

# Electronic properties of 1D and 2D materials

Yury Holubeu \*

June 19, 2024

This draft is not aimed for distribution.

Electronic properties of 1D and 2D materials and applications are discussed in details. Links below show the contents of solved [problems](#), description of [edge states](#), boundary [charge fluctuations](#) by Pletyukhov et al. [Rational boundary charge in 1D](#), [The kondo lattice and weak AFM](#) by Doniach (1977), [Tomonaga's Model and singularity of X-Ray Spectra of Metals](#) by Schotte, Schotte (1969), [excitons in metals](#) by Manan (1967) and [CDW in high magnetic field](#) by Grigoriev, Lyubshin.

## Contents

<b>1</b>	<b>Preface and main motivation</b>	<b>4</b>
<b>I</b>	<b>1D and 2D electronic properties in a Nutshell</b>	<b>5</b>
1.	Main Formulas and Ideas (?)	5
1.1	Final properties of 1D electron models	5
1.2	Dirac 1D model main formulas	5
1.3	Rationality of boundary charge in a nutshell	10
1.4	Main Green function methods	12
1.5	Tight-binding models in a nutshell	12
2.	Boundary Charge Fluctuations main ideas (???????)	13
2.1	Symmetries and nearsightedness principle main ideas	13
2.2	Rational boundary charge in one-dimensional systems with interaction and disorder by Pletyukhov et. al. main ideas	20
2.3	Application to single-channel and nearest-neighbor hopping models by Pletyukhov et. al.	20
<b>II</b>	<b>Theory</b>	<b>21</b>
<b>2</b>	<b>Rational boundary charge in 1D systems with interaction and disorder</b>	<b>21</b>
1.	Main Ideas (???????)	21
2.	Introduction	21
3.	General framework	26
3.1	Hamiltonian, boundary and interface charges	26
3.2	Rational quantization of boundary charge	28
4.	Application to single-channel and nearest-neighbor hopping models	34
4.1	Weyl semimetal physics at half-filling	34
4.2	Low-energy theory	38
4.3	Summary and outlook	45
5.	Appendixes	46
5.1	Local vs. non-local symmetries	46
5.2	Stability of NSP: DMRG analysis	48
5.3	Stability of NSP: 1-channel systems	49
5.4	Symmetries for single-channel and nearest-neighbor hopping models	53
5.5	Boundary charge for Dirac model (!!!!!!)	54
5.6	Boundary charge at zero gap	55
5.7	Interface charge for Dirac model	56
5.8	Finite smaller systems	58

---

\*yura.winter@gmail.com

<b>3</b>	<b>Universality of Boundary Charge Fluctuations</b>	<b>58</b>
0.1	Introduction . . . . .	59
0.2	Model and Boundary charge fluctuations . . . . .	60
0.3	Boundary charge fluctuations . . . . .	61
0.4	Single-channel case . . . . .	61
0.5	Two- and Three-Dimensional Systems . . . . .	64
0.6	Conclusion . . . . .	65
1.	Appendixes . . . . .	65
1.1	Numerical results of Fig.4b . . . . .	65
1.2	Boundary charge fluctuations . . . . .	65
1.3	Noninteracting and clean models . . . . .	70
1.4	SSH model with disorder . . . . .	75
1.5	Higher dimensions . . . . .	77
1.6	Special phase transitions . . . . .	81
1.7	$L_p$ dependence . . . . .	82
<b>4</b>	<b>Surface charge theorem and topological constraints for edge states</b>	<b>82</b>
1.	Introduction . . . . .	83
2.	Spectral properties . . . . .	89
2.1	The model . . . . .	89
2.2	Bloch eigenstates for the infinite system . . . . .	92
2.3	Energy dispersion . . . . .	97
2.4	Scattering states for half-infinite system . . . . .	100
2.5	Edge states for half-infinite system . . . . .	101
2.6	Topological constraints for edge states . . . . .	105
3.	Boundary charge and density . . . . .	110
3.1	Definition and splitting of boundary charge . . . . .	111
3.2	Particle-hole duality . . . . .	116
3.3	Density and localization of boundary charge . . . . .	116
4.	The surface charge theorem . . . . .	119
5.	Topological indices . . . . .	123
6.	Universal properties . . . . .	127
6.1	Physical picture . . . . .	127
6.2	Invariant and boundary charge for a single band . . . . .	133
6.3	Total invariant and boundary charge . . . . .	138
7.	Summary and outlook . . . . .	140
8.	Appendixes . . . . .	142
8.1	Complex hoppings . . . . .	142
8.2	Choice of phase-dependence . . . . .	143
8.3	Useful identities . . . . .	144
8.4	Expansion of $N_k$ around poles . . . . .	146
8.5	Topological constraints for edge states . . . . .	149
8.6	Diophantine equation . . . . .	150
8.7	Construction of all edge state configurations . . . . .	151
8.8	Density . . . . .	152
8.9	Asymptotic values . . . . .	153
8.10	Tuning of edge states via choice of the phase-dependence . . . . .	155
<b>III</b>	<b>Examples and Solved Problems</b>	<b>156</b>
9.	Typical examples . . . . .	156
<b>IV</b>	<b>Other topics</b>	<b>157</b>

<b>5</b>	<b>CDW state high magnetic field quasi-1D mean field by Grigoriev, Lyubshin</b>	<b>157</b>
0..1	Main idea of the article and what is important here? . . . . .	157
0..2	Introduction . . . . .	157
1.	The Model and the mean-field theory of CDW in magnetic field. . . . .	159
1..1	The model . . . . .	159
1..2	Mean field approach . . . . .	160
1..3	The Cosine Phase . . . . .	161
1..4	The Double Cosine Phase . . . . .	163
1..5	Free energy of cosine and double-cosine phases . . . . .	165
2.	The phase diagram . . . . .	166
2..1	Transition line $T_c(H)$ and the normal-CDW <sub>0</sub> -CDW <sub>x</sub> tricritical point . . . . .	167
2..2	$T_c(H)$ transition line at perfect nesting . . . . .	168
2..3	Metal-CDW transition lines and tricritical points at finite $t'_b$ . . . . .	169
2..4	Cosine and double-cosine phases at perfect nesting . . . . .	170
3.	Discussion . . . . .	173
<b>6</b>	<b>Some other articles</b>	<b>176</b>
1.	Excitons in Metals: Infinite Hole Mass by Manan (1967) . . . . .	176
1..1	I. INTRODUCTION . . . . .	177
1..2	II. THE INTERBAND OPTICAL CONDUCTIVITY . . . . .	178
1..3	III. HIGHER-ORDER INTERACTIONS . . . . .	180
1..4	A. Perturbation Expansion . . . . .	180
1..5	B. Solution to the Vertex Equation . . . . .	181
1..6	IV. DISCUSSION . . . . .	184
1..7	APPENDIX: THE IRREDUCIBLE INTERACTIONS . . . . .	185
2.	Tomonaga's Model and the Threshold Singularity of X-Ray Spectra of Metals by Schotte, Schotte (1969) . . . . .	186
2..1	1. INTRODUCTION . . . . .	186
2..2	2. FORMULATION OF THE PROBLEM IN TERMS OF TOMONAGA'S BOSONS . . . . .	187
2..3	3. TRANSITION PROBABILITY IN THE DENSITY WAVE MODEL . . . . .	190
3.	The kondo lattice and weak antiferromagnetism by Doniach (1977) . . . . .	192
3..1	1. The Kondo lattice . . . . .	192
3..2	2. A one-dimensional analog-the "Kondo necklace" . . . . .	193
3..3	3. A mean field solution for the Kondo necklace . . . . .	193
3..4	4. The excitation spectrum near the AF to Kondo transition . . . . .	194
3..5	5. Application to real systems . . . . .	194
3..6	Appendix . . . . .	195
<b>V</b>	<b>Adds</b>	<b>197</b>
1.	Дополнения . . . . .	197
1..1	Литература . . . . .	197
1..2	Refs rational . . . . .	197
1..3	Refs universality ... . . . .	202
1..4	Refs magn field . . . . .	203
1..5	References by S. Doniach (1977) . . . . .	206
1..6	References by Pletyukhov et. al. (?) . . . . .	206
1..7	References by Pletyukhov et. al. 2 (?) . . . . .	210
1..8	References from "Fluctuations of boundary" (?) . . . . .	211

# 1 Preface and main motivation

## Amazing facts

(I'll reveal it later)

## Puzzles for motivation

(I'll reveal it later)

## Part I

# 1D and 2D electronic properties in a Nutshell

## 1. Main Formulas and Ideas (?)

### 1.1 Final properties of 1D electron models

(here are plots and very imp formulas)

### 1.2 Dirac 1D model main formulas

#### Models, wave functions and energies of Dirac model

**Wave functions and energies in infinite modes (!!!!!)** Lest's consider 1D model of electrons in the potential

$$u(x) = u_0 \cos(2p_F x + \varphi)$$

(????? write here the initial hamiltonian!!!)

We decompose:

$$\psi(x) = \psi_1(x)e^{ip_F x} + \psi_2(x)e^{-ip_F x},$$

$$\mathcal{H}_0 = \int \psi^\dagger(x) \left( -\frac{\nabla^2}{2m} - \frac{p_F^2}{2m} \right) \psi(x) dx \approx -iv_F \int \left( \psi_1^\dagger(x) \frac{\partial \psi_1}{\partial x} - \psi_2^\dagger(x) \frac{\partial \psi_2}{\partial x} \right) dx$$

$$U(x) = g \partial u / \partial x \quad \hat{\mathcal{H}}_{\text{int}} = \int U(x) \psi^\dagger(x) \psi(x) dx \approx \Delta \int (\psi_1^\dagger(x) \psi_2(x) + \text{h.c.}) dx$$

$$\hat{H} = \begin{pmatrix} v_F \hat{k} & \Delta \\ \Delta^* & -v_F \hat{k} \end{pmatrix} \quad \hat{k} = -i\partial/\partial x; \quad \Delta = gu_0 p_F e^{i\varphi};$$

$$\varepsilon_\pm(k) = \pm \sqrt{v_F^2 k^2 + \Delta^2},$$

$$\delta E_{\text{el}} = E_{\text{el}}(\Delta) - E_{\text{el}}(\Delta = 0) = -2 \sum_k \left( \sqrt{v_F^2 k^2 + |\Delta|^2} - v_F |k| \right) \quad (\text{divergent})$$

$$\delta E_{\text{el}}|_{k < k_F} = -\frac{L|\Delta|^2}{\pi v_F} \left( \ln \frac{2\varepsilon_*}{|\Delta|} - \frac{1}{2} \right), \quad \varepsilon_* := v_F p_F$$

Also comparesment with energy of modulation of lattice  $E_{\text{latt}} = \frac{\rho c^2}{2} \int \left( \frac{\partial u}{\partial x} \right)^2 dx = \rho c^2 p_F^2 u_0^2 L = \frac{L|\Delta|^2}{g^2}$  can give us the maximum value of the gap:

$$\Delta_0 = \varepsilon_* \exp(-\pi v_F / g^2)$$

#### Wave functions and energies in two connected different 1D chains

In this Appendix we consider an interface between two Dirac models according to the Hamiltonian (2.106), where the phase  $\alpha(x)$  of the gap parameter depends on  $x$ . We will prove the Goldstone-Wilczek formula (2.107) for the interface charge for the particular choice  $\alpha(x) = \alpha_R \Theta(x) + \alpha_L \Theta(-x)$ . We define the parameter  $\delta\alpha = \alpha_R - \alpha_L$ . The eigenstates follow from the equation

$$[-iv_F \sigma_3 \partial_x + |\Delta|(\sigma_+ e^{i\alpha(x)} + \sigma_- e^{-i\alpha(x)})] \underline{\psi}(x) = \epsilon \underline{\psi}(x), \quad (1.1)$$

with  $\sigma_{\pm} = \frac{1}{2}(\sigma_x \pm i\sigma_y)$  and  $\underline{\psi}(x)^T = (R(x), L(x))^T$ . Like in the case of the semi-infinite Dirac model discussed in Appendix 5.5, we separate the spectrum of the Hamiltonian into two regions: I)  $|\epsilon| < |\Delta|$ , and II)  $|\epsilon| > |\Delta|$ .

The bound state solution appears for  $\sin(\delta\alpha/2) > 0$  with energy  $\epsilon = -|\Delta| \cos(\delta\alpha/2)$ , and is given by

$$\underline{\psi}^I(x) = \sqrt{\frac{\kappa}{2}} \begin{pmatrix} 1 \\ -e^{-i\frac{\alpha_R + \alpha_L}{2}} \end{pmatrix} e^{-\kappa|x|}, \quad (1.2)$$

with  $\kappa = \frac{|\Delta|}{v_F} \sin(\delta\alpha/2)$ . If it is occupied it gives an integer contribution to the interface charge.

For each energy  $|\epsilon_k| > |\Delta|$ , the extended eigenstates can be chosen as scattering states within two scattering channels. The first one (denoted by the index  $r$ )

$$\underline{\psi}_k^{(r)}(x) = \frac{\Theta(-x)}{\sqrt{2\pi}} [\underline{\chi}_{L,k} e^{ikx} + r_k \underline{\chi}_{L,-k} e^{-ikx}] + \frac{\Theta(x)}{\sqrt{2\pi}} t_k \underline{\chi}_{R,k} e^{ikx} \quad (1.3)$$

$$\underline{\psi}_k^{(l)}(x) = \frac{\Theta(x)}{\sqrt{2\pi}} [\underline{\chi}_{R,-k} e^{-ikx} + r'_k \underline{\chi}_{R,k} e^{ikx}] + \frac{\Theta(-x)}{\sqrt{2\pi}} t'_k \underline{\chi}_{L,-k} e^{-ikx} \quad (1.4)$$

$$\underline{\chi}_{R/L,k} := \frac{1}{\sqrt{2\epsilon(\epsilon - v_F k)}} \begin{pmatrix} -|\Delta| e^{i\alpha_{R/L}} \\ v_F k - \epsilon \end{pmatrix} \quad \epsilon = \pm\epsilon_k \quad (1.5)$$

$$\underline{\psi}_k^{(r/l)}(0^+) = \underline{\psi}_k^{(r/l)}(0^-). \quad (1.6)$$

$$r_k = r'_k = \frac{|\Delta|(e^{i\alpha_R} - e^{i\alpha_L})}{(\epsilon - v_F k)e^{i\alpha_L} - (\epsilon + v_F k)e^{i\alpha_R}}, \quad t_k = \frac{2v_F k e^{i\alpha_L}}{(\epsilon + v_F k)e^{i\alpha_R} - (\epsilon - v_F k)e^{i\alpha_L}}, \quad t'_k = \frac{2v_F k e^{i\alpha_R}}{(\epsilon + v_F k)e^{i\alpha_R} - (\epsilon - v_F k)e^{i\alpha_L}}. \quad (1.7)$$

By an explicit calculation one can readily verify the fulfillment of the unitarity conditions

$$|t_k|^2 + |r_k|^2 = |t'_k|^2 + |r'_k|^2 = 1, \quad (1.8)$$

$$t^* r'_k + r_k^* t'_k = 0. \quad (1.9)$$

For a filled valence band we choose  $\epsilon = -\epsilon_k$  and identify the extended states' contribution to the interface charge

$$Q_I^{\text{II}} = \int_{-\infty}^{\infty} dx f(x) \int_0^{\infty} dk \left( |\psi_k^{(r)}(x)|^2 + |\psi_k^{(l)}(x)|^2 - \frac{2}{\pi} \right) = \frac{|\Delta|}{\pi} \int_0^{\infty} dx f(x) \int_{-\infty}^{\infty} dk \frac{e^{2ikx}}{\epsilon_k} r'_k. \quad (1.10)$$

As one can conclude from (2.189), this quantity periodically depends on  $\delta\alpha = \alpha_R - \alpha_L$ . Evaluating (2.194) for  $\delta\alpha \in (0, 2\pi)$  with  $f(x) = e^{-\eta|x|}$ ,  $\eta \rightarrow 0^+$ , we obtain

$$Q_I^{\text{II}} = \frac{\delta\alpha}{2\pi} - 1. \quad (1.11)$$

Putting the chemical potential at the bottom of the conduction band, we receive an additional contribution  $Q_I^{\text{I}} = 1$  from the edge state (2.184), which is present for every value of  $\delta\alpha$ , and obtain the resulting expression (2.107) for the total interface charge.

### Wave functions and energies in isolated impurity (??????? write hamiltonian!!!)

$$V_{\text{imp}} = V_0 c^\dagger(x_{\text{imp}}) c(x_{\text{imp}}) \simeq V_0 \begin{pmatrix} c_+^\dagger(x_{\text{imp}}) & c_-^\dagger(x_{\text{imp}}) \end{pmatrix} \begin{pmatrix} 1 & e^{-2ik_F^{(\nu)} x_{\text{imp}}} \\ e^{2ik_F^{(\nu)} x_{\text{imp}}} & 1 \end{pmatrix} \begin{pmatrix} c_+(x_{\text{imp}}) \\ c_-(x_{\text{imp}}) \end{pmatrix} =$$

$$= \delta(x - x_{\text{imp}}) V_0 (\sigma_0 + \cos(\vartheta) \sigma_1 + \sin(\vartheta) \sigma_2), \quad \vartheta := 2k_F^{(\nu)} x_{\text{imp}}$$

$$\partial_x \psi(x) x_{\text{imp}} \approx -ig \delta(x - x_{\text{imp}}) \sigma_3 (\sigma_0 + \cos(\vartheta) \sigma_1 + \sin(\vartheta) \sigma_2) \psi(x), \quad g := \frac{V_0}{v_F^{(\nu)}}.$$

Scattering states :

$$\psi_{\bar{k}\tau}^{(\nu,l)}(x) = \frac{\Theta(-\delta x)}{\sqrt{2\pi}} \left[ \chi_{\bar{k}\tau}^{(\nu)} e^{i\bar{k}\delta x} + r_{\bar{k}\tau}^{(\nu,l)} \chi_{-\bar{k}\tau}^{(\nu)} e^{-i\bar{k}\delta x} \right] + \frac{\Theta(\delta x)}{\sqrt{2\pi}} t_{\bar{k}\tau}^{(\nu,l)} \chi_{\bar{k}\tau}^{(\nu)} e^{i\bar{k}\delta x}, \quad \bar{k} \in [0, \infty)$$

$$\psi_{\bar{k}\tau}^{(\nu,r)}(x) = \frac{\Theta(-\delta x)}{\sqrt{2\pi}} t_{\bar{k}\tau}^{(\nu,r)} \chi_{-\bar{k}\tau}^{(\nu)} e^{-i\bar{k}\delta x} + \frac{\Theta(\delta x)}{\sqrt{2\pi}} \left[ \chi_{-\bar{k}\tau}^{(\nu)} e^{-i\bar{k}\delta x} + r_{\bar{k}\tau}^{(\nu,r)} \chi_{\bar{k}\tau}^{(\nu)} e^{i\bar{k}\delta x} \right], \quad \delta x = x - x_{\text{imp}}$$

$$t_{\bar{k}\tau}^{(\nu,r)} = t_{\bar{k}\tau}^{(\nu,l)} = \frac{v_F^{(\nu)} \bar{k}}{v_F^{(\nu)} \bar{k} + ig \left( \epsilon_{\bar{k}\tau}^{(\nu)} + \Delta^{(\nu)} \cos(\bar{\gamma}^{(\nu)}) \right)},$$

$$r_{\bar{k}\tau}^{(\nu,r/l)} = -\frac{ig\tau \left( \Delta^{(\nu)} + \epsilon_{\bar{k}\tau}^{(\nu)} \cos(\bar{\gamma}^{(\nu)}) \mp iv_F^{(\nu)} \bar{k} \sin(\bar{\gamma}^{(\nu)}) \right)}{v_F^{(\nu)} \bar{k} + ig \left( \epsilon_{\bar{k}\tau}^{(\nu)} + \Delta^{(\nu)} \cos(\bar{\gamma}^{(\nu)}) \right)}, \quad \bar{\gamma}^{(\nu)} = \gamma^{(\nu)} + \vartheta.$$

$$\left| t_{\bar{k}\tau}^{(\nu,l)} \right|^2 + \left| r_{\bar{k}\tau}^{(\nu,l)} \right|^2 = \left| t_{\bar{k}\tau}^{(\nu,r)} \right|^2 + \left| r_{\bar{k}\tau}^{(\nu,r)} \right|^2 = 1, \quad \left( t_{\bar{k}\tau}^{(\nu,l)} \right)^* r_{\bar{k}\tau}^{(\nu,r)} + \left( r_{\bar{k}\tau}^{(\nu,l)} \right)^* t_{\bar{k}\tau}^{(\nu,r)} = 0.$$

(???????????????? write energy of the system!!!!)

Impurity-localized bound state :

$$\psi_{\text{imp}}^{(\nu)}(x) = \left[ \Theta(\delta x) W^{(\nu,r)} \begin{pmatrix} e^{i\gamma^{(\nu)}} \\ -\text{sgn}(g) e^{-i\varphi_{\text{imp}}^{(\nu)}} \end{pmatrix} + \Theta(-\delta x) W^{(\nu,l)} \begin{pmatrix} e^{i\gamma^{(\nu)}} e^{-i\varphi_{\text{imp}}^{(\nu)}} \\ -\text{sgn}(g) \end{pmatrix} \right] e^{-\kappa_{\text{imp}}^{(\nu)} |\delta x|}.$$

$$\kappa_{\text{imp}}^{(\nu)} := \frac{|g| \Delta^{(\nu)}}{(1+g^2) v_F^{(\nu)}} \left( \sqrt{1+g^2 \sin^2(\bar{\gamma}^{(\nu)})} - \text{sgn}(g) \cos(\bar{\gamma}^{(\nu)}) \right),$$

$$e^{i\varphi_{\text{imp}}^{(\nu)}} := \frac{1}{1+ig} \left( \sqrt{1+g^2 \sin^2(\bar{\gamma}^{(\nu)})} + i|g| \cos(\bar{\gamma}^{(\nu)}) \right),$$

$$W^{(\nu,r/l)} := \sqrt{\frac{\kappa_{\text{imp}}^{(\nu)}}{2} \left( 1 \pm \frac{|g| \sin(\bar{\gamma}^{(\nu)})}{\sqrt{1+g^2 \sin^2(\bar{\gamma}^{(\nu)})}} \right)}.$$

$$\epsilon_{\text{imp}}^{(\nu)} = -\frac{\Delta^{(\nu)}}{1+g^2} \left( g^2 \cos(\bar{\gamma}^{(\nu)}) + \text{sgn}(g) \sqrt{1+g^2 \sin^2(\bar{\gamma}^{(\nu)})} \right).$$

### Wave functions and energies in the presence of an edge

$$\left[ -iv_F \sigma_3 \partial_x + |\Delta| (\sigma_+ e^{i\alpha} + \sigma_- e^{-i\alpha}) \right] \underline{\psi}(x) = \epsilon \underline{\psi}(x), \quad (1.12)$$

with  $\sigma_\pm = \frac{1}{2}(\sigma_x \pm i\sigma_y)$ ,  $\underline{\psi}(x) = (R(x), L(x))^T$ , and the boundary condition  $R(0) + L(0) = 0$ .

If  $|\epsilon| < |\Delta|$ , there is a single bound state if  $\sin \alpha > 0$ :

$$\epsilon = -|\Delta| \cos \alpha$$

$$\underline{\psi}^{\text{I}}(x) = \sqrt{\kappa} \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-\kappa x}, \quad \kappa := \frac{|\Delta| |\sin \alpha|}{v_F} \quad (1.13)$$

If  $|\epsilon| > |\Delta|$  there is a continuum of scattering states with each  $k \in [0, \infty)$  with to the two bands:

$$\epsilon_{k,\pm} = \pm \sqrt{v_F^2 k^2 + |\Delta|^2} \equiv \pm \epsilon_k$$

For lower (valence) band:

$$\underline{\psi}_k(x) = \frac{1}{\sqrt{2\pi N_k}} \left[ \begin{pmatrix} -|\Delta| e^{i\alpha} \\ v_F k + \epsilon_k \end{pmatrix} e^{ikx} - s_k \begin{pmatrix} -|\Delta| e^{i\alpha} \\ -v_F k + \epsilon_k \end{pmatrix} e^{-ikx} \right], \quad (1.14)$$

$$N_k := |\Delta|^2 + (v_F k + \epsilon_k)^2 = 2\epsilon_k(\epsilon_k + v_F k), \quad s_k := \frac{|\Delta| e^{i\alpha} - v_F k - \epsilon_k}{|\Delta| e^{i\alpha} + v_F k - \epsilon_k}, \quad |s_k|^2 = \frac{\epsilon_k + v_F k}{\epsilon_k - v_F k}, \quad s_k s_{-k} = 1. \quad (1.15)$$

If there is an infinite-strength impurity  $g \rightarrow \infty$  we obtain

$$\begin{aligned} t_{\bar{k}\tau}^{(\nu,r)} &= t_{\bar{k}\tau}^{(\nu,l)} \rightarrow 0, \\ r_{\bar{k}\tau}^{(\nu,r/l)} &\rightarrow -\frac{\tau \left( \Delta^{(\nu)} + \epsilon_{\bar{k}\tau}^{(\nu)} \cos(\bar{\gamma}^{(\nu)}) \mp i v_F^{(\nu)} \bar{k} \sin(\bar{\gamma}^{(\nu)}) \right)}{\left( \epsilon_{\bar{k}\tau}^{(\nu)} + \Delta^{(\nu)} \cos(\bar{\gamma}^{(\nu)}) \right)}, \\ s_{\bar{k}\tau}^{(\nu)} &= \frac{\Delta^{(\nu)} e^{i\gamma^{(\nu)}} - v_F^{(\nu)} \bar{k} + \epsilon_{\bar{k}\tau}^{(\nu)}}{\Delta^{(\nu)} e^{i\gamma^{(\nu)}} + v_F^{(\nu)} \bar{k} + \epsilon_{\bar{k}\tau}^{(\nu)}} \end{aligned}$$

For impurity after  $g \rightarrow \infty$ , the above expressions boil down to

$$\begin{aligned} \kappa_{\text{imp}} &= \frac{\Delta^{(\nu)}}{v_F^{(\nu)}} |\sin(\bar{\gamma}^{(\nu)})|, \quad e^{i\varphi_{\text{imp}}} = \cos(\bar{\gamma}^{(\nu)}) - i |\sin(\bar{\gamma}^{(\nu)})| \\ W^{(\nu,r/l)} &= \sqrt{\frac{\kappa_{\text{imp}}}{2} \left( 1 \pm \frac{\sin(\bar{\gamma}^{(\nu)})}{|\sin(\bar{\gamma}^{(\nu)})|} \right)} \\ |s_{\bar{k}\tau}^{(\nu)}|^2 &= \frac{\epsilon_{\bar{k}\tau}^{(\nu)} - v_F^{(\nu)} \bar{k}}{\epsilon_{\bar{k}\tau}^{(\nu)} + v_F^{(\nu)} \bar{k}}, \quad s_{\bar{k}\tau}^{(\nu)} s_{-\bar{k}\tau}^{(\nu)} = 1 \\ \psi_e^{(\nu)}(x) &= \sqrt{\kappa_e^{(\nu)}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-\kappa_e^{(\nu)} x}, \quad \kappa_e^{(\nu)} = \frac{\Delta^{(\nu)}}{v_F^{(\nu)}} \sin(\gamma^{(\nu)}). \end{aligned}$$

Assuming that the chemical potential is located at the bottom of the conduction band, the bound state is occupied for  $0 < \alpha < \pi$ , and all valence band states  $\underline{\psi}_k$  are filled. Neglecting the strongly oscillating parts (providing unimportant corrections of  $O(\frac{\Delta}{v_F k_F}) \ll 1$ ), the contribution of each eigenstate to the density is given by

$$\rho_\psi(x) = \underline{\psi}^\dagger(x) \underline{\psi}(x) = |R(x)|^2 + |L(x)|^2. \quad (1.16)$$

We denote the contributions of the eigenstates  $\underline{\psi}^{\text{I}}$  and  $\underline{\psi}_k$  to the physical density by  $\rho_{\text{I}}(x)$  and  $\rho_k(x)$ , respectively. This gives for the total density relative to the average bulk density  $\bar{\rho}$

$$\rho(x) - \bar{\rho} = \rho_{\text{I}}(x) + \delta\rho_{\text{II}}(x), \quad (1.17)$$

$$\delta\rho_{\text{II}}(x) = \int_0^\infty dk \left[ \rho_k(x) - \frac{1}{\pi} \right], \quad (1.18)$$



$$Q_B = \int_0^\infty dx [\rho(x) - \bar{\rho}] f(x) = Q_B^I + Q_B^{II}; \quad Q_B^I = \int_0^\infty dx \rho_I(x) f(x), \quad Q_B^{II} = \int_0^\infty dx \delta \rho_{II}(x) f(x). \quad (1.19)$$

For the envelope function  $f(x)$  we choose the form  $f(x) = e^{-\eta x}$  with infinitesimally small  $\eta \rightarrow 0^+$ .

The bound state is occupied for  $0 < \alpha < \pi$  and gives an integer contribution to the boundary charge

$$Q_B^I = \int_0^\infty dx |\underline{\psi}^I(x)|^2 = \Theta_{0 < \alpha < \pi}. \quad (1.20)$$

This proves Eq. (2.100).

To calculate the scattering part  $Q_B^{II}$  to the boundary charge we use (2.165), (4.41), (2.167), and (2.168) and find after a straightforward calculation

$$\delta \rho_{II}(x) = -\frac{|\Delta|}{2\pi} \int_{-\infty}^\infty dk \frac{e^{2ikx}}{\epsilon_k} \frac{|\Delta| - \epsilon_k \cos \alpha - i v_F k \sin \alpha}{\epsilon_k - |\Delta| \cos \alpha}.$$

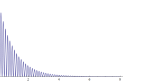
Inserting this result in (2.174) and performing the integration over  $x$  we obtain

$$Q_B^{II} = -\frac{1}{4} - \frac{|\Delta|}{4\pi} \int_{-\infty}^\infty dk \frac{v_F \sin \alpha}{\epsilon_k (\epsilon_k - |\Delta| \cos \alpha)} = -\frac{1}{4} + \frac{\ln(-e^{i\alpha})}{2\pi i}. \quad (1.21)$$

This proves Eq. (2.101).

## Densities and Charge of Dirac models

(!!! это очень скоро буду выводить!!!)  
Density of boundary electrons is:

$$\rho(x)_{\sin \gamma > 0} = \kappa e^{-2\kappa x} \langle (1 \quad -1) [\sigma_0 + e^{2ik_F^{(\nu)} x} \sigma_- + e^{-2ik_F^{(\nu)} x} \sigma_+] \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rangle = 2\kappa e^{-2\kappa x} (1 - \cos(k_F x)) \approx$$


$$\begin{aligned} \rho_F^{(s)}(x) &= -\frac{\operatorname{arcsinh}(2\xi^{(\nu)} \Lambda_\nu)}{2\pi \xi^{(\nu)}} \cos(2k_F^{(\nu)} x + \gamma^{(\nu)}) + \operatorname{Re} \left\{ \frac{e^{-i(2k_F^{(\nu)} x + \gamma^{(\nu)})}}{\xi^{(\nu)}} \int_{-\infty}^\infty dq \frac{e^{iqx/\xi^{(\nu)}}}{4\pi \bar{\epsilon}_q} \frac{e^{-i\gamma^{(\nu)}} - q - \bar{\epsilon}_q}{e^{-i\gamma^{(\nu)}} + q - \bar{\epsilon}_q} \right\} \\ &\simeq -\frac{\operatorname{arcsinh}(2\xi^{(\nu)} \Lambda_\nu)}{2\pi \xi^{(\nu)}} \cos(2k_F^{(\nu)} x + \gamma^{(\nu)}) - \frac{1}{2\xi^{(\nu)}} \frac{\cos(\gamma^{(\nu)})}{1 - \sin(\gamma^{(\nu)})} \sqrt{\frac{\xi^{(\nu)}}{2\pi x}} \sin(2k_F^{(\nu)} x + \gamma^{(\nu)}) e^{-x/\xi^{(\nu)}}, \end{aligned}$$

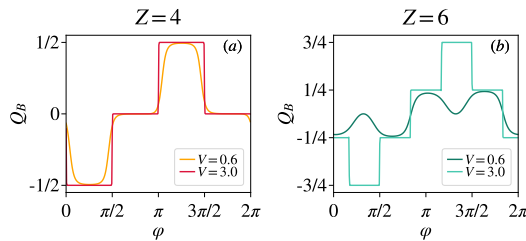


Figure 1: The same as the center row of Fig. 10, but for small  $N = 24$  and two values of the  $V$ .

### 1.3 Rationality of boundary charge in a nutshell

#### Quantisation and topology of a charge

	$\bar{Q}_B \bmod(e)$	Transformation
$T_n$	$Q_B + n\bar{\rho}$	translation $ m\rangle \rightarrow  m+n\rangle$
$\Pi$	$-Q_B$	unit-cell-local inversion
$U$	$Q_B$	site-local (anti-)unitary
$\Pi_n = T_n U \Pi$	$-Q_B + n\bar{\rho}$	unitary/time-reversal
$S_n = T_{-n} U$	$Q_B - n\bar{\rho}$	chiral/particle-hole

Table 1: Transformations  $Q_B \rightarrow \bar{Q}_B$  of the boundary charge under the elementary transformations  $T_n$  (translation by  $n$  lattice sites towards the boundary),  $\Pi$  (local inversion within each unit cell, where the unit cell is defined as the one starting at the boundary of a semi-infinite system),  $U$  (unitary or anti-unitary operations within the channel space of a single site), and combinations of these transformations defining the operations  $\Pi_n = T_n U \Pi$  and  $S_n = T_{-n} U$ . Except  $U$  all transformation rules are  $\bmod(e)$  due to the possible occurrence of edge states.  $\bar{\rho} = e\frac{\nu}{Z}$  is the average charge per site which is a rational number in the insulating regime.  $\nu$  denotes the number of filled bands and  $Z$  is the number of lattice sites of a unit cell (with  $N_c$  channels per site). If  $U$  is unitary (anti-unitary),  $\Pi_n$  and  $S_n$  are unitary (anti-unitary) operations. Highlighted in color are the transformation rules that need to be compared between Tables 2 and 3 to obtain the rational boundary charge.

	Symmetry	$\bar{Q}_B$	Quantization
$\Pi_n$	$\Pi_n H \Pi_n^\dagger = H$	$Q_B$	$\left. \begin{array}{l} \left. \begin{array}{l} \Pi_n H \Pi_n^\dagger = H \\ S_n H S_n^\dagger = -H \\ \& \frac{1}{2}\text{-filling} \end{array} \right\} \right\} Q_B = \frac{1}{2}n\bar{\rho} \bmod\left(\frac{e}{2}\right)$
$S_n$	$S_n H S_n^\dagger = -H$ & $\frac{1}{2}$ -filling	$-Q_B \bmod(e)$	

$$Q_B = \frac{1}{2}n\bar{\rho} \bmod\left(\frac{e}{2}\right). \quad (1.22)$$

$$Q_B \xrightarrow{T_n} \bar{Q}_B = Q_B + n\bar{\rho} \bmod(e), \quad (1.23)$$

$$Q_B \xrightarrow{\Pi} \bar{Q}_B = -Q_B \bmod(e). \quad (1.24)$$

$$Q_I = Q_B^L + Q_B^R \bmod(e) \quad (1.25)$$

$$= \delta n\bar{\rho} \bmod(e). \quad (1.26)$$

$$Q_B = \frac{\alpha}{2\pi}e + \frac{e}{4} \bmod(e). \quad (1.27)$$

### Boundary charge at zero gap

For the tight-binding model  $H_0$ , given by (2.64), restricted to the semi-infinite system  $m > 0$ , the eigenfunctions are given by (we set  $a = 1$ )

$$\psi_k(m) = \frac{1}{\sqrt{2\pi}}(e^{ikm} - e^{-ikm}), \quad (1.28)$$

with  $0 < k < \pi$ . For filling  $\bar{\rho} = k_F/\pi$ , this leads to the following charge  $\rho(m)$  at site  $m$

$$\begin{aligned} \rho(m) &= \int_0^{k_F} dk |\psi_k(m)|^2 \\ &= -\frac{1}{2\pi} \int_{-k_F}^{k_F} dk e^{2ikm} + \bar{\rho}. \end{aligned} \quad (1.29)$$

Inserting this result in the formula (2.16) for the boundary charge  $Q_B \equiv Q_B^R$ , we get

$$Q_B = -\frac{1}{4\pi} \int_{-k_F}^{k_F} dk \sum_{m=-\infty}^{\infty} e^{2ikm} f(m) + \frac{k_F}{2\pi}. \quad (1.30)$$

Choosing  $f(m) = e^{-\eta|m|}$ , we find  $\sum_{m=-\infty}^{\infty} e^{2ikm} f(m) = \pi\delta(k)$  and obtain for the boundary charge of  $H_0$  at zero gap

$$Q_B = -\frac{1}{4} + \frac{1}{2}\bar{\rho}. \quad (1.31)$$

This single-band model can be differently represented in terms of uniform unit cells with  $Z$  sites. This is especially useful, if we have in mind to add a  $Z$ -periodic perturbation on top of  $H_0$  (2.64). In the new representation, the single cosine band folds into  $Z$  bands with the reduced Brillouin zone (RBZ)  $[-\frac{\pi}{Z}, \frac{\pi}{Z})$ , the adjacent bands touching each other either in the center or at the edges of the RBZ. Choosing  $k_F/\pi$  of the original model to be rational,  $\frac{k_F}{\pi} = \frac{\nu}{Z}$ , we occupy  $\nu$  bands in the folded representation, and (2.181) then reads

$$Q_B = -\frac{1}{4} + \frac{\nu}{2Z}. \quad (1.32)$$

Adding a  $Z$ -periodic perturbation generically opens  $Z - 1$  gaps between all  $Z$  bands. Having the chemical potential in the  $\nu$ th gap, we can evaluate the correction to (2.182) due to the perturbation by means of the low-energy theory developed in Appendix 5.5.

This consideration clarifies the physical meaning of Eq. (2.105).

### Finite smaller systems

In this Appendix we show that the quantization of the boundary charge according to Fig. 10 is already visible for a tight-binding chain of  $\sim 20$  sites. As demonstrated in Fig. 16 for  $N = 24$  lattice sites the quantization can be demonstrated robustly as long as larger  $V$  can be accessed such that the localization length becomes small compared to the lattice size.

The results shown in Figs. 10 and 16 can be easily understood in the atomic limit  $V \gg t$ : The dominant contribution to  $Q_B$  comes from the polarization charge  $Q_P$  (2.133), while an eventual integer-valued Friedel charge contribution (2.132) is exactly cancelled by edge state contributions. To compute  $Q_P$ , we use the elaborated expression (see Ref. [87] for details)

$$Q_P = -\frac{1}{Z} \sum_{\alpha=1}^{Z/2} \sum_{j=1}^Z j \left( |\chi^{(\alpha)}(j)|^2 - \frac{1}{Z} \right), \quad (1.33)$$

where the occupied bands  $\epsilon^{(\alpha)}$  are approximately given by the potential components  $v_{\tilde{j}} < 0$  (one can even associate  $\tilde{j}$  with the band index  $\alpha$  sorting  $v_{\tilde{j}}$ 's in the ascending order for each value

of  $\varphi$ ). The corresponding eigenstate  $\chi^{(\alpha)}(j)$  possesses the only unity component  $\chi^{(\alpha)}(\tilde{j}) = 1$ , while  $\chi^{(\alpha)}(j \neq \tilde{j}) = 0$ . The plateau values in the discussed figures then immediately follow from (2.196). [It can so happen that two eigenstates  $v_{\tilde{j}_1}(\varphi)$  and  $v_{\tilde{j}_2}(\varphi)$  become degenerate at some value of  $\varphi$ , and then it is necessary to consider  $\frac{1}{\sqrt{2}}\{\chi^{(\alpha_1)}(j) \pm \chi^{(\alpha_2)}(j)\}$  for the eigenstates. This, however, does not alter the plateau value of  $Q_B$ .]

### 1.4 Main Green function methods

#### Main relations of Green function

$$G_{\lambda,\lambda'}^{(0)}(x, x' | t) = -i\Theta(t) \left\langle \left\{ c_{\lambda}(x | t), c_{\lambda'}^{\dagger}(x') \right\} \right\rangle_{\rho_0},$$

$$\rho(x) = -\frac{1}{\pi} \text{Im} \int d\omega \Theta(\mu - \omega) \text{tr}\{G(x, x)\},$$

$$\rho^{(0)}(x) = -\frac{1}{\pi} \text{Im} \int d\omega \Theta(\mu - \omega) \text{tr}\{G^{(0)}(x, x)\},$$

### 1.5 Tight-binding models in a nutshell

#### Main formulas of TB

$$\mathcal{H} = - \sum_n \sum_{j=1}^{Z-1} \sum_{\lambda,\lambda'=1}^{N_c} \left( (t_j)_{\lambda,\lambda'}^{\dagger} b_{n,j,\lambda}^{\dagger} b_{n,j+1,\lambda'} + \text{h.c.} \right) - \sum_n \sum_{\lambda,\lambda'=1}^{N_c} \left( (t_Z)_{\lambda,\lambda'}^{\dagger} b_{n,Z,\lambda}^{\dagger} b_{n+1,1,\lambda'} + \text{h.c.} \right) +$$

$$+ \sum_n \sum_{j=1}^Z \sum_{\lambda,\lambda'=1}^{N_c} (v_j)_{\lambda,\lambda'} b_{n,j,\lambda}^{\dagger} b_{n,j,\lambda'}.$$

$$H = - \sum_n \sum_{j=1}^{Z-1} \left( |n, j\rangle t_j^{\dagger} \langle n, j+1| + \text{h.c.} \right) - \sum_n \left( |n, Z\rangle t_Z^{\dagger} \langle n+1, 1| + \text{h.c.} \right) + \sum_n \sum_{j=1}^Z |n, j\rangle v_j \langle n, j|,$$

$$H = - \sum_m \left( |m\rangle t_m^{\dagger} \langle m| + \text{h.c.} \right) + \sum_m |m\rangle v_m \langle m|.$$

$$H_0 = -t \sum_m (|m+1\rangle \langle m| + \text{h.c.})$$

$$H' = \sum_m |m\rangle \delta v_m \langle m| - \sum_m (|m+1\rangle \delta t_m \langle m| + \text{h.c.}) \quad a \equiv 1$$

#### Tight-binding model as a discrete limit of continuum theory

$$\int dx \rightarrow a \sum_m.$$

$$b(ma) = \sqrt{a} c(ma).$$

$$\int dx c^{\dagger}(x) \frac{(p + A(x))^2}{2m^*} c(x) \simeq -t \sum_m (b^{\dagger}(ma) U_{ma}^{\dagger} b(ma+a) + b^{\dagger}(ma+a) U_{ma} b(ma)) + 2t \sum_m b^{\dagger}(ma) b(ma),$$

$$\int dx c^{\dagger}(x) \tilde{V}(x) c(x) \simeq \sum_m b^{\dagger}(ma) \tilde{V}(ma) b(ma).$$

$$\mathcal{H} \simeq - \sum_m \sum_{\lambda, \lambda'=1}^{N_c} \left( (t_m^\dagger)_{\lambda, \lambda'} b_\lambda^\dagger(ma) b_{\lambda'}(ma+a) + \text{h.c.} \right) + \sum_m \sum_{\lambda, \lambda'=1}^{N_c} (v_m)_{\lambda, \lambda'} b_\lambda^\dagger(ma) b_{\lambda'}(ma),$$

here  $t_m = tU_{ma} = t_{m+Z}$ ,  $v_m = \tilde{V}(ma) + 2t = v_{m+Z} = v_m^\dagger$ .

### Main formulas of SSH model

(это же в конд средах будет написано, скоро напишу, просто не так это сейчас важно)

## 2. Boundary Charge Fluctuations main ideas (???????)

(суть статьи Кирилла)

### 2.1 Symmetries and nearsightedness principle main ideas

#### Local vs. non-local symmetries

In this Appendix we provide a summary of our conventions to distinguish between local and non-local symmetries. Although this being standard (see, e.g., Ref. [23]), conventions sometimes differ in the literature and the material might be helpful for readers not so familiar with the precise definitions of the various symmetries.

For a given Hamiltonian  $H$ , there are four kinds of symmetries, depending on whether the symmetry operation commutes/anticommutes with  $H$  and whether it is unitary or anti-unitary

$$UHU^\dagger = H \quad , \quad SHS^\dagger = -H \quad , \quad (1.34)$$

$$THT^\dagger = H \quad , \quad CHC^\dagger = -H \quad . \quad (1.35)$$

Here,  $U$  and  $S$  are unitary operators, whereas  $T$  and  $C$  are anti-unitary operators.  $S$  is called a chiral symmetry,  $T$  a time-reversal symmetry, and  $C$  a particle-hole (or charge conjugation) symmetry. The anti-unitary symmetries  $T$  and  $C$  consist of a combination of unitary operations  $U_T$  and  $U_C$  with complex conjugation  $K$ :  $T = U_T K$  and  $C = U_C K$ . The operation  $K$  of complex conjugation requires a basis in which it is defined. Here, we take always the real-space representation in terms of  $|m\sigma\rangle$ , where  $m$  is the lattice site index and  $\sigma$  the channel index.

To distinguish local from non-local symmetries one needs to specify the unit cell and write the total Hilbert space as a direct product of the space within the unit cell (labeled by the site index  $j = 1, \dots, Z$  and the channel index  $\sigma = 1, \dots, N_c$  for each site) and the space of all unit cells labeled by the integer  $n$ . In the 1-particle subspace, the tight-binding model (2.7) can then be alternatively written as

$$H = \sum_{n, \tau} \underline{h}(\tau) \otimes |n + \tau\rangle \langle n| \quad , \quad (1.36)$$

where  $\underline{h}(\tau)$  are  $ZN_c \times ZN_c$ -matrices describing the coupling of unit cell  $n$  with unit cell  $n + \tau$  (the lattice site index  $m$  used in (2.7) is related to  $n$  and  $j$  by  $m = Z(n - 1) + j$ ; note that  $\tau$  has a different meaning compared to  $\delta$  used in (2.7), the same applies for the symbol  $\underline{h}$ ). A local symmetry is then defined by a symmetry with respect to the Hamiltonian  $\underline{h}(\tau)$  (i.e., it acts only within the space of a single unit cell) and, in addition, is independent of  $\tau$

$$U\underline{h}(\tau)U^\dagger = \underline{h}(\tau) \quad , \quad S\underline{h}(\tau)S^\dagger = -\underline{h}(\tau) \quad , \quad (1.37)$$

$$T\underline{h}(\tau)T^\dagger = \underline{h}(\tau) \quad , \quad C\underline{h}(\tau)C^\dagger = -\underline{h}(\tau) \quad . \quad (1.38)$$

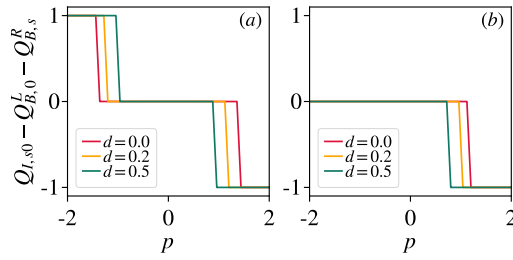


Figure 2: Stability of Eq. (2.22) with respect to random disorder. The figure shows  $Q_{I,s0} - Q_{B,0}^L - Q_{B,s}^R$  for the same parameters as used in the left column ( $Z = 4$ ) of Fig. 10 at  $\varphi = \pi/4$ , where the gap is maximal (but at finite size  $N = 1000$ ). The relative shift of the chains left and right to the interface are (a)  $s = 0$ , and (b)  $s = 1$ . We take half-filling instead of  $\mu = 0$  here. Random disorder drawn from a uniform distribution  $[-d/2, d/2)$  is added to the onsite potentials and  $p$  describes changes to the interface properties (see main text for details). As the properties of the interface are swept through  $Q_{I,s0} - Q_{B,0}^L - Q_{B,s}^R$  only changes mod(1).

Using the Fourier transform  $\tilde{h}(k) = \sum_{\tau} h(\tau) e^{-ik\tau}$ , with real quasimomentum  $-\pi < k < \pi$ , this can also be written as

$$U\tilde{h}(k)U^{\dagger} = \tilde{h}(k) \quad , \quad S\tilde{h}(k)S^{\dagger} = -\tilde{h}(k) \quad , \quad (1.39)$$

$$T\tilde{h}(-k)T^{\dagger} = \tilde{h}(k) \quad , \quad C\tilde{h}(-k)C^{\dagger} = -\tilde{h}(k) \quad . \quad (1.40)$$

Within our convention, a non-local symmetry can not be written in this form. There are three possibilities: (1) The non-local symmetry can be written as a local one by taking another choice of the unit cell. (2) The non-local symmetry acts within the space of a single unit cell but depends on  $\tau$  (or, equivalently, on the quasimomentum  $k$ ). (3) The non-local symmetry does not act within the space of a single unit cell whatever choice one takes for the unit cell, i.e., can only be written with respect to the total Hamiltonian  $H$ . For  $n \neq 0$ , the symmetries  $\Pi_n$  and  $S_n$  defined in Eq. (2.35) are non-local symmetries within our definition. Examples for cases (1) and (2) are discussed in the paragraph following Eq. (2.40) via special cases for the symmetry  $\Pi_n$ . The case (1) is discussed extensively at the end of Section 3.2 when the local symmetries  $\Pi_0$  or  $S_0$  are present but not with respect to the unit cell starting at the boundary of the semi-infinite system. For  $n \neq 0$ , the symmetry  $S_n$  is an example for case (3).

### Stability of NSP: DMRG analysis

In this Appendix we analyse the influence of static random disorder and short-ranged electron-electron interaction on the boundary and interface charge by using exact diagonalization and DMRG. For particular examples we demonstrate that the interface charge (2.22) is independent of the interface coupling  $V_I$  (up to an integer), and we show that Eqs. (2.26) and (4.154) for the boundary charge are generically valid.

We start with the interface charge and demonstrate in Figs. 13 and Figs. 14 that Eq. (2.22) holds even in the presence of random disorder as well as short-ranged electron-electron interaction, respectively. We consider an interface of the following form: Take initially two decoupled chains of the form as defined by Eq. (2.50). We want to include changes where the potential form of the right chain is shifted in the variable  $\varphi$  with respect to the left one by an integer multiple of  $\frac{2\pi}{Z}$ . This means  $\varphi \rightarrow \varphi + s\frac{2\pi}{Z}$ , which effectively shifts the right lattice by  $s$  sites compared to the left one. In Eq. (2.22) this means that  $n = s$  and  $n' = 0$ .

To define a single parameter  $p$  which changes the interface's properties continuously we consider the link between the rightmost site of the left lattice to the leftmost site of the right lattice to be  $t_{\text{link}} = |p|/2$  and add an onsite potential of size  $p$  to both of these sites. Therefore,  $p = 0$  is the decoupled case of two chains without an additional onsite potential at the edge and for negative  $p$  charges tend to get trapped at the interface, while for positive  $p$  they are pushed out. We add a quenched disorder following Eq. (2.15) for the results in Fig. 13 and a electron-electron interaction following Eq. (2.62) for the results in Fig. 14. Since we concentrate

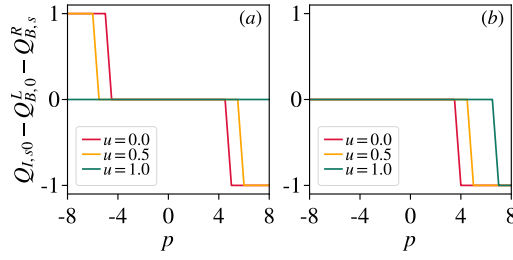


Figure 3: Stability of Eq. (2.22) with respect to interactions. The figure shows  $Q_{I,s0} - Q_{B,0}^L - Q_{B,s}^R$  for the same parameters as used in the left column ( $Z = 4$ ) of Fig. 10 at  $\varphi = \pi/4$ , where the gap is maximal (but at finite size  $N = 200$  and with larger  $V/t = 1.2$ ). The relative shift of the chains left and right to the interface are (a)  $s = 0$ , and (b)  $s = 1$ . We take half-filling instead of  $\mu = 0$  here.  $p$  describes changes to the interface properties (see main text for details). As the properties of the interface are swept through  $Q_{I,s0} - Q_{B,0}^L - Q_{B,s}^R$  only changes mod(1).

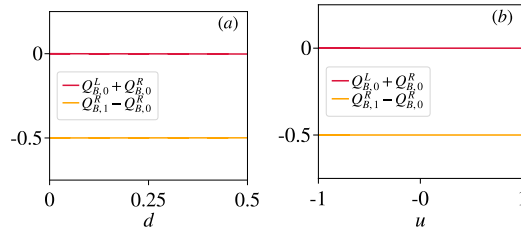


Figure 4: Stability of Eqs. (2.26) and (4.154) with respect to (a) random disorder and (b) interaction. We work at half filling such that  $\bar{\rho} = 1/2$ .  $Q_{B,1}^R - Q_{B,0}^R = \bar{\rho} \bmod(1)$  is shown to demonstrate Eq. (2.26), while  $Q_{B,0}^L + Q_{B,0}^R = 0 \bmod(1)$  illustrates Eq. (4.154). The parameters are the same as in Fig. 13.

on nearest-neighbor interaction we additionally scale the interaction over the interface bond by  $p$ , such that  $p = 0$  is the limit of two decoupled chains. Clearly Eq. (2.22) remains valid in both cases.

Next we study the influence of static random disorder and short-ranged electron-electron interaction on the transformation laws (2.26) and (4.154) of the boundary charge under translations and local inversion, see Figs. 15(a,b). Up to rather large values of the disorder and the electron-electron interaction both transformations laws remain perfectly valid, as expected from the NSP.

### Stability of NSP: 1-channel systems

In this Appendix we demonstrate the validity of Eq. (2.22) (with  $n = n' = 0$ ) for a model of two non-interacting single-channel nearest-neighbor chains coupled with each other via a tunable hopping amplitude. It is explicitly shown that (2.22) holds for any strength of the link.

Let us consider the Hamiltonian of the infinite chain  $H = H_R + H_L + V_I$  consisting of the three parts

$$H_R = \sum_{n=1}^{\infty} \left\{ |n\rangle\langle n| \otimes h(0) + |n+1\rangle\langle n| \otimes h(1) + |n\rangle\langle n+1| \otimes h(-1) \right\}, \quad (\text{right semi-inf. chain})$$

$$H_L = \sum_{n=-\infty}^0 \left\{ |n\rangle\langle n| \otimes h(0) + |n\rangle\langle n-1| \otimes h(1) + |n-1\rangle\langle n| \otimes h(-1) \right\}, \quad (\text{left semi-inf. chain})$$

$$V_I = \lambda \left[ |n=1\rangle\langle n=0| \otimes h(1) + |n=0\rangle\langle n=1| \otimes h(-1) \right], \quad (\text{tunneling})$$

Here, in contrast to the lattice site index  $m$ , the index  $n$  enumerates unit cells. Both  $H_R$

and  $H_L$  describe the lattices with the same structure of a unit cell, which is encoded in

$$h(0) = \sum_{j=1}^Z v_j |j\rangle\langle j| - \sum_{j=1}^{Z-1} t_j (|j\rangle\langle j+1| + |j+1\rangle\langle j|),$$

$h(1) = -t_Z |j=1\rangle\langle j=Z|$ , and  $h(-1) = h^\dagger(1)$ , i.e. characterized by  $Z$  sites  $j = 1, \dots, Z$  per unit cell, a single orbital (channel) per site, and by the same values for hoppings  $t_j$  and onsite potentials  $v_j$ . The tunneling amplitude  $\lambda t_Z$  between the two subsystems is quantified by the real-valued parameter  $\lambda \geq 0$ . Its special values  $\lambda = 0$  and  $\lambda = 1$  correspond to the cases of the two decoupled semi-infinite chains and the translationally invariant infinite chain, respectively. A restoration of the translational symmetry in the latter case is guaranteed by the perfect matching of the unit cells touching each other at the interface.

Due to the same structure, the Hamiltonians  $H_R$  and  $H_L$  appear to be isospectral, and their extended eigenstates can be therefore labeled by the same band index  $\alpha$  and quasimomentum  $k$  on the both sides from the interface. Moreover, these quantum numbers can be also used for a construction of scattering eigenstates of the coupled system, since eigenenergies  $\epsilon_k^{(\alpha)}$  of the extended states remain independent of  $\lambda$ , and they can be ultimately evaluated from the bulk Hamiltonian, i.e. at  $\lambda = 1$ . On the basis of this observation, we make the following ansatz for the two distinct scattering eigenstates additionally labeled by either  $r$  or  $l$ :

$$\psi_k^{(\alpha,r)}(n, j) = \frac{\Theta_{n \leq 0}}{\sqrt{2\pi}} [\chi_k^{(\alpha)}(j) e^{ikn} + r_k^{(\alpha)} \chi_{-k}^{(\alpha)}(j) e^{-ikn}] + \frac{\Theta_{n \geq 1}}{\sqrt{2\pi}} t_k^{(\alpha)} \chi_k^{(\alpha)}(j) e^{ikn}, \quad (1.41)$$

$$\psi_k^{(\alpha,l)}(n, j) = \frac{\Theta_{n \geq 1}}{\sqrt{2\pi}} [\chi_{-k}^{(\alpha)}(j) e^{-ikn} + r'_k{}^{(\alpha)} \chi_k^{(\alpha)}(j) e^{ikn}] + \frac{\Theta_{n \leq 0}}{\sqrt{2\pi}} t'_k{}^{(\alpha)} \chi_{-k}^{(\alpha)}(j) e^{-ikn}, \quad (1.42)$$

with  $k \in [0, \pi]$ . Both  $\psi_k^{(\alpha,r)}$  and  $\psi_k^{(\alpha,l)}$  as well as the bulk Bloch states  $\chi_{\pm k}^{(\alpha)}$  correspond to the eigenenergy  $\epsilon_k^{(\alpha)}$ . In the following, we focus on the band  $\alpha$  and omit the band index for brevity.

Inserting the ansatz (2.125) and (2.126) into the eigenvalue problem, we establish the scattering matrix

$$S_k = \begin{pmatrix} t_k & r'_k \\ r_k & t'_k \end{pmatrix}. \quad (1.43)$$

Its components read

$$t_k = t'_k = \lambda \frac{e^{2i\phi_k} - 1}{e^{2i\phi_k} - \lambda^2}, \quad (1.44)$$

$$r_k = \frac{\lambda^2 - 1}{e^{2i\phi_k} - \lambda^2} e^{2i\varphi_k(Z)}, \quad (1.45)$$

$$r'_k = \frac{\lambda^2 - 1}{e^{2i\phi_k} - \lambda^2} e^{2i\phi_k} e^{-2i\varphi_k(Z)}, \quad (1.46)$$

where  $\phi_k = \varphi_k(Z) - \varphi_k(1) - k$  is a gauge-invariant phase difference expressed in terms of the gauge-dependent phases  $\varphi_k(j)$  of the complex-valued components  $\chi_k(j)$ . By a direct calculation one can confirm the unitarity property  $S_k^\dagger S_k = 1$ , which implies both the orthogonality of  $\psi_k^{(\alpha,r)}$  and  $\psi_k^{(\alpha,l)}$  as well as their proper normalization.

The interface charge  $Q_I^{(\alpha)}$  associated with the band  $\alpha$ :

$$Q_I^{(\alpha)} = Q_F^{(\alpha)} + Q_P^{(\alpha)}, \quad (\text{Friedel and polarization parts}) \quad (1.47)$$

$$Q_F^{(\alpha)} = \sum_{m=-\infty}^{\infty} \left[ \rho^{(\alpha)}(m) - \rho_{\text{bulk}}^{(\alpha)}(m) \right] f(m), \quad (1.48)$$

$$Q_P^{(\alpha)} = \sum_{m=-\infty}^{\infty} \left[ \rho_{\text{bulk}}^{(\alpha)}(m) - \frac{1}{Z} \right] f(m), \quad (1.49)$$



with  $m = Z(n-1) + j$ . Here,  $\rho_{\text{bulk}}^{(\alpha)}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} dk |\chi_k^{(\alpha)}(j)|^2$  is the contribution from band  $\alpha$  to the charge at site  $m$  from the bulk Hamiltonian. As shown in Eq. (2.30), the polarization part to the interface charge vanishes  $Q_P^{(\alpha)} = 0$ . In turn, the Friedel part  $Q_F^{(\alpha)} \equiv Q_F$  amounts to

$$Q_F = \sum_{n=-\infty}^0 \int_{-\pi}^{\pi} \frac{dk}{2\pi} r_k^* \sum_{j=1}^Z \chi_k^2(j) e^{2i(k-i\eta)n} + \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} \frac{dk}{2\pi} r'_k \sum_{j=1}^Z \chi_k^2(j) e^{2i(k+i\eta)n} \quad (1.50)$$

$$= -1 + \int_{-\pi}^{\pi} \frac{dk}{2\pi} (r'_k - r_k^*) \sum_{j=1}^Z \chi_k^2(j) \frac{ie^{ik}}{2 \sin k}, \quad (1.51)$$

where  $\eta \rightarrow 0^+$  is a convergence factor. The last equality is only valid for  $\lambda^2 \neq 1$ , since the limits  $\lambda \rightarrow 1$  and  $\eta \rightarrow 0^+$  do not commute. In the translationally invariant case  $\lambda = 1$  there is no reflection at the interface, and one simply gets  $Q_F = 0$ .

In the following we prove that in general  $Q_F$  takes only integer values for arbitrary  $\lambda$ .

Let us introduce the two gauges: I)  $\chi_k^I$  with  $e^{2i\phi_k(Z)} = 1$ , i.e. the last component is real; II)  $\chi_k^{II}$  with  $e^{2i\phi_k(1)} = 1$ , i.e. the first component is real. Apparently,  $\chi_k^{II} = e^{i\phi_k + ik} \chi_k^I$ .

Next, we express the quantity  $Q_F + 1$  in the mixed form

$$\int_{-\pi}^{\pi} \frac{dk}{2\pi} \left[ -\frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \sum_{j=1}^Z [\chi_k^I(j)]^2 \frac{ie^{ik}}{2 \sin k} + \frac{\lambda^2 - 1}{\lambda^2 - e^{-2i\phi_k}} e^{-2i\phi_k} \sum_{j=1}^Z [\chi_k^{II}(j)]^2 \frac{ie^{-ik}}{2 \sin k} \right]. \quad (1.52)$$

In Ref. [87] we established that the components of the Bloch state in the gauge I have the form

$$\chi_k^I(j) = \frac{f_j^I e^{-ik} + g_j^I}{\sqrt{N_k^I}}, \quad 1 \leq j \leq Z-1, \quad (1.53)$$

$$\chi_k^I(Z) = \frac{s}{\sqrt{N_k^I}}, \quad (1.54)$$

where  $f_j^I$ ,  $g_j^I$ , and  $s$  are real-valued polynomial functions of  $\epsilon_k$ , and  $N_k^I = s^2 + \sum_{j=1}^{Z-1} |f_j^I e^{-ik} + g_j^I|^2$ . In that paper we also noted the following relations

$$-\text{Im}[\chi_k^{I\dagger} \frac{d}{dk} \chi_k^I] = \sum_{j=1}^{Z-1} \frac{(f_j^I)^2 + f_j^I g_j^I \cos k}{N_k^I}, \quad (1.55)$$

$$\begin{aligned} \sum_{j=1}^Z [\chi_k^I(j)]^2 \frac{ie^{ik}}{2 \sin k} &= \frac{ie^{ik}}{2 \sin k} + \sum_{j=1}^{Z-1} \frac{(f_j^I)^2 + f_j^I g_j^I e^{ik}}{N_k^I} \\ &= -\text{Im}[\chi_k^{I\dagger} \frac{d}{dk} \chi_k^I] + \frac{ie^{ik}}{2 \sin k} + \frac{i \sin k}{N_k^I} \sum_{j=1}^{Z-1} f_j^I g_j^I. \end{aligned} \quad (1.56)$$

In addition, it is also possible to derive the relation

$$\frac{\sin k}{N_k^I} \sum_{j=1}^{Z-1} f_j^I g_j^I = -\frac{1}{2s} \frac{ds}{dk}. \quad (1.57)$$

Hence,

$$\begin{aligned} \sum_{j=1}^Z [\chi_k^I(j)]^2 \frac{ie^{ik}}{2 \sin k} &= -\text{Im}[\chi_k^{I\dagger} \frac{d}{dk} \chi_k^I] - \frac{1}{2} \\ &\quad + \frac{i \cos k}{2 \sin k} - \frac{i}{2s} \frac{ds}{dk}. \end{aligned} \quad (1.58)$$

Let us now establish similar relations for  $\chi_k^{\text{II}}$ . We note that the Bloch state

$$\bar{\chi}_k^{\text{II}} = \begin{pmatrix} 0 & e^{-ik} 1_{(Z-1) \times (Z-1)} \\ 1 & 0 \end{pmatrix} \chi_k^{\text{II}} \quad (1.59)$$

is the eigenstate corresponding to the re-defined unit cell, which begins with the site 2, has the pre-last site  $Z$ , and ends with the site 1. Moreover, the component  $\bar{\chi}_k^{\text{II}}(Z)$  is real, and then by analogy with (2.143) it holds

$$\begin{aligned} \sum_{j=1}^Z [\bar{\chi}_k^{\text{II}}(j)]^2 \frac{ie^{ik}}{2 \sin k} &= -\text{Im}[\bar{\chi}_k^{\text{II}\dagger} \frac{d}{dk} \bar{\chi}_k^{\text{II}}] - \frac{1}{2} \\ &+ \frac{i \cos k}{2 \sin k} - \frac{i}{2\bar{s}} \frac{d\bar{s}}{dk}, \end{aligned} \quad (1.60)$$

where  $\bar{s}$  is a part of the representation for  $\bar{\chi}_k^{\text{II}}$ , which is analogous to (2.138), (2.139).

From (2.144) it follows that

$$\text{Im}[\bar{\chi}_k^{\text{II}\dagger} \frac{d}{dk} \bar{\chi}_k^{\text{II}}] = \text{Im}[\chi_k^{\text{II}\dagger} \frac{d}{dk} \chi_k^{\text{II}}] + |\bar{\chi}_k^{\text{II}}(Z)|^2 - 1, \quad (1.61)$$

and thus we find that

$$\begin{aligned} \sum_{j=1}^Z [\chi_k^{\text{II}}(j)]^2 \frac{ie^{-ik}}{2 \sin k} &= \sum_{j=1}^Z [\bar{\chi}_k^{\text{II}}(j)]^2 \frac{ie^{ik}}{2 \sin k} + |\bar{\chi}_k^{\text{II}}(Z)|^2 \\ &= -\text{Im}[\chi_k^{\text{II}\dagger} \frac{d}{dk} \chi_k^{\text{II}}] + \frac{1}{2} + \frac{i \cos k}{2 \sin k} - \frac{i}{2\bar{s}} \frac{d\bar{s}}{dk}. \end{aligned} \quad (1.62)$$

Inserting (2.143) and (2.147) into (2.137) and accounting the symmetry properties of integrands under the reflection  $k \rightarrow -k$ , we obtain

$$Q_F + 1 = \int_{-\pi}^{\pi} \frac{dk}{2\pi} \left\{ \text{Re} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \right] \left( \text{Im}[\chi_k^{\text{I}\dagger} \frac{d}{dk} \chi_k^{\text{I}}] - \text{Im}[\chi_k^{\text{II}\dagger} \frac{d}{dk} \chi_k^{\text{II}}] + 1 \right) + \text{Im} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \right] \left( \cot k - \frac{1}{2} \frac{d \ln(s\bar{s})}{dk} \right) \right\}. \quad (1.63)$$

From the transformation between the two gauges we obtain the relation

$$\text{Im}[\chi_k^{\text{I}\dagger} \frac{d}{dk} \chi_k^{\text{I}}] - \text{Im}[\chi_k^{\text{II}\dagger} \frac{d}{dk} \chi_k^{\text{II}}] = -1 - \frac{d\phi_k}{dk}. \quad (1.64)$$

A less obvious identity

$$s\bar{s} \sin^2 \phi_k = g_1^{\text{I}} \bar{f}_{Z-1}^{\text{II}} \sin^2 k \equiv \left( \prod_{j=1}^{Z-1} t_j^2 \right) \sin^2 k \quad (1.65)$$

follows from the identifications

$$\frac{\bar{s}}{\sqrt{N_k^{\text{II}}}} e^{-i\phi_k} = \frac{f_1^{\text{I}} + g_1^{\text{I}} e^{ik}}{\sqrt{N_k^{\text{I}}}}, \quad (1.66)$$

$$\frac{s}{\sqrt{N_k^{\text{I}}}} e^{i\phi_k} = \frac{\bar{f}_{Z-1}^{\text{II}} e^{-ik} + \bar{g}_{Z-1}^{\text{II}}}{\sqrt{N_k^{\text{II}}}}, \quad (1.67)$$

and the observation  $g_1^{\text{I}} = \bar{f}_{Z-1}^{\text{II}} = \prod_{j=1}^{Z-1} t_j$  which can be made on the basis of expressions quoted in Ref. [87]. Differentiating (2.150) with respect to  $k$  yields

$$\cot k - \frac{1}{2} \frac{d \ln(s\bar{s})}{dk} = \cot \phi_k \frac{d\phi_k}{dk} \quad (1.68)$$

With help of (2.149) and (2.153) we cast (2.148) to

$$Q_F + 1 = \int_{-\pi}^{\pi} \frac{dk}{2\pi} \left\{ -\operatorname{Re} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \right] \frac{d\phi_k}{dk} + \operatorname{Im} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \right] \cot \phi_k \frac{d\phi_k}{dk} \right\}. \quad (1.69)$$

Making the change of the integration variable  $k \rightarrow \phi_k$  and accounting possible multiple windings of the phase  $\phi_k$ , which are quantified by the integer winding number  $\operatorname{wn}[\phi_k] = \int_{-\pi}^{\pi} \frac{dk}{2\pi i} e^{-i\phi_k} \frac{d}{dk} e^{i\phi_k}$ , we express

$$\begin{aligned} Q_F + 1 &= \operatorname{wn}[\phi_k] \int_{-\pi}^{\pi} \frac{d\phi}{2\pi} \left\{ -\operatorname{Re} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi}} e^{2i\phi} \right] + \operatorname{Im} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi}} e^{2i\phi} \right] \cot \phi \right\} \\ &= \operatorname{wn}[\phi_k] \operatorname{sign}(\lambda^2 - 1). \end{aligned} \quad (1.70)$$

For the two decoupled chains ( $\lambda = 0$ ), we obtain

$$\begin{aligned} Q_F^{(\alpha)} &= -1 - \operatorname{wn}[\phi_k^{(\alpha)}] \\ &= \operatorname{wn}[\varphi_k^{(\alpha)}(1) - \varphi_k^{(\alpha)}(Z)], \end{aligned} \quad (1.71)$$

i.e. an integer number. This result persists in the whole range  $0 \leq \lambda < 1$ .

We conclude that the total interface charge  $Q_I(\lambda)$ , which might also include integer edge state contributions, is a sum of integers for any  $\lambda$ , and therefore  $Q_I(\lambda) = 0 \pmod{1}$ . On the other hand, since  $Q_B^R + Q_B^L = Q_I(\lambda = 0)$  by definition, we find  $Q_B^R + Q_B^L = Q_I(\lambda) \pmod{1}$ , in agreement with (2.4) for the model discussed in this Appendix.

### Symmetries for single-channel and nearest-neighbor hopping models

In this Appendix we prove the symmetry conditions (2.52) and (2.53) for the special case of a tight-binding model with one channel  $N_c = 1$  and nearest-neighbor hopping  $\delta = 0, \pm 1$ . In this case, the model is parametrized by  $Z$  on-site potentials  $v_m = v_{m+Z}$  and  $Z$  nearest-neighbor hoppings  $t_m = t_{m+Z}$  defined by

$$v_m = h_m(0) = v_m^* \quad , \quad t_m = -h_m(1) = -h_{m+1}(-1)^*. \quad (1.72)$$

Without loss of generality one can choose all  $t_m > 0$  real and positive since possible phases can be gauged away by a unitary transformation (see, e.g., Appendix A in Ref. [87] for a proof). The unitary transformation  $U_m$  must be a phase factor

$$U_m = e^{i\varphi_m} \quad , \quad \varphi_m = \varphi_{m+Z}. \quad (1.73)$$

Inserting these equations in the symmetry condition (2.39) for  $\Pi_n$  we find

$$v_m = v_{Z-m-n+1}, \quad (1.74)$$

$$t_m = e^{-i(\varphi_{Z-m-n} - \varphi_{Z-m-n+1})} t_{Z-m-n}. \quad (1.75)$$

Since  $t_m$  and  $t_{Z-m-n}$  are both positive this can only be fulfilled for  $U_m = U_{m+1}$  which is just a homogeneous and trivial phase factor. Therefore, we can set  $U_m = 1$  and find the condition (2.52).

Considering the other symmetry condition (2.40) for  $S_n$  we find

$$v_m = -v_{m-n}, \quad (1.76)$$

$$t_m = -e^{-i(\varphi_{m-n+1} - \varphi_{m-n})} t_{Z-m-n}. \quad (1.77)$$

Since  $t_m$  and  $t_{Z-m-n}$  are both positive this requires  $U_m = -U_{m+1}$  which, up to an unimportant common phase factor, is only realized for  $U_m = (-1)^m$ . This proves the condition (2.53).

## 2..2 Rational boundary charge in one-dimensional systems with interaction and disorder by Pletyukhov et. al. main ideas

### A. Hamiltonian, boundary and interface charges

### B. Rational quantization of boundary charge

#### 1. Translations

#### 2. Local inversion

#### 3. Non-local symmetries

## 2..3 Application to single-channel and nearest-neighbor hopping models by Pletyukhov et. al.

### Weyl semimetal physics at half-filling

### B. Low-energy theory 17

#### 1. Noninteracting Dirac model

#### 2. Interacting Dirac model

# Part II

## Theory

### 2 Rational boundary charge in 1D systems with interaction and disorder

1. Main Ideas (???????)
2. Introduction

#### Abstract

We study the boundary charge  $Q_B$  of generic semi-infinite one-dimensional insulators with translational invariance and show that non-local symmetries (i.e., including translations) lead to rational quantizations  $p/q$  of  $Q_B$ . In particular, we find that (up to an unknown integer) the quantization of  $Q_B$  is given in integer units of  $\frac{1}{2}\bar{\rho}$  and  $\frac{1}{2}(\bar{\rho}-1)$ , where  $\bar{\rho}$  is the average charge per site (which is a rational number for an insulator). This is a direct generalization of the known half-integer quantization of  $Q_B$  for systems with local inversion or local chiral symmetries to any rational value. Quite remarkably, this rational quantization remains valid even in the presence of short-ranged electron-electron interactions as well as static random disorder (breaking translational invariance). This striking stability can be traced back to the fact that local perturbations in insulators induce only local charge redistributions. We establish this result with complementary methods including density matrix renormalization group calculations, bosonization methods, and exact solutions for particular lattice models. Furthermore, for the special case of half-filling  $\bar{\rho} = \frac{1}{2}$ , we present explicit results in single-channel and nearest-neighbor hopping models and identify Weyl semimetal physics at gap closing points. Our general framework also allows us to shed new light on the well-known rational quantization of soliton charges at domain walls.

Charge fractionalization is a striking phenomenon which emerges in a variety of condensed matter systems of high interest such as the fractional quantum Hall effect [1, 2, 3, 4, 5], Luttinger liquids [6, 7, 8, 9, 10], and topological insulators [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. For its emergence a fundamental mechanism has been established via fractionally charged domain walls separating two systems with the same bulk spectrum but in different topological phases. This was analysed for a one-dimensional (1D) spinless Fermi gas coupled to a bosonic field with broken symmetry [22] and in polyacetylene chains due to electron-phonon coupling [23, 24]. This mechanism was further analysed for more general setups [35, 26] and a simple physical picture was proposed [27] to explain the fractional charge unit  $\frac{1}{2}$  via a  $Z$ -fold degenerate ground state generated by a charge-density wave (CDW) of wavelength  $\lambda = Za$  ( $a$  is the lattice constant) [28]. Within continuum field theories [29, 30, 31, 32] the fractional part of the soliton (or interface) charge  $Q_I$  was shown to be given by the Goldstone-Wilczek formula [31]  $Q_I = \frac{\delta\alpha}{2\pi}e \bmod(e)$ , where  $\delta\alpha$  is the phase difference between the two CDWs right and left to the interface. This interface charge is of purely topological nature, i.e., independent of the precise parameter values determining the domain wall. In addition, fluctuations of the soliton charges were analysed in continuum and lattice models showing that the fractional charge is a well-defined quantity [41, 34, 43, 36, 45, 38, 17].

Besides quantized soliton charges, charge quantization has also been studied at the boundary of topological insulators. Previous works focused on the special case of local inversion or local chiral symmetry as well as on non-interacting and clean systems, where the boundary charge  $Q_B$  is quantized in half-integer units. This was shown via the quantization of the Zak-Berry

phase  $\gamma$  in units of  $\pi$  [42, 24], which is related to the boundary charge by  $Q_B = -e\frac{\gamma}{2\pi} \bmod(e)$  [33, 31, 44, 35, 37, 41, 48, 39, 40]. The quantization of the Zak-Berry phase in the presence of local inversion symmetry has led to the notion of topological crystalline insulators (TCI) [17, 18, 19, 20, 21, 22], extending the standard classification schemes of topological insulators [24, 23, 25, 26, 27, 28, 29, 30, 31, 32], which are based on chiral, time reversal, and particle-hole symmetries only. In addition, combining local symmetries with translations (so-called non-local or non-symmorphic symmetries) new possibilities for TCIs have been predicted for 2D and 3D systems [66, 67, 68, 69, 70, 71].

The central topic of the present work is the generalization of the half-integer quantization of the boundary charge  $Q_B$  to any rational value  $\frac{p}{q}$  in generic 1D insulators. We will relate the occurrence of a rational boundary charge (RBC) to *non-local* symmetries, i.e. symmetries which can not be defined within the space of a single unit cell (see Appendix 5.1 for a summary of the precise definitions). Quite remarkably, in the presence of these symmetries, we will show that RBC can be easily understood in terms of the universal changes of  $Q_B$  under translations and local inversion. Stability of RBC is demonstrated since these transformation laws are not violated in the presence of short-ranged electron-electron interaction or static random disorder.

	$\bar{Q}_B \bmod(e)$	Transformation
$T_n$	$Q_B + n\bar{\rho}$	translation $ m\rangle \rightarrow  m+n\rangle$
$\Pi$	$-Q_B$	unit-cell-local inversion
$U$	$Q_B$	site-local (anti-)unitary
$\Pi_n = T_n U \Pi$	$-Q_B + n\bar{\rho}$	unitary/time-reversal
$S_n = T_{-n} U$	$Q_B - n\bar{\rho}$	chiral/particle-hole

Table 2: Transformations  $Q_B \rightarrow \bar{Q}_B$  of the boundary charge under the elementary transformations  $T_n$  (translation by  $n$  lattice sites towards the boundary),  $\Pi$  (local inversion within each unit cell, where the unit cell is defined as the one starting at the boundary of a semi-infinite system),  $U$  (unitary or anti-unitary operations within the channel space of a single site), and combinations of these transformations defining the operations  $\Pi_n = T_n U \Pi$  and  $S_n = T_{-n} U$ . Except  $U$  all transformation rules are  $\bmod(e)$  due to the possible occurrence of edge states.  $\bar{\rho} = e\frac{\nu}{Z}$  is the average charge per site which is a rational number in the insulating regime.  $\nu$  denotes the number of filled bands and  $Z$  is the number of lattice sites of a unit cell (with  $N_c$  channels per site). If  $U$  is unitary (anti-unitary),  $\Pi_n$  and  $S_n$  are unitary (anti-unitary) operations. Highlighted in color are the transformation rules that need to be compared between Tables 2 and 3 to obtain the rational boundary charge.

Our general and unified framework for the RBC is set up for generic 1D tight-binding models with any size  $Za$  of the unit cell and any number  $N_c$  of channels per site (like, e.g., described by spin, several orbitals, etc.). Importantly, our analytical study is not based on the representation of  $Q_B$  in terms of the Zak-Berry phase. It relies exclusively on the fundamental property of insulators, namely that local perturbations by external fields lead only to local charge redistributions, i.e., the corrections beyond a typical length scale  $\xi$  are exponentially small. This does not affect the fractional part of the boundary charge, since  $Q_B$  is defined via a macroscopic average on scales much larger than  $\xi$ . In addition, bound states (localized at the boundary) crossing the chemical potential due to the local perturbation, can only lead to a change of  $Q_B$  by an integer number. This local behavior, also known as the *nearsightedness principle* [72, 28] (NSP), is responsible for the universal features of topological insulators and

is also connected to the bulk-boundary correspondence [74, 75, 76, 77, 78, 79, 80, 81]. Furthermore, the same principle is responsible for charge pumping [50, 83] and leads to exponential localization of the excess density at boundaries and interfaces [48], such that  $Q_B$  and  $Q_I$  are well-defined quantities for insulators. Since the NSP is also valid for interacting and disordered systems, we can expect high stability of our results against short-ranged electron-electron interactions and static random disorder, as long as the gap of the insulator in which the chemical potential lies is not closed. Besides the general expectation of stability we will also support the NSP by numerical calculations based on density matrix renormalization group (DMRG) methods in the presence of short-ranged electron-electron interactions, by exact diagonalizations for static random disorder, and by analytical results for the particular example of two coupled non-interacting single-channel and nearest-neighbor hopping models. In addition, we establish analytically the stability against short-ranged interactions in continuum models by using the bosonization method.

	Symmetry	$\bar{Q}_B$	Quantization
$\Pi_n$	$\Pi_n H \Pi_n^\dagger = H$	$Q_B$	$\left. \vphantom{\begin{matrix} \Pi_n \\ S_n \end{matrix}} \right\} Q_B = \frac{1}{2}n\bar{\rho} \bmod\left(\frac{e}{2}\right)$
$S_n$	$S_n H S_n^\dagger = -H$ & $\frac{1}{2}$ -filling	$-Q_B \bmod(e)$	

Table 3: Transformations  $Q_B \rightarrow \bar{Q}_B$  of the boundary charge if the Hamiltonian  $H$  fulfils a symmetry by either commuting with  $\Pi_n$  or anti-commuting with  $S_n$ . For the symmetry  $S_n$  one needs in addition half-filling  $\bar{\rho} = e\frac{N_c}{2}$ . If  $U$  is a unitary (anti-unitary) operation,  $\Pi_n$  is a unitary (time-reversal) symmetry and  $S_n$  is a chiral (particle-hole) symmetry. For  $n = 0$  the operations  $\Pi_0$  and  $S_0$  are local symmetries acting within the space of a single unit-cell. For  $n \neq 0$  they are non-local symmetries, see Appendix 5.1 for our conventions to distinguish between local and non-local symmetries. By identifying the values for  $\bar{Q}_B$  from Tables 2 and 3 one obtains straightforwardly the rational quantization values  $Q_B = \frac{1}{2}n\bar{\rho} \bmod\left(\frac{e}{2}\right)$ . Highlighted in color are the transformation rules that need to be compared between Tables 2 and 3 to obtain the rational boundary charge

To sketch our derivation of RBC we have summarized our main results in the two tables 2 and 3. Table 2 lists the transformation of  $Q_B$  under basic operations, in particular under translation  $T_n$  by  $n$  lattice sites towards the boundary and under local inversion  $\Pi$  within each unit cell. Together with site-local transformations  $U$  leaving the boundary charge invariant, we define the two central operations  $\Pi_n = T_n U \Pi$  and  $S_n = T_{-n} U$ , which are non-local for  $n \neq 0$  since they contain a translation. Table 3 states the change of  $Q_B$  when the Hamiltonian has an explicit non-local symmetry by either commuting with  $\Pi_n$  or anti-commuting with  $S_n$  (for the symmetry  $S_n$  one needs in addition half-filling). Comparing the transformations of  $Q_B$  under  $\Pi_n$  and  $S_n$  stated in the two tables (marked in the same color) we arrive at the central result of RBC

$$Q_B = \frac{1}{2}n\bar{\rho} \bmod\left(\frac{e}{2}\right). \quad (2.1)$$

Here,  $\bar{\rho} = e\frac{\nu}{Z}$  is the average charge per site which is a rational multiple of  $e$  in the insulating regime, where  $\nu$  is the number of filled bands and  $Z$  is the number of sites of a unit cell. The trivial case of a local symmetry is  $n = 0$  leading to the well-known  $\frac{1}{2}$ -integer quantization of  $Q_B$ . Taking all integers  $n \neq 0$  into account we find that  $Q_B$  can take all rational quantization values. Due to the  $\bmod\left(\frac{e}{2}\right)$ -part our quantization rule shows that  $Q_B$  can always be written as a combination of multiples of two elementary quantization units:  $\frac{1}{2}\bar{\rho}$  and  $\frac{1}{2}(\bar{\rho} - 1)$ .



Besides the presence of a non-local symmetry of the Hamiltonian as stated in table 3, the central part of the proof of RBC is the transformation of  $Q_B$  under the two elementary operations  $T_n$  and  $\Pi$  of translations and local inversion according to table 2. They are the basic ingredients for the understanding of *all* universal properties of  $Q_B$  and are given by

$$Q_B \xrightarrow{T_n} \bar{Q}_B = Q_B + n\bar{\rho} \mod(e), \quad (2.2)$$

$$Q_B \xrightarrow{\Pi} \bar{Q}_B = -Q_B \mod(e). \quad (2.3)$$

Both transformation laws will be shown in this work to be ultimately related to the NSP which demonstrates their stability under short-ranged electron-electron interaction and static random disorder. Eq. (2.2) is a straightforward consequence of charge conservation since on average the charge  $n\bar{\rho}$  is moved into the boundary when the translation is described via an adiabatic process (up to an integer charge arising from edge states crossing the chemical potential during the adiabatic process). It has been used in a variety of recent works on single-channel and nearest-neighbor tight-binding models to analyse the universal phase-dependence  $Q_B(\varphi)$  as a function of a phase  $\varphi$  describing a continuous shift of the lattice towards the boundary [17, 18, 86, 87, 88]. Eq. (2.3) is a fundamental transformation which is based on the simple observation that local inversion of a semi-infinite system with a left boundary turns it to the same semi-infinite system with a right boundary [17]. Simple arguments based on the NSP will then show that the sum of these two boundary charges must be zero up to an integer charge.

The fact that the two elementary transformations (2.2) and (2.3) together with a non-local symmetry property of the Hamiltonian under  $\Pi_n$  or  $S_n$  explain both the RBC and its stability under interactions and disorder in a straightforward way is the central result of this work. We note that the interaction and the disorder have to fulfil the non-local symmetry property as well for our proof to be valid. Whereas homogeneous density-density interaction terms are obviously invariant under translations  $T_n$ , local inversion  $\Pi$ , and site-local transformations  $U$ , it might not be the case for some fixed disorder configuration. However, for random disorder the symmetry will be fulfilled on average and our numerical results confirm that the RBC is stable in the presence of random disorder. In addition, when the density-density interaction is not homogeneous, it is expected that it follows precisely the symmetry constraints imposed by the modulation of the on-site potentials and hopping terms.

Interestingly, we will show that the two universal transformation laws (2.2) and (2.3) shed also new light on the quantization of the interface charge. If the two lattices right and left to the interface have the same bulk spectrum and are only shifted relative to each other by  $\delta n$  sites, they are connected by the transformation  $T_{\delta n}\Pi$ . Therefore, if the two lattices are not connected to each other, one finds from (2.2) and (2.3) that the boundary charge  $Q_B^L$  of the left lattice is related to the boundary charge  $Q_B^R$  of the right lattice by  $Q_B^L = -Q_B^R + \delta n\bar{\rho} \mod(e)$ . Using the NSP, turning on some local coupling between the two lattices does not change the fractional part of the interface charge such that  $Q_I$  follows generically from

$$Q_I = Q_B^L + Q_B^R \mod(e) \quad (2.4)$$

$$= \delta n\bar{\rho} \mod(e). \quad (2.5)$$

As a result, we have extended the Goldstone-Wilczek formula to a discrete lattice and, in addition, have shown that it is stable in the presence of short-ranged electron-electron interactions and static random disorder.

We expect our results of RBC to be observable in experiments. Recent experiments in cold atom systems demonstrated that it is possible to get direct access to the boundary charge via the Zak-Berry phase [89] and to measure soliton charges of the SSH model [90]. In addition, concrete proposals for measuring topological solitons in solid state systems have been made such as carbon nanotubes [91], graphene nanoribbons [92], and Rashba nanowires [93, 94]. Here,



scanning single-electron transistor techniques allow for the direct measurement of local charges [95, 96, 97, 98, 99]. Moreover, the occurrence of interface states due to the quantization of the Zak phase has been measured in phononic crystals [99]. Besides these materials promising candidates to measure the boundary charge are quantum dot arrays as proposed in Ref. [17]. Similiar to cold atom systems, quantum dot arrays have the particular advantage of control over all parameters to implement on demand the specific non-local symmetries needed for RBC.

As an interesting application of our general framework we will discuss the case of a single-channel (i.e.,  $N_c = 1$ ) and nearest-neighbor hopping model. Of particular interest is the case of half-filling,  $\bar{\rho} = \frac{e}{2}$ , where one obtains from (2.1) the two universal quantization classes  $Q_B = \frac{e}{2} \bmod(\frac{e}{2})$  and  $Q_B = \frac{e}{4} \bmod(\frac{e}{2})$ . The first is the usual one present also for local inversion or local chiral symmetries. In contrast, the second was to the best of our knowledge not discussed before and is only possible for a *non-local* symmetry. We present an explicit realization of these classes in terms of a lattice model with equal hopping amplitudes and a harmonic modulation of the on-site potentials. Controlling the offset of the modulation by a phase-variable  $\varphi$  this model is of relevance for the integer quantum Hall effect (IQHE) (where  $\varphi$  corresponds to the transverse quasimomentum in a 2D quantum Hall setup) [18]. At half-filling (where  $Z$  must be even to open a gap), the model has the non-local chiral symmetry  $S_{Z/2} = T_{-Z/2}U$  with  $U|m\rangle = (-1)^m|m\rangle$  ( $|m\rangle$  denotes the state at lattice site  $m$ ). According to (2.1) this leads to the quantization values  $Q_B = \frac{Z}{8}e \bmod(\frac{e}{2})$ , i.e., the two quantization classes in terms of  $\frac{e}{2}$  or  $\frac{e}{4}$  are obtained for  $Z = 4, 8, 12, \dots$  and for  $Z = 2, 6, 10, \dots$ , respectively. The model has the advantage that the chiral symmetry  $S_{Z/2}$  holds for any phase  $\varphi$  of the potential modulation. This pins  $Q_B(\varphi)$  to quantized plateaus which change abruptly by  $\pm\frac{e}{2}$  at gap closing points. This leads to Weyl semimetal physics since edge modes connecting the gap closing points play the role of Dirac arcs. Despite the fact that in this case the Chern number vanishes (leading to zero Hall current), we find a non-trivial quantization effect of the boundary charge  $Q_B$ .

For single-channel and nearest-neighbor hopping models with very small gaps we will also set up a low-energy continuum theory via a Dirac model in 1+1 dimensions with a complex gap parameter  $\Delta = |\Delta|e^{i\alpha}$ , in analogy to the study of interface charges via the Goldstone-Wilczek formula. For a semi-infinite system we obtain the following universal result for the boundary charge:

$$Q_B = \frac{\alpha}{2\pi}e + \frac{e}{4} \bmod(e). \quad (2.6)$$

Interestingly, the boundary charge is insensitive to the gap size and reveals a linear behavior as function of the phase of the gap parameter. If the original lattice model is at half-filling and pure potential modulation is realized (as discussed above for the Weyl case), we find the symmetry  $S_{Z/2}$  for any phase  $\varphi$  of the CDW [100]. In the corresponding continuum model we will show that the parameter  $\alpha$  is obtained from the interference of two paths connecting right and left movers at the two Fermi points  $\pm k_F$ . For the two classes  $Z = 4, 8, 12, \dots$  and  $Z = 2, 6, 10, \dots$ , we find a phase-locking effect pinning  $\alpha$  to odd or even multiples of  $\frac{\pi}{2}$ , respectively. These two cases correspond to the two quantization classes of  $Q_B$  in terms of  $\frac{e}{2}$  or  $\frac{e}{4}$ , respectively, proving consistency of the continuum theory with our general framework.

The paper is organized as follows. Section 3. is devoted to the general framework to realize RBC in generic 1D insulators. We describe the model and the definition of boundary and interface charges in Section 3.1, and the RBC is analysed in Section 3.2. The basic transformation laws (2.2) and (2.3) are derived in Sections 3.2 and 3.2. Combining the two transformations we find the Goldstone-Wilczek formula (2.5) for the interface charge. In Section 3.2 we combine the transformation laws with non-local symmetries of the Hamiltonian and prove the central result (2.1). We proceed in Section 4. with an application of our general framework to the case of single-channel and nearest-neighbor hopping models. In Section 4.1 we describe Weyl semimetal physics at half-filling and discuss the connection to the IQHE by analysing

the universal phase-dependence of the boundary charge, the Diophantine equation, and the Hall current in the presence of a gap closing. An effective low-energy description of boundary and interface charges in terms of a continuum Dirac model in  $1 + 1$  dimensions is provided in Section 4.2. The derivation of the model in the noninteracting and interacting case is given in Sections 4.2 and 4.2, respectively. The universal formula (2.6) for the boundary charge and the Goldstone-Wilczek formula for the interface charge are presented in Section 4.2. We close with a summary and outlook in Section 7..

Throughout this work we use units such that  $\hbar = e = a = 1$ .

### 3. General framework

In this section we describe the general framework to derive the central transformations (2.2) and (2.3) of the boundary charge  $Q_B$  under translations and local inversion, respectively. We identify the non-local symmetries  $\Pi_n$  and  $S_n$  leading to the rational quantization values (2.1) of the boundary charge. In addition, we show that the Goldstone-Wilczek formula (2.5) for interface charges follows straightforwardly from the transformation laws.

#### 3.1 Hamiltonian, boundary and interface charges

We consider a generic translationally invariant tight-binding model in 1D with arbitrary short-ranged hopping and  $N_c$  channels per lattice site. For the infinite (bulk) case, the single-particle Hamiltonian reads

$$H_{\text{bulk}} = \sum_{m=-\infty}^{\infty} \sum_{\delta=-\delta_{\text{max}}}^{\delta_{\text{max}}} \underline{c}_{m+\delta}^\dagger \underline{h}_{\underline{m}}(\delta) \underline{c}_m. \quad (2.7)$$

Here,  $m$  denotes the lattice site index and  $\delta_{\text{max}}$  is the range of the hopping. The components  $c_{m\sigma}$  of the  $N_c$ -dimensional vector  $\underline{c}_m$  annihilate an electron on site  $m$  in channel  $\sigma = 1, \dots, N_c$ .  $\underline{h}_{\underline{m}}(\delta)$  is a generic  $N_c \times N_c$ -matrix describing the coupling between the channels of lattice site  $m$  and  $m + \delta$ . Translational invariance and hermiticity require the properties

$$\underline{h}_{\underline{m}}(\delta) = \underline{h}_{\underline{m}+Z}(\delta), \quad (2.8)$$

$$\underline{h}_{\underline{m}}^\dagger(\delta) = \underline{h}_{\underline{m}+\delta}(-\delta), \quad (2.9)$$

where  $Z$  is the number of lattice sites of a unit cell. Semi-infinite systems extending to the right or left side are defined by the Hamiltonians  $H_{R/L,n}$  by starting/ending the bulk Hamiltonian  $H_{\text{bulk}}$  at site  $m = n + 1$  and  $m = n$ , respectively, see Fig. 5 for illustration. Since the numeration of the sites is arbitrary one can alternatively label the sites by  $m = 1, 2, \dots$  for the right lattice and by  $m = 0, -1, -2, \dots$  for the left lattice, and formally define in compact notation

$$H_{R,n} = \sum_{m=1}^{\infty} \sum_{\substack{\delta \\ m+\delta > 0, |\delta| \leq \delta_{\text{max}}}} \underline{c}_{m+\delta}^\dagger \underline{h}_{\underline{m}+n}(\delta) \underline{c}_m, \quad (2.10)$$

$$H_{L,n} = \sum_{m=-\infty}^0 \sum_{\substack{\delta \\ m+\delta \leq 0, |\delta| \leq \delta_{\text{max}}}} \underline{c}_{m+\delta}^\dagger \underline{h}_{\underline{m}+n}(\delta) \underline{c}_m. \quad (2.11)$$



Figure 5: Illustration of the semi-infinite Hamiltonians  $H_{R,n}$  (right figure) and  $H_{L,n}$  (left figure). The Hamiltonian  $H_{R,n}$  is obtained from the bulk Hamiltonian  $H_{\text{bulk}}$  by starting it at site  $m = n + 1$ , whereas  $H_{L,n}$  is obtained by ending  $H_{\text{bulk}}$  at site  $m = n$ .

If an interface is studied between the two semi-infinite systems  $H_{R,n}$  on the right side and  $H_{L,n'}$  on the left side, we take any short-ranged coupling  $V_I$  defined within the interface region  $M_L \leq m \leq M_R$  (with  $|M_{L,R}| \sim O(Z)$ )

$$V_I = \sum_{M_L \leq m, m' \leq M_R} \underline{c}_m^\dagger \underline{v}_{mm'} \underline{c}_{m'}, \quad (2.12)$$

and define the total Hamiltonian by

$$H_{I,nn'} = H_{R,n} + H_{L,n'} + V_I. \quad (2.13)$$

For the study of the stability against short-ranged electron-electron interactions, we take a density-density interaction of the form

$$V_{ee} = \frac{1}{2} \sum_{m \neq m'} u(m - m') \hat{\rho}(m) \hat{\rho}(m'), \quad (2.14)$$

where  $u(m) = u(-m)$  is only nonzero for  $|m| \leq m_{\text{max}} \sim O(1)$  and  $\hat{\rho}(m) = \underline{c}_m^\dagger \underline{c}_m$ . If a semi-infinite lattice is studied, the sum over the sites is restricted to the corresponding regions. In the presence of an interface the short-ranged electron-electron interaction between the sites of the left and right lattices is also a coupling term which can be included alternatively in  $V_I$  by adding these many-particle terms to the interface coupling.

We will also test stability with respect to static random disorder. In this case we include a quenched onsite disorder

$$H_{\text{dis}} = \sum_m d(m) \hat{\rho}(m), \quad (2.15)$$

where  $d(m)$  is drawn from a uniform distribution  $d(m) \in [-d/2, d/2]$ . Of course, more complicated (channel-dependent or hopping) forms of disorder can be considered, which, however, will not change the main thrust of the arguments presented here.

In the insulating regime, where the excess density falls off exponentially on scale  $\xi$  into the bulk [48, 87] starting from a boundary or interface, the observables of interest are the boundary charges  $Q_{B,n}^R$  and  $Q_{B,n'}^L$  of the semi-infinite systems described by  $H_{R,n}$  and  $H_{L,n'}$ , respectively, and the interface charge  $Q_{I,nn'}$  of the Hamiltonian  $H_{I,nn'}$ , defined by a macroscopic average on scales much larger than  $Z$  and  $\xi$

$$Q_{B,n}^R = \sum_{m=1}^{\infty} (\rho(m) - \bar{\rho}) f(m), \quad (2.16)$$

$$Q_{B,n'}^L = \sum_{m=-\infty}^0 (\rho(m) - \bar{\rho}) f(m), \quad (2.17)$$

$$Q_{I,nn'} = \sum_{m=-\infty}^{\infty} (\rho(m) - \bar{\rho}) f(m). \quad (2.18)$$

Here,  $\rho(m) = \langle \hat{\rho}(m) \rangle$  is the total charge at site  $m$  in a grand canonical ensemble with respect to the corresponding Hamiltonians  $H_{R,n}$ ,  $H_{L,n'}$ , and  $H_{I,nn'}$ , respectively. We assume zero temperature and the chemical potential  $\mu$  to lie in some given band gap of the insulator. Further,  $\bar{\rho}$  is the average charge per site for the translationally invariant bulk Hamiltonian (2.7) defined by

$$\bar{\rho} = \frac{1}{Z} \sum_{j=1}^Z \rho_{\text{bulk}}(j) = \frac{\nu}{Z}, \quad (2.19)$$

where  $\nu$  is the filling factor defined as the number of occupied bands. The function  $f(m)$  is the envelope function of a charge measurement probe falling off smoothly from unity to zero, see Fig. 34.

To simplify notations, if no index  $n$  is displayed, we assume implicitly  $n = 0$ , i.e.,

$$H_{R/L} \equiv H_{R/L,0}, \quad H_I \equiv H_{I,00}, \quad (2.20)$$

$$Q_B^{R/L} \equiv Q_{B,0}^{R/L}, \quad Q_I \equiv Q_{I,00}. \quad (2.21)$$

We note that we have not used this convention in the introductory part (to avoid too many sub-indices at the beginning) where we denoted by  $Q_B^{R/L}$  and  $Q_I$  the boundary and interface charges for the systems  $H_{R/L,n}$  or  $H_{I,nn'}$  under consideration. Furthermore, we used implicitly the convention  $Q_B \equiv Q_B^R$  in the introductory part.

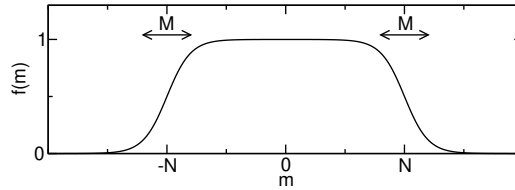


Figure 6: Sketch of the envelope function  $f(m)$  with  $N \gg M \gg Z, \xi$ .

Due to the NSP, it is expected that  $Q_{I,nn'}$  is independent of the coupling  $V_I$  (up to an integer). For particular examples we demonstrate this in Appendix 5.2 via exact diagonalization and DMRG calculations in the presence of static random disorder or short-ranged electron-electron interaction. Moreover, we show in Appendix 5.3 analytically that  $Q_{I,nn'}$  is independent of the size of a single link between two noninteracting and 1-channel nearest-neighbor hopping models. As a result we find [see Eq. (2.4)] that the interface charge is the independent sum of the boundary charges of the left and right lattice

$$Q_{I,nn'} = Q_{B,n}^R + Q_{B,n'}^L \mod(1). \quad (2.22)$$

In addition, we conclude that also the boundary charge does not change, when a local perturbation is added close to the boundary. In Section 3.2 we will furthermore show that the interface charge depends only on the relative difference  $\delta n = n' - n$ , i.e.,

$$Q_{I,nn'} = Q_{I,\delta n}, \quad \delta n = n - n', \quad (2.23)$$

see the Goldstone-Wilczek formula Eq. (2.5) or Eq. (2.33) below.

### 3.2 Rational quantization of boundary charge

In this section we show how the boundary charge transforms under translations and local inversion according to Eqs. (2.2) and (2.3). Together with certain symmetry requirements of

the Hamiltonian we will determine the rational quantization values for the boundary charge given by Eq. (2.1). In addition, we show that the Goldstone-Wilczek formula Eq. (2.5) for the interface charge is based only on the fundamental transformation laws of the boundary charge. For the special case of noninteracting and clean systems, we show in the Supplemental Material (SM) [26] that the same results can also be obtained from the transformation of the Zak-Berry phase under translations and local inversion.

### Translations

A translation  $T_n = T_{-n}^\dagger$  of the lattice by  $n$  sites to the left is defined by the operation

$$\underline{h}_{\underline{m}}(\delta) \xrightarrow{T_n} \underline{h}_{\underline{m}+n}(\delta). \quad (2.24)$$

Applying this to Eqs. (2.10) and (2.11), we evidently get

$$H_R \xrightarrow{T_n} H_{R,n}, \quad H_L \xrightarrow{T_n} H_{L,n}, \quad (2.25)$$

as illustrated in Fig. 7. Performing the transformation (2.24) via an adiabatic process, the lattice is shifted as a whole by  $n$  sites to the left, i.e., due to charge conservation, the average charge  $n\bar{\rho}$  will move into the left boundary of the right lattice and will move out of the right boundary of the left lattice. Since the boundary charges  $Q_{B,n}^{R/L}$  are defined as a macroscopic average via Eqs. (2.16) and (2.17), we get

$$Q_{B,n}^{R/L} = Q_B^{R/L} \pm n\bar{\rho} \pmod{1}. \quad (2.26)$$

Together with Eq. (2.25) this provides the following transformation of the boundary charges under translation:

$$Q_B^{R/L} \xrightarrow{T_n} \bar{Q}_B^{R/L} = Q_B^{R/L} \pm n\bar{\rho} \pmod{1}, \quad (2.27)$$

which proves Eq. (2.2). These relations are exact and do not depend on the presence or absence of short-ranged electron-electron interaction or random disorder, see Appendix 5.2 and bosonization studies in Section 4.2. They are based on the same arguments as charge pumping [50, 83] and have been extensively used recently for noncyclic adiabatic processes to analyse the universal average slope of the phase-dependence of the boundary charge [17, 18, 86, 87]. The unknown integer arises since bound states (at the boundaries) can cross the chemical potential during the adiabatic process leading to discrete integer jumps of the boundary charge.

Alternatively, Eq. (2.26) can also be derived directly from the NSP. Since local perturbations at the boundary do not change  $Q_B^R$  (up to an integer) we can add to  $H_R$  an infinitely high potential on the first  $n$  sites such that  $\rho(m) = 0$  for  $m = 1, 2, \dots, n$ . This leaves for the boundary charge from Eq. (2.16) only the contribution  $-n\bar{\rho}$  for the first  $n$  sites and  $Q_{B,n}^R$  for the rest. Using the invariance due to the NSP this gives  $Q_B^R = Q_{B,n}^R - n\bar{\rho}$  leading to Eq. (2.26). In an analogous way one can prove Eq. (2.26) for  $Q_B^L$  by starting from  $H_{B,n}^L$  and putting an infinite potential on the last  $n$  sites.

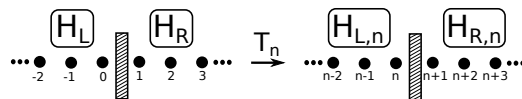


Figure 7: Illustration of the translation by  $n$  lattice sites for the left and right semi-infinite lattice. In both cases the lattice is moved to the left. This means that for  $H_R \rightarrow H_{R,n}$  the lattice is moved by  $n$  sites *towards* the boundary whereas, for  $H_L \rightarrow H_{L,n}$ , the lattice is moved by  $n$  sites *away* from the boundary. As a consequence, the boundary charge increases or decreases by  $n\bar{\rho}$ , respectively.

### Local inversion

Local inversion  $\Pi$  is defined as a symmetry operation performing an inversion in each unit cell separately, where the unit cell is defined such that it starts at the boundary. Since the unit cell defined in this way depends on the index  $n$  of the Hamiltonians  $H_{R/L,n}$  we discuss the case  $n = 0$  in the following (see the discussion at the end of Section 3.2 for transformations defined with respect to other choices of the unit cell). In this case the unit cell starting at the boundary consists of the sites  $m = 1, \dots, Z$  and is identical to the one for the bulk Hamiltonian  $H_{\text{bulk}}$ . In this subspace the transformation  $\Pi$  is formally defined by

$$\underline{h}_m(\delta) \xrightarrow{\Pi} \underline{h}_{Z-m+1}(-\delta) = [\underline{h}_{Z-m+1-\delta}(\delta)]^\dagger, \quad (2.28)$$

where we have used the hermiticity condition (2.9) in the last equality. Using the periodicity condition (2.8), this defines  $\Pi$  for all  $m$ . Applying an inversion turns the semi-infinite system  $H_R$  with a left boundary obviously to the semi-infinite system  $H_L$  with a right boundary

$$H_R \xrightarrow{\Pi} H_L, \quad (2.29)$$

as illustrated in Fig. 8. Taking the Hamiltonian (2.13) without any coupling  $V_I = 0$  between the left and right lattice, the interface charge is obviously given by  $Q_I = Q_B^L + Q_B^R$ . On the other hand, taking for  $V_I$  exactly the coupling corresponding to the bulk Hamiltonian (2.7) we get a translational invariant lattice everywhere with a periodic bulk density  $\rho_{\text{bulk}}(m) = \rho_{\text{bulk}}(m + Z)$ . The corresponding interface charge  $Q_{I,\text{bulk}}$  vanishes since

$$\begin{aligned} Q_{I,\text{bulk}} &= \sum_m (\rho_{\text{bulk}}(m) - \bar{\rho}) f(m) \\ &= \sum_n \sum_{j=1}^Z (\rho_{\text{bulk}}(j) - \bar{\rho}) f(Zn + j) \\ &= \sum_n f(Zn) \sum_{j=1}^Z (\rho_{\text{bulk}}(j) - \bar{\rho}) \\ &\quad + \sum_n f'(Zn) \sum_{j=1}^Z (\rho_{\text{bulk}}(j) - \bar{\rho}) j = 0, \end{aligned} \quad (2.30)$$

where we have used  $f(Zn + j) \approx f(Zn) + f'(Zn)j$  in the second step and  $\sum_n f'(Zn) \approx \int dx f'(Zx) = 0$  together with the definition (2.19) of  $\bar{\rho}$  in the third step. The approximations become exact in the limit of infinite parameters  $N$  and  $M$  defining the smoothness of the envelope function via Fig. 34. Finally, due to the NSP, the interface charge can only change mod(1) when switching on  $V_I$ , leading to

$$Q_B^L + Q_B^R = 0 \pmod{1}. \quad (2.31)$$

Together with (2.29) this provides the following transformation of the boundary charges under inversion

$$Q_B^{R/L} \xrightarrow{\Pi} \bar{Q}_B^{R/L} = -Q_B^{R/L} \pmod{1}, \quad (2.32)$$

which proves Eq. (2.3). The universal relation (4.154) has also been found recently in Ref. [87] for the special case of a noninteracting single-channel and nearest-neighbor hopping model, where the unknown integer has been specified for a single band. Again, we emphasize that it is also valid in the presence of short-ranged electron-electron interaction and random disorder



since we only used the NSP to derive it, see also Appendix 5.2 and bosonization studies in Section 4.2.

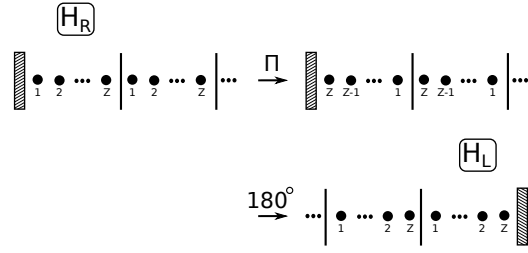


Figure 8: Illustration of local inversion by indicating the lattice sites of the unit cells. Inverting the lattice sites within a unit cell via the interchange  $m \leftrightarrow Z - m + 1$  and turning the system by  $180^\circ$  one finds that  $H_R$  is transformed to  $H_L$ .

We note that, by inserting the relations (2.26) and (4.154) into the formula (2.22) for the interface charge, we obtain straightforwardly the Goldstone-Wilczek formula (2.5) for a discrete lattice

$$Q_{I,nn'} = (n - n') \bar{\rho} \pmod{1}, \quad (2.33)$$

with  $\delta n \equiv n - n'$ . We conclude that the charge quantization at interfaces separating regions in different topological phases with the same bulk spectrum is fundamentally related to the NSP and the transformation laws of the boundary charge under translation and local inversion. This provides a very elegant proof of the Goldstone-Wilczek formula and shows at the same time that it is stable under short-ranged electron-electron interaction and static random disorder.

### Non-local symmetries

If the transformations (2.27) and (2.32) of the boundary charges  $Q_B^{R/L}$  under translations and local inversion are combined with explicit symmetries of the Hamiltonian, one can straightforwardly prove the rational quantization of the boundary charge. We note that symmetries are always defined with respect to the bulk Hamiltonian  $H_{\text{bulk}}$  and we discuss the consequences for the boundary charges  $Q_B^{R/L}$ , where the unit cell starting/ending at the boundary is the same as the one chosen for  $H_{\text{bulk}}$ . We define *local* symmetries by transformations acting only in the space of a single unit cell as they are used in the usual 10-fold classification schemes of TIs [24, 23, 25, 26, 27, 28, 29, 30, 31, 32] in terms of local chiral, time-reversal, and particle-hole symmetries. The case when the local symmetry is defined with respect to another choice of the unit cell is always discussed separately if relevant, see also the detailed discussion at the end of this section. For a summary of the precise definitions and our conventions to distinguish between local and non-local symmetries we refer to Appendix 5.1.

First, we note that all (anti-)unitary transformations  $U_m = U_{m+Z}$  acting only in the channel space of lattice site  $m$  commute with the operator  $\underline{c}_m^\dagger \underline{c}_m$  and, therefore, leave the boundary charge invariant. These are transformations defined by

$$\underline{h}_m(\delta) \xrightarrow{U} U_{m+\delta}^\dagger \underline{h}_m(\delta) U_m. \quad (2.34)$$

Secondly, we define two fundamental operations  $\Pi_n$  and  $S_n$  by combining  $U$  with local inversion and translations

$$\Pi_n = T_n U \Pi, \quad S_n = T_{-n} U. \quad (2.35)$$

$\Pi_n$  and  $S_n$  are defined in such a way that the boundary charge transforms according to

$$Q_B^{R/L} \xrightarrow{\Pi_n} \bar{Q}_B^{R/L} = -Q_B^{R/L} \pm n\bar{\rho} \mod(1), \quad (2.36)$$

$$Q_B^{R/L} \xrightarrow{S_n} \bar{Q}_B^{R/L} = Q_B^{R/L} \mp n\bar{\rho} \mod(1), \quad (2.37)$$

where we have used (2.27) and (2.32).

Finally, we assume that the Hamiltonian either commutes with  $\Pi_n$  or anticommutes with  $S_n$

$$\underline{h}_m(\delta) \xrightarrow{\Pi_n} \underline{h}_m(\delta) \quad \text{or} \quad \underline{h}_m(\delta) \xrightarrow{S_n} -\underline{h}_m(\delta). \quad (2.38)$$

Using Eqs. (2.24), (2.28), and (2.34) this requires one of the following symmetry conditions

$$\Pi_n : \quad \underline{h}_m(\delta) = U_{Z-m-n+1-\delta}^\dagger \cdot [\underline{h}_{Z-m-n+1-\delta}(\delta)]^\dagger U_{Z-m-n+1}, \quad (2.39)$$

$$S_n : \quad \underline{h}_m(\delta) = -U_{m-n+\delta}^\dagger \underline{h}_{m-n}(\delta) U_{m-n}. \quad (2.40)$$

In contrast to all previous relations this defines a certain *non-local* symmetry which the Hamiltonian has to fulfil. As we will show below the symmetry implies rational quantization values of the boundary charge. When  $U$  is a unitary transformation, (2.39) denotes a unitary symmetry  $\Pi_n$  and (2.40) a chiral symmetry  $S_n$ . Similarly, when  $U$  is an anti-unitary transformation, (2.39) denotes a time-reversal symmetry  $\Pi_n$  and (2.40) a particle-hole symmetry  $S_n$ . Both symmetries are *non-local* for  $n \neq 0$  since they involve translations. A special case is  $\Pi_n$  which, except for  $Z$  even and  $n$  odd, can be turned into a local symmetry but with respect to another choice of the unit cell. This follows since  $\Pi_n$  is an inversion symmetry around the axis  $m = \frac{1}{2}(Z - n + 1)$  (for  $n$  even) or  $m = Z - \frac{1}{2}(n - 1)$  (for  $n$  odd). This leads to a local site-inversion symmetry for  $Z$  odd, and to a local bond-inversion symmetry for both  $Z$  and  $n$  even. However, for  $Z$  even and  $n$  odd, we obtain a site-inversion symmetry but the corresponding unit cell contains only half of the sites at the boundaries which is not possible for tight-binding models, see Fig. 9 for illustration. Using the terminology of symmorphic and non-symmorphic symmetries one should call  $\Pi_n$  a symmorphic symmetry depending on the quasimomentum  $k$  (for  $Z$  even and  $n$  odd, sometimes also called unconventional non-symmorphic symmetry) since it returns to the same lattice site when applied twice [102, 103, 104, 105], whereas  $S_n$  is a non-symmorphic symmetry [106].

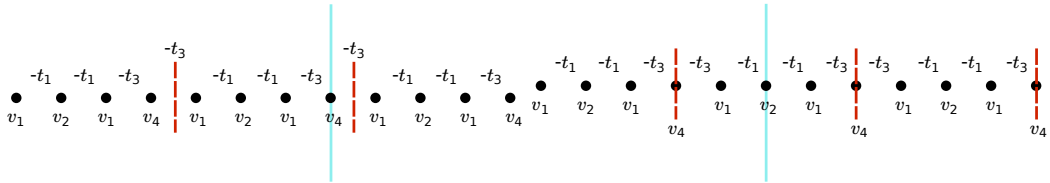


Figure 9: Visualization of the nonlocal inversion symmetry  $\Pi_n$  for  $n = 1$  and  $Z = 4$  for a nearest-neighbor hopping model with one channel per site.  $v_j$  denote the on-site potentials and  $-t_j$  are the hoppings. The red dashed vertical lines indicate the boundaries between the unit cells and the blue solid vertical line is the symmetry axis of inversion symmetry. In the lower figure the unit cell is redefined such that the symmetry becomes a local one. However, this is not possible for a tight-binding model since the unit cell contains only one half of a site at the boundaries.

For the (anti-)unitary symmetry  $\Pi_n$ , the boundary charge can obviously not change since the Hamiltonian is invariant. For the particle-hole or chiral symmetry  $S_n$  each eigenstate  $|\psi\rangle$



of the Hamiltonian with energy  $\epsilon$  has a corresponding eigenstate  $S_n|\psi\rangle$  with negative energy  $-\epsilon$ . Therefore, the operation  $S_n$  transforms the boundary charge  $Q_B^-$  (with  $Q_B \equiv Q_B^{R/L}$ ) from all states with negative energy to the boundary charge  $Q_B^+$  from all states with positive energy. When all states are filled we get  $\rho(m) = \bar{\rho} = N_c$  giving zero boundary charge, i.e.,  $Q_B^+ + Q_B^- = 0$ . At half-filling  $\bar{\rho} = \frac{N_c}{2}$ , the chemical potential is located somewhere in the gap near zero energy. In this case the boundary charge is given by  $Q_B = Q_B^- \bmod(1)$ , where the integer arises from edge states contributing an integer number to the boundary charge. As a consequence, up to an integer, we find that the boundary charge changes sign under  $S_n$  at half-filling. Therefore, we get the following relations

$$Q_B^{R/L} \xrightarrow{\Pi_n} \bar{Q}_B^{R/L} = Q_B^{R/L}, \quad (2.41)$$

$$Q_B^{R/L} \xrightarrow{S_n} \bar{Q}_B^{R/L} = -Q_B^{R/L} \bmod(1), \text{ for } \bar{\rho} = \frac{N_c}{2}. \quad (2.42)$$

Taking these equations together with (2.36) and (2.37), we arrive for both symmetries at the following rational quantization values of the boundary charge

$$Q_B^{R/L} = \pm \frac{1}{2} n \bar{\rho} \bmod\left(\frac{1}{2}\right), \quad (2.43)$$

which proves our central result (2.1).

Eq. (2.43) shows that the quantization of  $Q_B^{R/L}$  can always be written as

$$Q_B^{R/L} = n_1 \frac{1}{2} \bar{\rho} + n_2 \frac{1}{2} (\bar{\rho} - 1), \quad (2.44)$$

with some integers  $n_1$  and  $n_2$ . As a consequence, the quantization units of the boundary charge are given by  $\frac{1}{2}\bar{\rho}$  and  $\frac{1}{2}(\bar{\rho} - 1)$  in contrast to the quantization unit  $\bar{\rho}$  of interface charges, see Eq. (2.33). Furthermore, the quantization of  $Q_B^{R/L}$  requires certain symmetries of the bulk Hamiltonian, whereas the quantization of  $Q_I$  is only related to a symmetry relation between the lattices left and right to the interface, which are connected by  $T_{\delta n}\Pi$ . For the special case  $n = 0$ , one recovers from (2.43) the known quantization of the boundary charge in half-integer units [42, 24] in the presence of local symmetries.

One can ask the delicate question what happens if a local symmetry is not defined with respect to the unit cell starting at the boundary but with respect to any choice of the unit cell. This is equivalent to the question how the boundary charge changes when the semi-infinite system is cut off at a different site such that it starts with site  $n' + 1$  (for  $H_R$ ) or ends with site  $n'$  (for  $H_L$ ). According to the transformation (2.27) of  $Q_B^{R/L}$  under translations this leads to a shift of  $Q_B^{R/L}$  by  $\pm n' \bar{\rho} \bmod(1)$ . Therefore, for systems with local symmetries with respect to any definition of the unit cell, one can realize all quantization values

$$Q_B^{R/L} = \pm n' \bar{\rho} \bmod\left(\frac{1}{2}\right), \quad (2.45)$$

where the  $\bmod(\frac{1}{2})$  contribution stems from local symmetries defined with respect to the unit cell starting at the boundary. This provides the quantization units  $\bar{\rho}$  and  $\frac{1}{2}$ . In contrast, the presence of symmetries which are *non-local* with respect to any choice of the unit cell provides the interesting new possibility of a realization of the quantization unit  $\frac{1}{2}\bar{\rho}$ .

For given  $Z$  but arbitrary filling factor  $\nu = 1, \dots, N_c Z - 1$ , one can also analyse the conditions if  $n \frac{1}{2} \bar{\rho} \bmod(\frac{1}{2})$  can lead to new rational quantization values compared to  $n' \bar{\rho} \bmod(\frac{1}{2})$ . This is the case if the equation

$$\frac{1}{2} n \bar{\rho} = n' \bar{\rho} - m' \frac{1}{2} \quad (2.46)$$

can not be solved for any integers  $n'$  and  $m'$ . To analyse this we insert  $\bar{\rho} = \frac{\nu}{Z}$  and find that (2.46) is equivalent to

$$n\nu = 2n'\nu - m'Z \quad (2.47)$$

For  $n$  even, this equation is solved by  $n' = \frac{n}{2}$  and  $m' = 0$ . For  $n$  odd and  $Z$  odd, it is solved by  $n' = \frac{1}{2}(Z + n)$  and  $m' = \nu$ . However, for  $n$  odd and  $Z$  even, (2.47) can not be solved for  $\nu$  odd. Since the equation does not change when  $Z$  and  $\nu$  are multiplied with the same integer  $l$ , we conclude that new quantization classes occur for non-local symmetries if

$$Z = 2ql \quad , \quad \nu = (2p - 1)l \quad , \quad (2.48)$$

with two integers  $q$  and  $p$ . Since  $\frac{1}{2}\bar{\rho}$  changes by  $\frac{1}{2}$  if  $\nu$  changes by  $Z$ , we conclude that the new quantization values of the boundary charge due to non-local symmetries are given by

$$Q_B \rightarrow n \frac{2p - 1}{4q} \quad , \quad (2.49)$$

with  $n$  odd,  $q = 1, 2, \dots$ , and  $p = 1, 2, \dots, q$ .

## 4. Application to single-channel and nearest-neighbor hopping models

In this section we discuss a concrete realization of the rational quantization values of the boundary charge in the special case of single-channel and nearest-neighbor hopping models. We present the discussion of discrete lattice models in Section 4.1, where we will concentrate on Weyl semimetal physics occurring at half-filling and discuss the relevance for the IQHE. In Section 4.2 we will develop a low-energy theory in terms of a Dirac model in  $1 + 1$  dimensions and discuss the universal properties of the boundary and interface charge exactly for the noninteracting case and via the bosonization method in the presence of short-ranged electron-electron interactions. Throughout this section we concentrate on the boundary charge  $Q_B^R$  and denote it by  $Q_B$ .

### 4.1 Weyl semimetal physics at half-filling

In Refs. [86, 87] the case of single-channel and nearest-neighbor tight-binding models has been studied analytically for any value of  $Z$  and for generic modulations of the on-site potentials  $v_m = h_m(0)$  and the hoppings  $t_m = -h_m(1)$  (which can all be chosen real and positive  $t_m > 0$ , see Appendix A of Ref. [87]). In these references the universal phase-dependence of the boundary charge  $Q_B(\varphi)$  has been studied for arbitrary  $2\pi$ -periodic functions  $v_m(\varphi)$  and  $t_m(\varphi)$ . The special case

$$v_m(\varphi) = -2V \cos\left(\frac{2\pi}{Z}m + \varphi\right) \quad , \quad t_m(\varphi) = t \quad (2.50)$$

has been considered in Refs. [17, 18], in particular due to the relevance for the IQHE, see the detailed discussion in Ref. [18]. In this case,  $\varphi$  can be interpreted as the transverse quasimomentum in a 2D quantum Hall setup [18] and  $Z$  corresponds to the magnetic length. Whereas Ref. [18] has discussed small filling factors  $\nu$  with finite Chern number, the particular interest in this work is the case of half-filling,  $\nu = \frac{Z}{2}$  and  $\bar{\rho} = \frac{1}{2}$ , where  $Z$  is even to open a gap. Due to (2.43), the following two universal quantization classes are possible for  $Q_B \equiv Q_B^R$  at half-filling in case certain symmetries are fulfilled

$$Q_B = \frac{1}{2} \bmod\left(\frac{1}{2}\right) \quad \text{or} \quad Q_B = \frac{1}{4} \bmod\left(\frac{1}{2}\right) \quad . \quad (2.51)$$

The first  $\frac{1}{2}$ -class is the usual known one which occurs also in the presence of local inversion or local chiral symmetries. The second  $\frac{1}{4}$ -class is a novel one which requires essentially non-local symmetries. For half-filling, we will explain in the following that it is possible that the quantization of  $Q_B(\varphi)$  persists for all phases  $\varphi$  and Weyl semimetal physics occurs with  $Q_B$  jumping by  $\pm\frac{1}{2}$  at gap closing points. This is of relevance for the IQHE. Whereas the Chern number and, therefore, the Hall current vanishes, the boundary charge shows an interesting quantization feature.

In Appendix 5.4 the symmetry conditions (2.39) and (2.40) are explicitly evaluated for the special case of single-channel and nearest-neighbor hopping models. For the symmetry  $\Pi_n = T_n U \Pi$  one obtains  $U_m = 1$  and the condition

$$v_m = v_{Z-m-n+1} \quad , \quad t_m = t_{Z-m-n} \quad , \quad (2.52)$$

whereas for  $S_n = T_{-n} U$  one finds  $U_m = (-1)^m$  together with

$$v_m = -v_{m-n} \quad , \quad t_m = t_{m-n} \quad . \quad (2.53)$$

The non-local chiral symmetry  $S_n$  has the advantage that it can be fulfilled for all phases  $\varphi$ , whereas  $\Pi_n$  leads to a rational quantization of the boundary charge only at certain values of  $\varphi$ . Therefore, we concentrate in the following on  $S_n$ , where an interesting application relevant for the IQHE at half-filling can be formulated.

Applying the symmetry condition (2.53) twice one finds  $v_m = v_{m-2n}$  implying that  $Z = 2n$  defines the wavelength of the modulation which is identical to the number of sites of the unit cell (the hopping has the wavelength  $Z/2$ ). Therefore, the translation  $T_{-n} = T_{Z/2}$  shifts the lattice by half of the unit cell length, typical for non-symmorphic symmetries. For  $n = \frac{Z}{2}$  and  $\bar{\rho} = \frac{1}{2}$ , we get from (2.43) the quantization values

$$Q_B = \frac{Z}{8} \mod\left(\frac{1}{2}\right) \quad , \quad (2.54)$$

leading to the  $\frac{1}{2}$ -class for  $Z = 4, 8, 12, \dots$  and to the novel  $\frac{1}{4}$ -class for  $Z = 2, 6, 10, \dots$

For  $n = \frac{Z}{2}$ , a concrete realization of (2.53) for all phases is given by the pure potential modulation model (2.50) with constant hopping. Other more complicated realizations are also possible but do not lead to qualitative differences. This model has the advantage that a phase shift of  $\varphi$  by  $\frac{2\pi}{Z}$  shifts the lattice by one site towards the boundary, i.e.,  $Q_B$  changes by  $\bar{\rho} \mod(1)$  according to (2.26). This must be a half-integer for  $\bar{\rho} = \frac{1}{2}$ . Furthermore, since  $Q_B(\varphi)$  is quantized for all  $\varphi$  and since edge states crossing the chemical potential during the shift can change  $Q_B$  only by an integer value, we conclude that there must be necessarily a gap closing point in any phase interval of size  $\frac{2\pi}{Z}$ . Between the gap closing points  $Q_B$  is quantized due to the symmetry and edge modes connecting the gap closing points play the role of Dirac arcs, see Figs. 10(a1,a2). Therefore, we call this the *Weyl case*. At a gap closing point  $Q_B$  jumps by  $\pm\frac{1}{2}$  such that (2.26) is fulfilled, see Figs. 10(b1,b2). This is also demonstrated in Figs. 10(c1,c2), where we show the integer invariant  $I(\varphi)$ , defined by

$$I(\varphi) = \Delta Q_B(\varphi) - \frac{\nu}{Z} \in \{-1, 0\} \quad , \quad (2.55)$$

$$\Delta Q_B(\varphi) = Q_B\left(\varphi + \frac{2\pi}{Z}\right) - Q_B(\varphi) \quad . \quad (2.56)$$

According to Refs. [86, 87] this invariant fulfils the topological constraint  $I \in \{-1, 0\}$  due to charge conservation of particles and holes. We note that this property is not changed when a gap closing point appears during the shift of the lattice by one site. The phase-dependence of

the model parameters can always be chosen such that no gap closing appears during the shift without changing the parameters before and after the shift, see Ref. [87].

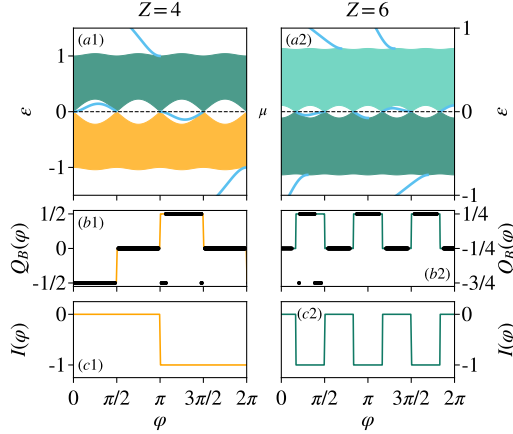


Figure 10: Phase-dependence of (a) band structure, (b) boundary charge  $Q_B$ , and (c) invariant  $I(\varphi) = Q_B(\varphi + \frac{2\pi}{Z}) - Q_B(\varphi) - \bar{\rho}$  for the model (2.50) with  $V = 0.5$  and  $t = 1$  at half-filling  $\bar{\rho} = \frac{1}{2}$ , and  $Z = 4, 6$  (left/right panel), calculated for a semi-infinite system. In (a) the bands  $\alpha = \frac{Z}{2}, \frac{Z}{2} + 1$  are shown together with edge modes connecting the gap closing points, where  $Q_B$  jumps by  $\pm \frac{1}{2}$ . The invariant is quantized to  $I \in \{-1, 0\}$ . Due to the nonlocal chiral symmetry  $v_m(\varphi) = -v_{m \pm \frac{Z}{2}}(\varphi)$ , quantization of  $Q_B$  occurs in units of  $\frac{1}{2}$  for  $Z = 4$ , whereas for  $Z = 6$  we get  $Q_B = \frac{1}{4} \bmod(1/2)$ . The black symbols (mainly overlaying the lines) in (b) show the case with random disorder for a finite system for additional staggered onsite-disorder drawn from a uniform distribution  $(-0.025, 0.025]$  for a finite system of  $Z \cdot 10^5$  lattice sites.

One can also generalize the universal form of  $Q_B(\varphi)$ , derived in Refs. [17, 18, 86, 87], to the case of gap closings. When the gap is non-zero for all phases, the form is given by

$$Q_B(\varphi) = f(\varphi) + \frac{M_- - M_+}{2\pi} \varphi + F(\varphi), \quad (2.57)$$

where  $f(\varphi)$  is a nonuniversal smooth  $\frac{2\pi}{Z}$ -periodic function, and

$$F(\varphi) = \sum_{\sigma=\pm} \sum_{i=1}^{M_\sigma} \sigma \Theta(\varphi - \varphi_i^\sigma) \quad (2.58)$$

describes the discrete jumps of  $Q_B$  by  $\pm 1$  when edge states cross the chemical potential at  $\varphi_i^\pm$  from above or below, respectively.  $M_\pm$  denotes the total number of edge states moving below/above the chemical potential when the phase changes by  $2\pi$ . The average slope from the linear term determines the Chern number  $C_\nu = M_- - M_+$  [51, 54, 55, 56, 87]. Moreover, Eqs. (2.55) and (2.57) imply the relation

$$C_\nu = \nu - s_\nu Z, \quad (2.59)$$

where  $s_\nu = \Delta F(\varphi) - I(\varphi)$  is a phase-independent integer characteristic for each gap. This relation is equivalent to the Diophantine equation [54, 55, 56], a central relation for the bulk-boundary correspondence of the IQHE.

The universal form (2.57) remains valid in the presence of gap closing points for the Weyl case, with the only difference that the jumps of  $Q_B$  have size  $\pm \frac{1}{2}$  at a gap closing since the charge of the edge state is distributed symmetrically among the two bands [87]. Thus, Eq. (2.57) remains the same, we only have to add a factor  $\frac{1}{2}$  in Eq. (2.58) for the terms in the sum

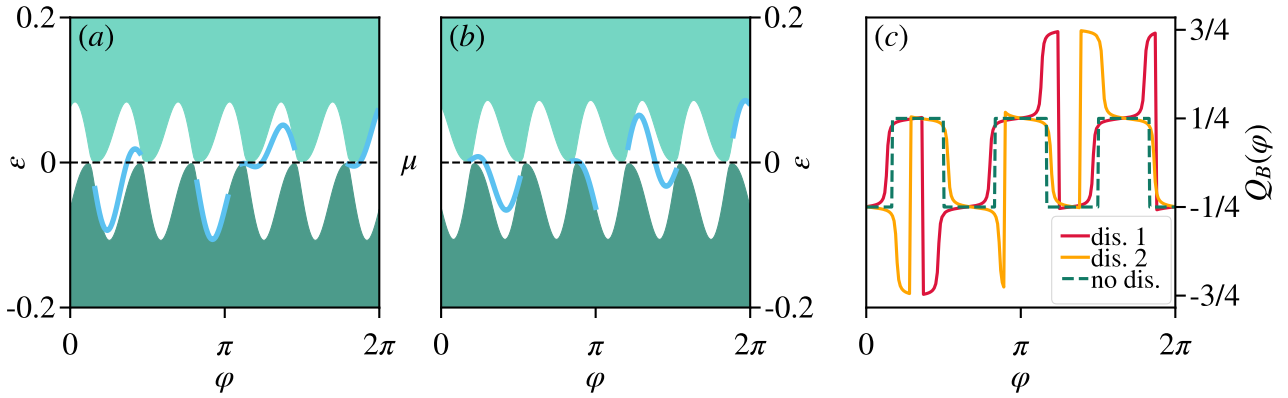


Figure 11: (a) and (b) show the phase-dependence of the band structure and the edge states for  $Z = 6$  and the same parameters as in Fig. 10 but with additional periodic disorder for the on-site potentials on the scale 0.2 (which is chosen rather strong to visualize the gap openings), calculated for a semi-infinite system. As a result, 3 edge states run either from the valence to the conduction band or vice versa. In (c) the phase-dependence of the boundary charge  $Q_B(\varphi)$  is shown for the two disordered configurations and compared to the clean case, demonstrating stability of the quantization of  $Q_B$  between the jumps.

corresponding to the jumps at gap closings and, correspondingly, count only the contribution  $\pm \frac{1}{2}$  to  $M_{\pm}$ . As a consequence, the Diophantine equation (2.59) remains also valid in the Weyl case, but  $s_{\nu}$  becomes half-integer. E.g., for Fig. 10, we have  $s_{\nu} = \frac{1}{2}$  which gives with  $\nu = \frac{Z}{2}$  a vanishing Chern number  $C_{\nu} = 0$ .

As shown in Ref. [18] the Hall current  $I_x$  for a 2D quantum Hall system (with periodic boundary conditions in  $y$ -direction, described by the azimuthal direction of a cylinder topology) along the direction  $x$  of the effective 1D system in response to a perpendicularly applied voltage bias  $V_y$  is related to  $Q_B(\varphi)$  by

$$I_x = -\frac{d}{dt} Q_{\text{edge}}^{(2D)}(t) = -\frac{d}{dt} \sum_{n=1}^M Q_B \left( \frac{2\pi}{M} \left( n + \frac{\Phi(t)}{\Phi_0} \right) \right), \quad (2.60)$$

where  $Q_{\text{edge}}^{(2D)}$  is the charge along the edge of the physical 2D system,  $\Phi_0 = h/e$  denotes the flux quantum, and  $V_y = -\frac{d}{dt} \Phi(t)$  is generated by a time-dependent magnetic flux  $\Phi(t)$  applied through the cylinder. The discrete values  $k_y = \frac{2\pi}{M}n$  describe the perpendicular quasimomentum along the azimuthal direction of the cylinder (with  $M$  lattice sites around the cylinder). Inserting (2.57) one finds that the  $\frac{2\pi}{Z}$ -periodic and smooth part  $f(\varphi)$  and the piecewise constant function  $F(\varphi)$  provide a  $\Phi$ -independent contribution to the sum (up to discrete jumps at particular values of  $\Phi$  from  $F(\varphi)$ ). Therefore, the Hall current probes only the linear term of (2.57), leading to a quantized Hall conductance  $\sigma_{xy}$  in terms of the Chern number

$$I_x = \sigma_{xy} V_y, \quad \sigma_{xy} = \frac{e}{h} C_{\nu}. \quad (2.61)$$

Therefore, for the Weyl case discussed above, the Hall conductance vanishes.

A delicate question concerns the unknown function  $f(\varphi)$  in (2.57). When  $Q_B(\varphi)$  is quantized for all  $\varphi$  we get  $M_+ = M_-$  and  $f(\varphi)$  is a constant determining the quantization value. However, when a small symmetry-breaking term in the form of periodic disorder is added to the on-site potentials, the gap will open at the Weyl points, and one obtains a discontinuous jump to a finite Chern number  $C_{\nu} = M_- - M_+$ , see Figs. 11(a,b) for  $Z = 6$ . The gap opens slightly and  $\frac{Z}{2}$  edge states move either from the valence to the conduction band or vice versa, giving rise to two different Chern numbers  $\pm \frac{Z}{2}$ , with a corresponding jump of the Hall current. In addition to

the linear term  $\frac{C_\nu}{2\pi}\varphi$ , also the functions  $f(\varphi)$  and  $F(\varphi)$  jump discontinuously such that all three terms on the right hand side of Eq. (2.57) are unstable against small symmetry-breaking terms for *all* phases. However, the boundary charge determined by the sum of all three terms remains a stable quantity between the jumps as shown in Fig. 11(c). This shows that the quantization values of  $Q_B(\varphi)$  at fixed  $\varphi$  between the gap closing points are well-defined and stable quantities accessible by experiments for 1D systems.

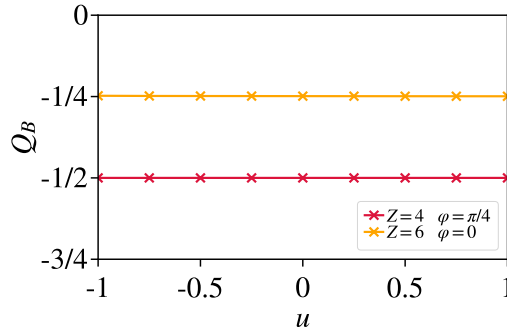


Figure 12: Stability of the rational quantization of the boundary charge upon inclusion of nearest neighbor interaction. The figure shows the boundary charge  $Q_B$  for the same parameters (but finite size  $N = 1000$ ) as used in Fig. 10 at two values of  $\varphi = \pi/4$  and  $\varphi = 0$ , where the gap is maximal, for  $Z = 4$  and  $Z = 6$ , respectively.

In Fig. 10 we add an analysis of the effects of non-periodic disorder in the panels showing  $Q_B(\varphi)$ . We overlay the results obtained without disorder (lines) with those calculated in the presence of a quenched onsite disorder (symbols) of quite moderate strength  $d/t = 0.05$ . The two results mainly overlap, besides of at values of  $\varphi$  close to gap closings where small energy shifts and gap openings induced by the disorder might change the value of quantization mod(1), similar to the effects of weak periodic disorder as shown in Fig. 11(c) (where the disorder strength is chosen much stronger to visualize the effects on the spectrum in Figs. 11(a,b)). Therefore, the overall quantization of  $1/2$  and  $1/4$  for  $Z = 4$  and  $Z = 6$ , respectively, remains perfectly intact mod(1).

Finally, we study the robustness of the quantization with respect to adding short-ranged electron-electron interaction at half-filling ( $\mu = 0$  in the non-interacting case above). We add the particle-hole symmetric version of Eq. (2.14) as the interaction

$$V_{ee} = u \sum_m \left( \hat{\rho}_m - \frac{1}{2} \right) \left( \hat{\rho}_{m+1} - \frac{1}{2} \right) \quad (2.62)$$

to the Hamiltonian and study the resulting interacting quantum many-body system using DMRG. This particle-hole symmetric formulation is chosen for convenience of implementation only. The results for  $\varphi$  being at the maximal single particle gap are summarized in Fig. 12 for  $Z = 4$  and  $Z = 6$  and the other parameters as in Fig. 10. As shown the quantization of  $Q_B$  remains perfectly intact upon including this local interaction up to rather large values of the interaction strength  $u$ .

## 4.2 Low-energy theory

In this section we discuss the low-energy theory in terms of a Dirac model in 1+1 dimensions, following closely the treatment in Refs. [39, 17, 18], where low-energy models have been derived from lattice models for the special case of potential modulation with constant hopping. Here, we discuss the general derivation of Dirac models and present an exact formula for the complex



gap parameter entering the low-energy model. We discuss in detail the restrictions for the phase of the gap parameter in the presence of non-local symmetries or for the special case of half-filling. We present exact formulas for the boundary and interface charge for the noninteracting Dirac model and prove the consistency with our general framework. Furthermore, we discuss the stability under short-ranged electron-electron interaction by using the bosonization method. Whereas the gap is significantly renormalized by interactions [112, 113, 114, 39], it turns out that the boundary charge is insensitive to the gap size but depends only the phase of the gap parameter in a linear fashion which is not influenced by interactions. Since the low-energy model is most conveniently written in continuum space, we explicitly write the lattice spacing  $a$  at the appropriate places and do not set it to one in this section.

### Noninteracting Dirac model

We first split the noninteracting single-channel nearest-neighbor hopping model with on-site potentials  $v_m$  and hoppings  $-t_m$  into two parts by writing  $t_m = t + \delta t_m$

$$H_{\text{bulk}} = H_0 + H', \quad (2.63)$$

$$H_0 = -t \sum_m c_{m+1}^\dagger c_m + h.c., \quad (2.64)$$

$$H' = \sum_m v_m c_m^\dagger c_m - \sum_m \delta t_m (c_{m+1}^\dagger c_m + h.c.). \quad (2.65)$$

$H'$  describes the external modulation which, just for illustration (other cases can be treated analogously), we take to be harmonic with wavelength  $\lambda_{\text{ex}} = Za$

$$v_m = -2\delta v \cos(k_{\text{ex}}ma + \varphi_v), \quad (2.66)$$

$$\delta t_m = -\delta t \cos(k_{\text{ex}}ma + \varphi_t), \quad (2.67)$$

where  $k_{\text{ex}} = \frac{2\pi}{\lambda_{\text{ex}}}$  is the wavevector of the external modulation.  $H_0$  can be easily diagonalized in terms of plane waves  $|k\rangle = \sqrt{\frac{a}{2\pi}} \sum_m e^{ikma} |m\rangle$ , leading to

$$H_0 = \int_{-\pi/a}^{\pi/a} dk \epsilon_k^{(0)} \tilde{c}_k^\dagger \tilde{c}_k, \quad (2.68)$$

with

$$\epsilon_k^{(0)} = -2t \cos(ka) \quad , \quad -\frac{\pi}{a} < k < \frac{\pi}{a}, \quad (2.69)$$

$$\tilde{c}_k^\dagger = \sqrt{\frac{a}{2\pi}} \sum_m e^{ikma} c_m^\dagger. \quad (2.70)$$

In the representation  $|k\rangle$  of the exact single-particle eigenstates of  $H_0$ , the matrix elements of  $H'$  can be straightforwardly calculated and one obtains

$$\begin{aligned} \langle k | H' | k' \rangle &= \delta(k - k' - k_{\text{ex}}) \left\{ -\delta v e^{i\varphi_v} + \frac{\delta t}{2} (e^{-ika} + e^{ik'a}) e^{i\varphi_t} \right\} \\ &= \delta(k - k' + k_{\text{ex}}) \left\{ -\delta v e^{-i\varphi_v} + \frac{\delta t}{2} (e^{-ika} + e^{ik'a}) e^{-i\varphi_t} \right\}. \end{aligned} \quad (2.71)$$

These matrix elements lead to  $Z - 1$  gap openings labeled by  $\nu = 1, 2, \dots, Z - 1$  at wavevectors  $\pm k_F^{(\nu)}$  with  $2k_F^{(\nu)} = \nu k_{\text{ex}}$ . The gaps are generated in  $\nu'$ -th order perturbation theory in  $H'$  with  $\nu' = \min\{\nu, Z - \nu\}$  [115] and are of the order

$$|\Delta_\nu| \sim t \left( \frac{\max\{|\delta v|, |\delta t|\}}{t} \right)^{\nu'}. \quad (2.72)$$

In the following we will concentrate on a certain gap with index  $\nu$  and write  $k_F \equiv k_F^{(\nu)}$  and  $\Delta \equiv \Delta_\nu$  for brevity. Our aim is to develop an effective low-energy model for energies close to the Fermi energy  $\epsilon_F = \epsilon_{k_F}^{(0)} = -2t \cos(k_F a)$ . Using Brillouin-Wigner perturbation theory the coupling between the states close to the two Fermi points  $\pm k_F$  can be described by an effective Hamiltonian

$$H'_{\text{eff}} = P \left( H' + H' Q \frac{1}{\epsilon_F - Q H_{\text{bulk}} Q} Q H' \right) P, \quad (2.73)$$

where  $P$  projects on the low-energy sector and  $Q = 1 - P$ . It is then straightforward to see that for  $|k|, |k'| \ll k_F$  one obtains a coupling between the two Fermi points via  $\nu' - 1$  virtual intermediate states described by the matrix element

$$\langle k_F + k | H'_{\text{eff}} | -k_F + k' \rangle = \Delta \delta(k - k'), \quad (2.74)$$

where

$$\Delta = |\Delta| e^{i\alpha} \quad (2.75)$$

is a complex gap parameter with negligible  $k$ -dependence. Using Eq. (2.71) one finds after a straightforward calculation

$$\Delta \equiv \Delta_\nu = \begin{cases} \Delta_\nu^- & \text{for } \nu < \frac{Z}{2} \\ \Delta_\nu^+ & \text{for } \nu > \frac{Z}{2} \\ \Delta_\nu^- + \Delta_\nu^+ & \text{for } \nu = \frac{Z}{2} \end{cases}, \quad (2.76)$$

with

$$\Delta_\nu^+ = -(-1)^{Z-\nu} \Delta_{Z-\nu}^- \Big|_{\substack{\varphi_v \rightarrow -\varphi_v \\ \varphi_t \rightarrow -\varphi_t + k_{\text{ex}} + \pi}}, \quad (2.77)$$

and

$$\Delta_\nu^- = \prod_{s=1}^{\nu-1} \frac{1}{\epsilon_F - \epsilon_{-k_F + s k_{\text{ex}}}^{(0)}} \prod_{l=1}^{\nu} \left\{ -\delta v e^{i\varphi_v} + \frac{\delta t}{2} \left[ e^{i(k_F - l k_{\text{ex}})a} + e^{-i(k_F - l k_{\text{ex}})a} e^{-i k_{\text{ex}} a} \right] e^{i\varphi_t} \right\}. \quad (2.78)$$

Most importantly, the gap parameter  $\Delta_{Z/2}$  at half-filling (only possible for  $Z$  even) is determined by an interference of two processes. This will become important for the quantization values of the boundary charge (see below).

Splitting the field operator  $\psi(ma) \equiv \frac{1}{\sqrt{a}} c_m$  in slowly varying right and left moving fields  $R(x)$  and  $L(x)$  via

$$\psi(x) = R(x) e^{i k_F x} + L(x) e^{-i k_F x}, \quad (2.79)$$

and inserting this decomposition in the effective Hamiltonian  $H_{\text{eff}} = H_0 + H'_{\text{eff}}$ , one finds after neglecting strongly oscillating terms the final result for the low-energy Hamiltonian in the form of a Dirac Hamiltonian in  $1+1$  dimensions

$$H_{\text{eff}} = \int dk \underline{\psi}_k^\dagger \{ v_F k \sigma_z + |\Delta| \cos \alpha \sigma_x - |\Delta| \sin \alpha \sigma_y \} \underline{\psi}_k, \quad (2.80)$$



where  $v_F = 2ta \sin(k_F a)$  and

$$\underline{\psi}_k = \frac{1}{\sqrt{2\pi}} \int dx e^{-ikx} \underline{\psi}(x), \quad \underline{\psi}(x) = \begin{pmatrix} R(x) \\ L(x) \end{pmatrix}. \quad (2.81)$$

For equal phases  $\varphi_v = \varphi_t = \varphi$  we can directly see from (2.77) and (2.78), that the phase  $\alpha$  of the gap parameter is a linear function of  $\varphi$  for  $\nu \neq \frac{Z}{2}$

$$\alpha = \pm \nu \varphi + \text{const} \quad \text{for} \quad \nu \lesseqgtr \frac{Z}{2}, \quad (2.82)$$

where the constant term is non-universal but independent of  $\varphi$ . As we will see in Section 4.2 this leads to a universal linear behaviour of the boundary charge as function of  $\varphi$ .

For the case of half-filling  $\nu = \frac{Z}{2}$  (only possible for  $Z$  even) it is more complicated due to the interference effect from two paths. If we take potential modulation only, i.e.,  $\delta t = 0$ , as considered in Section 4.1, we get from (2.77) and (2.78)

$$\Delta_{Z/2}^- = |\Delta_{Z/2}^-| (-1)^{Z/2} e^{i\frac{Z}{2}\varphi}, \quad (2.83)$$

$$\Delta_{Z/2}^+ = -|\Delta_{Z/2}^-| e^{-i\frac{Z}{2}\varphi}, \quad (2.84)$$

which gives for the sum

$$\Delta_{Z/2} = |\Delta_{Z/2}^-| \begin{cases} 2i \sin(\frac{Z}{2}\varphi) & \text{for } \frac{Z}{2} \text{ even} \\ -2 \cos(\frac{Z}{2}\varphi) & \text{for } \frac{Z}{2} \text{ odd} \end{cases}. \quad (2.85)$$

Therefore, due to the interference of the two paths, one obtains for the phase  $\alpha$  at half-filling an interesting phase-locking effect

$$\alpha = \begin{cases} \frac{\pi}{2} + \pi \Theta[-\sin(\frac{Z}{2}\varphi)] & \text{for } \frac{Z}{2} \text{ even} \\ \pi \Theta[\cos(\frac{Z}{2}\varphi)] & \text{for } \frac{Z}{2} \text{ odd} \end{cases}, \quad (2.86)$$

which, as we will see in Section 4.2, explains the pinning of the boundary charge to certain quantization values.

Finally, we analyse the restrictions for the values of  $\alpha$  in the presence of the non-local symmetries  $\Pi_n$  or  $S_n$  discussed in Section 3.2. This can be done directly by using the general definition of  $\Delta$  via (2.74) without using some special form for  $v_m$  and  $t_m$ . As shown in Appendix 5.4 the two symmetries act on the lattice sites as

$$\Pi_n |m\rangle = |Z - m - n + 1\rangle, \quad (2.87)$$

$$S_n |m\rangle = (-1)^m |m - n\rangle, \quad (2.88)$$

which implies the following transformation in quasimomentum space

$$\Pi_n |k\rangle = e^{ik(Z-n+1)a} |-k\rangle, \quad (2.89)$$

$$S_n |k\rangle = e^{ik(n+\pi)a} |k + \pi\rangle. \quad (2.90)$$

Together with (2.74) this implies for the unitary symmetry  $\Pi_n$  and the chiral symmetry  $S_n$  the following condition for the gap parameter

$$\Pi_n : \Delta = e^{-2ik_F(Z-n+1)a} \Delta^*, \quad (2.91)$$

$$S_n : \Delta = -e^{-2ik_F n a} \Delta^*. \quad (2.92)$$

Using  $2k_F a = \nu k_{\text{ex}} a = 2\pi\bar{\rho}$  and, for the symmetry  $S_n$  (which requires half-filling),  $k_F a = \frac{\pi}{2}$ , we get for both symmetries the following pinning of  $\alpha$

$$\alpha = (n-1)\pi\bar{\rho} \pmod{\pi}. \quad (2.93)$$

As shown in Section 4.2, this will prove consistency of the low-energy approach with the quantization of the boundary charge according to our general framework.

### Interacting Dirac model

Let us now proceed to include electron-electron interaction effects within the short-ranged density-density interaction form (2.14). Using standard bosonization techniques [116, 117, 118] we find within the low-energy model

$$H = H_{eff} + V_{ee} = \frac{v}{2} \int dx \left\{ K [\hat{\Pi}(x)]^2 + \frac{1}{K} [\partial_x \hat{\varphi}(x)]^2 \right\} + (-1)^p \frac{|\Delta|}{2\pi a} \int dx \cos(\sqrt{4\pi} \hat{\varphi}(x) - \alpha - \pi/2), \quad (2.94)$$

where the Luttinger liquid parameter  $K$  and the renormalized Fermi velocity  $v$  are defined by

$$K = \left( 1 + \frac{2u_1 - u_2}{\pi v_F} \right)^{-1/2}, \quad v = v_F / K, \quad (2.95)$$

with  $u_1 = a \sum_m u(m)$  and  $u_2 = a \sum_m u(m) e^{2ik_F m a}$  corresponding to forward and backward scattering processes, respectively. Here,  $u(m - m')$  describes the short-ranged Coulomb interaction between the densities at site  $m$  and  $m'$ , see Eq. (2.14).  $\hat{\Pi}(x) = \partial_x [\hat{\varphi}_+(x) - \hat{\varphi}_-(x)]$  and  $\hat{\varphi}(x) = \hat{\varphi}_+(x) + \hat{\varphi}_-(x)$  are canonically conjugate momentum and field variables. The chiral boson fields  $\hat{\varphi}_\pm(x)$  are related to the right and left movers  $\hat{\psi}_+(x) \equiv R(x)$  and  $\hat{\psi}_-(x) \equiv L(x)$  via

$$\hat{\psi}_p(x) = \frac{1}{\sqrt{4\pi a}} e^{ip\sqrt{4\pi}\hat{\varphi}_p(x)}, \quad (2.96)$$

where  $\frac{1}{a}$  is used for the momentum cutoff (or  $v/a$  for the high-energy cutoff). A subtlety is the phase shift by  $\frac{\pi}{2}$  in the cosine term of (2.94) and the undetermined prefactor  $(-1)^p$ . This is related to the commutator  $[\hat{\varphi}_+(x), \hat{\varphi}_-(x')] = \frac{i}{4}(2p + 1)$  which arises from the zero mode phases to ensure the anticommutation relation of left and right movers. Here, the value  $p$  is an arbitrary integer which will be finally determined by comparing the boundary charge with the exact solution for the noninteracting Dirac model.

In order to get an insight into the physics of the interacting system, it is instructive to perform a perturbative renormalization group analysis [39] using standard operator product expansion techniques [117]. Reducing the high-energy cutoff  $\Lambda$  (with initial value  $\Lambda_0 = v/a$ ) we obtain the following flow equations for the gap  $|\Delta|$  and the Luttinger parameter  $K$ :

$$\frac{d|\Delta|}{d\ell} = (1 - K)|\Delta|, \quad d\ell = -\frac{d\Lambda}{\Lambda}, \quad (2.97)$$

$$\frac{dK}{d\ell} = -c \frac{a_\Lambda^2 K^2 |\Delta|^2}{2\pi v^2}, \quad (2.98)$$

with  $a_\Lambda = v/\Lambda$  and some unimportant constant  $c \sim O(1)$  which depends on the RG procedure. As one can see, for repulsive interactions ( $K < 1$ ) the cosine term in (2.94) is a relevant perturbation and the system flows into the gapped phase while the interaction grows under the RG. This fact allows one to conclude that the fluctuations of the cosine term are getting effectively frozen in the renormalized theory and is of crucial importance in the determination of the boundary charge in the interacting theory in the following. Although it seems that the gap grows to infinity under the RG flow (with  $K$  shrinking to zero), we note that the flow equations can only be trusted until the cutoff  $\Lambda$  reaches a critical scale  $\Lambda_c \sim |\Delta|_c$  (or  $a_c \sim v/|\Delta|_c$ ), with  $|\Delta|_c = |\Delta|_{\Lambda=\Lambda_c}$ , at which the flow has to be truncated.

At half-filling  $k_F a = \frac{\pi}{2}$ , in principle, one also has to consider the Umklapp scattering process in the bosonized Hamiltonian. However, the Umklapp term is RG relevant only for strong two-particle interactions with  $K < 1/2$  [118]. Bearing this in mind, we are not going to focus on the Umklapp process in the following by confining ourselves to moderate electron-electron interactions.

### Boundary and interface charge (!!!!!!!!!!!!!!!)

**Boundary charge  $Q_B$  of the noninteracting Dirac model on the  $x > 0$**  Let us turn to the discussion on the boundary and interface charge quantization in the low-energy description. We start with the boundary charge  $Q_B$  of the noninteracting Dirac model on the half-line  $x > 0$  with vanishing boundary condition  $\psi(0) = R(0) + L(0) = 0$ , given by

$$H_{\text{eff}} = \int_0^\infty dx \underline{\psi}^\dagger(x) \left\{ v_F(-i\partial_x)\sigma_z + |\Delta| \cos \alpha \sigma_x - |\Delta| \sin \alpha \sigma_y \right\} \underline{\psi}(x). \quad (2.99)$$

In order to study the quantization of  $Q_B$ , one has to construct the single-particle eigenstates of the Hamiltonian. As shown in Appendix 5.5 there are two distinct types of admissible eigenstates. First of all, there are two scattering states (labeled by the nonnegative momentum  $k$ ) living at energies  $\epsilon_{k,\pm} = \pm \sqrt{v_F^2 k^2 + |\Delta|^2}$ , corresponding to conduction and valence bands, respectively. In addition, under the condition  $\sin \alpha > 0$ , inside the gap, there exists a single edge state sitting at an energy  $-|\Delta| \cos \alpha$  [39]. Assuming that the chemical potential of the system lies at the bottom of the conduction band, we show in Appendix 5.5 that the respective contributions of the edge and scattering states are

$$Q_B^{\text{edge}}(\alpha) = \Theta_{0 < \alpha < \pi} \equiv \begin{cases} 1 & \text{for } 0 < \alpha < \pi \\ 0 & \text{otherwise} \end{cases}, \quad (2.100)$$

$$Q_B^{\text{scatt}}(\alpha) = \frac{\ln(-e^{i\alpha})}{2\pi i} - \frac{1}{4}. \quad (2.101)$$

Combining both contributions, one immediately arrives at the following universal result

$$Q_B(\alpha) = \frac{\alpha}{2\pi} + \frac{1}{4} \quad (2.102)$$

for  $-\pi < \alpha < \pi$ , and periodic continuation to the other intervals. For arbitrary position of the chemical potential in the gap one has to use the result  $Q_B^{\text{edge}}(\alpha) = \theta(\mu + |\Delta| \cos \alpha) \Theta_{0 < \alpha < \pi}$  for the edge state charge, i.e., one has to add the integer term  $-\theta(-\mu - |\Delta| \cos \alpha) \Theta_{0 < \alpha < \pi}$  to (2.102). This proves Eq. (2.6) stated in the introduction.

Let us emphasize that the low energy result accurately reproduces the conclusions based on the microscopic theory. For instance, consider the case  $\nu \geq \frac{Z}{2}$ ,  $\varphi_v = \varphi_t = \varphi$ , so that according to (2.82),  $\alpha = \pm \nu \varphi + \text{const}$ , leading to the universal linear behaviour of  $Q_B$  as a function of  $\varphi$  [17, 18, 86, 87]. At half-filling  $\nu = \frac{Z}{2}$ , where  $\alpha$  is pinned to the values given by (2.86), one arrives at the following quantization of  $Q_B$

$$Q_B(\varphi) = \begin{cases} \frac{1}{2} + \frac{1}{2} \Theta(-\sin(\frac{Z}{2}\varphi)), & \text{for } \frac{Z}{2} \text{ even} \\ \frac{1}{4} + \frac{1}{2} \Theta(\cos(\frac{Z}{2}\varphi)), & \text{for } \frac{Z}{2} \text{ odd} \end{cases}, \quad (2.103)$$

again showing complete agreement with the microscopic prediction (2.54). Finally, in the presence of the symmetries  $\Pi_n$  or  $S_n$ , where  $\alpha$  is pinned to  $\alpha = (n-1)\pi\bar{\rho} \bmod(\pi)$  according to (2.93), one finds

$$Q_B = n\frac{1}{2}\bar{\rho} - \frac{1}{2}\bar{\rho} + \frac{1}{4} \bmod\left(\frac{1}{2}\right). \quad (2.104)$$

Comparing with the exact solution (2.43) there is a difference given by the constant  $-\frac{1}{2}\bar{\rho} + \frac{1}{4}$  which vanishes only at half-filling. This difference can be traced back to the fact that the Dirac model contains an infinite set of high-energy states which are unphysical. Interestingly, this constant can be shown to be given by the negative boundary charge of the original lattice model  $H_0$  on a half line at zero gap, see Eq. (2.181) in Appendix 5.6. Thus, we get the following relation between the boundary charge of the Dirac model and the exact one

$$Q_B^{\text{Dirac}} = Q_B^{\text{exact}} - Q_B^{\text{exact}}|_{\Delta=0}. \quad (2.105)$$

**Charge quantization for the noninteracting Dirac model.** Next we study the interface charge quantization for the noninteracting Dirac model. Now the diagonalization problem is formulated on the entire real line, however,  $\alpha$  is now allowed to be a function of position  $\alpha = \alpha(x)$  with Hamiltonian

$$H_{\text{eff}} = \int dx \underline{\psi}^\dagger(x) \left\{ v_F(-i\partial_x)\sigma_z + |\Delta| \cos[\alpha(x)]\sigma_x - |\Delta| \sin[\alpha(x)]\sigma_y \right\} \underline{\psi}(x). \quad (2.106)$$

In particular, we make the following choice  $\alpha(x) = \Theta(x)\alpha_R + \Theta(-x)\alpha_L$  and define  $\delta\alpha = \alpha_R - \alpha_L$ . As shown in Appendix 5.7, one concludes that there are two different types of states to consider, the scattering states, as well as the in-gap states localized at the interface. The bound state is present for  $\sin(\delta\alpha/2) > 0$  with energy  $-|\Delta| \cos(\delta\alpha/2)$  and contributes unity to the total interface charge (if occupied). As opposed to the semi-infinite problem, the valence and conduction ( $\epsilon_{k,\pm} = \pm\sqrt{v_F^2 k^2 + |\Delta|^2}$ ) band states are now two-fold degenerate. Indeed, for a given energy, one always has two distinct scattering channels, the one where particles scatter from left to right, and the opposite one, where particles scatter from right to left, and hence the degeneracy. Assuming that the chemical potential is located at the bottom of the conduction band, we show in Appendix 5.7 that the interface charge follows the Goldstone-Wilczek formula

$$Q_I = \frac{\delta\alpha}{2\pi} \quad (2.107)$$

for  $\delta\alpha \in (0, 2\pi)$ . Values of  $Q_I$  on other intervals are to be found from its periodic dependence on  $\delta\alpha$ . Similiar to the boundary charge one has to add the integer term  $-\theta(-\mu - |\Delta| \cos(\delta\alpha/2))\Theta_{0 < \delta\alpha/2 < \pi}$  for a chemical potential with arbitrary position in the gap.

Let us now proceed by studying the effects of the electron-electron interaction on the quantization of the boundary charge. We take the bosonized Hamiltonian (2.94) on the semi-infinite part  $x > 0$ , together with the boundary condition

$$0 = \psi(0) = \frac{1}{\sqrt{4\pi a}} \left( e^{i\sqrt{4\pi}\hat{\varphi}_+(0)} + e^{-i\sqrt{4\pi}\hat{\varphi}_-(0)} \right). \quad (2.108)$$

This requires  $i\sqrt{4\pi}\hat{\varphi}_-(0) = i\pi(2q-1) - i\sqrt{4\pi}\hat{\varphi}_+(0)$  with some integer  $q$ , leading to the following boundary condition for  $\hat{\varphi}(0) = \hat{\varphi}_+(0) + \hat{\varphi}_-(0)$  (see also Refs. [119, 39])

$$\hat{\varphi}(0) = \frac{1}{2\sqrt{\pi}}(2q-1)\pi. \quad (2.109)$$

As we have seen in Section 4.2, the gap  $|\Delta|$  increases under the RG flow, effectively freezing the quantum fluctuations of  $\hat{\varphi}$  such that the cosine term in (2.94) is minimized in the bulk. This leads to the following asymptotic value

$$\hat{\varphi}(\infty) = \frac{1}{2\sqrt{\pi}} \left( \alpha + \frac{\pi}{2} - p\pi + (2s-1)\pi \right), \quad (2.110)$$

with another integer  $s$ . With the help of bosonization identities, we deduce that the boundary charge may be related to the difference of the values of  $\hat{\varphi}(x)$  at  $x = \infty$  and  $x = 0$ :

$$\begin{aligned} Q_B &= \frac{1}{\sqrt{\pi}} \int_0^\infty dx \langle \partial_x \hat{\varphi}(x) \rangle \\ &= \frac{1}{\sqrt{\pi}} [\hat{\varphi}(\infty) - \hat{\varphi}(0)]. \end{aligned} \quad (2.111)$$

Inserting (2.109) and (2.110) we thus conclude

$$Q_B = \frac{\alpha}{2\pi} + \frac{1}{4} - \frac{1}{2}p \mod(1). \quad (2.112)$$

Comparing this result with the exact solution (2.102) without interaction we find that we have to choose  $p = 0$ . This proves the stability of the boundary charge under short-ranged electron-electron interaction within the low-energy model.

A similar calculation may be done in the case of an interface charge quantization by using an arbitrary function  $\alpha(x)$  describing the interface. In this case we get

$$\begin{aligned} Q_I &= \frac{1}{\sqrt{\pi}} \int dx \langle \partial_x \hat{\varphi} \rangle \\ &= \frac{1}{\sqrt{\pi}} [\hat{\varphi}(\infty) - \hat{\varphi}(-\infty)]. \end{aligned} \quad (2.113)$$

By using the same procedure as above, we arrive at the following result

$$Q_I = \frac{\alpha(\infty) - \alpha(-\infty)}{2\pi} \mod(1). \quad (2.114)$$

which in the case of  $\alpha(x) = \alpha_R \Theta(x) + \alpha_L \Theta(-x)$ , reduces to the non-interacting result, and thus again shows the robustness of the interface charge quantization. A similar expression has been found in Ref. [31].

### 4.3 Summary and outlook

For generic 1D insulators we have provided in this work a complete analysis of symmetry conditions to realize rational quantizations of the boundary charge. We obtained two interesting results: (a) *Any* rational quantization  $\frac{p}{q}$  can be realized if non-local symmetries involving translations are taken into account. (b) Besides the quantization unit  $\frac{1}{2}$  known from local symmetries we identified a new quantization unit  $\frac{1}{2}\bar{\rho}$ , where  $\bar{\rho}$  is the average charge per site. This has to be contrasted to the known quantization unit  $\bar{\rho}$  for interface charges.

Both the quantization of the boundary and the interface charge were shown to follow straightforwardly from the transformation laws of the boundary charge under translations and local inversion. These fundamental principles are physically very intuitive and were rigorously related to the intriguing property of insulators that local perturbations lead only to local charge redistributions. Therefore, all our results were proven to be stable against static random disorder and short-ranged electron-electron interaction. We demonstrated this explicitly by using exact diagonalization, DMRG methods, and bosonization calculations. In addition, the stability of the quantization of the boundary charge was recently analysed via functional renormalization group (fRG) studies for the interacting Rice-Mele model and the same conclusions were found [22]. Besides the boundary charge also other quantities were studied with fRG for this model like the full density profile and the precise form of edge states, where interaction effects have a more subtle effect. In the future it will be of interest to study also other quantities like density-density correlation functions in the presence of a boundary. In addition, fluctuations of the boundary charge are of relevance. While the overall size of fluctuations is expected to be small [41, 34, 43, 36, 45, 38, 17] when the gap is finite, it will be of interest to reveal universal properties of the fluctuations and to study their topological nature [121].

In addition to the general framework we have provided in this work an interesting application to identify a novel quantization class  $\frac{e}{4}$  in the special case of single-channel and nearest-neighbor hopping models at half-filling. As function of the phase variable controlling the offset of the potential modulation we found Weyl physics close to gap closing points and demonstrated the

stability of the quantization of the boundary charge in contrast to the Hall current. We suggest such systems to be realizable in cold atom systems [89, 90], in carbon based materials [91, 92] or phononic crystals [99]. Other promising candidates could be quantum dot arrays as outlined in Ref. [17], where control over all model parameters is possible. As shown in Appendix 5.8, the quantization of the boundary charge is already visible for an array size of  $\sim 20 - 30$  dots, which is within experimental reach.

As shown in Refs. [17, 18, 86, 87] the transformation law of the boundary charge under translations is also responsible for the quantization of the average linear slope of the boundary charge which is of fundamental importance for the understanding of the integer quantum Hall effect [18]. For a finite system of size commensurable with the unit cell size, it was found in Ref. [17] that the sum of the boundary charges at the left and right end of the system is zero (up to an integer). This is equivalent to the result proven rigorously in this work that the boundary charge changes sign under local inversion. The fact that the transformation laws are also responsible for rational quantization values of the boundary charge demonstrates the topological nature of the boundary charge and its usefulness for the characterization of topological insulators. This is of particular advantage compared to other topological indices, since the transformation laws are perfectly valid in the presence of disorder and interactions, as demonstrated in the present work.

Of further interest is the specification of the unknown integers in the transformation laws. They are related in a subtle way to bound states occurring at boundaries and interfaces. Therefore, their knowledge is of importance to establish a link between the boundary charge and the appearance of bound states. This question has been analysed recently in Refs. [86, 87] for the special case of noninteracting single-channel and nearest-neighbor hopping models. If only one band is occupied (i.e.,  $\bar{\rho} = \frac{1}{2}$ ) it was shown that the difference  $(n - n')\frac{1}{2} - (Q_{B,n}^R - Q_{B,n'}^R)$  is a quantized topological index related to the winding number of the gauge-invariant phase difference of the Bloch wave function between site  $m = n$  and  $m = n'$ . The same index describes the quantity  $(n - n')\frac{1}{2} + (Q_{B,n-1}^L - Q_{B,n'-1}^L)$ . In addition, it was found that the sum of the boundary charges left and right to a common boundary is given by the winding number of the phase-difference of the Bloch wave function between the first and last site of the unit cell starting at the boundary [122]. As a result, the topological index defined via the winding number of the phase difference of the Bloch wave function between different sites has a direct physical meaning and controls the transformation laws of the boundary charge in a unique way. Therefore, it will be of high interest in the future to find analogous rules for multi-channel systems via non-abelian versions of these winding numbers [123].

The framework developed in the present work can be straightforwardly generalized to other systems with a conserved quantity like, e.g., the boundary spin occurring in superconducting systems [124] or spin systems [125]. The underlying foundation for the transformation laws of the boundary charge is charge conservation and the presence of a gap. Therefore, if the spin in a certain direction is a conserved quantity analogous quantizations of the boundary spin are expected for insulating materials in the presence of symmetries. The same applies for the quantization at interfaces. Moreover, via dimensional reduction, we expect our results to be also of relevance for higher-dimensional systems.

## 5. Appendixes

### 5.1 Local vs. non-local symmetries

In this Appendix we provide a summary of our conventions to distinguish between local and non-local symmetries. Although this being standard (see, e.g., Ref. [23]), conventions sometimes differ in the literature and the material might be helpful for readers not so familiar with the precise definitions of the various symmetries.



For a given Hamiltonian  $H$ , there are four kinds of symmetries, depending on whether the symmetry operation commutes/anticommutes with  $H$  and whether it is unitary or anti-unitary

$$UHU^\dagger = H \quad , \quad SHS^\dagger = -H \quad , \quad (2.115)$$

$$THT^\dagger = H \quad , \quad CHC^\dagger = -H \quad . \quad (2.116)$$

Here,  $U$  and  $S$  are unitary operators, whereas  $T$  and  $C$  are anti-unitary operators.  $S$  is called a chiral symmetry,  $T$  a time-reversal symmetry, and  $C$  a particle-hole (or charge conjugation) symmetry. The anti-unitary symmetries  $T$  and  $C$  consist of a combination of unitary operations  $U_T$  and  $U_C$  with complex conjugation  $K$ :  $T = U_T K$  and  $C = U_C K$ . The operation  $K$  of complex conjugation requires a basis in which it is defined. Here, we take always the real-space representation in terms of  $|m\sigma\rangle$ , where  $m$  is the lattice site index and  $\sigma$  the channel index.

To distinguish local from non-local symmetries one needs to specify the unit cell and write the total Hilbert space as a direct product of the space within the unit cell (labeled by the site index  $j = 1, \dots, Z$  and the channel index  $\sigma = 1, \dots, N_c$  for each site) and the space of all unit cells labeled by the integer  $n$ . In the 1-particle subspace, the tight-binding model (2.7) can then be alternatively written as

$$H = \sum_{n,\tau} \underline{h}(\tau) \otimes |n+\tau\rangle\langle n| \quad , \quad (2.117)$$

where  $\underline{h}(\tau)$  are  $ZN_c \times ZN_c$ -matrices describing the coupling of unit cell  $n$  with unit cell  $n+\tau$  (the lattice site index  $m$  used in (2.7) is related to  $n$  and  $j$  by  $m = Z(n-1) + j$ ; note that  $\tau$  has a different meaning compared to  $\delta$  used in (2.7), the same applies for the symbol  $\underline{h}$ ). A local symmetry is then defined by a symmetry with respect to the Hamiltonian  $\underline{h}(\tau)$  (i.e., it acts only within the space of a single unit cell) and, in addition, is independent of  $\tau$

$$U\underline{h}(\tau)U^\dagger = \underline{h}(\tau) \quad , \quad S\underline{h}(\tau)S^\dagger = -\underline{h}(\tau) \quad , \quad (2.118)$$

$$T\underline{h}(\tau)T^\dagger = \underline{h}(\tau) \quad , \quad C\underline{h}(\tau)C^\dagger = -\underline{h}(\tau) \quad . \quad (2.119)$$

Using the Fourier transform  $\tilde{\underline{h}}(k) = \sum_\tau \underline{h}(\tau)e^{-ik\tau}$ , with real quasimomentum  $-\pi < k < \pi$ , this can also be written as

$$U\tilde{\underline{h}}(k)U^\dagger = \tilde{\underline{h}}(k) \quad , \quad S\tilde{\underline{h}}(k)S^\dagger = -\tilde{\underline{h}}(k) \quad , \quad (2.120)$$

$$T\tilde{\underline{h}}(-k)T^\dagger = \tilde{\underline{h}}(k) \quad , \quad C\tilde{\underline{h}}(-k)C^\dagger = -\tilde{\underline{h}}(k) \quad . \quad (2.121)$$

Within our convention, a non-local symmetry can not be written in this form. There are three possibilities: (1) The non-local symmetry can be written as a local one by taking another choice of the unit cell. (2) The non-local symmetry acts within the space of a single unit cell but depends on  $\tau$  (or, equivalently, on the quasimomentum  $k$ ). (3) The non-local symmetry does not act within the space of a single unit cell whatever choice one takes for the unit cell, i.e., can only be written with respect to the total Hamiltonian  $H$ . For  $n \neq 0$ , the symmetries  $\Pi_n$  and  $S_n$  defined in Eq. (2.35) are non-local symmetries within our definition. Examples for cases (1) and (2) are discussed in the paragraph following Eq. (2.40) via special cases for the symmetry  $\Pi_n$ . The case (1) is discussed extensively at the end of Section 3.2 when the local symmetries  $\Pi_0$  or  $S_0$  are present but not with respect to the unit cell starting at the boundary of the semi-infinite system. For  $n \neq 0$ , the symmetry  $S_n$  is an example for case (3).

### 5.2 Stability of NSP: DMRG analysis

In this Appendix we analyse the influence of static random disorder and short-ranged electron-electron interaction on the boundary and interface charge by using exact diagonalization and DMRG. For particular examples we demonstrate that the interface charge (2.22) is independent of the interface coupling  $V_I$  (up to an integer), and we show that Eqs. (2.26) and (4.154) for the boundary charge are generically valid.

We start with the interface charge and demonstrate in Figs. 13 and Figs. 14 that Eq. (2.22) holds even in the presence of random disorder as well as short-ranged electron-electron interaction, respectively. We consider an interface of the following form: Take initially two decoupled chains of the form as defined by Eq. (2.50). We want to include changes where the potential form of the right chain is shifted in the variable  $\varphi$  with respect to the left one by an integer multiple of  $\frac{2\pi}{Z}$ . This means  $\varphi \rightarrow \varphi + s\frac{2\pi}{Z}$ , which effectively shifts the right lattice by  $s$  sites compared to the left one. In Eq. (2.22) this means that  $n = s$  and  $n' = 0$ .

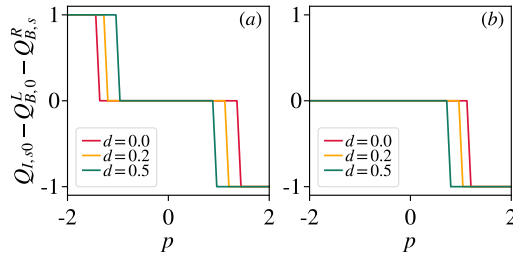


Figure 13: Stability of Eq. (2.22) with respect to random disorder. The figure shows  $Q_{I,s0} - Q_{B,0}^L - Q_{B,s}^R$  for the same parameters as used in the left column ( $Z = 4$ ) of Fig. 10 at  $\varphi = \pi/4$ , where the gap is maximal (but at finite size  $N = 1000$ ). The relative shift of the chains left and right to the interface are (a)  $s = 0$ , and (b)  $s = 1$ . We take half-filling instead of  $\mu = 0$  here. Random disorder drawn from a uniform distribution  $[-d/2, d/2)$  is added to the onsite potentials and  $p$  describes changes to the interface properties (see main text for details). As the properties of the interface are swept through  $Q_{I,s0} - Q_{B,0}^L - Q_{B,s}^R$  only changes mod(1).

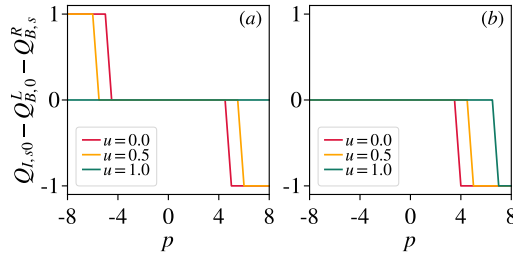


Figure 14: Stability of Eq. (2.22) with respect to interactions. The figure shows  $Q_{I,s0} - Q_{B,0}^L - Q_{B,s}^R$  for the same parameters as used in the left column ( $Z = 4$ ) of Fig. 10 at  $\varphi = \pi/4$ , where the gap is maximal (but at finite size  $N = 200$  and with larger  $V/t = 1.2$ ). The relative shift of the chains left and right to the interface are (a)  $s = 0$ , and (b)  $s = 1$ . We take half-filling instead of  $\mu = 0$  here.  $p$  describes changes to the interface properties (see main text for details). As the properties of the interface are swept through  $Q_{I,s0} - Q_{B,0}^L - Q_{B,s}^R$  only changes mod(1).

To define a single parameter  $p$  which changes the interface's properties continuously we consider the link between the rightmost site of the left lattice to the leftmost site of the right lattice to be  $t_{\text{link}} = |p|/2$  and add an onsite potential of size  $p$  to both of these sites. Therefore,  $p = 0$  is the decoupled case of two chains without an additional onsite potential at the edge and for negative  $p$  charges tend to get trapped at the interface, while for positive  $p$  they are



pushed out. We add a quenched disorder following Eq. (2.15) for the results in Fig. 13 and a electron-electron interaction following Eq. (2.62) for the results in Fig. 14. Since we concentrate on nearest-neighbor interaction we additionally scale the interaction over the interface bond by  $p$ , such that  $p = 0$  is the limit of two decoupled chains. Clearly Eq. (2.22) remains valid in both cases.

Next we study the influence of static random disorder and short-ranged electron-electron interaction on the transformation laws (2.26) and (4.154) of the boundary charge under translations and local inversion, see Figs. 15(a,b). Up to rather large values of the disorder and the electron-electron interaction both transformations laws remain perfectly valid, as expected from the NSP.

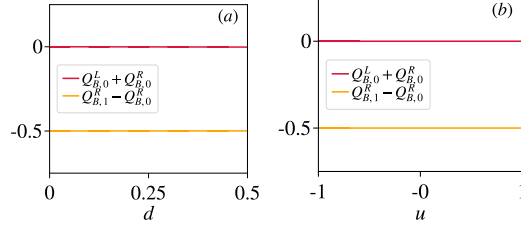


Figure 15: Stability of Eqs. (2.26) and (4.154) with respect to (a) random disorder and (b) interaction. We work at half filling such that  $\bar{\rho} = 1/2$ .  $Q_{B,1}^R - Q_{B,0}^R = \bar{\rho} \bmod(1)$  is shown to demonstrate Eq. (2.26), while  $Q_{B,0}^L + Q_{B,0}^R = 0 \bmod(1)$  illustrates Eq. (4.154). The parameters are the same as in Fig. 13.

### 5.3 Stability of NSP: 1-channel systems

In this Appendix we demonstrate the validity of Eq. (2.22) (with  $n = n' = 0$ ) for a model of two non-interacting single-channel nearest-neighbor chains coupled with each other via a tunable hopping amplitude. It is explicitly shown that (2.22) holds for any strength of the link.

Let us consider the Hamiltonian of the infinite chain  $H = H_R + H_L + V_I$  consisting of the three parts

$$H_R = \sum_{n=1}^{\infty} \{ |n\rangle\langle n| \otimes h(0) + |n+1\rangle\langle n| \otimes h(1) + |n\rangle\langle n+1| \otimes h(-1) \}, \quad (2.122)$$

$$H_L = \sum_{n=-\infty}^0 \{ |n\rangle\langle n| \otimes h(0) + |n\rangle\langle n-1| \otimes h(1) + |n-1\rangle\langle n| \otimes h(-1) \}, \quad (2.123)$$

$$V_I = \lambda [ |n=1\rangle\langle n=0| \otimes h(1) + |n=0\rangle\langle n=1| \otimes h(-1) ], \quad (2.124)$$

which describe the right semi-infinite chain, the left semi-infinite chain, and the tunneling between them, respectively. Here, in contrast to the lattice site index  $m$ , the index  $n$  enumerates unit cells. Both  $H_R$  and  $H_L$  describe the lattices with the same structure of a unit cell, which is encoded in  $h(0) = \sum_{j=1}^Z v_j |j\rangle\langle j| - \sum_{j=1}^{Z-1} t_j (|j\rangle\langle j+1| + |j+1\rangle\langle j|)$ ,  $h(1) = -t_Z |j=1\rangle\langle j=Z|$ , and  $h(-1) = h^\dagger(1)$ , i.e. characterized by  $Z$  sites  $j = 1, \dots, Z$  per unit cell, a single orbital (channel) per site, and by the same values for hoppings  $t_j$  and onsite potentials  $v_j$ . The tunneling amplitude  $\lambda t_Z$  between the two subsystems is quantified by the real-valued parameter  $\lambda \geq 0$ . Its special values  $\lambda = 0$  and  $\lambda = 1$  correspond to the cases of the two decoupled semi-infinite chains and the translationally invariant infinite chain, respectively. A restoration of the translational

symmetry in the latter case is guaranteed by the perfect matching of the unit cells touching each other at the interface.

Due to the same structure, the Hamiltonians  $H_R$  and  $H_L$  appear to be isospectral, and their extended eigenstates can be therefore labeled by the same band index  $\alpha$  and quasimomentum  $k$  on the both sides from the interface. Moreover, these quantum numbers can be also used for a construction of scattering eigenstates of the coupled system, since eigenenergies  $\epsilon_k^{(\alpha)}$  of the extended states remain independent of  $\lambda$ , and they can be ultimately evaluated from the bulk Hamiltonian, i.e. at  $\lambda = 1$ . On the basis of this observation, we make the following ansatz for the two distinct scattering eigenstates additionally labeled by either  $r$  or  $l$ :

$$\begin{aligned} \psi_k^{(\alpha,r)}(n,j) &= \frac{\Theta_{n \leq 0}}{\sqrt{2\pi}} [\chi_k^{(\alpha)}(j) e^{ikn} + r_k^{(\alpha)} \chi_{-k}^{(\alpha)}(j) e^{-ikn}] \\ &\quad + \frac{\Theta_{n \geq 1}}{\sqrt{2\pi}} t_k^{(\alpha)} \chi_k^{(\alpha)}(j) e^{ikn}, \end{aligned} \quad (2.125)$$

$$\begin{aligned} \psi_k^{(\alpha,l)}(n,j) &= \frac{\Theta_{n \geq 1}}{\sqrt{2\pi}} [\chi_{-k}^{(\alpha)}(j) e^{-ikn} + r_k'^{(\alpha)} \chi_k^{(\alpha)}(j) e^{ikn}] \\ &\quad + \frac{\Theta_{n \leq 0}}{\sqrt{2\pi}} t_k'^{(\alpha)} \chi_{-k}^{(\alpha)}(j) e^{-ikn}, \end{aligned} \quad (2.126)$$

with  $k \in [0, \pi]$ . Both  $\psi_k^{(\alpha,r)}$  and  $\psi_k^{(\alpha,l)}$  as well as the bulk Bloch states  $\chi_{\pm k}^{(\alpha)}$  correspond to the eigenenergy  $\epsilon_k^{(\alpha)}$ . In the following, we focus on the band  $\alpha$  and omit the band index for brevity.

Inserting the ansatz (2.125) and (2.126) into the eigenvalue problem, we establish the scattering matrix

$$S_k = \begin{pmatrix} t_k & r_k' \\ r_k & t_k' \end{pmatrix}. \quad (2.127)$$

Its components read

$$t_k = t_k' = \lambda \frac{e^{2i\phi_k} - 1}{e^{2i\phi_k} - \lambda^2}, \quad (2.128)$$

$$r_k = \frac{\lambda^2 - 1}{e^{2i\phi_k} - \lambda^2} e^{2i\varphi_k(Z)}, \quad (2.129)$$

$$r_k' = \frac{\lambda^2 - 1}{e^{2i\phi_k} - \lambda^2} e^{2i\phi_k} e^{-2i\varphi_k(Z)}, \quad (2.130)$$

where  $\phi_k = \varphi_k(Z) - \varphi_k(1) - k$  is a gauge-invariant phase difference expressed in terms of the gauge-dependent phases  $\varphi_k(j)$  of the complex-valued components  $\chi_k(j)$ . By a direct calculation one can confirm the unitarity property  $S_k^\dagger S_k = 1$ , which implies both the orthogonality of  $\psi_k^{(\alpha,r)}$  and  $\psi_k^{(\alpha,l)}$  as well as their proper normalization.

The interface charge  $Q_I^{(\alpha)}$  associated with the band  $\alpha$  consists of the Friedel part  $Q_F^{(\alpha)}$  and the polarization part  $Q_P^{(\alpha)}$  by using the following splitting based on Eq. (2.18)

$$Q_I^{(\alpha)} = Q_F^{(\alpha)} + Q_P^{(\alpha)}, \quad (2.131)$$

$$Q_F^{(\alpha)} = \sum_{m=-\infty}^{\infty} \left[ \rho^{(\alpha)}(m) - \rho_{\text{bulk}}^{(\alpha)}(m) \right] f(m), \quad (2.132)$$

$$Q_P^{(\alpha)} = \sum_{m=-\infty}^{\infty} \left[ \rho_{\text{bulk}}^{(\alpha)}(m) - \frac{1}{Z} \right] f(m), \quad (2.133)$$

with  $m = Z(n-1) + j$ . Here,  $\rho_{\text{bulk}}^{(\alpha)}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} dk |\chi_k^{(\alpha)}(j)|^2$  is the contribution from band  $\alpha$  to the charge at site  $m$  from the bulk Hamiltonian. As shown in Eq. (2.30), the polarization part to the interface charge vanishes  $Q_P^{(\alpha)} = 0$ . In turn, the Friedel part  $Q_F^{(\alpha)} \equiv Q_F$  amounts to

$$Q_F = \sum_{n=-\infty}^0 \int_{-\pi}^{\pi} \frac{dk}{2\pi} r_k^* \sum_{j=1}^Z \chi_k^2(j) e^{2i(k-i\eta)n} \quad (2.134)$$

$$+ \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} \frac{dk}{2\pi} r_k' \sum_{j=1}^Z \chi_k^2(j) e^{2i(k+i\eta)n} \quad (2.135)$$

$$= -1 + \int_{-\pi}^{\pi} \frac{dk}{2\pi} (r_k' - r_k^*) \sum_{j=1}^Z \chi_k^2(j) \frac{ie^{ik}}{2 \sin k}, \quad (2.136)$$

where  $\eta \rightarrow 0^+$  is a convergence factor. The last equality is only valid for  $\lambda^2 \neq 1$ , since the limits  $\lambda \rightarrow 1$  and  $\eta \rightarrow 0^+$  do not commute. In the translationally invariant case  $\lambda = 1$  there is no reflection at the interface, and one simply gets  $Q_F = 0$ .

In the following we prove that in general  $Q_F$  takes only integer values for arbitrary  $\lambda$ .

Let us introduce the two gauges: I)  $\chi_k^{\text{I}}$  with  $e^{2i\varphi_k(Z)} = 1$ , i.e. the last component is real; II)  $\chi_k^{\text{II}}$  with  $e^{2i\varphi_k(1)} = 1$ , i.e. the first component is real. Apparently,  $\chi_k^{\text{II}} = e^{i\phi_k + ik} \chi_k^{\text{I}}$ .

Next, we express the quantity  $Q_F + 1$  in the mixed form

$$\begin{aligned} \int_{-\pi}^{\pi} \frac{dk}{2\pi} \left[ -\frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \sum_{j=1}^Z [\chi_k^{\text{I}}(j)]^2 \frac{ie^{ik}}{2 \sin k} \right. \\ \left. + \frac{\lambda^2 - 1}{\lambda^2 - e^{-2i\phi_k}} e^{-2i\phi_k} \sum_{j=1}^Z [\chi_k^{\text{II}}(j)]^2 \frac{ie^{-ik}}{2 \sin k} \right]. \end{aligned} \quad (2.137)$$

In Ref. [87] we established that the components of the Bloch state in the gauge I have the form

$$\chi_k^{\text{I}}(j) = \frac{f_j^{\text{I}} e^{-ik} + g_j^{\text{I}}}{\sqrt{N_k^{\text{I}}}}, \quad 1 \leq j \leq Z-1, \quad (2.138)$$

$$\chi_k^{\text{I}}(Z) = \frac{s}{\sqrt{N_k^{\text{I}}}}, \quad (2.139)$$

where  $f_j^{\text{I}}$ ,  $g_j^{\text{I}}$ , and  $s$  are real-valued polynomial functions of  $\epsilon_k$ , and  $N_k^{\text{I}} = s^2 + \sum_{j=1}^{Z-1} |f_j^{\text{I}} e^{-ik} + g_j^{\text{I}}|^2$ . In that paper we also noted the following relations

$$-\text{Im}[\chi_k^{\text{I}\dagger} \frac{d}{dk} \chi_k^{\text{I}}] = \sum_{j=1}^{Z-1} \frac{(f_j^{\text{I}})^2 + f_j^{\text{I}} g_j^{\text{I}} \cos k}{N_k^{\text{I}}}, \quad (2.140)$$

$$\begin{aligned} \sum_{j=1}^Z [\chi_k^{\text{I}}(j)]^2 \frac{ie^{ik}}{2 \sin k} &= \frac{ie^{ik}}{2 \sin k} + \sum_{j=1}^{Z-1} \frac{(f_j^{\text{I}})^2 + f_j^{\text{I}} g_j^{\text{I}} e^{ik}}{N_k^{\text{I}}} \\ &= -\text{Im}[\chi_k^{\text{I}\dagger} \frac{d}{dk} \chi_k^{\text{I}}] + \frac{ie^{ik}}{2 \sin k} + \frac{i \sin k}{N_k^{\text{I}}} \sum_{j=1}^{Z-1} f_j^{\text{I}} g_j^{\text{I}}. \end{aligned} \quad (2.141)$$

In addition, it is also possible to derive the relation

$$\frac{\sin k}{N_k^{\text{I}}} \sum_{j=1}^{Z-1} f_j^{\text{I}} g_j^{\text{I}} = -\frac{1}{2s} \frac{ds}{dk}. \quad (2.142)$$

Hence,

$$\begin{aligned} \sum_{j=1}^Z [\chi_k^I(j)]^2 \frac{ie^{ik}}{2 \sin k} &= -\text{Im}[\chi_k^{I\dagger} \frac{d}{dk} \chi_k^I] - \frac{1}{2} \\ &+ \frac{i \cos k}{2 \sin k} - \frac{i}{2s} \frac{ds}{dk}. \end{aligned} \quad (2.143)$$

Let us now establish similar relations for  $\chi_k^{\text{II}}$ . We note that the Bloch state

$$\bar{\chi}_k^{\text{II}} = \begin{pmatrix} 0 & e^{-ik} 1_{(Z-1) \times (Z-1)} \\ 1 & 0 \end{pmatrix} \chi_k^{\text{II}} \quad (2.144)$$

is the eigenstate corresponding to the re-defined unit cell, which begins with the site 2, has the pre-last site  $Z$ , and ends with the site 1. Moreover, the component  $\bar{\chi}_k^{\text{II}}(Z)$  is real, and then by analogy with (2.143) it holds

$$\begin{aligned} \sum_{j=1}^Z [\bar{\chi}_k^{\text{II}}(j)]^2 \frac{ie^{ik}}{2 \sin k} &= -\text{Im}[\bar{\chi}_k^{\text{II}\dagger} \frac{d}{dk} \bar{\chi}_k^{\text{II}}] - \frac{1}{2} \\ &+ \frac{i \cos k}{2 \sin k} - \frac{i}{2\bar{s}} \frac{d\bar{s}}{dk}, \end{aligned} \quad (2.145)$$

where  $\bar{s}$  is a part of the representation for  $\bar{\chi}_k^{\text{II}}$ , which is analogous to (2.138), (2.139).

From (2.144) it follows that

$$\text{Im}[\bar{\chi}_k^{\text{II}\dagger} \frac{d}{dk} \bar{\chi}_k^{\text{II}}] = \text{Im}[\chi_k^{\text{II}\dagger} \frac{d}{dk} \chi_k^{\text{II}}] + |\bar{\chi}_k^{\text{II}}(Z)|^2 - 1, \quad (2.146)$$

and thus we find that

$$\begin{aligned} \sum_{j=1}^Z [\chi_k^{\text{II}}(j)]^2 \frac{ie^{-ik}}{2 \sin k} &= \sum_{j=1}^Z [\bar{\chi}_k^{\text{II}}(j)]^2 \frac{ie^{ik}}{2 \sin k} + |\bar{\chi}_k^{\text{II}}(Z)|^2 \\ &= -\text{Im}[\chi_k^{\text{II}\dagger} \frac{d}{dk} \chi_k^{\text{II}}] + \frac{1}{2} + \frac{i \cos k}{2 \sin k} - \frac{i}{2\bar{s}} \frac{d\bar{s}}{dk}. \end{aligned} \quad (2.147)$$

Inserting (2.143) and (2.147) into (2.137) and accounting the symmetry properties of integrands under the reflection  $k \rightarrow -k$ , we obtain

$$\begin{aligned} Q_F + 1 &= \int_{-\pi}^{\pi} \frac{dk}{2\pi} \left\{ \text{Re} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \right] \right. \\ &\times \left( \text{Im}[\chi_k^{I\dagger} \frac{d}{dk} \chi_k^I] - \text{Im}[\chi_k^{\text{II}\dagger} \frac{d}{dk} \chi_k^{\text{II}}] + 1 \right) \\ &\left. + \text{Im} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \right] \left( \cot k - \frac{1}{2} \frac{d \ln(s\bar{s})}{dk} \right) \right\}. \end{aligned} \quad (2.148)$$

From the transformation between the two gauges we obtain the relation

$$\text{Im}[\chi_k^{I\dagger} \frac{d}{dk} \chi_k^I] - \text{Im}[\chi_k^{\text{II}\dagger} \frac{d}{dk} \chi_k^{\text{II}}] = -1 - \frac{d\phi_k}{dk}. \quad (2.149)$$

A less obvious identity

$$s\bar{s} \sin^2 \phi_k = g_1^I \bar{f}_{Z-1}^{\text{II}} \sin^2 k \equiv \left( \prod_{j=1}^{Z-1} t_j^2 \right) \sin^2 k \quad (2.150)$$

follows from the identifications

$$\frac{\bar{s}}{\sqrt{N_k^\Pi}} e^{-i\phi_k} = \frac{f_1^I + g_1^I e^{ik}}{\sqrt{N_k^I}}, \quad (2.151)$$

$$\frac{s}{\sqrt{N_k^I}} e^{i\phi_k} = \frac{\bar{f}_{Z-1}^\Pi e^{-ik} + \bar{g}_{Z-1}^\Pi}{\sqrt{N_k^\Pi}}, \quad (2.152)$$

and the observation  $g_1^I = \bar{f}_{Z-1}^\Pi = \prod_{j=1}^{Z-1} t_j$  which can be made on the basis of expressions quoted in Ref. [87]. Differentiating (2.150) with respect to  $k$  yields

$$\cot k - \frac{1}{2} \frac{d \ln(s\bar{s})}{dk} = \cot \phi_k \frac{d\phi_k}{dk}. \quad (2.153)$$

With help of (2.149) and (2.153) we cast (2.148) to

$$\begin{aligned} Q_F + 1 = \int_{-\pi}^{\pi} \frac{dk}{2\pi} \left\{ -\operatorname{Re} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \right] \frac{d\phi_k}{dk} \right. \\ \left. + \operatorname{Im} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi_k}} e^{2i\phi_k} \right] \cot \phi_k \frac{d\phi_k}{dk} \right\}. \end{aligned} \quad (2.154)$$

Making the change of the integration variable  $k \rightarrow \phi_k$  and accounting possible multiple windings of the phase  $\phi_k$ , which are quantified by the integer winding number  $\operatorname{wn}[\phi_k] = \int_{-\pi}^{\pi} \frac{dk}{2\pi i} e^{-i\phi_k} \frac{d}{dk} e^{i\phi_k}$ , we express

$$\begin{aligned} Q_F + 1 &= \operatorname{wn}[\phi_k] \int_{-\pi}^{\pi} \frac{d\phi}{2\pi} \left\{ -\operatorname{Re} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi}} e^{2i\phi} \right] \right. \\ &\quad \left. + \operatorname{Im} \left[ \frac{\lambda^2 - 1}{\lambda^2 - e^{2i\phi}} e^{2i\phi} \right] \cot \phi \right\} \\ &= \operatorname{wn}[\phi_k] \operatorname{sign}(\lambda^2 - 1). \end{aligned} \quad (2.155)$$

For the two decoupled chains ( $\lambda = 0$ ), we obtain

$$\begin{aligned} Q_F^{(\alpha)} &= -1 - \operatorname{wn}[\phi_k^{(\alpha)}] \\ &= \operatorname{wn}[\varphi_k^{(\alpha)}(1) - \varphi_k^{(\alpha)}(Z)], \end{aligned} \quad (2.156)$$

i.e. an integer number. This result persists in the whole range  $0 \leq \lambda < 1$ .

We conclude that the total interface charge  $Q_I(\lambda)$ , which might also include integer edge state contributions, is a sum of integers for any  $\lambda$ , and therefore  $Q_I(\lambda) = 0 \pmod{1}$ . On the other hand, since  $Q_B^R + Q_B^L = Q_I(\lambda = 0)$  by definition, we find  $Q_B^R + Q_B^L = Q_I(\lambda) \pmod{1}$ , in agreement with (2.4) for the model discussed in this Appendix.

#### 5.4 Symmetries for single-channel and nearest-neighbor hopping models

In this Appendix we prove the symmetry conditions (2.52) and (2.53) for the special case of a tight-binding model with one channel  $N_c = 1$  and nearest-neighbor hopping  $\delta = 0, \pm 1$ . In this case, the model is parametrized by  $Z$  on-site potentials  $v_m = v_{m+Z}$  and  $Z$  nearest-neighbor hoppings  $t_m = t_{m+Z}$  defined by

$$v_m = h_m(0) = v_m^* \quad , \quad t_m = -h_m(1) = -h_{m+1}(-1)^*. \quad (2.157)$$

Without loss of generality one can choose all  $t_m > 0$  real and positive since possible phases can be gauged away by a unitary transformation (see, e.g., Appendix A in Ref. [87] for a proof). The unitary transformation  $U_m$  must be a phase factor

$$U_m = e^{i\varphi_m} \quad , \quad \varphi_m = \varphi_{m+Z}. \quad (2.158)$$

Inserting these equations in the symmetry condition (2.39) for  $\Pi_n$  we find

$$v_m = v_{Z-m-n+1}, \quad (2.159)$$

$$t_m = e^{-i(\varphi_{Z-m-n} - \varphi_{Z-m-n+1})} t_{Z-m-n}. \quad (2.160)$$

Since  $t_m$  and  $t_{Z-m-n}$  are both positive this can only be fulfilled for  $U_m = U_{m+1}$  which is just a homogeneous and trivial phase factor. Therefore, we can set  $U_m = 1$  and find the condition (2.52).

Considering the other symmetry condition (2.40) for  $S_n$  we find

$$v_m = -v_{m-n}, \quad (2.161)$$

$$t_m = -e^{-i(\varphi_{m-n+1} - \varphi_{m-n})} t_{Z-m-n}. \quad (2.162)$$

Since  $t_m$  and  $t_{Z-m-n}$  are both positive this requires  $U_m = -U_{m+1}$  which, up to an unimportant common phase factor, is only realized for  $U_m = (-1)^m$ . This proves the condition (2.53).

### 5.5.5 Boundary charge for Dirac model (!!!!!)

In this Appendix we determine all eigenstates of the semi-infinite Dirac model (2.99) and prove Eqs. (2.100) and (2.101). We start with solving the eigenvalue equation

$$\left[ -iv_F \sigma_3 \partial_x + |\Delta|(\sigma_+ e^{i\alpha} + \sigma_- e^{-i\alpha}) \right] \underline{\psi}(x) = \epsilon \underline{\psi}(x), \quad (2.163)$$

with  $\sigma_{\pm} = \frac{1}{2}(\sigma_x \pm i\sigma_y)$ ,  $\underline{\psi}(x) = (R(x), L(x))^T$ , and the boundary condition  $R(0) + L(0) = 0$ . There are two distinct spectral regions: I)  $|\epsilon| < |\Delta|$ , and II)  $|\epsilon| > |\Delta|$ . In the region I we find a single bound state solution for  $\sin \alpha > 0$  at energy  $\epsilon = -|\Delta| \cos \alpha$ , whose wavefunction is given by

$$\underline{\psi}^I(x) = \sqrt{\kappa} \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-\kappa x}, \quad (2.164)$$

with  $\kappa = \frac{|\Delta| |\sin \alpha|}{v_F}$ . In the second (II) spectral region we find a continuum of scattering states labeled by the momentum  $k \in [0, \infty)$  and corresponding to the two bands with energies  $\epsilon_{k,\pm} = \pm \sqrt{v_F^2 k^2 + |\Delta|^2} \equiv \pm \epsilon_k$ . The eigenstates of the lower (valence) band have the following form

$$\underline{\psi}_k(x) = \frac{1}{\sqrt{2\pi N_k}} \left[ \begin{pmatrix} -|\Delta| e^{i\alpha} \\ v_F k + \epsilon_k \end{pmatrix} e^{ikx} - s_k \begin{pmatrix} -|\Delta| e^{i\alpha} \\ -v_F k + \epsilon_k \end{pmatrix} e^{-ikx} \right], \quad (2.165)$$

with the normalization factor

$$N_k = |\Delta|^2 + (v_F k + \epsilon_k)^2 = 2\epsilon_k(\epsilon_k + v_F k), \quad (2.166)$$

and

$$s_k = \frac{|\Delta| e^{i\alpha} - v_F k - \epsilon_k}{|\Delta| e^{i\alpha} + v_F k - \epsilon_k}. \quad (2.167)$$

We note the helpful properties

$$|s_k|^2 = \frac{\epsilon_k + v_F k}{\epsilon_k - v_F k}, \quad s_k s_{-k} = 1. \quad (2.168)$$

Assuming that the chemical potential is located at the bottom of the conduction band, the bound state is occupied for  $0 < \alpha < \pi$ , and all valence band states  $\underline{\psi}_k$  are filled. Neglecting the

strongly oscillating parts (providing unimportant corrections of  $O(\frac{\Delta}{v_F k_F}) \ll 1$ ), the contribution of each eigenstate to the density is given by

$$\rho_\psi(x) = \underline{\psi}^\dagger(x) \underline{\psi}(x) = |R(x)|^2 + |L(x)|^2. \quad (2.169)$$

We denote the contributions of the eigenstates  $\underline{\psi}^I$  and  $\underline{\psi}_k$  to the physical density by  $\rho_I(x)$  and  $\rho_k(x)$ , respectively. This gives for the total density relative to the average bulk density  $\bar{\rho}$

$$\rho(x) - \bar{\rho} = \rho_I(x) + \delta\rho_{II}(x), \quad (2.170)$$

$$\delta\rho_{II}(x) = \int_0^\infty dk \left[ \rho_k(x) - \frac{1}{\pi} \right], \quad (2.171)$$

and, according to the definition (2.16), the boundary charge follows from

$$Q_B = \int_0^\infty dx [\rho(x) - \bar{\rho}] f(x) = Q_B^I + Q_B^{II}, \quad (2.172)$$

$$Q_B^I = \int_0^\infty dx \rho_I(x) f(x), \quad (2.173)$$

$$Q_B^{II} = \int_0^\infty dx \delta\rho_{II}(x) f(x). \quad (2.174)$$

For the envelope function  $f(x)$  we choose the form  $f(x) = e^{-\eta x}$  with infinitesimally small  $\eta \rightarrow 0^+$ .

The bound state is occupied for  $0 < \alpha < \pi$  and gives an integer contribution to the boundary charge

$$Q_B^I = \int_0^\infty dx |\underline{\psi}^I(x)|^2 = \Theta_{0 < \alpha < \pi}. \quad (2.175)$$

This proves Eq. (2.100).

To calculate the scattering part  $Q_B^{II}$  to the boundary charge we use (2.165), (4.41), (2.167), and (2.168) and find after a straightforward calculation

$$\begin{aligned} \delta\rho_{II}(x) &= -\frac{|\Delta|}{2\pi} \int_{-\infty}^\infty dk \frac{e^{2ikx}}{\epsilon_k} \\ &\quad \times \frac{|\Delta| - \epsilon_k \cos \alpha - i v_F k \sin \alpha}{\epsilon_k - |\Delta| \cos \alpha}. \end{aligned} \quad (2.176)$$

Inserting this result in (2.174) and performing the integration over  $x$  we obtain

$$\begin{aligned} Q_B^{II} &= -\frac{1}{4} - \frac{|\Delta|}{4\pi} \int_{-\infty}^\infty dk \frac{v_F \sin \alpha}{\epsilon_k (\epsilon_k - |\Delta| \cos \alpha)} \\ &= -\frac{1}{4} + \frac{\ln(-e^{i\alpha})}{2\pi i}. \end{aligned} \quad (2.177)$$

This proves Eq. (2.101).

### 5.6 Boundary charge at zero gap

For the tight-binding model  $H_0$ , given by (2.64), restricted to the semi-infinite system  $m > 0$ , the eigenfunctions are given by (we set  $a = 1$ )

$$\psi_k(m) = \frac{1}{\sqrt{2\pi}} (e^{ikm} - e^{-ikm}), \quad (2.178)$$



with  $0 < k < \pi$ . For filling  $\bar{\rho} = k_F/\pi$ , this leads to the following charge  $\rho(m)$  at site  $m$

$$\begin{aligned}\rho(m) &= \int_0^{k_F} dk |\psi_k(m)|^2 \\ &= -\frac{1}{2\pi} \int_{-k_F}^{k_F} dk e^{2ikm} + \bar{\rho}.\end{aligned}\quad (2.179)$$

Inserting this result in the formula (2.16) for the boundary charge  $Q_B \equiv Q_B^R$ , we get

$$Q_B = -\frac{1}{4\pi} \int_{-k_F}^{k_F} dk \sum_{m=-\infty}^{\infty} e^{2ikm} f(m) + \frac{k_F}{2\pi}. \quad (2.180)$$

Choosing  $f(m) = e^{-\eta|m|}$ , we find  $\sum_{m=-\infty}^{\infty} e^{2ikm} f(m) = \pi\delta(k)$  and obtain for the boundary charge of  $H_0$  at zero gap

$$Q_B = -\frac{1}{4} + \frac{1}{2}\bar{\rho}. \quad (2.181)$$

This single-band model can be differently represented in terms of uniform unit cells with  $Z$  sites. This is especially useful, if we have in mind to add a  $Z$ -periodic perturbation on top of  $H_0$  (2.64). In the new representation, the single cosine band folds into  $Z$  bands with the reduced Brillouin zone (RBZ)  $[-\frac{\pi}{Z}, \frac{\pi}{Z})$ , the adjacent bands touching each other either in the center or at the edges of the RBZ. Choosing  $k_F/\pi$  of the original model to be rational,  $\frac{k_F}{\pi} = \frac{\nu}{Z}$ , we occupy  $\nu$  bands in the folded representation, and (2.181) then reads

$$Q_B = -\frac{1}{4} + \frac{\nu}{2Z}. \quad (2.182)$$

Adding a  $Z$ -periodic perturbation generically opens  $Z-1$  gaps between all  $Z$  bands. Having the chemical potential in the  $\nu$ th gap, we can evaluate the correction to (2.182) due to the perturbation by means of the low-energy theory developed in Appendix 5.5.

This consideration clarifies the physical meaning of Eq. (2.105).

### 5.7 Interface charge for Dirac model

In this Appendix we consider an interface between two Dirac models according to the Hamiltonian (2.106), where the phase  $\alpha(x)$  of the gap parameter depends on  $x$ . We will prove the Goldstone-Wilczek formula (2.107) for the interface charge for the particular choice  $\alpha(x) = \alpha_R\Theta(x) + \alpha_L\Theta(-x)$ . We define the parameter  $\delta\alpha = \alpha_R - \alpha_L$ . The eigenstates follow from the equation

$$[-iv_F\sigma_3\partial_x + |\Delta|(\sigma_+e^{i\alpha(x)} + \sigma_-e^{-i\alpha(x)})]\underline{\psi}(x) = \epsilon\underline{\psi}(x), \quad (2.183)$$

with  $\sigma_{\pm} = \frac{1}{2}(\sigma_x \pm i\sigma_y)$  and  $\underline{\psi}(x)^T = (R(x), L(x))^T$ . Like in the case of the semi-infinite Dirac model discussed in Appendix 5.5, we separate the spectrum of the Hamiltonian into two regions: I)  $|\epsilon| < |\Delta|$ , and II)  $|\epsilon| > |\Delta|$ .

The bound state solution appears for  $\sin(\delta\alpha/2) > 0$  with energy  $\epsilon = -|\Delta|\cos(\delta\alpha/2)$ , and is given by

$$\underline{\psi}^I(x) = \sqrt{\frac{\kappa}{2}} \begin{pmatrix} 1 \\ -e^{-i\frac{\alpha_R+\alpha_L}{2}} \end{pmatrix} e^{-\kappa|x|}, \quad (2.184)$$

with  $\kappa = \frac{|\Delta|}{v_F} \sin(\delta\alpha/2)$ . If it is occupied it gives an integer contribution to the interface charge.

For each energy  $|\epsilon_k| > |\Delta|$ , the extended eigenstates can be chosen as scattering states within two scattering channels. The first one (denoted by the index  $r$ )

$$\begin{aligned} \underline{\psi}_k^{(r)}(x) = & \frac{\Theta(-x)}{\sqrt{2\pi}} [\underline{\chi}_{L,k} e^{ikx} + r_k \underline{\chi}_{L,-k} e^{-ikx}] \\ & + \frac{\Theta(x)}{\sqrt{2\pi}} t_k \underline{\chi}_{R,k} e^{ikx} \end{aligned} \quad (2.185)$$

represents the scattering of a wave incident on the interface from the left. The second scattering eigenstate (denoted by the index  $l$ )

$$\begin{aligned} \underline{\psi}_k^{(l)}(x) = & \frac{\Theta(x)}{\sqrt{2\pi}} [\underline{\chi}_{R,-k} e^{-ikx} + r'_k \underline{\chi}_{R,k} e^{ikx}] \\ & + \frac{\Theta(-x)}{\sqrt{2\pi}} t'_k \underline{\chi}_{L,-k} e^{-ikx} \end{aligned} \quad (2.186)$$

represents the scattering of a wave incident on the interface from the right. In above expressions,  $k \in [0, \infty)$  stands for the momentum quantum number, and

$$\underline{\chi}_{R/L,k} = \frac{1}{\sqrt{2\epsilon(\epsilon - v_F k)}} \begin{pmatrix} -|\Delta| e^{i\alpha_{R/L}} \\ v_F k - \epsilon \end{pmatrix} \quad (2.187)$$

are the normalized Bloch eigenstates of the right-sided ( $x > 0$ ) and left-sided ( $x < 0$ ) bulk Hamiltonians with eigenenergies  $\epsilon = \pm\epsilon_k$ .

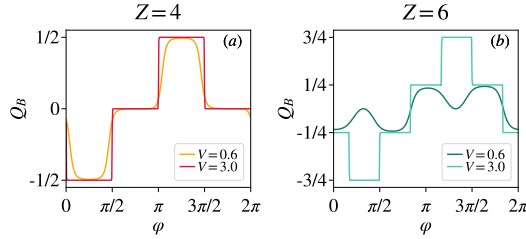


Figure 16: The same as the center row of Fig. 10, but for small  $N = 24$  and two values of the  $V$ .

The scattering amplitudes  $r_k, t_k$  and  $r'_k, t'_k$  can be determined from the continuity condition at the interface

$$\underline{\psi}_k^{(r/l)}(0^+) = \underline{\psi}_k^{(r/l)}(0^-). \quad (2.188)$$

This results in the expressions

$$r_k = r'_k = \frac{|\Delta|(e^{i\alpha_R} - e^{i\alpha_L})}{(\epsilon - v_F k)e^{i\alpha_L} - (\epsilon + v_F k)e^{i\alpha_R}}, \quad (2.189)$$

$$t_k = \frac{2v_F k e^{i\alpha_L}}{(\epsilon + v_F k)e^{i\alpha_R} - (\epsilon - v_F k)e^{i\alpha_L}}, \quad (2.190)$$

$$t'_k = \frac{2v_F k e^{i\alpha_R}}{(\epsilon + v_F k)e^{i\alpha_R} - (\epsilon - v_F k)e^{i\alpha_L}}. \quad (2.191)$$

By an explicit calculation one can readily verify the fulfillment of the unitarity conditions

$$|t_k|^2 + |r_k|^2 = |t'_k|^2 + |r'_k|^2 = 1, \quad (2.192)$$

$$t^* r'_k + r_k^* t'_k = 0. \quad (2.193)$$

For a filled valence band we choose  $\epsilon = -\epsilon_k$  and identify the extended states' contribution to the interface charge

$$\begin{aligned} Q_I^{\text{II}} &= \int_{-\infty}^{\infty} dx f(x) \int_0^{\infty} dk \left( |\psi_k^{(r)}(x)|^2 + |\psi_k^{(l)}(x)|^2 - \frac{2}{\pi} \right) \\ &= \frac{|\Delta|}{\pi} \int_0^{\infty} dx f(x) \int_{-\infty}^{\infty} dk \frac{e^{2ikx}}{\epsilon_k} r'_k. \end{aligned} \quad (2.194)$$

As one can conclude from (2.189), this quantity periodically depends on  $\delta\alpha = \alpha_R - \alpha_L$ . Evaluating (2.194) for  $\delta\alpha \in (0, 2\pi)$  with  $f(x) = e^{-\eta|x|}$ ,  $\eta \rightarrow 0^+$ , we obtain

$$Q_I^{\text{II}} = \frac{\delta\alpha}{2\pi} - 1. \quad (2.195)$$

Putting the chemical potential at the bottom of the conduction band, we receive an additional contribution  $Q_I^{\text{I}} = 1$  from the edge state (2.184), which is present for every value of  $\delta\alpha$ , and obtain the resulting expression (2.107) for the total interface charge.

### 5.8 Finite smaller systems

In this Appendix we show that the quantization of the boundary charge according to Fig. 10 is already visible for a tight-binding chain of  $\sim 20$  sites. As demonstrated in Fig. 16 for  $N = 24$  lattice sites the quantization can be demonstrated robustly as long as larger  $V$  can be accessed such that the localization length becomes small compared to the lattice size.

The results shown in Figs. 10 and 16 can be easily understood in the atomic limit  $V \gg t$ : The dominant contribution to  $Q_B$  comes from the polarization charge  $Q_P$  (2.133), while an eventual integer-valued Friedel charge contribution (2.132) is exactly cancelled by edge state contributions. To compute  $Q_P$ , we use the elaborated expression (see Ref. [87] for details)

$$Q_P = -\frac{1}{Z} \sum_{\alpha=1}^{Z/2} \sum_{j=1}^Z j \left( |\chi^{(\alpha)}(j)|^2 - \frac{1}{Z} \right), \quad (2.196)$$

where the occupied bands  $\epsilon^{(\alpha)}$  are approximately given by the potential components  $v_{\tilde{j}} < 0$  (one can even associate  $\tilde{j}$  with the band index  $\alpha$  sorting  $v_{\tilde{j}}$ 's in the ascending order for each value of  $\varphi$ ). The corresponding eigenstate  $\chi^{(\alpha)}(j)$  possesses the only unity component  $\chi^{(\alpha)}(\tilde{j}) = 1$ , while  $\chi^{(\alpha)}(j \neq \tilde{j}) = 0$ . The plateau values in the discussed figures then immediately follow from (2.196). [It can so happen that two eigenstates  $v_{\tilde{j}_1}(\varphi)$  and  $v_{\tilde{j}_2}(\varphi)$  become degenerate at some value of  $\varphi$ , and then it is necessary to consider  $\frac{1}{\sqrt{2}}\{\chi^{(\alpha_1)}(j) \pm \chi^{(\alpha_2)}(j)\}$  for the eigenstates. This, however, does not alter the plateau value of  $Q_B$ .]

## 3 Universality of Boundary Charge Fluctuations

### Abstract

We establish the quantum fluctuations  $\Delta Q_B^2$  of the charge  $Q_B$  accumulated at the boundary of an insulator as an integral tool to characterize phase transitions where a direct gap closes (and reopens), typically occurring for insulators with topological properties. The power of this characterization lies in its capability to treat different kinds of insulators on equal footing; being applicable to transitions between topological and non-topological band, Anderson, and Mott insulators alike. In the vicinity of the phase transition we find a universal scaling  $\Delta Q_B^2(E_g)$  as function of the gap size  $E_g$  and determine its generic form in various dimensions. For prototypical phase transitions with a massive Dirac-like bulk spectrum we demonstrate a scaling with the inverse gap in one dimension and a logarithmic one in two dimensions.

### 0.1 Introduction

In the last few decades studies concerning topological phases of matter, i.e. phases not characterized by a Landau-type of order parameter, have moved to the vanguard of condensed matter research [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. A topological phase transition separates two insulating phases with different topological properties and is typically accompanied by a band inversion at a special point in quasimomentum space where two bands are directly coupled. Whereas standard metal-insulator transitions are described via localization theories [13, 14], a topological phase transition probes specific low-energy features and is characterized by a closing and reopening of a direct gap, accompanied by a change of a topological index. Independent of whether a topological index remains the same or not at such a transition, the fundamental question arises how to embed these special phase transition into conventional ones, where the fluctuations of an appropriate observable diverge at the transition, accompanied by the divergence of a characteristic length scale. Close to the transition such a diverging length scale is naturally given by  $\xi_g = v_F/E_g$ , where  $v_F$  is a typical velocity and  $E_g$  denotes the gap size. Going one step further this poses the interesting issue whether fluctuations reveal universal scaling laws as function of  $\xi_g$  (or, equivalently,  $E_g$ ).

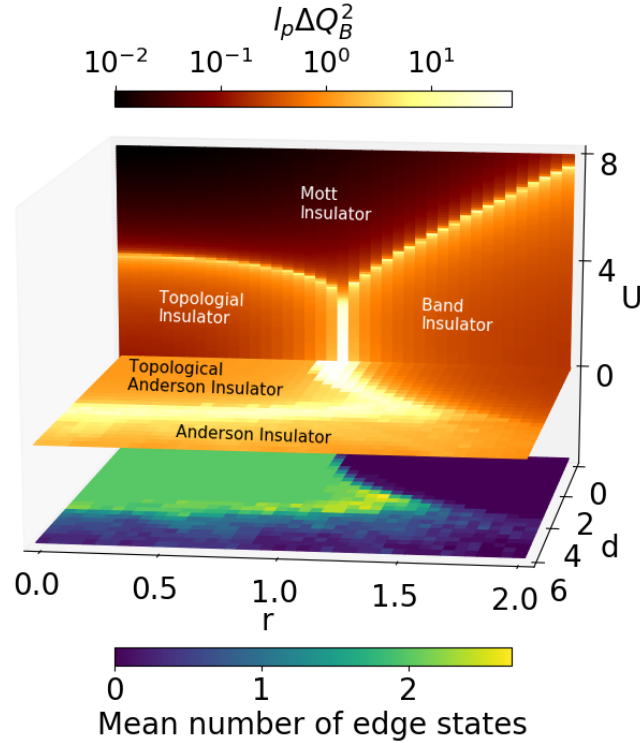


Figure 17: Topological phase diagram characterized by the fluctuations  $l_p \Delta Q_B^2$  (top) and by the number of zero-energy edge states (bottom) for the SSH model studied experimentally in [15].  $d$  is disorder strength and  $U$  denotes nearest-neighbor Coulomb interaction.  $r = t_1/t_2 < 1$  defines the topological region for  $d = U = 0$ . Phase boundaries between topological and non-topological band, Anderson, and Mott insulators are all well characterized by strongly enhanced fluctuations [16].

Recently it has been proposed that the boundary charge  $Q_B$  accumulated at a  $D - 1$ -dimensional flat surface of a  $D$ -dimensional insulator probes universal properties of topological insulators at low energies [17, 18, 19, 20, 21, 22]. Close to the transition point, it was demonstrated for one-dimensional, single-channel models that  $Q_B$  directly probes the phase of the gap parameter (in units of  $2\pi$ ) independent of the gap size, and reveals half-integer jumps at Weyl semimetal-like transitions [19, 20, 21]. Therefore, one expects strong fluctuations of  $Q_B$

at a (topological) phase transition and it is quite surprising that these fluctuations have so far not drawn much attention [17].

We remedy this substantial oversight in this letter and demonstrate that the fluctuations  $\Delta Q_B^2 = \langle \hat{Q}_B^2 \rangle - \langle \hat{Q}_B \rangle^2$  of the boundary charge themselves are the key to addressing universal properties of (topological) phase transitions. We identify a *universal regime*  $l_p \gg \xi_g \gg a$  where  $l_p \Delta Q_B^2(\xi_g)$  is a universal function of  $\xi_g$ , i.e., independent of the microscopic details of the charge measurement probe, described by a macroscopic length scale  $l_p$  on which the probe loses the contact to the sample (see below). Universality implies independence from high-energy properties, relevant on the scale of the lattice spacing  $a$ . In the regime close to the phase transition ( $\xi_g \rightarrow \infty$ ) we find that  $l_p \Delta Q_B^2$  diverges in one and two dimensions, quite analog to divergent fluctuations in conventional phase transitions. Therefore, we suggest the fluctuations of  $Q_B$  as a useful and measurable tool to probe the phase diagram of topological insulators. In Fig. 17 we begin by a compelling demonstration of the power of the suggested characterization focusing on the prototypical Su-Schrieffer-Heeger (SSH) model [23, 24] at half-filling including hopping disorder (experimentally studied in Ref. [15]) and nearest neighbor Coulomb interaction. Details of the model are postponed to Eq. (3.2), however, the general physics is dictated by the topological index  $r$  being the ratio of the two hopping amplitudes in the SSH model. Without disorder and interaction, the topologically non-trivial phase transitions to a trivial one at  $r = 1$ . Including disorder in the hoppings  $d$  a topological Anderson insulator is stabilized even beyond  $r = 1$  for weak disorder, while for strong disorder a trivial Anderson insulator is found. In the presence of strong enough Coulomb interaction  $U$  a Mott insulator is established (for an interpretation of the Mott transition as a topological one see Ref. [25]). All of the different phase boundaries between topological and non-topological band, Anderson, and Mott insulators are signalled by diverging boundary charge fluctuations  $l_p \Delta Q_B^2 \sim \xi_g$ . Our approach thus unifies transitions between all of these different classes of single-particle and correlated insulators. In addition, below, we find that  $l_p \Delta Q_B^2$  shows a universal scaling as function of  $E_g$  for a variety of models. We find a striking dependence on the dimensionality of the system which we exemplify for a massive Dirac-like low-energy spectrum. We report a typical scaling with the inverse gap in one dimension, logarithmic scaling in two-dimensional systems, and a monotonic increase of the fluctuations to a finite value at zero-gap in three dimensions.

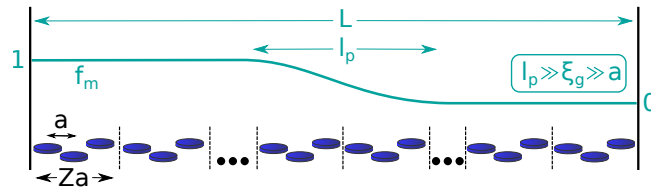


Figure 18: Pictorial representation of the system (bottom) and the definition of the envelope of the charge probe  $f_m$  (top). Here,  $l_p$  defines the length on which the probe smoothly loses contact to the sample (region where envelope varies smoothly from 1 to 0). We also define the lattice spacing  $a$  and the unit cell size  $Za$ . Implicitly, we assume that the fall-off of  $f_m$  fits into the system size  $L$  (the center of the fall-off is irrelevant). The universal regime is defined by  $l_p \gg \xi_g = v_F/E_g \gg a$ .

## 0.2 Model and Boundary charge fluctuations

We consider a generic and finite  $D$  dimensional tight-binding model with a  $D-1$  dimensional flat surface. In the direction perpendicular to the surface we consider  $N_s = L/a$  lattice sites labeled by  $m = 1, \dots, N_s$  with open boundary conditions, where  $L$  denotes the system size. We take an arbitrary extend and periodic boundary conditions in the remaining (transverse) directions. We then define an effective one-dimensional chain with two ends by absorbing the

remaining directional degrees of freedom into a (large) multi-channel character on each site defining  $N_B$  transverse channels labeled by  $\sigma = 1, \dots, N_B$  (which can additionally include, e.g., spin or orbital degrees of freedom as well). The size of the unit cell of the effective one-dimensional chain is denoted by  $Za$ , and  $x = ma$  defines the position of lattice site  $m$ ; see Fig. 18 (bottom). We consider zero temperature and fixed particle number  $N$ , and concentrate on the low-energy limit where the gap  $E_g$  is assumed to be small compared to the bandwidth or, equivalently,  $\xi_g = v_F/E_g \gg a$ . Generalizations are discussed in the SM [26]. We choose units  $\hbar = e = 1$ .

### 0.3 Boundary charge fluctuations

The boundary charge is a macroscopic observable measured on scales much larger than the microscopic scale  $\xi_g \gg a$ . We describe the macroscopic average by an envelope function  $f_m$  characteristic for a charge measurement probe, which falls off smoothly from unity to zero on the macroscopic length scale  $l_p \gg \xi_g$ , see Fig. 18 (top). The boundary charge operator at one end (determined by the envelope function  $f_m$  falling off from that end) of the system is defined [17, 20] by  $\hat{Q}_B = \sum_{m=1}^{N_s} f_m(\hat{\rho}_m - N/N_s)$ , where  $\hat{\rho}_m = \sum_{\sigma} a_{m\sigma}^\dagger a_{m\sigma}$  is the charge operator at site  $m$  summed over all  $N_B$  channels. The fluctuations  $\Delta Q_B^2 = \langle \hat{Q}_B^2 \rangle - \langle \hat{Q}_B \rangle^2$  can straightforwardly be expressed via the correlation function  $C_{mm'} = \langle \hat{\rho}_m \hat{\rho}_{m'} \rangle - \langle \hat{\rho}_m \rangle \langle \hat{\rho}_{m'} \rangle$  by exploiting the exact sum rule  $\sum_{m'=1}^{N_s} C_{mm'} = 0$ . We obtain  $\Delta Q_B^2 = -\frac{1}{2} \sum_{m,m'=1}^{N_s} (f_m - f_{m'})^2 C_{mm'}$ . Employing that  $C_{mm'}$  decays exponentially for  $|m - m'| \gg \xi_g$ , we find that the fluctuations are finite in the thermodynamic limit and the correlation function  $C_{mm'}$  can be replaced by the bulk correlation function  $C_{mm'}^{\text{bulk}} \equiv a^2 C_{\text{bulk}}(x, x')$  as we have  $l_p \gg \xi_g$ . Expanding  $f_m - f_{m'}$  up to first order in  $m - m'$  and averaging the correlation function over the unit cell [denoted by  $\bar{C}_{\text{bulk}}(x, x') = \bar{C}_{\text{bulk}}(x - x')$ ], we find in the universal limit  $l_p \gg \xi_g \gg a$  the result (see the SM [26] for details)

$$l_p \Delta Q_B^2 = -\frac{1}{2} \int dx x^2 \bar{C}_{\text{bulk}}(x) + O(\xi_g^2/l_p), \quad (3.1)$$

where we defined  $l_p^{-1} = \int dx f'(x)^2$  with  $f(ma) \equiv f_m$ . Our result requires only the condition that  $\bar{C}_{\text{bulk}}(x)$  decays exponentially for distances above the scale  $\xi_g$ . This is expected generically in the insulating regime due to the nearsightedness principle [27, 28]. Here,  $\xi_g$  should be considered as an upper limit for the decay length of  $\bar{C}_{\text{bulk}}(x)$ , in multi-channel or interacting models it is generically expected that  $\bar{C}_{\text{bulk}}(x)$  consists of a linear combination of many exponentially decaying terms with different length scales  $\xi_\sigma \lesssim \xi_g$  (see below the discussion of higher-dimensional systems).

We note that our central result (3.1) is independent of the scale  $l_p$  (besides the condition  $l_p \gg \xi_g$ ), offering a high degree of flexibility to measure and calculate the universality of boundary charge fluctuations. E.g., in cold atom systems, one can probe them either directly via the density profile or the correlation function [29]. Alternatively, for the special choice  $f_m = 1 - m/N_s$  (where  $l_p = L$ ), we get  $\hat{Q}_B = -\hat{P}/L$  with  $\hat{P} = a \sum_{m=1}^{N_s} m(\hat{\rho}_m - N/N_s)$  denoting the bulk polarization operator, the fluctuations of which are also discussed within localization theories [30]. Our result  $l_p \Delta Q_B^2 = \Delta P^2/L$  provides a *surface fluctuation theorem* connecting boundary and bulk fluctuations in a universal way. This can be viewed as the fluctuation-based analog of the celebrated surface charge theorem [31, 32, 19, 20].

### 0.4 Single-channel case

A theory as general as the one outlined above can be put to the test in a plethora of applications. We start with the most simple single-channel case  $N_B = 1$  and nearest-neighbor hoppings, where numerically exact results in the clean or disordered case (by diagonalization



of the single particle problem) as well as in the presence of interactions can be obtained with relative ease (in the interacting case by use of density matrix renormalization group approaches). We consider the following model

$$H = - \sum_{m=1}^{N_s-1} (t_m + w_m) (a_{m+1}^\dagger a_m + \text{h.c.}) + \sum_{m=1}^{N_s} v_m \rho_m + U \sum_{m=1}^{N_s} (\rho_m - 1/2) (\rho_{m+1} - 1/2), \quad (3.2)$$

where  $t_m = t_{m+Z}$  and  $v_m = v_{m+Z}$  are periodically modulated nearest-neighbor hoppings and on-site potentials, respectively,  $w_m$  describes bond disorder drawn from a uniform distribution  $w_m \in [-d_m/2, d_m/2]$  with  $d_m = d_{m+Z}$ , and  $U \geq 0$  is a nearest-neighbor repulsive interaction.

The phase diagram of this model in the SSH limit [23, 24] at half-filling (choosing  $Z = 2$  and  $v_m = 0$ ), and its characterization in terms of the boundary charge fluctuations were already discussed above; see Fig. 17. Varying the interaction strength  $U$  as well as hopping disorder  $d = d_1 = 2d_2$  gap closings indicated by strongly enhanced boundary charge fluctuations are found. In the  $(r, d)$ -plane at finite disorder and  $U = 0$  we show that our characterization in terms of the boundary charge fluctuations is perfectly consistent with the number of edge states, thus demonstrating perfect agreement with the theoretical [33] and experimental [15] findings. At finite  $U$  the transition to the correlated Mott insulator is more involved and classification schemes are rare. The Mott insulator is characterized by a charge density wave instability due to Umklapp processes [34] generating a staggered on-site potential. This potential breaks the chiral symmetry of the SSH model and leads to a non-topological phase. The boundary charge fluctuations provide a valuable tool to find also this transition line; compare Fig. 17. From exact solutions one point of this transition line into the Mott insulator is known to be at  $r = 1$  ( $t_1 = t_2$ ),  $U/t_1 = 2$ , which is in perfect agreement with the boundary charge fluctuation characterization.

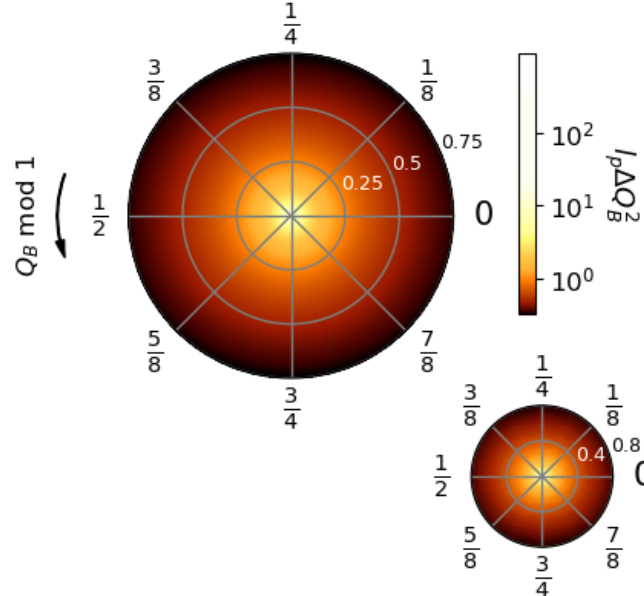


Figure 19: Polar color plot of  $l_p \Delta Q_B^2$  as function of  $Q_B \bmod(1)$  (polar component) and the gap  $E_g$  (radial component) for model (3.2) with  $Z = 2$ ,  $a = 1$ ,  $N/N_s = 1/2$ ,  $t = (t_1 + t_2)/2 = 1$ ,  $d = 0$ , and  $U = 0$  (upper panel) or  $U = 0.5$  (lower panel). The data points are obtained by taking  $t_{1/2} = t \pm \Delta_0/2 \cos(\phi)$ ,  $v_{1/2} = \pm \Delta_0 \sin(\phi)$ , and varying the parameters  $\Delta_0$  and  $\phi$  in the intervals  $\Delta_0 \in [0, 0.375]$  and  $\phi \in [0, 2\pi]$ . For  $U = 0$  the analytic result ( $N_s, l_p \rightarrow \infty$ ) is used, while for  $U = 0.5$  we set  $N_s = 1000$ , and  $l_p = 400$ .



Whereas the boundary charge fluctuations depend strongly on the gap size, the boundary charge  $Q_B$  itself is sensitive to the phase of the gap parameter in one-dimensional, single-channel models [21]. This suggests the polar plot of Fig. 19, where we show the fluctuations  $l_p \Delta Q_B^2$  in dependence of the gap  $E_g$  (radial component) and the boundary charge  $Q_B \bmod(1)$  (polar component) for the model of Eq. (3.2) with  $Z = 2$  and  $d = 0$ , both for  $U = 0$  and finite  $U$ , choosing a variety of parameters to define the staggered on-site potentials  $v_1 = -v_2$  and hoppings  $t_{1/2}$ . This corresponds to the noninteracting and interacting Rice-Mele model [35, 22]. The radially symmetric value of  $l_p \Delta Q_B^2$  indicates that the fluctuations depend only on the gap's absolute value but not on  $Q_B$  and that they strongly enhance at the center  $E_g \rightarrow 0$ . We expect this feature to be generic for one-dimensional, single-channel models in the low-energy regime.

For the noninteracting and clean Rice-Mele model  $Z = 2$  we find analytically the exact result (see the SM for details [26])  $l_p \Delta Q_B^2 = a(t_1^2 + t_2^2)/(4E_g \sqrt{E_g^2/4 + 4t_1 t_2})$ , with the gap  $E_g = 2\sqrt{v^2 + (t_1 - t_2)^2}$  and  $v = v_1 = -v_2$ . In the vicinity of phase transitions  $E_g \ll t = (t_1 + t_2)/2$ , we obtain the universal scaling

$$l_p \Delta Q_B^2(E_g) \xrightarrow{E_g \ll t} \frac{v_F}{8E_g} = \frac{\xi_g}{8}, \quad (3.3)$$

where  $v_F = 2ta$  denotes the Fermi velocity. For arbitrary value of  $Z$  and generic modulations of the nearest-neighbor hoppings and the on-site potentials we confirm this universal scaling for the chemical potential located in any gap. We obtain this result by using the exact eigenstates of a low-energy massive Dirac model in  $1 + 1$  dimensions, as proposed in Ref. [21] (see the SM [26] for details). Furthermore, we find that  $\xi_g = v_F/E_g$  is the exponential decay length of the correlation function  $\bar{C}_{\text{bulk}}(x)$ . For more exotic models which can not be described by a Dirac model in the low-energy regime, we show in the SM [26] that also other scalings are in principle possible. For the noninteracting and clean SSH model, we numerically confirm the scaling  $l_p \Delta Q_B^2 = v_F/(8E_g)$  in Fig. 20 a) and find that it holds up to surprisingly large gaps even beyond the applicability range of the low-energy theory. This relation holds also in the presence of disorder, at least for not too strong disorder  $d \lesssim 2 - 3$ , where a Born approximation [36] can be used to define a renormalized gap  $E_g = 2|\bar{t}_1 - \bar{t}_2|$ , with  $\bar{t}_1 = t_1 - \theta(t_1 - t_2)d_1^2/(12t_1)$  and  $\bar{t}_2 = t_2 - \theta(t_2 - t_1)d_2^2/(12t_2)$  denoting renormalized hoppings (see the SM for details [26]). In Fig. 20 b) we show the scaling for different  $U$  and  $d = 0$  and find that they collapse to the universal  $\sim 1/E_g$  if one allows for a  $U$  dependent non-universal prefactor [26]. In this case the gap is significantly increased by interactions [37, 38, 39, 21, 22].

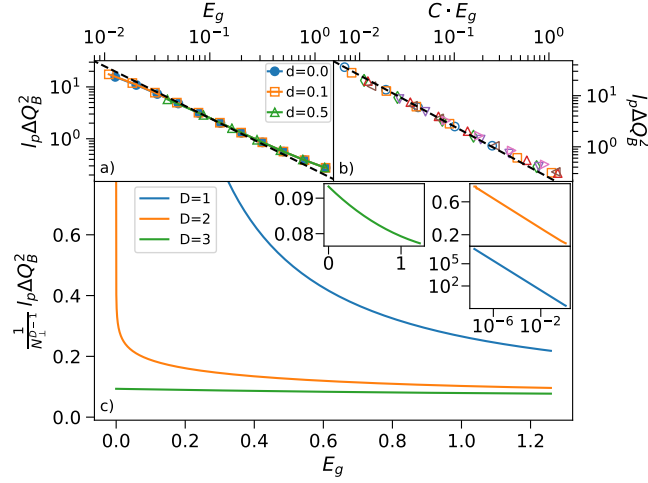


Figure 20: (a-b) Scaling of  $l_p \Delta Q_B^2$  with the gap  $E_g$  for model (3.2) on double-logarithmic scale for (a)  $U = 0$  and varying disorder strength  $d$  (averaged over 20 disorder configurations) and (b)  $d = 0$  and varying Coulomb interaction strength  $U$  (other parameters as in Fig. 17). The dashed line indicates  $v_F/(8E_g) = 1/(4E_g)$ . In (b)  $U = (0.0, 0.4, 0.8, 1.2, 1.6, 2.0, 2.4)$  are given by (blue, orange, green, red, purple, brown, pink) symbols. The non-universal prefactors  $C(U)$  for the collapse are given in the SM [26]. In panels (a) and (b) we choose  $N_s = 1000$  and  $l_p = 400$ . (c) Scaling of  $l_p \Delta Q_B^2 / N_\perp^{D-1}$  from the analytic result ( $N_s, l_p \rightarrow \infty$ ) with the gap  $E_g = 2|t_1 - t_2|$  for various dimensions  $D$  for the SSH model ( $D = 1$ ) and for the higher-dimensional models discussed in the main text ( $D = 2, 3$ ) with  $t = (t_1 + t_2)/2 = t_y = t_z = 1$ . The insets show the same results on different scales.

### 0.5 Two- and Three-Dimensional Systems

To analyze the scaling in two and three dimensions  $D = 2, 3$  for noninteracting and clean systems we use translational invariance in the transverse direction and consider  $N_B = N_\perp^D$  transverse quasimomenta  $\underline{k}_\perp$  as a channel index. The fluctuations of  $Q_B$  can then be calculated as an independent sum  $\Delta Q_B^2 = \sum_{\underline{k}_\perp} \Delta Q_B^2(\underline{k}_\perp)$ , with  $\underline{k}_\perp = k_y$  for  $D = 2$  and  $\underline{k}_\perp = (k_y, k_z)$  for  $D = 3$ . For each fixed  $\underline{k}_\perp$ , we consider an effective one-dimensional, single-channel system and get from Eq. (3.3) in the low-energy regime  $l_p \Delta Q_B^2(\underline{k}_\perp) = v_F(\underline{k}_\perp)/(8E_g(\underline{k}_\perp))$ , corresponding via Eq. (3.1) to an independent term of  $\bar{C}_{\text{bulk}}(x)$  decaying on length scale  $\xi(\underline{k}_\perp) = v_F(\underline{k}_\perp)/E_g(\underline{k}_\perp) \leq \xi_g$ . The momentum dependence of the effective gap  $E_g(\underline{k}_\perp)$  can be estimated for a typical massive Dirac-like spectrum in  $D + 1$  dimensions:  $E_g(\underline{k}_\perp) \approx 2\sqrt{\bar{v}_F^2 k_\perp^2 + E_g^2/4}$ , where  $E_g$  denotes the spectral gap, and we have neglected the weak momentum dependence of  $v_F(\underline{k}_\perp) \approx \bar{v}_F$ . In the thermodynamic limit  $\sum_{\underline{k}_\perp} \rightarrow (N_\perp a/(2\pi))^{D-1} \int_{-\pi/a}^{\pi/a} d^{D-1} k_\perp$ , we can estimate the scaling of the fluctuations. In two dimensions, we obtain a logarithmic scaling  $N_\perp^{-1} l_p \Delta Q_B^2 \sim (\bar{v}_F/W) \ln(W/E_g) \sim a \ln(\xi_g/a)$ , where  $W$  defines a high-energy cutoff scale for  $|\bar{v}_F k_y|$ . In contrast, for three dimensions we obtain a monotonic increase for the fluctuations with decreasing gap but a finite value in the zero-gap limit.

Systems illustrating this generic behavior can be realized, e.g., in cold atom systems [40]. As an example we consider an SSH model in  $x$ -direction (with alternating hoppings  $t_{1,2}$ ), constant nearest-neighbor hoppings in transverse direction (denoted by  $t_{y,z}$ ), and a homogeneous magnetic field of size  $B$  in  $z$ -direction (for  $D = 2$ ) or in the  $y$ - and  $z$ -direction (for  $D = 3$ ). For the simplest case that the magnetic length is given by  $\lambda_B = 2a$ , we obtain in the Landau gauge an effective one-dimensional Rice-Mele model with  $E_g(\underline{k}_\perp) = 2\sqrt{v(\underline{k}_\perp)^2 + (t_1 - t_2)^2}$ , where  $v(\underline{k}_\perp) = 2t_y \cos(k_y a)$  in  $D = 2$  or  $v(\underline{k}_\perp) = 2t_y \cos(k_y a) + 2t_z \cos(k_z a)$  in  $D = 3$  (see the SM [26] for details). Using the exact result for the Rice-Mele model to calculate  $l_p \Delta Q_B^2(\underline{k}_\perp)$ ,

one can perform the integral over  $k_{\perp}$ , and finds for the fluctuations as function of the gap  $E_g = 2|t_1 - t_2|$  the result shown in Fig. 20 c). The logarithmic scaling in  $D = 2$  is perfectly preserved even for large gaps, suggesting the boundary charge fluctuations to be also useful as an indicator for phase transitions in two dimensions. We emphasize that for  $D = 3$ , we observe only a weak increase of the fluctuations with a finite value at zero gap. Therefore, for three dimensions, the fluctuations are only a weak indicator for the transition. Although this limits the universally diverging behavior to one- and two-dimensional systems, future studies should address whether also the non-divergent enhancement carries valuable information of the type of transition passed. In addition, it will of interest to study how generic the proposed decoupling in transverse modes will persist in higher-dimensional interacting and disordered systems.

## 0.6 Conclusion

We have established the boundary charge and its fluctuations as a measurable tool to probe topological properties of insulators. Whereas the boundary charge takes the role of a phase and jumps by  $e/2$  at a topological phase transition [21], the complementary fluctuations are strongly enhanced in one and two dimensions and reveal a universal scaling as function of the gap size. In contrast to the number of topological edge states which is controlled by non-fluctuating topological indices, we found that the universal scaling properties of the fluctuations do not depend on whether a topological index changes at the transition but rely exclusively on the characteristic band structure of insulators with topological properties. Importantly, this characterization scheme can be applied to band, Anderson, and Mott insulators alike. An intriguing avenue of future research concerns the question whether the characterization proposed here is also useful in the context of topological superconductors, for which simple models do not fulfill charge conservation.

# 1. Appendixes

## 1.1 Numerical results of Fig.4b

$U$	0.0	0.4	0.8	1.2	1.6	2.0	2.4
$C$	1.0	0.92	0.88	0.96	0.89	0.96	1.18

Constants  $C$  used in the collapse of Fig. 4 (b) of the main text.

To achieve the universal collapse shown in Fig. 4 (b) of the main text we have numerically determined the constants  $C$  listed in Table 1.1.

## 1.2 Boundary charge fluctuations

Here we present a general analysis of the boundary charge fluctuations without assuming that the gap is small compared to the band width. This means that the length scale  $\xi_g = v_F/E_g$  is not assumed to be much larger than the lattice spacing  $a$ . As in the main part of the letter we assume fixed particle number  $N$  (i.e., a canonical ensemble) and zero temperature here (other cases are discussed in Section 1.2). We take an effectively one-dimensional system with  $N_s$  lattice sites and  $N_B$  channels per site, labeled by  $m = 1, \dots, N_s$  and  $\sigma = 1, \dots, N_B$ , respectively. Besides different spins, orbitals and other flavors, the channel indices include also the degrees of freedom in transverse direction, i.e., parallel to the surface. The size of the one-dimensional unit cell is denoted by  $Za$ . For the real space position of a lattice site we write  $x = ma$ , and  $L = N_s a$  defines the system size perpendicular to the boundary. Finally, we take the thermodynamic limit  $N_s, N \rightarrow \infty$  such that the average charge per site  $\bar{\rho} = N/N_s$  is kept constant. We use units  $\hbar = e = 1$ .

In Section 1.2 we analyse the relation of the boundary charge fluctuations to the second moment of the longitudinal density-density correlation function. In Section 1.2 we discuss the case of a grandcanonical ensemble and finite temperature.

### Relation to the density-density correlation function

Our aim is to calculate the fluctuations of the boundary charge operator at one end of the system defined by

$$\hat{Q}_B = \sum_{m=1}^{N_s} f_m (\hat{\rho}_m - \bar{\rho}), \quad (3.4)$$

where  $\hat{\rho}_m = \sum_{\sigma} a_{m\sigma}^{\dagger} a_{m\sigma}$  is the charge operator at site  $m$ , and  $a_{m\sigma}^{\dagger}$  creates a fermion on site  $m$  in channel  $\sigma$ . The macroscopic average is described by an envelope function  $f_m \equiv f(x) = 1 - \theta_{l_p}(x - L_p)$ , with  $L \gtrsim L_p + l_p/2 \gtrsim l_p \gg \xi_g$ , where  $\theta_{\delta x}(x)$  is some representation of the  $\theta$ -function broadened with  $\delta x$ . The scale  $L_p$  describes the length of a charge measurement probe and  $l_p$  is the scale on which the probe smoothly loses the contact to the sample. By convention, we define the scale  $l_p$  by the integral

$$l_p^{-1} = \int dx [f'(x)]^2. \quad (3.5)$$

The envelope function is assumed to be smooth on the microscopic length scales  $\xi_g$  and  $a$ , i.e.,  $l_p \gg \xi_g, a$ . As shown below the length scales  $L$ ,  $L_p$ , and  $l_p$  can even be of the same order of magnitude, provided that  $|L - L_p - l_p/2| \gtrsim O(\xi_g)$  and  $|L_p - l_p/2| \gtrsim O(\xi_g)$ . This condition means that the fall-off of the envelope function fits into the system size (up to  $O(\xi_g)$ ). Otherwise, as we will see below, the length scale  $L_p$  does not enter the final solution for the boundary charge fluctuations.

Defining the correlation function

$$C_{mm'} = \langle \hat{\rho}_m \hat{\rho}_{m'} \rangle - \langle \hat{\rho}_m \rangle \langle \hat{\rho}_{m'} \rangle, \quad (3.6)$$

we can express the fluctuations  $\Delta Q_B^2 = \langle \hat{Q}_B^2 \rangle - \langle \hat{Q}_B \rangle^2$  as

$$\Delta Q_B^2 = \sum_{m,m'=1}^{N_s} f_m f_{m'} C_{mm'}. \quad (3.7)$$

In the following we call  $C_{mm'}$  the density-density correlation function of the effectively one-dimensional system but it should be kept in mind that it is the correlation between the charges  $\rho_m$  and  $\rho_{m'}$  including the sum over *all* channel indices. In particular, this includes also the sum over all transverse quasimomenta. Therefore, for higher-dimensional systems, it describes rather the correlation between the total charge of two stripes and not the correlation between the densities at two points in real space.

Using the sum rule (which is exact at fixed particle number)

$$\sum_{m'=1}^{N_s} C_{mm'} = 0, \quad (3.8)$$

together with  $C_{mm'} = C_{m'm}$ , we can replace  $f_m f_{m'} \rightarrow -(1/2)(f_m - f_{m'})^2$  in (3.7) and obtain the very useful form

$$\Delta Q_B^2 = -\frac{1}{2} \sum_{m,m'=1}^{N_s} (f_m - f_{m'})^2 C_{mm'}. \quad (3.9)$$

This formula is very helpful since the correlation function  $C_{mm'}$  is exponentially small for an insulator for  $|x - x'| \gg \xi_g$ , with  $x = ma$  and  $x' = m'a$ . Therefore, only the part  $|x - x'| \lesssim \xi_g$  is relevant. Since  $f_m \equiv f(x)$  varies slowly on the scale  $\xi_g$  due to  $\xi_g \ll l_p$ , we can expand  $f_m - f_{m'} \approx f'(\frac{x+x'}{2})(x - x')$  in (3.9). Using in addition that the derivative  $f'(x) \sim 1/l_p$  is only non-zero for  $x = L_p + O(l_p/2)$ , we find that both  $x, x' \sim L_p + O(l_p/2) \gg \xi_g$  are located away from the boundary far beyond the length scale  $\xi_g$ . This holds even in the case when  $L_p$  and  $l_p/2$  have the same order of magnitude (up to  $O(\xi_g)$ ) since the contribution from all  $x, x' \sim \xi_g$  gives a contribution  $\lesssim (\frac{\xi_g}{l_p})^2$  to the fluctuations. Therefore, we can replace the correlation function  $C_{mm'}$  by its bulk value

$$C_{mm'}^{\text{bulk}} = a^2 C_{\text{bulk}}(x, x'), \quad (3.10)$$

and obtain for  $l_p \gg \xi_g$

$$\begin{aligned} \Delta Q_B^2 &= -\frac{1}{2}a^2 \sum_{m,m'=1}^{N_s} \left[ f'(\frac{x+x'}{2}) \right]^2 \\ &\quad \times (x - x')^2 C_{\text{bulk}}(x, x') + O(\frac{\xi_g}{l_p})^2. \end{aligned} \quad (3.11)$$

As a result, the double sum scales as  $a^2 \sum_{m,m'} \lesssim \xi_g l_p$  and one can see that the fluctuations are finite in the thermodynamic limit. In this formula the bulk correlation function can be calculated in the thermodynamic limit for any ensemble and for any boundary condition. Note that this is not possible in the form (3.7) since the sum rule (3.8) does no longer hold exactly in a grandcanonical ensemble at finite temperature (see the discussion in Section 1.2).

Using  $m = Z(n - 1) + j$ , where  $j = 1, \dots, Z$  denotes the site index within a unit cell and  $n = 1, 2, \dots$  labels the unit cells, we can write  $C_{\text{bulk}}(x, x') = C_{jj'}^{\text{bulk}}(Zna, Zn'a)$ . Due to translational invariance on the size  $Za$  of a unit cell (which holds on average also in the presence of random disorder) the bulk correlation function depends only on the difference  $n - n'$  and we can write

$$C_{\text{bulk}}(x, x') = C_{jj'}^{\text{bulk}}(Z(n - n')a). \quad (3.12)$$

Using  $f'(\frac{x+x'}{2}) = f'(Zn_s a)(1 + O(a/l_p))$ , with  $n_s = \frac{n+n'}{2}$ , we find that the sum over  $n_s$  in (3.11) gives  $\sum_{n_s} f'(Zn_s a)^2 = \frac{1}{Zal_p}(1 + O(Za/l_p))$  according to (3.5). Again neglecting boundary effects  $\sim O(\frac{\xi_g}{l_p})^2$ , we obtain

$$\begin{aligned} l_p \Delta Q_B^2 &= -\frac{a}{2Z} \sum_n \sum_{jj'} \\ &\quad \times (Zna + (j - j')a)^2 C_{jj'}^{\text{bulk}}(Zna). \end{aligned} \quad (3.13)$$

We note that this result is exact when performing the limits  $L, L_p, l_p \rightarrow \infty$  with  $L \gtrsim L_p + l_p/2 \gtrsim l_p$ . Eq. (3.13) establishes a universal relation of the boundary charge fluctuations to the density-density correlation function of the bulk, including all microscopic details of the unit cell. As one can see only the product  $l_p \Delta Q_B^2$  is related in a universal way to the bulk correlation function  $C_{jj'}^{\text{bulk}}(Zna)$ , where  $l_p$  is defined by Eq. (3.5) in terms of the envelope function. Since the bulk correlation function can be calculated in any ensemble and with any boundary condition this equation is the most convenient starting point to calculate the boundary charge fluctuations very efficiently from bulk quantities without any need to determine the complicated eigenfunctions of finite or half-infinite systems.

As one can see from the proof only the condition  $l_p \gg \xi_g$  enters together with the property that the support of the function  $f'(x)$  fits into the system size such that

$$-\int_0^L dx f'(x) = 1 + O(\xi_g/l_p), \quad (3.14)$$

$$\int_0^L dx [f'(x)]^2 = l_p^{-1} \left(1 + O(\xi_g/l_p)\right). \quad (3.15)$$

This is the reason why the length scales  $L$ ,  $L_p$ ,  $l_p$  can all be of the same order of magnitude (in the sense defined above) without changing the leading order contribution of the fluctuations. This is very helpful for numerical calculations and the experimental observability since only the condition  $l_p \gg \xi_g \gg a$  has to be fulfilled to find universal properties of the boundary charge fluctuations.

To proceed we note at this point that all universal properties of  $\Delta Q_B^2$  derived in this section follow from two fundamental properties of  $C_{jj'}^{\text{bulk}}(x)$  which have to be checked for the concrete model under consideration

$$C_{jj'}^{\text{bulk}}(x) = \sum_{\sigma} \frac{1}{\xi_{\sigma}^2} g_{jj'}^{(\sigma)}(x/\xi_{\sigma}) e^{-x/\xi_{\sigma}}, \quad (3.16)$$

$$s^2 g_{jj'}^{(\sigma)}(s) \sim O(1) \quad \text{for } |s| < 1. \quad (3.17)$$

The first condition (3.16) states exponential decay and indicates that the correlation function is in general a linear combination of many terms, each with its own decay length  $\xi_{\sigma} < \xi_g$ . This occurs generically in the presence of channel indices describing flavor degrees of freedom (spin, orbital, etc.) or the transverse quasimomentum (see Section 1.5). Due to the second condition (3.17) the pre-exponential function should scale as  $g_{jj'}^{(\sigma)}(s) \sim 1/s^2$  for  $|s| < 1$ . We find this scaling independent of the microscopic details of the model. For  $\xi_g \sim a$  this property is obvious since there is only a single length scale. In the low-energy regime  $\xi_g \gg a$ , we show in Section 1.3 explicitly that the two properties are fulfilled for single-channel and noninteracting models. The physical reason for the general case is obvious. For  $|x| \sim a \ll \xi_{\sigma} < \xi_g$  one probes high-energy scales where the gap is unimportant. Therefore, the correlation function will scale  $\sim 1/a^2$ . For  $|x| \sim \xi_{\sigma} \gg a$ , the lattice spacing does not play any role and a low-energy continuum theory is possible to describe the corresponding term of the correlation function. This theory is expected to depend mainly on a single length scale  $\xi_{\sigma}$ , such that the corresponding term of the correlation function will scale  $\sim 1/\xi_{\sigma}^2$ . In contrast, for  $|x| \gg \xi_{\sigma}$  or  $|s| \gg 1$ , the scaling depends crucially on the low-energy properties of the model and a pre-exponential power-law with an interaction dependent exponent is expected. The latter is difficult to determine for interacting systems. For clean single-channel systems one obtains  $1/|s|$  in the noninteracting case, see Section 1.3. However, this regime is of no relevance for the fluctuations since the corresponding term of the correlation function is exponentially small for  $|x| \gg \xi_{\sigma}$ . We note that the exponential decay property (3.16) is also valid for the correlation function  $C_{mm'}$  with a boundary but the scaling of the pre-exponential function for  $n, n'$  close to the boundary might be more subtle.

Using (3.16) and (3.17) one can estimate the order of magnitude of the fluctuations (3.13) as

$$l_p \Delta Q_B^2 = \sum_{\sigma} c_{\sigma} \xi_{\sigma} \equiv N_B \bar{\xi} \lesssim N_B \xi_g, \quad (3.18)$$

with  $c_{\sigma} \sim O(1)$ , and  $\bar{\xi} = \frac{1}{N_B} \sum_{\sigma} c_{\sigma} \xi_{\sigma} \leq \xi_g$  defining some average exponential decay length. This result shows that the boundary charge fluctuations  $\Delta Q_B \lesssim \sqrt{N_B \xi_g / l_p} \ll \sqrt{N_B}$  are always much



smaller than the boundary charge  $Q_B \sim N_B$ , even for  $N_B \sim O(1)$ , showing that the boundary charge is a well-defined observable for  $l_p \gg \xi_g$  [17], in analogy to interface charges studied in Refs. [[41, 42, 43, 44, 45, 46]].

In the low-energy regime when  $\xi_g \gg Za$ , one can neglect all terms in (3.18) with  $\xi_\sigma \sim Za \ll \xi_g$ . For the terms with  $\xi_\sigma \gg Za$ , we can neglect the part  $(j - j')a \sim Za \ll \xi_\sigma \sim Zna$  in (3.11) and the sum can be replaced by an integral. This leads to the compact formula

$$l_p \Delta Q_B^2 = -\frac{1}{2} \int dx x^2 \bar{C}_{\text{bulk}}(x), \quad (3.19)$$

where  $\bar{C}_{\text{bulk}}(x)$  is the correlation function averaged over  $j$  and  $j'$

$$\bar{C}_{\text{bulk}}(x) = \frac{1}{Z^2} \sum_{jj'} C_{jj'}^{\text{bulk}}(x). \quad (3.20)$$

### Numerics in the infinite system size limit $N_s \rightarrow \infty$

One of the many important implications of the previous subsection is that the boundary charge fluctuations  $l_p \Delta Q_B^2$  can be extracted either directly or equivalently (in the  $L_p, l_p \rightarrow \infty$  limit) from the right hand side of Eq. (3.13) via the bulk correlation functions. In the translationally invariant case this can be utilized to significantly speed up the numerical determination of  $l_p \Delta Q_B^2$  as one can work directly in the desired limit of  $N_s \rightarrow \infty$ , which is often more convenient. We use this in Fig. 1 of the main text to determine the non-zero  $U$ , but  $d = 0$  data from a highly efficient infinite system size density matrix renormalization group approach. With this the problem is directly phrased in the correct limit of  $N_s, L_p, l_p \rightarrow \infty$ , with  $N_s a \gtrsim L_p + l_p/2 \gtrsim l_p \gg \xi_g$ , but the evaluation of Eq. (3.13) requires the evaluation of an infinite sum  $\sum_n$ . As usual in numerical approaches we truncate this infinite sum at finite, but large index by  $\sum_{n=-\infty}^{\infty} \rightarrow \sum_{n=-n_c}^{n_c}$ , where  $n_c$  needs to be converged to  $n_c a \gg \xi_g$  (akin, but not exactly equal to keeping a finite  $l_p$ ). We choose  $n_c = 499$  for the data shown in Fig. 1 of the main text.

### Finite temperature and grandcanonical ensemble

For a canonical ensemble at fixed particle number, all our results are also valid at finite temperature  $T$ , provided that  $T \ll \Delta$  is much smaller than the gap  $E_g = 2\Delta$ . In this case the finite temperature corrections to  $\Delta Q_B^2$  can be shown to be exponentially small of relative order  $\sim \sqrt{T/\Delta} e^{-\Delta/T}$ , see Section 1.3. This changes for a grandcanonical ensemble where the sum rule (3.8) is no longer fulfilled exactly. For an estimation we consider a non-interacting system and find after a straightforward calculation using Wick's theorem

$$\begin{aligned} \sum_{m'=1}^{N_s} C_{mm'} &= \\ &= \sum_s n_F(\epsilon_s) [1 - n_F(\epsilon_s)] \sum_\sigma |\psi_s(m\sigma)|^2 \\ &\sim N_B \frac{T}{W} e^{-\Delta/T}, \end{aligned} \quad (3.21)$$

where  $\psi_s(m\sigma)$  are the single-particle eigenstates with energy  $\epsilon_s$ ,  $n_F(\epsilon_s)$  denotes the Fermi function, and  $W$  is the band width. As a result we find that in the steps to get Eq. (3.9) from (3.7), the contributions from  $f_m^2$  and  $f_{m'}^2$  lead to corrections of order

$$\Delta Q_B^2(T) - \Delta Q_B^2(T=0) \sim N_B \frac{L_p}{a} \frac{T}{W} e^{-\Delta/T}. \quad (3.22)$$



With  $\Delta Q_B^2(T=0) \sim N_B \frac{\bar{\xi}}{l_p}$ , and the thermal length  $L_T \sim v_F/T \sim aW/T$  we conclude that the temperature dependent correction is of relative order

$$\frac{\Delta Q_B^2(T) - \Delta Q_B^2(T=0)}{\Delta Q_B^2(T=0)} \sim \frac{L_p l_p}{\bar{\xi}^2} \frac{\bar{\xi}}{L_T} e^{-\Delta/T}. \quad (3.23)$$

However, even though  $L_p l_p \gg \xi_g^2 > \bar{\xi}^2$ , we expect these corrections to be very small at low temperatures  $\bar{\xi} < \xi_g \ll L_T$  due to the exponentially small factor  $e^{-\Delta/T}$ . As a consequence we conclude that all our central results remain valid for a grandcanonical ensemble as well.

### 1.3 Noninteracting and clean models

In this section we analyse the special case of noninteracting and clean systems. In Section 1.3 we provide general reasons for the properties (3.16) and (3.17) by expressing the correlation function via the propagator. For the special case of the Rice-Mele model we present the exact solution for the fluctuations of the boundary charge in Section 1.3. Finally, in Section 1.3 we present the generic low-energy theory for all single channel models in the low-energy regime in terms of a noninteracting Dirac model following the ideas of Ref. [21].

#### Properties of the density-density correlation function and the propagator

Here we provide qualitative reasons why the properties (3.16) and (3.17) are fulfilled. Due to translational invariance perpendicular to the effectively one-dimensional system (i.e., parallel to the boundary), we consider in the following a fixed value for the transverse quasimomentum, restricting the sum over the channels only to a finite and small number  $N_c$  of other flavor degree of freedom.

For any noninteracting and clean lattice model in a grandcanonical ensemble one can use Wick's theorem and obtains for the density-density correlation function

$$C_{mm'}^{\text{bulk}} = \delta_{mm'} \rho_m - \sum_{\sigma, \sigma'=1}^{N_c} |\langle a_{m\sigma}^\dagger a_{m'\sigma'} \rangle|^2. \quad (3.24)$$

The propagator  $\langle a_{m\sigma}^\dagger a_{m'\sigma'} \rangle$  can be expressed via the single-particle Bloch eigenfunctions

$$\psi_k^{(\alpha)}(m, \sigma) = \frac{1}{\sqrt{2\pi}} \chi_k^{(\alpha)}(j, \sigma) e^{ikn}, \quad (3.25)$$

with energy  $\epsilon_k^{(\alpha)}$ , where  $\alpha$  is the band index and  $-\pi < k < \pi$  denotes the quasimomentum in units of the inverse lattice spacing. With  $a_{m\sigma}^\dagger = \sum_{\alpha} \int_{-\pi}^{\pi} dk c_{k\alpha}^\dagger \psi_k^{(\alpha)}(m, \sigma)$  and  $\langle c_{k\alpha}^\dagger c_{k'\alpha'} \rangle = \delta_{\alpha\alpha'} \delta(k - k') n_F(\epsilon_k^{(\alpha)})$ , we obtain

$$\begin{aligned} \langle a_{m\sigma}^\dagger a_{m'\sigma'} \rangle &= \sum_{\alpha} \int_{-\pi}^{\pi} dk \\ &\times \psi_k^{(\alpha)}(m, \sigma)^* \psi_k^{(\alpha)}(m', \sigma') n_F(\epsilon_k^{(\alpha)}), \end{aligned} \quad (3.26)$$

which, for the case of small temperatures  $T \ll \Delta$  and by inserting (3.25), can be written as

$$\begin{aligned} \langle a_{m\sigma}^\dagger a_{m'\sigma'} \rangle &= \sum_{\alpha=1}^{\nu} \int_{-\pi}^{\pi} \frac{dk}{2\pi} \\ &\times \chi_k^{(\alpha)}(j, \sigma)^* \chi_k^{(\alpha)}(j', \sigma') e^{-ik(n-n')}. \end{aligned} \quad (3.27)$$

Here  $\sum_{\alpha=1}^{\nu}$  denotes the sum over the occupied bands. The Bloch vectors  $\chi_k^{(\alpha)}$  depend parametrically on  $e^{ik}$  and on the dispersion  $\epsilon_k^{(\alpha)}$ . The latter has a branching point in the complex plane with an imaginary part  $\text{Im}(k) \sim a/\xi_c$  [47, 48, 49, 20], where  $\xi_c$  is some length scale averaged over all channel indices. Closing the integration contour in the complex plane this gives the exponential decay of the propagator  $\sim e^{-Z|n-n'|a/\xi_c}$ , leading via (3.24) to the corresponding exponential decay (3.16) of the correlation function. For scales  $|n - n'|a \sim \xi_c$  and  $\xi_c \gg a$  we get  $k \sim |n - n'|^{-1} \sim a/\xi_c \ll 1$  such that all  $e^{ik} \approx 1$  and  $\epsilon_k^{(\alpha)} \sim \Delta$ . Therefore, the integral  $\int dk \sim a/\xi_c$  and the propagator scales in the same way. For very short scales  $|n - n'| \sim O(1)$  and  $\xi_c \gg a$ , the gap is not relevant and the propagator is of  $O(1)$ . The same occurs if  $|n - n'| \sim O(1)$  and  $\xi_c \sim a$ . This proves (3.17).

Finally we note that it is quite useful to express the propagator (3.27) and the fluctuations via the density matrix introduced in Ref. [[49]]. For the propagator we get

$$\langle a_{m\sigma}^\dagger a_{m'\sigma'} \rangle = \frac{Z}{2\pi} \int_{-\pi/Z}^{\pi/Z} d\bar{k} \, \hat{n}_{j'\sigma',j\sigma}(\bar{k}) e^{-i\bar{k}(m-m')}, \quad (3.28)$$

where  $\bar{k} = k/Z$ ,

$$\bar{\chi}_k^{(\alpha)}(j, \sigma) = e^{i\frac{k}{Z}(Z-j)} \chi_k^{(\alpha)}(j, \sigma), \quad (3.29)$$

and the density matrix (written as an operator in unit cell space)

$$\hat{n}(\bar{k}) = \sum_{\alpha=1}^{\nu} |\bar{\chi}_k^{(\alpha)}\rangle \langle \bar{\chi}_k^{(\alpha)}| = \hat{n}(\bar{k})^\dagger. \quad (3.30)$$

We note the properties

$$\hat{n}(\bar{k})^2 = \hat{n}(\bar{k}), \quad (3.31)$$

$$\text{tr} \hat{n}(\bar{k}) = \nu, \quad (3.32)$$

which follow immediately from the orthogonality relation  $\langle \bar{\chi}_k^{(\alpha)} | \bar{\chi}_k^{(\alpha')} \rangle = \delta_{\alpha\alpha'}$ .

Using (3.28) one finds after integration by parts

$$\begin{aligned} (m - m') \langle a_{m\sigma}^\dagger a_{m'\sigma'} \rangle &= -i \frac{Z}{2\pi} \int_{-\pi/Z}^{\pi/Z} d\bar{k} \\ &\times \left[ \partial_{\bar{k}} \hat{n}_{j'\sigma',j\sigma}(\bar{k}) \right] e^{-i\bar{k}(m-m')}. \end{aligned} \quad (3.33)$$

Therefore, from (3.13) and (3.24) one can write the fluctuations in terms of the density matrix as

$$l_p \Delta Q_B^2 = \frac{a}{2} \int_{-\pi/Z}^{\pi/Z} \frac{d\bar{k}}{2\pi} \text{tr} \{ [\partial_{\bar{k}} \hat{n}(\bar{k})]^2 \}. \quad (3.34)$$

The formula (3.34) can be equivalently rewritten in terms of the Bloch momentum  $k$  (instead of  $\bar{k}$ ):

$$l_p \Delta Q_B^2 = \frac{Za}{2} \int_{-\pi}^{\pi} \frac{dk}{2\pi} \text{tr} \{ [\partial_k \hat{n}(k)]^2 \}. \quad (3.35)$$

At finite temperature we have to use the following form of the density matrix

$$\hat{n}(\bar{k}) \rightarrow \tilde{n}(\bar{k}) = \sum_{\alpha} |\bar{\chi}_k^{(\alpha)}\rangle \langle \bar{\chi}_k^{(\alpha)}| n_F(\epsilon_k^{(\alpha)}). \quad (3.36)$$

### Example: Rice-Mele model

The Rice-Mele model is defined by  $Z = 2$  with two hoppings  $t_{1/2} > 0$  and staggered on-site potentials  $v = v_1 = -v_2$ . The Bloch Hamiltonian  $h_k$  in the two-dimensional unit cell reads

$$h_k = \begin{pmatrix} v & -t_1 - t_2 e^{-ik} \\ -t_1 - t_2 e^{ik} & -v \end{pmatrix} \quad (3.37)$$

and has eigenvalues

$$\epsilon_k^{(\pm)} = \pm \epsilon_k = \pm \sqrt{\Delta^2 + 4t_1 t_2 \cos^2 \frac{k}{2}}, \quad (3.38)$$

where  $\Delta = \sqrt{v^2 + (t_1 - t_2)^2}$  is half the energy gap  $E_g = 2\Delta$  between the valence  $\epsilon_k^{(-)}$  and the conduction  $\epsilon_k^{(+)}$  bands. The corresponding eigenstates read

$$\chi_k^{(\pm)} = \frac{1}{\sqrt{2\epsilon_k(\epsilon_k \mp v)}} \begin{pmatrix} t_1 + t_2 e^{-ik} \\ v \mp \epsilon_k \end{pmatrix}. \quad (3.39)$$

The expectation value for the boundary charge has been analysed in all detail in Ref. [[22]]. To calculate the fluctuations of the boundary charge we first insert the eigenstates (3.39) in (4.186), and find from (3.30) for the density matrix of the valence band

$$\hat{n}(k) = \frac{1}{2\epsilon_k} \begin{pmatrix} \epsilon_k - v & t_1 e^{\frac{i}{2}k} + t_2 e^{-\frac{i}{2}k} \\ t_1 e^{-\frac{i}{2}k} + t_2 e^{\frac{i}{2}k} & \epsilon_k + v \end{pmatrix}. \quad (3.40)$$

Computing the  $k$ -derivative we obtain

$$\begin{aligned} & \partial_k \hat{n}(k) \\ &= \partial_k \left( \frac{1}{2\epsilon_k} \right) \begin{pmatrix} -v & t_1 e^{\frac{i}{2}k} + t_2 e^{-\frac{i}{2}k} \\ t_1 e^{-\frac{i}{2}k} + t_2 e^{\frac{i}{2}k} & v \end{pmatrix} \\ &+ \frac{i}{4\epsilon_k} \begin{pmatrix} 0 & t_1 e^{\frac{i}{2}k} - t_2 e^{-\frac{i}{2}k} \\ -t_1 e^{-\frac{i}{2}k} + t_2 e^{\frac{i}{2}k} & 0 \end{pmatrix}. \end{aligned} \quad (3.41)$$

Using  $\partial_k \left( \frac{1}{\epsilon_k} \right) = \frac{t_1 t_2 \sin k}{\epsilon_k^3}$ , we evaluate

$$\begin{aligned} & \frac{1}{2} \int_{-\pi}^{\pi} \frac{dk}{2\pi} \text{tr} \{ [\partial_k \hat{n}(k)]^2 \} = \int_{-\pi}^{\pi} \frac{dk}{2\pi} \frac{1}{16\epsilon_k^6} \\ & \times [(\Delta^2 + 4t_1 t_2)^2 (t_1 - t_2)^2 \cos^2 \frac{k}{2} \\ & + \Delta^4 (t_1 + t_2)^2 \sin^2 \frac{k}{2} + 4v^2 t_1^2 t_2^2 \sin^2 k]. \end{aligned} \quad (3.42)$$

Performing this integral, we obtain from (3.35)

$$l_p \Delta Q_B^2 = a \frac{t_1^2 + t_2^2}{8\Delta} \frac{1}{\sqrt{\Delta^2 + 4t_1 t_2}}. \quad (3.43)$$

In the wide-band limit  $\frac{t_1 + t_2}{2} \equiv t \gg \Delta$  we estimate

$$l_p \Delta Q_B^2 \approx \frac{ta}{8\Delta} = \frac{v_F}{16\Delta}, \quad (3.44)$$

with  $v_F = 2ta$ .

**Finite temperature** At finite temperature (but still at fixed particle number) we replace  $\hat{n}(k)$  by  $\tilde{n}(k)$  defined in (3.36) (here we use  $k$  instead of  $\bar{k} = k/2$ ). For a two-band model ( $\alpha = \pm$ ) with the particle-hole symmetry  $\epsilon_k^{(-)} = -\epsilon_k^{(+)} \equiv -\epsilon_k$ , we use the completeness relation  $\sum_{\alpha=\pm} |\bar{\chi}_k^{(\alpha)}\rangle \langle \bar{\chi}_k^{(\alpha)}| = \hat{1}$  in order to simplify (3.36)

$$\tilde{n}(k) = n_F(-\epsilon_k) \hat{n}(k) + n_F(\epsilon_k) (\hat{1} - \hat{n}(k)). \quad (3.45)$$

Hence

$$\begin{aligned} \partial_k \tilde{n}(k) &= [n_F(-\epsilon_k) - n_F(\epsilon_k)] \partial_k \hat{n}(k) \\ &+ \{[\hat{1} - \hat{n}(k)] n'_F(\epsilon_k) - \hat{n}(k) n'_F(-\epsilon_k)\} \frac{d\epsilon_k}{dk}. \end{aligned} \quad (3.46)$$

Squaring this expression and using the properties (3.31), (3.32) as well as  $\text{tr}[\hat{n}(k) \partial_k \hat{n}(k)] = \text{tr}[(\hat{1} - \hat{n}(k)) \partial_k \hat{n}(k)] = 0$ , we find the wide-band limit expression

$$\begin{aligned} l_p \Delta Q_B^2 &\approx \\ &\approx Z a \int_{-\infty}^{\infty} \frac{dk}{2\pi} \left\{ \frac{v_F^2 \Delta^2}{4a^2 \epsilon_k^4} [n_F(-\epsilon_k) - n_F(\epsilon_k)]^2 \right. \\ &\quad \left. + \frac{1}{2} [(n'_F(-\epsilon_k))^2 + (n'_F(\epsilon_k))^2] \left( \frac{d\epsilon_k}{dk} \right)^2 \right\}. \end{aligned} \quad (3.47)$$

Note that the first term in the integrand follows from its finite-band counterpart in (3.42) in the limit under consideration.

Choosing  $\mu = 0$ , we evaluate (3.47)

$$l_p \Delta Q_B^2 \approx 2 \frac{v_F}{\Delta} \int_0^{\infty} \frac{dx}{2\pi} \left\{ \frac{\tanh^2(\frac{\Delta}{2T} \sqrt{x^2 + 1})}{2(1 + x^2)^2} + \frac{(\Delta/T)^2}{8 \cosh^4(\frac{\Delta}{2T} \sqrt{x^2 + 1})} \frac{x^2}{x^2 + 1} \right\}, \quad (3.48)$$

with the rescaled integration variable  $x = \frac{v_F k}{a \Delta}$ . This function decays monotonically in  $T$ , and at  $T \gtrsim \Delta$  it behaves like  $\sim \frac{v_F}{12\pi T}$ . At zero  $T$  we recover (3.44), finding additionally the low-temperature correction

$$\approx -\frac{2v_F}{\Delta \sqrt{2\pi \Delta/T}} e^{-\Delta/T}. \quad (3.49)$$

### Low-energy theory for single channel models

For the case of a noninteracting and clean single channel lattice model in the limit of small gap  $E_g \ll W$ , where  $\xi_g \gg a$ , one can describe the low-energy physics by an effective Dirac Hamiltonian in 1 + 1 dimensions [21]

$$\begin{aligned} H_{\text{bulk}} &= \int dx \underline{\psi}^\dagger(x) \{ -i v_F \partial_x \sigma_z \\ &\quad + \Delta \cos \gamma \sigma_x - \Delta \sin \gamma \sigma_y \} \underline{\psi}(x), \end{aligned} \quad (3.50)$$

where  $E_g = 2\Delta$  is the gap size,  $k_F = \pi\nu/(Za)$  is the Fermi momentum at which the gap opens,  $v_F = 2ta \sin(k_F a)$  denotes the Fermi velocity ( $t \sim W$  is the average hopping), and  $\underline{\psi}(x)$  is a two-component field consisting of slowly varying right- and left-moving fields  $\psi_\pm(x)$  such that the physical field operator can be expressed as

$$\psi(x) = \sum_{p=\pm} \psi_p(x) e^{ipk_F x}. \quad (3.51)$$

The variable  $\gamma$  describes the phase of the order parameter such that  $\Delta e^{i\gamma}$  describes the transition matrix element from  $-k_F$  to  $k_F$ , see Ref. [21].

The Dirac model has two bands with dispersion  $\pm\epsilon_k = \pm\sqrt{v_F^2 k^2 + \Delta^2}$ . For a chemical potential in the gap and  $T \ll \Delta$  all states of the valence band are filled. The eigenstates of the valence band states are given by

$$\underline{\psi}_k(x) = \frac{1}{\sqrt{2\pi N_k}} \begin{pmatrix} -\Delta e^{i\gamma} \\ v_F k + \epsilon_k \end{pmatrix} e^{ikx}, \quad (3.52)$$

with normalization factor  $N_k = \Delta^2 + (v_F k + \epsilon_k)^2 = 2\epsilon_k(\epsilon_k + v_F k)$ . Using this result one can straightforwardly calculate the propagators

$$\begin{aligned} \langle \psi_p^\dagger(x) \psi_{p'}(x') \rangle &= \int dk \langle \psi_{k,p}^\dagger(x) \psi_{k,p'}(x') \rangle \\ &= \frac{1}{2} \delta(x - x') \delta_{pp'} + \frac{1}{4\pi\xi_g} \left[ i\sigma_z K_1 \left( \frac{x - x'}{2\xi_g} \right) \right. \\ &\quad \left. - (\cos \gamma \sigma_x + \sin \gamma \sigma_y) K_0 \left( \frac{x - x'}{2\xi_g} \right) \right]_{pp'}, \end{aligned} \quad (3.53)$$

where  $K_\nu$  denotes the modified Bessel function of the second kind,  $\sigma_i$  are the Pauli matrices, and  $\xi_g = \frac{v_F}{2\Delta} = \frac{v_F}{E_g}$  is the exponential decay length. Using (3.51) and omitting the divergent contribution at  $x = x'$  (which can not be determined from a low-energy model) we find for the density-density correlation function the form

$$\begin{aligned} C_{\text{bulk}}(x, x') &= -|\langle \psi^\dagger(x) \psi(x') \rangle|^2 \\ &= -\frac{1}{4\pi^2 \xi_g^2} \left\{ K_1 \left( \frac{x - x'}{2\xi_g} \right) \sin[k_F(x - x')] \right. \\ &\quad \left. - K_0 \left( \frac{x - x'}{2\xi_g} \right) \cos[k_F(x + x') + \gamma] \right\}^2. \end{aligned} \quad (3.54)$$

To calculate  $C_{jj'}^{\text{bulk}}(Z(n - n')a)$  of the original lattice model, we insert  $x = ma$  and  $x' = m'a$  into this equation. Using  $m = Z(n - 1) + j$  and  $m' = Z(n' - 1) + j'$  together with  $k_F = \pi\nu/(Za)$ , one finds (except for  $n = 0$  and  $j = j'$ ) the final result

$$\begin{aligned} C_{jj'}^{\text{bulk}}(Zn) &= -\frac{1}{4\pi^2 \xi_g^2} \\ &\times \left\{ K_1 \left( \frac{Zn + j - j'}{2\xi_g} \right) \sin\left[\pi \frac{\nu}{Z}(j - j')\right] \right. \\ &\quad \left. - K_0 \left( \frac{Zn + j - j'}{2\xi_g} \right) \cos\left[\pi \frac{\nu}{Z}(j + j') + \gamma\right] \right\}^2. \end{aligned} \quad (3.55)$$

Comparing this analytical result in the low-energy limit with exact numerical ones for the original lattice model we find for small gaps a surprisingly perfect agreement even for small values of  $n$ . Using the asymptotic forms

$$K_0(s) \rightarrow \begin{cases} \sqrt{\frac{\pi}{2|s|}} e^{-|s|} & \text{for } |s| \gg 1 \\ -\ln |s| & \text{for } |s| \ll 1 \end{cases} \quad (3.56)$$

$$K_1(s) \rightarrow \begin{cases} \text{sign}(s) \sqrt{\frac{\pi}{2|s|}} e^{-|s|} & \text{for } |s| \gg 1 \\ \frac{1}{s} & \text{for } |s| \ll 1 \end{cases}, \quad (3.57)$$

we find that the properties (3.16) and (3.17) are fulfilled. Averaging the correlation function over  $j$  and  $j'$  according to (3.20) we find for  $|n| \gg 1$  (where we can neglect  $j - j'$  in the argument of the Bessel functions in (3.55)) the compact form

$$\bar{C}_{\text{bulk}}(x) \approx -\frac{1}{8\pi^2\xi_g^2} \left\{ \left[ K_0 \left( \frac{x}{2\xi_g} \right) \right]^2 + \left[ K_1 \left( \frac{x}{2\xi_g} \right) \right]^2 \right\}, \quad (3.58)$$

with the following asymptotics for small and large  $|x|$

$$\bar{C}_{\text{bulk}}(x) \rightarrow -\frac{1}{2\pi} \begin{cases} \frac{1}{2\xi_g|x|} e^{-|x|/\xi_g} & \text{for } |x| \gg \xi_g \\ \frac{1}{\pi x^2} & \text{for } |x| \ll \xi_g \end{cases}. \quad (3.59)$$

Inserting the result (3.58) in (3.19) gives the following result for the boundary charge fluctuations close to the phase transition point

$$l_p \Delta Q_B^2 = \frac{\xi_g}{8} = \frac{v_F}{16\Delta}. \quad (3.60)$$

This result generalizes the result (3.44) obtained for the Rice-Mele model to any single channel model.

#### 1.1.4 SSH model with disorder

Here we treat the disordered Su-Schrieffer-Heeger (SSH) model in Born approximation to calculate the gap at moderate disorder strength. The infinite bulk model is defined by the single-particle Hamiltonian

$$h = h_0 + V, \quad (3.61)$$

$$h_0 = -\sum_m t_m |m+1\rangle\langle m| + \text{h.c.}, \quad (3.62)$$

$$V = -\sum_m w_m |m+1\rangle\langle m| + \text{h.c.}, \quad (3.63)$$

where  $t_m = t_{m+2}$  are alternating hoppings with  $t_{1/2} > 0$ , and  $w_m$  describes random bond disorder taken from a uniform distribution  $w_m \in [-d_m/2, d_m/2]$ , with  $d_m = d_{m+2} > 0$  describing the strength of disorder on site  $m$ . The disorder-averaged propagator can be written in terms of the self-energy as

$$\bar{G}(E) = \left( \prod_m \frac{1}{d_m} \int_{-d_m/2}^{d_m/2} dw_m \right) \frac{1}{E - h} = \frac{1}{E - h_0 - \Sigma(E)}. \quad (3.64)$$

From the definition we get the useful properties (for any  $E$  in the complex plane)

$$\bar{G}(E)^\dagger = \bar{G}(E^*) \quad , \quad \Sigma(E)^\dagger = \Sigma(E^*), \quad (3.65)$$

or

$$\bar{G}(E)^T = \bar{G}^*(E) \quad , \quad \Sigma(E)^T = \Sigma^*(E), \quad (3.66)$$

where  $A(E)^T$  is the transposed matrix and  $A^*(E)$  denotes the conjugate complex of the matrix (without taking the conjugate complex of  $E$ ). Due to translational invariance we can

write the SSH Hamiltonian  $h_0$  and the self-energy  $\Sigma(E)$  in diagonal form with respect to the quasimomentum  $k$

$$h_0 = \int_{-\pi}^{\pi} dk |k\rangle \langle k| \otimes h_k^{(0)}, \quad (3.67)$$

$$\Sigma(E) = \int_{-\pi}^{\pi} dk |k\rangle \langle k| \otimes \Sigma_k(E), \quad (3.68)$$

where

$$\langle n|k\rangle = \frac{1}{\sqrt{2\pi}} e^{ikn} \quad (3.69)$$

are plane waves with respect to the unit cell index  $n$ , and  $h_k^{(0)}$  and  $\Sigma_k(E)$  are  $2 \times 2$ -matrices within the 2-dimensional unit cell space. For the SSH model we get the form

$$h_k^{(0)} = \begin{pmatrix} 0 & -t_1 - t_2 e^{-ik} \\ -t_1 - t_2 e^{ik} & 0 \end{pmatrix}. \quad (3.70)$$

In this notation matrix elements of the free propagator  $g(E) = 1/(E - h_0)$  can be written as

$$g(E)_{mm'} = \hat{g}(E, n - n')_{jj'}, \quad (3.71)$$

where  $m = 2(n - 1) + j$ ,  $m' = 2(n' - 1) + j'$  and

$$\hat{g}(E, n) = \int_{-\pi}^{\pi} \frac{dk}{2\pi} e^{ikn} \frac{1}{E - h_k^{(0)}} = \int_{-\pi}^{\pi} \frac{dk}{2\pi} e^{ikn} \frac{E - (t_1 + t_2 \cos k)\sigma_x - (t_2 \sin k)\sigma_y}{E^2 - t_1^2 - t_2^2 - 2t_1 t_2 \cos k} \quad (3.72)$$

is a  $2 \times 2$ -matrix with  $\sigma_{x,y,z}$  denoting the Pauli matrices.

In standard Born approximation (see, e.g., Ref. [36]) we find for the nonvanishing matrix elements

$$\Sigma(E)_{m,m+1} = \frac{d_m^2}{12} g(E)_{m+1,m}, \quad (3.73)$$

$$\Sigma(E)_{m+1,m} = \frac{d_m^2}{12} g(E)_{m,m+1}, \quad (3.74)$$

$$\Sigma(E)_{m,m} = \frac{d_m^2}{12} g(E)_{m+1,m+1} + \frac{d_{m-1}^2}{12} g(E)_{m-1,m-1}. \quad (3.75)$$

Since, according to (3.71) and (3.72), the diagonal elements  $g(E)_{mm}$  of the propagator are independent of  $m$ , (3.75) describes only a constant shift of the energy leading to the renormalized energy

$$\tilde{E} = E \left( 1 - \frac{d_1^2 + d_2^2}{12} \int_{-\pi}^{\pi} \frac{dk}{2\pi} \frac{1}{E^2 - t_1^2 - t_2^2 - 2t_1 t_2 \cos k} \right), \quad (3.76)$$

such that the averaged propagator in Born approximation can be written as

$$\bar{G}(E) = \frac{1}{\tilde{E} - h(E)}, \quad (3.77)$$

with  $h(E) = h_0 + \Sigma(E)$ . Using the matrix elements of the self-energy in Born approximation we can write

$$h(E) = - \sum_m \left\{ t_m(E) |m+1\rangle \langle m| + t_m^*(E) |m\rangle \langle m+1| \right\}, \quad (3.78)$$



where  $t_m(E) = t_{m+2}(E)$  and

$$t_1(E) = t_1 - \Sigma(E)_{21} = t_1 - \frac{d_1^2}{12} g(E)_{12}, \quad (3.79)$$

$$t_2(E) = t_2 - \Sigma(E)_{32} = t_2 - \frac{d_2^2}{12} g(E)_{23}, \quad (3.80)$$

and  $t_j^*(E)$  follows from the conjugate complex (but leaving  $E$  invariant). Using (3.71) we get  $g(E)_{12} = \hat{g}(E, 0)_{12}$  and  $g(E)_{23} = \hat{g}(E, -1)_{21}$  which, together with (3.72), leads to

$$g(E)_{12} = \int_{-\pi}^{\pi} \frac{dk}{2\pi} \frac{t_1 + t_2 e^{ik}}{|t_1 + t_2 e^{ik}|^2 - E^2}, \quad (3.81)$$

$$g(E)_{23} = \int_{-\pi}^{\pi} \frac{dk}{2\pi} \frac{t_2 + t_1 e^{ik}}{|t_2 + t_1 e^{ik}|^2 - E^2}. \quad (3.82)$$

At zero energy we get

$$g(0)_{12} = \frac{1}{t_1} \theta(t_1 - t_2), \quad (3.83)$$

$$g(0)_{23} = \frac{1}{t_2} \theta(t_2 - t_1), \quad (3.84)$$

leading to the following final result for the gap in the presence of disorder

$$\begin{aligned} E_g &= 2|t_1(0) - t_2(0)| \\ &= |2(t_1 - t_2) - \frac{d_1^2}{6t_1} \theta(t_1 - t_2) + \frac{d_2^2}{6t_2} \theta(t_2 - t_1)|. \end{aligned} \quad (3.85)$$

For  $t_1 > t_2$  this leads to the gap closing condition

$$r = \frac{t_1}{t_2} = \frac{1}{2} \left\{ 1 + \sqrt{1 + \frac{d_1^2}{3t_2^2}} \right\}, \quad (3.86)$$

which agrees rather well with the numerical result up to  $d_1 \sim 2 - 3$ . For  $t_1 < t_2$ , Born approximation turns out to be insufficient, which will be studied in a future work [50].

### 1.5 Higher dimensions

In this section we calculate the fluctuations for noninteracting and clean models in higher dimensions  $D = 2, 3$ . We start with  $D = 2$  and combine a standard 2D quantum Hall insulator [51] with an additional modulation of the on-site potentials and nearest-neighbor hoppings in  $x$ -direction. Such models were studied extensively in Refs. [[22, 21, 17, 18, 19, 20]] to study the expectation value of the boundary charge. We start from a 2D tight-binding model with sites labeled by  $(m, \sigma)$  with  $m = 1, \dots, N_s$  in  $x$ -direction and  $\sigma = 1, \dots, N_{\perp}$  in  $y$ -direction. The lattice spacing  $a = a_x = a_y$  is assumed to be the same in both directions. We take open boundary conditions in  $x$ -direction and periodic ones in  $y$ -direction. We consider a constant magnetic field  $B$  perpendicular to the sample in Landau gauge  $\underline{A}(m, \sigma) = (0, Bm, 0)$ . As a result, the hopping in  $y$ -direction from  $(m, \sigma) \rightarrow (m, \sigma + s)$  acquires a phase factor  $e^{i2\pi sm/Z}$ , where  $\lambda_B = Za$  is the magnetic length defined by  $\lambda_B/a = Z = \Phi_0/(Ba^2)$ , where  $\Phi_0 = hc/e$  denotes the flux quantum. For simplicity we assume that  $Z$  is an integer. If an additional flux  $\Phi$  is applied through the hole of the cylinder when the system is deformed to a ring in  $y$ -direction, an additional phase factor  $e^{-is\theta/N_{\perp}}$  with  $\theta = 2\pi\Phi/\Phi_0$  has to be considered for the hopping in  $y$ -direction by  $s$  sites.

Taking real and negative nearest-neighbor hoppings  $-t_m = -t_{m+Z}$  in  $x$ -direction and  $-t_y$  in  $y$ -direction, together with real on-site potentials  $v_m = v_{m+Z}$ , we arrive at the following 2D tight-binding model

$$\begin{aligned}
H = & \sum_{m=1}^{N_s} \sum_{\sigma=1}^{N_\perp} v_m a_{m\sigma}^\dagger a_{m\sigma} \\
& - \sum_{m=1}^{N_s-1} \sum_{\sigma=1}^{N_\perp} t_m (a_{m+1,\sigma}^\dagger a_{m\sigma} + \text{h.c.}) \\
& - t_y \sum_{m=1}^{N_s} \sum_{\sigma=1}^{N_\perp} \left( e^{i(2\pi m/Z - \theta/N_\perp)} a_{m,\sigma+1}^\dagger a_{m\sigma} + \text{h.c.} \right), \tag{3.87}
\end{aligned}$$

with  $a_{m\sigma} = a_{m,\sigma+N_\perp}$  due to periodic boundary conditions in  $y$ -direction. Using Fourier transform for the modes in this direction

$$a_{m\sigma} = \frac{1}{\sqrt{N_\perp}} \sum_{k_y} e^{ik_y a \sigma} c_{mk_y}, \tag{3.88}$$

with

$$k_y = \frac{2\pi}{N_\perp a} s, \quad s = 1, \dots, N_\perp, \tag{3.89}$$

we can write the Hamiltonian as an independent sum over all transverse quasimomenta

$$H = \sum_{k_y} H(k_y) \tag{3.90}$$

$$\begin{aligned}
H(k_y) = & \sum_{m=1}^{N_s} \bar{v}_m (k_y a + \theta/N_\perp) c_{mk_y}^\dagger c_{mk_y} \\
& - \sum_{m=1}^{N_s-1} t_m (c_{m+1,k_y}^\dagger c_{mk_y} + \text{h.c.}), \tag{3.91}
\end{aligned}$$

with

$$\bar{v}_m(\varphi) = v_m - 2t_y \cos\left(\frac{2\pi}{Z}m - \varphi\right). \tag{3.92}$$

The charge operator summed over all transverse modes is given by

$$\hat{\rho}_m = \sum_{\sigma=1}^{N_\perp} a_{m\sigma}^\dagger a_{m\sigma} = \sum_{k_y} c_{mk_y}^\dagger c_{mk_y}. \tag{3.93}$$

Since the Hamiltonian does not couple the transverse modes, the fluctuations of the boundary charge operator (3.4) can be written as an independent sum over the transverse modes

$$\Delta Q_B^2 = \sum_{k_y} \Delta Q_B^2(k_y), \tag{3.94}$$

where  $\Delta Q_B^2(k_y)$  describe the boundary charge fluctuations of the effectively 1-dimensional Hamiltonian  $H(k_y)$  at fixed  $k_y$ . Denoting by  $2\Delta(k_y)$  the gap of this Hamiltonian and by

$v_F(k_y)$  the Fermi velocity, we can take in the low-energy limit of small gap compared to the band width the result (3.60), leading for  $N_\perp \rightarrow \infty$  to the integral

$$\begin{aligned} l_p \Delta Q_B^2 &= \sum_{k_y} \frac{v_F(k_y)}{16\Delta(k_y)} \\ &\rightarrow \frac{N_\perp a}{2\pi} \int_{-\pi/a}^{\pi/a} dk_y \frac{v_F(k_y)}{16\Delta(k_y)}. \end{aligned} \quad (3.95)$$

To evaluate the result (3.95) explicitly, we consider an illustrative example in terms of the Rice-Mele model, with  $Z = 2$  and  $v_m = 0$ . Using (3.92) we obtain

$$\bar{v}_m(\varphi) = (-1)^m v(\varphi), \quad v(\varphi) = -2t_y \cos(\varphi). \quad (3.96)$$

The gap opens up at  $k_F = \pi/(2a)$  which corresponds to half-filling. In the low-energy limit  $t_y, |t_1 - t_2| \ll t = (t_1 + t_2)/2$ , we obtain  $v_F(k_y) = 2ta$  and

$$\Delta(k_y) = \sqrt{[v(k_y a + \theta/N_\perp)]^2 + |t_1 - t_2|^2}. \quad (3.97)$$

Inserting in (3.95) and calculating the integral leads to the final result

$$l_p \Delta Q_B^2 = \frac{N_\perp}{8\pi} \frac{v_F}{\sqrt{4t_y^2 + \Delta^2}} K\left(\frac{4t_y^2}{4t_y^2 + \Delta^2}\right), \quad (3.98)$$

where  $K(p) = \int_0^{\pi/2} dx [1 - p \sin^2(x)]^{-1/2}$  is the elliptic integral of first kind, and

$$E_g = 2\Delta = 2 \min_{k_y} \Delta(k_y) = 2|t_1 - t_2| \quad (3.99)$$

denotes the gap for the 2-dimensional system. As expected the fluctuations are independent of the phase  $\theta$ .

For  $\Delta \ll t_y$ , we can use the asymptotics  $K(p) \rightarrow \frac{1}{2} \ln(16/(1-p))$  for  $p \rightarrow 1$ , and obtain a logarithmic scaling of the fluctuations as function of the gap close to the phase transition

$$l_p \Delta Q_B^2 \xrightarrow{\Delta \ll t_y} \frac{N_\perp}{16\pi} \frac{v_F}{t_y} \ln \frac{8t_y}{\Delta}. \quad (3.100)$$

For the Rice-Mele model, we can also use the exact result (3.43) to calculate the fluctuations at fixed  $k_y$ . Performing the integral over  $k_y$  this leads to the following result for the total fluctuations

$$l_p \Delta Q_B^2 = N_\perp a \frac{4t^2 + \Delta^2}{32\pi t_y^2} I\left(\frac{\Delta}{2t_y}, \frac{t}{t_y}\right), \quad (3.101)$$

with

$$I(a, b) = \frac{1}{b\sqrt{1+a^2}} K\left(\frac{b^2 - a^2}{b^2(1+a^2)}\right). \quad (3.102)$$

To compare with (3.98) we can rewrite this result as

$$l_p \Delta Q_B^2 = \frac{N_\perp}{8\pi} \frac{2ta(1 + \frac{\Delta^2}{4t^2})}{\sqrt{4t_y^2 + \Delta^2}} K\left(\frac{t_1 t_2}{t^2} \frac{4t_y^2}{4t_y^2 + \Delta^2}\right), \quad (3.103)$$

and find that they agree for  $\Delta = |t_1 - t_2| \ll t = (t_1 + t_2)/2$ .

In the low-energy regime one can also analyse the boundary charge analytically. It is given as an independent sum over the boundary charges of the effective one-dimensional models

$$Q_B(\theta) = \sum_{k_y} Q_B^{1D}(k_y a + \frac{\theta}{N_\perp}). \quad (3.104)$$

Using (3.89) and the periodicity  $Q_B^{1D}(\varphi) = Q_B^{1D}(\varphi + 2\pi)$ , this can also be written as

$$Q_B(\theta) = \frac{N_\perp}{2\pi} \sum_{l=-\infty}^{\infty} e^{il\theta} \int_0^{2\pi} d\varphi Q_B^{1D}(\varphi) e^{-ilN_\perp \varphi}. \quad (3.105)$$

At zero chemical potential the boundary charge of the one-dimensional model can be calculated from the low-energy analysis of Ref. [[20]] as

$$Q_B^{1D}(\varphi) = \frac{\gamma(\varphi)}{2\pi} + \frac{1}{4} - \theta_{\pi/2 < \gamma(\varphi) < \pi}, \quad (3.106)$$

where  $-\pi < \gamma(\varphi) < \pi$  and

$$\Delta(\varphi) e^{i\gamma(\varphi)} = v(\varphi) + i(t_2 - t_1). \quad (3.107)$$

Since  $t_{1,2}$  are independent of the phase,  $\gamma(\varphi) = \gamma(\varphi + 2\pi)$  is bounded to an interval smaller than  $2\pi$  and the phase factor  $e^{i\gamma(\varphi)}$  can not wind around the origin in the complex plane. As a consequence,  $Q_B^{1D}(\varphi)$  is a smooth and periodic function of  $\varphi$  and the integral in (3.105) can only contribute for  $l = 0$ . Therefore, for this special model, the boundary charge is independent of  $\theta$  and the Hall current is zero. This result holds not only in the low-energy regime but also at large gap.

Although the topology does not change at the phase transition point  $r = t_1/t_2 = 1$ , there is a discontinuous change of the appearance of edge states in the gap. For  $r > 1$ , there is no edge state for any  $k_y$ , whereas, for  $r < 1$ , an edge state appears for all  $k_y$  at energy  $-v(k_y a + \theta/N_\perp) = 2t_y \cos(k_y a + \theta/N_\perp)$  which never touches the band edges. Therefore, the gap has to close at the transition, and the fluctuations show the same characteristic scaling as for topological phase transitions where the gap closing is induced by a change of the Chern number.

The above analysis can easily be generalized to a 3D-system by choosing a nearest-neighbor hopping  $t_z$  in  $z$ -direction and adding a magnetic field of size  $B$  in  $y$ -direction. Omitting the additional flux  $\Phi$ , we can write the Hamiltonian as an independent sum over all  $k_y$  and  $k_z$

$$H = \sum_{k_y, k_z} H(k_y, k_z) \quad (3.108)$$

$$H(k_y, k_z) = \sum_{m=1}^{N_s} \bar{v}_m(k_y, k_z) c_{mk_y k_z}^\dagger c_{mk_y k_z} - \sum_{m=1}^{N_s-1} t_m (c_{m+1, k_y k_z}^\dagger c_{mk_y k_z} + \text{h.c.}), \quad (3.109)$$

with

$$\bar{v}_m(k_y, k_z) = v_m - 2t_y \cos\left(\frac{2\pi}{Z}m - k_y a\right) - 2t_z \cos\left(\frac{2\pi}{Z}m - k_z a\right). \quad (3.110)$$

Taking again the special case  $Z = 2$  and  $v_m = 0$ , we obtain for each fixed  $(k_y, k_z)$  an effective Rice-Mele model with potential  $v(k_y, k_z) = \bar{v}_1(k_y, k_z) = -\bar{v}_2(k_y, k_z)$ , where

$$v(k_y, k_z) = 2t_y \cos(k_y a) + 2t_z \cos(k_z a). \quad (3.111)$$

Using the exact result (3.43) for the calculation of the fluctuations at fixed  $(k_y, k_z)$  and integrating over the transverse momenta one arrives straightforwardly at the following result for the total fluctuations

$$l_p \Delta Q_B^2 = N_\perp^2 a^2 \frac{4t^2 + \Delta^2}{32\pi^2 t_y^2} \int_0^{\pi/a} dk_z I(g(k_z), h(k_z)) , \quad (3.112)$$

with

$$g(k_z) = \frac{1}{2t_y} \sqrt{t_z^2 \cos^2(k_z a) + \Delta^2} , \quad (3.113)$$

$$h(k_z) = \frac{1}{2t_y} \sqrt{t_z^2 \cos^2(k_z a) + 4t^2} , \quad (3.114)$$

and  $I(a, b)$  is defined in (3.102). This leads to a convergent result for the fluctuations in the zero gap limit  $\Delta \rightarrow 0$ .

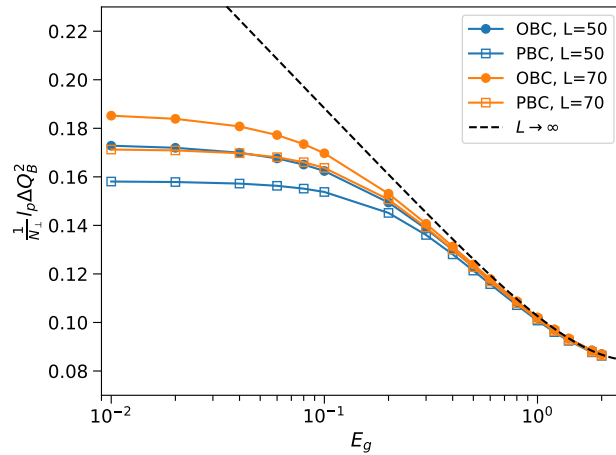


Figure 21: Comparison of different system sizes  $L = N_\perp a = N_s a$  and boundary conditions of the two dimensional model discussed in the text. The parameters are given by  $Z = 2$ ,  $v_m = 0$ ,  $t_{1/2} = 1.0 \pm \Delta/2$ ,  $t_y = 1.0$ ,  $\theta = 0$ ,  $L_p = L/2$  and  $l_p = (24, 34)$  whereby  $\Delta$  is varied in the interval  $[0.005, 1.0]$ .

To access the influence of different boundary conditions at finite system size we consider the two-dimensional model described above on a square lattice with  $L = N_\perp a = N_s a$ . Figure 21 summarizes the finite size scaling (symbols) compared to the infinite limit (dashed line) for open (OBC) and periodic (PBC) boundary conditions. As the system size is increased the universal result is approached for increasingly smaller gaps  $E_g$ .

### 1.6 Special phase transitions

Sometimes an expansion of the Bloch Hamiltonian  $h_k$  around the quasimomentum where the gap opens does not contain linear terms due to special symmetry conditions. An exemplary model of this type has been discussed in Ref. [[52]]. It contains only constant and quadratic terms in  $k$ :

$$h_k = \begin{pmatrix} \Delta + \alpha k^2 & \Gamma k^2 \\ \Gamma k^2 & -\Delta - \beta k^2 \end{pmatrix}. \quad (3.115)$$

The gap between the two bands is  $E_g = 2\Delta$ . The Bloch eigenstates depend only on the combination  $\frac{\Delta}{k^2}$ . It follows immediately that the imaginary part  $\kappa$  of the branching point appears to be  $\kappa \sim \sqrt{E_g}$ . Thus,  $l_p \Delta Q_B^2 \sim E_g^{-1/2}$ .

This can be generalized to the case where the minimal nonvanishing order of  $k$  in the expansion of  $h_k$  is  $l$ . This gives straightforwardly the scaling  $l_p \Delta Q_B^2 \sim \xi \sim E_g^{-1/l}$ .

We note in passing that in quasi-two-dimensional models the band structure of the form (3.115) can be realized in the multilayer graphene with a special stacking [53]. However, as has been shown in Section 1.5, for a  $2D$ -system the fluctuations follow from an integration over the transverse momentum  $k_y$ . Assuming close to the gap opening a general dispersion relation for the conduction band of the form

$$\epsilon(k_x, k_y) = (\Delta^{2/l} + ck_y^2)^{l/2}, \quad (3.116)$$

the solution of  $\epsilon(k_x, k_y) = 0$  for fixed  $k_y$  gives in the complex plane the solution  $k_x = 2i/\xi(k_y)$ , with

$$\xi(k_y) = \frac{2\sqrt{c}}{\sqrt{\Delta^{2/l} + ck_y^2}}. \quad (3.117)$$

The gap at fixed  $k_y$  is given by  $E_g(k_y) = 2\epsilon(0, k_y) = 2(\Delta^{2/l} + ck_y^2)^{l/2}$ , leading to an exotic scaling for the fluctuations of  $Q_B$  at fixed  $k_y$

$$l_p \Delta Q_B^2(k_y) \sim \xi(k_y) \sim E_g(k_y)^{-1/l}. \quad (3.118)$$

However, this does not change the logarithmic scaling of the total fluctuations in terms of the overall gap  $E_g = 2\Delta$

$$l_p \Delta Q_B^2 \sim \int_{-\pi/a}^{\pi/a} dk_y \xi(k_y) \stackrel{E_g \rightarrow 0}{\sim} |\ln(E_g/W)|, \quad (3.119)$$

where  $W$  defines the high-energy cutoff in terms of the band width.

### 1.7 $L_p$ dependence

## 4 Surface charge theorem and topological constraints for edge states

An analytical study of one-dimensional nearest-neighbor tight-binding models by Mikhail Pletyukhov et al.

### Abstract

For a wide class of noninteracting tight-binding models in one dimension we present an analytical solution for all scattering and edge states on a half-infinite system. Without assuming any symmetry constraints we consider models with nearest-neighbor hoppings and one orbital per site but arbitrary size of the unit cell and generic modulations of on-site potentials and hoppings. The solutions are parametrized by determinants which can be straightforwardly calculated from recursion relations. We show that this representation allows for an elegant analytic continuation to complex quasimomentum consistent with previous treatments for continuum models. Two important analytical results are obtained based on the explicit knowledge of all eigenstates. (1) An explicit proof of the surface charge theorem is presented including a unique relationship between the boundary charge  $Q_B^{(\alpha)}$  of a single band  $\alpha$  and the bulk polarization in terms of the Zak-Berry phase. In particular, the Zak-Berry phase is determined within a special gauge of the Bloch states such that no unknown integer is left. This establishes a precise form of a bulk-boundary correspondence relating the *boundary charge of a single band* to bulk properties. (2) We

derive a topological constraint for the phase-dependence of the edge state energies, where the phase variable describes a continuous shift of the lattice towards the boundary. The topological constraint is shown to be equivalent to the quantization of a topological index  $I = \Delta Q_B - \bar{\rho} \in \{-1, 0\}$  introduced in an accompanying letter [arXiv:1911.06890]. Here  $\Delta Q_B$  is the change of the boundary charge  $Q_B$  for a given chemical potential in the insulating regime when the lattice is shifted by one site towards the boundary, and  $\bar{\rho}$  is the average charge per site (both in units of the elementary charge  $e = 1$ ). This establishes an interesting link between universal properties of the boundary charge and edge state physics discussed within the field of topological insulators. In accordance with previous results for continuum systems, we also establish the localization of the boundary charge and determine the explicit form of the density given by an exponential decay and a pre-exponential function following a power-law with generic exponent  $-1/2$  at large distances.

## 1. Introduction

Triggered by the discovery of the quantum Hall effect [1, 51], the study of insulating materials has received considerable interest in the last decade due to the development of the field of topological insulators (TIs) with interesting edge states appearing in the gap which might be useful for quantum information processing [[3]-[7]], see Refs. [[8]-[14]] for reviews and textbooks. Since edge states describe interesting physical phenomena happening at the boundary of a system, this field is ultimately related to the study of the density at the boundary for an insulating system, where not only the edge states but also the scattering states have a nontrivial effect. That the boundary charge  $Q_B$  (in units of the elementary charge  $e = 1$ ) for a given chemical potential in the insulating regime does not only consist of the number  $Q_E$  of occupied edge states and can appear in quantized fractionalized units is well known and has a long history [15]. For systems with *local* [16] inversion symmetry  $Q_B$  is generically quantized in half-integer units and the field of topological crystalline insulators [[17]-[22]] has been put forward recently, extending the standard classification schemes of TIs [[23]-[32]]. An important step towards a generic discussion of  $Q_B$  has been undertaken within the so-called modern theory of polarization (MTP) [[33]-[40]], see Ref. [[41]] for a recent textbook review over the field. In summary, the MTP is based on two fundamental ingredients put forward in Refs. [[33, 31]]: (1) the establishment of a unique definition of the bulk polarization of an insulating crystal in terms of the Zak-Berry phase [42] and (2) the proof of the *surface charge theorem* which provides a relation between  $Q_B$  and the bulk polarization. Most importantly, this relation is not restricted to any symmetry constraints and holds for generic models in all dimensions. It provides an interesting variant of the common bulk-boundary correspondence discussed within the field of TIs, where a relation between bulk topological invariants (like Chern and winding numbers) for infinite systems and the number of zero-energy edge states for systems with a boundary is requested [[43]-[46]]. In analogy, the MTP provides a first step to set up such a relation between the bulk Zak-Berry phase and  $Q_B$  (instead of  $Q_E$ ). However, a limitation of the MTP is the fact that the Zak-Berry phase is not a gauge-invariant quantity. There is still a freedom to add any phase factor  $e^{i\varphi_{\vec{k}}} = e^{i\varphi_{\vec{k}+\vec{G}}}$  to the Bloch waves which depends on the quasimomentum vector  $\vec{k}$  and is periodic when  $\vec{k}$  is shifted by a vector  $\vec{G}$  of the reciprocal lattice. The winding of this phase factor gives rise to an arbitrary integer for the Zak-Berry phase. Moreover, also the proof of the surface charge theorem in terms of counting Wannier function centers in a certain volume does neither control the number of edge states nor the precise number of exponentially localized Wannier states close to the boundary (with wave functions differing significantly from those of the infinite system) in terms of bulk quantities. Both are hard to determine analytically for generic systems and often have to be calculated numerically. As a consequence, an important topological integer is left unknown in the surface charge theorem and it is even not clear how to relate this unknown integer to  $Q_E$ . This makes



it very difficult to bridge the field of TIs discussing edge states with the MTP concentrating on  $Q_B$ .

Another important topic discussed in the literature is to identify *universal* properties of  $Q_B$ . For systems with local inversion symmetry it is known that the Zak-Berry phase is quantized in half-integer units of  $2\pi$  [42] such that, due to the surface charge theorem,  $Q_B$  is also quantized in half-integer units. The same quantization appears at half-filling for systems with local chiral symmetry like, e.g., for the Su-Schrieffer-Heeger (SSH) model [47]. Away from symmetry constraints a finite one-dimensional (1D) tight-binding model with nearest neighbor hopping and a harmonic on-site potential with a wave-length commensurate with the lattice has been studied numerically [17]. In the insulating regime a very stable and almost linear behaviour of  $Q_B(\varphi)$  (up to discrete jumps arising from edge states crossing the chemical potential) as function of a phase variable  $\varphi$  controlling the offset of the potential was revealed. Later on this work was extended to other fillings [18], the stability against random disorder was demonstrated, and it was shown that the slope is universal and can be related to the quantized Hall conductance. Obviously this slope can not be explained by edge state physics and is triggered by scattering states. An intuitive and very simple physical argument has been put forward [17, 18] in terms of charge conservation which, for a half-infinite system, can be formulated as follows: if  $a$  denotes the lattice spacing, the unit cell consists of  $Z$  sites for a given commensurate wave-length  $\lambda = Za$  of the potential. This leads to an average charge  $\bar{\rho} = \frac{\nu}{Z}$  per site when  $\nu$  bands are filled. Shifting the potential continuously by one site towards the boundary via a phase change by  $\frac{2\pi}{Z}$  one expects in an adiabatic picture that on average the charge  $\bar{\rho}$  will be shifted into the boundary leading to an increase of the boundary charge  $Q_B$  by exactly the same amount. This is fundamentally related to the fact that  $Q_B$  is defined via a macroscopic average on scales much larger than  $Za$ , analog to the definition of the macroscopic charge density in classical electrodynamics (see, e.g., Chapter 4.5.1 in Ref. [[41]]). As a result, for large  $Z$  or, equivalently, in the large wave-length limit of the potential, one expects that  $Q_B(\varphi)$  will be almost a linear function with a universal slope  $\frac{Z\bar{\rho}}{2\pi} = \frac{\nu}{2\pi}$  on average. However, similar to the surface charge theorem described above, this physical argument misses an unknown integer. Since  $Q_B(\varphi) = Q_B(\varphi + 2\pi)$  must be periodic when the lattice is shifted by one unit cell, the charge increase by  $\nu$  due to the average linear slope must be compensated by the net difference of edge states moving above and those moving below the chemical potential during the phase interval  $2\pi$ . It is known from the integer quantum Hall effect (IQHE) that this number is given by the sum  $C_\nu = \sum_{\alpha=1}^{\nu} C^{(\alpha)}$  of the Chern numbers of the occupied bands, which, due to the Diophantine equation [54, 55, 56], is given by  $C_\nu = \nu - s_\nu Z$ , where  $s_\nu$  is another integer topological index characteristic for gap  $\nu$ . To fulfil charge conservation it is therefore possible that the average linear slope can in general take all values  $\frac{\nu - s_\nu Z}{2\pi}$ . Furthermore, since the slope appears only on average and large deviations can appear for higher lying gaps and for cases when the phase-dependence of the model parameters is quite strong, a precise definition of universality is required.

The fact that an important topological integer is missing within the MTP and within the discussion of universal properties of  $Q_B(\varphi)$  was the motivation of the accompanying letter [52] to fix the appropriate gauge for the Zak-Berry phase and to identify a novel topological index to characterize universal properties of  $Q_B(\varphi)$ . In a first step, 1D nearest-neighbor tight-binding models with one orbital per site on a half-infinite system were studied, with arbitrary size  $Z$  of the unit cell and generic on-site potentials  $v_j$  and hoppings  $t_j$  within a unit cell, where  $j = 1, \dots, Z$ , see Fig. 22 for illustration. The  $\varphi$ -dependence of the parameters was chosen such that the phase-shift  $\Delta\varphi = \frac{2\pi}{Z}$  describes a shift of the lattice by one site towards the boundary, i.e.,  $v_{j+1}(\varphi) = v_j(\varphi + \frac{2\pi}{Z})$  and  $t_{j+1}(\varphi) = t_j(\varphi + \frac{2\pi}{Z})$ . For constant hopping and in the continuum limit ( $Z \rightarrow \infty$ ,  $a \rightarrow 0$ , such that the length  $Za$  of a unit cell stays constant), one obtains at low filling the whole class of 1D solid state systems with generic periodic potentials

$V(x) = V(x + Za)$ . For the special case of cosine modulations with respect to  $\varphi$ , the whole class of generalized Aubry-André-Harper models [53] is covered, discussed extensively in the context of topological insulators, the IQHE, photonic crystals, and cold atom systems. It was argued in Ref. [[52]] that, based on the detailed knowledge of the eigenstates described in the present manuscript, a gauge of the Bloch states can be found such that the boundary charge  $Q_B^{(\alpha)}$  of a single band  $\alpha$  can be related in a unique way to the Zak-Berry phase fixing the unknown constant of the surface charge theorem for a *single* band. It was also shown that this gauge is related in a very natural way to the phase factors of the partial Bloch waves defining the scattering eigenstates of the half-infinite system. Concerning the universal properties of the phase-dependence of  $Q_B^{(\alpha)}(\varphi)$ , it turned out that the representation in terms of the Zak-Berry phase is very helpful to show that the change of  $Q_B^{(\alpha)}$  under a shift of the lattice by one site towards the boundary can be written as  $\Delta Q_B^{(\alpha)}(\varphi) = I_\alpha(\varphi) + \frac{1}{Z}$ . Here,  $I_\alpha$  is a gauge invariant topological index quantized in integer units, which was shown to be identical to the winding number  $w_\alpha = -I_\alpha$  of a fundamental phase  $\theta_k^{(\alpha)}$  given by the phase difference of the Bloch wave function between the sites right and left to the position of the boundary defining the half-infinite system. Interestingly, concerning the boundary charge  $Q_B$  at fixed chemical potential (which includes the sum  $\sum_{\alpha=1}^\nu Q_B^{(\alpha)}$  of the occupied bands together with the number  $Q_E$  of occupied edge states) physical arguments were presented that the quantized topological index  $I(\varphi) = \Delta Q_B(\varphi) - \bar{\rho} = \sum_{\alpha=1}^\nu I_\alpha(\varphi) + \Delta Q_E(\varphi)$  can only take the two possible values  $I \in \{-1, 0\}$ . Here, the value  $I = 0$  is associated with charge conservation of particles, which is the argument described above leading to  $\Delta Q_B = \bar{\rho}$ . [17] The other value  $I = -1$  leading to  $\Delta Q_B = \bar{\rho} - 1$  is associated in an analog way with charge conservation of holes. We note that the duality between particles and holes is based only on the Pauli principle and needs no further symmetry constraint. Since charge conservation of particles and holes is fulfilled for any model it is quite remarkable that these simple physical ingredients are sufficient to describe an important universal feature of  $Q_B(\varphi)$ . Without explicit proof it was argued in Ref. [[52]] that other values of  $I$  are not possible since it is always possible to choose the  $\varphi$ -dependence of the model parameters in such a way that no edge states cross the chemical potential during the shift of the lattice by one site (keeping the model parameters fixed before and after the shift). Moreover, it was shown in Ref. [[52]] that the quantization of  $I$  together with the periodicity  $Q_B(\varphi) = Q_B(\varphi + 2\pi)$  determines the generic form of  $Q_B(\varphi)$  such that the average linear slope is given in accordance with the Diophantine equation by  $\frac{\nu - s_\nu Z}{2\pi}$ , where  $s_\nu$  is a topological index characterizing a topological constraint for the phase-dependence of the edge state energies. This constraint was stated as

$$\Delta F(\varphi) = s_\nu + I(\varphi) \in \{s_\nu - 1, s_\nu\}, \quad (4.1)$$

where  $\Delta F(\varphi)$  is the difference of the number of edge states moving below and those moving above the chemical potential during the shift of the lattice by one site. This means that the topological index  $s_\nu$  appears at two different places in the  $\varphi$ -dependence of  $Q_B$ , in the average linear slope as well as in the number of discrete jumps appearing when edge states cross the chemical potential. For the calculation of the topological index  $I(\varphi)$  the influence of these two dependencies cancel each other such that its quantization can be explained by charge conservation of particles and holes alone without involving any edge state physics.

The purpose of this work is to present the exact solution for all eigenstates of the considered class of half-infinite tight-binding models and to provide a rigorous proof of all statements of the accompanying letter regarding the unique formulation of the surface charge theorem for a single band and the derivation of the topological constraint (4.1). The solutions for the scattering and edge states are presented in terms of sub-determinants of the matrix  $h_k - \epsilon$ , where  $h_k$  is the Hamiltonian in quasimomentum space and  $\epsilon$  is the energy of the eigenstate, and we show how to calculate the sub-determinants efficiently from a set of recursion relations. To the best of our

knowledge, this representation has not been stated before and provides a very efficient analytical and numerical tool to obtain all eigenstates and physical observables like the density and the boundary charge. We note that in contrast to many other approaches trying to find effective analytical or numerical solutions for tight-binding models on finite systems,[58] half-infinite systems have the advantage that the quasimomentum is continuous and the thermodynamic limit has already been carried out such that the two ends of the system can no longer effect each other. In such systems, following general arguments put forward in a recent article [57] the solutions of the infinite system for complex quasimomentum form generically a basis to construct all scattering and edge states for a half-infinite 1D system (up to special bifurcation points where additional solutions with pre-exponential power-laws have to be taken into account). Therefore, the use of complex quasimomentum is helpful for any system to calculate all scattering and edge states of a half-infinite system (even beyond the considered class of models in this work). However, we note that Ref. [[57]] just describes a general scheme for this construction and does not present an explicit representation of the Bloch states in terms of the quasimomentum (which requires the explicit solution of the Schrödinger equation).

Concerning the definition of the Zak-Berry phase we will elaborate on two different ways which can be chosen for the form of the Bloch wave in tight-binding models with periodic modulation of the parameters. The choice used in Ref. [[52]] is given by

$$\psi_{k,\text{bulk}}^{(\alpha)}(n, j) = \frac{1}{\sqrt{2\pi}} \chi_k^{(\alpha)}(j) e^{ikn}, \quad (4.2)$$

where  $n$  is the unit cell index and the  $Z$ -dimensional vector  $\chi_k = \chi_{k+2\pi}$  is periodic in  $k$  and describes the form of the Bloch wave function within the unit cell (we have set the lattice spacing  $a = 1$  such that  $-\pi \leq k < \pi$ ). The other possible choice used standardly in solid state physics is

$$\psi_{k,\text{bulk}}^{(\alpha)}(m) = \frac{1}{\sqrt{2\pi}} \bar{\chi}_k^{(\alpha)}(j) e^{i \frac{k}{Za} ma}, \quad (4.3)$$

where  $m = Z(n - 1) + j$  is the lattice site index and  $\bar{\chi}_k^{(\alpha)}(j) = \chi_k^{(\alpha)}(j) e^{ik \frac{Z-j}{Z}}$  is no longer periodic in  $k$ . Here,  $\frac{k}{Za}$  plays the role of the quasimomentum and  $ma$  is the position in real space. Defining the Zak-Berry phase  $\bar{\gamma}_\alpha$  of band  $\alpha$  with respect to  $\bar{\chi}_k^{(\alpha)}$  (as it is standardly done within the MTP) we will prove the central result

$$Q_B^{(\alpha)} = -\frac{\bar{\gamma}_\alpha}{2\pi} - \frac{Z-1}{2Z}, \quad (4.4)$$

provided that the gauge is chosen such that the last component  $\chi_k^{(\alpha)}(Z) = \bar{\chi}_k^{(\alpha)}(Z)$  is real. This gauge is ultimately related to the boundary condition for the half-infinite system since the scattering states get the form  $\psi_k^{(\alpha)}(n, j) \sim \chi_k^{(\alpha)}(j) e^{ikn} - \text{h.c.}$  in this gauge (here,  $n = 1, j = 1$  denotes the first site of a half-infinite system with a left boundary or, alternatively,  $n = 0, j = Z - 1$  is the last site of a half-infinite system with a right boundary, see Fig. 22). Eq. (4.4) is the unique formulation of the surface charge theorem for a single band where the second term on the right hand side describes the polarization of the ions. We note that a similar unique formulation of the surface charge theorem for the boundary charge  $Q_B = \sum_{\alpha=1}^{\nu} Q_B^{(\alpha)} + Q_E$  at fixed chemical potential is not possible since it is not known how to relate the number  $Q_E$  of edge states to a bulk quantity in the absence of any symmetry constraints. Only the total number  $Q_E^{\text{tot}}$  of edge states (independent of whether they are occupied or not) is given by the quantized sum  $\sum_{\alpha=1}^Z \gamma_\alpha$  over all Zak-Berry phases defined with respect to  $\chi_k^{(\alpha)}$  in the particular gauge described above.

Concerning the rigorous proof of the topological constraint (4.1) we will use two alternative ways. One is based on the explicit solution of the edge state wave functions in terms of the

sub-determinants of  $h_k - \epsilon$ . This provides concrete conditions how the energy  $\epsilon(\varphi)$  of the edge states can cross the chemical potential as function of the phase. We visualize this by a convenient diagrammatic language where  $\Delta F$  can be identified with an effective topological charge which, due to certain diagrammatic rules can only take two possible values  $s_\nu - 1$  or  $s_\nu$ . Based on the topological constraint we also derive useful rules how the phase-dependence of the edge state energies can look like in general. We complement this by a second proof based on the analytic continuation of Bloch states to complex quasimomentum, which is straightforward by using the explicit solution for the eigenstates. In accordance with previous approaches for half-infinite continuum systems [47] we find that the Bloch states have a pole in the complex plane corresponding to the edge state solutions. In addition branch cuts appear separating different bands on the real axis, see also Ref. [[60]] for continuum systems. We demonstrate that, as function of  $\varphi$ , the position of the edge state poles oscillate around the branch cuts, in accordance with similar findings using analytic continuations based on transfer matrices within the IQHE [56, 43, 61]. We will show that these oscillations are essential to prove that the phase-dependence of the model parameters within a phase-interval of size  $\frac{2\pi}{Z}$  can be always chosen such that no edge state crosses the chemical potential in a certain gap. This input was used in the accompanying letter [52] to justify that edge states are not the driving force standing behind the topological constraint but are instead followers adjusting to the phase-dependence of the model parameters in such a way that charge conservation for particles and holes is fulfilled.

Besides the explicit solution for all eigenstates of the considered class of tight-binding models and the rigorous proof of the central results (4.1) and (4.4), we elaborate on a number of further issues in this work. Besides the standard case where the wavelength  $\lambda$  of the modulations is identical to the length  $Za$  of the unit cell, we will also discuss the case  $\lambda = Za/p$  with a rational number  $Z/p$ . This is a standard choice within the discussion of the IQHE and is also covered by our analysis for the boundary charge. We find that the topological constraint (4.1) and the universal form of the phase-dependence of the boundary charge is not changed but, in accordance with the IQHE [[54]-[56]], we find that the Chern number is given by the Diophantine equation  $pC_\nu = \nu - s_\nu Z$ .

Another result concerns a useful universal relationship between the boundary charges  $Q_B^{R,(\alpha)}$  and  $Q_B^{L,(\alpha)}$  of a single band for a half-infinite system with a left/right boundary, respectively. Here, the system with a left boundary is starting with site  $j = 1$  of the unit cell, whereas the one with a right boundary ends with  $j = Z - 1$  (see Fig. 22). We note that this special relation between the left and the right boundary corresponds to the situation where, in each gap and for each value of the phase  $\varphi$ , there is always exactly one edge state either belonging to the left or to the right boundary, see also Refs. [[51, 56]]. In this case we obtain

$$Q_B^{R,(\alpha)} + Q_B^{L,(\alpha)} = -\frac{Z-1}{Z}, \quad (4.5)$$

which, in terms of the corresponding Zak-Berry phases, is equivalent to  $\bar{\gamma}_\alpha^R + \bar{\gamma}_\alpha^L = 0$  according to (4.4).

Besides the boundary charge we discuss also the density in the insulating regime and analyse its localization which is essential for a unique definition of the boundary charge via a macroscopic average. In accordance with similar results for continuum systems [48] we find that the density of an insulator falls off exponentially fast from the boundary to the bulk value of the infinite system. As in Ref. [[48]] we show that each edge state leaves a fingerprint in the scattering state density which has exactly the same form as the edge state but with opposite sign, see also Ref. [[47]]. Technically this arises from a pole contribution of the Friedel density in complex quasimomentum space, where the Friedel density is that part which leads to the well-known  $2k_F$ -oscillations for impurities in metallic systems [63]. The remaining part of the density follows from a branch cut contribution which leads to an exponentially decreasing contribution with localization length proportional to the product of a typical velocity and

the inverse gap between the valence and conduction band. In addition to previous results for continuum systems, we show that a pre-exponential function occurs falling off generically with  $1/\sqrt{n}$  at large distances, where  $n$  labels the unit cells. This result looks very similar to a corresponding power-law found for the one-particle density matrix of an infinite system, see Ref. [[49]].

We note that the discussion of the phase-dependence of the bulk polarization and the pumped charge is quite common in the literature, see e.g. Refs. [[50, 56, 51]] and other works on generalized twisted boundary conditions [44, 45]. However, to the best of our knowledge, a unique relation between the boundary charge and the bulk polarization has never been stated, and *universal* properties of the phase-dependence have been reported only with respect to a complete cycle when the phase is changed by  $2\pi$ . The latter is related to the Chern number or to the number of windings of the edge state pole around the branch cuts mentioned above [56, 43, 61]. In contrast, the quantized winding number  $w_\alpha(\varphi) = -I_\alpha(\varphi) = -\Delta Q_B^{(\alpha)}(\varphi) + \frac{1}{Z}$ , introduced in the accompanying letter [52], describes universal features for any given value of the phase  $\varphi \in [0, 2\pi]$  and has a direct relation to the boundary charge. Therefore, in this work, we will also present a detailed comparison of this winding number to the Chern number and the Zak-Berry phase and we will show explicitly that it contains more information compared to the topological invariants discussed so far in the literature.

The work is organized as follows. After introducing the model in Section 2.1 we present in Sections 2.2 and 2.3 the solution of the Bloch states and the energy dispersion for the infinite system, together with the analytic continuation to complex quasimomentum. The solutions for the scattering and edge states for the half-infinite system are provided in Sections 2.4 and 2.5. Based on the explicit conditions how to determine the edge states we present in Section 2.6 a rigorous derivation of the topological constraint (4.1) for the phase-dependence of the edge state energies and present rules for its visualization. The definition of the boundary charge and the derivation of the unique relation (4.5) between the boundary charges of half-infinite systems with a left or right boundary is described in Section 3.1. The particle-hole duality is reviewed in Section 3.2. Section 3.3 is devoted to the calculation of the density for a half-infinite insulator, where we demonstrate the localization of the boundary charge and derive the power-law for the pre-exponential function. In Section 4. we present the unique formulation of the surface charge theorem (4.4) and discuss the Zak-Berry phases in the two representations (4.2) and (4.3) of the Bloch wave. Furthermore, we derive the universal relation between the change of the boundary charge of a single band under a shift of the lattice by an arbitrary number of sites and the winding number associated with the phase difference of the Bloch wave function between the corresponding sites. Section 5. presents the comparison of this winding number to the Zak-Berry phase and the Chern number. The physical picture underlying the universal properties of the boundary charge is the topic of Section 6.1 which reviews the derivation proposed in Ref. [[52]]. In addition, via the analytic continuation of Bloch states, we will discuss in this section why the phase-dependence of the model parameters can always be chosen such that no edge state crosses the chemical potential in a certain gap. The rigorous derivation for the possible quantization values of the topological invariant of a single band and for the invariant in the presence of a fixed chemical potential, together with the relation to the topological constraint (4.1), is presented in Sections 6.2 and 6.3, respectively. These Sections contain also the derivation of the consequences for the phase-dependence of the boundary charge and the case where the wave-length of the modulations is given by a rational number  $Z/p$  in units of the lattice spacing  $a$ . We close with a summary and outlook in Section 7..

Throughout this work we use units  $\hbar = e = a = 1$ .



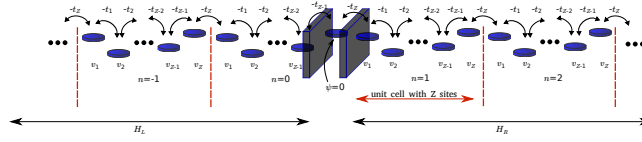


Figure 22: Sketch of the two half-infinite tight-binding models under consideration. The system is either extending to the right or to the left side, described by the Hamiltonians  $H_R$  and  $H_L$ , respectively. The unit cells are labelled by  $n = 0, \pm 1, \pm 2, \dots$ , the sites within a unit cell by  $j = 1, 2, \dots, Z$ . The absolute position of a site is labelled by  $m = Z(n - 1) + j$ . The right/left bar indicates the boundary of  $H_{R/L}$ . The site indicated between the two bars at  $n = 0, j = Z$  or  $m = 0$  is an artificial site where the eigenstate has to fulfil the boundary condition  $\psi(n = 0, j = Z) = \psi(m = 0) = 0$ .

## 2. Spectral properties

In this section we introduce the class of tight-binding models under consideration and present the analytical solution for all Bloch eigenstates of the infinite system and all scattering and edge states of a half-infinite system with a left or a right boundary. The exact result for all eigenstates will form the basis for a rigorous proof of the unique formulation of the surface charge theorem (4.4) in Section 4.. In addition, we present the analytic continuation of the energy dispersion and the Bloch eigenstates to complex quasimomentum, analog to corresponding representations for continuum systems [60, 47]. Of central importance is the determination of the conditions for the appearance of edge states and the analysis of their phase-dependence when the lattice is shifted by a continuous phase variable  $\varphi$ . We develop a graphical representation to determine the precise topological constraints (4.1) for the edge states in Section 2.6 which will turn out to be the basis for the rigorous proof of the quantization values of the invariant  $I(\varphi)$  characterizing universal properties of the boundary charge as introduced in Ref. [[52]], see Section 6..

### 2.1 The model

We consider a generic nearest-neighbor tight-binding model with one orbital per site on a half-infinite system, extending either to the right or to the left side, described by Hamiltonians  $H_{R/L}$ , respectively, see Fig. 22(a,b) for a sketch of the system. The unit cells are labelled by  $n = 0, \pm 1, \pm 2, \dots$ . Each unit cell consists of  $Z$  sites labelled by  $j = 1, \dots, Z$ . The absolute position of a site is labelled by the index  $m = Z(n - 1) + j \equiv (n, j)$ . We take generic on-site potentials  $\hat{v}_m \equiv \hat{v}_{nj} \equiv v_j$  and hoppings  $\hat{t}_m \equiv \hat{t}_{nj} \equiv t_j$  that depend only on the site index  $j$  within the unit cell. The Hamiltonians  $H_{R/L}$  in the 1-particle subspace are given by

$$H_R = \sum_{m \geq 1} \{ \hat{v}_m |m\rangle \langle m| - [\hat{t}_m |m\rangle \langle m+1| + \text{h.c.}] \} , \quad (4.6)$$

$$H_L = \sum_{m \leq -1} \{ \hat{v}_m |m\rangle \langle m| - [\hat{t}_m |m\rangle \langle m-1| + \text{h.c.}] \} . \quad (4.7)$$

The difference between  $H_R$  and  $H_L$  is the way the unit cell is cut off at the boundary. Whereas for  $H_R$  the complete unit cell  $n = 1$  is included, for  $H_L$  the site  $j = Z$  of the first unit cell  $n = 0$  is not included. With this choice the boundary condition for both Hamiltonians becomes the same at  $m = 0$  or at  $n = 0$  and  $j = Z$ , see below.

To guarantee  $H_{R/L} = H_{R/L}^\dagger$  the on-site potentials  $v_j = v_j^*$  must be real. The hoppings  $t_j = t_j^* > 0$  are chosen real and positive since possible phases can be gauged away by a unitary transformation which does not influence the density, see Appendix 8.1. As a consequence the Hamiltonian is real and is invariant under time reversal transformation  $T = K$  with  $T^2 = 1$ , where  $K$  denotes the operator of complex conjugation in the real-space basis of the lattice

sites. We emphasize that this is *not* a symmetry constraint imposed on top of the considered tight-binding models but can always be achieved after a unitary transformation. Furthermore, this symmetry is not relevant for inducing any topological properties in 1D. According to the classification scheme of topological insulators [[23]-[32]], only chiral symmetry or particle-hole symmetry with  $C^2 = 1$  can induce nontrivial topology in 1D systems. For  $T^2 = 1$  these two symmetries are equivalent and are only fulfilled if the potentials  $v_j$  are all the same. Also inversion symmetry is a special case, where  $v_j = v_{Z-j+1}$  and  $t_j = t_{Z-j}$  (with  $t_0 \equiv t_Z$ ) has to be fulfilled.

For convenience the average of all  $v_j$  is defined as zero energy and we define by  $t$  the average over all  $t_j$

$$\frac{1}{Z} \sum_{j=1}^Z v_j = 0 \quad , \quad \frac{1}{Z} \sum_{j=1}^Z t_j = t. \quad (4.8)$$

All  $v_j = v_j(\varphi)$  and  $t_j = t_j(\varphi)$  are taken as function of a phase variable  $0 \leq \varphi < 2\pi$ , which shifts the lattice continuously towards the boundary for  $H_R$  and away from the boundary for  $H_L$ , such that a phase change by  $\frac{2\pi}{Z}$  corresponds to a shift by one lattice site, i.e.,

$$v_{j+1}(\varphi) = v_j \left( \varphi + \frac{2\pi}{Z} \right) \quad , \quad t_{j+1}(\varphi) = t_j \left( \varphi + \frac{2\pi}{Z} \right). \quad (4.9)$$

Generically, this is achieved by using the form

$$v_j(\varphi) = V F_v \left( \frac{2\pi}{Z} j + \varphi \right), \quad (4.10)$$

$$t_j(\varphi) = t + \delta t F_t \left( \frac{2\pi}{Z} j + \varphi \right), \quad (4.11)$$

where  $F_v(\varphi) = F_v(\varphi + 2\pi)$  and  $F_t(\varphi) = F_t(\varphi + 2\pi)$  are some real and periodic functions of order of  $O(1)$ . In Appendix 8.2 we describe different ways how we have chosen generic and random forms for the two functions  $F_v$  and  $F_t$  used in many figures.

For the case where the wavelength of the modulations is given by some rational and non-integer number  $\lambda = Z/p$ , with  $p$  being some positive integer, the phase-dependence is chosen as

$$v_j(\varphi) = V F_v \left( \frac{2\pi p}{Z} j + \varphi \right), \quad (4.12)$$

$$t_j(\varphi) = t + \delta t F_t \left( \frac{2\pi p}{Z} j + \varphi \right). \quad (4.13)$$

In this form the parameters are again periodic under a phase change by  $2\pi$  but a shift of the lattice by one site happens on the different scale  $2\pi p/Z$  such that after a phase change by  $2\pi p$  the system has undergone all possible ways of how to define the boundary. We can map this parametrization on the form of Eqs. (4.12) and (4.13) by the rescaling  $\varphi' \equiv \varphi/p$ ,  $F'_v(\varphi') = F_v(\varphi)$  and  $F'_t(\varphi') = F_t(\varphi)$  such that the parameters as function of  $\varphi'$  get the same form as above

$$v_j(\varphi') = V F'_v \left( \frac{2\pi}{Z} j + \varphi' \right), \quad (4.14)$$

$$t_j(\varphi) = t + \delta t F'_t \left( \frac{2\pi}{Z} j + \varphi' \right). \quad (4.15)$$

Therefore this case is also covered by our general ansatz and will not be treated separately.



The eigenfunctions of the half-infinite tight-binding models (4.6) and (4.7) will be constructed in terms of the Bloch eigenstates of the infinite system defined by the bulk Hamiltonian written in compact form as

$$H_{\text{bulk}} = \sum_{n=-\infty}^{\infty} \{ |n\rangle\langle n| \otimes h(0) + |n+1\rangle\langle n| \otimes h(1) + |n\rangle\langle n+1| \otimes h(-1) \} , \quad (4.16)$$

where we sum over all unit cells  $n = -\infty, \dots, \infty$ . Here,  $h(0)$  and  $h(\pm 1)$  are  $Z \times Z$ -matrices in unit cell space describing the Hamiltonian within a unit cell or the hopping from unit cell  $n \rightarrow n+1$  or  $n+1 \rightarrow n$ , respectively. They are given by (zero matrix elements are not shown)

$$h(0) = \sum_{j=1}^Z v_j |j\rangle\langle j| - \sum_{j=1}^{Z-1} (t_j |j+1\rangle\langle j| + \text{h.c.}) \\ \equiv \begin{pmatrix} v_1 & -t_1 & & & \\ -t_1 & v_2 & -t_2 & & \\ & -t_2 & \ddots & \ddots & \\ & & \ddots & \ddots & -t_{Z-1} \\ & & & -t_{Z-1} & v_Z \end{pmatrix} \quad (4.17)$$

and

$$h(1) = -t_Z |1\rangle\langle Z| = \begin{pmatrix} & -t_Z \\ & \end{pmatrix} = [h(-1)]^\dagger . \quad (4.18)$$

The bulk Hamiltonian is translationally invariant and can be diagonalized by the Fourier transform

$$|k\rangle = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} e^{ikn} |n\rangle, \quad (4.19)$$

where  $-\pi \leq k < \pi$  is the quasimomentum. We get

$$H_{\text{bulk}} = \int_{-\pi}^{\pi} dk |k\rangle\langle k| \otimes h_k, \quad (4.20)$$

with the Bloch Hamiltonian

$$h_k = h(0) + e^{-ik} h(1) + e^{ik} h(-1) \quad (4.21)$$

$$= \begin{pmatrix} v_1 & -t_1 & & & -t_Z e^{-ik} \\ -t_1 & v_2 & -t_2 & & \\ & -t_2 & \ddots & \ddots & \\ & & \ddots & \ddots & -t_{Z-1} \\ -t_Z e^{ik} & & & -t_{Z-1} & v_Z \end{pmatrix}. \quad (4.22)$$

For later convenience we write  $h_k$  in the form

$$h_k = \begin{pmatrix} A & b_k \\ b_{-k}^T & v_Z \end{pmatrix}, \quad (4.23)$$

where we have defined

$$A = \begin{pmatrix} v_1 & -t_1 & & & \\ -t_1 & v_2 & -t_2 & & \\ & -t_2 & \ddots & \ddots & \\ & & \ddots & \ddots & -t_{Z-1} \\ & & & -t_{Z-1} & v_{Z-1} \end{pmatrix} \quad (4.24)$$

and the  $(Z - 1)$ -dimensional column vector

$$b_k = \begin{pmatrix} -t_Z e^{-ik} \\ 0 \\ \vdots \\ 0 \\ -t_{Z-1} \end{pmatrix}. \quad (4.25)$$

Once an eigenstate  $|\psi\rangle$  for the Hamiltonian  $H_{\text{bulk}}$  of the infinite system has been found

$$H_{\text{bulk}}|\psi\rangle = \epsilon|\psi\rangle, \quad (4.26)$$

we get also an eigenstate of the half-infinite systems  $H_{R/L}$  if the boundary condition

$$\psi(n = 0, j = Z) = \psi(m = 0) = 0 \quad (4.27)$$

and the asymptotic condition

$$\lim_{n \rightarrow \infty} \psi(n, j) \sim e^{ikn}, \quad \text{Im}(k) \geq 0 \quad (4.28)$$

for  $H_R$  and

$$\lim_{n \rightarrow -\infty} \psi(n, j) \sim e^{ikn}, \quad \text{Im}(k) \leq 0 \quad (4.29)$$

for  $H_L$  are fulfilled. As a consequence, we will see in Section 2.5 that, for given phase  $\varphi$ , we find always exactly one edge state, either for  $H_R$  or  $H_L$ , consistent with the study in Ref. [[56]] for the TKNN model [51].

## 2.2 Bloch eigenstates for the infinite system

We first determine the eigenstates of the infinite system

$$H_{\text{bulk}}|\psi_{k,\text{bulk}}^{(\alpha)}\rangle = \epsilon_k^{(\alpha)}|\psi_{k,\text{bulk}}^{(\alpha)}\rangle, \quad (4.30)$$

where  $\psi_{k,\text{bulk}}^{(\alpha)}$  is the Bloch eigenstate of band  $\alpha = 1, \dots, Z$ . We take the Bloch form

$$\psi_{k,\text{bulk}}^{(\alpha)}(n, j) = \frac{1}{\sqrt{2\pi}} \chi_k^{(\alpha)}(j) e^{ikn}, \quad (4.31)$$

where  $\chi_k^{(\alpha)} = \chi_{k+2\pi}^{(\alpha)}$  are the normalized Bloch states described by  $Z$ -dimensional column vectors, which are chosen periodic in  $k$  (for other representations of the Bloch wave function see the discussion in Section 4.). We will also choose the gauge such that  $\chi_k^{(\alpha)}(Z)$  is real (which will be relevant for the boundary condition for a half-infinite system, see Sections 2.4 and 2.5). They are eigenstates of the Bloch Hamiltonian  $h_k$

$$h_k \chi_k^{(\alpha)} = \epsilon_k^{(\alpha)} \chi_k^{(\alpha)}. \quad (4.32)$$

We note that  $h_k$  is diagonalizable for all complex  $k$  except at special branching points of the dispersion where a pair of two eigenstates merge together, see Section 2.3. Since  $(h_k)^T = h_{-k}$ , we find that  $(\chi_{-k}^{(\alpha)})^T$  are the left eigenvectors of  $h_k$  for any complex  $k$

$$(\chi_{-k}^{(\alpha)})^T h_k = \epsilon_{-k}^{(\alpha)} (\chi_{-k}^{(\alpha)})^T, \quad (4.33)$$

Since the eigenvalues of left and right eigenvectors are the same, we get

$$\epsilon_k^{(\alpha)} = \epsilon_{-k}^{(\alpha)} \quad (4.34)$$

for any complex  $k$ . The orthogonality and completeness relation can be written as

$$(\chi_{-k}^{(\alpha)})^T \chi_k^{(\alpha')} = \delta_{\alpha\alpha'} \quad , \quad \sum_{\alpha=1}^Z \chi_k^{(\alpha)} (\chi_{-k}^{(\alpha)})^T = \mathbb{1}. \quad (4.35)$$

Furthermore, since  $(h_k)^* = h_{-k^*}$  and  $h_k = h_{k+2\pi}$  for any complex  $k$ , and using the fact that the spectrum is non-degenerate (see below), we get for any complex  $k$

$$\epsilon_k^{(\alpha)} = \left( \epsilon_{-k^*}^{(\alpha)} \right)^*, \quad (4.36)$$

$$\epsilon_k^{(\alpha)} = \epsilon_{k+2\pi}^{(\alpha)}, \quad (4.37)$$

and we can choose the gauge such that

$$\chi_k^{(\alpha)}(j) = \left( \chi_{-k^*}^{(\alpha)}(j) \right)^*, \quad (4.38)$$

$$\chi_k^{(\alpha)}(j) = \chi_{k+2\pi}^{(\alpha)}(j). \quad (4.39)$$

We note that the normalization and orthogonality defined in terms of the left and right eigenvectors is very essential to find a convenient analytic continuation to complex quasimomentum. To construct the exact eigenstates for any complex  $k$  we use the ansatz

$$\chi_k^{(\alpha)} = \frac{1}{\sqrt{N_k^{(\alpha)}}} \begin{pmatrix} a_k^{(\alpha)} \\ s(\epsilon_k^{(\alpha)}) \end{pmatrix}, \quad (4.40)$$

where  $a_k^{(\alpha)}$  is a  $(Z-1)$ -dimensional column vector, and

$$N_k^{(\alpha)} = N_{-k}^{(\alpha)} = (a_{-k}^{(\alpha)})^T a_k^{(\alpha)} + \left[ s(\epsilon_k^{(\alpha)}) \right]^2 \quad (4.41)$$

guarantees normalization. The last component  $\sim s(\epsilon_k^{(\alpha)})$  plays a special role and determines the gauge of the Bloch state needed for a precise definition of the Zak-Berry phase to obtain a unique formulation of the surface charge theorem, see Eq. (4.4). We choose a gauge such that  $\chi_k^{(\alpha)}(Z)$  is real for real  $k$ . Furthermore, for any complex  $k$ , we will use a representation such that the normalization  $N_k^{(\alpha)}$  and all components  $a_k^{(\alpha)}(j)$  are *analytic* functions in the complex plane up to branch cuts arising from the dispersion relation  $\epsilon_k^{(\alpha)}$  which occurs as a parameter in all quantities. We show that such a representation is possible since the Hamiltonian  $h_k$  is an analytic function of  $k$ . As we will see this representation has many advantages and we propose it to be useful for the analysis of generic models even going beyond the single-channel case analysed in the present work.

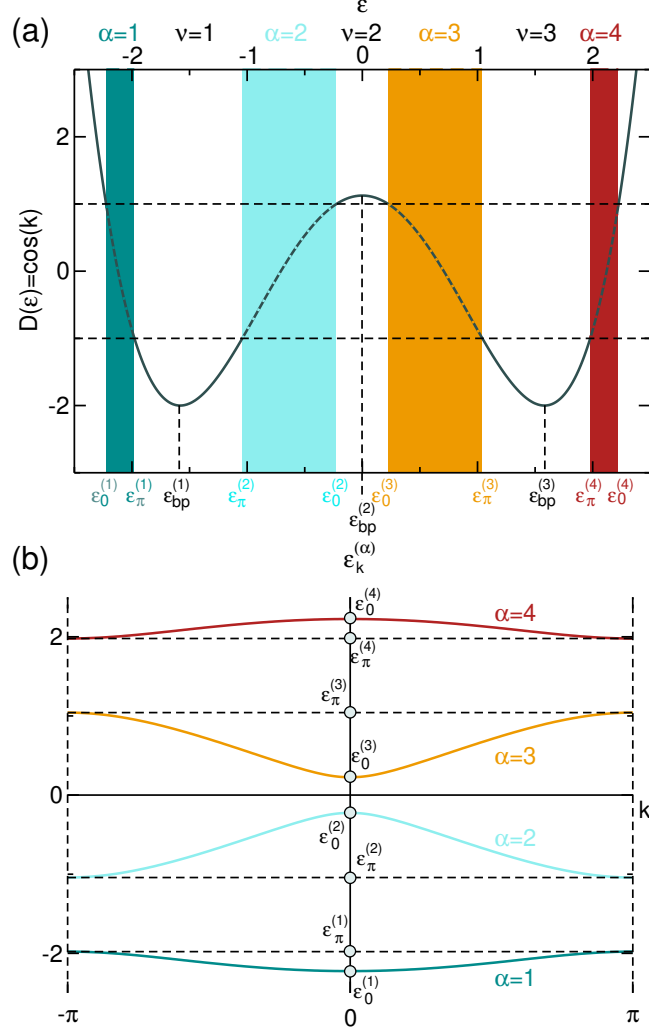


Figure 23: (a) The function  $D(\epsilon)$  for  $v_j = -\cos(2\pi j/Z)$  and  $t_j = 1$ . The bands are formed in the regions where  $D(\epsilon_k^{(\alpha)}) = \cos(k) \leq 1$ , with  $Z - 1$  gaps  $\nu = 1, \dots, Z - 1$  in between. The band edges are given by  $\epsilon_0^{(\alpha)}$  and  $\epsilon_\pi^{(\alpha)}$ , where  $\epsilon_0^{(\alpha)}$  is the band top/bottom for even/odd  $\alpha$ .  $\epsilon_{\text{bp}}^{(\nu)}$  denote the energies where  $\frac{d}{d\epsilon}D(\epsilon_{\text{bp}}^{(\nu)}) = 0$  and  $|D(\epsilon_{\text{bp}}^{(\nu)})| = |\cos(k_{\text{bp}}^{(\nu)})| > 1$ . They correspond to complex  $k$ -values  $k_{\text{bp}}^{(\nu)} = \pm i\kappa_{\text{bp}}^{(\nu)}$  and  $k_{\text{bp}}^{(\nu)} = \pm(\pi + i\kappa_{\text{bp}}^{(\nu)})$  for even and odd  $\nu$ , respectively, with  $\kappa_{\text{bp}}^{(\nu)} \geq 0$ , see Eq. (4.68). (b) The band structure for the same parameters as function of the quasimomentum  $k$ .

Inserting (4.40) into the eigenvalue problem (4.32) and using the form (4.23) of  $h_k$ , we obtain the two equations

$$(A - \epsilon_k^{(\alpha)})a_k^{(\alpha)} = -s(\epsilon_k^{(\alpha)})b_k \quad (4.42)$$

$$b_{-k}^T a_k^{(\alpha)} = -s(\epsilon_k^{(\alpha)})\bar{v}_Z(\epsilon_k^{(\alpha)}), \quad (4.43)$$

where we defined

$$\bar{v}_j(\epsilon) = v_j - \epsilon. \quad (4.44)$$

The first equation can be solved explicitly by

$$a_k^{(\alpha)} = -B(\epsilon_k^{(\alpha)})b_k, \quad B(\epsilon) = \frac{s(\epsilon)}{A - \epsilon}. \quad (4.45)$$

The matrix  $B(\epsilon)$  is a well-defined matrix even for  $\det(A - \epsilon) = 0$  if we take

$$s(\epsilon) = \det(A - \epsilon). \quad (4.46)$$

For this choice, the matrix elements  $B_{jj'}(\epsilon)$  are given by  $(-1)^{j+j'}$  times the subdeterminant of  $A - \epsilon$  where the row  $j'$  and the column  $j$  are omitted. Examining these subdeterminants based on the definition (4.24) of the  $A$ -matrix one finds for  $1 \leq j \leq j' \leq Z - 1$

$$\begin{aligned} B_{jj'}(\epsilon) &= B_{j'j}(\epsilon) \\ &= (t_j t_{j+1} \cdots t_{j'-1}) d_{1,j-1}(\epsilon) d_{j'+1,Z-1}(\epsilon), \end{aligned} \quad (4.47)$$

where  $t_j t_{j+1} \cdots t_{j'-1} \equiv 1$  for  $j = j'$ , and the determinants  $d_{ij}(\epsilon)$  are defined for  $i \leq j$  by

$$d_{ij}(\epsilon) = \det(A^{(ij)} - \epsilon), \quad (4.48)$$

with

$$A^{(ij)} = \begin{pmatrix} v_i & -t_i & & & \\ -t_i & v_{i+1} & -t_{i+1} & & \\ & -t_{i+1} & \ddots & \ddots & \\ & & \ddots & \ddots & -t_{j-1} \\ & & & -t_{j-1} & v_j \end{pmatrix}. \quad (4.49)$$

By convention, we define  $d_{ij} = 1$  for  $i > j$ . Using this result in Eq. (4.45), one obtains with Eq. (4.25) the explicit solution for the components  $j = 1, \dots, Z - 1$  of the Bloch state

$$a_k^{(\alpha)} = f(\epsilon_k^{(\alpha)})e^{-ik} + g(\epsilon_k^{(\alpha)}), \quad (4.50)$$

where  $f(\epsilon)$  and  $g(\epsilon)$  are  $(Z - 1)$ -dimensional column vectors with components

$$f_j(\epsilon) = t_1 \cdots t_{j-1} t_Z d_{j+1,Z-1}(\epsilon), \quad (4.51)$$

$$g_j(\epsilon) = \frac{\bar{t}^Z}{t_1 \cdots t_{j-1} t_Z} d_{1,j-1}(\epsilon). \quad (4.52)$$

Here,  $\bar{t}$  is defined by the geometric mean

$$\bar{t} \equiv (t_1 t_2 \cdots t_Z)^{1/Z}. \quad (4.53)$$

The determinants  $d_{ij}$  can be calculated from the recursion relations (we omit the argument  $\epsilon$ )

$$d_{ij} = \bar{v}_i d_{i+1,j} - t_i^2 d_{i+2,j} \quad (4.54)$$

$$= \bar{v}_j d_{i,j-1} - t_{j-1}^2 d_{i,j-2}, \quad (4.55)$$

together with  $d_{ii} = \bar{v}_i$ . The two determinants  $d_{2,Z-1}$  and  $d_{1,Z-2}$  will play a special role for the determination of the edge states, see Section 2.5. For later convenience, we define

$$\tilde{d}_{2,Z-1}(\epsilon) = \frac{t_Z^2}{\bar{t}^Z} d_{2,Z-1}(\epsilon), \quad (4.56)$$

$$\tilde{d}_{1,Z-2}(\epsilon) = \frac{t_{Z-1}^2}{\bar{t}^Z} d_{1,Z-2}(\epsilon). \quad (4.57)$$

Further useful properties of the determinants are listed and proven in Appendix 8.3. This Appendix contains also very helpful identities for the derivatives of the determinants and the  $B$ -matrix. In particular we find the relation

$$f(\epsilon)^T \frac{d}{d\epsilon} g(\epsilon) = g(\epsilon)^T \frac{d}{d\epsilon} f(\epsilon), \quad (4.58)$$

which will be needed in Section 4. to show the relation between the Friedel charge and the Zak-Berry phase.

Furthermore, we show in Appendix 8.3 that one can write all components  $a_k^{(\alpha)}(j)$ , with  $j = 2, \dots, Z-1$ , in terms of  $a_k^{(\alpha)}(1)$  via

$$\begin{aligned} t_1 \cdots t_{j-1} a_k^{(\alpha)}(j) &= d_{1,j-1}(\epsilon_k^{(\alpha)}) a_k^{(\alpha)}(1) \\ &\quad - d_{2,j-1}(\epsilon_k^{(\alpha)}) s(\epsilon_k^{(\alpha)}) t_Z e^{-ik}, \end{aligned} \quad (4.59)$$

where  $a_k^{(\alpha)}(1)$  follows from (4.50) as

$$a_k^{(\alpha)}(1) = \frac{\bar{t}^Z}{t_Z} \left\{ \tilde{d}_{2,Z-1}(\epsilon_k^{(\alpha)}) e^{-ik} + 1 \right\}. \quad (4.60)$$

This component will play a central role to define the quantized invariant in Section 6.2. Since we work in a gauge where  $\chi_k^{(\alpha)}(Z)$  is real for real  $k$ , we note that its phase corresponds to the gauge invariant phase-difference of the Bloch state  $\chi_k^{(\alpha)}(j)$  between the first and last site of a unit cell.

We note that Eq. (4.43) is not an independent equation and is automatically fulfilled when  $\epsilon_k^{(\alpha)}$  is an eigenvalue of  $h_k$ , i.e., when  $\det(h_k - \epsilon_k^{(\alpha)}) = 0$ . This follows from the relation

$$\det(h_k - \epsilon) = -b_{-k}^T B(\epsilon) b_k + s(\epsilon) \bar{v}_Z(\epsilon), \quad (4.61)$$

together with (4.45).

Eqs. (4.41), (4.46), (4.50), (4.51), and (4.52) show that all quantities  $N_k^{(\alpha)}$ ,  $a_k^{(\alpha)}(j)$ , and  $s(\epsilon_k^{(\alpha)})$  can be written as analytic functions  $\mathcal{F}(\epsilon, k)$  in  $\epsilon$  and  $k$ , with  $\epsilon \equiv \epsilon_k^{(\alpha)}$ . Therefore, the analytic continuation of the eigenstates follows straightforwardly from the analytic continuation of the dispersion  $\epsilon_k^{(\alpha)}$  which will be discussed in the next subsection. Only at the special points  $N_k^{(\alpha)} = 0$  additional poles appear for the Bloch state. As we will see in Section 2.5 they are related to edge states, see also Ref. [[47]] with similar results for continuum models.



Figure 24: Two choices for the analytic continuation of the dispersion for the same parameters as in Fig. 23. (a) The analytic continuation defining  $\epsilon_k$ , with  $\epsilon_k = \epsilon_k^{(\alpha)}$  for  $(\alpha - 1)\pi < |k| < \alpha\pi$  and  $\epsilon_k = \epsilon_k^{(Z)}$  for  $|k| > Z\pi$  on the real axis. Branch cuts occur at  $\pm\nu\pi + i\kappa$ , with  $|\kappa| < \kappa_{\text{bp}}^{(\nu)}$  and  $\nu = 1, \dots, Z - 1$ . (b) and (c) show the analytic continuation of  $\epsilon_k^{(\alpha)}$  for  $\alpha = 2, 3$ , with  $\epsilon_k^{(\alpha)} = \epsilon_{k+2\pi}^{(\alpha)}$  taken on the whole real axis. Branch cuts are located at  $k_{\text{bc}}^{(\nu)} + i\kappa$  and  $(k_{\text{bc}}^{(\nu)})^* - i\kappa$  (not shown), with  $|\kappa| > 0$  and  $\nu = 1, 2$  (for  $\alpha = 2$ ) and  $\nu = 2, 3$  (for  $\alpha = 3$ ), together with corresponding ones shifted by multiples of  $2\pi$ . Between the branch cuts we have indicated in the boxes the relation to the analytic continuation chosen in (a). As one can see the values of the dispersion left and right to the branch cut starting at  $k_{\text{bp}}^{(2)}$  are interchanged for  $\alpha = 2, 3$ .

### 2.3 Energy dispersion

To obtain the dispersion  $\epsilon_k^{(\alpha)}$ , we rewrite the condition  $\det(h_k - \epsilon) = 0$  by using (4.61) and inserting the form (4.25) of  $b_k$  together with the matrix elements (4.47) of the  $B$ -matrix. Using the recursion relation  $d_{1Z} = \bar{v}_Z d_{1,Z-1} - t_{Z-1}^2 d_{1,Z-2}$ , one finds after a straightforward calculation the condition

$$\cos(k) = D(\epsilon), \quad (4.62)$$

$$D(\epsilon) \equiv \frac{1}{2t^Z} \{d_{1Z}(\epsilon) - t_Z^2 d_{2,Z-1}(\epsilon)\}. \quad (4.63)$$

In Appendix 8.3 we prove many helpful representations for the function  $D(\epsilon)$  and its derivatives.  $D(\epsilon)$  is a polynomial of degree  $Z$  in  $\epsilon$  with real coefficients and the asymptotic behaviour

$$\lim_{|\epsilon| \rightarrow \infty} D(\epsilon) = \frac{(-\epsilon)^Z}{2t^Z}. \quad (4.64)$$

Therefore, for any given complex  $k$ , Eq. (4.62) has  $Z$  solutions  $\epsilon = \epsilon_k^{(\alpha)}$ , with  $\alpha = 1, \dots, Z$ , which fulfil the properties (4.34), (4.37) and (4.36). For real values  $-\pi \leq k < \pi$ , the Hamiltonian  $h_k$  is hermitian and has always  $Z$  real eigenvalues  $\epsilon_k^{(\alpha)}$ , with  $\alpha = 1, \dots, Z$ , corresponding to the different bands which we label from bottom to top. In this case, the graphical solution of Eq. (4.62) is shown in Fig. 23(a). From Eq. (4.64) we get  $D(\epsilon) > 0$  for  $\epsilon \rightarrow -\infty$ . Therefore, the bottom of the first band is always at  $k = 0$ . Since for each  $k$  there are  $Z$  real solutions of  $D(\epsilon) = \cos(k)$ , the function  $D(\epsilon)$  must be monotonous in each segment where  $|\cos(k)| \leq 1$ . As a consequence, we obtain always  $Z - 1$  gaps labelled by  $\nu = 1, \dots, Z - 1$  and the band dispersion  $\epsilon_k^{(\alpha)}$  for  $\alpha$  even/odd is a monotonously decreasing/increasing function for  $0 < k < \pi$  with band region defined by  $\epsilon_{\pi/0}^{(\alpha)} < \epsilon_k^{(\alpha)} < \epsilon_{0/\pi}^{(\alpha)}$ , see Fig. 23(b). Occasionally, two adjacent bands might touch at  $k = 0$  or  $k = \pm\pi$  leading to gap closings, but the classification in  $Z$  bands remains.

The analytic continuation of the dispersion can be obtained from Eq. (4.62), analog to Ref. [[60]]. This is achieved by starting from some dispersion  $\epsilon_k$  on the real axis (see below for two convenient choices) and solving a differential equation for  $\frac{d\epsilon_k}{dk}$  along an arbitrary path in the complex plane. Inserting  $\epsilon = \epsilon_k$  in Eq. (4.62) and taking the derivative with respect to  $k$



we get

$$\frac{d\epsilon_k}{dk} = -\frac{\sin(k)}{D'(\epsilon_k)}. \quad (4.65)$$

It follows that the branching points  $k_{\text{bp}}^{(\nu)}$  are given by the condition

$$\cos(k_{\text{bp}}^{(\nu)}) = D(\epsilon_{\text{bp}}^{(\nu)}), \quad (4.66)$$

$$D'(\epsilon_{\text{bp}}^{(\nu)}) \equiv \frac{dD}{d\epsilon}(\epsilon_{\text{bp}}^{(\nu)}) = 0, \quad (4.67)$$

such that  $\frac{d\epsilon_k}{dk} \rightarrow \pm\infty$  diverges for  $k \rightarrow k_{\text{bp}}^{(\nu)}$ . Since  $D'(\epsilon)$  is a polynomial of degree  $Z - 1$ , Eq. (4.67) has  $Z - 1$  solutions  $\epsilon_{\text{bp}}^{(\nu)}$ , with  $\nu = 1, \dots, Z - 1$ . As shown in Fig. 23(a), each  $\epsilon_{\text{bp}}^{(\nu)}$  corresponds to a certain gap and has a real value between the top of band  $\nu$  and the bottom of band  $\nu + 1$ , with  $D(\epsilon_{\text{bp}}^{(\nu)}) = \cos(k_{\text{bp}}^{(\nu)}) \geq 1$  ( $\leq -1$ ) for  $\nu$  even (odd). By convention we define  $k_{\text{bp}}^{(\nu)}$  as

$$k_{\text{bp}}^{(\nu)} = i\kappa_{\text{bp}}^{(\nu)} + \begin{cases} 0 & \text{for } \nu \text{ even} \\ \pi & \text{for } \nu \text{ odd} \end{cases}, \quad (4.68)$$

with  $\kappa_{\text{bp}}^{(\nu)} \geq 0$ . We note that the branching points appear always 4-fold as  $\pm k_{\text{bp}}^{(\nu)}$  and  $\pm(k_{\text{bp}}^{(\nu)})^*$ . Furthermore, each  $k_{\text{bp}}^{(\nu)}$  is defined mod( $2\pi$ ) but if all of these replicas appear depends on the specific initial condition on the real axis which will be discussed in the following.

Expanding Eq. (4.62) around the branching point  $k = k_{\text{bp}}^{(\nu)}$  and using  $\sin(k_{\text{bp}}^{(\nu)}) = i(-1)^\nu \sinh(\kappa_{\text{bp}}^{(\nu)})$  from (4.68) and  $D''(\epsilon_{\text{bp}}^{(\nu)}) = -(-1)^\nu |D''(\epsilon_{\text{bp}}^{(\nu)})|$  from Fig. 23(a) we find (for  $\alpha = \nu$  or  $\alpha = \nu + 1$ )

$$\epsilon_k^{(\alpha)} = \epsilon_{\text{bp}}^{(\nu)} + \left( 2 \frac{\sinh(\kappa_{\text{bp}}^{(\nu)})}{|D''(\epsilon_{\text{bp}}^{(\nu)})|} \right)^{1/2} \sqrt{i(k - k_{\text{bp}}^{(\nu)})}. \quad (4.69)$$

Depending on how the branch cut of the square root is chosen one obtains different ways to define the analytic continuation. Taking the following initial condition on the real axis to solve the differential equation (4.65)

$$\epsilon_k = \epsilon_k^{(\alpha)} \quad \text{for } (\alpha - 1)\pi < |k| < \alpha\pi, \quad (4.70)$$

$$\epsilon_k = \epsilon_k^{(Z)} \quad \text{for } |k| > Z\pi, \quad (4.71)$$

one obtains branch cuts connecting  $\pm\pi + i\kappa_{\text{bp}}^{(\nu)}$  with  $\pm\pi - i\kappa_{\text{bp}}^{(\nu)}$ , separating the band dispersions  $\epsilon_k^{(\nu)}$  and  $\epsilon_k^{(\nu+1)}$  on the real axis, see Fig. 24(a). This defines a common function  $\epsilon_k$  in the complex plane where the different band dispersion are connected analytically. By convention, if the index ( $\alpha$ ) is not written in the following,  $\epsilon_k$  denotes this function in the complex plane. Obviously, in this representation the periodicity condition (4.37) is no longer fulfilled but the properties (4.34) and (4.36) remain valid

$$\epsilon_k = \epsilon_{-k} = (\epsilon_{-k^*})^* \neq \epsilon_{k+2\pi}. \quad (4.72)$$

This choice for the analytic continuation has the advantage that all band dispersions are included but it is not very convenient to use it for calculating integrals  $\int_{-\pi}^{\pi} dk$  by closing the integration contour in the upper half since the integrals around the branch cuts are hard to evaluate.

An alternative way is to define an analytic continuation for each band  $\epsilon_k^{(\alpha)}$  separately by using the initial condition

$$\epsilon_k = \epsilon_k^{(\alpha)} \quad \text{for} \quad -\infty < k < \infty \quad (4.73)$$

on the real axis, with  $\epsilon_k^{(\alpha)} = \epsilon_{k+2\pi}^{(\alpha)}$ . With this choice for the analytic continuation of  $\epsilon_k^{(\alpha)}$  all properties stated in (4.34), (4.36), and (4.37) remain valid for any  $k$  in the complex plane. For each given  $\alpha$ , two branching points at  $k_{\text{bp}}^{(\alpha-1)}$  and  $k_{\text{bp}}^{(\alpha)}$  appear (mod( $2\pi$ ), for  $\alpha = 1, Z$  only one branching point is present). Choosing the branch cut of the square root in Eq. (4.69) on the negative real axis, the branch cuts of  $\epsilon_k^{(\alpha)}$  are pointing into the direction of the positive (negative) imaginary axis if  $\text{Im}(k) > 0$  ( $\text{Im}(k) < 0$ ), see Figs. 24(b,c) for  $\alpha = 2, 3$ .

Since the initial conditions for the two choices of the analytic continuation of  $\epsilon_k$  and  $\epsilon_k^{(\alpha)}$  are the same on the real axis for  $(\alpha - 1)\pi < |k| < \alpha\pi$  and for  $|k| > Z\pi$ , we obtain

$$\epsilon_k = \epsilon_k^{(\alpha)} \quad \text{for} \quad (\alpha - 1)\pi < |\text{Re}(k)| < \alpha\pi, \quad (4.74)$$

$$\epsilon_k = \epsilon_k^{(Z)} \quad \text{for} \quad |\text{Re}(k)| > Z\pi, \quad (4.75)$$

Using in addition  $\epsilon_k^{(\alpha)} = \epsilon_{k+2\pi}^{(\alpha)} = \epsilon_{-k}^{(\alpha)}$  and  $\epsilon_k = \epsilon_{-k}$ , we can relate the analytic continuation of  $\epsilon_k^{(\alpha)}$  to the one of  $\epsilon_k$ , see Fig. 24(b,c). This shows that the values left and right to the common branch cuts of  $\epsilon_k^{(\alpha)}$  and  $\epsilon_k^{(\alpha+1)}$  starting at  $k_{\text{bp}}^{(\alpha)}$  are interchanged for the bands  $\alpha$  and  $\alpha + 1$ . In Section 3.3 we will use this result to show that the branch cut contributions to the Friedel density cancel for adjacent bands.

We note that the differential equation (4.65) determining the analytic continuation can also be written in an alternative way by taking the derivative  $\frac{d}{dk}$  of  $(h_k - \epsilon_k)\chi_k = 0$ . Together with the form (4.21) of  $h_k$  we find

$$\begin{aligned} & -\frac{d\epsilon_k}{dk}\chi_k + (it_Z e^{-ik}|1\rangle\langle Z| - it_Z e^{ik}|Z\rangle\langle 1|)\chi_k \\ & + (h_k - \epsilon_k)\frac{d}{dk}\chi_k = 0. \end{aligned}$$

Multiplying from the left with  $\chi_{-k}^T$  and using (4.35) together with the form (4.40) of  $\chi_k$  we get

$$\frac{d\epsilon_k}{dk} = it_Z \frac{s(\epsilon_k)}{N_k} (e^{-ik}a_{-k}(1) - e^{ik}a_k(1)).$$

Finally, inserting Eq. (4.60) for  $a_k(1)$  we obtain

$$\frac{d\epsilon_k}{dk} = 2\bar{t}^Z \frac{s(\epsilon_k)}{N_k} \sin(k). \quad (4.76)$$

Comparing with (4.65) we find the useful relation

$$N_k = -2\bar{t}^Z s(\epsilon_k) D'(\epsilon_k). \quad (4.77)$$

For band  $\alpha$  and  $-\pi < k < \pi$ , this gives a relation for the sign of  $s(\epsilon_k^{(\alpha)})$  since  $N_k^{(\alpha)} > 0$  and the sign of  $D'(\epsilon_k^{(\alpha)})$  is given by  $(-1)^\alpha$  (see Fig. 23), providing

$$\text{sign}(s(\epsilon_k^{(\alpha)})) = -(-1)^\alpha \quad \text{for} \quad -\pi < k < \pi. \quad (4.78)$$

This result will be used in Section 2.5 to prove that each gap hosts exactly one edge state.

The condition  $N_k = 0$  defines the points where  $\sum_{j=1}^Z [\chi_k(j)]^2$  has a singularity. From Eq. (4.77) it follows that  $s(\epsilon_k) = 0$  or  $D'(\epsilon_k) = 0$  has to be fulfilled at such a point. The latter condition corresponds to the branching points of  $\epsilon_k$ . The condition  $s(\epsilon_k) = 0$  corresponds to the pole positions of edge states, as will be discussed in Section 2.5. We summarize

$$\begin{aligned} s(\epsilon_k) = 0, D'(\epsilon_k) \neq 0 \\ \Leftrightarrow \text{edge pole of } \sum_{j=1}^Z [\chi_k(j)]^2, \end{aligned} \quad (4.79)$$

$$\begin{aligned} s(\epsilon_k) \neq 0, D'(\epsilon_k) = 0 \\ \Leftrightarrow \text{branching point of } \sum_{j=1}^Z [\chi_k(j)]^2. \end{aligned} \quad (4.80)$$

We note that, due to (4.77) and (4.69), the second case leads to  $N(\epsilon_k) \sim \epsilon_k - \epsilon_{\text{bp}}^{(\nu)} \sim \sqrt{i(k - k_{\text{bp}}^{(\nu)})}$  close to the branching point. Therefore, the factor  $1/N(\epsilon_k)$  is integrable and does not have a pole but a branching point at this position. In contrast, the case  $s(\epsilon_k) = D'(\epsilon_k) = 0$  are special points where the edge pole and the branching point of  $\sum_{j=1}^Z [\chi_k(j)]^2$  merge together to a branching pole, where  $N(\epsilon_k) \sim k - k_{\text{bp}}^{(\nu)}$ , see Eq. (4.91) and Appendix 8.4.

## 2.4 Scattering states for half-infinite system

Once we have found the Bloch eigenstates  $\psi_{k,\text{bulk}}^{(\alpha)}(n, j)$  for  $H_{\text{bulk}}$  of the infinite system via (4.30), we can find a scattering eigenstate for  $H_{R/L}$  of the half-infinite system at the same energy  $\epsilon_k^{(\alpha)}$ , given by

$$H_{R/L}|\psi_k^{(\alpha)}\rangle = \epsilon_k^{(\alpha)}|\psi_k^{(\alpha)}\rangle, \quad (4.81)$$

$$\psi_k^{(\alpha)}(n, j) = \frac{1}{\sqrt{2\pi}} \left\{ \chi_k^{(\alpha)}(j) e^{ikn} - \chi_{-k}^{(\alpha)}(j) e^{-ikn} \right\}. \quad (4.82)$$

The boundary condition (4.27) is fulfilled since  $\chi_k^{(\alpha)}(Z) = \chi_{-k}^{(\alpha)}(Z) = s(\epsilon_k^{(\alpha)})/\sqrt{N_k}$  is real, which is the gauge we have used in Section 2.2. The  $k$ -values for the eigenstates of the half-infinite system are restricted to  $0 < k < \pi$ .

We note that the orthogonality and completeness relation of the eigenstates of the infinite system

$$\sum_{n=-\infty}^{\infty} \sum_{j=1}^Z \psi_{k,\text{bulk}}^{(\alpha)}(n, j)^* \psi_{k',\text{bulk}}^{(\alpha')}(n, j) = \delta_{\alpha\alpha'} \delta(k - k'), \quad (4.83)$$

$$\sum_{\alpha=1}^Z \int_{-\pi}^{\pi} dk \psi_{k,\text{bulk}}^{(\alpha)}(n, j) \psi_{k,\text{bulk}}^{(\alpha)}(n', j') = \delta_{nn'} \delta_{jj'} \quad (4.84)$$

implies only the orthogonality relation for the eigenstates of the half-infinite system

$$\sum_{n=-\infty}^{\infty} \sum_{j=1}^Z \psi_k^{(\alpha)}(n, j)^* \psi_{k'}^{(\alpha')}(n, j) = \delta_{\alpha\alpha'} \delta(k - k'), \quad (4.85)$$

since the edge states will also contribute to the completeness relation, see Section 2.5.

The result (4.82) can also be viewed as a consequence of scattering theory: the wave function consists of a superposition of an incoming and an outgoing wave. The boundary condition (4.27) is very simple here since we have considered only one orbital per site. For the multi-channel case, the eigenstates of  $H_{R/L}$  will also involve exponentially decaying parts.

### 2.5 Edge states for half-infinite system

In this section we construct explicitly all edge states for a half-infinite system and discuss the relation of various determinants to the complex quasimomentum corresponding to the edge states. These relations will turn out to be essential to derive the topological constraints for the phase-dependence of the edge state energies in the subsequent Section 2.6.

To find edge states for  $H_{R/L}$ ,

$$H_{R/L}|\psi_{k_e}^e\rangle = \epsilon_{k_e}|\psi_{k_e}^e\rangle, \quad (4.86)$$

$$\psi_{k_e}^e(n, j) = \chi_{k_e}^e(j)e^{ik_en}, \quad (4.87)$$

we look for solutions of Bloch states with  $\text{Im}(k_e) \geq 0$  and  $s(\epsilon_{k_e}) = \chi_{k_e}^e(Z) = 0$  to fulfil the boundary and asymptotic conditions (4.27), (4.28) and (4.29). In contrast to  $\chi_{k_e}$ , we parametrize them with a different normalization factor

$$\chi_{k_e}^e = \frac{1}{\sqrt{N_{k_e}^e}} \begin{pmatrix} a_{k_e} \\ 0 \end{pmatrix}, \quad (4.88)$$

but with the same vector  $a_{k_e}$ . Below we will find that  $e^{-ik_e}$  is real [see Eq. (4.93)] and, therefore, it follows from (4.50) that also the vector  $a_{k_e}$  and the edge state wave function  $\psi_{k_e}^e(n, j)$  are real. The normalization  $N_{k_e}^e$  is defined such that the edge state wave function (4.87) is normalized, which means  $\sum_{n=1}^{\infty} \sum_{j=1}^Z [\psi_{k_e}^e(n, j)]^2 = 1$  for an edge state of  $H_R$  and  $\sum_{n=-\infty}^0 \sum_{j=1}^Z [\psi_{k_e}^e(n, j)]^2$  for an edge state of  $H_L$ , leading to

$$N_{k_e}^e = \text{sign}(\text{Im}(k_e)) \frac{a_{k_e}^T a_{k_e}}{e^{-2ik_e} - 1}. \quad (4.89)$$

In contrast,  $N_{k_e}$  can not be the correct normalization since we get  $N_{k_e} = 0$  from  $s(\epsilon_{k_e}) = 0$  and (4.77), such that  $\sum_{j=1}^Z [\chi_k(j)]^2$  has a pole at  $k = k_e$ . Therefore, the normalization factor  $N_k$  of  $\chi_k$  is *not* analytically continued to the normalization factor  $N_{k_e}^e$  of  $\chi_{k_e}^e$ . The reason is that the Hamiltonian  $h_k$  is non-hermitian for complex  $k$  and has different right and left eigenstates given by  $\chi_k$  and  $\chi_{-k}$ , respectively. These states get a completely different analytic continuation. In particular, at the edge pole,  $\chi_{k_e} \rightarrow \infty$  and  $\chi_{-k_e} \rightarrow 0$ , such that the orthogonality and completeness relation (4.35) remain valid. In Appendix 8.4 we will show that an expansion of  $N_k$  around the edge pole  $k_e$  gives the result

$$N_k = \text{sign}(\text{Im}(k_e)) i N_{k_e}^e (k - k_e) + O(k - k_e)^2, \quad (4.90)$$

which will be used in Section 3.3 to prove that the edge state density is cancelled by the pole contribution of the Friedel density of that band which belongs to this edge state (see below for the definition of this correspondence). This result applies as long as the edge pole is isolated. If it agrees with a branching point  $k_e = k_{\text{bp}}^{(\nu)}$  of a certain gap  $\nu$ , we show in Appendix 8.4 that (4.90) gets an additional factor 2 and the corrections are of  $O(k - k_e)^{3/2}$

$$N_k = \text{sign}(\text{Im}(k_e)) 2i N_{k_e}^e (k - k_e) + O(k - k_e)^{3/2}. \quad (4.91)$$

This will be needed to show in Section 3.3 that the pole contribution of the Friedel density of band  $\alpha = \nu$  or band  $\alpha = \nu + 1$  cancels only half of the edge state density.

Since  $s(\epsilon_{k_e}) = 0$  we get from (4.42)

$$A a_{k_e} = \epsilon_{k_e} a_{k_e}. \quad (4.92)$$

This eigenvalue problem for the  $(Z - 1)$ -dimensional and hermitian matrix  $A$  has exactly  $Z - 1$  solutions with real eigenvalues  $\epsilon_{k_e}$ . Therefore, we find  $Z - 1$  edge states, each of them either

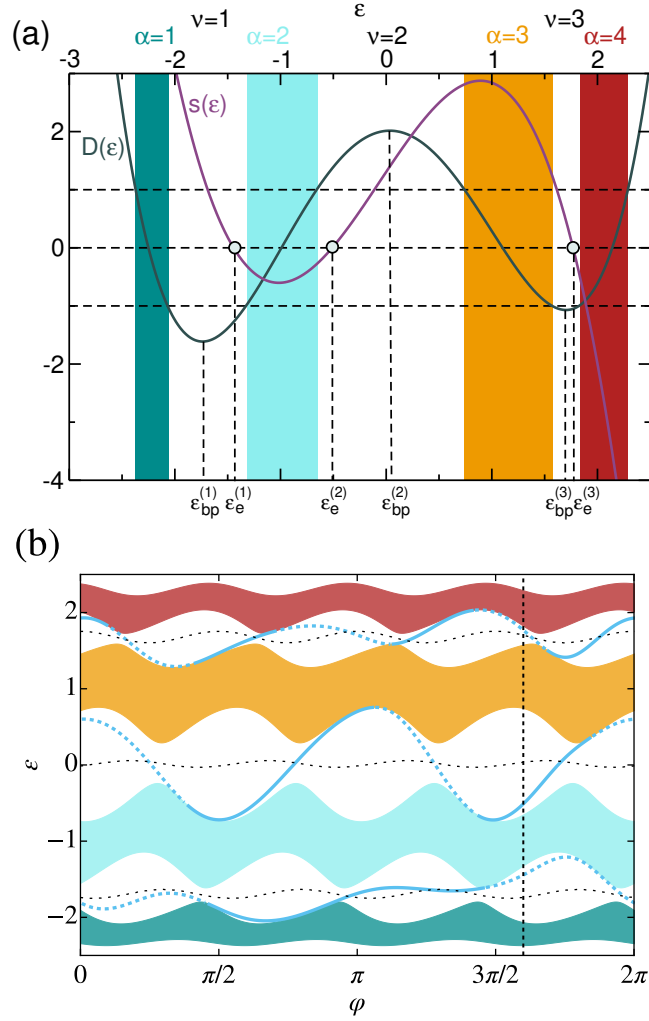


Figure 25: (a) Determination of the edge state energies  $\epsilon_e^{(\nu)}$  from the condition  $s(\epsilon_e^{(\nu)}) = 0$  for  $Z = 4$ ,  $V = 0.5$ ,  $t = 1.1$ ,  $\delta t = 0.1$ ,  $\varphi = 1.6\pi$ , and three random Fourier coefficients for the real functions  $F_v$  and  $F_t$  in Eqs. (4.10) and (4.11) according to the form (4.273), see Supplemental Material for the precise parameters [26]. As explained in the main text the position of the edge state energies is always located in the gaps, each gap hosting exactly one edge state. The sign of  $s(\epsilon)$  is given by  $(-1)^{\alpha+1}$  in the energy regions of band  $\alpha$ , see Eq. (4.78). Depending on whether  $\epsilon_e^{(\nu)} \leq \epsilon_{bp}^{(\nu)}$ , the edge states result from the analytic continuation of band  $\nu$  or  $\nu + 1$ . (b) The band structure as function of the phase variable  $\varphi$  for the same parameters as in (a). The band structure is periodic under a change of the phase variable by  $2\pi/Z$ . The edge state energies  $\epsilon_e^{(\nu)}(\varphi)$  for  $\nu = 1, 2, 3$  are shown by blue solid/dashed lines corresponding to edge states of  $H_R/H_L$  with  $\text{Im}(k_e^{(\nu)}) \geq 0$ . As can be seen each gap hosts exactly one edge state and the edge states change the boundary when they touch the bands. Approximately in the middle of each gap we show by black dashed lines the phase-dependence of the energies  $\epsilon_{bp}^{(\nu)}(\varphi)$  at the branching points for  $\nu = 1, 2, 3$ . For  $\epsilon_e^{(\nu)} \leq \epsilon_{bp}^{(\nu)}$  the edge states belong to the analytical continuation of band  $\nu$  or  $\nu + 1$ , respectively. In (b) we have indicated by the vertical dashed line the phase value  $\varphi = 1.6\pi$  used in (a) and where the analytic structure and the position of the edge poles is shown in Fig. 26.

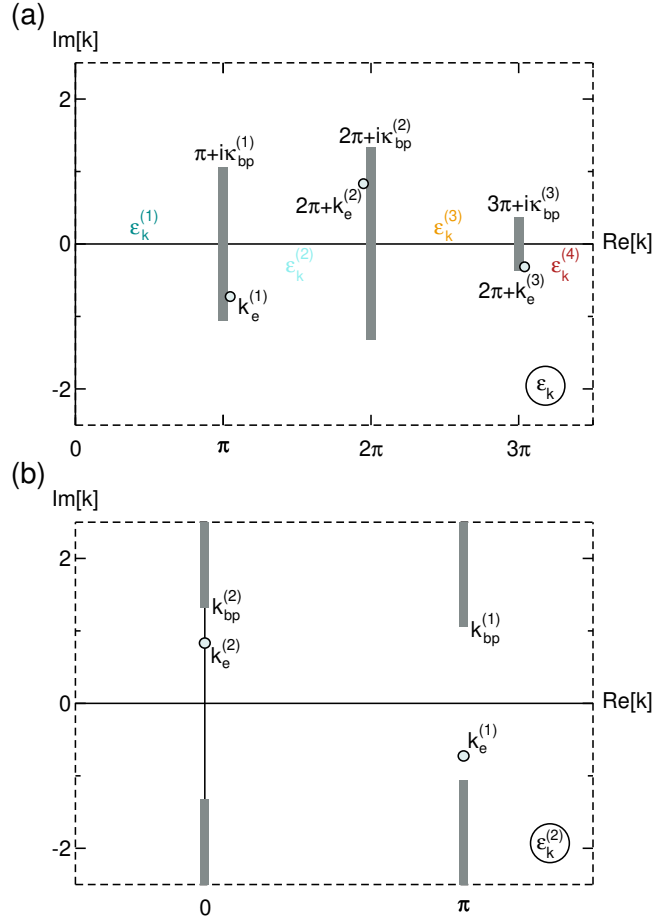


Figure 26: The analytic continuation of (a)  $[\chi_k(j)]^2$  and (b)  $[\chi_k^{(2)}(j)]^2$  analog to Fig. 24(a,b) but for the parameters of Fig. 25 at the particular phase  $\varphi = 1.6\pi$  indicated in Fig. 25(b) by a vertical dashed line. In addition we have indicated the edge pole positions  $k_e^{(\nu)}$ , with  $\nu = 1, 2, 3$ . For  $\text{Im}(k_e^{(\nu)}) \geq 0$ , they correspond to edge states of either  $H_R$  or  $H_L$ , respectively. In (a) an edge pole lying left (right) to the branch cut located at  $\nu\pi$  belongs to the analytic continuation of band  $\nu$  ( $\nu + 1$ ). The edge poles move around the branch cuts as function of  $\varphi$ , see snapshots and videos provided in the Supplemental Material [26].

corresponding to  $H_R$  or  $H_L$  depending on the sign of  $\text{Im}(k_e)$ . This sign will be determined below, see Eqs. (4.103) and (4.104). The energies of these  $Z - 1$  edge states are distributed among the  $Z - 1$  gaps  $\nu = 1, \dots, Z - 1$  of the bulk spectrum, each gap hosting exactly one edge state. This follows from (4.78) since this equation implies that the sign of  $s(\epsilon_k^{(\alpha)})$  is alternating with the band index  $\alpha$  such that in each gap there must be at least one solution with  $s(\epsilon_{k_e}) = \det(A - \epsilon_{k_e}) = 0$ , see Fig. 25(a) for illustration; also cf. Appendix C of Ref. [[61]] for an alternative proof. Therefore, we label the edge states by the same index  $\nu$  and denote them by  $\chi_e^{(\nu)} \equiv \chi_{k_e^{(\nu)}}^e$  with energies  $\epsilon_e^{(\nu)} \equiv \epsilon_{k_e^{(\nu)}}$ , where  $k_e^{(\nu)}$  denotes the complex quasimomentum of the edge state. From Fig. 25(a) we get analog to (4.68)

$$k_e^{(\nu)} = i\kappa_e^{(\nu)} + \begin{cases} 0 & \text{for } \nu \text{ even} \\ \pi & \text{for } \nu \text{ odd} \end{cases}. \quad (4.93)$$

In Fig. 25(b) we show an example of the phase-dependence of the band structure together with the edge state energies  $\epsilon_e^{(\nu)}$  and the energies  $\epsilon_{bc}^{(\nu)}$  at the branching points, with  $\nu = 1, 2, 3$ . The band structure is periodic under a phase change by  $2\pi/Z$  since a shift of the whole lattice by one site does not change the bulk spectrum. In each gap one edge state is present which changes

from  $H_R$  to  $H_L$  at the point where it touches one of the bands (where  $\text{Im}(k_e^{(\nu)}) = 0$ ), indicated by solid and dashed lines of the edge states in Fig. 25(b), respectively. We note that the edge states of  $H_R$  or  $H_L$  can be shifted by  $\pm 2\pi/Z$  as function of  $\varphi$  if one shifts the boundaries of  $H_R$  or  $H_L$  shown in Fig. 22 by one site. In Fig. 26(a,b) we show the analytic continuation corresponding to the band structure of Fig. 25(b) for a particular value of the phase. We indicate the positions of the three edge poles  $k_e^{(\nu)}$  for  $\nu = 1, 2, 3$  in the complex plane. We note that the values for  $k_e^{(\nu)}$  are sitting on top of the branch cuts shown in Fig. 26(a). Therefore, they should be shifted slightly to the left/right of the branch cuts, depending on whether  $\epsilon_e^{(\nu)} \leq \epsilon_{\text{bp}}^{(\nu)}$  or, equivalently, whether the edge state belongs to the analytic continuation of band  $\nu$  or band  $\nu + 1$ . When the phase variable  $\varphi$  changes, the edge poles move around the branch cuts in Fig. 26(b) as shown in snapshots and videos, available in the Supplemental Material [26]. As can be seen from the dispersion relation  $\epsilon_e^{(\nu)}(\varphi)$  of the edge states in Fig. 25, the edge pole encircles the branch cut between band  $\nu$  and  $\nu + 1$  by an integer number when the phase has changed by  $2\pi$ . In Section 6.1 we will explain that  $\nu + nZ$  edge states of  $H_R$  connect band  $\nu$  and  $\nu + 1$ , with  $n = 0, \pm 1, \dots$ , running either all upwards or downwards for  $\nu + nZ \geq 0$ , respectively. This means that the edge pole of gap  $\nu$  runs  $\nu + nZ$  times around the branch cut, either clockwise or counter-clockwise for  $\nu + nZ \geq 0$ .

In the following we omit the index  $\nu$  for simplicity, i.e., use  $k_e \equiv k_e^{(\nu)}$ . From (4.43) we get

$$b_{-k_e}^T a_{k_e} = 0, \quad (4.94)$$

and we calculate  $a_{k_e}$  via (4.45) in terms of the well-defined matrix  $B(\epsilon_{k_e})$

$$a_{k_e} = -B(\epsilon_{k_e})b_{k_e}. \quad (4.95)$$

Using (4.59) and (4.60) this leads to the explicit solution

$$t_1 \cdots t_{j-1} a_{k_e}(j) = d_{1,j-1}(\epsilon_{k_e}) a_{k_e}(1), \quad (4.96)$$

$$a_{k_e}(1) = \frac{\bar{t}^Z}{t_Z} \left\{ \tilde{d}_{2,Z-1}(\epsilon_{k_e}) e^{-ik_e} + 1 \right\}. \quad (4.97)$$

Using the form (4.25) of  $b_{k_e}$ , we get from (4.94) the relation  $t_{Z-1} a_{k_e}(Z-1) = -t_Z e^{ik_e} a_{k_e}(1)$ . From (4.96) we get for  $j = Z-1$  the result  $t_{Z-1} a_{k_e}(Z-1) = -\frac{t_Z}{\bar{t}^Z} d_{1Z}(\epsilon_{k_e}) a_{k_e}(1)$ . Since  $a_{k_e} \neq 0$ , this gives the relation

$$d_{1Z}(\epsilon_{k_e}) = \bar{t}^Z e^{ik_e} \quad (4.98)$$

for all edge states.

Furthermore, from  $s(\epsilon_{k_e}) = d_{1,Z-1}(\epsilon_{k_e}) = 0$  and the properties (4.55) and (4.288), we get  $(t_1 \cdots t_{Z-2})^2 = d_{1,Z-2}(\epsilon_{k_e}) d_{2,Z-1}(\epsilon_{k_e})$  and  $d_{1Z}(\epsilon_{k_e}) = -t_{Z-1}^2 d_{1,Z-2}(\epsilon_{k_e})$ , leading with (4.98) to

$$\tilde{d}_{2,Z-1}(\epsilon_{k_e}) = -e^{-ik_e} \quad (4.99)$$

$$\tilde{d}_{1,Z-2}(\epsilon_{k_e}) = -e^{ik_e} \quad (4.100)$$

$$d_{1,Z-1}(\epsilon_{k_e}) = 0, \quad (4.101)$$

where  $\tilde{d}_{1,Z-2}$  and  $\tilde{d}_{2,Z-1}$  have been defined in (4.56) and (4.57), respectively.

From (4.97) and (4.99) we conclude that

$$a_{-k_e} = 0, \quad (4.102)$$

which is consistent with  $N_{k_e} = 0$ , see (4.41).



Using the result (4.99) for  $\tilde{d}_{2,Z-1}(\epsilon_{k_e})$  we can determine the sign of  $\text{Im}(k_e)$  and decide whether the edge state is an eigenstate of  $H_R$  or  $H_L$

$$\underline{k_e = i\kappa_e} : \quad \tilde{d}_{2,Z-1}(\epsilon_{k_e}) \leq -1 \Leftrightarrow \text{Im}(k_e) \geq 0 \quad (4.103)$$

$$\underline{k_e = \pi + i\kappa_e} : \quad \tilde{d}_{2,Z-1}(\epsilon_{k_e}) \geq 1 \Leftrightarrow \text{Im}(k_e) \geq 0. \quad (4.104)$$

As a consequence, when the determinant  $\tilde{d}_{2,Z-1}(\epsilon_{k_e(\varphi)})$  runs through the points  $\pm 1$  as function of the phase variable  $\varphi$  which determines the position  $k_e(\varphi)$  of the edge pole, an edge state is changing from  $H_R$  to  $H_L$  or vice versa. The phase-dependence of the edge state energies will be discussed in all detail in the next section.

## 2.6 Topological constraints for edge states

In this section we will derive the topological constraints for the phase-dependence of the edge state energies proving rigorously the central result Eq. (4.1). We will develop a simple diagrammatic representation to visualize the constraints and derive some rules how the phase-dependence of the edge state energies looks like.

To study the energy of the edge states as function of the phase  $\varphi$ , we fix some gap  $\nu = 1, \dots, Z-1$  (not written in the following) and define the phase-dependence of the edge state energy via

$$\epsilon_e(\varphi) = \epsilon_{k_e(\varphi)}(\varphi), \quad (4.105)$$

where  $k_e(\varphi)$  denotes the phase-dependence of the complex quasimomentum determining the edge state and  $\epsilon_k(\varphi)$  denotes the phase-dependence of the dispersion relation via the parameters  $v_j(\varphi)$  and  $t_j(\varphi)$  defining the microscopic model. We first note the important property that the dispersion can never be the same at  $\varphi$  and  $\varphi + \frac{2\pi}{Z}$  when the edge states at both values belong either to  $H_R$  or to  $H_L$

$$\begin{aligned} \text{Im}[k_e(\varphi)] &= \text{Im} \left[ k_e \left( \varphi + \frac{2\pi}{Z} \right) \right] \Rightarrow \\ &\Rightarrow \quad \epsilon_e(\varphi) \neq \epsilon_e \left( \varphi + \frac{2\pi}{Z} \right). \end{aligned} \quad (4.106)$$

To show this we use (4.9) and find

$$\tilde{d}_{2,Z-1}(\epsilon, \varphi) = \tilde{d}_{1,Z-2} \left( \epsilon, \varphi + \frac{2\pi}{Z} \right), \quad (4.107)$$

where the dependence on  $\varphi$  again indicates the one from the parameters  $v_j(\varphi)$  and  $t_j(\varphi)$ . Assuming  $\epsilon_e(\varphi) = \epsilon_e(\varphi + \frac{2\pi}{Z})$  we get from (4.99), (4.100) and (4.107)

$$\begin{aligned} -e^{-ik_e(\varphi)} &= \tilde{d}_{2,Z-1}(\epsilon_e(\varphi), \varphi) \\ &= \tilde{d}_{1,Z-2}(\epsilon_e(\varphi), \varphi + \frac{2\pi}{Z}) \\ &= \tilde{d}_{1,Z-2}(\epsilon_e(\varphi + \frac{2\pi}{Z}), \varphi + \frac{2\pi}{Z}) \\ &= -e^{ik_e(\varphi + \frac{2\pi}{Z})}. \end{aligned}$$

This is only possible for  $\text{Im}[k_e(\varphi)] \neq \text{Im}[k_e(\varphi + \frac{2\pi}{Z})]$  which proves (4.106).

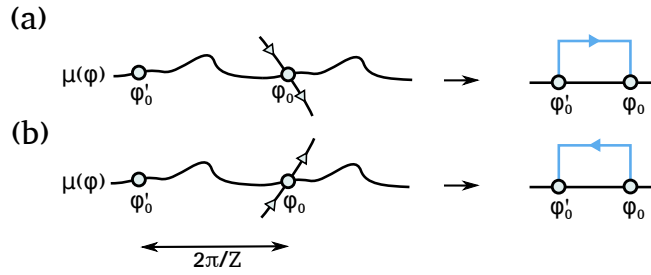


Figure 27: edge states of  $H_R$  crossing the chemical potential  $\mu(\varphi) = \mu(\varphi + \frac{2\pi}{Z})$  at  $\varphi = \varphi_0$  either from (a) above or (b) below. At  $\varphi'_0 = \varphi_0 - \frac{2\pi}{Z}$  it is not allowed that an edge state of  $H_R$  crosses the chemical potential. To the right we show the way how we visualize the two different possibilities by contractions (in blue color).

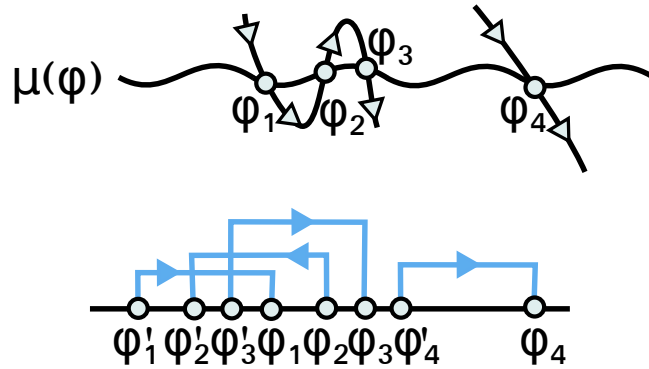


Figure 28: Visualization of edge states of  $H_R$  crossing  $\mu(\varphi)$  at  $\varphi = \varphi_i$ , with  $i = 1, 2, 3, 4$ , in terms of sequences of contractions. We defined  $\varphi'_i = \varphi_i - \frac{2\pi}{Z}$  such that each contraction has a fixed length  $\frac{2\pi}{Z}$ .

From (4.106) we can deduce a first important property how edge states of  $H_R$  (analog for  $H_L$ ) can cross any phase-dependent chemical potential

$$\mu(\varphi) = \mu\left(\varphi + \frac{2\pi}{Z}\right) \quad (4.108)$$

chosen somewhere in gap  $\nu$  and periodic with period  $\frac{2\pi}{Z}$ . When an edge state of  $H_R$  fulfils  $\epsilon_e(\varphi_0) = \mu(\varphi_0)$ , it is not possible that an edge state of  $H_R$  can cross the chemical potential at  $\varphi \pm \frac{2\pi}{Z}$ . Considering only edge states of  $H_R$  we show in Figs. 27(a,b) how we visualize the two possibilities of an edge state crossing the chemical potential at  $\varphi = \varphi_0$  either from above or below via contractions. The phase  $\varphi_0$  of the right vertex of a contraction denotes the position of the edge state crossing  $\mu(\varphi_0)$ . Each contraction has a fixed length  $\frac{2\pi}{Z}$  such that the left vertex has phase  $\varphi'_0 = \varphi_0 - \frac{2\pi}{Z}$ . In this picture the property (4.106) is equivalent to the fact that two vertices can never lie on top of each other. Each way the edge states of  $H_R$  can cross the chemical potential can then be visualized by a sequence of contractions, see Fig. 28. This visualization will turn out to be very convenient to formulate and prove many of the following topological constraints.

If a certain topological configuration of contractions occurs for some chosen  $\mu(\varphi)$  it can not change when choosing a different chemical potential since two vertices are never allowed to coincide. An exception are cases when the chemical potential is moved through a local minimum or maximum of the phase-dependence of the edge state energy. In this case a pair of two contractions with different directions fall on top of each other and are eliminated (or created). However, as explained below in all detail, such pairs do not change any of the topological constraints discussed in the following. Therefore, without loss of generality, we

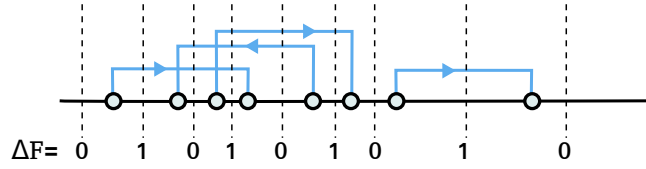


Figure 29: Graphical rule how to determine the virtual topological charge  $\Delta F(\varphi)$  at some phase  $\varphi$  lying between two vertices. Drawing a vertical cut at this position (dashed line) one has to take the number of right-going minus the number of left-going contraction lines through this vertical cut.

choose for  $\mu(\varphi)$  the top of the lower band  $\alpha = \nu$

$$\mu(\varphi) = \epsilon_{k_0}^{(\nu)}(\varphi), \quad (4.109)$$

where  $k_0 = 0$  for  $\nu$  even and  $k_0 = \pi$  for  $\nu$  odd. For this choice the edge states of  $H_R$  can just enter or leave the band, their connection below  $\mu(\varphi)$  is just meant as a guide for the eye to formulate certain topological constraints derived in the following.

In Appendix 8.5 we will prove from the rules (4.99) and (4.100) for the occurrence of edge states the following topological condition for the allowed edge state configurations of  $H_R$

Topological constraint:

$$\text{Outgoing and incoming vertices must alternate.} \quad (4.110)$$

Together with the fact that the contractions have a fixed length  $\frac{2\pi}{Z}$  and have to be ordered on an interval of size  $2\pi$  with periodic boundary conditions, we can construct all possible edge state configurations. None of them can be excluded in principle and it depends on the model under consideration which of them appear. E.g., the configuration shown in Fig. 28 is obviously an allowed one consistent with (4.110). The topological constraint can also be formulated in terms of a virtual topological charge  $\Delta F(\varphi)$  defined by

$$\Delta F(\varphi) = F(\varphi + \frac{2\pi}{Z}) - F(\varphi), \quad (4.111)$$

$$F(\varphi) = \sum_{\sigma=\pm} \sum_{i=1}^{M_\sigma} \sigma \theta(\varphi - \varphi_{i\sigma}), \quad (4.112)$$

where  $\varphi_{i\sigma}$  are the phase values where edge states of  $H_R$  enter/leave the band and  $M_\pm$  is the total number of entering/leaving edge states. Using the following form of the topological charge

$$\Delta F(\varphi) = \sum_{\sigma=\pm} \sum_{i=1}^{M_\sigma} \sigma \theta(\varphi'_{i\sigma} < \varphi < \varphi_{i\sigma}), \quad (4.113)$$

one finds that  $\Delta F(\varphi)$  can be read off for  $\varphi$  lying between two adjacent vertices of a certain configuration by making a virtual vertical cut at  $\varphi$  and taking the number of right-going minus the number of left-going contraction lines crossing the vertical cut, see Fig. 29 for an example. The topological constraint (4.110) can then alternatively be formulated as a constraint for the topological charge

$$\Delta F(\varphi) \in \{s-1, s\}, \quad (4.114)$$

where  $s$  is a phase-independent integer characteristic of each configuration. E.g., in Fig. 29 we get  $s = 1$ . We note that by reversing all direction of the contractions we change the sign of

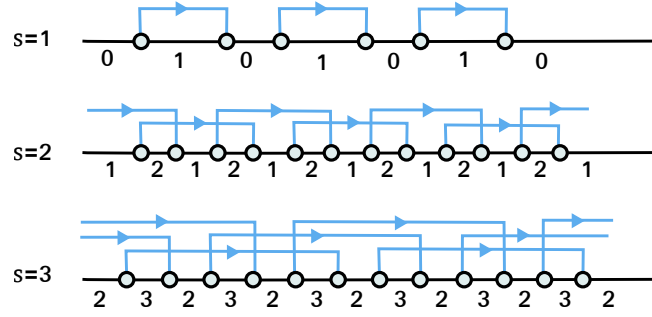


Figure 30: The allowed configurations for  $s = 1$ ,  $s = 2$ , and  $s = 3$  when all contractions have the same direction. The topological charge  $\Delta F \in \{s - 1, s\}$  is indicated between all adjacent vertices. Changing the direction of all contractions leads to the configurations with  $s' = -s + 1 = 0, -1, -2$ , see Eq. (4.115).

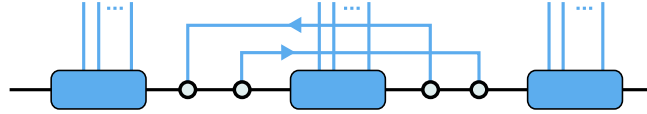


Figure 31: The construction how a pair of two contractions with different directions is inserted into an old configuration, formed by the three boxes. The condition is that between the two right and the two left vertices of the new contractions no other vertex of the old configuration is allowed to appear. Depending on the old configuration it might be necessary to invert the two directions of the new contractions in order to get an allowed configuration after inserting the two new contractions.

the topological charge and get again an allowed configuration. Due to (4.114) this operation changes the parameter  $s$  to  $s' = -s + 1$

$$-\Delta F(\varphi) \in \{s' - 1, s'\} \quad , \quad s' = -s + 1 \quad . \quad (4.115)$$

We note at this point that, following Ref. [[52]], we will derive in Section 6. the topological constraint (4.114) by a completely different route in terms of physically intuitive arguments using charge conservation and particle-hole duality only without involving any edge state physics. There, we will see that the difference  $I = \Delta F - s$  is a topological invariant which can be related to the physical observable of the boundary charge  $Q_B$  via

$$I(\varphi) = \Delta F(\varphi) - s = \Delta Q_B(\varphi) - \bar{\rho} \in \{-1, 0\} \quad , \quad (4.116)$$

where  $\Delta Q_B(\varphi) = Q_B(\varphi + \frac{2\pi}{Z}) - Q_B(\varphi)$  is the difference of the boundary charge between the shifted and unshifted lattice and  $\bar{\rho} = \frac{\nu}{Z}$  is the average charge per site for the infinite system without a boundary. To get a feeling which configurations are possible for the edge states according to the constraint (4.110), we proceed by providing an iterative scheme how more complicated configurations can be obtained from simpler ones. As a starting point we consider the possible configurations when all contractions have the same direction. These configuration are shown in Fig. 30(a,b,c) for different values of  $s = 1, 2, 3$ . The construction for larger values is obvious and the ones for  $s = 0, -1, -2, \dots$  are obtained by reversing the directions of all contractions according to Eq. (4.115). In terms of the edge states these configurations correspond to the cases where all edge states either enter or leave the band. For  $s = 1$  ( $s = 0$ ) the entering (leaving) points of adjacent edge states have a distance larger than  $\frac{2\pi}{Z}$  and increasing (decreasing)  $s$  by one means that  $Z$  additional edge states enter (leave) the band. For these configurations we proof in Appendix 8.6 the Diophantine equation [[54]-[56]]

$$M = M_- - M_+ = \nu - sZ \quad (4.117)$$

$$\Leftrightarrow -M = M_+ - M_- = \nu' - s'Z \quad , \quad (4.118)$$

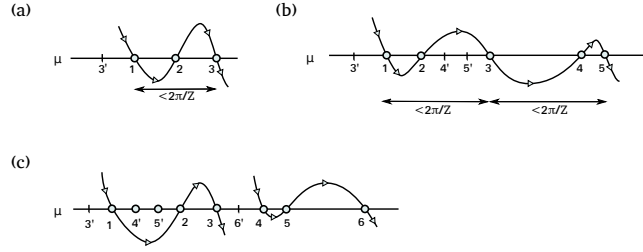


Figure 32: Illustration of configurations which arise when one adds subsequently pairs of edge states crossing  $\mu$  from above and below. For simplicity we have taken a constant  $\mu$  independent of  $\varphi$ . Here, the points  $i$  indicate the phase points  $\phi_i$  and  $i'$  the shifted ones  $\phi'_i = \phi_i - \frac{2\pi}{Z}$ . (a) Adding the pair (2, 3) to an edge state crossing  $\mu$  at 1 from above leads to an additional oscillation around  $\mu$  with length smaller than  $\frac{2\pi}{Z}$  such that the edge state remains to run from the upper to the lower band. (b) Adding another pair (4, 5) leads to a second oscillation of length smaller than  $\frac{2\pi}{Z}$ . Note that (4', 5') lies here between 2 and 3 such that the topological constraint is fulfilled. (c) Two consecutive edge states with one additional oscillation each. Here the shifted points 4' and 5' of the right edge state are located between the crossing points 1 and 2 of the left edge state. Note that the topological constraint does not allow 4' and 5' to lie between 2 and 3. This gives the rule that if the crossing points  $i$  and  $j$  are connected by an edge state with a maximum (minimum) in between then the shifted points  $i'$  and  $j'$  must be located in an interval where the edge state has also a maximum (minimum). The connection of the edge states between adjacent crossing points is just a guide for the eye, these connections can also be cut (e.g., when  $\mu$  is identical to the band edge this is even necessary).

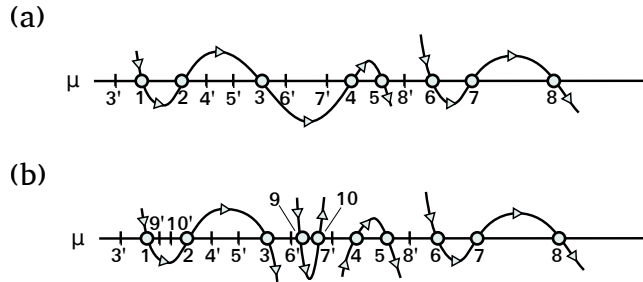


Figure 33: (a) Two consecutive edge states with two (one) oscillation of the left (right) edge state around  $\mu$ , analog to the cases discussed in Fig. 32. (b) A more complicated configuration where we insert the pair (9, 10) between 6' and 7' in (a) and cut the edge mode connecting 3 and 4. As a result we get four consecutive edge states, where the two outer ones connect the upper with the lower band whereas the two middle ones return to the same band, the left (right) one to the upper (lower) band. Note that the crossing points 1 and 2 are essential, otherwise there would be a mismatch between the consecutive points 3' and 9' which correspond both to outgoing vertices. The crossing points 4 and 5 are necessary to get a new configuration compared to Fig. 32. Without 4 and 5 one can connect the edge state between 10 and 6 and gets two consecutive edge states oscillating around  $\mu$ . Furthermore the crossing points 6 and 7 are needed since without 6' and 7' there is a mismatch between 3 and 9 and between 10 and 4 corresponding both to incoming or outgoing vertices, respectively. Therefore, *all* crossing points are essential to construct such an involved configuration.

with  $\nu' = Z - \nu$  and  $1 \leq \nu, \nu' \leq Z - 1$ . That  $\nu$  is precisely the index corresponding to the considered gap will be shown in Section 6.. As shown below in this section the Diophantine equation holds not only for the case when all contractions have the same direction but for *all* allowed configurations.

The way to obtain all configurations by mixing contractions with different directions is then quite obvious. As shown in Appendix 8.7 they are obtained iteratively by inserting into a given configuration one pair of contractions with different directions such that no other vertex appears between the two right and the two left vertices of the two new contractions, see Fig. 31 for illustration. Obviously by choosing the direction of the two new contractions appropriately one obtains again an allowed configuration and proceeding in this way every configuration can be obtained, see the proof in Appendix 8.7. We note from this construction that neither the number  $M = M_- - M_+$  nor the integer  $s$  are changed by adding a new pair of contractions. Therefore, the Diophantine equations (4.117) and (4.118) remain valid.

To visualize what kinds of edge state configuration are generated by this construction we have shown in Fig. 32(a,b) what happens if one adds pairs of edge states crossing the chemical potential from different sides one after the other to a single crossing point. This leads to more and more oscillations of the edge state around the chemical potential where each additional oscillation must have a length smaller than  $\frac{2\pi}{Z}$ . We note that there is no need to connect the edge state between two adjacent crossing points, all these lines can be optionally cut. The topological constraint only fixes the allowed configuration of the crossing points. Therefore, it is possible that not all edge states connect the two bands in the same direction, both the numbers  $M_+$  and  $M_-$  can be unequal to zero. In particular, when  $\mu$  is chosen as the band edge (either the bottom of the upper band or the top of the lower band) it is even necessary to leave out the connection lines above or below  $\mu$ . In Fig. 32(c) and Fig. 33(a) we have shown how several edge states oscillating around the chemical potential can arise. If we connect all possible adjacent crossing points we find that all edge states run from one band to the other (either all downwards or all upwards). Furthermore we find the rule that if two crossing points  $i$  and  $j$  are connected by an edge state with a maximum (minimum) of the edge state between these two points, then the shifted points  $i'$  and  $j'$  must be located in an interval where the edge state has also a maximum (minimum), at least in the case when an edge mode is present there. How more complicated configurations can appear where not all edge states connect the bands in the same direction is shown in Fig. 33(b). Here, two consecutive edge states return to the same band, one to the upper and the other to the lower band. However, these are rather exotic configurations which in practice occur very unlikely, except for very special functions  $F_v$  and  $F_\gamma$  with many random Fourier components defining the model.

### 3. Boundary charge and density

In this section we will define the boundary charge via a macroscopic average over the microscopic density, similar to Ref. [[17]] and Chapter 4.5.1 in Ref. [[41]]. As proposed in Ref. [[52]] this definition allows for a gauge invariant decomposition of the boundary charge in a Friedel, polarization, and edge part. We present all the necessary formulas for the Friedel and polarization part in terms of the Bloch states useful for many further investigations. We also review the particle-hole duality used in Ref. [[52]]. In addition to Ref. [[52]], we will also present the calculation of the boundary charge for a half-infinite system with a right boundary and establish a unique relation between the boundary charges of  $H_R$  and  $H_L$ . In Section 3.3 we will present an analytical calculation of the density based on the analytic continuation of Bloch states, similar to previous treatments for continuum systems [48]. We prove the exponential localization of the boundary charge, calculate the localization length, and show that a nontrivial pre-exponential function appears which falls off with a generic exponent  $1/\sqrt{n}$

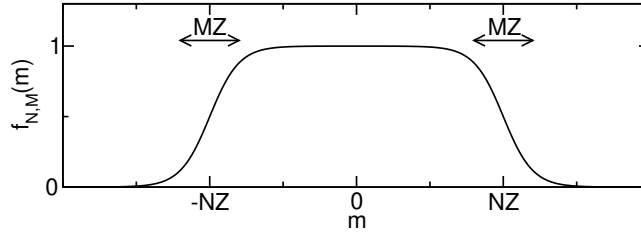


Figure 34: Sketch of the envelope function  $f_{N,M}(m)$  with  $N \gg M \gg 1$ .

for large distances.

### 3.1.1 Definition and splitting of boundary charge

Following Ref. [[17]], we define the boundary charges  $Q_B^{R/L}$  corresponding to the half-infinite systems  $H_{R/L}$  via the macroscopic average

$$Q_B^{R/L} = \lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} \sum_{m=\pm 1}^{\pm \infty} (\rho(m) - \bar{\rho}) f_{N,M}(m), \quad (4.119)$$

where  $\rho(m) = \langle |m\rangle \langle m| \rangle_{H_{R/L}}$  is the density at site  $m$  for the ground state of  $H_{R/L}$  and  $f_{N,M}(m)$  is an envelope function of a charge measuring device which, at site  $m \sim NZ$ , varies smoothly from unity to zero on a scale  $MZ$  much larger than the scale  $Z$  of a unit cell, i.e.,  $N \gg M \gg 1$ , see Fig. 34 for a sketch of the envelope function.  $\bar{\rho}$  is defined as the average density per site of the infinite system

$$\bar{\rho} = \frac{1}{Z} \sum_{j=1}^Z \rho_{\text{bulk}}(j), \quad (4.120)$$

where  $\rho_{\text{bulk}}(j) = \langle |nj\rangle \langle nj| \rangle_{H_{\text{bulk}}}$  is the bulk density in the ground state of the infinite system described by  $H_{\text{bulk}}$ . This bulk density is independent of the unit cell index  $n$  but can depend on the site index  $j$  within a unit cell. In the following we consider the insulating regime and assume that the chemical potential  $\mu = \mu_\nu$  is located in gap  $\nu$ , i.e., the bands  $\alpha = 1, 2, \dots, \nu$  are filled. In this case we get from (4.31)

$$\rho_{\text{bulk}}(j) = \sum_{\alpha=1}^{\nu} \rho_{\text{bulk}}^{(\alpha)}(j), \quad (4.121)$$

$$\begin{aligned} \rho_{\text{bulk}}^{(\alpha)}(j) &= \int_{-\pi}^{\pi} dk |\psi_{k,\text{bulk}}^{(\alpha)}(n, j)|^2 \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} dk |\chi_k^{(\alpha)}(j)|^2, \end{aligned} \quad (4.122)$$

and, due to the normalization (4.83) of the Bloch states,

$$\bar{\rho} = \frac{\nu}{Z}. \quad (4.123)$$

To get a well-defined expression independent of the envelope function we use the central property that the density of the half-infinite system approaches the bulk density very far away from the boundary

$$\lim_{n \rightarrow \infty} \rho(n, j) = \rho_{\text{bulk}}(j). \quad (4.124)$$



As we will show in Section 3.3 this happens on an exponential scale  $\xi_B \sim 1/\kappa_e^{(\nu)}, 1/\kappa_{bp}^{(\nu)}$ , see Eq. (4.175). The limit  $N, M \rightarrow \infty$  in Eq. (4.119) has to be understood in the sense  $NZ \gg \xi$  and  $MZ \gg Z$ , i.e., the measuring device should be smooth on the scale of the unit cell and should probe the charge on a scale much larger than the localization length  $\xi_B$  of the boundary charge. This guarantees that the deviation of the experimental measurement from the mathematical limit in Eq. (4.119) is exponentially small. Splitting the summand of (4.119) in two terms via  $\rho(m) - \bar{\rho} = (\rho(m) - \rho_{\text{bulk}}(j)) + (\rho_{\text{bulk}}(j) - \bar{\rho})$ , we can then set  $f_{N,M}(m) \approx 1$  for the first term. For the second term we expand  $f_{N,M}(m) = f_{N,M}(Z(n-1)+j) \approx f_{N,M}(Zn) + f'_{N,M}(Zn)(-Z+j)$  up to linear order which becomes exact in the limit  $N \gg M \rightarrow \infty$ . Using in addition  $\rho_{R/L}(m=0) = 0$  and  $\sum_{j=1}^Z (\rho_{\text{bulk}}(j) - \bar{\rho}) = 0$  together with

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} \sum_{n=1}^{\infty} f'(Zn) = -\frac{1}{Z}, \quad (4.125)$$

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} \sum_{n=-\infty}^0 f'(Zn) = \frac{1}{Z}, \quad (4.126)$$

we get the decomposition

$$Q_B^{R/L} = \sum_{m=\pm 1}^{\pm \infty} (\rho(m) - \rho_{\text{bulk}}(j)) + Q_P^{R/L}, \quad (4.127)$$

where

$$Q_P^R = -\frac{1}{Z} \sum_{j=1}^Z j (\rho_{\text{bulk}}(j) - \bar{\rho}), \quad (4.128)$$

$$\begin{aligned} Q_P^L &= \frac{1}{Z} \sum_{j=1}^Z j (\rho_{\text{bulk}}(j) - \bar{\rho}) - (\rho_{\text{bulk}}(Z) - \bar{\rho}) \\ &= \frac{1}{Z} \sum_{j=1}^{Z-1} j (\rho_{\text{bulk}}(j) - \bar{\rho}) \end{aligned} \quad (4.129)$$

is the contribution from the bulk polarization to the boundary charges, analog to the surface charge of a dielectric medium in classical electrodynamics. We note that this decomposition is analog to decompositions to define the boundary charge within the MTP, see e.g. Eq. (4.85) in Chapter 4.5 of Ref. [[41]]. The essential difference is that we have chosen a particular representation of the polarization charge such that the remaining part can be expressed uniquely via the Friedel and edge state charge, see below.

Using (4.120) and (4.121) we can split the polarization charge into the contributions of the individual bands

$$Q_P^{R/L} = \sum_{\alpha=1}^{\nu} Q_P^{R/L,(\alpha)}, \quad (4.130)$$

with

$$Q_P^{R,(\alpha)} \equiv Q_P^{(\alpha)}, \quad (4.131)$$

$$Q_P^{L,(\alpha)} = -Q_P^{(\alpha)} - \rho_{\text{bulk}}^{(\alpha)}(Z) + \frac{1}{Z}, \quad (4.132)$$

$$Q_P^{(\alpha)} = -\frac{1}{Z} \sum_{j=1}^Z j \left( \rho_{\text{bulk}}^{(\alpha)}(j) - \frac{1}{Z} \right). \quad (4.133)$$

This gives for the sum

$$Q_P^{L,(\alpha)} + Q_P^{R,(\alpha)} = -\rho_{\text{bulk}}^{(\alpha)}(Z) + \frac{1}{Z}. \quad (4.134)$$

Finally, we split the total density

$$\rho(m) = \rho_{\text{band}}(m) + \rho_{\text{edge}}(m) \quad (4.135)$$

into the contribution of the filled bands and the edge states and define the Friedel density via

$$\rho_F(m) = \rho_{\text{band}}(m) - \rho_{\text{bulk}}(j) = \sum_{\alpha=1}^{\nu} \rho_F^{(\alpha)}(m), \quad (4.136)$$

where

$$\rho_F^{(\alpha)}(n, j) = \rho_{\text{band}}^{(\alpha)}(n, j) - \rho_{\text{bulk}}^{(\alpha)}(j) \quad (4.137)$$

is the contribution from a single band. Thus, together with  $\rho_{\text{band}}(m=0) = 0$ , we arrive at the final splitting

$$Q_B^{R/L} = Q_F^{R/L} + Q_E^{R/L} + Q_P^{R/L}, \quad (4.138)$$

where

$$Q_F^{R/L} = \sum_{m=\pm 1}^{\pm\infty} \rho_F(m), \quad Q_E^{R/L} = \sum_{m=\pm 1}^{\pm\infty} \rho_{\text{edge}}(m) \quad (4.139)$$

define the contributions to the boundary charge from the Friedel and edge charge densities of  $H_{R/L}$ . Using (4.136) we can split  $Q_F^{R/L}$  into the contributions of the individual bands

$$Q_F^{R/L} = \sum_{\alpha=1}^{\nu} Q_F^{R/L,(\alpha)}, \quad Q_F^{R/L,(\alpha)} = \sum_{m=\pm 1}^{\pm\infty} \rho_F^{(\alpha)}(m). \quad (4.140)$$

Using (4.122) together with the form (4.82) of the eigenfunctions  $\psi_k^{(\alpha)}(n, j)$  of band  $\alpha$  of  $H_{\text{bulk}}$ , we can write the Friedel density of band  $\alpha$  in the form

$$\begin{aligned} \rho_F^{(\alpha)}(n, j) &= \int_0^\pi dk |\psi_k^{(\alpha)}(n, j)|^2 - \int_{-\pi}^\pi dk |\psi_{k,\text{bulk}}^{(\alpha)}(n, j)|^2 \\ &= -\frac{1}{2\pi} \int_{-\pi}^\pi dk \left( \chi_k^{(\alpha)}(j) \right)^2 e^{2ikn}. \end{aligned} \quad (4.141)$$

This gives for  $Q_F^{R,(\alpha)}$  the form

$$Q_F^{R,(\alpha)} \equiv Q_F^{(\alpha)}, \quad (4.142)$$

$$Q_F^{(\alpha)} = -\frac{1}{2\pi} \sum_{n=1}^{\infty} \int_{-\pi}^\pi dk (\chi_k^{(\alpha)})^T \chi_k^{(\alpha)} e^{2ikn}. \quad (4.143)$$

Using (4.40) and (4.41) we can write

$$\begin{aligned} (\chi_k^{(\alpha)})^T \chi_k^{(\alpha)} &= \frac{(a_k^{(\alpha)})^T a_k^{(\alpha)} + s(\epsilon_k^{(\alpha)})^2}{(a_k^{(\alpha)})^T a_{-k}^{(\alpha)} + s(\epsilon_k^{(\alpha)})^2} \\ &= 1 + \frac{1}{N_k^{(\alpha)}} (a_k^{(\alpha)})^T (a_k^{(\alpha)} - a_{-k}^{(\alpha)}). \end{aligned} \quad (4.144)$$

The unity on the r.h.s. does not contribute to (4.143). The second term can be evaluated with the form (4.50) for  $a_k$  leading to

$$\begin{aligned} (a_k^{(\alpha)})^T (a_k^{(\alpha)} - a_{-k}^{(\alpha)}) &= \\ &= \left( (f^T f)(\epsilon_k^{(\alpha)}) + (f^T g)(\epsilon_k^{(\alpha)}) e^{ik} \right) (e^{-2ik} - 1). \end{aligned} \quad (4.145)$$

Inserting (4.144) and (4.145) in (4.143) and using

$$\sum_{n=1}^{\infty} e^{2ikn} (e^{-2ik} - 1) = \sum_{n=0}^{\infty} e^{2ikn} - \sum_{n=1}^{\infty} e^{2ikn} = 1, \quad (4.146)$$

we get the compact form

$$Q_F^{(\alpha)} = -\frac{1}{2\pi} \int_{-\pi}^{\pi} dk \frac{(f^T f)(\epsilon_k^{(\alpha)}) + (f^T g)(\epsilon_k^{(\alpha)}) \cos(k)}{N_k^{(\alpha)}}, \quad (4.147)$$

which allows for a straightforward numerical evaluation.

A similar analysis yields for the Friedel charge  $Q_F^{L,(\alpha)}$

$$\begin{aligned} Q_F^{L,(\alpha)} - \rho_{\text{bulk}}^{(\alpha)}(Z) &= \sum_{m=-\infty}^0 \rho_F^{(\alpha)}(m) \\ &= -\frac{1}{2\pi} \sum_{n=-\infty}^0 \int_{-\pi}^{\pi} dk (\chi_k^{(\alpha)})^T \chi_k^{(\alpha)} e^{2ikn} \\ &= -\frac{1}{2\pi} \sum_{n=-\infty}^0 \int_{-\pi}^{\pi} dk e^{2ikn} \cdot \\ &\quad \cdot \left\{ 1 + \frac{(f^T f)(\epsilon_k^{(\alpha)}) + (f^T g)(\epsilon_k^{(\alpha)}) e^{ik}}{N_k^{(\alpha)}} (e^{-2ik} - 1) \right\}. \end{aligned} \quad (4.148)$$

The unity in the brackets contributes only for  $n = 0$  and gives  $-1$ . For the second term we use

$$\sum_{n=-\infty}^0 e^{2ikn} (e^{-2ik} - 1) = \sum_{n=-\infty}^{-1} e^{2ikn} - \sum_{n=-\infty}^0 e^{2ikn} = -1 \quad (4.149)$$

and, by using (4.147), we get  $-Q_F^{(\alpha)}$  for this term since the  $\sin(k)$  part of  $e^{ik}$  does not contribute due to  $\epsilon_k^{(\alpha)} = \epsilon_{-k}^{(\alpha)}$  and  $N_k^{(\alpha)} = N_{-k}^{(\alpha)}$ . Thus, we get from (4.148) the relation

$$Q_F^{L,(\alpha)} + Q_F^{R,(\alpha)} = -1 + \rho_{\text{bulk}}^{(\alpha)}(Z). \quad (4.150)$$

Defining the boundary charges of band  $\alpha$  by

$$Q_B^{R/L,(\alpha)} = Q_F^{R/L,(\alpha)} + Q_P^{R/L,(\alpha)}, \quad (4.151)$$

we get from (4.134) and (4.150) the universal result

$$Q_B^{L,(\alpha)} + Q_B^{R,(\alpha)} = -\frac{Z-1}{Z}. \quad (4.152)$$

This is an interesting relation between the boundary charges at the left and right boundary stated in Eq. (4.5) in the introduction. In Section 4. we will see that it leads to a corresponding relation between the Zak-Berry phases for  $H_L$  and  $H_R$ , see

As shown in Section 2.5 each gap  $\nu = 1, \dots, Z - 1$  contains either an edge state of  $H_L$  or one of  $H_R$ . Since the chemical potential  $\mu_\nu$  is located somewhere in gap  $\nu$  we get for all  $1 \leq \nu \leq Z - 1$

$$Q_E^R + Q_E^L = \nu - 1 + \theta(\mu_\nu - \epsilon_e^{(\nu)}). \quad (4.153)$$

Using (4.130), (4.138), (4.140), (4.151), (4.152) and (4.153), we obtain finally the following useful relation between the boundary charges of  $H_{R/L}$

$$Q_B^L + Q_B^R = \bar{\rho} - 1 + \theta(\mu_\nu - \epsilon_e^{(\nu)}). \quad (4.154)$$

For the case when the chemical potential is lying above the highest band all states are filled and we get  $\rho(m) = \bar{\rho} = 1$  such that, using the definition (4.119), we get zero boundary charge

$$Q_B^L = Q_B^R = 0 \quad \text{for } \nu = Z. \quad (4.155)$$

Since the two boundary charges are not independent, we will discuss in the following only the boundary charge  $Q_B^R$  of  $H_R$  and use the simplified notation  $Q_B \equiv Q_B^R$ ,  $Q_F \equiv Q_F^R$ ,  $Q_P \equiv Q_P^R$  and  $Q_E \equiv Q_E^R$  such that

$$Q_B = Q_F + Q_P + Q_E = \sum_{\alpha=1}^{\nu} Q_B^{(\alpha)} + Q_E, \quad (4.156)$$

where

$$Q_B^{(\alpha)} = Q_F^{(\alpha)} + Q_P^{(\alpha)} \quad (4.157)$$

is the boundary charge of a single band and

$$Q_F = \sum_{\alpha=1}^{\nu} Q_F^{(\alpha)} \quad , \quad Q_P = \sum_{\alpha=1}^{\nu} Q_P^{(\alpha)}. \quad (4.158)$$

Here,  $Q_F^{(\alpha)}$  and  $Q_P^{(\alpha)}$  are defined in Eqs. (4.143) and (4.133). If all states are filled we get from (4.155) that  $Q_B = 0$  or

$$\sum_{\alpha=1}^Z Q_B^{(\alpha)} = -Q_E^{\text{tot}}, \quad (4.159)$$

where  $Q_E^{\text{tot}}$  denotes the total number of edge states of  $H_R$  in all gaps  $\nu = 1, \dots, Z - 1$ . Furthermore, due to  $\sum_{\alpha=1}^Z \rho_{\text{bulk}}^{(\alpha)}(j) = 1$ , we get from (4.133) that the sum of the polarization charges of all bands is zero

$$\sum_{\alpha=1}^Z Q_P^{(\alpha)} = 0. \quad (4.160)$$

Using (4.157) and (4.159) this implies that the sum of the Friedel charges of all bands is quantized in integer units given by the negative charge of all edge states

$$\sum_{\alpha=1}^Z Q_F^{(\alpha)} = -Q_E^{\text{tot}}. \quad (4.161)$$

### 3.2 Particle-hole duality

For the discussion of the universal properties of the boundary charge  $Q_{B,p} \equiv Q_B$  of the particles it will turn out to be very important to look also at the boundary charge  $Q_{B,h}$  of the holes. To define this quantity we note that the particle and hole charge densities are given by

$$\rho_p(m) = \rho(m), \quad (4.162)$$

$$\rho_h(m) = -(1 - \rho(m)) = \rho(m) - 1. \quad (4.163)$$

In the same way the average particle and hole charge densities per site are given for the infinite system by

$$\bar{\rho}_p = \bar{\rho}, \quad \bar{\rho}_h = \bar{\rho} - 1. \quad (4.164)$$

The corresponding boundary charges for particles and holes are then defined via (4.119) as

$$Q_{B,p} = \lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} \sum_{m=\pm 1}^{\pm \infty} (\rho_p(m) - \bar{\rho}_p) f_{N,M}(m), \quad (4.165)$$

$$Q_{B,h} = \lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} \sum_{m=\pm 1}^{\pm \infty} (\rho_h(m) - \bar{\rho}_h) f_{N,M}(m). \quad (4.166)$$

Using Eqs. (4.162-4.166) we find the important property that the boundary charges of the particles and holes are the same

$$Q_B = Q_{B,p} = Q_{B,h}. \quad (4.167)$$

We emphasize that looking at the same physics from the hole point of view is not just reproducing the same in a different language, it involves in addition the Pauli principle and is related to the fact that the boundary charge is zero when all states are filled, see Eq. (4.155). When inverting the occupation of all states by populating all states above  $\mu_\nu$  instead of below  $\mu_\nu$  the charge density changes to  $\rho(m) \rightarrow 1 - \rho(m)$  due to the Pauli principle. Thus, the hole charge density  $\rho_h(m) = \rho(m) - 1$  can also be viewed as the negative charge density after population inversion which describes a new physical situation providing new information about the properties of the system. This will be very essential for the discussion in Section 6.1. We note that the two different viewpoints in terms of particles and holes have nothing to do with particle-hole symmetry but just involves the physics of the Pauli principle.

### 3.3 Density and localization of boundary charge

In this section we discuss the localization of the boundary charge  $Q_B$  of  $H_R$  given by  $Q_B = Q_F + Q_E + Q_P$  with

$$\begin{aligned} Q_F + Q_E &= \sum_{m=1}^{\infty} [\rho_F(m) + \rho_{\text{edge}}(m)] \\ &= \sum_{m=1}^{\infty} [\rho(m) - \rho_{\text{bulk}}(j)], \end{aligned} \quad (4.168)$$

see Eqs. (4.135), (4.136), (4.138) and (4.139). The polarization part  $Q_P \equiv Q_P^R$ , given by (4.128), is by definition a contribution to the boundary charge localized on the scale  $Z$  of the unit cell since it occurs for all scales  $N, M \gg 1$  of the envelope function  $f_{N,M}(m)$  of the charge measurement probe. In contrast, the Friedel density  $\rho_F(m)$  and the edge state density  $\rho_{\text{edge}}(m)$

can have localization lengths at the boundary much larger than the length scale of one unit cell  $\xi_F, \xi_e \gg Z$ . We expect that the two length scales of  $\rho_F$  and  $\rho_{\text{edge}}$  are of the same order  $\xi_F \sim \xi_e$  since, due to charge conservation, an edge state leaving/entering a certain band by adiabatically changing the phase variable  $\varphi$  will decrease/increase the charge of this band by one and it is very unlikely that this happens nonlocally. The localization length  $\xi_e$  can be arbitrarily large since the localization length  $\xi_e^{(\nu)} \sim 1/\kappa_e^{(\nu)}$  of an edge state of  $H_R$  located in gap  $\nu$  will go to infinity when its energy  $\epsilon_e^{(\nu)}$  approaches the band edges. Correspondingly, also the localization length of the Friedel density can be very large. However, the localization length  $\xi_B$  of the boundary charge or of the difference of the *total* density between the half-infinite and infinite system  $\rho(m) - \rho_{\text{bulk}}(j) = \rho_F(m) + \rho_{\text{edge}}(m)$  will turn out to be rather small since the edge state density is cancelled by a corresponding contribution of the Friedel density. A similar result has been obtained for the density of half-infinite continuum systems discussed in Ref. [[48]] based on the analytic continuation introduced in Ref. [[47]]. We summarize the results here and refer to Appendix 8.8 for the technical details.

If the chemical potential  $\mu_\nu \equiv \mu$  is placed in gap  $\nu$ ,  $\rho_F(m) + \rho_{\text{edge}}(m)$  results from summing up all Friedel densities of the filled bands  $\alpha = 1, \dots, \nu$ , together with all edge states from gaps  $\nu' = 1, \dots, \nu - 1$  and the one from gap  $\nu$  if it is occupied, i.e., if  $\epsilon_e^{(\nu)} < \mu_\nu$ . This gives

$$\rho(m) - \rho_{\text{bulk}}(j) = \sum_{\alpha=1}^{\nu} \rho_F^{(\alpha)}(m) + \rho_{\text{edge}}(m), \quad (4.169)$$

with

$$\rho_{\text{edge}}(m) = \sum_{\alpha=1}^{\nu-1} \rho_{\text{edge}}^{(\alpha)}(m) + \rho_{\text{edge}}^{(\nu)}(m) \theta(\mu_\nu - \epsilon_e^{(\nu)}), \quad (4.170)$$

where  $\rho_{\text{edge}}^{(\nu')}(n, j) = [\psi_{k_e^{(\nu')}}^e(n, j)]^2$  denotes the density of the edge state in gap  $\nu'$ . To calculate  $\rho_F^{(\alpha)}(m)$  we close the integration of Eq. (4.143) in the upper half. According to Fig. 26 one can split the integration into branch cut contributions starting at  $k_{\text{bc}}^{\alpha, \alpha-1}$  and integrations around the positions  $k_e^{\alpha, \alpha-1}$  of the edge state poles (if present in the upper half). As shown in Appendix 8.8 one obtains for the pole contribution

$$\begin{aligned} \rho_{F,P}^{(\alpha)}(m) = & \\ & - \rho_{\text{edge}}^{(\alpha)}(m) \theta(\epsilon_{\text{bc}}^{(\alpha)} - \epsilon_e^{(\alpha)}) \delta_{Z>\alpha} \\ & - \rho_{\text{edge}}^{(\alpha-1)}(m) \theta(\epsilon_e^{(\alpha-1)} - \epsilon_{\text{bc}}^{(\alpha-1)}) \delta_{\alpha>1}, \end{aligned} \quad (4.171)$$

As a result the pole contribution of the Friedel density of band  $\alpha$  cancels exactly the edge state density of those edge states which belong to band  $\alpha$ . For the total charge one would have expected this due to charge conservation but that it happens locally for each lattice site even if the edge state energies are far away from the band edges is quite surprising. It shows that edge states are not the only special effects happening at the boundary: if they appear, they always leave a corresponding fingerprint in the density of the scattering states at the boundary.

Summing all branch cut contributions over the occupied bands it turns out (see Appendix 8.8) that the common ones of adjacent bands  $\alpha$  and  $\alpha + 1$  starting at  $k_{\text{bp}}^{(\alpha)}$  cancel each other exactly since the values of the integrand left and right to the branch cut are interchanged, see Fig. 24(b,c). What remains is only the branch cut contribution from the valence band  $\alpha = \nu$

in gap  $\nu$ , which can be written as

$$\begin{aligned}\rho_{F,\text{bc}}(n, j) &= \sum_{\alpha=1}^{\nu} \rho_{F,\text{bc}}^{(\alpha)}(n, j) \\ &= \frac{1}{\pi} e^{-2\kappa_{\text{bp}}^{(\nu)} n} \text{Im} \int_0^{\infty} d\kappa \left[ \chi_{k_{\text{bp}}^{(\nu)} + i\kappa + 0^+}^{(\nu)}(j) \right]^2 e^{-2\kappa n}.\end{aligned}\quad (4.172)$$

As a result the localization length of the branch cut contribution of the total Friedel density is given by

$$\xi_{F,\text{bc}} \sim 1/\kappa_{\text{bp}}^{(\nu)}. \quad (4.173)$$

The pole contribution  $\rho_{F,P}(m) = \sum_{\alpha=1}^{\nu} \rho_{F,P}^{(\alpha)}(m)$  of the total Friedel density will cancel the density from all occupied edge states in gaps  $\nu' = 1, \dots, \nu - 1$ . However, the sum of the pole contribution of the Friedel density from gap  $\nu$  and the density of the edge state in gap  $\nu$  is only zero if both are present, i.e., for  $\epsilon_e^{(\nu)} < \mu_\nu, \epsilon_{\text{bp}}^{(\nu)}$ , or if both are absent, i.e., for  $\epsilon_e^{(\nu)} > \mu_\nu, \epsilon_{\text{bp}}^{(\nu)}$ . Therefore we obtain the following final result for the sum of the Friedel and edge state density

$$\begin{aligned}\rho(n, j) - \rho_{\text{bulk}}(j) &= \rho_F(n, j) + \rho_{\text{edge}}(n, j) \\ &= \rho_{F,\text{bc}}(n, j) + \\ &\quad + [\psi_{k_e^{(\nu)}}^e(n, j)]^2 \cdot \begin{cases} 1 & \text{for } \epsilon_{\text{bp}}^{(\nu)} < \epsilon_e^{(\nu)} < \mu_\nu \\ -1 & \text{for } \mu_\nu < \epsilon_e^{(\nu)} < \epsilon_{\text{bp}}^{(\nu)} \\ 0 & \text{otherwise} \end{cases}.\end{aligned}\quad (4.174)$$

Thereby, the second term on the r.h.s. can only occur if  $\epsilon_e^{(\nu)}$  corresponds to the energy of an edge state of  $H_R$ , i.e., if  $\text{Im}(k_e^{(\nu)}) > 0$ . Since the localization length of  $\psi_{k_e^{(\nu)}}^e(n, j)$  is given by  $\xi_e^{(\nu)} \sim 1/\kappa_e^{(\nu)}$  we get together with (4.173) the following compact form for the localization length  $\xi_B$  of the boundary charge

$$\xi_B = \begin{cases} 1/\kappa_e^{(\nu)} & \text{for } \mu_\nu < \epsilon_e^{(\nu)} < \epsilon_{\text{bp}}^{(\nu)} \text{ or } \epsilon_{\text{bp}}^{(\nu)} < \epsilon_e^{(\nu)} < \mu_\nu \\ 1/\kappa_{\text{bp}}^{(\nu)} & \text{for } \epsilon_e^{(\nu)} < \mu_\nu, \epsilon_{\text{bp}}^{(\nu)} \text{ or } \epsilon_e^{(\nu)} > \mu_\nu, \epsilon_{\text{bp}}^{(\nu)} \end{cases}.\quad (4.175)$$

As a surprising result we find the important property that the density and the localization length of the boundary charge depend only on the properties of the valence band  $\alpha = \nu$  and its analytic continuation into gap  $\nu$  (which is controlled by the  $k$ -dependence of the Bloch states close to the top of the valence band) and on the properties of the last edge state in gap  $\nu$  between the valence and conduction band. There is no dependence on the bands  $\alpha = 1, \dots, \nu - 1$  and the properties of the edge states in gaps  $\nu' = 1, \dots, \nu - 1$ . Since  $\kappa_e^{(\nu)} < \kappa_{\text{bp}}^{(\nu)}$  we find that  $\xi_B$  can only become very large if the chemical potential  $\mu_\nu$  approaches the top (bottom) of band  $\alpha = \nu$  ( $\alpha = \nu + 1$ ) and is slightly below (above) the edge state energy.

Our result is not changed when the edge state energy is identical to the energy at the branching point  $\epsilon_e^{(\nu')} = \epsilon_{\text{bp}}^{(\nu')}$  for some  $\nu' = 1, \dots, \nu$  such that a branching pole arises for the analytic continuation of the bands  $\alpha = \nu'$  and  $\alpha = \nu' + 1$ . Due to (4.91) the contribution from the branching pole of each of these bands will cancel only half of the density of the edge state, such that the sum of both cancels the whole one. This means that for all gaps  $\nu' = 1, \dots, \nu - 1$  the edge state density is again cancelled by the pole contribution of all Friedel densities. If  $\epsilon_e^{(\nu)} = \epsilon_{\text{bp}}^{(\nu)}$  happens for the gap in which the chemical potential is located, we have to add in the result (4.174) a factor  $\frac{1}{2}$  in front of  $[\psi_{k_e^{(\nu)}}^e(n, j)]^2$ .



Finally, we note that the asymptotic behaviour of  $\rho_{F,\text{bc}}(n, j)$  for  $n \gg 1$  can be evaluated from (4.172) by expanding  $N_k^{(\nu)}$  around  $k = k_{\text{bp}}^{(\nu)}$  with the help of (4.69) and (4.77). After a straightforward calculation one obtains

$$\begin{aligned} \rho_{F,\text{bc}}(n, j) &\stackrel{n \gg 1}{\sim} \frac{1}{\sqrt{n}} e^{-2\kappa_{\text{bp}}^{(\nu)} n} \\ &\times \begin{cases} a_{k_{\text{bp}}^{(\nu)}}^{(\nu)}(j)^2 & \text{for } j = 1, \dots, Z-1 \\ s(\epsilon_{k_{\text{bp}}^{(\nu)}})^2 & \text{for } j = Z \end{cases}. \end{aligned} \quad (4.176)$$

This gives rise to a universal asymptotic behaviour  $\sim \frac{1}{\sqrt{n}}$  of the pre-exponential function of the density. We note that similar observations have been reported for matrix elements of the one-particle density matrix  $\rho(x, x') = \frac{a}{2\pi} \int_{-\pi/a}^{\pi/a} dk \psi_{-k, \text{bulk}}(x) \psi_{k, \text{bulk}}(x')$  of infinite continuum systems in the asymptotic regime  $|x - x'| \rightarrow \infty$  [49] but it is so far not clear how the results are related to each other.

## 4. The surface charge theorem

In this section we will present the central proof of the unique formulation of the surface charge theorem (4.4) for a single band, relating the boundary charge to the bulk polarisation in terms of the Zak-Berry phase evaluated in a particular gauge. In particular, we will discuss the difference of the Zak-Berry phases defined with respect to the two different ways (4.2) and (4.3) to represent the Bloch wave. Based on this theorem we will present an alternative proof for the universal relation (4.5) between the boundary charges and the Zak-Berry phases of the half-infinite systems  $H_R$  and  $H_L$  with a left or a right boundary, respectively. Furthermore, we will use this theorem to relate the change of the boundary charge under a shift of the lattice by one site to the winding number of the phase difference of the Bloch wave function between adjacent sites, analog to Ref. [[52]] but generalizing it to a shift by an arbitrary number of sites.

The Zak-Berry phase  $\gamma_\alpha$  for a certain band  $\alpha$  is defined by

$$\gamma_\alpha = i \int_{-\pi}^{\pi} dk (\chi_k^{(\alpha)})^\dagger \frac{d}{dk} \chi_k^{(\alpha)} \quad (4.177)$$

$$= -\text{Im} \int_{-\pi}^{\pi} dk (\chi_{-k}^{(\alpha)})^T \frac{d}{dk} \chi_k^{(\alpha)}, \quad (4.178)$$

where we used (4.38) and (4.39) together with partial integration to derive the second equation. Using the form (4.40) of the Bloch state we find with the help of  $a_{-k}^{(\alpha)} = (a_k^{(\alpha)})^*$  and  $N_k^{(\alpha)} = N_{-k}^{(\alpha)} = (N_k^{(\alpha)})^*$

$$\gamma_\alpha = - \int_{-\pi}^{\pi} dk \frac{1}{N_k^{(\alpha)}} \text{Im} \left\{ (a_{-k}^{(\alpha)})^T \frac{d}{dk} a_k^{(\alpha)} \right\}. \quad (4.179)$$

Inserting the form (4.50) for  $a_k$  and using the central property (4.58) we get

$$\begin{aligned} -\text{Im} \left\{ (a_{-k}^{(\alpha)})^T \frac{d}{dk} a_k^{(\alpha)} \right\} &= \\ &= -\frac{1}{2i} \left( (a_{-k}^{(\alpha)})^T \frac{d}{dk} a_k^{(\alpha)} + (k \rightarrow -k) \right) \\ &= (f^T f)(\epsilon_k^{(\alpha)}) + (f^T g)(\epsilon_k^{(\alpha)}) \cos(k), \end{aligned} \quad (4.180)$$

Therefore, the Zak-Berry phase can be written in the form

$$\gamma_\alpha = \int_{-\pi}^{\pi} dk \frac{(f^T f)(\epsilon_k^{(\alpha)}) + (f^T g)(\epsilon_k^{(\alpha)}) \cos(k)}{N_k^{(\alpha)}}. \quad (4.181)$$

Comparing with (4.143) we get the result

$$Q_F^{(\alpha)} = -\frac{\gamma_\alpha}{2\pi}. \quad (4.182)$$

Combining this result with (4.159) we find that the total number of edge states of  $H_R$  is related to the total Zak-Berry phase

$$Q_E^{\text{tot}} = \frac{\gamma_{\text{tot}}}{2\pi} \quad , \quad \gamma_{\text{tot}} = \sum_{\alpha=1}^Z \gamma_\alpha. \quad (4.183)$$

We note that the Zak-Berry phase of an individual band is neither gauge invariant nor quantized. However, the total Zak-Berry phase

$$\begin{aligned} \gamma_{\text{tot}} &= i \int_{-\pi}^{\pi} dk \text{Tr} U_k^\dagger \frac{d}{dk} U_k \\ &= i \int_{-\pi}^{\pi} dk \frac{d}{dk} \ln(\det U_k) \end{aligned} \quad (4.184)$$

can be written as a winding number of the determinant of the unitary matrix  $U_k = (\chi_k^{(1)} \cdots \chi_k^{(Z)})$ . Therefore, the total Zak-Berry phase is quantized in units of  $2\pi$ . This is consistent with our result (4.183) since the total number of edge states must be quantized in integer units. However, we note that also the total Zak-Berry phase is not gauge-invariant and the precise relation (4.183) holds only in the particular gauge where  $\chi_k^{(\alpha)}(Z)$  is real.

Also the sum of the Friedel and polarization charge of a single band can be written in terms of a Zak-Berry phase. To achieve this we use a different gauge of the Bloch state  $\chi_k^{(\alpha)}$  by writing the Bloch wave function (4.31) in the form (4.3) which is the standard one within solid state physics

$$\psi_{k,\text{bulk}}^{(\alpha)}(m) = \frac{1}{\sqrt{2\pi}} \bar{\chi}_k^{(\alpha)}(j) e^{ikm/Z}, \quad (4.185)$$

with  $m = Z(n-1) + j$  labelling the lattice site and

$$\bar{\chi}_k^{(\alpha)}(j) = \chi_k^{(\alpha)}(j) e^{ik \frac{Z-j}{Z}}. \quad (4.186)$$

This form can be written in the standard form

$$\psi_{\tilde{k},\text{bulk}}^{(\alpha)}(x) = \frac{1}{\sqrt{2\pi}} u_{\tilde{k}}(x) e^{i\tilde{k}x}, \quad (4.187)$$

with  $\tilde{k} \equiv \frac{k}{L}$ ,  $x \equiv ma$ , and  $u_{\tilde{k}}(x) = u_{\tilde{k}}(x+L) \equiv \bar{\chi}_k(j)$ , where  $L = Za$  denotes the length of the unit cell. We note that  $\bar{\chi}_k^{(\alpha)} \neq \bar{\chi}_{k+2\pi}^{(\alpha)}$  is no longer periodic in  $k$ , except for the component  $\bar{\chi}_k^{(\alpha)}(Z) = \chi_k^{(\alpha)}(Z)$ . It fulfils the condition

$$\bar{\chi}_{k+2\pi}(j) = \bar{\chi}_k(j) e^{-i \frac{2\pi}{Z} j}, \quad (4.188)$$

which is equivalent to  $u_{\tilde{k} + \frac{2\pi}{L}}(x) = u_{\tilde{k}}(x)e^{-i\frac{2\pi}{L}x}$  such that  $\psi_{\tilde{k}, \text{bulk}}^{(\alpha)}$  is periodic under  $\tilde{k} \rightarrow \tilde{k} + \frac{2\pi}{L}$ . This is the standard gauge chosen for the definition of the Zak-Berry phase within the MTP

$$\bar{\gamma}_\alpha = i \int_{-\pi}^{\pi} dk (\bar{\chi}_k^{(\alpha)})^\dagger \frac{d}{dk} \bar{\chi}_k^{(\alpha)}. \quad (4.189)$$

Inserting (4.186) and using  $\frac{1}{2\pi} \int_{-\pi}^{\pi} dk |\chi_k^{(\alpha)}(j)|^2 = \rho_{\text{bulk}}^{(\alpha)}(j)$ , we find after some straightforward manipulations

$$\begin{aligned} -\frac{\bar{\gamma}_\alpha}{2\pi} &= -\frac{\gamma_\alpha}{2\pi} - \frac{1}{Z} \sum_{j=1}^Z j \rho_{\text{bulk}}^{(\alpha)}(j) + \sum_{j=1}^Z \rho_{\text{bulk}}^{(\alpha)}(j) \\ &= Q_F^{(\alpha)} + Q_P^{(\alpha)} - \frac{1}{Z^2} \sum_{j=1}^Z j + 1, \end{aligned} \quad (4.190)$$

leading to the central result of a unique formulation of the surface charge theorem for a single band

$$\boxed{Q_B^{(\alpha)} = -\frac{\bar{\gamma}_\alpha}{2\pi} - \frac{Z-1}{2Z}}, \quad (4.191)$$

where the second term on the right hand side  $-\frac{Z-1}{2Z} = \frac{1}{Z} \sum_{j=1}^Z (j-Z)\frac{1}{Z}$  can be written as a certain representation for the polarization of the ions (note that  $e = 1$  denotes the charge of the electron). This result applies to  $H_R$ , i.e., for a left boundary where the Zak-Berry phase  $\bar{\gamma}_\alpha^R \equiv \bar{\gamma}_\alpha$  is defined with respect to the unit cell  $j = 1, \dots, Z$ , starting with  $j = 1$  at the boundary. The result for  $H_L$  with a right boundary can be obtained in the same way by defining the Zak-Berry phase  $\bar{\gamma}_\alpha^L$  with respect to the unit cell  $j = Z-1, Z-2, \dots, 1, Z$ , starting with  $j = Z-1$  at the boundary, and reversing the direction of the Bloch wave. Relabelling the sites of  $H_L$  such that they look the same as for  $H_R$  we obtain after a straightforward consideration that the Bloch wave function for  $H_L$  can be written as  $\psi_{k, \text{bulk}}^{L,(\alpha)}(n, j) = \chi_k^{L,(\alpha)}(j)e^{ikn}$  with

$$\chi_k^{L,(\alpha)}(j) = \begin{cases} \chi_{-k}^{R,(\alpha)}(Z-j)e^{-ik} & \text{for } j = 1, \dots, Z-1 \\ \chi_{-k}^{R,(\alpha)}(Z) & \text{for } j = Z \end{cases}, \quad (4.192)$$

where  $\chi_k^{R,(\alpha)} \equiv \chi_k^{(\alpha)}$ . This leads with (4.186) to

$$\bar{\chi}_k^{L,(\alpha)}(j) = \bar{\chi}_{-k}^{R,(\alpha)}(Z-j), \quad (4.193)$$

where  $\bar{\chi}_k^{R,(\alpha)}(0) \equiv \bar{\chi}_k^{R,(\alpha)}(Z)$ . Using the definition (4.189) of the Zak phases  $\bar{\gamma}_\alpha^{R/L}$  with respect to  $\bar{\chi}_k^{R/L,(\alpha)}(j)$  we get the following universal relationship between the Zak phases for left/right boundaries

$$\boxed{\bar{\gamma}_\alpha^R + \bar{\gamma}_\alpha^L = 0}. \quad (4.194)$$

Since the boundary charges  $Q_B^{R,L(\alpha)}$  of  $H_{R/L}$  are given by

$$Q_B^{R,L(\alpha)} = -\frac{\bar{\gamma}_\alpha^{R,L}}{2\pi} - \frac{Z-1}{2Z}, \quad (4.195)$$

we find that (4.194) is equivalent to the universal relationship between the boundary charges for left/right boundaries

$$\boxed{Q_B^{L,(\alpha)} + Q_B^{R,(\alpha)} = -\frac{Z-1}{Z}}, \quad (4.196)$$

which is consistent with (4.152).

The surface charge theorem (4.191) is very helpful to understand how the boundary charge of a single band changes when we change the position of the boundary by  $\Delta m$  lattice sites to the right, or, equivalently, shift the lattice by  $\Delta m$  sites to the left. This is achieved by changing the phase  $\varphi \rightarrow \varphi + \frac{2\pi}{Z}\Delta m$ . To fulfil the boundary condition for the half-infinite system we need that  $\bar{\chi}_k^{(\alpha)}(Z) = \chi_k^{(\alpha)}(Z)$  is real. Since, for the shifted system, the site  $m = \Delta m$  is the last site  $j = Z$  of the unit cell, we have to take the following Bloch wave function for the shifted system

$$\tilde{\psi}_{k,\text{bulk}}(m) = e^{-i\theta_k^{(\alpha)}} \psi_{k,\text{bulk}}^{(\alpha)}(m + \Delta m), \quad (4.197)$$

where  $\theta_k^{(\alpha)}$  is the phase of  $\psi_{k,\text{bulk}}^{(\alpha)}(\Delta m)$ . Inserting (4.185) and using (4.189) we find after a straightforward algebra that this leads to a change of  $\bar{\gamma}_\alpha$  by

$$\bar{\gamma}_\alpha \rightarrow \bar{\gamma}_\alpha - 2\pi \left( \frac{\Delta m}{Z} - w_{\alpha,\Delta m} \right), \quad (4.198)$$

where  $w_{\alpha,\Delta m} = w[\theta_k^{(\alpha)}]$  is the winding number of the phase factor  $e^{i\theta_k^{(\alpha)}}$

$$w[\theta_k^{(\alpha)}] = \frac{1}{2\pi i} \int_{-\pi}^{\pi} dk e^{-i\theta_k^{(\alpha)}} \frac{d}{dk} e^{i\theta_k^{(\alpha)}}. \quad (4.199)$$

Together with (4.191) this gives for the change of the boundary charge

$$\boxed{Q_B^{(\alpha)}(\varphi + \frac{2\pi}{Z}\Delta m) - Q_B^{(\alpha)}(\varphi) = \frac{\Delta m}{Z} - w_{\alpha,\Delta m}}. \quad (4.200)$$

Since  $\psi_{k,\text{bulk}}^{(\alpha)}(m = 0)$  has been chosen real we note that  $\theta_k^{(\alpha)}$  is the gauge invariant phase difference of the Bloch wave function between site  $m = \Delta m$  and  $m = 0$ . In an analogous way we find the same result when the unshifted half-infinite system starts at  $m = m'$  and the shifted one at  $m = m' + \Delta m$ . In this case, the phase difference of the Bloch wave function between site  $m = m' + \Delta m$  and  $m = m'$  enters into the winding number. We note that (4.200) respects the periodicity  $Q_B(\varphi + 2\pi) = Q_B(\varphi)$  for  $\Delta m = Z$  since  $\theta_k^{(\alpha)} = k$  and  $w[k] = 1$  for this case.

The central result (4.200) provides a very nice interpretation of the winding numbers of phase differences of the Bloch wave function between two different sites in terms of the change of the boundary charge when one shifts the boundary. It will be discussed in all detail in Section 6. for  $\Delta m = 1$

$$w_\alpha \equiv w_{\alpha,1}, \quad (4.201)$$

together with the possible values for  $w_\alpha$ . We note that this winding number can also be expressed by the winding number  $\bar{w}_\alpha$  of the phase  $\varphi_k^{(\alpha)}(1)$  of the first component  $\chi_k^{(\alpha)}(1)$ . Since  $\theta_k^{(\alpha)} = k + \varphi_k^{(\alpha)}(1)$  we get

$$w_\alpha = 1 + \bar{w}_\alpha, \quad \bar{w}_\alpha = w[\varphi_k^{(\alpha)}(1)]. \quad (4.202)$$

## 5. Topological indices

Since the winding number  $w_\alpha$  introduced in the preceding Section 4. plays a very central role for universal properties of the boundary charge and is directly related to the phase variable of the Bloch wave function, we present in this section a comparison of this winding number to other topological indices used for the classification of topological insulators, in particular, to the Zak-Berry phase  $\gamma_\alpha$  and to the Chern number  $C^{(\alpha)}$  of a single band. This provides a more detailed analysis compared to Ref. [[52]] and shows that the winding number  $w_\alpha$  contains much more information compared to the Zak-Berry phase and the Chern number.

First, we note that  $w_\alpha$  is a gauge invariant and quantized topological index irrespective of any symmetry constraints. This is in contrast to the Zak-Berry phase  $\gamma_\alpha$  which is only quantized in case of a local inversion symmetry as was first shown in Ref. [[42]]. Local inversion symmetry is defined by a unitary operator  $\Pi$  acting on the local system with the property

$$\Pi h(\delta) \Pi^\dagger = h(-\delta) \quad \Leftrightarrow \quad \Pi h_k \Pi^\dagger = h_{-k}. \quad (4.203)$$

For our model a concrete realization is given by

$$\Pi|j\rangle = |Z - j + 1\rangle, \quad (4.204)$$

which requires

$$v_j = v_{Z-j+1} \quad , \quad t_j = t_{Z-j}, \quad (4.205)$$

with  $t_0 \equiv t_Z$ . To show the quantization we note that (4.203) implies that  $\Pi\chi_{-k}^{(\alpha)}$  is also an eigenvector of  $h_k$  with the same eigenvalue. As a consequence  $\chi_k^{(\alpha)}$  and  $\Pi\chi_{-k}^{(\alpha)}$  must agree up to a phase factor  $e^{i\eta_k^{(\alpha)}}$

$$\Pi\chi_{-k}^{(\alpha)} = e^{i\eta_k^{(\alpha)}} \chi_k^{(\alpha)}. \quad (4.206)$$

According to the definition (4.199) we denote the quantized winding number of this phase factor by  $w[\eta_k^{(\alpha)}]$ , whereas  $\gamma[\chi_k^{(\alpha)}]$  defines the Zak-Berry phase of the Bloch state  $\chi_k^{(\alpha)}$ . Using  $\chi_{-k}^{(\alpha)} = (\chi_k^{(\alpha)})^*$  we get

$$\begin{aligned} \gamma_\alpha &= \gamma[\chi_k^{(\alpha)}] = \gamma[e^{-i\eta_k^{(\alpha)}} \Pi(\chi_k^{(\alpha)})^*] \\ &= 2\pi w[\eta_k^{(\alpha)}] + \gamma[\Pi(\chi_k^{(\alpha)})^*] \\ &= 2\pi w[\eta_k^{(\alpha)}] + \gamma[(\chi_k^{(\alpha)})^*] \\ &= 2\pi w[\eta_k^{(\alpha)}] - \gamma[\chi_k^{(\alpha)}] = 2\pi w[\eta_k^{(\alpha)}] - \gamma_\alpha, \end{aligned} \quad (4.207)$$

leading to the quantization in units of  $\pi$  which was the central result of Ref. [[42]]

$$\gamma_\alpha = \pi w[\eta_k^{(\alpha)}]. \quad (4.208)$$

The relationship to the winding number  $w_\alpha$  is obtained by combining (4.206) with the concrete definition (4.204) of the operator  $\Pi$  leading to

$$\begin{aligned} e^{i\eta_k^{(\alpha)}} \chi_k^{(\alpha)}(Z) &= (\Pi\chi_{-k}^{(\alpha)})(Z) \\ &= \chi_{-k}^{(\alpha)}(1) = (\chi_k^{(\alpha)}(1))^*. \end{aligned} \quad (4.209)$$

Since  $\chi_k^{(\alpha)}(Z)$  is real this implies  $e^{i\eta_k^{(\alpha)}} = e^{-i\varphi_k^{(\alpha)}(1)}$  and  $w[\eta_k^{(\alpha)}] = -w[\varphi_k^{(\alpha)}(1)]$ , where  $\varphi_k^{(\alpha)}(1)$  is the phase of the first component  $\chi_k^{(\alpha)}(1)$ . Together with (4.202) this leads to the central result

$$\begin{aligned} w_\alpha &= 1 + w[\varphi_k^{(\alpha)}(1)] \\ &= 1 - w[\eta_k^{(\alpha)}] = 1 - \frac{\gamma_\alpha}{\pi}, \end{aligned} \quad (4.210)$$

where we used (4.208) in the last step. As a consequence we find in the special case of local inversion symmetry that the quantization of the Zak-Berry phase in units of  $\pi$  is directly related to the quantization of the winding number  $w_\alpha$  which, up to our knowledge, is an interesting fact not noticed previously. It shows that the winding number  $w_\alpha$  is a much more general topological index which can be also used in the *absence* of inversion symmetry, in contrast to the standard class of topological crystalline insulators [[17]-[22]] which rely on the quantization of the Zak-Berry phase. The Zak-Berry phase is not a gauge invariant quantity and, in the absence of inversion symmetry, it is even not quantized.

In summary we conclude that the Zak-Berry phase  $\gamma_\alpha$  itself is only a useful physical concept to characterize topological properties in case of local inversion symmetry. In this case its quantization in units of  $\pi$  leads via (4.182) to a quantization of the Friedel charge  $Q_F^{(\alpha)}$  in half-integer units. Since the polarization charge (4.133) is zero in the presence of local inversion symmetry, due to the property  $\rho_{\text{bulk}}^{(\alpha)}(j) = \rho_{\text{bulk}}^{(\alpha)}(Z - j + 1)$ , this means that also the boundary charge  $Q_B^{(\alpha)}$  is quantized in half-integer units which can be measured experimentally. In the absence of local inversion symmetry the winding number  $w_\alpha$  is a more general index to characterize topological properties and is related to the universal properties of the *change* of the boundary charge when the boundary is shifted to a different position.

In contrast to the Zak-Berry phase  $\gamma_\alpha$  the Chern number  $C^{(\alpha)}$  is a topological index which, similar to  $w_\alpha$ , is also quantized and gauge invariant in the absence of any symmetry constraints (cf. Refs. [[43, 66]]). It is defined via an integral over the Berry curvature  $F^{(\alpha)}$  as

$$C^{(\alpha)} = \frac{1}{2\pi} \int_{-\pi/2}^{3\pi/2} dk \int_0^{2\pi} d\varphi F^{(\alpha)}(k, \varphi), \quad (4.211)$$

with

$$F^{(\alpha)} = \partial_k A_\varphi^{(\alpha)} - \partial_\varphi A_k^{(\alpha)}. \quad (4.212)$$

Here, the vector  $\vec{A}^{(\alpha)} = (A_k^{(\alpha)}, A_\varphi^{(\alpha)})$  is the Berry connection defined by

$$A_k^{(\alpha)} = i(\chi_k^{(\alpha)})^\dagger \partial_k \chi_k^{(\alpha)}, \quad A_\varphi^{(\alpha)} = i(\chi_k^{(\alpha)})^\dagger \partial_\varphi \chi_k^{(\alpha)}. \quad (4.213)$$

We note that the Chern number is gauge invariant and does not change when one replaces  $\chi_k^{(\alpha)}$  by the different gauge  $\bar{\chi}_k^{(\alpha)}$  defined in (4.186). This leads to an additional term  $\partial_\varphi \sum_j |\chi_k^{(\alpha)}(j)|^2 \frac{Z-j}{Z}$  for the Berry curvature which gives zero when inserted in (4.211).

We have written the integral over  $k$  in (4.211) from  $k = -\pi/2$  to  $k = 3\pi/2$  since the singularities of the Berry connection appear at the entering/leaving points  $\varphi_{i\pm}^{(\alpha)}$  of the edge states which occur for  $k = k_0$ , with  $k_0 = 0$  or  $k_0 = \pi$ , see Fig. 35. This follows from the fact that all analytic quantities  $a_{k_0}^{(\alpha)}(j)$ ,  $s(\epsilon_{k_0}^{(\alpha)})$  and  $N_{k_0}^{(\alpha)}$  are zero for these particular phase values but the Bloch state remains finite. In addition, at this point the phases of all components of the Bloch state become the same and identical to the phase of the first component  $\chi_k^{(\alpha)}(1) = a_k^{(\alpha)}(1)/\sqrt{N_k}$ . This is very important to prove the quantization of the Chern number and its relation to  $w_\alpha$ . To show this we first note that close to one of the singularities  $(k_0, \varphi_{i\sigma_i}^{(\alpha)})$  we get up to linear order in  $k - k_e^{(\nu)}$  and  $\varphi - \varphi_{i\sigma_i}^{(\alpha)}$  (compare with (4.329))

$$\epsilon_k^{(\alpha)}(\varphi) \approx \epsilon_e^{(\nu)}(\varphi), \quad (4.214)$$

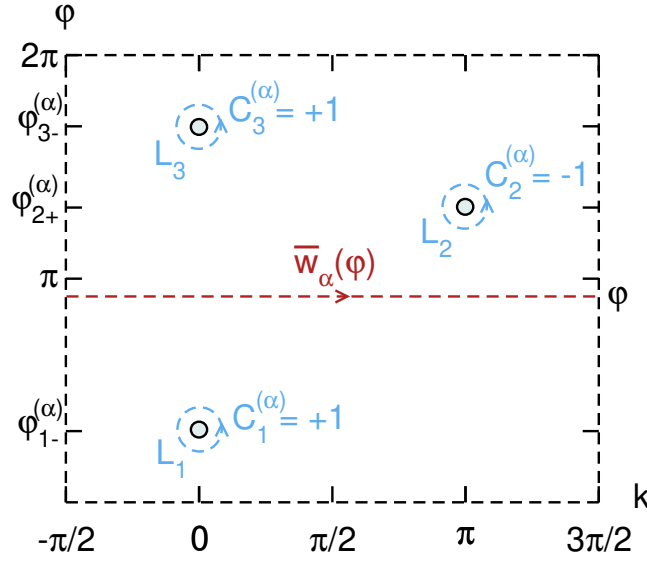


Figure 35: Singularities of the Berry connection in  $(k, \varphi)$ -space at the points where edge states enter or leave band  $\alpha$ . We have shown one edge state entering the band at  $\varphi_{2+}^{(\alpha)}$  for  $k = \pi$  and two edge states leaving the band at  $\varphi_{1-}^{(\alpha)}$  and  $\varphi_{3-}^{(\alpha)}$  for  $k = 0$ . The Chern number  $C^{(\alpha)} = \sum_i C_i^{(\alpha)}$  is given by the negative sum over all closed integrals over the paths  $L_i$  of the Berry connection around the singularities at  $(k_0, \varphi_{i\pm}^{(\alpha)})$ , with  $k_0 = 0$  or  $k_0 = \pi$ . According to Eq. (4.217), this is identical to the sum over the windings  $C_i^{(\alpha)}$  of the phase of  $a_k^{(\alpha)}(1)$  around the singularities. The dashed line indicates the path along which the winding number  $\bar{w}_\alpha(\varphi)$  is defined which determines the invariant  $I_\alpha(\varphi)$  via Eq. (4.232). When  $\varphi$  crosses  $\varphi_{i\pm}^{(\alpha)}$  from below the winding number  $\bar{w}_\alpha(\varphi)$  jumps by  $\pm 1$  according to Eq. (4.219) leading to  $C_i^{(\alpha)} = \mp 1$ , see Eq. (4.218). Note that  $\bar{w}_\alpha(\varphi)$  jumps also by  $\mp 1$  at the phase values  $\varphi_{i\pm}^{(\alpha)} - \frac{2\pi}{Z}$  according to Eq. (4.219) such that  $\bar{w}_\alpha(\varphi) = \bar{w}_\alpha(\varphi + 2\pi)$  is fulfilled.



where  $\nu$  is the index of the edge state corresponding to gap  $\nu = \alpha$  ( $\nu = \alpha - 1$ ) if the edge state enters/leaves at the top (bottom) of the band. This follows since the edge state leaves/enters the band in a smooth way as function of  $\varphi$  together with  $\partial_k \epsilon_k^{(\alpha)} = 0$  at  $k = k_0$ , see (4.65). Inserting (4.214) in (4.59) and using  $s(\epsilon_e^{(\nu)}) = 0$  we find

$$t_1 \cdots t_{j-1} a_k^{(\alpha)}(j) \approx d_{1,j-1}(\epsilon_e^{(\nu)}) a_k^{(\alpha)}(1) \quad (4.215)$$

up to linear order in  $k - k_e^{(\nu)}$  and  $\varphi - \varphi_{i\sigma_i}^{(\alpha)}$ , showing that close to all singularities the phases of all components are identical to the phase  $\varphi_k^{(\alpha)}(1)$  of the first component of the Bloch state. Using the normalization  $\sum_{j=1}^Z |\chi_k^{(\alpha)}(j)|^2 = 1$ , this means, that for  $(k, \varphi)$  close to the singularity  $(k_0, \varphi_{i\sigma_i}^{(\alpha)})$ , we get for the Berry connection vector

$$\vec{A}^{(\alpha)} \approx -(\partial_k, \partial_\varphi) \varphi_k^{(\alpha)}(1). \quad (4.216)$$

Taking this result together with Stokes theorem, relating the Chern number (4.211) to the sum over the closed integrals of the Berry connection around the singularities, we get

$$C^{(\alpha)} = - \sum_i \oint_{L_i} \vec{A}^{(\alpha)} d\vec{l} = \sum_i \oint_{L_i} d\varphi_k^{(\alpha)}(1), \quad (4.217)$$

where  $L_i$  denotes the counterclockwise closed curve around the corresponding singularity (see Fig. 35). As a result the Chern number is given by the sum  $C^{(\alpha)} = \sum_i C_i^{(\alpha)}$  over the winding numbers  $C_i^{(\alpha)}$  of  $a_k^{(\alpha)}(1)$  around the singularity at  $(k_0, \varphi_{i\sigma_i}^{(\alpha)})$ . We note that this winding number has to be distinguished from the winding number  $\bar{w}_\alpha(\varphi)$  defined in (4.202) which describes the winding of  $a_k^{(\alpha)}(1)$  along the path  $k = 0 \rightarrow k = 2\pi$  at fixed phase  $\varphi$ , see Fig. 35. Obviously the two are related by

$$C_i^{(\alpha)} = \bar{w}_\alpha(\varphi_{i\sigma_i}^{(\alpha)} - 0^+) - \bar{w}_\alpha(\varphi_{i\sigma_i}^{(\alpha)} + 0^+), \quad (4.218)$$

i.e.,  $C_i^{(\alpha)}$  is identical to the *negative jump* of  $\bar{w}_\alpha(\varphi)$  when  $\varphi$  crosses a point  $\varphi_{i\sigma_i}^{(\alpha)}$  where an edge state enters/leaves the band.

Since  $\bar{w}_\alpha(\varphi)$  can be related via (4.200) and (4.202) to the change of the boundary charge as

$$\bar{w}_\alpha(\varphi) = -1 - \frac{1}{Z} - Q_B^{(\alpha)}(\varphi + \frac{2\pi}{Z}) + Q_B^{(\alpha)}(\varphi), \quad (4.219)$$

we find that the jump of  $\bar{w}_\alpha(\varphi)$  is given by the jump of  $Q_B^{(\alpha)}(\varphi)$  at  $\varphi = \varphi_{i\sigma_i}^{(\alpha)}$  which is identical to  $\sigma_i$ . We note that  $Q_B^{(\alpha)}(\varphi + \frac{2\pi}{Z})$  does not jump at  $\varphi = \varphi_{i\sigma_i}^{(\alpha)}$  since no edge state can appear at  $\varphi = \varphi_{i\sigma_i}^{(\alpha)} + \frac{2\pi}{Z}$ , see Fig. 27 and the discussion in Section 2.6. Denoting by  $M_\pm^{(\alpha)}$  the total number of entering/leaving points of edge states, we find

$$\begin{aligned} C^{(\alpha)} &= \sum_i C_i^{(\alpha)} = - \sum_i \sigma_i \\ &= M_-^{(\alpha)} - M_+^{(\alpha)} \equiv M^{(\alpha)}. \end{aligned} \quad (4.220)$$

Due to (4.219) we note that  $\bar{w}_\alpha(\varphi)$  jumps also at all phase values  $\varphi_{i\sigma} - \frac{2\pi}{Z}$  such that  $\bar{w}_\alpha(\varphi) = \bar{w}_\alpha(\varphi + 2\pi)$  is a periodic function and the total Chern number can not be calculated by the overall jump of  $\bar{w}_\alpha$  across a certain phase interval. Note that the relation (4.220) was originally established in Ref. [[43]], and it is often referred to as the bulk-boundary correspondence.

In summary, we find that the winding number  $w_\alpha(\varphi)$  contains much more information compared to the Chern number  $C^{(\alpha)}$ . The Chern number is a phase-integrated quantity which

measures the sum of the jumps of  $w_\alpha(\varphi)$  at the phase values  $\varphi = \varphi_{i,\pm}$  where edge states enter or leave the band. Neither the value of  $w_\alpha$  itself nor the precise positions of the jumps enter into the Chern number. If  $\varphi$  is an effective phase resulting from the quasimomentum  $k_y$  perpendicular to the boundary of a two-dimensional system, it can be shown that the total Chern number of all filled bands  $C_\nu = \sum_{\alpha=1}^\nu C^{(\alpha)}$  is related to the plateau values of the transverse conductance [51]. In contrast,  $w_\alpha(\varphi)$  measures via (4.219) the whole phase-dependence of the change  $\Delta Q_B^{(\alpha)}(\varphi)$  of the boundary charge when the phase is changed by  $\frac{2\pi}{Z}$ .

We note that the winding numbers indicated in Fig. 35 along the various paths defining  $C_i^{(\alpha)}$  and  $\bar{w}_\alpha$  are the windings of the phase of the first component  $a_k^{(\alpha)}(1)$  of the Bloch state. The corresponding line integrals of the Berry connection are only the same (up to a sign) for the curves  $L_i$  (defining  $C_i^{(\alpha)}$ ) but *not* for the horizontal lines. The latter integral would result in the Zak-Berry phase (4.177) and one can also define the Chern number via the sum of the jumps of the Zak-Berry phase at the phase values  $\varphi = \varphi_{i,\pm}$  where edge states enter or leave the band. This can also be interpreted as a winding number [44] but we emphasize that, in contrast to  $w_\alpha(\varphi)$ , the Zak-Berry phase  $\gamma_\alpha(\varphi)$  itself is *not* a winding number and is neither quantized nor gauge-invariant as outlined above.

Finally, we note that when the Chern number is summed over all filled bands, we find that many terms cancel since the number of entering (leaving) modes at the bottom of band  $\alpha$  is identical to the number of leaving (entering) modes at the top of band  $\alpha - 1$ . Therefore, in consistency with many previous works [51, 43, 55, 56, 61, 66], only those edge states entering/leaving the top of band  $\alpha = \nu$  remain and the following result is obtained for the total Chern number

$$\begin{aligned} C_\nu &= \sum_{\alpha=1}^\nu C^{(\alpha)} = M_-^{(\nu)} - M_+^{(\nu)} \\ &= M_-(\mu_\nu) - M_+(\mu_\nu) = M_\nu, \end{aligned} \quad (4.221)$$

where  $M_\pm(\mu_\nu)$  denotes the total number of edge states crossing the chemical potential  $\mu_\nu$  from above/below, already introduced in (4.111). According to the discussion in Section 2.6 we note that the difference  $M_\nu = M_-(\mu_\nu) - M_+(\mu_\nu)$  does not depend on the precise position of the chemical potential inside the gap but only on the index  $\nu$  of the gap.

## 6. Universal properties

The topic of this section are the universal properties of the boundary charge as proposed in Ref. [[52]] that will be proven rigorously here based on the topological constraints for edge states derived in Section 2.6. We review the physical picture based on charge conservation and particle-hole duality as in Ref. [[52]] and provide a proof of the essential ingredient that the phase-dependence of the model parameters can always be chosen such that no edge states cross the chemical potential in a certain gap within a phase interval of size  $\frac{2\pi}{Z}$ . We provide a rigorous proof of the two central results of Ref. [[52]] why the invariants defined for a single band or for a given chemical potential are quantized and which values are allowed. In addition, we treat the case where the wavelength of the modulations is any rational number and discuss the expectations for multi-channel systems.

### 6.1 Physical picture

We start with physical arguments what we expect for the change of  $Q_B$  if the lattice is shifted by one site towards the boundary. This means that the lattice of  $H_R$  starts not with site  $(n, j) = (1, 1)$  (as in Fig. 22) but with  $(n, j) = (1, 2)$ . Formally, we achieve this by changing the phase variable  $\varphi$  by  $\frac{2\pi}{Z}$  such that  $v_j \rightarrow v_{j+1}$  and  $t_j \rightarrow t_{j+1}$ . This is a very fundamental

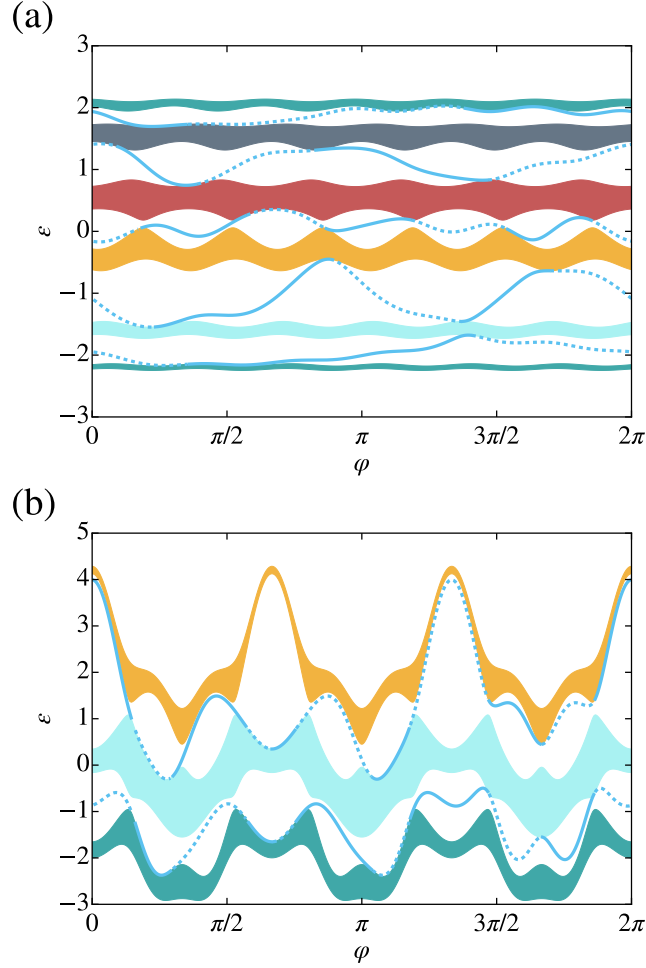


Figure 36: (a) The band structure and edge states analog to Fig. 25(b) for  $Z = 6$ ,  $V = 0.5$ ,  $t = 1$ ,  $\delta t = 0.1$  and three random Fourier coefficients for the real functions  $F_v$  and  $F_t$  in Eqs. (4.10) and (4.11) according to the form (4.273), see Supplemental Material for the precise parameters [26]. For the gaps  $\nu = 1, 2, 3$  there are  $\nu$  edge states of  $H_R$  (blue solid lines) moving upwards whereas, for the gaps  $\nu = 4, 5$ , there are  $Z - \nu$  edge states of  $H_R$  moving downwards. For band  $\alpha = 4$  all edge states of  $H_R$  are entering the band,  $Z - \alpha = 2$  from above and  $\alpha - 1 = 3$  from below, i.e., in total  $Z - 1 = 5$  edge states. (b) The band structure and edge states analog to Fig. 25(b) for  $Z = 3$ ,  $V = 0.5$ ,  $t = 1$ ,  $\delta t = 0$  and five random Fourier coefficients for the real function  $F_v$  in Eq. (4.10) according to the form (4.273), see Supplemental Material for the precise parameters [26]. Several edge states of  $H_R$  return to the same band in both gaps  $\nu = 1, 2$ . For band  $\alpha = 2$  all edge states of  $H_R$  are entering the band (disregarding the ones which return to the same band),  $Z - \alpha = 1$  from above and  $\alpha - 1 = 1$  from below, i.e., in total  $Z - 1 = 2$  edge states.

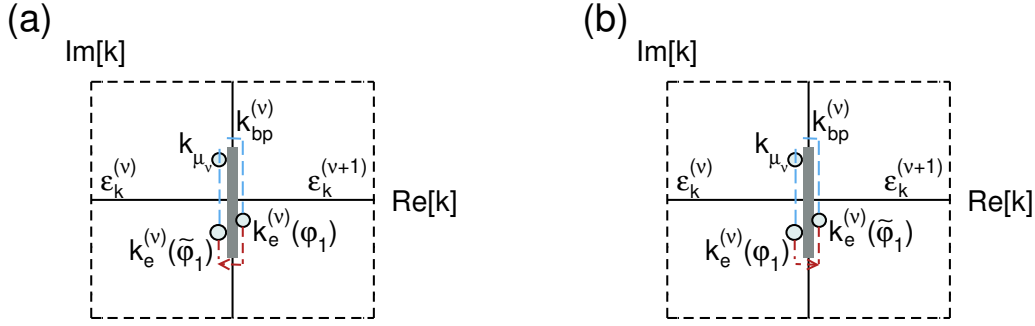


Figure 37: Sketch of the branch cut separating band  $\nu$  from  $\nu+1$ , corresponding to the analytic continuation of Fig. 26(a). The position of  $k_{\mu_\nu}$  corresponds to  $\mu_\nu$  and is defined via  $\epsilon_{k_{\mu_\nu}} = \mu_\nu$  and  $\text{Im}(k_{\mu_\nu}) > 0$ . The positions of  $k_e^{(\nu)}(\varphi_1)$  and  $k_e^{(\nu)}(\tilde{\varphi}_1)$  correspond to the edge state poles at the phases  $\varphi_1$  and  $\tilde{\varphi}_1 = \varphi_1 + \frac{2\pi}{Z}$ , respectively. Two possibilities are shown in (a) and (b) for their position. Obviously, for both cases one can choose the orientation of the path  $k_e^{(\nu)}(\varphi)$  of the edge state pole for all phases  $\varphi_1 < \varphi < \tilde{\varphi}_1$  in such a way that no crossing through  $k_{\mu_\nu}$  occurs (red part of the contour). We have not indicated in the figure the weak dependence of  $k_{\mu_\nu}$  and  $k_{bp}^{(\nu)}$  on  $\varphi$ .

question of how observables defined at the boundary depend on the way one cuts off an infinite system to define the boundary. Since the boundary charge contains also the charge  $Q_E$  of the edge states this is also related to the question of how the appearance and energy of edge states depend on the way one defines the boundary. As can be seen already from Fig. 25(b) the appearance and energy of edge states of  $H_R$  (shown as solid blue lines) depend crucially on the phase variable  $\varphi$ , a fact very well known from the integer QHE in 2D systems (where the quasimomentum  $k_y$  in  $y$ -direction plays the role of  $\varphi$ ) [51, 56]. An illustrative but very fundamental example in this respect is the SSH model for  $Z = 2$  and  $v_j = 0$ . This model is parametrized only by two hopping parameters  $t_1$  and  $t_2$  and it is obviously not important for the bulk properties whether  $t_1 > t_2$  or  $t_1 < t_2$ , one can just interchange the two hoppings to get one or the other case. If the system starts with site  $(n, j) = (1, 1)$  it is obvious that  $t_1 < t_2$  leads to the appearance of an edge state of  $H_R$  (since an electron can not leave the first site of the system in the limiting case  $t_1 = 0$ ). Similarly, if the system starts with site  $(n, j) = (1, 2)$  one needs the condition  $t_1 > t_2$  to get an edge state of  $H_R$ . This property is generic for any  $Z$ , edge states of  $H_R$  appear only in certain regions of  $\varphi$ . The underlying physics for the appearance of edge states as function of  $\varphi$  is very obvious and is related to charge pumping and charge conservation [50, 51]. If the phase variable is changed adiabatically in time by  $2\pi$ , the charge  $\nu$  of a whole unit cell (corresponding to the number  $\nu$  of filled bands when the chemical potential  $\mu_\nu$  is located in gap  $\nu$ ) has been moved into the boundary. Since the boundary charge does not change for a phase change by  $2\pi$  (the Hamiltonian  $H_R$  is exactly the same) the charge  $\nu$  must be taken away by exactly  $\nu$  edge states which move above the chemical potential during this process and move the charge to higher bands, see Fig. 25(b). As can be seen for the gaps  $\nu = 4, 5$  in Fig. 36(a) it can also happen that  $Z - \nu$  edge states of  $H_R$  run downwards from band  $\alpha = \nu + 1$  to band  $\alpha = \nu$ . This can be understood by an adiabatic pumping process in terms of the hole picture described in Section 3.2. Due to Eq. (4.164) on average the hole charge  $\nu - Z$  is shifted into the boundary when the phase changes by  $2\pi$ . Since the boundary charges of the holes and particles are the same this means that  $Z - \nu$  edge states have to move below the chemical potential to compensate this charge. Furthermore, it can happen that edge states return to the same band, see Fig. 36(b). Which case appears, depends on the model parameters and how the phase-dependence is chosen via the functions  $F_v$  and  $F_t$  in Eqs. (4.10) and (4.11).

The two ways of how our Gedankenexperiment of charge pumping can be interpreted in terms of particles and holes is very fundamental for an intuitive understanding of the physics. That each configuration of the edge states in a certain gap can be explained either by the particle or hole version of our Gedankenexperiment but not by the other relies on the Pauli principle which has no classical analog. Whereas in the particle picture the boundary charge *increases* by  $\nu$  during a phase change of  $2\pi$  and has to be compensated by  $\nu$  edge states moving *above*  $\mu_\nu$ , in the hole picture the boundary charge *decreases* by  $Z - \nu$  and has to be compensated by  $Z - \nu$  edge states moving *below*  $\mu_\nu$ . This will give rise to two completely different line shapes of the boundary charge as function of the phase variable  $\varphi$ , see Fig. 38(c) for an example which will be discussed below in all detail.

One can also think of occupying only one single band  $\alpha$  such that one charge is moved into the boundary after a phase change of  $2\pi$ . Therefore, one expects that during this process in total one edge state has to leave this band. In fact, in Figs. 25(b) and 36(a,b) it is the case for most of the bands that either  $\alpha - 1$  edge states enter into the band bottom and  $\alpha$  edge states leave from the band top or  $Z - \alpha$  edge states enter into the band top and  $Z - \alpha + 1$  edge states leave from the band bottom, such that in both cases one edge state leaves in total. However, in each of the figures one band is special in the sense that  $Z - 1$  edge states enter the band,  $Z - \alpha$  ones from above and  $\alpha - 1$  ones from below. This case can be understood in the hole picture since the hole charge  $1 - Z$  has been moved into the boundary after a phase change of  $2\pi$  which has to be compensated by  $Z - 1$  edge states entering the band.

We now consider the more fundamental issue of how the boundary charge changes when we move the system only by *one* site towards the boundary. We call this change

$$\begin{aligned}\Delta Q_B &\equiv \Delta Q_B(\varphi, \mu_\nu) \\ &= Q_B(\varphi + \frac{2\pi}{Z}, \mu_\nu) - Q_B(\varphi, \mu_\nu),\end{aligned}\tag{4.222}$$

which depends on the phase  $\varphi$  and the chemical potential  $\mu_\nu$  in gap  $\nu$  (note that edge states of  $H_R$  can be present in the gap such that  $Q_B$  depends on the precise value of  $\mu_\nu$  in gap  $\nu$ ). In this case we expect that *on average* the charge  $\bar{\rho} = \frac{\nu}{Z}$  is moved into the boundary. Since  $Q_B$  is defined via a *macroscopic average* on length scales much larger than the size of a unit cell, we expect the same for  $\Delta Q_B = \bar{\rho}$  if, in addition, no edge state is moving above/below  $\mu_\nu$  during the shift. So far we have only involved the physics of classical charge conservation to get this result. If we take in addition the Pauli principle into account and look at the same process from the hole point of view on average the hole charge  $\Delta Q_{B,h} = \bar{\rho}_h = \bar{\rho} - 1$  is moved into the boundary during the shift, which gives  $\Delta Q_B = \Delta Q_{B,h} = \bar{\rho} - 1$ . Therefore, we expect that the following two universal values are possible for the change of  $\Delta Q_B$

$$\Delta Q_B(\varphi, \mu_\nu) \in \{\bar{\rho}, \bar{\rho} - 1\}.\tag{4.223}$$

It is very important to realize that this result does *not* involve any edge state physics but relies only on classical charge conservation together with the Pauli principle. If additional edge states move above/below  $\mu_\nu$  during the shift one expects Eq. (4.223) to change mod(1). However, for a given set of parameters  $\{v_j(\varphi_1), t_j(\varphi_1)\}$  at phase  $\varphi = \varphi_1$  and, correspondingly, via Eq. (4.9) also at all phases shifted by multiples of  $\frac{2\pi}{Z}$ , the value of  $\Delta Q_B(\varphi_1, \mu_\nu)$  does *not* depend on the particular choice of the phase dependence of  $v_j(\varphi)$  and  $t_j(\varphi)$  for all phases  $\varphi_1 < \varphi < \varphi_1 + \frac{2\pi}{Z}$ . For our 1-channel model with a non-degenerate spectrum it is shown in Appendix 8.10 that, for any given phase interval of size  $\frac{2\pi}{Z}$ , the phase-dependence can always be chosen such that no edge state of  $H_R$  crosses the chemical potential  $\mu_\nu$  in gap  $\nu$  as function of  $\varphi$  in this interval. This is related to the fact that the complex quasimomentum  $k_e^{(\nu)}(\varphi)$  of the edge state pole moves around the branch cut defined between band  $\nu$  and  $\nu + 1$  as function of  $\varphi$ , see Fig. 26(a) and movies provided in the Supplemental Material [26]. This movement

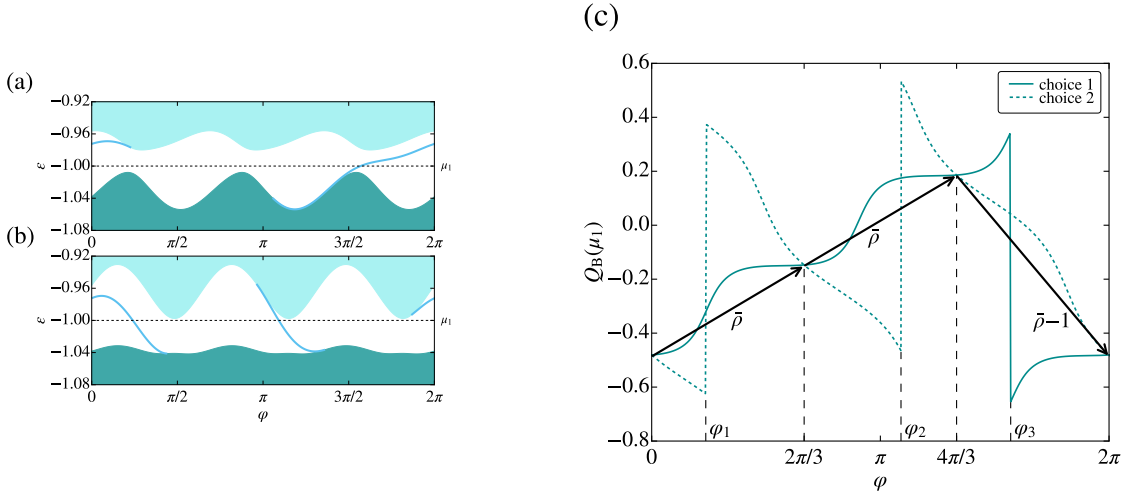


Figure 38: In (a) and (b) we show the phase-dependence of the band structure of the first two bands and the edge states of  $H_R$  in the first gap (blue lines) for  $Z = 3$ ,  $V = 0.1$ ,  $t = 1$ ,  $\delta t = 0.1$  for two functions  $F_v$  and  $F_t$  in Eqs. (4.10) and (4.11) taken from (4.276) with fixed and random parameters for  $v_j(0)$  and  $t_j(0) - t$  at phase  $\varphi = 0$  but different choices for the phase-dependence in between [via the random parameters  $v_j^{(1)}$  and  $t_j^{(1)}$  used in (4.282)], see the Supplemental Material [26] for the concrete parameters. In (a) one edge state appears running from the lower to the upper band whereas in (b) two edge states run from the upper to the lower band. In (c) the phase dependence of  $Q_B$  is shown for the two choices of (a) and (b) with  $\mu_1$  located in the first gap [see dashed line in (a) and (b)]. At  $\varphi = 0$  and  $\varphi = 2\pi/3$ , we get the change  $\Delta Q_B(0) = \Delta Q_B(2\pi/3) = \bar{\rho} = \frac{1}{3}$  (see the two left arrow). For the first choice  $Q_B$  increases on average by  $\frac{1}{Z} = \frac{1}{3}$  on both intervals whereas for the second choice  $Q_B$  decreases on average by  $\frac{1}{Z} - 1 = -\frac{2}{3}$  and obtains a jump by  $+1$  at  $\varphi = \varphi_{1,2}$  where an edge state moves below  $\mu_1$ . For  $\varphi = 4\pi/3$ , we get  $\Delta Q_B(4\pi/3) = \bar{\rho} - 1 = -\frac{2}{3}$  (see right arrow). Here, for the first choice, an edge state moves above  $\mu_1$  at  $\varphi = \varphi_3$  leading to a jump of  $Q_B$  by  $-1$  whereas for the second choice no edge state is involved.

can have both orientations depending on the choice of the phase-dependence of the parameters. Defining the complex quasimomentum  $k_{\mu_\nu}$  corresponding to the chemical potential uniquely via  $\epsilon_{k_{\mu_\nu}} = \mu_\nu$  and  $\text{Im}(k_{\mu_\nu}) > 0$ , the condition that an edge state of  $H_R$  does not cross  $\mu_\nu$  for all  $\varphi_1 < \varphi < \varphi_1 + \frac{2\pi}{Z}$  is equivalent to the condition that  $k_e^{(\nu)}(\varphi)$  does not cross  $k_{\mu_\nu}(\varphi)$  in this interval. This can always be achieved by choosing the phase-dependence of the parameters such that the orientation of the movement of  $k_e^{(\nu)}(\varphi)$  has the appropriate sign, see Fig. 37(a,b). As a result we find that the two values predicted in Eq. (4.223) are the only allowed values for the single-channel case and are *not* related to any edge state physics.

As an example we demonstrate in Fig. 38 that a different parametrization for the phase-dependence gives rise to very different interpretations of the change of the boundary charge. In Fig. 38(a,b) we show the edge states of  $H_R$  in the first gap  $\nu = 1$  for  $Z = 3$  and for given parameter sets of  $t_j(\varphi)$  and  $v_j(\varphi)$  at all phases  $\varphi = 2\pi j/3$ , with  $j = 1, \dots, 3$ , but with two different choices of how the phase-dependence is defined in between. In Fig. 38(a) one edge state is moving from the lower to the upper band whereas in Fig. 38(b) two edge states are moving from the upper to the lower band. Fig. 38(c) shows the corresponding phase-dependence of the boundary charge  $Q_B$  for the two choices when the chemical potential  $\mu_1$  is located in the first gap [dashed line in Figs. 38(a,b)]. Obviously, at  $\varphi = 2\pi j/3$ , with  $j = 1, 2, 3$ , the value of  $Q_B$  must be the same for the two choices but the line shape in between is completely different. Comparing the two points  $\varphi = 0, \frac{2\pi}{3}$  or  $\varphi = \frac{2\pi}{3}, \frac{4\pi}{3}$ , we get  $\Delta Q_B = \bar{\rho} = \frac{1}{3}$  in both cases [see the two left arrows in Fig. 38(c)]. For the first choice  $Q_B$  is increasing monotonously by  $\frac{1}{3}$  without



any edge state involved. In contrast, for the second choice,  $Q_B$  decreases monotonously by  $\bar{\rho} - 1 = -\frac{2}{3}$  but at  $\varphi = \varphi_{1,2}$  an edge state moves below  $\mu_1$  such that  $Q_B$  gets a discontinuous jump by  $+1$  and the same result is obtained for  $\Delta Q_B = \frac{1}{3}$ . Similarly, comparing  $Q_B$  between  $\varphi = \frac{4\pi}{3}$  and  $\varphi = 2\pi$  we get  $\Delta Q_B = \bar{\rho} - 1 = -\frac{2}{3}$  for both choices. Here, the situation is the other way around, for the first choice an edge state moves above  $\mu_1$  at  $\varphi = \varphi_3$  whereas for the second choice no edge state is involved, leading again to the same net result for  $\Delta Q_B$ . Therefore, we conclude that just by looking at the change of  $Q_B$  when the system is cut off at a different site at the boundary, it is not unambiguous to interpret the value of  $\Delta Q_B$  in terms of edge state physics. Instead, the correct interpretation is in terms of charge conservation and the Pauli principle as explained above, the edge states just play the role of “followers” and can appear in one or the other form depending on the concrete choice of the phase-dependence.

The result (4.223) holds if the chemical potential  $\mu_\nu$  is located in gap  $\nu$ . We note that it is not essential that the chemical potential is a constant, it can as well have a phase dependence provided it is  $\frac{2\pi}{Z}$ -periodic:  $\mu_\nu(\varphi) = \mu_\nu(\varphi + \frac{2\pi}{Z})$ . This is necessary for band structures like in Fig. 36(b) where two bands can not be separated by a fixed energy. If we consider the boundary charge  $Q_B(\varphi, \mu_\nu, \mu_{\nu'}) = Q_B(\varphi, \mu_\nu) - Q_B(\varphi, \mu_{\nu'})$  of all states lying between two energies  $\mu_{\nu'}$  and  $\mu_\nu$  located in gaps  $\nu' < \nu$  and  $\nu$ , respectively, the corresponding change is obtained by taking the difference of (4.223) for  $\nu$  and  $\nu'$

$$\begin{aligned} \Delta Q_B(\varphi, \mu_\nu, \mu_{\nu'}) &= \Delta Q_B(\varphi, \mu_\nu) - \Delta Q_B(\varphi, \mu_{\nu'}) \\ &\in \{\bar{\rho} - \bar{\rho}', \bar{\rho} - \bar{\rho}' \pm 1\} \end{aligned} \quad (4.224)$$

with  $\bar{\rho} = \frac{\nu}{Z}$  and  $\bar{\rho}' = \frac{\nu'}{Z}$ , giving rise to three possible values. For example, for a single band  $\alpha$  [with  $\mu_\nu(\varphi)$  chosen as the band top and  $\mu_{\nu'}(\varphi) = \mu_{\nu-1}(\varphi)$  as the band bottom] we obtain

$$\Delta Q_B^{(\alpha)}(\varphi) \in \left\{ \frac{1}{Z}, \frac{1}{Z} \pm 1 \right\}. \quad (4.225)$$

Here, we see that not only the values  $\frac{1}{Z}$  and  $\frac{1}{Z} - 1$  are possible, corresponding to the particle and hole picture, respectively, when no edge state enters/leaves the band during the shift. For a single band the edge states can enter and leave at the bottom or the top of the band. Therefore, not only the edge pole encircling the branch cut between band  $\alpha$  and  $\alpha + 1$  is relevant but also the one between band  $\alpha - 1$  and  $\alpha$ . It is obvious that the phase dependence can not always be chosen such that *both* edge poles avoid crossing the energy of the band edges. As a result the edge states have to be taken into account and another value becomes possible for  $\Delta Q_B^{(\alpha)}$ . Comparing (4.225) with (4.200) for  $\Delta m = 1$  we see that the winding number  $w_\alpha$  of the phase difference of the Bloch wave function between site  $m = 1$  and  $m = 0$  can only take three possible values

$$w_\alpha \in \{0, \pm 1\}. \quad (4.226)$$

This result will be proven analytically in the next Section 6.2.

Finally, we note that the result (4.223) holds only for 1-channel tight-binding models where the spectrum is non-degenerate. For multi-channel systems with  $N_c$  weakly coupled channels several edge states encircle each branch cut of Fig. 37 and it is no longer possible to choose the phase-dependence of  $v_j(\varphi)$  and  $t_j(\varphi)$  in such a way that no edge state crosses  $\mu_\nu$  on a phase interval of size  $\frac{2\pi}{Z}$ . [67] For  $N_c$  weakly coupled channels (such that we still have  $Z - 1$  gaps)  $N_c\nu$  bands are filled when the chemical potential is located in gap  $\nu$ . Therefore, instead of Eqs. (4.162) and (4.163), we get the following result for the particle and hole charge densities

$$\rho_p(m) = \rho(m) \quad , \quad \rho_h(m) = \rho(m) - N_c, \quad (4.227)$$

$$\bar{\rho}_p = \bar{\rho} \quad , \quad \bar{\rho}_h = \bar{\rho} - N_c, \quad (4.228)$$



with  $\bar{\rho} = N_c \nu / Z$ . Since the particle and hole boundary charges are again the same we find the following two possibilities for the change of the boundary charge when no edge states move above/below  $\mu_\nu$  during the shift by one lattice site

$$\Delta Q_B = \Delta Q_{B,p} = \bar{\rho}_p = \bar{\rho}, \quad (4.229)$$

$$\Delta Q_B = \Delta Q_{B,h} = \bar{\rho}_h = \bar{\rho} - N_c. \quad (4.230)$$

Since this result can only be changed mod(1) when edge states move above or below  $\mu_\nu$  during the shift it is reasonable that the allowed values are given by

$$\Delta Q_B(\varphi, \mu_\nu) \in \{\bar{\rho}, \bar{\rho} - 1, \dots, \bar{\rho} - N_c\}, \quad (4.231)$$

since this is obviously the limiting result for vanishing coupling between the channels (where we can just add up (4.223) independently). The case of several channels will be discussed in more detail in a future work [67].

### 6.2 Invariant and boundary charge for a single band

In this section we consider the case of a single band. We define the invariant

$$I_\alpha(\varphi) = \Delta Q_B^{(\alpha)} - \frac{1}{Z} = -w_\alpha(\varphi), \quad (4.232)$$

which is identical to the negative winding number and is an integer for all  $\varphi$ . We start with the proof of (4.226), which states that the invariant can only take the values

$$I_\alpha \in \{0, \pm 1\}. \quad (4.233)$$

We first use (4.202) to relate  $w_\alpha = 1 + \bar{w}_\alpha$  to the winding number  $\bar{w}_\alpha$  of the phase  $\varphi_k^{(\alpha)}(1)$  of the first component  $\chi_k^{(\alpha)}(1)$  of the Bloch state. (4.233) then requires the proof of

$$\bar{w}_\alpha \in \{0, -1, -2\}. \quad (4.234)$$

To show that only these particular values are allowed we use the form (4.60) for  $a_k^{(\alpha)}(1) = |a_k^{(\alpha)}(1)|e^{i\varphi_k^{(\alpha)}(1)}$ , which shows that  $\bar{w}_\alpha$  is the winding of the complex number  $q_k^{(\alpha)} = (t_Z/\bar{t}^Z)a_k^{(\alpha)}(1)$  (when  $k$  changes from zero to  $2\pi$ ) with

$$q_k^{(\alpha)} = \tilde{d}_k^{(\alpha)} e^{-ik} + 1, \quad (4.235)$$

$$\tilde{d}_k^{(\alpha)} = \tilde{d}_{-k}^{(\alpha)} = \tilde{d}_{2,Z-1}(\epsilon_k^{(\alpha)}). \quad (4.236)$$

The winding is determined from the number of crossings of  $q_k^{(\alpha)}$  through the positive and negative real axis. This happens for

$$(1) \ k = 0 : \quad q_0^{(\alpha)} = d_0^{(\alpha)} + 1 \quad (4.237)$$

$$(2) \ k = \pi : \quad q_\pi^{(\alpha)} = -d_\pi^{(\alpha)} + 1 \quad (4.238)$$

$$(3) \ \tilde{d}_k^{(\alpha)} = 0 : \quad q_k^{(\alpha)} = q_{-k}^{(\alpha)} = 1. \quad (4.239)$$

It can not happen that no crossing appears on the positive real axis since this would require  $d_0^{(\alpha)} < -1$ ,  $d_\pi^{(\alpha)} > 1$  and  $d_k^{(\alpha)} \neq 0$  for all  $k$ . This is obviously not possible. The winding number

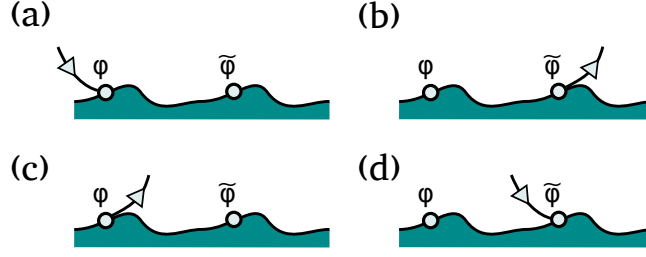


Figure 39: Four scenarios how edge states can enter/leave a band for phase  $\varphi$  and  $\tilde{\varphi} = \varphi + \frac{2\pi}{Z}$ . We have shown a band top but the same behaviour occurs for a band bottom. The quasimomentum at the band edge can be either  $k_0 = 0$  or  $k_0 = \pi$ . For (a) and (d) an edge state enters at  $\varphi$  and  $\tilde{\varphi}$ , respectively, such that the sign of  $\frac{d}{d\varphi} d_1^{(\alpha)}$  is given by  $(-1)^{k_0/\pi}$  at the entering point according to (4.335). Analog, for (b) and (c) the edge state leaves such that the sign is given by  $-(-1)^{k_0/\pi}$  at the leaving point. Together with (4.240) and (4.334) this gives the correct correlation to the jump of  $I_\alpha$  by  $-1$  for (a) and (b) and by  $+1$  for (c) and (d), see explanation in the main text.

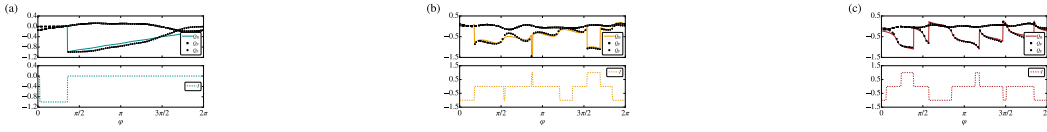


Figure 40: phase-dependence of  $Q_B^\alpha$ ,  $Q_F^\alpha$ ,  $Q_P^\alpha$  and  $I_\alpha$  for  $\alpha = 1, 3, 4$  in (a), (b) and (c), respectively, using the parameters of Fig. 36(a).

then follows from the number of crossings through the negative real axis (note that the number of crossings through  $q_k^{(\alpha)} = 1$  for  $k \neq 0, \pi$  is an even number due to  $q_k^{(\alpha)} = q_{-k}^{(\alpha)}$ )

$$\bar{w}_\alpha = \begin{cases} 0 & \text{for } \tilde{d}_0^{(\alpha)} > -1, \tilde{d}_\pi^{(\alpha)} < 1 \\ -1 & \text{for } \tilde{d}_0^{(\alpha)} < -1, \tilde{d}_\pi^{(\alpha)} < 1 \\ -1 & \text{for } \tilde{d}_0^{(\alpha)} > -1, \tilde{d}_\pi^{(\alpha)} > 1 \\ -2 & \text{for } \tilde{d}_0^{(\alpha)} < -1, \tilde{d}_\pi^{(\alpha)} > 1 \end{cases}. \quad (4.240)$$

This proves Eq. (4.233).

We now determine the cases where the invariant jumps as function of  $\varphi$ . This are the points where either  $\tilde{d}_0^{(\alpha)}(\varphi) = -1$  or  $\tilde{d}_\pi^{(\alpha)}(\varphi) = 1$ . In both cases we get  $d_1^{(\alpha)}(\varphi) = \tilde{d}_{k_0}^{(\alpha)}(\varphi) = -e^{-ik_0}$ , where  $k_0 = 0, \pi$  correspond to values at the band edges and  $d_1^{(\alpha)}(\varphi)$  has been defined in (4.326). According to (4.338) this means that an edge state appears at  $\varphi$  or at  $\tilde{\varphi} = \varphi + \frac{2\pi}{Z}$ . Thus, we conclude that the invariant  $I_\alpha$  can only jump by  $\pm 1$  at phase  $\varphi$  when an edge state enters/leaves the band at phase  $\varphi$  or at  $\tilde{\varphi}$ . This is quite obvious and consistent with the definition (4.232) of the invariant which can only jump when the boundary charge jumps at  $\varphi$  or  $\tilde{\varphi}$ . Due to charge conservation, the latter happens precisely when an edge state leaves/enters the band.

One can also understand that the jump of  $I_\alpha(\varphi)$  by  $\mp 1$  corresponds to the cases that either  $Q_B(\varphi)$  jumps by  $\pm 1$  (i.e., an edge state enters/leaves band  $\alpha$  at  $\varphi$ ) or  $Q_B(\tilde{\varphi})$  jumps by  $\mp 1$  (i.e., an edge state leaves/enters band  $\alpha$  at  $\tilde{\varphi}$ ), see Fig. 39(a-d). We have depicted the four possibilities of an edge state entering/leaving at phase  $\varphi$  or at  $\tilde{\varphi} = \varphi + \frac{2\pi}{Z}$  in Figs. 39(a-d). For the cases shown in Figs. 39(a,b) and Figs. 39(c,d) the invariant changes by  $-1$  and  $+1$ , respectively. Using (4.335) we see that  $\text{sign} \left\{ \frac{d}{d\varphi} d_1^{(\alpha)}(\varphi) \right\} = \pm (-1)^{k_0/\pi}$  for Fig. 39(a) and Fig. 39(c), respectively, where an edge state enters/leaves at  $\varphi$ , and  $\text{sign} \left\{ \frac{d}{d\varphi} d_1^{(\alpha)}(\tilde{\varphi}) \right\} = \mp (-1)^{k_0/\pi}$  for Fig. 39(b) and Fig. 39(d), respectively, where an edge state leaves/enters at  $\tilde{\varphi}$ . Using (4.334), this gives  $\text{sign} \left\{ \frac{d}{d\varphi} d_1^{(\alpha)}(\varphi) \right\} = \pm (-1)^{k_0/\pi}$  for Figs. 39(a,b) and Figs. 39(c,d), respectively. From (4.240)

we conclude that the winding number  $\bar{w}_\alpha$  jumps by  $\pm 1$  for these two cases, respectively, which corresponds to the invariant jumping by  $\mp 1$  since  $I_\alpha = -1 - \bar{w}_\alpha$ , see Eqs. (4.232) and (4.202).

The fact that the boundary charge  $Q_B^{(\alpha)}$  can only have discontinuous jumps by  $\pm 1$  when edge states enter/leave the band at the phase values  $\varphi_{i\pm}^{(\alpha)}$ , with  $i = 1, \dots, M_\pm^{(\alpha)}$ , leads together with (4.225) to the following shape of the phase-dependence of  $Q_B^{(\alpha)}$

$$Q_B^{(\alpha)}(\varphi) = f_\alpha(\varphi) + \frac{M^{(\alpha)}}{2\pi}\varphi + F_\alpha(\varphi), \quad (4.241)$$

where

$$f_\alpha(\varphi) = f_\alpha\left(\varphi + \frac{2\pi}{Z}\right) \quad (4.242)$$

is an unknown non-universal  $\frac{2\pi}{Z}$ -periodic function and

$$F_\alpha(\varphi) = \sum_{\sigma=\pm} \sum_{i=1}^{M_\sigma^{(\alpha)}} \sigma \theta(\varphi - \varphi_{i\sigma}^{(\alpha)}) \quad (4.243)$$

is the part describing the discontinuous jumps from edge states entering/leaving the band. Here,

$$M^{(\alpha)} = M_-^{(\alpha)} - M_+^{(\alpha)} \quad (4.244)$$

is the net number of edge states leaving the band for a phase change by  $2\pi$  which is given by the Chern number  $C^{(\alpha)}$  of band  $\alpha$ , see (4.220). Using the form (4.241) we get  $\Delta Q_B^{(\alpha)}(\varphi) = M^{(\alpha)}/Z + \Delta F_\alpha(\varphi)$  which, together with  $\Delta Q_B^{(\alpha)}(\varphi) = I_\alpha(\varphi) + \frac{1}{Z}$  leads to

$$M^{(\alpha)} = 1 - s^{(\alpha)}Z, \quad (4.245)$$

where

$$s^{(\alpha)} = \Delta F_\alpha(\varphi) - I_\alpha(\varphi) \quad (4.246)$$

is a characteristic and phase independent integer for band  $\alpha$ . Eq. (4.245) just describes charge conservation. When the phase changes by  $2\pi$  the charge  $1 \bmod(Z)$  is pumped into the boundary which has to be taken away by a corresponding number  $M^{(\alpha)}$  of edge states. This equation is also called the Diophantine equation discussed within the integer QHE [[54]-[56]], here derived for a single band. To get (4.241), we have only used that the invariant  $I_\alpha$  is an integer. An additional topological constraint how the edge states can enter/leave a band follow from the allowed values  $I_\alpha \in \{0, \pm 1\}$ . For given  $s^{(\alpha)}$  we get

$$\Delta F_\alpha(\varphi) \in \{s^{(\alpha)}, s^{(\alpha)} \pm 1\}. \quad (4.247)$$

It is important to notice that  $M^{(\alpha)}$  and, consequently, also the integer  $s^{(\alpha)}$  depend crucially on the choice of the functions  $F_v$  and  $F_t$  defining the model parameters via Eqs. (4.10) and (4.11). As we have discussed in Section 6.1 for given parameters at phase  $\varphi = 0$  the phase-dependence can always be chosen such that, for a given band  $\alpha$ , one of the following two cases can be realized

$$M^{(\alpha)} = 1 \quad \Leftrightarrow \quad s^{(\alpha)} = 0, \quad (4.248)$$

$$M^{(\alpha)} = 1 - Z \quad \Leftrightarrow \quad s^{(\alpha)} = 1. \quad (4.249)$$

This are also the most frequently obtained values for rather smooth functions  $F_v(\varphi)$  and  $F_t(\varphi)$  where either of the two cases occurs for *all* bands. Additional multiples of  $Z$  are obtained for  $M^{(\alpha)}$  if the phase-dependence contains higher Fourier components which is rather exotic. Therefore, in the following we will only consider a phase-dependence where one of the two cases of Eqs. (4.248) and (4.249) occurs for each band.

We note that for a rational wave length  $\lambda = \frac{Z}{p}$  of the modulation where the phase parametrization is chosen according to Eqs. (4.12) and (4.13) we get analog results for the invariant but it has to be defined differently since a shift by one lattice site corresponds to a phase change  $\frac{2\pi p}{Z}$

$$I_\alpha(\varphi) = \Delta Q_B^{(\alpha)}(\varphi) - \frac{1}{Z} \in \{0, \pm 1\}, \quad (4.250)$$

$$\Delta Q_B^{(\alpha)}(\varphi) = Q_B^{(\alpha)}(\varphi + \frac{2\pi p}{Z}) - Q_B^{(\alpha)}(\varphi). \quad (4.251)$$

The form (4.241) of the phase-dependence of  $Q_B^{(\alpha)}(\varphi) = Q_B^{(\alpha)}(\varphi + 2\pi)$  does not change, in particular the function  $f_\alpha(\varphi)$  has precisely the same periodicity (4.242) with period  $\frac{2\pi}{Z}$ . The latter is proven as follows. First, shifting the lattice by one site via a phase change of  $\frac{2\pi p}{Z}$  requires the periodicity  $f_\alpha(\varphi) = f_\alpha(\varphi + \frac{2\pi p}{Z})$  due to (4.250) and the smoothness of  $f_\alpha(\varphi)$ . Furthermore the overall system does not change when  $\varphi$  changes by  $2\pi$  leading to the second condition  $f_\alpha(\varphi) = f_\alpha(\varphi + 2\pi)$ . Since  $p$  and  $Z$  are incommensurate one can always find integers  $n$  and  $m$  such that  $mp = 1 + nZ$  or  $\frac{2\pi}{Z} = 2\pi n + \frac{2\pi p}{Z}m$ . As a consequence, the function  $f_\alpha(\varphi)$  must have the period  $\frac{2\pi}{Z}$ . The only equation which changes is the Diophantine equation (4.245) which gets the different form

$$pM^{(\alpha)} = 1 - s^{(\alpha)}Z, \quad (4.252)$$

where  $p$  and  $s^{(\alpha)}$  must be such that  $M^{(\alpha)}$  is an integer. Again we note that this can always be solved if  $Z$  and  $p$  are incommensurate. The topological constraint formulated via Eqs. (4.246) and (4.247) remains the same.

In Fig. 40(a)-(c) we show the phase-dependence of the boundary charge  $Q_B^{(\alpha)}$ , the Friedel charge  $Q_F^{(\alpha)}$ , the polarization charge  $Q_P^{(\alpha)}$ , and the invariant  $I_\alpha$  for the bands  $\alpha = 1, 3, 4$  of Fig. 36(a). In Fig. 36(a) we find  $M_-^{(\alpha)} - M_+^{(\alpha)} = 1$  for  $\alpha \neq 4$  and  $M_-^{(4)} - M_+^{(4)} = 1 - Z$ , corresponding to the particle and hole picture, respectively. Therefore, except for  $\alpha = 4$ , the linear term in Eq. (4.241) leads to an average increase of  $Q_B(\alpha)$  by  $\frac{1}{Z}$  on the phase interval  $\frac{2\pi}{Z}$ . On top of this linear function edge states can enter or leave the band leading to discontinuous jumps of  $Q_B^{(\alpha)}$  by  $\pm 1$  and the function  $f_\alpha(\varphi)$  can lead to further oscillations with period  $\frac{2\pi}{Z}$ . As can be seen in Fig. 40(a) for  $\alpha = 1$  the invariant  $I_1$  takes only the values  $I_1 \in \{0, -1\}$  since no edge states can enter/leave at the band bottom. Up to a jump by  $-1$  when an edge state leaves the band  $Q_B^{(1)}$  is almost a linear function with average slope  $\frac{1}{2\pi}$ . However, this holds only for the sum of the Friedel and polarization charge, neither  $Q_F^{(1)}$  nor  $Q_P^{(1)}$  alone show any linear behaviour on average. In Fig. 40(b) we show the phase-dependence for band  $\alpha = 3$ . The average slope of  $Q_B^{(3)}$  is the same but the function  $f_3(\varphi)$  is rather large and influences the result strongly, leading even to a negative slope of  $Q_B^{(3)}$  between the jumps. The invariant can take all values  $I_3 \in \{0, \pm 1\}$  since edge states can enter/leave from both sides of the band. The same happens for band  $\alpha = 4$  shown in Fig. 40(c). However, in this case, the average slope of  $Q_B^{(4)}$  is negative corresponding to the hole picture. Here, all edge states enter the band leading to  $Z - 1$  discontinuous jumps by  $+1$  which have to be compensated by the large negative slope  $\frac{1-Z}{2\pi}$  of the linear term.

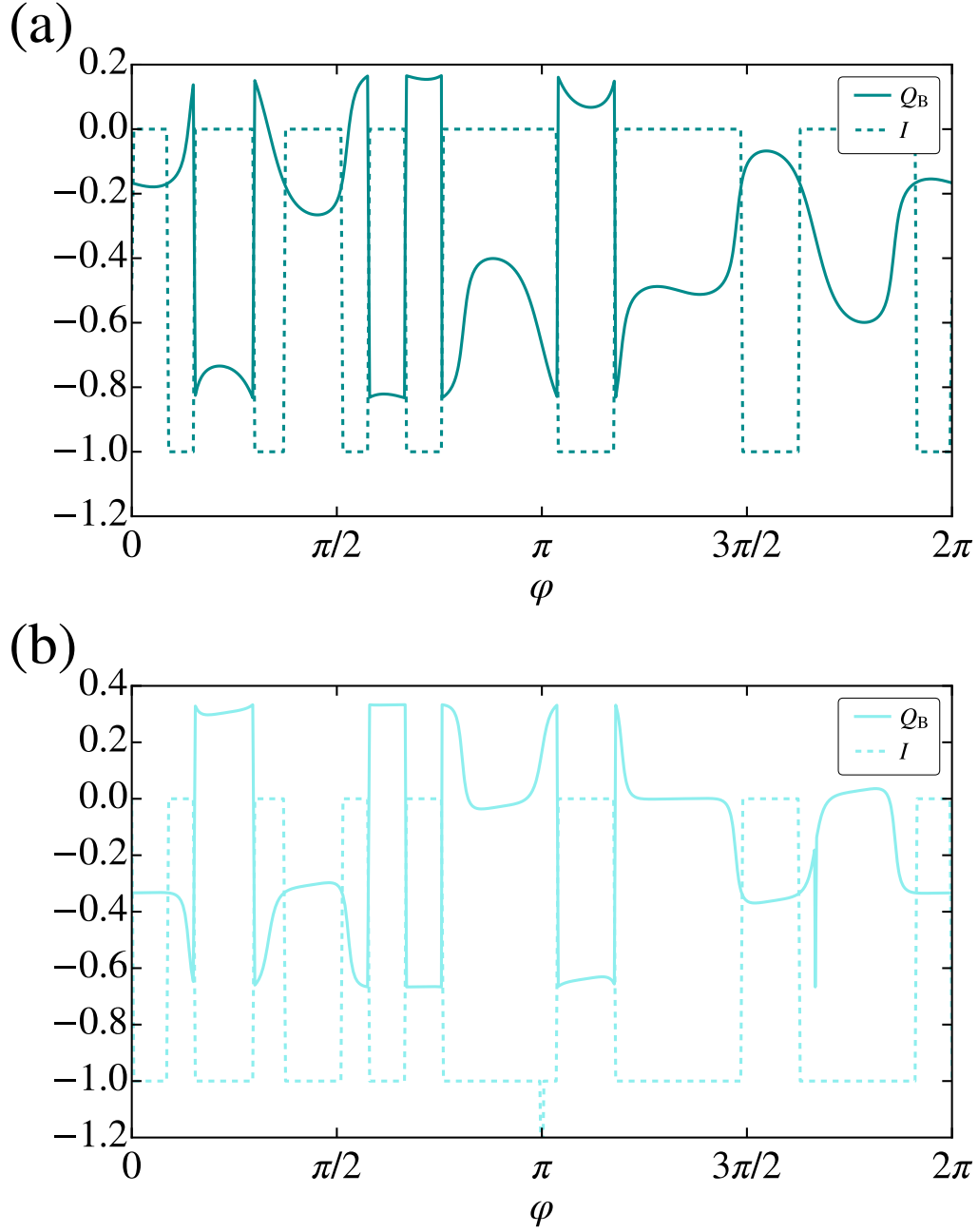


Figure 41: phase-dependence of boundary charge  $Q_B(\varphi, \mu_\nu)$  and invariant  $I(\varphi, \mu_\nu)$  for  $Z = 3$  and (a)  $\nu = 1$ , (b)  $\nu = 2$  for the parameters of Fig. 36(b).

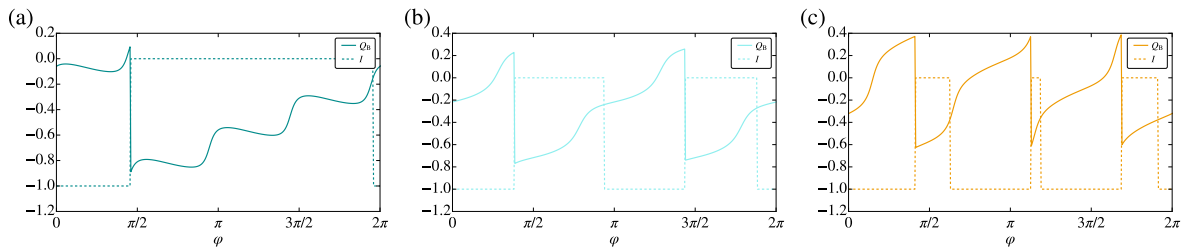


Figure 42: phase-dependence of boundary charge  $Q_B(\varphi, \mu_\nu)$  and invariant  $I(\varphi, \mu_\nu)$  for  $Z = 4$  and (a)  $\nu = 1$ , (b)  $\nu = 2$ , (c)  $\nu = 3$  for the parameters of Fig. 25(b).

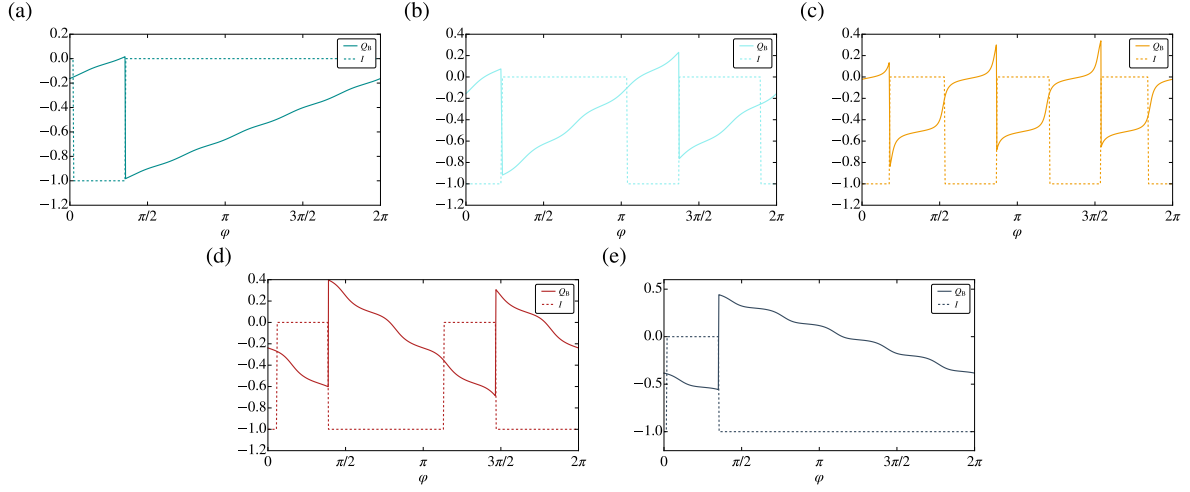


Figure 43: phase-dependence of boundary charge  $Q_B(\varphi, \mu_\nu)$  and invariant  $I(\varphi, \mu_\nu)$  for  $Z = 6$  and (a)  $\nu = 1$ , (b)  $\nu = 2$ , (c)  $\nu = 3$ , (d)  $\nu = 4$ , (e)  $\nu = 5$  for the parameters of Fig. 36(a).

### 6.3 Total invariant and boundary charge

We now consider the case where the chemical potential  $\mu_\nu$  is placed somewhere in gap  $\nu$  and calculate the total boundary charge  $Q_B(\varphi, \mu_\nu)$ . All our results are also valid if the chemical potential is chosen phase-dependent as long as it is  $\frac{2\pi}{Z}$ -periodic

$$\mu_\nu(\varphi) = \mu_\nu \left( \varphi + \frac{2\pi}{Z} \right). \quad (4.253)$$

Therefore, we can take for  $\mu_\nu$  also the top of band  $\alpha = \nu$  or the bottom of band  $\alpha = \nu + 1$  since  $\epsilon_k^{(\alpha)}(\varphi) = \epsilon_k^{(\alpha)}(\varphi + \frac{2\pi}{Z})$  is  $\frac{2\pi}{Z}$ -periodic in the phase. The boundary charge  $Q_B$  is given via (4.156) by summing over the boundary charges  $Q_B^{(\alpha)}$  of the occupied bands together with the charge of the occupied edge states. Therefore, we get for the change under a phase shift by  $\frac{2\pi}{Z}$

$$\begin{aligned} \Delta Q_B(\varphi, \mu_\nu) &= \sum_{\alpha=1}^{\nu} \Delta Q_B^{(\alpha)}(\varphi) + \Delta Q_E(\varphi, \mu_\nu) \\ &= \sum_{\alpha=1}^{\nu} I_\alpha(\varphi) + \frac{\nu}{Z} + \Delta Q_E(\varphi, \mu_\nu). \end{aligned} \quad (4.254)$$

Using (4.232) this can be expressed by a total invariant  $I(\varphi, \mu_\nu)$  via

$$I(\varphi, \mu_\nu) \equiv \Delta Q_B(\varphi, \mu_\nu) - \frac{\nu}{Z} \quad (4.255)$$

$$= \sum_{\alpha=1}^{\nu} I_\alpha(\varphi) + \Delta Q_E(\varphi, \mu_\nu) \quad (4.256)$$

$$= - \sum_{\alpha=1}^{\nu} w_\alpha(\varphi) + \Delta Q_E(\varphi, \mu_\nu). \quad (4.257)$$

As a result the total invariant is an integer or, with  $\bar{\rho} = \frac{\nu}{Z}$ , that  $\Delta Q_B - \bar{\rho}$  is an integer. To proof (4.223) we have to show that the invariant can only take two values

$$I(\varphi, \mu_\nu) \in \{0, -1\}. \quad (4.258)$$

To show this we first state the generic form of the phase-dependence of  $Q_B$ . As explained in detail in Section 3.3 the boundary charge can only jump if an edge state moves below or above

$\mu_\nu$ . We denote all these phase values by  $\varphi_{i\pm}(\mu_\nu)$ , with  $i = 1, \dots, M_\pm(\mu_\nu)$ , which depend on the choice of the chemical potential (for simplicity we do not indicate a possible phase-dependence of  $\mu_\nu$ ). However, we note that the difference

$$M_\nu = M_-(\mu_\nu) - M_+(\mu_\nu) \quad (4.259)$$

depends only on  $\nu$  since  $M_\nu$  describes the number of those edge states connecting the bands  $\alpha = \nu$  and  $\alpha = \nu + 1$  when the phase changes by  $2\pi$  counted positive (negative) when all these edge states move upwards (downwards). As shown in (4.221)  $M_\nu = C_\nu$  is identical to the sum over the Chern numbers of all filled bands  $\alpha = 1, \dots, \nu$ . Since  $\Delta Q_B(\varphi, \mu_\nu) - \frac{\nu}{Z}$  is an integer for all phases  $\varphi$ , the phase-dependence is similar to (4.241) of a single band

$$Q_B(\varphi, \mu_\nu) = f(\varphi, \mu_\nu) + \frac{M_\nu}{2\pi} \varphi + F(\varphi, \mu_\nu), \quad (4.260)$$

where

$$f(\varphi, \mu_\nu) = f\left(\varphi + \frac{2\pi}{Z}, \mu_\nu\right) \quad (4.261)$$

is an unknown non-universal  $\frac{2\pi}{Z}$ -periodic function and

$$F(\varphi, \mu_\nu) = \sum_{\sigma=\pm} \sum_{i=1}^{M_\sigma(\mu_\nu)} \sigma \theta[\varphi - \varphi_{i\sigma}(\mu_\nu)] \quad (4.262)$$

is the part describing the discontinuous jumps from edge states moving below/above  $\mu_\nu$ . This is precisely the definition we introduced in (4.111) where we used this quantity to define the topological charge  $\Delta F(\varphi, \mu_\nu)$ . Analog to (4.245) we get from (4.260) and (4.255) the Diophantine equation

$$M_\nu = \nu - s_\nu Z, \quad (4.263)$$

where

$$s_\nu = \Delta F(\varphi, \mu_\nu) - I(\varphi, \mu_\nu) \quad (4.264)$$

is a characteristic and phase independent integer for gap  $\nu$ . Similar to  $M_-^{(\alpha)} - M_+^{(\alpha)}$  and  $s^{(\alpha)}$  we note that  $M_\nu$  and  $s_\nu$  depend crucially on the choice of the functions  $F_v$  and  $F_t$  to define the parameters of the model via Eqs. (4.10) and (4.11). From (4.264) we conclude that the property for the invariant to take only the values  $I \in \{0, -1\}$  is equivalent to the following topological constraint of how the edge states can move below/above  $\mu_\nu$

$$\Delta F(\varphi, \mu_\nu) \in \{s_\nu, s_\nu - 1\}, \quad (4.265)$$

such that the integer  $s_\nu$  is related via (4.263) to  $M_\nu$ . This is precisely the topological constraint which we have proven in Section 2.6 via the explicit conditions how edge states can appear in the gaps, see (4.114), from which we have also obtained the Diophantine equation (4.117). Here, we have obtained the Diophantine equation (4.263) in a different way from the quantization of the invariant but, in addition, we have proven that the parameter  $\nu$  is indeed the index of the gap under consideration. We conclude that the approach described in Section 6.1 to derive the topological constraint, following along the lines of Ref. [[52]], is indeed correct and can be rigorously proven. This means that charge conservation and particle-hole duality together with the fact that the phase-dependence of the parameters can always be chosen such that no edge states cross  $\mu$  in a certain gap  $\nu$  within some phase interval of size  $\frac{2\pi}{Z}$  describe precisely the right



physical picture. This is a very surprising and remarkable fact since the two ways to derive the same result are based on a complete different approach. Whereas in Section 2.6 the proof is based on the precise way the edge states can run between the bands, in Ref. [[52]] the proof is based on the *absence* of edge states crossing  $\mu$ . This shows very clearly that it is *not* the physics of edge states which drives the topological constraint but rather the edge modes are followers which have to adjust to a certain choice of the phase-dependence (or the way the boundary is shifted continuously through the lattice) in order to respect the topological constraint.

Similar to (4.248) we note that for rather smooth functions  $F_v(\varphi)$  and  $F_t(\varphi)$  the integer  $s_\nu$  is typically given by one of the following cases for each gap  $\nu$

$$M_\nu = \nu \quad \Leftrightarrow \quad s_\nu = 0, \quad (4.266)$$

$$M_\nu = \nu - Z \quad \Leftrightarrow \quad s_\nu = 1. \quad (4.267)$$

$M_\nu = \nu$  means that  $\nu$  edge states move upwards in gap  $\nu$ , corresponding to the particle picture where the charge  $\bar{\rho}$  is moved into the boundary under a phase change of  $2\pi$  which has to be compensated by  $\nu$  edge states moving above  $\mu_\nu$ .  $M_\nu = \nu - Z$  is the case where  $Z - \nu$  edge states move downwards, describing the case where the hole charge  $\bar{\rho} - 1$  is moved into the boundary which has to be compensated by  $Z - \nu$  edge states moving below  $\mu_\nu$ .

Analog to (4.251) and (4.252) we note that for rational wave lengths with phase parametrization according to Eqs. (4.12) and (4.13) all equations remain the same only the invariant is redefined as

$$I(\varphi, \mu_\nu) = \Delta Q_B(\varphi, \mu_\nu) - \frac{\nu}{Z} \in \{0, -1\}, \quad (4.268)$$

$$\Delta Q_B(\varphi, \mu_\nu) = Q_B(\varphi + \frac{2\pi p}{Z}, \mu_\nu) - Q_B(\varphi, \mu_\nu), \quad (4.269)$$

and the Diophantine equation changes to

$$pM_\nu = \nu - s_\nu Z. \quad (4.270)$$

In Figs. 41(a,b), Figs. 42(a-c) and Figs. 43(a-e) we show the phase-dependence of the boundary charge and the invariant for  $Z = 3, 4, 6$  corresponding to the band structures of Fig. 36(b), Fig. 25(b) and Fig. 36(a), respectively. For each case we consider all possibilities for the gap  $\nu = 1, \dots, Z - 1$  and take the chemical potential  $\mu_\nu$  at the top of band  $\alpha = \nu$  (if  $\mu_\nu$  is placed somewhere in the gap only the position of the jumps from edge states are shifted). The three terms on the r.h.s. of Eq. (4.260) can be clearly identified. The part  $F(\varphi, \mu_\nu)$  shows up in the discontinuous jumps of  $Q_B(\varphi, \mu_\nu)$  by  $\pm 1$  at the phase values  $\varphi = \varphi_{i\pm}$  where edge states enter or leave the top of band  $\alpha = \nu$ . Correspondingly, the invariant  $I(\varphi, \mu_\nu)$  jumps at  $\varphi = \varphi_{i\pm}$  by  $\mp 1$  and at  $\varphi = \varphi_{i\pm} - \frac{2\pi}{Z}$  by  $\pm 1$ . The invariant is always given by the two values  $I \in \{0, -1\}$ , even for the rather chaotic case  $Z = 3$  of Figs. 41(a,b), where many edge states return to the same band. The linear term  $\frac{M_\nu}{2\pi}\varphi$  on the r.h.s. of (4.260) is visible by an overall positive or negative average slope of  $Q_B$  in Figs. 42(a-c) for  $Z = 4$  and in Figs. 43(a-e) for  $Z = 6$ . A positive slope occurs on average for  $M_\nu = \nu = 1, 2, 3$  in Figs. 42(a-c) and Figs. 43(a-c) whereas a negative slope is observed on average for  $M_\nu = \nu - Z$  and  $\nu = 4, 5$  in Figs. 43(d,e). The non-universal  $\frac{2\pi}{Z}$ -periodic function  $f(\varphi, \mu_\nu)$  turns out to be large for small  $Z$  or for gaps  $\nu$  close to  $Z/2$ . When  $Z$  is large the boundary charge is almost a linear function for the gaps  $\nu$  close to 1 or  $Z$ , see  $\nu = 1, 5$  in Figs. 43(a,e).

## 7. Summary and outlook

In this work we have presented a rigorous basis for a unique relationship between the boundary charge and the bulk Zak-Berry phase together with the analytical understanding of universal

properties of the boundary charge for a wide class of half-infinite nearest-neighbor tight-binding models with one channel per site in 1D beyond symmetry constraints. The proposed representation of the exact eigenstates and their analytic continuation, underlying essentially our analytical treatment, might be of interest for a wider class of models including multi-channel cases. We addressed the very fundamental issue of the topological constraints for the edge states when the boundary of a half-infinite system is shifted to a different position and the relation to the universal change of the boundary charge. We introduced a measurable topological invariant to characterize this dependence and revealed a link to the winding number of a fundamental phase, namely the phase difference of the Bloch wave for the infinite system between the sites left and right to the boundary. We analysed this winding number in comparison to other topological indices classifying topological systems and found that the winding number contains more information and probes this phase directly. Another important insight of this work is the proof that the derived universal properties can indeed be described by using charge conservation of particles and holes alone, as it was proposed in Ref. [[52]]. The edge states were shown to play the role of followers obeying certain topological constraints such that particles and holes fulfil charge conservation at the same time. This reflects a very simple physical picture analog to charge pumping that piling up particle or hole charge at the boundary by shifting the lattice towards a boundary leads inevitably to a linear increase of the boundary charge until edge states entering or leaving the band guarantee that the charge does not change when a whole unit cell has been shifted into the boundary. As a consequence, edge states are driven by this mechanism, leaving a fingerprint in the density of the scattering states as we have demonstrated, in agreement with previous works on continuum systems, by calculating the pole contributions of the Friedel density. Besides this “edge” part of the density of the scattering states we have also analysed the part from branch cuts leading to a different localization length and a nontrivial pre-exponential power-law that deserves further investigations, in particular in the presence of interactions.

So far our results refer to all single-channel models falling into the wide class of commensurate Aubry-André-Harper models but with generic modulation functions for the potentials and hoppings. However, since our proposed representation and analytic continuation of Bloch states is quite general and since the principle of charge conservation of particle and holes is always fulfilled, we expect that our results can be generalized to multi-channel systems as well [67]. As we already outlined in this work, one expects a weakening of the topological constraint in the sense that, if  $N_c$  channels are present, the topological invariant can take  $N_c + 1$  different values. An interesting question for future research will be the development of a non-Abelian version of the winding number and the determination of the precise gauge of the Bloch states such that an unambiguous link can be set up between the boundary charge and the Zak-Berry phase.

Although the scattering states of the half-infinite system consist of a linear combination of an incoming and outgoing plane wave for the single-channel model under consideration here, we have found that the scattering states have a nontrivial influence on the density and the boundary charge. This is even expected to be more dramatic for multi-channel systems since, in this case, the scattering states have to fulfil more boundary conditions which can only be fulfilled if they contain, in addition to the purely oscillating waves, also exponentially decaying contributions. This has been shown in a recent article on STM setups for probing the spectral density at the boundary of a Floquet topological insulator [68]. There it was shown, that the exponentially decaying contributions of the scattering states lead to a dramatic effect for the STM signal at bifurcation points which are even more pronounced and stable than the ones from topological states in the gap. It is quite intuitive that the exponential localization of the boundary charge will certainly hold always for an insulator (since a typical velocity divided by the inverse gap is the only relevant length scale besides the localization length of the edge

states) but it is not obvious how the pre-exponential function will look like in the presence of several channels.

Since the principle of charge conservation of particle and holes is not violated by interactions we expect that our results are quite robust against disorder and Coulomb interactions. The stability against disorder has already been demonstrated in Ref. [[52]] provided that disorder is so weak that the gaps are not closed. Weak Coulomb interaction can be treated very effectively by functional renormalization group methods [69] or bosonization techniques [39, 71]. An interesting issue concerns the properties of the boundary charge for the case of strong Coulomb interaction, where bosonization methods have suggested the generation of charge and spin density wave instabilities, possibly relevant for the occurrence of fractional charges as they appear, e.g., for the fractional quantum Hall effect [72].

An issue touched only slightly in this work is a precise discussion of the unknown and non-universal function  $f(\varphi, \mu_\nu)$  in (4.260). For rather smooth choices of the phase-dependence of the model parameters this function is observed to be rather small for not too small  $Z$  and for sufficiently large gaps located at rather low or high energies. This behaviour is expected to drastically change when the system is close to special symmetry points, where the boundary charge is quantized for all phases at particular filling factors, leading to a completely flat curve for  $Q_B(\varphi)$ , up to discrete jumps from edge states. This happens typically when the gap closes at particular values of the phase corresponding to Weyl semimetal physics [73]. Breaking the symmetry slightly leads only to a small deviation from a flat curve although the Chern number  $C_\nu = M_\nu$  determining the slope of the linear term in (4.260) might be nonzero. In this case the function  $f(\varphi, \mu_\nu)$  is very strong and plays a very important role. As a result this function seems to have two tendencies driven by two different physical mechanism. When it is small the boundary charge adjusts to charge conservation of particles and holes for all values of the phase leading to a nearly perfect linear form of  $Q_B(\varphi)$  between the jumps. On the other hand, it can be driven by symmetry constraints, which have the tendency to adjust the phase-dependence to certain quantized values of the boundary charge. Which mechanism wins depends on how strong certain symmetries are broken as will be discussed in a future work [73].

## 8. Appendixes

### 8.1 Complex hoppings

Here we show that the phases of the hoppings of any nearest-neighbor tight-binding model with one orbital per site defined on a half-infinite system can be gauged away by a unitary transformation. Starting from the generic Hamiltonian

$$H = \sum_{m=1}^{\infty} \{ \hat{v}_m |m\rangle \langle m| - (\hat{t}_m e^{i\theta_m} |m+1\rangle \langle m| + \text{h.c.}) \}, \quad (4.271)$$

where  $\hat{v}_m = \hat{v}_m^*$ ,  $\hat{t}_m = \hat{t}_m^*$ , and  $\theta_m = \theta_m^*$  are real, we define a unitary transformation  $U|m\rangle = e^{i\phi_m} |m\rangle$  via the phases  $\phi_1 = 0$ ,  $\phi_2 = \theta_1$ ,  $\phi_3 = \theta_2 + \phi_2$ , etc., such that  $\theta_m = \phi_{m+1} - \phi_m$ . It follows that

$$U^\dagger H U = \sum_{m=1}^{\infty} \{ \hat{v}_m |m\rangle \langle m| - (\hat{t}_m |m+1\rangle \langle m| + \text{h.c.}) \} \quad (4.272)$$

contains only real hoppings. The unitary transformation gives each site only a phase factor, i.e., the density on each site remains invariant. Furthermore, we find that the hoppings can be chosen positive  $\hat{t}_m > 0$ . The case  $\hat{t}_m = 0$  is excluded since it would correspond to a finite system.

## 8.2 Choice of phase-dependence

Here, we present two different ways how we parametrize the two real and periodic functions  $F_\gamma(\varphi) = F_\gamma(\varphi + 2\pi)$ , with  $\gamma = v, t$ , in Eqs. (4.10) and (4.11) in case we take a random choice in the figures.

The first choice is to take a random periodic and real function of the form

$$F_\gamma(\varphi) = 2 \sum_{n=1}^{N_\gamma} r_\gamma^n \cos(n\varphi + \theta_\gamma^n), \quad (4.273)$$

where  $0 < r_\gamma^n < 1$ ,  $\theta_\gamma^n$  are  $2N_\gamma$  random and real parameters. Due to (4.8) there is no Fourier component with  $n = 0$ .

The second choice consists in fixing the values of  $\gamma_j(0) \equiv v_j(0), t_j(0)$  for all  $j = 1, \dots, Z$  at phase  $\varphi = 0$  via

$$v_j(0) = V v_j^{(0)} \quad , \quad t_j(0) = t + \delta t t_j^{(0)} \quad (4.274)$$

by the real parameters  $\gamma_j^{(0)} = F_\gamma(2\pi j/Z)$  (for  $\gamma = v, t$ ), with zero average

$$\frac{1}{Z} \sum_{j=1}^Z \gamma_j^{(0)} = 0 \quad (4.275)$$

due to (4.8). The phase-dependence is then chosen in a random way via the function

$$F_\gamma(\varphi) = \text{Re} \left\{ \sum_{n=1}^Z F_\gamma^n e^{in\varphi} \right\} + G_\gamma(\varphi), \quad (4.276)$$

where

$$F_\gamma^n = \frac{1}{Z} \sum_{j=1}^Z \gamma_j^{(0)} e^{-in2\pi j/Z}, \quad (4.277)$$

is the discrete Fourier transform of  $\gamma_j^{(0)}$ , with  $n = 1, \dots, Z$ , and  $G_\gamma(\varphi)$  is some random periodic function with  $G_\gamma(2\pi j/Z) = 0$ . Therefore, by construction the condition  $\gamma_j^{(0)} = F_\gamma(2\pi j/Z)$  is fulfilled. The second term involving the function  $G_\gamma$  is introduced since the first term is zero for  $\varphi = \frac{\pi}{Z}(2j+1)$

$$F_\gamma \left( \frac{\pi}{Z}(2j+1) \right) = G_\gamma \left( \frac{\pi}{Z}(2j+1) \right). \quad (4.278)$$

This follows by inserting (4.277) in (4.276) and performing the sum over  $n$  with the result

$$F_\gamma(\varphi) - G_\gamma(\varphi) = \frac{1}{Z} \sum_{j'=1}^Z \gamma_{j'}^{(0)} \sin\left(\frac{\varphi Z}{2}\right) \left\{ -\sin\left(\frac{\varphi Z}{2}\right) + \cos\left(\frac{\varphi Z}{2}\right) \cot\left(\frac{\varphi}{2} - \frac{\pi j'}{Z}\right) \right\}. \quad (4.279)$$

Inserting  $\varphi = \frac{\pi}{Z}(2j+1)$  and using (4.275) we find (4.278). For  $G_\gamma = 0$  this means that all potentials and hopping are the same at these phase values

$$\begin{aligned} G_\gamma = 0 & \Rightarrow \\ v_j \left( \frac{\pi}{Z}(2j+1) \right) &= 0 \quad , \quad t_j \left( \frac{\pi}{Z}(2j+1) \right) = t, \end{aligned} \quad (4.280)$$

leading to the special case of gap closings. Therefore, to cover the generic case the second term  $G_\gamma(\varphi)$  is chosen randomly via

$$G_\gamma(\varphi) = \text{Im} \sum_{n=1}^Z G_\gamma^n e^{in\varphi}, \quad (4.281)$$

where

$$G_\gamma^n = \frac{1}{Z} \sum_{j=1}^Z \gamma_j^{(1)} e^{-in2\pi j/Z} \quad (4.282)$$

is defined analog to (4.277) but with some different real and random numbers for  $\gamma_j^{(1)}$ . Explicitly one obtains analog to (4.279)

$$G_\gamma(\varphi) = \frac{1}{Z} \sum_{j'=1}^Z \gamma_{j'}^{(1)} \sin\left(\frac{\varphi Z}{2}\right) \cdot \left\{ \cos\left(\frac{\varphi Z}{2}\right) - \sin\left(\frac{\varphi Z}{2}\right) \cot\left(\frac{\varphi}{2} - \frac{\pi j'}{Z}\right) \right\}. \quad (4.283)$$

### 8.3 Useful identities

We first proof (4.59) which in terms of the components  $\chi_k^{(\alpha)}(j)$  of the Bloch state reads for  $j = 2, \dots, Z-1$

$$t_1 \cdots t_{j-1} \chi_k^{(\alpha)}(j) = d_{1,j-1} \chi_k^{(\alpha)}(1) - d_{2,j-1} \chi_k^{(\alpha)}(Z) t_Z e^{-ik}, \quad (4.284)$$

where we omitted for simplicity the dependence of the determinants on  $\epsilon_k^{(\alpha)}$ . First we note that the eigenvalue equation (4.32) together with the form (4.21) of the Bloch Hamiltonian implies the recurrence relation

$$-t_j \chi_k^{(\alpha)}(j+1) + \bar{v}_j \chi_k^{(\alpha)}(j) - t_{j-1} \chi_k^{(\alpha)}(j-1) = 0, \quad (4.285)$$

for  $j = 2, \dots, Z-1$ , together with

$$-t_1 \chi_k^{(\alpha)}(2) + \bar{v}_1 \chi_k^{(\alpha)}(1) - t_Z e^{-ik} \chi_k^{(\alpha)}(Z) = 0, \quad (4.286)$$

where  $\bar{v}_j = v_j - \epsilon_k^{(\alpha)}$  has been defined in (4.44). (4.286) gives (4.284) for  $j = 2$ . The other values are obtained by induction. Assuming that (4.284) holds for all  $j = 2, \dots, l$ , we find for  $j = l+1$

$$\begin{aligned} t_1 \cdots t_l \chi_k^{(\alpha)}(l+1) &= \\ &= t_1 \cdots t_{l-1} [\bar{v}_l \chi_k^{(\alpha)}(l) - t_{l-1} \chi_k^{(\alpha)}(l-1)] \\ &= \bar{v}_l [d_{1,l-1} \chi_k^{(\alpha)}(1) - d_{2,l-1} \chi_k^{(\alpha)}(Z) t_Z e^{-ik}] - \\ &\quad - t_{l-1}^2 [d_{1,l-2} \chi_k^{(\alpha)}(1) - d_{2,l-2} \chi_k^{(\alpha)}(Z) t_Z e^{-ik}] \\ &= d_{1l} \chi_k^{(\alpha)}(1) - d_{2l} \chi_k^{(\alpha)}(Z) t_Z e^{-ik}, \end{aligned} \quad (4.287)$$

where we used (4.285) in the first step, (4.284) in the second step, and the recurrence relation (4.55) for the determinants in the last step. This is identical to (4.284) for  $j = l+1$ .

Next we proof

$$(t_1 \cdots t_{j-1})^2 d_{j+1,Z-1} = d_{1,j-1} d_{2,Z-1} - d_{2,j-1} d_{1,Z-1}, \quad (4.288)$$

with  $2 \leq j \leq Z-1$ . This relation follows directly from inserting (4.50) together with (4.51) and (4.52) in (4.59).

To proof the relation

$$\frac{d}{d\epsilon} d_{ij}(\epsilon) = - \sum_{j'=i}^j d_{i,j'-1}(\epsilon) d_{j'+1,j}(\epsilon), \quad (4.289)$$

for  $i \leq j$ , we use the property

$$\frac{d}{d\epsilon} \det R(\epsilon) = \text{Tr} \frac{\det R(\epsilon)}{R(\epsilon)} \frac{d}{d\epsilon} R(\epsilon), \quad (4.290)$$

valid for any matrix  $R(\epsilon)$ . Taking  $R(\epsilon) = A^{(ij)} - \epsilon$  with  $A^{(ij)}$  defined in (4.49), we get with (4.48)

$$\begin{aligned} \frac{d}{d\epsilon} d_{ij}(\epsilon) &= -\text{Tr} \frac{\det(A^{(ij)} - \epsilon)}{A^{(ij)} - \epsilon} \\ &= - \sum_{j'=i}^j \left( \frac{\det(A^{(ij)} - \epsilon)}{A^{(ij)} - \epsilon} \right)_{j'j'}, \end{aligned} \quad (4.291)$$

where we labelled the matrix elements of the  $j-i+1$ -dimensional matrix  $A^{(ij)}$  with the index  $j' = i, \dots, j$ . Using in analogy to (4.47) (with  $1 \rightarrow i$  and  $Z-1 \rightarrow j$ ) that

$$\left( \frac{\det(A^{(ij)} - \epsilon)}{A^{(ij)} - \epsilon} \right)_{j'j'} = d_{i,j'-1}(\epsilon) d_{j'+1,j}(\epsilon), \quad (4.292)$$

we arrive at (4.289).

Using (4.289) we now proof the relation (4.58) for the vectors  $f(\epsilon)$  and  $g(\epsilon)$ . Using the definitions (4.51) and (4.52) we get

$$\begin{aligned} f^T \frac{d}{d\epsilon} g &= \bar{t}^Z t_Z^2 \sum_{j=1}^{Z-1} d_{j+1,Z-1} \frac{d}{d\epsilon} d_{1,j-1} \\ &= \bar{t}^Z t_Z^2 \sum_{j=1}^{Z-1} d_{j+1,Z-1} \sum_{k=1}^{j-1} d_{1,k-1} d_{k+1,j-1} \\ &= \bar{t}^Z t_Z^2 \sum_{1 \leq k < j \leq Z-1} d_{1,k-1} d_{k+1,j-1} d_{j+1,Z-1} \end{aligned} \quad (4.293)$$

$$\begin{aligned} g^T \frac{d}{d\epsilon} f &= \bar{t}^Z t_Z^2 \sum_{j=1}^{Z-1} d_{1,j-1} \frac{d}{d\epsilon} d_{j+1,Z-1} \\ &= \bar{t}^Z t_Z^2 \sum_{j=1}^{Z-1} d_{1,j-1} \sum_{k=j+1}^{Z-1} d_{j+1,k-1} d_{k+1,Z-1} \\ &= \bar{t}^Z t_Z^2 \sum_{1 \leq j < k \leq Z-1} d_{1,j-1} d_{j+1,k-1} d_{k+1,Z-1}. \end{aligned} \quad (4.294)$$

Interchanging  $j \leftrightarrow k$  in (4.294) we find that it agrees with (4.293), which proves (4.58).

We now derive useful relations for the derivative of the  $B$ -matrix defined by (4.45) and (4.46). Taking successively the derivatives  $\frac{d^n}{d\epsilon^n}$  of

$$(A - \epsilon)B(\epsilon) = B(\epsilon)(A - \epsilon) = s(\epsilon), \quad (4.295)$$

we find for  $n = 1, 2, \dots$  with  $B^{(n)} \equiv B^{(n)}(\epsilon) \equiv \frac{d^n}{d\epsilon^n} B(\epsilon)$  and  $s^{(n)} \equiv s^{(n)}(\epsilon) \equiv \frac{d^n}{d\epsilon^n} s(\epsilon)$

$$\begin{aligned} -nB^{(n-1)} + (A - \epsilon)B^{(n)} &= \\ &= -nB^{(n-1)} + B^{(n)}(A - \epsilon) = s^{(n)}. \end{aligned} \quad (4.296)$$

Multiplying this equation from the left or right with  $B$  and using (4.295) we find in addition

$$\begin{aligned} -nBB^{(n-1)} + sB^{(n)} &= \\ &= -nB^{(n-1)}B + sB^{(n)} = s^{(n)}B. \end{aligned} \quad (4.297)$$

These relations imply the following useful identities

$$(A - \epsilon)B^{(n)} = B^{(n)}(A - \epsilon), \quad (4.298)$$

$$BB^{(n-1)} = B^{(n-1)}B, \quad (4.299)$$

$$-nB^{(n-1)} + (A - \epsilon)B^{(n)} = s^{(n)}, \quad (4.300)$$

$$-nBB^{(n-1)} + sB^{(n)} = s^{(n)}B. \quad (4.301)$$

Next we try to set up useful identities of the function  $D(\epsilon_k) = \cos(k)$  defined in (4.62) and (4.63), together with its derivatives. Using the form (4.25) of the vector  $b_k$  we find

$$\begin{aligned} b_k^T B b_{-k} &= t_Z^2 B_{11} + t_{Z-1}^2 B_{Z-1, Z-1} + 2t_Z t_{Z-1} D B_{1, Z-1} \\ &= t_Z^2 B_{11} + t_{Z-1}^2 B_{Z-1, Z-1} + 2\bar{t}^Z D, \end{aligned} \quad (4.302)$$

where we used (4.47) to get  $B_{1, Z-1} = t_1 \cdots t_{Z-1}$  in the last step. Taking the derivatives of this equation w.r.t.  $\epsilon_k$  we get with  $B_{1, Z-1}^{(n)} = 0$  for  $n = 1, 2, \dots$

$$\frac{d^n}{d\epsilon_k^n} \{b_k^T B b_{-k}\} = b_k^T B^{(n)} b_{-k} + 2\bar{t}^Z D^{(n)}. \quad (4.303)$$

From (4.43) and (4.45) we get

$$s(v_Z - \epsilon_k) = b_k^T B b_{-k}, \quad (4.304)$$

Using (4.303) and taking derivatives of this equation w.r.t.  $\epsilon$  we get for  $n = 1, 2, \dots$

$$\begin{aligned} -2s^{(n-1)} + s^{(n)}(v_Z - \epsilon_k) &= \\ &= b_k^T B^{(n)} b_{-k} + 2\bar{t}^Z D^{(n)}. \end{aligned} \quad (4.305)$$

#### 8.4 Expansion of $N_k$ around poles

To prove Eqs. (4.90) and (4.91) we expand  $N_k$  around the quasimomentum  $k = k_e \equiv k_e^{(\nu)}$  of the edge pole up to linear order where, according to Section 2.5, we have the conditions

$$s(\epsilon_{k_e}) = 0 \quad , \quad a_{-k_e} = 0, \quad (4.306)$$

and

$$(A - \epsilon_{k_e})a_{k_e} = 0 \quad , \quad b_{-k_e}^T a_{k_e} = 0. \quad (4.307)$$

In the following we use the subindex e to indicate that quantities are evaluated at  $\epsilon_{k_e}$ , e.g.,  $B_e \equiv B(\epsilon_{k_e})$ .



We first assume that this is an isolated pole and does not agree with the branching point  $k_e^{(\nu)} \neq k_{\text{bp}}^{(\nu)}$ . This means that  $\frac{d\epsilon_k}{dk}|_{k=k_e}$  is finite, i.e.,  $\epsilon_k - \epsilon_{k_e} \sim O(k - k_e)$ . Therefore, the term  $[s(\epsilon_k)]^2$  in  $N_k$  is of  $O(k - k_e)^2$  and can be neglected. We get

$$\begin{aligned} N_k &= a_k^T a_{-k} + [s(\epsilon_k)]^2 \\ &\approx a_{k_e}^T \frac{d}{dk} (a_{-k})|_{k=k_e} (k - k_e) \\ &= -a_{k_e}^T \frac{d}{dk} (B(\epsilon_k) b_{-k})|_{k=k_e} (k - k_e) \\ &= -\frac{d\epsilon_k}{dk}|_{k=k_e} a_{k_e}^T B_e^{(1)} b_{-k_e} (k - k_e) - a_{k_e}^T B_e \frac{d}{dk} (b_{-k})|_{k=k_e} (k - k_e). \end{aligned} \quad (4.308)$$

The first term of the last equation is zero since, due to the Faddeev-LeVerrier algorithm, the matrix  $B(\epsilon)$  can be written as a polynomial of degree  $Z - 2$  in the matrix  $A - \epsilon$ . This holds also for its derivatives since  $\frac{d}{d\epsilon}(A - \epsilon) = -1$ . Therefore, by using (4.307), we get no contribution of this term. For the second term we use

$$\frac{d}{dk} (b_{-k}) = \frac{i}{e^{-2ik} - 1} (b_k - b_{-k}), \quad (4.309)$$

and get with  $B_e b_{k_e} = -a_{k_e}$  and  $B_e b_{-k_e} = -a_{-k_e} = 0$  the result

$$N_k = \frac{i}{e^{-2ik_e} - 1} a_{k_e}^T a_{k_e} (k - k_e) + O(k - k_e)^2, \quad (4.310)$$

which proves Eq. (4.90).

Next we consider the case when an edge pole agrees with a branching point  $k_e \equiv k_e^{(\nu)} = k_{\text{bp}}^{(\nu)}$ . In this case we have in addition to (4.306) the condition

$$D_e^{(1)} \equiv \frac{dD}{d\epsilon}(\epsilon_{k_e}) = 0. \quad (4.311)$$

Therefore, we have to be more careful since expanding the equation  $D(\epsilon_k) = \cos(k)$  around  $k = k_e$  we get

$$\frac{1}{2} D_e^{(2)} (\epsilon_k - \epsilon_{k_e})^2 \approx -\sin(k_e) (k - k_e), \quad (4.312)$$

i.e., we have to expand all terms up to  $(\epsilon_k - \epsilon_{k_e})^2 \sim (k - k_e)$ . We get

$$\begin{aligned} N_k &= a_k^T a_{-k} + s(\epsilon_k)^2 \\ &= -a_k^T B(\epsilon_k) b_{-k} + s(\epsilon_k)^2 \\ &\approx -a_k^T B(\epsilon_k) b_{-k_e} + (s_e^{(1)})^2 (\epsilon_k - \epsilon_{k_e})^2 \\ &\quad - a_{k_e}^T B_e \frac{d}{dk} (b_{-k})|_{k=k_e} (k - k_e). \end{aligned} \quad (4.313)$$

The last term has already been evaluated above and leads to (4.310). For the first term we use

$$\begin{aligned} -a_k^T B(\epsilon_k) b_{-k_e} &= b_k^T B(\epsilon_k) B(\epsilon_k) b_{-k_e} \\ &= (-a_{k_e}^T + b_{k_e}^T B_e^{(1)} (\epsilon_k - \epsilon_{k_e}) + O(\epsilon_k - \epsilon_{k_e})^2) B(\epsilon_k) b_{-k_e}, \end{aligned}$$

where we used  $k - k_e \sim (\epsilon_k - \epsilon_{k_e})^2$  in the last step, see (4.312). The first term in the bracket gives zero analog to the first term on the r.h.s. of Eq. (4.308) since  $B(\epsilon_k)$  can be written

as a power series in  $A - \epsilon_k$  or in  $A - \epsilon_{k_e}$ . In addition, since  $B_{k_e} b_{-k_e} = -a_{-k_e} = 0$  we get  $B(\epsilon_k) b_{-k_e} \sim (\epsilon_k - \epsilon_{k_e})$  and obtain

$$-a_k^T B(\epsilon_k) b_{-k_e} \approx b_{k_e}^T (B_e^{(1)})^2 b_{-k_e} (\epsilon - \epsilon_{k_e})^2. \quad (4.314)$$

Inserting this result in (4.313) we get

$$\begin{aligned} N_k &= (b_{k_e}^T (B_e^{(1)})^2 b_{-k_e} + (s_e^{(1)})^2) (\epsilon_k - \epsilon_{k_e})^2 + \\ &+ \frac{i}{e^{-2ik_e} - 1} a_{k_e}^T a_{k_e} (k - k_e) + O(\epsilon_k - \epsilon_{k_e})^3. \end{aligned} \quad (4.315)$$

Finally, we prove that the first term on the r.h.s. is identical to the second one. We rewrite the second term by inserting

$$\begin{aligned} a_{k_e} &= -B_e b_{k_e} \\ &= a_{-k_e} - B_e (b_{k_e} - b_{-k_e}) \\ &= -2it_Z \sin(k_e) B_e e_1, \end{aligned}$$

where we used (4.306) and (4.25) in the last equality, and defined the  $Z - 1$ -dimensional unit vector  $e_1 = (10 \cdots 0)^T$ . This gives  $a_{k_e}^T a_{k_e} = -4t_Z^2 \sin^2(k_e) ((B_e)^2)_{11}$  and the second term on the r.h.s. of Eq. (4.315) can be written with the help of (4.312) as

$$\begin{aligned} \frac{i}{e^{-2ik_e} - 1} a_{k_e}^T a_{k_e} (k - k_e) &= \\ &= -t_Z^2 e^{ik_e} ((B_e)^2)_{11} D_e^{(2)} (\epsilon_k - \epsilon_{k_e})^2. \end{aligned} \quad (4.316)$$

The matrix element  $((B_e)^2)_{11}$  can be rewritten with the help of (4.301) for  $n = 1$  and  $k = k_e$ , and using (4.47), (4.56) and (4.99)

$$\begin{aligned} ((B_e)^2)_{11} &= -s_e^{(1)} (B_e)_{11} = -s_e^{(1)} d_{2,Z-1}(\epsilon_{k_e}) \\ &= -s_e^{(1)} \frac{\bar{t}^Z}{t_Z^2} \tilde{d}_{2,Z-1}(\epsilon_{k_e}) = s_e^{(1)} \frac{\bar{t}^Z}{t_Z^2} e^{-ik_e}. \end{aligned} \quad (4.317)$$

With the help of (4.305) for  $n = 1, 2$  and  $k = k_e$  we get

$$s_e^{(1)} (v_Z - \epsilon_{k_e}) = b_{k_e}^T B_e^{(1)} b_{-k_e}, \quad (4.318)$$

$$-2s_e^{(1)} + s_e^{(2)} (v_Z - \epsilon_{k_e}) = b_{k_e}^T B_e^{(2)} b_{-k_e} + 2\bar{t}^Z D_e^{(2)}. \quad (4.319)$$

Using (4.317), (4.318) and (4.319), we can calculate the factor in front of  $(\epsilon_k - \epsilon_{k_e})^2$  on the r.h.s. of Eq. (4.316)

$$\begin{aligned} -t_Z^2 e^{ik_e} ((B_e)^2)_{11} D_e^{(2)} &= -s_e^{(1)} \bar{t}^Z D_e^{(2)} = -s_e^{(1)} \left( -s_e^{(1)} + \frac{1}{2} s_e^{(2)} (v_Z - \epsilon_e) - \frac{1}{2} b_{k_e}^T B_e^{(2)} b_{-k_e} \right) \\ &= (s_e^{(1)})^2 + \frac{1}{2} b_{k_e}^T (-s_e^{(2)} B_e^{(1)} + s_e^{(1)} B_e^{(2)}) b_{-k_e} \end{aligned} \quad (4.320)$$

Finally, we use (4.300) for  $n = 1, 2$  and  $k = k_e$

$$s_e^{(1)} = -B_e + (A - \epsilon_{k_e}) B_e^{(1)}, \quad (4.321)$$

$$s_e^{(2)} = -2B_e^{(1)} + (A - \epsilon_{k_e}) B_e^{(2)}, \quad (4.322)$$

which leads to

$$\begin{aligned} -s_e^{(2)} B_e^{(1)} + s_e^{(1)} B_e^{(2)} &= 2(B_e^{(1)})^2 - B_e B_e^{(2)} \\ &= 2(B_e^{(1)})^2 - B_e^{(2)} B_e, \end{aligned} \quad (4.323)$$

where we used (4.299) in the last step. Inserting this result in (4.320) and using  $B_e b_{-k_e} = -a_{-k_e} = 0$  we arrive at

$$\begin{aligned} -t_Z^2 e^{ik_e} ((B_e)^2)_{11} D_e^{(2)} &= \\ &= (s_e^{(1)})^2 + b_{k_e}^T (B_e^{(1)})^2 b_{-k_e}. \end{aligned} \quad (4.324)$$

This proves that the first and second term on the r.h.s. of Eq. (4.315) are the same and we get the final result of Eq. (4.91)

$$N_k = \frac{2i}{e^{-2ik_e} - 1} a_{k_e}^T a_{k_e} (k - k_e) + O(\epsilon_k - \epsilon_{k_e})^3. \quad (4.325)$$

### 8.5 Topological constraints for edge states

In this Appendix we prove the central Eq. (4.110) for the topological constraint of edge state that outgoing and incoming vertices must alternate. In a first step we discuss in more detail the phase-dependence of following determinants evaluated with the energy at the top of the band (i.e.,  $k_0 = 0$  for  $\nu$  even and  $k_0 = \pi$  for  $\nu$  odd)

$$d_1(\varphi) \equiv \tilde{d}_{2,Z-1}(\epsilon_{k_0}(\varphi), \varphi), \quad (4.326)$$

$$d_2(\varphi) \equiv \tilde{d}_{1,Z-2}(\epsilon_{k_0}(\varphi), \varphi), \quad (4.327)$$

where we have left out the band index for simplicity. Using (4.107) and the periodicity of the band dispersion  $\epsilon_k(\varphi) = \epsilon_k(\varphi + \frac{2\pi}{Z})$ , we note the relation

$$d_1(\varphi) = d_2\left(\varphi + \frac{2\pi}{Z}\right). \quad (4.328)$$

When an edge state of  $H_R$  enters or leaves at  $\varphi = \varphi_0$  the phase-dependence of the edge state energy will connect smoothly to the band dispersion with the same first derivative

$$\epsilon'_e(\varphi_0) = \epsilon'_{k_0}(\varphi_0) \quad . \quad (4.329)$$

This means that up to linear order in  $\varphi - \varphi_0$  the two dispersion are the same  $\epsilon_e = \epsilon_{k_0} + O(\varphi - \varphi_0)^2$ . Thus, up to this order we can replace  $\epsilon_{k_0}(\varphi)$  by  $\epsilon_e(\varphi)$  in (4.326) and (4.327) and get with the help of (4.99) and (4.100)

$$d_1(\varphi) = -e^{-ik_e(\varphi)} + O(\varphi - \varphi_0)^2, \quad (4.330)$$

$$d_2(\varphi) = -e^{ik_e(\varphi)} + O(\varphi - \varphi_0)^2. \quad (4.331)$$

Taking the derivate of these relations at  $\varphi = \varphi_0$  and using  $k_e(\varphi) = k_0 + i\kappa_e(\varphi)$  together with  $\kappa_e(\varphi_0) = 0$  and (4.328) we get

$$d_1(\varphi_0) = d_2(\varphi_0) = d_1\left(\varphi_0 - \frac{2\pi}{Z}\right) = -e^{-ik_0}, \quad (4.332)$$

$$d'_1(\varphi_0) = -\kappa'_e(\varphi_0) e^{-ik_0}, \quad (4.333)$$

$$d'_1(\varphi_0) = -d'_2(\varphi_0) = -d'_1\left(\varphi_0 - \frac{2\pi}{Z}\right). \quad (4.334)$$

Since  $\kappa'_e(\varphi_0) \geq 0$  corresponds to an edge state of  $H_R$  leaving/entering the band, we get with  $e^{-ik_0} = (-1)^{k_0/\pi}$  from (4.333) the central property

$$\begin{aligned} \text{Entering/leaving edge state of } H_R &\Rightarrow \\ \text{sign}\{d'_1(\varphi_0)\} &= \pm(-1)^{k_0/\pi}. \end{aligned} \quad (4.335)$$

In summary, we conclude that the determinant  $d_1(\varphi)$  has always the same value  $-e^{-ik_0}$  at the points  $\varphi = \varphi_0$  where edge states of  $H_R$  enter or leave the band, the derivative  $d'_1(\varphi)$  of the determinant has different sign for entering and leaving edge states at  $\varphi = \varphi_0$ , and  $d'_1(\varphi)$  has a different sign for  $\varphi = \varphi_0$  and  $\varphi = \varphi'_0 = \varphi_0 - \frac{2\pi}{Z}$ . In terms of the contractions this means that

$$\begin{aligned} \text{(I) } d_1(\varphi) &\text{ has the same value } -e^{-ik_0} \text{ at all} \\ &\text{vertices,} \end{aligned} \quad (4.336)$$

$$\begin{aligned} \text{(II) } \text{sign}\{d'_1(\varphi)\} &= \pm(-1)^{k_0/\pi} \text{ for} \\ &\text{outgoing/incoming vertices,} \end{aligned} \quad (4.337)$$

and these two properties hold irrespective of whether the vertices are the right or left vertices of a contraction.

In addition, we get the property that if the determinant fulfils the condition  $d_1(\varphi) = -e^{-ik_0}$ , either an edge state of  $H_R$  must enter/leave at  $\varphi$  or at  $\varphi + \frac{2\pi}{Z}$ , i.e., in terms of the contraction

$$\text{(III) } d_1(\varphi) = -e^{-ik_0} \Rightarrow \text{a vertex appears at } \varphi. \quad (4.338)$$

To prove this we first note that  $e^{-ik_0} = e^{ik_0}$  due to  $k_0 = 0, \pi$ . With (4.60) this implies  $a_{k_0}(1) = 0$  and (4.59) leads to

$$t_1 \cdots t_{j-1} a_{k_0}(j) = -d_{2,j-1}(\epsilon_{k_0}) s(\epsilon_{k_0}) t_Z e^{-ik_0}, \quad (4.339)$$

for  $j = 2, \dots, Z-1$ . We consider two cases: (a)  $s(\epsilon_{k_0}) = 0$  and (b)  $s(\epsilon_{k_0}) \neq 0$ . For case (a) an edge state of  $H_R$  enters/leaves the band and  $a_{k_0} = 0$  which is consistent with (4.102) since the two quasimomenta  $\pm k_0$  are equivalent. However, note that the Bloch vector  $\chi_{k_0}$  is still well-defined since  $N_{k_0}$  is also zero. For case (b) we get  $a_{k_0} \neq 0$  and  $N_{k_0} \neq 0$  such that  $\chi_{k_0}(1) = a_{k_0}(1)/\sqrt{N_{k_0}} = 0$ . This means that we have found an edge state when the boundary of  $H_R$  is defined between the sites  $m = 1$  and  $m = 2$  which corresponds to the shifted system at  $\varphi + \frac{2\pi}{Z}$ . Therefore, in this case an edge state must enter/leave the band at phase  $\varphi + \frac{2\pi}{Z}$ .

Since the derivative of the determinant must alternate between two consecutive points where it takes the same value, we get from the three properties (4.336), (4.337) and (4.338) the central condition that outgoing and incoming vertices must alternate, which proves Eq. (4.110).

## 8.6 Diophantine equation

Here we prove the Diophantine equation (4.117) for the case when all contraction lines go to the right, see Fig. 30. We split the phase interval of size  $2\pi$  in  $Z$  subintervals of size  $\frac{2\pi}{Z}$ . Starting at  $\varphi = 0$  with an outgoing vertex, we find in the first subinterval  $s$  incoming vertices including the one at  $\varphi = \frac{2\pi}{Z}$ . At phase  $\varphi = \frac{2\pi}{Z} + 0^+$  the topological charge is  $s-1$ , corresponding to  $s-1$  right-going contractions. Therefore, in the second subinterval we obtain  $s-1$  incoming vertices. At all following phase values  $\varphi = \frac{2\pi n}{Z} + 0^+$ , with  $n = 2, \dots, Z-1$ , the topological charge is either  $s-1$  or  $s$ . This gives either  $s$  or  $s-1$  incoming vertices for the remaining subintervals. As a consequence, we obtain for the maximal and minimal value of the number  $M_+$  of incoming vertices the result

$$M_+^{\max} = 2s - 1 + s(Z - 2) = -1 + sZ, \quad (4.340)$$

$$M_+^{\min} = 2s - 1 + (s - 1)(Z - 2) = 1 - Z + sZ, \quad (4.341)$$

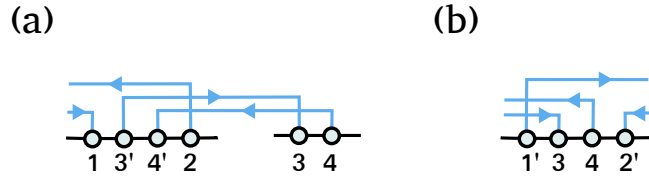


Figure 44: (a) The two vertices 1 and 2 are assumed to be a pair of left-going vertices with shortest distance where 1 is an incoming and 2 an outgoing vertex. As shown it is not possible that two other right-going vertices 3' and 4' occur between 1 and 2 since the two vertices 3 and 4 would form a pair with shorter distance compared to the pair (1, 2). (b) Here we show the two right-going vertices 1' and 2' corresponding to 1 and 2 of (a) which have the same distance. As shown it is not possible that two left-going vertices 3 and 4 can appear in between since they would have a shorter distance compared to the pair (1, 2).

which gives the following result for  $M = M_- - M_+ = -M_+$

$$M^{\min} = -M_+^{\max} = 1 - sZ, \quad (4.342)$$

$$M^{\max} = -M_+^{\min} = Z - 1 - sZ. \quad (4.343)$$

The proof when all contractions go to the left is analog and gives the same result for  $M = M_-$ . This proves (4.117).

### 8.7 Construction of all edge state configurations

Here we show that all configurations of contractions can be constructed by starting from the ones where all contractions have the same direction and iteratively adding pairs of contractions with different directions according to Fig. 31. To prove this statement we go the other way around and show that we can eliminate all such pairs iteratively starting from any given configuration until we end in a configuration where all contractions have the same direction.

Taking any configuration we first identify that pair (1, 2) of left-going vertices, one of them outgoing and the other incoming, which have shortest distance such that no other left-going vertex is allowed to occur in between, see the two vertices 1 and 2 in Fig. 44(a). W.l.o.g. we assume that vertex 1 is incoming and 2 is outgoing (the other case can be proven analog). We first prove that it is also not possible that right-going vertices can occur between 1 and 2. In Fig. 44(a) we have shown a pair (3', 4') of two right-going vertices between 1 and 2. These two vertices are connected to (3, 4). Thereby, it is not possible that 4 = 2 since otherwise the contraction between 4' and 4 would be shorter than the one between 3' and 3 (note that 3 can not occur between 1 and 2 since we assumed that no other left-going vertex can be in this interval). As a consequence we find that the two vertices 3 and 4 have a shorter distance compared to 1 and 2 which leads to a contradiction since the pair (1, 2) was assumed to be the one with the shortest distance. Similarly, we can prove that no other vertex can occur between 1' and 2'. First, a right-going vertex 3' is not allowed between 1' and 2' since this would lead to a left-going vertex 3 between 1 and 2 which is not allowed. As shown in Fig. 44(b) it is also not allowed that two left-going vertices 3 and 4 can occur between 1' and 2' since their distance would be smaller than the one between 1 and 2, again leading to a contradiction. As a consequence, we find that no other vertex can appear between 1 and 2 and between 1' and 2' such that this pair can be taken out without violating the topological constraint. Proceeding in this way one finds that all pairs of contractions with different directions can be eliminated until one arrives at a configuration where all left-going vertices are either incoming or outgoing vertices. These are the configurations where all contractions have the same direction, see Fig. 30.

### 8.8 Density

Here we present the technical details of Section 3.3 to calculate the density of a half-infinite system, analog to Ref. [[48]] for a system in continuum. First we analyse the Friedel density  $\rho_F^{(\alpha)}(n, j)$  of a certain band  $\alpha$  starting from (4.143) and, by using the periodicity  $\chi_k^{(\alpha)} = \chi_{k+2\pi}^{(\alpha)}$ , shift the integration region to  $\int_{-\pi/2}^{3\pi/2} dk$ . We close the integration contour in the upper half via the closed path  $\gamma$  shown in Fig. 45 and obtain

$$\rho_F^{(\alpha)}(n, j) = -\frac{1}{2\pi} \oint_{\gamma} dk \left[ \chi_k^{(\alpha)}(j) \right]^2 e^{2ikn}. \quad (4.344)$$

To obtain this we have used that the two contributions from  $3\pi/2 \rightarrow 3\pi/2 + i\infty$  and  $-\pi/2 + i\infty \rightarrow -\pi/2$  cancel each other due to the periodicity of  $\chi_k^{(\alpha)} = \chi_{k+2\pi}^{(\alpha)}$ . To show that the asymptotic part from  $3\pi/2 + i\infty \rightarrow -\pi/2 + i\infty$  is zero we need to analyse the asymptotic form of the integrand for  $k = x + iy$  with  $x, y$  real and  $y \rightarrow \infty$ . Using the results (4.354) and (4.362) derived in Appendix 8.9, we get for  $n \geq 1$

$$\begin{aligned} [\chi_k^{(\alpha)}(j)]^2 e^{2ikn} &\rightarrow \frac{1}{Z} (t_1 \dots t_{j-1})^2 t_Z^2 (-\epsilon_k^{(\alpha)})^{-2j} e^{2ik(n-1)} \\ &\sim e^{2ik(n-1+\frac{j}{Z})} \rightarrow 0. \end{aligned} \quad (4.345)$$

Using the analytic continuation of  $[\chi_k^{(\alpha)}(j)]^2$  according to Fig. 26(b), we can deform the contour  $\gamma$  to a sum over the contours  $\gamma_{bc}^{(\nu)}$  around the branch cuts starting at  $k_{bp}^{(\nu)}$  together with the contours  $\gamma_e^{(\nu)}$  encircling edge poles, with  $\nu \in \{\alpha - 1, \alpha\}$  and  $1 \leq \nu \leq Z - 1$ . Thereby, the pole lying below  $k_{bp}^{(\nu)}$  occurs only under the conditions

$$\text{Im}(\kappa_e^{(\nu)}) > 0, \quad (4.346)$$

$$\epsilon_{bp}^{(\nu)} \leq \epsilon_e^{(\nu)} \quad \text{for} \quad \nu = \begin{cases} \alpha - 1 \\ \alpha \end{cases}. \quad (4.347)$$

The first condition requires the edge state to be an eigenstate of  $H_R$  and the second condition is necessary since edge poles arising from the analytic continuation of band  $\alpha$  must either lie in gap  $\alpha - 1$  with energy above  $\epsilon_{bp}^{(\alpha-1)}$  or in gap  $\alpha$  with energy below  $\epsilon_{bp}^{(\alpha)}$ , see Fig. 25(b). Eq. (4.344) for the Friedel density can therefore be split into two parts from branch cut integrations and pole contributions

$$\rho_F^{(\alpha)}(n, j) = \rho_{F,bc}^{(\alpha)}(n, j) + \rho_{F,p}^{(\alpha)}(n, j), \quad (4.348)$$

with

$$\rho_{F,bc}^{(\alpha)}(n, j) = -\frac{1}{2\pi} \sum_{\substack{\nu=\alpha, \alpha-1 \\ 1 \leq \nu \leq Z-1}} \oint_{\gamma_{bc}^{(\nu)}} dk \left[ \chi_k^{(\alpha)}(j) \right]^2 e^{2ikn}, \quad (4.349)$$

$$\rho_{F,p}^{(\alpha)}(n, j) = -\frac{1}{2\pi} \sum'_{\nu=\alpha, \alpha-1} \oint_{\gamma_p^{(\nu)}} dk \left[ \chi_k^{(\alpha)}(j) \right]^2 e^{2ikn}, \quad (4.350)$$

where the prime at the last  $\sum'$  refers to the conditions stated in (4.346) and (4.347). To evaluate the branch cut contribution we write  $k = k_{bp}^{(\nu)} + i\kappa \pm 0^+$ , with  $\kappa > 0$ , and use  $\chi_{k_{bp}^{(\nu)} + i\kappa - 0^+}^{(\alpha)}(j) =$

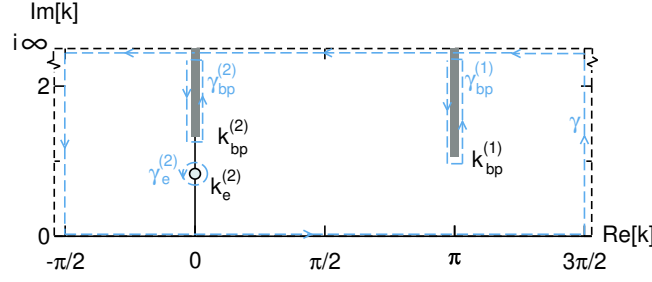


Figure 45: Closing of the integration contour in the upper half via the path  $\gamma$  to calculate the Friedel density  $\rho_F^{(2)}(n, j)$  of band  $\alpha = 2$ , corresponding to the analytic continuation shown in Fig. 26(b). The two contributions from  $3\pi/2 \rightarrow 3\pi/2 + i\infty$  and  $-\pi/2 + i\infty \rightarrow -\pi/2$  cancel each other due to the periodicity of  $\chi_k^{(2)} = \chi_{k+2\pi}^{(2)}$ . The asymptotic part from  $3\pi/2 + i\infty \rightarrow -\pi/2 + i\infty$  is zero. Using the analytic properties from Fig. 26(b) the integration along  $\gamma$  is the same as the sum of the two integrals along  $\gamma_{bc}^{(1)}$  and  $\gamma_{bc}^{(2)}$  surrounding the two branch cuts, together with the integral along  $\gamma_e^{(2)}$  encircling the edge pole at  $k_e^{(2)}$ .

$[\chi_{k_{bp}^{(\nu)} + i\kappa + 0^+}^{(\alpha)}(j)]^*$  which follows from (4.38), (4.39) and (4.68). This gives

$$\begin{aligned} \rho_{F,bc}^{(\alpha)}(n, j) &= \frac{1}{\pi} \sum_{\substack{\nu=\alpha, \alpha-1 \\ 1 \leq \nu \leq Z-1}} e^{-2\kappa_{bp}^{(\nu)} n} \\ &\times \text{Im} \int_0^\infty d\kappa \left[ \chi_{k_{bp}^{(\nu)} + i\kappa + 0^+}^{(\alpha)}(j) \right]^2 e^{-2\kappa n}. \end{aligned} \quad (4.351)$$

Summing up all branch cut contributions from the occupied bands  $\alpha = 1, \dots, \nu$ , we find that the contributions from adjacent bands  $\alpha$  and  $\alpha + 1$  for the common branch cut starting at  $k_{bp}^{(\alpha)}$  will exactly cancel each other since the values of the integrand left and right due to branch cut are interchanged, see Fig. 24(b,c). What remains is only the branch cut from band  $\alpha = \nu$  in gap  $\nu$ , leading to Eq. (4.172).

To calculate the edge pole contribution we use (4.88) and (4.90) to get for  $k \rightarrow k_e^{(\nu)}$

$$\begin{aligned} \chi_k^{(\alpha)}(j)^2 &\rightarrow \frac{N_{k_e^{(\nu)}}^e}{N_k^{(\alpha)}} \left[ \chi_{k_e^{(\nu)}}^e(j) \right]^2 \\ &\rightarrow \frac{-i}{k - k_e^{(\nu)}} \left[ \chi_{k_e^{(\nu)}}^e(j) \right]^2. \end{aligned} \quad (4.352)$$

This gives for (4.350) with (4.87)

$$\rho_{F,p}^{(\alpha)}(n, j) = - \sum'_{\nu=\alpha, \alpha-1} [\psi_{k_e^{(\nu)}}^e(n, j)]^2. \quad (4.353)$$

As a result the pole contribution of the Friedel density of band  $\alpha$  cancels exactly the edge state density of those edge states which belong to band  $\alpha$ , see Eq. (4.171).

### 8.9 Asymptotic values

In this Appendix we determine the asymptotic form of various quantities for  $k = x + iy$  with  $x, y$  real and  $y \rightarrow \infty$ . Since  $e^{-ik} = e^{-ix}e^y \rightarrow \infty$  and  $e^{ik} = e^{ix}e^{-y} \rightarrow 0$  we find from (4.62) that  $e^{-ik} \rightarrow \frac{1}{i^Z} (-\epsilon_k^{(\alpha)})^Z$  or

$$(-\epsilon_k^{(\alpha)})^Z \rightarrow \bar{i}^Z e^{-ik}. \quad (4.354)$$



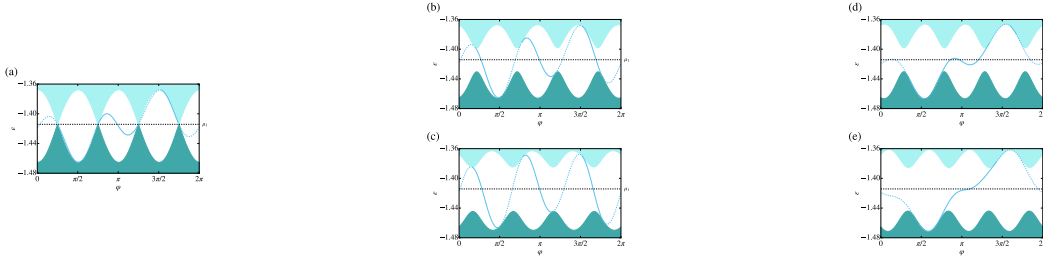


Figure 46: The phase-dependence of the edge state energy (solid blue line for  $H_R$  and dashed blue line for  $H_L$ ) in the first gap  $\nu = 1$  for  $Z = 4$  and  $V = 0.3$ ,  $t = 1$ ,  $\delta t = 0.2$ , for various functions  $F_v$  and  $F_t$  in Eqs. (4.10) and (4.11) taken from (4.276) with fixed and random parameters for  $v_j(0)$  and  $t_j(0) - t$  at phase  $\varphi = 0$  but different choices for the phase-dependence in between (via the random parameters  $\gamma_j^{(1)}$  for  $\gamma = v, t$  used in (4.282)), see the Supplemental Material [26] for the concrete parameters. The chemical potential  $\mu_1$  is indicated by a dashed black line. In (a) we have chosen  $\gamma_j^{(1)} = 0$  which leads to a gap closing at  $\varphi = \frac{\pi}{4}$ . In (c) the  $\gamma_j^{(1)}$  are by a factor of 2 larger compared to (b). (d) and (e) correspond to (b) and (c), respectively, but the sign of all  $\gamma_j^{(1)}$  has been reversed. As can be seen for the choices in (d) and (e) no edge state of  $H_R$  crosses  $\mu_1$  in the phase interval  $[0, \frac{2\pi}{Z}] = [0, \frac{\pi}{2}]$ .

We note that this asymptotic condition has no unique solution for the individual bands  $\epsilon_k^{(\alpha)}$ . A detailed analysis (not important for the following) gives

$$\epsilon_k^{(\alpha)} \rightarrow -\bar{t} e^{-ik/Z} e^{i2\pi\alpha/Z}. \quad (4.355)$$

As a result of (4.354) we get  $\epsilon_k^{(\alpha)} \rightarrow \infty$  such that all determinants have the asymptotic condition  $d_{nm}(\epsilon_k^{(\alpha)}) \rightarrow (-\epsilon_k^{(\alpha)})^{m-n+1}$  for  $m - n + 1 \geq 0$  (note that  $d_{nm} \equiv 1$  for  $n > m$ ). Using (4.50), (4.51), and (4.52) this gives for  $j = 1, \dots, Z - 1$

$$a_k^{(\alpha)}(j) \rightarrow t_1 \cdots t_{j-1} t_z (-\epsilon_k^{(\alpha)})^{Z-1-j} e^{-ik}, \quad (4.356)$$

$$a_{-k}^{(\alpha)}(j) \rightarrow \frac{\bar{t}^Z}{t_1 \cdots t_{j-1} t_z} (-\epsilon_k^{(\alpha)})^{j-1}. \quad (4.357)$$

Furthermore, from (4.46), we get  $s(\epsilon_k^{(\alpha)}) \rightarrow (-\epsilon_k^{(\alpha)})^{Z-1}$  or with (4.354)

$$\left[ s(\epsilon_k^{(\alpha)}) \right]^2 \rightarrow \bar{t}^Z (-\epsilon_k^{(\alpha)})^{Z-2} e^{-ik}. \quad (4.358)$$

Since  $a_k^{(\alpha)}(1) \rightarrow \bar{t}^Z (-\epsilon_k^{(\alpha)})^{Z-2} e^{-ik}$  is largest and  $a_k^{(\alpha)}(j) a_{-k}^{(\alpha)}(j) \rightarrow \bar{t}^Z (-\epsilon_k^{(\alpha)})^{Z-2} e^{-ik}$ , we obtain

$$\begin{aligned} (a_k^{(\alpha)})^T a_k^{(\alpha)} + \left[ s(\epsilon_k^{(\alpha)}) \right]^2 &\rightarrow [a_k^{(\alpha)}(1)]^2 \\ &\rightarrow \bar{t}_Z^2 (-\epsilon_k^{(\alpha)})^{2(Z-2)} e^{-ik}, \end{aligned} \quad (4.359)$$

$$\begin{aligned} N_k^{(\alpha)} &= (a_k^{(\alpha)})^T a_{-k}^{(\alpha)} + \left[ s(\epsilon_k^{(\alpha)}) \right]^2 \\ &\rightarrow Z \bar{t}^Z (-\epsilon_k^{(\alpha)})^{Z-2} e^{-ik}. \end{aligned} \quad (4.360)$$

Using (4.144) this gives for  $\sum_{j=1}^Z [\chi_k^{(\alpha)}(j)]^2 = (\chi_k^{(\alpha)})^T \chi_k^{(\alpha)}$  the asymptotic form

$$\sum_{j=1}^Z [\chi_k^{(\alpha)}(j)]^2 \rightarrow \frac{1}{Z} \bar{t}_Z^2 (-\epsilon_k^{(\alpha)})^{-2} e^{-2ik}. \quad (4.361)$$

For  $[\chi_k^{(\alpha)}(j)]^2 = \frac{1}{N_k^{(\alpha)}}[a_k^{(\alpha)}(j)]^2$  for  $j = 1, \dots, Z - 1$  and  $[\chi_k^{(\alpha)}(Z)]^2 = \frac{1}{N_k^{(\alpha)}} \left[ s(\epsilon_k^{(\alpha)}) \right]^2$  we get from (4.356), (4.358), and (4.360)

$$[\chi_k^{(\alpha)}(j)]^2 \rightarrow \frac{1}{Z} (t_1 \dots t_{j-1})^2 t_Z^2 (-\epsilon_k^{(\alpha)})^{-2j} e^{-2ik}. \quad (4.362)$$

### 8.10 Tuning of edge states via choice of the phase-dependence

Here we show that the phase-dependence of  $v_j(\varphi)$  and  $t_j(\varphi)$  in the phase interval  $0 < \varphi < \frac{2\pi}{Z}$  for given parameters  $v_j(0), t_j(0)$  and  $v_j(\frac{2\pi}{Z}) = v_{j+1}(0), t_j(\frac{2\pi}{Z}) = t_{j+1}(0)$  at the boundaries  $\varphi = 0, \frac{2\pi}{Z}$  can always be chosen such that no edge state crosses  $\mu_\nu$  in this interval. This can be achieved by using the parametrization (4.276) for the phase-dependence. For  $G_\gamma = 0$  a gap closing occurs in the middle of the interval at  $\varphi = \frac{\pi}{Z}$  where all  $v_j(\frac{\pi}{Z}) = 0$  and  $t_j(\frac{\pi}{Z}) = t$ , see (4.280). The spectrum of the edge state between the first and second band is shown in Fig. 46(a) for  $Z = 4$ . At the gap closing at  $\varphi = \frac{\pi}{4}$  the edge state changes from an edge state of  $H_L$  to an edge state of  $H_R$ . In Fig. 46(b,c) we show the same for a slightly nonzero function  $G_\gamma$  by choosing a random set of parameters  $\gamma_j^{(1)}$  in (4.282) on two scales  $\sim 0.05$  and  $\sim 0.1$ , respectively. In Fig. 46(d,e) we have changed the sign of all  $\gamma_j^{(1)}$ . As can be seen in Figs. 46(b,c) close to  $\varphi \sim \frac{\pi}{4}$  the edge state of  $H_R$  is taken upwards together with the upper band whereas in Figs. 46(d,e) it is taken downwards together with the lower band. This has the effect that the edge state of  $H_R$  does no longer cross the chemical potential in Figs. 46(d,e). This behaviour is generic and can be observed for any  $Z$  and for any initial parameters for  $\gamma_j^{(0)}$  and for all kinds of random parameters for  $\gamma_j^{(1)}$ . Changing the sign of  $G_\gamma$  corresponds precisely to the change of the orientation of the edge pole encircling the branch cut in order to avoid the crossing of  $k_e^{(\nu)}(\varphi)$  through the value of  $k_{\mu_\nu}$  corresponding to the chemical potential, see Fig. 37. For arbitrary  $\nu$  we find that this point describes the crossover from the edge state encircling the branch cut  $\nu$  times in clockwise direction to the case of  $Z - \nu$  windings with counter-clockwise orientation. This corresponds to the two cases of  $\nu$  edge states moving upwards compared to  $Z - \nu$  edge states moving downwards, see Eqs. (4.266) and (4.267). We note that additional windings by multiples of  $Z$  do not occur [corresponding to values  $s \neq 0, 1$  in the Diophantine equation (4.263)] since with our choice (4.276) of the phase parametrization we have only chosen at most  $Z$  Fourier modes for  $F_\gamma(\varphi)$ .

## Part III

# Examples and Solved Problems

### 9. Typical examples

## Part IV

# Other topics

## 5 CDW state high magnetic field quasi-1D mean field by Grigoriev, Lyubshin

### 0..1 Main idea of the article and what is important here?

(пока не знаю.)

### 0..2 Introduction

The phase diagram and the structure of CDW state in high magnetic field in quasi-1D materials: mean-field approach

P. D. Grigoriev

D. S. Lyubshin

#### Abstract

We develop the mean-field theory of a charge-density wave (CDW) state in magnetic field and study the properties of this state below the transition temperature. We show that the CDW state with shifted wave vector in high magnetic field (CDW<sub>x</sub> phase) has at least double harmonic modulation on the most part of the phase diagram. In the perfect nesting case the single harmonic CDW state with shifted wave vector exists only in a very narrow region near the tricritical point where the fluctuations are very strong. We show that the transition from CDW<sub>0</sub> to CDW<sub>x</sub> state below the critical temperature is accompanied by a jump of the CDW order parameter and of the wave vector rather than by their continuous increase. This implies a first order transition between these CDW states and explains the strong hysteresis accompanying this transition in many experiments. We examine how the phase diagram changes in the case of imperfect nesting.

The properties of metals with a charge-density-wave (CDW) ground state attract great attention since the fifties (see, e.g., monographs [[1, 2]]). In quasi-1D metals, the Fermi surface (FS) consists of two slightly warped sheets separated by  $2k_F$  and roughly possesses the nesting property that leads to the Peierls instability and favors the formation of the charge- or spin-density-wave (SDW) state at low temperature.

The mean-field description is known to be unable to describe strictly 1D conductors, where the non-perturbative methods and exactly solvable models are of great importance.[3] However, in most materials manifesting CDW or SDW phenomena the nonzero value of the electron transfer integral between conducting chains and the 3D character of the electron-electron interactions and lattice elasticity reduce the deviations from the mean-field solution and also make most of the methods and exactly solvable models developed for strictly 1D case inapplicable.[4] The effect of fluctuations in Q1D metals and their influence on the mean-field description of the Peierls transition in these metals has been considered in a number of papers (see e.g. [[5, 6]] and references therein). It was shown that the interchain dispersion of electrons strongly damps the fluctuation and validate the mean-field description. In Q1D organic metals,[7] at which our present study is mainly aimed, the free electron dispersion near the Fermi level is, approximately, given by

$$\varepsilon_\sigma(\mathbf{k}) = \hbar v_F(|k_x| - k_F) - t_\perp(\mathbf{k}_\perp) - \sigma H, \quad (5.1)$$

where  $v_F$  and  $k_F$  are the Fermi velocity and Fermi momentum in the chain ( $x$ ) direction,  $H \equiv \mu_B B$  is the Zeeman energy,  $\mu_B$  is the Bohr magneton and  $B$  is the external magnetic

field. The perpendicular-to-chain term,  $t_{\perp}(\mathbf{k}_{\perp})$ , is much greater than the energy scale of the CDW(SDW) transition temperature,  $T_{c0}$ . Only the "imperfect nesting" part,  $t'_b$ , of  $t_{\perp}(\mathbf{k}_{\perp})$  is of the order of  $T_{c0}$  (see Eqs. (5.2), (5.3)). Hence, the criterion for the mean-field theory to be applicable, [5, 6]  $t_{\perp} \gg T_{c0}$ , is reliably satisfied in most Q1D organic metals.

The mean-field description of the CDW properties is, in many aspects, very similar to the BCS theory. The pairing of two electrons in superconductors is replaced in CDW by the pairing of an electron with the hole on the opposite sheet of the Fermi surface. The charge and spin coupling constants in CDW (see interaction Hamiltonian (5.5)) are analogous to coupling constants in spin-singlet and spin-triplet channels in superconductivity. The CDW phase with shifted nesting vector is similar to the non-uniform superconducting phase (LOFF phase). [9, 8] However, there are several important differences between these two many-particle effects. The first difference is that the formation of a gap in the electron spectrum in CDW leads to an insulating rather than superconducting state. This happens due to the pinning of the CDW condensate by crystal imperfections. Hence, the CDW state does not reveal a superfluid current. [1, 2] Other differences appear, e.g., in the excitation spectrum. In particular, the lowest energy excitations in magnetic field or at imperfect nesting are not always the electron pairs as in BCS theory but may be the "soliton kinks". [10, 11, 12, 13, 14]

The theoretical investigation of the CDW/SDW properties at a mean-field level comprises two main branches. The first focuses on the transition line from the metallic states to CDW/SDW state using susceptibility calculations. This allows to include many additional factors into the theoretical model, such as different free electron dispersion relations, spin and orbital effects of the external magnetic field, applied pressure etc. It helped to discover and to explain many beautiful effects such as the field-induced spin- and charge-density waves (FISDW), [18, 19, 20, 21] the increase of the transition temperature due to the one-dimensionization of the electron spectrum in high magnetic field, [18, 23] the CDW state with shifted nesting vector [19, 24] etc. However, all these results cannot be continued below the transition temperature since they are based on the electron susceptibility calculations using metallic-state electron Green's functions. In particular, the calculation of [[24]] predicts an appearance of the  $CDW_x$  phase with shifted wave vector as the magnetic field exceeds some critical value. This calculation gives the metal- $CDW_x$  transition line,  $T_c(H)$ , and the dependence of the optimal shift  $q_x(H)$  of CDW wave vector on magnetic field at this transition line. However, this calculation does not extend below the transition temperature, and, hence, cannot be used to describe the properties and the phase diagram inside the CDW phase. For example, it neither describes the structure of the  $CDW_x$  phase below  $T_c(H)$  nor gives the  $CDW_0$ - $CDW_x$  transition line  $H_{c1}(T)$  or the kind of this transition. The mean-field study of CDW in Ref. [[22]] is applicable only in weak field when the CDW wave vector does not shift from its zero-field value.

The second branch of investigation involves the soliton physics. It appeared first to describe the ground state of polyacetylene [10, 11] and then developed into a rather large activity (see Refs. [[15, 16]] for a review). In particular, the soliton structure and the energy spectrum of CDW state in external magnetic field were considered [17]. However, all these results were only derived at zero temperature and perfect nesting and are not applicable at temperatures of the order of transition temperature,  $T \sim T_c$ . The finite-temperature phase transition to soliton phase has been considered in Ref. [[14]]. However, that phase diagram refers to zero field and finite shift of the chemical potential from the value  $\mu_0$  corresponding to the commensurate CDW. [14] The analysis of the CDW phase diagram in magnetic field has been performed in the case of perfect nesting and at the electron density close to half-filling. [25] However, this approach is also unable to describe the incommensurate case far from half-filling. In particular, it suggests the one-harmonic modulation of charge density, which is usually unstable in high magnetic field (see below).

Further theoretical description of the CDW state in magnetic field became important last

years because of the intensive experimental study of the CDW state in strongly anisotropic organic metals. [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36]

In the present paper we study the CDW phase diagram and the properties of the CDW state at finite temperature below the transition point in the mean-field approximation. This study links the results of susceptibility calculations and the results from the soliton approach. We take into account the spin effect of the external magnetic field and consider the CDW phase with shifted nesting vector. The "antinesting" term in the electron dispersion is also taken into account. We show that the  $\text{CDW}_x$  state with shifted wave vector, proposed in Ref. [[24]], has double harmonic modulation almost everywhere in the phase diagram (see Fig. 49,51). At perfect nesting,  $t'_b = 0$ , the single harmonic  $\text{CDW}_x$  state exists only in a very narrow region near the tricritical point. We show that the transition from  $\text{CDW}_0$  to  $\text{CDW}_x$  state below the critical temperature is accompanied by a jump of the CDW wave vector rather than by its continuous increase. This implies a first order transition between the CDW states and explains the strong hysteresis observed at the kink transition.[28, 29, 32, 34]

Besides the Zeeman splitting, external magnetic field affects the orbital electron motion. First, an external magnetic field perpendicular to conducting layers leads to one-dimensionization of the electron spectrum in quasi-1D metals [23] that improves the nesting property of the Fermi surface. Second, due to the strong scattering by the nesting vector, electrons in quasi-1D metal in magnetic field may form close orbits in momentum space [19]. These orbital effects increase the CDW or SDW transition temperature and may lead to the field-induced spin-density waves [18] (for a review, see [[7]]). A similar effect for the charge density waves was also proposed [21]. The interplay between spin and orbital effects of the magnetic field is quite interesting. It leads, for example, to a series of phase transitions between the CDW states with quantized nesting vector.[32]

In our present study we disregard the orbital effects of magnetic field. This limitation, however, is not very restrictive. If the magnetic field is parallel to the conducting layers, it produces only the Zeeman splitting and no orbital effect. The orbital quantization effects are, usually, more subtle than the spin effect. As an illustration, the quantized nesting phases can be observed only at very low temperature [32] because the effects of nesting quantization are strongly damped at  $\hbar\omega_c = e\hbar H/m^*c \ll 2\pi T, 2\pi\hbar/\tau, \Delta$ . This damping is somewhat similar to that of the magnetic quantum oscillations. The one-dimensionalization of electron spectrum in magnetic field can be effectively taken into account in the present study by introducing the magnetic field dependence of the imperfect nesting transfer integral,  $t'(H)$ , in the electron dispersion relation (5.2). A situation which is mathematically equivalent to the Zeeman splitting of magnetic field without its orbital effect arises when there are two slightly different Q1D chain system in the same compound with coupling between the chains of different type. The quantities  $U_c + U_s$  and  $U_c - U_s$  play the role of the electron coupling constant inside the chains and between the chains of different type, respectively. Such a system occurs in many compounds as TTF-type organic metals (see, e.g., the discussion on page 196 in [[16]]). Two slightly different types of chains occur also in  $\alpha\text{-(Per)}_2\text{M(mnt)}_2$ . [37, 35, 36] In the systems with molecular chains of two types, however, the difference between optimal CDW wave vectors on different chains is fixed by the molecular structure and does not vary as in the case of external magnetic field which gradually separates the Fermi surfaces with two different spin components.

## 1. The Model and the mean-field theory of CDW in magnetic field.

### 1.1 The model

We consider a quasi-1D metal with dispersion (5.1) and  $t_\perp(\mathbf{k}_\perp)$  given by the tight-binding model:

$$t_\perp(\mathbf{k}_\perp) = -2t_b \cos(k_y b) - 2t'_b \cos(2k_y b) - 2t_c \cos(k_z c_z), \quad (5.2)$$

where  $b$  and  $c_z$  are the lattice constants in  $y$ - and  $z$ -directions respectively. The dispersion along the  $z$ -axis is assumed to be much weaker than the dispersion along  $y$ -direction. Therefore, we omit the second harmonic  $\propto \cos(2k_z c_z)$  in the dispersion relation (5.2). Since the terms  $2t_c \cos(k_z c_z)$  and  $2t_b \cos(k_y b)$  do not violate the perfect nesting condition

$$\varepsilon(\mathbf{p} + \mathbf{Q}) = -\varepsilon(\mathbf{p}), \quad (5.3)$$

they do not influence the physics discussed below unless the nesting vector become shifted in  $y$ - $z$  plane. We do not consider such a shift in the present study.

The electron Hamiltonian is

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}},$$

with the free-electron part

$$\hat{H}_0 = \sum_{\mathbf{k}\sigma} \varepsilon_\sigma(\mathbf{k}) a_\sigma^\dagger(\mathbf{k}) a_\sigma(\mathbf{k}) \quad (5.4)$$

and the interaction part

$$\begin{aligned} \hat{H}_{\text{int}} = \frac{1}{2} \sum_{\mathbf{k}\mathbf{k}'\mathbf{Q}\sigma\sigma'} V_{\sigma\sigma'}(\mathbf{Q}) a_\sigma^\dagger(\mathbf{k} + \mathbf{Q}) a_\sigma(\mathbf{k}) \\ \times a_{\sigma'}^\dagger(\mathbf{k}' - \mathbf{Q}) a_{\sigma'}(\mathbf{k}'). \end{aligned} \quad (5.5)$$

This Hamiltonian does not include the orbital effect of magnetic field on quasi-1D electron spectrum which may be important in the case of substantially imperfect nesting and strong magnetic field perpendicular to the easy-conducting plane (see introduction). Later on we will be interested mainly in the interaction at the wave vector  $\mathbf{Q}$  close to the so-called nesting vector  $\mathbf{Q}_0$ , which is usually chosen as

$$\mathbf{Q}_0 = (\pm 2k_F, \pi/b, \pi/c). \quad (5.6)$$

The deviations  $\mathbf{Q} - \mathbf{Q}_0$  that will be considered below are of the order of  $\max\{H, t'_b\}/\hbar v_F \ll k_F$ , so that for such small deviations the interaction function

$$V_{\sigma\sigma'}(\mathbf{Q}) \approx V_{\sigma\sigma'}(\mathbf{Q} = \mathbf{Q}_0) = U_c - U_s \sigma \sigma' \quad (5.7)$$

The coupling constants  $U_c$  and  $U_s$  are the same as in Ref. [[24]]. Subscripts  $c$  and  $s$  distinguish charge and spin coupling constants. The Hamiltonian (5.5) can be obtained from the extended Hubbard model when the e-e scattering by the momenta close to the nesting wave vector  $\mathbf{Q}_0$  is only taken in to account.

## 1.2 Mean field approach

First, we note that the terms with  $\mathbf{Q} = 0$  in (5.5) only renormalizes the chemical potential and will not be omitted in subsequent calculations (all sums over  $\mathbf{Q}$  do not include  $\mathbf{Q} = 0$ ). We introduce the thermodynamic Green's function

$$g_\sigma(\mathbf{k}', \mathbf{k}, \tau - \tau') = \langle T_\tau \{ a_\sigma^\dagger(\mathbf{k}', \tau') a_\sigma(\mathbf{k}, \tau) \} \rangle, \quad (5.8)$$

where the operators are taken in the Heisenberg representation, and the average

$$D_{\mathbf{Q}\sigma} = \sum_{\mathbf{k}} g_\sigma(\mathbf{k} - \mathbf{Q}, \mathbf{k}, -0) - \delta(\mathbf{Q}, 0) n_\sigma. \quad (5.9)$$

After defining

$$\Delta_{\mathbf{Q}\sigma} = \sum_{\sigma'} (U_c - U_s \sigma \sigma') D_{\mathbf{Q}\sigma'} \quad (5.10)$$



one obtains in the mean-field approximation

$$\hat{H}_{\text{int}} = \sum_{\mathbf{Q}\mathbf{k}\sigma} a_{\sigma}^{\dagger}(\mathbf{k} + \mathbf{Q}) a_{\sigma}(\mathbf{k}) \Delta_{\mathbf{Q}\sigma} - \frac{1}{2} \sum_{\mathbf{Q}\sigma} D_{-\mathbf{Q}\sigma} \Delta_{\mathbf{Q}\sigma}. \quad (5.11)$$

Hermiticity of the Hamiltonian requires  $\Delta_{-\mathbf{Q}\sigma} = \Delta_{\mathbf{Q}\sigma}^*$ . It is now straightforward to write down the equations of motion which in the frequency representation take the form

$$\begin{aligned} [i\omega - \varepsilon_{\sigma}(\mathbf{k})] g_{\sigma}(\mathbf{k}', \mathbf{k}, \omega) - \sum_{\mathbf{Q}} \Delta_{\mathbf{Q}\sigma} g_{\sigma}(\mathbf{k}', \mathbf{k} - \mathbf{Q}, \omega) \\ = \delta(\mathbf{k}', \mathbf{k}). \end{aligned} \quad (5.12)$$

### 1.3 The Cosine Phase

Now we consider the solution with  $\Delta_{\mathbf{k}\sigma} \neq 0$  only for  $\mathbf{k} = \pm \mathbf{Q}$ , where  $\mathbf{Q} = 2k_F \mathbf{e}_x + (\pi/b) \mathbf{e}_y + \mathbf{q}$  with  $|\mathbf{q}| \ll k_F$ . If we neglect the scattering into the states with  $|k_x| \gtrsim 2k_F$ , the equations (5.12) decouple: for  $k_x > 0$  one has

$$\begin{pmatrix} i\omega - \varepsilon_{\sigma}(\mathbf{k}) & -\Delta_{\mathbf{Q}\sigma} \\ -\Delta_{-\mathbf{Q}\sigma} & i\omega - \varepsilon_{\sigma}(\mathbf{k} - \mathbf{Q}) \end{pmatrix} \hat{G} = \hat{I}, \quad (5.13)$$

where

$$\hat{G} \equiv \begin{pmatrix} g_{\sigma}(\mathbf{k}, \mathbf{k}, \omega) & g_{\sigma}(\mathbf{k} - \mathbf{Q}, \mathbf{k}, \omega) \\ g_{\sigma}(\mathbf{k}, \mathbf{k} - \mathbf{Q}, \omega) & g_{\sigma}(\mathbf{k} - \mathbf{Q}, \mathbf{k} - \mathbf{Q}, \omega) \end{pmatrix} \quad (5.14)$$

and  $\hat{I}$  is the  $2 \times 2$  identity matrix. The equation for  $k_x < 0$  may be obtained via a substitution  $\mathbf{Q} \rightarrow -\mathbf{Q}$ . Introducing the notations

$$\varepsilon_{\sigma}^{\pm}(\mathbf{k}', \mathbf{k}) = \frac{\varepsilon_{\sigma}(\mathbf{k}') \pm \varepsilon_{\sigma}(\mathbf{k})}{2} \quad (5.15)$$

and using  $\Delta_{\mathbf{Q}\sigma} \Delta_{-\mathbf{Q}\sigma} = |\Delta_{\mathbf{Q}\sigma}|^2$  from (5.13) one has

$$\begin{aligned} g_{\sigma}(\mathbf{k} - \mathbf{Q}, \mathbf{k}, \omega) &= -\frac{\Delta_{\mathbf{Q}\sigma}}{[\omega + i\varepsilon_{\sigma}(\mathbf{k})][\omega + i\varepsilon_{\sigma}(\mathbf{k} - \mathbf{Q})] + |\Delta_{\mathbf{Q}\sigma}|^2} \\ &= -\frac{\Delta_{\mathbf{Q}\sigma}}{[\omega + i\varepsilon_{\sigma}^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})]^2 + [\varepsilon_{\sigma}^-(\mathbf{k}, \mathbf{k} - \mathbf{Q})]^2 + |\Delta_{\mathbf{Q}\sigma}|^2}. \end{aligned} \quad (5.16)$$

The consistency equation therefore is

$$\Delta_{\mathbf{Q}\sigma'} = -T \sum_{\mathbf{k}\omega\sigma} \frac{(U_c - U_s \sigma \sigma') \Delta_{\mathbf{Q}\sigma}}{[\omega + i\varepsilon_{\sigma}^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})]^2 + [\varepsilon_{\sigma}^-(\mathbf{k}, \mathbf{k} - \mathbf{Q})]^2 + |\Delta_{\mathbf{Q}\sigma}|^2}. \quad (5.17)$$

where  $\omega$  takes the values  $\pi T(2n + 1)$ ,  $n \in \mathbb{Z}$ . Using the identity

$$T \sum_{\omega} \frac{1}{(\omega - i\alpha)^2 + a^2} = \frac{1}{4a} \left\{ \text{th} \frac{a - \alpha}{2T} + \text{th} \frac{a + \alpha}{2T} \right\}$$

this may be rewritten as

$$\Delta_{\mathbf{Q}\sigma'} = -\frac{1}{4} \sum_{\mathbf{k}\sigma} \frac{(U_c - U_s \sigma \sigma') \Delta_{\mathbf{Q}\sigma}}{\sqrt{|\Delta_{\mathbf{Q}\sigma}|^2 + [\varepsilon_{\sigma}^-(\mathbf{k}, \mathbf{k} - \mathbf{Q})]^2}} \left\{ \text{th} \frac{E_{\sigma,+}(\mathbf{k})}{2T} - \text{th} \frac{E_{\sigma,-}(\mathbf{k})}{2T} \right\}, \quad (5.18)$$

where

$$E_{\sigma,\pm}(\mathbf{k}) = \varepsilon_{\sigma}^+(\mathbf{k}, \mathbf{k} - \mathbf{Q}) \pm \sqrt{|\Delta_{\mathbf{Q}\sigma}|^2 + [\varepsilon_{\sigma}^-(\mathbf{k}, \mathbf{k} - \mathbf{Q})]^2} \quad (5.19)$$

is the energy spectrum in the cosine phase. From (5.2) it follows that for the right-moving electrons

$$\varepsilon_{\sigma}^+(\mathbf{k}, \mathbf{k} - \mathbf{Q}) = \frac{\hbar v_F q_x}{2} + 2t_b \sin \frac{q_y b}{2} \sin \left( k_y b - \frac{q_y b}{2} \right) - 2t'_b \cos q_y b \cos (2k_y b - q_y b) - \sigma H, \quad (5.20)$$

$$\varepsilon_{\sigma}^-(\mathbf{k}, \mathbf{k} - \mathbf{Q}) = \hbar v_F (k_x - k_F - q_x/2) - 2t_b \cos \frac{q_y b}{2} \cos \left( k_y b - \frac{q_y b}{2} \right) + 2t'_b \sin q_y b \sin (2k_y b - q_y b). \quad (5.21)$$

It is, however, more convenient to perform the momentum summation in Eq. (5.17) first.

Extending the limits of the summation over  $k_x$  to infinity, one has

$$\Delta_{\mathbf{Q}\sigma'} = \frac{\pi \nu_F |U_c| T}{2} \sum_{\omega} \left\langle \frac{(1 + \nu \sigma \sigma') \Delta_{\mathbf{Q}\sigma}}{\sqrt{[\omega + i\varepsilon_{\sigma}^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})]^2 + |\Delta_{\mathbf{Q}\sigma}|^2}} \right\rangle_{k_y}, \quad (5.22)$$

where the branch of the square root with positive real part is implied, the angular brackets stand for the averaging over all values of  $k_y$ , and we introduced the standard notation

$$\nu = -U_s/U_c \quad (5.23)$$

for the coupling ratio and

$$\nu_F = \frac{L_x}{\pi \hbar v_F} \quad (5.24)$$

for the density of states on the Fermi level per one spin component.

Expansion to the third order in  $\Delta_{\mathbf{Q}\sigma}$  yields

$$\Delta_{\mathbf{Q}\sigma'} = \sum_{\sigma} \frac{1 + \nu \sigma \sigma'}{2} (K_{\sigma}^{(1)} - |\Delta_{\mathbf{Q}\sigma}|^2 K_{\sigma}^{(3)c}) \Delta_{\mathbf{Q}\sigma} \quad (5.25)$$

(the superscript “c” stands for “cosine”) with

$$K_{\sigma}^{(1)} = \pi \nu_F |U_c| T \sum_{\omega} \left\langle \frac{\text{sign } \omega}{\omega + i\varepsilon_{\sigma}^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})} \right\rangle_{k_y}, \quad (5.26)$$

$$K_{\sigma}^{(3)c} = \frac{\pi \nu_F |U_c| T}{2} \sum_{\omega} \left\langle \frac{\text{sign } \omega}{(\omega + i\varepsilon_{\sigma}^+(\mathbf{k}, \mathbf{k} - \mathbf{Q}))^3} \right\rangle_{k_y}. \quad (5.27)$$

The second-order transition line corresponds to

$$\det \begin{pmatrix} \frac{1+\nu}{2} K_+^{(1)} - 1 & \frac{1-\nu}{2} K_-^{(1)} \\ \frac{1-\nu}{2} K_+^{(1)} & \frac{1+\nu}{2} K_-^{(1)} - 1 \end{pmatrix} = 0. \quad (5.28)$$

The left hand side of this equation is a function of  $\nu, T, H$  and  $\mathbf{q}$ . The normal to CDW phase transition occurs at some optimal value of  $\mathbf{q}$  which corresponds to the minimum of the free energy of the CDW state, that at the transition point is equivalent to the maximum of  $T_c$  or to the minimum of l.h.s. of Eq. (5.28).

Unrestricted summation over  $k_x$  introduced divergence into the summation over  $\omega$  in  $K_\sigma^{(1)}$ , which may be renormalized by introducing the zero-field and zero- $t'_b$  transition temperature  $T_{c0}$ . To obtain a closed expression for the latter, we note that for  $H = 0$  one has  $K_+^{(1)} = K_-^{(1)}$ , and the matrix in (5.28) has eigenvalues  $K_\sigma^{(1)} - 1$  and  $\nu K_\sigma^{(1)} - 1$ . Therefore, at  $H = 0$   $K_\sigma^{(1)}(T = T_c(t'_b), H = 0) = 1$ . From (5.26) at  $t'_b = 0$  after imposing a cutoff at  $\omega \sim E_F$  one obtains  $K^{(1)} = \nu_F |U_c| \ln(2\gamma E_F / \pi T)$ , and

$$T_{c0} = \frac{2\gamma E_F}{\pi} \exp\left(-\frac{1}{\nu_F |U_c|}\right), \quad (5.29)$$

in agreement with Ref. [[39]] and with the usual BCS expression. We may now write

$$K_\sigma^{(1)} = 1 + \nu_F |U_c| \left[ \ln \frac{T_{c0}}{T} + \pi T \sum_{\omega} \left( \left\langle \frac{\text{sign } \omega}{\omega + i\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})} \right\rangle_{k_y} - \frac{\text{sign } \omega}{\omega} \right) \right]. \quad (5.30)$$

The kernels  $K_\sigma^{(1)}$  and  $K_\sigma^{(3)c}$  can be further simplified in terms of the digamma function

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x):$$

$$K_\sigma^{(1)} = 1 + \nu_F |U_c| \left[ \ln \frac{T_{c0}}{T} + \psi\left(\frac{1}{2}\right) - \left\langle \text{Re } \psi\left(\frac{1}{2} + \frac{i\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})}{2\pi T}\right) \right\rangle_{k_y} \right], \quad (5.31)$$

$$K_\sigma^{(3)c} = -\frac{\nu_F |U_c|}{16\pi^2 T^2} \left\langle \text{Re } \psi''\left(\frac{1}{2} + \frac{i\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})}{2\pi T}\right) \right\rangle_{k_y}. \quad (5.32)$$

To actually find the fields  $\Delta_{\mathbf{Q}\sigma}$  just below the transition point to the leading order, one has to make the substitution  $K_\sigma^{(1)} \rightarrow K_\sigma^{(1)} - |\Delta_{\mathbf{Q}\sigma}|^2 K_\sigma^{(3)c}$  in (5.28). The ratio  $\Delta_{\mathbf{Q}-} : \Delta_{\mathbf{Q}+}$  is real and at the transition point is the column ratio in (5.28),

$$\Delta_{\mathbf{Q}-} : \Delta_{\mathbf{Q}+} = \left( \frac{1+\nu}{2} K_+^{(1)} - 1 \right) : \left( \frac{\nu-1}{2} K_-^{(1)} \right) = \left( \frac{\nu-1}{2} K_+^{(1)} \right) : \left( \frac{1+\nu}{2} K_-^{(1)} - 1 \right). \quad (5.33)$$

Introducing the (real) ratio  $\alpha \equiv |\Delta_{\mathbf{Q}-}|^2 : |\Delta_{\mathbf{Q}+}|^2$  we get

$$|\Delta_{\mathbf{Q}\sigma}^c|^2 = \frac{\alpha^{(1-\sigma)/2} \left( \nu K_+^{(1)} K_-^{(1)} - \frac{\nu+1}{2} (K_+^{(1)} + K_-^{(1)}) + 1 \right)}{K_+^{(3)c} \left( \nu K_-^{(1)} - \frac{\nu+1}{2} \right) + \alpha K_-^{(3)c} \left( \nu K_+^{(1)} - \frac{\nu+1}{2} \right)}. \quad (5.34)$$

The superscript "c" stands for the cosine phase. The CDW cosine phase with  $q_x \neq 0$  was analyzed at the transition line by means of the susceptibility calculation and called the CDW<sub>x</sub> phase, while the phase with  $q_x = 0$  was called CDW<sub>0</sub> [24].

#### 1.4 The Double Cosine Phase

Now we consider the solution with  $\Delta_{\mathbf{k}\sigma} \neq 0$  for  $\mathbf{k} = \pm \mathbf{Q}_0 \pm \mathbf{q}$ , where  $\mathbf{Q}_0 = 2k_F \mathbf{e}_x + (\pi/b) \mathbf{e}_y$  and  $q \ll k_F$ . Strictly speaking, there exist no self-consistent solutions with only four harmonics present or all others damped by a factor of  $\Delta_{\mathbf{Q}\sigma}/E_F$ . However, the modulations given above have equal second-order transition temperatures, and immediately below the transition point other harmonics are damped by a factor of  $\Delta_{\mathbf{Q}\sigma}/T$  or  $\Delta_{\mathbf{Q}\sigma}/H$ .

To obtain the leading-order expressions for the fields  $\Delta_{\mathbf{Q}\sigma}$  in the vicinity of the transition line, we rewrite the equations (5.12) in the "matrix" form

$$GG_0^{-1} - GF = I, \quad (5.35)$$

where

$$\begin{aligned} G(\mathbf{k}', \mathbf{k}) &= g_\sigma(\mathbf{k}', \mathbf{k}), \quad G_0(\mathbf{k}', \mathbf{k}) = \frac{\delta(\mathbf{k}', \mathbf{k})}{i\omega - \varepsilon_\sigma(\mathbf{k})}, \\ F(\mathbf{k}', \mathbf{k}) &= \sum_{\mathbf{Q}} \Delta_{\mathbf{Q}\sigma} \delta(\mathbf{k}', \mathbf{k} - \mathbf{Q}). \end{aligned} \quad (5.36)$$

The solution is

$$G = G_0 + G_0 F G_0 + G_0 F G_0 F G_0 + \dots \quad (5.37)$$

Omitting the contribution from the virtual states with the momentum  $|k_x| \gtrsim 2k_F$ , we obtain the consistency equation to the third order in  $\Delta_{\mathbf{Q}\sigma}$ :

$$\begin{aligned} \Delta_{\mathbf{Q}_0+\mathbf{q},\sigma'} &= T \sum_{\mathbf{k}\omega\sigma} (U_c - U_s \sigma \sigma') \times \left[ \frac{\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma}}{(i\omega - \varepsilon_\sigma(\mathbf{k}))(i\omega - \varepsilon_\sigma(\mathbf{k} - \mathbf{Q}_0 - \mathbf{q}))} + \right. \\ &+ \frac{\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma} |\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma}|^2}{(i\omega - \varepsilon_\sigma(\mathbf{k}))(i\omega - \varepsilon_\sigma(\mathbf{k} - \mathbf{Q}_0 - \mathbf{q}))(i\omega - \varepsilon_\sigma(\mathbf{k}))(i\omega - \varepsilon_\sigma(\mathbf{k} - \mathbf{Q}_0 - \mathbf{q}))} + \\ &+ \frac{\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma} |\Delta_{\mathbf{Q}_0-\mathbf{q},\sigma}|^2}{(i\omega - \varepsilon_\sigma(\mathbf{k}))(i\omega - \varepsilon_\sigma(\mathbf{k} - \mathbf{Q}_0 - \mathbf{q}))(i\omega - \varepsilon_\sigma(\mathbf{k} - 2\mathbf{q}))(i\omega - \varepsilon_\sigma(\mathbf{k} - \mathbf{Q}_0 - \mathbf{q}))} + \\ &\left. + \frac{\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma} |\Delta_{\mathbf{Q}_0-\mathbf{q},\sigma}|^2}{(i\omega - \varepsilon_\sigma(\mathbf{k}))(i\omega - \varepsilon_\sigma(\mathbf{k} - \mathbf{Q}_0 + \mathbf{q}))(i\omega - \varepsilon_\sigma(\mathbf{k}))(i\omega - \varepsilon_\sigma(\mathbf{k} - \mathbf{Q}_0 - \mathbf{q}))} \right]. \end{aligned} \quad (5.38)$$

The equation for the other two harmonics may be obtained via a substitution  $\mathbf{q} \rightarrow -\mathbf{q}$ .

Performing the integration over  $k_x$ , we arrive at

$$\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma'} = \sum_{\sigma} \frac{1 + \nu \sigma \sigma'}{2} \left( K_\sigma^{(1)} - |\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma}|^2 K_\sigma^{(3)c} - |\Delta_{\mathbf{Q}_0-\mathbf{q},\sigma}|^2 K_\sigma^{(3)d} \right) \Delta_{\mathbf{Q}_0+\mathbf{q},\sigma} \quad (5.39)$$

(the superscript “d” stands for “double cosine”) with  $K^{(1)}$  and  $K^{(3)c}$  given by (5.31) and (5.32), and

$$\begin{aligned} K_\sigma^{(3)d} &= \pi \nu_F |U_c| T \sum_{\omega} \left\langle \frac{(\omega - i\sigma H) \operatorname{sign} \omega}{[(\omega + i\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q}))(\omega - i\varepsilon_\sigma^-(\mathbf{k}, \mathbf{k} - \mathbf{Q}))]^2} \right\rangle_{k_y} = \\ &= -\frac{\nu_F |U_c|}{4\pi \hbar \nu_F q_x T} \left\langle \left[ \operatorname{Im} \psi' \left( \frac{1}{2} + \frac{i\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})}{2\pi T} \right) + \operatorname{Im} \psi' \left( \frac{1}{2} + \frac{i\varepsilon_\sigma^-(\mathbf{k}, \mathbf{k} - \mathbf{Q})}{2\pi T} \right) \right] \right\rangle_{k_y}. \end{aligned} \quad (5.40)$$

The function  $K_\sigma^{(3)d}$  does not depend on  $\sigma$  and in what follows we omit this subscript.

At only a longitudinal shift of the CDW wave vector ( $q_y = 0$ ) the  $k_y$ -dependence of  $\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})$  is symmetric, and the functions  $K_\sigma^{(1)}$ ,  $K_\sigma^{(3)c}$  and  $K^{(3)d}$  (being dependent only on the shift wave vector  $q_x$ ) possess the symmetry:

$$\begin{aligned} K_\sigma^{(1)}(q_x) &= K_{-\sigma}^{(1)}(-q_x), \\ K_\sigma^{(3)c}(q_x) &= K_{-\sigma}^{(3)c}(-q_x), \\ K^{(3)d}(-q_x) &= K^{(3)d}(q_x). \end{aligned} \quad (5.41)$$

This symmetry follows directly from Eqs. (5.31), (5.32) and (5.40) and the properties of digamma function:  $\operatorname{Re} \psi^{(n)}(a + ib) = \operatorname{Re} \psi^{(n)}(a - ib)$ ,  $\operatorname{Im} \psi^{(n)}(a + ib) = -\operatorname{Im} \psi^{(n)}(a - ib)$ .

This symmetry is not a consequence of the expansion in powers of  $\Delta_\sigma$ : the consistency equation (5.22) also does not change under the transformation

$$q_x \rightarrow -q_x, \sigma \rightarrow -\sigma. \quad (5.42)$$

The ratio  $|\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma}|^2 : |\Delta_{\mathbf{Q}_0+\mathbf{q},-\sigma}|^2$  near the transition point is still given by (5.33) accounting for the property (5.41). Hence,  $|\Delta_{\mathbf{Q}_0-\mathbf{q},\sigma}|^2 : |\Delta_{\mathbf{Q}_0-\mathbf{q},-\sigma}|^2 = |\Delta_{\mathbf{Q}_0+\mathbf{q},-\sigma}|^2 : |\Delta_{\mathbf{Q}_0+\mathbf{q},\sigma}|^2 \equiv \alpha$ . As in derivation of (5.34) one may rewrite the system of equations (5.39) on  $\Delta$  as two equations (5.28) with replacement  $K_\sigma^{(1)}(\mathbf{Q}_0 \pm \mathbf{q}_x) \rightarrow K_{\pm\sigma}^{(1)} - |\Delta_{\mathbf{Q}_0 \pm \mathbf{q},\sigma}|^2 K_{\pm\sigma}^{(3)c} - |\Delta_{\mathbf{Q}_0 \mp \mathbf{q},\sigma}|^2 K^{(3)d}$ .

Defining the ratio  $\beta \equiv |\Delta_{\mathbf{Q}_0-\mathbf{q},-}|^2 : |\Delta_{\mathbf{Q}_0+\mathbf{q},+}|^2$  we can extend the derivation of (5.34) to get

$$|\Delta_{\mathbf{Q}_0+\mathbf{q},+}|^2 = \frac{\left(\nu K_+^{(1)} K_-^{(1)} - \frac{\nu+1}{2}(K_+^{(1)} + K_-^{(1)}) + 1\right)}{\left(K_+^{(3)c} + \alpha\beta K^{(3)d}\right) \left(\nu K_-^{(1)} - \frac{\nu+1}{2}\right) + \left(\alpha K_-^{(3)c} + \beta K^{(3)d}\right) \left(\nu K_+^{(1)} - \frac{\nu+1}{2}\right)}, \quad (5.43)$$

and from the  $\mathbf{Q}_0 - \mathbf{q}$  part of (5.39)

$$|\Delta_{\mathbf{Q}_0-\mathbf{q},-}|^2 = \frac{\beta \left(\nu K_+^{(1)} K_-^{(1)} - \frac{\nu+1}{2}(K_+^{(1)} + K_-^{(1)}) + 1\right)}{\left(\beta K_+^{(3)c} + \alpha K^{(3)d}\right) \left(\nu K_-^{(1)} - \frac{\nu+1}{2}\right) + \left(\alpha\beta K_-^{(3)c} + K^{(3)d}\right) \left(\nu K_+^{(1)} - \frac{\nu+1}{2}\right)}. \quad (5.44)$$

Dividing (5.43) by (5.44) we obtain a linear equation on  $\beta$ :

$$\frac{\left(\beta K_+^{(3)c} + \alpha K^{(3)d}\right) \left(\nu K_-^{(1)} - \frac{\nu+1}{2}\right) + \left(\alpha\beta K_-^{(3)c} + K^{(3)d}\right) \left(\nu K_+^{(1)} - \frac{\nu+1}{2}\right)}{\left(K_+^{(3)c} + \alpha\beta K^{(3)d}\right) \left(\nu K_-^{(1)} - \frac{\nu+1}{2}\right) + \left(\alpha K_-^{(3)c} + \beta K^{(3)d}\right) \left(\nu K_+^{(1)} - \frac{\nu+1}{2}\right)} = 1, \quad (5.45)$$

which gives  $\beta = 1$ . The values  $\beta = 0$  or  $\beta = \infty$  in (5.43) and (5.44) correspond to cosine phase. Hence, in the double cosine phase the symmetry (5.42) is not broken, while the transition to cosine phase breaks this symmetry.

### 1.5.5 Free energy of cosine and double-cosine phases

One can easily write down the expressions for the free energies of cosine and double-cosine phases valid to the second order in the energy gap  $|\Delta_\sigma|$ .

From (5.11) we have for the free energy

$$F_{\text{CDW}} - F_n = \frac{1}{2} \sum_{\mathbf{Q}\sigma} D_{-\mathbf{Q}\sigma} \Delta_{\mathbf{Q}\sigma}. \quad (5.46)$$

To the second order in  $\Delta_{\mathbf{Q}\sigma}$  this rewrites as

$$\begin{aligned} F_{\text{CDW}} - F_n &= -\frac{\pi\nu_F T}{4} \sum_{\mathbf{Q}\omega\sigma} \left\langle \frac{\Delta_{\mathbf{Q}\sigma} \Delta_{-\mathbf{Q}\sigma} \text{sign } \omega}{\omega + i\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q})} \right\rangle_{k_y} \\ &= -\frac{1}{4|U_c|} \sum_{\mathbf{Q}\sigma} K_\sigma^{(1)} |\Delta_{\mathbf{Q}\sigma}|^2. \end{aligned} \quad (5.47)$$

The phase with the most negative r.h.s of (5.46) wins; positive values of the r.h.s. correspond to first-order transitions. From Eq. (5.47) using Eq. (5.41) we have

$$F_c - F_n = -\frac{K_+^{(1)} + \alpha K_-^{(1)}}{2|U_c|} |\Delta_{\mathbf{Q}_0+\mathbf{q}_{\mathbf{x},+}}|^2 \quad (5.48)$$

and

$$F_{2c} - F_n = -\frac{K_+^{(1)} + \alpha K_-^{(1)}}{|U_c|} |\Delta_{\mathbf{Q}_0 + \mathbf{q}_{x,+}}^{2c}|^2. \quad (5.49)$$

The quantity  $K_+^{(1)} + \alpha K_-^{(1)}$  depends on  $\mathbf{q}$  but is always positive near the metal-CDW transition line  $T_c(H)$  where this transition is of the second kind.

Since  $\text{CDW}_c$  and  $\text{CDW}_{2c}$  phases have the same transition temperature, the only way to determine which phase takes place is to compare their free energies, that near the transition line  $T_c(H)$  are given by the formulas (5.48) and (5.49). The double cosine phase wins if the ratio

$$r_F \equiv \frac{F_{2c} - F_n}{(F_c - F_n)} > 1, \quad (5.50)$$

where the values of the functions  $F_{2c}(T, H, \mathbf{q})$  and  $F_c(T, H, \mathbf{q})$  must be taken at the optimal value of the wave vector  $\mathbf{q}$ , that should be found by minimization of these free energy functions at each point of  $T, H$  phase diagram and for each of the two phases ( $\text{CDW}_c$  and  $\text{CDW}_{2c}$ ) separately.

Below the transition temperature the optimal shift vectors  $\mathbf{q}$  could be different for the cosine and double cosine phases (they minimize different free energy functions  $F_c$  and  $F_{2c}$ ). However, at the transition temperature  $T_c(H)$  the  $\text{CDW}_c$  and  $\text{CDW}_{2c}$  phases have the same optimal value of  $\mathbf{q}$  which is determined by the minimum of the left-hand side of Eq. (5.28). Hence, on the transition line  $T_c(H)$  Eq. (5.50) simplifies to

$$r_F = \frac{2|\Delta_{\mathbf{Q}_0 + \mathbf{q}_{x,+}}^{2c}|^2}{|\Delta_{\mathbf{Q}_0 + \mathbf{q}_{x,+}}^c|^2} > 1, \quad (5.51)$$

or after substitution of (5.34) and (5.43)

$$r_F = \frac{2 \left[ K_+^{(3)c} \left( \nu K_-^{(1)} - \frac{\nu+1}{2} \right) + \alpha K_-^{(3)c} \left( \nu K_+^{(1)} - \frac{\nu+1}{2} \right) \right]}{\left( K_+^{(3)c} + \alpha K_-^{(3)d} \right) \left( \nu K_-^{(1)} - \frac{\nu+1}{2} \right) + \left( \alpha K_-^{(3)c} + K_+^{(3)d} \right) \left( \nu K_+^{(1)} - \frac{\nu+1}{2} \right)} > 1. \quad (5.52)$$

These formulas will be used later to determine the phase diagram.

## 2. The phase diagram

In this section we only consider the longitudinal modulation ( $q_y = 0$ ) of the CDW wave vector. The CDW phase with  $q_y \neq 0$  ( $\text{CDW}_y$  phase) may appear for certain dispersion functions  $t_\perp(\mathbf{k}_\perp)$ . However, for the tight-binding model (5.2) the  $\text{CDW}_y$  phase is not expected to take place near the transition temperature. Near the critical pressure (when  $t'_b \approx t_b^*$  and the SDW-metal phase transition takes place) in zero field the  $\text{SDW}_2$  phase was predicted using the susceptibility calculation from the normal state.[40] However, this phase was predicted only near  $T = 0$ . Moreover, the absolute instability line calculated in [[40]] does not give the actual transition line at low temperature since this transition takes place as the first-order phase transition.

To be more sure that we can disregard appearance of the  $\text{CDW}_y$  phase in our consideration near the transition temperature we performed the calculation of the optimal shift vector  $\mathbf{q}$  at the transition line  $T_c(H)$  in a large range of parameters. We swept the following three-dimensional range of parameters:  $-0.3 \leq \nu \leq 0.9$ ,  $0 \leq t'_b/t_b^* < 1$  and  $0 \leq \mu_B H \leq 2T_{c0}$  and have not found that the shift of  $Q_y$  leads to higher  $T_c$  anywhere inside this range.

This does not contradict the prediction of the  $CDW_y$  [24], since according to this paper the  $CDW_y$  phase appears at  $\nu < 0$  essentially due to the orbital effect of the magnetic field (see Fig. 7d of Ref. [[24]]). Below we will only consider the range  $0 \leq \nu < 1$  which is close to the experimental situation in organic metals and disregard the appearance of  $CDW_y$  phase.

### 2.1 Transition line $T_c(H)$ and the normal- $CDW_0$ - $CDW_x$ tricritical point

The normal-to- $CDW$  phase transition line is given by Eq. (5.28) irrespective of to which  $CDW$  phase this transition occurs. The shift of the wave vector  $q_x$  which enters Eq. (5.28) corresponds to the maximum value of  $T_c$  at given magnetic field  $H$ . The behavior of  $T_c(H)$  and  $q_x(H)$  at the transition temperature has been analyzed in Ref. [[24]]. Diagonalization of the  $M_3, M_4$  part of the susceptibility matrix (6) in Ref. [[24]] corresponds to the diagonalization of the matrix in Eq. (5.28). On the transition line one of the eigenvalues of these matrices is zero. In the case of perfect nesting equation (5.28) gives the same result for  $T_c(H)$  as Eq. (17) of Ref. [[24]] (see fig. 7(a) of [[24]]).

For  $q_x = 0$  ( $CDW_0$  phase)  $K_+^{(1)} = K_-^{(1)}$  and Eq. (5.28) simplifies to  $K_+^{(1)} = 1$ , that gives for the transition temperature  $T_c = T_c(H)$  the well-known equation

$$\ln \left( \frac{T_c(H, t'_b)}{T_{c0}} \right) = \psi \left( \frac{1}{2} \right) - \left\langle \operatorname{Re} \psi \left( \frac{1}{2} - \frac{i\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q}_0)}{2\pi T_c} \right) \right\rangle_{k_y}, \quad (5.53)$$

where

$$\varepsilon_\sigma^+(\mathbf{k}, \mathbf{k} - \mathbf{Q}_0) = -2t'_b \cos(2k_y b) - \sigma H. \quad (5.54)$$

In the Eq. (5.53) one can take either of the values  $\sigma = \pm 1$  since this equation does not depend on the sign of  $\sigma$ . The transition line  $T_c(H)$  from normal to  $CDW_0$  phase does not depend on the value of  $\nu$ .

The tricritical point (where the  $CDW_0$  and  $CDW_x$  phases have the same transition temperature) is given by equation  $\partial^2 D(\nu, T, H, q_x) / \partial q_x^2 = 0$ , where  $D(\nu, T, H, q_x)$  is the left-hand side of Eq. (5.28). At this point  $q_x = 0, \Rightarrow K_+^{(1)} = K_-^{(1)} = 1$  and  $\partial^n K_+^{(1)} / \partial q_x^n = (-1)^n \partial^n K_-^{(1)} / \partial q_x^n$ , and the equation for the tricritical point becomes

$$\frac{\partial^2 K^{(1)}}{(\partial q_x)^2} = -\frac{2\nu}{1-\nu} \left( \frac{\partial K^{(1)}}{\partial q_x} \right)^2. \quad (5.55)$$

Substituting (5.31) and (5.32) this equation rewrites as

$$\begin{aligned} & - \left\langle \operatorname{Re} \psi'' \left( \frac{1}{2} - \frac{i\varepsilon_\uparrow^+(\mathbf{k}, \mathbf{k} - \mathbf{Q}_0)}{2\pi T_c} \right) \right\rangle_{k_y} = \\ & \eta \left[ \left\langle \operatorname{Im} \psi' \left( \frac{1}{2} - \frac{i\varepsilon_\uparrow^+(\mathbf{k}, \mathbf{k} - \mathbf{Q}_0)}{2\pi T_c} \right) \right\rangle_{k_y} \right]^2, \end{aligned} \quad (5.56)$$

where

$$\eta = 2\nu\nu_F |U_c| / (1 - \nu). \quad (5.57)$$

Together with Eq. (5.53) this equation allows to find the tricritical point.



### 2.2 $T_c(H)$ transition line at perfect nesting

The case of perfect nesting ( $t'_b = 0$ ) is equivalent to the strictly 1D case in many mathematical aspects. However, the 3D nature of the compound (its energy spectrum, e-e interaction and lattice elasticity) preserves, and the mean-field approach still works. The analysis in this subsection (at  $t'_b = 0$ ) is simple and performed analytically. It helps to understand some qualitative features of the CDW phase diagram in magnetic field.

At perfect nesting and purely longitudinal modulation of CDW ( $q_y = 0$ ) the expressions for  $K_{\sigma}^{(1)}$ ,  $K^{(3)d}$  and  $K_{\sigma}^{(3)c}$  simplify to

$$K_{\sigma 0}^{(1)} = 1 + \nu_F |U_c| \left[ \ln \frac{T_{c0}}{T} + \psi \left( \frac{1}{2} \right) - \operatorname{Re} \psi \left( \frac{1}{2} + \frac{ih_{\sigma}}{2\pi T} \right) \right], \quad (5.58)$$

$$K_0^{(3)d} = -\frac{\nu_F |U_c|}{4\pi \hbar v_F q_x T} \left[ \operatorname{Im} \psi' \left( \frac{1}{2} + \frac{ih_{\sigma}}{2\pi T} \right) + \operatorname{Im} \psi' \left( \frac{1}{2} + \frac{ih_{-\sigma}}{2\pi T} \right) \right], \quad (5.59)$$

and

$$K_{\sigma 0}^{(3)c} = -\frac{\nu_F |U_c|}{16\pi^2 T^2} \operatorname{Re} \psi'' \left( \frac{1}{2} + \frac{ih_{\sigma}}{2\pi T} \right), \quad (5.60)$$

where

$$h_{\sigma} = \frac{\hbar v_F q_x}{2} - \sigma H. \quad (5.61)$$

At  $q_x = 0$ , Eq. (5.59) even more simplifies:

$$K_{00}^{(3)d} = -\frac{\nu_F |U_c|}{8\pi^2 T^2} \operatorname{Re} \psi'' \left( \frac{1}{2} - \frac{iH}{2\pi T} \right) = 2K_{\sigma 00}^{(3)c}. \quad (5.62)$$

One can now derive a simple formula for the transition line  $T_c(H)$  in the high field limit  $H \gg T_c(H)$ . As  $H \rightarrow \infty$ ,  $h_{\sigma}$  in (5.61) goes to zero for one spin component and to  $-2\sigma H$  for the other. Using the limit expansion of the digamma function,  $\operatorname{Re} \psi(1/2 + ix) = \ln x + O(1/x)$ ,  $x \rightarrow \infty$ , from (5.58) we have at  $H/\pi T_c(H) \gg 1$

$$\begin{aligned} K_+^{(1)} &\approx 1 + \nu_F |U_c| \ln(T_{c0}/T) \\ K_-^{(1)} &\approx 1 + \nu_F |U_c| \ln(\pi T_{c0}/4\gamma H). \end{aligned} \quad (5.63)$$

Equation (5.28) rewrites as

$$K_+^1 = \frac{(1 + \nu)K_-^{(1)}/2 - 1}{\nu K_-^{(1)} - (1 + \nu)/2},$$

that in the limit  $H/\pi T_c(H) \gg 1$  becomes

$$\ln(T_{c0}/T) = \frac{\ln(4\gamma H/\pi T_{c0})}{\eta \ln(4\gamma H/\pi T_{c0}) + 1}. \quad (5.64)$$

At  $\eta \ln(4\gamma H/\pi T_{c0}) \gg 1$  this simplifies to

$$T_c(H \rightarrow \infty) = T_{c0} \exp(-1/\eta). \quad (5.65)$$

In the intermediate interval  $\pi T_c(H) \ll 4\gamma H \ll \pi T_{c0} \exp(1/\eta)$  we get

$$T_c(H) \approx \pi T_{c0}^2/(4\gamma H). \quad (5.66)$$

Hence, at perfect nesting in the limit  $H \rightarrow \infty$  the transition temperature tends to a finite value which is natural since it corresponds to the CDW instability for only one spin component.

However, the behavior of  $T_c(H)$  given by formulas (5.64), (5.65) and (5.66) strongly changes at finite "antinesting" term  $t'_b$  (see below).

### 2.3 Metal-CDW transition lines and tricritical points at finite $t'_b$

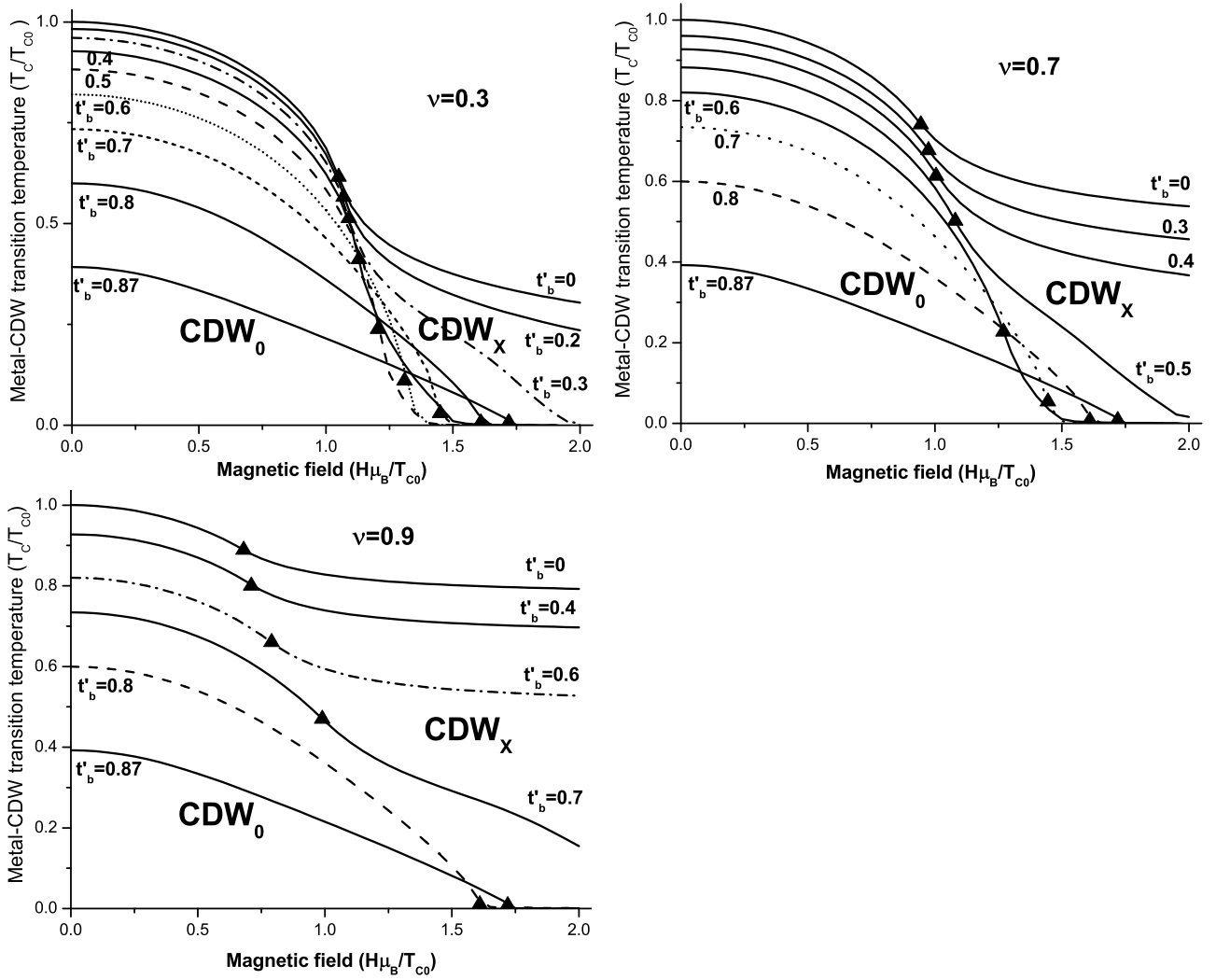


Figure 47: Transition lines  $T_c(H)$  at different values of  $t'_b$  at three different values of the coupling constant ratio  $\nu$ . The values of  $t'_b$  is given on the figures in units of  $T_{c0}$ ; for the tight-binding dispersion (5.2) the critical value of  $t'_b$  at which CDW disappears without magnetic field is  $t'^*_b = \Delta_0/2 \approx 0.88T_{c0}$ . The triangles show the tricritical points Normal- $CDW_0$ - $CDW_x$ .

Using the Eq. (5.28) and Eq. (5.56) we calculated the transition lines  $T_c(H)$  and the tricritical points at different values of  $t'_b$  and  $\nu$ . The results are shown in Fig. 47. From this figure we see that the transition line  $T_c(H)$  depends strongly on the both parameters  $t'_b$  and  $\nu$ , and one cannot determine the value of  $t'_b$  from the  $T_c(H)$  transition line if one does not know the value of  $\nu$ . Hence, to determine experimentally what values take  $t'_b$  and  $\nu$  in a particular compound, one has to perform two independent tests. For example, one can study the pressure dependence of the  $T_c(H)$  lines at two different tilt angles of magnetic field: the first measurement when the magnetic field is parallel to the conducting  $x$ - $y$  plane and the formulas derived above in neglect of orbital effects are valid, the second measurement in the very strong magnetic field perpendicular to the conducting  $x$ - $y$  plane, when the orbital effects are so strong that the one-dimensionization of the electron dispersion takes place and the formulas for perfect nesting (see sec. 3.2) become approximately valid. One can also determine the value of  $t'_b$  independently from the transport measurement and then compare it to the critical value of  $t'^*_b = \Delta_0/2 \approx 0.88T_{c0}$  at which the CDW state is damped without magnetic field.

There is one interesting common feature on all diagrams in Figs. 47. At each value of  $\nu$  there

is a critical value of  $t'_b < t_b^*$  above which the  $\text{CDW}_x$  phase disappears, i.e. as magnetic field increases the transition from  $\text{CDW}_0$  to metal state instead of the  $\text{CDW}_x$  phase takes place up to the lowest temperature. This critical value of  $t'_b$  may shift due to the "one-dimensionization" effect of magnetic field.

We have also checked if the region of cosine phase increases considerably after finite "antinetting" term is taken into account. The results of this study at the transition line are given in Fig. 48. These results indicate that the region of cosine phase does not increase considerably at finite  $t'_b$  for the tight-binding dispersion.

## 2.4 Cosine and double-cosine phases at perfect nesting

To determine which of the two phases ( $\text{CDW}_c$  or  $\text{CDW}_{2c}$ ) wins we have to compare their free energies given in section 2.5.

On the transition line this reduces to the evaluation of the ratio (5.52). At the tricritical point (when  $q_x = 0$ ) this ratio can be easily evaluated using Eq. (5.62):

$$r_{F\text{triple}} = 2/3 < 1. \quad (5.67)$$

Hence, near the tricritical point the cosine phase wins. In the limit of high field ( $H \gg T_c$ ) one has  $\hbar v_F q_x \rightarrow 2H$  and  $h_\sigma \rightarrow (1 - \sigma)H$ . In this limit  $K_+^{(3)c} \approx -(\nu_F |U_c| / 16\pi^2 T^2) \text{Re } \psi''(1/2) \gg K_-^{(3)c}, K^{(3)d}$ , and from (5.52) we get

$$r_F \rightarrow 2 \text{ at } H/T_c \rightarrow \infty. \quad (5.68)$$

Hence, at high magnetic field the double cosine phase wins. These two simple estimates suggest that the cosine and double-cosine phases both appear on the phase diagram.

The boundary between  $\text{CDW}_c$  and  $\text{CDW}_{2c}$  phases on the transition line  $T_c(H)$  is given by the equation  $r_F(T, H, \nu) = 1$ , which rewrites as

$$\begin{aligned} & \left( K_+^{(3)c} - \alpha K^{(3)d} \right) \left( \nu K_-^{(1)} - \frac{\nu + 1}{2} \right) = \\ & = \left( K^{(3)d} - \alpha K_-^{(3)c} \right) \left( \nu K_+^{(1)} - \frac{\nu + 1}{2} \right). \end{aligned} \quad (5.69)$$

This equation is valid also for nonzero  $t'_b$  and together with Eq. (5.28) allows to determine the second tricritical point  $(H_{c2}(\nu), T_c(H_{c2}))$  where the normal,  $\text{CDW}_c$  and  $\text{CDW}_{2c}$  phases meet.

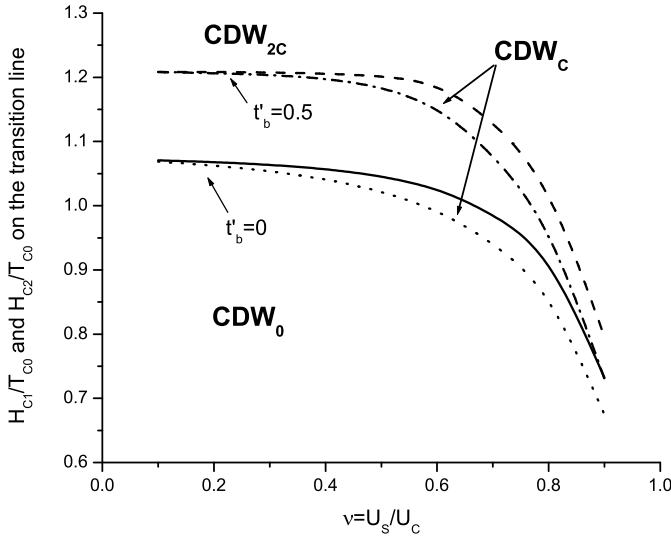


Figure 48: The magnetic field values  $H_{c1}(\nu)$  and  $H_{c2}(\nu)$  at two tricritical points Normal- $\text{CDW}_0$ - $\text{CDW}_c$  and Normal- $\text{CDW}_c$ - $\text{CDW}_{2c}$  correspondingly as function of coupling constant ratio  $\nu = U_s/U_c$  for two different values of  $t'_b$ . The solid and dot lines denote  $H_{c2}(\nu)$  and  $H_{c1}(\nu)$  at perfect nesting ( $t'_b = 0$ ) while the dash and dash-dot lines denote  $H_{c2}(\nu)$  and  $H_{c1}(\nu)$  at finite value of the "ant nesting" term ( $t'_b = 0.5T_{c0}$ ). The cosine phase exists in the areas between  $H_{c1}(\nu)$  and  $H_{c2}(\nu)$  lines at the same value of  $t'_b$ .

In Fig. 48 we plot two triple-point values  $H_{c1}(\nu)$  and  $H_{c2}(\nu)$  of magnetic field which are given by equations (5.56) and (5.69) correspondingly (together with Eq. (5.28)) at two different values of  $t'_b = 0$  and  $t'_b = 0.5T_{c0}$ .  $H_{c1}(\nu)$  and  $H_{c2}(\nu)$  determine the tricritical points Normal- $\text{CDW}_0$ - $\text{CDW}_c$  and Normal- $\text{CDW}_c$ - $\text{CDW}_{2c}$ . From this figure we see that the range of the cosine phase is very narrow (the difference  $H_{c2} - H_{c1}$  does not exceed 5% of  $H_{c1}$ ) and strongly depends on  $\nu$ .

At perfect nesting two tricritical points  $H_{c1}$  and  $H_{c2}$  coincide at  $\nu = 0$ . At this "double-triple" point the denominator in formulas (5.34) and (5.43) is zero. This means, that one should expand up to the higher orders in  $|\Delta^2|$  than fourth and that at this point the critical fluctuations are very strong.

From Fig. 48 we see that (i) on the phase diagram of CDW in magnetic field there are at least 3 different CDW states:  $\text{CDW}_0$ ,  $\text{CDW}_c$  and  $\text{CDW}_{2c}$ , (ii) at the transition line  $T_c(H)$  the region of the  $\text{CDW}_c$  phase is quite narrow for various values of  $t'_b$  and is usually sandwiched between the  $\text{CDW}_0$  and  $\text{CDW}_{2c}$  phases as the magnetic field increases.

To analyze how this picture evolves below the transition line  $T_c(H)$  we performed a numerical calculation of the free energies of  $\text{CDW}_c$  and  $\text{CDW}_{2c}$  phases using formulas (5.48), (5.49), (5.43), (5.34). These formulas are valid if  $\Delta \ll \pi T, H$ , that covers a narrow region below  $T_c(H)$ .

The computation performed in the case of perfect nesting shows that the region of the  $\text{CDW}_c$  phase becomes even more narrow as the temperature decreases and disappears at  $T \approx 0.95T_c(H)$  (see Fig. 49). Hence, the cosine phase at perfect nesting exists only in a very small region of the phase diagram where it is strongly smeared by the critical fluctuations.

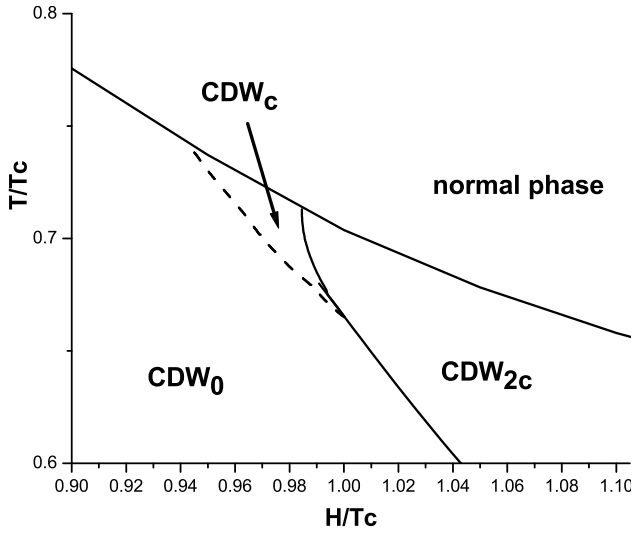


Figure 49: Phase diagram in  $H - T$  coordinates at  $\nu = 0.7$  at perfect nesting near the tricritical point. The  $CDW_c$  region disappears rapidly as temperature decreases below  $T_c$ .

The similar phase diagram appears in surface superconductors in parallel magnetic field.[41] The cosine and double-cosine CDW phases correspond to the helical and cosine (stripe) superconducting phases respectively, and the total  $CDW_x$  phase corresponds to the non-uniform superconducting LOFF state.[8, 9] However, our situation differs from that in Ref. [[41]], where the Rashba term in the electron dispersion relation plays an important role in obtaining the phase diagram similar to that on Fig. 49. The number of harmonics and coupling constant in our case is twice larger than that in Ref. [[41]]. However, the analogy between CDW and surface superconductors in magnetic field is rather deep (see the discussion section).

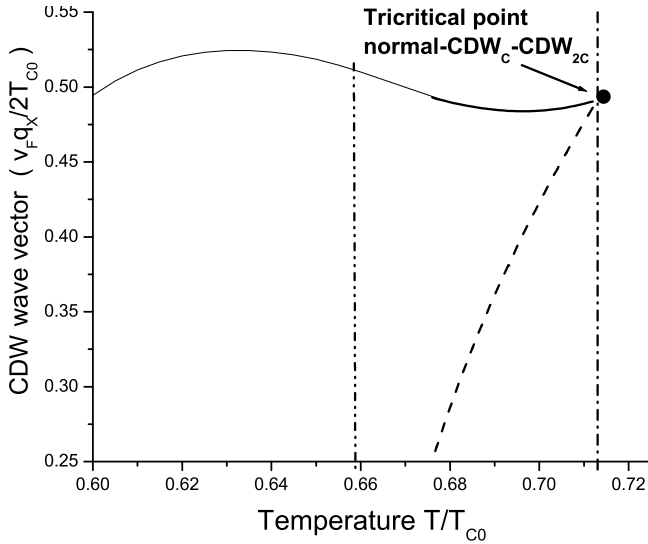


Figure 50: The shifts  $q_x$  of wave vectors of the phases  $CDW_{2c}$  (solid line) and  $CDW_c$  (dash line) along the  $CDW_c$ - $CDW_{2c}$  transition line (at  $0.713 > T/T_{c0} > 0.66$ ) or  $CDW_0$ - $CDW_{2c}$  transition line (at  $T/T_{c0} < 0.66$ ) as function of temperature. These shifts coincide only at the normal- $CDW_c$ - $CDW_{2c}$  tricritical point at  $T/T_{c0} = 0.713$ . The difference between the solid and dash lines gives the jump of the CDW wave vector on the  $CDW_c$ - $CDW_{2c}$  transition line that makes this transition of the first kind and leads to hysteresis. The dash-dot line stands for the normal-CDW transition temperature, and the dash-double-dot line shows the temperature of  $CDW_0$ - $CDW_c$ - $CDW_{2c}$  tricritical point.

On the transition line  $CDW_c$ - $CDW_{2c}$  the energy gaps and the optimal shifts  $q_x$  of the CDW

wave vectors of these two phases become substantially different as the temperature decreases below  $T_c$  (see Figs. 50). This means that the transition from  $\text{CDW}_c$  to  $\text{CDW}_{2c}$  is of the first order. The transition line between  $\text{CDW}_0$  and  $\text{CDW}_{2c}$  is also of the first order, and the  $\text{CDW}_{2c}$  nesting vector differs from  $Q_0 = 2k_F$  already on the transition line (see Fig. 50). This fact explains the huge hysteresis in magnetization and magnetoresistance observed in  $\alpha\text{-(BEDT-TTF)}_2\text{KHg(SCN)}_4$  at the kink transition.[29, 30, 32] The first-order phase transition driven by magnetic field is usually accompanied by hysteresis. The jump of the nesting vector on the  $\text{CDW}_0\text{-CDW}_{2c}$  transition line itself leads to huge hysteresis. The CDW and its wave vector are pinned by the impurities and crystal imperfections.[1] As magnetic field increases through the kink transition point  $H = H_c(T)$ , the jump of the CDW wave vector forces the CDW condensate to move. Due to the pinning this motion is hysteretic since at each time moment the CDW condensate finds some local minimum of the impurity potential. The larger jump of the CDW wave vector at the kink transition the greater the hysteresis is. This jump decreases with increasing temperature and comes to zero on the transition line  $T_c(H)$ . The hysteresis shows a similar temperature dependence. The hysteresis reduces with heating also due to the thermal activation processes which reduce the pinning.

### 3. Discussion

The electron charge and spin densities and the lattice distortion differ considerably in the cosine  $\text{CDW}_c$  and double-cosine  $\text{CDW}_{2c}$  phases that can help to distinguish these two states experimentally. In the double-cosine phase the charge modulation is the sum of two cosine distortions that leads to the beats of the charge density wave:

$$\begin{aligned}\rho_C^{2c}(x) &= \nu_F(\Delta_+ + \Delta_-)\{\cos[(Q_0 + q_x)x + \phi_1] \\ &\quad + \cos[(Q_0 - q_x)x + \phi_2]\} \\ &= 2\nu_F(\Delta_+ + \Delta_-)\cos[Q_0x + (\phi_1 + \phi_2)/2] \\ &\quad \times \cos[q_x x + (\phi_1 - \phi_2)/2].\end{aligned}\tag{5.70}$$

The phase shifts  $\phi_1$  and  $\phi_2$  may depend on coordinate because of the pinning of CDW by impurities. Usually,  $\Delta_+(Q)$  and  $\Delta_-(Q)$  in  $\text{CDW}_{2c}$  state differ substantially, and the charge density wave in magnetic field is accompanied by a spin density wave. The spin-density modulation in the double-cosine phase is given by

$$\begin{aligned}\rho_S^{2c}(x) &= \nu_F(\Delta_+ - \Delta_-)\{\cos[(Q_0 + q_x)x + \phi_1] \\ &\quad - \cos[(Q_0 - q_x)x + \phi_2]\} \\ &= -2\nu_F(\Delta_+ - \Delta_-)\sin[Q_0x + (\phi_1 + \phi_2)/2] \\ &\quad \times \sin[q_x x + (\phi_1 - \phi_2)/2].\end{aligned}\tag{5.71}$$

In the cosine phase both charge and spin densities have one-cosine modulations,

$$\begin{aligned}\rho_C^c(x) &= \nu_F(\Delta_+ + \Delta_-)\cos[(Q_0 + q_x)x + \phi_1] \\ \rho_S^c(x) &= \nu_F(\Delta_+ - \Delta_-)\cos[(Q_0 + q_x)x + \phi_1].\end{aligned}\tag{5.72}$$

Thus, the charge modulations in the cosine and double-cosine phases can be experimentally distinguished by X-ray or Raman scattering, and the spin modulation can be detected by muon or neutron scattering experiments.

The energy spectrum in the cosine and double-cosine phases also differ strongly. In the cosine phase the energy spectrum (5.19) is asymmetric with respect to the spin components since the energy gaps  $\Delta_\sigma$  differ for two spin components  $\sigma$ . This means that under external

electric field the spin current is produced in addition to the charge current since the charge is transferred by electrons with predominantly one spin component. The degree of current polarization depends on the shift of CDW wave vector, and, hence, can be controlled by the external magnetic field. This property of the cosine phase may find applications in spintronics. The double-cosine phase is symmetric in spin components, and its energy spectrum has at least two gaps for each spin component. The symmetry (5.42) preserves in the double-cosine phase while in the cosine phase it is spontaneously broken. Thus, the one-cosine and double-cosine CDW states differ substantially in their thermodynamic, transport and optical properties.

At low temperature  $T \ll T_c$  the expansion in (5.37) is not applicable and the additional harmonics  $\mathbf{Q} + (2n+1)\mathbf{q}_x$  with integer  $n$  appear in the double-cosine solution of the consistency equation. These harmonics make the charge modulation in the  $\text{CDW}_{2c}$  phase at low temperature essentially nonsinusoidal in space and, possibly, containing soliton walls. At low temperature the quasi-particles differs substantially from the pair excitations with activation energy  $2\Delta$  suggested by the simple mean-field description. Even at finite inter-chain electron dispersion the variational methods using the soliton solutions of the consistency equations show [16] that the lowest-energy excitations do not resemble the one-electron-type quasi-particles but involve many electrons and are accompanied by a "soliton" kinks in the coordinate dependence of the order-parameter  $\Delta_\sigma(x)$ . These excitations in the incommensurate CDW have spin 1/2, zero charge and the energy  $2\Delta/\pi$  for a 1D chain without magnetic field. In external magnetic field their activation may become energetically favorable at  $H > 2\Delta_0/\pi$  [17] that corresponds to the phase-transition from  $\text{CDW}_0$  to the  $\text{CDW}_x$  phase. Unfortunately, there is no a complete solution of this problem at nonzero temperature at present time. The calculation at the electron density close to half-filling [25] does not describe properly the incommensurate case far from the half-filling. In particular, it considers only the one-cosine modulation of the charge density. As we have seen, it is usually energetically less favorable than the double-cosine phase in high magnetic field. Therefore, we do not present an exact phase diagram in the whole temperature and magnetic field range.

Qualitatively, for the tight-binding dispersion one may expect the following picture. The cosine phase does not appear on the phase diagram besides the small region near the tricritical point as shown on Fig. 49. One can easily show that in the case of perfect nesting at  $T \rightarrow 0$  the cosine phase is always unstable toward the formation of the double-cosine modulation. However, an observation of the cosine phase would be very interesting because of its spin asymmetry in thermodynamic and transport properties. For this reason some further calculation of the phase diagram for various electron dispersion relations would be interesting. The transition from  $\text{CDW}_0$  to  $\text{CDW}_{2c}$  phase at low temperature comes about as the appearance of soliton kinks (or soliton walls). [17] The soliton phase described in Ref. [[17]] has two forbidden bands (energy gaps) which makes it similar to the double-cosine phase. With increase of magnetic field the density of kinks increases, and the order parameter comes continuously to the sinusoidal modulation close to that in the double-cosine phase. At perfect nesting the  $\text{CDW}_0$ - $\text{CDW}_{2c}$  boundary may be obtained qualitatively at arbitrary temperature as a smooth link between Fig. 49 and the point  $H_c = 2\Delta_0/\pi \approx 1.1T_{c0}$  (see Fig. 51).



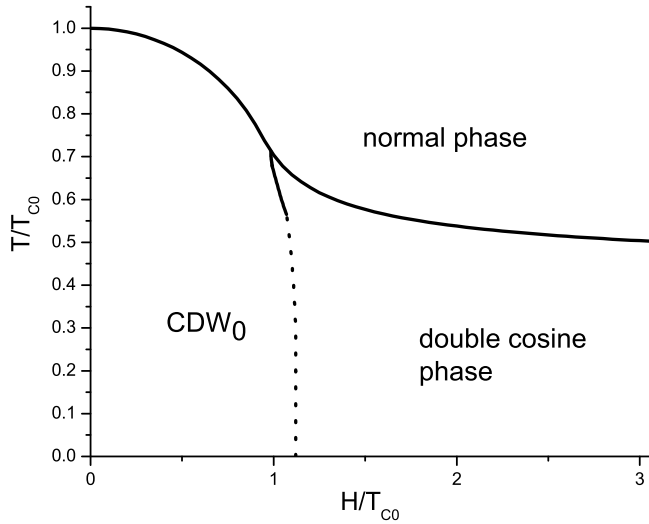


Figure 51: Phase diagram in  $H - T$  coordinates at perfect nesting in the whole essential region of CDW at  $\nu = 0.7$ . The solid lines have been calculated from Eqs. (5.28), (5.48), (5.49), (5.34) and (5.43). The  $CDW_c$  region is so narrow that it can hardly be distinguished on the total phase diagram. The dot line is a smooth link between the  $CDW_0$ - $CDW_{2c}$  boundary near  $T_c$  and its value  $H_c = 2\Delta_0/\pi \approx 1.12T_{c0}$  at  $T = 0$ .

In the case of non-uniform LOFF state the susceptibility calculation in metal phase only gives the length of the optimal wave vector  $|\mathbf{q}_{opt}|$  of the order-parameter modulation. To determine the most energetically favorable combination of  $\mathbf{q}$  :  $|\mathbf{q}| = |\mathbf{q}_{opt}|$  one has to consider the third-order terms in  $|\Delta|$  in the consistency equation. To determine which phase,  $CDW_c$  or  $CDW_{2c}$ , wins we performed the similar procedure. Our procedure takes into account some peculiarities of the CDW state such as the influence of the "antinesting" harmonic in the electron dispersion.

We have already mentioned the resemblance of the phase diagram obtained for CDW (Fig. 49) with that of layered superconductors with Rashba term.[41] The  $CDW_0$ - $CDW_{2c}$  transition at low temperature also has many common features with the transition from uniform superconducting state to LOFF state in layered superconductors.[42] In that case at the first critical field  $H_{c1}$  the formation of a single soliton kink also becomes energetically favorable, and the functional shape of this soliton is well approximated by  $\Delta(x) = \Delta_0 \tanh(x/x_0)$  both in  $CDW_{2c}$  and LOFF states at low temperature. However, there are some important differences between these two systems. In the case of superconductivity the phase transition to the LOFF state was predicted to be of the second kind [42], and the concentration of soliton walls increases gradually from zero with the increase of the magnetic field. In Ref. [[17]] the first-order phase transition was predicted between the uniform CDW and soliton phases. Our calculation also suggests the first-order transition between  $CDW_0$  and  $CDW_{2c}$  phases. The first-order phase transition from metal to the LOFF phase was predicted by both the mean-field [42] and renormalization-group studies.[?] The renormalization group analysis suggests [?] that the enhanced fluctuations of the LOFF state, associated with additional broken symmetries, are important to make this transition of the first order. We do not consider the effect of fluctuations. According to the criterion in Refs. [[5, 6]] the effects of fluctuations must not destroy the mean-field solution in organic metals where the interlayer transfer integral  $t_b$  is, usually, much larger than  $T_{c0}$ . The transition from the normal state to the  $CDW_x$  phase observed on experiments [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36] seems also to be of the second order. Further analysis of the  $CDW_0$ - $CDW_x$  phase diagram in the case of imperfect nesting at other dispersion relation can give more accurate description and even new qualitative features.

Our results are aimed to help to analyze the rather complicated phase diagram of CDW

state in high magnetic field observed in a number of organic metals, such as  $\alpha$ -(BEDT-TTF)<sub>2</sub>KHg(SCN)<sub>4</sub> [26, 27, 28, 29, 30, 31, 32, 33, 34] (see also a discussion in [[38]]), (Per)<sub>2</sub>M(mnt)<sub>2</sub> (M being Au, Pt, Cu) [35, 36, 37] etc, where the transition from CDW<sub>0</sub> to high-field CDW<sub>x</sub> phase is within the experimentally achievable magnetic fields. The above investigation may also be applied to analyze the phase diagram in nonorganic compounds with CDW ground state [2] (the attainable magnetic field in pulsed magnets is already comparable to the transition temperatures in some of these compounds). An accurate quantitative description and comparison with the experimental observations in Refs. [[26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36]] may require the substitution of a more realistic electron dispersion relation for each compound in the formulas derived above. With the two-harmonic tight-binding dispersion our study already explains several qualitative features of the CDW state in high magnetic field. First, we show that the transition from CDW<sub>0</sub> to CDW<sub>x</sub> state is of the first kind and is accompanied by the substantial jumps of the energy gap in electron spectrum and of the nesting vector that results in strong hysteresis. This fact being observed in many experiments on  $\alpha$ -(BEDT-TTF)<sub>2</sub>KHg(SCN)<sub>4</sub> has not received a theoretical substantiation before. Second, we propose and prove the appearance of double-cosine CDW above the kink field which is consistent with the low temperature soliton solution. Third, we describe some properties of the cosine and double-cosine phases.

To summarize, we have developed a mean-field theory of the CDW state in magnetic field which takes into account imperfect nesting and the shift of the CDW wave vector. This allows a detailed study of the CDW properties and the phase diagram in high magnetic field below the transition temperature. Our analysis gives the link between the previous theoretical results based on the susceptibility calculation [24] and the soliton solutions at zero temperature.[17, 16] Although our study of the CDW<sub>2c</sub> state is only applicable close to the transition line  $T_c(H)$  where the ration  $\Delta_\sigma/T_c$  can be considered as a small parameter, even an investigation in this region allows to make some important conclusions about the structure of CDW at high magnetic field. We have shown that the CDW<sub>x</sub> state at high magnetic field has predominantly a double-cosine modulation with wave vectors  $\mathbf{Q}_\pm = \mathbf{Q}_0 \pm \mathbf{q}_x$ . At perfect nesting the one-cosine CDW with shifted wave vector exists (according to the mean-field calculations) only in a very narrow region near the tricritical point (see Fig. 49 and Fig. 51). The finite second ("antinesting") harmonic in the electron dispersion (5.2) does not change this picture considerably. However, more specific dispersions or other perturbations may substantially change the phase diagram. We also suggest an interesting peculiarity of the cosine phase — its spin asymmetry that allows to make a controllable spin current.

## 6 Some other articles

### 1. Excitons in Metals: Infinite Hole Mass by Manan (1967)

#### Abstract

The optical conductivity is evaluated for interband transitions between a flat valence band and a parabolic conduction band. The conduction band is filled with electrons to a Fermi energy  $\mu_F$ . The conductivity is calculated assuming that the electron-hole interaction is attractive, static, and short range. The final-state interactions between the electron and hole cause a divergence in the conductivity at the interband threshold. This divergence appears to go as a power law. For this case of an infinite hole mass, the exciton binding energy vanishes, since the singularity in the scattering amplitude occurs just at threshold.

### 1.1 I. INTRODUCTION

PREVIOUS calculations <sup>1-4</sup> have discussed the important contribution of final-state interactions to the interband optical transition in metals or heavily doped semiconductors. The electron-hole interactions can, in some cases, significantly alter the shape and strength of the interband absorption threshold. Virtual plasmon excitations, introduced by the dynamic screening, are also an important source of final-state interactions. <sup>4</sup>

The present calculation is concerned with the finalstate interactions which contribute to a specific and simple type of interband transition. This has the valence band flat and infinitely narrow-i.e., the hole mass is infinite. This band is separated from the conduction band by a gap  $E_G$ . The conduction band is parabolic, and is filled up to a Fermi energy  $\mu_F$ . The electron-hole interaction is assumed to be attractive, static, and short-range. Our neglect of plasmon effects arising from dynamic screening means that the model does not describe a real metal. Yet the calculations may still have applications in interpreting the soft x-ray spectra of metals. These considerations are discussed in Sec. IV.

The present model is interesting because the exciton problem is directly related to the donor problem for an ionized impurity. In the conventional calculation of the electron's self-energy from ionized impurity scattering, <sup>5</sup> the exclusion-principle factors all cancel out. Any bound state, if one exists, must occur beneath the bottom of the conduction band. Yet in the exciton case, exclusion-principle effects do limit the final-state scattering. <sup>3</sup> This causes logarithmic singularities in the electron-hole scattering amplitude, which suggests the existence of a bound exciton state just below the interband threshold. The present calculation investigates in detail this disparity between the exciton- and donorstate calculations.

We have found that the exciton binding energy is zero when the hole mass is infinite. That is, the singularity in scattering amplitude occurs right at the interband threshold. The interband oscillator strength diverges as an inverse power law at threshold. The strength of this power-law divergence is determined by the strength of the coupling constant for electronhole interactions. These results appear to be the exact, infinite order, solution which is deduced from our results, which extend to third order.

Since we are primarily concerned with the analytical properties of the various logarithmic divergences, it was decided to use the contact model for the electron-hole interaction. This enormously simplifies the investigation of the singular characteristics of the final-state interactions. The attractive electron-hole interaction is characterized by a constant  $\Delta$  [ $\Delta \sim N_F |V_0|$ ] up to a cutoff energy  $\xi_0$ . In terms of these constants, the solution to the optical conductivity has the form

$$\sigma \sim \frac{1}{\omega} \left( \frac{\xi_0}{\omega - \omega_c} \right)^{2\Delta} \theta(\omega - \omega_c) \quad (1.1)$$

where  $\omega_c = E_G + \mu_F$  is the threshold frequency. This result has the power-law divergence discussed above. Our solution is just valid to "logarithmic accuracy." This means that only the important powers of the parameter

---

0

• Present address: Institute of Theoretical Science, University of Oregon, Eugene, Oregon.

<sup>1</sup> G. D. Mahan, Phys. Rev. 153, 882 (1967).

<sup>2</sup> A. W. Overhauser, Phys. Rev. 156, 844 (1967).

<sup>3</sup> G. D. Mahan, Phys. Rev. Letters 18, 448 (1967).

<sup>4</sup> R. A. Weiner, Ph.D. thesis, Harvard University, 1967 (unpublished); G. D. Mahan, Phys. Letters 24A, 708 (1967)

<sup>5</sup> R. H. Parmenter, Phys. Rev. 104, 22 (1956).

$$L(\omega) = \Delta \ln |(\omega - \omega_c) / \xi_0| \quad (1.2)$$

are retained when investigating the logarithmic divergences of the electron-hole scattering function.

In the solutions to the Kondo problem, the phrase "logarithmic accuracy" also implies that one does not keep track of the phases of the logarithmic singularity. This has led to a controversy<sup>6</sup> over the position and nature of the Kondo bound-state pole. This does not happen in the present calculation. By using a linear vertex equation, it is quite easy to keep track of the phases. A bound state, if any existed, would have the pole falling on the real  $\omega$  axis at  $\omega < \omega_c$ . In fact, it appears that one can similarly find the phase of the scattering integral in the Kondo problem by using a linear vertex equation. This possibility is discussed in Sec. IV.

## 1.2 II. THE INTERBAND OPTICAL CONDUCTIVITY

The optical properties of a metal are determined by the dielectric function<sup>7,8</sup>

$$\epsilon(\omega) = 1 + 4\pi i \sigma(\omega) / \omega. \quad (2.1)$$

The Kubo formula is used to evaluate the conductivity  $\sigma(\omega)$ . For an isotropic system at zero temperature,<sup>9,10</sup>

$$\sigma(\omega) = \frac{2}{3\omega} \int_0^\infty dt e^{i\omega t} \langle [j_\nu(0), j_\nu(t)] \rangle \quad (2.2)$$

For interband transitions, we assume that the  $\mathbf{p}$  operator in  $\mathbf{j}$  operates between Bloch functions. This matrix element is taken to be independent of energy, as is done in most calculations. The correlation function (2.2) is most easily evaluated in the Matsubara formalism.<sup>10</sup> This will be discussed first for an arbitrary set of bands, and then for the particular example of interest.

Denote  $C_{j\mathbf{k}}^\dagger$  as the creation operator for an electron in band  $j$  with momentum  $\mathbf{k}$ . We define the correlation function

$$\begin{aligned} \Pi(i\omega) &= \frac{1}{V} \sum_{\mathbf{k}, \mathbf{k}', i, j, n, m} \int_0^\beta d\tau e^{i\omega(\tau - \tau')} \\ &\times \left\langle T_\tau C_{j\mathbf{k}}(\tau) C_{i\mathbf{k}}^\dagger(\tau) C_{m\mathbf{k}'}(\tau') C_{n\mathbf{k}'}^\dagger(\tau') \right\rangle \end{aligned} \quad (2.3)$$

The retarded form of this correlation function  $\Pi_{\text{ret}}(\omega)$  is obtained by setting  $i\omega \rightarrow \omega + i\delta$ . In terms of this retarded function, the interband conductivity is

$$\begin{aligned} \text{Re } \sigma(\omega) &= \frac{e^2 \langle p \rangle^2}{3\omega m^2} A(\omega) \\ A(\omega) &= -2 \text{Im } \Pi_{\text{ret}}(\omega) \end{aligned}$$

The function  $A(\omega)$  has a simple physical interpretation. The correlation function (2.3) can be evaluated by a diagrammatic expansion as in Fig. 1, where dotted lines are the electron-hole interactions. If one ignores these interactions, then the correlation function is just given by the first term [Fig. 1(a)]. Call this function  $\Pi^{(0)}(i\omega)$ , and

<sup>06</sup> S. D. Silverstein and C. B. Duke, Phys. Rev. Letters 18, 695 (1967); H. Suhl, *ibid.* 18, 743 (1967).

$$A^{(0)}(\omega) = -2 \text{Im} \Pi_{\text{ret}}^{(0)}(\omega)$$

$$A^{(0)}(\omega) = 2\pi \sum_{ij} \int_{E_i < \mu_f, E_j > \mu_\mu} \frac{d^3 k}{(2\pi)^3} \quad (2.4)$$

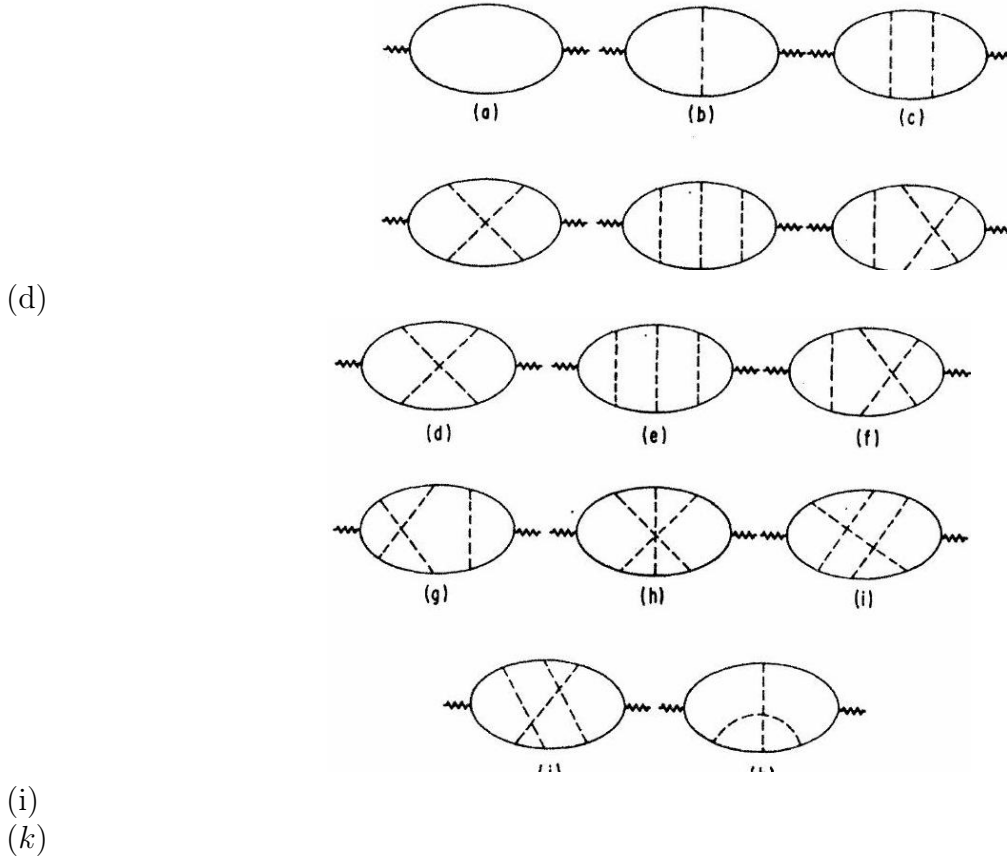


Fig. 1. Some types of correlation functions which contribute to the final-state interactions. The dashed lines represent the static, attractive, electron-hole interactions.

The function  $A^{(0)}(\omega)$  is proportional to the joint density of states. Specifically,

$$A^{(0)}(\omega) = \pi \rho(\omega)$$

Previous calculations<sup>7,8</sup> of the optical properties of metals have assumed that the transition probability is proportional to the joint density of states. This is no longer adequate when final-state interactions are important,<sup>1</sup> since then  $A(\omega)$  is quite different than  $A^{(0)}(\omega)$ .

These equations simplify when applied to the present problem. For a contact interaction, the retarded correlation function is

$$\Pi_{\text{ret}}(\omega) = \int_0^{\xi_0} \frac{d\xi}{\omega - \omega_c + \xi + i\delta} \Gamma(\xi, \omega) \quad (2.5)$$

where  $\xi$  is the conduction electron's energy as measured from its Fermi surface. The vertex function  $\Gamma(\xi, \omega)$  is generated by the diagrammatic expansion of Fig. 1. This procedure is

<sup>07</sup> H. Ehrenreich and H. R. Philipp, Phys. Rev. 128, 1622 (1962);

B. R. Cooper, H. Ehrenreich, and H. R. Philipp, *ibid.* 138, A494 (1965).

<sup>8</sup>H. Ehrenreich, H. R. Philipp, and B. Segall, Phys. Rev. 132, 1918 (1963).

<sup>9</sup> W. Kohn, in Optical Properties and Electronic Structure of Metals and Alloys, edited by F. Abeles (John Wiley & Sons, Inc., New York, 1966).

<sup>10</sup> A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, Quantum Field Theory in Statistical Physics (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963).

discussed in the next section. The poles and structure of  $\Gamma(\xi, \omega)$  determine the nature of the interband absorption. This is the interesting aspect of optical experiments, since one directly measures the scattering function.

### 1.3 III. HIGHER-ORDER INTERACTIONS

#### 1.4 A. Perturbation Expansion

We begin the discussion of higher-order interactions by examining their contribution to a perturbation expansion for the correlation function. A solution to the correlation function is given in the following sections. The perturbation-series result is useful for providing a qualitative introduction to the effects of these higher interaction terms. It also provides an expansion against which to check the exact result after it is obtained.

As discussed above, the perturbation-series results will only be evaluated to logarithmic accuracy. The result is a power series in  $L(\omega)$  given in (1.2). When the electron-hole potential is an attractive Coulomb potential with Fermi Thomas screening, then the coupling parameter is

$$\Delta \cong (r_s/12) \ln(1 + 6/r_s)$$

where  $r_s$  is the standard density parameter. It was shown in Ref. 1 that the contribution of the ladder diagrams to the interband absorption may be expressed as

$$A_L(\omega) = A^{(0)}(\omega) \left| 1 - \frac{\Lambda_2(\omega)}{1 + \Lambda_1(\omega)} \right|^2 \quad (3.1)$$

To logarithmic accuracy, the terms appearing in (3.1) are

$$\begin{aligned} \Lambda_1(\omega) &= \Lambda_2(\omega) = L(\omega) - i\pi\Delta\theta(\omega - \omega_c), \\ A^{(0)}(\omega) &= 2\pi\theta(\omega - \omega_c). \end{aligned} \quad (3.2)$$

Expanding (3.1), and ignoring the terms  $\text{Im } \Lambda$ , gives

$$A_L(\omega) = A^{(0)}(\omega) \{1 - 2L(\omega) + 3L(\omega)^2 - 4L(\omega)^3 \dots\}$$

The terms in the expansion originate from the ladder diagrams in Figs. 1(a), 1(b), 1(c), and 1(e). The term  $-2L(\omega)$  corresponds to Fig. 1(b). Since it is the only vertex term of order  $\Delta$ , this logarithmic singularity cannot be canceled by any other diagram. The first important higher-order interaction is the cross diagram Fig. 1(d). When the hole mass is infinite, this contributes to the absorption an amount

$$A_{1d} = -A^{(0)}(\omega)L(\omega)^2$$

Note that this only partially cancels the  $3L^2(\omega)$  contribution from Fig. 1(c). The third-order diagrams (e) to (j) in Fig. 1 contribute in units of  $L^3$ :  $-4, \frac{4}{3}, \frac{4}{3}, -\frac{2}{3}, \frac{1}{3}, \frac{1}{3}$ . A consideration of all terms in Fig. 1 from (a) to (j) gives

$$A(\omega) = A^{(0)}(\omega) \left\{ 1 - 2L + 2L^2 - \frac{4}{3}L^3 \dots \right\} \quad (3.3)$$

It is evident that these logarithmic singularities remain in the conductivity even when higher-order interactions are included. When the hole mass is finite, the nonladder diagrams do not

contribute logarithmic singularities of the same order as the ladder terms. This is what makes the infinite hole mass an interesting special case of study, since here all of the diagrams of the kind in Fig. 1 from (a) to (j) contribute significant logarithmic singularities. In the impurity scattering calculation, the logarithmic singularities from the nonladder diagrams exactly cancel those from the ladder terms. This obviously does not occur in the present exciton calculation.

Vertex corrections such as Fig. 1(k) have not been included in this analysis. In this exciton problem, they do not contribute logarithmic singularities of the type in (3.3). This diagram was important in the recent analysis by Luttinger and Kohn<sup>11</sup> of a superconductive instability in low-temperature electron gases. The present calculation differs from theirs in several respects. Since both of their particle propagators are electrons in the Fermi sea, they have a symmetry which is lacking in the electron-hole problem. Also, they only found instabilities in states of odd angular momentum, while we are explicitly concerned with relative  $s$  states.

The series in (3.3) appears to be

$$e^{-2L} = 1 - 2L + 2L^2 - \frac{4}{3}L^3 + \dots \quad (3.4)$$

Of course, one cannot guarantee that the series (3.3) is an exponential without evaluating the series to all orders. However, the solution of the scattering integral, which is given below, also suggests that this is the correct series limit. Thus we proceed on the assumption that the series (3.3) is just an exponential. Since the correlation function is an analytic function of  $\omega$ , the real and imaginary parts are related. Using the definition of  $L(\omega)$ , this further suggests that our result is

$$A(\omega) = \Delta^{-1} \text{Im} \times \exp \left[ -2\Delta \ln \left| \frac{\omega - \omega_c}{\xi_0} \right| + 2\Delta \pi i \theta(\omega - \omega_c) \right] \quad (3.5)$$

$$A(\omega) = \frac{\theta(\omega - \omega_c)}{\Delta} \left| \frac{\xi_0}{\omega - \omega_c} \right|^{2\Delta} \sin(2\Delta\pi).$$

The distressing observation that this is negative for  $\Delta > \frac{1}{2}$  appears to be an artifact of our approximation, as is discussed below. Note that the result does become  $A^{(0)}(\omega)$  as  $\Delta \rightarrow 0$ . The divergence at the critical frequency is a power law, rather than logarithmic. This result could only be obtained from an infinite-order solution, which in this case has only been guessed. It also shows that there is no bound state in the infinite hole mass case. Essentially, the binding energy has become zero since the singularity occurs precisely at  $\omega = \omega_c$ .

### 1.5 B. Solution to the Vertex Equation

We now proceed to solve for the correlation function (2.5). For the irreducible interactions of interest, it is easy to show that one can write a linear vertex equation

$$\Gamma(\xi, \omega) = 1 + \int_0^{\xi_0} \frac{d\xi'}{\omega - \omega_c + \xi' + i\delta} \Gamma(\xi', \omega) V(\xi, \xi', \omega) \quad (3.6)$$

The potential  $V$  is the sum of all irreducible interactions. For the present calculation, we have retained contribu-

<sup>11</sup> J. M. Luttinger, Phys. Rev. 150, 202 (1966); W. Kohn and J. M. Luttinger, Phys. Rev. Letters 15, 524 (1966).

tions to  $V$  up to third order in  $\Delta$



$$V(\xi, \xi', \omega) = - \left\{ \Delta + \Delta^2 \ln \left( \frac{\xi + \xi' - \omega + \omega_c}{\xi_0} \right) - \Delta^3 L_2 \left( \frac{\xi}{\xi + \xi' - \omega + \omega_c} \right) - \Delta^3 L_2 \left( \frac{\xi'}{\xi + \xi' - \omega + \omega_c} \right) \right\} \quad (3.7)$$

where  $L_2(z)$  is the Euler dilogarithmic function

$$L_2(z) = - \int_0^z \frac{dt}{t} \ln(1-t)$$

The imaginary parts of  $V$  are found by the condition, determined by causality, that the imaginary part of  $\omega$  be infinitesimally real in the region of branch cuts. The first term  $\Delta$  is the direct interaction [Fig. 1(b)]. The second term is the exchange interaction [ Fig. 1(d) ]. The terms of order  $\Delta^3$  are the sum of the three interactions h, i, and j in Fig. 1. The result (3.7) is derived in the Appendix.

The vertex equation (3.6) is solved by a slight modification of Noyes's <sup>12</sup> method. Define two functions  $t(\omega)$  and  $f(\xi, \omega)$  by

$$\Gamma(\xi, \omega) = t(\omega) f(\xi, \omega) \quad (3.8)$$

with the additional normalization requirement that

$$f(\xi_1, \omega) = 1 \quad (3.9)$$

for some value of  $\xi = \xi_1$ . Then these equations can be manipulated to give

$$f(\xi, \omega) = 1 + \int_0^{\xi_0} \frac{d\xi'}{\omega - \omega_c + \xi' + i\delta} f(\xi', \omega) \times [V(\xi', \xi_1, \omega) - V(\xi', \xi, \omega)], \quad (3.10)$$

$$t(\omega) = 1/[1 + \Lambda(\omega)]$$

$$\Lambda(\omega) = \int_0^{\xi_0} \frac{d\xi'}{\omega - \omega_c + \xi' + i\delta} f(\xi', \omega) V(\xi', \xi_1, \omega). \quad (3.11)$$

Equation (3.10) provides a method of systematically evaluating  $f(\xi, \omega)$  as a power series in  $\Delta$ . Then one evaluates  $\Lambda(\omega)$  and  $\Pi(\omega)$ . The poles of  $\Pi(\omega)$  specify the nature of the bound state.

The potentials in (3.7) have the property

$$\text{Im } V(\xi, \xi', \omega) = 0, \quad \xi, \xi' > 0, \quad \omega < \omega_c \quad (3.12)$$

Although this has been explicitly verified only up to and including third-order interactions, it seems reasonable to conclude that it is generally true. This makes  $f(\xi, \omega)$  entirely real for  $\omega < \omega_c$  and  $\xi > 0$ . This means that a bound-state pole, if one exists for  $\omega < \omega_c$ , falls on the real axis.

Equations (3.8) to (3.10) are first solved for  $\omega < \omega_c$ . The normalizing condition (2.9) is chosen at  $\xi_1 = \omega_c - \omega$ .

Evaluating (3.10) and (3.11) up to order  $\Delta^3$  gives

---

<sup>12</sup> H. P. Noyes, Phys. Rev. Letters 15, 538 (1965); K. I. Kowalski, *ibid.* 15, 798 (1965).

$$f(\xi, \omega) = 1 - \Delta^2 \left[ L_2 \left( \frac{\xi}{\omega - \omega_c} \right) - L_2(-1) \right] \quad (3.13)$$

$$\Lambda(\omega) = L(\omega) + \frac{1}{2!} L(\omega)^2 + \frac{1}{3!} L(\omega)^3 + O(\Delta^4) \quad (3.14)$$

With the knowledge of  $f(\xi, \omega)$ , the full correlation function in (2.5) can be evaluated for  $\omega < \omega_c$ ,

$$\Pi(\omega) = \frac{1}{\Delta} \frac{L + L^3/3! + O(\Delta^5)}{1 + L + L^2/2! + L^3/3! + O(\Delta^4)} \quad (3.15)$$

The denominator series appears to be  $\exp(L)$ , while the numerator we take to be  $\sinh(L)$ . Assuming that we have correctly guessed the result, we get

$$\Pi(\omega) = \frac{1}{2\Delta} [1 - \exp(-2L)] \quad (3.16a)$$

$$\Pi(\omega) = \frac{1}{2\Delta} \left[ 1 - \left( \frac{\xi_0}{\omega - \omega_c} \right)^{2\Delta} \right] \quad (3.16b)$$

The singularity occurs at  $\omega = \omega_c$ . There is no boundstate pole, but just a cut which starts at  $\omega = \omega_c$ .

Equations (3.8) to (3.10) have also been solved for  $\omega > \omega_c$ . Here one must keep track of both the real and imaginary parts of the various functions. It is convenient to set  $\xi_1 = \omega - \omega_c + \delta$  in (3.9). The calculation is tedious, but the result, correct to logarithmic accuracy, is for  $\omega > \omega_c$

$$\Pi(\omega) = \frac{1}{\Delta} \frac{L + L^3/3! - i\pi\Delta(1 + L^2/2!)}{1 + L + L^2/2! + L^3/3! - i\pi\Delta(1 + L + L^2/2!)}$$

This gives for  $A(\omega)$

$$A(\omega) = 2\pi\theta(\omega - \omega_c) \left\{ (1 + L + L^2/2! + L^3/3!)^2 + \Delta^2\pi^2 (1 + L + L^2/2!)^2 \right\}^{-1}$$

If we assume that the series is exponential, then

$$A(\omega) = \frac{2\pi\theta(\omega - \omega_c)}{1 + \Delta^2\pi^2} \left( \frac{\xi_0}{\omega - \omega_c} \right)^{2\Delta} \quad (3.17)$$

This result appears slightly different from (3.5). Both equations contain the same  $(\omega - \omega_c)^{-2\Delta}$  divergence. The other constant terms differ because we have not properly kept track of them in deriving either (3.5) or (3.17). For instance, consider the integral

$$\begin{aligned} \operatorname{Re} \int_0^{\xi_0} \frac{d\xi}{\omega - \omega_c + \xi} \ln \left| \frac{\xi}{\xi_0} \right| &= \frac{1}{2} \ln^2 \left| \frac{\omega - \omega_c}{\xi_0} \right| \\ &\quad - \frac{1}{6} \pi^2 - \frac{1}{4} \pi^2 \theta(\omega - \omega_c) + O\left(\frac{\omega - \omega_c}{\xi_0}\right) \end{aligned}$$

In solving the problem to logarithmic accuracy, we have been ignoring the constant terms such as  $\pi^2/6$  and  $\pi^2/4$ . The existence of such terms explains the differences between (3.17) and (3.5). The difficulty with including such constant terms is that the problem has two singular regions, one at  $\xi \sim 0$  and one at  $\xi \sim \xi_0$ . The unphysical singularity at  $\xi \sim \xi_0$  is an artifact of

the contact model. Yet, the constant terms cannot be calculated accurately without treating the singularities at  $\xi \sim \xi_0$  in some arbitrary fashion. The fact that these constant terms may be appreciable when  $\Delta$  is not small may explain the disturbing observation that (3.5) is negative for  $\Delta > \frac{1}{2}$ .

## 1.6 IV. DISCUSSION

A solution has been obtained to the correlation function specifying interband optical absorption for the case of an infinite hole mass. The analytic structure of the divergences near the absorption threshold has been investigated using a contact model for the attractive Coulomb interaction. We found that both the perturbation-series result, and the solution to the scattering integral, suggest that the divergence is a power law. If  $\omega$  is the photon frequency, then

$$\sigma(\omega) \sim \frac{1}{\omega} \left( \frac{\xi_0}{\omega - \omega_c} \right)^{2\Delta} \theta(\omega - \omega_c)$$

where  $\omega_c$  is the threshold frequency,  $\xi_0$  is a cutoff energy, and  $\Delta$  is the strength of the attractive electron-hole interaction.

The absorption and emission spectra of soft x rays constitute a physical system which approximates the present model. The emission spectra should have a similar divergence since, as discussed by Suna,<sup>13</sup> the emission and absorption differ only by factors relating to the occupation factors. Some x-rays data<sup>14,15</sup> do show structure at threshold. Some of these cases, such as magnesium, are satisfactorily explained in terms of density-of-states effects. But the  $L$  and  $M$  absorption edges of<sup>14,15</sup> Ni, Cu, and Zn do show peaks at threshold which apparently are not caused by the density of states. Although this feature could be caused by exciton processes, it is sobering to note that the spectra also have much other unexplained structure, such as the high-energy tail in  $K$  emission.

For interband transition in real metals, final-state interactions involving the virtual excitation of plasmons are also important. These plasmons arise from the dynamic nature of the screening of the electron-hole Coulomb interaction. These dynamic interactions appear to be more important than the statically screened interaction in interband optical transitions in real metals. In these cases, however, the statically screened Coulomb interaction does not lead to logarithmic singularities. Nevertheless, it might be necessary to include the plasmon processes before the present calculations can be applied to describe real x-ray transitions.

One can also envision processes where the hole is in a  $d$  state, and where the important final-state interactions are caused by an  $s-d$  interaction. The nature of exciton states in this process is related to the Kondo problem.<sup>16,17</sup> For this interaction, the linear vertex equation (3.6) becomes a matrix equation because of spin.

$$\begin{aligned}\Gamma &= \Gamma_s + \boldsymbol{\sigma} \cdot \mathbf{s}\Gamma_v \\ V &= V_s + \boldsymbol{\sigma} \cdot \mathbf{s}V_v\end{aligned}$$

Each of the higher-order irreducible interactions, such as Fig. 1(d), contributes both to  $V_s$  and  $V_v$ . Since the analytic properties of the irreducible interactions are the same for the exciton and Kondo problems, then (3.12) applies. If one solves  $\Gamma = N/D$ , then  $N$  and  $D$  are both real for  $\omega < \omega_c$ . Hence, the zeros of  $D$  for  $\omega < \omega_c$  give poles in  $\Gamma$  which correspond to a donor state. Abrikosov's<sup>17</sup> solution to the Kondo problem indicates that such poles do exist.

<sup>13</sup> A. Suna, Phys. Rev. 135, A111 (1964).

<sup>14</sup> L. G. Parratt, Rev. Mod. Phys. 31, 616 (1959).

<sup>15</sup> D. H. Tomboulia, in Handbuch der Physik, edited by S. Flügge (Springer-Verlag, Berlin, 1957), Vol. 30, p. 246.

Note added in manuscript. J. J. Hopfield<sup>18</sup> has suggested that the ground state of an electronic system with and without an impurity are orthogonal. P. W. Anderson<sup>19</sup> has recently suggested a proof of this theorem, and notes that this causes some transitions to have divergences. His functions diverge as a power law with an exponent

$$\epsilon = \frac{1}{3\pi^2} \sum_l (2l+1) \sin^2 \delta_l = \Delta^2/3$$

This exponent of  $\Delta^2/3$  is quite different, and much weaker, than the present result  $2\Delta$ . Thus, Anderson's calculation predicts divergences which are much smaller than those discussed here.

### 1.7 APPENDIX: THE IRREDUCIBLE INTERACTIONS

Equation (3.6) indicates that the correlation function can be evaluated using a linear integral equation. This is only true because we have assumed that the basic electron-hole interaction is static. That one can use a linear integral equation of this form is deduced by writing out explicitly the lowest six or eight terms. Here we will elaborate on this by deriving (3.7) for the irreducible interactions.

The two equations (3.6) and (3.7) will first be written in terms of momentum variables. This is conventionally the way these results are expressed, and it is instructive

<sup>16</sup> K. Yoshida, Phys. Rev. 147, 223 (1966).

<sup>17</sup> A. A. Abrikosov, Physics 2, 5 (1965); 2, 61 (1965).

<sup>18</sup> J. J. Hopfield (private communication).

<sup>19</sup> P. W. Anderson, Phys. Rev. Letters 18, 1049 (1967).

to see them before one invokes the contact model.

$$\Gamma(\mathbf{p}, \omega) = 1 + \int \frac{d^3 p'}{(2\pi)^3} \times \frac{[1 - n_F(p')]}{\omega - \omega_c + \xi'_p + i\delta} \Gamma(\mathbf{p}', \omega) V(\mathbf{p}, \mathbf{p}', \omega) \quad (\text{A1})$$

$$V(\mathbf{p}, \mathbf{p}', \omega) = \sum_m V_m(\mathbf{p}, \mathbf{p}', \omega).$$

The potential is the sum of all irreducible interactions. The  $m$  subscript corresponds to Fig. 1(m); thus  $V_b$  is the irreducible interaction first introduced in Fig. 1(b).

$$V_b(\mathbf{p}_1, \mathbf{p}_2, \omega) = \mathcal{U}(\mathbf{p}_1 - \mathbf{p}_2)$$

$$V_d(\mathbf{p}_1, \mathbf{p}_2, \omega) = - \int \frac{d^3 p_3}{(2\pi)^3} \frac{n_F(p_3) \mathcal{U}(\mathbf{p}_1 - \mathbf{p}_3) \mathcal{U}(\mathbf{p}_3 - \mathbf{p}_2)}{\omega - \omega_c + \xi_3 - \xi_1 - \xi_2 + i\delta}$$

$$V_h(\mathbf{p}_1, \mathbf{p}_2, \omega) = \int \frac{d^3 p_3 d^3 p_4}{(2\pi)^6} \frac{\mathcal{U}(\mathbf{p}_1 - \mathbf{p}_3) \mathcal{U}(\mathbf{p}_3 - \mathbf{p}_4) ((\mathbf{p}_4 - \mathbf{p}_2) n_F(p_3) n_F(p_4))}{(\omega - \omega_c - \xi_1 - \xi_2 + \xi_3 + i\delta) (\omega - \omega_c - \xi_1 - \xi_2 + \xi_4 + i\delta)}$$

$$V_i(\mathbf{p}_1, \mathbf{p}_2, \omega) = - \int \frac{d^3 p_3 d^3 p_4}{(2\pi)^6} \frac{\mathcal{U}(\mathbf{p}_1 - \mathbf{p}_3) \mathcal{U}(\mathbf{p}_3 - \mathbf{p}_4) \mathcal{U}(\mathbf{p}_4 - \mathbf{p}_2) n_F(p_3) (1 - n_F(p_4))}{(\omega - \omega_c - \xi_1 - \xi_2 + \xi_3 + i\delta) (\omega - \omega_c - \xi_1 - \xi_4 + \xi_3 + i\delta)}$$

$$V_j(\mathbf{p}_1, \mathbf{p}_2, \omega) = - \int \frac{d^3 p_3 d^3 p_4}{(2\pi)^6} \frac{\mathcal{U}(\mathbf{p}_1 - \mathbf{p}_3) \mathcal{U}(\mathbf{p}_3 - \mathbf{p}_4) \mathcal{U}(\mathbf{p}_4 - \mathbf{p}_2) (1 - n_F(p_3)) n_F(p_4)}{(\omega - \omega_c - \xi_2 - \xi_3 + \xi_4 + i\delta) (\omega - \omega_c - \xi_1 - \xi_2 + \xi_4 + i\delta)}$$

where

$$\xi_n = (p_n^2 - p_F^2) / 2m, \quad n = 1, 2, 3, 4$$

In the contact model, one replaces

$$\int \frac{d^3 p}{(2\pi)^3} \mathbf{u} \rightarrow -\Delta \int_{-\xi_0}^{\xi_0} d\xi$$

where  $\mathfrak{U}$  is attractive and  $\Delta > 0$ . This immediately takes (A1) into (3.6). Similarly, the irreducible interactions become

$$V_b(\xi, \xi', \omega) = -\Delta,$$

$$V_d(\xi, \xi', \omega) = -\Delta^2 \int_{-\xi_0}^0 \frac{d\xi_3}{\omega - \omega_c + \xi_3 - \xi - \xi' + i\delta} = -\Delta^2 \ln \left( \frac{\xi + \xi' - \omega + \omega_c - i\delta}{\xi_0} \right)$$

$$V_h(\xi, \xi', \omega) = -\Delta^3 \left\{ \int_{-\xi_0}^0 \frac{d\xi_3}{\omega - \omega_c + \xi_3 - \xi - \xi' + i\delta} \right\}^2 = -\Delta^3 \ln^2 \left( \frac{\xi + \xi' - \omega + \omega_c - i\delta}{\xi_0} \right)$$

$$\begin{aligned} V_i + V_j &= \Delta^3 \int_{-\xi_0}^0 d\xi_3 \int_0^{\xi_0} d\xi_4 \frac{1}{\omega - \omega_c + \xi_3 - \xi - \xi' + i\delta} \left[ \frac{1}{\omega - \omega_c - \xi_4 + \xi_3 - \xi + i\delta} + \frac{1}{\omega - \omega_c - \xi_4 + \xi_3 - \xi'} \right] \\ &= \Delta^3 \left\{ \ln^2 \left( \frac{\xi + \xi' - \omega + \omega_c - i\delta}{\xi_0} \right) + L_2 \left( \frac{\xi}{\xi + \xi' - \omega + \omega_c - i\delta} \right) + L_2 \left( \frac{\xi'}{\xi + \xi' - \omega + \omega_c - i\delta} \right) \right\} \end{aligned}$$

The sum of these terms gives (3.7).

## 2. Tomonaga's Model and the Threshold Singularity of X-Ray Spectra of Metals by Schotte, Schotte (1969)

### Abstract

Singularities near the threshold of the soft x-ray spectra of metals have been predicted by Mahan and have recently been calculated by Nozières et al. using the model of a localized core hole. We show that the singular behavior can be understood in terms of density waves of the conduction electrons which are excited when, in the absorption process, the core hole is created, providing an attractive potential for the conduction electrons. For the description of the conduction electrons in terms of density waves, Tomonaga's model is adopted.

### 2.1 1. INTRODUCTION

IN the x-ray absorption process, a deep-lying core electron is excited into the conduction band by an incoming x-ray photon. The core hole left behind acts as a one-body potential on the conduction electrons. A simplified model has been used to describe the situation <sup>1</sup>:

$$H = \sum_k \epsilon_k a_k^\dagger a_k + E_0 b^\dagger b + \frac{1}{N} \sum_{k,k'} V_{kk'} a_k^\dagger a_{k'} b b^\dagger \quad (1)$$

Here the hole (described by  $b^\dagger, b$ ) is represented by a single nondegenerate level  $E_0$  with infinite lifetime. It is assumed to interact with the free conduction electrons (described by  $a_k^\dagger, a_k$ ) via a contact potential  $V_{kk'} = V$ ; the interaction between conduction electrons has been neglected.

We confine ourselves to the discussion of the adsorption process. In this case the transition role according to the Golden Rule is

$$W(\omega) = 2\pi \sum_f \left| \sum_k w_k(\omega) \langle f | a_k^\dagger b | i \rangle \right|^2 \delta(\tilde{E}_i + \omega - \tilde{E}_f) \quad (2)$$

For simplicity the matrix elements of the dipole operator  $w_k(\omega)$  will be regarded as constant. In an equivalent one-body description, the transition rate is given by

$$W(\omega) = 2\pi w^2 \sum_f \left| \left\langle f_{n+1} \left| \frac{1}{\sqrt{N}} \sum_k a_k^\dagger \right| i_n \right\rangle \right|^2 \times \delta(E_i - E_f + E_0 + \omega) \quad (3)$$

The initial state

$$|i_n\rangle = \prod_{k=0}^{k_F} a_k^\dagger |0\rangle$$

is the  $n$ -particle ground state of the Hamiltonian

$$H_i = \sum_k \epsilon_k a_k^\dagger a_k \quad (4)$$

with the energy  $E_i = \tilde{E}_i - E_0$ . The final states  $|f_{n+1}\rangle$  are the  $(n+1)$ -particles eigenstates of the Hamiltonian

$$H_f = \sum_k \epsilon_k a_k^\dagger a_k + \frac{V}{N} \sum_{k,k'} a_k^\dagger a_{k'}. \quad (5)$$

The final states  $|f_{n+1}\rangle$  hold the clue to the problem. According to Anderson,<sup>2</sup> the overlap between the initial and any final state containing a finite number of electron-hole pairs is zero in the limit of infinite volume.

So one has the choice of doing the calculations with finite volume, performing the limiting process to infinite volume in the end, or of circumventing the problem to determine the final states by using Green's-function techniques.<sup>1</sup> However, the physical origin of the singular behavior of the transition rate near the threshold<sup>1,3</sup> does not seem to be clearly understood.

In this work, we offer another way of calculating the response of the free electrons to the sudden switching on of the core hole potential. We will use the Tomonaga model<sup>4</sup> by which we can describe the excitations of the Fermi sea in terms of density waves. On the one hand, we circumvent the difficulty posed by the Anderson theorem, and, on the other hand, we have a plausible physical interpretation of what is going on.

## 2.2 FORMULATION OF THE PROBLEM IN TERMS OF TOMONAGA'S BOSONS

We introduce the density operator

$$\rho_k = \sum_{k_1=0}^{k_D-k} \frac{1}{\sqrt{N}} a_{k_1}^\dagger a_{k_1+k}, \quad (6)$$

$$\rho_{-k} = \sum_{k_1=k}^{k_D} \frac{1}{\sqrt{N}} a_{k_1}^\dagger a_{k_1-k}, \quad k \geq 0$$

---

0

- Supported by the Deutsche Forschungs-Gemeinschaft. Supported by the U. S. National Science Foundation and the Office of Naval Research.

† U. S. gratefully acknowledges a Travel Grant from the Deutsche Forschungs-Gemeinschaft. On leave of absence of the Institut für Theoretische Physik der Universität zu Köln.

<sup>1</sup> P. Nozières and C. T. de Dominicis, Phys. Rev. 178, 1084 (1969); B. Roulet, I. Gavoret, and P. Nozières, ibid. 178, 1252 (1969).

and consider a band of width  $D$ , assuming a constant density of states within the band.  $k = 0$  and  $k_D = k$  are the momenta at the bottom and at the top of band, respectively. The electron energies are given by

$$\epsilon_k = (k - k_F) / \rho \quad (7)$$

where  $\rho$  is the density of states.

Actually, in the Tomonaga model the free-particle energy has to be proportional to the momentum. Another important feature of the Tomonaga model is its one-dimensionality. But the problem considered here can be looked upon as a one-dimensional one because it involves only  $s$ -wave scattering. So the summation in (6) over  $k_1$  is to be understood as a summation over energy shells. We examine now the commutation relations of the  $\rho_k$ 's:

$$[\rho_k, \rho_{k'}] = 0 \text{ for } k, k' > 0 \text{ and } k, k' < 0 \quad (8)$$

$$\begin{aligned} [\rho_k, \rho_{-k'}] &= \frac{1}{N} \sum_{k_1=0}^{k'} a_{k_1}^\dagger a_{k_1+k-k'} - \frac{1}{N} \sum_{k_1=k_D-k}^{k-k+k'} a_{k_1}^\dagger a_{k_1+k-k'} \\ [\rho_k, \rho_{-k}] &= \frac{1}{N} \sum_{k_1=0}^k a_{k_1}^\dagger a_{k_1} - \frac{1}{N} \sum_{k_1=k_D-k}^{k_D} a_{k_1}^\dagger a_{k_1} \end{aligned} \quad (9)$$

We adopt Tomonaga's approximation and substitute these complicated commutation relations by simpler ones:

$$[\rho_k, \rho_{-k'}] = k \delta_{kk'} \quad (11)$$

For a detailed discussion of this important step we refer to Tomonaga's paper. In short, the simpler commutation relations may be used if only a certain subspace of all possible states has to be considered. These are the states which do not have unoccupied levels deep in the bottom of the band or occupied levels high above the Fermi energy at the top of the band, such that the parts left out of the commutator do not contribute very much. In addition, we have to assume that  $V$  be very small compared to the bandwidth. On the one hand we had to consider a contact potential to keep mathematics simple, but, on the other hand, a contact potential causes excitations deep below and high above the Fermi level (or short density waves), which is explicitly excluded in the model. So we must expect all our results to be valid only for small  $V$ .

With (11) we calculate the commutator

$$[H_i, \rho_k] = -(k/\rho) \rho_k \quad (12)$$

where use has been made of the linear energy-momentum dependence, and we conclude that  $H_i$  has the form <sup>5</sup>

$$\tilde{H}_i = \sum_{k>0} \frac{1}{\rho} \rho_{-k} \rho_k \quad (13)$$

The transcription of  $H_f$  [Eq. (5)] in terms of bosons also leads to a very simple expression,

---

<sup>02</sup> P. W. Anderson, Phys. Rev. Letters 18, 1049 (1967).

<sup>3</sup> G. D. Mahan, Phys. Rev. 163, 612 (1967).

<sup>4</sup> S. Tomonaga, Progr. Theoret. Phys. (Kyoto) 5, 544 (1950). 479



$$\tilde{H}_f = \sum_{k>0}^1 \rho_{-k} \rho_k + \frac{V}{\sqrt{N}} \sum_{k>0} (\rho_k + \rho_{-k}) \quad (14)$$

It is convenient to introduce the normalized boson operators ( $k > 0$ )

$$b_k = (1/\sqrt{k})\rho_k, \quad b_k^\dagger = (1/\sqrt{k})\rho_{-k}, \quad (15)$$

which obey the usual boson commutation relations

$$[b_k, b_{k'}^\dagger] = \delta_{kk'} \quad (16)$$

In terms of bosons

$$\tilde{H}_i = \sum_{k>0} k b_k^\dagger b_k, \quad (17)$$

and

$$H_f = \sum_{k>0} \frac{k}{\rho} \left( b_k^\dagger + \frac{V}{\sqrt{(kN)}} \right) \left( b_k + \frac{V}{\sqrt{(kN)}} \right) - \frac{V^2 \rho}{N} \sum_k 1.$$

$\tilde{H}_i$  and  $\tilde{H}_f$  describe a set of harmonic oscillators. The effect of the potential is a shift of the zero points of the harmonic oscillators. As the transformation procedure is only correct up to a constant we will drop the constant term on the right-hand side of Eq. (18) in the following: In order to calculate the transition rate (3) we have to express the operator

$$a^\dagger = \frac{1}{\sqrt{N}} \sum_{k_1} a_{k_1}^\dagger$$

in terms of harmonic-oscillator coordinates. This is a rather delicate problem. First we examine the commutation relations between  $a^\dagger$  and the density operators  $\rho_k$ . We find

$$[\rho_k, a^\dagger] = \frac{1}{\sqrt{N}} a^\dagger - \frac{1}{N} \sum_{k_1=k_D-k}^{k_D} a_{k_1}^\dagger \quad \text{for } k > 0 \quad (19)$$

$$[\rho_{-k}, a^\dagger] = \frac{1}{\sqrt{N}} a^\dagger - \frac{1}{N} \sum_{k_1=0}^k a_{k_1}^\dagger \quad \text{for } k > 0 \quad (20)$$

We simplify these commutation relations in the spirit of Tomonaga's approximation, i.e., the matrix elements

$$\left\langle f_{n+1} \left| \frac{1}{N} \sum_{k_1=0}^k a_{k_1}^\dagger \right| i_n \right\rangle$$

are considered to be small. This is correct within the Tomonaga model and means that the energy of the final states has to be small compared to the bandwidth. So, we shall use

$$[\rho_k, a^\dagger] = (1/\sqrt{N}) a^\dagger \quad \text{for all } k > 0 \quad (21)$$

instead of the exact commutation relations [Eqs. (19) and (20)]. Defining

$$U = \exp \left[ \sum_k \alpha_k (b_k^\dagger - b_k) \right], \quad \alpha_k \text{ real} \quad (22)$$

---

<sup>05</sup> The tacit assumption is the completeness of the  $k$ , which is only fulfilled in the subspace of  $s$ -wave states.

we have a canonical transformation which acts as a translation operator on any function of  $b_k^\dagger$  and  $b_k$ , namely,

$$U^\dagger b_k U = b_k + \alpha_k \quad \text{or} \quad [b_k, U] = \alpha_k U \quad (23)$$

This relation is identical to (21), if we put  $\alpha_k = 1/\sqrt{(kN)}$  and

$$a^\dagger \propto \exp \left[ \sum_k \frac{1}{\sqrt{(kN)}} (b_k^\dagger - b_k) \right] \quad (24)$$

### 2.3 3. TRANSITION PROBABILITY IN THE DENSITY WAVE MODEL

Before calculating the transition rate, let us ask how to interpret the Golden Rule [Eq. (3)] in the new description. The initial state is the lowest eigenstate of a set of harmonic oscillators [Eq. (17)]-the analog to the "quiescent" Fermi sea. The final states are all possible excitations of another set of oscillators with shifted zero points [ see Eq. (18)]. The ground state of this second set is also the analog to a quiescent Fermi sea, which is (in the language of the fermion description) a Slater determinant of scattering waves. Excitations in both oscillator systems can be pictured as density waves.

According to (24),  $a^\dagger$  acts as a shift operator distorting, say, the initial quiescent Fermi sea. That is, by injecting the core-state electron into the band, a local inhomogeneity of the electron density is created, by which each oscillator is elongated by a small amount. At the same time, the oscillators have to react to the core-hole potential  $V$  produced in the absorption process. This is described by the zero-point shift of the second set of oscillators. As will be seen later these two effects are additive. This picture of the problem has some similarity to the "small polaron," where now the real lattice is replaced by the band electrons, and the real phonons are replaced by Tomonaga's "phonons."

Instead of the transition rate [Eq. (3)], we calculate the correlation function

$$\mathcal{F}(t) = \left\langle i \left| \exp \left( +i\tilde{H}_i t \right) a \exp \left( -i\tilde{H}_f t \right) a^\dagger \right| i \right\rangle \quad (25)$$

which is related to  $W(\omega)$  by

$$W(\omega) \propto \text{Im} -\frac{i}{\pi} \int_0^\infty e^{i(\omega+E_0)t} \mathcal{F}(t) dt \quad (26)$$

The canonical transformation

$$U_V = \exp \left[ V \rho \sum_k \frac{1}{\sqrt{(kN)}} (b_k^\dagger - b_k) \right] \quad (27)$$

transforms  $\tilde{H}_i$  into  $\tilde{H}_f$  :

$$\tilde{H}_f = U_V^\dagger \tilde{H}_i U_V \quad (28)$$

so that

$$\mathcal{F}(t) = \left\langle i \left| \exp \left( +i\tilde{H}_i t \right) a U_V^\dagger \exp \left( -i\tilde{H}_i t \right) U_V a^\dagger \right| i \right\rangle, \quad (29)$$

with

$$Ua^\dagger \equiv B = \exp \left[ (1 + V\rho) \sum_k \frac{1}{\sqrt{(kN)}} (b_k^\dagger - b_k) \right] \quad (30)$$

and

$$B(t) = \exp \left[ (1 + V\rho) \sum_k \frac{1}{\sqrt{(kN)}} \times \left( b_k^\dagger e^{+i(k/\rho)t} - b_k e^{-i(k/\rho)t} \right) \right] \quad (31)$$

We have

$$\mathfrak{F}(t) = \langle i | B^\dagger(t) B(0) | i \rangle \quad (32)$$

which is equal to <sup>6</sup>

$$\mathfrak{F}(t) = \exp \left( (1 + V\rho)^2 \frac{1}{N} \sum_{k>0} \frac{1}{k} (e^{-i(k/\rho)t} - 1) \right); \quad (33)$$

converting the sum in an integral, introducing a cutoff <sup>7</sup>

$$\frac{1}{N} \sum_{k>0}^{k_{\max}} \frac{1}{k} (e^{-i(k/\rho)t} - 1) = \int_0^{itk_{\max}(1/\rho)} \frac{e^{-x} - 1}{x} dx \quad (34)$$

we find as the leading term for large  $t$

$$\mathfrak{F}(t) \sim \left[ t^{(1+V\rho)^2} \right]^{-1}. \quad (35)$$

Inserting  $-V\rho = \delta_B/\pi$ , where  $\delta_B$  is the phase shift in Born approximation, we finally get for the energy dependence near the threshold

$$W(\omega) \sim (\omega + E_0)^{-[2\delta_B/\pi - (\delta_B/\pi)^2]} \quad (36)$$

Thus, we have recovered the main feature of Nozière's result, namely, the singular threshold behavior of the response function. But in the correct answer  $\delta_B$  is replaced by  $\delta$ , the exact phase shift. In this context it is amusing to note that in the Luttinger model, <sup>8</sup> where an infinite energy spectrum with a linear dispersion is assumed from the start,  $\delta_B$  is the correct phase shift. The Schrödinger equation (for  $s$  waves) for a linear dispersion reads

$$[-(i/\rho)(d/dx) + V(x)]\Psi(x) = (E - E_F)\Psi(x) \quad (37)$$

and the solution is given by

$$\Psi(x) = \exp \left[ ikx - i\rho \int_0^x V(x) dx \right] \quad (38)$$

so that

$$\Psi(x) \sim e^{ikx+i\delta} \quad \text{as} \quad x \rightarrow \infty \quad (39)$$

with

$$\delta = -\rho \int_0^\infty V(x) dx = -\pi\rho V \quad (40)$$

•

### 3. The kondo lattice and weak antiferromagnetism by Doniach (1977)

#### Abstract

By considering a one-dimensional analog of a system of conduction electrons exchange coupled to a localized spin in each cell of a lattice, it is suggested that a second-order transition from an antiferromagnetic to a Kondo spincompensated ground state will occur as the exchange coupling constant  $J$  is increased to a critical value  $J_c$ . For systems in which  $J \leq J_c$ , a very weak sublattice magnetization may occur as a result of nearly complete spin-compensation.

#### 3.1 1. The Kondo lattice

One of the fundamental mysteries in the magnetic behavior of fractional valence compounds—of which examples are  $\beta\text{Ce}$ ,  $\text{CeAl}_2$ , and  $\text{SmS}$  under pressure—is the lack of magnetic order at low temperatures despite the good evidence for unpaired f-spins (based on atomic volume). A number of authors [1-3] have suggested that Kondo pairing will occur in concentrated systems of local moments, even though there is of the order of one magnetic atom per unit cell, a situation to which previous theoretical studies of the Kondo effect, confined to isolated magnetic impurities in a non magnetic host, would not apply.

In this paper a start is made on exploring the theory of a system in which there is a local spin/conduction-electron exchange interaction in every cell of a crystal. This will be termed a "Kondo lattice." The main result, based on the theory of a one-dimensional analog, is that there should be a second-order transition at zero temperature, as the exchange coupling is varied, between an antiferromagnetic ground state for weak  $J$  and a Kondo-like state, in which the local moments are quenched, for strong  $J$ .

The fact that the transition is second order implies that the sublattice moment order parameter goes to zero in a continuous fashion as  $J$  is increased to a critical value,  $J_c$ . Thus, for  $J$  near to, and slightly smaller than  $J_c$ , there will exist antiferromagnets with very weak, "nearly quenched" moments, even though the "f electrons are in a state with a well-defined local non-zero spin state.

A qualitative understanding of the existence of this transition may be had by comparing the binding energy of a Kondo singlet

$$W_K \sim N(0)^{-1} e^{-1/N(0)J}$$

with that of a Ruderman-Kittel-Yosida antiferromagnetic state

$$W_{\text{AF}} \sim C J^2 N(0),$$

where  $N(0)$  is the density of conduction-electron states and  $C$  is a dimensionless constant depending on details of the band structure. For  $J_N(0)$  less than a critical value the RKY state dominates, while above this, the Kondo singlet binding dominates [4] (see fig. 1).

<sup>06</sup> To evaluate an expression of the form  $\langle e^A e^B \rangle$  [compare (32)] one makes use of the well-known relations  $e^A e^B = e^{A+B+(1/2)[A,B]}$ , which holds if  $[A,B]$  is a  $c$  number, and  $\langle e^{L(b^\dagger, b)} \rangle = \exp \frac{1}{2} \langle L^2(b^\dagger, b) \rangle$ , where  $L$  is any linear combination of Bose operators.

<sup>7</sup> According to the Tomonaga model,  $k_{\text{max}} \approx \frac{1}{2} k_f$ .

<sup>8</sup> J. M. Luttinger, J. Math. Phys. 4, 1154 (1963).

0

• Research supported in part by the Army Research Office, Durham, NC.

### 3.2 2. A one-dimensional analog-the "Kondo necklace"

In order to make progress on studying the Kondo lattice theoretically, the study of the following model is suggestive:

$$\begin{aligned} H &= J \sum_i \mathbf{S}_i \cdot \boldsymbol{\tau}_i + W \sum_i (\tau_i^x \tau_{i+1}^x + \tau_i^y \tau_{i+1}^y) \\ &= H_J + H_W. \end{aligned} \quad (3)$$

Here the one-dimensional electron gas has been replaced by a set of psuedo spins  $\boldsymbol{\tau}_i$  on a linear

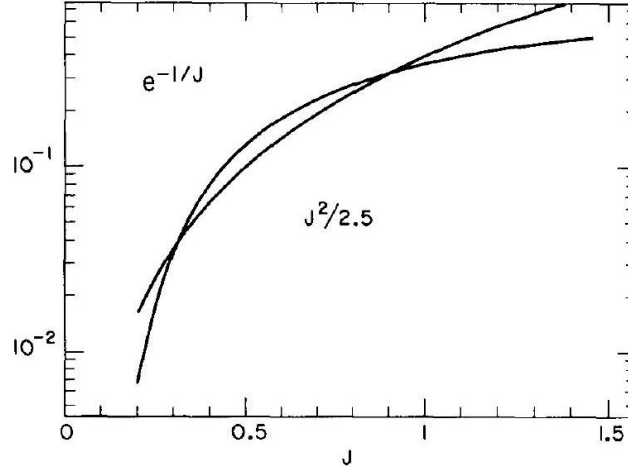


Fig. 1. Comparison of AF with Kondo energies.

lattice. ( $\mathbf{S}_i$  and  $\boldsymbol{\tau}_i$  are independent sets of Pauli spin-  $\frac{1}{2}$  operators.) As shown by Luther and Peschel [5], the one-dimensional spinless fermion gas (Luttinger model) is equivalent under the Jordan-Wigner transformation to a chain of spins,  $H_w$ . Note that the pseudospin chain coupled to the set of localized spins  $\mathbf{S}$  is not identical to the original Kondo hamiltonian. In particular, the rotational invariance of the original hamiltonian is broken. However, in the weak coupling limit the scaling behavior of an individual spin "bead",  $\mathbf{S}$ , on a chain of pseudospins  $\boldsymbol{\tau}_i$ , is found to be equivalent to that of the original hamiltonian (for a single Kondo spin) on making the Yuval-Anderson transformation [6, 7] (see Appendix). I think the reason that this works is that in the Kondo hamiltonian in one dimension, the spin  $\mathbf{S}$  couples mainly to the spin density modes of the one-dimensional electron gas, so the principal excitations involved in the weak coupling limit are those of a set of spin density bosons [8].

### 3.3 3. A mean field solution for the Kondo necklace

Although the mean field approximation is quantitatively misleading in one dimension, it serves to estimate what kind of symmetry breaking is important as a function of the ratio  $J/W$ . The actual value of the critical coupling for the transition  $(J/W)_c$  will be lower than the mean field value on account of fluctuation effects.

A variational trial function for the onedimensional antiferromagnet is

$$\Psi_{AF} = \prod \varphi_A^i \varphi_B^{i+1}$$

where

$$\varphi_A = as + bt^- + ct^0 + dt^+,$$

$$\varphi_B = -as - dt + ct' - bt^+,$$

are trial functions for the  $A$  and  $B$  sublattices and  $s, t$  are the singlet and triplet states of an isolated cell of the Kondo necklace

$$s = 1/\sqrt{2}(\alpha d - \beta u), \quad t^\circ = 1/\sqrt{2}(\alpha d + \beta u), \\ t^- = \beta d, \quad t^+ = \alpha u,$$

in which  $(\alpha, \beta), (u, d)$  refer to the up and down eigenstates of the  $S^z$  and  $\tau^z$  operators, respectively. A study of the general variational problem (4) leads to the result that  $c = 0$  and  $d = -h$ , leading to the expression to be minimized (the operators are defined by  $\tau^* = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ . etc.):  $\langle H \rangle = 2Jb^2 - 3Ja^2 - 8Wa^2h^2$

subject to

$$a^2 + 2b^2 = 1.$$

This leads to a result for the sub-lattice order parameter of

$$\langle S' \rangle_A = -\langle S^x \rangle_B = 4ab = \sqrt{1 - (J/W)}^2$$

so that  $(J/W)_c = 1$ . Below this value the sublattice magnetization grows in a continuous fashion as  $J/W$  is decreased [13]. Above this value, the singlet state  $II s_i$  is the ground state, i.e. the system is in a quenched Kondo-like state for which  $\langle \tau^2 \rangle = 0$  (still corresponding to a half-filled fermion "band"). In general, fluctuation effects will be expected to strongly reduce the critical ratio.

The model here is to be contrasted with an apparently similar one due to Blume [9] for a system consisting of a set of singlet ground states plus triplet exciton excited states which are exchange coupled to each other but not to the singlet states. Blume's model has a first-order transition to a magnetic ground state as the exchange parameter is increased. In contrast, the present model mixes the singlet states with the triplet via  $H_w$ , thus leading to the above second-order transition (in the mean field approximation).

### 3.4 4. The excitation spectrum near the AF to Kondo transition

The above model has the property that its spin density excitation spectrum develops an energy gap (the singlet-triplet splitting) for  $J > J_c$ . (In the metallic case there would still be charge density modes with zero excitation energy at the Fermi surface.) As  $J \rightarrow J_c$ , this gap will reduce to zero, leading to large spin-fluctuation effects.

To examine the character of the excitation spectrum in an RPA-like calculation (a dynamic extension of the above mean field calculation), consider the propagator

$$g_{ij}(t) = i\theta(t) \langle \tilde{\tau}_i^x(t) \tilde{\tau}_j^x(0) \rangle.$$

where  $\tilde{\tau}^x(t)$  is a Heisenberg operator. The time

dependence may be formally expanded in powers of  $H_W$  using Schrödinger perturbation theory, and the resulting terms evaluated using semi-invariants [10]. By keeping only terms involving second-order semi-invariants,

$$g^0(t) = i\theta(t) \langle \tau_i^x(t) \tau_i^x(0) \rangle,$$

where  $\tau^x(t) = e^{iH_{\delta t} \tau^x} e^{-iH_{\mu t}}$ , and neglecting all higher order semi-invariants, one finds, for terms in which intermediate time labels are not permuted, an expression for the Fourier transform of  $g_{ij}(t)$  :

$$g^{-1}(q, \omega) = \omega - 4J - 4W \cos qa,$$

where  $a$  is the interatomic spacing. So the effective gap in the spectrum at the antiferromagnetic zone boundary,  $q = \pm\pi/a$ , is  $(4J - 4W)$  which tends to zero as  $J/W \rightarrow 1$ .

### 3.5 5. Application to real systems

The replacement of the electron gas by a set of spins used in the above discussion cannot be generalized to three dimensions. Furthermore, fractional valence effects in which the f-state electron is hopping in and out of a d-band have not been considered at all. So the above results

should not be taken literally; but I believe they are suggestive of what might happen in real three-dimensional systems.

The fact that Kondo-like quenching of local moments appears to occur for fractional valence systems is consistent with the above ideas on the empirical ground that only when the f-level is degenerate with the d-band is the effective Schrieffer-Wolff exchange interaction likely to be strong enough to satisfy the above criterion for a non magnetic ground state of  $JN(0) = 0(1)$ .

I am grateful to J. Fields, M. Weinstein, and P. Anderson for helpful comments. This work was started while I was on leave at the Université de Paris, Orsay. I would like to thank J. Friedel and other members of the Groupe de Physique des Solides for their hospitality.

### 3.6 Appendix

Equivalence of the weak coupling scaling behavior of the pseudospin model and that of the Kondo problem

To show the equivalence of the scaling behavior for a system consisting of a single local spin,  $S$ , coupled to a chain of pseudospins, the ground state energy may be written as an expansion in powers of the transverse coupling  $J_\perp$  following Yuval and Anderson. The hamiltonian is expressed as the sum of

$$H_0 = W \sum_i (\tau_i^x \tau_{i+1}^x + \tau_i^y \tau_{i+1}^y) + JS^z \tau_0^z$$

and

$$H_1 = J_\perp (S^+ \tau_0^- + S^- \tau_0^+).$$

In the weak coupling approximation, the ground to ground state amplitude

$$F(t) = \langle \Psi_0 \uparrow | e^{iH_0 t} T \left( \exp i \int_0^t H_1(\tau) d\tau \right) | \Psi_0 \uparrow \rangle$$

can be calculated for leading asymptotic behavior using the boson representation [following Luther and Peschel (eq. 15)] with

$$\tau_0^+(t) = \exp(-\phi_1(t)/2 - \phi_2(t)/2),$$

where  $\phi_1$  and  $\phi_2$  are boson operators for forward and backward bands of the Luttinger model

$$\phi_{1,2}(x=0) = \mp 2\pi L^{-1} \sum_k \frac{1}{k} \rho_{1,2}(k) e^{-\alpha|k|/2},$$

in which  $\rho_{1,2}(k)$  are boson density operators in terms of which the pseudospin hamiltonian is diagonal:

$$\begin{aligned} H_{\text{ps}} &= \sum_i (\tau_i^x \tau_{i+1}^x + \tau_i^y \tau_{i+1}^y) \\ &= 2\pi L^{-1} \sum_k [\rho_1(-k) \rho_1(k) + \rho_2(-k) \rho_2(k)] \end{aligned} \quad (\text{A5})$$

where

$$[\rho_1(-k), \rho_1(k')] = kL \delta_{kk'}/2\pi.$$

Terms in the expansion of (A2) in powers of  $H_1$  involve ground state averages of which the second order term is

$$F_2(t_1 t_2) = (iJ_1)^2 \langle \tau^+(t_2) e^{iH_0(t_2-t_1)} \tau^-(t_1) \rangle.$$

To lowest order in  $J, \tau^Z$  appearing in  $H_0$  may be replaced by

$$\tau^z \rightarrow \rho_1 + \rho_2 \quad (\text{A7})$$

The backscattering terms in Luther and Peschel, eq. (10), do not contribute to lowest order in  $J$ .]Eq.(A6)thenfactorizesintotwo identical terms for band 1 and band 2 operators, leading to

$$F_2(t_1 t_2) = G(t_1 t_2)^2$$



where

$$G(t_1 t_2) = \exp [(1 - J/W) \ln |t_1 - t_2|].$$

As explained by Schotte and Schotte [11] this is the Born approximation for the corresponding X-ray problem [12]. Similar results may be obtained for the general term in the expansion of  $F(t)$ , leading to the weak coupling limit of the Yuval-Anderson formula in which the Kondo problem is expressed as the partition function of a set of charges interacting via a logarithmic potential of strength  $(2 - 2\epsilon)$ , where  $\epsilon = J/W$  in the weak coupling limit.

## Part V

# Adds

### 1. Дополнения

#### 1..1 Литература

#### 1..2 Refs rational

## References

- [1] R.B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983).
- [2] B.I. Halperin, Phys. Rev. Lett. **52**, 1583 (1984).
- [3] D. Arovas, J.R. Schrieffer, and F. Wilczek, Phys. Rev. Lett. **53**, 722 (1984).
- [4] L. Saminadayar, D.C. Glattli, Y. Jin, and B. Etienne, Phys. Rev. Lett. **79**, 2526 (1997).
- [5] R. de-Picciotto, M. Reznikov, M. Heiblum, V. Umansky, G. Bunin, and D. Mahalu, Nature **389**, 162 (1997).
- [6] K.-V. Pham, M. Gabay, and P. Lederer, Phys. Rev. B **61**, 16397 (2000).
- [7] K.-I. Imara, K.-V. Pham, P. Lederer, and F. Piéchon, Phys. Rev. B **66**, 035313 (2002).
- [8] H. Steinberg, G. Barak, A. Yacoby, L.N. Pfeiffer, K.W. West, B.I. Halperin, and K. Le Hur, Nat. Phys. **4**, 116 (2008).
- [9] H. Kamata, N. Kumada, M. Hashisaka, K. Muraki, and T. Fujisawa, Nat. Nanotech. **9**, 177 (2014).
- [10] H. Inoue, A. Grivnin, N. Ofek, I. Neder, M. Heiblum, V. Umansky, and D. Mahalu, Phys. Rev. Lett. **112**, 166801 (2014).
- [11] E.J. Bergholtz and A. Karlhede, Phys. Rev. Lett. **94**, 026802 (2005).
- [12] A. Seidel, H. Fu, D.-H. Lee, J.M. Leinaas, and J. Moore, Phys. Rev. Lett. **95**, 266405 (2005).
- [13] A. Seidel and D.-H. Lee, Phys. Rev. Lett. **97**, 056804 (2006).
- [14] E.J. Bergholtz, J. Kailasvuori, E. Wikberg, T.H. Hansson, and A. Karlhede, Phys. Rev. B **74**, 081308(R) (2006).
- [15] D.-H. Lee, G.-M. Zhang, and T. Xiang, Phys. Rev. Lett. **99**, 196805 (2007).
- [16] X.-L. Qi, T.L. Hughes, and S.-C. Zhang, Nature **4**, 273 (2008).
- [17] X.-L. Qi, T.L. Hughes, and S.-C. Zhang, Phys. Rev. B **78**, 195424 (2008).
- [18] S. Ryu, C. Mudry, C.Y. Hou, and C. Chamon, Phys. Rev. B **80**, 205319 (2009).
- [19] N. Goldman, I. Satija, P. Nikolic, A. Bermudez, M.A. Martin-Delgado, M. Lewenstein, and I.B. Spielman, Phys. Rev. Lett. **105**, 255302 (2010).

- [20] J. Klinovaja and D. Loss, Phys. Rev. B **92** 121410(R) (2015).
- [21] A. Przysiężna, O. Dutta, and J. Zakrzewski, New J. of Phys. **17**, 013018 (2015).
- [22] R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).
- [23] W.P. Su, J.R. Schrieffer, and A.J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979); *ibid.* Phys. Rev. B **22**, 2099 (1980).
- [24] R. Jackiw and J.R. Schrieffer, Nucl. Phys. B **190**, 253 (1981).
- [25] M.J. Rice and E.J. Mele, Phys. Rev. Lett. **49**, 1455 (1982).
- [26] S. Kivelson, Phys. Rev. B **28**, 2653 (1983).
- [27] W.P. Su and J.R. Schrieffer, Phys. Rev. Lett. **46**, 738 (1981).
- [28] A.J. Heeger, S. Kivelson, J.R. Schrieffer, and W.-P. Su, Rev. Mod. Phys. **60**, 781 (1988).
- [29] E. Witten, Phys. Letters B **86**, 283 (1979).
- [30] H. Takayama, Y.R. Lin-Liu, and K. Maki, Phys. Rev. B **21**, 2388 (1980).
- [31] J. Goldstone and F. Wilczek, Phys. Rev. Lett. **47**, 986 (1981).
- [32] R. Jackiw and G. Semenoff, Phys. Rev. Lett. **50**, 439 (1983).
- [33] S.A. Kivelson and J.R. Schrieffer, Phys. Rev. B **25**, 6447 (1982).
- [34] J.S. Bell and R. Rajaraman, Phys. Lett. B **116**, 151 (1982).
- [35] S.A. Kivelson, Phys. Rev. B **26**, 4269 (1982).
- [36] J.S. Bell and R. Rajaraman, Nucl. Phys. B **220**, 1 (1983).
- [37] Y. Frishman and B. Horowitz, Phys. Rev. B **27**, 2565 (1983).
- [38] R. Jackiw, A.K. Kerman, I. Klebanov, and G. Semenoff, Nucl. Phys. B **225**, 233 (1983).
- [39] J.-H. Park, G. Yang, J. Klinovaja, P. Stano, and D. Loss, Phys. Rev. B **94**, 075416 (2016).
- [40] J. Zak, Phys. Rev. Lett. **48**, 359 (1982); *ibid.* **62**, 2747 (1989).
- [41] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. W. Ludwig, New J. Phys. **12**, 065010 (2010).
- [42] R.D. King-Smith and D. Vanderbilt, Phys. Rev. B(R) **47**, 1651 (1993).
- [43] D. Vanderbilt and R.D. King-Smith, Phys. Rev. B **48**, 4442 (1993).
- [44] R. Resta, Ferroelectrics **136**, 51 (1992); *ibid.* Europhys. Lett. **22**, 133 (1993).
- [45] R. Resta, Rev. Mod. Phys. **66**, 899 (1994).
- [46] N. Marzari, A.A. Mostofi, J.R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012).
- [47] D. Vanderbilt, *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators*, (Cambridge University Press, 2018).

- [48] G. Ortiz and R.M. Martin, Phys. Rev. B **49**, 14202 (1994).
- [49] J.-W. Rhim, J. Behrends and J.H. Bardarson, Phys. Rev. B **95**, 035421 (2017).
- [50] G. van Miert and C. Ortix, Phys. Rev. B **96**, 235130 (2017).
- [51] T. L. Hughes, E. Prodan, and B. A. Bernevig, Phys. Rev. B **83**, 245132 (2011).
- [52] C.-K. Chiu, H. Yao, and S. Ryu, Phys. Rev. B **88**, 075142 (2013).
- [53] K. Shiozaki and M. Sato, Phys. Rev. B **90**, 165114 (2014).
- [54] A. Alexandradinata, Zhijun Wang, and B. A. Bernevig, Phys. Rev. X **6**, 021008 (2016).
- [55] L. Trifunovic and P. Brouwer, Phys. Rev. B **96**, 195109 (2017).
- [56] A. Lau, C. Ortix, and J. van den Brink, Phys. Rev. Lett. **115**, 216805 (2015); A. Lau, J. van den Brink, and C. Ortix, Phys. Rev. B **94**, 165164 (2016); A. Lau and C. Ortix, Eur. Phys. J. Spec. Top. **227**, 1309 (2018).
- [57] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).
- [58] A. Kitaev, in Advances in Theoretical Physics, edited by V. Lebedev, and M. Feigel'man, AIP Conf. Proc. No. 1134 (AIP, New York, 2009), p. 22.
- [59] R.-J. Slager, A. Mesaros, V. Juričić, and J. Zaanen, Nat. Phys. **9**, 98 (2012).
- [60] P. Jadaun, D. Xiao, Q. Niu, and S. K. Banerjee, Phys. Rev. B **88**, 085110 (2013).
- [61] C.-K. Chiu, H. Yao, and S. Ryu, Phys. Rev. B **88**, 075142 (2013).
- [62] F. Zhang, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **111**, 056403 (2013).
- [63] W. A. Benalcazar, J.C.Y. Teo, and T. L. Hughes, Phys. Rev. B **89**, 224503 (2014).
- [64] T. Morimoto and A. Furusaki, Phys. Rev. B **88**, 125129 (2013).
- [65] M. Diez, D. I. Pikulin, I.C. Fulga, and J. Tworzydło, New J. Phys. **17**, 043014 (2015).
- [66] C.-X. Liu, R.-X. Zhang, and B.K. VanLeeuwen, Phys. Rev. B **90**, 085304 (2014).
- [67] S.M. Young and C.L. Kane, Phys. Rev. Lett. **115**, 126803 (2015).
- [68] Z. Wang, A. Alexandradinata, R.J. Cava, and B.A. Bernevig, Nature **532**, 189 (2016).
- [69] K. Shiozaki, M. Sato, and K. Gorni, Phys. Rev. B **93**, 195413 (2016).
- [70] P.-Y. Chang, O. Erlen, and P. Coleman, Nat. Phys. **13**, 794 (2017).
- [71] B.J. Wieder, B. Bradlyn, Z. Wang, J. Cano, Y. Kim, H.-S.D. Kim, A.M. Rappe, C.L. Kane, and B.A. Bernevig, Science **361**, 246 (2018).
- [72] W. Kohn, Phys. Rev. Lett. **76** 3168 (1996).
- [73] E. Prodan and W. Kohn, PNAS **102**, 11638 (2005).
- [74] L. Fidkowski, T.S. Jackson, and I. Klich, Phys. Rev. Lett. **107**, 036601 (2011).

- [75] R.S.K. Mong and V. Shivamoggi, Phys. Rev. B **83**, 125109 (2011).
- [76] V. Gurarie, Phys. Rev. B **83**, 085426 (2011).
- [77] A.M. Essin and V. Gurarie, Phys. Rev. B **84**, 125132 (2011).
- [78] T. Fukui, K. Shiozaki, T. Fujiwara, and S. Fujimoto, J. Phys. Soc. Jpn. **81**, 114602 (2012).
- [79] Y. Yu, Y.-S. Wu, and X. Xie, Nucl. Phys. B **916**, 550 (2017).
- [80] J.-W. Rhim, J.H. Bardarson, and R.-J. Slager, Phys. Rev. B **97**, 115143 (2018).
- [81] M. Silveirinha, Phys. Rev. X **9**, 011037 (2019).
- [82] D.J. Thouless, Phys. Rev. B **27**, 6083 (1983).
- [83] Q. Niu and D.J. Thouless, J. Phys. A **17**, 2453 (1984).
- [84] C. Kallin and B.I. Halperin, Phys. Rev. B **29**, 2175 (1984).
- [85] M. Thakurathi, J. Klinovaja, and D. Loss, Phys. Rev. B **98**, 245404 (2018).
- [86] M. Pletyukhov, D.M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, arXiv:1911.06890, to appear in Phys. Rev. B (RC).
- [87] M. Pletyukhov, D.M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, arXiv:1911.06886, to appear in Phys. Rev. B.
- [88] P. Mara, R. Citro, and C. Ortix, Phys. Rev. B **91**, 125411 (2015).
- [89] M. Atala, M. Aidelsburger, J.T. Barreiro, D. Abanin, T. Kitagawa, E. Demler, and I. Bloch, Nat. Phys. **9**, 795 (2013).
- [90] E.J. Meier, F.A. An, and B. Gadway, Nat. Comm. **7**, 13986 (2016).
- [91] Y. Efroni, S. Ilani, and E. Berg, Phys. Rev. Lett. **119**, 147704 (2017).
- [92] S.-R.E. Yang, Nanomaterials **9**, 885 (2019).
- [93] J. Klinovaja, P. Stano, and D. Loss, Phys. Rev. Lett. **109**, 236801 (2012).
- [94] D. Rainis, A. Saha, J. Klinovaja, L. Trifunovic, and D. Loss, Phys. Rev. Lett. **112**, 196803 (2014).
- [95] M.J. Yoo, T.A. Fulton, H.F. Hess, R.L. Willett, L.N. Dunkleberger, R.J. Chichester, L.N. Pfeiffer, and K.W. West, Science **276** 579 (1997).
- [96] S.H. Tessmer, P.I. Glicofridis, R.C. Ashoori, L.S. Levitov, and M.R. Melloch, Nature **392**, 51 (1998).
- [97] G. Finkelstein, P.I. Glicofridis, R.C. Ashoori, and M. Shayegan, Science **289**, 90 (2000).
- [98] G. Ben-Shach, A. Haim, I. Appelbaum, Y. Oreg, A. Yacoby, and B.I. Halperin, Phys. Rev. B **91**, 045403 (2015).
- [99] M. Xiao, G. Ma, Z. Yang, P. Sheng, Z.Q. Zhang, and C.T. Chan, Nat. Phys. **11**, 240 (2015).

- [100] We note that  $\varphi$  is the precise definition of the phase of the CDW, whereas  $\alpha$  is the phase of the complex gap parameter entering the Dirac model. In general they are different. Away from half-filling, they are related by  $\alpha = \pm\nu\varphi + \text{const}$ , see Eq. (2.82). However, at half-filling,  $\alpha$  has a different meaning, it can be even pinned independent of  $\varphi$ , see Eq. (2.86).
- [101] see Supplemental Material, where the transformation laws for the boundary charge are derived via the representation in terms of the Zak-Berry phase.
- [102] B.-J. Yang, T.A. Bojesen, T. Morimoto, and A. Furusaki, Phys. Rev. B **95**, 075135 (2017).
- [103] W. Brzezicki and M. Cuoco, Phys. Rev. B **95**, 155108 (2017).
- [104] J. Zhang, Y.-H. Chan, C.-K. Chiu, M.G. Vergniory, L.M. Schoop, and A.P. Schnyder, Phys. Rev. Mater. **2**, 074201 (2018).
- [105] M. Malard, P.E. de Brito, S. Östlung, and H. Johannesson, Phys. Rev. B **98**, 165127 (2018).
- [106] Y.-X. Zhao and A.P. Schnyder, Phys. Rev. B **94**, 195109 (2016).
- [107] D.J. Thouless, M. Kohmoto, M.P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).
- [108] I. Dana, Y. Avron, and J. Zak, J. Phys. C **18**, L679 (1985).
- [109] M. Kohmoto, Phys. Rev. B **39**, 11943 (1989); *ibid.*, J. Phys. Soc. Jpn. **61**, 2645 (1992).
- [110] Y. Hatsugai, Phys. Rev. B **48**, 11851 (1993).
- [111] S. Gangadharaiah, L. Trifunovic, and D. Loss, Phys. Rev. Lett. **108**, 136803 (2012).
- [112] S. Kivelson, H. B. Thacker and W.-K. Wu, Phys. Rev. B **31**, 3785 (1985).
- [113] B. Horovitz and J. Sólyom, Phys. Rev. B **32**, 2681 (1985).
- [114] W.-K. Wu and S. Kivelson, Phys. Rev. B **33**, 8546 (1986).
- [115] J. Klinovaja and D. Loss, Eur. Phys. J. B **87**, 171 (2014).
- [116] A. Gogolin, A. Nersesyan, and A. Tsvelik, Bosonization approach to strongly correlated systems. Cambridge: Cambridge University Press, (1998).
- [117] J. von Delft and H. Schoeller, Annalen Phys. **7**, 225 (1998).
- [118] T. Giamarchi, Quantum physics in one dimension. Oxford: Clarendon Press, (2010).
- [119] A. E. Mattsson, S. Eggert, and H. Johannesson, Phys. Rev. B **56**, 15615 (1997).
- [120] Y.-T. Lin, D.M. Kennes, M. Pletyukhov, H. Schoeller, and V. Meden, submitted to Phys. Rev. B, unpublished.
- [121] K. Piasotski, M. Pletyukhov, J. Klinovaja, D. Loss, H. Schoeller, and D.M. Kennes, unpublished.

- [122] In Ref. [87] it was shown that  $Q_{B,0}^R + Q_{B,Z-1}^L = -1 + \frac{1}{Z}$ . In addition, it was shown that  $Q_{B,Z-1}^L = Q_{B,0}^L - \frac{Z-1}{Z} - w$ , where  $w$  is the winding number of the phase difference of the Bloch wave function between the first and last site of the unit cell. This gives the result  $Q_{B,0}^R + Q_{B,0}^L = w$ .
- [123] N. Müller, D.M. Kennes, K. Piasotski, M. Pletyukhov, J. Klinovaja, D. Loss, and H. Schoeller, unpublished.
- [124] M. Serina, D. Loss, and J. Klinovaja, Phys. Rev. B **98**, 035419 (2018).
- [125] M. Estarellas, I. D’Amico, and T. Spiller, Sci. Rep. **7**, 42904 (2017).

### 1.3 Refs universality ...

## References

- [1] J.-H. Park, G. Yang, J. Klinovaja, P. Stano, and D. Loss, Phys. Rev. B **94**, 075416 (2016).
- [2] M. Thakurathi, J. Klinovaja, and D. Loss, Phys. Rev. B **98**, 245404 (2018).
- [3] M. Pletyukhov D. M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, Phys. Rev. B **101**, 165304 (2020).
- [4] M. Pletyukhov D. M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, Phys. Rev. B **101**, 161106(R) (2020).
- [5] Y.-T. Lin, D.M. Kennes, M. Pletyukhov, C. Weber, H. Schoeller, and V. Meden, arXiv:2004.06325, to be published in Phys. Rev. B.
- [6] M. Pletyukhov, D. M. Kennes, K. Piasotski, J. Klinovaja, D. Loss, and H. Schoeller, arXiv:2004.00463, to be published in Phys. Rev. Research.
- [7] S.A. Kivelson and J.R. Schrieffer, *Fractional charge, a sharp quantum observable*, Phys. Rev. B **25**, 6447 (1982).
- [8] J.S. Bell and R. Rajaraman, *On solitons with half integral charge*, Phys. Lett. B **116**, 151 (1982).
- [9] S.A. Kivelson, *Wannier functions in one-dimensional disordered systems: Application to fractionally charged solitons*, Phys. Rev. B **26**, 4269 (1982).
- [10] J.S. Bell and R. Rajaraman, *On states, on a lattice, with half-integral charge*, Nucl. Phys. B **220**, 1 (1983).
- [11] Y. Frishman and B. Horovitz, *Charge fluctuations and fractional charge of fermions in 1 + 1 dimensions*, Phys. Rev. B **27**, 2565 (1983).
- [12] R. Jackiw, A.K. Kerman, I. Klebanov, and G. Semenoff, *Fluctuations of fractional charge in soliton anti-soliton systems*, Nucl. Phys. B **225**, 233 (1983).
- [13] J.J. Rehr and W. Kohn, Phys. Rev. B **10**, 448 (1974).
- [14] C. Kallin and B.I. Halperin, Phys. Rev. B **29**, 2175 (1984).
- [15] L. He and D. Vanderbilt, Phys. Rev. Lett. **86**, 5341 (2001).



- [16] N.D.M. Hine and W.M.C. Foulkes, J. Phys. C **19**, 506212 (2007).
- [17] H. Min and A.H. MacDonald, Progr. of Theor. Phys. Supplement **176**, 227 (2008).
- [18] C.W. Groth, M. Wimmer, A.R. Akhmerov, J. Tworzydło, and C.W.J. Beenakker, Phys. Rev. Lett. **103**, 196805 (2009).
- [19] C.S. Weber *et al.*, in preparation.
- [20] D.J. Thouless, M. Kohmoto, M.P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).
- [21] R. Resta, Phys. Rev. Lett. **96**, 137601 (2006).

#### 1.4 Refs magn field

## References

- [1] G. Grüner, Density waves in Solids Perseus Publishing; 1st edition (January 15, 2000).
- [2] P. Monceau, Electronic Properties of Inorganic Quasi One-Dimensional Compounds, D. Reidel Pub. Co.; (April 2002).
- [3] See e.g., the books: E. Fradkin, Field Theories of Condensed Matter Systems, Westview Press (March, 1998); Alexander O. Gogolin, Alexander A. Nersisyan, Alexei M. Tsvelik, Bosonization Approach to Strongly Correlated Systems, Cambridge University Press (December 10, 1998).
- [4] One illustrating example that the strictly 1D models are not applicable to describe actual Q1D compounds could be the phase diagram of low-temperature ground state in coordinates of coupling constants. The renorm-group solution of the 1D models with  $g_{1,2}$  coupling constants ( $g_1$  and  $g_2$  being the backward and forward scattering amplitudes respectively) (see, e.g., D. Senechal, cond-mat/9908262) gives the gapless Luttinger-liquid state at  $|g_1| < 2g_2$  while in actual compounds no such a state is usually observed.
- [5] B. Horovitz, H. Gutfreund and M. Weger, Phys Rev B. **12**, 3174 (1975)
- [6] Ross H. McKenzie, Phys Rev B. **52**, 16428 (1995)
- [7] T. Ishiguro, K. Yamaji, and G. Saito, Organic Superconductors, 2<sup>nd</sup> edition, Springer-Verlag Berlin Heidelberg, 1998.
- [8] P. Fulde and A. Ferrel, Phys. Rev. **135**, A550 (1964).
- [9] A. I. Larkin and Yu. N. Ovchinnikov, Sov. Phys. JETP **20**, 762 (1965).
- [10] W.P. Su, J. R. Schrieffer and A. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979); Phys. Rev. B **22**, 2099 (1980).
- [11] S.A. Brazovskij, Sov. Phys. JETP **51**, 342 (1980).
- [12] S.A. Brazovskii, L.P. Gor'kov, J.R. Schrieffer, Physica Scripta **25**, 423 (1982).
- [13] S.A. Brazovskii, L.P. Gor'kov, A.G. Lebed', Sov. Phys. JETP **56**, 683 (1982) [Zh. Eksp. Teor. Fiz. **83**, 1198 (1982)].

- 
- [14] A.I. Buzdin and V.V. Tugushev, Sov. Phys. JETP **58**, 428 (1983) [Zh. Eksp. Teor. Fiz. **85**, 735 (1983)].
- [15] W.P. Su and J. R. Schrieffer, Physics in One Dimension/ Ed. by J. Bernascony and T. Schneider, Springer series in Solid State Sciences, Berlin, Heidelberg and New York: Springer, 1981.
- [16] S.A. Brazovskij and N.N. Kirova, Sov. Sci. Rev. A Phys., volume **5**, p. 99 (1984).
- [17] S.A. Brazovskii, I.E. Dzyaloshinskii and N.N. Kirova, Sov. Phys. JETP **54**, 1209 (1981) [ZhETF 81, 2279 (1981)].
- [18] L. P. Gor'kov and A. G. Lebed, J. Phys. (Paris) Lett. **45**, L433 (1984).
- [19] G. Montambaux, M. Héritier, and P. Lederer, Phys. Rev. Lett. **55**, 2078 (1985).
- [20] J. F. Kwak, J. E. P.M. Chaikin et al., Phys. Rev. Lett. **56**, 972 (1986).
- [21] A. G. Lebed, JETP Lett. **78**, 138 (2003).
- [22] W. Dieterich and P.Fulde, Z. Phys. **265**, 239 (1973).
- [23] P. M. Chaikin, J. Phys. I France **6**, 1875 (1996).
- [24] D. Zanchi, A. Bjeliš, and G. Montambaux, Phys. Rev. B **53**, 1240 (1996).
- [25] M. Fujita, K. Machida and H. Nakanishi, J. Phys. Soc Jpn., **54**, 3820 (1985).
- [26] M.V. Kartsovnik and V.N. Laukin, J. Phys. I France **6**, 1753 (1996).
- [27] N. Biskup, J.A.A.J. Perenboom, J.S. Qualls, and J.S. Brooks, Solid State Commun. **107**, 503 (1998).
- [28] P. Christ, W. Biberacher, M. V. Kartsovnik, E. Steep, E. Balthes, H. Weiss, and H. Muller, JETP Lett. **71**, 303 (2000) [Pisma Zh. Eksp. Teor. Fiz. **71**, 437 (2000)].
- [29] J. S. Qualls, L. Balicas, J. S. Brooks et al., Phys. Rev. B **62**, 10008 (2000).
- [30] D. Andres, M. V. Kartsovnik, W. Biberacher et al., Phys. Rev. B **64**, 161104(R) (2001).
- [31] N. Harrison, C. H. Mielke, A. D. Christianson, J.S. Brooks, and M. Tokumoto, Phys. Rev. Lett. **86**, 1586 (2001).
- [32] D. Andres, M. V. Kartsovnik, P. D. Grigoriev, W. Biberacher, and H. Müller, Phys. Rev. B **68**, 201101(R) (2003).
- [33] K. Maki, B. Dörra, M. V. Kartsovnik et al., Phys. Rev. Lett. **90**, 256402 (2003).
- [34] N. Harrison, J. Singleton, A. Bangura et al., Phys. Rev. B **69**, 165103 (2004).
- [35] D. Graf, J. S. Brooks, E. S. Choi et al., Phys. Rev. B **69**, 125113 (2004).
- [36] D. Graf, E. S. Choi, J. S. Brooks et al., Phys. Rev. Lett. **93**, 076406 (2004).
- [37] E. B. Lopes, M. J. Matos, R. T. Henriques, M. Almeida, and J. Dumas Phys. Rev. B **52**, R2237 (1995).
- [38] R.McKenzie, cond-mat/9706235 (1998)

- 
- [39] A. Bjeliš, D. Zanchi, and G. Montambaux, “Pauli and orbital effects of magnetic field on charge density waves”, `cond-mat/9909303`.
  - [40] Y.Hasegawa and H. Fukuyama, J. Phys. Soc. Japan **55**, 3978 (1986).
  - [41] O. V. Dimitrova and M. V. Feigel'man, JETP Lett. 78, 637 (2003)
  - [42] H. Burkhardt and D. Rainer, Ann. Physik **3**, 181 (1994).

## References

- [1] Monceau, Pierre: Electronic properties of inorganic quasi-one-dimensional compounds, volume 1. Springer, 1985.
- [2] Pletyukhov, Mikhail, Kennes, Dante M., Piasotski, Kiryl, Klinovaja, Jelena, Loss, Daniel, and Schoeller, Herbert: Rational boundary charge in one-dimensional systems with interaction and disorder. Physical Review Research, 2(3), September 2020, ISSN 2643-1564. <http://dx.doi.org/10.1103/PhysRevResearch.2.033345>.
- [3] Piasotski, Kiryl, Müller, Niclas, Kennes, Dante M., Schoeller, Herbert, and Pletyukhov, Mikhail: Universal properties of boundary and interface charges in multichannel one-dimensional cont. Physical Review B, 106(16), October 2022, ISSN 2469-9969. <http://dx.doi.org/10.1103/PhysRevB.106.165405>.

### 1.5 References by S. Doniach (1977)

- [1] A. Edelstein, Phys. Lett. 27 A (1968) 614.
- [2] For other references see B. Cornut and B. Coqblin. Phys. Rev. B 5 (1972) 4541.
- [3] S.H. Liu, P. Burgardt, K.A. Gschneider and S. Iegvold. J. Phys. F. 6(1976) I. 55
- [4] In fig. 1, the RKY binding again takes over at large  $J$ . but this is misleading since the weak coupling formula (1) breaks down in this regime.
- [5] A. Luther and I. Peschel, Phys. Rev. B 12 (1975) 3908.
- [6] G. Yuval and P.W. Anderson. Phys. Rev. B 1 (1970) 1.522 .
- (7) P.W. Anderson. G. Yuval and 1).R. Hamann, Phys Rev B 1 (1970) 4464
- [8] A. Luther and V. Emery, Phys. Rev. Lett. 33 (1974) 589 .
- [9] M. Blume. Phys. Rev. 141(1966)517
- [10] R.B. Stinchcombe, (i. Horwitz. F. Englert and R. Brout, Phys. Rev. 130 ) (1963) 155
- [11] K.D. Schotte and U. Schotte. Phys. Rev. 182 (1969) 479.
- [12] See also S. Doniach and F.H. Sondheimer, Green's Functions for Solid State Physicists (Benjamin. 1974) ch. 9.

11.3| At finite temperatures an effective Néel temperature is expected to go linearly to zero in  $J \rightarrow J$  from below: S. Doniach. Proceedings of the Rochester (conference on Valence Instabilities. editor R. D. Parks. (Plenum Press. New Yorks to be published in 1977.

### 1.6 References by Pletyukhov et. al. (?)

(??? from which article they are?)

## References

- [1] K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).
- [2] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).
- [3] B. A. Volkov and O. A. Pankratov, Pis'ma Zh. Eksp. Teor. Fiz. **42**, 145 (1985) [JETP Lett. **42**, 178 (1985)].
- [4] O. A. Pankratov, S. V. Pakhomov, and B. A. Volkov, Solid State Commun. **61**, 93 (1987).

- [5] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
- [6] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).
- [7] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).
- [8] M. König, S. Wiedmann, C. Brune, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).
- [9] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008).
- [10] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
- [11] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
- [12] B. A. Bernevig, *Topological Insulators and Topological Superconductors*, Princeton University Press (2013).
- [13] G. Tkachov, *Topological Insulators: The Physics of Spin Helicity in Quantum Transport*, (Pan Stanford, 2015).
- [14] J. K. Asbóth, L. Oroszlány, and A. Pályi, *A Short Course on Topological Insulators*, Lecture Notes in Physics, Springer 2016.
- [15] R. Jackiw and J. Schrieffer, Nucl. Phys. B **190**, 253 (1981); W. P. Su and J. R. Schrieffer, Phys. Rev. Lett. **46**, 738 (1981); M. J. Rice and E. J. Mele, Phys. Rev. Lett. **49**, 1455 (1982); R. Jackiw and G. Semenoff, Phys. Rev. Lett. **50**, 439 (1983); S. Ryu, C. Mudry, C.-Y. Hou, and C. Chamon, Phys. Rev. B **80**, 205319 (2009); J. Klinovaja and D. Loss, Phys. Rev. Lett. **110**, 126402 (2013); R. Wakatsuki, M. Ezawa, Y. Tanaka, and N. Nagaosa, Phys. Rev. B **90**, 014505 (2014).
- [16] Here, *local* means that the symmetry operation can be defined within a unit cell.
- [17] T. L. Hughes, E. Prodan, and B. A. Bernevig, Phys. Rev. B **83**, 245132 (2011).
- [18] C.-K. Chiu, H. Yao, and S. Ryu, Phys. Rev. B **88**, 075142 (2013).
- [19] K. Shiozaki and M. Sato, Phys. Rev. B **90**, 165114 (2014).
- [20] A. Alexandradinata, Zhijun Wang, and B. A. Bernevig, Phys. Rev. X **6**, 021008 (2016).
- [21] L. Trifunovic and P. Brouwer, Phys. Rev. B **96**, 195109 (2017).
- [22] A. Lau, C. Ortix, and J. van den Brink, Phys. Rev. Lett. **115**, 216805 (2015); A. Lau, J. van den Brink, and C. Ortix, Phys. Rev. B **94**, 165164 (2016); A. Lau and C. Ortix, Eur. Phys. J. Spec. Top. **227**, 1309 (2018).
- [23] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).
- [24] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. W. Ludwig, New J. Phys. **12**, 065010 (2010).
- [25] A. Kitaev, in Advances in Theoretical Physics, edited by V. Lebedev, and M. Feigel'man, AIP Conf. Proc. No. 1134 (AIP, New York, 2009), p. 22.

- [26] R.-J. Slager, A. Mesaros, V. Juričić, and J. Zaanen, Nat. Phys. **9**, 98 (2012).
- [27] P. Jadaun, D. Xiao, Q. Niu, and S. K. Banerjee, Phys. Rev. B **88**, 085110 (2013).
- [28] C.-K. Chiu, H. Yao, and S. Ryu, Phys. Rev. B **88**, 075142 (2013).
- [29] F. Zhang, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **111**, 056403 (2013).
- [30] W. A. Benalcazar, J.C.Y. Teo, and T. L. Hughes, Phys. Rev. B **89**, 224503 (2014).
- [31] T. Morimoto and A. Furusaki, Phys. Rev. B **88**, 125129 (2013).
- [32] M. Diez, D. I. Pikulin, I.C. Fulga, and J. Tworzydło, New J. Phys. **17**, 043014 (2015).
- [33] R.D. King-Smith and D. Vanderbilt, Phys. Rev. B(R) **47**, 1651 (1993).
- [34] D. Vanderbilt and R.D. King-Smith, Phys. Rev. B **48**, 4442 (1993).
- [35] R. Resta, Rev. Mod. Phys. **66**, 899 (1994).
- [36] K.N. Kudin and R. Car, J. of Chem. Phys. **126**, 234101 (2007).
- [37] N. Marzari, A.A. Mostofi, J.R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012).
- [38] N.A. Spaldin, J. of Solid State Chem. **195**, 2 (2012).
- [39] J.-W. Rhim, J. Behrends and J.H. Bardarson, Phys. Rev. B **95**, 035421 (2017).
- [40] G. van Miert and Carmine Ortix, Phys. Rev. B **96**, 235130 (2017).
- [41] D. Vanderbilt, *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators*, (Cambridge University Press, 2018).
- [42] J. Zak, Phys. Rev. Lett. **48**, 359 (1982); J. Zak, Phys. Rev. Lett. **62**, 2747 (1989).
- [43] Y. Hatsugai, Phys. Rev. Lett. **71**, 3697 (1993).
- [44] X.-L. Qi, Y.-S. Wu, and S.-C. Zhang, Phys. Rev. B **74**, 045125 (2006).
- [45] P. Delplace, D. Ullmo, and G. Montambaux, Phys. Rev. B **84**, 195452 (2011).
- [46] L. Fidkowski, T.S. Jackson, and I. Klich, Phys. Rev. Lett. **107**, 036601 (2011); R.S.K. Mong and V. Shivamoggi, Phys. Rev. B **83**, 125109 (2011); V. Gurarie, Phys. Rev. B **83**, 085426 (2011); A.M. Essin and V. Gurarie, Phys. Rev. B **84**, 125132 (2011); T. Fukui, K. Shiozaki, T. Fujiwara, and S. Fujimoto, J. Phys. Soc. Jpn. **81**, 114602 (2012); Y. Yu, Y.-S. Wu, and X. Xie, Nucl. Phys. B **916**, 550 (2017); J.-W. Rhim, J.H. Bardarson, and R.-J. Slager, Phys. Rev. B **97**, 115143 (2018); M. Silveirinha, Phys. Rev. X **9**, 011037 (2019).
- [47] W.P. Su, J.R. Schrieffer, and A.J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979);
- [48] J.-H. Park, G. Yang, J. Klinovaja, P. Stano, and D. Loss, Phys. Rev. B **94**, 075416 (2016).
- [49] M. Thakurathi, J. Klinovaja, and D. Loss, Phys. Rev. B **98**, 245404 (2018).
- [50] D. J. Thouless, Phys. Rev. B **27**, 6083 (1982).
- [51] Y. Hatsugai and T. Fukui, Phys. Rev. B **94**, 041102(R) (2016).

- 
- [52] M. Pletyukhov, D.M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, arXiv:1911.06890, submitted to Phys. Rev. Lett.
- [53] Y. Lahini, R. Pugatch, F. Pozzi, M. Sorel, R. Morandotti, N. Davidson, and Y. Silberberg, Phys. Rev. Lett. **103**, 013901 (2009); M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, Science **349**, 842 (2015); S. Ganeshan, K. Sun, and S. Das Sarma, Phys. Rev. Lett. **110**, 180403 (2013).
- [54] I. Dana, Y. Avron, and J. Zak, J. Phys. C **18**, L679 (1985).
- [55] M. Kohmoto, Phys. Rev. B **39**, 11943 (1989); *ibid.*, J. Phys. Soc. Jpn. **61**, 2645 (1992).
- [56] Y. Hatsugai, Phys. Rev. B **48**, 11851 (1993).
- [57] D. M. Kennes, N. Müller, M. Pletyukhov, C. Weber, C. Bruder, F. Hassler, J. Klinovaja, D. Loss, and H. Schoeller, Phys. Rev. B **100**, 041103 (2019).
- [58] A. Alase, E. Cobanera, G. Ortiz, and L. Viola, Phys. Rev. B **96**, 195133 (2017); F. K. Kunst, G. van Miert, and E. J. Bergholtz, Phys. Rev. B **99**, 085427 (2019).
- [59] J.J. Rehr and W. Kohn, Phys. Rev. B **10**, 448 (1974).
- [60] W. Kohn, Phys. Rev. **115**, 809 (1959).
- [61] Y. Hatsugai, J. Phys.: Condens. Matter **9**, 2507 (1997).
- [62] C. Kallin and B.I. Halperin, Phys. Rev. B **29**, 2175 (1984).
- [63] J. Friedel, Del Nuovo Cimento **2**, 287 (1958).
- [64] L. He and D. Vanderbilt, Phys. Rev. Lett. **86**, 5341 (2001).
- [65] see Supplemental Material, where the parameters used in the Figures are listed.
- [66] M. Kohmoto, Ann. Phys. **160**, 343 (1985).
- [67] N. Müller, D.M. Kennes, M. Pletyukhov, J. Klinovaja, D. Loss, and H. Schoeller, in preparation.
- [68] N. Müller, D.M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, arXiv:1911.02295, submitted to Phys. Rev. B.
- [69] Y.-T. Lin, D.M. Kennes, V. Meden, and H. Schoeller, in preparation.
- [70] S. Gangadharaiah, L. Trifunovic, and D. Loss, Phys. Rev. Lett. **108**, 136803 (2012).
- [71] K. Piasotski, D.M. Kennes, M. Pletyukhov, J. Klinovaja, D. Loss, V. Meden, and H. Schoeller, in preparation.
- [72] J. Klinovaja, Y. Tserkovnyak, and D. Loss, Phys. Rev. B **91**, 085426 (2015).
- [73] M. Pletyukhov, D.M. Kennes, K. Piasotski, J. Klinovaja, D. Loss, and H. Schoeller, in preparation.



### 1..7 References by Pletyukhov et. al. 2 (?)

## References

- [17] J.-H. Park, G. Yang, J. Klinovaja, P. Stano, and D. Loss, *Phys. Rev. B* **94**, 075416 (2016).
- [41] S. Kivelson and J. R. Schrieffer, *Phys. Rev. B* **25**, 6447 (1982).
- [42] R. Rajaraman and J. Bell, *Physics Letters B* **116**, 151 (1982).
- [43] S. Kivelson, *Phys. Rev. B* **26**, 4269 (1982).
- [44] J. Bell and R. Rajaraman, *Nucl. Phys. B* **220**, 1 (1983).
- [45] Y. Frishman and B. Horovitz, *Phys. Rev. B* **27**, 2565 (1983).
- [46] R. Jackiw, A. Kerman, I. Klebanov, and G. Semenoff, *Nuclear Physics B* **225**, 233 (1983).
- [21] M. Pletyukhov, D. M. Kennes, K. Piasotski, J. Klinovaja, D. Loss, and H. Schoeller, *Phys. Rev. Research* **2**, 033345 (2020).
- [47] J. J. Rehr and W. Kohn, *Phys. Rev. B* **10**, 448 (1974).
- [48] C. Kallin and B. I. Halperin, *Phys. Rev. B* **29**, 2175 (1984).
- [49] L. He and D. Vanderbilt, *Phys. Rev. Lett.* **86**, 5341 (2001).
- [20] M. Pletyukhov, D. M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, *Phys. Rev. B* **101**, 165304 (2020).
- [22] Y.-T. Lin, D. M. Kennes, M. Pletyukhov, C. S. Weber, H. Schoeller, and V. Meden, *Phys. Rev. B* **102**, 085122 (2020).
- [36] C. W. Groth, M. Wimmer, A. R. Akhmerov, J. Tworzydło, and C. W. J. Beenakker, *Phys. Rev. Lett.* **103**, 196805 (2009).
- [50] C.S. Weber *et al.*, in preparation.
- [51] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).
- [18] M. Thakurathi, J. Klinovaja, and D. Loss, *Phys. Rev. B* **98**, 245404 (2018).
- [19] M. Pletyukhov, D. M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, *Phys. Rev. B* **101**, 161106(R) (2020).
- [52] N. Hine and M. Foulkes, *Journal of Physics: Condensed Matter* **19**, 506212 (2007).
- [53] H. Min and A. H. MacDonald, *Progress of Theoretical Physics Supplement* **176**, 227 (2008).

## 1..8 References from “Fluctuations of boundary” (?)

### References

- [1] B. A. Volkov and O. A. Pankratov, JETP Lett. **42**, 178 (1985).
- [2] O. Pankratov, S. Pakhomov, and B. Volkov, Solid State Communications **61**, 93 (1987).
- [3] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
- [4] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757–1761 (2006).
- [5] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).
- [6] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).
- [7] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature **452**, 970 (2008).
- [8] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
- [9] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
- [10] B. A. Bernevig and T. L. Hughes, Topological Insulators and Topological Superconductors, stu - student edition ed. (Princeton University Press, 2013).
- [11] G. Tkachov, Topological insulators: The physics of spin helicity in quantum transport (Pan Stanford, 2015).
- [12] J. K. Asbóth, L. Oroszlány, and A. Pályi, Lecture Notes in Physics (2016), 10.1007/978-3-319-25607-8.
- [13] W. Kohn, Phys. Rev. **133**, A171 (1964).
- [14] R. Resta, Journal of Physics: Condensed Matter **14**, R625 (2002).
- [15] E. Meier, F. An, A. Dauphin, M. Maffei, P. Massignan, T. Hughes, and B. Gadway, Science **362**, 929 (2018).
- [16] The other parameters are  $N_s = 400$ ,  $l_p = 200$  for  $d \neq 0$  and  $N_s, l_p \rightarrow \infty$  for  $d = 0$  (see SM [26] for details how the infinite system size limit can be taken numerically),  $t_2 = 1$ ,  $t_1 = r$ ,  $a = 1$  and we work at half-filling. The results are averaged over 20 disorder configurations. For the details of the model see Eq. (3.2).
- [17] J.-H. Park, G. Yang, J. Klinovaja, P. Stano, and D. Loss, Phys. Rev. B **94**, 075416 (2016).
- [18] M. Thakurathi, J. Klinovaja, and D. Loss, Phys. Rev. B **98**, 245404 (2018).
- [19] M. Pletyukhov, D. M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, Phys. Rev. B **101**, 161106(R) (2020).
- [20] M. Pletyukhov, D. M. Kennes, J. Klinovaja, D. Loss, and H. Schoeller, Phys. Rev. B **101**, 165304 (2020).
- [21] M. Pletyukhov, D. M. Kennes, K. Piasotski, J. Klinovaja, D. Loss, and H. Schoeller, Phys. Rev. Research **2**, 033345 (2020).

- 
- [22] Y.-T. Lin, D. M. Kennes, M. Pletyukhov, C. S. Weber, H. Schoeller, and V. Meden, *Phys. Rev. B* **102**, 085122 (2020).
- [23] W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698 (1979).
- [24] W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. B* **22**, 2099 (1980).
- [25] S. Sen, P. J. Wong, and A. K. Mitchell, *Phys. Rev. B* **102**, 081110(R) (2020).
- [26] See Supplemental Material for the  $C$ s used in Fig 4, details about the calculations in the non-interacting cases considered, as well as a discussion of finite temperature effects, which includes Refs. [41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53].
- [27] W. Kohn, *Phys. Rev. Lett.* **76**, 3168 (1996).
- [28] E. Prodan and W. Kohn, *Proceedings of the National Academy of Sciences* **102**, 11635 (2005).
- [29] I. Bloch, J. Dalibard, and W. Zwerger, *Rev. Mod. Phys.* **80**, 885 (2008).
- [30] C. Sgierovello, M. Peressi, and R. Resta, *Phys. Rev. B* **64**, 115202 (2001).
- [31] D. Vanderbilt and R. D. King-Smith, *Phys. Rev. B* **48**, 4442 (1993).
- [32] D. Vanderbilt, *Berry phases in electronic structure theory: electric polarization, orbital magnetization* (Cambridge University Press, Cambridge, 2018).
- [33] I. Mondragon-Shem, T. L. Hughes, J. Song, and E. Prodan, *Phys. Rev. Lett.* **113**, 046802 (2014).
- [34] T. Giamarchi, *Quantum Physics in One Dimension*, International Series of Monographs on Physics (Oxford University Press, Oxford, 2003).
- [35] M. J. Rice and E. J. Mele, *Phys. Rev. Lett.* **49**, 1455 (1982).
- [36] C. W. Groth, M. Wimmer, A. R. Akhmerov, J. Tworzydło, and C. W. J. Beenakker, *Phys. Rev. Lett.* **103**, 196805 (2009).
- [37] S. Kivelson, H. B. Thacker, and W.-K. Wu, *Phys. Rev. B* **31**, 3785 (1985).
- [38] B. Horovitz and J. Sólyom, *Phys. Rev. B* **32**, 2681 (1985).
- [39] S. Gangadharaiah, L. Trifunovic, and D. Loss, *Phys. Rev. Lett.* **108**, 136803 (2012).
- [40] N. R. Cooper, J. Dalibard, and I. B. Spielman, *Rev. Mod. Phys.* **91**, 015005 (2019).
- [41] S. Kivelson and J. R. Schrieffer, *Phys. Rev. B* **25**, 6447 (1982).
- [42] R. Rajaraman and J. Bell, *Physics Letters B* **116**, 151 (1982).
- [43] S. Kivelson, *Phys. Rev. B* **26**, 4269 (1982).
- [44] J. Bell and R. Rajaraman, *Nucl. Phys. B* **220**, 1 (1983).
- [45] Y. Frishman and B. Horovitz, *Phys. Rev. B* **27**, 2565 (1983).
- [46] R. Jackiw, A. Kerman, I. Klebanov, and G. Semenoff, *Nuclear Physics B* **225**, 233 (1983).

- 
- [47] J. J. Rehr and W. Kohn, *Phys. Rev. B* **10**, 448 (1974).
  - [48] C. Kallin and B. I. Halperin, *Phys. Rev. B* **29**, 2175 (1984).
  - [49] L. He and D. Vanderbilt, *Phys. Rev. Lett.* **86**, 5341 (2001).
  - [50] C.S. Weber *et al.*, in preparation.
  - [51] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).
  - [52] N. Hine and M. Foulkes, *Journal of Physics: Condensed Matter* **19**, 506212 (2007).
  - [53] H. Min and A. H. MacDonald, *Progress of Theoretical Physics Supplement* **176**, 227 (2008).