Communications in
**Mathematical**
**Physics**

# Dimerization in Quantum Spin Chains with $O(n)$ Symmetry

**Jakob E. Björnberg**[1] , **Peter Mühlbacher**[2] , **Bruno Nachtergaele**[3] ,
**Daniel Ueltschi**[2]

[1] Department of Mathematics, Chalmers University of Technology and the University of Gothenburg,
Göteborg, Sweden. E-mail: jakob.bjornberg@gu.se

[2] Department of Mathematics, University of Warwick, Coventry CV4 7AL, UK. E-mail: peter@muehlbacher.me;
daniel@ueltschi.org

[3] Department of Mathematics and Center for Quantum Mathematics and Physics, University of California,
Davis, Davis, CA 95616, USA. E-mail: bxn@math.ucdavis.edu

**Abstract:** We consider quantum spins with $S \geq 1$, and two-body interactions with
$O(2S+1)$ symmetry. We discuss the ground state phase diagram of the one-dimensional
system. We give a rigorous proof of dimerization for an open region of the phase diagram,
for $S$ sufficiently large. We also prove the existence of a gap for excitations.

## Contents

## 1. Introduction

Over the course of almost a century of studying quantum spin chains, physicists and mathematicians have uncovered a wide variety of interesting physical phenomena and in the process invented an impressive arsenal of new mathematical techniques and structures. Nevertheless, our understanding of these simplest of quantum many-body systems is still far from complete. For many models of interest we have only partial information about the ground state phase diagram, the nature of the phase transitions, and the spectrum of excitations. We consider here a family of spin systems with two-body interactions, where interactions are translation invariant and $O(2S+1)$ invariant. We investigate the ground state phase diagram, looking for ground states that possess less symmetry than the interactions. Our main result is a rigorous proof of dimerization (where translation invariance is broken) in a region of the phase diagram with $S$ large enough (Theorem 1.1). We also prove exponential clustering (Theorem 1.2) and the existence of a gap (Theorem 1.4).

The family of models is introduced in Sect. 1.1; the phase diagram for general $S \geq 1$ is described in Sect. 1.2; the case $S = 1$ has received a lot of attention and we discuss it explicitly in Sect. 1.3; our result about dimerization is stated in Sect. 1.4.

The $O(n)$ models have a graphical representation which we describe in Sect. 2. We use it to define a "contour model" in Sect. 3 where contours are shown to have small weights. This allows to use the method of cluster expansion and prove dimerization in Sect. 4.

*1.1. A family of quantum spin chains with $O(n)$-invariant interactions.* We consider a family of quantum spin chains consisting of $2\ell$ spins of magnitude $S$ defined by a nearest-neighbor Hamiltonian $H_\ell$ acting on the Hilbert space $\mathcal{H}_\ell = (\mathbb{C}^n)^{\otimes 2\ell}$, with $n = 2S + 1 \geq 2$, of the form

$$H_\ell = \sum_{x=-\ell+1}^{\ell-1} h_{x,x+1}, \tag{1.1}$$

where $h_{x,x+1}$ denotes a copy of $h = h^* \in M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ acting on the nearest neighbor pair at sites $x$ and $x + 1$.

We are interested in the family of interactions

$$h = uT + vQ, \quad u, v \in \mathbb{R}, \tag{1.2}$$

where $T$ is the transposition operator defined by $T(\phi \otimes \varphi) = \varphi \otimes \phi$, for $\phi, \varphi \in \mathbb{C}^n$, and $Q$ is the orthogonal projection onto the one-dimensional subspace of $\mathbb{C}^n \otimes \mathbb{C}^n$ spanned by a vector of the form

$$\psi = \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{n} e_\alpha \otimes e_\alpha, \tag{1.3}$$

for some orthornormal basis $\{e_\alpha | \alpha = 1, \ldots, n\}$ of $\mathbb{C}^n$.

The spectrum of $h$ is easy to find. $T$ has the eigenvalues 1 and $-1$, corresponding to the symmetric and antisymmetric subspaces of $\mathbb{C}^n \otimes \mathbb{C}^n$, whose dimensions are $n(n+1)/2$ and $n(n-1)/2$, respectively. Since $\psi$ is symmetric, the eigenvalues of $h$ are $u+v$, $u$, $-u$.

Let $R$ be a linear transformation represented by an orthogonal matrix in the basis $\{e_\alpha\}$, meaning $\langle e_\alpha, RR^T e_\beta \rangle = \delta_{\alpha\beta}$. This amounts to defining a specific representation of $O(n)$ on the system under consideration. It is then straightforward to check $(R \otimes R)\psi = \psi$. It follows that $R \otimes R$ commutes with $Q = |\psi\rangle\langle\psi|$. Since $T$ also commutes with $R \otimes R$, the Hamiltonians with interaction $h$ given in (1.2) have a local $O(n)$ symmetry. This family of models is in fact, up to a trivial additive constant, the most general translation-invariant nearest neighbor Hamiltonian for spins of dimension $n$ and with a translation-invariant local $O(n)$ symmetry.

To make contact with previous results in the literature, it is useful to note a couple of equivalent forms of the spin chains we consider. First, for integer values of $S$, that is odd dimensions $n$, consider the orthonormal basis $\{e_\alpha\}$, relabeled by $\alpha = -S, \dots, S$, and related to the standard eigenbasis of the third spin matrix $S^{(3)}$, satisfying $S^{(3)}|\alpha\rangle = \alpha|\alpha\rangle$, as follows: for $\alpha = 0$ take $e_0 = i^S|0\rangle$, and for $\alpha > 0$ define

$$e_\alpha = \frac{i^{S-\alpha}}{\sqrt{2}}(|\alpha\rangle + |-\alpha\rangle), \quad e_{-\alpha} = \frac{i^{S-\alpha+1}}{\sqrt{2}}(|\alpha\rangle - |-\alpha\rangle). \tag{1.4}$$

Then, we have

$$\psi = \phi := \frac{1}{\sqrt{n}} \sum_{\alpha=-S}^{S} (-1)^{S-\alpha}|\alpha, -\alpha\rangle, \tag{1.5}$$

which is the $SU(2)$ singlet vector in the standard spin basis. The transposition operator $T$ is of course not affected by any translation-invariant local basis change. Therefore, for odd $n$, and with a simple change of basis, the family of interactions (1.2) is seen to be equivalent to

$$\tilde{h} = uT + vP, \quad u, v \in \mathbb{R}, \tag{1.6}$$

where $P$ is the orthogonal projection onto the singlet state $\phi$.

The case of even $n$ is different. Interactions $h$ and $\tilde{h}$ are not unitarily equivalent. But the model with interaction $\tilde{h}$ is nonetheless interesting and we discuss it in Appendix A. We also prove dimerization and a gap in this case, see Theorem 1.3 and Theorem 1.4.

For $n \geq 2$, $u = 0$, and $v = -1$, this is the much studied $-P^{(0)}$ spin chain [1,3,4,6, 12,18,19].

*1.2. Ground state phase diagram for general $n \geq 3$.* We start with the phase diagram for arbitrary $n \geq 3$ and discuss the special case $n = 3$ in Sect. 1.3. The ground state phase diagram of the spin chain with nearest-neighbor interactions $h_{x,x+1} = uT_{x,x+1} + vQ_{x,x+1}$ is depicted in Fig. 1. It can be broadly divided into four domains.

The domain formed by the quadrant $u \leq 0$, $v \geq 0$ (blue region in Fig. 1) is ferromagnetic. There are many ground states and they minimize $h_{x,x+1}$ for all $x$; that is, they are frustration-free. The ground state energy per bond is equal to $u$. Indeed, let $\varphi = \sum_\alpha c_\alpha e_\alpha$ with $\sum_\alpha |c_\alpha|^2 = 1$. It is clear that $|\varphi \otimes \varphi\rangle$ is eigenstate of $T$ with eigenvalue 1; further, we have

$$\langle \varphi \otimes \varphi | Q | \varphi \otimes \varphi \rangle = \frac{1}{n}\left|\sum_\alpha c_\alpha^2\right|^2. \tag{1.7}$$
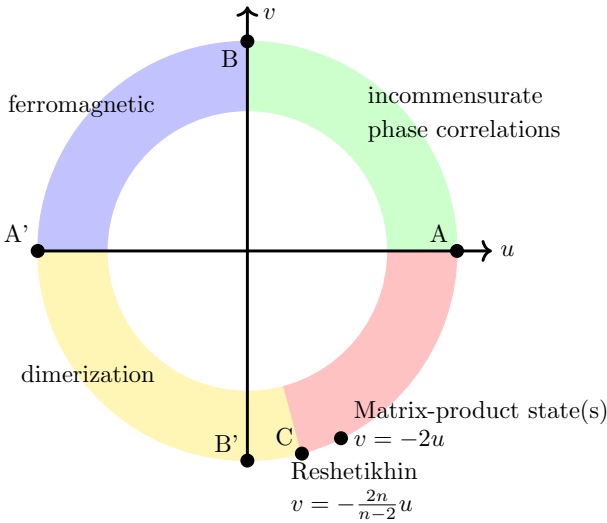
**Fig. 1.** Ground state phase diagram for the chain with nearest-neighbor interactions $uT + vQ$ for $n \geq 3$. Our main result, Theorem 1.1, is a proof of dimerization in an open region around the point B'

The latter is zero when $\sum_\alpha c_\alpha^2 = 0$. Since $Q$ is a projector, such a state is eigenstate with eigenvalue 0. Notice that the state $R\varphi$ also satisfies this condition, for all orthogonal transformation $R$.

The product state $\otimes_{x=-\ell+1}^{\ell} \varphi$ is then a ground state of $h_{x,x+1}$ with eigenvalue $u$, for all $x$. In addition to these product states, we can obviously take linear combinations.

The next domain is the arc-circle between $(u, v) = (-1, 0)$ and the "Reshetikhin point" with $v = -\frac{2n}{n-2}u$ (yellow region in Fig. 1), which features dimerization. In order to see that dimerization is plausible as soon as $v < 0$, let

$$\varphi_{x,x+1} = \sqrt{1 - \varepsilon^2}\, |S, S\rangle + \frac{\varepsilon}{\sqrt{n-1}} \sum_{\alpha=-S}^{S-1} |\alpha, \alpha\rangle. \tag{1.8}$$

and consider the (partially) dimerized state $\varphi_{-\ell+1,-\ell+2} \otimes \varphi_{-\ell+3,-\ell+4} \otimes \ldots$ For $\varepsilon = 0$, this is a product state, but for $\varepsilon \neq 0$ it is not. Roughly half the edges, namely the edges $x, x+1$ with $x = -\ell + 1, -\ell + 3, \ldots$, are dimerized and their energy is

$$\langle \varphi_{x,x+1} | u T_{x,x+1} + v Q_{x,x+1} | \varphi_{x,x+1} \rangle = u + \frac{v}{n}(1 + 2\sqrt{n-1}\varepsilon) + O(\varepsilon^2). \tag{1.9}$$

The non-dimerized edges contribute

$$\langle \varphi_{x-1,x} \otimes \varphi_{x+1,x+2} | u T_{x,x+1} + v Q_{x,x+1} | \varphi_{x-1,x} \otimes \varphi_{x+1,x+2} \rangle = u + \frac{v}{n} + O(\varepsilon^2). \tag{1.10}$$

The average energy per bond of the state $\varphi_{x,x+1}$ is then $u + \frac{v}{n} + v\frac{\sqrt{n-1}}{n}\varepsilon$, up to $O(\varepsilon^2)$ corrections. When $v < 0$ the optimal product states have energy $u + \frac{v}{n}$ (using (1.7) with $\sum_\alpha c_\alpha^2 = 1$), so the partially dimerized state (1.8) has lower energy when $\varepsilon$ is positive and small.

Our main result is that dimerization does occur in an open domain around the point B', provided $n$ is sufficiently large, see Theorem 1.1. This extends the results of [3,18], valid at the point B'.

Then comes the domain formed by the arc-circle between the Reshetikhin point $v = -\frac{2n}{n-2}u$ and $(u, v) = (1, 0)$ (red region in Fig. 1). For $n$ odd a unique translation-invariant ground state is expected.

This domain contains several interesting special cases. The direction $(u = 1, v = 0)$ is the the $SU(n)$ generalization of the spin-1/2 Bethe-ansatz solvable Heisenberg model studied by Sutherland and others [23]. The direction $v = -\frac{2n}{n-2}u$ was solved by Reshetikhin [21] (this generalizes the Takhtajan–Babujian model for $n = 3$). These models are gapless. The direction $v = -2u$ is a frustration free point and the ground states are given matrix-product states. For odd $n$, these are generalizations of the AKLT model. The ground state for the infinite chain is unique and is in the Haldane phase. For even $n$, there are two matrix-product ground states that break the translation invariance of the chain down to period 2 [27].

The final domain is the quadrant $u, v > 0$. The ground states are expected to have slow decaying correlations with incommensurate phase correlations. That is, spin-spin correlations between sites 0 and $x$ are expected to be of the form $|x|^{-r} \cos(\omega|x|)$ for $|x|$ large, and where $r, \omega$ depend on the parameters $u, v$ [10].

It is perhaps worth mentioning that the phase diagram for spatial dimensions other than 1 is quite different. Dimerization is not expected. Instead, the system displays various forms of magnetic long-range orders (ferromagnetic, spin nematic, Néel, …). See [29] for results about magnetic ordering for all $n \geq 2$ and for parameters that correspond to the dimerized phase here.

*1.3. The $S = 1$ model ($n = 3$).* For $n = 3$, the family of models is equivalent to the familiar spin-1 chain with bilinear and biquadratic interactions. The latter is most often parametrized by an angle $\phi$ as follows:

$$\cos\phi\, \vec{S}_x \cdot \vec{S}_{x+1} + \sin\phi\, (\vec{S}_x \cdot \vec{S}_{x+1})^2 = 3(\sin\phi - \cos\phi)P + \cos\phi\, T + \sin\phi\, I. \quad (1.11)$$

We can apply the change of basis that is the inverse of Eq. (1.4), namely

$$|0\rangle = -i\, e_0, \quad |1\rangle = \tfrac{1}{\sqrt{2}}(e_1 - i\, e_{-1}), \quad |-1\rangle = \tfrac{1}{\sqrt{2}}(e_1 + i\, e_{-1}). \quad (1.12)$$

Then the interaction is given by (1.11) but with the operator $Q$ instead of $P$.

The ground state phase diagram with parameter $\phi$ is depicted in Fig. 2. The domains and the points are the same as in Fig. 1. The ferromagnetic domain corresponds to $\phi \in (\frac{\pi}{2}, \frac{5\pi}{4})$, and the model is frustration-free in this range. Among the ground states, there is a family of product states that shows that the $O(3)$ symmetry of the Hamiltonian is spontaneously broken. As a consequence, the Goldstone Theorem [14] implies that there are gapless excitations above the ground state in this region. The dimerization domain is $\phi \in (\frac{5\pi}{4}, \frac{7\pi}{4})$. The next domain is $\phi \in (-\frac{\pi}{4}, \frac{\pi}{4})$ with unique, translation-invariant ground states. Finally, the domain $\phi \in (\frac{\pi}{4}, \frac{\pi}{2})$ is expected to display states with slow decay of correlations, with incommensurate phase.

There are several points where exact and/or rigorous information is available: (i) $\phi \in [0, \pi/2]$ with $\tan\phi = 1/3$, it is the spin-1 AKLT chain [2] with interaction $\tilde{h}$ given by the orthogonal projection on the spin-2 states. In the thermodynamic limit, it has a unique ground state of Matrix Product form with a non-vanishing spectral gap and exact
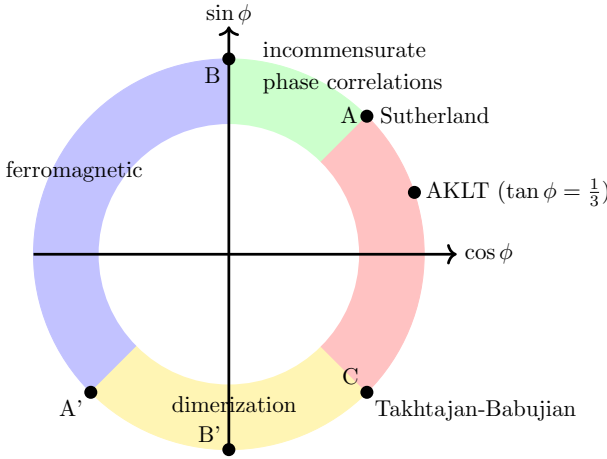
**Fig. 2.** Ground state phase diagram for the $S = 1$ chain with nearest-neighbor interactions $\cos\phi\,\vec{S}_x \cdot \vec{S}_{x+1} + \sin\phi\,(\vec{S}_x \cdot \vec{S}_{x+1})^2$. The domains and the points are the same as those in Fig. 1

exponential decay of correlations; (ii) the two points with $\tan\phi = 1$, A and A' in Fig. 2, have $SU(3)$ symmetry and are often referred to as the Sutherland model [23]. An exact solution for the ground state at $\phi = -3\pi/4$ is gapless and highly degenerate, while for $\phi = \pi/4$ is believed to be a unique critical state with gapless excitations; (iii) the point $\phi = -\pi/4$ is the Bethe-ansatz solvable Takhtajan–Babujian model [5,24], which is also gapless; (iv) the point $\phi = -\pi/2$, is the $-P^{(0)}$ spin-1 chain, already mentioned above. Aizenman, Duminil-Copin, and Warzel proved that it has two dimerized (2-periodic) ground states with exponential decay of correlations [3]; all evidence indicates that these states are gapped.

Let us briefly comment on higher spatial dimensions. Dimerization is not expected. Various rigorous results about magnetic long-range order have been established: for $\phi = 0$ [8]; for $\phi \gtrsim \frac{5\pi}{4}$ [25,29]; and for $\phi \lesssim 0$ [15]. Recently, the model on the complete graph has been studied by Ryan using methods based on the Brauer algebra [22], which plays a role in the representation theory of the orthogonal groups analogous to that of the symmetric group for the general linear groups.
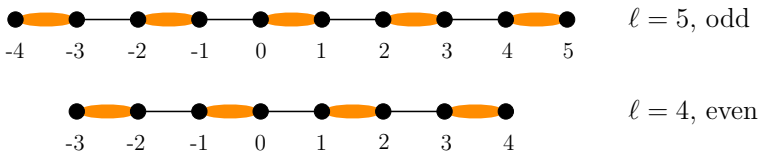
*1.4. Our result about dimerization.* Let us introduce the operators $L^{\alpha,\alpha'}$, $1 \leq \alpha < \alpha' \leq n$, that are generators of the Lie algebra $\mathfrak{o}(n)$:

$$L^{\alpha,\alpha'} = |\alpha\rangle\langle\alpha'| - |\alpha'\rangle\langle\alpha|. \tag{1.13}$$

And for $x \in \{-\ell+1, \ldots, \ell\}$, let $L_x^{\alpha,\alpha'}$ be the operator in $\mathcal{H}_\ell$ that acts as $L^{\alpha,\alpha'}$ at the site $x$, and as the identity elsewhere.

**Theorem 1.1** *There exist constants $n_0, u_0, c > 0$ (independent of $\ell$) such that for $n > n_0$ and $|u| < u_0$, we have that for all $1 \leq \alpha < \alpha' \leq n$,*

$$\lim_{\beta\to\infty}\left[\langle L_0^{\alpha,\alpha'} L_1^{\alpha,\alpha'}\rangle_{\ell,\beta,u} - \langle L_{-1}^{\alpha,\alpha'} L_0^{\alpha,\alpha'}\rangle_{\ell,\beta,u}\right] > c \quad \text{for all } \ell \text{ odd;}$$

$$\lim_{\beta\to\infty}\left[\langle L_0^{\alpha,\alpha'} L_1^{\alpha,\alpha'}\rangle_{\ell,\beta,u} - \langle L_{-1}^{\alpha,\alpha'} L_0^{\alpha,\alpha'}\rangle_{\ell,\beta,u}\right] < -c \quad \text{for all } \ell \text{ even.}$$

**Fig. 3.** Illustration for dimerization. Depending on whether $\ell$ is even or odd, the site $x = 0$ is more entangled with its left or its right neighbor

Theorem 1.1 establishes the existence of at least two distinct infinite-volume ground states, close to the point B' of the phase diagram (see Fig. 3). Notice that the same result holds if we replace the operators $L_0^{\alpha,\alpha'} L_1^{\alpha,\alpha'}$ with spin operators $S_0^{(3)} S_1^{(3)}$, diagonal in the basis $\{e_\alpha\}$.

We expect that there are exactly two extremal ground states, precisely given by limits $\ell \to \infty$ along odd or even integers. We also expect that, if we take the chain to be $\{-\ell, -\ell + 1, \ldots, \ell\}$, the corresponding infinite-volume ground state is equal to the average of the two extremal states.

The next result shows that the ground state retains the $O(n)$ symmetry of the system, that there is no magnetic long range order. This is indeed an attribute of dimerisation.

**Theorem 1.2** *There exist constants $n_0, u_0, c_1, c_2, C > 0$ (independent of $\ell$) such that for $n > n_0$ and $|u| < u_0$, we have*

$$\lim_{\beta \to \infty} \left| \langle L_x^{\alpha,\alpha'} \, e^{-tH_\ell} \, L_y^{\alpha,\alpha'} \, e^{tH_\ell} \rangle_{\ell,\beta,u} \right| \leq C \, e^{-c_1|x-y|-c_2|t|}$$

*for all $\ell \in \mathbb{N}$, all $x, y \in \{-\ell + 1, \ldots, \ell\}$, all $1 \leq \alpha < \alpha' \leq n$, and all $t \in \mathbb{R}$.*

Dimerization has been established in [3,18] at the point B' in the phase diagrams of Figs. 1 and 2. The earlier result [18] uses the loop representation of [4] combined with a Peierls argument; it holds for $S \geq 8$ (or $n \geq 17$). The second result, due to Aizenman, Duminil-Copin and Warzel, remarkably holds for all $S \geq 1$ ($n \geq 3$), i.e., for all values of $S$ (or $n$) where dimerization is expected. It uses the loop representation and random cluster representation of [4] as well as recent results for the two-dimensional random cluster model [7,20].

Away from the point B' these methods do not apply. In this article we use the loop representation of [29], which combines those of [4,26], in order to get a contour model; see Theorem 2.1. The loop representation involves a probability measure for $u, v \leq 0$ only; it involves a signed measure otherwise. This is described in Sect. 2. For large $n$, typical configurations involve many loops, that are short loops located on all the dimerized edges. We define contours to be excitations with respect to this background. It is possible to obtain a contour model with piecewise compatible contours, that is suitable for a cluster expansion (Sect. 3). This method is robust regarding signs and it allows to intrude in the region with positive parameter $u$. Proving that the expansion converges is difficult, since the cost of excitations is entropic rather than energetic. This is done in Sect. 4. This allows to establish dimerization in the loop model, see Theorem 4.7. It is equivalent to Theorem 1.1, thus proving our main result. Theorem 1.2 is proved in Sect. 4.4.

The interaction that is responsible for dimerization is the operator $Q_{x,x+1}$ and we prove that dimerization is stable under perturbations of this interaction by $uT_{x,x+1}$, with $|u|$ sufficiently small. It should be possible to prove stability under more general

perturbations that are not necessarily invariant under the group $O(n)$. Since the unperturbed model is *not* frustration-free, this does not follow from the recent result about the stability of gapped phases with discrete symmetry breaking in [17], which requires the frustration-free property. For translation-invariant perturbations by $O(n)$ invariant next-nearest neighbor or further terms, the methods of this paper should generalize in a straightforward manner.

We now discuss the case of the spin chain with Hamiltonian

$$\tilde{H}_\ell = \sum_{x=-\ell+1}^{\ell-1} \left( u T_{x,x+1} + v P_{x,x+1} \right), \tag{1.14}$$

where $P$ is projection onto the singlet state (recall (1.6)). We have a similar result about dimerization. Let $S^{(i)}$, $i = 1, 2, 3$, be the spin operators that are the generators of the $SU(2)$ symmetry group for $\tilde{H}_\ell$. In the basis $|\alpha\rangle$ where $P$ is the projection onto the vector $\phi$ in (1.5), we can choose $S^{(3)}$ such that $S^{(3)}|\alpha\rangle = \alpha|\alpha\rangle$. Let

$$\langle S_x^{(i)} S_y^{(i)} \rangle^{\tilde{}}_{\ell,\beta,u} = \frac{1}{\mathrm{Tr}\, \mathrm{e}^{-\beta \tilde{H}_\ell}} \mathrm{Tr}\, S_x^{(i)} S_y^{(i)}\, \mathrm{e}^{-\beta \tilde{H}_\ell}. \tag{1.15}$$

**Theorem 1.3** *Let $v = -1$, and $i \in \{1, 2, 3\}$. There exist constants $n_0, u_0, c > 0$ (independent of $\ell$) such that for $n > n_0$ and $|u| < u_0$, we have*

$$\lim_{\beta \to \infty} \left[ \langle S_0^{(i)} S_1^{(i)} \rangle^{\tilde{}}_{\ell,\beta,u} - \langle S_{-1}^{(i)} S_0^{(i)} \rangle^{\tilde{}}_{\ell,\beta,u} \right] > c \quad \text{for all } \ell \text{ odd;}$$

$$\lim_{\beta \to \infty} \left[ \langle S_0^{(i)} S_1^{(i)} \rangle^{\tilde{}}_{\ell,\beta,u} - \langle S_{-1}^{(i)} S_0^{(i)} \rangle^{\tilde{}}_{\ell,\beta,u} \right] < -c \quad \text{for all } \ell \text{ even.}$$

When $n$ is odd this theorem is equivalent to Theorem 1.1, as the correlations of spin operators are the same as correlations of operators $L_{x,y}^{\alpha,\alpha'}$, up to some factors. In the case where $n$ is even, this is no longer the case and the proof needs to be adapted; the modifications are described in Appendix A.

*1.5. Gap for excitations.* Let $E_0^{(\ell)} < E_1^{(\ell)} < \dots$ be the eigenvalues of $H_\ell$, and $\tilde{E}_0^{(\ell)} < \tilde{E}_1^{(\ell)} < \dots$ be the eigenvalues of $\tilde{H}_\ell$. The gaps are defined as

$$\begin{aligned} \Delta^{(\ell)} &= E_1^{(\ell)} - E_0^{(\ell)}, \\ \tilde{\Delta}^{(\ell)} &= \tilde{E}_1^{(\ell)} - \tilde{E}_0^{(\ell)}. \end{aligned} \tag{1.16}$$

The gaps are obviously positive but the question is whether they are so uniformly in $\ell$.

**Theorem 1.4** *There exist constants $n_0, u_0, c > 0$ (independent of $\ell$) such that for $n > n_0$ and $|u| < u_0$, we have*

*(a) The multiplicities of $E_0^{(\ell)}$ and $\tilde{E}_0^{(\ell)}$ are equal to 1. (That is, ground states are unique.)*
*(b) $\Delta^{(\ell)} \geq c$ and $\tilde{\Delta}^{(\ell)} \geq c$ for all $\ell$.*

Recall that the chain is $\{-\ell + 1, \ldots, \ell\}$ and it always contains an even number of sites. Our theorem does not cover the chains with odd numbers of sites, although we expect the corresponding Hamiltonians to be gapped as well.

The spatial exponential decay proved in Theorem 1.2 is also a consequence of Theorem 1.4, due to the Exponential Clustering Theorem (see the simultaneous articles [9,16]). Our proof here is motivated by [11]. For the model $H_\ell$ it can be found in Sect. 5. It relies on a loop and contour representation, and on cluster expansions, as for the proof of dimerization. The modifications for $\tilde{H}_\ell$ are discussed in the appendix.

## 2. Graphical Representation for $O(n)$ Models

Consider the one-dimensional graph consisting of the $2\ell$ vertices $V_\ell := \{-\ell + 1, \ldots, \ell\}$ and the edges $E_\ell := \{(x, x + 1) : -\ell + 1 \leq x \leq \ell - 1\}$. Fix $\beta > 0$. To each vertex and edge of this graph we associate a periodic time interval $T_\beta = (-\beta, \beta)_{\text{per}}$ to obtain a set of *space-time vertices* $\overline{V}_{\ell,\beta} := V_\ell \times T_\beta$ as well as a set of *space-time edges* $\overline{E}_{\ell,\beta} := E_\ell \times T_\beta$.

By a *configuration* $\omega$ we mean a finite subset of $\overline{E}_{\ell,\beta}$, each point of $\omega$ receiving a *mark* $\times$ or $\sqsupset\!\sqsubset$. The points of $\omega$ will collectively be called *links*, those marked $\times$ being referred to as *crosses* and those marked $\sqsupset\!\sqsubset$ as *double-bars*. We write $\omega = (\omega_{\sqsupset\!\sqsubset}, \omega_{\times})$ and denote the set of all such (link) configurations $\Omega_{\ell,\beta}$.

To every configuration $\omega \in \Omega_{\ell,\beta}$ corresponds a set of *loops*; see Fig. 4 for an illustration. A loop $l$ is a closed, injective trajectory

$$[0, L]_{\text{per}} \rightarrow \overline{V}_\ell t \mapsto l(t) = (v(t), T(t)),$$

such that $x(t)$ is piecewise constant and $T'(t) \in \{\pm 1\}$. We call $L \equiv |l|$ the *length* of $l$, that is the smallest $L > 0$ in the above equation. A jump occurs at $t \in [0, L]$ provided that $\{x(t-), x(t+)\} \times T(t)$ contains a link. We have $T'(t+) = -T'(t-)$ in case that link is a double bar and $T'(t+) = T'(t-)$ in case it is a cross. We identify loops with identical support and we occasionally abuse notation and identify a loop with the set of links it traverses. The number of loops in a configuration $\omega$ is denoted $\mathcal{L}(\omega)$. The number of links in a configuration $\omega$ is denoted by $\#\omega$. Similarly the number of double bars is denoted $\#\omega_{\sqsupset\!\sqsubset}$ and the number of crosses is denoted by $\#\omega_{\times}$.

For $u \in \mathbb{R}$, we define the following *signed measure* on the set $\Omega_{\ell,\beta}$ of link configurations $\omega$:

$$\mathrm{d}\bar{\rho}_u(\omega) = u^{\#\omega_{\times}} \mathrm{d}^{\otimes\#\omega} x, \tag{2.1}$$

where $\mathrm{d}x$ is the Lebesgue measure on $\overline{E}_{\ell,\beta}$. We also introduce the following normalized measure $\rho_u$, satisfying $\rho_u(\Omega_{\ell,\beta}) = 1$:

$$\mathrm{d}\rho_u(\omega) = \mathrm{e}^{-(1+u)2\beta|E_\ell|} \mathrm{d}\bar{\rho}_u(\omega) \tag{2.2}$$

If $u$ is positive, the measure $\rho_u$ is a positive measure and hence a probability measure; in fact, under this measure $\omega$ has the distribution of a Poisson point process with intensity $u$ for crosses $\times$ and intensity 1 for double-bars $\sqsupset\!\sqsubset$. But we also allow small, negative $u$. Let

$$Z_{\ell,\beta,n,u} := \int_{\Omega_{\ell,\beta}} \mathrm{d}\rho_u(\omega)\, n^{\mathcal{L}(\omega)-\#\omega_{\sqsupset\!\sqsubset}}. \tag{2.3}$$

This loop model is equivalent to the quantum spin system, and the next result is an instance of this equivalence. The equivalence goes back to Tóth [26] and Aizenman–Nachtergaele [4] for special choices of the parameters; the general case of the interaction (1.2) is due to [29]. Note that it holds for arbitrary finite graphs, not only for chains.

We write $x \nleftrightarrow y$ to characterize the set of configurations $\omega$ where $(x, 0)$ and $(y, 0)$ belong to distinct loops; $x \overset{+}{\longleftrightarrow} y$ where the top of $(x, 0)$ is connected to the bottom of $(y, 0)$; and $x \overset{-}{\longleftrightarrow} y$ where the top of $(x, 0)$ is connected to the top of $(y, 0)$ (see [29, Fig. 2] for an illustration).

**Theorem 2.1** *For the Hamiltonian* (1.1) *with* $h_{x,x+1} = -u T_{x,x+1} - Q_{x,x+1}$, *we have that*

*(a)* $\mathrm{Tr}\, \mathrm{e}^{-2\beta H_\ell} = \mathrm{e}^{2\beta(1+u)|E_\ell|} Z_{\ell,\beta,n,u}$.
*(b) For all* $1 \leq \alpha < \alpha' \leq n$, *we have*

$$
\mathrm{Tr}\, L_x^{\alpha,\alpha'} L_y^{\alpha,\alpha'} \mathrm{e}^{-2\beta H_\ell}
$$
$$
= \frac{2}{n} \mathrm{e}^{2\beta(1+u)|E_\ell|} \int_{\Omega_{\ell,\beta}} \mathrm{d}\rho_u(\omega)\, n^{\mathcal{L}(\omega) - \#\omega_{\rightleftharpoons}} \big( \mathbb{1}[x \overset{-}{\longleftrightarrow} y] - \mathbb{1}[x \overset{+}{\longleftrightarrow} y] \big).
$$

The sign of the parameter $u$ in the definition of the interaction has indeed changed; but the theorem holds for arbitrary real (or even complex) parameters. Theorem 2.1 can also be formulated for the interaction $h_{x,x+1} = -u T_{x,x+1} - v Q_{x,x+1}$, by inserting the factor $v^{\#\omega_{\rightleftharpoons}}$ inside the integrals.

*Proof.* The proof of (a) can be found in [29, Theorem 3.2] and (b) is similar, so we only sketch it here. Let $\Sigma(\omega)$ be the set of "space-time spin configurations" that are constant along the loops (so that $|\Sigma(\omega)| = n^{\mathcal{L}(\omega)}$). By a standard Feynman-Kac expansion, we get

$$
\mathrm{Tr}\, \mathrm{e}^{-2\beta H_\ell} = \mathrm{e}^{2\beta(1+u)|E_\ell|} \int_{\Omega_{\ell,\beta}} \mathrm{d}\rho_u(\omega)\, n^{-\#\tilde{\omega}_{\rightleftharpoons}} \sum_{\sigma \in \Sigma(\omega)} 1. \tag{2.4}
$$

We recognize the partition function in (2.3), so we get (a).

For (b) we need a modified set of space-time spin configurations where the spin value must jump from $\alpha$ to $\alpha'$, or from $\alpha'$ to $\alpha$, at the points $(x, 0)$ and $(y, 0)$. Let $\Sigma_{x,y}^{\alpha,\alpha'}$ be this set. We then have

$$
\mathrm{Tr}_x^{\alpha,\alpha'} L_y^{\alpha,\alpha'} \mathrm{e}^{-2\beta H_\ell} = \mathrm{e}^{2\beta(1+u)|E_\ell|} \int_{\Omega_{\ell,\beta}} \mathrm{d}\rho_u(\omega)\, n^{-\#\tilde{\omega}_{\rightleftharpoons}}
$$
$$
\sum_{\sigma \in \Sigma_{x,y}^{\alpha,\alpha'}(\omega)} \langle \sigma_{x,0+}| L_x^{\alpha,\alpha'} |\sigma_{x,0-}\rangle \langle \sigma_{y,0+}| L_y^{\alpha,\alpha'} |\sigma_{y,0-}\rangle. \tag{2.5}
$$

It is necessary that $(x, 0)$ and $(y, 0)$ belong to the same loop in order to get a nonzero contribution. Further, we have

$$
\langle \sigma_{x,0+}| L_x^{\alpha,\alpha'} |\sigma_{x,0-}\rangle \langle \sigma_{y,0+}| L_y^{\alpha,\alpha'} |\sigma_{y,0-}\rangle = \begin{cases} -1 & \text{if } x \overset{+}{\longleftrightarrow} y, \\ +1 & \text{if } x \overset{-}{\longleftrightarrow} y. \end{cases} \tag{2.6}
$$

Since $|\Sigma_{x,y}^{\alpha,\alpha'}(\omega)| = \frac{2}{n} n^{\mathcal{L}(\omega)}$, we get (b). $\qquad \square$

From now on and to the end of this article we work with the loop model.

*Remark 2.2* (*Intuition*). It is helpful to think of $\rho_u$ as an a-priori measure on a gas of loops, and rewrite the integrand $n^{\mathcal{L}(\omega)-\#\omega_{\leftrightharpoons}}$ as $e^{-(\log n)H(\omega)}$, with 'Hamiltonian'

$$-H(\omega) := \mathcal{L}(\omega) - \#\omega_{\leftrightharpoons}, \tag{2.7}$$

and inverse temperature $\log n$. Thinking of $n$ as large, the Laplace principle tells us that 'typical' configurations should maximise $n^{\mathcal{L}(\omega)-\#\omega_{\leftrightharpoons}}$. Our goal is to write $Z_{\ell,\beta,n,u}$ as a dominant contribution from such maximizers, and some excitations.

We end this section with the following remark about working with a signed measure. Since the (possibly signed) measure $\rho_u$ is closely related to the probability measure $\rho_1$, it is easy to see that any event $A$ satisfying $\rho_1(A) = 0$ also has zero measure under $\rho_u$. In fact, we have the following slightly stronger property:

**Lemma 2.3** *If $A$ is an event such that $\rho_1(A) = 0$ and $f : \Omega_{\ell,\beta} \to \mathbb{R}$ is a $\rho_1$-integrable function, then for any $u \in \mathbb{R}$ we have that*

$$\int_A d\rho_u(\omega) f(\omega) = 0. \tag{2.8}$$

*Proof.* Using (2.1) and (2.2) it is easy to see that

$$\left| \int_A d\rho_u(\omega) f(\omega) \right| \leq C \int_A d\rho_1(\omega) |f(\omega)| = 0, \tag{2.9}$$

for some finite constant $C$ depending only on $u, \ell, \beta$. □

As a consequence, we may assume that crosses and double-bars occur at different times, also when $u < 0$ and the measure $\rho_u$ carries signs. We implicitly used this property when defining loops.

## 3. The Contour Model

*3.1. Contours.* We classify loops as follows, see Fig. 4. A loop is *contractible* if it can be continuously deformed to a point and *winding* otherwise. Not all loops are contractible since our time interval $T_\beta$ is periodic. A loop is *long* if it visits three or more distinct vertices or if it is winding; it is *short* otherwise.
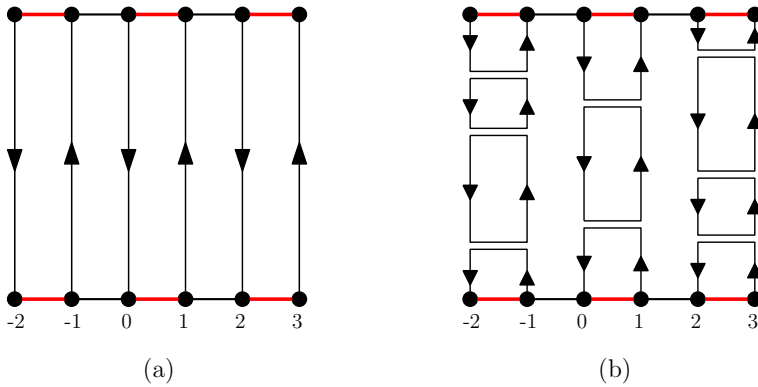
We define a *canonical orientation* of the space-time vertices $\overline{V}_{\ell,\beta}$, using the directions up ($\uparrow$) and down ($\downarrow$), by orienting the leftmost space-time vertex $\{-\ell+1\} \times T_\beta$ down $\downarrow$ and requiring that neighbouring space-time vertices have opposite orientations; see Fig. 5. We write $V_\ell^\uparrow := \{x \in V_\ell : x + \ell$ is even$\}$ for the set of vertices with up-orientation, and $V_\ell^\downarrow := \{x \in V_\ell : x + \ell$ is odd$\}$ for the set of vertices with down-orientation, and introduce the following subsets of the edge-set $E_\ell$:

$$\begin{aligned} E_\ell^+ &:= \left\{ (x, x+1) \in E_\ell : x \in V_\ell^\downarrow, x+1 \in V_\ell^\uparrow \right\}, \\ E_\ell^- &:= E_\ell \setminus E_\ell^+ = \left\{ (x, x+1) \in E_\ell : x \in V_\ell^\uparrow, x+1 \in V_\ell^\downarrow \right\}. \end{aligned} \tag{3.1}$$

We define $\overline{E}_{\ell,\beta}^+$ and $\overline{E}_{\ell,\beta}^-$, as well as $\overline{V}_{\ell,\beta}^\uparrow$ and $\overline{V}_{\ell,\beta}^\downarrow$, analogously.
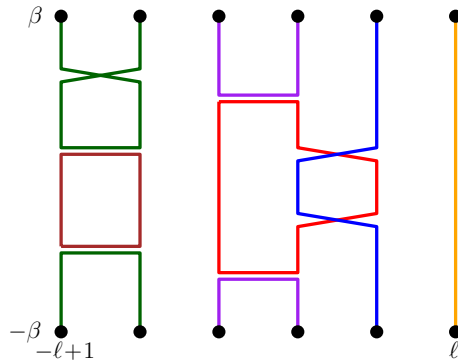
**Fig. 4.** A configuration $\omega$ consisting of three short loops (green, brown, purple), and three long loops (red, blue, orange) two of which are winding loops (blue, orange)



**Fig. 5.** **a** The canonical orientation of $\overline{V}_{\ell,\beta}$ with the set $E_\ell^+$ highlighted red. **b** a configuration $\omega$ with many short loops; these are positively oriented under the canonical orientation

These definitions are motivated as follows. We expect that 'typical' configurations $\omega$ contain many short loops. To maximize the number of short loops one places only double-bars in $\overline{E}_\ell^+$, as in Fig. 5b. The canonical orientation is chosen so that all the short loops in such a configuration are positively oriented (i.e. counter-clockwise). The canonical orientation will be useful in classifying the excitations away from such 'typical' $\omega$. Also note that if the origin 0 belongs to a short, positively oriented loop, then we have $0 \leftrightarrow 1$ for $\ell$ odd and $0 \leftrightarrow -1$ for $\ell$ even. To prove our main result Theorem 1.1 we will essentially argue that the origin is likely to belong to a short, positively oriented loop.

Given a loop $l$ in a configuration $\omega$, we define a *segment* of $l$ as a trajectory of $l$ between two times $0 \leq s_1 < s_2 \leq L(l)$ when $l$ passes through height $\beta$. That is to say, $l(s_1) = (v_1, \beta), l(s_2) = (v_2, \beta)$ for some $v_i \in V_\ell$, while $l$ does not pass through height $\beta$ in times $t \in (s_1, s_2)$. We say that a segment is *spanning* if for every $t \in T_\beta$ there exists a $v = v(t) \in V_\ell$ such that the segment traverses $(v, t)$. Note that a spanning segment is not necessarily part of a winding loop. See Fig. 6.

**Definition 3.1** (*Contours*). We say that two loops are *connected* if they share a link or both are winding. A *contour* is then a maximally connected set of long loops.
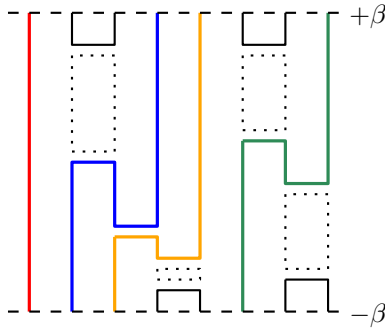
**Fig. 6.** The leftmost and rightmost loops are winding loops with one spanning segment each. The loop in the middle is contractible. There are four spanning segments in total
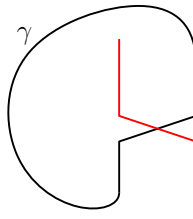


**Fig. 7.** A cross is traversed by a contour $\gamma$. If the red loop visits a third vertex, it is a long loop; otherwise it must be a winding loop. In both cases, it is actually part of $\gamma$

*Remark 3.2* For later reference, we note here that any cross $\rtimes$ which is traversed by some loop in a contour is necessarily traversed both ways by the contour; see Fig. 7.

A contour which contains at least one winding loop will be called a *winding contour*. See Fig. 8.

We need a notion of *interior* of a contour, and for this it is useful to regard our configuration $\omega$ as living in the bi-infinite cylinder $C_\beta = \mathbb{R} \times T_\beta$. More precisely, given $\omega$ we consider the subset $\overline{\omega}$ of $C_\beta$ obtained as the union of (i) $\overline{V}_\beta$ embedded in $C_\beta$ in the natural way, and (ii) the links of $\omega$ embedded as straight line segments connecting adjacent points of $\overline{V}_\beta$. Note that, in the embedding $\overline{\omega}$, crosses and double-bars are embedded in the same way. For a loop $l$ of $\omega$, define its *support $S(l)$* as the subset of $\overline{\omega}$ traced out by $l$, meaning the union of the vertical and horizontal line segments of $\overline{\omega}$ corresponding to the intervals of $\overline{V}_\beta$ and the links of $\omega$ traversed by $l$. For a contour $\gamma$ of $\omega$ we then make the following definitions.

- The *support $S(\gamma)$* is the union of the supports $S(l)$ of the loops $l$ belonging to $\gamma$. Note that $S(\gamma)$ is a closed subset of $C_\beta$.
- The *exterior $E(\gamma)$* is the union of the unbounded connected components of $C_\beta \setminus S(\gamma)$. Note that $E(\gamma)$ is open.
- The *interior $I(\gamma) := C_\beta \setminus \overline{E(\gamma)}$*. Note that $I(\gamma)$ is an open set.
- The *boundary $B(\gamma) := \overline{E(\gamma)} \setminus E(\gamma)$* which is a closed set.
- The (vertical) *length $|\gamma|$* of a contour as the sum of the (vertical) lengths of its loops, $|\gamma| := \sum_{l \in \gamma} |l|$.

These notions are illustrated in Figs. 9, 10 and 11.
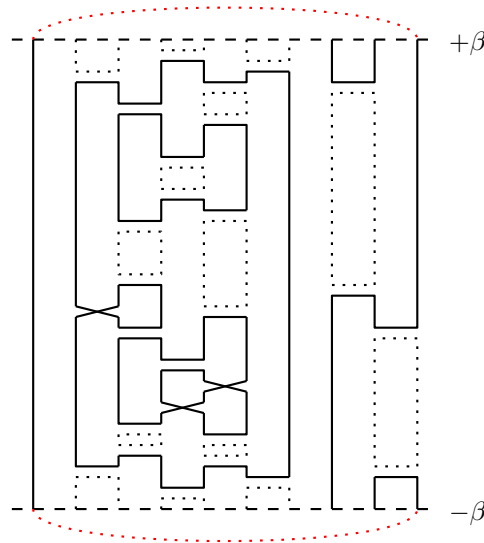
**Fig. 8.** Two contours: One winding contour, consisting of two winding loops, and one consisting of four long, but contractible loops
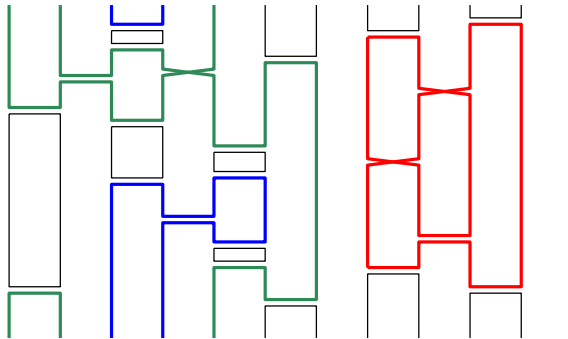


**Fig. 9.** A configuration $\omega$ with three contours highlighted green, blue and red. The green contour consists of two winding loops

Having defined $I(\gamma)$ as a subset of the cylinder $C_\beta$, we may also regard $I(\gamma)$ (or more precisely, its closure $\overline{I(\gamma)}$) as a subset of $\overline{E}_{\ell,\beta}$ by identifying a point $(x, x+1) \times \{t\} \in \overline{E}_{\ell,\beta}$ with the closed line-segment from $(x, t)$ to $(x + 1, t)$ in $C_\beta$. Similarly, $S(\gamma)$ and $B(\gamma)$ may be regarded as subsets of $\overline{V}_{\ell,\beta} \cup \omega$. We freely switch between these points of view.

Fixing a contour $\gamma$, note that the boundary $B(\gamma)$ consists of a collection of closed curves and horizontal line segments (of length 1). We use the canonical orientation of $\overline{V}_{\ell,\beta}$ to orient each vertical segment of $B(\gamma)$. It is not hard to see that this gives a consistent orientation of all the closed curves constituting $B(\gamma)$. (This follows from Remark 3.2.) Recall the standard notion of a *positively oriented curve* as one whose interior is always on the left.

**Definition 3.3** (*Type of a contour*). We say that the contour $\gamma$ is of *positive type* if the canonical orientation of $B(\gamma)$ is *positive* in the sense that $I(\gamma)$ is on the left of each
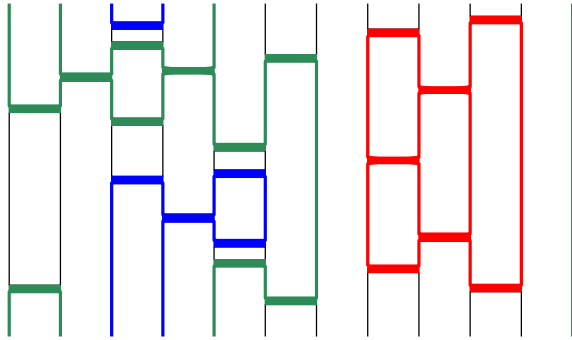
**Fig. 10.** The corresponding embedding $\overline{\omega} \subseteq C_\beta = T_\beta \times \mathbb{R}$, with the supports $S(\gamma)$ of the contours highlighted with the corresponding colors
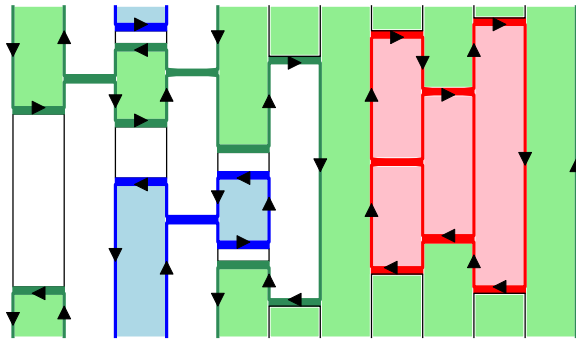


**Fig. 11.** The interiors of the corresponding contours with the boundaries $B(\gamma)$ receiving the canonical orientation. The green and blue contours are of positive type (interiors $I(\gamma)$ on the left) while the red contour is of negative type (interior on the right)

closed curve of $B(\gamma)$. Otherwise we say that $\gamma$ is of *negative type* (being of negative type is equivalent to the interior being on the right).

*Remark 3.4* Suppose that $\omega \in \Omega_{\ell,\beta}$ is such that a given point $\bar{v} \in \overline{V}_{\ell,\beta}$ is not on or inside any contour, that is to say

$$\bar{v} \in \bigcap_{\gamma \in \Gamma(\omega)} E(\gamma), \tag{3.2}$$

where $E(\gamma)$ is the exterior of $\gamma$ defined above. Then we have that $\bar{v}$ is on a *positively oriented* short loop. Indeed, this is related to the fact that all external contours are of positive type, see Lemma 3.6.

*3.2. Domains and admissibility of contours.* We now introduce several notations and definitions pertaining to contours and how they relate to each other. First, given $\omega \in \Omega_{\ell,\beta}$ we define $\Gamma(\omega) = \{\gamma_1, \ldots, \gamma_k\}$ as the set of contours in the configuration $\omega$. Here, and in what follows, a contour may be identified with the set of links it traverses. The collection of all possible contours will be denoted $X_{\ell,\beta} = \bigcup_{\omega \in \Omega_{\ell,\beta}} \Gamma(\omega)$, and we write
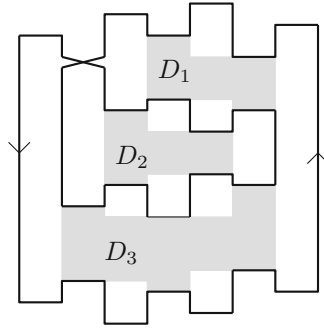
**Fig. 12.** A contour $\gamma$ of positive type, containing three domains $D_1, D_2, D_3$. Domains $D_1$ and $D_3$ are of negative type, while $D_2$ is of positive type

$X_{\ell,\beta}^+ \subseteq X_{\ell,\beta}$ for the collection of *positive-type* contours. We write

$$\mathfrak{X}_{\ell,\beta} = \bigcup_{k \geq 0} \binom{X_{\ell,\beta}}{k} \quad \text{and} \quad \mathfrak{X}_{\ell,\beta}^+ = \bigcup_{k \geq 0} \binom{X_{\ell,\beta}^+}{k} \tag{3.3}$$

for the set of finite collections of contours, respectively positive-type contours. Elements of $\mathfrak{X}_{\ell,\beta}$ and of $\mathfrak{X}_{\ell,\beta}^+$ will usually be denoted by $\Gamma$. It is important to note that far from every such set $\Gamma$ of contours can be obtained as $\Gamma(\omega)$ for some $\omega \in \Omega_{\ell,\beta}$; in fact, we will devote some effort to identifying criteria under which such an $\omega$ does indeed exist. We say that $\Gamma \in \mathfrak{X}_{\ell,\beta}$ is *admissible* if $\Gamma = \Gamma(\omega)$ for some $\omega \in \Omega_{\ell,\beta}$, and write $\mathfrak{A}_{\ell,\beta} = \Gamma(\Omega_{\ell,\beta})$ for the collection of admissible sets of contours.

Recall that the interior $I(\gamma)$ of a contour $\gamma$ is by definition an open subset of the cylinder $C_\beta$. Also recall that we regard $\overline{E}_{\ell,\beta}$ as a closed subset of $C_\beta$ by identifying a point $(x, x+1) \times \{t\} \in \overline{E}_{\ell,\beta}$ with the closed line-segment from $(x, t)$ to $(x+1, t)$. We now define the *(interior) domains* of $\gamma$ as follows.

**Definition 3.5** A domain $D$ of $\gamma$ is a subset of $\overline{E}_{\ell,\beta} \cap I(\gamma)$ which, when regarded as a subset of $C_\beta$ as above, is connected, satisfies $D \cap S(\gamma) = \varnothing$, and is maximal with these properties.

We define the *type* of a domain in a similar way to the type of a contour. Namely, we orient the (topological) boundary of $D$ consistently with the canonical orientation of $\overline{V}_{\ell,\beta}$ and say that $D$ is of *positive type* if this is a positive orientation (interior on the left), and of *negative type* otherwise. See Fig. 12.

Given two contours $\gamma$ and $\gamma'$, we say that $\gamma$ is a *descendant* of $\gamma'$, writing $\gamma \prec \gamma'$, if $S(\gamma) \subseteq D$ for some domain $D$ of $\gamma'$. Given $\Gamma \in \mathfrak{X}_{\ell,\beta}$ and $\gamma, \gamma' \in \Gamma$, we say that $\gamma$ is an *immediate descendant* of $\gamma'$ in $\Gamma$ if $\gamma \prec \gamma'$ and there is no $\overline{\gamma} \in \Gamma$ satisfying both $\gamma \prec \overline{\gamma}$ and $\overline{\gamma} \prec \gamma'$. It is important to note that the notion of being an *immediate* descendant depends not only on the two contours $\gamma$ and $\gamma'$ but on the set $\Gamma$; in other words, immediate descendancy cannot be checked in a pairwise manner. If $\gamma \in \Gamma$ is not the descendant of any other contour $\gamma' \in \Gamma$ then we say that $\gamma$ is an *external* contour; this notion is also dependent on the set $\Gamma$.

Note that the unique (if it exists) winding contour is always external since a winding loop cannot be in the interior of any contractible loop.
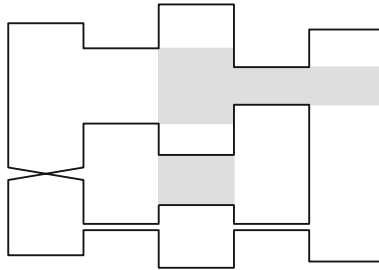
**Fig. 13.** A contour $\gamma$ and its two appropriately shifted domains (shaded areas). The lower one was not moved since it already was positive type. The upper one was shifted one column to the right. If a $\gamma' \in X_{\ell,\beta}$ gets placed inside it, $S^{-1}(\Gamma = (\gamma, \gamma'))$ will return an admissible collection of contours

**Lemma 3.6** *Fix $\Gamma \in \mathfrak{X}_{\ell,\beta}$. Then $\Gamma$ is admissible, i.e. $\Gamma \in \mathfrak{A}_{\ell,\beta}$, if and only if the following hold:*

*(1) all external contours in $\Gamma$ are of positive type;*
*(2) for any pair of distinct contours $\gamma, \gamma' \in \Gamma$ we have that either $\overline{I(\gamma)} \cap \overline{I(\gamma')} = \varnothing$ or $\gamma \prec \gamma'$ or $\gamma' \prec \gamma$;*
*(3) if $\gamma$ is an immediate descendant of $\gamma'$, in a domain $D$ of $\gamma'$, then the types of $\gamma$ and of $D$ coincide;*
*(4) there exists at most one winding contour $\gamma \in \Gamma$.*

*Proof.* It is easy to see that the four conditions above hold for any admissible $\Gamma = \Gamma(\omega)$. To show the converse, we construct an explicit $\omega \in \Omega_{\ell,\beta}$ with $\Gamma(\omega) = \Gamma$. Starting from the empty configuration $\omega_0 = \varnothing \in \Omega_{\ell,\beta}$, add all links of all external contours and then place a double bar at height 0, say, on each $e \in E_\ell^+$ that does not have any link on it. This defines $\omega_1$ such that $\Gamma(\omega_1)$ is precisely the set of external contours of $\Gamma$. Next, add the links of all contours which are immediate descendants of external contours. This does not create any new long loops apart from those in these contours because their types coincide with those of the domains they are in. Iterate this procedure until there are no more contours left to add. $\qquad\square$

An important prerequisite for applying a cluster expansion is to be able to verify the admissibility of a set of contours in a pairwise manner. As indicated above, and in the light of Lemma 3.6, this is not directly possible since the notion of being an immediate descendant depends on the whole set $\Gamma$. We get around this issue by introducing a notion of *compatibility* which applies to sets of positive-type contours $\Gamma \in \mathfrak{X}_{\ell,\beta}^+$, and which can be checked in a pairwise manner. We then show that there is a bijective correspondence between admissible and compatible sets of contours.

The bijective correspondence referred to above involves *shifting* contours and rests on the simple observation that if $\gamma$ is a negative-type contour, then $\gamma' = \gamma + (1, 0)$ (i.e. $\gamma$ translated to the right one unit) is a positive-type contour.

Given a positive-type contour $\gamma \in X_{\ell,\beta}^+$ with domains $D_1(\gamma), \ldots, D_k(\gamma) \subseteq I(\gamma)$, we define the *appropriately shifted domains* $D_i^+(\gamma)$ of $\gamma$ by

$$D_i^+(\gamma) = \begin{cases} D_i(\gamma), & \text{if } D_i(\gamma) \text{ is of positive type,} \\ D_i(\gamma) + (1, 0), & \text{otherwise.} \end{cases} \tag{3.4}$$

Note that while $D_i^+(\gamma) \subseteq \overline{I(\gamma)}$, a shifted domain may intersect the boundary $B(\gamma)$. See Fig. 13.
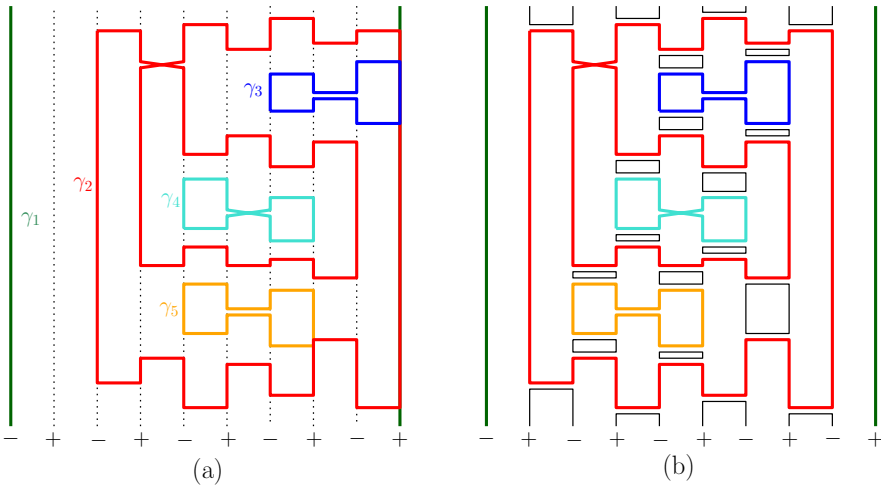
**Fig. 14.** (a) A compatible set of contours $\Gamma = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5\}$. We have, for example, $\sigma_\Gamma(\gamma_3) = 2$ since $\gamma_3$ lies in a shifted domain of both $\gamma_1$ and $\gamma_2$, while $\sigma_\Gamma(\gamma_4) = 1$ since it lies in a shifted domain of $\gamma_1$ only. (b) The admissible set $\Sigma(\Gamma)$

**Definition 3.7** Given two positive-type contours $\gamma, \gamma' \in X_{\ell,\beta}^+$, we say that $\gamma$ and $\gamma'$ are *compatible* if one of the following hold:

(1) $\overline{I(\gamma)} \cap \overline{I(\gamma')} = \varnothing$, or
(2) $S(\gamma) \subseteq D_i^+(\gamma')$ for some $i$, or
(3) $S(\gamma') \subseteq D_i^+(\gamma)$ for some $i$, or
(4) at least one of $\gamma$ or $\gamma'$ is not a winding contour.

We define

$$\delta(\gamma, \gamma') = \begin{cases} 1 & \text{if } \gamma, \gamma' \text{ are compatible} \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Finally, we let $\mathfrak{C}_{\ell,\beta}^+ \subseteq \mathfrak{X}_{\ell,\beta}^+$ to be the collection of all pairwise compatible sets of positive-type contours; that is, $\Gamma = \{\gamma_1, \ldots, \gamma_k\} \in \mathfrak{X}_{\ell,\beta}^+$ belongs to $\mathfrak{C}_{\ell,\beta}^+$ if $\prod_{1 \le i < j \le k} \delta(\gamma_i, \gamma_j) = 1$.

A compatible set $\Gamma$ is generally itself not admissible, since compatible contours may overlap but admissible contours may not. Intuitively, one obtains an admissible set of contours from a compatible set by 'shifting back' the appropriately shifted domains $D_i^+(\gamma)$ and the contours they contain. For nested contours, the shift is performed iteratively; see Fig. 14.

More formally, define the *shift* $\Sigma : \mathfrak{C}_{\ell,\beta}^+ \to \mathfrak{X}_{\ell,\beta}$ as follows. First, given $\Gamma \in \mathfrak{C}_{\ell,\beta}^+$ and $\gamma \in \Gamma$, write $\sigma_\Gamma(\gamma)$ for the number of contours $\gamma' \in \Gamma \setminus \{\gamma\}$ such that $\gamma \subseteq D_i^+(\gamma') \ne D_i(\gamma')$. This represents the number of times $\gamma$ is shifted to the right in order to obtain the compatible set $\Gamma$ from an admissible set of contours. We define

$$\Sigma(\Gamma) = \{\gamma - (\sigma_\Gamma(\gamma), 0) : \gamma \in \Gamma\}. \tag{3.6}$$

**Lemma 3.8** *The shift $\Sigma$ is a bijection from $\mathfrak{C}^+_{\ell,\beta}$, the collection of compatible sets of contours, to $\mathfrak{A}_{\ell,\beta}$, the collection of admissible sets of contours.*

*Proof.* It is easy to construct an inverse $\Sigma^{-1}$ of $\Sigma$ on $\mathfrak{A}_{\ell,\beta}$, as follows. Given $\Gamma \in \mathfrak{A}_{\ell,\beta}$, start with an external contour $\gamma$ (which is of positive type by Lemma 3.6) and form its appropriately shifted domains $D^+_i(\gamma)$. In doing so, shift also the descendants of $\gamma$ along with their domains. Note that all the immediate descendants of $\gamma$ are then mapped to positive type contours. Then iteratively continue this procedure for the (shifted) immediate descendants of $\gamma$. The resulting set $\Sigma^{-1}(\Gamma)$ then satisfies Definition 3.7.

It remains to show that $\Sigma(\Gamma) \in \mathfrak{A}_{\ell,\beta}$ for all $\Gamma \in \mathfrak{C}^+_{\ell,\beta}$, i.e. that $\Sigma(\Gamma)$ satisfies Lemma 3.6. Compatibility ensures that there is at most one winding contour. It is clear that external contours are of positive type since they are not shifted. For $\gamma, \gamma'$ with disjoint interiors, this property is preserved by $\Sigma$; if $\gamma \subseteq D^+_i(\gamma')$ then the shifting ensures that the images of $\gamma, \gamma'$ under $\Sigma$ satisfy $\gamma \prec \gamma'$, while the relative amounts by which the contours are shifted ensures that the types of immediate descendants in $\Sigma(\Gamma)$ coincide with the types of the relevant domains. $\qquad\square$

We close this subsection with a simple lemma about counting the amount of 'available space' for short loops in a configuration $\omega$, in terms of the lengths of the contours. For $\Gamma \in \mathfrak{A}_{\ell,\beta}$, we define the *free set* $F(\Gamma) \subseteq \overline{E}_{\ell,\beta}$ as the space-time edges where we can add links without modifying the contours in $\Gamma$ or creating new ones.

**Lemma 3.9** *Let $\Gamma \in \mathfrak{A}_{\ell,\beta}$ be an admissible set of contours. Then*

$$|F(\Gamma)| = |\overline{E}^+_{\ell,\beta}| - \tfrac{1}{2}\textstyle\sum_{\gamma \in \Gamma} |\gamma|. \tag{3.7}$$

*Proof.* We need to show that $2|\overline{E}^+_{\ell,\beta}| = 2|F(\Gamma)| + \sum_{\gamma \in \Gamma} |\gamma|$. Note that $2|\overline{E}^+_{\ell,\beta}| = |\overline{V}_{\ell,\beta}|$ and that $2|F(\Gamma)|$ equals the total length of all the short loops. But any point in $\overline{V}_{\ell,\beta}$ lies either on a contour or on a short loop, thus $|\overline{V}_{\ell,\beta}| = 2|F(\Gamma)| + \sum_{\gamma \in \Gamma} |\gamma|$, as required. $\qquad\square$

*3.3. Decomposition of $H(\omega)$.* Recall from (2.7) the quantity $-H(\omega) = \mathcal{L}(\omega) - |\omega_{\rightleftharpoons}|$. We now show that $H(\omega)$ can be decomposed as a sum over contours and we prove bounds on the summands. To this end, for a loop $l$ let $\mathcal{T}(l)$ denote the number of *turns* that $l$ makes; symbolically $\mathcal{T}(l) = \#\smile + \#\frown$. For a contour $\gamma$, write $\mathcal{T}(\gamma)$ for the total number of U-turns of all loops in $\gamma$. Next define the function $h : X_{\ell,\beta} \to \mathbb{Z}$ by

$$h(\gamma) = \mathcal{L}(\gamma) - \tfrac{1}{2}\mathcal{T}(\gamma) \tag{3.8}$$

where $\mathcal{L}(\gamma)$ denotes the number of loops in the contour $\gamma$.

**Lemma 3.10** *For $\omega \in \Omega_{\ell,\beta}$ with contours $\Gamma = \Gamma(\omega)$ we have $-H(\omega) = \sum_{\gamma \in \Gamma} h(\gamma)$.*

*Proof.* Since every double-bar of $\omega$ accounts for exactly two turns (of either one or two loops), we have

$$-H(\omega) = \sum_l \left(1 - \tfrac{1}{2}\mathcal{T}(l)\right) \tag{3.9}$$

where the sum is over all loops $l$ in the configuration $\omega$. The result now follows from the observation that short, non-winding loops make exactly two turns. $\qquad\square$

Write $\#\gamma_\sqcup$ for the number of double-bars visited by $\gamma$ and $\#\gamma_\ltimes$ for the number of crosses.

**Lemma 3.11** *For contours $\gamma$ without crosses, the function $h : X_{\ell,\beta} \to \mathbb{Z}$ satisfies*

$$h(\gamma) \leq -\tfrac{1}{3}\#\gamma + 2\ell\, \mathrm{I\!I}\{\gamma \text{ has a spanning segment}\}. \tag{3.10}$$

Note that the constant $-\tfrac{1}{3}$ is tight for the smallest non-winding contours with six double-bars and no crosses, while for larger contours the constant may be taken closer to $-\tfrac{1}{2}$. As to the indicator function, we will see that contours containing spanning segments become very rare asymptotically.

*Proof.* Write $\mathcal{W}(l)$ for the number of winding segments in $l$ and $\mathcal{W}(\gamma) = \sum_{l \in \gamma} \mathcal{W}(l)$. We claim that it suffices to show that $h(\gamma) \leq r(\gamma)$ where

$$r(\gamma) = -\tfrac{1}{3}\mathcal{T}(\gamma) + \mathcal{W}(\gamma). \tag{3.11}$$

Indeed, $r(\gamma)$ is bounded above by the right-hand-side of (3.10) for the following reasons:

- double-bars visited twice by $\gamma$ count twice in $\mathcal{T}(\gamma)$ but only once in $\#\gamma = \#\gamma_\sqcup$, while those visited once by $\gamma$ count once in both, meaning that $\mathcal{T}(\gamma) \geq \#\gamma$;
- $\mathcal{W}(\gamma) \leq 2\ell\, \mathrm{I\!I}\{\gamma \text{ has a spanning segment}\}$ since each point of the form $(x, 0) \in \overline{V}_{\ell,\beta}$ is visited by at most one winding segment.

Next, the claimed inequality $h(\gamma) \leq r(\gamma)$ is equivalent to:

$$\mathcal{T}(\gamma) + 6\mathcal{W}(\gamma) \geq 6\mathcal{L}(\gamma). \tag{3.12}$$

To establish (3.12), first note that both sides are additive over loops. Thus it suffices to show that any long or winding loop $l$ satisfies

$$\mathcal{T}(l) + 6\mathcal{W}(l) \geq 6. \tag{3.13}$$

If $\mathcal{W}(l) \geq 1$ this is clear, hence we may assume that the loop is non-winding. A long, non-winding loop which traverses only double-bars necessarily makes at least 6 turns, see Fig. 15. This proves (3.13) and hence the claim.  □

**Lemma 3.12** *For all contours $\gamma \in X_{\ell,\beta}$ we have*

$$h(\gamma) \leq 2\ell\, \mathrm{I\!I}\{\gamma \text{ has a spanning segment}\}. \tag{3.14}$$

*In particular $h(\gamma) \leq 0$ for all non-winding contours.*

*Proof* Note that $h(\gamma)$ is additive over loops $l \in \gamma$. So since there can be at most $2\ell$ spanning segments, it suffices to show for every loop that $1 - \tfrac{1}{2}(\#\sqcup + \#\sqcap) \leq \mathrm{I\!I}\{l \text{ has a spanning segment}\}$. This is clearly true.  □

**Lemma 3.13** *For $\gamma \in X_{\ell,\beta}$, all $u$ with $|u| \leq 1$, and all $\kappa > 0$, we have*

$$n^{h(\gamma)}|u|^{\#\gamma_\ltimes} \leq \min(n, |u|^{-\kappa/2})^{-(\frac{1}{3}-\kappa)\#\gamma} n^{2\ell\, \mathrm{I\!I}\{\gamma \text{ has a spanning segment}\}}. \tag{3.15}$$

*Proof.* If $\gamma$ has no spanning segment and $\#\gamma_{\times} \geq \frac{\kappa}{2}(\frac{1}{3} - \kappa)\#\gamma$, the claim follows from $h(\gamma) \leq 0$ (Lemma 3.12). Now consider the case where $\gamma$ has no spanning segment and $\#\gamma_{\times} \leq \frac{\kappa}{2}(\frac{1}{3} - \kappa)\#\gamma$. If $\#\gamma < (\frac{\kappa}{2}(\frac{1}{3} - \kappa))^{-1}$, then $\#\gamma_{\times} = 0$ and thus we may apply Lemma 3.11 to get the desired bound in this case. So assume now that $1 \leq \frac{\kappa}{2}(\frac{1}{3} - \kappa)\#\gamma$ (and still that $\gamma$ has no spanning segment and $\#\gamma_{\times} \leq \frac{\kappa}{2}(\frac{1}{3} - \kappa)\#\gamma$).

Let $\Gamma$ denote the collection of contours and small loops obtained by removing all crosses from $\gamma$, and let $m$ denote the number of small loops in $\Gamma$. Since the removal of a cross can only create at most one more loop, we have that $m \leq \#\gamma_{\times} + 1 \leq \frac{\kappa}{2}(\frac{1}{3} - \kappa)\#\gamma + 1 \leq \kappa(\frac{1}{3} - \kappa)\#\gamma$, and that

$$h(\gamma) \leq h(\Gamma) + \#\gamma_{\times} \leq h(\Gamma) + \frac{\kappa}{2}(\frac{1}{3} - \kappa)\#\gamma. \tag{3.16}$$

Since every short loop uses at most two double bars, the number of double bars belonging to contours of $\Gamma$ is at least

$$\#\gamma_{=} - 2m \geq \left(1 - \frac{\kappa}{2}(\frac{1}{3} - \kappa)\right)\#\gamma - 2\kappa(\frac{1}{3} - \kappa)\#\gamma = \left(1 - \frac{5}{2}\kappa(\frac{1}{3} - \kappa)\right)\#\gamma. \tag{3.17}$$

Applying Lemma 3.11, this allows us to conclude that

$$h(\Gamma) \leq -\frac{1}{3}\left(1 - \frac{5}{2}\kappa(\frac{1}{3} - \kappa)\right)\#\gamma. \tag{3.18}$$

Combining (3.16) and (3.18) we conclude that

$$h(\gamma) \leq -\left[\frac{1}{3} - 3\kappa(\frac{1}{3} - \kappa)\right]\#\gamma \leq -(\frac{1}{3} - \kappa)\#\gamma. \tag{3.19}$$

It remains to show the claim for $\gamma$ with a spanning segment. To this end, consider $\omega$ obtained as follows: Denote by $\omega_0$ a configuration of links such that its set of contours $\Gamma(\omega_0) = \{\gamma\}$. Now add $|E_\ell^+| = \ell$ double bars at the same height, exactly one per column in $E_\ell^+$. Denote this configuration by $\omega$ and note that $\Gamma(\omega)$ does not contain any winding contours. Observe that the number of crosses is unchanged and we added $\ell$ links, hence changing the number of loops by at most $\ell$. Using these observations and Lemma 3.10 we thus get

$$h(\gamma) = -H(\omega_0) = -H(\omega) + \mathcal{E} = \sum_{\gamma' \in \Gamma(\omega)} h(\gamma') + \mathcal{E}, \tag{3.20}$$

where $\mathcal{E}$ is an error that is bounded by $|\mathcal{E}| \leq 2\ell$ and all $\gamma' \in \Gamma(\omega)$ are non-winding so that the previous bounds apply. This concludes the proof. $\qquad\square$

## 4. Proof of Dimerization

*4.1. Setting of the cluster expansion.* We summarize the main results of the method of cluster expansion as we need it. The following setting and theorem was proposed in [28], extending the results of [13] to the continuous setting and general repulsive interactions.

Let $\Gamma$ be a measurable space, $\eta$ a complex measure on $\Gamma$ such that $|\eta|(\Gamma) < \infty$, where $|\eta|$ is the total variation (absolute value) of $\eta$. Let $\zeta$ be a symmetric function $\Gamma \times \Gamma \to \mathbb{C}$ such that $|1 + \zeta(\gamma, \gamma')| \leq 1$ for all $\gamma, \gamma' \in \Gamma$. Define the partition function $Z$ by

$$Z = \sum_{k \geq 0} \frac{1}{k!} \int d\eta(\gamma_1) \dots \int d\eta(\gamma_k) \prod_{1 \leq i < j \leq k} \left(1 + \zeta(\gamma_i, \gamma_j)\right). \tag{4.1}$$
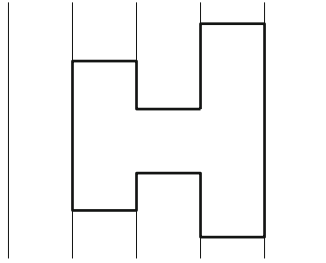
**Fig. 15.** A long, non-winding loop makes at least 6 turns

Finally, define the cluster function

$$\varphi(\gamma_1, \ldots, \gamma_k) = \begin{cases} 1 & \text{if } k = 1, \\ \frac{1}{k!} \sum_G \prod_{\{i,j\} \in G} \zeta(\gamma_i, \gamma_j) & \text{otherwise,} \end{cases} \tag{4.2}$$

where the sum is over *connected graphs* of $k$ elements, and the product is over the edges of $G$. Then we have the following expressions and estimates.

**Theorem 4.1** *Assume that there exist functions $a, b : \Gamma \to [0, \infty)$ such that for all $\gamma \in \Gamma$, we have the following Kotecký-Preiss criterion*

$$\int d|\eta|(\gamma') |\zeta(\gamma, \gamma')| e^{a(\gamma') + b(\gamma')} \leq a(\gamma). \tag{4.3}$$

*(Also, assume that $\int d|\eta|(\gamma) e^{a(\gamma) + b(\gamma)} < \infty$.) Then we have the following.*

*(a) The partition function is equal to*

$$Z = \exp\left\{ \sum_{k \geq 1} \int d\eta(\gamma_1) \ldots \int d\eta(\gamma_k) \, \varphi(\gamma_1, \ldots, \gamma_k) \right\},$$

*where the combined sum and integral converges absolutely.*
*(b) For all $\gamma_1 \in \Gamma$,*

$$1 + \sum_{k \geq 2} k \int d|\eta|(\gamma_2) \ldots \int d|\eta|(\gamma_k) \left( \sum_{i=1}^{k} |\zeta(\gamma, \gamma_i)| \right) |\varphi(\gamma_1, \ldots, \gamma_k)| e^{b(\gamma_1) + \cdots + b(\gamma_k)}$$
$$\leq e^{a(\gamma_1)}.$$

*(c) For all $\gamma \in \Gamma$,*

$$\sum_{k \geq 1} \int d|\eta|(\gamma_1) \ldots \int d|\eta|(\gamma_k) \left( \sum_{i=1}^{k} |\zeta(\gamma, \gamma_i)| \right) |\varphi(\gamma_1, \ldots, \gamma_k)| e^{b(\gamma_1) + \cdots + b(\gamma_k)}$$
$$\leq a(\gamma).$$

This theorem can be found in [28], see Theorems 1 and 3 there, as well as Eqs (18) and (19). Notice that the term $b(\gamma)$ is not usually part of the Kotecký–Preiss criterion and is not needed for convergence of the cluster expansion. But it gives better estimates, see (b) and (c) above, which are most helpful in proving exponential decay.

*4.2. Cluster expansion for the partition function.* Let us return to our loop model. We start with the partition function (2.3), namely

$$Z_{\ell,\beta,n,u} = e^{-(1+u)|\overline{E}_{\ell,\beta}|} \int_{\Omega_{\ell,\beta}} d\bar{\rho}_u(\omega) n^{\mathcal{L}(\omega)-\#\omega_{\leftrightarrows}} \tag{4.4}$$

where $d\bar{\rho}_u(\omega) = u^{\#\omega_{\times}} d^{\otimes\#\omega} x$ is given in (2.1). Since we identify contours $\gamma \in X_{\ell,\beta}$ with the links they are made up of, $d\bar{\rho}_u(\gamma)$ is also well defined. We define

$$\tilde{w}(\gamma) := e^{-(1+u)\frac{1}{2}|\gamma|} n^{h(\gamma)} u^{\#\gamma_{\times}}, \tag{4.5}$$

where $h(\gamma)$ is defined in (3.8). Let $\mathcal{L}(\Gamma) = \{l : \exists \gamma \in \Gamma : l \in \gamma\}$ be the set of loops in a (not necessarily admissible) collection of contours $\Gamma$. Let $Y_{\ell,\beta} \subseteq X_{\ell,\beta}$ the set of contours $\gamma$ (not necessarily admissible) consisting of two adjacent winding loops not traversing any links and let $\mathcal{Y}_{\ell,\beta} := \{\Gamma \subseteq Y_{\ell,\beta} : \gamma \cap \gamma' = \emptyset, \ \forall \gamma \neq \gamma' \in \Gamma\}$. Now let

$$w(\gamma) := \sum_{\tilde{\gamma} \in g(\gamma)} \tilde{w}(\tilde{\gamma})(-e^{-2\beta})^{\frac{\#\gamma\backslash\tilde{\gamma}}{2}}, \tag{4.6}$$

where

$$g(\gamma) = \{\tilde{\gamma} \subseteq \gamma \mid \exists \Gamma' \in \mathcal{Y}_{\ell,\beta} : \gamma = \tilde{\gamma} \cup \mathcal{L}(\Gamma')\} \tag{4.7}$$

is the set of contours $\tilde{\gamma}$ such that $\gamma$ can be obtained by adding pairs of adjacent, winding loops not traversing any links (those that come from having an "empty good column") to $\tilde{\gamma}$ and $\#\gamma \setminus \tilde{\gamma}$ denotes the number of loops that are in $\gamma$, but not in $\tilde{\gamma}$ – necessarily an even number by the definition of $\mathcal{Y}_{\ell,\beta}$.

Note that for $\gamma \in X_{\ell,\beta}^{\text{nw}}$ we have $g(\gamma) = \{\gamma\}$, so $w(\gamma) = \tilde{w}(\gamma)$.

**Proposition 4.2** *We have, for any $u \in \mathbb{R}$,*

$$Z_{\ell,\beta,n,u} = e^{-(1+u)|\overline{E}_\ell|} \sum_{k\geq 0} \frac{1}{k!} \int_{X_\ell^+} d\bar{\rho}_1(\gamma_1) \dots \int_{X_\ell^+} d\bar{\rho}_1(\gamma_k) \left(\prod_{i=1}^k w(\gamma_i)\right) \prod_{1\leq i<j\leq k} \delta(\gamma_i, \gamma_j).$$

*Proof.* First note that $d\bar{\rho}_u$ factorises, i.e. for $\omega_1, \omega_2 \in \Omega_{\ell,\beta}$ sharing no links, we have $d\bar{\rho}_u(\omega_1 \cup \omega_2) = d\bar{\rho}_u(\omega_1)d\bar{\rho}_u(\omega_2)$. In particular, for any admissible set $\{\gamma^1, \dots, \gamma^k\} \in \mathfrak{A}_{\ell,\beta}$ of contours,

$$d\bar{\rho}_u(\gamma^1 \cup \dots \cup \gamma^k) = \prod_{i=1}^k d\bar{\rho}_u(\gamma^i). \tag{4.8}$$

Now let $\Gamma_0 \in \mathfrak{C}_{\ell,\beta}^+$ denote a fixed set of *compatible positive-type* contours and let

$$\mathcal{A}(\Gamma_0) = \{\omega \in \Omega_{\ell,\beta} : \Gamma(\omega) = \Sigma(\Gamma_0)\} \tag{4.9}$$

denote the set of link-configurations $\omega$ that induce the set of contours $\Gamma_0$ without adding any new contours. By considering the admissible set $\Gamma(\omega)$ and its shift $\Gamma_0 = \Sigma^{-1}(\Gamma(\omega)) = \{\gamma^1, \dots, \gamma^k\}$, we conclude that

$$Z_{\ell,\beta,n,u} = e^{-(1+u)|\overline{E}_\ell|} \sum_{k\geq 0} \frac{1}{k!} \int_{X_{\ell,\beta}^+} d\bar{\rho}_1(\gamma^1) u^{\#\gamma_{\times}^1} \dots \int_{X_{\ell,\beta}^+} d\bar{\rho}_1(\gamma^k) u^{\#\gamma_{\times}^k}$$

$$\prod_{1\leq i<j\leq k} \delta(\gamma^i, \gamma^j) \int_{\mathcal{A}(\Gamma_0)} d\bar{\rho}_u(\omega \setminus \Sigma(\Gamma_0)) n^{\mathcal{L}(\omega)-\#\omega_{\leftrightarrows}}. \tag{4.10}$$

We also used Remark 3.2, which tells us that crosses are never 'shared' between distinct contours or between a contour and a short loop. (Note that in the last integral, we have the measure $\bar{\rho}_u$ rather than $\bar{\rho}_1$ as in the other integrals.)

Next, applying Lemma 3.10 to write $\mathcal{L}(\omega) - \#\omega_{\leftrightarrows} = \sum_{i=1}^{k} h(\gamma^i)$, we obtain

$$\int_{\mathcal{A}(\Gamma_0)} \mathrm{d}\bar{\rho}_u\big(\omega \setminus \Sigma(\Gamma_0)\big) n^{\mathcal{L}(\omega) - \#\omega_{\leftrightarrows}} = \prod_{i=1}^{k} n^{h(\gamma^i)} \int_{\mathcal{A}(\Gamma_0)} \mathrm{d}\bar{\rho}_u\big(\omega \setminus \Sigma(\Gamma_0)\big). \quad (4.11)$$

For $\Gamma \in \mathfrak{C}_{\ell,\beta}^+$ denote by $\mathfrak{W}(\Gamma)$ the set of columns of $\bar{E}_{\ell,\beta}$ where adding a double bar at any height would not change the set of contours $\Sigma(\Gamma)$. Using Lemma 3.9 and recalling that $|T_\beta| = 2\beta$, we get

$$\int_{\mathcal{A}(\Gamma_0)} \mathrm{d}\bar{\rho}_u\big(\omega \setminus \Sigma(\Gamma_0)\big) = \mathrm{e}^{(1+u)|F(\Gamma(\omega))|} \big( \mathrm{e}^{-|T_\beta|} \,(\mathrm{e}^{|T_\beta|} - 1)\big)^{|\mathfrak{W}(\Gamma_0)|}$$

$$= \mathrm{e}^{(1+u)|\bar{E}_\ell^+|} \Big( \prod_{\gamma \in \Gamma_0} \mathrm{e}^{-(1+u)\frac{1}{2}|\gamma|} \Big)(1 - \mathrm{e}^{-2\beta})^{|\mathfrak{W}(\Gamma_0)|}. \quad (4.12)$$

Denote by $\gamma_w(\Gamma)$ the unique winding contour in $\Gamma$, if it exists, and $\emptyset$ otherwise. For all $\Gamma \in \mathfrak{C}_{\ell,\beta}^+$ we have

$$(1 - \mathrm{e}^{-2\beta})^{|\mathfrak{W}(\Gamma)|} = \sum_{W \subseteq \mathfrak{W}(\Gamma)} \prod_{w \in W} (-\mathrm{e}^{-2\beta})$$

$$= \sum_{\Gamma' \in \mathcal{Y}_{\ell,\beta}} \prod_{\gamma' \in \Gamma'} (-\mathrm{e}^{-2\beta}) \times \Big( \mathbb{1}\{\gamma_w' \in \mathcal{A}_{\ell,\beta}\} \prod_{\gamma \in \Gamma \setminus \{\gamma_w(\Gamma)\}} \delta(\gamma, \gamma_w') \Big)$$

$$= \int_{\mathfrak{X}_{\ell,\beta}} \mathrm{d}\bar{\rho}_1(\Gamma') \Big[ \prod_{\gamma' \in \Gamma'} (-\mathrm{e}^{-2\beta} \, \mathbb{1}\{\gamma' \in Y_{\ell,\beta}\}) \Big]$$

$$\times \mathbb{1}\{\gamma_w' \in \mathcal{A}_{\ell,\beta}\} \prod_{\gamma \in \Gamma \setminus \{\gamma_w(\Gamma)\}} \delta(\gamma, \gamma_w') \prod_{\gamma, \gamma' \in \Gamma'} \mathbb{1}\{\gamma \cap \gamma' = \emptyset\},$$

$$(4.13)$$

where $\gamma_w' \equiv \gamma_w'(\Gamma, \Gamma') := \gamma_w(\Gamma) \cup \mathcal{L}(\Gamma')$. For the last equation we used that $\bar{\rho}_1$ is just the counting measure on subsets of $Y_{\ell,\beta}$ since these loops do not traverse any links. Intuitively this amounts to summing over "admissible extensions" $\Gamma'$ of some given set of contours $\Gamma$ and assigning a different weight to these extensions. But instead of integrating over one set of contours and then another set of contours that are treated differently, we might also integrate over one set of contours and then decide which weight to give to each part of the contour. More rigorously, we combine Eqs (4.10), (4.11), (4.12) and (4.13) to get

$$Z_{\ell,\beta,n,u} \, \mathrm{e}^{(1+u)|\bar{E}_\ell|}$$

$$= \int_{\mathfrak{C}_{\ell,\beta}^+} \mathrm{d}\bar{\rho}_1(\Gamma) \Big( \prod_{\gamma \in \Gamma} \tilde{w}(\gamma) \Big)(1 - \mathrm{e}^{-2\beta})^{|\mathfrak{W}(\Gamma)|}$$

$$= \int_{\mathfrak{C}_{\ell,\beta}^+} \mathrm{d}\bar{\rho}_1(\Gamma) \Big( \prod_{\gamma \in \Gamma} \tilde{w}(\gamma) \Big) \int_{\mathfrak{X}_{\ell,\beta}} \mathrm{d}\bar{\rho}_1(\Gamma') \Big[ \prod_{\gamma' \in \Gamma'} (-\mathrm{e}^{-2\beta} \, \mathbb{1}\{\gamma' \in Y_{\ell,\beta}\}) \Big]$$

$$\times \, \mathbb{I}\{\gamma'_w \in \mathcal{A}_{\ell,\beta}\} \prod_{\gamma \in \Gamma \setminus \{\gamma_w(\Gamma)\}} \delta(\gamma, \gamma'_w) \prod_{\gamma, \gamma' \in \Gamma'} \mathbb{I}\{\gamma \cap \gamma' = \emptyset\}$$

$$= \int_{\mathfrak{C}^+_{\ell,\beta}} d\bar\rho_1(\Gamma) \prod_{\gamma \in \Gamma} \sum_{\tilde\gamma \in g(\gamma)} \tilde{w}(\tilde\gamma) (-e^{-2\beta})^{\frac{\#\gamma \setminus \tilde\gamma}{2}}. \tag{4.14}$$

$\square$

In what follows we estimate integrals over contours $\gamma$ which intersect a given point or interval. Winding and non-winding contours are treated separately; we will actually see that winding contours play a very limited role for $\beta$ large. We write $X^{\mathrm{w}}_{\ell,\beta} \subseteq X_{\ell,\beta}$ and $X^{\mathrm{nw}}_{\ell,\beta} \subseteq X_{\ell,\beta}$ for the sets of winding and non-winding contours, respectively. We also write $X^{\mathrm{w}}_{\ell,\beta}(k) \subseteq X_{\ell,\beta}$ for the set of winding contours which traverse exactly $k$ links, and we write $X^{\mathrm{nw}}_{\ell,\beta}(\bar{v}, k) \subseteq X_{\ell,\beta}$ for the set of non-winding contours $\gamma$, which traverse $k$ links and which visit the point $\bar{v} \in \overline{V}_\ell$. Similarly, if $I = [(v, s), (v, t)] \subseteq \overline{V}_{\ell,\beta}$ is an interval we write $X^{\mathrm{nw}}_{\ell,\beta}(I, k)$ for the set of non-winding contours $\gamma$, which traverse $k$ links and which intersect $I$.

**Lemma 4.3.** *Fix any $\ell \in \mathbb{N}, \beta > 0 \, c > 0$ and points $(v, s), (v, t) \in \overline{V}_\ell$ with $s < t$. Write $I = [(v, s), (v, t)]$. Then we have*

$$\int_{X^{\mathrm{nw}}_{\ell,\beta}(I,k)} d\bar\rho_1(\gamma) \, e^{-c|\gamma|} \leq 8^{k-1} c^{-k} (1 + c|I|), \tag{4.15}$$

$$\int_{X^{\mathrm{w}}_{\ell,\beta}(k)} d\bar\rho_1(\gamma) \, e^{-c|\gamma|} \leq (2\ell + 1)^{2\ell+2} (k+1)^{2\ell} 8^{k-2} c^{-k}. \tag{4.16}$$

*Proof.* Let us start with the case of non-winding contours and the case when $I = \{\bar{v}\}$ contains only one point. Elements $\gamma \in X^{\mathrm{nw}}_{\ell,\beta}(\bar{v}, k)$ may be encoded using tuples $(t_1, \ldots, t_k, l_1, \ldots, l_{k-1}) \in \mathbb{R}^k_+ \times (\{\times, \rightleftharpoons\} \times \{\mathrm{L}, \mathrm{R}\} \times \{1, 2\})^{k-1}$, as follows.

- Consider a walker started at $\bar{v}$ and travelling upwards until it first encounters the endpoint of a link; store the vertical distance traversed as $t_1$.
- This link can go to the left, L, or to the right R; it can be a double bar $\rightleftharpoons$ or a cross $\times$; and it can be traversed by loops in $\gamma$ once, 1, or twice, 2. Store this information as $l_1$.
- Having crossed the link, our walker follows $\gamma$ and records vertical distances until *previously unexplored* links as $t_i$ and information about those links as $l_i$, as before.
- If a loop is closed and there are still links that are traversed twice by loops in $\gamma$, but have only been visited once by our walker, the walker continues walking from such a link and recording $t_i$ and $l_i$ as before (we fix some arbitrary rule for selecting the link and the direction of travel).
- This procedure is iterated until the entire contour has been traversed.

See Fig. 16 for an illustration.

Noting that $t_1 + \cdots + t_k \leq |\gamma|$ and that the number of options for $\{l_i\}^{k-1}_{i=1}$ is bounded by $8^{k-1}$, we get

$$\int_{X^{\mathrm{nw}}_{\ell,\beta}(\bar{v},k)} d\bar\rho_1(\gamma) \, e^{-c|\gamma|} \leq 8^{k-1} \int_0^\infty dt_1 \ldots \int_0^\infty dt_k \, e^{-c(t_1 + \cdots + t_k)} = 8^{k-1} c^{-k}. \tag{4.17}$$

Next, we may apply a similar argument to obtain that, for $\varepsilon > 0$ small enough,
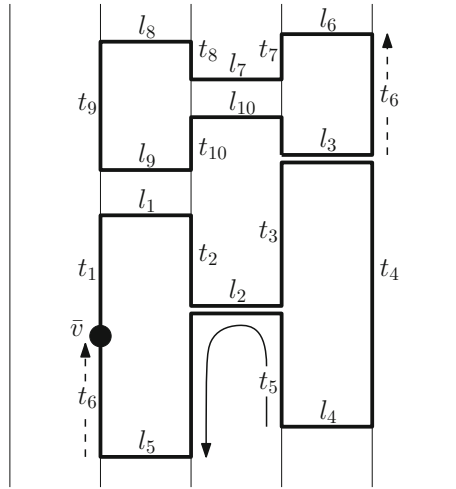
**Fig. 16.** Illustration of $t_1, \ldots, t_k, l_1, \ldots, l_{k-1}$. Note, for example, that $t_5$ is not merely the distance between the fourth and fifth links, and that $t_6$ does not readily admit an interpretation as distance between links at all

$$\int_{X_{\ell,\beta}^{\mathrm{nw}}((v,t),k)\setminus X_{\ell,\beta}^{\mathrm{nw}}((v,t+\varepsilon),k)} \mathrm{d}\bar{\rho}_1(\gamma)\, \mathrm{e}^{-c|\gamma|} \leq 8^{k-1}c^{-k+1}\frac{1-\mathrm{e}^{-\varepsilon c}}{c} \leq 8^{k-1}c^{-k+1}\varepsilon. \quad (4.18)$$

Indeed, for a contour $\gamma$ which visits $(v,t)$ but not $(v,t+\varepsilon)$, we must have $t_1 \leq \varepsilon$ in the encoding above, and replacing the integral over $t_1 \in [0,\infty)$ with an integral over $t_1 \in [0,\varepsilon]$ gives the claim. Next, to deduce (4.15) from (4.17) and (4.18), we argue as follows. If $\gamma$ visits $I = [(v,s),(v,t)]$, then either $\gamma$ contains the endpoint $(v,t)$, or there are $r \in (s,t)$ and $\varepsilon > 0$ such that $(v,r) \in \gamma$ but $(v,r+\varepsilon) \notin \gamma$. Using (4.17), the first possibility accounts for the first term $8^{k-1}c^{-k}$ in (4.15). The other possibility accounts for the second term, which one may, for example, see by using a fine dyadic discretization of the interval $I$ and passing to the limit using (4.18) and the monotone convergence theorem.

For winding contours $\gamma$, recall that they consist of $r_1 \leq 2\ell + 1$ winding loops with a finite number of contractible long loops attached to at least one of them. In particular there are at most $r \leq r_1 \leq 2\ell + 1$ winding loops that do not share a link and are not connected via a sequence of long, but contractible loops. Let us denote these by $\gamma_1, \ldots, \gamma_r$, and the numbers of links they each visit by $k_1, \ldots, k_r$, respectively, where $k = \sum_{i=1}^r k_i$. There are $r$ vertices $v_1, \ldots, v_r \in V_\ell$ such that $\gamma_i$ visits $(v_i, 0)$. Summing over the possibilities for $r, v_1, \ldots, v_r$, as well as $k_1, \ldots, k_r$, and applying the argument for (4.17) to each $\gamma_i$, we obtain

$$\int_{X_{\ell,\beta}^{\mathrm{w}}(k)} \mathrm{d}\bar{\rho}_1(\gamma)\, \mathrm{e}^{-c|\gamma|} \leq \sum_{r=1}^{2\ell+1} \binom{2\ell+1}{r} \sum_{\substack{k_1,\ldots,k_r \geq 0 \\ k_1+\cdots+k_r = k}} 8^{k-2}c^{-k}$$

$$\leq (2\ell+1)(2\ell+1)^{2\ell+1} \max_{1\leq r\leq 2\ell+1} \left|\left\{(k_1,\ldots,k_r) \in \mathbb{N}^r : \sum_i k_i = k\right\}\right| 8^{k-2}c^{-k}$$

$$\leq (2\ell+1)^{2\ell+2}(k+1)^{2\ell}8^{k-2}c^{-k}. \quad (4.19)$$

$\square$

In order to ensure the convergence of the cluster expansion, we need to check that interactions between contours are small so as to satisfy the Kotecký-Preiss criterion in Eq. (4.3). For $a_1, a_2, b_1, b_2 \geq 0$, let us introduce

$$a(\gamma) = a_1|\gamma| + a_2 \#\gamma, \qquad b(\gamma) = b_1|\gamma| + b_2 \#\gamma, \tag{4.20}$$

where $\#\gamma$ denotes the number of links visited by $\gamma$. Then we have the following bound.

**Lemma 4.4** (Kotecký–Preiss criterion). *Let $w(\gamma)$ be as in (4.6). Then there exist $n_0$, $u_0$, $a_1, a_2, b_1, b_2 > 0$ (independent of $\ell, n, u$), and $\beta_0(\ell, n)$, such that for $n > n_0$, $|u| < u_0$, and $\beta > \beta_0(\ell, n)$, we have for any $\ell$ and any $\gamma_0 \in X_{\ell,\beta}^+$ that*

$$\int_{X_{\ell,\beta}^+} d\bar{\rho}_1(\gamma)|w(\gamma)| e^{a(\gamma)+b(\gamma)} \left(1 - \delta(\gamma, \gamma_0)\right) \leq a(\gamma_0). \tag{4.21}$$

*Proof.* Let us alleviate the notation by introducing

$$\bar{w}(\gamma) := |w(\gamma)| e^{a(\gamma)+b(\gamma)}. \tag{4.22}$$

We use Lemma 3.13 with $\kappa$ such that $\frac{1}{3} - \kappa = \frac{1}{4}$ and with $u_0$ such that $u_0^{-\kappa/2} = n_0$, and we set

$$c_1 = \frac{1-u_0}{2} - a_1 - b_1, \qquad c_2 = \frac{1}{4} \log n_0 - a_2 - b_2. \tag{4.23}$$

Clearly,

$$\int_{X_{\ell,\beta}^+} d\bar{\rho}_1(\gamma)\bar{w}(\gamma)(1 - \delta(\gamma, \gamma_0)) \leq \int_{X_{\ell,\beta}^w} d\bar{\rho}_1(\gamma)\bar{w}(\gamma)$$
$$+ \int_{X_{\ell,\beta}^{nw}} d\bar{\rho}_1(\gamma)\bar{w}(\gamma)(1 - \delta(\gamma, \gamma_0)). \tag{4.24}$$

Let us first consider the contribution of winding contours. Notice that

$$\mathbb{I}\{\gamma \text{ has a spanning segment}\} \leq |\gamma|/(2\beta). \tag{4.25}$$

For $\gamma \in X_{\ell,\beta}^w$, note that $|w(\gamma)| \leq |g(\gamma)|\tilde{w}_{u=-u_0}(\gamma)$ where $g(\gamma)$ is given in (4.7). Here $|g(\gamma)| \leq 2^{|E_\ell|}$, which is some constant depending only on $\ell$. Hence, also for $\gamma \in X_{\ell,\beta}^w$ we get $\bar{w}(\gamma) \leq e^{-c_1|\gamma|} e^{-c_2\#\gamma}$ for $\beta \equiv \beta(\ell, n)$ large enough.

Using Lemma 4.3 and the fact that a winding contour $\gamma$ satisfies $|\gamma| \geq \frac{1}{2}|\gamma| + \beta$, the first term on the right in (4.24) satisfies

$$\int_{X_{\ell,\beta}^w} d\bar{\rho}_1(\gamma)\bar{w}(\gamma) \leq e^{-c_1\beta} \sum_{k\geq 0} e^{-c_2 k} \int_{X_{\ell,\beta}^w(k)} d\bar{\rho}_1(\gamma) e^{-\frac{c_1}{2}|\gamma|}$$
$$\leq (2\ell+1)^{2\ell+2} e^{-c_1\beta} \sum_{k\geq 0}(k+1)^{2\ell}\left(\frac{16}{c_1} e^{-c_2}\right)^k \leq c(\ell, n) e^{-c'\beta}, \tag{4.26}$$

with some absolute constant $c' > \frac{1}{4}$ and $c(\ell, n) < \infty$ for $n_0$ sufficiently large such that the geometric series converges. In particular (4.26) gets arbitrarily small for $\beta$, $n_0$ large enough.

We now turn to non-winding contours. We have

$$
\begin{aligned}
\bar{w}(\gamma) &= e^{-\frac{1+u}{2}|\gamma|+(a_1+b_1)|\gamma|}\, n^{h(\gamma)}|u|^{\#\gamma}\!\rtimes e^{(a_2+b_2)\#\gamma} \\
&\leq e^{-(\frac{1-u_0}{2}-a_1-b_1)|\gamma|}\, n_0^{-\frac{1}{4}\#\gamma}\, e^{(a_2+b_2)\#\gamma} \\
&= e^{-c_1|\gamma|}\, e^{-c_2\#\gamma}\,.
\end{aligned}
\tag{4.27}
$$

Note that if $\delta(\gamma,\gamma_0)=0$ then $\gamma$ and $\gamma_0$ intersect somewhere on $\overline{V}_{\ell,\beta}$. We may decompose the subset of $\overline{V}_{\ell,\beta}$ visited by $\gamma_0$ as a union of closed intervals $I_1,\ldots,I_m$ where $m\leq\#\gamma_0$ is the number of links of $\gamma_0$. Noting also that a non-winding contour $\gamma$ has at least 5 links, we obtain from Lemma 4.3 that

$$
\begin{aligned}
\int_{X_{\ell,\beta}^{\mathrm{nw}}} \mathrm{d}\bar{\rho}_1(\gamma)\bar{w}(\gamma)(1-\delta(\gamma,\gamma_0)) &\leq \sum_{j=1}^{m}\sum_{k\geq 5} e^{-c_2 k}\int_{X_{\ell,\beta}^{\mathrm{nw}}(I_j,k)} \mathrm{d}\bar{\rho}_1(\gamma)\, e^{-c_1|\gamma|} \\
&\leq \sum_{j=1}^{m}\sum_{k\geq 5} e^{-c_2 k}\, 8^{k-1} c_1^{-k}\big(1+c_1|I_j|\big) \\
&\leq \Big(\tfrac{1}{8}\#\gamma_0 + \tfrac{c_1}{8}|\gamma_0|\Big)\sum_{k\geq 5}\Big(\tfrac{8}{c_1}\, e^{-c_2}\Big)^k.
\end{aligned}
\tag{4.28}
$$

The Lemma 4.4 holds true provided that $\beta$ is large enough, and

$$
\frac{1}{8}\sum_{k\geq 5}\Big(\tfrac{8}{c_1}\, e^{-c_2}\Big)^k \leq a_2,\qquad \frac{c_1}{8}\sum_{k\geq 5}\Big(\tfrac{8}{c_1}\, e^{-c_2}\Big)^k \leq a_1.
\tag{4.29}
$$

Both conditions are fulfilled for $n_0$ (and therefore $c_2$) large enough. $\qquad\square$

We will also need an estimate on the integral of contours that contain or surround a given point.

**Corollary 4.5** *For any $\varepsilon>0$, there exists $n_0,u_0,a_1,a_2,b_1,b_2>0$ (independent of $\ell,n,u$) such that for $n>n_0$, $|u|<u_0$, $\beta>\beta_0(\ell,n)$, we have*

$$
\int_{X_{\ell,\beta}^{+}} \mathrm{d}\bar{\rho}_1(\gamma)|w(\gamma)|\, e^{a(\gamma)+b(\gamma)}\, \mathbb{1}\{(0,0)\in\overline{I(\gamma)}\} \leq \varepsilon.
\tag{4.30}
$$

*Proof.* We proceed as in Eq. (4.24), so that it suffices to bound the contribution from winding and non-winding contours separately. Recall the definition of $\bar{w}$ in (4.22). Using Eq. (4.26), we can make the contribution from winding contours arbitrarily small, say $\varepsilon/2$, by choosing $\beta\equiv\beta(\ell,n)$ sufficiently large, i.e. we have

$$
\int_{X_{\ell,\beta}^{+}} \mathrm{d}\bar{\rho}_1(\gamma)\bar{w}(\gamma)\mathbb{1}\{(0,0)\in\overline{I(\gamma)}\} \leq \int_{X_{\ell,\beta}^{\mathrm{nw}}} \mathrm{d}\bar{\rho}_1(\gamma)\bar{w}(\gamma)\mathbb{1}\{(0,0)\in\overline{I(\gamma)}\} + \frac{\varepsilon}{2}.
\tag{4.31}
$$

If $\gamma$ is non-winding and has $k$ links, then it must pass by a site at time 0 at distance less than $k/2$ from $(0,0)$. Thus we have the bound

$$
\int_{X_{\ell,\beta}^{\mathrm{nw}}} \mathrm{d}\bar{\rho}_1(\gamma)\bar{w}(\gamma)\mathbb{1}\{(0,0)\in\overline{I(\gamma)}\} \leq \int_{X_{\ell,\beta}^{\mathrm{nw}}} \mathrm{d}\bar{\rho}_1(\gamma)\bar{w}(\gamma)\#\gamma\,\mathbb{1}\{(0,0)\in\gamma)\}.
\tag{4.32}
$$

This can be shown to be arbitrarily small, say less than $\varepsilon/2$, when $n$ and $\beta$ are large, as in the previous lemma. $\qquad\square$

Let $\mathcal{C}_k$ denote the set of connected (undirected) graphs with vertex set $\{1, \ldots, k\}$ and define

$$\varphi(\gamma_1, \ldots, \gamma_k) = \begin{cases} 1, & \text{if } k = 1, \\ \frac{1}{k!} \sum_{G \in \mathcal{C}_k} \prod_{ij \in G} (\delta(\gamma_i, \gamma_j) - 1), & \text{if } k \geq 2, \end{cases} \tag{4.33}$$

where the product in the second line is over the edges of $G$. The following is the main consequence of Theorem 4.1, which holds because of Lemma 4.4.

**Proposition 4.6.** *(Cluster expansion of the partition function) For parameters as in Lemma 4.4, the following sum converges absolutely:*

$$\Phi_{\ell,\beta} := \sum_{m \geq 1} \int_{X_{\ell,\beta}^+} \mathrm{d}\bar{\rho}_1(\gamma_1) \ldots \int_{X_{\ell,\beta}^+} \mathrm{d}\bar{\rho}_1(\gamma_k) \left( \prod_{i=1}^{k} w(\gamma_i) \right) \varphi(\gamma_1, \ldots, \gamma_m), \tag{4.34}$$

*and we have that*

$$\mathrm{e}^{(1+u)|\overline{E}_{\ell,\beta}^-|} Z_{\ell,\beta,n,u} = \exp\left(\Phi_{\ell,\beta}\right). \tag{4.35}$$

Notice that $\Phi_{\ell,\beta}$ depends on $n$ and $u$ as well.

*4.3. Dimerization.* Let us introduce the (signed) measure $\mu_{\ell,\beta,n,u}$ such that the integral of a function $f : \Omega_{\ell,\beta} \to \mathbb{C}$ is given by

$$\mu_{\ell,\beta,n,u}(f) = \frac{1}{Z_{\ell,\beta,n,u}} \int_{\Omega_{\ell,\beta}} \mathrm{d}\rho_u(\omega) \, n^{\mathcal{L}(\omega) - \#\omega_{\rightleftharpoons}} f(\omega). \tag{4.36}$$

The next theorem can be understood as dimerization in the loop model. Together with Theorem 2.1 it implies Theorem 1.1.

**Theorem 4.7** *For any $c > 0$, there exist $n_0, u_0 > 0$ such that for all $n > n_0$ and $|u| < u_0$, we have*

*(a) For all $\ell$ even:*

$$\liminf_{\beta \to \infty} \mu_{\ell,\beta,n,u}(0 \overset{-}{\longleftrightarrow} -1) > 1 - c, \quad \text{and} \quad \limsup_{\beta \to \infty} |\mu_{\ell,\beta,n,u}(0 \overset{+}{\longleftrightarrow} -1)| < c;$$

$$\limsup_{\beta \to \infty} |\mu_{\ell,\beta,n,u}(0 \overset{+}{\longleftrightarrow} 1)| < c, \quad \text{and} \quad \limsup_{\beta \to \infty} |\mu_{\ell,\beta,n,u}(0 \overset{-}{\longleftrightarrow} 1)| < c. \tag{4.37}$$

*(b) For all $\ell$ odd:*

$$\liminf_{\beta \to \infty} \mu_{\ell,\beta,n,u}(0 \overset{-}{\longleftrightarrow} 1) > 1 - c, \quad \text{and} \quad \limsup_{\beta \to \infty} |\mu_{\ell,\beta,n,u}(0 \overset{+}{\longleftrightarrow} 1)| < c;$$

$$\limsup_{\beta \to \infty} |\mu_{\ell,\beta,n,u}(0 \overset{+}{\longleftrightarrow} -1)| < c, \quad \text{and} \quad \limsup_{\beta \to \infty} |\mu_{\ell,\beta,n,u}(0 \overset{-}{\longleftrightarrow} -1)| < c.$$

$$\tag{4.38}$$

Notice that the limits $\beta \to \infty$ actually exist; this could be established using the correspondence with the quantum spin system, where convergence is clear. This is less visible in the loop model, though, hence the use of lim sup and lim inf so we do not need to prove it.

*Proof.* Assume without loss of generality that $\ell$ is odd. The case of $\ell$ even works similarly. Let $\mathcal{O}$ (for *outside*) denote the event that $(0, 0)$ is not on or inside any contour, that is

$$\mathcal{O} = \left\{ \omega \in \Omega_{\ell,\beta} : (0, 0) \in \bigcap_{\gamma \in \Gamma(\omega)} E(\gamma) \right\}. \tag{4.39}$$

By Remark 3.4 and the fact that $\mu_{\ell,\beta,n,u}(1) = 1$,

$$\mu_{\ell,\beta,n,u}(0 \overset{-}{\longleftrightarrow} 1) = 1 - \mu_{\ell,\beta,n,u}(\mathbb{1}\{0 \overset{-}{\longleftrightarrow} 1\}^c) = 1 - \mu_{\ell,\beta,n,u}(\mathbb{1}_{\mathcal{O}^c}\mathbb{1}\{0 \overset{-}{\longleftrightarrow} 1\}^c). \tag{4.40}$$

We also have

$$\mu_{\ell,\beta,n,u}(0 \overset{+}{\longleftrightarrow} 1) = \mu_{\ell,\beta,n,u}(\mathbb{1}_{\mathcal{O}^c}\mathbb{1}\{0 \overset{+}{\longleftrightarrow} 1\}^c),$$
$$\mu_{\ell,\beta,n,u}(0 \overset{\pm}{\longleftrightarrow} -1) = \mu_{\ell,\beta,n,u}(\mathbb{1}_{\mathcal{O}^c}\mathbb{1}\{0 \overset{\pm}{\longleftrightarrow} -1\}^c). \tag{4.41}$$

Then Theorem 4.7 follows from the next lemma (Lemma 4.8), with the function $f$ being an indicator function.                                                                                $\square$

Given a set of compatible contours $\Gamma \in \mathfrak{C}^+_{\ell,\beta}$ and its "shifted version" $\Sigma(\Gamma) \in \mathfrak{A}_{\ell,\beta}$, we identify their contours in the natural way; i.e. for every $\gamma \in \Gamma$ there exists a unique $\Sigma(\gamma; \Gamma) \in \Sigma(\Gamma)$ that is obtained by shifting $\gamma$. Every $\Gamma \in \mathfrak{C}^+_{\ell,\beta}$ can now be uniquely decomposed into $\Gamma = \Gamma_0 \dot{\cup} \Gamma \setminus \Gamma_0$ with

$$\Gamma_0 := \{\gamma \in \Gamma \mid (0, 0) \in \overline{I(\Sigma(\gamma; \Gamma))}\}, \tag{4.42}$$

where $I(\gamma)$ is the interior of $\gamma$, defined in Sect. 3. This decomposition will be useful in the proof of the following lemma.

**Lemma 4.8** *Let $g : \Omega_{\ell,\beta} \to \mathbb{R}$ be a function that, for every $\omega$, only depends on the contours $\gamma \in \Gamma(\omega) \in \mathfrak{A}_{\ell,\beta}$ that surround or contain $(0, 0)$. Assume that $|g| \leq 1$. Then for every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}, u_0 > 0$ such that for all $\ell, n > n_0, |u| < u_0$, and $\beta \equiv \beta(\ell, n)$ large enough,*

$$|\mu_{\ell,\beta,n,u}(\mathbb{1}_{\mathcal{O}^c}g)| < \varepsilon. \tag{4.43}$$

*Proof.* We have

$$\mu_{\ell,\beta,n,u}(\mathbb{1}_{\mathcal{O}^c}g) = \frac{Z_{\ell,\beta,n,u}[\mathcal{O}^c; g]}{Z_{\ell,\beta,n,u}}, \tag{4.44}$$

where

$$Z_{\ell,\beta,n,u}[\mathcal{O}^c; g]$$
$$= e^{-(1+u)|\overline{E}_{\ell,\beta}|} \int_{\mathfrak{C}^+_{\ell,\beta}\setminus\{\emptyset\}} d\bar{\rho}_1(\Gamma_0) g(\Gamma_0) \left( \prod_{\gamma \in \Gamma_0} w(\gamma) \right) \mathbb{1}\left\{ (0, 0) \in \overline{I(\Sigma(\gamma; \Gamma_0))} \; \forall \gamma \in \Gamma_0 \right\}$$
$$\times \sum_{m \geq 0} \frac{1}{m!} \int_{X^+_{\ell,\beta}} d\bar{\rho}_1(\gamma_1) \dots \int_{X^+_{\ell,\beta}} d\bar{\rho}_1(\gamma_m) \left( \prod_{i=1}^m w_{\Gamma_0}(\gamma_i) \right) \prod_{1 \leq i < j \leq m} \delta(\gamma_i, \gamma_j). \tag{4.45}$$

Notice that the contour weights in the last line depend on $\Gamma_0$ and are defined as

$$w_{\Gamma_0}(\gamma) := w(\gamma)\,\mathbb{I}\{(0,0) \notin \overline{I(\Sigma(\gamma;\Gamma_0))}\} \prod_{\gamma_0 \in \Gamma_0} \delta(\gamma,\gamma_0). \tag{4.46}$$

Intuitively, we first integrate over all contours surrounding $(0,0)$ (after shifting, they are called $\Gamma_0$) and then we integrate out the remaining contours that are compatible with $\Gamma_0$.

The second line of Eq. (4.45) has the structure of a partition function. Since $|w_{\Gamma_0}(\gamma_i)| \le |w(\gamma_i)|$, Lemma 4.4 holds for the modified weights too, and therefore also the suitable modification of Proposition 4.6. This is then equal to $\exp\big(\Phi_{\ell,\beta}(\Gamma_0)\big)$ where

$$\Phi_{\ell,\beta}(\Gamma_0) := \sum_{m \ge 1} \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_1) \ldots \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_m) \Big(\prod_{i=1}^m w_{\Gamma_0}(\gamma_i)\Big) \varphi(\gamma_1,\ldots,\gamma_m). \tag{4.47}$$

Notice that the sum in $\Phi_{\ell,\beta}(\Gamma_0)$ converges absolutely. Then

$$\mu(\mathbb{I}_{\mathcal{O}^c}g) = \int_{\mathfrak{C}_{\ell,\beta}^+ \setminus \{\emptyset\}} d\bar\rho_1(\Gamma_0) g(\Gamma_0)\Big(\prod_{\gamma \in \Gamma_0} w(\gamma)\Big)$$

$$\mathbb{I}\big\{(0,0) \in \overline{I(\Sigma(\gamma;\Gamma_0))}\ \forall \gamma \in \Gamma_0\big\} \exp\big\{\Phi_{\ell,\beta}(\Gamma_0) - \Phi_{\ell,\beta}\big\} \tag{4.48}$$

Let $\delta_0(\gamma,\Gamma_0) := \mathbb{I}\{(0,0) \notin \overline{I(\Sigma(\gamma;\Gamma_0))}\} \prod_{\gamma_0 \in \Gamma_0} \delta(\gamma,\gamma_0)$ be the indicator function for $\gamma$ being a contour that is compatible with $\Gamma_0$ and should not be part of $\Gamma_0$. Then

$$\Phi_{\ell,\beta} - \Phi_{\ell,\beta}(\Gamma_0) = \sum_{m \ge 1} \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_1) \ldots \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_m)\Big(\prod_{i=1}^m w(\gamma_i)\Big)\varphi(\gamma_1,\ldots,\gamma_m)$$

$$\mathbb{I}\big\{\exists i \le m : \delta_0(\gamma_i,\Gamma_0) = 0\big\}. \tag{4.49}$$

We bound these "corrections coming from contours not in $\Gamma_0$" as follows:

$$|\Phi_{\ell,\beta} - \Phi_{\ell,\beta}(\Gamma_0)|$$

$$\le \sum_{m \ge 1} \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_1) \ldots \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_m)$$

$$\Big(\prod_{i=1}^m |w(\gamma_i)|\Big)|\varphi(\gamma_1,\ldots,\gamma_m)|\,\mathbb{I}\{\exists i \le m : \delta_0(\gamma_i,\Gamma_0) = 0\}$$

$$\le \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_1)|w(\gamma_1)|(1 - \delta_0(\gamma_1,\Gamma_0))\Big(1 + \sum_{m \ge 2} m \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_2) \ldots$$

$$\int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_m)|\varphi(\gamma_1,\ldots,\gamma_m)|\Big)$$

$$\le \int_{X_{\ell,\beta}^+} d\bar\rho_1(\gamma_1)|w(\gamma_1)|\,e^{a(\gamma_1)}\,(1 - \delta_0(\gamma_1,\Gamma_0)). \tag{4.50}$$

The last inequality follows from Theorem 4.1 (b). It is easy to see from the definition of $\delta_0$ that $1 - \delta_0(\gamma_1, \Gamma_0) \leq \sum_{\gamma_0 \in \Gamma_0} (1 - \delta(\gamma_0, \gamma_1)) + \mathbb{I}\{(0, 0) \in \overline{I(\Sigma(\gamma_1; \Gamma_0))}\}$. Thus

$$|\Phi_{\ell,\beta} - \Phi_{\ell,\beta}(\Gamma_0)| \leq \sum_{\gamma_0 \in \Gamma_0} a(\gamma_0) + \int_{X_{\ell,\beta}^+} d\bar{\rho}_1(\gamma_1) |w(\gamma_1)| e^{a(\gamma_1)} \mathbb{I}\{(0, 0) \in \overline{I(\gamma_1)}\}.$$

$$(4.51)$$

We used Lemma 4.4 for the first summand and translation invariance for the second summand. Using Corollary 4.5, we can bound the second summand by $\varepsilon$, arbitrarily (and uniformly in $\ell, n, u$) small, for $n_0, \beta$ large enough.

Plugging these bounds back into (4.48) and using $|g| \leq 1$, we get

$$\left|\mu_{\ell,\beta,n,u}(\mathbb{I}_{\mathcal{O}^c} g)\right| \leq e^{\varepsilon} \int_{\mathfrak{C}_{\ell,\beta}^+ \setminus \{\emptyset\}} d\bar{\rho}_1(\Gamma_0) \prod_{\gamma_0 \in \Gamma_0} \left(|w(\gamma_0)| e^{a(\gamma_0)}\right)$$

$$\mathbb{I}\{(0, 0) \in \overline{I(\Sigma(\gamma_0; \Gamma_0))} \,\forall \gamma_0 \in \Gamma_0\}$$

$$\leq e^{\varepsilon} \sum_{m \geq 1} \int_{X_{\ell,\beta}} d\bar{\rho}_1(\gamma_1) \bar{w}(\gamma_1) \mathbb{I}\{(0, 0) \in \overline{I(\gamma_1)}\} \cdots$$

$$\int_{X_{\ell,\beta}} d\bar{\rho}_1(\gamma_m) \bar{w}(\gamma_m) \mathbb{I}\{(0, 0) \in \overline{I(\gamma_m)}\}$$

$$\leq 2 e^{\varepsilon} \sum_{m \geq 1} \varepsilon^m = 2 e^{\varepsilon} \left(\frac{\varepsilon}{1 - \varepsilon}\right).$$

$$(4.52)$$

For the second inequality we use the fact that all admissible $\Sigma(\Gamma_0) \in \mathfrak{A}_{\ell,\beta}$ such that $(0, 0) \in \overline{I(\Sigma(\gamma_0; \Gamma_0))} \,\forall \gamma_0 \in \Gamma_0$ can be written uniquely as $\{\gamma_1, \ldots, \gamma_m\} \in \mathfrak{A}_{\ell,\beta}$ with $(0, 0) \in \overline{I(\gamma_m)}$ and $\gamma_i$ surrounding $\gamma_j$ whenever $i < j$. In particular all $\gamma_i$ must surround $(0, 0)$. The last inequality follows from applying Corollary 4.5 for each of the nested integrals, with a factor of 2 since we integrate over $X_{\ell,\beta}$ instead of $X_{\ell,\beta}^+$.

Since $\varepsilon$ can be made arbitrarily small by taking $n_0, \beta$ large, we get the lemma. $\quad\square$

### 4.4. Proof of exponential decay of correlations.

We now turn to exponential decay. Theorem 1.2 is an immediate consequence of the following result about loop correlations.

**Theorem 4.9** *There exists an $n_0, u_0, C, c_1, c_2 > 0$ (independent of $\ell, n, u$) such that for $n > n_0, |u| < u_0$, we have*

$$\left.\begin{array}{r}\left|\mu_{\ell,\beta,n,u}\big((x, s) \leftrightarrow (y, t)\big)\right| \\[4pt] \left|\mu_{\ell,\beta,n,u}\big((x, s) \overset{+}{\longleftrightarrow} (y, t)\big)\right| \\[4pt] \left|\mu_{\ell,\beta,n,u}\big((x, s) \overset{-}{\longleftrightarrow} (y, t)\big)\right|\end{array}\right\} \leq C e^{-c_1|x-y| - c_2|s-t|}.$$

$$(4.53)$$

*for all $\ell \in \mathbb{N}$, all $x, y \in \{-\ell + 1, \ldots, \ell\}$, and all $s, t \in \mathbb{R}$.*

*Proof.* All three bounds can be proved in the same way; here we only discuss the first one. We closely follow the proof of Lemma 4.8 and assume $x = s = 0$ for notational convenience.

It turns out that the proof for $|x - y| \leq 1$ present uninformative technical difficulties. On the other hand, exponential decay can be easily proved in the equivalent quantum

model by expanding the trace in the basis of eigenvectors of the Hamiltonian and by using the existence of a spectral gap (which is proved in the next section).

So it is enough to consider here $|x - y| > 1$. This allows us to write $\mathbb{I}\{(x, s) \leftrightarrow (y, t)\}(\omega) = \mathbb{I}_{\mathcal{O}^c}(\omega)g(\omega)$ with a function $g$ such that $|g| \leq 1$ and that only depends on the contours $\gamma \in \Gamma(\omega) \in \mathfrak{A}_{\ell,\beta}$ that surround or contain $(x, s)$. (Recall that we assumed $(x, s)$ to be the origin $(0, 0)$ — otherwise one would simply have to redefine $\mathcal{O}$ and $\Gamma_0$ to depend on $(x, s)$.) We proceed as in Lemma 4.8 to get

$$\mu_{\ell,\beta,n,u}\big((x, s) \leftrightarrow (y, t)\big) = \frac{Z_{\ell,\beta,n,u}[\mathcal{O}^c; g]}{Z_{\ell,\beta,n,u}}, \tag{4.54}$$

where $Z_{\ell,\beta,n,u}[\mathcal{O}^c; g]$ is given as in Eq. (4.45). Proceeding exactly the same way, we get the analogue of Eq. (4.52), namely

$$e^{c_1|x-y|+c_2|s-t|}\,|\mu_{\ell,\beta,n,u}(\mathbb{I}_{\mathcal{O}^c}g)| \leq e^{c_1|x-y|+c_2|s-t|}\,e^{\varepsilon}$$

$$\int_{\mathfrak{C}_{\ell,\beta}^+\setminus\{\emptyset\}} d\bar\rho_1(\Gamma_0)g(\Gamma_0) \prod_{\gamma_0 \in \Gamma_0} \big(|w(\gamma_0)|\,e^{a(\gamma_0)}\big)$$

$$\times\, \mathbb{I}\{(x, s) \in \overline{I(\Sigma(\gamma_0; \Gamma_0))}\forall\gamma_0 \in \Gamma_0\}$$

$$\leq e^{\varepsilon} \int_{\mathfrak{C}_{\ell,\beta}^+\setminus\{\emptyset\}} d\bar\rho_1(\Gamma_0) \prod_{\gamma_0 \in \Gamma_0} \big(|w(\gamma_0)|\,e^{a(\gamma_0)+b(\gamma_0)}\big)$$

$$\mathbb{I}\{(x, s) \in \overline{I(\Sigma(\gamma_0; \Gamma_0))}\forall\gamma_0 \in \Gamma_0\}$$

$$\leq 2\,e^{\varepsilon}\left(\frac{\varepsilon}{1-\varepsilon}\right). \tag{4.55}$$

Here we chose $\varepsilon \in (0, 1)$ to be a constant, independent of all other parameters $n, \beta, \ell, u$.

For all $\gamma \in \Gamma_0$ such that $g(\Gamma_0) \neq 0$ (hence $g(\Gamma_0) = 1$) we have $|\gamma| \geq 2|s - t|$ and $\#\gamma \geq 2|x - y|$. Recall the function $b$ of Lemma 4.4. Choosing $c_1 = 2b_1, c_2 = 2b_2$, we get the second inequality. Corollary 4.5 then allows us to proceed as in Eq. (4.52), which gives the last inequality. □

## 5. Proof of the Spectral Gap

We follow the method of Kennedy and Tasaki [11] and show that the method of cluster expansion can be used to prove the existence of a positive spectral gap. Indeed, it implies the validity of the following lemma (recall that $Z_{\ell,\beta} = \text{Tr }e^{-2\beta H_\ell}$).

**Lemma 5.1.** *There exists $n_0, u_0, c > 0$ (independent of $\ell, \beta, n, u$) and $C_\ell$ (independent of $\beta, n, u$) such that for all $n \geq n_0$ and $|u| \leq u_0$, we have for all $\beta \geq \frac{1}{2}$ that*

$$\big|E_0^{(\ell)} + \tfrac{1}{2\beta}\log Z_{\ell,\beta}\big| \leq C_\ell\,e^{-\beta c}.$$

*Proof.* We check that, for all $1 \leq \beta < \beta'$, we have

$$\big|\tfrac{1}{2\beta}\Phi_{\ell\beta} - \tfrac{1}{2\beta'}\Phi_{\ell\beta'}\big| \leq C_\ell\,e^{-\beta c}. \tag{5.1}$$

Since $-\frac{1}{2\beta}\log Z_{\ell,\beta} = 2\ell(1 + u) - \frac{1}{2\beta}\log \Phi_{\ell,\beta}$, we get the lemma by taking the limit $\beta' \to \infty$.
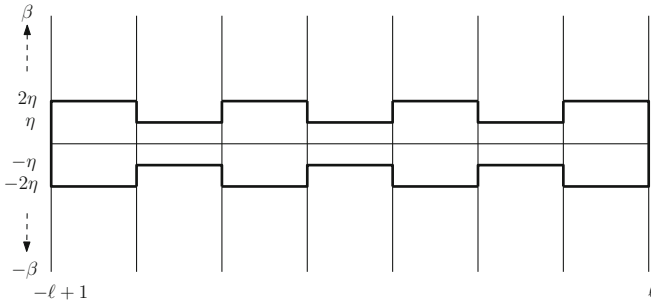
**Fig. 17.** The contour $\gamma^{(\eta)}$ used in the proof of Lemma 5.1

Let $\mathfrak{C}_\beta^+$ denote the set of clusters in $V_\ell \times T_\beta$, i.e. the sequences of contours $\Gamma = (\gamma_1, \ldots, \gamma_k)$, $\gamma_i \in X_{\ell,\beta}^+$, such that $\varphi(\Gamma) \neq 0$. For $t \in T_\beta$, let $\mathbb{1}_t(\Gamma)$ be the indicator that the cluster $\Gamma$ crosses the line $V_\ell \times \{t\}$. Let $\mathfrak{L}(\Gamma) \in [0, 2\beta]$ be the vertical length of the cluster $\Gamma$:

$$\mathfrak{L}(\Gamma) = \int_{-\beta}^{\beta} \mathbb{1}_t(\Gamma) \, dt. \tag{5.2}$$

Then, from the cluster expansion of Proposition 4.6,

$$\frac{1}{2\beta} \Phi_{\ell,\beta} = \frac{1}{2\beta} \int_{\mathfrak{C}_\beta^+} d\bar\rho_1(\Gamma) w(\Gamma) \varphi(\Gamma) = \frac{1}{2\beta} \int_{\mathfrak{C}_\beta^+} d\bar\rho_1(\Gamma) \int_{-\beta}^{\beta} dt \frac{\mathbb{1}_t(\Gamma)}{\mathfrak{L}(\Gamma)} w(\Gamma) \varphi(\Gamma)$$

$$= \frac{1}{2\beta} \int_{-\beta}^{\beta} dt \int_{\mathfrak{C}_\beta^+} d\bar\rho_1(\Gamma) \frac{w(\Gamma) \varphi(\Gamma)}{\mathfrak{L}(\Gamma)} \mathbb{1}_t(\Gamma) = \int_{\mathfrak{C}_\beta^+} d\bar\rho_1(\Gamma) \frac{w(\Gamma) \varphi(\Gamma)}{\mathfrak{L}(\Gamma)} \mathbb{1}_0(\Gamma). \tag{5.3}$$

We used Fubini's theorem to exchange the integrals, and time translation invariance in the last step. The last expression is convenient to cancel terms for different $\beta$s; for $\beta < \beta'$, we have

$$\frac{1}{2\beta} \Phi_{\ell,\beta} - \frac{1}{2\beta'} \Phi_{\ell,\beta'} = \int_{\mathfrak{C}_\beta^+} d\bar\rho_1(\Gamma) \frac{w(\Gamma) \varphi(\Gamma)}{\mathfrak{L}(\Gamma)} \mathbb{1}_0(\Gamma) - \int_{\mathfrak{C}_{\beta'}^+} d\bar\rho_1(\Gamma) \frac{w(\Gamma) \varphi(\Gamma)}{\mathfrak{L}(\Gamma)} \mathbb{1}_0(\Gamma)$$

$$= \int_{\mathfrak{C}_\beta^+} d\bar\rho_1(\Gamma) \frac{w(\Gamma) \varphi(\Gamma)}{\mathfrak{L}(\Gamma)} \mathbb{1}_0(\Gamma) \mathbb{1}_{\mathfrak{L}(\Gamma)=2\beta}$$

$$- \int_{\mathfrak{C}_{\beta'}^+} d\bar\rho_1(\Gamma) \frac{w(\Gamma) \varphi(\Gamma)}{\mathfrak{L}(\Gamma)} \mathbb{1}_0(\Gamma) 1_{\mathfrak{L}(\Gamma)\geq 2\beta}. \tag{5.4}$$

Indeed, the contribution of clusters with $\mathfrak{L}(\Gamma) < 2\beta$ has precisely canceled.

We can use estimates from cluster expansions in order to bound the terms above. For $\eta > 0$, let us introduce the contour $\gamma^{(\eta)}$ that surrounds the horizontal axis at time 0, as shown in Fig. 17. Its vertical length goes to 0 as $\eta \to 0$. We have $a(\gamma^{(\eta)}) = 2(2\ell - 1)a_1 + 4(\ell + 1)\eta a_2$; and if $\eta < 2\beta$, we have

$$\mathbb{1}_0(\Gamma) \leq 1 - \delta(\gamma^{(\eta)}, \Gamma) \leq \sum_{\gamma \in \Gamma} \left(1 - \delta(\gamma^{(\eta)}, \gamma)\right). \tag{5.5}$$

Using $1 \leq 2\beta \leq \sum_{\gamma \in \Gamma} |\gamma|$, which holds for contours such that $1_{\mathfrak{L}(\Gamma) \geq 2\beta} = 1$, we have, using Lemma 4.4 and the estimate in Theorem 4.1 (c) that

$$
\begin{aligned}
e^{2b_2\beta} &\left| \int_{\mathfrak{C}_\beta^+} d\bar{\rho}_1(\Gamma) \frac{w(\Gamma)\varphi(\Gamma)}{\mathfrak{L}(\Gamma)} \mathbb{1}_0(\Gamma) \mathbb{1}_{\mathfrak{L}(\Gamma)=2\beta} \right| \\
&\leq \int_{\mathfrak{C}_\beta^+} d\bar{\rho}_1(\Gamma) |w(\Gamma)| e^{b_2 \sum_{\gamma \in \Gamma} |\gamma|} |\varphi(\Gamma)| \sum_{\gamma \in \Gamma} \left(1 - \delta(\gamma^{(\eta)}, \gamma)\right) \\
&\leq 2(2\ell - 1)a_1 + 4(\ell + 1)\eta a_2. \qquad (5.6)
\end{aligned}
$$

The other term in the right side of (5.4) can be estimated in the same way, giving the same bound. This proves Lemma 5.1 with $C_\ell = 4(2\ell - 1)a_1$ (we can take $\eta \to 0$) and $c = 2b_2$. □

*Proof of Theorem 1.4.* With $m_i^{(\ell)}$ the multiplicity of the eigenvalue $E_i^{(\ell)}$ (satisfying $\sum_{i \geq 0} m_i^{(\ell)} = n^{2\ell}$), we have

$$
Z_{\ell,\beta} = \sum_{i \geq 0} m_i^{(\ell)} e^{-2\beta E_i^{(\ell)}} = m_0^{(\ell)} e^{-2\beta E_0^{(\ell)}} \left( 1 + \sum_{i \geq 1} \frac{m_i^{(\ell)}}{m_0^{(\ell)}} e^{-2\beta(E_i^{(\ell)} - E_0^{(\ell)})} \right). \quad (5.7)
$$

Thus

$$
-\tfrac{1}{2\beta} \log Z_{\ell,\beta} = E_0^{(\ell)} - \tfrac{1}{2\beta} \log m_0^{(\ell)} - \tfrac{1}{2\beta} R(\ell, \beta), \qquad (5.8)
$$

where

$$
R(\ell, \beta) = \log\left( 1 + \sum_{i \geq 1} \frac{m_i^{(\ell)}}{m_0^{(\ell)}} e^{-2\beta(E_i^{(\ell)} - E_0^{(\ell)})} \right) \geq \log\left( 1 + \frac{m_1^{(\ell)}}{m_0^{(\ell)}} e^{-2\beta\Delta^{(\ell)}} \right) \geq e^{-3\beta\Delta^{(\ell)}},
$$
$$\qquad (5.9)$$

for $\beta$ large enough (depending on $\ell$). On the other hand, Lemma 5.1 implies that

$$
-\tfrac{1}{2\beta} \log Z_{\ell,\beta} = E_0^{(\ell)} + R'(\ell, \beta), \qquad (5.10)
$$

where $|R'(\ell, \beta)| \leq C_\ell e^{-\beta c}$. We then have

$$
-\tfrac{1}{2\beta} \log m_0^{(\ell)} - \tfrac{1}{2\beta} R(\ell, \beta) = R'(\ell, \beta). \qquad (5.11)
$$

Using the bound $R(\ell, \beta) \leq n^{2\ell} e^{-2\beta\Delta^{(\ell)}}$ where $\Delta^{(\ell)} > 0$, and looking at the asymptotic $\beta \to \infty$, we see that $m_0^{(\ell)} = 1$. Next, using (5.9), we get

$$
\tfrac{1}{2\beta} e^{-3\beta\Delta^{(\ell)}} \leq C_\ell e^{-\beta c} \qquad (5.12)
$$

for all $\beta$ sufficiently large; this implies that $\Delta^{(\ell)} \geq \tfrac{1}{3}c$, uniformly in $\ell, n, u$. □

**Publisher's Note**   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## A. Appendix A. The Interaction $uT + vP$ When $n$ is Even

For $n$ odd, the interactions $uT + vP$ and $uT + vQ$ are related by the unitary transformation of Eq. (1.4). This holds for models defined on arbitrary graphs or lattices.

We now discuss the case of $n$ even. As we shall see, we need to restrict ourselves to bipartite graphs (of which the chain is of course an example). We work with the $S^{(3)}$-eigenbasis $e_\alpha := |\alpha\rangle$ with $\alpha = -S, -S+1, \ldots, S$. To begin, we define a unitary $V$ by setting

$$V|\alpha\rangle = (-1)^{S-\alpha} |-\alpha\rangle. \tag{A.1}$$

With $\psi$ the vector of (1.3) and $\phi$ the vector of (1.5), we have

$$\phi = (\mathbb{I} \otimes V)\psi. \tag{A.2}$$

Therefore, since $P$ is the projection onto $\phi$,

$$P = (\mathbb{I} \otimes V)Q(\mathbb{I} \otimes V^*). \tag{A.3}$$

Since $T\psi = \psi$ and $T\phi = -\phi$, we have $TQT = Q$ and $TPT = P$. Using these properties we find

$$(V \otimes \mathbb{I})Q(V^* \otimes \mathbb{I}) = (V \otimes \mathbb{I})TQT(V^* \otimes \mathbb{I}) = T(\mathbb{I} \otimes V)Q(\mathbb{I} \otimes V^*)T = P. \tag{A.4}$$

Both models are translation-invariant although the unitary that relates them is not:

$$(V \otimes \mathbb{I} \otimes V \cdots \otimes \mathbb{I}) \left[ \sum_{x=-\ell+1}^{\ell-1} Q_{x,x+1} \right] (V^* \otimes \mathbb{I} \otimes V^* \cdots \otimes \mathbb{I}) = \sum_{x=-\ell+1}^{\ell-1} P_{x,x+1}. \tag{A.5}$$

Let $\tilde{T}$ be the transformation of the operator $T$. We have

$$\tilde{T} = (\mathbb{I} \otimes V)T(\mathbb{I} \otimes V^*) = (\mathbb{I} \otimes V)(V^* \otimes \mathbb{I})T = -(V \otimes V)T. \tag{A.6}$$

Let us summarize the above considerations by the following proposition. We define the new Hamiltonian $H'_\ell = \sum_{x=-\ell+1}^{\ell-1} \left( u\tilde{T}_{x,x+1} + vQ_{x,x+1} \right)$.

**Proposition A.1** *For $n$ even, the interaction $uT + vP$ is unitarily equivalent with $u\tilde{T} + vQ$. The Hamiltonian $\tilde{H}_\ell$ defined in (1.14) is unitarily equivalent to $H'_\ell$.*

Notice that, when $u = 0$, the $Q$-model and the $P$-model are unitarily equivalent for all $n$. The proposition is stated for chains, but it clearly holds for arbitrary bipartite graphs. Next, we derive a loop representation for the model $H'_\ell$.

**Proposition A.2** *There exists a function $s(l)$ from the set of loops to $\pm 1$ such that for all $n \geq 2$,*

*(a) Tr* $e^{-2\beta H'_\ell} = e^{2\beta(1+u)|E_\ell|} \int d\rho_u(\omega) n^{\mathcal{L}(\omega)-|\omega_{\rightleftharpoons}|} \prod_{loop\, l\, in\, \omega} s(l).$

*(b) For $i = 1, 2, 3$,*

$$Tr\, S_x^{(i)} S_y^{(i)}\, e^{-2\beta H'_\ell} = \frac{n^2-1}{12}\, e^{2\beta(1+u)|E_\ell|} \int_{\Omega_{\ell,\beta}} d\rho_u(\omega)\, n^{\mathcal{L}(\omega)-|\omega_{\rightleftharpoons}|}$$

$$\times \left( \mathbb{1}[x \overset{+}{\longleftrightarrow} y] - \mathbb{1}[x \overset{-}{\longleftrightarrow} y] \right) \prod_{loop\, l\, in\, \omega} s(l).$$

Proposition A.2 is stated for chains, but it actually holds for arbitrary *bipartite* graphs (unlike Theorem 2.1 which holds for all finite graphs). For odd $n$ the signs $s(l)$ are all equal to $+1$.

*Proof.* First, observe that the number of crosses along the trajectory of a loop, is even (here, if a cross is traversed twice in a loop, it counts as two). Indeed, the total number of crosses and double-bars along the trajectory is even because the graph is bipartite; and the number of double-bars is even because the number of changes in vertical direction is even; so the number of crosses is also even.

The expansion of the operator $e^{-2\beta H'_\ell}$ can be made in terms of configurations $\omega$, and of "space-time spin configurations" (see [29]). The space-time spin configurations that are compatible with $\omega$ have the property that their value on a loop is $\pm\alpha$ for some $\alpha = -S, \ldots, S$, the changes of signs occurring when traversing crosses (and any such choice results in a possible space-time spin configuration because the number of crosses along a loop trajectory is even).

Proceeding as in Theorem 2.1, we find that

$$\text{Tr}\, e^{-2\beta H'_\ell} = e^{2\beta(1+u)|E_\ell|} \int d\rho_u(\omega) n^{\mathcal{L}(\omega)-|\omega_{\rightleftharpoons}|} s(\omega), \tag{A.7}$$

where $s(\omega)$ is an overall sign: $s(\omega) = \pm 1$. Notice that, since $\tilde{T}$ involves a minus sign, there is no need to change the sign of $u$ in the interaction as in Theorem 2.1.

The signs are due to the action of operators $V$. We can collect the signs for each loop individually. Consider two successive crosses. If the vertical direction is the same (which is the case if there is an even number of double-bars between them), we get the factor

$$(-1)^{S-\alpha}(-1)^{S+\alpha} = (-1)^{2S} = -1. \tag{A.8}$$

If the vertical direction is opposite (which is the case if there is an odd number of double-bars between them), the factor is

$$(-1)^{S-\alpha}(-1)^{S-\alpha} = 1. \tag{A.9}$$

This is illustrated in Fig. 18. The value of $s(l)$ is the product of these factors. Notice that the sign does not depend on the value of $\alpha$ in the loop. This proves item (a) of the proposition.
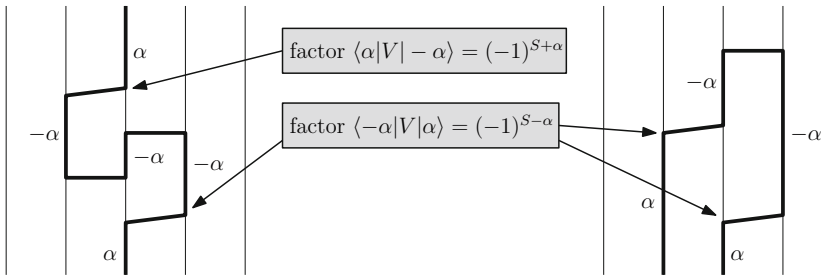
**Fig. 18.** Signs arising when traversing crosses. Left: the crosses are separated by an even number of double bars which yields the factor $(-1)^{S-\alpha}(-1)^{S+\alpha} = -1$. Right: the crosses are separated by an odd number of double bars which yields the factor 1

The spin correlations are the same for all $i = 1, 2, 3$ by symmetry and it is enough to consider $i = 3$. This is identical to [29, Theorem 3.5 (a)] except for the signs (the claim there was restricted to odd $n$ where $s(l) = +1$). Using space-time spin configurations, we have

$$\mathrm{Tr}\, S_x^{(i)} S_y^{(i)}\, \mathrm{e}^{-2\beta H_\ell'} = \mathrm{e}^{2\beta(1+u)|E_\ell|} \int_{\Omega_{\ell,\beta}} \mathrm{d}\rho_u(\omega)\, n^{-|\omega_{\boxminus}|}\Big(\prod_{\text{loop } l \text{ in } \omega} s(l)\Big) \sum_{\sigma \in \Sigma(\omega)} \sigma_{x,0}\sigma_{y,0}.$$

(A.10)

We used the fact that the signs do not depend on the spin values of the loops. If $(x, 0)$ and $(y, 0)$ do not belong to the same loop, the sum over $\sigma$ is zero. If $(x, 0)$ and $(y, 0)$ belong to the same loop and the connection is $x \overset{+}{\longleftrightarrow} y$, then $\sigma_{x,0} = \sigma_{y,0}$ and the sum gives $\frac{3}{S}(S + 1)n = \frac{1}{12}(n^2 - 1)n$. If the connection is $x \overset{-}{\longleftrightarrow} y$, then $\sigma_{x,0} = -\sigma_{y,0}$ and we get minus the same factor. This gives the identity (b).  $\square$

We can now prove Theorem 1.3.

*Proof of Theorem 1.3.* By Proposition A.1, the claims of Theorem 1.3 are equivalent to proving dimerization in the model with Hamiltonian $H_\ell'$. We use the loop representation of Proposition A.2. We can then retrace the steps of the proof of Theorem 4.7. In doing so, note that all short loops $l$ have $s(l) = +1$, while long or winding loops have $s(l) = \pm 1$. We incorporate the latter factors in the weights $w(\gamma)$ of the contours, see (4.6). Therefore, the only difference is that the weights of contours have possibly other signs. All bounds are the same, though, and the cluster expansion gives the same result.  $\square$

The proof of the gap for $\tilde{H}_\ell$ is exactly the same as the proof for $H_\ell$ described in Sect. 5.

## References

1. Affleck, I.: Exact results on the dimerisation transition in $su(n)$ antiferromagnetic chains. J. Phys. Condens. Matter **2**, 405–415 (1990)
2. Affleck, I., Kennedy, T., Lieb, E.H., Tasaki, H.: Valence bond ground states in isotropic quantum antiferromagnets. Commun. Math. Phys. **115**, 477–528 (1988)
3. Aizenman, M., Duminil-Copin, H., Warzel, S.: Dimerization and Néel order in different quantum spin chains through a shared loop representation. Ann. Henri Poincaré **21**, 2737–2774 (2020)
4. Aizenman, M., Nachtergaele, B.: Geometric aspects of quantum spin states. Commun. Math. Phys. **164**, 17–63 (1994)

5. Babujian, H.: Exact solution of the one-dimensional isotropic Heisenberg chain with arbitrary spins S. Phys. Lett. **90A**, 479–482 (1982)
6. Barber, M.N., Batchelor, M.T.: Spectrum of the biquadratic spin-1 antiferromagnetic chain. Phys. Rev. B **40**, 4621–4626 (1989)
7. Duminil-Copin, H., Li, J.-H., Manolescu, I.: Universality for the random-cluster model on isoradial graphs. Electron. J. Probab. **23**, 1–70 (2018)
8. Dyson, F.J., Lieb, E.H., Simon, B.: Phase transitions in quantum spin systems with isotropic and non-isotropic interactions. J. Stat. Phys. **18**, 335–383 (1978)
9. Hastings, M.B., Koma, T.: Spectral gap and exponential decay of correlations. Commun. Math. Phys. **265**, 781–804 (2006)
10. Itoi, C., Kato, M.-H.: Extended massless phase and the Haldane phase in a spin-1 isotropic antiferromagnetic chain. Phys. Rev. B **55**, 8295–8303 (1997)
11. Kennedy, T., Tasaki, H.: Hidden symmetry breaking and the Haldane phase in $S = 1$ quantum spin chains. Commun. Math. Phys. **147**, 431–484 (1992)
12. Klümper, A.: The spectra of $q$-state vertex models and related antiferromagnetic quantum spin chains the spectra of q-state vertex models and related antiferromagnetic quantum spin chains. J. Phys. A Math. Gen. **23**, 809–823 (1990)
13. Kotecký, R., Preiss, D.: Cluster expansion for abstract polymer models. Commun. Math. Phys. **103**, 491–498 (1986)
14. Landau, L.J., Perez, J.F., Wreszinski, W.F.: Energy gap, clustering, and the Goldstone theorem in statistical mechanics. J. Stat. Phys. **26**, 755–766 (1981)
15. Lees, B.: Existence of Néel order in the S=1 bilinear-biquadratic Heisenberg model via random loops. Commun. Math. Phys. **347**, 83–101 (2016)
16. Nachtergaele, B., Sims, R.: Lieb-Robinson bounds and the exponential clustering theorem. Commun. Math. Phys. **265**, 119–130 (2006)
17. Nachtergaele, B., Sims, R., Young, A.: Quasi-Locality Bounds for Quantum Lattice Systems. Part II. Perturbations of Frustration-Free Spin Models with Gapped Ground States, arXiv:2010.15337 (2020)
18. Nachtergaele, B., Ueltschi, D.: A direct proof of dimerization in a family of $SU(n)$-invariant quantum spin chains. Lett. Math. Phys. **107**, 1629–1647 (2017)
19. Nepomechie, R.I., Pimenta, R.A.: Universal Bethe ansatz solution for the Temperley-Lieb spin chain. Nucl. Phys. B **910**, 910–928 (2016)
20. Ray, G., Spinka, Y.: A short proof of the discontinuity of phase transition in the planar random-cluster model with $q > 4$. Commun. Math. Phys. **378**, 1977–1988 (2020)
21. Reshetikhin, N.Y.: A method of functional equations in the theory of exactly solvable quantum systems. Lett. Math. Phys. **7**, 205–213 (1983)
22. Ryan, K.: The free energy of a class of spin 1/2 and 1 quantum spin systems on the complete graph, arXiv:2011.07007 (2020)
23. Sutherland, B.: Model for a multicomponent quantum system. Phys. Rev. B **12**, 3795–3805 (1975)
24. Takhtajan, L.A.: The picture of low-lying excitations in the isotropic Heisenberg chain of arbitrary spins. Phys. Lett. **87A**, 479–482 (1982)
25. Tanaka, K., Tanaka, A., Idokagi, T.: Long-range order in the ground state of the S=1 isotropic bilinear-biquadratic exchange Hamiltonian. J. Phys. A **34**, 8767–8780 (2001)
26. Tóth, B.: Improved lower bound on the thermodynamic pressure of the spin 1/2 Heisenberg ferromagnet. Lett. Math. Phys. **28**, 75–84 (1993)
27. Tu, H.-H., Zhang, G.-M., Xiang, T.: Class of exactly solvable $SO(n)$ symmetric spin chains with matrix product ground states. Phys. Rev. B **78**, 094404 (2008)
28. Ueltschi, D.: Cluster expansions and correlation functions. Moscow Math. J. **4**(2), 511–522 (2004)
29. Ueltschi, D.: Random loop representations for quantum spin systems. J. Math. Phys. **54**(083301), 1–40 (2013)