STOCHASTIC PDE LIMIT OF THE SIX VERTEX MODEL

IVAN CORWIN, PROMIT GHOSAL, HAO SHEN, AND LI-CHENG TSAI

Abstract. We study the stochastic six vertex model and prove that under weak asymmetry scaling (i.e., when the parameter $\Delta \to 1^+$ so as to zoom into the ferroelectric/disordered phase critical point) its height function fluctuations converge to the solution to the Kardar–Parisi–Zhang (KPZ) equation. We also prove that the one-dimensional family of stochastic Gibbs states for the symmetric six vertex model converge under the same scaling to the stationary solution to the stochastic Burgers equation. We achieve this through a new Markov duality method which shows how duality can be leveraged to prove previously inaccessible results in hydrodynamics and SPDE limits.

Our starting point is an exact microscopic Hopf–Cole transform for the stochastic six vertex model which follows from the model’s known one-particle Markov self-duality. Given this transform, the crucial step is to establish self-averaging for specific quadratic function of the transformed height function. We use the model’s two-particle self-duality to produce explicit expressions (as Bethe ansatz contour integrals) for conditional expectations from which we extract time-decorrelation and hence self-averaging in time. The crux of the Markov duality method is that the entire convergence result reduces to precise estimates on the one-particle and two-particle transition probabilities.

Contents

1. Introduction 2
1.1. KPZ equation as a limit of the stochastic six vertex model 2
1.2. Stochastic Burgers equation as a limit of symmetric six vertex model 5
1.3. KPZ equation as a limit of ASEP 9
1.4. Markov duality method 11
1.5. Further literature 12
2. Stochastic and symmetric six vertex models 15
2.1. Stochastic six vertex model as a particle system and its height function 15
2.2. Stochastic Gibbs states for the symmetric six vertex model 18
3. Self duality for stochastic six vertex model 18
4. Hopf–Cole transform: reformulation of Theorem 1.1 20
4.1. Microscopic Hopf–Cole transform 20
4.2. The SHE 23
5. Proof of Theorems 1.1* and 1.6 24
5.1. Proof of Theorem 1.1* 25
5.2. Proof of Theorem 1.6 29
6. Estimating the two-point semigroup 31
6.1. Estimating the free part $V_{\epsilon}^{fr}$ 32
6.2. Estimating the interacting part $V_{\epsilon}^{in}$, an overview 35
6.3. Estimating the interacting part $V_{\epsilon}^{in}$, the $(+-)$-case 40
6.4. Estimating the interacting part $V_{\epsilon}^{in}$, the $(-+)$-case 44
6.5. Estimating the interacting part $V_{\epsilon}^{in}$, the $(++)$-case 47
7. Controlling the quadratic variation: Proof of Proposition 5.3 51
7.1. Expanding the quadratic variation 51
7.2. Time decorrelation via duality 56
Appendix A. Quadratic variation in ASEP 59

Key words and phrases. six vertex model, Kardar–Parisi–Zhang equation, interacting particle system, Markov duality.
1. Introduction

The Six Vertex (6V) model and the Kardar–Parisi–Zhang (KPZ) equation are widely studied models in equilibrium and non-equilibrium statistical mechanics. In this paper we demonstrate how a certain scaling limit of the former model converges to the later equation. This limit comes from scaling into the critical point dividing the ferroelectric and disordered phases of the model. Our results apply for both the stochastic and symmetric 6V models (Theorems 1.1 and 1.6 respectively). The technical core of this paper is the Markov duality method which we introduce and develop. One-particle duality allows us to perform a microscopic Hopf-Cole transform of the model’s height function process into a discrete stochastic heat equation, and prove tightness of that resulting equation. Two-particle duality controls the quadratic variation of the martingale part and proves precise self-averaging in time.

The structure of this introduction is as follows: Section 1.1 introduces the stochastic 6V model and records our first main result, it convergence to the KPZ equation (Theorem 1.1). Section 1.2 introduces the symmetric 6V model and records our second main result, the convergence of the one-parameter family of stochastic Gibbs states to the stationary solution to the stochastic Burgers equation (Theorem 1.6). This section also describes the model with external fields and how the stochastic Gibbs states arise in the (conjectural) phase diagram for the model’s Gibbs states. Section 1.3 recalls how the KPZ equation arises as a scaling limit for ASEP (a well studied continuous time limit of the stochastic 6V model). The purpose of this is to highlight (in the simplest case possible) the key technical challenge in proving such results—self average of the quadratic variation. Section 1.4 briefly introduces our new Markov duality method approach in the context of ASEP. This approach is developed fully for the stochastic 6V model in the main body of the paper. Section 1.5 provides a brief review of related literature studying the symmetric and stochastic 6V models, KPZ equation, and Markov dualities.

1.1. KPZ equation as a limit of the stochastic six vertex model. The stochastic 6V model is a discrete time interacting particle system introduced in 1992 by Gwa and Spohn [GS92]. The model depends on two parameters $b_1, b_2 \in (0,1)$ which are used to define (positive) weights on six type of vertices—see the top row of Figure 1. Treating the solid lines entering a vertex from below or the left as inputs and those exits above or to the right as outputs, these vertex are conservative (i.e., the number of input lines equals the number of output lines) and stochastic (i.e., for fixed inputs, the sum of weights over outputs is always 1, and the individual weights are non-negative). Given a down-right path in $\mathbb{Z}^2$ and a specification of boundary condition inputs along the path, the stochastic 6V model is a measure on the vertices to the up and right of the path, or equivalently a measure on the collection of solid lines which leave the boundary inputs and continue in the up and right directions. The measure is defined recursively: starting with vertices with inputs given, the outputs are randomly chosen amongst all possible outputs with probabilities given by the associated vertex weights. The left-side of Figure 2 illustrates when the boundary condition inputs are specified on the coordinate axes for the first quadrant. See Section 2.1 for a more precise definition of the model (including a bi-infinite version) and Section 1.5.2 for a brief review of related literature.

| Non-crossing paths | \[ \] | \[ \] | \[ \] | \[ \] | \[ \] | \[ \] |
|--------------------|------|------|------|------|------|------|
| Stochastic weights | 1    | 1    | $b_1$ | $b_2$ | $1 - b_1$ | $1 - b_2$ |
| Symmetric weights  | $a$  | $a$  | $b$  | $b$  | $c$  | $c$  |
| Asymmetric weights | $e^{-H-V}a$ | $e^{H+V}a$ | $e^{-H+V}b$ | $e^{H-V}b$ | $c$  | $c$  |

Figure 1. Six vertices with their stochastic, symmetric and asymmetric weights.

If the boundary condition inputs are specified entirely on the horizontal axis, it is natural to think of vertical solid lines as particles evolving in time (as measured by the $y$-coordinate) via the following
Markovian update. Start with left-most particle\(^1\). With probability \(b_1\) it stays put, and with \(1 - b_1\) it moves one to the right. The particle continues to move right with probability \(b_2\) per step until it either stops, or it hits the next particle. When no collision happens, repeat these rules for the next particle to the right. If a collision occurs, the moving particle stops at that site and the next particle starts moving to the right with probability 1, and continues to move with probability \(b_2\) (as usual). See Section 1.5.2 for a discussion of some limit of the stochastic 6V model.

![Figure 2](image)

**Figure 2.** Left: Particle trajectories for the stochastic 6V model with boundary condition inputs along the coordinate axes. Right: Periodic boundary conditions.

Define the height function \(N(t,x)\) for the stochastic 6V model to be equal to the net number of particles which have moved across the time-space line between \((0,0)\) and \((t,x)\) (i.e., summing 1 for each left-to-right move and \(-1\) for each right-to-left move—see Figure 3). For a precise definition as well as a construction of \(N(t,x)\) for bi-infinite configurations, see Section 2.1. Given such \(N(t,x)\), we first linearly interpolate in \(x \in \mathbb{Z}\) and then linearly interpolate in \(t \in \mathbb{Z}_{\geq 0}\) to make \(N(t,x) \in C([0,\infty),C(\mathbb{R}))\). Hereafter, we endow the space \(C(\mathbb{R})\) and \(C([0,\infty),C(\mathbb{R}))\) the topology of uniform convergence over compact subsets, and write \(\Rightarrow\) for the weak convergence of probability laws.

![Figure 3](image)

**Figure 3.** The stochastic 6V particle trajectories and associated height functions. Here we assume a left-most particle and label the first five particles \(x_1,\ldots,x_5\). The lines represent their temporal trajectories. The dark grey numbers represent the height function \(N(t,y)\) for different regions. The height function changes value when crossing particle trajectories (increasing as one crosses from left to right).

\(^1\)If there is no left-most particle, the dynamics can be still be defined with some care—see Section 2.1.
Our main result for the stochastic 6V model states that, under weak asymmetry scaling where \( b_1 \in (0, 1) \) is fixed and \( \tau = b_2/b_1 = e^{-\sqrt{\varepsilon}} \to 1 \), an analog of (1.25) holds for \( N(t, x) \). To setup notations, we fix any density \( \rho \in (0, 1) \) hereafter, and let

\[
\lambda = \frac{1 - b_2 \tau^{-\rho}}{b_1 - (b_1 + b_2 - 1) \tau^{-\rho}} = \frac{1 - b_1 \tau^{-1-\rho}}{b_1 - (b_1 + b_1 \tau - 1) \tau^{-\rho}},
\]

\[
\mu = \frac{\tau^{-\rho}(1 - b_1)(1 - b_2)}{(b_1 - (b_1 + b_2 - 1) \tau^{-\rho})(1 - b_2 \tau^{-\rho})} = \frac{\tau^{-\rho}(1 - b_1)(1 - b_1 \tau)}{(b_1 - (b_1 + b_1 \tau - 1) \tau^{-\rho})(1 - b_1 \tau^{-1-\rho})}.
\]

The reason of choosing these values of the parameters \( \lambda, \mu \) will be clear in Section 4.1. Specifically, under the weak asymmetry scaling where \( b_1 \in (0, 1) \) fixed and \( \tau = \tau_\varepsilon = b_2/b_1 := e^{-\sqrt{\varepsilon}} \), we have \( \lambda = \lambda_\varepsilon \) and \( \mu = \mu_\varepsilon \), which, up to first order in \( \sqrt{\varepsilon} \), read

\[
\lambda_\varepsilon = 1 - \rho \sqrt{\varepsilon} + O(\varepsilon),
\]

\[
\mu_\varepsilon = 1 + \frac{b_1 - 2b_2 \rho}{b_1 - 1} \sqrt{\varepsilon} + O(\varepsilon).
\]

We adopt standard notation \( O(a) \) to denote a generic quantity such that \( \sup_{0<\varepsilon<1} |O(a)| \varepsilon^{-1} < \infty \). Recall the KPZ equation (see Section 4.2 for its definition; Sections 1.3 and 1.5.3 a literature review).

\[
\partial_t \mathcal{H}(t, x) = \frac{\nu_\varepsilon}{2} \partial^2_x \mathcal{H}(t, x) - \frac{\kappa_\varepsilon}{2} (\partial_x \mathcal{H}(t, x))^2 + \sqrt{D_\varepsilon} \xi(t, x),
\]

with coefficients

\[
\nu_\varepsilon := \frac{2b_1}{1 - b_1}, \quad \kappa_\varepsilon := \frac{2b_1}{1 - b_1}, \quad D_\varepsilon := \frac{2b_1 \rho (1 - \rho)}{1 - b_1}.
\]

**Theorem 1.1.** Consider the stochastic 6V model, with parameter \( b_1 > b_2 \in (0, 1) \).

(a) **(Near stationary initial conditions)** Fix a density \( \rho \in (0, 1) \). With \( \varepsilon \downarrow 0 \) denoting a scaling parameter, we start the stochastic 6V model from a sequence of initial conditions \( \{N_\varepsilon(0, x)\}_{\varepsilon>0} \), and let \( N_\varepsilon(t, x) \) denote the resulting height function. Assume that \( \{N_\varepsilon(0, x)\}_{\varepsilon>0} \) is near stationary with density \( \rho \) (Definition 4.4), and that for some \( C(\mathbb{R}) \)-valued process \( \mathcal{H}^{ic}(x) \),

\[
\sqrt{\varepsilon}(N_\varepsilon(0, \varepsilon^{-1} - 1) - \rho \varepsilon^{-1} x) \Rightarrow \mathcal{H}^{ic}(x), \quad \text{in } C(\mathbb{R}).
\]

Then, under the weak asymmetry scaling where \( b_1 \in (0, 1) \) is fixed, \( \tau = \tau_\varepsilon = b_2/b_1 := e^{-\sqrt{\varepsilon}} \), and \( \lambda \) and \( \mu \) depend on \( \varepsilon \) as in (1.3) and (1.4), we have

\[
\sqrt{\varepsilon}(N_\varepsilon(\varepsilon^{-2} t, \varepsilon^{-1} x + \mu_\varepsilon \varepsilon^{-2} t) - \rho(\varepsilon^{-1} x + \mu_\varepsilon \varepsilon^{-2} t)) - \varepsilon^{-2} t \log \lambda_\varepsilon \Rightarrow \mathcal{H}(t, x),
\]

\[
\text{in } C([0, \infty), C(\mathbb{R})),
\]

where \( \mathcal{H}(t, x) \) is the Hopf–Cole solution (defined in Section 4.2) of the KPZ equation (1.5) with initial condition \( \mathcal{H}^{ic}(x) \).

(b) **(Step initial condition)** Start the stochastic 6V model from the step initial condition \( N(0, x) = (x)_+ := \max(0, x) \), and let \( N_\varepsilon(t, x) \) denote the resulting height function. Fix \( \rho \in (0, 1) \). Under the weak asymmetry scaling where \( b_1 \in (0, 1) \) is fixed, \( \tau = \tau_\varepsilon = b_2/b_1 := e^{-\sqrt{\varepsilon}} \), and \( \lambda \) and \( \mu \) depend on \( \varepsilon \) as in (1.3) and (1.4), we have

\[
\sqrt{\varepsilon}(N_\varepsilon(\varepsilon^{-2} t, \varepsilon^{-1} x + \mu_\varepsilon \varepsilon^{-2} t) - \rho(\varepsilon^{-1} x + \mu_\varepsilon \varepsilon^{-2} t)) - \varepsilon^{-2} t \log \lambda_\varepsilon - \log \frac{\rho(1-\rho)}{\sqrt{\varepsilon}} \Rightarrow \mathcal{H}(t, x),
\]

\[
\text{in } C((0, \infty), C(\mathbb{R})),
\]

where \( \mathcal{H}(t, x) \) is the Hopf–Cole solution of the KPZ equation (1.5) with narrow wedge initial condition (see Section 4.2).

**Remark 1.2.** It is worth remarking on the freedom to choose arbitrary \( \rho \in (0, 1) \) in the theorem. For the near stationary initial conditions, \( \rho \) controls the density of particles (or vertical lines) as well as the characteristic velocity around which we focus. For step initial data, \( \rho \) determines a velocity within
the rarefaction fan (and gives the density around that velocity). Previous KPZ equation limit results
for ASEP where limited to $\rho = 1/2$ since the arguments become more complicated in a moving frame
or with a disproportionate number of particles to holes.

**Remark 1.3.** [CT17] proves KPZ equation convergence for a portion of the class of higher spin
stochastic vertex models [CP16]. Those models fall into two types – those with spin $I, J \in \mathbb{Z}_{\geq 1}$ in
which the number of particles or arrows per edge is bounded by $I$ or $J$ (depending on the edge’s
orientation) and those with non-integer spin in which there may be an infinite number of particles or
arrows per edge. [CT17] analyzed this second class, specifically under scaling in which the expected
number of particles per site diverges with $\varepsilon$. This simplifies analysis quite dramatically since [CT17]
is able to Taylor expand the quadratic martingale in the density yielding an analysis which completely
avoids the key complexity which we encounter here. The stochastic 6V model comes from taking
$I = J = 1$ and hence the number of particles per site is either 0 or 1. Though we do not address the
general $I, J \in \mathbb{Z}_{\geq 1}$ class herein, we expect our Markov duality method is applicable there.

**Remark 1.4.** Plugging the expansions (1.3) and (1.4) for $\lambda_\varepsilon$ and $\mu_\varepsilon$ into (1.8), one can see that the
two terms of vertical shifting of height function, namely $-\sqrt{\varepsilon \rho(\mu_\varepsilon \varepsilon^{-2}t)}$ and $-\varepsilon^{-2}t \log \lambda_\varepsilon$, are both of
order $O(\varepsilon^{-3/2})$; but their order $O(\varepsilon^{-3/2})$ parts cancel out. Therefore (1.8) states that the rescaled
and tilted height function subtracting $O(\varepsilon^{-1})t$ converges to the solution to KPZ equation.

**Proof sketch.** Proposition 4.1 provides an exact microscopic Hopf-Cole transform through which the
stochastic 6V model height process is relates to a microscopic Stochastic Heat Equation (SHE). This
transformation is readily seen as a consequence of the (one-particle) Markov self-duality given in
Corollary 3.4. Theorem 1.1* proves convergence of this microscopic SHE to the continuum SHE.
When translated back into the stochastic 6V model height function, this implies Theorem 1.1.

The proof of Theorem 1.1* boils down to showing tightness and identifying the limiting linear and
quadratic martingale problem. The first two items follow in a standard manner from moment bounds
provided by Proposition 5.1. Controlling the quadratic variation is the hard part. Proposition 5.3
does this by proving a form of self-averaging for the quadratic variation (which itself is a quadratic
in the solution to the microscopic SHE). The proof of the self-average relies upon the two-particle
duality through Proposition 4.3. That duality reduces the calculation of conditional expectations
to computations involving the transition probability for a two-particle version of the stochastic 6V
model. Such transition formulas can be written explicitly using Bethe ansatz—see Proposition 3.5
or the formula in (4.11). Proposition 6.1 contains very precise estimates on the two-point transition
probabilities which are proved via involved steepest descent analysis on the double-contour integral
formulas encoding these probabilities.

In Sections 1.3 and 1.4 (and Appendix A) we explain how these ideas work in the simpler the context
of ASEP. For ASEP, there are other methods which can be used to prove self-averaging. Presently,
the Markov duality method we introduce here is the only approach which works for the 6V model.

1.2. **Stochastic Burgers equation as a limit of symmetric six vertex model.** The symmetric
6V model is a foundational model in 2D equilibrium statistical mechanics. It is defined with respect
to a pre-imposed a choice of boundary condition on a compact domain in $\mathbb{Z}^2$, e.g. periodic boundary
condition on a rectangular domain as in Figure 2. Then, one chooses an assignment of vertices inside
the domain which fit together (i.e. output lines match input lines from vertices to the right or above)
with probability proportional to the product of vertex weights. These weights are specified by $a, b, c > 0$
in (fact, by scaling only two of these matter) as in Figure 1 and the model is called symmetric since
reflecting the vertices over the diagonal does not change their weight. To go from such a product of
weights to a probability requires dividing by a normalizing constant (also called a partition function)
which is the sum over all configurations of the product of weights. The need to normalize was not
present in the case of stochastic weights as it equals 1 there.

1.2.1. **Conjectural phase diagram for symmetric six vertex model Gibbs states.** How does the symmetric
6V model behave as the mesh size goes to zero? Is there a limit shape? How does the height
function fluctuate around it? How much do boundary conditions or external fields effect these limits? These questions are intertwined with understanding the extremal, translation invariant, ergodic infinite volume Gibbs states (or simply Gibbs states for short) and their associated free energies. These can be thought of as distributions on configurations of vertices on $\mathbb{Z}^2$ which satisfy the symmetric 6V Gibbs property—the marginal distribution restricted to any compact subdomain, given the state of the boundary vertices, is given by the above symmetric 6V model probability prescription (i.e., product over weights of vertices normalized to be a probability distribution).

While much has been conjectured about the symmetric 6V Gibbs states (e.g. their phase diagram, free energy, uniqueness, and fluctuations) very little has been proved—see Section 1.5.1 for some further discussion. The description we provide below (which is not used in any proofs) can be found, for instance, in [Nol92, BS95, Res10] and is essentially conjectural. We include it here to motivate the importance of studying the stochastic Gibbs states as we do in Section 1.2.2.

The Gibbs states for the symmetric 6V model (with a given choice of $a, b, c$) are believed to arise as infinite volume limits of the periodic boundary condition asymmetric 6V model in which there are horizontal and vertical external fields of strength $H, V \in \mathbb{R}$ (see Figure 1). These fields reward the occurrence of horizontal or vertical lines by factors of $e^{H/2}$ and $e^{V/2}$ and penalize the absence of lines by $e^{-H/2}$ and $e^{-V/2}$. Consider any rectangle enclosed in the interior of the fundamental domain of the periodic model. Then, regardless of the choices of external fields, conditioned on the vertices on the boundary of the rectangle, the configuration inside is given by the symmetric, zero-field 6V model weights. This is because all possible vertex configurations inside the rectangle have the same number of vertical and horizontal lines. This is analogous to the fact that for a simple random walk with drift, the marginal distribution of the walk given a fixed starting and ending level is drift-independent.

![Figure 4](image-url)

**Figure 4.** 6V model with parameters $(a, b, c) \approx (0.201, 0.1, 1)$ (or $u, \eta = .1$) and $\Delta \approx 1.005$. (a): Phase diagram mapping $(H, V)$ onto Gibbs states. (b): Average density of horizontal and vertical lines $(h, v)$ accessible as $(H, V)$ varies. The A1 phase maps to $(h, v) = (1, 1)$, A2 to $(0, 0)$, B2 to $(0, 1)$, B1 to $(1, 0)$. The disordered phase D2 maps to the grey area above the diagonal in the $(h, v)$ plot, and D1 to the reflected area. The disordered phase extends asymptotically vertically and horizontally so as to separate the A and B phases. The conical points are where D2/D1, A1 and A2 touch. This point maps to the entire boundary of the white lens around the $(h, v)$ diagonal. Inside the lens there are no (extremal) Gibbs states with those specified densities.

Gibbs state are believed to be uniquely indexed by their average density $(h, v) \in [0, 1]^2$ of horizontal and vertical lines (respectively). It is not necessary that every $(h, v)$ will have a corresponding Gibbs
state which realizes those densities. [Res10] describes the conjectural mapping (derived based on Bethe ansatz calculations) between \((H, V)\) and \((h, v)\). The nature of this mapping depends on the parameter

\[
\Delta = \frac{a^2 + b^2 - c^2}{2ab}.
\]  

(1.9)

We will focus on the case when \(\Delta > 1\) and \(a > b + c\) (the other possible case when \(\Delta > 1\) is \(b > a + c\) and that can be recovered by a simple transformation of vertices) in which the conjectural phase diagram is given in Figure 4^2 (see also the caption beneath). There are four frozen phases which arise when \(H\) and \(V\) are sufficiently positive or negative. Between them are *disordered* phases which map onto values of \((h, v)\) in the grey region. [Nie84] (see more recently [KMSW17]) conjectured that the fluctuations in the disordered phase are log-correlated and related to the Gaussian free field (or central charge 1 CFT). Such a result has only been proved at the free-fermion \((\Delta = 0)\) point [Ken00, Ken01, Ken09].

In Figure 4(A) the disordered regions D1 and D2 terminate near the origin at *conical points* connected by a line between the A1 and A2 phases. In Figure 4(B) these conical points are mapped to the entire boundary between the grey disordered phase and the white excluded phase (the lens around the diagonal which do not have corresponding extremal Gibbs states). The different conical Gibbs states arise depending on the angle in the \((H, V)\)-plane along which one approaches the conical point. [BS95] argued that the one-parameter family of Gibbs states which arise at these conical points coincides with the one-parameter family of so-called *stochastic* Gibbs states.

1.2.2. *Stochastic Gibbs states and their scaling limits*. Whereas even the existence of disordered Gibbs states is only conjectural for \(\Delta \neq 0\), the one-parameter family of stochastic Gibbs states (which enjoy the symmetric \((a, b, c)\) Gibbs property) is readily constructed owing to their connection with the stochastic 6V model. Fix \((a, b, c)\) and consider the stochastic 6V model with parameters^3

\[
b_1 = \frac{b}{a}(\Delta + \sqrt{\Delta^2 - 1}), \quad \text{and} \quad b_2 = \frac{b}{a}(\Delta - \sqrt{\Delta^2 - 1}).
\]

(1.10)

Choose \((h, v) \in [0, 1]^2\) such that

\[
\frac{v}{1 - v}(1 - b_2) = \frac{h}{1 - h}(1 - b_1).
\]

(1.11)

There is a one-parameter family of solutions \((h, v)\) to this relation. Consider boundary condition inputs for the stochastic 6V model on the first quadrant where with probability \(h\) there are horizontal lines coming in from the \(y\)-axis, and with probability \(v\) there are vertical lines coming in from the \(x\)-axis (all these occur independently). [Agg16] (see also Lemma 2.6 below) proves that this boundary condition is stationary so that if one shifts the coordinates of the origin into the third quadrant, the marginal distribution restricted to the first quadrant remain unchanged. Shifting the origin back to \((-\infty, -\infty)\) defines the stochastic Gibbs state with line densities \((h, v)\), which we denote by \(SG(h, v)\). Figure 5 illustrates such a stochastic Gibbs state.

The one-parameter family of stochastic Gibbs states coincide with the densities which are conjectured to arise from the conical point (i.e. the boundary of the white lens in Figure 4)^4. Let us briefly make this matching to the formula for that lens boundary given in [RS16]. When \(\Delta > 1\) and \(a > b + c\), Baxter introduced a convenient (projective) parameterization of \((a, b, c)\):

\[
a = \sinh(u + \eta), \quad b = \sinh(u), \quad c = \sinh(\eta),
\]

(1.12)

with \(u, \eta > 0\). Note in particular \(\Delta = \cosh(\eta)\). In terms of this parameterization, the conjectural (see, for example, [RS16]) one-parameter family of conical Gibbs states have horizontal and vertical line

---

^2When \(|\Delta| < 1\) the conical points in the phase diagram disappear and the two disordered phases merge. When \(\Delta < -1\) a new antiferroelectric phase emerges for \(H, V\) near zero. The associated Gibbs state is composed of diagonal bands of zig-zags made up only the \(c\)-type vertices.

^3This relation can be reversed to give \(\Delta = \frac{b_1 + b_2}{2b_2}\).

^4In fact, (1.11) only gives bottom boundary of the lens. The other boundary comes from applying the diagonal symmetry of the symmetric model.
densities given by the relation
\[ h = \frac{v(1 \pm \tanh(u + \eta))}{1 \pm \tanh(u + \eta)(2v - 1)}. \tag{1.13} \]
and the conical points arises from choosing \((H,V) = (\pm \eta/2, \mp \eta/2)\).

Our main theorem on symmetric 6V model describes the large scale behaviors of the stochastic Gibbs state when \(\Delta \downarrow 1\)—that is, when we zoom into the ferroelectric-disorder interface.

A natural quantity describing large scale behavior of Gibbs states is the empirical distributions of vertical or horizontal lines. We will focus on vertical lines, and the analogous result on horizontal lines is obtained through exchanging \(x\)- and \(y\)-axes. Given a tiling on \(\mathbb{Z}^2\) by the six vertices from Figure 1, for each point \((x,y) \in \mathbb{Z}^2\), we let \(u(x,y)\) denote the indicator function for having an incoming (i.e., from below) vertical line. More explicitly,
\[ u(x,y) = 1\{(x,y) \text{ is tiled with } \begin{array}{c} \rightarrow \\ \downarrow \\ \leftarrow \end{array}\}. \tag{1.14} \]
For a fixed \(v \in (0,1)\) average density of vertical lines, we define the scaled empirical distribution \(U_\varepsilon\), acting on \(f \in C^\infty_c(\mathbb{R}^2)\) (\(C^\infty\) with compact support) as
\[ \langle U_\varepsilon, f \rangle := \varepsilon^{\frac{5}{2}} \sum_{x,y \in \mathbb{Z}} (u(x,y) - v) f(\varepsilon^{-1}x - \mu_\varepsilon \varepsilon^{-2}y, \varepsilon^{-2}y). \tag{1.15} \]
Here \(\mu_\varepsilon\) is the proper centering of the reference frame in order the observe KPZ-type fluctuations. In terms of Baxter’s projective parametrization (1.12), for fixed \(u > 0\), \(\mu_\varepsilon\) is obtained by matching \(b_1, b_2\) into \(u, \eta\) and via (1.10)–(1.12) in (1.2), and set \(\eta = \eta_\varepsilon = \frac{1}{2}\sqrt{\varepsilon}\) and \(\tau = \tau_\varepsilon = e^{-\sqrt{\varepsilon}}\).

Informally speaking, the \(\varepsilon \downarrow 0\) limit of the empirical distribution \(U_\varepsilon\) is described by the stationary solution of the Stochastic Burgers Equation (SBE):
\[ \partial_t U = \nu_s \partial_{xx} U + \kappa_s \partial_x(U^2) + \sqrt{D_s} \partial_x \xi. \tag{1.16} \]

To formulate our result precisely, first note that the solution \(U\) of the SBE (1.16) is a distribution (i.e., generalized function) valued process. In the following we will work with the space \(C^{-1}(\mathbb{R}^2)\) of distributions. For \(f \in C^\infty_c(\mathbb{R}^2)\), write \(f_\delta(x,y) := f(\delta^{-1}x,y)\) for the corresponding scaled function. This scaling probes only the regularity in \(x\). For linear functionals \(U,U'\) on \(C^\infty_c(\mathbb{R}^2)\), define
\[ \|U\|_{C^{-1}(\mathbb{R}^2),[-\ell,\ell]^2} := \sup \{ \|U(f_\delta)\| : \delta \in (0,1), f \in C^\infty_c(\mathbb{R}^2), \text{ supp}(f) \subset [-\ell,\ell]^2, \|f\|_\infty + \|\partial_x f\|_\infty \leq 1\}, \tag{1.17} \]
\[ d_{C^{-1}(\mathbb{R}^2)}(U,U') := \sum_{\ell=1}^\infty (2^{-\ell} \wedge \|U - U'\|_{C^{-1}(\mathbb{R}^2),[-\ell,\ell]^2}). \tag{1.18} \]

\[5\text{In [RS16], } t = 2h - 1 \text{ and } s = 2v - 1.\]
The space $C^{-1}(\mathbb{R}^2)$ consists of linear functional $U : C_c^\infty(\mathbb{R}^2) \rightarrow \mathbb{R}$ satisfying $d_{C^{-1}(\mathbb{R}^2)}(U, 0) < \infty$, endowed with the metric $d_{C^{-1}(\mathbb{R}^2)}(\cdot, \cdot)$.

To define the stationary solution of the SBE (1.16), consider the Hopf–Cole solution $\mathcal{H}_{\text{stat}}(t, x) \in C([0, \infty), C(\mathbb{R}))$ of the KPZ equation (1.5), with initial condition

$$\mathcal{H}_{\text{stat}}(0, x) = \sqrt{\rho(1 - \rho)} B(x), \quad \rho = v,$$

(1.19)

where $B(x)$ denote a two-sided standard Brownian motions. It is known [BG97, FQ15] that the Brownian motion (1.19) is quasi-stationary for the KPZ equation (1.5). That is, $\mathcal{H}_{\text{stat}}(t_0, \cdot) - \mathcal{H}_{\text{stat}}(t_0, 0) \xrightarrow{\text{law}} \sqrt{\rho(1 - \rho)} B(\cdot)$, for any $t_0 \in [0, \infty)$. This and the uniqueness of Hopf–Cole solutions implies that

$$\mathcal{H}_{\text{stat}}(t + t_0, x) - \mathcal{H}_{\text{stat}}(t_0, 0) \xrightarrow{\text{law}} \mathcal{H}_{\text{stat}}(t, x), \quad \text{as } C([0, \infty), C(\mathbb{R})) \text{-valued processes}$$

for any $t_0 > 0$. Utilizing (1.20), we show in Section 5.2 that the centered height process $(\mathcal{H}_{\text{stat}}(t, x) - \mathcal{H}_{\text{stat}}(t, 0))$ can in fact be extended to all values of $t > -\infty$.

**Proposition 1.5.** There exists a $C(\mathbb{R}, C(\mathbb{R}))$-valued process $\mathcal{K}(t, x)$ such that, for any fixed $t_0 \in \mathbb{R}$,

$$\mathcal{K}(t - t_0, x) \xrightarrow{\text{law}} \mathcal{H}_{\text{stat}}(t, x) - \mathcal{H}_{\text{stat}}(t_0, 0), \quad \text{as } C([0, \infty), C(\mathbb{R})) \text{-valued processes in } (t, x).$$

(1.21)

Note that in the above proposition $\mathcal{K}(t, x)$ is a process with $t \in \mathbb{R}$. Given this, the solution of $\mathcal{U}$ of the SBE is defined as

$$\mathcal{U} : C_c^\infty(\mathbb{R}^2) \rightarrow \mathbb{R}, \quad (\mathcal{U}, f) := - \int_{\mathbb{R}^2} \partial_x f(x, y) \mathcal{K}(y, x) dx dy.$$  

(1.22)

Given that $\mathcal{H}_{\text{stat}} \in C(\mathbb{R}_+ \times \mathbb{R})$, it is straightforward to check $\mathcal{U} \in C^{-1}(\mathbb{R}^2)$.

The following is our main result on symmetric 6V model:

**Theorem 1.6.** For fixed $u > 0$ and $\eta = \eta_\varepsilon = \frac{1}{2} \sqrt{\varepsilon}$, consider the symmetric 6V model with vertex weights given by Baxter’s (projective) parametrization (1.12). Fix further $v \in (0, 1)$, and let $h = h_\varepsilon$ be given by (1.13) with the choice of $-\eta = \eta_\varepsilon = \frac{1}{2} \sqrt{\varepsilon}$.

Consider the empirical distribution (1.15) of the stochastic Gibbs state $\mathcal{SG}(b_1, b_2; h_\varepsilon, v)$, where $b_1, b_2$ are given in terms of $u, \eta_\varepsilon$ through (1.10)–(1.12). As $\varepsilon \rightarrow 0$ (whereby $\Delta = \cosh(\eta_\varepsilon) \downarrow 1$), with $\mathcal{U}$ given as in (1.22), we have that

$$U_\varepsilon \Longrightarrow \mathcal{U} \quad \text{in } C^{-1}(\mathbb{R}^2).$$

**Remark 1.7.** In order to see the SBE limit here it is necessary to look along the characteristic (in the sense of Burger’s equation) direction. In (1.15), this is reflected in the slope $\mu_\varepsilon$ which is essentially given by the derivative of $h(v)$ evaluated at $v$.

**Proof sketch.** This result is proved in Section 5.2. Since the stochastic Gibbs states come from a suitably chosen stochastic 6V model, we can apply Theorem 1.1 to prove convergence. The convergence is for positive times, but using the stationarity, we can extend it easily to all time.

1.3. **KPZ equation as a limit of ASEP.** Stochastic Partial Differential Equations (SPDEs) describe the evolution of systems in the presence of random noise. The construction and approximation theory for non-linear SPDEs has attracted significant attention and enjoyed major breakthroughs in recent years (see, for instance, [BG97, Hai13, Hai14, GP17a, GJ14, GP17b]). Such equations are believed to describe the fluctuations of microscopic systems around their hydrodynamic limits.

The KPZ equation is a model for random growth processes, interacting particle systems, and directed polymers [Cor12, QS15]. Writing $\mathcal{H}(t, x)$ for the height at time $t \geq 0$ above $x \in \mathbb{R}$, the equation reads:

$$\partial_t \mathcal{H}(t, x) = \frac{\kappa}{2} \partial_x^2 \mathcal{H}(t, x) - \frac{\nu}{2} (\partial_x \mathcal{H}(t, x))^2 + \sqrt{D} \xi(t, x),$$

(1.23)

where $\xi(t, x)$ denotes the Gaussian space-time white noise, and $\kappa \neq 0 \in \mathbb{R}$ and $\nu, D > 0$ are constants measuring the strength of each term in (1.23).
Making sense of (1.23) is confounded by the non-linearity—solutions are rough enough that this does not make classical sense. The simplest, though indirect, approach is through the Hopf–Cole transform—one simply defines $H(t,x) = -\nu \kappa \log Z(t,x)$ where $Z$ solves the SHE (with multiplicative noise)$^6$:

$$\partial_t Z(t,x) = \frac{\nu}{2} \partial^2_x Z(t,x) + \frac{\kappa D}{\nu} \xi(t,x) Z(t,x).$$

(1.24)

There are two other definitions which have been introduced recently and yield equivalent solutions: energy solutions [GJ14, GP17a] and the regularity structures [Hai14]/paracontrolled distributions [GP17b] (these last notions are for periodic $x \in [0,1]$). See also renormalization group techniques in [Kup16].

How does the KPZ equation arise from microscopic systems? Fixing $(b,z) \in \mathbb{R}^2$ and letting (for the moment) $H_\varepsilon(t,x) := \varepsilon^b \mathcal{H}(\varepsilon^{-2} t, \varepsilon^{-1} x)$ one sees that $H_\varepsilon$ satisfies a version of (1.23) with scaled coefficients (see, for instance, [Qua11]). There are no choices for $(b,z)$ besides $(0,0)$ which leave the equation invariant. One may, however, simultaneously scale coefficients in (1.23) to compensate for the effects of the $(b,z)$-scaling. This is a proxy for understanding how discrete models may converge to (1.23) when one performs $(b,z)$-scaling while also scaling model parameters to effectively tune coefficients. This is called weak scaling, and significant efforts have sought to show weak KPZ universality, meaning that general classes of processes converge to (1.23) under such weak scaling.

Even though the focus of this work is on the 6V model, we focus for the moment on ASEP since it is a simpler process and allows us to cleanly identify the key challenge in proving the KPZ equation limit for the stochastic 6V model. The Asymmetric Simple Exclusion Process (ASEP) is a continuous-time particle system in which particles inhabit sites indexed by $\mathbb{Z}$ and jump left and right according to continuous time exponential clocks with rates $\ell \ge 0$ and $r \ge 0$ (fix $\ell \ge r$ and $\ell + r = 1$) subject to exclusion (jumps to occupied sites are suppressed). The ASEP height function $N_{\text{ASEP}}(t,x)$ is defined just as for the stochastic 6V model and has $1/0$ slopes entering occupied/vacant sites (see Figure 6). ASEP arises as a continuous time limit of the stochastic 6V model when $b_1 = \varepsilon \ell$, $b_2 = \varepsilon r$, time is scale to be $\varepsilon^{-1} t$ and particles are viewed in a moving frame with velocity $\varepsilon^{-1}$ (see [BCG16, Agg17]).

![Figure 6. ASEP particle configuration with the associated height function above it. Left jumps correspond to adding a rhombus and right jumps do the opposite.](image)

The ASEP was the first discrete space system proved to converge to the KPZ equation: [BG97] proved that for nearly stationary initial condition with density $\rho = \frac{1}{2}$ (Definition 4.4), under weak asymmetry scaling where $\ell - r = \sqrt{\varepsilon}$,

$$\sqrt{\varepsilon} \left( N_{\text{ASEP}}(\varepsilon^{-2} t, \varepsilon^{-1} x) - \frac{1}{2} \varepsilon^{-1} x - \frac{1}{4} \varepsilon^{-2} t \right) \xrightarrow{\varepsilon \to 0} \mathcal{H}(t,x),$$

(1.25)

as a space-time process. The starting point for this result was an observation in [Gar88] that ASEP admits a microscopic Hopf–Cole transform:

Setting $\tau = r/\ell$, and $Q(t,x) = \tau N_{\text{ASEP}}(t,x)$, $dQ(t,x) = L_{\tau,\ell}^1 Q(t,x) + Q(t,x) dM(t,x)$.

(1.26)

Here $L_{\tau,\ell}^1$ is the generator of a simple continuous time random walk with left and right jump rates given by $r$ and $\ell$ (note the exchange in left and right rates), and $dM(t,x)$ is a martingale with explicit quadratic variation (see Appendix A).

$^6$The positivity and well-posedness of (1.24) follows classical methods, see [Cor12, Qua11] for further details.
The convergence in (1.25) is shown not at the level of the height function, but rather its exponential, by showing that the above microscopic SHE (1.26) converges under the scalings in (1.25) to its continuum version (1.24). Given tightness of the exponential process (which follows from detailed estimates on the random walk transition probability), the convergence to (1.24) is achieved via martingale problems (see Section 5.1). That is, the SHE is uniquely characterized by a linear and quadratic martingale problem which, respectively, identify the drift and the noise.

Convergence of the linear problem follows easily by approximating $L^1_{r,\ell}$ with the Laplacian. The convergence of the quadratic problem is rather involved and ultimately boils down to showing that

\[
\nabla Q(t, x + 1) \nabla Q(t, x) \text{ self-averages in } t.
\]

Such expressions arise from the quadratic variation of the $dM(t, x)$. Here $(\nabla f)(x) = f(x + 1) - f(x)$. In (1.27), “self-averaging” refers to a phenomena where the moments of the integral of the expression over a long time interval of length $O(\varepsilon^{-2})$ will vanish as $\varepsilon \to 0$, see (A.4). For ASEP, this phenomena is explained more in Appendix A, in particular, see (A.10). In the case of stochastic six vertex model, the precise statement of “self-averaging” is given in Proposition 5.3. See Remark 5.5.

The statement (1.27) is natural from the perspective of hydrodynamic limit theory. Indeed, [Qua11] demonstrated how the replacement lemma (i.e., local equilibrium) can be used to prove (1.27). The proof in [BG97] proceeded through a different, iterative scheme. Roughly speaking, it seeks to close a sequences of inequalities starting from (1.26). Crucial to the closing of inequalities (and hence to this scheme as a whole) is a non-trivial summation identity for the random walk transition probability.

1.4. Markov duality method. The Markov duality method that we introduce and develop in this article provides a new way to obtain optimal control over the conditional expectation of the expression in (1.27) (and related terms). More importantly, the method also applies to the general class of discrete time stochastic vertex models introduced in [CP16]—in particular, to the stochastic 6V model. Presently, none of the other methods used for KPZ equation convergence results seem to be applicable to the stochastic 6V model. The quadratic variation for the stochastic 6V model takes a more complicated form (as in (4.9)–(4.10)) than that of ASEP. This being the case, the approach of [BG97] for closing inequalities does not appear to generalize.

Hydrodynamic theory methods like energy solutions [GJ14, GP17a] or the approach to self-averaging given in [Qua11] relies heavily upon continuous time Markov process methods. In fact, hydrodynamic theory for discrete-time processes is not particularly well-developed as many of the basic tools that work in continuous time fail to generalize. The model considered here is updated sequentially in discrete time (see Section 2.1), so, from the perspective of Markov chains, the update of each particle depends on configurations of infinitely many other particles. This intricate feature further impedes generalizing methods of continuous time Markov process and hydrodynamic limit theory.

Other methods like regularity structures [Hai14], paraccontrolled distributions [GP17b] and renormalization group methods [Kup16] have not yet been sufficiently developed to deal with processes that are driven by a process-dependent noise. More precisely, this refers to the fact that the martingale in (1.26) have a $Q$-dependent quadratic variation. Additionally, those methods are presently restricted to periodic boundary conditions. The Markov duality method works for discrete time processes with general initial condition on the full line. Its shortcoming is that it requires the existence of (at least $k = 1, 2$) Markov dualities like below. See Section 1.5.3 for further discussion on literature related to KPZ equation convergence results.

The microscopic Hopf–Cole transform [Gar88] is the $k = 1$ case of ASEP Markov duality [BCS14]:

\[
\text{For } k \geq 1 \text{ and } \bar{x} = (x_1 < \cdots < x_k) \in \mathbb{Z}^k, \quad \frac{d}{dt} \mathbb{E} \left[ \prod_{i=1}^{k} Q(t, x_i) \right] = L^k_{r,\ell} \mathbb{E} \left[ \prod_{i=1}^{k} Q(t, x_i) \right].
\]

Here $\mathbb{E}$ is the expectation of the ASEP height process, and $L^k_{r,\ell}$ acts on $\bar{x}$ as the space-reversed generator of $k$-particle ASEP with locations $\bar{x}$. For $k = 1$, removing expectations yields (1.26). Replacing $Q(t, x)$ by its discrete derivative $\bar{Q}(t, x) := Q(t, x) - Q(t, x - 1)$ yields a similar duality due to [Sch97].
The Markov duality method uses the $Q$ and $\tilde{Q}$ duality for $k = 2$ to prove convergence of the discrete quadratic martingale problem to that of the SHE. For example, the key term in (1.27) can be rewritten as $\tilde{Q}(t, x + 1)\tilde{Q}(t, x)$ and duality shows that for $x_1 < x_2$ and $t > s$,

$$
\mathbb{E}[\tilde{Q}(t, x_1)\tilde{Q}(t, x_2) | \mathscr{F}(s)] = \sum_{y_1 < y_2} p_{t-s}(\vec{x} \rightarrow \vec{y})\tilde{Q}(s, y_1)\tilde{Q}(s, y_2)
$$

where $p_{t-s}(\vec{x} \rightarrow \vec{y})$ is the two-particle space-reversed ASEP transition probability from $\vec{x} = (x_1, x_2)$ to $\vec{y} = (y_1, y_2)$ in time $t - s$. Once in this form, the discrete differentiation can be transferred to the transition probabilities and the proof of self-averaging reduces to fine estimates on such derivatives of the two-particle heat kernel. In essence, duality turns a hydrodynamic problem (involving the local equilibration in the collective behavior of many particles) into a diffusive problem (involving the fluctuations of a handful of particles).

The Bethe ansatz (for ASEP, see [TW08, TW11] or Appendix A) provides a means to extract very precise estimates for finite particle system transition probabilities. We also remark that whereas previous results on the KPZ equation limit for ASEP have assumed density near $1/2$, the duality method works just as well for any density and for any moving frame in the rarefaction fan.

The major downside of our Markov duality method is that such dualities like (1.28) do not hold for generic systems and their occurrence is often due to algebraic structures which are not very flexible to perturbations (see Section 1.5.4 for further discussion). However, it was shown in [CP16, Kua17] that the stochastic 6V model (in fact its higher spin generalizations too) enjoy the same sort of duality as in (1.28) (see Section 3). We see the main technical accomplishment of this paper to be the development and implementation of this new duality method to control the quadratic martingale.

1.5. Further literature.

1.5.1. Symmetric six vertex model. Introduced in 1935 by Pauling [Pau35] as a model for 2D ice and then in its general form in 1941 by Slater [Sla41] to model potassium dihydrogen phosphate, the symmetric 6V model has found many applications across physics and mathematics as well as prompted the discovery of new algebraic structures such as quantum groups and new symmetric functions. The 6V model was exactly solved in Lieb’s breakthrough work [Lie67] which was the first time the ideas of Bethe ansatz were applied to a statistical mechanics model. This work immediately (e.g. [Sut67, YY66]) opened up the field to many important developments including coordinate/algebraic Bethe ansatz, quantum groups, domain-wall boundary conditions, connections to symmetric functions—see the reviews/books [Bax89, Nol92, Fad96, KBI93, JM93, Res10, BL14, Gau14, Koz15, BP15a]). The results of this paper probe the behavior of the 6V model as $\Delta \searrow 1$. There are many other interesting phase transitions in the 6V model—for instance when $a = b$ (i.e., the Fierz, or F model—studied first in [Rys63]), as $c \rightarrow 2a$ (or equivalently $\Delta \rightarrow -1$) there is a remarkable infinite order phase transition in the free energy (see [LW72] for further information).

1.5.2. Stochastic six vertex model. Study of this special case of the asymmetric 6V model was initiated in 1992 by Gwa and Spohn [GS92]. However, the relation between the conical points and the stochastic 6V model was conjectured in 1995 by Bukman and Shore [BS95], though there was earlier discussion about the existence of these conical points in [JS84].

The study of the stochastic 6V model was recently reinitiated in [BCG16] wherein they proved the prediction from [GS92] that the stochastic 6V model was in the KPZ universality class. This was demonstrated at the level of convergence of the one-point distribution (to the GUE Tracy-Widom distribution) for a special boundary condition on the first quadrant with no lines coming from the $y$-axis and no anti-lines coming from the $x$-axis (i.e., step initial condition). This result did not involve any special weak scaling, hence convergence to the GUE Tracy-Widom distribution and not the one-point distribution for the KPZ equation. [AB16, Agg16] then extended the one-point convergence to other initial condition, including the stationary case (i.e., the stochastic Gibbs state).
In that case, [Agg16] computed an exact one-point formula and proved convergence to the stationary KPZ distribution (the Baik-Rains distribution) in the characteristic direction. In principle one could take the weakly asymmetric scaling limit of that formula and match it with the formula for the stationary KPZ equation proved in [BCFV15] (though that would only prove a one-point convergence result, as opposed to the process level result herein). In a similar spirit, [BO17] showed that under weakly asymmetric scaling, one point distribution of the stochastic 6V model converges to that of the KPZ equation (see also [BG16]). The scaling considered in [BO17] is different than here—essentially they also tune $b_1, b_2 \to 1$ (herein they converge to a value strictly less than 1). It is quite likely that our approach could apply under the scaling used in [BO17], though we do not pursue that here.

[BBCW17] recently studied a half-space version of the stochastic 6V model and demonstrated that its one-point asymptotics match the prediction from other models in the KPZ universality class. It may be possible to adapt methods from [CS16] to connect the half-space stochastic 6V model to the KPZ equation under weakly asymmetric scaling, though we do not pursue that here.

The stochastic 6V model admits a higher spin analog wherein more than one line can move along each edge in $\mathbb{Z}^2$ (i.e., multiple particles can occupy the same site, or move together). These models have recently been studied in [CP16, BP16] and admit some similar asymptotics as the stochastic 6V model. The Markov duality method should also apply to these models (as they all enjoy the same duality as the stochastic 6V model).

There are other limits of the stochastic 6V model besides the KPZ equation and ASEP, e.g. Hall-Littlewood PushTASEP [BP15b, BCG16, BBW16, Gho17] and Brownian motions with oblique reflection [SS15]. Another limit considered in parallel to the present paper is in the work of [BG18]. They consider a different type of limit in which $b_1$ and $b_2$ both tend to 1. [BG18] proves a law of large numbers and some Gaussian fluctuation results under this scaling. Moreover, they conjecture (and prove in a certain low density regime) convergence to the stochastic telegraph equation—a linear hyperbolic SPDE driven by additive space-time white noise. It would be natural to try to prove this conjecture using the Markov duality method developed herein, and to try to fill-out the scaling limits which sit between our results and those of [BG18].

1.5.3. Kardar-Parisi-Zhang equation. The KPZ equation (1.23) was introduced in 1986 by Kardar, Parisi and Zhang [KPZ86]. In 1995 Bertini and Cancrini [BC95] provided the first justification for the Hopf–Cole solution to the KPZ equation. Bertini and Giacomin [BG97] soon after proved the first discrete convergence result (for ASEP) to the KPZ equation. This converge result has recently been extended in works of [ACQ11, Qua11]. [DT16] extended the convergence result to certain non-nearest-neighbor exclusion processes which do not satisfy an exact microscopic Hopf–Cole transform.

The first convergence result to the KPZ equation for a discrete time particle system was recently proved in [CT17]. The systems considered therein were infinite spin versions of the higher spin vertex models studied in [CP16]. The scaling there was different than the weakly asymmetric scaling here. In particular, the number of particles per site diverges under their scaling. This simplified the study of the quadratic martingale problem considerably. In particular, due to the divergence of the number of particles per site, the key bound which plays a central role in this work and in that of [BG97] becomes straightforward and does not require any sort of trick to control. Other recent KPZ equation convergence works, following the style of [BG97], have included the ASEP-($q,j$) [CST16], Hall-Littlewood PushTASEP [Gho17], and open ASEP [CS16, Par17, Lab17].

The energy solution method for KPZ equation convergence was initiated in the work of the Jara and Gonçalves [GJ10] (cf. [Ass13]). Initially this approach only provided tightness and it was not known whether energy solutions were unique. Uniqueness (and hence the identification with the Hopf–Cole solution) was proved in [GP17a]. This approach has been applied to prove that a wide variety of particle systems converge to the KPZ equation, see [GJ14, GJS15, FGS16, GJ13, GJ16, GPS17]. Those results require stationary initial condition and the method of proof relies heavily upon having well-developed hydrodynamic theory estimates available.
Regularity structures and paracontrolled distributions provide another route to prove convergence results to the KPZ equation. These notions of solutions were introduced by Hairer [Hai13, Hai14] and Gubinelli and Perkowski [GP17b] (cf. [GIP15]), and have since been used to prove convergence for some space-time regularized versions of the equation [HS17, HQ15, DGP17]. [HM15, CM16, EH17] has recently developed a discrete space-time version of regularity structures, which may prove useful in demonstrating convergence of various discrete processes to the KPZ equation. It is worth noting that presently due to technical challenges involved with going to the full line, these works on the KPZ equation using regularity structures or paracontrolled distributions are restricted to periodic spatial coordinates. Finally, there is also a renormalization group method which has been applied to the KPZ equation in [Kup16], though it also is also restricted to a periodic setting.

1.5.4. Markov duality. Markov dualities are extremely useful notions within probability. An early example of a self-duality was for the simple symmetric exclusion processes (SSEP) [Lig05] where it played a key role in proving that only extremal, translation invariant, ergodic invariant distributions of SSEP on $\mathbb{Z}^d$ are the Bernoulli product distributions. Whereas that duality applied, in fact to SSEP on any graphs, asymmetric particle system dualities seem to be much more rigid and dependant upon algebraic structures only present for one spatial dimension. The first such example was found in [Sch97] where the $\tilde{Q}$ version of the duality in (1.28) was first discovered based on the affine quantum group $U_q[sl_2]$ symmetry of ASEP (see also [SS94]). The self duality of ASEP has played an important role in demonstrating that ASEP belongs to the KPZ universality class (see, for instance, [BCS14, Cor14] and the reference therein).

Recently, a generalized version of ASEP (called ASEP-(q,j)) which enjoys a generalization of the ASEP self-duality was introduced in [CGRS16] based on higher spin representations of $U_q[sl_2]$. Self duality has been also proved [BS15, Kua16] in certain multi-species versions of ASEP using higher rank quantum group symmetries in the spirit of [CGRS16].

The stochastic 6V model (as well as higher spin vertex models) duality was discovered and proved in [CP16] (see [Kua17] for an algebraic proof of some of the dualities from [CP16] based on properties of the $R$ matrix and quantum group co-product). It is this duality for the stochastic 6V model that plays a pivotal role in this paper and is discussed in more detail in Section 3.

Outline. In Section 2 we give a brief discussion the stochastic and symmetric 6V models. This includes the definition of the stochastic model with bi-infinite configurations, the construction of stochastic Gibbs states, and how they fit into the stochastic and symmetric models. Then, to setup the premise of our analysis, in Section 3 we recall the self-duality of the stochastic model, and in Section 4, we introduce the microscopic Hopf–Cole transform. Specifically, once the transform is introduced, Theorem 1.1 on the convergence of the stochastic model to KPZ naturally translates into the corresponding, equivalent statement in terms of convergence toward the SHE, Theorem 1.1*. In Section 5, we settle the main results Theorems 1.1* and 1.6 while assuming Proposition 5.3. The latter is a statement on self-averaging of the relevant quadratic variation. Proving Proposition 5.3 makes up the core of our analysis. In Section 6, we perform steepest-decent-type analysis on the given contour integral formula for the semigroup. The analysis produces estimates on the semigroup and its gradients, jointly over all relevant points in spacetime. Then, in Section 7, we incooperate these estimates into the stochastic model via duality and prove Proposition 5.3.

To make connection with ASEP, in Appendix A, we briefly recall its Hopf–Cole transform and the structure of the relevant martingale. Given this setup, we explain how, for ASEP, our duality approach could serve as an alternative to the approach of [BG97] for controlling the quadratic variation.

Acknowledgements. The authors wish to thank J. Dubedat, P. di Francesco, and T. Spencer for discussions about the disordered phase of the 6V model and A. Aggarwal for discussions about his work on this subject. I. Corwin was partially supported by the NSF through DMS:1664650, and the Packard Foundation through a Packard Fellowship for Science and Engineering. H. Shen is partially supported by the NSF through DMS:1712684. L-C. Tsai is partially supported by the NSF through
2. Stochastic and symmetric six vertex models

2.1. Stochastic six vertex model as a particle system and its height function.

2.1.1. Defining the left-finite process. In [BCG16, Section 2], the stochastic 6V model is defined on
the first quadrant $\mathbb{Z}^2_{\geq 0}$ by first specifying the configuration of lines coming from the bottom and left
boundary and then inductively filling in the quadrant. Specifically, once it is determined whether
lines are entering a given vertex from below and from the left, the stochastic weights in Figure 1
specify the probability according to which one chooses (independently over vertices) the outgoing line
configuration. Proceeding recursively in this manner defines the stochastic 6V model distribution on
the entire quadrant (for the given boundary condition).

If we restrict ourselves to boundary conditions where there are no lines coming from the left bound-
ary, then the lines from the bottom can be seen as the trajectories of a discrete time sequential update
exclusion-type particle system. Under this interpretation, time is measured by the
$y$-axis, and the particles are identified with vertical lines and their moves are identified with the horizontal lines. We
define below this particle system and allow particles to start anywhere on $\mathbb{Z}$ as long as there is always
a left-most particle. After doing that, we explain how to extend our definition to two-sided infinite
particle configurations (as will be necessary to state our main results).

**Definition 2.1.** For $w \in \mathbb{Z}$ define the space of left-finite ordered particle configurations with left-most
label $w$ to be

$$\Xi_{\geq w} := \{ \vec{x} = (\ldots = x_{w - 1} < x_w < x_{w + 1} < \ldots) : x_i \in \mathbb{Z} \cup \{-\infty\}, \text{ for } i \in \mathbb{Z}_{\geq w} \}. \quad (2.1)$$

Here $x_i$ represents the location of the particle labeled $i$. Notice that we have placed a virtual particle
$x_{w - 1}$ at $-\infty$. We allow $\Xi_{\geq w}$ to contain configurations with infinitely many particles as well as finitely
many particles. In the later case, there will be some $w'$ such that $x_i = +\infty$ for all $i > w'$.

Having defined our state space $\Xi_{\geq w}$ we proceed to describe the discrete time Markov chain $(\vec{x}(t))_{t \in \mathbb{Z}_{\geq 0}}$
where $\vec{x}(t) \in \Xi_{\geq w}$ for each $t$. Fix $b_1, b_2 \in (0, 1)$ and let

$$\tau = b_2/b_1 \in (0, 1)$$

denote their ratio. We will assume that $b_2 < b_1$ so that $\tau \in (0, 1)$ throughout. The algebraic results
do not generally depend on this, but when we perform asymptotics we will use this asymmetry. Given
$\vec{x}(t)$, we choose $\vec{x}(t + 1)$ according to the following sequential (left to right) procedure. For each $i \geq w$
(starting with $i = w$ and progressing sequentially to $i = w + 1$, $i = w + 2$, etc), choose $x_i(t + 1)$ so
that (recall that $x_j(t + 1)$ for $j < i$ have already been updated)

(a) if $x_{i - 1}(t + 1) < x_i(t)$, then

$$\mathbb{P}(x_i(t + 1) = x_i(t) + j) = \begin{cases} b_1, & \text{if } j = 0; \\ (1 - b_1)(1 - b_2)b_2^{j - 1}, & \text{if } 1 \leq j \leq x_{i + 1}(t) - x_i(t) - 1; \\ (1 - b_1)b_2^{j - 1}, & \text{if } j = x_{i + 1}(t) - x_i(t); \\ 0, & \text{otherwise}; \end{cases}$$

(b) if $x_{i - 1}(t + 1) = x_i(t)$, then

$$\mathbb{P}(x_i(t + 1) = x_i(t) + j) = \begin{cases} (1 - b_2)b_2^{j - 1}, & \text{if } 1 \leq j < x_{i + 1}(t) - x_i(t); \\ b_2^{j - 1}, & \text{if } j = x_{i + 1}(t) - x_i(t); \\ 0, & \text{otherwise}. \end{cases}$$

Since we have assumed the convention $x_{w - 1}(t) = -\infty$, the particle $x_w$ is always updated by rule (a).
In words, sequentially (starting with particle \( x_n \)) each particle \( x_i \) wakes up and moves one to the right with probability \( 1 - b_1 \). Once awake, the particle continues moving right with probability \( b_2 \) for each step. If \( x_i \) eventually moves into the location occupied already by \( x_{i+1} \), then \( x_i \) stops moving and stays put, while \( x_{i+1} \) is forced to wake up and move one to the right (after which it continues with the probability \( b_2 \) rule as above). Once the particle \( x_i \) stops, that is its new position \( x_i(t+1) \).

To each state \( \vec{x}(t) \in \mathbb{X}_{\geq w} \), we may associate occupation variables and a height function as follows: Define the \( \{0,1\} \)-valued occupation variables
\[
\eta(t,y) := 1\{x_n(t)=y \text{ for some } n \in \mathbb{Z}_{\geq w}\}
\]  
(2.2)
where the indicator function is 1 if the site \( y \) is occupied by a particle at time \( t \), and 0 otherwise. Likewise, define the height function
\[
N(t,y) := N_y(\vec{x}(t)) = N_0(\vec{x}(0)).
\] (We have centered \( N \) so that \( N(0,0) = 0 \).) In the above definition, we have used the following notation. For \( y \in \mathbb{Z} \), \( N_y : \mathbb{X}_{\geq w} \to \mathbb{Z}_{\geq w-1} \) and (for later use) \( \eta_y : \mathbb{X}_{\geq w} \to \{0,1\} \) are defined by\(^7\)
\[
N_y(\vec{x}) := \max \left\{ n : x_n \leq y \right\} \quad \text{and} \quad \eta_y(\vec{x}) := N_y(\vec{x}) - N_{y-1}(\vec{x}).
\] (2.3)
In particular, one has \( N_{\eta_y}(\vec{x}) = n \), and \( N_y(\vec{x}) - w = 1 \) if \( y \) is to the left of all particles in \( \vec{x} \). It follows that \( N(t,y) - N(t,y-1) = \eta(t,y) \), so that the space-time level-lines of \( N(t,y) \) correspond with the trajectories of \( \vec{x}(t) \). See Figure 3 for an illustration.

Under the dynamics described above in Definition 2.1, the height function \( t \mapsto N(\cdot,t) \) evolves in \( t \) as a Markov chain\(^8\). We may describe its transitions explicitly.

**Definition 2.2.** Let \( X \sim \text{Ber} (\rho) \) mean that \( X \) is a Bernoulli random variable taking values in \( \{0,1\} \) with \( \mathbb{P}(X = 1) = \rho \). Let \( \{B(t,y,\eta), B'(t,y;\eta) : t \in \mathbb{Z}_{\geq 0}, y \in \mathbb{Z}, \eta \in \{0,1\}\} \) denote a countable collection of independent Bernoulli variables, with \( B(t,y,\eta) \sim \text{Ber}(1-b_1^\eta) \) and \( B'(t,y,\eta) \sim \text{Ber}(b_2^{1-\eta}) \).

Using the Bernoulli random variables from the above definition we see that
\[
N(t+1,y) \overset{\text{law}}{=} \begin{cases} N(t,y) - B(t,y;\eta(t,y)), & \text{if } N(t+1,y-1) = N(t,y-1) - 1, \\ N(t,y) - B'(t,y;\eta(t,y)), & \text{if } N(t+1,y-1) = N(t,y-1). \end{cases}
\] (2.4)

2.1.2. Defining the bi-infinite process. Since the stochastic 6V model is sequentially updated, it is not a priori clear how to define it when there are infinitely many particles to the left and right of the origin. [CT17] showed that it is possible to restate the stochastic 6V model in terms of a parallel update rule which readily admits a bi-infinite extension. We restate this result below as well as include a convergence result showing how to approximate the bi-infinite process with left-finite ones.

**Definition 2.3.** Denote the space of bi-infinite order particle configurations by
\[
\mathbb{X} = \{ \cdots < x_{-1} < x_0 < x_1 < \cdots : x_i \in \mathbb{Z} \cup \{-\infty, +\infty\} \}.
\]
Notice that we have included left and right finite configurations in \( \mathbb{X} \) by having imaginary particles at \( -\infty \) or \( +\infty \).

**Lemma 2.4.** Consider a bi-infinite configuration \( \vec{x} \in \mathbb{X} \) and let \( \vec{x}_{\geq w} = (x_i : i \geq w) \in \mathbb{X}_{\geq w} \) for any \( w \in \mathbb{Z} \). Let \( N(0,y) = N_y(\vec{x}) = N_0(\vec{x}) \) and \( N(0,y) = N_y(\vec{x}_{\geq w}(t)) = N_0(\vec{x}_{\geq w}(0)) \) where \( \vec{x}_{\geq w}(t) \) is the stochastic 6V Markov chain at time \( t \) with initial condition \( \vec{x}_{\geq w} \). Likewise, let \( \eta(0,y) = N(0,y) - \eta_y(\vec{x}(t)) \). We distinguish the notation \( \eta(t,y) \) as a process and the notation \( \eta_y \) as a function on particle configurations \( \vec{x} \) merely for convenience.

\(^7\)Note that \( \eta(t,y) = \eta_y(\vec{x}(t)) \). \(^8\)The notation \( t \mapsto N(\cdot,t) \) means the function which takes \( t \) to height functions (defined on the full line). By convention, such height functions will have height \( w - 1 \) asymptotically as \( y \) goes to \( -\infty \). The Markov chain is defined on that state-space.
Consider the stochastic 6V model with parameters \( \alpha, \beta, \gamma, \nu \). Let \( B(t, y, \eta) \) and \( B'(t, y, \eta) \) be as in Definition 2.5. Then for any \( t \in \mathbb{Z}_{\geq 0} \) and \( z, y \in \mathbb{Z} \), we have that

\[
N^w(t, y) - N^w(t + 1, y) = \sum_{y' = -\infty}^{-y} \prod_{z} \left( B(t, z, \eta^w(t, y)) - B'(t, z, \eta^w(t, y)) \right) B'(t, y', \eta^w(t, y)).
\]

Furthermore for any \( y \in \mathbb{Z} \), as \( w \to -\infty \), \( N^w(t, y) \to N(t, y) \) in \( L^p \) for all \( p \geq 1 \) and in probability. The limit \( N(t, y) \) is specified inductively in \( t \) (with \( t = 0 \) as the base case) by the (convergent) relation

\[
N(t, y) - N(t + 1, y) = \sum_{y' = -\infty}^{-y} \prod_{z} \left( B(t, z, \eta(t, y)) - B'(t, z, \eta(t, y)) \right) B'(t, y', \eta(t, y))
\]

and hence satisfies (2.4). From \( N(t, y) \) we define \( \eta(t, y) = N(t, y) - N(t, y - 1) \), and we may uniquely define \( \vec{x}(t) \) so that the particles of \( \vec{x}(t) \) track the level lines of \( N(t, y) \).

Proof. The result is a special case of the statement and proof of [CT17, Lemma 2.3 and Remark 2.5]. In [CT17] the authors consider a more general higher-spin version of the stochastic 6V model [CP16] with arbitrary horizontal spin \( J \) as well as parameters \( \alpha, \beta, \gamma, \nu \). Our stochastic 6V model corresponds with taking \( J = 1 \) (spin-1/2), \( \nu = 1/q = \tau \), and matching \( b_1 = \frac{1 + q}{1 + \alpha} \) and \( b_2 = \frac{\alpha + q}{1 + \alpha} \).

Unless specified otherwise, the stochastic 6V model now means the bi-infinite version of Lemma 2.4.

2.1.3. Stationary initial condition. A key aspect of studying an interacting particle system is to identify its stationary distributions, in particular those which are translation invariant and ergodic. These distributions are the first step towards identifying the hydrodynamic equations and non-universal constants which arise in the KPZ scaling theory (see, for instance, [Spo14] and references therein). For ASEP these are characterized by one parameter \( \rho \in [0, 1] \) and given by product distribution \( \text{Ber}(\rho) \) on occupation variables. The same distributions turn out to be stationary of the stochastic 6V model. In fact, as shown in [Agg16], the stationary stochastic 6V model enjoys a sort of stationarity along down-right paths very much akin to that of certain exactly solvable directed polymer and last passage percolation models (see, for instance, [Sep12]).

Definition 2.5. Consider the stochastic 6V model with parameters \( b_1, b_2 \). Choose \( (h, v) \in [0, 1]^2 \) such that \( \frac{1}{1 - h} (1 - b_2) = \frac{h}{1 - h} (1 - b_1) \). The stationary stochastic 6V model on the first quadrant is defined relative to \( (h, v) \) by specifying that on the \( y \)-axis (\( x \)-axis) horizontal (vertical) lines enter from the boundary independently with probabilities \( h \) (\( v \)).

Lemma 2.6. Consider the stationary stochastic 6V model on the first quadrant from Definition 2.5. Then, along any fixed down-right lattice path in the first quadrant (i.e., a collection of vertices in \( \mathbb{Z}^2_{\geq 0} \) so that each vertex follows the previous one by adding \((1, 0) \) or \((0, -1) \) to its coordinates) the sequence of incoming line occupancy variables (i.e., whether a horizontal or vertical line enters the vertex along the path) are independent and incoming horizontal lines are present with probability \( h \) while incoming vertical lines are present with probability \( v \). Consequently, we can define the stationary stochastic 6V model on all of \( \mathbb{Z}^2 \) by taking the distributional limit as \( n \to \infty \) of the model on the first quadrant with the origin shifted to \((-n, -n)\). We refer to this distribution (of vertex configurations on \( \mathbb{Z}^2 \)) as the stochastic Gibbs states with densities \((h, v)\), and denote it by \( \mathcal{SG}(b_1, b_2; h, v) \).

Proof. This is the content of [Agg16, Lemma A.2].

The distribution \( \mathcal{SG}(b_1, b_2; h, v) \) does not treat the \( x \)-axis and \( y \)-axis directions differently. In terms of the particle process interpretation for the stochastic 6V model, this stationary distribution corresponds to starting with particles independently at each site of \( \mathbb{Z} \) with probability \( v \). The parameter \( h = h(v) \) then corresponds to the probability that a particle crosses a given vertical column at a given time, and the stationarity says that these events are all independent. The function \( h(v) \) is called the flux.
2.2. Stochastic Gibbs states for the symmetric six vertex model. As discussed in the introduction, the stochastic Gibbs states constructed in Lemma 2.6 are Gibbs states for a symmetric 6V model in the ferroelectric phase with parameters matched accordingly.

**Proposition 2.7.** Consider positive \((a, b, c)\) such that \(a > b + c\) and such that \(\Delta > 1\) (recall \(\Delta\) from (1.9)). Let \(b_1, b_2\) be given as in (1.10), and \((h, v) \in [0, 1]^2\) satisfy \(\frac{v}{1-v}(1-b_2) = \frac{h}{1-h}(1-b_1)\). Then, the stationary stochastic 6V distribution \(SG(b_1, b_2; h, v)\) from Lemma 2.6 is a extremal, translation invariant, ergodic infinite volume Gibbs state for the symmetric 6V model on \(\mathbb{Z}^2\) with weights \((a, b, c)\), and \((h, v)\) gives the density of horizontal and vertical lines under this Gibbs state.

**Proof.** A version of this result seems to have been first observed in [BS95]. More recently, it appeared in [RS16]; [Agg16, Proposition A.3] provides a proof. □

### 3. Self duality for stochastic six vertex model

The Markov duality method we introduce in this paper for showing convergence of the stochastic 6V model to the KPZ equation relies upon the model’s self-duality (in particular the one and two-particle duality), which we present in this section. This result was first proved for the stochastic 6V model with left-finite initial condition in [CP16]. We recall that result first, and then extend it by approximation to the bi-infinite stochastic 6V model defined in Lemma 2.4.

Let us first recall the general definition of Markov duality.

**Definition 3.1.** Given two Markov chains (in discrete time) or processes (in continuous time) \(x(t) \in X\) and \(y(t) \in Y\), we say \(x(t)\) and \(y(t)\) are dual with respect to a duality function \(H : X \times Y \to \mathbb{R}\) if for all \(x \in X\), \(y \in Y\) and \(t \geq 0\)

\[
\mathbb{E}^x[H(x(t), y)] = \mathbb{E}^y[H(x, y(t))].
\]

Here, \(\mathbb{E}^x\) denotes the expectation when the process \(x(t)\) has been started with the initial condition \(x(0) = x\), and \(\mathbb{E}^y\) likewise for the \(y\) variables.

Our stochastic 6V self duality theorem is actually a duality between the stochastic 6V model and its \(k\)-particle space reversal \((k \geq 1\) is arbitrary), which we define now.

**Definition 3.2.** Let \(\mathcal{Y}^k = \{(y_1 < \cdots < y_k) \in \mathbb{Z}^k\}\) denote the state space of ordered \(k\)-particle configurations (sometimes called a discrete Weyl chamber). The reversed stochastic 6V (or \(\tilde{S}6V\)) model with \(k\)-particles is the Markov chain \(\tilde{y}(t) = (y_1(t) < \cdots < y_k(t)) \in \mathcal{Y}^k\) defined such that \(-\tilde{y}(t) := (-y_k(t) < \cdots < -y_1(t)) \in \mathcal{Y}^k\) evolves according to the stochastic 6V dynamics given in Definition 2.1. For \(\tilde{x}, \tilde{y} \in \mathcal{Y}^k\), let \(\mathbb{P}_{\tilde{S}6V}(\tilde{x} \to \tilde{y}; t)\) denote the transition probability that the reversed stochastic 6V started from \(\tilde{y}(0) = \tilde{x}\) has \(\tilde{y}(t) = \tilde{y}\). Likewise, we let \(\mathbb{P}_{\tilde{S}6V}^\tau(\tilde{x} \to \tilde{y}; t)\) denote the transition probability that the (usual) stochastic 6V started from \(\tilde{y}(0) = \tilde{x}\) has \(\tilde{y}(t) = \tilde{y}\).

**Proposition 3.3.** Fix \(k \in \mathbb{Z}_{\geq 1}\), \(w \in \mathbb{Z}\) and parameters \(b_1, b_2 \in (0, 1)\) with \(b_2 < b_1\) (and recall that \(\tau = b_2/b_1\)). Let \(\tilde{x}(t) \in X_{\geq w}\) denote the stochastic 6V model with left-finite configurations (recall Definition 2.1, as well as the notation \(N_y(\tilde{x})\) and \(\eta_y(\tilde{x})\) defined therein) and let \(\tilde{y}(t) \in \mathcal{Y}^k\) denote the (reversed) \(\tilde{S}6V\) model with \(k\)-particles (Definition 3.2). Then \(\tilde{x}(t)\) and \(\tilde{y}(t)\) are dual with respect to the following two duality functions (recall Definition 3.1)

\[
H(\tilde{x}, \tilde{y}) := \prod_{i=1}^{k} \tau^{N_{\eta_y}(\tilde{x})}, \quad \text{and} \quad \tilde{H}(\tilde{x}, \tilde{y}) := \prod_{i=1}^{k} \eta_{\tau^{-(i+1)}}(\tilde{x}) \tau^{-N_{\eta_y}(\tilde{x})}.
\]

**Proof.** This is a special case of the dualities proved for the higher spin stochastic vertex models in [CP16, Theorem 2.23] (see also Section 5.5 therein). Note that the corresponding duality function \(\tilde{H}\) (called \(\tilde{G}_n(\tilde{g}, \tilde{n})\) therein) takes a slight different form here. Under current notation, the duality function in [CP16, Theorem 2.22] corresponds to \(\tilde{H}'(\tilde{x}, \tilde{y}) := \prod_{i=1}^{k} \eta_y(\tilde{x}) \tau^{N_{\eta_y}(\tilde{x})}\). One readily sees that

\[
\tilde{H}(\tilde{x}, (y_1, \ldots, y_k)) = \tau^k \tilde{H}'(\tilde{x}, (y_1 + 1, \ldots, y_k + 1)),
\]
so the duality of the latter readily implies that of the former.

For our applications, we need to extend this duality to the bi-infinite stochastic 6V model. This is accomplished by appealing to the approximation result given in Lemma 2.4. Let \((\mathcal{F}(t))_{t \in \mathbb{Z}_{\geq 0}}\) denote the canonical filtration of the stochastic 6V model.

**Corollary 3.4.** Fix \(k \in \mathbb{Z}_{\geq 1}, b_1, b_2 \in (0, 1)\) with \(b_2 < b_1\), and let \(\tau = b_2/b_1\). The result of Proposition 3.3 also hold for the bi-infinite stochastic 6V model \(\bar{x}(t) \in \mathbb{X}\). In particular, letting \(N(t, y)\) denote the height function associated in Lemma 2.4 to \(\mathcal{F}(t)\), and recall the reversed stochastic 6V model transition probability \(\mathbb{P}_{S\bar{6}V}\) from Definition 3.2, this implies that

\[
\mathbb{E}\left[\prod_{i=1}^{k} \tau^{N(t+s,y_i)} \bigg| \mathcal{F}(s)\right] = \sum_{\vec{y} \in \mathbb{Y}^k} \mathbb{P}_{S\bar{6}V}(\vec{y} \rightarrow \vec{y}'; t) \prod_{i=1}^{k} \tau^{N(s,y'_i)}
\]

\[
= \sum_{\vec{y} \in \mathbb{Y}^k} \mathbb{P}_{S\bar{6}V}(\vec{y} \rightarrow \vec{y}'; t) \prod_{i=1}^{k} \tau^{N(s,y'_i)},
\]

\[
\mathbb{E}\left[\prod_{i=1}^{k} \eta(t+s,y_i+1)\tau^{N(t+s,y_i)} \bigg| \mathcal{F}(s)\right] = \sum_{\vec{y} \in \mathbb{Y}^k} \mathbb{P}_{S\bar{6}V}(\vec{y} \rightarrow \vec{y}'; t) \prod_{i=1}^{k} \eta(s,y'_i+1)\tau^{N(s,y'_i)}
\]

\[
= \sum_{\vec{y} \in \mathbb{Y}^k} \mathbb{P}_{S\bar{6}V}(\vec{y} \rightarrow \vec{y}'; t) \prod_{i=1}^{k} \eta(s,y'_i+1)\tau^{N(s,y'_i)}.
\]

Above, the expectation is over the height function \(N(t+s, \cdot, \cdot)\) conditioned on its values \(N(s, \cdot, \cdot)\) at time \(s\), and \(\eta\) is coupled to \(N\) so that \(\eta(t, y) = N(t, y) - N(t, y-1)\).

**Proof.** We will give the proof for the \(H\) duality as the \(\bar{H}\) duality follows identically. Without loss of generality we assume that \(s = 0\). It suffices also to show that the duality holds for just \(t = 1\) since general \(t\) follows inductively.

Recall the notation \(\bar{x}_{\geq w}\) and \(\bar{x}_{\geq w}(t)\) from Lemma 2.4 for the bi-infinite stochastic 6V model cutoff to be left-finite with first particle \(x_w\). Applying the duality in Proposition 3.3 implies that

\[
\mathbb{E}\left[\prod_{i=1}^{k} \tau^{N_{y_i}(\bar{x}_{\geq w}(1))} \bigg| \mathcal{F}(s)\right] = \sum_{\vec{y} \in \mathbb{Y}^k} \mathbb{P}_{S\bar{6}V}(\vec{y} \rightarrow \vec{y}'; 1) \prod_{i=1}^{k} \tau^{N_{y'_i}(\bar{x}_{\geq w})},
\]

where the expectation is over \(\bar{x}_{\geq w}(t)\) at \(t = 1\) with initial condition \(\bar{x}_{\geq w}\) at \(t = 0\). In order to prove the corollary, we must show that taking \(w \rightarrow -\infty\), both sides of the above equation converge to their bi-infinite version. The left-hand side converges as \(w \rightarrow -\infty\) to \(\mathbb{E}\left[\prod_{i=1}^{k} \tau^{N_{y_i}(\bar{x}(1))}\right]\). This is because, by Lemma 2.4 \(N_{y_i}(\bar{x}_{\geq w}(t))\) converges in probability to \(N_{y_i}(\bar{x}(t))\) and in a single time step \(N_{y_i}(\bar{x}_{\geq w}(t))\) may change by at most one, hence the argument of the expectation is a bounded function. To show the right-hand side convergence, we bound (for some constant \(C < \infty\))

\[
\mathbb{P}_{S\bar{6}V}(\vec{y} \rightarrow \vec{y}'; 1) \leq C \prod_{i=1}^{k} b_i^{|y_i - y'_i|} \mathbb{1}_{\{y_i \geq y'_i\}}
\]

and then use the fact that \(\tau^{-1}b_2 = b_1 < 1\) to apply dominated convergence. The above bound follows since for the reversed stochastic 6V model \(\bar{y}(t)\), the increments (up to a minus sign) \(- (y_i(0) - y_i(1))\) can be stochastically upper bounded by \(1 + \text{geo}(b_2)\), where \(\text{geo}(b_2)\) is a geometric random variable with values in \(\mathbb{Z}_{\geq 0}\) with parameter \(b_2\). This proves the first identity for the \(H\) duality.

For the second identity, by definition of space-reversed stochastic 6V model, we have

\[
\mathbb{P}_{S\bar{6}V}(\vec{y} \rightarrow \vec{y}'; t) = \mathbb{P}_{S\bar{6}V}(\vec{y} \rightarrow -\vec{y}'; t),
\]
where, for \( \vec{y} = (y_1 < \ldots < y_k) \in \mathcal{Y}^k \), \(-\vec{y} := (-y_k < \ldots < -y_1) \in \mathcal{Y}^k \) denotes the space-reversed configuration. Further, the stochastic 6V model enjoys a space-time reversal symmetry:

\[
P_{\text{6V}}^{-}(\vec{y} \rightarrow \vec{y}'; t) = P_{\text{6V}}^{-}(\vec{y}' \rightarrow \vec{y}; t).
\]

To see this, notice that \((\vec{y}, -\vec{y}') \mapsto (\vec{y}', \vec{y})\) amounts to a vertical and horizontal flip in the vertex model configuration. Under such flips, the weights for \((+, +, +, -)\) remain unchanged, while the weights for \((-, +)\) swap. Given fixed initial and terminal conditions \((\vec{y}, \vec{y}')\), it is readily checked that 6V measures are invariant under the prescribed swap. From these considerations we conclude

\[
P_{\text{6V}}^{-}(\vec{y} \rightarrow \vec{y}'; t) = P_{\text{6V}}^{-}(\vec{y} \rightarrow -\vec{y}'; t) = P_{\text{6V}}^{-}(\vec{y}' \rightarrow \vec{y}; t).
\]

This proves the second claimed identity.

Owing to its Bethe ansatz solvability, the \(k\)-particle (reversed) stochastic 6V model admits explicit integral formulas for transition probabilities. We will make use of the below formula (and subsequent calculations involving it) we will use \(\vec{x}\) and \(\vec{y}\) to denote \(k\)-particle configurations (as opposed to \(\vec{y}\) and \(\vec{y}'\) as in our discussion on duality).

**Proposition 3.5.** Fix \(k \in \mathbb{Z}_{\geq 1}\) and parameters \(b_1, b_2 \in (0, 1)\) with \(b_2 < b_1\). Then for any \(\vec{x}, \vec{y} \in \mathcal{Y}^k\) (where \(\mathcal{Y}^k\) is the discrete Weyl Chamber defined in Definition 3.2) and \(t \in \mathbb{Z}_{\geq 0}\),

\[
P_{\text{6V}}^{-}(\vec{y} \rightarrow \vec{x}; t) = U(\vec{y}, \vec{x}; t)
\]

where \(U(\vec{y}, \vec{x}; t)\) is defined for all \(\vec{y}, \vec{x} \in \mathbb{Z}^k\) by

\[
U(\vec{y}, \vec{x}; t) = \oint_{C_r} \cdots \oint_{C_r} \sum_{\sigma \in \mathfrak{S}_k} (-1)^{\sigma} \prod_{1 \leq i < j \leq k} (\vec{y}_\sigma(i)) z_{\sigma(i)} z_{\sigma(j)} (\vec{z}_\sigma)^{\mathcal{D}(\vec{z})} \frac{dz_i}{2\pi i}.
\]

Here \(C_r\) is a circular contour (counter-clockwise oriented) centered at the origin with a large enough radius \(r\) so as to include all poles of the integrand, \(\mathfrak{S}_k\) is the set of all permutations on the set \(\{1, \ldots, k\}\), \((-1)^{\sigma} \in \{-1, 1\}\) is the sign of the permutation, and

\[
\tilde{\mathcal{F}}(z_i, z_j, \sigma) := \frac{1 - (1 + \tau^{-1}) z_{\sigma(i)} + \tau^{-1} z_{\sigma(i)} z_{\sigma(j)}}{1 - (1 + \tau^{-1}) z_i + \tau^{-1} z_i z_j}, \quad \mathcal{D}(z) := \frac{b_1 + (1 - b_1 - b_2) z^{-1}}{1 - b_2 z^{-1}}.
\]

**Proof.** This is a special case of [BCG16, Theorem 3.6, Eq. (26)] with \(c_1 = 1 - b_1, c_2 = 1 - b_2\) and \(a_1 = a_2 = 1\). \(\square\)

### 4. Hopf–Cole Transform: Reformulation of Theorem 1.1

One-particle \(H\) duality (Proposition 3.3) implies that \(E[\tau^{N(t,x)}]\) solves the evolution equation for a one-particle stochastic 6V model. As is true for general finite variance homogeneous random walks on \(\mathbb{Z}\), this evolution equation is a discrete heat equation and after proper centering and scaling, it will go to the continuous heat equation on \(\mathbb{R}\). In this section we describe (see Proposition 4.1) the martingale part that is left when ones does not take expectations, as well as the proper centering of the process \(\tau^{N(t,x)}\) that gives \(Z(t,x)\), the microscopic Hopf–Cole transform of \(N(t,x)\).

Given such a transform, we reformulate the convergence to KPZ equation (i.e., Theorem 1.1) as an equivalent statement of convergence to SHE (see Theorem 1.1*).

#### 4.1. Microscopic Hopf–Cole Transform

Recall that \(\rho \in (0, 1)\) is a fixed parameter representing the average density. Referring back to Theorem 1.1, we notice that the convergence results involve centering and tilting of the height function \(N(t,x)\). Our first step here is hence to introduce the corresponding centering and tilting of \(\tau^{N(t,x)}\). To setup notation, consider the stochastic 6V model...
with a single particle starting from \( x = 0 \). This is simply a discrete-time random walk \( X(t) = S(1) + \ldots + S(t) \), with i.i.d. increment \( S(1), \ldots, S(t) \) that has distribution \( S(t) \text{law} S \), where

\[
\mathbb{P}(S = n) = \begin{cases} 
(1 - b_1)(1 - b_2)b_2^{n-1}, & \text{when } n > 0, \\
b_1, & \text{when } n = 0,
\end{cases} \quad (4.1)
\]

Now, with \( N(t, x) \) being tilted by \(-\rho x\) in (1.8), we consider the analogous tilt of \( S \):

\[
\mathbb{P}(S' = n) := \lambda \mathbb{E}[\tau^{-\rho S}1_{\{S = n\}}] = \lambda \tau^{-\rho \mu} \mathbb{P}(S = n),
\]

(4.2)

The parameter \( \lambda = (\mathbb{E}[\tau^{-\rho S}])^{-1} \) is in place to ensure (4.2) defines a random variable, and the variable \( S' \) has mean \( \mu = \mathbb{E}[S'] > 0 \). From (4.1), it is straightforward to check\(^9\) that \( \lambda \) and \( \mu \) are given by (1.1)–(1.2). We further consider the corresponding centered variable \( R := S' - \mu \). With \( \mu \) being the centering parameter (in Theorem 1.1) that set the reference frame alone the characteristics, we let

\[
\Xi(t) = Z - t\mu
\]

denote a shifted integer lattice to accommodate the centering by \( \mu \). Under these notation, we define (microscopic) Hopf–Cole (i.e. Gärtner) transform of the stochastic 6V model as

\[
Z(t, x) := \lambda^t N(t, x + \mu) - \rho(x + \mu), \quad x \in \Xi(t),
\]

(4.3)

where \( \lambda \) and \( \mu \) are given in (1.1)–(1.2).

It is straightforward to verify that the \( k = 1 \) duality for \( \tau^{N_s(x)} \) given in Proposition 3.3 implies

\[
\mathbb{E}[Z(t + 1, x - \mu) - Z(t, x) | \mathcal{F}(t)] = (LZ(t))(x - \mu),
\]

(4.4)

where \( L \) denotes the generator of the random walk associated with \( R \). More precisely, let

\[
p(x) := \mathbb{P}(R = x) = \begin{cases} 
\lambda(1 - b_1)(1 - b_2)b_2^{x+\mu-1}\tau^{-\rho(x+\mu)}, & \text{when } x + \mu \in \mathbb{Z}_{>0}, \\
\lambda b_1, & \text{when } x + \mu = 0,
\end{cases} \quad (4.5)
\]

denote the probability mass function of \( R \), and let \( p \) act on functions \( f : \Xi(t) \to \mathbb{R} \) as

\[
(pf)(x) := \sum_{y \in \Xi(t)} p(x - y)f(y), \quad x \in \Xi(t + 1).
\]

We likewise define the generator

\[
(Lf)(x) := \sum_{y \in \Xi(t)} (p(x - y) - 1_{x+\mu=y})f(y), \quad x \in \Xi(t + 1).
\]

(4.6)

The operator \( p \) and generator \( L \) map a function on the lattice \( \Xi(t) \) to a function on the lattice \( \Xi(t + 1) \); this is the only dependence of \( p \) and \( L \) on \( t \) and we drop this dependence in our notation.

Equation (4.4) states that \( Z(t + 1, x - \mu) - Z(t, x) - (LZ(t))(x - \mu) := M(t, x) \) is an \( \mathcal{F} \)-martingale increment. We next calculate its quadratic variation.

**Proposition 4.1.** For any \( t \in \mathbb{Z}_{\geq 0} \) and \( x \in \Xi(t) \), we have

\[
Z(t + 1, x - \mu) - Z(t, x) = (LZ(t))(x - \mu) + M(t, x),
\]

(4.7)

where \( M(t, x) \) is an \( \mathcal{F} \)-martingale increment, i.e., \( \mathbb{E}[M(t, x) | \mathcal{F}(t)] = 0, \) \( t \in \mathbb{Z}_{\geq 0}, \) with

\[
\mathbb{E}[M(t, x_1)M(t, x_2) | \mathcal{F}(t)] = (b_1\tau^{-1}b_2^{x_1-x_2}\Theta_1(t, x_1 \land x_2)\Theta_2(t, x_1 \land x_2),
\]

(4.8)

\[
\Theta_1(t, x) := \lambda\tau^{-1}Z(t, x) - (pZ(t))(x - \mu),
\]

(4.9)

\[
\Theta_2(t, x) := -\lambda Z(t, x) + (pZ(t))(x - \mu).
\]

(4.10)

\(^9\) The computation for \( \lambda \) simply boils down to a geometric series. The computation for \( \mu \) boils down to a sum of the form \( \sum_{n \geq 0} (n + 1)(b_2\tau^{-r})^n \); this multiplied by \( (1 - b_2\tau^{-r}) \) again gives a geometric series.
Proof. The result is a special case of the statement and proof of [CT17, Proposition 2.6]. In [CT17] the authors consider a more general higher-spin version of the stochastic 6V model [CP16] with arbitrary non-negative integer valued horizontal spin $J$ as well as parameters $\alpha, q, \nu$. Our stochastic 6V corresponds with taking $J = 1$ and $\nu = 1/q = \tau$ therein, and matching $b_1 \mapsto \frac{1}{1 + q}$ and $\tau^{-\rho} \mapsto \rho$. \hfill \Box

More generally, for $k \geq 2$, $Z(t, x)$ inherits a duality from $\tau^{N(t,x)}$, analogous to Corollary 3.4 and Proposition 3.5. The analogous semigroup integral formulas are obtained by a centering and tilting of $U$ (as in Proposition 3.5). We state the duality and integral formula result for $Z$ only for $k = 2$ (as we will only need that case). For $y_1 < y_2 \in \mathbb{E}(s)$ and for $x_1 < x_2 \in \mathbb{E}(s + t)$, we define

$$V((y_1, y_2), (x_1, x_2); t) := \oint_{C_r} \oint_{C_r} \bigg( z_1^{x_1-y_1+\lfloor \mu t - \lfloor \mu t \rfloor \rfloor} z_2^{x_2-y_2+\lfloor \mu t - \lfloor \mu t \rfloor \rfloor} - z_2^{x_2-y_2+\lfloor \mu t - \lfloor \mu t \rfloor \rfloor} z_1^{x_1-y_1+\lfloor \mu t - \lfloor \mu t \rfloor \rfloor} \bigg)^2 \prod_{i=1}^{2} \frac{\mathcal{D}(t, z_i) dz_i}{2 \pi i z_i}. \quad (4.11)$$

Here $C_r$ is a counter-clockwise oriented, circular contour that is centered at origin, with a large enough radius $r$ so as to include all poles of the integrand, and

$$\mathcal{F}(z_1, z_2) := \frac{1 + \tau^{-1}2\rho z_1 z_2 - (1 + \tau^{-1})\rho^2 z_2}{1 - 1 + 2\rho z_1 z_2 - (1 + \tau^{-1})\rho^2 z_2}, \quad (4.12)$$

$$\mathcal{D}(t, z) := z^{\lfloor \mu t \rfloor} \left( \frac{b_1 + (1 - b_1 - b_2)/(\tau^\rho z)}{1 - b_2/(\tau^\rho z)} \right)^t. \quad (4.13)$$

Remark 4.2. One could rewrite the formula (4.11) in a seemingly simpler form:

$$V((y_1, y_2), (x_1, x_2); t) = \oint_{C_r} \oint_{C_r} \bigg( z_1^{x_1-y_1} z_2^{x_2-y_2} - \mathcal{F}(z_1, z_2) z_1^{x_1-y_1} z_2^{x_2-y_2} \bigg)^2 \prod_{i=1}^{2} \frac{\mathcal{D}(z_i) dz_i}{2 \pi i z_i},$$

where $\mathcal{D}(z) := z^{\mu} \lambda b_1 + (1 - b_1 - b_2)/(\tau^\rho z)$. The expression, however, involves non-integer powers of $z_i$, because $x_i - y_j \notin \mathbb{Z}$ and $\mu \notin \mathbb{Z}$ in general, and having non-integer powers is undesirable for our analysis in the sequel. With $x_i - y_i \in \mathbb{E}(t)$, we have that $x_i - y_j + (\lfloor \mu t \rfloor - \lfloor \mu t \rfloor \rfloor \rfloor) \in \mathbb{Z}$, so the formula (4.11) involves only integer powers of $z_i$.

Adopt the shorthand notation for centered occupation variables:

$$\eta_c(t, x) := \eta(t, x + \mu t), \quad \eta_c^+(t, x) := \eta_c(t, x + 1), \quad x \in \mathbb{E}(t). \quad (4.14)$$

Proposition 4.3. With $Z$ being the Hopf–Cole transform of the stochastic 6V model with parameter $b_1 > b_2 \in (0, 1)$, for all $x_1 < x_2 \in \mathbb{E}(s + t)$ and $t, s \in \mathbb{Z}_{\geq 0}$, we have

$$E \left[ Z(t+s,x_1)Z(t+s,x_2) \bigg| \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \mathbb{E}(s)} V((y_1, y_2), (x_1, x_2); t) Z(s, y_1)Z(s, y_2), \quad (4.15)$$

$$E \left[ (\eta_c^+ Z)(t+s,x_1)(\eta_c^+ Z)(t+s,x_2) | \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \mathbb{E}(s)} V((y_1, y_2), (x_1, x_2); t) (\eta_c^+ Z)(s, y_1)(\eta_c^+ Z)(s, y_2). \quad (4.16)$$

Proof. Recall from (4.3) that $Z(t, x)$ is obtained from $\tau^{N(t,x+\mu t)}$ through centering and tilting. Translate the $k = 2$ duality from Corollary 3.4 and Proposition 3.5 through the centering and tilting. We see that (4.15)–(4.16) holds for

$$V((y_1, y_2), (x_1, x_2); t) = \lambda^{2t} e^{-\rho(x_1 + x_2 - y_1 - y_2 + 2\mu t)} U \left( (y_1 + \mu s, y_2 + \mu s), (x_1 + \mu(t + s), x_2 + \mu(t + s)); t \right). \quad (4.17)$$
Our goal here is to show that this $V$ (given in (4.17)) is indeed written as the contour integral in (4.11). Referring back to the formula (3.2) for $U$, with (4.17), here we have

$$V((y_1, y_2), (x_1, x_2); t) = \oint_{C_r} \oint_{C_{r}} \left( (\tau^{-\rho} z_1)^{x_1 \cdot y_1 + \mu t \cdot |\mu t|} (\tau^{-\rho} z_2)^{x_2 - y_2 + \mu t \cdot |\mu t|} \right)$$

$$- \frac{1 - (1 + \tau^{-1}) z_2 + \tau^{-1} z_1 z_2}{1 - (1 + \tau^{-1}) z_1 + \tau^{-1} z_1 z_2} \left( (\tau^{-\rho} z_1)^{x_2 - y_1 + \mu t \cdot |\mu t|} (\tau^{-\rho} z_2)^{x_1 \cdot y_2 + \mu t \cdot |\mu t|} \right) \prod_{i=1}^{2} \frac{\mathcal{D}(z_i, t) dz_i}{2\pi i z_i},$$

where, with $U(z)$ defined in Proposition 3.5, $\mathcal{D}(t, z) := (\tau^{-\rho} z)^{|\mu t|} \lambda^{t} \mathcal{D}(z)^{t}$. Given this, the claimed result now follows by the change of variable $\tau^{-\rho} z_i := \bar{z}_i$. \hfill \Box

4.2. The SHE. Proposition 4.1 states that $Z$ solves a discrete-time, discrete space SPDE. Examining this equation suggests that, under appropriate scaling, $Z$ should converge to the solution of the SHE:

$$\partial_t Z(t, x) = \frac{\nu_s}{2} \partial_x^2 Z(t, x) + \frac{\kappa_s \sqrt{D_s}}{\nu_s} \xi(t, x) Z(t, x).$$

(4.18)

The coefficients $\nu_s$, $\kappa_s$ and $D_s$ are given in (1.6). (Although $\nu_s = \kappa_s$, we prefer to write the equation as above to better track the limiting coefficients.)

To formulate the convergence to SHE precisely, recall that a $C([0, \infty), C(\mathbb{R}))$-valued process $Z$ is a mild solution of (4.18) with initial condition $Z^{ic}(x)$ if

$$Z(t, x) = \int_{\mathbb{R}} p(\nu_s, t, x-y) Z^{ic}(y) dy + \int_{0}^{t} \int_{\mathbb{R}} p(\nu_s(t-s), x-y) Z(s, y) \frac{\kappa_s \sqrt{D_s}}{\nu_s} \xi(s, y) ds dy,$$

for all $t \in [0, \infty)$ and $x \in \mathbb{R}$. Given non-negative $Z^{ic} \in C(\mathbb{R})$ that is not identically zero, the SHE permits a unique mild solution that stays positive for all $t > 0$. See, for example, [Cor12, Proposition 2.5] and the references therein. With the SHE being an informal exponentiation of the KPZ equation, we say $H$ is a Hopf–Cole solution of the KPZ equation (1.5) if

$$e^{-\frac{2s}{\sqrt{\tau}} H(t, x)} = e^{-H(t, x)}$$

(4.19)

is a mild solution of (4.18). So far our discussion has been for a $C(\mathbb{R})$-valued $Z^{ic}$, which is the proper setup for near stationary initial conditions (defined in the following). To accommodate the step initial condition, $\eta(0, x) = 1_{\{x \geq 0\}}$, we need to also consider the SHE starting from delta function $\delta(x)$. The mild solution is defined analogously:

$$Z(t, x) = p(\nu_s, t, x-y) + \int_{0}^{t} \int_{\mathbb{R}} p(\nu_s(t-s), x-y) Z(s, y) \frac{\kappa_s \sqrt{D_s}}{\nu_s} \xi(s, y) ds dy,$$

for $t > 0$ and $x \in \mathbb{R}$. It is standard to show that, for delta initial condition, there exists a unique $C((0, \infty), C(\mathbb{R}))$-valued solution $Z$, which is positive. For such $Z$, we then define $H(t, x) := \log(Z(t, x))$ as the solution of the KPZ equation (1.5) with narrow wedge initial condition.

As discussed above Theorem 1.1, we will prove convergence to the Hopf-Cole solution to the KPZ equation under weak asymmetry scaling, where

$$\rho \in (0, 1), b_1 \in (0, 1)$$

are fixed, $\tau = \tau_{\varepsilon} = b_2 / b_1 = b_2^\rho / b_1 := e^{-\sqrt{\varepsilon}}$ and $(\lambda, \mu) = (\lambda_{\varepsilon}, \mu_{\varepsilon})$ are defined in (1.1)(1.2) which behave asymptotically as (1.3)(1.4). Under this scaling, the microscopic Hopf-Cole transform (4.3) reads

$$Z(t, x) = Z_{\varepsilon}(t, x) := e^{\frac{1}{4} \log \lambda_{\varepsilon} - \sqrt{\varepsilon} (N_{\varepsilon}(t, x + \mu_{\varepsilon} t) - \rho(x + \mu_{\varepsilon} t))}, \quad x \in \Xi(t).$$

(4.20)

Hereafter, we adopt the standard notation $\|X\|_n := (\mathbb{E}[|X|^n])^{\frac{1}{n}}$, and say for all $\varepsilon > 0$ small enough if the referred statement holds for all $\varepsilon \in (0, \varepsilon_0)$, for some generic but fixed threshold $\varepsilon_0 > 0$ that may change from line to line. Following [BG97], we define near stationary initial conditions for the stochastic 6V model:
**Definition 4.4.** Fix any density parameter $\rho \in (0, 1)$. With $\varepsilon \downarrow 0$ being the scaling parameter, consider a sequence of possibility random initial conditions $\{N_\varepsilon(0, x)\}_{\varepsilon > 0}$, and let $Z_\varepsilon(0, x)$ denote the corresponding Hopf–Cole transformed initial data defined through (4.3). We say the initial condition is **near stationary with density** $\rho$ if, for any given $n < \infty$ and $\alpha \in (0, \frac{1}{2})$, there exist constant $C = C(n, \alpha)$ and $u = u(n, \alpha)$, such that
\[
\|Z_\varepsilon(0, x)\|_n \leq C \exp (u\varepsilon|x|),
\]
\[
\|Z_\varepsilon(0, x) - Z_\varepsilon(0, x')\|_n \leq C \left(\varepsilon|x - x'|\right)^{\alpha} \exp \left(u\varepsilon(|x| + |x'|)\right),
\]
for all $x, x' \in \mathbb{Z}$, and small enough $\varepsilon > 0$.

We now state our result on the convergence of $Z(t, x)$ to the SHE. Due to the round-about definition of the Hopf–Cole solution (4.19), it is readily checked (see (4.20)) that, Theorem 1.1* in the following formulation of Theorem 1.1. Given $Z(t, x)$, $t \in \mathbb{Z}_{\geq 0}$, $x \in \Xi(t)$, we first linearly interpolate in $x$ and then linearly interpolate in $t$ to obtain a $C([0, \infty), \mathbb{R})$-valued process.

**Theorem 1.1*. Consider the stochastic 6V model, with parameter $b_1 > b_2 \in (0, 1)$.

(a) **(Near stationary initial conditions)** Fix a density $\rho \in (0, 1)$. Start the stochastic 6V model from a sequence (parameterized by $\varepsilon$) of near stationary with density $\rho$ initial conditions, and let $Z_\varepsilon(t, x)$ denote the resulting Hopf–Cole transform. If, for some $C(\mathbb{R})$-valued process $\mathcal{Z}^{ic}$, we have
\[
Z_\varepsilon(0, \varepsilon^{-1}x) \Rightarrow \mathcal{Z}^{ic}(x), \quad \text{in } C(\mathbb{R}),
\]
then, under the weak asymmetry scaling we have
\[
Z_\varepsilon(\varepsilon^{-2}t, \varepsilon^{-1}x) \Rightarrow \mathcal{Z}(t, x), \quad \text{in } C([0, \infty), C(\mathbb{R})),
\]
where $\mathcal{Z}(t, x)$ is the mild solution of the SHE (4.18) with initial condition $\mathcal{Z}^{ic}(x)$.

(b) **(Step initial condition)** Start the stochastic 6V model from the step initial condition $N(0, x) = (x)_{+}$, and let $Z_\varepsilon(t, x)$ denote the resulting Hopf–Cole transform. Let $\rho \in (0, 1)$ be fixed. Under the weak asymmetry scaling we have
\[
\frac{\rho(1 - \rho)}{\varepsilon} Z_\varepsilon(\varepsilon^{-2}t, \varepsilon^{-1}x) \Rightarrow \mathcal{Z}(t, x), \quad \text{in } C([0, \infty), C(\mathbb{R})),
\]
where $\mathcal{Z}(t, x)$ is the mild solution of the SHE (4.18) with delta initial condition $\delta(x)$.

5. **Proof of Theorems 1.1* and 1.6

The framework of the proof of Theorem 1.1* is standard. We first establish tightness, and then identify the limit through martingale problems. As noted earlier, the major technical step here is to establish self-averaging of the quadratic variation. We state this as Proposition 5.3 (postponing its proof to Section 7) and give the rest of the proof of Theorem 1.1* in Section 5.1.

Given Theorem 1.1 (or equivalently Theorem 1.1*), Theorem 1.6 follows as a rather straightforward consequence. In Section 5.2, we establish Theorem 1.6.

Hereafter, we will be assuming either the weak asymmetry scaling $\tau = \tau_\varepsilon = e^{-\sqrt{\varepsilon}}$ (for the stochastic model), or the scaling $\eta = \eta_\varepsilon = \frac{1}{2}\sqrt{\varepsilon}$ (for the symmetric model under Baxter’s projective parametrization (1.12)). To highlight this dependence, for parameters we write $\lambda = \lambda_\varepsilon$, $\mu = \mu_\varepsilon$, etc. On the other hand, to simplify notation, for processes we often omit this dependence, and write $Z_\varepsilon = Z$, etc. We also adopt the notation $C(\alpha, \beta, \ldots) < \infty$ for a generic deterministic finite constant that may change from line to line, but depend only on the designated variables $\alpha, \beta, \ldots$. The dependence on $(\rho, b_1) \in (0, 1)^2$ will not be indicated as they are fixed throughout the article.

---

10This is different from exponentiating the interpolated height function. Nevertheless, under the weak asymmetry scaling $\tau = \exp(-\sqrt{\varepsilon})$, it is straightforward to verify that the difference between these two interpolation schemes is negligible as $\varepsilon \to 0$. 
5.1. **Proof of Theorem 1.1***. Given the existing literature (e.g., [CT17]), the proof presented here is standard except for Proposition 5.3 on the self-averaging of quadratic variation, whose proof is postponed to Section 7. Hence, we only sketch the proof: highlighting main ideas, and referring to existing works where the required techniques are found.

5.1.1. **Part (a): near stationary initial conditions.** Fix throughout this subsection some near stationary initial condition \( Z(0, x) \) (note that we omit the \( \epsilon \) dependence as declared earlier). Let \( Z(t, x) \) denote the corresponding Hopf–Cole transform evolving under the stochastic 6V model (with the weak asymmetry scaling, as declared earlier). The first step of the proof is to establish moment bounds on \( Z(t, x) \). Recall that \( \| \cdot \|_n := \left( \mathbb{E}[|X|^n] \right)^{1/n} \).

**Proposition 5.1.** For any \( \alpha \in (0, \frac{1}{2}) \), \( n \in \mathbb{Z}_{>0} \), and \( T < \infty \), there exist \( C := C(\alpha, T) < \infty \) and \( u = u(n, \alpha) < \infty \) such that

\[
\| Z(t, x) \|_n \leq e^{u|\epsilon|},
\]

\[
\| Z(t, x) - Z(t, x') \|_n \leq C(\epsilon|x - x'|)^{2\alpha} e^{u|\epsilon|x'|},
\]

\[
\| Z(t, x) - Z(t', x) \|_n \leq C(\epsilon^2|t - t'|)^{\alpha} e^{u|\epsilon|x'|}.
\]

for any \( t, t' \in [0, \epsilon^{-2}T] \) and \( x, x' \in \mathbb{R} \).

**Sketch of proof.** The proof follows standard iterative calculations. This starts with a series of heat kernel estimates. More precisely, with \( p(x) \) being the (one-step) transition probability as in (4.5), we define the \( t \)-step transition probability:

\[
p^t(x) := \left( \prod_{i=0}^{t-1} p \right)(x) = \sum_{x_1, x_2, \ldots, x_t = x} p(x_1) \cdots p(x_t), \quad x \in \Xi(t).
\]

Using the type of estimates in [CT17, Lemma 4.3] and [BG97, Section A], one establishes the following estimates on \( p^t \). For any \( u, T < \infty \) and \( \alpha \in (0, 1/4) \), there exists \( C = C(T, \alpha) \), such that, for all \( t \leq \epsilon^{-2}T \) and \( x, x' \in \Xi(t) \),

\[
\sum_{x \in \Xi(t)} p^t(x) \exp(u\epsilon|x|) \leq C, \quad \sum_{x \in \Xi(t)} |x|^{2\alpha} p^t(x) \exp(u\epsilon|x|) \leq C t^{2\alpha},
\]

\[
p^t(x) \leq C(t + 1)^{-1/2}, \quad |p^t(x) - p^t(x')| \leq C|x - x'|^{4\alpha(t + 1)^{-(2\alpha + 1/2)}}.
\]

Given such estimates, the next step is to rewrite (4.7) in integrated form as

\[
Z(t + 1, x - \mu_\epsilon) = \left( pZ(t) \right)(x - \mu_\epsilon) + M(t, x), \quad x \in \Xi(t),
\]

which, upon iterating in \( t \) from \( t = 0 \), gives

\[
Z(t, x) = \left( p^t Z(0) \right)(x) + \sum_{s=0}^{t-1} \left( p^{t-s-1} M(s) \right)(x + \mu_\epsilon), \quad t \in \mathbb{Z}_{\geq 0}, \quad x \in \Xi(t).
\]

Given this equation, one bounds the moment of \( Z(t, x) \) in terms of \( Z(0, x) \). With the aid of the preceding estimates on \( p^t \) and Burkholder’s inequality, using techniques analogous to those in [CT17, Lemma 4.3-4.4], one obtains the desired bounds (5.1)–(5.3) for \( t, t' \in \mathbb{Z}_{\geq 0} \) and \( x, x' \in \Xi(t) \). Extension to real value \( t, t', x, x' \) follows easily as the process \( Z(t, x) \) is extended in \( (t, x) \) to real values through linear interpolation.

An immediate consequence of Proposition 5.1 is the tightness of \( Z(\epsilon^{-2} \cdot, \epsilon^{-1} \cdot) \).

**Corollary 5.2.** The collection of processes \( \{ Z(\epsilon^{-2} \cdot, \epsilon^{-1} \cdot) \}_{\epsilon > 0} \) is tight in \( C([0, \infty), C(\mathbb{R})) \).

**Proof.** This follows from the Kolmogorov–Chentsov criterion as in [CT17, Proof of Proposition 2.12].

□
Given the tightness result, it remains to show that the limit points are the mild solution of SHE. We achieve this through martingale problems. Recall from [BG97] that, we say a \( C(0, \infty), C(\mathbb{R}) \)-valued process \( Z(t, x) \) solves the **martingale problem** associated with the SHE (4.18) if, for any given \( T < \infty \), there exists \( C(T) < \infty \) such that

\[
\sup_{t \in [0, T]} \sup_{x \in \mathbb{R}} e^{-|x|C(T)} \mathbb{E} \left[ Z^2(t, x) \right] < \infty, \tag{5.4}
\]

and if, for any \( \phi \in C_c^\infty(\mathbb{R}) \),

\[
t \mapsto \mathcal{M}_\phi(t) := \left( \int_{\mathbb{R}} \phi(x) Z(s, x) dx \right)_{s=0}^{s=t} - \nu_x s - \frac{\nu_x^2}{2} \int_0^t \frac{\phi''(x)}{x} Z(s, x) ds dx, \tag{5.5}
\]

\[
t \mapsto \mathcal{N}_\phi(t) := \frac{D_x \nu_x^2}{\nu_x^2} \int_0^t \int_{\mathbb{R}} \phi^2(x) Z^2(s, x) ds dx \tag{5.6}
\]

are local martingales. It is shown in [BG97] that any solution \( Z \) of the prescribed martingale problem is a solution\(^{11}\) of the SHE (4.18). Moreover, they show that there is a unique such solution.

Hence, it suffices to show that any limit point of \( Z(\varepsilon^{-2} \cdot, \varepsilon^{-1} \cdot) \) solves the martingale problem. As mentioned earlier, the major technical step occurs in establishing (5.6) (i.e., the quadratic martingale problem), where we need self-averaging of the quadratic variation. We now state the desired estimate on such self-averaging. To this end, recall the expressions \( \Theta_1, \Theta_2 \) from (4.9)–(4.10), which are associated with the quadratic variation of the martingale increment \( M \) in Proposition 4.1.

**Proposition 5.3.** Given any fixed \( T < \infty \), we have that, for all \( t \in [0, \varepsilon^{-2} T] \cap \mathbb{Z}, x_* \in \mathbb{Z}, \) and all \( \varepsilon > 0 \) small enough,

\[
\left\| e^2 \sum_{s=0}^t \left( \varepsilon^{-1} \Theta_1 \Theta_2 - \frac{2\beta \rho (1 - \rho)}{1 + \beta} Z^2(s, x_* + \mu_s) \right) \right\|_2 \leq \varepsilon \frac{1}{2} C(T) e^{C \varepsilon |x_*|}. \tag{5.7}
\]

**Remark 5.4.** In (5.7), we compensate the space variable \( x_* \) by \( \mu_s, \mu_s \) to ensure the resulting variables is in \( \Xi(s) \).

**Remark 5.5.** Proposition 5.3 demonstrates a self-averaging upon integrating over long time interval, namely, the quadratic variation of the martingale \( M(t, x) \) subtracting the leading order term (that is, a constant multiple of \( Z^2 \)), vanishes as \( \varepsilon \to 0 \). This is not obvious at all and is the linchpin of the analysis of the present paper. The remainder of this subtraction is given in Lemma 7.2, which consists of terms of the form \((\varepsilon^{-\frac{1}{2}} \nabla Z)(t, x_1) Z(t, x_2), (\varepsilon^{-\frac{1}{2}} \nabla Z)(t, x_1)(\varepsilon^{-\frac{1}{2}} \nabla Z)(t, x_2) \) for \( x_1 \sim x_2 \). By the definition of \( Z \), see (4.20), \( \nabla Z \) behaves as \( \varepsilon \frac{1}{2} Z \), so these remainder terms seem to be of the same order as the leading order term. Self-averaging is key to showing that they are, in fact, of lower order. The proof of Proposition 5.3 is given in Section 7, which relies on duality argument in Section 7 and estimates of two-point transition kernels given in Section 6. The heuristic on how duality and estimates of transition kernels lead to the proof of such a self-averaging is discussed in Appendix A with the simpler example of ASEP.

Postponing the proof of this proposition to Section 7, we now finish the proof of Theorem 1.1*(a):

**Proposition 5.6.** Any limit point of \( \{ Z(\varepsilon^{-2} \cdot, \varepsilon^{-1} \cdot) \}_{\varepsilon > 0} \) solves the martingale problems (5.5) and (5.6).

**Sketch of Proof.** Fix a limit point \( Z \), and, after passing to a subsequence, we assume \( Z(\varepsilon^{-2} \cdot, \varepsilon^{-1} \cdot) \) converges in distribution to \( Z \).

The condition (5.4) is readily verified from the moment bounds in Proposition 5.1.

We now turn to verifying the condition (5.5), i.e., showing that \( \mathcal{M}_\phi \) is a local martingale. To this end, fixing a test function \( \phi \in C_c^\infty(\mathbb{R}) \), we consider the discrete, microscopic analog of \( \mathcal{M}_\phi \):

\[
M_\phi(t) := \varepsilon \sum_{x \in \Xi(s)} \phi(\varepsilon x) Z(s, x)_{s=0}^{s=t} + \varepsilon \sum_{s=1}^{t} \sum_{x \in \Xi(s)} \phi(\varepsilon x) \left( \mathcal{L}Z(s - 1) \right)(x),
\]

\(^{11}\)A weak solution in fact. But it is standard to show that solving (4.18) in the weak and mild senses are equivalent.
where $L$ is the generator of a random walk defined in (4.6). From Proposition 4.1, we have

$$M_\phi(t) = \varepsilon \sum_{s=0}^{t-1} \sum_{x \in \mathbb{Z}} \phi(\varepsilon x)M(s, x),$$

where $M(s, x)$ is an $\mathcal{F}$-martingale increment. In particular, $(M_\phi(t))_{t \in \mathbb{Z}_{\geq 0}}$ is a martingale. Using standard argument from [CT17, Proof of Proposition 2.13], one approximates the continuous-time process $M_\phi(t)$ with the discrete-time process $M_\phi(t)$ as $\varepsilon \to 0$, whereby showing that the former is a local martingale. The factor $\nu_*$ arises as the variance of the associated random walk. More precisely, from (4.5) and (1.3)-(1.4), with $b_2 = e^{-\sqrt{\varepsilon} b_1}$, we calculate

$$\text{Var}(R_\varepsilon) = \mu_\varepsilon^2 \lambda \varepsilon b_1 + \sum_{n \geq 1} (n - \mu_\varepsilon)^2 \lambda \varepsilon (1 - b_1)(1 - b_2^n)(b_2^n - n \tau_\varepsilon^{1-n})$$

$$= \mu_\varepsilon^2 \lambda \varepsilon b_1 + \lambda \varepsilon (1 - b_1)(1 - b_2^n)\tau_\varepsilon^{1-n} - \frac{2\mu_\varepsilon + 1}{(1 - b_2^n \tau_\varepsilon^{1-n})^2} + \frac{2}{(1 - b_2^n \tau_\varepsilon^{1-n})^3}$$

$$\longrightarrow \nu_* = \frac{2b_1}{1 - b_1}, \quad \text{as } \varepsilon \to 0. \quad (5.8)$$

Here we used the fact that the sum over $n$ multiplied by a factor $(1 - b_2^n \tau_\varepsilon^{1-n})^2$ gives a quantity that can be summed as geometric series.

The proof of (5.6) follows by a similar approximation scheme. Specifically,

$$t \longmapsto M_\phi(t) - \langle M_\phi \rangle(t), \quad t \in \mathbb{Z}_{\geq 0}$$

is an $\mathcal{F}$-martingale, where $\langle M_\phi \rangle(t)$ is the quadratic variation of $M_\phi(t)$, given by

$$\langle M_\phi \rangle(t) := \sum_{s=1}^{t} \mathbb{E}[(M_\phi(s) - M_\phi(s-1))^2|\mathcal{F}(t)].$$

The major step here is to argue that $\langle M_\phi \rangle(t)$ is well-approximated by a discrete analog of

$$\frac{D_x^2}{\nu_\varepsilon^2} \int_0^t \int_{\mathbb{Z}} (Z^2 \phi^2)(s, x)dsdx.$$ 

To this end, using (4.8), we calculate $\langle M_\phi \rangle(t)$ as

$$\langle M_\phi \rangle(t) = \varepsilon^2 \sum_{s=0}^{t-1} \phi(\varepsilon x)\phi(\varepsilon x')(b_1 e^{-\sqrt{\varepsilon}(1-\rho)}|x' - x|\Theta_1(t, x \land x')\Theta_2(t, x \land x')).$$

With $b_1 < 1$, the factor $(b_1 e^{-\sqrt{\varepsilon}(1-\rho)}|x' - x|)$ introduces an exponential decay in $|x - x'|$. Using the continuity of $\phi$, it is standard to show that the previous expression is well-approximated by the corresponding expression where $\phi(\varepsilon x)\phi(\varepsilon x')$ is replaced by $\phi^2(\varepsilon(x \land x'))$. More precisely, letting $\mathcal{E}_\varepsilon(t)$ denote a generic process such that

$$\lim_{\varepsilon \to 0} \sup_{t \in \mathbb{Z}, |x| \leq T} \|\mathcal{E}_\varepsilon(t)\|_2 = 0, \quad \text{for any given } T < \infty, \quad (5.9)$$

the continuity of $\phi$ gives that

$$\langle M_\phi \rangle(t) = \varepsilon^2 \sum_{s=0}^{t-1} \phi^2(\varepsilon(x \land x'))(b_1 e^{-\sqrt{\varepsilon}(1-\rho)}|x' - x|\Theta_1(t, x \land x')\Theta_2(t, x \land x') + \mathcal{E}_\varepsilon(t)).$$

With $\sum_{y \in \mathbb{Z}}(b_1 e^{-\sqrt{\varepsilon}(1-\rho)})|y| = \frac{1+b_1 e^{-\sqrt{\varepsilon}(1-\rho)}}{1-b_1 e^{-\sqrt{\varepsilon}(1-\rho)}} \rightarrow \frac{1+b_1}{1-b_1}$, we now have

$$\langle M_\phi \rangle(t) - \frac{1+b_1}{1-b_1} \varepsilon^2 \sum_{s=0}^{t-1} \sum_{x \in \mathbb{Z}(s)} \varepsilon^{-1} \Theta_1(t, x)\Theta_2(t, x)\phi^2(\varepsilon x) = \mathcal{E}_\varepsilon(t). \quad (5.10)$$
Further, fixing some large enough \( L < \infty \) with \( \text{supp}(\phi) \subset [-L, L] \), we have

\[
\| \varepsilon^2 \sum_{s=0}^{t} \varepsilon \sum_{x \in \Xi(s)} (\varepsilon^{-1}\Theta_1\Theta_2 - \frac{2b_1(1-\rho)}{1+b_1} Z^2)(s, x)\phi^2(\varepsilon x) \|_2
\]

\[
= \| \varepsilon \sum_{x \in \mathbb{Z}} \varepsilon^2 \sum_{s=0}^{t} (\varepsilon^{-1}\Theta_1\Theta_2 - \frac{2b_1(1-\rho)}{1+b_1} Z^2)(s, x, \mu_{\varepsilon}s - [\mu_{\varepsilon}s])\phi^2(\varepsilon(x + \mu_{\varepsilon}s - [\mu_{\varepsilon}s])) \|_2
\]

\[
\leq C(L, \phi) \sup_{s, \in [-\varepsilon L, \varepsilon L] \cap \mathbb{Z}} \| \varepsilon^2 \sum_{s=0}^{t} (\varepsilon^{-1}\Theta_1\Theta_2 - \frac{2b_1(1-\rho)}{1+b_1} Z^2)(s, x, \mu_{\varepsilon}s - [\mu_{\varepsilon}s]) \|_2.
\]

The last expression, by Proposition 5.3, is bounded by \( C(T, \varepsilon L, \phi)\varepsilon^{\frac{1}{2}} \), for all \( t \in \mathbb{Z} \cap [0, \varepsilon^{-2}T] \), for each fixed time horizon \( T < \infty \). Consequently,

\[
\varepsilon^{2} \sum_{s=0}^{t} \varepsilon \sum_{x \in \Xi(s)} (\varepsilon^{-1}\Theta_1\Theta_2 - \frac{2b_1(1-\rho)}{1+b_1} Z^2)(s, x)\phi^2(\varepsilon x) = \mathcal{E}_\varepsilon(t).
\]

Inserting this into (5.10), together with

\[
\frac{2b_1(1-\rho)}{1+b_1} \left( \frac{1}{1-b_1} = \frac{D_s\kappa}{\nu_s^2} \right),
\]

we now arrive at

\[
\langle M_\phi \rangle(t) - \frac{D_s\kappa^2}{\nu_s^2} \varepsilon^2 \sum_{s=0}^{t-1} \varepsilon \sum_{x \in \Xi(s)} \phi^2(\varepsilon x) Z^2(s, x) = \mathcal{E}_\varepsilon(t).
\]

(5.11)

So far, we have only shown that the expression (5.11) converges to zero (in \( L^2 \)) pointwise in \( t \), (i.e., (5.9)). Given the moment bounds from Proposition 5.1, a standard argument leverages such pointwise convergence to convergence at process level, yielding

\[
\sup_{t \in \mathbb{Z} \cap [0, \varepsilon^{-2}T]} \left| \langle M_\phi \rangle(t) - \frac{D_s\kappa^2}{\nu_s^2} \varepsilon^2 \sum_{s=0}^{t-1} \varepsilon \sum_{x \in \Xi(s)} Z^2(s, x)\phi^2(\varepsilon x) \right| \rightarrow_p 0.
\]

This, together with the discrete-to-continuous approximation argument from [CT17, Proof of Proposition 2.13], verifies that (5.6) is a local martingale.

5.1.2. Part (b): step initial condition. Consider \( \hat{Z}(t, x) := \frac{\rho(1-\rho)}{\sqrt{\varepsilon}} Z(t, x) \) under the step initial condition \( N(0, x) = (x)_+ \). From (4.3),

\[
\hat{Z}(t, x) = \begin{cases} \frac{\rho(1-\rho)}{\sqrt{\varepsilon}} e^{-\sqrt{\varepsilon}(1-\rho)} & \text{for } x \geq 0, \\ \frac{\rho(1-\rho)}{\sqrt{\varepsilon}} e^{-\sqrt{\varepsilon}x} & \text{for } x < 0. \end{cases}
\]

In particular \( \sum_{x \in \mathbb{Z}} \hat{Z}(0, x) = \frac{\rho(1-\rho)}{\sqrt{\varepsilon}} (\frac{1}{1-e^{-\sqrt{\varepsilon}(1-\rho)}} + \frac{e^{-\sqrt{\varepsilon}0}}{1-e^{-\sqrt{\varepsilon}0}}) \rightarrow 1 \). This together with the exponential decay (in \( |x| \)) of \( \hat{Z}(0, x) \) shows that \( \hat{Z}(0, \varepsilon^{-1}x) \) converges to \( \delta(x) \). With this and the convergence result for near stationary initial conditions (i.e., part (a)), following the argument of [ACQ11, Section 3], part (b) is an immediate consequence of the following moment bounds: for any given \( T < \infty \), \( n \in \mathbb{Z}_{>0} \), and \( \alpha \in (0, \frac{1}{2}) \), there exists \( C = C(T, n, \alpha) < \infty \) such that

\[
\| \hat{Z}(t, x) \|_n \leq \frac{C}{(\varepsilon^{-2}t)^{\frac{\alpha}{2}}}, \quad \| \hat{Z}(t, x) - \hat{Z}(t, x') \|_n \leq \frac{C(\varepsilon|x-x'|)^{\alpha}}{(\varepsilon^{-2}t)^{\frac{\alpha}{2}}},
\]

(5.12)

for all \( t \in (0, \varepsilon^{-2}T] \), \( x, x' \in \mathbb{R} \). Given the estimates on \( \rho^t \) in the proof of Proposition 5.1, the bounds (5.12) is established by the same types of calculations as in [CT17, Proof of Proposition 2.14] and [ACQ11, Section 3]. We omit the details here.
5.2. Proof of Theorem 1.6. Recall that Proposition 1.5 asserts an extension of the stationary solution of the SBE to all values of \( t > -\infty \). We begin by giving this construction.

Proof of Proposition 1.5. The construction of \( \mathcal{K} \) follows standard, Kolmogorov-type argument. To begin with, given (1.20), we have that
\[
(\mathcal{H}_{\text{stat}}(t, \cdot) - \mathcal{H}_{\text{stat}}(t, 0))_{t \geq 0} =: (\tilde{\mathcal{K}}(t, \cdot))_{t \geq 0}
\]
is a stationary (in \( t \)) process. Consider the space \( \mathcal{X} := \prod_{\mathbb{R}} C(\mathbb{R}) \), endowed with the product \( \sigma \)-algebra and with the product topology. For each \( t_1 < \ldots < t_n \in \mathbb{R} \), we define a probability distribution \( \mathbb{P}_{t_1, \ldots, t_n} \) on \( \prod_{\{t_1, \ldots, t_n\}} C(\mathbb{R}) \) given by that of
\[
(\tilde{\mathcal{K}}(0, \cdot), \tilde{\mathcal{K}}(t_2 - t_1, \cdot), \ldots, \tilde{\mathcal{K}}(t_n - t_1, \cdot)).
\]

Thanks to the stationarity of \( \mathcal{K}(t, \cdot) \), the laws \( \mathbb{P}_{t_1, \ldots, t_n} \) are consistent among \( \{t_1 < \ldots < t_n\} \in \mathbb{R} \). Thus, the Kolmogorov extension theorem gives an \( \mathcal{X} \)-valued process \( \tilde{\mathcal{K}}(t, x) \), such that, for any \( t_0 \in \mathbb{R} \),
\[
\tilde{\mathcal{K}}(t - t_0, \cdot) = \mathcal{H}_{\text{stat}}(t, \cdot) - \mathcal{H}_{\text{stat}}(t, 0), \quad \text{in finite dimensional (in } t \text{) distributions.} \tag{5.13}
\]

The next step is to further construct a continuous version of \( \tilde{\mathcal{K}} \). That is, a \( C(\mathbb{R}, C(\mathbb{R})) \)-valued process that shares the same finite dimensional (in \( t \)) distributions as \( \tilde{\mathcal{K}}(t, x) \). To this end, for each \( n \in \mathbb{Z}_{\geq 0} \), we construct a \( C(\mathbb{R}, C(\mathbb{R})) \)-valued process \( \mathcal{K}_n \) by setting \( \mathcal{K}_n(\frac{i}{2^n}, x) := \tilde{\mathcal{K}}(\frac{i}{2^n}, x) \), for \( i \in \mathbb{Z} \), and linearly interpolate in \( t \). For such dyadic approximations, given any fixed \( [t_1, t_2] \times [x_1, x_2] := D \subset \mathbb{R}^2 \), we have that
\[
\sup_{(t,x) \in D} \left| \mathcal{K}_n(t, x) - \mathcal{K}_{n+m}(t, x) \right| \\
\leq \sup \left\{ \left| \tilde{\mathcal{K}}(t, x) - \tilde{\mathcal{K}}(s, x) \right| : s, t \in [t_1, t_2] \cap 2^{-(m+n)}\mathbb{Z}, \ |t - s| \leq 2^{-n}, \ x \in [x_1, x_2] \right\}.
\]

As \( \mathcal{H}_{\text{stat}} \) is continuous, with (5.13), we see that the r.h.s. converges to zero in distribution (and hence converges to zero in probability) as \( (n, m) \to (\infty, \infty) \). This being the case, using the first Borel–Cantelli lemma, it is standard to construct a subsequence of \( \{\mathcal{K}_n\}_n \) that is almost surely Cauchy in \( C(\mathbb{R}, C(\mathbb{R})) \). The resulting limiting process \( \mathcal{K} \in C(\mathbb{R}, C(\mathbb{R})) \) gives the desired continuous version of \( \tilde{\mathcal{K}} \). With \( \mathcal{K} \) and \( \mathcal{H}_{\text{stat}} \) both being continuous, the desired property (1.21) follows from (5.13). \( \square \)

We now prove Theorem 1.6.

Proof of Theorem 1.6. Recall the definition of \( \| \cdot \|_{C^{-1}(\mathbb{R}^2), [-\ell, \ell]^2} \) from (1.17). Referring to (1.18), we see that \( U_x \to U \) in \( C^{-1}(\mathbb{R}^2) \), if and only if, for any fixed \( \ell \in \mathbb{Z}_{>0} \), \( \|U_x - U\|_{C^{-1}(\mathbb{R}^2), [-\ell, \ell]^2} \to 0 \). With this in mind, we henceforward fix \( \ell \in \mathbb{Z}_{>0} \). Further, even though the relevant test functions in (1.17) have support in \( [-\ell, \ell]^2 \), with both \( \mathcal{U} \) and stochastic Gibbs state being translation invariant in \( y \), after a suitable translation, we assume without lost of generality that relevant test functions are supported in \( (x, y) \in [-\ell, \ell] \times [0, 3\ell] \).

The next step is to translate the statements regarding the symmetric model into the context of the stochastic model. Recall that, for a given \((a, b, c)\)-symmetric 6V model with \( \Delta > 1 \), defining \( b_1, b_2 \in (0, 1) \) by inverting the relation (1.10), the stochastic Gibbs state \( \mathcal{S}_\rho(b_1, b_2; h, v) \) for the \((a, b, c)\)-symmetric model is equivalent to the \((b_1, b_2)\)-stochastic model in its stationary measure. Here \((h, v) \in (0, 1)^2 \) is an one parameter family of parameters satisfying (1.13), and the corresponding stationary measure for the \((b_1, b_2)\)-stochastic model is the product Bernoulli measure \( \bigotimes_{x \in \mathbb{Z}} \text{Ber}(\rho) \) with \( \rho := v \). While for the symmetric model we have used coordinates \((x, y)\) for the \( x \) and \( y \) axes, for the stochastic model we will use \((x, t)\) with \( y \) replaced by \( t \) to represent the temporal axis. We will also tend to write these coordinates as \((t, x)\) with time first and then space. The purpose of the shifting in \( y \) described above is to ensure that \( t \geq 0 \) for the stochastic model.
Recall from (1.14)–(1.15) that \( u(x, y) \) denote the indicator of an incoming vertical line, and that \( U_{\varepsilon} \) is the corresponding empirical measure. Under this mapping between the symmetric and stochastic models, the former becomes the occupation variable \( u(x, y) = 1_{\{\text{having a particle at } (t = y, x)\}} = \eta(y, x) \).

Fix \( f \in C^\infty(\mathbb{R}^2) \) with support \( (x, y) \in [-\ell, \ell] \times [0, 3\ell] \). With \( \eta(y, x) := N(y, x) - N(y, x - 1) \), we have

\[
\langle U_{\varepsilon}, f \rangle = \varepsilon^2 \sum_{x, y \in \mathbb{Z}} (N(y, x) - N(y, x - 1) - \rho) f(\varepsilon^{-1} x - \mu \varepsilon^{-2} y, \varepsilon^{-2} y). \tag{5.14}
\]

From Theorem 1.1, we know that the centered scaled height function

\[
\tilde{N}(t, x) := \sqrt{\varepsilon}\left(N_{\varepsilon}(\varepsilon^{-2} t, \varepsilon^{-1} x + \mu \varepsilon^{-2} t) - \rho (\varepsilon^{-1} x + \mu \varepsilon^{-2} t) - \varepsilon^{-2} t \log \lambda_{\varepsilon}\right),
\]

converges. In (5.14), we can substitute in \( \tilde{N} \) for \( N \), switch from the \((x, y)\)-coordinates to \((t, x)\), and apply summation by parts in \( x \). This gives

\[
\langle U_{\varepsilon}, f \rangle = -\varepsilon^2 \sum_{t \in \mathbb{Z}_{\geq 0}} \left( \varepsilon \sum_{x \in \Xi(t)} \varepsilon^{-1}(\tilde{N}(t, x) - \tilde{N}(t, x - \varepsilon)) f(x, t) \right)
= -\varepsilon^2 \sum_{t \in \mathbb{Z}_{\geq 0}} \left( \varepsilon \sum_{x \in \Xi(t)} \tilde{N}(t, x)(\varepsilon(f(x + \varepsilon, t) - f(x, t))) \right). \tag{5.15}
\]

The last expression is indeed similar to \( \langle U, f \rangle \) defined in (1.22), with integrations replaced by sums, and derivative on \( f \) replaced by difference. Recall that \( \eta(0, x) \) is linearly interpolated onto \( (t, x) \in \mathbb{R}^+ \times \mathbb{R} \) to give a \( C(\mathbb{R}^+, C(\mathbb{R})) \)-valued process. This being the case, we further write

\[
\langle U_{\varepsilon}, f \rangle = -\int_0^{\infty} \int_{\mathbb{R}} \tilde{N}(t, x) \partial_x f(x, t) dx dt + A_{\varepsilon}(t, x),
\]

where \( A_{\varepsilon}(t, x) \) denotes a residue term with \( |A_{\varepsilon}(t, x)| \leq \sqrt{\varepsilon}C(\ell)(||f||_\infty + ||\partial_x f||_\infty) \).

Recall that, here, the stochastic model starts from Bernoulli initial condition

\[
(\eta(0, x))_x \sim \bigotimes_{x \in \mathbb{Z}} \text{Ber}(\rho), \quad N(0, x) := \sum_{y \in [0, x]} (\eta(0, y) - \rho).
\]

It is standard to check that such an initial condition indeed satisfies the conditions in Definition 4.4. Further, as \( \varepsilon \to 0 \), we have \( \tilde{N}(0, \cdot) = \sqrt{\rho(1 - \rho)}B(\cdot) \) in \( C(\mathbb{R}) \), where \( B \) denotes a standard Brownian motion. Given these properties, Theorem 1.1 asserts that

\[
\tilde{N}(\cdot, \cdot) \Rightarrow \mathcal{H}_{\text{stat}}(\cdot, \cdot), \quad \text{in } C(\mathbb{R}^+, C(\mathbb{R})).
\]

By Skorokhod’s representation theorem, we further assume that this convergence holds in probability under a suitable coupling of \( \tilde{N} \) and \( \mathcal{H}_{\text{stat}} \), whereby

\[
\sup_{t \in [0, 3\ell]} \sup_{x \in [-\ell, \ell]} |\tilde{N}(t, x) - \mathcal{H}_{\text{stat}}(t, x)| \longrightarrow p 0. \tag{5.16}
\]

Recall that \( f_{\delta}(x, y) := f(\delta^{-1} x, y) \). Now, under the aforementioned coupling, take the difference of (1.22) and (5.15), and replace \( f \) with \( f_{\delta} \). This gives

\[
|\langle U_{\varepsilon} - U, f_{\delta} \rangle| \leq ||\partial_x f_{\delta}||_\infty \sup_{t \in [0, 3\ell]} \sup_{x \in [-\ell, \ell]} |\tilde{N}(t, x) - \mathcal{H}_{\text{stat}}(t, x) + C(\ell)\varepsilon(\partial_x f_{\delta})_\infty + \|f\|_\infty)
= \delta^{-1}||\partial_x f||_\infty \sup_{t \in [0, 3\ell]} \sup_{x \in [-\ell, \ell]} |\tilde{N}(t, x) - \mathcal{H}_{\text{stat}}(t, x) + C(\ell)\varepsilon(\delta^{-1}||\partial_x f||_\infty + \|f\|_\infty).
\]

As this holds true for all \( f \in C^\infty(\mathbb{R}^2) \) with \( \text{supp}(f) \subset [\ell, \ell] \times [0, 3\ell] \), referring to (1.17), we see that

\[
||U_{\varepsilon} - U||_{C^{-1}(\mathbb{R}^2), [-\ell, \ell]^2} \leq \sup_{t \in [0, 3\ell]} \sup_{x \in [-\ell, \ell]} |\tilde{N}(t, x) - \mathcal{H}_{\text{stat}}(t, x)| + C(\ell)\varepsilon.
\]
Taking $\varepsilon \to 0$, we thus conclude $\|U_\varepsilon - U\|_{C^{-1}([-\ell,\ell]^2)} \to 0$. This being true for arbitrary $\ell \in \mathbb{Z}_{>0}$, we conclude the desired result: $d_{C^{-1}(\mathbb{R}^2)}(U_\varepsilon, U) \to 0$.

6. Estimating the two-point semigroup

Recall from (4.11) that $V_\varepsilon$ denotes the semigroup for the two-point functions of $Z$, where we put $\varepsilon$ in the notation of $V_\varepsilon$ to emphasize the dependence. In order to complete the proof of Theorem 1.1*, it remains to prove Proposition 5.3. The proof will be carried out in Section 7 with the aid of duality. Key to this proof is certain estimates on $V_\varepsilon$ and its gradients, which are the subjects of this section.

Recall that $\nabla f(x) := f(x + 1) - f(x)$ denotes discrete gradient. In the sequel we use notation such as $\nabla_x$ to highlight the variable on which the gradient acts. Recall that $V_\varepsilon((y_1, y_2), (x_1, x_2); t)$ is related to the stochastic $6V$ model only within the Weyl chamber: $x_1 < x_2$. Thus, for expressions such as

\[ \nabla_{x_1} V_\varepsilon((y_1, y_2), (x_1, x_2); t) = V_\varepsilon(\{(y_1, y_2), (x_1 + 1, x_2); t\}) - V_\varepsilon(\{(y_1, y_2), (x_1, x_2); t\}) \]

to be relevant, we must impose an additional constraint $x_1 + 1 < x_2$. In this case we say $(x_1, x_2, y_1, y_2)$ is in the $\nabla$-Weyl chamber, which is understood with respect to whichever gradient is taken.

The goal of this section is to establish:

**Proposition 6.1.** For any $\alpha, T \in (0, \infty)$, there exist constants $C(\alpha, T), C(\alpha) > 0$ such that

\[
|V_\varepsilon((y_1, y_2), (x_1, x_2); t)| \leq \frac{C(\alpha, T)}{t + 1} e^{-\alpha |x_1 - y_1| + |x_2 - y_2|},
\]

\[
|\nabla_{x_j} V_\varepsilon((y_1, y_2), (x_1, x_2); t)|, \quad |\nabla_{y_j} V_\varepsilon((y_1, y_2), (x_1, x_2); t)| \leq \frac{C(\alpha, T)}{(t + 1)^{3/2}} e^{-\alpha |x_1 - y_1| + |x_2 - y_2|},
\]

for all $x_1 < x_2 \in \mathbb{E}(t + s)$, $y_1 < y_2 \in \mathbb{E}(s)$, $s, t \in [0, \varepsilon^{-2} T] \cap \mathbb{Z}$, $j = 1, 2$, and $(x_1, x_2, y_1, y_2)$ in their respective Weyl or $\nabla$-Weyl chamber.

In proving Proposition 6.1, it is convenient to consider ‘small $t$’ and ‘large enough $t$’ separately. More precisely, in the following we use the phrase for large enough $t$ if the referred statement holds for all $t \geq t_0$, for some generic threshold $t_0 < \infty$ that may change from line to line, but depends only on $\alpha$ and $T$. This is not to be confused with the global assumption $t \leq \varepsilon^{-2} T$.

The case with $t \leq t_0$ is simple. Let us first settle it.

**Proof of Proposition 6.1, the case with $t \leq t_0 = t_0(\alpha, T)$.** Fix an arbitrary $t_0 < \infty$, and assume $t \leq t_0$ throughout the proof. Since $(t + 1)$ is bounded away from zero and infinity, it suffices to show

\[ |V_\varepsilon((y_1, y_2), (x_1, x_2); t)| \leq C(t_0) e^{-\frac{1}{t_0}|(x_1 - y_1) + |x_2 - y_2|} \].

(6.1)

From this the desired estimates on $|V|$ and $|\nabla V|$ both follow.

Instead of directly proving this bound for $V_\varepsilon$, let us first consider $U$ and prove that

\[ |U((y_1, y_2), (x_1, x_2); t)| \leq C(t_0) e^{-\frac{1}{t_0}|(x_1 - y_1) + |x_2 - y_2|} \].

(6.2)

Recall from Proposition 3.5 that $U((y_1, y_2), (x_1, x_2); t) = \mathbb{P}_{6V}^{(y_1, y_2)}((y_1, y_2) \rightarrow (x_1, x_2); t)$ denotes the transition probability of stochastic $6V$ particle system with two particles. Here we will appeal to the probabilistic interpretation of $U = \mathbb{P}_{6V}^{(y_1, y_2)}$ and not rely upon contour integral formulas. Let $(x_1(t) < x_2(t)) \in \mathbb{Z}^2$ denote the time $t$ locations of the particles, starting from $x_i(0) = y_i$. To show (6.2), it suffices to show such a statement with $t = 1$. To see this, observe that $U((y_1, y_2), (x_1, x_2); t)$ can be written as a $t$-fold convolution of one-step transition probabilities. The convolution can be expanded into a sum over all trajectories $(x_1(\cdot), x_2(\cdot))$ with $x_i(0) = y_i$ and $x_i(t) = x_i$. The contribution to each trajectory can be bounded by $t$ products of the one-step bound, leading to the contribution $C^t e^{-\frac{1}{t_0}|(x_1 - y_1) + |x_2 - y_2|}$ for some $C > 0$. (Note that the exponential terms came from telescoping.) The total number of trajectories to sum over is upper-bounded by $(x_1 - y_1)^t(\varepsilon^{-2} T)^t$ which, for $t < t_0$, is bounded by $C(t_0)|x_1 - y_1|^t|x_2 - y_2|^t$. Combining these two bounds and using that $t < t_0$, we arrive
The $t = 1$ version of (6.2) is easy shown directly from the definition of the dynamics of the stochastic 6V model. Finally, recall that $V_\varepsilon$ is related to $U$ through (4.17). Given that $\lambda_\varepsilon \to 1$, $\mu_\varepsilon \to 1$, $\tau_\varepsilon \to 1$, and $t \leq t_0$, the preceding bound on $|U|$ immediately yields the desired result (6.1).

Having settled Proposition 6.1 for short time, we now turn to the case for large enough $t$. For this we appeal to the contour integral representation, and analyze the integrals therein. To begin with, considering separately the four cases distinguished by the signs of $F$ and $\varepsilon$, we breakdown the proof of Proposition 6.1 into proving:

**Proposition 6.2.** For any $\alpha, T \in (0, \infty)$ and $t_0 = t_0(\alpha, T)$, there exist $C(\alpha, T), C(\alpha) > 0$ such that

(a) $\left| V_{\varepsilon}^f((y_1, y_2), (x_1, x_2); t) \right| \leq \frac{C(\alpha, T)}{t + 1} e^{-\alpha|x_1-y_1|+|x_2-y_2|\sqrt{t+1+C(\alpha)}}$,

(b) $\left| \nabla_{x_1} V_{\varepsilon}^f((y_1, y_2), (x_1, x_2); t) \right|, \left| \nabla_{y_1} V_{\varepsilon}^f((y_1, y_2), (x_1, x_2); t) \right| \leq \frac{C(\alpha, T)}{(t+1)^{3/2}} e^{-\alpha|x_1-y_1|+|x_2-y_2|\sqrt{t+1+C(\alpha)}}$,

(c) $\left| V_{\varepsilon}^{in}((y_1, y_2), (x_1, x_2); t) \right| \leq \frac{C(\alpha, T)}{t + 1} e^{-\alpha|x_1-y_1|+|x_2-y_2|\sqrt{t+1+C(\alpha)}}$,

(d) $\left| \nabla_{x_1} V_{\varepsilon}^{in}((y_1, y_2), (x_1, x_2); t) \right|, \left| \nabla_{y_1} V_{\varepsilon}^{in}((y_1, y_2), (x_1, x_2); t) \right| \leq \frac{C(\alpha, T)}{(t+1)^{3/2}} e^{-\alpha|x_1-y_1|+|x_2-y_2|\sqrt{t+1+C(\alpha)}}$,

for all $x_1 < x_2 \in \Xi(t + s), y_1 < y_2 \in \Xi(s), s \in [0, e^{-2T}] \cap \mathbb{Z}, t \in [t_0, e^{-2T}] \cap \mathbb{Z}, j = 1, 2$, and $(x_1, x_2, y_1, y_2)$ in their respective Weyl or $\nabla$-Weyl chamber.

Note that in Proposition 6.2(c)–(d), the pairing of $x_i$'s and $y_j$'s is swapped compared to Proposition 6.1. This arises naturally from the contour integral structure of $V_{\varepsilon}^{in}$, and in fact gives a stronger bound than the one in the original pairing. To see this, under the assumption $x_1 < x_2$ and $y_1 < y_2$, considering separately the four cases distinguished by the signs of $x_1 - y_1$ and $x_2 - y_2$, we check that

\[
|x_1 - y_1| + |x_2 - y_2| = (x_1 - y_1) + (x_2 - y_2) = (x_1 - y_1) + (x_2 - y_1) \leq |x_1 - y_2| + |x_2 - y_1|,
\]

\[
|x_1 - y_1| + |x_2 - y_2| = (y_1 - x_1) + (y_2 - x_2) = (y_1 - x_2) + (y_2 - x_1) \leq |y_1 - x_2| + |y_2 - x_1|,
\]

\[
|x_1 - y_1| + |x_2 - y_2| = (x_1 - y_1) + (y_2 - x_2) \leq (x_1 - y_1) + (y_2 - x_1) \leq |x_2 - y_1| + |y_2 - x_1|,
\]

\[
|x_1 - y_1| + |x_2 - y_2| = (y_1 - x_1) + (x_2 - y_2) \leq (y_1 - x_1) + (x_2 - y_1) \leq |y_2 - x_1| + |x_2 - y_1|.
\]

Throughout the rest of this section, we fix an exponent $\alpha \in (0, \infty)$, a time horizon $T \in (0, \infty)$, and assume $t \leq e^{-2T}$ is large enough. In the sequel we will frequently use polar coordinates $z = re^{i\theta}$ to parametrize complex numbers. Throughout this section we will operate under convention $\theta \in (-\pi, \pi)$.

### 6.1. Estimating the Free Part $V_{\varepsilon}^f$

Let us explain the strategy before starting the estimate. We plan to deform $\mathcal{C}_r \times \mathcal{C}_r$ to some suitable contours, along which we easily extract the spatial exponential decay. To this end, for $\beta \in \mathbb{R}$ set

\[
u(t, \beta) := \exp\left(\frac{\beta}{\sqrt{t+1+C(\alpha)}}\right).
\]

We fixed the constant $C_\ast \in (0, \infty)$ large enough so that $\nu(t, \beta) \geq \exp(-1/C_\ast) \geq 1 + \frac{\beta}{2}$. This is to avoid the pole of $\mathcal{D}_\varepsilon(t, z)$ (given in (4.13)) at $z = b_t e^{\sqrt{\theta}(\rho-1)}$. Now, let $\text{sgn}(x) := 1_{\{x > 0\}}$ denote the sign.
function, and let
\[ r_i = u(t, -\text{sgn}(x_i - y_i)\alpha) \]
where \( \alpha \in (0, \infty) \) is the parameter given in Proposition 6.2. Along the contour \((z_1, z_2) \in C_{r_1} \times C_{r_2}\), we have the desired exponential decay:
\[ |z_i|^x_i y_i = \exp\left(-\frac{\alpha|x_i - y_i|}{\sqrt{t + 1 + \alpha C_1}}\right). \]

Given the exponential decay, we still need to show that each of the remaining integrals (for \( i = 1, 2 \))
\[ \int_{-\pi}^{\pi} |D(t, z_i(\theta_i))| \frac{d\theta_i}{2\pi |z_i(\theta_i)|} \]
are bounded by \((t + 1)^{-\frac{1}{2}}C\). This is achieved by steepest decent analysis. Under weak asymmetry scaling, the function \( D_\varepsilon(t, z) \) (given in (4.13)) reads
\[ D_\varepsilon(t, z) = z^{[\mu \varepsilon t]} \left( \frac{1 - b_1 e^{\sqrt{\rho}(t - 1)}}{b_1 + e^{\sqrt{\rho} - b_1 e^{\sqrt{\rho} - b_1 e^{\sqrt{\rho}(t - 1) - 1}}} - 1 - b_1 e^{\sqrt{\rho}(t - 1) - 1}} \right)^t. \] (6.6)

As we show in Lemma 6.3 below, along the contour \( C_{r_1} \), under the polar parametrization \( z_i = r_i e^{i\theta_i} \),
- \( |D_\varepsilon(t, z_i(\theta_i))| \) has Gaussian decay in \( \theta_i \) of the form \( \exp(-\frac{1}{\varepsilon} \theta_i^2 (t + 1)) \) in a neighborhood of \( \theta_i = 0 \),
- \( |D_\varepsilon(t, z_i(\theta_i))| \) has an exponential decay in \( t \) of the form \( \exp(-\frac{1}{\varepsilon} (t + 1)) \) away from \( \theta_i = 0 \).

The first bullet point is done by Taylor expansion, and relies only on local properties of \( D_\varepsilon(t, z_i) \) and \( C_{r_1} \) near \( \theta_i = 0 \). The second bullet point holds because of global properties of \( D(t, z_i) \). More precisely, set
\[ D_\varepsilon(z) := \frac{b_1 z + 1 - 2b_1}{1 - b_1/z}. \] (6.7)

Referring to the definition (4.13) of \( D_\varepsilon(t, z) \), with \( \mu \varepsilon \to 1 \) as \( \varepsilon \to 0 \), we have that
\[ \lim_{(t, z) \to (\infty, 0)} |D_\varepsilon(t, z)|^{\frac{1}{t}} = |D_\varepsilon(z)|, \quad \text{uniformly over } z \in C_1. \]

Now, with \( r_i = u(t, \pm \alpha) \to 1 \) as \( t \to \infty \), we see that the second bullet point cannot hold unless
\[ |D_\varepsilon(z)| < 1, \quad \forall z \in C_1 \setminus \{1\}. \] (SD.C1)

Conditions of the type (SD.C1) will turn out to be decisive in showing that steepest decent analysis works. (SD.C1) can be verified by interpreting \( D_\varepsilon(z) \) as a probability generating function \( \mathbb{E}[z^{\lambda \varepsilon t}] \) of a random variable \( X \). We will, instead, verify (SD.C1) (Lemma 6.3.1) by viewing \( D_\varepsilon(z) \) as a rational function and directly calculating its modulus along the unit circle \( C_1 \). This approach has the advantage of generalizing to the case for the interacting part \( V_{\varepsilon}^{in} \).

We now begin the estimate, starting with the steepest-decent-type bound on \( |D_\varepsilon(t, z)| \).

**Lemma 6.3.** Given any \( \beta \in \mathbb{R} \) and \( T < \infty \), there exists \( C(\beta, T), C > 0 \) such that
\[ |D_\varepsilon(t, z)| \leq C(\beta, T) \exp\left(-\frac{1}{\varepsilon} \theta^2 (t + 1)\right), \quad \text{with } z = u(t, \beta) e^{i\theta} \in C_{u(t, \beta)}, \]
for all \( \theta \in (-\pi, \pi] \), large enough \( t \leq T \), and small enough \( \varepsilon > 0 \).

**Proof.** Our first step is to recognize \( D_\varepsilon(t, z) \) as the \( t \)-th power of a given function. To this end, referring to (4.13), observe that
\[ D_\varepsilon(z) := D_\varepsilon(t, z)^t = z^{[\mu \varepsilon t]} \lambda_\varepsilon \frac{b_1 + (1 - b_1 - b_1^2)(\tau_\varepsilon z)^{-1}}{1 - b_1^2(\tau_\varepsilon z)^{-1}}. \]

Indeed, \( D_\varepsilon(z) \) has a \( t \)-dependence through \( z^{[\mu \varepsilon t]} \), but since \( \mu \varepsilon \to 1 \) as \( \varepsilon \to 0 \), we expect the \( t \)-dependence to be ‘weak’ and hence suppress it in notation. Due to the non-integer power \( z^{[\mu \varepsilon t]} \), the function \( D_\varepsilon(z) \) is not meromorphic on \( C \). However, since \( \mu \varepsilon \to 1 \) as \( \varepsilon \to 0 \), there exists a fixed neighborhood \( O \) of \( z = 1 \), such that \( D_\varepsilon(z) \) is analytic on \( z \in O \). Throughout the proof we will operate on \( O \) whenever referring to the function \( D_\varepsilon(z) \).
As in the statement of Lemma 6.3, set \( z(\theta) = u(t, \beta)e^{i\theta} \). The proof follows a three-step procedure:

(Zero \( \theta \)) Show that \( |\mathcal{D}_\epsilon(z(0))| \leq \exp(C(\beta, T)\frac{1}{t+1}) \), for all \( t \leq \epsilon^{-2}T \) large enough and \( \epsilon > 0 \) small enough.

Note that the right hand side of this bound also ‘weakly’ depends on \( t \) for \( t \) sufficiently large.

(Small \( \theta \)) Show that there exists \( \theta_0 > 0 \), such that \( |\mathcal{D}_\epsilon(z(\theta))| \leq |\mathcal{D}_\epsilon(z(0))|\exp(-\frac{\epsilon^2}{C}) \), for all \( |\theta| \leq \theta_0 \), and \( \epsilon > 0 \) small enough.

(Large \( \theta \)) Show that \( |\mathcal{D}_\epsilon(t, z(\theta))| \leq \exp(-\frac{1}{t}) \), for \( |\theta| > \theta_0 \), \( t \geq 0 \) large enough, and \( \epsilon > 0 \) small enough.

Once these have been established, with \( \mathcal{D}_\epsilon(t, z) = \mathcal{D}_\epsilon(z)^t \), the desired result follows immediately. Our task is hence to carry out the steps (Zero \( \theta \)), (Small \( \theta \)), and (Large \( \theta \)).

(Zero \( \theta \)): First, since the function \( \mathcal{D}_\epsilon(z) \) is invoked here, let us check that the claimed assumption \( z(0) \in O \) holds. Indeed, with \( u(t, \beta) \to 1 \) as \( t \to \infty \), we have that \( z(0) \in O \), for all \( t \) large enough.

Recall that \( R_\epsilon := S'_\epsilon - \mu_\epsilon \), and that \( S'_\epsilon \) is defined in (4.1)–(4.2) with \( \mu_\epsilon = \mathbb{E}(S'_\epsilon) \). One readily checks that \( \mathcal{D}_\epsilon(z) = z^{\frac{|\mu_\epsilon t|}{t}} - \mu_\epsilon \mathbb{E}[z^{-R_\epsilon}], z \in O \). Given this, it is straightforward to calculate

\[
\partial_z \left( \log \mathcal{D}_\epsilon(z) \right) = \frac{|\mu_\epsilon t|}{t} - \mu_\epsilon - \frac{\mathbb{E}[R_\epsilon z^{-R_\epsilon} - 1]}{\mathbb{E}[z^{-R_\epsilon}]} \tag{6.8a}
\]

\[
\partial_z^2 \left( \log \mathcal{D}_\epsilon(z) \right) = \frac{\mathbb{E}[R_\epsilon(R_\epsilon + 1)z^{-R_\epsilon + 1}]}{\mathbb{E}[z^{-R_\epsilon}]} - \left( \frac{\mathbb{E}[R_\epsilon z^{-R_\epsilon - 1}]}{\mathbb{E}[z^{-R_\epsilon}]^2} \right)^2 \tag{6.8b}
\]

\[
\left| \partial_z^2 \left( \log \mathcal{D}_\epsilon(z) \right) \right| \leq C, \tag{6.8c}
\]

for all \( z \in O \). Using (6.8a)–(6.8c) we see that \( \left| \partial_z \left( \log \mathcal{D}_\epsilon(z) \right) \right|_{z=1} \leq |t^{-1}| \) and \( \left| \partial_z^2 \left( \log \mathcal{D}_\epsilon(z) \right) \right|_{z=1} \leq C \) for some \( C > 0 \). Using this, along with \( \log \mathcal{D}_\epsilon(1) = 0 \), we may Taylor expand around \( z = 1 \) and bound \( \left| \log \mathcal{D}_\epsilon(z) \right| \leq t^{-1}|z - 1| + C|z - 1|^2 \). Now, set \( z = z(0) = u(t, \beta) \), and use the fact that \( |u(t, \beta) - 1| \leq C(\beta, T)(t + 1)^{-1/2} \) to bound (after exponentiating)

\[
\mathcal{D}_\epsilon(z(0)) \leq \exp \left( (u(t, \beta) - 1| + C|u(t, \beta) - 1|^2 \right) \leq \exp(C(\beta, T)\frac{1}{t+1}).
\]

(Small \( \theta \)): First, with \( u(t, \beta) \to 1 \) as \( t \to \infty \), it is readily verified that there exists a small enough \( \theta_0 > 0 \) such that the assumption \( z(\theta) \in O \) holds for all \( |\theta| \leq \theta_0 \) and \( t \) large enough. From (6.8a)–(6.8c), we calculate (recall \( \nu_\ast \) from (1.6))

\[
\partial_\theta(\log \mathcal{D}_\epsilon(z(\theta)))|_{\theta=0} \in i\mathbb{R},
\]

\[
\lim_{\epsilon \to 0} \partial_\theta^2(\log \mathcal{D}_\epsilon(z(\theta)))|_{\theta=0} = -u(t, \beta)^2 \lim_{\epsilon \to 0} \text{Var}(R_\epsilon) \leq -\frac{1}{\epsilon} \nu_\ast,
\]

\[
\left| \partial_\theta^2(\log \mathcal{D}_\epsilon(z(\theta))) \right| \leq C.
\]

Given these properties, Taylor expanding \( \log \mathcal{D}_\epsilon(t, z(\theta)) \) in \( \theta \) around \( \theta = 0 \) to the second order yields

\[
\text{Re} \left[ \log \mathcal{D}_\epsilon(t, z(\theta)) - \log \mathcal{D}_\epsilon(t, z(0)) \right] \leq -\frac{1}{\epsilon} \theta^2, \quad |\theta| \leq \theta_0,
\]

for some fixed \( \theta_0 > 0 \). Further exponentiating this gives the desired result

\[
\left| \mathcal{D}_\epsilon(z(\theta)) \right| \leq \left| \mathcal{D}_\epsilon(z(0)) \right|e^{-\frac{1}{\epsilon} \theta^2}, \quad \forall |\theta| \leq \theta_0,
\]

and \( \epsilon > 0 \) small enough.

(Large \( \theta \)): Recall the definition of \( \mathcal{D}_\epsilon(z) \) from (6.7). With \( \mu_\epsilon \to 1 \) as \( \epsilon \to 0 \), referring to the expression (6.6) for \( \mathcal{D}_\epsilon(t, z) \), we readily verify that

\[
\lim_{(t, \epsilon) \to (\infty, 0)} \mathcal{D}_\epsilon(t, z(\theta)) = \mathcal{D}_\ast(e^{i\theta}), \quad \text{uniformly over } \theta \in (-\pi, \pi]. \tag{6.9}
\]

The r.h.s. of (6.9) leads us to want to show (SD.\( \mathcal{C}_1 \)). To verify (SD.\( \mathcal{C}_1 \)), we calculate

\[
\left| \mathcal{D}_\ast(e^{i\theta}) \right|^2 = \left( 1 + \frac{b_1(w + w^{-1} - 2)}{1 - b_1 w^{-1}} \right) \left( 1 + \frac{b_1(w^{-1} + w - 2)}{1 - b_1 w} \right)|_{w=e^{i\theta}}
\]

\[
= 1 + \frac{(w + w^{-1} - 2)(2b_1 + 2 - (b_1^2 + 1)(w + w^{-1}))}{|1 - b_1 w|^2}|_{w=e^{i\theta}}
\]
This calculation shows $|\mathcal{D}_\varepsilon(e^{i\theta})| < 1 - \frac{1}{c}$ for $|\theta| > \theta_0$. Combining with (6.9) gives the desired result:

$$
|\mathcal{D}_\varepsilon(t, z(\theta))|^{1/2} \leq 1 - \frac{1}{c}, \quad \forall |\theta| > \theta_0,
$$

for $t \leq \varepsilon^{-2}T$ large enough, and $\varepsilon > 0$ small enough.

\[\square\]

**Proof of Proposition 6.2(a)–(b).** As declared previously, we deform the contours $C_\varepsilon \times C_\varepsilon \mapsto C_{\varepsilon_1} \times C_{\varepsilon_2}$, where $r_i := u(-\text{sgn}(x_i - y_i)\alpha)$. With $r_i \geq \frac{1 + b_i}{2}$ as explained below (6.5), the deformation does not cross any pole, and gives

$$
V_{\varepsilon}^{r} = \frac{2}{2\pi i R_1} \int_{C_{\varepsilon_1}} \frac{z_i}{z_i - y_i + (\mu t - |\mu t|)}\mathcal{D}_\varepsilon(t, z_i)dz_i.
$$

(6.10)

Along the new contour $C_{\varepsilon_1}$, we have the desired exponential decay $|z_i|^{-\alpha y} = \exp\left(-\frac{\alpha|x_i - y_i|}{\sqrt{1 + \alpha C_\varepsilon}}\right)$. Hence, under the parametrization $z_i = r_i e^{i\theta_0}$, we have

$$
|V_{\varepsilon}^{r}| \leq e^{-\alpha(x_i - y_i + |x_i - y_i|)} \frac{2}{2\pi i R_1} \int_{-\pi}^{\pi} |\mathcal{D}_\varepsilon(t, z_i)|d\theta_i = C(\alpha, T) e^{-\alpha(x_i - y_i + |x_i - y_i|)} \frac{1}{(t + 1)^{3/2}}.
$$

This gives the desired estimate on $|V_{\varepsilon}^{r}|$.

As for the gradients, first note that taking $\nabla_{x_j}$ or $\nabla_{y_j}$ in (6.10) introduces a factor of $z_j^{\pm} - 1$, i.e.,

$$
\nabla_{x_j} V_{\varepsilon}^{r} = \frac{2}{2\pi i R_1} \int_{C_{\varepsilon_1}} \int_{C_{\varepsilon_2}} (z_j - 1) \frac{z_i}{z_i - y_i + (\mu t - |\mu t|)}\mathcal{D}_\varepsilon(t, z_i)dz_i,
$$

$$
\nabla_{y_j} V_{\varepsilon}^{r} = \frac{2}{2\pi i R_1} \int_{C_{\varepsilon_1}} \int_{C_{\varepsilon_2}} (z_j - 1) \frac{z_i}{z_i - y_i + (\mu t - |\mu t|)}\mathcal{D}_\varepsilon(t, z_i)dz_i.
$$

With $r_i = u(t, \pm\alpha)$, we have $|z_j^{\pm} - 1| \leq \frac{C(\alpha)}{t + 1} + \theta_j$ for $z_j = r_i e^{i\theta}$. Using this bound and the preceding procedure for bounding $|V_{\varepsilon}^{r}|$, we obtain the desired estimate on the gradients:

$$
|\nabla_{x_j} V_{\varepsilon}^{r}|, \quad |\nabla_{y_j} V_{\varepsilon}^{r}| \leq C(\alpha, T) e^{-\alpha(x_i - y_i + |x_i - y_i|)} \int_{\mathbb{R}^2} \left(\frac{1}{t + 1} + \theta_j\right) \frac{2}{(t + 1)^{3/2}}.
$$

\[\square\]

6.2. **Estimating the interacting part $V^{in}$, an overview.** In this subsection, we give an overview of the strategy for estimating $V^{in}_\varepsilon$. Compared to the estimate for $V^{r}_\varepsilon$, the major difference is that the expression $\mathcal{F}_\varepsilon(z_1, z_2)$ introduces a pole during contour deformations. More explicitly, under weak asymmetry scaling, $\mathcal{F}_\varepsilon(z_1, z_2)$ (defined in (4.12)) reads

$$
\mathcal{F}_\varepsilon(z_1, z_2) = \frac{1 + e^{\sqrt{1(2\rho - 1)}z_1}z_2}{1 + e^{\sqrt{1(2\rho - 1)}z_1}z_2} - \frac{e^{\sqrt{1(2\rho - 1)}z_1}z_2}{e^{\sqrt{1(2\rho - 1)}z_1}z_2}.
$$

(6.11)

This expression has a pole at $z_2 = p_\varepsilon(z_1)$, where

$$
p_\varepsilon(z) := (e^{\sqrt{1(\rho - 1)}} + e^{\sqrt{1(\rho - 1)}})z - e^{\sqrt{1(2\rho - 1)}}z^{-1}.
$$

(6.12)
For the variable $z_1$, we will devise a suitable contour $\Gamma(\mu,\varepsilon)$, on a case-by-case basis depending on the signs of $x_2 - y_1$. Starting with the expression (6.4), we deform the contours in two steps. First, with $z_2 \in C_\varepsilon$ being fixed, we deform the contour of $z_1$: $C_\varepsilon \mapsto \Gamma(\mu,\varepsilon)$. For the suitable $\Gamma(\mu,\varepsilon)$ so constructed in the sequel, we will check that

no pole is crossed during the deformation $z_1 \in C_\varepsilon \mapsto \Gamma(\mu,\varepsilon)$, if $\varepsilon$ is large enough. \hfill (No Pole)

In particular, here $\varepsilon$ must be so large that $C_\varepsilon$ contains $p_\varepsilon(\Gamma(\mu,\varepsilon))$. Next, for the $z_2$-contour, consider

$$r_2 := u(t, \text{sgn}(x_1 - y_2)k_2\alpha), \quad r_2' := u(t, \text{sgn}(x_1 - y_2)2k_2\alpha), \quad r_2'' := u(t, \text{sgn}(x_1 - y_2)3k_2\alpha),$$

where $k_2 \in \mathbb{Z}_{>0}$ is an auxiliary parameter, irrelevant for the general discussion in this subsection. With $z_1 \in \Gamma(\mu,\varepsilon)$ being fixed, we shrink the contour of $z_2$ from the large circle $C_\varepsilon$ to $C_{\tilde{r}_2(z_1)}$, where the radius $\tilde{r}_2(z_1)$ depends on the location of $p_\varepsilon(z_1)$, given by

$$\tilde{r}_2(z_1) := 1_{\{|p_\varepsilon(z_1)| \leq r_2\}}(r_2' \lor r_2'').$$

That is, for a fixed $z_1 \in \Gamma(\mu,\varepsilon)$, we examine the location of $p_\varepsilon(z_1)$, and if it sits outside of $C_{\tilde{r}_2}$, we shrink the large circle $z_2 \in C_\varepsilon$ to a smaller circle with radius $r_2' \lor r_2'' \leq r_2'$, otherwise shrink $C_\varepsilon$ to a circle with radius $r_2' \lor r_2'' > r_2'$.

During the second deformation $z_2 \in C_\varepsilon \mapsto C_{\tilde{r}_2(z_1)}$, we cross a pole at $z_2 = p_\varepsilon(z_1)$ if $r_2' < |p_\varepsilon(z_1)|$.

This is a simple pole from the term $\tilde{F}_\varepsilon(z_1, z_2)$, with

$$\text{Res}_{z_2 = p_\varepsilon(z_1)} \tilde{F}_\varepsilon(z_1, z_2) = \left( e^{\sqrt{\varepsilon}(\rho - 1)} + e^{\sqrt{-\varepsilon}\rho} \right) \left( \frac{p_\varepsilon(z_1)}{z_1} - 1 \right).$$

Set $J_\varepsilon(t, z) := D_\varepsilon(t, z)D_\varepsilon(t, p_\varepsilon(z))$ and

$$J(z) := z_1^{x_2 - y_1 - 1 + (\mu t - [\mu t])} p_\varepsilon(z_1)^{x_1 - y_2 + 1 + (\mu t - [\mu t])} - z_1^{x_2 - y_1 + (\mu t - [\mu t])} p_\varepsilon(z_1)^{x_1 - y_2 + (\mu t - [\mu t])}.$$

For each fixed $z_1 \in \Gamma(\mu,\varepsilon)$, applying the residue theorem to calculate the resulting expression after the deformation $z_2 \in C_\varepsilon \mapsto C_{\tilde{r}_2(z_1)}$, we have

$$V^{\text{in}}_\varepsilon = V_\text{blk} + V_\text{res},$$

where $V_\text{blk}$ and $V_\text{res}$ respectively contribute the ‘bulk’ and ‘residue’ parts of the deformed integral:

$$V_\text{blk} := \int_{\Gamma(\mu,\varepsilon)} \left( \int_{C_{\tilde{r}_2(z_1)}} z_2^{x_2 - y_1 + (\mu t - [\mu t]) - 1} p_\varepsilon(z_1)^{x_1 - y_2 + (\mu t - [\mu t])} \tilde{F}_\varepsilon(z_1, z_2) \frac{D_\varepsilon(t, z_2)dz_2}{2\pi i z_2} \right) \frac{D_\varepsilon(t, z_1)dz_1}{2\pi i z_1}, \quad \text{(6.16)}$$

$$V_\text{res} := \int_{\Gamma(\mu,\varepsilon)} 1_{\{|p_\varepsilon(z_1)| > r_2\}} (e^{\sqrt{\varepsilon}(\rho - 1)} + e^{\sqrt{-\varepsilon}\rho}) J(z) \frac{J_\varepsilon(t, z)dz_1}{2\pi i z_1 p_\varepsilon(z_1)}.$$ 

The integral in (6.16) is iterated because $\tilde{r}_2(z_1)$ depends on $z_1$.

Recall that $|\tilde{F}_\varepsilon(z_1, z_2)| = \infty$ at $z_2 = p_\varepsilon(z_1)$. By having $\tilde{r}_2(z_1)$ as in (6.14), we avoid the point $z_2 = p_\varepsilon(z_1)$ in the integral (6.16). More precisely, from (6.14), together with (6.5), we have that

$$|z_2 - p_\varepsilon(z_1)| \geq \left( |r_2' - r_2''| \lor |r_2' - r_2''| \right) \geq \frac{1}{\sqrt{\varepsilon} + \varepsilon}, \quad (z_1, z_2) \in \Gamma(\mu,\varepsilon) \times C_{\tilde{r}_2(z_1)}.$$

(Alternatively, one could also fix the radius $\tilde{r}_2(z_1) = r_2'$ for the $z_2$-contour. The resulting integrand in (6.16) in this case has a singularity at $z_2 = p_\varepsilon(z_1)$, which is integrable over $(z_1, z_2) \in \Gamma(\mu,\varepsilon) \times C_{\tilde{r}_2}$.

Proceeding this way however, requires elaborated estimates near the singularly jointly as $(t, \varepsilon)$ varies. We avoid doing so by constructing $\tilde{r}_2(z_1)$ in such a way that (6.18) holds.)

The contour $\Gamma(\mu,\varepsilon)$ needs be constructed in such a way that both $V_\text{blk}$ and $V_\text{res}$ are controlled by steepest decent analysis. In particular, a steepest decent condition analogous to (SD,C1) needs to hold here. To formulate the condition, assume that $\Gamma(\mu,\varepsilon)$ converges to a limiting contour $\Gamma_*$ as $(t, \varepsilon) \to (\infty, 0)$. Given $\lim_{\varepsilon \to 0} p_\varepsilon(z) = 2 - z^{-1}$ from (6.12), we define

$$J_\varepsilon(z) := D_\varepsilon(z)D_\varepsilon(2 - z^{-1}) = \frac{b_1 z + 1 - 2b_1}{1 - b_1 z^{-1}} + 1 - 2b_1 \frac{1}{1 - b_1/(2 - z^{-1})}.$$ 

(6.19)
The analogous steepest decent condition we must check here is
\[ |D_s(z)| < 1 \text{ for all } z \in \Gamma \setminus \{1\}, \quad |H_s(z)| < 1 \text{ for all } z \in \Gamma \setminus \{1\}. \]

(Figure 7) The figures show where the designated function is larger (darker) or smaller (lighter) than 1 in absolute value, for \( b_1 = 0.7 \). The unit circle is shown for comparison.

Figure 7 shows the region in \( \mathbb{C} \) where \( |D_s(z)| < 1 \) and where \( |H_s(z)| < 1 \), for \( b_1 = 0.7 \). In particular, we see that \( |H_s(z)| < 1 \) fails for a portion of the unit circle \( \mathcal{C}_1 \). This being the case, we need to devise a different type of contour than the contour \( \mathcal{C}_{r_1} \) in the preceding subsection. We begin with a prototype
\[ \mathcal{M} := \{ z : |z - \frac{1}{2}| = \frac{1}{2} \}. \]

This contour \( \mathcal{M} \) satisfies the steepest decent condition
\[ |D_s(z)| < 1 \text{ for all } z \in \mathcal{M} \setminus \{1\}, \quad |H_s(z)| < 1 \text{ for all } z \in \mathcal{M} \setminus \{1\}, \]
which we verify now.

Proof of (SD.M). First, express \( D_s(z) \) and \( H_s(z) \) (defined in (6.7) and (6.19)) as
\[
D_s(z) = \frac{b_1 z + 1 - 2b_1}{1 - b_1 z^{-1}} = 1 + \frac{b_1 z + b_1 z^{-1} - 2b_1}{1 - b_1 z^{-1}}, \\
H_s(z) = \frac{b_1 z + 1 - 2b_1}{1 - b_1 z^{-1}} = \frac{b_1(2 - z^{-1}) + 1 - 2b_1}{2 - b_1 - z^{-1}} = 1 + \frac{2b_1 z + 2b_1 z^{-1} - 4b_1}{2 - b_1 - z^{-1}}.
\]

under the parametrization \( z(\theta) := \frac{1 + e^{i\theta}}{2} \in \tilde{\mathcal{C}} \), we calculate
\[
|D_s\left(\frac{1 + e^{i\theta}}{2}\right)|^2 = \left(1 + \frac{b_1(w - 1)^2}{2(w + 1 - 2b_1)}\right)\left(1 + \frac{b_1(w^{-1} - 1)^2}{2(w^{-1} + 1 - 2b_1)}\right)|_{w=e^{i\theta}} = 1 + \frac{b_1(w - 2 + w^{-1})(2 - 3b_1)(w + w^{-1}) + 4 - 2b_1}{2(w^{-1} + 1 - 2b_1)^2}|_{w=e^{i\theta}} = 1 - \frac{b_1(1 - \cos \theta)(2 - b_1 + (2 - 3b_1) \cos \theta)}{|(w^{-1} + 1 - 2b_1)|^2}.
\]

\[
|H_s\left(\frac{1 + e^{i\theta}}{2}\right)|^2 = \left(1 + \frac{b_1(w - 1)^2}{(2 - b_1)w - b_1}\right)\left(1 + \frac{b_1(w^{-1} - 1)^2}{(2 - b_1)w^{-1} - b_1}\right)|_{w=e^{i\theta}} = 1 + \frac{4b_1(1 - b_1)(w - 2 + w^{-1})}{|(2 - b_1)w - b_1|^2}|_{w=e^{i\theta}}.
\]
\[
= 1 - \frac{8b_1(1 - b_1)(1 - \cos \theta)}{|(2 - b_1)e^{i \theta} - b_1|^2}.
\]

It is now readily checked that these expressions are strictly less than 1 for all \( \theta \in (\pi, \pi) \setminus \{0\} \) (and \( b_1 \in (0, 1) \)), which gives exactly the desired properties. \( \square \)

**Figure 8.** The contour \( \mathcal{M}' \) and its parametrization.

**Figure 9.** The contour \( \mathcal{M}'' \) and its parametrization.

Even though \( \mathcal{M} \) enjoys the desired property (SD.\( \mathcal{M} \)), it cuts through the point \( z = 0 \). This could cause issues, as the integrals (6.16)–(6.17) generally contain poles at \( z_1 = 0 \). To circumvent this problem, we consider modifications \( \mathcal{M}' \) and \( \mathcal{M}'' \) of \( \mathcal{M} \):

\[
\mathcal{M}' = \mathcal{M}'(u_*) := \partial (\{ |z| \leq 1 \} \cap \{ |z - \frac{1}{2} | \leq \frac{1}{2} + u_* \}),
\]

\[
\mathcal{M}'' = \mathcal{M}''(u_*) := \partial (\{ |z - \frac{1}{2} | \leq \frac{1}{2} \} \cup \{ |z - u_* | \leq 2u_* \}),
\]

clockwise oriented; see Figures 8–9. Here \( u_* \in (0, \frac{1}{3} \wedge b_1) \) is a parameter, which we fix in Lemma 6.4 so that the resulting contours \( \mathcal{M}' \) and \( \mathcal{M}'' \) also enjoy the steepest decent condition. We now verify the steepest decent condition for \( \mathcal{M}' \) and \( \mathcal{M}'' \).

**Lemma 6.4.** There exists \( u_* \in (0, \frac{1}{3} \wedge b_1) \) such that, for the contours \( \mathcal{M}'(u_*) \) and \( \mathcal{M}''(u_*) \) we have

\[
|\mathcal{D}_s(z)| < 1, \quad |\tilde{\mathcal{D}}_s(z)| < 1 \quad z \in \mathcal{M}' \setminus \{1\},
\]

(SD.\( \mathcal{M}' \))
Straightforward calculation gives

We will show that for all small enough $u > 0$,

$$|\mathcal{M}_*(z)| < 1, \quad |\mathcal{H}_s(z)| < 1 \quad z \in \mathcal{M}'' \setminus \{1\},$$

(SD.$\mathcal{M}$)

Proof. We will show that for all small enough $u > 0$,

$$|\mathcal{M}_*(z)| < 1, \quad |\mathcal{H}_s(z)| < 1 \quad z \in \mathcal{M}'(u) \setminus \{1\},$$

$$|\mathcal{M}_*(z)| < 1, \quad |\mathcal{H}_s(z)| < 1 \quad z \in \mathcal{M}''(u) \setminus \{1\}.$$  

We begin with the statement for $\mathcal{M}''(u)$. Indeed, this contour differs from $\mathcal{M}$ only in the neighborhood $O(3u) := \{z \in \mathbb{C} : |z| < 3u\}$ of $z = 0$. This being the case, instead of the entire contour $\mathcal{M}''(u)$, we need only to consider the part $\mathcal{M}''(u) \cap O(3u)$. We already know from (SD.$\mathcal{M}$) that $|\mathcal{M}_*(0)| < 1$ and $|\mathcal{H}_s(0)| < 1$. It is readily checked from (6.7) and (6.19) that $\mathcal{M}_*(z)$ and $\mathcal{H}_s(z)$ are continuous at $z = 0$, hence we see that $|\mathcal{M}_*(z)| < 1, |\mathcal{H}_s(z)| < 1$ holds on $z \in \mathcal{M}''(u) \cap O(3u)$ for all small enough $u > 0$.

We now turn to $\mathcal{M}'(u)$. Let us first analyze the local behaviors of $\mathcal{M}_*(z)$ and $\mathcal{H}_s(z)$ near $z = 1$. Straightforward calculation gives

$$\mathcal{D}_s(1) = 1, \quad \partial_z \mathcal{D}_s(1) = 0, \quad \partial_z^2 \mathcal{D}_s(1) = \nu_s, \quad \mathcal{H}_s(1) = 1, \quad \partial_z \mathcal{H}_s(1) = 0, \quad \partial_z^2 \mathcal{H}_s(1) = 2\nu_s,$$

so Taylor expansion of $\mathcal{D}_s(z)$ around $z = 1$ gives $1 + \frac{1}{2}\nu_s(z-1)^2$ up the second order, and Taylor expansion of $\mathcal{H}_s(z)$ around $z = 1$ gives $1 + \nu_s(z-1)^2$ up the second order. The expression $\nu_s(z-1)^2$ is real and negative along the vertical direction: $z - 1 \in \mathbb{R}$. Since $\mathcal{D}_s(z)$ and $\mathcal{H}_s(z)$ are analytic in a neighborhood of $z = 1$, we have

$$|\mathcal{D}_s(z)|, \quad |\mathcal{H}_s(z)| \leq 1 - \frac{1}{\nu_s}|z|, \quad \forall z \in \mathcal{A},$$

where $\mathcal{A} := \{z = v e^{i\phi} : v \in [0, v_0], |\phi + \frac{\pi}{2}| \leq \phi_0\}$ is an ‘hourglass-shape’ region centered at $z = 1$, and $v_0, \phi_0 > 0$ are fixed. See Figure 10. This property ensures that $|\mathcal{D}_s|, |\mathcal{H}_s| < 1$ within $\mathcal{A} \setminus \{1\}$, so instead of the entire contour $\mathcal{M}'(u)$, it suffices to consider the part $(\mathcal{M}'(u) \setminus \mathcal{A})$.

Instead of $(\mathcal{M}'(u) \setminus \mathcal{A})$, let us first consider $(\mathcal{M} \setminus \mathcal{A})$. Since the contour $\mathcal{M}$ passes through the point $z = 1$ vertically, under the parametrization $z(\theta) = \frac{1-v e^{i\theta}}{2}$, the part $(\mathcal{M} \setminus \mathcal{A})$ avoids a neighborhood of $\theta = 0$. This being the case, referring to the calculations (6.20), we see that

$$\sup_{z \in \mathcal{M} \setminus \mathcal{A}} |\mathcal{D}_s(z)| < 1, \quad \sup_{z \in \mathcal{M} \setminus \mathcal{A}} |\mathcal{H}_s(z)| < 1.$$ 

Let $\text{dist}(A, B) := \inf\{|z_1 - z_2| : z_1 \in A, z_2 \in B\}$ denotes the distance of two sets $A, B \subset \mathbb{C}$. Referring to the definition (6.21) of $\mathcal{M}'(u)$, we see that

$$\lim_{u \downarrow 0} \text{dist}((\mathcal{M} \setminus \mathcal{A}), (\mathcal{M}'(u) \setminus \mathcal{A})) = 0.$$ 

Further, it is readily verified (from (6.7) and (6.19)) that $\mathcal{D}_s$ and $\mathcal{H}_s$ are uniformly continuous $\mathcal{M}$. These properties together give

$$\lim_{u \downarrow 0} \left( \sup_{z \in \mathcal{M}'(u) \setminus \mathcal{A}} |\mathcal{D}_s(z)| \right) = \sup_{z \in \mathcal{M} \setminus \mathcal{A}} |\mathcal{D}_s(z)| < 1, \quad \lim_{u \downarrow 0} \left( \sup_{z \in \mathcal{M}'(u) \setminus \mathcal{A}} |\mathcal{H}_s(z)| \right) = \sup_{z \in \mathcal{M} \setminus \mathcal{A}} |\mathcal{H}_s(z)| < 1,$$

which concludes the proof. \hfill \Box
In the following subsections we prove Proposition 6.2(c)–(d), namely establishing the desired estimates on $V_\varepsilon^{in}$ and its gradients. To this end, we treat separately the cases distinguished by the signs of $x_2 - y_1$ and $x_1 - y_2$, which we refer to as the $(+\,-)$, $(-\,-)$, and $(+\,+)\text{-cases}$:

- $x_2 - y_1 > 0$ and $x_1 - y_2 \leq 0$, the $(+\,-)$-case;
- $x_2 - y_1 \leq 0$ and $x_1 - y_2 \leq 0$, the $(-\,-)$-case;
- $x_2 - y_1 > 0$ and $x_1 - y_2 > 0$, the $(+\,+)\text{-case}$.

The $(-\,\,\,)\text{-case}$ (i.e., $x_2 - y_1 \leq 0$ and $x_1 - y_2 > 0$) is irrelevant due the assumption $x_1 < x_2$ and $y_1 < y_2$.

Let us introduce one more convention about Taylor expansion which will be used in the subsequent arguments. Recall the assumption $t \leq \varepsilon - 2T$ which ensures that $\varepsilon \leq C(T)(t + 1)^{-1/2}$. At times we will Taylor expand expressions in the variables $(\sqrt{\varepsilon}, \frac{1}{\sqrt{t+1}})$. In the course of doing so, we adopt the following ordering convention in light of the aforementioned condition on $\varepsilon$.

**Definition 6.5.** To Taylor expand a given expression $f(\sqrt{\varepsilon}, \frac{1}{\sqrt{t+1}})$, we assign $\sqrt{\varepsilon}$ the order of $(t + 1)^{-1/4}$. For example, Taylor expansion of $f(\sqrt{\varepsilon}, \frac{1}{\sqrt{t+1}})$ up to order $1$ reads

$$f(0,0) + \partial_1 f(0,0) \sqrt{\varepsilon} + \frac{1}{2} \partial^2_1 f(0,0) \varepsilon + \partial_2 f(0,0) \frac{1}{\sqrt{t+1}}.$$

6.3. **Estimating the interacting part $V_\varepsilon^{in}$, the $(+\,-)$-case.** We begin by constructing the contour $\Gamma(t, \varepsilon)$. For the $(+\,-)$-case considered here, $\Gamma(t, \varepsilon)$ is constructed as perturbation of $M'$. More precisely, recall the definition of $u(t, \beta)$ from (6.5). For $\beta \in \mathbb{R}$, set

$$M'(t, \beta) := \partial \{ |z| \leq u(t, \beta) \} \cap \{ |z - \frac{1}{2}| \leq \frac{1}{2} \},$$

(6.23)
clockwise oriented; see Figure 11 and compare it with Figure 8. Under these notation, we set

$$\Gamma(t, \varepsilon) := M'(t, -k_1 \alpha),$$

where $k_1 = k_1(\alpha, T) \in \mathbb{Z}_{>0}$ is an auxiliary parameter to be fixed later.

---

**Figure 11.** The contour $M'(t, \beta)$ and its parametrization. The figure shows the case $\beta < 0$.

Hereafter we parametrize $z_1 = z_1(\theta_1) \in M'(t, -k_1 \alpha)$ as depicted in Figure 11. As for the $z_2$-contour, we fix $k_2 := 1$ in (6.13). Recalling $\tilde{r}_2(z_1)$ from (6.14), we parametrize $z_2(\theta) := \tilde{r}_2(z_1)e^{i\theta_2} \in \mathcal{C}_{\tilde{r}_2(z_1)}$.

The parameter $k_1 \in \mathbb{Z}_{>0}$ is to ensures that

$$r'_2 \geq p_\varepsilon(z_1(0)) + \frac{1}{\sqrt{t+1}} \in \mathbb{R}.$$

(6.24)

---

\(^{12}\)Here $\Gamma(t, \varepsilon)$ does not depend on $\varepsilon$, but we keep this notation to be consistent throughout all cases.
Lemma 6.6. Given any \( \varepsilon \) that do have the desired Gaussian decay of (6.25) \( z \) be relevant toward controlling the integral \( V \). From this, together with \( \varepsilon \leq \frac{C(T)}{\sqrt{t+1}} \) under current assumptions, we see that the condition (6.24) holds for a large enough \( k_1 = k_1(\alpha, T) \), and we fix such a \( k_1 \in \mathbb{Z}_{>0} \) hereafter.

The purpose of imposing the condition (6.24) is to control the region \( \{ z_1 : |p_\varepsilon(z_1)| > r_2' \} \), as will be relevant toward controlling the integral \( \mathbf{V}_{\text{res}}(6.17) \). Under the aforementioned parametrization \( z_1 = z_1(\theta) \), the condition (6.24) ensures a lower bound on \( |\theta| \) for which \( |p_\varepsilon(z_1(\theta))| > r_2' \). That is,

\[
|p_\varepsilon(z_1(\theta_1))| > r_2' \quad \text{holds only if} \quad |\theta_1| \geq \frac{1}{C(\alpha)(t+1)^{1/4}}.
\]

Proof of (6.25). Set \( f(\theta_1) := |p_\varepsilon(z_1(\theta_1))| - r_2' \). Our goal is to obtain a lower bound on those \( |\theta_1| \) such that \( f(\theta_1) \geq 0 \). Given the explicit expression \( p_\varepsilon(z_1(\theta_1)) = e^{\varepsilon(\rho-1)} + e^{\varepsilon\rho} - e^{\varepsilon(2\rho-1)}u(t, k_1 \alpha) e^{-i\theta_1} \), one readily checks that \( \frac{d}{d\theta_1} f(0) = 0 \), and that \( |\frac{d}{d\theta_1} f(\theta_1)| \leq C(\alpha), \) for all \( (\theta_1, t, \varepsilon) \in (-\pi, \pi) \times \mathbb{Z}_{\geq 0} \times (0, 1) \). Taylor expanding \( f(\theta_1) \) accordingly as

\[
f(\theta_1) = f(0) + \int_0^{\theta_1} (\theta_1 - \theta) \frac{d}{d\theta} f(\theta) d\theta,
\]

we see \( f(\theta_1) \geq 0 \) only if \( f(0) + C(\alpha)\theta_1^2 \geq 0 \). Now, the condition (6.24) ensures that \( f(0) \leq -\frac{1}{\sqrt{t+1}} \). From this we conclude (6.25).

Recall that \( \mathbf{H}_\varepsilon(t, z) := \mathbf{D}_\varepsilon(t, z) \mathbf{D}_\varepsilon(t, p_\varepsilon(z)) \). Let us check that, along the contour \( \mathcal{M}'(t, -k_1 \alpha) \), we do have the desired Gaussian decay of \( |\mathbf{D}_\varepsilon| \) and \( |\mathbf{H}_\varepsilon| \).

Lemma 6.6. Given any \( T \in (0, \infty) \) and \( \beta \in \mathbb{R} \),

\[
|\mathbf{D}_\varepsilon(t, z)|, \quad |\mathbf{H}_\varepsilon(t, z)| \leq C(\beta, T) \exp(-\frac{\theta^2}{C}(t+1)), \quad z = z(\theta) \in \mathcal{M}'(t, \beta),
\]

for all \( \theta \in (-\pi, \pi) \), large enough \( t \leq \varepsilon^{-2}T \), and small enough \( \varepsilon > 0 \).

Proof. The proof follows the same three-step procedure as the proof of Lemma 6.3. Given the identities (6.8a)–(6.8c), the proof of the first two steps (Zero \( \theta \))–(Small \( \theta \)) follows the same argument via Taylor expansion as in Lemma 6.3, and we do not repeat it here.

We now focus on establishing the last step (Large \( \theta \)). First, the contour \( \mathcal{M}'(t, \beta) \) converges, as \( t \to \infty \), to \( \mathcal{M}' \). More precisely, write \( z_{\mathcal{M}'(t, \beta)}(\theta; t, \beta) \) and \( z_{\mathcal{M}'(t, \beta)}(\theta) \) for the respectively polar parametrization as depicted in Figures 11 and 8. We have \( \lim_{t \to \infty} z_{\mathcal{M}'(t, \beta)}(\theta; t, \beta) = z_{\mathcal{M}'(t, \beta)}(\theta) \), uniformly over \( \theta \in (-\pi, \pi) \). This being the case, from the given expressions (6.6), (6.7), and (6.19) of \( \mathbf{D}_\varepsilon(t, z), \mathbf{D}_\varepsilon(z), \) and \( \mathbf{H}_\varepsilon(z) \), it is readily checked that

\[
\lim_{t \to \infty} \left| \mathbf{D}_\varepsilon(t, z_{\mathcal{M}'(t, \beta)}(\theta)) \right|^\frac{1}{2} = \left| \mathbf{D}_\varepsilon(z_{\mathcal{M}'(t, \beta)}(\theta)) \right|, \quad \lim_{t \to \infty} \left| \mathbf{H}_\varepsilon(t, z_{\mathcal{M}'(t, \beta)}(\theta)) \right|^\frac{1}{2} = \left| \mathbf{H}_\varepsilon(z_{\mathcal{M}'(t, \beta)}(\theta)) \right|,
\]

uniformly over \( \theta \in (-\pi, \pi) \). The limiting expressions on the r.h.s. put us into the considerations of the steepest decent condition (SD\( \mathcal{M}' \)), which has been verified in Lemma 6.4. From this we conclude the desired conclusion: there exists \( t_0 < \infty \) such that, for any given \( \theta_0 > 0 \),

\[
\left| \mathbf{D}_\varepsilon(t, z) \right|^\frac{1}{2} \leq 1 - \frac{1}{C(\theta_0)}, \quad \left| \mathbf{D}_\varepsilon(t, z) \right|^\frac{1}{2} \leq 1 - \frac{1}{C(\theta_0)}, \quad \forall z = z_{\mathcal{M}'(t, \beta)}(\theta) \in \mathcal{M}'(t, \beta), \quad |\theta| \geq \theta_0.
\]

\( \Box \)

We have all the necessary ingredients for estimating \( \mathbf{V}_{\text{res}}^\in \).
Proof of Proposition 6.2(c)–(d), the (+−)-case, with large enough $t$. The proof begins with the contour deformation described in Section 6.2. Let us check the condition (No Pole). For a fixed $z_2 \in \mathcal{C}_r$, the integrand in (6.4) has poles in $z_1 = 0$, $z_1 = e^{\sqrt{T}(r-1)}b_1$, and $p_\varepsilon(z_1) = z_2$. Referring to the definition (6.23) of $\mathcal{M}'(t, -k_1\alpha)$ (or Figure 11), we see that the first two poles are contained in $\mathcal{M}'(t, -k_1\alpha)$. As for the pole $p_\varepsilon(z_1) = z_2$, the function $p_\varepsilon(z)$ (defined in (6.12)) is uniformly bounded (in $\varepsilon, z$) away from $z = 0$. This being the case, by making $r$ large enough, we ensure that $|p_\varepsilon(z_1)| < r = |z_2|$ throughout the contour deformation $z_1 \in \mathcal{C}_r \mapsto \mathcal{M}'(t, -k_1\alpha)$. Having checked the condition (No Pole), we are now given the decomposition $V^\text{in}_\varepsilon = V_{\text{blk}} + V_{\text{res}}$. The proof amounts to bounding $V_{\text{blk}}$ and $V_{\text{res}}$, as well as their gradients.

We begin with $V_{\text{blk}}$ (6.16). The proof consists of a sequence of bounds on terms appearing in the integrand (6.16). In the following we assume $z_1 = z_1(\theta_1) \in \mathcal{M}'(t, -k_1\alpha)$ and $z_2 = z_1(\theta_2) \in \mathcal{C}_{r_2}(z_1)$. \(V_{\text{blk}}, z_1\) Show that $|z_1|^{x_2-y_1+\mu_\varepsilon t} - |\mu_\varepsilon t| \leq \exp(-\frac{\alpha|x_2-y_1|}{\sqrt{T}+1+C(\alpha)})$.

Referring to the definition (6.23) of $\mathcal{M}'(t, -k_1\alpha)$ (or Figure 11), we see that $\mathcal{M}'(t, -k_1\alpha)$ is contained in $\mathcal{C}_{u(t,-k_1\alpha)}$, so $|z_1|^{x_2-y_1+\mu_\varepsilon t} - |\mu_\varepsilon t| \leq u(t, -k_1\alpha)|x_2-y_1| \leq C(\alpha)e^{-\frac{\alpha|x_2-y_1|}{\sqrt{T}+1+C(\alpha)}}.$ \(V_{\text{blk}}, z_2\) Show that $|z_2|^{x_1-y_2+\mu_\varepsilon t} - |\mu_\varepsilon t| \leq C(\alpha)\exp(-\frac{\alpha|x_1-y_2|}{\sqrt{T}+1+C(\alpha)})$.

Recall the current assumption $x_1 - y_2 \leq 0$. The power $x_1 - y_2 + \mu_\varepsilon t - |\mu_\varepsilon t|$ would have a definitive sign (i.e., non-positive) if we offset it by $-|\mu_\varepsilon t - |\mu_\varepsilon t||$. Since $|z_2| \leq C(\alpha)$ is bounded along its contour $z_2 \in \mathcal{C}_{r_2}(z_1)$, offsetting the exponent costs only a factor of $C(\alpha)$:

$|z_2|^{x_1-y_2+\mu_\varepsilon t} - |\mu_\varepsilon t| \leq C(\alpha)|z_2|^{x_1-y_2} = C(\alpha)|z_2|^{-|x_1-y_2|}.$

Recall the definitions of the $\eta_2$’s and of $\tilde{\eta}_2$ from (6.13)–(6.14), and recall that $k_2 := 1$ here. We see that $\tilde{\eta}_2(z_2) \geq u(t, \alpha)$, so $|z_2|^{x_1-y_2+\mu_\varepsilon t} - |\mu_\varepsilon t| \leq C(\alpha)u(t, \alpha)^{-|x_1-y_2|} \leq C(\alpha)e^{-\frac{\alpha|x_1-y_2|}{\sqrt{T}+1+C(\alpha)}}.$ \(V_{\text{blk}}, \tilde{\eta}_2\) Show that $|\tilde{\eta}_2(z_1, z_2)| \leq C(\alpha)(1 + |\theta_1|\sqrt{T+1} + |\theta_2|\sqrt{T+1})$.

Recall the expression (6.11) of $\tilde{\eta}_2$, and rewrite it as

$\tilde{\eta}_2(z_1, z_2) = 1 + (e^{\sqrt{T}(\rho-1)} + e^{\sqrt{T}\rho}) \frac{z_2/z_1 - 1}{z_2 - p_\varepsilon(z_1)}.$ \(V_{\text{blk}, \mathcal{D}_\varepsilon}\) Show that $|\mathcal{D}_\varepsilon(z_1)| \leq C(\alpha, T) \exp\left(-\frac{\theta_2}{\alpha}(t+1)\right)$.

This is the content of Lemma 6.6.

Expressing (6.16) as an integral over $(\theta_1, \theta_2) \in (-\pi, \pi)^2$, and inserting the bounds from \(V_{\text{blk}}, z_1\)–\(V_{\text{blk}, \mathcal{D}_\varepsilon}\) into the resulting expression, we arrive at

$|V_{\text{blk}}| \leq C(\alpha, T) \int_{-\pi}^\pi \int_{-\pi}^\pi e^{-\frac{\alpha|x_2-y_1|+|x_1-y_2|}{\sqrt{T}+1+C(\alpha)}} \left(1 + \sqrt{T+1}|\theta_1| + \sqrt{T+1}|\theta_2|\right)e^{-\frac{\theta_2}{\alpha}(t+1)}d\theta_1.$
Performing the change of variables $\sqrt{t+\mathcal{J}_\theta} \mapsto \theta_1$, and extending the integration domain to $\mathbb{R}^2$ (which only increases its value) we obtain the desired bound on $|V_{\text{blk}}|$

$$|V_{\text{blk}}| \leq C(\alpha, T)e^{-\frac{\alpha(x_2-y_2|x_1-y_1)}{\sqrt{t+1+C(\alpha)}}} \frac{1}{t+1} \int_{\mathbb{R}^2} (1 + |\theta_1| + |\theta_2|)e^{-\frac{\theta_1^2}{t+1}}d\theta_1 = C(\alpha, T)e^{-\frac{\alpha(x_2-y_2|x_1-y_1)}{\sqrt{t+1+C(\alpha)}}}.$$ 

As for $V_{\text{res}}$, the proof similarly consists of bounds on terms involved in the integral (6.17). In the following we always assume $z_1 = z_1(\theta_1) \in \mathcal{M}'(t, -k_1 \alpha)$.

\(V_{\text{res}, \frac{1}{z_1 p_z}}\) Show that $\frac{1}{|p_z(z_1)| |z_1|} \leq C(\alpha)$:

Referring to the definition (6.23) of $\mathcal{M}'(t, -k_1 \alpha)$ (or Figure 11), we see that $|z_1|$ is bounded away from 0 and $\infty$ along $\mathcal{M}'(t, -k_1 \alpha)$. This being the case, referring to the definition (6.12) of $p_z(z)$, we see that the same holds for $|p_z(z_1)|$. Hence the claim follows.

\(V_{\text{res}, 3}\) Show that $1_{\{p_z(z_1) > r_2^2\}} |\mathcal{J}(z_1)| \leq \exp(-\frac{\alpha(x_2-y_2|x_1-y_1)}{\sqrt{t+1+C(\alpha)}})$:

Recall from (6.15) that $\mathcal{J}(z_1)$ consists of products of powers of $z_1$ and $p_z(z_1)$. As argued in the previous step ($V_{\text{res}, \frac{1}{z_1 p_z}}$), the terms $|z_1|, |z_1|^{-1}, |p_z(z_1)|^{-1} \leq C(\alpha)$ are bounded along $\mathcal{M}'(t, -k_1 \alpha)$. This being the case, we alter the powers in (6.15) by some fixed amount, at the cost of $C(\alpha)$, and write

$$1_{\{p_z(z_1) > r_2^2\}} |\mathcal{J}(z_1)| \leq C(\alpha)1_{\{p_z(z_1) > r_2^2\}} |z_1| |x_2-y_2| |p_z(z_1)|^{-|x_1-y_1|}.$$ 

In the last expression, using $|z_1| \leq u(t, -k_1 \alpha)$ (as argued in ($V_{\text{blk}, z_1}$)) and the given constraint $|p_z(z_1)| > r_2^2 = u(t, 2\alpha)$, we obtained the desired property:

$$1_{\{p_z(z_1) > r_2^2\}} |\mathcal{J}(z_1)| \leq C(\alpha)u(t, -k_1 \alpha)^{-|x_2-y_1|} u(t, \alpha)^{-|x_1-y_1|} \leq C(\alpha)e^{-\frac{\alpha(x_2-y_2|x_1-y_1)}{\sqrt{t+1+C(\alpha)}}}.$$ 

\(V_{\text{res}, \beta_2}\) Show that $|\beta_2(z_1)| \leq C(\alpha, T) \exp(-\frac{\theta_t^2}{C(t+1)})$:

This is the content of Lemma 6.6.

Express (6.17) as an integral over $\theta_1 \in (-\pi, \pi]$, and insert the bounds from ($V_{\text{res}, \frac{1}{z_1 p_z}}$)–($V_{\text{res}, \beta_2}$) into the resulting expression. Together the derived lower bound (6.25) on $|\theta_1|$ gives

$$|V_{\text{res}}| \leq C(\alpha, T)e^{-\frac{\alpha(x_2-y_2|x_1-y_1)}{\sqrt{t+1+C(\alpha)}}} \int_{(-\pi, \pi]} 1\{|\theta_1| \geq \frac{1}{C(\alpha)\theta_t}\}e^{-\frac{\theta_t^2}{C(\alpha)\theta_t^2}}d\theta_1.$$ 

Extending the integration domain to $\mathbb{R}$, and performing a change of variable $\sqrt{t+\mathcal{J}_\theta} \mapsto \theta_1$ yields

$$|V_{\text{res}}| \leq C(\alpha, T)e^{-\frac{\alpha(x_2-y_2|x_1-y_1)}{\sqrt{t+1+C(\alpha)}}} \frac{1}{t+1} \int_{\mathbb{R}} 1\{|\theta_1| \geq \frac{(t+1)^{1/4}}{C(\alpha)}\}e^{-\frac{\theta_t^2}{(t+1)^{1/4}}}d\theta_1.$$ 

Here, unlike in the case for $V_{\text{res}}^p$, we get $\frac{1}{\sqrt{t+1}}$ instead of $\frac{1}{t+1}$ in front of the integral. This insufficiency is compensated by having the constraint $|\theta_1| \geq (t+1)^{1/4}/C(\alpha)$. Indeed,

$$\int_{\mathbb{R}} 1\{|\theta_1| \geq \frac{(t+1)^{1/4}}{C(\alpha)}\}e^{-\frac{\theta_t^2}{(t+1)^{1/4}}}d\theta_1 \leq \exp\left(-\frac{1}{C(\alpha)}(t+1)^{1/4}\right),$$ 

and fractional exponentials such as $\exp\left(-\frac{1}{C(\alpha)}(t+1)^{1/4}\right)$ decay faster than any power $(t+1)^{-n}$. From this we conclude the desired bound on $|V_{\text{res}}|$:

$$|V_{\text{res}}| \leq C(\alpha, T) \frac{1}{t+1} e^{-\frac{\alpha(x_2-y_2|x_1-y_1)}{\sqrt{t+1+C(\alpha)}}}.$$ 

So far we have derived bounds on $|V_{\text{blk}}|$ and $|V_{\text{res}}|$, and this concludes the proof of Proposition 6.2(c).

Part (d) amounts to performing similar estimates on the gradients, e.g., $|\nabla_{x_j} V_{\text{blk}}|$ and $|\nabla_{x_j} V_{\text{res}}|$. Taking a gradient merely introduces a factor of $(z_j^\pm - 1)$ in the contour integrals (6.16)–(6.17). It is straightforward to check that

$$|z_j^\pm - 1| \leq \frac{1}{\sqrt{t+1}} + |\theta_j|, \quad z_1 = z_1(\theta) \in \mathcal{M}'(t, -k_1 \alpha), \quad z_2 = z_2(\theta) \in C_{\theta_2}(z_1). \quad (6.28)$$
Incorporate this bound into the preceding analysis gives the desired bounds on the gradients. Compared to the bounds on $|V_{blk}|$ and $|V_{res}|$, an additional factor of $\frac{1}{\sqrt{t+1}}$ arises due to (6.28). □

6.4. **Estimating the interacting part $V^{\text{in}}$**, the $(-\cdot)$-case. The case considered here is more involved than the $(+-)$-case: we face a conflict in the choice of the $z_1$-contour. As discussed in Section 6.2, in order to control the term $\delta_z(t, z)$ in $V_{res}$ by steepest decent analysis, we favor contours of the type $M(t, \beta)$. On the other hand, with $x_2 - y_1 \leq 0$ under current assumptions, we need $|z_1| > 1$ in $V_{blk}$ to obtain the desired spatial exponential decay $\exp(-\frac{a|x_2-y_1|}{\sqrt{t+1}C(\alpha)})$. Referring to the definition (6.23) of $M(t, -k_1\alpha)$ (or Figure 11), we see that $|z_1| > 1$ fails for a portion of $M(t, \beta)$, regardless of the sign of $\beta$—i.e., the bulk part $V_{blk}$ and the residue part $V_{res}$ favor different contours.

In view of the preceding discussion, we choose

$$\Gamma(t, \varepsilon) := \mathcal{C}_{u(t, 3\alpha)},$$

which is preferred for controlling $V_{blk}$ but not $V_{res}$, and then, re-deforming contour $\mathcal{C}_{u(t, 3\alpha)} \to M(t, 3\alpha)$ in $V_{res}$. Let us check that doing so does not cross a pole.

**Lemma 6.7.** For all $t > 0$ large enough and $\varepsilon > 0$ small enough, we have $V_{res} = V'_{res}$, where

$$V'_{res} := \int_{M(t, 3\alpha)} 1_{\{|p_\varepsilon(z_1)| > r_2\}} \mathcal{J}(z_1) \frac{1}{z_1p_\varepsilon(z_1)} \delta_z(t, z_1)dz_1$$

(6.29)

is the same as $V_{res}$ except the contour is replaced by $M(t, 3\alpha)$.

**Proof.** Referring to the definition (6.23) of $M(t, 3\alpha)$ (or Figure 11), we see that the difference $\mathcal{C}_{u(t, 3\alpha)} - M(t, 3\alpha)$ is the boundary of the crescent

$$G(t) := \{z \in \mathbb{C} : |z| \leq u(t, 3\alpha)\} \setminus \{z \in \mathbb{C} : |z - \frac{1}{2}| < \frac{1}{2} + u_t\}.$$ 

See Figure 12. We write $\partial G(t)$ for the boundary, counterclockwise oriented. This gives

$$V_{res} - V'_{res} = \frac{e^{\sqrt{\varepsilon}(\rho-1)} + e^{\sqrt{\varepsilon}\rho}}{2\pi i} \int_{\partial G(t)} 1_{\{|p_\varepsilon(z_1)| > r_2\}} \mathcal{J}(z_1) \frac{1}{z_1p_\varepsilon(z_1)} \delta_z(t, z_1)dz_1.$$ 

Along $\partial G(t)$, the indicator $1_{\{|p_\varepsilon(z_1)| > r_2\}}$ is in fact irrelevant. More precisely, let us check that, for all $t$ large enough,

$$|p_\varepsilon(z)| > r_2, \quad z \in G(t).$$

(6.30)

Referring to the definition (6.12) of $p_\varepsilon(z)$, we have $\lim_{\varepsilon \to 0} p_\varepsilon(z) := p_\varepsilon(z) = 2 - z^{-1}$. It is straightforward to verify that

$$\inf \{ |p_\varepsilon(z)| : \text{Re}(z) \leq 1 - v \} > 1, \quad \forall v > 0.$$ 

(6.31)

Referring to Figure 12, we see that $G(t)$ is contained in $\mathcal{C}_{u(t, 3\alpha)}$ and avoids a fixed neighborhood of $z = 1$. This being the case, with $u(t, 3\alpha) \to 1$ as $\varepsilon \to 0$, we indeed have that $G(t) \subset \{z : \text{Re}(z) \leq 1 - v\}$, for some small enough $v > 0$, so the claim (6.30) holds.

Given (6.30), we drop the indicator $1_{\{|p_\varepsilon(z_1)| > r_2\}}$ and write

$$V'_{res} - V_{res} = \frac{e^{\sqrt{\varepsilon}(\rho-1)} + e^{\sqrt{\varepsilon}\rho}}{2\pi i} \int_{\partial G(t)} \mathcal{J}(z_1) \frac{1}{z_1p_\varepsilon(z_1)} \delta_z(t, z_1)dz_1.$$ 

Our goal is to show that the integral is zero. To this end, set $q_\varepsilon(z) := zp_\varepsilon(z) := (e^{\sqrt{\varepsilon}\rho} + e^{\sqrt{\varepsilon}(\rho-1)})z - e^{\sqrt{\varepsilon}(2\rho-1)}$, recall the definition of $\mathcal{J}(z_1)$ from (6.15), that $\delta_z(t, z_1) := \mathcal{D}_z(t, z_1)\mathcal{D}_\varepsilon(t, p_\varepsilon(z_1))$, and recall the definition of $\mathcal{D}_\varepsilon(t, z)$ from (4.13). We express the integrand of the last integral as

$$\frac{\mathcal{J}(z_1)}{z_1p_\varepsilon(z_1)} \delta_z(t, z_1) = (z_1^{x_2-y_1} - (x_1-y_2)z + (\mu t - |\mu t|)) (z_1^{x_2-y_1} - (x_1-y_2)z + (\mu t - |\mu t|)) q_\varepsilon(z_1)^{x_1-y_2-1+|\mu t|},$$

$$\left(\lambda_\varepsilon \left(\frac{z_1b_1 + (1-b_1-b_2^\varepsilon)\tau_\varepsilon^{-\rho}}{z_1 - b_2^\varepsilon \tau_\varepsilon^{-\rho}}\right) t \right)^{t} \left(\lambda_\varepsilon \left(\frac{p_\varepsilon(z_1)b_1 + (1-b_1-b_2^\varepsilon)\tau_\varepsilon^{-\rho}}{p_\varepsilon(z_1) - b_2^\varepsilon \tau_\varepsilon^{-\rho}}\right) \right)^{t}.$$ 

(6.32)
It suffices to check that this expression has no poles within $z_1 \in \mathcal{G}(t)$. The assumption $x_1 < x_2$ and $y_1 < y_2$ ensures that $(x_2 - y_1) - (x_1 - y_2) \geq 2$. Thus, the expression (6.32) can only have poles at $q_{\varepsilon}^{-1}(0)$, $b_{\varepsilon}^{2} \tau_{\varepsilon}^{-\rho}$, or $p_{\varepsilon}^{-1}(b_{\varepsilon}^{2} \tau_{\varepsilon}^{-\rho})$. With $b_{\varepsilon}^{2} \rightarrow b_1, \tau_{\varepsilon} \rightarrow 1$, and $p_{\varepsilon}(z) \rightarrow 2 - z^{-1}$, we have

$$q_{\varepsilon}^{-1}(0) \rightarrow \frac{1}{2}, \quad b_{\varepsilon}^{2} \tau_{\varepsilon}^{-\rho} \rightarrow b_1, \quad p_{\varepsilon}^{-1}(b_{\varepsilon}^{2} \tau_{\varepsilon}^{-\rho}) \rightarrow 2 - b_1, \quad \text{as } \varepsilon \rightarrow 0.$$ 

Referring to Figure 12, we see that $\frac{1}{2}$, $b_1$, and $2 - b_1$, all sit strictly outside of $\mathcal{G}(t)$. Consequently, no poles enter into $\mathcal{G}(t)$ as long as $t > 0$ is large enough and $\varepsilon > 0$ is small enough. 

Having introduced the contours $\mathcal{C}_{u(t,3\alpha)}$ and $\mathcal{M}(t,3\alpha)$, hereafter we write $z_1(\theta_1) = u(t,3\alpha)e^{i\theta_1} \in \mathcal{C}_{u(t,3\alpha)}$, and write $\tilde{z}_1(\theta_1) \in \mathcal{M}(t,3\alpha)$ for the parametrization depicted in Figure 11.

To control $V_{\text{res}}$ in the following, similarly to the $(+-)$-case done previously, we need the analogous condition (6.24) to hold:

$$r_2' \geq p_{\varepsilon}(\tilde{z}_1(0)) + \frac{1}{\sqrt{t} + 1} \in \mathbb{R}. \quad (6.24')$$

We achieve this by making the auxiliary parameter $k_2 \in \mathbb{Z}_{>0}$ in (6.13) large enough. Recall from Definition 6.5 the announced convention on Taylor expansion, and expand the expression $r_2' - p_{\varepsilon}(\tilde{z}_1(0)) = u(t,2k_2\alpha) - p_{\varepsilon}(u(t,3\alpha))$ in $(\sqrt{\varepsilon}, \frac{1}{\sqrt{t} + 1})$ to the leading order in $\frac{1}{\sqrt{t} + 1}$ to get

$$z_2(0) - p_{\varepsilon}(\tilde{z}_1(0)) = 0 \cdot \sqrt{\varepsilon} - \rho(1 - \rho)\varepsilon + \frac{(k_2 - 3\alpha)}{\sqrt{t} + 1} + \ldots .$$

With $\varepsilon \leq \frac{C(T)}{\sqrt{t} + 1}$, from the expansion we see that (6.24') does hold for some large enough $k_2 = k_2(\alpha,T)$, and we fix such a $k_2 \in \mathbb{Z}_{>0}$ hereafter. Given this condition, following the same procedure of deriving (6.25) as in the $(+-)$-case, here we have

$$|p_{\varepsilon}(\tilde{z}_1(\theta_1))| > r_2' \text{ holds only if } |\theta_1| \geq \frac{1}{C(\alpha)(t + 1)^{1/4}}. \quad (6.25')$$

**Proof of Proposition 6.2(c)--(d), the $(-+)$-case, with large enough $t$.** The proof begins with the contour deformation described in Section 6.2. The condition (No Pole) is checked the same way as in the $(+-)$-case, which gives the decomposition $V_{\text{in}} = V_{\text{blk}} + V_{\text{res}}$. We next perform the aforementioned re-deformation $\mathcal{C}_{u(t,3\alpha)} \mapsto \mathcal{M}(t,3\alpha)$ in $V_{\text{res}}$. Lemma 6.7 ensures that no pole is crossed during this step, giving $V_{\text{in}}^{-} = V_{\text{blk}} + V_{\text{res}}^{-}$.

The proof amounts to bounding $V_{\text{blk}}, V_{\text{res}}^{-}$, and their gradients. We begin with $V_{\text{blk}}$, given by the integral expression (6.16). In the following we check a sequence of bounds on terms involved in (6.16), and we always assume $z_1 = z_1(\theta_1) \in \mathcal{C}_{u(t,3\alpha)}$ and $z_2 = z_2(\theta_2) \in \mathcal{C}_{\tilde{T}_2(z_1)}$ in the course of doing so.

**($V_{\text{blk}}, z_1$)** Show that $|z_1|^{|x_2 - y_1 + \mu t| - |\mu t|} \leq \exp(-\frac{\alpha|x_2 - y_1|}{\sqrt{t} + 1 + C(\alpha)})$:

This is so because $|z_1| = u(t,3\alpha)$ and $x_2 - y_1 \leq 0$ under current assumptions.
We now turn to $V_\epsilon$. Show that
\[ |z| \leq C(\alpha) (1 + |\theta_1| \sqrt{t+1} + \theta_2 \sqrt{t+1}) \]
This is established by the same argument as in the (++)-case. We do not repeat it here.

Given $(V_{\text{blk}} z_1) - (V_{\text{blk}} D_\epsilon)$, the desired bound on $V_{\text{blk}}$ follows by inserting the bounds into (6.16), and integrating the result. The procedure is the same as the (++)-case, and we do not repeat it here.

We now turn to $V_{\text{res}}$. In the following we always assume $\tilde{z}_1 = z_1(\theta_1) \in M'(t, 3\alpha)$.

\[ (V_{\text{res}} - \frac{1}{\tilde{z}_1}) |_{p_{\epsilon}(\tilde{z}_1)} \leq C(\alpha) \]
Referring to the definition (6.23) of $M(t, 3\alpha)$ (or Figure 11), we see that $|\tilde{z}_1|$ is bounded away from 0 and $\infty$ along $M'(t, 3\alpha)$. This being the case, referring to the definition (6.12) of $p_{\epsilon}(z)$, the same holds for $|p_{\epsilon}(\tilde{z}_1)|$.

\[ (V_{\text{res}} - \frac{1}{\tilde{z}_1}) |_{p_{\epsilon}(\tilde{z}_1)} \leq C(\alpha) \]
Recall from (6.15) that $\mathcal{J}(z_1)$ consists of products of powers of $\tilde{z}_1$ and $p_{\epsilon}(\tilde{z}_1)$. As argued in the previous step $(V_{\text{res}} - \frac{1}{\tilde{z}_1})$, the terms $|\tilde{z}_1|, |\tilde{z}_1|^{-1}, |p_{\epsilon}(\tilde{z}_1)|, |p_{\epsilon}(\tilde{z}_1)|^{-1} \leq C(\alpha)$ are bounded along $M'(t, 3\alpha)$. This being the case, we alter the powers (6.15) in by some fixed amount, at the cost of $C(\alpha)$, and write
\[
|\mathcal{J}(\tilde{z}_1)| \leq C(\alpha) |\tilde{z}_1|^{-|x_2-y_1|} |p_{\epsilon}(\tilde{z}_1)|^{-|x_1-y_2|}.
\]

Set $n_1 := |x_2-y_1|$ and $n_2 := |x_1-y_2|$. Instead of bounding $|\tilde{z}_1|^{-n_1}$ and $|p_{\epsilon}(\tilde{z}_1)|^{-n_2}$ separately, here we need to ‘bundle’ part of them together. The assumption $y_1 < y_2, x_1 < x_2$ in the (--) case yields $n_2 > n_1$. Given this, we write
\[
|\mathcal{J}(\tilde{z}_1)| \leq C(\alpha) |\tilde{z}_1|^{-|x_2-y_1|} |p_{\epsilon}(\tilde{z}_1)|^{-|x_1-y_2|} = C(\alpha) |\tilde{z}_1|^{-n_1} |p_{\epsilon}(\tilde{z}_1)|^{-n_2}.
\]
We claim that, for all $t \leq \varepsilon^{-2}T$ large enough and $\varepsilon > 0$ small enough,
\[
|\tilde{z}_1| \geq u(t, \alpha), \quad |p_{\epsilon}(\tilde{z}_1)| \geq u(t, 2\alpha), \quad \tilde{z}_1 \in M'(t, 3\alpha).
\]
Once these bounds are established, it follows that
\[
|\mathcal{J}(\tilde{z}_1)| \leq C(\alpha) u(t, \alpha)^{-n_1} u(t, 2\alpha)^{-n_2-n_1} \leq C(\alpha) e^{-\frac{2n_1}{\sqrt{t+1}+C(\alpha)}} e^{-\frac{2n_2-n_1}{\sqrt{t+1}+C(\alpha)}}.
\]
This concludes the desired bound on $|\mathcal{J}(\tilde{z}_1)|$, and it hence suffices to verify the claim (6.33).

Recall from (6.23) that $M'(t, 3\alpha)$ is given by $C_{u(t, 3\alpha)}$ near $u = 1$, and the rest by $\tilde{M} := \{|z - \frac{1}{2}| = \frac{1}{2} + u_\ast\}$. With this in mind, let us check the bounds separately on $C_{u(t, 3\alpha)}$ and $\tilde{M}$.

We begin with $C_{u(t, 3\alpha)}$. Adopt the parametrization $C_{u(t, 3\alpha)} \ni z_1(\theta_1) = u(t, 3\alpha)e^{i\theta_1}$ and write
\[
\tilde{z}_1|_{p_{\epsilon}(\tilde{z}_1)} = u(t, 3\alpha)e^{i\theta_1}(e^{\sqrt{\rho}(p-1)} - e^{-\sqrt{\rho}(p-1)}) - e^{\sqrt{2}(2\rho-1)},
\]
\[
p_{\epsilon}(\tilde{z}_1) = (e^{\sqrt{\rho}(p-1)} - e^{-\sqrt{\rho}(p-1)}) - e^{\sqrt{2}(2\rho-1)}u(t, -3\alpha)e^{-i\theta_1}.
\]
As $\theta_1$ varies, the r.h.s. of (6.34)–(6.35) trace out circles, denoted by $\tilde{C}(t, \varepsilon)$ and $\tilde{C}'(t, \varepsilon)$ respectively. The circle $\tilde{C}(t, \varepsilon)$ is centered at a point in $(-\infty, 0)$. For such circles, the nearest point to the origin occurs at the right-end. This gives
\[
\inf_{\tilde{z}_1 \in C_{u(t, 3\alpha)}} |z_1| \leq u(t, 3\alpha)(e^{\sqrt{\rho}(p-1)} - e^{-\sqrt{\rho}(p-1)}) - e^{\sqrt{2}(2\rho-1)}.
\]
A similarly geometric reasoning gives
\[
\inf_{\tilde{z}_1 \in C_{u(t, 3\alpha)}} |p_{\epsilon}(\tilde{z}_1)| = (e^{\sqrt{\rho}(p-1)} - e^{-\sqrt{\rho}(p-1)}) - e^{\sqrt{2}(2\rho-1)}u(t, -3\alpha).
\]
To bound the r.h.s., under the leading order announced in Definition 6.5, we Taylor expand the r.h.s. in \(\sqrt{\varepsilon}, \frac{1}{\sqrt{t+1}}\) up to the leading order in \(\frac{1}{\sqrt{t+1}}\) to get
\[
\begin{align*}
(u(t, 3\alpha)(e^{\sqrt{\varepsilon}(\rho-1)} - e^{\sqrt{\varepsilon}(\rho-1)}) - e^{\sqrt{2}(\rho-1)}) &= 1 + 0 \cdot \sqrt{\varepsilon} + \rho(1 - \rho)\varepsilon + \frac{3\alpha}{\sqrt{t+1}} + \ldots, \\
(e^{\sqrt{\varepsilon}(\rho-1)} - e^{\sqrt{2}(\rho-1)}) - e^{\sqrt{2}(\rho-1)}(u(t, -3\alpha)) &= 1 + 0 \cdot \sqrt{\varepsilon} + \rho(1 - \rho)\varepsilon + \frac{3\alpha}{\sqrt{t+1}} + \ldots.
\end{align*}
\]
From this, together with \(\varepsilon \leq \frac{C(T)}{\sqrt{t+1}}\) (because \(t \leq \varepsilon^{-2}T\)), we see that the desired bounds \(\bar{z}_1p_\varepsilon(\bar{z}_1) \geq u(t, \alpha), |p_\varepsilon(\bar{z}_1)| \geq u(t, 2\alpha)\) hold on \(C(u(t, 3\alpha))\), for all large enough \(t \leq \varepsilon^{-2}T\) and small enough \(\varepsilon > 0\).

We now turn to \(\bar{M}\). Recall that \(p_\varepsilon(z) := 2 - z^{-1}\) denotes the \(\varepsilon \downarrow 0\) limit of \(p_\varepsilon(z)\). Along the contour \(\bar{M} := \{z - \frac{1}{2} = \frac{1}{2} + u_\varepsilon\}\) we have \(|zp_\varepsilon(z)| = 1 + 2u_\varepsilon > 1\). This being the case, the bound \(|\bar{z}_1p_\varepsilon(\bar{z}_1)| \geq u(t, \alpha)\) holds on \(\bar{M}\) for large enough \(t\).

The other bound \(|p_\varepsilon(\bar{z}_1)| \geq u(t, 2\alpha)\) actually does not hold on the entire \(\bar{M}\), but relevant to our purpose is the part \(\bar{M} \cap \mathcal{M}'(t, \beta) := \mathcal{M}'(t, \beta)\). Referring to the definition (6.23) of \(\mathcal{M}'(t, \beta)\), we see that \(\mathcal{M}'(t, \beta) \subseteq \{z : \text{Re}(z) < 1 - \delta\}\) holds for some fixed \(\delta > 0\), for all \(t > 0\) large enough. This and (6.31) together gives the desired bound \(|p_\varepsilon(\bar{z}_1)| \geq u(t, 2\alpha)\) on \(\mathcal{M}'(t, \beta)\).

((V'\_res,\mathcal{N}_\varepsilon))) Show that \(|\mathcal{N}(\varepsilon z)| \leq C(\alpha, T) \exp(-\frac{\theta^2}{4}(t + 1))\):
This is the content of Lemma 6.6.

Given \((V'\_res,\bar{z}_1p_\varepsilon)\)–\((V'\_res,\bar{z}_1p_\varepsilon)\), and the derived constraint (6.25) on \(|\theta_1|\), the desired bound on \(V'\_res\) follows the same integration procedure as the \((++-\))-case.

As for the gradient, similarly to the \((++-\))-case, here we have
\[
|z_j^\pm - 1| \leq \sqrt{\varepsilon} + |\theta_j|, \quad z_1 = z_1(\theta) \in C(u(t, 3\alpha)) \text{ or } \mathcal{M}'(t, 3\alpha), \quad z_2 = z_2(\theta) \in C_\varepsilon(z_1) \text{,} \quad (6.28')
\]
Incorporate this bound into the preceding analysis gives the desired bounds on the gradients. \(\square\)

6.5. Estimating the interacting part \(V'_{\text{in}}\), the \((++-\))-case. Before heading to the construction of \(\Gamma(t, \varepsilon)\), we begin with some general discussion that motivates the construction. As it turns out, the analysis of the residue part \(V'_{\text{res}}\) favors contours of the type:
\[
\mathcal{N}(t, \varepsilon, \beta) := \left\{z : \left| \frac{1}{e^{\sqrt{\varepsilon}(\rho-1)} + e^{\sqrt{\varepsilon}(\rho-1)}} \right| = \frac{u(t, \beta)e^{\sqrt{2}(\rho-1)}}{e^{\sqrt{\varepsilon}(\rho-1)} + e^{\sqrt{2}(\rho-1)}} \right\}.
\]
First, it is readily checked that \(\mathcal{N}(t, \varepsilon, \beta)\) is the \(u(t, \beta)\)-level set of \(|zp_\varepsilon(z)|\), i.e.,
\[
\mathcal{N}(t, \varepsilon, \beta) = \left\{z : p_\varepsilon(z) = u(t, \beta)\right\}.
\]
This property is useful toward extracting the spatial exponential decay \(\exp(-\frac{\alpha(|z_2 - y_2| + |z_1 - y_2|)}{\sqrt{t+1} + C(\alpha)})\) from \(V'_{\text{res}}\). Further, \(\mathcal{N}(t, \varepsilon, \beta)\) is itself a circle, and, as \((t, \varepsilon) \to (\infty, 0)\), converges to \(\mathcal{M} := \{z - 1 = 1\}\).

With \(\mathcal{M}\) satisfying the steepest decent condition (SD,\(\mathcal{M}\)), it is conceivable that \(\mathcal{N}_\varepsilon(z_1)\) will be controlled along the contour \(\Gamma(t, \varepsilon, \beta)\).

However, if we choose \(\Gamma(t, \varepsilon)\) to be \(\mathcal{N}(t, \varepsilon, \beta)\) (with \(\beta \in \mathbb{R}\), for all \(\rho > \frac{1}{2}\), the first stage of contour deformation \(C_\varepsilon \to \Gamma(t, \varepsilon)\) will inevitably cross a pole at \(p_\varepsilon(z_1) = z_2\) no matter how large \(r\) is. To avoid this issue, we consider a modification \(\mathcal{M}'(t, \varepsilon, \beta)\) of \(\mathcal{N}(t, \varepsilon, \beta)\). This modification is similar to how we modified \(\mathcal{M}\) to get \(\mathcal{M}'\). Recall that \(u_\varepsilon > 0\) is a fixed parameter in the definition of \(\mathcal{M}'\) and \(\mathcal{M}'\) (see (6.21)–(6.22)). We set
\[
\mathcal{M}'(t, \varepsilon, \beta) := \partial \left\{\left|z - u_\varepsilon\right| \leq 2u_\varepsilon\right\} \cup \left\{\left|z - \frac{e^{\sqrt{\varepsilon}(\rho-1)}}{e^{\sqrt{\varepsilon}(\rho-1)} + e^{\sqrt{\varepsilon}(\rho-1)}} \leq \frac{u(t, \beta)}{e^{\sqrt{\varepsilon}(\rho-1)} + e^{\sqrt{\varepsilon}(\rho-1)}}\right\},
\]
counterwise oriented. See Figure 13.

We now define the \(z_1\)-contour
\[
\Gamma(t, \varepsilon) := \mathcal{M}'(t, \varepsilon, -k_1\alpha).
\]
For all $t > 0$ large enough and $\varepsilon > 0$ small enough, we have $V_{\text{res}} = V'_{\text{res}}$, where
\begin{align*}
V''_{\text{res}} := \oint_{\mathcal{N}(t, \varepsilon, -k_1 \alpha)} 1_{\{|p_{z}(z_1)| > r'_2\}} \left( \frac{e^{\sqrt{\varepsilon}(\rho - 1)} + e^{\sqrt{\varepsilon} \rho}}{2\pi i p_{z}(z_1)} \right) \frac{\mathfrak{J}(z_1)}{z_1} dz_1
\end{align*}
is the same as $V_{\text{res}}$ except the $z_1$-contour is replaced by $\mathcal{N}(t, \varepsilon, -k_1 \alpha)$.

Proof. Referring to the definitions (6.36)-(6.37) of $\mathcal{N}(t, \varepsilon, -k_1 \alpha)$ and $\mathcal{M}''(t, \varepsilon, -k_1 \alpha)$ (see also Figure 13), we see that the difference $\mathcal{N}(t, \varepsilon, -k_1 \alpha) - \mathcal{M}''(t, \varepsilon, -k_1 \alpha)$ is the boundary of the crescent
\begin{align*}
\mathcal{G}(t, \varepsilon) := \{ z \in \mathbb{C} : |z - u_s| \leq 2u_s \} \setminus \left\{ z - \frac{e^{\sqrt{\varepsilon}(2\rho - 1)}}{e^{\sqrt{\varepsilon} \rho} + e^{\sqrt{\varepsilon}(\rho - 1)}} \leq \frac{u(t, -k_1 \alpha)}{e^{\sqrt{\varepsilon} \rho} + e^{\sqrt{\varepsilon}(\rho - 1)}} \right\}.
\end{align*}

See Figure 14. With $\partial \mathcal{G}(t, \varepsilon)$ denoting the boundary, counterclockwise oriented, we have
\begin{align*}
V_{\text{res}} - V''_{\text{res}} = \frac{e^{\sqrt{\varepsilon}(\rho - 1)} + e^{\sqrt{\varepsilon} \rho}}{2\pi i} \oint_{\partial \mathcal{G}(t, \varepsilon)} 1_{\{|p_{z}(z_1)| > r'_2\}} \mathfrak{J}(z_1) \frac{1}{z_1 p_{z}(z_1)} \mathfrak{J}_z(t, z_1) dz_1.
\end{align*}
Also, with $\cup \mathcal{G}(t, \varepsilon) \subset \{|z| \leq 3u_s\}$ and with $3u_s < 1$, we see from (6.31) that
\begin{align*}
|p_{z}(z)| > r'_2, \quad z \in \mathcal{G}(t, \varepsilon),
\end{align*}
for all $t > 0$ large enough and $\varepsilon > 0$ small enough. We hence drop the indicator $1_{\{|p_{z}(z_1)| > r'_2\}}$ and write
\begin{align*}
V_{\text{res}} - V''_{\text{res}} = \frac{e^{\sqrt{\varepsilon}(\rho - 1)} + e^{\sqrt{\varepsilon} \rho}}{2\pi i} \oint_{\partial \mathcal{G}(t)} \mathfrak{J}(z_1) \frac{\mathfrak{J}_z(t, z_1)}{z_1 p_{z}(z_1)} dz_1.
\end{align*}

It suffices to check that the integrand $\frac{\mathfrak{J}(z_1)}{z_1 p_{z}(z_1)} \mathfrak{J}_z(t, z_1)$ has no poles within $z_1 \mathcal{G}(t, \varepsilon)$. This was carried out in the proof of Lemma 6.7 already. There we found that the $\varepsilon \to 0$ limit of the poles occurs at $\frac{1}{2}, b_1,$ and $2 - b_1$. With $u_s < \frac{1}{2} \land b_1$, these points sit strictly outside of $\mathcal{G}(t, \varepsilon)$ for all $t > 0$ large enough.
and $\varepsilon > 0$ small enough. Hence, no poles of $\frac{\partial(z_1)}{\partial \theta_1}(z_1)$ enters $G(t, \varepsilon)$, as long as $t > 0$ is large enough and $\varepsilon > 0$ is small enough.

Having introduce the contours $M''(t, \varepsilon, -k_1\alpha)$ and $N(t, \varepsilon, -k_1\alpha)$, hereafter we write $z_1(\theta_1) \in M''(t, \varepsilon, -k_1\alpha)$ for the parametrization depicted in Figure 13, and write $\tilde{z}_1(\theta_1) \in N(t, \varepsilon, -k_1\alpha)$ for the parametrization give in (6.36). We now turn to the auxiliary parameter $k_1 = k_1(\alpha) \in \mathbb{Z}_{\geq 2}$. Similar to previous cases, the parameter $k_1$ is chosen large enough to ensure that

$$r_2' = u(-2\alpha) \geq p_\varepsilon(\tilde{z}_1(0)) + \frac{1}{\sqrt{t+1}} \in \mathbb{R}.$$  

Such a condition holds for a large enough $k_1 = k_1(\alpha, T)$, as can be checked by the same calculations by Taylor expansion as in the $(+-)$-case. We do not repeat the calculations, and fix such $k_1 \in \mathbb{Z}_{\geq 2}$.

Given this condition, using the same argument for obtaining (6.25) in the $(+-)$-case, here we have

$$|p_\varepsilon(\tilde{z}_1(\theta_1))| > r_2'$$ holds only if $|\theta_1| \geq \frac{1}{C(\alpha)(t+1)^{1/4}}.$  \hspace{1cm} (6.25’)

Let us check that, along the contour $z_1 \in M''(t, \varepsilon, -k_1\alpha)$ and $z_1 \in N(t, -k\alpha)$, and we do have the desired Gaussian decay of $|D_{\varepsilon}|$ and $|H_{\varepsilon}|$.

**Lemma 6.9.** Given any $T \in (0, \infty)$ and $\beta \in \mathbb{R}$,

$$|D_{\varepsilon}(t, z)|, \quad |D_{\varepsilon}(t, z)| \leq \frac{C(\beta, T)}{\varepsilon^2(t+1)}, \quad z \in M''(t, \varepsilon, \beta),$$

$$|D_{\varepsilon}(t, z)|, \quad |D_{\varepsilon}(t, z)| \leq \frac{C(\beta, T)}{\varepsilon^2(t+1)}, \quad z \in N(t, \varepsilon, \beta),$$

for all $\theta \in (-\pi, \pi]$, large enough $t \leq \varepsilon^{-2}T$, and small enough $\varepsilon > 0$.

**Proof.** The proof follows the same three-step procedure as the proof of Lemma 6.3. Given the identities (6.8a)--(6.8c), the proof of the first two steps $(Zero \ \theta)$--$(Small \ \theta)$ follows the same argument via Taylor expansion as in Lemma 6.3, and we do not repeat it here. As for the last step $(Large \ \theta)$, as argued in the proof of Lemma 6.6, it amounts to checking the corresponding limiting condition. Recall that $M = \{|z - \frac{1}{2}| = \frac{1}{2}\}$ and recall the definition of $M''$ from (6.22). It is readily checked that $M''(t, \varepsilon, \beta)$ converges uniformly to $M''$ as $(t, \varepsilon) \rightarrow (\infty, 0)$, under their respective polar parametrization, and similarly $N(t, \varepsilon, \beta)$ converges uniformly to $M$ as $(t, \varepsilon) \rightarrow (\infty, 0)$. This being the case, the proof reduces to checking the steepest decent condition (SD.$M$) and (SD.$M''$), which have been verified. \hfill $\square$

We have all the necessary ingredients for estimating $V_{\varepsilon}^a$. 

---

**Figure 14.** The region $G(t, \varepsilon)$. 

---
Proof of Proposition 6.2(c)–(d), the (++)-case, with large enough $t$. The proof begins with the contour deformation described in Section 6.2. The condition (No Pole) is checked by the same argument in the (−−)-case, which gives the decomposition $V^\text{in}_{\varepsilon} = V_{\text{blk}} + V_{\text{res}}$. We next perform the aforementioned re-deformation $\mathcal{M}''(t, \varepsilon, -k_1 \alpha) \rightarrow \mathcal{N}(t, \varepsilon, -k_1 \alpha)$ in $V_{\text{res}}$. Lemma 6.8 ensures that no pole is crossed during this step, giving $V^\text{out}_{\varepsilon} = V_{\text{blk}} + V_{\text{res}}$.

The proof amounts to bounding $V_{\text{blk}}$, $V''_{\text{res}}$, and their gradients. We begin with $V_{\text{blk}}$. In the following we check a sequence of bounds on terms involved in (6.16), and we always assume $z_1 = z_1(\theta_1) \in \mathcal{M}''(t, \varepsilon, -k_1 \alpha)$ and $z_2 = z_2(\theta_2) \in \mathcal{C}_{\varepsilon}(z_1)$ in the course of doing so.

$(V_{\text{blk}, z_1})$ Show that $|z_1|^{x_2-y_1+\mu t-|\mu t|} \leq \exp(-\frac{\alpha |x_2-y_1|}{\sqrt{t+1}+C(\alpha)})$:

With $x_2 - y_1 > 0$ under current assumptions, we need an upper bound on $|z_1|$. To this end, instead of $z_1 \in \mathcal{M}''(t, \varepsilon, -k_1 \alpha)$, let us first consider $\tilde{z}_1 \in \mathcal{N}(t, \varepsilon, -k_1 \alpha)$. This contour $\mathcal{N}(t, \varepsilon, -k_1 \alpha)$ is a circle with a center in $(0, \infty)$. For such circles, the farthest point to the origin occurs at the right-end. This gives

$$\sup_{\tilde{z}_1(\theta_1) \in \mathcal{N}(t, \varepsilon, -k_1 \alpha)} |\tilde{z}_1(\theta_1)| = \tilde{z}_1(0) = \frac{e^{\sqrt{\varepsilon}(2\rho-1)} + u(t, -k_1 \alpha)}{e^{\sqrt{\varepsilon}\rho} + e^{\sqrt{\varepsilon}(p-1)}}.$$ 

Recall from Definition 6.5 the announced convention on Taylor expansion, and expand the last expression in $(\sqrt{\varepsilon}, \frac{1}{\sqrt{t+1}})$ up to the leading order in $\frac{1}{\sqrt{t+1}}$. This gives

$$\sup_{\tilde{z}_1 \in \mathcal{N}(t, \varepsilon, -k_1 \alpha)} |\tilde{z}_1| = 1 + 0 \cdot \sqrt{\varepsilon} - \frac{1}{8} \rho(1-\rho) \varepsilon - \frac{k_1 \alpha}{\sqrt{t+1}} + \ldots,$$

With $k_1 \geq 2$, and $\varepsilon \leq \frac{C(T)}{\sqrt{t+1}}$ under current assumptions, we have

$$\sup_{\tilde{z}_1 \in \mathcal{N}(t, \varepsilon, -k_1 \alpha)} |\tilde{z}_1| \leq u(t, -\alpha), \quad (6.39)$$

for all large enough $t$.

Now, recall from (6.37) that $\mathcal{M}''(t, \varepsilon, -k_1 \alpha)$ differs from $\mathcal{N}(t, \varepsilon, -k_1 \alpha)$ only in $\{|z| \leq 2u_\ast\} \subset \{|z| \leq 3u_\ast\}$. With $3u_\ast < 1$, the bound (6.39) readily implies

$$\sup_{z_1(\theta_1) \in \mathcal{M}''(t, \varepsilon, -k_1 \alpha)} |\tilde{z}_1(\theta_1)| \leq u(t, -\alpha) \lor (3u_\ast) = u(t, -\alpha),$$

for all $t$ large enough. Consequently, $|z_1|^{x_2-y_1+\mu t-|\mu t|} \leq u(t, -\alpha)|x_2-y_1| \leq \exp(-\frac{\alpha |x_2-y_1|}{\sqrt{t+1}+C(\alpha)})$.

$(V_{\text{blk}, z_2})$ Show that $|z_2|^{x_1-y_2+\mu t-|\mu t|} \leq C(\alpha) \exp(-\frac{\alpha |x_1-y_2|}{\sqrt{t+1}+C(\alpha)})$:

With $k_2 := 1$ and with $\tilde{r}_2$ defined in (6.14), we have $|z_2| \leq u(t, -\alpha)$. This and the assumption $x_1 - y_2 > 0$ gives the desired claim.

$(V_{\text{blk}, \mathcal{D}_\varepsilon})$ Show that $|\mathcal{D}_\varepsilon(z_1, z_2)| \leq C(\alpha)(1 + |\theta_1 - \theta_2| \sqrt{t+1})$:

This bound is established by the same argument as in the (−−)-case. We do not repeat it here.

$(V_{\text{blk}, \mathcal{D}_\varepsilon})$ Show that $|\mathcal{D}_\varepsilon(z_1)| \leq C(\alpha, T) \exp(-\frac{\theta_2^2}{C}(t + 1))$:

This is the content of Lemma 6.9.

Given $(V_{\text{blk}, z_1})$–$(V_{\text{blk}, \mathcal{D}_\varepsilon})$, the desired bound on $V_{\text{blk}}$ follows by inserting the bounds into (6.16), and integrating the result. The procedure is the same as the (−−)-case, and we do not repeat it here.

We now turn to $V''_{\text{res}}$. In the following we always assume $\tilde{z}_1 = \tilde{z}_1(\theta_1) \in \mathcal{N}(t, \varepsilon, -k_1 \alpha)$.

$(V''_{\text{res}, \frac{1}{|\mathcal{D}_\varepsilon(z_1)|}})$ Show that $\frac{1}{|\mathcal{D}_\varepsilon(z_1)|} \leq C(\alpha)$:

This is true because $|\mathcal{D}_\varepsilon(z_1)| = u(t, -k_1 \alpha)$.

$(V''_{\text{res}, |\mathcal{D}_\varepsilon(z_1)|})$ Show that $|\mathcal{D}_\varepsilon(z_1)| \leq C(\alpha) \exp(-\frac{\alpha |x_2-y_1|+|x_1-y_2|}{\sqrt{t+1}+C(\alpha)})$:

Set $n_1 := |x_2 - y_1|$ and $n_2 := |x_1 - y_2|$. The assumption $y_1 < y_2$, $x_1 < x_2$ in the (++)-case
yields \( n_1 - 2 \geq n_2 \geq 0 \). Given this, recalling the definition of \( \mathcal{J} \) from (6.15), we write
\[
|\mathcal{J}(\tilde{z}_1)| \leq (|\tilde{z}_1|^{n_1 - n_2 - 2} + |\tilde{z}_1|^{n_1 - n_2})|\tilde{z}_1|p_{\varepsilon}(\tilde{z}_1)|^{n_2}.
\]
Given the bound (6.39) on \( |\tilde{z}_1| \) and given that \( p_{\varepsilon}(\tilde{z}_1)|\tilde{z}_1| = u(t, -k_1 \alpha) \), we have
\[
|\mathcal{J}(\tilde{z}_1)| \leq 2u(t, -\alpha)^{n_1 - n_2 - 2}u(t, -k_1 \alpha)^{n_2}.
\]

With \( k_1 \geq 2 \), the desired result follows:
\[
|\mathcal{J}(\tilde{z}_1)| \leq C(\alpha)e^{-\alpha(n_1 - n_2)}e^{-2\alpha(n_1 - n_2)} = C(\alpha)e^{-\alpha(n_1 - n_2)}.
\]

**(V"_{\text{res}}\mathcal{J}_{\varepsilon})** Show that \( |\mathcal{J}_{\varepsilon}(\tilde{z}_1)| \leq C(\alpha, T)\exp(-\frac{\alpha^2}{C}(t + 1)) \):
This is the content of Lemma 6.9.

Given \((V"_{\text{res}}, \frac{1}{\varepsilon^2})\)\((V"_{\text{res}}, \mathcal{J}_{\varepsilon})\), and the derived constraint (6.25) on \(|\theta_1|\), the desired bound on \( V_{\text{res}} \) follows the same integration procedure as the \((+-)\)-case.

As for the gradient, similarly to the \((+-)\)-case, here we have
\[
|z_j^+ - 1| \leq \frac{1}{\sqrt{\varepsilon + 1}} + |\theta_1|, \ z_1 = z_1(\theta) \in M''(\varepsilon, -k_1 \alpha) \text{ or } N(\varepsilon, -k_1 \alpha), \ z_2 = z_2(\theta) \in C_{\tilde{z}_2}(z_1). \quad (6.28''')
\]
Incorporate this bound into the preceding analysis gives the desired bounds on the gradients. \( \Box \)

### 7. Controlling the Quadratic Variation: Proof of Proposition 5.3

Based on the estimates from Section 6 and the duality of the stochastic 6V model from Section 3, here we prove Proposition 5.3.

#### 7.1. Expanding the Quadratic Variation

The first step toward proving Proposition 5.3 is to find an expression for \( \varepsilon^{-\frac{1}{2}}\Theta_1(t, x)\Theta_2(t, x) \) that exposes the limiting behavior \( 2b_0p_{\varepsilon}(1 - \rho)\sqrt{T^2}(t, x) \). Recall the definition of \( \Theta_1(t, x) \) and \( \Theta_2(t, x) \) from (4.9)–(4.10). With \( \sum_{i=0}^{\infty}p_{\varepsilon}(i - \mu) = 1 \), we rewrite them as
\[
\varepsilon^{-\frac{1}{2}}\Theta_1(t, x) = \varepsilon^{-\frac{1}{2}}(\lambda_x \tau_x^{-1} - 1)Z(t, x) - \varepsilon^{-\frac{1}{2}}\sum_{i=0}^{\infty}p_{\varepsilon}(i - \mu)(Z(t, x) - Z(t, x)),
\]
\[
\varepsilon^{-\frac{1}{2}}\Theta_2(t, x) = \varepsilon^{-\frac{1}{2}}(1 - \lambda_x)Z(t, x) + \varepsilon^{-\frac{1}{2}}\sum_{i=0}^{\infty}p_{\varepsilon}(i - \mu)(Z(t, x) - Z(t, x)).
\]

In order to extract the relevant limiting behaviors, in the sequel we will perform a sequence of expansions on the r.h.s. of (7.1)–(7.2). Here, let us prepare some notation to express various error terms throughout the subsequent expansions. We use \( G_{\varepsilon}(t, x_1, \ldots, x_n; x) \) to denote a generic (random) process that has a uniform exponential decay off the point \( x \); and use \( B_{\varepsilon}(t, x_1, \ldots, x_n) \) to denote a generic uniformly bounded (random) process. More precisely, there exists deterministic \( a > 0, C < \infty \) such that, for all \( \varepsilon \in (0, 1), t \in \mathbb{Z}_{\geq 0}, x_1, \ldots, x_n, x \in \Xi(t), \)
\[
|G_{\varepsilon}(t, x_1, \ldots, x_n; x)| \leq C \exp(-a|x_1 - x| - \ldots - a|x_1 - x|),
\]
\[
|B_{\varepsilon}(t, x_1, \ldots, x_n)| \leq C.
\]

With these notation we write \( X_{\text{bddd}}(t, x) \) for a generic expression of the form
\[
X_{\text{bddd}}(t, x) = \sum_{x_1, x_2 \in \Xi(t)} G_{\varepsilon}(t, x_1, x_2; x)Z(t, x_1)Z(t, x_2),
\]
where ‘bdd’ stands for ‘bounded’. In the sequel \( G_{\varepsilon}, B_{\varepsilon} \) and \( X_{\text{bddd}} \) may differ from line to line, as they refer to generic expressions of the declared type. Under this notation, we view expression of the type \( \varepsilon^uX_{\text{bddd}}(t, x), u > 0, \text{small and negligible} \).

We will also consider expressions that involve gradients. To motivate the definitions of the following expressions, let us first consider an expansion of \( \nabla Z(t, x) \). Recall that \( \nabla f(x) := f(x + 1) - f(x) \) denotes the (forward) discrete gradient, and recall from (4.14) that \( \eta_{\varepsilon}(t, x) \in \{0, 1\}, x \in \Xi(t) \), denote
the centered occupation variable. Referring back to the definition (4.3) of $Z$, with $\tau_\varepsilon = \exp(-\sqrt{\varepsilon})$, we see that $\nabla Z(t, x) = (e^{-\sqrt{\varepsilon}(\eta_\varepsilon^+(t, x) - \rho)} - 1)Z(t, x)$. Taylor expanding the exponential gives

$$
\varepsilon^{-\frac{1}{2}} \nabla Z(t, x) = - (\eta_\varepsilon^+(t, x)) + \rho Z(t, x) + \sqrt{\varepsilon} B_\varepsilon(t, x) Z(t, x),
$$

(7.4)

In particular,

$$
\varepsilon^{-\frac{1}{2}} \nabla Z(t, x) = B_\varepsilon(t, x) Z(t, x).
$$

(7.5)

Such a bound (7.5) is pointwise. As it turn out, after a suitable time averaging, expressions that involve $\varepsilon^{-\frac{1}{2}} \nabla$ acting on $Z$ decay to zero (except for a product of two $\varepsilon^{-\frac{1}{2}} \nabla$ evaluated at the same site, see (7.8) and Lemma 7.1 below). The underlying mechanism arises from the structure for the semigroup $V_\varepsilon$: referring to Proposition 6.1, we see that $V_\varepsilon$ gains an extra factor $(t+1)^{-\frac{1}{2}}$ upon taking gradient. This being the case, we view expressions of the type

$$
Z_{\nabla}(t, x_1, x_2) := (\varepsilon^{-\frac{1}{2}} \nabla Z(t, x_1)) Z(t, x_2)
$$
as small, and consider generic linear combinations of them

$$
\mathcal{Y}_{\nabla}(t, x) = \sum_{x_1, x_2 \in \Xi(t)} \gamma_\varepsilon(t, x_1, x_2; x) Z_{\nabla}(t, x_1, x_2),
$$

(7.6)

with some deterministic coefficients $\gamma_\varepsilon(t, x_1, x_2; x)$ that decay exponentially off $x$:

$$
|\gamma_\varepsilon(t, x_1, x_2; x)| \leq C \exp(-a|x_1 - x| - a|x_2 - x|).
$$

(7.7)

We will also consider generic expressions that involves two pieces of gradient:

$$
\mathcal{Y}_{\nabla, \nabla}(t, x) = \sum_{x_1 < x_2 \in \Xi(t)} \gamma_\varepsilon(t, x_1, x_2; x)(\varepsilon^{-\frac{1}{2}} \nabla Z)(t, x_1)(\varepsilon^{-\frac{1}{2}} \nabla Z)(t, x_2),
$$

(7.8)

for some generic deterministic coefficients $\gamma_\varepsilon(t, x_1, x_2; x)$ satisfying (7.7), (and may differ from line to line in the sequel).

Note that in (7.8), the sum ranges over distinct $x_1$ and $x_2$. In fact, diagonal terms $x_1 = x_2$ contains non-negligible contributions:

**Lemma 7.1.** We have that

$$
(\varepsilon^{-\frac{1}{2}} \nabla Z)^2(t, x) - \rho(1 - \rho) Z^2(t, x) = -Z_{\nabla}(t, x, x+1) + \varepsilon^{\frac{3}{2}} B_\varepsilon(t, x) Z^2(t, x).
$$

Proof. To expose the relevant contribution from this expression, we appeal the expansion (7.4) of $\varepsilon^{-\frac{1}{2}} \nabla Z$, square it, followed by using $\eta_\varepsilon^2 = \eta_\varepsilon$. This gives (recall $\eta_\varepsilon$ from (4.14))

$$
(\varepsilon^{-\frac{1}{2}} \nabla Z(t, x))^2 = \left( - \eta_\varepsilon(t, x+1) Z(t, x) + \rho Z(t, x) + \varepsilon^{\frac{1}{2}} B_\varepsilon(t, x) Z(t, x) \right)^2
$$

$$
= \left( \eta_\varepsilon(t, x+1) Z^2(t, x) - 2 \rho \eta_\varepsilon(t, x+1) Z^2(t, x) + \rho^2 Z^2(t, x) + \varepsilon^{\frac{1}{2}} (B_\varepsilon(t, x) Z^2(t, x)) \right)
$$

$$
= \left( (-1 + \rho^2) \eta_\varepsilon^2 Z^2 + \rho^2 Z^2 + \varepsilon^{\frac{1}{2}} B_\varepsilon Z^2 \right)
$$

(7.9)

Use (7.4) in reverse: $\eta_\varepsilon^+ Z = -\varepsilon^{-\frac{1}{2}} \nabla Z + \rho Z + \varepsilon^{\frac{1}{2}} B_\varepsilon Z$, we rewrite the expression $\eta_\varepsilon^2 Z^2$ as $-(\varepsilon^{-\frac{1}{2}} \nabla Z) + \rho Z^2 + \varepsilon^{\frac{1}{2}} B_\varepsilon Z^2$. Inserting this into the last displayed equation gives the desired result. \hfill \Box

Having introduced the necessary notation and tools, we now begin to expand $\Theta_1$ and $\Theta_2$.

**Lemma 7.2.** We have that

$$
\varepsilon^{-1} \Theta_1(t, x) \Theta_2(t, x) - \frac{2b_1 \rho (1 - \rho)}{1 + b_1} Z^2(t, x) = \sqrt{\varepsilon} \chi_{b_1}(t, x) + \mathcal{Y}_{\nabla}(t, x) + \mathcal{Y}_{\nabla, \nabla}(t, x).
$$

(7.10)
Proof. The starting point of the proof is the expressions (7.1)–(7.2) for \( \Theta_1(t,x) \) and \( \Theta_2(t,x) \). First, from (1.3) and \( \tau_\varepsilon^{-1} = e^{\sqrt{\varepsilon}} \), we have that \( \varepsilon^{-\frac{1}{2}} (\lambda_\varepsilon \tau_\varepsilon^{-1} - 1) = (1 - \rho) + O(\varepsilon^{\frac{1}{2}}) \) and that \( \varepsilon^{-\frac{1}{2}} (1 - \lambda_\varepsilon) = \rho + O(\varepsilon^{\frac{1}{2}}) \). Given this, in (7.1)–(7.2) we replace \( \varepsilon^{-\frac{1}{2}} (\lambda_\varepsilon \tau_\varepsilon^{-1} - 1) \) with \( (1 - \rho) \) and replace \( \varepsilon^{-\frac{1}{2}} (1 - \lambda_\varepsilon) \) with \( \rho \), up to errors of the form \( \varepsilon^{\frac{1}{2}} \mathcal{B}_\varepsilon(t,x) \). Further, telescope the expression \( Z(t,x - i) - Z(t,x) \) into \(-\nabla Z(t,x - i) - \nabla Z(t,x - i + 1) - \ldots - \nabla Z(t,x - 1) \). This, combined with (1.3), gives

\[
e^{-\frac{1}{2}} \Theta_1(t,x) = (1 - \rho) Z(t,x) + \sum_{i=0}^{\infty} \sum_{0 < j \leq i} p_\varepsilon(i - \mu_\varepsilon) e^{-\frac{1}{2}} \nabla Z(t,x - j) + \varepsilon^{\frac{1}{2}} \mathcal{B}_\varepsilon(t,x) Z(t,x),
\]

\[
e^{-\frac{1}{2}} \Theta_2(t,x) = \rho Z(t,x) - \sum_{i=0}^{\infty} \sum_{0 < j \leq i} p_\varepsilon(i - \mu_\varepsilon) e^{-\frac{1}{2}} \nabla Z(t,x - j) + \varepsilon^{\frac{1}{2}} \mathcal{B}_\varepsilon(t,x) Z(t,x).
\]

To simplify notation, set \( u_\varepsilon(j) := \sum_{i=j}^{\infty} p_\varepsilon(i - \mu_\varepsilon) \), we write

\[
e^{-\frac{1}{2}} \Theta_1(t,x) = (1 - \rho) Z(t,x) + \sum_{j=1}^{\infty} u_\varepsilon(j) e^{-\frac{1}{2}} \nabla Z(t,x - j) + \varepsilon^{\frac{1}{2}} \mathcal{B}_\varepsilon(t,x) Z(t,x).
\]  
\[
e^{-\frac{1}{2}} \Theta_2(t,x) = \rho Z(t,x) - \sum_{j=1}^{\infty} u_\varepsilon(j) e^{-\frac{1}{2}} \nabla Z(t,x - j) + \varepsilon^{\frac{1}{2}} \mathcal{B}_\varepsilon(t,x) Z(t,x).
\]

The next step is to take the product of (7.9)–(7.10). Let \( A_{1,Z}, A_{1,\nabla}, A_{1,\text{err}} \) denote the respective terms on the r.h.s. of (7.9), and similarly \( A_{2,Z}, A_{2,\nabla}, A_{2,\text{err}} \) for (7.10). In the following we expand

\[
e^{-1} \Theta_1(t,x) \Theta_2(t,x) = (A_{1,Z} + A_{1,\nabla} + A_{1,\text{err}}) (A_{2,Z} + A_{2,\nabla} + A_{2,\text{err}}),
\]

and analyze the resulting terms.

- Indeed, \( A_{1,Z} A_{2,Z} = \rho (1 - \rho) Z^2(t,x) \).
- Next, the term \( A_{1,Z} A_{2,\nabla} + A_{1,\nabla} A_{2,Z} \).

Indeed \( A_{1,Z} A_{2,\nabla} + A_{1,\nabla} A_{2,Z} \) is a linear combination of \( Z(t,x) e^{-\frac{1}{2}} \nabla Z(t,x - j) \), with coefficients \((2\rho - 1) u_\varepsilon(j) \). Let us check that \( u_\varepsilon(j) \) decays exponentially in \(|j|\). Referring back to (4.5), with \( \mu_\varepsilon, \lambda_\varepsilon \to 1 \) as \( \varepsilon \to 0 \), the kernel \( p_\varepsilon \) decays geometrically, uniformly over \( \varepsilon \in (0,1) \):

\[
p_\varepsilon(x) \leq C b_1^j |x|.
\]

From this we see that

\[
|u_\varepsilon(j)| = \sum_{i \geq j} p_\varepsilon(i - \mu_\varepsilon) \leq C \|j| b_1^j \| \leq C e^{-\frac{1}{2} \log b_1 ||j||}.
\]

Given this property (7.12), we conclude that \( A_{1,Z} A_{2,\nabla} + A_{1,\nabla} A_{2,Z} \) is a linear combination of \( Z(t,x) e^{-\frac{1}{2}} \nabla Z(t,x - j) \), with deterministic coefficients that decay exponentially in \(|j|\), whereby

\[
A_{1,Z} A_{2,\nabla} + A_{1,\nabla} A_{2,Z} = \mathcal{N}(t,x).
\]

- We now turn to \( A_{1,\nabla} A_{2,\nabla} \).

With \( A_{1,\nabla} \) and \( A_{2,\nabla} \) both being sums, in the produce of \( A_{1,\nabla} A_{2,\nabla} \), we separate the diagonal and off-diagonal term. Off-diagonal terms form a linear combination of \( e^{-\frac{1}{2}} \nabla Z(x - j) e^{-\frac{1}{2}} \nabla Z(x - j') \), \( j \neq j' \), with coefficient \( u_\varepsilon(j) u_\varepsilon(j') \). Thanks to (7.12), this coefficient decays exponentially in \(|j| + |j'|\).

This being the case, off-diagonal terms jointly contribute an expression of the type \( \mathcal{N}(t,x) \). We hence keep track of only the diagonal terms, and write

\[
A_{1,\nabla} A_{2,\nabla} = -\sum_{j=1}^{\infty} u_\varepsilon(j)^2 (\nabla Z(t,x - j))^2 + \mathcal{N}(t,x).
\]
Lastly, everything else: \((A_{1,Z} + A_{1,\var})A_{2,\err} + A_{1,\err}(A_{2,Z} + A_{2,\var}) + A_{1,\err}A_{2,\err}\).

First, by (7.5), in \(A_{i,\var}\) we replace each \(\varepsilon^{-\frac{1}{2}}\nabla Z(t, x - j)\) with \(B_{\varepsilon}(t, x - j)\). Once this is done, expanding the expression \((A_{1,Z} + A_{1,\var})A_{2,\err} + A_{1,\err}(A_{2,Z} + A_{2,\var}) + A_{1,\err}A_{2,\err}\) gives

\[\varepsilon^{-\frac{1}{2}}(\text{linear combination of } B_{\varepsilon}(t, x - j)B_{\varepsilon}(t, x - j')) Z(t, x)^2.\]

Thanks to (7.12), the coefficients within the linear combination decays exponentially in \(|j| + |j'|\). This gives

\[(A_{1,Z} + A_{1,\var})A_{2,\err} + A_{1,\err}(A_{2,Z} + A_{2,\var}) + A_{1,\err}A_{2,\err} = \varepsilon^{\frac{1}{2}}X_{\bdd}(t, x).\]

Given the preceding discussion, we now have

\[\varepsilon^{-1}\Theta_1(t, x)\Theta_2(t, x) = \rho(1 - \rho)Z^2(t, x) + \sqrt{\varepsilon}X_{\bdd}(t, x) + \mathcal{Y}_{\var}(t, x) + \mathcal{Y}_{\var,\var}(t, x) - \sum_{j=1}^{\infty} u_\varepsilon(j)^2(\varepsilon^{-\frac{1}{2}}\nabla Z(t, x - j))^2.\]  

(7.13)

As shown in Lemma 7.1, the last term in (7.13) contains a non-negligible contribution to \(Z^2(t, x)\). The rest of the proof consists of extracting this contribution. First, using Lemma 7.1, we write

\[\sum_{j=1}^{\infty} u_\varepsilon(j)^2(\varepsilon^{-\frac{1}{2}}\nabla Z(t, x - j))^2 - \rho(1 - \rho)A = \mathcal{Y}_{\var}(t, x) + \varepsilon^{\frac{1}{2}}X_{\bdd}(t, x),\]  

(7.14)

where \(A := \sum_{j=1}^{\infty} u_\varepsilon(j)^2Z^2(t, x - j)\). The focus now is on the term \(A\). We argue that, replacing \(Z(t, x - j)\) with \(\hat{Z}(t, x)\) in \(A\) only produces an affordable error. To see this, write

\[|Z(t, x - j) - \hat{Z}(t, x)| = |e^{\sqrt{\varepsilon}\sum_{i=0}^{j-1}(\eta_{\varepsilon}(t, x - i)) - \rho} - 1|Z(t, x) \leq \sqrt{\varepsilon}|j\varepsilon^{\frac{1}{2}}|Z(t, x).\]  

(7.15)

Now, write \(Z(t, x - j)\) as \(Z(t, x) + Z(t, x - j) - \hat{Z}(t, x)\), with the aid of (7.15) and (7.12), we have

\[A = Z^2(t, x)\sum_{j=1}^{\infty} u_\varepsilon(j)^2 + \varepsilon^{\frac{1}{2}}B_{\varepsilon}(t, x)Z^2(t, x).\]  

(7.16)

With (1.3), and \(b_2 = e^{-\sqrt{\varepsilon}}b_1\), a straightforward calculation from (4.5) gives

\[\sum_{j=1}^{\infty} u_\varepsilon(j)^2 = \frac{1 - b_1}{1 + b_1} + O(\sqrt{\varepsilon}).\]

Using this in (7.16), and inserting the result back into (7.14), we conclude

\[\sum_{j=1}^{\infty} u_\varepsilon(j)^2(\varepsilon^{-\frac{1}{2}}\nabla Z(t, x - j))^2 - \rho(1 - \rho)\frac{1 - b_1}{1 + b_1}Z^2(t, x) = \mathcal{Y}_{\var}(t, x) + \varepsilon^{\frac{1}{2}}X_{\bdd}(t, x).\]

This together with (7.13) gives the desired result. \(\square\)

Lemma 7.2 provides the relevant decomposition of \(\varepsilon^{-1}\Theta_1\Theta_2\) into its limiting expression and residual terms. While we do expect the residual terms \(\varepsilon^{\frac{1}{2}}X_{\bdd}, \mathcal{Y}_{\var},\) and \(\mathcal{Y}_{\var,\var}\) to tend to zero, bounds on the last two terms are not immediate. To see this, recall from Proposition 4.3 that the duality functions for the stochastic 6V model are \(Z(s, x_1)Z(s, x_2)\) and \((\eta_{\varepsilon}^1 Z)(s, x_1)(\eta_{\varepsilon}^2 Z)(s, x_2),\) for \(x_1 < x_2\). On the other hand, the expressions \(\mathcal{Y}_{\var}\) and \(\mathcal{Y}_{\var,\var}\) (as in (7.6) and (7.8)) are linear combinations of \(Z(s, x_1)Z(s, x_2)\), that generally involve \(x_1 = x_2\).

To circumvent this ‘diagonal’ issue, recalling from (7.7) that \(\gamma_\varepsilon\) denotes generic deterministic coefficients with an exponential decay, we consider a slight modification \(X_{\var}\) of \(\mathcal{Y}_{\var}\), which is the same type of expressions with an additional constraint \(|x_1 - x_2| > 1\):

\[X_{\var}(t, x) = \sum_{x_1, x_2 \in \Xi(t), |x_2 - x_1| > 1} \gamma_\varepsilon(t, x_1, x_2)Z_{\var}(s, x_1, x_2).\]
Next, set
\[ \tilde{Z}(t, x_1, x_2) := (\eta^+_c Z)(t, x_1)(\eta^+_c Z)(t, x_2) - \rho^2 Z(t, x_1)Z(t, x_2). \] (7.17)
In place of \( Y_{\nabla, \nabla} \), we consider expressions \( X_{\tilde{Z}} \) of the type
\[ X_{\tilde{Z}}(t, x) = \sum_{x_1 < x_2 \in \Xi(t)} \gamma_{\varepsilon}(t, x_1, x_2; x) \tilde{Z}(t, x_1, x_2). \] (7.18)
The next lemma allows us to trade in \( Y_{\nabla, \nabla} \) for \( X_{\nabla} \) and \( X_{\tilde{Z}} \).

Lemma 7.3. We have that
\[ Y_{\nabla, \nabla}(t, x) = X_{\tilde{Z}}(t, x) + Y_{\nabla}(t, x) + \varepsilon^{\frac{1}{2}} X_{b\ddot{d}}(t, x), \] (7.19)
\[ Y_{\nabla}(t, x) = X_{\nabla}(t, x) + \varepsilon^{\frac{1}{2}} X_{b\ddot{d}}(t, x). \] (7.20)

Proof. Indeed, \( Y_{\nabla, \nabla}(t, x) \) denotes a generic linear combination of
\[ A := (\varepsilon^{-\frac{1}{2}} \nabla Z)(t, x_1)(\varepsilon^{-\frac{1}{2}} \nabla Z)(t, x_2), \quad x_1 < x_2, \]
and \( X_{\tilde{Z}}(t, x) \) denotes a generic linear combination of \( \tilde{Z}(t, x_1, x_2), \ x_1 < x_2 \). This being the case, to prove (7.19), it suffices to show that \( A - \tilde{Z}(t, x_1, x_2) \) is written as a linear combination of \( Z_{\nabla}(t, x_1, x_2) \) and negligible terms that carry an outstanding \( \varepsilon^{\frac{1}{2}} \) factor. To this end, we use (7.4) to expand
\[ A = (- \eta^+_c Z + \rho Z + \varepsilon^{\frac{1}{2}} B_\varepsilon Z)(t, x_1)(- \eta^+_c Z + \rho Z + \varepsilon^{\frac{1}{2}} B_\varepsilon Z)(t, x_2) \]
\[ = \tilde{Z}(t, x_1, x_2) + \rho(- \eta^+_c Z + \rho Z)(t, x_1)Z(t, x_2) + \rho Z(t, x_1)(- \eta^+_c Z + \rho Z)(t, x_2) \]
\[ + \varepsilon^{\frac{1}{2}} B_\varepsilon(t, x_1, x_2)Z(t, x_1)Z(t, x_2). \] (7.21)
In (7.21), further use (7.4) in reverse to write \( - \eta^+_c Z + \rho Z = \varepsilon^{\frac{1}{2}} \nabla Z + \varepsilon^{\frac{1}{2}} B_\varepsilon Z \). We get
\[ A - \tilde{Z}(t, x_1, x_2) = \rho Z_{\nabla}(t, x_1, x_2) + \rho Z_{\nabla}(t, x_2, x_1) + \varepsilon^{\frac{1}{2}} B_\varepsilon(t, x_1, x_2)Z(t, x_1)Z(t, x_2). \]

This gives the desired result (7.19).

As for (7.20), recall that both \( X_{\nabla} \) and \( Y_{\nabla} \) denote generic linear combinations of the same terms. The only difference is in that the former misses those terms with \( |x_1 - x_2| \leq 1 \). Consequently, the result (7.20) follows once we show
\[ Z(t, x + 1)(\varepsilon^{-\frac{1}{2}} \nabla Z(t, x)) = Z(t, x + 2)(\varepsilon^{-\frac{1}{2}} \nabla Z(t, x)) + \varepsilon^{\frac{1}{2}} (B_\varepsilon Z^2)(t, x), \]
\[ Z(t, x)(\varepsilon^{-\frac{1}{2}} \nabla Z(t, x + 1)) = Z(t, x - 1)(\varepsilon^{-\frac{1}{2}} \nabla Z(t, x + 1)) + \varepsilon^{\frac{1}{2}} (B_\varepsilon Z^2)(t, x), \]
\[ Z(t, x)(\varepsilon^{-\frac{1}{2}} \nabla Z(t, x)) = Z(t, x - 2)(\varepsilon^{-\frac{1}{2}} \nabla Z(t, x)) + \varepsilon^{\frac{1}{2}} (B_\varepsilon Z^2)(t, x). \]

Going from the l.h.s. to the r.h.s. amounts to changing \( Z(t, x + 1) \mapsto Z(t, x + 2) \) or changing \( Z(t, x) \mapsto Z(t, x - 1) \); note that the \( \nabla Z \) factor is never changed. Thanks to (7.4), these changes introduce only error of the form \( \varepsilon^{\frac{1}{2}} (B_\varepsilon Z)(t, x) \). Also, by (7.5), \( \varepsilon^{-\frac{1}{2}} \nabla Z(t, x) = (B_\varepsilon Z)(t, x), \varepsilon^{-\frac{1}{2}} \nabla Z(t, x + 1) = (B_\varepsilon Z)(t, x) \). Hence, the overall error caused by the aforementioned changes is indeed of the form \( \varepsilon^{\frac{1}{2}} (B_\varepsilon Z^2)(t, x) \).

\[ \square \]

Lemmas 7.2 and 7.3 immediately yield

Corollary 7.4. We have
\[ \varepsilon^{-1} \Theta_1(t, x)\Theta_2(t, x) - \frac{2b_1(1 - \rho)}{1 + b_1} Z^2(t, x) = \sqrt{\varepsilon} X_{b\ddot{d}}(t, x) + X_{\nabla}(t, x) + X_{\tilde{Z}}(t, x). \]
7.2. Time decorrelation via duality. Given the decomposition from Corollary 7.4, our goal toward proving Proposition 5.3 is to argue that, each type of expression on the r.h.s. is negligible as \( \varepsilon \to 0 \). This is straightforward for \( \sqrt{\varepsilon} X_{\text{bd}(t,x)} \) due to the outstanding \( \varepsilon^{1/2} \) factor. On the other hand, as mentioned earlier, the terms \( X_{\text{V}} \) and \( X_{\text{Z}} \) converge to zero only after time averaging. This being the case, with \( X_{\text{V}} \) and \( X_{\text{Z}} \) being linear combinations of \( Z_{\text{V}} \) and \( Z_{\text{Z}} \), we direct our focus onto bounding

\[
B_{X_{\text{V}}}(\ell, x_1^*, x_2^*) := \mathbb{E} \left( \left( \varepsilon^2 \sum_{s=0}^{\ell-1} Z_{\text{V}}(s, x_1^*(s), x_2^*(s)) \right)^2 \right),
\]

\[
B_{X_{\text{Z}}}(\ell, x_1^*, x_2^*) := \mathbb{E} \left( \left( \varepsilon^2 \sum_{s=0}^{\ell-1} \tilde{Z}(s, x_1^*(s), x_2^*(s)) \right)^2 \right),
\]

for \( \ell \in \mathbb{Z} \cap [0, \varepsilon^{-2}T] \) and \( x_1^* \neq x_2^* \in \mathbb{Z} \), and \( x_i^*(s) := x_i^* - \mu_{\varepsilon} s + |\mu_{\varepsilon} s| \in \Xi(s) \). These expressions are expanded into conditional expectations as

\[
B_{X_{\text{V}}}(\ell, x_1^*, x_2^*) = \varepsilon^4 \left( \sum_{s_1 < s_2 < \ell} + \sum_{s_1 = s_2 < \ell} \right) \mathbb{E} \left[ \mathbb{E} \left[ Z_{\text{V}}(s_2, x_1, x_2) | \mathcal{F}(s_1) \right] Z_{\text{V}}(s_1, x_1, x_2) \right],
\]

\[
B_{X_{\text{Z}}}(\ell, x_1^*, x_2^*) = \varepsilon^4 \left( \sum_{s_1 < s_2 < \ell} + \sum_{s_1 = s_2 < \ell} \right) \mathbb{E} \left[ \mathbb{E} \left[ \tilde{Z}(s_2, x_1, x_2) | \mathcal{F}(s_1) \right] \tilde{Z}(s_1, x_1, x_2) \right],
\]

where \( x_i := x_i^* - \mu_{\varepsilon} s_i + |\mu_{\varepsilon} s_i| \) and the notation \( (\sum + \sum)(\cdot) := \sum(\cdot) + \sum(\cdot) \).

Given (7.24)–(7.25), we set out to bound the following conditional expectations

\[
\mathbb{E} \left[ (Z_{\text{V}}(t + s, x_1, x_2) | \mathcal{F}(s)) Z_{\text{V}}(s, x_1, x_2) \right], \quad \mathbb{E} \left[ \tilde{Z}(t + s, x_1, x_2) | \mathcal{F}(s) \right] \tilde{Z}(s, x_1, x_2),
\]

and show that they decay as \( t \) becomes large. We begin by relating these conditional expectations to the semigroup \( V_{\varepsilon} \) via duality. Recall that \( \nabla_x \) denotes the discrete gradient acting on a designated variable \( x \).

Lemma 7.5. Let \( t, s \in \mathbb{Z}_{\geq 0} \). For all \( x_1 + 1 < x_2 \in \Xi(t) \), we have

\[
\mathbb{E} \left[ Z_{\text{V}}(t, x_1, x_2) | \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \Xi(s)} \varepsilon^{-1/2} \nabla_{y_1} V_{\varepsilon}((y_1, y_2), (x_1, x_2); t) Z(s, y_1) Z(s, y_2),
\]

\[
\mathbb{E} \left[ Z_{\text{V}}(t, x_2, x_1) | \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \Xi(s)} \varepsilon^{-1/2} \nabla_{y_2} V_{\varepsilon}((y_1, y_2), (x_1, x_2); t) Z(s, y_1) Z(s, y_2).
\]

For all \( x_1 < x_2 \in \Xi(t) \), with

\[
V_{\text{V}+\text{V}}((y_1, y_2), (x_1, x_2); t) := \nabla_{y_1} V_{\varepsilon}((y_1 - 1, y_2), (x_1, x_2); t) + \nabla_{y_2} V_{\varepsilon}((y_1, y_2 - 1), (x_1, x_2); t),
\]

we have

\[
\mathbb{E} \left[ \tilde{Z}(t + s, x_1, x_2) | \mathcal{F}(s) \right] = - \sum_{y_1 + 1 < y_2 \in \Xi(s)} \varepsilon^{-1/2} V_{\text{V}+\text{V}}((y_1, y_2), (x_1, x_2); t) Z(s, y_1) Z(s, y_2)
\]

\[
+ \sum_{y_1 + 1 < y_2} \varepsilon^{1/2} V_{\varepsilon}((y_1, y_2), (x_1, x_2); t) B_{\varepsilon}(s, y_1, y_2) Z(s, y_1) Z(s, y_2)
\]

\[
+ \sum_{i < j} \left( \sum_{y \in \Xi(s)} V_{\varepsilon}((y + i, y + j), (x_1, x_2); t) B_{\varepsilon}(s, y) Z(s, y + i') Z(s, y + j') \right).
\]

Remark 7.6. Recall the discussion regarding \( \nabla \)-Weyl chamber from the beginning of Section 6. With the assumption \( x_1 + 1 < x_2 \), the expressions in (7.26)–(7.27) that involve \( \nabla \nabla_{\varepsilon} \) are indeed within their \( \nabla \)-Weyl chambers, and similarly for those in (7.28).
Lemma 7.7. \textit{Given }T < \infty, \textit{there exists }u = u(T) < \infty \textit{ such that, for all } s, t \in [0, \varepsilon^{-2}T] \cap \mathbb{Z} \textit{ and } x_1, x_2 \in \Xi(t),\]

\[1_{\{|x_1 - x_2| > 1\}} \mathbb{E} \left[ \mathbb{E} \left( Z(t + s, x_1, x_2) \middle| \mathcal{F}(s) \right) Z(t, x_1, x_2) \right] \leq C(T) \varepsilon^{\frac{1}{2}} e^{u \varepsilon(|x_1| + |x_2|)};\]

\[\mathbb{E} \left[ \mathbb{E} \left( Z(t + s, x_1, x_2) \middle| \mathcal{F}(s) \right) Z(t, x_1, x_2) \right] \leq C(T) \varepsilon^{\frac{1}{2}} e^{u \varepsilon(|x_1| + |x_2|)}.\]

\textbf{Proof.} The moment bound (5.1) from Proposition 5.1 gives that \(\mathbb{E}[Z(s, y)^4] \leq C(T)e^{u \varepsilon|y|},\) for some fixed \(u = u(T) \in (0, \infty).\) This together with the Cauchy–Schwarz inequality gives

\[\mathbb{E}[Z(s, x_1)Z(s, x_2)Z(s, y_1)Z(s, y_2)] \leq C(T)e^{u \varepsilon(|x_1| + |x_2| + |y_1| + |y_2|)}.\]
To alleviate notation, in the following we often write $V_\varepsilon((y_1, y_2), (x_1, x_2); t) = V_\varepsilon$. Multiply both sides of (7.26)–(7.27) by $Z_\varepsilon(s, x_1, x_2)$. Incorporating both the cases $x_1 + 1 < x_2$ and $x_2 + 1 < x_2$, we write

\[
\mathbb{E} \left[ Z_\varepsilon(t + s, x_1, x_2) | \mathcal{F}(s) \right] Z_\varepsilon(s, x_1, x_2) 1_{\{|x_1 - x_2| > 1\}} \leq C(T) \sum_{y_1 < y_2 \in \Xi(s)} \varepsilon^{-\frac{1}{2}} (|\nabla_{x_1} V_\varepsilon| + |\nabla_{x_2} V_\varepsilon|) Z(s, y_1) Z(s, y_2) |Z_\varepsilon(s, x_1, x_2)|.
\]

Note that (7.26)–(7.27) by $C \sqrt{T}$, as seen from in (7.33). Insert these bounds into (7.31)–(7.32). With $E \leq Z(s, x_1, x_2)$, we get

\[
\mathbb{E} \left[ Z_\varepsilon(t + s, x_1, x_2) | \mathcal{F}(s) \right] Z_\varepsilon(s, x_1, x_2) 1_{\{|x_1 - x_2| > 1\}} \leq C(T) \sum_{y_1 < y_2 \in \Xi(s)} \varepsilon^{-\frac{1}{2}} (|\nabla_{x_1} V_\varepsilon| + |\nabla_{x_2} V_\varepsilon|) Z(s, y_1) Z(s, y_2) Z(s, x_1) Z(s, x_2),
\]

where, in the last inequality, we used (7.5) to write $|Z_\varepsilon(s, x_1, x_2)| \leq C Z(s, x_1) Z(s, x_2)$. Take expectation on both sides using (7.30). For $f : (y_1 < \ldots < y_n) \in \Xi(s)^n \mapsto f(\bar{y}) \in \mathbb{R}$, set

\[
[f]_u := \sum_{y_1 < \ldots < y_n} \left| f(y_1, \ldots, y_n) \right| e^{u(|y_1| + \ldots + |y_n|)}.
\]

We then obtain

\[
\mathbb{E} \left[ \mathbb{E} (Z_\varepsilon(t + s, x_1, x_2) | \mathcal{F}(s) ) Z_\varepsilon(s, x_1, x_2) \right] 1_{\{|x_1 - x_2| > 1\}} \leq e^{u(1+|x_1|+|x_2|)} C(T) \varepsilon^{-\frac{1}{2}} (|\nabla_{x_1} V_\varepsilon|_{u\varepsilon} + |\nabla_{x_2} V_\varepsilon|_{u\varepsilon}).
\]

Note that $[\nabla_{x_1} V_\varepsilon]_{u\varepsilon}, [\nabla_{x_2} V_\varepsilon]_{u\varepsilon}$ are only sums over $y_1 < y_2 \in \Xi(s)$ and are thus still functions of $x_1, x_2$.

A similar procedure starting with (7.28) gives

\[
\mathbb{E} \left[ \mathbb{E} (\bar{Z}(t + s, x_1, x_2) | \mathcal{F}(s) ) \bar{Z}(s, x_1, x_2) \right] \leq e^{u(1+|x_1|+|x_2|)} C(T) \varepsilon^{-\frac{1}{2}} (|\nabla_{x_1} V_\varepsilon|_{u\varepsilon} + |\nabla_{x_2} V_\varepsilon|_{u\varepsilon}) + \varepsilon^\frac{1}{2} |\nabla_{\varepsilon}| + \sum_{|i|, |j| \leq 3} (\nabla_{x_i} V_{\varepsilon, i, j})_{u\varepsilon},
\]

where $\nabla_{x,i} V_{\varepsilon, i, j}(y) := \nabla_{x}[V_{\varepsilon}((y+i, y+j), (x_1, x_2); t)]$.

With $t \leq \varepsilon^{-2} T$, we set $\alpha := 3u \sqrt{T}$ so that $e^{u \frac{x}{\sqrt{T+1}+C(\alpha)}} = \frac{\alpha x}{\sqrt{T+1}+C(\alpha)} e^{2u \varepsilon} > 2u \varepsilon$, for all $\varepsilon > 0$ small enough. For such an exponent $\alpha$, we indeed have

\[
\sum_{y \in \Xi(s)} e^{\alpha y} e^{u \varepsilon |y|} \leq e^{u |x|} \sum_{y \in \Xi(s)} e^{\alpha y} e^{u \varepsilon |x-y|} \leq e^{u \varepsilon |x|} \sum_{y \in \Xi(s)} e^{\alpha y} \leq C(\alpha) e^{u \varepsilon |x|} (t + 1)^\frac{1}{2},
\]

for all $\varepsilon > 0$ small enough. Now, apply the estimates on $|\nabla_{\varepsilon}|$ and $|\nabla_{\varepsilon} V_\varepsilon|$ from Proposition 6.1 with this exponent $\alpha$. We get

\[
[\nabla_{x_1} V_\varepsilon]_{u\varepsilon}, [\nabla_{x_2} V_\varepsilon]_{u\varepsilon} \leq \frac{C(\alpha, T)}{(t+1)^{1/2}} e^{u(1+|x_1|+|x_2|)},
\]

\[
[V_\varepsilon]_{u\varepsilon} \leq C(\alpha, T)e^{u(1+|x_1|+|x_2|)},
\]

\[
[V_{\varepsilon, i, j}]_{u\varepsilon} \leq \frac{C(\alpha, T, i, j)}{(t+1)^{1/2}} e^{u(1+|x_1|+|x_2|)}.
\]

Here, upon taking $[\nabla_{\varepsilon}]_{u\varepsilon}$, the sums over $y_1$ and over $y_2$ of $V_\varepsilon((y_1, y_2), (x_1, x_2), t)$ each produces a factor of $(t+1)^{1/2}$, as seen from in (7.33). Insert these bounds into (7.31)–(7.32). With $\frac{\varepsilon^{-\frac{1}{2}}}{\sqrt{T+1}} + \varepsilon^\frac{1}{2} + \frac{1}{\sqrt{T+1}} \leq \frac{C(T) \varepsilon^{-\frac{1}{2}}}{\sqrt{T+1}}$, and with $\alpha = \alpha(u, T)$, we conclude the desired result. \hfill \square

Recall the definitions of $B_{X_\varepsilon}$ and $B_{X_\varepsilon}$ from (7.22)–(7.23). We are now ready to derive the relevant bounds on these quantities.
Corollary 7.8. Fix $T < \infty$, let $t \in \mathbb{Z} \cap [0, \varepsilon^{-2}T]$ and $x_{s}^{*} \neq x_{r}^{*} \in \mathbb{Z}$. We have

$$B_{X_{t}}(t, x_{1}^{*}, x_{2}^{*}), \quad B_{X_{2}}(t, x_{1}^{*}, x_{2}^{*}) \leq C(T) \varepsilon^{\frac{1}{2}} e^{ue(|x_{1}^{*}|+|x_{2}^{*}|)}.$$  

Proof. This follows by inserting the bounds from Lemma 7.7 into (7.24)–(7.25):

$$B_{X_{t}}(t, x_{1}^{*}, x_{2}^{*}) \leq \varepsilon^{4} \left( 2 \sum_{s_{1}<s_{2}<t} C(T) \frac{e^{-\frac{1}{2}}}{\sqrt{s_{2}-s_{1}+1}} e^{ue(|x_{1}^{*}|+|x_{2}^{*}|)} \right) \leq C(T) \varepsilon^{\frac{1}{2}} e^{ue(|x_{1}^{*}|+|x_{2}^{*}|)}.$$  

Note that, with $|x_{i} - x_{r}^{*}| \leq 1$, we replaced $x_{i}$ with $x_{r}^{*}$ at the cost of increasing the constant $C(T)$ by factors of $e^{ue} \leq C(T)$ (with $u = u(T)$). \qed

We are now ready to prove Proposition 5.3.

Proof of Proposition 5.3. Fix $T < \infty, t \in [0, \varepsilon^{-2}T] \cap \mathbb{Z}$, and $x_{s} \in \mathbb{Z}$, and write $x_{s}(s) := x_{s} - \mu_{\varepsilon} s + |\mu_{\varepsilon} s|$. Given the decomposition in Corollary 7.4, it suffices to prove that

$$\left\| \varepsilon^{2} \sum_{s=0}^{t} A(s, x_{s}(s)) \right\|_{2} \leq \varepsilon^{\frac{1}{2}} C(T) e^{C|x_{s}|},$$  

(7.34)

for $A(s, x) = \varepsilon^{\frac{1}{2}} \mathcal{X}_{\text{bdd}}(t, x), \mathcal{X}(t, x)$, and $\mathcal{X}_{2}(t, x)$, and for all $\varepsilon > 0$ small enough.

Recall from (7.3) that $\mathcal{X}_{\text{bdd}}(t, x)$ denotes a generic linear combination of $Z(t, x_{1})Z(t, x_{2})$, with random but uniformly exponentially decay coefficients $G_{\varepsilon}(t, x_{1}, x_{2}; x)$. Consequently,

$$\left\| \varepsilon^{2} \sum_{s=0}^{t} \varepsilon^{\frac{1}{2}} \mathcal{X}_{\text{bdd}}(s, x_{s}(s)) \right\|_{2} \leq \varepsilon^{\frac{1}{2}} \left( \sum_{s=0}^{t} \sum_{x_{1}, x_{2} \in \mathbb{Z}} e^{-\frac{1}{2}((x_{1}-x_{s}(s))+(x_{2}-x_{s}(s)))} \|Z(s, x_{1})Z(s, x_{2})\|_{2} \right).$$

Given this, together with $|x_{s}(s) - x_{s}| \leq 1$, the statement (7.34) for $A(s, x) = \varepsilon^{\frac{1}{2}} \mathcal{X}_{\text{bdd}}(t, x)$ readily follows from the moment bound (5.1) in Proposition 5.1.

Next, recall that $\mathcal{X}(t, x)$ and $\mathcal{X}_{2}(t, x)$ denote generic linear combinations of $Z(t, x_{1}, x_{2})$ and $Z(t, x_{1}, x_{2})$ with some deterministic coefficients (7.7) that decay exponentially off $x$. This gives

$$\left\| \varepsilon^{2} \sum_{s=0}^{t} \mathcal{X}(s, x_{s}(s)) \right\|_{2} \leq \sum_{x_{1}^{*}<x_{2}^{*} \in \mathbb{Z}} 1_{\{|x_{1}^{*}-x_{2}^{*}|>1\}} e^{-\frac{1}{2}(|x_{1}^{*}(s)-x_{s}(s)|+|x_{2}^{*}(s)-x_{s}(s)|)} B_{\mathcal{X}}(t, x_{1}^{*}, x_{2}^{*})^{1/2},$$

$$\left\| \varepsilon^{2} \sum_{s=0}^{t} \mathcal{X}_{2}(s, x_{s}(s)) \right\|_{2} \leq \sum_{x_{1}^{*}<x_{2}^{*} \in \mathbb{Z}} e^{-\frac{1}{2}(|x_{1}^{*}(s)-x_{s}(s)|+|x_{2}^{*}(s)-x_{s}(s)|)} B_{\mathcal{X}_{2}}(t, x_{1}^{*}, x_{2}^{*})^{1/2},$$

where $x_{1}^{*}(s) := x_{1}^{*} - \mu_{\varepsilon} s + |\mu_{\varepsilon} s|$. Given this, the statement (7.34) for $A(s, x) = \mathcal{X}(t, x)$, and $\mathcal{X}_{2}(t, x)$, readily follows from the bounds in Corollary 7.8. \qed

APPENDIX A. QUADRATIC VARIATION IN ASEP

In this appendix we expand upon the brief discussion from Sections 1.3 and 1.4 and explain how our Markov duality method can be applied to ASEP, which is a simpler limit of the stochastic 6V model. We will not carry out the necessary analysis, but rather just point to the main steps.

Recall that ASEP is an interacting particle system on $\mathbb{Z}$, where particles inhabit sites index by $\mathbb{Z}$ and jump left and right according to continuous time exponential clocks with rates $\ell > 0$ and $r > 0$ subject to exclusion (jumps to occupied sites are suppressed). We will assume that $\ell + r = 1$ and set $\tau := r/\ell$. The ASEP height function $N_{\text{ASEP}}(t, x)$ has 1/0 slopes entering occupied/vacant sites as depicted in Figure 6. For ASEP with near-stationary initial data of density $\rho = \frac{1}{2}$ we define a variant\footnote{This follows immediately from (1.26) by a simple tilting and centering.}.
of the Hopf–Cole transform of $N_{\text{ASEP}}(t, x)$ by

$$Z_{\text{ASEP}}(t, x) := \tau^{N_{\text{ASEP}}(t,x)} e^{\frac{t}{2}(1 - 2\sqrt{r})}, \quad t \in [0, \infty), \ x \in \mathbb{Z}.$$ 

This solves the following microscopic SHE:

$$dZ_{\text{ASEP}}(t, x) = \sqrt{r} \Delta Z_{\text{ASEP}}(t, x) dt + dM(t, x), \quad (A.1)$$

where $\Delta f(x) := f(x + 1) + f(x - 1) - 2f(x)$ denotes the discrete Laplacian, and, for each $x \in \mathbb{Z}$, the process $t \mapsto M(t, x)$ is a martingale.

Under weak asymmetry scaling, i.e., $\tau = \tau_\varepsilon := e^{-\sqrt{\varepsilon}}$ and $(t, x) \mapsto (\varepsilon^{-2}t, \varepsilon^{-1}x)$, an informal scaling argument applied to (A.1) indicates that the equation should converge to the continuum SHE. Key to establishing this convergence is the identification of the limiting quadratic variation of $M(t, x)$. Under weak asymmetry scaling, the optional quadratic variation of $M(t, x)$ reads

$$d\langle M(t, x), M(t, x') \rangle = \varepsilon 1_{\{x=x'\}} \left( \left( 1 + \frac{\varepsilon}{2} B_\varepsilon(t, x) \right) Z^2_{\text{ASEP}}(t, x) + \tilde{F}_\varepsilon(t, x) \right) dt, \quad (A.2)$$

where, following notations in Section 7, $B_\varepsilon(t, x)$ is a generic, uniformly bounded process, and

$$\tilde{F}_\varepsilon(t, x) := \varepsilon^{-\frac{1}{2}} \nabla Z_{\text{ASEP}}(t, x)e^{-\frac{1}{2}} \nabla Z_{\text{ASEP}}(t, x - 1). \quad (A.3)$$

Referring to the r.h.s. of (A.2), we see that $\varepsilon^{-\frac{1}{2}} B_\varepsilon(t, x)$ is indeed negligible compared to the constant $\frac{1}{2}$ factor. Key to identifying the limiting behavior is to argue that $\tilde{F}(t, x)$ is also negligible. With $\nabla Z_{\text{ASEP}}(t, x) = (e^{-\sqrt{\varepsilon}(t,x+1)} - 1)Z_{\text{ASEP}}(t, x)$, we indeed have $\tilde{F}_\varepsilon(t, x) = B_\varepsilon(t, x)Z^2_{\text{ASEP}}(t, x)$, i.e., pointwise bounded up to a multiplicative factor of $Z^2_{\text{ASEP}}(t, x)$. On the other hand, it is conceivable that this term $\tilde{F}(t, x)$ does not tend to zero pointwise, i.e., $\tilde{F}(t, x) \not\rightarrow 0$. The crux of the convergence result is to prove that this term converges to zero after time-averaging:

$$E \left[ \left( \varepsilon^2 \int_0^T \tilde{F}_\varepsilon(t, x) dt \right)^2 \right] \rightarrow 0. \quad (A.4)$$

This is first achieved in [BG97] by showing the decay as $t$ becomes large of the conditional expectation

$$E \left[ \tilde{F}_\varepsilon(t + s, x) | \mathcal{F}(s) \right],$$

where $\mathcal{F}$ denotes the canonical filtration of ASEP. Roughly speaking, the estimate starts by using (A.1) to develop a sequence of inequality that bounds the conditional expectation. ‘Closing’ the series of inequality relies crucially on an identity [BG97, (A.6)] for the (semi)-discrete heat kernel. We do not know of a way to generalize this approach from [BG97] to the stochastic 6V model setting.

Here we provide an alternative approach via duality. The Markov duality method also begins with bounding conditional expectations. However, instead of trying to close a sequence of inequalities, this method provides direct access to the conditional expectations. First, the expression $\tilde{F}_\varepsilon(t, x)$ is not convenient for our purpose. Use $\nabla Z_{\text{ASEP}}(t,x) = (e^{-\sqrt{\varepsilon}(\eta^+(t,x)-\frac{1}{2})} - 1) Z_{\text{ASEP}}(t, x)$ where $\eta^+(t, x) := \eta(t, x + 1)$, and Taylor expand

$$\nabla Z_{\text{ASEP}}(t,x) = \sqrt{\varepsilon} (\frac{1}{2} - \eta^+(t,x)) Z_{\text{ASEP}} + \varepsilon B_\varepsilon(t, x) Z_{\text{ASEP}}, \quad (A.5)$$

where $B_\varepsilon(t, x)$ stands for a generic uniformly bounded process as in Section 7. We can then write

$$\tilde{F}_\varepsilon(t, x) = F_\varepsilon(t, x) + \varepsilon^2 B_\varepsilon(t, x) Z^2_{\text{ASEP}},$$

where

$$F_\varepsilon(t, x) = \frac{1}{2} Z_\nabla(t, x - 1, x) + \frac{1}{2} Z_\nabla(t, x - 1, x + 1) + \tilde{Z}(t, x - 1, x), \quad (A.6)$$

where, following the notation in Section 7,

$$Z_\nabla(t, x_1, x_2) := (e^{-\frac{1}{2}} \nabla Z_{\text{ASEP}}(t,x_1)) Z_{\text{ASEP}}(t, x_2),$$

$$\tilde{Z}(t, x_1, t_2) := (\eta^+ Z_{\text{ASEP}})(t, x_1)(\eta^+ Z_{\text{ASEP}})(t, x_2) - \frac{1}{4} Z_{\text{ASEP}}(t, x_1) Z_{\text{ASEP}}(t, x_2).$$

To see (A.6), we use (A.5) just as in the proof of Lemma 7.3 (for the stochastic 6V model):

$$\tilde{F}_\varepsilon(t, x) = \left( (\frac{1}{2} - \eta^+) Z_{\text{ASEP}} \right)(t, x) \left( (\frac{1}{2} - \eta^+) Z_{\text{ASEP}} \right)(t, x - 1) + \varepsilon^2 B_\varepsilon(t, x) Z^2_{\text{ASEP}}(t, x)$$
In the last step, we replace \( Z_{\text{ASEP}}(t, x) \). Specifically, the notation which we used the notation

\[
\frac{1}{2}((\frac{1}{2} - \eta^+)Z_{\text{ASEP}}(t, x)Z_{\text{ASEP}}(t, x - 1) + \frac{1}{2}((\frac{1}{2} - \eta^+)Z_{\text{ASEP}}(t, x - 1)Z_{\text{ASEP}}(t, x)) + \epsilon^2 B(x, t)Z_{\text{ASEP}}(t, x) = \text{r.h.s of (A.6)} + \epsilon^2 B(x, t)Z_{\text{ASEP}}(t, x).
\]

In the last step, we replace \( Z_{\text{ASEP}}(t, x) \) with \( Z_{\text{ASEP}}(t, x \pm 1) \) costing error of order \( \epsilon^2 B(x, t)Z_{\text{ASEP}}(t, x) \).

As mentioned in Section 1.4, ASEP enjoys self-duality via the functions \( Q \) and \( \tilde{Q} \) defined therein. Specifically, the \( k = 2 \) duality translates (after tilting and centering) into the following statement, in which we used the notation

\[
V_{\text{ASEP}}((y_1, y_2), (x_1, x_2); t) := e^{2t(1 - 2\sqrt{t})} e^{-\frac{1}{2} (x_1 + x_2 - y_1 - y_2)^2} P_{\text{ASEP}}((y_1, y_2) \rightarrow (x_1, x_2); t).
\]

**Proposition A.1.** For all \( x_1 < x_2 \in \mathbb{Z} \) and \( t, s \in [0, \infty) \), we have

\[
E\left[Z_{\text{ASEP}}(t + s, x_1)Z_{\text{ASEP}}(t + s, x_2) \mid \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \mathbb{Z}} V_{\text{ASEP}}((y_1, y_2), (x_1, x_2); t) Z_{\text{ASEP}}(s, y_1)Z_{\text{ASEP}}(s, y_2), \tag{A.7}
\]

\[
E\left[(\eta^+ Z_{\text{ASEP}})(t + s, x_1)(\eta^+ Z_{\text{ASEP}})(t + s, x_2) \mid \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \mathbb{Z}} V_{\text{ASEP}}((y_1, y_2), (x_1, x_2); t) (\eta^+ Z_{\text{ASEP}})(s, x_1)(\eta^+ Z_{\text{ASEP}})(s, x_2). \tag{A.8}
\]

Proposition A.1 provides the necessary ingredients for expressing conditional expectations for the relevant quantities. Specifically, with \( \tilde{Z}(t, x - 1, x) \) being an linear combination the two observables in (A.7) and in (A.8) at \( (x_1, x_2) = (x - 1, x) \) we have

\[
E\left[\tilde{Z}(t + s, x - 1, x) \mid \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \mathbb{Z}} V_{\text{ASEP}}((y_1, y_2), (x - 1, x); t) \tilde{Z}(t, y_1, y_2). \tag{A.9}
\]

Likewise, \( Z_{\nabla}(t, x, x-1) \) is the difference of \( Z_{\text{ASEP}}(t, x+1)Z_{\text{ASEP}}(t, x-1) \) and \( Z_{\text{ASEP}}(t, x)Z_{\text{ASEP}}(t, x-1) \). Taking the difference of (A.7) for \((x_1, x_2) = (x + 1, x - 1)\) and for \((x, x - 1)\) gives

\[
E\left[Z_{\nabla}(t + s, x, x - 1) \mid \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \mathbb{Z}} \epsilon^{-\frac{1}{2}} \nabla_{x_1} V_{\text{ASEP}}((y_1, y_2), (x_1, x_2); t) \big|_{(x_1, x_2) = (x, x-1)} Z_{\text{ASEP}}(s, y_1)Z_{\text{ASEP}}(s, y_2),
\]

where \( \nabla_{x_1} \) denotes the discrete (forward) gradient acting on the variable \( x_1 \). Similarly,

\[
E\left[Z_{\nabla}(t + s, x - 1, x + 1) \mid \mathcal{F}(s) \right] = \sum_{y_1 < y_2 \in \mathbb{Z}} \epsilon^{-\frac{1}{2}} \nabla_{x_1} V_{\text{ASEP}}((y_1, y_2), (x_1, x_2); t) \big|_{(x_1, x_2) = (x-1, x+1)} Z_{\text{ASEP}}(s, y_1)Z_{\text{ASEP}}(s, y_2).
\]

From the perspective of duality, roughly speaking, the mechanism of decay in \( t \rightarrow \infty \) arises from the discrete gradient \( \nabla_{x_1} \). The semigroup \( V_{\text{ASEP}} \) behaves similar to (two copies of) the heat kernel, so that \( \sum_{y_1 < y_2} V_{\text{ASEP}}((y_1, y_2), (x_1, x_2); t) = O(1) \), and each gradient of \( V_{\text{ASEP}} \) effectively produces a factor of \( t^{-1/2} \) for large \( t \). Under the scaling \( \epsilon^{-2} \) of time, namely \( t^{-1/2} \approx \epsilon^{-1} \), we expect to trade in \( \epsilon^{-1/2} \nabla \) for \( \epsilon^{-1/2} \epsilon^{-1} = \epsilon^{-1/2} \rightarrow 0 \). In other words, the key heuristic is that the l.h.s of (A.4) behaves as

\[
E\left[\left(\epsilon^2 \int_0^{\epsilon^{-2}T} \tilde{F}(t, x) dt \right)^2 \right] \approx \epsilon^4 \int_0^{\epsilon^{-2}T} \int_0^{\epsilon^{-2}T} \frac{\epsilon^{-1/2}}{\sqrt{t_1 - t_2}} dt_1 dt_2 \approx \epsilon^{\frac{1}{2}} \rightarrow 0. \tag{A.10}
\]

Note that the identity (A.9) in its current form does not involve gradients of \( V_{\text{ASEP}} \). This identity can, however, be rewritten via Taylor expansion and summation by parts in a form that exposes the decay in \( t \rightarrow \infty \). We do not perform this procedure here, and direct the readers to Lemma 7.5, where
the exact same procedure in carried out for the stochastic 6V model. Specifically, the identity (7.28) therein holds with \( (V_{ASEP}, Z_{ASEP}, Z) \) in place of \( (V_\epsilon, Z, \Xi(s)) \), and with \( s, t \in [0, \infty) \) instead of \( Z_{\geq 0} \).

Given the preceding discussion, the task for bounding conditional expectations boils down to estimating the semigroup \( V_{ASEP} \) and its gradients. Thanks to Bethe ansatz, \( V_{ASEP} \) permits an explicit, analyzable formula in terms double contour integrals. Under weak asymmetry scaling, we write \( V_{ASEP} = V_\epsilon,ASEP \) and the formula reads

\[
V_{\epsilon,ASEP}((y_1, y_2), (x_1, x_2); t) := \oint_{C_\epsilon} \oint_{C_\epsilon} \left( z_1^{x_1-y_1} z_2^{x_2-y_2} - \delta_{\epsilon}^{ASEP}(z_1, z_2) z_1^{x_1-y_1} z_2^{x_2-y_2} \right) \prod_{i=1}^2 \frac{e^{\epsilon z_i^{ASEP}(z_i)} dz_i}{2\pi i z_i},
\]

where \( C_\epsilon \) is a counter-clockwise oriented, circular contour centered at origin, with a large enough radius \( r \) so as to include all poles of the integrand, and

\[
\delta_{\epsilon}^{ASEP}(z_1, z_2) := \frac{1 + z_1 z_2 - (e^{-\frac{1}{2}\sqrt{\epsilon}} + e^{\frac{1}{2}\sqrt{\epsilon}}) z_2}{1 + z_1 z_2 - (e^{-\frac{1}{2}\sqrt{\epsilon}} + e^{\frac{1}{2}\sqrt{\epsilon}}) z_1}, \quad \epsilon^{ASEP}(z_1) := \sqrt{\epsilon r}(z + z^{-1} - 2).
\]

This contour integral formula is amenable to steepest decent analysis. Careful analysis jointly in \( (x_1, x_2, y_1, y_2, t) \) should produce the relevant estimates on \( V_{\epsilon,ASEP} \) and its gradient (the result and proof should be analogous to Proposition 6.1). We do not pursue this analysis here.

REFERENCES

[AB16] A. Aggarwal and A. Borodin. Phase transitions in the ASEP and stochastic six-vertex model. arXiv:1607.08684, 2016.
[ACQ11] G. Amir, I. Corwin, and J. Quastel. Probability distribution of the free energy of the continuum directed random polymer in 1 + 1 dimensions. Commun. Pure Appl. Math., 64(4):466–537, 2011.
[Agg16] A. Aggarwal. Current fluctuations of the stationary ASEP and six-vertex model. Duke Math. J., 167:269–384, 2016.
[Agg17] A. Aggarwal. Convergence of the stochastic six-vertex model to the ASEP. Mathematical Physics, Analysis and Geometry, 20(2):3, 2017.
[Ass13] S. Assing. A rigorous equation for the Cole–Hopf solution of the conservative KPZ equation. Stoch. PDEs: Anal. Comp., 1(2):365–388, 2013.
[Bax89] R. J. Baxter. Exactly solved models in statistical mechanics. Academic Press, Inc., London, 1989.
[BBCW17] G. Barraquand, A. Borodin, I. Corwin, and M. Wheeler. Stochastic six-vertex model in a half-quadrant and half-line open ASEP. arXiv:1704.04309, 2017.
[BBW16] A. Borodin, A. Bufetov, and M. Wheeler. Between the stochastic six vertex model and Hall-Littlewood processes. arXiv:1611.09486, 2016.
[BC95] L. Bertini and N. Cancrini. The stochastic heat equation: Feynman–Kac formula and intermittence. J. Statist. Phys., 78(5-6):1377–1401, 1995.
[BCFV15] A. Borodin, I. Corwin, P. Ferrari, and B. Vető. Height fluctuations for the stationary KPZ equation. Math. Phys. Anal. Geo., 18(20):1–95, 2015.
[BCG16] A. Borodin, I. Corwin, and V. Gorin. Stochastic six-vertex model. Duke Math. J., 165(3):563–624, 2016.
[BCS14] A. Borodin, I. Corwin, and T. Sasamoto. From duality to determinants for \( q \)-TASEP and ASEP. Ann. Probab., 42(6):2314–2382, 2014.
[BG97] L. Bertini and G. Giacomin. Stochastic Burgers and KPZ equations from particle systems. Commun. Math. Phys., 183(3):571–607, 1997.
[BG16] A. Borodin and V. Gorin. Moments match between the KPZ equation and the Airy point process. SIGMA Sym. Integr. Geom. Meth. Appl., 12:102, 2016.
[BG18] A. Borodin and V. Gorin. A stochastic telegraph equation from the six-vertex model. arXiv:1803.09137, 2018.
[BL14] P. Bleher and K. Liechty. Random Matrices and the Six-Vertex Model. CRM Monograph Series. American Mathematical Society, 2014.
[BO17] A. Borodin and G. Olshanski. The ASEP and determinantal point processes. Commun. Math. Phys., 353:853–903, 2017.
[BP15a] A. Borodin and L. Petrov. Integrable probability: stochastic vertex models and symmetric functions. Stochastic Processes and Random Matrices: Lecture Notes of the Les Houches Summer School, 104, 2015.
[BP15b] A. Bufetov and L. Petrov. Law of Large Numbers for infinite random matrices over a finite field. Selecta Math., 21:1271–1338, 2015.
I. Corwin, Departments of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027  
E-mail address: corwin@math.columbia.edu

P. Ghosal, Departments of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027  
E-mail address: pg2475@columbia.edu

H. Shen, Departments of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027  
E-mail address: pkushenha@gmail.com

L.-C. Tsai, Departments of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027  
E-mail address: lctsai.math@gmail.com