\rangle W^*$ and $R_{V_0}(k) = W R_0(k) W^*$, with W unitary).

$$
\begin{aligned}
A_I &= \chi \langle D_{V_0}\rangle R_{V_0}(k) \langle D_{V_0}\rangle \chi \\
&= \chi W \langle D_0\rangle W^* W R_0(k) W^* W \langle D_0\rangle W^* \chi \\
&= \chi W R_0(k) \langle D_0\rangle^2 W^* \chi \\
&= \chi R_{V_0}(k) \langle D_{V_0}\rangle^2 \chi.
\end{aligned}
$$

Then, applying the identity $\langle D_{V_0}\rangle^2 = I - \Delta + V_0$, one obtains

$$
A_I = (1+k^2)\chi R_{V_0}(k)\chi + \chi^2.
$$

Finally, using (3.11) together with (B.4), one has the pointwise bound

$$
|R_{V_0}(x,y;k)| \leq C\langle x\rangle\langle y\rangle
$$

with C uniform in k. It follows that for $f \in L^2$,

$$
\left|\chi R_{V_0}(k)\chi f\right|_{L^2} = \left|\chi(x) \int_\zeta R_{V_0}(x,\zeta;k)\chi(\zeta) f(\zeta)\, \mathrm{d}\zeta\right|_{L^2_x} \leq C \left|\chi(x)\langle x\rangle\right|^2_{L^2_x} |f|_{L^2},
$$

so that A_I is bounded from L^2 to L^2 , with

$$
\left\|A_I\right\|_{L^2\to L^2} \leq C\left(\left|\chi(x)\langle x\rangle\right|^2_{L^2_x} + |\chi|_{L^\infty}\right). \tag{C.9}
$$

Before carrying on with estimating the term $A_{II}^{(a)}$, let us state the following Lemma.

LEMMA C.3. *Let K be defined for $(\xi,\eta) \in \mathbb{R}\times\mathbb{R}$ by*

$$
K(\xi,\eta) \equiv (\langle\xi\rangle - \langle\eta\rangle) \int_\zeta \overline{\Psi(\zeta;\xi)}\, \Psi(\zeta;\eta)\, \chi(\zeta)\, d\zeta \tag{C.10}
$$

Then $K(\xi,\eta)$ satisfies the following upper bounds:

$$
| K(\xi,\eta) | \leq \frac{C_\chi}{1+|\xi-\eta|}, \tag{C.11}
$$

$$
| \partial_\eta K(\xi,\eta) | + | \partial_\xi K(\xi,\eta) | \leq \frac{C'_\chi}{1+|\xi-\eta|}, \tag{C.12}
$$

with the C_χ and C'_χ constants depending on the function χ with

$$
C_\chi \equiv C \left(\sum_{j=0}^{2} \left\|\langle\zeta\rangle^j\, \partial^j_\zeta \chi\right\|_{L^1_\zeta} + \left\|\langle\zeta\rangle^2\, \chi\right\|_{L^1_\zeta} + \left\|\langle\zeta\rangle^2\, \chi\right\|_{L^\infty_\zeta}\right)
$$

$$C'_\chi \equiv C\ \left(\sum_{j=0}^{2} \left\| \langle\zeta\rangle^{j+1}\ \partial_\zeta^j \chi \right\|_{L^1_\zeta} + \left\| \langle\zeta\rangle^3\ \chi \right\|_{L^1_\zeta} + \left\| \langle\zeta\rangle^3\ \chi \right\|_{L^\infty_\zeta} \right)$$

Proof. We consider the case where $\xi \geq 0$ and $\eta \geq 0$. The other cases follow similarly. Therefore, one has

$$K(\xi,\eta)\ = (\langle\xi\rangle - \langle\eta\rangle)\ I(\xi,\eta), \quad \text{with}$$

$$I(\xi,\eta) \equiv \int_\zeta \overline{\Psi(\zeta;\xi)}\Psi(\zeta;\eta)\chi(\zeta)\ \mathrm{d}\zeta = \int_\zeta e^{i\zeta(\eta-\xi)}\overline{t(\xi)m_+(\zeta;\xi)}t(\eta)m_+(\zeta;\eta)\chi(\zeta)\ \mathrm{d}\zeta. \tag{C.13}$$

Throughout the proof, we will use extensively the uniform bounds on m_+ given in (B.4).

First, by the uniform boundedness of $t(\xi)$ and $\langle\zeta\rangle^{-1}\ m_+(\zeta,\xi)$ in ζ and ξ, one has

$$|I(\xi,\eta)| \leq |t(\xi)||t(\eta)|\ \int_\zeta\ |\langle\zeta\rangle^{-1}m_+(\zeta;\eta)\ \langle\zeta\rangle^{-1}m_+(\zeta;\xi)|\ \langle\zeta\rangle^2|\chi(\zeta)|\ \mathrm{d}\zeta\ \leq C\ \|\langle\zeta\rangle^2\chi\|_{L^1_\zeta}. \tag{C.14}$$

For $|\eta-\xi| \geq 1$ we write

$$I(\xi,\eta) \equiv \frac{1}{(i(\eta-\xi))^2}\overline{t(\xi)}t(\eta)\ \int_\zeta \left(\ \frac{d^2}{\mathrm{d}\zeta^2}\ e^{i\zeta(\eta-\xi)}\ \right)\ \overline{m_+(\zeta;\eta)}\ m_+(\zeta;\xi)\ \chi(\zeta)\ \mathrm{d}\zeta \tag{C.15}$$

$$= \frac{1}{(i(\eta-\xi))^2}\ \overline{t(\xi)}t(\eta)\int_\zeta e^{i\zeta(\eta-\xi)}\ \frac{d^2}{\mathrm{d}\zeta^2}\ \left(\ \overline{m_+(\zeta;\eta)}\ m_+(\zeta;\xi)\ \chi(\zeta)\right)\ \mathrm{d}\zeta. \tag{C.16}$$

The most singular terms in the integrand of (C.16) are those containing $\partial_\zeta^2 m_+$. In particular, recall the relation $\partial_x^2 m_+ = -2ik\partial_x m_+ + V_0 m_+$, where V_0 contains Dirac mass singularities. Thus, for $|\xi-\eta| \geq 1$ we have

$$|I(\xi,\eta)|\ \leq\ C\ \left(\sum_{j=0}^{2} \left\| \langle\zeta\rangle^j\ \partial_\zeta^j\chi \right\|_{L^1_\zeta} + \|\langle\zeta\rangle\ \chi\|_{L^1_\zeta} + \left\| \langle\zeta\rangle^2\ \chi \right\|_{L^\infty} \right)\ \cdot\ \frac{1}{|\xi-\eta|^2} \tag{C.17}$$

Applying (C.14) for $|\eta-\xi| \leq 1$ and (C.17) for $|\eta-\xi| \geq 1$ yields

$$|I(\xi,\eta)|\ \leq\ C_\chi \frac{1}{1+|\xi-\eta|^2}.$$

Finally, since $|K(\xi,\eta)|\ =\ |I(\xi,\eta)|\,|\langle\xi\rangle - \langle\eta\rangle|\ \leq\ C\,|I(\xi,\eta)|\,|\xi-\eta|$, multiplication by $|\xi-\eta|$ implies (C.11).

Using the same method as previously, one obtains similarly

$$|\partial_\eta I(\xi,\eta)|\ \leq\ C'_\chi \frac{1}{1+|\xi-\eta|^2}.$$

Finally, one has $|\partial_\eta K(\xi,\eta)|\ \leq\ |\partial_\eta I(\xi,\eta)|\,|\langle\xi\rangle - \langle\eta\rangle|\ + |I(\xi,\eta)|$, so that we deduce the first part of (C.12). By symmetry, one obtains the same estimate for $\partial_\xi K(\xi,\eta)$, which concludes the proof of Lemma C.3. □

Proposition C.4. *$A_{II}^{(a)} \equiv [\langle D_{V_0}\rangle, \chi] R_{V_0} \langle D_{V_0}\rangle \chi$ is bounded $L^2 \to L^2$, i.e.*

$$\|\ [\langle D_{V_0}\rangle, \chi] R_{V_0} \langle D_{V_0}\rangle \chi\ g\ \|_{L^2}\ \leq\ C\ \|g\|_{L^2}, \quad g \in L^2(\mathbb{R}). \tag{C.18}$$

Proof. Our strategy is as follows. We view the operator $A_{II}^{(a)}$ as a composition of two operators

$$A_{II}^{(a)}\ =\ [\langle D_{V_0}\rangle, \chi]\ \circ\ R_{V_0}\langle D_{V_0}\rangle \chi$$

and first find a representation of each operator with respect to the distorted Fourier basis. We then directly prove the boundedness of $A_{II}^{(a)} : L^2 \mapsto L^2$ using this spectral representation and an appropriate frequency localization argument.

In terms of the distorted Fourier transform, one has

$$\begin{aligned}[\langle D_{V_0}\rangle, \chi] f(x) &= [\langle D_{V_0}\rangle, \chi]\left(\int_\eta \Psi(x;\eta)\mathcal{F}[f](\eta)\ \mathrm{d}\eta\right)\\ &= \int_\eta \mathcal{F}[f](\eta)\Big(\langle D_{V_0}\rangle(\chi(x)\Psi(x;\eta)) - \chi\langle D_{V_0}\rangle\Psi(x;\eta)\Big)\ \mathrm{d}\eta, \end{aligned} \tag{C.19}$$

Now, since $\langle D_{V_0}\rangle \Psi(x;\eta) = \langle\eta\rangle\Psi(x;\eta)$, one has

$$\begin{aligned}\langle D_{V_0}\rangle(\chi(x)\Psi(x;\eta)) &= \int_\xi \Psi(x;\xi)\int_\zeta \langle D_{V_0}\rangle(\chi\Psi(\cdot;\eta))(\zeta)\overline{\Psi(\zeta;\xi)}\ \mathrm{d}\zeta\ \mathrm{d}\xi\\ &= \int_\xi \Psi(x;\xi)\int_\zeta \chi(\zeta)\Psi(\zeta;\eta)\langle D_{V_0}\rangle\overline{\Psi(\cdot;\xi)}(\zeta)\ \mathrm{d}\zeta\ \mathrm{d}\xi\\ &= \int_\xi \Psi(x;\xi)\int_\zeta \chi(\zeta)\Psi(\zeta;\eta)\langle\xi\rangle\overline{\Psi(\zeta;\xi)}\ \mathrm{d}\zeta\ \mathrm{d}\xi.\end{aligned}$$

Therefore, we finally deduce

$$\begin{aligned}[\langle D_{V_0}\rangle, \chi] f(x) &= \int_\eta \mathcal{F}[f](\eta)\left(\int_\xi \Psi(x;\xi)\int_\zeta \Psi(\zeta;\eta)\overline{\Psi(\zeta;\xi)}\chi(\zeta)\left(\langle\xi\rangle - \langle\eta\rangle\right)\ \mathrm{d}\zeta\ \mathrm{d}\xi\ \right)\mathrm{d}\eta\\ &= \int_\xi \Psi(x;\xi)\int_\eta \left(\langle\xi\rangle - \langle\eta\rangle\right)\int_\zeta \overline{\Psi(\zeta;\xi)}\Psi(\zeta;\eta)\chi(\zeta)\mathcal{F}[f](\eta)\ \mathrm{d}\eta\mathrm{d}\xi.\end{aligned} \tag{C.20}$$

To represent the operator $R_{V_0}\langle D_{V_0}\rangle\chi$ in terms of the distorted Fourier basis we note:

$$\begin{aligned}\mathcal{F}[R_{V_0}\langle D_{V_0}\rangle\chi g](\eta) &= \int_\eta \overline{\Psi(z;\eta)}\ (R_{V_0}\langle D_{V_0}\rangle\chi g)(z)\ \mathrm{d}z\\ &= \int_z (R_{V_0}\langle D_{V_0}\rangle\overline{\Psi(z;\eta)})\chi(z)g(z)\ \mathrm{d}z\\ &= \int_z \frac{\langle\eta\rangle}{\eta^2-k^2}\ \overline{\Psi(z;\eta)}\chi(z)g(z)\ \mathrm{d}z\\ &= \frac{\langle\eta\rangle}{\eta^2-k^2}\ \mathcal{F}[\chi g](\eta).\end{aligned} \tag{C.21}$$

Combining (C.20) and (C.21), one has

$$\begin{aligned}[\langle D_{V_0}\rangle, \chi]R_{V_0}\langle D_{V_0}\rangle\chi g(x) &= \int_\xi \Psi(x;\xi)\int_\eta \left(\langle\xi\rangle - \langle\eta\rangle\right)\int_\zeta \overline{\Psi(\zeta;\xi)}\Psi(\zeta;\eta)\chi(\zeta)\ \mathrm{d}\zeta\\ &\qquad \frac{\langle\eta\rangle}{\eta^2-k^2}\mathcal{F}[\chi g](\eta)\ \mathrm{d}\eta\ \mathrm{d}\xi\\ &= \int_\xi \Psi(x;\xi)T^{II}[g](\xi)\mathrm{d}\xi.\end{aligned} \tag{C.22}$$

By the Plancherel Theorem, the L^2 estimate of $A_{II}^{(a)}$ is equivalent to the bound

$$\left\|T^{II}[g]\right\|_{L^2} = \left\|\int_\eta\int_\zeta \overline{\Psi(\zeta;\xi)}\ \Psi(\zeta;\eta)\ \chi(\zeta)\ \mathrm{d}\zeta\frac{\left(\langle\xi\rangle - \langle\eta\rangle\right)\langle\eta\rangle}{\eta^2-k^2}\ \mathcal{F}[\chi g](\eta)\ \mathrm{d}\eta\right\|_{L^2_\xi} \leq C\|g\|_{L^2} \tag{C.23}$$

We now proceed with a proof of (C.23). First we define $\varphi_{|\kappa|<\delta_0}$ to be the positive smooth function satisfying

(C.24) $\varphi_{|\kappa|<\delta_0}$ equal to one for $|\kappa| \leq \delta_0/2$, zero for $|\kappa| > \delta_0$ and symmetric about $\kappa = 0$.

We use φ to localize at frequencies near $\eta = \pm k$ and frequencies away from $\eta = \pm k$.

$$T^{II}[g] \equiv T^{II}_{near}[g] + T^{II}_{far}[g] \tag{C.25}$$

where

$$T^{II}_{far}[g](\xi) \equiv \int_\eta K(\xi,\eta)\ \frac{\langle\eta\rangle}{\eta^2-k^2}\ \left[1-\varphi_{||\eta|-|k||<\delta_0}(\eta)\right]\mathcal{F}[\chi g](\eta)\ \mathrm{d}\eta, \tag{C.26}$$

$$T^{II}_{near}[g](\xi) \equiv \int_\eta K(\xi,\eta)\ \frac{\langle\eta\rangle}{\eta^2-k^2}\ \varphi_{||\eta|-|k||<\delta_0}(\eta)\ \mathcal{F}[\chi g](\eta)\ \mathrm{d}\eta, \tag{C.27}$$

with K defined as in (C.10) by

$$K(\xi,\eta) \equiv (\langle\xi\rangle - \langle\eta\rangle)\ \int_\zeta \overline{\Psi(\zeta;\xi)}\ \Psi(\zeta;\eta)\ \chi(\zeta)\ \mathrm{d}\zeta.$$

Bound on $T^{II}_{far}[g](\xi)$: We bound the expression

$$T^{II}_{far}[g](\xi) \equiv \int_\eta K(\xi,\eta)\ \frac{\langle\eta\rangle}{\eta^2-k^2}\ \left[1-\varphi_{||\eta|-|k||<\delta_0}(\eta)\right]\mathcal{F}[\chi g](\eta)\ \mathrm{d}\eta. \tag{C.28}$$

By Lemma C.3, $K(\xi,\eta)$ satisfies the following pointwise bound, which is valid for all $\xi,\eta\in\mathbb{R}$:

$$\mid K(\xi,\eta)\mid\ \leq\ C_\chi\ \frac{1}{1+|\xi-\eta|}.$$

Recall now the special case of Young's inequality:

$$\|h\star g\|_2 \leq \|h\|_2\ \|g\|_1\ .$$

This, together with the pointwise bound of $K(\xi,\eta)$, yields:

$$\begin{aligned}
\|\ T^{II}_{far}[g]\ \|_2 &= \left\| \int K(\xi,\eta)\ \frac{\langle\eta\rangle}{\eta^2-k^2}\ \left[1-\varphi_{||\eta|-|k||<\delta_0}(\eta)\right]\ \mid\mathcal{F}[\chi g](\eta)\mid\ \mathrm{d}\eta\ \right\|_{L^2_\xi}\\
&\leq\ C_\chi\ \left\|\frac{1}{\langle\eta\rangle}\ \right\|_{L^2_\eta}\ \left\|\frac{\langle\eta\rangle}{\eta^2-k^2}\left[1-\varphi_{||\eta|-|k||<\delta_0}(\eta)\right]\mathcal{F}[\chi g](\eta)\ \right\|_{L^1_\eta}\\
&\leq\ C_\chi\ \left\|\frac{1}{\langle\eta\rangle}\ \right\|^2_{L^2_\eta}\ \|\mathcal{F}[\chi g]\ \|_{L^2_\eta}\ \leq\ C_\chi\ \|\ \chi g\ \|_{L^2_\eta}\\
&\leq\ C_\chi\ \|\chi\|_{L^\infty}\ \|g\|_{L^2}.
\end{aligned}$$

Bound on $T^{II}_{near}[g](\xi)$:

$$\begin{aligned}
T^{II,\varepsilon}_{near}[g](\xi) &\equiv \int_\eta K(\xi,\eta)\ \frac{\langle\eta\rangle}{\eta^2-k^2}\varphi_{\varepsilon\leq||\eta|-|k||<\delta_0}(\eta)\ \mathcal{F}[\chi g](\eta)\ \mathrm{d}\eta\\
&= \frac{1}{2k}\ \int_\eta K(\xi,\eta)\ \langle\eta\rangle\ \varphi_{\varepsilon\leq||\eta|-|k||<\delta_0}(\eta)\ \left(\frac{\mathcal{F}[\chi g](\eta)}{\eta-k}\ -\ \frac{\mathcal{F}[\chi g](\eta)}{\eta+k}\ \right)\mathrm{d}\eta\\
&\equiv \int \Lambda^\varepsilon(\xi,\eta)\ \frac{\mathcal{F}[\chi g](\eta)}{\eta-k}\mathrm{d}\eta\ +\ \int \Lambda^\varepsilon(\xi,\eta)\ \frac{\mathcal{F}[\chi g](\eta)}{\eta+k}\ \mathrm{d}\eta,
\end{aligned}$$

where

$$\Lambda^\varepsilon(\xi,\eta) \equiv \frac{1}{2k}\ \langle\eta\rangle\ \varphi_{\varepsilon\leq||\eta|-|k||<\delta_0}(\eta)\ K(\xi,\eta). \tag{C.29}$$

and $K(\xi,\eta)$ is displayed in (C.10). Note that by Lemma C.3,

$$| \Lambda^\varepsilon(\xi,\eta) | \ \leq \ C_\chi \frac{1}{1+|\xi-\eta|}\varphi_{\varepsilon\leq||\eta|-|k||<\delta_0}(\eta). \tag{C.30}$$

We bound the first term in the above expansion of $T^{II,\varepsilon}_{near}$. The second term is treated similarly. We have

$$\int \Lambda^\varepsilon(\xi,\eta)\ \frac{\mathcal{F}[\chi g](\eta)}{\eta-k}\mathrm{d}\eta \ = \ \mathcal{S}^\varepsilon(\xi) \ + \ \mathcal{E}^\varepsilon(\xi) + \ \mathcal{R}^\varepsilon(\xi), \quad \text{where}$$

$$\mathcal{S}^\varepsilon(\xi) \equiv \Lambda^\varepsilon(\xi,k)\ \int \frac{\mathcal{F}[\chi g](\eta)}{\eta-k}\ \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}\ \mathrm{d}\eta,$$

$$\mathcal{E}^\varepsilon(\xi) \equiv \ \int (\Lambda^\varepsilon(\xi,\eta) \ - \ \Lambda^\varepsilon(\xi,k))\ \frac{\mathcal{F}[\chi g](\eta)}{\eta-k}\ \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}\ \mathrm{d}\eta,$$

$$\mathcal{R}^\varepsilon(\xi) \equiv \ \int \Lambda^\varepsilon(\xi,\eta)\ \frac{\mathcal{F}[\chi g](\eta)}{\eta-k}\mathbf{1}_{|\eta-k|\geq\delta_0/4}\ \mathrm{d}\eta.$$

One bounds $\mathcal{R}^\varepsilon$ using (C.30) by

$$\begin{aligned} \| \ \mathcal{R}^\varepsilon \ \|_{L^2} \ &\leq \ \frac{4C_\chi}{\delta_0}\left\| \frac{1}{1+|\eta|}\ \right\|_{L^2_\eta} \|\varphi_{\varepsilon\leq||\eta|-|k||<\delta_0}(\eta)\mathcal{F}[\chi g](\eta)\|_{L^1} \\ &\leq \ \frac{4C_\chi}{\delta_0}\|\varphi_{\varepsilon\leq||\eta|-|k||<\delta_0}\|_{L^2_\eta}\|\chi\|_{L^\infty}\ \|g\|_{L^2}. \end{aligned} \tag{C.31}$$

Moreover, we have

$$\begin{aligned} \left| \ \frac{\Lambda^\varepsilon(\xi,\eta)-\Lambda^\varepsilon(\xi,k)}{\eta-k}\ \right| \ &\leq \ \left| \ \partial_\eta \ \Lambda^\varepsilon(\xi,\eta) \ |_{\eta=\tilde\eta\in\{\varepsilon\leq|\eta-k|\leq\delta_0/4\}} \ \right| \\ &\leq \mathbf{1}_{|\eta-k|\leq\delta_0}\ (\ |\ \partial_\eta \ K(\xi,\eta)\ |\ \langle\eta\rangle \ + \ |K(\xi,\eta)|\ )\,. \end{aligned} \tag{C.32}$$

From the estimates of Lemma C.3, and using Young's inequality, one deduces

$$\| \ \mathcal{E}^\varepsilon \ \|_{L^2} \ \leq \ C'_\chi \ \|\chi\|_{L^\infty}\|g\|_{L^2}. \tag{C.33}$$

We treat the singular integral $\mathcal{S}^\varepsilon$ as follows. By antisymmetry of the function $(\eta-k)^{-1}\ \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}(\eta)$ we have

$$\int \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}\ \frac{1}{\eta-k}\ \mathcal{F}[\chi g](\eta)\ \mathrm{d}\eta = \ \int \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}\ \frac{\mathcal{F}[\chi g](\eta)-\mathcal{F}[\chi g](k)}{\eta-k}\ \ \mathrm{d}\eta.$$

Moreover, we have

$$\left| \ \frac{\mathcal{F}[\chi g](\eta)-\mathcal{F}[\chi g](k)}{\eta-k}\ \right| \ \leq \ \left| \ \partial_\eta \ \mathcal{F}[\chi g](\eta) \ |_{\eta=\tilde\eta\in\{\varepsilon\leq|\eta-k|\leq\delta_0/4\}} \ \right|. \tag{C.34}$$

By the uniform boundedness of $\langle\zeta\rangle^{-2}\partial_\eta m_+(\zeta,\eta)$ and $\langle\zeta\rangle^{-2}\partial_\eta m_-(\zeta,\eta)$ in $\mathbb{R}\times\mathbb{R}$, we have that

$$\begin{aligned} |\partial_\eta\ (\ \mathcal{F}[\chi g](\eta)\ )| &= \left| \int_\zeta \partial_\eta\overline{\Psi(\zeta;\eta)}\chi(\zeta)g(\zeta)\ \mathrm{d}\zeta \right| \\ &\leq \ \sup_{(\zeta;\eta)\in\mathbb{R}\times\mathbb{R}} |\langle\zeta\rangle^{-2}\partial_\eta\ \Psi(\zeta;\eta)|\|\langle\zeta\rangle^2\chi\|_{L^2_\zeta}\|g\|_{L^2} \ \leq \ C\ \|\langle\zeta\rangle^2\chi\|_{L^2_\zeta}\|g\|_{L^2}. \end{aligned}$$

Therefore,

$$\left| \ \int \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}(\eta)\ \frac{1}{\eta-k}\ \mathcal{F}[\chi g](\eta)\ \mathrm{d}\eta\ \right| \ \leq \ C\ \|\langle\zeta\rangle^2\chi\|_{L^2_\zeta}\|g\|_{L^2},$$

from which it follows that

$$| \ \mathcal{S}^\varepsilon(\xi) \ | \ \leq \ C\ \|\langle\zeta\rangle^2\chi\|_{L^2_\zeta}\|g\|_{L^2} \ \ |\Lambda^\varepsilon(\xi,k)\ | \ \leq \ C_\chi\frac{1}{1+|\xi-k|}\ \|g\|_{L^2}. \tag{C.35}$$

Thus we have from (C.31), (C.33) and (C.35),

$$\| T^{II}_{near}[g] \|_2 \leq C'_\chi \|g\|_2.$$

Using the estimates of $T^{II}_{far}[g]$ and $T^{II}_{near}[g]$ yields (C.23). Therefore, $A^{(a)}_{II}$ is bounded from L^2 to L^2. This completes the proof of Proposition C.5. □

Proposition C.5. *$A^{(b)}_{II} \equiv \chi \langle D_{V_0}\rangle R_{V_0}(k))([\chi, \langle D_{V_0}\rangle])$ is bounded from L^2 to L^2.*

This follows from Proposition C.4 and duality.

Finally, we consider the operator $A_{III} \equiv [\langle D_{V_0}\rangle, \chi] \circ R_{V_0}(k) \circ [\chi, \langle D_{V_0}\rangle]$.

Proposition C.6. *The operator A_{III} is bounded from L^2 to L^2.*

Proof. By (C.19), one has

$$\begin{aligned} A_{III}[g](x) = & \int_\xi \Psi(x;\xi) \int_\eta (\langle\xi\rangle - \langle\eta\rangle) \int_\zeta \overline{\Psi(\zeta;\xi)} \Psi(\zeta;\eta)\chi(\zeta)\, \mathrm{d}\zeta \\ & \frac{1}{\eta^2 - k^2} \mathcal{F}\left[\, [\langle D_{V_0}\rangle, \chi] g \,\right](\eta)\, \mathrm{d}\eta\, \mathrm{d}\xi \\ = & \int_\xi \Psi(x;\xi) \int_\eta K(\xi,\eta) \frac{1}{\eta^2 - k^2} \mathcal{F}\left[\, [\langle D_{V_0}\rangle, \chi] g \,\right](\eta)\, \mathrm{d}\eta\, \mathrm{d}\xi \\ = & \int_\xi \Psi(x;\xi) \int_\eta K(\xi,\eta) \frac{1}{\eta^2 - k^2} \int_\theta K(\eta,\theta) \mathcal{F}[g](\theta)\, d\theta\, \mathrm{d}\eta\, \mathrm{d}\xi \end{aligned} \tag{C.36}$$

By the Plancherel Theorem, the L^2 estimate of A_{III} is equivalent to the bound

$$\| T^{III}[g] \|_{L^2} = \left\| \int_\eta K(\xi,\eta) \frac{1}{\eta^2 - k^2} \int_\theta K(\eta,\theta) \mathcal{F}[g](\theta)\, d\theta\, \mathrm{d}\eta \right\|_{L^2_\xi} \leq C \|g\|_{L^2}. \tag{C.37}$$

We now proceed with a proof of (C.37). We use $\varphi_{|\kappa|<\delta_0}$, defined as in (C.24), to localize at frequencies near $\eta = \pm k$ and frequencies away from $\eta = \pm k$.

$$T^{III}[g] \equiv T^{III}_{near}[g] + T^{III}_{far}[g] \tag{C.38}$$

where

$$T^{III}_{far}[g](\xi) \equiv \int_\eta K(\xi,\eta) \left(1 - \varphi_{||\eta|-|k||<\delta_0}(\eta)\right) \frac{1}{\eta^2 - k^2} \int_\theta K(\eta,\theta) \mathcal{F}[g](\theta)\, d\theta\, \mathrm{d}\eta \tag{C.39}$$

$$T^{III}_{near}[g](\xi) \equiv \int_\eta K(\xi,\eta)\ \varphi_{||\eta|-|k||<\delta_0}(\eta)\ \frac{1}{\eta^2 - k^2} \int_\theta K(\eta,\theta) \mathcal{F}[g](\theta)\, d\theta\, \mathrm{d}\eta \tag{C.40}$$

Bound on $T^{III}_{far}[g](\xi)$: We recall Lemma C.3, stating that $K(\xi,\eta)$ satisfies the following upper bound:

$$| K(\xi,\eta) | \leq C_\chi \frac{1}{1 + |\xi - \eta|},$$

with $C_\chi \equiv C \left(\sum_{j=0}^2 \left\| \langle\zeta\rangle^j\, \partial^j_\zeta \chi \right\|_{L^1_\zeta} + \left\| \langle\zeta\rangle^2\, \chi \right\|_{L^1_\zeta} + \left\| \langle\zeta\rangle^2\, \chi \right\|_{L^\infty} \right)$. Therefore, one has the pointwise estimate

$$\left| \int_\theta K(\eta,\theta) \mathcal{F}[g](\theta)\, d\theta \right| \leq \left\| \frac{1}{1 + |\eta - \theta|} \right\|_{L^2_\theta} \|\mathcal{F}[g]\|_{L^2} \leq \left\| \frac{1}{1 + |\cdot|} \right\|_{L^2_\theta} \|g\|_{L^2}.$$

Moreover, for $||\eta|-|k||>\delta_0$, one has $\big(1-\varphi_{||\eta|-|k||<\delta_0}(\eta)\big)\ \ |\eta^2-k^2|^{-1}\ \in\ L^1$. Therefore, by Young's inequality,

$$\big\|T^{III}_{far}[g]\big\|_{L^2}\leq C\left\|\frac{1}{1+|\cdot|}\right\|^2_{L^2}\left\|\big(1-\varphi_{||\eta|-|k||<\delta_0}(\eta)\big)\,\frac{1}{|\eta^2-k^2|}\right\|_{L^1}\|g\|_{L^2}\leq C_\chi\,\|g\|_{L^2}\,.$$

Bound on $T^{III}_{near}[g](\xi)$:

$$\begin{aligned}T^{III,\varepsilon}_{near}[g](\xi)\ &\equiv\ \int_\eta K(\xi,\eta)\ \ \varphi_{||\eta|-|k||<\delta_0}(\eta)\ \ \frac{1}{\eta^2-k^2}\ \int_\theta K(\eta,\theta)\mathcal{F}[g](\theta)\ d\theta\ \mathrm{d}\eta\\ &\equiv\int\Lambda^\varepsilon(\xi,\eta)\ \frac{1}{\eta-k}\int_\theta K(\eta,\theta)\mathcal{F}[g](\theta)\ d\theta\mathrm{d}\eta\\ &\qquad+\ \Lambda^\varepsilon(\xi,\eta)\ \frac{1}{\eta+k}\int_\theta K(\eta,\theta)\mathcal{F}[g](\theta)\ d\theta\mathrm{d}\eta\ ,\end{aligned}$$

with $\Lambda^\varepsilon(\xi,\eta)\ \equiv\ \frac{1}{2k}\ K(\xi,\eta)\ \varphi_{\varepsilon\leq||\eta|-|k||<\delta_0}(\eta)$.

Note that by Lemma C.3,

$$|\ \Lambda^\varepsilon(\xi,\eta)\ |\ \leq\ C_\chi\frac{1}{1+|\xi-\eta|}\varphi_{\varepsilon\leq||\eta|-|k||<\delta_0}(\eta). \tag{C.41}$$

We bound the first term in the above expansion of $T^{III,\varepsilon}_{near}$. The second term is treated similarly. We have

$$\begin{aligned}\int\Lambda^\varepsilon(\xi,\eta)\ \frac{\mathcal{F}[\chi g](\eta)}{\eta-k}\mathrm{d}\eta\ &=\mathcal{S}^\varepsilon(\xi)\ +\ \mathcal{E}^\varepsilon(\xi)+\ \mathcal{R}^\varepsilon(\xi),\quad\text{where}\\ \mathcal{S}^\varepsilon(\xi)\equiv&\Lambda^\varepsilon(\xi,k)\ \int\ \frac{1}{\eta-k}\int_\theta K(\eta,\theta)\mathcal{F}[g](\theta)\ d\theta\ \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}\ \mathrm{d}\eta,\\ \mathcal{E}^\varepsilon(\xi)\equiv&\int\frac{\Lambda^\varepsilon(\xi,\eta)-\Lambda^\varepsilon(\xi,k)}{\eta-k}\int_\theta K(\eta,\theta)\mathcal{F}[g](\theta)\ d\theta\ \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}\ \mathrm{d}\eta,\\ \mathcal{R}^\varepsilon(\xi)\equiv&\int\Lambda^\varepsilon(\xi,\eta)\ \frac{1}{\eta-k}\int_\theta K(\eta,\theta)\mathcal{F}[g](\theta)\ d\theta\ \mathbf{1}_{|\eta-k|\geq\delta_0/4}\ \mathrm{d}\eta.\end{aligned}$$

As in the proof of Proposition C.4, the kernel of the integral operators defining $\mathcal{E}^\varepsilon$ and $\mathcal{R}^\varepsilon$ are non-singular, and we have uniformly in ϵ:

$$\|\mathcal{E}^\varepsilon\|_{L^2}+\|\mathcal{R}^\varepsilon\|_{L^2}\ \leq\ C'_\chi\|g\|_{L^2}.$$

We treat the singular integral $\mathcal{S}^\varepsilon$ as follows. By antisymmetry of the function $(\eta-k)^{-1}\ \mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}(\eta)$ we have

$$\begin{aligned}\mathcal{S}(\xi)=\ &\Lambda(\xi,k)\int\mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}(\eta)\ \int_\theta\frac{K(\eta,\theta)}{\eta-k}\ \mathcal{F}[g](\theta)\ d\theta\ \mathrm{d}\eta\\ =\ &\int\mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}(\eta)\ \int_\theta\frac{K(\eta,\theta)-K(k,\theta)}{\eta-k}\ \mathcal{F}[g](\theta)\ d\theta\ \mathrm{d}\eta.\end{aligned}$$

Moreover, Lemma C.3 leads to

$$\left|\ \frac{K(\eta,\theta)-K(k,\theta)}{\eta-k}\ \right|\ \leq\ \left|\ \partial_\eta\ K(\eta,\theta)\ |_{\eta=\tilde\eta\in\{\varepsilon\leq|\eta-k|\leq\delta_0/4\}}\ \right|\ \leq\ C'_\chi\frac{1}{1+|\eta-\theta|}. \tag{C.42}$$

Therefore, by Cauchy-Schwarz inequality,

$$\left|\int_\theta\frac{K(\eta,\theta)-K(k,\theta)}{\eta-k}\ \mathcal{F}[g](\theta)\ d\theta\ \right|\leq\ C'_\chi\|g\|_{L^2},$$

from which it follows that

$$|\ \mathcal{S}^\varepsilon(\xi)\ |\ \leq\ |\Lambda^\varepsilon(\xi,k)|\ \|\mathbf{1}_{\varepsilon\leq|\eta-k|\leq\delta_0/4}\|_{L^1_\eta}\ \|g\|_{L^2}\ \leq\ C'_\chi\frac{1}{1+|\xi-k|}\ \|g\|_{L^2}. \tag{C.43}$$

Thus we have $\left\| T^{III}_{near}[g] \right\|_2 \leq C'_\chi \|g\|_2$. Using the estimates of $T^{III}_{far}[g]$ and $T^{III}_{near}[g]$ yields (C.37). Therefore, A_{III} is bounded from L^2 to L^2. This completes the proof of Proposition C.6, and hence the proof of Proposition 5.5. □

Références

Partie A : Ondes internes de gravité en océanographie

[1] Alvarez-Samaniego, B., and Lannes, D. Large time existence for 3D water-waves and asymptotics. *Invent. Math. 171*, 3 (2008), 485–541.

[2] Alvarez-Samaniego, B., and Lannes, D. A Nash-Moser theorem for singular evolution equations. Application to the Serre and Green-Naghdi equations. *Indiana Univ. Math. J. 57*, 1 (2008), 97–131.

[3] Ambrose, D. M. Well-posedness of vortex sheets with surface tension. *SIAM J. Math. Anal. 35*, 1 (2003), 211–244 (electronic).

[4] Ambrose, D. M., and Masmoudi, N. Well-posedness of 3D vortex sheets with surface tension. *Commun. Math. Sci. 5*, 2 (2007), 391–430.

[5] Apel, J., Ostrovsky, L., Stepanyants, Y., and Lynch, J. Internal solitons in the ocean and their effect on underwater sound. *The Journal of the Acoustical Society of America 121*, 2 (2007), 695–722.

[6] Attouch, H., Buttazzo, G., and Michaille, G. *Variational analysis in Sobolev and BV spaces*, vol. 6 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006.

[7] Barros, R., and Gavrilyuk, S. Dispersive nonlinear waves in two-layer flows with free surface. II. Large amplitude solitary waves embedded into the continuous spectrum. *Stud. Appl. Math. 119*, 3 (2007), 213–251.

[8] Barros, R., Gavrilyuk, S. L., and Teshukov, V. M. Dispersive nonlinear waves in two-layer flows with free surface. I. Model derivation and general properties. *Stud. Appl. Math. 119*, 3 (2007), 191–211.

[9] Benjamin, T. B. Internal waves of permanent form in fluids of great depth. *J. Fluid Mech. 29*, 3 (1967), 559–592.

[10] Benjamin, T. B., Bona, J. L., and Mahony, J. J. Model equations for long waves in nonlinear dispersive systems. *Philos. Trans. Roy. Soc. London Ser. A 272*, 1220 (1972), 47–78.

[11] Benjamin, T. B., and Bridges, T. J. Reappraisal of the Kelvin-Helmholtz problem. I. Hamiltonian structure. *J. Fluid Mech. 333* (1997), 301–325.

[12] Besse, C. Schéma de relaxation pour l'équation de Schrödinger non linéaire et les systèmes de Davey et Stewartson. *C.R. Acad. Sci. Paris Sér. I*, 12 (1998), 1427–1432.

[13] Besse, C., and Bruneau, C. H. Numerical study of elliptic-hyperbolic Davey-Stewartson system : dromions simulation and blow-up. *Math. Models Methods Appl. Sci. 8*, 8 (1998), 1363–1386.

[14] Bona, J. L., Chen, M., and Saut, J.-C. Boussinesq equations and other systems for small-amplitude long waves in nonlinear dispersive media. I. Derivation and linear theory. *J. Nonlinear Sci. 12*, 4 (2002), 283–318.

[15] Bona, J. L., Chen, M., and Saut, J.-C. Boussinesq equations and other systems for small-amplitude long waves in nonlinear dispersive media. II. The nonlinear theory. *Nonlinearity 17*, 3 (2004), 925–952.

[16] Bona, J. L., Colin, T., and Lannes, D. Long wave approximations for water waves. *Arch. Ration. Mech. Anal. 178*, 3 (2005), 373–410.

[17] Bona, J. L., Lannes, D., and Saut, J.-C. Asymptotic models for internal waves. *J. Math. Pures Appl. (9) 89*, 6 (2008), 538–566.

[18] Bona, J. L., and Zhang, B.-Y. The initial-value problem for the forced Korteweg-de Vries equation. *Proc. Roy. Soc. Edinburgh Sect. A 126*, 3 (1996), 571–598.

[19] Boussinesq, J. Théorie de l'intumescence liquide appelée onde solitaire ou de translation se propageant dans un canal rectangulaire. *C.R. Acad. Sci. Paris Sér. A-B 72* (1871), 755–759.

[20] Boussinesq, J. Théorie des ondes et des remous qui se propagent le long d'un canal rectangulaire horizontal, en communiquant au liquide contenu dans ce canal des vitesses sensiblement pareilles de la surface au fond. *J. Math. Pures Appl. 17*, 2 (1872), 55–108.

[21] Bresch, D., and Renardy, M. Well-posedness of two-layer shallow water flow between two horizontal rigid plates. To appear in *Nonlinearity*.

[22] Camassa, R., Choi, W., Michallet, H., Rusas, P.-O., and Sveen, J. K. On the realm of validity of strongly nonlinear asymptotic approximations for internal waves. *J. Fluid Mech. 549* (2006), 1–23.

[23] Camassa, R., and Holm, D. D. An integrable shallow water equation with peaked solitons. *Phys. Rev. Lett. 71*, 11 (1993), 1661–1664.

[24] Camassa, R., and Wu, T. Y.-t. Stability of forced steady solitary waves. *Philos. Trans. Roy. Soc. London Ser. A 337*, 1648 (1991), 429–466.

[25] Camassa, R., and Wu, T. Y.-t. Stability of some stationary solutions for the forced KdV equation. *Phys. D 51*, 1-3 (1991), 295–307.

[26] Chazel, F. Influence of bottom topography on long water waves. *M2AN Math. Model. Numer. Anal. 41*, 4 (2007), 771–799.

[27] Chazel, F. On the Korteweg-de Vries approximation for uneven bottoms. *Eur. J. Mech. B Fluids 28*, 2 (2009), 234–252.

[28] Cheng, C.-H. A., Coutand, D., and Shkoller, S. On the motion of vortex sheets with surface tension in three-dimensional Euler equations with vorticity. *Comm. Pure Appl. Math. 61*, 12 (2008), 1715–1752.

[29] Choi, W., Barros, R., and Jo, T.-C. A regularized model for strongly nonlinear internal solitary waves. *J. Fluid Mech. 629* (2009), 73–85.

[30] Choi, W., and Camassa, R. Weakly nonlinear internal waves in a two-fluid system. *J. Fluid Mech. 313* (1996), 83–103.

[31] Choi, W., and Camassa, R. Fully nonlinear internal waves in a two-fluid system. *J. Fluid Mech. 396* (1999), 1–36.

[32] Colliander, J., Keel, M., Staffilani, G., Takaoka, H., and Tao, T. Sharp global well-posedness for KdV and modified KdV on $\mathbb{R}$ and $\mathbb{T}$. *J. Amer. Math. Soc. 16*, 3 (2003), 705–749 (electronic).

[33] Craig, W. An existence theory for water waves and the Boussinesq and Korteweg-de Vries scaling limits. *Comm. Partial Differential Equations 10*, 8 (1985), 787–1003.

[34] Craig, W., and Groves, M. D. Normal forms for wave motion in fluid interfaces. *Wave Motion 31*, 1 (2000), 21–41.

[35] Craig, W., Guyenne, P., and Kalisch, H. Hamiltonian long-wave expansions for free surfaces and interfaces. *Comm. Pure Appl. Math. 58*, 12 (2005), 1587–1641.

[36] Craig, W., Sulem, C., and Sulem, P.-L. Nonlinear modulation of gravity waves : a rigorous approach. *Nonlinearity 5*, 2 (1992), 497–522.

[37] de Saint-Venant, B. Théorie du mouvement non-permanent des eaux, avec application aux crues des rivières et à l'introduction des marées dans leur lit. *C.R. Acad. Sci. Paris 73* (1871), 147–154.

[38] Dingemans, M. W. *Water Wave Propagation Over Uneven Bottoms : Non-linear wave propagation*, vol. 13 of *Advanced Series on Ocean Engineering*. World Scientific, Cornell Univ., Hollister Hall, 1997.

[39] Djordjevic, V. D., and Redekopp, L. G. The fission and disintegration of internal solitary waves moving over two-dimensional topography. *J. Phys. Oceanogr. 8*, 6 (1978), 1016–1024.

[40] Duchêne, V. Asymptotic shallow water models for internal waves in a two-fluid system with a free surface. *SIAM J. Math. Anal. 42*, 5 (2010), 2229–2260.

[41] Duchêne, V. Boussinesq/Boussinesq systems for internal waves with a free surface, and the KdV approximation. To appear in Math. Model. Numer. Anal. (M2AN). Arxiv preprint :1007.3116.

[42] Durufle, M., and Israwi, S. A numerical study of variable depth KdV equations and generalizations of Camassa-Holm-like equations. Preprint, available at `http://hal.archives-ouvertes.fr/hal-00454495/en/`.

[43] EKMAN, V. W. On dead water. *Sci. Results Norw. North Polar Expedi. 1893-96 5*, 15 (1904), 1–152.

[44] FUNAKOSHI, M., AND OIKAWA, M. The resonant interaction between a long internal gravity wave and a surface gravity wave packet. *J. Phys. Soc. Japan 52*, 6 (1983), 1982–1995.

[45] FUNAKOSHI, M., AND OIKAWA, M. Long internal waves of large amplitude in a two-layer fluid. *J. Phys. Soc. Japan 55*, 1 (1986), 128–144.

[46] GARDNER, C. S., GREENE, J. M., KRUSKAL, M. D., AND MIURA, R. M. Method for solving the Korteweg-deVries equation. *Phys. Rev. Lett. 19*, 19 (1967), 1095–1097.

[47] GILL, A. E. *Atmosphere-ocean dynamics*, vol. 30 of *International geophysics series*. Academic Press, 1982.

[48] GREEN, A. E., AND NAGHDI, P. M. A derivation of equations for wave propagation in water of variable depth. *J. Fluid Mech. 78*, 02 (1976), 237–246.

[49] GRIMSHAW, R., PELINOVSKY, E., AND TALIPOVA, T. The modified korteweg–de vries equation in the theory of large-amplitude internal waves. *Nonlinear Processes Geophys. 4*, 4 (1997), 237–250.

[50] GUYENNE, P. Large-amplitude internal solitary waves in a two-fluid model. *C.R. Mec. 334*, 6 (2006), 341–346.

[51] GUYENNE, P., LANNES, D., AND SAUT, J.-C. Well-posedness of the Cauchy problem for models of large amplitude internal waves. *Nonlinearity 23*, 2 (2010), 237–275.

[52] HELFRICH, K. R., AND MELVILLE, W. K. Long nonlinear internal waves. vol. 38 of *Annu. Rev. Fluid Mech.* Palo Alto, CA, 2006, pp. 395–425.

[53] HEREMAN, W. A. *Encyclopedia of Complexity and Systems Science*. Springer Verlag, 2009, ch. Shallow Water Waves and Solitary Waves, pp. 8112–8125.

[54] ISRAWI, S. Variable depth KdV equations and generalizations to more nonlinear regimes. *M2AN Math. Model. Numer. Anal. 44*, 2 (2010), 347–370.

[55] KADOMTSEV, B. B., AND PETVIASHVILI, V. I. On the stability of solitary waves in weakly dispersing media. *Sov. Phys. Dokl. 15* (1970), 539–541.

[56] KAKUTANI, T., AND YAMASAKI, N. Solitary waves on a two-layer fluid. *J. Phys. Soc. Japan 45*, 2 (1978), 674–679.

[57] KANO, T., AND NISHIDA, T. A mathematical justification for Korteweg-de Vries equation and Boussinesq equation of water surface waves. *Osaka J. Math. 23*, 2 (1986), 389–413.

[58] KATO, T., AND PONCE, G. Commutator estimates and the Euler and Navier-Stokes equations. *Comm. Pure Appl. Math. 41*, 7 (1988), 891–907.

[59] Kenig, C. E., Ponce, G., and Vega, L. Well-posedness and scattering results for the generalized Korteweg-de Vries equation via the contraction principle. *Comm. Pure Appl. Math. 46*, 4 (1993), 527–620.

[60] Keulegan, G. H. Characteristics of internal solitary waves. *J. Res. Nat. Bur. Stand 51* (1953), 133–140.

[61] Koop, C. G., and Butler, G. An investigation of internal solitary waves in a two-fluid system. *J. Fluid Mech. 112* (1981), 225–251.

[62] Korteweg, D. J., and De Vries, G. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. *Philos. Mag. 5*, 39 (1895), 422–443.

[63] Kostyukov, A. *Theory of ship waves and wave resistance*. Leningrad : Sudostroyeniye, 1959. In Russian. English translation : Iowa City, Effective Communications Inc. (1968).

[64] Kruskal, M. D., Miura, R. M., Gardner, C. S., and Zabusky, N. J. Korteweg-de Vries equation and generalizations. V. Uniqueness and nonexistence of polynomial conservation laws. *J. Mathematical Phys. 11* (1970), 952–960.

[65] Kubota, T., Ko, D., and Dobbs, L. Weakly-nonlinear, long internal gravity waves in stratified fluids of finite depth. *J. Hydronautics 12*, 4 (1978), 157–165.

[66] Lamb, H. On waves due to a travelling disturbance, with an application to waves in superposed fluids. *Philos. Mag. 6*, 31 (1916), 386–398.

[67] Lannes, D. Secular growth estimates for hyperbolic systems. *J. Differential Equations 190*, 2 (2003), 466–503.

[68] Lannes, D. Well-posedness of the water-waves equations. *J. Amer. Math. Soc. 18*, 3 (2005), 605–654 (electronic).

[69] Lannes, D. Sharp estimates for pseudo-differential operators with symbols of limited smoothness and commutators. *J. Funct. Anal. 232*, 2 (2006), 495–539.

[70] Lannes, D. A stability criterion for two-fluid interfaces and applications. Arxiv preprint :1005.4565, 2010.

[71] Lannes, D., and Saut, J.-C. Weakly transverse Boussinesq systems and the Kadomtsev-Petviashvili approximation. *Nonlinearity 19*, 12 (2006), 2853–2875.

[72] Lax, P. D. Integrals of nonlinear equations of evolution and solitary waves. *Comm. Pure Appl. Math. 21*, 5 (1968), 497–490.

[73] Lebeau, G. Régularité du problème de Kelvin-Helmholtz pour l'équation d'Euler 2d. *ESAIM Control Optim. Calc. Var. 8* (2002), 801–825 (electronic). A tribute to J. L. Lions.

[74] LEE, S.-J. *Generation of long water waves by moving disturbances*. PhD thesis, California Institute of Technology, 1985.

[75] LEE, S.-J., YATES, G. T., AND WU, T. Y. Experiments and analyses of upstream-advancing solitary waves generated by moving disturbances. *J. Fluid Mech. 199*, -1 (1989), 569–593.

[76] LEONE, C., SEGUR, H., AND HAMMACK, J. L. Viscous decay of long internal solitary waves. *Phys. Fluids 25*, 6 (1982), 942–944.

[77] LONG, R. R. Long waves in a two-fluid system. *J. Meteorol. 13* (1956), 70–74.

[78] LONG, R. R. On the Boussinesq approximation and its role in the theory of internal waves. *Tellus 17*, 1 (1965), 46–52.

[79] LU, D.-Q., AND CHEN, T.-T. Surface and interfacial gravity waves induced by an impulsive disturbance in a two-layer inviscid fluid. *Journal of Hydrodynamics, Series B 21*, 1 (2009), 26–33.

[80] MALEEWONG, M., GRIMSHAW, R., AND ASAVANANT, J. Free surface flow under gravity and surface tension due to an applied pressure distribution. II. Bond number less than one-third. *Eur. J. Mech. B Fluids 24*, 4 (2005), 502–521.

[81] MAL′TSEVA, Z. L. Unsteady long waves in a two-layer fluid. *Dinamika Sploshn. Sredy*, 93-94 (1989), 96–110.

[82] MATSUNO, Y. A unified theory of nonlinear wave propagation in two-layer fluid systems. *J. Phys. Soc. Japan 62*, 6 (1993), 1902–1916.

[83] MÉTIVIER, G. *Small viscosity and boundary layer methods. Theory, stability analysis, and applications*. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser Boston Inc., Boston, MA, 2004.

[84] MÉTIVIER, G. *Para-differential calculus and applications to the Cauchy problem for nonlinear systems*, vol. 5 of *Centro di Ricerca Matematica Ennio De Giorgi (CRM) Series*. Edizioni della Normale, Pisa, 2008.

[85] MICHALLET, H., AND BARTHÉLEMY, E. Ultrasonic probes and data processing to study interfacial solitary waves. *Exp. Fluids 22*, 5 (1997), 380–386.

[86] MICHALLET, H., AND BARTHÉLEMY, E. Experimental study of interfacial solitary waves. *J. Fluid Mech. 366* (1998), 159–177.

[87] MILLER, R. L. An experimental study of form and transitions in impulse waves. In *Proc. Syrup. Long Waves* (1972), University of Delaware, pp. 33–61.

[88] MILOH, T., TULIN, M., AND ZILMAN, G. Dead-water effects of a ship moving in stratified seas. *J. Offshore Mech. Arct. Eng. 115*, 2 (1993), 105–110.

[89] MIYATA, M. An internal solitary wave of large amplitude. *La mer 23*, 2 (1985), 43–48.

[90] Miyata, M. Long internal waves of large amplitude. In *Nonlinear Water Waves : IUTAM Symposium* (Tokyo, Aug 1987), Springer, pp. 399–405.

[91] Motygin, O. V., and Kuznetsov, N. G. The wave resistance of a two-dimensional body moving forward in a two-layer fluid. *J. Engrg. Math. 32*, 1 (1997), 53–72.

[92] Nansen, F. *The Norwegian North polar expedition 1893-1896*. Nansen Fund, 1900.

[93] Nguyen, H. Y., and Dias, F. A Boussinesq system for two-way propagation of interfacial waves. *Phys. D 237*, 18 (2008), 2365–2389.

[94] Nguyen, T., and Yeung, R. Steady-wave sytems in a two-layer fluid of finite depth. In *Proc. 12th Int. Workshop Water Waves and Floating Bodies* (March 1997), pp. 115–119.

[95] Ono, H. Algebraic solitary waves in stratified fluids. *J. Phys. Soc. Japan 39*, 4 (1975), 1082–1091.

[96] Ostrovsky, L. A., and Grue, J. Evolution equations for strongly nonlinear internal waves. *Phys. Fluids 15*, 10 (2003), 2934–2948.

[97] Ostrovsky, L. A., and Stepanyants, Y. A. Do internal solitons exist in the ocean ? *Rev. Geophys. 27*, 3 (1989), 293–310.

[98] Ostrovsky, L. A., and Stepanyants, Y. A. Internal solitons in laboratory experiments : comparison with theoretical models. *Chaos 15*, 3 (2005), 037111, 1–28.

[99] Peters, A. D., and Stoker, J. J. Solitary waves in liquids having non-constant density. *Comm. Pure Appl. Math. 13* (1960), 115–164.

[100] Rayleigh, J. W. S. On waves. *Philos. Mag. 1*, 5 (1876), 251–271.

[101] Russel, J. S. Report on waves. In *Report of the fourteenth meeting of the British Association for the Advancement of Science* (1845), pp. 311–390.

[102] Sakai, T., and Redekopp, L. G. Models for strongly-nonlinear evolution of long internal waves in a two-layer stratification. *Nonlinear Processes in Geophysics 14*, 1 (2007), 31–47.

[103] Schneider, G. Approximation of the Korteweg-de Vries equation by the nonlinear Schrödinger equation. *J. Differential Equations 147*, 2 (1998), 333–354.

[104] Schneider, G., and Wayne, C. E. The long-wave limit for the water wave problem. I. The case of zero surface tension. *Comm. Pure Appl. Math. 53*, 12 (2000), 1475–1535.

[105] Segur, H., and Hammack, J. L. Soliton models of long internal waves. *J. Fluid Mech. 118* (1982), 285–304.

[106] Shatah, J., and Zeng, C. A priori estimates for fluid interface problems. *Comm. Pure Appl. Math. 61*, 6 (2008), 848–876.

[107] SHEN, S. S. P. Forced solitary waves and hydraulic falls in two-layer flows. *J. Fluid Mech. 234* (1992), 583–612.

[108] SRETENSKII, L. On internal waves at the interface of two fluids with application to a dead water phenomenon. *Zh. Geofiz. 4* (1934), 332–370.

[109] SULEM, C., AND SULEM, P.-L. Finite time analyticity for the two- and three-dimensional Rayleigh-Taylor instability. *Trans. Amer. Math. Soc. 287*, 1 (1985), 127–160.

[110] SULEM, C., SULEM, P.-L., BARDOS, C., AND FRISCH, U. Finite time analyticity for the two- and three-dimensional Kelvin-Helmholtz instability. *Comm. Math. Phys. 80*, 4 (1981), 485–516.

[111] TAYLOR, M. E. *Partial differential equations. III Nonlinear equations*, vol. 117 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1997.

[112] TEN, I., AND KASHIWAGI, M. Hydrodynamics of a body floating in a two-layer fluid of finite depth. Part 1. Radiation problem. *J. Mar. Sci. Technol. 9*, 3 (2004), 127–141.

[113] VASSEUR, R. Ondes d'interface dans les fluides stratifiés. Master's thesis, Ecole Normale Supérieure de Lyon, 2008.

[114] VASSEUR, R., MERCIER, M., AND DAUXOIS, T. Resurrecting dead-water phenomenon. *Nonlinear Processes Geophys. 18* (2011), 193–208.

[115] WALKER, L. R. Interfacial solitary waves in a two-fluid medium. *Phys. Fluids 16*, 11 (1973), 1796–1804.

[116] WU, D., AND WU, T. Three-dimensional nonlinear long waves due to moving surface pressure. In *Proc. 14th Symp. Naval Hydrodyn.* (Ann Arbor, Michigan, 1983), National Academy Press, pp. 103–129.

[117] WU, J., AND CHEN, B. Unsteady ship waves in shallow water of varying depth based on Green-Naghdi equation. *Ocean Engineering 30*, 15 (2003), 1899–1913.

[118] WU, T. Y. Long waves in ocean and coastal waters. *J. Eng. Mech. 107*, 3 (1981), 501–522.

[119] WU, T. Y. Generation of upstream advancing solitons by moving disturbances. *J. Fluid Mech. 184* (1987), 75–99.

[120] YEUNG, R. W., AND NGUYEN, T. C. Waves generated by a moving source in a two-layer ocean of finite depth. *J. Engrg. Math. 35*, 1-2 (1999), 85–107.

[121] ZABUSKY, N. J., AND KRUSKAL, M. D. Interaction of "solitons" in a collisionless plasma and the recurrence of initial states. *Phys. Rev. Lett. 15*, 6 (1965), 240–243.

[122] ZAKHAROV, V. E. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *J. Appl. Mech. Tech. Phys. 9* (1968), 190–194.

Partie B : Propagation des ondes à travers un milieu non homogène

[123] Development and applications of materials exhibiting photonic band gaps. *J. Opt. Soc. Am. B : Opt. Phys. 10*, 2 (Feb 1993), 280–280. Special issue.

[124] Abramowitz, M., and Stegun, I. A. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55 of *National Bureau of Standards Applied Mathematics Series*. U.S. Government Printing Office, Washington, D.C., 1964.

[125] Agmon, S. Spectral properties of Schrödinger operators and scattering theory. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 2*, 2 (1975), 151–218.

[126] Allaire, G. Homogenization and two-scale convergence. *SIAM J. Math. Anal. 23*, 6 (1992), 1482–1518.

[127] Allaire, G. Periodic homogenization and effective mass theorems for the Schrödinger equation. In *Quantum transport*, vol. 1946 of *Lecture Notes in Math.* Springer, Berlin, 2008, pp. 1–44.

[128] Allaire, G., and Amar, M. Boundary layer tails in periodic homogenization. *ESAIM Control Optim. Calc. Var. 4* (1999), 209–243 (electronic).

[129] Bensoussan, A., Lions, J.-L., and Papanicolaou, G. *Asymptotic analysis for periodic structures*, vol. 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1978.

[130] Berezin, F. A., and Shubin, M. A. *The Schrödinger equation*, vol. 66 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1991. Translated from the 1983 Russian edition.

[131] Choi, J., Lin, T., Sun, S. M., and Whang, S. Supercritical surface waves generated by negative or oscillatory forcing. *Discrete Contin. Dyn. Syst. Ser. B 14*, 4 (2010), 1313–1335.

[132] Choi, J. W., Sun, S. M., and Whang, S. Supercritical surface gravity waves generated by a positive forcing. *Eur. J. Mech. B Fluids 27*, 6 (2008), 750–770.

[133] Collin, R. E. *Field theory of guided waves*. International Series in Pure and Applied Physics. Mc-Graw-Hill Book Co., Inc., New York, 1960.

[134] D'Ancona, P., and Fanelli, L. L^p-boundedness of the wave operator for the one dimensional Schrödinger operator. *Comm. Math. Phys. 268*, 2 (2006), 415–438.

[135] Deift, P., and Trubowitz, E. Inverse scattering on the line. *Comm. Pure Appl. Math. 32*, 2 (1979), 121–251.

[136] Duchêne, V., Marzuola, J. L., and Weinstein, M. I. Wave operator bounds for 1-dimensional Schrödinger operators with singular potentials and applications. *J. Math. Phys. 52* (2011), 013505 :1–17.

[137] DUCHÊNE, V., AND WEINSTEIN, M. I. Scattering, homogenization and interface effects for oscillatory potentials with strong singularities. To appear in (SIAM) Multiscale Model. Simul. Arxiv preprint :1010.2694.

[138] FOLLAND, G. B. *Introduction to partial differential equations*, second ed. Princeton University Press, Princeton, NJ, 1995.

[139] FUKUIZUMI, R., AND JEANJEAN, L. Stability of standing waves for a nonlinear Schrödinger equation with a repulsive Dirac delta potential. *Discrete Contin. Dyn. Syst. 21*, 1 (2008), 121–136.

[140] FUKUIZUMI, R., OHTA, M., AND OZAWA, T. Nonlinear Schrödinger equation with a point defect. *Ann. Inst. H. Poincaré Anal. Non Linéaire 25*, 5 (2008), 837–845.

[141] GÉRARD-VARET, D., AND MASMOUDI, N. Homogenization in polygonal domains. To appear in J. Europ. Math. Soc., 2008.

[142] GÉRARD-VARET, D., AND MASMOUDI, N. Homogenization and boundary layer. Preprint, available at the following address : `http://www.math.nyu.edu/faculty/masmoudi/homog_Varet3.pdf`, 2010.

[143] GOLOWICH, S. E., AND WEINSTEIN, M. I. Homogenization expansion for resonances of microstructured photonic waveguides. *J. Opt. Soc. Am. B 20*, 4 (Apr 2003), 633–647.

[144] GOLOWICH, S. E., AND WEINSTEIN, M. I. Scattering resonances of microstructures and homogenization theory. *Multiscale Model. Simul. 3*, 3 (2005), 477–521 (electronic).

[145] GOODMAN, R. H., HOLMES, P. J., AND WEINSTEIN, M. I. Strong nls soliton-defect interactions. *Physica D 192*, 3-4 (2004), 215–248.

[146] GRIFFITHS, D. J., AND STEINKE, C. A. Waves in locally periodic media. *Am. J. Phys. 69*, 2 (2001), 137–154.

[147] GRIFFITHS, D. J., AND TAUSSIG, N. F. Scattering from a locally periodic potential. *Am. J. of Phys. 60*, 10 (1992), 883–888.

[148] HOLMER, J., MARZUOLA, J., AND ZWORSKI, M. Soliton splitting by external delta potentials. *J. Nonlinear Sci. 17*, 4 (2007), 349–367.

[149] HOLMER, J., AND ZWORSKI, M. Slow soliton interaction with delta impurities. *J. Mod. Dyn. 1*, 4 (2007), 689–718.

[150] HÖRMANDER, L. *The analysis of linear partial differential operators. II Differential operators with constant coefficients*. Classics in Mathematics. Springer-Verlag, 2005.

[151] INOUE, K., AND OHTAKA, K. *Photonic crystals : physics, fabrication, and applications*. Springer Verlag, 2004.

[152] JACKSON, R. K., AND WEINSTEIN, M. I. Geometric analysis of bifurcation and symmetry breaking in a Gross-Pitaevskii equation. *J. Statist. Phys. 116*, 1-4 (2004), 881–905.

[153] Jikov, V. V., Kozlov, S. M., and Oleĭnik, O. A. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, 1994.

[154] Joannopoulos, J. D., Meade, R. D., and Winn, J. N. *Photonic Crystals. Molding the Flow of Light*. Princeton University Press, New-York, 1995.

[155] Kato, T. *Perturbation theory for linear operators*. Die Grundlehren der mathematischen Wissenschaften, Band 132. Springer-Verlag New York, Inc., New York, 1966.

[156] Kiang, D. Multiple scattering by a dirac comb. *Am. J. Phys. 42*, 9 (1974), 785–787.

[157] Krieger, J., and Schlag, W. Stable manifolds for all monic supercritical focusing nonlinear Schrödinger equations in one dimension. *J. Amer. Math. Soc. 19*, 4 (2006), 815–920.

[158] Le Coz, S., Fukuizumi, R., Fibich, G., Ksherim, B., and Sivan, Y. Instability of bound states of a nonlinear Schrödinger equation with a Dirac potential. *Phys. D 237*, 8 (2008), 1103–1128.

[159] Lippmann, B. A., and Schwinger, J. Variational principles for scattering processes. I. *Physical Rev. (2) 79* (1950), 469–480.

[160] Maas, L., and van Haren, J. Worden mooi-weer verdrinkingen door dood-water veroorzaakt. *Meteorologica 15* (2006), 211–216. In Dutch.

[161] Marzuola, J., Metcalfe, J., and Tataru, D. Strichartz estimates and local smoothing estimates for asymptotically flat Schrödinger equations. *J. Funct. Anal. 255*, 6 (2008), 1497–1553.

[162] Marzuola, J. L., and Weinstein, M. I. Long time dynamics near the symmetry breaking bifurcation for nonlinear Schrödinger/Gross-Pitaevskii equations. *Discrete Contin. Dyn. Syst. 28*, 4 (2010), 1505–1554.

[163] Milewski, P., Tabak, E., Turner, C., Rosales, R., and Menzaque, F. Nonlinear stability of two-layer flows. *Commun. Math. Sci. 2*, 3 (2004), 427–442.

[164] Moskow, S., and Vogelius, M. First-order corrections to the homogenized eigenvalues of a periodic composite medium. A convergence proof. *Proc. Roy. Soc. Edinburgh Sect. A 127*, 6 (1997), 1263–1299.

[165] Newton, R.G. Low-energy scattering for medium range potentials. *J. Math. Phys. 27* (1986), 2720–2730.

[166] Protopopov, B. E. Upstream generation of solitons : Numerical analysis of the dependence on key parameters. *J. Appl. Mech. Tech. Phys. 34* (1993), 85–90.

[167] Reed, M., and Simon, B. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1978.

[168] Reed, M., and Simon, B. *Methods of modern mathematical physics. III. Scattering theory*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1979.

[169] SANTOSA, F., AND VOGELIUS, M. First-order corrections to the homogenized eigenvalues of a periodic composite medium. *SIAM J. Appl. Math. 53*, 6 (1993), 1636–1668.

[170] SCHECHTER, M. *Operator methods in quantum mechanics*. North-Holland Publishing Co., New York, 1981.

[171] SHAMPINE, L. F., AND REICHELT, M. W. The MATLAB ODE suite. *SIAM J. Sci. Comput. 18*, 1 (1997), 1–22. Dedicated to C. William Gear on the occasion of his 60th birthday.

[172] SHEN, H. *Two PDE Problems from Electromagnetics*. PhD thesis, New York University, New York, 2007.

[173] SNYDER, A. W., AND LOVE, J. D. *Optical waveguide theory*. Science paperbacks. Chapman and Hall, London, 1983.

[174] STEIN, E. M. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.

[175] SULEM, C., AND SULEM, P.-L. *The nonlinear Schrödinger equation. Self-focusing and wave collapse*, vol. 139 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1999.

[176] TANG, S., AND ZWORSKI, M. Potential scattering on the real line. Lecture notes, available at `http://math.berkeley.edu/~zworski/tz1.pdf`.

[177] TARTAR, L. *The general theory of homogenization*, vol. 7 of *Lecture Notes of the Unione Matematica Italiana*. Springer-Verlag, Berlin, 2009.

[178] WEDER, R. The $W_{k,p}$-continuity of the Schrödinger wave operators on the line. *Comm. Math. Phys. 208*, 2 (1999), 507–520.

[179] YABLONOVITCH, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett. 58*, 20 (May 1987), 2059–2062.

[180] YABLONOVITCH, E., GMITTER, T. J., AND LEUNG, K. M. Photonic band structure : The face-centered-cubic case employing nonspherical atoms. *Phys. Rev. Lett. 67*, 17 (Oct 1991), 2295–2298.

[181] YAJIMA, K. The $W^{k,p}$-continuity of wave operators for Schrödinger operators. *J. Math. Soc. Japan 47*, 3 (1995), 551–581.