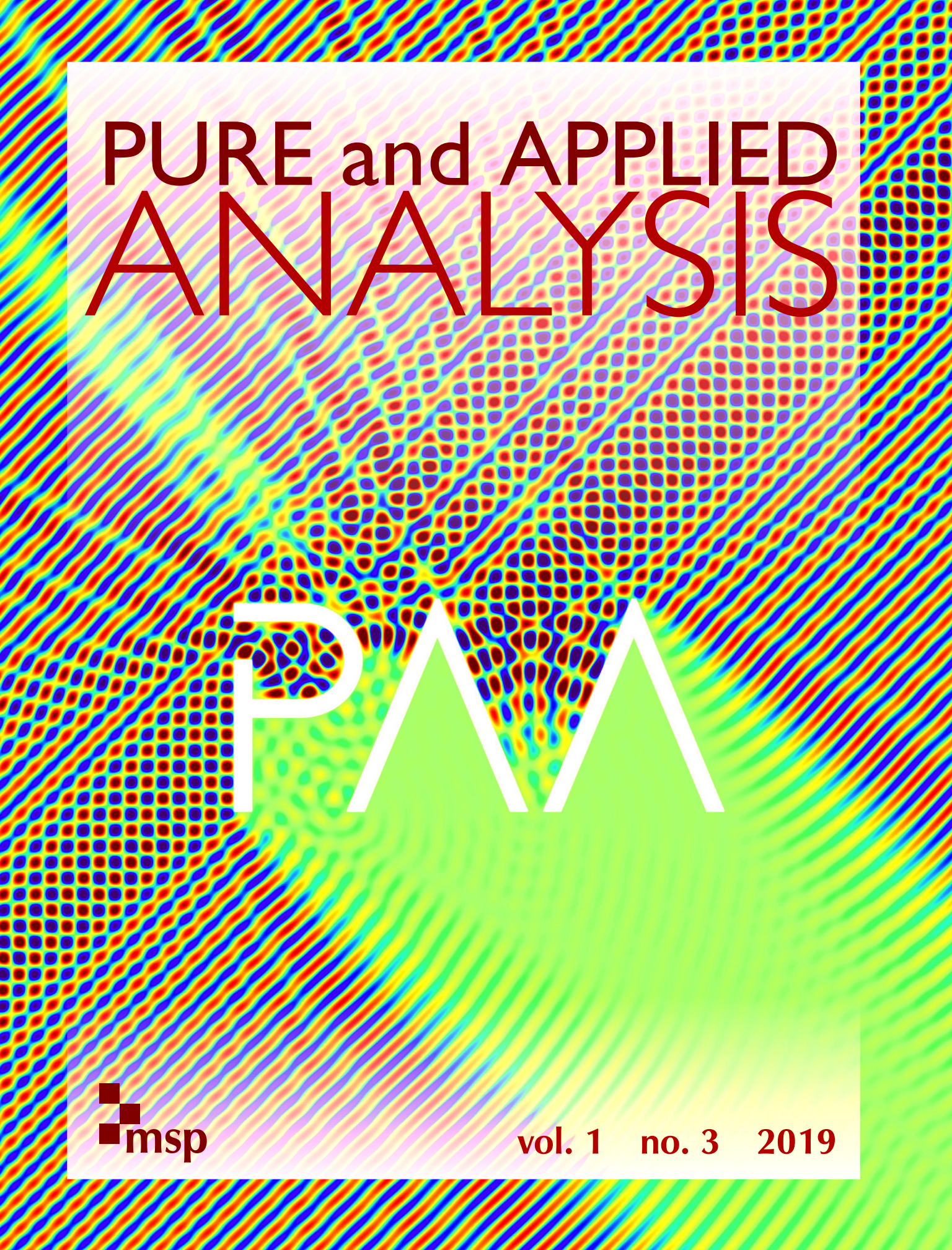

vol. 1 no. 3 2019

PURE and APPLIED
ANALYSIS

msp.org/paa

Cover image: The figure shows the outgoing scattered field produced by scattering a plane wave, coming from the northwest, off of the (stylized) letters P A A. The total field satisfies the homogeneous Dirichlet condition on the boundary of the letters. It is based on a numerical computation by Mike O'Neil of the Courant Institute.

See inside back cover or msp.org/paa for submission instructions.

The subscription price for 2019 is US $495/year for the electronic version, and $555/year (+$25, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Pure and Applied Analysis (ISSN 2578-5885 electronic, 2578-5893 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

PAA peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

PURE and APPLIED ANALYSIS

Vol. 1, No. 3, 2019
dx.doi.org/10.2140/paa.2019.1.327

msp

POSITIVITY, COMPLEX FIOS, AND TOEPLITZ OPERATORS

LEWIS A. COBURN, MICHAEL HITRIK AND JOHANNES SJÖSTRAND

We establish a characterization of complex linear canonical transformations that are positive with respect to a pair of strictly plurisubharmonic quadratic weights. As an application, we show that the boundedness of a class of Toeplitz operators on the Bargmann space is implied by the boundedness of their Weyl symbols.

1. Introduction and statement of results

The notion of a positive complex Lagrangian manifold, introduced in [Hörmander 1971], has long played an important role in microlocal analysis and spectral theory. Restricting the attention to the linear case, relevant for this work, let us recall that a complex Lagrangian plane $\Lambda \subset \mathbb{C}^{2n}$ is said to be positive if we have

$$\frac{1}{i}\sigma(\rho, \mathcal{C}(\rho)) \geq 0, \quad \rho \in \Lambda. \tag{1-1}$$

Here σ is the complex symplectic form on $\mathbb{C}^{2n}$ and $\mathcal{C} : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ is the antilinear map of complex conjugation. Let us mention here several familiar problems, where considerations of positive Lagrangian manifolds are essential. These include the spectral analysis and resolvent estimates for elliptic quadratic differential operators [Sjöstrand 1974; Hitrik et al. 2013], the study of spectral instability and pseudospectra for semiclassical nonnormal operators [Hörmander 1960; Dencker et al. 2004], as well as the construction of Gaussian beam quasimodes for semiclassical self-adjoint operators of principal type, associated with closed elliptic trajectories [Ralston 1976; Babich and Buldyrev 1991].

In [Sjöstrand 1982], one of us introduced and developed the notion of positivity of a complex Lagrangian space relative to a strictly plurisubharmonic quadratic weight, which is the starting point for the present work. To recall this notion, we let Φ_0 be a real-valued strictly plurisubharmonic quadratic form on $\mathbb{C}^n$ and let us introduce the real linear subspace

$$\Lambda_{\Phi_0} = \left\{ \left(x, \frac{2}{i}\frac{\partial \Phi_0}{\partial x}(x) \right) : x \in \mathbb{C}^n \right\} \subset \mathbb{C}^{2n}. \tag{1-2}$$

We can view Λ_{Φ_0} as the image of the real phase space $\mathbb{R}^{2n} \subset \mathbb{C}^{2n}$ under a suitable complex linear canonical transformation on $\mathbb{C}^{2n}$, and in particular we notice that Λ_{Φ_0} is maximally totally real. In

MSC2010: 32U05, 32W25, 35S30, 47B35, 70H15.

Keywords: positive Lagrangian plane, positive canonical transformation, strictly plurisubharmonic quadratic form, Fourier integral operator in the complex domain, Toeplitz operator.

analogy with the discussion above, we say that a complex linear Lagrangian space $\Lambda \subset \mathbb{C}^{2n}$ is positive relative to Λ_{Φ_0} provided that the natural analog of (1-1) holds,

$$\frac{1}{i}\sigma(\rho, \iota_{\Phi_0}(\rho)) \geq 0, \quad \rho \in \Lambda. \tag{1-3}$$

Here the map of complex conjugation $\mathcal{C}$ has been replaced by the unique antilinear involution $\iota_{\Phi_0} : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ such that $\iota_{\Phi_0}|_{\Lambda_{\Phi_0}} = 1$. A result of [Sjöstrand 1982] establishes a complete characterization of complex Lagrangians that are positive relative to Λ_{Φ_0} — see also Theorem 2.1 below.

In this work, we shall be mainly concerned with positive complex canonical transformations. Indeed, the main goal of the present work is to provide a characterization of positive complex linear canonical transformations relative to plurisubharmonic weights, and to consider Fourier integral operators (FIOs) in the complex domain associated to positive canonical transformations, establishing a link between such operators and Toeplitz operators. In particular, it seems that the point of view of complex FIOs allows us to shed some new light on some basic questions in the theory of Toeplitz operators. We would like to emphasize here that the original motivation for attempting to establish a link between FIOs in the complex domain and Toeplitz operators came from a talk delivered by Coburn at the conference "Complex and functional analysis and their interactions with harmonic analysis" at the Mathematical Research and Conference Center, Będlewo, June 2017.

We shall now proceed to define the notion of a complex linear canonical transformation which is positive relative to a strictly plurisubharmonic quadratic weight, and to state our main results. In fact, proceeding in the spirit of the discussion above, it will be more transparent to introduce the notion of positivity relative to a pair of strictly plurisubharmonic quadratic forms rather than relative to a single one. Thus, let Φ_1, Φ_2 be two strictly plurisubharmonic quadratic forms on $\mathbb{C}^n$ with the corresponding antilinear involutions ι_{Φ_1}, ι_{Φ_2}. Let $\kappa : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ be a complex linear canonical transformation, $\kappa^*\sigma = \sigma$. We say that κ is positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ provided that

$$\frac{1}{i}\big(\sigma(\kappa(\rho), \iota_{\Phi_1}\kappa(\rho)) - \sigma(\rho, \iota_{\Phi_2}(\rho))\big) \geq 0, \quad \rho \in \mathbb{C}^{2n}. \tag{1-4}$$

The positivity of κ relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ is said to be strict provided that the inequality in (1-4) is strict for all $0 \neq \rho \in \mathbb{C}^{2n}$. Let us remark that in the case when the positivity is taken relative to the real phase space $\mathbb{R}^{2n}$, see (1-1), such canonical transformations were studied in [Hörmander 1983, 1995]; see also the recent works [Pravda-Starov et al. 2018; Aleman and Viola 2018].

We can now state the first main result of this work.

Theorem 1.1. *Let $\kappa : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ be a complex linear canonical transformation and let Φ_1, Φ_2 be strictly plurisubharmonic quadratic forms on $\mathbb{C}^n$. The canonical transformation κ is positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ precisely when we have*

$$\kappa(\Lambda_{\Phi_2}) = \Lambda_{\Phi}, \tag{1-5}$$

where Φ is a strictly plurisubharmonic quadratic form such that $\Phi \leq \Phi_1$.

Remark. The definition (1-4) of a positive canonical transformation is a direct adaptation of the corresponding notion of positivity due to Hörmander [1983; 1995] to the weighted setting. One advantage of the consideration of the general case of a pair of weights Φ_1, Φ_2 is that we can let κ be the identity in (1-4) and get an invariant notion of the positivity of one plurisubharmonic weight compared to another, in view of Theorem 1.1.

Our second main result is concerned with applications of Theorem 1.1 to the study of Toeplitz operators in the Bargmann space

$$H_{\Phi_0}(\mathbb{C}^n) = L^2(\mathbb{C}^n, e^{-2\Phi_0}L(dx)) \cap \mathrm{Hol}(\mathbb{C}^n),$$

where Φ_0 is a strictly plurisubharmonic quadratic form on $\mathbb{C}^n$ and $L(dx)$ is the Lebesgue measure on $\mathbb{C}^n$. See also (A-1). Specifically, we shall be concerned with the continuity properties of (in general unbounded) Toeplitz operators of the form

$$\mathrm{Top}(e^{2q}) = \Pi_{\Phi_0} \circ e^{2q} \circ \Pi_{\Phi_0} : H_{\Phi_0}(\mathbb{C}^n) \to H_{\Phi_0}(\mathbb{C}^n), \tag{1-6}$$

where q is a complex-valued quadratic form on $\mathbb{C}^n$ and

$$\Pi_{\Phi_0} : L^2(\mathbb{C}^n, e^{-2\Phi_0}L(dx)) \to H_{\Phi_0}(\mathbb{C}^n)$$

is the orthogonal projection. Sufficient conditions for the boundedness of $\mathrm{Top}(e^{2q})$ are provided in the following result.

Theorem 1.2. *Let Φ_0 be a strictly plurisubharmonic quadratic form on $\mathbb{C}^n$ and let q be a quadratic form on $\mathbb{C}^n$ such that*

$$2\,\mathrm{Re}\, q(x) < \Phi_{\mathrm{herm}}(x) := \tfrac{1}{2}(\Phi_0(x) + \Phi_0(ix)), \quad x \neq 0, \tag{1-7}$$

$$\partial_x \partial_{\bar{x}}(\Phi_0 - q) \neq 0. \tag{1-8}$$

Let $a \in C^\infty(\Lambda_{\Phi_0})$ be the Weyl symbol of the Toeplitz operator $\mathrm{Top}(e^{2q})$. Assume that $a \in L^\infty(\Lambda_{\Phi_0})$. Then the Toeplitz operator

$$\mathrm{Top}(e^{2q}) : H_{\Phi_0}(\mathbb{C}^n) \to H_{\Phi_0}(\mathbb{C}^n)$$

is bounded.

Remark. Let us remark that Theorem 1.2 is closely related to the conjecture of [Berger and Coburn 1994; Coburn 2019] stating that a Toeplitz operator is bounded on $H_{\Phi_0}(\mathbb{C}^n)$ precisely when its Weyl symbol is bounded on Λ_{Φ_0}. Theorem 1.2 can therefore be regarded as establishing the sufficiency part of the conjecture in the special case when the Toeplitz symbol is of the form $\exp(2q)$, where q is a complex-valued quadratic form on $\mathbb{C}^n$, satisfying (1-7), (1-8).

Remark. As we shall see in Section 4, the strict inequality in condition (1-7) guarantees that the operator $\mathrm{Top}(e^{2q})$ is densely defined, and it seems difficult to weaken. Notice also that the Hermitian form Φ_{herm} in (1-7) is positive definite on $\mathbb{C}^n$, thanks to the strict plurisubharmonicity of Φ_0.

The plan of the paper is as follows. In Section 2, we establish the necessity part of Theorem 1.1, by means of direct geometric arguments, relying on some general results of [Sjöstrand 1982]; see also

[Caliceti et al. 2012; Hitrik and Sjöstrand 2018]. The proof of Theorem 1.1 is completed in Section 3, where we have found it convenient to introduce explicitly a Fourier integral operator in the complex domain quantizing the canonical transformation κ satisfying (1-5), when verifying the positivity of κ. Applications to Toeplitz operators are given in Section 4, where Theorem 1.2 is established. Appendix A is devoted to some elementary remarks concerning integral representations for linear continuous maps between weighted spaces of holomorphic functions, which can be regarded as a version of the Schwartz kernel theorem in this setting. These representations are to be applied in the main text when deriving a Bergman-type representation for our complex FIOs. Finally, Appendix B, for use in Section 4, characterizes boundedness properties of operators given as Weyl quantizations of symbols of the form $e^{iF(x,\xi)}$, where F is a holomorphic quadratic form on $\mathbb{C}^{2n}$.

2. Positive Lagrangian planes and positive canonical transformations in the H_Φ-setting

Let Φ_0 be a strictly plurisubharmonic quadratic form on $\mathbb{C}^n$. Associated to Φ_0 is the I-Lagrangian R-symplectic linear manifold Λ_{Φ_0}, given by

$$\Lambda_{\Phi_0} = \left\{ \left(x, \frac{2}{i}\frac{\partial \Phi_0}{\partial x}(x) \right) : x \in \mathbb{C}^n \right\} \subset \mathbb{C}^{2n}. \tag{2-1}$$

The linear manifold Λ_{Φ_0} is maximally totally real, and we let ι_{Φ_0} be the unique antilinear involution

$$\iota_{\Phi_0} : \mathbb{C}^{2n} \to \mathbb{C}^{2n} \tag{2-2}$$

such that the restriction of ι_{Φ_0} to Λ_{Φ_0} is the identity. For future reference, we may recall the explicit description of the involution ι_{Φ_0} given in [Hitrik and Sjöstrand 2018],

$$\left(y, \frac{2}{i}(\Phi_{0,xx}'' y + \Phi_{0,x\bar{x}}'' \bar{x}) \right) \mapsto \left(x, \frac{2}{i}(\Phi_{0,xx}'' x + \Phi_{0,x\bar{x}}'' \bar{y}) \right). \tag{2-3}$$

We also have

$$\iota_{\Phi_0} : \left(y, \frac{2}{i}\overline{\partial_y \Psi_0(x,\bar{y})} \right) \mapsto \left(x, \frac{2}{i}\partial_x \Psi_0(x,\bar{y}) \right), \tag{2-4}$$

where $\Psi_0(x,y)$ is the polarization of Φ_0, i.e., the unique holomorphic quadratic form on $\mathbb{C}^n_x \times \mathbb{C}^n_y$ such that $\Psi_0(x,\bar{x}) = \Phi_0(x)$.

Let $\Lambda \subset \mathbb{C}^{2n}$ be a $\mathbb{C}$-Lagrangian space, i.e., a complex linear subspace such that $\dim_{\mathbb{C}} \Lambda = n$ and $\sigma|_\Lambda = 0$. Here σ is the standard symplectic form on $\mathbb{C}^{2n}$. Let us consider the Hermitian form

$$b(\nu,\mu) = \frac{1}{i}\sigma(\nu, \iota_{\Phi_0}(\mu)), \quad \nu,\mu \in \mathbb{C}^{2n}. \tag{2-5}$$

We say that Λ is positive relative to Λ_{Φ_0} if the Hermitian form (2-5) is positive semidefinite when restricted to Λ,

$$b(\mu,\mu) \geq 0, \quad \mu \in \Lambda. \tag{2-6}$$

The positivity is said to be strict if the form b in (2-5) is positive definite along Λ. As remarked in the introduction, this notion is a direct adaptation of the corresponding notion of positivity due to Hörmander

[1971], where in place of $(\Lambda_{\Phi_0}, \iota_{\Phi_0})$ we have $(\mathbb{R}^{2n}, \mathcal{C})$, with $\mathcal{C}$ being the antilinear map of complex conjugation.

Remark. It is easy to see and is established in [Caliceti et al. 2012; Hitrik and Sjöstrand 2018] that the Hermitian form b is nondegenerate along Λ precisely when Λ and Λ_{Φ_0} are transversal.

Our starting point is the following well-known result; see [Sjöstrand 1982; Caliceti et al. 2012; Hitrik and Sjöstrand 2018].

Theorem 2.1. *A $\mathbb{C}$-Lagrangian space Λ is positive relative to Λ_{Φ_0} if and only if $\Lambda = \Lambda_\Psi$, where Ψ is a pluriharmonic quadratic form such that $\Psi \leq \Phi_0$.*

The proof of Theorem 2.1 given in [Sjöstrand 1982; Caliceti et al. 2012; Hitrik and Sjöstrand 2018] discusses the case of strictly positive Lagrangian planes only and depends on the general fact that the set of all $\mathbb{C}$-Lagrangian spaces which are strictly positive relative to Λ_{Φ_0} is a connected component in the set of all $\mathbb{C}$-Lagrangian spaces that are transversal to Λ_{Φ_0}. Here we shall give a more direct proof, using the explicit description of the involution ι_{Φ_0}, given in (2-3), (2-4). Let $\Lambda \subset \mathbb{C}^{2n}$ be $\mathbb{C}$-Lagrangian, positive relative to Λ_{Φ_0}. It follows from (2-3), as explained in [Sjöstrand 1982; Hitrik and Sjöstrand 2018], that the fiber $\{(0, \xi); \xi \in \mathbb{C}^n\}$ is strictly negative relative to Λ_{Φ_0}, in the sense that the Hermitian form b in (2-5) is negative definite along the fiber, and therefore Λ is necessarily of the form $\xi = \partial_x \varphi(x)$, where φ is a holomorphic quadratic form on $\mathbb{C}^n$. It follows that

$$\Lambda = \Lambda_\Psi, \tag{2-7}$$

where $\Psi = -\operatorname{Im} \varphi$ is pluriharmonic quadratic. We shall now see that $\Psi \leq \Phi_0$, and to this end, let us consider the decomposition

$$\Phi_0 = \Phi_{\text{herm}} + \Phi_{\text{plh}}, \tag{2-8}$$

where

$$\Phi_{\text{herm}}(x) = \Phi''_{0,\bar{x}x} x \cdot \bar{x} \tag{2-9}$$

is positive definite Hermitian and

$$\Phi_{\text{plh}}(x) = \operatorname{Re}(\Phi''_{0,xx} x \cdot x) \tag{2-10}$$

is pluriharmonic. Let

$$A = \frac{2}{i}(\Phi_{\text{plh}})''_{xx} = \frac{2}{i}(\Phi_0)''_{xx},$$

and let us consider the complex linear "vertical" canonical transformation

$$\kappa_A(y, \eta) = (y, \eta + Ay). \tag{2-11}$$

We have

$$\kappa_A(\Lambda_{\Phi_{\text{herm}}}) = \Lambda_{\Phi_0}, \tag{2-12}$$

and letting $\iota_{\Phi_{\text{herm}}}$ be the antilinear involution associated to $\Lambda_{\Phi_{\text{herm}}}$, it is then clear that

$$\iota_{\Phi_{\text{herm}}} = \kappa_A^{-1} \circ \iota_{\Phi_0} \circ \kappa_A. \tag{2-13}$$

It follows that Λ is positive relative to Λ_{Φ_0} precisely when

$$\kappa_A^{-1}(\Lambda) = \Lambda_{\Psi - \Phi_{\rm plh}}$$

is positive relative to $\Lambda_{\Phi_{\rm herm}}$, and when proving Theorem 2.1 we may assume therefore that the pluriharmonic part of Φ_0 vanishes. In this discussion, we are also allowed to perform complex linear changes of variables in $\mathbb{C}^n$, which correspond to canonical transformations of the form $\kappa_C : (y, \eta) \mapsto (C^{-1}y, C^t\eta)$, where C is an invertible complex $n \times n$ matrix. We have $\kappa_C(\Lambda_{\Phi_0}) = \Lambda_{\Phi_1}$, $\Phi_1(x) = \Phi_0(Cx)$, and it follows therefore that when establishing Theorem 2.1 it suffices to consider the model case when

$$\Phi_0(x) = \tfrac{1}{2}|x|^2. \tag{2-14}$$

An application of (2-3) shows that the involution ι_{Φ_0} is then given by

$$(y, \eta) \mapsto \left(\frac{1}{i}\bar{\eta}, \frac{1}{i}\bar{y}\right), \tag{2-15}$$

and therefore

$$b(\mu, \mu) = \frac{1}{i}\sigma(\mu, \iota_{\Phi_0}(\mu)) = |x|^2 - |\xi|^2, \quad \mu = (x, \xi) \in \mathbb{C}^{2n}. \tag{2-16}$$

When $\mu \in \Lambda = \Lambda_\Psi$, we write $\xi = (2/i)\partial_x\Psi(x) = \partial_x\varphi(x)$, $\Psi(x) = -\operatorname{Im}\varphi$, where φ is a quadratic holomorphic form, and therefore if Λ is positive relative to Λ_{Φ_0}, then (2-16) shows that

$$|\varphi''_{xx}x| \le |x|, \qquad x \in \mathbb{C}^n \iff \|\varphi''_{xx}\| \le 1. \tag{2-17}$$

We get

$$\Psi(x) = -\operatorname{Im}\varphi(x) \le \tfrac{1}{2}|\varphi''_{xx}x \cdot x| \le \tfrac{1}{2}|x|^2 = \Phi_0(x), \quad x \in \mathbb{C}^n. \tag{2-18}$$

Conversely, let Λ be $\mathbb{C}$-Lagrangian of the form $\Lambda = \Lambda_\Psi$, where Ψ is pluriharmonic quadratic such that $\Psi \le \Phi_0$. Let us write $\Psi = -\operatorname{Im}\varphi$, where φ is a holomorphic quadratic form. We shall now see that Λ_Ψ is positive relative to Λ_{Φ_0}, and it follows from the remarks above that it suffices to verify the positivity in the model case when Φ_0 is given by (2-14), so that we have

$$\Psi(x) = -\operatorname{Im}\varphi(x) \le \Phi_0(x) = \tfrac{1}{2}|x|^2. \tag{2-19}$$

Writing

$$-\operatorname{Im}\varphi''_{xx}x \cdot x \le |x|^2, \tag{2-20}$$

replacing x by $e^{i\theta}x$ and varying $\theta \in \mathbb{R}$, we get

$$|\varphi''_{xx}x \cdot x| \le |x|^2, \quad x \in \mathbb{C}^n. \tag{2-21}$$

Next, writing

$$\varphi''_{xx}x \cdot y = \tfrac{1}{4}\big(\varphi''_{xx}(x+y)\cdot(x+y) - \varphi''_{xx}(x-y)\cdot(x-y)\big),$$

we get, using (2-21),

$$|\varphi''_{xx}x \cdot y| \le \tfrac{1}{4}(|x+y|^2 + |x-y|^2) = \tfrac{1}{2}(|x|^2 + |y|^2). \tag{2-22}$$

Replacing $x \mapsto \lambda^{1/2}x,\ y \mapsto \lambda^{-1/2}y,\ \lambda > 0$, we get

$$|\varphi''_{xx}x \cdot y| \leq \frac{1}{2}\Big(\lambda|x|^2 + \frac{1}{\lambda}|y|^2\Big), \tag{2-23}$$

and choosing $\lambda = |y|/|x|$, assuming for simplicity that $x \neq 0,\ y \neq 0$, we obtain

$$|\varphi''_{xx}x \cdot y| \leq |x||y|.$$

Hence, $\|\varphi''_{xx}\| \leq 1$ and the positivity of Λ_Ψ relative to Λ_{Φ_0} follows from (2-16), (2-17). The proof of Theorem 2.1 is complete.

Remark. Closely related to the proof of Theorem 2.1 given above is the normal form for strictly plurisubharmonic quadratic forms, given in Lemma 5.1 of [Hörmander 1997]; see also [Harvey and Wells 1973].

Let Φ_1, Φ_2 be two strictly plurisubharmonic quadratic forms on $\mathbb{C}^n$ and let $\kappa : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ be a complex linear canonical transformation which is positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$, in the sense of (1-4). In the remainder of this section, we shall establish the necessity part of Theorem 1.1, while the sufficiency is discussed in Section 3. To this end, let us observe first that the linear I-Lagrangian R-symplectic manifold $\kappa(\Lambda_{\Phi_2})$ is transversal to the fiber $\{(0,\xi) : \xi \in \mathbb{C}^n\}$. Indeed, we have in view of (1-4),

$$\frac{1}{i}\sigma(\rho, \iota_{\Phi_1}(\rho)) \geq 0, \quad \rho \in \kappa(\Lambda_{\Phi_2}), \tag{2-24}$$

while, as recalled above, we know from [Sjöstrand 1982; Hitrik and Sjöstrand 2018] that the fiber is strictly negative relative to Λ_{Φ_1}. It follows that $\kappa(\Lambda_{\Phi_2}) = \Lambda_\Phi$, where Φ is a real quadratic form such that the Levi form $\bar\partial\partial\Phi$ is nondegenerate. When verifying that Φ is (necessarily strictly) plurisubharmonic, we claim that it suffices to do so when the pluriharmonic part of Φ_2 vanishes. Indeed, introducing the decomposition (2-8), with the quadratic form Φ_2 in place of Φ_0 and considering the canonical transformation κ_A given in (2-11), we see, using also (2-13), that κ is positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ precisely when $\kappa_A^{-1} \circ \kappa \circ \kappa_A$ is positive relative to $(\Lambda_{\Phi_1 - \Phi_{2,\mathrm{plh}}}, \Lambda_{\Phi_{2,\mathrm{herm}}})$. Here $\Phi_{2,\mathrm{plh}}$ and $\Phi_{2,\mathrm{herm}}$ are the pluriharmonic and the Hermitian parts of Φ_2, respectively. Here it is also helpful to notice that

$$\iota_{\Phi_1 - \Phi_{2,\mathrm{plh}}} = \kappa_A^{-1} \circ \iota_{\Phi_1} \circ \kappa_A.$$

To summarize, if we know that the generating function of the linear I-Lagrangian R-symplectic manifold

$$\kappa_A^{-1} \circ \kappa \circ \kappa_A(\Lambda_{\Phi_{2,\mathrm{herm}}})$$

is plurisubharmonic, then the same property is also enjoyed by the generating function of $\kappa(\Lambda_{\Phi_2})$. In what follows we shall assume therefore that

$$\Phi_{2,xx} = \Phi_{2,\bar x\,\bar x} = 0. \tag{2-25}$$

As above, in this discussion, we are also allowed to perform complex linear changes of variables in $\mathbb{C}^n$, which correspond to canonical transformations of the form $(y,\eta) \mapsto (C^{-1}y, C^t\eta)$, where C is an invertible

complex $n \times n$ matrix. Such canonical transformations preserve the plurisubharmonicity of the generating functions, and similarly to the proof of Theorem 2.1, it suffices therefore to consider the case when

$$\Phi_2(x) = \tfrac{1}{2}|x|^2. \tag{2-26}$$

Theorem 2.1 then shows that the $\mathbb{C}$-Lagrangian plane given by $\{(x,\xi) \in \mathbb{C}^{2n} : \xi = 0\}$ is strictly positive relative to Λ_{Φ_2}, and therefore $\kappa(\{(x,\xi) \in \mathbb{C}^{2n} : \xi = 0\})$ is strictly positive relative to Λ_{Φ_1}, in view of the positivity of κ. Another application of Theorem 2.1 gives that

$$\kappa(\{(x,\xi) \in \mathbb{C}^{2n} : \xi = 0\}) = \Lambda_\Psi, \tag{2-27}$$

where the quadratic form Ψ is pluriharmonic, with $\Psi \le \Phi_1$.

Let $\phi(x, y, \theta)$ be a holomorphic quadratic form on $\mathbb{C}^n_x \times \mathbb{C}^n_y \times \mathbb{C}^N_\theta$, which is a nondegenerate phase function in the sense of Hörmander, generating the graph of κ. It follows from (2-27), as explained in [Caliceti et al. 2012], that the quadratic form

$$\mathbb{C}^n \times \mathbb{C}^N \ni (y, \theta) \mapsto -\operatorname{Im}\phi(0, y, \theta) \tag{2-28}$$

is nondegenerate, and since it is pluriharmonic, the signature is necessarily $(n+N, n+N)$. Recalling that

$$\kappa(\Lambda_{\Phi_2}) = \Lambda_\Phi, \tag{2-29}$$

we see, using [Caliceti et al. 2012], that the quadratic form

$$(y, \theta) \mapsto -\operatorname{Im}\phi(0, y, \theta) + \Phi_2(y) \tag{2-30}$$

is nondegenerate as well. We would like to conclude that the signature of the quadratic form in (2-30) is also $(n+N, n+N)$, and to that end, we follow [Sjöstrand 1982] and consider the continuous deformation

$$[0, 1] \ni t \mapsto -\operatorname{Im}\phi(0, y, \theta) + t\Phi_2(y). \tag{2-31}$$

Using (2-16) we see that

$$\frac{1}{i}\sigma(\mu, \iota_{\Phi_2}(\mu)) \ge 0, \quad \mu \in \Lambda_{t\Phi_2}, \quad 0 \le t \le 1. \tag{2-32}$$

It follows as before that the I-Lagrangian manifold $\kappa(\Lambda_{t\Phi_2})$ is transversal to the fiber, $0 \le t \le 1$, and therefore we conclude that the nondegeneracy of the quadratic forms in (2-31) is maintained along the deformation $0 \le t \le 1$. Recalling that the set of nondegenerate quadratic forms of a fixed given signature is a connected component in the set of all nondegenerate quadratic forms, we conclude that the signature of the quadratic form in (2-30) is $(n+N, n+N)$. Now, as explained in [Caliceti et al. 2012], the quadratic form Φ in (2-29) is given by

$$\Phi(x) = \operatorname{vc}_{y,\theta}(-\operatorname{Im}\phi(x, y, \theta) + \Phi_2(y)), \tag{2-33}$$

where $\operatorname{vc}_{y,\theta}$ stands for the critical value with respect to y, θ, and we conclude by the fundamental lemma of [Sjöstrand 1982], see also [Hitrik and Sjöstrand 2018], that Φ is plurisubharmonic. (As already observed, the plurisubharmonicity of Φ is necessarily strict.)

We shall next see that $\Phi \leq \Phi_1$, and when doing so it will be convenient to discuss the following auxiliary result first, which may be of some independent interest.

Proposition 2.2. *Let $\kappa : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ be a complex linear canonical transformation which is positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$. If Φ_2 is strictly convex then κ has a generating function $\varphi(x, \eta)$ which is a holomorphic quadratic form such that*

$$\kappa : (\varphi'_\eta(x, \eta), \eta) \mapsto (x, \varphi'_x(x, \eta)). \tag{2-34}$$

Proof. It suffices to show that the map

$$\pi : \operatorname{graph}(\kappa) \ni (x, \xi; y, \eta) \mapsto (x, \eta) \in \mathbb{C}^{2n}$$

is bijective, i.e., injective. Let $(0, \xi; y, 0) \in \operatorname{Ker}(\pi)$ so that $\kappa : (y, 0) \mapsto (0, \xi)$. Let us consider the Hermitian forms

$$b_j(v, \mu) = \frac{1}{i}\sigma(v, \iota_{\Phi_j}(\mu)), \quad j = 1, 2.$$

The strict convexity of Φ_2 together with Theorem 2.1 implies

$$b_2((y, 0), (y, 0)) \asymp |y|^2, \quad y \in \mathbb{C}^n, \tag{2-35}$$

and the strict negativity of the fiber with respect to Λ_{Φ_1} gives

$$b_1((0, \xi), (0, \xi)) \asymp -|\xi|^2, \quad \xi \in \mathbb{C}^n.$$

Hence by the positivity of κ, we get

$$0 \leq b_1((0, \xi), (0, \xi)) - b_2((y, 0), (y, 0)) \asymp -(|\xi|^2 + |y|^2).$$

It follows that $(y, \xi) = 0$ and we conclude that π is injective. □

Remark. Suppose that the assumptions of Proposition 2.2 hold. The holomorphic quadratic form $\varphi(x, \theta) - y \cdot \theta$ is then a nondegenerate phase function generating the graph of κ.

Let us now turn to the proof of the fact that

$$\Phi \leq \Phi_1. \tag{2-36}$$

It follows from the remarks above that it suffices to verify (2-36) when the pluriharmonic part of Φ_2 vanishes, and since we are again allowed to perform complex linear changes of variables in $\mathbb{C}^n$, as before, we conclude that it suffices to consider the case when Φ_2 is given by (2-26). Proposition 2.2 applies and there exists therefore a holomorphic quadratic form $\varphi(x, \theta)$ such that

$$\kappa : (\varphi'_\theta(x, \theta), \theta) \mapsto (x, \varphi'_x(x, \theta)). \tag{2-37}$$

We shall now express the positivity of κ relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ in terms of the generating function φ. To this end, we shall first obtain an explicit expression for the Hermitian form

$$\frac{1}{i}\sigma((y, \eta), \iota_{\Phi_1}(y, \eta)), \quad (y, \eta) \in \mathbb{C}^{2n},$$

where we write

$$\Phi_1(x) = \tfrac{1}{2} L\bar{x}\cdot x + \mathrm{Re}(Ax\cdot x), \quad L = 2\Phi''_{1,x\bar{x}}, \quad A = \Phi''_{1,xx}. \tag{2-38}$$

Here L is Hermitian positive definite and performing a unitary transformation, we may assume, for simplicity, that L is diagonal, with real positive diagonal elements. A simple computation using (2-3) shows that

$$\frac{1}{i}\sigma((y,\eta),\iota_{\Phi_1}(y,\eta)) = L\bar{y}\cdot y + (2Ay - i\eta)\cdot x, \tag{2-39}$$

where

$$L\bar{x} = i\eta - 2Ay,$$

and therefore we get

$$\frac{1}{i}\sigma((y,\eta),\iota_{\Phi_1}(y,\eta)) = L\bar{y}\cdot y - L^{-1}(2iAy+\eta)\cdot\overline{(2iAy+\eta)}. \tag{2-40}$$

Using also (2-37), we conclude that κ is positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ precisely when

$$L^{-1}(\varphi'_x + 2iAx)\cdot\overline{(\varphi'_x + 2iAx)} + |\varphi'_\theta(x,\theta)|^2 \le L\bar{x}\cdot x + |\theta|^2, \quad (x,\theta)\in\mathbb{C}^{2n}. \tag{2-41}$$

It is now easy to conclude the proof of the necessity part of Theorem 1.1, using (2-41). It follows from (2-33) that we can write

$$\Phi(x) = \mathrm{vc}_{y,\theta}\big(-\mathrm{Im}(\varphi(x,\theta) - y\cdot\theta) + \Phi_2(y)\big). \tag{2-42}$$

At the unique critical point $(y(x),\theta(x))$, we have

$$y = \varphi'_\theta(x,\theta), \tag{2-43}$$

$$\frac{2}{i}\frac{\partial\Phi_2}{\partial y}(y) = \theta \iff \theta = \frac{1}{i}\bar{y}. \tag{2-44}$$

Injecting (2-44) into (2-42), we get

$$\Phi(x) = -\,\mathrm{Im}\,\varphi(x,\theta) - \tfrac{1}{2}|\theta|^2, \quad \theta = \theta(x), \tag{2-45}$$

and in view of (2-38), it suffices therefore to establish the inequality

$$-2\,\mathrm{Im}\,\varphi(x,\theta) \le L\bar{x}\cdot x + |\theta|^2 + 2\,\mathrm{Re}(Ax\cdot x), \quad (x,\theta)\in\mathbb{C}^{2n}. \tag{2-46}$$

When verifying (2-46), we write, using the Euler homogeneity relation,

$$2\varphi(x,\theta) = \varphi'_x(x,\theta)\cdot x + \varphi'_\theta(x,\theta)\cdot\theta, \tag{2-47}$$

and therefore,

$$-2\,\mathrm{Im}\,\varphi(x,\theta) = -\,\mathrm{Im}\big((\varphi'_x(x,\theta) + 2iAx)\cdot x + \varphi'_\theta(x,\theta)\cdot\theta\big) + 2\,\mathrm{Re}(Ax\cdot x). \tag{2-48}$$

An application of the Cauchy–Schwarz inequality with respect to the positive definite Hermitian forms $(x,y)\mapsto L^{-1}x\cdot\bar{y}$, $(x,y)\mapsto x\cdot\bar{y}$ together with the inequality $ab \le \frac{1}{2}a^2 + \frac{1}{2}b^2$ allows us to conclude

that the first term in the right-hand side of (2-48) does not exceed

$$\tfrac{1}{2}\big(L^{-1}(\varphi'_x + 2iAx)\cdot\overline{(\varphi'_x + 2iAx)} + L\bar{x}\cdot x + |\varphi'_\theta(x,\theta)|^2 + |\theta|^2\big).$$

The inequality (2-46) follows, in view of (2-41). The proof of the necessity part of Theorem 1.1 is complete.

Remark. In the context of Theorem 1.1, assume that $\Phi_1 = \Phi_2 =: \Phi_0$ and let us write

$$\Phi_0(x) = \sup_{y\in\mathbb{R}^n}(-\operatorname{Im}\varphi(x,y)), \tag{2-49}$$

where $\varphi(x,y)$ is a holomorphic quadratic form on $\mathbb{C}^n_x\times\mathbb{C}^n_y$ such that $\det\varphi''_{xy}\neq 0$ and $\operatorname{Im}\varphi''_{yy} > 0$. In the special case when Φ_0 is given by (2-26), we can take

$$\varphi(x,y) = i\big(\tfrac{1}{2}x^2 + \sqrt{2}x\cdot y + \tfrac{1}{2}y^2\big).$$

The complex canonical transformation

$$\kappa_\varphi : \mathbb{C}^{2n}\ni(y,-\varphi'_y(x,y))\mapsto(x,\varphi'_x(x,y))\in\mathbb{C}^{2n} \tag{2-50}$$

maps $\mathbb{R}^{2n}$ bijectively onto Λ_{Φ_0}, see [Hitrik and Sjöstrand 2018], and it exchanges the complex conjugation map $\mathcal{C}$ and the involution ι_{Φ_0}. Setting

$$\tilde{\kappa} = \kappa_\varphi^{-1}\circ\kappa\circ\kappa_\varphi, \tag{2-51}$$

we see that the complex linear canonical transformation $\tilde{\kappa}$ is positive in the sense of [Hörmander 1995],

$$\frac{1}{i}\big(\sigma(\tilde{\kappa}(\rho),\mathcal{C}\tilde{\kappa}(\rho)) - \sigma(\rho,\mathcal{C}(\rho))\big)\geq 0,\quad \rho\in\mathbb{C}^{2n}. \tag{2-52}$$

An application of Proposition 5.10 of [Hörmander 1995] allows us to conclude therefore that the map $\tilde{\kappa}$ enjoys the factorization

$$\tilde{\kappa} = \tilde{\kappa}_1\circ\tilde{\kappa}_2\circ\tilde{\kappa}_3, \tag{2-53}$$

where $\tilde{\kappa}_1$ and $\tilde{\kappa}_3$ are real linear canonical maps and the map $\tilde{\kappa}_2$ is of the form

$$\tilde{\kappa}_2 = \exp(-iH_{\tilde{q}}), \tag{2-54}$$

where $\tilde{q}$ is a quadratic form with $\operatorname{Re}\tilde{q}\geq 0$ on $\mathbb{R}^{2n}$ — see also the discussion in the proof of Proposition 5.12 of [Hörmander 1995]. We obtain the factorization

$$\kappa = \kappa_1\circ\kappa_2\circ\kappa_3, \tag{2-55}$$

where we have

$$\kappa_j : \Lambda_{\Phi_0}\to\Lambda_{\Phi_0},\quad j = 1,3, \tag{2-56}$$

and

$$\kappa_2 = \exp(-iH_q), \tag{2-57}$$

where q is a holomorphic quadratic form on $\mathbb{C}^{2n}$ such that $\operatorname{Re}q\geq 0$ along Λ_{Φ_0}. The representation (2-55) can be used to give an alternative proof of the basic inequality $\Phi\leq\Phi_0$ in Theorem 1.1, in this special case.

3. Positivity and Fourier integral operators

The purpose of this section is to establish the sufficiency part of Theorem 1.1. To this end, let Φ_1, Φ_2 be two strictly plurisubharmonic quadratic forms on $\mathbb{C}^n$ and let $\kappa : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ be a complex linear canonical transformation. Assume that

$$\kappa(\Lambda_{\Phi_2}) = \Lambda_\Phi, \tag{3-1}$$

where Φ is a strictly plurisubharmonic quadratic form such that

$$\Phi \le \Phi_1. \tag{3-2}$$

We shall establish the positivity of κ relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ by making a judicious choice of a nondegenerate phase function generating the graph of κ, and to this end, it will be convenient to consider a metaplectic Fourier integral operator associated to κ. Let therefore $\varphi(x, y, \theta)$ be a holomorphic quadratic form on $\mathbb{C}^n_x \times \mathbb{C}^n_y \times \mathbb{C}^N_\theta$, which is a nondegenerate phase function in the sense of Hörmander, generating the graph of κ. It follows from [Caliceti et al. 2012] that the plurisubharmonic quadratic form

$$\mathbb{C}^n \times \mathbb{C}^N \ni (y, \theta) \mapsto -\operatorname{Im}\varphi(0, y, \theta) + \Phi_2(y) \tag{3-3}$$

is nondegenerate of signature $(n + N, n + N)$. We conclude, following [Sjöstrand 1982; Caliceti et al. 2012] that the Fourier integral operator

$$Au(x) = \iint e^{i\varphi(x,y,\theta)} a u(y)\, dy\, d\theta, \quad a \in \mathbb{C}, \tag{3-4}$$

quantizing κ, can be realized by means of a good contour and we obtain a bounded linear map,

$$A : H_{\Phi_2}(\mathbb{C}^n) \to H_\Phi(\mathbb{C}^n). \tag{3-5}$$

Here

$$H_{\Phi_2}(\mathbb{C}^n) = \operatorname{Hol}(\mathbb{C}^n) \cap L^2(\mathbb{C}^n, e^{-2\Phi_2} L(dx)),$$

with $H_\Phi(\mathbb{C}^n)$ having an analogous definition.

We shall now discuss a Bergman-type representation of the bounded operator in (3-5); see also [Melin and Sjöstrand 2003] for a related discussion. To this end, let us recall from Theorem A.1 that we can write

$$Au(x) = \int K_A(x, \bar{y}) u(y)\, e^{-2\Phi_2(y)}\, L(dy) =: \tilde{A}u(x). \tag{3-6}$$

Here the kernel $K_A(x, z)$ is holomorphic on $\mathbb{C}^n_x \times \mathbb{C}^n_z$, with

$$y \mapsto \overline{K(x, \bar{y})} \in H_{\Phi_2}(\mathbb{C}^n),$$

uniquely determined by (3-6). If $u \in L^2_{\Phi_2}(\mathbb{C}^n) = L^2(\mathbb{C}^n, e^{-2\Phi_2} L(dx))$ is orthogonal to $H_{\Phi_2}(\mathbb{C}^n)$, we see from (3-6) that $\tilde{A}u = 0$. Hence the operator $\tilde{A}$ in (3-6) is a well-defined linear continuous map

$$\tilde{A} : L^2_{\Phi_2}(\mathbb{C}^n) \to H_{\Phi_2}(\mathbb{C}^n).$$

Furthermore, $\tilde{A}$ extends to a map: $\mathcal{E}'(\mathbb{C}^n) \to \mathrm{Hol}(\mathbb{C}^n)$ and we have

$$K_A(x,\bar{y})e^{-2\Phi_2(y)} = (\tilde{A}\delta_y)(x), \tag{3-7}$$

where $\delta_y \in \mathcal{E}'(\mathbb{C}^n)$ is the delta function at y. Let next $\Pi_2 : L^2_{\Phi_2}(\mathbb{C}^n) \to H_{\Phi_2}(\mathbb{C})$ be the orthogonal projection and let us recall from [Hitrik and Sjöstrand 2018] that the operator Π_2 is given by

$$\Pi_2 u(x) = a_2 \int e^{2\Psi_2(x,\bar{y})-\Phi_2(y)} u(y)\, L(dy), \quad a_2 > 0. \tag{3-8}$$

Here Ψ_2 is the polarization of Φ_2, i.e., a holomorphic quadratic form on $\mathbb{C}^{2n}_{x,y}$ such that $\Psi_2(x,\bar{x}) = \Phi_2(x)$. We get $\tilde{A}\delta_y = \tilde{A}\Pi_2\delta_y = A\Pi_2\delta_y$, and it follows from (3-7) that

$$K_A(x,\bar{y}) = A(a_2 e^{2\Psi_2(\cdot,\bar{y})})(x). \tag{3-9}$$

From [Hitrik and Sjöstrand 2018], let us recall the basic property

$$2\,\mathrm{Re}\,\Psi_2(x,\bar{y}) - \Phi_2(x) - \Phi_2(y) \sim -|x-y|^2$$

on $\mathbb{C}^n_x \times \mathbb{C}^n_y$, and in particular we have

$$2\,\mathrm{Re}\,\Psi_2(x,\bar{y}) \le \Phi_2(x) + \Phi_2(y). \tag{3-10}$$

It follows that

$$-\,\mathrm{Im}\,\varphi(0,\tilde{y},\theta) + 2\,\mathrm{Re}\,\Psi_2(\tilde{y},0) \le -\,\mathrm{Im}\,\varphi(0,\tilde{y},\theta) + \Phi_2(\tilde{y}). \tag{3-11}$$

Here, as observed in (3-3), the right-hand side is a nondegenerate plurisubharmonic quadratic form of signature $(n+N, n+N)$, and since the left-hand side is pluriharmonic, we conclude that it is also nondegenerate of signature $(n+N, n+N)$. Writing

$$-\,\mathrm{Im}\,\varphi(0,\tilde{y},\theta) + 2\,\mathrm{Re}\,\Psi_2(\tilde{y},0) = \mathrm{Re}(i\varphi(0,\tilde{y},\theta) + 2\Psi_2(\tilde{y},0)),$$

we conclude that the holomorphic quadratic form

$$\mathbb{C}^n \times \mathbb{C}^N \ni (\tilde{y},\theta) \mapsto i\varphi(0,\tilde{y},\theta) + 2\Psi_2(\tilde{y},0)$$

is nondegenerate. It follows that the holomorphic function

$$\mathbb{C}^n \times \mathbb{C}^N \ni (\tilde{y},\theta) \mapsto i\varphi(x,\tilde{y},\theta) + 2\Psi_2(\tilde{y},z)$$

has a unique critical point which is nondegenerate for each $(x,z) \in \mathbb{C}^n \times \mathbb{C}^n$. An application of exact (quadratic) stationary phase allows us therefore to conclude that

$$K_A(x,\bar{y}) = \hat{a}e^{2\Psi(x,\bar{y})}, \quad \hat{a} \in \mathbb{C}. \tag{3-12}$$

Here $\Psi(x,z)$ is a holomorphic quadratic form on $\mathbb{C}^{2n}$ given by

$$2\Psi(x,z) = \mathrm{vc}_{\tilde{y},\theta}(i\varphi(x,\tilde{y},\theta) + 2\Psi_2(\tilde{y},z)). \tag{3-13}$$

Let us now make the following basic observation.

Proposition 3.1. *The holomorphic quadratic form* $\Psi(x,z)$ *given in* (3-13) *satisfies*

$$2\,\mathrm{Re}\,\Psi(x,\bar{y}) \le \Phi(x)+\Phi_2(y), \quad (x,y)\in \mathbb{C}^n_x\times\mathbb{C}^n_y. \tag{3-14}$$

Proof. It will be more convenient to verify that

$$2\,\mathrm{Re}\,\Psi(x,y) \le \Phi(x)+\Phi_2^*(y), \quad (x,y)\in \mathbb{C}^n_x\times\mathbb{C}^n_y, \tag{3-15}$$

where $\Phi_2^*(y)=\Phi_2(\bar{y})$. A direct calculation shows that

$$\frac{2}{i}\partial_y\Phi_2^*(y) = -\overline{\frac{2}{i}(\partial_y\Phi_2)(\bar{y})},$$

or equivalently,

$$\frac{2}{i}\partial_y(\Phi_2^*)(\bar{y}) = -\overline{\frac{2}{i}(\partial_y\Phi_2)(y)}.$$

It follows that the antilinear involution

$$\Gamma : \mathbb{C}^{2n}\ni(y,\eta)\mapsto(\bar{y},-\bar{\eta})\in\mathbb{C}^{2n} \tag{3-16}$$

maps Λ_{Φ_2} bijectively onto $\Lambda_{\Phi_2^*}$. We conclude in view of (3-1) that

$$\kappa\circ\Gamma : \Lambda_{\Phi_2^*}\to\Lambda_\Phi, \tag{3-17}$$

and let us consider the graph of the map in (3-17), $\mathrm{Graph}(\kappa\circ\Gamma)\cap(\Lambda_\Phi\times\Lambda_{\Phi_2^*})$. Here $\Lambda_\Phi\times\Lambda_{\Phi_2^*}=\Lambda_{\Phi(x)+\Phi_2^*(y)}$ is I-Lagrangian and R-symplectic for the standard symplectic form

$$d\xi\wedge dx + d\eta\wedge dy \tag{3-18}$$

on $\mathbb{C}^{2n}_{x,\xi}\times\mathbb{C}^{2n}_{y,\eta}$ and we claim that $\mathrm{Graph}(\kappa\circ\Gamma)\cap(\Lambda_\Phi\times\Lambda_{\Phi_2^*})$ is Lagrangian for the symplectic form in (3-18), restricted to $\Lambda_\Phi\times\Lambda_{\Phi_2^*}$. This can be seen by a direct computation: when $(t,s)\in\Lambda_{\Phi_2^*}\times\Lambda_{\Phi_2^*}$ we have, writing σ for the standard symplectic form on $\mathbb{C}^{2n}$,

$$\sigma\big(\kappa(\Gamma(t)),\kappa(\Gamma(s))\big)+\sigma(t,s)=\sigma(\Gamma(t),\Gamma(s))+\sigma(t,s)=-\overline{\sigma(t,s)}+\sigma(t,s)=0,$$

since $\sigma(t,s)$ is real. Here we have also used that, by a straightforward computation,

$$\sigma(\Gamma t,\Gamma s)=-\overline{\sigma(t,s)}. \tag{3-19}$$

It is then well known that $\pi_{x,y}(\mathrm{Graph}(\kappa\circ\Gamma)\cap(\Lambda_\Phi\times\Lambda_{\Phi_2^*}))$, the projection of $\mathrm{Graph}(\kappa\circ\Gamma)\cap(\Lambda_\Phi\times\Lambda_{\Phi_2^*})$ in $\mathbb{C}^{2n}_{x,y}$, is maximally totally real; see [Melin and Sjöstrand 2003].

We now come to check (3-15). To this end, we observe that (3-13) gives

$$2\partial_x\Psi(x,y)=i\,\partial_x\varphi(x,\tilde{y},\theta), \tag{3-20}$$

$$2\partial_y\Psi(x,y)=2\partial_y\Psi_2(\tilde{y},y), \tag{3-21}$$

where

$$\partial_\theta\varphi(x,\tilde{y},\theta)=0, \quad \partial_{\tilde{y}}\varphi(x,\tilde{y},\theta)+\frac{2}{i}\partial_{\tilde{y}}\Psi_2(\tilde{y},y)=0. \tag{3-22}$$

We shall consider (3-20), (3-21) at the points $(x, y) \in \pi_{x,y}(\mathrm{Graph}(\kappa \circ \Gamma) \cap \Lambda_\Phi \times \Lambda_{\Phi_2^*})$, which corresponds to $\tilde{y} = \bar{y}$ in (3-22). Using (3-22) together with the fact that

$$\partial_{\tilde{y}} \Psi_2(\tilde{y}, \bar{\tilde{y}}) = \partial_{\tilde{y}} \Phi_2(\tilde{y}),$$

and (3-21) together with the fact that

$$(\partial_y \Psi_2)(\bar{y}, y) = \partial_y \Phi_2^*(y),$$

we conclude that at the points

$$(x, y) \in \pi_{x,y}(\mathrm{Graph}(\kappa \circ \Gamma) \cap \Lambda_\Phi \times \Lambda_{\Phi_2^*}),$$

the following equalities hold:

$$\partial_x \Psi(x, y) = \partial_x \Phi(x), \quad \partial_y \Psi(x, y) = \partial_y \Phi_2^*(y). \tag{3-23}$$

In other words,

$$\partial_x(\Phi(x) - 2\,\mathrm{Re}\,\Psi(x, y)) = \partial_y(\Phi_2^*(y) - 2\,\mathrm{Re}\,\Psi(x, y)) = 0,$$

along $\pi_{x,y}(\mathrm{Graph}(\kappa \circ \Gamma) \cap \Lambda_\Phi \times \Lambda_{\Phi_2^*})$, and thus the gradient of the real-valued function

$$F(x, y) = \Phi(x) + \Phi_2^*(y) - 2\,\mathrm{Re}\,\Psi(x, y) \tag{3-24}$$

vanishes on $\pi_{x,y}(\mathrm{Graph}(\kappa \circ \Gamma) \cap \Lambda_\Phi \times \Lambda_{\Phi_2^*})$. It follows that the strictly plurisubharmonic quadratic form $F(x, y)$ vanishes to the second order along

$$\pi_{x,y}(\mathrm{Graph}(\kappa \circ \Gamma) \cap \Lambda_\Phi \times \Lambda_{\Phi_2^*}), \tag{3-25}$$

and since the latter is maximally totally real, we get $F \geq 0$, thus implying (3-15). □

Remark. The strictly plurisubharmonic quadratic form $F(x, y)$ in (3-24) vanishes to the second order along the maximally totally real subspace (3-25), and therefore the conclusion that $F \geq 0$ can be strengthened to

$$F(x, y) \asymp \mathrm{dist}\big((x, y), \pi_{x,y}(\mathrm{Graph}(\kappa \circ \Gamma) \cap \Lambda_\Phi \times \Lambda_{\Phi_2^*})\big)^2.$$

Let us now return to the Bergman-type representation of the Fourier integral operator A in (3-4) quantizing κ. Combining (3-6) and (3-12), we get

$$Au(x) = \iint \check{a} e^{2(\Psi(x,\bar{y}) - \Phi_2(y))} u(y)\, dy\, d\bar{y} \tag{3-26}$$

for some $\check{a} \in \mathbb{C}$. This can be viewed as a Fourier integral operator

$$Au(x) = \iint \check{a} e^{2(\Psi(x,\theta) - \Psi_2(y,\theta))} u(y)\, dy\, d\theta, \tag{3-27}$$

where we take the integration contour $\theta = \bar{y}$ in (3-27).

Since $\partial_y \partial_\theta \Psi_2(y, \theta)$ is nondegenerate, the phase function

$$\phi(x, y, \theta) = \frac{2}{i}(\Psi(x, \theta) - \Psi_2(y, \theta)) \tag{3-28}$$

is nondegenerate in the sense of Hörmander, and the canonical transformation κ takes the form

$$\kappa : \Big(y, \frac{2}{i}\partial_y \Psi_2(y,\theta)\Big) \mapsto \Big(x, \frac{2}{i}\partial_x \Psi(x,\theta)\Big), \quad \text{with } \partial_\theta \Psi(x,\theta) = \partial_\theta \Psi_2(y,\theta). \tag{3-29}$$

We may also notice here that if we define

$$\kappa_\Psi : \Big(\theta, -\frac{2}{i}\partial_\theta \Psi(y,\theta)\Big) \mapsto \Big(y, \frac{2}{i}\partial_y \Psi(y,\theta)\Big)$$

and κ_{Ψ_2} similarly, then $\kappa = \kappa_\Psi \circ \kappa_{\Psi_2}^{-1}$.

The discussion so far shows that the canonical transformation κ enjoying the mapping properties (3-1), (3-2), admits a nondegenerate phase function of the form (3-28), where the quadratic form Ψ satisfies

$$2\,\mathrm{Re}\,\Psi(x,\bar{y}) \le \Phi_1(x) + \Phi_2(y), \quad (x,y) \in \mathbb{C}_x^n \times \mathbb{C}_y^n. \tag{3-30}$$

The positivity of κ relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ is then implied by the following general result.

Proposition 3.2. *Let κ be a canonical transformation satisfying* (3-1) *and let us consider a metaplectic Fourier integral operator of the form* (3-26), *or equivalently* (3-27), *associated to κ. Then the following conditions are equivalent*:

(i) *κ is positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$ in the sense of* (1-4):

$$\frac{1}{i}\sigma(t_1, \iota_{\Phi_1} t_1) - \frac{1}{i}\sigma(t_2, \iota_{\Phi_2} t_2) \ge 0, \quad \textit{whenever } t_1 = \kappa(t_2),\ t_2 \in \mathbb{C}^{2n}. \tag{3-31}$$

(ii) *$\Lambda_{2\,\mathrm{Re}\,\Psi(x,\bar{y})}$ is positive relative to $\Lambda_{\Phi_1(x)+\Phi_2(y)}$.*

(iii) *$2\,\mathrm{Re}\,\Psi(x,\bar{y}) - \Phi_1(x) - \Phi_2(y) \le 0$ on $\mathbb{C}_x^n \times \mathbb{C}_y^n$.*

Proof. The equivalence (ii)$\Leftrightarrow$(iii) follows from Theorem 2.1, so it suffices to show the equivalence (i)$\Leftrightarrow$(ii).

Clearly, (iii) is equivalent to

$$2\,\mathrm{Re}\,\Psi(x,y) - \Phi_1(x) - \Phi_2^*(y) \le 0 \quad \text{on } \mathbb{C}_{x,y}^{2n}, \tag{3-32}$$

where $\Phi_2^*(y) = \Phi_2(\bar{y})(= \overline{\Phi_2(\bar{y})})$, and by Theorem 2.1(ii) is equivalent to

$$\Lambda_{2\,\mathrm{Re}\,\Psi(x,y)} \text{ is positive relative to } \Lambda_{\Phi_1(x)+\Phi_2^*(y)}. \tag{3-33}$$

We have

$$\begin{aligned}\Lambda_{2\,\mathrm{Re}\,\Psi} &= \Big\{\Big(x, \frac{2}{i}\partial_x 2\,\mathrm{Re}\,\Psi(x,y); y, \frac{2}{i}\partial_y 2\,\mathrm{Re}\,\Psi(x,y)\Big)\Big\} \\ &= \Big\{\Big(x, \frac{2}{i}\partial_x \Psi(x,y); y, \frac{2}{i}\partial_y \Psi(x,y)\Big)\Big\},\end{aligned} \tag{3-34}$$

and (3-33) means that

$$\frac{1}{i}\sigma(t_1, \iota_{\Phi_1} t_1) + \frac{1}{i}\sigma(t_2, \iota_{\Phi_2^*} t_2) \ge 0 \quad \text{for all } (t_1,t_2) \in \Lambda_{2\,\mathrm{Re}\,\Psi}. \tag{3-35}$$

Here, we shall relate the involutions $\iota_{\Phi_2^*}$ and ι_{Φ_2}. From (2-4) let us recall that ι_{Φ_2} is given by

$$\iota_{\Phi_2} : \Big(y, \frac{2}{i}\overline{\partial_y \Psi_2(x,\bar{y})}\Big) \mapsto \Big(x, \frac{2}{i}\partial_x \Psi_2(x,\bar{y})\Big). \tag{3-36}$$

We also know that the antilinear involution Γ, given in (3-16), maps Λ_{Φ_2} bijectively onto $\Lambda_{\Phi_2^*}$, and since $\iota_{\Phi_2}, \iota_{\Phi_2^*}$ are the unique antilinear maps equal to the identity on Λ_{Φ_2} and $\Lambda_{\Phi_2^*}$ respectively, it follows that

$$\iota_{\Phi_2^*} = \Gamma \iota_{\Phi_2} \Gamma. \tag{3-37}$$

From (3-19), let us recall that

$$\frac{1}{i}\sigma(\Gamma t, \Gamma s) = \overline{\frac{1}{i}\sigma(t, s)},$$

so using (3-37), we find that the second term in (3-35) is equal to

$$\frac{1}{i}\sigma(t_2, \Gamma\iota_{\Phi_2}\Gamma t_2) = \overline{\frac{1}{i}\sigma(\Gamma t_2, \iota_{\Phi_2}\Gamma t_2)} = \frac{1}{i}\sigma(\Gamma t_2, \iota_{\Phi_2}\Gamma t_2) = -\frac{1}{i}\sigma(\iota_{\Phi_2}\Gamma t_2, \Gamma t_2),$$

where we also used the fact that $(1/i)\sigma(t, \iota_{\Phi_2} t)$ is real. Hence (3-33) is equivalent, via (3-35), to

$$\frac{1}{i}\sigma(t_1, \iota_{\Phi_1} t_1) - \frac{1}{i}\sigma(\iota_{\Phi_2}\Gamma t_2, \Gamma t_2) \geq 0 \quad \text{for all } (t_1, t_2) \in \Lambda_{2\operatorname{Re}\Psi}. \tag{3-38}$$

From (3-36), we get

$$\iota_{\Phi_2}\Gamma : \Big(\bar{y}, -\overline{\frac{2}{i}\partial_y \Psi_2(x, \bar{y})}\Big) \mapsto \Big(x, \frac{2}{i}\partial_x \Psi_2(x, \bar{y})\Big);$$

i.e.,

$$\iota_{\Phi_2}\Gamma : \Big(\theta, \frac{2}{i}\partial_\theta \Psi_2(y, \theta)\Big) \mapsto \Big(y, \frac{2}{i}\partial_y \Psi_2(y, \theta)\Big), \tag{3-39}$$

where we changed the notation slightly for convenience.

Write

$$\Lambda_{2\operatorname{Re}\Psi} \ni (t_1, t_2) = \Big(x, \frac{2}{i}\partial_x \Psi(x, \theta); \theta, \frac{2}{i}\partial_\theta \Psi(x, \theta)\Big),$$

and put $t_3 = \iota_{\Phi_2}\Gamma t_2$, so that by (3-39)

$$t_3 = \Big(y, \frac{2}{i}\partial_y \Psi_2(y, \theta)\Big),$$

where

$$\Big(\theta, \frac{2}{i}\partial_\theta \Psi(x, \theta)\Big) = \Big(\theta, \frac{2}{i}\partial_\theta \Psi_2(y, \theta)\Big).$$

Comparing with (3-29), we see that $t_1 = \kappa(t_3)$. Since $\Gamma t_2 = \iota_{\Phi_2}^2 \Gamma t_2 = \iota_{\Phi_2} t_3$, we see that (3-38) is equivalent to

$$\frac{1}{i}\sigma(t_1, \iota_{\Phi_1} t_1) - \frac{1}{i}\sigma(t_3, \iota_{\Phi_2} t_3) \geq 0, \quad \text{when } t_1 = \kappa(t_3), \tag{3-40}$$

which is precisely (3-31) up to a change of notation. This completes the proof of the equivalence (i)$\Leftrightarrow$(ii) and of the proposition. □

Combining Propositions 3.1 and 3.2, we see that the proof of the sufficiency part of Theorem 1.1 is now complete.

Remark. Let $\kappa : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ be a complex linear canonical transformation such that (3-1) holds, where Φ_2, Φ are strictly plurisubharmonic. It follows from (3-23) that the holomorphic quadratic form $\Psi(x, y)$

depends only on κ and on the weights Φ_2, Φ, but not on the choice of a nondegenerate phase function $\varphi(x, y, \theta)$, $(x, y, \theta) \in \mathbb{C}^n_x \times \mathbb{C}^n_y \times \mathbb{C}^N_\theta$ such that

$$\Lambda'_\varphi = \mathrm{Graph}(\kappa),$$

where

$$\Lambda'_\varphi = \{(x, \varphi'_x(x, y, \theta); y, -\varphi'_y(x, y, \theta)) : \varphi'_\theta(x, y, \theta) = 0\}.$$

It follows that if $\psi(x, y, w)$, $(x, y, w) \in \mathbb{C}^n_x \times \mathbb{C}^n_y \times \mathbb{C}^{N'}_w$, is a second nondegenerate phase function such that

$$\Lambda'_\varphi = \Lambda'_\psi = \mathrm{Graph}(\kappa),$$

then both φ and ψ give rise to the same Fourier integral operators, realized as bounded linear maps: $H_{\Phi_2}(\mathbb{C}^n) \to H_\Phi(\mathbb{C}^n)$.

We shall finish this section by making some remarks concerning metaplectic Fourier integral operators in the complex domain, associated to canonical transformations that are strictly positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$. Let

$$\kappa : \mathbb{C}^{2n} \to \mathbb{C}^{2n} \tag{3-41}$$

be a complex linear canonical transformation which is strictly positive relative to $(\Lambda_{\Phi_1}, \Lambda_{\Phi_2})$. According to Theorem 1.1, we then have

$$\kappa(\Lambda_{\Phi_2}) = \Lambda_\Phi, \tag{3-42}$$

where Φ is a strictly plurisubharmonic quadratic form on $\mathbb{C}^n$ such that

$$\Phi_1(x) - \Phi(x) \asymp |x|^2, \quad x \in \mathbb{C}^n. \tag{3-43}$$

Let

$$Tu(x) = \iint e^{i\phi(x,y,\theta)} a u(y)\, dy\, d\theta, \quad a \in \mathbb{C},$$

be a Fourier integral operator associated to κ. As discussed above, it follows from [Caliceti et al. 2012; Sjöstrand 1982] that the operator T can be realized by means of a suitable good contour and we then obtain a bounded operator

$$T : H_{\Phi_2}(\mathbb{C}^n) \to H_\Phi(\mathbb{C}^n). \tag{3-44}$$

It follows from (3-43) that the inclusion map $H_\Phi(\mathbb{C}^n) \to H_{\Phi_1}(\mathbb{C})$ is compact, and the operator $T : H_{\Phi_2}(\mathbb{C}^n) \to H_{\Phi_1}(\mathbb{C}^n)$ is therefore compact. The following sharpening is essentially well known; see [Aleman and Viola 2018].

Proposition 3.3. *The operator*

$$T : H_{\Phi_2}(\mathbb{C}^n) \to H_{\Phi_1}(\mathbb{C}^n)$$

is of trace class, with the singular values $s_j(T)$ satisfying

$$s_j(T) = \mathcal{O}(j^{-\infty}). \tag{3-45}$$

Proof. Let q be a holomorphic quadratic form on $\mathbb{C}^{2n}$ such that its restriction to Λ_{Φ_1} is real positive definite. Let us introduce the Weyl quantization of q, the operator $Q = q^w(x, D_x)$. The quadratic differential operator Q is self-adjoint on $H_{\Phi_1}(\mathbb{C}^n)$ with discrete spectrum, and let us consider the metaplectic Fourier integral operator e^{tQ}, $0 \le t \le t_0 \ll 1$, acting on the space $H_\Phi(\mathbb{C}^n)$. Using some well-known arguments, explained in detail in [Hérau et al. 2005; Hitrik and Pravda-Starov 2009; Hitrik et al. 2018], we see that, for $t \in [0, t_0]$ with $t_0 > 0$ small enough, the operator e^{tQ} is bounded,

$$e^{tQ} : H_\Phi(\mathbb{C}^n) \to H_{\Phi_t}(\mathbb{C}^n), \tag{3-46}$$

where Φ_t is a strictly plurisubharmonic quadratic form on $\mathbb{C}^n$, depending smoothly on $t \ge 0$ small enough, such that

$$\Phi_t(x) = \Phi(x) + \mathcal{O}(t)|x|^2. \tag{3-47}$$

Combining this observation with (3-43) we conclude that there exists $\delta > 0$ small enough such that the operator

$$e^{\delta Q} T : H_{\Phi_2}(\mathbb{C}^n) \to H_{\Phi_1}(\mathbb{C}^n) \tag{3-48}$$

is bounded. Writing

$$T = e^{-\delta Q} e^{\delta Q} T, \tag{3-49}$$

and applying the Ky Fan inequalities, we get

$$s_j(T) \le s_j(e^{-\delta Q}) \| e^{\delta Q} T \|_{\mathcal{L}(H_{\Phi_2}, H_{\Phi_1})} = \mathcal{O}(j^{-\infty}).$$

Here we have also used the fact that the singular values of the compact positive self-adjoint operator $e^{-\delta Q}$ on $H_{\Phi_1}(\mathbb{C}^n)$ satisfy

$$s_j(e^{-\delta Q}) = \mathcal{O}(j^{-\infty}).$$

It follows that T is of trace class and the proof of the proposition is complete. □

4. Applications to Toeplitz operators

The purpose of this section is to apply the point of view of Fourier integral operators in the complex domain, developed in the previous sections, to the study of Toeplitz operators in the Bargmann space, establishing Theorem 1.2.

Let Φ_0 be a strictly plurisubharmonic quadratic form on $\mathbb{C}^n$ and let $p : \mathbb{C}^n \to \mathbb{C}$ be measurable. Associated to p is the Toeplitz operator

$$\mathrm{Top}(p) = \Pi_{\Phi_0} \circ p \circ \Pi_{\Phi_0} : H_{\Phi_0}(\mathbb{C}^n) \to H_{\Phi_0}(\mathbb{C}^n). \tag{4-1}$$

Here

$$\Pi_{\Phi_0} : L^2(\mathbb{C}^n, e^{-2\Phi_0} L(dx)) \to H_{\Phi_0}(\mathbb{C}^n)$$

is the orthogonal projection. We shall always assume that when equipped with the natural domain

$$\mathcal{D}(\mathrm{Top}(p)) = \{u \in H_{\Phi_0}(\mathbb{C}^n) : pu \in L^2(\mathbb{C}^n, e^{-2\Phi_0} L(dx))\}, \tag{4-2}$$

the operator $\mathrm{Top}(p)$ becomes densely defined.

For future reference, let us recall the link between the Toeplitz and Weyl quantizations on $\mathbb{C}^n$. Let $p \in L^\infty(\mathbb{C}^n)$, say. Then we have

$$\mathrm{Top}(p) = a^w(x, D_x), \tag{4-3}$$

where $a \in C^\infty(\Lambda_{\Phi_0})$ is given by

$$a\left(x, \frac{2}{i}\frac{\partial \Phi_0}{\partial x}(x)\right) = \left(\exp\left(\tfrac{1}{4}(\Phi''_{0,x\bar{x}})^{-1}\partial_x \cdot \partial_{\bar{x}}\right)p\right)(x), \quad x \in \mathbb{C}^n. \tag{4-4}$$

See [Guillemin 1984; Sjöstrand 1996]. Here $-(\Phi''_{0,x\bar{x}})^{-1}\partial_x \cdot \partial_{\bar{x}}$ is a constant coefficient second-order differential operator on $\mathbb{C}^n$ whose symbol is the positive definite quadratic form

$$\tfrac{1}{4}(\Phi''_{0,x\bar{x}})^{-1}\bar{\xi} \cdot \xi > 0, \quad 0 \neq \xi \in \mathbb{C}^n \simeq \mathbb{R}^{2n},$$

and therefore the operator in (4-4) can be regarded as the forward heat flow acting on p.

In this section we shall be concerned with the question of when an operator of the form $\mathrm{Top}(p)$ is bounded,

$$\mathrm{Top}(p) \in \mathcal{L}(H_{\Phi_0}(\mathbb{C}^n), H_{\Phi_0}(\mathbb{C}^n)),$$

and following [Berger and Coburn 1994], in doing so we shall only consider Toeplitz symbols of the form

$$p = e^{2q}, \tag{4-5}$$

where q is a complex-valued quadratic form on $\mathbb{C}^n$. Let us first proceed to give an explicit criterion, guaranteeing that when equipped with the domain (4-2), the operator $\mathrm{Top}(e^{2q})$ is densely defined. Recalling the decomposition (2-8) and considering the unitary map

$$H_{\Phi_0}(\mathbb{C}^n) \ni u \mapsto ue^{-f} \in H_{\Phi_{\mathrm{herm}}}(\mathbb{C}^n), \quad f(x) = \Phi''_{0,xx}x \cdot x,$$

we may observe that the space $e^f\mathcal{P}(\mathbb{C}^n) = \{e^f p : p \in \mathcal{P}(\mathbb{C}^n)\}$ is dense in $H_{\Phi_0}(\mathbb{C}^n)$. Here $\mathcal{P}(\mathbb{C}^n)$ is the space of holomorphic polynomials on $\mathbb{C}^n$. It follows that

$$e^f\mathcal{P}(\mathbb{C}^n) \subset \mathcal{D}(\mathrm{Top}(e^{2q})),$$

so that $\mathrm{Top}(e^{2q})$ is densely defined, provided that

$$2\,\mathrm{Re}\, q(x) < \Phi_{\mathrm{herm}}(x), \tag{4-6}$$

in the sense of quadratic forms on $\mathbb{C}^n$.

Recalling (3-8), we may write

$$\mathrm{Top}(e^{2q})u(x) = C\int e^{2(\Psi_0(x,\bar{y})-\Phi_0(y))}e^{2q(y,\bar{y})}u(y)\,dy\,d\bar{y}, \quad u \in \mathcal{D}(\mathrm{Top}(e^{2q})). \tag{4-7}$$

Here $C > 0$ and Ψ_0 is the polarization of Φ_0. Similarly to (3-27), we get

$$\mathrm{Top}(e^{2q})u(x) = C\iint_\Gamma e^{2(\Psi_0(x,\theta)-\Psi_0(y,\theta)+q(y,\theta))}u(y)\,dy\,d\theta, \tag{4-8}$$

where Γ is the contour in $\mathbb{C}^{2n}$, given by $\theta = \bar{y}$. Here the holomorphic quadratic form

$$F(x, y, \theta) = \frac{2}{i}(\Psi_0(x, \theta) - \Psi_0(y, \theta) + q(y, \theta)) \tag{4-9}$$

is a nondegenerate phase function in the sense of Hörmander, in view of the fact that $\det \Psi''_{0,x\theta} \neq 0$, and therefore the operator $\mathrm{Top}(e^{2q})$ in (4-8) can be viewed as a metaplectic Fourier integral operator associated to a suitable canonical relation $\subset \mathbb{C}^{2n} \times \mathbb{C}^{2n}$. We have the formal factorization

$$\mathrm{Top}(e^{2q}) = AB,$$

where

$$Av(x) = \int e^{2\Psi_0(x,\theta)} v(\theta)\, d\theta, \quad Bu(\theta) = \int e^{-2\widetilde{\Psi}_0(y,\theta)} u(y)\, dy, \tag{4-10}$$

and where we have written $\widetilde{\Psi}_0(y, \theta) = \Psi_0(y, \theta) - q(y, \theta)$. Here the operator A, formally, is an elliptic Fourier integral operator associated to the canonical transformation

$$\Big(\theta, -\frac{2}{i}\partial_\theta \Psi_0(x, \theta)\Big) \mapsto \Big(x, \frac{2}{i}\partial_x \Psi_0(x, \theta)\Big).$$

It follows that the canonical relation associated to $\mathrm{Top}(e^{2q})$ is the graph of a canonical transformation if and only if this is the case for the Fourier integral operator B. We conclude that the operator $\mathrm{Top}(e^{2q})$ in (4-8) is associated to a canonical transformation precisely when

$$\partial_y \partial_\theta \widetilde{\Psi}_0 \neq 0. \tag{4-11}$$

The condition (4-11) is equivalent to the assumption (1-8) in Theorem 1.2. The canonical transformation is then given by

$$\kappa : (y, -\partial_y F(x, y, \theta)) \mapsto (x, \partial_x F(x, y, \theta)), \quad \partial_\theta F(x, y, \theta) = 0. \tag{4-12}$$

Example. In the following discussion, we shall revisit the family of examples discussed in Section 6 of [Berger and Coburn 1994] and show how the point of view of Fourier integral operators in the complex domain, developed above, allows one to recover the main findings of Section 6 of that paper, obtained there by means of a direct computation.

Let $\Phi_0(x) = \frac{1}{2}|x|^2$ and $q = \frac{1}{2}\lambda|y|^2$, $\lambda \in \mathbb{C}$ with $\mathrm{Re}\,\lambda < \frac{1}{2}$. Here the restriction on $\mathrm{Re}\,\lambda$ implies that (4-6) holds, so that the operator $\mathrm{Top}(e^{2q})$ is densely defined in $H_{\Phi_0}(\mathbb{C}^n)$. We have

$$\Psi_0(x, y) = \tfrac{1}{2} x \cdot y,$$

and the phase function F in (4-9) is given by

$$F(x, y, \theta) = \frac{2}{i}\Big(\frac{1}{2}x \cdot \theta - \Big(\frac{1-\lambda}{2}\Big) y \cdot \theta\Big). \tag{4-13}$$

In particular, the condition (4-11) is satisfied and we may then compute the canonical transformation κ associated to the corresponding Fourier integral operator $\mathrm{Top}(e^{2q})$ in (4-8).

The critical set C_F of the phase F is given by $\partial_\theta F = 0 \Longleftrightarrow x = (1-\lambda)y$, and the corresponding canonical transformation κ is of the form

$$\kappa : (y, -\partial_y F(x,y,\theta)) \mapsto (x, \partial_x F(x,y,\theta)), \quad (x,y,\theta) \in C_F. \tag{4-14}$$

It follows that κ is given by

$$\kappa : (y,\eta) \mapsto \Big((1-\lambda)y, \frac{\eta}{1-\lambda}\Big). \tag{4-15}$$

We shall now determine when the canonical transformation κ is positive relative to Λ_{Φ_0}, which can be done by a direct computation: it follows from (2-4) that the involution ι_{Φ_0} is given by

$$\iota_{\Phi_0} : (y,\eta) \mapsto \Big(\frac{1}{i}\bar\eta, \frac{1}{i}\bar y\Big), \tag{4-16}$$

and therefore, we may compute,

$$\begin{aligned}\frac{1}{i}\sigma(\kappa(y,\eta), \iota_{\Phi_0}\kappa(y,\eta)) &= \frac{1}{i}\sigma\bigg(\Big((1-\lambda)y, \frac{\eta}{1-\lambda}\Big), \Big(\frac{1}{i}\frac{\bar\eta}{1-\bar\lambda}, \frac{1}{i}(1-\bar\lambda)\bar y\Big)\bigg)\\ &= |1-\lambda|^2|y|^2 - \frac{|\eta|^2}{|1-\lambda|^2}.\end{aligned} \tag{4-17}$$

Similarly, we have

$$\frac{1}{i}\sigma((y,\eta), \iota_{\Phi_0}(y,\eta)) = |y|^2 - |\eta|^2. \tag{4-18}$$

Combining (4-17), (4-18) we see that the κ is positive relative to Λ_{Φ_0} if and only if

$$|1-\lambda| \geq 1. \tag{4-19}$$

This condition occurs in [Berger and Coburn 1994, pp. 581–582] (with the inessential difference that in the discussion in that paper one considers $\Phi_0(x) = \frac{1}{4}|x|^2$), where it is verified that the operator $\mathrm{Top}(e^{2q})$ is in $\mathcal{L}(H_{\Phi_0}(\mathbb{C}^n), H_{\Phi_0}(\mathbb{C}^n))$ precisely when (4-19) holds.

In the case when the strict inequality holds in (4-19), the canonical transformation κ in (4-15) is strictly positive relative to Λ_{Φ_0} and it follows from Proposition 3.3 that the Toeplitz operator $\mathrm{Top}(e^{2q})$ is of trace class on $H_{\Phi_0}(\mathbb{C}^n)$.

We shall now proceed to discuss the "boundary" case when

$$|1-\lambda| = 1. \tag{4-20}$$

In this case, using (4-15) we immediately see that $\kappa(\Lambda_{\Phi_0}) = \Lambda_{\Phi_0}$, and therefore we conclude, in view of [Caliceti et al. 2012; Sjöstrand 1982], that the operator

$$\mathrm{Top}(e^{2q}) : H_{\Phi_0}(\mathbb{C}^n) \to H_{\Phi_0}(\mathbb{C}^n) \tag{4-21}$$

is bounded, with a bounded two-sided inverse.

We claim next that the operator in (4-21) is in fact unitary when (4-20) holds, and when verifying the unitarity, it will be convenient to pass to the Weyl quantization, computing the Weyl symbol of $\mathrm{Top}(e^{2q})$.

It follows from (4-4) that

$$a\Big(x, \frac{2}{i}\frac{\partial\Phi_0}{\partial x}(x)\Big) = \Big(\exp\Big(\frac{\Delta}{8}\Big)e^{2q}\Big)(x) = \Big(\frac{2}{\pi}\Big)^n \int_{\mathbb{C}^n} e^{-2|x-y|^2} e^{\lambda|y|^2}\, L(dy). \tag{4-22}$$

Here Δ is the Laplacian on $\mathbb{C}^n \simeq \mathbb{R}^{2n}$. Computing the Gaussian integral in (4-22) by the exact version of stationary phase, we get, see also [Berger and Coburn 1994],

$$a\Big(x, \frac{2}{i}\frac{\partial\Phi_0}{\partial x}(x)\Big) = \Big(\frac{2}{2-\lambda}\Big)^n \exp\Big(\frac{2\lambda}{2-\lambda}|x|^2\Big). \tag{4-23}$$

Here we may notice that

$$\mathrm{Re}\Big(\frac{2\lambda}{2-\lambda}\Big) = 0,$$

when (4-20) holds, reflecting the fact that the associated canonical transformation in (4-15) is "real" in this case. We conclude that the Weyl symbol of the Toeplitz operator $\mathrm{Top}(e^{2q})$ is given by

$$a(x,\xi) = \Big(\frac{2}{2-\lambda}\Big)^n \exp(iF(x,\xi)), \quad F(x,\xi) = \frac{2\lambda}{2-\lambda}x\cdot\xi, \tag{4-24}$$

so that

$$\mathrm{Top}(e^{2q}) = \Big(\frac{2}{2-\lambda}\Big)^n (\exp(iF))^w. \tag{4-25}$$

We have $(\mathrm{Im}\, F)|_{\Lambda_{\Phi_0}} = 0$ and an application of Proposition 5.11 of [Hörmander 1995] together with the metaplectic invariance of the Weyl quantization allows us to conclude that the operator

$$\sqrt{\det(I-\mathcal{F}/2)}(\exp(iF))^w : H_{\Phi_0}(\mathbb{C}^n) \to H_{\Phi_0}(\mathbb{C}^n) \tag{4-26}$$

is unitary. Here $\mathcal{F}$ is the Hamilton map of F, i.e., the matrix of the (linear) Hamilton field H_F, and it remains therefore to check that

$$\sqrt{\det(I-\mathcal{F}/2)} = \Big(\frac{2}{2-\lambda}\Big)^n e^{i\theta}, \quad \theta\in\mathbb{R}. \tag{4-27}$$

To this end, we compute using (4-24),

$$\mathcal{F}/2 = \frac{\lambda}{2-\lambda}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad I-\mathcal{F}/2 = \frac{2}{2-\lambda}\begin{pmatrix}1-\lambda & 0\\ 0 & 1\end{pmatrix},$$

and (4-27) follows, thanks to (4-20). We conclude therefore that the Toeplitz operator $\mathrm{Top}(e^{2q})$ is unitary on $H_{\Phi_0}(\mathbb{C}^n)$, when $\mathrm{Re}\,\lambda < \frac{1}{2}$ and (4-20) holds. The unitarity property has also been observed in [Berger and Coburn 1994].

Remark. In the case when $\mathrm{Re}\,\lambda < \frac{1}{2}$, $|1-\lambda| > 1$, we observed that the operator $\mathrm{Top}(e^{2q})$ is of trace class on $H_{\Phi_0}(\mathbb{C}^n)$, and we get, using (4-24) and the metaplectic invariance of the Weyl quantization,

$$\mathrm{tr}\,\mathrm{Top}(e^{2q}) = \frac{1}{(2\pi)^n}\iint_{\Lambda_{\Phi_0}} a\frac{(\sigma|_{\Lambda_{\Phi_0}})^n}{n!},$$

where a is given in (4-24).

We are now ready to discuss the proof of Theorem 1.2. It follows from Theorem 1.1 and the discussion in this section that it suffices to check that the canonical transformation (4-12) associated to the operator $\mathrm{Top}(e^{2q})$ is positive relative to Λ_{Φ_0}. To this end, let us consider the Weyl symbol of $\mathrm{Top}(e^{2q})$, given by (4-4),

$$a(x,\xi) = \big(\exp\big(\tfrac{1}{4}(\Phi''_{0,x\bar{x}})^{-1}\partial_x\cdot\partial_{\bar{x}}\big)e^{2q}\big)(x), \quad (x,\xi)\in\Lambda_{\Phi_0}. \tag{4-28}$$

A simple computation of the inverse Fourier transform of a real Gaussian shows that

$$a(x,\xi) = C_{\Phi_0}\int_{\mathbb{C}^n} \exp(-4\Phi_{\mathrm{herm}}(x-y))e^{2q(y)}\,L(dy), \quad C_{\Phi_0}\neq 0. \tag{4-29}$$

Here the convergence of the integral in (4-29) is guaranteed by (4-6). In view of the exact version of stationary phase, it is therefore clear that

$$a(x,\xi) = C\exp(iF(x,\xi)), \quad (x,\xi)\in\Lambda_{\Phi_0}, \tag{4-30}$$

for some constant $C\neq 0$, where F is a holomorphic quadratic form on $\mathbb{C}^{2n}$. Proposition B.1 shows that the positivity of κ in (4-12) relative to Λ_{Φ_0} is equivalent to the fact that the Weyl symbol in (4-30) is such that $\mathrm{Im}\,F|_{\Lambda_{\Phi_0}}\geq 0 \Longleftrightarrow \exp(iF)\in L^\infty(\Lambda_{\Phi_0})$. The proof of Theorem 1.2 is complete.

Appendix A: Schwartz kernel theorem in the H_Φ-setting

In this appendix we shall make some elementary remarks concerning integral representations for linear continuous maps between weighted spaces of holomorphic functions. Such observations are essentially well known; see for instance [Peetre 1990].

Let $\Omega_j\subset\mathbb{C}^{n_j}$ be open, $j=1,2$, and let $\Phi_j\in C(\Omega_j;\mathbb{R})$. We introduce the weighted spaces

$$H_{\Phi_j}(\Omega_j) = \mathrm{Hol}(\Omega_j)\cap L^2(\Omega_j, e^{-2\Phi_j}L(dy_j)), \quad j=1,2, \tag{A-1}$$

where $L(dy_j)$ is the Lebesgue measure on $\mathbb{C}^{n_j}$. When viewed as closed subspaces of $L^2(\Omega_j, e^{-2\Phi_j}L(dy_j))$, the spaces $H_{\Phi_j}(\Omega_j)$ are separable complex Hilbert spaces and the natural embeddings $H_{\Phi_j}(\Omega_j)\to \mathrm{Hol}(\Omega_j)$ are continuous. Here the space $\mathrm{Hol}(\Omega_j)$ is equipped with its natural Fréchet space topology of locally uniform convergence. Let

$$T: H_{\Phi_1}(\Omega_1)\to H_{\Phi_2}(\Omega_2) \tag{A-2}$$

be a linear continuous map. Let us also write $\overline{\Omega}_1 = \{z\in\mathbb{C}^{n_1} : \bar{z}\in\Omega_1\}$.

Theorem A.1. *There exists a unique function $K(x,z)\in\mathrm{Hol}(\Omega_2\times\overline{\Omega}_1)$ such that*

$$\Omega_1\ni y\mapsto \overline{K(x,\bar{y})}\in H_{\Phi_1}(\Omega_1) \tag{A-3}$$

for each $x\in\Omega_2$, and

$$Tf(x) = \int_{\Omega_1} K(x,\bar{y})f(y)e^{-2\Phi_1(y)}\,L(dy), \quad f\in H_{\Phi_1}(\Omega_1). \tag{A-4}$$

We also have

$$\Omega_2\ni x\mapsto K(x,z)\in H_{\Phi_2}(\Omega_2) \tag{A-5}$$

for each $z\in\overline{\Omega}_1$.

When proving Theorem A.1, we observe that it follows from the remarks above that for each $x \in \Omega_2$, the linear form

$$H_{\Phi_1}(\Omega_1) \ni f \mapsto (Tf)(x) \in \mathbb{C} \tag{A-6}$$

is continuous, and there exists therefore a unique element $k_x \in H_{\Phi_1}(\Omega_1)$ such that for all $f \in H_{\Phi_1}(\Omega_1)$ we have

$$Tf(x) = (f, k_x)_{\Phi_1}, \quad x \in \Omega_2. \tag{A-7}$$

Here and in what follows $(\cdot,\cdot)_{\Phi_j}$ stands for the scalar product in the space $H_{\Phi_j}(\Omega_j),\ j = 1, 2$.

Letting (e_j) be an orthonormal basis for $H_{\Phi_1}(\Omega_1)$, we may write with convergence in $H_{\Phi_1}(\Omega_1)$, for each $x \in \Omega_2$ fixed,

$$k_x = \sum_{j=1}^{\infty} (k_x, e_j)_{\Phi_1} e_j = \sum_{j=1}^{\infty} \overline{Te_j(x)} e_j. \tag{A-8}$$

By Parseval's formula we get

$$\|k_x\|^2_{\Phi_1} = \sum_{j=1}^{\infty} |Te_j(x)|^2, \quad x \in \Omega_2. \tag{A-9}$$

Here we know that

$$\|k_x\|_{\Phi_1} = \sup_{\|f\|_{\Phi_1} \le 1} |Tf(x)|, \tag{A-10}$$

and it follows that the function $\Omega_2 \ni x \mapsto \|k_x\|_{\Phi_1}$ is locally bounded. Let us now make the following elementary observation: Let $\Omega \subset \mathbb{C}^n$ be open and let $f_n \in \mathrm{Hol}(\Omega)$ be such that the series

$$\sum_{n=1}^{\infty} |f_n(z)|^2 \tag{A-11}$$

converges for each $z \in \Omega$, with the sum being locally integrable in Ω. Then the series converges locally uniformly in Ω. Indeed, let us write

$$\sum_{n=1}^{\infty} |f_n(z)|^2 =: F(z) \in L^1_{\mathrm{loc}}(\Omega).$$

Let $K \subset \Omega$ be compact and let ω be an open neighborhood of K such that $K \subset \omega \Subset \Omega$. Then by Cauchy's integral formula and the Cauchy–Schwarz inequality we have

$$\sup_K |f_n|^2 \le \mathcal{O}_{K,\omega}(1) \|f_n\|^2_{L^2(\omega)}.$$

We get therefore the uniform bound

$$\sum_{n=1}^{N} \sup_K |f_n|^2 \le \mathcal{O}_{K,\omega}(1) \|F\|_{L^1(\omega)}, \quad N = 1, 2, \ldots,$$

implying the locally uniform convergence of (A-11).

It follows that (A-9) holds with locally uniform convergence in $x \in \Omega_2$, and in particular the function $\Omega_2 \ni x \mapsto \|k_x\|^2_{\Phi_1}$ is continuous plurisubharmonic. We may therefore conclude that the series in (A-8) converges locally uniformly in $\Omega_1 \times \Omega_2$. Letting

$$K(x,z) := \overline{k_x(\bar{z})} = \sum_{j=1}^{\infty} Te_j(x)\overline{e_j(\bar{z})}, \tag{A-12}$$

we conclude that $K \in \mathrm{Hol}(\Omega_2 \times \overline{\Omega}_1)$ is such that (A-3) and (A-4) hold, and these properties characterize the kernel K uniquely.

When verifying (A-5), we let $\tilde{k}_x \in H_{\Phi_2}(\Omega_2)$ be the reproducing kernel for $H_{\Phi_2}(\Omega_2)$. We may then write, when $f \in H_{\Phi_1}(\Omega_1),\ x \in \Omega_2$,

$$Tf(x) = (Tf, \tilde{k}_x)_{\Phi_2} = (f, T^*\tilde{k}_x)_{\Phi_1}, \tag{A-13}$$

and therefore,

$$k_x = T^*\tilde{k}_x. \tag{A-14}$$

Here

$$T^* : H_{\Phi_2}(\Omega_2) \to H_{\Phi_1}(\Omega_1)$$

is the adjoint of T. Letting (f_j) be an orthonormal basis for $H_{\Phi_2}(\Omega_2)$ and recalling that

$$\tilde{k}_x = \sum_{j=1}^{\infty} \overline{f_j(x)}\, f_j, \tag{A-15}$$

we get

$$k_x(y) = \sum_{j=1}^{\infty} \overline{f_j(x)} T^* f_j(y), \tag{A-16}$$

Therefore,

$$K(x,\bar{y}) = \sum_{j=1}^{\infty} f_j(x)\overline{T^* f_j(y)}.$$

and we see that (A-5) follows. We also get

$$\|K(\cdot,\bar{y})\|^2_{\Phi_2} = \sum_{j=1}^{\infty} |T^* f_j(y)|^2. \tag{A-17}$$

Remark. It follows from (A-9) that $T \in \mathcal{L}(H_{\Phi_1}(\Omega_1), H_{\Phi_2}(\Omega_2))$ is of Hilbert–Schmidt class precisely when

$$\iint_{\Omega_1\times\Omega_2} |K(x,\bar{y})|^2 e^{-2(\Phi_1(y)+\Phi_2(x))}\, L(dy)\, L(dx) < \infty.$$

Remark. An alternative proof of Theorem A.1 can be obtained by applying the Schwartz kernel theorem directly to the linear continuous map

$$\Pi_{\Phi_2} T \Pi_{\Phi_1} : L^2(\Omega_1, e^{-2\Phi_1} L(dy_1)) \to L^2(\Omega_2, e^{-2\Phi_2} L(dy_2)).$$

Here

$$\Pi_{\Phi_j} : L^2(\Omega_j, e^{-2\Phi_j} L(dy_j)) \to H_{\Phi_j}(\Omega_j)$$

is the orthogonal projection. Writing the Schwartz kernel of $\Pi_{\Phi_2} T \Pi_{\Phi_1}$ in the form $K(x, \bar{y})e^{-2\Phi_1(y)}$, we see that K should satisfy $\partial_{\bar{x}} K(x, \bar{y}) = 0$. Now the distribution kernel of the adjoint $\Pi_{\Phi_1} T^* \Pi_{\Phi_2}$ is given by $\overline{K(y, \bar{x})}e^{-2\Phi_2(y)}$, and it follows that $\partial_{\bar{x}}(\overline{K(y, \bar{x})}) = 0$. We get $\partial_x(K(y, \bar{x})) = 0$, so that $(\partial_{\bar{y}} K)(y, \bar{x}) = 0 \Longleftrightarrow \partial_{\bar{y}} K(x, y) = 0$. We conclude that $K(x, y)$ is holomorphic in (x, y).

Appendix B. Positivity and Weyl quantization

The purpose of this appendix is to characterize the boundedness of the Weyl quantization of a symbol of the form $\exp(iF(x, \xi))$, where F a complex quadratic form, in the H_Φ-setting. See also [Hörmander 1995] for a related discussion in the context of L^2-boundedness.

Let $F = F(x, \xi)$ be a complex-valued holomorphic quadratic form on $\mathbb{C}^{2n}$ and let us consider formally the Weyl quantization of $e^{iF(x,\xi)}$,

$$Au(x) = \mathrm{Op}^w(e^{iF})u(x) = \frac{1}{(2\pi)^n} \iint e^{i((x-y)\cdot\theta + F((x+y)/2,\theta))} u(y)\, dy\, d\theta. \tag{B-1}$$

The holomorphic quadratic form $(x - y)\cdot\theta + F\big(\frac{1}{2}(x + y), \theta\big)$ is a nondegenerate phase function in the sense of Hörmander and generates a canonical relation

$$\kappa : (y, \eta) \mapsto (x, \xi), \tag{B-2}$$

given by

$$\begin{aligned} x &= \frac{x+y}{2} - \tfrac{1}{2} F'_\xi\left(\frac{x+y}{2}, \theta\right), \quad \xi = \theta + \tfrac{1}{2} F'_x\left(\frac{x+y}{2}, \theta\right), \\ y &= \frac{x+y}{2} + \tfrac{1}{2} F'_\xi\left(\frac{x+y}{2}, \theta\right), \quad \eta = \theta - \tfrac{1}{2} F'_x\left(\frac{x+y}{2}, \theta\right). \end{aligned} \tag{B-3}$$

The graph is parametrized by $\rho = \big(\frac{1}{2}(x + y), \theta\big) \in \mathbb{C}^{2n}$ and (B-2), (B-3) take the form

$$\kappa : \rho + \tfrac{1}{2} H_F(\rho) \mapsto \rho - \tfrac{1}{2} H_F(\rho), \tag{B-4}$$

where $H_F(\rho) = (F'_\xi(\rho), -F'_x(\rho))$ is the Hamilton field of F at ρ.

We shall now give a criterion for when κ in (B-4) is a canonical transformation. Recall that $H_F(\rho) = \mathcal{F}\rho$, where

$$\mathcal{F} = \begin{pmatrix} F''_{\xi x} & F''_{\xi\xi} \\ -F''_{xx} & -F''_{x\xi} \end{pmatrix}$$

is the fundamental matrix of F (usually appearing as the linearization of a Hamilton vector field, which in our case is already linear). We have

$$\mathcal{F} = JF'', \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad F'' = \begin{pmatrix} F''_{xx} & F''_{x\xi} \\ F''_{\xi x} & F''_{\xi\xi} \end{pmatrix},$$

and we notice that $J^2 = -1,\ J^\top = -J$. Then (B-4) is the relation

$$(1 + \mathcal{F}/2)\rho \mapsto (1 - \mathcal{F}/2)\rho. \tag{B-5}$$

Now $\mathcal{F}$ is antisymmetric with respect to the bilinear form $\sigma(\nu, \mu) = J\nu \cdot \mu$; hence $1 - \mathcal{F}/2$ is bijective if and only if its transpose $1 + \mathcal{F}/2$ with respect to σ is bijective. We conclude that the following three statements are equivalent:

(i) κ is a canonical transformation.

(ii) $1 - \mathcal{F}/2$ is bijective.

(iii) $1 + \mathcal{F}/2$ is bijective.

In what follows, we shall assume that (i)–(iii) hold.

Let Φ_0 be a strictly plurisubharmonic quadratic form on $\mathbb{C}^n$ and let $\iota_{\Phi_0} : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ be the corresponding antilinear involution, i.e., the unique antilinear map which is equal to the identity on Λ_{Φ_0}. We shall now proceed to characterize the positivity of the canonical transformation κ in (B-4) relative to Λ_{Φ_0}. Let

$$[\mu, \nu] = \tfrac{1}{2} b(\mu, \nu), \tag{B-6}$$

where $b(\mu, \nu)$ has been defined in (2-5). It is a Hermitian form and κ is positive relative to Λ_{Φ_0} precisely when

$$[\nu, \nu] \geq [\mu, \mu] \quad \text{for all } \nu, \mu \text{ with } \nu = \kappa(\mu). \tag{B-7}$$

By (B-4) this is equivalent to

$$\big[\rho - \tfrac{1}{2} H_F(\rho), \rho - \tfrac{1}{2} H_F(\rho)\big] \geq \big[\rho + \tfrac{1}{2} H_F(\rho), \rho + \tfrac{1}{2} H_F(\rho)\big], \quad \rho \in \mathbb{C}^{2n},$$

or equivalently,

$$\mathrm{Re}[H_F(\rho), \rho] \leq 0, \quad \rho \in \mathbb{C}^{2n}. \tag{B-8}$$

To simplify the following discussion, we shall make use of the invariance (exact Egorov theorem) under conjugation of A in (B-1) with a unitary metaplectic Fourier integral operator $U : L^2(\mathbb{R}^n) \to H_{\Phi_0}(\mathbb{C}^n)$ with the associated canonical transformation κ_U, mapping $\mathbb{R}^{2n}$ onto Λ_{Φ_0}. The operator $B = U^{-1} A U$ is the Weyl quantization of e^{iG}, where $G = F \circ \kappa_U$. Also $\iota_{\Phi_0} = \kappa_U \mathcal{C} \kappa_U^{-1}$, where $\mathcal{C}$ is the involution associated to $\mathbb{R}^{2n}$, which is just the map of ordinary complex conjugation. By abuse of notation we write F also for the pull back $F \circ \kappa_U$ and we continue the discussion in the case when Λ_{Φ_0} has been replaced with $\mathbb{R}^{2n}$ and ι_{Φ_0} with $\mathcal{C}$, $\mathcal{C}(\rho) = \bar{\rho}$. In this setting, (B-8) becomes

$$\mathrm{Im}\, \sigma(F'_\xi(\rho), -F'_x(\rho); \bar{x}, \bar{\xi}) \leq 0 \quad \text{for all } \rho = (x, \xi) \in \mathbb{C}^{2n};$$

i.e.,

$$\mathrm{Im}(F'_x(x, \xi) \cdot \bar{x} + F'_\xi(x, \xi) \cdot \bar{\xi}) \geq 0, \quad (x, \xi) \in \mathbb{C}^{2n},$$

or even more simply,

$$\mathrm{Im}(F''_{\rho\rho} \rho \cdot \bar{\rho}) \geq 0.$$

Writing $\rho = \mu + i\nu,\ \mu, \nu \in \mathbb{R}^{2n}$, we see that the last inequality is equivalent to

$$\operatorname{Im} F''\mu \cdot \mu + \operatorname{Im} F''\nu \cdot \nu \geq 0;$$

i.e.,

$$\operatorname{Im} F'' \geq 0;$$

i.e.,

$$\operatorname{Im} F \geq 0 \quad \text{on } \mathbb{R}^{2n}.$$

By the metaplectic invariance it follows that the positivity condition (B-7) is equivalent to

$$\operatorname{Im} F \geq 0 \quad \text{on } \Lambda_{\Phi_0}, \tag{B-9}$$

now with the original F.

Remark. The condition (B-9) is quite natural since we know that for ordinary symbols instead of e^{iF}, the natural contour of integration in (B-1) should be

$$\theta = \frac{2}{i}\partial_x \Phi\left(\frac{x+y}{2}\right);$$

see [Sjöstrand 1996; Hitrik and Sjöstrand 2018].

We summarize the discussion in this section in the following result.

Proposition B.1. *Let F be a holomorphic quadratic form on $\mathbb{C}^{2n}$ such that the fundamental matrix of F does not have the eigenvalues ± 2. Let Φ_0 be a strictly plurisubharmonic quadratic form on $\mathbb{C}^n$. The canonical transformation associated to the Fourier integral operator $\operatorname{Op}^w(e^{iF})$ is positive relative to Λ_{Φ_0} precisely when*

$$\operatorname{Im} F|_{\Lambda_{\Phi_0}} \geq 0. \tag{B-10}$$

In particular, if (B-10) *holds, then the operator*

$$\operatorname{Op}^w(e^{iF}) : H_{\Phi_0}(\mathbb{C}^n) \to H_{\Phi_0}(\mathbb{C}^n)$$

is bounded.

Acknowledgements

Hitrik would like to express his sincere and profound gratitude to the Institut de Mathématiques de Bourgogne at the Université de Bourgogne for the kind hospitality in August–September 2017, where part of this project was conducted. We are grateful to the referee for helpful suggestions and remarks.

References

[Aleman and Viola 2018] A. Aleman and J. Viola, "On weak and strong solution operators for evolution equations coming from quadratic operators", *J. Spectr. Theory* **8**:1 (2018), 33–121. MR Zbl

[Babich and Buldyrev 1991] V. M. Babič and V. S. Buldyrev, *Short-wavelength diffraction theory: asymptotic methods*, Springer Series on Wave Phenomena **4**, Springer, 1991. MR Zbl

[Berger and Coburn 1994] C. A. Berger and L. A. Coburn, "Heat flow and Berezin–Toeplitz estimates", *Amer. J. Math.* **116**:3 (1994), 563–590. MR Zbl

[Caliceti et al. 2012] E. Caliceti, S. Graffi, M. Hitrik, and J. Sjöstrand, "Quadratic $\mathcal{PT}$-symmetric operators with real spectrum and similarity to self-adjoint operators", *J. Phys. A* **45**:44 (2012), art. id. 444007. MR Zbl

[Coburn 2019] L. Coburn, "Fock space, the Heisenberg group, heat flow and Toeplitz operators", pp. 1–15 in *Handbook of analytic operator theory*, edited by K. Zhu, CRC Press, Boca Raton, FL, 2019.

[Dencker et al. 2004] N. Dencker, J. Sjöstrand, and M. Zworski, "Pseudospectra of semiclassical (pseudo-)differential operators", *Comm. Pure Appl. Math.* **57**:3 (2004), 384–415. MR Zbl

[Guillemin 1984] V. Guillemin, "Toeplitz operators in n dimensions", *Integral Equations Operator Theory* **7**:2 (1984), 145–205. MR Zbl

[Harvey and Wells 1973] F. R. Harvey and R. O. Wells, Jr., "Zero sets of non-negative strictly plurisubharmonic functions", *Math. Ann.* **201** (1973), 165–170. MR Zbl

[Hérau et al. 2005] F. Hérau, J. Sjöstrand, and C. C. Stolk, "Semiclassical analysis for the Kramers–Fokker–Planck equation", *Comm. Partial Differential Equations* **30**:4-6 (2005), 689–760. MR Zbl

[Hitrik and Pravda-Starov 2009] M. Hitrik and K. Pravda-Starov, "Spectra and semigroup smoothing for non-elliptic quadratic operators", *Math. Ann.* **344**:4 (2009), 801–846. MR Zbl

[Hitrik and Sjöstrand 2018] M. Hitrik and J. Sjöstrand, "Two minicourses on analytic microlocal analysis", pp. 483–540 in *Algebraic and analytic microlocal analysis* (Evanston, IL, 2013), edited by M. Hitrik et al., Springer Proc. Math. Stat. **269**, Springer, 2018.

[Hitrik et al. 2013] M. Hitrik, J. Sjöstrand, and J. Viola, "Resolvent estimates for elliptic quadratic differential operators", *Anal. PDE* **6**:1 (2013), 181–196. MR Zbl

[Hitrik et al. 2018] M. Hitrik, K. Pravda-Starov, and J. Viola, "From semigroups to subelliptic estimates for quadratic operators", *Trans. Amer. Math. Soc.* **370**:10 (2018), 7391–7415. MR Zbl

[Hörmander 1960] L. Hörmander, "Differential equations without solutions", *Math. Ann.* **140** (1960), 169–173. MR Zbl

[Hörmander 1971] L. Hörmander, "On the existence and the regularity of solutions of linear pseudo-differential equations", *Enseignement Math.* (2) **17** (1971), 99–163. MR Zbl

[Hörmander 1983] L. Hörmander, "L^2 estimates for Fourier integral operators with complex phase", *Ark. Mat.* **21**:2 (1983), 283–307. MR Zbl

[Hörmander 1995] L. Hörmander, "Symplectic classification of quadratic forms, and general Mehler formulas", *Math. Z.* **219**:3 (1995), 413–449. MR Zbl

[Hörmander 1997] L. Hörmander, "On the Legendre and Laplace transformations", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **25**:3-4 (1997), 517–568. MR Zbl

[Melin and Sjöstrand 2003] A. Melin and J. Sjöstrand, "Bohr–Sommerfeld quantization condition for non-selfadjoint operators in dimension 2", pp. 181–244 in *Autour de l'analyse microlocale*, edited by G. Lebeau, Astérisque **284**, Soc. Math. France, Paris, 2003. MR Zbl

[Peetre 1990] J. Peetre, "The Berezin transform and Ha-plitz operators", *J. Operator Theory* **24**:1 (1990), 165–186. MR Zbl

[Pravda-Starov et al. 2018] K. Pravda-Starov, L. Rodino, and P. Wahlberg, "Propagation of Gabor singularities for Schrödinger equations with quadratic Hamiltonians", *Math. Nachr.* **291**:1 (2018), 128–159. MR Zbl

[Ralston 1976] J. V. Ralston, "On the construction of quasimodes associated with stable periodic orbits", *Comm. Math. Phys.* **51**:3 (1976), 219–242. MR Zbl

[Sjöstrand 1974] J. Sjöstrand, "Parametrices for pseudodifferential operators with multiple characteristics", *Ark. Mat.* **12** (1974), 85–130. MR Zbl

[Sjöstrand 1982] J. Sjöstrand, "Singularités analytiques microlocales", pp. 1–166 in Astérisque **95**, Soc. Math. France, Paris, 1982. MR Zbl

[Sjöstrand 1996] J. Sjöstrand, "Function spaces associated to global I-Lagrangian manifolds", pp. 369–423 in *Structure of solutions of differential equations* (Katata/Kyoto, 1995), edited by M. Morimoto and T. Kawai, World Sci., River Edge, NJ, 1996. MR Zbl

Received 3 Jul 2018. Revised 26 Feb 2019. Accepted 4 May 2019.

LEWIS A. COBURN: lcoburn@buffalo.edu
Department of Mathematics, SUNY at Buffalo, Buffalo, NY, United States

MICHAEL HITRIK: hitrik@math.ucla.edu
Department of Mathematics, University of California, Los Angeles, CA, United States

JOHANNES SJÖSTRAND: johannes.sjostrand@u-bourgogne.fr
IMB, Université de Bourgogne, Dijon, France

PURE and APPLIED
ANALYSIS

Vol. 1, No. 3, 2019
dx.doi.org/10.2140/paa.2019.1.359

msp

MICROLOCAL ANALYSIS OF FORCED WAVES

SEMYON DYATLOV AND MACIEJ ZWORSKI

Dedicated to Richard Melrose on the occasion of his 70th birthday

We use radial estimates for pseudodifferential operators to describe long-time evolution of solutions to $iu_t - Pu = f$, where P is a self-adjoint zeroth-order pseudodifferential operator satisfying hyperbolic dynamical assumptions and where f is smooth. This is motivated by recent results of Colin de Verdière and Saint-Raymond (2019) concerning a microlocal model of internal waves in stratified fluids.

1. Introduction

Colin de Verdière and Saint-Raymond [2019] recently found an interesting connection between modeling of internal waves in stratified fluids and spectral theory of zeroth-order pseudodifferential operators on compact manifolds. In other problems of fluid mechanics, relevance of such operators has been known for a long time, for instance in [Ralston 1973]. We refer to [Colin de Verdière and Saint-Raymond 2019] for pointers to current physics literature on internal waves and for numerical and experimental illustrations.

The purpose of this note is to show how the main result of [Colin de Verdière and Saint-Raymond 2019] (see also [Colin de Verdière 2018]) follows from the now standard radial estimates for pseudodifferential operators. In particular, we avoid the use of Mourre theory, normal forms and Fourier integral operators and do not assume that the subprincipal symbols vanish. We also relax some geometric assumptions. The conclusions are formulated in terms of Lagrangian regularity in the sense of [Hörmander 1985a, §25.1]. We illustrate the results with numerical examples. There are many possibilities for refinements but we restrict ourselves to applying off-the-shelf results at this stage.

Radial estimates were introduced by Melrose [1994] for the study of asymptotically Euclidean scattering and have been developed further in various settings. We only mention some of the more relevant ones: scattering by zeroth-order potentials (very close in spirit to the problems considered in [Colin de Verdière and Saint-Raymond 2019]) by Hassell, Melrose, and Vasy [Hassell et al. 2004], asymptotically hyperbolic scattering by Vasy [2013] (see also [Dyatlov and Zworski 2016, Chapter 5] and [Zworski 2016]) and by Datchev and Dyatlov [2013], in general relativity by Vasy [2013], Dyatlov [2012] and Hintz and Vasy [2018], and in hyperbolic dynamics by Dyatlov and Zworski [2016]. Particularly useful here is [Haber and Vasy 2015], which generalized some of the results of [Hassell et al. 2004]. A very general version of radial estimates is presented "textbook style" in Section E.4 of [Dyatlov and Zworski 2019], henceforth abbreviated [DZ19].

MSC2010: 35A27.
Keywords: forced waves, spectral theory, pseudodifferential operators, radial estimates.

1A. ***The main result.*** Motivated by internal waves in linearized fluids Colin de Verdière and Saint-Raymond [2019] considered long-time behavior of solutions to

$$\begin{gathered}(i\partial_t - P)u(t) = f, \quad u(0) = 0, \quad f \in C^\infty(M),\\ P \in \Psi^0(M), \quad P = P^*,\end{gathered} \tag{1}$$

where M is a closed surface and P satisfies dynamical assumptions presented in Section 1B. By changing P to $P - \omega_0$ we can change f to the more physically relevant oscillatory forcing term, $e^{-i\omega_0 t} f$.

Since the solution $u(t)$ is given by

$$u(t) = -i \int_0^t e^{-isP} f \, ds = P^{-1}(e^{-itP} - 1) f \tag{2}$$

(where the operator $P^{-1}(e^{-itP} - 1)$ is well-defined for all t using the spectral theorem), the properties of the spectrum of P play a crucial role in the description of the long-time behavior of $u(t)$. Referring to Section 1B for the precise assumptions we state:

Theorem. *Suppose that the operator P satisfies assumptions* (5), (8) *below and that $0 \notin \mathrm{Spec}_{\mathrm{pp}}(P)$. Then, for any $f \in C^\infty(M)$, the solution to* (1) *satisfies*

$$u(t) = u_\infty + b(t) + \epsilon(t), \quad \|b(t)\|_{L^2} \le C, \quad \|\epsilon(t)\|_{H^{-1/2-}} \to 0, \quad t \to \infty, \tag{3}$$

where (denoting by $H^{-\frac{1}{2}-}$ the intersection of the spaces $H^{-\frac{1}{2}-\epsilon}$ over $\epsilon > 0$)

$$u_\infty \in I^0(M; \Lambda_0^+) \subset H^{-\frac{1}{2}-}(M) \tag{4}$$

and $I^0(M; \Lambda_0^+)$ is the space of Lagrangian distributions of order 0 *(see Section 4A) associated to the attracting Lagrangian Λ_0^+ defined in* (9)*.*

The proof gives other results obtained in [Colin de Verdière and Saint-Raymond 2019]. In particular, we see that in the neighborhood of 0 the spectrum of P is absolutely continuous except for finitely many eigenvalues with smooth eigenfunctions — see Section 3B.

In the case of general Morse–Smale flows (allowing for fixed points), Colin de Verdière [2018, Theorem 4.3] used a hybrid of Mourre estimates (in particular their finer version given by Jensen, Mourre, and Perry [Jensen et al. 1984]) and of the radial estimates [DZ19, §E.4] to obtain a version of (3) with an estimate on $\mathrm{WF}(u_\infty)$. At this stage the purely microlocal approach of this paper would only give $\|\epsilon(t)\|_{H^{-3/2-}} \to 0$.

1B. ***Assumptions on P.*** We assume that M is a compact surface without boundary and $P \in \Psi^0(M)$ is a zeroth-order pseudodifferential operator with principal symbol $p \in S^0(T^*M \setminus 0; \mathbb{R})$ which is homogeneous (of order 0) and has 0 as a regular value. We also assume that for some smooth density, $dm(x)$, on M, P is self-adjoint:

$$\begin{gathered}P \in \Psi^0(M), \qquad P = P^* \quad \text{on } L^2(M, dm(x)),\\ p := \sigma(P), \qquad p(x, t\xi) = p(x, \xi), \quad t > 0, \qquad dp|_{p^{-1}(0)} \ne 0.\end{gathered} \tag{5}$$

The homogeneity assumption on p can be removed as the results of [DZ19, §E.4] and [Dyatlov and Zworski 2017] we use do not require it. That would however complicate the statement of the dynamical assumptions.

We use the notation of [DZ19, §E.1.3], denoting by $\overline{T}^*M$ the fiber-radially compactified cotangent bundle. Define the quotient map for the $\mathbb{R}^+$ action, $(x,\xi)\mapsto(x,t\xi)$, $t>0$,

$$\kappa : \overline{T}^*M \setminus 0 \to \partial\overline{T}^*M. \tag{6}$$

Denote by $|\xi|$ the norm of a covector $\xi\in T^*_xM$ with respect to some fixed Riemannian metric on M. The rescaled Hamiltonian vector field $|\xi|H_p$ commutes with the $\mathbb{R}^+$ action and

$$X := \kappa_*(|\xi|H_p) \quad \text{is tangent to} \quad \Sigma := \kappa(p^{-1}(0)). \tag{7}$$

Note that Σ is an orientable surface since it is defined by the equation $p=0$ in the orientable 3-manifold $\partial\overline{T}^*M$.

We now recall the dynamical assumption made in [Colin de Verdière and Saint-Raymond 2019]:

$$\text{The flow of } X \text{ on } \Sigma \text{ is a Morse–Smale flow with } \textit{no} \text{ fixed points.} \tag{8}$$

For the reader's convenience we recall the definition of Morse–Smale flows generated by X on a surface Σ (see [Nikolaev and Zhuzhoma 1999, Definition 5.1.1]):

(1) X has a finite number of fixed points, all of which are hyperbolic.

(2) X has a finite number of hyperbolic limit cycles.

(3) There are no separatrix connections between saddle fixed points.

(4) Every trajectory different from (1) and (2) has unique trajectories (1) or (2) as its α, ω-limit sets.

As stressed in [Colin de Verdière and Saint-Raymond 2019], Morse–Smale flows enjoy stability and genericity properties — see [Nikolaev and Zhuzhoma 1999, Theorem 5.1.1]. At this stage, following [Colin de Verdière and Saint-Raymond 2019], we make the strong assumption that there are no fixed points. By the Poincaré–Hopf theorem, that forces Σ to be a union of tori. Under the assumption (8), the flow of X on Σ has an attractor L_0^+, which is a union of closed attracting curves. We define the following *conic Lagrangian submanifold* of $T^*M\setminus 0$ (see [Hörmander 1985a, §21.2] and Lemma 2.1):

$$\Lambda_0^+ := \kappa^{-1}(L_0^+). \tag{9}$$

1C. ***Examples.*** We illustrate the result with two simple examples on $M := \mathbb{T}^2 = \mathbb{S}^1\times\mathbb{S}^1$, where $\mathbb{S}^1 = \mathbb{R}/(2\pi\mathbb{Z})$. Define $D := (1/i)\partial$. Consider first

$$\begin{gathered} P := \langle D\rangle^{-1} D_{x_2} - 2\cos x_1, \quad p = |\xi|^{-1}\xi_2 - 2\cos x_1, \\ |\xi|H_p = -\frac{\xi_1\xi_2}{|\xi|^2}\partial_{x_1} + \frac{\xi_1^2}{|\xi|^2}\partial_{x_2} - 2(\sin x_1)|\xi|\partial_{\xi_1}, \\ \Lambda_0^+ = \big\{\big(\pm\tfrac{\pi}{2}, x_2; \xi_1, 0\big) : x_2\in\mathbb{S}^1,\ \pm\xi_1 < 0\big\}. \end{gathered} \tag{10}$$

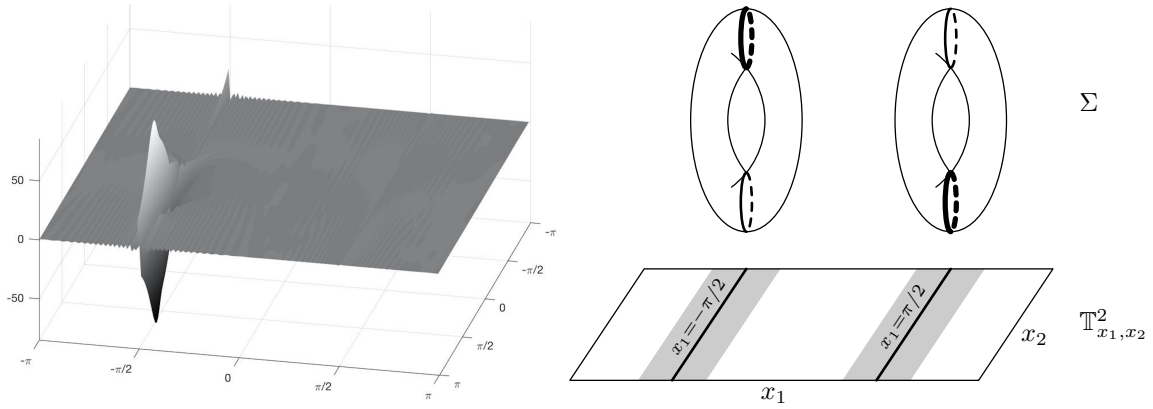

Figure 1. On the left: the plot of the real part of $u(50)$ for $P = \langle D\rangle^{-1} D_{x_2} + 2\cos x_1$ on $\mathbb{T}^2$ and f given by a smooth bump function centered at $\left(-\frac{\pi}{2}, 0\right)$. We see the singularity formation on the line $x_1 = -\frac{\pi}{2}$. On the right: $\Sigma := \kappa(p^{-1}(0)) \subset \partial\overline{T}^*\mathbb{T}^2$. The attracting Lagrangian, Λ_0^+, comes from the highlighted circles. See Section 1C for a discussion of the examples shown in the figures.

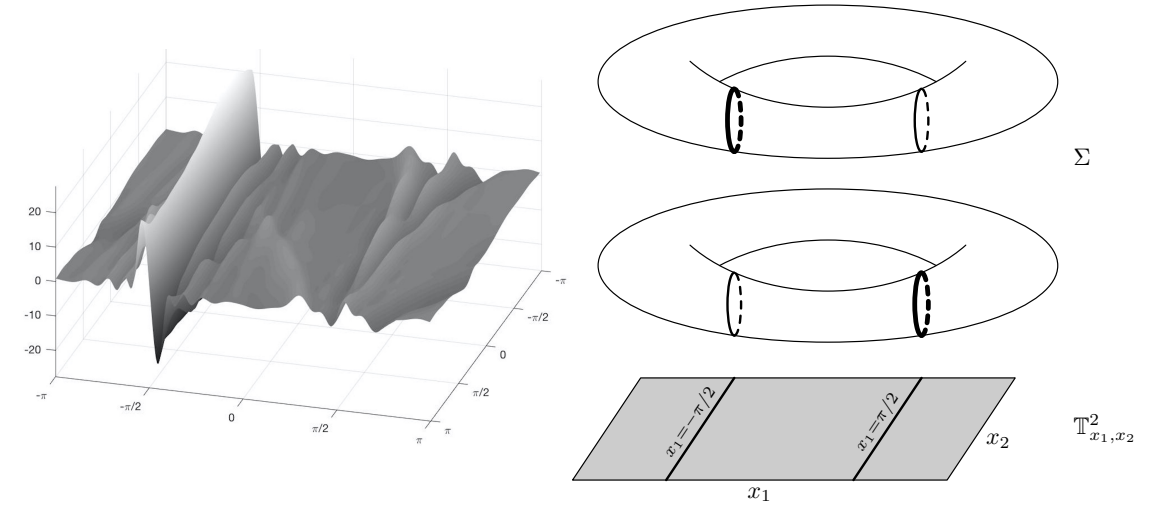

Figure 2. On the left: the plot of the real part of $u(50)$ for P given by (11) and f given by a smooth bump function centered at $\left(-\frac{\pi}{2}, 0\right)$. We see the singularity formation on the line $x_1 = -\frac{\pi}{2}$ and the slower formation of singularity at $x_1 = \frac{\pi}{2}$. On the right: $\Sigma := \kappa(p^{-1}(0))$. The attracting Lagrangian Λ_0^+ comes from the highlighted circles.

In this case $\kappa(p^{-1}(0))$, with κ given in (6), is a union of two tori which do *not* cover $\mathbb{T}^2$ (and thus does not satisfy the assumptions of [Colin de Verdière and Saint-Raymond 2019] but is covered by the treatment here, and in [Colin de Verdière 2018]). See Figure 1 for the plot of $\Re u(t)$, $t = 50$, and for a schematic visualization of $\Sigma = \kappa(p^{-1}(0))$.

Our result applies also to the closely related operator

$$P := \langle D\rangle^{-1} D_{x_2} - \tfrac{1}{2}\cos x_1, \quad p = |\xi|^{-1}\xi_2 - \tfrac{1}{2}\cos x_1,$$
$$|\xi| H_p = -\frac{\xi_1\xi_2}{|\xi|^2}\partial_{x_1} + \frac{\xi_1^2}{|\xi|^2}\partial_{x_2} - \tfrac{1}{2}\sin x_1 |\xi|\partial_{\xi_1}. \tag{11}$$

The attracting Lagrangians are the same but the energy surface $\kappa(p^{-1}(0))$ consists of two tori covering $\mathbb{T}^2$ (and hence satisfying the assumptions of [Colin de Verdière and Saint-Raymond 2019]) — see Figure 2.

2. Geometric structure of attracting Lagrangians

In this section we prove geometric properties of the attracting and repulsive Lagrangians for the flow $e^{t|\xi|H_p}$ where p satisfies (8).

2A. ***Sink and source structure.*** Let $\Sigma(\omega) := \kappa(p^{-1}(\omega))$. If $\delta > 0$ is sufficiently small then stability of Morse–Smale flows (and the stability of nonvanishing of X) shows that (8) is satisfied for $\Sigma(\omega)$, $|\omega| \le 2\delta$. Let $L_\omega^\pm \subset \Sigma(\omega)$ be the attractive $(+)$ and repulsive $(-)$ hyperbolic cycles for the flow of X on $\Sigma(\omega)$. We first establish dynamical properties needed for the application of radial estimates in Section 3:

Lemma 2.1. *L_ω^+ is a radial sink and L_ω^- a radial source for the Hamiltonian flow of $|\xi|(p-\omega) = |\xi|\sigma(P-\omega)$ in the sense of [DZ19, Definition E.50]. The conic submanifolds*

$$\Lambda_\omega^\pm := \kappa^{-1}(L_\omega^\pm) \subset T^*M \setminus 0$$

are Lagrangian.

Remark. It is not true that $L_\omega^\pm$ are radial sinks/sources for the Hamiltonian flow of $p-\omega$ since [DZ19, Definition E.50] requires convergence of all nearby Hamiltonian trajectories, not just those on the characteristic set $p^{-1}(\omega)$. See Remark 3 following [DZ19, Definition E.50] for details. The singular behavior of $|\xi|$ at $\xi = 0$ is irrelevant here since we are considering a neighborhood of the fiber infinity.

Proof. We consider the case of L_ω^+ as that of L_ω^- is similar. To simplify the formulas below we put $\omega := 0$. To see that Λ_0^+ is a Lagrangian submanifold we note that H_p and $\xi\partial_\xi$ are tangent to Λ_0^+ and independent (since X does not vanish on L_0^+). Denoting the symplectic form by σ, we have $\sigma(H_p, \xi\partial_\xi) = -dp(\xi\partial_\xi) = 0$; that is, σ vanishes on the tangent space to Λ_0^+.

We next show that L_0^+ is a radial sink. For simplicity assume that it consists of a single attractive closed trajectory of X of period $T > 0$; in particular $e^{TX} = I$ on L_0^+. Define the vector field

$$Y := H_{|\xi|p},$$

which is homogeneous of order 0 on $T^*M \setminus 0$ and thus extends smoothly to the fiber-radial compactification $\overline{T}^*M \setminus 0$; see [DZ19, Proposition E.5]. We have $Y = X$ on $\partial\overline{T}^*M \cap p^{-1}(0)$; thus $L_0^+ \subset \partial\overline{T}^*M$ is a closed trajectory of Y of period T.

Fix arbitrary $(x_0, \xi_0) \in L_0^+$ and define the linearized Poincaré map $\mathcal{P}$ induced by $de^{TY}(x_0, \xi_0)$ on the quotient space $T_{(x_0,\xi_0)}(\overline{T}^*M)/\mathbb{R}Y_{(x_0,\xi_0)}$. The adjoint map $\mathcal{P}^*$ acts on covectors in $T^*_{(x_0,\xi_0)}(\overline{T}^*M)$

which annihilate $Y_{(x_0,\xi_0)}$. To prove that L_0^+ is a radial sink it suffices to show that the spectral radius of $\mathcal{P}$ is strictly less than 1.

Put $\rho := |\xi|^{-1}$, which is a boundary-defining function on $\overline{T}^*M$; then $\Sigma = \partial\overline{T}^*M \cap p^{-1}(0)$ is given by $\{p = 0,\ \rho = 0\}$. Since $Y = X$ on Σ and L_0^+ is an attractive cycle for X on Σ, we have

$$\mathcal{P}|_{\ker(dp)\cap\ker(d\rho)} = c_1 \quad \text{for some } c_1 \in \mathbb{R},\ \ |c_1| < 1.$$

Since Y is tangent to $\partial\overline{T}^*M = \rho^{-1}(0)$, we have $Y\rho = f_2\rho$ for some $f_2 \in C^\infty(\overline{T}^*M \setminus 0; \mathbb{R})$. Recalling that $Y = H_{|\xi|p}$, we compute $Yp = pH_{|\xi|}p = -pH_p(\rho^{-1}) = f_2 p$. Setting $c_2 := f_2(x_0, \xi_0)$ we then have

$$\mathcal{P}^*(dp(x_0,\xi_0)) = c_2 dp(x_0,\xi_0), \quad \mathcal{P}^*(d\rho(x_0,\xi_0)) = c_2 d\rho(x_0,\xi_0).$$

Thus $\mathcal{P}$ has eigenvalues c_1, c_2, c_2. On the other hand, e^{TY} preserves the symplectic density $|\sigma\wedge\sigma|$, which has the form $\rho^{-3}d$ vol for some density d vol on $\overline{T}^*M$ which is smooth up to the boundary. Taking the limit of this statement at (x_0,ξ_0) we obtain $\det\mathcal{P} = \det de^{TY}(x_0,\xi_0) = c_2^3$. It follows that $c_1 = c_2$ and thus $\mathcal{P}$ has spectral radius $|c_1| < 1$ as needed. □

For future use we define the conic hypersurfaces in $T^*M \setminus 0$

$$\Lambda^\pm := \bigcup_{|\omega|<2\delta} \Lambda_\omega^\pm. \tag{12}$$

2B. ***Geometry of Lagrangian families.*** We next establish some facts about families of Lagrangian submanifolds which do not need the dynamical assumptions (8). Instead we assume that

- $p : T^*M \setminus 0 \to \mathbb{R}$ is homogeneous of order 0;
- $\Lambda \subset T^*M \setminus 0$ is a conic hypersurface;
- $dp|_{T\Lambda} \neq 0$ everywhere;
- the Hamiltonian vector field H_p is tangent to Λ.

Under these assumptions, the sets

$$\Lambda_\omega := \Lambda \cap p^{-1}(\omega)$$

are two-dimensional conic submanifolds of $T^*M \setminus 0$. Moreover, similarly to Lemma 2.1, each Λ_ω is Lagrangian. Indeed, if G is a (local) defining function of Λ, namely $G|_\Lambda = 0$ and $dG|_\Lambda \neq 0$, then H_p being tangent to Λ implies

$$\{p, G\} = 0 \quad \text{on } \Lambda. \tag{13}$$

Thus H_p, H_G form a tangent frame on Λ_ω and $\sigma(H_p, H_G) = 0$ on Λ, where σ denotes the symplectic form.

Since $\xi\partial_\xi$ is tangent to each Λ_ω, for any choice of local defining function G of Λ we can write

$$\xi\partial_\xi = \Phi H_p + \Theta H_G \quad \text{on } \Lambda \tag{14}$$

for some functions Φ, Θ on Λ. Since the one-dimensional subbundle $\mathbb{R}H_G \subset T\Lambda$ is invariantly defined, we see that $\Phi \in C^\infty(\Lambda; \mathbb{R})$ does not depend on the choice of G.

The function Φ is homogeneous of order 1. Indeed, we can choose G to be homogeneous of order 1, which implies that $[\xi\partial_\xi, H_G] = 0$; we also have $[\xi\partial_\xi, H_p] = -H_p$. By taking the commutator of both sides of (14) with $\xi\partial_\xi$, we see that $\xi\partial_\xi\Phi = \Phi$. Similarly we see that Θ is homogeneous of order 0.

On the other hand, taking the commutators of both sides of (14) with H_p and H_G and using the following consequence of (13),

$$[H_p, H_G] = H_{\{p,G\}} \in \mathbb{R}H_G \quad \text{on } \Lambda,$$

we get the identities

$$H_p\Phi \equiv 1, \quad H_G\Phi \equiv 0 \quad \text{on } \Lambda. \tag{15}$$

The function Φ is related to the ω-derivative of a generating function of Λ_ω (see (45)):

Lemma 2.2. *Assume that Λ_ω is locally given (in some coordinate system on M) by*

$$\Lambda_\omega = \{(x,\xi)\colon x = \partial_\xi F(\omega,\xi),\ \xi \in \Gamma_0\}, \tag{16}$$

where $\xi \mapsto F(\omega,\xi)$ is a family of homogeneous functions of order 1 and $\Gamma_0 \subset \mathbb{R}^2 \setminus 0$ is a cone. Then we have

$$\partial_\omega F(\omega,\xi) = -\Phi(\partial_\xi F(\omega,\xi),\xi). \tag{17}$$

Proof. Let G be a (local) defining function of Λ. Taking the ∂_ξ-component of (14) at a point $\zeta := (\partial_\xi F(\omega,\xi),\xi) \in \Lambda$ we have

$$\xi = -\Phi(\zeta)\,\partial_x p(\zeta) - \Theta(\zeta)\,\partial_x G(\zeta). \tag{18}$$

On the other hand, differentiating in ω the identities

$$p(\partial_\xi F(\omega,\xi),\xi) = \omega, \quad G(\partial_\xi F(\omega,\xi),\xi) = 0$$

we get

$$\langle \partial_x p(\zeta), \partial_\xi\partial_\omega F(\omega,\xi)\rangle = 1, \quad \langle \partial_x G(\zeta), \partial_\xi\partial_\omega F(\omega,\xi)\rangle = 0. \tag{19}$$

Combining (18) and (19) we arrive at

$$\langle \xi, \partial_\xi\partial_\omega F(\omega,\xi)\rangle = -\Phi(\zeta) = -\Phi(\partial_\xi F(\omega,\xi),\xi),$$

which implies (17) since the function $\xi \mapsto \partial_\omega F(\omega,\xi)$ is homogeneous of order 1. □

Now we specialize to the Lagrangian families used in this paper. We start with a sign condition on Φ which will be used in Section 5:

Lemma 2.3. *Suppose that for $\Lambda = \Lambda^+$ or $\Lambda = \Lambda^-$, with $\Lambda^\pm$ given in* (12), *we define $\Phi^\pm$ using* (14). *Then for some constant $c > 0$,*

$$\pm\Phi^\pm(x,\xi) \geq c|\xi| \quad \text{on } \Lambda^\pm. \tag{20}$$

Proof. We consider the case of Φ^+ as the case of Φ^- is handled by replacing p with $-p$. Recall from Lemma 2.1 that each $L^+_\omega = \kappa(\Lambda^+ \cap p^{-1}(\omega))$ is a radial sink for the flow $e^{t|\xi|H_p}$. Take $(x,\xi) \in \Lambda^+$ with

$|\xi|$ large. Then (with S^*M denoting the cosphere bundle with respect to any fixed metric on M)

$$e^{-tH_p}(x,\xi) \in S^*M \quad \text{for some } t>0, \ \ t\sim|\xi|. \tag{21}$$

Recall from (15) that $H_p\Phi^+=1$ on Λ^+. Thus

$$\Phi^+(x,\xi)=\Phi^+(e^{-tH_p}(x,\xi))+t\geq c|\xi|-C.$$

It follows that $\Phi^+(x,\xi)\geq c|\xi|$ for large $|\xi|$; since Φ^+ is homogeneous of order 1, this inequality then holds on the entire Λ^+. □

We next construct adapted global defining functions of $\Lambda^\pm$ used in Section 4B:

Lemma 2.4. *Let $\Lambda^\pm$ be defined in* (12). *Then there exist $G_\pm\in C^\infty(T^*M\setminus 0;\mathbb{R})$ such that*

(1) *$G_\pm$ are homogeneous of order* 1;

(2) *$G_\pm|_{\Lambda^\pm}=0$ and $dG_\pm|_{\Lambda^\pm}\neq 0$;*

(3) *$H_pG_\pm=a_\pm G_\pm$ in a neighborhood of $\Lambda^\pm$, where $a_\pm\in C^\infty(T^*M\setminus 0;\mathbb{R})$ are homogeneous of order -1 and $a_\pm|_{\Lambda^\pm}=0$.*

Proof. We construct G_+, with G_- constructed similarly. Fix some function $\widetilde{G}_+$ which satisfies conditions (1) and (2) of the present lemma. It exists since Λ^+ is conic and orientable (each of its connected components is diffeomorphic to $[-\delta,\delta]\times\mathbb{S}^1\times\mathbb{R}^+$). Let Θ_+ be defined in (14):

$$\xi\partial_\xi=\Phi_+H_p+\Theta_+H_{\widetilde{G}_+} \quad \text{on } \Lambda^+. \tag{22}$$

Commuting both sides of (14) with $\xi\partial_\xi$ we see that Θ_+ is homogeneous of order 0. Moreover Θ_+ does not vanish on Λ^+ since H_p is not radial (since the flow of X in (7) has no fixed points). Choose G_+ satisfying conditions (1) and (2) and such that

$$G_+=\Theta_+\widetilde{G}_+ \quad \text{near } \Lambda^+.$$

Then (22) gives

$$\xi\partial_\xi=\Phi_+H_p+H_{G_+} \quad \text{on } \Lambda^+. \tag{23}$$

We have $H_pG_+|_{\Lambda^+}=0$ (since H_p is tangent to Λ^+); therefore $H_pG_+=a_+G_+$ near Λ^+ for some function a_+. Commuting both sides of (23) with H_p and using that $H_p\Phi_+\equiv 1$ on Λ^+ from (15) we have

$$H_p=[H_p,\xi\partial_\xi]=H_p+[H_p,H_{G_+}]=H_p+H_{\{p,G_+\}}=H_p+a_+H_{G_+} \quad \text{on } \Lambda^+.$$

Since H_{G_+} does not vanish on Λ^+, this gives $a_+|_{\Lambda^+}=0$ as needed. □

One application of Lemma 2.4 is the existence of an H_p-invariant density on $\Lambda^\pm$:

Lemma 2.5. *There exist densities $\nu_\omega^\pm$ on $\Lambda^\pm_\omega$, $\omega\in[-\delta,\delta]$, such that*

- *$\nu_\omega^\pm$ are homogeneous of order* 1, *that is, $\mathcal{L}_{\xi\partial_\xi}\nu_\omega^\pm=\nu_\omega^\pm$;*
- *$\nu_\omega^\pm$ are invariant under H_p, that is, $\mathcal{L}_{H_p}\nu_\omega^\pm=0$.*

Proof. In the notation of Lemma 2.4 define $\nu_\omega^\pm$ by $|\sigma\wedge\sigma| = |dp\wedge dG_\pm|\times\nu_\omega^\pm$, where σ is the symplectic form. The properties of $\nu_\omega^\pm$ follow from the identities

$$\mathcal{L}_{\xi\partial_\xi}\sigma = \sigma, \quad \mathcal{L}_{\xi\partial_\xi} dp = 0, \quad \mathcal{L}_{\xi\partial_\xi} dG_\pm = dG_\pm, \quad \mathcal{L}_{H_p}\sigma = 0$$

and the following statement which holds on $\Lambda^\pm$:

$$\mathcal{L}_{H_p}(dp\wedge dG_\pm) = dp\wedge d(a_\pm G_\pm) = 0. \qquad \square$$

3. Resolvent estimates

Here we recall the radial estimates as presented in [DZ19, §E.4] specializing to the setting of Section 1B. We use the notation of [DZ19, Appendix E] and we write $\|u\|_s := \|u\|_{H^s(M)}$.

Since we are not in the semiclassical setting of [DZ19, §E.4] we will only use the usual notion of the wave front set: for $u\in\mathscr{D}'(M)$, $\mathrm{WF}(u)\subset T^*M\setminus 0$ — see [DZ19, Exercise E.16]. Similarly, for $A\in\Psi^k(M)$ we denote by $\mathrm{ell}(A)\subset T^*M\setminus 0$ its (nonsemiclassical) elliptic set. Both sets are conic.

3A. ***Radial estimates uniformly up to the real axis.*** Since L_ω^- is a radial source we can apply [DZ19, Theorem E.52] (with $h:=1$) to the operator

$$\widetilde{P}_\epsilon := \widetilde{P} - i\epsilon\langle D\rangle \in \Psi^1(M), \quad \widetilde{P} := \langle D\rangle^{\frac12}(P-\omega)\langle D\rangle^{\frac12}, \quad 0\le\epsilon\ll 1.$$

Here, since $\widetilde{P}$ is self-adjoint, the threshold regularity condition [DZ19, (E.4.39)] is satisfied for $\widetilde{P}$ with any $s>0$. Strictly speaking, one has to modify the proof of [DZ19, Theorem E.52] to include the anti-self-adjoint part $-i\epsilon\langle D\rangle$, which has a favorable sign but is of the same differential order as $\widetilde{P}$. (In [loc. cit.] it was assumed that the principal symbol of P is real-valued near L_ω^-.) More precisely, we put $\boldsymbol{P} := \widetilde{P}$ and $f := \widetilde{P}_\epsilon u$ (instead of $f := \widetilde{P}u$) in [DZ19, Theorem E.52]. Since $\widetilde{P}_\epsilon$ satisfies the sign condition for propagation of singularities [DZ19, Theorem E.47], it suffices to check that the positive commutator estimate [DZ19, Lemma E.49] holds. For that we write

$$\Im\langle f, G^*Gu\rangle_{L^2} = \Im\langle \widetilde{P}u, G^*Gu\rangle_{L^2} - \epsilon\Re\langle\langle D\rangle u, G^*Gu\rangle_{L^2}. \tag{24}$$

Here $G\in\Psi^s(M)$ is the quantization of an escape function used in the proof of [DZ19, Lemma E.49]; recall that we put $h:=1$. We now estimate the additional term in (24):

$$\begin{aligned} -\Re\langle\langle D\rangle u, G^*Gu\rangle_{L^2} &= -\|\langle D\rangle^{\frac12}Gu\|_{L^2}^2 + \langle\Re(G^*[\langle D\rangle, G])u, u\rangle_{L^2} \\ &\le C\|B_1 u\|_{s-1/2}^2 + C\|u\|_{H^{-N}}^2, \end{aligned}$$

where B_1 satisfies the properties in the statement of [DZ19, Lemma E.49] and in the last line we used that $G^*[\langle D\rangle, G]\in\Psi^{2s}(M)$ has purely imaginary principal symbol and thus $\Re(G^*[\langle D\rangle, G])\in\Psi^{2s-1}(M)$. The rest of the proof of [DZ19, Lemma E.49] applies without changes. See also [Dyatlov and Guillarmou 2016, Lemma 3.7].

Applying the radial estimate in [DZ19, Theorem E.52] for the operator $\widetilde{P}_\epsilon = \langle D\rangle^{\frac12}(P-\omega-i\epsilon)\langle D\rangle^{\frac12}$ to $\langle D\rangle^{-\frac12}u$ we see that for every $\widetilde{B}_-\in\Psi^0(M)$, $\Lambda^-\subset\mathrm{ell}(\widetilde{B}_-)$, there exists $A_-\in\Psi^0(M)$, $\Lambda^-\subset\mathrm{ell}(A_-)$,

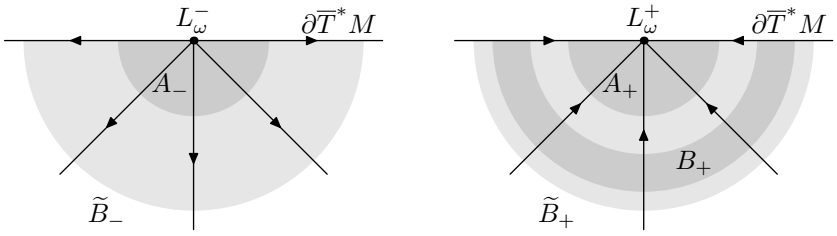

Figure 3. An illustration of the supports of the operators appearing in (25) (left: radial sources) and (26) (right: radial sinks). The horizontal line on the top denotes $\partial\overline{T}^*M$; the arrows denote flow lines of $|\xi|H_p$.

such that

$$\begin{gathered}\|A_-u\|_s \le C\|\widetilde{B}_-(P-\omega-i\epsilon)u\|_{s+1} + C\|u\|_{-N},\\ u\in C^\infty(M),\quad s>-\tfrac12,\quad |\omega|\le\delta,\quad \epsilon\ge 0,\end{gathered}\tag{25}$$

where C does not depend on ϵ,ω and N can be chosen arbitrarily large. The supports of A_-, $\widetilde{B}_-$ are shown in Figure 3.

The inequality (25) can be extended to a larger class of distributions (as opposed to $u\in C^\infty(M)$): it suffices that $\widetilde{B}_-(P-\omega-i\epsilon)u\in H^{s+1}(M)$ and that $A_-u\in H^{s'}(M)$ for some $s'>-\frac12$. See Remark 5 after [DZ19, Theorem E.52] or [Dyatlov and Zworski 2016, Proposition 2.6; Vasy 2013, Proposition 2.3].

Similarly we have estimates near radial sinks [DZ19, Theorem E.54] for L_ω^+. Namely, for every $\widetilde{B}_+\in\Psi^0(M)$, $\Lambda^+\subset\operatorname{ell}(\widetilde{B}_+)$, there exist $A_+, B_+\in\Psi^0(M)$, such that $\Lambda^+\subset\operatorname{ell}(A_+)$, $\operatorname{WF}(B_+)\cap\Lambda^+=\varnothing$, and

$$\begin{gathered}\|A_+u\|_s \le C\|\widetilde{B}_+(P-\omega-i\epsilon)u\|_{s+1} + C\|B_+u\|_s + C\|u\|_{-N},\\ u\in C^\infty(M),\quad s<-\tfrac12,\quad |\omega|\le\delta,\quad \epsilon\ge 0,\end{gathered}\tag{26}$$

where C does not depend on ϵ,ω and N can be chosen arbitrarily large. The inequality is also valid for distributions u such that $\widetilde{B}_+(P-\omega-i\epsilon)u\in H^{s+1}(M)$ and $B_+u\in H^s(M)$ and it then provides (unconditionally) $A_+u\in H^s(M)$ — see Remark 2 after [DZ19, Theorem E.54] or [Dyatlov and Zworski 2016, Proposition 2.7; Vasy 2013, Proposition 2.4].

Away from radial points we have the now standard propagation results of Duistermaat and Hörmander [DZ19, Theorem E.47]: if $A, B, \widetilde{B}\in\Psi^0(M)$ and for each $(x,\xi)\in\operatorname{WF}(A)$ there exists $T\ge 0$ such that

$$e^{-T|\xi|H_p}(x,\xi)\in\operatorname{ell}(B),\quad e^{-t|\xi|H_p}(x,\xi)\in\operatorname{ell}(\widetilde{B}),\quad 0\le t\le T,$$

then

$$\begin{gathered}\|Au\|_s \le C\|\widetilde{B}(P-\omega-i\epsilon)u\|_{s+1} + C\|Bu\|_s + C\|u\|_{-N},\\ u\in C^\infty(M),\quad s\in\mathbb{R},\quad |\omega|\le\delta,\quad \epsilon\ge 0,\end{gathered}\tag{27}$$

with C independent of ϵ,ω. We also have the elliptic estimate [DZ19, Theorem E.33]: (27) holds with $B=0$ if $\operatorname{WF}(A)\cap p^{-1}([-\delta,\delta])=\varnothing$ and $\operatorname{WF}(A)\subset\operatorname{ell}(\widetilde{B})$.

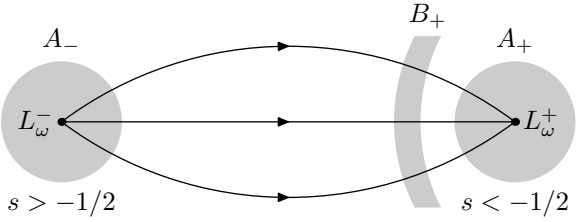

Figure 4. A schematic representation of the flow $e^{t|\xi|H_p}$ on the fiber infinity $\partial\overline{T}^*M$ intersected with the energy surface $p^{-1}(\omega)$, with the regularity thresholds for the estimates (25) and (26).

Let us now consider

$$u_\epsilon = u_\epsilon(\omega) := (P - \omega - i\epsilon)^{-1} f, \quad f \in C^\infty(M), \quad |\omega| \leq \delta, \quad \epsilon > 0.$$

For any fixed $\epsilon > 0$, $P - \omega - i\epsilon \in \Psi^0(M)$ is an elliptic operator (its principal symbol equals $p - \omega - i\epsilon$ and p is real-valued); thus by elliptic regularity $u_\epsilon \in C^\infty(M)$. Combining (25), (26) and (27) we see that for any $\beta > 0$

$$\|u_\epsilon\|_{-1/2-\beta} \leq C\|f\|_{1/2+\beta} + C\|u_\epsilon\|_{-N}, \tag{28}$$

and that

$$\|Au_\epsilon\|_s \leq C\|f\|_{s+1} + C\|u_\epsilon\|_{-N}, \quad \mathrm{WF}(A) \cap \Lambda^+ = \varnothing, \quad s > -\tfrac{1}{2}. \tag{29}$$

Here the constant C depends on β, s but does not depend on ϵ, ω. Indeed, by our dynamical assumption (8) every trajectory $e^{t|\xi|H_p}(x,\xi)$ with $(x,\xi) \in p^{-1}([-\delta,\delta]) \setminus \Lambda^+$ converges to Λ^- as $t \to -\infty$ (see Figure 4). Applying (27) with $B := A_-$ and using (25) we get (29). Putting $A := B_+$ in (29) and using (26) we get (28).

In particular, we obtain a regularity statement for the limits of the family (u_ϵ):

$$\text{there exist } \epsilon_j \to 0,\ u \in \mathscr{D}'(M) \text{ such that } u_{\epsilon_j} \xrightarrow{\mathscr{D}'(M)} u \quad \Longrightarrow \quad u \in H^{-\frac{1}{2}-}(M), \quad \mathrm{WF}(u) \subset \Lambda^+. \tag{30}$$

Note also that every u in (30) solves the equation $(P - \omega)u = f$.

3B. ***Regularity of eigenfunctions.*** Motivated by (30) we have the following regularity statement. The proof is an immediate modification of the proof of [Dyatlov and Zworski 2017, Lemma 2.3]: replace P there by $A^{-1}(P - \omega)A^{-1}$, where $A \in \Psi^{-\frac{1}{2}}(M)$ is elliptic, self-adjoint on $L^2(M, dm(x))$ (same density with respect to which P is self-adjoint) and invertible. We record this as:

Lemma 3.1. *Suppose that P satisfies* (5) *and* (8). *Then for ω sufficiently small and for $u \in \mathscr{D}'(M)$*

$$(P - \omega)u \in C^\infty, \quad \mathrm{WF}(u) \subset \Lambda^+, \quad \Im\langle (P - \omega)u, u\rangle \geq 0, \quad |\omega| \leq \delta,$$

implies that $u \in C^\infty(M)$.

In particular this shows that if $(P-\omega)u = 0$ and $\mathrm{WF}(u) \subset \Lambda^+$ then $u \in L^2$; that is, ω lies in the point spectrum $\mathrm{Spec}_{\mathrm{pp}}(P)$. Radial estimates then show that the number of such ω's is finite in a neighborhood of 0:

Lemma 3.2. *Under the assumptions* (5) *and* (8), *with δ sufficiently small*,

$$\begin{gathered} |\operatorname{Spec}_{\mathrm{pp}}(P)\cap[-\delta,\delta]|<\infty,\\ (P-\omega)u=0,\quad u\in L^2(M),\quad |\omega|\le\delta \quad\Longrightarrow\quad u\in C^\infty(M). \end{gathered} \tag{31}$$

Proof. If $u\in L^2(M)$ then the threshold assumption in (25) is satisfied for $P-\omega$ near Λ^- and for $-(P-\omega)$ near Λ^+. Using the remark about regularity after (25), as well as (27) away from sinks and sources, we conclude that

$$\|u\|_s\le C\|u\|_{-N} \tag{32}$$

for any s and N. That implies that $u\in C^\infty(M)$. Now, suppose that there exists an infinite set of L^2 eigenfunctions with eigenvalues in $[-\delta,\delta]$:

$$(P-\omega_j)u_j=0,\quad \langle u_k,u_j\rangle_{L^2(M)}=\delta_{kj},\quad |\omega_j|\le\delta.$$

Since $u_j\rightharpoonup 0$, weakly in L^2, we have $u_j\to 0$ strongly in H^{-1}. But this contradicts (32) applied with $s=0$ and $N=1$. □

From now on we make the assumption that P has no eigenvalues in $[-\delta,\delta]$:

$$\operatorname{Spec}_{\mathrm{pp}}(P)\cap[-\delta,\delta]=\varnothing. \tag{33}$$

By Lemma 3.2 we see that (33) holds for δ small enough as long as $0\notin\operatorname{Spec}_{\mathrm{pp}}(P)$.

3C. ***Limiting absorption principle.*** Using results of Sections 3A–3B we obtain a version of the limiting absorption principle sufficient for proving (3). Radial estimates can also easily give existence of $(P-\omega-i0)^{-1}: H^{\frac12+}(M)\to H^{-\frac12-}(M)$ but we restrict ourselves to the simpler version and follow [Melrose 1994, §14]. The only modification lies in replacing scattering asymptotics by the regularity result given in Lemma 3.1.

Lemma 3.3. *Suppose that P satisfies* (5), (8), *and* (33). *Then for $|\omega|\le\delta$ and $f\in C^\infty(M)$, the limit*

$$(P-\omega-i\epsilon)^{-1}f\xrightarrow{H^{-1/2-}(M)}(P-\omega-i0)^{-1}f,\quad \epsilon\to 0+,$$

exists. This limit is the unique solution to the equation

$$(P-\omega)u=f,\quad \mathrm{WF}(u)\subset\Lambda^+, \tag{34}$$

and the map $\omega\mapsto(P-\omega-i0)^{-1}f\in H^{-\frac12-}(M)$ is continuous in $\omega\in[-\delta,\delta]$.

Remark. Replacing P with $-P$ we see that there is also a limit

$$(P-\omega+i\epsilon)^{-1}f\xrightarrow{H^{-1/2-}(M)}(P-\omega+i0)^{-1}f,\quad \epsilon\to 0+,$$

which satisfies (34) with Λ^+ replaced by Λ^-.

Proof. We first note that Lemma 3.1 and the spectral assumption (33) imply that (34) has no more than one solution. By (30), if a (distributional) limit $(P-\omega-i\epsilon_j)^{-1}f$, $\epsilon_j\to 0$, exists then it solves (34).

To show that the limit exists, put $u_\epsilon := (P-\omega-i\epsilon)^{-1} f$ and suppose first that $\|u_\epsilon\|_{-\frac{1}{2}-\alpha}$ is not bounded as $\epsilon \to 0+$ for some $\alpha > 0$. Hence there exists $\epsilon_j \to 0+$ such that $\|u_{\epsilon_j}\|_{-\frac{1}{2}-\alpha} \to \infty$. Putting $v_j := u_{\epsilon_j}/\|u_{\epsilon_j}\|_{-\frac{1}{2}-\alpha}$ we obtain

$$(P-\omega-i\epsilon_j)v_j = f_j, \quad \|v_j\|_{-\frac{1}{2}-\alpha} = 1, \quad f_j \xrightarrow{C^\infty(M)} 0. \tag{35}$$

Applying (28) with $N = \frac{1}{2}+\alpha$ we see that v_j is bounded in $H^{-\frac{1}{2}-\beta}(M)$ for any $\beta > 0$. Since $H^{-\frac{1}{2}-\beta}(M) \hookrightarrow H^{-\frac{1}{2}-\alpha}(M)$, we know $\beta < \alpha$ is compact and can assume, by passing to a subsequence, that $v_j \to v$ in $H^{-\frac{1}{2}-\alpha}(M)$. Then $(P-\omega)v = 0$ and the same reasoning that led to (30) shows that $\mathrm{WF}(v) \subset \Lambda^+$. Thus v solves (34) with $f \equiv 0$, implying that $v \equiv 0$. This gives a contradiction with the normalization $\|v_j\|_{-\frac{1}{2}-\alpha} = 1$.

We conclude that u_ϵ is bounded in $H^{-\frac{1}{2}-\alpha}(M)$ for all $\alpha > 0$. But then similarly to the previous paragraph $(u_\epsilon)_{\epsilon\to 0}$ is precompact in $H^{-\frac{1}{2}-\alpha}(M)$ for all $\alpha > 0$. Since every limit point has to be the (unique) solution to (34), we see that u_ϵ converges to that solution as $\epsilon \to 0+$ in $H^{-\frac{1}{2}-\alpha}(M)$.

As for continuity in ω, we note that the above proof gives the stronger statement

$$(P-\omega_j-i\epsilon_j)^{-1} f \xrightarrow{H^{-1/2-}(M)} (P-\omega-i0)^{-1} f \tag{36}$$

for all $\epsilon_j \to 0+$, $\omega_j \to \omega$, and $|\omega_j| \le \delta$. □

In Section 4B we will need the following upgraded version of Lemma 3.3:

Lemma 3.4. *Suppose that P satisfies* (5), (8), *and* (33). *Let $s < -\frac{1}{2}$ and $g \in H^{s+1}(M)$, $\mathrm{WF}(g) \subset \Lambda^+$, where Λ^+ is defined by* (12). *Then for $|\omega| \le \delta$ the limit*

$$(P-\omega-i\epsilon)^{-1} g \xrightarrow{H^{s-}(M)} (P-\omega-i0)^{-1} g, \quad \epsilon \to 0+, \tag{37}$$

exists, and $\mathrm{WF}((P-\omega-i0)^{-1}g) \subset \Lambda^+$. *In particular, for $k \ge 1$ and $f \in C^\infty(M)$ the limit*

$$(P-\omega-i\epsilon)^{-k} f \xrightarrow{H^{-k+1/2-}(M)} (P-\omega-i0)^{-k} f, \quad \epsilon \to 0+, \tag{38}$$

exists. Finally, $(P-\omega-i0)^{-1} f \in C^k_\omega([-\delta,\delta]; H^{-k-\frac{1}{2}-}(M))$, *with*

$$\partial^k_\omega (P-\omega-i0)^{-1} f = k!\,(P-\omega-i0)^{-k-1} f.$$

Proof. We follow closely the proof of Lemma 3.3 and put $u_\epsilon := (P-\omega-i\epsilon)^{-1} g$. Since $P-\omega-i\epsilon$ is elliptic for every $\epsilon > 0$, we have $u_\epsilon \in H^{s+1}(M)$ and $\mathrm{WF}(u_\epsilon) \subset \mathrm{WF}(g) \subset \Lambda^+$, so it remains to establish uniformity as $\epsilon \to 0+$. We use the following version of (29) (which follows from the same proof): for every $A \in \Psi^0(M)$ with $\mathrm{WF}(A) \cap \Lambda^+ = \varnothing$ there exists $\widetilde{B} \in \Psi^0(M)$ with $\mathrm{WF}(\widetilde{B}) \cap \Lambda^+ = \varnothing$ such that

$$\|Au_\epsilon\|_{s'} \le C\|\widetilde{B} g\|_{s'+1} + C\|u_\epsilon\|_{-N}, \quad s' > -\tfrac{1}{2}, \tag{39}$$

where the constant C does not depend on ω, ϵ. We also have the following version of (28): there exists $B' \in \Psi^0(M)$ with $\mathrm{WF}(B') \cap \Lambda^+ = \varnothing$ such that

$$\|u_\epsilon\|_s \le C\|g\|_{s+1} + C\|B' g\|_1 + C\|u_\epsilon\|_{-N}, \quad s < -\tfrac{1}{2}. \tag{40}$$

Here the norms $\|\widetilde{B}g\|_{s'+1}$ and $\|B'g\|_1$ are finite since $\mathrm{WF}(g)\subset\Lambda^+$. From (39) and (40) we get regularity for limit points of u_{ϵ_j} similarly to (30):

$$\text{there exist } \epsilon_j\to 0+,\ u\in\mathscr{D}'(M) \text{ such that } u_{\epsilon_j}\xrightarrow{\mathscr{D}'(M)} u \quad\Longrightarrow\quad u\in H^s(M),\quad \mathrm{WF}(u)\subset\Lambda^+.$$

The existence of the limit (37) follows as in the proof of Lemma 3.3, replacing $-\frac{1}{2}$ by s in Sobolev space orders; here $u=(P-\omega-i0)^{-1}g$ is the unique solution to

$$(P-\omega)u=g,\quad \mathrm{WF}(u)\subset\Lambda^+.$$

Iterating this argument, we get existence of the limit (38) and continuous dependence of $(P-\omega-i0)^{-k}f\in H^{-k+\frac{1}{2}-}$ on $\omega\in[-\delta,\delta]$ similarly to (36), with $u=(P-\omega-i0)^{-k}f$ being the unique solution to

$$(P-\omega)^k u=f,\quad \mathrm{WF}(u)\subset\Lambda^+.$$

It remains to show differentiability in ω. For simplicity we assume that $\omega=0$ and show that for $f\in C^\infty(M)$,

$$\partial_\omega[(P-\omega-i0)^{-1}f]|_{\omega=0}=(P-\omega-i0)^{-2}f\quad\text{in } H^{-\frac{3}{2}-}. \tag{41}$$

The case of higher derivatives is handled by iteration. To show (41) we define $u_\epsilon(\omega):=(P-\omega-i\epsilon)^{-1}f$ and write for $\omega\neq 0$, with limits in $H^{-\frac{3}{2}-}$,

$$\begin{aligned}\frac{u_0(\omega)-u_0(0)}{\omega}&=\lim_{\epsilon\to 0+}\frac{u_\epsilon(\omega)-u_\epsilon(0)}{\omega}=\lim_{\epsilon\to 0+}(P-\omega-i\epsilon)^{-1}(P-i\epsilon)^{-1}f\\&=(P-\omega-i0)^{-1}(P-i0)^{-1}f.\end{aligned} \tag{42}$$

To show the last equality above we first note that the family $(P-\omega-i\epsilon)^{-1}(P-i\epsilon)^{-1}f$ is precompact in $H^{-\frac{3}{2}-\alpha}(M)$ for any $\alpha>0$ as follows from iterating (40). By (39) every limit point u of this family as $\epsilon\to 0+$ satisfies $P(P-\omega)u=f$, $\mathrm{WF}(u)\subset\Lambda$ and thus equals $(P-\omega-i0)^{-1}(P-i0)^{-1}f$. Finally, letting $\omega\to 0$ in (42) we get (41). □

4. Lagrangian structure of the resolvent

We now describe the Lagrangian structure of the resolvent refining the results of [Haber and Vasy 2015] in our special case. To start, we briefly review basic theory of Lagrangian distributions following [Hörmander 1985b, §25.1].

4A. ***Lagrangian distributions.*** Let M be a compact surface and $\Lambda_0\subset T^*M\setminus 0$ a conic Lagrangian submanifold without boundary. Denote by $I^s(M;\Lambda_0)\subset\mathcal{D}'(M)$ the space of Lagrangian distributions of order s on M associated to Λ_0. It has the following properties:

(1) $I^s(M;\Lambda_0)\subset H^{-\frac{1}{2}-s-}(M)$.

(2) For all $u\in I^s(M;\Lambda_0)$ we have $\mathrm{WF}(u)\subset\Lambda_0$.

(3) If $\Lambda_1\subset\Lambda_0$ is an open conic subset and $u\in I^s(M;\Lambda_0)$, then $u\in I^s(M;\Lambda_1)$ if and only if $\mathrm{WF}(u)\subset\Lambda_1$.

(4) For all $A \in \Psi^k(M)$ and $u \in I^s(M;\Lambda_0)$ we have $Au \in I^{s+k}(M;\Lambda_0)$.

(5) If additionally $\sigma(A)|_{\Lambda_0} = 0$, then $Au \in I^{s+k-1}(M;\Lambda_0)$.

Define

$$I^{s+}(M;\Lambda_0) := \bigcap_{s'>s} I^{s'}(M;\Lambda_0).$$

A simple example on a torus (in the notation of Section 1C) is given by

$$u(x) := \big(x_1 - \tfrac{\pi}{2} - i0\big)^{-1}\varphi(x), \quad \varphi \in C_c^\infty(B(0,1)), \quad u \in I^0(\mathbb{T}^2;\Lambda_0^+) \subset H^{-\frac{1}{2}-}(\mathbb{T}^2), \tag{43}$$

where Λ_0^+ is given in (10).

To define Lagrangian distributions we use Melrose's iterative characterization [Hörmander 1985b, Definition 25.1.1]: $u \in \mathcal{D}'(M)$ lies in $I^{s+}(M;\Lambda_0)$ if and only if $\mathrm{WF}(u) \subset \Lambda_0$ and

$$A_1 \cdots A_\ell\, u \in H^{-\frac{1}{2}-s-}(M) \quad \text{for any } A_1,\dots,A_\ell \in \Psi^1(M),\ \ \sigma(A_j)|_{\Lambda_0} = 0. \tag{44}$$

Note that [Hörmander 1985b] uses Besov spaces ${}^\infty H^s$. However, this does not make a difference in (44) since $H^s \subset {}^\infty H^s \subset H^{s'}$ for all $s' < s$; see [Hörmander 1985a, Proposition B.1.2].

We also need oscillatory integral representations for Lagrangian distributions. Assume that in some local coordinate system on M, Λ_0 is given by

$$\Lambda_0 = \{(x,\xi)\colon x = \partial_\xi F(\xi),\ \xi \in \Gamma_0\}, \tag{45}$$

where $\Gamma_0 \subset \mathbb{R}^2 \setminus 0$ is an open cone and $F : \Gamma_0 \to \mathbb{R}$ is homogeneous of order 1. (Every Lagrangian can be locally written in this form after a change of base, x, variables — see [Hörmander 1985a, Theorem 21.2.16]. Using a pseudodifferential partition of unity we can write every Lagrangian distribution as a sum of expressions of the form (46).) Then $u \in I^s(M;\Lambda_0)$ if and only if u can be written (modulo a C^∞ function) as

$$u(x) = \int_{\Gamma_0} e^{i(\langle x,\xi\rangle - F(\xi))}\, a(\xi)\, d\xi, \tag{46}$$

where $a(\xi) \in C^\infty(\mathbb{R}^2)$ is a symbol of order $s - \frac{1}{2}$, namely

$$|\partial_\xi^\alpha a(\xi)| \le C_\alpha \langle\xi\rangle^{s-\frac{1}{2}-|\alpha|}, \quad \xi \in \mathbb{R}^2, \tag{47}$$

and a is supported in a closed cone contained in Γ_0. See [Hörmander 1985b, Proposition 25.1.3]. An equivalent way of stating (46) is in terms of the Fourier transform $\hat{u}$: $e^{iF(\xi)}\hat{u}(\xi)$ is a symbol, that is, satisfies estimates (47).

We finally review properties of the principal symbol of a Lagrangian distribution, used in the proof of Lemma 4.5 below, referring the reader to [loc. cit., Chapter 25] for details. The principal symbol of a Lagrangian distribution, u, with values in half-densities, $u \in I^s(M,\Lambda;\Omega_M^{\frac{1}{2}})$, is the equivalence class

$$\sigma(u) \in S^{s+\frac{1}{2}}(\Lambda;\mathcal{M}_\Lambda \otimes \Omega_\Lambda^{\frac{1}{2}})/S^{s-\frac{1}{2}}(\Lambda;\mathcal{M}_\Lambda \otimes \Omega_\Lambda^{\frac{1}{2}}),$$

see [loc. cit., Theorem 25.1.9], where:

- $\Omega_\Lambda^{\frac12}$ is the line bundle of half-densities on Λ.
- $\mathcal{M}_\Lambda$ is the Maslov line bundle; it has a finite number of prescribed local frames with ratios of any two prescribed frames given by a constant of absolute value 1. Consequently it has a canonical inner product and does not enter into the calculations below.
- $S^k(\Lambda;\mathcal{M}_\Lambda\otimes\Omega_\Lambda^{\frac12})$ is the space of sections in $C^\infty(\Lambda;\mathcal{M}_\Lambda\otimes\Omega_\Lambda^{\frac12})$ which are symbols of order k, defined using the dilation operator $(x,\xi)\mapsto(x,\lambda\xi)$, $\lambda>0$; see the discussion on [Hörmander 1985b, page 13]. In the parametrization (46) we have $\sigma(u|dx|^{\frac12})=(2\pi)^{-\frac12}a(\xi)|d\xi|^{\frac12}$. The factor $|d\xi|^{\frac12}$ accounts for the difference in the order of the symbol.

If $P\in\Psi^\ell(M;\Omega_M^{\frac12})$ satisfies $\sigma(P)|_\Lambda=0$ and $u\in I^s(M,\Lambda;\Omega_M^{\frac12})$ then

$$Pu\in I^{s+\ell-1}(M,\Lambda;\Omega_M^{\frac12}),\quad \sigma(Pu)=\frac{1}{i}L\sigma(u), \tag{48}$$

where L is a first-order differential operator on $C^\infty(\Lambda;\mathcal{M}_\Lambda\otimes\Omega_\Lambda^{\frac12})$ with principal part H_p. Equation (48) is the *transport equation* for P (the *eikonal equation* corresponds to $\sigma(P)|_\Lambda=0$)—see [loc. cit., Theorem 25.2.4]. If P is self-adjoint, then its subprincipal symbol is real-valued by [Hörmander 1985a, Theorem 18.1.34] and thus by [Hörmander 1985b, (25.2.12)]

$$L^*=-L\quad\text{on } L^2(\Lambda;\mathcal{M}_\Lambda\otimes\Omega_\Lambda^{\frac12}). \tag{49}$$

4B. ***Lagrangian regularity.*** We now establish Lagrangian regularity for elements in the range of the operators $(P-\omega\mp i0)^{-1}$ constructed in Section 3C:

Lemma 4.1. *Suppose that P satisfies* (5), (8), *and* (33). *Let $f\in C^\infty(M)$ and*

$$u^\pm(\omega):=(P-\omega\mp i0)^{-1}f\in H^{-\frac12-}(M),\quad |\omega|\le\delta.$$

Then $u^\pm(\omega)\in I^0(M;\Lambda_\omega^\pm)$. Moreover, the symbols of $u^\pm(\omega)$ depend smoothly on ω:

$$u^\pm(\omega)\in C^\infty_\omega([-\delta,\delta];I^0(M;\Lambda_\omega^\pm)), \tag{50}$$

where the precise meaning of (50) *is explained in Lemma 4.4 below* ((67) *and Remark* 2).

Remark. Lemma 4.1 is similar to [Haber and Vasy 2015, Theorems 1.7 and 6.3]. There are two differences: that paper makes the assumption that the Hamiltonian field H_p is radial on $\Lambda_\omega^\pm$ (which is not true in our case) and it also does not prove smooth dependence of the symbols of $u^\pm(\omega)$ on ω. Because of these we give a self-contained proof of Lemma 4.1 below, noting that the argument is simpler in our situation.

We focus on the case of $u^+(\omega)$, with regularity of $u^-(\omega)$ proved by replacing P, ω with $-P$, $-\omega$, respectively. By Lemma 3.4 we have for every $k\ge0$

$$u^+(\omega)\in C^k_\omega([-\delta,\delta];H^{-k-\frac12-}(M)),\quad \mathrm{WF}(\partial_\omega^k u^+(\omega))\subset\Lambda^+, \tag{51}$$

where the wavefront set statement is uniform in ω.

To upgrade (51) to Lagrangian regularity, we use the criterion (44), applying first-order operators W and $D_\omega - Q$ to $u^+(\omega)$ (see Lemma 4.3 below). Here,

$$W, Q \in \Psi^1(M), \quad \sigma(W) = G_+, \quad \sigma(Q)|_{\Lambda^+} = \Phi_+, \tag{52}$$

where G_+ is the defining function of Λ^+ constructed in Lemma 2.4 and Φ_+ is defined in (14). The operator $D_\omega - Q$, where $D_\omega := (1/i)\partial_\omega$, is used to establish smoothness in ω.

Our proof uses the following corollary of (26):

$$\begin{gathered} \text{if} \quad Z \in \Psi^{-1}(M), \quad \sigma(Z)|_{\Lambda^+} = 0, \quad s < -\tfrac{1}{2} \quad \text{then} \\ v \in \mathcal{D}'(M), \quad \mathrm{WF}(v) \subset \Lambda^+, \quad (P + Z - \omega)v \in H^{s+1} \quad \Longrightarrow \quad v \in H^s. \end{gathered} \tag{53}$$

The addition of Z does not change the validity of (26) since it is a subprincipal term whose symbol vanishes on Λ^+; see [DZ19, Theorem E.54].

We also use the following identity valid for any operators A, B on $\mathcal{D}'(M)$:

$$B^m A = \sum_{j=0}^{m} \binom{m}{j} (\mathrm{ad}_B^j A) B^{m-j}, \quad \mathrm{ad}_B A := [B, A], \ \ \mathrm{ad}_B^0 A := A. \tag{54}$$

The first step of the proof is to establish regularity with respect to powers of W:

Lemma 4.2. *Assume that $v \in \mathcal{D}'(M)$ satisfies for some $\ell \geq 0$ and $s < -\frac{1}{2}$*

$$\mathrm{WF}(v) \subset \Lambda^+, \quad W^j (P - \omega) v \in H^{s+1} \ \text{for } j = 0, \dots, \ell. \tag{55}$$

Then $W^\ell v \in H^s$, where W is defined in (52)*.*

Proof. We argue by induction on ℓ. For $\ell = 0$ the lemma follows immediately from (53). We thus assume that $\ell > 0$ and the lemma is true for all smaller values of ℓ; in particular $W^k v \in H^s$ for $0 \leq k \leq \ell - 1$. Using (54) we write

$$W^\ell (P - \omega) = (P - \omega) W^\ell + \sum_{j=1}^{\ell} \binom{\ell}{j} (\mathrm{ad}_W^j P) W^{\ell - j}. \tag{56}$$

We recall from Lemma 2.4 that near Λ^+ we have $H_{G_+} p = -a_+ G_+$, where a_+ is homogeneous of order -1 and $a_+|_{\Lambda^+} = 0$. Therefore for $j \geq 1$ we have $H_{G_+}^j p = -(H_{G_+}^{j-1} a_+) G_+$ near Λ^+. Motivated by this we take

$$B_j \in \Psi^{-1}(M), \quad \sigma(B_j) = (-1)^{j-1} i^j H_{G_+}^{j-1} a_+, \quad 1 \leq j \leq \ell.$$

Then, for $1 \leq j \leq \ell$

$$\mathrm{ad}_W^j P = B_j W + R_j, \quad R_j \in \Psi^{-1} \text{ microlocally near } \Lambda^+. \tag{57}$$

Combining (56) and (57) we get

$$(P - \omega) W^\ell = W^\ell (P - \omega) - \sum_{j=1}^{\ell} \binom{\ell}{j} (B_j W^{\ell + 1 - j} + R_j W^{\ell - j}). \tag{58}$$

Applying both sides of (58) to v and using that $W^k v \in H^s$ for $0 \le k \le \ell - 1$ and that $W^\ell (P - \omega) v \in H^{s+1}$ we get

$$(P + \ell B_1 - \omega) W^\ell v \in H^{s+1}.$$

Since $\sigma(B_1) = i a_+$ vanishes on Λ^+, we apply (53) to conclude that $W^\ell v \in H^s$ as needed. □

Since $(P - \omega) u^+(\omega) = f \in C^\infty(M)$, Lemma 4.2 implies that

$$W^\ell u^+(\omega) \in H^{-\frac{1}{2}-}(M) \quad \text{for all } \ell \ge 0. \tag{59}$$

This can be generalized as follows:

$$A_1 \cdots A_\ell u^+(\omega) \in H^{-\frac{1}{2}-}(M) \quad \text{for all } A_1, \dots, A_\ell \in \Psi^1(M), \ \ \sigma(A_j)|_{\Lambda^+} = 0. \tag{60}$$

To see (60), we argue by induction on ℓ. We have $\sigma(A_j) = \tilde a_j G_+$ near $\mathrm{WF}(u^+(\omega)) \subset \Lambda^+$ for some $\tilde a_j$ which is homogeneous of order 0. Taking $\widetilde A_j \in \Psi^0(M)$ with $\sigma(\widetilde A_j) = \tilde a_j$ we have

$$A_j = \widetilde A_j W + \widetilde R_j \quad \text{where } \widetilde R_j \in \Psi^0(M) \text{ microlocally near } \mathrm{WF}(u^+(\omega)).$$

Then we can write $A_1 \cdots A_\ell u^+(\omega)$ as the sum of two kinds of terms (plus a C^∞ remainder):

- the term $\widetilde A_1 \cdots \widetilde A_\ell W^\ell u^+(\omega)$, which lies in $H^{-\frac{1}{2}-}(M)$ by (59), and
- terms of the form $A_1' \cdots A_m' u^+(\omega)$, where $0 \le m \le \ell - 1$, $A_j' \in \Psi^1(M)$, and $\sigma(A_j')|_{\Lambda^+} = 0$, which lie in $H^{-\frac{1}{2}-}(M)$ by the inductive hypothesis.

From (60) we can deduce (similarly to the proof of Lemma 4.4 below) that $u^+(\omega) \in I^{0+}(M; \Lambda_\omega^+)$ for each $\omega \in [-\delta, \delta]$. To obtain the smooth dependence of the symbol of $u^+(\omega)$ on ω we generalize (59) by additionally applying powers of $D_\omega - Q$:

Lemma 4.3. *For all integers $\ell, m \ge 0$ we have*

$$W^\ell (D_\omega - Q)^m u^+(\omega) \in H^{-\frac{1}{2}-}(M), \quad |\omega| \le \delta, \tag{61}$$

and the corresponding norms are bounded uniformly in ω.

Proof. We argue by induction on m, with the case $m = 0$ following from (59). Put

$$u_j(\omega) := (D_\omega - Q)^j u^+(\omega) \in \mathcal{D}'(M), \quad 0 \le j \le m.$$

By (51) we have $\mathrm{WF}(u_j(\omega)) \subset \Lambda^+$ for all j. Moreover, by the inductive hypothesis

$$W^\ell u_j(\omega) \in H^{-\frac{1}{2}-}(M) \quad \text{for all } \ell, \ \ 0 \le j \le m - 1. \tag{62}$$

Put

$$Y := [P - \omega, D_\omega - Q] = -i - [P, Q] \in \Psi^0(M)$$

and note that since $\sigma(Q)|_{\Lambda^+} = \Phi_+$ and $H_p \Phi_+ \equiv 1$ on Λ^+ by (15),

$$\sigma(Y)|_{\Lambda^+} = 0. \tag{63}$$

Moreover, by (15) we have $H_{G_+}\Phi_+ \equiv 0$ on Λ^+; thus the Hamiltonian vector field H_{Φ_+} is tangent to Λ^+. This implies that

$$\sigma(\mathrm{ad}_Q^j Y) = (-i)^j H_{\Phi_+}^j \sigma(Y) \equiv 0 \quad \text{on } \Lambda^+ \text{ for all } j \geq 0. \tag{64}$$

Applying (54) with $A := P - \omega$ and $B := D_\omega - Q$ to $u^+(\omega)$ we get

$$(P-\omega)u_m(\omega) = (D_\omega - Q)^m f + \sum_{j=1}^{m} (-1)^{j-1} \binom{m}{j} (\mathrm{ad}_Q^{j-1} Y) u_{m-j}(\omega). \tag{65}$$

Since $f \in C^\infty$ does not depend on ω, we have $(D_\omega - Q)^m f \in C^\infty$. Next, by the inductive hypothesis (62) we have $W^\ell u_{m-j}(\omega) \in H^{-\frac{1}{2}-}$ for all $\ell \geq 0$ and $1 \leq j \leq m$. Arguing similarly to (60) and using (64) we see that $W^\ell (\mathrm{ad}_Q^{j-1} Y) u_{m-j}(\omega) \in H^{\frac{1}{2}-}$ as well (here $\mathrm{ad}_Q^{j-1} Y \in \Psi^0(M)$ which explains the stronger regularity). Thus (65) implies

$$W^\ell (P-\omega) u_m(\omega) \in H^{\frac{1}{2}-}(M) \quad \text{for all } \ell \geq 0.$$

Now Lemma 4.2 gives $W^\ell u_m(\omega) \in H^{-\frac{1}{2}-}$ for all $\ell \geq 0$ as needed.

Finally, uniformity of (61) in ω follows immediately from the proof since the estimates (51) and (26) that we used are uniform in ω. □

We now deduce from Lemma 4.3 that $u^+(\omega)$ has microlocal oscillatory integral representations (46) with symbols depending smoothly on ω. This shows the weaker version of (50) with I^0 replaced by I^{0+}.

Lemma 4.4. *Assume that $\mathcal{U} \subset T^*M \setminus 0$ is an open conic set such that $\Lambda_\omega^+ \cap \mathcal{U}$ are given in the form* (16) *in some local coordinate system on M:*

$$\Lambda_\omega^+ \cap \mathcal{U} = \{(x,\xi) : x = \partial_\xi F(\omega,\xi),\ \xi \in \Gamma_0\}, \quad |\omega| \leq \delta, \tag{66}$$

where $\xi \mapsto F(\omega,\xi)$ is homogeneous of order 1 and $\Gamma_0 \subset \mathbb{R}^2 \setminus 0$ is an open cone. Let $A \in \Psi^0(M)$, $\mathrm{WF}(A) \subset \mathcal{U}$. Then,

$$Au^+(\omega,x) = \int_{\Gamma_0} e^{i(\langle x,\xi\rangle - F(\omega,\xi))}\, a(\omega,\xi)\, d\xi + C^\infty_{\omega,x}, \quad |\omega| \leq \delta, \tag{67}$$

where $a(\omega,\xi)$ is a smooth in ω family of symbols of order $-\frac{1}{2}+$ in ξ supported in a closed cone inside Γ_0, see (47).

Remarks. (1) The statement (67) means that $u^+(\omega)$ can be represented as (46), *microlocally* in every closed cone contained in $\mathcal{U}$.

(2) When (67) holds for every choice of parametrization (66) we write

$$u^+(\omega) \in C^\infty_\omega([-\delta,\delta]; I^{0+}(M;\Lambda_\omega^+)),$$

with the analogous notation in the case of $u^-(\omega)$. That explains the statement of Lemma 4.1.

Proof. Since $(P-\omega)u^+(\omega) = f \in C^\infty(M)$, it follows from Lemma 4.3 that for all $m, \ell, r \geq 0$

$$(D_\omega - Q)^m W^\ell (P-\omega)^r u^+(\omega) \in H^{-\frac{1}{2}-}(M).$$

This can be generalized as follows:

$$(D_\omega - Q(\omega))^m A_1(\omega)\cdots A_\ell(\omega)u^+(\omega) \in H^{-\frac{1}{2}-}(M) \tag{68}$$

for all m and all $A_1(\omega),\dots,A_\ell(\omega), Q(\omega)\in\Psi^1(M)$ depending smoothly on $\omega\in[-\delta,\delta]$ and such that $\sigma(A_j(\omega))|_{\Lambda_\omega^+}=0$, $\sigma(Q(\omega))|_{\Lambda_\omega^+}=\Phi_+$. The proof is similar to the proof of (60), using the decomposition

$$A_j(\omega)=A_j'(\omega)W+A_j''(\omega)(P-\omega)+R_j(\omega),\quad \text{where } R_j(\omega)\in\Psi^0 \text{ microlocally near } \mathrm{WF}(u^+(\omega)),$$

for some $A_j'(\omega), A_j''(\omega)\in\Psi^0(M)$ depending smoothly on $\omega\in[-\delta,\delta]$.

Since $\mathrm{WF}(A\partial_\omega^k u^+(\omega))\subset\Lambda^+\cap p^{-1}([-\delta,\delta])\cap\mathcal{U}$ for all k, by the Fourier inversion formula we can write $Au^+(\omega)$ in the form (67) for some $a(\omega,\xi)$ which is smooth in ω,ξ and supported in $\xi\in\Gamma_1$, where $\Gamma_1\subset\Gamma_0$ is some closed cone. It remains to show the following growth bounds as $\xi\to\infty$: for every $\epsilon>0$

$$\langle\xi\rangle^{-\frac{1}{2}+|\alpha|-\epsilon}\partial_\omega^m\partial_\xi^\alpha a(\omega,\xi)\in L^\infty_\omega([-\delta,\delta];L^2_\xi(\mathbb{R}^2)). \tag{69}$$

(From (69) one can get L^∞_ξ bounds using Sobolev embedding as in the proof of [Hörmander 1985b, Proposition 25.1.3].)

Denote by $\mathcal{I}(a)$ the integral on the right-hand side of (67). By Lemma 2.2 we have $\partial_\omega F(\omega,\xi)=-\Phi_+(\partial_\xi F(\omega,\xi),\xi)$; therefore we may take $Q(\omega):=-\partial_\omega F(\omega,D_x)$ to be a Fourier multiplier. The operators

$$A_{jk}(\omega):=D_{x_k}((\partial_{\xi_j}F)(\omega,D_x)-x_j),\quad j,k\in\{1,2\},$$

lie in Ψ^1 and satisfy $\sigma(A_{jk}(\omega))|_{\Lambda_\omega^+}=0$. We have

$$(D_\omega-Q(\omega))\mathcal{I}(a)=\mathcal{I}(D_\omega a),\quad A_{jk}(\omega)\mathcal{I}(a)=\mathcal{I}(\xi_k D_{\xi_j}a).$$

Also, if $\mathcal{I}(a)\in H^{-\frac{1}{2}-}$ uniformly in ω, then $\langle\xi\rangle^{-\frac{1}{2}-\epsilon}a(\omega,\xi)\in L^\infty_\omega([-\delta,\delta];L^2_\xi(\mathbb{R}^2))$. Applying (68) with the operators $D_\omega-Q(\omega)$ and $A_{jk}(\omega)$ we get (69), finishing the proof. □

We finally show the stronger statement of Lemma 4.1 (with I^0 instead of I^{0+}) using the transport equation satisfied by the principal symbol:

Lemma 4.5. *We have*

$$u^+(\omega)\in C^\infty_\omega([-\delta,\delta];I^0(M;\Lambda_\omega^+));$$

that is, (67) *holds where* $a(\omega,\xi)$ *is a symbol of order* $-\frac{1}{2}$ *in* ξ.

Proof. In our setting $P\in\Psi^0(M)$ is self-adjoint with respect to a smooth density on M — see (5). Using that density to trivialize the half-density bundle we obtain a self-adjoint operator $P\in\Psi^0(M;\Omega_M^{\frac{1}{2}})$.

Let $a^+\in S^{\frac{1}{2}+}(\Lambda_\omega^+;\mathcal{M}_{\Lambda_\omega^+}\otimes\Omega^{\frac{1}{2}}_{\Lambda_\omega^+})$ be a representative of $\sigma(u^+(\omega))$. Using the transport equation (48) and $(P-\omega)u^+(\omega)=f\in C^\infty(M)$, we have

$$b^+:=La^+\in S^{-\frac{3}{2}+}(\Lambda_\omega^+;\mathcal{M}_{\Lambda_\omega^+}\otimes\Omega^{\frac{1}{2}}_{\Lambda_\omega^+}), \tag{70}$$

where L is a first-order differential operator on $C^\infty(\Lambda_\omega^+;\mathcal{M}_{\Lambda_\omega^+}\otimes\Omega^{\frac{1}{2}}_{\Lambda_\omega^+})$ with principal part given by H_p and $L^*=-L$ by (49).

We trivialize $\Omega^{\frac{1}{2}}_{\Lambda^+_\omega}$ using the density ν^+_ω constructed in Lemma 2.5 and write

$$a^+ = \tilde{a}^+\sqrt{\nu^+_\omega}, \quad b^+ = \tilde{b}^+\sqrt{\nu^+_\omega},$$

where $\tilde{a}^+ \in S^{0+}(\Lambda^+_\omega; \mathcal{M}_{\Lambda^+_\omega})$, $\tilde{b}^+ \in S^{-2+}(\Lambda^+_\omega; \mathcal{M}_{\Lambda^+_\omega})$. By (70) we have

$$(H_p + V)\tilde{a}^+ = \tilde{b}^+, \tag{71}$$

where H_p naturally acts on sections of the locally constant bundle $\mathcal{M}_{\Lambda^+_\omega}$ and $V \in C^\infty(\Lambda^+_\omega)$ is homogeneous of order -1. Moreover, since $L^* = -L$ we have

$$\Re V = \tfrac{1}{2}(\mathcal{L}_{H_p}\nu^+_\omega)/\nu^+_\omega = 0$$

using Lemma 2.5.

By (71) for all $(x,\xi) \in \Lambda^+_\omega$ and $t \geq 0$ we have

$$\tilde{a}^+(x,\xi) = (e^{-t(H_p+V)}\tilde{a}^+)(x,\xi) + \int_0^t (e^{-s(H_p+V)}\tilde{b}^+)(x,\xi)\,ds. \tag{72}$$

Since $\Re V = 0$ we have $|e^{-t(H_p+V)}\tilde{a}^+(x,\xi)| = |\tilde{a}^+(e^{-tH_p}(x,\xi))|$ and the same is true for $\tilde{b}^+$.

Take $(x,\xi) \in \Lambda^+_\omega$ with $|\xi|$ large. As in (21) choose $t \geq 0$, $t \sim |\xi|$, such that $e^{-tH_p}(x,\xi) \in S^*M$; we next apply (72). The first term on the right-hand side is bounded uniformly as $\xi \to \infty$. The same is true for the second term since the function under the integral is $\mathcal{O}((t-s)^{-2+})$. It follows that $\tilde{a}^+(x,\xi)$ is bounded as $\xi \to \infty$.

Since $[\xi\partial_\xi, H_p + V] = -H_p - V$, we have for all j

$$(H_p + V)(\xi\partial_\xi)^j\tilde{a}^+ = (\xi\partial_\xi + 1)^j\tilde{b}^+ \in S^{-2+}(\Lambda^+_\omega; \mathcal{M}_{\Lambda^+_\omega}). \tag{73}$$

It follows that $(H_p + V)^\ell(\xi\partial_\xi)^j\tilde{a}^+ = \mathcal{O}(\langle\xi\rangle^{-\ell})$ for all j, ℓ: the case $\ell = 0$ follows from (72) applied to (73) and the case $\ell \geq 1$ follows directly from (73). Since $\xi\partial_\xi$ and H_p form a frame on Λ^+_ω, we have $\tilde{a}^+ \in S^0(\Lambda^+_\omega; \mathcal{M}_{\Lambda^+_\omega})$, which implies that $u^+_\omega \in I^0(M; \Lambda^+_\omega)$. □

Remark. It is instructive to consider the transport equation (71) in the microlocal model used in [Colin de Verdière and Saint-Raymond 2019]: near a model sink

$$\Lambda^+_\omega = \{(-\omega, x_2; \xi_1, 0) : \xi_1 > 0\} \subset T^*(\mathbb{R}_{x_1} \times \mathbb{S}^1_{x_2}) \subset 0$$

(see the global examples in Section 1C) we consider $p(x,\xi) := \xi_1^{-1}\xi_2 - x_1$. We are then solving $(p(x,D) - \omega)u^+(\omega) \equiv 0$ microlocally near Λ^+_ω, see [DZ19, Definition E.29], and for that we expand the symbol on u^+_ω into Fourier modes in x_2,

$$u^+_\omega(x) = \frac{1}{2\pi}\int_{\mathbb{R}}\sum_{n\in\mathbb{Z}} \hat{a}^+_\omega(n,\xi_1)\,e^{i(x_1+\omega)\xi_1}\,e^{inx_2}\,d\xi_1, \quad a^+_\omega = \sum_{n\in\mathbb{Z}} \hat{a}^+_\omega(n,\xi_1)\,e^{inx_2}|d\xi_1 dx_2|^{\frac{1}{2}}.$$

The Fourier coefficients should satisfy $(\xi_1^{-1} n + D_{\xi_1})\tilde{a}^+_\omega(n,\xi_1) = 0$ for $\xi_1 > 1$ and $\tilde{a}^\omega_+(n,\xi_1) = 0$ for $\xi_1 < -1$. Hence the symbol is given by

$$a^+_\omega = \tilde{a}^+(\omega)|dx_2 d\xi_1|^{\frac{1}{2}}, \quad \tilde{a}^+(x_2,\xi_1) = \sum_{n\in\mathbb{Z}} \xi_1^{-in} a_n(\omega)\, e^{inx_2}, \quad a_n(\omega) = \mathcal{O}(\langle n\rangle^{-\infty}).$$

Hence, the symbol is very "nonclassical" in the sense that it does not have an expansion in powers of ξ_1. In the general case an analogous conclusion follows from the structure of (71).

5. An asymptotic result

We now place ourselves in the setting of Lemma 4.1 and assume that $u(\omega) \in C^\infty_\omega([-\delta,\delta]; I^0(M;\Lambda_\omega))$ in the sense described in Lemma 4.5, where $\Lambda_\omega = \Lambda^+_\omega$ or $\Lambda_\omega = \Lambda^-_\omega$. We are interested in the asymptotic behavior as $t \to \infty$ of

$$I(t) := \int_0^t \int_{\mathbb{R}} e^{-is\omega} \varphi(\omega)\, u(\omega)\, d\omega\, ds \in \mathcal{D}'(M), \quad \varphi \in C^\infty_{\mathrm{c}}((-\delta,\delta)). \tag{74}$$

We have the following local asymptotic result.

Lemma 5.1. *Suppose that $u(\omega) \in \mathcal{D}'(\mathbb{R}^2)$ is given by*

$$u(\omega) = u(\omega,x) = \frac{1}{(2\pi)^2} \int_{\Gamma_0} e^{i(\langle x,\xi\rangle - F(\omega,\xi))}\, a(\omega,\xi)\, d\xi, \tag{75}$$

where Γ_0, F, and a satisfy the general conditions in (67)*. Suppose also that*

$$\epsilon\, \partial_\omega F(\omega,\xi) < 0, \quad \epsilon = \pm, \quad \xi \in \Gamma_0, \quad |\omega| \le \delta. \tag{76}$$

Then as $t \to \infty$,

$$\begin{gathered} I(t) = u_\infty + b(t) + v(t), \qquad \|b(t)\|_{H^{1/2-}} \le C, \qquad v(t) \to 0 \quad \text{in } H^{-\frac{1}{2}-}(\mathbb{R}^2), \\ u_\infty = \begin{cases} 2\pi\, \varphi(0)\, u(0), & \epsilon = +, \\ 0, & \epsilon = -. \end{cases} \end{gathered} \tag{77}$$

Proof. We start by remarking that we can assume that the amplitude a is supported away from $\xi = 0$. The remaining contribution can be absorbed into $b(t)$: if $a = a(\omega,\xi) = 0$ for $|\xi| > C$ then

$$\begin{aligned} \hat{w}(t,\xi) &:= \int_0^t \int_{\mathbb{R}} e^{-is\omega} e^{-iF(\omega,\xi)}\, a(\omega,\xi)\, \varphi(\omega)\, d\omega\, ds \\ &= \int_0^t \int_{\mathbb{R}} [(1+s^2)^{-1}(1+D^2_\omega) e^{-is\omega}]\, e^{-iF(\omega,\xi)}\, a(\omega,\xi)\, \varphi(\omega)\, d\omega\, ds, \end{aligned}$$

which by integration by parts in ω is bounded in t and compactly supported in ξ.

Since $u(\omega,x)$ has nice structure on the Fourier transform side it is natural to consider the Fourier transform of $x \mapsto I(t)(x)$, $J(t,\xi) := \mathcal{F}_{x\to\xi} I(t)$, where

$$J(t,\xi) = \frac{1}{h} \int_0^{ht} \int_{\mathbb{R}} e^{-\frac{i}{h}(F(\omega,\eta) + r\omega)}\, a\left(\omega, \frac{\eta}{h}\right) \varphi(\omega)\, d\omega\, dr, \quad \xi = \frac{\eta}{h}, \quad \eta \in \mathbb{S}^1. \tag{78}$$

From the assumptions on a we have $J(t,\xi)=0$ unless $\eta\in\Gamma_1$, where $\Gamma_1\subset\Gamma_0$ is a closed cone. The phase in $J(t)$ is stationary when

$$\omega=0,\quad r=r(\eta):=-\partial_\omega F(0,\eta). \tag{79}$$

From (76), $\partial_\omega F(\omega,\eta)\neq 0$ and this means that for some $\gamma>0$,

$$|r+\partial_\omega F(\omega,\eta)|>c\langle r\rangle,\quad \eta\in\mathbb{S}^1\cap\Gamma_1,\quad |\omega|\le\delta,\quad |r|\notin\Big(\gamma,\frac{1}{\gamma}\Big). \tag{80}$$

Let $\chi\in C_{\mathrm{c}}^\infty((\gamma/2,2/\gamma);[0,1])$ be equal to 1 on $(\gamma,1/\gamma)$. Using integration by parts based on

$$h^N\big(-(r+\partial_\omega F(\omega,\eta))^{-1}D_\omega\big)^N e^{-\frac{i}{h}(F(\omega,\eta)+r\omega)}=e^{-\frac{i}{h}(F(\omega,\eta)+r\omega)},$$

and (80), we see that, by taking $N\ge 2$,

$$\frac{1}{h}\int_0^{ht}\int_{\mathbb{R}}(1-\chi(r))\,e^{-\frac{i}{h}(F(\omega,\eta)+r\omega)}\,a\Big(\omega,\frac{\eta}{h}\Big)\varphi(\omega)\,d\omega\,dr=\mathcal{O}(h^{N-1}),$$

uniformly in $t\ge 0$. Hence, for all N

$$J(t)=\widetilde{J}(t)+\mathcal{F}_{x\mapsto\xi}\,u_0(t),\quad \sup_{t\ge 0}\|u_0(t)\|_{H^N}\le C_N,$$

$$\widetilde{J}(t,\xi):=\frac{1}{h}\int_0^{ht}\int_{\mathbb{R}}\chi(r)\,e^{-\frac{i}{h}(F(\omega,\eta)+r\omega)}\,a\Big(\omega,\frac{\eta}{h}\Big)\varphi(\omega)\,d\omega\,dr,\quad \xi=\frac{\eta}{h},\ \ \eta\in\mathbb{S}^1.$$

When $ht\ge 2/\gamma$, we have $\widetilde{J}(t,\xi)=\widetilde{J}(\infty,\xi)$ due to the support property of χ. In particular this implies that $\widetilde{J}(t,\xi)\to\widetilde{J}(\infty,\xi)$ as $t\to\infty$ pointwise in ξ. We apply the standard method of stationary phase to $\widetilde{J}(\infty)$ noting that

$$-\partial^2_{\omega,r}(F(\omega,\eta)+r\omega)=\begin{bmatrix}-\partial^2_\omega F & -1\\ -1 & 0\end{bmatrix},\quad \operatorname{sgn}\partial^2_{\omega,r}(F(\omega,\eta)-r\omega)=0.$$

Therefore

$$\widetilde{J}(\infty,\xi)=\begin{cases}2\pi\,a(0,\xi)\,\varphi(0)\,e^{-iF(0,\xi)}+\mathcal{O}(\langle\xi\rangle^{-\frac{3}{2}+}), & \partial_\omega F(0,\xi)<0,\\ \mathcal{O}(\langle\xi\rangle^{-\infty}), & \partial_\omega F(0,\xi)>0.\end{cases} \tag{81}$$

Hence to obtain (77) all we need to show is that $\widetilde{J}(t,\xi)=\mathcal{O}(\langle\xi\rangle^{-\frac{1}{2}+})$ uniformly in t as then by dominated convergence,

$$\langle\xi\rangle^{-\frac{1}{2}-}\widetilde{J}(t)\xrightarrow{L^2(\mathbb{R}^2,d\xi)}\langle\xi\rangle^{-\frac{1}{2}-}\widetilde{J}(\infty),\quad t\to+\infty,$$

that is,

$$\widetilde{I}(t):=\mathcal{F}^{-1}_{\xi\to x}\,\widetilde{J}(t)\xrightarrow{H^{-1/2-}(\mathbb{R}^2)}\mathcal{F}^{-1}_{\xi\to x}\,\widetilde{J}_\infty(t),\quad t\to+\infty.$$

Here the $\mathcal{O}(\langle\xi\rangle^{-\frac{3}{2}+})$ remainder in (81) can be put into $b(t)$ in (77).

The uniform boundedness of $\widetilde{J}(t,\xi)$ is a consequence of the following simple lemma:

Lemma 5.2. *Suppose that $A = A(s,\omega) \in C_c^\infty(\mathbb{R}^2)$ and $G \in C^\infty(\mathbb{R};\mathbb{R})$. Then as $h \to 0$*

$$L(h) := \int_0^\infty \int_{\mathbb{R}} e^{\frac{i}{h}(G(\omega)+s\omega)} A(s,\omega)\, d\omega\, ds = \mathcal{O}\Big(h \log\Big(\frac{1}{h}\Big)\Big). \tag{82}$$

Proof. We define

$$B(\sigma,\omega) := \int_0^\infty e^{is\sigma} A(s,\omega)\, ds, \quad B(\sigma,\omega) = i\sigma^{-1} A(0,\omega) + \mathcal{O}(\sigma^{-2}), \quad |\sigma| \to \infty.$$

Hence,

$$\begin{aligned} L(h) &= \int_{\mathbb{R}} e^{\frac{i}{h}G(\omega)} B\Big(\frac{\omega}{h}, \omega\Big)\, d\omega = h \int_{\mathbb{R}} e^{\frac{i}{h}G(hw)} B(w, hw)\, dw \\ &= \mathcal{O}(h) \int_{|w| \le \frac{C}{h}} \frac{dw}{1+|w|} = \mathcal{O}\Big(h \log\Big(\frac{1}{h}\Big)\Big), \end{aligned}$$

proving (82). (In fact we see that the estimate is sharp: if we take $G \equiv 0$ and A which is *odd* in ω, one does have logarithmic growth.) □

To use the lemma to show the bound $\widetilde{J}(t,\xi) = \mathcal{O}(\langle\xi\rangle^{-\frac{1}{2}+})$, uniformly in $t \ge 0$, it suffices to consider the case $ht \le 2/\gamma$, since otherwise $\widetilde{J}(t,\xi) = \widetilde{J}(\infty,\xi)$. As before, we write $\xi = \eta/h$ where $\eta \in \mathbb{S}^1$. Then

$$\widetilde{J}(t,\xi) = \frac{1}{h} \int_0^\infty \int_{\mathbb{R}} e^{\frac{i}{h}(s\omega - ht\omega - F(\omega,\eta))}\, \chi(ht-s)\, a\Big(\omega, \frac{\eta}{h}\Big)\, \varphi(\omega)\, d\omega\, ds.$$

We now apply Lemma 5.2 with $A(s,\omega) := h^{\alpha-\frac{1}{2}} \chi(ht-s) a(\omega,\eta/h)\varphi(\omega)$, $\alpha > 0$ (and arbitrary), and $G(\omega) = -ht\omega - F(\omega,\eta)$ to obtain, $\widetilde{J}(t) = \mathcal{O}(h^{\frac{1}{2}-\alpha}\log(1/h)) = \mathcal{O}(\langle\xi\rangle^{-\frac{1}{2}+2\alpha})$, which concludes the proof. □

6. Proof of the Main Theorem

In the approach of [Colin de Verdière and Saint-Raymond 2019] the decomposition of $u(t)$ is obtained using (2) and proving that, for φ supported in a neighborhood of 0,

$$P^{-1}(e^{-itP} - 1)\varphi(P) f \xrightarrow{H^{-1/2-}(M)} -(P - i0)^{-1}\varphi(P) f, \quad t \to \infty, \tag{83}$$

which makes formal sense if we think in terms of distributions. The rigorous argument requires finer aspects of Mourre theory developed by Jensen, Mourre, and Perry [Jensen et al. 1984].

Here we take a more geometric approach and use Lemmas 3.3 and 4.1 to study the behavior of $u(t)$. Fix $\delta > 0$ small enough so that the results of Section 2A, as well as (33), hold. Fix $\varphi \in C_c^\infty((-\delta,\delta))$ such that $\varphi = 1$ near 0. By (2), the spectral theorem, and Stone's formula (see for instance [DZ19, Theorem B.8]) we have

$$\begin{aligned} u(t) &= -i \int_0^t e^{-isP}\varphi(P) f\, ds + P^{-1}(e^{-itP} - 1)(1 - \varphi(P)) f \\ &= \frac{1}{2\pi} \int_0^t \int_{\mathbb{R}} e^{-is\omega}\, \varphi(\omega)(u^-(\omega) - u^+(\omega))\, d\omega\, ds + b_1(t), \end{aligned} \tag{84}$$

where $\|b_1(t)\|_{L^2} \le C$ for all $t \ge 0$ and $u^\pm(\omega) := (P - \omega \mp i0)^{-1} f \in H^{-\frac{1}{2}-}(M)$ are defined in Lemma 3.3.

By Lemma 4.1 we have $u^{\pm}(\omega) \in C^{\infty}_{\omega}([-\delta,\delta]; I^0(M;\Lambda^{\pm}_{\omega}))$. The main result (3), (4) then follows from Lemma 5.1. Here we use a pseudodifferential partition of unity to write $u^{\pm}(\omega)$ as a finite sum of oscillatory integrals (75) and the geometric condition (76) follows from Lemmas 2.2 and 2.3. We obtain $u_\infty = -u^+(0)$, which is consistent with (83).

Acknowledgements

This note is a result of a "groupe de travail" on [Colin de Verdière and Saint-Raymond 2019] conducted in Berkeley in February and March of 2018. We would like to thank the participants of that seminar and in particular Thibault de Poyferré for explaining the fluid-mechanical motivation to us. Thanks go also to András Vasy for a helpful discussion of results of [Haber and Vasy 2015]. We are also grateful to Michał Wrochna for pointing out to us a mistake in Lemma 2.1 — see the remark following that lemma — and to the anonymous referee for many suggestions to improve the manuscript. This research was conducted during the period Dyatlov served as a Clay Research Fellow and Zworski was supported by the National Science Foundation grant DMS-1500852 and by a Simons Fellowship.

References

[Colin de Verdière 2018] Y. Colin de Verdière, "Spectral theory of pseudo-differential operators of degree 0 and application to forced linear waves", 2018. To appear in *Anal. PDE*. arXiv

[Colin de Verdière and Saint-Raymond 2019] Y. Colin de Verdière and L. Saint-Raymond, "Attractors for two dimensional waves with homogeneous Hamiltonians of degree 0", *Comm. Pure Appl. Math.* (online publication May 2019).

[Datchev and Dyatlov 2013] K. Datchev and S. Dyatlov, "Fractal Weyl laws for asymptotically hyperbolic manifolds", *Geom. Funct. Anal.* **23**:4 (2013), 1145–1206. MR Zbl

[Dyatlov 2012] S. Dyatlov, "Asymptotic distribution of quasi-normal modes for Kerr–de Sitter black holes", *Ann. Henri Poincaré* **13**:5 (2012), 1101–1166. MR Zbl

[Dyatlov and Guillarmou 2016] S. Dyatlov and C. Guillarmou, "Pollicott–Ruelle resonances for open systems", *Ann. Henri Poincaré* **17**:11 (2016), 3089–3146. MR Zbl

[Dyatlov and Zworski 2016] S. Dyatlov and M. Zworski, "Dynamical zeta functions for Anosov flows via microlocal analysis", *Ann. Sci. Éc. Norm. Supér.* (4) **49**:3 (2016), 543–577. MR Zbl

[Dyatlov and Zworski 2017] S. Dyatlov and M. Zworski, "Ruelle zeta function at zero for surfaces", *Invent. Math.* **210**:1 (2017), 211–229. MR Zbl

[Dyatlov and Zworski 2019] S. Dyatlov and M. Zworski, *Mathematical theory of scattering resonances*, Graduate Studies in Mathematics **200**, American Mathematical Society, 2019. To appear; available at https://tinyurl.com/dyatzwor.

[Haber and Vasy 2015] N. Haber and A. Vasy, "Propagation of singularities around a Lagrangian submanifold of radial points", *Bull. Soc. Math. France* **143**:4 (2015), 679–726. MR Zbl

[Hassell et al. 2004] A. Hassell, R. Melrose, and A. Vasy, "Spectral and scattering theory for symbolic potentials of order zero", *Adv. Math.* **181**:1 (2004), 1–87. MR Zbl

[Hintz and Vasy 2018] P. Hintz and A. Vasy, "The global non-linear stability of the Kerr–de Sitter family of black holes", *Acta Math.* **220**:1 (2018), 1–206. MR Zbl

[Hörmander 1985a] L. Hörmander, *The analysis of linear partial differential operators, III: Pseudodifferential operators*, Grundlehren der Math. Wissenschaften **274**, Springer, 1985. MR Zbl

[Hörmander 1985b] L. Hörmander, *The analysis of linear partial differential operators, IV: Fourier integral operators*, Grundlehren der Math. Wissenschaften **275**, Springer, 1985. MR Zbl

[Jensen et al. 1984] A. Jensen, E. Mourre, and P. Perry, "Multiple commutator estimates and resolvent smoothness in quantum scattering theory", *Ann. Inst. H. Poincaré Phys. Théor.* **41**:2 (1984), 207–225. MR Zbl

[Melrose 1994] R. B. Melrose, "Spectral and scattering theory for the Laplacian on asymptotically Euclidian spaces", pp. 85–130 in *Spectral and scattering theory* (Sanda, Japan, 1992), edited by M. Ikawa, Lecture Notes in Pure and Appl. Math. **161**, Dekker, New York, 1994. MR Zbl

[Nikolaev and Zhuzhoma 1999] I. Nikolaev and E. Zhuzhoma, *Flows on 2-dimensional manifolds: an overview*, Lecture Notes in Math. **1705**, Springer, 1999. MR Zbl

[Ralston 1973] J. V. Ralston, "On stationary modes in inviscid rotating fluids", *J. Math. Anal. Appl.* **44** (1973), 366–383. MR Zbl

[Vasy 2013] A. Vasy, "Microlocal analysis of asymptotically hyperbolic and Kerr–de Sitter spaces", *Invent. Math.* **194**:2 (2013), 381–513. MR Zbl

[Zworski 2016] M. Zworski, "Resonances for asymptotically hyperbolic manifolds: Vasy's method revisited", *J. Spectr. Theory* **6**:4 (2016), 1087–1114. MR Zbl

Received 3 Jul 2018. Revised 18 Apr 2019. Accepted 19 Apr 2019.

SEMYON DYATLOV: dyatlov@math.berkeley.edu
Department of Mathematics, University of California, Berkeley, CA, United States

and

Department of Mathematics, MIT, Cambridge, MA, United States

MACIEJ ZWORSKI: zworski@math.berkeley.edu
Department of Mathematics, University of California, Berkeley, CA, United States

PURE and APPLIED ANALYSIS

Vol. 1, No. 3, 2019
dx.doi.org/10.2140/paa.2019.1.385

msp

CHARACTERIZATION OF EDGE STATES IN PERTURBED HONEYCOMB STRUCTURES

ALEXIS DROUOT

This paper is a mathematical analysis of conduction effects at interfaces between insulators. Motivated by work of Haldane and Raghu (2008), we continue the study of a linear PDE initiated by Fefferman, Lee-Thorp, and Weinstein (2016). This PDE is induced by a continuous honeycomb Schrödinger operator with a line defect.

This operator exhibits remarkable connections between topology and spectral theory. It has essential spectral gaps about the Dirac point energies of the honeycomb background. In a perturbative regime, Fefferman, Lee-Thorp, and Weinstein constructed edge states: time-harmonic waves propagating along the interface, localized transversely. At leading order, these edge states are adiabatic modulations of the Dirac-point Bloch modes. Their envelopes solve a Dirac equation that emerges from a multiscale procedure.

We develop a scattering-oriented approach that derives *all* possible edge states, at arbitrary precision. The key component is a resolvent estimate connecting the Schrödinger operator to the emerging Dirac equation. We discuss topological implications via the computation of the spectral flow, or edge index.

1. Introduction and results

A central branch of condensed matter physics studies energy propagation between dissimilar media. In favorable conditions, the interface acts like a unidirectional channel for electronic transport: the material is conducting in the edge direction but remains insulating transversely. In experiments, this property is remarkably robust: it persists even if the interface becomes bent, sharp or disordered. The first theoretical investigations concerned the quantum Hall effect [Ando et al. 1975; von Klitzing et al. 1980; Halperin 1982; Thouless et al. 1982; Hatsugai 1993]. The research has since focused on topological insulators

MSC2010: primary 35P15; secondary 35P25, 35Q40, 35Q41.
Keywords: edge states, graphene, Dirac points, Schrödinger operators.

[Kane and Mele 2005a; 2005b; Fu et al. 2007; Moore and Balents 2007; Hsieh et al. 2008; Roy 2009; Zhang et al. 2009; Jotzu et al. 2014], together with their applications in electronics, photonics, acoustics, mechanics and geophysics [Khanikaev et al. 2007; Yu et al. 2008; Wang et al. 2008; Singha et al. 2011; Rechtsman et al. 2013; Nash et al. 2015; Brendel et al. 2017; Delplace et al. 2017; Ozawa et al. 2018; Perrot et al. 2018].

Energy transport along the interface may be interpreted as a bifurcation phenomenon. In certain periodic materials, the introduction of an edge forces Bloch modes to bifurcate into edge states: time-harmonic waves propagating along rather than across the edge. This seemingly goes back to Tamm [1932], who looked at bifurcations from local extrema in the band spectrum. Shockley [1939] next studied bifurcations from linear crossings in the band spectrum on a one-dimensional example. In contrast with Tamm's work, Shockley's analysis applies to insulators with narrow energy gaps. It was later discovered that Shockley's states may be topologically protected: *they may persist against large local perturbations.*

Honeycomb structures are invariant under $\frac{2\pi}{3}$-rotation and spatial inversion. These symmetries generate Dirac points: conical degeneracies in the band spectrum. Impurities breaking spatial inversion split the dispersion surfaces away and open energy gaps: the material transits from a metal to an insulator. Here we analyze interface effects at the junction of two such insulators.

Motivated by [Haldane and Raghu 2008; Raghu and Haldane 2008], Fefferman, Lee-Thorp and Weinstein [Fefferman et al. 2016b] introduced a PDE that models parity-breaking perturbations of a *continuous* honeycomb lattice (see Section 1A–1B). The perturbed operator exhibits (a) an edge that separates two asymptotically periodic near-honeycomb structures; (b) gaps in the essential spectrum centered at Dirac point energies of the honeycomb background. Under a spectral condition on the unperturbed operator (see [Fefferman et al. 2016b, §1.3] and Section 1C), Fefferman, Lee-Thorp and Weinstein designed edge states as adiabatic modulations of the Dirac-point Bloch modes. Their envelopes are eigenvectors of a Dirac operator produced via a multiscale procedure. See [Fefferman et al. 2016b, Theorem 7.3].

Here, we follow instead a scattering approach. We recover the results of [Fefferman et al. 2016a; 2016b]. In addition, we obtain

- a resolvent estimate connecting the initial PDE to the emerging Dirac equation,
- the complete characterization of edge states in the energy gap,
- full expansions of the edge states at all order in the size of the perturbation.

See Sections 1E and 3C for precise statements.

The full identification of edge states represents the most significant advance. It allows for topological interpretation of the results. In Section 1G, we compute the signed number of eigenvalues that move across Dirac point energies when the edge-parallel quasimomentum runs from 0 to 2π. This is a topological invariant of the system — called spectral flow or edge index — and it vanishes here. This calculation confirms numerical simulations [Raghu and Haldane 2008; Fefferman et al. 2016a; Lee-Thorp et al. 2019]. It corroborates the prediction of the Kitaev table [Kitaev 2009; Ryu et al. 2010], combined with the bulk-edge correspondence: breaking spatial inversion while keeping time-reversal invariance does not create protected edge states.

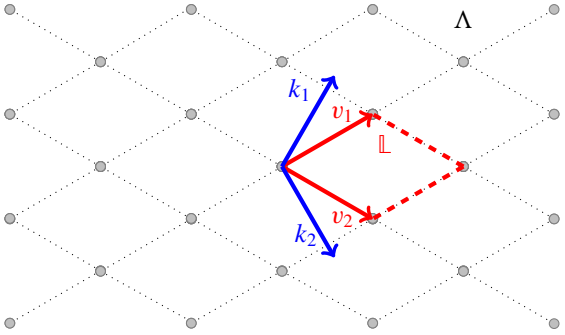

Figure 1. The equilateral lattice with its generating vectors v_1, v_2 and dual vectors k_1, k_2 together with the fundamental cell $\mathbb{L}$.

In the last part of the work, we consider a magnetic analog of the operator studied in [Fefferman et al. 2016a; 2016b], similar to those of [Raghu and Haldane 2008; Haldane and Raghu 2008; Lee-Thorp et al. 2019]. It models time-reversal breaking instead of parity breaking. We show that the corresponding spectral flow equals either 2 or -2. This confirms the existence of at least two topologically protected, unidirectionally propagating waves along the edge; see [Haldane and Raghu 2008] and the Kitaev table [Kitaev 2009; Ryu et al. 2010], as well as the numerical results [Raghu and Haldane 2008; Lee-Thorp et al. 2019].

1A. ***Periodic operators and Dirac points.*** We start with a description of honeycomb potentials as in [Fefferman and Weinstein 2012]. Let Λ be the equilateral $\mathbb{Z}^2$-lattice. It is generated by two vectors v_1 and v_2, given in canonical coordinates by

$$v_1 = a\begin{bmatrix}\sqrt{3}\\ 1\end{bmatrix}, \quad v_2 = a\begin{bmatrix}\sqrt{3}\\ -1\end{bmatrix}, \tag{1-1}$$

where $a > 0$ is a constant such that $\mathrm{Det}[v_1, v_2] = 1$. The dual basis k_1, k_2 consists of two vectors in $(\mathbb{R}^2)^*$ which satisfy $\langle k_i, v_j\rangle = \delta_{ij}$. (See Figure 1.) The dual lattice is $\Lambda^* = \mathbb{Z}k_1 \oplus \mathbb{Z}k_2$. The corresponding fundamental cell and dual fundamental cell are

$$\mathbb{L} \stackrel{\text{def}}{=} \{sv_1 + s'v_2 : s, s' \in [0, 1)\}, \quad \mathbb{L}^* \stackrel{\text{def}}{=} \{\tau k_1 + \tau' k_2 : \tau, \tau' \in [0, 2\pi)\}. \tag{1-2}$$

Definition 1.1. We say that $V \in C^\infty(\mathbb{R}^2, \mathbb{R})$ is a honeycomb potential if:

- V is Λ-periodic: $V(x + w) = V(x)$ for $w \in \Lambda$.
- V is even: $V(x) = V(-x)$.
- V is invariant under the $\frac{2\pi}{3}$-rotation

$$V(Rx) = V(x), \quad R \stackrel{\text{def}}{=} \frac{1}{2}\begin{bmatrix}-1 & \sqrt{3}\\ -\sqrt{3} & -1\end{bmatrix}.$$

A simple example of honeycomb potential is the periodization of a radial function over the lattice

$$\left(\frac{v_1 + v_2}{3} + \Lambda\right) \cup \left(\frac{2v_1 + 2v_2}{3} + \Lambda\right);$$

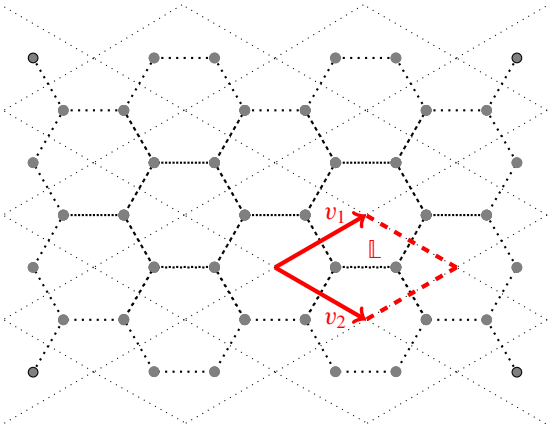

Figure 2. If each gray circle supports the same radial function (with respect to the center of the circle), the resulting potential has the honeycomb symmetry.

see Figure 2. Given a honeycomb potential V, we will study spatially delocalized perturbations of the (unbounded) Schrödinger operator

$$P_0 \overset{\text{def}}{=} -\Delta + V : L^2(\mathbb{R}^2, \mathbb{C}) \to L^2(\mathbb{R}^2, \mathbb{C}),$$

with domain $H^2(\mathbb{R}^2, \mathbb{C})$. This operator is periodic with respect to Λ. This allows us to apply Floquet–Bloch theory; see [Reed and Simon 1978, §XIII]: P_0 leaves the space

$$L^2_\xi \overset{\text{def}}{=} \{u \in L^2_{\text{loc}}(\mathbb{R}^2, \mathbb{C}) : u(x+w) = e^{i\langle \xi, w\rangle} u(x),\ w \in \Lambda\}, \quad \xi \in \mathbb{R}^2,$$

invariant. The space L^2_ξ is Hilbertian when equipped with the Hermitian form

$$\langle f, g\rangle_{L^2_\xi} \overset{\text{def}}{=} \int_{\mathbb{L}} \overline{f(x)} g(x)\, dx.$$

Let $P_0(\xi)$ be formally equal to $P_0 = -\Delta + V$, but acting on L^2_ξ. It has compact resolvent and discrete spectrum — denoted below by $\Sigma_{L^2_\xi}(P_0(\xi))$ — depending on ξ:

$$\lambda_{0,1}(\xi) \le \lambda_{0,2}(\xi) \le \cdots \le \lambda_{0,j}(\xi) \le \cdots.$$

The maps $\xi \in \mathbb{R}^2 \mapsto \lambda_{0,j}(\xi)$ are called dispersion surfaces of P_0. The L^2-spectrum of P_0 consists of the ranges of the dispersion surfaces: it equals

$$\Sigma_{L^2}(P_0) = \bigcup_{\xi \in \mathbb{R}^2} \Sigma_{L^2_\xi}(P_0(\xi)) = \{\lambda_{0,j}(\xi) : j \ge 1,\ \xi \in \mathbb{R}^2\}.$$

We now discuss Dirac points. Roughly speaking, they correspond to the conical degeneracies in the band spectrum of P_0.

Definition 1.2. A pair $(\xi_\star, E_\star) \in \mathbb{R}^2 \times \mathbb{R}$ is a Dirac point of $P_0 = -\Delta + V$ if:

(i) $E_\star$ is an $L^2_{\xi_\star}$-eigenvalue of $P_0(\xi_\star)$ of multiplicity 2;

(ii) There exists an orthonormal basis $\{\phi_1, \phi_2\}$ of $\ker_{L^2_{\xi_\star}}(P_0(\xi_\star) - E_\star)$ such that

$$\phi_1(Rx) = e^{2i\pi/3}\phi_1(x), \quad \phi_2(x) = \overline{\phi_1(-x)}, \quad \phi_2(Rx) = e^{-2i\pi/3}\phi_2(x). \tag{1-3}$$

(iii) There exist $j_\star \geq 1$ and $v_F > 0$ such that for ξ close to $\xi_\star$,

$$\lambda_{0,j_\star}(\xi) = E_\star - v_F \cdot |\xi - \xi_\star| + O(\xi - \xi_\star)^2,$$
$$\lambda_{0,j_\star+1}(\xi) = E_\star + v_F \cdot |\xi - \xi_\star| + O(\xi - \xi_\star)^2.$$

When V is a honeycomb potential, [Fefferman and Weinstein 2012] showed that $P_0 = -\Delta + V$ generically admits Dirac points $(\xi_\star, E_\star)$. We refer to that paper for details and to Section 2C for a review of their results. Because of (1-3), $(\xi_\star, E_\star)$ must satisfy

$$\xi_\star \in \{\xi_\star^A, \xi_\star^B\} \bmod 2\pi\Lambda^*, \quad \xi_\star^A \stackrel{\text{def}}{=} \tfrac{2\pi}{3}(2k_1 + k_2), \quad \xi_\star^B \stackrel{\text{def}}{=} \tfrac{2\pi}{3}(k_1 + 2k_2). \tag{1-4}$$

See Figure 3. Symmetries impose that $(\xi_\star^A, E_\star)$ is a Dirac point of P_0 if and only if $(\xi_\star^B, E_\star)$ is a Dirac point of P_0. We call the pair (ϕ_1, ϕ_2) of (1-3) a Dirac eigenbasis.

As observed in [Fefferman and Weinstein 2012], Dirac points are stable against small perturbations preserving spatial inversion (parity) and time-reversal symmetry (conjugation). Conversely, breaking parity (while keeping conjugation invariance) generically opens spectral gaps about Dirac point energies. For $\delta \neq 0$, we introduce the operator

$$\begin{gathered} P_\delta \stackrel{\text{def}}{=} P_0 + \delta W = -\Delta + V + \delta W, \quad \text{where} \\ W \in C^\infty(\mathbb{R}^2, \mathbb{R}), \qquad W(x+w) = W(x), \quad w \in \Lambda, \qquad W(-x) = -W(x). \end{gathered} \tag{1-5}$$

We will assume in the rest of the paper that the nondegeneracy condition

$$\vartheta_\star \stackrel{\text{def}}{=} \langle \phi_1, W\phi_1 \rangle_{L^2_{\xi_\star}} \neq 0 \tag{1-6}$$

holds. This condition is generic in the sense that it excludes only a hyperplane of potentials W in the space of odd, smooth, Λ-periodic functions. Under (1-6), if $(\xi_\star, E_\star)$ is a Dirac point of P_0, then the operator $P_\delta(\xi_\star)$ (equal to P_δ, but acting on $L^2_{\xi_\star}$) admits an $L^2_{\xi_\star}$-spectral gap centered at $E_\star$:

$$\operatorname{dist}(\Sigma_{L^2_{\xi_\star}}(P_\delta(\xi_\star)), E_\star) = \vartheta_F \cdot \delta + O(\delta^2), \quad \vartheta_F \stackrel{\text{def}}{=} |\vartheta_\star|.$$

This gap has width $2\vartheta_F \cdot \delta + O(\delta^2)$; see Figure 3. This is a simple fact proved via perturbation analysis; see, e.g., [Fefferman and Weinstein 2012, Remark 9.2] or Section 4B. Whether this $L^2_{\xi_\star}$-spectral gap extends to a *global* L^2-gap of P_δ depends on the global behavior of the dispersion surfaces of P_0; see [Fefferman et al. 2016b, §1.3 and §8]. When it does, the operators $P_{\pm\delta}$ describe insulators at energy $E_\star$ with a narrow gap centered at $E_\star$. These materials are parity-breaking perturbations of the *metal* modeled by P_0.

1B. ***Edges and the model.*** We now describe the model of Fefferman, Lee-Thorp, and Weinstein [Fefferman et al. 2016a; 2016b] for honeycomb operators with an edge. Fix $v = a_1 v_1 + a_2 v_2 \in \Lambda$, with $a_1, a_2 \in \mathbb{Z}$

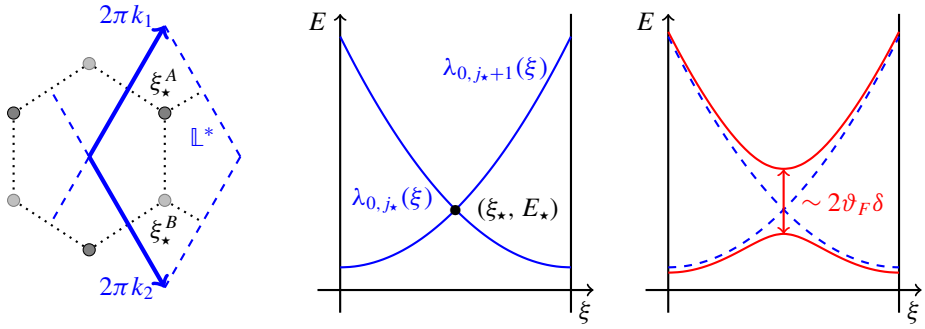

Figure 3. The picture on the left represents the Dirac points $\xi_\star^A$ and $\xi_\star^B$ inside a dual fundamental cell $\mathbb{L}^*$. The two pictures on the right represent the bifurcation of a Dirac point $(\xi_\star, E_\star)$ to an open gap on a one-dimensional section of the Brillouin zone.

relatively prime, representing the direction of an edge $\mathbb{R}v$. We introduce $v' \in \Lambda$ and $k, k' \in \Lambda^*$ such that

$$\begin{gathered} v' \overset{\text{def}}{=} b_1 v_1 + b_2 v_2, \quad a_1 b_2 - a_2 b_1 = 1, \quad b_1, b_2 \in \mathbb{Z}, \\ k \overset{\text{def}}{=} b_2 k_1 - b_1 k_2, \quad k' \overset{\text{def}}{=} -a_2 k_1 + a_1 k_2. \end{gathered} \tag{1-7}$$

The pairs (v, v') and (k, k') are dual to one another and span Λ and Λ^*. See Section 2E.

Recall that $P_{\pm\delta} = -\Delta + V \pm \delta W$. Fefferman, Lee-Thorp, and Weinstein [Fefferman et al. 2016a; 2016b] analyzed an operator $\mathscr{P}_\delta$ that describes an adiabatic transition from $P_{-\delta}$ to P_δ transversely to the edge $\mathbb{R}v$. Specifically,

$$\mathscr{P}_\delta \overset{\text{def}}{=} P_0 + \delta \cdot \kappa_\delta \cdot W = -\Delta + V + \delta \cdot \kappa_\delta \cdot W.$$

Above, the function $\kappa_\delta \in C^\infty(\mathbb{R}^2, \mathbb{R})$ is an adiabatic modulation of a domain wall $\kappa \in C^\infty(\mathbb{R}, \mathbb{R})$ along $\mathbb{R}v$:

$$\kappa_\delta(x) = \kappa(\delta\langle k', x\rangle), \quad \exists L > 0, \ \ \kappa(t) = \begin{cases} -1 & \text{when } t \le -L, \\ 1 & \text{when } t \ge L. \end{cases} \tag{1-8}$$

The operator $\mathscr{P}_\delta$ is a Schrödinger operator with potential represented in Figure 9. It models the soft junction of two insulators modeled by $P_{\pm\delta}$ along the interface $\mathbb{R}v$.

Although $\mathscr{P}_\delta$ is not periodic with respect to Λ, it is periodic with respect to $\mathbb{Z}v$ because $\langle k', v\rangle = 0$. For every $\zeta \in \mathbb{R}$, $\mathscr{P}_\delta$ acts as an unbounded operator on

$$L^2[\zeta] \overset{\text{def}}{=} \big\{u \in L^2_{\text{loc}}(\mathbb{R}^2, \mathbb{C}) : u(x+v) = e^{i\zeta} u(x), \ \textstyle\int_{\mathbb{R}^2/\mathbb{Z}v} |u(x)|^2\, dx < \infty\big\}, \tag{1-9}$$

with domain $H^2[\zeta]$ — defined according to (1-9). Let $\mathscr{P}_\delta[\zeta]$ be the resulting operator.

We continue the analysis of [Fefferman et al. 2016a; 2016b]: we study the electronic properties of the material modeled by $\mathscr{P}_\delta$. We investigate whether energy propagates along the edge $\mathbb{R}v$. This boils down to studying *edge states* of $\mathscr{P}_\delta$. These are time-harmonic waves propagating along $\mathbb{R}v$ and localized transversely to $\mathbb{R}v$. Mathematically, they are the $L^2[\zeta]$-eigenvectors of $\mathscr{P}_\delta[\zeta]$. Such states correspond to diffusionless electronic channels along $\mathbb{R}v$; they have great potential in technological applications.

1C. ***The no-fold condition of Fefferman, Lee-Thorp, and Weinstein.*** We set $\zeta_\star = \langle \xi_\star, v\rangle$ and $\zeta_\star^J = \langle \xi_\star^J, v\rangle$. Thanks to (1-4),

$$\zeta_\star^A = \tfrac{2\pi}{3}(2a_1 + a_2), \quad \zeta_\star^B = \tfrac{2\pi}{3}(a_1 + 2a_2). \tag{1-10}$$

Hence, $\zeta_\star \in \{0, \frac{2\pi}{3}, \frac{4\pi}{3}\} \bmod 2\pi\mathbb{Z}$. Recall the no-fold condition [Fefferman et al. 2016b, §1.3].

Definition 1.3. The no-fold condition holds along the edge $\mathbb{R}v$ at $\zeta_\star$ if

$$\forall j \geq 1,\ \forall \tau \in \mathbb{R}, \quad \lambda_{0,j}(\zeta_\star k + \tau k') = E_\star \quad \Longrightarrow \quad j \in \{j_\star, j_\star + 1\} \quad \text{and} \quad \tau = \langle \xi_\star, v'\rangle \bmod 2\pi.$$

The essential spectrum of $\mathscr{P}_\delta[\zeta_\star]$ is obtained from the (essential) spectra of the bulk operators $P_{\pm\delta}[\zeta_\star]$ (the operators formally equal to $P_{\pm\delta}$, but acting on $L^2[\zeta_\star]$). These are conjugated under spatial inversion. Therefore they have the same spectrum. From Floquet–Bloch theory,

$$\Sigma_{L^2[\zeta_\star],\mathrm{ess}}(\mathscr{P}_\delta[\zeta_\star]) = \Sigma_{L^2[\zeta_\star]}(P_\delta[\zeta_\star]) = \bigcup_{\xi \in \zeta_\star k + \mathbb{R}k'} \Sigma_{L^2_\xi}(P_\delta(\xi)).$$

If $(\xi_\star, E_\star)$ is a Dirac point of P_0 and $\vartheta_\star \neq 0$, then for small δ, $P_{\pm\delta}(\xi)$ has an L^2_ξ-spectral gap centered at $E_\star$ when ξ is $O(\delta)$-away from $\xi_\star$ — see, e.g., Section 4B. The no-fold condition requires this gap to extend to an $L^2[\zeta_\star]$-spectral gap of $P_{\pm\delta}[\zeta_\star]$.

The no-fold condition holds for $|V|_\infty$ sufficiently small and the zigzag edge $a_1 = 1,\ a_2 = 0$ [Fefferman et al. 2016b, Theorem 8.2]. It holds for $|V|_\infty$ sufficiently large and edges satisfying $a_1 \neq a_2 \bmod 3$ [Fefferman et al. 2018, Corollary 6.3]. It may fail in physically relevant cases. See, e.g., the case of certain low-contrast potentials and the zigzag edge [Fefferman et al. 2016b, Theorem 8.4] and armchair-type edges $v = a_1 v_1 + a_2 v_2$, where $a_1 - a_2 = 0 \bmod 3$ [Fefferman et al. 2018, Remark 6.5] or Section 2E. In particular, if the no-fold condition holds, (1-10) and $a_1 - a_2 \neq 0 \bmod 3$ prescribe the possible values of $\zeta_\star$:

$$\zeta_\star \in \{\zeta_\star^A, \zeta_\star^B\} = \{\tfrac{2\pi}{3}, \tfrac{4\pi}{3}\} \bmod 2\pi\mathbb{Z}.$$

1D. ***The multiscale approach of [Fefferman et al. 2016b] and the Dirac operator.*** Let $(\xi_\star, E_\star)$ be a Dirac point of P_0 and (ϕ_1, ϕ_2) be a Dirac eigenbasis (see Definition 1.2). The map

$$\eta \in \mathbb{R}^2 \mapsto 2\langle \phi_1, (\eta \cdot D_x)\phi_2\rangle \in \mathbb{C} \tag{1-11}$$

is linear. Because of rotational invariance of $P_0 = -\Delta + V$, the map (1-11) acts (as an application from $\mathbb{C}$ to $\mathbb{C}$) like a complex multiplication:

$$\exists \nu_\star \in \mathbb{C} \setminus \{0\}, \quad \forall \eta \in \mathbb{R}^2 \equiv \mathbb{C}, \quad \nu_\star \eta = 2\langle \phi_1, (\eta \cdot D_x)\phi_2\rangle_{L^2_{\xi_\star}}.$$

See Section 2C. Recall that $\vartheta_\star = \langle \phi_1, W\phi_1\rangle_{L^2_{\xi_\star}} \neq 0$ and that κ satisfies (1-8). In this section, we review the role of the (unbounded) Dirac operator

$$\not{D}_\star = \begin{bmatrix} 0 & \nu_\star k' \\ \overline{\nu_\star k'} & 0 \end{bmatrix} D_t + \vartheta_\star \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \kappa : L^2(\mathbb{R}, \mathbb{C}^2) \to L^2(\mathbb{R}, \mathbb{C}^2)$$

in the analysis of Fefferman, Lee-Thorp, and Weinstein [Fefferman et al. 2016b].

When $\vartheta_\star \neq 0$, [loc. cit.] produces arbitrarily accurate quasimodes of $\mathscr{P}_\delta[\zeta_\star]$ via a multiscale approach. These are pairs $(u_\delta, E_\delta) \in H^2_{\zeta_\star} \times \mathbb{R}$ satisfying

$$(\mathscr{P}_\delta[\zeta_\star] - E_\delta)u_\delta = O_{L^2[\zeta_\star]}(\delta^\infty), \quad E_\delta = E_\star + \delta E_1 + O(\delta^2).$$

They are power series in δ whose coefficients solve a hierarchy of equations of orders $1, \delta, \delta^2, \ldots$. The operator $\not{D}_\star$ appears in the equation of order δ. This equation admits a solution if and only if E_1 is an eigenvalue of $\not{D}_\star$; see [loc. cit., §6].

The operator $\not{D}_\star$ has essential spectrum equal to $(-\infty, \vartheta_F] \cup [\vartheta_F, \infty)$. It has an odd number of eigenvalues $\{\vartheta_j\}_{j=-N}^N$ in $(-\vartheta_F, \vartheta_F)$, simple and symmetric about 0:

$$\vartheta_{-N} < \cdots < \vartheta_{-1} < \vartheta_0 = 0 < \vartheta_1 < \cdots < \vartheta_N, \quad \vartheta_{-j} = -\vartheta_j.$$

In particular, 0 is always an eigenvalue of $\not{D}_\star$. We refer to see Section 3B for details.

When the no-fold condition holds, [loc. cit.] uses a sophisticated Lyapounov–Schmidt reduction to prove that each eigenvalue ϑ_j of $\not{D}_\star$ seeds an $L^2[\zeta_\star]$-eigenvalue of $\mathscr{P}_\delta[\zeta_\star]$, with energy $E_\star + \delta\vartheta_j + O(\delta^2)$. They show that to leading order, the corresponding eigenvector equals the first term produced by the multiscale approach: it is

$$\alpha_1(\delta\langle k', x\rangle) \cdot \phi_1(x) + \alpha_2(\delta\langle k', x\rangle) \cdot \phi_2(x) + O_{H^2_{\zeta_\star}}(\delta^{1/2}), \quad (\not{D}_\star - \vartheta_j)\begin{bmatrix}\alpha_1 \\ \alpha_2\end{bmatrix} = 0.$$

In other words, they validate mathematically the formal multiscale procedure at leading order. But some questions persist:

- Is the multiscale procedure rigorously valid at all orders?
- Do the eigenvalues of $\not{D}_\star$ seed *all* eigenvalues of $\mathscr{P}_\delta[\zeta_\star]$ near $E_\star$?
- How can the relation between $\mathscr{P}_\delta[\zeta_\star]$ and $\not{D}_\star$ be clarified?

The present work responds to these questions.

1E. ***Results.*** Our first result relates the resolvents of $\mathscr{P}_\delta[\zeta_\star]$ and $\not{D}_\star$. It requires the operator Π and its adjoint Π^*, defined as

$$\Pi : L^2(\mathbb{R}^2/\mathbb{Z}v, \mathbb{C}^2) \to L^2(\mathbb{R}, \mathbb{C}^2), \qquad (\Pi f)(t) \stackrel{\text{def}}{=} \int_0^1 f(sv + tv')\, ds,$$
$$\Pi^* : L^2(\mathbb{R}, \mathbb{C}^2) \to L^2(\mathbb{R}^2/\mathbb{Z}v, \mathbb{C}^2), \quad (\Pi^* g)(x) \stackrel{\text{def}}{=} g(\langle k', x\rangle),$$

and the dilation $\mathcal{U}_\delta$ defined as

$$\mathcal{U}_\delta : L^2(\mathbb{R}, \mathbb{C}^2) \to L^2(\mathbb{R}, \mathbb{C}^2), \quad (\mathcal{U}_\delta f)(t) \stackrel{\text{def}}{=} f(\delta t).$$

Recall that V is a honeycomb potential — see Definition 1.1; $W \in C^\infty(\mathbb{R}^2, \mathbb{R})$ breaks spatial inversion — see (1-5); and $\kappa \in C^\infty(\mathbb{R}, \mathbb{R})$ is a domain wall function — see (1-8). We make the following assumptions:

(H1) $(\xi_\star, E_\star)$ is a Dirac point of $P_0 = -\Delta + V$ — see Definition 1.2 — with $\xi_\star \in \mathbb{L}^*$.

(H2) The no-fold condition — Definition 1.3 — holds.

(H3) The nondegeneracy assumption $\vartheta_\star \neq 0$ holds — see (1-6).

Theorem 1.4. *Assume* (H1)–(H3) *hold and fix* $\epsilon > 0$. *There exists* $\delta_0 > 0$ *such that if*

$$\delta \in (0, \delta_0), \quad z \in \mathbb{D}(0, \vartheta_F - \epsilon), \quad \operatorname{dist}(\Sigma_{L^2}(\not{D}_\star), z) \geq \epsilon, \quad \lambda = E_\star + \delta z$$

then $\mathscr{P}_\delta[\zeta_\star] - \lambda$ *is invertible and*

$$(\mathscr{P}_\delta[\zeta_\star] - \lambda)^{-1} = \frac{1}{\delta} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^\top \cdot \Pi^* \mathcal{U}_\delta \cdot (\not{D}_\star - z)^{-1} \cdot \mathcal{U}_\delta^{-1} \Pi \cdot \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}} + \mathscr{O}_{L^2[\zeta_\star]}(\delta^{-1/3}). \tag{1-12}$$

The leading-order term in (1-12) comes with a coefficient $1/\delta$: the remainder term $\mathscr{O}_{L^2[\zeta_\star]}(\delta^{-1/3})$ is subleading when $z \in \mathbb{D}(0, \vartheta_F - \epsilon)$. Hence, Theorem 1.4 shows that the resolvents of $\mathscr{P}_\delta[\zeta_\star]$ and of $\not{D}_\star$ behave similarly, after suitable conjugations.

Theorem 1.4 applies to a spectral range that spans — modulo ϵ — the entire spectral gap of $\mathscr{P}_\delta[\zeta_\star]$ about $E_\star$. The next result describes the spectrum of $\mathscr{P}_\delta[\zeta_\star]$ in the essential spectral gap in terms of the eigenvalues

$$\vartheta_{-N} < \cdots < \vartheta_{-1} < \vartheta_0 = 0 < \vartheta_1 < \cdots < \vartheta_N$$

of the Dirac operator $\not{D}_\star$. Let X be the function space equal to

$$\{f \in C^\infty(\mathbb{R}^2 \times \mathbb{R}, \mathbb{C}) : \forall t \in \mathbb{R},\ f(\cdot, t) \in L^2_{\xi_\star} \text{ and } \exists a > 0,\ \sup e^{a|t|} |f(x,t)| < \infty\}. \tag{1-13}$$

Corollary 1.5. *Assume* (H1)–(H3) *hold and fix* $\vartheta_\sharp \in (\vartheta_N, \vartheta_F)$. *There exists* $\delta_0 > 0$ *such that for* $\delta \in (0, \delta_0)$ *the operator* $\mathscr{P}_\delta[\zeta_\star]$ *has exactly* $2N+1$ *eigenvalues* $\{E_{\delta,j}\}_{j \in [-N,N]}$ *in* $[E_\star - \vartheta_\sharp \delta, E_\star + \vartheta_\sharp \delta]$ *that are all simple.*

The associated eigenpairs $(E_{\delta,j}, u_{\delta,j})$ *admit full two-scale expansions in powers of* δ:

$$\begin{aligned} E_{\delta,j} &= E_\star + \vartheta_j \cdot \delta + a_2 \cdot \delta^2 + \cdots + a_M \cdot \delta^M + O(\delta^{M+1}), \\ u_{\delta,j}(x) &= f_0(x, \delta\langle k', x\rangle) + \delta \cdot f_1(x, \delta\langle k', x\rangle) + \cdots + \delta^M \cdot f_M(x, \delta\langle k', x\rangle) + o_{H^k}(\delta^M). \end{aligned}$$

In the above:

- M *and* k *are any integers*; H^k *is the* k*-th order Sobolev space.*
- *The terms* $a_m \in \mathbb{R}$, $f_m \in X$ *are recursively constructed via multiscale analysis.*
- *The leading-order term* f_0 *satisfies*

$$f_0(x,t) = \alpha_1(t) \cdot \phi_1(x) + \alpha_2(t) \cdot \phi_2(x), \quad (\not{D}_\star - \vartheta_j) \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = 0.$$

This corollary (a) mathematically validates the multiscale procedure of [Fefferman et al. 2016b] at all orders in δ, and (b) shows that all eigenvectors of $\mathscr{P}_\delta[\zeta_\star]$ are induced by the modes of $\not{D}_\star$. See Figures 4 and 5. In particular, (a) improves the result of Fefferman, Lee-Thorp, and Weinstein [loc. cit.] to arbitrary order in δ. From a general point of view, (b) represents the most important advance. It characterizes edge states topologically. It opens the way for mathematical proofs of the bulk-edge correspondence in *continuous* honeycomb structures. See Section 1G and [Drouot 2019] for further details.

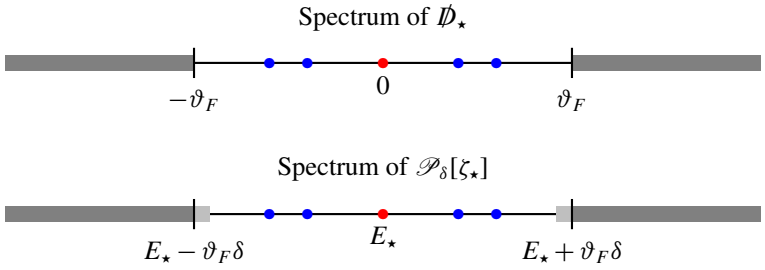

Figure 4. Eigenvalues of $\not{D}_\star$ in $(-\vartheta_F, \vartheta_F)$ (top) and eigenvalues of $\mathscr{P}_\delta$ in the spectral gap containing $E_\star$ (bottom). An approximate rescaling equal to $z \mapsto E_\star + \delta z + O(\delta^2)$ maps the top to the bottom. The red dots represent the zero eigenvalue of $\not{D}_\star$ and the corresponding one for $\mathscr{P}_\delta$. Theorem 1.4 and Corollary 1.5 do not apply in the lighter gray area near the essential spectrum.

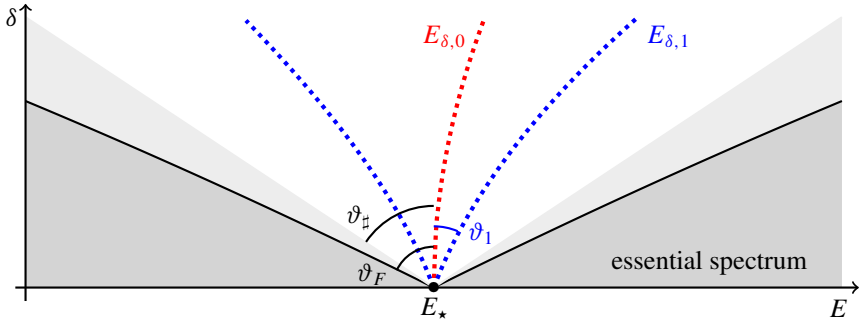

Figure 5. Discrete eigenvalues of $\not{D}_\star$ seed the bifurcation of eigenvalues of $\mathscr{P}_\delta$ (red dotted curves) from the Dirac point energy $E_\star$ (at $\delta = 0$) of P_0 as δ increases away from zero. The slopes of these curves at $\delta = 0$ (blue lines) are given by the eigenvalues of $\not{D}_\star$.

1F. ***Extension to quasimomenta near $\zeta_\star$.*** Corollary 1.5 predicts that for $\delta \in (0, \delta_0)$, $\mathscr{P}_\delta[\zeta_\star]$ has precisely $2N+1$ eigenvalues near $E_\star$. A general perturbation argument shows that $\mathscr{P}_\delta[\zeta]$ also has $2N+1$ eigenvalues for ζ close enough to $\zeta_\star$. However this argument does not specify quantitatively how close ζ needs to be to $\zeta_\star$.

We prove generalizations of Theorem 1.4 and Corollary 1.5 that hold for ζ at distance $O(\delta)$ from $\zeta_\star$; see Section 3C for statements. We show that the eigenvalues of $\mathscr{P}_\delta[\zeta_\star + \mu\delta]$ lying near $E_\star$ and of the Dirac operator

$$\not{D}(\mu) \overset{\text{def}}{=} \begin{bmatrix} 0 & v_\star k' \\ v_\star k' & 0 \end{bmatrix} D_t + \mu \begin{bmatrix} 0 & v_\star \ell \\ v_\star \ell & 0 \end{bmatrix} + \vartheta_\star \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \kappa, \quad \ell \overset{\text{def}}{=} k - \frac{\langle k, k' \rangle}{|k'|^2} k',$$

are $O(\delta^2)$-away after the rescaling $z \mapsto E_\star + \delta z$.

Interestingly enough, the spectrum of $\not{D}(\mu)$ can be derived from that of $\not{D}_\star = \not{D}(0)$; see Section 3B and Figure 6. We observe that $\not{D}(\mu)$ has a topologically protected mode that bifurcates linearly from the zero

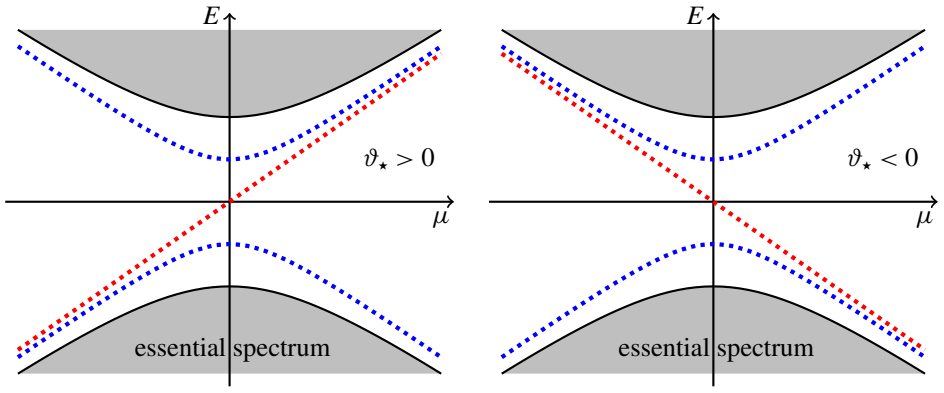

Figure 6. The spectrum of $\not{D}(\mu)$ as a function of μ. The topologically protected eigenvalue (in red) bifurcate linearly, while the nontopologically protected eigenvalues (in blue) bifurcate quadratically.

mode of $\not{D}_\star$. This suggests that under the $\mathscr{P}_\delta$ time-dependent evolution, L^2-wave packets formed from the topologically protected mode of $\not{D}(\mu)$ propagate dispersionless along the edge for a very long time.

All other modes of $\not{D}(\mu)$ are nontopologically protected and bifurcate quadratically from the modes of $\not{D}_\star$. L^2-wave packets formed from such modes should have a shorter lifetime. This suggests that topologically protected modes are more robust even in the time-dependent situation.

1G. ***A topological perspective.*** Recall that $k' \in \Lambda^*$ is the dual direction transverse to an edge $\mathbb{R}v$ and that $\lambda_{0,j}(\xi)$ are the dispersion surfaces of a honeycomb Schrödinger operator P_0. Let $(\xi_\star, E_\star) = (\xi_\star, \lambda_{0,j_\star}(\xi_\star))$ denote a Dirac point of P_0. We introduce an assumption (H4) that extends (H3) to values $\zeta \neq \zeta_\star$. It asks for the $j_\star$-th $L^2[\zeta]$-gap of $P_0[\zeta]$ to be open when $\zeta \notin \left\{\frac{2\pi}{3}, \frac{4\pi}{3}\right\} \bmod 2\pi\mathbb{Z}$.

(H4) For every $\zeta \notin \left\{\frac{2\pi}{3}, \frac{4\pi}{3}\right\} \bmod 2\pi\mathbb{Z}$, for every $\tau, \tau' \in \mathbb{R}$,

$$\lambda_{0,j_\star}(\zeta k + \tau k') < \lambda_{0,j_\star+1}(\zeta k + \tau' k').$$

Assumption (H4) holds for nonarmchair-type edges ($a_1 \neq a_2 \bmod 3$) and high-contrast potentials: see [Fefferman et al. 2018, Theorem 6.1 and Remark 6.5]. This follows from two general phenomena:

- Schrödinger operators with multiple-well potentials approach their tight binding limits as the depth of the wells increases [Harrell 1979; Helffer and Sjöstrand 1984; 1985; 1987; Simon 1984; Martinez 1987; 1988; Outassourt 1987; Carlsson 1990; Fefferman et al. 2018; Fefferman and Weinstein 2018];
- Wallace's tight binding model of honeycomb lattices [1947] satisfies a suitable version of (H4).

When (H1)–(H4) hold and δ is sufficiently small, the $j_\star$-th $L^2[\zeta]$-gap of $P_\delta[\zeta]$ is open. This allows us to define the spectral flow of the family

$$\zeta \in [0, 2\pi] \mapsto \mathscr{P}_\delta[\zeta]$$

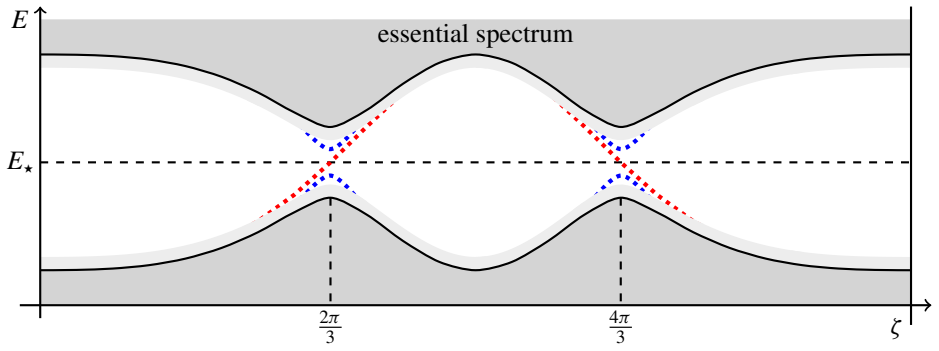

Figure 7. The spectrum of $\mathscr{P}_\delta[\zeta]$ as a function of ζ. The dark gray region represents the essential spectrum. The dotted curves are the eigenvalues of $\mathscr{P}_\delta[\zeta]$ (the edge state energies). Zooming about δ^{-1} times near $\left(\frac{2\pi}{3}, E_\star\right)$ or $\left(\frac{4\pi}{3}, E_\star\right)$ produces Figure 6. Because of complex conjugation, $\vartheta_\star^A = -\vartheta_\star^B$: near $\frac{2\pi}{3}$ (resp. $\frac{4\pi}{3}$), the red curves move upwards (resp. downwards). This results in a spectral flow cancellation.

in the $j_\star$-th $L^2[\zeta]$-gap. It is the signed number of $L^2[\zeta]$-eigenvalues of $\mathscr{P}_\delta[\zeta]$ crossing the $j_\star$-th gap downwards as ζ runs from 0 to 2π; see, e.g., [Waterstraat 2017, §4]. Corollary 3.3 in Section 3C allows one to count precisely these eigenvalues. It leads to:

Corollary 1.6. *Assume that* (H1)–(H4) *hold for both Dirac points* $(\xi_\star^A, E_\star)$ *and* $(\xi_\star^B, E_\star)$. *There exists* $\delta_0 > 0$ *such that for all* $\delta \in (0, \delta_0)$, *the spectral flow of* $\mathscr{P}_\delta$ *in the* $j_\star$*-th* $L^2[\zeta]$*-gap vanishes.*

This is because $\vartheta_\star^A$ and $\vartheta_\star^B$ are opposite — where $\vartheta_\star^J$ corresponds to $\vartheta_\star$ for the Dirac point $(\xi_\star^J, E_\star)$. See Figure 7. The spectral flow is a topological invariant: it does not change if a 2π-periodic family of compact operators $H^2[\zeta] \to L^2[\zeta]$ is added to $\mathscr{P}_\delta[\zeta]$. Hence Corollary 1.6 is very robust. However, it is a disappointing result: it suggests that the edge states of Corollary 1.5 shall not be topologically stable. We conjecture:

Conjecture. *Assume that* (H1)–(H4) *hold for both Dirac points* $(\xi_\star^A, E_\star)$ *and* $(\xi_\star^B, E_\star)$. *There exists* $\delta_0 > 0$ *such that for every* $\delta \in (0, \delta_0)$ *there exists a family* $\zeta \in \mathbb{R} \mapsto B_\delta(\zeta)$ *such that*:

- $B_\delta(\zeta)$ *is a compact operator* $H^2[\zeta] \to L^2[\zeta]$.
- $B_\delta(\zeta)$ *depends continuously on* ζ *(with respect to the operator norm on* $H^2[\zeta] \to L^2[\zeta]$*) and* $B_\delta(\zeta + 2\pi) = B_\delta(\zeta)$ *for every* $\zeta \in \mathbb{R}$.
- $\mathscr{P}_\delta[\zeta] + B_\delta(\zeta) : H^2[\zeta] \to L^2[\zeta]$ *has no eigenvalues in the essential spectral gap containing* $E_\star$.

On a positive note, our approach also applies to magnetic Schrödinger operators

$$\begin{gathered} \mathbb{P}_\delta = -(\nabla_{\mathbb{R}^2} + i\delta \cdot \kappa_\delta \cdot \mathbb{A})^2 + V, \\ \mathbb{A} \in C^\infty(\mathbb{R}^2, \mathbb{R}^2), \qquad \mathbb{A}(x+w) = \mathbb{A}(x), \quad w \in \Lambda, \qquad \mathbb{A}(-x) = -\mathbb{A}(x). \end{gathered} \tag{1-14}$$

The asymptotic operators for $\langle k', x\rangle$ near $\pm\infty$ are equal to

$$-(\nabla_{\mathbb{R}^2} + i\delta\mathbb{A})^2 + V. \tag{1-15}$$

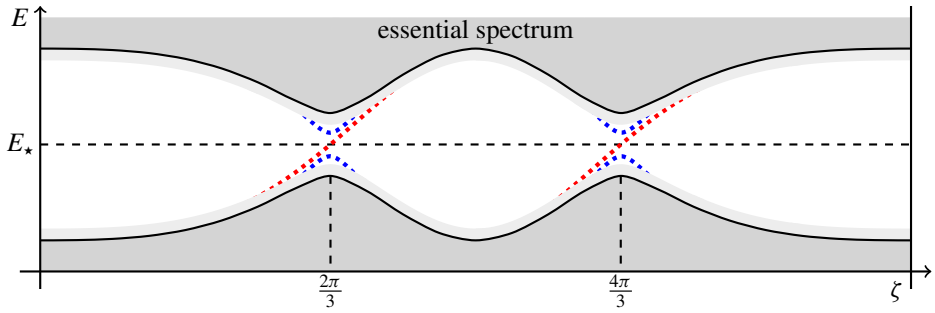

Figure 8. The spectrum of the magnetic-like perturbation $\mathbb{P}_\delta$ of P_0 for positive $\theta_\star$. The topologically protected mode of the Dirac operator induces precisely two edge-state energy curves. In contrast with Figure 7, $\theta_\star^A = \theta_\star^B$: both red curves move upwards. The resulting spectral flow is -2, indicating topologically protected states.

From a physical point of view, (1-15) models quantum particles in a magnetic field $\delta B = \delta(\partial_1 \mathbb{A}_2 - \partial_2 \mathbb{A}_1)$ (oriented in the direction $e_3 \in \mathbb{R}^3$ orthogonal to $\mathbb{R}^2$) and an electric field ∇V. Therefore $\mathbb{P}_\delta$ represents particles evolving in a near-periodic electromagnetic background, with magnetic field varying adiabatically along $\mathbb{R}v'$ from $-\delta B$ to δB. Note that the magnetic flux of B vanishes because B is periodic.

We can see (1-14) as a perturbation of $-\Delta + V$ by

$$\delta \cdot \kappa_\delta \cdot \mathbb{W}, \quad \mathbb{W} \stackrel{\text{def}}{=} \mathbb{A} \cdot D_x + D_x \cdot \mathbb{A},$$

modulo a term of order δ^2. The perturbation $\mathbb{W}$ no longer breaks spatial inversion; instead it breaks time-reversal symmetry (complex conjugation). See [Raghu and Haldane 2008; Haldane and Raghu 2008; Lee-Thorp et al. 2019] for related models. We replace (H3) with:

(H3′) The nondegeneracy condition $\theta_\star \stackrel{\text{def}}{=} \langle \phi_1, \mathbb{W}\phi_1 \rangle_{L^2_{\xi_\star}} \neq 0$ holds.

When (H1), (H2) and (H3′) hold, the operator $\mathbb{P}_\delta[\zeta_\star]$ has an essential spectral gap centered at $E_\star$, of width of order δ — similarly to $\mathscr{P}_\delta[\zeta_\star]$. If moreover (H4) holds, then we can define the spectral flow of the family $\zeta \mapsto \mathbb{P}_\delta[\zeta]$.

Corollary 1.7. *Assume that* (H1), (H2), (H3′) *and* (H4) *hold for both Dirac points* $(\xi_\star^A, E_\star)$ *and* $(\xi_\star^B, E_\star)$. *There exists* $\delta_0 > 0$ *such that for all* $\delta \in (0, \delta_0)$, *the spectral flow of* $\mathbb{P}_\delta$ *equals* $-2 \cdot \operatorname{sgn}(\theta_\star)$.

Corollary 1.7 shows that $\mathbb{P}_\delta$ admits two topologically protected edge states; see Figure 8. This corroborates results of [Haldane and Raghu 2008; Raghu and Haldane 2008], where two quasimodes are produced via a multiscale approach. They were not proved to be topologically protected there: a statement in the spirit of Corollary 3.3 is missing. The authors perform a formal computation of the bulk index: they show that it should equal 2 or -2. We studied rigorously the bulk aspects of our problem in the recent work [Drouot 2019].

1H. ***Strategy.*** Our proof has three essential components:

• The simplest step consists in deriving Corollary 1.5 from Theorem 1.4; see Section 3C. Theorem 1.4 is used to count the exact number $2N+1$ of eigenvalues in the essential spectral gap (slightly away from the edges). We derive the full expansion of edge states in powers of δ using (a) the formal multiscale procedure of [Fefferman et al. 2016b] to produce $2N+1$, almost orthogonal, arbitrarily accurate quasimodes, and (b) a general selfadjoint principle that implies that these quasimodes must all be near genuine eigenvectors.

• We derive resolvent estimates for the bulk operators $P_{\pm\delta}[\zeta_\star]$. We first obtain resolvent estimates for the operators $P_{\pm\delta}(\xi) : H^2_\xi \to L^2_\xi$ in Section 4. We prove that near $(\xi_\star, E_\star)$, these operators essentially behave like Pauli matrices. In Section 5 we integrate these estimates along the dual edge $\zeta_\star k + \mathbb{R}k'$ and derive the expansion

$$(P_{\pm\delta}[\zeta_\star]-\lambda)^{-1} = \frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top \Pi^*\cdot\mathcal{U}_\delta(\not{D}_{\star,\pm}-z)^{-1}\mathcal{U}_\delta^{-1}\cdot\Pi\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}).$$

Above, Π and $\mathcal{U}_\delta$ are the operators introduced in Section 1E, and $\not{D}_{\star,\pm}$ are the formal limits of $\not{D}_\star$ as t goes to $\pm\infty$.

• We use a sophisticated version of the Lippmann–Schwinger principle to connect the resolvents of $\mathscr{P}_\delta[\zeta_\star]$ and of $P_{\pm\delta}[\zeta_\star]$. This requires us to construct a parametrix for $\mathscr{P}_\delta[\zeta_\star]$. After algebraic manipulations — essentially cyclicity arguments — homogenization effects take place and produce the operator $\not{D}_\star$. This leads to the resolvent estimate of Theorem 1.4.

1I. ***Relation to earlier work.*** The mechanism responsible for the production of edge states is the bifurcation of eigenvalues from the edge of the continuous spectrum. Such problems have a long history: see, e.g., [Tamm 1932; Schockley 1939; Simon 1976; Deift and Hempel 1986; Figotin and Klein 1997; Borisov 2007; 2011; 2015; Borisov and Gadyl'shin 2008; Parzygnat et al. 2010; Hoefer and Weinstein 2011; Zelenko 2016] for states generated by defects in periodic backgrounds; and [Golowich and Weinstein 2005; Borisov and Gadyl'shin 2006; Duchêne and Weinstein 2011; Duchêne et al. 2014; Dimassi 2016; Dimassi and Duong 2017; Drouot 2018a; 2018c; 2018d; Duchêne and Raymond 2018] for localized highly oscillatory perturbations.

Fefferman, Lee-Thorp and Weinstein [Fefferman et al. 2016a; 2016b] produced the closest results to our analysis. They were the first to prove existence of edge states for continuous honeycomb lattices in the small/adiabatic regime $\delta \to 0$. They built on their own work [Fefferman et al. 2014; 2017], where they proved existence of defect states for dislocated one-dimensional materials.

Our work improves and extends [Fefferman et al. 2016a; 2016b] in the following ways:

- It connects the resolvents of $\mathscr{P}_\delta[\zeta]$ and $\not{D}(\mu)$.
- It provides full expansions of edge states in powers of δ.
- It identifies *all* edge states with energy near Dirac point energies.

The third point allows for the topological interpretation of the results in terms of the spectral flow of $\zeta \mapsto \mathscr{P}_\delta[\zeta]$. This is a robust invariant of the system, also called the edge index. We conjecture that the modes of $\mathscr{P}_\delta[\zeta]$ should not be topologically protected: the edge index vanishes. However, for the

magnetic operator $\mathbb{P}_\delta[\zeta]$ introduced in (1-14), two such states are topologically protected: they persist under large (suitable) deformations.

We refer to [Gérard et al. 1991; Panati et al. 2003; Watson et al. 2017; Watson and Weinstein 2018] for the study of similar operators with perturbations that vary adiabatically in all directions and to [De Nittis and Lein 2011; 2014; Cornean et al. 2015; 2017a; 2017b] for analysis of perturbations small with respect to the inverse scale of variation. The scaling studied here is peculiar: the perturbation varies adiabatically in one direction only.

Our strategy generalizes the one-dimensional work [Drouot et al. 2018], developed to improve the results of [Fefferman et al. 2014; 2017]. The construction of genuine edge states from quasimodes in Section 3C follows the classical approach of [Drouot et al. 2018, §3.3]. We derive the fiberwise resolvent estimates for $P_{\pm\delta}(\xi)$ in Sections 4A–4B as in [loc. cit., §4.1-4.2]. We did not prove resolvent estimates in [loc. cit.]; we used instead Fredholm determinants.

We pushed the analysis of [Drouot et al. 2018] further in [Drouot 2018b]. There, we showed that the defect states of [Fefferman et al. 2014; 2017] are topologically stable in the following sense. The model embeds naturally in a one-parameter family of *dislocated systems*, related to [Post 2003; Korotyaev 2000; Hempel and Kohlmann 2011a; 2011b; Dohnal et al. 2009; Hempel et al. 2015]. We compute the spectral flow in terms of bulk quantities. We show that it is equal to the bulk index — the Chern number of a Bloch eigenbundle for the bulk. Hence, [Drouot 2018b] provides a novel continuous setting where the bulk-edge correspondence holds — adding to [Kellendonk and Schulz-Baldes 2004a; 2004b; Taarabt 2014; Fukui et al. 2012; Bal 2017; 2018; Bourne and Rennie 2018]. A similar strategy has been developed in [Drouot 2019] to deal with magnetic honeycomb operators.

1J. ***Further perspectives.*** Our results stimulate future lines of research:

• Armchair-type edges are edges such that the associated dual line $\zeta_\star k + \mathbb{R}k'$ passes through both Dirac momenta $\xi_\star^A$ and $\xi_\star^B$. They correspond to the directions

$$v = a_1 v_1 + a_2 v_2, \quad v_1 \wedge v_2 = 1, \quad a_1 = a_2 \bmod 3; \tag{1-16}$$

see Section 2E. The no-fold condition barely fails for such edges: $\mathscr{P}_\delta[\zeta_\star]$ still has an essential gap in, say, the sharp-contrast regime. See [Fefferman et al. 2018, Corollary 6.3]. We expect our techniques to be robust enough to handle such edges. In particular, a 2×2 block of uncoupled Dirac operators should emerge in the resolvent estimates.

• This work may open the way to prove the no-fold conjecture of Fefferman, Lee-Thorp, and Weinstein [Fefferman et al. 2016b]. It predicts that long-lived *resonant* edge states should appear when the no-fold condition fails. This is supported by the existence of highly accurate localized quasimodes, still produced by the formal multiscale procedure of [loc. cit.]. See [Gérard and Sigal 1992; Stefanov and Vodev 1996; Tang and Zworski 1998; Stefanov 1999; 2000; Gannot 2015] for the relation between quasimodes and resonances in other settings.

• The eigenvalue curve $\zeta \mapsto E_{\delta,0}^\zeta$ of $\mathscr{P}_\delta[\zeta]$ corresponding to the topologically protected mode of $\not{D}(\mu)$ intersects $E_\star$ transversely. See the red curves in Figure 7. This contrasts with the eigenvalue curves

$\zeta \mapsto E^{\zeta}_{\delta,j}$, $j \neq 0$, which exhibit quadratic extrema near $\zeta_\star$; see the blue curves in Figure 7. This indicates that L^2-wave packets constructed from the topologically protected modes of $\not{D}(\mu)$ should have a longer lifetime. Mathematical and experimental investigations of this phenomenon would be interesting. The techniques could lead to a time-dependent analysis of quasimodes when the no-fold condition fails. See [Gérard and Sigal 1992] for a related investigation in the shape resonance context and [Carles et al. 2004; Ablowitz and Zhu 2012; 2013; Fefferman and Weinstein 2014; Arbunich and Sparber 2018] for related investigations in gapless settings.

• In [Drouot 2019], we investigate the relation between the bulk and edge indices of $\mathscr{P}_\delta[\zeta]$ or $\mathbb{P}_\delta[\zeta]$, as in [Haldane and Raghu 2008]. The bulk-edge correspondence is widely unexplored in continuous, asymptotically periodic settings: apart from [Bourne and Rennie 2018; Drouot 2018b], the only investigations concern the quantum Hall effect [Kellendonk and Schulz-Baldes 2004a; 2004b; Taarabt 2014]. The discrete setting is better understood [Kellendonk et al. 2002; Elgart et al. 2005; Graf and Porta 2013; Avila et al. 2013; Bal 2017; Shapiro 2017; Braverman 2018; Graf and Shapiro 2018; Graf and Tauber 2018; Shapiro and Tauber 2018]. It would also be nice to study it in quantum graph models of graphene — see [Kuchment and Post 2007; Becker and Zworski 2019; Becker et al. 2018; Lee 2016] for setting and spectral results.

• The recent numerical approach [Thicke et al. 2018] could be applied to $\mathbb{P}_\delta$ as δ increases away from 0. Corollary 1.7 shows that two edge states persist as long as the gap remains open. However their qualitative description (Corollary 7.4) should progressively break down as δ increases. It would be interesting to investigate numerically how their shape changes.

Notation. Here is a list of notation used in this work:

• If $z \in \mathbb{C}$, then $\bar{z}$ denotes its complex conjugate and $|z|$ its modulus. We will sometimes identify a vector $x = [x_1, x_2]^\top \in \mathbb{R}^2$ with the complex number $x_1 + ix_2$.

• $\mathbb{S}^1 \subset \mathbb{C}$ is the circle $\{z \in \mathbb{C} : |z| = 1\}$.

• $\mathbb{D}(z, r) \subset \mathbb{C}$ denotes the disk centered at $z \in \mathbb{C}$ of radius r.

• If $E, F \subset \mathbb{C}$, then $\mathrm{dist}(E, F)$ denotes the Euclidean distance between E and F.

• D_x is the operator $(1/i)[\partial_{x_1}, \partial_{x_2}]^\top = (1/i)\nabla$.

• L^2 denotes the space of square-summable functions and H^s are the classical Sobolev spaces.

• If $\mathcal{H}$ and $\mathcal{H}'$ are Hilbert spaces and $\psi \in \mathcal{H}$, we write $|\psi|_{\mathcal{H}}$ for the norm of $\mathcal{H}$; if $A : \mathcal{H} \to \mathcal{H}'$ is a bounded operator, the operator norm of A is

$$\|A\|_{\mathcal{H}\to\mathcal{H}'} \overset{\text{def}}{=} \sup_{|\psi|_{\mathcal{H}}=1} |A\psi|_{\mathcal{H}'}.$$

If $\mathcal{H} = \mathcal{H}'$, we simply write $\|A\|_{\mathcal{H}} = \|A\|_{\mathcal{H}\to\mathcal{H}}$.

• If $\psi_\epsilon \in \mathcal{H}$ and $f : \mathbb{R}\setminus\{0\} \to \mathbb{R}$, we write $\psi_\epsilon = O_{\mathcal{H}}(f(\epsilon))$ when there exists $C > 0$ such that $|\psi_\epsilon|_{\mathcal{H}} \leq Cf(\epsilon)$ for $\epsilon \in (0, 1]$. If $A_\epsilon : \mathcal{H} \to \mathcal{H}$ is a linear operator and $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$, we write $A_\epsilon = \mathscr{O}_{\mathcal{H}\to\mathcal{H}'}(f(\epsilon))$ when there exists $C > 0$ such that $\|A_\epsilon\|_{\mathcal{H}\to\mathcal{H}'} \leq Cf(\epsilon)$ for $\epsilon \in (0, 1]$. If $\mathcal{H} = \mathcal{H}'$, we simply write $A_\epsilon = \mathscr{O}_{\mathcal{H}}(f(\epsilon))$.

• We denote the spectrum of a (possibly unbounded) operator A on $\mathcal{H}$ by $\Sigma_{\mathcal{H}}(A)$. It splits into an essential part $\Sigma_{\mathcal{H},\mathrm{ess}}(A)$ and a discrete part $\Sigma_{\mathcal{H},\mathrm{d}}(A)$.

• Λ is the lattice $\mathbb{Z}v_1 \oplus \mathbb{Z}v_2$ — see Section 1A. An edge is a line $\mathbb{R}v \subset \mathbb{R}^2$, with $v = a_1v_1 + a_2v_2 \in \Lambda$, a_1, a_2 relatively prime integers. We associate to v vectors v', k and k' via (1-7).

• The space L^2_ξ consists of ξ-quasiperiodic functions with respect to Λ:

$$L^2_\xi \stackrel{\text{def}}{=} \{u \in L^2_{\mathrm{loc}}(\mathbb{R}^2, \mathbb{C}) : u(x+w) = e^{i\langle \xi, w\rangle}u(x),\ w \in \Lambda\}.$$

• $\ell \in (\mathbb{R}^2)^*$ is the projection of k orthogonally to k':

$$\ell \stackrel{\text{def}}{=} k - \frac{\langle k, k'\rangle}{|k'|^2}k'.$$

• $L^2[\zeta]$ is the space

$$L^2[\zeta] \stackrel{\text{def}}{=} \left\{u \in L^2_{\mathrm{loc}}(\mathbb{R}^2, \mathbb{C}) : u(x+v) = e^{i\zeta}u(x),\ \textstyle\int_{\mathbb{R}^2/\mathbb{Z}v} |u(x)|^2\,dx < \infty\right\}.$$

• $V \in C^\infty(\mathbb{R}^2, \mathbb{R})$ is a honeycomb potential — see Definition 1.1.

• $W \in C^\infty(\mathbb{R}^2, \mathbb{R})$ is Λ-periodic and odd — see (1-5).

• P_δ is the operator $-\Delta + V + \delta W$ on L^2; for $\xi \in \mathbb{R}^2$, $P_\delta(\xi)$ is the operator formally equal to P_δ but acting on L^2_ξ. For $\zeta \in \mathbb{R}$, $P_\delta[\zeta]$ is the operator formally equal to P_δ but acting on $L^2[\zeta]$.

• $\mathscr{P}_\delta$ is the operator $-\Delta + V + \delta \cdot \kappa_\delta \cdot W$ on L^2, where $\kappa_\delta(x) = \kappa(\delta\langle k', x\rangle)$ and κ is a domain-wall function — see (1-8). $\mathscr{P}_\delta[\zeta]$ is the operator formally equal to $\mathscr{P}_\delta$ but acting on $L^2[\zeta]$.

• $(\xi_\star, E_\star)$ denotes a Dirac point of $P_0 = -\Delta + V$, associated to a Dirac eigenbasis (ϕ_1, ϕ_2) — see Definition 1.2.

• $\zeta_\star$ is the real number $\langle \xi_\star, v\rangle$.

• $\xi^A_\star, \xi^B_\star, \zeta^A_\star, \zeta^B_\star$ are defined in (1-4) and (1-10), respectively.

• $v_\star$ is a complex number associated to $(\xi_\star, E_\star)$ and to the Dirac eigenbasis (ϕ_1, ϕ_2) such that $|v_\star| = v_F$ — see Section 2C.

• $\vartheta_\star = \langle \phi_1, W\phi_1\rangle_{L^2_{\xi_\star}}$ is always assumed to be nonzero; we also define $|\vartheta_\star| = \vartheta_F$.

• The Pauli matrices are

$$\boldsymbol{\sigma_1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{\sigma_2} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \boldsymbol{\sigma_3} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

These matrices satisfy $\boldsymbol{\sigma_j}^2 = \mathrm{Id}$ and $\boldsymbol{\sigma_i}\boldsymbol{\sigma_j} = -\boldsymbol{\sigma_j}\boldsymbol{\sigma_i}$ for $i \neq j$.

2. Honeycomb potentials, Dirac points and edges

2A. ***Equilateral lattice.*** We review briefly the definitions of Section 1A. The equilateral lattice Λ is $\Lambda = \mathbb{Z}v_1 \oplus \mathbb{Z}v_2$ given in canonical coordinates by

$$v_1 = a\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}, \quad v_2 = a\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix},$$

where $a > 0$ is a constant such that $\mathrm{Det}[v_1, v_2] = 1$. Let $k_1, k_2 \in (\mathbb{R}^2)^*$ be dual vectors: $\langle k_i, v_j \rangle = \delta_{ij}$. Identifying $(\mathbb{R}^2)^*$ with $\mathbb{R}^2$ via the scalar product,

$$[k_1, k_2] \cdot [v_1, v_2] = \mathrm{Id} \quad \Longrightarrow \quad [k_1, k_2] = [v_1, v_2]^{-1} = \frac{1}{6a} \begin{bmatrix} \sqrt{3} & \sqrt{3} \\ 3 & -3 \end{bmatrix}.$$

Our definition does not involve a factor 2π — in contrast with some other conventions. The fundamental cell $\mathbb{L}$ and dual cell $\mathbb{L}^*$ are

$$\mathbb{L} \overset{\text{def}}{=} \{t_1 v_1 + t_2 v_2 : t_1, t_2 \in [0, 1)\}, \quad \mathbb{L}^* \overset{\text{def}}{=} \{\tau_1 k_1 + \tau_2 k_2 : \tau_1, \tau_2 \in [0, 2\pi)\}.$$

2B. ***Symmetries.*** Recall that the space of ξ-quasiperiodic functions is

$$L^2_\xi \overset{\text{def}}{=} \{u \in L^2_{\mathrm{loc}}(\mathbb{R}^2, \mathbb{C}) : u(x + w) = e^{i\langle \xi, w \rangle} u(x),\ w \in \Lambda\}.$$

We introduce three operators: $\mathcal{R}$ (rotation); $\mathcal{I}$ (spatial inversion); and $\mathcal{C}$ (complex conjugation). These are given by

$$\mathcal{R}u(x) = u(Rx), \quad R \overset{\text{def}}{=} \frac{1}{2} \begin{bmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}, \qquad \mathcal{I}u(x) = u(-x), \qquad \mathcal{C}u(x) = \overline{u(x)}.$$

We study the action of these operators on the spaces L^2_ξ. Note that $Rv_1 = -v_2$ and $Rv_2 = v_1 - v_2$. Hence, R leaves Λ invariant. If $u \in L^2_\xi$ then

$$\begin{aligned} (\mathcal{R}u)(x + v) &= u(Rx + Rv) = e^{i\langle \xi, Rv \rangle} (\mathcal{R}u)(x) = e^{i\langle R^* \xi, v \rangle} (\mathcal{R}u)(x), \\ (\mathcal{I}u)(x + v) &= u(-x - v) = e^{-i\langle \xi, v \rangle} (\mathcal{I}u)(x), \\ (\mathcal{C}u)(x + v) &= \overline{u(x + v)} = e^{-i\langle \xi, v \rangle} (\mathcal{C}u)(x). \end{aligned}$$

It follows that

$$\mathcal{R} L^2_\xi = L^2_{R^{-1}\xi}, \quad \mathcal{I} L^2_\xi = L^2_{-\xi}, \quad \mathcal{C} L^2_\xi = L^2_{-\xi}. \tag{2-1}$$

Let $\xi^A_\star$ and $\xi^B_\star$ be given by (1-4):

$$\xi^A_\star = \tfrac{2\pi}{3}(2k_1 + k_2), \quad \xi^B_\star = \tfrac{2\pi}{3}(k_1 + 2k_2).$$

We observe that

$$R^{-1}\xi^A_\star = \xi^A_\star + 2\pi(k_1 + k_2), \quad R^{-1}\xi^B_\star = \xi^B_\star + 2\pi k_1.$$

In particular, $R^{-1}\xi_\star = \xi_\star \bmod 2\pi\Lambda^*$ when $\xi_\star \in \{\xi^A_\star, \xi^B_\star\}$. Thanks to (2-1), we see that the space $L^2_{\xi_\star}$ is $\mathcal{R}$-invariant. Since $\mathcal{R}^3 = \mathrm{Id}$, we deduce that $\mathcal{R} : L^2_{\xi_\star} \to L^2_{\xi_\star}$ has three eigenvalues: 1, τ, $\bar\tau$ with $\tau = e^{2i\pi/3}$. Since $\mathcal{R}$ is a unitary operator, $L^2_{\xi_\star}$ admits an orthogonal decomposition

$$L^2_{\xi_\star} = L^2_{\xi_\star, 1} \oplus L^2_{\xi_\star, \tau} \oplus L^2_{\xi_\star, \bar\tau}, \quad L^2_{\xi_\star, z} \overset{\text{def}}{=} \ker_{L^2_{\xi_\star}} (\mathcal{R} - z).$$

The operator $\mathcal{C}\mathcal{I}$ maps $L^2_{\xi_\star}$ to itself. If $u \in L^2_{\xi_\star, \tau}$ then

$$\mathcal{R}(\mathcal{C}\mathcal{I}u)(x) = \overline{u(-Rx)} = \overline{\tau \cdot u(-x)} = \bar\tau \cdot (\mathcal{C}\mathcal{I}u)(x).$$

Therefore $\mathcal{C}\mathcal{I} L^2_{\xi_\star, \tau} = L^2_{\xi_\star, \bar\tau}$.

2C. ***Dirac points.*** We recall that $P_0 = -\Delta + V$, where V is a honeycomb potential — see Definition 1.1. We denote by

$$\lambda_{0,1}(\xi) \le \lambda_{0,2}(\xi) \le \cdots \le \lambda_{0,j}(\xi) \le \cdots \tag{2-2}$$

the dispersion surfaces of P_0, i.e., the L^2_ξ-eigenvalues of $P_0(\xi)$. Conical intersections in the band spectrum (2-2) are called Dirac points — see Definition 1.2. Fefferman and Weinstein [2012] — see also [Colin de Verdière 1991; Grushin 2009; Berkolaiko and Comech 2018; Lee 2016; Keller et al. 2018; Ammari et al. 2018] for related perspectives — showed the following result:

Theorem 2.1 [Fefferman and Weinstein 2012, Theorem 5.1]. *Let $V_0 \in C^\infty(\mathbb{R}^2, \mathbb{R})$ be a honeycomb potential such that*

$$\int_{\mathbb{L}} e^{-i(k_1+k_2)x} V_0(x)\, dx \neq 0. \tag{2-3}$$

There exists a closed countable set $\mathcal{S} \subset \mathbb{R}$ such that for every $t \in \mathbb{R} \setminus \mathcal{S}$, the operator $-\Delta_{\mathbb{R}^2} + tV_0$ admits Dirac points

$$(\xi_\star, E_\star) \in \{\xi_\star^A, \xi_\star^B\} \times \mathbb{R}, \quad \xi_\star^A \stackrel{\text{def}}{=} \tfrac{2\pi}{3}(2k_1 + k_2), \quad \xi_\star^B \stackrel{\text{def}}{=} \tfrac{2\pi}{3}(k_1 + 2k_2).$$

This result shows that P_0 generically admits Dirac points: the condition (2-3) excludes a hyperplane in the space of honeycomb potentials; the "bad" set $\mathcal{S}$ is countable and accounts for extraordinary cases, e.g., higher multiplicity of $E_\star$ or quadratic intersections of dispersion surfaces. When P_0 admits Dirac points, the eigenspace $\ker_{L^2_{\xi_\star}}(P_0(\xi_\star) - E_\star)$ is spanned by an orthonormal basis $\{\phi_1, \phi_2\}$, with

$$\phi_1 \in L^2_{\xi_\star,\tau}, \quad \phi_2 = \overline{\mathcal{I}\phi_1} \in L^2_{\xi_\star,\bar\tau}.$$

We call (ϕ_1, ϕ_2) a Dirac eigenbasis. It is unique modulo the $\mathbb{S}^1$-action $(\phi_1, \phi_2) \mapsto (\omega\phi_1, \bar\omega\phi_2)$, $\omega \in \mathbb{S}^1$.

Lemma 2.2. *Let $(\xi_\star, E_\star)$ be a Dirac point of P_0 with Dirac eigenbasis (ϕ_1, ϕ_2). Then*

$$\langle \phi_1, D_x\phi_1\rangle_{L^2_{\xi_\star}} = \langle \phi_2, D_x\phi_2\rangle_{L^2_{\xi_\star}} = 0.$$

In addition, there exists $v_\star \in \mathbb{C}$ with $|v_\star| = v_F$ such that for all $\eta \in \mathbb{R}^2$ (canonically identified with a complex number),

$$2\langle \phi_1, (\eta \cdot D_x)\phi_2\rangle_{L^2_{\xi_\star}} = v_\star\eta, \quad 2\langle \phi_2, (\eta \cdot D_x)\phi_1\rangle_{L^2_{\xi_\star}} = \overline{v_\star\eta}.$$

This lemma can be deduced from [Fefferman et al. 2016b, Proposition 4.5]. We include a proof in Appendix A.1. It relies on some algebraic relations relating P_0, $\mathcal{R}$ and $\mathcal{I}$, and on perturbation theory of eigenvalues.

2D. ***Breaking the symmetry.*** We will consider Schrödinger operators $P_\delta = -\Delta + V + \delta W$, where

$$W \in C^\infty(\mathbb{R}^2, \mathbb{R}), \qquad W(x + w) = W(x), \quad w \in \Lambda, \qquad W(x) = -W(-x).$$

Lemma 2.3. *Let $(\xi_\star, E_\star)$ be a Dirac point of P_0 with Dirac eigenbasis (ϕ_1, ϕ_2) — see Definition 1.2. Then $\langle \phi_1, W\phi_2\rangle_{L^2_{\xi_\star}} = \langle \phi_2, W\phi_1\rangle_{L^2_{\xi_\star}} = 0$. Furthermore,*

$$\vartheta_\star \stackrel{\text{def}}{=} \langle \phi_1, W\phi_1\rangle_{L^2_{\xi_\star}} = -\langle \phi_2, W\phi_2\rangle_{L^2_{\xi_\star}}.$$

See the proofs of [Fefferman et al. 2016b, (6.19), (6.20)] or Appendix A.1. These identities rely on $\mathcal{I}$ being an isometry. If $\omega \in \mathbb{S}^1$, the change $(\phi_1, \phi_2) \mapsto (\omega\phi_1, \bar{\omega}\phi_2)$ of Dirac eigenbasis leaves $\vartheta_\star$ invariant.

2E. ***Edges.*** Let a_1 and a_2 be two relatively prime integers and $v = a_1 v_1 + a_2 v_2$. Introduce $v' = b_1 v_1 + b_2 v_2$, where $a_1 b_2 - a_2 b_1 = 1$. The vectors v and v' span Λ:

$$\begin{aligned} b_1 v - a_1 v' &= (b_1 a_2 - a_1 b_2) v_2 = -v_2, \\ b_2 v - a_2 v &= (b_2 a_1 - a_2 b_1) v_1 = v_1. \end{aligned} \tag{2-4}$$

Let k and k' be dual vectors. We claim that $k = b_2 k_1 - b_1 k_2$ and $k' = -a_2 k_1 + a_1 k_2$:

$$\begin{aligned} \langle k, v\rangle &= b_2 a_1 - b_1 a_2 = 1, \quad \langle k, v'\rangle = -a_2 a_1 + a_1 a_2 = 0, \\ \langle k', v\rangle &= b_2 b_1 - b_1 b_2 = 0, \quad \langle k', v'\rangle = -a_2 b_1 + a_1 b_2 = 1. \end{aligned}$$

Let $(\xi_\star^A, E_\star)$ be a Dirac point in the sense of Definition 1.2 and $\mathbb{R}v$ be an edge. Assume that $\xi_\star^B$ belongs to the dual edge $\zeta_\star^A k + \mathbb{R}k' \bmod 2\pi\Lambda^*$. In this case we can write $\xi_\star^B = \zeta_\star^A k + \tau k'$, with $\tau \neq \langle \xi_\star^A, v'\rangle \bmod 2\pi\mathbb{Z}$. Since $\lambda_{0,j_\star}(\xi_\star^B) = E_\star$, the no-fold condition fails when $\xi_\star^B \in \zeta_\star^A k + \mathbb{R}k' \bmod 2\pi\Lambda^*$ (see Definition 1.3). Given the expressions (1-4) of $\xi_\star^A$ and $\xi_\star^B$ and (1-7) of v', this arises precisely when

$$\frac{2a_1 + a_2}{3} - \frac{a_1 + 2a_2}{3} \in \mathbb{Z} \quad \Longleftrightarrow \quad a_2 - a_1 \in 3\mathbb{Z}.$$

In particular, if the no-fold condition holds then $a_1 - a_2 \neq 0 \bmod 3$. This implies that $\{\zeta_\star^A, \zeta_\star^B\} = \left\{\frac{2\pi}{3}, \frac{4\pi}{3}\right\} \bmod 2\pi\mathbb{Z}$ because of (1-10).

3. The characterization of edge states

This work studies the eigenvalues of the operator

$$\mathscr{P}_\delta[\zeta] = -\Delta + V + \delta \cdot \kappa_\delta \cdot W : L^2[\zeta] \to L^2[\zeta].$$

Above, κ_δ is a domain-wall function — see (1-8) — and $L^2[\zeta]$ is the space (1-9). The operator $\mathscr{P}_\delta[\zeta]$ is a Schrödinger operator that interpolates between $P_\delta[\zeta]$ at $-\infty$ and $P_\delta[\zeta]$ at $+\infty$. See Figure 9. In this section we review the multiscale approach of [Fefferman et al. 2016a; 2016b] and we derive Corollary 1.5 assuming Theorem 1.4, in a slightly more general setting.

3A. ***The formal multiscale approach.*** The eigenvalue problem for $\mathscr{P}_\delta[\zeta]$ is

$$\begin{cases} (-\Delta + V(x) + \delta\kappa_\delta(x) W(x) - E_\delta) u_\delta = 0, \\ u_\delta(x + v) = e^{i\zeta} u_\delta(x), \end{cases} \qquad \int_{\mathbb{R}^2/\mathbb{Z}v} |u_\delta(x)|^2 \, dx < \infty. \tag{3-1}$$

The multiscale procedure of Fefferman, Lee-Thorp, and Weinstein [Fefferman et al. 2016b, §6] produces approximate solutions of (3-1). We review it below.

We first observe that if we write a function $u_\delta \in C^\infty(\mathbb{R}^2, \mathbb{C})$ as

$$u_\delta(x) = U_\delta(x, \delta\langle k', x\rangle), \quad U_\delta \in C^\infty(\mathbb{R}^2 \times \mathbb{R}, \mathbb{C}), \tag{3-2}$$

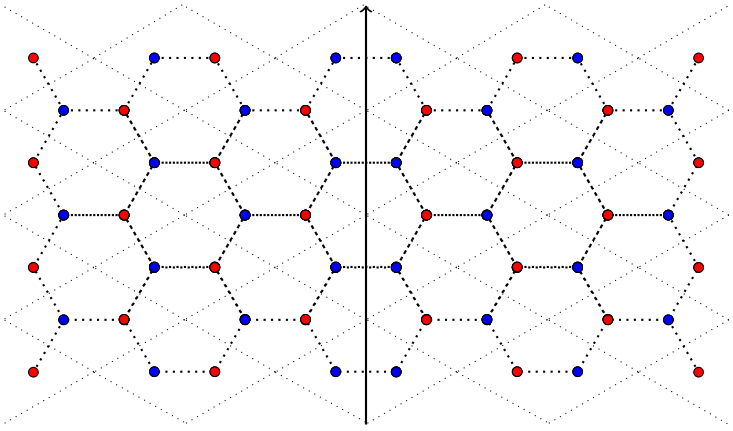

Figure 9. $\mathscr{P}_\delta[\zeta]$ is a Schrödinger operator with a typical potential represented above, with the zigzag edge $v_1 - v_2$. Each red (resp. blue) circle supports an atomic (e.g., radial) potential. The resulting potential is not periodic with respect to Λ; rather it is periodic with respect to $\mathbb{Z}v$.

then u_δ solves (3-1) if and only if U_δ solves

$$\begin{cases} ((D_x + \delta k' D_t)^2 + V(x) + \delta\kappa(t)W(x) - E_\delta)U_\delta = 0, \\ U_\delta(x+v,t) = e^{i\zeta}U_\delta(x,t), \end{cases} \qquad \int_{\mathbb{R}^2/\mathbb{Z}v} |U_\delta(x, \delta\langle k', x\rangle)|^2\,dx < \infty. \tag{3-3}$$

We now produce approximate solutions to the system (3-3) when ζ is near $\zeta_\star = \langle \xi_\star, v\rangle$. We fix $(\xi_\star, E_\star)$ a Dirac point of P_0 and we write $\zeta = \zeta_\star + \mu\delta$, $\zeta_\star = \langle \xi_\star, v\rangle$. We make an ansatz for U_δ and E_δ:

$$U_\delta(x,t) = e^{i\mu\delta\langle \ell, x\rangle} \cdot \left(\sum_{j=1,2} \alpha_j(t)\cdot\phi_j(x) + \delta\cdot V_\delta(x,t) \right), \qquad E_\delta = E_\star + \vartheta\delta + O(\delta^2), \tag{3-4}$$

where

- (ϕ_1, ϕ_2) is a Dirac eigenbasis for $(\xi_\star, E_\star)$ — see Definition 1.2;
- α_1, α_2 are smooth, exponentially decaying functions on $\mathbb{R}$, to be specified below;
- $V_\delta \in X$ — the space defined in (1-13).
- $\ell = k - (\langle k', k\rangle/|k'|^2)k'$ is the projection of k to the orthogonal of $\mathbb{R}k'$;
- $\vartheta \in \mathbb{R}$ is a real number that will be specified below.

Since $\phi_1, \phi_2 \in L^2_{\xi_\star}$, $V_\delta \in X$ and $\alpha_1, \alpha_2 \in L^2(\mathbb{R})$, the ansatz (3-4) implies

$$U_\delta(x+v,t) = e^{i\zeta}U_\delta(x,t), \qquad \int_{\mathbb{R}^2/\mathbb{Z}v} |U_\delta(x, \delta\langle k', x\rangle)|^2\,dx < \infty.$$

In particular the boundary and decay conditions of (3-3) hold under (3-4).

The eigenvalue problem (3-3) becomes a hierarchy of equations, obtained by identifying terms of orders $1, \delta, \delta^2, \ldots$. Since $(P_0 - E_\star)\phi_j = 0$, the equation for the terms of order 1 is automatically satisfied.

The equation for the terms of order δ is

$$e^{i\mu\delta\langle\ell,x\rangle}(P_0-E_\star)V_\delta(x,t)+e^{i\mu\delta\langle\ell,x\rangle}(2(k'\cdot D_x)D_t+\kappa(t)W(x)-\vartheta)\sum_{j=1,2}\alpha_j(t)\phi_j(x)$$
$$+2\mu e^{i\mu\delta\langle\ell,x\rangle}(\ell\cdot D_x)\sum_{j=1,2}\alpha_j(t)\phi_j(x)=0. \quad (3\text{-}5)$$

Note that for every $t\in\mathbb{R}$, $(P_0-E_\star)V_\delta(\cdot,t)$ is orthogonal to ϕ_1 and ϕ_2. Therefore, for this system to have a solution, we must have for every $t\in\mathbb{R}$ and $k=1,2$,

$$\Big\langle \phi_k, \big(2(k'\cdot D_x)D_t+2\mu(\ell\cdot D_x)+\kappa(t)W-\vartheta\big)\sum_{j=1,2}\alpha_j(t)\cdot\phi_j\Big\rangle_{L^2_{\xi_\star}}=0. \quad (3\text{-}6)$$

The scalar products $\langle\phi_j,(k'\cdot D_x)\phi_k\rangle_{L^2_{\xi_\star}}$, $\langle\phi_j,(\ell\cdot D_x)\phi_k\rangle_{L^2_{\xi_\star}}$ and $\langle\phi_j,W\phi_k\rangle_{L^2_{\xi_\star}}$ appear in the solvability condition (3-6). They were computed in Lemmas 2.2 and 2.3. Using these formulas, (3-6) simplifies to

$$(\not{D}(\mu)-\vartheta)\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}=0,\quad \not{D}(\mu)\overset{\text{def}}{=}\begin{bmatrix}0 & \nu_\star k'\\ \overline{\nu_\star k'} & 0\end{bmatrix}D_t+\mu\begin{bmatrix}0 & \nu_\star\ell\\ \overline{\nu_\star\ell} & 0\end{bmatrix}+\vartheta_\star\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\kappa.$$

This system has exponentially decaying solutions $[\alpha_1,\alpha_2]^\top$ if and only if ϑ is an eigenvalue of $\not{D}(\mu)$. Under this condition, (3-5) has a solution V_δ. In other words, this constructs a function U_δ such that (3-3) is satisfied modulo $O_X(\delta^2)$, meaning that for some $a,C>0$ and all $\delta\in(0,1)$

$$\sup_{\mathbb{R}\times\mathbb{R}^2} e^{a|t|}\cdot\big|\big((D_x+\delta k' D_t)^2+V(x)+\delta\kappa(t)W(x)-E_\delta\big)U_\delta\big|\le C\delta^2.$$

We can iterate this procedure to arbitrarily high orders in δ. It produces a function U_δ such that (3-3) is satisfied modulo $O_X(\delta^M)$ for any M. Identifying U_δ with u_δ according to (3-2), this procedure produces for any M and any eigenvalue ϑ of $\not{D}(\mu)$ a function $u_{\delta,M}$ that solves

$$(\mathscr{P}_\delta[\zeta]-E_\delta)u_{\delta,M}=O_X(\delta^M),\quad E_\delta=E_\star+\delta\vartheta+O(\delta^2).$$

This is an approximate solution to the eigenvalue problem (3-1).

It is natural to ask whether these approximate solutions are close to eigenvectors. The work [Fefferman et al. 2016b] shows that this holds at first order in δ. Below we state results that imply that this holds at *any* order in δ. This dramatically refines the main result of [loc. cit.]. Our approach relies on resolvent estimates rather than by-hand construction of eigenvectors. It comes with further improvements of [loc. cit.]:

- the precise counting of eigenvalues of $\mathscr{P}_\delta[\zeta]$;
- an estimate that connects the resolvents of $\mathscr{P}_\delta[\zeta]$ and $\not{D}(\mu)$.

These results are stated in Section 3C and first require a spectral analysis of $\not{D}(\mu)$.

3B. ***The Dirac operator $\not{D}(\mu)$.*** The Dirac operator

$$\not{D}(\mu)=\begin{bmatrix}0 & \nu_\star k'\\ \overline{\nu_\star k'} & 0\end{bmatrix}D_t+\mu\begin{bmatrix}0 & \nu_\star\ell\\ \overline{\nu_\star\ell} & 0\end{bmatrix}+\vartheta_\star\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\kappa$$

emerges in the multiscale analysis of [Fefferman et al. 2016b]. We saw that its eigenvalues are particularly relevant in the construction of approximate eigenvectors of $\mathscr{P}_\delta[\zeta]$, $\zeta = \zeta_\star + \delta\mu$. In this section we relate the spectra of $\not{D}(\mu)$ and $\not{D}_\star = \not{D}(0)$.

Lemma 3.1. *The essential and discrete spectra of $\not{D}_\star$ and $\not{D}(\mu)$ are related through*

$$\Sigma_{L^2,\mathrm{ess}}(\not{D}(\mu)) = \mathbb{R} \setminus \big(-\sqrt{\vartheta_F^2 + \mu^2 \cdot v_F^2|\ell|^2}, \sqrt{\vartheta_F^2 + \mu^2 \cdot v_F^2|\ell|^2}\big),$$
$$\Sigma_{L^2,\mathrm{d}}(\not{D}(\mu)) = \big\{\mu \cdot v_F|\ell| \cdot \operatorname{sgn}(\vartheta_\star) : \pm\sqrt{\vartheta_j^2 + \mu^2 \cdot v_F^2|\ell|^2} \text{ with } 0 \neq \vartheta_j \in \Sigma_{L^2,\mathrm{d}}(\not{D}_\star)\big\}.$$

All the eigenvalues of $\not{D}_\star$ and $\not{D}(\mu)$ are simple.

The proof of Lemma 3.1 relies on a supersymmetry: there exists a 2×2 matrix $\boldsymbol{m_2}$ such that $\boldsymbol{m_2}^2 = \mathrm{Id}$ and $\boldsymbol{m_2}\not{D}_\star = -\boldsymbol{m_2}\not{D}_\star$. We postpone it to Appendix A.2. We also mention that $\not{D}_\star$ may have more than one eigenvalue — see [Lu et al. 2018]. For a general perspective for applications of supersymmetries in spectral theory, see [Cycon et al. 1987, §6–12].

3C. ***Parallel quasimomentum near $\zeta_\star$.*** We are now ready to state the main result of our work. Recall that the assumptions (H1)–(H3) were introduced in Section 1E and that Π, Π^* and $\mathcal{U}_\delta$ are defined by

$$\Pi : L^2(\mathbb{R}^2/\mathbb{Z}v, \mathbb{C}^2) \to L^2(\mathbb{R}, \mathbb{C}^2), \qquad (\Pi f)(t) \stackrel{\text{def}}{=} \textstyle\int_0^1 f(sv + tv')\, ds,$$
$$\Pi^* : L^2(\mathbb{R}, \mathbb{C}^2) \to L^2(\mathbb{R}^2/\mathbb{Z}v, \mathbb{C}^2), \quad (\Pi^* g)(x) \stackrel{\text{def}}{=} g(\langle k', x\rangle),$$
$$\mathcal{U}_\delta : L^2(\mathbb{R}, \mathbb{C}^2) \to L^2(\mathbb{R}, \mathbb{C}^2), \qquad (\mathcal{U}_\delta f)(t) \stackrel{\text{def}}{=} f(\delta t).$$

Theorem 3.2. *Assume that the assumptions* (H1)–(H3) *hold. Fix $\mu_\sharp > 0$ and $\epsilon > 0$. There exists $\delta_0 > 0$ such that if*

$$\mu \in (-\mu_\sharp, \mu_\sharp), \quad \delta \in (0, \delta_0), \quad z \in \mathbb{D}\big(0, \sqrt{\vartheta_F^2 + \mu^2 \cdot v_F^2|\ell|^2} - \epsilon\big), \quad \operatorname{dist}(\Sigma_{L^2}(\not{D}(\mu)), z) \geq \epsilon,$$
$$\zeta = \zeta_\star + \delta\mu, \quad \lambda = E_\star + \delta z$$

then $\mathscr{P}_\delta[\zeta] - \lambda$ is invertible and its resolvent $(\mathscr{P}_\delta[\zeta] - \lambda)^{-1}$ equals

$$\frac{1}{\delta} \cdot \begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle} \cdot \Pi^* \mathcal{U}_\delta \cdot (\not{D}(\mu) - z)^{-1} \cdot \mathcal{U}_\delta^{-1} \Pi \cdot e^{i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}).$$

It suffices to take $\mu = 0$ in Theorem 3.2 to derive Theorem 1.4.

Corollary 3.3. *Assume* (H1)–(H3) *hold and fix $\vartheta_\sharp \in (\vartheta_N, \vartheta_F)$ and $\mu_\sharp > 0$. There exists $\delta_0 > 0$ such that for*

$$\delta \in (0, \delta_0), \quad \mu \in (-\mu_\sharp, \mu_\sharp), \quad \zeta = \zeta_\star + \delta\mu,$$

the operator $\mathscr{P}_\delta[\zeta]$ has exactly $2N + 1$ eigenvalues $\{E_{\delta,j}^\zeta\}_{j\in[-N,N]}$ in

$$\big[E_\star - \delta\sqrt{\vartheta_\sharp^2 + \mu^2 \cdot v_F^2|\ell|^2},\ E_\star + \delta\sqrt{\vartheta_\sharp^2 + \mu^2 \cdot v_F^2|\ell|^2}\,\big].$$

These eigenvalues are simple. Furthermore, for each $j \in [-N, N]$, the eigenpairs $(E^\zeta_{\delta,j}, u^\zeta_{\delta,j})$ admit full expansions in powers of δ:

$$E^\zeta_{\delta,j} = E_\star + \vartheta^\mu_j \cdot \delta + a^\mu_2 \cdot \delta^2 + \cdots + a^\mu_M \cdot \delta^M + O(\delta^{M+1}),$$
$$u^\zeta_{\delta,j}(x) = e^{i(\zeta-\zeta_\star)\langle \ell, x\rangle}\big(f^\mu_0(x, \delta\langle k', x\rangle) + \cdots + \delta^M \cdot f_M(x, \delta\langle k', x\rangle)\big) + o_{H^k}(\delta^M).$$

In the above expansions:

- *M and k are any integers; H^k is the k-th order Sobolev space.*
- *ϑ^μ_j is the j-th eigenvalue of $\not{D}(\mu)$, described in Lemma 3.1.*
- *The terms $a^\mu_m \in \mathbb{R}$ and $f^\mu_m \in X$ are recursively constructed via the multiscale analysis of [Fefferman et al. 2016b] — see Section 3A.*
- *The leading-order term f^μ_0 satisfies*

$$f^\mu_0(x,t) = \alpha^\mu_1(t)\phi_1(x) + \alpha^\mu_2(t)\phi_2(x), \quad (\not{D}(\mu) - \vartheta^\mu_j)\begin{bmatrix}\alpha^\mu_1\\ \alpha^\mu_2\end{bmatrix} = 0.$$

Proof of Corollary 3.3 assuming Theorem 3.2. In order to locate eigenvalues of $\mathscr{P}_\delta[\zeta]$, it suffices to integrate the resolvent on contours enclosing regions where Theorem 3.2 does not apply.

Let ϑ_j be an eigenvalue of $\not{D}(\mu)$ and $\epsilon > 0$ so that $\not{D}(\mu)$ has no other eigenvalues in $\mathbb{D}(\vartheta_j, \epsilon)$. We compute the residue

$$\frac{1}{2\pi i}\oint_{\partial\mathbb{D}(E_\star+\delta\vartheta_j,\epsilon\delta)} (\lambda - \mathscr{P}_\delta(\zeta))^{-1}\, d\lambda. \tag{3-7}$$

This is the projector on the spectrum of $\mathscr{P}_\delta(\zeta)$ that is enclosed by $\partial\mathbb{D}(E_\star + \delta\vartheta_j, \epsilon\delta)$. Because of Theorem 3.2 and the relation $\lambda = E_\star + \delta z$, $d\lambda = \delta\, dz$, (3-7) equals

$$\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle}\cdot \Pi^* \mathcal{U}_\delta \cdot \frac{1}{2\pi i}\oint_{\partial\mathbb{D}(\vartheta_j,\epsilon)} (z - \not{D}(\mu))^{-1}\, dz \cdot \mathcal{U}_\delta^{-1}\Pi \cdot e^{i\mu\delta\langle \ell, x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(\delta^{2/3}).$$

The residue

$$\frac{1}{2\pi i}\oint_{\partial\mathbb{D}(\vartheta_j,\epsilon)} (z - \not{D}(\mu))^{-1}\, dz$$

is a rank-1 projector on $\ker_{L^2}(\not{D} - \vartheta_j)$. We write it as $\alpha^\zeta \otimes \alpha^\zeta$, where $|\alpha^\zeta|_{L^2} = 1$. We deduce that the residue (3-7) equals

$$\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle}\cdot \Pi^* \mathcal{U}_\delta \cdot \alpha \otimes \alpha \cdot \mathcal{U}_\delta^{-1}\Pi \cdot e^{i\mu\delta\langle \ell, x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(\delta^{2/3}) = v^\zeta_0 \otimes v^\zeta_0 + \mathscr{O}_{L^2[\zeta]}(\delta^{2/3}),$$

where

$$v^\zeta_0 \overset{\text{def}}{=} \delta^{1/2}\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle}\cdot \Pi^* \mathcal{U}_\delta \cdot \alpha.$$

Above we used that $(\mathcal{U}_\delta^{-1})^* = \delta \cdot \mathcal{U}_\delta$.

We deduce that (3-7) is a projector that takes the form $v_0^\zeta \otimes v_0^\zeta + \mathscr{O}_{L^2[\zeta]}(\delta^{2/3})$, where $|v_0^\zeta|_{L^2[\zeta]} = 1$. In particular, it is nonzero. Moreover, it has rank at most 1. Indeed, normalized vectors in its range must be of the form $v_0^\zeta + O_{L^2[\zeta]}(\delta^{2/3})$; therefore two of them cannot be orthogonal for δ sufficiently small. We deduce that (3-7) has rank exactly 1: $\mathscr{P}_\delta[\zeta]$ has exactly one eigenvalue in $\mathbb{D}(E_\star + \delta\vartheta_j, \epsilon\delta)$.

The rest of the proof is identical to [Drouot et al. 2018, Proof of Corollary 1]. It relies on

- the fact that $\mathscr{P}_\delta[\zeta]$ has exactly one eigenvalue in the disk enclosed by $\partial\mathbb{D}(E_\star + \delta\vartheta_j, \epsilon\delta)$ — proved just above;
- a general variational argument that shows that an approximate eigenpair (ψ, E) for a selfadjoint problem that has only one eigenvalue near E must be close to a genuine eigenpair — see [Drouot et al. 2018, Lemma 3.1];
- the construction of arbitrarily accurate approximate eigenpairs thanks to the multiscale procedure of [Fefferman et al. 2016b] — see Section 3A.

We refer to [Drouot et al. 2018, Proof of Corollary 1] for details. □

Most of the rest of the paper is devoted to the proof of Theorem 3.2.

4. The Bloch resolvent

Recall that V is a honeycomb potential — see Definition 1.1 — and that $W \in C^\infty(\mathbb{R}^2, \mathbb{R})$ is odd and Λ-periodic. In this section we study the resolvent of $P_\delta(\xi)$, the operator formally equal to $P_\delta = -\Delta + V + \delta W$ but acting on quasiperiodic spaces L^2_ξ.

Under the no-fold condition, we prove in Lemma 4.1 that $(P_\delta(\xi) - z)^{-1}$ is subdominant away from the Dirac quasimomenta $\xi_\star$. The situation is more subtle near $\xi_\star$. In Lemma 4.2 we show that when the nondegeneracy assumption (1-6) holds and (ξ, λ) is near a Dirac point $(\xi_\star, E_\star)$, $(P_\delta(\xi) - \lambda)^{-1}$ behaves like the resolvent of a rank-2 operator.

4A. ***Resolvent away from Dirac momenta.*** We recall that $\mathbb{L}$ is the fundamental cell associated to the generators v_1 and v_2; see (1-2). Given $\xi \in \mathbb{R}^2$, we define $\rho(\xi)$ as

$$\rho(\xi) \stackrel{\text{def}}{=} \operatorname{dist}(\xi + 2\pi\Lambda^*, \zeta_\star k + \mathbb{R}k').$$

Lemma 4.1. *Assume that the assumptions* (H1) *and* (H2) *hold. Let $c > 0$. There exist $\delta_0, \epsilon_0 > 0$ such that if*

$$\delta \in (0, \delta_0), \quad \xi \in \mathbb{L}^*, \quad \rho(\xi) \le \epsilon_0, \quad |\xi - \xi_\star| \ge \delta^{1/3}, \quad \lambda \in \mathbb{D}(E_\star, c\delta) \tag{4-1}$$

then $P_\delta(\xi) - \lambda$ is invertible and

$$\|(P_\delta(\xi) - \lambda)^{-1}\|_{L^2_\xi \to H^2_\xi} = O(\delta^{-1/3}).$$

Proof. 1. We first show that there exists $\epsilon_0 > 0$ such that

$$\xi \in \mathbb{L}^* \setminus \{\xi_\star\}, \ \ \rho(\xi) \le \epsilon_0, \quad \Longrightarrow \quad \lambda_{0,j_\star}(\xi) < E_\star - 2\epsilon_0 \cdot |\xi - \xi_\star|. \tag{4-2}$$

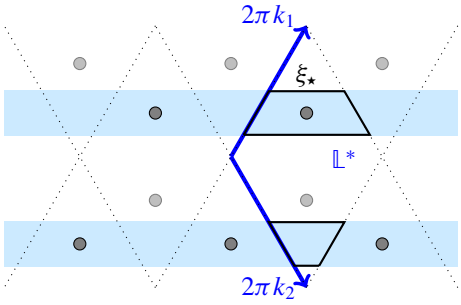

Figure 10. If $v = v_1 - v_2$ is the zigzag edge then $k' = k_1 + k_2$. An ϵ_0-neighborhood of the dual line $\zeta_\star k + \mathbb{R}k'$ is represented above as the blue strip. Lemma 4.1 applies to quasimomenta in the area enclosed in black. This domain of validity extends by periodicity to the blue strips away from $\xi_\star$ mod $2\pi\Lambda^*$.

Indeed, if this does not hold then we can find ξ_n such that

$$\xi_n \in \mathbb{L}^* \setminus \{\xi_\star\}, \quad \rho(\xi_n) \leq \frac{1}{n}, \quad \lambda_{0,j_\star}(\xi_n) \geq E_\star - \frac{2}{n} \cdot |\xi - \xi_\star|.$$

Since $\xi_n \in \mathbb{L}^*$, we know ξ_n is bounded. There exists a subsequence $\xi_{\varphi(n)}$ of ξ_n that converges to an element ξ_∞ in the closure of $\mathbb{L}^*$, with $\rho(\xi_\infty) = 0$. Because $\lambda_{0,j_\star}$ is continuous, we have $\lambda_{0,j_\star}(\xi_\infty) \geq E_\star$. Since $\rho(\xi_\infty) = 0$, there exist $\eta \in \Lambda^*$ and $\tau_0 \in \mathbb{R}$ such that

$$\xi_\infty + 2\pi\eta = \zeta_\star k + \tau_0 k'.$$

We look at the function $\varphi(\tau) \overset{\text{def}}{=} \lambda_{0,j_\star}(\zeta_\star k + \tau k')$. It is 2π-periodic and it equals $E_\star$ precisely when $\tau = \langle \xi_\star, v' \rangle$ mod 2π because of (H2). Moreover,

$$\varphi(\langle \xi_\star, v' \rangle + \epsilon) = E_\star - v_F|\epsilon k'| + O(\epsilon^2).$$

Therefore, the intermediate value theorem shows that $\varphi(\tau) < E_\star$ unless $\tau = \langle \xi_\star, v' \rangle$ mod 2π. We deduce that $\tau_0 = \langle \xi_\star, v' \rangle$ mod 2π. Hence $\xi_\infty = \xi_\star$ mod $2\pi\Lambda^*$. Since ξ_∞ is in the closure of $\mathbb{L}^*$, we know $\xi_\infty = \xi_\star$. Since it also belongs to $\zeta_\star k + \mathbb{R}k'$, we have $\xi_\infty = \xi_\star$. Since $\xi_\star$ is a Dirac point, we deduce

$$E_\star - \frac{2}{\varphi(n)} \cdot |\xi_{\varphi(n)} - \xi_\star| \leq \lambda_{0,j_\star}(\xi_{\varphi(n)}) \leq E_\star - v_F \cdot |\xi_{\varphi(n)} - \xi_\star| + O(\xi_{\varphi(n)} - \xi_\star)^2.$$

This cannot hold for large n, unless $\xi_{\varphi(n)} = \xi_\star$, which is excluded. We deduce that (4-2) holds. A similar argument implies that

$$\xi \in \mathbb{L}^* \setminus \{\xi_\star\}, \quad \rho(\xi) \leq \epsilon_0 \quad \Longrightarrow \quad \lambda_{0,j_\star+1}(\xi) > E_\star + 2\epsilon_0 \cdot |\xi - \xi_\star|. \tag{4-3}$$

2. From (4-2) and (4-3), we deduce that for $\delta > 0$,

$$\xi \in \mathbb{L}^*, \quad \rho(\xi) \leq \epsilon_0, \quad |\xi - \xi_\star| \geq \delta^{1/3} \quad \Longrightarrow \quad \begin{cases} \lambda_{0,j_\star}(\xi) < E_\star - 2\epsilon_0\delta^{1/3}, \\ \lambda_{0,j_\star+1}(\xi) > E_\star + 2\epsilon_0\delta^{1/3}. \end{cases}$$

In particular, if $c > 0$ is given and $\lambda \in \mathbb{D}(E_\star, c\delta)$ then

$$\xi \in \mathbb{L}^*, \quad \rho(\xi) \le \epsilon_0, \quad |\xi - \xi_\star| \ge \delta^{1/3} \quad \Longrightarrow \quad \begin{cases} \operatorname{Re}(\lambda_{0,j_\star}(\xi) - \lambda) < c\delta - 2\epsilon_0\delta^{1/3}, \\ \operatorname{Re}(\lambda_{0,j_\star+1}(\xi) - \lambda) > 2\epsilon_0\delta^{1/3} - c\delta. \end{cases}$$

In particular, when δ_0 is sufficiently small, $\delta \in (0, \delta_0)$ and $\lambda \in \mathbb{D}(E_\star, c\delta)$,

$$\xi \in \mathbb{L}^*, \quad \rho(\xi) \le \epsilon_0, \quad |\xi - \xi_\star| \ge \delta^{1/3} \quad \Longrightarrow \quad \begin{cases} \operatorname{Re}(\lambda_{0,j_\star}(\xi) - \lambda) < -\epsilon_0\delta^{1/3}, \\ \operatorname{Re}(\lambda_{0,j_\star+1}(\xi) - \lambda) > \epsilon_0\delta^{1/3}. \end{cases}$$

Since the dispersion surfaces are labeled in increasing order, we deduce that if (4-1) is satisfied then

$$\operatorname{dist}(\Sigma_{L^2_\xi}(P_0(\xi)), \lambda) \ge \epsilon_0\delta^{1/3}, \quad (P_0(\xi) - \lambda)^{-1} = \mathscr{O}_{L^2_\xi}(\delta^{-1/3}).$$

We derived the estimate on $(P_0(\xi) - \lambda)^{-1}$ using the spectral theorem.

3. Assume that (4-1) holds. Thanks to step 1, $P_0(\xi) - \lambda$ is invertible and

$$P_\delta(\xi) - \lambda = P_0(\xi) - \lambda + \delta W = (P_0(\xi) - \lambda) \cdot (\mathrm{Id} + (P_0(\xi) - \lambda)^{-1}\delta W).$$

The second term equals $\mathrm{Id} + \mathscr{O}_{L^2_\xi}(\delta^{2/3})$. In particular it is invertible by a Neumann series for δ sufficiently small, with uniformly bounded inverse. We deduce that $P_\delta(\xi) - \lambda$ is invertible with inverse $\mathscr{O}_{L^2_\xi}(\delta^{-1/3})$.

4. To conclude we must show that the inverse of $P_\delta(\xi) - \lambda$ is $\mathscr{O}_{L^2_\xi \to H^2_\xi}(\delta^{-1/3})$. This is a standard consequence of the elliptic estimate: using $\delta = O(1)$, $\lambda = O(1)$, we see that for any $f \in H^2_\xi$,

$$|f|_{H^2_\xi} \le |f|_{L^2_\xi} + |\Delta f|_{H^2_\xi} \le C|f|_{L^2_\xi} + |(P_\delta(\xi) - \lambda)f|_{H^2_\xi}.$$

We apply this inequality to $f = (P_\delta(\xi) - \lambda)^{-1}u$ to deduce that

$$\|(P_\delta(\xi) - \lambda)^{-1}\|_{L^2_\xi \to H^2_\xi} \le C\|(P_\delta(\xi) - \lambda)^{-1}\|_{L^2_\xi} + 1.$$

In particular, the estimate $\mathscr{O}_{L^2_\xi}(\delta^{-1/3})$ proved in step 3 improves automatically to a bound $\mathscr{O}_{L^2_\xi \to H^2_\xi}(\delta^{-1/3})$. This completes the proof. □

4B. ***Resolvent near Dirac momenta.*** Fix a Dirac point $(\xi_\star, E_\star)$ of $P_0(\xi)$ and assume that $\vartheta_\star$ — defined in (1-6) — is nonzero. Identify $\xi - \xi_\star \in \mathbb{R}^2$ with the corresponding complex number and introduce the 2×2 matrix $M_\delta(\xi)$,

$$M_\delta(\xi) \stackrel{\text{def}}{=} \begin{bmatrix} E_\star + \delta\vartheta_\star & \nu_\star \cdot (\xi - \xi_\star) \\ \overline{\nu_\star \cdot (\xi - \xi_\star)} & E_\star - \delta\vartheta_\star \end{bmatrix}.$$

Lemma 4.2. *Let $\theta \in (0, 1)$. If*

$$\delta > 0, \quad \xi \in \mathbb{R}^2, \quad \vartheta_F \stackrel{\text{def}}{=} |\vartheta_\star| \ne 0, \quad \lambda \in \mathbb{D}\big(E_\star, \theta\sqrt{\vartheta_F^2 \cdot \delta^2 + \nu_F^2 \cdot |\xi - \xi_\star|^2}\big) \tag{4-4}$$

then the matrix $M_\delta(\xi) - \lambda$ is invertible and

$$\|(M_\delta(\xi) - \lambda)^{-1}\|_{\mathbb{C}^2} = O((\delta + |\xi - \xi_\star|)^{-1}).$$

Proof. The matrix $M_\delta(\xi)$ is Hermitian. It has eigenvalues

$$\mu_\delta^\pm(\xi) \overset{\text{def}}{=} E_\star \pm \sqrt{\vartheta_F^2 \cdot \delta^2 + \nu_F^2 \cdot |\xi - \xi_\star|^2}.$$

If (4-4) holds then the eigenvalues $\mu_\delta^\pm(\xi) - \lambda$ of $M_\delta(\xi) - \lambda$ satisfy

$$|\mu_\delta^\pm(\xi) - \lambda| \geq (1-\theta)\sqrt{\vartheta_F^2 \cdot \delta^2 + \nu_F^2 \cdot |\xi - \xi_\star|^2} \geq \frac{1-\theta}{\sqrt{2}} \cdot (\nu_F \cdot |\xi - \xi_\star| + \vartheta_F \cdot \delta).$$

By the spectral theorem, we deduce that $(M_\delta(\xi) - \lambda)^{-1}$ exists and has operator-norm bounded by $O((|\xi - \xi_\star| + \delta)^{-1})$. □

Introduce the operator

$$\Pi_0(\xi) : L^2_\xi \to \mathbb{C}^2, \quad \Pi_0(\xi)u \overset{\text{def}}{=} \begin{bmatrix} \langle e^{i\langle \xi - \xi_\star, x\rangle}\phi_1, u\rangle_{L^2_\xi} \\ \langle e^{i\langle \xi - \xi_\star, x\rangle}\phi_2, u\rangle_{L^2_\xi} \end{bmatrix}. \tag{4-5}$$

Lemma 4.3. *Assume that the assumptions* (H1) *and* (H3) *hold. Let* $\theta \in (0,1)$. *There exists* $\delta_0 > 0$ *such that if*

$$\delta \in (0, \delta_0), \quad |\xi - \xi_\star| \leq \delta^{1/3}, \quad \lambda \in \mathbb{D}\big(E_\star, \theta\sqrt{\vartheta_F^2 \cdot \delta^2 + \nu_F^2 \cdot |\xi - \xi_\star|^2}\big) \tag{4-6}$$

then $P_\delta(\xi) - \lambda$ *is invertible and*

$$(P_\delta(\xi) - \lambda)^{-1} = \Pi_0(\xi)^* \cdot (M_\delta(\xi) - \lambda)^{-1} \cdot \Pi_0(\xi) + \mathscr{O}_{L^2_\xi \to H^2_\xi}(1).$$

Proof. 1. Introduce the ξ-dependent family of vector spaces

$$\mathscr{V}(\xi) = \mathbb{C} \cdot e^{i\langle \xi - \xi_\star, x\rangle}\phi_1 \oplus \mathbb{C} \cdot e^{i\langle \xi - \xi_\star, x\rangle}\phi_2.$$

We split L^2_ξ as $\mathscr{V}(\xi) \oplus \mathscr{V}(\xi)^\perp$. With respect to this decomposition, we write $P_\delta(\xi)$ as a block-by-block operator:

$$P_\delta(\xi) - \lambda = \begin{bmatrix} A_\delta(\xi) - \lambda & B_\delta(\xi) \\ C_\delta(\xi) & D_\delta(\xi) - \lambda \end{bmatrix}. \tag{4-7}$$

We use below $\langle \cdot, \cdot \rangle$ to denote the L^2_ξ-scalar product.

2. We show that

$$B_\delta(\xi) = \mathscr{O}_{\mathscr{V}(\xi)^\perp \to \mathscr{V}(\xi)}(\delta + |\xi - \xi_\star|), \quad C_\delta(\xi) = \mathscr{O}_{\mathscr{V}(\xi) \to \mathscr{V}(\xi)^\perp}(\delta + |\xi - \xi_\star|). \tag{4-8}$$

Note that $C_\delta(\xi) = B_\delta(\xi)^*$; hence we just have to estimate $B_\delta(\xi)$, i.e., show that

$$u \in \mathscr{V}(\xi)^\perp, \ \ |u|_{L^2_\xi} = 1 \quad \Longrightarrow \quad \langle e^{i\langle \xi - \xi_\star, x\rangle}\phi_j, P_\delta(\xi)u\rangle = O(\delta + |\xi - \xi_\star|), \tag{4-9}$$

where the implicit constant does not depend on u. We have

$$\begin{aligned}
&\langle e^{i\langle \xi - \xi_\star, x\rangle}\phi_j, P_\delta(\xi)u\rangle \\
&\quad = \langle P_\delta(\xi) \cdot e^{i\langle \xi - \xi_\star, x\rangle}\phi_j, u\rangle = \langle (-\Delta + V + \delta W) \cdot e^{i\langle \xi - \xi_\star, x\rangle}\phi_j, u\rangle \\
&\quad = \langle e^{i\langle \xi - \xi_\star, x\rangle}(-\Delta + V)\phi_j, u\rangle + \langle [-\Delta, e^{i\langle \xi - \xi_\star, x\rangle}]\phi_j, u\rangle + \delta\langle We^{i\langle \xi - \xi_\star, x\rangle}\phi_j, u\rangle \\
&\quad = (E_\star + |\xi - \xi_\star|^2)\langle e^{i\langle \xi - \xi_\star, x\rangle}\phi_j, u\rangle + 2\langle e^{i\langle \xi - \xi_\star, x\rangle}(\xi - \xi_\star) \cdot D_x\phi_j, u\rangle + \delta\langle We^{i\langle \xi - \xi_\star, x\rangle}\phi_j, u\rangle.
\end{aligned}$$

The first bracket vanishes because $u \in \mathscr{V}(\xi)^\perp$. The second and third brackets are $O(\xi - \xi_\star)$ and $O(\delta)$, respectively — and this holds uniformly in u with $|u|_{L^2_\xi} = 1$. This gives (4-9), itself implying (4-8).

3. Here we prove that if (4-6) is satisfied then

$$D_\delta(\xi) - \lambda : \mathscr{V}(\xi)^\perp \cap H^2_\xi \to \mathscr{V}(\xi)^\perp \cap L^2_\xi \text{ is invertible} \quad \text{and} \quad (D_\delta(\xi) - \lambda)^{-1} = \mathscr{O}_{\mathscr{V}(\xi)^\perp}(1).$$

It suffices to construct an operator $E_\delta(\xi, \lambda) : \mathscr{V}(\xi)^\perp \to \mathscr{V}(\xi)^\perp$ such that

$$E_\delta(\xi, \lambda) = \mathscr{O}_{\mathscr{V}(\xi)^\perp}(1), \quad E_\delta(\xi, \lambda) \cdot (D_\delta(\xi) - \lambda) = \mathrm{Id}_{\mathscr{V}(\xi)^\perp} + \mathscr{O}_{\mathscr{V}(\xi)^\perp}(\delta + |\xi - \xi_\star|). \tag{4-10}$$

The space $\mathscr{V}(\xi_\star) = \ker_{L^2_{\xi_\star}}(P_0(\xi_\star) - E_\star)$ has dimension 2; $P_0(\xi)$ depends smoothly on ξ in the sense that $e^{-i\xi x} \cdot P_0(\xi) \cdot e^{i\xi x}$ forms a smooth family of operators $H^2_0 \to L^2_0$. Therefore, there exist $\eta > 0$ and $\epsilon > 0$ such that

$$|\xi - \xi_\star| \le \epsilon \quad \Longrightarrow \quad P_0(\xi) \text{ has precisely two eigenvalues in } [E_\star - \eta, E_\star + \eta]. \tag{4-11}$$

See [Kato 1980, §VII.1.3, Theorem 1.8]. Let $\mathscr{W}(\xi)$ be the vector space spanned by the two eigenvectors of $P_0(\xi)$ with energy in $[E_\star - \eta, E_\star + \eta]$. Let $Q_0(\xi)$ be the operator formally equal to $P_0(\xi)$ but acting on $\mathscr{W}(\xi)^\perp$. From (4-11), for $|\xi - \xi_\star| \le \epsilon$, the spectrum of $Q_0(\xi)$ consists of the eigenvalues of $P_0(\xi)$ outside $[E_\star - \eta, E_\star + \eta]$. The spectral theorem implies that if δ_0 is small enough, under (4-6),

$$Q_0(\xi) - \lambda : \mathcal{W}(\xi)^\perp \cap H^2_\xi \to \mathcal{W}(\xi)^\perp \cap L^2_\xi \text{ is invertible} \quad \text{and} \quad (Q_0(\xi) - \lambda)^{-1} = \mathscr{O}_{\mathscr{W}(\xi)^\perp}(1). \tag{4-12}$$

Let $J(\xi) : \mathscr{V}(\xi)^\perp \to \mathscr{W}(\xi)^\perp$ be obtained by orthogonally projecting an element $u \in \mathscr{V}(\xi)^\perp \subset L^2_\xi$ to $\mathscr{W}(\xi)^\perp$. We set

$$E_\delta(\xi, \lambda) \stackrel{\text{def}}{=} J(\xi)^* \cdot (Q_0(\xi) - \lambda)^{-1} \cdot J(\xi) : \mathscr{V}(\xi)^\perp \to \mathscr{V}(\xi)^\perp.$$

The first estimate of (4-10) is satisfied because of (4-12). We want to check the second estimate. Observe that

$$\begin{aligned} E_\delta(\xi, \lambda) \cdot (D_\delta(\xi) - \lambda) &= E_\delta(\xi, \lambda) \cdot \pi_{\mathscr{V}(\xi)^\perp}(P_0(\xi) - \lambda + \delta W) \\ &= J(\xi)^* \cdot (Q_0(\xi) - \lambda)^{-1} \cdot J(\xi) \cdot \pi_{\mathscr{V}(\xi)^\perp}(P_0(\xi) - \lambda) + \mathscr{O}_{\mathscr{V}(\xi)}(\delta). \end{aligned} \tag{4-13}$$

Above, $\pi_{\mathscr{V}(\xi)^\perp} : L^2_\xi \to L^2_\xi$ is the orthogonal projection from L^2_ξ to $\mathscr{V}(\xi)^\perp$, also seen as an operator $L^2_\xi \mapsto \mathscr{V}(\xi)^\perp$. We introduce similarly $\pi_{\mathcal{W}(\xi)^\perp}$. Then

$$\begin{aligned} J(\xi) \cdot \pi_{\mathscr{V}(\xi)^\perp} &= \pi_{\mathscr{W}(\xi)^\perp} \cdot (\mathrm{Id} - \pi_{\mathscr{V}(\xi)}) = \pi_{\mathscr{W}(\xi)^\perp} - (\mathrm{Id} - \pi_{\mathscr{W}(\xi)}) \cdot \pi_{\mathscr{V}(\xi)} \\ &= \pi_{\mathscr{W}(\xi)^\perp} - (\pi_{\mathscr{V}(\xi)} - \pi_{\mathscr{W}(\xi)}) \cdot \pi_{\mathscr{V}(\xi)}. \end{aligned} \tag{4-14}$$

The individual eigenvectors associated to the eigenvalues of $P_0(\xi)$ in $[E_\star - \eta, E_\star + \eta]$ do not depend smoothly on ξ but the projector $\pi_{\mathscr{W}(\xi)}$ depends smoothly on ξ — see [Kato 1980, §VII1.3, Theorem 1.7]. Since $\mathscr{V}(\xi_\star) = \mathscr{W}(\xi_\star)$, this implies $\pi_{\mathscr{V}(\xi)} - \pi_{\mathscr{W}(\xi)} = \mathscr{O}_{L^2_\xi}(\xi - \xi_\star)$. We deduce that

$$J(\xi) \cdot \pi_{\mathscr{V}(\xi)^\perp} = \pi_{\mathscr{W}(\xi)^\perp} + \mathscr{O}_{\mathscr{W}(\xi)^\perp}(\xi - \xi_\star). \tag{4-15}$$

We combine (4-13) and (4-15) to obtain

$$\begin{aligned} E_\delta(\xi,\lambda)\cdot(D_\delta(\xi)-\lambda) &= J(\xi)^*\cdot(Q_0(\xi)-\lambda)^{-1}\cdot\pi_{\mathscr{W}(\xi)^\perp}(P_0(\xi)-\lambda)+\mathscr{O}_{L^2_\xi}(\delta)\\ &= J(\xi)^*\pi_{\mathscr{W}(\xi)^\perp}+\mathscr{O}_{\mathscr{V}(\xi)^\perp}(\delta+|\xi-\xi_\star|). \end{aligned}$$

The operator $J(\xi)^*$ takes an element in $\mathscr{W}(\xi)^\perp$ and projects it to $\mathscr{V}(\xi)^\perp$. By the same argument as (4-14) and (4-15) (inverting $\mathscr{V}(\xi)$ and $\mathscr{W}(\xi)$),

$$J(\xi)^*\pi_{\mathscr{W}(\xi)^\perp}=\pi_{\mathscr{V}(\xi)^\perp}+O_{\mathscr{V}(\xi)^\perp}(\xi-\xi_\star).$$

We conclude that the second estimate of (4-10) is satisfied. It follows that $D_\delta(\xi)-\lambda:\mathscr{V}(\xi)^\perp\to\mathscr{V}(\xi)^\perp$ is invertible under (4-6).

4. We now study $A_\delta(\xi)-\lambda$. This operator acts on the two-dimensional space $\mathscr{V}(\xi)$; its matrix in the basis $\{e^{i\langle\xi-\xi_\star,x\rangle}\phi_1, e^{i\langle\xi-\xi_\star,x\rangle}\phi_2\}$ is

$$\begin{bmatrix} \langle e^{i\langle\xi-\xi_\star,x\rangle}\phi_1,(P_\delta(\xi)-\lambda)e^{i\langle\xi-\xi_\star,x\rangle}\phi_1\rangle & \langle e^{i\langle\xi-\xi_\star,x\rangle}\phi_1,(P_\delta(\xi)-\lambda)e^{i\langle\xi-\xi_\star,x\rangle}\phi_2\rangle\\ \langle e^{i\langle\xi-\xi_\star,x\rangle}\phi_2,(P_\delta(\xi)-\lambda)e^{i\langle\xi-\xi_\star,x\rangle}\phi_1\rangle & \langle e^{i\langle\xi-\xi_\star,x\rangle}\phi_2,(P_\delta(\xi)-\lambda)e^{i\langle\xi-\xi_\star,x\rangle}\phi_2\rangle \end{bmatrix}. \tag{4-16}$$

We observe that

$$\begin{aligned} e^{-i\langle\xi-\xi_\star,x\rangle}(P_\delta(\xi)-\lambda)e^{i\langle\xi-\xi_\star,x\rangle} &= P_\delta(\xi_\star)-\lambda+[e^{-i\langle\xi-\xi_\star,x\rangle},-\Delta]e^{i\langle\xi-\xi_\star,x\rangle}\\ &= P_\delta(\xi_\star)-\lambda+[\Delta,e^{-i\langle\xi-\xi_\star,x\rangle}]e^{i\langle\xi-\xi_\star,x\rangle}\\ &= P_\delta(\xi_\star)-\lambda+2((\xi-\xi_\star)\cdot D_x)-|\xi-\xi_\star|^2. \end{aligned}$$

Therefore the matrix elements in (4-16) are given by

$$\begin{aligned} \big\langle e^{i\langle\xi-\xi_\star,x\rangle}\phi_j,(P_\delta(\xi)-\lambda)e^{i\langle\xi-\xi_\star,x\rangle}\phi_k\big\rangle &= \big\langle\phi_j,\big(P_\delta(\xi_\star)+2(\xi-\xi_\star)\cdot D_x-\lambda-|\xi-\xi_\star|^2\big)\phi_k\big\rangle\\ &= (E_\star-|\xi-\xi_\star|^2-\lambda)\delta_{jk}+\big\langle\phi_j,(\delta W+2(\xi-\xi_\star)\cdot D_x)\phi_k\big\rangle. \end{aligned}$$

We deduce from Lemmas 2.2 and 2.3 that the matrix (4-16) is equal to $M_\delta(\xi)-\lambda+O(\xi-\xi_\star)^2$. Using a Neumann series argument based on (4-16), when (4-6) holds, $A_\delta(\xi)-\lambda$ is invertible, and

$$(A_\delta(\xi)-\lambda)^{-1}=\Pi_0(\xi)^*\cdot(M_\delta(\xi)-\lambda)^{-1}\cdot\Pi_0(\xi)+\mathscr{O}_{\mathscr{V}(\xi)}\bigg(\frac{|\xi-\xi_\star|^2}{\delta^2+|\xi-\xi_\star|^2}\bigg). \tag{4-17}$$

Because of Lemma 4.2, we also observe that

$$(A_\delta(\xi)-\lambda)^{-1}=\mathscr{O}_{\mathscr{V}(\xi)}((\delta+|\xi-\xi_\star|)^{-1}). \tag{4-18}$$

5. Schur's lemma allows us to invert block-by-block operators of the form (4-7) under certain conditions on the blocks; see [Drouot et al. 2018, Lemma 4.1] for the version needed here. We need to verify that

$$\begin{gathered} A_\delta(\xi)-\lambda:\mathscr{V}(\xi)\to\mathscr{V}(\xi)\ \text{ is invertible,}\\ D_\delta(\xi)-\lambda-C_\delta(\xi)\cdot(A_\delta(\xi)-\lambda)^{-1}\cdot B_\delta(\xi):\mathscr{V}(\xi)^\perp\to\mathscr{V}(\xi)^\perp\ \text{ is invertible.} \end{gathered} \tag{4-19}$$

The first statement holds because of step 4. Regarding the second statement, we observe that because of (4-8) and (4-18),

$$C_\delta(\xi)\cdot(A_\delta(\xi)-\lambda)^{-1}\cdot B_\delta(\xi)=\mathscr{O}_{\mathscr{V}(\xi)^\perp}(\delta+|\xi-\xi_\star|)=\mathscr{O}_{\mathscr{V}(\xi)^\perp}(\delta^{1/3}).$$

Because of step 3, $D_\delta(\xi)-\lambda$ is invertible and its inverse is $\mathscr{O}_{\mathscr{V}(\xi)^\perp}(1)$. Therefore a Neumann-series argument shows that the second statement in (4-19) holds. It also shows that the inverse is $\mathscr{O}_{\mathscr{V}(\xi)^\perp}(1)$.

We apply Schur's lemma — see [Drouot et al. 2018, Lemma 4.1]. From (4-7), we obtain that $P_\delta(\xi)-\lambda : H^2_\xi \to L^2_\xi$ is invertible when (4-6) holds, and moreover

$$(P_\delta(\xi)-\lambda)^{-1}=\begin{bmatrix}(A_\delta(\xi)-\lambda)^{-1} & 0\\ 0 & 0\end{bmatrix}+\mathscr{O}_{L^2_\xi}(1).$$

Using (4-17) and the projector (4-5), we deduce that

$$(P_\delta(\xi)-\lambda)^{-1}=\Pi_0(\xi)^*\cdot(M_\delta(\xi)-\lambda)^{-1}\cdot\Pi_0(\xi)+\mathscr{O}_{L^2_\xi}(1). \tag{4-20}$$

The error term in (4-20) improves automatically to $\mathscr{O}_{L^2_\xi\to H^2_\xi}(1)$ because of elliptic regularity — see the argument at the end of the proof of Lemma 4.1. □

5. The bulk resolvent along the edge

Let $v\in\Lambda$ be the direction of an edge. We define accordingly v', k, k' and ℓ — see Section 2E. For $\zeta\in\mathbb{R}$, we set

$$L^2[\zeta]\stackrel{\text{def}}{=}\big\{u\in L^2_{\text{loc}}(\mathbb{R}^2,\mathbb{C}) : u(x+v)=e^{i\zeta}u(x),\ \textstyle\int_{\mathbb{R}^2/\mathbb{Z}v}|u(x)|^2\,dx<\infty\big\}.$$

Let $P_\delta[\zeta]$ be the operator formally equal to P_δ but acting on $L^2[\zeta]$. We are interested in the resolvent of $P_\delta[\zeta]$ for δ small and ζ near $\zeta_\star=\langle\xi_\star, v\rangle$. We recall

$$\begin{aligned}
&\Pi : L^2(\mathbb{R}^2/\mathbb{Z}v,\mathbb{C}^2)\to L^2(\mathbb{R},\mathbb{C}^2), && (\Pi f)(t)\stackrel{\text{def}}{=}\textstyle\int_0^1 f(sv+tv')\,ds,\\
&\Pi^* : L^2(\mathbb{R},\mathbb{C}^2)\to L^2(\mathbb{R}^2/\mathbb{Z}v,\mathbb{C}^2), && (\Pi^* g)(x)\stackrel{\text{def}}{=} g(\langle k', x\rangle),\\
&\mathcal{U}_\delta : L^2(\mathbb{R},\mathbb{C}^2)\to L^2(\mathbb{R},\mathbb{C}^2), && (\mathcal{U}_\delta f)(t)\stackrel{\text{def}}{=} f(\delta t).
\end{aligned} \tag{5-1}$$

Let $\not{D}_\pm(\mu) : H^1(\mathbb{R},\mathbb{C}^2)\to L^2(\mathbb{R},\mathbb{C}^2)$ be the formal limits of $\not{D}(\mu)$ as $t\to\pm\infty$:

$$\not{D}_\pm(\mu)\stackrel{\text{def}}{=}\begin{bmatrix}\vartheta_\star & \nu_\star k'\\ \nu_\star k' & -\vartheta_\star\end{bmatrix}D_t+\mu\begin{bmatrix}0 & \nu_\star\ell\\ \nu_\star\ell & 0\end{bmatrix}\pm\begin{bmatrix}\vartheta_\star & 0\\ 0 & -\vartheta_\star\end{bmatrix}.$$

The main result of this section relates the resolvent of $P_{\pm\delta}[\zeta]$ at $E_\star+\delta z$ to that of $\not{D}_\pm(\mu)$ at z for small enough δ. The assumptions (H1)–(H3) were defined in Section 1E.

Theorem 5.1. *Assume that the assumptions* (H1)–(H3) *hold and fix* $\mu_\sharp>0$ *and* $\theta\in(0,1)$. *There exists* $\delta_0>0$ *such that if*

$$\delta\in(0,\delta_0),\quad \mu\in(-\mu_\sharp,\mu_\sharp),\quad z\in\mathbb{D}\big(0,\theta\sqrt{\vartheta_F^2+\mu^2\cdot\nu_F^2|\ell|^2}\big),$$
$$\zeta=\zeta_\star+\delta\mu,\quad \lambda=E_\star+\delta z$$

then the operators $P_{\pm\delta}[\zeta]-\lambda : H^2[\zeta]\to L^2[\zeta]$ *are invertible. Furthermore,*

$$(P_{\pm\delta}[\zeta]-\lambda)^{-1} = S_{\pm\delta}(\mu,z)+\mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}),$$
$$(k'\cdot D_x)(P_{\pm\delta}[\zeta]-\lambda)^{-1} = S^D_{\pm\delta}(\mu,z)+\mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}),$$

where

$$S_{\pm\delta}(\mu,z)\overset{\text{def}}{=}\frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^{\top} e^{i\mu\delta\langle\ell,x\rangle}\Pi^*\cdot\mathcal{U}_\delta(\not{D}_\pm(\mu)-z)^{-1}\mathcal{U}_\delta^{-1}\cdot\Pi e^{-i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}},$$
$$S^D_{\pm\delta}(\mu,z)\overset{\text{def}}{=}\frac{1}{\delta}\cdot\begin{bmatrix}(k'\cdot D_x)\phi_1\\ (k'\cdot D_x)\phi_2\end{bmatrix}^{\top} e^{i\mu\delta\langle\ell,x\rangle}\Pi^*\cdot\mathcal{U}_\delta(\not{D}_\pm(\mu)-z)^{-1}\mathcal{U}_\delta^{-1}\cdot\Pi e^{-i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}.$$

5A. ***Strategy.*** We first observe that it suffices to prove Theorem 5.1 for $P_\delta[\zeta]$. Indeed, to go from $P_\delta[\zeta]$ to $P_{-\delta}[\zeta]$ we simply replace W with $-W$. The only parameter to change is $\vartheta_\star$, which becomes $-\vartheta_\star$. This simply transforms $\not{D}_+(\mu)$ to $\not{D}_-(\mu)$.

To prove Theorem 5.1, we decompose $P_\delta[\zeta]$ fiberwise using the operators $P_\delta(\xi)$ (formally equal to P_δ but acting on L^2_ξ). Specifically

$$P_\delta[\zeta]=\frac{1}{2\pi}\int^{\oplus}_{\mathbb{R}/(2\pi\mathbb{Z})} P_\delta(\zeta k+\tau k')\cdot d\tau=\frac{1}{2\pi}\int^{\oplus}_{[0,2\pi]} P_\delta(\zeta k+\tau k')\cdot d\tau.$$

When $P_\delta[\zeta]-\lambda$ is invertible, we are interested in the resolvent

$$(P_\delta[\zeta]-\lambda)^{-1}=\frac{1}{2\pi}\int^{\oplus}_{[0,2\pi]}(P_\delta(\zeta k+\tau k')-\lambda)^{-1}\,d\tau. \tag{5-2}$$

The fiber resolvents $(P_\delta(\zeta k+\tau k')-\lambda)^{-1}$ were studied in Section 4. We first show that if $\zeta k+\tau k'$ satisfies $\rho(\zeta k+\tau k')\geq\delta^{1/3}$ then this quasimomentum does not contribute significantly to the resolvent $(P_\delta[\zeta]-\lambda)^{-1}$.

Then we study the contributions from quasimomenta $\zeta k+\tau k'$ at distance at most $\delta^{1/3}$ from $\xi_\star$. The Dirac operator $\not{D}_+(\mu)$ emerges from a rescaled direct integration of the dominant rank-2 matrix exhibited in Lemma 4.3.

5B. ***Reduction to*** $\boldsymbol{\zeta k+\tau k'}$ ***near*** $\boldsymbol{\xi_\star}$**.** We start the proof of Theorem 5.1. Below $\theta\in(0,1)$ and $\mu_\sharp>0$ are fixed numbers. Let n be the integer such that

$$\langle\xi_\star,v'\rangle\in[2\pi n,2\pi n+2\pi).$$

Write $\xi=\zeta k+\tau k',\ \tau\in[2\pi n,2\pi n+2\pi)$, and introduce

$$I\overset{\text{def}}{=}\{\tau\in[2\pi n,2\pi n+2\pi):|\xi-\xi_\star|\leq\delta^{1/3}\},\qquad I^c\overset{\text{def}}{=}[2\pi n,2\pi n+2\pi)\setminus I.$$

Observe that $\rho(\xi)=\delta|\mu\ell|$. In particular for δ small enough $\rho(\xi)$ is smaller than the threshold ϵ_0 given by Lemma 4.1. That lemma yields

$$\begin{aligned}(P_\delta[\zeta]-\lambda)^{-1}&=\frac{1}{2\pi}\int^{\oplus}_{\tau\in I}(P_\delta(\zeta k+\tau k')-\lambda)^{-1}\,d\tau+\frac{1}{2\pi}\int^{\oplus}_{\tau\in I^c}(P_\delta(\zeta k+\tau k')-\lambda)^{-1}\,d\tau\\&=\frac{1}{2\pi}\int^{\oplus}_{\tau\in I}(P_\delta(\zeta k+\tau k')-\lambda)^{-1}\,d\tau+\mathscr{O}_{L^2[\zeta]\to H^2[\zeta]}(\delta^{-1/3}).\end{aligned} \tag{5-3}$$

We would like to apply Lemma 4.3 to the leading term of (5-3). We check the assumptions: we must verify that when λ belongs to the range allowed in Theorem 5.1, λ belongs to the range required by Lemma 4.3. This is equivalent to

$$\mathbb{D}\big(E_\star, \theta\delta\sqrt{\vartheta_F^2 + \mu^2 \cdot \nu_F^2|\ell|^2}\big) \subset \mathbb{D}\big(E_\star, \theta\sqrt{\vartheta_F^2\delta^2 + \nu_F^2 \cdot |\xi - \xi_\star|^2}\big). \tag{5-4}$$

To check that (5-4) holds, we observe that

$$\begin{aligned} |\xi - \xi_\star|^2 &= |k'|^2(\tau - \tau_\star)^2 + \mu^2\delta^2|\ell|^2, \\ \tau_\star &\stackrel{\text{def}}{=} \langle \xi_\star, v'\rangle - \mu\delta\frac{\langle k, k'\rangle}{|k'|^2}, \quad \ell \stackrel{\text{def}}{=} k - \frac{\langle k, k'\rangle}{|k'|^2}k'. \end{aligned} \tag{5-5}$$

This implies

$$\theta\delta\sqrt{\vartheta_F^2 + \mu^2 \cdot \nu_F^2|\ell|^2} = \theta\sqrt{\vartheta_F^2\delta^2 + \mu^2\nu_F^2 \cdot |\ell|^2\delta^2} \le \theta\sqrt{\vartheta_F^2\delta^2 + \nu_F^2 \cdot |\xi - \xi_\star|^2}.$$

Therefore we can apply Lemma 4.3 to the leading term of (5-3). It shows that

$$\begin{aligned} &P_\delta[\zeta] = T_\delta[\zeta] + \mathscr{O}_{L^2[\zeta]\to H^2[\zeta]}(\delta^{-1/3}), \\ &T_\delta[\zeta] \stackrel{\text{def}}{=} \frac{1}{2\pi}\int_{\tau\in I}^{\oplus} \Pi_0(\zeta k + \tau k')^* \cdot (M_\delta(\zeta k + \tau k') - \lambda)^{-1} \cdot \Pi_0(\zeta k + \tau k')\, d\tau. \end{aligned} \tag{5-6}$$

Because of (5-5), $\tau_\star = \langle \xi_\star, v'\rangle + O(\delta)$. From Section 2E and the definition of n, we have $\langle \xi_\star, v'\rangle \in \big\{2\pi n + \frac{2\pi}{3}, 2\pi n + \frac{4\pi}{3}\big\}$. Hence $\tau_\star$ is in the interior of I for δ sufficiently small. It follows that I is an interval centered at $\tau_\star$:

$$I = [\tau_\star - \delta\cdot\alpha_\delta, \tau_\star + \delta\cdot\alpha_\delta], \quad \alpha_\delta \stackrel{\text{def}}{=} \frac{\sqrt{\delta^{2/3} - \mu^2 \cdot \nu_F^2|\ell|^2\delta^2}}{|k'|\delta} = \frac{\delta^{-2/3}}{|k'|} + O(\delta^{2/3}). \tag{5-7}$$

We make the substitution $\tau \mapsto \tau_\star + \delta\tau$. The vector $\zeta k + \tau k'$ becomes $\zeta k + (\tau_\star + \delta\tau)k' = \xi_\star + \delta(\tau k' + \mu\ell)$, the interval I becomes $[-\alpha_\delta, \alpha_\delta]$, $d\tau$ becomes $\delta\, d\tau$ and

$$\begin{aligned} M_\delta(\zeta k + \delta(\tau k' + \mu\ell)) &= E_\star + \delta\begin{bmatrix} \vartheta_\star & \nu_\star(\tau k' + \mu\ell) \\ \nu_\star(\tau k' + \mu\ell) & -\vartheta_\star \end{bmatrix}, \\ (M_\delta(\zeta k + \delta(\tau k' + \mu\ell)) - \lambda)^{-1} &= \frac{1}{\delta}\begin{bmatrix} \vartheta_\star - z & \nu_\star(\tau k' + \mu\ell) \\ \nu_\star(\tau k' + \mu\ell) & -\vartheta_\star - z \end{bmatrix}^{-1}, \quad z \stackrel{\text{def}}{=} \frac{\lambda - E_\star}{\delta}. \end{aligned}$$

We deduce that $T_\delta[\zeta]$ equals

$$\frac{1}{2\pi}\int_{|\tau|<\alpha_\delta}^{\oplus} \Pi_0(\xi_\star + \delta(\tau k' + \mu\ell))^* \cdot \begin{bmatrix} \vartheta_\star - z & \nu_\star(\tau k' + \mu\ell) \\ \nu_\star(\tau k' + \mu\ell) & -\vartheta_\star - z \end{bmatrix}^{-1} \cdot \Pi_0(\xi_\star + \delta(\tau k' + \mu\ell)) \cdot d\tau.$$

Thanks to the definition (5-7) of Π_0,

$$\Pi_0(\xi_\star + \delta(\tau k' + \mu\ell))u = \begin{bmatrix} \langle e^{i\delta\langle \tau k' + \mu\ell, x\rangle}\phi_1, u\rangle \\ \langle e^{i\delta\langle \tau k' + \mu\ell, x\rangle}\phi_2, u\rangle \end{bmatrix}.$$

We conclude that the operator $T_\delta[\zeta]$ has kernel

$$\frac{1}{2\pi}\begin{bmatrix}\phi_1(x)\\ \phi_2(x)\end{bmatrix}^\top \cdot \int_{|\tau|\leq\alpha_\delta}\begin{bmatrix}\dfrac{\vartheta_\star - z}{\nu_\star(\tau k'+\mu\ell)} & \nu_\star(\tau k'+\mu\ell)\\ -\vartheta_\star - z\end{bmatrix}^{-1} e^{i\delta\langle \tau k'+\mu\ell, x-y\rangle}\, d\tau \cdot \overline{\begin{bmatrix}\phi_1(y)\\ \phi_2(y)\end{bmatrix}}. \tag{5-8}$$

5C. ***Kernel identities and proof of Theorem 5.1.*** Recall that Π, Π^* and $\mathcal{U}_\delta$ are defined in (5-1).

Lemma 5.2. *There exists $C>0$ such that for every $\delta\in(0,1)$, the following holds. Let $\Psi\in L^\infty(\mathbb{R}, M_2(\mathbb{C}))$, possibly depending on δ, and A_Ψ be the operator with kernel*

$$(x,y)\mapsto \begin{bmatrix}\phi_1(x)\\ \phi_2(x)\end{bmatrix}^\top \cdot \frac{1}{2\pi}\int_{\mathbb{R}} \Psi(\tau) e^{i\delta\langle \tau k'+\mu\ell, x-y\rangle}\, d\tau \cdot \overline{\begin{bmatrix}\phi_1(y)\\ \phi_2(y)\end{bmatrix}}.$$

Then A_Ψ is bounded on $L^2[\zeta]$ with $\|A_\Psi\|_{L^2[\zeta]}\leq C\delta^{-1}|\Psi|_\infty$, and

$$A_\Psi = \frac{1}{\delta}\cdot \begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{i\mu\delta\langle \ell, x\rangle}\Pi^*\cdot \mathcal{U}_\delta \Psi(D_t)\mathcal{U}_\delta^{-1}\cdot \Pi e^{-i\mu\delta\langle \ell, x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}. \tag{5-9}$$

If in addition $\tau\cdot\Psi\in L^\infty(\mathbb{R}, M_2(\mathbb{C}))$ then $(k'\cdot D_x)A_\Psi$ is bounded on $L^2[\zeta]$ with

$$\|(k'\cdot D_x)A_\Psi\|_{L^2[\zeta]}\leq C\delta^{-1}|\Psi|_\infty + C|\tau\cdot\Psi|_\infty$$

and

$$(k'\cdot D_x)A_\Psi = \frac{1}{\delta}\cdot \begin{bmatrix}(k'\cdot D_x)\phi_1\\ (k'\cdot D_x)\phi_2\end{bmatrix}^\top e^{i\mu\delta\langle \ell, x\rangle}\Pi^*\cdot \mathcal{U}_\delta \Psi(D_t)\mathcal{U}_\delta^{-1}\cdot \Pi e^{-i\mu\delta\langle \ell, x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(|\langle\tau\rangle\cdot\psi|_\infty).$$

Proof. 1. We first note that the operator $\delta^{-1}\cdot\mathcal{U}_\delta\Psi(D_t)\mathcal{U}_\delta^{-1}$ has kernel

$$(t,t')\in\mathbb{R}\times\mathbb{R}^2\mapsto \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\delta\tau(t-t')}\Psi(\tau)\cdot d\tau. \tag{5-10}$$

Let δ_0 denote the Dirac mass. We claim that the operator Π has kernel

$$(t', y)\in\mathbb{R}\times\mathbb{R}^2/(\mathbb{Z}v)\mapsto \delta_0(\langle k', y\rangle - t'). \tag{5-11}$$

Fix $f\in C_0^\infty(\mathbb{R}^2/\mathbb{Z}v, \mathbb{C}^2)$. The integral

$$\int_{\mathbb{R}^2/\mathbb{Z}v}\delta_0(\langle k', y\rangle - t')f(y)\,dy$$

is well-defined. We perform the substitution $y\mapsto(\langle k, y\rangle, \langle k', y\rangle)$; the inverse substitution is $(s,t)\mapsto sv+tv'$; the Jacobian determinant is $dy = \mathrm{Det}[v,v']\cdot ds\,dt$. Since v, v' are related to v_1, v_2 by (2-4) and $\mathrm{Det}[v_1, v_2]=1$ because of (1-1), $\mathrm{Det}[v,v']=1$. The above integral becomes

$$\int_{\mathbb{R}^2/\mathbb{Z}e_1}\delta_0(t-t')f(sv+tv')\,ds\,dt = \int_{\mathbb{R}/\mathbb{Z}} f(sv+t'v')\,ds.$$

We recover the formula (5-1) for Πf. From (5-11), we deduce that the kernel of Π^* is

$$(x,t)\in\mathbb{R}^2/\mathbb{Z}v\times\mathbb{R}\mapsto\delta_0(\langle k', x\rangle - t). \tag{5-12}$$

To obtain (5-10), we compose the kernels (5-11), (5-10) and (5-12). This forces t to be $\langle k', x\rangle$ and t' to be $\langle k', y\rangle$. Hence the operator $\delta^{-1}\cdot\Pi^*\cdot\mathcal{U}_\delta\Psi(D_t)\mathcal{U}_\delta^{-1}\cdot\Pi$ has kernel

$$(x, y)\mapsto\frac{1}{2\pi}\int_{\mathbb{R}} e^{i\delta\tau\langle k', x-y\rangle}\Psi(\tau)\cdot d\tau.$$

This implies (5-9).

2. We prove the $L^2[\zeta]$-bound. The operator Π maps $L^2(\mathbb{R}^2/\mathbb{Z}v, \mathbb{C})$ to $L^2(\mathbb{R}, \mathbb{C}^2)$, independently of δ. Its adjoint maps $L^2(\mathbb{R}, \mathbb{C}^2)$ to $L^2(\mathbb{R}^2/\mathbb{Z}v, \mathbb{C})$, independently of δ. The dilations $\mathcal{U}_\delta$ and $\mathcal{U}_\delta^{-1}$ map $L^2(\mathbb{R}, \mathbb{C}^2)$ to itself, with bounds $\delta^{-1/2}$ and $\delta^{1/2}$, respectively. The operator $\Psi(D_t)$ is a Fourier multiplier; hence it maps $L^2(\mathbb{R}, \mathbb{C}^2)$ to itself, with bound $|\Psi|_\infty$. Combining all these bounds together we get

$$\|A_\Psi\|_{L^2[\zeta]}\le C\delta^{-1}|\Psi|_\infty.$$

3. We observe that the operator $(k'\cdot D_x)A_\Psi$ has kernel

$$\begin{bmatrix}(k'\cdot D_x)\phi_1(x)\\(k'\cdot D_x)\phi_2(x)\end{bmatrix}^\top\cdot\frac{1}{2\pi}\int_{\mathbb{R}}\Psi(\tau)e^{i\delta\langle\tau k'+\mu\ell, x-y\rangle}\,d\tau\cdot\overline{\begin{bmatrix}\phi_1(y)\\\phi_2(y)\end{bmatrix}}$$
$$+\begin{bmatrix}\phi_1(x)\\\phi_2(x)\end{bmatrix}^\top\cdot\frac{1}{2\pi}\int_{\mathbb{R}}\Psi(\tau)\cdot\tau\delta|k'|^2e^{it\delta\langle\tau k'+\mu\ell, x-y\rangle}\,d\tau\cdot\overline{\begin{bmatrix}\phi_1(y)\\\phi_2(y)\end{bmatrix}}.$$

Above, we used that $\ell\cdot k'=0$. These two terms are kernels of operators studied in steps 1 and 2. The first one has $L^2[\zeta]$-operator norm controlled by $C\delta^{-1}|\Psi|_\infty$ and the second one by $C|\tau\cdot\Psi|_\infty$. □

Lemma 5.3. *Let $\vartheta_\sharp\in(0,\vartheta_F)$. There exists $C>0$ such that for any $z\in\mathbb{D}(0,\vartheta_\sharp)$, the following holds. Let $\Psi_0:\mathbb{R}\to M_2(\mathbb{C})$ be given by*

$$\Psi_0(\tau)\overset{\text{def}}{=}\begin{bmatrix}\vartheta_\star-z & \nu_\star(\tau k'+\mu\ell)\\ \overline{\nu_\star(\tau k'+\mu\ell)} & -\vartheta_\star-z\end{bmatrix}^{-1}.\tag{5-13}$$

Then $\tau\cdot\Psi_0\in L^\infty(\mathbb{R}, M_2(\mathbb{C}))$ and for every $a\ge0$,

$$\sup_{|\tau|\ge a}\|\Psi_0(\tau)\|_{\mathbb{C}^2}\le Ca^{-1},\qquad\sup_{|\tau|\ge a}\|\tau\Psi_0(\tau)\|_{\mathbb{C}^2}\le C.\tag{5-14}$$

To prove Lemma 5.3, it suffices to observe that

$$\Psi_0(\tau)=-\frac{1}{|\vartheta_\star-z|^2+\nu_F^2|\ell|^2\mu^2+\nu_F^2|k'|^2\tau^2}\begin{bmatrix}\vartheta_\star-z & \nu_\star(\tau k'+\mu\ell)\\ \overline{\nu_\star(\tau k'+\mu\ell)} & -\vartheta_\star-z\end{bmatrix}.$$

In particular, $\Psi_0(\tau)=\mathscr{O}_{\mathbb{C}^2}(\tau^{-1})$. This yields the bounds (5-14). Let $\not{D}_+(\mu):H^1(\mathbb{R}, \mathbb{C}^2)\to L^2(\mathbb{R}, \mathbb{C}^2)$ be the Dirac operator defined by

$$\not{D}_+(\zeta)\overset{\text{def}}{=}\begin{bmatrix}\vartheta_\star & \nu_\star k'\\ \overline{\nu_\star k'} & -\vartheta_\star\end{bmatrix}D_t+\mu\begin{bmatrix}0 & \nu_\star\ell\\ \overline{\nu_\star\ell} & 0\end{bmatrix}+\begin{bmatrix}\vartheta_\star & 0\\ 0 & -\vartheta_\star\end{bmatrix}.$$

We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1. 1. Because of (5-6), it suffices to prove Theorem 5.1 when $P_\delta[\zeta]$ is replaced by $T_\delta[\zeta]$. We first apply Lemma 5.2 with Ψ_0 given by (5-13). It shows that

$$A_{\Psi_0} \overset{\text{def}}{=} \frac{1}{\delta} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^\top e^{i\mu\delta\langle \ell, x\rangle} \Pi^* \cdot \mathcal{U}_\delta (\not{D}_+(\mu) - z)^{-1} \mathcal{U}_\delta^{-1} \cdot \Pi e^{-i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}}$$

has kernel

$$(x, y) \mapsto \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \end{bmatrix}^\top \cdot \frac{1}{2\pi} \int_{\mathbb{R}} \Psi_0(\tau) e^{i\delta\langle \tau k' + \mu\ell, x-y\rangle}\, d\tau \cdot \overline{\begin{bmatrix} \phi_1(y) \\ \phi_2(y) \end{bmatrix}}.$$

2. We now apply Lemma 5.2 with $\Psi_1(\tau) \overset{\text{def}}{=} \mathbb{1}_{\mathbb{R}\setminus[-\alpha_\delta,\alpha_\delta]}(\tau) \cdot \Psi_0(\tau)$ (recall that $\alpha_\delta = |k'|^{-1}\delta^{-2/3} + O(\delta^{2/3})$ was defined in (5-7)). It shows that A_{Ψ_1} has kernel

$$\begin{bmatrix} \phi_1(x) \\ \phi_2(x) \end{bmatrix}^\top \cdot \frac{1}{2\pi} \int_{|\tau|\geq\alpha_\delta} \Psi_0(\tau) e^{i\delta\langle \tau k' + \mu\ell, x-y\rangle}\, d\tau \cdot \overline{\begin{bmatrix} \phi_1(y) \\ \phi_2(y) \end{bmatrix}}.$$

Thanks to the bounds of Lemma 5.3, $A_{\Psi_1} = \mathscr{O}_{L^2[\zeta]}(\delta^{-1}\alpha_\delta^{-1}) = \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3})$.

3. When we subtract the kernel of A_{Ψ_1} from the kernel of A_{Ψ_0}, we get the kernel of $T_\delta[\zeta]$; see (5-8). This shows that $T_\delta[\zeta] = A_{\Psi_0} - A_{\Psi_1}$. Hence

$$T_\delta[\zeta] = \frac{1}{\delta} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^\top e^{i\mu\delta\langle \ell, x\rangle} \Pi^* \cdot \mathcal{U}_\delta (\not{D}_+(\mu) - z)^{-1} \mathcal{U}_\delta^{-1} \cdot \Pi e^{-i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}).$$

4. Lemma 5.2 and the bounds of Lemma 5.3 imply that

$$(k' \cdot D_x) A_{\Psi_0} = \frac{1}{\delta} \cdot \begin{bmatrix} (k' \cdot D_x)\phi_1 \\ (k' \cdot D_x)\phi_2 \end{bmatrix}^\top e^{i\mu\delta\langle \ell, x\rangle} \Pi^* \cdot \mathcal{U}_\delta (\not{D}_+(\mu) - z)^{-1} \mathcal{U}_\delta^{-1} \cdot \Pi e^{-i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(1).$$

It also implies that $(k' \cdot D_x) A_{\Psi_1} = \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3})$. We conclude that

$$(k' \cdot D_x) T_\delta[\zeta] = \frac{1}{\delta} \cdot \begin{bmatrix} (k' \cdot D_x)\phi_1 \\ (k' \cdot D_x)\phi_2 \end{bmatrix}^\top e^{i\mu\delta\langle \ell, x\rangle} \Pi^* \cdot \mathcal{U}_\delta (\not{D}_+(\mu) - z)^{-1} \mathcal{U}_\delta^{-1} \cdot \Pi e^{-i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}). \quad \square$$

6. The resolvent of the edge operator

Recall that κ is a domain wall function — see (1-8) — and introduce the operator

$$\mathscr{P}_\delta = -\Delta + V + \delta \cdot \kappa_\delta \cdot W, \quad \kappa_\delta(x) = \kappa(\delta\langle k', x\rangle).$$

We denote by $\mathscr{P}_\delta[\zeta]$ the operator formally equal to $\mathscr{P}_\delta$ but acting on $L^2[\zeta]$. In this section we prove Theorem 3.2: we connect the resolvent of $\mathscr{P}_\delta[\zeta]$ to that of the Dirac operator $\not{D}(\mu)$ emerging from the multiscale analysis of [Fefferman et al. 2016b]:

$$\not{D}(\mu) = \begin{bmatrix} 0 & v_\star k' \\ \overline{v_\star k'} & 0 \end{bmatrix} D_t + \mu \begin{bmatrix} 0 & v_\star \ell \\ \overline{v_\star \ell} & 0 \end{bmatrix} + \vartheta_\star \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \kappa.$$

The strategy is as follows. We first prove a formula for $(\mathscr{P}_\delta[\zeta] - \lambda)^{-1}$ in terms of the asymptotic operators $(P_{\pm\delta}[\zeta] - \lambda)^{-1}$. We then apply Theorem 5.1 to exhibit the leading-order term in this formula.

We use a cyclicity argument to simplify this leading-order term. An averaging effect emerges as the driving phenomenon connecting $\mathscr{P}_\delta[\zeta]$ to $\not{D}(\mu)$. This yields Theorem 3.2.

6A. ***Parametrix.*** We first construct a parametrix for $\mathscr{P}_\delta[\zeta]-\lambda$. Introduce

$$\mathscr{Q}_\delta(\zeta,\lambda)\stackrel{\text{def}}{=}\sum_{\pm}\chi_{\pm,\delta}\cdot(P_{\pm\delta}[\zeta]-\lambda)^{-1},\quad \chi_\pm\stackrel{\text{def}}{=}\frac{1\pm\kappa}{2}. \tag{6-1}$$

This operator is well-defined — and depends holomorphically on λ — as long as $\lambda\notin\Sigma_{L^2[\zeta]}(P_\delta[\zeta])$. Formally speaking, it behaves asymptotically like $(\mathscr{P}_\delta[\zeta]-\lambda)^{-1}$.

A calculation similar to [Drouot et al. 2018, §5.2] yields

$$(\mathscr{P}_\delta[\zeta]-\lambda)\mathscr{Q}_\delta(\zeta,\lambda)=\mathrm{Id}+\mathscr{K}_\delta(\zeta,\lambda),$$

where

$$\mathscr{K}_\delta(\zeta,\lambda)=\frac{\delta}{2}\big(2(D_t\kappa)_\delta\cdot(k'\cdot D_x)+\delta|k'|^2(D_t^2\kappa)_\delta+(\kappa_\delta^2-1)W\big)\big((P_\delta[\zeta]-\lambda)^{-1}-(P_{-\delta}[\zeta]-\lambda)^{-1}\big).$$

This identity shows that if $\mathrm{Id}+\mathscr{K}_\delta(\zeta,\lambda)$ is invertible then $\mathscr{P}_\delta[\zeta]-\lambda$ is invertible. When this holds, $(\mathscr{P}_\delta[\zeta]-\lambda)^{-1}$ has an expression in terms of $\mathscr{Q}_\delta(\zeta,\lambda)$ and $\mathscr{K}_\delta(\zeta,\lambda)$:

$$(\mathscr{P}_\delta[\zeta]-\lambda)^{-1}=\mathscr{Q}_\delta(\zeta,\lambda)\cdot(\mathrm{Id}+\mathscr{K}_\delta(\zeta,\lambda))^{-1}.$$

The operators $\mathscr{Q}_\delta(\zeta,\lambda)$ and $\mathscr{K}_\delta(\zeta,\lambda)$ have expressions in terms of $(P_{\pm\delta}[\zeta]-\lambda)^{-1}$. An application of Theorem 5.1 estimates $\mathscr{Q}_\delta(\zeta,\lambda)$ and $\mathscr{K}_\delta(\zeta,\lambda)$, assuming

$$\begin{gathered}\delta\in(0,\delta_0),\quad \mu\in(-\mu_\sharp,\mu_\sharp),\quad z\in\mathbb{D}\big(0,\sqrt{\vartheta_F^2+\mu^2\cdot v_F^2|\ell|^2}\big),\\ \lambda=E_\star+\delta z,\quad \zeta=\zeta_\star+\delta\mu.\end{gathered} \tag{6-2}$$

We introduce the operator

$$R_0(\mu,z):L^2(\mathbb{R},\mathbb{C}^2)\to H^2(\mathbb{R},\mathbb{C}^2),$$
$$R_0(\mu,z)\stackrel{\text{def}}{=}(\not{D}_+(\mu)^2-z^2)^{-1}=(v_F^2|k'|^2D_t^2+\mu^2|v_\star\ell|^2+\vartheta_F^2-z^2)^{-1}.$$

It is well-defined when z is away from the spectrum of $\not{D}_\pm(\mu)$ — in particular when

$$z\in\mathbb{D}\big(0,\sqrt{\vartheta_F^2+\mu^2\cdot v_F^2|\ell|^2}\big).$$

Lemma 6.1. *Let $\mu_\sharp>0$, $\theta\in(0,1)$. There exists $\delta_0>0$ such that under the assumptions of Theorem 5.1, $\mathscr{K}_\delta(\zeta,\lambda)$ and $\mathscr{Q}_\delta(\zeta,\lambda)$ admit the expansions*

$$\mathscr{K}_\delta(\zeta,\lambda)=\mathcal{K}_\delta(\mu,z)+\mathscr{O}_{L^2[\zeta]}(\delta^{2/3}),\quad \mathscr{Q}_\delta(\zeta,\lambda)=\mathcal{Q}_\delta(\mu,z)+\mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}).$$

Above, $\mathcal{K}_\delta(\mu,z)$ is equal to

$$\vartheta_\star\left(2(D_t\kappa)_\delta\begin{bmatrix}k'\cdot D_x\phi_1\\ k'\cdot D_x\phi_2\end{bmatrix}^\top+(\kappa_\delta^2-1)W\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top\right)e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\mathcal{U}_\delta\boldsymbol{\sigma_3}\cdot R_0(\mu,z)\cdot\mathcal{U}_\delta^{-1}\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}$$

and

$$\mathcal{Q}_\delta(\mu,z)\stackrel{\text{def}}{=}\frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\mathcal{U}_\delta\cdot(\not{D}(\mu)+z)\cdot R_0(\mu,z)\cdot\mathcal{U}_\delta^{-1}\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}.$$

The proof is a calculation using the relation between $\mathscr{Q}_\delta(\zeta,\lambda)$ and $\mathscr{K}_\delta(\zeta,\lambda)$ with the edge resolvents $(P_{\pm\delta}[\zeta]-\lambda)^{-1}$, and the expansions of these resolvents provided by Theorem 5.1. We defer it to Appendix A.3.

6B. ***Weak convergence.*** We are interested in the eigenvalues of $\mathscr{P}_\delta[\zeta]$. We previously studied eigenvalue problems in seemingly different situations [Drouot 2018a; 2018c; 2018d], as well as in a one-dimensional analog [Drouot et al. 2018]. The proofs of these results rely on a *cyclicity* principle: if A and B are two matrices then the nonzero eigenvalues of AB and BA are equal (together with their multiplicity).

Although the leading-order terms $\mathcal{K}_\delta(\mu,z)$ and $\mathcal{Q}_\delta(\mu,z)$ have complicated expressions, they exhibit a structure favorable to applying the cyclicity principle. This will provide a simple formula for the product

$$\mathcal{Q}_\delta(\mu,z)\cdot(\mathrm{Id}+\mathcal{K}_\delta(\mu,z))^{-1}$$

and complete the proof of Theorem 3.2.

A preliminary step is the computation of a weak limit that arises when permuting factors in $\mathcal{K}_\delta(\mu,z)$: the operator $L^2(\mathbb{R},\mathbb{C}^2)\to L^2(\mathbb{R},\mathbb{C}^2)$ given by

$$\begin{aligned}\vartheta_\star\mathcal{U}_\delta^{-1}\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}\cdot\Big(2(D_t\kappa)_\delta\cdot\begin{bmatrix}k'\cdot D_x\phi_1\\ k'\cdot D_x\phi_2\end{bmatrix}^\top+(\kappa_\delta^2-1)W\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top\Big)\cdot e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\mathcal{U}_\delta\boldsymbol{\sigma_3}\\ =\vartheta_\star\mathcal{U}_\delta^{-1}\cdot\Pi\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}\Big(2(D_t\kappa)_\delta\cdot\begin{bmatrix}k'\cdot D_x\phi_1\\ k'\cdot D_x\phi_2\end{bmatrix}^\top+(\kappa_\delta^2-1)W\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top\Big)\boldsymbol{\sigma_3}\Pi^*\cdot\mathcal{U}_\delta.\end{aligned}\tag{6-3}$$

Lemma 6.2. *The operator* (6-3) *is a multiplication operator by a function* $\mathscr{U}^\delta:\mathbb{R}\to M_2(\mathbb{C})$ *with two-scale structure*:

$$\mathscr{U}^\delta(t)=\mathscr{U}\Big(\frac{t}{\delta},t\Big),\quad \mathscr{U}\in C_0^\infty(\mathbb{R}/\mathbb{Z}\times\mathbb{R},M_2(\mathbb{C})).\tag{6-4}$$

The function $\mathscr{U}^\delta$ *converges weakly to*

$$\mathscr{U}^0\in C_0^\infty(\mathbb{R},M_2(\mathbb{C})),\quad \mathscr{U}^0(t)\overset{\text{def}}{=}\vartheta_F^2(\kappa(t)^2-1)+\vartheta_\star\begin{bmatrix}0 & -\nu_\star k'\\ \overline{\nu_\star k'} & 0\end{bmatrix}(D_t\kappa)(t).\tag{6-5}$$

Finally, if $\mathscr{U}^\delta-\mathscr{U}^0$ *is seen as a multiplication operator from* H^1 *to* H^{-1},

$$\mathscr{U}^\delta-\mathscr{U}^0=\mathscr{O}_{H^1\to H^{-1}}(\delta).\tag{6-6}$$

Proof. 1. We set

$$\begin{aligned}F(x,t)&\overset{\text{def}}{=}\vartheta_\star\overline{\begin{bmatrix}\phi_1(x)\\ \phi_2(x)\end{bmatrix}}\Big((D_t\kappa)(t)\cdot\begin{bmatrix}2k'\cdot D_x\phi_1(x)\\ 2k'\cdot D_x\phi_2(x)\end{bmatrix}^\top+(\kappa(t)^2-1)W(x)\begin{bmatrix}\phi_1(x)\\ \phi_2(x)\end{bmatrix}^\top\Big)\boldsymbol{\sigma_3},\\ F^\delta(x)&\overset{\text{def}}{=}F(x,\delta\langle k',x\rangle).\end{aligned}$$

Fix $g\in C_0^\infty(\mathbb{R},\mathbb{C}^2)$. The action of the operator (6-3) on g is given by

$$\begin{aligned}(\mathcal{U}_\delta^{-1}\cdot\Pi F^\delta\Pi^*\cdot\mathcal{U}_\delta g)(t)&=\int_0^1 F^\delta\Big(sv+\frac{t}{\delta}v'\Big)g\Big(\Big\langle k',\delta\Big(sv+\frac{t}{\delta}v'\Big)\Big\rangle\Big)\,ds\\ &=\int_0^1 F\Big(sv+\frac{t}{\delta}v',t\Big)g(t)ds=\int_0^1 F\Big(sv+\frac{t}{\delta}v',t\Big)\,ds\cdot g(t).\end{aligned}$$

Therefore (6-3) is the multiplication operator by

$$\mathscr{U}^\delta(t) \stackrel{\text{def}}{=} \int_0^1 F\Big(sv + \frac{t}{\delta}v', t\Big)\, ds.$$

Note that F is Λ-periodic in x and compactly supported in t. Therefore $\mathscr{U}^\delta$ has the two-scale structure (6-4):

$$\mathscr{U}^\delta(t) = \mathscr{U}\Big(\frac{t}{\delta}, t\Big), \quad \mathscr{U}(\tau, t) \stackrel{\text{def}}{=} \int_0^1 F(sv + \tau v', t)\, ds. \tag{6-7}$$

2. The function $\mathscr{U}$ is periodic in the first variable and compactly supported in the second one. Therefore the weak limit of $\mathscr{U}^\delta$ is

$$\mathscr{U}^0(t) \stackrel{\text{def}}{=} \int_0^1 \mathscr{U}(\tau, t)\, d\tau = \int_0^1\int_0^1 F(sv + \tau v', t)\, d\tau\, ds = \int_{\mathbb{L}} F(x, t)\, dx. \tag{6-8}$$

In the last inequality, we changed variables: $sv + \tau v'$ became $sv_1 + \tau v_2$ (with Jacobian equal to 1); hence $[0, 1]^2$ became $\mathbb{L}$, the fundamental cell of $\mathbb{R}^2/\Lambda$ given in (1-2). Going back to the definition of F, we end up with

$$\begin{aligned}\mathscr{U}^0(t) &= \left(\begin{bmatrix}\vartheta_F^2 & 0\\ 0 & -\vartheta_F^2\end{bmatrix}(\kappa(t)^2 - 1) + \vartheta_\star\begin{bmatrix}0 & v_\star k'\\ \overline{v_\star k'} & 0\end{bmatrix}(D_t\kappa)(t)\right)\boldsymbol{\sigma_3}\\ &= \vartheta_F^2(\kappa(t)^2 - 1) + \vartheta_\star\begin{bmatrix}0 & -v_\star k'\\ \overline{v_\star k'} & 0\end{bmatrix}(D_t\kappa)(t).\end{aligned}$$

3. We show the quantitative estimate (6-6). Since $\mathscr{U}^\delta$ and $\mathscr{U}^0$ are functions on $\mathbb{R}$,

$$\|\mathscr{U}^\delta - \mathscr{U}^0\|_{H^1\to H^{-1}} \le C|\mathscr{U}^\delta - \mathscr{U}^0|_{H^{-1}}.$$

See, e.g., [Drouot 2018c, Lemma 2.1]. Recall that $\mathscr{U}^\delta$ is related to $\mathscr{U}$ via (6-7). The function $\mathscr{U}$ is periodic in the first variable and compactly supported in the second variable. We write a Fourier decomposition of $\mathscr{U}$:

$$\mathscr{U}(\tau, t) = \sum_{m\in\mathbb{Z}} b_m(t)e^{2i\pi m\tau}, \quad b_m(t) \stackrel{\text{def}}{=} \int_0^1 e^{-2i\pi m\tau'}\mathscr{U}(t, \tau')\, d\tau'.$$

Because of (6-7) and (6-8),

$$\mathscr{U}^\delta(t) - \mathscr{U}^0(t) = \sum_{m\neq 0} b_m(t)e^{2i\pi mt/\delta}.$$

In other words, $\mathscr{U}^\delta - \mathscr{U}^0$ has a highly oscillatory structure. The coefficients b_m are smooth functions of t. Their Sobolev norms decay rapidly since $\mathscr{U}$ depends smoothly on τ. We can then conclude as in the proof of [Drouot 2018a, Lemma 3.1]. □

The function $\mathscr{U}^0$ is an effective potential that arises as the homogenized limit of $\mathscr{U}^\delta$. It appears in the Dirac operator $\not{D}(\mu)$. Indeed, a computation shows that

$$\not{D}(\mu)^2 = v_F^2|k'|^2D_t^2 + \mu^2\cdot v_F^2|\ell|^2 + \vartheta_F^2\kappa^2 + \vartheta_\star\begin{bmatrix}0 & -v_\star k'\\ \overline{v_\star k'} & 0\end{bmatrix}(D_t\kappa).$$

Because of (6-5), we deduce that

$$\not{D}(\mu)^2 = \nu_F^2 |k'|^2 D_t^2 + \mu^2 \cdot \nu_F^2 |\ell|^2 + \vartheta_F^2 + \mathscr{U}^0. \tag{6-9}$$

We will apply this identity in the next section.

6C. ***A cyclicity argument.*** The next result is stated abstractly. It relies on the cyclicity principle.

Lemma 6.3. *Let A, B, C, D, E be bounded operators*:

$$\begin{gathered} A : H^1(\mathbb{R}, \mathbb{C}^2) \to L^2[\zeta], \quad B : L^2[\zeta] \to L^2(\mathbb{R}, \mathbb{C}^2), \\ C : L^2(\mathbb{R}, \mathbb{C}^2) \to L^2[\zeta], \quad D : L^2(\mathbb{R}, \mathbb{C}^2) \to H^1(\mathbb{R}, \mathbb{C}^2), \\ E : L^2(\mathbb{R}, \mathbb{C}^2) \to L^2(\mathbb{R}, \mathbb{C}^2). \end{gathered}$$

Assume that for some $M \geq 1$:

(a) *The operator norms of A, B, C, D, E are bounded by M.*

(b) *The operator* $\mathrm{Id} + DED : L^2(\mathbb{R}, \mathbb{C}^2) \to L^2(\mathbb{R}, \mathbb{C}^2)$ *is invertible and*

$$\|(\mathrm{Id} + DED)^{-1}\|_{L^2(\mathbb{R}, \mathbb{C}^2)} \leq M.$$

(c) *The following estimate holds*:

$$\epsilon \overset{\text{def}}{=} \|D(BC - E)D\|_{L^2(\mathbb{R}, \mathbb{C}^2)} \leq \frac{1}{2M}.$$

Then the operator $\mathrm{Id} + CD^2B : L^2[\zeta] \to L^2[\zeta]$ *is invertible*,

$$\begin{gathered} \|(\mathrm{Id} + CD^2B)^{-1}\|_{L^2[\zeta]} \leq 3M^5, \quad \text{and} \\ \|AD^2B \cdot (\mathrm{Id} + CD^2B)^{-1} - AD \cdot (\mathrm{Id} + DED)^{-1} \cdot DB\|_{L^2[\zeta]} \leq 2M^6 \epsilon. \end{gathered} \tag{6-10}$$

Proof. Below we use L^2 and H^1 to denote $L^2(\mathbb{R}, \mathbb{C}^2)$ and $H^1(\mathbb{R}, \mathbb{C}^2)$.

1. Recall that $\mathrm{Id} + CD^2B = \mathrm{Id} + CD \cdot DB : L^2[\zeta] \to L^2[\zeta]$ is invertible if and only if $\mathrm{Id} + DB \cdot CD : L^2 \to L^2$ is invertible. In this case, the inverses are related via

$$(\mathrm{Id} + CD^2B)^{-1} = \mathrm{Id} - CD(\mathrm{Id} + DB \cdot CD)^{-1} DB. \tag{6-11}$$

Because of (b), $\mathrm{Id} + DED$ is invertible and

$$\begin{aligned} \mathrm{Id} + DB \cdot CD &= \mathrm{Id} + DED + D(BC - E)D \\ &= (\mathrm{Id} + DED) \cdot (\mathrm{Id} + (\mathrm{Id} + DED)^{-1} \cdot D(BC - E)D). \end{aligned} \tag{6-12}$$

Because of both (b) and (c),

$$\|(\mathrm{Id} + DED)^{-1} \cdot D(BC - E)D\|_{L^2} \leq \tfrac{1}{2}.$$

This implies that $\mathrm{Id} + (\mathrm{Id} + DED)^{-1} \cdot D(BC - E)D$ is invertible by a Neumann series; the inverse has operator norm controlled by 2. Thanks to (6-12), $\mathrm{Id} + DBCD$ is invertible and the inverse has norm

controlled by $2M$. Hence $\mathrm{Id}+CD^2B$ is invertible. Thanks to (6-11) and (a),

$$\|(\mathrm{Id}+CD^2B)^{-1}\|_{L^2[\zeta]} \le 1 + M^2 \cdot 2M \cdot M^2 \le 3M^5.$$

This proves the first estimate of (6-10).

2. Observe that

$$(\mathrm{Id}+DBCD)^{-1} - (\mathrm{Id}+DED)^{-1} = (\mathrm{Id}+DED)^{-1} \cdot D(E-BC)D \cdot (\mathrm{Id}+DBCD)^{-1}.$$

Because of the bounds proved in step 1 and of (c),

$$\|(\mathrm{Id}+DBCD)^{-1} - (\mathrm{Id}+DED)^{-1}\|_{L^2} \le 2M^2\epsilon. \tag{6-13}$$

We write

$$\begin{aligned} AD^2B \cdot (\mathrm{Id}+CD^2B)^{-1} &= AD^2B \cdot (\mathrm{Id} - CD(\mathrm{Id}+DBCD)^{-1}DB) \\ &= AD \cdot DB - AD \cdot DBCD(\mathrm{Id}+DBCD)^{-1} \cdot DB \\ &= AD \cdot (\mathrm{Id} - DBCD(\mathrm{Id}+DBCD)^{-1}) \cdot DB \\ &= AD \cdot (\mathrm{Id}+DBCD)^{-1} \cdot DB. \end{aligned}$$

The operator norms of $AD : L^2 \to L^2[\zeta]$ and $DB : L^2[\zeta] \to H^1$ are each bounded by M^2 because of (a). We deduce from (6-13) that

$$\|AD^2B \cdot (\mathrm{Id}+CD^2B)^{-1} - AD \cdot (\mathrm{Id}+DED)^{-1} \cdot DB\|_{L^2[\zeta]} \le 2M^6\epsilon.$$

This proves the second estimate of (6-13), hence completes the proof of the lemma. □

We would like to apply Lemma 6.3 with the choices

$$\begin{aligned} &A \overset{\text{def}}{=} \delta^{1/2} \cdot \begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle}\Pi^* \cdot \mathcal{U}_\delta(\not{D}(\mu)+z), \quad B \overset{\text{def}}{=} \frac{1}{\delta^{1/2}} \cdot \mathcal{U}_\delta^{-1}\Pi e^{i\mu\delta\langle \ell, x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}, \\ &C \overset{\text{def}}{=} \delta^{1/2}\vartheta_\star\left(2(D_t\kappa)_\delta \cdot \begin{bmatrix}k' \cdot D_x\phi_1\\ k' \cdot D_x\phi_2\end{bmatrix}^\top + (\kappa_\delta^2 - 1)W\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top\right) \cdot e^{-i\mu\delta\langle \ell, x\rangle}\Pi^* \cdot \mathcal{U}_\delta\boldsymbol{\sigma_3}, \\ &D = (\not{D}_+(\mu)^2 - z^2)^{-1/2} = R_0(\mu, z)^{1/2}, \quad E = \mathscr{U}^0. \end{aligned} \tag{6-14}$$

These operators are manufactured so that

$$\mathcal{Q}_\delta(\mu, z) = \frac{1}{\delta}AD^2B, \quad \mathcal{K}_\delta(\mu, z) = CD^2B; \tag{6-15}$$

see the formula of Lemma 6.1. Recall that $\mathscr{U}^\delta, \mathscr{U}^0$ were defined in Lemma 6.2 and observe that $BC = \mathscr{U}^\delta \to E = \mathcal{U}^0$ (for the operator norm $H^1 \to H^{-1}$). This provides the favorable setting needed for the use of the cyclicity argument (Lemma 6.3).

The definition of D requires some precision. Let $\varphi(\omega) = (\omega^2 - z^2)^{-1/2}$, where the square root is holomorphic on $\mathbb{C} \setminus (-\infty, 0]$. If $|z| < \sqrt{\vartheta_F^2 + \mu^2 \cdot \nu_F^2|\ell|^2}$ and

$$\omega \in \Sigma_{L^2}(\not{D}_+(\mu)) = \mathbb{R} \setminus \left[-\sqrt{\vartheta_F^2 + \mu^2 \cdot \nu_F^2|\ell|^2}, \sqrt{\vartheta_F^2 + \mu^2 \cdot \nu_F^2|\ell|^2}\right].$$

then $\mathrm{Re}(\omega^2 - z^2) > 0$. Hence φ is well-defined on the spectrum of $\not{D}_+(\mu)$. This allows to define $D = \varphi(\not{D}_+(\mu))$ using the spectral theorem.

Lemma 6.4. *Fix $\epsilon_1 > 0$, $\mu_\sharp \in \mathbb{R}$. There exists $\delta_0 > 0$ such that if*

$$\begin{gathered} \delta \in (0, \delta_0), \quad \mu \in (-\mu_\sharp, \mu_\sharp), \\ z^2 \in \mathbb{D}(0, \vartheta_F^2 + \mu^2 \cdot v_F^2 |\ell|^2), \quad \mathrm{dist}(\Sigma_{L^2}(\not{D}(\mu)^2), z^2) \geq \epsilon_1^2 \end{gathered} \tag{6-16}$$

then $(\mathrm{Id} + DED)^{-1}$ *and* $(\mathrm{Id} + CD^2B)^{-1}$ *are invertible on* $L^2[\zeta]$. *Moreover*,

$$\begin{aligned} AD^2B \cdot (\mathrm{Id} + CD^2B)^{-1} &= AD \cdot (\mathrm{Id} + DED)^{-1} \cdot DB + \mathscr{O}_{L^2[\zeta]}(\delta), \\ (\mathrm{Id} + CD^2B)^{-1} &= \mathscr{O}_{L^2[\zeta]}(1). \end{aligned} \tag{6-17}$$

Proof. Below we use L^2 and H^1 to denote $L^2(\mathbb{R}, \mathbb{C}^2)$ and $H^1(\mathbb{R}, \mathbb{C}^2)$. The equation (6-17) is a consequence of Lemma 6.3, assuming that the assumptions (a), (b) and (c) hold with a constant M independent of δ, μ, z satisfying (6-16).

1. We first verify (a). We observe that the only singular dependence of A, B, C and E is in δ. It arises only in the operators $\delta^{1/2}\mathcal{U}_\delta$ and $\delta^{-1/2}\mathcal{U}_\delta^{-1}$, which are both isometries on L^2. In addition,

$$\mathrm{dist}(\Sigma_{L^2}(\not{D}(\mu)^2), z^2) \geq \epsilon_1^2 \quad \Longrightarrow \quad \mathrm{dist}(\Sigma_{L^2}(\not{D}_+(\mu)^2), z^2) \geq \epsilon_1^2.$$

Therefore D is controlled by ϵ_1^{-2}, and (a) holds independently of δ, μ, z satisfying (6-16).

2. From the definition (6-14) of D, we know D is invertible. Therefore we can write

$$\mathrm{Id} + DED = D(D^{-2} + E)D.$$

Moreover, thanks to (6-9),

$$D^{-2} + E = \not{D}(\mu)^2 - z^2. \tag{6-18}$$

When z satisfies the condition of (6-16), the operator $\not{D}(\mu)^2 - z^2$ is invertible. This comes with the bound

$$\|(\not{D}(\mu)^2 - z^2)^{-1}\|_{L^2} \leq \frac{1}{\epsilon_1^2}.$$

This is independent of δ: (b) holds.

3. The operator D maps L^2 to H^1 and H^{-1} to L^2, with uniformly bounded norm in μ, z satisfying (6-16). Therefore (c) holds — possibly after shrinking δ_0 — if

$$\|BC - E\|_{H^1 \to H^{-1}} = O(\delta). \tag{6-19}$$

We observe that $BC = \mathscr{U}^\delta$ and recall that $E = \mathscr{U}^0$. Therefore (6-19) reduces to the quantitative estimate (6-6) proved in Lemma 6.2.

4. Because of steps 1, 2 and 3, we can apply Lemma 6.3. It yields Lemma 6.4. □

According to this lemma, when (6-16) holds, $\mathrm{Id} + \mathcal{K}_\delta(\mu, z)$ is invertible. Hence

$$\mathcal{Q}_\delta(\mu, z) \cdot (\mathrm{Id} + \mathcal{K}_\delta(\mu, z))^{-1}$$

is well-defined. Thanks to (6-15),

$$\begin{aligned}\mathcal{Q}_\delta(\mu,z)\cdot(\mathrm{Id}+\mathcal{K}_\delta(\mu,z))^{-1} &= \frac{1}{\delta}AD\cdot(\mathrm{Id}+DED)^{-1}\cdot DB+\mathscr{O}_{L^2[\zeta]}(1)\\ &= \frac{1}{\delta}AD\cdot D^{-1}(D^{-2}+E)^{-1}D^{-1}\cdot DB+\mathscr{O}_{L^2[\zeta]}(1)\\ &= \frac{1}{\delta}A\cdot(D^{-2}+E)^{-1}\cdot B+\mathscr{O}_{L^2[\zeta]}(1).\end{aligned}$$

We now plug in the formula (6-14) for A, B, C, D, E, and we use the relation (6-18). This yields

$$\begin{aligned}&\mathcal{Q}_\delta(\mu,z)\cdot(\mathrm{Id}+\mathcal{K}_\delta(\mu,z))^{-1}\\ &\quad=\frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\mathcal{U}_\delta\cdot(\not{D}(\mu)+z)\cdot(\not{D}(\mu)^2-z^2)^{-1}\cdot\mathcal{U}_\delta^{-1}\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}+\mathscr{O}_{L^2[\zeta]}(1)\\ &\quad=\frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle\ell,x\rangle}\cdot\Pi^*\mathcal{U}_\delta\cdot(\not{D}(\mu)-z)^{-1}\cdot\mathcal{U}_\delta^{-1}\Pi\cdot e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}+\mathscr{O}_{L^2[\zeta]}(1).\end{aligned}\tag{6-20}$$

We are now ready to prove Theorem 3.2.

Proof of Theorem 3.2. 1. Fix $\epsilon>0$ and $\mu_\sharp>0$. Fix $z\in\mathbb{C}$ satisfying

$$z\in\mathbb{D}\Big(0,\sqrt{\vartheta_F^2+\mu^2\cdot\nu_F^2|\ell|^2}-\frac{\epsilon}{3}\Big),\quad \mathrm{dist}(\Sigma_{L^2}(\not{D}(\mu)^2),z^2)\geq\frac{\epsilon^2}{9}.\tag{6-21}$$

Note that this does not quite correspond to the assumptions of Theorem 3.2. Instead it is a stronger form of the assumptions of Lemma 6.4 with $\epsilon_1=\epsilon/3$. The equation (6-21) implies that $\mathrm{Id}+\mathcal{K}_\delta(\mu,z)$ is invertible. Apply Lemma 6.1 with

$$\theta=1-\frac{\epsilon}{3\sqrt{\vartheta_F^2+\mu^2\cdot\nu_F^2|\ell|^2}}.$$

It implies that

$$\mathrm{Id}+\mathscr{K}_\delta(\zeta,\lambda)=\mathrm{Id}+\mathcal{K}_\delta(\mu,z)+\mathscr{O}_{L^2[\zeta]}(\delta^{2/3}).$$

Hence — after possibly shrinking δ_0 — the operator $\mathrm{Id}+\mathscr{K}_\delta(\zeta,\lambda)$ is invertible. The inverses of $\mathrm{Id}+\mathscr{K}_\delta(\zeta,\lambda)$ and $\mathrm{Id}+\mathcal{K}_\delta(\mu,z)$ are related via

$$(\mathrm{Id}+\mathscr{K}_\delta(\zeta,\lambda))^{-1}=(\mathrm{Id}+\mathcal{K}_\delta(\mu,z))^{-1}+\mathscr{O}_{L^2[\zeta]}(\delta^{2/3}),$$

because $(\mathrm{Id}+\mathcal{K}_\delta(\mu,z))^{-1}=(\mathrm{Id}+CD^2B)^{-1}$ is uniformly bounded by Lemma 6.4. It follows that under (6-21), $\mathscr{P}_\delta[\zeta]-\lambda$ is invertible and

$$(\mathscr{P}_\delta[\zeta]-\lambda)^{-1}=\mathscr{Q}_\delta(\zeta,\lambda)\cdot(\mathrm{Id}+\mathscr{K}_\delta(\zeta,\lambda))^{-1}.$$

2. Observe that $\mathcal{Q}_\delta(\mu,z)=\mathscr{O}_{L^2[\zeta]}(\delta^{-1})$: this comes from the relation between $\mathscr{Q}_\delta(\zeta,\lambda)$ and $\mathcal{Q}_\delta(\mu,z)$ provided by Lemma 6.1. We deduce that $\mathscr{P}_\delta[\zeta]-\lambda$ is invertible and

$$(\mathscr{P}_\delta[\zeta]-\lambda)^{-1}=\mathcal{Q}_\delta(\mu,z)\cdot(\mathrm{Id}+\mathcal{K}_\delta(\mu,z))^{-1}+\mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}).$$

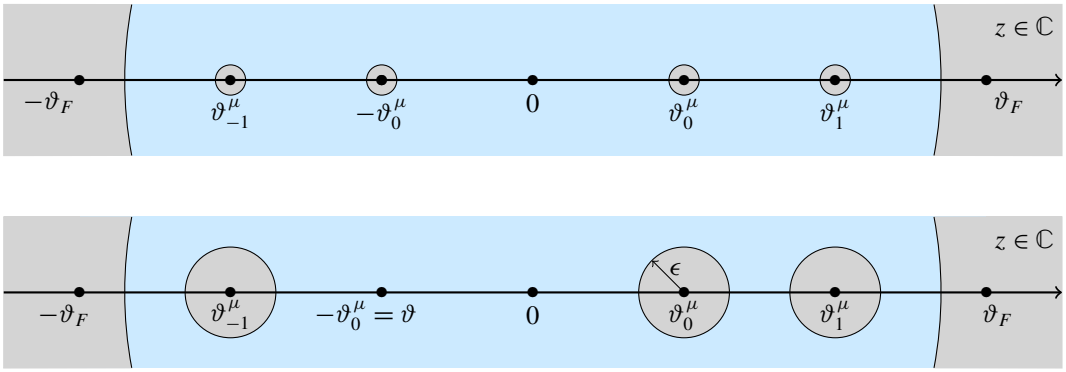

Figure 11. The top blue area represents the domain of validity of (6-22) provided by steps 1 and 2. The bottom blue area represents the domain of validity of (6-22) as specified by Theorem 3.2. In step 3 we prove that (6-22) holds near $\vartheta = -\vartheta_0^{\mu}$, at the price of increasing $\epsilon/3$ to ϵ.

Thanks to (6-20), this simplifies to

$$(\mathscr{P}_\delta[\zeta]-\lambda)^{-1} = \frac{1}{\delta}\cdot \begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^{\top} e^{-i\mu\delta\langle \ell, x\rangle}\Pi^* \mathcal{U}_\delta \cdot (\not{D}(\mu)-z)^{-1}\cdot \mathcal{U}_\delta^{-1}\Pi e^{i\mu\delta\langle \ell, x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}). \quad (6\text{-}22)$$

See Figure 11.

3. The estimate (6-22) is valid as long as z satisfies (6-21). There is a subtlety here: (6-21) does not quite correspond to the assumption of Theorem 3.2. To conclude the proof, we must show that (6-21) is unnecessarily strong. In other words, we assume in these final steps that

$$z \in \mathbb{D}\big(0, \sqrt{\vartheta_F^2 + \mu^2\cdot \nu_F^2|\ell|^2} - \epsilon\big), \quad \operatorname{dist}(\Sigma_{L^2}(\not{D}(\mu)), z) \geq \epsilon, \qquad \operatorname{dist}(\Sigma_{L^2}(\not{D}(\mu)^2), z^2) < \frac{\epsilon^2}{9}.$$

The third condition implies

$$\operatorname{dist}(\Sigma_{L^2}(\not{D}(\mu)), z) < \frac{\epsilon}{3} \quad \text{or} \quad \operatorname{dist}(\Sigma_{L^2}(-\not{D}(\mu)), z) < \frac{\epsilon}{3}.$$

From the second condition, we deduce that $\operatorname{dist}(\Sigma_{L^2}(-\not{D}(\mu)), z) < \epsilon/3$. The spectra of $\not{D}(\mu)$ and $-\not{D}(\mu)$ differ by at most one eigenvalue:

$$\Sigma_{L^2}(-\not{D}(\mu)) \setminus \Sigma_{L^2}(\not{D}(\mu)) \subset \{\vartheta\}, \quad \vartheta \overset{\text{def}}{=} -\mu\cdot \nu_F|\ell|\cdot \operatorname{sgn}(\vartheta_\star), \quad (6\text{-}23)$$

see Lemma 3.1. Hence, z must belong to $\mathbb{D}(\vartheta, \epsilon/3)$.

4. Because of step 3, the proof of Theorem 3.2 is complete if we can show that (6-22) holds when

$$z \in \mathbb{D}\big(0, \sqrt{\vartheta_F^2 + \mu^2\cdot \nu_F^2|\ell|^2} - \epsilon\big), \quad \operatorname{dist}(\Sigma_{L^2}(\not{D}(\mu)), z) \geq \epsilon, \qquad z \in \mathbb{D}\Big(\vartheta, \frac{\epsilon}{3}\Big).$$

Fix $s \in \partial\mathbb{D}(\vartheta, \epsilon/3)$. Then, $|z-s| < 2\epsilon/3$. This implies that

$$s \in \mathbb{D}\Big(0, \sqrt{\vartheta_F^2 + \mu^2\cdot \nu_F^2|\ell|^2} - \frac{\epsilon}{3}\Big), \quad \operatorname{dist}(\Sigma_{L^2}(\not{D}(\mu)), s) \geq \frac{\epsilon}{3}, \quad |\vartheta - s| = \frac{\epsilon}{3}.$$

Because of (6-23), s satisfies

$$\operatorname{dist}(\Sigma_{L^2}(\not{D}(\mu)), s) \geq \frac{\epsilon}{3}, \quad \operatorname{dist}(\Sigma_{L^2}(-\not{D}(\mu)), s) = \frac{\epsilon}{3} \quad \Longrightarrow \quad \operatorname{dist}(\Sigma_{L^2}(\not{D}(\mu)^2), s) \geq \frac{\epsilon^2}{9}.$$

In particular, s satisfies (6-21).

Therefore steps 1 and 2 apply to $s \in \partial\mathbb{D}(\vartheta, \epsilon/3)$. They yield

$$(\mathscr{P}_\delta[\zeta] - E_\star - \delta s)^{-1} = \frac{1}{\delta} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle} \Pi^* \mathcal{U}_\delta \cdot (\not{D}(\mu) - s)^{-1} \cdot \mathcal{U}_\delta^{-1} \Pi e^{i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}} + \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}). \quad (6\text{-}24)$$

Note that $(\not{D}(\mu) - s)^{-1}$ has no poles in the disk $\mathbb{D}(\vartheta, \epsilon/3)$: otherwise z could not be at distance at least ϵ from $\Sigma_{L^2}(\not{D}(\mu))$. Thus, integrating (6-24) over the circle $\partial\mathbb{D}(\vartheta, \epsilon/3)$,

$$\frac{1}{2\pi i} \oint_{\partial\mathbb{D}(\vartheta, \epsilon/3)} (\mathscr{P}_\delta[\zeta] - E_\star - \delta s)^{-1}\, ds = \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}). \quad (6\text{-}25)$$

We substitute $\lambda = E_\star + \delta s$ in (6-25) to get

$$\frac{1}{2\pi i} \oint_{\partial\mathbb{D}(E_\star + \delta\vartheta, \epsilon\delta/3)} (\mathscr{P}_\delta[\zeta] - \lambda)^{-1}\, d\lambda = \mathscr{O}_{L^2[\zeta]}(\delta^{2/3}). \quad (6\text{-}26)$$

Equation (6-26) implies that $(\mathscr{P}_\delta[\zeta] - \lambda)^{-1}$ cannot have a pole in $\mathbb{D}(E_\star + \vartheta\delta, \epsilon\delta/3)$. Indeed, since $\mathscr{P}_\delta[\zeta]$ is selfadjoint, the nonzero residues of its resolvent are nonzero projectors, and hence have $L^2[\zeta]$-operator norm at least equal to 1.

We deduce that $s \mapsto (\mathscr{P}_\delta[\zeta] - E_\star - \delta s)^{-1}$ is holomorphic in the disk $\mathbb{D}(\vartheta, \epsilon/3)$, and so is the leading term in (6-24). Their difference is bounded by $\mathscr{O}_{L^2[\zeta]}(\delta^{-1/3})$ on the boundary of the disk. By the maximum principle, this difference is $\mathscr{O}_{L^2[\zeta]}(\delta^{-1/3})$ also inside the disk. This shows that (6-24) holds when s is in the disk $\mathbb{D}(\vartheta, \epsilon/3)$. Equivalently (6-22) holds when $z \in \mathbb{D}(\vartheta, \epsilon/3)$. This completes the proof of Theorem 3.2. □

7. A topological perspective

7A. ***The role of $\vartheta_\star^A$ and $\vartheta_\star^B$ in the spectral flow.*** Assume that P_0 has Dirac points $(\xi_\star^A, E_\star)$ and $(\xi_\star^B, E_\star)$ — where $\xi_\star^A$ and $\xi_\star^B$ were defined in (1-4). Following Definition 1.2, these Dirac points are associated to Dirac eigenbases (ϕ_1^A, ϕ_2^A) and (ϕ_1^B, ϕ_2^B):

$$\phi_1^J \in L^2_{\xi_\star^J, \tau},\ \phi_2^J \in L^2_{\xi_\star^J, \bar\tau}, \quad J = A, B, \qquad \text{and} \qquad \vartheta_\star^J = \langle \phi_1^J, W\phi_1^J \rangle_{L^2_{\xi_\star^J}}. \quad (7\text{-}1)$$

We recall that $\vartheta_\star^J$ does not depend on the choice of Dirac eigenbasis satisfying (7-1). The next result is a key identity — see also [Lee-Thorp et al. 2019, §7.1].

Lemma 7.1. *The identity $\vartheta_\star^A + \vartheta_\star^B = 0$ holds.*

Proof. 1. We claim that $\mathcal{I}\phi_1^A \in L^2_{\xi_\star^B, \tau}$. Thanks to (1-4),

$$-\xi_\star^A = -\tfrac{2\pi}{3}(2k_1 + k_2) = \tfrac{2\pi}{3}(k_1 + 2k_2) = \xi_\star^B \bmod 2\pi\Lambda^*.$$

Because $\phi_1^A \in L^2_{\xi_\star^A,\tau}$,

$$(\mathcal{I}\phi_1^A)(x+w) = \phi_1^A(-x-w) = e^{-i\langle \xi_\star^A, w\rangle}(\mathcal{I}\phi_1^A)(x) = e^{i\langle \xi_\star^B, w\rangle}(\mathcal{I}\phi_1^A)(x),$$
$$(\mathcal{R}\mathcal{I}\phi_1^A)(x) = \phi_1^A(-Rx) = \tau\phi_1^A(-x) = \tau(\mathcal{I}\phi_1^A)(x).$$

It follows that $\mathcal{I}\phi_1^A \in L^2_{\xi_\star^B,\tau}$ — as claimed. The same calculation shows that $\mathcal{I}\phi_2^A \in L^2_{\xi_\star^B,\bar\tau}$.

The operator P_0 is $\mathcal{I}$-invariant. Thus, $\mathcal{I}\phi_1^A$ and $\mathcal{I}\phi_2^A$ form an orthonormal basis of $\ker_{L^2_{\xi_\star}}(P_0(\xi_\star^B) - E_\star)$, and $(\mathcal{I}\phi_1^A, \mathcal{I}\phi_2^A)$ is a Dirac eigenbasis for $(\xi_\star^B, E_\star)$.

2. Because W is odd and $\vartheta_\star^B$ does not depend on the choice of Dirac eigenbasis,

$$\vartheta_\star^B = \langle \mathcal{I}\phi_1^A, W\mathcal{I}\phi_1^A\rangle_{L^2_{\xi_\star^B}} = -\langle \phi_1^A, W\phi_1^A\rangle_{L^2_{\xi_\star^A}} = -\vartheta_\star^A. \qquad \square$$

Recall the assumption (H4): for every $\zeta \notin \{\frac{2\pi}{3}, \frac{4\pi}{3}\} \bmod 2\pi\mathbb{Z}$ and $\tau, \tau' \in \mathbb{R}$,

$$\lambda_{0,j_\star}(\zeta k + \tau k') < \lambda_{0,j_\star+1}(\zeta k + \tau' k').$$

Lemma 7.2. *Assume* (H1)–(H4) *hold for both $\xi_\star^A$ and $\xi_\star^B$. There exists a function $E \in C^0(\mathbb{R}/(2\pi\mathbb{Z}), \mathbb{R})$ with $E(\zeta_\star^A) = E(\zeta_\star^B) = E_\star$ and such that*

$$\forall \zeta \in \mathbb{R}, \quad E(\zeta) \notin \Sigma_{L^2[\zeta],\mathrm{ess}}(\mathscr{P}_\delta[\zeta]).$$

Moreover, there exist $\mu_\flat > 0$ and $\delta_0 > 0$ such that if

$$\delta \in (0, \delta_0), \quad \zeta \in [0, 2\pi], \quad \big|\zeta - \tfrac{2\pi}{3}\big| \geq \mu_\flat\delta, \quad \big|\zeta - \tfrac{4\pi}{3}\big| \geq \mu_\flat\delta,$$

then the operator $\mathscr{P}_\delta[\zeta]$ has no spectrum in $[E(\zeta) - \delta, E(\zeta) + \delta]$.

Proof. 1. Set $r(\zeta) = \mathrm{dist}\big(\zeta, \{\frac{2\pi}{3}, \frac{4\pi}{3}\}\big)$. We first show that there exists $a > 0$ such that for $\zeta \in [0, 2\pi]$,

$$\inf_{\tau,\tau'\in\mathbb{R}} (\lambda_{0,j_\star+1}(\zeta k + \tau' k') - \lambda_{0,j_\star}(\zeta k + \tau k')) \geq 4a \cdot r(\zeta). \tag{7-2}$$

Otherwise, we can find $\zeta_n \in [0, 2\pi]$, $\tau_n, \tau_n' \in \mathbb{R}$, such that

$$\lambda_{0,j_\star+1}(\zeta_n k + \tau_n' k') - \lambda_{0,j_\star}(\zeta_n k + \tau_n k') \leq \frac{r(\zeta_n)}{n} = \frac{1}{n}\cdot \mathrm{dist}\big(\zeta_n, \{\tfrac{2\pi}{3}, \tfrac{4\pi}{3}\}\big). \tag{7-3}$$

Using periodicity of the dispersion curves, we can assume that τ_n, τ_n' both live in $[0, 2\pi]$. In particular we can pass to converging subsequences: there exist ζ_∞, τ_∞ and τ_∞' with

$$\lambda_{0,j_\star}(\zeta_\infty k + \tau_\infty k') = \lambda_{0,j_\star+1}(\zeta_\infty k + \tau_\infty' k'). \tag{7-4}$$

Because of (H4), $\zeta_\infty \in \{\frac{2\pi}{3}, \frac{4\pi}{3}\} = \{\zeta_\star^A, \zeta_\star^B\} \bmod 2\pi$. In the proof of Lemma 4.1, we showed that

$$\zeta_\star \in \{\tfrac{2\pi}{3}, \tfrac{4\pi}{3}\}, \quad \tau, \tau' \in \mathbb{R} \quad \Longrightarrow \quad \lambda_{0,j_\star}(\zeta_\star k + \tau k') \leq E_\star, \quad \lambda_{0,j_\star+1}(\zeta_\infty k + \tau' k') \geq E_\star.$$

Thanks to (7-4), we deduce that $\lambda_{0,j_\star+1}(\zeta_\infty k + \tau_\infty' k') = E_\star = \lambda_{0,j_\star}(\zeta_\infty k + \tau_\infty k')$. The no-fold condition implies that $\zeta_\infty k + \tau_\infty k' = \zeta_\infty k + \tau_\infty' k' = \xi_\star$, where $\xi_\star \in \{\xi_\star^A, \xi_\star^B\}$ is a Dirac-point momentum. In particular,

$\zeta_n k + \tau_n' k'$ and $\zeta_n k + \tau_n k'$ both converge to $\xi_\star$. We deduce that for n sufficiently large,

$$\lambda_{0,j_\star+1}(\zeta_n k + \tau_n' k') - \lambda_{0,j_\star}(\zeta_n k + \tau_n k') \geq v_F |\zeta_n k + \tau_n' k' - \xi_\star| \geq v_F |k'| \cdot r(\zeta_n),$$

because $\langle \xi_\star, v \rangle \in \left\{ \frac{2\pi}{3}, \frac{4\pi}{3} \right\}$. This contradicts (7-3). We deduce that (7-2) holds for some $a > 0$. Without loss of generalities, we assume below that $a < v_F|\ell|$.

2. Define

$$E(\zeta) \stackrel{\text{def}}{=} 2a \cdot r(\zeta) + \sup_{\tau \in \mathbb{R}} \lambda_{0,j_\star}(\zeta k + \tau k').$$

This is a continuous, 2π-periodic function. Observe that for every $\xi \in \zeta k + \mathbb{R}k'$

$$\lambda_{0,j_\star}(\xi) \leq E(\zeta) - 2a \cdot r(\zeta) \leq E(\zeta) + 2a \cdot r(\zeta) \leq \lambda_{0,j_\star+1}(\xi). \tag{7-5}$$

Assume that $a \cdot r(\zeta) \geq \delta$ and that $\lambda \in [E(\zeta) - \delta, E(\zeta) + \delta]$. Since the dispersion surfaces are labeled in increasing order, we deduce that

$$\xi \in \zeta k + \mathbb{R}k' \quad \Longrightarrow \quad \operatorname{dist}(\Sigma_{L^2_\xi}(P_0(\xi)), \lambda) \geq a \cdot r(\zeta).$$

The reconstruction formula (5-2) and the spectral theorem yield

$$a \cdot r(\zeta) \geq \delta, \ \ \lambda \in [E(\zeta) - \delta, E(\zeta) + \delta] \quad \Longrightarrow \quad \|(P_0[\zeta] - \lambda)^{-1}\|_{L^2[\zeta]} \leq \frac{1}{a \cdot r(\zeta)}. \tag{7-6}$$

3. We now observe that under the assumptions of (7-6),

$$\mathscr{P}_\delta[\zeta] - \lambda = (P_0[\zeta] - \lambda) \cdot (\mathrm{Id} + \delta \cdot (P_0[\zeta] - \lambda)^{-1} \cdot \kappa_\delta W). \tag{7-7}$$

Because of (7-6) and since κ, W are in L^∞, there exist $\delta_0 > 0$ and $\mu_\flat > 0$ with

$$\delta \in (0, \delta_0), \ \ \zeta \in [0, 2\pi], \ \ r(\zeta) \geq \mu_\flat \delta \quad \Longrightarrow \quad \|\delta \cdot (P_0[\zeta] - \lambda)^{-1} \cdot \kappa_\delta W\|_{L^2[\zeta]} \leq \tfrac{1}{2}.$$

In particular, the second factor in the right-hand side of (7-7) is invertible via a Neumann series. We deduce that $\mathscr{P}_\delta[\zeta] - \lambda$ is invertible. This implies that $\mathscr{P}_\delta[\zeta]$ has no spectrum in $[E(\zeta) - \delta, E(\zeta) + \delta]$, as long as $r(\zeta) \geq \mu_\flat \delta$.

4. It remains to show that $E(\zeta)$ is not in the essential spectrum of $\mathscr{P}_\delta[\zeta]$, independently of ζ. Because of step 3, this holds for every ζ such that $r(\zeta) \geq \mu_\flat \delta$. Fix ζ such that $r(\zeta) < \mu_\flat \delta$. Let $\xi_\star$ be a Dirac point closest to $\zeta k + \mathbb{R}k'$: the distance between $\xi_\star$ and the line $\zeta k + \mathbb{R}k'$ is $r(\zeta)|\ell|$. Because of (7-5),

$$\lambda_{0,j_\star}(\zeta k + \tau k') + 2a \cdot r(\zeta) \leq E(\zeta) \leq \lambda_{0,j_\star+1}(\zeta k + \tau k') - 2a \cdot r(\zeta).$$

Since $\xi_\star$ is a Dirac point, we get

$$E_\star - (v_F|\ell| - 2a) \cdot r(\zeta) + O(r(\zeta)^2) \leq E(\zeta) \leq E_\star + (v_F|\ell| + 2a) \cdot r(\zeta) + O(r(\zeta)^2).$$

Hence, for δ sufficiently small,

$$E(\zeta) \in [E_\star - (v_F|\ell| - a) \cdot r(\zeta), E_\star + (v_F|\ell| - a) \cdot r(\zeta)].$$

Fix $\theta \in (0,1)$ such that $v_F|\ell| - a = \theta v_F|\ell|$; θ exists because $a \in (0, v_F|\ell|)$. Then

$$E(\zeta) \in \mathbb{D}\big(E_\star, \theta\sqrt{\vartheta_F^2\delta^2 + r(\zeta)^2 \cdot v_F^2|\ell|^2}\big).$$

Apply Theorem 5.1 with $\mu_\sharp > \mu_\flat$: for δ sufficiently small and $|\zeta - \zeta_\star| < \mu_\sharp\delta$, we have $E(\zeta) \notin \Sigma_{L^2[\zeta],\mathrm{ess}}(P_{\pm\delta}(\zeta))$. This implies that $E(\zeta)$ is not in the essential spectrum of $\mathscr{P}_\delta[\zeta]$ as long as $r(\zeta) < \mu_\flat\delta$, which concludes the proof. □

Lemma 7.2 allows us to define the spectral flow of the family $\zeta \mapsto \mathscr{P}_\delta[\zeta]$ as ζ runs from 0 to 2π: it is the signed number of eigenvalues of $\mathscr{P}_\delta[\zeta]$ that cross the curve $\zeta \mapsto E(\zeta)$ (with downwards crossings counted positively). Because $\mathscr{P}_\delta[\zeta]$ depends periodically on ζ, the spectral flow of $\mathscr{P}_\delta$ is a topological invariant. We refer to [Waterstraat 2017, §4] for an introduction to spectral flow. We are now ready to prove Corollary 1.6.

Proof of Corollary 1.6. We split $[0, 2\pi]$ in three parts: $[0, 2\pi] = I_A \cup I_B \cup I_0$ with

$$I_J \overset{\text{def}}{=} [\zeta_\star^J - \mu_\flat\delta, \zeta_\star^J + \mu_\flat\delta], \quad J = A, B, \qquad I_0 \overset{\text{def}}{=} [0, 2\pi] \setminus (I_A \cup I_B),$$

where we identified $\zeta_\star^J$ with their reduction modulo $2\pi\mathbb{Z}$. The spectral flow of $\zeta \in I_0 \mapsto \mathscr{P}_\delta[\zeta]$ through $E_\star$ vanishes because of Lemma 7.2.

In order to compute the spectral flow of $\zeta \in I_J \mapsto \mathscr{P}_\delta[\zeta]$ through $E_\star$, we fix $\mu_\sharp > \mu_\flat$, $\vartheta_\sharp > \vartheta_N$ and we apply Corollary 3.3. This result allows us to precisely count the number $N_\pm^J$ of eigenvalues of $\mathscr{P}_\delta[\zeta_\star^J \pm \mu_\flat\delta]$ in the set

$$\mathcal{E} \overset{\text{def}}{=} \big[E_\star - \delta\sqrt{\vartheta_\sharp^2 + \mu_\flat^2 \cdot v_F^2|\ell|^2}, E_\star\big]$$

in terms of the number of eigenvalues $2N+1$ of the Dirac operator $\not{D}(\mu)$. Thanks to Lemma 3.1, we find

$$N_-^J = N+1, \quad N_+^J = N \quad \text{if } \vartheta_\star^J > 0, \qquad N_-^J = N, \quad N_+^J = N+1 \quad \text{if } \vartheta_\star^J < 0.$$

In particular, the spectral flow of $\zeta \in I_J \mapsto \mathscr{P}_\delta[\zeta]$ through $E_\star$ is $N_+^J - N_-^J = -\operatorname{sgn}(\vartheta_\star^J)$ — see, e.g., [Waterstraat 2017, §4.1]. Since $\vartheta_\star^A$ and $\vartheta_\star^B$ have opposite sign, the spectral flow of the whole family $\zeta \in [0, 2\pi] \mapsto \mathscr{P}_\delta[\zeta]$ vanishes. □

7B. ***Magnetic perturbations of honeycomb Schrödinger operators.*** Let V be a honeycomb potential and $\mathbb{A} \in C^\infty(\mathbb{R}^2, \mathbb{R}^2)$ be Λ-periodic, odd and real-valued. Set

$$\mathbb{P}_\delta \overset{\text{def}}{=} (D_x + \delta \cdot \kappa_\delta \cdot \mathbb{A})^2 + V.$$

This operator is a nonlocal perturbation of $P_0 = -\Delta + V$, where $\delta \cdot \kappa_\delta \cdot \mathbb{A}$ plays the role of a perturbing magnetic potential. We introduce similarly to $\mathscr{P}_\delta[\zeta]$ the operator $\mathbb{P}_\delta[\zeta]$ formally equal to $\mathbb{P}_\delta$ but acting on $L^2[\zeta]$. We observe that

$$\begin{aligned}\mathbb{P}_\delta[\zeta] &= -\Delta + V + \delta \cdot \kappa_\delta \cdot (\mathbb{A}D_x + D_x\mathbb{A}) + \delta^2((k' \cdot D_x\kappa)_\delta + \kappa_\delta^2|\mathbb{A}|^2) \\ &= P_0 + \delta \cdot \kappa_\delta \cdot \mathbb{W} + \mathscr{O}_{L^2[\zeta]}(\delta^2) \quad \text{where } \mathbb{W} \overset{\text{def}}{=} \mathbb{A} \cdot D_x + D_x \cdot \mathbb{A}. \end{aligned} \tag{7-8}$$

We first state a simple analog of Lemma 2.3:

Lemma 7.3. *Let* $(\xi_\star, E_\star)$ *be a Dirac point of* P_0 *with Dirac eigenbasis* (ϕ_1, ϕ_2) *— see Definition 1.2. Then* $\langle \phi_1, \mathbb{W}\phi_2\rangle_{L^2_{\xi_\star}} = \langle \phi_2, \mathbb{W}\phi_1\rangle_{L^2_{\xi_\star}} = 0$. *Furthermore,*

$$\theta_\star \stackrel{\text{def}}{=} \langle \phi_1, \mathbb{W}\phi_1\rangle_{L^2_{\xi_\star}} = -\langle \phi_2, \mathbb{W}\phi_2\rangle_{L^2_{\xi_\star}}.$$

See Appendix A.1 or [Lee-Thorp et al. 2019, Proposition 5.1] for the proof. Below we state Corollary 7.4, which is the analog of Corollary 3.3 for the magnetic operator $\mathbb{P}_\delta[\zeta]$. We assume:

(H3′) The nondegeneracy condition $\theta_\star \neq 0$ holds.

When (H3′) holds and δ is small enough, the operator $\mathbb{P}_\delta[\zeta]$ has an essential spectral gap centered at $E_\star$ of width $\sim \delta$. Indeed (H3′) implies that $P_0 + \delta\cdot\kappa_\delta\cdot\mathbb{W}$ admits such a gap — as for $\mathscr{P}_\delta$ when (H3) is satisfied. This gap can only be moved by $O(\delta^2)$ under the perturbation $\mathscr{O}_{L^2[\zeta]}(\delta^2)$ of (7-8). We introduce the operator

$$\mathbb{D}(\mu) \stackrel{\text{def}}{=} \begin{bmatrix} 0 & \nu_\star k' \\ \overline{\nu_\star k'} & 0 \end{bmatrix} D_t + \mu \begin{bmatrix} 0 & \nu_\star \ell \\ \overline{\nu_\star \ell} & 0 \end{bmatrix} + \theta_\star \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \kappa.$$

We denote by $\{\theta_j^\mu\}_{j=-n}^n$ its eigenvalues. They are all simple — see Lemma 3.1 — and lie in $(-\theta_F, \theta_F)$, where $\theta_F = |\theta_\star|$.

Corollary 7.4. *Assume that* (H1), (H2) *and* (H3′) *hold and fix* $\theta_\sharp \in (\theta_N, \theta_F)$ *and* $\mu_\sharp > 0$. *There exists* $\delta_0 > 0$ *such that for*

$$\delta \in (0, \delta_0), \quad \mu \in (-\mu_\sharp, \mu_\sharp), \quad \zeta = \zeta_\star + \delta\mu,$$

the operator $\mathbb{P}_\delta[\zeta]$ *has exactly* $2n+1$ *eigenvalues* $\{\lambda^\zeta_{\delta,j}\}_{j\in[-n,n]}$ *in*

$$\big[E_\star - \delta\sqrt{\theta_\sharp^2 + \mu^2\cdot v_F^2|\ell|^2},\ E_\star + \delta\sqrt{\theta_\sharp^2 + \mu^2\cdot v_F^2|\ell|^2}\,\big].$$

These eigenvalues are simple. Furthermore, for each $j \in [-N, N]$, *the eigenpairs* $(\lambda^\zeta_{\delta,j}, v^\zeta_{\delta,j})$ *admit full expansions in powers of* δ:

$$\lambda^\zeta_{\delta,j} = E_\star + \theta_j^\mu\cdot\delta + b_2^\mu\cdot\delta^2 + \cdots + b_M^\mu\cdot\delta^M + O(\delta^{M+1}),$$
$$v^\zeta_{\delta,j}(x) = e^{i(\zeta-\zeta_\star)\langle \ell, x\rangle}\big(g_0^\mu(x, \delta\langle k', x\rangle) + \cdots + \delta^M\cdot g_M^\mu(x, \delta\langle k', x\rangle)\big) + o_{H^k}(\delta^M).$$

In the above expansions:

- M *and* k *are any integers*; H^k *is the* k*-th order Sobolev space.*
- θ_j^μ *is the* j*-th eigenvalue of* $\mathbb{D}(\mu)$.
- *The terms* $b_m^\mu \in \mathbb{R}$, $g_m^\mu \in X$ *are recursively constructed via multiscale analysis.*
- *The leading-order term* g_0^μ *satisfies*

$$g_0^\mu(x,t) = \beta_1^\mu(t)\phi_1(x) + \beta_2^\mu(t)\phi_2(x), \quad (\mathbb{D}(\mu) - \theta_j^\mu)\begin{bmatrix}\beta_1^\mu \\ \beta_2^\mu\end{bmatrix} = 0.$$

The proof is identical to that of Theorem 3.2 and Corollary 3.3; we do not reproduce it here. Let $\theta_\star^J$ be the coefficient $\theta_\star$ associated to the Dirac point $(\xi_\star^J, E_\star)$. The main difference between $\mathscr{P}_\delta[\zeta]$ and $\mathbb{P}_\delta[\zeta]$ lies in the next identity — see also [Lee-Thorp et al. 2019, §7.1].

Lemma 7.5. *The identity $\theta_\star^A = \theta_\star^B$ holds.*

Proof. Because of step 1 in the proof of Lemma 7.1, $(\mathcal{I}\phi_1^A, \mathcal{I}\phi_2^A)$ is a Dirac eigenbasis for $(\xi_\star^B, E_\star)$. Since $\theta_\star^B$ does not depend on the choice of Dirac eigenbasis and $\mathbb{W}$ commutes with $\mathcal{I}$,

$$\theta_\star^B = \langle \mathcal{I}\phi_1^A, \mathbb{W}\mathcal{I}\phi_1^A\rangle_{L^2_{\xi_\star^B}} = \langle \phi_1^A, \mathbb{W}\phi_1^A\rangle_{L^2_{\xi_\star^A}} = \theta_\star^A. \qquad \square$$

Corollary 1.7 has the same proof as Corollary 1.6. We find that the spectral flow of $\mathbb{P}_\delta$ in the $j_\star$-th gap as ζ runs from 0 to 2π is equal to

$$-\operatorname{sgn}(\theta_\star^A) - \operatorname{sgn}(\theta_\star^B) = -2 \cdot \operatorname{sgn}(\theta_\star).$$

Appendix

A.1. ***Proofs of some identities.*** We prove the identities relating the Dirac eigenbasis and W. Similar proofs arise in [Fefferman et al. 2016b; 2017; Lee-Thorp et al. 2019].

Proof of Lemma 2.2. Below we use $\langle \cdot, \cdot \rangle$ instead of $\langle \cdot, \cdot \rangle_{L^2_{\xi_\star}}$ to simplify notations.

1. We first analyze the (2-vector) $\langle \phi_1, D_x\phi_1\rangle$. We observe that $\langle \phi_1, D_x\phi_1\rangle \in \mathbb{R}^2$ because D_x is selfadjoint. Since $\phi_1 \in L^2_{\xi_\star,\tau}$,

$$\langle \phi_1, D_x\phi_1\rangle = \langle \mathcal{R}\phi_1, \mathcal{R}D_x\phi_1\rangle = \langle \tau\phi_1, \mathcal{R}D_x\mathcal{R}^{-1}\cdot\tau\phi_1\rangle = \langle \phi_1, (\mathcal{R}D_x\mathcal{R}^{-1})\cdot\phi_1\rangle.$$

As $\mathcal{R}D_x\mathcal{R}^{-1} = R^{-1}D_x$, we conclude that $\langle \phi_1, D_x\phi_1\rangle$ is either 0 or an eigenvector of R. Since the latter cannot be real, we conclude $\langle \phi_1, D_x\phi_1\rangle = 0$. The same argument applies to $\langle \phi_2, D_x\phi_2\rangle$.

2. We now analyze $\langle \phi_1, D_x\phi_2\rangle$. Since $\phi_1 \in L^2_{\xi_\star,\tau}$ and $\phi_2 \in L^2_{\xi_\star,\bar\tau}$

$$\langle \phi_1, D_x\phi_2\rangle = \langle \mathcal{R}\phi_1, \mathcal{R}D_x\phi_2\rangle = \langle \tau\phi_1, \mathcal{R}D_x\mathcal{R}^{-1}\cdot\bar\tau\phi_2\rangle = \bar\tau^2\langle \phi_1, (\mathcal{R}D_x\mathcal{R}^{-1})\cdot\phi_2\rangle.$$

As $\mathcal{R}D_x\mathcal{R}^{-1} = R^{-1}D_x$ and $\bar\tau^2 = \tau$, we deduce $R\langle \phi_1, D_x\phi_2\rangle = \tau\langle \phi_1, D_x\phi_2\rangle$. This yields $\langle \phi_1, D_x\phi_2\rangle \in \ker_{\mathbb{C}^2}(R-\tau)$. This eigenspace is $\mathbb{C}\cdot[1, i]^\top$; thus there exists $\nu_\star \in \mathbb{C}$ with

$$2\langle \phi_1, D_x\phi_2\rangle = \nu_\star \cdot \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

If we identify the point $\eta = (\eta_1, \eta_2) \in \mathbb{R}^2$ with $\eta_1 + i\eta_2 \in \mathbb{C}$, then

$$2\langle \phi_1, (\eta\cdot D_x)\phi_2\rangle = 2\langle \phi_1, (\eta_1 D_{x_1} + \eta_2 D_{x_2})\phi_2\rangle = \nu_\star\eta_1 + i\nu_\star\eta_2 = \nu_\star\eta.$$

Above $\nu_\star\eta$ denotes the multiplication of $\nu_\star$ with $\eta = \eta_1 + i\eta_2$. Taking the complex conjugate of this identity and observing that $\eta\cdot D_x$ is a selfadjoint operator, we get

$$2\langle \phi_2, (\eta\cdot D_x)\phi_1\rangle = \overline{\nu_\star\eta}.$$

3. It remains to show that $|\nu_\star| = v_F$. Fix $\eta \in \mathbb{R}^2$ with $|\eta| = 1$. Because of perturbation theory of eigenvalues, the operator $P_0(\xi_\star + t\eta)$ has precisely two eigenvalues near $E_\star$ when t is sufficiently small —

see [Kato 1980, §VII1.3, Theorem 1.8]. Because $(\xi_\star, E_\star)$ is a Dirac point of P_0, they are

$$E_\star \pm v_F t + O(t^2). \tag{A-1}$$

Let $\xi = \xi_\star + t\eta$. We want to construct approximate eigenvectors of $P_0(\xi)$. Let $a, b \in \mathbb{C}^2$, $\mu \in \mathbb{R}$, and $v \in H^2_{\xi_\star}$, with $v = O_{H^2_{\xi_\star}}(1)$ uniformly in t. Then

$$\begin{aligned} e^{-it\langle\eta,x\rangle}(P_0 - E_\star + \mu t)e^{it\langle\eta,x\rangle} \cdot (a\phi_1 + b\phi_2 + tv) \\ &= ((D_x + t\eta)^2 + V - E_\star + \mu t)(a\phi_1 + b\phi_2 + tv) \\ &= t(P_0 - E_\star)v + t(2\eta \cdot D_x + \mu)(a\phi_1 + b\phi_2) + O_{L^2_{\xi_\star}}(t^2). \end{aligned} \tag{A-2}$$

We now construct v such that

$$(P_0 - E_\star)v + (2\eta \cdot D_x + \mu)(a\phi_1 + b\phi_2) = 0. \tag{A-3}$$

This equation admits a solution if and only if $(2\eta \cdot D_x + \mu)(a\phi_1 + b\phi_2)$ is orthogonal to ϕ_1 and ϕ_2. This solvability condition is equivalent to

$$\begin{cases} \langle \phi_1, (2\eta \cdot D_x + \mu)(a\phi_1 + b\phi_2)\rangle = 0, \\ \langle \phi_2, (2\eta \cdot D_x + \mu)(a\phi_1 + b\phi_2)\rangle = 0, \end{cases} \quad \Longleftrightarrow \quad \begin{cases} v_\star \eta \cdot b + \mu a = 0, \\ \overline{v_\star \eta} \cdot b + \mu a = 0. \end{cases} \tag{A-4}$$

A nontrivial solution of (A-4) exists if and only if

$$\operatorname{Det} \begin{bmatrix} \mu & v_\star \eta \\ \overline{v_\star \eta} & \mu \end{bmatrix} = 0 \quad \Longleftrightarrow \quad |v_\star \eta|^2 = |v_\star|^2 = \mu^2.$$

Therefore, when $\mu = |v_\star|$, we can construct $(a, b) \neq (0, 0)$ satisfying (A-4) for $\mu = \pm|v_\star|$. With this choice, (A-3) admits a solution v. It follows from (A-2) that

$$(P_0(\xi) - E_\star + |v_\star|t) \cdot e^{it\langle\eta,x\rangle}(a\phi_1 + b\phi_2 + tv) = O(t^2).$$

In other words, we constructed an $O(t^2)$-accurate quasimode for $P_0(\xi)$, with energy $E_\star + |v_\star|t$. A general principle — see, e.g., [Drouot et al. 2018, Lemma 3.1] — implies that $P_0(\xi)$ has an eigenvalue at $E_\star - |v_\star|t + O(t^2)$. Because of (A-1), this eigenvalue must be $E_\star - v_F t + O(t^2)$. This implies $|v_\star| = v_F$ and completes the proof. □

Proof of Lemma 2.3. Below we use $\langle \cdot, \cdot \rangle$ instead of $\langle \cdot, \cdot \rangle_{L^2_{\xi_\star}}$ to simplify notation. We start by proving the first identity. Since $\mathcal{I}$ is an isometry and $\mathcal{I}\phi_2 = \bar{\phi}_1$,

$$\langle \phi_2, W\phi_1 \rangle = \langle \mathcal{I}\phi_2, \mathcal{I}W\mathcal{I}\phi_1 \rangle = -\langle \bar{\phi}_1, W\bar{\phi}_2 \rangle = -\langle \phi_2, W\phi_1 \rangle.$$

This implies $\langle \phi_2, W\phi_1 \rangle = 0$. Using that W is real-valued, $\langle \phi_1, W\phi_2 \rangle = 0$ as well. We prove now the second identity: for the same reasons as above,

$$\langle \phi_1, W\phi_1 \rangle = \langle \mathcal{I}\phi_1, \mathcal{I}W\mathcal{I}\phi_1 \rangle = -\langle \bar{\phi}_2, W\bar{\phi}_2 \rangle = -\langle \phi_2, W\phi_2 \rangle. \qquad \square$$

Proof of Lemma 7.3. Below we use $\langle \cdot, \cdot \rangle$ instead of $\langle \cdot, \cdot \rangle_{L^2_{\xi_\star}}$ to simplify notations. We start by proving the first identity. Since $\mathcal{I}$ is an isometry from $L^2_{\xi_\star^A}$ to $L^2_{\xi_\star^B}$ and $\mathcal{I}\phi_2 = \bar{\phi}_1$, $\mathcal{I}\mathbb{W} = \mathbb{W}\mathcal{I}$,

$$\langle \phi_2, \mathbb{W}\phi_1 \rangle = \langle \mathcal{I}\phi_2, \mathcal{I}\mathbb{W}\mathcal{I}\phi_1 \rangle = \langle \bar{\phi}_1, \mathbb{W}\bar{\phi}_2 \rangle.$$

Moreover, $\overline{\mathbb{W}} = -\mathbb{W}$ because $\mathbb{A}$ is real-valued and $D_x = (1/i)\nabla$. Therefore,

$$\langle \phi_2, \mathbb{W}\phi_1\rangle = -\langle \bar{\phi}_1, \bar{\mathbb{W}}\bar{\phi}_2\rangle = -\langle \mathbb{W}\phi_2, \phi_1\rangle = -\langle \phi_2, \mathbb{W}\phi_1\rangle.$$

We used in the last equality the selfadjointness of $\mathbb{W}$. We deduce $\langle \phi_2, \mathbb{W}\phi_1\rangle = 0$. Similarly, $\langle \phi_1, \mathbb{W}\phi_2\rangle = 0$.

We prove now the second identity: for the same reasons as above,

$$\langle \phi_1, \mathbb{W}\phi_1\rangle = \langle \mathcal{I}\phi_1, \mathcal{I}\mathbb{W}\phi_1\rangle = -\langle \bar{\phi}_2, \mathbb{W}\bar{\phi}_2\rangle = -\langle \bar{\phi}_2, \overline{\mathbb{W}\phi}_2\rangle = -\langle \phi_2, \mathbb{W}\phi_2\rangle. \qquad \square$$

A.2. ***Spectrum of the Dirac operator.***

Proof of Lemma 3.1. 1. Introduce the matrices

$$\boldsymbol{m_1} = \frac{1}{\nu_F|k'|}\begin{bmatrix} 0 & \nu_\star k' \\ \overline{\nu_\star k'} & 0 \end{bmatrix}, \quad \boldsymbol{m_2} = \frac{1}{\nu_F|\ell|}\begin{bmatrix} 0 & \nu_\star \ell \\ \overline{\nu_\star \ell} & 0 \end{bmatrix}, \quad \boldsymbol{m_3} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Note that $\boldsymbol{m_j}^2 = \mathrm{Id}$. Moreover, the matrices $\boldsymbol{m_j}$ anticommute: $\boldsymbol{m_j m_k} + \boldsymbol{m_k m_j} = 0$ when $j \neq k$. Indeed, $\boldsymbol{m_1 m_2} + \boldsymbol{m_2 m_1}$ equals

$$\frac{1}{\nu_F|k'|\cdot\nu_F|\ell|}\begin{bmatrix} \nu_\star k'\cdot\overline{\nu_\star\ell} + \nu_\star\ell\cdot\overline{\nu_\star k'} & 0 \\ 0 & \nu_\star k'\cdot\overline{\nu_\star\ell} + \nu_\star\ell\cdot\overline{\nu_\star k'} \end{bmatrix} = \frac{2\,\mathrm{Re}(k'\bar{\ell})}{|\ell k'|} = 0,$$

because $\mathrm{Re}(k'\bar{\ell}) = \langle k', \ell\rangle = 0$. With this notation,

$$\not{D}(\mu) = \nu_F|k'|\boldsymbol{m_1}D_t + \mu\cdot\nu_F|\ell|\boldsymbol{m_2} + \vartheta_\star\boldsymbol{m_3}\kappa = \not{D}_\star + \mu\cdot\nu_F|\ell|\boldsymbol{m_2}.$$

2. The formula for the essential spectrum is derived by looking at those of the asymptotic operators:

$$\not{D}_\pm(\mu) \overset{\text{def}}{=} \nu_F|k'|\boldsymbol{m_1}D_t + \mu\cdot\nu_F|\ell|\boldsymbol{m_2} \pm \vartheta_\star\boldsymbol{m_3}.$$

These are Fourier multipliers. Their essential spectrum corresponds to the possible eigenvalues of their symbol as the Fourier parameter runs through $\mathbb{R}$. We find

$$\Sigma_{L^2,\mathrm{ess}}(\not{D}_\pm(\mu)) = \mathbb{R}\setminus\big(-\sqrt{\vartheta_F^2 + \mu^2\cdot\nu_F^2|\ell|^2}, \sqrt{\vartheta_F^2 + \mu^2\cdot\nu_F^2|\ell|^2}\big).$$

3. We start by studying the bifurcation of the zero mode of $\not{D}_\star = \not{D}(0)$. This mode satisfies the equation $\not{D}(0)u = 0$ or equivalently

$$(\nu_F|k'|\partial_t + \vartheta_\star i\boldsymbol{m_1 m_3}\kappa)u = 0.$$

The matrix $i\boldsymbol{m_1 m_3}$ has eigenvalues ± 1. Let u_0 be an eigenvector of $i\boldsymbol{m_1 m_3}$ associated with the eigenvalue $\mathrm{sgn}(\vartheta_\star)$ and set

$$u(t) = u_0\cdot\exp\bigg(-\frac{\vartheta_F}{\nu_F|k'|}\int_0^t \kappa(s)\,ds\bigg).$$

A direct calculation shows that u is an eigenvector of $\not{D}(0)$.

We claim that $\boldsymbol{m_2}u_0 = \mathrm{sgn}(\vartheta_\star)u_0$. Since $i\boldsymbol{m_1 m_3}u_0 = \mathrm{sgn}(\vartheta_\star)u_0$,

$$i\boldsymbol{m_2 m_1 m_3}u_0 = \mathrm{sgn}(\vartheta_\star)\boldsymbol{m_2}u_0, \quad i\boldsymbol{m_2 m_1 m_3} = \frac{i}{|k'\ell|}\begin{bmatrix} \ell\bar{k}' & 0 \\ 0 & -\bar{\ell}k' \end{bmatrix}.$$

Recall that $\mathrm{Re}(\ell\bar{k}') = 0$ because ℓ and k' are orthogonal. Therefore $-\bar{\ell}k' = \ell\bar{k}'$, and we deduce that

$$\mathrm{sgn}(\vartheta_\star)\boldsymbol{m_2}u_0 = -\frac{i}{|k'\ell|}k'\bar{\ell}u_0 \quad \Longrightarrow \quad \boldsymbol{m_2}u_0 = \mathrm{sgn}(\mathrm{Im}(k'\bar{\ell}))\cdot\mathrm{sgn}(\vartheta_\star)u_0.$$

We recall that $k' = -a_2k_1 + a_1k_2$, $k = b_2k_1 - b_1k_2$, $a_2b_1 - b_2a_1 = 1$ — see Section 2E. Hence

$$\mathrm{Im}(k'\bar{\ell}) = \mathrm{Det}[k, k'] = (a_2b_1 - b_2a_1)\cdot\mathrm{Det}[k_1, k_2] = 1 > 0.$$

We deduce that $\boldsymbol{m_2}u_0 = \mathrm{sgn}(\vartheta_\star)u_0$ and $\boldsymbol{m_2}u = \mathrm{sgn}(\vartheta_\star)u$.

We recall that $\not{D}(\mu) = \not{D}_\star + \mu\cdot v_F|\ell|\boldsymbol{m_2}$, $\not{D}_\star u = 0$ and obtain

$$\not{D}(\mu)u = \mu\cdot v_F|\ell|\cdot\mathrm{sgn}(\vartheta_\star)u.$$

This shows that $\mu\cdot v_F|\ell|\cdot\mathrm{sgn}(\vartheta_\star)$ is an eigenvalue of $\not{D}(\mu)$.

4. Let $\vartheta_j > 0$ be an eigenvalue of $\not{D}_\star$. Since $\boldsymbol{m_2}\not{D}_\star = -\not{D}_\star\boldsymbol{m_2}$, we deduce that $-\vartheta_j$ is also eigenvalue of $\not{D}_\star$. The respective eigenvectors are denoted by f_+, f_- and are related via $\boldsymbol{m_2}f_+ = f_-$. We look for an eigenpair $(E, a_+f_+ + a_-f_-)$ of $\not{D}(\mu) = \not{D}_\star + \mu\cdot v_F|\ell|\boldsymbol{m_2}$: it suffices to solve the equation

$$\begin{aligned}(\not{D}_\star + \mu v_F|\ell|\boldsymbol{m_2})\sum_\pm a_\pm f_\pm = E\sum_\pm a_\pm f_\pm \quad &\Longleftrightarrow \quad \sum_\pm \pm\vartheta_j a_\pm f_\pm + \mu\cdot v_F|\ell|a_\pm f_\mp = E\sum_\pm a_\pm f_\pm \\ &\Longleftrightarrow \quad (\vartheta_j\boldsymbol{\sigma_3} + \mu\cdot v_F|\ell|\boldsymbol{\sigma_1})\begin{bmatrix}a_+\\ a_-\end{bmatrix} = Ea.\end{aligned}$$

This is equivalent to (E, a) being an eigenpair of $\vartheta_j\boldsymbol{\sigma_3} + \mu\cdot v_F|\ell|\boldsymbol{\sigma_1}$. Thus we conclude that $E = \pm\sqrt{\vartheta_j^2 + \mu^2\cdot v_F^2|\ell|^2}$ are both eigenvalues of $\not{D}(\mu)$.

5. So far we only showed that the eigenvalues of $\not{D}_\star$ induce eigenvalues of $\not{D}(\mu)$. We must prove the converse statement. Without loss of generality, $\mu \neq 0$. We first deal with eigenvalues of $\not{D}(\mu)$ which *apparently* do not bifurcate from the zero mode of $\not{D}_\star$. That is, we assume first that (E, f) is an eigenpair of $\not{D}(\mu) = \not{D}_\star + \mu\cdot v_F|\ell|\boldsymbol{m_2}$, with $E \neq \mathrm{sgn}(\vartheta_\star)\cdot v_F|\ell|\mu$.

We first claim that f and $g = \boldsymbol{m_2}f$ are linearly independent. Otherwise, we would have $f = \boldsymbol{m_2}f$ or $f = -\boldsymbol{m_2}f$ because $\boldsymbol{m_2}^2 = \mathrm{Id}$. This would imply respectively in the first and second cases

$$\not{D}_\star f = (E - \mu\cdot v_F|\ell|)f \quad \text{or} \quad \not{D}_\star f = (E + \mu\cdot v_F|\ell|)f. \tag{A-5}$$

In particular, f is an eigenvector of $\not{D}_\star$ with $\boldsymbol{m_2}f$ and f colinear. Because of step 3 it must be a zero mode of $\not{D}_\star$. Because of step 2 we must have $\boldsymbol{m_2}f = \mathrm{sgn}(\vartheta_\star)f$. Going back to (A-5), $E = \mathrm{sgn}(\vartheta_\star)\cdot v_F|\ell|$, which contradicts our assumption.

We now look for an eigenpair of $\not{D}_\star$ in the form $(\vartheta_j, af + bg)$. We get the equation

$$\begin{aligned}\not{D}_\star(af + bg) = \vartheta_j(af + bg) \quad &\Longleftrightarrow \quad a(Ef - \mu\cdot v_F|\ell|g) + b(\mu\cdot v_F|\ell|f - Eg) = \vartheta_j(af + bg) \\ &\Longleftrightarrow \quad (E\boldsymbol{\sigma_1} + i\mu\cdot v_F|\ell|\boldsymbol{\sigma_2})\begin{bmatrix}a\\ b\end{bmatrix} = \vartheta_j\begin{bmatrix}a\\ b\end{bmatrix}.\end{aligned}$$

Hence, ϑ is an eigenvalue of $E\boldsymbol{\sigma_1} + i\mu\cdot v_F|\ell|\boldsymbol{\sigma_2}$; equivalently, $\vartheta_j = \pm\sqrt{E^2 - \mu^2\cdot v_F^2|\ell|^2}$.

6. To conclude we deal with the case of an eigenpair (E, f) of $\not{D}(\mu)$ with $E = \mu \cdot v_F|\ell| \cdot \operatorname{sgn}(\vartheta_\star)$ — i.e., when E seemingly bifurcates from the zero mode of $\not{D}_\star$.

We claim that f and $g = \boldsymbol{m_2} f$ are colinear. Otherwise, following the last part of step 5, we would be able to construct $[a, b]^\top$, an eigenvector of $\operatorname{sgn}(\vartheta_\star)\boldsymbol{\sigma_1} + i\boldsymbol{\sigma_2}$ such that $af + bg$ is an eigenvector of $\not{D}_\star$. The matrix $\operatorname{sgn}(\vartheta_\star)\boldsymbol{\sigma_1} + i\boldsymbol{\sigma_2}$ has only one eigenvector, which is either $[0, 1]^\top$ or $[1, 0]^\top$. Therefore either f or g — but not both — is an eigenvector of $\not{D}_\star$. This implies that f or g is a zero eigenvector of $\not{D}_\star$. In particular, f and $\boldsymbol{m_2} f$ (or g and $\boldsymbol{m_2} g$) are colinear — which is a contradiction.

It follows that $f = \boldsymbol{m_2} f$ or $f = -\boldsymbol{m_2} f$. If $\boldsymbol{m_2} f = \operatorname{sgn}(\vartheta_\star) f$, we are done. In the other case, we deduce the existence of an eigenpair $(f, 2\mu \cdot v_F|\ell| \cdot \operatorname{sgn}(\vartheta_\star))$ of $\not{D}_\star$. This would require f and $\boldsymbol{m_2} f$ to be colinear, which is impossible. This completes the proof of the converse statement.

7. The argument presented in steps 5 and 6 shows that the eigenvalues of $\not{D}(\mu)$ and $\not{D}_\star$ have the same multiplicity. Appendix C of [Drouot et al. 2018] shows that $\not{D}_\star$ has only simple eigenvalues. □

A.3. *A calculation.*

Proof of Lemma 6.1. 1. From Theorem 5.1, when (6-2) is satisfied,

$$(P_\delta[\zeta] - \lambda)^{-1} \pm (P_{-\delta}[\zeta] - \lambda)^{-1} = S_\delta(\mu, z) \pm S_{-\delta}(\mu, z) + \mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}).$$

A calculation yields

$$S_\delta(\mu, z) \pm S_{-\delta}(\mu, z) = \frac{1}{\delta} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle} \Pi^* \cdot \mathcal{U}_\delta \big((\not{D}_+(\mu) - z)^{-1} \pm (\not{D}_-(\mu) - z)^{-1} \big) \mathcal{U}_\delta^{-1} \cdot \Pi e^{i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}}.$$

We now compute the resolvent difference $(\not{D}_+(\mu) - z)^{-1} \pm (\not{D}_-(\mu) - z)^{-1}$. We have

$$(\not{D}_+(\mu) - z)^{-1} + (\not{D}_-(\mu) - z)^{-1} = 2 \begin{bmatrix} z & \nu_\star k' D_t + \mu \nu_\star \ell \\ \overline{\nu_\star k'} D_t + \mu \overline{\nu_\star \ell} & z \end{bmatrix} R_0(\mu, z),$$

$$(\not{D}_+(\mu) - z)^{-1} - (\not{D}_-(\mu) - z)^{-1} = 2\vartheta_\star \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} R_0(\mu, z) = 2\vartheta_\star \boldsymbol{\sigma_3} R_0(\mu, z).$$

Above we recall that $R_0(\mu, z) = (v_F^2 |k'|^2 D_t^2 + \mu^2 \cdot v_F^2 |\ell|^2 + \vartheta_F^2 - z^2)^{-1}$. This implies that

$$\begin{aligned} &S_\delta(\mu, z) + S_{-\delta}(\mu, z) \\ &\quad = \frac{2}{\delta} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle} \Pi^* \cdot \mathcal{U}_\delta \begin{bmatrix} z & \nu_\star k' D_t + \mu \nu_\star \ell \\ \overline{\nu_\star k'} D_t + \mu \overline{\nu_\star \ell} & z \end{bmatrix} R_0(\mu, z) \mathcal{U}_\delta^{-1} \cdot \Pi e^{i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}} \end{aligned}$$

and

$$S_\delta(\mu, z) - S_{-\delta}(\mu, z) = \frac{2}{\delta} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle} \Pi^* \cdot \mathcal{U}_\delta \vartheta_\star \boldsymbol{\sigma_3} R_0(\mu, z) \mathcal{U}_\delta^{-1} \cdot \Pi e^{i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}}.$$

We similarly obtain

$$(k' \cdot D_x)(S_\delta(\mu, z) - S_{-\delta}(\mu, z)) = \frac{2}{\delta} \cdot \begin{bmatrix} (k' \cdot D_x)\phi_1 \\ (k' \cdot D_x)\phi_2 \end{bmatrix}^\top e^{-i\mu\delta\langle \ell, x\rangle} \Pi^* \cdot \mathcal{U}_\delta \vartheta_\star \boldsymbol{\sigma_3} R_0(\mu, z) \mathcal{U}_\delta^{-1} \cdot \Pi e^{i\mu\delta\langle \ell, x\rangle} \overline{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}}.$$

2. From the definition of $\mathscr{K}_\delta[\zeta](z)$, we see that

$$\begin{aligned}\mathscr{K}_\delta[\zeta](z) &= \tfrac{1}{2}\big([-\Delta,\kappa_\delta]+\delta(\kappa_\delta^2-1)W\big)\big((P_\delta[\zeta]-\lambda)^{-1}-(P_{-\delta}[\zeta]-\lambda)^{-1}\big)\\ &= \tfrac{1}{2}\big(2(D_t\kappa)_\delta\cdot(k'\cdot D_x)+\delta(\kappa_\delta^2-1)W\big)\big(S_\delta(\mu,z)-S_{-\delta}(\mu,z)\big)+\mathscr{O}_{L^2[\zeta]}(\delta^{2/3}).\end{aligned}$$

Thanks to step 1, the leading-order term is

$$\begin{aligned}&\mathcal{K}_\delta(\mu,z)\\ &\overset{\text{def}}{=}\vartheta_\star\left(2(D_t\kappa)_\delta\cdot\begin{bmatrix}k'\cdot D_x\phi_1\\ k'\cdot D_x\phi_2\end{bmatrix}^\top+(\kappa_\delta^2-1)W\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top\right)\cdot e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\cdot\ \mathcal{U}_\delta\boldsymbol{\sigma_3}R_0(\mu,z)\mathcal{U}_\delta^{-1}\cdot\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}.\end{aligned}$$

3. Because of the definition (6-1) and Theorem 5.1,

$$\begin{aligned}\mathscr{Q}_\delta(\zeta,\lambda) &= \frac{1}{2}\cdot((P_\delta[\zeta]-\lambda)^{-1}+(P_{-\delta}[\zeta]-\lambda)^{-1})+\frac{\kappa_\delta}{2}\cdot((P_\delta[\zeta]-\lambda)^{-1}-(P_{-\delta}[\zeta]-\lambda)^{-1})\\ &= \frac{1}{2}(S_\delta(\mu,z)+S_{-\delta}(\mu,z))+\frac{\kappa_\delta}{2}\cdot(S_\delta(\mu,z)-S_{-\delta}(\mu,z))+\mathscr{O}_{L^2[\zeta]}(\delta^{-1/3}).\end{aligned}$$

Thanks to the first step, the leading-order term is

$$\begin{aligned}\mathcal{Q}_\delta(\mu,z)\overset{\text{def}}{=}\frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\cdot\mathcal{U}_\delta\cdot\begin{bmatrix}z & \nu_\star k'D_t+\mu\nu_\star\ell\\ \overline{\nu_\star k'}D_t+\mu\overline{\nu_\star\ell} & z\end{bmatrix}R_0(\mu,z)\cdot\mathcal{U}_\delta^{-1}\cdot\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}\\ +\kappa_\delta\cdot\frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\cdot\mathcal{U}_\delta\cdot\vartheta_\star\boldsymbol{\sigma_3}R_0(\mu,z)\cdot\mathcal{U}_\delta^{-1}\cdot\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}.\end{aligned}$$

A key identity is $\kappa_\delta\Pi^*\mathcal{U}_\delta=\Pi^*\mathcal{U}_\delta\kappa$. Therefore, we deduce that

$$\mathcal{Q}_\delta(\mu,z)=\frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\cdot\mathcal{U}_\delta\cdot\begin{bmatrix}\vartheta_\star\kappa+z & \nu_\star k'D_t+\mu\nu_\star\ell\\ \overline{\nu_\star k'}D_t+\mu\overline{\nu_\star\ell} & -\vartheta_\star\kappa+z\end{bmatrix}R_0(\mu,z)\cdot\mathcal{U}_\delta^{-1}\cdot\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}.$$

The operator

$$\not{D}(\mu)+z=\begin{bmatrix}\vartheta_\star\kappa+z & \nu_\star k'D_t+\mu\nu_\star\ell\\ \overline{\nu_\star k'}D_t+\mu\overline{\nu_\star\ell} & -\vartheta_\star\kappa+z\end{bmatrix}$$

emerges and we end up with

$$\mathcal{Q}_\delta(\mu,z)=\frac{1}{\delta}\cdot\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}^\top e^{-i\mu\delta\langle\ell,x\rangle}\Pi^*\cdot\mathcal{U}_\delta(\not{D}(\mu)+z)\cdot R_0(\mu,z)\mathcal{U}_\delta^{-1}\cdot\Pi e^{i\mu\delta\langle\ell,x\rangle}\overline{\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}}.\qquad\square$$

Acknowledgements

I would like to thank Michael Weinstein for his in-depth introduction to the subject and for suggesting the project. Support from the Simons Foundation through M. Weinstein's Math+X investigator award #376319 and from NSF DMS-1800086 are gratefully acknowledged.

References

[Ablowitz and Zhu 2012] M. J. Ablowitz and Y. Zhu, "Nonlinear waves in shallow honeycomb lattices", *SIAM J. Appl. Math.* **72**:1 (2012), 240–260. MR Zbl

[Ablowitz and Zhu 2013] M. J. Ablowitz and Y. Zhu, "Nonlinear wave packets in deformed honeycomb lattices", *SIAM J. Appl. Math.* **73**:6 (2013), 1959–1979. MR Zbl

[Ammari et al. 2018] H. Ammari, B. Fitzpatrick, H. Lee, E. O. Hiltunen, and S. Yu, "Honeycomb-lattice Minnaert bubbles", preprint, 2018. arXiv

[Ando et al. 1975] T. Ando, Y. Matsumoto, and Y. Uemura, "Theory of Hall effect in a two-dimensional electron system", *J. Phys. Soc. Japan* **39**:2 (1975), 279–288.

[Arbunich and Sparber 2018] J. Arbunich and C. Sparber, "Rigorous derivation of nonlinear Dirac equations for wave propagation in honeycomb structures", *J. Math. Phys.* **59**:1 (2018), art. id. 011509. MR Zbl

[Avila et al. 2013] J. C. Avila, H. Schulz-Baldes, and C. Villegas-Blas, "Topological invariants of edge states for periodic two-dimensional models", *Math. Phys. Anal. Geom.* **16**:2 (2013), 137–170. MR Zbl

[Bal 2017] G. Bal, "Topological protection of perturbed edge states", preprint, 2017. arXiv

[Bal 2018] G. Bal, "Continuous bulk and interface description of topological insulators", preprint, 2018. arXiv

[Becker and Zworski 2019] S. Becker and M. Zworski, "Magnetic oscillations in a model of graphene", *Comm. Math. Phys.* **367**:3 (2019), 941–989. MR

[Becker et al. 2018] S. Becker, R. Han, and S. Jitomirskaya, "Cantor spectrum of graphene in magnetic fields", preprint, 2018. arXiv

[Berkolaiko and Comech 2018] G. Berkolaiko and A. Comech, "Symmetry and Dirac points in graphene spectrum", *J. Spectr. Theory* **8**:3 (2018), 1099–1147. MR Zbl

[Borisov 2007] D. I. Borisov, "Some singular perturbations of periodic operators", *Teoret. Mat. Fiz.* **151**:2 (2007), 207–218. In Russian; translated in *Theor. Math. Phys.* **151**:2 (2007), 614–624. MR Zbl

[Borisov 2011] D. I. Borisov, "On the spectrum of a two-dimensional periodic operator with a small localized perturbation", *Izv. Ross. Akad. Nauk Ser. Mat.* **75**:3 (2011), 29–64. In Russian; translated in *Izv. Math.* **75**:3 (2011), 471–505. MR Zbl

[Borisov 2015] D. I. Borisov, "On the band spectrum of a Schrödinger operator in a periodic system of domains coupled by small windows", *Russ. J. Math. Phys.* **22**:2 (2015), 153–160. MR Zbl

[Borisov and Gadyl'shin 2006] D. I. Borisov and R. R. Gadyl'shin, "The spectrum of the Schrödinger operator with a rapidly oscillating compactly supported potential", *Teoret. Mat. Fiz.* **147**:1 (2006), 58–63. In Russian; translated in *Theor. Math. Phys.* **147**:1 (2006), 496–500. MR Zbl

[Borisov and Gadyl'shin 2008] D. I. Borisov and R. R. Gadyl'shin, "On the spectrum of a periodic operator with small localized perturbation", *Izv. Ross. Akad. Nauk Ser. Mat.* **72**:4 (2008), 37–66. In Russian; translated in *Izv. Math.* **72**:4 (2008), 659–688). MR Zbl

[Bourne and Rennie 2018] C. Bourne and A. Rennie, "Chern numbers, localisation and the bulk-edge correspondence for continuous models of topological phases", *Math. Phys. Anal. Geom.* **21**:3 (2018), art. id. 16. MR Zbl

[Braverman 2018] M. Braverman, "The spectral flow of a family of Toeplitz operators", preprint, 2018. arXiv

[Brendel et al. 2017] C. Brendel, V. Peano, O. Painter, and F. Marquardt, "Snowflake topological insulator for sound waves", preprint, 2017. arXiv

[Carles et al. 2004] R. Carles, P. A. Markowich, and C. Sparber, "Semiclassical asymptotics for weakly nonlinear Bloch waves", *J. Statist. Phys.* **117**:1-2 (2004), 343–375. MR Zbl

[Carlsson 1990] U. Carlsson, "An infinite number of wells in the semi-classical limit", *Asymptotic Anal.* **3**:3 (1990), 189–214. MR Zbl

[Colin de Verdière 1991] Y. Colin de Verdière, "Sur les singularités de van Hove génériques", pp. 99–110 in *Analyse globale et physique mathématique* (Lyon, 1989), Mém. Soc. Math. France (N.S.) **46**, Soc. Math. France, Paris, 1991. MR Zbl

[Cornean et al. 2015] H. D. Cornean, V. Iftimie, and R. Purice, "Peierls substitution and magnetic pseudo-differential calculus", preprint, 2015. arXiv

[Cornean et al. 2017a] H. D. Cornean, B. Helffer, and R. Purice, "Low lying spectral gaps induced by slowly varying magnetic fields", *J. Funct. Anal.* **273**:1 (2017), 206–282. MR Zbl

[Cornean et al. 2017b] H. D. Cornean, B. Helffer, and R. Purice, "Peierls' substitution for low lying spectral energy windows", preprint, 2017. arXiv

[Cycon et al. 1987] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon, *Schrödinger operators with application to quantum mechanics and global geometry*, Springer, 1987. MR Zbl

[De Nittis and Lein 2011] G. De Nittis and M. Lein, "Applications of magnetic ΨDO techniques to SAPT", *Rev. Math. Phys.* **23**:3 (2011), 233–260. MR Zbl

[De Nittis and Lein 2014] G. De Nittis and M. Lein, "Effective light dynamics in perturbed photonic crystals", *Comm. Math. Phys.* **332**:1 (2014), 221–260. MR Zbl

[Deift and Hempel 1986] P. A. Deift and R. Hempel, "On the existence of eigenvalues of the Schrödinger operator $H - \lambda W$ in a gap of $\sigma(H)$", *Comm. Math. Phys.* **103**:3 (1986), 461–490. MR Zbl

[Delplace et al. 2017] P. Delplace, J. B. Marston, and A. Venaille, "Topological origin of equatorial waves", *Science* **358**:6366 (2017), 1075–1077. MR Zbl

[Dimassi 2016] M. Dimassi, "Semi-classical asymptotics for the Schrödinger operator with oscillating decaying potential", *Canad. Math. Bull.* **59**:4 (2016), 734–747. MR Zbl

[Dimassi and Duong 2017] M. Dimassi and A. T. Duong, "Scattering and semi-classical asymptotics for periodic Schrödinger operators with oscillating decaying potential", *Math. J. Okayama Univ.* **59**:1 (2017), 149–174. MR Zbl

[Dohnal et al. 2009] T. Dohnal, M. Plum, and W. Reichel, "Localized modes of the linear periodic Schrödinger operator with a nonlocal perturbation", *SIAM J. Math. Anal.* **41**:5 (2009), 1967–1993. MR Zbl

[Drouot 2018a] A. Drouot, "Bound states for rapidly oscillatory Schrödinger operators in dimension 2", *SIAM J. Math. Anal.* **50**:2 (2018), 1471–1484. MR Zbl

[Drouot 2018b] A. Drouot, "The bulk-edge correspondence for continuous dislocated systems", preprint, 2018. arXiv

[Drouot 2018c] A. Drouot, "Resonances for random highly oscillatory potentials", *J. Math. Phys.* **59**:10 (2018), art. id. 101506. MR Zbl

[Drouot 2018d] A. Drouot, "Scattering resonances for highly oscillatory potentials", *Ann. Sci. Éc. Norm. Supér.* (4) **51**:4 (2018), 865–925. MR Zbl

[Drouot 2019] A. Drouot, "The bulk-edge correspondence for continuous honeycomb lattices", preprint, 2019. arXiv

[Drouot et al. 2018] A. Drouot, C. L. Fefferman, and M. I. Weinstein, "Defect modes for dislocated periodic media", preprint, 2018. arXiv

[Duchêne and Raymond 2018] V. Duchêne and N. Raymond, "Spectral asymptotics for the Schrödinger operator on the line with spreading and oscillating potentials", *Doc. Math.* **23** (2018), 599–636. MR Zbl

[Duchêne and Weinstein 2011] V. Duchêne and M. I. Weinstein, "Scattering, homogenization, and interface effects for oscillatory potentials with strong singularities", *Multiscale Model. Simul.* **9**:3 (2011), 1017–1063. MR Zbl

[Duchêne et al. 2014] V. Duchêne, I. Vukićević, and M. I. Weinstein, "Scattering and localization properties of highly oscillatory potentials", *Comm. Pure Appl. Math.* **67**:1 (2014), 83–128. MR Zbl

[Elgart et al. 2005] A. Elgart, G. M. Graf, and J. H. Schenker, "Equality of the bulk and edge Hall conductances in a mobility gap", *Comm. Math. Phys.* **259**:1 (2005), 185–221. MR Zbl

[Fefferman and Weinstein 2012] C. L. Fefferman and M. I. Weinstein, "Honeycomb lattice potentials and Dirac points", *J. Amer. Math. Soc.* **25**:4 (2012), 1169–1220. MR Zbl

[Fefferman and Weinstein 2014] C. L. Fefferman and M. I. Weinstein, "Wave packets in honeycomb structures and two-dimensional Dirac equations", *Comm. Math. Phys.* **326**:1 (2014), 251–286. MR Zbl

[Fefferman and Weinstein 2018] C. L. Fefferman and M. I. Weinstein, "Edge states of continuum Schrödinger operators for sharply terminated honeycomb structures", preprint, 2018. arXiv

[Fefferman et al. 2014] C. L. Fefferman, J. P. Lee-Thorp, and M. I. Weinstein, "Topologically protected states in one-dimensional continuous systems and Dirac points", *Proc. Natl. Acad. Sci. USA* **111**:24 (2014), 8759–8763. MR Zbl

[Fefferman et al. 2016a] C. L. Fefferman, J. P. Lee-Thorp, and M. I. Weinstein, "Bifurcations of edge states: topologically protected and non-protected in continuous 2D honeycomb structures", *2D Materials* **3**:1 (2016), art. id. 014008.

[Fefferman et al. 2016b] C. L. Fefferman, J. P. Lee-Thorp, and M. I. Weinstein, "Edge states in honeycomb structures", *Ann. PDE* **2**:2 (2016), art. id. 12. MR Zbl

[Fefferman et al. 2017] C. L. Fefferman, J. P. Lee-Thorp, and M. I. Weinstein, *Topologically protected states in one-dimensional systems*, Mem. Amer. Math. Soc. **1173**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

[Fefferman et al. 2018] C. L. Fefferman, J. P. Lee-Thorp, and M. I. Weinstein, "Honeycomb Schrödinger operators in the strong binding regime", *Comm. Pure Appl. Math.* **71**:6 (2018), 1178–1270. MR Zbl

[Figotin and Klein 1997] A. Figotin and A. Klein, "Localized classical waves created by defects", *J. Statist. Phys.* **86**:1-2 (1997), 165–177. MR Zbl

[Fu et al. 2007] L. Fu, C. L. Kane, and E. J. Mele, "Topological insulators in three dimensions", *Phys. Rev. Lett.* **98**:10 (2007), art. id. 106803.

[Fukui et al. 2012] T. Fukui, K. Shiozaki, T. Fujiwara, and S. Fujimoto, "Bulk-edge correspondence for Chern topological phases: a viewpoint from a generalized index theorem", *J. Phys. Soc. Japan* **81**:11 (2012), art. id. 114602.

[Gannot 2015] O. Gannot, "From quasimodes to resonances: exponentially decaying perturbations", *Pacific J. Math.* **277**:1 (2015), 77–97. MR Zbl

[Gérard and Sigal 1992] C. Gérard and I. M. Sigal, "Space-time picture of semiclassical resonances", *Comm. Math. Phys.* **145**:2 (1992), 281–328. MR Zbl

[Gérard et al. 1991] C. Gérard, A. Martinez, and J. Sjöstrand, "A mathematical approach to the effective Hamiltonian in perturbed periodic problems", *Comm. Math. Phys.* **142**:2 (1991), 217–244. MR

[Golowich and Weinstein 2005] S. E. Golowich and M. I. Weinstein, "Scattering resonances of microstructures and homogenization theory", *Multiscale Model. Simul.* **3**:3 (2005), 477–521. MR Zbl

[Graf and Porta 2013] G. M. Graf and M. Porta, "Bulk-edge correspondence for two-dimensional topological insulators", *Comm. Math. Phys.* **324**:3 (2013), 851–895. MR Zbl

[Graf and Shapiro 2018] G. M. Graf and J. Shapiro, "The bulk-edge correspondence for disordered chiral chains", *Comm. Math. Phys.* **363**:3 (2018), 829–846. MR Zbl

[Graf and Tauber 2018] G. M. Graf and C. Tauber, "Bulk-edge correspondence for two-dimensional Floquet topological insulators", *Ann. Henri Poincaré* **19**:3 (2018), 709–741. MR Zbl

[Grushin 2009] V. V. Grushin, "Application of the multiparameter theory of perturbations of Fredholm operators to Bloch functions", *Mat. Zametki* **86**:6 (2009), 819–828. MR Zbl

[Haldane and Raghu 2008] F. D. M. Haldane and S. Raghu, "Possible realization of directional optical waveguides in photonic crystals with broken time-reversal symmetry", *Phys. Rev. Lett.* **100**:1 (2008), art. id. 013904.

[Halperin 1982] B. I. Halperin, "Quantized Hall conductance, current-carrying edge states, and the existence of extended states in a two-dimensional disordered potential", *Phys. Rev. B* **25**:4 (1982), 2185–2190.

[Harrell 1979] E. M. Harrell, "The band-structure of a one-dimensional, periodic system in a scaling limit", *Ann. Physics* **119**:2 (1979), 351–369. MR Zbl

[Hatsugai 1993] Y. Hatsugai, "Chern number and edge states in the integer quantum Hall effect", *Phys. Rev. Lett.* **71**:22 (1993), 3697–3700. MR Zbl

[Helffer and Sjöstrand 1984] B. Helffer and J. Sjöstrand, "Multiple wells in the semiclassical limit, I", *Comm. Partial Differential Equations* **9**:4 (1984), 337–408. MR Zbl

[Helffer and Sjöstrand 1985] B. Helffer and J. Sjöstrand, "Puits multiples en limite semi-classique, II: Interaction moléculaire, symétries, perturbation", *Ann. Inst. H. Poincaré Phys. Théor.* **42**:2 (1985), 127–212. MR Zbl

[Helffer and Sjöstrand 1987] B. Helffer and J. Sjöstrand, "Effet tunnel pour l'équation de Schrödinger avec champ magnétique", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **14**:4 (1987), 625–657. MR

[Hempel and Kohlmann 2011a] R. Hempel and M. Kohlmann, "Spectral properties of grain boundaries at small angles of rotation", *J. Spectr. Theory* **1**:2 (2011), 197–219. MR Zbl

[Hempel and Kohlmann 2011b] R. Hempel and M. Kohlmann, "A variational approach to dislocation problems for periodic Schrödinger operators", *J. Math. Anal. Appl.* **381**:1 (2011), 166–178. MR Zbl

[Hempel et al. 2015] R. Hempel, M. Kohlmann, M. Stautz, and J. Voigt, "Bound states for nano-tubes with a dislocation", *J. Math. Anal. Appl.* **431**:1 (2015), 202–227. MR Zbl

[Hoefer and Weinstein 2011] M. A. Hoefer and M. I. Weinstein, "Defect modes and homogenization of periodic Schrödinger operators", *SIAM J. Math. Anal.* **43**:2 (2011), 971–996. MR Zbl

[Hsieh et al. 2008] D. Hsieh, D. Qian, A. L. Wray, Y. Y. Xia, Y. Hor, R. Cava, and M. Z. Hasan, "A topological Dirac insulator in a quantum spin Hall phase", *Nature* **452** (2008), 970–974.

[Jotzu et al. 2014] G. Jotzu, M. Messer, R. Desbuquois, M. Lebrat, T. Uehlinger, D. Greif, and T. Esslinger, "Experimental realization of the topological Haldane model with ultracold fermions", *Nature* **515** (2014), 237–240.

[Kane and Mele 2005a] C. L. Kane and E. J. Mele, "Quantum spin Hall effect in graphene", *Phys. Rev. Lett.* **95**:22 (2005), art. id. 226801.

[Kane and Mele 2005b] C. L. Kane and E. J. Mele, "Z_2 topological order and the quantum spin Hall effect", *Phys. Rev. Lett.* **95**:14 (2005), art. id. 146802.

[Kato 1980] T. Kato, *Perturbation theory for linear operators*, Grundlehren der Mathematischen Wissenschaften **132**, Springer, 1980. Zbl

[Kellendonk and Schulz-Baldes 2004a] J. Kellendonk and H. Schulz-Baldes, "Boundary maps for C^*-crossed products with $\mathbb{R}$ with an application to the quantum Hall effect", *Comm. Math. Phys.* **249**:3 (2004), 611–637. MR Zbl

[Kellendonk and Schulz-Baldes 2004b] J. Kellendonk and H. Schulz-Baldes, "Quantization of edge currents for continuous magnetic operators", *J. Funct. Anal.* **209**:2 (2004), 388–413. MR Zbl

[Kellendonk et al. 2002] J. Kellendonk, T. Richter, and H. Schulz-Baldes, "Edge current channels and Chern numbers in the integer quantum Hall effect", *Rev. Math. Phys.* **14**:1 (2002), 87–119. MR Zbl

[Keller et al. 2018] R. T. Keller, J. L. Marzuola, B. Osting, and M. I. Weinstein, "Spectral band degeneracies of $\frac{\pi}{2}$-rotationally invariant periodic Schrödinger operators", *Multiscale Model. Simul.* **16**:4 (2018), 1684–1731. MR Zbl

[Khanikaev et al. 2007] A. B. Khanikaev, S. H. Mousavi, W.-K. Tse, M. Kargarian, A. H. MacDonald, and G. Shvets, "Topological insulators with inversion symmetry", *Phys. Rev. B* **76**:4 (2007), art. id. 045302.

[Kitaev 2009] A. Kitaev, "Periodic table for topological insulators and superconductors", pp. 22–30 in *Advances in theoretical physics: Landau Memorial Conference* (Chernogolokova, Russia, 2008), edited by V. Lebedev and M. Feigel'man, AIP Conference Proceedings **1134**, AIP, 2009. Zbl

[von Klitzing et al. 1980] K. von Klitzing, G. Dorda, and M. Pepper, "New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance", *Phys. Rev. Lett.* **45**:6 (1980), 494–497.

[Korotyaev 2000] E. Korotyaev, "Lattice dislocations in a 1-dimensional model", *Comm. Math. Phys.* **213**:2 (2000), 471–489. MR Zbl

[Kuchment and Post 2007] P. Kuchment and O. Post, "On the spectra of carbon nano-structures", *Comm. Math. Phys.* **275**:3 (2007), 805–826. MR Zbl

[Lee 2016] M. Lee, "Dirac cones for point scatterers on a honeycomb lattice", *SIAM J. Math. Anal.* **48**:2 (2016), 1459–1488. MR Zbl

[Lee-Thorp et al. 2019] J. P. Lee-Thorp, M. I. Weinstein, and Y. Zhu, "Elliptic operators with honeycomb symmetry: Dirac points, edge states and applications to photonic graphene", *Arch. Ration. Mech. Anal.* **232**:1 (2019), 1–63. MR Zbl

[Lu et al. 2018] J. Lu, A. B. Watson, and M. I. Weinstein, "Dirac operators and domain walls", preprint, 2018. arXiv

[Martinez 1987] A. Martinez, "Estimations de l'effet tunnel pour le double puits, I", *J. Math. Pures Appl.* (9) **66**:2 (1987), 195–215. MR Zbl

[Martinez 1988] A. Martinez, "Estimations de l'effet tunnel pour le double puits, II: États hautement excités", *Bull. Soc. Math. France* **116**:2 (1988), 199–229. MR Zbl

[Moore and Balents 2007] J. E. Moore and L. Balents, "Topological invariants of time-reversal-invariant band structures", *Phys. Rev. B* **75**:12 (2007), art. id. 121306.

[Nash et al. 2015] L. M. Nash, D. Kleckner, A. Read, V. Vitelli, A. M. Turner, and W. T. M. Irvine, "Topological mechanics of gyroscopic metamaterials", *Proc. Natl. Acad. Sci. USA* **112**:47 (2015), 14495–14500.

[Outassourt 1987] A. Outassourt, "Comportement semi-classique pour l'opérateur de Schrödinger à potentiel périodique", *J. Funct. Anal.* **72**:1 (1987), 65–93. MR Zbl

[Ozawa et al. 2018] T. Ozawa, H. M. Price, A. Amo, N. Goldman, M. Hafezi, L. Lu, M. Rechtsman, D. Schuster, J. Simon, O. Zilberberg, and I. Carusotto, "Topological photonics", preprint, 2018. arXiv

[Panati et al. 2003] G. Panati, H. Spohn, and S. Teufel, "Space-adiabatic perturbation theory", *Adv. Theor. Math. Phys.* **7**:1 (2003), 145–204. MR

[Parzygnat et al. 2010] A. Parzygnat, K. Lee, Y. Avniel, and S. Johnson, "Sufficient conditions for two-dimensional localization by arbitrarily weak defects in periodic potentials with band gaps", *Phys. Rev. B* **81**:15 (2010), art. id. 155324.

[Perrot et al. 2018] M. Perrot, P. Delplace, and A. Venaille, "Topological transition in stratified fluids", preprint, 2018. arXiv

[Post 2003] O. Post, "Eigenvalues in spectral gaps of a perturbed periodic manifold", *Math. Nachr.* **261/262** (2003), 141–162. MR Zbl

[Raghu and Haldane 2008] S. Raghu and F. D. M. Haldane, "Analogs of quantum-Hall-effect edge states in photonic crystals", *Phys. Rev. A* **78**:3 (2008), art. id. 033834.

[Rechtsman et al. 2013] M. C. Rechtsman, J. M. Zeuner, Y. Plotnik, Y. Lumer, D. Podolsky, F. Dreisow, S. Nolte, M. Segev, and A. Szameit, "Photonic Floquet topological insulators", *Nature* **496** (2013), 196–200.

[Reed and Simon 1978] M. Reed and B. Simon, *Methods of modern mathematical physics, IV: Analysis of operators*, Academic, New York, 1978. MR Zbl

[Roy 2009] R. Roy, "Topological phases and the quantum spin Hall effect in three dimensions", *Phys. Rev. B* **79**:19 (2009), art. id. 195322.

[Ryu et al. 2010] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. W. Ludwig, "Topological insulators and superconductors: tenfold way and dimensional hierarchy", *New J. Phys.* **12** (2010), art. id. 065010.

[Schockley 1939] W. Schockley, "On the surface states associated with a periodic potential", *Phys. Rev.* **317**:4 (1939), 317–323.

[Shapiro 2017] J. Shapiro, "The bulk-edge correspondence in three simple cases", preprint, 2017. arXiv

[Shapiro and Tauber 2018] J. Shapiro and C. Tauber, "Strongly disordered Floquet topological systems", preprint, 2018. arXiv

[Simon 1976] B. Simon, "The bound state of weakly coupled Schrödinger operators in one and two dimensions", *Ann. Physics* **97**:2 (1976), 279–288. MR Zbl

[Simon 1984] B. Simon, "Semiclassical analysis of low lying eigenvalues, III: Width of the ground state band in strongly coupled solids", *Ann. Physics* **158**:2 (1984), 415–420. MR Zbl

[Singha et al. 2011] A. Singha, M. Gibertini, B. Karmakar, S. Yuan, M. Polini, G. Vignale, M. I. Kastnelson, A. Pinczuk, L. N. Pfeiffer, K. W. West, and V. Pellegrini, "Two-dimensional Mott–Hubbard electrons in an artificial honeycomb lattice", *Science* **332**:6034 (2011), 1176–1179.

[Stefanov 1999] P. Stefanov, "Quasimodes and resonances: sharp lower bounds", *Duke Math. J.* **99**:1 (1999), 75–92. MR Zbl

[Stefanov 2000] P. Stefanov, "Resonances near the real axis imply existence of quasimodes", *C. R. Acad. Sci. Paris Sér. I Math.* **330**:2 (2000), 105–108. MR Zbl

[Stefanov and Vodev 1996] P. Stefanov and G. Vodev, "Neumann resonances in linear elasticity for an arbitrary body", *Comm. Math. Phys.* **176**:3 (1996), 645–659. MR Zbl

[Taarabt 2014] A. Taarabt, "Equality of bulk and edge Hall conductances for continuous magnetic random Schrödinger operators", preprint, 2014. arXiv

[Tamm 1932] I. Tamm, "Über eine mögliche Art der Elektronenbindung an Kristalloberflächen", *Phys. Z. Sowjetunion* **1** (1932), 733–746. Zbl

[Tang and Zworski 1998] S.-H. Tang and M. Zworski, "From quasimodes to resonances", *Math. Res. Lett.* **5**:3 (1998), 261–272. MR Zbl

[Thicke et al. 2018] K. Thicke, A. Watson, and J. Lu, "Computation of bound states of semi-infinite matrix Hamiltonians with applications to edge states of two-dimensional materials", preprint, 2018. arXiv

[Thouless et al. 1982] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, "Quantized Hall conductance in a two-dimensional periodic potential", *Phys. Rev. Lett.* **49**:6 (1982), 405–408.

[Wallace 1947] P. R. Wallace, "The band theory of graphite", *Phys. Rev.* **71**:9 (1947), 622–634. Zbl

[Wang et al. 2008] Z. Wang, Y. D. Chong, J. D. Joannopoulos, and M. Soljačić, "Reflection-free one-way edge modes in a gyromagnetic photonic crystal", *Phys. Rev. Lett.* **100**:1 (2008), art. id. 013905.

[Waterstraat 2017] N. Waterstraat, "Fredholm operators and spectral flow", *Rend. Semin. Mat. Univ. Politec. Torino* **75**:1 (2017), 7–51. MR

[Watson and Weinstein 2018] A. Watson and M. I. Weinstein, "Wavepackets in inhomogeneous periodic media: propagation through a one-dimensional band crossing", *Comm. Math. Phys.* **363**:2 (2018), 655–698. MR Zbl

[Watson et al. 2017] A. B. Watson, J. Lu, and M. I. Weinstein, "Wavepackets in inhomogeneous periodic media: effective particle-field dynamics and Berry curvature", *J. Math. Phys.* **58**:2 (2017), art. id. 021503. MR Zbl

[Yu et al. 2008] Z. Yu, G. Veronis, Z. Wang, and S. Fan, "One-way electromagnetic waveguide formed at the interface between a plasmonic metal under a static magnetic field and a photonic crystal", *Phys. Rev. Lett.* **100**:2 (2008), art. id. 023902.

[Zelenko 2016] L. Zelenko, "Virtual bound levels in a gap of the essential spectrum of the weakly perturbed periodic Schrödinger operator", *Integral Equations Operator Theory* **85**:3 (2016), 307–345. MR Zbl

[Zhang et al. 2009] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, "Topological insulators in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 with a single Dirac cone on the surface", *Nature Phys.* **5** (2009), 438–442.

Received 21 Dec 2018. Revised 26 Feb 2019. Accepted 4 Apr 2019.

ALEXIS DROUOT: alexis.drouot@gmail.com
Department of Mathematics, Columbia University, New York, NY, United States

PURE and APPLIED
ANALYSIS

Vol. 1, No. 3, 2019
dx.doi.org/10.2140/paa.2019.1.447

msp

MULTIDIMENSIONAL NONLINEAR GEOMETRIC OPTICS FOR TRANSPORT OPERATORS WITH APPLICATIONS TO STABLE SHOCK FORMATION

JARED SPECK

In $n \geq 1$ spatial dimensions, we study the Cauchy problem for a genuinely nonlinear quasilinear transport equation coupled to a quasilinear symmetric hyperbolic subsystem of a rather general type. For an open set (relative to a suitable Sobolev topology) of regular initial data that are close to the data of a simple plane wave, we give a sharp, constructive proof of shock formation in which the transport variable remains bounded but its first-order Cartesian coordinate partial derivatives blow up in finite time. Moreover, we prove that, at least at the low derivative levels, the singularity does not propagate into the symmetric hyperbolic variables: they and their first-order Cartesian coordinate partial derivatives remain bounded, even though they interact with the transport variable all the way up to its singularity. The formation of the singularity is tied to the finite-time degeneration, relative to the Cartesian coordinates, of a system of geometric coordinates adapted to the characteristics of the transport operator. Two crucial features of the proof are that relative to the geometric coordinates, all solution variables remain smooth, and that the finite-time degeneration coincides with the intersection of the transport characteristics. Compared to prior shock formation results in more than one spatial dimension, in which the blowup occurred in solutions to quasilinear wave equations, the main new features of the present work are: (i) we develop a theory of nonlinear geometric optics for transport operators, which is compatible with the coupling and which allows us to implement a quasilinear geometric vector field method, even though the regularity properties of the corresponding eikonal function are less favorable compared to the wave equation case and (ii) we allow for a full quasilinear coupling; i.e., the principal coefficients in all equations are allowed to depend on all solution variables.

MSC2010: primary 35L67; secondary 35L45.

Keywords: blowup, characteristics, eikonal equation, eikonal function, simple wave, singularity, stable blowup, vector field method, wave breaking.

1. Introduction

The study of quasilinear hyperbolic PDE systems is one of the most classical pursuits in mathematics, and it is also among the most active. Such systems are of intense theoretical interest, in no small part due to the fact that their study lies at the core of the revered field of nonlinear hyperbolic conservation laws (more generally "balance laws"); we refer readers to [Dafermos 2010] for a detailed discussion of the history of nonlinear hyperbolic balance laws as well as a comprehensive introduction to the main results of the field and the main techniques behind their proofs, with an emphasis on the case of one spatial dimension. The subject of quasilinear hyperbolic systems is of physical interest as well since they are used to model a vast range of physical phenomena.

A fundamental issue surrounding the study of the initial value problem for such PDEs is that solutions can develop singularities in finite time, starting from regular initial data. In one spatial dimension, the theory is in a rather advanced state, and in many cases, the known well-posedness results are able to accommodate the formation of shock singularities as well as their subsequent interactions; see the aforementioned work of Dafermos. The advanced status of the one-space-dimensional theory is highly indebted to the availability of estimates in the space of functions of bounded variation (BV). In contrast, Rauch [1986] showed that for quasilinear hyperbolic systems in more than one spatial dimension, well-posedness in BV class *generally does not hold*. For this reason, energy estimates in L^2-based Sobolev spaces play an essential role in multiple spatial dimensions, and as a consequence, even the question of whether or not there is stable singularity formation (starting from regular initial data) can be exceptionally challenging. That is, in proving a constructive shock formation result in more than one spatial dimension, one cannot avoid the exacting task of deriving energy estimates that hold up to the singularity; below we will elaborate on this difficulty.

In view of the remarks above, it is not surprising that the earliest blowup results for quasilinear hyperbolic PDEs in more than one spatial dimension without symmetry assumptions were not constructive, but were instead based on proofs by contradiction, with influential contributions coming from, for example, John [1981] for a class of wave equations and Sideris [1984; 1985] for a class of hyperbolic systems in the former work and for the compressible Euler equations in the latter. The main idea of the proofs was to show that for smooth solutions with suitable initial data, certain spatially averaged quantities satisfy ordinary differential inequalities that force them to blow up, contradicting the assumption of smoothness.

Although the blowup results mentioned in the previous paragraph are compelling, their chief drawback is that they provide no information about the nature of the singularity, other than an upper bound on the solution's classical lifespan. In particular, such results are not useful if one aims to extract sharp information about the blowup mechanism and blowup time, or if one aims to uniquely continue the solution past the singularity in a weak sense. In contrast, many state-of-the-art blowup results for hyperbolic PDEs yield a detailed description of the singularity formation, even in the challenging setting of more than one spatial dimension. This is especially true for results on the formation of shocks starting from smooth initial conditions, a topic that has enjoyed remarkable progress in the last decade, as we describe in Section 1G. Our main results are in this vein, our motivation being to advance the rigorous mathematical

theory of the formation of shocks. We recall that a shock singularity[1] is such that some derivative of the solution blows up in finite time, while the solution itself remains bounded. Shock singularities are of interest in part due to their rather mild nature, which leaves open the hope that one might be able to extend the solution uniquely past the shock, in a weak sense, under suitable selection criteria. In the case of the relativistic Euler equations and the compressible Euler equations in multiple spatial dimensions, this hope has been realized in the form of Christodoulou's recent breakthrough resolution [2019] of the *restricted shock development* problem without symmetry assumptions; see Section 1G2 for further discussion.

We now provide a very rough statement of our results; see Theorem 1.5 on page 453 for a more detailed summary and Theorem 10.1 on page 509 for the complete statements.

Theorem 1.1 (stable shock formation (very rough version)). *In an **arbitrary number** of spatial dimensions, there are many quasilinear hyperbolic PDE systems, comprising a transport equation satisfying an appropriate genuinely nonlinear-type assumption coupled to a symmetric hyperbolic subsystem, such that the following occurs: there exists an open set of initial data **without symmetry assumptions** such that the transport variable remains bounded but its first derivatives blow up in finite time. More precisely, the derivatives of the transport variable in directions **tangent** to the transport characteristics remain bounded, while its derivative with respect to any unit-length **transversal** vector field blows up. Moreover, the singularity does not propagate into the symmetric hyperbolic variables; they remain bounded, as do their first derivatives in **all** directions.*[2]

Remark 1.2 (rescaling the transversal derivative so as to "cancel" the blowup). We note already that a key part of the proof is showing the derivative of the transport variable in the transversal direction $\check{X}$ also remains bounded. This does not contradict Theorem 1.1 for the following reason: the vector field $\check{X}$ is constructed so that its Cartesian components go to 0 as the shock forms, in a manner that exactly compensates for the blowup of an "order-unity-length" transversal derivative of the transport variable. Roughly, the situation can be described as follows, where Ψ is the transport variable and the remaining quantities will be rigorously defined later in the article: $|X\Psi|$ blows up,[3] $|\check{X}\Psi|$ remains bounded, $\check{X} = \mu X$, and the weight μ vanishes for the first time at the shock; one could say that $|X\Psi|$ blows up like C/μ as $\mu \downarrow 0$, where C is the size of $|\check{X}\Psi|$ at the shock; see Section 1F4 for a more in-depth discussion of this point.

Remark 1.3 (the heart of the proof and the kind of initial data under study). The heart of the proof of Theorem 1.1 is to control the singular terms and to show that the shock actually happens, i.e., that chaotic interactions do not prevent the shock from forming or cause a more severe kind of singularity. In an effort to focus on only the singularity formation, we have chosen to study the simplest nontrivial set of initial data to which our methods apply: perturbations of the data corresponding to simple plane symmetric waves (see Section 1D for further discussion), where we assume plentiful initial Sobolev regularity.

[1] The formation of a shock is sometimes referred to as "wave breaking".

[2] Our proof allows for the possibility that the second-order Cartesian coordinate partial derivatives of the symmetric hyperbolic variables might blow up at the locations of the transport variable singularities.

[3] Here and throughout, if Z is a vector field and f is a scalar function, then $Zf := Z^\alpha \partial_\alpha f$ is the derivative of f in the direction Z.

The corresponding solutions do not experience dispersion, so there are no time or radial weights in our estimates. We will describe the initial data in more detail in Section 1F3.

Remark 1.4 (extensions to other kinds of hyperbolic subsystems). From our proof, one can infer that the assumption of symmetric hyperbolicity for the subsystem from Theorem 1.1 is in itself not important; we therefore anticipate that similar shock formation results should hold for systems comprising quasilinear transport equations coupled to many other types of hyperbolic subsystems, such as wave equations or regularly hyperbolic, in the sense of [Christodoulou 2000], subsystems.

1A. ***Paper outline.*** • In the remainder of Section 1, we give a more detailed description of our main results, summarize the main ideas behind the proofs, place our work in context by discussing prior works on shock formation, and summarize some of our notation.

• In Section 2, we precisely define the class of systems to which our main results apply.

• In Section 3, we construct the majority of the geometric objects that play a role in our analysis. We also derive evolution equations for some of the geometric quantities.

• In Section 4, we derive energy identities.

• In Section 5, we state the number of derivatives that we use to close our estimates, state our size assumptions on the data, and state bootstrap assumptions that are useful for deriving estimates.

• In Section 6, we derive pointwise estimates for solutions to the evolution equations and their derivatives, up to top order.

• In Section 7, we derive some properties of the change of variables map from geometric to Cartesian coordinates.

• In Section 8, which is the main section of the paper, we derive a priori estimates for all of the quantities under study.

• In Section 9, we provide some continuation criteria that, in the last section, we use to show that the solution survives up to the shock.

• In Section 10, we state and prove the main theorem.

1B. ***The role of nonlinear geometric optics in proving Theorem 1.1.*** In prior constructive stable shock formation results in more than one spatial dimension (which we describe in Section 1G2), the blowup occurred in the derivatives of a solution to a quasilinear wave equation. In the present work, the blowup occurs in the derivatives of the solution to the transport equation. The difference is significant in that to obtain a sharp picture of shock formation, one must rely on a geometric version of the vector field method that is precisely tailored to the family of characteristics whose intersection is tied to the blowup. The key point is that the basic regularity properties of the characteristics and the geometric vector fields (which seem essential for the proofs) that are adapted to them are different in the wave equation and transport equation cases. In fact, in the transport equation case, *the Cartesian components of the geometric vector fields are one degree less differentiable compared to the wave equation case*; see (1F.1) for the set of

geometric vector fields that we use in the present article. Although one might anticipate that the reduced regularity will lead to new complications, in the present paper, we are able to handle the loss of regularity using a strategy that, in fact, leads to simplifications compared to the wave equation case: roughly by treating the first-order *Cartesian* coordinate partial derivatives of the symmetric hyperbolic variables as new unknowns, we are able to *allow the loss of regularity* in the geometric vector fields. The fact that we can allow the loss is ultimately tied to the fact that in the present article, the variable that forms a shock is a solution to a first-order equation (in contrast to the case of wave equations). We emphasize that our approach is considerably different from, and in some ways simpler than, approaches that have been taken in proving shock formation in solutions to quasilinear wave equations, a context in which the known proofs fundamentally rely on avoiding[4] the loss of regularity; while the special structure of wave equations indeed allows one to avoid the loss of regularity in the eikonal function and the corresponding geometric vector fields, the known approaches to avoiding the loss introduce enormous technical complications into the analysis. We will discuss these fundamental points in more detail in Sections 1E and 1F5.

Although the blowup mechanism for solutions to the transport equations under study is broadly similar to the Riccati-type mechanism that drives singularity formation in the simple one-space-dimensional example of Burgers' equation[5] (see Section 1D for related discussion), the proof of our main theorem is much more complicated, owing in part to the aforementioned difficulty of having to derive energy estimates in multiple spatial dimensions. The overall strategy of our proof is to construct a system of geometric coordinates adapted to the transport characteristics, *relative to which the solution remains smooth*, in part because the geometric coordinates "hide"[6] the Riccati-type term mentioned above. In more than one spatial dimension, the philosophy of constructing geometric coordinates to regularize the problem of shock formation seems to have originated in Alinhac's work [1999a; 1999b; 2001] on quasilinear wave equations; see Section 1G2 for further discussion. As will become abundantly clear, our construction of the geometric coordinates and other related quantities is tied to the following fundamental ingredient in our approach: our development of a theory of *nonlinear geometric optics for quasilinear transport equations*, tied to an eikonal function, that is compatible with full quasilinear coupling to the symmetric hyperbolic subsystem. We use nonlinear geometric optics to construct geometric vector field differential operators (see (1F.1)) adapted to the characteristics as well as to detect the singularity formation. By "compatible", we mean, especially, from the perspective of regularity considerations. Indeed, in any situation in which one uses nonlinear geometric optics to study a quasilinear hyperbolic PDE system, *one must ensure that the regularity of the corresponding eikonal function is consistent with that of the solution*. By "full quasilinear coupling", we mean that in the systems that we study, the *principal coefficients in all equations are allowed to depend on all solution variables*.

[4]Actually, as we describe in Section 1G2, Alinhac's approach handles the loss of regularity through a Nash–Moser iteration scheme. However, Alinhac's Nash–Moser approach suffers from some technical limitations that seem to obstruct one's ability to track the behavior of the solution up to the boundary of the maximal development. In turn, this poses an obstacle to even properly setting up the shock development problem; see Section 1G2 for further discussion.

[5]The Riccati term appears after one spatial differentiation of Burgers' equation.

[6]In one spatial dimension, this is sometimes referred to as "straightening out the characteristics" via a change of coordinates.

Upon introducing nonlinear geometric optics into the problem, we encounter the following key difficulty, which we alluded to above:

> Some of the geometric vector fields that we construct (see (1F.1)) have Cartesian components that are one degree less differentiable than the transport variable, as we explain in Section 1F5.

On the one hand, due to the full quasilinear coupling, it seems that we must use the geometric vector fields when commuting the symmetric hyperbolic subsystem to obtain higher-order estimates; this allows us to avoid generating uncontrollable commutator error terms involving "bad derivatives" (i.e., in directions transversal to the transport characteristics) of the shock-forming transport variable. On the other hand, the loss of regularity of the Cartesian components of the geometric vector fields leads, at the top-order derivative level, to commutator error terms in the symmetric hyperbolic subsystem that are uncontrollable in that they have insufficient regularity. To overcome this difficulty, we employ the following strategy:

> We never commute the symmetric hyperbolic subsystem a top-order number of times with a pure string of geometric vector fields; instead, we first commute the symmetric hyperbolic subsystem with a single Cartesian coordinate partial derivative, and then follow up the Cartesian derivative with commutations by the geometric vector fields.

The strategy above allows us to avoid the loss of a derivative, but it generates commutator error terms depending on a single Cartesian coordinate partial derivative, which are dangerous because they are transversal to the transport characteristics. Indeed, the first-order Cartesian coordinate partial derivatives of the transport variable blow up at the shock. Fortunately, by using a weight[7] adapted to the characteristics, we are able to control such error terms featuring a single Cartesian differentiation, all the way up to the singularity.

We close this subsection by providing some remarks on prior implementations of nonlinear geometric optics in the study of the maximal development[8] of initial data for quasilinear hyperbolic PDEs without symmetry assumptions. The approach was pioneered by Christodoulou and Klainerman [1993] in their celebrated proof of the stability of Minkowski spacetime as a solution to the Einstein vacuum equations.[9] Since perturbative global existence results for hyperbolic PDEs typically feature estimates with "room to spare", in many cases, it is possible to close the proofs by relying on a version of *approximate* nonlinear geometric optics, which features approximate eikonal functions whose level sets approximate the characteristics. The advantage of using approximate eikonal functions is that is that their regularity theory is typically very simple. For example, such an approach was taken by Lindblad and Rodnianski [2010] in their proof of the stability of the Minkowski spacetime relative to wave coordinates. Their proof was less precise than Christodoulou and Klainerman's but significantly shorter since, unlike Christodoulou and Klainerman, Lindblad and Rodnianski relied on approximate eikonal functions whose level sets were

[7]The weight is the quantity μ from Remark 1.2, and we describe it in detail below.

[8]The maximal development is, roughly, the largest possible classical solution that is uniquely determined by the data. Readers can consult [Sbierski 2016; Wong 2013] for further discussion.

[9]Roughly, [Christodoulou and Klainerman 1993] contains a small-data global existence result for Einstein's equations.

standard Minkowski light cones; in particular, Lindblad and Rodnianski were able to close their proof using C^∞ vector fields tied to the background Minkowskian geometry.

The use of eikonal functions for proving shock formation for quasilinear wave equations in more than one spatial dimension without symmetry assumptions was pioneered by Alinhac [1999a; 1999b; 2001], and his approach was later remarkably sharpened/extended by Christodoulou [2007]. In contrast to global existence problems, in proofs of shock formation without symmetry assumptions, the *use of an eikonal function adapted to the true characteristics* (*as opposed to approximate ones*) *seems essential*, since the results yield that the singularity formation exactly coincides with the intersection of the characteristics. One can also draw an analogy between works on shock formation and works on low regularity well-posedness for quasilinear wave equations, such as [Klainerman and Rodnianski 2003; 2005; Smith and Tataru 2005; Klainerman et al. 2015], where the known proofs fundamentally rely on eikonal functions whose levels sets are true characteristics.

1C. ***A more precise statement of the main results.*** For the systems under study, we assume that the number of spatial dimensions is $n \geq 1$, where n is arbitrary. For convenience, we study the dynamics of solutions in spacetimes of the form $\mathbb{R} \times \Sigma$, where

$$\Sigma = \mathbb{R} \times \mathbb{T}^{n-1} \tag{1C.1}$$

is the spatial manifold and $\mathbb{T}^{n-1}$ is the standard $(n-1)$-dimensional torus (i.e., $[0, 1)^{n-1}$ with the endpoints identified and equipped with the usual smooth orientation). The factor $\mathbb{T}^{n-1}$ in (1C.1) will correspond to perturbations away from plane symmetry. Our assumption on the topology of Σ is for technical convenience only; since our results are localized in spacetime, one could derive similar stable blowup results for arbitrary spatial topology.[10] Throughout, $\{x^\alpha\}_{\alpha=0,\dots,n}$ are a fixed set of Cartesian spacetime coordinates on $\mathbb{R}\times\Sigma$, where $t := x^0 \in \mathbb{R}$ is the time coordinate, $\{x^i\}_{i=1,\dots,n}$ are the spatial coordinates on Σ, $x^1 \in \mathbb{R}$ is the "noncompact space coordinate", and $\{x^i\}_{i=2,\dots,n}$ are standard locally defined coordinates on $\mathbb{T}^{n-1}$ such that $(\partial_2, \dots, \partial_n)$ is a positively oriented frame. We denote the Cartesian coordinate partial derivative vector fields by $\partial_\alpha := \frac{\partial}{\partial x^\alpha}$, and we sometimes use the alternate notation $\partial_t := \partial_0$. Note that the vector fields $\{\partial_\alpha\}_{\alpha=0,\dots,n}$ can be globally defined so as to form a smooth frame, even though the $\{x^i\}_{i=2,\dots,n}$ are only locally defined. For mathematical convenience, in our main results, we consider *nearly plane symmetric solutions*, where by our conventions, exact plane symmetric solutions depend only on t and x^1. We now roughly summarize our main results; see Theorem 10.1 for precise statements.

Theorem 1.5 (stable shock formation (rough version)).
Assumptions. Consider the following coupled system[11] *with initial data posed on the constant-time hypersurface* $\Sigma_0 := \{0\} \times \mathbb{R} \times \mathbb{T}^{n-1} \simeq \mathbb{R} \times \mathbb{T}^{n-1}$:

$$L^\alpha(\Psi, v)\, \partial_\alpha \Psi = 0, \tag{1C.2a}$$

$$A^\alpha(\Psi, v)\, \partial_\alpha v = 0, \tag{1C.2b}$$

[10] However, assumptions on the data that lead to shock formation generally must be adapted to the spatial topology.

[11] Throughout we use Einstein's summation convention. Greek lowercase "spacetime" indices vary over $0, 1, \dots, n$, while Latin lowercase "spatial" indices vary over $1, 2, \dots, n$.

where Ψ is a scalar function, $v = (v^1, \dots, v^M)$ is an array (M is arbitrary), and the A^α are symmetric $M \times M$ matrices. Assume that $L^1(\Psi, v)$ satisfies a genuinely nonlinear-type condition tied to its dependence on Ψ (specifically condition (2B.1)*) and that for small Ψ and v, the constant-time hypersurfaces Σ_t and the $\mathcal{P}_u$ are* ***spacelike***[12] *for the subsystem* (1C.2b)*. Here and throughout, the $\mathcal{P}_u$ are L-characteristics, which are the family of (solution-dependent) hypersurfaces equal to the level sets of the eikonal function u, that is, the solution to the eikonal equation (see footnote 3 regarding the notation) $Lu = 0$ with the initial condition $u|_{\Sigma_0} = 1 - x^1$.*

To close the proof, we make the following assumptions on the data, which we propagate all the way up to the singularity:

> *Along Σ_0, the array v, all of its derivatives, and the $\mathcal{P}_u$-tangential derivatives of Ψ are small* ***relative***[13] *to quantities constructed out of a first-order $\mathcal{P}_u$-transversal derivative of Ψ (see Section 5D for the precise smallness assumptions, which involve geometric derivatives). Moreover, along $\mathcal{P}_0$, all derivatives of v up to top order are relatively small.*

Conclusions. *There exists an open set (relative to a suitable Sobolev topology) of data that are close to the data of a simple plane wave (where a simple plane wave is such that $\Psi = \Psi(t, x^1)$ and $v \equiv 0$), given along the unity-thickness subset Σ_0^1 of Σ_0 and a finite portion of $\mathcal{P}_0$, such that the solution behaves as follows*:

> $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$ *blows up in finite time, while $|\Psi|$, $\{|v^J|\}_{1 \le J \le M}$, and $\{|\partial_\alpha v^J|\}_{0 \le \alpha \le n, 1 \le J \le M}$ remain uniformly bounded.*

The blowup coincides with the intersection of the $\mathcal{P}_u$, which in turn is precisely characterized by the vanishing of the inverse foliation density $\mu := 1/\partial_t u$ of the $\mathcal{P}_u$, which satisfies $\mu|_{\Sigma_0^1} \approx 1$; see Figure 1 for a picture in which a shock is about to form (in the region up top, where μ is small). Moreover, one can complete (t, u) to form a geometric coordinate system $(t, u, \vartheta^2, \dots, \vartheta^n)$ on spacetime with the following key property, central to the proof:

> *No singularity occurs in Ψ, $\{v^J\}_{1 \le J \le M}$, $\{\partial_\alpha v^J\}_{0 \le \alpha \le n, 1 \le J \le M}$, or their derivatives with respect to the geometric coordinates*[14] *up to top order.*

Put differently, the problem of shock formation can be transformed into an equivalent problem in which one proves nondegenerate estimates relative to the geometric coordinates and, at the same time, proves that the geometric coordinates degenerate in a precise fashion with respect to the Cartesian coordinates as $\mu \downarrow 0$.

Remark 1.6 (nontrivial interactions all the way up to the singularity). We emphasize that in Theorem 1.5, v can be nonzero at the singularity in $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$. This means, in particular, that the problem cannot be reduced to the study of blowup for the much easier case of a decoupled scalar transport equation.

[12] This means that $A^\alpha \omega_\alpha$ is positive definite, where the one-form ω is conormal to the surface and satisfies $\omega_0 > 0$.

[13] We also assume an absolute smallness condition on $\|\Psi\|_{L^\infty(\Sigma_0)}$.

[14] In practice, we will derive estimates for the derivatives of the solution with respect to the vector fields depicted in Figure 1.

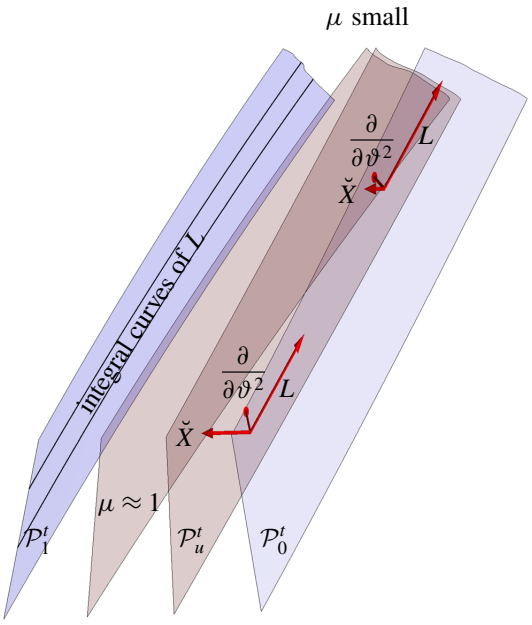

Figure 1. The dynamics until close to the time of the shock when $n = 2$.

Remark 1.7 (extensions to allow for semilinear terms). We expect that the results of Theorem 1.5 could be extended to allow for the presence of arbitrary smooth semilinear terms on the right-hand sides of (1C.2a)–(1C.2b) that are functions of (Ψ, v). The extension would be straightforward to derive for semilinear terms that vanish when $v = 0$ (for example, terms of type $v \cdot \Psi$). The reason is that our main results imply that such semilinear terms remain small, in suitable norms, up to the shock. In fact, such semilinear terms completely vanish for the exact simple waves whose perturbations we treat in Theorem 1.5; see Section 1D for further discussion of simple waves. Consequently, a set of initial data similar to the one from Theorem 1.5 would also lead to the formation of a shock in the presence of such semilinear terms. In contrast, for semilinear terms that do not vanish when $v = 0$ (for example, terms of type Ψ^2), the analysis would be more difficult and the assumptions on the data might have to be changed to produce shock-forming solutions. In particular, such semilinear terms can, at least for data with Ψ large, radically alter the behavior of some solutions. This can be seen in the simple model problem of the inhomogeneous Burgers-type equation $\partial_t \Psi + \Psi \, \partial_x \Psi = \Psi^2$. This equation admits the family of ODE-type blowup solutions $\Psi_{(\mathrm{ODE});T}(t) := (T - t)^{-1}$, whose singularity is much more severe than the shocks that typically form when the semilinear term Ψ^2 is absent.

Remark 1.8 (description of a portion of the maximal development). We expect that the approach we take in proving our main theorem is precise enough that it can be extended to yield sharp information about the behavior of the solution up the boundary of the maximal development, as Christodoulou [2007, Chapter 15] did in his related work (which we describe in more detail in Section 1G2). For brevity, we do not pursue this issue in the present article. However, in the detailed version of our main results (i.e., Theorem 10.1), we set the stage for the possible future study of the maximal development by proving a

"one-parameter family of results", indexed by $U_0 \in (0, 1]$; one would need to vary U_0 to study the maximal development. Here and throughout, U_0 corresponds to an initial data region $\Sigma_0^{U_0}$ of thickness U_0; see Figure 2 on page 472 and Section 1F2 for further discussion. For $U_0 = 1$, which is implicitly assumed in Theorem 1.5, a shock forms in the maximal development of the data given along[15] $\Sigma_0^{U_0} \cup \mathcal{P}_0$. However, for small U_0, a shock does not necessarily form in the maximal development of the data given along $\Sigma_0^{U_0} \cup \mathcal{P}_0$ within the amount of time that we attempt to control the solution.

1D. ***Further discussion on simple plane symmetric waves.*** Theorem 1.5 shows, roughly, that the well-known stable blowup of $\partial_x \Psi$ in solutions to the one-space-dimensional Burgers' equation

$$\partial_t \Psi + \Psi \, \partial_x \Psi = 0 \tag{1D.1}$$

is stable under a full quasilinear coupling of (1D.1) to other hyperbolic subsystems, under perturbations of the coefficients in the transport equation, and under increasing the number of spatial dimensions. We now further explain what we mean by this. A special case of Theorem 1.5 occurs when $v \equiv 0$ and Ψ depends only on t and x^1 (plane symmetry). In this simplified context, the blowup of $\max_{\alpha=0,1} |\partial_\alpha \Psi|$ for solutions to (1C.2a) can be proved using a simple argument based on the method of characteristics, similar to the argument that is typically used to prove blowup in the case of Burgers' equation. Solutions with $v \equiv 0$ are sometimes referred to as *simple waves* since they can be described by a single nonzero scalar component. From this perspective, we see that Theorem 1.5 yields the stability of simple plane wave shock formation for the transport variable in solutions to the system (1C.2a)–(1C.2b).

1E. ***The main new ideas behind the proof.*** The proof of Theorem 1.5 is based in part on ideas used in earlier works on shock formation in more than one spatial dimension. We review these works in Section 1G. Here we summarize the two most novel aspects behind the proof of Theorem 1.5.

• (nonlinear geometric optics for transport equations) As in all prior shock formation results in more than one spatial dimension, our proof relies on nonlinear geometric optics, that is, the eikonal function u. The use of an eikonal function is essentially the method of characteristics implemented in more than one spatial dimension. All of the prior works were such that the blowup occurred in a solution to a quasilinear wave equation and thus the theory of nonlinear geometric optics was adapted to the corresponding "wave characteristics". In this article, we advance the theory of nonlinear geometric optics for transport equations. Although the theory is simpler in some ways, compared to the case of wave equations, it is, as our prior discussion has suggested, also more degenerate in the following sense: *the regularity theory for the eikonal function u is less favorable in that u is one degree less differentiable in some directions compared to the case of wave equations*. We therefore must close the proof of Theorem 1.5 under this decreased differentiability. We defer further discussion of this point until Section 1F5. Here, we will simply further motivate our use of nonlinear geometric optics in proving shock formation.

First, we note that in more than one spatial dimension, it does not seem possible to close the proof using only the Cartesian coordinates; indeed, Theorem 1.5 shows that the blowup of Ψ precisely corresponds to

[15] Actually, we only need to specify the data along the subset $\Sigma_0^{U_0} \cup \mathcal{P}_0^{2\mathring{A}_*^{-1}}$ of $\Sigma_0^{U_0} \cup \mathcal{P}_0$; see Section 1F2 for discussion of this subset and the data-dependent parameter $\mathring{A}_*$.

the vanishing of the inverse foliation density μ of the characteristics, which is equivalent to the blowup of $\partial_t u$. Hence, it is difficult to imagine how a sharp, constructive proof of stable blowup would work without referencing an eikonal function. In view of these considerations, we construct a geometric coordinate system $(t, u, \vartheta^2, \dots, \vartheta^n)$ adapted to the transport operator vector field L and prove that Ψ, v, $V_\alpha := \partial_\alpha v$, and their geometric coordinate partial derivatives remain regular all the way up to the singularity in $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$. The blowup of $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$ occurs because the change of variables map between geometric and Cartesian coordinates *degenerates*, which is in turn tied to the vanishing of μ; the Jacobian determinant of this map is in fact proportional to μ; see Lemma 3.25. The coordinate t is the standard Cartesian time function. The geometric coordinate function u is the eikonal function described in Theorem 1.5. The initial condition $u|_{\Sigma_0} = 1 - x^1$ is adapted to the approximate plane symmetry of the initial data. We similarly construct the "geometric torus coordinates" $\{\vartheta^j\}_{j=2,\dots,n}$ by solving $L\vartheta^j = 0$ with the initial condition $\vartheta^j|_{\Sigma_0} = x^j$. The main challenge is to derive regular estimates relative to the geometric coordinates for all quantities, including the solution variables and geometric quantities constructed out of the geometric coordinates.

• (full quasilinear coupling) Because we are able to close the proof with decreased regularity for u (compared to the case of wave equations), we are able to handle full quasilinear coupling between all solution variables. This is an interesting advancement over prior works, where the principal coefficients in the evolution equation for the shock-forming variable were allowed to depend only on the shock-forming variable itself and on other solution variables that satisfy a wave equation with the *same principal part* as the shock-forming variable; i.e., in (1C.2a), we allow $L^\alpha = L^\alpha(\Psi, v)$, where the principal part of the evolution equation (1C.2b) for v is *distinct* (by assumption) from L.

1F. ***A more detailed overview of the proof.*** In this subsection, we provide an overview of the proof of our main results. Our analysis is based in part on some key ideas originating in earlier works, which we review in Section 1G. Our discussion in this subsection is, at times, somewhat loose; our rigorous analysis begins in Section 2.

1F1. *Setup and geometric constructions.* In Sections 2–3, we construct the geometric coordinate system $(t, u, \vartheta^2, \dots, \vartheta^n)$ described in Section 1B, which is central for all that follows. We also construct many related geometric objects, including the inverse foliation density μ (see Definition 3.5 for the precise definition) of the characteristics $\mathcal{P}_u$ of the eikonal function u, i.e., of the level sets of u. As we mentioned earlier, our overall strategy is to show that the solution remains regular with respect to the geometric coordinates, all the way up to the top derivative level, to show that μ vanishes in finite time, and to show that the vanishing of μ is exactly tied to the blowup of $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$. It turns out that when deriving estimates, it is important to replace the geometric coordinate partial derivative vector field $\frac{\partial}{\partial u}$ with a Σ_t-tangent vector field that we denote by $\breve{X}$, which is similar to $\frac{\partial}{\partial u}$ but generally not parallel to it; see Figure 1 on page 455 for a picture of $\breve{X}$. In the context of the present paper, the main advantage of $\breve{X}$ is that it enjoys the following key property: the vector field $X = \mu^{-1}\breve{X}$ has Cartesian components that remain uniformly bounded, all the way up to the shock. Put differently, we have $\breve{X} = \mu X$, where we will show that X is a vector field of order-unity Euclidean length (and thus the Euclidean length of $\breve{X}$ is $\mathcal{O}(\mu)$). We further explain the significance

of this in Section 1F4, when we outline the proof that the shock forms. In total, when deriving estimates for the derivatives of quantities, we differentiate them with respect to elements of the vector field frame

$$\mathscr{Z} := \{L, \check{X}, {}^{(2)}\Theta, \dots, {}^{(n)}\Theta\}, \tag{1F.1}$$

which spans the tangent space of spacetime at each point with $\mu > 0$. Here, ${}^{(i)}\Theta := \frac{\partial}{\partial \vartheta^i}$, L is the vector field from (1C.2a) and, by construction, we have $L = \frac{\partial}{\partial t}$ (see (3C.5)). The vector fields L and ${}^{(i)}\Theta$ are tangent to the $\mathcal{P}_u$, while $\check{X}$ is transversal and normalized by $\check{X}u = 1$ (see (3C.6)); see Figure 1 on page 455 for a picture of the frame. Note that since $\check{X}$ is of length $\mathcal{O}(\mu)$, the uniform boundedness of $|\check{X}\Psi|$ is consistent with the formation of a singularity in $|X\Psi|$ and thus in the Cartesian coordinate partial derivatives of Ψ when $\mu \downarrow 0$; see Section 1F4 for further discussion of this point.

We now highlight a crucial ingredient in our proof, alluded to earlier: we treat the Cartesian coordinate partial derivatives of v^J as independent unknowns V_α^J, defined by

$$V_\alpha^J := \partial_\alpha v^J. \tag{1F.2}$$

As we stressed already in Section 1B, our reliance on V_α^J *allows us to avoid commuting* (1C.2b) *up to top order with elements of* $\mathscr{Z}$, which allows us to avoid certain top-order commutator terms that would result in the loss of a derivative. Moreover, as we noted in Theorem 1.5, a key aspect of our framework is our proof that the quantities V_α remain bounded up to the singularity in $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$. To achieve this, we will control V_α by studying its evolution equation subsystem $A^\beta \, \partial_\beta V_\alpha = -(\partial_\alpha A^\beta) V_\beta$, whose inhomogeneous terms are controllable under the scope of our approach.

1F2. *A more precise description of the spacetime regions under study.* For convenience, we study only the future portion of the solution that is completely determined by the data lying in the subset $\Sigma_0^{U_0} \subset \Sigma_0$ of thickness U_0 and on a portion of the transport characteristic $\mathcal{P}_0$, where $0 < U_0 \leq 1$ is a parameter, fixed until Theorem 10.1; see Figure 2 on page 472. We will study spacetime regions such that $0 \leq u \leq U_0$, where u is the eikonal function described above. We have introduced the parameter U_0 because one would need to allow U_0 to vary in order to study the behavior of the solution up the boundary of the maximal development, as we mentioned in Remark 1.8.

In our analysis, we will use a bootstrap argument in which we only consider times t with $0 \leq t < 2\mathring{A}_*^{-1}$, where $\mathring{A}_* > 0$ is a data-dependent parameter described in Section 1F3 (see also Definition 5.1). Our main theorem shows that if $U_0 = 1$, then a shock forms at a time equal to a small perturbation of $\mathring{A}_*^{-1}$; see Section 1F4 for an outline of the proof. For this reason, in proving our main results, we only take into account the portion of the data lying in $\Sigma_0^{U_0}$ and in the subset $\mathcal{P}_0^{2\mathring{A}_*^{-1}}$ of the characteristic $\mathcal{P}_0$; from domain-of-dependence considerations, one can infer that only this portion can influence the solution in the regions under study.

Remark 1.9. For the remainder of Section 1F, we will suppress further discussion of U_0 by setting $U_0 = 1$.

1F3. *Data-size assumptions, bootstrap assumptions, and pointwise estimates.* In Section 5, we state our assumptions on the data and formulate bootstrap assumptions that are useful for deriving estimates. Our assumptions on the data involve the parameters $\mathring{\alpha} > 0$, $\mathring{\epsilon} \geq 0$, $\mathring{A} > 0$, and $\mathring{A}_* > 0$, where, for our proofs

to close, $\mathring{\alpha}$ must be chosen to be small in an absolute sense and $\mathring{\epsilon}$ must be chosen to be small in a *relative sense* compared to $\mathring{A}^{-1}$ and $\mathring{A}_*$ (see Section 5D for a precise description of the required smallness). The following remarks capture the main ideas behind the data-size parameters:

(1) $\mathring{\alpha} = \|\Psi\|_{L^\infty(\Sigma_0^1)}$ is the size of Ψ.

(2) $\mathring{\epsilon}$ is the size, in appropriate norms, of the derivatives of Ψ up to top order in which *at least one $\mathcal{P}_u$-tangential differentiation occurs*, and of v, V and *all of their derivatives* up to top order with respect to the elements of the vector field frame $\mathscr{Z}$ defined in (1F.1). We emphasize that we will study perturbations of plane symmetric shock-forming solutions such that $\mathring{\epsilon} = 0$. That is, the case $\mathring{\epsilon} = 0$ corresponds to a plane symmetric simple wave in which $v \equiv 0$. We state the total number of derivatives that we use to close the estimates in Sections 5A and 5B2. We also highlight that to close our proof, we never need to differentiate any quantity with more than one copy of the $\mathcal{P}_u$-transversal vector field $\check{X}$. This approach is possible in part because of the following crucial fact, proved in Lemma 3.22: commuting the elements of the frame $\mathscr{Z}$ with each other yields a vector field belonging to $\mathrm{span}\{{}^{(2)}\Theta, \dots, {}^{(n)}\Theta\}$.

(3) $\mathring{A} = \|\check{X}\Psi\|_{L^\infty(\Sigma_0^1)}$ is the size of the $\mathcal{P}_u$-transversal derivative of Ψ.

(4) $\mathring{A}_* = \sup_{\Sigma_0^1}[\mathcal{G}\check{X}\Psi]_-$ is a modified measure of the size of the $\mathcal{P}_u$-transversal derivative of Ψ, where $\mathcal{G} \neq 0$ is a coefficient determined by the nonlinearities and $[f]_- := |\min\{f, 0\}|$.

(5) When $t = 0$, other geometric quantities that we use in studying solutions obey similar size estimates, where any differentiation of a quantity with respect to a $\mathcal{P}_u$-tangential vector field leads to $\mathcal{O}(\mathring{\epsilon})$-smallness; see Lemma 5.5. A crucial exception occurs for $L\mu$, which initially is of relatively large size $\mathcal{O}(\mathring{A})$ in view of its evolution equation $L\mu \sim \check{X}\Psi + \cdots$ (see (3G.1a) for the precise evolution equation).

(6) The relative smallness of $\mathring{\epsilon}$ corresponds to initial data that are close to that of a simple plane symmetric wave, as we described in Section 1D.

One of the main steps in our analysis is to propagate the size assumptions above all the way up to the shock. To this end, on a region of the form $(t, u, \vartheta) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0] \times \mathbb{T}^{n-1}$, we make L^∞-type bootstrap assumptions that capture the expectation that the size assumptions stated above hold. In particular, *the bootstrap assumptions capture our expectation that no singularity will form in any quantity relative to the geometric coordinates*. Moreover, since $V_\alpha^J = \partial_\alpha v^J$, the bootstrap assumptions for the smallness[16] of V *capture our expectation that the Cartesian coordinate partial derivatives of v should remain bounded*; indeed, this is a key aspect of our proof that we use to control various error terms depending on V. As we mentioned earlier, a crucial point is that we have set the problem up so that the shock forms at time $T_{(\mathrm{Lifespan})} < 2\mathring{A}_*^{-1}$. Therefore, we make the assumption

$$0 < T_{(\mathrm{Boot})} < 2\mathring{A}_*^{-1}, \tag{1F.3}$$

which leaves us with ample margin of error to show that a shock forms. In particular, in view of (1F.3), we can bound factors of t, $\exp(t)$, etc. by a constant $C > 0$ depending on $\mathring{A}_*^{-1}$, and the estimates will

[16] We note that the bootstrap assumptions refer to a parameter $\varepsilon > 0$ that, in our main theorem, we will show is controlled by $\mathring{\epsilon}$; for brevity, we will avoid further discussion of ε until Section 5C2.

close as long as $\mathring{\epsilon}$ is sufficiently small; see Section 1H for further discussion on our conventions regarding the dependence of constants C.

In Section 6, with the help of the bootstrap assumptions and data-size assumptions described above, we commute all evolution equations, including (1C.2a)–(1C.2b) and evolution equations for μ and related geometric quantities, with elements of the $\mathscr{Z}$ up to top order and derive pointwise estimates for the error terms. Actually, due to the special structures of the equations relative to the geometric coordinates, *we never need to commute the evolution equations satisfied by v, V, or μ with the transversal vector field $\check{X}$.* Moreover, for the other geometric quantities, we need to commute their evolution equations *at most once* with $\check{X}$. We clarify, however, that we commute all equations many times with the elements of the $\mathcal{P}_u$-tangential subset $\mathscr{P} := \{L, {}^{(2)}\Theta, \dots, {}^{(n)}\Theta\}$.

1F4. *Sketch of the formation of the shock.* Let us assume that the bootstrap assumptions and pointwise estimates described in Section 1F3 hold for a sufficiently long amount of time. We will sketch how they can be used to give a simple proof of shock formation, that is, that $\mu \downarrow 0$ and $\partial\Psi$ blows up. The main estimates in this regard are provided by Lemma 6.8; here we sketch them. First, using (3G.1a), the bootstrap assumptions, and the pointwise estimates, we deduce the following evolution equation for the inverse foliation density: $L\mu(t,u,\vartheta) = [\mathcal{G}\check{X}\Psi](t,u,\vartheta) + \cdots$, where the "blowup coefficient" $\mathcal{G} \neq 0$ was described in Section 1F3 and "$\cdots$" denotes small error terms, which we ignore here. Next, we note the following pointwise estimate, which falls under the scope of the discussion in Section 1F3: $L(\mathcal{G}\check{X}\Psi) = \cdots$ (smallness is gained since L is a $\mathcal{P}_u$-tangential differentiation). Recalling that $L = \frac{\partial}{\partial t}$, we use the fundamental theorem of calculus and the smallness of $L(\mathcal{G}\check{X}\Psi)$ to deduce $[\mathcal{G}\check{X}\Psi](t,u,\vartheta) = [\mathcal{G}\check{X}\Psi](0,u,\vartheta) + \cdots$. Inserting this estimate into the one above for $L\mu$, we find that $L\mu(t,u,\vartheta) = [\mathcal{G}\check{X}\Psi](0,u,\vartheta) + \cdots$. From the fundamental theorem of calculus and the initial condition $\mu(0,u,\vartheta) = 1 + \cdots$, we obtain $\mu(t,u,\vartheta) = 1 + t[\mathcal{G}\check{X}\Psi](0,u,\vartheta) + \cdots$. From this estimate and the definition of $\mathring{A}_*$, we obtain $\min_{(u,\vartheta)\in[0,1]\times\mathbb{T}^{n-1}} \mu(t,u,\vartheta) = 1 - t\mathring{A}_* + \cdots$. Hence, μ vanishes for the first time at $T_{(\text{Lifespan})} = \mathring{A}_*^{-1} + \cdots$, as desired. Moreover, the reasoning used above can easily be extended to show that $|\check{X}\Psi|(t,u,\vartheta) \gtrsim 1$ at any point (t,u,ϑ) such that $\mu(t,u,\vartheta) < \frac{1}{4}$. Recalling that $\check{X} = \mu X$, where X has order-unity Euclidean length, we see that the following holds:

$$|X\Psi| \text{ must blow up like } C/\mu \text{ as } \mu \downarrow 0.$$

This argument shows, in particular, that the vanishing of μ *exactly coincides with the blowup of* $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$.

1F5. *Considerations of regularity.* This subsubsection is an interlude in which we highlight some issues tied to considerations of regularity. Our discussion will distinguish the problem of shock formation for transport equations from the (by now) well-understood case of quasilinear wave equations, which we further describe in Section 1G2. To illustrate the issues, we will highlight some features of our analysis, with a focus on derivative counts. In Lemma 3.21, we derive the following evolution equation for the Cartesian components of ${}^{(i)}\Theta$: $L\,{}^{(i)}\Theta^j = {}^{(i)}\Theta L^j$, where ${}^{(i)}\Theta = \frac{\partial}{\partial \vartheta^i}$. Recalling that $L = \frac{\partial}{\partial t}$, that $V_\alpha^J = \partial_\alpha v^J$, and that L^j is a smooth function of (Ψ, v), we infer, from standard energy estimates for transport equations, that ${}^{(i)}\Theta^j$ should have the same degree of Sobolev differentiability as $\partial\Psi$ and V. In

particular, we expect that ${}^{(i)}\Theta^j$ should be *one degree less differentiable* than Ψ. For similar reasons, μ, V, and some other geometric quantities that play a role in our analysis are also one degree less differentiable than Ψ. The following point is crucial for our approach:

> We are able to close the energy estimates for Ψ up to top order even though, upon commuting Ψ's transport equation, we generate error terms that depend on the "less differentiable" quantities.

That is, in controlling Ψ, we must carefully ensure that all error terms feature an allowable amount of regularity. Moreover, the same care must be taken throughout the paper, by which we mean that we must ensure that we can close the estimates for all quantities using a consistent number of derivatives. In particular, we stress that it is precisely due to considerations of the regularity of the Cartesian components of ${}^{(i)}\Theta$ and $\check{X}$ that we have introduced the quantities $V_\alpha^J = \partial_\alpha v^J$, as we explained in Section 1F1.

In the case of quasilinear wave equations with principal part $(g^{-1})^{\alpha\beta}(\Psi)\partial_\alpha\partial_\beta\Psi$, the derivative counts are different. For example, the inverse foliation density μ enjoys the *same* Sobolev regularity as the wave equation solution variable Ψ in directions tangent to the characteristics, a gain of one tangential derivative compared to the present work. Moreover, for quasilinear wave equations, a similar gain in tangential differentiability also holds for some other key geometric objects, which we will not describe here. The gain is available because certain special combinations of quantities constructed out of the eikonal function and the wave equation solution variable satisfy an unexpectedly good evolution equation, with source terms that have better-than-expected regularity; see Section 1G2 or the survey article [Holzegel et al. 2016] for further discussion. Moreover, this gain seems *essential* for closing some of the top-order energy estimates in the wave equation case, the reason being that one must commute the geometric vector fields through the *second-order* wave operator, which eats up the gain. As we explain in Section 1G2, one pays a steep price in gaining back the derivative: the resulting energy estimates allow for possible energy blowup at the high geometric derivative levels (a potential phenomenon that is related to, but distinct from, the formation of a shock), a difficulty which we do not encounter in the present work.

We close this subsubsection by again highlighting that we are able to prove shock formation for systems with full quasilinear coupling (in the sense explained in the second paragraph of Section 1B) precisely because we are able to close our estimates using geometric quantities that are one degree less differentiable than Ψ, and that the viability of allowing the loss of differentiability leads to simplifications in the proof compared to the case of quasilinear wave equations. In contrast, in the case of quasilinear wave equations, due to the apparent necessity of avoiding a loss of differentiability in various geometric quantities, it does not seem possible to prove shock formation for general systems of quasilinear wave equations with multiple propagation speeds; the special combinations of quantities mentioned in the previous paragraph, which are needed to close the geometric energy estimates in the case of quasilinear wave equations, seem to be unstable under a full quasilinear coupling of multiple speed wave systems. Here is one representative manifestation of this issue: the problem of multispace-dimensional shock formation for covariant wave equation systems (see footnote 24 on page 467 regarding the notation) of the form

$$\Box_{g_1(\Psi_1,\Psi_2)}\Psi_1 = 0, \tag{1F.4a}$$

$$\Box_{g_2(\Psi_1,\Psi_2)}\Psi_2 = 0, \tag{1F.4b}$$

where g_1 and g_2 are Lorentzian metrics,[17] is open whenever $g_1 \neq g_2$, even though shock formation for systems with $g_1 = g_2$ and for scalar equations $\Box_{g(\Psi)}\Psi = 0$ is well-understood [Speck 2016]. We note, however, that stable shock formation has been understood for some wave equation systems such that the quasilinear part of the shock-forming variable's wave equation has a decoupled structure. Specifically, in [Speck 2018], in two spatial dimensions, we proved a stable shock formation result for the variable Ψ_1 for systems in the unknowns (Ψ_1, Ψ_2) of the form

$$\Box_{g_1(\Psi_1)}\Psi_1 = \mathcal{N}_1(\Psi_1, \partial\Psi_1, \Psi_2, \partial\Psi_2), \tag{1F.5a}$$

$$(g_2^{-1})^{\alpha\beta}(\Psi_1, \Psi_2, \partial\Psi_2)\,\partial_\alpha\partial_\beta\Psi_2 = \mathcal{N}_2(\Psi_1, \partial\Psi_1, \Psi_2, \partial\Psi_2), \tag{1F.5b}$$

under appropriate assumptions on the semilinear terms $\mathcal{N}_1$ and $\mathcal{N}_2$ as well as the assumption that the wave propagation speed corresponding to g_1 is faster than the wave propagation speed corresponding to g_2, i.e., that Ψ_1 is the "fastest wave variable"; see Section 1G2 for further discussion of this result. We clarify that a key structural feature, exploited in [Speck 2018], is that in (1F.5a), the metric g_1 corresponding to the shock-forming variable Ψ_1 depends only on Ψ_1; this is tantamount to the assumption of partial decoupling of the most difficult quasilinear terms.

1F6. *Energy estimates.* In Section 8, we derive the main technical estimates of the article: energy estimates up to top order for Ψ, v, V, μ, and related geometric quantities. Energy estimates are an essential ingredient in the basic regularity theory of quasilinear hyperbolic systems in multiple spatial dimensions, and in this article, they are also important because they yield improvements of our bootstrap assumptions described in Section 1F3. We now describe the energies, which we construct in Section 4. To control the transport variable Ψ, we construct geometric energies along Σ_t. To control the symmetric hyperbolic variables v and V, we construct μ-weighted energies along Σ_t as well as *non-μ-weighted* energies along the transport characteristics $\mathcal{P}_u$. With Σ_t^u defined to be the subset of Σ_t in which the eikonal function takes on values in between 0 and u and $\mathcal{P}_u^t$ defined to be the subset of $\mathcal{P}_u$ corresponding to times between 0 and t, our energies $\mathbb{E}^{(\text{Shock})}[P\Psi](t,u), \ldots$, and our characteristic fluxes $\mathbb{E}^{(\text{Regular})}[V](t,u), \ldots$ satisfy, with $P \in \mathscr{P} = \{L, {}^{(2)}\Theta, \ldots, {}^{(n)}\Theta\}$ (see Section 4 for the details)

$$\mathbb{E}^{(\text{Shock})}[P\Psi](t,u) := \int_{\Sigma_t^u} (P\Psi)^2\, d\vartheta\, du', \tag{1F.6a}$$

$$\mathbb{E}^{(\text{Regular})}[v](t,u) \approx \int_{\Sigma_t^u} \mu|v|^2\, d\vartheta\, du', \qquad \mathbb{F}^{(\text{Regular})}[v](t,u) \approx \int_{\mathcal{P}_u^t} |v|^2\, d\vartheta\, dt', \tag{1F.6b}$$

$$\mathbb{E}^{(\text{Regular})}[V](t,u) \approx \int_{\Sigma_t^u} \mu|V|^2\, d\vartheta\, du', \quad \mathbb{F}^{(\text{Regular})}[V](t,u) \approx \int_{\mathcal{P}_u^t} |V|^2\, d\vartheta\, dt'. \tag{1F.6c}$$

In our analysis, we of course must also control various higher-order energies, but here we ignore this issue. The degenerate μ weights featured in $\mathbb{E}^{(\text{Regular})}[v]$ and $\mathbb{E}^{(\text{Regular})}[V]$ arise from expressing the standard energy for symmetric hyperbolic systems in terms of the geometric coordinates; roughly, the weight μ appears because Σ_t is transversal to the $\mathcal{P}_u$ and because dx^1 is "well-approximated by" $\mu\, du'$.

[17] That is, for $i = 1, 2$, the matrix of Cartesian components of g_i has signature $(-, +, \ldots, +)$.

For controlling certain error integrals that arise in the energy identities, *it is crucial that the characteristic fluxes* $\mathbb{F}^{(\text{Regular})}[v]$ *and* $\mathbb{F}^{(\text{Regular})}[V]$ *do not feature any degenerate* μ *weight*. These characteristic fluxes are positive definite only because our structural assumptions on the equations ensure that the propagation speed of v and V is strictly slower than that of Ψ (see (2C.1) for the precise assumptions). Readers can consult Lemma 4.2 and its proof to better understand the role of these assumptions.

We now outline the derivation of the energy estimates; see Section 8 for precise statements and proofs. Let us define[18] the controlling quantity $\mathbb{W}(t, u)$ to be the sum of the energies and characteristic fluxes in (1F.6a)–(1F.6c) and their analogs up to the top derivative level (corresponding to differentiations with respect to the geometric vector fields). The initial data that we study in our main theorem satisfy (by assumption) $\mathbb{W}(0, 1) \lesssim \mathring{\epsilon}^2$ and $\mathbb{W}(2\mathring{A}_*^{-1}, 0) \lesssim \mathring{\epsilon}^2$, where $\mathring{\epsilon}$ is the small parameter described in Section 1F3. We again stress that $\mathbb{W}(t, u) \equiv 0$ for simple plane waves.

Next, we note that energy identities, based on applying the divergence theorem on the geometric coordinate region $[0, t] \times [0, u] \times \mathbb{T}^{n-1}$, together with the pointwise estimates for error terms mentioned in Section 1F3, lead to the inequality

$$\mathbb{W}(t, u) \leq C\mathring{\epsilon}^2 + C \int_{t'=0}^{t} \int_{u'=0}^{u} \int_{\mathbb{T}^{n-1}} \{|P\Psi|^2 + |v|^2 + |V|^2\}(t', u', \vartheta)\, d\vartheta\, du'\, dt' + \cdots, \tag{1F.7}$$

where the terms "$\cdots$" depend on the geometric derivatives of Ψ, v, and V up to top order and the derivatives of various geometric quantities up to top order; the terms "$\cdots$" can be bounded using arguments similar to the ones we sketch below, so we will not discuss them further here. In view of the definition of $\mathbb{W}$, we deduce the following inequality from (1F.7):

$$\mathbb{W}(t, u) \leq C\mathring{\epsilon}^2 + C \int_{t'=0}^{t} \mathbb{W}(t', u)\, dt' + C \int_{u'=0}^{u} \mathbb{W}(t, u')\, du' + \cdots. \tag{1F.8}$$

Then from (1F.8) and Gronwall's inequality with respect to t and u, we conclude, ignoring the terms "$\cdots$" and taking into account (1F.3), that the following a priori estimate holds for $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$ (see Proposition 8.6 for the details):

$$\mathbb{W}(t, u) \lesssim \mathring{\epsilon}^2 \exp(C\mathring{A}_*^{-1}) \lesssim \mathring{\epsilon}^2. \tag{1F.9}$$

The estimate (1F.9) represents the realization of our hope that the solution remains regular relative to the geometric coordinates, up to the top derivative level.

We now stress the following key point: the characteristic fluxes $\mathbb{F}^{(\text{Regular})}[v]$ and $\mathbb{F}^{(\text{Regular})}[V]$ *are needed to control the terms* $|v|^2 + |V|^2$ *on the right-hand side of* (1F.7); without the characteristic fluxes, instead of the term $C \int_{u'=0}^{u} \mathbb{W}(t, u')\, du'$ on the right-hand side of (1F.8), we would instead have the term $C \int_{t'=0}^{t} \mathbb{W}(t', u)/(\min_{\Sigma_{t'}^u} \mu)\, dt'$, whose denominator vanishes as the shock forms. Such a term would have led to a priori estimates allowing for the possibility that at all derivative levels, the geometric energies blow up as the shock forms. This in turn would have been inconsistent with the bootstrap assumptions

[18] Our definition of $\mathbb{W}(t, u)$ given here is schematic. See Definition 8.1 for the precise definition of the controlling quantity, which we denote by $\mathbb{Q}(t, u)$.

described in Section 1F3 and would have obstructed our approach of showing that the solution remains regular relative to the geometric coordinates.

1F7. *Combining the estimates.* Once we have obtained the a priori energy estimates, we can derive improvements of our L^∞-type bootstrap assumptions via Sobolev embedding (see Corollary 8.8). These steps, together with the estimates from Section 1F4 showing that μ vanishes in finite time, are the main steps in the proof of the main theorem. We need a few additional technical results to complete the proof, including some results guaranteeing that the geometric and Cartesian coordinates are diffeomorphic up to the shock (see Section 7) and some fairly standard continuation criteria (see Section 9), which in total ensure that the solution survives up to the shock. We combine all of these results in Section 10, where we prove the main theorem.

1G. ***Connections to prior work.*** Many aspects of the approach outlined in Section 1F have their genesis in earlier works, which we now describe.

1G1. *Results in one spatial dimension.* In one spatial dimension and in symmetry classes whose PDEs are effectively one-dimensional, there are many results, by now considered classical, that use the method of characteristics to exhibit the formation of shocks in initially smooth solutions to various quasilinear hyperbolic systems. Important examples include Riemann's work [1860] (in which he developed the method of Riemann invariants), Lax's proof [1964] of stable blowup for 2×2 genuinely nonlinear systems via the method of Riemann invariants, Lax's blowup results [1972; 1973] for scalar conservation laws, John's extension [1974] of Lax's work to systems in one spatial dimension with more than two unknowns (which required the development of new ideas since the method of Riemann invariants does not apply), and the recent work of Christodoulou and Perez [2016], in which they significantly sharpened [John 1974]. The main obstacle to extending the results mentioned above to more than one spatial dimension is that one must complement the method of characteristics with an ingredient that, due to the singularity formation, is often accompanied by enormous technical complications: energy estimates that are adapted to and that hold up to the singularity. We further explain these technical complications in the next subsubsection.

1G2. *Results in more than one spatial dimension.* The first breakthrough results on shock formation in more than one spatial dimension without symmetry assumptions were proved by Alinhac [1999a; 1999b; 2001] for small-data solutions to scalar quasilinear wave equations of the form

$$(g^{-1})^{\alpha\beta}(\partial\Phi)\,\partial_\alpha\,\partial_\beta\Phi = 0 \tag{1G.1}$$

that fail to satisfy the null condition. Here, $g(\partial\Phi)$ is a Lorentzian metric equal to the Minkowski metric plus an error term of size $\mathcal{O}(\partial\Phi)$. As we do in this paper, Alinhac constructed a set of geometric coordinates tied to an eikonal function u, which in the context of his problems was a solution the fully nonlinear eikonal equation

$$(g^{-1})^{\alpha\beta}(\partial\Phi)\,\partial_\alpha u\,\partial_\beta u = 0. \tag{1G.2}$$

Much like in our work here, the level sets of u are characteristic hypersurfaces for (1G.1). They are also known, in the context of Lorentzian geometry, as *null hypersurfaces*, in view of their intimate connection

to the g-null[19] vector field $-(g^{-1})^{\alpha\beta}\,\partial_\beta u$. In his works, Alinhac identified a set of small compactly supported initial data satisfying a nondegeneracy condition such that $\max_{\alpha,\beta=0,\dots,n}|\partial_\alpha\partial_\beta\Phi|$ blows up in finite time due to the intersection of the characteristics, while $|\Phi|$ and $\max_{\alpha=0,\dots,n}|\partial_\alpha\Phi|$ remain bounded. Moreover, relative to the geometric coordinates, Φ and $\{\partial_\alpha\Phi\}_{\alpha=0,\dots,n}$ remain smooth, except possibly at the very high derivative levels (we will elaborate upon this just below).

In proving his results, Alinhac faced three serious difficulties. We will focus only on the case of three spatial dimensions, though Alinhac obtained similar results in two spatial dimensions. The first difficulty is that for small data, solutions to (1G.1) experience a long period of dispersive decay, which seems to work against the formation of a shock and which necessitated the application of Klainerman's commuting vector field method [1985; 1986] in which the vector fields have time and radial weights. We stress that such dispersive behavior is not exhibited by the solutions that we study in this article and hence our vector fields do not feature time or radial weights. Alinhac showed that after an era[20] of dispersive decay, the nonlinearity in (1G.1) takes over and drives the formation of the shock.

The second main difficulty faced by Alinhac is that to follow the solution up the singularity, it seems necessary to commute the equations with geometric vector fields constructed out of the eikonal function, and these vector fields seem to lead to the loss of a derivative when commuted through the wave operator. Specifically, the geometric vector fields Z have Cartesian components that depend on ∂u, and hence commuting them through the wave equation (1G.2) leads to an equation of the schematic form $(g^{-1})^{\alpha\beta}(\partial\Phi)\,\partial_\alpha\partial_\beta(Z\Phi) = \partial^2 Z\cdot\partial\Phi+\cdots$. The difficulty is that standard wave equation energy estimates suggest, due to the source term $\partial^2 Z$, that Φ enjoys only the same Sobolev regularity as $Z\sim\partial u$, whereas standard energy estimates for the eikonal equation (1G.2) only allow one to prove that ∂u enjoys the same Sobolev regularity as $\partial^2\Phi$; this *suggests*, misleadingly, that the approach of using vector fields constructed out of an eikonal function will lead to the loss of a derivative. To overcome this difficulty, Alinhac obtained the nonlinear solution, up to the shock, as the limit of iterates that solve singular linearized problems, and he used a rather technical Nash–Moser iteration scheme featuring a free boundary in order to recover the loss of a derivative. For technical reasons, his reliance on the Nash–Moser iteration allowed him to follow "most" small-data solutions to the constant-time hypersurface of first blowup, and not further. More precisely, his approach only allowed him to treat "nondegenerate" data such that the first singularity is *isolated* in the constant-time hypersurface of first blowup. We again emphasize that in our work here, we encounter a similar difficulty concerning the regularity of the geometric vector fields, but since our PDE systems are first-order, we are able to overcome it in a different way, in fact by *allowing* for reduced regularity in the geometric vector fields; see Sections 1B and 1F5.

The third and most challenging difficulty encountered by Alinhac is the following: when proving energy estimates relative to the geometric coordinates, it seems necessary to rely on energies that feature degenerate weights that vanish as the shock forms; the weights are direct analogs of the inverse foliation density μ from Theorem 1.5. These weights make it difficult to control certain error terms in the energy

[19]That is, if $\hat{L}^\alpha := -(g^{-1})^{\alpha\beta}\,\partial_\beta u$, then by (1G.2), we have $g(\hat{L},\hat{L}) = 0$.

[20]Roughly the era of dispersive decay lasts for a time interval of length $\exp(c/\epsilon)$, where ϵ is the size of the data in a weighted Sobolev norm.

identities, which in turn leads to a priori estimates allowing for the following possibility: as the shock forms, the high-order energies might blow up like a positive power of[21] $1/\mu$. We stress that the possible high-order energy blowup encountered by Alinhac occurs relative to the geometric coordinates and is related to — but *distinct* from — the formation of the shock singularity (in which $\max_{\alpha,\beta=0,\ldots,n} |\partial_\alpha \partial_\beta \Phi|$ blows up). To close the proof, Alinhac had to show that the possible high-order geometric energy blowup does not propagate down too far to the lower geometric derivative levels, i.e., that the solution remains fairly smooth relative to the geometric coordinates. This "descent scheme" costs many derivatives, and for this reason, the data must belong to a Sobolev space of rather high order for the estimates to close. We stress that although the energies that we use in the present paper also contain the same degenerate μ weights, we encounter different kinds of error terms in our energy estimates, tied in part to the fact that our systems are first-order and tied in part to our strategy of estimating the quantity V_α^J defined by (1F.2). For this reason, our a priori energy estimates relative to the geometric coordinates are regular in that *even the top-order geometric energies remain uniformly bounded up to the shock*.

In the remarkable work [Christodoulou 2007], Alinhac's shock formation results are significantly sharpened for the quasilinear wave equations of irrotational (i.e., vorticity-free) relativistic fluid mechanics in three spatial dimensions, which form a subclass of wave equations of type (1G.1). These wave equations arise from formulating the relativistic Euler equations in terms of a fluid potential Φ, which is possible when the vorticity vanishes. The equations studied by Christodoulou enjoy special features that he exploited in his proofs, such as having an Euler–Lagrange formulation with a Lagrangian that is invariant under the Poincaré group. The main results proved by Christodoulou are as follows: (i) there is an open (relative to a Sobolev space of high, nonexplicit order) set of small[22] data such that the only possible singularities that can form in the solution are shocks driven by the intersection of the characteristics; (ii) there is an open subset of the data from (i), not restricted by nondegeneracy assumptions of the type imposed by Alinhac, such that a shock does in fact form in finite time; and (iii) for those solutions that form shocks, Christodoulou gave a complete description of the maximal classical development of the data near the singularity, which intersects the future of the constant-time hypersurface of first blowup. His sharp description of the maximal development seems necessary for even properly setting up the *shock development problem*. This is the problem of uniquely locally continuing the solution past the singularity to the Euler equations in a *weak sense*, a setting in which one must also construct the "shock hypersurface", across which the solution jumps (the solution is smooth on either side of the shock hypersurface). The shock development problem in relativistic fluid mechanics was solved in spherical symmetry in [Christodoulou and Lisibach 2016] and, in yet another breakthrough work [Christodoulou 2019], for the nonrelativistic compressible Euler equations and the relativistic Euler equations without symmetry assumptions in a restricted case (known as the restricted shock development problem) such that the jump in entropy across the shock hypersurface was ignored. The work [Christodoulou 2019] is

[21] In the context of wave equations, μ is often defined as follows: $\mu = -1/((g^{-1})^{\alpha\beta}\, \partial_\alpha u\, \partial_\beta t)$, where t is the Cartesian time function.

[22] In the context of [Christodoulou 2007], "small" means a small perturbation of the nontrivial constant-state fluid solutions, which take the form $\Phi = kt$, where $k > 0$ is a constant.

the first of its type in more than one spatial dimension. We remark that in one spatial dimension, there are general results of this type. For example, for the existence of (weak — but unique under suitable admissibility criteria) solutions to strictly hyperbolic systems in one spatial dimension with *small total variation* (a context that allows for the presence of and interaction of "small" shock waves), we refer readers to the aforementioned work [Dafermos 2010, Chapter XV].

Compared to Alinhac's approach, the main technical improvement afforded by Christodoulou's approach [2007] to proving shock formation is that it avoids the loss of a derivative through a sharper, more direct method; instead of using Alinhac's Nash–Moser scheme, Christodoulou found special combinations of geometric quantities that satisfy good evolution equations, and he combined them with elliptic estimates on codimension-two spacelike hypersurfaces.[23] This approach to avoiding the loss of a derivative in wave equation eikonal functions originated in the aforementioned proof [Christodoulou and Klainerman 1993] of the stability of Minkowski spacetime, and it was extended by Klainerman and Rodnianski [2003] to the case of general scalar quasilinear wave equations in their study of low-regularity well-posedness for wave equations of the form $-\partial_t^2\Psi + g^{ab}(\Psi)\,\partial_a\partial_b\Psi = 0$. In total, Christodoulou's approach allowed him to control the solution up to the shock using a traditional "forwards" approach, without the free boundary found in Alinhac's iteration scheme. However, as in Alinhac's work, Christodoulou's energy estimates allowed for the possibility that the high-order energies might blow up as the shock forms. Therefore, like Alinhac, Christodoulou had to give a separate, technical argument to show that any high-order energy singularity does not propagate down too far to the lower geometric derivative levels.

In [Speck 2016], we extended Christodoulou's sharp shock formation results to the case of general quasilinear wave equations of type (1G.1) in three spatial dimensions that fail to satisfy the null condition, to the case of covariant wave equations of the type[24] $\Box_{g(\Psi)}\Psi = 0$ that fail to satisfy the null condition, and to inhomogeneous versions of these wave equations featuring "admissible" semilinear terms. Similar results were proved in [Christodoulou and Miao 2014] for a subset of these equations, namely those wave equations arising from nonrelativistic compressible fluid mechanics with vanishing vorticity. All of the results mentioned so far in this subsubsection are explained in detail in the survey article [Holzegel et al. 2016].

In the wake of the results above, there have been significant further advancements, which we now describe. In [Speck et al. 2016], we extended the shock formation results of [Speck 2016] to a new, physically relevant regime of initial conditions for wave equations in two spatial dimensions such that the solutions are close to simple outgoing plane symmetric waves, much like the setup of the present article. For the initial conditions studied in [Speck et al. 2016], the solutions do not experience dispersive decay. Hence, we used a new analytic framework to control the solution up to the shock, based on "close-to-simple-plane-wave"-type smallness assumptions on the data that are similar in spirit to the assumptions that we make on the data in the present article. The results of [Speck et al. 2016] can be viewed as an extension, to the case of quasilinear wave equations without symmetry assumptions, of the aforementioned blowup results of [Lax 1964] for 2×2 genuinely nonlinear systems, and as an extension of well-known blowup results for first-order quasilinear scalar conservation laws in an arbitrary number

[23] These codimension-two surfaces are analogs of the $(n-1)$-dimensional tori $\mathcal{T}_{t,u}$ from Definition 3.2.

[24] $\Box_g$ is the covariant wave operator of g. Relative to arbitrary coordinates, $\Box_g\Psi = (1/\sqrt{|\det g|})\,\partial_\alpha(\sqrt{|\det g|}(g^{-1})^{\alpha\beta}\,\partial_\beta\Psi)$.

of spatial dimensions; see, for example, [Dafermos 2010, Section 6.1] for a discussion of finite-time shock formation for scalar equations on $\mathbb{R}^{1+n}$ of the form $\partial_t \Phi + \sum_{a=1}^{n} \partial_a [G(\Phi)] = 0$ under appropriate assumptions on the nonlinearity G and the initial data. For special classes of wave equations in three spatial dimensions with cubic nonlinearities, Miao and Yu [2017] proved similar shock formation results for a set of large initial data featuring a single scaling parameter, similar to the short pulse ansatz exploited in the breakthrough work [Christodoulou 2009] on the formation of trapped surfaces in solutions to the Einstein vacuum equations. For the same wave equations studied in [Miao and Yu 2017], Miao [2018] recently made a related-but-distinct ansatz on the initial data and proved the existence of an open set of solutions that exist classically on the time interval $(-\infty, T_{(\text{Shock})})$ but blow up at time $T_{(\text{Shock})} \approx -1$.

All of the works mentioned above concern systems whose characteristics have a simple structure: they correspond to a single wave operator. We now describe some recent shock formation results in which the systems have more complicated principal parts, leading to multiple speeds of propagation and distinct families of characteristics. The first result of this type without symmetry assumptions was our joint work [Luk and Speck 2018] with J. Luk, which concerned the compressible Euler equations in two spatial dimensions under an arbitrary[25] barotropic[26] equation of state. Specifically, in [Luk and Speck 2018], we extended the shock formation results of [Christodoulou and Miao 2014] for the compressible Euler equations to allow for the presence of small amounts of vorticity at the location of the singularity. The vorticity satisfies a transport equation and, as it turns out, remains Lipschitz with respect to the Cartesian coordinates, all the way up to the shock. More precisely, the shock occurs in the "sound wave part" of the system rather than in the vorticity, and, as in all prior works, the shock is driven by the intersection of a family of characteristic hypersurfaces corresponding to a Lorentzian metric (known as the *acoustical metric* in the context of fluid mechanics). In particular, [Luk and Speck 2018] yielded the first proof of stable shock formation without symmetry assumptions in solutions to a hyperbolic system featuring multiple speeds, where all solution variables were allowed to interact up to the singularity.

The results proved in [Luk and Speck 2018] were based on a new wave-transport-div-curl formulation of the compressible Euler equations under a barotropic equation of state, which we derived in [Luk and Speck 2016]. The new formulation exhibits remarkable null structures and regularity properties, tied in part to the availability of elliptic estimates for the vorticity in three spatial dimensions (vorticity stretching does not occur in two spatial dimensions, and in its absence, one does not need elliptic estimates to control the vorticity). In a forthcoming work, we will extend the shock formation results of [Luk and Speck 2018] to the much more difficult case of three spatial dimensions, where to control the vorticity up to top order in a manner compatible with the wave part of the system, one must rely on the elliptic estimates, which allow one to show that the vorticity is exactly as differentiable as the velocity with respect to geometric vector fields adapted to the sound wave characteristics. In [Speck 2017], we extended the results of [Luk and Speck 2016] to allow for an arbitrary equation of state in which the pressure

[25] There is one exceptional equation of state, known as that of the Chaplygin gas, to which the results of [Luk and Speck 2018] do not apply. In one spatial dimension, the resulting PDE system is *totally linearly degenerate*, and it is widely believed that shocks do not form in (initially smooth) solutions to such systems.

[26] A barotropic equation of state is such that the pressure is a function of the density.

depends on the density and entropy. The formulation of the equations in [Speck 2017] exhibits further remarkable properties that, in our forthcoming work, we will use to prove a stable shock formation result in three spatial dimensions in which the vorticity and entropy are allowed to be nonzero at the singularity. In the work [Speck 2018] (which we mentioned at the end of Section 1F5), in two spatial dimensions, we proved the first stable shock formation result for systems of quasilinear wave equations featuring *multiple wave speeds* of propagation; i.e., the systems featured more than one distinct quasilinear wave operator. The main result yielded an open set of data such that the "fastest" wave forms a shock in finite time, while the remaining solution variables remain regular up to the singularity in the fast wave, much like in Theorem 1.5. The initial conditions were perturbations of simple plane waves, similar to the setup for the case of the scalar wave equations studied in [Speck et al. 2016] and similar to the setup of the present article. The main new difficulty that we faced in [Speck 2018] is that the geometric vector fields adapted to the shock-forming fast wave, which seem to be an essential ingredient for following the fast wave all the way to its singularity, exhibit very poor commutation properties with the slow wave operator. Indeed, commuting the geometric vector fields all the way through the slow wave operator produces error terms that are uncontrollable, both from the point of view of regularity and from the point of view of the strength of the singular commutator terms that this generates. To overcome this difficulty, we relied on a first-order reformulation of the slow wave equation which, though somewhat limiting in the precision it affords, allows us to avoid commuting all the way through the slow wave operator and hence to avoid the uncontrollable error terms.

1H. ***Notation, index conventions, and conventions for "constants".*** We now summarize some of our notation. Some of the concepts referred to here are defined later in the article. Throughout, $\{x^\alpha\}_{\alpha=0,1,\dots,n}$ denote the standard Cartesian coordinates on spacetime $\mathbb{R}\times\Sigma$, where $x^0\in\mathbb{R}$ is the time variable and $(x^1, x^2, \dots, x^n)\in\Sigma=\mathbb{R}\times\mathbb{T}^{n-1}$ are the space variables. We denote the corresponding Cartesian partial derivative vector fields by $\partial_\alpha =: \frac{\partial}{\partial x^\alpha}$ (the $\frac{\partial}{\partial x^\alpha}$ are globally defined and smooth even though $\{x^i\}_{i=2}^n$ are only locally defined) and we often use the alternate notation $t := x^0$ and $\partial_t := \partial_0$.

• Lowercase Greek spacetime indices α, β, etc. correspond to the Cartesian spacetime coordinates and vary over $0, 1, \dots, n$. Lowercase Latin spatial indices a, b, etc. correspond to the Cartesian spatial coordinates and vary over $1, 2, \dots, n$. An exception to the latter rule occurs for the geometric torus coordinate vector fields ${}^{(i)}\Theta$ from (3A.5), in which the labeling index i varies over $2, \dots, n$. Uppercase Latin indices such as J correspond to the components v^J of the array of symmetric hyperbolic variables and typically vary from 1 to M.

• We use Einstein's summation convention in that repeated indices are summed over their respective ranges.

• Unless otherwise indicated, all quantities in our estimates that are not explicitly under an integral are viewed as functions of the geometric coordinates (t, u, ϑ) of Definition 3.4. Unless otherwise indicated, quantities under integrals have the functional dependence established below in Definition 3.26.

• If Q_1 and Q_2 are two operators, then $[Q_1, Q_2] = Q_1Q_2 - Q_2Q_1$ denotes their commutator.

- $A \lesssim B$ means that there exists $C > 0$ such that $A \leq CB$.
- $A \approx B$ means that $A \lesssim B$ and $B \lesssim A$.
- $A = \mathcal{O}(B)$ means that $|A| \lesssim |B|$.
- Constants such as C and c are free to vary from line to line. *These constants, as well as implicit constants, are allowed to depend in an increasing, continuous fashion on the data-size parameters $\mathring{A}$ and $\mathring{A}_*^{-1}$ from Section 5B. However, the constants can be chosen to be independent of the parameters $\mathring{\alpha}$, $\mathring{\epsilon}$, and ε whenever the following conditions hold:* (i) *$\mathring{\epsilon}$ and ε are sufficiently small relative to* 1*, relative to $\mathring{A}^{-1}$, and relative to $\mathring{A}_*$, and* (ii) *$\mathring{\alpha}$ is sufficiently small relative to* 1, in the sense described in Section 5D.
- Constants $C_{\blacklozenge}$ are also allowed to vary from line to line, but unlike C and c, the $C_{\blacklozenge}$ are universal in that, as long as $\mathring{\alpha}$, $\mathring{\epsilon}$, and ε are sufficiently small relative to 1, they do not depend on $\mathring{\alpha}$, ε, $\mathring{\epsilon}$, $\mathring{A}$, or $\mathring{A}_*^{-1}$.
- $A = \mathcal{O}_{\blacklozenge}(B)$ means that $|A| \leq C_{\blacklozenge}|B|$, with $C_{\blacklozenge}$ as above.
- $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ respectively denote the standard floor and ceiling functions.

2. Rigorous setup of the problem and fundamental definitions

In this section, we state the equations that we will study and state our basic assumptions on the nonlinearities.

2A. ***Statement of the equations.*** Our main results concern systems in $1+n$ spacetime dimensions and $1+M$ unknowns of the form

$$L\Psi = 0, \tag{2A.1a}$$

$$A^{\alpha}\, \partial_{\alpha} v = 0, \tag{2A.1b}$$

where, in our main theorem, the scalar function Ψ forms a shock, $M \geq 1$ is an integer,[27]

$$v := (v^J)_{J=1,\ldots,M} \tag{2A.2}$$

denotes the "symmetric hyperbolic variables" (whose first-order Cartesian coordinate partial derivatives will remain bounded up to the singularity in $\max_{\alpha=0,\ldots,n} |\partial_{\alpha}\Psi|$), L is a vector field whose Cartesian components are given smooth functions of Ψ and v, that is, $L^{\alpha} = L^{\alpha}(\Psi, v)$, and the A^{α} are *symmetric* $M \times M$ matrices whose components $A_I^{\alpha;J} = A_J^{\alpha;I}$ are given smooth functions of Ψ and v. Note that (2A.1b) is equivalent to the M scalar equations $A_J^{\alpha;I}\, \partial_{\alpha} v^J = 0$, where $1 \leq I \leq M$, and we sum over the repeated occurrences of α and J. For convenience, we assume the normalization conditions

$$L^0 \equiv 1, \tag{2A.3a}$$

$$L^1|_{(\Psi,v)=(0,0)} = 1. \tag{2A.3b}$$

More generally, if we were to assume that $(L^0|_{(\Psi,v)=(0,0)}, L^1|_{(\Psi,v)=(0,0)}) \neq (0,0)$, then we could achieve (2A.3a)–(2A.3b) by performing a linear change of coordinates in the (t, x^1)-plane and then dividing (2A.1a) by a scalar.

[27] Our results also apply in the case $M = 0$, though we omit discussion of this simpler case.

As we stressed in the introduction, an essential aspect of our analysis is that we treat the Cartesian coordinate partial derivatives of v^J as independent quantities. For this reason, we define

$$V_\alpha^J := \partial_\alpha v^J, \quad V_\alpha := (V_\alpha^J)_{1\le J\le M}, \quad V := (V_\alpha^J)_{0\le\alpha\le n,1\le J\le M}. \tag{2A.4}$$

As a straightforward consequence of (2A.1b) and definition (2A.4), we obtain the following evolution equation for V_α:

$$A^\beta\,\partial_\beta V_\alpha = -(\partial_\alpha A^\beta)V_\beta. \tag{2A.5}$$

2B. ***The genuinely nonlinear-type assumption.*** Recall that we can view $L^1 = L^1(\Psi, v)$. To ensure that shocks can form in nearly plane symmetric solutions, we assume that

$$\left.\frac{\partial L^1}{\partial\Psi}\right|_{(\Psi,v)=(0,0)} \neq 0. \tag{2B.1}$$

By continuity, it follows from (2B.1) that $\frac{\partial L^1}{\partial\Psi} \neq 0$ whenever $|\Psi| + |v|$ is sufficiently small.

2C. ***Assumptions on the speed of propagation for the symmetric hyperbolic subsystem.*** In this subsection, we state our assumptions on the speed of propagation for the symmetric hyperbolic subsystem (2A.1b). Specifically, we assume that the matrices

$$A^0|_{(\Psi,v)=(0,0)} \text{ and } A^0|_{(\Psi,v)=(0,0)} - A^1|_{(\Psi,v)=(0,0)} \text{ are positive definite.} \tag{2C.1}$$

We now explain the significance of (2C.1). The positivity of $A^0|_{(\Psi,v)=(0,0)}$ ensures that for solution values near the "background state" $(\Psi, v) = (0, 0)$, the hypersurfaces Σ_t are spacelike for (2A.1b), that is, for the evolution equation satisfied by the non-shock-forming variable v. By (2A.3a), the Σ_t are also spacelike for (2A.1a); i.e., L is transversal to Σ_t. The positivity of $A^0|_{(\Psi,v)=(0,0)} - A^1|_{(\Psi,v)=(0,0)}$ will ensure that for solution values near the background state, hypersurfaces close to the flat planes $\{t - x^1 = \text{const.}\}$ are spacelike for (2A.1b). This assumption is significant because for the solutions that we will study, we will construct (in Section 3A) a family $\{\mathcal{P}_u\}_{u\in[0,1]}$ of hypersurfaces that are characteristic for (2A.1a) (that is, for the operator L) and that are close to the flat planes $\{t - x^1 = \text{const.}\}$. Put differently, the $\mathcal{P}_u$ will be characteristic for the evolution equation for Ψ but spacelike for the evolution equation for v, which essentially means that for solution values near the background state, Ψ propagates at a strictly faster speed than v (and also strictly faster than V, since the principal coefficients in the evolution equations for v and V_α are the same).

3. Geometric constructions

In this section, we define/construct most of the geometric objects that we use to analyze solutions. We defer the construction of our L^2-type energies until Section 4.

3A. ***The eikonal function and the geometric coordinates.*** In this subsection, we construct the geometric coordinates that we use to follow the solution all the way to the shock. The most important of these is the eikonal function.

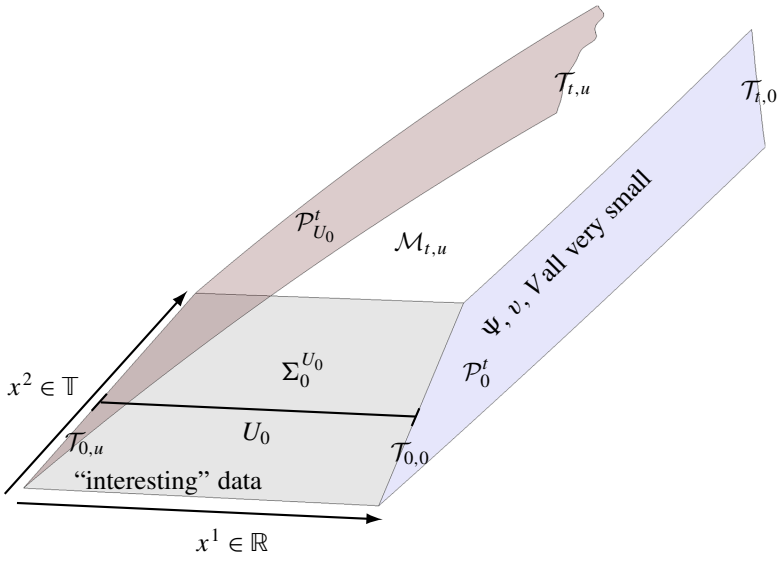

Figure 2. The spacetime region under study in the case $n = 2$.

Definition 3.1. The eikonal function is the solution u to the following transport initial value problem, where L is the transport operator vector field from (2A.1a):

$$Lu = 0, \quad u|_{\Sigma_0} = 1 - x^1. \tag{3A.1}$$

For reasons described in Remark 1.8 and Section 1F2, we now fix a real parameter U_0 satisfying

$$0 < U_0 \leq 1. \tag{3A.2}$$

We will restrict out attention to spacetime regions with $0 \leq u \leq U_0$.

Our analysis will take place on the following subsets of spacetime, which are tied to the eikonal function; see Figure 2 for a picture of the setup.

Definition 3.2. We define the following subsets of spacetime, where $x := (x^1, x^2, \ldots, x^n)$ denotes a point in $\mathbb{R} \times \mathbb{T}^{n-1}$ and (t, x) denotes a point in $\mathbb{R} \times \mathbb{R} \times \mathbb{T}^{n-1}$:

$$\Sigma_{t'} := \{(t, x) \mid t = t'\}, \tag{3A.3a}$$

$$\Sigma_{t'}^{u'} := \{(t, x) \mid t = t', \ 0 \leq u(t, x) \leq u'\}, \tag{3A.3b}$$

$$\mathcal{P}_{u'} := \{(t, x) \mid u(t, x) = u'\}, \tag{3A.3c}$$

$$\mathcal{P}_{u'}^{t'} := \{(t, x) \mid 0 \leq t \leq t', \ u(t, x) = u'\}, \tag{3A.3d}$$

$$\mathcal{T}_{t',u'} := \mathcal{P}_{u'}^{t'} \cap \Sigma_{t'}^{u'} = \{(t, x) \mid t = t', \ u(t, x) = u'\}, \tag{3A.3e}$$

$$\mathcal{M}_{t',u'} := \bigcup_{u \in [0,u']} \mathcal{P}_u^{t'} \cap \{(t, x) \mid 0 \leq t < t'\}. \tag{3A.3f}$$

We refer to the Σ_t and Σ_t^u as "constant time slices", the $\mathcal{P}_u^t$ as "characteristics", and the $\mathcal{T}_{t,u}$ as "tori". Note that $\mathcal{M}_{t,u}$ is "open-at-the-top" by construction.

To complete the geometric coordinate system, we now construct local coordinates on the tori $\mathcal{T}_{t,u}$.

Definition 3.3. We define the local geometric torus coordinates $(\vartheta^2, \dots, \vartheta^n)$ to be the solutions to the following initial value problems, where L is the transport operator vector field from (2A.1a):

$$L\vartheta^i = 0, \quad \vartheta^i|_{\Sigma_0} = x^i, \quad (i = 2, 3, \dots, n). \tag{3A.4}$$

Note that we can view $(\vartheta^2, \dots, \vartheta^n)$ as locally defined coordinates on $\mathcal{T}_{t,u} \simeq \mathbb{T}^{n-1}$.

Definition 3.4. We refer to $(t, u, \vartheta^2, \dots, \vartheta^n)$ as the geometric coordinates, and we set $\vartheta := (\vartheta^2, \dots, \vartheta^n)$. We denote the corresponding partial derivative vector fields by

$$\frac{\partial}{\partial t}, \quad \frac{\partial}{\partial u}, \quad {}^{(i)}\Theta := \frac{\partial}{\partial \vartheta^i}, \quad (i = 2, \dots, n). \tag{3A.5}$$

Note that the ${}^{(i)}\Theta$ are $\mathcal{T}_{t,u}$-tangent by construction. Moreover, we note even though the coordinate functions ϑ^i are only locally defined on $\mathcal{T}_{t,u}$, the vector fields $\{{}^{(i)}\Theta\}_{i=2,\dots,n}$ can be defined so as to form a smooth (relative to the geometric coordinates) global positively oriented frame on $\mathcal{T}_{t,u}$.

3B. ***The inverse foliation density.*** We now define $\mu > 0$, the inverse foliation density of the characteristics $\mathcal{P}_u$. When μ goes to 0, the characteristics intersect and, as our main theorem shows, $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$ blows up. That is, $\mu \downarrow 0$ *signifies the formation of a shock singularity*.

Definition 3.5 (inverse foliation density). We define $\mu > 0$ as follows:

$$\mu := \frac{1}{\partial_t u}. \tag{3B.1}$$

We observe that from (2A.3a)–(2A.3b) and (3A.1), it follows that when $|\Psi| + |v|$ is sufficiently small (as will be the case in our main theorem), we have

$$\mu|_{\Sigma_0} = 1 + \mathcal{O}_\blacklozenge(|\Psi|) + \mathcal{O}_\blacklozenge(|v|). \tag{3B.2}$$

In particular, if Ψ and v are initially small, then μ is initially close to 1.

3C. ***Vector fields and one-forms adapted to the characteristics and the blowup coefficient.*** In this subsection, we construct various vector fields and one-forms that are adapted to the characteristics $\mathcal{P}_u$. We also derive some of their basic properties. We also define the blowup coefficient $\mathcal{G}$ which, when nonzero, signifies the genuinely nonlinear nature of the transport equation (2A.1a).

Definition 3.6 (the eikonal function gradient one-forms). We define λ and ξ to be the one-forms with the following Cartesian components ($0 \le \alpha \le n$, $1 \le j \le n$):

$$\lambda_\alpha := \mu\, \partial_\alpha u, \tag{3C.1a}$$

$$\xi_0 := 0, \quad \xi_j := \mu\, \partial_j u. \tag{3C.1b}$$

Remark 3.7. From (3B.1) and (3C.1a), we deduce that

$$\lambda_0 = 1. \tag{3C.2}$$

The following definition captures the strength of the coefficient of the main term that drives the shock formation (as is evidenced by the estimates (6C.8a)–(6C.8b)). The definition is adapted to the x^1-direction since, in our main theorem, we study solutions with approximate plane symmetry (where by plane symmetric solutions, we mean ones that depend only on t and x^1).

Definition 3.8 (the blowup coefficient). Viewing $L^1 = L^1(\Psi, v)$, we define the blowup coefficient $\mathcal{G}$ as

$$\mathcal{G} := \frac{\partial L^1}{\partial \Psi} \xi_1. \tag{3C.3}$$

Remark 3.9 ($\mathcal{G} \neq 0$). The solutions that we will study will be such that ξ_1 is a small perturbation of -1; see definition (3D.3d) and the estimate (6C.7a). Hence, from (2B.1), it follows that $\mathcal{G} \neq 0$ for $|\Psi| + |v|$ sufficiently small (as will be the case for the solutions under study).

In the next definition, we define a pair of $\mathcal{P}_u$-transversal vector fields that we use to study the solution.

Definition 3.10 ($\mathcal{P}_u$-transversal vector fields). We define the Cartesian components of the Σ_t-tangent vector fields X and $\check{X}$ as follows ($1 \leq j \leq n$):

$$X^j := -L^j, \tag{3C.4a}$$

$$\check{X}^j := \mu X^j = -\mu L^j. \tag{3C.4b}$$

We now derive some basic properties of L and $\check{X}$.

Lemma 3.11 (basic properties of L and $\check{X}$). *Relative to the geometric coordinates, we have*

$$L = \frac{\partial}{\partial t}. \tag{3C.5}$$

Moreover, the following identity holds:

$$\check{X} u = 1. \tag{3C.6}$$

Finally, there exists a $\mathcal{T}_{t,u}$-tangent vector field Ξ such that

$$\check{X} = \frac{\partial}{\partial u} - \Xi. \tag{3C.7}$$

Proof. To prove (3C.5), we note that $Lu = L\vartheta^j = 0$ by construction. Also taking into account (2A.3a), we conclude (3C.5).

To prove (3C.6), we first use the eikonal equation (3A.1) and the assumption (2A.3a) to deduce the identity $\partial_t u = -L^a \partial_a u$. Multiplying this identity by μ and appealing to definition (3B.1), we deduce that $1 = -\mu L^a \partial_a u$, which, in view of definition (3C.4b), yields (3C.6). The existence of a $\mathcal{T}_{t,u}$-tangent vector field such that (3C.7) holds then follows as a simple consequence of (3C.6) and the identity $\check{X} t = 0$ (that is, the fact that $\check{X}$ is Σ_t-tangent). □

Lemma 3.12 (basic identities for the eikonal function gradient one-forms). *The following identities hold*:

$$L^\alpha \lambda_\alpha = 0, \quad L^a \xi_a = -1, \tag{3C.8a}$$

$$X^\alpha \lambda_\alpha = 1, \quad X^a \xi_a = 1. \tag{3C.8b}$$

Moreover, if Y is any $\mathcal{T}_{t,u}$-tangent vector field, then

$$Y^\alpha \lambda_\alpha = 0, \quad Y^a \xi_a = 0. \tag{3C.8c}$$

Proof. The identities in (3C.8a) are a straightforward consequence of (3A.1), definitions (3C.1a)–(3C.1b), (2A.3a), and (3C.2). The identities in (3C.8b) follow from (3C.4b), (3C.6), definitions (3C.1a)–(3C.1b), and the fact that $X^0 = 0$. To obtain (3C.8c), we first note that for $\mathcal{T}_{t,u}$-tangent vector fields Y, we have $Y \in \operatorname{span}\{{}^{(i)}\Theta\}_{i=2,\dots,n}$ and thus $Yu := Y^\alpha \partial_\alpha u = 0$. The identities in (3C.8c) follow from this fact, definitions (3C.1a)–(3C.1b), and the fact that $Y^0 = 0$. □

To obtain estimates for the solution's derivatives, we will commute the equations with the vector fields belonging to the following sets.

Definition 3.13. We define the following sets of commutation vector fields:

$$\mathscr{Z} := \{L, \check{X}, {}^{(2)}\Theta, {}^{(3)}\Theta, \dots, {}^{(n)}\Theta\}, \tag{3C.9a}$$

$$\mathscr{P} := \{L, {}^{(2)}\Theta, {}^{(3)}\Theta, \dots, {}^{(n)}\Theta\}. \tag{3C.9b}$$

Remark 3.14. Note that $\mathscr{P}$ consists of precisely the $\mathcal{P}_u$-tangent elements of $\mathscr{Z}$.

3D. ***Perturbed parts of various scalar functions.*** In this subsection, we define the perturbed parts of various scalar functions that we have constructed. The perturbed quantities, which are decorated with the subscript or superscript "Small", vanish for the background solution $(\Psi, v) = (0, 0)$.

Definition 3.15 (the perturbed parts of various scalar functions). Let L be the vector field from (2A.1a), let $\{{}^{(i)}\Theta\}_{i=2,\dots,n}$ be the geometric torus vector fields from (3A.5), and let ξ be the one-form defined in (3C.1b). We define the following "background" quantities, which are constants ($j = 1, \dots, n$):

$$\tilde{L}^j := L^j|_{(\Psi,v)=(0,0)}, \tag{3D.1a}$$

$$\widetilde{X}^j := X^j|_{(\Psi,v)=(0,0)} = -L^j|_{(\Psi,v)=(0,0)}. \tag{3D.1b}$$

In (3D.1a)–(3D.1b), we are viewing L^j and X^j to be functions of (Ψ, v) (this is possible for X^j by (3C.4a)). Note that by (2A.3b) and (3C.4a), we have

$$\tilde{L}^1 = 1, \quad \widetilde{X}^1 = -1. \tag{3D.2}$$

We also define the perturbed quantities

$$L^j_{(\text{Small})} := L^j - \tilde{L}^j, \tag{3D.3a}$$

$$X^j_{(\text{Small})} := X^j - \widetilde{X}^j = -L^j_{(\text{Small})}, \tag{3D.3b}$$

$${}^{(i)}\Theta^j_{(\text{Small})} := {}^{(i)}\Theta^j - \delta^{ij}, \tag{3D.3c}$$

$$\xi_j^{(\text{Small})} := \xi_j + \delta_j^1, \tag{3D.3d}$$

where the second equality in (3D.3b) follows from (3C.4a) and δ^{ij} and δ_j^1 are standard Kronecker deltas.

3E. ***Arrays of unknowns and schematic notation.*** We use the following arrays for convenient shorthand notation.

Definition 3.16 (shorthand notation for various solution variables). We define the following arrays γ and $\underline{\gamma}$ of scalar functions:

$$\gamma := (\Psi, v^J, V_\alpha^J, \xi_i^{(\text{Small})}, {}^{(j)}\Theta_{(\text{Small})}^k)_{0\le\alpha\le n,\, 1\le i,k\le n,\, 2\le j\le n,\, 1\le J\le M}, \tag{3E.1a}$$

$$\underline{\gamma} := (\mu, \Psi, v^J, V_\alpha^J, \xi_i^{(\text{Small})}, {}^{(j)}\Theta_{(\text{Small})}^k)_{0\le\alpha\le n,\, 1\le i,k\le n,\, 2\le j\le n,\, 1\le J\le M}. \tag{3E.1b}$$

Remark 3.17 (schematic functional dependence). In the remainder of the article, we use the notation $\mathrm{f}(s_1, s_2, \dots, s_m)$ to schematically depict an expression that depends smoothly on the scalar functions $s_1, s_2, \dots, s_m$. Note that in general, $\mathrm{f}(0) \neq 0$.

Remark 3.18 (the meaning of the symbol P). Throughout, P schematically denotes a differential operator that is tangent to the characteristics $\mathcal{P}_u$, typically L or ${}^{(i)}\Theta$. We use such notation when the precise details of P are not important.

3F. ***Cartesian partial derivatives in terms of geometric vector fields.*** In the next lemma, we expand the vector fields $\{\partial_\alpha\}_{\alpha=0,\dots,n}$ in terms of the geometric commutation vector fields.

Lemma 3.19 (Cartesian partial derivatives in terms of geometric vector fields). *There exist smooth scalar functions* $\mathrm{f}_{ij}(\gamma)$ *such that the Cartesian vector fields* ∂_α *can be expanded as follows in terms of the elements of the set* $\mathscr{Z}$ *defined in* (3C.9a) *whenever* $|\gamma|$ *is sufficiently small, where* ξ_j *is defined in* (3C.1b):

$$\partial_t = L + X, \tag{3F.1a}$$

$$\partial_j = \xi_j X + \sum_{i=2}^n \mathrm{f}_{ij}(\gamma)\, {}^{(i)}\Theta \quad (1 \le j \le n). \tag{3F.1b}$$

Proof. Equation (3F.1a) follows from (2A.3a), (3C.4a), and the fact that $X^0 = 0$.

To prove (3F.1b), we first note that for any fixed j with $1 \le j \le n$, since ∂_j is Σ_t-tangent and since $\{X, {}^{(2)}\Theta, \dots, {}^{(n)}\Theta\}$ spans the tangent space of Σ_t, there exist unique (j-dependent) scalars $\alpha_1, \dots, \alpha_n$ such that $\partial_j = \alpha_1 X + \sum_{i=2}^n \alpha_i\, {}^{(i)}\Theta$. Using both sides of this expansion to differentiate the eikonal function u and using (3C.4b) and (3C.6), we obtain the identity $\partial_j u = \alpha_1 \mu^{-1}$. In view of definition (3C.1b), we conclude that $\alpha_1 = \xi_j$, as is stated on the right-hand side of (3F.1b). Next, for $1 \le j, k \le n$, we allow both sides of the expansion to differentiate the Cartesian coordinate x^k to obtain the identity $\delta_j^k = \alpha_1 X^k + \sum_{i=2}^n \alpha_i\, {}^{(i)}\Theta^k$. For fixed j, we can view this as an identity whose left-hand side is the n-dimensional vector with components $(\delta_j^1, \dots, \delta_j^n)^\top$ and whose right-hand side is equal to the product of a matrix $M_{n\times n}$ and the n-dimensional vector $(\alpha_1, \dots, \alpha_n)^\top$, where $\top$ denotes transpose. From Definition 3.15, we see that

$$M_{n\times n} = \left(\begin{array}{c|c} -1 & \mathbf{0}_{1\times(n-1)} \\ \hline *_{(n-1)\times 1} & \mathbb{I}_{(n-1)\times(n-1)} \end{array}\right) + M_{n\times n}^{(\text{Small})},$$

where the entries of $*_{(n-1)\times 1}$ are of the schematic form $\mathrm{f}(\gamma)$ and the entries of $M^{(\mathrm{Small})}_{n\times n}$ are of the schematic form $\gamma\mathrm{f}(\gamma)$ (and thus are small when $|\gamma|$ is small). Hence, when $|\gamma|$ is small, we can invert $M_{n\times n}$ to conclude that the α_i are smooth functions of γ, which completes the proof of (3F.1b). □

3G. ***Evolution equations for the Cartesian components of various geometric quantities.*** In this subsection, we derive transport equations for the Cartesian components of various geometric quantities that are adapted to the characteristics $\mathcal{P}_u$. Later, we will use these transport equations to derive estimates for these quantities.

Lemma 3.20 (transport equations for μ, ξ_j, and $\xi_j^{(\mathrm{Small})}$). *The scalar functions μ, ξ_j, and $\xi_j^{(\mathrm{Small})}$, which are defined respectively in* (3B.1), (3C.1b), *and* (3D.3d), *satisfy the following transport equations, where the scalar functions $\mathrm{f}_{ij}(\gamma)$ are as in Lemma 3.19 ($i = 2, \dots, n,\ j = 1, \dots, n$):*

$$L\mu = (\check{X}L^a)\xi_a + \mu(LL^a)\xi_a, \tag{3G.1a}$$

$$L\xi_j = L\xi_j^{(\mathrm{Small})} = (LL^a)\xi_a\xi_j - \sum_{i=2}^{n} \mathrm{f}_{ij}(\gamma)({}^{(i)}\Theta L^a)\xi_a. \tag{3G.1b}$$

Moreover, there exist functions that are smooth whenever $|\gamma|$ is sufficiently small and that are schematically denoted by f *such that the following initial conditions hold along Σ_0:*

$$\mu|_{\Sigma_0} = 1 + (\Psi, v)\cdot \mathrm{f}(\Psi, v), \tag{3G.2a}$$

$$\xi_j|_{\Sigma_0} = \{-1 + (\Psi, v)\cdot \mathrm{f}(\Psi, v)\}\delta_j^1, \tag{3G.2b}$$

$$\xi_j^{(\mathrm{Small})}|_{\Sigma_0} = (\Psi, v)\cdot \mathrm{f}(\Psi, v)\delta_j^1. \tag{3G.2c}$$

Proof. Differentiating the eikonal equation (3A.1) with ∂_α and using (2A.3a), we obtain

$$L\,\partial_\alpha u = -(\partial_\alpha L^a)\,\partial_a u. \tag{3G.3}$$

Setting $\alpha = 0$ in (3G.3) and appealing to definition (3B.1), we deduce

$$L\mu = \mu(\partial_t L^a)(\mu\,\partial_a u). \tag{3G.4}$$

From (3G.4), (3F.1a), (3C.4b), and definition (3C.1b), we conclude (3G.1a).

Next, we set $\alpha = j$ in (3G.3), multiply the equation by μ, and use definition (3C.1b) and (3G.4) to compute that

$$\begin{aligned} L(\mu\,\partial_j u) &= -(\partial_j L^a)(\mu\,\partial_a u) + (\partial_t L^a)(\mu\,\partial_a u)(\mu\,\partial_j u) \\ &= -(\partial_j L^a)\xi_a + (\partial_t L^a)\xi_a\xi_j. \end{aligned} \tag{3G.5}$$

From (3G.5) and (3F.1a)–(3F.1b), we conclude (3G.1b).

To prove (3G.2a), we use (2A.3a)–(2A.3b), (3A.1), definition (3B.1), (3D.2), and (3D.3a) to obtain $(1/\mu)|_{\Sigma_0} = \partial_t u|_{\Sigma_0} = -L^a\,\partial_a u|_{\Sigma_0} = L^1|_{\Sigma_0} = 1 + (\Psi, v)\cdot \mathrm{f}(\Psi, v)$, from which (3G.2a) easily follows (when $|\Psi|$ and $|v|$ are small). To prove (3G.2b), we use definition (3C.1b) and the argument above to deduce that $\xi_j|_{\Sigma_0} = -(\mu\delta_j^1)|_{\Sigma_0} = \{-1 + (\Psi, v)\cdot \mathrm{f}(\Psi, v)\}\delta_j^1$, as desired. Equation (3G.2c) then follows from (3G.2b) and definition (3D.3d). □

In the next lemma, we derive transport equations for the Cartesian components of the geometric torus coordinate partial derivative vector fields.

Lemma 3.21 (transport equations for the Cartesian components of ${}^{(i)}\Theta$). *The Cartesian components* ${}^{(i)}\Theta^j$ *of the* $\mathcal{T}_{t,u}$*-tangent vector fields from* (3A.5) *and their perturbed parts* ${}^{(i)}\Theta^j_{(\text{Small})}$ *defined in* (3D.3c) *are solutions to the following transport equation initial value problem*:

$$L\,{}^{(i)}\Theta^j = {}^{(i)}\Theta L^j, \quad {}^{(i)}\Theta^j|_{\Sigma_0} = \delta^{ij}, \tag{3G.6a}$$

$$L\,{}^{(i)}\Theta^j_{(\text{Small})} = {}^{(i)}\Theta L^j, \quad {}^{(i)}\Theta^j|_{\Sigma_0} = 0, \tag{3G.6b}$$

where δ^{ij} *is the standard Kronecker delta.*

Proof. L and ${}^{(i)}\Theta$ are geometric coordinate partial derivative vector fields and they therefore commute: $[L, {}^{(i)}\Theta] = 0$. Relative to Cartesian coordinates, the vanishing commutator can be expressed as $L\,{}^{(i)}\Theta^j = {}^{(i)}\Theta L^j$, which is the desired evolution equation in (3G.6a). Next, we observe that along Σ_0, ${}^{(i)}\Theta = \partial_i$ by construction. Hence, ${}^{(i)}\Theta^j|_{\Sigma_0} = {}^{(i)}\Theta|_{\Sigma_0} x^j = \partial_i x^j = \delta^{ij}$, which yields the initial condition (3G.6a). Equation (3G.6b) then follows from definition (3D.3c) and (3G.6a). □

3H. ***Vector field commutator properties.*** In this subsection, we derive some basic properties of various vector field commutators.

Lemma 3.22. *The following vector fields are* $\mathcal{T}_{t,u}$*-tangent* ($i = 2, \dots, n$):

$$[L, \check{X}], \quad [L, {}^{(i)}\Theta], \quad [\check{X}, {}^{(i)}\Theta], \quad (i = 2, \dots, n). \tag{3H.1}$$

Moreover, *there exist smooth functions*, *denoted by subscripted versions of* f, *such that the following identities hold whenever* $|\gamma|$ *is sufficiently small* (*see Remark 3.18 regarding the notation*) ($i, i_1, i_2 = 2, \dots, n$):

$$[L, {}^{(i)}\Theta] = [{}^{(i_1)}\Theta, {}^{(i_2)}\Theta] = 0, \tag{3H.2a}$$

$$[L, \check{X}] = \sum_{i=2}^{n} \mathrm{f}_i(\underline{\gamma}, L\Psi, \check{X}\Psi)\,{}^{(i)}\Theta, \tag{3H.2b}$$

$$[\check{X}, {}^{(i)}\Theta] = \sum_{j=2}^{n} \mathrm{f}_{ij}(\underline{\gamma}, \check{X}\gamma, P\Psi, P\mu)\,{}^{(j)}\Theta. \tag{3H.2c}$$

Proof. Since (3C.5) implies that L is a geometric coordinate partial derivative vector field and since, by definition, the same is true of ${}^{(i)}\Theta$, we conclude (3H.2a).

To prove (3H.2b), we first use (3C.5), (3C.6), and the fact that $\check{X}$ is Σ_t-tangent to deduce that $[L, \check{X}]t = [L, \check{X}]u = 0$. Hence, $[L, \check{X}]$ is $\mathcal{T}_{t,u}$-tangent. Therefore, there exist unique scalars α_i such that the following identity holds for $j = 1, 2, \dots, n$: $[L, \check{X}]^j = \sum_{i=2}^n \alpha_i\,{}^{(i)}\Theta^j$. Next, we use the fact that $L^a = \mathrm{f}(\Psi, v)$, (3C.4a)–(3C.4b), and the evolution equation (3G.1a) to deduce the schematic identity $[L, \check{X}]^j = L(\mu X^j) - \check{X}L^j = \mathrm{f}(\underline{\gamma}, L^\alpha V_\alpha, \check{X}^a V_a, L\Psi, \check{X}\Psi) = \mathrm{f}(\underline{\gamma}, L\Psi, \check{X}\Psi)$. Next, considering the index range $2 \le j \le n$, we view the identity $[L, \check{X}]^j = \sum_{i=2}^n \alpha_i\,{}^{(i)}\Theta^j$ as an identity whose left-hand side is the $(n-1)$-dimensional vector with Cartesian components equal to $([L, \check{X}]^2, \dots, [L, \check{X}]^n)^\top$ and whose

right-hand side is the product of the $(n-1)\times(n-1)$ matrix $M_{(n-1)\times(n-1)} := ({}^{(i)}\Theta^j)_{i,j=2,\dots,n}$ and the $(n-1)$-dimensional vector $(\alpha_2,\dots,\alpha_n)^\top$, where $\top$ denotes transpose. From definition (3D.3c), we see that $M_{(n-1)\times(n-1)}$ is equal to the identity matrix plus an error matrix whose components are of the schematic form $\gamma\mathrm{f}(\gamma)$. In particular, $M_{(n-1)\times(n-1)}$ is invertible whenever $|\gamma|$ is sufficiently small. Hence, $(\alpha_2,\dots,\alpha_n)^\top$ is the product of a matrix, whose components are of the form $\mathrm{f}(\gamma)$ and the vector $([L,\check{X}]^2,\dots,[L,\check{X}]^n)^\top$, whose components are of the form $\mathrm{f}(\underline{\gamma}, L\Psi, \check{X}\Psi)$. This completes the proof of (3H.2b). The identity (3H.2c) can be proved in a similar fashion and we omit the details. □

Corollary 3.23 (evolution equation for Ξ^j). *There exist functions that are smooth whenever $|\gamma|$ is sufficiently small and that are schematically denoted by indexed versions of* f *such that the Cartesian components Ξ^j $(j=1,\dots,n)$ of the $\mathcal{T}_{t,u}$-tangent vector field Ξ from* (3C.7) *satisfy the evolution equation*

$$L\Xi^j = \sum_{i=2}^{n} \Xi^a \mathrm{f}_{ia}(\gamma)\,{}^{(i)}\Theta L^j - \sum_{i=2}^{n} \mathrm{f}_i(\underline{\gamma}, L\Psi, \check{X}\Psi)\,{}^{(i)}\Theta^j \tag{3H.3}$$

and the initial condition

$$\Xi^j|_{\Sigma_0} = \mathrm{f}^j(\Psi, \upsilon), \tag{3H.4}$$

where the f_{ia} *on the right-hand side of* (3H.3) *are as in* (3F.1b), *and the second sum on the right-hand side of* (3H.3) *is precisely the sum on the right-hand side of* (3H.2b).

Proof. From (3C.5) and (3C.7), we deduce that $[L,\Xi]^j = -[L,\check{X}]^j$. Considering the Cartesian components of both sides of this equation and using (3H.2b), we obtain $L\Xi^j = \Xi^a\,\partial_a L^j - \sum_{i=2}^n \mathrm{f}_i(\underline{\gamma}, L\Psi, \check{X}\Psi)\,{}^{(i)}\Theta^j$. Finally, we use (3F.1b) to substitute for ∂_a in the expression $\Xi^a\,\partial_a L^j$, and we use (3C.8c) to deduce that the component $\Xi^a\xi_a X L^j$ vanishes. In total, this yields (3H.3).

To prove (3H.4), we use (3C.7) to deduce that $\Xi^j = \Xi x^j = \frac{\partial}{\partial u}x^j - \check{X}^j$. In view of the way in which the geometric coordinates were constructed, along Σ_0, we have $\frac{\partial}{\partial u} = -\partial_1$. Moreover, in view of (3C.4a)–(3C.4b) and (3G.2a), we deduce that $\check{X}^j|_{\Sigma_0} = \check{X}x^j|_{\Sigma_0} = (\mu X^j)|_{\Sigma_0} = \mu|_{\Sigma_0}\mathrm{f}(\Psi,\upsilon) = \mathrm{f}(\Psi,\upsilon)$, where f depends on j. Combining the calculations above, we conclude (3H.4). □

3I. ***The change of variables map.*** In this subsection, we define the change of variables map from geometric to Cartesian coordinates and derive some of its basic properties.

Definition 3.24. We define $\Upsilon : \mathbb{R}\times\mathbb{R}\times\mathbb{T}^{n-1} \to \mathbb{R}\times\mathbb{R}\times\mathbb{T}^{n-1}$ to be the change of variables map from geometric to Cartesian coordinates; i.e., $\Upsilon^\alpha(t,u,\vartheta^2,\dots,\vartheta^n) = x^\alpha$.

Lemma 3.25 (basic properties of the change of variables map). *The following identities hold, where L is the vector field from* (2A.1a), *the* ${}^{(i)}\Theta$ *are the vector fields from* (3A.5), $\check{X}$ *is the vector field from* (3C.4b), *and* Ξ *is the vector field from* (3C.7):

$$\frac{\partial\Upsilon}{\partial(t,u,\vartheta^2,\dots,\vartheta^n)} := \frac{\partial(x^0,x^1,x^2,\dots,x^n)}{\partial(t,u,\vartheta^2,\dots,\vartheta^n)} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ L^1 & \mu X^1+\Xi^1 & {}^{(2)}\Theta^1 & {}^{(3)}\Theta^1 & \dots & {}^{(n)}\Theta^1 \\ L^2 & \mu X^2+\Xi^2 & {}^{(2)}\Theta^2 & {}^{(3)}\Theta^2 & \dots & {}^{(n)}\Theta^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ L^n & \mu X^n+\Xi^n & {}^{(2)}\Theta^n & {}^{(3)}\Theta^n & \dots & {}^{(n)}\Theta^n \end{pmatrix}. \tag{3I.1}$$

Moreover, there exists a smooth function of γ vanishing at $\gamma = 0$, schematically denoted by $\gamma \mathrm{f}(\gamma)$, such that

$$\det \frac{\partial(x^0, x^1, x^2, \ldots, x^n)}{\partial(t, u, \vartheta^2, \ldots, \vartheta^n)} = \frac{\partial(x^1, x^2, \ldots, x^n)}{\partial(u, \vartheta^2, \ldots, \vartheta^n)} = -\mu\{1 + \gamma \mathrm{f}(\gamma)\}. \tag{3I.2}$$

Similarly, the following identity holds:

$$\det \frac{\partial(x^2, \ldots, x^n)}{\partial(\vartheta^2, \ldots, \vartheta^n)} = 1 + \gamma \mathrm{f}(\gamma). \tag{3I.3}$$

Proof. The first column of (3I.1) is a simple consequence of (3C.5) and the fact that $Lx^\alpha = L^\alpha$. The second column of (3C.4b) follows similarly from the fact that $\breve{X}$ is Σ_t-tangent (i.e., $\breve{X}t = 0$), (3C.4b), and (3C.7). The remaining $n - 1$ columns of (3C.4b) follow similarly from the fact that the vector fields ${}^{(i)}\Theta$ are Σ_t-tangent.

The first equality in (3I.2) is a simple consequence of (3I.1). To derive the second equality in (3I.2), we first note that since $\Xi \in \mathrm{span}\{{}^{(i)}\Theta\}_{i=2,\ldots,n}$, we can delete Ξ from the matrix on the right-hand side of (3I.1) without changing its determinant. It follows that

$$\text{left-hand side of (3I.2)} = \mu \det \begin{pmatrix} X^1 & {}^{(2)}\Theta^1 & {}^{(3)}\Theta^1 & \ldots & {}^{(n)}\Theta^1 \\ X^2 & {}^{(2)}\Theta^2 & {}^{(3)}\Theta^2 & \ldots & {}^{(n)}\Theta^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X^n & {}^{(2)}\Theta^n & {}^{(3)}\Theta^n & \ldots & {}^{(n)}\Theta^n \end{pmatrix}.$$

In view of Definition 3.15 and definition (3E.1a), we see that the previous expression is equal to μ times the determinant of $M_{n\times n} + M_{n\times n}^{(\mathrm{Small})}$, where $M_{n\times n}$ and $M_{n\times n}^{(\mathrm{Small})}$ are the matrices from the proof of Lemma 3.19. Using arguments similar to the ones given in the proof of Lemma 3.19, we conclude the identity (3I.2). The identity (3I.3) can be proved via a similar argument, and we omit the details. □

3J. ***Integration forms and integrals.*** In this subsection, we define quantities connected to the two kinds of integration that we use in our analysis: integration with respect to the geometric coordinates and integration with respect to the Cartesian coordinates. In Remark 3.29, we clarify why both kinds of integration play a role in our analysis and why geometric integration is the most important for our analysis. In Lemma 3.30, we quantify the relationship between the two kinds of integration.

3J1. *Geometric integration.*

Definition 3.26 (geometric forms and related integrals). Relative to the geometric coordinates of Definition 3.4, we define the following forms:[28]

$$\begin{aligned} d\vartheta &:= d\vartheta^2 \cdots d\vartheta^n, & d\underline{\varpi} &:= d\vartheta du', \\ d\overline{\varpi} &:= d\vartheta dt', & d\varpi &:= d\vartheta du' dt'. \end{aligned} \tag{3J.1}$$

[28]Throughout the paper, we blur the distinction between the (nonnegative) integration measure $d\vartheta$ and the corresponding form $d\vartheta^2 \wedge \cdots \wedge d\vartheta^n$, and similarly for the other quantities appearing in (3J.1). The precise meaning will be clear from context.

If f is a scalar function, then we define

$$\int_{\mathcal{T}_{t,u}} f\, d\vartheta := \int_{\vartheta\in\mathbb{T}^{n-1}} f(t,u,\vartheta)\, d\vartheta, \tag{3J.2a}$$

$$\int_{\Sigma_t^u} f\, d\underline{\varpi} := \int_{u'=0}^{u}\int_{\vartheta\in\mathbb{T}^{n-1}} f(t,u',\vartheta)\, d\vartheta\, du', \tag{3J.2b}$$

$$\int_{\mathcal{P}_u^t} f\, d\overline{\varpi} := \int_{t'=0}^{t}\int_{\vartheta\in\mathbb{T}^{n-1}} f(t',u,\vartheta)\, d\vartheta\, dt', \tag{3J.2c}$$

$$\int_{\mathcal{M}_{t,u}} f\, d\varpi := \int_{t'=0}^{t}\int_{u'=0}^{u}\int_{\vartheta\in\mathbb{T}^{n-1}} f(t',u',\vartheta)\, d\vartheta\, du'\, dt'. \tag{3J.2d}$$

3J2. *Cartesian integration.*

Definition 3.27 (the one-form H). Let λ be the one-form from Definition 3.6. We define H to be the one-form with the following Cartesian components:

$$H_\nu := \frac{1}{(\delta^{\alpha\beta}\lambda_\alpha\lambda_\beta)^{1/2}}\lambda_\nu, \tag{3J.3}$$

where $\delta^{\alpha\beta}$ is the standard inverse Euclidean metric on $\mathbb{R}\times\Sigma$ (that is, $\delta^{\alpha\beta} = \mathrm{diag}(1,1,\dots,1)$ relative to the Cartesian coordinates). Note that H is the Euclidean-unit-length conormal to $\mathcal{P}_u$.

Definition 3.28 (Cartesian coordinate volume and area forms and related integrals). We define

$$d\mathcal{M} := dx^1dx^2\cdots dx^n dt, \quad d\Sigma := dx^1dx^2\cdots dx^n, \quad d\mathcal{P} \tag{3J.4}$$

to be, respectively, the standard volume form on $\mathcal{M}_{t,u}$ induced by the Euclidean metric[29] on $\mathbb{R}\times\Sigma$, the standard area form induced on Σ_t^u by the Euclidean metric on $\mathbb{R}\times\Sigma$, and the standard area form induced on $\mathcal{P}_u^t$ by the Euclidean metric on $\mathbb{R}\times\Sigma$.

We define the integrals of functions f with respect to the forms above in analogy with the way that we defined the integrals (3J.2a)–(3J.2d). For example,

$$\int_{\Sigma_t^U} f\, d\Sigma := \int_{\{(x^1,\dots,x^n)\,|\,0\le u(t,x^1,\dots,x^n)\le U\}} f(t,x^1,\dots,x^n)\, dx^1\cdots dx^n,$$

where $u(t,x^1,\dots,x^n)$ is the eikonal function.

Remark 3.29 (the role of the Cartesian forms). We never *estimate* integrals involving the Cartesian forms; before deriving estimates, we will always use Lemma 3.30 below in order to replace the Cartesian forms with the geometric ones of Definition 3.26; we use the Cartesian forms only when deriving energy *identities* relative to the Cartesian coordinates, in which the Cartesian forms naturally appear.

[29] By definition, the Euclidean metric has the components $\mathrm{diag}(1,1,\dots,1)$ relative to the standard Cartesian coordinates $(t,x^1,x^2,\dots,x^n)$ on $\mathbb{R}\times\Sigma$.

3J3. *Comparison between the Cartesian integration measures and the geometric integration measures.* In the next lemma, we quantify the relationship between the Cartesian integration measures and the geometric integration measures.

Lemma 3.30. *There exist scalar functions, schematically denoted by* $\mathrm{f}(\gamma)$*, that are smooth for* $|\gamma|$ *sufficiently small and such that the following relationship holds between the geometric integration measures corresponding to Definition 3.26 and the Cartesian integration measures corresponding to Definition 3.28, where all of the measures are nonnegative* (*see footnote 28*):

$$d\mathcal{M} = \mu\{1+\gamma \mathrm{f}(\gamma)\}\, d\varpi, \quad d\Sigma = \mu\{1+\gamma \mathrm{f}(\gamma)\}\, d\underline{\varpi}, \quad d\mathcal{P} = \{\sqrt{2}+\gamma \mathrm{f}(\gamma)\}\, d\overline{\varpi}. \tag{3J.5}$$

Proof. We prove only the identity $d\mathcal{P} = \{\sqrt{2}+\gamma \mathrm{f}(\gamma)\}\, d\overline{\varpi}$ since the other two identities in (3J.5) are a straightforward consequence of Lemma 3.25 (in particular, the Jacobian determinant[30] expressions in (3I.2)). Throughout this proof, we view $d\overline{\varpi}$ (see (3J.1)) as the n-form $dt \wedge d\vartheta^2 \wedge \cdots \wedge d\vartheta^n$ on $\mathcal{P}_u$, where $dt \wedge d\vartheta^2 = dt \otimes d\vartheta^2 - d\vartheta^2 \otimes dt$, etc. Similarly, we view $d\mathcal{P}$ as the n-form induced on $\mathcal{P}_u$ by the standard Euclidean metric on $\mathbb{R}\times\Sigma$. Then relative to Cartesian coordinates, we have $d\mathcal{P} = (dx^0 \wedge dx^1 \wedge \cdots \wedge dx^n)\cdot W$, where W is the future-directed Euclidean normal to $\mathcal{P}_u$ and $(dx^0 \wedge dx^1 \wedge \cdots \wedge dx^n)\cdot W$ denotes contraction of W against the first slot of $dx^0 \wedge dx^1 \wedge \cdots \wedge dx^n$. Note that $W^\alpha = \delta^{\alpha\beta} H_\beta$, where H_α is defined in (3J.3) and $\delta^{\alpha\beta} = \mathrm{diag}(1, 1, \ldots, 1)$ is the standard inverse Euclidean metric on $\mathbb{R}\times\Sigma$. Since $d\overline{\varpi}$ and $d\mathcal{P}$ are proportional and since $(dt \wedge d\vartheta^2 \wedge \cdots \wedge d\vartheta^n)\cdot(L \otimes {}^{(2)}\Theta \otimes \cdots \otimes {}^{(n)}\Theta) = 1$, it suffices to show that $\{\sqrt{2}+\gamma \mathrm{f}(\gamma)\} = (dx^0 \wedge dx^1 \wedge \cdots \wedge dx^n)\cdot(W \otimes L \otimes {}^{(2)}\Theta \otimes \cdots \otimes {}^{(n)}\Theta)$. To proceed, we note that $(dx^0 \wedge dx^1 \wedge \cdots \wedge dx^n)\cdot(W \otimes L \otimes {}^{(2)}\Theta \otimes \cdots \otimes {}^{(n)}\Theta)$ is equal to the determinant of the $(1+n)\times(1+n)$ matrix

$$N := \begin{pmatrix} W^0 & L^0 & 0 & \cdots & 0 \\ W^1 & L^1 & {}^{(2)}\Theta^1 & \cdots & {}^{(n)}\Theta^1 \\ \vdots & \vdots & \vdots & & \vdots \\ W^n & L^n & {}^{(2)}\Theta^n & \cdots & {}^{(n)}\Theta^n \end{pmatrix}.$$

From (2A.3a)–(2A.3b), Definition 3.6, (3C.2), Definition 3.15, definition (3E.1a), definition (3J.3), and the relation $W^\alpha = \delta^{\alpha\beta} H_\beta$, it follows that

$$N = \left(\begin{array}{cc|c} \frac{\sqrt{2}}{2} & 1 & \\ -\frac{\sqrt{2}}{2} & 1 & \mathbf{0}_{2\times(n-1)} \\ \hline *_{(n-1)\times 2} & & \mathbb{I}_{(n-1)\times(n-1)} \end{array}\right) + N^{(\mathrm{Small})},$$

where the entries of the submatrix $*_{(n-1)\times 2}$ are of the schematic form $\mathrm{f}(\gamma)$, $\mathbb{I}_{(n-1)\times(n-1)}$ is the identity matrix, and $N^{(\mathrm{Small})}$ is a matrix whose entries are all of the schematic form $\gamma \mathrm{f}(\gamma)$, where f is smooth. From these facts and the basic properties of the determinant, we conclude that $\det N = \sqrt{2}+\gamma \mathrm{f}(\gamma)$, which is the desired identity. □

[30] Note that the minus sign in (3I.2) does not appear in (3J.5) since we are viewing (3J.5) as a relationship between integration measures.

3K. ***Notation for repeated differentiation.*** In this subsection, we define some notation that we use when performing repeated differentiation.

Definition 3.31. Recall that the commutation vector field sets $\mathscr{Z}$ and $\mathscr{P}$ are defined in Definition 3.13. We label the $n+1$ vector fields in $\mathscr{Z}$ as follows: $Z_{(1)} = L$, $Z_{(2)} = {}^{(2)}\Theta$, $Z_{(3)} = {}^{(3)}\Theta$, $\dots$, $Z_{(n)} = {}^{(n)}\Theta$, $Z_{(n+1)} = \breve{X}$. Note that $\mathscr{P} = \{Z_{(1)}, Z_{(2)}, \dots, Z_{(n)}\}$. We define the following vector field operators:

- If $\vec{I} = (\iota_1, \iota_2, \dots, \iota_N)$ is a multi-index of order $|\vec{I}| := N$ with $\iota_1, \iota_2, \dots, \iota_N \in \{1, 2, \dots, n+1\}$, then $\mathscr{Z}^{\vec{I}} := Z_{(\iota_1)} Z_{(\iota_2)} \cdots Z_{(\iota_N)}$ denotes the corresponding N-th order differential operator. We write $\mathscr{Z}^N$ rather than $\mathscr{Z}^{\vec{I}}$ when we are not concerned with the structure of $\vec{I}$, and we sometimes omit the superscript when $N = 1$.
- If $\vec{I} = (\iota_1, \iota_2, \dots, \iota_N)$, then $\vec{I}_1 + \vec{I}_2 = \vec{I}$ means that $\vec{I}_1 = (\iota_{k_1}, \iota_{k_2}, \dots, \iota_{k_m})$ and $\vec{I}_2 = (\iota_{k_{m+1}}, \iota_{k_{m+2}}, \dots, \iota_{k_N})$, where $1 \le m \le N$ and $k_1, k_2, \dots, k_N$ is a permutation of $1, 2, \dots, N$.
- Sums such as $\vec{I}_1 + \vec{I}_2 + \cdots + \vec{I}_K = \vec{I}$ have an analogous meaning.
- $\mathcal{P}_u$-tangent operators such as $\mathscr{P}^{\vec{I}}$ are defined analogously, except in this case we have $\iota_1, \iota_2, \dots, \iota_N \in \{1, 2, \dots, n\}$. We write $\mathscr{P}^N$ rather than $\mathscr{P}^{\vec{I}}$ when we are not concerned with the structure of $\vec{I}$, and we sometimes omit the superscript when $N = 1$.

3L. ***Notation involving multi-indices.*** In defining our main L^2-controlling quantity (see Definition 8.1), we will refer to the following set of multi-indices.

Definition 3.32 (a set of $\mathscr{Z}$-multi-indices). We define $\mathcal{I}_*^{[1,N];1}$ to be the set of $\mathscr{Z}$ multi-indices $\vec{I}$ (in the sense of Definition 3.31) such that (i) $1 \le |\vec{I}| \le N$, (ii) $\mathscr{Z}^{\vec{I}}$ contains *at least one factor* belonging to $\mathscr{P} = \{L, {}^{(2)}\Theta, {}^{(3)}\Theta, \dots, {}^{(n)}\Theta\}$, and (iii) $\mathscr{Z}^{\vec{I}}$ contains no more than one factor of $\breve{X}$.

3M. ***Norms.*** In this subsection, we define the norms that we use in studying the solution.

Definition 3.33 (pointwise norms). We define the following pointwise norms for arrays $v = (v^J)_{1 \le J \le M}$ and $V = (V_\alpha^J)_{0 \le \alpha \le n, 1 \le J \le M}$:

$$|v| := \sum_{J=1}^{M} |v^J|, \quad |V_\alpha| := \sum_{J=1}^{M} |V_\alpha^J|, \quad |V| := \sum_{J=1}^{M} \sum_{\alpha=0}^{n} |V_\alpha^J|. \tag{3M.1}$$

We will use the following L^2 and L^∞ norms in our analysis.

Definition 3.34 (L^2 and L^∞ norms). In terms of the geometric forms of Definition 3.26, we define the following norms for scalar or array-valued functions w:

$$\|w\|_{L^2(\mathcal{T}_{t,u})}^2 := \int_{\mathcal{T}_{t,u}} |w|^2 \, d\vartheta, \quad \|w\|_{L^2(\Sigma_t^u)}^2 := \int_{\Sigma_t^u} |w|^2 \, d\underline{\varpi}, \quad \|w\|_{L^2(\mathcal{P}_u^t)}^2 := \int_{\mathcal{P}_u^t} |w|^2 \, d\overline{\varpi}, \tag{3M.2a}$$

$$\begin{aligned} \|w\|_{L^\infty(\mathcal{T}_{t,u})} &:= \operatorname*{ess\,sup}_{\vartheta \in \mathbb{T}^{n-1}} |w|(t, u, \vartheta), \\ \|w\|_{L^\infty(\Sigma_t^u)} &:= \operatorname*{ess\,sup}_{(u', \vartheta) \in [0,u] \times \mathbb{T}^{n-1}} |w|(t, u', \vartheta), \\ \|w\|_{L^\infty(\mathcal{P}_u^t)} &:= \operatorname*{ess\,sup}_{(t', \vartheta) \in [0,t] \times \mathbb{T}^{n-1}} |w|(t', u, \vartheta). \end{aligned} \tag{3M.2b}$$

3N. ***Strings of commutation vector fields and vector field seminorms.*** We will use the following shorthand notation to capture the relevant structure of our vector field differential operators and to schematically depict estimates.

Definition 3.35. • $\mathscr{Z}^{N;1} f$ denotes an arbitrary string of N commutation vector fields in $\mathscr{Z}$ (see (3C.9a)) applied to f, where the string contains *at most* one factor of the $\mathcal{P}_u^t$-transversal vector field $\check{X}$. We sometimes write Zf instead of $\mathscr{Z}^{1;1} f$.

• $\mathscr{P}^N f$ denotes an arbitrary string of N commutation vector fields in $\mathscr{P}$ (see (3C.9b)) applied to f. Consistent with Remark 3.18, we sometimes write Pf instead of $\mathscr{P}^1 f$.

• For $N \geq 1$, $\mathscr{Z}_*^{N;1} f$ denotes an arbitrary string of N commutation vector fields in $\mathscr{Z}$ applied to f, where the string contains *at least* one $\mathcal{P}_u$-tangent factor and *at most* one factor of $\check{X}$. We also set $\mathscr{Z}_*^{0;0} f := f$.

• For $N \geq 1$, $\mathscr{P}_*^N f$ denotes an arbitrary string of N commutation vector fields in $\mathscr{P}$ applied to f, where the string contains *at least one factor* belonging to the geometric torus coordinate partial derivative vector field set $\{{}^{(2)}\Theta, {}^{(3)}\Theta, \dots, {}^{(n)}\Theta\}$ or *at least two factors* of L.

Remark 3.36 (another way to think about operators $\mathscr{P}_*^N$). For exact simple plane wave solutions, if $N \geq 1$ and f is *any* of the quantities that we must estimate, then we have $\mathscr{P}_*^N f \equiv 0$.

We also define seminorms constructed out of sums of the strings of vector fields above:

• $|\mathscr{Z}^{N;1} f|$ simply denotes the magnitude of one of the $\mathscr{Z}^{N;1} f$ as defined above (there is no summation).

• $|\mathscr{Z}^{\leq N;1} f|$ is the *sum* over all terms of the form $|\mathscr{Z}^{N';1} f|$ with $N' \leq N$ and $\mathscr{Z}^{N';1} f$ as defined above. We sometimes write $|\mathscr{Z}^{\leq 1} f|$ instead of $|\mathscr{Z}^{\leq 1;1} f|$.

• $|\mathscr{Z}^{[1,N];1} f|$ is the sum over all terms of the form $|\mathscr{Z}^{N';1} f|$ with $1 \leq N' \leq N$ and $\mathscr{Z}^{N';1} f$ as defined above.

• Sums such as $|\mathscr{P}^{\leq N} f|$, $|\mathscr{P}_*^{[1,N]} f|$, etc. are defined analogously.

• Seminorms such as $\|\mathscr{Z}_*^{[1,N];1} f\|_{L^\infty(\Sigma_t^u)}$ and $\|\mathscr{P}_*^{[1,N]} f\|_{L^\infty(\Sigma_t^u)}$ (see (3M.2)) are defined analogously.

Remark 3.37. In our forthcoming estimates, terms that do not make sense are assumed to be absent. For example in the case $N = 1$, all terms on the right-hand side of (6B.3) are absent except for the term $|\mathscr{P}^{\leq N-1} V|$.

Remark 3.38 (remarks on the symbol "$*$"). Some operators in Definition 3.35 are decorated with a $*$. These operators involve $\mathcal{P}_u$-tangent differentiations that often lead to a gain in smallness in the estimates. More precisely, the operators $\mathscr{P}_*^N$ always lead to a gain in smallness, while the operators $\mathscr{Z}_*^{N;1}$ lead to a gain in smallness except perhaps when they are applied to μ (because $L\mu$ is not generally small).

4. Energy identities

In this section, we define the building block energies and characteristic fluxes that we use to control the solution in L^2 and derive their basic coerciveness properties. We then derive energy identities involving the building blocks. Later in the article, in Definition 8.1, we will use the building blocks to define the main L^2-controlling quantity.

4A. ***Energies and characteristic flux definitions.***

Definition 4.1 (energies and characteristics fluxes). In terms of the geometric forms of Definition 3.26, we define the energy $\mathbb{E}^{(\mathrm{Shock})}[\,\cdot\,]$, which is a functional of scalar-valued functions f, as

$$\mathbb{E}^{(\mathrm{Shock})}[f](t,u) := \int_{\Sigma_t^u} f^2 \, d\underline{\varpi}. \tag{4A.1}$$

In terms of the Cartesian forms of Definition 3.28 and the Euclidean-unit-length one-form H_α defined in (3J.3), we define the energy $\mathbb{E}^{(\mathrm{Regular})}[\,\cdot\,]$ and characteristic flux $\mathbb{F}^{(\mathrm{Regular})}[\,\cdot\,]$, which are functionals of $\mathbb{R}^M$-valued functions w, as

$$\mathbb{E}^{(\mathrm{Regular})}[w](t,u) := \int_{\Sigma_t^u} \delta_{JK} A_I^{0;J}(\Psi, v) w^I w^K \, d\Sigma, \tag{4A.2a}$$

$$\mathbb{F}^{(\mathrm{Regular})}[w](t,u) := \int_{\mathcal{P}_u^t} \delta_{JK} A_I^{\alpha;J}(\Psi, v) H_\alpha w^I w^K \, d\mathcal{P}, \tag{4A.2b}$$

where δ_{JK} is the standard Kronecker delta.

Lemma 4.2 (coerciveness of the energies and characteristic fluxes for the symmetric hyperbolic variables). *If $|\gamma|$ is sufficiently small, then the energy and the characteristic flux from Definition 4.1 enjoy the following coerciveness*:

$$\mathbb{E}^{(\mathrm{Regular})}[w](t,u) \approx \int_{\Sigma_t^u} \mu \delta_{JK} w^J w^K \, d\underline{\varpi}, \tag{4A.3a}$$

$$\mathbb{F}^{(\mathrm{Regular})}[w](t,u) \approx \int_{\mathcal{P}_u^t} \delta_{JK} w^J w^K \, d\overline{\varpi}, \tag{4A.3b}$$

where δ_{JK} is the standard Kronecker delta.

Proof. From the arguments given in the proof of Lemma 3.30, it follows that the one-form H_α defined in (3J.3) can be decomposed as $H_\alpha = \delta_\alpha^0 - \delta_\alpha^1 + H_\alpha^{(\mathrm{Small})}$, where $H_\alpha^{(\mathrm{Small})} = \gamma \mathrm{f}(\gamma)$. Hence, from (2C.1), it follows that when $|\gamma|$ is sufficiently small, we have $\delta_{JK} A_I^{0;J} w^I w^K \approx \delta_{JK} w^J w^K$ and $\delta_{JK} A_I^{\alpha;J} H_\alpha w^I w^K \approx \delta_{JK} w^J w^K$. Appealing to definitions (4A.2a)–(4A.2b) and using the integration measure relationships stated in (3J.5), we conclude (4A.3a)–(4A.3b). □

4B. ***Energy-characteristic flux identities.*** The integral identities in the following proposition form the starting point for our L^2 analysis of solutions. A crucial point is that the left-hand side of (4B.4) features the characteristic flux $\mathbb{F}^{(\mathrm{Regular})}[\,\cdot\,](t,u)$, which by (4A.3b) can be used to control v and V on the characteristic hypersurfaces $\mathcal{P}_u^t$ *without any degenerate μ weight*.

Proposition 4.3 (energy-characteristic flux identities). *Let $L = L^\alpha(\Psi, v)\,\partial_\alpha$ be the vector field from* (2A.1a) *and let f be a solution to the inhomogeneous transport equation*

$$Lf = \mathfrak{F}. \tag{4B.1}$$

Then the following integral identity holds for the energy defined in (4A.1):

$$\mathbb{E}^{(\mathrm{Shock})}[f](t,u) = \mathbb{E}^{(\mathrm{Shock})}[f](0,u) + 2\int_{\mathcal{M}_{t,u}} f\mathfrak{F} \, d\varpi. \tag{4B.2}$$

Moreover, let $A_J^{\alpha;I}(\Psi, v)$ *be the components of the symmetric matrices from* (2A.1b) *and let* w *be a solution to the* (*linear-in-*w) *inhomogeneous symmetric hyperbolic system*

$$\mu A_J^{\alpha;I}\, \partial_\alpha w^J = \mathfrak{F}^I. \tag{4B.3}$$

Then there exist smooth functions, schematically denoted by f, *such that the following integral identity holds for the energy and characteristic flux defined in* (4A.2a)–(4A.2b):

$$\begin{aligned}\mathbb{E}^{(\mathrm{Regular})}[w](t,u) + \mathbb{F}^{(\mathrm{Regular})}[w](t,u) &= \mathbb{E}^{(\mathrm{Regular})}[w](0,u) + \mathbb{F}^{(\mathrm{Regular})}[w](t,0) \\ &\quad + 2\int_{\mathcal{M}_{t,u}} \{1+\gamma \mathrm{f}(\gamma)\}\, \delta_{JK} \mathfrak{F}^J w^K \, d\varpi \\ &\quad + \int_{\mathcal{M}_{t,u}} \mathrm{f}_{JK}(\underline{\gamma}, \check{X}\Psi, P\Psi) w^J w^k \, d\varpi, \end{aligned} \tag{4B.4}$$

where δ_{JK} *is the standard Kronecker delta.*

Proof. The identity (4B.2) is a simple consequence of (4B.1) since $L = \frac{\partial}{\partial t}$ relative to the geometric coordinates (t, u, ϑ).

To prove (4B.4), we define the following vector field relative to the Cartesian coordinates: $\mathscr{J}^\alpha := \delta_{JK} A_I^{\alpha;J} w^I w^K$. Using (4B.3) and the symmetry assumption $A_J^{\alpha;I} = A_I^{\alpha;J}$, we derive (relative to the Cartesian coordinates) the following divergence identity: $\mu\, \partial_\alpha \mathscr{J}^\alpha = 2\delta_{JK}\mathfrak{F}^J w^K + \delta_{JK}(\mu\, \partial_\alpha A_I^{\alpha;J}) w^I w^K$. We now apply the divergence theorem to the vector field $\mathscr{J}$ on the region $\mathcal{M}_{t,u}$, where we use the Cartesian coordinates, the Euclidean metric $\delta^{\alpha\beta} := \mathrm{diag}(1, 1, \ldots, 1)$ on $\mathbb{R}\times\Sigma$, and the Cartesian forms of Definition 3.28 in all computations. Also using that the future-directed Euclidean conormal to Σ_t has Cartesian components δ_α^0 and that the future-directed Euclidean conormal to $\mathcal{P}_u^t$ has Cartesian components H_α (see Definition 3.27), we deduce

$$\begin{aligned}&\int_{\Sigma_t^u} \delta_{JK} A_I^{0;J} w^I w^K \, d\Sigma + \int_{\mathcal{P}_u^t} \delta_{JK} A_I^{\alpha;J} H_\alpha w^I w^K \, d\mathcal{P} \\ &\quad = \int_{\Sigma_0^u} \delta_{JK} A_I^{0;J} w^I w^K \, d\Sigma + \int_{\mathcal{P}_0^t} \delta_{JK} A_I^{\alpha;J} H_\alpha w^I w^K \, d\mathcal{P} \\ &\qquad + \int_{\mathcal{M}_{t,u}} \{2\delta_{JK}\mathfrak{F}^J w^K + \delta_{JK}(\mu\, \partial_\alpha A_I^{\alpha;J}) w^I w^K\} \frac{d\mathcal{M}}{\mu}. \end{aligned} \tag{4B.5}$$

Next, using Lemma 3.19 and definition (3E.1b), we can express the integrand $\delta_{JK}(\mu\, \partial_\alpha A_I^{\alpha;J}) w^I w^K$ on the right-hand side of (4B.5) in the following schematic form: $\mathrm{f}_{JK}(\underline{\gamma}, \check{X}\Psi, P\Psi) w^J w^k$. Also using Lemma 3.30 to express the integration measure $d\mathcal{M}/\mu$ on the right-hand side of (4B.5) as $\{1+\gamma \mathrm{f}(\gamma)\}\, d\varpi$ and appealing to definitions (4A.2a)–(4A.2b), we arrive at the desired identity (4B.4). □

5. The number of derivatives, data-size assumptions, bootstrap assumptions, smallness assumptions, and running assumptions

In this section, we state the number of derivatives that we use to close the forthcoming estimates, state our assumptions on the size of the data, formulate bootstrap assumptions that we use to derive estimates,

and describe our smallness assumptions. In Section 5E, we explain why there exist data that satisfy the assumptions.

5A. ***The number of derivatives.*** Throughout the rest of the paper, N_{Top} and N_{Mid} denote two fixed positive integers satisfying the following relations, where n is the number of spatial dimensions:

$$N_{\mathrm{Top}} \geq n+5, \quad N_{\mathrm{Mid}} := \lceil \tfrac{1}{2} N_{\mathrm{Top}} \rceil + 1. \tag{5A.1}$$

The solutions that we will study are such that, roughly, the order $\leq N_{\mathrm{Top}}$ derivatives of Ψ (with respect to suitable strings of geometric vector fields) are uniformly bounded in the norm $\|\cdot\|_{L^2(\Sigma_t^u)}$ and the order $\leq N_{\mathrm{Mid}}$ derivatives of Ψ are uniformly bounded in the norm $\|\cdot\|_{L^\infty(\Sigma_t^u)}$. The remaining quantities that we must estimate satisfy similar bounds but, in some cases, they are one degree less differentiable. The definitions in (5A.1) are convenient in the sense that they will lead to the following: when we derive L^2 estimates for error term products in the commuted equations, all factors in the product except at most one will be uniformly bounded in the norm $\|\cdot\|_{L^\infty(\Sigma_t^u)}$.

5B. ***Data-size assumptions.*** In this subsection, we state our assumptions on the size of the data.

5B1. *The data-size parameter that controls the time of shock formation.* We start with the definition of a data-size parameter $\mathring{A}_*$, which is tied to the time of first shock formation. More precisely, our main theorem shows that $\max_{\alpha=0,\dots,n} |\partial_\alpha \Psi|$ blows up at a time approximately equal to $\mathring{A}_*^{-1}$.

Definition 5.1 (the crucial quantity that controls the time of shock formation). We define $\mathring{A}_*$ as

$$\mathring{A}_* := \sup_{\Sigma_0^1} [\mathcal{G} \breve{X} \Psi]_-, \tag{5B.1}$$

where $\mathcal{G} \neq 0$ (see Remark 3.9) is the blowup coefficient from Definition 3.8 and $[f]_- := |\min\{f, 0\}|$.

Remark 5.2 (functional dependence of $\mathcal{G}$ along Σ_0). Note that by (3G.2b) and the fact that $L^1 = L^1(\Psi, v)$, we can view $\mathcal{G}$, along Σ_0, as a function of $\Psi|_{\Sigma_0}$ and $v|_{\Sigma_0}$.

Remark 5.3 (positivity of $\mathring{A}_*$). Our main theorem relies on the assumption that $\mathring{A}_* > 0$. Thus, we will make this assumption for the rest of the article.

5B2. *Data-size assumptions.* For technical convenience, we assume that the solution is C^∞ with respect to the Cartesian coordinates along the "data hypersurfaces" $\Sigma_0^{U_0}$ and $\mathcal{P}_u^{2\mathring{A}_*^{-1}}$. However, to close our estimates, we only need to make assumptions on various Sobolev and Lebesgue norms of the data, where the norms are defined in terms of the geometric coordinates and the commutation vector fields $\mathscr{Z}$ defined in (3C.9a). In this subsubsection, we state the norm assumptions, which involve three parameters, denoted by $\mathring{\alpha}$, $\mathring{\epsilon}$, and $\mathring{A}$, that complement the parameter $\mathring{A}_*$. We note that $\mathring{A}$ does not need to be small, and that the same is true for the parameter $\mathring{A}_*$ from Definition 5.1. We will describe our smallness assumptions on $\mathring{\alpha}$ and $\mathring{\epsilon}$ in Section 5D.

We assume that the data satisfy the following size assumptions (see Section 3N regarding the vector field differential operator notation).

L^2 assumptions along Σ_0^1.

$$\|\mathscr{Z}_*^{[1,N_{\mathrm{Top}};1}\Psi\|_{L^2(\Sigma_0^1)},\ \|\mathscr{Z}^{\leq N_{\mathrm{Top}}-1;1}v\|_{L^2(\Sigma_0^1)},\ \|\mathscr{Z}^{\leq N_{\mathrm{Top}}-1;1}V\|_{L^2(\Sigma_0^1)} \leq \mathring{\epsilon}. \tag{5B.2}$$

L^∞ assumptions along Σ_0^1.

$$\|\Psi\|_{L^\infty(\Sigma_0^1)} \leq \mathring{\alpha}, \tag{5B.3a}$$

$$\|\mathscr{Z}_*^{[1,N_{\mathrm{Mid}}];1}\Psi\|_{L^\infty(\Sigma_0^1)},\ \|\mathscr{Z}^{\leq N_{\mathrm{Mid}}-1;1}v\|_{L^\infty(\Sigma_0^1)},\ \|\mathscr{Z}^{\leq N_{\mathrm{Mid}}-1;1}V\|_{L^\infty(\Sigma_0^1)} \leq \mathring{\epsilon}, \tag{5B.3b}$$

$$\|\check{X}\Psi\|_{L^\infty(\Sigma_0^1)} \leq \mathring{A}. \tag{5B.3c}$$

Assumptions along $\mathcal{P}_0^{2\mathring{A}_*^{-1}}$.

$$\|\mathscr{Z}^{\leq N_{\mathrm{Top}}-1;1}v\|_{L^2(\mathcal{P}_0^{2\mathring{A}_*^{-1}})},\ \|\mathscr{Z}^{\leq N_{\mathrm{Top}}-1;1}V\|_{L^2(\mathcal{P}_0^{2\mathring{A}_*^{-1}})} \leq \mathring{\epsilon}. \tag{5B.4}$$

Assumptions along $\mathcal{T}_{0,u}$. We assume that for $u \in [0,1]$, we have

$$\|\mathscr{P}^{\leq N_{\mathrm{Top}}-2}v\|_{L^2(\mathcal{T}_{0,u})},\ \|\mathscr{P}^{\leq N_{\mathrm{Top}}-2}V\|_{L^2(\mathcal{T}_{0,u})} \leq \mathring{\epsilon}. \tag{5B.5}$$

Remark 5.4. Roughly, we will study solutions that are perturbations of nontrivial solutions with $\mathring{\epsilon} = 0$. Note that $\mathring{\epsilon} = 0$ corresponds to a simple plane symmetric wave, as we described in Section 1D. Note also that $\mathring{\alpha}$, $\mathring{A}_*$, and $\mathring{A}$ are generally nonzero for simple plane symmetric waves.

5B3. *Estimates for the initial data of the remaining geometric quantities.* To close our proof, we will have to estimate the scalar functions μ, $\xi_j^{(\mathrm{Small})}$, ${}^{(i)}\Theta^j_{(\mathrm{Small})}$, and Ξ^j featured in the array (3E.1b) and definition (3C.7). In this subsubsection, as a preliminary step, we estimate the size of their data along Σ_0^1.

Lemma 5.5 (estimates for the data of μ, $\xi_j^{(\mathrm{Small})}$, ${}^{(i)}\Theta^j_{(\mathrm{Small})}$, and Ξ^j). *Under the data-size assumptions of Section 5B2, there exists a constant $C > 0$ depending on the parameter $\mathring{A}$ from* (5B.3c) *and a constant $C_\blacklozenge > 0$ that* ***does not depend on*** *$\mathring{A}$ such that the following estimates hold for the scalar functions μ, $\xi_j^{(\mathrm{Small})}$, ${}^{(i)}\Theta^j_{(\mathrm{Small})}$ and Ξ^j defined in Definitions 3.5 and 3.15 and* (3C.7), *whenever $\mathring{\alpha}$ and $\mathring{\epsilon}$ are sufficiently small* (*see Section 3N regarding the vector field notation*):

$$\|\mathscr{P}_*^{[1,N_{\mathrm{Top}}-1]}\mu\|_{L^2(\Sigma_0^1)} \leq C\mathring{\epsilon}, \tag{5B.6a}$$

$$\|\mu - 1\|_{L^\infty(\Sigma_0^1)} \leq C_\blacklozenge(\mathring{\alpha} + \mathring{\epsilon}), \tag{5B.6b}$$

$$\|L\mu\|_{L^\infty(\Sigma_0^1)} \leq C, \tag{5B.6c}$$

$$\|\mathscr{P}_*^{[1,N_{\mathrm{Mid}}-1]}\mu\|_{L^\infty(\Sigma_0^1)} \leq C\mathring{\epsilon}, \tag{5B.6d}$$

$$\|\mathscr{Z}_*^{[1,N_{\mathrm{Top}}-1];1}\xi_j^{(\mathrm{Small})}\|_{L^2(\Sigma_0^1)} \leq C\mathring{\epsilon}, \tag{5B.7a}$$

$$\|\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_0^1)} \leq C_\blacklozenge(\mathring{\alpha} + \mathring{\epsilon})\delta_j^1, \tag{5B.7b}$$

$$\|\mathscr{Z}_*^{[1,N_{\mathrm{Mid}}-1];1}\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_0^1)} \leq C\mathring{\epsilon}, \tag{5B.7c}$$

$$\|\check{X}\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_0^1)} \leq C, \tag{5B.7d}$$

$$\| \mathscr{Z}^{\leq N_{\mathrm{Top}}-1;1(i)} \Theta^j_{(\mathrm{Small})} \|_{L^2(\Sigma_0^1)} \leq C\mathring{\epsilon}, \tag{5B.8a}$$

$$\| \mathscr{Z}^{\leq N_{\mathrm{Mid}}-1;1(i)} \Theta^j_{(\mathrm{Small})} \|_{L^\infty(\Sigma_0^1)} \leq C\mathring{\epsilon}, \tag{5B.8b}$$

$$\| \mathscr{P}^{\leq N_{\mathrm{Top}}-1} \Xi^j \|_{L^2(\Sigma_0^1)} \leq C, \tag{5B.9a}$$

$$\| \mathscr{P}^{\leq N_{\mathrm{Mid}}-1} \Xi^j \|_{L^\infty(\Sigma_0^1)} \leq C. \tag{5B.9b}$$

Remark 5.6 (the "nonsmall" quantities). Note that the only estimates not featuring the smallness parameters $\mathring{\alpha}$ or $\mathring{\epsilon}$ are (5B.6c), (5B.7d), (5B.9a), and (5B.9b).

Proof sketch. We only sketch the proof since it is standard but has a tedious component that is similar to other analysis that we carry out later: commutator estimates of the type proved in Lemma 6.2, based on the vector field commutator identities (3H.2a)–(3H.2c).

To proceed, we use Lemmas 3.20 and 3.21, Corollary 3.23, and the fact that L^α and X^α are smooth functions of (Ψ, v) (the latter by (3C.4a)) to deduce the following schematic relationships, which hold along Σ_0 (where f is smooth):

$$(\mu - 1)|_{\Sigma_0} = (\Psi, v) \cdot \mathrm{f}(\Psi, v), \tag{5B.10}$$

$$\xi_j^{(\mathrm{Small})}|_{\Sigma_0} = (\Psi, v) \cdot \mathrm{f}(\Psi, v)\delta_j^1, \tag{5B.11}$$

$${}^{(i)}\Theta^j_{(\mathrm{Small})}|_{\Sigma_0} = 0, \tag{5B.12}$$

$$\Xi^j|_{\Sigma_0} = \mathrm{f}(\Psi, v), \tag{5B.13}$$

as well as the following evolution equations, also written in schematic form (where $P \in \mathscr{P}$):

$$L\mu = \mathrm{f}(\gamma)\check{X}\Psi + \mu\mathrm{f}(\gamma)L\Psi + \mu\mathrm{f}(\gamma)V, \tag{5B.14}$$

$$L\xi_j^{(\mathrm{Small})} = \mathrm{f}(\gamma)P\Psi + \mathrm{f}(\gamma)V, \tag{5B.15}$$

$$L\,{}^{(i)}\Theta^j_{(\mathrm{Small})} = \mathrm{f}(\gamma)P\Psi + \mathrm{f}(\gamma)V, \tag{5B.16}$$

$$L\Xi^j = (\Xi^1, \ldots, \Xi^n) \cdot \mathrm{f}(\gamma, P\Psi) + \mathrm{f}(\underline{\gamma}, L\Psi, \check{X}\Psi). \tag{5B.17}$$

By repeatedly differentiating (5B.14)–(5B.17) with the elements of $\mathscr{Z}$ and using the commutator identities (3H.2a)–(3H.2c), we can algebraically express all quantities that we need to estimate in terms of the derivatives of μ, $\xi_j^{(\mathrm{Small})}$, ${}^{(i)}\Theta^j_{(\mathrm{Small})}$, and Ξ^j with respect to the (Σ_t-tangent) vector fields in $\{\check{X}, {}^{(2)}\Theta, \ldots, {}^{(n)}\Theta\}$ and the $\mathscr{Z}$ derivatives of Ψ, v, and V. Then using (5B.10)–(5B.13), we can express, along Σ_0, the derivatives of μ, $\xi_j^{(\mathrm{Small})}$, ${}^{(i)}\Theta^j_{(\mathrm{Small})}$ and Ξ^j with respect to the elements of $\{\check{X}, {}^{(2)}\Theta, \ldots, {}^{(n)}\Theta\}$ in terms of the derivatives of Ψ and v with respect to the elements of $\{\check{X}, {}^{(2)}\Theta, \ldots, {}^{(n)}\Theta\}$. The estimates (5B.6a)–(5B.9b) then follow from these algebraic expressions, the data-size assumptions (5B.2)–(5B.3c), and the standard Sobolev calculus. We stress that the identities (3H.2a)–(3H.2c) show that commutator terms contain a factor involving a differentiation with respect to one of the ${}^{(i)}\Theta$, which, in view of our data-size assumptions from Section 5B2, leads to a gain in $\mathcal{O}(\mathring{\epsilon})$ smallness for all commutator terms. □

5C. ***Bootstrap assumptions.*** In this subsection, we state the bootstrap assumptions that we use to control the solution.

5C1. *$T_{(\text{Boot})}$, the positivity of μ, and the diffeomorphism property of Υ.* We now state some basic bootstrap assumptions. We start by fixing a real number $T_{(\text{Boot})}$ with

$$0 < T_{(\text{Boot})} \leq 2\mathring{A}_*^{-1}, \tag{5C.1}$$

where $\mathring{A}_* > 0$ (see Remark 5.3) is the data-dependent parameter from Definition 5.1.

We assume that on the spacetime domain $\mathcal{M}_{T_{(\text{Boot})},U_0}$ (see (3A.3f)), we have

$$\mu > 0. \tag{$\boldsymbol{BA}\mu > 0$}$$

Inequality ($\boldsymbol{BA}\mu > 0$) essentially means that no shocks are present in $\mathcal{M}_{T_{(\text{Boot})},U_0}$.

We also assume that

$$\begin{gathered}\text{the change of variables map } \Upsilon \text{ from Definition 3.24 is a diffeomorphism from} \\ [0, T_{(\text{Boot})}) \times [0, U_0] \times \mathbb{T}^{n-1} \text{ onto its image.}\end{gathered} \tag{5C.2}$$

5C2. *Fundamental L^∞ bootstrap assumptions.* In this section, we state our fundamental L^∞ bootstrap assumptions. We will derive strict improvements of the fundamental bootstrap assumptions in Corollary 8.8, on the basis of a priori energy estimates and Sobolev embedding.

Fundamental bootstrap assumptions for v and V. We assume that the following inequalities hold for $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$ ($\alpha = 0, \ldots, n,\ J = 1, \ldots, M$):

$$\| \mathscr{P}^{\leq N_{\text{Mid}}-1} v^J \|_{L^\infty(\Sigma_t^u)},\ \| \mathscr{P}^{\leq N_{\text{Mid}}-1} V_\alpha^J \|_{L^\infty(\Sigma_t^u)} \leq \varepsilon, \tag{5C.3}$$

where $\varepsilon > 0$ is a small bootstrap parameter (see Section 5D for discussion on the required smallness).

5C3. *Auxiliary bootstrap assumptions.* In addition to the fundamental bootstrap assumptions, we find it convenient to make auxiliary bootstrap assumptions, which we state in this subsubsection. We will derive strict improvements of the auxiliary bootstrap assumptions in Proposition 6.5.

Auxiliary bootstrap assumptions for Ψ. We assume that the following inequalities hold for $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:

$$\| \Psi \|_{L^\infty(\Sigma_t^u)} \leq \mathring{\alpha} + \varepsilon^{1/2}, \tag{5C.4a}$$

$$\| \mathscr{Z}_*^{[1,N_{\text{Mid}};]1} \Psi \|_{L^\infty(\Sigma_t^u)} \leq \varepsilon^{1/2}, \tag{5C.4b}$$

$$\| \check{X} \Psi \|_{L^\infty(\Sigma_t^u)} \leq \mathring{A} + \varepsilon^{1/2}. \tag{5C.4c}$$

Auxiliary bootstrap assumptions for v and V. We assume that the following inequalities hold for $(t, u) \in [0, T_{(\text{Boot})}) \times [0, U_0]$:

$$\| \mathscr{Z}^{\leq N_{\text{Mid}}-1;1} v \|_{L^\infty(\Sigma_t^u)} \leq \varepsilon^{1/2}, \tag{5C.5a}$$

$$\| \mathscr{Z}^{\leq N_{\text{Mid}}-1;1} V \|_{L^\infty(\Sigma_t^u)} \leq \varepsilon^{1/2}. \tag{5C.5b}$$

Auxiliary bootstrap assumptions for μ, $\xi_j^{(\mathrm{Small})}$, and ${}^{(j)}\Theta^k_{(\mathrm{Small})}$. We assume that the following inequalities hold for $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\|\mu\|_{L^\infty(\Sigma_t^u)} \leq 1 + 2\mathring{A}_*^{-1} \|\mathcal{G}\check{X}\Psi\|_{L^\infty(\Sigma_0^u)} + \mathring{\alpha}^{1/2} + \varepsilon^{1/2}, \tag{5C.6a}$$

$$\|L\mu\|_{L^\infty(\Sigma_t^u)} \leq \|\mathcal{G}\check{X}\Psi\|_{L^\infty(\Sigma_0^u)} + \varepsilon^{1/2}, \tag{5C.6b}$$

$$\|\mathscr{P}_*^{[1,N_{\mathrm{Mid}}-1]}\mu\|_{L^\infty(\Sigma_t^u)} \leq \varepsilon^{1/2}, \tag{5C.6c}$$

where $\mathcal{G} \neq 0$ (see Remark 3.9) is the blowup coefficient from Definition 3.8, $\|\mathcal{G}\check{X}\Psi\|_{L^\infty(\Sigma_0^u)} \leq C_\blacklozenge \mathring{A}$, and $C_\blacklozenge > 0$ is a constant with the parameter-dependence properties described in Section 1H.

Moreover, we assume that

$$\|\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_t^u)} \leq \mathring{\alpha}^{1/2} + \mathring{\epsilon}^{1/2}, \tag{5C.7a}$$

$$\|\mathscr{Z}_*^{[1,N_{\mathrm{Mid}}-1];1}\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_t^u)} \leq \varepsilon^{1/2}, \tag{5C.7b}$$

$$\|\check{X}\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_t^u)} \leq \|\check{X}\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_0^u)} + \varepsilon^{1/2}, \tag{5C.7c}$$

$$\|\mathscr{Z}^{\leq N_{\mathrm{Mid}}-1;1(j)}\Theta^k_{(\mathrm{Small})}\|_{L^\infty(\Sigma_t^u)} \leq \varepsilon^{1/2}. \tag{5C.7d}$$

5D. ***Smallness assumptions.*** For the remainder of the article, when we say that "statement X holds whenever A is small relative to B", we mean that there is a particular continuous increasing function $f : (0,\infty) \to (0,\infty)$ such that statement X holds whenever $A < f(B)$. The functions f are allowed to vary throughout the article. To avoid lengthening the paper, we often avoid explicitly specifying the form of f.

To ensure that all of the statements needed for our main results hold, we will make the following smallness assumptions, where we will continually adjust the required smallness in order to close our estimates:

- The bootstrap parameter ε and the data smallness parameter $\mathring{\epsilon}$ from Section 5B2. are small relative to 1.
- ε and $\mathring{\epsilon}$ are small relative to $\mathring{A}^{-1}$, where $\mathring{A}$ is the data-size parameter from Section 5B2.
- ε and $\mathring{\epsilon}$ are small relative to the data-size parameter $\mathring{A}_*$ from Definition 5.1.
- The data-size parameter $\mathring{\alpha}$ from Section 5B2 is small relative to 1.
- $\mathring{\epsilon} \leq \varepsilon < \mathring{\alpha}$.

The first two assumptions will allow us to treat error terms of size ε and $\varepsilon\mathring{A}$ as small quantities. The third assumption is relevant because the expected blowup time is approximately $\mathring{A}_*^{-1}$, and the assumption will allow us to show that various error products featuring a small factor ε in fact remain small for $t < 2\mathring{A}_*^{-1}$, which is plenty of time for us to show that a shock forms. The smallness assumption on $\mathring{\alpha}$ ensures that the solution remains within the regime of hyperbolicity of the equations and that $\mathcal{G} \neq 0$, where $\mathcal{G}$ is the blowup coefficient from Definition 3.8.

5E. ***Existence of data satisfying the size assumptions.*** We now outline a proof that there exists an open set of data satisfying the size assumptions of Section 5B and the smallness assumptions of Section 5D. Since the assumptions are stable under Sobolev perturbations, it is enough to exhibit data corresponding to plane symmetric solutions, that is, solutions that depend only on t and x^1. This means that along Σ_0, it is enough to exhibit appropriate data that depend only on x^1. To exhibit data for Ψ, we simply let $f(x^1)$ be any smooth nontrivial function that is compactly supported in Σ_0^1, and we set $\Psi(0, x^1, \dots, x^n) := \kappa f(x^1)$, where κ is a real parameter. We then take vanishing data for v, so that, as a consequence of the evolution equation (2A.1b), we have $v \equiv 0$ and $V \equiv 0$. With the help of these facts, it is straightforward to check that by choosing κ to be sufficiently small in magnitude, we can satisfy all of the desired assumptions. More precisely, by construction, we have $\mathring{\epsilon} = 0$, and by choosing $|\kappa|$ to be small, we can ensure that the quantity $\mathring{\alpha} > 0$ on the right-hand side of (5B.3a) is as small as we want.

5F. ***Basic assumptions, facts, and estimates that we use silently.*** In this subsection, we state some basic assumptions and conventions that we silently use throughout the rest of the paper when deriving estimates.

(1) All of the estimates that we derive hold on the bootstrap region $\mathcal{M}_{T_{(\mathrm{Boot})},U_0}$. Moreover, in deriving estimates, we rely on the data-size assumptions and bootstrap assumptions from Sections 5B–5C, and the smallness assumptions of Section 5D.

(2) All quantities that we estimate can be controlled in terms of the quantities featured in the array $\underline{\gamma}$ from definition (3E.1b), the Cartesian components Ξ^j of the $\mathcal{T}_{t,u}$-tangent vector field Ξ from (3C.7), and the $\mathscr{Z}$-derivatives of these quantities.

(3) We typically use the Leibniz rule for vector field differentiation when deriving pointwise estimates for the $\mathscr{Z}$-derivatives derivatives of products of the schematic form $\prod_{i=1}^m p_i$. Our derivative counts are such that after any product is differentiated, all factors except at most one are uniformly bounded in L^∞ on $\mathcal{M}_{T_{(\mathrm{Boot})},U_0}$.

(4) The constants $C > 0$ in all of our estimates are allowed to depend on the data-size parameters $\mathring{A}$ and $\mathring{A}_*^{-1}$, as we described in Section 1H.

(5) The constants $C_\blacklozenge > 0$ do not depend on $\mathring{A}$ or $\mathring{A}_*$, as we described in Section 1H.

(6) We use the convention for nonsensical terms mentioned in Remark 3.37.

5G. ***Omission of the independent variables in some expressions.*** We use the following notational conventions in the rest of the article:

- Many of our pointwise estimates are stated in the form

$$|f_1| \lesssim F(t)|f_2|$$

for some function F. Unless we otherwise indicate, it is understood that both f_1 and f_2 are evaluated at the point with geometric coordinates (t, u, ϑ).

- Unless we otherwise indicate, in integrals $\int_{\mathcal{T}_{t,u}} f \, d\vartheta$, we view the integrand f as a function of (t, u, ϑ), and ϑ is the integration variable.

• Unless we otherwise indicate, in integrals $\int_{\Sigma_t^u} f \, d\underline{\varpi}$, we view the integrand f as a function of (t, u', ϑ), and (u', ϑ) are the integration variables.

• Unless we otherwise indicate, in integrals $\int_{\mathcal{P}_u^t} f \, d\overline{\varpi}$, we view the integrand f as a function of (t', u, ϑ), and (t', ϑ) are the integration variables.

• Unless we otherwise indicate, in integrals $\int_{\mathcal{M}_{t,u}} f \, d\varpi$, we view the integrand f as a function of (t', u', ϑ), and (t', u', ϑ) are the integration variables.

6. Pointwise estimates and improvements of the auxiliary bootstrap assumptions

In this section, we use the data-size assumptions and bootstrap assumptions of Section 5 to derive pointwise and L^∞ estimates for various quantities. The main result is Proposition 6.5. In particular, the results of this section yield strict improvements of the auxiliary bootstrap assumptions of Section 5C3.

Remark 6.1. Throughout this section, we silently use the conventions described in Section 5F. Moreover, N_{Top} and N_{Mid} denote the integers from Section 5A.

6A. ***Commutator estimates.*** We start by providing some commutator estimates that we will use throughout the analysis.

Lemma 6.2. *Let $1 \leq N \leq N_{\mathrm{Top}}$ be an integer, let $\vec{I}$ be a multi-index for the set $\mathscr{P}$ of $\mathcal{P}_u$-tangent commutation vector fields such that $|\vec{I}| = N$, and let $\vec{J}$ be any permutation of $\vec{I}$ (in particular, $|\vec{I}| = |\vec{J}| = N \leq N_{\mathrm{Top}}$). Then the following identity for scalar functions f holds*:

$$\mathscr{P}^{\vec{I}} f - \mathscr{P}^{\vec{J}} f = 0. \tag{6A.1}$$

Let $1 \leq N \leq N_{\mathrm{Top}}$ be an integer. Then the following commutator estimate for scalar functions f holds (*see Definition 3.35 regarding the vector field notation*):

$$|[L, \mathscr{Z}^{N;1}]f| \lesssim |\mathscr{P}_*^{[1,N]} f| + \underbrace{|\mathscr{P}_*^{[1,\lfloor N/2 \rfloor]} f||\mathscr{Z}_*^{[1,N];1}\Psi|}_{\text{absent if } N=1} + \underbrace{|\mathscr{P}_*^{[1,\lfloor N/2 \rfloor]} f||\mathscr{P}_*^{[1,N-1]}\underline{\gamma}|}_{\text{absent if } N=1}. \tag{6A.2}$$

Let $2 \leq N \leq N_{\mathrm{Top}}$ be an integer, let $\vec{I} \in \mathcal{I}_*^{[1,N];1}$ (see Definition 3.32), and let $\vec{J}$ be any permutation of $\vec{I}$. Then the following commutator estimate for scalar functions f holds:

$$|\mathscr{Z}^{\vec{I}} f - \mathscr{Z}^{\vec{J}} f| \lesssim |\mathscr{P}_*^{[1,N-1]} f| + |\mathscr{P}_*^{[1,\lfloor N/2 \rfloor]} f||\mathscr{Z}_*^{[1,N-1];1}\gamma| + |\mathscr{P}_*^{[1,\lfloor N/2 \rfloor]} f||\mathscr{P}_*^{[1,N-1]}\underline{\gamma}|. \tag{6A.3}$$

Proof. Equation (6A.1) is a trivial consequence of the commutation identity (3H.2a).

The estimate (6A.2) is a straightforward consequence of the commutation identities (3H.2a)–(3H.2b) and the bootstrap assumptions.

Similarly, the estimate (6A.3) is a straightforward consequence of the commutation identities (3H.2a)–(3H.2c) and the bootstrap assumptions. □

6B. ***Transversal derivatives in terms of tangential derivatives.*** The next lemma, which is algebraic in nature, plays a crucial role in controlling v and V. Roughly, the lemma shows that the $\check{X}$ derivative of these quantities can be expressed in terms of their $\mathcal{P}_u$-tangential derivatives plus error terms. In particular, this means that we do not have to commute the evolution equations for v and V with $\check{X}$ in order to control $\check{X}v$ and $\check{X}V_\alpha$; we can instead use the equations to algebraically solve for the $\check{X}$ derivative. This is important because commuting these equations (which must be weighted with μ to avoid singular error terms) with $\check{X}$ would generate the error term $\check{X}\mu$, which is uncontrollable based on the degree of $\check{X}$-differentiability that we have imposed on Ψ.

Lemma 6.3 (algebraic expressions for transversal derivatives of v and V in terms of their tangential derivatives). *There exist smooth functions of* γ, *schematically denoted by* f, *such that the following algebraic identities hold whenever* $|\gamma|$ *is sufficiently small* (*where* $P \in \mathscr{P}$ *and* $Z \in \mathscr{Z}$):

$$\check{X}v = \mu \mathrm{f}(\gamma) V, \tag{6B.1}$$

$$\check{X}V_\alpha = \mathrm{f}(\underline{\gamma}) P V + \mathrm{f}(\underline{\gamma}, Z\Psi) V. \tag{6B.2}$$

Proof. To prove (6B.1), we first multiply (2A.1b) by μ and use Lemma 3.19 to obtain the following identity, whose right-hand side is written in schematic form: $\mu(A^0 + A^a\xi_a)Xv = \mu \mathrm{f}(\gamma) Pv = \mu \mathrm{f}(\gamma) P^\alpha V_\alpha = \mu \mathrm{f}(\gamma) V$. Next, using Definition 3.15, we see that $\mu(A^0 + A^a\xi_a)Xv = (A^0 - A^1 + A_{(\mathrm{Small})})\check{X}v$, where $A^0 - A^1$ is a matrix whose entries are of the schematic form $\mathrm{f}(\gamma)$ and $A_{(\mathrm{Small})}$ is a matrix whose entries are of the schematic form $\gamma \mathrm{f}(\gamma)$. From these facts and the assumption (2C.1), it follows that whenever $|\gamma|$ is sufficiently small, the matrix $A^0 - A^1 + A_{(\mathrm{Small})}$ is invertible. From this fact, the desired identity (6B.1) easily follows.

The proof of (6B.2) is based on (2A.5) and is similar but requires one new ingredient: we use Lemma 3.19 to (schematically) express the right-hand side of (2A.5) as $\mathrm{f}(\underline{\gamma}, Z\Psi)V$. □

With the help of Lemmas 6.2 and 6.3, we now derive pointwise estimates showing that the derivatives of v and V involving up to one $\check{X}$ differentiation can be controlled in terms of quantities that do not depend on the $\check{X}$ derivatives of v and V.

Lemma 6.4 (pointwise estimates for transversal derivatives of v and V in terms of their tangential derivatives). *The following estimates hold for* $1 \leq N \leq N_{\mathrm{Top}}$:

$$\begin{aligned} |\mathscr{Z}^{N;1} v| \lesssim{}& |\mathscr{Z}_*^{[1,N-1];1}\Psi| + |\mathscr{P}^{[1,N-1]}v| + |\mathscr{P}^{\leq N-1}V| \\ &+ \sum_{j=1}^{n} |\mathscr{Z}_*^{[1,N-1];1}\xi_j^{(\mathrm{Small})}| + \sum_{i=2}^{n}\sum_{j=1}^{n} |\mathscr{Z}_*^{[1,N-1];1(i)}\Theta^j_{(\mathrm{Small})}| + |\mathscr{P}_*^{[1,N-1]}\mu|. \end{aligned} \tag{6B.3}$$

Moreover, the following estimates hold for $1 \leq N \leq N_{\mathrm{Top}} - 1$:

$$\begin{aligned} |\mathscr{Z}^{N;1} V| \lesssim{}& |\mathscr{Z}_*^{[1,N];1}\Psi| + |\mathscr{P}^{[1,N-1]}v| + |\mathscr{P}^{\leq N}V| \\ &+ \sum_{j=1}^{n} |\mathscr{Z}_*^{[1,N-1];1}\xi_j^{(\mathrm{Small})}| + \sum_{i=2}^{n}\sum_{j=1}^{n} |\mathscr{Z}_*^{[1,N-1];1(i)}\Theta^j_{(\mathrm{Small})}| + |\mathscr{P}_*^{[1,N-1]}\mu|. \end{aligned} \tag{6B.4}$$

Proof. We will prove (6B.3)–(6B.4) simultaneously by using induction in N. The base case $N = 1$ can be handled using the same arguments given below and we omit these details. We therefore assume the induction hypothesis that (6B.3)–(6B.4) have been proved with $N-1$ in the role of N; to prove (6B.3)–(6B.4) in the case N, we first consider an order-N operator of the form $\mathscr{P}^{N-1}\check{X}$. Using (6B.2), we deduce that $\mathscr{P}^{N-1}\check{X}V_\alpha = \mathscr{P}^{N-1}\{\mathrm{f}(\underline{\gamma})PV + \mathrm{f}(\underline{\gamma}, Z\Psi)V\}$. From this expression and the bootstrap assumptions, we deduce that $|\mathscr{P}^{N-1}\check{X}V_\alpha| \lesssim$ the right-hand side of (6B.4) as desired. Then using the commutator estimate (6A.3) and the bootstrap assumptions, we can arbitrarily permute the vector field factors in $\mathscr{P}^{N-1}\check{X}V_\alpha$ up to error terms that are pointwise bounded in magnitude by $\lesssim$ the right-hand side of (6B.4) plus error terms of the form $|\mathscr{Z}_*^{[1,N-1];1}v| + |\mathscr{Z}_*^{[1,N-1];1}V|$, which (by the induction hypothesis) have already been shown to be bounded by $\lesssim$ the right-hand side of (6B.4). We have therefore obtained the desired bounds for V in the case that $\mathscr{Z}^{N;1}$ contains a factor of $\check{X}$. In the case that the operator $\mathscr{Z}^{N;1}$ contains a factor of $\check{X}$, the estimate (6B.3) for v follows similarly with the help of (6B.1). To prove (6B.3) in the case that the operator $\mathscr{Z}^{N;1}$ does not contain a factor of $\check{X}$, that is, that $\mathscr{Z}^{N;1} = \mathscr{P}^N$, we first write $\mathscr{P}^N v = \mathscr{P}^{N-1}(P^\alpha \partial_\alpha v) = \mathscr{P}^{N-1}(P^\alpha V_\alpha) = \mathscr{P}^{N-1}(\mathrm{f}(\gamma)V_\alpha)$. From this expression and the bootstrap assumptions, we bound the magnitude of the right-hand side of this equation by $\lesssim$ the right-hand side of (6B.3) as desired. In the case that $\mathscr{Z}^{N;1}$ does not contain a factor of $\check{X}$, that is, that $\mathscr{Z}^{N;1} = \mathscr{P}^N$, the estimate (6B.4) is trivial. We have therefore closed the induction. We clarify that in the final step, we allow $N = N_{\mathrm{Top}}$ in (6B.3), but not in (6B.4). □

6C. ***Pointwise estimates and improvements of the auxiliary bootstrap assumptions.*** We now state and prove the main result of this section.

Proposition 6.5 (pointwise estimates and improvements of the auxiliary bootstrap assumptions). *Let N_{Top} and N_{Mid} be the integers fixed in Section 5A. If $N \le N_{\mathrm{Top}}$, then the following estimates hold* (*see Section 3N regarding the vector field differential operator notation*).

Pointwise estimates for the commuted evolution equations of Ψ, v and V.

$$|L\mathscr{Z}^{N;1}\Psi| \lesssim |\mathscr{Z}_*^{[1,N];1}\Psi| + |\mathscr{Z}_*^{[1,N-1];1}\gamma| + |\mathscr{P}_*^{[1,N-1]}\underline{\gamma}|. \tag{6C.1}$$

Similarly, if $1 \le N \le N_{\mathrm{Top}}$, then the following pointwise estimates hold:

$$|\mu A^\alpha \partial_\alpha \mathscr{P}^{N-1}v| \lesssim |\mathscr{Z}_*^{[1,N];1}\Psi| + |\mathscr{Z}_*^{[1,N-1];1}\gamma| + |\mathscr{P}_*^{[1,N-1]}\underline{\gamma}|, \tag{6C.2a}$$

$$|\mu A^\alpha \partial_\alpha \mathscr{P}^{N-1}V_\alpha| \lesssim |\mathscr{Z}_*^{[1,N];1}\Psi| + |\mathscr{Z}_*^{[1,N-1];1}\gamma| + |\mathscr{P}_*^{[1,N-1]}\underline{\gamma}| + |V|. \tag{6C.2b}$$

Pointwise estimates for the commuted evolution equations of $\xi_j^{(\mathrm{Small})}$, ${}^{(i)}\Theta^j_{(\mathrm{Small})}$, and μ. If $1 \le N \le N_{\mathrm{Top}}$, then the following estimates hold:

$$|L\mathscr{Z}^{N-1;1}\xi_j^{(\mathrm{Small})}| \lesssim |\mathscr{Z}_*^{[1,N];1}\Psi| + |\mathscr{Z}_*^{[1,N-1];1}\gamma| + |\mathscr{P}_*^{[1,N-1]}\underline{\gamma}| + |V|, \tag{6C.3a}$$

$$|L\mathscr{Z}^{N-1;1\,(i)}\Theta^j_{(\mathrm{Small})}| \lesssim |\mathscr{Z}_*^{[1,N];1}\Psi| + |\mathscr{Z}_*^{[1,N-1];1}\gamma| + |\mathscr{P}_*^{[1,N-1]}\underline{\gamma}| + |V|. \tag{6C.3b}$$

Furthermore, if $2 \le N \le N_{\mathrm{Top}}$, then the following estimates hold:

$$|L\mathscr{P}^{N-1}\mu| \lesssim |\mathscr{Z}_*^{[1,N];1}\Psi| + |\mathscr{P}^{[1,N-1]}\gamma| + |\mathscr{P}_*^{[1,N-1]}\underline{\gamma}| + |V|. \tag{6C.4}$$

L^∞ estimates for Ψ. In addition, the following estimates hold:

$$\|\Psi\|_{L^\infty(\Sigma_t^u)} \leq \mathring{\alpha} + C\varepsilon, \tag{6C.5a}$$

$$\|\mathscr{Z}_*^{[1,N_{\mathrm{Mid}}];1}\Psi\|_{L^\infty(\Sigma_t^u)} \leq C\varepsilon, \tag{6C.5b}$$

$$\|\check{X}\Psi\|_{L^\infty(\Sigma_t^u)} \leq \mathring{A} + C\varepsilon. \tag{6C.5c}$$

L^∞ estimates for v and V. Moreover, the following estimates hold:

$$\|\mathscr{Z}^{\leq N_{\mathrm{Mid}}-1;1}v\|_{L^\infty(\Sigma_t^u)} \leq C\varepsilon, \tag{6C.6a}$$

$$\|\mathscr{Z}^{\leq N_{\mathrm{Mid}}-1;1}V\|_{L^\infty(\Sigma_t^u)} \leq C\varepsilon. \tag{6C.6b}$$

L^∞ estimates for $\xi_j^{(\mathrm{Small})}$, ${}^{(i)}\Theta^j_{(\mathrm{Small})}$, and μ. The following estimates hold:

$$\|\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_t^u)} \leq C_{\blacklozenge}\mathring{\alpha}\delta_j^1 + C\varepsilon, \tag{6C.7a}$$

$$\|\mathscr{Z}_*^{[1,N_{\mathrm{Mid}}-1];1}\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_t^u)} \leq C\varepsilon, \tag{6C.7b}$$

$$\|\check{X}\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_t^u)} \leq \|\check{X}\xi_j^{(\mathrm{Small})}\|_{L^\infty(\Sigma_0^u)} + C\varepsilon, \tag{6C.7c}$$

$$\|\mathscr{Z}^{\leq N_{\mathrm{Mid}}-1;1\,(i)}\Theta^j_{(\mathrm{Small})}\|_{L^\infty(\Sigma_t^u)} \leq C\varepsilon, \tag{6C.7d}$$

$$\|\mathscr{P}_*^{[1,N_{\mathrm{Mid}}-1]}\mu\|_{L^\infty(\Sigma_t^u)} \leq C\varepsilon. \tag{6C.7e}$$

Sharp estimates for μ and $L\mu$. In addition, the following pointwise estimates hold:

$$\begin{aligned}\mu(t,u,\vartheta) &= 1 + t[\mathcal{G}\check{X}\Psi](0,u,\vartheta) + \mathcal{O}_{\blacklozenge}(\mathring{\alpha}) + \mathcal{O}(\varepsilon)\\ &= 1 + t[\mathcal{G}\check{X}\Psi](t,u,\vartheta) + \mathcal{O}_{\blacklozenge}(\mathring{\alpha}) + \mathcal{O}(\varepsilon),\end{aligned} \tag{6C.8a}$$

$$\begin{aligned}L\mu(t,u,\vartheta) &= [\mathcal{G}\check{X}\Psi](0,u,\vartheta) + \mathcal{O}(\varepsilon)\\ &= \{\mathcal{G}|_{(\Psi,v)=(0,0)} + \mathcal{O}_{\blacklozenge}(\mathring{\alpha})\}\check{X}\Psi(t,u,\vartheta) + \mathcal{O}(\varepsilon),\end{aligned} \tag{6C.8b}$$

where the blowup coefficient $\mathcal{G}$ is defined in Definition 3.8 and, in view of Remark 5.2 and (3G.2b), $\mathcal{G}|_{(\Psi,v)=(0,0)} = -\frac{\partial L^1}{\partial \Psi}\big|_{(\Psi,v)=(0,0)}$.

Moreover,

$$\|\mu\|_{L^\infty(\Sigma_t^u)} \leq 1 + 2\mathring{A}_*^{-1}\|\mathcal{G}\check{X}\Psi\|_{L^\infty(\Sigma_0^u)} + C_{\blacklozenge}\mathring{\alpha} + C\varepsilon, \tag{6C.9a}$$

$$\|L\mu\|_{L^\infty(\Sigma_t^u)} \leq \|\mathcal{G}\check{X}\Psi\|_{L^\infty(\Sigma_0^u)} + C\varepsilon. \tag{6C.9b}$$

Estimates for Ξ^j. Finally, if $1 \leq N \leq N_{\mathrm{Top}}$, then the following estimates hold for the Cartesian components Ξ^j of the $\mathcal{T}_{t,u}$-tangent vector field Ξ from (3C.7):

$$|L\mathscr{P}^{\leq N-1}\Xi^j| \lesssim |\mathscr{P}^{\leq N-1}\Xi^j| + |\mathscr{Z}_*^{\leq N;1}\Psi| + |\mathscr{P}_*^{[1,N-1]}\underline{\gamma}| + 1, \tag{6C.10a}$$

$$\|\mathscr{P}^{\leq N_{\mathrm{Mid}}-1}\Xi^j\|_{L^\infty(\Sigma_t^u)} \lesssim 1. \tag{6C.10b}$$

Remark 6.6 (strict improvements of the auxiliary bootstrap assumptions). The L^∞ estimates of Proposition 6.5 provide, in particular, strict improvements of the auxiliary bootstrap assumptions of Section 5C3 whenever $\mathring{\alpha}$ and ε are sufficiently small.

Proof of Proposition 6.5. See Section 5F for some comments on the analysis. We start by noting that the order in which we prove estimates is important. Throughout the proof, we use the phrase "conditions on the data" to mean the assumptions from Section 5B2 for the data of Ψ, v, and V, as well as the estimates from Lemma 5.5 for the data of μ, $\xi_j^{(\text{Small})}$, ${}^{(i)}\Theta^j_{(\text{Small})}$, and Ξ^j. We also silently use the last item on page 491.

Proof of (6C.1): The estimate follows from the evolution equation (2A.1a), the commutator estimate (6A.2), and the bootstrap assumptions.

Proof of (6C.4): We first schematically write (3G.1a) as $L\mu = \mathrm{f}(\gamma)\check{X}\Psi + \mathrm{f}(\underline{\gamma})L\Psi + \mathrm{f}(\underline{\gamma})V$. Hence, using (6A.1), we deduce $L\mathscr{P}^{N-1}\mu = \mathscr{P}^{N-1}\{\mathrm{f}(\gamma)\check{X}\Psi + \mathrm{f}(\underline{\gamma})L\Psi + \mathrm{f}(\underline{\gamma})V\}$. The desired bound (6C.4) then follows from this equation and the bootstrap assumptions (we stress that the assumption $N \geq 2$ is needed for this estimate).

Proof of (6C.3a) and (6C.3b): We first schematically write (3G.1b) as $L\xi_j^{(\text{Small})} = \mathrm{f}(\gamma)P\Psi + \mathrm{f}(\gamma)V$. Hence, $L\mathscr{Z}^{N-1;1}\xi_j^{(\text{Small})} = [L, \mathscr{Z}^{N-1;1}]\xi_j^{(\text{Small})} + \mathscr{Z}^{N-1;1}\{\mathrm{f}(\gamma)P\Psi + \mathrm{f}(\gamma)V\}$. To bound the magnitude of the term $\mathscr{Z}^{N-1;1}\{\cdots\}$ by $\lesssim$ the right-hand side of (6C.3a), we use the bootstrap assumptions. To bound the commutator term $[L, \mathscr{Z}^{N-1;1}]\xi_j^{(\text{Small})}$, we also use (6A.2). The estimate (6C.3b) can be proved in the same way as the estimate (6C.3a), since by (3G.6b), ${}^{(i)}\Theta^j_{(\text{Small})}$ obeys a schematically identical evolution equation: $L\,{}^{(i)}\Theta^j_{(\text{Small})} = \mathrm{f}(\gamma)P\Psi + \mathrm{f}(\gamma)V$.

Proof of (6C.5b), (6C.7b), (6C.7d), and (6C.7e): We set

$$q = q(t,u,\vartheta) := |\mathscr{Z}_*^{[1,N_{\text{Mid}};1;1}\Psi| + \sum_{j=1}^{n} |\mathscr{Z}_*^{[1,N_{\text{Mid}}-1];1;1}\xi_j^{(\text{Small})}| + \sum_{i=2}^{n}\sum_{j=1}^{n} |\mathscr{Z}^{\leq N_{\text{Mid}}-1;1\,(i)}\Theta^j_{(\text{Small})}| + |\mathscr{P}_*^{[1,N_{\text{Mid}}-1]}\mu|. \quad (6C.11)$$

From (6C.1), (6C.3a)–(6C.4), the pointwise estimates of (6B.3), the fundamental bootstrap assumptions (5C.3), and the fundamental theorem of calculus, we deduce, in view of the fact that $L = \frac{\partial}{\partial t}$, that $|q(t,u,\vartheta)| \leq |q(0,u,\vartheta)| + c\int_{s=0}^{t} |q(s,u,\vartheta)|\,ds + C\varepsilon$. Moreover, the conditions on the data imply that $|q(0,u,\vartheta)| \leq C\varepsilon$. Hence, from Gronwall's inequality, we deduce that $|q(t,u,\vartheta)| \lesssim \varepsilon\exp(ct) \lesssim \varepsilon$, which implies all four of the desired bounds.

Proof of (6C.5a), (6C.5c), (6C.7a), and (6C.7c): To prove (6C.5a), we first use the fundamental theorem of calculus to obtain $|\Psi|(t,u,\vartheta) \leq |\Psi|(0,u,\vartheta) + \int_{s=0}^{t} |L\Psi|(s,u,\vartheta)\,ds$. The estimate (6C.5b) implies that the time integral in the previous inequality is $\lesssim \varepsilon$. In view of the conditions on the data, we conclude (6C.5a). The remaining three estimates can be proved similarly with the help of the estimates (6C.5b) and (6C.7b).

Proof of (6C.6a)–(6C.6b): These estimates follow from the pointwise estimates (6B.3)–(6B.4), the fundamental bootstrap assumptions (5C.3), and the estimates (6C.5b), (6C.7b), (6C.7d), and (6C.7e).

Proof of (6C.2a)–(6C.2b): We first use Lemma 3.19 to deduce the schematic relation

$$\mu\,\partial_\alpha = \mathrm{f}(\gamma)\check{X} + \mu\mathrm{f}(\gamma)P = \mathrm{f}(\gamma)\check{X} + \mathrm{f}(\underline{\gamma})P. \quad (6C.12)$$

Next, using (6C.12), the definition $\partial_\alpha v = V_\alpha$, and the fact that for $Z \in \mathscr{Z}$ we have $Z^\alpha = \mathrm{f}(\underline{\gamma})$, we deduce that $\mu \times$ the right-hand side of (2A.5) $= \mathrm{f}(\underline{\gamma}, Z\Psi)V$. Therefore, commuting $\mu \times$ (2A.5) with $\mathscr{P}^{N-1}$, we obtain

$$\mu A^\alpha \, \partial_\alpha \mathscr{P}^{N-1} V_\alpha = [\mathrm{f}(\gamma)\check{X}, \mathscr{P}^{N-1}]V_\alpha + [\mathrm{f}(\underline{\gamma})P, \mathscr{P}^{N-1}]V_\alpha + \mathscr{P}^{N-1}\{\mathrm{f}(\underline{\gamma}, Z\Psi)V\}. \tag{6C.13}$$

Using the bootstrap assumptions, we deduce that $|\mathscr{P}^{N-1}\{\mathrm{f}(\underline{\gamma}, Z\Psi)V\}| \lesssim$ the right-hand side of (6C.2b) as desired. To bound the commutator term $|[\mathrm{f}(\underline{\gamma})P, \mathscr{P}^{N-1}]V_\alpha|$, we use the bootstrap assumptions and the commutator identity (6A.1). To bound the commutator term $|[\mathrm{f}(\gamma)\check{X}, \mathscr{P}^{N-1}]V_\alpha|$, we use the bootstrap assumptions, the commutator estimate (6A.3), and the pointwise estimate (6B.4). We have therefore proved (6C.2b). The estimate (6C.2a) can be proved in a similar fashion starting from (2A.1b) and with the help of (6B.3); we omit the details.

Proof of (6C.8b): A special case of (6C.7e) is the estimate $LL\mu(t, u, \vartheta) = \mathcal{O}(\varepsilon)$. From this bound and the fundamental theorem of calculus, we deduce $L\mu(t, u, \vartheta) = L\mu(0, u, \vartheta) + \mathcal{O}(\varepsilon)$. Next, we use the identity $(\check{X}L^a)\xi_a = -(\check{X}L^1)\xi_1 + \sum_{a=2}^n (\check{X}L^a)\xi_a^{(\mathrm{Small})}$, definition (3C.3), and the conditions on the data to decompose (3G.1a) at time 0 as

$$L\mu(0, u, \vartheta) = -[\mathcal{G}\check{X}\Psi](0, u, \vartheta) + \mathcal{O}(\varepsilon). \tag{6C.14}$$

We next note that fundamental theorem of calculus yields

$$[\mathcal{G}\check{X}\Psi](t, u, \vartheta) = [\mathcal{G}\check{X}\Psi](0, u, \vartheta) + \int_{s=0}^t L[\mathcal{G}\check{X}\Psi](s, u, \vartheta)\, ds. \tag{6C.15}$$

Since the estimates (6C.5b) and (6C.7b) and the bootstrap assumptions imply that $L[\mathcal{G}\check{X}\Psi] = \mathcal{O}(\varepsilon)$, we deduce from (6C.15) that $[\mathcal{G}\check{X}\Psi](t, u, \vartheta) = [\mathcal{G}\check{X}\Psi](0, u, \vartheta) + \mathcal{O}(\varepsilon)$. Moreover, in view of Remark 5.2 and our data assumptions (5B.3a)–(5B.3b), we have, by Taylor expanding, the estimate $\mathcal{G}(0, u, \vartheta) := \mathcal{G}|_{(\Psi(0,u,\vartheta), v(0,u,\vartheta))} = \mathcal{G}|_{(\Psi,v)=(0,0)} + \mathcal{O}_\blacklozenge(\mathring{\alpha}) + \mathcal{O}(\varepsilon)$. Combining these estimates, we arrive at both of the bounds stated in (6C.8b).

Proof of (6C.8a): Using the fundamental theorem of calculus (as in (6C.15)) and the initial condition $\mu|_{\Sigma_0} = 1 + \mathcal{O}_\blacklozenge(\mathring{\alpha}) + \mathcal{O}(\varepsilon)$, which follows from (3B.2) and the conditions on the data, we obtain $\mu(t, u, \vartheta) = 1 + \mathcal{O}_\blacklozenge(\mathring{\alpha}) + \mathcal{O}(\varepsilon) + \int_{s=0}^t L\mu(s, u, \vartheta)\, ds$. Substituting the right-hand side of (6C.8b) (evaluated at (s, u, ϑ)) for the integrand $L\mu(s, u, \vartheta)$, we arrive at the first estimate stated in (6C.8a). To obtain the second estimate stated in (6C.8a), we use the first estimate and the bound $[\mathcal{G}\check{X}\Psi](t, u, \vartheta) = [\mathcal{G}\check{X}\Psi](0, u, \vartheta) + \mathcal{O}(\varepsilon)$ noted in the previous paragraph.

Proof of (6C.9a) and (6C.9b): Estimate (6C.9a) follows easily from (6C.8a) and the fact that $0 < t < 2\mathring{A}_*^{-1}$. Similarly, (6C.9b) follows easily from (6C.8b).

Proof of (6C.10a)–(6C.10b): Using (3H.3) and (6A.1), we deduce the following schematic identity: $L\mathscr{P}^{N-1}\Xi^j = \mathscr{P}^{N-1}\{\Xi^a[\mathrm{f}(\gamma)V + \mathrm{f}(\gamma)P\Psi]\} + \mathscr{P}^{N-1}\{\mathrm{f}(\underline{\gamma}, Z\Psi)\}$. From this identity and the bootstrap

assumptions, we deduce

$$\max_{1\le j\le n}|L\mathscr{P}^{\le N-1}\Xi^j| \lesssim \max_{1\le j\le n}|\mathscr{P}^{\le N-1}\Xi^j| + \max_{1\le j\le n}|\mathscr{P}^{\le \lfloor (N-1)/2\rfloor}\Xi^j|\{|\mathscr{Z}_*^{\le N;1}\Psi|+|\mathscr{P}_*^{[1,N-1]}\underline{\gamma}|+1\} + |\mathscr{Z}_*^{\le N;1}\Psi|+|\mathscr{P}_*^{[1,N-1]}\underline{\gamma}|+1. \quad (6C.16)$$

In particular, from (6C.16) and the bootstrap assumptions, we deduce

$$\max_{1\le j\le n}|L\mathscr{P}^{\le N_{\mathrm{Mid}}-1}\Xi^j| \lesssim \max_{1\le j\le n}|\mathscr{P}^{\le N_{\mathrm{Mid}}-1}\Xi^j| + 1. \quad (6C.17)$$

Moreover, from the conditions on the data, we deduce that $\max_{1\le j\le n}|\mathscr{P}^{\le N_{\mathrm{Mid}}-1}\Xi^j|(0,u,\vartheta)\lesssim 1$. Recalling that $L=\frac{\partial}{\partial t}$, we now use this data bound, (6C.17), and Gronwall's inequality in $\max_{1\le j\le n}|\mathscr{P}^{\le N_{\mathrm{Mid}}-1}\Xi^j|$ to deduce that $\max_{1\le j\le n}\|\mathscr{P}^{\le N_{\mathrm{Mid}}-1}\Xi^j\|_{L^\infty(\Sigma_t^u)}\lesssim 1$, which is the desired bound (6C.10b). Finally, from (6C.16) and (6C.10b), we conclude (6C.10a). □

6D. ***Estimates closely tied to the formation of the shock.*** In this subsection, we prove a lemma that lies at the heart of showing that μ vanishes in finite time and that its vanishing coincides with the blowup of $\max_{\alpha=0,\dots,n}|\partial_\alpha\Psi|$. Roughly, the lemma shows that when μ is small, $\check{X}\Psi$ must be quantitatively large in magnitude and that $\check{X}\Psi$ has a sign that forces μ to continue shrinking (the latter fact is important in that $\check{X}\Psi$ is the dominant term in the evolution equation (3G.1a) for μ).

We start by defining a quantity that captures the "worst-case" behavior of μ along Σ_t^u.

Definition 6.7. We define the following quantity, where μ is the inverse foliation density in Definition 3.5:

$$\mu_\star(t,u) := \min_{\Sigma_t^u}\mu. \quad (6D.1)$$

We now prove the main result of this subsection.

Lemma 6.8 ($|\check{X}\Psi|$ is large when μ is small). *The following implication holds*:

$$\mu(t,u,\vartheta)<\tfrac{1}{4} \quad\Longrightarrow\quad [\mathcal{G}\check{X}\Psi](t,u,\vartheta) < -\tfrac{1}{4}\mathring{A}_*, \quad (6D.2)$$

where the blowup coefficient $\mathcal{G}\neq 0$ (*see Remark 3.9*) *is defined in Definition 3.8 and the data-size parameter* $\mathring{A}_*$ *is defined in Definition 5.1.*

In addition,

$$\mu(t,u,\vartheta)<\tfrac{1}{4} \quad\Longrightarrow\quad |X\Psi|(t,u,\vartheta) > \frac{1}{8|\widetilde{\mathcal{G}}|}\frac{1}{\mu(t,u,\vartheta)}\mathring{A}_*, \quad (6D.3)$$

where the constant $\widetilde{\mathcal{G}} := \mathcal{G}|_{(\Psi,v)=(0,0)}$ *is the blowup coefficient evaluated at the background value of* $(\Psi,v)=(0,0)$ (*see Remark 5.2 and note that, as we mentioned just below* (6C.8b), $\widetilde{\mathcal{G}} = -\frac{\partial L^1}{\partial\Psi}\big|_{(\Psi,v)=(0,0)}$).

Finally, when $U_0=1$, *the quantity* $\mu_\star$ *defined in* (6D.1) *satisfies the estimate*

$$\mu_\star(t,1) = 1 - t\mathring{A}_* + \mathcal{O}_\blacklozenge(\mathring{\alpha}) + \mathcal{O}(\varepsilon). \quad (6D.4)$$

Proof. From the second estimate stated in (6C.8a), we deduce that if $\mu(t,u,\vartheta)<\frac{1}{4}$, then $t[\mathcal{G}\check{X}\Psi](t,u,\vartheta) < -\frac{3}{4}+\mathcal{O}_\blacklozenge(\mathring{\alpha})+\mathcal{O}(\varepsilon)$. From this bound and the fact that $0\le t<T_{(\mathrm{Boot})}<2\mathring{A}_*^{-1}$, we conclude (6D.2).

To prove (6D.3), we first use the fundamental theorem of calculus to deduce

$$\mathcal{G}(t,u,\vartheta) = \mathcal{G}(0,u,\vartheta) + \int_{s=0}^{t} L\mathcal{G}(s,u,\vartheta)\,ds. \tag{6D.5}$$

Since the estimates (6C.5b) and (6C.7b) and the bootstrap assumptions imply that $L\mathcal{G} = \mathcal{O}(\varepsilon)$, we find from (6D.5) that $\mathcal{G}(t,u,\vartheta) = \mathcal{G}(0,u,\vartheta) + \mathcal{O}(\varepsilon)$. Moreover, in view of Remark 5.2 and our data assumptions (5B.3a)–(5B.3b), we have, by Taylor expanding, the estimate $\mathcal{G}(0,u,\vartheta) := \mathcal{G}|_{(\Psi(0,u,\vartheta),v(0,u,\vartheta))} = \widetilde{\mathcal{G}} + \mathcal{O}_{\blacklozenge}(\mathring{\alpha}) + \mathcal{O}(\varepsilon)$. It follows that $\mathcal{G}(t,u,\vartheta) = \widetilde{\mathcal{G}} + \mathcal{O}_{\blacklozenge}(\mathring{\alpha}) + \mathcal{O}(\varepsilon)$. Using this estimate to substitute for the factor $\mathcal{G}(t,u,\vartheta)$ in the second inequality in (6D.2) and then taking the absolute value of the resulting inequality, we deduce that if $\mu(t,u,\vartheta) < \frac{1}{4}$, then

$$|\check{X}\Psi|(t,u,\vartheta) > \frac{1}{4\left\{|\widetilde{\mathcal{G}}| + \mathcal{O}_{\blacklozenge}(\mathring{\alpha}) + \mathcal{O}(\varepsilon)\right\}}\mathring{A}_{*}.$$

Dividing both sides of this inequality by $\mu(t,u,\vartheta)$ and appealing to (3C.4b), we arrive at (6D.3).

To prove (6D.4), we use the first line of (6C.8a) to deduce $\mu(t,u,\vartheta) = 1 + t[\mathcal{G}\check{X}\Psi](0,u,\vartheta) + \mathcal{O}_{\blacklozenge}(\mathring{\alpha}) + \mathcal{O}(\varepsilon)$. Taking the minimum of both sides of this estimate over $(u,\vartheta) \in [0,1] \times \mathbb{T}^{n-1}$ and appealing to Definitions 5.1 and 6.7, we conclude (6D.4). □

7. Estimates for the change of variables map

In this section, we derive estimates for the change of variables map Υ from Definition 3.24. The main result is Proposition 7.3, which will serve as a technical ingredient in our proof that the solution exists up until the first shock. Roughly, the proposition shows that if μ remains bounded from below strictly away from 0, then Υ can be extended to a diffeomorphism on the closure of the bootstrap domain.

7A. ***Control of the components of the change of variables map.*** In this subsection, we provide two preliminary lemmas that yield estimates for the components of Υ.

Lemma 7.1 (bounds for geometric coordinate partial derivatives of functions in terms of geometric vector field derivatives). *For $K \in \{0,1\}$, the following estimate holds for scalar functions f:*

$$\sum_{i_0+i_1+\cdots+i_n\le 1} \left\| \left(\frac{\partial}{\partial t}\right)^{i_0+K} \left(\frac{\partial}{\partial u}\right)^{i_1} \left(\frac{\partial}{\partial \vartheta^2}\right)^{i_2} \cdots \left(\frac{\partial}{\partial \vartheta^n}\right)^{i_n} f \right\|_{L^\infty(\Sigma_t^u)} \lesssim \|\mathscr{Z}^{\le 1+K;1} f\|_{L^\infty(\Sigma_t^u)}. \tag{7A.1}$$

Proof. From (3C.7) and (3F.1b), the fact that Ξ is $\mathcal{T}_{t,u}$-tangent, and (3C.8c), we deduce the identity

$$\frac{\partial}{\partial u} = \check{X} + \Xi^a\,\partial_a = \check{X} + \sum_{i=2}^{n} \Xi^a \mathrm{f}_{ia}(\gamma)^{(i)}\Theta.$$

From this identity and the L^∞ estimates of Proposition 6.5 (in particular the estimate (6C.10b)), it follows that $\frac{\partial}{\partial u}$ is a linear combination of the elements of $\mathscr{Z}$ with coefficients that are bounded in the norm $\|\cdot\|_{L^\infty(\Sigma_t^u)}$ by $\lesssim 1$. The estimate (7A.1) is a straightforward consequence of this fact and the facts that $L = \frac{\partial}{\partial t} \in \mathscr{Z}$ and ${}^{(i)}\Theta = \frac{\partial}{\partial \vartheta^i} \in \mathscr{Z}$. □

We now show that Υ can be extended to a function defined on the closure of the bootstrap domain that belongs to several function spaces.

Lemma 7.2 (a preliminary extension result for the change of variables map). *The components $\Upsilon^\alpha(t, u, \vartheta)$ of the change of variables map from Definition 3.24 extend to the compact domain $[0, T_{(\mathrm{Boot})}] \times [0, U_0] \times \mathbb{T}^{n-1}$ with the following regularity ($i = 2, \dots, n$, $\alpha = 0, \dots, n$):*

$$\Upsilon^\alpha, \ \frac{\partial}{\partial \vartheta^i}\Upsilon^\alpha \in \bigcap_{k=0,1} C^k\big([0, T_{(\mathrm{Boot})}], W^{1-k,\infty}([0, U_0] \times \mathbb{T}^{n-1})\big).$$

Moreover, the following estimates[31] *hold for $(t, u) \in [0, T_{(\mathrm{Boot})}] \times [0, U_0]$, where $C = C(\mathring{A})$:*

$$\sum_{i_0+i_1+\cdots+i_n \le 1} \left\| \left(\frac{\partial}{\partial t}\right)^{i_0} \left(\frac{\partial}{\partial u}\right)^{i_1} \left(\frac{\partial}{\partial \vartheta^2}\right)^{i_2} \cdots \left(\frac{\partial}{\partial \vartheta^n}\right)^{i_n} \Upsilon^\alpha \right\|_{L^\infty(\Sigma_t^u)} \le C, \tag{7A.2a}$$

$$\sum_{\substack{i_0+i_1+\cdots+i_n \le 2 \\ 1 \le i_2+\cdots+i_n}} \left\| \left(\frac{\partial}{\partial t}\right)^{i_0} \left(\frac{\partial}{\partial u}\right)^{i_1} \left(\frac{\partial}{\partial \vartheta^2}\right)^{i_2} \cdots \left(\frac{\partial}{\partial \vartheta^n}\right)^{i_n} \Upsilon^\alpha \right\|_{L^\infty(\Sigma_t^u)} \le C\varepsilon. \tag{7A.2b}$$

Proof. We will show that the following estimates hold for $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:

$$\sum_{K=0}^{1} \sum_{i_0+i_1+\cdots+i_n \le 1} \left\| \left(\frac{\partial}{\partial t}\right)^{i_0+K} \left(\frac{\partial}{\partial u}\right)^{i_1} \left(\frac{\partial}{\partial \vartheta^2}\right)^{i_2} \cdots \left(\frac{\partial}{\partial \vartheta^n}\right)^{i_n} \Upsilon^\alpha \right\|_{L^\infty(\Sigma_t^u)} \lesssim 1, \tag{7A.3}$$

$$\sum_{K=0}^{1} \sum_{\substack{i_0+i_1+\cdots+i_n \le 2 \\ 1 \le i_2+\cdots+i_n}} \left\| \left(\frac{\partial}{\partial t}\right)^{i_0+K} \left(\frac{\partial}{\partial u}\right)^{i_1} \left(\frac{\partial}{\partial \vartheta^2}\right)^{i_2} \cdots \left(\frac{\partial}{\partial \vartheta^n}\right)^{i_n} \Upsilon^\alpha \right\|_{L^\infty(\Sigma_t^u)} \lesssim \varepsilon. \tag{7A.4}$$

Since $L = \frac{\partial}{\partial t}$ relative to geometric coordinates, all results of the lemma then follow as straightforward consequences of (7A.3)–(7A.4), the fundamental theorem of calculus, and the completeness of the spaces $W^{j,\infty}([0, U_0] \times \mathbb{T}^{n-1})$ for $j = 0, 1$.

Using (7A.1), we see that to establish (7A.3), it suffices to show that

$$\|\mathscr{Z}^{\le 2;1}\Upsilon^\alpha\|_{L^\infty(\Sigma_t^u)} \lesssim 1. \tag{7A.5}$$

To derive (7A.5), we first clarify that Υ^α can be identified with the Cartesian coordinate x^α, viewed as a function of $(t, u, \vartheta^2, \dots, \vartheta^n)$. To bound x^α, we note that $Lx^\alpha = L^\alpha = \mathrm{f}(\Psi, v)$. Hence, the bootstrap assumptions imply that $\|Lx^\alpha\|_{L^\infty(\Sigma_t^u)} \lesssim 1$. From this estimate and the fundamental theorem of calculus (as in (6C.15)), we conclude (see footnote 31) that $\|x^\alpha\|_{L^\infty(\Sigma_t^u)} \lesssim 1$ as desired. Next, we note that for $P \in \mathscr{P}$, we have $Px^\alpha = P^\alpha = \mathrm{f}(\gamma)$ and $\check{X}x^\alpha = \check{X}^\alpha = \mathrm{f}(\gamma)$. Hence, to complete the proof of (7A.5), we need only to show that $\|\mathscr{P}^{\le 1}\mathrm{f}(\gamma)\|_{L^\infty(\Sigma_t^u)} \lesssim 1$ and $\|\mathscr{Z}^{\le 1;1}\mathrm{f}(\gamma)\|_{L^\infty(\Sigma_t^u)} \lesssim 1$. These bounds are simple

[31] The L^∞ estimate for the torus coordinates $x^i \in \mathbb{T}$ (where $i = 2, \dots, n$) stated in (7A.2a) should be interpreted as the statement that for each fixed $i \in \{2, \dots, n\}$ and $(u, \vartheta) \in [0, U_0] \times \mathbb{T}^{n-1}$, the Euclidean distance traveled by the curves $t \to x^i(t, u, \vartheta)$, $t \in [0, T_{(\mathrm{Boot})}]$, in the universal covering space $\mathbb{R}$ of $\mathbb{T}$ is uniformly bounded.

consequences of the bootstrap assumptions. We have therefore proved (7A.3). The estimate (7A.4) can be proved using a similar argument and we omit the details. □

7B. ***The diffeomorphism properties of the change of variables map.*** We now derive the main result of Section 7.

Proposition 7.3 (sufficient conditions for Υ to be a global diffeomorphism). *If*

$$\inf_{(t,u)\in[0,T_{(\mathrm{Boot})})\times[0,U_0]} \mu_\star(t,u) > 0, \tag{7B.1}$$

then the change of variables map Υ extends to a global diffeomorphism from $[0, T_{(\mathrm{Boot})}] \times [0, U_0] \times \mathbb{T}^{n-1}$ onto its image with the following regularity ($i = 2, \dots, n,\ \alpha = 0, \dots, n$):

$$\Upsilon^\alpha,\ {}^{(i)}\Theta\Upsilon^\alpha \in \bigcap_{k=0,1} C^k\big([0, T_{(\mathrm{Boot})}],\, W^{1-k,\infty}([0, U_0] \times \mathbb{T}^{n-1})\big). \tag{7B.2}$$

Proof. By the bootstrap assumption (5C.2), Υ is a diffeomorphism from $[0, T_{(\mathrm{Boot})}) \times [0, U_0] \times \mathbb{T}^{n-1}$ onto its image $\mathcal{M}_{T_{(\mathrm{Boot})},U_0}$. In addition, Lemma 7.2 implies that each component Υ^α extends to a function of the geometric coordinates satisfying (7B.2). Next, we use (3I.2), the L^∞ estimates of Proposition 6.5, and the assumption (7B.1) to deduce that the Jacobian determinant of Υ is uniformly bounded in magnitude from above and below away from 0 on $[0, T_{(\mathrm{Boot})}] \times [0, U_0] \times \mathbb{T}^{n-1}$. Hence, from the inverse function theorem, we deduce that Υ extends as a local diffeomorphism from $[0, T_{(\mathrm{Boot})}] \times [0, U_0] \times \mathbb{T}^{n-1}$ onto its image. Therefore, to complete the proof of the lemma, we need only to show that Υ is injective on the domain $[0, T_{(\mathrm{Boot})}] \times [0, U_0] \times \mathbb{T}^{n-1}$. Since Υ is a diffeomorphism on the domain $[0, T_{(\mathrm{Boot})}) \times [0, U_0] \times \mathbb{T}^{n-1}$, it suffices to show that $\Upsilon(T_{(\mathrm{Boot})}, u_1, \vartheta_1) \neq \Upsilon(T_{(\mathrm{Boot})}, u_2, \vartheta_2)$ whenever $(u_i, \vartheta_i) \in [0, U_0] \times \mathbb{T}^{n-1}$ and $(u_1, \vartheta_1) \neq (u_2, \vartheta_2)$.

We first show that if $u_1 \neq u_2$, then $\Upsilon(T_{(\mathrm{Boot})}, u_1, \vartheta_1) \neq \Upsilon(T_{(\mathrm{Boot})}, u_2, \vartheta_2)$. To this end, we observe that from definitions (3C.1b) and (3D.3d), the estimates (6C.9a) and (6C.7a), and the assumption (7B.1), it follows that $\sum_{a=1}^n |\partial_a u|$ is uniformly bounded from above and from below, strictly away from 0. It follows that no two distinct (closed) characteristic hypersurface portions $\mathcal{P}_{u_1}^{T_{(\mathrm{Boot})}}$ and $\mathcal{P}_{u_2}^{T_{(\mathrm{Boot})}}$ can intersect, which yields the desired result.

To finish the proof of the lemma, we must show that $\Upsilon(T_{(\mathrm{Boot})}, u, \vartheta_1) \neq \Upsilon(T_{(\mathrm{Boot})}, u, \vartheta_2)$ whenever $u \in [0, U_0]$ and $\vartheta_1 \neq \vartheta_2$. That is, we must show that for each fixed $u \in [0, U_0]$, the map υ defined by $\upsilon(\vartheta) := \Upsilon(T_{(\mathrm{Boot})}, u, \vartheta)$ is an injection from $\mathbb{T}^{n-1}$ onto its image. To this end, for each fixed $u \in [0, U_0]$, we consider the family of t-parametrized maps $\tilde{\upsilon}(t, \cdot)$ (where $t \in [0, T_{(\mathrm{Boot})}]$) defined to be the last $n-1$ components of $\Upsilon(t, u, \cdot)$; that is, $\tilde{\upsilon}(t; \vartheta) := \big(\Upsilon^2(t, u, \cdot), \Upsilon^3(t, u, \cdot), \dots, \Upsilon^n(t, u, \cdot)\big)$ (recall that Υ^i can be identified with the local Cartesian coordinate x^i). Note that $\tilde{\upsilon}(t, \cdot)$ can be viewed as a map from the domain $\mathbb{T}^{n-1}$ (equipped with the geometric coordinates $(\vartheta^2, \dots, \vartheta^n)$) to the target $\mathbb{T}^{n-1}$ (equipped with the Cartesian coordinates $(x^2, \dots, x^n)$). Since Υ is continuous on $[0, T_{(\mathrm{Boot})}] \times [0, U_0] \times \mathbb{T}^{n-1}$, it follows that υ is homotopic to the degree-one[32] map $\tilde{\upsilon}(0, \cdot)$ by the homotopy $\tilde{\upsilon}(t; \vartheta)$. Hence, it is a basic result of degree theory (see, for example, [Lee 2013, Proposition 17.36]) that $\tilde{\upsilon}(t, \cdot)$ is also a degree-one map. In

[32] $\tilde{\upsilon}(0, \cdot)$ is degree-one because $x^i(0, u, \vartheta^2, \dots, \vartheta^n) = \vartheta^i$ for $i = 2, \dots, n$ by construction.

particular, $\upsilon(\,\cdot\,) = \tilde{\upsilon}(T_{(\mathrm{Boot})}, \cdot\,)$ is degree-one. Next, we note that Lemma 7.2 implies that $\Upsilon^j(T_{(\mathrm{Boot})}, u, \cdot\,)$ can be viewed as a C^1 function of $(\vartheta^2, \dots, \vartheta^n) \in \mathbb{T}^{n-1}$ and that by (3D.3c) and (7A.2b), for $i, j = 2, \dots, n$, we have ${}^{(i)}\Theta\Upsilon^j(T_{(\mathrm{Boot})}, u, \vartheta^2, \dots, \vartheta^n) = \delta^{ij} + {}^{(i)}\Theta^j_{(\mathrm{Small})}(T_{(\mathrm{Boot})}, u, \vartheta^2, \dots, \vartheta^n) = \delta^{ij} + \mathcal{O}(\varepsilon)$, where δ^{ij} is the standard Kronecker delta. From this estimate and the degree-one property of $\upsilon(\,\cdot\,) = \tilde{\upsilon}(T_{(\mathrm{Boot})}, \cdot\,)$, we deduce[33] that for sufficiently small ε, $\upsilon(\,\cdot\,)$ is a bijection[34] from $\mathbb{T}^{n-1}$ to $\mathbb{T}^{n-1}$. In particular, υ is injective, which is the desired result. □

8. Energy estimates and strict improvements of the fundamental bootstrap assumptions

In this section, we derive the main estimates of the paper: a priori energy estimates that hold up to top order on the bootstrap region. The main ingredients in the proofs are the energy identities of Section 4 and the pointwise estimates of Proposition 6.5. As a corollary, we also derive strict improvements of the fundamental L^∞ bootstrap assumptions of Section 5C2.

8A. ***Definition of the fundamental L^2-controlling quantity.*** We start by defining the coercive quantity that we use to control the solution in L^2 up to top order.

Definition 8.1 (the main coercive L^2-controlling quantity). In terms of the energy-characteristic flux quantities of Definition 4.1 and the multi-index set $\mathcal{I}_*^{[1,N_{\mathrm{Top}};1}$ of Definition 3.32, we define

$$\begin{aligned}\mathbb{Q}(t,u) := \sup_{(t',u')\in[0,t]\times[0,u]} \max\Big\{ &\max_{\vec{I}\in\mathcal{I}_*^{[1,N_{\mathrm{Top}};1}} \mathbb{E}^{(\mathrm{Shock})}[\mathscr{Z}^{\vec{I}}\Psi](t',u'),\\ &\max_{|\vec{I}|\le N_{\mathrm{Top}}-1} \Big\{\mathbb{E}^{(\mathrm{Regular})}[\mathscr{P}^{\vec{I}} f](t',u') + \mathbb{F}^{(\mathrm{Regular})}[\mathscr{P}^{\vec{I}} f](t',u')\Big\}\Big\}. \qquad (8\mathrm{A}.1)\\ &f\in\{v^J\}_{1\le J\le M}\cup\{V^J_\alpha\}_{0\le\alpha\le n;1\le J\le M}\end{aligned}$$

8B. ***Coerciveness of the fundamental L^2-controlling quantity.*** In the next lemma, we exhibit the coerciveness properties of $\mathbb{Q}(t,u)$.

Lemma 8.2 (coerciveness of $\mathbb{Q}(t,u)$). *The following estimates hold*:

$$\sup_{(t',u')\in[0,t]\times[0,u]} \|\mathscr{Z}_*^{[1,N_{\mathrm{Top}};1}\Psi\|_{L^2(\Sigma^{u'}_{t'})} \le \mathbb{Q}^{1/2}(t,u), \qquad (8\mathrm{B}.1)$$

$$\sup_{(t',u')\in[0,t]\times[0,u]} \|\sqrt{\mu}\mathscr{P}^{\le N_{\mathrm{Top}}-1}v\|_{L^2(\Sigma^{u'}_{t'})} \le C\mathbb{Q}^{1/2}(t,u), \qquad (8\mathrm{B}.2\mathrm{a})$$

$$\sup_{(t',u')\in[0,t]\times[0,u]} \|\sqrt{\mu}\mathscr{P}^{\le N_{\mathrm{Top}}-1}V\|_{L^2(\Sigma^{u'}_{t'})} \le C\mathbb{Q}^{1/2}(t,u), \qquad (8\mathrm{B}.2\mathrm{b})$$

[33] Recall that if $f : \mathbb{T}^{n-1} \to \mathbb{T}^{n-1}$ is a C^1 surjective map without critical points, then f is degree-one if for $p, q \in \mathbb{T}^{n-1}$, $1 = \sum_{p\in f^{-1}(q)}(\operatorname{sign}\det df(p))$, where $df(p)$ denotes the differential of f at p and the $df(p)$ are computed relative to an atlas corresponding to the smooth orientation on $\mathbb{T}^{n-1}$ chosen at the beginning of the article. It is a basic fact of degree theory (see, for example, [Lee 2013, Theorem 17.35]) that the sum is independent of q. Note that in the context of the present argument, the components of the $(n-1)\times(n-1)$ matrix $df(\,\cdot\,)$ are ${}^{(i)}\Theta\Upsilon^j(T_{(\mathrm{Boot})}, u, \cdot\,)$ $(i, j = 2, 3, \dots, n)$.

[34] The surjective property of this map is easy to deduce.

$$\sup_{(t',u')\in[0,t]\times[0,u]} \|\mathscr{P}^{\leq N_{\mathrm{Top}}-1} v\|_{L^2(\mathcal{P}^{t'}_{u'})} \leq C\mathbb{Q}^{1/2}(t,u), \tag{8B.3a}$$

$$\sup_{(t',u')\in[0,t]\times[0,u]} \|\mathscr{P}^{\leq N_{\mathrm{Top}}-1} V\|_{L^2(\mathcal{P}^{t'}_{u'})} \leq C\mathbb{Q}^{1/2}(t,u). \tag{8B.3b}$$

Proof. Lemma 8.2 follows from Definition 8.1, Definition 4.1, Lemma 4.2, and the L^∞ estimates of Proposition 6.5 (which provide the smallness of γ that is assumed, for example, in the hypotheses of Lemma 4.2). □

8C. ***Sobolev embedding.*** The main result of this subsection is Lemma 8.4, a Sobolev embedding result which shows that the norm $\|\cdot\|_{L^\infty(\Sigma^u_t)}$ of v and V and their $\mathcal{P}_u$-tangential derivatives up to mid-order is controlled by $\mathbb{Q}$. In Corollary 8.8, we will use the lemma as an ingredient in our derivation of strict improvements of the fundamental L^∞ bootstrap assumptions. As a preliminary step, we provide the following lemma, in which we derive some L^2 estimates for v, V, and their derivatives along the codimension-two tori $\mathcal{T}_{t,u}$.

Lemma 8.3 (L^2 control of the non-shock-forming variables on $\mathcal{T}_{t,u}$). *The following estimates hold for* $0 \leq \alpha \leq n$ *and* $1 \leq J \leq M$:

$$\|\mathscr{P}^{\leq N_{\mathrm{Top}}-2} v^J\|_{L^2(\mathcal{T}_{t,u})},\ \|\mathscr{P}^{\leq N_{\mathrm{Top}}-2} V^J_\alpha\|_{L^2(\mathcal{T}_{t,u})} \leq C\mathring{\epsilon} + C\mathbb{Q}^{1/2}(t,u). \tag{8C.1}$$

Proof. We first note the following estimate for scalar functions f, which follows from differentiating under the integral and using Young's inequality:

$$\frac{\partial}{\partial t}\|f\|^2_{L^2(\mathcal{T}_{t,u})} = 2\int_{\mathcal{T}_{t,u}} fLf\, d\vartheta \leq \|f\|^2_{L^2(\mathcal{T}_{t,u})} + \|Lf\|^2_{L^2(\mathcal{T}_{t,u})}. \tag{8C.2}$$

Integrating (8C.2) from time 0 to time t, we find that

$$\|f\|^2_{L^2(\mathcal{T}_{t,u})} \leq \|f\|^2_{L^2(\mathcal{T}_{0,u})} + \int_{s=0}^{t} \|f\|^2_{L^2(\mathcal{T}_{s,u})}\, ds + \|Lf\|^2_{L^2(\mathcal{P}^t_u)}. \tag{8C.3}$$

From (8C.3) and Gronwall's inequality, we deduce that

$$\|f\|^2_{L^2(\mathcal{T}_{t,u})} \leq C\|f\|^2_{L^2(\mathcal{T}_{0,u})} + C\|Lf\|^2_{L^2(\mathcal{P}^t_u)}. \tag{8C.4}$$

We now apply (8C.4) with the role of f played by $\mathscr{P}^{\leq N_{\mathrm{Top}}-2} v^J$ and $\mathscr{P}^{\leq N_{\mathrm{Top}}-2} V^J_\alpha$. In view of the data-size assumptions (5B.5) and the bounds $\|L\mathscr{P}^{\leq N_{\mathrm{Top}}-2} v\|^2_{L^2(\mathcal{P}^t_u)} \lesssim \mathbb{Q}(t,u)$ and $\|L\mathscr{P}^{\leq N_{\mathrm{Top}}-2} V\|^2_{L^2(\mathcal{P}^t_u)} \lesssim \mathbb{Q}(t,u)$, which follow from (8B.3a)–(8B.3b), we arrive at the desired estimate (8C.1). □

We now prove the main result of this subsection.

Lemma 8.4 (L^∞ control of the non-shock-forming variables up to mid-order in terms of $\mathbb{Q}$). *The following estimates hold*:

$$\|\mathscr{P}^{\leq N_{\mathrm{Mid}}-1} v\|_{L^\infty(\Sigma^u_t)},\ \|\mathscr{P}^{\leq N_{\mathrm{Mid}}-1} V_\alpha\|_{L^\infty(\Sigma^u_t)} \leq C\mathring{\epsilon} + C\mathbb{Q}^{1/2}(t,u). \tag{8C.5}$$

Proof. Standard Sobolev embedding on $\mathbb{T}^{n-1}$ yields the following estimate for scalar functions f:

$$\|f\|_{L^\infty(\mathcal{T}_{t,u})} \lesssim \|f\|_{L^2(\mathcal{T}_{t,u})} + \sum_{K=1}^{\lfloor (n+1)/2 \rfloor} \sum_{Y_{(1)},\ldots,Y_{(K)}\in\{{}^{(i)}\Theta\}_{i=2,3,\ldots,n}} \|Y_{(1)}\cdots Y_{(K)} f\|_{L^2(\mathcal{T}_{t,u})}. \tag{8C.6}$$

The desired estimate (8C.5) now follows from (8C.6), (8C.1), and (5A.1), where the last of these equations in particular implies that $N_{\mathrm{Mid}} - 1 + \lfloor (n+1)/2 \rfloor \leq N_{\mathrm{Top}} - 2$. □

8D. ***Preliminary L^2 estimates for μ, $\xi_j^{(\mathrm{Small})}$, and ${}^{(i)}\Theta^j_{(\mathrm{Small})}$.*** In the next lemma, we bound the L^2 norms of the derivatives of the quantities μ, $\xi_j^{(\mathrm{Small})}$, and ${}^{(i)}\Theta^j_{(\mathrm{Small})}$ in terms of $\mathbb{Q}$. This serves as a preliminary step for our forthcoming derivation of L^2 estimates for Ψ, v, and V, since μ, $\xi_j^{(\mathrm{Small})}$, and ${}^{(i)}\Theta^j_{(\mathrm{Small})}$ appear as source terms in their commuted evolution equations (as is shown by the right-hand sides of (6C.1)–(6C.2b)).

Lemma 8.5 (L^2 estimates for μ, $\xi_j^{(\mathrm{Small})}$, and ${}^{(i)}\Theta^j_{(\mathrm{Small})}$ in terms of $\mathbb{Q}$). *The following estimates hold for $2 \leq i \leq n$, $1 \leq j \leq n$, and $(t,u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$, where $\mathbb{Q}$ is defined in Definition 8.1*:

$$\| \mathscr{P}_*^{[1,N_{\mathrm{Top}}-1]} \mu \|_{L^2(\Sigma_t^u)} \leq C\mathring{\epsilon} + C\mathbb{Q}^{1/2}(t,u), \tag{8D.1a}$$

$$\| \mathscr{Z}_*^{[1,N_{\mathrm{Top}}-1];1} \xi_j^{(\mathrm{Small})} \|_{L^2(\Sigma_t^u)} \leq C\mathring{\epsilon} + C\mathbb{Q}^{1/2}(t,u), \tag{8D.1b}$$

$$\| \mathscr{Z}^{[1,N_{\mathrm{Top}}-1];1\,(i)} \Theta^j_{(\mathrm{Small})} \|_{L^2(\Sigma_t^u)} \leq C\mathring{\epsilon} + C\mathbb{Q}^{1/2}(t,u). \tag{8D.1c}$$

Proof. See Section 5F for some comments on the analysis. We set

$$q = q(t,u) := \| \mathscr{P}_*^{[1,N_{\mathrm{Top}}-1]} \mu \|^2_{L^2(\Sigma_t^u)} + \sum_{j=1}^n \| \mathscr{Z}_*^{[1,N_{\mathrm{Top}}-1];1} \xi_j^{(\mathrm{Small})} \|^2_{L^2(\Sigma_t^u)} + \sum_{i=2}^n \sum_{j=1}^n \| \mathscr{Z}^{[1,N_{\mathrm{Top}}-1];1\,(i)} \Theta^j_{(\mathrm{Small})} \|^2_{L^2(\Sigma_t^u)}. \tag{8D.2}$$

The estimates from Lemma 5.5 for the data of μ, $\xi_j^{(\mathrm{Small})}$, and ${}^{(i)}\Theta^j_{(\mathrm{Small})}$ imply that $q(0,u) \leq C\mathring{\epsilon}^2$. Hence, from the pointwise estimates (6C.3a)–(6C.3b) and (6C.4), the pointwise estimates (6B.3)–(6B.4), Definition 3.16, Young's inequality, the energy identity (4B.2), and Lemma 8.2, we deduce that

$$\begin{aligned}
q(t,u) &\leq C\mathring{\epsilon}^2 + C \sum_{j=1}^n \int_{\mathcal{M}_{t,u}} |\mathscr{Z}_*^{[1,N_{\mathrm{Top}}-1];1} \xi_j^{(\mathrm{Small})}|^2 \, d\varpi \\
&\quad + C \sum_{i=2}^n \sum_{j=1}^n \int_{\mathcal{M}_{t,u}} |\mathscr{Z}_*^{[1,N_{\mathrm{Top}}-1];1\,(i)} \Theta^j_{(\mathrm{Small})}|^2 \, d\varpi \\
&\quad + C \int_{\mathcal{M}_{t,u}} |\mathscr{P}_*^{[1,N_{\mathrm{Top}}-1]} \mu|^2 \, d\varpi + C \int_{\mathcal{M}_{t,u}} |\mathscr{Z}_*^{[1,N_{\mathrm{Top}}];1} \Psi|^2 \, d\varpi \\
&\quad + C \int_{\mathcal{M}_{t,u}} |\mathscr{P}^{\leq N_{\mathrm{Top}}-1} v|^2 \, d\varpi + C \int_{\mathcal{M}_{t,u}} |\mathscr{P}^{\leq N_{\mathrm{Top}}-1} V|^2 \, d\varpi \\
&\leq C\mathring{\epsilon}^2 + C \int_{s=0}^t q(s,u) \, ds + C \int_{s=0}^t \mathbb{Q}(s,u) \, ds + C \int_{u'=0}^u \mathbb{Q}(t,u') \, du' \\
&\leq C\mathring{\epsilon}^2 + C \int_{s=0}^t q(s,u) \, ds + C\mathbb{Q}(t,u).
\end{aligned} \tag{8D.3}$$

From (8D.3) and Gronwall's inequality, we conclude the bound $q(t,u) \leq C\mathring{\epsilon}^2 + C\mathbb{Q}(t,u)$, from which the estimates (8D.1a)–(8D.1c) easily follow. □

8E. ***The main a priori estimates.*** In the next proposition, we derive our main a priori energy estimates.

Proposition 8.6 (the main a priori estimates). *There exists a constant $C > 0$ such that under the data-size assumptions of Section 5B2, the bootstrap assumptions of Section 5C2, and the smallness assumptions of Section 5D, the following estimates hold for $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, U_0]$:*

$$\mathbb{Q}(t, u) \leq C\mathring{\epsilon}^2 + C\int_{s=0}^{t} \mathbb{Q}(s, u)\, ds + C\int_{u'=0}^{u} \mathbb{Q}(t, u')\, du'. \tag{8E.1}$$

Moreover, as a consequence of (8E.1)*, the following estimate holds for $(t, u) \in [0, T_{(\mathrm{Boot})}) \times [0, 1]$:*

$$\mathbb{Q}(t, u) \leq C\mathring{\epsilon}^2. \tag{8E.2}$$

Remark 8.7 (a top-order L^2 estimate for v). From the pointwise estimate (6B.3), the bootstrap assumptions, Lemma 8.5, and (8E.2), one can easily obtain the bound $\| \mathscr{Z}^{\leq N_{\mathrm{Top}}; 1} v \|_{L^2(\Sigma_t^u)} \leq C\mathring{\epsilon}$, which is a gain of one derivative for v compared to what is directly implied by (8E.2). Similarly, we could gain a derivative for v in the L^∞ estimate (8E.8) below. However, we have no need for these gains of a derivative, so we will ignore them for the remainder of the paper.

Proof of Proposition 8.6. Proof of (8E.1): We first derive energy inequalities for Ψ and its derivatives. Let $\vec{I} \in \mathcal{I}_*^{[1, N_{\mathrm{Top}}]; 1}$ (see Definition 3.32). From definitions (3E.1a)–(3E.1b), the energy identity (4B.2), the data-size assumption (5B.2), the pointwise estimate (6C.1), the estimates (6B.3)–(6B.4), and Young's inequality, we deduce

$$\begin{aligned}
\mathbb{E}^{(\mathrm{Shock})}[\mathscr{Z}^{\vec{I}}\Psi](t, u) \leq{}& C\mathring{\epsilon}^2 + C\int_{\mathcal{M}_{t,u}} |\mathscr{Z}_*^{[1, N_{\mathrm{Top}}]; 1}\Psi|^2\, d\varpi + C\int_{\mathcal{M}_{t,u}} |\mathscr{P}^{\leq N_{\mathrm{Top}}-1} v|^2\, d\varpi \\
&+ C\int_{\mathcal{M}_{t,u}} |\mathscr{P}^{\leq N_{\mathrm{Top}}-1} V|^2\, d\varpi + C\int_{\mathcal{M}_{t,u}} |\mathscr{P}_*^{[1, N_{\mathrm{Top}}-1]}\mu|^2\, d\varpi \\
&+ C\sum_{j=1}^{n}\int_{\mathcal{M}_{t,u}} |\mathscr{Z}_*^{[1, N_{\mathrm{Top}}-1]; 1}\xi_j^{(\mathrm{Small})}|^2\, d\varpi \\
&+ C\sum_{i=2}^{n}\sum_{j=1}^{n}\int_{\mathcal{M}_{t,u}} |\mathscr{Z}_*^{[1, N_{\mathrm{Top}}-1]; 1(i)}\Theta^j_{(\mathrm{Small})}|^2\, d\varpi.
\end{aligned} \tag{8E.3}$$

From Lemmas 8.2 and 8.5, and (8E.3), we deduce

$$\mathbb{E}^{(\mathrm{Shock})}[\mathscr{Z}^{\vec{I}}\Psi](t, u) \leq C\mathring{\epsilon}^2 + C\int_{s=0}^{t} \mathbb{Q}(s, u)\, ds + C\int_{u'=0}^{u} \mathbb{Q}(t, u')\, du'. \tag{8E.4}$$

We now derive a similar energy inequality for v, V, and their derivatives. Specifically, using definitions (3E.1a)–(3E.1b), the energy-characteristic flux identity (4B.4), the data-size assumptions (5B.2) and (5B.4), the pointwise estimates (6C.2a)–(6C.2b), the estimates (6B.3)–(6B.4), Lemmas 8.2 and 8.5, and the L^∞ estimates of Proposition 6.5, we deduce that for $|\vec{I}| \leq N_{\mathrm{Top}} - 1$, we have, for any $f \in$

$\{v^J\}_{1\le J\le M}\cup\{V_\alpha^J\}_{0\le\alpha\le n;1\le J\le M}$, the estimate

$$\mathbb{E}^{(\mathrm{Regular})}[\mathscr{P}^{\vec{I}}f](t,u)+\mathbb{F}^{(\mathrm{Regular})}[\mathscr{P}^{\vec{I}}f](t,u)\le C\mathring{\epsilon}^2+C\int_{s=0}^{t}\mathbb{Q}(s,u)\,ds+C\int_{u'=0}^{u}\mathbb{Q}(t,u')\,du'. \quad (8E.5)$$

From (8E.4), (8E.5), and Definition 8.1, we conclude the desired bound (8E.1).

Proof of (8E.2): With $c>0$ a real parameter to be chosen below, we define

$$\mathbb{Q}_c(t,u):=\sup_{(\hat{t},\hat{u})\in[0,t]\times[0,u]}\{\exp(-c\hat{t})\exp(-c\hat{u})\mathbb{Q}(\hat{t},\hat{u})\}. \quad (8E.6)$$

Using (8E.1) and the simple inequality $\int_{y'=0}^{y}\exp(cy')\,dy'\le(1/c)\exp(cy)$, we deduce that for $(\hat{t},\hat{u})\in[0,t]\times[0,u]\subset[0,T_{(\mathrm{Boot})})\times[0,U_0]$, the following estimate holds:

$$\begin{aligned}\exp(-c\hat{t})\exp(-c\hat{u})\mathbb{Q}(\hat{t},\hat{u})&\le C\exp(-c\hat{t})\exp(-c\hat{u})\mathring{\epsilon}^2\\&\quad+C\exp(-c\hat{t})\exp(-c\hat{u})\times\Big\{\sup_{t'\in[0,\hat{t}]}\exp(-ct')\mathbb{Q}(t',\hat{u})\Big\}\times\int_{t'=0}^{\hat{t}}\exp(ct')\,dt'\\&\quad+C\exp(-c\hat{t})\exp(-c\hat{u})\times\Big\{\sup_{u'\in[0,\hat{u}]}\exp(-cu')\mathbb{Q}(\hat{t},u')\Big\}\times\int_{u'=0}^{\hat{u}}\exp(cu')\,du'\\&\le C\mathring{\epsilon}^2+\frac{2C}{c}\sup_{(t',u')\in[0,\hat{t}]\times[0,\hat{u}]}\{\exp(-ct')\exp(-cu')\mathbb{Q}(t',u')\}, \quad (8E.7)\end{aligned}$$

where the constant C on the right-hand side of (8E.7) can be chosen to be independent of $c>0$. From (8E.7) and definition (8E.6), we deduce that $\mathbb{Q}_c(t,u)\le C\mathring{\epsilon}^2+(2C/c)\mathbb{Q}_c(t,u)$. Hence, fixing $c:=c'>2C$, we deduce that $\mathbb{Q}_{c'}(t,u)\le C'\mathring{\epsilon}^2$. From this bound and the definition of $\mathbb{Q}_{c'}$, it follows that for $(t,u)\in[0,T_{(\mathrm{Boot})})\times[0,U_0]$, we have $\mathbb{Q}(t,u)\le C'\exp(c't)\exp(c'u)\mathring{\epsilon}^2\le C''\mathring{\epsilon}^2$, where C'' depends on C', c', and $\mathring{A}_*^{-1}$ (in view of the bootstrap assumption (5C.1)). This is precisely the desired bound (8E.2). □

Corollary 8.8 (improvement of the fundamental L^∞ bootstrap assumptions). *For $0\le\alpha\le n$ and $1\le J\le M$, the following estimates hold*:

$$\|\mathscr{P}^{\le N_{\mathrm{Mid}}-1}v^J\|_{L^\infty(\Sigma_t^u)},\ \|\mathscr{P}^{\le N_{\mathrm{Mid}}-1}V_\alpha^J\|_{L^\infty(\Sigma_t^u)}\le C\mathring{\epsilon}. \quad (8E.8)$$

In particular, if $C\mathring{\epsilon}<\varepsilon$, then the estimate (8E.8) *is a strict improvement of the fundamental bootstrap assumption* (5C.3).

Proof. Estimate (8E.8) follows from the energy estimate (8E.2) and the Sobolev embedding result (8C.5). □

9. Continuation criteria

In this section, we provide a proposition that yields continuation criteria. We will use the proposition during the proof of the main theorem (Theorem 10.1), specifically as an ingredient in showing that the solution survives until the shock.

Proposition 9.1 (continuation criteria). *Let $(\Psi, v^1, \dots, v^M)$ be a smooth solution to the system* (2A.1a)–(2A.1b) *satisfying the size assumptions*[35] *on Σ_0^1 and $\mathcal{P}_0^{2\mathring{A}_*^{-1}}$ stated in Section 5B as well as the smallness assumptions stated in Section 5D. Let $T_{(\text{Local})} \in (0, 2\mathring{A}_*^{-1})$ and $U_0 \in (0, 1]$, and assume that the solution exists classically on the ("open-at-the-top") spacetime region $\mathcal{M}_{T_{(\text{Local})},U_0}$ (where $\mathcal{M}_{T_{(\text{Local})},U_0}$ is defined in* (3A.3f)) *that is completely determined by the data on $\Sigma_0^{U_0} \cup \mathcal{P}_0^{2\mathring{A}_*^{-1}}$ (see Figure 2 on page 472). Let u be the eikonal function that satisfies the eikonal equation initial value problem* (3A.1), *let μ be the inverse foliation density of the characteristics $\mathcal{P}_u$ defined in* (3B.1), *and let $\lambda_\alpha = \mu\, \partial_\alpha u$ (as in* (3C.1a)). *Assume that $\mu > 0$ on $\mathcal{M}_{T_{(\text{Local})},U_0}$ and that the change of variables map Υ from geometric to Cartesian coordinates (see Definition 3.24) is a diffeomorphism from $[0, T_{(\text{Local})}) \times [0, U_0] \times \mathbb{T}^{n-1}$ onto $\mathcal{M}_{T_{(\text{Local})},U_0}$ such that for $i = 2, \dots, n$ and $\alpha = 0, \dots, n$, we have*

$$\Upsilon^\alpha,\ {}^{(i)}\Theta\Upsilon^\alpha \in \bigcap_{k=0,1} C^k\big([0, T_{(\text{Local})}),\, W^{1-k,\infty}([0, U_0] \times \mathbb{T}^{n-1})\big). \tag{9.1}$$

Let $\mathcal{H} \subset \mathbb{R} \times \mathbb{R}^M \times \mathbb{R}^{1+n}$ be the set of arrays $(\widetilde{\Psi}, \tilde{v}, \tilde{\lambda})$ such that the following two conditions hold:

- *The Cartesian components $L^i(\Psi, v)$ $(i = 1, \dots, n)$ and the $M \times M$ matrices $A^\alpha(\Psi, v)$ $(\alpha = 0, \dots, n)$ are smooth functions for (Ψ, v) belonging to a neighborhood of $(\widetilde{\Psi}, \tilde{v})$.*
- *$A^0(\Psi, v)$ and $A^\alpha(\Psi, v)\lambda_\alpha$ are positive definite matrices for (Ψ, v, λ) belonging to a neighborhood of $(\widetilde{\Psi}, \tilde{v}, \tilde{\lambda})$.*

Assume that none of the following four breakdown scenarios occur:

(1) $\inf_{\mathcal{M}_{T_{(\text{Local})},U_0}} \mu = 0$.

(2) $\sup_{\mathcal{M}_{T_{(\text{Local})},U_0}} \mu = \infty$.

(3) *There exists a sequence $p_n \in \mathcal{M}_{T_{(\text{Local})},U_0}$ such that $(\Psi(p_n), v(p_n), \lambda(p_n))$ escapes every compact subset of $\mathcal{H}$ as $n \to \infty$.*

(4) $\sup_{\mathcal{M}_{T_{(\text{Local})},U_0}} \max_{\alpha=0,1,\dots,n}\{|\partial_\alpha \Psi| + |V_\alpha|\} = \infty$, *where $V_\alpha^J = \partial_\alpha v^J$.*

In addition, assume that the following condition is satisfied:

(5) *The change of variables map Υ extends to the compact set $[0, T_{(\text{Local})}] \times [0, U_0] \times \mathbb{T}^{n-1}$ as a diffeomorphism onto its image that enjoys the regularity properties* (9.1) *with $[0, T_{(\text{Local})})$ replaced by $[0, T_{(\text{Local})}]$.*

Then there exists a $\Delta > 0$ such that Ψ, v, V, u, μ, λ, and all of the other geometric quantities defined throughout the article can be uniquely extended (where Ψ, v, u, and μ are smooth solutions to their evolutions equations) to a strictly larger region of the form $\mathcal{M}_{T_{(\text{Local})}+\Delta,U_0}$ into which their Sobolev regularity along $\Sigma_0^{U_0}$ and $\mathcal{P}_0^{2\mathring{A}_*^{-1}}$ (described in Section 5B) is propagated.[36] *Moreover, if Δ is sufficiently small, then none of the four breakdown scenarios occur in the larger region, and Υ extends to*

[35] Recall that even though we make size assumptions only for certain Sobolev norms, for technical convenience, we have assumed that the data on Σ_0^1 and $\mathcal{P}_0^{2\mathring{A}_*^{-1}}$ are C^∞.

[36] Put differently, the same norms that are finite along $\Sigma_0^{U_0}$ and $\mathcal{P}_0^{2\mathring{A}_*^{-1}}$ (as stated in Section 5B) are also finite along Σ_t^u and $\mathcal{P}_u^t$ for $(t, u) \in [0, T_{(\text{Local})} + \Delta] \times [0, U_0]$.

$[0, T_{(\text{Local})} + \Delta] \times [0, U_0] \times \mathbb{T}^{n-1}$ *as a diffeomorphism onto its image that enjoys the regularity properties* (9.1) *with* $[0, T_{(\text{Local})})$ *replaced by* $[0, T_{(\text{Local})} + \Delta]$.

Discussion of proof. The proof of Proposition 9.1 is mostly standard. A sketch of a similar result was provided in [Speck 2016, Proposition 21.1.1], so here, we only mention the main ideas. Criterion (3) is connected to avoiding a breakdown in hyperbolicity of the equation. Criterion (4) is a standard criterion used to locally continue the solution relative to the Cartesian coordinates. Criteria (1) and (2) and the assumption (5) for Υ are connected to ruling out the blowup of u, degeneracy of the change of variables map, and degeneracy of the region $\mathcal{M}_{T_{(\text{Local})},U_0}$. In particular, criteria (1) and (2) play a role in a proving that $\sum_{a=1}^{n} |\partial_a u|$ is uniformly bounded from above and strictly from below away from 0 on $\mathcal{M}_{T_{(\text{Local})},U_0}$ (the proof was essentially given in the proof of Proposition 7.3). □

10. The main theorem

We now prove the main result of the paper.

Theorem 10.1 (stable shock formation). *Let n denote the number of spatial dimensions, let N_{Top} and N_{Mid} be positive integers satisfying* (5A.1), *and let $\mathring{\alpha} > 0$, $\mathring{\epsilon} \geq 0$, $\mathring{A} > 0$, and $\mathring{A}_* > 0$ be the data-size parameters from Section 5B. For each $U_0 \in (0, 1]$ (as in* (3A.2)), *let*

$$T_{(\text{Lifespan});U_0} := \sup\Big\{t \in [0,\infty) \,\Big|\, \text{the solution exists classically on } \mathcal{M}_{t;U_0} \text{ and } \Upsilon \text{ is a diffeomorphism from } [0,t) \times [0,U_0] \times \mathbb{T}^{n-1} \text{ onto its image}\Big\},$$

where Υ is the change of variables map from Definition 3.24. If $\mathring{\alpha}$ is sufficiently small relative to 1 *and if $\mathring{\epsilon}$ is sufficiently small relative to* 1, $\mathring{A}^{-1}$, *and $\mathring{A}_*$ in the sense explained in Section 5D, then the following conclusions hold, where all constants can be chosen to be independent of U_0 (see Section 1H for our conventions regarding the dependence of constants on the various parameters).*

Dichotomy of possibilities. One of the following mutually disjoint possibilities must occur, where $\mu_\star(t,u) = \min_{\Sigma_t^u} \mu$ (as in (6D.1)) *and μ is the inverse foliation density of the transport characteristics $\mathcal{P}_u$ from Definition 3.5*:

(I) $T_{(\text{Lifespan});U_0} > 2\mathring{A}_*^{-1}$. *In particular, the solution exists classically on the spacetime region* $\operatorname{cl}\mathcal{M}_{2\mathring{A}_*^{-1},U_0}$, *where* cl *denotes closure. Furthermore,* $\inf\{\mu_\star(s,U_0) \mid s \in [0, 2\mathring{A}_*^{-1}]\} > 0$.

(II) $0 < T_{(\text{Lifespan});U_0} \leq 2\mathring{A}_*^{-1}$, *and*

$$T_{(\text{Lifespan});U_0} = \sup\Big\{t \in [0, 2\mathring{A}_*^{-1}) \,\Big|\, \inf\{\mu_\star(s,U_0) \mid s \in [0,t)\} > 0\Big\}. \tag{10.1}$$

In addition, case (II) *occurs when $U_0 = 1$, and we have the estimate*[37]

$$T_{(\text{Lifespan});1} = \{1 + \mathcal{O}_\blacklozenge(\mathring{\alpha}) + \mathcal{O}(\mathring{\epsilon})\}\mathring{A}_*^{-1}. \tag{10.2}$$

Case (I). *The energy estimates of Proposition 8.6 and the L^∞ estimates of Corollary 8.8 hold on* $\operatorname{cl}\mathcal{M}_{2\mathring{A}_*^{-1},U_0}$. *The same is true for the estimates of Lemma 6.4 and Proposition 6.5, but with all factors ε*

[37] See Section 1H regarding our use of the symbol $\mathcal{O}_\blacklozenge$.

on the right-hand side of all inequalities replaced by $C\mathring{\epsilon}$. *Moreover, for* μ *and the quantities from Definition 3.15, the following estimates hold for* $2 \leq i \leq n$, $1 \leq j \leq n$, *and* $(t, u) \in [0, 2\mathring{A}_*^{-1}] \times [0, U_0]$ (*see Section 3N regarding the differential operator notation*):

$$\| \mathscr{P}_*^{[1, N_{\mathrm{Top}}-1]} \mu \|_{L^2(\Sigma_t^u)} \leq C\mathring{\epsilon}, \tag{10.3a}$$

$$\| \mathscr{Z}_*^{[1, N_{\mathrm{Top}}-1]; 1} \xi_j^{(\mathrm{Small})} \|_{L^2(\Sigma_t^u)} \leq C\mathring{\epsilon}, \tag{10.3b}$$

$$\| \mathscr{Z}^{[1, N_{\mathrm{Top}}-1]; 1(i)} \Theta_{(\mathrm{Small})}^j \|_{L^2(\Sigma_t^u)} \leq C\mathring{\epsilon}. \tag{10.3c}$$

Case (II). *The energy estimates of Proposition 8.6 and the* L^∞ *estimates of Corollary 8.8 hold on* $\mathcal{M}_{T_{(\mathrm{Lifespan});U_0}, U_0}$, *as do the estimates of Lemma 6.4 and Proposition 6.5 with all factors* ε *on the right-hand side of all inequalities replaced by* $C\mathring{\epsilon}$. *Moreover, the estimates* (10.3a)–(10.3c) *hold for* $(t, u) \in [0, T_{(\mathrm{Lifespan});U_0}) \times [0, U_0]$. *In addition, the scalar functions* $\mathscr{Z}^{\leq N_{\mathrm{Mid}}-1; 1}\Psi$, $\mathscr{Z}^{\leq N_{\mathrm{Mid}}-2; 1} v^J$, $\mathscr{Z}^{\leq N_{\mathrm{Mid}}-2; 1} V_\alpha^J$, $\mathscr{P}^{\leq N_{\mathrm{Mid}}-2}\mu$, $\mathscr{Z}^{\leq N_{\mathrm{Mid}}-2; 1}\xi_j$, $\mathscr{Z}^{\leq N_{\mathrm{Mid}}-2; 1(i)}\Theta^j$, $\mathscr{Z}^{\leq N_{\mathrm{Mid}}-2; 1} L^i$, $\mathscr{P}^{\leq N_{\mathrm{Mid}}-2}\check{X}^i$, *and* $\mathscr{Z}^{\leq N_{\mathrm{Mid}}-2; 1} X^i$ *extend to* $\Sigma_{T_{(\mathrm{Lifespan});U_0}}^{U_0}$ *as functions of the geometric coordinates* (t, u, ϑ) *belonging to the space* $C\big([0, T_{(\mathrm{Lifespan});U_0}], L^\infty([0, U_0] \times \mathbb{T}^{n-1})\big)$.

Moreover, let $\Sigma_{T_{(\mathrm{Lifespan});U_0}}^{U_0;(\mathrm{Blowup})}$ *be the subset of* $\Sigma_{T_{(\mathrm{Lifespan});U_0}}^{U_0}$ *defined by*

$$\Sigma_{T_{(\mathrm{Lifespan});U_0}}^{U_0;(\mathrm{Blowup})} := \{(T_{(\mathrm{Lifespan});U_0}, u, \vartheta) \mid \mu(T_{(\mathrm{Lifespan});U_0}, u, \vartheta) = 0\}. \tag{10.4}$$

Then for each point $(T_{(\mathrm{Lifespan});U_0}, u, \vartheta) \in \Sigma_{T_{(\mathrm{Lifespan});U_0}}^{U_0;(\mathrm{Blowup})}$, *there exists a past neighborhood*[38] *containing it such that the following lower bound holds in the neighborhood*:

$$|X\Psi(t, u, \vartheta)| \geq \frac{1}{8\mathring{A}_*} \frac{1}{|\widetilde{\mathcal{G}}|\mu(t, u, \vartheta)}, \tag{10.5}$$

where $\widetilde{\mathcal{G}} := \mathcal{G}|_{(\Psi, v)=(0,0)}$ *is the blowup coefficient of Definition 3.8, evaluated at the background value of* $(\Psi, v) = (0, 0)$ (*see Remark 5.2 and note that, as we mentioned just below* (6C.8b), $\widetilde{\mathcal{G}} = -\frac{\partial L^1}{\partial \Psi}\big|_{(\Psi, v)=(0,0)}$). *In* (10.5), $1/(8|\widetilde{\mathcal{G}}|\mathring{A}_*)$ *is a* ***positive***[39] *data-dependent constant, and the* $\mathcal{T}_{t,u}$*-transversal,* Σ_t*-tangent vector field* X *is of order-unity Euclidean length*: $C^{-1} \leq \delta_{ab} X^a X^b \leq C$, *where* δ_{ij} *is the standard Kronecker delta. In particular,* $X\Psi$ *blows up like* $1/\mu$ *at all points in* $\Sigma_{T_{(\mathrm{Lifespan});U_0}}^{U_0;(\mathrm{Blowup})}$. *Conversely, at all points* $(T_{(\mathrm{Lifespan});U_0}, u, \vartheta) \in \Sigma_{T_{(\mathrm{Lifespan});U_0}}^{U_0} \backslash \Sigma_{T_{(\mathrm{Lifespan});U_0}}^{U_0;(\mathrm{Blowup})}$, *we have*

$$|X\Psi(T_{(\mathrm{Lifespan});U_0}, u, \vartheta)| < \infty. \tag{10.6}$$

Proof. Let $C' > 1$ be a constant. We will enlarge C' as needed throughout the proof. We define

$$T_{(\mathrm{Max});U_0} := \text{the supremum of the set of times } T_{(\mathrm{Boot})} \in [0, 2\mathring{A}_*^{-1}] \text{ such that:} \tag{10.7}$$

- Ψ, v^J, V_α^J, u, μ, $\xi_j^{(\mathrm{Small})}$, ${}^{(i)}\Theta_{(\mathrm{Small})}^j$, and all of the other quantities defined throughout the article exist classically on $\mathcal{M}_{T_{(\mathrm{Boot})}, U_0}$.

[38] By a past neighborhood, we mean an open set of points (t, u, ϑ) intersected with the slab $[0, T_{(\mathrm{Lifespan});U_0}] \times \mathbb{R} \times \mathbb{T}^{n-1}$.

[39] See Remarks 3.9 and 5.3.

- The change of variables map Υ from Definition 3.24 is a (global) diffeomorphism from $[0, T_{(\mathrm{Boot})}) \times [0, U_0] \times \mathbb{T}^{n-1}$ onto its image $\mathcal{M}_{T_{(\mathrm{Boot})}, U_0}$ satisfying
$$\Upsilon^\alpha,\ \frac{\partial}{\partial \vartheta^i}\Upsilon^\alpha \in \bigcap_{k=0,1} C^k\big([0, T_{(\mathrm{Boot})}),\, W^{1-k,\infty}([0, U_0] \times \mathbb{T}^{n-1})\big).$$
- $\inf\{\mu_\star(t, U_0) \mid t \in [0, T_{(\mathrm{Boot})})\} > 0$ (see Definition 6.7).
- The fundamental L^∞ bootstrap assumptions (5C.3) hold with $\varepsilon := C'\mathring{\epsilon}$ for $(t, u) \in \times[0, T_{(\mathrm{Boot})}) \times [0, U_0]$.

By standard local well-posedness for quasilinear hyperbolic systems (see, for example, [Ringström 2009, Part I]), if $\mathring{\alpha}$ and $\mathring{\epsilon}$ are sufficiently small in the sense explained in Section 5D and C' is sufficiently large, then $T_{(\mathrm{Max});U_0} > 0$. Under the same smallness/largeness assumptions, by Corollary 8.8, the bootstrap assumptions (5C.3) are not saturated for $(t, u) \in [0, T_{(\mathrm{Max});U_0}) \times [0, U_0]$. For this reason, all estimates proved throughout the article on the basis of the bootstrap assumptions in fact hold on $\mathcal{M}_{T_{(\mathrm{Boot})}, U_0}$ with ε replaced by $C\mathring{\epsilon}$. We use this fact throughout the remainder of the proof without further remark. In particular, the estimates of Proposition 6.5 hold for $(t, u) \in [0, T_{(\mathrm{Max});U_0}) \times [0, U_0]$ with all factors ε on the right-hand side of all inequalities replaced by $C\mathring{\epsilon}$. Moreover, by inserting the energy estimates of Proposition 8.6 into the right-hand sides of the estimates of Lemma 8.5, we conclude that the estimates (10.3a)–(10.3c) hold for $(t, u) \in [0, T_{(\mathrm{Max});U_0}) \times [0, U_0]$.

We now establish the dichotomy of possibilities. We first show that if
$$\inf\{\mu_\star(t, U_0) \mid t \in [0, T_{(\mathrm{Max});U_0})\} > 0, \tag{10.8}$$
then $T_{(\mathrm{Max});U_0} = 2\mathring{A}_*^{-1}$. To proceed, we assume for the sake of deriving a contradiction that (10.8) holds but that $T_{(\mathrm{Max});U_0} < 2\mathring{A}_*^{-1}$. Then from (10.8) and Proposition 7.3, we see that if $\mathring{\alpha}$ and $\mathring{\epsilon}$ are sufficiently small, then Υ extends to a global diffeomorphism from $[0, T_{(\mathrm{Max});U_0}] \times [0, U_0] \times \mathbb{T}$ onto its image that enjoys the regularity (7B.2) (with $T_{(\mathrm{Boot})}$ replaced by $T_{(\mathrm{Max});U_0}$ in (7B.2)). Also using the assumption (2C.1), Definition 3.6, definition (3D.3d), and the estimates of Proposition 6.5, we see that none of the four breakdown scenarios of Proposition 9.1 occur on $\mathcal{M}_{T_{(\mathrm{Max});U_0}, U_0}$. Hence, by Proposition 9.1, we can classically extend the solution to a region of the form $\mathcal{M}_{T_{(\mathrm{Max});U_0}+\Delta, U_0}$, with $\Delta > 0$ and $T_{(\mathrm{Max});U_0} + \Delta < 2\mathring{A}_*^{-1}$, such that all of the properties defining $T_{(\mathrm{Max});U_0}$ hold for the larger time $T_{(\mathrm{Max});U_0} + \Delta$. This contradicts the definition of $T_{(\mathrm{Max});U_0}$ and in fact implies that if (10.8) holds and if $\mathring{\alpha}$ and $\mathring{\epsilon}$ are sufficiently small, then (I) $T_{(\mathrm{Max});U_0} = 2\mathring{A}_*^{-1}$ and $T_{(\mathrm{Lifespan});U_0} > 2\mathring{A}_*^{-1}$. The only other possibility is: (II) $\inf\{\mu_\star(t, U_0) \mid t \in [0, T_{(\mathrm{Max});U_0})\} = 0$.

We now aim to show that case (II) corresponds to the formation of a shock singularity in the constant-time hypersurface subset $\Sigma^{U_0}_{T_{(\mathrm{Max});U_0}}$. We first derive the statements regarding the quantities that extend to $\Sigma^{U_0}_{T_{(\mathrm{Lifespan});U_0}}$ as elements of the space $C\big([0, T_{(\mathrm{Lifespan});U_0}], L^\infty([0, U_0] \times \mathbb{T})\big)$. Here we will prove the desired results with $T_{(\mathrm{Max});U_0}$ in place of $T_{(\mathrm{Lifespan});U_0}$; in the next paragraph, we will show that $T_{(\mathrm{Max});U_0} = T_{(\mathrm{Lifespan});U_0}$. Let q denote any of the quantities $\mathscr{Z}^{\le N_{\mathrm{Mid}}-1;1}\Psi, \dots, \mathscr{Z}^{\le N_{\mathrm{Mid}}-2;1}X^i$ that, in the theorem, are stated to extend. From the estimates of Lemma 6.4 and Proposition 6.5, we deduce that $\|Lq\|_{L^\infty(\Sigma_t^{U_0})}$ is uniformly bounded for $0 \le t < T_{(\mathrm{Max});U_0}$. Using this fact, the fact that $L = \frac{\partial}{\partial t}$, the

fundamental theorem of calculus, and the completeness of the space $L^\infty([0,U_0]\times\mathbb{T})$, we conclude that q extends to $\Sigma^{U_0}_{T_{(\mathrm{Max});U_0}}$ as a function of the geometric coordinates (t,u,ϑ) belonging to the space $C\big([0,T_{(\mathrm{Max});U_0}],L^\infty([0,U_0]\times\mathbb{T})\big)$, as desired.

We now show that the classical lifespan is characterized by (10.1) and that $T_{(\mathrm{Max});U_0}=T_{(\mathrm{Lifespan});U_0}$. To this end, we first use (6D.3) and the continuous extension properties proved in the previous paragraph to deduce (10.5). Also using Definition 3.15, the schematic relation $X^j_{(\mathrm{Small})}=\gamma\mathrm{f}(\gamma)$, and the L^∞ estimates of Proposition 6.5, we deduce that $C^{-1}\leq\delta_{ab}X^aX^b\leq C$. That is, the vector field X is of order-unity Euclidean length. From this estimate and (10.5), we deduce that at points in $\Sigma^{U_0}_{T_{(\mathrm{Max});U_0}}$ where μ vanishes, $|X\Psi|$ blows up like $1/\mu$. Hence, $T_{(\mathrm{Max});U_0}$ is the classical lifespan. That is, we conclude that $T_{(\mathrm{Max});U_0}=T_{(\mathrm{Lifespan});U_0}$, and we obtain the characterization (10.1) of the classical lifespan. The estimate (10.6) follows from the estimate (6C.5c), the fact that $\breve{X}=\mu X$, and the continuous extension properties proved in the previous paragraph.

Finally, to obtain (10.2), we use (6D.4) to conclude that $\mu_\star(t,1)$ vanishes for the first time when $t=\{1+\mathcal{O}_\blacklozenge(\mathring{\alpha})+\mathcal{O}(\mathring{\epsilon})\}\mathring{A}_*^{-1}$. We have therefore proved the theorem. □

Acknowledgements

Speck gratefully acknowledges support from NSF grant # DMS-1162211, from NSF CAREER grant DMS-1454419, from a Sloan Research Fellowship provided by the Alfred P. Sloan foundation, and from a Solomon Buchsbaum grant administered by the Massachusetts Institute of Technology.

References

[Alinhac 1999a] S. Alinhac, "Blowup of small data solutions for a class of quasilinear wave equations in two space dimensions, II", *Acta Math.* **182**:1 (1999), 1–23. MR Zbl

[Alinhac 1999b] S. Alinhac, "Blowup of small data solutions for a quasilinear wave equation in two space dimensions", *Ann. of Math.* (2) **149**:1 (1999), 97–127. MR Zbl

[Alinhac 2001] S. Alinhac, "The null condition for quasilinear wave equations in two space dimensions, II", *Amer. J. Math.* **123**:6 (2001), 1071–1101. MR Zbl

[Christodoulou 2000] D. Christodoulou, *The action principle and partial differential equations*, Annals of Math. Studies **146**, Princeton Univ. Press, 2000. MR Zbl

[Christodoulou 2007] D. Christodoulou, *The formation of shocks in 3-dimensional fluids*, Eur. Math. Soc., Zürich, 2007. MR Zbl

[Christodoulou 2009] D. Christodoulou, *The formation of black holes in general relativity*, Eur. Math. Soc., Zürich, 2009. MR Zbl

[Christodoulou 2019] D. Christodoulou, *The shock development problem*, Eur. Math. Soc., Zürich, 2019. MR Zbl

[Christodoulou and Klainerman 1993] D. Christodoulou and S. Klainerman, *The global nonlinear stability of the Minkowski space*, Princeton Math. Series **41**, Princeton Univ. Press, 1993. MR Zbl

[Christodoulou and Lisibach 2016] D. Christodoulou and A. Lisibach, "Shock development in spherical symmetry", *Ann. PDE* **2**:1 (2016), art. id. 3. MR Zbl

[Christodoulou and Miao 2014] D. Christodoulou and S. Miao, *Compressible flow and Euler's equations*, Surveys of Modern Math. **9**, Int. Press, Somerville, MA, 2014. MR Zbl

[Christodoulou and Perez 2016] D. Christodoulou and D. R. Perez, "On the formation of shocks of electromagnetic plane waves in non-linear crystals", *J. Math. Phys.* **57**:8 (2016), art. id. 081506. MR Zbl

[Dafermos 2010] C. M. Dafermos, *Hyperbolic conservation laws in continuum physics*, 3rd ed., Grundlehren der Math. Wissenschaften **325**, Springer, 2010. MR Zbl

[Holzegel et al. 2016] G. Holzegel, S. Klainerman, J. Speck, and W. W.-Y. Wong, "Small-data shock formation in solutions to 3D quasilinear wave equations: an overview", *J. Hyperbolic Differ. Equ.* **13**:1 (2016), 1–105. MR Zbl

[John 1974] F. John, "Formation of singularities in one-dimensional nonlinear wave propagation", *Comm. Pure Appl. Math.* **27** (1974), 377–405. MR Zbl

[John 1981] F. John, "Blow-up for quasilinear wave equations in three space dimensions", *Comm. Pure Appl. Math.* **34**:1 (1981), 29–51. MR Zbl

[Klainerman 1985] S. Klainerman, "Uniform decay estimates and the Lorentz invariance of the classical wave equation", *Comm. Pure Appl. Math.* **38**:3 (1985), 321–332. MR Zbl

[Klainerman 1986] S. Klainerman, "The null condition and global existence to nonlinear wave equations", pp. 293–326 in *Nonlinear systems of partial differential equations in applied mathematics, I* (Santa Fe, NM, 1984), edited by B. Nicolaenko et al., Lectures in Appl. Math. **23**, Amer. Math. Soc., Providence, RI, 1986. MR Zbl

[Klainerman and Rodnianski 2003] S. Klainerman and I. Rodnianski, "Improved local well-posedness for quasilinear wave equations in dimension three", *Duke Math. J.* **117**:1 (2003), 1–124. MR Zbl

[Klainerman and Rodnianski 2005] S. Klainerman and I. Rodnianski, "Rough solutions of the Einstein-vacuum equations", *Ann. of Math.* (2) **161**:3 (2005), 1143–1193. MR Zbl

[Klainerman et al. 2015] S. Klainerman, I. Rodnianski, and J. Szeftel, "The bounded L^2 curvature conjecture", *Invent. Math.* **202**:1 (2015), 91–216. MR Zbl

[Lax 1964] P. D. Lax, "Development of singularities of solutions of nonlinear hyperbolic partial differential equations", *J. Math. Phys.* **5** (1964), 611–613. MR Zbl

[Lax 1972] P. D. Lax, "The formation and decay of shock waves", *Amer. Math. Monthly* **79**:3 (1972), 227–241. MR Zbl

[Lax 1973] P. D. Lax, *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, CBMS Regional Conf. Series Appl. Math. **11**, Soc. Indust. Appl. Math., Philadelphia, 1973. MR Zbl

[Lee 2013] J. M. Lee, *Introduction to smooth manifolds*, 2nd ed., Graduate Texts in Math. **218**, Springer, 2013. MR Zbl

[Lindblad and Rodnianski 2010] H. Lindblad and I. Rodnianski, "The global stability of Minkowski space-time in harmonic gauge", *Ann. of Math.* (2) **171**:3 (2010), 1401–1477. MR Zbl

[Luk and Speck 2016] J. Luk and J. Speck, "The hidden null structure of the compressible Euler equations and a prelude to applications", 2016. To appear in *J. Hyperbolic Differ. Equ.* arXiv

[Luk and Speck 2018] J. Luk and J. Speck, "Shock formation in solutions to the 2D compressible Euler equations in the presence of non-zero vorticity", *Invent. Math.* **214**:1 (2018), 1–169. MR Zbl

[Miao 2018] S. Miao, "On the formation of shock for quasilinear wave equations with weak intensity pulse", *Ann. PDE* **4**:1 (2018), art. id. 10. MR Zbl

[Miao and Yu 2017] S. Miao and P. Yu, "On the formation of shocks for quasilinear wave equations", *Invent. Math.* **207**:2 (2017), 697–831. MR Zbl

[Rauch 1986] J. Rauch, "BV estimates fail for most quasilinear hyperbolic systems in dimensions greater than one", *Comm. Math. Phys.* **106**:3 (1986), 481–484. MR Zbl

[Riemann 1860] B. Riemann, "Über die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite", *Abh. Königlichen Ges. Wiss. Göttingen* **8** (1860), 43–66.

[Ringström 2009] H. Ringström, *The Cauchy problem in general relativity*, Eur. Math. Soc., Zürich, 2009. MR Zbl

[Sbierski 2016] J. Sbierski, "On the existence of a maximal Cauchy development for the Einstein equations: a dezornification", *Ann. Henri Poincaré* **17**:2 (2016), 301–329. MR Zbl

[Sideris 1984] T. C. Sideris, "Formation of singularities in solutions to nonlinear hyperbolic equations", *Arch. Ration. Mech. Anal.* **86**:4 (1984), 369–381. MR Zbl

[Sideris 1985] T. C. Sideris, "Formation of singularities in three-dimensional compressible fluids", *Comm. Math. Phys.* **101**:4 (1985), 475–485. MR Zbl

[Smith and Tataru 2005] H. F. Smith and D. Tataru, "Sharp local well-posedness results for the nonlinear wave equation", *Ann. of Math.* (2) **162**:1 (2005), 291–366. MR Zbl

[Speck 2016] J. Speck, *Shock formation in small-data solutions to 3D quasilinear wave equations*, Math. Surveys and Monographs **214**, Amer. Math. Soc., Providence, RI, 2016. MR Zbl

[Speck 2017] J. Speck, "A new formulation of the $3D$ compressible Euler equations with dynamic entropy: remarkable null structures and regularity properties", preprint, 2017. To appear in *Arch. Ration. Mech. Anal.* arXiv

[Speck 2018] J. Speck, "Shock formation for $2D$ quasilinear wave systems featuring multiple speeds: blowup for the fastest wave, with non-trivial interactions up to the singularity", *Ann. PDE* **4**:1 (2018), art. id. 6. MR Zbl

[Speck et al. 2016] J. Speck, G. Holzegel, J. Luk, and W. Wong, "Stable shock formation for nearly simple outgoing plane symmetric waves", *Ann. PDE* **2**:2 (2016), art. id. 10. MR Zbl

[Wong 2013] W. W.-Y. Wong, "A comment on the construction of the maximal globally hyperbolic Cauchy development", *J. Math. Phys.* **54**:11 (2013), art. id. 113511. MR Zbl

Received 17 Feb 2019. Revised 14 Apr 2019. Accepted 22 Apr 2019.

JARED SPECK: jared.speck@vanderbilt.edu
Department of Mathematics, Vanderbilt University, Nashville, TN, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles are usually in English or French, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not refer to bibliography keys. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and a Mathematics Subject Classification for the article, and, for each author, affiliation (if appropriate) and email address.

Format. Authors are encouraged to use LaTeX and the standard amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages — Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc. — allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PURE and APPLIED ANALYSIS

vol. 1 no. 3 2019

2578-5893 (2019)1:3;1-4