

## The Haldane model and its localization dichotomy

Giovanna Marcelli, Domenico Monaco, Massimo Moscolari and  
Gianluca Panati

*Dedicated to Gianfausto, on the occasion of his 85<sup>th</sup> birthday*

**Abstract.** *Gapped periodic quantum systems exhibit an interesting Localization Dichotomy, which emerges when one looks at the localization of the optimally localized Wannier functions associated to the Bloch bands below the gap. As recently proved, either these Wannier functions are exponentially localized, as it happens whenever the Hamiltonian operator is time-reversal symmetric, or they are delocalized in the sense that the expectation value of  $|\mathbf{x}|^2$  diverges. Intermediate regimes are forbidden.*

*Following the lesson of our Maestro, to whom this contribution is gratefully dedicated, we find useful to explain this subtle mathematical phenomenon in the simplest possible model, namely the discrete model proposed by Haldane [10]. We include a pedagogical introduction to the model and we explain its Localization Dichotomy by explicit analytical arguments. We then introduce the reader to the more general, model-independent version of the dichotomy proved in [19].*

### 1 Introduction

Gianfausto Dell'Antonio has been always transmitting to younger collaborators the attitude to understand – and explain – a mathematical phenomenon in the simplest possible model which still captures its essential features. Remembering his recommendation, we devote this contribution to explain a recent, model-independent result – namely the *Localization Dichotomy* for gapped periodic quantum systems, proved in [19] – by illustrating its essential features in a simple, but yet physically relevant, discrete model.

We consider the model proposed by Haldane in [10], which has become one of the paradigmatic models to describe Chern insulators, a subclass of topological insulators [1, 6, 12]. Haldane argued that the essential ingredient in the Quantum Hall Effect (QHE) is the breaking of time-reversal symmetry, an effect that can be obtained either by an external magnetic field (as in a QHE setup) or, alternatively,

---

2010 Mathematics Subject Classification: 81Q70, 81V70, 47A56, 47A10.

Keywords: Periodic Schrödinger operators, Chern insulators, Haldane model, Quantum Anomalous Hall Effect, Bloch frames, Wannier functions.

© The Author(s) 2018. This article is an open access publication.

by some mechanism internal to the sample, as e. g. the presence of strong magnetic dipole moments of the ionic cores. In Haldane's words [10]:

*“While the particular model presented here is unlikely to be directly physically realizable, it indicates that, at least in principle, the QHE can be placed in the wider context of phenomena associated with broken time-reversal invariance, and does not necessarily require external magnetic fields, but could occur as a consequence of magnetic ordering in a quasi-two-dimensional system.”*

Remarkably, the first sentence turned out to be too pessimistic: after three decades, Chern insulators predicted in [10] have been experimentally synthesized as crystalline solids [3, 4, 5] and the Haldane model can also be physically simulated by Bose-Einstein condensates in suitably arranged optical lattices.

In this paper, we first provide a pedagogical introduction to the Haldane model, which is here presented in the first-quantization formalism, as opposed to most of the physics literature, which uses instead a second-quantization language. In Section 3, we recall the definition of *Bloch functions* and of the *Chern number* associated to an isolated Bloch band, and we exhibit, in the Haldane model, a Bloch function producing a non-zero Chern number and having a singular derivative: more precisely, its  $H^1$ -norm diverges. This situation exemplifies a recent model-independent mathematical result [19], which shows that a non-zero Chern number indeed forces a divergence of the  $H^1$ -norm of the corresponding Bloch functions in any Bloch gauge. We explain in Section 4 how the latter divergence reflects into the delocalization of the corresponding *Wannier functions*, and we illustrate to the reader the more general *Localization Dichotomy* mentioned above. A natural question is whether the previous result – whose formulation heavily relies on periodicity – can be recast in the broader contest of non-periodic models. Some preliminary results in this direction are in preparation [15].

We hope that the introductory style of this contribution will be useful to fill the linguistic gap between mathematics and physics, as they represent a unity in the scientific vision of the person to whom the paper is dedicated.

**Dedication.** The senior author of this paper moved his first steps into the scientific world under the precious guidance of Gianfausto Dell'Antonio. From his example, as a scientist and as a human being, he learned not only how to do mathematics, but how to be a Mathematical Physicist. We all – authors of different generations – consider Gianfausto as our *Maestro*, and we gratefully acknowledge the unvaluable contribution he gave to the development of Quantum Mathematics in Italy over more than half a century.

## 2 The Haldane model and its symmetries

The tight-binding model proposed by Haldane [10] has become a paradigm in solid-state physics, as it is presumably the simplest physically-reasonable model which

is invariant by lattice-translations (a unitary  $\mathbb{Z}^2$ -symmetry) and simultaneously breaks, for some values of the parameters  $(\phi, M)$  labeling the model, time-reversal symmetry (an antiunitary  $\mathbb{Z}_2$ -symmetry). In view of that, it has become one of the most popular models to study materials in the Altland-Zirnbauer symmetry class A, which includes Quantum Hall systems and Chern insulators [1, 6, 12].

The Haldane model is usually presented by using a second-quantization formalism [6, 8, 24], which makes it difficult to readers unfamiliar with the latter to appreciate the simplicity and elegance of the essential ideas. Since second quantization is not needed at all to describe non-interacting electrons, we review in this Section the essential features of the Haldane model, in a pedagogical style, by using the usual language of discrete Schrödinger operators (*i. e.* a first-quantization formalism).

## 2.1 The honeycomb structure

The Haldane model describes independent electrons on a honeycomb structure<sup>1</sup>  $\mathcal{C} \subset \mathbb{R}^2$ , illustrated in Figure 1. The structure is characterized by the *displacement vectors*

$$\mathbf{d}_1 = d \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix}, \quad \mathbf{d}_2 = d \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}, \quad \mathbf{d}_3 = d \begin{pmatrix} -1 & 0 \end{pmatrix} = -\mathbf{d}_1 - \mathbf{d}_2,$$

where  $d$  is the smallest distance between two points of  $\mathcal{C}$ . The periodicity of the structure is expressed by the *periodicity vectors*

$$\mathbf{a}_1 = \mathbf{d}_2 - \mathbf{d}_3, \quad \mathbf{a}_2 = \mathbf{d}_3 - \mathbf{d}_1, \quad \mathbf{a}_3 = \mathbf{d}_1 - \mathbf{d}_2 = -\mathbf{a}_1 - \mathbf{a}_2. \quad (2.1)$$

The vectors  $\mathbf{a}_i$  generate a Bravais lattice  $\Gamma := \text{Span}_{\mathbb{Z}}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} \cong \mathbb{Z}^2$  where one  $\mathbf{a}_i$  is redundant as it is an integer linear combination of the other two. Any point  $\mathbf{x} \in \mathcal{C}$  can be written by using a Bravais lattice vector and one of the  $\mathbf{d}_i$  vectors. It is then sufficient to pick two  $\mathbf{a}_i$ -vectors and one  $\mathbf{d}_i$ -vector to generate the whole crystal. This choice, which is often called a *dimerization* of  $\mathcal{C}$ , is not unique, as illustrated in Figure 2. The above procedure is equivalent to the choice of a periodicity cell that contains two non-equivalent sites  $A$  and  $B$  (black and white dots in Figure 1, respectively), and is a fundamental cell w.r.t. the action of  $\Gamma$ . Hence, each choice of a periodicity cell provides an identification  $\mathcal{C} \cong \Gamma \times \{\mathbf{0}, \nu\}$ , where  $\nu$  is one of the displacement vectors, yielding an isomorphism<sup>2</sup>  $\ell^2(\mathcal{C}) \cong \ell^2(\Gamma) \otimes \mathbb{C}^2 \cong \ell^2(\Gamma, \mathbb{C}^2)$ . We will often use this “dimerization isomorphism” and the following typographic convention:

<sup>1</sup>The physics literature usually refers to the latter as a “honeycomb lattice”. We prefer to avoid here this ambiguous use of the word “lattice”, since this word has a precise meaning in mathematics: a lattice is a discrete subgroup of  $(\mathbb{R}^d, +)$  with maximal rank. The ambiguity does not arise when speaking about the *Bravais lattice*, which is a lattice for both physicists and mathematicians.

<sup>2</sup>From an abstract viewpoint, we are just using the fact that the  $L^2$ -functor, from measure spaces to Hilbert spaces, preserves the product structure, mapping the cartesian product into the tensor product.

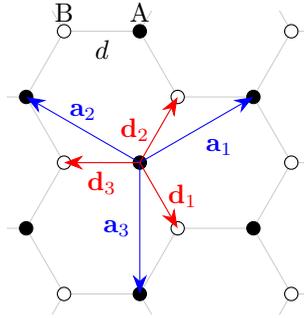


Figure 1: The honeycomb structure, with the displacement vectors  $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$  and the periodicity vectors  $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$  (color online).

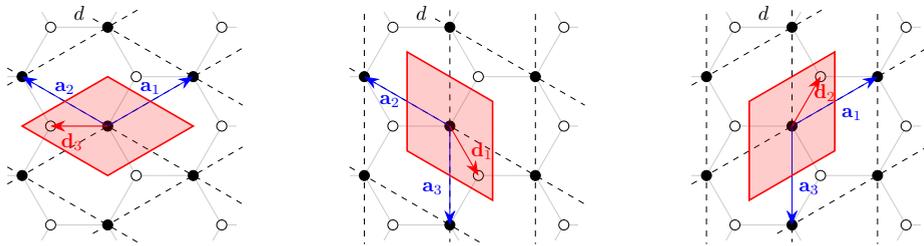


Figure 2: Three possible dimerizations of the honeycomb structure, corresponding to three different periodicity cells (color online).

- a small letter for a function  $\psi \in \ell^2(\mathcal{C})$ , with complex values  $\psi_{\mathbf{x}}$  for  $\mathbf{x} \in \mathcal{C}$ ;
- capital letter for a function  $\Psi \in \ell^2(\Gamma, \mathbb{C}^2)$ ; we make use of a pseudo-spin notation, namely

$$\Psi_{\gamma} = \begin{pmatrix} \psi_{\gamma,A} \\ \psi_{\gamma,B} \end{pmatrix} \quad \text{for } \gamma \in \Gamma,$$

where the labels  $A$  and  $B$  refer respectively to the sublattices  $\Gamma_A$  and  $\Gamma_B$ .

Finally, notice that the honeycomb structure has an interesting *inversion symmetry*, namely a reflection w.r.t. a specific line, which exchanges the role of the sublattices  $\Gamma_A$  and  $\Gamma_B$ . Thus, it yields a  $\mathbb{Z}_2$ -symmetry which can be easily broken by adding an on-site  $\Gamma$ -periodic potential which distinguishes between  $\Gamma_A$  and  $\Gamma_B$ . The latter procedure corresponds to a variation of the parameter  $M$  in the Haldane Hamiltonian, to be introduced shortly, and to the transition from graphene to boron-nitride sheets in physical reality.

## 2.2 The Hamiltonian

The Haldane model is defined, in a first quantization formalism, through a Hamiltonian operator acting on  $\ell^2(\mathcal{C}) \cong \ell^2(\Gamma, \mathbb{C}^2)$ , and depending on two real parameters  $(\phi, M)$ , with  $\phi \in (-\pi, \pi]$  representing a magnetic flux and  $M \in \mathbb{R}$  corresponding to an on-site energy which distinguishes among the two sublattices  $\Gamma_A$  and  $\Gamma_B$ .

The translation operator  $T_{\mathbf{u}}$ , corresponding to a translation by  $\mathbf{u} \in \mathbb{R}^2$ , is defined by

$$(T_{\mathbf{u}}\psi)_{\mathbf{x}} = \begin{cases} \psi_{\mathbf{x}-\mathbf{u}} & \text{if } \mathbf{x} - \mathbf{u} \in \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } \psi \in \ell^2(\mathcal{C}). \quad (2.2)$$

Moreover, we denote by  $\chi_A$  (resp.  $\chi_B$ ) the characteristic function of the sublattice  $\Gamma_A$  (resp.  $\Gamma_B$ ). For example, in the three dimerizations appearing in Figure 2, one has  $\Gamma_A = \Gamma$  and  $\Gamma_B = \Gamma + \nu$ , where  $\nu \in \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$  depends on the chosen dimerization.

Equipped with this notation, one defines the Haldane operator  $H \equiv H_{(\phi, M)}$  acting in  $\ell^2(\mathcal{C})$  (*i. e.* without reference to a specific dimerization) as a sum of three terms

$$H = H_{\text{NN}} + H_{\text{NNN}} + V. \quad (2.3)$$

The nearest neighbor (NN) term is defined – by using the displacement vectors – by

$$H_{\text{NN}} = t_1 \sum_{j=1}^3 (T_{\mathbf{d}_j} + T_{-\mathbf{d}_j}) \quad \text{with } t_1 \in \mathbb{R}. \quad (2.4)$$

The next nearest neighbor (NNN) term uses instead the periodicity vectors and reads

$$H_{\text{NNN}} = t_2(\cos \phi) \sum_{j=1}^3 (T_{\mathbf{a}_j} + T_{-\mathbf{a}_j}) + t_2(i \sin \phi)(\chi_A - \chi_B) \sum_{j=1}^3 (T_{\mathbf{a}_j} - T_{-\mathbf{a}_j}) \quad (2.5)$$

with  $t_2 \in \mathbb{R}$ . The last term is a potential that distinguishes sites in sublattices  $\Gamma_A$  and  $\Gamma_B$ , namely

$$V_{\mathbf{x}} = M(\chi_A - \chi_B)_{\mathbf{x}} = \begin{cases} +M & \text{if } \mathbf{x} \in \Gamma_A \\ -M & \text{if } \mathbf{x} \in \Gamma_B. \end{cases} \quad (2.6)$$

**Remark 2.1** (Comparison with the honeycomb Hofstadter model). By analogy with the Hofstadter model [13], one might be tempted to replace the NNN term by the more symmetric expression

$$\begin{aligned} \tilde{H}_{\text{NNN}} &= t_2 \sum_{j=1}^3 (e^{i\phi} T_{\mathbf{a}_j} + e^{-i\phi} T_{-\mathbf{a}_j}) \\ &= t_2(\cos \phi) \sum_{j=1}^3 (T_{\mathbf{a}_j} + T_{-\mathbf{a}_j}) + t_2(i \sin \phi) \sum_{j=1}^3 (T_{\mathbf{a}_j} - T_{-\mathbf{a}_j}). \end{aligned} \quad (2.7)$$

Notice, however, that the latter operator does not distinguish between the sublattices  $\Gamma_A$  and  $\Gamma_B$ , yielding an operator which acts diagonally on the  $\mathbb{C}^2$ -factor in  $\ell^2(\Gamma) \otimes \mathbb{C}^2$ . The operator (2.5) acts instead in a non-diagonal way, and offers the opportunity to model subtler physical effects.

One can easily check that the Haldane model enjoys some relevant symmetries:

- (i)  $\Gamma$ -periodicity: indeed, one checks that  $[T_\gamma, H] = 0$  for every  $\gamma \in \Gamma$ ;
- (ii)  $\frac{2\pi}{3}$ -rotation symmetry: indeed  $[U_R, H] = 0$  where  $U_R$  is defined as usual by  $(U_R\psi)_{\mathbf{x}} = \psi_{R^{-1}\mathbf{x}}$ , with  $R \in \text{SO}(2)$  a rotation by a  $\frac{2\pi}{3}$  angle in the plane;
- (iii) broken time-reversal symmetry (TRS): for  $\phi \in \{0, \pi\}$  the Hamiltonian commutes with the time-reversal operator, given by complex conjugation in  $\ell^2(\mathcal{C})$ . As far as  $\sin \phi \neq 0$ , TRS is broken, as it clearly appears from (2.5).

### 2.3 The Fourier decomposition

The  $\Gamma$ -periodicity of the model allows to use Fourier transform or, more intrinsically, the Bloch-Floquet decomposition.

Since the Fourier transform unitarily maps  $\ell^2(\mathbb{Z}^d)$  into  $L^2(\mathbb{T}^d)$ , after a choice of dimerization one obtains an isomorphism  $\ell^2(\mathcal{C}) \cong \ell^2(\Gamma, \mathbb{C}^2) \cong L^2(\mathbb{T}_*^2, \mathbb{C}^2)$  where the torus  $\mathbb{T}_*^2 = \mathbb{R}^2/\Gamma^*$ , called *Brillouin torus* by physicists, is defined as a quotient by the reciprocal or dual lattice

$$\Gamma^* = \{ \mathbf{k} \in \mathbb{R}^2 : \mathbf{k} \cdot \gamma \in 2\pi\mathbb{Z} \text{ for all } \gamma \in \Gamma \}. \quad (2.8)$$

We choose any dimerization such that the sublattices are identified with  $\Gamma$  and  $\Gamma + \nu$ , respectively, for a suitable  $\nu \in \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$  (compare Figure 2). With this convention, an isomorphism is exhibited by

$$(\mathcal{F}_\nu\psi)(\mathbf{k}) = \sum_{\gamma \in \Gamma} e^{-i\mathbf{k} \cdot \gamma} \Psi_{-\gamma} = \sum_{\gamma \in \Gamma} e^{-i\mathbf{k} \cdot \gamma} \begin{pmatrix} \psi_{-\gamma} \\ \psi_{-\gamma+\nu} \end{pmatrix}. \quad (2.9)$$

The operator  $\mathcal{F}_\nu$  establishes a unitary transformation

$$\mathcal{F}_\nu: \ell^2(\mathcal{C}) \rightarrow \mathcal{H} := L^2(\mathbb{T}_*^2, \mathbb{C}^2), \quad (2.10)$$

where  $\mathcal{H}$  is equipped with the inner product (notice the normalization)

$$\langle \varphi_1, \varphi_2 \rangle_{\mathcal{H}} := \frac{1}{|\mathbb{T}_*^2|} \int_{\mathbb{T}_*^2} d\mathbf{k} \langle \varphi_1(\mathbf{k}), \varphi_2(\mathbf{k}) \rangle_{\mathbb{C}^2}.$$

Every operator  $A$  acting in  $\ell^2(\mathcal{C})$  which is  $\Gamma$ -periodic, in the sense that

$$[A, T_\gamma] = 0 \text{ for every } \gamma \in \Gamma, \quad (2.11)$$

is conjugated to an operator  $\mathcal{F}_\nu A \mathcal{F}_\nu^{-1} =: A_\nu$  acting in  $\mathcal{H}$ . Notice that  $A_\nu$  is decomposable<sup>3</sup> in the sense that

$$(A_\nu \varphi)(\mathbf{k}) = A(\mathbf{k})\varphi(\mathbf{k}) \quad \text{for all } \mathbf{k} \in \mathbb{T}_*^2$$

where  $\mathbb{T}_*^2 \ni \mathbf{k} \mapsto A(\mathbf{k}) \in \mathcal{B}(\mathbb{C}^2)$  due to (2.10). Moreover, the  $\Gamma$ -periodicity of  $A$  reflects in the following property:

$$A(\mathbf{k} + \lambda) = A(\mathbf{k}) \quad \text{for all } \lambda \in \Gamma^*, \mathbf{k} \in \mathbb{T}_*^2. \quad (2.12)$$

The latter is understood as an equality of matrices. The matrix  $A(\mathbf{k})$  is called the *fiber of the operator  $A$*  at the point  $\mathbf{k}$ , and we use the notation  $A \longleftrightarrow A(\mathbf{k})$  to indicate the correspondence between the  $\Gamma$ -periodic operator  $A$  and the operator  $\mathcal{F}_\nu A \mathcal{F}_\nu^{-1}$ , acting in  $L^2(\mathbb{T}_*^2, \mathbb{C}^2)$ , given fiberwise by the multiplication operator times the (matrix-valued) function  $A(\mathbf{k})$ . Notice that everything above depends – in general – on the choice of a dimerization, as the subscript in  $\mathcal{F}_\nu$  suggests.

The Haldane Hamiltonian (2.3) is  $\Gamma$ -periodic, and its fibers  $H(\mathbf{k})$  over the Brillouin torus can be conveniently decomposed on the Pauli basis  $\{\sigma_0 = \mathbb{I}, \sigma_1, \sigma_2, \sigma_3\}$  as

$$H(\mathbf{k}) = \sum_{j=0}^3 R_j(\mathbf{k}) \sigma_j.$$

It is easy to show that

$$R_0(\mathbf{k}) = 2t_2(\cos \phi) \sum_{j=1}^3 \cos(\mathbf{k} \cdot \mathbf{a}_j), \quad (2.13)$$

$$R_3(\mathbf{k}) = M - 2t_2(\sin \phi) \sum_{j=1}^3 \sin(\mathbf{k} \cdot \mathbf{a}_j). \quad (2.14)$$

Indeed, one exploits the fact that the Fourier transform intertwines the translation operator  $T_\gamma$ , for  $\gamma \in \Gamma$ , with the multiplication times  $e^{i\mathbf{k} \cdot \gamma} \mathbb{I}$ . Since  $T_{\mathbf{a}_j} \longleftrightarrow e^{i\mathbf{k} \cdot \mathbf{a}_j} \mathbb{I}$ , one concludes that

$$T_{\mathbf{a}_j} + T_{-\mathbf{a}_j} \longleftrightarrow 2 \cos(\mathbf{k} \cdot \mathbf{a}_j) \mathbb{I}$$

which immediately gives (2.13). Analogously, since  $(\chi_A - \chi_B)T_{\mathbf{a}_j} \longleftrightarrow e^{i\mathbf{k} \cdot \mathbf{a}_j} \sigma_3$ , one concludes that

$$(\chi_A - \chi_B)(T_{\mathbf{a}_j} - T_{-\mathbf{a}_j}) \longleftrightarrow 2i \sin(\mathbf{k} \cdot \mathbf{a}_j) \sigma_3$$

which gives (2.14). Notice that the previous terms do not depend on a specific choice of the dimerization, provided one of the sublattices agrees with  $\Gamma$ .

---

<sup>3</sup>For the sake of brevity we omit the dependence of the operator  $A_\nu$  on the dimerization procedure, i. e. we remove the subscript  $\nu$ .

As for the off-diagonal terms, one has however to be more careful, since the computation *does depend on the choice of “the” periodicity cell*, as pointed out for example in [2, 7]. We make here the specific choice

$$Y = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 \text{ with } \alpha_j \in [-\frac{1}{2}, +\frac{1}{2}]\} \quad (2.15)$$

so that  $Y \cap \mathcal{C} = \{\mathbf{0}, \mathbf{d}_3\}$ , as illustrated in the first panel in Figure 2. One has that  $\mathcal{C} \cong \Gamma \times \{\mathbf{0}, \mathbf{d}_3\}$  as a measure space, and the dimerization isomorphism is exhibited by

$$\Psi_\gamma = \begin{pmatrix} \psi_{\gamma,A} \\ \psi_{\gamma,B} \end{pmatrix} = \begin{pmatrix} \psi_{\gamma+\mathbf{0}} \\ \psi_{\gamma+\mathbf{d}_3} \end{pmatrix}. \quad (2.16)$$

With this specific choice, the remaining terms are

$$R_1(\mathbf{k}) = t_1 (1 + \cos(\mathbf{k} \cdot \mathbf{a}_1) + \cos(\mathbf{k} \cdot \mathbf{a}_2)), \quad (2.17)$$

$$R_2(\mathbf{k}) = t_1 (\sin(\mathbf{k} \cdot \mathbf{a}_1) - \sin(\mathbf{k} \cdot \mathbf{a}_2)). \quad (2.18)$$

These expressions are easily derived. By using (2.2), one computes

$$\begin{aligned} (T_{+\mathbf{d}_3} \psi)_\gamma &= \begin{pmatrix} \psi_{\gamma-\mathbf{d}_3} \\ \psi_{(\gamma+\mathbf{d}_3)-\mathbf{d}_3} \end{pmatrix} = \begin{pmatrix} 0 \\ \psi_{\gamma,A} \end{pmatrix} \\ (T_{-\mathbf{d}_3} \Psi)_\gamma &= \begin{pmatrix} \psi_{\gamma+\mathbf{d}_3} \\ \psi_{(\gamma+\mathbf{d}_3)+\mathbf{d}_3} \end{pmatrix} = \begin{pmatrix} \psi_{\gamma,B} \\ 0 \end{pmatrix}. \end{aligned}$$

Thus  $T_{\mathbf{d}_3} + T_{-\mathbf{d}_3} = \mathbf{1} \otimes \sigma_1$ , so that the Fourier transform  $\mathcal{F}_{\mathbf{d}_3}$  yields

$$T_{\mathbf{d}_3} + T_{-\mathbf{d}_3} \longleftrightarrow \mathbf{1} \otimes \sigma_1. \quad (2.19)$$

The coordinate  $j = 3$  is privileged in view of our choice of the periodicity cell.

As for the next term, one uses that  $\mathbf{a}_1 = \mathbf{d}_2 - \mathbf{d}_3$  so that

$$\begin{aligned} (T_{+\mathbf{d}_2} \psi)_\gamma &= \begin{pmatrix} \psi_{\gamma-\mathbf{d}_3-\mathbf{a}_1} \\ \psi_{(\gamma+\mathbf{d}_3)-\mathbf{d}_3-\mathbf{a}_1} \end{pmatrix} = \begin{pmatrix} 0 \\ \psi_{\gamma-\mathbf{a}_1,A} \end{pmatrix} \\ (T_{-\mathbf{d}_2} \psi)_\gamma &= \begin{pmatrix} \psi_{\gamma+\mathbf{d}_3+\mathbf{a}_1} \\ \psi_{(\gamma+\mathbf{d}_3)+\mathbf{d}_3+\mathbf{a}_1} \end{pmatrix} = \begin{pmatrix} \psi_{\gamma+\mathbf{a}_1,B} \\ 0 \end{pmatrix}. \end{aligned}$$

After Fourier transform one obtains

$$(\mathcal{F}_{\mathbf{d}_3}(T_{\mathbf{d}_2} + T_{-\mathbf{d}_2})\psi)(\mathbf{k}) = \begin{pmatrix} 0 & e^{-i\mathbf{k} \cdot \mathbf{a}_1} \\ e^{+i\mathbf{k} \cdot \mathbf{a}_1} & 0 \end{pmatrix} (\mathcal{F}_{\mathbf{d}_3}\psi)(\mathbf{k}).$$

Analogously, in view of  $\mathbf{a}_2 = \mathbf{d}_3 - \mathbf{d}_1$  one has

$$\begin{aligned} (T_{+\mathbf{d}_1} \psi)_\gamma &= \begin{pmatrix} \psi_{\gamma-\mathbf{d}_3+\mathbf{a}_2} \\ \psi_{(\gamma+\mathbf{d}_3)-\mathbf{d}_3+\mathbf{a}_2} \end{pmatrix} = \begin{pmatrix} 0 \\ \psi_{\gamma+\mathbf{a}_2,A} \end{pmatrix} \\ (T_{-\mathbf{d}_1} \psi)_\gamma &= \begin{pmatrix} \psi_{\gamma+\mathbf{d}_3-\mathbf{a}_2} \\ \psi_{(\gamma+\mathbf{d}_3)+\mathbf{d}_3-\mathbf{a}_2} \end{pmatrix} = \begin{pmatrix} \psi_{\gamma-\mathbf{a}_2,B} \\ 0 \end{pmatrix} \end{aligned}$$

which gives

$$(\mathcal{F}_{\mathbf{d}_3}(T_{\mathbf{d}_1} + T_{-\mathbf{d}_1})\psi)(\mathbf{k}) = \begin{pmatrix} 0 & e^{+i\mathbf{k}\cdot\mathbf{a}_2} \\ e^{-i\mathbf{k}\cdot\mathbf{a}_2} & 0 \end{pmatrix} (\mathcal{F}_{\mathbf{d}_3}\psi)(\mathbf{k}).$$

Summarizing the information above, one concludes that

$$\sum_{j=1}^3 (T_{\mathbf{d}_j} + T_{-\mathbf{d}_j}) \iff \begin{pmatrix} 0 & 1 + e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{+i\mathbf{k}\cdot\mathbf{a}_2} \\ 1 + e^{+i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2} & 0 \end{pmatrix} \tag{2.20}$$

which immediately gives (2.17) and (2.18).

### 3 Bloch functions and their singularities

In this Section, we will be interested in studying the spectral properties of the Haldane Hamiltonian, which we rewrite as

$$H(\mathbf{k}) = \sum_{j=0}^3 R_j(\mathbf{k}) \sigma_j = \begin{pmatrix} R_0(\mathbf{k}) + R_3(\mathbf{k}) & \overline{R(\mathbf{k})} \\ R(\mathbf{k}) & R_0(\mathbf{k}) - R_3(\mathbf{k}) \end{pmatrix},$$

where we have abbreviated

$$R(\mathbf{k}) := R_1(\mathbf{k}) + i R_2(\mathbf{k}) = t_1(1 + e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2})$$

(compare (2.20)). It is then immediate to see that the eigenvalues of  $H(\mathbf{k})$  are given by

$$E_{\pm}(\mathbf{k}) := R_0(\mathbf{k}) \pm \sqrt{\sum_{j=1}^3 R_j(\mathbf{k})^2} = R_0(\mathbf{k}) \pm \sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2}.$$

These two energy bands will not overlap (that is,  $E_-(\mathbf{k}) \leq E_+(\mathbf{k})$  for all  $\mathbf{k} \in \mathbb{R}^2$ ) provided  $t_2/t_1 < 1/3$ , which we will assume hereafter. The bands can still touch at the points in the Brillouin torus which are determined by the equation

$$\sum_{j=1}^3 R_j(\mathbf{k})^2 = 0 \iff R(\mathbf{k}) = 0 \text{ and } R_3(\mathbf{k}) = 0.$$

We see then that there are (at most) two such points in the Brillouin torus, usually labeled  $\mathbf{K}$  and  $\mathbf{K}'$ , determined by the zeroes of  $R$ : these are obtained from the conditions

$$e^{i\mathbf{K}'\cdot\mathbf{a}_1} = e^{i2\pi/3} \text{ and } e^{-i\mathbf{K}'\cdot\mathbf{a}_2} = e^{-i2\pi/3}, \quad e^{i\mathbf{K}\cdot\mathbf{a}_1} = e^{-i2\pi/3} \text{ and } e^{-i\mathbf{K}\cdot\mathbf{a}_2} = e^{i2\pi/3},$$

which in particular imply  $\mathbf{K}' = -\mathbf{K} \pmod{\Gamma^*}$ . Since locally around these points the dispersion of the energy bands is linear when they produce band intersections, *i. e.*

$E_{\pm}(\mathbf{k}) = E_{\pm}(\mathbf{K}^{\sharp}) \pm v_F |\mathbf{k} - \mathbf{K}^{\sharp}| + O(|\mathbf{k} - \mathbf{K}^{\sharp}|^2)$  for  $\mathbf{K}^{\sharp} \in \{\mathbf{K}, \mathbf{K}'\}$ , the points  $\mathbf{K}$  and  $\mathbf{K}'$  are usually called *Dirac points*. The equation  $R_3(\mathbf{k}) = 0$  then determines the locus in the space of parameters  $(\phi, M)$  where either  $\mathbf{K}$  or  $\mathbf{K}'$  (or both) are points of degeneracy for the eigenvalues of the Haldane Hamiltonian, namely

$$R_3(\mathbf{K}) = M + 3\sqrt{3}t_2 \sin \phi, \quad R_3(\mathbf{K}') = M - 3\sqrt{3}t_2 \sin \phi.$$

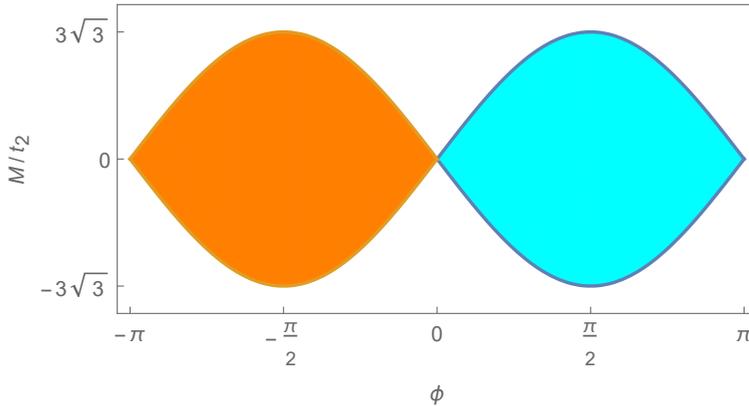


Figure 3: The topological phase diagram of the Haldane model. In cyan, the region  $\{R_3(\mathbf{K}) > 0, R_3(\mathbf{K}') < 0\}$ , characterized by a Chern number  $c_1 = -1$ ; in orange, the region  $\{R_3(\mathbf{K}) < 0, R_3(\mathbf{K}') > 0\}$ , characterized by a Chern number  $c_1 = +1$  (color online). In the rest of the phase diagram,  $c_1 = 0$ .

We see that the parameter space  $(\phi, M)$  gets divided into four regions where the Hamiltonian is gapped (see Figure 3), characterized by the signs of  $R_3(\mathbf{K})$  and  $R_3(\mathbf{K}')$ . We will show now how it is possible to assign a *topological label* to each of the four gapped phases, determining also the “quantum anomalous Hall conductivity” of the Haldane model for all parameters in the region. To this end, it is convenient to introduce the eigenvector  $u_-(\mathbf{k})$ , that is, the *Bloch function*, associated to the lower band  $E_-(\mathbf{k})$  of the Haldane Hamiltonian. This reads

$$u_-(\mathbf{k}) = N(\mathbf{k})^{-1} \begin{pmatrix} \sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2} - R_3(\mathbf{k}) \\ -R(\mathbf{k}) \end{pmatrix},$$

where  $N(\mathbf{k}) := \left[ 2\sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2} \left( \sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2} - R_3(\mathbf{k}) \right) \right]^{1/2}$  is a normalizing factor ensuring  $\|u_-(\mathbf{k})\|_{\mathbb{C}^2} = 1$  for all  $\mathbf{k} \in \mathbb{R}^2$  (compare [8, Appendix B]). The *Bloch gauge* (that is, the phase within the complex one-dimensional eigenspace associated to the lower energy band) is chosen so that the first component  $u_{-,1}(\mathbf{k})$  is real.

If  $\mathbf{K}^{\sharp}$  denotes either of the Dirac points, then  $R(\mathbf{K}^{\sharp})$  vanishes, as  $\mathbf{K}$  and  $\mathbf{K}'$  are precisely the zeroes of  $R$ , while  $R_3(\mathbf{K}^{\sharp}) \neq 0$  due to the gap condition. Conse-

quently,

$$\sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2} - R_3(\mathbf{k}) \Big|_{\mathbf{k}=\mathbf{K}^\sharp} = |R_3(\mathbf{K}^\sharp)| - R_3(\mathbf{K}^\sharp),$$

and similarly

$$N(\mathbf{K}^\sharp) = [2|R_3(\mathbf{K}^\sharp)| (|R_3(\mathbf{K}^\sharp)| - R_3(\mathbf{K}^\sharp))]^{1/2}.$$

We see that  $u_-(\mathbf{k})$  may have singularities at the Dirac points, depending on the signs of  $R_3(\mathbf{K})$  and  $R_3(\mathbf{K}')$ . In particular, it holds that  $u_-(\mathbf{k})$  is *singular at  $\mathbf{K}$*  in the region  $\{R_3(\mathbf{K}) > 0, R_3(\mathbf{K}') < 0\}$  of the parameter space  $(\phi, M)$  (depicted in cyan in Figure 3), while it is *analytic on the whole Brillouin torus* in the region  $\{R_3(\mathbf{K}) < 0, R_3(\mathbf{K}') < 0\}$  (the lower white region in Figure 3). The qualitative features of this singularity (or lack thereof) are illustrated in Figures 4 and 5.

To investigate further the singularity of  $u_-(\mathbf{k})$ , we restrict our attention to parameters  $(\phi, M)$  so that  $R_3(\mathbf{K}) > 0$  and  $R_3(\mathbf{K}') < 0$ . As discussed, in this region  $\mathbf{K}$  is the only singular point of  $u_-(\mathbf{k})$ . By rewriting, after a few simple algebraic manipulations,

$$u_-(\mathbf{k}) = \frac{1}{\sqrt{2}} \left( \begin{array}{c} \frac{|R(\mathbf{k})|}{(R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2)^{1/4} \left( \sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2} + R_3(\mathbf{k}) \right)^{1/2}} \\ - \frac{R(\mathbf{k})}{|R(\mathbf{k})|} \left( 1 + \frac{R_3(\mathbf{k})}{\sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2}} \right)^{1/2} \end{array} \right),$$

we see that in this region the first component of  $u_-(\mathbf{k})$  is smooth, while it is the second component that has a singularity, due to the explicit dependence on the phase of  $R(\mathbf{k})$ . This implies in particular that, locally around  $\mathbf{k} = \mathbf{K}$ ,  $u_{-,2}(\mathbf{k})$  is homogeneous of degree zero in the radial coordinate  $r = |\mathbf{k} - \mathbf{K}|$ , so that the derivatives of  $u_{-,2}(\mathbf{k})$  have a  $(1/r)$ -singularity, making the  $H^1$ -norm of  $u_-$

$$\begin{aligned} \|u_-\|_{H^1} &:= \left( \|u_-\|_{L^2}^2 + \|\partial_{k_1} u_-\|_{L^2}^2 + \|\partial_{k_2} u_-\|_{L^2}^2 \right)^{1/2} \\ &= \left[ \int_{\mathbb{T}_*^2} (\|u_-(\mathbf{k})\|_{\mathbb{C}^2}^2 + \|\partial_{k_1} u_-(\mathbf{k})\|_{\mathbb{C}^2}^2 + \|\partial_{k_2} u_-(\mathbf{k})\|_{\mathbb{C}^2}^2) d\mathbf{k} \right]^{1/2} \end{aligned}$$

divergent. From the same type of homogeneity argument, one can also deduce that all the fractional Sobolev norms  $\|u_-\|_{H^s}$  for  $s \in [0, 1)$  are instead finite.

The singularity of  $u_-(\mathbf{k})$  at  $\mathbf{k} = \mathbf{K}$  carries also a topological information. This can be accessed by means of the *Berry connection*, defined as the differential 1-form

$$\mathcal{A} := \text{Im} \langle u_-(\mathbf{k}), du_-(\mathbf{k}) \rangle_{\mathbb{C}^2} = \sum_{j=1}^2 \text{Im} \langle u_-(\mathbf{k}), \partial_{k_j} u_-(\mathbf{k}) \rangle_{\mathbb{C}^2} dk_j.$$

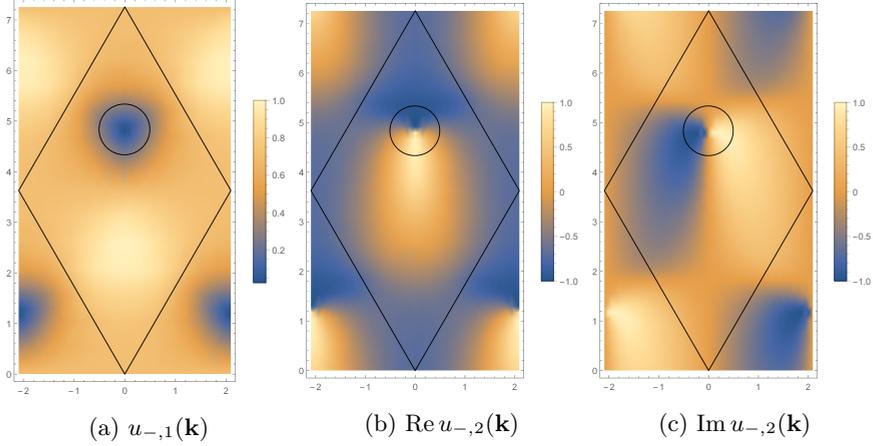
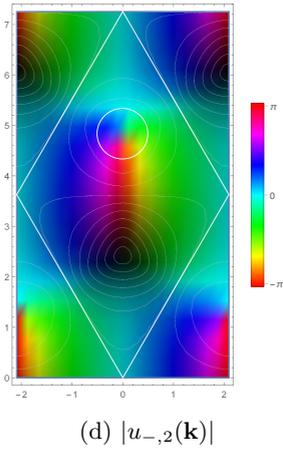


Figure 4: Density plots for the components of  $u_{-}(\mathbf{k})$  (color online). The parameters chosen to produce these plots are as follows:  $d = 1$  for the lattice constant,  $t_1 = 1$ ,  $t_2 = 1/4$ ,  $M = 0$ ,  $\phi = \pi/2$ . The rhomboidal region is the Brillouin zone  $\{k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 : k_1, k_2 \in [0, 1]\}$ , where the vectors  $\mathbf{b}_1, \mathbf{b}_2$  spanning the dual lattice  $\Gamma^*$  are determined by the conditions  $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$ . The circle points to the position of the Dirac point  $\mathbf{K}$ . The rapid change of both  $\text{Re } u_{-,2}(\mathbf{k})$  and  $\text{Im } u_{-,2}(\mathbf{k})$  around  $\mathbf{k} = \mathbf{K}$  are evident from (4b) and (4c), respectively, signalling a discontinuity. Instead,  $u_{-,1}(\mathbf{k})$  is seen to be regular (and actually vanishing) at  $\mathbf{k} = \mathbf{K}$  from (4a). In the last plot (4d), contour lines for the absolute value  $|u_{-,2}(\mathbf{k})|$  are plotted, while the color code indicates the value of the argument of the phase of  $u_{-,2}(\mathbf{k})$  as in the legend. In agreement with the previous comments, it is possible to see a phase singularity of  $u_{-,2}(\mathbf{k})$  around  $\mathbf{k} = \mathbf{K}$ , while the absolute value  $|u_{-,2}(\mathbf{k})| = \sqrt{1 - u_{-,1}(\mathbf{k})^2}$  remains smooth.



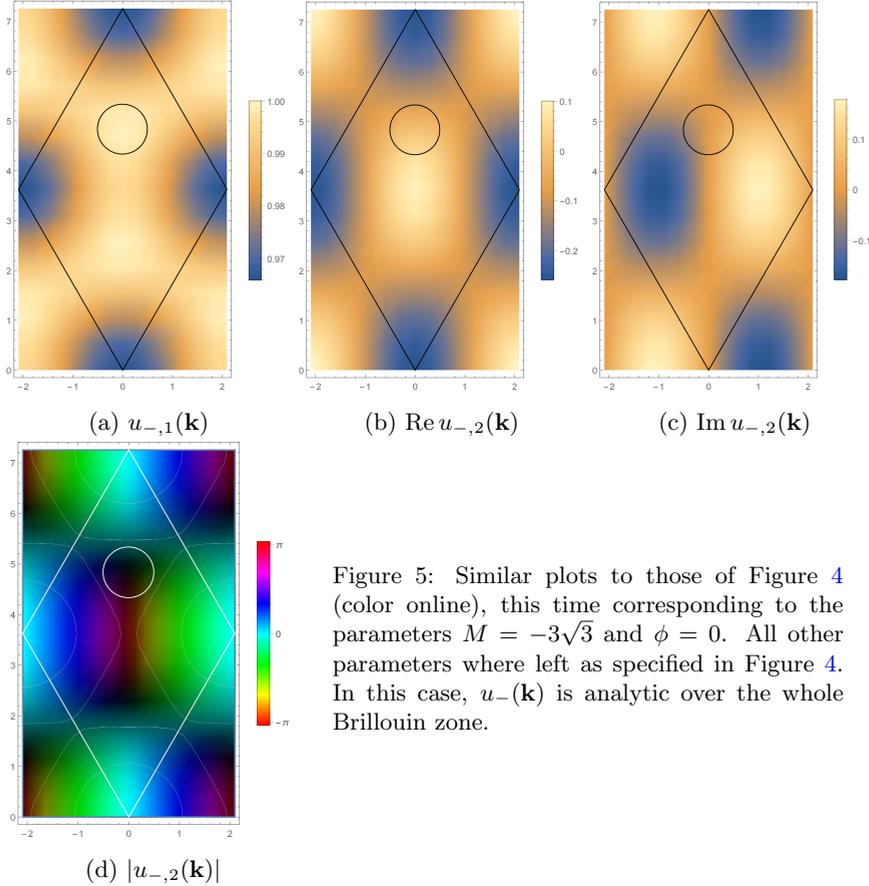


Figure 5: Similar plots to those of Figure 4 (color online), this time corresponding to the parameters  $M = -3\sqrt{3}$  and  $\phi = 0$ . All other parameters were left as specified in Figure 4. In this case,  $u_-(\mathbf{k})$  is analytic over the whole Brillouin zone.

We argue as above: the  $(1/r)$ -singularity of the derivatives of  $u_{-,2}(\mathbf{k})$  around  $\mathbf{k} = \mathbf{K}$  is integrable (even though not square-integrable), so that the integral of  $\mathcal{A}$  around a small loop  $\ell_\varepsilon$  (say, of diameter  $\varepsilon \ll 1$ ) encircling the singularity of  $u_-(\mathbf{k})$  stays bounded even in the limit  $\varepsilon \rightarrow 0$ . Denoting by  $D_\varepsilon$  the region bounded by the loop  $\ell_\varepsilon$  (which then bounds also  $\mathbb{T}_*^2 \setminus D_\varepsilon$ , as the Brillouin torus is closed), in the limit of a very small loop one obtains from Stokes' theorem

$$\lim_{\varepsilon \rightarrow 0} \oint_{\ell_\varepsilon} \mathcal{A} = - \lim_{\varepsilon \rightarrow 0} \oint_{\partial(\mathbb{T}_*^2 \setminus D_\varepsilon)} \mathcal{A} = - \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{T}_*^2 \setminus D_\varepsilon} d\mathcal{A} = - \int_{\mathbb{T}_*^2} \mathcal{F}. \quad (3.1)$$

In the last step, we introduced the *Berry curvature* 2-form

$$\begin{aligned} \mathcal{F} &:= d\mathcal{A} = 2 \operatorname{Im} \langle \partial_{k_1} u_-(\mathbf{k}), \partial_{k_2} u_-(\mathbf{k}) \rangle dk_1 \wedge dk_2 \\ &= -i \operatorname{Tr}_{\mathbb{C}^2} (P_-(\mathbf{k}) [\partial_{k_1} P_-(\mathbf{k}), \partial_{k_2} P_-(\mathbf{k})]) dk. \end{aligned} \quad (3.2)$$

The last equality (see e. g. [19, Lemma 7.2]) shows that  $\mathcal{F}$  can be expressed directly

in terms of the family of projections

$$P_-(\mathbf{k}) := |u_-(\mathbf{k})\rangle \langle u_-(\mathbf{k})| = \frac{1}{2\sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2}} \begin{pmatrix} \sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2} - R_3(\mathbf{k}) & -R(\mathbf{k}) \\ -\bar{R}(\mathbf{k}) & \sqrt{R_3(\mathbf{k})^2 + |R(\mathbf{k})|^2} + R_3(\mathbf{k}) \end{pmatrix}$$

on the eigenspace corresponding to the lower energy band. Contrary to the Bloch function, these projections depend analytically on  $\mathbf{k}$  over the whole Brillouin torus, making it possible to compute the last limit in (3.1).

The Berry curvature is a geometric object. In fact, its integral over the Brillouin torus is an integer multiple of  $2\pi$ :

$$c_1 := \frac{1}{2\pi} \int_{\mathbb{T}_*^2} \mathcal{F} \in \mathbb{Z}. \quad (3.3)$$

This integer, called the *Chern number*, is the topological invariant which underlies the quantization of the (anomalous) Hall conductivity in Chern insulators [10, 4, 3, 5] and quantum Hall insulators [25, 9]. In the specific case under investigation of the Haldane Hamiltonian, the four regions of parameters  $(\phi, M)$  in which  $H(\mathbf{k})$  is gapped can be labelled by the Chern number [10]: with reference to the colors of Figure 3, the Chern number can be computed, *e.g.* starting from (3.1), to be  $c_1 = -1$  for the cyan region,  $c_1 = +1$  for the orange region, and  $c_1 = 0$  for the two white regions. In analogy with the thermodynamical phases of statistical mechanics, one then speaks of *topological phases of matter* distinguished by different topological invariants, and refers to Figure 3 as the *topological phase diagram* for the Haldane Hamiltonian.

**Remark 3.1.** It is interesting to notice that the topological content associated to singularities of the Bloch function at the Dirac points persists, in an appropriate sense, also in the *gapless* regime. If for example the parameters  $(\phi, M)$  are threaded from the cyan region to the lower white region of Figure 3, passing through a point in parameter space  $(\phi_*, M_*)$  where  $R_3(\mathbf{K}) = 0$ , then at  $(\phi_*, M_*)$  not only the Bloch function  $u_-$  but also the projection  $P_-$  becomes singular at  $\mathbf{K}$ . Nonetheless, the topological charge exchanged through the gapless phase can be quantified by means of a local topological invariant, the *eigenspace vorticity*, associated to family of projections  $P_-$  around the singular point  $\mathbf{K}$  [17]. In the situation described above, this eigenspace vorticity equals  $\Delta c_1 = 0 - (-1) = 1$ .

## 4 The localization dichotomy for periodic insulators

It is astounding to discover the predictive power of the Haldane model. In fact, it turns out that the features discussed in the previous Section are completely generic in two and three dimensions: indeed, the close connection between the

structure of the singularities of the Bloch functions and the topology of the associated eigenspaces persists in a much wider context and for more general models.

This was recently proved and quantified in a precise way in [19]. To formulate the main result, we need to set up the more general framework. Let  $d \leq 3$ . The configuration space of a crystalline system is modeled by the space  $X$ , which can be either  $\mathbb{R}^d$  or a  $d$ -dimensional crystalline structure (e.g. the honeycomb structure presented in Section 2):  $X$  carries an action of the lattice  $\Gamma \simeq \mathbb{Z}^d$  by translations, which is assumed to lift to translation operators  $T_\gamma \in U(L^2(X))$ ,  $\gamma \in \Gamma$ .

Associated to these translation operators, there is a *Bloch–Floquet–Zak transform*

$$\mathcal{U}: L^2(X) \rightarrow L^2_\tau(\mathbb{B}; L^2_{\text{per}}(Y)) \simeq \int_{\mathbb{B}}^\oplus L^2_{\text{per}}(Y) d\mathbf{k},$$

defined by

$$(\mathcal{U}\psi)(\mathbf{k}, \mathbf{y}) := \sum_{\gamma \in \Gamma} e^{-i\mathbf{k} \cdot (\mathbf{y} - \gamma)} (T_\gamma \psi)(\mathbf{y}), \quad \mathbf{k} \in \mathbb{B}, \mathbf{y} \in Y, \quad (4.1)$$

on suitable  $\psi \in L^2(X)$ . Here  $\mathbb{B}$  stands for the fundamental cell of the dual lattice  $\Gamma^*$  (the *Brillouin zone* in the physics literature),  $Y$  stands for the fundamental cell of the lattice  $\Gamma$  (compare (2.15)), and  $L^2_\tau(\mathbb{B}; L^2_{\text{per}}(Y))$  is the Hilbert space

$$L^2_\tau(\mathbb{B}; L^2_{\text{per}}(Y)) := \left\{ \begin{array}{l} u \in L^2_{\text{loc}}(\mathbb{R}^d; L^2_{\text{loc}}(\mathbb{R}^d)) : \\ u(\mathbf{k} + \lambda, \mathbf{y}) = (\tau_\lambda u)(\mathbf{k}, \mathbf{y}) := e^{-i\lambda \cdot \mathbf{y}} u(\mathbf{k}, \mathbf{y}) \\ \text{and } T_\gamma u(\mathbf{k}, \cdot) = u(\mathbf{k}, \cdot) \\ \forall \mathbf{k} \in \mathbb{R}^d, \mathbf{y} \in \mathbb{R}^d, \lambda \in \Gamma^*, \gamma \in \Gamma \end{array} \right\}$$

of functions of the *Bloch momentum*  $\mathbf{k}$  and of the degrees of freedom in the unit cell  $\mathbf{y}$  which are quasi-periodic ( $\tau$ -covariant) in  $\mathbf{k}$  and periodic in  $\mathbf{y}$  (see [18] for details). For crystalline structures of the type described in Section 2,  $\mathcal{U}$  coincides with the Fourier transform (2.9) up to the extra phase factor  $e^{-i\mathbf{k} \cdot \mathbf{y}}$  in (4.1), which turns periodic functions of  $\mathbf{k}$  into quasi-periodic, but makes the boundary conditions on the unit cell  $Y$  in direct space  $\mathbf{k}$ -independent (namely, exactly periodic). A Bloch–Floquet–Zak transform is defined by (4.1) also in the continuous case  $X = \mathbb{R}^d$ , where  $T_\gamma$  can be the standard translation  $(T_\gamma \psi)(\mathbf{y}) := \psi(\mathbf{y} - \gamma)$  or, more interestingly, a *magnetic translation* generated by a uniform magnetic field with flux per unit cell which is commensurate to the flux quantum (equal to  $2\pi$  in Hartree units), see [26] and the discussion in [19, Sec. 3]. Also in this case we will denote by  $A \longleftrightarrow A(\mathbf{k})$  the correspondence between a periodic operator  $A$  on  $L^2(X)$  such that  $[A, T_\gamma] = 0$  for all  $\gamma \in \Gamma$  and its decomposition into fibers in the Bloch–Floquet–Zak representation:  $\mathcal{U} A \mathcal{U}^{-1} = \int_{\mathbb{B}}^\oplus A(\mathbf{k}) d\mathbf{k}$ .

Now that the framework of crystalline systems is clear, we can formulate the main hypothesis of the central result from [19], which abstracts the predominant features of the Haldane Hamiltonian described in the previous Sections.

**Assumption.** Let  $H$  be a periodic self-adjoint operator on  $L^2(X)$  with  $H \longleftrightarrow H(\mathbf{k})$  where  $H(\mathbf{k})$  defines a family of operators on  $L^2_{\text{per}}(Y)$  such that

- (i)  $\{H(\boldsymbol{\kappa})\}_{\boldsymbol{\kappa} \in \mathbb{C}^d}$  defines an *entire analytic family in the sense of Kato* with compact resolvent [23];
- (ii) the family is  $\tau$ -*covariant*, that is,  $H(\mathbf{k} + \lambda) = \tau_\lambda H(\mathbf{k}) \tau_\lambda^{-1}$  for all  $\mathbf{k} \in \mathbb{R}^d$  and  $\lambda \in \Gamma^*$ ;
- (iii) the family is *gapped*, namely there exists a set  $\mathcal{I} \subset \mathbb{N}$  with  $|\mathcal{I}| = m < \infty$  such that

$$\inf_{\mathbf{k} \in \mathbb{R}^d} \inf_{\substack{n \in \mathcal{I} \\ m \in \mathbb{N} \setminus \mathcal{I}}} |E_n(\mathbf{k}) - E_m(\mathbf{k})| \geq g > 0$$

where  $\sigma(H(\mathbf{k})) = \{E_n(\mathbf{k})\}_{n \in \mathbb{N}}$  denotes the spectrum of  $H(\mathbf{k})$  (consisting of discrete eigenvalues, the *Bloch bands*, by the compact resolvent assumption).

In the discrete case (e.g. for the Haldane Hamiltonian), the regularity assumption is easy to verify, as it is equivalent in position space to having sufficiently fast decaying hoppings between different sites of the crystal (say, exponential in the distance between the sites), and thus is in particular satisfied whenever the hopping Hamiltonian has finite range, as often happens in applications. For (magnetic, periodic) Schrödinger operators, there are standard  $L^p$ -regularity assumptions on the electro-magnetic potentials that guarantee analyticity of the corresponding fiber Hamiltonians [23].

Notice moreover that the gap assumption allows to define the family of spectral projections  $P(\mathbf{k})$  onto the spectral island  $\sigma_0(\mathbf{k}) := \{E_n(\mathbf{k}) : n \in \mathcal{I}\}$ , for example through the *Riesz formula*

$$P(\mathbf{k}) = \frac{i}{2\pi} \oint_{C(\mathbf{k})} (H(\mathbf{k}) - z)^{-1} dz,$$

where  $C(\mathbf{k})$  is a positively oriented contour in the complex energy plane, locally constant in  $\mathbf{k}$ , which lies in the resolvent set of  $H(\mathbf{k})$  and encircles only the eigenvalues in  $\sigma_0(\mathbf{k})$ . This family of projections is then  $\tau$ -covariant and depends analytically on  $\boldsymbol{\kappa} \in \Omega_\alpha$ , where  $\Omega_\alpha \subset \mathbb{C}^d$  is a complex strip of half-width  $\alpha > 0$  around the “real axis”  $\mathbb{R}^d \subset \mathbb{C}^d$  [22, Prop. 2.1].

As in (3.3), we can define the *Chern numbers* associated to  $\{P(\mathbf{k})\}_{\mathbf{k} \in \mathbb{R}^d}$  as

$$c_1(P)_{ij} := \frac{1}{2\pi} \int_{\mathbb{B}_{ij}} \text{Tr}_{L^2_{\text{per}}(Y)} (P(\mathbf{k}) [\partial_{k_i} P(\mathbf{k}), \partial_{k_j} P(\mathbf{k})]) dk_i \wedge dk_j, \quad 1 \leq i < j \leq d, \tag{4.2}$$

where  $\mathbb{B}_{ij} \subset \mathbb{B}$  is the 2-dimensional sub-torus of  $\mathbb{B}$  where the coordinate different from the  $i$ -th and  $j$ -th is fixed (e.g. to zero).

We are finally able to state the main result from [19], generalizing the analysis on the Haldane Hamiltonian from the previous Section.

**Theorem 4.1** ([19]). *Let  $H \longleftrightarrow H(\mathbf{k})$  be as in the above Assumption, and  $P(\mathbf{k})$  be the spectral projection on the gapped spectral island of  $H(\mathbf{k})$ . Then for all  $s \in [0, 1)$  there exists a Bloch frame  $\{u_1, \dots, u_m\} \subset H^s_\tau(\mathbb{B}; L^2_{\text{per}}(Y))$  for  $\{P(\mathbf{k})\}_{\mathbf{k} \in \mathbb{R}^d}$ , namely a set of functions  $u_a \in H^s_{\text{loc}}(\mathbb{R}^d; L^2_{\text{per}}(Y))$  such that*

$$u_a(\mathbf{k} + \lambda) = \tau_\lambda u_a(\mathbf{k}), \quad \langle u_a, u_b \rangle_{L^2} = \delta_{ab}, \quad \text{and} \quad P(\mathbf{k}) = \sum_{a=1}^m |u_a(\mathbf{k})\rangle \langle u_a(\mathbf{k})|.$$

Moreover, the following statements are equivalent:

- (i) there exists a Bloch frame in  $H^1_\tau(\mathbb{B}; L^2_{\text{per}}(Y))$ ;
- (ii) there exists a Bloch frame in  $C^\omega_\tau(\Omega_\alpha; L^2_{\text{per}}(Y))$ , the space of  $\tau$ -covariant analytic functions on  $\Omega_\alpha$  with values in  $L^2_{\text{per}}(Y)$ ;
- (iii) the Chern numbers  $c_1(P)_{ij}$ ,  $1 \leq i < j \leq d$ , defined in (4.2), vanish.

The above result can be interpreted as a *Localization–Topology Correspondence*, having implications also for the transport properties of the model under scrutiny for a crystalline insulator. To better clarify this point, we need to introduce one further notion. Given a periodic Hamiltonian  $H \longleftrightarrow H(\mathbf{k})$  as in the Assumption above, denote by  $P = \mathcal{U}^{-1} \left( \int_{\mathbb{B}}^\oplus P(\mathbf{k}) d\mathbf{k} \right) \mathcal{U}$  the periodic projection on  $L^2(X)$  onto the subspace corresponding to the isolated spectral island in momentum space. The Hamiltonian  $H$  has generically absolutely continuous spectrum (given by  $\sigma(H) = \{ \lambda \in \mathbb{R} : \lambda = E_n(\mathbf{k}) \text{ for some } n \in \mathbb{N}, \mathbf{k} \in \mathbb{R}^d \}$ ), so it is not possible in general to find a basis of the range of  $P$  given by eigenstates of the Hamiltonian. Nonetheless, if  $\{u_a(\mathbf{k})\}_{1 \leq a \leq m}$  is an orthonormal basis for  $\text{Ran } P(\mathbf{k})$  — a Bloch frame, in the terminology introduced above — then it is possible to define (*composite*) *Wannier functions* [16] by

$$w_a(\mathbf{y} - \gamma) := (\mathcal{U}^{-1} u_a)(\mathbf{y} - \gamma) = \frac{1}{|\mathbb{B}|} \int_{\mathbb{B}} e^{i\mathbf{k} \cdot (\mathbf{y} - \gamma)} u_a(\mathbf{k}, \mathbf{y}) d\mathbf{k}, \quad \begin{array}{l} 1 \leq a \leq m, \\ \mathbf{y} \in Y, \gamma \in \Gamma. \end{array}$$

The functions  $w_a$  will automatically be in  $\text{Ran } P \subset L^2(X)$ , and so will the translates  $T_\gamma w_a$  by periodicity of  $P$ . One can then check [14] that  $\{T_\gamma w_a\}_{\gamma \in \Gamma, 1 \leq a \leq m}$  constitutes an orthonormal basis for  $\text{Ran } P$  if the Bloch frame is  $\tau$ -covariant. *Localized* Wannier functions are found to describe accurately the orbitals of the crystalline insulator [16], and it is hence important to understand their decay properties at infinity. Since the Bloch–Floquet–Zak transform shares with the standard Fourier transform the property of intertwining the multiplication operator by  $\mathbf{x}$  on  $L^2(X)$  and the gradient  $\nabla_{\mathbf{k}}$  with respect to the crystal momentum, one can read off these decay properties of Wannier functions by looking at the smoothness with respect to  $\mathbf{k}$  of the corresponding Bloch frame. More precisely, it holds that

$$\begin{aligned} \langle \mathbf{x} \rangle^s w_a \in L^2(X) &\iff u_a \in H^s(\mathbb{B}; L^2_{\text{per}}(Y)), \quad s \geq 0, \\ e^{\beta|\mathbf{x}|} w_a \in L^2(X), \forall \beta \in [0, \alpha) &\iff u_a \in C^\omega(\Omega_\alpha; L^2_{\text{per}}(Y)), \end{aligned}$$

where we have denoted  $\langle \mathbf{x} \rangle := (1 + |\mathbf{x}|^2)^{1/2}$ .

The existence of a basis of well-localized (say, exponentially) Wannier functions signals the absence of charge transport in the crystal; on the contrary, a power-law decay of the Wannier functions is an indication of topological transport. If the Hall conductivity is non-zero, one then expects Wannier functions to be poorly localized. This is exactly the content of the above Theorem, which can be recast in terms of Wannier functions as a *Localization Dichotomy*: either Wannier functions are exponentially localized (and this happens exactly when the Hall conductivity vanishes), or they are delocalized in the sense that they yield an infinite expectation of the squared position operator  $|\mathbf{x}|^2$ ; no intermediate regimes of decay are allowed. The precise result is as follows.

**Theorem 4.2** (Localization Dichotomy [19]). *Let  $H$  be as in the above Assumption, and  $P$  be the spectral projection on the gapped spectral island. Then for all  $s \in [0, 1)$  there exists a Wannier basis for  $\text{Ran } P$ , that is, an orthonormal basis  $\{T_\gamma w_a\}_{\gamma \in \Gamma, 1 \leq a \leq m}$  of  $\text{Ran } P$ , such that*

$$\sup_{\gamma \in \Gamma} \int_X \langle \mathbf{x} - \gamma \rangle^{2s} |(T_\gamma w_a)(\mathbf{x})|^2 d\mathbf{x} \leq C_s < \infty \quad \text{for all } s \in [0, 1).$$

Moreover, the following statements are equivalent:

(i) *there exists a Wannier basis such that*

$$\sup_{\gamma \in \Gamma} \int_X \langle \mathbf{x} - \gamma \rangle^2 |(T_\gamma w_a)(\mathbf{x})|^2 d\mathbf{x} \leq C_1 < \infty;$$

(ii) *there exists a Wannier basis such that*

$$\sup_{\gamma \in \Gamma} \int_X e^{2\beta|\mathbf{x}-\gamma|} |(T_\gamma w_a)(\mathbf{x})|^2 d\mathbf{x} \leq C_\omega < \infty \quad \text{for all } \beta \in [0, \alpha);$$

(iii) *the Chern numbers  $c_1(P)_{ij}$ ,  $1 \leq i < j \leq d$ , defined in (4.2), vanish.*

We sketch here the main ideas from the proof of Theorem 4.1: for the detailed argument, the reader is referred to [19].

The first part consists in exhibiting a Bloch frame which is in  $H_\tau^s$  for all  $s \in [0, 1)$ . In  $2d$ , this is obtained via *parallel transport*, a procedure which allows to construct a smooth ( $C^\infty$ ) and  $\tau$ -covariant Bloch frame on the 1-dimensional boundary of the Brillouin zone  $\mathbb{B}$ , and to extend this to the interior. The end result is a Bloch frame which is  $\tau$ -covariant and smooth except at one point in the Brillouin zone. The technique of parallel transport gives a precise control also on the type of singularity of the constructed Bloch frame, which is seen to be consistent with the claimed  $H^s$ -regularity (the derivatives of the Bloch functions have a  $(1/r)$ -divergence at the singular point). This situation should be compared

with the Bloch function for the Haldane Hamiltonian exhibited in the previous Section. In  $3d$ , one needs to further extend an already singular datum at the 2-dimensional boundary of the Brillouin zone to the 3-dimensional “bulk”: this can be done again by parallel transport, and produces this time lines of singularities, which dictate in turn the  $H^s$ -regularity in the statement of the Theorem.

The next part of the proof requires to show that if a Bloch frame in  $H^1_\tau$  exists, then the Chern numbers of the family of projections vanish. The proof relies on a very subtle approximation of  $H^1_\tau$  frames by  $C^\infty_\tau$  frames. The subtlety lies in the fact that the space of frames in an Hilbert space is a non-linear manifold; the approximation of Sobolev maps with values in a manifold by regular maps becomes more involved, and requires in general certain topological conditions to be satisfied (see [11] and [19, App. B]). Nonetheless, in our setting  $H^1$  maps can indeed be approximated by  $C^\infty$  ones; when calculating an “approximate” Berry curvature (3.2) with the regular frames, its integrals over the tori  $\mathbb{B}_{ij}$  are zero, so that in the limit the Chern numbers for the family of projections  $P(\mathbf{k})$  must also vanish. It is then well-known [21, 22] how to modify the  $H^1$ -regular Bloch frame to an analytic one, provided the Chern numbers vanish.

As a side remark, note how in  $2d$  the “threshold” Sobolev regularity  $H^1$  coincides also with the “threshold” of the Sobolev embedding  $H^s \hookrightarrow C^0$ , which holds for  $s > 1$ . Geometric arguments, based on the theory of vector bundles, yield that a non-zero Chern number forbids the existence of  $\tau$ -covariant *continuous* Bloch frames [21]: Theorem 4.1 improves this result, claiming that also Bloch frames in  $H^1_\tau$  cannot exist when the Chern numbers are non-vanishing. In  $3d$ , the result is even more stringent, as the threshold for the Sobolev embedding of  $H^s$  into continuous functions is at  $s = 3/2$ .

**Acknowledgements.** We are grateful to Clément Tauber for many useful discussions on the related Kane-Mele model, and for his precious help with some figures.

## References

- [1] Ando, Y.: Topological insulator materials. *J. Phys. Soc. Jpn.* **82**, 102001 (2013)
- [2] Bena, C., Montambaux, G.: Remarks on the tight-binding model of graphene. *New J. Phys.* **11**, 095003 (2009)
- [3] Bestwick, A.J., Fox, E.J., Kou, X., Pan, L., Wang, K.L., Goldhaber-Gordon, D.: Precise quantization of the anomalous Hall effect near zero magnetic field. *Phys. Rev. Lett.* **114**, 187201 (2015)
- [4] Chang, C.Z. *et al.*: Experimental Observation of the Quantum Anomalous Hall Effect in a Magnetic Topological Insulator. *Science* **340**, 167–170 (2013)
- [5] Chang, C.Z. *et al.*: High-precision realization of robust quantum anomalous Hall state in a hard ferromagnetic topological insulator. *Nat. Mat.* **14**, 473 (2015)
- [6] Fruchart, M., Carpentier, D.: An introduction to topological insulators. *C. R. Phys.* **14**, 779–815 (2013)
- [7] Fruchart, M., Carpentier, D., Gawedzki, K.: Parallel Transport and Band Theory in Crystals. *EPL* **106** 60002 (2014)

- [8] Giuliani, A., Mastropietro, V., Porta, M.: Universality of Hall conductivity in interacting electron systems. *Commun. Math. Phys.* **349**, 1107–1161 (2017)
- [9] Graf, G.M.: Aspects of the Integer Quantum Hall Effect. In: *Spectral Theory and Mathematical Physics: A Festschrift in Honor of Barry Simon's 60th Birthday*, No. 76 in *Proceedings of Symposia in Pure Mathematics*, pages 429–442. American Mathematical Society, Providence (2007)
- [10] Haldane, F.D.M.: Model for a Quantum Hall Effect without Landau levels: condensed-matter realization of the “parity anomaly”. *Phys. Rev. Lett.* **61**, 2017 (1988)
- [11] Hang, F., Lin, F.H.: Topology of Sobolev mappings II. *Acta Math.* **191**, 55–107 (2003)
- [12] Hasan, M.Z., Kane, C.L.: Colloquium: Topological Insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010)
- [13] Hofstadter, D.R.: Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields. *Phys. Rev. B* **14**, 2239–2249 (1976)
- [14] Kuchment, P.: An overview of periodic elliptic operators. *Bull. AMS* **53**, 343–414 (2016)
- [15] Marcelli, G., Moscolari, M., Panati, G.: Localization implies Chern triviality. In preparation (2019)
- [16] Marzari, N., Mostofi, A., Yates, J., Souza, I., Vanderbilt, D.: Maximally localized Wannier functions: Theory and applications. *Rev. Mod. Phys.* **84**, 1419–1475 (2012)
- [17] Monaco, D., Panati, G.: Topological invariants of eigenvalue intersections and decrease of Wannier functions in graphene. *J. Stat. Phys.* **155**, 1027–1071 (2014)
- [18] Monaco, D., Panati, G.: Symmetry and localization in periodic crystals: triviality of Bloch bundles with a fermionic time-reversal symmetry. *Acta Appl. Math.* **137**, 185–203 (2015)
- [19] Monaco, D., Panati, G., Pisante, A., Teufel, S.: Optimal decay of Wannier functions in Chern and Quantum Hall insulators. *Comm. Math. Phys.* **359**, 61–100 (2018)
- [20] Nenciu, A., Nenciu, G.: Existence of exponentially localized Wannier functions for nonperiodic systems. *Commun. Math. Phys.* **190**, 541–548 (1998)
- [21] Panati, G.: Triviality of Bloch and Bloch-Dirac bundles. *Ann. Henri Poincaré* **8**, 995–1011 (2007)
- [22] Panati, G., Pisante, A.: Bloch bundles, Marzari-Vanderbilt functional and maximally localized Wannier functions. *Commun. Math. Phys.* **322**, 835–875 (2013)
- [23] Reed M., Simon, B.: *Methods of Modern Mathematical Physics. Volume IV: Analysis of Operators*. Academic Press, New York (1978)
- [24] Santoro, G.E.: Lectures notes on Non-equilibrium quantum systems. <http://indico.ictp.it/event/7644/material/2/2.pdf> (accessed December 10, 2018)
- [25] Thouless, D.J., Kohmoto, M., Nightingale, M.P., de Nijs, M.: Quantized Hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).
- [26] Zak, J.: Magnetic translation group. *Phys. Rev.* **134**, A1602 (1964)

Received: 23 November 2018.

Accepted: 18 December 2018.

Giovanna Marcelli

*Fachbereich Mathematik, Eberhard Karls Universität Tübingen. Auf der Morgenstelle 10, 72076 Tübingen (DE).*

giovanna.marcelli@uni-tuebingen.de

Domenico Monaco

*Dipartimento di Matematica e Fisica, Università degli Studi di Roma Tre. Largo San Leonardo Murialdo 1, 00146 Rome (IT).*

dmonaco@mat.uniroma3.it

Massimo Moscolari

*Dipartimento di Matematica, “La Sapienza” Università di Roma. Piazzale Aldo Moro 2, 00185 Rome (IT).*

moscolari@mat.uniroma1.it

Gianluca Panati

*Dipartimento di Matematica, “La Sapienza” Università di Roma. Piazzale Aldo Moro 2, 00185 Rome (IT).*

panati@mat.uniroma1.it

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.