Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 1–27

Large Deviations for the Asymmetric Simple Exclusion Process

Srinivasa R.S. Varadhan

Abstract.

We explain the large deviation behavior of the totally asymmetric simple exclusion process in one dimension.

§1. Introduction

So far, large deviations from hydrodynamic scaling have been worked out only for systems under diffusive scaling. Large deviation results are presented here for the Totally Asymmetric Simple Exclusion Process or TASEP in one dimension. This work was carried out by Leif Jensen [2] in his PhD dissertation submitted to New York University in the year 2000 and is available at the website

http://www.math.columbia.edu/~jensen/thesis.html

We will present here a detailed sketch of the derivation of the upper bound and a rough outline of how the lower bound is established.

§2. Hydrodynamic limit of TASEP

The Model.

We have a particle system on the integers $\mathbf{Z}$ or (in the periodic case) on $\mathbf{Z}_N$, the integers moduo N. The configuration is $\eta = \{\eta_x : x \in \mathbf{Z}\}$ or $\{\eta_x : x \in \mathbf{Z}_N\}$. The evolution of $\eta(t) = \{\eta_x(t)\}$ is governed by the generator

$$(\mathcal{L}f)(\eta) = \sum_x \eta_x(1 - \eta_{x+1})[f(\eta^{x,x+1}) - f(\eta)]$$

Received December 31, 2002.
Revised May 19, 2003.
Partially Supported by NSF grant 0104343.

where

$$\eta_z^{x,y} = \begin{cases} \eta_z & if \quad z \neq x, y \\ \eta_y & if \quad z = x \\ \eta_x & if \quad z = y \end{cases}$$

This corresponds to the process where the particles independently wait for an exponential time and then jump one step to the right if the site is free. Otherwise they wait for another exponential time. All the particles are doing this simultaneously and independently.

The Scaling.

For each N we consider an initial configuration $\eta_{x,N}$, that may or may not be random. We consider these models for $N \to \infty$. Assume that for some deterministic density function $\rho_0(\xi)$, $0 \leq \rho_0(\cdot) \leq 1$, and every test function $J(\cdot)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum J(\frac{x}{N})\eta_{x,N} = \int J(\xi)\rho_0(\xi)d\xi$$

The limit is taken in probability in the random case. The class of test functions are continuous functions with compact support in $\mathbf{R}$, if we started with $\mathbf{Z}$ and the periodic unit interval $\mathbf{S}$, if we started with $\mathbf{Z}_N$.

Time is speeded up by a factor of N, i.e. the process is viewed at time Nt or equivalently the generator is multiplied by a factor of N. This introduces in a natural way a probability measure P_N on the space of trajectories $\{\eta_x(t) : x \in Z_N \text{ or } Z, t \geq 0\}$.

Theorem 2.1. *(The law of large numbers.) For any $t > 0$, there exists a deterministic density function $\rho(t, \cdot)$, on $\mathbf{R}$ or $\mathbf{S}$ as the case may be, such that*

$$\lim_{N\to\infty} \frac{1}{N} \sum J(\frac{x}{N})\ \eta_x(t) = \int J(\xi)\ \rho(t,\xi)d\xi$$

in probability for every suitable test function. The density $\rho(t, x)$ is determined as the unique weak solution of

$$\rho_t(t,\xi) + [\rho(t,\xi)(1-\rho(t,\xi))]_\xi = 0 \tag{1}$$

with initial condition $\rho(0, \cdot) = \rho_0(\cdot)$, that satisfies the 'entropy condition'.

Remark 2.2. *The entropy condition can be stated in many equivalent forms. For example if $\rho(t, \cdot)$ is a smooth solution, then for any smooth function $h(r)$*

$$[h(\rho(t,\xi))]_t = h'(\rho(t,\xi))\rho_t(t,\xi) = -h'(\rho(t,\xi))(1-2\rho(t,\xi))\rho_\xi(t,\xi)$$

or

$$[h(\rho(t,\xi))]_t + [g(\rho(t,\xi))]_\xi = 0 \tag{2}$$

where g and h are related by

$$g'(r) = h'(r)(1-2r) \tag{3}$$

If $\rho(t,\cdot)$ is only a weak solution, then equation (2) may not hold even weakly. A weak solution of equation (1) is said to satisfy the entropy condition if for every convex function h and the corresponding g defined by equation (3),

$$[h(\rho(t,\xi))]_t + [g(\rho(t,\xi))]_\xi \le 0 \tag{4}$$

holds as a distribution on $[0,T]\times\mathbf{R}$ or $[0,T]\times\mathbf{S}$ as the case may be. Then for any initial value, the weak solution satisfying the entropy condition exists and is unique. The density profile of the TASEP converges to this unique solution.

We will not prove it here. For the special case when the sites are $\mathbf{Z}$ and $\eta_{x,N}(0) = 1$ for $x \le 0$ and 0 otherwise was carried out by Rost [4], who proved that in this case the solution $\rho(t,\xi)$ is the rarefaction wave,

$$\rho(t,\xi) = \begin{cases} 1 & if \quad \xi \le -t \\ \frac{t-\xi}{2t} & if \quad -t \le \xi \le t \\ 0 & if \quad \xi \ge t \end{cases}$$

and the density of the TASEP converges to it. Seppäläinen in [5] obtained a representation of the TASEP with arbitrary initial conditions in terms of a family of coupled processes with initial conditions of Rost type and was able to reduce the general case to the Rost case.

If we look at special solutions of the form

$$\rho(t,\xi) = \begin{cases} \rho & if \quad \xi \le 0 \\ 1-\rho & if \quad \xi \ge 0 \end{cases}$$

then this will be an entropic solution only when $\rho \le \frac{1}{2}$. In particular if $\rho = 1$, although the initial profile in the Rost case is a stationary weak solution it is not entropic. On the other hand if we hold the lead particle from jumping, then nothing can move. So with probability e^{-Nt}, the Rost initial profile can remain intact up to time t. This illustrates that non-entropic solutions are relevant for large deviations.

§3. Large Deviations. Some super exponential estimates

The validity of hydrodynamical scaling depends on some basic facts. We will state them in the periodic case. The needed modifications when we have the entire **Z** are obvious. The 'one block estimate' allows one to replace the microscopic flux by its expectations, given the densities over blocks of size $2k+1$. If

$$\mathcal{E}(N,k,t) = \frac{1}{N}\int_0^t \sum_x |e_{N,k,x}(s)|ds$$

where

$$e_{N,k,x}(s) = |\frac{1}{2k+1}\sum_{y:|y-x|\leq k} \eta_y(s)(1-\eta_{y+1}(s)) - \bar{\eta}_x^k(s)(1-\bar{\eta}_x^k(s))|$$

and $\bar{\eta}_x^k = \frac{1}{2k+1}\sum_{y:|y-x|\leq k}\eta_y$, then

$$\lim_{k\to\infty}\lim_{N\to\infty} E^{P_N}\Big[\mathcal{E}(N,k,t)\Big] = 0$$

The expectation is taken here with respect to the measure P_N that corresponds to some initial profile on the periodic lattice Z_N and evolves according to TASEP dynamics in the speeded up time scale. Then the two block estimate allows one to replace $\bar{\eta}_x^k$ with large k by $\bar{\eta}_x^{N\epsilon}$ with a small ϵ. One can exhibit this in many ways. For instance, if we define,

$$\mathcal{D}(N,\epsilon,k,t) = \int_0^t [\frac{1}{N}\sum_x [\bar{\eta}_{x,N}^k(s)]^2 - \frac{1}{N}\sum_x [\bar{\eta}_{x,N}^{N\epsilon}(s)]^2]ds$$

then, by proving

$$\lim_{\epsilon\to 0}\lim_{k\to\infty}\lim_{N\to\infty} E^{P_N}\Big[\mathcal{D}(N,\epsilon,k,t)\Big] = 0$$

one can establish that any limit of the empirical density is a weak solution of equation (1).

Remark 3.1. *Because of finite propagation speed, basically the effect of any change in a region is only felt over a finite macroscopic domain. This allows us to go back and forth between the periodic and the nonperiodic cases without much effort. If we take the domain large enough then the probability of any effect outside is superexponentially small. So even for large deviations, one can go back and forth.*

Theorem 3.2. *One has the super exponential 'one block' and 'two block estimates'. For any $\delta > 0$,*

$$\limsup_{k\to\infty}\limsup_{N\to\infty}\frac{1}{N}\log P\Big[\mathcal{E}(N,k,t)\geq\delta\Big]=-\infty \tag{5}$$

$$\limsup_{\epsilon\to 0}\limsup_{k\to\infty}\limsup_{N\to\infty}\frac{1}{N}\log P\Big[\mathcal{D}(N,\epsilon,k,t)\geq\delta\Big]=-\infty \tag{6}$$

Sketch of proof: We look at the periodic case. The Dirichlet form

$$D(p)=\sum_{x,\eta}[\sqrt{p(\eta^{x,x+1})}-\sqrt{p(\eta)}]^2$$

can be used in conjunction with the Feynman-Kac formula to provide the first estimate. This is not any different from the symmetric case. The fact that the scaling factor is N and not N^2 does not affect the estimate. It only matters that it is large.

The second estimate on the other hand is a bit tricky. In the symmetric case the proof uses the full strength of the factor N^2, and does not work here. Instead the proof is carried out in several steps. First one proves that there is an exponential error bound, for large deviations from the hydrodynamical limit in the Rost case, by explicit computation. This is not hard and can be done by just following Rost's proof carefully. Then this is extended to arbitrary initial conditions by following through Seppäläinen's proof. One then notices that, by convexity, if $\mathcal{D}(N,\epsilon,k,t)$ does not go to zero, and the one block estimate holds, then the hydrodynamic limit cannot hold. Therefore the two block estimate holds with exponential error probability. Finally a bootstrap argument is used to improve the exponential error probability to a superexponential estimate. The space time region of size $N\times N$ is divided into ℓ^2 grids of size $\frac{N}{\ell}\times\frac{N}{\ell}$. The probability of a significant violation in the two block estimate is $e^{-c\frac{N}{\ell}}$ for one grid. The grids do not influence each other that much. Now the usual Bernoulli large deviation estimate yields a multiplication of the exponent by a factor ℓ^2, that equals the number of grids. If we pick ℓ large we are done.

Corollary 3.3. *Outside the set of weak solutions the probability measure P_N decays superexponentially fast.*

It is then natural to expect that the rate function for large deviations will be a measure of how 'nonentropic' the weak solution is.

§4. Macroscopic and Microscopic Entropies.

A microstate on the configurations on $\mathbf{Z}_N$ is a probability distribution $p_N(\eta)$ on the configurations $\eta \in \{0,1\}^{\mathbf{Z}_N}$. Its entropy (relative to the uniform distribution) is defined as

$$H_N(p_N) = N\log 2 + \sum_\eta p_N(\eta)\log[p_N(\eta)]$$

For a macroscopic density profile $\rho(\xi)$, the corresponding entropy function is defined by

$$\mathcal{H}(\rho(\cdot)) = \log 2 + \int_{\mathbf{S}} [\rho(\xi)\log\rho(\xi) + (1-\rho(\xi))\log(1-\rho(\xi))]d\xi$$

If p_N has asymptotic profile ρ, in the sense that

$$\lim_{N\to\infty}\frac{1}{N}\sum J(\frac{x}{N})\eta_x = \int J(\xi)\rho(\xi)d\xi$$

in probability with respect to p_N, then by Jensen's inequality

$$\liminf_{N\to\infty}\frac{1}{N}H_N(p_N) \geq \mathcal{H}(\rho(\cdot))$$

We need a result of Kosygina [3] that asserts that under certain additional conditions the equality holds, i.e.

$$\lim_{N\to\infty}\frac{1}{N}H_N(p_N) = \mathcal{H}(\rho(\cdot))$$

Two conditions are needed.

- The Dirichlet form is "small"

$$D_N(p_N) = \sum_{x,\eta}[\sqrt{p(\eta^{x,x+1})} - \sqrt{p(\eta)}]^2 = o(N)$$

- The two block estimate holds.

$$\lim_{\epsilon\to 0}\lim_{k\to\infty}\lim_{N\to\infty} E^{p_N}\left[\mathcal{D}(N,\epsilon,k)\right] = 0$$

where

$$\mathcal{D}(N,\epsilon,k) = \frac{1}{N}\sum_x[\bar{\eta}^k_{x,N}]^2 - \frac{1}{N}\sum_x[\bar{\eta}^{N\epsilon}_{x,N}]^2$$

The proof uses the fact that the control of Dirichlet form allows us to estimate $\frac{1}{N}H_N(p_N)$ by

$$\log 2 + E^{P_N}\left[\frac{1}{N}\sum_x [\bar{\eta}_x^k \log \bar{\eta}_x^k + (1-\bar{\eta}_x^k)\log(1-\bar{\eta}_x^k)]\right]$$

and the two block estimate allows k to be replaced by $N\epsilon$ and if the law of large numbers holds then we easily pass to $\mathcal{H}(\rho(\cdot))$, providing the upper bound. The lower bound as we mentioned is essentially Jensen's inequality.

With some additional work the following theorem due to Kosygina can be proved.

Theorem 4.1. *Consider the evolution according to TASEP in the periodic case with any initial conditions. Suppose the hydrodynamic limit holds with some profile $\rho(t,\xi)$. Then for any $\delta > 0$*

$$\limsup_{N\to\infty} \sup_{\delta\le s\le t} |\frac{1}{N}H_N(p_N(s)) - \mathcal{H}(\rho(s,\cdot))| = 0$$

Idea of proof: The discussion above will allow us to control it for most times s. But the entropy is monotone and cannot fluctuate wildly.

Remark 4.2. *Actually the theorem Kosygina will continue to hold even if we modify the dynamics by changing the rates, replacing in the speeded up scale N by $N\lambda_{x,x+1}(s,\eta)$, provided the relative entropy of the modified process with respect to the unperturbed process remains bounded by CN. This is because the estimates on the Dirichlet form, usually obtained by differentiating the entropy at time t, with respect to t can still be derived. Because the two block estimates has superexponential error estimates for the unperturbed process, they will continue to hold for the perturbed process which has relative entropy bounded by CN. Since the proof of Theorem 4.1 depends only on estimates on the Dirichlet form and two block estimates, the Theorem will continue to hold even when we perturb.*

Remark 4.3. *If for some p_N with profile ρ the entropy relation*

$$\limsup_{N\to\infty} |\frac{1}{N}H_N(p_N) - \mathcal{H}(\rho(\cdot))| = 0$$

holds, then from the super additivity of the entropy function over disjoint blocks, one has for the marginal $p_{N,B}$ of p_N on any block of size $N(b-a)$

say from $[Na, Nb]$,

$$\limsup_{N\to\infty} |\frac{1}{N} H_N(p_{N,B}) - \int_a^b h(\rho(\xi))d\xi| = 0$$

§5. Large Deviation. The rate function

The basic space on which we will carry out the large deviation is the space $\Omega = C[[0,T],\mathcal{M}]$ of continuous maps $\rho(t,d\xi)$ of $[0,T]$ into the space $\mathcal{M}$ of nonnegative measures on S. Although under P_N, $\rho(t,d\xi)$ consists of atoms with mass $\frac{1}{N}$, because of exclusion any conceivable limit will be supported on $\rho(t,d\xi)$ that have densities $\rho(t,\xi)d\xi$ that satisfy $0 \le \rho(t,\xi) \le 1$ for all $(t,\xi) \in [0,T] \times \mathbf{S}$ and are weakly continuous as mappings of $[0,T] \to \mathcal{M}$.

The rate function $\mathcal{I}(\rho(\cdot,\cdot))$ is defined as $+\infty$ if $\rho(\cdot,\cdot)$ is not a weak solution of

$$\rho_t + [\rho(1-\rho)]_\xi = 0$$

If it is a weak solution, then

$$\begin{aligned}\mathcal{I}(\rho(\cdot,\cdot)) &= \int_{0+0}^{T-0}\int_{\mathbf{S}} \big[[h(\rho(\cdot,\cdot))]_t + [g(\rho(\cdot,\cdot))]_\xi\big]^+ dt d\xi \\ &= \sup_{J\in\mathcal{J}} \int_0^T \int_{\mathbf{S}} J(t,\xi)\big[[h(\rho(\cdot,\cdot))]_t + [g(\rho(\cdot,\cdot))]_\xi\big] dt d\xi \\ &= -\inf_{J\in\mathcal{J}} \int_0^T \int_{\mathbf{S}} \big[J_t(t,\xi)h(\rho(\cdot,\cdot)) + J_\xi(t,\xi)g(\rho(\cdot,\cdot))\big] dt d\xi\end{aligned}$$

Here $h(r) = r\log r + (1-r)\log(1-r)$ and $g(r)$ as defined by equation (3) is

$$g(r) = r(1-r)\log\frac{r}{(1-r)} - r$$

and

$$\mathcal{J} = \{J(\cdot,\cdot) : 0 \le J(\cdot,\cdot) \le 1, J(0,\cdot) \equiv J(T,\cdot) \equiv 0\}$$

It is interesting to note that the set of weak solutions of nonlinear equations is in general not weakly closed. However a result on compensated compactness, that can be found in Tartar [6], tells us that the set C_ℓ of weak solutions for which $\mathcal{I}(\rho(\cdot,\cdot)) \le \ell$ is in fact compact in the strong topology, guaranteeing that the rate function is indeed lower semi continuous. It is easy to check uniform modulus of continuity in time in the weak topology. So the rate function in fact does have compact level sets.

§6. Upper Bounds

For upper bounds we will use the formulation of Ellis and Dupuis [1]. Suppose $\eta_{x,N}$ is a deterministic initial condition with a profile $\rho_0(\xi)$.

Theorem 6.1. *Suppose P_N is the measure on the configuration space $\{\eta_x(t)\}$ induced by the TASEP and Q_N is such that $Q_N << P_N$ and the measure $\widehat{Q}_N$ induced by Q_N on Ω converges to the degenerate distribution at $\rho(\cdot,\cdot) \in \Omega$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} H(Q_N|P_N) \geq \mathcal{I}(\rho(\cdot,\cdot))$$

Remark 6.2. *This is easily seen to be equivalent to the standard upper bound LDP estimate.*

The proof is broken up into several lemmas.

Lemma 6.3. *Without loss of generality we can assume that Q_N is Markov with rates $N\widehat{\lambda}(t,x,\eta)$.*

Proof. Consider the probability distribution $q_N(t,\eta)$ of $\eta(t)$ at time t under Q_N. We have

$$\frac{1}{N}\sum_x J(\frac{x}{N})\eta_x \to \int J(\xi)\rho(t,\xi)d\xi$$

in probability with respect to $q_N(t,\eta)$. The process Q_N has some rates $N\lambda_N(t,x,\omega)$ of particles jumping from x to $x+1$, that may depend on the past history upto time t. This comes from general martingale theory. One can write the formal generator

$$(\mathcal{L}_{t,\omega}f)(\eta) = N\sum_x \lambda(t,x,\omega)\eta_x(1-\eta_{x+1})[f(\eta^{x,x+1}) - f(\eta)]$$

and with respect to Q_N,

$$f(\eta(t)) - f(\eta(0)) - \int_0^t (\mathcal{L}_{s,\omega}f)(\eta(s))ds$$

will be martingales. By Girsanov formula one can calculate on $[0,T]$,

$$\frac{1}{N}H(Q_N|P_N) = E^{Q_N}\left[\int_0^T [\sum_x \eta_x(t)(1-\eta_{x+1}(t))\theta(\lambda(t,x,\omega))]dt\right]$$

where $\theta(\lambda) = \lambda\log\lambda - \lambda + 1$. If we replace $\lambda(t,x,\omega)$ by its conditional expectation

$$\widehat{\lambda}(t,x,\eta) = E^{Q_N}\left[\lambda(t,x,\omega)|\eta(t)\right]$$

we see that

$$E^{Q_N}[f(\eta(t)) - f(\eta(0))] = E^{Q_N}\left[\int_0^T (\widehat{\mathcal{L}}_t f)(\eta(t))dt\right]$$

with

$$(\widehat{\mathcal{L}}_t f)(\eta) = N\sum_x \widehat{\lambda}(t,x,\eta)\eta_x(1-\eta_{x+1})[f(\eta^{x,x+1}) - f(\eta)]$$

In other words $q_N(t,\eta)$ is the solution of the forward equation corresponding to $\widehat{\mathcal{L}}$. On the other hand, since $\theta(\lambda)$ is a convex function of λ, by Jensen's inequality,

$$\begin{aligned} &E^{Q_N}\left[\eta_x(t)(1-\eta_{x+1}(t))\theta(\lambda(t,x,\omega))\right] \\ &\qquad\qquad \geq E^{Q_N}\left[\eta_x(t)(1-\eta_{x+1}(t))\theta(\widehat{\lambda}(t,x,\omega))\right] \end{aligned}$$

The Markov process with $\widehat{\mathcal{L}}_t$ as generator has the same marginals at time t as Q_N and will work as well. In other words for our theorem we can assume with out loss of generality that Q_N is Markov with rates $N\widehat{\lambda}(x,t,\eta)$. Q.E.D.

Consider the joint probability distribution $q_{N,x,k}(t,\eta)$ at the $2k+1$ sites $[x-k,\ldots,x+k]$ of $\{\eta_y\}$ under $q_N(t,\eta)$. We think of it as function of η that depends on the variables $\{\eta_y : |y-x| \leq k\}$.

We let

$$H(N,x,k,t) = \frac{1}{N}\sum_{\eta\in[0,1]^{2k+1}} q_{N,x,k}(t,\eta)\log q_{N,x,k}(t,\eta)$$

and compute

$$\begin{aligned} H_t(N,x,k,t) &= \frac{1}{N}\sum_{\eta\in[0,1]^{2k+1}} \dot{q}_{N,x,k}(t,\eta)[1+\log q_{N,x,k}(t,\eta)] \\ &= \frac{1}{N}\sum_{\eta\in[0,1]^{2k+1}} \dot{q}_{N,x,k}(t,\eta)\log q_{N,x,k}(t,\eta) \\ &= \frac{1}{N}\sum_{\eta\in[0,1]^{N}} \dot{q}_{N}(t,\eta)\log q_{N,x,k}(t,\eta) \end{aligned}$$

Using the forward equation $\dot{q}_N(t,\eta) = N(\mathcal{L}_t^* q_N)(t,\eta)$, we get

$$\begin{aligned} &H_t(N,x,k,t) \\ &= \sum_{\eta\in[0,1]^N} q_N(t,\eta)\mathcal{L}_t[\log q_{N,x,k}(t,\eta)] \\ &= \sum_{\substack{\eta\in[0,1]^N \\ x-k-1\le y\le x+k}} q_N(t,\eta)\widehat{\lambda}(t,y,\eta)e_{y,y+1}(\eta)\log\frac{q_{N,x,k}(t,\eta^{y,y+1})}{q_{N,x,k}(t,\eta)} \end{aligned}$$

where $e_{y,y+1}(\eta) = \eta_y(1-\eta_{y+1})$. We use the inequality

$$\lambda\log y \le \lambda\log\lambda - \lambda + 1 + e^a y - 1 - a\lambda$$

with the choice of $a = a_{N,x,y,k}(\eta)$ to be made later.

$$\begin{aligned} &H_t(N,x,k,t) \le \\ &\sum_{\substack{\eta\in[0,1]^N \\ x-k-1\le y\le x+k}} q_N(t,\eta)e_{y,y+1}(\eta)[\widehat{\lambda}(t,y,\eta)\log\widehat{\lambda}(t,y,\eta) - \widehat{\lambda}(t,y,\eta) + 1] \\ &+ \sum_{\substack{\eta\in[0,1]^N \\ x-k-1\le y\le x+k}} q_N(t,\eta)e_{y,y+1}(\eta)\frac{e^{a_{N,x,y,k}}q_{N,x,k}(t,\eta^{y,y+1}) - q_{N,x,k}(t,\eta)}{q_{N,x,k}(t,\eta)} \\ &- \sum_{\substack{\eta\in[0,1]^N \\ x-k-1\le y\le x+k}} q_N(t,\eta)e_{y,y+1}(\eta)\widehat{\lambda}(t,y,\eta)a_{N,x,y,k} \end{aligned}$$

We rewrite this as

$$\begin{aligned} &\sum_{\substack{\eta\in[0,1]^N \\ x-k-1\le y\le x+k}} q_N(t,\eta)e_{y,y+1}(\eta)[\widehat{\lambda}(t,y,\eta)\log\widehat{\lambda}(t,y,\eta) - \widehat{\lambda}(t,y,\eta) + 1] \\ &\qquad \ge H_t(N,x,k,t) - A_1(N,x,k,t) + A_2(N,x,k,t) \end{aligned}$$

where

$$\begin{aligned} &A_1(N,x,k,t) = \\ &\sum_{\substack{\eta\in[0,1]^N \\ x-k-1\le y\le x+k}} q_N(t,\eta)e_{y,y+1}(\eta)\frac{e^{a_{N,x,y,k}}q_{N,x,k}(t,\eta^{y,y+1}) - q_{N,x,k}(t,\eta)}{q_{N,x,k}(t,\eta)} \end{aligned}$$

and

$$A_2(N,x,k,t) = \sum_{\substack{\eta\in[0,1]^N \\ x-k-1\le y\le x+k}} q_N(t,\eta)e_{y,y+1}(\eta)\widehat{\lambda}(t,y,\eta)a_{N,x,y,k}$$

We now multiply by $J(t,\frac{x}{N})$, where $J\in\mathcal{J}$, sum over x, integrate with respect to t from 0 to T and finally multiply by $\frac{1}{(2k+2)N}$,

$$\begin{aligned}
&\frac{1}{N}H(Q_N|P_N)\\
&\ge \int_0^T \frac{1}{(2k+2)N}\sum_{x\in\mathbf{Z}_N} J(t,\frac{x}{N})d\Big[\sum_{\eta\in[0,1]^{2k+1}} q_{N,x,k}(t,\eta)\log q_{N,x,k}(t,\eta)\Big]\\
&\quad - E^{Q_N}\Bigg[\int_0^T\Bigg[\frac{1}{2k+2}\sum_{\substack{x,y\\ x-k-1\le y\le x+k}} J(t,\frac{x}{N})e_{y,y+1}(\eta)\times\\
&\qquad\qquad \frac{e^{a_{N,x,y,k}}q_{N,x,k}(t,\eta^{y,y+1})-q_{N,x,k}(t,\eta)}{q_{N,x,k}(t,\eta)}\Bigg]dt\Bigg]\\
&\quad + E^{Q_N}\Bigg[\int_0^T\Bigg[\frac{1}{2k+2}\sum_{\substack{x,y\\ x-k-1\le y\le x+k}} J(t,\frac{x}{N})e_{y,y+1}(\eta)\times\\
&\qquad\qquad \widehat{\lambda}(t,y,\eta)a_{N,x,y,k}\Bigg]dt\Bigg]\\
&= T_1(N,J(\cdot,\cdot),k)-T_2(N,J(\cdot,\cdot),k)+T_3(N,J(\cdot,\cdot),k)
\end{aligned}$$

Now we have to analyse the terms on the right. Let us look at

$$\begin{aligned}
&T_1(N,J(\cdot,\cdot),k)=\\
&\int_0^T \frac{1}{(2k+2)N}\sum_{x\in\mathbf{Z}_N} J(t,\frac{x}{N})d\Big[\sum_{\eta\in[0,1]^{2k+1}} q_{N,x,k}(t,\eta)\log q_{N,x,k}(t,\eta)\Big]
\end{aligned}$$

Integrating by parts,

$$\begin{aligned}
&T_1(N,J(\cdot,\cdot),k)=\\
&-\int_0^T \sum_{x\in\mathbf{Z}_N} J_t(t,\frac{x}{N})\Big[\frac{1}{(2k+2)N}\sum_{\eta\in[0,1]^{2k+1}} q_{N,x,k}(t,\eta)\log q_{N,x,k}(t,\eta)\Big]dt
\end{aligned}$$

We pick $k=N\epsilon$ and let $\epsilon\to 0$ at the end.

Lemma 6.4. *If for $t \in [0,T]$, $q_N(t,\eta)$ leads to the profile $\rho(t,\cdot)$, then*

$$\lim_{\epsilon \to 0} \lim_{N \to \infty}$$

$$\int_0^T \sum_{x \in \mathbf{Z}_N} J_t(t, \frac{x}{N}) \left[\frac{1}{2N^2\epsilon} \sum_{\eta \in [0,1]^{2N\epsilon}} q_{N,x,N\epsilon}(t,\eta) \log q_{N,x,N\epsilon}(t,\eta) \right] dt$$

$$= \int_0^T \int_{\mathbf{S}} J_t(t,\xi) h(\rho(t,\xi)) dt d\xi$$

Proof. Let us consider the quantity

$$H_N(t,x,\epsilon) = \log 2 + \frac{1}{2N\epsilon} \sum_{\eta \in [0,1]^{2N\epsilon+1}} q_{N,x,N\epsilon}(t,\eta) \log q_{N,x,N\epsilon}(t,\eta)$$

and the measure

$$\mu_N(t,\epsilon) = \frac{1}{N} \sum_x H_N(t,x,\epsilon) \delta_{\frac{x}{N}}$$

We need to prove the weak convergence of

$$\lim_{\epsilon \to 0} \lim_{N \to \infty} \mu_N(t,\epsilon) dt = h(\rho(t,\xi)) dt d\xi$$

Since we are looking at relative entropy with respect to a product measure, i.e. uniform distribution on $[0,1]^{\mathbf{Z}_N}$, it is easy to see that

$$\liminf_{\epsilon \to 0} \liminf_{N \to \infty} \mu_N(t,\epsilon) dt \geq h(\rho(t,\xi)) dt d\xi$$

in view of the remark at the end of the last section. On the other hand the total mass of $\mu_N(t,\epsilon)$ is dominated by the total entropy

$$\log 2 + \frac{1}{N} \sum_{\eta \in [0,1]^{\mathbf{Z}_N}} q_N(t,\eta) \log q_N(t,\eta)$$

and we are done. Q.E.D.

Now we try to control $T_2(N, J(\cdot,\cdot), N\epsilon) - T_2(N, J(\cdot,\cdot), N\epsilon)$ which is more difficult. The interior terms with $x - k - 1 < y < x + k$ are easy. We choose $a_{N,x,y,k} = 0$.

$$E^{Q_N}\left[\frac{1}{2N\epsilon}\sum_{x-k-1<y<x+k} e_{y,y+1}(\eta)\frac{q_{N,x,N\epsilon}(t,\eta^{y,y+1}) - q_{N,x,N\epsilon}(t,\eta)}{q_{N,x,N\epsilon}(t,\eta)}\right]$$

$$= \sum_{\eta\in[0,1]^{\mathbf{Z}_{2N\epsilon}}} \frac{1}{2N\epsilon}\sum_{x-k-1<y<x+k} e_{y,y+1}(\eta) \times$$

$$[q_{N,x,N\epsilon}(t,\eta^{y,y+1}) - q_{N,x,N\epsilon}(t,\eta)]$$

$$= \sum_{\eta\in[0,1]^{\mathbf{Z}_{2N\epsilon}}} \frac{1}{2N\epsilon}\sum_{x-k-1<y<x+k} [\eta_{y+1} - \eta_y] q_{N,x,N\epsilon}(t,\eta)$$

If we carry out a summation by parts in x and integration over t, this leads in the limit to

$$-\int_0^T \int_{\mathbf{S}} J_\xi(t,\xi)\rho(t,\xi)dtd\xi$$

We look next at the boundary terms. Note that $k = [N\epsilon]$. The boundary terms equal $B = B_1 + B_2 + B_3 + B_4$

$$B_1 = -\,E^{Q_N}\left[\int_0^T \left[\frac{1}{2N\epsilon}\sum_x J(t,\frac{x}{N})\eta_{x-k-1}(1-\eta_{x-k}) \times \right.\right.$$

$$\left.\left.\frac{e^{a_{N,x,-,k}} q_{N,x,k}(t,\eta^{x-k-1,x-k}) - q_{N,x,k}(t,\eta)}{q_{N,x,k}(t,\eta)}\right] dt\right]$$

$$B_2 = -\,E^{Q_N}\left[\int_0^T \left[\frac{1}{2N\epsilon}\sum_x J(t,\frac{x}{N})\eta_{x+k}(1-\eta_{x+k+1}) \times \right.\right.$$

$$\left.\left.\frac{e^{a_{N,x,+,k}} q_{N,x,k}(t,\eta^{x+k,x+k+1}) - q_{N,x,k}(t,\eta)}{q_{N,x,k}(t,\eta)}\right] dt\right]$$

$$B_3 = +\,E^{Q_N}\left[\int_0^T \left[\frac{1}{2N\epsilon}\sum_x J(t,\frac{x}{N})\eta_{x-k-1}(1-\eta_{x-k}) \times \right.\right.$$

$$\left.\left.\widehat{\lambda}(t,x-k-1,\eta)a_{N,x,-,k}\right] dt\right]$$

$$B_4 = + E^{Q_N}\left[\int_0^T \left[\frac{1}{2N\epsilon}\sum_x J(t,\frac{x}{N})\eta_{x+k}(1-\eta_{x+k+1}) \times \right.\right.$$
$$\left.\left. \widehat{\lambda}(t,x+k,\eta)a_{N,x,+,k}\right]dt\right]$$

We would like to make the choice of $a_{N,x,-,k} = -u(t,\frac{x-k-1}{N})$ and $a_{N,x,+,k} = u(t,\frac{x+k}{N})$ for some smooth u. We can combine B_3 and B_4 and write

$$\begin{aligned}
&B_3 + B_4\\
&= E^{Q_N}\left[\int_0^T \left[\frac{1}{2N\epsilon}\sum_x [J(t,\frac{x+k+1}{N}) - J(t,\frac{x-k}{N})] \times \right.\right.\\
&\qquad\qquad \left.\left.\eta_x(1-\eta_{x+1})\widehat{\lambda}(t,x,\eta)u(t,\frac{x}{N})\right]dt\right]\\
&= E^{Q_N}\left[\int_0^T \left[\frac{1}{2N\epsilon}\sum_x [J(t,\frac{x+k+1}{N}) - J(t,\frac{x-k}{N})] \times \right.\right.\\
&\qquad\qquad \left.\left.\eta_x(1-\eta_{x+1})u(t,\frac{x}{N})\right]dt\right]\\
&\quad + E^{Q_N}\left[\int_0^T \left[\frac{1}{2N\epsilon}\sum_x [J(t,\frac{x+k+1}{N}) - J(t,\frac{x-k}{N})] \times \right.\right.\\
&\qquad\qquad \left.\left.[\widehat{\lambda}(t,x,\eta) - 1]u(t,\frac{x}{N})\right]dt\right]\\
&\simeq \frac{1}{2\epsilon}\int_0^T\int_{\mathbf{S}}[J(t,\xi+\epsilon) - J(t,x-\epsilon)]\rho(t,\xi)(1-\rho(t,\xi))u(t,\xi)dtd\xi\\
&\quad + Error
\end{aligned}$$

The error term is dominated by

$$CE^{Q_N}\left[\int_0^T \left[\frac{1}{N}\sum_x |\widehat{\lambda}(t,x,\eta(t)) - 1|\right]dt\right]$$

For any $\theta > 0$, there is a constant C_θ such that

$$|\lambda - 1| \le \theta + C_\theta[\lambda\log\lambda - \lambda + 1]$$

Therefore

$$Error \le C\theta + \frac{C_\theta}{N^2}H(Q_N|P_N)$$

We will get an estimate on B_1. The term B_2 is similar. B_1 is estimated by

$$\frac{1}{2N\epsilon}E^{Q_N}\left[\int_0^T \sum_x \eta_{x-1}(1-\eta_x)\left|e^{-u(t,\frac{x}{N})}\frac{q_{N,x,x+2k}(t,\eta^{x-1,x})}{q_{N,x,x+2k}(t,\eta)}-1\right|dt\right]$$

The quantity

$$R_N = \eta_{x-1}(1-\eta_x)\frac{q_{N,x,x+2k}(t,\eta^{x-1,x})}{q_{N,x,x+2k}(t,\eta)}$$

has to be looked at carefully. Take $x = 0$. If we denote $q_{N,x,x+2k}(t,\eta)$ by $f_N(\eta_0,\eta_1,\cdots,\eta_{2k})$ then

$$\begin{aligned} R_N &= \eta_1(1-\eta_0)\frac{f_N(1,\eta_1,\cdots,\eta_{2k})}{f_N(0,\eta_1,\cdots,\eta_{2k})} = \eta_1(1-\eta_0)\frac{p_N(1|\eta_1,\cdots,\eta_{2k})}{p_N(0|\eta_1,\cdots,\eta_{2k})} \\ &\simeq \eta_1(1-\eta_0)\frac{\rho(0)}{1-\rho(0)} \end{aligned}$$

Therefore it follows that

$$\limsup_{N\to\infty} B_1 \le \frac{1}{2\epsilon}\int_0^T\int_{\mathbf{S}} \rho(t,\xi)(1-\rho(t,\xi))\left|\frac{e^{-u(t,\xi)}\rho(t,\xi)}{1-\rho(t,\xi)}-1\right|dtd\xi$$

and similarly

$$\limsup_{N\to\infty} B_2 \le \frac{1}{2\epsilon}\int_0^T\int_{\mathbf{S}} \rho(t,\xi)(1-\rho(t,\xi))\left|\frac{e^{u(t,\xi)}(1-\rho(t,\xi))}{\rho(t,\xi)}-1\right|dtd\xi$$

If we let u approach $\log\frac{\rho(t,\xi)}{1-\rho(t,\xi)}$ both B_1 and B_2 tend to 0 for any positive ϵ. Finally we let $\epsilon \to 0$.

$$\lim_{\epsilon\to 0}\lim_{N\to\infty}[B_3+B_4] = \int_0^T\int_{\mathbf{S}} J_\xi(t,\xi)\rho(t,\xi)(1-\rho(t,\xi))\log\frac{\rho(t,\xi)}{1-\rho(t,\xi)}dtd\xi$$

This proves

$$\begin{aligned} &\liminf_{N\to\infty}\frac{1}{N}H(Q_N|P_N) \ge \\ &\quad -\int_0^T\int_{\mathbf{S}} J_t(t,\xi)h(\rho(t,\xi))dtd\xi + \int_0^T\int_{\mathbf{S}} J_\xi(t,\xi)\rho(t,\xi)dtd\xi \\ &\quad -\int_0^T\int_{\mathbf{S}} J_\xi(t,\xi)\rho(t,\xi)(1-\rho(t,\xi))\log\frac{\rho(t,\xi)}{1-\rho(t,\xi)}dtd\xi \\ &= -\int_0^T\int_{\mathbf{S}}[J_t(t,\xi)h(\rho(t,\xi)) + J_\xi(t,\xi)g(\rho(t,\xi))]dtd\xi \end{aligned}$$

Since J is arbitrary, we are done.

§7. Lower Bounds

The situation with the lower bounds is not totally satisfactory. Ideally one should construct an explicit perturbation of the rates that produces a particular profile, the entropy cost of such a perturbation being approximately equal to the rate function for such a large deviation. This one is not able to do at this time. The best we can do is to prove the existence of such perturbations and construct them implicitly. Even this we can do only to produce a single non-entropic shock traveling at a constant speed. By patching together, one can possibly handle a finite number of shocks of varying speeds, even crossing each other forming caustics. However one does not see at the moment how to produce a 'general' non-entropic weak solution, partly because one does not know what it is. Ideally there would be an approximation theorem allowing us to pass from a solution with a finite number of shocks to a general weak solution with a finite large deviation rate.

We will sketch the proof for the simple case of a stationary non-entropic shock at 0 starting from a special initial configuration. Suppose we are given on $\mathbf{Z}$ an initial configuration of particles where every site $x \leq 0$ is filled and every site $x > 0$ is empty. We wish to perturb the standard speeded up TASEP dynamics with new rates $N\lambda_N(t, x, \eta)$, such that for the modified process Q_N, for every test function J with compact support and every $t \in [0, T]$, we have in probability,

$$\lim_{N\to\infty} \sum_x J(\frac{x}{N})\eta_x(t) = \int_{-\infty}^{\infty} J(\xi)\rho(t,\xi)d\xi$$

where $\rho(t, \xi)$ is the following special weak solution.

$$\rho(t,\xi) = \begin{cases} 1 & if \quad x \leq -t \\ \frac{t-x}{2t} & if \quad -t \leq x \leq -t(2\rho-1) \\ \rho & if \quad -t(2\rho-1) \leq x \leq 0 \\ 1-\rho & if \quad 0 \leq x \leq t(2\rho-1) \\ \frac{t-x}{2t} & if \quad t(2\rho-1) \leq x \leq t \\ 0 & if \quad x \geq t \end{cases} \tag{7}$$

Here $\rho > \frac{1}{2}$ and there is a non-entropic shock at 0 where the density jumps from a higher value of $\rho > \frac{1}{2}$ to the lower value of $1 - \rho < \frac{1}{2}$. The rate function for this profile in the interval $[0, T]$ is proportional to T, i.e. equals $c(\rho)T$, where

$$c(\rho) = 2\rho - 1 - 2\rho(1-\rho)\log\frac{\rho}{1-\rho} \tag{8}$$

The problem is to find rates $N\lambda_N(x, \eta)$ such that the process with these new rates has a law of large numbers with the profile $\rho(t, \xi)$ given by (7) and achieve this with an entropy cost that is roughly $c(\rho)N$. We know from the upper bound that we cannot do better. Since we want to slow down particles at or near 0, ideally cutting down the rate at 0 should do it. If we slow down the rate at 0 to $N\lambda$ with some fixed $\lambda < 1$, holding all other rates at N, we will produce a profile of the type we want with some $\rho = \rho(\lambda)$ that is hard to determine, except in the trivial case of $\lambda = 0, \rho = 1, c(0) = 1$. The cost is surely not going to be optimal. We can however lower the rate at several points around 0, depending on the current configuration of particles. The new rates $N\lambda_N(x, \eta)$ will do the trick. We will implicitly construct them. We will then have to see how this will work for any initial condition. After that we need to modify the construction for shocks moving with constant velocity. Then patch things together for one or more shocks with non constant jumps and non constant velocities that may cross each other.

The idea for a single shock is simple enough. A non-entropic shock is entropic if time is reversed. Let us begin with a generator of a TASEP with jumps to the left rather than to the right. The generator is given by

$$(\widehat{\mathcal{L}}f)(\eta) = N\sum_x \eta_{x+1}(1-\eta_x)[f(\eta^{x,x+1}) - f(\eta)]$$

We start with some initial configuration at $t = 0$, that produces the density profile of $\rho(T, \xi)$ specified by equation (7). The hydrodynamical scaling limit will be an entropic solution of

$$\widehat{\rho}_t - [\widehat{\rho}(1-\widehat{\rho})]_x = 0$$

with $\widehat{\rho}_T(0, \xi) = \rho(T, \xi)$. This is seen to be the time reversal of the profile given in (7).

$$\widehat{\rho}_T(t, \xi) = \rho(T - t, \xi)$$

for $0 \le t \le T$ and $\xi \in R$. If we now take the process Q_N corresponding to $\widehat{\mathcal{L}}$ and reverse time to get trajectories $\eta(T - t)$, the new process will have some generator $\mathcal{L}_{N,T,t}$ that is time inhomogeneous and nearly impossible to compute. However it does have the advantage that it has a hydrodynamical limit with a profile that is the time reversed version of $\widehat{\rho}_T(t,\xi)$ which is of course $\rho(t,\xi)$. The entropy will match, because while the forward motion is losing entropy at the shock, the time reversed motion will put it back and this is done by the new rates for the reversed process. If we do not waste entropy at microscopic scale, then the book keeping at micro and macro levels match and will give us the lower bound for large deviations. However the rates for $\widehat{\mathcal{L}}_{N,T,t}$ are too messy and one has to make it independent of T, N and t, and localize it, so that it is transportable and can be used as a module that we can use at any place and time to slow the flow, which is all that any non-entropic solution is expected to do. We start with a fairly general simple calculation.

Let P be a time homogeneous Markov process with trajectories $x(t)$ in a finite time interval $[0,T]$, on a finite state space with generator

$$(Af)(x) = \sum_y c(x,y)f(y).$$

Let $\pi(t,x)$ be the marginal distributions in the time interval $[0,T]$. We denote by $C(x) = -c(x,x) = \sum_{y\neq x} c(x,y)$. Let $\widehat{P}_T$ be the process that corresponds to the time reversed trajcetories $y(t) = x(T-t)$. Although $\widehat{P}_T$ is a Markov process, it is in general time inhomogeneous and will have a generator that depends on the marginals $\pi(\cdot,\cdot)$. We denote its time dependent generator by

$$(\widehat{\mathcal{A}}_{T,t}f)(x) = \sum_{y\neq x} \widehat{c}_T(t,x,y)f(y)$$

and

$$\widehat{C}_T(x) = -\widehat{c}_T(t,x,x) = \sum_y \widehat{c}_T(t,x,y)$$

We can also reverse the generator and define $\widehat{\mathcal{A}}$ as

$$(\widehat{\mathcal{A}}f)(x) = \sum_y \widehat{c}(x,y)f(y)$$

with $\widehat{c}(x,y) = c(y,x)$ for $x \neq y$ and

$$\widehat{C}(x) = -\widehat{c}(x,x) = \sum_{y\neq x} \widehat{c}(x,y) = \sum_{y\neq x\cdot} c(y,x)$$

We denote by $\widehat{Q}_T$, the process with generator $\widehat{\mathcal{A}}$ with initial distribution $\pi(T,\cdot)$.

Theorem 7.1. *We have the following simple formula connecting the function*

$$H(t) = \sum_x \pi(t,x) \log \pi(t,x)$$

and the relative entropy $H(\widehat{P}_T, \widehat{Q}_T)$.

$$H(\widehat{P}_T | \widehat{Q}_T) = H(0) - H(T) + E^{\widehat{Q}_T}\left[\int_0^T [\widehat{C}(x(t)) - C(x(t))]dt\right] \tag{9}$$

Proof. The probabilities $\pi(t,x)$ satisfy the forward equation

$$\frac{d\pi(t,y)}{dt} = \sum_{x \neq y} c(x,y)\pi(t,x) - C(y)\pi(t,y)$$

The time reversed process $\widehat{P}_T$ defined by $y(t) = x(T-t)$ will have marginals $\pi(T-t,y)$ and some generator

$$(\widehat{\mathcal{A}}_{T,t} f)(x) = \sum_y \widehat{c}_T(t,x,y) f(y)$$

Of course

$$\begin{aligned} \frac{d\pi(T-t,y)}{dt} &= -\sum_{x\neq y} c(x,y)\pi(T-t,x) + C(y)\pi(t,y) \\ &= \sum_{x \neq y} \widehat{c}_T(t,x,y)\pi(T-t,x) - \widehat{C}_T(t,y)\pi(T-t,y) \end{aligned}$$

Actually it is not hard to see that for $x \neq y$

$$\pi(T-t,x)\widehat{c}_T(t,x,y) = \widehat{c}(x,y)\pi(T-t,y)$$

and

$$\widehat{C}_T(t,x) = \frac{1}{\pi(T-t,x)} \sum_{y \neq x} \widehat{c}(x,y)\pi(T-t,y)$$

Our goal is to compute the relative entropy

$$\begin{aligned} H(\widehat{P}_T | \widehat{Q}_T) = \int_0^T \sum_x \pi(T-t,x) \sum_{y:y\neq x} [c_T(t,x,y) \log \frac{c_T(t,x,y)}{\widehat{c}(x,y)} \\ - c_T(t,x,y) + \widehat{c}(x,y)]dt \end{aligned}$$

$$
\begin{aligned}
&H(\widehat{P}_T|\widehat{Q}_T)\\
&\quad= \int_0^T \Bigg[\sum_x \pi(T-t,x)\Bigg[\Bigg[\sum_{y:y\neq x} \frac{c(y,x)\pi(T-t,y)}{\pi(T-t,x)} \log \frac{\pi(T-t,y)}{\pi(t-t,x)}\Bigg]\\
&\qquad\qquad\qquad\qquad\qquad\qquad - C_T(t,x) + \widehat{C}(x)\Bigg]\Bigg] dt\\
&\quad= \int_0^T \sum_{x\neq y} c(y,x)\Bigg[\pi(T-t,y)\log\frac{\pi(T-t,y)}{\pi(T-t,x)}\\
&\qquad\qquad\qquad\qquad\qquad - \pi(T-t,y) + \pi(T-t,x)\Bigg] dt\\
&\quad= \int_0^T \sum_{x\neq y} c(y,x)\Bigg[\pi(t,y)\log\frac{\pi(t,y)}{\pi(t,x)} - \pi(t,y) + \pi(t,x)\Bigg] dt
\end{aligned}
$$

We begin by differentiating $H(t) = \sum_y \pi(t,y)\log\pi(t,y)$.

$$
\begin{aligned}
H'(t) &= \frac{d}{dt}\sum_y \pi(t,y)\log\pi(t,y) = \sum_y \pi(t,y)(\mathcal{A}\log\pi(t,\cdot))(y)\\
&= \sum_y \pi(t,y)\Bigg[[\sum_{x\neq y} c(y,x)\log\pi(t,x)] - C(y)\log\pi(t,y)\Bigg]\\
&= \sum_{x\neq y} c(y,x)[\pi(t,y)\log\pi(t,x) - \pi(t,y)\log\pi(t,y)]\\
&= -\sum_{x\neq y} c(y,x)\pi(t,y)\log\frac{\pi(t,y)}{\pi(t,x)}
\end{aligned}
$$

This proves (9). Q.E.D.

Let us start the backward TASEP $\widehat{\mathcal{L}}$, with an initial distribution μ_N concentrated on the finite set

$$
\Omega_{N,L} = \Big\{\eta : \eta_x = 1 \text{ for } x < -NL \quad\text{and}\quad \eta_x = 0 \text{ for } x \geq NL\Big\}
$$

for some $L \geq T$. Our initial distribution μ_N will be a Bernoulli with $\mu_N[\eta_x = 1] = \rho(T, \frac{x}{N})$ given in equation (7). Assume $L \geq T$. Then the

TASEP will have profile $\rho(T-t,\xi)$ and at time $t=T$ end up at $\rho(0,\xi)$. The time reversed process will now be a perturbation of the TASEP going in the right direction with the profile we need. (7.1). We note that

$$\begin{aligned}\widehat{C}(\eta)-C(\eta)&=N\sum_x \eta_x(1-\eta_{x+1})-N\sum_x \eta_{x+1}(1-\eta_x)\\&=N\sum_x[\eta_x-\eta_{x+1}]\equiv N.\end{aligned}$$

Moreover $H(T)=0$ and

$$\begin{aligned}\frac{1}{NT}H(0)&\simeq\frac{1}{T}\int_{-T}^{T}[\rho(T,\xi)\log\rho(T,\xi)+(1-\rho(T,\xi))\log(1-\rho(T,\xi))]d\xi\\&=2(2\rho-1)[\rho\log\rho+(1-\rho)\log(1-\rho)]\\&\quad+2\int_{2\rho-1}^{1}[\frac{1-\xi}{2}\log\frac{1-\xi}{2}+\frac{1+\xi}{2}\log\frac{1+\xi}{2}]d\xi\\&=2\rho-2-2\rho(1-\rho)\log\frac{\rho}{1-\rho}\end{aligned}$$

The relative entropy can now be computed using formula (9) and is seen to be asymptotic to CTN with

$$\begin{aligned}C=c(\rho)&=2\rho-2-2\rho(1-\rho)\log\frac{\rho}{1-\rho}+1\\&=2\rho-1-2\rho(1-\rho)\log\frac{\rho}{1-\rho}\end{aligned}$$

agreeing with (8).

This perturbation is neither stationary in time nor local in nature. We need to modify it.

The special initial configuration of particles at every site $x\le 0$ and no particles at any site $x>0$ will be denoted by $\bar{\eta}$. Let $N(T)$ be the number of transitions from 0 to 1 during $[0,T]$ for the TASEP. Let P be the unpertubed TASEP from this special configurartion. Our initial goal is to construct, for each given ρ a perturbed measure Q_ρ such that $Q_\rho << P$, with

$$E^{Q_\rho}[N(T)]\simeq\rho(1-\rho)T$$

and

$$H(Q_\rho|P)\simeq Tc(\rho)$$

with $c(\rho)$ given by (8). We wish to do this with a local, time independent perturbation at least approximately. We can work out the algebra and restate it as trying to make for $a < \frac{1}{4}$,

$$E^{Q_\rho}[N(T)] \simeq aT$$

with an entropy cost not exceeding

$$I(a) = \sqrt{1-4a} - 2a \log \frac{1+\sqrt{1-4a}}{1-\sqrt{1-4a}} \tag{10}$$

We consider for $\sigma > 0$,

$$U(\sigma, t, \eta) = E^\eta[e^{-\sigma N(t)}]$$

where η is the initial configurartion. First we note that by a simple coupling argument

$$U(\sigma, t, \bar\eta) \le U(\sigma, t, \eta) \le U(\sigma, t, \bar\eta) e^{\sigma g(\eta)}$$

with

$$g(\eta) = \sum_{x \le 0} (1 - \eta_x) + \sum_{x > 0} \eta_x$$

for all configurations η with only a finite number of ocupied sites x with $x > 0$ and finite number of empty sites x with $x \le 0$. By Markov property, if

$$A(\sigma, t) = \inf_\eta U(\sigma, t, \eta) = U(\sigma, t, \bar\eta)$$

then $A(\sigma, t+s) \ge A(\sigma, t) A(\sigma, s)$ and $-\log A(t)$ is subadditive and

$$\lim_{t \to \infty} \frac{\log A(\sigma, t)}{t} = \sup_t \frac{\log A(\sigma, t)}{t} = -\lambda(\sigma) \tag{11}$$

exists. Moreover

$$e^{-t\lambda(\sigma) - \theta(t)} \le U(\sigma, t, \bar\eta) \le U(\sigma, t, \eta) \le e^{-t\lambda(\sigma) + \sigma g(\eta)}$$

where $\theta(t) = o(t)$ as $t \to \infty$. We can write down a differential equation satisfied by $U(\sigma, t, \eta)$

$$\frac{\partial U(\sigma, t, \eta)}{\partial t} = (\mathcal{L}_\sigma U)(\sigma, t, \eta)$$

with

$$U(\sigma, 0, \eta) = 1$$

The genertaor $\mathcal{L}_\sigma$ is obtained by a combination of Girsanov formula and Feynman-Kac formula. It takes the form

$$(\mathcal{L}_\sigma U)(\eta) = \sum_x c_{x,x+1}\eta_x(1-\eta_{x+1})[U(\eta^{x,x+1}) - U(\eta)] - (1-e^{-\sigma})U(\eta)$$

where $c_{x,x+1} = 1$ for $x \neq 0$ and $c_{0,1} = e^{-\sigma}$.

Theorem 7.2. *Let $\sigma > 0$ be given. For each $\epsilon > 0$, there exists a positive local function $V = V_{\sigma,\epsilon}(\eta)$ that satisfies*

$$(\mathcal{L}_\sigma V)(\eta) + (\lambda(\sigma) + \epsilon)V(\eta) \geq 0$$

for all η.

Proof. As a first step we produce a function that is continuous, i.e. depends weakly on far away sites and then approximate it to get a local function. Our first choice is

$$W(\eta) = \frac{1}{t_0}\int_0^{t_0} \exp[(\lambda(\sigma) + \frac{\epsilon}{2})t]U(\sigma,t,\eta)dt$$

Because of the lower bound on U we can assume that t_0 is large enough so that for all η,

$$e^{[\lambda(\sigma)+\frac{\epsilon}{2}]t_0}U(\sigma,t_0,\eta) \geq 1$$

Let us compute $\mathcal{L}_\sigma W$.

$$\begin{aligned}(\mathcal{L}_\sigma W)(\eta) &= \frac{1}{t_0}\int_0^{t_0} \exp[(\lambda(\sigma) + \frac{\epsilon}{2})t](\mathcal{L}_\sigma U)(\sigma,t,\eta)dt \\ &= \frac{1}{t_0}\int_0^{t_0} \exp[(\lambda(\sigma) + \frac{\epsilon}{2})t]U_t(\sigma,t,\eta)dt \\ &= \frac{1}{t_0}[e^{[\lambda(\sigma)+\frac{\epsilon}{2}]t_0}U(\sigma,t_0,\eta) - 1] - (\lambda(\sigma) + \frac{\epsilon}{2})W(\eta) \\ &\geq -(\lambda(\sigma) + \frac{\epsilon}{2})W(\eta)\end{aligned}$$

Since t_0 is finite, W depends weakly on far away sites, and can be nicely approximated by a V that is local. Q.E.D.

The next step is to use $V = V_{\sigma,\epsilon}$ to construct our perturbations. These perturbations cost entropy but will limit the flow between 0 and 1. There is a trade off and σ is the parameter that will control this trade off. Opitmality in the trade off is reached as $\epsilon \to 0$. We begin by defining the rates

$$c_{x,x+1}(\sigma,\epsilon,\eta) = c_{x,x+1}(\sigma)\frac{V(\eta^{x,x+1})}{V(\eta)}$$

Note that $c_{x,x+1}(\sigma) = 1$ for $x \neq 0$ and $c_{0,1} = e^{-\sigma}$. The corresponding perturbed evolution

$$(\mathcal{L}_{\sigma,\epsilon} f)(\eta) = \sum_x c_{x,x+1}(\sigma,\epsilon,\eta)\eta_x(1-\eta_{x+1})[f(\eta^{x,x+1}) - f(\eta)]$$

is local. Let us speed up by N, and do the hydrodynamic scaling for this perturbation. Let Q_N be the perturbed process and P_N be the original process both rescaled and with the special initial configuration $\bar{\eta}$.

Theorem 7.3. *For every N, we have on the interval* $[0,T]$,

$$H(Q_N|P_N) \leq E^{Q_N}\Big[\log V(\eta(T)) - \log V(\eta(0)) - \sigma N(T)\Big] + NT[\lambda(\sigma)+\epsilon]$$

Proof.

$$\frac{1}{N}H(Q_N|P_N) = E^{Q_N}\left[\int_0^T \phi_{\sigma,\epsilon}(\eta(t))dt\right]$$

where

$$\phi_{\sigma,\epsilon}(\eta) = e_{x,x+1}(\eta)[c_{x,x+1}(\sigma,\epsilon,\eta)\log c_{x,x+1}(\sigma,\epsilon,\eta) - c_{x,x+1}(\sigma,\epsilon,\eta) + 1]$$

We use our definition of $c_{x,x+1}(\sigma,\epsilon,\eta)$ and an easy calculation to get

$$\phi_{\sigma,\epsilon}(\eta) = (\mathcal{L}_{\sigma,\epsilon}\log V)(\eta) - \sigma\, e_{0,1}(\eta)c_{0,1}(\sigma,\epsilon,\eta) - \frac{(\mathcal{L}_\sigma V)(\eta)}{V(\eta)}$$

The proof is completed by integrating with respect to t and taking expectations, noting

$$NE^{Q_N}\left[\int_0^T (\mathcal{L}_{\sigma,\epsilon}\log V)(\eta(t))dt\right]$$
$$= E^{Q_N}\left[\int_0^T [\log V(\eta(T)) - \log V(\eta(0))]\right]$$

Q.E.D.

Now the rest of the argument is relatively straight forward. First, we need a lemma.

Lemma 7.4. *The limit $\lambda(\sigma)$ defined in (11) satisfies*

$$\lambda(\sigma) = \inf_{\rho\geq\frac{1}{2}}[\sigma\rho(1-\rho) + c(\rho)] = \inf_{a\leq\frac{1}{4}}[a\sigma + I(a)]$$

where $I(a)$ is as in (10).

Proof. *Upper bound*: If $F(\rho(\cdot,\cdot))$ is the flow through the origin during $[0,T]$ for a weak solution $\rho(\cdot,\cdot)$, then from the upper bound already established

$$T\,\lambda(\sigma) \geq -\inf_{\rho(\cdot,\cdot)}\Big[\sigma\,F(\rho(\cdot,\cdot)) + I(\rho(\cdot,\cdot))\Big]$$

If one fixes $F(\rho(\cdot,\cdot)) = aT = \rho(1-\rho)T$, the infimum of $I(\rho(\cdot,\cdot))$ is shown by a variational argument to equal $Tc(\rho) = TI(a)$.

Lower bound: By a simple calculation using Jensen's inequality

$$\begin{aligned}\log E^{P_N}\Big[e^{-\sigma N(T)}\Big] &= \log E^{Q_N}\Big[e^{-\sigma N(T)}\frac{dP_N}{dQ_N}\Big]\\ &= \log E^{Q_N}\Big[e^{-\sigma N(T)+\log[\frac{dP_N}{dQ_N}]}\Big]\\ &= \log E^{Q_N}\Big[e^{-\sigma N(T)-\log[\frac{dQ_N}{dP_N}]}\Big]\\ &\geq -E^{Q_N}\Big[\sigma N(T) + \log[\frac{dQ_N}{dP_N}]\Big]\\ &= -E^{Q_N}\big[\sigma N(T)\big] - H(Q_N|P_N)\end{aligned}$$

We can take any Q_N and we pick it as the time reversal of the backward TASEP. Our earlier calculations establish the lower bound of $Tc(\rho)$ for the relative entropy and $T\rho(1-\rho)$ for the flow. One checks that $I(a)$ is a strictly convex function of a. Q.E.D.

This proves that if we perturb by $\mathcal{L}_{\sigma,\epsilon}$ and take the limit as $\epsilon \to 0$, the profiles we get will satisfy the entropy condition, will have flow at 0 limited by Ta and the realtive entropy will be bounded by $TI(a)$, where a is dual to σ.

The final step in the proof is to prove that the only profile, that satisfies the entropy condition away from 0, has the rate function bounded by $TI(a)$ and the flow through the origin bounded by Ta, is given by (7) with $\rho > \frac{1}{2}$ chosen so that $a = \rho(1-\rho)$.

References

[1] Dupuis, Paul.; Ellis, Richard S. A weak convergence approach to the theory of large deviations. Wiley Series in Probability and Statistics: A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York,

[2] Jensen, Leif. Large deviations of TASEP. PhD thesis, NYU, 2000.

[3] Kosygina, Elena. The behavior of the specific entropy in the hydrodynamic scaling limit. Ann. Probab. 29 (2001), no. 3, 1086–1110.

[4] Rost, H. Nonequilibrium behaviour of a many particle process: density profile and local equilibria. Z. Wahrsch. Verw. Gebiete 58 (1981), no. 1, 41–53.

[5] Seppäläinen, T. Coupling the totally asymmetric simple exclusion process with a moving interface. I Brazilian School in Probability (Rio de Janeiro, 1997). Markov Process. Related Fields 4 (1998), no. 4, 593–628.

[6] Tartar, L. Compacité par compensation: résultats et perspectives. (French) [Compensated compactness: results and perspectives] Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. IV (Paris, 1981/1982), 350–369, Res. Notes in Math., 84, Pitman, Boston, MA, 1983.

Courant Institute of mathematical Sciences
New York University
New York
USA

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 29–52

Random Path Representation and Sharp Correlations Asymptotics at High-Temperatures

Massimo Campanino [1], **Dmitry Ioffe** [2], **and Yvan Velenik**

Abstract.

We recently introduced a robust approach to the derivation of sharp asymptotic formula for correlation functions of statistical mechanics models in the high-temperature regime. We describe its application to the nonperturbative proof of Ornstein-Zernike asymptotics of 2-point functions for self-avoiding walks, Bernoulli percolation and ferromagnetic Ising models. We then extend the proof, in the Ising case, to arbitrary odd-odd correlation functions. We discuss the fluctuations of connection paths (invariance principle), and relate the variance of the limiting process to the geometry of the equidecay profiles. Finally, we explain the relation between these results from Statistical Mechanics and their counterparts in Quantum Field Theory.

§1. Introduction

In many situations, various quantities of interest can be represented in terms of path-like structures. This is the case, e.g., of correlations in various lattice systems, either in perturbative regimes (through a suitable expansion), or non-perturbatively, as in the ferromagnetic Ising models at supercritical temperatures. Many important questions about the fine asymptotics of these quantities can be reformulated as local limit theorems for these (essentially) one-dimensional objects. In [7], building upon the earlier works [15, 6], we proposed a robust non-perturbative approach to such a problem. It has already been applied successfully in the case of self-avoiding walks, Bernoulli percolation and Ising models.

Received January 17, 2003.
Revised April 14, 2003.

[1] Investigation supported by University of Bologna. Funds for selected research topics and by the MIUR national project "Stochastic processes ..."

[2] Supported by the Fund for the Promotion of Research at the Technion

We briefly review the results that have been thus obtained (see also [8] for a short description of the main ideas of the proof).

Self-avoiding walks. A self-avoiding path ω from 0 to $x \neq 0$ is a sequence of distinct sites $t_0 = 0, t_1, t_2, \ldots, t_n = x$ in $\mathbb{Z}^d$, with $|t_i - t_{i-1}| = 1, i = 1, \ldots, n$ (the restriction to nearest-neighbor jumps can be replaced by arbitrary, possibly weighted, jumps of finite range). Let $\beta < 0$, we are interested in the following quantity:

$$G_\beta^{\mathrm{SAW}}(x) \overset{\triangle}{=} \sum_{\omega:0\to x} e^{\beta|\omega|},$$

where the sum runs over all self-avoiding paths from 0 to x, and $|\omega|$ denotes the length of the path. $G_\beta^{\mathrm{SAW}}(x)$ is finite for all $\beta < \beta_{\mathrm{c}}^{\mathrm{SAW}}$, with $\beta_{\mathrm{c}}^{\mathrm{SAW}} > -\infty$. Actually, $\sum_{x\in\mathbb{Z}^d} G_\beta^{\mathrm{SAW}}(x)$ is finite if and only if $\beta < \beta_{\mathrm{c}}^{\mathrm{SAW}}$.

Bernoulli bond percolation. Let $\beta > 0$. We consider a family of i.i.d. $\{0,1\}$-valued random variables n_e, indexed by the bonds e between two nearest-neighbor sites of $\mathbb{Z}^d$ (again, restriction to nearest-neighbor sites can be dropped); $\mathrm{Prob}_\beta(n(e) = 1) = 1 - e^{-\beta}$. We say that 0 is connected to x ($0 \leftrightarrow x$) in a realization n of these random variables if there is a self-avoiding path ω from 0 to x such that $n_e = 1$ for all increments e along the path. We are interested in the following quantity:

$$G_\beta^{\mathrm{perc}}(x) \overset{\triangle}{=} \mathrm{Prob}_\beta(0 \leftrightarrow x).$$

The high-temperature region $\beta < \beta_{\mathrm{c}}^{\mathrm{perc}}$ is defined through

$$\beta_{\mathrm{c}}^{\mathrm{perc}} \overset{\triangle}{=} \sup\{\beta : \sum_{x\in\mathbb{Z}^d} G_\beta^{\mathrm{perc}}(x) < \infty\} > 0.$$

It is a deep result of [2] that the percolation transition is sharp, i.e.

$$\beta_c = \inf\{\beta : \mathrm{Prob}_\beta(0 \leftrightarrow \infty) > 0\}.$$

Ising model. Let $\beta > 0$. We consider a family of $\{-1,1\}$-valued random variables σ_x, indexed by the sites $x \in \mathbb{Z}^d$. Let $\Lambda_L = \{-L, \ldots, L\}^d$. The probability of a realization σ of the random variables $(\sigma_x)_{x\in\Lambda_L}$, with boundary condition $\overline{\sigma} \in \{-1,1\}^{\mathbb{Z}^d}$, is given by

$$\mu_{\beta,\overline{\sigma},L}(\sigma) \overset{\triangle}{=} (Z_{\beta,\overline{\sigma},L})^{-1} \exp\bigl[\beta \sum_{\substack{\{x,y\}\subset\Lambda_L\\|x-y|=1}} \sigma_x\sigma_y + \beta \sum_{\substack{x\in\Lambda_L, y\notin\Lambda_L\\|x-y|=1}} \sigma_x\overline{\sigma}_y\bigr].$$

(As for the two previous models, the nearest-neighbor restriction can be replaced by a – possibly weighted – finite-range assumption.) The set of limiting measures, as $L \to \infty$ and for any boundary conditions, is a simplex, whose extreme elements are the Gibbs states of the model. We define the high-temperature region as $\beta < \beta_c^{\text{Ising}}$, where

$$\beta_c^{\text{Ising}} = \sup\{\beta : \text{There is a unique Gibbs state at parameter } \beta\} > 0\,.$$

We are interested in the following quantity:

$$G_\beta^{\text{Ising}}(x) \triangleq \mathbb{E}_{\mu_\beta}[\sigma_0\sigma_x]\,,$$

where the expectation is computed with respect to any translation invariant Gibbs state μ_β (it is independent of which one is chosen). It is a deep result of [3] that the high-temperature region can also be characterized as the set of all β such that

$$\sum_{x\in\mathbb{Z}^d} G_\beta^{\text{Ising}}(x) < \infty\,.$$

We now discuss simultaneously these three models; to that end, we simply forget the model-specific superscripts, and simply write β_c or G_β. It can be shown that for all three models, for all $\beta < \beta_c$, the function $G_\beta(x)$ is actually exponentially decreasing in $|x|$, i.e. the corresponding inverse correlation length $\xi_\beta : \mathbb{R}^d \to \mathbb{R}$ satisfy

$$\xi_\beta(x) \triangleq \lim_{k\to\infty} -\frac{1}{k}\,\log G_\beta(\lfloor kx\rfloor) > 0\,,$$

where $\lfloor x\rfloor$ is the componentwise integer part of x. Obviously, ξ_β is positive-homogeneous, and it is not difficult to prove that it is convex; it is thus an equivalent norm on $\mathbb{R}^d$ (for $\beta < \beta_c$).

The main result of [15, 6, 7] is the derivation of the following sharp asymptotics for $G_\beta(x)$, as $|x| \to \infty$, for these three models, in the corresponding high-temperature regions.

Theorem 1.1. *Consider one of the models above, and let $\beta < \beta_c$. Then, uniformly as $|x| \to \infty$,*

$$G_\beta(x) = \frac{\Psi_\beta(n_x)}{\sqrt{|x|^{d-1}}}\, e^{-\xi_\beta(n_x)\,|x|}\,(1 + o(1))\,,$$

where $n_x = x/|x|$, and Ψ_β is strictly positive and analytic. Moreover, ξ_β is also an analytic function.

As a by-product of the proof of Theorem 1.1, we obtain the following results on the shape of the equidecay profiles,

$$\mathbf{U}_\beta \stackrel{\triangle}{=} \{x \in \mathbb{R}^d \,:\, \xi_\beta(x) \leq 1\}$$

and their polar, the Wulff shapes

$$\mathbf{K}_\beta \stackrel{\triangle}{=} \bigcap_{n \in \mathbb{S}^{d-1}} \{t \in \mathbb{R}^d \,:\, (t,n)_d \leq \xi_\beta(n)\}\,.$$

Theorem 1.2. *Consider one of the models above, and let $\beta < \beta_c$. Then $\mathbf{K}_\beta$ has a locally analytic, strictly convex boundary. Moreover, the Gaussian curvature κ_β of $\mathbf{K}_\beta$ is uniformly positive,*

$$\overline{\kappa}_\beta \stackrel{\triangle}{=} \min_{t \in \partial \mathbf{K}_\beta} \kappa_\beta(t) > 0\,. \tag{1}$$

By duality, $\partial \mathbf{U}_\beta$ is also locally analytic and strictly convex.

Remark 1.3. *In two dimensions $\mathbf{K}_\beta$ is reminiscent of the Wulff shape (and is exactly the low-temperature Wulff shape in the cases of the nearest-neighbor Ising and percolation models). Equation* (1) *is then called the positive stiffness condition; it is known to be equivalent to the following sharp triangle inequality* [14, 20]*: Uniformly in $u, v \in \mathbb{R}^2$*

$$\xi_\beta(u) + \xi_\beta(v) - \xi_\beta(u+v) \geq \overline{\kappa}_\beta\,(|u| + |v| - |u+v|)\,.$$

Theorem 1.1 can in fact easily be extended to arbitrary odd-odd correlation functions. We show this here in the most difficult case of ferromagnetic Ising models; namely, we establish exact asymptotic formula for correlation functions of the form $\mathbb{E}_{\mu_\beta}[\sigma_A \sigma_{B+x}]$, where A, B are finite subsets of $\mathbb{Z}^d$ with $|A|$ and $|B|$ odd, and for any $C \subset \mathbb{Z}^d$, $\sigma_C \stackrel{\triangle}{=} \prod_{y \in C} \sigma_y$. Notice that even-odd correlations are necessarily zero by symmetry. The case of even-even correlations is substantially more delicate though (already for the much simpler SAW model), in particular in low dimensions; we hope to come back to this issue in the future.

Theorem 1.4. *Consider the Ising model. Let $\beta < \beta_c^{\mathrm{Ising}}$, and let A and B be finite odd subsets of $\mathbb{Z}^d$. Then, uniformly in $|x| \to \infty$,*

$$\mathbb{E}_{\mu_\beta}[\sigma_A \sigma_{B+x}] = \frac{\Psi_\beta^{A,B}(n_x)}{\sqrt{|x|^{d-1}}}\, e^{-\xi_\beta(n_x)\,|x|}\,(1 + o(1))\,,$$

where $n_x = x/|x|$, and $\Psi_\beta^{A,B}$ is strictly positive and analytic.

We sketch the proof of this theorem in Section 4.

The main feature shared by the three models discussed above is that the function $G_\beta(x)$ can each time be written in the form

$$G_\beta(x) = \sum_{\lambda:\, 0 \to x} q_\beta(\lambda)\,, \tag{2}$$

where the sum runs over admissible path-like objects (SAW paths, percolation clusters, random-lines, see Section 3, respectively). The weights $q_\beta(\,\cdot\,)$ are supposed to be strictly positive and to possess a variation of the following four properties:

- **Strict exponential decay of the two-point function:** There exists $C_1 < \infty$ such that, for all $x \in \mathbb{Z}^d \setminus \{0\}$,

$$g(x) \;=\; \sum_{\lambda:\, 0 \to x} q(\lambda) \le C_1\, e^{-\xi(x)}\,, \tag{3}$$

  where $\xi(x) = -\lim_{k\to\infty}(k)^{-1} \log g(\lfloor kx \rfloor)$ is the inverse correlation length.
- **Finite energy condition:** For any pair of compatible paths λ and η define the conditional weight

$$q(\lambda \,|\, \eta) = q(\lambda \amalg \eta)/q(\eta)$$

  where $\lambda \amalg \eta$ denotes the concatenation of λ and η. Then there exists a universal finite constant $C_2 < \infty$ such that the conditional weights are controlled in terms of path sizes $|\lambda|$ as:

$$q(\lambda \,|\, \eta) \;\ge\; e^{-C_2|\lambda|}\,. \tag{4}$$

- **BK-type splitting property:** There exists $C_3 < \infty$, such that, for all $x, y \in \mathbb{Z}^d \setminus \{0\}$ with $x \neq y$,

$$\sum_{\lambda:\, 0 \to x \to y} q(\lambda) \le C_2 \sum_{\lambda:\, 0 \to x} q(\lambda) \sum_{\lambda:\, x \to y} q(\lambda)\,. \tag{5}$$

- **Exponential mixing :** There exists $C_4 < \infty$ and $\theta \in (0,1)$ such that, for any four paths λ, η, γ_1 and γ_2, with $\lambda \amalg \eta \amalg \gamma_1$ and $\lambda \amalg \eta \amalg \gamma_2$ both admissible,

$$\frac{q(\lambda \,|\, \eta \amalg \gamma_1)}{q(\lambda \,|\, \eta \amalg \gamma_2)} \le \exp\{C_4 \sum_{\substack{x\in\lambda \\ y \in \gamma_1 \cup \gamma_2}} \theta^{|x-y|}\}\,. \tag{6}$$

Many other models enjoy a graphical representation of correlation functions of the form (2). In perturbative regimes, cluster expansions provide a generic example. Non-perturbative examples include the random-cluster representation for Potts (and other) models [10], or random walk representation of N-vector models [11], etc... However, it might not always be easy, or even possible, to establish properties (3), (4), (5) and (6) for the corresponding weights, especially (5) which is probably the less robust one. It should however be possible to weaken the latter so that it only relies on some form of locally uniform mixing properties.

Road-map to the paper

In Section 2 we review and explain our probabilistic approach to the analysis of high temperature correlation functions. The point of departure is the random path representation formula (2), and the whole theory is built upon a study of the local fluctuation structure of the corresponding connection paths. One of the consequences is the validity of the invariance principle under the diffusive scaling, which we formulate in Theorem 2.2 below. For simplicity the discussion in Section 2 is restricted to the case of SAW-s, and hence the underlying local limit results are those about the sums of independent random variables. In the case of high temperature ferromagnetic Ising models the random line representation, which we shall briefly recall in Section 3, gives rise to path weights q_β which do not possess appropriate factorization properties. Nevertheless these weights satisfy conditions (3)-(6) and we conclude Section 3 with an explanation of how the problem of finding correlation asymptotics can be reformulated in terms of local limit properties of one dimensional systems generated by Ruelle operators for full shifts on countable alphabets. The proof of Theorem 1.4 is discussed in Section 4. Finally, in Section 5, we explain the relation between the problems discussed here, inspired by Statistical Physics, and their counterparts originating from the corresponding lattice Quantum Field Theories.

§2. Fluctuations of connection paths

In this section we describe local structure and large scale properties of connection paths conditioned to hit a distant point. In all three models above (SAW, percolation, Ising) the distribution of the connection paths converges, after the appropriate rescaling, to the $(d-1)$-dimensional Brownian bridge, and, from the probabilistic point of view, these results belong to the realm of classical Gaussian local limit analysis of one dimensional systems based on uniform analytic expansions of finite volume log-moment generating functions. An invariance principle for the sub-critical Bernoulli bond percolation has been established in [17] and for the phase separation line in the 2D nearest neighbour Ising model at any $\beta > \beta_c$ in [13]. In both cases the techniques and the ideas of [6] and [7] play the crucial role, and, in fact, the renormalization and the fluctuation analysis developed in the latter papers pertains to a large class of models which admit a random path type representation with path weights enjoying a suitable variation of (3)-(6). In particular, it should lead to a closed form theory of low temperature phase boundaries in two dimensions [16]. Note that different tools have been early employed in [9, 12].

For the sake of simplicity we shall sketch here the case of self-avoiding walks and shall try to stipulate the impact of the geometry of $\mathbf{K}_\beta$ on the magnitude of paths fluctuations in the corresponding directions.

Let $\hat{x} \in \mathbb{S}^{d-1}$ and the dual point $\hat{t} \in \partial\mathbf{K}_\beta$; $(\hat{t}, \hat{x}) = \xi_\beta(\hat{x})$, be fixed for the rest of the section. Consider the set $\mathcal{P}^n$ of all self-avoiding paths $\gamma : 0 \to \lfloor n\hat{x} \rfloor$, where for $y \in \mathbb{R}^d$ we define $\lfloor y \rfloor = (\lfloor y_1 \rfloor, \dots, \lfloor y_d \rfloor) \in \mathbb{Z}^d$. Finally, consider the following probability measure $\mathbb{P}^n_\beta$ on $\mathcal{P}^n$:

$$(7) \qquad \mathbb{P}^n_\beta(\gamma) = \frac{1}{\mathbf{Z}^n_\beta} \mathrm{e}^{\beta|\gamma|} 1_{\{\gamma \in \mathcal{P}^n\}}.$$

In order to explain and to formulate the invariance principle which holds under $\mathbb{P}^n_\beta$ we need, first of all, to readjust the notion of irreducible splitting of paths $\gamma \in \mathcal{P}^n$;

$$(8) \qquad \gamma = \lambda_L \amalg \lambda_1 \amalg \cdots \amalg \lambda_M \amalg \lambda_R.$$

Fix $\delta \in (0, 1)$ and a large enough renormalization scale K. Given a path $\lambda = (u_0, u_1, \dots, u_m)$ let us say that a point u_l; $0 < l < m$, is $\hat{x}$-correct break point of λ if the following two conditions hold:

A) $(u_j, \hat{x}) < (u_l, \hat{x}) < (u_i, \hat{x})$ for all $j < l < i$.

B) The remaining sub-path $(u_{l+1}, \dots, u_m)$ lies inside the set

$$2K\mathbf{U}_\beta(u_l) + \mathcal{C}_\delta(\hat{t}),$$

where $\mathbf{U}_\beta(z) = z + \mathbf{U}_\beta$, and the forward cone $\mathcal{C}_\delta(\hat{t})$ is defined as

$$(9) \qquad \mathcal{C}_\delta(\hat{t}) = \left\{ y \in \mathbb{R}^d \, : \, (y, \hat{t}) > (1 - \delta)\xi_\beta(y) \right\}.$$

Note that this definition depends on the parameters K and δ; as they are usually kept constant, we only write them explicitly when needed.

With $\hat{x} \in \mathbb{S}^{d-1}$, $\hat{t} \in \partial\mathbf{K}_\beta$, K and δ fixed as above let us say that a path λ is irreducible if it does not contain $\hat{x}$-correct break points. We use $\mathcal{S}$ to denote the set of all irreducible paths (modulo $\mathbb{Z}^d$-shifts). Define also the following three subsets of $\mathcal{S}$:

$$\mathcal{S}_L = \{\lambda = (u_0, \dots, u_m) \in \mathcal{S} \,:\, \forall l > 0\, (u_l, \hat{x}) < (u_m, \hat{x})\}$$

$$\mathcal{S}_R = \left\{ \begin{array}{r} \lambda = (u_0, \dots, u_m) \in \mathcal{S} \,:\, \forall l > 0\, (u_l, \hat{x}) > (u_0, \hat{x}) \text{ and} \\ \gamma \subset K\mathbf{U}_\beta(u_0) + \mathcal{C}_\delta(\hat{t}). \end{array} \right\} \tag{10}$$

$$\mathcal{S}_0 = \mathcal{S}_L \cap \mathcal{S}_R$$

For any $\gamma \in \mathcal{P}_n$ which has at least two $\hat{x}$-correct break points the decomposition (8) is unambiguously defined by the following set of conditions:

$$\lambda_L \in \mathcal{S}_L,\ \lambda_R \in \mathcal{S}_R \text{ and } \lambda_1, \dots, \lambda_M \in \mathcal{S}_0.$$

The only difference between (8) and the irreducible decomposition employed in [7] is that the break points here are defined with respect to the $\hat{x}$-orthogonal hyper-planes instead of $\hat{t}$-orthogonal hyper-planes. This is to ensure that the displacements along all the λ-paths which appear in (8) have positive projection on the direction of $\hat{x}$. More precisely, given a SAW path $\lambda = (u_0, \dots, u_m)$ let us define the displacement along λ as $V(\lambda) = u_m - u_0$. By the very definition of (8) all

$$V_L \triangleq V(\lambda_L),\ V_1 \triangleq V(\lambda_1), \dots,\ V_M \triangleq V(\lambda_M),\ V_R \triangleq V(\lambda_R).$$

belong to the (lattice) half-space $\{y \in \mathbb{Z}^d \,:\, (y, \hat{x}) > 0\}$. The renormalization calculus developed in [6, 7] implies:

Lemma 2.1. *For every $\beta < \beta_c$ and for any $\delta > 0$ there exists a finite scale $K_0 = K_0(\delta, \beta)$ and a number $\nu = \nu(\delta, \beta) > 0$, such that*

$$\sum_{\lambda \in \mathcal{S} \,:\, V(\lambda) = y} \mathrm{e}^{\beta|\lambda|} \leq \exp\left\{-(\hat{t}, y) - \nu|y|\right\}, \tag{11}$$

uniformly in $y \in \mathbb{Z}^d$.

Going back to the decomposition (8) notice that

$$V_L + V_1 + \cdots + V_M + V_R = \lfloor n\hat{x} \rfloor \tag{12}$$

for any $\gamma : 0 \to \lfloor n\hat{x} \rfloor$. Therefore, Lemma 2.1 and the Ornstein-Zernike formula of Theorem 1.1 yield:

$$\mathbb{P}_\beta^n \left(\max\{|V_L|, |V_1|, \dots, |V_M|, |V_R|\} > (\log n)^2 \right) = o\left(\frac{1}{n^\rho} \right), \tag{13}$$

for any $\rho > 0$. In particular, if for given $\gamma : 0 \to \lfloor n\hat{x} \rfloor$ one considers the piece-wise constant trajectory $\widehat{\gamma}$ through the vertices $0, V_L, V_L + V_1, \dots, \lfloor n\hat{x} \rfloor$, then the $\mathbb{R}^d$-Hausdorff distance between γ and $\widehat{\gamma}$ is bounded above as:

$$\mathbb{P}_\beta^n \left(\mathrm{d_H}(\gamma, \widehat{\gamma}) > (\log n)^2 \right) = o\left(\frac{1}{n^\rho} \right), \tag{14}$$

as well. Indeed, one needs only to control the fluctuation of λ_L in (8), the traversal deviations of paths in $\mathcal{S}_R$ are automatically under control by the cone confinement property (10).

Estimate (14) enables a formulation of the invariance principle for SAW γ in terms of the effective path $\widehat{\gamma}$. In its turn the invariance principle for $\widehat{\gamma}$ is a version of the conditional invariance principle for paths of random walks in $(d-1)$-dimensions with the direction of the target point $\hat{x}$ playing the role of time. It happens to be natural to choose the frame of the remaining $(d-1)$ spatial dimensions according to principal directions of curvature $\mathfrak{v}_1, \dots, \mathfrak{v}_{d-1}$ of $\partial\mathbf{K}_\beta$ at $\hat{t}$. In this way, in view of the positive $\hat{x}$-projection property of all the λ-path displacements in (8), the effective path $\widehat{\gamma} \subset \mathbb{R}^d$ could be parametrized in the orthogonal frame $(\hat{x}, \mathfrak{v}_1, \dots, \mathfrak{v}_{d-1})$ as a function $\widehat{X} : [0, n] \to \mathbb{R}^{d-1}$. As usual define the diffusive scaling $\widehat{X}_n(\cdot)$ of $\widehat{X}(\cdot)$ as

$$\widehat{X}_n(\tau) = \frac{1}{\sqrt{n}} \widehat{X}(\lfloor n\tau \rfloor).$$

Let $C_{0,0}[0,1]$ be the space of continuous $\mathbb{R}^{d-1}$-valued functions f on $[0,1]$ which satisfy the boundary conditions $f(0) = f(1) = 0$.

Theorem 2.2. *The distribution of $\widehat{X}_n(\cdot)$ under $\mathbb{P}_\beta^n$ weakly converges on $C_{0,0}[0,1]$ to the distribution of*

$$\left(\sqrt{\kappa_1} B_1(\cdot), \dots, \sqrt{\kappa_{d-1}} B_{d-1}(\cdot) \right), \tag{15}$$

where $B_1(\cdot), \dots, B_{d-1}(\cdot)$ are independent Brownian bridges on $[0,1]$ and $\kappa_1, \dots, \kappa_{d-1}$ are the principal curvatures of $\partial\mathbf{K}_\beta$ at $\hat{t}$.

Let us dwell on the probabilistic picture behind Theorems 1.1, 1.2 and 2.2: First of all, note that by Lemma 2.1

$$(16) \qquad \mathbb{Q}_0(y) = \mathrm{e}^{(\hat{t},y)} \sum_{\lambda \in \mathcal{S}_0 \,:\, V(\lambda)=y} \mathrm{e}^{\beta|\lambda|} \triangleq \mathrm{e}^{(\hat{t},y)} W_0(y)$$

is a (non-lattice) probability distribution on $\mathbb{Z}^d$ with exponentially decaying tails. Indeed, an alternative important way to think about $\mathbf{K}_\beta$ is as of the closure of the domain of convergence of the series

$$(17) \qquad t \in \mathbb{R}^d \mapsto \sum_{y \in \mathbb{Z}^d} \mathrm{e}^{(t,y)} G_\beta(y).$$

On the other hand, Lemma 2.1 ensures that the series $\mathbb{W}_0(t) \triangleq \sum \mathrm{e}^{(t,y)} W_0(y)$ converges in the ν-neighbourhood $B_\nu(\hat{t}) = \{t \,:\, |t-\hat{t}| < \nu\}$ of $\hat{t}$. In view of the decomposition (8) and Lemma 2.1,

$$(18) \qquad G_\beta(n\hat{x}) = O\left(\mathrm{e}^{-n\xi_\beta(\hat{x})-\nu n}\right) + \sum_{M=1}^{\infty} W_L * W_0^{*M} * W_R(n\hat{x}),$$

where we have assumed for the convenience of notation that $n\hat{x} \in \mathbb{Z}^d$, and

$$W_L(y) = \sum_{\lambda \in \mathcal{S}_L \,:\, V(\lambda)=y} \mathrm{e}^{\beta|\lambda|} \quad \text{and} \quad W_R(y) = \sum_{\lambda \in \mathcal{S}_R \,:\, V(\lambda)=y} \mathrm{e}^{\beta|\lambda|}.$$

As a result, the piece of the boundary $\partial\mathbf{K}_\beta$ inside $B_\nu(\hat{t})$ is implicitly given by

$$\partial\mathbf{K}_\beta \cap B_\nu(\hat{t}) = \{t \in B_\nu(\hat{t}) \,:\, \mathbb{W}_0(t) = 1\}.$$

In order to obtain the full claim of Theorem 1.2 one needs only to check the non-degeneracy of $\mathrm{Hess}(\mathbb{W}_0)$ at $\hat{t}$, which, in the case of SAW-s, is a direct consequence of the finite energy condition (4). Note, by the way, that since $\hat{x}$ is the normal direction to $\partial\mathbf{K}_\beta$ at $\hat{t}$, there exists a number $\alpha \in (0,\infty)$, such that

$$(19) \qquad \nabla\mathbb{W}_0(\hat{t}) = \alpha\hat{x}.$$

Multiplying both sides of (18) by $e^{n\xi_\beta(\hat{x})} = e^{(\hat{t}, n\hat{x})}$ we arrive to the following key representation of the two point function G_β:

(20)

$$e^{n\xi_\beta(\hat{x})}G_\beta(n\hat{x}) = e^{n(\hat{t},\hat{x})}G_\beta(n\hat{x}) = O\left(e^{-n\nu}\right)$$
$$+ \sum_{v_L, v_R \in \mathbb{Z}^d} \mathbb{Q}_L(v_L)\mathbb{Q}_R(v_R) \sum_{M=1}^{\infty} \mathbb{Q}_0(V_1 + \cdots + V_M = n\hat{x} - v_L - v_R),$$

where, similar to (16), we have defined $\mathbb{Q}_L(v) = e^{(\hat{t},v)}W_L(v)$ and, accordingly, $\mathbb{Q}_R(v) = e^{(\hat{t},v)}W_R(v)$.

Unlike $\mathbb{Q}_0$ the measures $\mathbb{Q}_L$ and $\mathbb{Q}_R$ are in general not probability but, by Lemma 2.1, they are finite and have exponentially decaying tails:

$$\sum_{|y|>n} \left(\mathbb{Q}_L(y) + \mathbb{Q}_R(y)\right) \leq e^{-\nu n/2}.$$

Since by (19) the expectation of V_l under $\mathbb{Q}_0$ equals to $\alpha\hat{x}$, the usual local limit CLT for $\mathbb{Z}^d$ random variables and the Gaussian summation formula imply that the right hand side in (20) equals to $c_1/\sqrt{n^{d-1}}$. Actually, a slightly more careful analysis along these line leads to the full analytic form of the Ornstein-Zernike formula as claimed in Theorem 1.1.

Let us explain now how the principal curvatures $\kappa_1, \ldots, \kappa_{d-1}$ of $\partial\mathbf{K}_\beta$ at $\hat{t}$ enter the picture: By the irreducible path representation and arguments completely similar to those just reproduced above, the total weight of all piece-wise constant paths $\widehat{\gamma} = \widehat{\gamma}(V_L, V_1, \ldots, V_M, V_R)$; $M = 1, 2, \ldots$, which pass through a point $v_n \in \mathbb{Z}^d$;

$$v_n = \lambda n\hat{x} + \sqrt{n}\sum_{l=1}^{d-1} a_l \mathfrak{v}_l \stackrel{\Delta}{=} \lambda n\hat{x} + \sqrt{n}\mathfrak{v},$$

equals to

$$\frac{c_2}{\sqrt{(\lambda(1-\lambda)n^2)^{(d-1)}}} e^{-\xi_\beta(v_n) - \xi_\beta(n\hat{x} - v_n)} \left(1 + o(1)\right),$$

where $c_2 > 0$ does not depend on $\lambda \in (0,1)$ and the coefficients $a_1, \ldots, a_{d-1}$. Comparing with the OZ formula for the full partition function G_β, we infer that

$$\mathbb{P}^n_\beta\left(v_n \in \widehat{\gamma}\right) = \frac{c_3 \exp\left\{-\left(\xi_\beta(v_n) + \xi_\beta(n\hat{x} - v_n) - \xi_\beta(n\hat{x})\right)\right\}}{\sqrt{(\lambda(1-\lambda)n)^{d-1}}} \left(1 + o(1)\right).$$

From now on we refer to Chapter 2.5 in [22] for the missing details in the arguments below. ξ_β is the support function of $\mathbf{K}_\beta$ and by Theorem 1.2 it is a smooth function. Thus, for every $v \in \mathbb{R}^d$ the gradient $\nabla\xi_\beta(v) \in \partial\mathbf{K}_\beta$ and $\xi_\beta(v) = (\nabla\xi_\beta(v), v)$ (in particular $\hat{t} = \nabla\xi_\beta(\hat{x}) = \nabla\xi_\beta(n\hat{x})$). Principal radii of the curvature $1/\kappa_1, \ldots, 1/\kappa_{d-1}$ of $\partial\mathbf{K}_\beta$ at $\hat{t}$ are the eigenvalues of the linear map

$$\mathrm{d}^2\xi_\beta|_{\hat{x}} \, : \, T_{\hat{x}}\mathbb{S}^{d-1} \mapsto T_{\hat{x}}\mathbb{S}^{d-1},$$

and $\mathfrak{v}_1, \ldots, \mathfrak{v}_{d-1} \in T_{\hat{x}}\mathbb{S}^{d-1}$ are the corresponding eigenvectors. Therefore,

$$\begin{aligned}
\xi_\beta(v_n) + \xi_\beta(n\hat{x} - v_n) - \xi_\beta(n\hat{x}) &= n\lambda\left(\xi_\beta\left(\hat{x} + \frac{1}{\lambda\sqrt{n}}\mathfrak{v}\right) - \xi_\beta(\hat{x})\right) \\
&+ n(1-\lambda)\left(\xi_\beta\left(\hat{x} - \frac{1}{(1-\lambda)\sqrt{n}}\mathfrak{v}\right) - \xi_\beta(\hat{x})\right) \\
&= \frac{1}{2\lambda}\left(\mathrm{d}^2\xi_\beta|_{\hat{x}}\mathfrak{v}, \mathfrak{v}\right) + \frac{1}{2(1-\lambda)}\left(\mathrm{d}^2\xi_\beta|_{\hat{x}}\mathfrak{v}, \mathfrak{v}\right) + O\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{2\lambda(1-\lambda)}\sum_{l=1}^{d-1}\frac{a_l^2}{\kappa_l} + O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Computations for higher order finite dimensional distributions follow a completely similar pattern.

§3. Random-line representation of Ising correlations

Correlation functions of ferromagnetic Ising models admit a very useful representation in terms of sums over weighted random paths, which is especially convenient for our purposes here. The two-point function formula (2) is a particular case. In this section, we recall how this representation is derived; we refer to [20] for details and additional results. In the end of the section we shall briefly indicate how (20) and, accordingly, the whole local limit analysis should be re-adjusted in order to incorporate the (dependent) case of Ising paths.

Although we use it for the infinite-volume Gibbs measure, it is convenient to derive the random path representation first for finite volumes, and then take the limit. As there is a single Gibbs state for the values of β we consider, it suffices to consider free boundary conditions (i.e. no interactions between spins inside the box and spins outside).

Given a set of edges B of the lattice $\mathbb{Z}^d$, we define the associated set of vertices as $V_B \overset{\triangle}{=} \{x \in \mathbb{Z}^d : \exists e \in B \text{ with } x \in e\}$ ($x \in e$ means that x is an endpoint of e). For any vertex $x \in V_B$, we define the *index* of x in B by $\operatorname{ind}(x, B) \overset{\triangle}{=} \sum_{e \in B} \mathbf{1}_{\{e \ni x\}}$. The *boundary* of B is defined by $\partial B \overset{\triangle}{=} \{x \in V_B : \operatorname{ind}(x, B) \text{ is odd}\}$.

In this context, the finite volume Gibbs measure is defined by

$$\mu_{B,\beta}(\sigma) \overset{\triangle}{=} Z_\beta(B)^{-1} \exp[-\beta \sum_{e=(x,y)\in B} \sigma_x \sigma_y],$$

and we use the standard notation $\langle \cdot \rangle_{B,\beta}$ to denote expectation w.r.t. this probability measure.

We fix an arbitrary total ordering of $\mathbb{Z}^d$. At each $x \in \mathbb{Z}^d$, we fix (in an arbitrary way) an ordering of the x-incident edges of the graph:

$$B(x) \overset{\triangle}{=} \{e \in B : \operatorname{ind}(x, \{e\}) > 0\} = \{e_1^x, \dots, e_{\operatorname{ind}(x,B)}^x\},$$

and for two incident edges $e = e_i \in B(x)$, $e' = e_j \in B(x)$ we say that $e \leq e'$ if the corresponding inequality holds for their sub-indices; $i \leq j$.

Let $A \subset V_B$ be such that $|A|$ is even; we write $\sigma_A \overset{\triangle}{=} \prod_{i \in A} \sigma_i$. Using the identity $e^{\beta \sigma_x \sigma_y} = \cosh(\beta)\big(1 + \sigma_x \sigma_y \tanh(\beta)\big)$, we obtain the following expression for the correlation function $\langle \sigma_A \rangle_{B,\beta}$,

$$\langle \sigma_A \rangle_{B,\beta} = Z_\beta(B)^{-1} \sum_{\substack{D \subset B \\ \partial D = A}} \prod_{e \in D} \tanh \beta,$$

where

$$Z_\beta(B) \overset{\triangle}{=} \sum_{\substack{D \subset B \\ \partial D = \emptyset}} \prod_{e \in D} \tanh \beta.$$

From $D \subset B$ with $\partial D = A$, we would like to extract a family of $|A|/2$ "self-avoiding paths" connecting pairs of sites of A. We apply the following algorithm:

STEP 0 Set $k = 1$ and $\Delta = \emptyset$.

STEP 1 Set $z_0^{(k)}$ to be the first site of A in the ordering of $\mathbb{Z}^d$ fixed above, $j = 0$, and update $A \overset{\triangle}{=} A \setminus \{z_0^{(k)}\}$.

STEP 2 Let $e_j^{(k)} = (z_j^{(k)}, z_{j+1}^{(k)})$ be the first edge in $B(z_j^{(k)}) \setminus \Delta$ (in the ordering of $B(z_j^{(k)})$ fixed above) such that $e_j^{(k)} \in D$. This defines $z_{j+1}^{(k)}$.

STEP 3 Update $\Delta \overset{\triangle}{=} \Delta \cup \{e \in B(z_j^{(k)}) : e \leq e_j^{(k)}\}$. If $z_{j+1}^{(k)} \in A$, then go to STEP 4. Otherwise update $j \overset{\triangle}{=} j+1$ and return to STEP 2.
STEP 4 Set $n^{(k)} = j+1$ and stop the construction of this path. Update $A \overset{\triangle}{=} A \setminus \{z_{j+1}^{(k)}\}$, $k \overset{\triangle}{=} k+1$ and go to STEP 1.

This procedure produces a sequence $(z_0^{(1)}, \dots, z_{n^{(1)}}^{(1)}, z_0^{(2)}, \dots, z_{n^{(|A|/2)}}^{(|A|/2)})$. Let $\bar{i} \overset{\triangle}{=} |A|/2 + 1 - i$, and set $w_k^{(i)} \overset{\triangle}{=} z_{n-k}^{(\bar{i})}$.

We, thus, constructed $|A|/2$ paths, γ_i, $i = 1, \dots, |A|/2$, given by[1]

$$\gamma_i \overset{\triangle}{=} \gamma_i(D) \overset{\triangle}{=} (w_0^{(i)}, \dots, w_{n^{(i)}}^{(i)})$$

connecting distinct pairs of points of A, and such that

- $(w_k^{(i)}, w_{k+1}^{(i)}) \in B$, $k = 0, \dots, n^{(i)} - 1$, $i = 1, \dots, |A|/2$
- $(z_k^{(i)}, z_{k+1}^{(i)}) \neq (z_l^{(j)}, z_{l+1}^{(j)})$ if $i \neq j$, or if $i = j$ but $k \neq l$.

(but $z_k^{(i)} = z_l^{(j)}$ is allowed). A family of contours $\underline{\gamma} = (\gamma_1, \dots, \gamma_{|A|/2})$ is (A, B)-admissible if it can be obtained from a set $D \subset B$ with $\partial D = A$, using this algorithm; in that case we write $\underline{\gamma} \sim (A, B)$. Notice that here the order of the paths is important: if γ_k is a path from x_k to y_k then we must have $y_1 > y_2 > \dots > y_{|A|/2}$. This is to ensure that we do not count twice the same configuration of paths.

The construction also yields a set of edges $\Delta(\underline{\gamma}) \overset{\triangle}{=} \Delta$. Observe that $\Delta(\underline{\gamma})$ is entirely determined by $\underline{\gamma}$ (and the order chosen for the sites and edges). In particular the sets $D \subset B$ giving rise to an (A, B)-admissible family $\underline{\gamma}$ are characterized by $\partial D = A$ and

$$D \cap \Delta(\underline{\gamma}) = \bigcup_{i=1}^{|A|/2} \gamma_i \,.$$

Therefore, for such sets, $\partial(D \setminus \Delta(\underline{\gamma})) = \emptyset$, and we can write

$$\langle \sigma_A \rangle_{\beta,B} = \sum_{\underline{\gamma} \sim (A,B)} q_{\beta,B}(\underline{\gamma}) \,,$$

where

$$q_{\beta,B}(\underline{\gamma}) = w(\underline{\gamma}) \, \frac{Z_\beta(B \setminus \Delta(\underline{\gamma}))}{Z_\beta(B)} \,,$$

---

[1] This backward construction of the lines turns out to be convenient for the reformulation in terms of Ruelle's formalism, see [7].

with

$$w(\underline{\gamma}) = \prod_{i=1}^{|A|/2} \prod_{k=1}^{n^{(i)}} \tanh \beta .$$

This is an instance of the *random-line representation* for correlation functions of the Ising model in B. It has been studied in detail in [19, 20] and is essentially equivalent (though the derivations are quite different) to the random-walk representation of [1]. We'll need a version of this representation when B is replaced by the set $\mathcal{E}(\mathbb{Z}^d)$ of all edges of $\mathbb{Z}^d$. To this end, we use the following result ([20], Lemmas 6.3 and 6.9): For all $\beta < \beta_{\mathrm{c}}$,

$$\langle \sigma_A \rangle_\beta = \sum_{\underline{\gamma} \sim A} q_\beta(\underline{\gamma}) , \tag{21}$$

where $q_\beta(\underline{\gamma}) \overset{\triangle}{=} \lim_{B_n \nearrow \mathcal{E}(\mathbb{Z}^d)} q_{\beta, B_n}(\underline{\gamma})$ is well defined.

It will also be useful to work with a more relaxed definition of admissibility, since we want to cut our paths into pieces, and the order of the resulting pieces might not correspond with the order of their endpoints. In general, given a path $\gamma = (x_1, x_2, \dots , x_n)$, we define $\Delta(\gamma) = \bigcup_{k=1}^{n} \{ e \in B(x_k) \, : \, e \leq (x_{k-1}, x_k) \}$. We say that a path $\gamma = (x_1, \dots , x_n)$ is admissible if $\{(x_1, x_2), \dots , (x_{k-1} x_k)\} \cap \Delta((x_k, \dots , x_n)) = \emptyset$ for all $2 \leq k \leq n-1$. Given a family of paths $\underline{\gamma} = (\gamma_1, \dots , \gamma_n)$, we define $\Delta(\underline{\gamma}) = \bigcup_{k=1}^{n} \Delta(\gamma_k)$. A family of admissible paths $\underline{\gamma}$ is then admissible if $(\gamma_1, \dots , \gamma_k) \cap \Delta((\gamma_{k+1}, \dots , \gamma_n)) = \emptyset$ for all $1 \leq k \leq n-1$. Notice that the order of the paths is still important ((γ_1, γ_2) can be admissible while (γ_2, γ_1) is not), but there are no constraint on the order of their endpoints. Indeed, they can even share endpoints. Observe that these definitions are identical to those above when restricted to the same setting.

We then have the following crucial inequality: Let $\underline{\gamma}$ be an admissible family of paths. Then

$$\sum_{\substack{\gamma_0 : x \to y \\ \gamma_0 \cap \Delta(\underline{\gamma}) = \emptyset}} q_\beta(\gamma_0, \underline{\gamma}) \leq q_\beta(\underline{\gamma}) \sum_{\gamma_0 : x \to y} q_\beta(\gamma_0) . \tag{22}$$

We give a brief proof. It is enough to consider the analogous statement in finite volumes B. Since $\Delta(\gamma_0, \underline{\gamma}) = \Delta(\gamma_0) \cup \Delta(\underline{\gamma})$ and $\gamma_0 \cap \Delta(\underline{\gamma}) = \emptyset$, we have

$$q_{\beta,B}(\gamma_0, \underline{\gamma}) = q_{\beta, B \setminus \Delta(\underline{\gamma})}(\gamma_0) \, q_{\beta,B}(\underline{\gamma}) \, .$$

Hence (22) follows simply from Griffiths' second inequality since

$$\sum_{\gamma_0 \subset B\setminus\Delta(\underline{\gamma})} q_{\beta,B\setminus\Delta(\underline{\gamma})}(\gamma_0) = \langle\sigma_x\sigma_y\rangle_{\beta,B\setminus\Delta(\underline{\gamma})} \leq \langle\sigma_x\sigma_y\rangle_\beta \,.$$

If the set A in (21) contains only two points, $A = \{x, y\}$, then we recover (2). The main difference between the SAW case considered in Section 2 and the case of sub-critical ferromagnetic Ising models is that the path weights q_β in (2) do not factorize: In general,

$$q_\beta(\gamma \amalg \lambda) \neq q_\beta(\gamma)q_\beta(\lambda).$$

Consequently, the displacement variables $V_1, V_2, \dots$ fail to be independent and the underlying local limit analysis should be generalized. The appropriate framework is that of the statistical mechanics of one dimensional systems generated by Ruelle operators for full shifts on countable alphabets. We refer to [7] for all the background material and here only sketch how the construction leads to the claims of Theorems 1.1, 1.2 and, after an appropriate re-definition of the measures $\mathbb{P}^n_\beta$, to the invariance principle stated in Theorem 2.2: As in the case of SAW-s fix a direction $\hat{x} \in \mathbb{S}^{d-1}$. The key renormalization result (Theorem 2.3 in [7]) which implies that the rate of decay of the irreducible connections is strictly larger than the rate of decay of the two point function G_β. In view of (8) this validates a representation of G_β as a sum of dependent random variables $V_L + V_1 + \dots + V_M + V_R$ with exponentially decaying tails. Namely, as in Section 2 let $\mathcal{S} = \mathcal{S}(K)$ be the set of all $\hat{x}$-irreducible paths. Then the following Ising analog of Lemma 2.1 holds:

Lemma 3.1. *For every $\beta < \beta_c$ and for any $\delta > 0$ there exists a finite scale $K_0 = K_0(\delta, \beta)$ and a number $\nu = \nu(\delta, \beta) > 0$, such that*

$$\sum_{\lambda\in\mathcal{S}\,:\,V(\lambda)=y} q_\beta(\lambda) \leq \exp\left\{-(\hat{t}, y) - \nu|y|\right\}, \tag{23}$$

uniformly in $y \in \mathbb{Z}^d$.

The above Lemma suggests that the main contribution to the sharp asymptotics of G_β comes from the weights of the paths $\lambda_1, \dots, \lambda_M$ in the decomposition (8). Accordingly, consider now the set $\mathcal{S}_0$ of cylindrical $\hat{x}$-irreducible paths which was introduced in Section 2. Given a finite collection $\lambda, \lambda_1, \dots, \lambda_M \in \mathcal{S}_0$ define the conditional weight

$$q_\beta(\lambda \,|\, \underline{\lambda}) = q_\beta(\lambda \,|\, \lambda_1 \amalg \cdots \amalg \lambda_M) = \frac{q_\beta(\lambda \amalg \lambda_1 \amalg \cdots \amalg \lambda_M)}{q_\beta(\lambda_1 \amalg \cdots \amalg \lambda_M)}.$$

By the crucial exponential mixing property (6) one is able to control the dependence of the conditional weights $q_\beta(\lambda \,|\, \underline{\lambda})$ on λ_M as follows:

$$\sup_{\lambda,\lambda_1,\dots,\lambda_{M-1}\in\mathcal{S}_0}\ \sup_{\lambda_M,\tilde{\lambda}_M\in\mathcal{S}_0} \frac{q_\beta(\lambda \,|\, \lambda_1 \amalg \cdots \amalg \lambda_M)}{q_\beta(\lambda \,|\, \lambda_1 \amalg \cdots \amalg \tilde{\lambda}_M)} \leq \mathrm{e}^{c_1\theta^M}. \tag{24}$$

In our formalism the set $\mathcal{S}_0$ plays the role of a countable alphabet. The estimate (24) enables the extension of the conditional weights $q_\beta(\lambda \,|\, \underline{\lambda})$ to the case of infinite strings $\underline{\lambda} = (\lambda_1, \lambda_2, \dots)$. Let $\mathfrak{S}_{0,\theta}$ be the set of all such strings endowed with the metrics

$$\mathrm{d}_\theta(\underline{\lambda}, \tilde{\underline{\lambda}}) = \theta^{\inf\{k : \lambda_k \neq \tilde{\lambda}_k\}}.$$

and let $\mathfrak{F}_{0,\theta}$ be the set of all bounded Lipschitz continuous functions on $\mathfrak{S}_{0,\theta}$.

As before we choose $\hat{t} \in \partial\mathbf{K}_\beta$ to be the dual direction to $\hat{x}$. Given a path $\lambda \in \mathcal{S}_0$ and a string $\underline{\lambda} \in \mathfrak{S}_{0,\theta}$ define the potential

$$\psi_\beta(\lambda \,|\, \underline{\lambda}) = \log q_\beta(\lambda \,|\, \underline{\lambda}) + \left(\hat{t}, V(\lambda)\right).$$

By (6) and Lemma 3.1 the operator

$$\mathcal{L}_z f(\underline{\lambda}) = \sum_{\lambda\in\mathcal{S}_0} \mathrm{e}^{\psi_\beta(\lambda \,|\, \underline{\lambda}) + (z, V(\lambda))} f(\lambda \amalg \underline{\lambda}), \tag{25}$$

is well defined and bounded on $\mathfrak{F}_{0,\theta}$ for every $z \in \mathbb{C}^d$ with $|z| < \nu$.

The dependent Ising analog of (20) is then given (see Section 3 of [7]) by

$$\begin{aligned} \mathrm{e}^{n\xi_\beta(\hat{x})} G_\beta(n\hat{x}) &= O\left(\mathrm{e}^{-n\nu}\right) \\ &+ \sum_{\mu\in\mathcal{S}_L} \sum_{\eta\in\mathcal{S}_R} q_\beta(\mu) q_\beta(\eta) \sum_{M=1}^{\infty} \mathbb{Q}_{0,M}^{\mu,\eta}(n\hat{x} - v_L - v_R). \end{aligned} \tag{26}$$

For each $M = 1, 2, \dots$ the family of weights $\{\mathbb{Q}_{0,M}^{\mu,\eta}\}$ is related to the family of operators $\{\mathcal{L}_z\}$ via the Fourier transform:

$$\sum_{y\in\mathbb{Z}^d} \mathrm{e}^{(z,y)} \mathbb{Q}_{0,M}^{\mu,\eta}(y) = \mathcal{L}_z^M w_{\mu,\eta}, \tag{27}$$

where the family $\{w_{\mu,\eta}\}$ is uniformly positive and uniformly bounded in $\mathfrak{F}_{0,\theta}$. In this way the analytic perturbation theory of the leading (that is lying on the spectral circle) eigenvalue of $\mathcal{L}_z$ enables the expansion of the logarithm of the right hand side in (26) which, in its turn, leads to classical Gaussian local limit results for the dependent sums $V_1 + \cdots + V_M$.

§4. Asymptotics of odd-odd correlations

In this section, we sketch the proof of Theorem 1.4. We do not give a complete, self-contained argument, since this would be too long, and would involve many repetitions from [7]. Instead, we provide the only required update as compared to the proof for 2-point functions given in the latter work. As such, this section should be considered as a complement, and we shall give exact references to the formulas in [7] whenever required.

As explained in Section 3, the correlation function $\langle \sigma_A \sigma_{B+x} \rangle_\beta$ admits a random-line representation of the form

$$\langle \sigma_A \sigma_{B+x} \rangle_\beta = \sum_{\underline{\gamma} \sim A \cup (B+x)} q_\beta(\underline{\gamma}) ,$$

where $\underline{\gamma}$ runs over families of compatible open contours connecting all the sites of $A \cup (B+x)$. Among the $\frac{1}{2}(|A|+|B|)$ paths of $\underline{\gamma}$, at least one must connect a site of A to a site of $B+x$. We first show that one can ignore the contribution of $\underline{\gamma}$ with more than one such connection (i.e. at least three of them). The first observation is that we have the following lower bound on the correlation function: By the second Griffiths' inequality,

$$\langle \sigma_A \sigma_{B+x} \rangle_\beta \geq \langle \sigma_{A \setminus \{y\}} \rangle_\beta \, \langle \sigma_{B \setminus \{z\}} \rangle_\beta \, \langle \sigma_y \sigma_{z+x} \rangle_\beta ,$$

where y and z are arbitrarily chosen sites of A and B respectively. Another application of the second Griffiths' inequality implies that

$$\langle \sigma_{A \setminus \{y\}} \rangle_\beta \, \langle \sigma_{B \setminus \{z\}} \rangle_\beta > 0 .$$

Moreover, we already know that

$$\langle \sigma_y \sigma_{z+x} \rangle_\beta = \Psi_\beta(n_x) \, |x|^{-(d-1)/2} \, e^{-\xi_\beta(x)} \, (1 + o(1)) .$$

But, applying (22), we obtain immediately that the contribution of families of paths $\underline{\gamma}$ with three or more connections between A and $B+x$ is bounded above by $C(A,B)\, e^{-3\xi_\beta(x)}$ and is therefore negligible.

We can henceforth safely assume that there is a single connection between A and $B+x$; we denote the corresponding path by γ, while the remaining paths are denoted by $\underline{\gamma}_A$ and $\underline{\gamma}_B$. We want to show that we can repeat the argument used for the two-point function in [7] in this more general setting. This is indeed quite reasonable since the paths in $\underline{\gamma}_A$ and $\underline{\gamma}_B$ should remain localized, and therefore the picture is still that of a single very long path as for 2-point functions. The main point is thus to prove sufficiently strong localization properties for the paths

$\underline{\gamma}_A$ and $\underline{\gamma}_B$, so as to ensure that an appropriate version Lemma 3.1 (see also Theorem 2.3 of [7]) still holds. The import of the latter lemma was to assert nice decay and decoupling properties of the integrated weights of the irreducible pieces in the decomposition of connection paths (8). Notice first that exactly the same decomposition can still be used here, provided we attach the paths in $\underline{\gamma}_A$ and $\underline{\gamma}_B$ to the corresponding leftmost and rightmost extremal pieces λ_L and λ_R, and keep the remaining intermediate cylindrical irreducible pieces unchanged. Apart from the compatibility requirements one then has to check that $\underline{\gamma}_B$ stays inside the forward cone containing λ_R, so that the crucial estimate (3.9) in [7] remains valid.

Let $y \in A$, $z \in B + x$, $\gamma : y \to z$, and let $\underline{\gamma}_A$, resp. $\underline{\gamma}_B$, denote the collections of remaining paths connecting pairs of sites in $A \setminus \{y\}$, respectively $x + B \setminus \{z\}$. For given collections $\underline{\gamma}_A$ and $\underline{\gamma}_B$ we define the irreducible decomposition of γ in precisely the same way as in (8), except for the extremal pieces $\lambda_L = (u_0^L, \dots, u_m^L)$ and $\lambda_R = (u_0^R, \dots, u_n^R)$, which have to satisfy the following modified set of conditions:

- $(u_k^L, \hat{x}) < (u_m^L, \hat{x})\ \forall\, k = 0, \dots, m-1$
- $(u_k^R, \hat{x}) > (u_0^R, \hat{x})\ \forall\, k = 1, \dots, n$
- $\underline{\gamma}_A$ must belong to the same $\hat{x}$-halfspace as λ_L and for any $\hat{x}$-break point u_k^L of λ_L the $\hat{x}$-orthogonal hyperplane through u_k^L intersects $\underline{\gamma}_A$.
- $\underline{\gamma}_B$ must belong to the same $\hat{x}$-halfspace as λ_R and for any $\hat{x}$-break point u_k^R of λ_R the $\hat{x}$-orthogonal hyperplane through u_k^R intersects $\underline{\gamma}_B$.
- $\underline{\gamma}_B$ must belong to $2K\mathbf{U}_\beta(u_0^R) + \mathcal{C}_\delta(t)$ (see (9)).

With a slight ambiguity of notation let us call compatible pairs $(\underline{\gamma}_A, \lambda_L)$ and $(\underline{\gamma}_B, \lambda_R)$ $\hat{x}$-irreducible if they satisfy all the conditions above. We then only have to check that

$$\sum_{\substack{(\underline{\gamma}_A, \lambda_L)\ \hat{x}\text{-irreducible} \\ \lambda_L : y \to u}} q_\beta(\underline{\gamma}_A, \lambda) \le e^{-(\hat{t}, u-y) - \nu |u-y|},$$

$$\sum_{\substack{(\underline{\gamma}_B, \lambda_R)\ \hat{x}\text{-irreducible} \\ \lambda_R : u \to z}} q_\beta(\underline{\gamma}_B, \lambda) \le e^{-(\hat{t}, z-u) - \nu |z-u|}$$

for some $\nu > 0$ and any $u \in \mathbb{Z}^d$. We only check the second statement since it is the most complicated one. Fix a large enough scale K. A site u of γ is a $(\hat{x}, \gamma_B, \delta)$-admissible break point if it is a $\hat{x}$-break point of γ

and, in addition,

$$\underline{\gamma}_B \subset 2K\mathbf{U}_\beta(u) + \mathcal{C}_\delta(\hat{t})\,.$$

Lemma 4.1. *Fix a forward cone parameter* $\delta \in (0, 1/4)$ *and a set* $B = \{y_1, z_1, \dots, y_n, z_n\}$; $B \subset \mathbb{Z}^d \setminus \{0\}$. *There exist a renormalization scale* K_0 *and positive numbers* $\epsilon = \epsilon(\delta, \beta)$, $\nu = \nu(\delta, \beta)$ *and* $N = N(\beta) < \infty$, *such that for all* $K \geq K_0$, *the upper bound*

$$\sum_{\substack{\lambda: -x \to 0 \\ \partial \underline{\gamma}_B}} q_\beta(\lambda, \underline{\gamma}_B) \mathbf{1}_{\{\lambda \textit{ has no } (\hat{x}, \underline{\gamma}_B)\textit{-admissible break points}\}} \leq N \mathrm{e}^{-(t,x)_d - \nu|x|}\,,$$

holds uniformly in the dual directions $t \in \partial \mathbf{K}_\beta$ *and in the starting points* $x \in \mathbb{Z}^d$. *In the first sum* $\underline{\gamma}_B, \lambda$ *runs over all admissible family of paths such that* $\lambda : -x \to 0$, *while* $\underline{\gamma}_B = (\gamma_1, \dots, \gamma_n)$ *satisfies* $\gamma_k : y_k \to z_k$.

Proof. Applying (22), we can assume that $x \in \mathcal{C}'_\nu(t)$, see the remark after Theorem 2.3 of [7]. Let $Q_C(x) = \{v \in \mathbb{Z}^d : |v| \leq |x|/C\}$ where C is some large enough constant. To simplify notations, we suppose that $B = \{y, z\}$, i.e. that $\underline{\gamma}_B \equiv \gamma : y \to z$. The general case is treated in the same way.

We first show that, typically, $\gamma \subset Q_C(x)$. Indeed, using again (22), we have that

$$\begin{aligned}
\sum_{\substack{(\lambda,\gamma) \\ \lambda:\, -x \to 0 \\ \gamma:\, y \to z}} q_\beta(\lambda, \gamma) \mathbf{1}_{\{\gamma \not\subset Q_C(x)\}} &\leq \sum_{u \in \partial Q_C(x)} \sum_{\substack{(\lambda, \gamma_1, \gamma_2) \\ \lambda:\, -x \to 0 \\ \gamma_1 :\, y \to u,\ \gamma_2 : u \to z}} q_\beta(\lambda, \gamma_1, \gamma) \\
&\leq \sum_{u \in \partial Q_C(x)} \langle \sigma_{-x} \sigma_0 \rangle_\beta \, \langle \sigma_y \sigma_u \rangle_\beta \, \langle \sigma_u \sigma_v \rangle_\beta \\
&\leq \frac{c_d |x|}{C} \, \mathrm{e}^{-c|x|} \, e^{-\xi_\beta(x)}\,.
\end{aligned}$$

We can therefore suppose that $\gamma \subset Q_C(x)$. Observe now that in the latter case

$$\begin{aligned}
&\{\lambda \text{ has no } (x, \gamma, 2\delta)\text{-break points}\} \\
&\qquad \subset \{\lambda \text{ has no } x\text{-break point } u \text{ with } (t, u) \leq -\tfrac{1}{2}(t, x)\} \\
&\qquad\qquad \overset{\triangle}{=} \mathcal{A}(t, K, \delta, x)\,,
\end{aligned}$$

provided that C is taken large enough. Indeed, would such a x-break point u exist then the cone $u + \mathcal{C}_{2\delta}(t)$ must contain the box $Q_C(x)$, hence also γ.

The probability of $\mathcal{A}(t, K, \delta, x)$ is estimated exactly as in the proof of Theorem 2.3 of [7]. Indeed, the presence of the path γ only affects an arbitrarily small fraction of the slabs $\mathcal{S}_k(t)$ introduced in the latter proof, provided C is taken large enough, so that the argument given there applies with no modifications. Q.E.D.

§5. Relation to Quantum Field Theories

There is an abundant literature devoted to the relation between Ising and other ferromagnetic type models to the Euclidean lattice quantum field theories, see e.g. [21, 18] or more recently [5, 4]; the latter article contains also an extensive bibliography on the subject. In this works the spins live on the integer lattice $\mathbb{Z}^{d+1}$ with one special direction, say $\vec{e}_1$, being visualized as the imaginary time axis. Thus, for example, the analyticity properties of the mixed Fourier transform

$$\mathbb{G}_\beta(p_1, i\mathbf{p}) = \sum_{x_1 \in \mathbb{Z}} \sum_{\mathbf{x} \in \mathbb{Z}^d} \mathrm{e}^{p_1 x_1 + i(\mathbf{p}, \mathbf{x})} G_\beta(x_1, \mathbf{x}), \tag{28}$$

$(p_1, \mathbf{p}) \in \mathbb{T} \times \mathbb{T}^d$, are related in this way to the question of existence of one particle states.

Below we shall briefly indicate how the the key probabilistic representation (20) leads to the following conclusion (see e.g Proposition 4.2 in [18], Theorem 2.3 in [21]): For every $\mathbf{p} \in \mathbb{T}^d$ define

$$\omega(\mathbf{p}) = -\lim_{n \to \infty} \frac{1}{n} \log \sum_{\mathbf{x} \in \mathbb{Z}^d} \mathrm{e}^{i(\mathbf{p}, \mathbf{x})} G_\beta(n, \mathbf{x}), \tag{29}$$

$\omega(\mathbf{p})$ being interpreted as the energy of a particle with momentum $\mathbf{p}$.

Theorem 5.1. *There exists a neighbourhood* $B_\delta = \{\mathbf{p} : |\mathbf{p}| < \delta\}$ *of the origin in* $\mathbb{R}^d$ *such that the function* $\mathbf{p} \mapsto \omega(\mathbf{p})$ *is real analytic on* B_δ. Hess (ω) (0) *is precisely the matrix of the second fundamental form of* $\partial \mathbf{K}_\beta$ *at* $\hat{t} = (\xi_\beta(\vec{e}_1), 0)$. *Furthermore, there exists* $\epsilon > 0$ *such that for every* $\mathbf{p} \in B_\delta$ *the function*

$$p_1 \mapsto \mathbb{G}_\beta(p_1, i\mathbf{p})$$

has a meromorphic extension to the disc $\{p_1 \in \mathbb{C} : |p_1 - \hat{p}_1| < \epsilon\}$; $\hat{p}_1 = \xi_\beta(\vec{e}_1)$, *with the only simple pole at* $p_1 = \omega(\mathbf{p})$.

In the sequel we use the notation introduced in Section 2. Because of the $\mathbb{Z}^d$-lattice symmetries the dual point $\hat{t} \in \partial \mathbf{K}_\beta$ of $\vec{e}_1$ is given by

$\hat{t} = (\hat{p}_1, 0)$ with $\hat{p}_1 = \xi_\beta(\vec{e}_1)$. Given $y \in \mathbb{Z}^{d+1}$ define (see (10))

$$W(y) = \sum_{\substack{\gamma:0\to y\\ \gamma\in\mathcal{S}}} \mathrm{e}^{\beta|\gamma|},$$

and let W_L, W_0 and W_R be defined as in Section 2. Summing up all the weights of irreducible paths in (8) we arrive to the following representation of G_β:

$$\begin{aligned} G_\beta(y) &= W(y) + \sum_{y_L+y_R=y} W_L(y_L)W_R(y_R) \\ &+ \sum_{y_L,y_R}\sum_{M=1}^{\infty} W_L(y_L)W_R(y_R)W_0^{*M}(y - y_L - y_R). \end{aligned} \tag{30}$$

Consider the mixed Fourier transforms

$$\mathbb{W}(p_1, \mathbf{p}) = \sum_{x_1\in\mathbb{Z}}\sum_{\mathbf{x}\in\mathbb{Z}^d} \mathrm{e}^{p_1x_1+(\mathbf{p},\mathbf{x})} W(x_1, \mathbf{x})$$

and

$$\mathbb{W}_b(p_1, \mathbf{p}) = \sum_{x_1\in\mathbb{Z}}\sum_{\mathbf{x}\in\mathbb{Z}^d} \mathrm{e}^{p_1x_1+(\mathbf{p},\mathbf{x})} W_b(x_1, \mathbf{x})\,;\ b = 0, L, R.$$

By Lemma 2.1 all four functions above are analytic in the complex neighbourhood $B_\nu^{\mathbb{C}}(\hat{t})$ of $\hat{t}$; $B_\nu^{\mathbb{C}}(\hat{t}) = \{(p_1, \mathbf{p}) : \sqrt{|p_1 - \hat{p}_1|^2 + |\mathbf{p}|^2} < \nu\}$. Thus, the extension of $\mathbb{G}_\beta(p_1, \mathbf{p})$ to $B_\nu^{\mathbb{C}}(\hat{t})$ is given by:

$$\mathbb{W}(p_1, \mathbf{p}) + \frac{\mathbb{W}_L(p_1, \mathbf{p})\mathbb{W}_R(p_1, \mathbf{p})}{1 - \mathbb{W}_0(p_1, \mathbf{p})}.$$

Consequently, the surface of poles of $\mathbb{G}_\beta(p_1, \mathbf{p})$ inside $B_\nu^{\mathbb{C}}(\hat{t})$ is given by the implicit equation

$$\mathbb{W}_0(p_1, \mathbf{p}) = 1. \tag{31}$$

As we have already seen in Section 2, the restriction of (31) to $(p_1, \mathbf{p}) \in \mathbb{R} \times \mathbb{R}^d$ defines the piece of the boundary $\partial\mathbf{K}_\beta$ inside $B_\nu(\hat{t})$. Since by (19) $\partial\mathbb{W}_0/\partial p_1(\hat{t}) \neq 0$ and, in addition, $\mathrm{Hess}(\mathbb{W}_0)(\hat{t})$ is non-degenerate, the analytic implicit function theorem implies that there exists $\delta > 0$, such that the equation (31) can be resolved for $\mathbf{p} \in B_\delta^{\mathbb{C}}(\hat{t}) \subset \mathbb{C}^d$ as

$$p_1 = \widetilde{\omega}(\mathbf{p}). \tag{32}$$

In particular, for $\mathbf{p} \in \mathbb{R}^d$ the equation (32) gives a parameterization of $\partial\mathbb{K}_\beta$ in the δ-neighbourhood of $\hat{t}$, and $\mathrm{Hess}(\widetilde{\omega})(0)$ is, indeed, the matrix of the second fundamental form of $\partial \mathbf{K}_\beta$ at $\hat{t}$. Finally, the 1-particle mass shell ω in (29) is recovered as $\omega(\mathbf{p}) = \widetilde{\omega}(i\mathbf{p})$.

In the case of ferromagnetic Ising models set $\vec{p} = (p_1, \mathbf{p})$ and readjust the definition (25) of the Ruelle operator $\mathcal{L}_z$ as $\widetilde{\mathcal{L}}_{\vec{p}} = \mathcal{L}_{\vec{p}-\hat{t}}$. Then $\widetilde{\mathcal{L}}_{\vec{p}}$ is well defined and bounded on $\mathfrak{F}_{0,\theta}$ for every $\vec{p} \in B_\nu^{\mathbb{C}}(\hat{t})$. It could be then shown that the surface of poles of $\mathbb{G}_\beta$ inside $B_\nu^{\mathbb{C}}(\hat{t})$ is implicitly given by

$$\tilde{\rho}_\beta(p_1, \mathbf{p}) \;=\; 1,$$

where $\tilde{\rho}_\beta(p_1, \mathbf{p})$ is the leading (lying on the spectral circle) eigenvalue of $\widetilde{\mathcal{L}}_{\vec{p}}$. Further analysis of the spectral properties of the family $\{\widetilde{\mathcal{L}}_{\vec{p}}\}$ reveals [7] that there exists $\epsilon > 0$, such that $\rho_\beta(p_1, \mathbf{p})$ is a simple pole of the corresponding resolvent for every $\vec{p} \in B_\epsilon^{\mathbb{C}}(\hat{t})$. In this way, the conclusion of Theorem 5.1 follows from the analytic perturbation theory of discrete spectra and from the conditional variance argument which ensures the non-degeneracy of $\mathrm{Hess}(\rho_\beta)(\hat{t})$.

References

[1] M. Aizenman (1982), *Geometric analysis of φ^4 fields and Ising models. I, II*, Comm. Math. Phys. **86**, no. 1, 1–48.

[2] M. Aizenman and D.J. Barsky (1987), *Sharpness of the phase transition in percolation models*, Comm. Math. Phys. **108**, no. 3, 489–526.

[3] M. Aizenman, D.J. Barsky and R. Fernández (1987), *The phase transition in a general class of Ising-type models is sharp*, J.Stat.Phys. **47**, 3/4, 342–374.

[4] F. Auil, J.C.A. Barata (2001), *Scattering and bound states in Euclidean lattice quantum field theories*, Ann.Inst.H. Poincaré **2**, 1065-1097.

[5] J.C.A. Barata, K. Fredenhagen (1991), *Particle scattering in Euclidean lattice field theories*, Comm.Math.Phys. **138**, 507-519.

[6] M. Campanino and D. Ioffe (1999) *Ornstein-Zernike Theory for the Bernoulli bond percolation on $\mathbb{Z}^d$*, Ann. Probab. **30**, no. 2, 652–682.

[7] M. Campanino, D. Ioffe D. and Y. Velenik (2003), *Ornstein-Zernike Theory for Finite-Range Ising Models Above T_c*, Probab. Theory Related Fields **125**, 305–349.

[8] M. Campanino, D. Ioffe D. and Y. Velenik (2003), *Rigorous Non-Perturbative Ornstein-Zernike Theory for Ising Ferromagnets*, Europhysics Letters **62**, no.2, 182–188.

[9] R. Durrett (1978), *On the shape of a random string*, Ann. Probab. **7**, (1979), 1014–1027.

[10] C.M. Fortuin, P.W. Kasteleyn (1972), *On the random-cluster model. I. Introduction and relation to other models*, Physica **57**, 536–564.

[11] J. Bricmont, A. El Mellouki, J. Fröhlich (1986), *Random surfaces in statistical mechanics: roughening, rounding, wetting,...*, J. Statist. Phys. **42**, no. 5-6, 743–798.

[12] Y. Higuchi (1979), *On some limit theorems related to the phase separation line in the two-dimensional Ising model*, Z. Wahrsch. Verw. Gebiete **50**, (1979), 287–315.

[13] L. Greenberg, D. Ioffe (2003), in preparation.

[14] D. Ioffe (1994), *Large deviations for the 2D Ising model: a lower bound without cluster expansions*, J. Statist. Phys. **74**, no. 1-2, 411–432.

[15] D. Ioffe (1998), *Ornstein-Zernike behaviour and analyticity of shapes for self-avoiding walks on* $\mathbb{Z}^d$, Mark.Proc.Rel.Fields **4**, 323–350.

[16] D. Ioffe, Y. Velenik, in preparation.

[17] E. Kovchegov (2002), *Brownian bridge asymptotics for the subcritical Bernoulli bond percolation*, preprint.

[18] P. Paes-Leme (1978), *Ornstein-Zernike and analyticity properties of classical lattice spin systems*, Ann.Physics **115**, 367-387.

[19] C.-E. Pfister, and Y. Velenik (1997), *Large Deviations and Continuum Limit in the 2D Ising Model*, Probab. Theory Related Fields **109**, 435–506.

[20] C.-E. Pfister, and Y. Velenik (1999), *Interface, surface tension and reentrant pinning transition in the 2D Ising model*, Comm. Math. Phys. **204**, no. 2, 269–312.

[21] R.S. Schor (1978), *The particle structure of* ν*-dimensional Ising models at low temperatures*, Comm.Math.Phys. **59**, 213-233.

[22] R. Schneider (1993), Convex Bodies: The Brunn-Minkowski Theory, Encyclopedia of Math. and its Applications **44**, Cambridge Univ. Press.

Massimo Campanino
Dipartimento di Matematica
Università di Bologna
Piazza di Porta S. Donato 5
I-40126 Bologna
Italy

Dmitry Ioffe
Faculty of Industrial Engineering
Technion
Haifa 3200
Israel

Yvan Velenik
UMR-CNRS 6632
Université de Provence
39 rue Joliot Curie
13453 Marseille
France

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 53–88

Spectral Gap Inequalities in Product Spaces with Conservation Laws

Pietro Caputo

Abstract.

Following an idea introduced by Carlen, Carvalho and Loss [7] we propose a general strategy to prove Poincaré inequalities in product spaces with one or more conservation laws. The method is shown to yield alternative proofs of well known results, such as the diffusive bounds for the spectral gap of generalized exclusion and zero range processes. Other models are also discussed, including anisotropic exclusion processes, simple exclusion with site–disorder and Ginzburg–Landau processes, where this approach provides sharp spectral gap estimates apparently inaccessible by previously known techniques.

§1. Introduction

The problem of determining the speed of convergence to equilibrium of conservative stochastic dynamics has motivated many investigations in recent years. In the context of reversible processes the simplest way to attack this question is by estimating the spectral gap of the corresponding Markov generators or – equivalently – by proving a Poincaré inequality. In this direction an important achievement are the diffusive estimates established for Kawasaki dynamics in high temperature lattice gases by Lu and Yau [21] and by Cancrini and Martinelli [3]. In this paper we confine ourselves to systems whose underlying equilibrium measure is *product* and the only remaining interaction is due to the global *conservation law.* Although this is certainly a radical simplification, we shall see that already in this class one finds interesting models for which traditional techniques apparently fail to give optimal spectral gap bounds.

The simplest model in this class is the *simple exclusion* process, for which sharp spectral gap estimates are well known, at least since the

---

Received December 24, 2002.
Revised February 14, 2003.

work of Quastel, [22]. Other conservative dynamics sharing the product–property are the so–called *generalized simple exclusion* processes and the *zero range* process. For these models the martingale approach of [21] was successfully applied by Landim, Sethuraman and Varadhan [20] to show that the spectral gap scales diffusively with the size of the system, uniformly in the conserved parameter. A rather complete picture of decay to equilibrium for the zero range process was then obtained by Janvresse, Landim, Quastel and Yau [14].

As already noted in [22], when the system is of product type it is natural to drop all geometrical constraints in the dynamics and consider processes where exchanges are performed along the edges of a complete graph rather than only along nearest neighbors edges. As we shall see in all the examples treated in this note, once one has a Poincaré inequality for this complete graph (mean–field) dynamics a straightforward comparison argument allows to derive diffusive scaling bounds for the local exchange dynamics.

An example of complete graph dynamics is the model proposed by Kac [15] to study trend to equilibrium for the Boltzmann equation. Spectral gap estimates for this process were investigated by Diaconis and Saloff-Coste [10], and by Janvresse [13]. The latter work catches the right shrinking–rate of the spectral gap by adapting the martingale approach of [21]. Recent remarkable work of Carlen, Carvalho and Loss [7, 8] however shows that spectral gap estimates for the Kac model can be sharpened considerably if one–site Poincaré inequalities in the martingale approach are replaced by a fine analysis of the spectrum of an auxiliary Markov process.

As observed in [8] their approach can be generalized to treat a broader class of models than just the Kac model. Our aim in this paper is to show that in principle some of the ideas of [7] apply to all conservative systems of product type. In the case of Kac and related models considered in [8] the spectrum of the auxiliary process can be computed rather explicitly in view of the special form of the probability measures involved. This is in general not the case for the models discussed here and the main technical ingredient in our estimates are uniform local expansions related to the central limit theorem.

Here is a plan of the paper. In section 2 we discuss the auxiliary dynamics introduced in [7] and outline a general strategy to prove uniform spectral gap estimates in product spaces with one or more conservation laws. Here we present explicit sufficient conditions to be checked in specific models. The known results on generalized exclusion and zero range processes mentioned above are re-derived in a compact way in section 3 and section 4, respectively. A simple instance of a model with many conservation laws is considered in section 5. Recent results on anisotropic exclusion and Ginzburg–Landau processes appearing in [5] and [4] are reviewed in section 6 and section 7, respectively. Finally in section 8 we prove a new estimate for the simple exclusion process with site disorder.

§2. A general strategy

Consider a generic probability space $(X, \mathcal{F}, \mu)$. In the applications to be discussed below we shall choose $X = \mathbb{N}, \mathbb{Z}$ or $\mathbb{R}$ depending on the specific model. For every $N \in \mathbb{N}$ denote by Ω_N the N-fold product of X, $\Omega_N = X^N$ and by $\mu_N = \mu^{\otimes N}$ the associated product measure. The conservation law is expressed in terms of a given measurable function $\xi : X \to \mathbb{R}$, with $\xi \in L^2(\mu)$. Namely, given a parameter $\rho \in \mathbb{R}$ to play the role of a density, we shall look at configurations $\eta = \{\eta_k\}_{k=1}^N \in \Omega_N$ such that $\sum_{k=1}^N \xi(\eta_k) = \rho N$. If we define $\xi_\rho = \xi - \rho$, we consider the measurable set

$$\Theta_{N,\rho} := \{\eta \in \Omega_N : \sum_{k=1}^{N} \xi_\rho(\eta_k) = 0\}. \tag{1}$$

Whenever it makes sense we define the canonical probability measure by conditioning on the event $\Theta_{N,\rho}$:

$$\nu_{N,\rho} = \mu_N(\ \cdot \mid \Theta_{N,\rho}). \tag{2}$$

The complete graph dynamics will be described by a Dirichlet form of the type

$$\mathcal{E}_{N,\rho}(f) = \frac{1}{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \nu_{N,\rho}\big[(v_{k,\ell} f)^2\big], \tag{3}$$

where $v_{k,\ell}$ are generic exchange operators to be specified in each model. For the moment we only require that $v_{k,k} = 0$, $k = 1, 2, \dots, N$. To carry a concrete example in mind we recall that the complete graph exclusion process is recovered in the case $X = \{0, 1\}$, $\mu = \mathrm{Be}(p)$, any $p \in (0, 1)$; $\xi(\eta_k) = \eta_k$, $[v_{k,\ell} f](\eta) = f(\eta^{k,\ell}) - f(\eta)$, with $\eta^{k,\ell}$ denoting the configuration η where η_k and η_ℓ have been exchanged.

We denote by $\mathrm{Var}_{N,\rho}(f)$ the usual variance of $f \in L^2(\Omega_N, \nu_{N,\rho})$ with respect to $\nu_{N,\rho}$. The Poincaré constant for fixed N and ρ is defined by

$$\gamma(N, \rho) = \sup_{f} \frac{\mathrm{Var}_{N,\rho}(f)}{\mathcal{E}_{N,\rho}(f)}, \tag{4}$$

with the supremum ranging over functions f in the domain of the Dirichlet form $\mathcal{E}_{N,\rho}$. Definition (4) is meaningful for all ergodic processes, i.e. when $\mathrm{Var}_{N,\rho}(f) > 0$ implies $\mathcal{E}_{N,\rho}(f) > 0$, and we set by convention $\gamma(N, \rho) = 0$ in all degenerate cases, i.e. when $\mathrm{Var}_{N,\rho}(f) = \mathcal{E}_{N,\rho}(f) = 0$ for all f, such as e.g. the exclusion process with $\rho \in \{0, 1\}$. We say that $\nu_{N,\rho}$ satisfies a *uniform Poincaré inequality* if $\sup_N \sup_\rho \gamma(N, \rho) < \infty$.

2.1. The auxiliary process

Let $\mathcal{F}_k$ denote the σ–algebra generated by the one–site variables η_k, $k = 1, \ldots, , N$. Following [7, 8] we consider the nonnegative stochastic operator $\mathcal{P} : L^2(\nu_{N,\rho}) \to L^2(\nu_{N,\rho})$ defined by

$$\mathcal{P}f = \frac{1}{N} \sum_{k=1}^{N} \nu_{N,\rho}(f \,|\, \mathcal{F}_k) \,. \tag{5}$$

Then $1 - \mathcal{P}$ can be interpreted as the generator of a new Markov process with reversible invariant measure $\nu_{N,\rho}$. This is completely independent of the actual dynamics defined by (3), but we will see in a moment that an estimate on the spectral gap of this process produces useful recursive bounds on the constants $\gamma(N, \rho)$. To gain some insight observe that by symmetry

$$\nu_{N,\rho}(\xi(\eta_k)|\eta_j) = \rho_{\eta_j} := \rho + \frac{\rho - \xi(\eta_j)}{N-1} \tag{6}$$

whenever $k \neq j$, so that

$$\nu_{N,\rho}(\xi_\rho(\eta_k) \,|\, \eta_j) = -\frac{1}{N-1}\, \xi_\rho(\eta_j) \,, \qquad k \neq j \,. \tag{7}$$

Here and in what follows we often write (with slight abuse) $\nu(f \,|\, \eta_j)$ for the function $\nu(f \,|\, \mathcal{F}_j)(\eta)$. It follows that any function of the form

$$f_\xi(\eta) = \sum_{k=1}^{N} \alpha_k \xi_\rho(\eta_k) \,, \qquad \alpha \in \mathbb{R}^N \tag{8}$$

satisfies

$$\mathcal{P}f_\xi = \frac{1}{N-1}\, f_\xi \,, \qquad (1 - \mathcal{P}) f_\xi = \frac{N-2}{N-1}\, f_\xi. \tag{9}$$

We formulate the needed spectral gap inequality as follows. We say that property (SGP) holds if there exists $C < \infty$, $\delta > 0$ such that for every $N \geqslant 3$, $\rho \in \mathbb{R}$ and $f \in L^2(\nu_{N,\rho})$ with $\nu_{N,\rho}(f) = 0$:

$$\nu_{N,\rho}(f(1 - \mathcal{P})f) \geqslant \frac{N-2}{N-1} \left[1 - CN^{-1-\delta}\right] \nu_{N,\rho}(f^2) \,. \tag{\textbf{SGP}}$$

We now turn to the implications of such a bound. A useful criterium to check the bound (SGP) in specific models will be developed in the next subsection. We define the constant

$$\gamma(N) := \sup_{\rho} \gamma(N, \rho) . \tag{10}$$

Proposition 2.1. *Assume $\gamma(N) < \infty$ for every $N \in \mathbb{N}$. If (SGP) holds then we have the uniform Poincaré inequality*

$$\sup_{N} \gamma(N) < \infty \tag{11}$$

Proof. It is sufficient to show that (SGP) implies a bound of the form

$$\gamma(N) \leqslant \left[1 + CN^{-1-\delta}\right] \gamma(N-1) , \tag{12}$$

with $C < \infty$ and $\delta > 0$ independent of ρ and N.

Take an arbitrary function[1] $f \in L^2(\nu_{N,\rho})$ with $\nu_{N,\rho}(f) = 0$. The conditional expectation $\nu_{N,\rho}(f \mid \eta_k)$ is identified with the average $\nu_{N-1,\rho_{\eta_k}}(f)$, where ρ_{η_k} is given in (6). For each k we then have the decomposition

$$\nu_{N,\rho}(f^2) = \nu_{N,\rho}\left[\mathrm{Var}_{N-1,\rho_{\eta_k}}(f)\right] + \nu_{N,\rho}\left[\nu_{N,\rho}(f \mid \eta_k)^2\right] .$$

Averaging over k:

$$\nu_{N,\rho}(f^2) = \frac{1}{N} \sum_{k=1}^{N} \nu_{N,\rho}\left[\mathrm{Var}_{N-1,\rho_{\eta_k}}(f)\right] + \nu_{N,\rho}\left[f \mathcal{P} f\right] \tag{13}$$

with the operator $\mathcal{P}$ defined in (5). By definition of the constants (10):

$$\begin{aligned} \mathrm{Var}_{N-1,\rho_{\eta_k}}(f) &\leqslant \gamma(N-1)\, \mathcal{E}_{N-1,\rho_{\eta_k}}(f) \\ &= \frac{\gamma(N-1)}{N-1} \sum_{j \neq k} \sum_{\ell \neq k} \nu_{N,\rho}\left[(v_{j,\ell} f)^2 \mid \mathcal{F}_k\right] \end{aligned} \tag{14}$$

From (13)–(14) and the identity

$$\frac{1}{N} \sum_{k=1}^{N} \nu_{N,\rho}\left[\mathcal{E}_{N-1,\rho_{\eta_k}}(f)\right] = \frac{N-2}{N-1} \mathcal{E}_{N,\rho}(f)$$

---

[1] In this proof we shall not be careful about questions of domains of the Dirichlet forms $\mathcal{E}_{N,\rho}$. It is however straightforward to settle these issues in all the following applications.

we obtain the estimate

$$\nu_{N,\rho}\big[f(1-\mathcal{P})f\big] \leqslant \frac{N-2}{N-1}\,\gamma(N-1)\,\mathcal{E}_{N,\rho}(f)\,. \tag{15}$$

Now (12) follows from (15) and the hypothesis (SGP). Q.E.D.

2.2. Reduction to one–dimensional process

As in [8] the spectrum of $\mathcal{P}$ can be studied in terms of the spectrum of a one–dimensional operator $\mathcal{K}$, see (16) below. Here we show that the estimate (SGP) is implied by a suitable spectral estimate on $\mathcal{K}$, see (SGK) below.

Let π_k be the canonical projection of Ω_N onto X given by $\pi_k\eta = \eta_k$. We call $\nu^1_{N,\rho}$ the one–site marginal of $\nu_{N,\rho}$, i.e. $\nu^1_{N,\rho} = \nu_{N,\rho}\circ\pi_1^{-1}$ is the distribution of η_1 under $\nu_{N,\rho}$. By permutation symmetry all one–site marginals coincide. Let $\mathcal{H}$ denote the Hilbert space $L^2(X,\nu^1_{N,\rho})$ and use $\langle\cdot,\cdot\rangle$ for the corresponding scalar product. Write also $\langle g\rangle$ for the mean of a function $g\in\mathcal{H}$ w.r.t. $\nu^1_{N,\rho}$. We write $\mathcal{H}_0$ for the subspace of $g\in\mathcal{H}$ such that $\langle g\rangle = 0$. We define the stochastic self–adjoint operator $\mathcal{K}:\mathcal{H}\to\mathcal{H}$ by the bilinear form:

$$\langle g,\mathcal{K}h\rangle = \nu_{N,\rho}\big[(g\circ\pi_1)(h\circ\pi_2)\big]\,, \qquad g,h\in\mathcal{H}\,. \tag{16}$$

The identity (7) shows that

$$\mathcal{K}\xi_\rho = -\frac{1}{N-1}\,\xi_\rho \tag{17}$$

for every ρ. Thus the spectrum of $\mathcal{K}$ always contains the eigenvalues $-\frac{1}{N-1}$ and 1. We say that property (SGK) holds if the rest of the spectrum of $\mathcal{K}$ is confined around zero within a neighborhood of radius $O(N^{-1-\delta})$ for some $\delta>0$ uniformly in N,ρ, i.e. if there exist constants $C<\infty,\delta>0$ such that for every N and ρ, for every $g\in\mathcal{H}_0$ satisfying $\langle g,\xi_\rho\rangle = 0$ one has

$$\big|\langle g,\mathcal{K}g\rangle\big| \leqslant C\,N^{-1-\delta}\,\langle g,g\rangle\,. \tag{\textbf{SGK}}$$

Lemma 2.2. *(SGK) implies (SGP).*

Proof. We define the closed subspace Γ of $L^2(\nu_{N,\rho})$ consisting of sums of mean–zero functions of a single variable:

$$\Gamma = \Big\{f\in L^2(\nu_{N,\rho}) :\ f = \sum_{k=1}^N g_k\circ\pi_k\,;\quad g_1,\dots,g_N\in\mathcal{H}_0\,,\ \Big\} \tag{18}$$

We first observe that $\mathcal{P}f \in \Gamma$ for every $f \in L^2(\nu_{N,\rho})$ with $\nu_{N,\rho}(f) = 0$. Therefore $\mathcal{P}f = 0$ whenever $f \in \Gamma^\perp$, f with mean zero. In particular we may restrict to $f \in \Gamma$ to prove (SGP).

Given $f \in \Gamma$, $f = \sum_k g_k \circ \pi_k$, we define $\varphi_f = \sum_k g_k$, a function in $\mathcal{H}_0$. A simple computation shows that

$$\nu_{N,\rho}(f^2) = \langle \varphi_f, \mathcal{K}\varphi_f \rangle + \sum_k \langle g_k, (1-\mathcal{K}) g_k \rangle \,, \tag{19}$$

where $\mathcal{K}$ is the operator defined in (16). Similarly one computes

$$\begin{aligned} \nu_{N,\rho}\big(f(1-\mathcal{P})f\big) = {} & \frac{N-2}{N} \langle \varphi_f, \mathcal{K}(1-\mathcal{K})\varphi_f \rangle \\ & + \frac{1}{N} \sum_k \langle g_k, (1-\mathcal{K})[(N-1)+\mathcal{K}] g_k \rangle \,. \end{aligned} \tag{20}$$

Consider now the subspace $\mathcal{S} \subset \Gamma$ of symmetric functions:

$$\mathcal{S} = \Big\{ f \in L^2(\nu_{N,\rho}) : \ f = \sum_{k=1}^N g \circ \pi_k \,, \ \ g \in \mathcal{H}_0 \Big\} \,. \tag{21}$$

Since $\mathcal{S}$ is invariant for $\mathcal{P}$, i.e. $\mathcal{P}\mathcal{S} \subset \mathcal{S}$ we may consider separately the cases $f \in \mathcal{S}$ and $f \in \mathcal{S}^\perp$, with $\mathcal{S}^\perp$ denoting the orthogonal complement in Γ. When $f \in \mathcal{S}$, $f = \sum_{k=1}^N g \circ \pi_k$ we have $\varphi_f = Ng$ and rearranging terms in (19) and (20) we obtain

$$\nu_{N,\rho}(f^2) = N(N-1) \langle g, [\mathcal{K} + \frac{1}{N-1}] g \rangle \tag{22}$$

$$\nu_{N,\rho}\big(f(1-\mathcal{P})f\big) = (N-1)^2 \langle g, [1-\mathcal{K}][\mathcal{K} + \frac{1}{N-1}] g \rangle \tag{23}$$

From (SGK) we see that $\mathcal{K} + \frac{1}{N-1}$ is nonnegative on the whole subspace $\mathcal{H}_0$. Moreover, since $f = 0$ when g is a multiple of ξ_ρ, we may then restrict to the case $\langle g, \xi_\rho \rangle = 0$. Writing $\tilde{g} = [\mathcal{K} + \frac{1}{N-1}]^{\frac{1}{2}} g$ and observing that $\langle \tilde{g} \rangle = 0$ and $\langle \tilde{g}, \xi_\rho \rangle = 0$, the assumption (SGK) implies

$$\begin{aligned} \nu_{N,\rho}\big(f(1-\mathcal{P})f\big) & \geqslant (N-1)^2 \big[1 - CN^{-1-\delta}\big] \langle \tilde{g}, \tilde{g} \rangle \\ & \geqslant \frac{N-2}{N-1} \big[1 - CN^{-1-\delta}\big] \nu_{N,\rho}(f^2) \,, \qquad f \in \mathcal{S} \,. \end{aligned} \tag{24}$$

We turn to study the case $f \in \mathcal{S}^\perp$. Let us first observe that one can assume without loss that $f \in \Gamma$ is such that $\langle \varphi_f, \xi_\rho \rangle = \sum_k \langle g_k, \xi_\rho \rangle = 0$. Indeed if $c = (N \langle \xi_\rho, \xi_\rho \rangle)^{-1} \sum_k \langle g_k, \xi_\rho \rangle$ and $\tilde{g}_k = g_k - c\xi_\rho$, we have

$\sum_k \tilde{g}_k \circ \pi_k = \sum_k g_k \circ \pi_k$ in $L^2(\nu_{N,\rho})$ since by the conservation law $\sum_k \xi_\rho \circ \pi_k = 0$. Now, for every $u \in \mathcal{S}$, $u = \sum_k u_0 \circ \pi_k$, with $u_0 \in \mathcal{H}_0$ one has

$$\nu_{N,\rho}(uf) = (N-1)\,\langle \varphi_f, [\mathcal{K} + \frac{1}{N-1}]u_0\rangle\,.$$

Thus $f \in \mathcal{S}^\perp$ implies that $[\mathcal{K}+\frac{1}{N-1}]\varphi_f$ is a constant in $\mathcal{H}$. Since $\langle \varphi_f \rangle = 0$ and $\langle \varphi_f, \xi_\rho\rangle = 0$, (SGK) implies $\varphi_f = 0$. Writing $\hat{g}_k = (1-\mathcal{K})^{\frac{1}{2}} g_k$, then (19) and (20) imply

$$\nu_{N,\rho}(f^2) = \sum_k \langle \hat{g}_k, \hat{g}_k\rangle \tag{25}$$

$$\nu_{N,\rho}\big(f(1-\mathcal{P})f\big) = \frac{1}{N}\sum_k \langle \hat{g}_k, [(N-1)+\mathcal{K}]\hat{g}_k\rangle \tag{26}$$

Since $\langle \hat{g}_k\rangle = 0$ for all k we use (SGK) to estimate

$$\langle \hat{g}_k, \mathcal{K}\hat{g}_k\rangle \geqslant -\frac{1}{N-1}\langle \hat{g}_k, \hat{g}_k\rangle\,.$$

From (25) and (26) we obtain

$$\nu_{N,\rho}\big(f(1-\mathcal{P})f\big) \geqslant \frac{N-2}{N-1}\nu_{N,\rho}(f^2)\,, \qquad f \in \mathcal{S}^\perp\,. \tag{27}$$

From (24) and (27) we obtain (SGP) and the proof is completed. Q.E.D.

2.3. Several conservation laws

In the case of more than one conservation law we are given an r–dimensional vector $\bar{\xi} = (\xi^1, \dots, \xi^r)$ of measurable functions $\xi^j : X \to \mathbb{R}$, for some positive integer r, and we require that

$$\sum_{k=1}^{N} \xi^j(\eta_k) = \rho^j N\,, \qquad j = 1, \dots, r$$

with $\bar{\rho} := (\rho^1, \dots, \rho^r)$ an assigned density vector. If we denote $\Theta_{N,\bar{\rho}}$ the event realizing simultaneously all the constraints above we then define the conditional probability measure

$$\nu_{N,\bar{\rho}} = \mu_N(\,\cdot \mid \Theta_{N,\bar{\rho}})\,. \tag{28}$$

With these notations the argument of Proposition 2.1 carries over with no change provided we replace ρ with $\bar{\rho}$. We observe that (17) now holds for every $\xi^j_{\rho^j}$, $j = 1, \dots, r$. Moreover, as in Lemma 2.2 one proves that (SGP) can be obtained as a consequence of (SGK), provided the latter condition is modified by requiring the spectral estimate for any $g \in \mathcal{H}_0$ which is orthogonal to all functions $\xi^j_{\rho^j}$ simultaneously. As a simple example of a system with several conservation laws we will discuss the colored exclusion process in section 5.

2.4. From complete graph to local exchanges

In many applications it is interesting to consider local versions of the conservative dynamics. In analogy with (3) we describe such local dynamics by the Dirichlet form

$$\mathcal{D}_{N,\rho}(f) = \sum_{k=1}^{N-1} \nu_{N,\rho}\big[(v_{k,k+1}f)^2\big]\,. \tag{29}$$

The standard tool to compare the forms $\mathcal{D}_{N,\rho}$ and $\mathcal{E}_{N,\rho}$ is what is often called (for obvious reasons) the moving–particle lemma. In this general setting we may state this as follows. We say that a moving–particle lemma holds, or simply that (MP) holds if there exists a constant $C < \infty$ such that for every N and ρ, every integer $n \leqslant N$ and every f one has

$$\nu_{N,\rho}\big[(v_{1,n}f)^2\big] \leqslant C\,n \sum_{k=1}^{n-1} \nu_{N,\rho}\big[(v_{k,k+1}f)^2\big]\,. \tag{\textbf{MP}}$$

A simple consequence of (MP) is the comparison estimate

$$\mathcal{E}_{N,\rho}(f) \leqslant C\,N^2\,\mathcal{D}_{N,\rho}(f)\,. \tag{30}$$

Thus, if we are able to prove the uniform Poincaré inequality (11) and (MP) holds we can infer uniform diffusive estimates for the local dynamics. These arguments can be generalized in a straightforward way to treat local dynamics in which particles are located at the sites of a box in a d–dimensional lattice $\mathbb{Z}^d$, any $d \geqslant 1$. Suppose for instance $N = L^d$, for some $L \in \mathbb{N}$, is the cardinality of the hypercube $\Lambda_L = \{1, \ldots, L\}^d \subset \mathbb{Z}^d$ and we are interested in a process defined by the Dirichlet form

$$\widetilde{\mathcal{D}}_{L,\rho}(f) = \sum_{\substack{x,y\in\Lambda_L:\\ |x-y|=1}} \nu_{N,\rho}\big[(v_{x,y}f)^2\big]\,, \tag{31}$$

where $|x| := \sum_{i=1}^d |x_i|$, $x \in \mathbb{Z}^d$. Then, assuming (MP), a straightforward path–counting argument gives the diffusive bound

$$\mathcal{E}_{N,\rho}(f) \leqslant C\,L^2\,\widetilde{\mathcal{D}}_{L,\rho}(f)\,. \tag{32}$$

We shall see that all the examples we consider hereafter do satisfy the (MP) property.

§3. Generalized exclusion

Here we take $X = \{0, 1, \ldots, R\}$, R a given integer, and μ a probability measure on X such that $p(n) := \mu(\eta_1 = n) > 0$ for all $n = 0, 1, \ldots, R$. Ω_N is the space of configurations $\eta = (\eta_k)$, with the interpretation that η_k is the number of particles at site k. Here $\xi(\eta_k) = \eta_k$ and the total number of particles is conserved. For any $\rho \in I_{R,N} := \{0, \frac{1}{N}, \frac{2}{N}, \ldots, R - \frac{1}{N}, R\}$ we have the canonical measure $\nu_{N,\rho}$ defined by (2).

The generalized exclusion process on the complete graph $\{1, 2, \ldots, N\}$ can be loosely described as follows. At each site a Poisson clock rings with rate 1. When site k rings we choose uniformly one of the sites, say j. If $k \neq j$, if site k contains at least one particle (i.e. $\eta_k > 0$) and site j is not saturated (i.e. $\eta_j < R$), a particle is moved from k to j with rate $c(\eta_k, \eta_j)$, otherwise nothing happens. The rates $c(\cdot, \cdot)$ are chosen in such a way that the resulting process is reversible w.r.t. $\nu_{N,\rho}$. A possible choice is for instance $c(\eta_j, \eta_k) = 1/[p(\eta_j)p(\eta_k)]$. In any case, assuming a uniform bound from above and below on the rates $c(\cdot, \cdot)$, the resulting Dirichlet form is controlled (up to multiplicative constants) in terms of the quadratic form

$$(33) \qquad \mathcal{E}_{N,\rho}(f) = \frac{1}{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \nu_{N,\rho}\big[(v_{k,\ell} f)^2\big], \qquad v_{k,\ell} f = f \circ T_{k,\ell} - f$$

where f is any real function on Ω_N and

$$\big(T_{k,\ell}\eta\big)_j = \begin{cases} \eta_k - 1 & \text{if } j = k\,,\ \eta_k > 0 \text{ and } \eta_\ell < R \\ \eta_\ell + 1 & \text{if } j = \ell\,,\ \eta_k > 0 \text{ and } \eta_\ell < R \\ \eta_j & \text{otherwise}\,. \end{cases}$$

As in Lemma A.2.8 of [16] (p.392) it is not difficult to prove that property (MP) holds for this model. In particular, by (30)–(32) the estimate of Theorem 3.1 below immediately implies the well known diffusive scaling estimate (as given e.g. in [16], Theorem A.2.1).

Theorem 3.1. *For every $R \in \mathbb{N}$ there exists $C < \infty$ such that*

$$\sup_{N \geqslant 2} \sup_{\rho \in I_{R,N}} \gamma(N, \rho) \leqslant C\,.$$

The proof of Theorem 3.1 is based on Proposition 2.1. We thus have to check that $\sup_\rho \gamma(N, \rho)$ is finite for all N and that property (SGP) holds.

The first requirement is easily seen to be satisfied. Namely for every fixed N and $\rho \in I_{R,N}$, $\rho \neq 0, R$, the process is ergodic, i.e. whenever

$f \in L^2(\nu_{N,\rho})$ is such that $\mathcal{D}_{N,\rho}(f) = 0$ then f is constant over $\Theta_{N,\rho}$. This implies that $\gamma(N, \rho) < \infty$. Since ρ can take only a finite number of values we have $\gamma(N) = \sup_\rho \gamma(N, \rho) < \infty$ for every fixed N.

To prove (SGP) we rely on Lemma 2.2. In this setting the operator $\mathcal{K}$ defined in (16) is a $(R+1) \times (R+1)$–matrix with entries

$$\mathcal{K}(n,m) = \nu_{N,\rho}(\eta_2 = m \,|\, \eta_1 = n)\,.$$

In order to simplify the notation we adopt the following shortcuts:

$$\nu(n) := \nu_{N,\rho}(\eta_1 = n)\,, \quad \nu(n,m) := \nu_{N,\rho}(\eta_1 = n, \eta_2 = m) \tag{34}$$

For any function $\varphi \in \mathcal{H}_0$ we have

$$\langle \varphi, \mathcal{K}\varphi \rangle = \sum_{n=0}^{R} \sum_{m=0}^{R} \nu(n)\nu(m)Q(n,m)\varphi(n)\varphi(m) \tag{35}$$

where we introduce the kernel

$$Q(n,m) = \frac{\nu(n,m) - \nu(n)\nu(m)}{\nu(n)\nu(m)} \tag{36}$$

The proof of (SGK) will be obtained by a careful examination of the kernel Q. If $\varphi \in \mathcal{H}_0$ is such that $\langle \varphi, \xi_\rho \rangle = 0$ as in the hypothesis of (SGK), then from (35) we have

$$\langle \varphi, \mathcal{K}\varphi \rangle = \sum_{n=0}^{R} \sum_{m=0}^{R} \nu(n)\nu(m) \Big[Q(n,m) + \frac{\xi_\rho(n)\xi_\rho(m)}{\sigma_\rho^2 N}\Big]\varphi(n)\varphi(m)\,,$$

where σ_ρ^2 refers to the grand–canonical variance at density ρ, see (39) below. Therefore (SGK) follows from the Schwarz' inequality and Proposition 3.2 below, which we prove in the next subsection.

Proposition 3.2. *For every $R \in \mathbb{N}$ there exists $C < \infty$ and $\delta > 0$ such that*

$$\sum_{n=0}^{R} \sum_{m=0}^{R} \nu(n)\nu(m) \Big[Q(n,m) + \frac{\xi_\rho(n)\xi_\rho(m)}{\sigma_\rho^2 N}\Big]^2 \leqslant C\,N^{-2-\delta} \tag{37}$$

3.1. Proof of Proposition 3.2

We start with some preliminaries. Let $\bar{\mu}_\alpha$, $\alpha > 0$, be the probability measure on X defined by

$$\bar{\mu}_\alpha(\eta_1 = k) = \frac{p(k)\alpha^k}{\bar{Z}_\alpha}\,, \qquad \bar{Z}_\alpha = \sum_{j=0}^{R} p(j)\alpha^j\,. \tag{38}$$

Let $\rho = \rho(\alpha)$ be the average number of particles according to $\bar{\mu}_\alpha$:

$$\rho = \frac{1}{\bar{Z}_\alpha} \sum_{k=1}^{R} kp(k)\alpha^k$$

Since the function $\rho : [0, \infty] \to [0, R]$ is strictly increasing, with $\rho'(\alpha) = \alpha^{-1}\mathrm{Var}_{\bar{\mu}_\alpha}(\eta_1)$, we can invert it to to find the function $\alpha(\rho) : [0, R] \to [0, \infty]$. From now on we shall write μ_ρ for the measure $\bar{\mu}_{\alpha(\rho)}$. We call σ_ρ^2 the variance

$$\sigma_\rho^2 = \mathrm{Var}_{\mu_\rho}(\eta_1) \tag{39}$$

Clearly $\sigma_\rho^2 \leqslant R^2/2$, and $\sigma_\rho^2 \to 0$ when $\rho \to 0$ or $\rho \to R$. Define $p_\rho(k) := \mu_\rho(\eta_1 = k)$. It is simple to check the following estimates, to be used for small density ρ:

(40)
$$p_\rho(0) = 1 - \rho + O(\rho^2)\,, \quad p_\rho(1) = \rho + O(\rho^2)\,, \quad p_\rho(k) = O(\rho^k)\,,\ k \geqslant 2\,.$$

In particular, $\sigma_\rho^2 = \rho + O(\rho^2)$, as $\rho \to 0$. By duality the same estimate holds with ρ replaced by $R - \rho$ when $\rho \to R$. The characteristic function of the rescaled variable ξ_ρ/σ_ρ is defined by

$$v_\rho(\zeta) = \mu_\rho\big(\exp(i\zeta\xi_\rho/\sigma_\rho)\big) \tag{41}$$

Lemma 3.3. *There exists $a = a(R) > 0$ such that for every $\rho \in (0, R)$*

$$|v_\rho(\zeta)| \leqslant e^{-a\zeta^2}\,, \qquad \zeta \in [-\pi\sigma_\rho, \pi\sigma_\rho]\,.$$

Proof. Observe that by the trigonometric identity $\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$:

$$\begin{aligned} |v_\rho(\zeta)|^2 &= \mu_\rho\big[\cos(\zeta\xi_\rho/\sigma_\rho)\big]^2 + \mu_\rho\big[\sin(\zeta\xi_\rho/\sigma_\rho)\big]^2 \\ &= \sum_{k=0}^{R}\sum_{j=0}^{R} p_\rho(k)p_\rho(j)\cos[\zeta(k-j)/\sigma_\rho]\,. \end{aligned}$$

Now estimate

$$\cos[\zeta(k-j)/\sigma_\rho] \leqslant \begin{cases} 1 & \text{if } |k-j| \neq 1 \\ 1 - \frac{2\zeta^2}{\pi^2\sigma_\rho^2} & \text{if } |k-j| = 1 \end{cases} \qquad |\zeta| \leqslant \pi\sigma_\rho$$

It follows that

$$|v_\rho(\zeta)|^2 \leqslant 1 - \frac{4\zeta^2}{\pi^2\sigma_\rho^2} \sum_{k=0}^{R-1} p_\rho(k)p_\rho(k+1)\,.$$

Using (40) it is easy to check that there exists $\delta = \delta(R) > 0$ such that uniformly in $\rho \in (0, R)$

$$\sum_{k=0}^{R-1} p_\rho(k) p_\rho(k+1) \geqslant \delta\, \sigma_\rho^2 .$$

We have shown that $|v_\rho(\zeta)|^2 - 1 \leqslant -2a\zeta^2$, with $a = 2\delta/\pi^2$. The lemma then follows from the elementary inequality $x \leqslant e^{\frac{1}{2}(x^2-1)}, x \in [0,1]$ applied to $x = |v_\rho(\zeta)|$. Q.E.D.

We now start the proof of Proposition 3.2. By particle–hole duality we may restrict to densities ρ satisfying $\rho \leqslant R/2$. It is convenient to consider separately two regimes of density.

The case $\frac{R}{2} \geqslant \rho \geqslant N^{-\frac{3}{4}}$. Denote by $\mu_{N,\rho}$ the product measure $\mu_\rho^{\otimes N}$ and recall the event $\Theta_{N,\rho}$ that the sum of the η's is ρN. Set $\tilde{v}_\rho(\zeta) = v_\rho(\zeta/\sqrt{N})$. By elementary Fourier transform we have

$$2\pi\sigma_\rho\sqrt{N}\mu_{N,\rho}(\Theta_{N,\rho}) = \int \mathrm{d}\zeta\, \tilde{v}_\rho(\zeta)^N . \tag{42}$$

Here and in the rest of this proof all the integrals are over the interval $[-\pi\sigma_\rho\sqrt{N}, \pi\sigma_\rho\sqrt{N}]$. Similarly

$$\nu(n) = \frac{p_\rho(n)}{2\pi\sigma_\rho\sqrt{N}\mu_{N,\rho}(\Theta_{N,\rho})} \int \mathrm{d}\zeta\, \tilde{v}_\rho(\zeta)^{N-1} e^{i\frac{\zeta}{\sigma_\rho\sqrt{N}}\bar{n}} \tag{43}$$

$$\nu(n,m) = \frac{p_\rho(n)p_\rho(m)}{2\pi\sigma_\rho\sqrt{N}\mu_{N,\rho}(\Theta_{N,\rho})} \int \mathrm{d}\zeta\, \tilde{v}_\rho(\zeta)^{N-2} e^{i\frac{\zeta}{\sigma_\rho\sqrt{N}}[\bar{n}+\bar{m}]} \tag{44}$$

where we use the shortcut notation $\bar{n} = \xi_\rho(n) = n - \rho$, $\bar{m} = \xi_\rho(m) = m - \rho$. We can then write

$$Q(m,n) = \frac{\nu(n,m) - \nu(n)\nu(m)}{\nu(n)\nu(m)} = \frac{\mathrm{NUM}}{\mathrm{DEN}} \tag{45}$$

with

$$\begin{aligned} \mathrm{NUM} := &\int \mathrm{d}t\, \tilde{v}_\rho(\zeta)^{N-2} e^{i\frac{\zeta}{\sigma_\rho\sqrt{N}}[\bar{n}+\bar{m}]} \int \mathrm{d}\zeta'\, \tilde{v}_\rho(\zeta')^N \\ &- \int \mathrm{d}\zeta\, \tilde{v}_\rho(\zeta)^{N-1} e^{i\frac{\zeta}{\sigma_\rho\sqrt{N}}\bar{n}} \int \mathrm{d}\zeta'\, \tilde{v}_\rho(\zeta')^{N-1} e^{i\frac{\zeta'}{\sigma_\rho\sqrt{N}}\bar{m}} \end{aligned} \tag{46}$$

and

$$\mathrm{DEN} := \int \mathrm{d}\zeta\, \tilde{v}_\rho(\zeta)^{N-1} e^{i\frac{\zeta}{\sigma_\rho\sqrt{N}}\bar{n}} \int \mathrm{d}\zeta'\, \tilde{v}_\rho(\zeta')^{N-1} e^{i\frac{\zeta'}{\sigma_\rho\sqrt{N}}\bar{m}}$$

Thanks to the bound of Lemma 3.3 we have $|\tilde{v}_\rho(\zeta)|^N \leqslant e^{-a\zeta^2}$. Therefore in the integrals above only the region $|\zeta| \leqslant C \log N$ (for some large but fixed C) has to be taken care of. We then observe that there exists $\delta > 0$ such that uniformly

$$\tilde{v}_\rho(\zeta) = 1 - \frac{\zeta^2}{2N} + O(N^{-1-\delta}), \qquad |\zeta| \leqslant C \log N. \tag{47}$$

Indeed, by expanding $\tilde{v}_\rho$ around the origin the third order error term is bounded from above by $C|\zeta|^3(\sigma_\rho\sqrt{N})^{-3}\mu_\rho(|\xi_\rho|^3)$. Observing that $\mu_\rho(|\xi_\rho|^3) \leqslant C\sigma_\rho^2$ and $\sigma_\rho^2 \geqslant C^{-1}\rho$ then (47) follows from the assumption $R/2 \geqslant \rho \geqslant N^{-3/4}$. Similarly one can write $\tilde{v}_\rho(\zeta)^N = e^{-\frac{1}{2}\zeta^2} + O(N^{-\delta})$ in the range $|\zeta| \leqslant C \log N$. This gives the uniform estimates

$$\begin{aligned} I_1 &:= \int \mathrm{d}\zeta\, \tilde{v}_\rho(\zeta)^N = \sqrt{2\pi} + O(N^{-\delta}) \\ I_2 &:= \int \mathrm{d}\zeta\, \zeta^2\, \tilde{v}_\rho(\zeta)^N = \sqrt{2\pi} + O(N^{-\delta}) \\ I_3 &:= \int \mathrm{d}\zeta\, \zeta\, \tilde{v}_\rho(\zeta)^N = O(N^{-\delta}) \end{aligned}$$

From (47) we also deduce

$$\begin{aligned} \tilde{v}_\rho(\zeta)^{N-2} &= \tilde{v}_\rho(\zeta)^N \left(1 + \frac{\zeta^2}{N} + O(N^{-1-\delta})\right) \\ \tilde{v}_\rho(\zeta)^{N-1} &= \tilde{v}_\rho(\zeta)^N \left(1 + \frac{\zeta^2}{2N} + O(N^{-1-\delta})\right) \end{aligned}$$

uniformly in the region $|\zeta| \leqslant C \log N$. We then expand

$$\begin{aligned} e^{i\frac{\zeta}{\sigma_\rho\sqrt{N}}\bar{n}} &= 1 + i\frac{\zeta\bar{n}}{\sigma_\rho\sqrt{N}} + u_n(\zeta) \\ e^{i\frac{\zeta}{\sigma_\rho\sqrt{N}}[\bar{n}+\bar{m}]} &= \Big(1 + i\frac{\zeta\bar{n}}{\sigma_\rho\sqrt{N}} + u_n(\zeta)\Big)\Big(1 + i\frac{\zeta\bar{m}}{\sigma_\rho\sqrt{N}} + u_m(\zeta)\Big) \end{aligned}$$

with error terms u_n satisfying $|u_n(\zeta)| \leqslant C\zeta^2\frac{\bar{n}^2}{N\sigma_\rho^2}$. When we plug all the previous identities into (46), after all the cancellations we arrive at

$$\begin{aligned} \mathrm{NUM} &= -\frac{\bar{n}\bar{m}}{\sigma_\rho^2 N}(I_1I_2 - I_3^2) + R_1(n,m) \\ &= -2\pi\frac{\bar{n}\bar{m}}{\sigma_\rho^2 N} + R_1(n,m) + R_2(n,m) \end{aligned} \tag{48}$$

with remainder terms satisfying

$$|R_1(n,m)| \leqslant C\frac{\bar{n}^2\bar{m}^2}{(\sigma_\rho^2 N)^2} + C\frac{(|\bar{n}|\bar{m}^2 + |\bar{m}|\bar{n}^2)}{(\sigma_\rho^2 N)^{3/2}} + O(N^{-1-\delta})$$

and $|R_2(n,m)| \leqslant CN^{-1-\delta}|\bar{n}|\,|\bar{m}|/\sigma_\rho^2$. Using the bounds $\nu(\xi_\rho^2) \leqslant C\sigma_\rho^2$ and $\nu(\xi_\rho^4) \leqslant C\sigma_\rho^2$ together with $\sigma_\rho^2 \geqslant C^{-1}N^{-\frac{3}{4}}$ we see that

$$\sum_{n,m} \nu(n)\nu(m)|R_i(n,m)|^2 \leqslant O(N^{-2-\delta})\,, \qquad i=1,2\,. \tag{49}$$

On the other hand similar reasoning implies

$$\mathrm{DEN} = 2\pi + O(N^{-\delta}) \tag{50}$$

In conclusion (37) follows from (48)–(50).

The case $\rho \leqslant N^{-\frac{3}{4}}$. We first check that

$$\sum_{\substack{n,m:\, n+m\,\geqslant\, 2,\\ nm\neq 1}} \nu(n)\nu(m)\, Q(n,m)^2 = O(N^{-2-\delta}) \tag{51}$$

To prove (51) we take advantage of the very thin tails of $\nu(n)$ in the range $\rho \leqslant N^{-3/4}$. By a standard argument using Lemma 3.3 (see e.g. the proof of Proposition 3.8 in [5]), from (43)–(44) and (40) one obtains

$$\nu(n) \;\leqslant\; C p_\rho(n) \tag{52}$$

and $\nu(n,m) \leqslant Cp_\rho(n)p_\rho(m)$, where C is a uniform constant. Therefore $\nu(n) = O(\rho^n)$ and $\nu(n,m) = O(\rho^{n+m})$. In the same way, writing $\nu(m\,|\,n) := \frac{\nu(m,n)}{\nu(n)}$, we have

$$\frac{\nu(n,m)^2}{\nu(n)\nu(m)} = \nu(m\,|\,n)\nu(n\,|\,m) \leqslant C\,\rho_m^n\,\rho_n^m \leqslant C\rho^{n+m}\,,$$

where $\rho_n = \rho + (\rho - n)/(N-1) \leqslant \rho N/(N-1)$. Therefore

$$\nu(n)\nu(m)Q(n,m)^2 \leqslant C\,\rho^{n+m}$$

In particular, $\sum_{n+m\,\geqslant\,3} \nu(n)\nu(m)Q(n,m)^2 \leqslant C\rho^3 \leqslant C\,N^{-9/4}$, since $\rho \leqslant N^{-3/4}$. On the other hand $Q(0,2) = O(\rho)$ since $\nu(0,2) = \nu(2) - \nu(\eta_1 = 2, \eta_2 \geqslant 1) = \nu(0)\nu(2) + O(\rho^3)$. It follows $\nu(0)\nu(2)Q(0,2)^2 =$

$O(\rho^4)$. This completes the proof of (51). In a similar way, using $\sigma_\rho^2 = \rho + O(\rho^2)$ one checks:

$$\sum_{\substack{n,m\colon n+m \geqslant 2,\\ nm\neq 1}} \nu(n)\nu(m)\left|\frac{\xi_\rho(n)\xi_\rho(m)}{\sigma_\rho^2 N}\right|^2 = O(N^{-2-\delta}) \tag{53}$$

It remains to prove that (37) holds when n and m are restricted to $\{0,1\}$:

(54)

$$\nu(n)\nu(m)\Big[Q(n,m) + \frac{\xi_\rho(n)\xi_\rho(m)}{\sigma_\rho^2 N}\Big]^2 = O(N^{-2-\delta}), \qquad n,m \in \{0,1\}.$$

Recall that $\nu(1) = \rho + O(\rho^2)$ and $\nu(0) = 1 - \rho + O(\rho^2)$. With $\rho_n = \rho - \xi_\rho(n)/(N-1)$ we then have

$$\nu(m\,|\,n) = \begin{cases} (1-\rho_n) + O(\rho^2) & m = 0 \\ \rho_n + O(\rho^2) & m = 1 \end{cases}$$

Therefore

$$Q(m,n) = \frac{\nu(m\,|\,n) - \nu(m)}{\nu(m)} = \begin{cases} \frac{\xi_\rho(n)}{(1-\rho)N} + O(\rho^2) & m = 0 \\ -\frac{\xi_\rho(n)}{\rho N} + O(\rho) & m = 1 \end{cases} \tag{55}$$

Since $\sigma_\rho^2 = \rho + O(\rho^2)$, (55) implies (54). This completes the proof of the proposition.

§4. Zero–range processes

The zero range processes fit the general setting of section 2. Here $X = \mathbb{N}$ and the variables η_k are interpreted as occupation numbers. The apriori probability measure μ is of the form

$$p(0) = \frac{1}{Z}; \qquad p(n) = \frac{1}{Z}\prod_{i=1}^{n}\frac{1}{c(i)}, \quad n \geqslant 1$$

where $p(n) := \mu(\eta_k = n)$, c is a given positive function on $\mathbb{N}_+$ to be interpreted as the rate of escape, see below, and Z is the normalization constant. We shall make assumptions which imply in particular that $c(n) \geqslant \delta n$ for some $\delta > 0$ and all $n \geqslant 1$ so that μ is always well defined (and has all exponential moments).

The conserved quantity is the total number of particles, so $\xi(\eta_k) = \eta_k$. The complete graph dynamics is described as follows. Each site $k \in \{1, \dots, N\}$ is equipped with a Poisson clock which rings at rate 1. When site k rings we choose uniformly another site, say j. If $k \neq j$ and $\eta_k > 0$ we move one particle from k to j with rate $c(\eta_k)$. The rate is independent of the configuration η outside site k, thus justifying the name zero range. The canonical measures $\nu_{N,\rho}$ are reversible since

$$c(n)p(n)p(m) = c(m+1)p(n-1)p(m+1)\,, \qquad n \geqslant 1\,,\ m \geqslant 0.$$

The Dirichlet form is then given by (3) with

$$v_{k,\ell} f(\eta) = \sqrt{c(\eta_k)/2}\,\left[f(T_{k,\ell}\eta) - f(\eta)\right] \tag{56}$$

where

$$\left(T_{k,\ell}\eta\right)_j = \begin{cases} \eta_k - 1 & \text{if } j = k\,,\ \eta_k \geqslant 1 \\ \eta_\ell + 1 & \text{if } j = \ell\,,\ \eta_k \geqslant 1 \\ \eta_j & \text{otherwise}\,. \end{cases}$$

We make two assumptions on the rate $c(\cdot)$:

- c is globally Lipschitz: There exists $a_1 < \infty$ such that

$$\sup_n |c(n+1) - c(n)| \leqslant a_1 \tag{H1}$$

- c grows at infinity: There exists $N_0 < \infty$ and $a_2 > 0$ such that

$$c(n) \geqslant c(m) + a_2\,, \qquad n \geqslant N_0 + m \tag{H2}$$

A very special case is $c(n) = n$, so that the measure μ is Poisson. In this case the process consists of ρN independent random walks on the complete graph and therefore a uniform Poincaré inequality is trivially obtained by tensorization. (H1) and (H2) are the assumptions considered by Landim, Sethuraman and Varadhan [20] and we shall use some key preliminary results of [20] to make our proof. Since the property (MP) discussed in section 2 is immediate for the zero range process (56) one can recover the main results of [20] using (30)–(32) and the theorem below.

Theorem 4.1. *Assume* (H1) *and* (H2). *There exists $C < \infty$ such that*

$$\sup_{N \geqslant 2}\ \sup_{\rho \in \mathbb{N}/N} \gamma(N, \rho) \leqslant C \tag{57}$$

We shall prove the theorem by checking the hypothesis of Proposition 2.1. We follow as closely as possible the analysis of the previous section. However more care is required here in view of the unboundedness of the variables η_k.

Let $\gamma(N)$ be defined as in (10). We first check that $\gamma(N) < \infty$ for all N. From Lemma 3.1 and Lemma 3.2 of [20] we have that for any $f \in L^2(\nu_{N,\rho})$ with $\nu_{N,\rho}(f) = 0$

$$\frac{1}{N}\sum_{k=1}^{N}\nu_{N,\rho}\big(\nu_{N,\rho}(f\,|\,\mathcal{F}_k)^2\big) \leqslant C\Big[\mathcal{E}_{N,\rho}(f) + \frac{1}{N}\sum_{k=1}^{N}\nu\big(\mathrm{Var}_{N-1,\rho_{\eta_k}}(f)\big)\Big]$$

for some uniform constant $C < \infty$. Thus from (13) and (14) we obtain in particular

$$\gamma(N) \leqslant C\,\gamma(N-1) + C\,, \qquad N \geqslant 2\,,$$

which clearly implies that $\gamma(N)$ is finite for every N, since $\gamma(1) = 0$.

Thanks to Lemma 2.2 we reduce the proof of (SGP) to the proof of estimate (SGK). As in (35) we write

$$\langle\varphi, \mathcal{K}\varphi\rangle = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\nu(n)\nu(m)Q(n,m)\varphi(n)\varphi(m) \tag{58}$$

with the kernel Q given by (36). In the next subsection we prove (SGK)

4.1. Proof of (SGK)

As in (38) we define the exponential family $\bar{\mu}_\alpha, \alpha > 0$ and the corresponding measures μ_ρ indexed by the density $\rho > 0$. The latter is given by $\rho = \mu_\rho(\eta_1)$ and a simple computation gives $\alpha(\rho) = \mu_\rho(c(\eta_1))$. The variance σ_ρ^2 is defined as in (39). As shown in [20], Lemma 5.1, the assumptions (H1) and (H2) imply the uniform bounds

$$\delta\rho \leqslant \sigma_\rho^2 \leqslant \delta^{-1}\,\rho\,. \tag{59}$$

for some $\delta \in (0,1)$. We distinguish two regimes according to the value of the density ρ. We speak of *low density* when $\rho < 1$ and of *high density* when $\rho \geqslant 1$. Note that the choice of the critical value 1 is purely conventional. For low densities we use the same strategy as in the previous section with only small modifications. In the case of high density we rely on the uniform local central limit theorem derived in [20], Theorem 6.1.

Low density. When $\rho \leqslant 1$ the system behaves in many respects like the model with cutoff considered in the previous section. In particular when

$\rho \to 0$ we have the same estimates as in (40). We are going to prove the following analogon of Proposition 3.2.

There exists $C < \infty$ and $\delta > 0$ such that for any $N \in \mathbb{N}$ and any $\rho \leqslant 1$

$$(60) \qquad \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \nu(n)\nu(m)\Big[Q(n,m) + \frac{\xi_\rho(n)\xi_\rho(m)}{\sigma_\rho^2 N}\Big]^2 \leqslant C\,N^{-2-\delta}$$

As seen in the previous section, this bound immediately implies (SGK) in the low density region $\rho \leqslant 1$.

Let v_ρ denote the characteristic function for the random variable ξ_ρ/σ_ρ, see (41). With the observations above, the estimate (59) and the argument of Lemma 3.3 one checks that there exists $a > 0$ independent of $\rho \leqslant 1$ such that

$$(61) \qquad |v_\rho(\zeta)| \leqslant e^{-a\zeta^2}\,, \qquad \zeta \in [-\pi\sigma_\rho, \pi\sigma_\rho]\,.$$

When $N^{-\frac{3}{4}} \leqslant \rho \leqslant 1$ the proof of the proposition goes as follows. We write $Q(n,m)$ as in (45). Expanding as in (47) we have the same estimates as in (48)–(50). The only exception is that (50) now holds in the following sense: for every $T > 0$ there exists $\delta > 0$ such that uniformly in $N^{-\frac{3}{4}} \leqslant \rho \leqslant 1$

$$(62) \qquad \sup_{\substack{n,m:\\ n+m\,\leqslant\, T\log N}} \big|\,\mathrm{DEN} - 2\pi\,\big| = O(N^{-\delta})\,.$$

In this way we have obtained

$$\sum_{\substack{n,m:\\ n+m\,\leqslant\, T\log N}} \nu(n)\nu(m)\Big[Q(n,m) + \frac{\xi_\rho(n)\xi_\rho(m)}{\sigma_\rho^2 N}\Big]^2 = O(N^{-2-\delta})$$

On the other hand, since $\nu(n) \leqslant Cp_\rho(n) \leqslant Ce^{-n/C}$ uniformly in N and $\rho \leqslant 1$, we have

$$\sum_{\substack{n,m:\\ n+m>T\log N}} \nu(n)\nu(m)\Big[Q(n,m) + \frac{\xi_\rho(n)\xi_\rho(m)}{\sigma_\rho^2 N}\Big]^2 = O(N^{-3})$$

provided T is sufficiently large (but independent of ρ and N). This proves the claim in the regime $N^{-\frac{3}{4}} \leqslant \rho \leqslant 1$.

When $\rho \leqslant N^{-\frac{3}{4}}$ we use exactly the same argument as in (51) and (55) which applies without modifications. This ends the proof of (SGK) in the case $\rho \leqslant 1$.

High density. Here the strategy above has to be modified since the Gaussian bound (61) does not hold anymore and one has to control every estimate uniformly as $\rho \to \infty$. The main tool is the uniform Edgeworth expansion derived in [20]. For $M \in \mathbb{N}$, we define

$$W_{M,\rho}(t) = \mu_{M,\rho}\Big(\sum_{k=1}^{M}(\eta_k - \rho) = -t\Big)\,. \tag{63}$$

In (63) and all expressions below when we write $W_{M,\rho}(t)$ we assume that $\rho M - t$ is a nonnegative integer. The following lemma is a straightforward consequence of [20], Theorem 6.1, part (b).

Lemma 4.2. *For any $\kappa < 1/6$ there exists $C < \infty$ such that for all $M \geqslant 1$ and all $\rho \geqslant 1$*

$$\sup_{|t| \leqslant \sigma_\rho M^\kappa} \Big| \sigma_\rho \sqrt{M} W_{M,\rho}(t) - \frac{e^{-\frac{t^2}{2\sigma_\rho^2 M}}}{\sqrt{2\pi}} \Big(1 + \frac{A_\rho t}{\sigma_\rho M} + \frac{B_\rho}{M}\Big) \Big| \leqslant C M^{-\frac{3}{2}}\,,$$

where A_ρ and B_ρ are real numbers with $\sup_{\rho \geqslant 1}(|A_\rho| + |B_\rho|) < \infty$.

We can express the kernel $Q(n,m)$ in terms of the probabilities (63):

$$Q(n,m) = \frac{W_{N-2,\rho}(\bar{n} + \bar{m}) W_{N,\rho}(0) - W_{N-1,\rho}(\bar{n}) W_{N-1,\rho}(\bar{m})}{W_{N-1,\rho}(\bar{n}) W_{N-1,\rho}(\bar{m})}\,, \tag{64}$$

where $\bar{n} = n - \rho$, $\bar{m} = m - \rho$. We fix $\kappa = 1/10$ and define the sets

$$\mathcal{T}_{N,\rho} = \{(n,m) \in \mathbb{N}^2 \ : \ |\bar{n}| + |\bar{m}| \leqslant \sigma_\rho N^\kappa\}$$

Let us agree to denote by $\varepsilon(N)$ anything which vanishes at least as $O(N^{-\frac{3}{2}})$ uniformly in the sets $\mathcal{T}_{N,\rho}$, $\rho \geqslant 1$. Thus the result of Lemma 4.2, with $M = N - 1$ and $t = \bar{n}$, can be written as

$$\sigma_\rho \sqrt{N-1} W_{N-1,\rho}(\bar{n}) = \frac{e^{-\frac{\bar{n}^2}{2\sigma_\rho^2 (N-1)}}}{\sqrt{2\pi}} \Big(1 + \frac{A_\rho \bar{n}}{\sigma_\rho (N-1)} + \frac{B_\rho}{N-1}\Big) + \varepsilon(N) \tag{65}$$

We use now (65) to write

$$\begin{aligned} &2\pi\sigma_\rho^2 (N-1) W_{N-1,\rho}(\bar{n}) W_{N-1,\rho}(\bar{m}) \\ &= e^{-\frac{\bar{n}^2 + \bar{m}^2}{2\sigma_\rho^2 (N-1)}} \Big(1 + \frac{A_\rho \bar{n}}{\sigma_\rho (N-1)} + \frac{B_\rho}{N-1}\Big)\Big(1 + \frac{A_\rho \bar{m}}{\sigma_\rho (N-1)} + \frac{B_\rho}{N-1}\Big) + \varepsilon(N) \\ &\qquad = e^{-\frac{\bar{n}^2 + \bar{m}^2}{2\sigma_\rho^2 N}} \Big(1 + \frac{A_\rho (\bar{n} + \bar{m})}{\sigma_\rho N} + \frac{2B_\rho}{N}\Big) + \varepsilon(N)\,. \end{aligned}$$

Furthermore, writing $q(N) = (N-1)/\sqrt{N(N-2)} = 1 + O(N^{-2})$, from Lemma 4.2, with $M = N-2$ and $t = \bar{n} + \bar{m}$, one has

$$
\begin{aligned}
&2\pi\sigma_\rho^2(N-1)W_{N-2,\rho}(\bar{n}+\bar{m})W_{N,\rho}(0) \\
&\quad = q(N)\, e^{-\frac{(\bar{n}+\bar{m})^2}{2\sigma_\rho^2(N-2)}}\Big(1+\frac{A_\rho(\bar{n}+\bar{m})}{\sigma_\rho(N-2)}+\frac{B_\rho}{N-2}\Big)\Big(1+\frac{B_\rho}{N}\Big) + \varepsilon(N) \\
&\quad = e^{-\frac{(\bar{n}+\bar{m})^2}{2\sigma_\rho^2 N}}\Big(1+\frac{A_\rho(\bar{n}+\bar{m})}{\sigma_\rho N}+\frac{2B_\rho}{N}\Big) + \varepsilon(N) \\
&\quad = e^{-\frac{\bar{n}^2+\bar{m}^2}{2\sigma_\rho^2 N}}\Big(1-\frac{\bar{n}\bar{m}}{\sigma_\rho^2 N}\Big)\Big(1+\frac{A_\rho(\bar{n}+\bar{m})}{\sigma_\rho N}+\frac{2B_\rho}{N}\Big) + \varepsilon(N)\,.
\end{aligned}
$$

Inserting in (64) we have obtained

$$
\sup_{\rho \geqslant 1}\ \sup_{(n,m)\in \mathcal{T}_{N,\rho}} \Big| Q(n,m) + \frac{\xi_\rho(n)\xi_\rho(m)}{\sigma_\rho^2 N}\Big| = O\big(N^{-\frac{3}{2}}\big) \tag{66}
$$

To conclude the proof of (SGK) in the case $\rho \geqslant 1$ it is therefore sufficient to prove

$$
\sum_{(n,m)\notin \mathcal{T}_{N,\rho}} \nu(n,m)|\varphi(n)|\,|\varphi(m)| \leqslant C\,N^{-\frac{3}{2}}\,\langle\varphi,\varphi\rangle\,, \tag{67}
$$

for any $\varphi \in \mathcal{H}$, uniformly over $\rho \geqslant 1$.

We first claim that for any $k \in \mathbb{N}$ there exists $C_k < \infty$ such that

$$
\nu\big(|\eta_1 - \rho_m| \geqslant T\sigma_\rho \mid \eta_2 = m\big) \leqslant C_k\, T^{-2k} \tag{68}
$$

for any $0 \leqslant m \leqslant \rho N/2$ and any $T > 0$, with $\rho_m = \rho + (\rho - m)/(N-1)$.

To prove (68) recall that there exists $C < \infty$ independent of ρ such that for every $n \in \mathbb{N}$ we have $\nu(n) \leqslant Cp_\rho(n)$ (this is a consequence of Lemma 4.2 if $\rho \geqslant 1$, otherwise see (52)). Therefore $\nu(n\,|\,m) \leqslant Cp_{\rho_m}(n)$ and

$$
\nu\big(|\eta_1 - \rho_m| \geqslant T\sigma_\rho \mid \eta_2 = m\big) \leqslant C\, m_{2k,\rho_m}\,(T\sigma_\rho)^{-2k}\,,
$$

where $m_{2k,\rho} := \mu_\rho\big[(\eta_1-\rho)^{2k}\big]$. From [20], Lemma 5.2, we know that

$$
M_k := \sup_{\rho \geqslant 1/2} \frac{m_{2k,\rho}}{\sigma_\rho^{2k}} < \infty\,,
$$

for every $k \in \mathbb{N}$. Since $m \leqslant \rho N/2$ implies $\rho_m \geqslant \rho/2 \geqslant 1/2$, the above yields

$$
\nu\big(|\eta_1 - \rho_m| \geqslant T\sigma_\rho \,|\, \eta_2 = m\big) \leqslant C\, M_k\,(T\sigma_\rho/\sigma_{\rho_m})^{-2k}\,.
$$

Now (68) follows since by (59) we have $\sigma_\rho/\sigma_{\rho_m} \geqslant \delta\sqrt{\rho/\rho_m}$ and, using $m \geqslant 0$, $\rho \geqslant \rho_m(N-1)/N$.

Once (68) is established we may prove (67) as follows. Observe that for any $N \geqslant 3$, $(n,m) \notin \mathcal{T}_{N,\rho}$ implies either $|n-\rho_m| \geqslant T\sigma_\rho$ or $|m-\rho_n| \geqslant T\sigma_\rho$, with $T = N^\kappa/4$. By (68) and the Schwarz' inequality we estimate, uniformly in $\rho \geqslant 1$:

$$\sum_{\substack{(n,m)\notin\mathcal{T}_{N,\rho}\\ n+m\,\leqslant\,\rho N/2}} \nu(n,m)|\varphi(n)|\,|\varphi(m)|$$

$$\leqslant 2 \sum_{m\,\leqslant\,\rho N/2} \nu(m)|\varphi(m)| \sum_{n:|n-\rho_m|\,\geqslant\,T\sigma_\rho} \nu(n\,|\,m)|\varphi(n)|$$

$$\leqslant \sqrt{C_k}\,T^{-k}\sum_m \nu(m)|\varphi(m)|\Big(\sum_n \nu(n\,|\,m)|\varphi(n)|^2\Big)^{\frac{1}{2}} \leqslant \sqrt{C_k}\,T^{-k}\,\langle\varphi,\varphi\rangle\,.$$

Since $T = N^\kappa/4$ we choose k such that $k\kappa > 3/2$ and (67) is proven under the additional requirement $n+m \leqslant \rho N/2$.

It remains to prove

$$\sum_{\substack{n,m:\\ n+m>\rho N/2}} \nu(n,m)|\varphi(n)|\,|\varphi(m)| \leqslant N^{-\frac{3}{2}}\langle\varphi,\varphi\rangle \tag{69}$$

This in turn follows from Schwarz' inequality and the uniform bound

$$\sum_{\substack{n,m:\\ n+m>\rho N/2}} \nu(n\,|\,m)\nu(m\,|\,n) \leqslant N^{-3}\,. \tag{70}$$

To establish (70) we write $\nu(n\,|\,m)\nu(m\,|\,n) \leqslant C\,p_{\rho_m}(n)p_{\rho_n}(m)$ and use the simple bounds $p_\rho(n) \leqslant e^{-n/C}$ valid for $n \geqslant C\rho$, where C is a sufficiently large constant. Recalling that $\rho \geqslant \rho_m(N-1)/N$, $m \geqslant 0$ this immediately implies (70) and therefore (69). This ends the proof of (SGK) in the high density region $\rho \geqslant 1$.

§5. Colored exclusion

In this section we consider a model with different kinds of particles, or particles of different colors, with the constraint that each site is occupied at most by a single particle and the number of particles of each kind is conserved. We set $X = \{0,1,\ldots,R\}$ with some positive integer R. If $\eta_k = 0$ we say that site k is empty while if $\eta_k = m$, $m \in \{1,\ldots,R\}$ we think of site k as being occupied by a particle with color m. The conservation laws are expressed in terms of the functions

$$\xi^m(\eta_k) = \mathbf{1}_{\{m\}}(\eta_k) = \begin{cases} 1 & \eta_k = m \\ 0 & \eta_k \neq m \end{cases} \tag{71}$$

so that the multicanonical measure $\nu_{N,\bar\rho}$ in (28) is obtained by conditioning on the event

$$\Theta_{N,\bar\rho} = \{\eta \in \Omega_N \,:\, \sum_{k=1}^{N} \xi^m(\eta_k) = \rho_m N\,,\ m = 1,\dots,R\}\,,$$

with $\bar\rho = (\rho_1,\dots,\rho_R)$ an assigned density vector with $\sum_{m=1}^{R} \rho_m \leqslant 1$. We say that $\bar\rho$ is trivial if $\rho_m \in \{0,1\}$ for every $m \in \{1,\dots,R\}$. The dynamics is given by random transpositions so that the Dirichlet form is

$$\mathcal{E}_{N,\bar\rho}(f) = \frac{1}{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} \nu_{N,\bar\rho}\big[(f \circ T_{k,\ell} - f)^2\big] \tag{72}$$

where $f \in L^2(\nu_{N,\bar\rho})$ and

$$\left(T_{k,\ell}\eta\right)_j = \begin{cases} \eta_k & \text{if } j = \ell \\ \eta_\ell & \text{if } j = k \\ \eta_j & \text{otherwise}\,. \end{cases}$$

This and related random transposition or card–shuffling models have been studied in great detail by Diaconis and Shashahani [11] with more elaborate techniques. The result we prove below is rather simple but it illustrates well the use of the general arguments outlined in section 2. Note that when $R = 1$ we have the usual exclusion process on the complete graph, sometimes called the Bernoulli–Laplace model. When $R = 2$ the model was studied by Quastel, [22].

Let $\gamma(N,\bar\rho)$ be the Poincaré constant associated to the couple $(N,\bar\rho)$, as in (4). Note that $\gamma(N,\bar\rho) = 0$ when $\bar\rho$ is trivial. Let $\rho^* = \rho^*(N)$ be the density vector corresponding to one particle only: $\rho_1^* = 1/N$ and $\rho_m^* = 0$, $m = 2,\dots,R$. When $\bar\rho = \rho^*$ we have a (rate 2) random walk on the complete graph and a direct computation shows that $\mathcal{E}_{N,\rho^*}(f) = 4\mathrm{Var}_{N,\rho^*}(f)$ for every f. Therefore $\gamma(N,\rho^*) = 1/4$.

Theorem 5.1. *For any $R \in \mathbb{Z}_+$, $N \geqslant 2$ and any density $\bar\rho$:*

$$\gamma(N,\bar\rho) \leqslant \gamma(N,\rho^*) = \frac{1}{4} \tag{73}$$

Proof. We shall use the notation (34). We write $\rho_0 = 1 - \sum_{m=1}^{R} \rho_m$ and $\xi_{\rho_m}^m(n) = \mathbf{1}_{\{m\}}(n) - \rho_m$ for every $m = 0, 1, \ldots, R$. Notice that by symmetry we have

$$\nu(m) = \rho_m\,, \qquad m = 0, 1, \ldots, R\,.$$

Therefore

$$\nu(m \,|\, n) = \frac{\nu(m,n)}{\nu(n)} = \frac{\rho_m N - \xi^m(n)}{N-1} = \rho_m - \frac{\xi_{\rho_m}^m(n)}{N-1}\,. \tag{74}$$

Take $\varphi \in \mathcal{H}_0$ and write

$$\begin{aligned} \langle \varphi, \mathcal{K}\varphi \rangle &= \sum_{n=0}^{R} \sum_{m=0}^{R} \nu(n) \big[\nu(m \,|\, n) - \nu(m)\big] \varphi(n)\varphi(m) \\ &= -\frac{1}{N-1} \sum_{m=0}^{R} \varphi(m) \langle \xi_{\rho_m}^m, \varphi \rangle \end{aligned} \tag{75}$$

From this we see that whenever $\varphi \in \mathcal{H}_0$ is orthogonal to all $\xi_{\rho_m}^m$ then $\langle \varphi, \mathcal{K}\varphi \rangle = 0$. From the analysis in Lemma 2.2 it follows that

$$\nu_{N,\bar\rho}\big(f(1-\mathcal{P})f\big) \geqslant \frac{N-2}{N-1} \nu_{N,\bar\rho}(f^2) \tag{76}$$

for every $f \in L^2(\nu_{N,\bar\rho})$ with $\nu_{N,\bar\rho}(f) = 0$ and any $N \geqslant 3$. Thus if $\gamma(N)$ denotes supremum of $\gamma(N, \bar\rho)$ over all possible values of $\bar\rho$, the argument of Proposition 2.1 gives $\gamma(N) \leqslant \gamma(2)$ for every $N \geqslant 3$. The theorem then follows since $\gamma(2) = \gamma(2, \rho^*) = 1/4$. Q.E.D.

§6. Anisotropic exclusion processes

Here we review recent results obtained in collaboration with F. Martinelli, [5, 6]. The model can be described in the general framework of section 2. We set $X = \{0,1\}^H$ where H is a positive integer to be interpreted as the height of the system. The measure μ is itself a product of Bernoulli measures

$$\begin{aligned} \mu &= \otimes_{h=1}^{H} \mu_h\,, & \mu_h &= \mathrm{Be}(p_h) \\ p_h &:= \frac{q^{2h}}{1+q^{2h}}\,, & q &\in (0,1)\,. \end{aligned} \tag{77}$$

Then $\Omega_N = X^N$ and a configuration $\eta = \{\eta_i\}_{i=1}^N$ is given in terms of its components $\eta_i = \{\alpha_{(i,h)}\}_{h=1}^H$, with $\alpha_{i,h} \in \{0,1\}$ interpreted as the presence or absence of a particle at site (i, h). The conservation law is given by

$$\xi(\eta_i) = \sum_{h=1}^{H} \alpha_{(i,h)} .$$

The canonical measures $\nu_{N,\rho} = \nu_{H,N,\rho}$ are defined as usual by (2) for every fixed value of H. We may think of identical particles placed at the sites of a two–dimensional cylindrical region $\Lambda = \{1, \dots, N\} \times \{1, \dots, H\}$. Each site can be occupied by at most one particle and the total number of particles is fixed. Since $q < 1$ there is anisotropy in the vertical axis and particles prefer to be at the bottom of Λ. The choice of the model (77) is motivated by interesting connections with anisotropic quantum spin chains, see [1, 5, 6, 17, 18] and references therein.

The dynamics can be described as follows. At each site of Λ we have an independent rate 1 Poisson clock. Suppose site (i, h) rings. If $h = H$ we do nothing. If $h < H$ we choose at random one of the sites $(j, h+1)$, $j = 1, \dots, N$. The occupation variables $\alpha_{(i,h)}$ and $\alpha_{(j,h+1)}$ are then exchanged with rate

$$c_{(i,h);(j,h+1)}(\alpha) = q^{\alpha_{(i,h)} - \alpha_{(j,h+1)}} . \tag{78}$$

That is if a particle is moving upwards the rate is q whereas if it is moving downwards the rate is q^{-1}. We thus obtain a process described by the Dirichlet form (3) with the exchange operators, for $i \neq j$

$$v_{i,j} f(\alpha) = \left(\frac{1}{2} \sum_{h=1}^{H-1} c_{(i,h);(j,h+1)}(\alpha) \left[f(\alpha^{(i,h);(j,h+1)}) - f(\alpha) \right]^2 \right)^{\frac{1}{2}} , \tag{79}$$

and $v_{i,i} f = 0$, where we write $\alpha^{(i,h);(j,h+1)}$ for the configuration in which the values of α at (i, h) and $(j, h+1)$ have been exchanged. Notice that the process is local in the vertical direction while it is nonlocal in the horizontal direction. One of the main results of [6] is that for every $q \in (0, 1)$ the relaxation time is bounded, uniformly in H, in N and in the number of particles.

Let us recall the definition of the Poincaré constant (4). In order to keep track of the dependence on H we write here $\gamma(H, N, \rho)$ instead of $\gamma(N, \rho)$.

Theorem 6.1. *For every $q \in (0, 1)$ there exists $C < \infty$ such that*

$$\sup_{N \geqslant 2} \sup_{H \geqslant 2} \sup_{\rho} \gamma(H, N, \rho) \leqslant C \tag{80}$$

The proof of Theorem 6.1 has been obtained by applying the arguments of Proposition 2.1 and Lemma 2.2. The crucial step in the proof of property (SGK) is a result analogous to Proposition 3.2. We refer to [6] for more details.

Some of the applications of Theorem 6.1, especially those to quantum Heisenberg models, are linked to the restriction of the process to horizontal sums of the basic variables $\alpha_{i,h}$ given by

$$\omega_h = \sum_{i=1}^{N} \alpha_{(i,h)} \,, \qquad h = 1, \dots, H$$

In view of the symmetries of the Dirichlet form $\mathcal{E}_{N,\rho}$ defined by (79) it is not hard to see that the restriction to the variables $\{\omega_h\}$ is again a Markov process. Indeed, the latter can be described as follows. Assign to each row $h = 1, \dots, H-1$ two independent exponentially distributed times (with mean 1), τ_-^h and τ_+^h. When τ_+^h rings the configuration ω is updated with rate $r_{+,h}(\omega) := q^{-1}(N - \omega_h)\omega_{h+1}/N$ to the configuration $\omega^{+,h}$ in which ω_h is increased by 1 and ω_{h+1} is decreased by 1 (while the rest is unchanged). When τ_-^h rings we do the reverse transition ($\omega \to \omega^{-,h}$: ω_h is decreased and ω_{h+1} increased) with rate $r_{-,h}(\omega) := q(N - \omega_{h+1})\omega_h/N$. We can write the Dirichlet form of this process as

(81)
$$\frac{1}{2}\sum_{h=1}^{H-1} \tilde{\nu}\left(r_{+,h}(\omega)\big[f(\omega^{+,h}) - f(\omega)\big]^2 + r_{-,h}(\omega)\big[f(\omega^{-,h}) - f(\omega)\big]^2\right)$$

where $\tilde{\nu}$ stands for the marginal of $\nu_{H,N,\rho}$ on the variables ω. A simple computation gives the probability $\tilde{\nu}(\omega)$ of a single ω compatible with the global constraint $\sum_h \omega_h = \rho N$:

$$\tilde{\nu}(\omega) = \frac{1}{\tilde{Z}} \prod_{h=1}^{H} \binom{N}{\omega_h} q^{2h\omega_h} \,. \tag{82}$$

The process (81) can be interpreted as describing relaxation of a nonnegative profile $\{\omega_h\}_{h=1}^H$ subject to a fixed area constraint. In view of the anisotropy the profile is strongly localized under the measure $\tilde{\nu}$, i.e. $\omega_h \approx N$ for heights h below ρ and $\omega_h \approx 0$ above ρ with high probability. By Theorem 6.1 relaxation to equilibrium in $L^2(\tilde{\nu})$ is exponentially fast uniformly in ρ.

In the case $N = 2$ the process (81) admits another interesting interpretation as a model for diffusion limited chemical reactions, see [1] and references therein. Namely describe the state $\omega_h = 2$ as the presence at h of a particle of type A, $\omega_h = 0$ as a particle of type B and $\omega_h = 1$ as the absence of particles. If n_A, n_B denote the size of the two populations we see that the difference $n_A - n_B$ is conserved and we have a model for asymmetric diffusion with creation and annihilation of the two species. Particles of type A have a constant drift towards the bottom while particles of type B have the same drift towards the top. Nearest neighbour pairs can produce the reaction $A + B \longrightarrow$ *inert* and the reverse reaction *inert* $\longrightarrow A + B$ with the appropriate rates. While Theorem 6.1 implies immediately a uniform lower bound on the spectral gap for this process, a direct proof of the result for the two-particle model seemed difficult to us.

§7. Ginzburg-Landau processes

Here we discuss a recent result ([4]) for the Ginzburg–Landau process. The model is obtained from the general setting in section 2 with $X = \mathbb{R}$ and $\xi(\eta_k) = \eta_k$. The single site probability distribution is of the form

$$\mu(\mathrm{d}\eta) = \frac{e^{-V(\eta)}}{Z}\,\mathrm{d}\eta\,, \tag{83}$$

where $V : \mathbb{R} \longrightarrow \mathbb{R}$ is a given function with $Z = \int e^{-V(\eta)}\mathrm{d}\eta < \infty$. Precise assumptions on V are specified below. The resulting canonical measure $\nu_{N,\rho}$ on the hyperplane $\sum_{k=1}^{N} \eta_k = \rho N$ is given by (2), for all $\rho \in \mathbb{R}$. We consider the process defined by the symmetric Dirichlet form $\mathcal{E}_{N,\rho}$ given in (3) with the choice

$$v_{k,\ell} f = \partial_k f - \partial_\ell f\,, \tag{84}$$

where $\partial_k f$ is the partial derivative of f along the k-th coordinate η_k. This yields an ergodic diffusion process on every ρ– hyperplane with reversible invariant measure $\nu_{N,\rho}$ given by (2). In the definition (4) of the Poincaré constant $\gamma(N, \rho)$ the supremum is taken over all smooth functions $f : \mathbb{R}^N \longrightarrow \mathbb{R}$.

The main result of [4] says that a uniform Poincaré inequality holds whenever V is of the form $V = \varphi + \psi$ with ψ a smooth bounded function and φ a strictly convex function satisfying some mild growth condition at infinity. To describe the latter we define the class Φ of functions $\varphi \in C^2(\mathbb{R}, \mathbb{R})$ with second derivative φ'' satisfying

- *Strict convexity:* There exists $\delta > 0$ such that $\varphi'' \geqslant \delta$.

- *Polynomial growth at infinity:* There exist constants $\beta_-, \beta_+ \in [0, \infty)$ and a constant $C \in [1, \infty)$ such that

$$\frac{1}{C} \leqslant \liminf_{x \to \infty} \frac{\varphi''(\pm x)}{x^{\beta_\pm}} \leqslant \limsup_{x \to \infty} \frac{\varphi''(\pm x)}{x^{\beta_\pm}} \leqslant C\,. \tag{85}$$

Clearly, any strictly convex polynomial belongs to Φ. The perturbation will be taken from the class Ψ of functions $\psi \in C^2(\mathbb{R}, \mathbb{R})$ such that $|\psi|_\infty < \infty$, $|\psi'|_\infty < \infty$ and $|\psi''|_\infty < \infty$.

Theorem 7.1. *Assume V is of the form $V = \varphi + \psi$ with $\varphi \in \Phi$ and $\psi \in \Psi$. Then*

$$\sup_{N \in \mathbb{N}} \sup_{\rho \in \mathbb{R}} \gamma(N, \rho) < \infty\,. \tag{86}$$

An immediate corollary of Theorem 7.1 is the uniform diffusive bound for the local dynamics (29). This follows from property (MP) and (30)–(32). Diffusive bounds for the spectral gap of Ginzburg–Landau processes are a key ingredient in the proof of hydrodynamic limits for the nongradient system considered by Varadhan [24]. When there is no perturbation ($\psi = 0$), Theorem 7.1 (without the additional requirement (85)) becomes an immediate consequence of the Brascamp–Lieb inequality [2], see [4]. Since perturbative arguments are very sensitive to the increasing number of dimensions, the case of nonconvex potentials is much more involved. Recently the uniform diffusive estimate has been obtained by Landim, Panizo, Yau [19] in the case $V(x) = ax^2 + \psi(x)$, $a > 0$ and ψ bounded. The results of [19] have been later generalized slightly by Chafai [9] . The proofs of both [19] and [9] are based on the martingale approach ([21]) and the method is sufficiently robust to yield the stronger logarithmic Sobolev inequality. These techniques seem to fail however in the case of non quadratic potentials - thus ruling out natural problems such as quartic potentials.

The proof of Theorem 7.1 is based on the general strategy outlined in Proposition 2.1 and Lemma 2.2. The delicate part of the work is to establish the bound required in condition (SGK). Formally the situation is similar to that encountered in previous sections, but here $\mathcal{K}$ is an integral operator and the technique has to be modified slightly. Moreover, contrary to the case of zero range processes discussed in section 4, here the variance σ_ρ^2 of the grand–canonical measures

$$\mu_\rho(\mathrm{d}x) = \frac{e^{-V(x) - \lambda_\rho x}}{Z_\rho}\, \mathrm{d}x \tag{87}$$

vanishes as $\rho \to \pm\infty$ as soon as φ'' is unbounded. In the above formula λ_ρ is determined as usual by the condition that $(Z_\rho)^{-1} \int x e^{-V(x)-\lambda_\rho x} \mathrm{d}x = \rho$. The technical hypothesis (85) is mainly used to control the speed of decay of σ_ρ^2. Using a uniform local central limit theorem for the measures (87) we prove in [4], Theorem 3.1, that there exists $C < \infty$ independent of ρ and N such that for every $f \in \mathcal{H}_0$ satisfying $\langle f, \xi_\rho \rangle = 0$ one has

$$|\langle f, \mathcal{K} f \rangle| \leqslant C N^{-\frac{3}{2}} \langle f, f \rangle . \tag{88}$$

§8. Exclusion with site–disorder

Here we consider the following non–homogeneous model. The single state space is $X = \{0,1\}$ and the conservation law is $\xi(\eta_k) = \eta_k$, interpreted as the presence or absence of a particle at k. In contrast to previous models here the measure μ is site–dependent. We choose for every $k \in \{1, \dots, N\}$, $N \in \mathbb{N}$, the Bernoulli measures $\mu_k = \mathrm{Be}(\omega_k)$:

$$\mu_k(\eta_k = 1) = \omega_k \,, \qquad \omega_k \in [\delta, 1-\delta] \,, \quad k = 1, \dots, N \tag{89}$$

Here $\delta \in (0, 1/2]$ is fixed and $\omega \in [\delta, 1-\delta]^N$ can be interpreted as a realization of a random field, as in [12, 23]. However, we shall not use any probabilistic structure behind the variables ω and our results will all be uniform in $\omega \in [\delta, 1-\delta]^N$. For every such ω, every $\rho \in [0,1]$, we define the (quenched) canonical measure

$$\nu_{N,\rho} = \bigotimes_{k=1}^{N} \mu_k \Big(\,\cdot\, \mid \sum_{\ell=1}^{N} \eta_\ell = \rho N \Big) . \tag{90}$$

The Dirichlet form of the complete graph dynamics is written as in (3) with the choice

$$v_{k,\ell} f = \sqrt{c_{k,\ell}(\eta)} \, \big[f(\eta^{k,\ell}) - f(\eta) \big] \,,$$

where as usual $\eta^{k,\ell}$ denotes the configuration where η_k and η_ℓ have been exchanged, and $c_{k,\ell}$ denotes the associated transition rate. A possible choice of the rates is e.g.

$$c_{k,\ell}(\eta) = \begin{cases} \omega_k(1-\omega_\ell) & (\eta_k, \eta_\ell) = (0,1) \\ \omega_\ell(1-\omega_k) & (\eta_k, \eta_\ell) = (1,0) \end{cases}$$

The result below applies to any choice of rates provided these are uniformly bounded from above and away from zero.

For every fixed ω we call $\gamma^\omega(N, \rho)$ the corresponding Poincaré constant as in (4).

Theorem 8.1. *For every $\delta \in (0, 1/2]$ there exists $C < \infty$ such that*

$$\sup_{N \geqslant 2} \sup_{\omega \in [\delta, 1-\delta]^N} \sup_{\rho \in (0,1)} \gamma^{\omega}(N, \rho) \leqslant C\,. \tag{91}$$

Theorem 8.1 is a useful tool in the proof of hydrodynamic limit for the site–disordered simple exclusion process, [12, 23]. One can check that the model described above satisfies the moving particle lemma (MP) of section 2. A little care is required here because of the inhomogeneous medium. We refer to Lemma 3.1 in [23] for details. Thus an immediate corollary of Theorem 8.1 is the diffusive bound on the spectral gap of the local dynamics, see (30)–(32).

8.1. Proof of Theorem 8.1

We use the iteration outlined in Proposition 2.1. From a comparison with the homogeneous case $\omega_k \equiv$ const. we see that $\sup_\omega \sup_\rho \gamma^\omega(N, \rho) \leqslant C^N$ for some $C < \infty$. This guarantees that the first hypothesis of the proposition is satisfied.

If $\mathcal{P}$ denotes the operator introduced in (5) we need to show that (SGP) holds, i.e. that for every $f \in L^2(\nu_{N,\rho})$ with $\nu_{N,\rho}(f) = 0$

$$\nu_{N,\rho}(f(1-\mathcal{P})f) \geqslant \frac{N-2}{N-1}\big[1 - CN^{-1-\epsilon}\big]\nu_{N,\rho}(f^2) \tag{92}$$

with independent constants $\epsilon > 0$, $C < \infty$. As seen in section 2 (see the proof of Lemma 2.2) it is sufficient to prove (92) for functions f of the form $f(\eta) = \sum_{k=1}^{N} g_k(\eta_k)$ with $g_k : X \to \mathbb{R}$ a mean–zero function. Since here $X = \{0, 1\}$, we must have $g_k = \alpha_k(\eta_k - \rho_k)$, $\rho_k := \nu_{N,\rho}(\eta_k)$, for some $\alpha_k \in \mathbb{R}$. That is, we shall prove (92) for functions of the form

$$f(\eta) = \sum_{k=1}^{N} \alpha_k \bar{\eta}_k\,, \qquad \alpha \in \mathbb{R}^N \tag{93}$$

with $\bar{\eta}_k := \eta_k - \rho_k$. We take f as in (93) and compute

$$\nu_{N,\rho}(f^2) = \sum_{k,\ell} \alpha_k \alpha_\ell \nu_{N,\rho}(\bar{\eta}_k \bar{\eta}_\ell) = \langle \tilde{\alpha}, Q\tilde{\alpha} \rangle \tag{94}$$

where we use the notation

$$Q_{k,\ell} := \frac{\nu_{N,\rho}(\bar{\eta}_k \bar{\eta}_\ell)}{\gamma_k \gamma_\ell}\,, \qquad \tilde{\alpha}_k := \gamma_k \alpha_k\,, \qquad \gamma_k^2 := \nu_{N,\rho}(\bar{\eta}_k^2) = \rho_k(1-\rho_k)$$

and $\langle v,w\rangle := \sum_{k=1}^{N} v_k w_k$ for the scalar product in $\mathbb{R}^N$. Observing that

$$\nu_{N,\rho}(\bar\eta_k \mid \eta_j) = \frac{\nu_{N,\rho}(\bar\eta_k\bar\eta_j)}{\gamma_j^2}\,\bar\eta_j$$

one obtains in a similar way

$$\nu_{N,\rho}\big(f\mathcal{P}f\big) = \frac{1}{N}\langle\tilde\alpha, Q^2\tilde\alpha\rangle\,.$$

Q is a non–negative matrix. Setting $\hat\alpha := Q^{\frac12}\tilde\alpha$ we have

$$\nu_{N,\rho}\big(f(1-\mathcal{P})f\big) = \langle\hat\alpha,(1-\frac{Q}{N})\hat\alpha\rangle\,,\qquad \nu_{N,\rho}\big(f^2\big) = \langle\hat\alpha,\hat\alpha\rangle\,. \tag{95}$$

We write now $\Gamma := 1-Q$, so that

$$\Gamma_{k,\ell} = \begin{cases} -\frac{\nu_{N,\rho}(\bar\eta_k\bar\eta_\ell)}{\gamma_k\gamma_\ell} & k\neq\ell \\ 0 & k=\ell \end{cases}$$

Then (95) reads

$$\nu_{N,\rho}\big(f(1-\mathcal{P})f\big) = \frac{N-1}{N}\,\langle\hat\alpha,(1+\frac{\Gamma}{N-1})\hat\alpha\rangle\,.$$

By (95), the claim (92) follows if we can prove

$$\Gamma \geqslant -CN^{-\epsilon}\,. \tag{96}$$

This in turn will follow from the next lemma.

Lemma 8.2. *There exists $C<\infty$, $\epsilon>0$ such that for all ω, N, ρ and $k\neq\ell$*

$$\Big|\Gamma_{k,\ell} - \frac{\beta_k\beta_\ell}{N}\Big| \leqslant CN^{-1-\epsilon} \tag{97}$$

with non–negative numbers $\beta_k = \beta_k(\omega,N,\rho)$, $k=1,\dots,N$ satisfying $\beta_k\leqslant C$ uniformly.

Assuming (97) we conclude

$$\begin{aligned}\langle v,\Gamma v\rangle &= \sum_k\sum_{\ell\neq k} v_k v_\ell\Big(\frac{\beta_k\beta_\ell}{N} + O(N^{-1-\epsilon})\Big)\\ &\geqslant -\frac{1}{N}\sum_k \beta_k^2 v_k^2 - CN^{-\epsilon}\sum_k v_k^2 \geqslant -C'N^{-\epsilon}\langle v,v\rangle\end{aligned}$$

with a constant $C'<\infty$. This gives (96).

8.2. Proof of Lemma 8.2

We start with some preliminaries. Let $p_{k,\rho}$ be the grand–canonical single site probabilities

$$p_{k,\rho} := \frac{\omega_k e^{\lambda}}{\omega_k e^{\lambda} + 1 - \omega_k}$$

where $\lambda = \lambda^{\omega}_{N,\rho}$ is a real number such that $\sum_{k=1}^{N} p_{k,\rho} = \rho N$. We set $\mu_{k,\rho} := \mathrm{Be}(p_{k,\rho})$ and call $\mu_{N,\rho} = \otimes_{k=1}^{N} \mu_{k,\rho}$ the corresponding grand–canonical measure. We also use the notations

$$\hat{\eta}_k = \eta_k - p_{k,\rho}\,, \quad \sigma^2_{k,\rho} = p_{k,\rho}(1 - p_{k,\rho})\,, \quad \sigma^2_{\rho} = \frac{1}{N}\sum_{k=1}^{N} \sigma^2_{k,\rho}$$

Since $\omega_k \in [\delta, 1-\delta]$ it is immediate to check that there exists $C = C(\delta) < \infty$ such that $p_{k,\rho} \leqslant C p_{\ell,\rho}$ for all k, ℓ and ρ. In particular for some $C = C(\delta) < \infty$ one has

$$C^{-1}\rho \leqslant p_{k,\rho} \leqslant C\rho\,, \quad C^{-1}\rho(1-\rho) \leqslant \sigma^2_{k,\rho} \leqslant C\rho(1-\rho) \tag{98}$$

Given $k, \ell \in \{1, \ldots, N\}$ consider the events

$$U_1 = \{\eta : \sum_{j \neq k,\ell} \eta_j = \rho N - 1\}\,, \quad U_2 = \{\eta : \sum_{j \neq k,\ell} \eta_j = \rho N - 2\}\,.$$

A simple computation shows that

$$\frac{\rho_k}{\rho_\ell} = \frac{\omega_k\big((1-\omega_\ell)\mu_{N,\rho}(U_1) + \omega_\ell e^{\lambda}\mu_{N,\rho}(U_2)\big)}{\omega_\ell\big((1-\omega_k)\mu_{N,\rho}(U_1) + \omega_k e^{\lambda}\mu_{N,\rho}(U_2)\big)} \tag{99}$$

From the bounds on ω we deduce that there exists $C = C(\delta) < \infty$ such that $\rho_k \leqslant C\rho_\ell$ and similarly

$$C^{-1}\rho \leqslant \rho_k \leqslant C\rho\,, \quad C^{-1}\rho(1-\rho) \leqslant \gamma_k^2 \leqslant C\rho(1-\rho)\,. \tag{100}$$

We turn to the proof of the lemma. By duality we may assume $\rho \leqslant 1/2$. We start with the case $1/2 \geqslant \rho \geqslant N^{-3/4}$. Let $\tilde{v}_{k,\rho}(\zeta)$ denote the characteristic function

$$\tilde{v}_{k,\rho}(\zeta) = \mu_{k,\rho}\Big[\exp\big(i\frac{\zeta\hat{\eta}_k}{\sigma_\rho\sqrt{N}}\big)\Big]\,.$$

Since by (98) $\sigma^2_{k,\rho} \geqslant C^{-1}\sigma^2_\rho$, the argument of Lemma 3.3 implies the Gaussian bound

$$|\tilde{v}_{k,\rho}(\zeta)| \leqslant e^{-a\zeta^2/N}\,, \quad |\zeta| \leqslant \pi\sigma_\rho\sqrt{N} \tag{101}$$

with some $a > 0$ only depending on δ. Using Fourier transform we see that

$$\nu_{N,\rho}(\bar{\eta}_k \bar{\eta}_\ell) = \frac{p_{k,\rho} p_{\ell,\rho} A_{k,\ell}}{B^2}$$

with

$$B := \int \mathrm{d}\zeta \prod_{k=1}^{N} \tilde{v}_{k,\rho}(\zeta)$$

and

$$\begin{aligned} A_{k,\ell} := & \int \mathrm{d}\zeta \, e^{i \frac{\zeta}{\sigma_\rho \sqrt{N}} (2 - p_{k,\rho} - p_{\ell,\rho})} \prod_{j \neq k,\ell} \tilde{v}_{j,\rho}(\zeta) \int \mathrm{d}\zeta' \prod_j \tilde{v}_{j,\rho}(\zeta') \\ (102) \qquad & - \int \mathrm{d}\zeta \, e^{i \frac{\zeta}{\sigma_\rho \sqrt{N}} (1 - p_{k,\rho})} \prod_{j \neq k} \tilde{v}_{j,\rho}(\zeta) \int \mathrm{d}\zeta' \, e^{i \frac{\zeta'}{\sigma_\rho \sqrt{N}} (1 - p_{\ell,\rho})} \prod_{j \neq \ell} \tilde{v}_{j,\rho}(\zeta') \, . \end{aligned}$$

Here all integrals are in the range $[-\pi\sigma_\rho\sqrt{N}, \pi\sigma_\rho\sqrt{N}]$. Using the hypothesis $\rho \geqslant N^{-3/4}$, the bounds (98) and the computation of section 3, see (47), we have

$$\tilde{v}_{j,\rho}(\zeta) = 1 - \frac{\zeta^2 \sigma_{j,\rho}^2}{2\sigma_\rho^2 N} + O\big(N^{-1-\epsilon}\big) \, , \qquad |\zeta| \leqslant C \log N \, .$$

Since $\sigma_\rho^2 = (\sum_k \sigma_{k,\rho}^2)/N$ we have

$$\prod_j \tilde{v}_{j,\rho}(\zeta) = 1 - \frac{\zeta^2}{2} + O\big(N^{-\epsilon}\big) \, , \quad |\zeta| \leqslant C \log N \, .$$

As in (50) we deduce

$$B = \sqrt{2\pi} + O(N^{-\epsilon})$$

Moreover

$$\prod_{j \neq k,\ell} \tilde{v}_{j,\rho}(\zeta) = \Big(1 + \frac{\zeta^2 (\sigma_{k,\rho}^2 + \sigma_{\ell,\rho}^2)}{2\sigma_\rho^2 N} + O\big(N^{-1-\epsilon}\big)\Big) \prod_j \tilde{v}_{j,\rho}(\zeta) \, ,$$

$$\prod_{j \neq k} \tilde{v}_{j,\rho}(\zeta) = \Big(1 + \frac{\zeta^2 \sigma_{k,\rho}^2}{2\sigma_\rho^2 N} + O\big(N^{-1-\epsilon}\big)\Big) \prod_j \tilde{v}_{j,\rho}(\zeta) \, , \qquad |\zeta| \leqslant C \log N \, .$$

If we plug these expansions in (102) and open all the brackets as in the derivation of (48) we obtain the estimate

$$p_{k,\rho} p_{\ell,\rho} A_{k,\ell} = -2\pi \frac{\sigma_{k,\rho}^2 \sigma_{\ell,\rho}^2}{\sigma_\rho^2 N} + O(\sigma_\rho^2 N^{-1-\epsilon})$$

uniformly in the case $\rho \geqslant N^{-3/4}$ ($\sigma_\rho^2 \geqslant C^{-1}N^{-3/4}$). Using $\gamma_k = O(\sigma_{k,\rho}) = O(\sigma_\rho)$ the Lemma follows with $\beta_k = \sigma_{k,\rho}^2/(\gamma_k\sigma_\rho) = O(1)$ by (98) and (100). This proves (8.2) in the case $\rho \geqslant N^{-3/4}$.

We now prove the lemma for densities $\rho \leqslant N^{-3/4}$. We set $\hat{\omega}_k := \omega_k/(1-\omega_k)$ and rewrite (99) as

$$\frac{\rho_k}{\rho_\ell} = \frac{\hat{\omega}_k}{\hat{\omega}_\ell}\,\frac{1+e^\lambda\hat{\omega}_\ell W}{1+e^\lambda\hat{\omega}_k W}\,, \qquad W := \frac{\mu_{N,\rho}(U_2)}{\mu_{N,\rho}(U_1)}\,. \tag{103}$$

When $\rho N = 1$ we have $W = 0$ and $\rho_k/\rho_\ell = \hat{\omega}_k/\hat{\omega}_\ell$. Suppose $\rho N \geqslant 2$. Define the event

$$V^m = \{\eta : \sum_{j\neq k,\ell,m} \eta_j = \rho N - 2\}\,.$$

Then

$$\mu_{N,\rho}(U_1) = \frac{1}{\rho N - 1}\sum_{m\neq k,\ell} p_{m,\rho}\mu_{N,\rho}(V^m)\,,$$

$$\mu_{N,\rho}(U_2) = \frac{1}{N(1-\rho)}\sum_{m\neq k,\ell} (1-p_{m,\rho})\mu_{N,\rho}(V^m)\,.$$

Since $p_{m,\rho} \geqslant C^{-1}\rho$ we see that $W = \mu_{N,\rho}(U_2)/\mu_{N,\rho}(U_1) \leqslant C$ uniformly. Using $e^\lambda \leqslant Cp_{k,\rho} \leqslant C'\rho$, from (103) we have

$$\frac{\rho_k}{\rho_\ell} = \frac{\hat{\omega}_k}{\hat{\omega}_\ell}\left[1 + O(\rho)\right]. \tag{104}$$

Summing over k in (104) we arrive at the estimate

$$\rho_\ell = \frac{\rho N\hat{\omega}_\ell}{\sum_k \hat{\omega}_k} + O(\rho^2)\,. \tag{105}$$

Set now $\rho_\ell^{(j)} := \nu_{N,\rho}(\eta_\ell \,|\, \eta_j = 1)$. From (105) applied to $N-1$ sites with $\rho N - 1$ particles:

$$\rho_\ell^{(j)} = \frac{(\rho N-1)\hat{\omega}_\ell}{\sum_{k\neq j}\hat{\omega}_k} + O(\rho^2) = \rho_\ell - \frac{\hat{\omega}_\ell}{\sum_k \hat{\omega}_k} + O(\rho^2)\,.$$

Since $\nu_{N,\rho}(\bar{\eta}_k\bar{\eta}_\ell) = -\rho_k(\rho_\ell - \rho_\ell^{(k)})$, $\Gamma_{k,\ell}$ can be written as

$$\Gamma_{k,\ell} = \frac{\rho_k\hat{\omega}_\ell}{\gamma_k\gamma_\ell\sum_j\hat{\omega}_j} + O(\rho^2) = \frac{\rho N\hat{\omega}_k\hat{\omega}_\ell}{\gamma_k\gamma_\ell(\sum_j\hat{\omega}_j)^2} + O(\rho^2)\,.$$

Since $\rho^2 \leqslant N^{-3/2}$, (97) follows with $\beta_k := N\hat{\omega}_k\sqrt{\rho}/\gamma_k(\sum_j \hat{\omega}_j)$. Then $\beta_k = O(1)$ by (100). This completes the proof of Lemma 8.2.

References

[1] F.C. Alcaraz, *Exact steady states of asymmetric diffusion and two-species annihilation with back reaction from the ground state of quantum spin models*, Intern. Journal of Modern Physics B, **8**, 3449-3461, 1994.

[2] H. J. Brascamp, E. H. Lieb, *On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation.* J. Funct. Anal. **22**, 366–389, 1976

[3] N. Cancrini, F. Martinelli, *On the spectral gap of Kawasaki dynamics under a mixing condition revisited.* J. Math. Phys. **41**, no. 3, 1391-1423, 2000

[4] P. Caputo, *Uniform Poincaré inequalities in unbounded conservative spin systems: The non–interacting case*, `preprint` 2002

[5] P. Caputo, F. Martinelli, *Asymmetric diffusion and the energy gap above the 111 ground state of the quantum XXZ model*, Comm. Math. Phys. **226**, 323–375, 2002

[6] P. Caputo, F. Martinelli, *Relaxation time of anisotropic simple exclusion processes and quantum Heisenberg models*, to appear in Ann. Appl. Probab.

[7] E. Carlen, M. C. Carvalho, M. Loss, *Many-Body Aspects of Approach to Equilibrium.* Sèminaire: Equations aux Dèrivées Partielles, Exp. No. XIX, Ecole Polytech., Palaiseau, 2001.

[8] E. Carlen, M.C. Carvalho, M. Loss, *Determination of the spectral gap in Kac's master equation and related stochastic evolutions*, `preprint` 2002

[9] D. Chafai, *Glauber versus Kawasaki for spectral gap and logarithmic Sobolev inequalities of some unbounded conservative spin systems*, `preprint` 2002, archived as `mp_arc 02-30`

[10] P. Diaconis, L. Saloff-Coste, *Bounds for Kac's master equation*, Comm. Math. Phys. **209**, 729–755, 2000

[11] P. Diaconis, M. Shahshahani, *Time to reach stationarity in the Bernoulli-Laplace diffusion model*, SIAM J. Math. Anal. **18**, no. 1, 208–218, 1987.

[12] A. Faggionato, F. Martinelli, *Hydrodynamic limit of a disordered lattice gas*, `preprint` 2002

[13] E. Janvresse, *Spectral gap for Kac's model of Boltzmann equation*, Ann. Probab. **29**, 288–304, 2001

[14] E. Janvresse, C. Landim, J. Quastel, H.-T. Yau, *Relaxation to equilibrium of conservative dynamics. I. Zero-range processes*, Ann. Probab. **27**, 325–360, 1999

[15] M. Kac, *Foundations of kinetic theory*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III, 171–197. University of California Press, 1956.

[16] C. Kipnis, C. Landim, *Scaling limits of interacting particle systems.* Grundlehren der Mathematischen Wissenschaften, 320. Springer-Verlag, Berlin, 1999

[17] T. Koma, B. Nachtergaele, *The spectral gap of the ferromagnetic XXZ chain*, Lett. Math. Phys. **40**, no. 1, 1–16, 1997.

[18] T. Koma, B. Nachtergaele, S. Starr, *The spectral gap for the ferromagnetic spin-J XXZ chain*, Adv. Theor. Math. Phys. **5**, 1047–1090, 2001

[19] C. Landim, G. Panizo, H. T. Yau, *Spectral gap and logarithmic Sobolev inequality for unbounded conservative spin systems*, Ann. Inst. H. Poincaré Probab. Statist. **38**, 739–777, 2002

[20] C. Landim, S. Sethuraman, S. R. S. Varadhan, *Spectral gap for zero-range dynamics.* Ann. Probab. **24**, 1871–1902, 1996

[21] S. T. Lu, H. T. Yau, *Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics.* Comm. Math. Phys. **156**, 399-433, 1993

[22] J. Quastel, *Diffusion of color in the simple exclusion process*, Comm. Pure Appl. Math. **45**, 623–679, 1992

[23] J. Quastel, H.–T. Yau, *Bulk diffusion in a system with site–disorder*, `preprint`

[24] S. R. S. Varadhan, *Nonlinear diffusion limit for a system with nearest neighbor interactions. II*, in Asymptotic problems in probability theory (Sanda/Kyoto, 1990), 75–128, Pitman Res. Notes Math. Ser., 283, Longman Sci. Tech. , 1993.

Dip. Matematica
Universita' di Roma Tre
L.go S. Murialdo 1
00146 Roma
Italy.
E-mail address: `caputo@mat.uniroma3.it`

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 89–114

Ten Explicit Criteria of One-Dimensional Processes

Mu-Fa Chen

Abstract.

The traditional ergodicity consists a crucial part in the theory of stochastic processes, plays a key role in practical applications. The ergodicity has much refined recently, due to the study on some inequalities, which are especially powerful in the infinite dimensional situation. The explicit criteria for various types of ergodicity for birth-death processes and one-dimensional diffusions are collected in Tables 8.1 and 8.2, respectively. In particular, an interesting story about how to obtain one of the criteria for birth-death processes is explained in details. Besides, a diagram for various types of ergodicity for general reversible Markov processes is presented.

The paper is organized as follows. First, we recall the study on an exponential convergence from different point of view in different subjects: probability theory, spectral theory and harmonic analysis. Then we show by examples the difficulties of the study and introduce the explicit criterion for the convergence, the variational formulas and explicit estimates for the convergence rates. Some comparison with the known results and an application are included. Next, we present ten (eleven) criteria for the two classes of processes, respectively, with some remarks. In particular, a diagram of various types of ergodicity for general reversible Markov processes is presented. For which, partial proofs are included in Appendix. Finally, we indicate a generalization to Banach spaces, this enables us to cover a large class of inequalities (equivalently, various types of ergodicity).

Let us begin with the paper by recalling the three traditional types of ergodicity.

Received November 21, 2002.

Revised December 20, 2002.

Research supported in part by NSFC (No. 10121101), RFDP and 973 Project.

§1. Three traditional types of ergodicity

Let $Q = (q_{ij})$ be a regular Q-matrix on a countable set $E = \{i, j, k, \cdots\}$. That is, $q_{ij} \geq 0$ for all $i \neq j$, $q_i := -q_{ii} = \sum_{j \neq i} q_{ij} < \infty$ for all $i \in E$ and Q determines uniquely a transition probability matrix $P_t = (p_{ij}(t))$ (which is also called a Q-process or a Markov chain). Denote by $\pi = (\pi_i)$ a stationary distribution of P_t: $\pi P_t = \pi$ for all $t \geq 0$. From now on, assume that the Q-matrix is irreducible and hence the stationary distribution π is unique. Then, the three types of ergodicity are defined respectively as follows.

$$\text{(1.1)} \qquad \textit{Ordinary ergodicity}: \quad \lim_{t\to\infty} |p_{ij}(t) - \pi_j| = 0$$

$$\text{(1.2)} \qquad \textit{Exponential ergodicity}: \quad \lim_{t\to\infty} e^{\hat{\alpha} t} |p_{ij}(t) - \pi_j| = 0$$

$$\text{(1.3)} \qquad \textit{Strong ergodicity}: \quad \lim_{t\to\infty} \sup_i |p_{ij}(t) - \pi_j| = 0$$

$$\Longleftrightarrow \lim_{t\to\infty} e^{\hat{\beta} t} \sup_i |p_{ij}(t) - \pi_j| = 0,$$

where $\hat{\alpha}$ and $\hat{\beta}$ are (the largest) positive constants and i, j varies over whole E. The equivalence in (1.3) is well known but one may refer to Proof (b) in the Appendix of this paper. These definitions are meaningful for general Markov processes once the pointwise convergence is replaced by the convergence in total variation norm. The three types of ergodicity were studied in a great deal during 1953–1981. Especially, it was proved that

strong ergodicity $\Longrightarrow$ exponential ergodicity $\Longrightarrow$ ordinary ergodicity.

Refer to Anderson (1991), Chen (1992, Chapter 4) and Meyn and Tweedie (1993) for details and related references. The study is quite complete in the sense that we have the following criteria which are described by the Q-matrix plus a test sequence (y_i) only, except the exponential ergodicity for which one requires an additional parameter λ.

Theorem 1.1 (Criteria). *Let $H \neq \emptyset$ be an arbitrary but fixed finite subset of E. Then the following conclusions hold.*

(1) *The process P_t is ergodic iff the system of inequalities*

$$\text{(1.4)} \qquad \begin{cases} \sum_j q_{ij} y_j \leq -1, & i \notin H \\ \sum_{i \in H} \sum_{j \neq i} q_{ij} y_j < \infty \end{cases}$$

has a nonnegative finite solution (y_i).

(2) *The process P_t is exponentially ergodic iff for some $\lambda > 0$ with $\lambda < q_i$ for all i, the system of inequalities*

$$\begin{cases} \sum_j q_{ij} y_j \le -\lambda y_i - 1, & i \notin H \\ \sum_{i \in H} \sum_{j \ne i} q_{ij} y_j < \infty \end{cases} \tag{1.5}$$

has a nonnegative finite solution (y_i).

(3) *The process P_t is strongly ergodic iff the system (1.4) of inequalities has a bounded nonnegative solution (y_i).*

The probabilistic meaning of the criteria reads respectively as follows:

$$\max_{i \in H} \mathbb{E}_i \sigma_H < \infty, \quad \max_{i \in H} \mathbb{E}_i e^{\lambda \sigma_H} < \infty \quad \text{and} \quad \sup_{i \in E} \mathbb{E}_i \sigma_H < \infty,$$

where $\sigma_H = \inf\{t \ge \text{the first jumping time} : X_t \in H\}$ and λ is the same as in (1.5). The criteria are not completely explicit since they depend on the test sequences (y_i) and in general it is often non-trivial to solve a system of infinite inequalities. Hence, one expects to find out some explicit criteria for some specific processes. Clearly, for this, the first candidate should be the birth-death process. Recall that for a birth-death process with state space $E = \mathbb{Z}_+ = \{0, 1, 2, \cdots\}$, its Q-matrix has the form: $q_{i,i+1} = b_i > 0$ for all $i \ge 0$, $q_{i,i-1} = a_i > 0$ for all $i \ge 1$ and $q_{ij} = 0$ for all other $i \ne j$. Along this line, it was proved by Tweedie (1981)(see also Anderson (1991) or Chen (1992)) that

$$S := \sum_{n \ge 1} \mu_n \sum_{j \le n-1} \frac{1}{\mu_j b_j} < \infty \Longrightarrow \text{Exponential ergodicity,} \tag{1.6}$$

where $\mu_0 = 1$ and $\mu_n = b_0 \cdots b_{n-1}/a_1 \cdots a_n$ for all $n \ge 1$. Refer to Wang (1980), Yang (1986) or Hou et al (2000) for the probabilistic meaning of S. The condition is explicit since it depends only on the rates a_i and b_i. However, the condition is not necessary. A simple example is as follows. Let $a_i = b_i = i^\gamma\ (i \ge 1)$ and $b_0 = 1$. Then the process is exponential ergodic iff $\gamma \ge 2$ (see Chen (1996)) but $S < \infty$ iff $\gamma > 2$. Surprisingly, the condition is correct for strong ergodicity.

Theorem 1.2 (Zhang, Lin and Hou (2000)).

$$S < \infty \Longleftrightarrow \textit{Strong ergodicity.}$$

Refer to Hou et al (2000). With a different proof, the result is extended by Y. H. Zhang (2001) to the single-birth processes with state space $\mathbb{Z}_+$. Here, the term "single birth"means that $q_{i,i+1} > 0$ for all $i \ge 0$

but $q_{ij} \geq 0$ can be arbitrary for $j < i$. Introducing this class of Q-processes is due to the following observation: If the first inequality in (1.4) is replaced by equality, then we get a recursion formula for (y_i) with one parameter only. Hence, there should exist an explicit criterion for the ergodicity (resp. uniqueness, recurrence and strong ergodicity). For (1.5), there is also a recursion formula but now two parameters are involved and so it is unclear whether there exists an explicit criterion or not for the exponential ergodicity.

Note that the criteria are not enough to estimate the convergence rate $\hat{\alpha}$ or $\hat{\beta}$ (cf. Chen (2000a)). It is the main reason why we have to come back to study the well-developed theory of Markov chains. For birth-death processes, the estimation of $\hat{\alpha}$ was studied by Doorn in a book (1981) and in a series of papers (1985, 1987, 1991). He proved, for instance, the following lower bound

$$\hat{\alpha} \geq \inf_{i\geq 0} \{a_{i+1} + b_i - \sqrt{a_i b_i} - \sqrt{a_{i+1} b_{i+1}}\},$$

which is exact when a_i and b_i are constant. The following formula for the lower bounds was implicated in his papers and rediscovered in a different point of view (in the study on spetral gap) by Chen (1996):

$$\hat{\alpha} = \sup_{v>0} \inf_{i\geq 0} \{a_{i+1} + b_i - a_i/v_{i-1} - b_{i+1} v_i\}.$$

Besides, the precise $\hat{\alpha}$ was determined by Doorn for four practical models. The main tool used in Doorn's study is the Karlin-Mcgregor's representation theorem, a specific spectral representation, involving heavy techniques. There is no explicit criterion for $\hat{\alpha} > 0$ ever appeared so far.

§2. The first (non-trivial) eigenvalue (spectral gap)

The birth-death processes have a nice property—symmetrizability: $\mu_i p_{ij}(t) = \mu_j p_{ji}(t)$ for all i, j and $t \geq 0$. Then, the matrix Q can be regarded as a self-adjoint operator on the real L^2-space $L^2(\mu)$ with norm $\|\cdot\|$. In other words, one can use the well-developed L^2-theory. For instance, one can study the L^2-exponential convergence given below. Assuming that $Z = \sum_i \mu_i < \infty$ and then setting $\pi_i = \mu_i / Z$. Then, the convergence means that

$$\|P_t f - \pi(f)\| \leq \|f - \pi(f)\| \leq e^{-\lambda_1 t} \tag{2.1}$$

for all $t \geq 0$, where $\pi(f) = \int f \mathrm{d}\pi$ and λ_1 is the first non-trivial eigenvalue (more precisely, the spectral gap) of $(-Q)$ (cf. Chen (1992, Chapter 9)).

The estimation of λ_1 for birth-death processes was studied by Sullivan (1984), Liggett (1989) and Landim, Sethuraman and Varadhan (1996) (see also Kipnis & Lamdin (1999)). It was used as a comparison tool to handle the convergence rate for some interacting particle systems, which are infinite-dimensional Markov processes. Here we recall three results as follows.

Theorem 2.1 (Sullivan (1984)). *Let c_1 and c_2 be two constants satisfying*

$$c_1 \ge \sup_{i\ge 1} \frac{\sum_{j\ge i} \mu_j}{\mu_i}, \qquad c_2 \ge \sup_{i\ge 1} \frac{\mu_i}{\mu_i a_i}.$$

Then $\lambda_1 \ge 1/4c_1^2 c_2$.

Theorem 2.2 (Liggett (1989)). *Let c_1 and c_2 be two constants satisfying*

$$c_1 \ge \sup_{i\ge 1} \frac{\sum_{j\ge i} \mu_j}{\mu_i a_i}, \qquad c_2 \ge \sup_{i\ge 1} \frac{\sum_{j\ge i} \mu_j a_j}{\mu_i a_i}.$$

Then $\lambda_1 \ge 1/4c_1 c_2$.

Theorem 2.3 (Liggett (1989)). *For bounded a_i and b_i, $\lambda_1 > 0$ iff (μ_i) has an exponential tail.*

The reason we are mainly interested in the lower bounds is that on the one hand, they are more useful in practice and on the other hand, the upper bounds are usually easier to obtain from the following classical variational formula.

$$\lambda_1 = \inf\{D(f) : \mu(f) = 0, \mu(f^2) = 1\},$$

where

$$D(f) = \frac{1}{2}\sum_{i,j} \mu_i q_{ij}(f_j - f_i)^2, \qquad \mathscr{D}(D) = \{f \in L^2(\mu) : D(f) < \infty\}$$

and $\mu(f) = \int f \mathrm{d}\mu$.

Let us now leave Markov chains for a while and turn to diffusions.

§3. One-dimensional diffusions

As a parallel of birth-death process, we now consider an elliptic operator $L = a(x)\mathrm{d}^2/\mathrm{d}x^2 + b(x)\mathrm{d}/\mathrm{d}x$ on the half line $[0, \infty)$ with $a(x) > 0$ everywhere. Again, we are interested in estimation of the principle eigenvalues, which consist of the typical, well-known Sturm-Liouville eigenvalue problem in the spectral theory. Refer to Egorov & Kondratiev (1996) for the present status of the study and references. Here, we mention two results, which are the most general ones we have ever known before.

Theorem 3.1. *Let $b(x) \equiv 0$ (which corresponds to the birth-death process with $a_i = b_i$ for all $i \geq 1$) and set $\delta = \sup_{x>0} x \int_x^\infty a^{-1}$. Here we omit the integration variable when it is integrated with respect to the Lebesgue measure. Then, we have*

1. *Kac & Krein (1958): $\delta^{-1} \geq \lambda_0 \geq (4\delta)^{-1}$, here λ_0 is the first eigenvalue corresponding to the Dirichlet boundary $f(0) = 0$.*
2. *Kotani & Watanabe (1982): $\delta^{-1} \geq \lambda_1 \geq (4\delta)^{-1}$.*

It is simple matter to rewrite the classical variational formula as (3.1) below. Similarly, we have (3.2) for λ_0.

Poincaré inequalities.

$$\lambda_1 : \quad \|f - \pi(f)\|^2 \leq \lambda_1^{-1} D(f) \tag{3.1}$$

$$\lambda_0 : \quad \|f\|^2 \leq \lambda_0^{-1} D(f), \quad f(0) = 0. \tag{3.2}$$

It is interesting that inequality (3.2) is a special but typical case of the weighted Hardy inequality discussed in the next section.

§4. Weighted Hardy inequality

The classical Hardy inequality goes back to Hardy (1920):

$$\int_0^\infty \left(\frac{f}{x}\right)^p \leq \left(\frac{p}{p-1}\right)^p \int_0^\infty f'^p, \quad f(0) = 0, f' \geq 0,$$

where the optimal constant was determined by Landau (1926). After a long period of efforts by analysts, the inequality was finally extended to the following form, called weighted Hardy inequality (Muckenhoupt (1972))

$$\int_0^\infty f^2 \mathrm{d}\nu \leq A \int_0^\infty f'^2 \mathrm{d}\lambda, \quad f \in C^1, f(0) = 0, \tag{4.1}$$

where ν and λ be nonnegative Borel measures.

The Hardy-type inequalities play a very important role in the study of harmonic analysis and have been treated in many publications. Refer to the books: Opic & Kufner (1990), Dynkin (1990), Mazya (1985) and the survey article Davies (1999) for more details. We will come back this inequality soon.

We have finished the overview of the study on the exponential convergence (equivalently, the Poincaré inequality) in the different subjects. In order to have a more concrete feeling about the the difficulties of the topic, we now introduce some simple examples.

§5. Difficulties

First, consider the birth-death processes with finite state space E. When $E = \{0, 1\}$, the Q-matrix becomes $Q = \begin{pmatrix} -b_0 & b_0 \\ a_1 & -a_1 \end{pmatrix}$. Then, it is trivial that $\lambda_1 = a_1 + b_0$. The result is nice since either a_1 or b_0 increases, so does λ_1. If we go one more step, $E = \{0, 1, 2\}$, then we have four parameters b_0, b_1 and a_1, a_2 and

$$\lambda_1 = 2^{-1}\big[a_1 + a_2 + b_0 + b_1 - \sqrt{(a_1 - a_2 + b_0 - b_1)^2 + 4a_1 b_1}\big].$$

Now, the role for λ_1 played by the parameters becomes ambiguous. When $E = \{0, 1, 2, 3\}$, we have six parameters: $b_0, b_1, b_2, a_1, a_2, a_3$. Then

$$\lambda_1 = \frac{D}{3} - \frac{C}{3 \cdot 2^{1/3}} + \frac{2^{1/3}\,(3\,B - D^2)}{3\,C},$$

where the quantities D, B and C are not too complicated:

$$\begin{aligned}
D &= a_1 + a_2 + a_3 + b_0 + b_1 + b_2,\\
B &= a_3\, b_0 + a_2\, (a_3 + b_0) + a_3\, b_1\\
&\qquad + b_0\, b_1 + b_0\, b_2 + b_1\, b_2 + a_1\, (a_2 + a_3 + b_2),\\
C &= \Big(A + \sqrt{4(3\,B - D^2)^3 + A^2}\,\Big)^{1/3}.
\end{aligned}$$

However, in the last expression, another quantity is involved:

$$\begin{aligned}
A = {} & -2\,a_1^3 - 2\,a_2^3 - 2\,a_3^3 + 3\,a_3^2\,b_0 + 3\,a_3\,b_0^2 - 2\,b_0^3 + 3\,a_3^2\,b_1\\
& - 12\,a_3\,b_0\,b_1 + 3\,b_0^2\,b_1 + 3\,a_3\,b_1^2 + 3\,b_0\,b_1^2 - 2\,b_1^3 - 6\,a_3^2\,b_2 + 6\,a_3\,b_0\,b_2\\
& + 3\,b_0^2\,b_2 + 6\,a_3\,b_1\,b_2 - 12\,b_0\,b_1\,b_2 3\,b_1^2\,b_2 - 6\,a_3\,b_2^2 + 3\,b_0\,b_2^2 + 3\,b_1\,b_2^2\\
& - 2\,b_2^3 + 3\,a_1^2\,(a_2 + a_3 - 2\,b_0 - 2\,b_1 + b_2)\\
& + 3\,a_2^2\,[a_3 + b_0 - 2\,(b_1 + b_2)]\\
& + 3\,a_2\,\big[a_3^2 + b_0^2 - 2\,b_1^2 - b_1\,b_2 - 2\,b_2^2\\
& \qquad\quad - a_3(4\,b_0 - 2\,b_1 + b_2) + 2\,b_0(b_1 + b_2)\big]\\
& + 3\,a_1\,\big[a_2^2 + a_3^2 - 2\,b_0^2 - b_0\,b_1 - 2\,b_1^2 - a_2(4\,a_3 - 2\,b_0 + b_1 - 2\,b_2)\\
& \qquad\quad + 2\,b_0\,b_2 + 2\,b_1\,b_2 + b_2^2 + 2\,a_3(b_0 + b_1 + b_2)\big].
\end{aligned}$$

Thus, the roles of the parameters are completely mazed! Of course, it is impossible to compute λ_1 explicitly when the size of the matrix is greater than five!

Next, we go to the estimation of λ_1. Consider the infinite state space $E = \{0, 1, 2, \cdots\}$. Denote by g and $D(g)$, respectively, the eigenfunction of λ_1 and the degree of g when g is polynomial. Three examples of the perturbation of λ_1 and $D(g)$ are listed in Table 1.1.

$b_i\,(i \geq 0)$	$a_i\,(i \geq 1)$	λ_1	$D(g)$
$i + c\,(c > 0)$	$2i$	1	1
$i+1$	$2i+3$	2	2
$i+1$	$2i + (4+\sqrt{2})$	3	3

Table 1.1 Three examples of the perturbation of λ_1 and $D(g)$

The first line is the well known linear model, for which $\lambda_1 = 1$, independent of the constant $c > 0$, and g is linear. Next, keeping the same birth rate, $b_i = i + 1$, changes the death rate a_i from $2i$ to $2i + 3$ (resp. $2i + 4 + \sqrt{2}$), which leads to the change of λ_1 from one to two (resp. three). More surprisingly, the eigenfunction g is changed from linear to quadratic (resp. triple). For the other values of a_i between $2i$, $2i + 3$ and $2i+4+\sqrt{2}$, λ_1 is unknown since g is non-polynomial. As seen from these examples, the first eigenvalue is very sensitive. Hence, in general, it is very hard to estimate λ_1.

Hopefully, I have presented enough examples to show the difficulties of the topic.

§6. Results about λ_1, $\hat{\alpha}$ and λ_0

It is position to state our results. To do so, define

$$\mathscr{W} = \{w : w_i \uparrow\uparrow, \pi(w) \geq 0\}, \quad Z = \sum_i \mu_i,$$

$$\delta = \sup_{i>0} \sum_{j \leq i-1} \frac{1}{\mu_j b_j} \sum_{j \geq i} \mu_j,$$

where $\uparrow\uparrow$ means strictly increasing. By suitable modification, we can define $\mathscr{W}'$ and explicit sequences δ_n and δ_n'. Refer to Chen (2001a) for details.

The next result provides a complete answer to the question proposed in Section 1.

Theorem 6.1. *For birth-death processes, the following assertions hold.*

(1) *Dual variational formulas:*

$$(6.1)\quad \lambda_1 = \sup_{w\in\mathscr{W}} \inf_{i\geq 0} \mu_i b_i(w_{i+1}-w_i)\bigg/\sum_{j\geq i+1}\mu_j w_j \qquad [\textit{Chen}\ (1996)]$$

$$(6.2)\quad = \inf_{w\in\mathscr{W}'} \sup_{i\geq 0} \mu_i b_i(w_{i+1}-w_i)\bigg/\sum_{j\geq i+1}\mu_j w_j \qquad [\textit{Chen}\ (2001a)]$$

(2) *Appoximating procedure and explicit bounds:*

$$Z\delta^{-1} \geq {\delta'_n}^{-1} \geq \lambda_1 \geq \delta_n^{-1} \geq (4\delta)^{-1} \textit{ for all } n \qquad [\textit{Chen}\ (2000b, 2001a)].$$

(3) *Explicit criterion:* $\lambda_1 > 0$ *iff* $\delta < \infty$ [*Miclo* (1999), *Chen* (2000*b*)].
(4) *Relation:* $\hat{\alpha} = \lambda_1$ [*Chen* (1991)].

In (6.1), only two notations are used: the sets $\mathscr{W}$ and $\mathscr{W}'$ of test functions (sequences). Clearly, for each test function, (6.1) gives us a lower bound of λ_1. This explains the meaning of "variational". Because of (6.1), it is now easy to obtain some lower estimates of λ_1, and in particular, one obtains all the lower bounds mentioned above. Next, by exchanging the orders of "sup" and "inf", we get (6.2) from (6.1), ignoring a slight modification of $\mathscr{W}$. In other words, (6.1) and (6.2) are dual of one to the other. For the explicit estimates "$\delta^{-1} \geq \lambda_0 \geq (4\delta)^{-1}$" and in particular for the criterion, one needs to find out a representative test function w among all $w \in \mathscr{W}$. This is certainly not obvious, because the test function w used in the formula is indeed a mimic of the eigenfunction (eigenvector) of λ_1, and in general, the eigenvalues and the corresponding eigenfunctions can be very sensitive, as we have seen from the above examples. Fortunately, there exists such a representative function with a simple form. We will illustrate the function in the context of diffusions in the second to the last paragraph of this section.

In parallel, for diffusions on $[0,\infty]$, define

$$C(x) = \int_0^x b/a, \qquad \delta = \sup_{x>0}\int_0^x e^{-C}\int_x^\infty e^C/a,$$
$$\mathscr{F} = \big\{f \in C[0,\infty)\cap C^1(0,\infty) : f(0) = 0 \text{ and } f'|_{(0,\infty)} > 0\big\}.$$

Theorem 6.2 (Chen (1999a, 2000b, 2001a)). *For diffusion on* $[0,\infty)$*, the following assertions hold.*

(1) *Dual variational formulas:*

$$\lambda_0 \geq \sup_{f\in\mathscr{F}} \inf_{x>0} e^{C(x)}f'(x) \Big/ \int_x^\infty fe^C/a \tag{6.3}$$

$$\lambda_0 \leq \inf_{f\in\mathscr{F}'} \sup_{x>0} e^{C(x)}f'(x) \Big/ \int_x^\infty fe^C/a \tag{6.4}$$

Furthermore, the signs of the equality in (6.3) *and* (6.4) *hold if both* a *and* b *are continuous on* $[0,\infty)$.

(2) *Appoximating procedure and explicit bounds: A decreasing sequence* $\{\delta_n\}$ *and an increasing sequence* $\{\delta_n'\}$ *are constructed explicitly such that*

$$\delta^{-1} \geq {\delta_n'}^{-1} \geq \lambda_0 \geq \delta_n^{-1} \geq (4\delta)^{-1} \qquad \textit{for all } n.$$

(3) *Explicit criterion:* λ_0 *(resp.* λ_1*)* > 0 *iff* $\delta < \infty$.

We mention that the above two results are also based on Chen and Wang (1997a).

To see the power of the dual variational formulas, let us return to the weighted Hardy's inequality.

Theorem 6.3 (Muckenhoupt (1972)). *The optimal constant* A *in the inequality*

$$\int_0^\infty f^2 \mathrm{d}\nu \leq A \int_0^\infty f'^2 \mathrm{d}\lambda, \quad f\in C^1, f(0)=0, \tag{6.5}$$

satisfies $B \leq A \leq 4B$, *where* $B = \sup_{x>0} \nu[x,\infty] \int_x^\infty (\mathrm{d}\lambda_{\mathrm{abs}}/\mathrm{dLeb})^{-1}$ *and* $\mathrm{d}\lambda_{\mathrm{abs}}/\mathrm{dLeb}$ *is the derivative of the absolutely continuous part of* λ *with respect to the Lebesgue measure.*

By setting $\nu = \pi$ and $\lambda = e^C\mathrm{d}x$, it follows that the criterion in Theorem 6.2 is a consequence of the Muckenhoupt's Theorem. Along this line, the criteria in Theorems 6.1 and 6.2 for a typical class of the processes were also obtained by Bobkov and Götze (1999a, b), in which, the contribution of an earlier paper by Luo (1992) was noted.

We now point out that the explicit estimates "$\delta^{-1} \geq \lambda_0 \geq (4\delta)^{-1}$" in Theorems 6.2 or 6.3 follow from our variational formulas immediately. Here we consider the lower bound "$(4\delta)^{-1}$" only, the proof for the upper bound "δ^{-1}" is also easy, in terms of (6.4).

Recall that $\delta = \sup_{x>0} \int_0^x e^{-C} \int_x^\infty e^C/a$. Set $\varphi(x) = \int_0^x e^{-C}$. By using the integration by parts formula, it follows that

$$\begin{aligned}\int_x^\infty \frac{\sqrt{\varphi}\, e^C}{a} &= -\int_x^\infty \sqrt{\varphi}\, \mathrm{d}\Big(\int_\bullet^\infty \frac{e^C}{a}\Big)\\ &\le \frac{\delta}{\sqrt{\varphi(x)}} + \frac{\delta}{2}\int_x^\infty \frac{\varphi'}{\varphi^{3/2}} \le \frac{2\delta}{\sqrt{\varphi(x)}}.\end{aligned}$$

Hence

$$I(\sqrt{\varphi})(x) = \frac{e^{-C(x)}}{(\sqrt{\varphi})'(x)} \int_x^\infty \frac{\sqrt{\varphi} e^C}{a} \le \frac{e^{-C(x)}\sqrt{\varphi(x)}}{(1/2)e^{-C(x)}} \cdot \frac{2\delta}{\sqrt{\varphi(x)}} = 4\delta.$$

This gives us the required bound by (6.3).

Theorem 6.2 can be immediately applied to the whole line or higher-dimensional situation. For instance, for Laplacian on compact Riemannian manifolds, it was proved by Chen & Wang (1997b) that

$$\lambda_1 \ge \sup_{f \in \mathscr{F}} \inf_{r \in (0,D)} I(f)(r)^{-1} =: \xi_1,$$

where $I(f)$ is the same as before but for some specific function $C(x)$. Thanks are given to the coupling technique which reduces the higher dimensional case to dimension one. We now have $\delta^{-1} \ge \delta_n'^{\,-1} \downarrow \ge \xi_1 \ge \delta_n^{-1} \uparrow \ge (4\delta)^{-1}$, similar to Theorem 6.2. Refer to Chen (2000b, 2001a) for details. As we mentioned before, the use of the test functions is necessary for producing sharp estimates. Actually, the variational formula enables us to improve a number of best known estimates obtained previously by geometers, but none of them can be deduced from the estimates "$\delta^{-1} \ge \xi_1 \ge (4\delta)^{-1}$". Besides, the approximating procedure enables us to determine the optimal linear approximation of ξ_1 in K:

$$\xi_1 \ge \frac{\pi^2}{D^2} + \frac{K}{2},$$

where D is the diameter of the manifold and K is the lower bound of Ricci curvature (cf., Chen, Scacciatelli and Yao (2001)). We have thus shown the value of our dual variational formulas.

§7. Three basic inequalities

Up to now, we have mainly studied the Poincaré inequality, i.e., (7.1) below. Naturally, one may study other inequalities, for instance, the logarithmic Sobolev inequality or the Nash inequality listed below.

(7.1)

Poincaré inequality: $\|f-\pi(f)\|^2 \le \lambda_1^{-1}D(f)$

(7.2)

Logarithmic Sobolev inequality: $\int f^2\log(|f|/\|f\|)\mathrm{d}\pi \le \sigma^{-1}D(f)$

(7.3)

Nash inequality: $\|f-\pi(f)\|^{2+4/\nu} \le \eta^{-1}D(f)\|f\|_1^{4/\nu}$

(for some $\nu > 0$).

Here, to save notation, σ (resp. η) denotes the largest constant so that (7.2) (resp. (7.3)) holds.

The importance of these inequalities is due to the fact that each inequality describes a type of ergodicity. First, (7.1) $\Longleftrightarrow$ (2.1). Next, the logarithmic Sobolev inequality implies (is indeed equivalent to, in the context of diffusions) the decay of the semigroup P_t to π exponentially in relative entropy with rate σ and the Nash inequality is equivalent to $\|P_t f - \pi(f)\| \le C\|f\|_1/t^{\nu/2}$.

§8. Criteria

Recently, the criteria for the last two inequalities as well as for the discrete spectrum (which means that there is no continuous spectrum and moreover, all eigenvalues have finite multiplicity) are obtained by Mao (2000, 2002a, b), based on the weighted Hardy's inequality. On the other hand, the main parts of Theorems 6.1 and 6.2 are extended to a general class of Banach spaces in Chen (2002a, d, e), which unify a large class inequalities and provide a unified criterion in particular. We can now summarize the results in Table 8.1. The table is arranged in such order that the property in the latter line is stranger than the former one, the only exception is that even though the strong ergodicity is often stronger than the logarithmic Sobolev inequality but they are not comparable in general (Chen (2002b)).

Birth-death processes
Transition intensity:

$$\begin{aligned} i \to i+1 &\quad \text{at rate} \quad b_i = q_{i,i+1} > 0\\ \to i-1 &\quad \text{at rate} \quad a_i = q_{i,i-1} > 0. \end{aligned}$$

Define

$$\mu_0 = 1, \quad \mu_n = \frac{b_0 \cdots b_{n-1}}{a_1 \cdots a_n}, \; n \geq 1; \quad \mu[i, k] = \sum_{i \leq j \leq k} \mu_j.$$

Property	**Criterion**
Uniqueness	$\sum_{n\geq 0} \frac{1}{\mu_n b_n} \mu[0, n] = \infty \quad (*)$
Recurrence	$\sum_{n\geq 0} \frac{1}{\mu_n b_n} = \infty$
Ergodicity	$(*)$ & $\mu[0, \infty) < \infty$
Exponential ergodicity L^2-exponential convergence	$(*)$ & $\sup_{n\geq 1} \mu[n, \infty) \sum_{j\leq n-1} \frac{1}{\mu_j b_j} < \infty$
Logarithmic Sobolev inequality	$(*)$ & $\sup_{n\geq 1} \mu[n, \infty) \log[\mu[n, \infty)^{-1}] \sum_{j\leq n-1} \frac{1}{\mu_j b_j} < \infty$
Strong ergodicity L^1-exponential convergence	$(*)$ & $\sum_{n\geq 0} \frac{1}{\mu_n b_n} \mu[n+1, \infty) = \sum_{n\geq 1} \mu_n \sum_{j\leq n-1} \frac{1}{\mu_j b_j} < \infty$
Nash inequality	$(*)$ & $\sup_{n\geq 1} \mu[n, \infty)^{(\nu-2)/\nu} \sum_{j\leq n-1} \frac{1}{\mu_j b_j} < \infty \; (\varepsilon)$

Table 8.1. Ten criteria for birth-death processes

Here, "$(*)$ & $\cdots$" means that one requires the uniqueness condition in the first line plus the condition "$\cdots$". The "(ε)" in the last line means that there is still a small gap from being necessary. In other words, when $\nu \in (0, 2]$, there is still no criterion for the Nash inequality.

Diffusion processes on $[0, \infty)$ with reflecting boundary
Operator:

$$L = a(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + b(x)\frac{\mathrm{d}}{\mathrm{d}x}.$$

Define

$$C(x) = \int_0^x b/a, \qquad \mu[x, y] = \int_x^y e^C/a.$$

For the Nash inequality, we have the same remark as before. The reason we have one more criterion here is due to the equivalence of the logarithmic Sobolev inequality and the exponential convergence in entropy. However, this is no longer true in the discrete case. In general, the logarithmic Sobolev inequality is stronger than the exponential convergence in entropy. A criterion for the exponential convergence in entropy for birth-death processes remains open (cf., Zhang and Mao (2000) and Mao and Zhang (2000)). The two equivalences in the tables come from the next diagram.

Property	**Criterion**
Uniqueness	$\int_0^\infty \mu[0,x]e^{-C(x)} = \infty \quad (*)$
Recurrence	$\int_0^\infty e^{-C(x)} = \infty$
Ergodicity	$(*)$ & $\mu[0,\infty) < \infty$
Exponential ergodicity L^2-exponential convergence	$(*)$ & $\sup_{x>0} \mu[x,\infty) \int_0^x e^{-C} < \infty$
Discrete spectrum	$(*)$ & $\lim_{n\to\infty} \sup_{x>n} \mu[x,\infty) \int_n^x e^{-C} = 0$
Logarithmic Sobolev inequality Exponential convergence in entropy	$(*)$ & $\sup_{x>0} \mu[x,\infty)\log[\mu[x,\infty)^{-1}]\int_0^x e^{-C} < \infty$
Strong ergodicity L^1-exponential convergence	$(*)$ & $\int_0^\infty \mu[x,\infty)e^{-C(x)} < \infty?$
Nash inequality	$(*)$ & $\sup_{x>0} \mu[x,\infty)^{(\nu-2)/\nu}\int_0^x e^{-C} < \infty(\varepsilon)$

Table 8.2. Eleven criteria for one-dimensional diffusions

§9. New picture of ergodic theory

Theorem 9.1. *Let $(E,\mathscr{E})$ be a measurable space with countably generated $\mathscr{E}$. Then, for a Markov processes with state space $(E,\mathscr{E})$, reversible and having transition probability densities with respect to a probability measure π, we have the diagram shown in Figure 9.1.*

$$\begin{tikzcd}
 & \textit{Nash inequality} \arrow[dl, Rightarrow] \arrow[dr, Rightarrow] & \\
\textit{Logarithmic Sobolev inequality} \arrow[d, Rightarrow] & & L^1\textit{-exponential convergence} \arrow[d, equal] \\
\textit{Exponential convergence in entropy} \arrow[d, Rightarrow] & & \pi\textit{-a.s. Strong ergodicity} \arrow[d, Rightarrow] \\
\textit{Poincaré inequality} \arrow[rr, Leftrightarrow] & {} \arrow[d, Rightarrow] & \pi\textit{-a.s. Exponential ergodicity} \\
 & L^2\textit{-algebraic ergodicity} \arrow[d, Rightarrow] & \\
 & \textit{Ordinary ergodicity} &
\end{tikzcd}$$

Fig. 9.1. Diagram of nine types of ergodicity

Here are some remarks about Figure 9.1.

(1) The importance of the diagram is obvious. For instance, by using the estimates obtained from the study on Poincaré inequality, based on the advantage on the analytic approach — the L^2-theory and the equivalence in the diagram, one can estimate exponentially ergodic convergence rates, for which, the known knowledge is still very limited. Actually, these two convergence rates are often coincided (cf. the proofs given in Appendix). In

particular, one obtains a criterion for the exponential ergodicity in dimension one, which has been opened for a long period. Conversely, one obtains immediately some criteria, which are indeed new, for Poincaré inequality to be held from the well-known criteria for the exponential ergodicity. Next, there is still very limited known knowledge about the L^1-spectrum, due to the structure of the L^1-space, which is only a Banach but not Hilbert space. Based on the probabilistic advantage and the identity in the diagram, from the study on the strong ergodicity, one learns a lot about the L^1-spectral gap of the generator.

(2) The L^2-*algebraic ergodicity* means that $\text{Var}(P_t f) \le CV(f)t^{1-q}$ $(t > 0)$ holds for some V having the properties: V is homogeneous of degree two (in the sense that $V(cf+d) = c^2V(f)$ for any constants c and d) and $V(f) < \infty$ for all functions f with finite support (cf. Liggett (1991)). Refer to Chen and Wang (2000), Röckner and Wang (2001) for the study on the L^2-algebraic convergence.

(3) The diagram is complete in the following sense: each single-directed implication can not be replaced by double-directed one. Moreover, the L^1-exponential convergence (resp., the strong ergodicity) and the logarithmic Sobolev inequality (resp., the exponential convergence in entropy) are not comparable.

(4) The reversibility is used in both of the identity and the equivalence. Without the reversibility, the L^2-exponential convergence still implies π-a.s. exponentially ergodic convergence.

(5) An important fact is that the condition "having densities" is used only in the identity of L^1-exponential convergence and π-a.s. strong ergodicity, without this condition, L^1-exponential convergence still implies π-a.s. strong ergodicity, and so the diagram needs only a little change (However, the reversibility is still required here). Thus, it is a natural open problem to remove this "density's condition".

(6) Except the identity and the equivalence, all the implications in the diagram are suitable for general Markov processes, not necessarily reversible, even though the inequalities are mainly valuable in the reversible situation. Clearly, the diagram extends the ergodic theory of Markov processes.

The diagram was presented in Chen (1999c, 2002b), originally stated mainly for Markov chains. Recently, the identity of L^1-exponential convergence and the π-a.s. strong ergodicity is proven by Mao (2002c). A counter-example of diffusion was constructed by Wang (2001) to show that the strong ergodicity does not imply the exponential convergence in entropy. Partial proofs of the diagram are given in Appendix.

§10. Go to Banach spaces

To conclude this paper, we indicate an idea to show the reason why we should go to the Banach spaces.

Theorem 10.1 (Varopoulos, N. (1985); Carlen, E. A., Kusuoka, S., Stroock, D. W. (1987); Bakry, D., Coulhon, T., Ledoux, M. and Saloff-Coste, L. (1995)). *When $\nu > 2$, the Nash inequality*

$$\|f - \pi(f)\|^{2+4/\nu} \le C_1 D(f)\|f\|_1^{4/\nu}$$

is equivalent to the Sobolev-type inequality

$$\|f - \pi(f)\|_{\nu/(\nu-2)}^2 \le C_2 D(f),$$

where $\|\cdot\|_p$ is the $L^p(\mu)$-norm.

In view of Theorem 10.1, it is natural to study the inequality

$$\|(f - \pi(f))^2\|_{\mathbb{B}} \le AD(f)$$

for a general Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}}, \mu)$. It is interesting that even for the general setup, we still have quite satisfactory results. Refer to Bobkov and Götze (1999a, b) and Chen (2002a, d, e) for details.

§11. Appendix: Partial proofs of Theorem 9.1

The detailed proofs and some necessary counterexamples were presented in Chen (1999c, 2002b) for reversible Markov processes, except the identity of the L^1-exponential convergence and π-a.s. strong ergodicity. Note that for discrete state spaces, one can rule out "a.s." used in the diagram. Here, we prove the new identity and introduce some more careful estimates for the general state spaces. The author would like to acknowledge Y. H. Mao for his nice ideas which are included in this appendix. The steps of the proofs are listed as follows.

(a) Nash inequality $\Longrightarrow$ L^1-exponential convergence and π-a.s. Strong ergodicity.
(b) L^1-exponential convergence $\Longleftrightarrow$ π-a.s. Strong ergodicity.
(c) Nash inequality $\Longrightarrow$ Logarithmic Sobolev inequality.
(d) L^2-exponential convergence $\Longrightarrow$ π-a.s. Exponential ergodicity.
(e) Exponential ergodicity $\Longrightarrow$ L^2-exponential convergence.

(a) **Nash inequality $\Longrightarrow L^1$-exponential convergence and π-a.s. Strong ergodicity [Chen (1999b)].** Denote by $\|\cdot\|_{p\to q}$ the operator's norm from $L^p(\pi)$ to $L^q(\pi)$. Note that

$$\begin{aligned}\text{Nash inequality} &\Longleftrightarrow \mathrm{Var}(P_t(f)) = \|P_t f - \pi(f)\|_2^2 \le C^2\|f\|_1^2/t^{q-1} \\ &\qquad\qquad (q := \nu/2+1) \\ &\Longleftrightarrow \|(P_t - \pi)f\|_2 \le C\|f\|_1/t^{(q-1)/2}. \\ &\Longleftrightarrow \|P_t - \pi\|_{1\to 2} \le C/t^{(q-1)/2}.\end{aligned}$$

Since $\|P_t - \pi\|_{1\to 1} \le \|P_t - \pi\|_{1\to 2}$, we have

$$\text{Nash inequality} \Longrightarrow L^1\text{-algebraic convergence.}$$

Furthermore, because of the semigroup property, the convergence of $\|\cdot\|_{1\to 1}$ must be exponential, we indeed have

$$\text{Nash inequality} \Longrightarrow L^1\text{-exponential convergence.}$$

In the symmetric case: $P_t - \pi = (P_t - \pi)^*$, and so

$$\|P_{2t} - \pi\|_{1\to\infty} \le \|P_t - \pi\|_{1\to 2}\|P_t - \pi\|_{2\to\infty} = \|P_t - \pi\|_{1\to 2}^2.$$

Hence, $\|P_t - \pi\|_{1\to\infty} \le C/t^{q-1}$. Thus,

$$\begin{aligned}&\operatorname{ess\,sup}_x\|P_t(x,\cdot) - \pi\|_{\mathrm{Var}} = \operatorname{ess\,sup}_x \sup_{|f|\le 1} |(P_t(x,\cdot) - \pi)f| \\ &\le \operatorname{ess\,sup}_x \sup_{\|f\|_1\le 1} |(P_t(x,\cdot) - \pi)f| = \sup_{\|f\|_1\le 1} \operatorname{ess\,sup}_x|(P_t(x,\cdot) - \pi)f| \\ &= \|P_t - \pi\|_{1\to\infty} \le C/t^{q-1} \to 0, \qquad \text{as } t\to\infty.\end{aligned}$$

This gives us the π-a.s. strong ergodicity.

(b) **L^1-exponential convergence $\Longleftrightarrow$ π-a.s. Strong ergodicity [Mao (2002c)].** Since $(L^1)^* = L^\infty \Longrightarrow \|P_t - \pi\|_{1\to 1} = \|P_t^* - \pi\|_{\infty\to\infty}$ and $P_t^*(x,\cdot) \ll \pi$, we have

$$\begin{aligned}\|P_t^* - \pi\|_{\infty\to\infty} &= \operatorname{ess\,sup}_x \sup_{\|f\|_\infty = 1} |(P_t^* - \pi)f(x)| \\ &= \operatorname{ess\,sup}_x \sup_{\sup|f|=1} |(P_t^* - \pi)f(x)| \\ &= \operatorname{ess\,sup}_x\|P_t^*(x,\cdot) - \pi\|_{\mathrm{Var}}.\end{aligned}$$

Hence, π-a.s. strong ergodicity is exactly the same as the L^1-exponential convergence. Without condition "$P_t^*(x,\cdot) \ll \pi$", the second equality becomes "$\geq$", and so we have in the general reversible case that

$$L^1\text{-exponential convergence} \Longrightarrow \pi\text{-a.s. Strong ergodicity.}$$

(c) **Nash inequality $\Longrightarrow$ Logarithmic Sobolev inequality** [**Chen (1999b)**]. Because $\|f\|_1 \leq \|f\|_p$ for all $p \geq 1$, we have $\|\cdot\|_{2\to 2} \leq \|\cdot\|_{1\to 2} \leq C/t^{(q-1)/2}$, and so

$$\text{Nash inequality} \Longrightarrow \text{Poincaré inequality} \Longleftrightarrow \lambda_1 > 0.$$

$$\|P_t\|_{p\to 2} \leq \|P_t\|_{1\to 2} \leq \|P_t - \pi\|_{1\to 2} + \|\pi\|_{1\to 2} < \infty, \qquad p \in (1,2).$$

The assertion now follows from [Bakry (1992); Theorem 3.6 and Proposition 3.9].

The remainder of the Appendix is devoted to the proof of the assertion:

$$(A1)\quad L^2\text{-exponential convergence} \Longleftrightarrow \pi\text{-a.s. Exponential ergodicity.}$$

Actually, this is done by Chen (2000a). Because, by assumption, the process is reversible and $P_t(x,\cdot) \ll \pi$. Set $p_t(x,y) = \frac{\mathrm{d}P_t(x,\cdot)}{\mathrm{d}\pi}(y)$. Then we have $p_t(x,y) = p_t(y,x)$, $\pi \times \pi$-a.s. (x,y). Hence

$$(A2)\qquad \int p_s(x,y)^2\pi(\mathrm{d}y) = \int p_s(x,y)p_s(y,x)\pi(\mathrm{d}y) = p_{2s}(x,x) < \infty$$

(Carlen et al (1987)).

This means that $p_t(x,\cdot) \in L^2(\pi)$ for all $t > 0$ and π-a.s. $x \in E$. Thus, by [Chen (2000a); Theorem 1.2] and the remarks right after the theorem, $(A1)$ holds.

The proof above is mainly based on the time-discrete analog result by Roberts and Rosenthal (1997). Here, we present a more direct proof of $(A2)$ as follows.

(d) L^2-exponential convergence $\Longrightarrow$ π-a.s. Exponential ergodicity [Chen (1991, 1998, 2000a)]. Let $\mu \ll \pi$. Then

$$\begin{aligned}
\|\mu P_t - \pi\|_{\mathrm{Var}} &= \sup_{|f|\le 1} |(\mu P_t - \pi) f| = \sup_{|f|\le 1} \left|\pi\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi} P_t f - f\right)\right| \\
&= \sup_{|f|\le 1} \left|\pi\left(f P_t^*\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right) - f\right)\right| \\
&= \sup_{|f|\le 1} \left|\pi\left[f\left(P_t^*\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\right)\right)\right]\right| \\
&\le \left\|P_t^*\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\right)\right\|_1 \le \left\|\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\right\|_2 e^{-t\,\mathrm{gap}(L^*)} \\
&= \left\|\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\right\|_2 e^{-t\,\mathrm{gap}(L)}.
\end{aligned} \tag{A3}$$

We now consider two cases separately.

In the reversible case with $P_t(x, \cdot) \ll \pi$, by $(A2)$, we have

$$\begin{aligned}
\|P_t(x,\cdot) - \pi\|_{\mathrm{Var}} &\le \left\|P_{t-s}\left(\frac{\mathrm{d}P_s(x,\cdot)}{\mathrm{d}\pi} - 1\right)\right\|_1 \\
&\le \|p_s(x,\cdot) - 1\|_2 e^{-(t-s)\,\mathrm{gap}(L)} \\
&= \left[\sqrt{p_{2s}(x,x) - 1}\, e^{s\,\mathrm{gap}(L)}\right] e^{-t\,\mathrm{gap}(L)}, \qquad t \ge s.
\end{aligned} \tag{A4}$$

Therefore, there exists $C(x) < \infty$ such that

$$\|P_t(x,\cdot) - \pi\|_{\mathrm{Var}} \le C(x) e^{-t\,\mathrm{gap}(L)}, \qquad t \ge 0, \quad \pi\text{-a.s. } (x). \tag{A5}$$

Denote by ε_1 be the largest ε such that $\|P_t(x,\cdot) - \pi\|_{\mathrm{Var}} \le C(x) e^{-\varepsilon t}$ for all t. Then $\varepsilon_1 \ge \mathrm{gap}(L) = \lambda_1$.

In the φ-irreducible case, without using the reversibility and transition density, from $(A3)$, one can still derive π-a.s. exponential ergodicity (but may have different rates). Refer to Roberts and Tweedie (2001) for a proof in the time-discrete situation (the title of the quoted paper is confused, where the term "L^1-convergence" is used for the π-a.s. exponentially ergodic convergence, rather than the standard meaning of L^1-exponential convergence used in this paper. These two types of convergence are essentially different as shown in Theorem 9.1). In other words, the reversibility and the existence of the transition density are not essential in this implication.

(e) π-a.s. Exponential ergodicity $\Longrightarrow$ L^2-exponential convergence [Chen (2000a), Mao (2002c)]. In the time-discrete case, a similar assertion was proved by Roberts and Rosenthal(1997) and so can be extended to the time-continuous case by using the standard technique [cf., Chen (1992), §4.4]. The proof given below provides more precise estimates. Let the σ-algebra $\mathscr{E}$ be countably generated. By Numemelin and P. Tuominen (1982) or [Numemelin (1984); Theorem 6.14 (iii)], we have in the time-discrete case that

$$\begin{aligned}&\pi\text{-a.s. geometrically ergodic convergence}\\ &\qquad\Longleftrightarrow \| \|P^n(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1 \text{ geometric convergence,}\end{aligned}$$

here and in what follows, the L^1-norm is taken with respect to the variable "$\bullet$". This implies in the time-continuous case that

$$\begin{aligned}&\pi\text{-a.s. exponentially ergodic convergence}\\ &\qquad\Longleftrightarrow \| \|P_t(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1 \text{ exponential convergence.}\end{aligned}$$

Assume that $\| \|P_t(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1 \le Ce^{-\varepsilon_2 t}$ with largest ε_2.

We now prove that $\| \|P_t(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1 \ge \|P_t-\pi\|_{\infty\to 1}$. Let $\|f\|_\infty = 1$. Then

$$\begin{aligned}\|(P_t-\pi)f\|_1 &= \int \pi(\mathrm{d}x)\Big|\int \Big[P_t(x,\mathrm{d}y)-\pi(\mathrm{d}y)\Big]f(y)\Big|\\ &\le \int \pi(\mathrm{d}x) \sup_{\|g\|_\infty\le 1}\Big|\int \Big[P_t(x,\mathrm{d}y)-\pi(\mathrm{d}y)\Big]g(y)\Big|\\ &= \| \|P_t(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1\\ &\qquad\qquad \text{(Need } P_t(x,\cdot)\ll\pi \text{ or reversibility!).}\end{aligned}$$

Next, we prove that $\|P_{2t}-\pi\|_{\infty\to 1} = \|P_t-\pi\|^2_{\infty\to 2}$ in the reversible case. We have

$$\begin{aligned}\|(P_t-\pi)f\|_2^2 &= ((P_t-\pi)f,(P_t-\pi)f) = (f,(P_t-\pi)^2 f)\\ &= (f,(P_{2t}-\pi)f) \le \|f\|_\infty \|(P_{2t}-\pi)f\|_1\\ &\le \|f\|_\infty^2 \|P_{2t}-\pi\|_{\infty\to 1}.\end{aligned}$$

Hence $\|P_{2t}-\pi\|_{\infty\to 1} \ge \|P_t-\pi\|^2_{\infty\to 2}$. The inverse inequality is obvious by using the semigroup property and symmetry: $\|P_{2t}-\pi\|_{\infty\to 1} \le \|P_t-\pi\|_{\infty\to 2}\|P_t-\pi\|_{2\to 1} = \|P_t-\pi\|^2_{\infty\to 2}$.

We remark that in general case, without reversibility, we have $\|P_t-\pi\|_{\infty\to 1} \ge \|P_t-\pi\|^2_{\infty\to 2}/2$. Actually,

$$\begin{aligned}\|(P_t-\pi)f\|_2^2 &\le \int |(P_t-\pi)f|^2\mathrm{d}\pi \le 2\|f\|_\infty \int |(P_t-\pi)f|\mathrm{d}\pi\\ &\le 2\|f\|_\infty^2 \|P_t-\pi\|_{\infty\to 1}, \qquad f\in L^\infty(\pi).\end{aligned}$$

Finally, assume that the process is reversible. We prove that $\lambda_1 = \text{gap}(L) \geq \varepsilon_2$. We have just proved that for every f with $\pi(f) = 0$ and $\|f\|_2 = 1$, $\|P_t f\|_2^2 \leq C\|f\|_\infty^2 e^{-2\varepsilon_2 t}$. Following [Wang (2000; Lemma 2.2), or Röckner and Wang (2001)], by the spectral representation theorem, we have

$$\begin{aligned}\|P_t f\|_2^2 &= \int_0^\infty e^{-2\lambda t}\text{d}(E_\lambda f, f)\\ &\geq \bigg[\int_0^\infty e^{-2\lambda s}\text{d}(E_\lambda f, f)\bigg]^{t/s} \qquad \text{(by Jensen inequality)}\\ &= \|P_s f\|_2^{2t/s}, \qquad t \geq s.\end{aligned}$$

Thus, $\|P_s f\|_2^2 \leq \Big[C\|f\|_\infty^2\Big]^{s/t} e^{-2\varepsilon_2 s}$. Letting $t \to \infty$, we get

$$\|P_s f\|_2^2 \leq e^{-2\varepsilon_2 s}, \qquad \pi(f) = 0,\ \|f\|_2 = 1,\ f \in L^\infty(\pi).$$

Since $L^\infty(\pi)$ is dense in $L^2(\pi)$, we have

$$\|P_s f\|_2^2 \leq e^{-2\varepsilon_2 s}, \quad s \geq 0, \quad \pi(f) = 0,\ \|f\|_2 = 1.$$

Therefore, $\lambda_1 \geq \varepsilon_2$. Q.E.D.

Remark A1. Note that when $p_{2s}(\cdot,\cdot) \in L^{1/2}(\pi)$ (in particular, when $p_{2s}(x,x)$ is bounded in x) for some $s > 0$, from ($A4$), it follows that there exists a constant C such that $\|\|P_t(\bullet,\cdot) - \pi\|_{\text{Var}}\|_1 \leq Ce^{-\lambda_1 t}$. Then, we have $\varepsilon_2 \geq \lambda_1$. Combining this with (e), we indeed have $\lambda_1 = \varepsilon_2$.

Remark A2. It is proved by Hwang et al (2002) that under mild condition, in the reversible case, $\lambda_1 = \varepsilon_1$. Refer also to Wang (2002) for related estimates.

Final remark. The main body of this paper is an updated version of Chen (2001c), which was written at the beginning stage of the study on seeking explicit criteria. The resulting picture is now quite complete and so the most parts of the original paper has to be changed, except the first section. This paper also refines a part of Chen (2002c).

Acknowledgement. This paper is based on the talks given at "Stochastic analysis on large scale interacting systems", Shonan Village Center, Hayama, Japan (July 17–26, 2002), "Stochastic analysis and statistical mechanics", Yukawa Institute, Kyoto University, Japan (July 29–30, 2002), the "First Sino-German Conference on Stochastic Analysis—A Satellite Conference of ICM 2002", The Sino-German Center, Beijing (August 29–September 3, 2002), and "Stochastic analysis in infinite dimensional spaces", RIMS, Kyoto University, Japan (November 6–8, 2002). The author is grateful for the kind invitation, financial support and the warm hospitality made by the organization committees: Profs. T. Funaki, H. Osada, N. Yosida, T. Kumagai and their colleagues and students; Profs. S. Albeverio, Z. M. Ma and M. Röckner; Profs. S. Aida, I. Shigekawa, Y. Takahashi and their colleagues.

The materials of the paper are included in a mini-course presented at the "Second Sino-French Colloquium in Probability and Applications", Wuhan, China (April 9–19, 2001), and included in the talks presented at Institute of Mathematics, Academy Sinica, Taipei (April 19–May 1, 2001), and in Department of Mathematics, Kyoto University, Kyoto, Japan (October 17, 2002–January 17, 2003). The author would like to acknowledge Prof. L. M. Wu; Profs. C. R. Hwang, T. S. Chiang, Y. S. Chow and S. J. Sheu; Profs. I. Shigekawa and Y. Takahashi for the financial support and the warm hospitality.

References

[1] Anderson, W. J. (1991), *Continuous-Time Markov Chains*, Springer Series in Statistics, Springer, Berlin.

[2] Bakry, D. (1992), *L'hypercontractivité et son utilisation en théorie des semigroupes*, LNM, **1581** Springer.

[3] Bakry, D., Coulhon, T., Ledoux, M., Saloff-Coste, L. (1995), *Sobolev inequalities in disguise*, Indiana Univ. Math. J. **44**(4), 1033–1074.

[4] Bobkov, S. G. and Götze, F. (1999a), *Discrete isoperimetric and Poincaré inequalities*, Prob. Th. Rel. Fields **114**, 245–277.

[5] Bobkov, S. G. and Götze, F. (1999b), *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163**, 1–28.

[6] Carlen, E. A., Kusuoka, S., Stroock, D. W. (1987), *Upper bounds for symmetric Markov transition functions*, Ann. Inst. Henri Poincaré **2**, 245–287.

[7] Chen, M. F. (1991), *Exponential L^2-convergence and L^2-spectral gap for Markov processes*, Acta Math. Sin. New Ser. **7**(1), 19–37.

[8] Chen, M. F. (1992), *From Markov Chains to Non-Equilibrium Particle Systems*, World Scientific, Singapore.

[9] Chen, M. F. (1996), *Estimation of spectral gap for Markov chains*, Acta Math. Sin. New Ser. **12**(4), 337-360.

[10] Chen, M. F. (1998), *Estimate of exponential convergence rate in total variation by spectral gap*, Acta Math. Sin. Ser. (A) 41:1 (Chinese Ed.), 1–6; Acta Math. Sin. New Ser. **14**(1), 9–16.

[11] Chen, M. F. (1999a), *Analytic proof of dual variational formula for the first eigenvalue in dimension one*, Sci. Sin. (A) **42**:8, 805–815.

[12] Chen, M. F. (1999b), *Nash inequalities for general symmetric forms*, Acta Math. Sin. Eng. Ser. **15**:3, 353–370.

[13] Chen, M. F. (1999c), *Eigenvalues, inequalities and ergodic theory* (II) , Advances in Math. **28**:6, 481–505.

[14] Chen, M. F. (2000a), *Equivalence of exponential ergodicity and L^2-exponential convergence for Markov chains*, Stoch. Proc. Appl. **87**, 281–297.

[15] Chen, M. F. (2000b), *Explicit bounds of the first eigenvalue*, Sci. in China, Ser. A, **43**(10), 1051–1059.

[16] Chen, M. F. (2001a), *Variational formulas and approximation theorems for the first eigenvalue in dimension one*, Sci. Chin. (A) **44**(4), 409–418.

[17] Chen, M. F. (2001b), *Ergodic Convergence Rates of Markov Processes — Eigenvalues, Inequalities and Ergodic Theory* [*Collection of papers,* 1993—2001], http://www.bnu.edu.cn/~chenmf/main_eng.htm.

[18] Chen, M. F. (2001c), *Explicit criteria for several types of ergodicity*, Chin. J. Appl. Prob. Stat. **17**(2), 1–8.

[19] Chen, M. F. (2002a), *Variational formulas of Poincaré-type inequalities in Banach spaces of functions on the line*, Acta Math. Sin. Eng. Ser. **18**(3), 417–436.

[20] Chen, M. F. (2002b), *A new story of ergodic theory*, in "Applied Probability", 25-34, eds. R. Chan et al., AMS/IP Studies in Advanced Mathematics, **26**, 2002.

[21] Chen, M. F. (2002c), *Ergodic Convergence Rates of Markov Processes — Eigenvalues, Inequalities and Ergodic Theory*, in Proceedings of "ICM 2002", **III**, 25–40, Higher Education Press, Beijing.

[22] Chen, M. F. (2002d), *Variational formulas of Poincaré-type inequalities for birth-death processes*, preprint, submitted to Acta Math. Sin. Eng. Ser.

[23] Chen, M. F. (2002e), *Variational formulas of Poincaré-type inequalities for one-dimensional processes*, to appear in "IMS volume dedicated to Rabi Bhattacharya".

[24] Chen, M. F., Scacciatelli, E. and Yao, L. (2002), *Linear approximation of the first eigenvalue on compact manifolds*, Sci. Sin. (A) **45**(4), 450–461.

[25] Chen, M. F. and Wang, F. Y. (1997a), *Estimation of spectral gap for elliptic operators*, Trans. Amer. Math. Soc., **349**(3), 1239-1267.

[26] Chen, M. F. and Wang, F. Y. (1997b), *General formula for lower bound of the first eigenvalue*, Sci. Sin. **40**(4), 384–394.

[27] Chen, M. F. and Wang, Y. Z. (2000), *Algebraic Convergence of Markov Chains*, to appear in Ann. Appl. Probab.

[28] Davies, E. B. (1999), *A review of Hardy inequality*, Operator Theory: Adv. & Appl. **110**, 55–67.

[29] Van Doorn, E. (1981), *Stochastic Monotonicity and Queuing Applications of Birth-Death Processes*, Lecture Notes in Statistics **4**, Springer.

[30] Van Doorn, E. (1985), *Conditions for exponential ergodicity and bounds for the decay parameter of a birth-death process*, Adv. Appl. Prob. **17**, 514–530.

[31] Van Doorn, E. (1987), *Representations and bounds for zeros of orthogonal polynomials and eigenvalues of sign-symmetric tri-diagonal matrices*, J. Approx. Th. **51**, 254–266.

[32] Van Doorn, E. (1991), *Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes*, Adv. Appl. Prob. **23**, 683–700.

[33] Dynkin, E. M. (1990), *EMS*, Springer-Valerg (1991), Berlin.

[34] Egorov, Y. and Kondratiev, V. (1996), *On Spectral Theory of Elliptic Operators*, Birkbäuser, Berlin.

[35] Hwang, C. R., Hwang-Ma, S. Y. and Sheu, S. J. (2002), *Accelerating diffusions*, preprint.

[36] Hardy, G. H. (1920), *Note on a theorem of Hilbert*, Math. Zeitschr. **6**, 314–317.

[37] Hou, Z. T. et al (2000) *Birth-death Processes*, Hunan Sci. Press, Hunan.

[38] Kac, I. S. and Krein, M. G. (1958), *Criteria for discreteness of the spectrum of a singular string*, Izv. Vyss. Učebn. Zaved. Mat. **2**, 136–153 (In Russian).

[39] Kipnis, C. and Landim, C. (1999), *Scaling Limits of Interacting Particle Systems*, Springer-Verlag, Berlin.

[40] Kotani, S. and Watanabe, S. (1982), *Krein's spectral theory of strings and generalized diffusion processes*, Lecture Notes in Math. **923**, 235–259.

[41] Landau, E. (1926), *A note on a theorem concerning series of positive term*, J. London Math. Soc. **1**, 38–39.

[42] Landim, C., Sethuraman, S. and Varadhan, S. R. S. (1996) , *Spectral gap for zero range dynamics*, Ann. Prob. **24**, 1871–1902.

[43] Liggett, T. M. (1989), *Exponential L^2 convergence of attractive reversible nearest particle systems*, Ann. Prob. **17**, 403–432.

[44] Liggett, T. M. (1991), *L^2 Rates of convergence for attractive reversible nearest particle systems: the critical case*, Ann. Prob. **19**(3), 935–959.

[45] Luo, J. H. (1992), *On discrete analog of Poincaré-type inequalities and density representation*, preprint (unpublished).

[46] Mao, Y. H. (2000), *On empty essential spectrum for Markov processes in dimension one*, preprint.

[47] Mao, Y. H. (2001), *Strong ergodicity for Markov processes by coupling methods*, preprint.

[48] Mao, Y. H. (2002a), *The logarithmic Sobolev inequalities for birth-death process and diffusion process on the line*, Chin. J. Appl. Prob. Statis. **18**(1), 94–100.

[49] Mao, Y. H. (2002b), *Nash inequalities for Markov processes in dimension one*, Acta. Math. Sin. Eng. Ser. **18**(1), 147–156.

[50] Mao, Y. H. (2002c), *In preparation.*

[51] Mao, Y. H. and Zhang, S. Y. (2000), *Comparison of some convergence rates for Markov process* (In Chinese), Acta. Math. Sin. **43**(6), 1019–1026.

[52] Mazya, V. G. (1985), *Sobolev Spaces*, Springer-Valerg.

[53] Meyn, S. P. and Tweedie, R. L. (1993), *Markov Chains and Stochastic Stability*, Springer, London.

[54] Miclo, L. (1999a), *Relations entre isopérimétrie et trou spectral pour les chaînes de Markov finies*, Prob. Th. Rel. Fields **114**, 431–485.

[55] Miclo, L. (1999b), *An example of application of discrete Hardy's inequalities*, Markov Processes Relat. Fields **5**, 319–330.

[56] Muckenhoupt, B. (1972), *Hardy's inequality with weights* Studia Math. **XLIV**, 31–38.

[57] Nummelin, E. (1984), *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge Univ. Press.

[58] Nummelin, E. and Tuominen, P. (1982), *Geometric ergodicity of Harris recurrent chains with applications to renewal theory*, Stoch. Proc. Appl. **12**, 187-202.

[59] Opic, B. and Kufner, A. (1990), *Hardy-type Inequalities*, Longman, New York.

[60] Roberts, G. O. and Rosenthal, J. S. (1997), *Geometric ergodicity and hybrid Markov chains*, Electron. Comm. Probab. **2**, 13–25.

[61] Roberts, G. O. and Tweedie, R. L. (2001), *Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains*, J. Appl. Probab. **38**(A), 37–41.

[62] Röckner, M. and Wang, F. Y. (2001), *Weak Poincaré inequalities and L^2-convergence rates of Markov semigroups*, J. Funct. Anal. **185**(2), 564–603.

[63] Sullivan, W. G. (1984), *The L^2 spectral gap of certain positive recurrent Markov chains and jump processes* Z. Wahrs. **67**, 387–398.

[64] Tweedie, R. L. (1981), *Criteria for ergodicity, exponential ergodicity and strong ergodicity of Markov processes*, J. Appl. Prob. **18**, 122–130.

[65] Varopoulos, N. (1985), *Hardy-Littlewood theory for semigroups*, J. Funct. Anal. **63**, 240–260.

[66] Wang, F. Y. (2000), *Functional inequalities, symmegroup properties and spectrum estimates*, Infinite Dim. Anal., Quantum Probab. and related Topics, **3**:2, 263–295.

[67] Wang, F. Y. (2001), *Convergence rates of Markov semigroups in probability distances*, preprint.

[68] Wang, F. Y. (2002), *Coupling,convergence rate of Markov processes and weak Poincaré inequalities*, Sci. in China (A), **45**:8, 975–983.

[69] Wang, Z. K. (1980), *Birth-death Processes and Markov Chains*, Science Press, Beijing (In Chinese).

[70] Yang, X. Q. (1986), *Constructions of Time-homogeneous Markov Processes with Denumerable States*, Hunan Sci. Press, Hunan.

[71] Zhang, S. Y. and Mao, Y. H. (2000), *Exponential convergence rate in Boltzman-Shannon entropy*, Sci. Sin. (A) **44**(3), 280–285.

[72] Zhang, Y. H. (2001), *Strong ergodicity for single birth processes*, J. Appl. Prob. **38**, 270–277.

Department of Mathematics
Beijing Normal University
Beijing 100875
The People's Republic of China
E-mail address: `mfchen@bnu.edu.cn`

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 115–142

Probabilistic Analysis of Directed Polymers in a Random Environment: a Review

Francis Comets, Tokuzo Shiga, and Nobuo Yoshida

Abstract.

Directed polymers in random environment can be thought of as a model of statistical mechanics in which paths of stochastic processes interact with a quenched disorder (impurities), depending on both time and space. We review here main results which have been obtained during the last fifteen years, with proofs to most of the results. The material covers the diffusive behavior of the polymers in weak disorder phase studied by J. Imbrie, T. Spencer, E. Bolthausen, R. Song and X. Y. Zhou [11, 3, 25], and localization of the paths in strong disordered phase recently obtained by P. Carmona, Y. Hu, and the authors of the present article [4, 5].

CONTENTS

1. Introduction

1.1. Physical background

We start with an informal description of the situation we will discuss in these notes. Imagine a hydrophilic polymer chain wafting in water. Due to the thermal fluctuation, the shape of the polymer should be understood as a random object. We now suppose that the water contains randomly placed hydrophobic molecules as impurities, which repel the

---

Received March 28, 2003.
Revised June 10, 2003.

hydrophilic monomers which the polymer consists of. The question we address here is:

(1.1)
How do the impurities affect the global shape of the polymer chain?

We try to answer this question in a mathematical framework. However, as is everywhere else in mathematical physics, it is very difficult to do so without compromising with a rather simplified picture of the real world. Here, our simplification goes as follows. We first suppress entanglement and U-turns of the polymer; we shall represent the polymer chain as a graph $\{(j,\omega_j)\}_{j=1}^n$ in $\mathbb{N} \times \mathbb{Z}^d$, so that the polymer is supposed to live in $(1+d)$-dimensional discrete lattice and to stretch in the direction of the first coordinate. Each point $(j,\omega_j) \in \mathbb{N} \times \mathbb{Z}^d$ on the graph stands for the position of j-th monomer in this picture. Secondly, we assume that, the transversal motion $\{\omega_j\}_{j=1}^n$ performs a simple random walk in $\mathbb{Z}^d$, if the impurities are absent. We then define the energy of the path $\{(j,\omega_j)\}_{j=1}^n$ by

$$-\beta \sum_{j=1}^{n} \eta(j,\omega_j) \; , \tag{1.2}$$

where $\beta = 1/T > 0$ is the inverse temperature and $\{\eta(n,x) : n \geq 1, \; x \in \mathbb{Z}^d\}$ is a real i.i.d. random variables, with $\eta(n,x)$ describing the presence (or strength) of an impurity at site (n,x). The typical shape $\{(j,\omega_j)\}_{j=1}^n$ of the polymer is then given by the one that minimizes the energy (1.2). Let us suppose for example that $\eta(n,x)$ takes two different values $+1$ ("presence of a water molecule at (n,x)") and -1 ("presence of the hydrophobic impurity at (n,x)"). Then, the energy of the polymer is increased by $+\beta$ each time a monomer is in contact with the impurity $(\eta(j,\omega_j) = -1)$. Therefore, the typical shape of the polymer for each given configuration of $\{\eta(j, \; x)\}$ is given by the one which tries to avoid the impurities as much as possible. The purpose of these notes is to introduce rigorous results which answer (1.1) roughly as follows.

- **(a):** If $d \geq 3$ and β small enough, the impurities do not affect the global shape of the polymer (*the weak disorder phase*).
- **(b):** If either (i): $d \leq 2$ and $\beta \neq 0$ or (ii): $d \geq 3$ and β large enough[1], then, the impurities change the global shape of the polymer drastically (*the strong disorder phase*).

---

[1] To be precise, there are some exceptions. See Remark 2.2.1 below.

1.2. Simple random walk model for directed polymers

We now put the informal description given in section 1.1 into a mathematical framework. As we mentioned before, the framework can be thought of as a model in statistical mechanics. However, no prior knowledge of statistical mechanics is needed in this paper. The model we consider here is defined as a random walk in a random environment. We first fix notation for the random walk and the random environment. Then, we introduce the polymer measure.

- *The random walk:* $(\{\omega_n\}_{n\geq 0}, P)$ is a simple random walk on the d-dimensional integer lattice $\mathbb{Z}^d$. More precisely, we let Ω be the path space $\Omega = \{\omega = (\omega_n)_{n\geq 0}; \omega_n \in \mathbb{Z}^d, n \geq 0\}$, $\mathcal{F}$ be the cylindrical σ-field on Ω, and, for all $n \geq 0$, $\omega_n : \omega \mapsto \omega_n$ be the projection map. We consider the unique probability measure P on $(\Omega, \mathcal{F})$ such that $\omega_1 - \omega_0, \dots, \omega_n - \omega_{n-1}$ are independent and

$$P\{\omega_0 = 0\} = 1, \quad P\{\omega_n - \omega_{n-1} = \pm\delta_j\} = (2d)^{-1}, \quad j = 1, 2, \dots, d,$$

where $\delta_j = (\delta_{kj})_{k=1}^d$ is the j-th vector of the canonical basis of $\mathbb{Z}^d$. In the sequel, $P[X]$ denotes the P-expectation of a r.v.(random variable) X on $(\Omega, \mathcal{F}, P)$.

- *The random environment:* $\eta = \{\eta(n,x) : n \in \mathbb{N}, \ x \in \mathbb{Z}^d\}$ is a sequence of r.v.'s which are real valued, non-constant, and i.i.d.(independent identically distributed) r.v.'s defined on a probability space $(H, \mathcal{G}, Q)$ such that

$$Q[\exp(\beta\eta(n,x))] < \infty \quad \text{for all } \beta \in \mathbb{R}. \tag{1.3}$$

Here, and in the sequel, $Q[Y]$ denotes the Q-expectation of a r.v. Y on $(H, \mathcal{G}, Q)$.

- *The polymer measure:* For any $n > 0$, define the probability measure μ_n on the path space $(\Omega, \mathcal{F})$ by

$$\mu_n(d\omega) = \frac{1}{Z_n} \exp\left(\beta \sum_{1\leq j\leq n} \eta(j, \omega_j)\right) P(d\omega), \tag{1.4}$$

where $\beta > 0$ is a parameter (the inverse temperature) and

$$Z_n = P\left[\exp\left(\beta \sum_{1\leq j\leq n} \eta(j, \omega_j)\right)\right] \tag{1.5}$$

is the normalizing constant (the partition function).

The polymer measure μ_n can be thought of as a Gibbs measure on the path space $(\Omega, \mathcal{F})$ with the Hamiltonian (1.2). We stress that the random environment η is contained in both Z_n and μ_n without being integrated out, so that they are r.v.'s on the probability space $(H, \mathcal{G}, Q)$. The polymer is attracted to sites where the random environment is positive, and repelled by sites where the environment is negative.

Remark 1.2.1. This model was originally introduced in physics literature [10] to mimic the phase boundary of Ising model subject to random impurities. Later on, the model reached the mathematics community [11, 3], where it was reformulated as above.

Here are two standard choices for the distribution of $\eta(n, x)$.

Example 1.2.1. *Bernoulli environment* ([3, 11, 25]); This is the case with

$$Q\{\eta(n,x) = -1\} = p > 0, \quad Q\{\eta(n,x) = +1\} = 1 - p > 0.$$

In the physical picture described in section 1.1, $\eta(n,x) = -1$ (resp. $\eta(n,x) = +1$) can be interpreted as the presence of a hydrophobic impurity (resp. a water molecule) at site (x, n).

Example 1.2.2. *Gaussian environment* ([4]); This is the case in which $\eta(n,x)$ is a standard normal random variable;

$$Q\{\eta(n,x) \in dt\} = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt.$$

2. Some typical results for the simple random walk model

In this section, we present some typical results for the simple random walk model. Here, we focus on the conceptual issues on these results and do not go into the proofs.

We now introduce an important quantity for this model, which appears in the assumptions of the results we present. Let $\lambda(\beta)$ be the logarithmic moment generating function of $\eta(n, x)$,

$$\lambda(\beta) = \ln Q[\exp(\beta\eta(n,x))], \quad \beta \in \mathbb{R}. \tag{2.1}$$

The function $\lambda(\beta)$ can be explicitly computed for some typical choice of the distribution of $\eta(n,x)$. For example, $\lambda(\beta) = \ln(pe^{-\beta} + (1-p)e^{\beta})$ for the Bernoulli environment (Example 1.2.2) and $\lambda(\beta) = \frac{1}{2}\beta^2$ for the Gaussian environment (Example 1.2.2).

2.1. The weak disorder phase

The results we present in this subsection show that the impurities do not change the transversal fluctuation of the polymer if $d \geq 3$ and β is small enough. We first recall the following fact about the return probability π_d for the simple random walk:

$$\pi_d \overset{\text{def.}}{=} P\{\omega_n = 0 \text{ for some } n \geq 1\} \begin{cases} = 1 & \text{if } d \leq 2, \\ < 1 & \text{if } d \geq 3. \end{cases} \tag{2.2}$$

More precisely, it is known that $\pi_{d+1} < \pi_d$ for all $d \geq 3$ [22, Lemma 1] and that $\pi_3 = 0.3405...$ [26, page 103]. In particular, $\pi_d \leq 0.3405...$ for all $d \geq 3$.

Theorem 2.1.1. **(The diffusive behavior;** [11, 3, 25]) *Suppose that $d \geq 3$ (hence $\pi_d < 1$) and that*

$$\gamma_1(\beta) \overset{\text{def.}}{=} \lambda(2\beta) - 2\lambda(\beta) < \ln(1/\pi_d). \tag{2.3}$$

Then,

$$\lim_{n \nearrow \infty} \mu_n[|\omega_n|^2]/n = 1 \quad Q\text{-}a.s. \tag{2.4}$$

Note that $\gamma_1(\beta)$ is increasing on $[0, \infty)$ and $\gamma_1(0) = 0$ so that the condition in (2.3) does hold if β is small. Proof of Theorem 2.1.1 is given in section 3.2

Example 2.1.1. Consider the Bernoulli environment (Example 1.2.1). In this case, it is not difficult to see from direct computations that $\lim_{\beta \nearrow \infty} \gamma_1(\beta) = -\ln(1-p)$. This shows that (2.3) holds for all $\beta \geq 0$ if $p < 1 - \pi_d$.

Example 2.1.2. Consider the Gaussian environment (Example 1.2.2). Then, $\gamma_1(\beta) = \beta^2$ and hence (2.3) holds if $\beta < \sqrt{\ln(1/\pi_d)}$.

Remark 2.1.1. The first rigorous proof of Theorem 2.1.1 was obtained by J. Z. Imbrie and T. Spencer [11] in the case of Bernoulli environment. Soon afterwards, a more transparent proof based on the martingale analysis was given by E. Bolthausen [3]. The martingale proof was then extended to general environment under condition (2.3) by R. Song and X. Y. Zhou [25]. By the argument in [3, 25], it is possible to get a much more precise statement than (2.4). In fact, under the same assumption in Theorem 2.1.1, the following central limit theorem holds;

for all $f \in C(\mathbb{R}^d)$ with at most polynomial growth at infinity,

(2.5)

$$\lim_{n \nearrow \infty} \mu_n \left[f\left(\omega_n/\sqrt{n}\right)\right] = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f\left(x/\sqrt{d}\right) \exp(-|x|^2/2)dx, \quad Q\text{-a.s.}$$

The diffusive behavior (2.4) follows from (2.5) by choosing $f(x) = |x|^2$. In [3], (2.5) is obtained for the Bernoulli environment only. However, with the help of the observation made in [25], it is not difficult to extend the central limit theorem to general environment under the assumption in Theorem 2.1.1. We will sketch the proof of (2.5) in Remark 3.2.4 below.

We now recall the following well known fact for the simple random walk, i.e., the case of $\beta = 0$;

$$\max_{x \in \mathbb{Z}^d} P\{\omega_n = x\} = \mathcal{O}(n^{-d/2}), \quad \text{as } n \nearrow \infty. \tag{2.6}$$

The decay rate $n^{-d/2}$ in (2.6) can be understood as a manifestation of the fact that the possible position of ω_n is spread over a ball in $\mathbb{Z}^d$ with radius const. × $\sqrt{n}$.

For $\beta \neq 0$, we can still prove (2.6) in a weaker form as follows.

Theorem 2.1.2. **(Delocalization;**[4, 5]**)** *Suppose that $d \geq 3$ and that β is small enough so that (2.3) holds. Then,*

$$\sum_{n \geq 1} \max_{x \in \mathbb{Z}^d} \mu_{n-1}\{\omega_n = x\}^2 < \infty, \quad Q\text{-}a.s. \tag{2.7}$$

and thus,

$$\lim_{n \nearrow \infty} \max_{x \in \mathbb{Z}^d} \mu_{n-1}\{\omega_n = x\} = 0, \quad Q\text{-}a.s. \tag{2.8}$$

Proof of Theorem 2.1.2 is given in section 3.3.

Remark 2.1.2. Theorem 2.1.2 was obtained for Gaussian environment by P. Carmona and Y. Hu [4] and then for general environment by F. Comets, T. Shiga and N. Yoshida [5].

2.2. The strong disorder phase

The result we present in this subsection shows that the behavior of the polymer is quite different from the usual random walk if either (i) $d = 1, 2$

and $\beta \neq 0$ or (ii) $d \geq 3$ and β is large[2] For this model, it is rather recent that the phenomena of this kind began to be studied rigorously.

We now present a result which is in sharp contrast with (2.6) and (2.8).

Theorem 2.2.1. **(Localization to the favorite sites** [4, 5]**)** *Suppose either that*

(i): $d = 1, 2$ *and* $\beta \neq 0$ *or*

(ii): $d \geq 1$ *and*

$$\gamma_2(\beta) \overset{\text{def.}}{=} \beta\lambda'(\beta) - \lambda(\beta) > \ln(2d). \tag{2.9}$$

Then, there exists a constant $c = c(d, \beta) > 0$ *such that*

$$\varlimsup_{n \nearrow \infty} \max_{x \in \mathbb{Z}^d} \mu_{n-1}\{\omega_n = x\} \geq c, \quad Q\text{-}a.s. \tag{2.10}$$

The bound (2.10) suggests that the position of ω_n, viewed under the polymer measure μ_{n-1}, is concentrated at a small region (the "favorite sites") with the size $\mathcal{O}(1)$ as $n \nearrow \infty$.

Note that γ_2 is increasing on $[0, \infty)$ and therefore that (2.9) holds for large enough β if

$$\lim_{\beta \nearrow \infty} \gamma_2(\beta) > \ln(2d). \tag{2.11}$$

We see from Theorem 2.1.2 and Theorem 2.2.1 that, if $d \geq 3$ and (2.11), then a phase transition occurs as β increases from the weak disorder phase to the strong disorder phase.

Theorem 2.2.1 under condition (ii) is proved in section 3.4. For the proof of this theorem under condition (i), we refer the reader to [4, 5].

Remark 2.2.1. For $d \geq 3$, there are exceptional choices of the distribution of $\eta(n, x)$ like the one discussed in Example 2.1.1, for which (2.10) does not hold even for large β (in fact, (2.8) holds for all β); to be on the safe side for this statement, one can consider unbounded environments, or bounded ones without mass on the top point of its support. In this case, one has (2.11), and hence (2.9) for large enough β. See Example 2.2.1 and Example 2.2.2 below.

Example 2.2.1. Consider the Bernoulli environment (Example 1.2.1). Then, it is not difficult to see from direct computations that $\lim_{\beta \nearrow \infty} \gamma_2(\beta) = \ln(1/(1-p))$. This shows that (2.9) holds for large enough β if $p > 1 - \frac{1}{2d}$.

---

[2] This is again, up to some exceptions See Remark 2.2.1 below.

Example 2.2.2. Consider the Gaussian environment (Example 1.2.1). Then, $\gamma_2(\beta) = \beta^2/2$ and hence (2.9) holds if $\beta > \sqrt{2\ln(2d)}$.

Remark 2.2.2. Theorem 2.2.1 was obtained for Gaussian environment by P. Carmona and Y. Hu [4] and then for general environment by F. Comets, T. Shiga and N. Yoshida [5].

2.3. The normalized partition function and its positivity in the limit

We now introduce an important process on $(H, \mathcal{G}, Q)$ ((2.12) below), which is a martingale in fact. The large time behavior of this process characterizes the phase diagram of this model and for this reason, many of results on the model can be best understood from the viewpoint of this process.

Define the *normalized partition function* by

$$W_n = \exp(-\lambda(\beta)n)Z_n, \quad n \geq 1. \tag{2.12}$$

We then have

Lemma 2.3.1. *The limit*

$$W_\infty = \lim_{n \nearrow \infty} W_n \tag{2.13}$$

exists Q-a.s. Moreover, there are only two possibilities for the positivity of the limit;

$$Q\{W_\infty > 0\} = 1 \ , \tag{2.14}$$

or

$$Q\{W_\infty = 0\} = 1 \ . \tag{2.15}$$

The proof of this lemma is standard and is given in section 3.1.

The above contrasting situations (2.14) and (2.15) can be considered as the characterization of the weak disorder phase and the strong disorder phase, respectively. In fact, as are shown in Theorem 3.3.1 below, (2.14) implies (2.7), while (2.15) implies a weaker form of (2.10) that $\sum_{n\geq 1} \max_{x\in\mathbb{Z}^d} \mu_{n-1}\{\omega_n = x\} = \infty$, Q-a.s. It is even expected that (2.14) implies (2.5) and that (2.15) implies (2.10).

We close this subsection with the following result, which, in consistency with what we discuss above, describes the basic phase diagram of the model.

Theorem 2.3.2. **(a):** *For $d \geq 3$, (2.3) implies (2.14).*

(b): *Either (i) or (ii) in Theorem 2.2.1 implies (2.15).*

Proofs of Theorem 2.3.2 (a) and (b) are given in Sections 3.2 and 3.4, respectively.

Remark 2.3.1. For $d \geq 3$, it is an interesting question to find a characterization of (2.14) (or (2.15)) in terms of the distribution of $\eta(n,x)$. As is shown in section 3.1, $(W_n)_{n\geq 1}$ is a mean-one, positive martingale on $(H, \mathcal{G}, Q)$. In this respect, this question has somewhat similar flavor to some other topics in the probability theory such as Kakutani's dichotomy for infinite product measure (e.g.,[8, page 244]), nontriviality of the limit of the normalized Galton-Watson process [1] and of multiplicative chaos [14].

3. Martingale analysis on the simple random walk model

This section is devoted to the proofs of the results introduced in the previous one. We define an increasing sequence of sub σ-fields of $\mathcal{G}$ by

$$\mathcal{G}_n = \sigma[\eta(j,x) \ ; \ j \leq n, \ x \in \mathbb{Z}^d], \quad n \geq 1. \tag{3.1}$$

A major technical advantage of the model is that we can relate objects of interest such as

$$\mu_n[|\omega_n|^2] \text{ and } \max_{x\in\mathbb{Z}^d} \mu_{n-1}\{\omega_n = x\}$$

to some martingale on $(H, \mathcal{G}, Q)$ with respect to the filtration $(\mathcal{G}_n)$. As is very easy to guess, what makes this possible is the independence of the environment $\{\eta(n,x)\}$, especially in the time parameter n. We will see from the arguments below, the martingale analysis plays a key role in everything we do.

3.1. Proof of Lemma 2.3.1

We first show that $(W_n)_{n\geq 1}$ is a mean-one, positive $(\mathcal{G}_n)$-martingale on $(H, \mathcal{G}, Q)$. Here and in what follows, we use the following notation.

$$\begin{aligned} e(n,x) &= e(n,x,\eta) = \exp\left(\beta\eta(n,x) - \lambda(\beta)\right), && (3.2)\\ e_{1,n} &= e_{1,n}(\omega,\eta) = \prod_{1\leq j\leq n} e(j,\omega_j). && (3.3)\end{aligned}$$

Note that $W_n = P[e_{1,n}]$ in this notation. For any fixed $\omega \in \Omega$, $e_{1,n}$ is the product of mean-one i.i.d. random variables on $(H, \mathcal{G}, Q)$ and hence is a mean-one, positive $(\mathcal{G}_n)$-martingale. This implies the martingale property of W_n. By the martingale convergence theorem, the limit W_∞ exists Q-a.s. It is clear that the event $\{W_\infty = 0\}$ is measurable with respect to the tail σ-field

$$\bigcap_{n\geq 1} \sigma[\eta(j,x) \ ; \ j \geq n, \ x \in \mathbb{Z}^d].$$

Therefore by Kolmogorov's zero-one law, only (2.14) and (2.15) are the possibilities. □

3.2. The second moment method

In this subsection, we give proofs to Theorem 2.1.1 and Theorem 2.3.2 (a). The proofs are based on the L^2 analysis of certain martingales on $(H, \mathcal{G}, Q)$. This approach was introduced by E. Bolthausen [3] and then investigated further by R. Song and X. Y. Zhou [25]. We summarize the main step in their analysis as Proposition 3.2.1 below. The proposition deals with a process $(M_n)_{n\geq 1}$ on $(H, \mathcal{G}, Q)$ of the form;

$$M_n = P[\varphi(n, \omega_n) e_{1,n}]. \tag{3.4}$$

Here, $e_{1,n}$ has been introduced by (3.3) and $\varphi : \mathbb{N} \times \mathbb{Z}^d \to \mathbb{R}$ is a function for which we assume the following properties:

(P1): There are constants $C_i, p \in [0, \infty)$, $i = 0, 1, 2$ such that

$$|\varphi(n,x)| \leq C_0 + C_1|x|^p + C_2 n^{p/2} \quad \text{for all } (n,x) \in \mathbb{N} \times \mathbb{Z}^d. \tag{3.5}$$

(P2): $\Phi_n \overset{\text{def.}}{=} \varphi(n, \omega_n)$, $n \geq 1$ is a martingale on $(\Omega, \mathcal{F}, P)$ with respect to the filtration

$$\mathcal{F}_n = \sigma[\omega_j \ ; \ j \leq n]. \tag{3.6}$$

It is easy to see from (P2) that $(M_n)_{n\geq 1}$ is a $(\mathcal{G}_n)$-martingale on $(H, \mathcal{G}, Q)$. The following proposition generalizes [3, Lemma 4] and [25, Theorem 2].

Proposition 3.2.1. *Consider the martingale $(M_n)_{n\geq 1}$ defined by (3.4). Suppose that $d \geq 3$ and that (2.3), (P1), (P2) are satisfied. Then, there exists $\kappa \in [0, p/2)$ such that*

$$\max_{0\leq j\leq n} |M_j| = \mathcal{O}(n^\kappa), \quad \textit{as } n \nearrow \infty, \ Q\textit{-a.s.} \tag{3.7}$$

If in addition, $p < \frac{1}{2}d - 1$, then

$$\lim_{n\nearrow\infty} M_n \ \textit{exists } Q\textit{-a.s. and in } L^2(Q). \tag{3.8}$$

Remark 3.2.1. As will be seen from the way (3.7) is used below, it is crucial that the divergence of the right-hand-side is strictly slower than $n^{p/2}$, and this is where the property (P2) is relevant. If we drop the property (P2) from the assumption of Proposition 3.2.1, we then have a larger bound:

$$M_n = \mathcal{O}(n^{p/2}), \quad \text{as } n \nearrow \infty, \ Q\text{-a.s.} \tag{3.9}$$

This larger bound from the weaker assumption can be obtained via Proposition 3.2.1 as in the proof of (2.1) in [3].

We will prove Proposition 3.2.1 later on. Before doing so, we explain how this proposition is used to derive the desired conclusions in Theorem 2.1.1 and in Theorem 2.3.2 (a).

- Theorem 2.3.2 (a) is proved by choosing $\varphi \equiv 1$ in Proposition 3.2.1. By (3.8), $M_n = W_n$ converges in $L^2(Q)$. In particular,

$$Q[W_\infty] = \lim_{n \nearrow \infty} Q[W_n] = 1.$$

This implies $Q\{W_\infty > 0\} > 0$ and hence that $Q\{W_\infty > 0\} = 1$ by the zero-one law.

- To prove (2.4), we take $\varphi(n, x) = |x|^2 - n$ (hence $p = 2$). Then, by Theorem 2.3.2 (a) and Proposition 3.2.1, there exists $\kappa \in [0, 1)$ such that

$$\mu_n[|\omega_n|^2] - n = P[\varphi(n, \omega_n)e_{1,n}]/W_n = \mathcal{O}(n^\kappa) \quad Q\text{-a.s.}$$

We now turn to the proof of Proposition 3.2.1. Here, we follow [25]. We present a key step in the proof as a lemma.

Lemma 3.2.2. *Suppose that $d \geq 3$ and that (2.3), (P1), (P2) are satisfied. Then,*

$$Q[M_n^2] = \mathcal{O}(b_n), \quad \textit{as } n \nearrow \infty, \ Q\textit{-a.s.} \tag{3.10}$$

where $b_n = 1$ if $p < \frac{d}{2} - 1$, $b_n = \ln n$ if $p = \frac{d}{2} - 1$, and $b_n = n^{p - \frac{d}{2} + 1}$ if $p > \frac{d}{2} - 1$.

Remark 3.2.2. The choice of b_n is made in order to have $\sum_{1 \leq j \leq n} j^{p - \frac{d}{2}} = \mathcal{O}(b_n)$. See (3.15) below for the reason of the power $p - \frac{d}{2}$.

Proof of Lemma 3.2.2: On the product space $(\Omega^2, \mathcal{F}^{\otimes 2})$, we consider the probability measure $P^{\otimes 2} = P^{\otimes 2}(d\omega, d\widetilde{\omega})$, that we will view as the distribution of the couple $(\omega, \widetilde{\omega})$ with $\widetilde{\omega} = (\widetilde{\omega}_k)_{k\geq 0}$ an independent copy of $\omega = (\omega_k)_{k\geq 0}$. We write $\chi_{i_1,...,i_k}$ for the indicator function of the event

$$\{\omega_{i_1} = \widetilde{\omega}_{i_1},\ \omega_{i_2} = \widetilde{\omega}_{i_2}\ , ...,\ \omega_{i_k} = \widetilde{\omega}_{i_k}\}.$$

We first expand the second moment $Q[M_n^2]$ as follows:

(3.11)

$$Q[M_n^2] = \Phi_0^2 + \sum_{1\leq k\leq n} (e^{\gamma_1(\beta)} - 1)^k \sum_{1\leq i_1<...<i_k\leq n} P^{\otimes 2}[\Phi_{i_k}(\omega)^2\chi_{i_1,...,i_k}].$$

To see this, we write M_n^2 in terms of the independent copy:

$$\begin{aligned} M_n^2 &= P[\Phi_n e_{1,n}]^2 \\ &= P^{\otimes 2}[\Phi_n(\omega)\Phi_n(\widetilde{\omega})e_{1,n}(\omega,\eta)e_{1,n}(\widetilde{\omega},\eta)]. \end{aligned} \tag{3.12}$$

It follows from (3.12) that

$$Q[M_n^2] = P^{\otimes 2}\left[\Phi_n(\omega)\Phi_n(\widetilde{\omega})Q[e_{1,n}(\omega,\eta)e_{1,n}(\widetilde{\omega},\eta)]\right]. \tag{3.13}$$

On the other hand, with notation (3.2), we have that

$$Q[e(\omega_j,\eta)e(\widetilde{\omega}_j,\eta)] = 1 + (e^{\gamma_1(\beta)} - 1)\chi_j,$$

ane hence that

$$\begin{aligned} Q[e_{1,n}(\omega,\eta)e_{1,n}(\widetilde{\omega},\eta)] &= \prod_{1\leq j\leq n}\Big(1 + (e^{\gamma_1(\beta)} - 1)\chi_j\Big) \\ &= 1 + \sum_{1\leq k\leq n} (e^{\gamma_1(\beta)} - 1)^k \sum_{1\leq i_1<...<i_k\leq n} \chi_{i_1,...i_k}. \end{aligned} \tag{3.14}$$

The expansion (3.11) is now obtained by inserting (3.14) into (3.13) and by the martingale property of Φ_n.

Let us fix $i_1, ..., i_k$ for a moment. We then have by (3.5) that

$$P^{\otimes 2}[\Phi_{i_k}(\omega)^2\chi_{i_1,...,i_k}] \leq 3C_1^2 A_{i_1,...,i_k} + 3(C_0^2 + C_2^2)B_{i_1,...,i_k},$$

where

$$A_{i_1,...,i_k} = P^{\otimes 2}[|\omega_{i_k}|^{2p}\chi_{i_1,...,i_k}],\quad B_{i_1,...,i_k} = i_k^p P^{\otimes 2}[\chi_{i_1,...,i_k}].$$

We now bound $A_{i_1,...,i_k}$ from above. As will be seen from the way it is done, the same bound (up to the multiplicative constant) is obtained for

$B_{i_1,\dots,i_k}$. Observe that

$$\begin{aligned} P^{\otimes 2}[|\omega_n|^{2p}\chi_n] &= \sum_{x\in\mathbb{Z}^d} P[|\omega_n|^{2p} : \omega_n = x]P[\omega_n = x] \\ &\leq Cn^{-\frac{d}{2}}P[|\omega_n|^{2p}] \\ &\leq Cn^{p-\frac{d}{2}}, \qquad (3.15) \end{aligned}$$

where we have used (2.6) on the second line. We write $j_\ell = i_\ell - i_{\ell-1}$, $\ell = 1,2,..,k$ with $i_0 = 0$. We then see from the Markov property and (3.15) that

$$\begin{aligned} &A_{i_1,\dots,i_k} \\ &\leq k^{2p-1}\sum_{1\leq \ell\leq k} P^{\otimes 2}[|\omega_{i_\ell} - \omega_{i_{\ell-1}}|^{2p}\chi_{i_1,\dots,i_k}] \\ &= k^{2p-1}\sum_{1\leq \ell\leq k}\left(\prod_{1\leq m<\ell} P^{\otimes 2}[\chi_{j_m}]\right)P^{\otimes 2}[|\omega_{j_\ell}|^{2p}\chi_{j_\ell}]\left(\prod_{\ell<m\leq k} P^{\otimes 2}[\chi_{j_m}]\right) \\ &\leq Ck^{2p-1}\sum_{1\leq \ell\leq k} j_\ell^{p-\frac{d}{2}}\prod_{\substack{1\leq m\leq k \\ m\neq \ell}} P^{\otimes 2}[\chi_{j_m}] \end{aligned}$$

Note that $\sum_{1\leq j\leq n} j^{p-\frac{d}{2}} = \mathcal{O}(b_n)$ and that $\sum_{j\geq 1} P^{\otimes 2}[\chi_j] = \frac{\pi_d}{1-\pi_d}$. Therefore, we obtain from what we have seen above that

$$\begin{aligned} &\sum_{1\leq i_1<\dots<i_k\leq n} P^{\otimes 2}[\Phi_{i_k}(\omega)^2\chi_{i_1,\dots,i_k}] \\ &\leq Ck^{2p-1}\sum_{1\leq \ell\leq k}\sum_{1\leq j_1\leq n}\cdots\sum_{1\leq j_k\leq n} j_\ell^{p-\frac{d}{2}}\prod_{\substack{1\leq m\leq k \\ m\neq \ell}} P^{\otimes 2}[\chi_{j_m}] \\ &\leq \mathcal{O}(b_n)k^{2p}\left(\frac{\pi_d}{1-\pi_d}\right)^{k-1}. \end{aligned}$$

By this and (3.11), we now arrive at

$$Q[M_n^2] \leq \Phi_0^2 + \mathcal{O}(b_n)\sum_{k\geq 1} k^{2p}(e^{\gamma_1(\beta)}-1)^k\left(\frac{\pi_d}{1-\pi_d}\right)^{k-1}.$$

The summation in k converges, thanks to the assumptions $d \geq 3$ and (2.3). This finishes the proof of (3.10). □

Remark 3.2.3. We see from the proof of (3.10) that

$$\sup_{n\geq 1} Q[W_n^2] = 1 + \sum_{k\geq 1}(e^{\gamma_1(\beta)} - 1)^k \left(\frac{\pi_d}{1-\pi_d}\right)^{k-1}.$$

This shows that $\sup_{n\geq 1} Q[W_n^2] < \infty$ if and only if $d \geq 3$ and (2.3) holds.

It is now, easy to complete the proof of Proposition 3.2.1. We set $M_n^* = \max_{0\leq j\leq n} |M_j|$ to simplify the notation. For (3.10), it is sufficient to prove that for any $\delta > 0$,

$$M_n^* = \mathcal{O}(n^\delta \sqrt{b_n}) \quad \text{as } n \nearrow \infty,\ Q\text{-a.s.,} \tag{3.16}$$

where b_n is the L^2-bound in Lemma 3.2.2. Moreover, by the monotonicity of M_n^* and the polynomial growth of $n^\delta\sqrt{b_n}$, it is enough to prove (3.16) along a subsequence $\{n^k : n \geq 1\}$ for some power $k \geq 2$. Now, take $k > 1/\delta$. We then have by Chebychev's inequality, Doob's inequality and Lemma 3.2.2 that

$$\begin{aligned}
Q\{M_{n^k}^* > n^{k\delta}\sqrt{b_{n^k}}\} &\leq Q\{M_{n^k}^* > n\sqrt{b_{n^k}}\} \\
&\leq Q[(M_{n^k}^*)^2]/(n^2 b_{n^k}) \\
&\leq 4Q[M_{n^k}^2]/(n^2 b_{n^k}) \\
&\leq Cn^{-2}.
\end{aligned}$$

Then, it follows from the Borel-Cantelli lemma that

$$Q\{M_{n^k}^* \leq n^{k\delta}\sqrt{b_{n^k}} \text{ for large enough } n\text{'s}\} = 1.$$

This ends the proof of (3.7).

The second statement (3.8) in Proposition 3.2.1 follows from Lemma 3.2.2 and the martingale convergence theorem. This completes the proof of Proposition 3.2.1. □

Remark 3.2.4. With Proposition 3.2.1 in hand, we are no longer far away from the central limit theorem (2.5). Following [3], we now explain a route to (2.5).

We let $a = (a_j)_{j=1}^d$ and $b = (b_j)_{j=1}^d$ denote multi indices in what follows. We will use standard notation $|a|_1 = a_1 + ... + a_d$, $x^a = x_1^{a_1} \cdots x_d^{a_d}$ and $(\frac{\partial}{\partial x})^a = \left(\frac{\partial}{\partial x_1}\right)^{a_1} \cdots \left(\frac{\partial}{\partial x_d}\right)^{a_d}$ for $x \in \mathbb{R}^d$. It is enough to prove (2.5) for any monomial of the form $f(x) = x^a$. We will do this by induction on $|a|_1$. We introduce

$$\begin{aligned}\varphi(n,x) &= \left(\frac{\partial}{\partial\theta}\right)^a \exp(\theta\cdot x - n\rho(\theta))\Big|_{\theta=0},\\ \psi(n,x) &= \left(\frac{\partial}{\partial\theta}\right)^a \exp\left(\theta\cdot x - n\frac{|\theta|^2}{2d}\right)\Big|_{\theta=0},\end{aligned}$$

where $\rho(\theta) = \ln\left(\frac{1}{d}\sum_{1\le j\le d}\cosh(\theta_j)\right)$. Clearly, the function φ satisfies (P1) and (P2) with $p = |a|_1$. On the other hand, we see from the definition of ψ that

$$(2\pi)^{-d/2}\int_{\mathbb{R}^d}\psi(1,x/\sqrt{d})e^{-|x|^2/2}dx = 0. \tag{3.17}$$

Moreover, it is not difficult to see [3, Lemma 3c] that $\varphi(n,x) = x^a + \varphi_0(n,x)$ and $\psi(n,x) = x^a + \psi_0(n,x)$ where

$$\varphi_0(n,x) = \sum_{\substack{|b|_1+2j\le|a|_1\\ j\ge1}} A_a(b,j)x^b n^j,\quad \psi_0(n,x) = \sum_{\substack{|b|_1+2j=|a|_1\\ j\ge1}} A_a(b,j)x^b n^j.$$

for some $A_a(b,j)\in\mathbb{R}$. In particular, φ_0 and ψ_0 have the same coefficients for $x^b n^j$ with $|b|_1 + 2j = |a|_1$. We now write $\mu_n[(\omega_n/\sqrt{n})^a]$ as

$$\begin{aligned}\mu_n[(\omega_n/\sqrt{n})^a] &= \frac{1}{W_n}P[\varphi(n,\omega_n)e_{1,n}]n^{-|a|_1/2}\\ &\quad -\frac{1}{W_n}P[\psi_0(1,\omega_n/\sqrt{n})e_{1,n}]\\ &\quad +\frac{1}{W_n}P[(\psi_0(n,\omega_n)-\varphi_0(n,\omega_n))\,e_{1,n}]n^{-|a|_1/2}\end{aligned}$$

As $n \nearrow \infty$, the second term converges to $(2\pi)^{-d/2}\int_{\mathbb{R}^d}(x/\sqrt{d})^a \times e^{-|x|^2/2}dx$ by the induction hypothesis and (3.17). The first and the third terms on the right-hand side vanish as $n \nearrow \infty$. In fact, we use Theorem 2.3.2 (a), Proposition 3.2.1 for the first term and Theorem 2.3.2 (a), (3.9) for the third term.

3.3. The replica overlap

In this subsection, we prove Theorem 2.1.2 as a consequence of Theorem 2.3.2 (a) and Theorem 3.3.1 below.

For $n \ge 1$, we introduce the following random variables on $(H, \mathcal{G}, Q)$;

$$I_n = \sum_{x\in\mathbb{Z}^d}\mu_{n-1}\{\omega_n = x\}^2,\quad J_n = \max_{x\in\mathbb{Z}^d}\mu_{n-1}\{\omega_n = x\}.$$

It is clear that

$$J_n^2 \leq I_n \leq J_n. \tag{3.18}$$

Both Theorem 2.1.2 and Theorem 2.2.1 deal with the large time behavior of J_n, $n \nearrow \infty$. As we will see below, I_n is better suited for the martingale analysis. For this reason, we will prove these theorems by studying I_n, rather than J_n itself.

We now mention to an interpretation of I_n. On the product space $(\Omega^2, \mathcal{F}^{\otimes 2})$, we consider the probability measure $\mu_n^{\otimes 2} = \mu_n^{\otimes 2}(d\omega, d\widetilde{\omega})$, that we will view as the distribution of the couple $(\omega, \widetilde{\omega})$ with $\widetilde{\omega} = (\widetilde{\omega}_k)_{k \geq 0}$ an independent copy of $\omega = (\omega_k)_{k \geq 0}$ with law μ_n. We then have that

$$I_n = \mu_{n-1}^{\otimes 2}(\omega_n = \widetilde{\omega}_n) \, . \tag{3.19}$$

Hence, the summation

$$\sum_{1 \leq k \leq n} I_k \tag{3.20}$$

is the expected amount of the overlap up to time n of two independent polymers in the same (fixed) environment. This can be viewed as an analogue to the so-called *replica overlap* often discussed in the context of disordered systems, e.g. mean field spin glass, and also of directed polymers on trees [7].

The large time behavior of (3.20) and the normalized partition function W_n are related as follows.

Theorem 3.3.1. *Let $\beta \neq 0$. Then,*

$$\{W_\infty = 0\} = \left\{ \sum_{n \geq 1} I_n = \infty \right\}, \quad Q\text{-a.s.} \tag{3.21}$$

Moreover, if $Q\{W_\infty = 0\} = 1$, there exist $c_1, c_2 \in (0, \infty)$ such that Q-a.s.,

$$c_1 \sum_{1 \leq k \leq n} I_k \leq -\ln W_n \leq c_2 \sum_{1 \leq k \leq n} I_k \qquad \textit{for large enough } n\textit{'s}. \tag{3.22}$$

We first note that Theorem 2.1.2 is now obtained as a consequence of Theorem 2.3.2 (a), (3.21) and (3.18).

Proof of Theorem 3.3.1: To conclude (3.21) and (3.22), it is enough to show the following (3.23) and (3.24):

$$\{W_\infty = 0\} \subset \left\{ \sum_{n \geq 1} I_n = \infty \right\}, \quad Q\text{-a.s.} \tag{3.23}$$

There are $c_1, c_2 \in (0,\infty)$ such that

$$\left\{ \sum_{n\geq 1} I_n = \infty \right\} \subset \{(3.22) \text{ holds}\}, \quad Q\text{-a.s.} \tag{3.24}$$

The proof of (3.23) and (3.24) are based on Doob's decomposition for the process $-\ln W_n$. It is convenient to introduce some more notation. For a sequence $(a_n)_{n\geq 0}$ (random or non-random), we set $\Delta a_n = a_n - a_{n-1}$ for $n \geq 1$. Let us now recall Doob's decomposition in this context; any $(\mathcal{G}_n)$-adapted process $X = \{X_n\}_{n\geq 0} \subset L^1(Q)$ can be decomposed in a unique way as

$$X_n = M_n(X) + A_n(X), \quad n \geq 1,$$

where $M(X)$ is an $(\mathcal{G}_n)$-martingale and

$$A_0 = 0, \quad \Delta A_n = Q[\Delta X_n | \mathcal{G}_{n-1}], \quad n \geq 1.$$

$M_n(X)$ and $A_n(X)$ are called respectively, the martingale part and compensator of the process X. If X is a square integrable martingale, then the compensator $A_n(X^2)$ of the process $X^2 = \{(X_n)^2\}_{n\geq 0} \subset L^1(Q)$ is denoted by $\langle X \rangle_n$ and is given by the following formula:

$$\Delta\langle X \rangle_n = Q[(\Delta X_n)^2 | \mathcal{G}_{n-1}]$$

Here, we are interested in the Doob's decomposition of $X_n = -\ln W_n$, whose martingale part and the compensator will be henceforth denoted M_n and A_n respectively

$$-\ln W_n = M_n + A_n. \tag{3.25}$$

To compute M_n and A_n, we introduce $U_n = \mu_{n-1}[e(n,\omega_n)] - 1$ (Recall (3.2)). It is then clear that

$$W_n / W_{n-1} = 1 + U_n \tag{3.26}$$

and hence that

$$\begin{aligned} \Delta A_n &= -Q[\ln(1+U_n)|\mathcal{G}_{n-1}], \\ \Delta M_n &= -\ln(1+U_n) + Q[\ln(1+U_n)|\mathcal{G}_{n-1}]. \end{aligned} \tag{3.27}$$

In particular,

$$\Delta\langle M \rangle_n \leq Q[\ln^2(1+U_n)|\mathcal{G}_{n-1}]. \tag{3.28}$$

We now claim that there is a constant $c \in (0, \infty)$ such that

$$\frac{1}{c} I_n \leq \Delta A_n \leq c I_n, \quad \Delta\langle M \rangle_n \leq c I_n. \tag{3.29}$$

Indeed, both follow from (3.27), (3.28) and Lemma 3.3.2 below; $\{e_i\}$, $\{\alpha_i\}$ and Q in the lemma play the roles of $\{e(n,z)\}_{|z|_1 \leq n}$, $\{\mu_{n-1}(\omega_n = z)\}_{|z|_1 \leq n}$ and $Q[\,\cdot\, |\mathcal{G}_{n-1}]$.

We now conclude (3.23) from (3.29) as follows (the equalities and the inclusions here being understood as Q-a.s.):

$$\begin{aligned} \left\{ \sum_{n \geq 1} I_n < \infty \right\} &\subset \{A_\infty < \infty, \ \langle M \rangle_\infty < \infty\} \\ &\subset \{A_\infty < \infty, \ \lim_{n \nearrow \infty} M_n \text{ exists and is finite}\} \\ &\subset \{W_\infty > 0\}. \end{aligned}$$

Here, on the second line, we have used a well-known property for martingales, e.g. [8, page 255, (4.9)].

Finally we prove (3.24). By (3.29), it is enough to show that

$$\{A_\infty = \infty\} \subset \left\{ \lim_{n \nearrow \infty} -\frac{\ln W_n}{A_n} = 1 \right\}, \quad Q\text{-a.s.} \tag{3.30}$$

Thus, let us suppose that $A_\infty = \infty$. If $\langle M \rangle_\infty < \infty$, then again by [8, page 255, (4.9)], $\lim_{n \nearrow \infty} M_n$ exists and is finite and therefore (3.30) holds. If, on the contrary, $\langle M \rangle_\infty = \infty$, then

$$-\frac{\ln W_n}{A_n} = \frac{M_n}{\langle M \rangle_n} \frac{\langle M \rangle_n}{A_n} + 1 \to 1 \quad Q\text{-a.s.}$$

by (3.29) and the law of large numbers for martingales, see [8, page 255, (4.10)]. This completes the proof of Theorem 3.3.1. □

Lemma 3.3.2. *Let e_i, $1 \leq i \leq m$ be positive, non-constant i.i.d. random variables on a probability space $(H, \mathcal{G}, Q)$ such that*

$$Q[e_1] = 1, \quad Q[e_1^3 + \ln^2 e_1] < \infty.$$

For $\{\alpha_i\}_{1 \leq i \leq m} \subset [0, \infty)$ such that $\sum_{1 \leq i \leq m} \alpha_i = 1$, define a centered random variable $U > -1$ by $U = \sum_{1 \leq i \leq m} \alpha_i e_i - 1$. Then, there exists a *constant* $c \in (0, \infty)$, *independent of* $\{\alpha_i\}_{1 \le i \le m}$ *such that*

$$\frac{1}{c} \sum_{1 \le i \le m} \alpha_i^2 \le Q\left[\frac{U^2}{2+U}\right], \tag{3.31}$$

$$\frac{1}{c} \sum_{1 \le i \le m} \alpha_i^2 \le -Q\left[\ln(1+U)\right] \le c \sum_{1 \le i \le m} \alpha_i^2, \tag{3.32}$$

$$Q\left[\ln^2(1+U)\right] \le c \sum_{1 \le i \le m} \alpha_i^2. \tag{3.33}$$

The readers are invited to try the proof of this lemma as an interesting exercise. A solution can be found in [5].

3.4. The fractional moment method

In this subsection, we prove Theorem 2.2.1(b) and Theorem 2.3.2(b). Both are obtained by dealing with the fractional moment $Q[W_n^\theta]$, $0 < \theta < 1$. To be more precise, we will prove that for some $\theta \in (0,1)$ and $a_n \nearrow \infty$,

$$\overline{\lim_{n \nearrow \infty}} \frac{1}{a_n} \ln Q[W_n^\theta] < 0. \tag{3.34}$$

Proof of Theorem 2.2.1 under condition (ii): We first assume (3.34) with $a_n = n$ for a moment to see that it implies (2.10). We then have by the Borel-Cantelli lemma that there is $c_3 \in (0, \infty)$ such that

$$\overline{\lim_{n \nearrow \infty}} \frac{1}{n} \ln W_n < -c_3, \quad Q\text{-a.s.} \tag{3.35}$$

Then, by (3.18) and (3.22) we conclude that

$$\begin{aligned} \overline{\lim_{n \nearrow \infty}} J_n &\ge \overline{\lim_{n \nearrow \infty}} \frac{1}{n} \sum_{1 \le k \le n} I_k \\ &\ge -\underline{\lim_{n \nearrow \infty}} \frac{1}{c_2 n} \ln W_n \\ &\ge c_3/c_2. \end{aligned}$$

We now turn to the proof of (3.34) with $a_n = n$. Recall the notation (3.2) and define

$$W_{n,m}^x = P\left[\prod_{1 \le j \le m} e(j+n, x+\omega_j)\right], \quad n, m \ge 1.$$

For $\theta \in (0,1)$, by the subadditive estimate $(u+v)^\theta \le u^\theta + v^\theta$, $u, v > 0$, we get

$$W_n^\theta \le (2d)^{-\theta} \sum_{x,|x|_1=1} e(1,x)^\theta (W_{1,n-1}^x)^\theta.$$

Since $W_{1,n-1}^x$ has the same law as W_{n-1},

$$Q[W_n^\theta] \le r(\theta) Q[W_{n-1}^\theta],$$

where $r(\theta) = (2d)^{1-\theta} Q[e(1,x)^\theta]$. Note that $\theta \mapsto \ln r(\theta)$ is convex, continuously differentiable, and that $\ln(2d) = \ln r(0) > \ln r(1) = 0$. Therefore $r(\theta) < 1$ for some $\theta \in (0,1)$ if and only if $0 < \left.\frac{d \ln r(\theta)}{d\theta}\right|_{\theta=1}$, which is equivalent to $\gamma_2(\beta) > \ln(2d)$. □

Proof of Theorem 2.3.2(b): We will check (3.34) where $a_n = n^{1/3}$ if $d = 1$ and $a_n = \sqrt{\ln n}$ if $d = 2$. In this respect, we first prove an auxiliary lemma.

Lemma 3.4.1. *For* $\theta \in [0,1]$ *and* $\Lambda \subset \mathbb{Z}^d$,

$$|\Lambda| Q\left[W_{n-1}^\theta I_n\right] \ge Q\left[W_{n-1}^\theta\right] - 2P(\omega_n \notin \Lambda)^\theta. \tag{3.36}$$

Proof: Repeating the argument in [19, page 453], we see that

$$\begin{aligned}
|\Lambda| I_n &\ge |\Lambda| \sum_{z \in \Lambda} \mu_{n-1}(\omega_n = z)^2 \\
&\ge \mu_{n-1}(\omega_n \in \Lambda)^2 \\
&= (1 - \mu_{n-1}(\omega_n \notin \Lambda))^2 \\
&\ge 1 - 2\mu_{n-1}(\omega_n \notin \Lambda) \\
&\ge 1 - 2\mu_{n-1}(\omega_n \notin \Lambda)^\theta.
\end{aligned}$$

Note also that

$$\begin{aligned}
Q\left[W_{n-1}^\theta \mu_{n-1}(\omega_n \notin \Lambda)^\theta\right] &\le Q\left[W_{n-1}\mu_{n-1}(\omega_n \notin \Lambda)\right]^\theta \\
&= P(\omega_n \notin \Lambda)^\theta.
\end{aligned}$$

We therefore see that

$$\begin{aligned}
|\Lambda| Q\left[W_{n-1}^\theta I_n\right] &\ge Q\left[W_{n-1}^\theta\right] - 2Q\left[W_{n-1}^\theta \mu_{n-1}(\omega_n \notin \Lambda)^\theta\right] \\
&\ge Q\left[W_{n-1}^\theta\right] - 2P(\omega_n \notin \Lambda)^\theta.
\end{aligned}$$

□

Assume now that $\theta \in (0,1)$, and define a function $f : (-1,\infty) \to [0,\infty)$ by

$$f(u) = 1 + \theta u - (1+u)^\theta.$$

It is then clear that there are constants $c_1, c_2 \in (0,\infty)$ such that

$$\frac{c_1 u^2}{2+u} \le f(u) \le c_2 u^2 \quad \text{for all } u \in (-1,\infty). \tag{3.37}$$

We see from (3.26), (3.37) and (3.31) that

$$\begin{aligned} Q[\Delta W_n^\theta | \mathcal{G}_{n-1}] &= W_{n-1}^\theta Q[(1+U_n)^\theta - 1 | \mathcal{G}_{n-1}] \\ &= -W_{n-1}^\theta Q[f(U_n)|\mathcal{G}_{n-1}] \\ &\le -c_3 W_{n-1}^\theta I_n. \end{aligned}$$

We therefore have by (3.36) that

$$QW_n^\theta \le \left(1 - \frac{c_3}{|\Lambda|}\right) Q\left[W_{n-1}^\theta\right] + \frac{2c_3}{|\Lambda|} P(\omega_n \notin \Lambda)^\theta. \tag{3.38}$$

For $d = 1$, set $\Lambda = (-n^{2/3}, n^{2/3}]$. Then,

$$P(\omega_n \notin \Lambda) = P\left(\left|\frac{\omega_n}{n^{1/2}}\right| \ge n^{1/6}\right) \le 2\exp(-\frac{n^{1/3}}{2}),$$

so that (3.38) reads,

$$QW_n^\theta \le \left(1 - \frac{c_3}{2n^{2/3}}\right) Q\left[W_{n-1}^\theta\right] + 4c_3 \exp(-\frac{\theta n^{1/3}}{2}).$$

It is not difficult to conclude (3.34) with $a_n = n^{1/3}$ from the above.

For $d = 2$, we set

$$\Lambda = (-n^{1/2} \ln^{1/4} n, n^{1/2} \ln^{1/4} n]^2$$

to get (3.34) with $a_n = \sqrt{\ln n}$ in a similar way as above. □

4. Some related models

The simple random walk model which we have discussed so far has a number of close relatives in the literature. We now mention some of them.

4.1. Gaussian random walk model

This model considered in M. Petermann [23] and by O. Mejane [20]. The polymer measure for this model is defined by the same expression (1.4). Here, however, the random walk $(\omega_n)_{n\geq 1}$ is the summation of independent Gaussian random variables in $\mathbb{R}^d$, i.e., Ω is replaced by $\Omega = \{\omega = (\omega_n)_{n\geq 0}; \omega_n \in \mathbb{R}^d, n \geq 0\}$ and P is the unique measure on $(\Omega, \mathcal{F})$ such that $\omega_1 - \omega_0, \dots, \omega_n - \omega_{n-1}$ are independent and

$$P\{\omega_0 = 0\} = 1, \quad P\{\omega_n - \omega_{n-1} \in dx\} = (2\pi)^{-d/2} \exp(-|x|^2/2)dx.$$

Moreover, as the random environment, one takes a random field

$$\{\eta(n,x) \ ; \ (n,x) \in \mathbb{N} \times \mathbb{R}^d\}$$

with a certain mild correlation in x variables. A major technical advantage in working with the Gaussian random walk rather than the simple random walk is the applicability of a Girsanov-type path transformation, which plays a key role in analyzing this model.

4.2. Brownian directed polymer

This model is introduced in [6] as a continuous model of directed polymers in random environment, defined in terms of Brownian motion and of a Poisson random measure. We first fix notation we use for the Brownian motion and Poisson random measure. Then, we introduce the polymer measure. We write $\mathbb{R}_+ = [0, \infty)$.

- *The Brownian motion:* Let $(\{\omega_t\}_{t\geq 0}, P)$ denote a d-dimensional standard Brownian motion. To be more specific, we let the measurable space $(\Omega, \mathcal{F})$ be $C(\mathbb{R}_+ \to \mathbb{R}^d)$ with the cylindrical σ-field, and P be the Wiener measure on $(\Omega, \mathcal{F})$ such that $P\{\omega_0 = 0\} = 1$.
- *The space-time Poisson random measure:* We let η denote the Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ with the unit intensity, defined on a probability space $(\mathcal{M}, \mathcal{G}, Q)$. To make the definitions more precisely, we let $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}^d)$ denote the class of Borel sets in $\mathbb{R}_+ \times \mathbb{R}^d$. Then, η is an integer valued random measure characterized by the following property: If $A_1, ..., A_n \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}^d)$ are disjoint and bounded, then

(4.1)
$$Q\left(\bigcap_{j=1}^{n} \{\eta(A_j) = k_j\}\right) = \prod_{j=1}^{n} \exp(-|A_j|)\frac{|A_j|^{k_j}}{k_j!} \quad \text{for } k_1, ..., k_n \in \mathbb{N}.$$

Here, $|A|$ denotes the Lebesgue measure of $\mathbb{R}^{d+1}$.

- *The polymer measure:* We let V_t denote a "tube" around the graph $\{(s, \omega_s)\}_{0<s\leq t}$ of the Brownian path,

$$V_t = V_t(\omega) = \{(s, x) \; ; \; s \in (0, t], \; x \in U(\omega_s)\}, \tag{4.2}$$

where $U(x) \subset \mathbb{R}^d$ is the closed ball with the unit volume, centered at $x \in \mathbb{R}^d$. For any $t > 0$, define a probability measure μ_t on the path space $(\Omega, \mathcal{F})$ by

$$\mu_t(d\omega) = \frac{\exp(\beta\eta(V_t))}{Z_t} P(d\omega), \tag{4.3}$$

where $\beta \in \mathbb{R}$ is a parameter and

$$Z_t = P[\exp(\beta\eta(V_t))] \, . \tag{4.4}$$

The Brownian motion model defined above can be thought of as a natural transposition of the simple random walk model into continuum setting.

Analogous results of Theorem 2.1.2, Theorem 2.2.1, Theorem 2.3.2, and Theorem 3.3.1 as well as an almost sure large deviation principle for the polymer measure are obtained for this model in [6]. The model allows application of stochastic calculus, with respect to both Brownian motion and Poisson process, leading to qualitative properties of the quenched Lyapunov exponent and handy formulas for the fluctuation of the free energy.

Another strong motivation for the present model is its relation to some *stochastic partial differential equations.* To describe the connection, it is necessary to relativize the partition function, by specifying the ending point of the Brownian motion at time t. Let $P[\cdot|\omega_t = y]$ be the distribution of the Brownian bridge starting at the origin at time 0 and ending at y at time t. Define

$$Z_t(y) = P[\exp(\beta\eta(V_t)) \, | \omega_t = y](2\pi t)^{-d/2} \exp\{-|y|^2/2t\} \, . \tag{4.5}$$

Then, by definition of the Brownian bridge,

$$Z_t = \int_{\mathbb{R}^d} Z_t(y) dy \, .$$

Similar to the Feynman-Kac formula, one obtains [6] the following stochastic heat equation (SHE) with multiplicative noise in a certain weak sense,

(4.6)

$$dZ_t(y) = \tfrac{1}{2}\Delta_y Z_t(y) dt + (e^\beta - 1) Z_{t^-}(y) \eta(dt \times U(y)) \, , \quad t \geq 0, y \in \mathbb{R}^d \, ,$$

where $dZ_t(y)$ denotes the time differential and $\Delta_y = (\frac{\partial}{\partial y^1})^2 + ... + (\frac{\partial}{\partial y^d})^2$.

In the literature, this equation has been extensively considered in the case of a Gaussian driving noise, instead of the Poisson process η here. Although we are able to prove (4.6) only in the weak sense, let us now pretend that (4.6) is true for all $y \in \mathbb{R}^d$. We would then see from Itô's formula that the function $h_t(y) = \ln Z_t(y)$ solves the Kardar-Parisi-Zhang equation (KPZ):

$$dh_t(y) = \tfrac{1}{2}\left(\Delta_y h_t(y) + |\nabla_y h_t(y)|^2\right) dt + \beta\eta(dt \times U(y)) .$$

We observe that, since h has jumps in the space variable y, the nonlinearity makes the precise meaning of this equation somewhat knotty. This equation was introduced in [15] to describe the long scale behavior of growing interfaces. More precisely, the fluctuations in the KPZ equation –driven by a δ-correlated, gaussian noise–, are believed to be non standard, and universal, i.e., the same as in a large class of microscopic models. See [17] for a detailed review of kinetic roughening of growth models within the physics literature, in particular to Section 5 for the status of this equation.

4.3. Crossing Brownian motion in a soft Poissonian potential

This model is studied by M. Wüthrich [30, 31, 32], see also [28]. The model investigated there is described in terms of Brownian motion and of Poisson points. However, the Brownian motion there is "undirected", in other words, the d-dimensional Brownian motion travels through the Poisson points distributed in space $\mathbb{R}^d$, not in space-time as in the Brownian directed polymer.

4.4. First and last passage percolation

The first (resp. last) passage percolation can be thought of as an analogue of directed polymers at $\beta = -\infty$ (resp. $\beta = +\infty$). In fact, we have for example that

$$\lim_{\beta \nearrow +\infty} \frac{1}{\beta} \ln Z_n = T_n^* \overset{\text{def.}}{=} \max_{\omega \in \Omega : \omega_0 = 0} \sum_{1 \le j \le n} \eta(j, \omega_j),$$

i.e., the maximal passage time T_n^* in the context of the directed last passage percolation can be obtained as a limit of the free energy of the directed polymer. It is expected and even partly vindicated that the properties of the path with minimal/maximal passage time has similar feature to the typical paths under the polymer measure [16, 21, 18]. A few exactly solvable models of directed last passage percolation have recently been worked out in dimension $d = 1$ [2, 12, 13]. Johansson [12] treats the case of geometrically distributed η's, and Baik, Deift and Johansson analyze some continuous space Poissonian directed last passage percolation model in connection with the longest increasing sequence of the random permutation [2, 13]. For these exactly solvable models, it is proven that the maximal passage time T_n^* has the following asymptotic form in law as $n \nearrow \infty$:

$$c_0 n + c_1 n^{1/3} X, \tag{4.7}$$

where c_i, $i = 1, 2$ are positive constants and X is a random variable with the Tracy-Widom distribution. As is well known, the Tracy-Widom distribution appeared in the literatures in connection with the Gaussian Unitary Ensemble [29]. Since then, it has increasingly realized that this distribution is universal as the scaling limit of many other related models. For this reason, we are tempted to believe that for $d = 1$ and $\beta \neq 0$, the free energy $\ln Z_n$ of the directed polymer has the same large time behavior as (4.7) with c_i, $i = 1, 2$ depending on β and the choice of η [27].

4.5. Other models

Directed polymers in random environment, at positive or zero temperature, relate – even better, can sometimes be exactly mapped – to a number of interesting models of growing random surfaces (directed invasion percolation, ballistic deposition, polynuclear growth, low temperature Ising models), and non equilibrium dynamics (totally asymmetric simple exclusion, population dynamics in random environment); We refer to the survey paper [17] by Krug and Spohn for detailed account on these models and their relations.

5. Critical exponents

We write $\xi(d)$ for the "wandering exponent",i.e., the critical exponent for the transversal fluctuation of the path, and $\chi(d)$ for the the critical exponent for the longitudinal fluctuation of the free energy. Their definitions are roughly

$$\sup_{0 \le j \le n} |\omega_j| \approx n^{\xi(d)} \quad \text{and} \quad \ln Z_n - Q[\ln Z_n] \approx n^{\chi(d)} \quad \text{as } n \nearrow \infty. \tag{5.1}$$

There are various ways to define rigorously these exponents, e.g. (0.6) and (0.10-11) in [30], (2.4) and (2.6-7-8) in [24], and the equivalence between these specific definitions are often non trivial. Here, we do not go into such subtlety and take (5.1) as "definitions". The polymer is said to be *diffusive* if $\xi(d) = 1/2$ and *super-diffusive* if $\xi(d) > 1/2$.

These exponents are investigated in the context of various other models and in a large number of papers. In particular, it is conjectured in physics literature that the *scaling identity* holds in any dimension,

$$\chi(d) = 2\xi(d) - 1, \quad d \geq 1, \tag{5.2}$$

and that the polymer is super-diffusive in dimension one;

$$\chi(1) = 1/3, \quad \xi(1) = 2/3. \tag{5.3}$$

See, e.g., [10],[9, (3.4),(5.11),(5.12)], [17, (5.19),(5.28)]. For some models of directed first passage percolation, K. Johansson [12, 13] proves (5.3), cf. (4.7).

On the other hand, other rigorous results prove (or suggest) for example that

$$\begin{aligned} \chi(d) &\geq 2\xi(d) - 1 \text{ for all } d \geq 1, && (5.4)\\ \xi(d) &\leq 3/4 \text{ for all } d \geq 1, && (5.5)\\ \xi(1) &> 1/2 && (5.6) \end{aligned}$$

M. Piza [24] discusses (5.4) –(5.6) for the simple random walk model. For the Gaussian random walk model, M. Petermann [23] proves (5.6), while O. Mejane [20] shows (5.5). F. Comets and N. Yoshida [6] discuss (5.4)–(5.6) in the framework of Brownian directed polymer. Critical exponents similar to the above are also discussed for the crossing Brownian motion in a soft Poissonian potential by M. Wüthrich [30, 31, 32] and for the first passage percolation by C. Licea, M. Piza and C. Newman [21, 18].

Acknowledgements: We would like to thank H. Spohn for nice discussions and variable remarks.

References

[1] Athreya, K. and Ney, P.: Branching Processes, Springer Verlag New York (1972).

[2] Baik, J.; Deift, P.; Johansson, K.: On the distribution of the length of the longest increasing subsequence of random permutations, J. Amer. Math. Soc. **12**, (1999), 1119-1178.

[3] Bolthausen, E., 1989: A note on diffusion of directed polymers in a random environment, Commun. Math. Phys. **123**, 529–534.

[4] Carmona, P., Hu Y., 2002: On the partition function of a directed polymer in a random environment. Probab.Theory Related Fields **124** (2002), no. 3, 431–457.

[5] Comets, F., Shiga, T., Yoshida, N. Directed Polymers in Random Environment: Path Localization and Strong Disorder, Preprint 2002, To appear in Bernoulli (2003).

[6] Comets, F., Yoshida, N. Brownian Directed Polymers in Random Environment Preprint 2003,

[7] Derrida, B. and Spohn, H. : Polymers on disordered trees, spin glasses, and traveling waves. *J. Statist. Phys.*, **51** (1988), 817–840.

[8] Durrett, R. :"Probability-Theory and Examples", 2nd Ed., Duxbury Press, 1995.

[9] Fisher, D. S., Huse, D. A. : Directed paths in random potential, Phys. Rev. B, **43**, 10 728 -10 742 (1991).

[10] Huse, D. A., Henley, C. L.: Pinning and roughening of domain wall in Ising systems due to random impurities, Phys. Rev. Lett. **54**, 2708–2711, (1985).

[11] Imbrie, J.Z., Spencer, T.: Diffusion of directed polymer in a random environment, J. Stat. Phys. **52**, Nos 3/4, (1988), 609-626.

[12] Johansson, K.: Shape fluctuations and random matrices. Comm. Math. Phys. **209**, (2000), no. 2, 437–476.

[13] Johansson, K.: Transversal fluctuations for increasing subsequences on the plane. Probab. Theory Related Fields **116** (2000), no. 4, 445–456.

[14] Kahane, J. P. and Peyriere, J. : Sur certaines martingales de Benoit Mandelbrot, Adv. in Math. **22**, 131–145, (1976).

[15] Kardar, M., Parisi, G., Zhang, Y.-C.: Dynamical scaling of growing interfaces, *Phys. Rev. Lett.* **56** (1986), 889–892.

[16] Kesten, H.: Aspect of first passage percolation, In École d'Éte de Probabilités de Saint-Flour XIV, Springer Lecture Notes in Mathematics, **1180**, 126–263, (1986).

[17] Krug, H. and Spohn, H.: Kinetic roughening of growing surfaces. In: *Solids Far from Equilibrium*, C. Godrèche ed., Cambridge University Press (1991).

[18] Licea, C., Newman, C., Piza, M., 1996: Superdiffusivity in first-passage percolation. Probab. Theory Related Fields **106**, no. 4, 559–591.

[19] Liggett, T.M., 1985: "Interacting Particle Systems", Springer Verlag, Berlin-Heidelberg-Tokyo.

[20] Mejane, O. 2002: Upper bound of a volume exponent for directed polymers in a random environment, preprint XXX.

[21] Newman, C., Piza, M.: Divergence of shape fluctuations in two dimensions, Ann. Probab. **23** (1995), No. 3, 977–1005.

[22] Olsen, P.and Song, R.: Diffusion of directed polymers in a strong random environment. J. Statist. Phys. **83** (1996),no. 3-4, 727–738.

[23] Petermann, M., 2000: Superdiffusivity of directed polymers in random environment, Ph.D. Thesis Univ. Zürich.

[24] Piza, M.S.T., 1997: Directed polymers in a random environment: some results on fluctuations, J. Statist. Phys. **89** , no. 3-4, 581–603.
[25] Song, R., Zhou, X.Y., 1996: A remark on diffusion on directed polymers in random environment, J. Statist. Phys. **85**, Nos.1/2, 277–289.
[26] Spitzer, F.: "Principles of Random Walks", Springer Verlag, New York, Heiderberg, Berlin (1976).
[27] Spohn, H.: Private communications.
[28] Sznitman, A.-S., : *Brownian Motion, Obstacles and Random Media*, Springer monographs in mathematics, Springer, 1998.
[29] Tracy, C. A. ; Widom, H.: Level spacing distributions and the Airy kernel, Comm. Math. Phys. **159**, (1994), 151–174.
[30] Wüthrich, Mario V. : Scaling identity for crossing Brownian motion in a Poissonian potential. *Probab. Theory Related Fields* **112** (1998), no. 3, 299–319.
[31] Wüthrich, Mario V. : Superdiffusive behavior of two-dimensional Brownian motion in a Poissonian potential. *Ann. Probab.* **26** (1998), no. 3, 1000–1015.
[32] Wüthrich, Mario V. : Fluctuation results for Brownian motion in a Poissonian potential. *Ann. Inst. H. Poincaré Probab. Statist.* **34** (1998), no. 3, 279–308.

Francis Comets
Université Paris 7
Mathématiques
Case 7012
2 place Jussieu
75251 Paris
France
E-mail address: `comets@math.jussieu.fr`

Tokuzo Shiga
Tokyo Institute of Technology
Oh-okayama, Meguroku
Tokyo 152-8551
Japan
E-mail address: `tshiga@math.titech.ac.jp`

Nobuo Yoshida
Division of Mathematics
Graduate School of Science
Kyoto University
Kyoto 606-8502
Japan
E-mail address: `nobuo@kusm.kyoto-u.ac.jp`

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 143–171

Entropy Pairs and Compensated Compactness for Weakly Asymmetric Systems

József Fritz

Abstract.

The hyperbolic (Euler) scaling limit of weakly asymmetric Ginzburg–Landau models with a single conservation law is investigated, weak asymmetry means that the microscopic viscosity of the system tends to infinity in a prescribed way during the hydrodynamic limit. The system is not attractive, its potential is a bounded perturbation of a quadratic function. The macroscopic equation reads as $\partial_t \rho + \partial_x S'(\rho) = 0$, where S is a convex function. The Tartar - Murat theory of compensated compactness is extended to microscopic systems, we prove weak convergence of the scaled density field to the set of weak solutions. In the attractive case of a convex potential this set consists of the unique entropy solution. Our main tool is the logarithmic Sobolev inequality of Landim, Panizo and Yau for continuous spins.

§1. Introduction and Main Result

In the last fifteen years a great progress has been made in the theory of hydrodynamic limits. Although the first papers [2] and [22] concern hyperbolic problems, most results are related to diffusive systems, see e.g. [13,29,30] and the monograph [15] with historical notes and further references. The main difficulty in hyperbolic problems comes from the breakdown of regularity and uniqueness of macroscopic solutions. In a smooth regime the relative entropy method of Yau [32] works well in quite general situations. Beyond shocks, however, only some attractive

---

Received January 6, 2003.

Revised June 3, 2003.

Partially supported by Hungarian Science Foundation Grants T26176 and T37685.

Mathematics Subject Classification: Primary 60K31, secondary 82C22.

Keywords: Ginzburg–Landau models, hyperbolic scaling, Lax entropy pairs, compensated compactness.

systems like asymmetric exclusions, zero range and stick processes are tractable, see [1,4,21,23,24] and also [16,31] on entropy and large deviations for asymmetric exclusion processes. The specific structure of these models is very important, and PDE techniques play an essential role in the proofs. The main purpose of this paper is to develop a general method for hyperbolic problems: we are going to extend the Tartar - Murat theory of compensated compactness to microscopic (stochastic) systems, see [26] and [19] for the first ideas, [14] or [25] for a systematic treatment of these advanced PDE techniques. Compensated compactness yields weak convergence of the scaled empirical density to the set of weak solutions to the macroscopic equations. A first exposition of the main ideas for stochastic systems was given in [9] in the case of asymmetric exclusions, and also for a lattice gas with two conservation laws. Here we study an asymmetric Ginzburg–Landau model with a single conservation law in details; we wanted to demonstrate that this method is really applicable. Another model, a two-component lattice gas with collisions is to be discussed in a forthcoming paper [11], see also [27,28] for a large class of two-component models. To have convergence of the scaled microscopic process to a well specified macroscopic solution, one has to supplement compensated compactness with the entropy condition of Lax and Kružkov implying the uniqueness of the limiting macroscopic solution. Unfortunately, we can only prove this condition for attractive Ginzburg–Landau models by adapting the coupling method of Rezakhanlou [21]. In another paper [10] we investigate non-attractive lattice gas models with a single conservation law. The structure of these systems allows us to verify also the entropy condition, thus we get convergence to a single entropy solution specified by its initial value.

Let $\eta_k(t) \in \mathbb{R}$ for $t \geq 0$, $k \in \mathbb{Z}$, and consider the following infinite system of stochastic differential equations as the evolution law of this continuous spin model. Given a potential $V(y) = y^2/2 + U(y)$ such that U, U', U'' are bounded,

$$\begin{aligned}(1.1) \qquad d\eta_k &= \tfrac{1}{2}\,(V'_{k-1} - V'_{k+1})\,dt \\ &\quad + \sigma(\varepsilon)\,(V'_{k+1} + V'_{k-1} - 2V'_k)\,dt + \sqrt{2\sigma(\varepsilon)}\,(dw_{k-1} - dw_k)\,,\end{aligned}$$

where w_k, $k \in \mathbb{Z}$ is a family of independent Wiener processes, $\sigma = \sigma(\varepsilon) > 1/2$ is the coefficient of microscopic viscosity; abbreviations like $V'_k := V'(\eta_k)$ are widely used also later on. The scaling parameter, $0 < \varepsilon \to 0$ of the hydrodynamic limit is interpreted as the spacing of the lattice in macroscopic units, hyperbolic scaling means that time is speeded up by a factor of $1/\varepsilon$. We shall let σ depend on ε during the limiting procedure in such a way that $\varepsilon\sigma(\varepsilon) \to 0$ as $\varepsilon \to 0$, thus the effect of the second difference, $\sigma\,(V'_{k+1}+V'_{k-1}-2V'_k)$ diminishes as $\varepsilon \to 0$. A technical condition, $\varepsilon\sigma^2(\varepsilon) \to \infty$ will be explained later.

Since the drift is Lipschitz continuous in a weighted ℓ^2 space, Ω of doubly infinite sequences $\eta = (\eta_k : k \in \mathbb{Z})$ with weights $e^{-|k|}$, for instance, the existence of unique strong solutions to (1.1) follows by a standard iteration procedure, and Ω carries a large class of probability measures, see e.g. [9] for further references. Let $\mathcal{F}_{k,l}$ denote the σ-field generated by $\eta_{k,l} := (\eta_j : k-l < j \le k)$, μ_t is the distribution of the evolved configuration $\eta(t)$, and $\mu_{t,k,l}$ denotes the distribution of $\eta_{k,l}(t)$. Short hand notation is to be used later in case of $k = n$ and $l = 2n$.

The total spin $\sum \eta_k$ is formally preserved by the evolution, and certain product measures λ_z with one dimensional marginal densities g_z, $z \in \mathbb{R}$,

$$g_z(y) := \exp\bigl(zy - V(y) - F(z)\bigr)\,, \quad F(z) := \log \int_{-\infty}^{\infty} e^{zy - V(y)}\,dy$$

are all stationary states. As a reference measure, $\lambda := \lambda_0$ will be used; we may and do assume that $F(0) = F'(0) = 0$. A converse statement on stationary states in a much stronger form will be needed, our main tool is the logarithmic Sobolev inequality of [17]; that is why we are assuming that V is a bounded perturbation of a quadratic function. The model is attractive if V is convex; we are interested in the general case when an effective coupling is not available.

Due to the asymmetric part $(1/2)(V'_{k-1} - V'_{k+1})$ of the drift, the model admits a hyperbolic scaling as specified below. In the absence of the stochastic term $\sqrt{2\sigma}\,(dw_k - dw_{k-1})$, (1.1) looks like a lattice approximation procedure for solving $\partial_t\rho + \partial_x V'(\rho) = 0$; the viscid part $\sigma\,(V'_{k+1} + V'_{k-1} - 2V'_k)$ is needed even in this deterministic situation to stabilize the algorithm, see [14,18,25]. However, in regions of the phase space where V is concave, the viscid correction plays an opposite role; the convexity of V is very important in the deterministic case. Moreover, the value of σ may depend on the initial condition.

The behavior of the stochastic model is similar, but more complex. For $\varepsilon > 0$ interpreted as the macroscopic spacing of the lattice, let $\rho_\varepsilon(t,x) := \eta_k(t/\varepsilon)$ if $|x - k\varepsilon| < \varepsilon/2$ denote the empirical process; we are interested in its limiting behavior as the scaling parameter $\varepsilon \to 0$. In view of the principle of local equilibrium, the true distribution, $\mu_{t/\varepsilon}$ of our process is close to a product measure with marginal densities g_z such that z does depend on space and time. Since $\lambda_z(\eta_k) = F'(z) = \rho$ is the expectation of η_k with respect to λ_z, while $\lambda_z(V'_k) = z = S'(\rho)$ if $\rho = F'(z)$, where $S(\rho) := \sup_z \{z\rho - F(z)\}$, $\partial_t\rho + \partial_x S'(\rho) \approx$

$\varepsilon\sigma\partial_x^2 S'(\rho)$ is expected for the asymptotic mean of ρ_ε as $\varepsilon \to 0$. This was proven in [7] for $\sigma = \sigma_0/\varepsilon$ and small U by means of the parabolic perturbation technique of [6], see also [12,4] on the weakly asymmetric exclusion process. Heuristic considerations of this kind suggest that the macroscopic equation becomes $\partial_t \rho + \partial_x S'(\rho) = 0$ if $\varepsilon\sigma \to 0$ during the hydrodynamic limit. This can be proven by means of the relative entropy method [32] when the macroscopic solution is smooth, even if $\sigma > 0$ is fixed. In case of an incompressible limit (perturbation of equilibrium) the initial configuration is changed during the scaling limit, see [20,24,28].

In a regime of shocks some new methods are needed, this is the subject of the present paper. Unfortunately, we are able to control oscillations of the empirical process only if the microscopic viscosity, σ goes to infinity as $\varepsilon \to 0$. More precisely, we are assuming that $\varepsilon\sigma \to 0$ but $\varepsilon\sigma^2 \to +\infty$ as $\varepsilon \to 0$. For example, $\sigma(\varepsilon) := \sqrt{\varepsilon}\log(1/\varepsilon)$ is an allowed choice. Let us remark that the concept of microscopic viscosity is plausible in many other cases, and conditions on its growth rate are the same as above, see [9,10,11], but Dittrich [4] only needs $\varepsilon\sigma^3 \to +\infty$ in case of asymmetric exclusions. If the generator, $\mathfrak{L}$ of a conservative process decomposes as $\mathfrak{L} = \mathfrak{L}_0 + \sigma\mathfrak{G}$, where $\mathfrak{G}$ is symmetric, then the parameter $\sigma > 0$ may be interpreted as the microscopic viscosity of the model. The paper [28] investigates perturbation of the equilibrium for a class of two-component hyperbolic models in a smooth regime; the order of microscopic viscosity of these models is the same as here. Also in such situations we use the term *weakly asymmetric system*; perhaps the phrase *large microscopic viscosity limit* would be more correct.

A locally square integrable $\rho \in L^2_{\mathrm{loc}}(\mathbb{R}^2_+)$ is a *weak solution* to $\partial_t \rho + \partial_x S'(\rho) = 0$ with initial value $\rho_0 \in L^2_{\mathrm{loc}}(\mathbb{R})$ if $\rho(t,x)$ satisfies

$$(1.2) \qquad \int_0^\infty \int_{-\infty}^\infty \left(\rho\psi_t' + S'(\rho)\psi_x'\right) dx\, dt + \int_{-\infty}^\infty \rho_0(x)\psi(0,x)\, dx = 0$$

for all test functions $\psi \in C_c^1(\mathbb{R}^2)$, where $\psi_u' := \partial_u \psi$, $C_c^k(\mathbb{R}^2)$ is the space of compactly supported $\psi : \mathbb{R}^2 \mapsto \mathbb{R}$ with k continuous derivatives, $\mathbb{R}^2_+ := [0,\infty)\times\mathbb{R}$, L^2_{loc} is the space of locally square integrable functions. It is easy to check that S'' is bounded, thus the definition above is not a senseless one. In fact, only the local integrability of ρ is needed in (1.2) because S' is linearly bounded, but we prefer an L^2 setting. In case of a single conservation law the *Lax entropy condition* is sufficient for the uniqueness of weak solutions, see [14] or [25]. For $\alpha > 0$ let $\mathcal{H}_\alpha$ denote the set of such couples (h, J) of continuously differentiable real functions that $|h(u)| + |J(u)| = O(|u|^\alpha)$ for large u, and $J' = h'S''$, that is $\partial_t h(\rho) + \partial_x J(\rho) = 0$ along classical solutions; (h, J) is called an *Lax entropy pair.* A weak solution, ρ satisfies the entropy condition if

$$\int_0^\infty \int_{-\infty}^\infty \big(h(\rho)\psi_t' + J(\rho)\psi_x'\big)\,dx\,dt + \int_{-\infty}^\infty h(\rho_0(x))\psi(0,x)\,dx \geq 0 \tag{1.3}$$

for all $0 \leq \psi \in C_c^1(\mathbb{R}^2)$ and $(h, J) \in \mathcal{H}_1$ with h convex. An equivalent version of the Lax inequality has been proposed by Kružkov, see [21,25], namely

$$\begin{aligned} \int_0^\infty \int_{-\infty}^\infty \big(|\rho - c|\,\psi_t' + |S'(\rho) - S'(c)|\,\psi_x'\big)\,dx\,dt \\ + \int_{-\infty}^\infty |\rho_0(x) - c|\,\psi(0,x)\,dx \geq 0 \end{aligned} \tag{1.4}$$

for all $0 \leq \psi \in C_c^1(R^2)$ and $c \in \mathbb{R}$; the flux of $h_c(u) := |u - c|$ can be chosen as $J_c(u) := |S'(u) - S'(c)|$.

The Lax inequality is motivated by the viscous approximation $\partial_t u_\varepsilon + \partial_x S'(u_\varepsilon) = \varepsilon\sigma\partial_x^2\Phi'(u_\varepsilon)$, where $\varepsilon\sigma \to 0$, because

$$\begin{aligned} \partial_t h(u_\varepsilon) + \partial_x J(u_\varepsilon) &= \varepsilon\sigma h'(u_\varepsilon)\partial_x^2\Phi'(u_\varepsilon) \\ &= \varepsilon\sigma\partial_x\big(h'(u_\varepsilon)\Phi''(u_\varepsilon)\partial_x u_\varepsilon\big) - \varepsilon\sigma h''(u_\varepsilon)\Phi''(u_\varepsilon)(\partial_x u_\varepsilon)^2, \end{aligned} \tag{1.5}$$

whence one can derive (1.3) with an appropriate choice of Φ; the most favored one is $\Phi(u) = u^2/2$. We see that there is a freedom in choosing the viscid correction $\varepsilon\sigma\partial_x^2\Phi'$, but $\Phi'' \geq 0$ is very important at this point; the structure of lattice approximation procedures and other numerical schemes is similar. The viscous correction must be elliptic in all cases. Stochastic models are structured in a more canonical way because they must have a stationary state; the form of (1.1) is dictated by this requirement. Calculations at the microscopic level follow the above scheme of the viscous approximation, that is why the strict convexity of V is needed for (1.3) or (1.4).

Several topologies shall be used in the study of the limit distribution of the empirical process ρ_ε. We shall see that ρ_ε is locally square integrable, thus the weak topology of $L^2_{\mathrm{loc}}(\mathbb{R}^2_+)$ will be the first one; the distribution P_ε of ρ_ε will be considered on this space. We are interested in the limiting behavior of the density field R_ε,

$$R_\varepsilon(\psi) := \int_0^\infty \int_{-\infty}^\infty \psi(t,x)\,\rho_\varepsilon(t,x)\,dx\,dt, \tag{1.6}$$

where $\psi \in C_c(\mathbb{R}^2)$ is compactly supported. The initial conditions are specified in terms of a family $\mu_{\varepsilon,0} : \varepsilon > 0$ of initial distributions. First of all, we have some $\rho_0 \in L^2_{\text{loc}}(\mathbb{R})$ such that

$$\lim_{\varepsilon\to 0}\int_{-\infty}^{\infty}\varphi(x)\rho_\varepsilon(0,x)\,dx = \int_{-\infty}^{\infty}\varphi(x)\rho_0(x)\,dx \tag{1.7}$$

in probability for all $\varphi \in C_c(\mathbb{R})$. The second condition tells that the entropy, S of $\mu_{\varepsilon,0}$ relative to $\lambda := \lambda_0$ is extensive. The entropy of a probability measure ν relative to λ is defined as $S[\nu|\lambda] := \nu(\log f)$ if $\nu \ll \lambda$ and $d\nu = f d\lambda$, $S = +\infty$ otherwise. Let $\mu_{\varepsilon,0,n}$ denote the restriction of $\mu_{\varepsilon,0}$ to $\mathcal{F}_{n,2n}$, and suppose that $f_n := d\mu_{\varepsilon,0,n}/d\lambda$ satisfies

$$S_n[\mu_{\varepsilon,0}|\lambda] := \int f_{\varepsilon,n}\log f_{\varepsilon,n}\,d\lambda \le C_0\, n \quad \forall\varepsilon > 0 \text{ and } n \in \mathbb{N}. \tag{1.8}$$

Our main result is

Theorem 1.1. *Suppose* (1.7), (1.8) *and specify* $\sigma = \sigma(\varepsilon)$ *such that* $\varepsilon\sigma(\varepsilon) \to 0$ *but* $\varepsilon\sigma^2(\varepsilon) \to +\infty$ *as* $\varepsilon \to 0$. *Then the family* P_ε *is tight as* $\varepsilon \to 0$, *and any limit distribution is concentrated on a set of weak solutions* (1.2) *to the macroscopic equation* $\partial_t\rho + \partial_x S'(\rho) = 0$. *Moreover, if* V *is strictly convex, then we have a weak solution* $\rho \in L^2_{loc}(\mathbb{R}^2_+)$ *such that*

$$\lim_{\varepsilon\to 0} R_\varepsilon(\psi) = R(\psi) := \int_0^\infty\int_{-\infty}^{\infty}\psi(t,x)\rho(t,x)\,dx\,dt$$

in probability for all $\psi \in C_c(\mathbb{R}^2)$; *this* ρ *is uniquely specified by its initial value* ρ_0 *and the entropy condition* (1.4).

The paper is organized as follows. Below and in Section 2 we are going to exhibit the main ideas of the argument. Section 3 summarizes some basic facts on the microscopic model, further technical details are added in Section 4. The proof is then completed in the last section.

The first main step of the proof is certainly the replacement of $V'(\rho_\varepsilon)$ with $S'(\rho_\varepsilon)$, this characteristic argument of hydrodynamic limits does not appear in PDE theory. The second step is then to show that the weak limit of $S'(\rho_\varepsilon)$ equals $S'(\rho)$, where ρ is the weak limit of ρ_ε. As we have learned from [13], the replacement of V' with S' can be done at a level of block averages. In case of a diffusive scaling the celebrated two-block estimate allows us to work with macroscopic blocks, thus the weak limit commutes with S'. This step is more difficult if we consider a hyperbolic problem because the two-block lemma extends to blocks of size $l = o(\sqrt{\sigma/\varepsilon})$ only, consequently there is no direct argument to identify the weak limit of $S'(\rho_\varepsilon)$. The concept of measure solution plays an important role at this point, see e.g. [3] on partial differential equations, and [29] on a first application to a microscopic system.

Let Θ denote the set of measurable families $\theta = \theta_{t,x} : (t,x) \in \mathbb{R}_+^2$ of probability measures on $\mathbb{R}$ such that $\theta_{t,x}(u^2)$ is locally integrable on $\mathbb{R}_+^2$. $\theta \in \Theta$ is a *measure solution* to (1.1) if

$$\int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \theta_{t,x}(d\rho)\big(\rho\psi_t' + S'(\rho)\psi_x'\big)\,dx\,dt = 0 \tag{1.9}$$

for all $\psi \in C_c^1(\mathbb{R}_+^2)$, the space of $\psi \in C^1(\mathbb{R}^2)$ such that supp ψ is contained in the interior of $\mathbb{R}_+^2$. Notice that the initial value has not been included in this definition. A function $u \in L^2_{\text{loc}}(\mathbb{R}_+^2)$ is represented by a family $\theta \in \Theta$ of Dirac measures such that $\theta_{t,x}$ is concentrated at the actual value $u(t,x)$ of u; this θ is called the *Young representation* of u. Moreover, any $\theta \in \Theta$ can be identified as a locally finite measure m_θ, $dm_\theta := dt\,dx\,\theta_{t,x}(du)$ on $\mathbb{R}_+^3 := \mathbb{R}_+^2 \times \mathbb{R}$; equip Θ with the associated weak topology. Therefore any weak solution is a measure solution, and the existence of measure solutions follows by a direct compactness argument. Compensated compactness is the tool for proving that any measure solution is actually a weak solution. We say that $\theta \in \Theta$ admits a *Tartar factorization* for a couple $(h_1, J_1), (h_2, J_2) \in \mathcal{H}_1$ of entropy pairs, if for almost every $(t,x) \in \mathbb{R}_+^2$ we have

$$\theta_{t,x}(h_1 J_2) - \theta_{t,x}(h_2 J_1) = \theta_{t,x}(h_1)\theta_{t,x}(J_2) - \theta_{t,x}(h_2)\theta_{t,x}(J_1)\,. \tag{1.10}$$

In the case of a single conservation law like $\partial_t \rho + \partial_x S'(\rho) = 0$, Tartar's factorization implies that θ is a family of Dirac measures, that is a weak solution. To get uniqueness of weak solutions we need the Kružkov inequality (1.4). The entropy condition can also be stated at the level of measure solutions,

$$\begin{aligned}\int_0^\infty \int_{-\infty}^\infty &\big(\theta_{t,x}(h_c)\psi_t' + \theta_{t,x}(J_c)\psi_x'\big)\,dx\,dt \\ &+ \int_{-\infty}^\infty \theta_{0,x}(h)\psi(0,x)\,dx \geq 0\end{aligned} \tag{1.11}$$

for all $0 \leq \psi \in C_c^1(\mathbb{R}^2)$ and for the Kružkov entropy pairs (h_c, J_c), $c \in \mathbb{R}$, see (1.4); the derivation of (1.11) is easier than that of (1.4). Let us remark that DiPerna [3] proves the uniqueness of measure solutions satisfying (1.11) without any reference to Tartar's factorization, but his initial condition is much stronger than that we do have here. Compensated compactness requires large microscopic viscosity, but it has an advantage from the point of view of uniqueness. Since we have weak solutions, (1.4) is sufficient, i.e. no continuity condition is needed at time zero, see [22].

Entropy pairs constitute additional conservation laws at the macroscopic level, but the microscopic model must be ergodic, thus it can not have any other conservation law than those we are a priori given. Therefore the Lax entropies exhibit rapid oscillations, they should be controlled by means of non-gradient tools as initiated by Varadhan [30].

§2. Compensated Compactness

The proof of Tartar's factorization is based on some functional analytic properties of the Lax entropy production $X := \partial_t h + \partial_x J$, we have to estimate X in various spaces. Let $||\varphi||$ denote the uniform norm, $||\varphi||_p$ is the L^p norm of $\varphi : \mathbb{R}^2 \mapsto \mathbb{R}$ for $p \geq 1$. The Sobolev space $H_{+1}(\mathbb{R}^2)$ is defined as the completion of $C_c^1(\mathbb{R}^2)$ with respect to $|| \cdot ||_{+1}$, $||\varphi||_{+1}^2 := ||\varphi||_2^2 + ||\varphi_t'||_2^2 + ||\varphi_x'||_2^2$, and $H_{-1}(\mathbb{R}^2)$ is the dual of H_{+1} with respect to $L^2(\mathbb{R}^2)$. Here and below we adopt a convention: if a function u is only defined on $\mathbb{R}_+^2$, then we extend its definition by setting $u(t,x) = 0$ for $t < 0$.

A first version of Tartar's theorem can be stated as follows. Let $(h_i, J_i) \in \mathcal{H}_1$ for $i = 1,2$, and set $X_{i,\varepsilon} := \partial_t h_i(u_\varepsilon) + \partial_x J_i(u_\varepsilon)$. Suppose that u_ε , $h_i(u_\varepsilon)$ and $J_i(u_\varepsilon)$ are all weakly convergent in $L^2(\mathbb{R}^2)$ as $\varepsilon \to 0$, while the Young representation θ_ε of u_ε tends to some $\theta \in \Theta$. If the set $\{X_{i,\varepsilon} : i = 1,2; \varepsilon > 0\}$ is relative compact in $H_{-1}(\mathbb{R}^2)$, then (1.10) holds true. The so called Murat lemma on the conditions of Tartar's theorem had certainly been motivated by (1.5). It states that if $h_i(u_\varepsilon)$ and $J_i(u_\varepsilon)$ are bounded in $L^p(\mathbb{R}^2)$ for some $p > 2$, and $X_{i,\varepsilon} = Y_{i,\varepsilon} + Z_{i,\varepsilon}$ such that $Z_{i,\varepsilon}$ is bounded in the space of finite signed measures on $\mathbb{R}^2$, while $Y_{i,\varepsilon}$ belongs to a compact set of $H_{-1}(\mathbb{R}^2)$, then $X_{i,\varepsilon}$ also lies in a compact subset of $H_{-1}(\mathbb{R}^2)$. Since the empirical process does not vanish at infinity, we have to localize the problem by multiplying X with a general $\phi \in C_c^2(\mathbb{R}^2)$; this step is also present in the original papers [26] and [19]. "Compensation" appears at two places. The factorization on the right hand side of (1.10) holds true only for the difference on the left, and $\partial_t h$, $\partial_x J$ alone are only bounded in H_{-1} , their sum does belong to a compact set.

In view of our project, we formulate and prove Tartar's factorization and the entropy inequality at the microscopic level, this will be done in terms of block averages. For any sequence ξ indexed by $\mathbb{Z}$,

$$(2.1)\qquad \bar{\xi}_{l,k} := \frac{1}{l}\sum_{j=0}^{l-1}\xi_{k-j} \quad\text{and}\quad \hat{\xi}_{l,k} := \frac{1}{l^2}\sum_{j=-l}^{l}||j|-l|\,\xi_{k+j}\,.$$

For example, $\bar{V}'_{l,k}$ refers to the sequence $V'_k = V'(\eta_k)$. The smooth averaging $\hat{\xi}_l$ seems to be convenient in analytic calculations, while the usual one, $\bar{\xi}_l$ is preferred in computing canonical expectations. The size $l = l(\varepsilon)$ of these blocks should be chosen in such a way that

$$(2.2)\qquad \limsup_{\varepsilon\to 0}\frac{\sigma(\varepsilon)}{\varepsilon l^3(\varepsilon)} < +\infty \quad\text{and}\quad \lim_{\varepsilon\to 0}\frac{l(\varepsilon)}{\sigma(\varepsilon)} = 0\,,$$

thus $\varepsilon l^2(\sigma) \to +\infty$ as $\varepsilon \to 0$. Since $\varepsilon\sigma(\varepsilon) \to 0$ and $\varepsilon\sigma^2(\varepsilon) \to +\infty$, $\sigma^2 = o(l^3)$. We see also that $(\sigma/\varepsilon)^{1/3} = O(l) = o(\sigma)$, thus the integer part of $\varepsilon^{-1/4}\sqrt{\sigma(\varepsilon)}$ is an acceptable choice for l. Because of some technical reasons, we modify the empirical process as $\hat{\rho}_\varepsilon(t,x) := \hat{\eta}_{l,k}(t/\varepsilon)$ if $|x - k\varepsilon| < \varepsilon/2$, $\hat{P}_\varepsilon$ denotes its distribution on $L^2_{\mathrm{loc}}(\mathbb{R}^2_+)$; from now on the block size $l = l(\varepsilon)$ is specified according to (2.2). In view of the Young representation, the empirical process $\hat{\rho}_\varepsilon$ can be considered also as a random element $\hat{\theta}_\varepsilon$ of Θ; the distribution, $\hat{P}_{\theta,\varepsilon}$ of $\hat{\theta}_\varepsilon$ is defined on this space. Of course, P_ε, $\hat{P}_\varepsilon$ and $\hat{P}_{\theta,\varepsilon}$ are not really different from each other, just the notion of weak convergence varies.

The microscopic version of entropy production $X = \partial_t h + \partial_x J$ is defined for $\psi \in C^1_c(\mathbb{R}^2_+)$ and $(h,J) \in \mathcal{H}_1$ by

$$(2.3)\qquad X_\varepsilon(\psi,h) := -\int_0^\infty\int_{-\infty}^\infty\bigl(h(\hat{\rho}_\varepsilon)\psi'_t + J(\hat{\rho}_\varepsilon)\psi'_x\bigr)\,dx\,dt\,,$$

remember that ψ is compactly supported in the interior of $\mathbb{R}^2_+$. We have

$$(2.4)\qquad \begin{aligned} X_\varepsilon(\psi,h) &= N_\varepsilon(\psi,h) + M_\varepsilon(\psi,h)\\ &\quad + \frac{1}{\varepsilon}\int_0^\infty\int_{-\infty}^\infty \psi(t,x)\bigl(\mathfrak{L}h(\hat{\rho}_\varepsilon) + \varepsilon\widetilde{\nabla}_\varepsilon J(\hat{\rho}_\varepsilon)\bigr)\,dx\,dt\,,\end{aligned}$$

where N_ε is a numerical error due to the lattice approximation of the space derivative, M_ε is a stochastic integral coming from the Ito lemma, and $\mathfrak{L} = \mathfrak{L}_0 + \sigma\mathfrak{G}$ is the generator of the microscopic process (1.1). On smooth cylinder functions $\varphi(\eta)$, $\mathfrak{L}_0$ and $\mathfrak{G}$ are acting as

$$\mathfrak{L}_0\varphi := -\sum_{k\in\mathbb{Z}}(\widetilde{\nabla}_1 V'_k)\partial_k\varphi\,,\quad \mathfrak{G}\varphi := \sum_{k\in\mathbb{Z}}(\nabla_1\partial_k - \nabla_1 V'_k)\nabla_1\partial_k\varphi\,,$$

where $\partial_k\varphi := \partial\varphi/\partial\eta_k$, $\nabla_l\xi_k := l^{-1}(\xi_{k+l} - \xi_k)$, $\widetilde{\nabla}_l := (1/2)(\nabla_l - \nabla_l^*)$, $\nabla_l^*\xi_k := l^{-1}(\xi_{k-l} - \xi_k)$, $\Delta_l := -\nabla_l^*\nabla_l$ for $l \in \mathbb{N}$. Note that $\nabla_1\hat{\xi}_l = \nabla_l\bar{\xi}_l$. The formalism is used also for functions as $\varepsilon\nabla_\varepsilon\varphi(x) := \varphi(x+\varepsilon) - \varphi(x)$, $\widetilde{\nabla}_\varepsilon\varphi(x) := (1/2\varepsilon)\big(\varphi(x+\varepsilon) - \varphi(x-\varepsilon)\big)$, and so on.

Mimicking integration by parts, the numerical error becomes

$$N_\varepsilon(\psi, h) = \int_0^\infty \int_{-\infty}^\infty J(\hat{\rho}_\varepsilon)(\widetilde{\nabla}_\varepsilon - \partial_x)\psi(t,x)\, dx\, dt. \tag{2.5}$$

The stochastic equations for $\hat{\eta}$ read as

$$d\hat{\eta}_{l,k} = -\widetilde{\nabla}_1\hat{V}'_{l,k}\, dt + \sigma\Delta_1\hat{V}'_{l,k}\, dt + \sqrt{2\sigma}\, \nabla_1^*\, d\hat{w}_{l,k}\, ,$$

thus scaling the noise as $\hat{\zeta}(t,x) := \sqrt{\varepsilon}\, \hat{w}_{l,k}(t/\varepsilon)$ if $|x - k\varepsilon| < \varepsilon/2$,

$$M_\varepsilon(\psi, h) = \sqrt{2\sigma\varepsilon} \int_{-\infty}^\infty \int_0^\infty \psi(t,x)h'(\hat{\rho}_\varepsilon)\nabla_\varepsilon^*\hat{\zeta}_\varepsilon(dt,x)\, dx\, . \tag{2.6}$$

Splitting $\mathfrak{L}/\varepsilon$ into its asymmetric and symmetric components, we obtain a decomposition $X_\varepsilon = N_\varepsilon + M_\varepsilon + X_{a,\varepsilon} + X_{s,\varepsilon}$, where

$$X_{a,\varepsilon}(\psi, h) := \frac{1}{\varepsilon}\int_0^\infty \int_{-\infty}^\infty \psi(t,x)\big(\mathfrak{L}_0 h(\hat{\rho}_\varepsilon) + \varepsilon\widetilde{\nabla}_\varepsilon J(\hat{\rho}_\varepsilon)\big)\, dx\, dt\, , \tag{2.7}$$

$$X_{s,\varepsilon}(\psi, h) := \frac{\sigma}{\varepsilon}\int_0^\infty \int_{-\infty}^\infty \psi(t,x)\, \mathfrak{G} h\big(\hat{\rho}_\varepsilon(t,x)\big)\, dx\, dt\, . \tag{2.8}$$

The main term here is certainly the asymmetric $X_{a,\varepsilon}(\psi, h)$.

Having in mind (1.5) and the Tartar - Murat theorems, we are looking for a decomposition of entropy production $X_\varepsilon(\psi, h) = Y_\varepsilon(\psi, h) + Z_\varepsilon(\psi, h)$ described as follows.

Proposition 2.1. *Let* $(h_1, J_1), (h_2, J_2) \in \mathcal{H}_1$, *and suppose that we are given some random functionals* $Y_\varepsilon(\psi, h_i)$, $Z_\varepsilon(\psi, h_i)$, $A_\varepsilon(\phi)$, $B_\varepsilon(\phi)$ *such that* $X_\varepsilon = Y_\varepsilon + Z_\varepsilon$, $A_\varepsilon(\phi)$ *and* $B_\varepsilon(\phi)$ *do not depend on* ψ , *moreover*

$$|Y_\varepsilon(\phi\psi, h_i)| \le A_\varepsilon(\phi)\|\psi\|_{+1}\, , \quad |Z_\varepsilon(\phi\psi, h_i)| \le B_\varepsilon(\phi)\|\psi\|$$

for each $\psi \in C_c^1(\mathbb{R}_+^2)$, $\phi \in C_c^2(\mathbb{R}^2)$, $i = 1, 2$ *and* $\varepsilon > 0$. *If* $\|\phi\rho_\varepsilon\|_2^2 \le B_\varepsilon(\phi)$, $\mathsf{E}A_\varepsilon(\phi) \to 0$ *and* $\limsup \mathsf{E}B_\varepsilon(\phi) < +\infty$ *as* $\varepsilon \to 0$, *then* $\widehat{P}_{\theta,\varepsilon} : \varepsilon > 0$ *is tight on* Θ , *and* (1.10) *holds true with probability one with respect to any weak limit point* $\widehat{P}_\theta$ *of* $\widehat{P}_{\theta,\varepsilon}$ *as* $\varepsilon \to 0$.

This is a microscopic (stochastic) synthesis of the fundamental results of L. Tartar and F. Murat on compensated compactness. We postpone its proof to the last section, the main problem is to verify the conditions; that is the content of Sections 3 and 4. The first part of Theorem 1.1 follows from Proposition 2.1 and Lemma 5.1.

For the Lax–Kružkov inequality we do not need bounds that are uniform in ψ, but the viscid term, $\sigma\Delta_1 V'$ of the microscopic evolution must be elliptic as a (nonlinear) operator on the configuration space.

Proposition 2.2. *Suppose all conditions of Theorem 1.1 including the strict convexity of V, then $\hat{P}_{\theta,\varepsilon} : \varepsilon > 0$ is a tight family with respect to the weak topology of Θ, and its weak limit distributions are concentrated on a set of measure solutions satisfying* (1.11).

The proof of this statement is based on the attractiveness of the microscopic process due to monotonicity of V'. Following [21], it is presented in Section 5. The proof of Theorem 1.1 is then completed by weak uniqueness of entropy solutions. The case of a general (non-convex) potential is a formidable open problem.

§3. The a priori bounds

This section summarizes some estimates based on relative entropy and its rate of production, the fundamental entropy inequality $\nu(\varphi) \leq S[\nu|\lambda] + \log\lambda(e^{\varphi})$ will be used several times. The Donsker–Varadhan rate function of a probability measure $\nu \ll \lambda$ with respect to a self-adjoint generator, $\mathfrak{S}$ of a Markov process in $L^2(\lambda)$ is a Dirichlet form $D[\nu|\lambda,\mathfrak{S}] := -4\lambda(\sqrt{f}\,\mathfrak{S}\sqrt{f})$ when $f := d\nu/d\lambda$; for technical details see [13],[15] or [9] with further references. We consider (1.1) with an arbitrary, but fixed value of $\sigma > 1/2$, $\mu_{t,n}$ is the restriction of the evolved measure, μ_t to $\mathcal{F}_{n,2n}$, and $f_n(t,\eta)$ denotes the λ-density of $\mu_{t,n}$, if any. Set $S_n(t) := S[\mu_{t,n}|\lambda]$, while $D_n(t) := D[\mu_{t,n}|\lambda,\mathfrak{G}_{n,2n}]$, where

$$\mathfrak{G}_{k,l}\,\varphi := \sum_{j=k-l+1}^{k-1} (\nabla_1\partial_j - \nabla_1 V_j')\nabla_1\partial_j\varphi$$

for smooth φ. If $0 < f_n$ is differentiable then

$$D_n(t) = 4\sum_{k=1-n}^{n-1}\int(\nabla_1\partial_k\sqrt{f_n})^2\,d\lambda = \sum_{k=1-n}^{n-1}\int\frac{1}{f_n}(\nabla_1\partial_k f_n)^2\,d\lambda\,.$$

First we derive an explicit bound for S_n and the time integral of D_n.

Lemma 3.1. *If* $S_n(0) \leq C_0\, n$ *then*

$$S_n(t) + \sigma \int_0^t D_n(s)\, ds \leq C_1 \left(t + \sqrt{n^2 + \sigma t}\right)$$

for all $n \in \mathbb{N}$, *where* C_1 *is a constant depending only on* C_0 *and* U.

Proof. We follow the proof of Proposition 1 in [8], only the main steps are presented. Remember that λ is preserved by the deterministic process generated by $\mathfrak{L}_0$, i.e. $\lambda(\mathfrak{L}_0\varphi) = 0$, while $\mathfrak{G}$ is symmetric in $L^2(\lambda)$, thus

$$\int \varphi \mathfrak{G} \psi\, d\lambda = -\sum_{k\in\mathbb{Z}} \int (\nabla_1 \partial_k \varphi) \nabla_1 \partial_k \psi\, d\lambda$$

for smooth cylinder functions φ and ψ. If $f_n > 0$ is smooth enough, then by a direct calculation

$$\begin{aligned} \partial_t S_n &= \int (\partial_t + \mathfrak{L}) \log f_n(t,\eta)\, \mu_t(d\eta) = \int f_{n+1} \mathfrak{L} \log f_n\, d\lambda \\ &= \sum_{k\in\mathbb{Z}} \int f_{n+1} (\tilde{\nabla}_1 V_k') \frac{\partial_k f_n}{f_n}\, d\lambda - \sigma \sum_{k\in\mathbb{Z}} \int (\nabla_1 \partial_k f_{n+1}) \frac{\nabla_1 \partial_k f_n}{f_n}\, d\lambda \\ &= -\sigma D_n - \sigma D_{\partial,n} + \sum_{k\in\mathbb{Z}} \int (f_{n+1} - f_n)(\tilde{\nabla}_1 V_k') \frac{\partial_k f_n}{f_n}\, d\lambda \\ &- \sigma \sum_{k\in\mathbb{Z}} \int (\nabla_1 \partial_k f_{n+1} - \nabla_1 \partial_k f_n) \frac{\nabla_1 \partial_k f_n}{f_n}\, d\lambda , \end{aligned}$$

where

$$D_{\partial,n}(t) := \int \frac{1}{f_n} (\partial_n f_n)^2\, d\lambda + \int \frac{1}{f_n} (\partial_{1-n} f_n)^2\, d\lambda$$

and $f_n = f_n(t,\eta)$. Both sums on the right hand side above consist only of boundary terms corresponding to $k = \pm n$, $\lambda(V_k') = 0\ \forall k$, and for $k = n+1$ or $k = -n$ we have

$$\int \varphi_n \partial_k f_{n+1}\, d\lambda = \int \varphi_n V_k' f_{n+1}\, d\lambda$$

whenever φ_n is $\mathfrak{F}_{n,2n}$ measurable. Denoting

$$B_n(t) := \frac{1}{2} \int \left(V_{n+1}' \partial_n f_n - V_{-n}' \partial_{1-n} f_n\right) \frac{f_{n+1}}{f_n}\, d\lambda ,$$

by an easy computation we arrive at

$$\begin{aligned}\partial_t S_n + \sigma\, D_n &= (1+2\sigma)\, B_n - \sigma D_{\partial,n} \\ &= B_n - \sigma \sum_{k=\pm n} \int (\nabla_1 \partial_k f_{n+1})\, \frac{\nabla_1 \partial_k f_n}{f_n}\, d\lambda \qquad (3.1) \\ &\leq B_n + \sigma \sqrt{D_{n+1} - D_n}\, \sqrt{D_{\partial,n}}\ ;\end{aligned}$$

at the final step $\nabla_1 \partial_n f_n = -\partial_n f_n$, $\nabla_1 \partial_{-n} f_n = \partial_{1-n} f_n$, the Schwarz inequality and convexity of D were used.

First of all we have to estimate B_n . For any probability measure ν, and $u \in \mathbb{R}$ we have an entropy bound

$$u\, \nu(V_k') \leq S[\nu|\lambda] + \log \lambda(e^{uV_k'}) \leq S[\nu|\lambda] + \frac{1}{2}\, \|V''\|\, u^2\,,$$

see (3.4) for the second inequality, whence by setting $u = \pm\sqrt{2S/\|V''\|}$ we obtain that $\nu^2(V_k') \leq 2\|V''\|\, S[\nu|\lambda]$. Let $\nu = \mu_t[\cdot|\mathfrak{F}_{n,2n}]$, again by Schwarz and convexity we get

$$B_n(t) \leq K_0 \sqrt{S_{n+1}(t) - S_n(t)}\, \sqrt{D_{\partial,n}(t)}\,.$$

In view of (3.1) there is nothing to prove if $(1+2\sigma)B_n \leq \sigma D_{\partial,n}$, but

$$\sigma D_{\partial,n} \leq 4B_n \leq 4K_0 \sqrt{(S_{n+1} - S_n) D_{\partial,n}}$$

in the opposite case, whence a system

$$\partial_t S_n + \sigma D_n \leq K_1 \big(S_{n+1} - S_n + \sigma\, \sqrt{S_{n+1} - S_n}\, \sqrt{D_{n+1} - D_n}\, \big)$$

of differential inequalities follows immediately, where K_1 depends only on $\|V''\|$. This system admits an explicit solution, see Lemma 3 in [8], the result is just the bound we have to prove. Since the final statement does not depend on smoothness of f_n any more, this restriction can be removed by a standard regularization. Q.E.D.

As a first consequence, from the entropy bound we get the moment condition $\limsup \mathsf{E}\|\phi\rho_\varepsilon\|_2^2 < +\infty$ of Proposition 2.1 for $\phi \in C_c^2(\mathbb{R}^2)$.

Lemma 3.2. *We have a universal constant C_2 such that*

$$\frac{1}{nt} \sum_{|k|<n} \int_0^t \int \eta_k^2\, d\mu_s\, ds \leq C_2 \Big(1 + \frac{t}{n} + \sqrt{1 + \sigma t/n^2} \Big)\,.$$

Proof. From the basic entropy inequality, $\nu(\varphi) \le S[\nu|\lambda] + \log \lambda(e^{\varphi})$, for any $\beta > 0$ we get

$$\frac{1}{n} \sum_{|k|<n} \mu_t(\eta_k^2) \le \frac{1}{\beta n} S_n(t) + \frac{2}{\beta} \log \lambda(e^{\beta \eta_k^2}) .$$

To estimate $\lambda(e^{\eta_k^2})$, let E_g denote expectation with respect to an $N(0, 2\beta)$ variable ζ, then $e^{\beta \eta_k^2} = \mathsf{E}_g e^{\zeta \eta_k}$, thus $\lambda(e^{\zeta \eta_k}) = e^{F(\zeta)}$, and $F(\zeta) \le (1/2)\|F''\|\,\zeta^2$ as $F(0) = F'(0) = 0$ by assumption. Since $F''(z)$ is just the variance of η_k under λ_z, $F''(z) \le \lambda_z((\eta_k - y)^2)$ e.g. if $z = V'(y)$. However, $(\eta_k - y)^2 \le a + b(V_k' - z)^2$ because $V''(x)$ is strictly positive for large $|x|$, while $\lambda_z((V_k' - z)^2) = \lambda_z(V_k'')$, we have $\|F''\| \le a + b\|V''\| < +\infty$. Finally,

$$\log \mathsf{E}_g e^{\gamma \zeta^2} = -\log \sqrt{1 - 4\gamma\beta} \le 4\gamma\beta \quad \text{whenever } 8\gamma\beta \le 1 , \tag{3.2}$$

which completes the proof via Lemma 3.1. Q.E.D.

The following lemma summarizes some results of [17]. For any linearly bounded $h \in C(\mathbb{R})$, and $\alpha_j \in \mathbb{R} : 0 \le j < l$ set $\hat{h}(\rho) := \lambda_z(h(\eta_k))$,

$$\phi_{l,k}(h, \alpha) := \sum_{j=0}^{l-1} \alpha_j \big(h(\eta_{k-j}) - \hat{h}(\bar{\eta}_{l,k})\big) ,$$

and $\Phi_h(\rho, u) := \log \lambda_z(e^{uh(\eta_k) - u\hat{h}(\rho)})$, where $z := S'(\rho)$.

Lemma 3.3. *We have positive constants l_0 and C_3 depending only on U such that if $l > l_0$, then any probability measure, ν on $\mathfrak{F}_{k,l}$ satisfies*

$$\beta \int \phi_{l,k}(h, \alpha)\, d\nu \le C_3 \left(1 + l^2 D[\nu|\lambda, \mathfrak{G}_{k,l}]\right) + \frac{1}{2} \log \int \exp\Big(\sum_{j=0}^{l-1} \Phi_h(\bar{\eta}_{l,k}, 2\beta\alpha_j)\Big)\, d\nu .$$

Proof. Given $\bar{\eta}_{l,k} = \rho$, denote $\bar{\nu}_{l,\rho}$ and $\bar{\lambda}_{l,\rho}$ the conditional distributions of $\eta_{k,l}$ under ν and λ, respectively. In view of LSI, which is Theorem 2.2 in [17], we have $S[\bar{\nu}_{l,\rho}|\bar{\lambda}_{l,\rho}] \le C_3' l^2 D[\bar{\nu}_{l,\rho}|\bar{\lambda}_{l,\rho}, \mathfrak{G}_{k,l}]$ for all ν and ρ with the same C_3', thus from the entropy bound

$$\beta \bar{\nu}_{l,\rho}(\phi_{l,k}) \le C_3'\, l^2 D[\bar{\nu}_{l,\rho}|\bar{\lambda}_{l,\rho}, \mathfrak{G}_{k,l}] + \log \bar{\lambda}_{l,\rho}(e^{\beta\phi_{l,k}}) . \tag{3.3}$$

Let $\bar{\lambda}_{l,m,\rho}$ denote the restriction of $\bar{\lambda}_{l,\rho}$ to $\mathfrak{F}_{k,m}$. If l is large enough, $z = S'(\rho)$ and $1 < m \le 1 + l/2$, then $d\bar{\lambda}_{l,m,\rho}/d\lambda_z$ is uniformly bounded in view Corollary 5.5 of [17]. Splitting $\phi_{l,k}(h,\alpha)$ into two pieces, by means of the Schwarz inequality we obtain that

$$\log \bar{\lambda}_{l,\rho}(e^{\beta\phi_{k,l}}) \leq \log C_3'' + \frac{1}{2} \log \lambda_z(e^{2\beta\phi_{l,k}}) \quad \text{if } z = S'(\rho)\,.$$

Since $\mathfrak{G}_{k,l}\,\bar{\eta}_{l,k} = 0$, we can integrate (3.3) with respect to ν ; notice that $D[\nu|\lambda_z, \mathfrak{G}_{k,l}]$ does not depend on z. Q.E.D.

From now on we are assuming that $l > l_0$ in all statements. On the rate of convergence to local equilibrium we have

Lemma 3.4. *There exists a universal constant* C_4 *such that*

$$\frac{1}{nl} \sum_{|k|<n} \int_0^t \int \left(\bar{V}'_{l,k} - S'(\bar{\eta}_{l,k})\right)^2 d\mu_s\, ds \leq C_4\, C_{t,n}(\sigma, l)\,,$$

where $C_{t,n}(\sigma,l) := t/l^2 + (l/\sigma)(1 + tn^{-1} + \sigma t n^{-2})$.

Proof. We apply Lemma 3.3 with $h = V'$, $\alpha_j = 1/l$ and $\beta = \beta_0 l$; for brevity we let $\phi = \phi_{l,k}(V', \alpha)$ and $\Phi(\rho, u) = \Phi_{V'}(\rho, u)$. First we show that $\lambda_z(e^{\beta\phi^2}) \leq C_4'$ if $z = S'(\bar{\eta}_{l,k})$ and β_0 is small. Since $e^{\beta\phi^2} = \mathsf{E}_g e^{\zeta\phi}$ if ζ is an $N(0, 2\beta)$ variable, and $\lambda_z(e^{\beta\phi}) = \mathsf{E}_g e^{l\Phi(\rho,\zeta/l)}$, the statement follows in the usual way by (3.2). Indeed, $\Phi(\rho, u) \leq \frac{1}{2}\,\|V''\|\,u^2$ for all $y \in \mathbb{R}$ because $\Phi(z, 0) = 0$ and, integrating by parts, we obtain a bound

$$\Phi_u'(\rho, u) = u \int V''(\eta_k) \exp\bigl(uV_k' - uz - \Phi(\rho, u)\bigr)\, d\lambda_z\,, \tag{3.4}$$

that is $|\Phi_u'(\rho, u)| \leq |u|\,\|V''\|$, whence $C_4' = O(\beta_0)$, thus

$$\beta_0 l \int \bigl(\bar{V}_{l,k} - S'(\bar{\eta}_{l,k})\, d\mu_s \leq C_3 + C_3 l^2 D[\mu_{s,k,l}|\lambda, \mathfrak{G}_{k,l}] + C_4'\,.$$

Doing summation for k and integrating with respect to time, the statement follows from Lemma 3.1 by subadditivity of D . Q.E.D.

Differences of various block averages are estimated by means of

Lemma 3.5. *Let* $\alpha_j \in \mathbb{R}$ *for* $0 \leq j < l$ *such that* $\sum \alpha_j = 0$, $\sum \alpha_j^2 \leq 1/l$, *and set* $\phi_{l,k}(1,\alpha) := \phi_{l,k}(h,\alpha)$ *when* $h(y) \equiv y$. *We have a universal* C_5 *such that*

$$\frac{1}{nl} \sum_{|k|<n} \int_0^t \int \phi_{l,k}^2(1,\alpha)\, d\mu_s\, ds \leq C_5\, C_{t,n}(\sigma, l)\,.$$

Proof. It is essentially the same as that of Lemma 3.4 with the only difference that at the final step, in the exponent we have

$$\sum_{j=0}^{l-1}\big(F(\alpha_j\zeta) - \alpha_j\zeta F'(z) - F(z)\big) = \frac{1}{2}\sum_{j=0}^{l-1} F''(\gamma_j)\alpha_j^2\zeta^2 \leq \frac{1}{2l}\|F''\|\zeta^2\,,$$

which completes the proof as $\|F''\| < +\infty$. Q.E.D.

The following lemma is essentially the two-block estimate of [13]. In particular, choosing $l = 2r$ we obtain a bound for $(\nabla_r \bar{V}'_{r,k})^2$.

Lemma 3.6. *We have a universal* C_6 *such that for* $2r \leq l$,

$$\frac{1}{nl}\sum_{|k|<n}\int_0^t \int (\bar{V}'_{r,k} - \bar{V}'_{l,k})^2\, d\mu_s\, ds \leq C_6\, C_{t,n}(\sigma, l, r)\,,$$

where $C_{t,n}(\sigma, l, r) := t/rl + (l/\sigma)(1 + tn^{-1} + \sigma t n^{-2})$.

Proof. This is a consequence of the previous lemma, but integrating by parts on the left hand side, it can directly be estimated by the Dirichlet form via the Schwarz inequality without any reference to LSI, see e.g. [8] for details. Q.E.D.

Now we are in a position to verify all conditions of Proposition 2.1.

§4. The Lax entropy production

We start with the explicit decomposition $X_\varepsilon = N_\varepsilon + M_\varepsilon + X_{a,\varepsilon} + X_{s,\varepsilon}$ of entropy production, see (2.3) and (2.5), (2.6), (2.7), (2.8). To get $X_\varepsilon = Y_\varepsilon + Z_\varepsilon$ as needed in Proposition 2.1, we split some terms into new ones, and each of them will be casted into one of two categories named by Y and Z according to the bound it satisfies. More precisely, a random functional $\Gamma_\varepsilon(\psi)$ is of type Y if for each $\phi \in C_c^2(\mathbb{R})$ we have a random bound $A_\varepsilon(\phi)$ such that $A_\varepsilon(\phi)$ does not depend on ψ,

$$|\Gamma_\varepsilon(\phi\psi)| \leq A_\varepsilon(\phi)\|\psi\|_{+1} \quad \text{and} \quad \lim_{\varepsilon\to 0} \mathsf{E}A(\phi) = 0.$$

Similarly, Γ is of type Z if

$$|\Gamma_\varepsilon(\phi\psi)| \leq A_\varepsilon(\phi)\|\psi\| \quad \text{and} \quad \limsup_{\varepsilon\to 0} \mathsf{E}A(\phi) < +\infty.$$

In case of terms of type Z we also indicate if the bound does, or does not vanish.

Throughout this section we deal with an entropy pair $(h, J) \in \mathcal{H}_1$ such that h' and h'' are bounded. All calculations are done at the microscopic level, thus the integral mean

$$\psi_k(t) := \frac{1}{\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} \psi(t\varepsilon, k\varepsilon + x)\phi(t\varepsilon, k\varepsilon + x)\, dx \tag{4.1}$$

appears at several places; the notation $H_k(t) := H(\bar{\eta}_{l,k}(t))$ shall also be used for functions $H \in C(\mathbb{R})$ like h, J, h', S'' and so on. $\phi \in C_c^2(\mathbb{R}^2)$ plays an explicit role only in

Lemma 4.1. *The stochastic integral M_ε is of type Y.*

Proof. This is the only case where we estimate the H_{-1} norm in a direct way by using Fourier transform; the underlying generalized function is just

$$m_\varepsilon(t, x) := \sqrt{2\varepsilon\sigma}\, h'(\hat{\rho}_\varepsilon(t, x))\phi(t, x)\partial_t \nabla_\varepsilon^* \hat{\zeta}_\varepsilon(t, x)\,.$$

In view of $|M_\varepsilon(\psi, h)| \le \|m_\varepsilon\|_{-1}\, \|\psi\|_{+1}$, we have to show that

$$\lim_{\varepsilon\to 0} \mathsf{E}\|m_\varepsilon\|_{-1}^2 = \lim_{\varepsilon\to 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\mathsf{E}|\tilde{m}_\varepsilon(\tau, \omega)|^2}{1 + \tau^2 + \omega^2}\, d\tau\, d\omega = 0\,, \tag{4.2}$$

where $\tilde{m}_\varepsilon$ denotes the Fourier transform of m_ε . In microscopic variables

$$\tilde{m}_\varepsilon(\tau, \omega) = \frac{\varepsilon\sqrt{2\sigma}}{l} \sum_{k\in\mathbb{Z}} \int_0^\infty \psi_k(t, \tau, \omega) h_k'(t)(d\bar{w}_{l,k-1} - d\bar{w}_{l,k+l-1})\,,$$

where $\psi_k(t, \tau, \omega)$ is defined by (4.1) with $\psi = (2\pi)^{-1} \exp(\imath t\tau + \imath x\omega)$. The sum of the integrands can be rewritten as a sum like $\sum \bar{\xi}_{l,k}\, dw_k$, thus a simple Ito calculus and $(\bar{\xi}_{l,k})^2 \le (\bar{\xi^2})_{l,k}$ result in

$$\begin{aligned}\mathsf{E}|\tilde{m}_\varepsilon(\tau, \omega)|^2 &\le \frac{4\sigma\varepsilon^2}{l^2} \sum_{k\in\mathbb{Z}} \int_0^\infty |\psi_k(t)\, h_k'(t)|^2\, dt \\ &\le \frac{4\sigma\varepsilon^2}{l^2} \|h'\|^2 \sum_{k\in\mathbb{Z}} \int_0^\infty |\psi_k(t)|^2\, dt\,.\end{aligned}$$

Of course, $\psi_k(t)$ is bounded, and it is zero if one of $|\varepsilon k|$ or εt exceeds some threshold depending on the support of ϕ. For large values of $|\omega|$ another bound of

$$\psi_k(t, \tau, \omega) = \frac{e^{\imath\tau t + \imath\omega k}}{2\pi\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} e^{\imath x\omega}\, \phi(\varepsilon t, \varepsilon k + x)\, dx$$

is needed. Integrating by parts we get

$$\begin{aligned}|\psi_k(t,\tau,\omega)| &\le \frac{1}{2\pi\varepsilon\omega}\left|\int_{-\varepsilon/2}^{\varepsilon/2} e^{\imath x\omega}\,\phi_x'(\varepsilon t,\varepsilon k+x)\,dx\right| \\ &+ \frac{1}{2\pi\varepsilon\omega}\left|e^{\imath\varepsilon\omega/2}\phi(\varepsilon t,\varepsilon k+\varepsilon/2)-e^{-\imath\varepsilon\omega/2}\phi(\varepsilon t,\varepsilon k-\varepsilon/2)\right| \\ &\le \frac{3}{4\pi|\omega|}\,\|\phi_x'\| + \frac{|\sin(\omega\varepsilon/2)|}{2\pi\varepsilon|\omega|}\,\|\phi\|\,,\end{aligned}$$

thus we have a constant, K_1 depending only on ϕ such that

$$|\psi_k(t,\tau,\omega)|^2 \le K_1\Psi_\varepsilon(\omega)\,, \quad \text{where} \quad \Psi_\varepsilon(\omega) := \min\left\{1,(\varepsilon\omega)^{-2}\right\}$$

and $0<\varepsilon<1$. Comparing the bounds above, we see that

$$\mathsf{E}|\tilde{m}_\varepsilon(\tau,\omega)|^2 \le K_2\|h'\|^2\,\frac{\sigma}{l^2}\,\Psi_\varepsilon(\omega)\,,$$

thus integrating (4.2) with respect to τ,

$$\mathsf{E}\|m_\varepsilon\|_{-1}^2 \le \frac{K_3\sigma}{l^2}\int_{-\infty}^{\infty}\frac{\Psi_\varepsilon(\omega)\,d\omega}{\sqrt{1+\omega^2}}$$

follows immediately, where K_3 is a new constant depending only on ϕ and $\|h'\|$. Integrating over the domain $|\omega|<1/\varepsilon$, the trivial bound $\Psi_\varepsilon(\omega)\le 1$ is sufficient, while $\Psi_\varepsilon(\omega)\le(\varepsilon\omega)^{-2}$ is used in the opposite case to conclude

$$\mathsf{E}|\tilde{m}_\varepsilon(\tau,\omega)|^2 \le K_4\,\frac{\sigma}{l^2}\,(1-\log\varepsilon)\,.$$

In view of (2.2) and its consequences we have $\sigma = o(l^{3/2})$ and $1/\varepsilon = o(l^2)$, thus the right hand side vanishes as $\varepsilon\to 0$. Q.E.D.

From now on we may suppress the dependence of our functionals on ϕ. In practice this simply means that we put $\phi\equiv 1$ and suppose that the support of ψ is contained in a rectangle $(-1,T)\times(-L,L)$, thus we need the estimates of Section 3 for $n<L/\varepsilon$ and $t<T/\varepsilon$ only. Introduce

$$Q_\varepsilon^* := \frac{\varepsilon}{l}\sum_{|k|<L/\varepsilon}\int_0^{T/\varepsilon} Q_k(t,l)\,dt\,, \tag{4.3}$$

where $Q_k(t,l) := (l\nabla_l\bar{\eta}_{k,l})^2 + (\hat{\eta}_{l,k+l}-\hat{\eta}_{l,k})^2 + (\hat{\eta}_{l,k}-\bar{\eta}_{l,k})^2 + (l\nabla_l\bar{V}_{l,k}')^2$, and

$$Z_\varepsilon^* := \frac{\varepsilon}{l}\sum_{|k|<L/\varepsilon}\int_0^{T/\varepsilon}\left(\bar{V}_{l,k}'(\eta(t)) - S'(\bar{\eta}_{l,k}(t))\right)^2 dt\,. \tag{4.4}$$

Moreover, set $C_\varepsilon(\sigma, l) := C_{t,n}(\sigma, l)$ when $t = T/\varepsilon$ and $n = L/\varepsilon$. In the rest of the paper we assume (2.2), thus $C_\varepsilon = O(l/\sigma)$ goes to 0 as $\varepsilon \to 0$. In view of the a priori bounds, $\mathsf{E}Q^*$ and $\mathsf{E}Z^*_\varepsilon$ are of order $C_\varepsilon(\sigma, l)$.

It is a bit surprising that a two-block lemma is needed to treat N_ε.

Lemma 4.2. *The numerical error N_ε is of type Y.*

Proof. Let $\psi_k(t)$ be as in (4.1) with $\phi \equiv 1$, then

$$N_\varepsilon(\psi, h) = \varepsilon \sum_{k\in\mathbb{Z}} \int_0^\infty \big(\widetilde{\nabla}_1 \psi_k - \varepsilon\, \widetilde{\nabla}_{\varepsilon/2} \psi(t\varepsilon, k\varepsilon)\big) J_k(t)\, dt\,.$$

Since $\widetilde{\nabla}_1 = \nabla_1 - (1/2)\Delta_1$, the integrand turns into $(1/2)(\nabla_1\psi_k)\nabla_1 J_k + \varphi_k \nabla_1^* J_k$, where $\varphi_k := \psi_k - \psi(t\varepsilon, k\varepsilon - \varepsilon/2)$ is an integral of ψ'_x. By the Schwarz inequality

$$\varphi_k(t) = \frac{1}{2\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} (\varepsilon - 2x) \psi'_x(t\varepsilon, k\varepsilon + x)\, dx = O(\sqrt{\varepsilon}) \| 1_{\varepsilon,k}\, \psi'_x(t\varepsilon, \cdot)\|_2\,,$$

where $1_{\varepsilon,k}(x)$ is the indicator of the interval $(k\varepsilon - \varepsilon/2, k\varepsilon + \varepsilon/2)$; the L^2 norm refers to space. Similarly,

$$\nabla_1 \psi_k(t) = \frac{1}{\varepsilon} \int_{-\varepsilon}^{\varepsilon} (\varepsilon - |x|) \psi'_x(t\varepsilon, k\varepsilon + x + \varepsilon/2)\, dx\,,$$

thus $\nabla_1\psi_k$ satisfies the same bound that φ_k does.

On the other hand, $\nabla_1^* J_k = J'(\gamma_k) \nabla_l^* \bar{\eta}_{l,k+l-1}$ with some intermediate value γ_k. Since $J' = h' S''$ is bounded, separating φ_k and $\nabla_1^* J_k$ by means of the Schwarz inequality, and doing the same with $\nabla_1 \psi_k$ and $\nabla_1 J_k$, we obtain that $N_\varepsilon^2 = O(\varepsilon)\, \|\psi'_x\|_2^2\, Q^*_\varepsilon$, that is N_ε is of type Y, and $\sqrt{\varepsilon l/\sigma}$ is its order. Q.E.D.

The next step is the only one where LSI is really needed.

Lemma 4.3. *The asymmetric functional, $X_{a,\varepsilon}$ reads as $X_{a,\varepsilon} = Y_{a,\varepsilon} + Z_{a,\varepsilon} + Q_{a,\varepsilon}$, where $Q_{a,\varepsilon}$ and $Z_{a,\varepsilon}$ are of type Z with a vanishing bound, $Y_{a,\varepsilon}$ is of type Y.*

Proof. Using earlier notation we have

$$X_{a,\varepsilon}(\psi, h) = \frac{\varepsilon}{2} \sum_{k\in\mathbb{Z}} \int_0^\infty (\psi_k + \psi_{k+1})(\nabla_1 J_k - h'_k\, \nabla_l \bar{V}'_{l,k})\, dt\,,$$

and $\nabla_1 J_k = h'_k\, S''_k\, \nabla_l \bar{\eta}_{l,k} + \frac{1}{2}\, J''(\gamma_k)(\nabla_l \bar{\eta}_{l,k})^2$ with some intermediate value γ_k. Moreover, $S''_k\, \nabla_l\, \bar{\eta}_{l,k} = \nabla_l S'(\bar{\eta}_{l,k}) + S'''(\tilde{\eta}''_k)(\hat{\eta}_{l,k} - \tilde{\eta}'_k) \nabla_l \bar{\eta}_{l,k}$, where $\tilde{\eta}'_k$ is a convex combination of $\bar{\eta}_{l,k+l}$ and $\bar{\eta}_{l,k}$, i.e. $|\hat{\eta}_{l,k} - \tilde{\eta}'_k| \leq$

$|\hat{\eta}_{l,k} - \bar{\eta}_{l,k}| + |\hat{\eta}_{l,k} - \bar{\eta}_{l,k-l}|$. Summarizing the calculations above, we get $X_{a,\varepsilon} = X^*_{a,\varepsilon} + Q_{a,\varepsilon}$, where

$$X^*_{a,\varepsilon} := \frac{\varepsilon}{2} \sum_{k\in\mathbb{Z}} \int_0^\infty (\psi_k + \psi_{k+1})\, h'_k\, \nabla_l (S'(\bar{\eta}_{l,k}) - \bar{V}'_{l,k})\, dt\,,$$

while the remainder, $Q_{a,\varepsilon}$ is a bilinear form of differences of block averages of size at most $2l+1$. Since $J'' = h''S'' + h'S'''$, and S''' is bounded in view of Lemma 5.1 in [17], the coefficients of $Q_{a,\varepsilon}$ are all uniformly bounded, consequently $Q_{a,\varepsilon} = O(\|\psi\|)\, Q^*_\varepsilon$. This means that $Q_{a,\varepsilon}$ is of type Z with a vanishing order of $C_\varepsilon(\sigma, l) = O(l/\sigma)$.

On the other hand, from $\nabla^*_l(\xi_k \xi'_k) = (\nabla^*_l \xi_k)\xi'_k + \xi_{k-l}\nabla^*_l \xi'_k$ we get $X^*_{a,\varepsilon} = Y_{a,\varepsilon} + Z_{a,\varepsilon}$, where

$$Y_{a,\varepsilon} := \frac{\varepsilon}{2} \sum_{k\in\mathbb{Z}} \int_0^\infty (\nabla^*_l \psi_k + \nabla^*_l \psi_{k+1})\, h'_k\, (S'(\bar{\eta}_{l,k}) - \bar{V}'_{l,k})\, dt\,,$$

$$Z_{a,\varepsilon} := \frac{\varepsilon}{2} \sum_{k\in\mathbb{Z}} \int_0^\infty (\psi_{k-l} + \psi_{k+1-l})\, (\nabla^*_l h'_k)\, (S'(\bar{\eta}_{l,k}) - \bar{V}'_{l,k})\, dt\,.$$

From the estimate of Lemma 4.2 for $\nabla_1 \psi$ it follows by convexity that $(\nabla_l \psi_k(t))^2 = O(l\varepsilon)\|1_{l\varepsilon,k}\, \psi'_x(t\varepsilon, \cdot)\|_2^2$. Separating the space gradients of ψ from $h'\,(S' - \bar{V}'_l)$ by means of the Schwarz inequality, we obtain that $|Y_{a,\varepsilon}|^2 \le \varepsilon^2 \|h'\|\, \|\psi'_x\|_2^2\, Z^*_\varepsilon$, thus $Y_{a,\varepsilon}$ is of type Y, and $\sqrt{l\varepsilon}\, C_\varepsilon^{1/2}(\sigma, l)$ is the order of its bound. Finally, $\nabla^*_l h'_k = h''(\gamma'_k)\nabla^*_l \hat{\eta}_{l,k}$, whence $|Z_{a,\varepsilon}|^2 \le \|h''\|\, \|\psi\|\, Q^*_\varepsilon\, Z^*_\varepsilon$, that is $Z_{a,\varepsilon}$ is of type Z with a vanishing bound of order $C_\varepsilon(\sigma, l)$. Q.E.D.

The symmetric form decomposes as $X_{s,\varepsilon} = X_{s1,\varepsilon} + X_{s2,\varepsilon}$, where

$$X_{s1,\varepsilon}(\psi, h) := -\sigma\varepsilon \sum_{k\in\mathbb{Z}} \int_0^\infty \nabla_1(\psi_k h'_k)\, (\nabla_l \bar{V}'_{l,k})\, dt\,,$$

$$X_{s2,\varepsilon}(\psi, h) := \frac{\sigma\varepsilon}{l^2} \sum_{k\in\mathbb{Z}} \int_0^\infty \psi_k h''_k\, (d\bar{w}_{l,k-1} - d\bar{w}_{l,k+l-1})^2$$

$$= \frac{2\sigma\varepsilon}{l^3} \sum_{k\in\mathbb{Z}} \int_0^\infty \psi_k h''_k\, dt\,.$$

In case of $X_{s1,\varepsilon}$ we write $\nabla_1(\psi_k h'_k) = \psi_k \nabla_1 h'_k + h'_{k+1} \nabla_1 \psi_k$ to get $X_{s1,\varepsilon} = Y_{s,\varepsilon} - Z_{s,\varepsilon}$, where

$$Z_{s,\varepsilon} := \sigma\varepsilon \sum_{k\in\mathbb{Z}} \int_0^\infty \psi_k \, (\nabla_1 h'_k) \, \nabla_l \bar{V}'_{l,k} \, dt \, ,$$

$$Y_{s,\varepsilon} := -\sigma\varepsilon \sum_{k\in\mathbb{Z}} \int_0^\infty (\nabla_1 \psi_k) h'_{k+1} \nabla_l \bar{V}'_{l,k} \, dt \, .$$

The symmetric part of entropy production is handled by means of

Lemma 4.4. *We have* $X_{s,\varepsilon} = Y_{s,\varepsilon} + X_{s2,\varepsilon} - Z_{s,\varepsilon}$, *where* $Y_{s,\varepsilon}$ *is of type* Y , $X_{s2,\varepsilon}$ *is of type* Z , *and* $\sigma/\varepsilon l^3$ *is the order of its bound.* $Z_{a,\varepsilon}$ *is also of type* Z , *but its bound does never vanish.*

Proof. Since h'' is bounded, $X_{s2,\varepsilon} = \|\psi\| \, O(\sigma/\varepsilon l^3)$ is of type Z . From $2xy \le x^2 + y^2$, and $\nabla_l h'_k = h''(\gamma''_k) \nabla_l \bar{\eta}_{l,k}$ we get

$$|Z_{s,\varepsilon}| \le \frac{\sigma}{2l} \, \|h''\| \, \|\psi\| (Q^*_\varepsilon + Z^*_\varepsilon) \, ,$$

see (4.3) and (4.4) for the definition of Q^*_ε and Z^*_ε . Therefore $Z_{s,\varepsilon}$ is of type Z, and the bound does not vanish. Finally, applying the Schwarz inequality as we did many times before, we have

$$|Y_{s,\varepsilon}|^2 \le \frac{\sigma^2 \varepsilon}{l} \, \|h'\|^2 \, \|\psi'_x\|_2^2 \, Q^*_\varepsilon \, ,$$

consequently $Y_{s,\varepsilon}$ is of type Y as $\varepsilon \sigma^2 l^{-1} \, C_\varepsilon(\sigma, l) = O(\varepsilon\sigma)$. Q.E.D.

§5. Completion of the proofs

Proposition 2.1, is a more or less direct consequence of the results of Section 4.

Proof of Proposition 2.1: Suppose first that $(h_i, J_i) \in \mathcal{H}_\alpha$, where $0 < \alpha < 1$, and h', h'' are bounded, then $\limsup \mathsf{E}\|\phi\rho_\varepsilon\|_2^2 < +\infty$ implies $h, J \in L^p(\mathbb{R}^2_+)$ with some $p > 2$. More precisely, the distributions of $h_i(\hat{\rho}_\varepsilon)$ and $J_i(\hat{\rho}_\varepsilon)$ are tight in the weak topology of $L^p_{\text{loc}}(\mathbb{R}^2_+)$. Similarly, the distributions of the functionals Y_ε and Z_ε are tight with respect to the weak local topology of H_{-1} and the space of measures, respectively. In view of the Skorohod embedding theorem, see Theorem 1.8 in Chapter 3 of [5], we can realize the associated weak convergence of probability measures as a.s. convergence on a suitably constructed probability space. In this setting the theorems of Tartar and Murat apply directly, so we have Tartar factorization for entropy pairs from $\mathcal{H}_\alpha$. The final statement follows by a direct approximation procedure. Q.E.D.

Tartar's factorization property is the input of

Lemma 5.1. *Let* $h_1(\rho) := \rho$, $J_1(\rho) := S'(\rho)$, $h_2(\rho) := S'(\rho)$ *and define* J_2 *by* $J_2(0) = 0$ *and* $J_2'(\rho) := S''^2(\rho)$. *If this couple of entropy pairs satisfies* (1.10), *then* $\theta_{t,x}$ *is almost everywhere a Dirac measure.*

Proof. The trivial case of a quadratic V can be excluded, thus there is no such interval where S'' is constant because S is analytic. Rearranging (1.10) we get

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q(u,v)\,\theta_{t,x}(du)\,\theta_{t,x}(dv) = 0 \quad \text{a.s. on } \mathbb{R}_+^2\,,$$

where $Q(u,v) := (u-v)\big(J_2(u)-J_2(v)\big) - \big(S'(u)-S'(v)\big)^2$. Since

$$J_2(u) - J_2(v) = (u-v)\int_0^1 S''^2(tu+(1-t)v)\,dt\,,$$

$$S'(u) - S'(v) = (u-v)\int_0^1 S''(tu+(1-t)v)\,dt\,,$$

$Q(u,v) > 0$ follows by the Schwarz inequality if $u \neq v$, which proves the Dirac property of θ. Q.E.D.

As a consequence, we have (1.2) with probability one with respect to any weak limit distribution of $\hat{\rho}_\varepsilon$. In view of Lemma 3.5 the same statement holds also true for the usual averages $\bar{\rho}_\varepsilon$ defined by $\bar{\rho}_\varepsilon(t,x) := \bar{\eta}_k(t/\varepsilon)$ if $|x-k\varepsilon| < \varepsilon/2$, and even $l = o(1/\varepsilon)$ is allowed; the lower bound $l \geq (\sigma/\varepsilon)^{1/3}$ is the relevant one.

To prove Proposition 2.2, we have to show that the contribution of terms $(\nabla_1 h(\hat{\eta}_{l,k}))\nabla_l \bar{V}'_{l,k}$ is not negative if h is convex. Despite of Lemma 3.4, this is not quite obvious. Fortunately, the Lax-Kružkov inequality does not require uniform bounds as compensated compactness does, weak limiting arguments are sufficient. Nevertheless, convexity of V seems to be essential at this point.

Proof of Proposition 2.2: Since V is convex by assumption, following [21] we can exploit the attractiveness of the process, see also [15] for some technical details. Let ζ denote the equilibrium process with initial distribution λ_z such that $F'(z) = c$. The original process, η is coupled to ζ simply by identifying the Wiener processes in their stochastic equations (1.1); the initial distribution is $\mu_{\varepsilon,0} \times \lambda_z$. It is remarkable that this coupled process admits a comparison principle: the set

$$\{(\eta,\zeta) : (\eta_{k+1}-\eta_k)(\zeta_{k+1}-\zeta_k) \geq 0\ \forall k \in \mathbb{Z}\}$$

is preserved by time. Introduce

$$\begin{aligned}
W_\varepsilon(\eta,\zeta,\psi) &:= \sum_{k\in\mathbb{Z}}\int_0^\infty \psi_t'(t,\varepsilon k)|\eta_k(t/\varepsilon)-\zeta_k(t/\varepsilon)|\,dx\,dt\,,\\
W_\varepsilon^*(\eta,\zeta,\psi) &:= \sum_{k\in\mathbb{Z}}\int_0^\infty \psi_x'(t,\varepsilon k)|V'(\eta_k(t/\varepsilon))-V'(\zeta_k(t/\varepsilon))|\,dx\,dt\,,\\
H_\varepsilon(\eta,\psi) &:= \int_0^\infty\int_{-\infty}^\infty \psi_t'(t,x)|\hat\rho_\varepsilon(t,x)-c|\,dx\,dt\,,\\
H_\varepsilon^*(\eta,\psi) &:= \int_0^\infty\int_{-\infty}^\infty \psi_t'(t,x)|S'(\hat\rho_\varepsilon(t,x))-S'(c)|\,dx\,dt\,;
\end{aligned}$$

we have to show that for all $c\in\mathbb{R}$ and $0\le\psi\in C_c^1(\mathbb{R}_+^2)$ we have

$$\lim_{\varepsilon\to0}\big(W_\varepsilon(\eta,\zeta,\psi)-H_\varepsilon(\eta,\psi)\big)=0\,,\tag{5.1}$$

$$\lim_{\varepsilon\to0}\big(W_\varepsilon^*(\eta,\zeta,\psi)-H_\varepsilon^*(\eta,\psi)\big)=0\,,\tag{5.2}$$

$$\liminf_{\varepsilon\to0}\big(W_\varepsilon(\eta,\zeta,\psi)+W_\varepsilon^*(\eta,\zeta,\psi)\big)\ge0\tag{5.3}$$

in the sense of stochastic convergence, see Section 3 of [21]; the crucial point is (5.3). To prove it, observe first that $\eta_k-\zeta_k$ is differentiable, and $V_{k-1}'-V_{k+1}'=2V_{k-1}'-2V_k'-V_{k-1}'-V_{k+1}'+2V_k'$, thus

$$\begin{aligned}
\partial_t|\eta_k-\zeta_k| &= \operatorname{sign}(\eta_k-\zeta_k)\big(V'(\eta_{k-1})-V'(\zeta_{k-1})-V'(\eta_k)+V'(\zeta_k)\big)\\
&+(\sigma-1/2)\operatorname{sign}(\eta_k-\zeta_k)\big(V'(\eta_{k-1})-V'(\zeta_{k-1})-V'(\eta_k)+V'(\zeta_k)\big)\\
&+(\sigma-1/2)\operatorname{sign}(\eta_k-\zeta_k)\big(V'(\eta_{k+1})-V'(\zeta_{k+1})-V'(\eta_k)+V'(\zeta_k)\big)\,.
\end{aligned}$$

Let $\chi_k(\eta,\zeta):=\operatorname{sign}(\eta_k-\zeta_k)\operatorname{sign}(\eta_{k+1}-\zeta_{k+1})$, by an elementary computation

$$\begin{aligned}
\partial_t|\eta_k-\zeta_k| &\le \chi_{k-1}|V'(\eta_{k-1})-V'(\zeta_{k-1})|-\chi_k|V'(\eta_k)-V'(\zeta_k)|\\
&+(\sigma-1/2)\,\chi_{k-1}\big(|V'(\eta_{k-1})-V'(\zeta_{k-1})|-|V'(\eta_k)-V'(\zeta_k)|\big)\\
&-(\sigma-1/2)\,\chi_k\big(|V'(\eta_k)-V'(\zeta_k)|-|V'(\eta_{k+1})-V'(\zeta_{k+1})|\big)\\
&-(\sigma-1/2)(2-\chi_{k-1}-\chi_k)|V'(\eta_k)-V'(\zeta_k)|\,.
\end{aligned}$$

Hence by rearranging the sums we get

$$W_\varepsilon+W_\varepsilon^*\ge R_\varepsilon(\eta,\zeta,\psi)+\varepsilon\sum_{k\in\mathbb{Z}}\int_0^\infty W_k(t/\varepsilon)\,dt\,,$$

where R_ε is a numerical error term, t is the macroscopic time, and

$$\begin{aligned}W_k(t) : = {} & \psi'_x(\varepsilon t, \varepsilon k)(\chi_k - 1)|V'(\eta_k) - V'(\zeta_k)| \\ & + (\sigma - 1/2)\psi'_x(t, \varepsilon k)\, \chi_k \, \nabla_1 |V'(\eta_k) - V'(\zeta_k)| \\ & + (1/\varepsilon)(\sigma - 1/2)\psi(\varepsilon t, \varepsilon k)(1 - \chi_k)|V'(\eta_k) - V'(\zeta_k)| \,.\end{aligned}$$

Since $\psi \in C^2(\mathbb{R}^2)$ may be assumed, R_ε goes to zero as $\varepsilon \to 0$, and $\chi_k \nabla_1$ on the second lime above can be replaced with $(\chi_k - 1)\nabla_1$, we see that the last nonnegative terms dominate the rest. Indeed, $\varepsilon|\psi'_x| = o(\psi)$ is certainly true if $\psi > 0$ vanishes in a suitable way as $|x| \to \infty$, whence the general case follows by a direct approximation procedure.

The proofs of (5.1) and (5.2) also follow [21]; but they turn out to be much simpler in our case. The formal generator of the coupled process reads as $\mathfrak{L}_{\eta,\zeta} := \mathfrak{L}_{0,\eta} + \mathfrak{L}_{0,\zeta} + \sigma\, \mathfrak{G}_{\eta,\zeta}$, where $\mathfrak{L}_{0,\eta}$ and $\mathfrak{L}_{0,\zeta}$ are identical copies of $\mathfrak{L}_0$ acting on the η and ζ components, respectively, while $\mathfrak{G}_{\eta,\zeta}$ is the generator of the coupled process (η, ζ) defined by

$$d\eta_k = \Delta_1 V'(\eta_k)\, dt + \sqrt{2}\nabla_1^* \, dw_k \,, \quad d\zeta_k = \Delta_1 V'(\zeta_k)\, dt + \sqrt{2}\nabla_1^* \, dw_k$$

with identical Wiener processes for both systems. In view of the Kolmogorov equation, for smooth cylinder functions

$$\mathsf{E}\phi(\eta(t), \zeta(t)) = \mathsf{E}\phi(\eta(0), \zeta(0)) + \mathsf{E} \int_0^t \mathfrak{L}_{\eta,\zeta}\phi(\eta(s), \zeta(s))\, ds \,.$$

Let $\bar{\nu}_\varepsilon$ denote the time average of the joint distribution of η and ζ from $t = 0$ to $t = 1/\varepsilon$. In view of the L^2 moment condition coming from Lemma 3.2, this family is tight, thus dividing the Kolmogorov equation by σ/ε, we see that its weak limit points are all stationary measures for the coupled process generated by $\mathfrak{G}_{\eta,\zeta}$. Performing a simultaneous averaging also in space, we obtain translation invariant limit distributions $\bar{\nu}$ that are stationary with respect to $\mathfrak{G}_{\eta,\zeta}$, and also satisfy the moment conditions $\bar{\nu}(\eta_k^2 + \zeta_k^2) = K < +\infty$. These statements follow immediately also from Theorem 1 of [8] without any averaging in space. The evaluation of W and W^* should be based on such a joint distribution $\bar{\nu}$, see [13] and [22].

To prove that $\chi_k = 1$ $\bar{\nu}$-a.s. for all $k \in \mathbb{Z}$, consider now the coupled process defined by $\mathfrak{G}_{\eta,\zeta}$, with $\bar{\nu}$ as its initial distribution. By elementary calculation we get

$$\begin{aligned}\partial_t|\eta_k - \zeta_k| = {} & -\nabla_1 \mathrm{sign}(\eta_k - \zeta_k) \cdot \nabla_1\big(V'(\eta_k) - V'(\zeta_k)\big) \\ & + \mathrm{sign}(\eta_{k+1} - \zeta_{k+1})\, \nabla_1\big(V'(\eta_k) - V'(\zeta_k)\big) \\ & - \mathrm{sign}(\eta_k - \zeta_k)\, \nabla_1\big(V'(\eta_{k-1}) - V'(\zeta_{k-1})\big) \,,\end{aligned}$$

where both sides are of mean zero with respect to $\bar{\nu}$ because of its stationarity. Summing for $k \in (-n, n)$ we see that the last two terms cancel each other, only two of them survives at the boundary. Therefore the translation invariance of $\bar{\nu}$ implies

$$\int \nabla_1 \mathrm{sign}(\eta_k - \zeta_k) \cdot \nabla_1 \big(V'(\eta_k) - V'(\zeta_k)\big)\, d\bar{\nu} = 0$$

for all $k \in \mathbb{Z}$, that is $\mathrm{sign}(\eta_k - \zeta_k)$ is a constant $\bar{\nu}$-a.s. This means that $\bar{\nu}[\chi_k = 1] = 1$ for all $k \in \mathbb{Z}$, thus we can get rid of the absolute values under the sums in the expressions of W and W^*. First we replace η and ζ in W_ε and W_ε^* with their large microscopic block averages $\bar{\eta}_r$ and $\bar{\zeta}_r$. Letting $\varepsilon \to 0$ first, and $r \to +\infty$ at the second step, we get c as the limit of $\bar{\zeta}_r$. Finally, Lemma 3.6 allows us to replace $\bar{\eta}_r$ with $\bar{\eta}_l$, where $l = l(\varepsilon)$ is the intermediate block size of (2.2). The replacement of $V'(\eta_k)$ with $S'(\bar{\eta}_{l,k})$ is the same, thus we can pass to (1.11) along subsequences. Q.E.D.

Proof of Theorem 1.1: In view of Skorohod's embedding, Lemma 5.1 and the a priori bounds, the empirical processes, $\hat{\rho}_\varepsilon$ and $\bar{\rho}_\varepsilon$ converge almost surely, and also in $L^1_{\mathrm{loc}}(\mathbb{R}^2_+)$ to the same $\rho \in L^2(\mathbb{R}^2_+)$ along subsequences. At the same time, $V'(\rho_\varepsilon)$ has the same weak limits as $S'(\bar{\rho}_\varepsilon)$ does, thus we have convergence to the set of weak solutions.

The uniqueness part is now a direct consequence of Proposition 2.2 and weak uniqueness of entropy solutions, see Kružkov's result, Theorem 2.3.5 in [25]. Although the proof there is written for bounded solutions only, the essential condition is bounded propagation, that is $\|S''\| < +\infty$. By means of the local L^2 bound we have, the argument extends to our case without any essential change. On the other hand, we have already derived from Proposition 2.1 and Lemma 5.1 that the measure solutions involved in (1.11) are all weak solutions, thus we have (1.4), too. Therefore any limit distribution of the empirical process $\hat{\rho}_\varepsilon$ is concentrated on the unique entropy solution specified by its initial value. In this way we have shown that if $\varepsilon \to 0$ then

$$\hat{R}_\varepsilon(\psi) := \int_0^\infty \int_{-\infty}^\infty \psi(t,x)\hat{\rho}_\varepsilon(t,x)\, dx\, dt$$

converges in probability for each $\psi \in C_c(\mathbb{R}^2)$ to $R(\psi)$ defined in Theorem 1.1. However, $R_\varepsilon(\psi)$ has the same limit. Q.E.D.

Concluding remarks: We are trying to present a brief and heuristic description of situations of hyperbolic scaling in which the method proposed here might apply, several principal open problems are also mentioned. We consider a microscopic Markov evolution generated by

$\mathfrak{L} = \mathfrak{L}_0 + \sigma(\varepsilon)\mathfrak{G}$ such that both $\mathfrak{L}_0$ and $\mathfrak{G}$ are Markov generators, and the conservative observables and the associated (equilibrium) Gibbs states of $\mathfrak{G}$ are all conserved also by $\mathfrak{L}_0$. The main component, $\mathfrak{L}_0$ is asymmetric (but not necessarily antisymmetric), while $\mathfrak{G}$ is symmetric with respect to the equilibrium states. The scaling parameter $\varepsilon > 0$ denotes the macroscopic unit of distance in space, $\sigma(\varepsilon) > 0$ is interpreted as the coefficient of microscopic viscosity, $\sigma(\varepsilon) \to +\infty$ and $\varepsilon\sigma(\varepsilon) \to 0$ as $\varepsilon \to 0$.

First of all we are assuming that $\mathfrak{L}_0$ admits Euler scaling with a resulting hyperbolic system of macroscopic conservation laws, that is we speed up time by a factor of $1/\varepsilon$, see e.g. [27] for a class of such models. In the absence of the symmetric stabilization $\sigma\mathfrak{G}$, these equations can be derived in a smooth regime only. In general, there is a good reason to expect that the effect of $(1/\varepsilon)\sigma(\varepsilon)\mathfrak{G}$ diminishes as $\varepsilon \to 0$ because $\varepsilon^{-2}\mathfrak{G}$ is the proper scaling of the symmetric $\mathfrak{G}$. In other words,

$$\varepsilon^{-1}\mathfrak{L} = \varepsilon^{-1}\mathfrak{L}_0 + (\varepsilon\sigma(\varepsilon))\,\varepsilon^{-2}\mathfrak{G}$$

resembles the scheme of small viscosity limit as $\varepsilon \to 0$; $\varepsilon\sigma(\varepsilon)$ is the coefficient of macroscopic viscosity.

Independently of the number of conservation laws, once we have LSI for $\mathfrak{G}$, there is a good chance to derive Tartar's factorization property for the limiting Young measures; $\varepsilon\sigma^2(\varepsilon) \to +\infty$ as $\varepsilon \to 0$ seems to be a general condition at this point, see [9,10,11]. It is not clear this time if this condition could be relaxed, or not. In the case of a single conservation law Tartar factorization is usually sufficient for the identification of measure solutions as weak solutions by using an argument like that of Lemma 5.1. The problem of two conservation laws is more delicate, a very nice model is discussed in [11]. In other cases we have to do something more for proving that measure solutions are weak solutions. Although there is a general theory of hyperbolic and genuinely nonlinear systems of two conservation laws in one space dimension initiated by DiPerna, additional difficulties emerge when we are working on stochastic models. Indeed, this theory requires at the very beginning that the limiting Young measure is compactly supported. Moreover, most physically motivated models have some singularities in the phase space of the macroscopic equations, general methods fail at such points. In PDE theory these difficulties are ruled out by restricting the initial values to singularity-free compact invariant regions, if any. However, it is not easy to establish the existence of such invariant regions in the case of microscopic systems. Coupling is an effective tool, but attractive evolutions do not allow two conservation laws.

Anyway, compensated compactness yields convergence of the empirical process to a set of weak solutions in several cases, so the next question is the uniqueness of the limit. If we have a single conservation law, the Lax–Kružkov entropy condition is sufficient for uniqueness, and coupling based on attractiveness is not the only way of proving it. For example, if $\mathfrak{G}$ is acting on the conservative observables like a discrete Laplacian, that is a *linear elliptic operator*, then the derivation of the Lax inequality (1.3) is only a question of direct computations. This is the case when $\mathfrak{L}_0$ describes interacting exclusions because then $\mathfrak{G}$ can be chosen as the generator of stirring, see [10,11]. There is a conflict of $\mathfrak{L}_0$ and $\mathfrak{G}$ if the cardinality of the individual phase space is bigger than three. For instance, the easy way mentioned above is only available for the trivial, linear Ginzburg–Landau model. It is not clear if attractiveness of $\mathfrak{G}$ were sufficient for the Lax–Kružkov inequality. Uniqueness for two conservation laws is certainly very hard, even in the simplest cases Oleinik type entropy conditions were needed for the Riemann invariants. The derivation of such one-sided uniform bounds on space gradients is really problematic for stochastic models.

Acknowledgement: I am indebted to Claudio Landim for useful discussions on LSI and uniform large deviation estimates.

References

[1] Benassi, A. and Fouque, J-P. (1987): Hydrodynamic limit for the asymmetric simple exclusion process. Ann. Probab. **15**:546–560.

[2] Boldrighini, C. and Dobrushin, R. L. and Sukhov, Yu. M. (1983):One-dimensional hard rod caricature of hydrodynamics. J. Stgatist. Phys. **31**:577-616.

[3] DiPerna, R. (1985): Measure-valued solutions to conservation laws. Arch. Rational Mech. Anal. **88**:223–270.

[4] Dittrich, P. (1992): Long-time behaviour of the weakly asymmetric exclusion process and the Burgers equation without viscosity. Math. Nachr. **155**:279–287.

[5] Ethier, S.N. and Kurtz, T.G.: Markov Processes, Characterization and convergence. Wiley, New York 1986.

[6] Fritz, J. (1987): On the hydrodynamic limit of a one-dimensional Ginzburg–Landau lattice model. J. Statist. Phys. **47**:551–572.

[7] Fritz, J. and Maes, Ch. (1988): Derivation of a hydrodynamic equation for Ginzburg–Landau models in an external field. Journ. Statist. Phys. **53**:1179–1206.

[8] Fritz, J. (1990): On the diffusive nature of entropy flow in infinite systems: Remarks to a paper by Guo, Papanicolau and Varadhan. Comm. Math. Phys. **133**:331–352.

[9] Fritz, J.: An Introduction to the Theory of Hydrodynamic Limits. Lectures in Mathematical Sciences **18.** The University of Tokyo, ISSN 0919–8180, Tokyo 2001.

[10] Fritz, J. and Nagy, Katalin (in preparation): On uniqueness of the Euler limit of one-component lattice gas models.

[11] Fritz, J. and Tóth, B. (2003): Derivation of the Leroux system as the hydrodynamic limit of a two-component lattice gas. Preprint: www.math.bme.hu/~jofri

[12] Gärtner, J. (1988): Convergence towards Burger's equation and propagation of chaos for weakly asymmetric exclusion processes. Stoch. Process. Appl. **27**:233–260.

[13] Guo, M.Z. and Papanicolaou, G.C. and Varadhan, S.R.S. (1988): Nonlinear diffusion limit for a system with nearest neighbor interactions. Comm. Math. Phys. **118**:31–59.

[14] Hörmander, L.: Lectures on Nonlinear Hyperbolic Differential Equation. Mathèmatiques & Applicationes 26, Springer, Berlin 1997.

[15] Kipnis, C. and Landim, C.: Scaling Limit of Interacting Particle Systems. Springer, Berlin 1999.

[16] Kosygina, Elena (2001): The behavior of the specific entropy in the hydrodynamic scaling limit. Ann. Probab. **29**:1086–1110.

[17] Landim, C. and Panizo, G. and Yau, H.T. (2002): Spectral gap and logarithmic Sobolev inequality for unbounded conservative spin systems. Ann. Inst. H. Poincare **38**: 739–777.

[18] Lax, P.: Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Wawes. SIAM, CBMS–NSF 11, 1973.

[19] Murat, F. (1978): Compacitè par compensation. Ann. Sci. Norm. Sup. Pisa **5**:489–507.

[20] Quastel, J. and Yau, H.T. (1998): Lattice gases, large deviations, and the incompressible Navier-Stokes equations. Ann. of Math. **148**:51–108.

[21] Rezakhanlou, F. (1991): Hydrodynamic limit for attractive particle systems on $\mathbf{Z}^d$. Comm. Math. Phys. **140**:417–448.

[22] Rost, H. (1981): Nonequilibrium behaviour of a many particle process: density profile and local equilibria. Z. Wahrsch. Verw. Gebiete **58**:41–53.

[23] Seppäläinen, T. (1999): Existence of hydrodynamics for the totally asymmetric simple K-exclusion process. Ann. Probab. **27**:361–415.

[24] Seppäläinen, T. (2001): Perturbalion of the equilibrium for a totally asymmetric stick process in one dimension. Ann. Probab. **29**:176–204.

[25] Serre, D.: Systems of Conservation Laws 1–2. Cambridge University Press, Cambridge 1999.

[26] Tartar, L.: Compensated compactness and applications to partial differential equations. Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. IV:136–212, Pitman, Boston 1979.

[27] Tóth, B. and Valkó, B. (2003): Onsager relations and Eulerian hydrodynamic limit for systems with several conservation laws. J. Statist. Phys. **112**: 497–521.

[28] Tóth, B. and Valkó, B. (preprint, 2003): Perturbation of singular equilibria for systems with two conservation laws – hydrodynamic limit.

[29] Varadhan, S.R.S. (1991): Scaling limits for interacting diffusions. Comm. Math. Phys. **135**:313–353.

[30] Varadhan, S.R.S.: Nonlinear diffusion limit for a system with nearest neighbor interactions II. Asymptotic problems in probability theory, (Sanda/Kyoto, 1990), 75–128, Longman, Harlow 1993.

[31] Varadhan, S.R.S. (2004): Large deviations for the asymmetric simple exclusion process. Advanced Studies in Pure Mathematics. **39**, "Stochastic Analysis on Large Scale Interacting Systems" : 1–27.

[32] Yau, H.T. (1991): Relative entropy and hydrodynamics of Ginzburg–Landau models. Lett. Math. Phys. **22**:63–80.

Institute of Mathematics
Budapest University of Technology and Economics
H-1111 Budapest
Egry József u. 1. H-42
Hungary
E-mail address: `jofri@math.bme.hu`

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 173–211

Large Deviations for $\nabla\varphi$ Interface Model and Derivation of Free Boundary Problems

Tadahisa Funaki and Hironobu Sakagawa

Abstract.

We consider the $\nabla\varphi$ interface model with weak self potential (one-body potential) under general Dirichlet boundary conditions on a large bounded domain and establish the large deviation principle for the macroscopically scaled interface height variables. As its application the law of large numbers is proved and the limit profile is characterized by a variational problem which was studied by Alt-Caffarelli [1], Alt-Caffarelli-Friedman [2] and others. The minimizers generate free boundaries inside the domain. We also discuss the $\nabla\varphi$ interface model with δ-pinning potential in one dimension.

§1. Introduction

Interfaces and variational problems.

It is one of the quite general and fundamental principles in physics that physically realizable phenomena may be characterized by variational problems. Such principle is expected to hold in the problem related to the phase coexistence and separation as well. Indeed, under the situation that two distinct pure phases like crystal/vapor coexist in space, hypersurfaces called interfaces are formed and separate these distinct phases at macroscopic level. The shape of the interface in equilibrium is assumed to minimize the anisotropic total surface energy. The corresponding solutions may be obtained by the so-called Wulff construction (see [5], [8] and references therein). The underlying variational problems change depending on the physical situations of interest.

In statistical mechanics, to derive the shape of the macroscopic interface, one need to determine its total surface energy based on statistical

---

Received January 24, 2003.
Revised May 28, 2003.
The first author is partially supported by JSPS, Grant-in-Aids for Scientific Research 14340029 and for Exploratory Research 13874015.

ensembles at microscopic level, which are formulated as Gibbs measures. This procedure can be accomplished by analyzing a proper scaling limit in the ensembles, which connects microscopic and macroscopic levels.

$\nabla\varphi$ interface model.

The basic microscopic model we study in this article is the $\nabla\varphi$ interface model, which is a continuous analogue of SOS type model. In this model, the interface is already considered as a microscopic object and described by height variables $\phi = \{\phi(x)\}$, the vertical distance of the surface measured from the points x on a fixed reference hyperplane located in the space (see [18], [19] for example). Assuming interfaces are formed in $d+1$ dimensional space, the variables ϕ are defined on a large bounded domain D_N in the d-dimensional square lattice $\mathbb{Z}^d$. Here D_N corresponds to the reference hyperplane which is discretized and $N \in \mathbb{Z}_+$ is the scaling parameter representing the ratio of the macroscopically typical length to the microscopic one.

Given strictly convex symmetric nearest neighbor interactions $V : \mathbb{R} \to \mathbb{R}$ and boundary conditions $\psi = \{\psi(x) \in \mathbb{R}; x \in \partial^+ D_N\}$, an interface energy $H_N^{\psi}(\phi)$ at microscopic level called Hamiltonian is assigned to each interface height variable $\phi = \{\phi(x) \in \mathbb{R}; x \in D_N\}$ on D_N as a sum of $V(\phi(x) - \phi(y))$ taken over all pairs of neighboring sites x and y in the domain $\overline{D_N}$. Here $\overline{D_N} = D_N \cup \partial^+ D_N$ is the closure of D_N, $\partial^+ D_N = \{x \notin D_N; |x-y| = 1 \text{ for some } y \in D_N\}$ is the outer boundary of D_N and $\phi(x) = \psi(x)$ for $x \in \partial^+ D_N$ in the sum; note that $x \notin D_N$ means $x \in \mathbb{Z}^d \setminus D_N$. We shall take $D_N = ND \cap \mathbb{Z}^d$ for a fixed bounded domain D in $\mathbb{R}^d$ having piecewise Lipschitz boundary ∂D, where $ND = \{N\theta \in \mathbb{R}^d; \theta \in D\}$; D is the macroscopic reference hyperplane while D_N is its microscopic correspondence.

Weak self potentials.

We further assume the space is filled by a media changing in the distances from D_N. Such situation can be realized by adding self potentials (one-body potentials) $U : D \times \mathbb{R} \to \mathbb{R}$ to the Hamiltonian which has therefore the following form:

$$(1.1) \quad H_N^{\psi,U}(\phi) = \sum_{x,y \in \overline{D_N}, |x-y|=1} V(\phi(x) - \phi(y)) + \sum_{x \in D_N} U(\frac{x}{N}, \phi(x)).$$

The first sum here is over all pairs of neighboring sites. Then the statistical ensemble for the height variables ϕ is defined by the finite volume

Gibbs measure on D_N

$$(1.2) \qquad \mu_N^{\psi,U}(d\phi) = \frac{1}{Z_N^{\psi,U}} \exp\{-H_N^{\psi,U}(\phi)\} \prod_{x\in D_N} d\phi(x),$$

where $Z_N^{\psi,U}$ is a normalization factor; note that $\mu_N^{\psi,U} \in \mathcal{P}(\mathbb{R}^{D_N})$, the family of all probability measures on $\mathbb{R}^{D_N}$. We shall sometimes regard $\mu_N^{\psi,U} \in \mathcal{P}(\mathbb{R}^{\overline{D_N}})$ by considering $\phi(x) = \psi(x)$ for $x \in \partial^+ D_N$ under $\mu_N^{\psi,U}$. We consider the case that U is represented as $U(\theta, r) = Q(\theta)W(r)$, where the function $Q : D \to [0,\infty)$ is bounded and the basic assumption on $W : \mathbb{R} \to \mathbb{R}$ is that the limits $\alpha = \lim_{r\to+\infty} W(r)$ and $\beta = \lim_{r\to-\infty} W(r)$ exist, and the values of W are always between α and β; see the conditions (Q1), (W1) and (W2) in Section 2. The self potential U is called weak since it is bounded. A typical example of W we have in mind throughout this paper is a function of the form

$$(1.3) \qquad W(r) = \beta 1_{\{r<0\}} + \alpha 1_{\{r\geq 0\}}, \quad r \in \mathbb{R}.$$

This potential describes the situation that the space is filled by two different media above and below the hyperplane D_N. If $\beta < \alpha$, the negative values are more favorable than the positive ones for the interface height variables ϕ under the Gibbs measures. In other words the interface is weakly attracted to the negative side, namely by the media below the hyperplane D_N.

Scaling limit and large deviations.

The aim of the present paper is to study the macroscopic behavior of the microscopic height variables ϕ under the Gibbs measures $\mu_N^{\psi,U}$ as $N \to \infty$. The scaling connecting microscopic and macroscopic levels is introduced by associating the macroscopic height variables $h^N = \{h^N(\theta); \theta \in D\}$ with ϕ as step functions (or their polilinear approximations (2.1)) on D, which satisfy

$$h^N(x/N) = N^{-1}\phi(x), \quad x \in D_N.$$

Note that both x- and ϕ-axis are rescaled by the same factor $1/N$, since the interface is located in the $d+1$ dimensional space. The boundary conditions ψ should be simultaneously scaled to have macroscopic limits $g(\theta), \theta \in \partial D$, see the conditions ($\psi$1), ($\psi$2) in Section 2. We shall prove that the law of large numbers holds for h^N distributed under $\mu_N^{\psi,U}$ as $N \to \infty$ and the limit $h = \{h(\theta); \theta \in D\}$ is characterized as the minimizer of the macroscopic total surface energy

$$(1.4) \qquad \int_D \sigma(\nabla h(\theta))\, d\theta - A \int_D Q(\theta) 1(h(\theta) \leq 0) d\theta$$

in the class of h having boundary condition g if the minimizer is unique, see Corollary 2.1. Here $\sigma = \sigma(u) \in \mathbb{R}$ is the so-called surface tension of the (macroscopic) surface with tilt $u \in \mathbb{R}^d$ (see (2.3) or [18]) and we assume $A = \alpha - \beta \geq 0$. When $A < 0$, the formula (1.4) should be slightly modified.

We shall actually establish the large deviation principle (LDP) for h^N under $\mu_N^{\psi,U}$, see Theorem 2.1. As its application, one can prove the law of large numbers. The variational problem characterizing the limit generates free boundaries inside D. Such variational problem was thoroughly studied by Alt and Caffarelli [1] for non-negative macroscopic boundary data g with $A > 0$ and by Alt, Caffarelli and Friedman [2] for general g especially when σ is quadratic: $\sigma(u) = |u|^2$, and by Weiss [26] for more general σ.

Bibliographical notes.

Our results are related to those obtained by Pfister and Velenik [24]. They considered the two dimensional Ising model at low temperature on a large box with attractive wall set at the bottom line. This line segment corresponds to our hyperplane D_N, although it has an effect of hard wall at the same time, since the interfaces separating $\pm$-phases can not go down beyond the bottom line in their setting. One of the motivations of [24] was to understand the so-called wetting or pinning/depinning transition.

The problem of the wetting transition is recently discussed for the $\nabla\varphi$ interface model as well by several authors. We shortly summarize the known results. The potential

$$U(\theta, r) = U(r) = -b1_{\{|r|\leq a\}}, \quad r \in \mathbb{R} \tag{1.5}$$

with $a, b > 0$ is called of square well type and yields a weak pinning effect to the interface near D_N, i.e. the level $\phi(x) = 0$. The limit as $a \downarrow 0$ keeping $s = 2a(e^b - 1)$ constant is called δ-pinning. Dunlop et al. [16] first proved the localization of the ϕ-field, namely the uniform boundedness in N of the expected height variables $E^{\mu_N^{0,U}}[|\phi(x)|]$ under the Gibbs measures $\mu_N^{0,U}$ with 0-boundary conditions or the existence of infinite volume limit of $\mu_N^{0,U}$ as $N \to \infty$, if the Hamiltonian contains arbitrarily weak pinning potentials U when $d = 2$ for quadratic V. This should be compared with the case without pinning (i.e. $U \equiv 0$) in which the localization occurs only when $d \geq 3$ and also compared with the case of strong pinning (or massive) potentials satisfying $\lim_{|r|\to\infty} U(r) = +\infty$ for which the localization occurs for all dimensions. The result of [16] is extended for general convex potential V by Deuschel and Velenik [15] later. In addition to the localization, the mass generation, namely the exponential decay of the correlations of the ϕ-field is shown by Ioffe and Velenik [20] for $d = 2$ with δ-pinning. Further precise estimates on the asymptotic behaviors of the mass and the degree of localization by means of the variances of the field as the pinning effect becomes smaller were established by Bolthausen and Velenik [9]. The basic assumption in our paper (W2) on the potential $W(r)$ unfortunately excludes the potential U of square well type given in (1.5).

When $U(r) = +\infty$ for $r < 0$, we say that the hard wall is settled at the level $\phi(x) = 0$ or at D_N. The ϕ-field can take only non-negative values. To discuss the wetting transition for the $\nabla\varphi$ interface model, the effects of the hard wall and the pinning near 0-level are introduced at the same time. Fisher [17] proved the existence of the wetting transition, namely the qualitative change in the localization/delocalization of the field depending on which of these two competitive effects dominate the other, when $d = 1$ for the SOS type discrete model. This result is extended by Caputo and Velenik [10] for $d = 2$. The precise path level behavior is discussed by Isozaki and Yoshida [21] when $d = 1$. Bolthausen et al. [7] showed that, contrarily when $d \geq 3$, no transition occurs and the field is always localized, i.e. only the phase of partial wetting appears. Note that the field on a hard wall is delocalized for all dimensions d if there is no pinning effect, i.e. $U \equiv 0$ for $r \geq 0$. The latter property is called entropic repulsion. Bolthausen and Ioffe [8] proved the law of large numbers in the partial wetting phase in 2-dimension (i.e. $d = 2$) under the Gibbs measures with 0-boundary conditions, hard wall, δ-pinning and quadratic V conditioned that the macroscopic total volume of the interfaces is kept constant. They derived the so-called Winterbottom shape in the limit and the variational problem characterizing it. The 1-dimensional case with general V was discussed by De Coninck et al. [11].

Our model only takes a special class of self potentials, in particular satisfying the condition (W2), into account and neglects the effect of the hard wall. Since the field can take negative values and the potential U has no strong singularity like hard wall, the situation becomes mild in a sense. On the other hand, this makes us possible to discuss the corresponding dynamics without making much effort, which will be discussed elsewhere; see also [23] for dynamics with general boundary conditions when $U \equiv 0$.

Organization of the paper.

In Section 2, the model is introduced in more precise way and the main results are stated. The proof of the large deviation principle is reduced to the case of $U \equiv 0$ in Section 3, since the potential U can be treated as a rather simple perturbation. The large deviation principle for general boundary conditions without the self potential U is proved in Sections 4 and 5. The case with 0-boundary conditions without U was discussed by Deuschel et al. [13]. Our main effort is therefore made for the treatment of the general boundary conditions. By a simple shift the problem can be reduced to the 0-boundary case, however with bond-depending interaction potentials. Finally, in Section 6, we prove the large deviation principle for δ-pinning case when $d = 1$ and Gaussian potential.

§2. Model and Results

Model and basic assumptions.

Recall that a bounded domain D in $\mathbb{R}^d$ with piecewise Lipschitz boundary is given and microscopic regions $D_N, \overline{D_N}$ and $\partial^+ D_N$, $N \in \mathbb{Z}_+$ in $\mathbb{Z}^d$ are defined from D. For a configuration $\phi = \{\phi(x); x \in D_N\} \in \mathbb{R}^{D_N}$ of the random interface on D_N and microscopic boundary condition $\psi = \{\psi(x); x \in \partial^+ D_N\} \in \mathbb{R}^{\partial^+ D_N}$, $\phi \vee \psi$ represents that on $\overline{D_N}$ which coincides with ϕ on D_N and ψ on $\partial^+ D_N$. For every $\Lambda \subset \mathbb{Z}^d$, Λ^* denotes the set of all directed bonds $b = \langle x, y \rangle$ in Λ, which are directed from y to x. We write $x_b = x$, $y_b = y$ for $b = \langle x, y \rangle$. For each $b \in (\mathbb{Z}^d)^*$ and $\phi = \{\phi(x); x \in \mathbb{Z}^d\}$, define $\nabla\phi(b) = \phi(x_b) - \phi(y_b)$. We also define $\nabla_j \phi(x) = \phi(x + e_j) - \phi(x)$, $1 \le j \le d$ for $x \in \mathbb{Z}^d$ where $e_j \in \mathbb{Z}^d$ is the j-th unit vector. $\nabla\phi(x) = \{\nabla_j \phi(x)\}_{1 \le j \le d}$ denotes vector field of height differences of ϕ.

The Hamiltonian on D_N with boundary condition ψ is defined by

$$H_N^\psi(\phi) = \frac{1}{2} \sum_{b \in \overline{D_N}^*} V(\nabla(\phi \vee \psi)(b)), \quad \phi \in \mathbb{R}^{D_N}.$$

Note that this coincides with the first term of (1.1). For the interaction potential V, we assume the following conditions:

(V1) $V \in C^2(\mathbb{R})$,
(V2) $V(\eta) = V(-\eta)$ for every $\eta \in \mathbb{R}$,
(V3) there exist $c_-, c_+ > 0$ such that $c_- \le V''(\eta) \le c_+$ for every $\eta \in \mathbb{R}$.

Next, let $U : D \times \mathbb{R} \to \mathbb{R}$ be a self potential which has an effect attracting the interface ϕ to the negative or positive side. We consider the case that U is decomposed as $U(\theta, r) = Q(\theta)W(r)$, where $Q : D \to [0, \infty)$, $W : \mathbb{R} \to \mathbb{R}$ and assume the following conditions:

(Q1) Q is non-negative, bounded and piecewise continuous,
(W1) W is measurable,
(W2) there exist $\alpha, \beta \in \mathbb{R}$ such that $\lim_{r\to+\infty} W(r) = \alpha$, $\lim_{r\to-\infty} W(r) = \beta$ and $\alpha \wedge \beta \le W(r) \le \alpha \vee \beta$ for every $r \in \mathbb{R}$ (in particular, W is bounded).

Then, $H_N^{\psi,U}(\phi) = H_N^{\psi}(\phi) + \sum_{x\in D_N} U(\frac{x}{N}, \phi(x))$ is the Hamiltonian (1.1) on D_N with boundary condition ψ and self potential U. The corresponding finite volume Gibbs measure $\mu_N^{\psi,U}$ on D_N is defined by (1.2). We shall denote $\mu_N^{\psi,0}$ by μ_N^{ψ}. In the Gaussian case i.e. $V(\eta) = \frac{1}{2}\eta^2$ and $U \equiv 0$, we shall denote it by $\mu_N^{\psi,*}$.

For $g \in C^\infty(\mathbb{R}^d)$, define $H_g^1(D) = \{h \in H^1(D); h - g\big|_D \in H_0^1(D)\}$. The function $g\big|_{\partial D}$ will be the macroscopic boundary condition. We assume the following conditions for the corresponding microscopic boundary condition $\psi \in \mathbb{R}^{\partial^+ D_N}$.

(ψ1) $\max\limits_{x\in\partial^+ D_N} |\psi(x)| \le CN$,
(ψ2) $\sum\limits_{x\in\partial^+ D_N} |\psi(x) - Ng(\frac{x}{N})|^{p_0} \le CN^d$ for some $C > 0$ and $p_0 > 2$.

Remark 2.1. *Since ∂D is piecewise Lipschitz and $g\big|_D \in C^\infty(\bar{D})$, by Theorem 8.7 and Theorem 8.9 of* [27]*, there exists a continuous linear trace operator $T_0 : H^1(D) \to H^{\frac{1}{2}}(\partial D)$ such that $T_0 u = u\big|_{\partial D}$ for every $u \in C^\infty(\bar{D})$ and it holds that $H_g^1(D) = \{h \in H^1(D); T_0 h = g\big|_{\partial D}\}$.*

Scaling and polilinear interpolation.

Our scaled random interface $\{h^N(\theta); \theta \in D\}$ is defined by polilinear interpolation of the macroscopically scaled height variables i.e. $h^N(\theta) = \frac{1}{N}\phi(x)$ for $\theta = \frac{x}{N}$, $x \in \overline{D_N}$ and

$$(2.1) \qquad h^N(\theta) = \sum_{\lambda\in\{0,1\}^d} \Big[\prod_{i=1}^d (\lambda_i\{N\theta_i\} + (1-\lambda_i)(1-\{N\theta_i\}))\Big] h^N\Big(\frac{[N\theta]+\lambda}{N}\Big),$$

for general $\theta \in D$, where $[\cdot]$ and $\{\cdot\}$ denote the integral and the fractional parts, respectively, see (1.17) of [13]. We also define the scaled profile $\{\bar{h}^N(\theta); \theta \in D\}$ by step function i.e. $\bar{h}^N(\theta) = \frac{1}{N}\phi([N\theta])$ for $\theta \in D$. Similarly, for each scalar lattice field $\{u(\frac{x}{N}); x \in D_N\}$, we will define $\{u^N(\theta); \theta \in D\}$ by $u^N(\theta) = u(\frac{x}{N})$ for $\theta = \frac{x}{N}$, $x \in D_N$ and by (2.1) for general $\theta \in D$ and $\{\bar{u}^N(\theta); \theta \in D\}$ by $\bar{u}^N(\theta) = u(\frac{[N\theta]}{N})$ for $\theta \in D$. Also, given a continuous function $f(\theta)$ of $\theta \in D$, we will define $\{f^N(\theta); \theta \in D\}$ and $\{\bar{f}^N(\theta); \theta \in D\}$ from scalar lattice field $\{f(\frac{x}{N}); x \in D_N\}$ as above. Using Jensen's inequality and elementary estimates, we can see that for each $p > 1$, there exists a constant $C_0 = C_0(d,p) > 0$ such that

$$C_0 \|\bar{u}^N\|_{\mathbb{L}^p(D)} \leq \|u^N\|_{\mathbb{L}^p(D)} \leq \|\bar{u}^N\|_{\mathbb{L}^p(D)}, \tag{2.2}$$

for every scalar lattice field $\{u(\frac{x}{N}); x \in D_N\}$.

LDP in the case with weak self potentials.

Now we are in the position to state the main result of this paper. The (normalized) surface tension with tilt $u \in \mathbb{R}^d$ is defined by

$$\sigma(u) = -\lim_{N\to\infty} \frac{1}{N^d} \log \frac{Z^{\psi_u}_{\Lambda_N}}{Z^0_{\Lambda_N}}, \tag{2.3}$$

where $Z^{\psi}_{\Lambda_N}$ is a partition function for $\mu^{\psi}_{\Lambda_N} (= \mu^{\psi,0}_{\Lambda_N})$ on $\Lambda_N = [1, N-1]^d \cap \mathbb{Z}^d$ and $\psi_u(x) = u \cdot x$, $x \in \overline{\Lambda_N}$ represents the u-tilted boundary condition (cf. [13], [18]). For $h \in H^1(D)$, define surface free energy (integrated surface tension)

$$\Sigma(h) = \int_D \sigma(\nabla h(\theta)) d\theta.$$

Theorem 2.1. *The family of random surfaces $\{h^N(\theta); \theta \in D\}$ distributed under $\mu^{\psi,U}_N$ satisfies the large deviation principle (LDP) on $\mathbb{L}^2(D)$ with speed N^d and the rate functional $I^U(h)$, that is, for every closed set $\mathcal{C}$ and open set $\mathcal{O}$ of $\mathbb{L}^2(D)$ we have that*

$$\limsup_{N\to\infty} \frac{1}{N^d} \log \mu^{\psi,U}_N (h^N \in \mathcal{C}) \leq -\inf_{h\in\mathcal{C}} I^U(h), \tag{2.4}$$

$$\liminf_{N\to\infty} \frac{1}{N^d} \log \mu^{\psi,U}_N (h^N \in \mathcal{O}) \geq -\inf_{h\in\mathcal{O}} I^U(h). \tag{2.5}$$

The functional $I^U(h)$ is given by

$$I^U(h) = \begin{cases} \Sigma^U(h) - \inf_{H^1_g(D)} \Sigma^U & \text{if } h \in H^1_g(D), \\ +\infty & \text{otherwise,} \end{cases}$$

where $\inf_{H^1_g(D)} \Sigma^U = \inf\{\Sigma^U(h); h \in H^1_g(D)\}$ *and*

$$\Sigma^U(h) = \Sigma(h) + \alpha \int_D Q(\theta) 1(h(\theta) > 0) d\theta + \beta \int_D Q(\theta) 1(h(\theta) < 0) d\theta \\ + (\alpha \wedge \beta) \int_D Q(\theta) 1(h(\theta) = 0) d\theta.$$

Remark 2.2. *By the proof of Theorem 2.1 (see* (3.8) *below), if* U *is given by* $U(\theta, r) = QW(r)$ *for some constant* $Q \geq 0$ *and* $W(r)$ *satisfies the condition* (W2) *with* $(\alpha, \beta) = (0, -A)$ *or* $(-A, 0)$ *for some* $A \geq 0$ *so that* $-A \leq W(r) \leq 0$ *for every* $r \in \mathbb{R}$, *then it holds that*

$$-AQ = -\lim_{N\to\infty} \frac{1}{N^d} \log \frac{Z^{0,U}_{\Lambda_N}}{Z^0_{\Lambda_N}}, \tag{2.6}$$

where the right hand side represents the difference of the free energies of the interface in the case with self potential and in the case without self potential. In this sense, $\Sigma^U(h)$ *above represents macroscopic total surface energy of the profile* h; *see also Remark 3.1 below.*

As a corollary of the upper bound (2.4) in Theorem 2.1, we obtain the following law of large numbers for $\{h^N(\theta); \theta \in D\}$ under $\mu^{\psi,U}_N$.

Corollary 2.1. *If* Σ^U *has a unique minimizer* $\bar{h}$ *in* $H^1_g(D)$, *then the law of large numbers holds under* $\mu^{\psi,U}_N$, *namely,*

$$\lim_{N\to\infty} \mu^{\psi,U}_N(\|h^N - \bar{h}\|_{\mathbb{L}^2(D)} > \delta) = 0,$$

for every $\delta > 0$.

Remark 2.3. **(Free boundary problems)** *If* $\sigma = \sigma(u)$ *is smooth enough (i.e.* $\sigma \in C^{2,\gamma}(\mathbb{R}^d), \gamma > 0$) *and if the free boundary* $\partial\{h > 0\}$ *of the minimizer* h *of* Σ^U *is locally* C^2, *then* h *satisfies the Euler equation* $\operatorname{div}\{\nabla\sigma(\nabla h)\} = 0$ *in* $D \setminus \partial\{h > 0\}$ *and the condition* $\Psi(\nabla h^+) - \Psi(\nabla h^-) = AQ$ *on the free boundary* $D \cap \partial\{h > 0\}$, *where* $\Psi(u) = u \cdot \nabla\sigma(u) - \sigma(u)$ *and* $A = (\alpha \vee \beta) - (\alpha \wedge \beta)$. *The Lipschitz continuity of the minimizer* h *and the regularity of its free boundary were studied by* [1], [2], [26] *and others. In our case, for the regularity of the surface tension,* $\sigma \in C^{1,1}(\mathbb{R}^d)$ *is only known in general, see* [18].

LDP for δ-pinning in one dimension.

The Gibbs measure with δ-pinning corresponds to the weak limit of the square-well pinning measure $\mu^{\psi,W}_N$ with $W(r) = -b1_{\{|r| \leq a\}}$ as

$a \downarrow 0, b \to \infty$ by keeping $2a(e^b - 1) = e^J$ for $J \in \mathbb{R}$ and has the following representation:

$$\mu_N^{\psi,J}(d\phi) = \frac{1}{Z_N^{\psi,J}} \exp\{-H_N^\psi(\phi)\} \prod_{x \in D_N} (e^J \delta_0(d\phi(x)) + d\phi(x)).$$

We regard $\mu_N^{\psi,J} \in \mathcal{P}(\mathbb{R}^{\overline{D_N}})$ by considering $\phi(x) = \psi(x)$ for $x \in \partial^+ D_N$ as before.

We study the large deviation principle for $\{h^N(\theta); \theta \in D\}$ under $\mu_N^{\psi,J}$ when $d = 1$ and with Gaussian potential i.e. $V(\eta) = \frac{1}{2}\eta^2$. Let $D = (0,1)$, $D_N = [1, N-1] \cap \mathbb{Z}$ and take the boundary condition $\psi(0) = aN$ and $\psi(N) = bN, a, b \in \mathbb{R}$. We shall denote $\mu_N^{\psi,J}$, $Z_N^{\psi,J}$, μ_N^ψ and Z_N^ψ as $\mu_N^{a,b,J}$, $Z_N^{a,b,J}$, $\mu_N^{a,b}$ and $Z_N^{a,b}$, respectively. Define

$$W_{a,b}(D) = \{h \in C([0,1]; \mathbb{R}); h(0) = a, h(1) = b\},$$

$$H_{a,b}^1(D) = \{h \in W_{a,b}(D); h \text{ is absolutely continuous and } h' \in \mathbb{L}^2(D)\}.$$

The space $W_{a,b}(D)$ is endowed with the topology determined by the sup-norm $\| \cdot \|_\infty$. Then, we have the following LDP.

Theorem 2.2. *Assume that $d = 1$ and $V(\eta) = \frac{1}{2}\eta^2$. Then the family of random surfaces $\{h^N(\theta); \theta \in D\}$ distributed under $\mu_N^{a,b,J}$ satisfies the large deviation principle on $W_{a,b}(D)$ (i.e. the upper and lower bounds for closed and open subsets of $W_{a,b}(D)$, respectively) with speed N and the rate functional given by*

$$I^J(h) = \begin{cases} \Sigma^J(h) - \inf_{H_{a,b}^1(D)} \Sigma^J & \text{if } h \in H_{a,b}^1(D), \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$\Sigma^J(h) = \frac{1}{2}\int_0^1 (h')^2(\theta)d\theta + \tau(J)|\{\theta \in D; h(\theta) = 0\}|,$$

and

$$\tau(J) = -\lim_{N\to\infty} \frac{1}{N} \log \frac{Z_N^{0,0,J}}{Z_N^{0,0}}, \tag{2.7}$$

note that $| \cdot |$ stands for the Lebesgue measure.

Remark 2.4. *The function $\tau(J)$ is the so-called pinning free energy. By the proof of Theorem 2.2 and Remark 6.1 below, one can see that the limit exists and $\tau(J) < 0$ for every $J \in \mathbb{R}$.*

§3. Proof of Theorem 2.1: LDP with Self Potentials

LDP without self potentials.

This section reduces the proof of Theorem 2.1 to the LDP for $\mu_N^{\psi}(=\mu_N^{\psi,0})$, i.e. the Gibbs measure without self potential. The case where the boundary condition $\psi \equiv 0$ was studied in [13].

Proposition 3.1. *The family of random surfaces $\{h^N(\theta); \theta \in D\}$ distributed under μ_N^{ψ} satisfies the large deviation principle on $\mathbb{L}^2(D)$ with speed N^d and the rate functional given by*

$$I(h) = \begin{cases} \Sigma(h) - \inf\limits_{H_g^1(D)} \Sigma & \text{if } h \in H_g^1(D), \\ +\infty & \text{otherwise.} \end{cases}$$

Treatment of boundary conditions.

One of the key observations for the proof of Proposition 3.1 is the following trivial identity:

$$\nabla(\phi \vee \psi)(b) = \nabla((\phi - \xi) \vee 0)(b) + \nabla(\xi \vee \psi)(b), \tag{3.1}$$

for every $\xi = \{\xi(x); x \in D_N\}$ and $b \in \overline{D_N}^*$. Now take ξ as $\xi(x) = Ng(\frac{x}{N})$ for $x \in D_N$ (and for $x \in \overline{D_N}$; recall $g \in C^\infty(\mathbb{R}^d)$) and define

$$\widetilde{H}_N^{\psi}(\phi) = \frac{1}{2} \sum_{b \in \overline{D_N}^*} V(\nabla(\phi \vee 0)(b) + \nabla(\xi \vee \psi)(b)).$$

Consider the finite volume Gibbs measure with Hamiltonian $\widetilde{H}_N^{\psi}(\phi)$ and 0-boundary condition:

$$\tilde{\mu}_N^{\psi}(d\phi) = \frac{1}{\widetilde{Z}_N^{\psi}} \exp\{-\widetilde{H}_N^{\psi}(\phi)\} \prod_{x \in D_N} d\phi(x).$$

Then the following LDP holds for $\tilde{\mu}_N^{\psi}$.

Proposition 3.2. *The family of random surfaces $\{h^N(\theta); \theta \in D\}$ distributed under $\tilde{\mu}_N^{\psi}$ satisfies the large deviation principle on $\mathbb{L}^2(D)$ with speed N^d and the rate functional given by*

$$\widetilde{I}(h) = \begin{cases} \widetilde{\Sigma}(h) - \inf\limits_{H_0^1(D)} \widetilde{\Sigma} & \text{if } h \in H_0^1(D), \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$\widetilde{\Sigma}(h) = \int_D \sigma(\nabla h(\theta) + \nabla g(\theta))d\theta.$$

We shall prove this proposition in Sections 4 and 5.

Proof of Proposition 3.1. Consider the continuous map $\Phi_g : \mathbb{L}^2(D) \to \mathbb{L}^2(D)$ given by $\Phi_g(h) = h + g$. It is easy to see that

$$I(h) = \inf\{\widetilde{I}(\tilde{h}); \tilde{h} \in \mathbb{L}^2(D), \Phi_g(\tilde{h}) = h\}.$$

Then by definitions of μ_N^ψ, $\tilde{\mu}_N^\psi$ and (3.1), Proposition 3.1 follows from the contraction principle (cf. [25], [14] and [12, Theorem 4.2.1]) and Proposition 3.2. Q.E.D.

Deduction of Theorem 2.1 from Proposition 3.1.

We shall prove Theorem 2.1 assuming that Proposition 3.2 and therefore Proposition 3.1 are shown. We only consider the case where $\alpha \geq \beta$. The case where $\alpha \leq \beta$ can be proved completely in an analogous manner or by turning the interfaces upside down by the map $\phi \mapsto -\phi$ and $\psi \mapsto -\psi$. The pinning potential $U(\theta, r) = Q(\theta)W(r)$ which satisfies the conditions (W1) and (W2) with $\alpha \geq \beta$ can be rewritten as $U(\theta, r) = Q(\theta)\alpha + Q(\theta)\widetilde{W}(r)$ and $\widetilde{W}(r)$ satisfies conditions (W1) and

(W2)′ there exists $A \geq 0$ such that $\lim_{r\to+\infty} W(r) = 0$, $\lim_{r\to-\infty} W(r) = -A$ and $-A \leq W(r) \leq 0$ for every $r \in \mathbb{R}$,

with $A = \alpha - \beta$. Since the contribution of the first term $Q(\theta)\alpha$ in $\exp\{-H_N^{\psi,U}(\phi)\}$ of $\mu_N^{\psi,U}$ cancels with the normalization factor, we only have to consider the case that W satisfies the conditions (W1) and (W2)′.

The following lemma allows us to replace the self potential part of the Hamiltonian by the integration of $-AQ$ on the domain where $g \in \mathbb{L}^2(D)$ is non-positive when the macroscopically scaled profile h^N is close enough to g. Note that g here represents a general function in $\mathbb{L}^2(D)$ and not the macroscopic boundary condition.

Lemma 3.1. *Assume the conditions* (Q1), (W1) *and* (W2)′ *on* $U(\theta, r) = Q(\theta)W(r)$. *Let* $g \in \mathbb{L}^2(D)$ *and* $0 < \delta < 1$ *be fixed. If* $h^N \in B_2(g, \delta) = \{h \in \mathbb{L}^2(D); \|h-g\|_{\mathbb{L}^2(D)} < \delta\}$ *for* N *large enough, then there exists some constant* $C > 0$ *such that*

$$\sum_{x\in D_N} U(\frac{x}{N}, \phi(x)) + N^d A \int_D Q(\theta)1(g(\theta) \leq -\delta^{\frac{1}{2}})d\theta \leq CN^d\delta,$$

for every N *large enough.*

Proof. There exists an approximating sequence $\{g_k\}_{k\geq 1} \subset C(D)$ of $g \in \mathbb{L}^2(D)$ such that $\|g_k - g\|_{\mathbb{L}^2(D)} \to 0$ as $k \to \infty$. Recall that one can define g_k^N (polilinear functions) and $\bar{g}_k^N$ (step functions) for $g_k \in C(D)$. Now, by (2.2), it holds that

$$\|\bar{h}^N - g\|_{\mathbb{L}^2(D)} \leq C\|h^N - g\|_{\mathbb{L}^2(D)} + a_{N,k},$$

for every $k \geq 1$, where

$$a_{N,k} = (C+1)\|g - g_k\|_{\mathbb{L}^2(D)} + C\|g_k - g_k^N\|_{\mathbb{L}^2(D)} + \|g_k - \bar{g}_k^N\|_{\mathbb{L}^2(D)},$$

which goes to 0 as $N \to \infty$ and $k \to \infty$. Hence,

$$\|\bar{h}^N - g\|_{\mathbb{L}^2(D)} < C\delta + a_{N,k}, \tag{3.2}$$

if $h^N \in B_2(g,\delta)$. The positive constants C in the estimates may change from line to line in the paper.

Now, for $\gamma > 0$, we rewrite

$$\begin{aligned}
&\sum_{x\in D_N} U(\frac{x}{N}, \phi(x)) + N^d A \int_D Q(\theta) 1(g(\theta) \leq -\gamma) d\theta \\
&\quad = N^d \int_D \big(W(N\bar{h}^N(\theta)) + A1(g(\theta) \leq -\gamma)\big) Q(\theta) d\theta \\
&\qquad + \Big\{ \sum_{x\in D_N} Q(\frac{x}{N}) W(N\bar{h}^N(\frac{x}{N})) - N^d \int_D W(N\bar{h}^N(\theta)) Q(\theta) d\theta \Big\} \\
&\quad \equiv S_1 + S_2.
\end{aligned}$$

For S_1, we divide the integration on D into the sum of those on three domains $\{g > -\gamma\} (\equiv \{\theta \in D; g(\theta) > -\gamma\})$, $\{g \leq -\gamma\} \cap C_{N,\gamma}^c$ and $\{g \leq -\gamma\} \cap C_{N,\gamma}$, where $C_{N,\gamma} = \{|\bar{h}^N - g| < \gamma/2\}$ and $C_{N,\gamma}^c = D \setminus C_{N,\gamma}$. The integration on $\{g > -\gamma\}$ is non-positive, because $Q \geq 0$, $W \leq 0$ and $A1(g(\theta) \leq -\gamma) = 0$ on this domain. Next, since (3.2) implies $|C_{N,\gamma}^c| < \frac{4}{\gamma^2}(C\delta + a_{N,k})^2$, we obtain

$$\int_{\{g\leq -\gamma\}\cap C_{N,\gamma}^c} \big| W(N\bar{h}^N(\theta)) + A1(g(\theta) \leq 0) \big| d\theta \leq \frac{K}{\gamma^2}(C\delta + a_{N,k})^2,$$

where $K = 4(\|W\|_\infty + A)$. On $\{g \leq -\gamma\} \cap C_{N,\gamma}$, we have $\bar{h}^N(\theta) < -\gamma/2$. By this fact and the assumption (W2)$'$, $\big| W(N\bar{h}^N(\theta)) + A1(g(\theta) \leq -\gamma) \big| \leq \delta$ holds for N large enough and we see that

$$\int_{\{g\leq -\gamma\}\cap C_{N,\gamma}} \big| W(N\bar{h}^N(\theta)) + A1(g(\theta) \leq -\gamma) \big| d\theta \leq \delta |D|.$$

Therefore, we obtain

$$S_1 \leq N^d \|Q\|_\infty \left(\frac{K}{\gamma^2}(C\delta + a_{N,k})^2 + \delta|D|\right),$$

for N large enough, every $k \geq 1$ and $\gamma > 0$. For S_2, we have

$$|S_2| \leq N^d \|W\|_\infty \int_D \left| Q\left(\frac{[N\theta]}{N}\right) - Q(\theta) \right| d\theta + O(N^{d-1}),$$

where $O(N^{d-1})$ is the boundary term. Finally, taking $\gamma = \delta^{\frac{1}{2}}$ and N, k large enough, we complete the proof. Q.E.D.

Under the condition (W2)$'$, the rate functional $\Sigma^U(h)$ has the form

$$\Sigma^U(h) = \Sigma(h) - A \int_D Q(\theta) 1(h(\theta) \leq 0) d\theta, \tag{3.3}$$

which coincides with (1.4), and enjoys the following properties.

Lemma 3.2. (1) *The functional $\Sigma^U(h)$ is lower semi-continuous on $\mathbb{L}^2(D)$.*
(2) *Let $\Sigma^U_-(h)$ be the functional defined by* (3.3) *with $1(h(\theta) \leq 0)$ replaced by $1(h(\theta) < 0)$. Then, for every open set $\mathcal{O}$ of $\mathbb{L}^2(D)$, we have that*

$$\inf_{h \in \mathcal{O}} \Sigma^U(h) = \inf_{h \in \mathcal{O}} \Sigma^U_-(h).$$

Proof. (1) Decomposing D into two domains $C_\gamma = \{|h - g| < \gamma\}$ and C_γ^c, in a similar way to the proof of Lemma 3.1, one can prove that

$$\int_D Q(\theta) 1(h(\theta) \leq 0) d\theta \leq \int_D Q(\theta) 1(g(\theta) \leq \gamma) d\theta + \|Q\|_\infty \frac{\delta^2}{\gamma^2},$$

for every $\gamma > 0$ if $h \in B_2(g, \delta)$. By this inequality and the property (strict convexity) of the surface tension (cf. [13, Lemma 3.6]):

$$\frac{1}{2} c_- |v - u|^2 \leq \sigma(v) - \sigma(u) - (v - u) \cdot (\nabla \sigma)(u) \leq \frac{1}{2} c_+ |v - u|^2, \tag{3.4}$$

for every $u, v \in \mathbb{R}^d$, it is easy to see the lower semi-continuity of $\Sigma^U(h)$ on $\mathbb{L}^2(D)$.
(2) Since $\Sigma^U(h) \leq \Sigma^U_-(h)$ is obvious for every $h \in \mathbb{L}^2(D)$, the conclusion follows once we can show that

$$\inf_{h \in \mathcal{O}} \Sigma^U(h) \geq \inf_{h \in \mathcal{O}} \Sigma^U_-(h). \tag{3.5}$$

To this end, for every $\varepsilon > 0$, take $h \in \mathcal{O}$ such that $\Sigma^U(h) \le \inf_{\mathcal{O}} \Sigma^U + \varepsilon$. We approximate such h by a sequence $\{h^n\}_{n\ge 1}$ defined by $h^n(\theta) = h(\theta) - f^n(\theta)$, where $f^n \in C_0^\infty(D)$ are functions such that $f^n(\theta) \equiv \frac{1}{n}$ on $D_n = \{\theta \in D; \text{dist } (\theta, \partial D) \ge \frac{1}{n}\}$ and $|\nabla f^n(\theta)| \le C$ with $C > 0$. Note that h^n satisfy the same boundary condition as h. Then, since $\lim_{n\to\infty} \Sigma(h^n) = \Sigma(h)$ (recall $h \in H_g^1(D)$) and

$$\begin{aligned} -A\int_D Q(\theta)1(h^n(\theta) < 0)d\theta &\le -A\int_{D_n} Q(\theta)1(h(\theta) < \frac{1}{n})d\theta \\ &\le -A\int_D Q(\theta)1(h(\theta) \le 0)d\theta + A\|Q\|_\infty |D \setminus D_n|, \end{aligned}$$

we obtain $\limsup_{n\to\infty} \Sigma_-^U(h^n) \le \Sigma^U(h)$. However, $\mathcal{O}$ is an open set of $\mathbb{L}^2(D)$, so that $h^n \in \mathcal{O}$ for n large enough and thus (3.5) is shown. Q.E.D.

Proof of Theorem 2.1. Step1 (lower bound). Let $g \in \mathbb{L}^2(D)$ and $\delta > 0$. Then, by Lemma 3.1 and the LDP lower bound for μ_N^ψ (Proposition 3.1), we have

$$\begin{aligned} &\liminf_{N\to\infty} \frac{1}{N^d} \log \frac{Z_N^{\psi,U}}{Z_N^\psi} \mu_N^{\psi,U}(h^N \in B_2(g,\delta)) \\ &\ge -\inf_{h\in B_2(g,\delta)} I(h) + A\int_D Q(\theta)1(g(\theta) \le -\delta^{\frac{1}{2}})d\theta - C\delta \\ &\ge -\{I(g) - A\int_D Q(\theta)1(g(\theta) \le -\delta^{\frac{1}{2}})d\theta\} - C\delta. \end{aligned}$$

Take now an arbitrary open set $\mathcal{O}$ of $\mathbb{L}^2(D)$. Then,

$$\begin{aligned} &\liminf_{N\to\infty} \frac{1}{N^d} \log \frac{Z_N^{\psi,U}}{Z_N^\psi} \mu_N^{\psi,U}(h^N \in \mathcal{O}) \\ &\ge -\{I(h) - A\int_D Q(\theta)1(h(\theta) \le -\delta^{\frac{1}{2}})d\theta\} - C\delta \end{aligned}$$

for every $h \in \mathcal{O}$ and $\delta > 0$ such that $B_2(h,\delta) \subset \mathcal{O}$. Letting $\delta \downarrow 0$, since $h \in \mathcal{O}$ is arbitrary, we have

$$\begin{aligned} (3.6) \qquad &\liminf_{N\to\infty} \frac{1}{N^d} \log \frac{Z_N^{\psi,U}}{Z_N^\psi} \mu_N^{\psi,U}(h^N \in \mathcal{O}) \\ &\ge -\inf_{h\in\mathcal{O}} \{I(h) - A\int_D Q(\theta)1(h(\theta) < 0)d\theta\}. \end{aligned}$$

However, by Lemma 3.2-(2), one can replace $1(h(\theta) < 0)$ with $1(h(\theta) \leq 0)$ in the right hand side of (3.6).

Step2 (*upper bound*). Let $g \in \mathbb{L}^2(D)$ and $\delta > 0$ be fixed. We define

$$\begin{aligned} L_N^+ &= N\{\theta \in D; g(\theta) > \delta^{\frac{1}{2}}\} \cap \mathbb{Z}^d, \\ L_N^- &= N\{\theta \in D; g(\theta) < -\delta^{\frac{1}{2}}\} \cap \mathbb{Z}^d, \\ I_N &= N\{\theta \in D; |g(\theta)| \leq \delta^{\frac{1}{2}}\} \cap \mathbb{Z}^d. \end{aligned}$$

By the assumption (W2)$'$ on W, for every $\varepsilon > 0$ there exists $K = K_\varepsilon > 0$ such that $W(r) \geq -(A-\varepsilon)1_{\{r \leq K\}} - \varepsilon$ for every $r \in \mathbb{R}$. Therefore, we have

$$\begin{aligned} &\exp\{-\sum_{x \in D_N} U(\frac{x}{N}, \phi(x))\} \\ &\leq \exp\{(A-\varepsilon)\sum_{x \in D_N} Q(\frac{x}{N})1(\phi(x) \leq K) + \varepsilon \sum_{x \in D_N} Q(\frac{x}{N})\} \\ &= \exp\{\varepsilon \sum_{x \in D_N} Q(\frac{x}{N})\} \sum_{\Lambda \subset D_N} \prod_{x \in \Lambda} (e^{(A-\varepsilon)Q(\frac{x}{N})} - 1)1(\phi(x) \leq K). \end{aligned}$$

Now, if $\phi(x) \leq K$ for $x \in L_N^+$, then $\frac{1}{N}\phi(x) - g(\frac{x}{N}) < -\frac{1}{2}\delta^{\frac{1}{2}}$ for N large enough. Thus, if $\phi(x) \leq K$ for every $x \in \Lambda \subset L_N^+$ on $\{h^N \in B_2(g,\delta)\}$, since $\|\bar{h}^N - \bar{g}^N\|_{\mathbb{L}^2(D)} < \frac{1}{C_0}(\delta + \|g - g^N\|_{\mathbb{L}^2(D)})$, we have for N large enough

$$\frac{2\delta^2}{C_0} > \frac{1}{N^d} \sum_{x \in D_N} (\frac{1}{N}\phi(x) - g(\frac{x}{N}))^2 > \frac{|\Lambda|\delta}{4N^d},$$

namely, $|\Lambda| < 8C_0^{-1}\delta N^d$, where $C_0 > 0$ is the constant appeared in (2.2). Combining these facts

$$\begin{aligned} &\exp\{-\varepsilon \sum_{x \in D_N} Q(\frac{x}{N})\} \frac{Z_N^{\psi,U}}{Z_N^\psi} \mu_N^{\psi,U}(h^N \in B_2(g,\delta)) \\ &\leq \sum_{\substack{\Lambda \subset L_N^+ \\ |\Lambda| < 8C_0^{-1}\delta N^d}} \prod_{x \in \Lambda} (e^{(A-\varepsilon)Q(\frac{x}{N})} - 1) \sum_{\Lambda' \subset I_N \cup L_N^-} \prod_{x \in \Lambda'} (e^{(A-\varepsilon)Q(\frac{x}{N})} - 1) \\ &\quad \times \frac{1}{Z_N^\psi} \int 1(h^N \in B_2(g,\delta))1(\phi(x) \leq K \text{ for every } x \in \Lambda \cup \Lambda') \\ &\qquad \times \exp\{-H_N^\psi(\phi)\} \prod_{x \in D_N} d\phi(x) \end{aligned}$$

$$\le \left(e^{(A-\varepsilon)\|Q\|_\infty}-1\right)^{8C_0^{-1}\delta N^d}\,\left|\{\Lambda\subset L_N^+; |\Lambda|<8C_0^{-1}\delta N^d\}\right| \\ \times \exp\{(A-\varepsilon)\sum_{x\in I_N\cup L_N^-} Q(\frac{x}{N})\}\mu_N^\psi(h^N\in B_2(g,\delta)).$$

By using Stirling's formula, we see that

$$\begin{aligned}\left|\{\Lambda\subset L_N^+; |\Lambda|<8C_0^{-1}\delta N^d\}\right| &\le \frac{(CN^d)^{8C_0^{-1}\delta N^d}}{(8C_0^{-1}\delta N^d)!}\\ &\le \frac{C}{\delta}N^d(\frac{C}{\delta})^{C\delta N^d}(1+o(1))\end{aligned}$$

as $N\to\infty$, for some constant $C>0$ independent of N and δ. Hence, by the LDP upper bound for the measure μ_N^ψ (Proposition 3.1), we obtain

$$\begin{aligned}&\limsup_{N\to\infty}\frac{1}{N^d}\log\frac{Z_N^{\psi,U}}{Z_N^\psi}\mu_N^{\psi,U}(h^N\in B_2(g,\delta))\\ &\le (A-\varepsilon)\int_D Q(\theta)1(g(\theta)\le\delta^{\frac{1}{2}})d\theta\\ &\quad-\inf_{h\in\bar{B}_2(g,\delta)} I(h)+C(\delta)+\varepsilon\int_D Q(\theta)d\theta,\end{aligned}$$

where $C(\delta)$ is a constant independent of N and goes to 0 as $\delta\to 0$. Then, by using the lower semi-continuity of $I(h)$ and the right-continuity of $\int_D Q(\theta)1(g(\theta)\le\delta^{\frac{1}{2}})d\theta$ in δ, we see that for every $g\in\mathbb{L}^2(D)$ and $\varepsilon>0$, there exists $\delta>0$ small enough such that

$$\begin{aligned}&\limsup_{N\to\infty}\frac{1}{N^d}\log\frac{Z_N^{\psi,U}}{Z_N^\psi}\mu_N^{\psi,U}(h^N\in B_2(g,\delta))\\ &\le -\{I(g)-A\int_D Q(\theta)1(g(\theta)\le 0)d\theta\}+\varepsilon.\end{aligned}$$

Therefore, the standard argument in the theory of LDP yields

$$\begin{aligned}(3.7)\qquad&\limsup_{N\to\infty}\frac{1}{N^d}\log\frac{Z_N^{\psi,U}}{Z_N^\psi}\mu_N^{\psi,U}(h^N\in \mathcal{C})\\ &\le -\inf_{h\in\mathcal{C}}\{I(h)-A\int_D Q(\theta)1(h(\theta)\le 0)d\theta\},\end{aligned}$$

for every compact set $\mathcal{C}$ of $\mathbb{L}^2(D)$. Since U is bounded, exponential tightness for $\mu_N^{\psi,U}$ can be proved in a similar way to those for μ_N^ψ which will be proved in Section 4 (see Remark 4.1 below). Thus, (3.7) holds for every closed set $\mathcal{C}$ of $\mathbb{L}^2(D)$.

Finally, taking $\mathcal{O} = \mathcal{C} = \mathbb{L}^2(D)$ in (3.6) (recall the remark subsequent to the estimate) and (3.7), we see that

$$\lim_{N\to\infty} \frac{1}{N^d} \log \frac{Z_N^{\psi,U}}{Z_N^{\psi}} = - \inf_{H_g^1(D)} \Sigma^U + \inf_{H_g^1(D)} \Sigma, \tag{3.8}$$

and this concludes the proof. Q.E.D.

Remark 3.1. *As we mentioned in Remark 2.2, if* U *is given by* $U(\theta, r) = QW(r)$ *for some constant* $Q \geq 0$ *and* $W(r)$ *(or* $W(-r)$*) satisfying the condition* (W2)$'$, *then (3.8) with* $D_N = \Lambda_N$ *yields the difference of the free energies of the interface in the case with and without self potentials, see (2.6). This can also be proved in the following way under the condition* (W2)$'$*: for every* $\varepsilon \in (0, A)$ *there exists* $K = K_\varepsilon > 0$ *such that* $W(r) \leq -(A-\varepsilon)1_{\{r \leq -K\}}$ *for every* $r \in \mathbb{R}$*. Therefore, we have*

$$\begin{aligned} \frac{Z_{\Lambda_N}^{0,U}}{Z_{\Lambda_N}^0} &= E^{\mu_{\Lambda_N}^0}\big[\exp\{-Q \sum_{x\in\Lambda_N} W(\phi(x))\}\big] \\ &\geq E^{\mu_{\Lambda_N}^0}\big[\exp\{(A-\varepsilon)Q \sum_{x\in\Lambda_{N,\varepsilon}} 1(\phi(x) \leq -K)\}\big] \\ &= E^{\mu_{\Lambda_N}^0}\big[\sum_{\Gamma\subset\Lambda_{N,\varepsilon}} (e^{(A-\varepsilon)Q}-1)^{|\Gamma|} 1(\phi(x) \leq -K \textit{ for every } x \in \Gamma)\big] \\ &\geq e^{(A-\varepsilon)Q|\Lambda_{N,\varepsilon}|} \mu_{\Lambda_N}^0(\phi(x) \leq -K \textit{ for every } x \in \Lambda_{N,\varepsilon}), \end{aligned}$$

where $\Lambda_{N,\varepsilon} = \{x \in \Lambda_N; \mathrm{dist}(x, \Lambda_N^c) \geq \varepsilon N\}$*. However,* [6, Proposition 2.1] *shows that the probability in the last line is bounded below by*

$$\exp\{-CN^{d-2}\log N(1+o(1))\},$$

as $N \to \infty$ *for some constant* $C > 0$ *independent of* N*. This implies*

$$\liminf_{N\to\infty} \frac{1}{N^d} \log \frac{Z_{\Lambda_N}^{0,U}}{Z_{\Lambda_N}^0} \geq AQ.$$

The opposite inequality is obvious, since $W(r) \geq -A$*.*

§4. Proof of Proposition 3.2: LDP without Self Potentials

Convergence of average profiles.

In this section, the proof of Proposition 3.2 will be given assuming the convergence of average profiles (Lemma 4.1). We shall follow the strategy of [13]. The only difference is that the Dirichlet boundary data $g\,|_{\partial D}$ is given from $g \in C^\infty(\mathbb{R}^d)$ in our case, while [13] treated the case of $g \equiv 0$. For $f \in C_0^\infty(D)$, set

$$H_{N,f}^\psi(\phi) = H_N^\psi(\phi) - \frac{1}{N}\sum_{x\in D_N} f(\frac{x}{N})\phi(x),$$
$$\widetilde{H}_{N,f}^\psi(\phi) = \widetilde{H}_N^\psi(\phi) - \frac{1}{N}\sum_{x\in D_N} f(\frac{x}{N})\phi(x),$$

and consider the following two Gibbs probability measures:

$$\mu_{N,f}^\psi(d\phi) = \frac{1}{Z_{N,f}^\psi}\exp\{-H_{N,f}^\psi(\phi)\}\prod_{x\in D_N} d\phi(x),$$
$$\tilde{\mu}_{N,f}^\psi(d\phi) = \frac{1}{\widetilde{Z}_{N,f}^\psi}\exp\{-\widetilde{H}_{N,f}^\psi(\phi)\}\prod_{x\in D_N} d\phi(x),$$

having the different boundary conditions $\phi(x) = \psi(x)$ and $\phi(x) = 0$ for $x \in \partial^+ D_N$, respectively; recall that ψ and g satisfy the conditions (ψ1), (ψ2). We write the averages of the profile h^N defined by (2.1) under $\mu_{N,f}^\psi$ and $\tilde{\mu}_{N,f}^\psi$ as $\bar{h}_{N,f}^\psi(\theta) = E^{\mu_{N,f}^\psi}[h^N(\theta)]$ and $\widetilde{h}_{N,f}^\psi(\theta) = E^{\tilde{\mu}_{N,f}^\psi}[h^N(\theta)]$, respectively. For $f \in \mathbb{L}^2(D)$, h_f denotes the unique weak solution $h = h(\theta)$ in $H_0^1(D)$ of the following elliptic partial differential equation:

$$\mathrm{div}\{(\nabla\sigma)(\nabla h(\theta) + \nabla g(\theta))\} = -f(\theta), \quad \theta \in D.$$

The crucial step in the proof of Proposition 3.2 is the following lemma.

Lemma 4.1.

$$\widetilde{h}_{N,f}^\psi \to h_f \textit{ in } H_0^1(D) \textit{ as } N \to \infty.$$

We shall prove this lemma in Section 5. Next, define

$$\Xi_{N,f}^\psi \equiv \frac{\widetilde{Z}_{N,f}^\psi}{\widetilde{Z}_N^\psi} = E^{\tilde{\mu}_N^\psi}\Big[\exp\{\frac{1}{N}\sum_{x\in D_N} f(\frac{x}{N})\phi(x)\}\Big].$$

Then, in a similar way to the proof of Theorem 1.1 of [13] , by calculating the functional derivative of $\widetilde{\Sigma}(h)$ and using the differentiation-integration trick (i.e. computing $\frac{d}{dt}\log \widetilde{Z}_{N,tf}^\psi$ and integrating it in $t \in [0,1]$), Lemma 4.1 yields the following lemma. The proof is omitted.

Lemma 4.2. *The limit* $\Lambda(f) \equiv \lim_{N\to\infty} \frac{1}{N^d} \log \Xi^{\psi}_{N,f}$ *exists and it holds that*

$$\begin{aligned}\Lambda(f) &= \int_D \int_0^1 h_{tf}(\theta) f(\theta) dt d\theta, \\ &= \sup_{h\in H_0^1(D)} \{\langle h, f\rangle - \widetilde{\Sigma}(h)\} + \inf_{H_0^1(D)} \widetilde{\Sigma}, \\ &= \langle h_f, f\rangle - \widetilde{\Sigma}(h_f) + \inf_{H_0^1(D)} \widetilde{\Sigma},\end{aligned}$$

where $\langle h, f\rangle = \int_D h(\theta) f(\theta) d\theta$.

Exponential tightness.

For the proof of the LDP upper bound in Proposition 3.2, we prepare the following lemma.

Lemma 4.3. *There exists* $\varepsilon > 0$ *such that*

$$\sup_{N\ge 1} \frac{1}{N^d} \log E^{\tilde{\mu}^{\psi}_{N,f}} \Big[\exp\{\varepsilon \sum_{x\in\overline{D_N}} (|h^N(\frac{x}{N})|^2 + |\nabla^N h^N(\frac{x}{N})|^2)\}\Big] < \infty,$$

where for a scalar lattice field $\{u(\frac{x}{N}); x \in \overline{D_N}\}$, $\nabla^N u(\frac{x}{N}) = \{\nabla^N_j u(\frac{x}{N})\}_{1\le j\le d}$ *denotes a discrete gradient of* u *defined by* $\nabla^N_j u(\frac{x}{N}) = N\{u(\frac{x+e_j}{N}) - u(\frac{x}{N})\}, 1 \le j \le d$.

Proof. Since D is bounded, by discrete Poincaré's inequality and the definition of h^N, we only have to prove that there exists $\varepsilon > 0$ such that

$$\sup_{N\ge 1} \frac{1}{N^d} \log E^{\tilde{\mu}^{\psi}_{N,f}} \Big[\exp\{\varepsilon \sum_{b\in\overline{D_N}^*} |\nabla\phi(b)|^2\}\Big] < \infty. \tag{4.1}$$

However, this is shown by a simple direct computation. Indeed, by the strict convexity of V, it is easy to see that

$$\begin{aligned}&\frac{1}{2} c_- H_N^{0,*}(\phi) - \frac{1}{4} c_- \sum_{b\in\overline{D_N}^*} (\nabla(\xi \vee \psi)(b))^2 \\ &\le \widetilde{H}^{\psi}_N(\phi) \le 2c_+ H_N^{0,*}(\phi) + \frac{1}{2} c_+ \sum_{b\in\overline{D_N}^*} (\nabla(\xi \vee \psi)(b))^2,\end{aligned}$$

where $H_N^{0,*}(\phi) = \frac{1}{4} \sum_{b\in\overline{D_N}^*} \big(\nabla(\phi \vee 0)(b)\big)^2$. Therefore, the expectation in (4.1) is bounded above by

$$\exp\Big\{\Big(\frac{c_-}{4}+\frac{c_+}{2}\Big)\sum_{b\in\overline{D_N}^*}|\nabla(\xi\vee\psi)(b)|^2\Big\}$$
$$\times\frac{\int\exp\big\{\big(4\varepsilon-\frac{c_-}{2}\big)H_N^{0,*}(\phi)+\frac{1}{N}\sum_{x\in D_N}f(\frac{x}{N})\phi(x)\big\}\prod_{x\in D_N}d\phi(x)}{\int\exp\big\{-2c_+H_N^{0,*}(\phi)+\frac{1}{N}\sum_{x\in D_N}f(\frac{x}{N})\phi(x)\big\}\prod_{x\in D_N}d\phi(x)}.$$

A simple Gaussian calculation yields

$$\int\exp\Big\{-\alpha H_N^{0,*}(\phi)+\frac{1}{N}\sum_{x\in D_N}f(\frac{x}{N})\phi(x)\Big\}\prod_{x\in D_N}d\phi(x)$$
$$=\Big(\frac{2\pi}{\alpha}\Big)^{\frac{|D_N|}{2}}\sqrt{\det(-\Delta_{D_N})}\exp\Big\{\frac{1}{2\alpha N^2}V_{N,f}\Big\},$$

for every $\alpha > 0$, where Δ_{D_N} is a discrete Laplacian on D_N with 0-boundary condition,

$$V_{N,f}=\Big(f(\frac{\cdot}{N}),\ (-\Delta_{D_N})^{-1}f(\frac{\cdot}{N})\Big)_{D_N}=\mathrm{Var}_{\mu_N^{0,*}}\Big(\sum_{x\in D_N}f(\frac{x}{N})\phi(x)\Big),$$

and $(\ \cdot\ ,\ \cdot\ )_{D_N}$ denotes $l^2(D_N)$-scalar product. Therefore, for every $0<\varepsilon<\frac{1}{8}c_-$, we obtain

$$\log E^{\tilde\mu^{\psi}_{N,f}}\Big[\exp\Big\{\varepsilon\sum_{b\in\overline{D_N}^*}|\nabla\phi(b)|^2\Big\}\Big]$$
$$\le C|D_N|+C\frac{1}{N^2}V_{N,f}+C\sum_{b\in\overline{D_N}^*}|\nabla(\xi\vee\psi)(b)|^2,$$

for some $C = C_\varepsilon > 0$ independent of N. However, $V_{N,f} = O(N^{d+2})$ (cf. [13, Lemma 2.8]) and

$$\sum_{b\in\overline{D_N}^*}|\nabla(\xi\vee\psi)(b)|^2$$
$$\le 2\sum_{b\in\overline{D_N}^*}|\nabla\xi(b)|^2+2\sum_{x\in\partial^+D_N}|\xi(x)-\psi(x)|^2=O(N^d),$$

as $N\to\infty$ by recalling the assumption on ψ and that $\xi(x) = Ng(\frac{x}{N})$ for $x\in\overline{D_N}$ with $g\big|_D\in C^\infty(\bar D)$. This concludes the proof of (4.1). Q.E.D.

Proof of Proposition 3.2.

Proof of Proposition 3.2; upper bound. For every $f \in C_0^\infty(D)$ and measurable set $\mathcal{E}$ of $\mathbb{L}^2(D)$, Chebyshev's inequality shows

$$\tilde{\mu}_N^\psi(h^N \in \mathcal{E}) \le \exp\{-N^d \inf_{h\in\mathcal{E}}\langle h, f\rangle\} E^{\tilde{\mu}_N^\psi}\left[\exp\{N^d\langle h^N, f\rangle\}\right]. \tag{4.2}$$

Noting that

$$N^d\langle h^N, f\rangle \le \frac{1}{N}\sum_{x\in\overline{D_N}} f(\frac{x}{N})\phi(x) + \frac{1}{N^2}\|\nabla f\|_\infty \sum_{x\in\overline{D_N}} |\phi(x)|,$$

and using Hölder's inequality, the expectation in the right hand side of (4.2) is bounded above by

$$E^{\tilde{\mu}_N^\psi}\left[\exp\{\frac{p}{N}\sum_{x\in\overline{D_N}} f(\frac{x}{N})\phi(x)\}\right]^{\frac{1}{p}} E^{\tilde{\mu}_N^\psi}\left[\exp\{\frac{q}{N^2}\|\nabla f\|_\infty \sum_{x\in\overline{D_N}} |\phi(x)|\}\right]^{\frac{1}{q}}$$
$$\equiv I_1^N \times I_2^N,$$

for $p, q > 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$. However, Lemmas 4.2 and 4.3 imply

$$\lim_{N\to\infty} \frac{1}{N^d}\log I_1^N = \frac{1}{p}\Lambda(pf),$$

and

$$\limsup_{N\to\infty} \frac{1}{N^d}\log I_2^N \le 0,$$

respectively. Hence, we have

$$\limsup_{N\to\infty} \frac{1}{N^d}\log \tilde{\mu}_N^\psi(h^N \in \mathcal{E}) \le -\inf_{h\in\mathcal{E}}\langle h, f\rangle + \frac{1}{p}\Lambda(pf).$$

Now, by (3.4), we can prove the continuity of h_f in $H_0^1(D)$ with respect to $f \in \mathbb{L}^2(D)$ (cf. [13, Section 3.5]). Therefore, by taking the limit $p \downarrow 1$ and infimum with respect to $f \in C_0^\infty(D)$, we obtain

$$\limsup_{N\to\infty} \frac{1}{N^d}\log \tilde{\mu}_N^\psi(h^N \in \mathcal{E}) \le -\sup_{f\in C_0^\infty(D)} \inf_{h\in\mathcal{E}}\{\langle h, f\rangle - \Lambda(f)\}.$$

Then by using Lemma 4.2, mini-max theorem (cf. [22, Appendix 2 Lemma 3.2]) and duality lemma (cf. [12, Lemma 4.5.8]), the standard argument yields the LDP upper bound for every compact set of $\mathbb{L}^2(D)$. This can be generalized for every closed set, since the exponential tightness of $\tilde{\mu}_{N,f}^\psi$ follows from Lemma 4.3. Q.E.D.

Remark 4.1. *Since the potential U is bounded, by recalling* (3.1) *and the assumption on ψ, we see that the estimate in Lemma 4.3 holds for $\mu_N^{\psi,U}$ in place of $\tilde{\mu}_{N,f}^{\psi}$ for some $\varepsilon_0 > 0$, which might be smaller than that in Lemma 4.3. In particular, the exponential tightness holds for $\mu_N^{\psi,U}$.*

Proof of Proposition 3.2; lower bound. By Lemmas 4.1 and 4.2, it is easy to see that

$$\lim_{N\to\infty} \frac{1}{N^d} H(\tilde{\mu}_{N,f}^{\psi} | \tilde{\mu}_N^{\psi}) = \widetilde{I}(h_f),$$

where $H(\tilde{\mu}_{N,f}^{\psi}|\tilde{\mu}_N^{\psi}) = E^{\tilde{\mu}_{N,f}^{\psi}}\big[\log \frac{d\tilde{\mu}_{N,f}^{\psi}}{d\tilde{\mu}_N^{\psi}}\big]$ is the relative entropy of $\tilde{\mu}_{N,f}^{\psi}$ with respect to $\tilde{\mu}_N^{\psi}$; see (5.4) in [13]. On the other hand, by Lemma 4.1, Brascamp-Lieb inequality (cf. [13, Lemma 2.8]) and the definition of $\tilde{h}_{N,f}^{\psi}$, one can prove that $\lim_{N\to\infty} E^{\tilde{\mu}_{N,f}^{\psi}}[\|h^N - h_f\|_{\mathbb{L}^2(D)}^2] = 0$ (cf. (1.39) in [13]), and this implies $\lim_{N\to\infty} \tilde{\mu}_{N,f}^{\psi}(h^N \in \mathcal{O}) = 1$ for every open set $\mathcal{O} \subset \mathbb{L}^2(D)$ satisfying $h_f \in \mathcal{O}$. Combining these two facts with the entropy inequality (cf. [14, Lemma 5.4.21]), we obtain

$$\liminf_{N\to\infty} \frac{1}{N^d} \log \tilde{\mu}_N^{\psi}(h^N \in \mathcal{O}) \geq - \inf_{\substack{f\in C_0^\infty(D) \\ \text{s.t. } h_f\in\mathcal{O}}} \widetilde{I}(h_f).$$

However, we can prove by (3.4) that if $h_{f_n} \to h$ in $H_0^1(D)$ as $n \to \infty$ for $\{f_n\} \subset C_0^\infty(D)$ then $\widetilde{I}(h_{f_n}) \to \widetilde{I}(h)$ as $n \to \infty$. This fact and the continuity of h_f in $H_0^1(D)$ with respect to $f \in \mathbb{L}^2(D)$ yield that $\inf_{\substack{f\in C_0^\infty(D) \\ \text{s.t. } h_f\in\mathcal{O}}} \widetilde{I}(h_f) = \inf_{h\in\mathcal{O}} \widetilde{I}(h)$ for every open set $\mathcal{O} \subset \mathbb{L}^2(D)$, which completes the proof of the LDP lower bound. Q.E.D.

§5. Proof of Lemma 4.1: Convergence of Average Profiles

Reduction to two lemmas (Lemmas 5.2 and 5.3).

In this section we shall prove Lemma 4.1. The following lemma follows from (3.4) (cf. [13, Lemma 3.7]).

Lemma 5.1. *Let $\{h_n\}_{n\geq 1}$ be a sequence of $H_0^1(D)$ and define $\widetilde{\Sigma}_f(h)$ $= \widetilde{\Sigma}(h) - \langle h, f\rangle$. If $\lim_{n\to\infty} \widetilde{\Sigma}_f(h_n) = \inf_{H_0^1(D)} \widetilde{\Sigma}_f$, then $h_n \to h_f$ in $H_0^1(D)$ as $n \to \infty$.*

Also by (3.4), we have

$$\begin{aligned}\widetilde{\Sigma}_f(q) - \widetilde{\Sigma}_f(\tilde{h}^{\psi}_{N,f}) & \\ \geq \int_D (\nabla q(\theta) - \nabla \tilde{h}^{\psi}_{N,f}(\theta)) \cdot (\nabla\sigma)(\nabla \tilde{h}^{\psi}_{N,f}(\theta) + \nabla g(\theta))d\theta & \\ - \int_D (q(\theta) - \tilde{h}^{\psi}_{N,f}(\theta)) f(\theta) d\theta, &\end{aligned}$$

for every $q \in C_0^\infty(D)$. Once we can prove that the right hand side goes to 0 as $N \to \infty$ for every $q \in C_0^\infty(D)$, we have $\lim_{N\to\infty} \widetilde{\Sigma}_f(\tilde{h}^{\psi}_{N,f}) = \inf_{H_0^1(D)} \widetilde{\Sigma}_f$. This combined with Lemma 5.1 completes the proof of Lemma 4.1. Hence, all we have to prove are the following two lemmas.

Lemma 5.2. *For every $q \in C_0^\infty(D)$,*

$$\lim_{N\to\infty} \int_D \nabla q(\theta) \cdot (\nabla\sigma)(\nabla \tilde{h}^{\psi}_{N,f}(\theta) + \nabla g(\theta))d\theta = \int_D q(\theta) f(\theta) d\theta.$$

Lemma 5.3.

$$\begin{aligned}\lim_{N\to\infty} \Big\{ \int_D \nabla \tilde{h}^{\psi}_{N,f}(\theta) \cdot (\nabla\sigma)(\nabla \tilde{h}^{\psi}_{N,f}(\theta) + \nabla g(\theta))d\theta & \\ - \int_D \tilde{h}^{\psi}_{N,f}(\theta) f(\theta) d\theta \Big\} = 0. &\end{aligned}$$

For the proof of Lemmas 5.2 and 5.3, we prepare several lemmas.

A priori bounds.

Lemma 5.4. *There exists some $p \in (2, p_0)$ such that*

$$\sup_{N\geq 1} \|\nabla \tilde{h}^{\psi}_{N,f}\|_{\mathbb{L}^p(D)} < \infty \quad \textit{and} \quad \sup_{N\geq 1} \|\nabla \bar{h}^{\psi}_{N,f}\|_{\mathbb{L}^p(D)} < \infty,$$

where $p_0 > 2$ is the constant appearing in the condition $(\psi 2)$.

Proof. We first prove the uniform $\mathbb{L}^p$ estimate for $\nabla \tilde{h}^{\psi}_{N,f}$. It is easy to see that

(5.1)

$$\begin{aligned} & V'\big(\nabla_j \phi(x) + \nabla_j(\xi \vee \psi)(x)\big) \\ & \quad - V'\big(\nabla_j \phi(x) - E^{\tilde{\mu}^{\psi}_{N,f}}[\nabla_j \phi(x)] + \nabla_j(\xi \vee \psi)(x)\big) \\ & = E^{\tilde{\mu}^{\psi}_{N,f}}[\nabla_j \phi(x)] \\ & \quad \times \int_0^1 V''\big(\nabla_j \phi(x) - (1-t)E^{\tilde{\mu}^{\psi}_{N,f}}[\nabla_j \phi(x)] + \nabla_j(\xi \vee \psi)(x)\big)dt, \end{aligned}$$

for every $1 \le j \le d$ and $x \in D_N$. For $x \in D_N$, define $A_N(x) = \{A_{N,i,j}(x)\}_{1\le i,j\le d}$ and $a_N(x) = \{a_{N,j}(x)\}_{1\le j\le d}$ by

$$\begin{aligned}
A_{N,j,j}(x) &= E^{\tilde\mu^\psi_{N,f}}\Big[\int_0^1 V''\big(\nabla_j\phi(x) - (1-t)E^{\tilde\mu^\psi_{N,f}}[\nabla_j\phi(x)] \\
&\qquad\qquad\qquad\qquad + \nabla_j(\xi\vee\psi)(x)\big)dt\Big], \\
A_{N,i,j}(x) &= 0 \quad \text{if } i \ne j, \\
a_{N,j}(x) &= E^{\tilde\mu^\psi_{N,f}}\Big[V'\big(\nabla_j\phi(x) - E^{\tilde\mu^\psi_{N,f}}[\nabla_j\phi(x)] + \nabla_j(\xi\vee\psi)(x)\big)\Big],
\end{aligned}$$

respectively. Then, taking $\mathrm{div}_N\{E^{\tilde\mu^\psi_{N,f}}[\,\cdot\,]\}$ of the both sides of (5.1), we have

$$\begin{aligned}
&\mathrm{div}_N\{A_N(x)\nabla^N\tilde h^\psi_{N,f}(\frac{x}{N})\} \\
&= -\mathrm{div}_N\{a_N(x)\} + \mathrm{div}_N\{E^{\tilde\mu^\psi_{N,f}}\big[V'\big(\nabla\phi(x) + \nabla(\xi\vee\psi)(x)\big)\big]\},
\end{aligned}$$

where $\mathrm{div}_N\alpha$ is defined by $\mathrm{div}_N\alpha(x) = N\sum_{j=1}^d(\alpha_j(x) - \alpha_j(x-e_j))$ for a vector lattice field $\alpha(x) = \{\alpha_j(x)\}_{1\le j\le d}$, $x \in \mathbb{Z}^d$. By calculating $\frac{\partial H^\psi_{N,f}}{\partial\phi(x)}$ and taking its expectation under $\mu^\psi_{N,f}$ as in the proof of (1.55) of [13], we obtain

$$\mathrm{div}_N\{E^{\mu^\psi_{N,f}}[V'(\nabla\phi(x))]\} = -f(\frac{x}{N}), \tag{5.2}$$

for every $x \in D_N$. By (3.1), the change of variable yields

$$\mathrm{div}_N\{E^{\tilde\mu^\psi_{N,f}}\big[V'\big(\nabla\phi(x) + \nabla(\xi\vee\psi)(x)\big)\big]\} = -f(\frac{x}{N}).$$

Therefore, $\{\tilde h^\psi_{N,f}(\frac{x}{N})\}$ satisfies the following discrete elliptic equation:

$$\mathrm{div}_N\{A_N(x)\nabla^N\tilde h^\psi_{N,f}(\frac{x}{N})\} = -\mathrm{div}_N\{a_N(x)\} - f(\frac{x}{N}),$$

for every $x \in D_N$. However, by the assumption on V, $A_N(x)$ satisfies the uniform ellipticity condition $c_-I \le A_N(x) \le c_+I$ for every $x \in D_N$. Hence, by the proof of Lemma 3.4 of [13], we know that there exist some $p > 2$ and $C < \infty$ such that

$$\|\nabla\tilde h^\psi_{N,f}\|_{\mathbb{L}^p(D)} \le C\big(\|a_N\|_{\mathbb{L}^p(D)} + \|f\|_{\mathbb{L}^p(D)}\big),$$

uniformly in N. Note that $\tilde\mu^\psi_{N,f}$ is endowed with 0-boundary condition.

Now, since V' is linearly growing, using the change of variable again, we have that

$$|a_{N,j}(x)| \le C\big(E^{\mu^{\psi}_{N,f}}\big[|\nabla_j\phi(x) - E^{\mu^{\psi}_{N,f}}[\nabla_j\phi(x)]|\big] + |\nabla_j(\xi \vee \psi)(x)|\big),$$

for some $C > 0$. Then, $\sum_{x\in D_N} |a_N(x)|^{p_0} = O(N^d)$ as $N \to \infty$ follows from the Brascamp-Lieb inequality and the assumptions on ψ as in the proof of Lemma 4.3. This proves the uniform $\mathbb{L}^p$ estimate for $\nabla\tilde{h}^{\psi}_{N,f}$.

The uniform $\mathbb{L}^p$ estimate for $\nabla\bar{h}^{\psi}_{N,f}$ follows from that for $\nabla\tilde{h}^{\psi}_{N,f}$, the change of variable and the assumptions on ψ. Q.E.D.

Lemma 5.5. *For every $e \in \mathbb{Z}^d$ with $|e| = 1$, we have*

$$\lim_{N\to\infty} \frac{1}{N^d} \sum_{x\in D_N} \Big|\nabla^N\tilde{h}^{\psi}_{N,f}\Big(\frac{x+e}{N}\Big) - \nabla^N\tilde{h}^{\psi}_{N,f}\Big(\frac{x}{N}\Big)\Big|^2 = 0, \tag{5.3}$$

$$\lim_{N\to\infty} \frac{1}{N^d} \sum_{x\in D_N} \Big|\nabla^N\bar{h}^{\psi}_{N,f}\Big(\frac{x+e}{N}\Big) - \nabla^N\bar{h}^{\psi}_{N,f}\Big(\frac{x}{N}\Big)\Big|^2 = 0. \tag{5.4}$$

Proof. We first prove (5.4) by following the argument for the proof of Lemma 3.1 of [13]. Define $I_N = \{x \in D_N; \mathrm{dist}(x, \mathbb{Z}^d \setminus D_N) \ge 2\}$, then the sum $\sum_{x\in D_N}$ in (5.4) can be divided into $\sum_{x\in I_N}$ and $\sum_{x\in D_N\setminus I_N}$. The boundary term $\sum_{x\in D_N\setminus I_N}$ is $o(N^d)$ as $N \to \infty$ by Lemma 5.4 and Hölder's inequality. For the interior term $\sum_{x\in I_N}$, the entropy argument (cf. [13, Proposition 2.10 and Lemma 3.2]) yields the desired result. Note that the variance of the field $\phi(x)$ does not depend on the boundary condition ψ under the Gaussian measure $\mu^{\psi,*}_N$.

Next, we shall prove (5.3). The boundary term $\sum_{x\in D_N\setminus I_N}$ is $o(N^d)$ as before. For the interior term, by (3.1), the change of variable yields

$$\nabla^N_j\bar{h}^{\psi}_{N,f}\Big(\frac{x}{N}\Big) = \nabla^N_j\tilde{h}^{\psi}_{N,f}\Big(\frac{x}{N}\Big) + \nabla_j(\xi \vee \psi)(x), \tag{5.5}$$

for every $1 \le j \le d$ and $x \in D_N$. The contribution from the first term is $o(N^d)$ by (5.4), while that coming from the second term: $\sum_{x\in I_N} \big|\nabla\xi(x+e) - \nabla\xi(x)\big|^2$ is also $o(N^d)$. This is because $\xi(x) = Ng(\frac{x}{N})$ and we have $\nabla_j\xi(x+e) - \nabla_j\xi(x) = \frac{1}{N}\nabla^N_j\nabla^N_e g(\frac{x}{N})$ for every $1 \le j \le d$ and $x \in D_N$. Q.E.D.

Local equilibria.

Next, let

$$\begin{aligned} \mathcal{X} &= \{\eta \in \mathbb{R}^{(\mathbb{Z}^d)^*}; \eta = \nabla\phi \text{ for some } \phi \in \mathbb{R}^{\mathbb{Z}^d}\}, \\ \mathcal{X}_r &= \{\eta \in \mathcal{X}; \sum_{b\in(\mathbb{Z}^d)^*} |\eta(b)|^2 e^{-2r|x_b|} < \infty\}, \quad r > 0, \end{aligned}$$

and define $Q_N \in \mathcal{M}_+(D \times \mathcal{X})$ and $V_N \in \mathcal{M}_+(\mathbb{R}^d \times \mathcal{X})$ by

$$\begin{aligned} Q_N(d\theta d\eta) &= \frac{1}{N^d} \sum_{x\in D_N} \delta_{\frac{x}{N}}(d\theta)\mu_{N,f}^{\psi,\nabla} \circ \tau_x^{-1}(d\eta), \\ V_N(dvd\eta) &= \frac{1}{N^d} \sum_{x\in D_N} \delta_{\nabla^N \bar{h}_{N,f}^{\psi}(\frac{x}{N})}(dv)\mu_{N,f}^{\psi,\nabla} \circ \tau_x^{-1}(d\eta), \end{aligned}$$

where $\mathcal{M}_+(\mathcal{E})$ stands for the class of all non-negative measures on $\mathcal{E}$, $\mu_{N,f}^{\psi,\nabla}(d\eta)$ is the distribution of $\eta = \nabla\phi$ on $\mathcal{X}$ under $\mu_{N,f}^{\psi}$ and $\tau_x : \mathcal{X} \to \mathcal{X}$ denotes the shift on $\mathbb{Z}^d$ defined by $(\tau_x\eta)(b) = \eta(b - x)$ for $b \in (\mathbb{Z}^d)^*$. We regard $\mu_{N,f}^{\psi} \in \mathcal{P}(\mathbb{R}^{\mathbb{Z}^d})$ by considering $\phi(x) = \psi(x)(= g(\frac{x}{N}))$ for $x \in \mathbb{Z}^d \setminus D_N$. We denote by $\mu_v^{\nabla}(d\eta)$, $v = (v_i)_{1\le i\le d} \in \mathbb{R}^d$ the unique $\nabla\phi$-Gibbs measure on $\mathcal{X}$ which is translation invariant, ergodic and satisfies $E^{\mu_v^{\nabla}}[\eta(b)^2] < \infty$ for every $b \in (\mathbb{Z}^d)^*$ and $E^{\mu_v^{\nabla}}[\eta(e_i)] = v_i$ for every $1 \le i \le d$ (cf. [18, Section 3]).

In a similar way to the proof of Lemma 4.3 of [13], we can prove the following lemma. Note again that the variance does not depend on the boundary condition ψ under the Gaussian measure $\mu_N^{\psi,*}$. The proof is omitted.

Lemma 5.6. *For each $r > 0$ both the families of measures $\{Q_N\}$ on $D\times\mathcal{X}_r$ and $\{V_N\}$ on $\mathbb{R}^d\times\mathcal{X}_r$ are tight. Moreover, for every limit point Q of $\{Q_N\}$, there exists $\nu_Q \in \mathcal{M}_+(D \times \mathbb{R}^d)$ such that Q is represented as*

$$Q(d\theta d\eta) = \int_{\mathbb{R}^d} \nu_Q(d\theta dv)\mu_v^{\nabla}(d\eta).$$

Similarly, for each limit point V of $\{V_N\}$, there exists $\nu_V \in \mathcal{M}_+(\mathbb{R}^d\times\mathbb{R}^d)$ such that V is represented as

$$V(dvd\eta) = \int_{\mathbb{R}^d} \nu_V(dvdu)\mu_u^{\nabla}(d\eta).$$

Now by Lemma 5.4, along some subsequence, $\{\nabla \tilde{h}^{\psi}_{N,f}(\theta)\}_N$ generates the family of Young measures $\tilde{\nu}(\theta, dv) \in \mathcal{P}(\mathbb{R}^d)$ i.e. it holds that

$$\lim_{N\to\infty} \int_D q(\theta)G(\nabla \tilde{h}^{\psi}_{N,f}(\theta))d\theta = \int_{D\times\mathbb{R}^d} q(\theta)G(v)\tilde{\nu}(\theta, dv)d\theta. \tag{5.6}$$

for every $q \in \mathbb{L}^\infty(D)$ and $G \in C_0(\mathbb{R}^d)$ (cf. [13, Section 4.3], [3]). Then, the following lemma holds.

Lemma 5.7. *If the subsequence $\{N\}$ is commonly taken, the limits ν_Q and ν_V which appear in Lemma 5.6 can be represented as*

$$\nu_Q(d\theta dv) = \tilde{\nu}(\theta, dv - \nabla g(\theta))d\theta, \tag{5.7}$$

and

$$\nu_V(dvdu) = \delta_v(du) \int_D \tilde{\nu}(\theta, dv - \nabla g(\theta))d\theta. \tag{5.8}$$

Proof. By following the argument in the proof of Lemma 4.4 of [13], we shall only prove (5.7). The second equality (5.8) can be proved in a similar manner. For (5.7), it is enough to show that

$$\int_{D\times\mathbb{R}^d} q(\theta)G(v)\nu_Q(d\theta dv) = \int_{D\times\mathbb{R}^d} q(\theta)G(v + \nabla g(\theta))\tilde{\nu}(\theta, dv)d\theta \tag{5.9}$$

for every $q \in C_0^\infty(D)$ and $G \in C_b^1(\mathbb{R}^d)$. In fact, since the ergodicity of μ_v^∇ implies

$$G(v) = \lim_{l\to\infty} E^{\mu_v^\nabla}[G(\mathrm{Av}_l\eta)],$$

where $\mathrm{Av}_l\eta = \frac{1}{(2l+1)^d}\sum_{x\in B_l}\eta(x) \in \mathbb{R}^d$, $B_l = [-l,\ l]^d \cap \mathbb{Z}^d$, we have by Lemma 5.6,

$$\begin{aligned}&\int_{D\times\mathbb{R}^d} q(\theta)G(v)\nu_Q(d\theta dv)\\ &\quad= \lim_{l\to\infty}\lim_{N\to\infty}\frac{1}{N^d}\sum_{x\in D_N} q(\frac{x}{N})E^{\mu^{\psi,\nabla}_{N,f}\circ\tau_x^{-1}}[G(\mathrm{Av}_l\eta)].\end{aligned}$$

If one can replace $E^{\mu^{\psi,\nabla}_{N,f}\circ\tau_x^{-1}}[G(\mathrm{Av}_l\eta)]$ with $G(\nabla^N\tilde{h}^{\psi}_{N,f}(\frac{x}{N}) + \nabla^N g(\frac{x}{N}))$, then the right hand side is equal to

$$\begin{aligned}&\lim_{N\to\infty}\frac{1}{N^d}\sum_{x\in D_N} q(\frac{x}{N})G(\nabla^N\tilde{h}^{\psi}_{N,f}(\frac{x}{N}) + \nabla^N g(\frac{x}{N}))\\ &\quad= \int_{D\times\mathbb{R}^d} q(\theta)G(v + \nabla g(\theta))\tilde{\nu}(\theta, dv)d\theta,\end{aligned}$$

which implies (5.9). The last equality follows from Proposition 4.2 of [13], Lemma 5.5 and the fact that the equation (5.6) holds for $G = G(v + \nabla g(\theta))$ instead of $G = G(v)$ by p.213 remark 3 of [3].

For the replacement above, we have

$$\begin{aligned}
&\Big| \frac{1}{N^d} \sum_{x\in D_N} q(\frac{x}{N}) E^{\mu^{\psi,\nabla}_{N,f}\circ\tau_x^{-1}}[G(\mathrm{Av}_l\eta)] \\
&\qquad - \frac{1}{N^d} \sum_{x\in D_N} q(\frac{x}{N}) G(\nabla^N \tilde{h}^{\psi}_{N,f}(\frac{x}{N}) + \nabla^N g(\frac{x}{N})) \Big| \\
&\le S_1 + S_2 + S_3,
\end{aligned}$$

where

$$\begin{aligned}
S_1 &= \Big| \frac{1}{N^d} \sum_{x\in D_N} q(\frac{x}{N}) \big\{ E^{\mu^{\psi,\nabla}_{N,f}\circ\tau_x^{-1}}[G(\mathrm{Av}_l\eta)] - G\big(E^{\mu^{\psi,\nabla}_{N,f}\circ\tau_x^{-1}}[\mathrm{Av}_l\eta]\big)\big\} \Big|, \\
S_2 &= \Big| \frac{1}{N^d} \sum_{x\in D_N} q(\frac{x}{N}) \big\{ G\big(E^{\mu^{\psi,\nabla}_{N,f}\circ\tau_x^{-1}}[\mathrm{Av}_l\eta]\big) - G\big(\nabla^N \bar{h}^{\psi}_{N,f}(\frac{x}{N})\big)\big\} \Big|, \\
S_3 &= \Big| \frac{1}{N^d} \sum_{x\in D_N} q(\frac{x}{N}) \big\{ G\big(\nabla^N \bar{h}^{\psi}_{N,f}(\frac{x}{N})\big) \\
&\qquad - G\big(\nabla^N \tilde{h}^{\psi}_{N,f}(\frac{x}{N}) + \nabla^N g(\frac{x}{N})\big)\big\} \Big| .
\end{aligned}$$

In a similar way to the proof of Lemma 4.4 of [13], we can prove that $S_1, S_2 \to 0$ as $N \to \infty$, $l \to \infty$. Also by (5.5),

$$\begin{aligned}
S_3 &= \Big| \frac{1}{N^d} \sum_{x\in D_N} q(\frac{x}{N}) \big\{ G\big(\nabla^N \tilde{h}^{\psi}_{N,f}(\frac{x}{N}) + \nabla(\xi \vee \psi)(x)\big) \\
&\qquad - G\big(\nabla^N \tilde{h}^{\psi}_{N,f}(\frac{x}{N}) + \nabla\xi(x)\big)\big\} \Big| \\
&\le \frac{1}{N^d} \sum_{x\in\partial^- D_N} \|q\|_\infty \|\nabla G\|_\infty |\nabla(\xi\vee\psi)(x) - \nabla\xi(x)|,
\end{aligned}$$

where $\partial^- D_N = \{x \in D_N; \mathrm{dist}(x, \mathbb{Z}^d \setminus D_N) = 1\}$. This goes to 0 as $N \to \infty$ by the assumptions on ψ. Q.E.D.

Proof of Lemmas 5.2 and 5.3.

We are now in the position to prove Lemmas 5.2 and 5.3.

Proof of Lemma 5.2. For every $q \in C_0^\infty(D)$, by (5.2) and summation by parts, we have

$$\begin{aligned}\int_D q(\theta)f(\theta)d\theta &= \lim_{N\to\infty}\frac{1}{N^d}\sum_{x\in D_N} q(\frac{x}{N})f(\frac{x}{N}) \\ &= \lim_{N\to\infty}\frac{1}{N^d}\sum_{x\in D_N} \nabla^N q(\frac{x}{N})\cdot E^{\mu_{N,f}^{\psi}}[V'(\nabla\phi(x))].\end{aligned}$$

Now by the definition of Q_N, Lemmas 5.6, 5.7 and the property of the surface tension $\frac{\partial\sigma}{\partial v_i}(v) = E^{\mu_v^\nabla}[V'(\nabla_i\phi(0))]$ for every $1 \le i \le d$ (cf. [18, Theorem 3.4 (iii)]), we obtain

$$\begin{aligned}\int_D q(\theta)f(\theta)d\theta &= \int_{D\times\mathcal{X}} \nabla q(\theta)\cdot E^{\mu_v^\nabla}[V'(\nabla\phi(0))]\nu_Q(d\theta dv) \\ &= \int_{D\times\mathbb{R}^d} \nabla q(\theta)\cdot(\nabla\sigma)(v+\nabla g(\theta))\tilde{\nu}(\theta,dv)d\theta \\ &= \lim_{N\to\infty}\int_D \nabla q(\theta)\cdot(\nabla\sigma)(\nabla\tilde{h}_{N,f}^{\psi}(\theta)+\nabla g(\theta))d\theta,\end{aligned}$$

Note that we can apply (5.6) for $G = G(v,\theta) = (\nabla\sigma)(v+\nabla g(\theta))$ instead of $G = G(v)$ by p.213 remark 3 of [3] and the property of the surface tension $|(\nabla\sigma)(u)| \le c(1+|u|)$ (cf. [18, Theorem 3.4 (v)]).

Q.E.D.

Proof of Lemma 5.3. By (5.2), summation by parts and (5.5), we have

$$\begin{aligned}&\lim_{N\to\infty}\int_D \tilde{h}_{N,f}^{\psi}(\theta)f(\theta)d\theta \\ &\quad= \lim_{N\to\infty}\frac{1}{N^d}\sum_{x\in D_N}\nabla^N\tilde{h}_{N,f}^{\psi}(\frac{x}{N})\cdot E^{\mu_{N,f}^{\psi}}[V'(\nabla\phi(x))] \\ &\quad= \lim_{N\to\infty}\frac{1}{N^d}\sum_{x\in D_N}\nabla^N\bar{h}_{N,f}^{\psi}(\frac{x}{N})\cdot E^{\mu_{N,f}^{\psi}}[V'(\nabla\phi(x))] \\ &\qquad - \lim_{N\to\infty}\frac{1}{N^d}\sum_{x\in D_N}\nabla(\xi\vee\psi)(x)\cdot E^{\mu_{N,f}^{\psi}}[V'(\nabla\phi(x))] \\ &\quad\equiv S_1 - S_2.\end{aligned}$$

Now, by the assumptions on V and ψ, it is easy to see that

$$S_2 = \lim_{N\to\infty}\frac{1}{N^d}\sum_{x\in D_N}\nabla^N g(\frac{x}{N})\cdot E^{\mu_{N,f}^{\psi}}[V'(\nabla\phi(x))],$$

since $\xi(x) = Ng(\frac{x}{N})$. Hence, by the proof of Lemma 5.2, we obtain

$$S_2 = \lim_{N\to\infty} \int_D \nabla g(\theta) \cdot (\nabla\sigma)(\nabla\tilde{h}^{\psi}_{N,f}(\theta) + \nabla g(\theta))d\theta.$$

Also, by Lemmas 5.6, 5.7 and the property of the surface tension σ, in a similar way to the proof of Lemma 5.2 we can prove that

$$S_1 = \lim_{N\to\infty} \int_D (\nabla\tilde{h}^{\psi}_{N,f}(\theta) + \nabla g(\theta)) \cdot (\nabla\sigma)(\nabla\tilde{h}^{\psi}_{N,f}(\theta) + \nabla g(\theta))d\theta.$$

Therefore, the proof is completed. Q.E.D.

§6. Proof of Theorem 2.2: LDP for δ-Pinning

Schilder's theorem.

Throughout this section, we assume that $d = 1$ and $V(\eta) = \frac{1}{2}\eta^2$. We first notice that the large deviation principle holds for $\{h^N(\theta); \theta \in D\}$ under $\mu^{a,b}_N$ on $W_{a,b}(D)$. Recall that the space $W_{a,b}(D)$ is endowed with the topology determined by the sup-norm.

Lemma 6.1. *For the family of distributions on the space $W_{a,b}(D)$ under $\mu^{a,b}_N$ of $\{h^N(\theta); \theta \in D\}$, the large deviation principle holds with a rate functional $I^{a,b}(h) := \Sigma(h) - \frac{1}{2}(b-a)^2$ where $\Sigma(h) = \frac{1}{2}\int_0^1 (h')^2(\theta)d\theta$.*

Proof. Let $w = \{w(x); x \in [0, N]\}$ be the one-dimensional standard Brownian motion starting at 0 and set $\bar{h}^N(\theta) := w(N\theta)/N, \theta \in [0,1]$. Then, by Schilder's theorem (see, e.g., Theorem 5.1 of [25]), the large deviation principle holds for $\{\bar{h}^N\}_N$ on $W_0 = \{h \in C([0,1];\mathbb{R}); h(0) = 0\}$ with the rate function $\Sigma(h)$. Define $\phi = \{\phi(x); x \in [0,N]\}$ from w as $\phi(x) = w(x) - xw(N)/N + (N-x)a + xb$. Then, $\{\phi(x); x \in D_N\}$ is $\mu^{a,b}_N$-distributed. Set $\tilde{h}^N(\theta) = \phi(N\theta)/N, \theta \in [0,1]$, and consider a mapping $\Phi : \bar{h} \in W_0 \mapsto \tilde{h} \in W_{a,b}(D)$ defined by

$$\Phi(\bar{h})(\theta) = \bar{h}(\theta) - \theta\bar{h}(1) + (1-\theta)a + \theta b.$$

Then, Φ is continuous and $\tilde{h}^N = \Phi(\bar{h}^N)$ holds. Therefore, by the contraction principle, the large deviation principle holds for $\{\tilde{h}^N\}_N$ with the rate functional $\tilde{\Sigma}(h) = \inf_{\bar{h}\in W_0:\Phi(\bar{h})=h} \Sigma(\bar{h})$, which coincides with $I^{a,b}(h)$. The proof of lemma is completed by showing a super exponential estimate for the difference between h^N and $\tilde{h}^N$ as in p.17 of [25]: For every

$\delta > 0$,

$$P\left(\|h^N - \tilde{h}^N\|_\infty \geq \delta\right) = \exp\left[-\frac{N^2\delta^2}{8} + o(N^2)\right],$$

as $N \to \infty$. Q.E.D.

Proof of Theorem 2.2.

Proof of Theorem 2.2. Step1 (*lower bound*). Let $\delta > 0$ and $g \in W_{a,b}(D)$ which satisfies the condition:

$$\text{(6.1)}\quad \begin{aligned} &|\{\theta \in D; g(\theta) = 0\}| > 0 \text{ and there exist disjoint intervals} \\ &\{I^j\}_{1 \leq j \leq K}, K < \infty \text{ such that } |\{\theta \in D; g(\theta) = 0\}| = \sum\nolimits_{j=1}^{K} |I^j| \\ &\text{and } g(\theta) = 0 \text{ if } \theta \in \bigcup\nolimits_{j=1}^{K} I^j, \end{aligned}$$

be fixed. Then, one can decompose $D \setminus \bigcup_{j=1}^{K} I^j = \bigcup_{j=1}^{K+1} L^j$ with disjoint intervals $\{L^j\}_{1 \leq j \leq K+1}$. We define $I_N^j = NI^j \cap \mathbb{Z}, L_N^j = NL^j \cap \mathbb{Z}, I_N = \bigcup_{j=1}^{K} I_N^j$ and $L_N = \bigcup_{j=1}^{K+1} L_N^j$. By expanding the product $\prod_{x \in D_N}(e^J \delta_0(d\phi(x)) + d\phi(x))$, we have

$$\begin{aligned} &\frac{Z_N^{a,b,J}}{Z_N^{a,b}} \mu_N^{a,b,J}(h^N \in B_\infty(g,\delta)) \\ &\quad = \sum_{\Lambda \subset D_N} e^{J|\Lambda^c|} \frac{Z_\Lambda^{a,b}}{Z_N^{a,b}} \mu_\Lambda^{a,b}(h^N \in B_\infty(g,\delta)) \\ &\quad \geq \sum_{L_N \subset \Lambda \subset D_N} e^{J|\Lambda^c|} \frac{Z_\Lambda^{a,b}}{Z_N^{a,b}} \mu_\Lambda^{a,b}(h^N \in B_\infty(g,\delta)) \\ &\quad = \sum_{A \subset I_N} e^{J|I_N \setminus A|} \frac{Z_{L_N \cup A}^{a,b}}{Z_N^{a,b}} \mu_{L_N \cup A}^{a,b}(h^N \in B_\infty(g,\delta)), \end{aligned}$$

where $B_\infty(g,\delta) = \{h \in W_{a,b}(D); \|h - g\|_\infty < \delta\}$ and $\mu_\Lambda^{a,b}$ is defined by

$$\begin{aligned} \mu_\Lambda^{a,b}(d\phi) = &\frac{1}{Z_\Lambda^{a,b}} \exp\{-\frac{1}{2} \sum_{b \in \bar{\Lambda}^*} V(\nabla(\phi \vee \tilde{\psi})(b))\} \\ &\times \prod_{x \in \Lambda} d\phi(x) \prod_{x \in \overline{D_N} \setminus \Lambda} \delta_{\tilde{\psi}(x)}(d\phi(x)), \end{aligned}$$

and $\tilde{\psi}(x) = \psi(x)$ if $x \in \partial^+ D_N = \{0, N\}$ (i.e. $\tilde{\psi}(0) = aN, \tilde{\psi}(N) = bN$), $\tilde{\psi}(x) = 0$ if $x \in D_N \setminus \Lambda$. The constant $Z_\Lambda^{a,b}$ is for normalization.

Now, write $I_N \setminus A = \{x_1, x_2, \cdots, x_k\}$, $1 \le x_1 < x_2 < \cdots < x_k \le N-1$ and define $l_1 = [1, x_1 - 1] \cap \mathbb{Z}$, $l_2 = [x_1 + 1, x_2 - 1] \cap \mathbb{Z}$, $\cdots$, $l_k = [x_{k-1}+1, x_k - 1] \cap \mathbb{Z}$, $l_{k+1} = [x_k + 1, N-1] \cap \mathbb{Z}$. Then, $\bigcup_{j=1}^{k+1} l_j = L_N \cup A$ and by the Markov property of the ϕ-field, we have

$$\begin{aligned}\mu_{L_N \cup A}^{a,b}(h^N \in B_\infty(g, \delta)) &\ge \mu_{L_N \cup A}^{a,b}\Big(\max_{x \in \bigcup_{j=1}^{k+1} l_j} |\frac{1}{N}\phi(x) - g(\frac{x}{N})| < \frac{1}{2}\delta\Big) \\ &= \prod_{j=1}^{k+1} \mu_{l_j}^{a_j, b_j}\big(\max_{x \in l_j} |\frac{1}{N}\phi(x) - g(\frac{x}{N})| < \frac{1}{2}\delta\big),\end{aligned}$$

for N large enough, where $a_j = a$ if $j = 1$, $a_j = 0$ otherwise, $b_j = b$ if $j = k+1$, $b_j = 0$ otherwise. We define $\Gamma = \{1 \le j \le k+1; l_j \supset L_N^i$ for some $1 \le i \le K+1\}$ and $\Gamma^c = \{1 \le j \le k+1\} \setminus \Gamma$. If $j \in \Gamma^c$, since $g(\frac{x}{N}) = 0$ for each $x \in l_j$, we have

$$\begin{aligned}\mu_{l_j}^{a_j, b_j}\big(\max_{x \in l_j} |\frac{1}{N}\phi(x) - g(\frac{x}{N})| < \frac{1}{2}\delta\big) &= \mu_{l_j}^{a_j, b_j}\big(\max_{x \in l_j} |\frac{1}{N}\phi(x)| < \frac{1}{2}\delta\big) \\ &\ge 1 - \sum_{x \in l_j} \mu_{l_j}^{0,0}\big(|\phi(x)| \ge \frac{1}{2}\delta N\big).\end{aligned}$$

However, it is easy to see that

$$\mu_{l_j}^{0,0}\big(|\phi(x)| \ge \frac{1}{2}\delta N\big) \le \exp\Big\{-\frac{(\frac{1}{2}\delta N)^2}{\mathrm{Var}_{\mu_{l_j}^{0,0}}(\phi(x))}\Big\} \le \exp\{-C\delta^2 N\},$$

for some $C > 0$ and we obtain

$$\prod_{j \in \Gamma^c} \mu_{l_j}^{a_j, b_j}\big(\max_{x \in l_j} |\frac{1}{N}\phi(x) - g(\frac{x}{N})| < \frac{1}{2}\delta\big) \ge 1 - N\exp\{-C\delta^2 N\}.$$

Next, for every closed interval $F \equiv [x_F, y_F] \subset [0,1]$, define

$$\begin{aligned}&B_\infty(g, \delta; F) = \{h \in C(F; \mathbb{R}); \sup_{\theta \in F} |h(\theta) - g(\theta)| < \delta\}, \\ &W_{a,b}(F) = \{h \in C(F; \mathbb{R}); h(x_F) = a, h(y_F) = b\}, \\ &H_{a,b}^1(F) = \{h \in W_{a,b}(F); h \text{ is absolutely continuous}, h' \in \mathbb{L}^2(F)\}.\end{aligned}$$

We also define $\tilde{l}_j = \frac{l_j}{N} \subset [0,1]$ for $j \in \Gamma$. Then, by the LDP lower bound for $\mu_N^{a,b}$ (Lemma 6.1), we know that

$$\begin{aligned}\prod_{j\in\Gamma}&\mu_{l_j}^{a_j,b_j}\big(\max_{x\in l_j}|\frac{1}{N}\phi(x)-g(\frac{x}{N})|<\frac{1}{2}\delta\big)\\ &\ge \exp\Big\{-N\big(\sum_{j\in\Gamma}\inf_{h\in B_\infty(g,\frac{1}{2}\delta;\tilde{l}_j)} I_{\tilde{l}_j}^{a_j,b_j}(h)+\varepsilon\big)\Big\}\\ &\ge \exp\Big\{-N\big(\Sigma(g)-\frac{1}{2}(\frac{a^2}{|\tilde{l}_1|}+\frac{b^2}{|\tilde{l}_{k+1}|})+\varepsilon\big)\Big\},\end{aligned}$$

for every $\varepsilon > 0$ and N large enough, where

$$I_F^{a,b}(h) = \begin{cases}\Sigma_F(h) - \frac{(b-a)^2}{2|F|} & \text{if } h \in H^1_{a,b}(F),\\ +\infty & \text{otherwise,}\end{cases}$$

and $\Sigma_F(h) = \frac{1}{2}\int_F (h')^2(\theta)d\theta$ for closed interval $F \subset [0,1]$. Recall that $\Sigma_{[0,1]}(h)$ coincides with $\Sigma(h)$. Therefore, we obtain

$$\begin{aligned}&\mu_{L_N\cup A}^{a,b}(h^N \in B_\infty(g,\delta))\\ &\quad\ge \exp\Big\{-N\big(\Sigma(g)-\frac{1}{2}(\frac{a^2}{|\tilde{l}_1|}+\frac{b^2}{|\tilde{l}_{k+1}|})+\varepsilon\big)\Big\}(1-Ne^{-CN\delta^2}).\end{aligned}$$

Note that this estimate holds for every choice of $A \subset I_N$ and for every N large enough, since $|\Gamma| \le K+1$ is independent of N. Also, simple calculation yields that

$$\begin{aligned}Z_{L_N\cup A}^{a,b} &= Z_{L_N\cup A}^{0,0}\exp\Big\{-\frac{N}{2}(\frac{a^2}{|\tilde{l}_1|}+\frac{b^2}{|\tilde{l}_{k+1}|})\Big\},\\ Z_N^{a,b} &= Z_N^{0,0}\exp\Big\{-\frac{N}{2}(b-a)^2\Big\}.\end{aligned}$$

Hence we obtain

$$\begin{aligned}(6.2)\qquad &\frac{Z_N^{a,b,J}}{Z_N^{a,b}}\mu_N^{a,b,J}(h^N \in B_\infty(g,\delta))\\ &\ge \sum_{A\subset I_N} e^{J|I_N\setminus A|}\frac{Z_{L_N\cup A}^{0,0}}{Z_N^{0,0}}\exp\big\{-N\big(I^{a,b}(g)+2\varepsilon\big)\big\},\end{aligned}$$

for every $\varepsilon > 0$ and N large enough.

Now, we can exactly calculate that $Z_N^{0,0} = \frac{(\sqrt{2\pi})^{N-1}}{\sqrt{N}}$ and this shows

$$1 \le \frac{Z_{L_N \cup A}^{0,0}}{Z_{L_N}^{0,0} Z_A^{0,0}} \le e^{a_N}, \tag{6.3}$$

for every $A \subset I_N$, where $a_N = o(N)$. Note that L_N consists of finite number of disjoint intervals of size $O(N)$. By using (6.3), it is easy to see that

$$\frac{Z_{I_N}^{0,0,J}}{Z_{I_N}^{0,0}} e^{-a_N} \le \sum_{A \subset I_N} e^{J|I_N \setminus A|} \frac{Z_{L_N \cup A}^{0,0}}{Z_N^{0,0}} \le \frac{Z_{I_N}^{0,0,J}}{Z_{I_N}^{0,0}} e^{a_N}. \tag{6.4}$$

The sub-additivity argument (cf. [8, Section 4.3], [18, Appendix II]) and the fact that $\frac{|I_N|}{N} \to |\{\theta \in D; g(\theta) = 0\}|$ as $N \to \infty$ yield that the limit $\tau(J)$ in (2.7) exists and it holds that

$$\lim_{N\to\infty} \frac{1}{N} \log \frac{Z_{I_N}^{0,0,J}}{Z_{I_N}^{0,0}} = -\tau(J)|\{\theta \in D; g(\theta) = 0\}|. \tag{6.5}$$

Combining (6.4), (6.5) with (6.2), we obtain

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \log \frac{Z_N^{a,b,J}}{Z_N^{a,b}} \mu_N^{a,b,J}(h^N \in B_\infty(g,\delta)) & \\ \ge -I^{a,b}(g) - \tau(J)|\{\theta \in D; g(\theta) = 0\}| & \\ \equiv -I^{a,b;J}(g), & \end{aligned} \tag{6.6}$$

for every $g \in W_{a,b}(D)$ satisfying the condition (6.1) and $\delta > 0$. In the case that $|\{\theta \in D; g(\theta) = 0\}| = 0$, we have only to take the sum $A = I_N$ in (6.2) and the same inequality as above is obtained.

However, for every open set $\mathcal{O}$ of $W_{a,b}(D)$, we have that

$$\inf_{g \in \mathcal{O}:(6.1)'} I^{a,b;J}(g) = \inf_{h \in \mathcal{O}} I^{a,b;J}(h), \tag{6.7}$$

where (6.1)$'$ means the condition (6.1) or $|\{\theta \in D; g(\theta) = 0\}| = 0$. Indeed, since the left hand side of (6.7) is larger than or equal to the right hand side, we may prove the reverse inequality only. To this end, for every $\varepsilon > 0$, take $h \in \mathcal{O}$ such that $I^{a,b;J}(h) \le \inf_{\mathcal{O}} I^{a,b;J} + \varepsilon$; note that $h \in H_{a,b}^1(D)$. Since $\mathcal{O}$ is open, one can find $\delta > 0$ such that $B_\infty(h,\delta) \subset \mathcal{O}$. Taking $n \ge 1$ such that $|\theta_1 - \theta_2| \le 1/n$ implies $|h(\theta_1) - h(\theta_2)| < \delta$, divide the interval $[0,1] = \cup_{k=1}^n \mathcal{J}_k, \mathcal{J}_k = [(k-1)/n, k/n]$ and set $\mathcal{J} = \cup_k \mathcal{J}_k$, the union of $\mathcal{J}_k$'s on which $h(\theta) \ne 0$. We now define a function $g = g(\theta)$, first on $\mathcal{J}$, by $g(\theta) = h(\theta)$. On $\mathcal{J}^c$, starting at points in $\partial\mathcal{J}$, $g(\theta) = h(\theta)$ up to $\bar{\theta}$'s such that $h(\bar{\theta}) = 0$, and set $g \equiv 0$ otherwise. Then, $g \in B_\infty(h, \delta) \subset \mathcal{O}$, $I^{a,b;J}(g) \le I^{a,b;J}(h)$ and g satisfies the condition (6.1)$'$. This proves (6.7). Therefore, from (6.6) and (6.7), we have

$$\liminf_{N\to\infty} \frac{1}{N} \log \frac{Z_N^{a,b,J}}{Z_N^{a,b}} \mu_N^{a,b,J}(h^N \in \mathcal{O}) \ge -\inf_{h\in\mathcal{O}} I^{a,b;J}(h), \tag{6.8}$$

for every open set $\mathcal{O}$ of $W_{a,b}(D)$.

Step2 (*upper bound*). Let $\delta > 0$ and $g \in W_{a,b}(D)$ which satisfies the condition:

$$\begin{aligned}&\text{for every } \gamma > 0 \text{ small enough, there exist disjoint}\\ &\text{intervals } \{I^j(\gamma)\}_{1\le j\le K}, K < \infty \text{ such that}\\ &\{\theta \in D; |g(\theta)| \le \gamma\} = \bigcup\nolimits_{j=1}^{K} I^j(\gamma),\end{aligned} \tag{6.9}$$

be fixed. Then, one can write $\{\theta \in D; |g(\theta)| > \gamma\} = \bigcup_{j=1}^{K+1} L^j(\gamma)$ for disjoint intervals $\{L^j(\gamma)\}_{1\le j\le K+1}$. We define $I_N^j = NI^j(\delta) \cap \mathbb{Z}, L_N^j = NL^j(\delta) \cap \mathbb{Z}, I_N = \bigcup_{j=1}^{K} I_N^j$ and $L_N = \bigcup_{j=1}^{K+1} L_N^j$. Since $\mu_\Lambda^{a,b}(h^N \in B_\infty(g,\delta)) = 0$ for $\Lambda \subset D_N$ such that $\Lambda \not\supset L_N$, we have

$$\begin{aligned}&\frac{Z_N^{a,b,J}}{Z_N^{a,b}} \mu_N^{a,b,J}(h^N \in B_\infty(g,\delta))\\ &\quad = \sum_{L_N\subset\Lambda\subset D_N} e^{J|\Lambda^c|} \frac{Z_\Lambda^{a,b}}{Z_N^{a,b}} \mu_\Lambda^{a,b}(h^N \in B_\infty(g,\delta))\\ &\quad = \sum_{A\subset I_N} e^{J|I_N\setminus A|} \frac{Z_{L_N\cup A}^{a,b}}{Z_N^{a,b}} \mu_{L_N\cup A}^{a,b}(h^N \in B_\infty(g,\delta)).\end{aligned}$$

Now, let $I_N \setminus A = \{x_1, x_2, \cdots, x_k\}$, $1 \le x_1 < x_2 < \cdots < x_k \le N-1$ and define $l_1, l_2, \cdots, l_k, l_{k+1}$ and Γ in the same way as in the proof of lower bound. Then, by the Markov property of the ϕ-field and the LDP upper bound for $\mu_N^{a,b}$ (Lemma 6.1), we have

$$\begin{aligned}\mu_{L_N\cup A}^{a,b}(h^N \in B_\infty(g,\delta)) &\le \mu_{L_N\cup A}^{a,b}\Big(\max_{x\in\bigcup_{j=1}^{k+1} l_j} |\frac{1}{N}\phi(x) - g(\frac{x}{N})| < \delta\Big)\\ &\le \prod_{j\in\Gamma} \mu_{l_j}^{a_j,b_j}\Big(\max_{x\in l_j} |\frac{1}{N}\phi(x) - g(\frac{x}{N})| < \delta\Big)\end{aligned}$$

$$\leq \exp\Big\{-N\big(\sum_{j\in\Gamma} \inf_{h\in\bar{B}_\infty(g,2\delta;\tilde{l}_j)} I_{\tilde{l}_j}^{a_j,b_j}(h) - \varepsilon\big)\Big\}$$

$$\leq \exp\Big\{-N\big(\inf_{\substack{h\in\bar{B}_\infty(g,2\delta) \\ h(0)=a,h(1)=b}} \Sigma(h) - \frac{1}{2}\big(\frac{a^2}{|\tilde{l}_1|} + \frac{b^2}{|\tilde{l}_{k+1}|}\big) - \varepsilon\big)\Big\},$$

for every $\varepsilon > 0$ and N large enough. Then, in a similar way to the proof of lower bound, we can prove that

$$\begin{aligned} (6.10) \qquad & \limsup_{N\to\infty} \frac{1}{N} \log \frac{Z_N^{a,b,J}}{Z_N^{a,b}} \mu_N^{a,b,J}(h^N \in B_\infty(g,\delta)) \\ & \leq - \inf_{h\in\bar{B}_\infty(g,2\delta)} I^{a,b}(g) - \tau(J)|\{\theta \in D; |g(\theta)| \leq \delta\}|, \end{aligned}$$

for every $g \in W_{a,b}(D)$ satisfying the condition (6.9) and $\delta > 0$. Note that I_N is defined by $N\{\theta \in D; |g(\theta)| \leq \delta\} \cap \mathbb{Z}$ in this case.

By using (6.10), the right-continuity of $|\{\theta \in D; |g(\theta)| \leq \delta\}|$ in δ and the fact that the set of $g \in W_{a,b}(D)$ satisfying the condition (6.9) is dense in $W_{a,b}(D)$, the similar argument to the proof of the upper bound of Theorem 2.1 yields that for every $g \in W_{a,b}(D)$ and $\varepsilon > 0$, there exists some $\delta > 0$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \log \frac{Z_N^{a,b,J}}{Z_N^{a,b}} \mu_N^{a,b,J}(h^N \in B_\infty(g,\delta)) \leq -I^{a,b;J}(g) + \varepsilon.$$

Since $\mu_N^{a,b,J}$ can be written as the superposition of $\mu_\Lambda^{a,b}, \Lambda \subset D_N$, exponential tightness for $\mu_N^{a,b,J}$ follows from the similar argument as before and the standard argument yields

$$(6.11) \qquad \limsup_{N\to\infty} \frac{1}{N} \log \frac{Z_N^{a,b,J}}{Z_N^{a,b}} \mu_N^{a,b,J}(h^N \in \mathcal{C}) \leq - \inf_{h\in\mathcal{C}} I^{a,b;J}(h),$$

for every closed set $\mathcal{C}$ of $W_{a,b}(D)$. The lower and upper bounds (6.8) and (6.11) conclude the proof. Q.E.D.

Remark 6.1. *By the proof above and* [8, Lemma 2.3.1 (a)] (*note that the argument given there can be extended to all $d \geq 1$*), *we know that*

$$\frac{Z_N^{0,0,J}}{Z_N^{0,0}} = \sum_{\Lambda\subset D_N} e^{J|\Lambda^c|} \frac{Z_\Lambda^{0,0}}{Z_N^{0,0}} \geq \sum_{\Lambda\subset D_N} e^{J|\Lambda^c|} e^{-C|\Lambda^c|} = (1 + e^{J-C})^{|D_N|}$$

for some constant $C > 0$. Therefore, $\tau(J) < 0$ for every $J \in \mathbb{R}$.

Acknowledgement.

The authors thank G.S. Weiss for pointing out the references [1], [2] and [26] to them. They also thank H. Spohn for helpful discussions.

References

[1] H.W. Alt and L.A. Caffarelli, *Existence and regularity for a minimum problem with free boundary*, J. Reine Angew. Math., **325** (1981), pp. 105–144.

[2] H.W. Alt, L.A. Caffarelli and A. Friedman, *Variational problems with two phases and their free boundaries*, Trans. Amer. Math. Soc., **282** (1984), pp. 431–461.

[3] J.M. Ball, *A version of the fundamental theorem for Young measures*, in *PDEs and continuum models of phase transitions (Nice, 1988)*, Springer–Verlag, Lecture Notes in Physics, **344** (1989), pp. 207–215.

[4] G. Ben Arous and J.-D. Deuschel, *The construction of the $d+1$-dimensional Gaussian droplet*, Commun. Math. Phys., **179** (1996), pp. 467–488.

[5] T. Bodineau, D. Ioffe and Y. Velenik, *Rigorous probabilistic analysis of equilibrium crystal shapes*, J. Math. Phys., **41** (2000), pp. 1033–1098.

[6] E. Bolthausen, J.-D. Deuschel and O. Zeitouni, *Entropic repulsion of the lattice free field*, Commun. Math. Phys., **170** (1995), pp. 417–443.

[7] E. Bolthausen, J.-D. Deuschel and O. Zeitouni, *Absence of a wetting transition for a pinned harmonic crystal in dimensions three and larger*, J. Math. Phys., **41** (2000), pp. 1211–1223.

[8] E. Bolthausen and D. Ioffe, *Harmonic crystal on the wall: a microscopic approach*, Commun. Math. Phys., **187** (1997), pp. 523–566.

[9] E. Bolthausen and Y. Velenik, *Critical behavior of the massless free field at the depinning transition*, Commun. Math. Phys., **223** (2001), pp. 161–203.

[10] A. Caputo and Y. Velenik, *A note on wetting transition for gradient fields*, Stoch. Proc. Appl., **87** (2000), pp. 107–113.

[11] J. De Coninck, F. Dunlop and V. Rivasseau, *On the microscopic validity of the Wulff construction and of the generalized Young equation*, Commun. Math. Phys., **121** (1989), pp. 401–419.

[12] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2nd edition, Springer-Verlag, Applications of Mathematics **38**, 1998.

[13] J.-D. Deuschel, G. Giacomin and D. Ioffe, *Large deviations and concentration properties for $\nabla\varphi$ interface models*, Probab. Theory Relat. Fields, **117** (2000), pp. 49–111.

[14] J.-D. Deuschel and D.W. Stroock, *Large Deviations*, Academic Press, Vol. 137 in Pure and Applied Mathematics, 1989.

[15] J.-D. DEUSCHEL AND Y. VELENIK, *Non-Gaussian surface pinned by a weak potential*, Probab. Theory Relat. Fields, **116** (2000), pp. 359–377.

[16] F. DUNLOP, J. MAGNEN, V. RIVASSEAU AND P. ROCHE, *Pinning of an interface by a weak potential*, J. Statis. Phys., **66** (1992), pp. 71–98.

[17] M.E. FISHER, *Walks, walls, wetting, and melting*, J. Statis. Phys., **34** (1984), pp. 667–729.

[18] T. FUNAKI AND H. SPOHN, *Motion by mean curvature from the Ginzburg-Landau $\nabla\phi$ interface models*, Commun. Math. Phys., **185** (1997), pp. 1–36.

[19] G. GIACOMIN, *Anharmonic lattices, random walks and random interfaces*, Recent Research Developments in Statistical Physics, Transworld Research Network, **1** (2000), pp. 97–118.

[20] D. IOFFE AND Y. VELENIK, *A note on the decay of correlations under δ-pinning*, Probab. Theory Relat. Fields, **116** (2000), pp. 379–389.

[21] Y. ISOZAKI AND N. YOSHIDA, *Weakly pinned random walk on the wall: pathwise descriptions of the phase transition*, Stoch. Proc. Appl., **96** (2001), pp. 261–284.

[22] C. KIPNIS AND C. LANDIM, *Scaling Limits of Interacting Particle Systems*, Springer-Verlag, Grundlehern der math. Wiss. **320**, 1999.

[23] T. NISHIKAWA, *Hydrodynamic limit for the Ginzburg-Landau $\nabla\phi$ interface model with a boundary condition*, to appear in Probab. Theory Rel. Fields, 2003.

[24] C.-E. PFISTER AND Y. VELENIK, *Interface, surface tension and reentrant pinning transition in the 2D Ising model*, Commun. Math. Phys., **204** (1999), pp. 269–312.

[25] S.R.S. VARADHAN, *Large Deviations and Applications*, SIAM, 1984.

[26] G.S. WEISS, *A free boundary problem for non-radial-symmetric quasilinear elliptic equations*, Adv. Math. Sci. Appl., **5** (1995), pp. 497–555.

[27] J. WLOKA, *Partial differential equations*, Cambridge Univ. Press, 1987.

Tadahisa Funaki
Graduate School of Mathematical Sciences
The University of Tokyo
3-8-1 Komaba, Meguro-ku
Tokyo 153-8914
JAPAN
E-mail address: `funaki@ms.u-tokyo.ac.jp`

Hironobu Sakagawa
Department of Mathematics
Faculty of Science and Technology
Keio University
3-14-1 Hiyoshi, Kouhoku-ku
Yokohama 223-8522
JAPAN
E-mail address: `sakagawa@math.keio.ac.jp`

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 213–232

A PDE Approach for Motion of Phase-Boundaries by a Singular Interfacial Energy

Mi-Ho Giga and Yoshikazu Giga

§1. Introduction

This is a review paper on geometric motions of phase boundaries like crystal surfaces when the interfacial energy is very singular. Such motions arise in nonequilibrium problem at low temperature. Our purpose is to review a macroscopic approach describing the phenomena by a partial differential equation (PDE) with singular diffusivity. Because of nonlocal effect of singular diffusivity the notion of solution itself is unclear. In this paper we focus the problem whether a solution of approximate parabolic problem converges to a 'solution' of PDE with the singular diffusivity. We do not intend to exhaust the references.

The equilibrium of a crystal shape is often explained as a solution of an anisotropic isoperimetric problem. The problem is described as follows. Let γ be a continuous function on $\mathbf{R}^n$ which is positively homogeneous of degree one, *i.e.*, $\gamma(\lambda p) = \lambda\gamma(p)$ for all $p \in \mathbf{R}^n$, $\lambda > 0$. Assume that $\gamma(p) > 0$ for $p \neq 0$. For an oriented hypersurface S with the orientation $\mathbf{n}$ (a unit normal vector field) in $\mathbf{R}^n$ let $I(S)$ be defined by

$$I(S) = \int_S \gamma(\mathbf{n})dS, \tag{1.1}$$

where dS denotes the surface element. The quantity $I(S)$ is called the interfacial energy and γ is called the *interfacial energy density* (depending upon the temperature τ through the structure of the crystal surface S).

Received February 4, 2003.
Revised May 8, 2003.
Partly Supported by the Grant-in-Aid for Scientific Research, No.12874024, No.14204011, the Japan Society for the Promotion of Science.

Anisotropic isoperimetric problem. Minimize $I(\partial D)$ among all open sets $D(\subset \mathbf{R}^n)$ with a fixed volume.

The answer is by now well-known. The interfacial energy $I(\partial D)$ is minimized if and only if D is a dilation of the Wulff shape $\mathcal{W}_\gamma$ defined by

$$\mathcal{W}_\gamma = \bigcap_{|\mathbf{m}|=1} \{x \in \mathbf{R}^n,\ x \cdot \mathbf{m} \leq \gamma(\mathbf{m})\}. \tag{1.2}$$

The reader is referred to [5], [34], [30] and references cited there for the development of the theory. This Wulff shape is considered as a shape of a crystal in an equilibrium state. The Wulff shape is always convex and closed, and its support function

$$\hat{\gamma}(p) = \sup\{p \cdot x; \quad x \in \mathcal{W}_\gamma\} \quad \text{for} \quad p \in \mathbf{R}^n$$

is the convex hull of γ. At low temperature the Wulff shape has a flat portion called a facet. In this case $\hat{\gamma}$ is not C^1 at the direction corresponding to the normal of the facet. We rather consider $\hat{\gamma}$ instead of γ, so we assume that γ is always convex, since $\mathcal{W}_\gamma = \mathcal{W}_{\hat{\gamma}}$.

The first variation of $I(\partial D)$ with respect to a variation of the volume of D is

$$H_\gamma = -\mathrm{div}_S((\nabla_p \gamma)(\mathbf{n})) \quad \text{with} \quad S = \partial D. \tag{1.3}$$

This is called the *weighted mean curvature* of S in the direction of $\mathbf{n}$, which is the unit outer normal vector field of ∂D. Here div_S denotes the surface divergence. If $\gamma(p) = |p|$, then H_γ is the usual mean curvature H. (We use the convention that H is the sum of all principal curvatures instead of its average.) The weighted mean curvature of $\partial\mathcal{W}_\gamma$ always equals $-(n-1)$ so $\mathcal{W}_\gamma$ substitutes the role of a unit sphere for usual mean curvature. If $H_\gamma = -(n-1)$, we expect that D is the Wulff shape but we do not know in general whether such a conjecture is settled except the case $\gamma(p) = |p|$ which is proposed by H. Hopf and solved affirmatively by [1]. If γ is not C^1 so that $\mathcal{W}_\gamma$ has a facet, we observe that H_γ should be a nonlocal quantity since otherwise $H_\gamma = 0$ on such a facet, since the second fundamental form equals zero on a facet.

In nonequilibrium state a phase-boundary such as a crystal surface moves. Its motion is often considered as a result of thermodynamical driving forces—variation of the free energy. A typical example is the mean curvature flow equation

$$V = H \quad \text{on} \quad S_t \tag{1.4}$$

proposed by Mullins [40] to describe the motion of the antiphase boundaries of grains in material sciences. Here, V denotes the normal velocity of evolving (embedded) (hyper) surface $\{S_t\}$ in the direction of $\mathbf{n}$; the parameter t denotes the time variable. The mean curvature is considered as the first variation of the area. To study a crystal growth problem anisotropic effect must be taken into account. For example one consider

$$V = M(\mathbf{n})(H_\gamma + C) \quad \text{on} \quad S_t \tag{1.5}$$

as proposed by [39]. Here C is a constant and $M(\mathbf{n})$ is a positive continuous function on the unit sphere S^{n-1}; H_γ is the weighted mean curvature defined in (1.3), which is considered as the first variation of I of (1.1). An axiomatic derivation of equations like (1.4) and (1.5) is found, for example, in [34]. Mathematical theory is well-developed for (1.4) and its generalization (1.5) if γ is smooth and convex. For example one is able to extend a solution beyond singularities (e.g. pinching) to a global-in-time solution by a level set method [43] (see also [46] and [42]) whose analytic foundation is established by [10], [13]; see [27], [30] and references cited there.

At low temperature τ the Frank diagram of $\gamma = \gamma^\tau$

$$\text{Frank}\gamma = \{p \in \mathbf{R}^n;\ \gamma(p) \leq 1\}$$

may have a corner whose position vector directs to the normal of $\mathcal{W}_\gamma$. (Frank γ is a convex conjugate (or polar) of $\mathcal{W}_\gamma$.) There is a critical temperature $\tau_R(\mathbf{n})$ (depending on $\mathbf{n}$) called *roughening temperature* such that there is a facet of $\mathcal{W}_\gamma$ with the normal $\mathbf{n}$ if and only if $\tau < \tau_R(\mathbf{n})$. The evolution law also depends on temperature explicitly. When the latent heat is negligible, its general form [34] is

$$V = f(\mathbf{n},\ H_\gamma + C) \quad \text{on} \quad S_t \tag{1.6}$$

for a given smooth function $f = f^\tau$ on $S^{n-1} \times \mathbf{R}$, which is nondecreasing in the second variable. The theory of crystal growth [11] says that if $\tau \leq \tau_R(\mathbf{n})$, then

$$\frac{\partial f^\tau}{\partial X}(\mathbf{n}, X) = 0 \quad \text{at} \quad X = 0$$

while $\tau > \tau_R(\mathbf{n})$,

$$\frac{\partial f^\tau}{\partial X}(\mathbf{n}, X) \neq 0 \quad \text{at} \quad X = 0.$$

So if $\tau > \tau_R(\mathbf{n})$, the equation (1.6) can be approximated by (1.5) at least for small $H_\gamma + C$. However, if $\tau \leq \tau_R(\mathbf{n})$, we have to study (1.6) directly. Evolutions with facets are also discussed in surface sciences; see [9], [47] and papers cited there.

If Frank γ has a corner, the definition of solution is not clear even for (1.5). If Frank γ is a convex polyhedra, γ is called a crystalline energy (density). If $n = 2$ and S_t is a planar curve, a notion of solution is proposed by [2] and [48] by restricting S_t as a spcecial polygonal curve. This evolution is called a *crystalline motion.* Based on variational and order-preserving structure the notion of solution is extended by [16], [19], [21] for (1.5) and (1.6), when S_t is a general graph-like curve (§2.2 and Appendix). It applies for general graph-like curves with general γ not necessarily crystalline. Even the level set approach handling non graph-like curves is extended to this situation in [23], [24] ; see also [28] for a review. By now it is known that the problem for $n = 2$ is well-posed although the diffusion effect included in H_γ is not local. To see the difficulty of the problem we assume that $n = 2$, $S_t = \{(x, y); y = u(x, t)\}$ and $\gamma(p_1, p_2) = |p_1| + |p_2|$ and observe that (1.5) with $M \equiv 1, C \equiv 0$ equals

$$u_t(1 + u_x^2)^{-1/2} = (u_x/|u_x|)_x,$$

where subscripts t and x of u denote the partial derivatives. It formally equals (2.3) since $(1 + p^2)^{1/2}\delta(p) = \delta(p)$, where δ denotes the Dirac delta function; the notion of solution to (2.3) is unclear at all. Similar equation

$$u_t = (u_x/|u_x|)_x + u_{xx}$$

has been proposed by H. Spohn [47], where he proposed a notion of solution based on free boundary value problems.

In this paper we focus the problem whether our solution of (1.6) with singular γ can approximated by a solution of approximate equation (1.6) with regular γ, when S_t is given as the graph of a function. We discuss this problem in Section 2. Except the last convergence (2.11) the results are known (cf. [16], [21], [23], [24] and papers cited there). For evolving curves the notions (§2.2 and Appendix) of a solution are consistent with an ansatz that the flat portion called facet (whose normal corresponds to the corner of the Frank diagram) stays as flat during the evolution. We call this ansatz facet-stay-as-facet hypothesis. This hypothesis is invoked to define crystalline motion. Our convergence results in Section 2 assert that the facet-stay-as-facet hypothesis is fulfilled for a limit of solutions of approximate problems. This has a strong contrast to the problem for evolving surfaces where a facet may break (Remark 2.3 (i)). So far for three-dimensional problem even local-in-time solvability is unknown even when γ is crystalline. In Section 3 we claim that a solution proposed by H. Spohn [47] is a solution in our sense so it is obtained as a limit of approximate problems. For more examples of solutions see [36], [26]. There are several other applications of equations with singular diffusivity. The reader is referred to [45], [29], [25], [49] as well as [36], [26].

In the thermal grooving problem it is often more important to study evolution by surface diffusion [41]. This corresponds to the fourth order problem $V = -\Delta H_\gamma$ [8]. Although there are several analytic results when $H_\gamma = H$, for singular γ there are no analytic results; except [47] where several special solutions are proposed; however several numerical results are available as in [44]. This type of problem seems to be important to study thermal smoothing of surface [9].

Before we conclude this introduction we would like to point out several relations between microscopic approach and macroscopic approach. For equilibrium problems macroscopic model is justified as a limit of several microscopic models [5]. There is roughening transition in microscopic model [15]. At the low temperature macroscopic interfacial energy obtained from microscopic models may have singularities so that the Wulff shape has a facet for $n \geq 3$ (while it has no facet when $n = 2$). However, for nonequilibrium problem, the convergence from microscopic to macroscopic is only known above the roughening temperature [18], [35] mainly because of lack of estimates; see also a nice review by T. Funaki [17]. The authors are grateful to Professor Tadahisa Funaki and Professor Herbert Spohn for valuable informative remarks.

§2. General convergence results

We are interested in studying the convergence of a solution when singular interfacial energy is approximated by a smooth energy. So far there are two systematic ways to study this kind of problems. One is based on comparison priciples and is considered as an extension [19], [21] of the theory of viscosity solutions [12]. The other one is based on the theory of nonlinear semigroups initiated by [37] (see also [3]). It reflects the variational structure. The first method is so far restricted to one space dimensional problem but it applies to equations without divergence structure. The second method applies to general space dimension but it is restricted to a gradient system, which has a divergence structure. We first discuss the first method.

2.1. Viscosity approach

We consider a fully nonlinear evolution equation in one space dimension of the form

$$u_t + F(u_x, \Lambda_W(u)) = 0, \quad x \in \mathbf{R}, \quad t > 0 \tag{2.1}$$

with $\Lambda_W(u)$ formally written as $(W'(u_x))_x$. Here W is a given convex function on $\mathbf{R}$ and C^2 outside a discrete set P. Thus the derivative of W may have jumps. The function $F : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is a given continuous function satisfying a monotonicity condition:

$$F(p, X) \leq F(p, Y) \quad \text{for} \quad X \geq Y, \tag{2.2}$$

so that the equation (2.2) is (degenerate) parabolic. (The value Λ_W is actually unchanged by adding an affine function to W but we denote it by Λ_W rather than by $\Lambda_{W''}$:) If $F(p, X) = -X$, then (2.1) is the heat equation when $W(p) = p^2/2$. If $W(p) = |p|$, the equation (2.1) formally becomes

$$u_t = 2\delta(u_x)u_{xx}, \tag{2.3}$$

and the quantity $\delta(u_x)u_{xx}$ is not well-defined even in the distribution sense for smooth u. So we need to introduce a new notion of solution. (For this particular example both the first and the second methods provide a suitable notion of a solution.) In [19] (see also [21]) a notion of solution called a *viscosity solution* for initial value problem of (2.1) is proposed and its unique existence is proved under periodic boundary condition to avoid extra technicality; see [19] for other boundary conditions. We shall recall its definition as well as that of Λ_W briefly in the Appendix. Fortunately, in various settings we have the convergence principle.

CVP. Assume that $F_\varepsilon \to F$ and $W_\varepsilon \to W$ locally uniformly as $\varepsilon \to 0$. For $\varepsilon > 0$ let $u^\varepsilon \in C([0, T) \times \mathbf{T})$ be a solution of

(2.4)

$$u_t + F_\varepsilon(u_x, \Lambda_{W_\varepsilon}(u)) = 0 \quad \text{in} \quad (0, T) \times \mathbf{R} \quad \text{with} \quad u|_{t=0} = u_0^\varepsilon \quad \text{in} \quad \mathbf{R}$$

with $u_0^\varepsilon \in C(\mathbf{T}), \mathbf{T} = \mathbf{R}/\omega\mathbf{Z}$, $\omega > 0$. If $u_0^\varepsilon \to u_0$ in $C(\mathbf{T})$, then u^ε convergences to some function u locally uniformly in $[0, T) \times \mathbf{T}$ (without taking a subsequence) and u is a unique solution of (2.1) with the initial data $u_0 \in C(\mathbf{T})$. (The constant T may be taken as $+\infty$.) (We should not assume uniform convergence of derivatives of W_ε so that W is allowed to be non -C^1.)

To state the convergence result rigorously we need to introduce a class of W and F.

> $\mathcal{E} = \{W; W$ is convex in $\mathbf{R}$ and W is C^2 except some discrete set P. Moreover, $\sup_{K \setminus P} W'' = C_K < \infty$ for every compact set K in $\mathbf{R}\}$.

Any piecewise linear, convex function W belongs to $\mathcal{E}$. Also $W(p) = a|p|/2+bp^2/2$ for $a, b > 0$ belongs to $\mathcal{E}$. Let $\mathcal{F}$ be the set of all continuous function F satisfying the monotonicity condition (2.2). We shall state a special version of the convergence result in [21] where F is allowed to depend on the time explicitly.

Theorem 2.1 *(Convergence). Assume that $F_\varepsilon, F \in \mathcal{F}$ and that W_ε, $W \in \mathcal{E}$. Then (CVP) holds for viscosity solutions.*

Of course, if W_ε and F_ε are smooth and the problem (2.4) is strictly parabolic with smooth initial data u_0^ε, then the classical theory [38] of parabolic equations provides a unique smooth solution u^ε for (2.4). So Theorem 2.1 justifies the notion of solution when W' may have jumps in the sense that the solution is the limit of classical solutions of the strictly parabolic problems. On the other hand, if W_ε is piecewise linear, and W is smooth, Theorem 2.1 also provides the convergence of the crystalline algorithm (proposed by [2] and [48]). Theorem 2.1 extends some aspects of earliear convergnece results [16], [33] of the algorithm for a general equation. The reader is referred to [21], [22] for details and generalizations. As in [22] we also have the convergence of derivatives.

Theorem 2.2 *(Convergence of derivaties). Assume that $F_\varepsilon, F \in \mathcal{F}$ and that $W_\varepsilon, W \in \mathcal{E}$. Under the situation of (CVP) assume furthermore that $u_{0xx}^\varepsilon (\varepsilon > 0)$ is a finite Radon measure with $\limsup_{\varepsilon\to 0} \|u_{0xx}^\varepsilon\|_1 < \infty$. Then*

$$\lim_{\varepsilon\to 0} \sup_{0\le t<T'} \|u_x^\varepsilon - u_x\|_{L^r(\mathbf{T})}(t) = 0$$

for every $r \in [1,\infty)$ and $0 < T' < T$. Here $\|\cdot\|_1$ denotes the total variation of the measure and u_{0xx}^ε represents the distributional second derivative of u_0^ε.

Remark 2.3. (i) So far this method does not apply to a spatially inhomogeneous equation or higher dimensional problems because of complexity of nonlocal curvatures. Despite proposal of several notions of solutions [31], [4], [32], the local existence of a solution approximated by smoother problem is not yet known.

(ii) Theorem 2.1 applies to the interface equation (1.6) when S_t is the graph of a spatially periodic function of one variable. In fact, if $S_t = \{y = u(t, x)\}$, then (1.6) can be written in the form of (2.1) with

$$W(p) = \gamma(-p, 1),$$
$$F(p, X) = -(1+p^2)^{1/2} f((-p(1+p^2)^{-1/2}, (1+p^2)^{-1/2}), X + C)$$

when $\mathbf{n}$ is taken upward, *i.e.*, $\mathbf{n} = (-u_x, 1)/(1+u_x^2)^{1/2}$. The weighted curvature H_γ of S_t in the direction of $\mathbf{n}$ at $(x_0, u(x_0))$ equals $\Lambda_W(u)(x_0)$. Thus **CVP** implies that the solution $\{S_t^\varepsilon\}$ of

$$V = f_\varepsilon(\mathbf{n}, H_{\gamma_\varepsilon}) \quad \text{on} \quad S_t^\varepsilon = \{y = u^\varepsilon(t,x)\}$$

converges to the solution of $\{S_t\}$ of (1.6) in the Hausdorff distance sense in $[0,T) \times \mathbf{T} \times \mathbf{R}$, provided that $f_\varepsilon \to f$, $\gamma_\varepsilon \to \gamma$ locally uniformly as $\varepsilon \to 0$ and that $S_0^\varepsilon \to S_0$ in $\mathbf{T} \times \mathbf{R}$ in the Hausdorff distance sense. Such a convergence result has been proved for closed curves in more general setting [24].

2.2. Variational approach

We consider a gradient system in a multi-dimensional space $\mathbf{R}^n$ under periodic boundary condition, *i.e.* in $\mathbf{T}^n = \Pi_{j=1}^n(\mathbf{R}/\omega_j\mathbf{Z}), \omega_j > 0 (j = 1, ..., n)$:

$$(2.5) \qquad u_t - \operatorname{div}((DW)(\nabla u)) = 0 \quad \text{in} \quad \mathbf{T}^n \times (0,\infty).$$

Here $W : \mathbf{R}^n \to \mathbf{R}$ be a convex function and DW denotes the gradient of W. If initial data u_0 is Lipschitz continuous, the maximum principle yields a priori bound

$$(2.6) \qquad ||\nabla u||_\infty(t) \le ||\nabla u_0||_\infty \quad \text{for all} \quad t \ge 0,$$

where $||\cdot||_\infty$ denotes the L^∞- norm in $L^\infty(\mathbf{T}^n)$. For example,

$$||\nabla u||_\infty(t) = \text{ess.sup}_{x\in\mathbf{T}^n}|\nabla u(x,t)|.$$

Let K be a number such that $||\nabla u_0||_\infty \le K$. We may modify $W(p)$ for $|p| \ge K+1$ so that $W(p)$ is coercive in the sense that

$$(2.7) \qquad \lim_{|p|\to\infty} W(p)/|p| = \infty$$

without changing the notion of a solution with initial data u_0 since (2.6) holds. This modification simplifies the formulation. As in [16] we formulate the problem by using subdifferentials. If we define the energy for $v \in H = L^2(\mathbf{T}^n)$ by

(2.8)

$$\varphi(v) = \begin{cases} \displaystyle\int_{\mathbf{T}^n} W(\nabla v)dx & \text{if } \nabla v \in L^1(\mathbf{T}^n) \text{ and } W(\nabla v) \in L^1(\mathbf{T}^n), \\ \infty & \text{otherwise,} \end{cases}$$

then φ is convex and lower semicontinuous in H as in [6]. (The coercivity assumption (2.7) is important to conclude the lower semicontinuity.)

In [16] only one dimensional problem is treated but we follow their approach. Then the initial value problem for (2.5) with $\|\nabla u_0\|_\infty \le K$ is formulated as an abstract ordinary differential equation for $u(t) = u(\cdot, t)$ in the Hilbert space $L^2(\mathbf{T}^n)$ with the standard inner product $< f, g >= \int_{\mathbf{T}^n} fgdx$:

$$\frac{du}{dt}(t) \in -\partial\varphi(u(t)) \quad \text{for} \quad t > 0, \quad u|_{t=0} = u_0, \tag{2.9}$$

where $\partial\varphi$ denotes the subdifferential of φ, *i.e.*,

$$\partial\varphi(v) = \{w \in H; \quad \varphi(v+h) - \varphi(v) \ge < w, h > \quad \text{holds for all} \quad h \in H\}.$$

A general theory guarantees that for $u_0 \in H$ satisfy $\|\nabla u_0\|_\infty \le K$ there is a unique solution $u \in C([0,T); H)$ of (2.9) (with (2.8)) which is absolutely continuous with values in H on $[\delta, T]$ as a function of t for every $\delta > 0, T > 0$; see [3]. We refer this u as the *solution of* (2.5) (in the *variational sense*) with initial data u_0. As in [16] a general stability theorem due to J. Watanabe [50] (see also [26]) based on a result of H. Brezis and A. Pazy [7] implies the following convergence result.

Theorem 2.4. *Assume that W^ε and W are convex in $\mathbf{R}^n$. Assume that $W^\varepsilon(p) = W(p)$ for $|p| \ge K + 1$ and satisfies (2.7). Assume that $W^\varepsilon \to W$ (locally uniformly) as $\varepsilon \to 0$. Let u^ε be the solution of*

$$u_t - div((DW^\varepsilon)(\nabla u)) = 0 \quad in \quad \mathbf{T}^n \times (0, \infty), \quad u|_{t=0} = u_0^\varepsilon$$

with $\|\nabla u_0^\varepsilon\|_\infty \le K$. Assume that $u_0^\varepsilon \to u_0$ in $L^2(\mathbf{T}^n)$ as $\varepsilon \to 0$. Then u^ε converges to a solution u of (2.5) with the initial data u_0 in the sense that for any $T > 0$

$$\lim_{\varepsilon \to 0} \sup_{0 \le t \le T} \|u^\varepsilon - u\|_{L^2(\mathbf{T}^n)}(t) = 0. \tag{2.10}$$

Remark 2.5.(i) Since $\|\nabla u^\varepsilon\|_\infty(t) \le K$ for all $t \ge 0$ (cf (2.6)), by Arzelà-Ascoli's theorem we also get the uniform convergence

$$\lim_{\varepsilon \to 0} \sup_{0 \le t \le T} \|u^\varepsilon - u\|_\infty(t) = 0. \tag{2.11}$$

The proof of (2.11) admitting (2.10) and (2.6) is elementery; see [25, Lemma 4.3].

(ii) This convergence result also asserts that the solution with singular energy is approximated by that of smoother problem.

(iii) This approach applies to spatially inhomogeneous equation of the form

$$u_t - \frac{1}{b(x)}\{\mathrm{div}(a(x)(DW)(\nabla u)) + C(x)\} = 0$$

as described in [26] and [20].
(iv) By Theorem 2.1 and 2.4 if both variational and viscosity notion of solution is available it must agree each other, since both solutions are obtained as the same limiting procedure [20].

§3. Examples

Example 1. We consider (2.3) with ω-periodic boundary condition, *i.e.*, in $\mathbf{T} \times (0, \infty)$ with $\mathbf{T} = \mathbf{R}/(\omega\mathbf{Z})$ and $\omega > 0$. Assume that the initial data

$$u_0(x) = \begin{cases} A(x), & 0 \le x \le \alpha_0, \\ h_0, & \alpha_0 \le x \le \beta_0, \\ B(x), & \beta_0 \le x \le \omega/2 \end{cases} \tag{3.1}$$

with $\alpha_0 \le \beta_0, A' > 0, B' < 0, A(0) = B(\omega/2) = 0, A(\alpha_0) = B(\beta_0) = h_0$. Here A and B are C^1 and $h_0 > 0$ is a constant. We also assume that $A' \le K, -B' \le K$ with some $K > 0$ so that u_0 is Lipschitz continuous. We extend u_0 to $[-\omega/2,0]$ as an odd function, and further extend u_0 as an ω-periodic function in $\mathbf{R}$, *i.e.*, a function on $\mathbf{T}$. The problem (2.3) with (3.1) is formulated as in (2.9) if we take φ in (2.8) with $W(p) = |p|$ for $p, |p| \le K + 1$ where $\|u_{0x}\|_\infty \le K$. The solution is explicitly written as follows; see [19],[26]. Let $h(t)$ be a function defined by

$$h(t) = S^{-1}(2t),\ S(k) = \int_k^{h_0} (B^{-1}(\eta) - A^{-1}(\eta))d\eta,$$

where -1 represents the inverse of a function. This $h(t)$ satisfies $h(0) = h_0$ and is decreasing in time. Moreover, $h(T) = 0$ for $T = S(0)/2$. We set

$$u(x,t) = \begin{cases} \min(h(t), u_0(x)), & t \le T, \quad x \in [0, \omega/2], \\ 0, & t > T \end{cases}$$

and extend $u(\cdot, t)$ to $[-\omega/2, 0]$ as an odd function, and further extend $u(\cdot, t)$ as an ω-periodic function in $\mathbf{R}$. It turns out that $u(x,t)$ is the unique solution of (2.3) with (3.1) (in the variational sense). Indeed, if we set

$$\alpha(t) = A^{-1}(h(t)), \quad \beta(t) = B^{-1}(h(t)) \quad \text{with} \quad \alpha(0) = \alpha_0,\ \beta(0) = \beta_0,$$

then

$$h_t(t) = -\frac{2}{\beta(t) - \alpha(t)}. \tag{3.2}$$

The right hand side can be interpreted as $-(W'(+0) - W'(-0))/$ {length of flat portion} where $W(p) = |p|$. See Definition A.3 in the Appendix. Since

$$u(x,t) = \begin{cases} A(x), & 0 \le x \le \alpha(t), \\ h(t), & \alpha(t) \le x \le \beta(t), \\ B(x), & \beta(t) \le x \le \omega/2 \end{cases} \tag{3.3}$$

for $t \le T$, it is not difficult to derive

$$u_t(\cdot, t) \in -\partial\varphi(u(\cdot, t)) \quad \text{for all} \quad t > 0$$

from (3.2) [36], [26]. Indeed, we fix $t \in [0, T)$ and set

$$\zeta(x) = \begin{cases} -1, & 0 \le x \le \alpha, \\ \frac{2}{\beta-\alpha}(x - \alpha) - 1, & \alpha \le x \le \beta, \\ 1, & \beta \le x \le \omega/2; \end{cases}$$

here we suppress t-dependence of α, β and ζ. We extend ζ to $[-\omega/2, 0]$ as an even function, and further extend ζ as an ω-periodic function in $\mathbf{R}$. We then observe that $u_t(x,t) = -\zeta_x(x)$ for $x \in (0, \omega)$. To show $u_t \in -\partial\varphi(u)$ it suffices to prove

$$\zeta_x \in \partial\varphi(v) \quad \text{with} \quad v(x) = u(x,t).$$

We observe that

$$-\zeta(x) \in \partial W(v_x(x)), \quad 0 \le x \le \omega,$$

where ∂W is the subdifferential of a convex function W on $\mathbf{R}$. In other words, $W(v_x(x) + q) - W(v_x(x)) \ge -\zeta(x)q$ for all $q \in \mathbf{R}$, $x \in [0, \omega]$. Thus by definition of ∂W

$$\begin{aligned} \varphi(v + h) \quad - \quad \varphi(v) &= \int_0^\omega \{W(v_x(x) + h_x(x)) - W(v_x(x))\} dx \\ &\ge \int_0^\omega -\zeta(x) h_x(x) dx \end{aligned}$$

for all $h \in L^2(\mathbf{T})$ with $h_x \in L^2(\mathbf{T})$, $W(h_x) \in L^1(\mathbf{T})$. Integrating by parts yields

$$\varphi(v + h) - \varphi(v) \ge \int_0^\omega \zeta_x h dx$$

so we conclude that $\zeta_x \in \partial\varphi(v)$. Thus, we conclude that $u_t(\cdot,t) \in -\partial\varphi(u(\cdot,t))$ for each $t \in (0,T)$. It is clear that this relation holds for all $t \geq T$ since $0 \in -\partial\varphi(0)$.

Note that u_t is a constant on each flat portion of u and its quantity depends on the length of the flat portion so is determined nonlocally. We also note that the flat portion (facet) instantaneously (spontaneously) formed when $\alpha_0 = \beta_0$. The speed of $\alpha(t), \beta(t)$ at $t = 0$ is infinite in this case. By the way we note that the speed (3.2) at the facet can be formally obtained by integrating (2.3) on interval $(\alpha(t) - 0, \beta(t) + 0)$ if one assumes the facet-stay-as-facet hypothesis (see [2],[48]).

Our convergence theorems (Theorems 2.1, 2.4 and Remark 2.5 (i)) in partiular imply that such a solution u is obtained as a local uniform limit of the solution u^ε of

$$(A) \qquad u_t = a_\varepsilon(u_x)u_{xx}, \quad u|_{t=0} = u_0$$

with a smooth positive function a_ε such that $a_\varepsilon \rightarrow 2\delta$ as a weak convergence of measures in $(-K-1, K+1)$ as $\varepsilon \rightarrow 0$. (We may assume that $a_\varepsilon(p) = 1$ for p with $|p| \geq K+2$.) Moreover, u is the viscosity solution as shown in [19].

Example 2. We give another example of an equation that a facet is spontaneously formed. We consider

$$(3.4) \qquad u_t = 2c_0\delta(u_x)u_{xx} + u_{xx}$$

instead of (2.3) with $c_0 > 0$. For initial value u_0 we again consider ω-periodic function in $\mathbf{R}$ defined in Example 1. Our equation (3.4) is formulated as (2.9) by taking φ of (2.8) by setting

$$W(p) = c_0|p| + |p|^2/2 \quad \text{for} \quad p \quad \text{with} \quad |p| \leq K+1.$$

In [47] H. Spohn solves the initial value problem for (3.4) with (3.1) by reducing it to the Stefan problem studied by [14] under a symmetry assumption

$$(3.5) \qquad u_0(x - \omega/4) = u_0(-x - \omega/4).$$

Since his proposed solution is expressed by a different dependent variable, it is a priori not clear that it is the solution in our sense. We shall recall his solution. Assume that u is of the form

$$\begin{cases} u(x,t) = h(t), & \alpha(t) \leq x \leq \omega_1, \quad \omega_1 = \omega/4, \\ u(0,t) = 0, & \\ u_x(x,t) > 0, & 0 \leq x \leq \alpha(t) \end{cases}$$

with some free boundary $\alpha(t)$ at least for small $t > 0$. By our symmetry assumption (3.5) it is natural to assume that $u(x-\omega_1, t) = u(-x-\omega_1, t)$. Differentiate $u_t = (W'(u_x))_x$ with $W(p) = c_0|p| + |p|^2/2$ in x formally and set $w = u_x$ to get

$$w_t = g(w)_{xx} \tag{3.6}$$

with g defined by

$$g(w) = \begin{cases} c_0 + w & \text{for} \quad w > 0, \\ -c_0 + w & \text{for} \quad w < 0. \end{cases}$$

We set $v = g(w)$ and observe that $v(x,t) > c_0$ for $x \in [0, \alpha(t))$. As in [47] we postulate v and v_x is continuous accross $x = \alpha(t)$ and $v(\alpha(t), t) = c_0$ for (small) $t > 0$. Since $u_x = 0$ on $(\alpha(t), 2\omega_1 - \alpha(t))$, it is natural to postulate $0 < v(x,t) < c_0$ for $(\alpha(t), \omega_1)$ by symmetry. Here the case $\alpha_0 = \beta_0$ is also allowed. By (3.6) v satisfies

$$v_t = v_{xx} \quad \text{for} \quad x \in (0, \alpha(t)), \tag{3.7}$$

$$0 = v_{xx} \quad \text{for} \quad x \in (\alpha(t), \omega_1). \tag{3.8}$$

Since $v(\omega_1, t) = 0$ by symmetry, the equation (3.8) yields

$$v(x,t) = c_0(\omega_1 - x)/(\omega_1 - \alpha(t)), \quad x \in (\alpha(t), \omega_1).$$

By continuity assumption of v_x we have

$$v_x(\alpha(t) - 0, t) = -c_0/(\omega_1 - \alpha(t)) \quad \text{for} \quad \text{(small)} \quad t > 0. \tag{3.9}$$

Thus we obtain the Stefan type problem (3.7), (3.9) with $v(\alpha(t), t) = c_0$. The boundary condition $v_x(0,t) = 0$ is provided by the symmetry assumption of u_0. If (v, α) satisfies these equations, $u(x,t)$ for $0 < x < \alpha(t)$ must satisfiy the heat equation. According to [14] this problem is solvable until $\alpha(t)$ becomes zero provided that A in u_0 is C^3 in $[0, \alpha_0]$. The free boundary $\alpha = \alpha(t)$ is C^1 for $t > 0$ and continuous up to $t = 0$. Thus our u has the property that $u \in C([0,T), L^2(\mathbf{T}))$ and that u is absolutely continuous on $[\delta, T-\delta]$ for $\delta > 0$. To see that u is a solution of (3.4) in our variational sense it suffices to show that

$$\begin{aligned} u_t(x,t) &= \frac{-2c_0}{\beta(t) - \alpha(t)} \\ &= -(W'(+0) - W'(-0))/\{\text{the length of flat portion}\}, \end{aligned}$$

for $x \in (\alpha(t), \beta(t))$ and for $t \in (0,T)$ with

$$T = \sup\{t; \quad \alpha(\tau) > 0 \quad \text{for} \quad \tau \in [0,t)\},$$

where $\beta(t) = 2\omega_1 - \alpha(t)$. In fact, as in Example 1 this speed relation together with $u_t = u_{xx}$ for $0 \leq x \leq \alpha(t)$ yields $u_t \in -\partial\varphi(u)$ for a.e. $t \in (0,T)$ by observing that for each $t \in (0,T)$

$$\zeta(x) = \begin{cases} -u_x, & 0 \leq x \leq \alpha, \\ \frac{2c_0}{\beta-\alpha}(x-\alpha) - c_0, & \alpha \leq x \leq \beta, \\ -u_x, & \beta \leq x \leq \omega/2 \end{cases}$$

fulfills $u_t = -\zeta_x$ and $\zeta_x \in \partial\varphi(v)$ with $v(x) = u(x,t)$, where we suppress t-dependence of α, β and ζ. (As in Example 1, we extend ζ as an ω-periodic function in $\mathbf{R}$.)

Since for $t \geq T$ we have $u \equiv 0$ so (2.9) is clearly fulfilled for $t \geq T$. In other words it suffices to prove that

$$\text{(3.10)} \qquad u_t(x,t) = -c_0/(\omega_1 - \alpha(t)), \quad \alpha(t) \leq x \leq \beta(t).$$

We integrate (3.7) with respect to $x \in (0, \alpha(s))$ and then the time variable $s \in (0,t)$. We observe that

$$\begin{aligned} \int_0^{\alpha(s)} v_{xx}(x,s)dx &= v_x(\alpha(s)-0,s) - 0 \\ &= -c_0/(\omega_1 - \alpha(s)) \end{aligned}$$

by (3.9) and that

$$\begin{aligned} &\int_0^t \left(\int_0^{\alpha(s)} v_t(x,s)dx \right) ds = \int_0^t \left(\int_0^{\alpha(s)} (v-c_0)_t(x,s)dx \right) ds \\ &= \int_0^{\alpha(t)} (v(x,t) - c_0)dx - \int_0^{\alpha(0)} (v(x,0) - c_0)dx, \end{aligned}$$

by changing the order of integration and $v(\alpha(s), s) = c_0$. Thus from (3.7) it follows that

$$\int_0^{\alpha(0)} (v(x,0) - c_0)dx = \int_0^{\alpha(t)} (v(x,t) - c_0)dx + c_0 \int_0^t (\omega_1 - \alpha(s))^{-1} ds.$$

Since $u(\alpha(t),t) = \int_0^{\alpha(t)} (v(x,t) - c_0)dx$, we have

$$\frac{d}{dt}(u(\alpha(t),t)) = \frac{d}{dt} \int_0^{\alpha(t)} (v(x,t) - c_0)dx = -\frac{c_0}{\omega_1 - \alpha(t)}.$$

Since $v(\alpha(t),t)=c_0$ so that $u_x(\alpha(t),t)=0$, this implies (3.10).

We thus conclude that Spohn's solution is the solution in our variational sense. Thus, again our convergence theorems (Theorem, 2.1, 2.4 and Remark 2.5(i)) in paticular implies that such a solution u can be obtained as a local uniform limit of the solution of the approximate equation (A) if $a_\varepsilon \to 2c_0\delta+1$ as $\varepsilon \to 0$. Moreover, it is the viscosity solution. Thus as noted in [21], [22] it can be approximated numerically by a *crystalline algorithm.* A similar remark also applies to Example 1.

If u_0 is concave in $[0,\omega/2]$, $u(x,t)$ is also concave in $[0,\omega/2]$ for $t \in [0,T]$. This can be proved by above approximation and the standard maximum principle. In this case our solution u of (3.4) is a subsolution of (2.3) on $(0,\omega/2)$. Thus by a comparison theorem [19] $u \equiv 0$ for $t > T_0$ with some $T_0 > 0$ since the solution of (2.3) vanishes in finite time. In [47] this phenomena has been proved by a different method under the assmption that $A'' < 0$ in $[0,\omega/2]$. In his situation α is monotone decreasing as shown in [47].

Appendix. Definition of viscosity solution and its existence

We recall the definition of viscosity solution for (2.1) and the existence theorem for the reader's convenience [19], [21]. In the appendix we assume $W \in \mathcal{E}$ and $F \in \mathcal{F}$. Let Ω be an open interval.

Definition A.1 (P-faceted). A function $f \in C(\Omega)$ is called *faceted* at $x_0 \in \Omega$ with slope p in Ω if f fulfills the following properties : there is a closed nontrivial finite interval I $(\subset \Omega)$ (called a *faceted region*) containing x_0 such that f agrees with an affine function

$$l_p(x) = p(x-x_0) + f(x_0)$$

in I and $f(x) \neq l_p(x)$ for all $x \in J \setminus I$ with some neighborhood $J(\subset \Omega)$ of I. The length of I is denoted by $L(f,x_0)$. For a discrete set P in $\mathbf{R}$ a function f is called *P-faceted* at x_0 in Ω if f is faceted at x_0 in Ω with some slope $p \in P$.

Definition A.2 (Space of test functions). Let P be the set of jump discontinuities of W'. Let $C^2_P(\Omega)$ be the set of all $f \in C^2(\Omega)$ such that f is P-faceted at x_0 in Ω whenever $f'(x_0) \in P$. For $Q := (0,T)\times\Omega$ with $T>0$ let $A_P(Q)$ be the set of all function on Q of the form

$$f(x)+g(t), \quad f \in C^2_P(\Omega), \quad g \in C^1(0,T).$$

Definition A.3 (Weighted curvature). Let P be the set of jump discontinuities of W'. Let x_0 be a point in Ω.

For $f \in C^2(\Omega)$ we set the value

$$\Lambda_W(f)(x_0) = W''(f'(x_0))f''(x_0)$$

if $f'(x_0) \notin P$, and

$$\Lambda_W(f)(x_0) = \frac{\chi(f, x_0)}{L(f, x_0)} \Delta(f'(x_0))$$

if $f'(x_0) \in P$ and f is P-faceted at x_0 in Ω. Here $\Delta(p) = W'(p+0) - W'(p-0)$ for $p \in P$, and $\chi(f, x_0)$ is the *transition number* defined by

$$\begin{cases} \chi = +1 & \text{if} \quad f \geq l_p \quad \text{in} \quad J, \\ \chi = -1 & \text{if} \quad f \leq l_p \quad \text{in} \quad J, \\ \chi = 0 & \text{otherwise} \end{cases}$$

for some neighborhood J of the faceted region I.

Definition A.4 (Viscosity solution). A real valued continuous function u on $Q = (0, T) \times \Omega$ is a *viscosity subsolution* of

$$u_t + F(u_x, \Lambda_W(u)) = 0 \quad \text{in} \quad Q \tag{1}$$

if

$$\psi_t(\hat{t}, \hat{x}) + F(\psi_x(\hat{t}, \hat{x}), \Lambda_W(\psi(\hat{t}, \cdot), \hat{x})) \leq 0 \tag{2}$$

whenever $(\psi, (\hat{t}, \hat{x})) \in A_P(Q) \times Q$ fullfills

$$\max_Q (u - \psi) = (u - \psi)(\hat{t}, \hat{x}).$$

A *viscosity supersolution* is defined by replacing max by min, and the inequality in (2) by the reverse one. If u is a sub- and supersolution, u is called a *viscosity solution.*

Theorem A.5 (Existence [19]). *Suppose that* $u_0 \in C(\mathbf{R})$ *is periodic with period* $\omega > 0$. *Then there exists a unique function* $u \in C([0, T] \times \mathbf{R})$ *(for any* $T > 0$*) that satisfies*
(i) u *is a viscosity solution of (1) in* $(0, T) \times \mathbf{R}$;
(ii)$u(0, x) = u_0(x)$ *for* $x \in \mathbf{R}$;
(iii) $u(t, x + \omega) = u(t, x)$ *for* $(t, x) \in [0, T) \times \mathbf{R}$.

References

[1] A. D. Alexandrov, Uniqueness theorems for surfaces in the large I, Vestnik Leningrad Univ. Math., **11** (1956), 5-17.

[2] S. B. Angenent and M. E. Gurtin, Multiphase thermomechanics with interfacial structure 2. Evolution of an isothermal interface, Arch. Rational Mech. Anal., **108** (1989), 323-391.

[3] V. Barbu, "Nonlinear Semigroups and Differential Equations in Banach spaces", Noordnoff, Groningen, 1976.

[4] G. Bellettini and M. Novaga, Approximation and comparison for nonsmooth anisotropic motion by mean curvature in $\mathbf{R}^N$, Math. Mod. Methods. Appl. Sci., **10** (2000), 1-10.

[5] T. Bodineau, D. Ioffe and Y. Velenik, Rigorous probabilistic analysis of equilibrium crystal shapes, J. Math, Phys., **41** (2000), 1033-1098.

[6] H. Brezis, Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equation, Contribution to Nonlinear Functional Analysis (ed. E. Zarantonello) Academic Press (1971), pp . 101-156.

[7] H. Brezis and A. Pazy, Convergence and approximation of semigroups of nonlinear operators in Banach spaces, J. Funct. Anal., **9** (1972), 63-74.

[8] J. W. Cahn and J. E. Taylor, Surface motion by surface diffusion, Acta Metallurgica **42** (1994), 1045-1063.

[9] A. Chame, F. Lancon, P. Politi, G. Renaud, I. Vilfan and J. Villain, Three mysteries in surface science, Int. J. Modern Phys. B, **11** (1997), 3657-3671.

[10] Y. -G. Chen, Y. Giga and S. Goto, Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations, J. Diff. Geom., **33** (1991), 749-786.

[11] A. A. Chernov, Morden Crystallography III, Crystal Growth, Solid-State Sciences **36**, Springer 1984 (especially §3.3).

[12] M. G. Crandall, H. Ishii and P. -L. Lions, User's guide to viscosity solutions of second order partial differential equation, Bull. Amer. Math. Soc., **27** (1992), 1-67.

[13] L. C. Evans and J. Spruck, Motion of level sets by mean curvature I, J. Diff. Geom., **33** (1991), 635-681.

[14] A. Fasano and M. Primicerio, Liquid flow in partially saturated porous media, J. Inst. Maths. Appls. **23** (1979), 503–517.

[15] J. Fröhlich and T. Spencer, The Kosterlitz-Thouless transition in two-dimensional abelian spin systems and the Coulomb gas, Commun. Math. Phys. **81** (1981), 527-602.

[16] T. Fukui and Y. Giga, Motion of a graph by nonsmooth weighted curvature, in "World Congress of Nonlinear Analysis '92", (V. Lakshmikantham, ed.), de Gruyter, Berlin, voll, 1996, pp. 47-56.

[17] T. Funaki, Probabilistic model for phase separation and interface equations, Sūgaku **50** (1998), 68-85. (Japanese) Sugaku Expositions **16** (2003), 97-116. (English translation)

[18] T. Funaki and H. Spohn, Motion by mean curvature from the Ginsburg-Landau $\nabla\phi$ interface models, Commun. Math. Phys., **185** (1997), 1-36.

[19] M.-H. Giga and Y. Giga, Evolving graphs by singular weighted curvature, Arch. Rational Mech. Anal. **141** (1998), 117-198.

[20] M.-H Giga and Y. Giga, A subdifferential interpretation of crystalline motion under nonuniform driving force, Proc. of the International Conference in Dynamical Systems and Differential Equations, Springfield Missouri, (1996), "Dynamical Systems and Differential Equations", (eds. W.-X. Chen and S.-C. Hu), Southwest Missouri Univ. vol.**1** (1998), pp.276-287.

[21] M.-H. Giga and Y. Giga, Stability for evolving graphs by nonlocal weighted curvature Commum. in Partial Differential Equations, **24** (1999), 109-184.

[22] M.-H. Giga and Y. Giga, Remarks on convergence of evolving graphs by nonlocal curvature, In : Progress in Partial Differential Equations vol. 1 (eds. H. Amann et al) Pitman Research Notes in Mathematics Series, **383** (1998), pp. 99-116, Longman, Essex, England.

[23] M.-H. Giga and Y. Giga, Crystalline and level set flow - Convergence of a crystalline algorithm for a general anisotropic curvature flow in the plane, Gakuto International Series, Math. Sciences and Appl. vol **13** (2000) Free Boundary Problems: Theory and Applications (ed. N. Kenmochi), pp.64-79.

[24] M.-H. Giga and Y. Giga, Generalized motion by nonlocal curvature in the plane, Arch. Rational. Mech. Anal., **159** (2001), 295-333.

[25] M.-H. Giga and Y. Giga, Minimal vertical singular diffusion preventing overturning for the Burges equation, Contem. Math., to appear.

[26] M.-H. Giga, Y. Giga and R. Kobayashi, Very singular diffusion equations, Proc. of Taniguchi Conf. on Math., Advanced Studies in Pure Mathematics **31** (2001), pp. 93-125.

[27] Y. Giga, A level set method for surface evolution equations, Sūgaku **47** (1995), 321-340; English translation, Sugaku Exposition, **10** (1997), 217-241.

[28] Y. Giga, Anisotropic curvature effect in interface dynamics, Sūgaku **52** (2000), 113-127; English translation, to appear.

[29] Y. Giga, Viscosity solutions with shocks, Comm. Pure Appl. Math., **55** (2002), 431-480.

[30] Y. Giga, "Surfaces Evolution Equations - a level set method", Hokkaido University Technical Report Series in Mathematics, ♯71, Sapporo, 2002, also Lipschitz Lecture Notes, **44**, University of Bonn, 2002.

[31] Y. Giga, M. E. Gurtin and J. Matias, On the dynamics of crystalline motions, Japan J. Indust. Appl. Math., **15** (1998), 7-50.

[32] Y. Giga, M. Paolini and P. Rybka, On the motion by singular interfacial energy, Japan J. Indust. Appl. Math., **18** (2001), 231-248.

[33] P. M. Girão and R. V. Kohn, Convergence of a crystalline algorithm for the heat equation in one dimension and for the motion of a graph by weighted curvature, Numer. Math., **67** (1994), 41-70.

[34] M. E. Gurtin, "Thermomechanics of Evolving Phase Boundaries in the Plane", Oxford, Clarendon Press, 1993.

[35] M. Katsoulakis and P. E. Souganidis, Interacting particle systems and generalized evolution of fronts, Arch. Rational Mech. Anal., **127** (1994), 133-157.

[36] R. Kobayashi and Y. Giga, Equations with singular diffusivity, J. Stat. Phys., **95** (1999), 1187-1220.

[37] Y. Kōmura, Nonlinear semi-groups in Hilbert space, J. Math. Soc. Japan, **19** (1967), 493-507.

[38] O. A. Ladyzhenskaya, V. A. Solonnikov and N. N. Ural'tseva, "Linear and Quasi-Linear Equation of Parabolic Type", AMS, 1968.

[39] H. Müller-Krumbhaar, T. W. Burkhardt and D. M. Kroll, A generalized kinetic equation for crystal growth, J. Crystal Growth, **38** (1977), 13-22.

[40] W. W. Mullins, Two-dimensional motion of idealized grain boundaries, J. Appl. Phys., **27** (1956), 900-904.

[41] W. W. Mullins, Grain boundary grooving by volume diffusion, Trans. Met. Soc., AIME **218** (1960), 354-361.

[42] S. Osher and R. Fedkiw, "Level set Methods and Dynamic Implicit Surfaces", Applied Math. Ser. 153 Springer, 2003.

[43] S. Osher and J. A. Sethian, Front propagation with curvature dependent speed: Algorithm based on Hamilton-Jacobi formulations, J. Comput. Phys., **79** (1988), 12-49.

[44] A. Roosen and J. E. Taylor, An algorithm for the computation of crystal growth in a diffusion field using fully faceted interfaces, J. Comp. Phys., **114** (1994), 113-128.

[45] L. I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D **60** (1992), 259-268.

[46] J. A. Sethian, "Level Set Methods, Evolving Interfaces in Geomety, Fluide Mechanics, Computer Vision, and Materials Science", Cambridge Univ. Press, 1996.

[47] H. Spohn, Surface dynamics below the roughening transition, J. Phys. I France, **3** (1993), 69-81.

[48] J. Taylor, Constructions and conjectures in crystalline nondifferential geometry, In : Differential Geometry (eds. B. Lawson and K. Tanenblat), Proceedings of the Conference on Differential Geometry, Rio de Janeiro, Pitman Monograph Surveys Pure Appl. Math., **52** (1991), 321-336.

[49] Y.-H. R. Tsai, Y. Giga and S. Osher, A level set approach for computing discontinuous solutions of a class of Hamilton-Jacobi equations, Math. Comp., **72** (2003), 159-181.

[50] J. Watanabe, Approximation of nonlinear problems of a certain type, in "Numerical Analysis of Evolution Equations", (H. Fujita and M.

Yamaguti, eds.), Lecture Notes in Num. Appl., 1, Kinokuniya, Tokyo, 1979, pp. 147-163.

Mi-Ho Giga and Yoshikazu Giga
Department of Mathematics
Hokkaido University
Sapporo 060-0810
Japan

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 233–281

The Dobrushin-Hryniv Theory for the Two-Dimensional Lattice Widom-Rowlinson Model

Yasunari Higuchi, Joshin Murai and Jun Wang

Abstract.

We consider the fluctuation of the phase boundary separating two phases of the Widom-Rowlinson model in the plane square lattice. The phase boundary is conditioned to have specified values of the area underneath and the height difference of two end points. Dobrushin and Hryniv studied the phase boundary of the Solid-on-Solid model [DH1] and of the Ising model [DH2], and obtained the central limit theorem for the fluctuation of the phase boundary from the Wulff profile. The phase boundary of the Ising model is well approximated by that of the Solid-on-Solid model with the aid of the cluster expansion. Their argument seems to be applicable to the general models which have polymer representation. We apply their theory to the Widom-Rowlinson model.

§1. Introduction

Let $\mathbf{Z}^2$ be the square lattice and let $\Lambda_{L,M}$ be the rectangle $[1, L-1] \times [-M, M]$ in $\mathbf{Z}^2$. We consider a system of particles in $\Lambda_{L,M}$. These particles are of two types, either A or B. There is strong repulsive interaction between particles of different types. Namely, a B particle can not occupy a site within distance $\sqrt{2}$ from a site where an A particle sits, and vice versa.

A *configuration* ω is a function from $\Lambda_{L,M}$ to $\{-1, 0, +1\}$. $\omega(x) = +1$ denotes that the site x is occupied by an A particle, $\omega(x) = -1$ denotes that x is occupied by a B particle and $\omega(x) = 0$ denotes that there is no particle at x. We say that a configuration ω is *feasible* if

---

Received April 10, 2003.
Revised June 24, 2003.
Supported by Grant-in-Aid for Scientific Research(B)No.12440027 and Grant-in-Aid for JSPS fellows No.00026.

$\omega(x)\omega(y) \geq 0$ for all pairs x, y with $|x - y| \leq \sqrt{2}$, where $|\cdot|$ denotes the Euclidean distance.

Let $\Omega_{L,M}$ denote the set of all feasible configurations in $\Lambda_{L,M}$. The Hamiltonian of our system is a function on $\Omega_{L,M}$ given by

$$H(\omega) = \sum_{x \in \Lambda_{L,M}} \mu(1 - \omega(x)^2) \tag{1.1}$$

for every $\omega \in \Omega_{L,M}$. Here, μ denotes the chemical potential.

Let $h > 0$ be fixed and assume that $M > Lh$. Then we can put the following boundary condition:

$$\eta^h(x) = \begin{cases} +1, & \text{if } x^2 > \lceil hx^1 \rceil, \\ 0, & \text{if } x^2 = \lceil hx^1 \rceil, \\ -1, & \text{otherwise,} \end{cases}$$

for every $x = (x^1, x^2) \in \partial\Lambda_{L,M} = [0, L] \times [-M - 1, M + 1] \setminus \Lambda_{L,M}$. Let $\Omega^h_{L,M}$ denote the set of all configurations ω in $\Omega_{L,M}$ such that $\omega \circ \eta^h$ is feasible, where $\omega \circ \eta$ is given by

$$\omega \circ \eta(x) = \begin{cases} \omega(x), & \text{if } x \in \Lambda_{L,M}; \\ \eta(x), & \text{if } x \in \partial\Lambda_{L,M}. \end{cases}$$

The conditional Gibbs distribution on $\Omega^h_{L,M}$ with the boundary condition η^h is given by

$$P^h_{L,M}(\omega) = \left(Z^h_{L,M}\right)^{-1} \exp\{-\mu|S^0(\omega)|\}, \tag{1.2}$$

where $S^0(\omega)$ is the set of points in $\Lambda_{L,M}$ such that ω takes 0 value, $|S|$ denotes the cardinality of a set S, and $Z^h_{L,M}$ is the normalizing constant, which we call the *partition function.*

For a feasible configuration ω, we call a connected component of $S^0(\omega)$ a *contour.* Among contours we can find a unique contour which connects $(0, 0)$ with $(L, \lceil hL \rceil)$. We call this the *separating contour* with the starting point $(0, 0)$ and the end point $(L, \lceil hL \rceil)$, and denote it by $\Gamma(\omega)$. Let $\mathcal{S}^h_{L,M}$ denote the collection

$$\{\Gamma(\omega); \omega \in \Omega^h_{L,M} \text{ is feasible}\}.$$

The aim of this paper is to investigate the fluctuation of the separating contour via Dobrushin-Hryniv theory.

the backbone

We say that a set $C \subset \mathbf{Z}^2$ is $*$*connected* if for every $x, y \in C$, there exist a sequence $z_0 = x, z_1, \ldots, z_m = y$ in C such that $|z_i - z_{i-1}| \leq \sqrt{2}$ for every $1 \leq i \leq n$. A *hole* of a connected set $C \subset \mathbf{Z}^2$ is a finite $*$connected component of $C^c = \mathbf{Z}^2 \setminus C$. Let $C_1, C_2, \dots, C_n$ be connected subsets of $\Lambda_{L,M}$. We say that contours $\{C_j\}$ are *compatible* if they are connected components of the set $\cup_{1 \leq j \leq n} C_j$. We also say that $\{C_j\}$ are *compatible* with a connected set D if $\{D, C_j\}$ are compatible for every j. Then the partition function $Z^h_{L,M}$ can be rewritten as

$$Z^h_{L,M} = \sum_{\Gamma \in \mathcal{S}^h_{L,M}} \sum_{\{C_j\}} 2^{N(\Gamma)} \exp\{-\mu|\Gamma|\} \prod_j (2^{N(C_j)} \exp\{-\mu|C_j|\}),$$

where the second summation is taken over compatible families $\{C_j\}$, which are compatible with Γ, $|\Gamma|$ is the number of points in Γ and $N(C)$ is the number of holes in C. Therefore, we can find μ_0 sufficiently large so that we have a cluster expansion (see [KP])

$$(1.3) \qquad Z^h_{L,M} = \sum_{\Gamma \in \mathcal{S}^h_{L,M}} \exp\{-\mu|\Gamma| + N(\Gamma) \ln 2 + \sum_{\substack{\Lambda \subset \Lambda_{L,M} \\ \Lambda\, c\, \Gamma}} \Phi(\Lambda)\}$$

for $\mu > \mu_0$, where $\Lambda\, c\, \Gamma$ denotes that Λ is compatible with Γ. Moreover, the function $\Phi(\Lambda)$ satisfies the estimate

$$(1.4) \qquad \sum_{\Lambda \ni 0} |\Phi(\Lambda)| e^{(\mu-\mu_0)|\Lambda|} < 1,$$

and $\Phi(\Lambda) = 0$ unless Λ is connected. Let

$$Z^+_{L,M} = \exp\{ \sum_{\Lambda \subset \Lambda_{L,M}} \Phi(\Lambda)\}.$$

Dividing both sides of (1.3) by $Z^+_{L,M}$, we have

$$(1.5) \qquad \frac{Z^h_{L,M}}{Z^+_{L,M}} = \sum_{\Gamma \in \mathcal{S}^h_{L,M}} \exp\{-\mu|\Gamma| + N(\Gamma) \ln 2 - \sum_{\substack{\Lambda \subset \Lambda_{L,M} \\ \Lambda\, i\, \Gamma}} \Phi(\Lambda)\},$$

where $\Lambda\, i\, \Gamma$ denotes that Λ is incompatible with Γ. We use the summand in the right hand side of (1.5) as a statistical weight of the separating contour Γ. Let $\Gamma \in \mathcal{S}^h_{L,M}$. We extract a self-avoiding path from Γ in the following way.

First we define an order of preference among four directions;
up > down> right > left.

This order naturally defines an order among self-avoiding paths connecting $(0,0)$ with $(L, \lceil hL \rceil)$. To be more precise, let $\pi = \{x_1, x_2, \dots, x_n\}$

and $\pi' = \{y_1, y_2, \dots, y_k\}$ be two self-avoiding paths connecting $(0,0)$ with $(L, \lceil hL \rceil)$. Let j_0 be the first number j such that $x_j \neq y_j$. We define that $\pi > \pi'$ if the direction of the ordered edge $\{x_{j_0-1}, x_{j_0}\}$ is preferred to the direction of the ordered edge $\{y_{j_0-1}, y_{j_0}\}$. Now, let

$$\Pi_\Gamma := \{\pi : \text{ self-avoiding path in } \Gamma \text{ connecting } (0,0) \text{ with } (L, \lceil hL \rceil)\}.$$

Let $\pi(\Gamma)$ be the unique maximal element of Π_Γ with respect to this order. We call $\pi(\Gamma)$ the *backbone* of Γ. This backbone will play the role of the phase separation line of the 2D Ising model.

For $\Gamma \in \mathcal{S}^h_{L,M}$, $\pi(\Gamma)$ separates $[0, L] \times [-M-1, M+1]$ into two $*$connected components. One is above $\pi(\Gamma)$ and the other is below $\pi(\Gamma)$. Let $a^-(\pi(\Gamma))$ and $a^+(\pi(\Gamma))$ be the number of points in $\mathbf{Z}^{2^*} \cap [0, L] \times [-M-1, M+1]$, which are below $\pi(\Gamma)$ and above $\pi(\Gamma)$, respectively. Here, $\mathbf{Z}^{2^*} = \mathbf{Z}^2 + (\frac{1}{2}, \frac{1}{2})$. The area $a(\pi(\Gamma))$ is defined by

$$a(\pi(\Gamma)) := a^-(\pi(\Gamma)) - a^+(\pi(\Gamma)). \tag{1.6}$$

This value is independent of M if M is sufficiently large.

free energy

If μ is sufficiently large, (1.5) has a limit as $M \to \infty$:

$$\lim_{M \to \infty} \frac{Z^h_{L,M}}{Z^+_{L,M}} = \sum_{\Gamma \in \mathcal{S}^h_L} \exp\left\{ -\mu|\Gamma| + N(\Gamma)\ln 2 - \sum_{\substack{\Lambda \subset \Lambda_{L,\infty} \\ \Lambda\, i\, \Gamma}} \Phi(\Lambda) \right\}, \tag{1.7}$$

where $\mathcal{S}^h_L := \cup_{M>0} \mathcal{S}^h_{L,M}$, $\Lambda_{L,\infty} := [1, L-1] \times (-\infty, \infty) \cap \mathbf{Z}^2$.

Let $W(\Gamma)$ be the weight in the right hand side of (1.7);

$$W(\Gamma) := \exp\left\{ -\mu|\Gamma| + N(\Gamma)\ln 2 - \sum_{\substack{\Lambda \subset \Lambda_{L,\infty} \\ \Lambda\, i\, \Gamma}} \Phi(\Lambda) \right\}$$

for $\Gamma \in \cup_{h \in \mathbf{R}} \mathcal{S}^h_L = \mathcal{S}_L$. For $\Gamma \in \mathcal{S}_L$, we denote by $A(\Gamma) = (0,0)$ and $B(\Gamma) = (L, k(\Gamma))$ the starting point and endpoint of Γ, respectively.

For $\zeta \in \mathbf{C}$, we define

$$\varphi(\zeta) := \lim_{L \to \infty} \frac{1}{L} \ln \sum_{\Gamma \in \mathcal{S}_L} e^{\mu\zeta k(\Gamma)} W(\Gamma). \tag{1.8}$$

if the limit exists. This is the free energy of the height of the last point of Γ. For $\Gamma \in \mathcal{S}_L$, we define $\{X_L(t);\ t \in [0,1]\} = \{X_L(t;\Gamma);\ t \in [0,1]\}$

by

$$\begin{cases} X_L\left(\frac{j}{L}\right) = \max\{k \in \mathbf{Z};\ (j,k) \in \pi(\Gamma)\}, \\ X_L(t) \ \ = (j+1-Lt)X_L\left(\frac{j}{L}\right) + (Lt-j)X_L\left(\frac{j+1}{L}\right) \quad (j \le Lt \le j+1) \end{cases}$$

Let P_L be the probability measure on $\mathcal{S}_L$ defined by

$$P_L(\Gamma) = \left[\sum_{\Gamma' \in \mathcal{S}_L} W(\Gamma')\right]^{-1} W(\Gamma). \tag{1.9}$$

Theorem There exists $\mu_1 > \mu_0$ such that for $\mu > \mu_1$, (1.9) is well defined on $\mathcal{S}_L$ and the followings hold.

Assume that for $h > 0$ and $a \ge \frac{h}{2}$ there exist a $\delta > 0$ and a pair $(\zeta_0, \zeta_1) \in \mathbf{R}^2$ with $\max\{|\zeta_0 + \zeta_1|, |\zeta_1|\} \le 1 - \frac{\delta}{\mu}$ such that

$$\frac{1}{\mu}\int_0^1 \nabla_{(\zeta_0,\zeta_1)}\varphi(\zeta_0(1-x)+\zeta_1)dx = (a,h).$$

Then the process

$$Y_L(t) := \frac{1}{\sqrt{L}}\left\{X_L(t) - \frac{L}{\mu}\int_0^t \varphi'(\zeta_0(1-x)+\zeta_1)dx\right\}$$

under $P_L(\cdot \mid a(\pi(\Gamma)) = \lceil aL^2 \rceil, k(\Gamma) = \lceil hL \rceil)$ converges

$$Y(t) = \frac{1}{\mu}\int_0^t \sqrt{\varphi''(\zeta_0(1-x)+\zeta_1)}dB(x)$$

conditioned that

$$\int_0^1 Y(t)dt = 0, \quad Y(1) = 0.$$

Here, $\{B(t)\}_{t\ge 0}$ is the one dimensional standard Brownian motion.

Remark Although $X_L(t)$ is defined by the backbone $\pi(\Gamma)$, the width (in the x^2 direction) of the separating contour Γ is negligible and, hence, the limiting process $Y(t)$ depends only on Γ. So, the choice of the backbone is for technical reasons only.

The proof of the theorem goes along the line of [DH1,2], and we regard our model as a perturbation of Solid-on-Solid(SOS) model. This SOS model corresponds to the ensemble of (site) self avoiding paths in $[0, L] \times \mathbf{Z}$ starting from $(0,0)$ and ending at a site in $\{x^1 = L\}$, which do not go back in the horizontal direction. Let us call such a path an *SOS path.* There are no $\{\Lambda_\alpha\}$'s for the SOS model.

An SOS path will be cut into simple polymers. A simple polymer is obtained from intersection of an SOS path with a vertical line $\{x^1 = j\}$ for some $1 \le j \le L$, shifted so that its starting point is at height zero. So, it has a form $\{(j,0),(j,1),\dots,(j,k)\}$ for some $k \ge 0$ or $\{(j,0),(j,-1),\dots,(j,k)\}$ for some $k < 0$.

Let

$$Q(\zeta) = \sum_{\substack{\xi:simple\ polymer \\ starting\ from\ (0,0)}} e^{\mu\zeta k(\xi)-\mu|\xi|}$$

where $k(\xi)$ and $|\xi|$ are the height of the endopoint of ξ and number of sites in ξ, respectively. Then

$$\sum_{\Gamma:SOS\ path\ in\ [0,L]\times\mathbf{Z}} e^{\mu}\zeta k(\Gamma)W(\Gamma) = Q(\zeta)^L.$$

We would like to show that

$$Q(\zeta)^{-L} \sum_{\Gamma\in\mathcal{S}_L} e^{\mu\zeta k(\Gamma)}W(\Gamma)$$

has a form;

$$\sum_{\substack{I_1,\dots,I_r\subset[0,L]; \\ disjoint\ intervals}} \prod_{j=1}^{r} X(I_j) \tag{1.10}$$

which admits a cluster expansion, and is equal to $e^{L\varphi_L(\zeta)}$ for some function $\hat{\varphi}_L$ analytic in ζ. Further, we need that the second derivative in ζ of $\hat{\varphi}_L$ is sufficiently small in absolute value compared to the second derivative (in ζ) of $\ln Q$ in order to show the non-degeneracy of the covariance of the limit process $Y(t)$.

These two points, i.e., a) existence and analyticity of the free energy and b) non-degeneracy of the limiting covariance are to be checked depending on our model. Remaining arguments are the same as in [DH1,2], and we present them for the sake of completeness.

Finally, recent progress of understanding the fluctuation of interfaces provides us a beautiful and systematic approach using the renewal theory ([Ioffe], [KH]). For our problem, it seems also possible to follow this new line. However, what we have to check are the same, and at this stage we are not able to present our result in a compact form following this general approach.

Acknowledgement. The authors thank D.Ioffe for many valuable comments and stimulating discussions. He pointed out that this approach

is possible for the continuum Widom-Rowlinson model, which should be true, but we have not completed the whole story, yet.

§2. Local limit theorem

We will first show the existence of the limit (1.8) and its analyticity. Let $\Gamma \in \mathcal{S}_L$, $A(\Gamma) = (0,0)$, $B(\Gamma) = (L, k(\Gamma))$ be its starting and ending points. Let $\pi(\Gamma)$ be the backbone of Γ connecting $A(\Gamma)$ with $B(\Gamma)$. We decompose $\Gamma \setminus \pi(\Gamma)$ into connected components $\{C_j\}_{j=1}^{s}$. As in [DH2] we expand

$$\exp\left\{- \sum_{\substack{\Lambda\subset\Lambda_{L,\infty} \\ \Lambda\, i\, \Gamma}} \Phi(\Lambda)\right\} = \sum_{n=0}^{\infty} \sum_{\substack{\Lambda_1,\dots,\Lambda_n\subset\Lambda_{L,\infty} \\ \Lambda_\nu\, i\, \Gamma}} \prod_{\nu=1}^{n} (e^{-\Phi(\Lambda_\nu)} - 1).$$

Then

$$\sum_{\Gamma\in\mathcal{S}_L} e^{\mu\zeta k(\Gamma)} W(\Gamma) \tag{2.1}$$

$$= \sum_{k=-\infty}^{\infty} \sum_{\substack{\pi;(0,0)\to(L,k) \\ self-avoiding \\ \pi\subset\Lambda_{L,\infty}}} \sum_{\substack{C_1,\dots,C_s;compatible \\ C_\nu\, i\, \pi,\ C_\nu\cap\pi=\emptyset \\ \pi;backbone\ of\ \pi\cup C_1\cup\dots\cup C_s}} \sum_{\substack{\Lambda_1,\dots\Lambda_t; \\ \Lambda_\alpha isconnected \\ \Lambda_\alpha\, i\, \pi\cup C_1\cup\dots\cup C_s}}$$

$$e^{\mu\zeta k} e^{-\mu|\pi|+N(\pi,C_1,\dots,C_s)\ln 2-\mu\sum_{\nu=1}^{s}|C_\nu|} \prod_{\alpha=1}^{t} (e^{-\Phi(\Lambda_\alpha)} - 1),$$

where $N(\pi, C_1, \dots, C_s)$ denotes the number of holes of $\pi \cup \cup_{\nu=1}^{s} C_\nu$.

polymers

Defining polymers is to cut the separating contour Γ into elementary pieces according to the additional information of $\{\Lambda_\alpha\}$. A simplest way to do it would be to cut γ at lines $\{x^1 = \ell + \frac{1}{2}\}$ of dual lattice such that they intersect only one edge of Γ and intersection with edges of Λ_α's is empty. But the resulting pieces, say polymers, do interact. Even a "simple polymer" can interact with some polymers.

For example, a part of Γ like Fig 1 will be separated into two parts: one having $\sqsubset$ shape and one point to the right of it. If instead of one point, there comes a simple polymer of height three to the right of $\sqsubset$, then they are put together and there is no natural way to cut them (Fig. 2).

Thus, in a natural way of cutting procedure, Γ will be cut into interacting polymers. This causes us to introduce a polymer chain below, working with which we can use usual cluster expansion. The idea is to

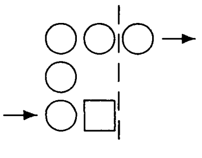

Fig. 1

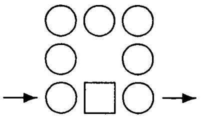

Fig. 2

treat a cluster of polymers interacting each other possibly through simple polymers which are at neiboring sites of such 'active' polymers.

Let $\hat{l} \leq \hat{r}$ be positive integers. A polymer ξ with base $[\hat{l}, \hat{r}]$ is a collection $\xi = (\gamma, C_1, \ldots, C_s, \Lambda_1, \ldots, \Lambda_t)$ such that

(a) γ is a self-avoiding path in $\{\hat{l} \leq x^1 \leq \hat{r}\}$ starting from $(\hat{l}, 0)$ and ending at a point $(\hat{r}, k)$ in $\{x^1 = \hat{r}\}$. Here, we understand γ as an edge set.

(b) $\{C_\nu\}_{\nu=1}^s$ is a compatible family of connected subsets of $\{x \in \Lambda_{L,\infty}; \hat{l} \leq x^1 \leq \hat{r}\}$ such that
   (b-1) $C_\nu \cap V(\gamma) = \emptyset$, where $V(\gamma)$ is the set of vertices in γ.
   (b-2) $C_\nu \cup V(\gamma)$ is connected.
   (b-3) γ is the backbone of $\gamma \cup C_1 \cup \ldots \cup C_s$ with starting point $(\hat{l}, 0)$ and endpoint $(\hat{r}, k)$.

(c) $\{\Lambda_\alpha\}_{\alpha=1}^t$ is a collection of connected subsets of $\{x \in \Lambda_{L,\infty}; \hat{l} \leq x^1 \leq \hat{r}\}$ such that

$$\Lambda_\alpha \, i \, V(\gamma) \cup \cup_{\nu=1}^s C_\nu .$$

Besides these conditions, we need a technical condition for a polymer. This condition is to subtract 'simple polymers' from the phase separating contour Γ as much as possible.

An edge e is called an *edge of* ξ if

$$e \in \gamma \cup \mathcal{E}(\cup_{\nu=1}^s C_\nu \cup \cup_{\alpha=1}^t \Lambda_\alpha) \cup \mathcal{E}(\gamma, \cup_{\nu=1}^s C_\nu \cup \cup_{\alpha=1}^t \Lambda_\alpha),$$

where $\mathcal{E}(\cup_{\nu=1}^s C_\nu \cup \cup_{\alpha=1}^t \Lambda_\alpha)$ is the set of nearest neighbor edges in $\cup_{\nu=1}^s C_\nu \cup \cup_{\alpha=1}^t \Lambda_\alpha$, and $\mathcal{E}(\gamma, \cup_{\nu=1}^s C_\nu \cup \cup_{\alpha=1}^t \Lambda_\alpha)$ is the set of edges that connect γ with $\cup_{\nu=1}^{s} C_\nu \cup \cup_{\alpha=1}^{t} \Lambda_\alpha$. An edge $e = \{x, y\}$ of ξ is *not admissible* if it is a horizontal edge in $\mathcal{E}(\gamma, \cup_{\nu=1}^{s} C_\nu \cup \cup_{\alpha=1}^{t} \Lambda_\alpha)$, such that

(1) The left vertex x is in a connected component D of $\cup_{\nu=1}^{s} C_\nu \cup \cup_{\alpha=1}^{t} \Lambda_\alpha$ and the right vertex y is in $V(\gamma)$,
(2) further, there exists a horizontal edge $e' = \{x', y'\}$ of ξ such that $x' \in V(\gamma)$ and $y' \in D$, where x' is the left vertex of e'.

Other edges of ξ are *admissible.* Also, we identify an edge $\{x, y\}$ of $\mathbf{Z}^2$ with the line segment connecting x and y. Now we introduce the remaining condition (d) for a polymer ξ.

(d) If $\hat{l} < \hat{r}$, then for $\hat{l} \leq j < \hat{r}$, $j \in \mathbf{N}$, the line $\ell_j = \{x^1 = j + \frac{1}{2}\}$ intersects at least two admissible edges of ξ.

We call γ the *backbone* of ξ. For two disjoint self-avoiding paths γ_1, γ_2 such that the starting point of γ_2 is nearest neighbor of the endpoint of γ_1, we can define the concatenation $\gamma_1 \circ \gamma_2$ of these paths by simply connecting them.

Let $\xi = (\gamma, C_1, \ldots, C_u, \Lambda_1, \ldots, \Lambda_v)$ and $\xi' = (\gamma', C_1', \ldots, C_w', \Lambda_1', \ldots, \Lambda_z')$ be two polymers with bases $[\hat{l}, \hat{r}]$ and $[\hat{l}', \hat{r}']$ $(\hat{l} \leq \hat{l}')$, respectively. We say that ξ and ξ' are *compatible* if either of the following conditions holds;

(1) $\hat{r} + 1 < \hat{l}'$,

(2) $\hat{l}' = \hat{r} + 1$, the backbone of

$$\tilde{\Gamma} := \gamma \cup C_1 \cup \ldots \cup C_u \cup (\gamma' + (0, k(\gamma))) \cup (C_1' + (0, k(\gamma))) \cup \ldots \cup (C_w' + (0, k(\gamma)))$$

is the concatenation $\gamma \circ (\gamma' + (0, k(\gamma)))$, and connected components of the set $\tilde{\Gamma} \setminus \gamma \circ (\gamma' + (0, k(\gamma)))$ are $\{C_1, \ldots, C_u, C_1' + (0, k(\gamma)), \ldots, C_w' + (0, k(\gamma))\}$. Here, $k(\gamma)$ is the hight of the endpoint of γ.

The family $\{\xi_p\}_{p=0}^{n+1}$ is compatible if ξ_p and $\xi_{p'}$ $(p \neq p')$ are compatible.

Let π be a self-avoiding path in $\Lambda_{L,\infty}$ connecting $(0,0)$ with $(L, k(\pi))$, $\{C_\nu\}_{\nu=1}^{s}$ be a compatible family of connected subsets of $\Lambda_{L,\infty}$ such that

(1) $C_\nu \, i \, \pi$ and $C_\nu \cap \pi = \emptyset$,
(2) π is the backbone of $V(\pi) \cup \cup_{\nu=1}^{s} C_\nu$.

Let also $\{\Lambda_\alpha\}_{\alpha=1}^{t}$ be a collection of connected subsets of $\Lambda_{L,\infty}$ such that $\Lambda_\alpha \, i \, \pi \cup \cup_{\nu=1}^{s} C_\nu$ for each α. We say that the line $\ell_j = \{x^1 = j + \frac{1}{2}\}$ $(0 \leq j \leq L-1)$ is a *cutting line* of $(\pi, \{C_\nu\}_{\nu=1}^{s}, \{\Lambda_\alpha\}_{\alpha=1}^{t})$ if ℓ_j intersects only one admissible edge of $(\pi, \{C_\nu\}_{\nu=1}^{s}, \{\Lambda_\alpha\}_{\alpha=1}^{t})$.

Let $\ell_0 < \ell_{j_1} < \ldots < \ell_{j_n} < \ell_{j_{n+1}} = \ell_{L-1}$ be all the cutting lines of $(\pi, \{C_\nu\}_{\nu=1}^{s}, \{\Lambda_\alpha\}_{\alpha=1}^{t})$. For each $m \in \{0, 1, \ldots, n+1\}$, there is a unique edge $e_m = \{B_m, A_{m+1}\}$ of π which intersects ℓ_{j_m}. Let γ_m be the portion of π starting from A_m and ending at B_m. Also let $\{C_\nu^{(m)}\}_{\nu=1}^{s(m)}$ and $\{\Lambda_\alpha^{(m)}\}_{\alpha=1}^{t(m)}$ be the set of elements of $\{C_\nu\}_{\nu=1}^{s}$ and $\{\Lambda_\alpha\}_{\alpha=1}^{t}$ such that they are subsets of $[j_{m-1}+1, j_m] \times (-\infty, \infty) \cap \mathbf{Z}^2$. Then $A_m = (j_{m-1}+1, p)$ for some $p \in \mathbf{Z}$. Thus we obtain the m-th polymer ξ_m by setting

$$\xi_m = (\gamma_m - (0,p), \{C_\nu^{(m)} - (0,p)\}_{\nu=1}^{s(m)}, \{\Lambda_\alpha^{(m)} - (0,p)\}_{\alpha=1}^{t(m)}).$$

By definition, $\{\xi_0, \xi_1, \dots, \xi_{n+1}\}$ are compatible.

For a polymer $\xi_m = (\gamma_m, \{C_\nu^{(m)}\}, \{\Lambda_\alpha^{(m)}\})$, let $k_m = k(\xi_m) = k(\gamma_m)$ be the hight of the endpoint of the self-avoiding path γ_m. Then the hight $k(\pi)$ of the endpoint of the original path π is given by

$$k(\pi) = \sum_{m=0}^{n+1} k(\gamma_m).$$

For a polymer $\xi = (\gamma, \{C_\nu\}_{\nu=1}^{u}, \{\Lambda_\alpha\}_{\alpha=1}^{v})$, set

$$\Psi(\xi) = e^{-\mu|\gamma| + N^*(\gamma, C_1, \dots, C_u)\ln 2 - \mu \sum_{\nu=1}^{u} |C_\nu|} \times \prod_{\alpha=1}^{v} (e^{-\Phi(\Lambda_\alpha)} - 1), \tag{2.2}$$

Where

$$\begin{aligned} N^*(\gamma, C_1, \dots, C_s) &= N(\gamma, C_1, \dots, C_s) \\ &+ N_l(\gamma, C_1, \dots, C_s) + N_r(\gamma, C_1, \dots, C_s) \end{aligned}$$

and $N_l(\gamma, C_1, \dots, C_s)$ is the number of new holes created by $V(\gamma) \cup \cup_{\nu=1}^{s} C_\nu$ and the line $\{x^1 = \hat{l} - 1\}$, where *base*$(\xi) = [\hat{l}, \hat{r}]$. Similarly, $N_r(\gamma, C_1, \dots, C_s)$ is the number of new holes created by $V(\gamma) \cup \cup_{\nu=1}^{s} C_\nu$ and the line $\{x^1 = \hat{r} + 1\}$.

A polymer ξ is called *simple* if *base*(ξ) is one point and $\xi = (\gamma, \emptyset, \emptyset)$. Thus, the weight $\Psi(\xi)$ is given by $\Psi(\xi) = e^{-\mu|\gamma|}$. A polymer ξ is called *decorated* if it is not simple.

A decorated polymer $\xi = (\gamma, \{C_\nu\}, \{\Lambda_\alpha\})$ with *base*$(\xi) = [\hat{l}, \hat{r}]$ is said *r-active* if there exists a simple polymer $\xi_1 = (\gamma_1, \emptyset, \emptyset)$ with *base*$(\xi_1) = \{\hat{r} + 1\}$ such that ξ_1 is incompatible with ξ or the concatenation of γ and γ_1 together with $\cup_\nu C_\nu$ produces a new hole. ξ is said *l-active* if there exists a simple polymer $\xi_2 = (\gamma_2, \emptyset, \emptyset)$ with *base*$(\xi_2) = \{\hat{l} - 1\}$ such that ξ_2 is incompatible with ξ or the concatenation of γ_2 and γ together with $\cup_\nu C_\nu$ produces a new hole. If ξ is both r-active and l-active, we call it *bi-active*. A *polymer chain* is a family of decorated polymers $\mathcal{C} = \{\xi_1, \dots, \xi_m\}$ such that

(1) $\{\xi_1, \dots, \xi_n\}$ are compatible.

(2) If $base(\xi_u) = [\hat{l}_u, \hat{r}_u]$, $1 \le u \le n$, then $\hat{l}_{u+1} = \hat{r}_u + 1$ or $\hat{r}_u + 2$ for every u.

(3) If $\hat{l}_{u+1} = \hat{r}_u + 2$ for some u, then ξ_u is r-active and ξ_{u+1} is l-active.

Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two polymer chains. We say that $\mathcal{C}_1$ and $\mathcal{C}_2$ are *compatible* if $\mathcal{C}_1 \cup \mathcal{C}_2$ is a compatible family of polymers, but it is not a polymer chain.

For a polymer chain $\mathcal{C} = \{\xi_1, \dots, \xi_m\}$, let

$$base(\mathcal{C}) = base(\xi_1) \cup \dots \cup base(\xi_m).$$

For a polymer ξ, we define

$$\hat{\Psi}(\xi; \zeta) := e^{\mu\zeta k(\xi)} \Psi(\xi) Q(\zeta)^{-|base(\xi)|},$$

where $|base(\xi)| = \hat{r} - \hat{l} + 1$ when $base(\xi) = [\hat{l}, \hat{r}]$, and $Q(\zeta)$ is the generating function of the hight of the endpoint of a simple polymer ;

$$Q(\zeta) = e^{-\mu} \sum_{k=-\infty}^{\infty} e^{\mu\zeta k} e^{-|k|\mu}.$$

Also, for a polymer chain $\mathcal{C} = \{\xi_1, \dots, \xi_m\}$, we put

$$\mathbf{F}_{\hat{\Psi}}(\mathcal{C}; \zeta) := \prod_{u=1}^{m} \hat{\Psi}(\xi_u; \zeta) \times \mathcal{J}_l(\xi_1) \mathcal{J}_r(\xi_m) \prod_{u=1}^{m-1} \mathcal{J}(\xi_u, \xi_{u+1}),$$

where for $base(\xi) = [\hat{l}, \hat{r}]$ and $base(\xi_1) = [c, d]$ with $c > \hat{r}$, $\mathcal{J}_l, \mathcal{J}_r, \mathcal{J}$ are defined in the following way.

$$\mathcal{J}_l(\xi) = \begin{cases} \displaystyle\sum_{\xi' c \xi}^{\hat{l}-1} \hat{\Psi}(\xi'; \zeta) 2^{N(\xi', \xi) - N_l(\xi, C_1, \dots, C_s)} & \text{if } \xi \text{ is l-active} \\ 1, & \text{otherwise,} \end{cases}$$

where $\displaystyle\sum_{\xi' c \xi}^{\hat{l}-1}$ means over simple polymers $\xi' = (\gamma', \emptyset, \emptyset)$ with base $\{\hat{l} - 1\}$ compatible with ξ, and $N(\xi', \xi)$ is the number of new holes created by the concatenation of γ' and γ together with $\cup_\nu C_\nu$, which is not larger than $N_l(\gamma, C_1, \dots, C_s)$. Similarly,

$$\mathcal{J}_r(\xi) = \begin{cases} \displaystyle\sum_{\xi' c \xi}^{\hat{r}+1} \hat{\Psi}(\xi'; \zeta) 2^{N(\xi, \xi') - N_r(\gamma, C_1, \dots, C_s)}, & \text{if } \xi \text{ is r-active} \\ 1, & \text{otherwise,} \end{cases}$$

and $\mathcal{J}(\xi,\xi_1)$ is defined in two cases.

(i) If $c = \hat{r}+2$, ξ is r-active and ξ_1 is l-active, then

$$\mathcal{J}(\xi,\xi_1) = \sum_{\xi' c\, \xi,\xi_1}^{\hat{r}+1} \hat{\Psi}(\xi';\zeta) 2^{N(\xi,\xi')+N(\xi',\xi_1)-N_r(\gamma,C_1,\dots,C_s)-N_l(\gamma_1,\tilde{C}_1,\dots,\tilde{C}_{s_1})},$$

(ii) If $c = \hat{r}+1$, and ξ and ξ' are compatible, then

$$\mathcal{J}(\xi,\xi_1) = 2^{N(\xi,\xi_1)-N_r(\gamma,C_1,\dots,C_s)-N_l(\gamma_1,\tilde{C}_1,\dots,\tilde{C}_{s_1})}.$$

Let $\mathcal{K}_L$ be the set of all decorated polymers with base in $[0,L]$, and $\mathcal{CP}_L$ be the set of polymer chains with base in $[0,L]$. Then we have

$$\frac{1}{Q(\zeta)^L}\sum_{\Gamma\in\mathcal{S}_L} e^{\mu\zeta k(\Gamma)}W(\Gamma) = \sum_{\substack{\mathcal{C}_1,\dots,\mathcal{C}_r\in\mathcal{CP}_L;\\ compatible}} \prod_{i=1}^{r}\mathbf{F}_{\hat{\Psi}}(\mathcal{C}_i;\zeta). \tag{2.3}$$

Lemma 2.1 Let $\delta>0$ be given. Then there exists $\mu_4>\mu_0$ such that for $\mu>\mu_4$, the free energy $\varphi(\zeta)$ in (1.8) exists and is analytic in ζ if $Re\zeta < 1-\frac{\delta}{\mu}$.

Proof. It is sufficient to show that

$$\frac{1}{L}\ln \sum_{\substack{\mathcal{C}_1,\dots,\mathcal{C}_r\in\mathcal{CP}_L;\\ compatible}} \prod_{i=1}^{r}\mathbf{F}_{\hat{\Psi}}(\mathcal{C}_i;\zeta)$$

converges as $L\to\infty$ and its limit $\hat{\varphi}(\zeta)$ is analytic for $Re\zeta < 1-\frac{\delta}{\mu}$. Then we have

$$\varphi(\zeta) = \hat{\varphi}(\zeta) + \ln Q(\zeta),$$

which is analytic in this region.

In order to verify the convergence and analyticity, we have to check that there exist functions $c^*, d^* : \mathcal{CP} = \{\mathcal{C};\ \text{polymer chain}\} \to [0,\infty)$ such that

$$\sum_{\mathcal{C}\in\mathcal{CP};\mathcal{C}\, i\, \mathcal{C}_0} e^{c^*(\mathcal{C})+d^*(\mathcal{C})}|\mathbf{F}_{\hat{\Psi}}(\mathcal{C};\zeta)| \le c^*(\mathcal{C}_0) \tag{2.4}$$

for any polymer chain $\mathcal{C}_0$ and for any $\zeta\in\mathbf{C}$ with $Re\zeta < 1-\frac{\delta}{\mu}$ (see e.g. [KP]). For a decorated polymer $\xi = (\gamma,\{C_\nu\},\{\Lambda_\alpha\})$, we put $c(\xi) = 3|base(\xi)|$ and

$$d(\xi) = \begin{cases} (\mu-\mu_4)|base(\xi)| + \frac{\delta}{6}|\gamma| - (\mu-\mu_2-1), & \text{if } |base(\xi)|\ge 2,\\ (\mu-\mu_4)|base(\xi)| + \frac{\delta}{6}|\gamma|, & \text{if } |base(\xi)| = 1. \end{cases}$$

Then we set

$$c^*(\mathcal{C}) = \sum_{\xi \in \mathcal{C}} c(\xi), \quad d^*(\mathcal{C}) = \sum_{\xi \in \mathcal{C}} d(\xi)$$

The constant μ_4 is specified later. We will first show that

$$\sum_{\xi \in \mathcal{K}_L; \xi i\, \xi_0} e^{c(\xi)+d(\xi)} |\hat{\Psi}(\xi; \zeta)| \leq c(\xi_0) \tag{2.5}$$

for every polymer ξ_0. Note first that

$$|\gamma| = N_v(\gamma) + N_h(\gamma) + 1, \tag{2.6}$$

where $N_v(\gamma)$ is the number of vertical edges in γ, and $N_h(\gamma)$ is the number of horizontal edges in γ. Also, by definition of decorated polymers, if *base*(ξ) is one point, then

$$N_h(\gamma) + \sum_{\nu=1}^{s} |C_\nu| + \sum_{\alpha=1}^{t} |\Lambda_\alpha| \geq 1, \tag{2.7a}$$

since either $\{C_\nu\}$ or $\{\Lambda_\alpha\}$ is non-empty if *base*(ξ) is one point. If $|base(\xi)| \geq 2$, then we have

$$N_h(\gamma) + \sum_{\nu=1}^{s} |C_\nu| + \sum_{\alpha=1}^{t} |\Lambda_\alpha| \geq 2(|base(\xi)| - 1). \tag{2.7b}$$

Let γ be a self-avoiding path such that it is the backbone of some decorated polymer with base $I = [\hat{l}, \hat{r}]$. We estimate the following sum.

$$G(\gamma) := \sum_{\xi; \gamma \, is\, the\, backbone\, of\, \xi} |\Psi(\xi) e^{\mu k(\gamma)\zeta}|.$$

From (1.4), $|\Phi(\Lambda)| \leq e^{-(\mu-\mu_0)|\Lambda|} < 1$ and therefore we have

$$|e^{-\Phi(\Lambda)} - 1| \leq e^{-(\mu-\mu_0-1)|\Lambda|}.$$

Using this, if $\hat{l} = \hat{r}$, i.e., $|I| = 1$, then we have $N^*(\gamma, C_1, \dots, C_s) = 0$ and

$$\begin{aligned} G(\gamma) &\le e^{-\mu|\gamma|} e^{\mu k(\gamma) Re\zeta} \sum_{\{C_\nu\};C_\nu\, i\, \gamma} e^{-\mu \sum_\nu |C_\nu|} \\ &\quad \times \sum_{\{\Lambda_\alpha\};\Lambda_\alpha\, i\, \gamma\cup C_1\cup\cdots\cup C_s} e^{-(\mu-\mu_0-1)\sum_\alpha |\Lambda_\alpha|} \\ &\le e^{-\mu|\gamma|+\mu k(\gamma)Re\zeta-(\mu-\mu_2-1)} \\ &\quad \times \sum_{\{C_\nu\};C_\nu\, i\, \gamma} e^{-\mu_2 \sum_\nu |C_\nu|} \\ &\quad \times \sum_{\{\Lambda_\alpha\};\Lambda_\alpha\, i\, \gamma\cup C_1\cup\cdots\cup C_s} e^{-(\mu_2-\mu_0)\sum_\alpha |\Lambda_\alpha|} \end{aligned} \tag{2.8}$$

The summation over $\{\Lambda_\alpha\}$ is estimated as follows.

$$\begin{aligned} &\sum_{\{\Lambda_\alpha\};\Lambda_\alpha\, i\, \gamma\cup C_1\cup\cdots\cup C_s} e^{-(\mu_2-\mu_0)\sum_\alpha |\Lambda_\alpha|} \\ &\le \sum_{t=0}^{\infty} \frac{1}{t!} \sum_{\Lambda_1\, i\, \gamma\cup C_1\cup\cdots\cup C_s} \cdots \sum_{\Lambda_t\, i\, \gamma\cup C_1\cup\cdots\cup C_s} e^{-(\mu_2-\mu_0)\sum_\alpha |\Lambda_\alpha|} \\ &\le \exp\left\{ 4|\gamma\cup C_1\cup\cdots\cup C_s| \sum_{\Lambda\ni 0;connected} e^{-(\mu_2-\mu_0)|\Lambda|} \right\} \\ &= \exp\left\{ (|\gamma| + \sum_\nu |C_\nu|) g_1(\mu_2,\mu_0) \right\}. \end{aligned}$$

Since there exist constants $K_1, \kappa > 0$ such that the number N_n of connected sets of n points in $\mathbf{Z}^2$ which contain the origin is bounded as

$$N_n \le K_1 \kappa^n \quad (n \ge 1),$$

we know that $g_1(\mu_2,\mu_0) = 4\sum_{\Lambda\ni 0;connected} e^{-(\mu_2-\mu_0)|\Lambda|}$ goes to zero exponentially fast as $\mu_2 \to \infty$. Thus, summing up the RHS of (2.8) over $\{\Lambda_\alpha\}$'s we obtain

$$\begin{aligned} G(\gamma) &\le e^{-(\mu-g_1(\mu_2,\mu_0))|\gamma|+\mu k(\gamma)Re\zeta} \\ &\quad \times e^{-(\mu-\mu_2-1)} \sum_{\{C_\nu\};C_\nu\, i\, \gamma} e^{-(\mu_2-g_1(\mu_2,\mu_0))\sum_\nu |C_\nu|} \\ &\le e^{-(\mu-g_1(\mu_2,\mu_0)-g_2(\mu_2,\mu_0))|\gamma|+\mu k(\gamma)Re\zeta-(\mu-\mu_2-1)}, \end{aligned}$$

where $g_2(\mu_2, \mu_0) = 4\sum_{C\ni 0;connected} e^{-(\mu_2 - g_1(\mu_2,\mu_0))|C|}$. If $\hat{r} > \hat{l}$, i.e., $|I| \geq 2$, then since $N^*(\gamma, C_1, \dots, C_s) \leq N_h(\gamma) + \sum_\nu |C_\nu|$, we have from (2.7b) as in (2.8),

$$
\begin{aligned}
(2.9) \qquad G(\gamma) \leq & e^{-\mu|\gamma|+\mu k(\gamma)Re\zeta} \sum_{\{C_\nu\};C_\nu\, i\, \gamma} e^{-\mu\sum_\nu |C_\nu|} 2^{N^*(\gamma,C_1,\dots,C_s)} \\
& \times \sum_{\{\Lambda_\alpha\};\Lambda_\alpha\, i\, \gamma\cup C_1\cup\dots C_s} e^{-(\mu-\mu_0-1)\sum_\alpha |\Lambda_\alpha|} \\
& \leq e^{-\mu|\gamma|+\mu k(\gamma)Re\zeta-(\mu-\mu_2-1)(2|I|-N_h(\gamma)-2)} 2^{N_h(\gamma)} \\
& \times \sum_{\{C_\nu\};\, C_\nu\, i\, \gamma} e^{-(\mu_2-\ln 2)\sum_\nu |C_\nu|} \\
& \times \sum_{\{\Lambda_\alpha\};\Lambda_\alpha\, i\, \gamma\cup C_1\cup\dots\cup C_s} e^{-(\mu_2-\mu_0)\sum_\alpha |\Lambda_\alpha|} \\
& \leq e^{-(\mu-g_1(\mu_2,\mu_0))|\gamma|+\mu k(\gamma)Re\zeta-(\mu-\mu_2-1)(2|I|-N_h(\gamma)-2)} \\
& \times \sum_{\{C_\nu\}} e^{-(\mu_2-g_1(\mu_2,\mu_0)-\ln 2)\sum_\nu |C_\nu|} e^{N_h(\gamma)\ln 2} \\
& \leq e^{-(\mu-g_1(\mu_2,\mu_0)-g_3(\mu_2,\mu_0))|\gamma|+\mu k(\gamma)Re\zeta} \\
& \times e^{-(\mu-\mu_2-1)(2|I|-N_h(\gamma)-2)+N_h(\gamma)\ln 2},
\end{aligned}
$$

where $g_3(\mu_2, \mu_0) = 4\sum_{C\ni 0;connected} e^{-(\mu_2 - g_1(\mu_2,\mu_0)-\ln 2)}$. We take μ_2 sufficiently large so that $g_1(\mu_2, \mu_0)$, $g_2(\mu_2, \mu_0)$ and $g_3(\mu_2, \mu_0)$ are all smaller than $\frac{\delta}{4}$.

Assume that $Re\zeta < 1 - \frac{\delta}{\mu}$. Then since $N_v(\gamma) \geq |k(\gamma)|$, from (2.6) we have

$$
(2.10) \qquad G(\gamma) \leq e^{-\frac{\delta}{2}N_v(\gamma)-(\mu_2-\frac{\delta}{2})(N_h(\gamma)+1)-(\mu-\mu_2-1)(2|I|-1)},
$$

if $|I| \geq 2$, and

$$
(2.11) \qquad G(\gamma) \leq e^{-\frac{\delta}{2}N_v(\gamma)-(\mu_2-\frac{\delta}{2})-2(\mu-\mu_2-1)}
$$

if $|I| = 1$. Since $c(\xi)$ and $d(\xi)$ depend only on the backbone γ, we write them $c(\gamma)$ and $d(\gamma)$. Then

$$
\begin{aligned}
(2.12) \qquad & \sum_{\xi;\gamma\, is\, the\, backbone\, of\, \xi} \left|\Psi(\xi)e^{\mu k(\gamma)\zeta}\right| e^{c(\xi)+d(\xi)} \\
& = G(\gamma)e^{c(\gamma)+d(\gamma)} \\
& \leq e^{-\frac{\delta}{3}N_v(\gamma)-(\mu_2-\frac{2\delta}{3})(N_h(\gamma)+1)} e^{-(\mu+\mu_4-2\mu_2-5)|base(\gamma)|},
\end{aligned}
$$

where $base(\gamma) = base(\xi)$ for any ξ such that γ is the backbone of ξ. Therefore we have for a fixed interval I,

$$\tag{2.13} \begin{aligned} &\sum_{base(\gamma)=I} G(\gamma)e^{c(\gamma)+d(\gamma)} \\ &\leq e^{-(\mu+\mu_4-2\mu_2-5)|I|} \sum_{base(\gamma)=I} e^{-\frac{\delta}{3}N_v(\gamma)-(\mu_2-\frac{2\delta}{3})(N_h(\gamma)+1)}. \end{aligned}$$

To estimate the RHS of (2.13) we separate γ into fragments following the idea of [DKS]. Let $\gamma = \{x_0, x_1, \dots, x_n\}$ be a self-avoiding path with $base(\gamma) = I$. Let $j_0 = 0$, and for $i \geq 1$, let

$$j_i := \min\{j > j_{i-1};\ \{x_{j-1}, x_j\} \text{ is a horizontal edge}\}.$$

Each vertical part $\{x_{j_{i-1}}, x_{j_{i-1}+1}, \dots, x_{j_i-1}\}$ of γ with the direction of the exit vector $\{x_{j_i-1}, x_{j_i}\}$ is called a fragment. For a fragment $f = \{\bar{x}_0, \bar{x}_1, \dots, \bar{x}_p\}$ with exit direction $e(f)$, we define

$$W(f) := e^{-\frac{\delta}{3}N_v(f)-(\mu_2-\frac{2\delta}{3})} = e^{-\frac{p\delta}{3}-(\mu_2-\frac{2\delta}{3})}.$$

Then the decomposition of γ into fragments $\{f_1, \dots, f_r\}$ leads to the identity

$$e^{-\frac{\delta}{3}N_v(\gamma)-(\mu_2-\frac{2\delta}{3})(N_h(\gamma)+1)} = \prod_{j=1}^{r} W(f_j).$$

Therefore we have

$$\begin{aligned} \sum_{\gamma;base(\gamma)=I} e^{-\frac{\delta}{3}N_v(\gamma)-(\mu_2-\frac{2\delta}{3})(N_h(\gamma)+1)} &= \sum_{r=|I|}^{\infty} \sum_{f_1,\dots,f_r} \prod_{j=1}^{r} W(f_j) \\ &\leq \sum_{r=|I|}^{\infty} \left(2 \sum_{k=-\infty}^{\infty} e^{-\frac{\delta}{3}|k|}\right)^r \times e^{-(\mu_2-\frac{2\delta}{3})r} \\ &= \frac{R(\mu_2,\delta)^{|I|}}{1-R(\mu_2,\delta)}, \end{aligned}$$

if μ_2 is sufficiently large. Thus, if $Re\zeta < 1 - \frac{\delta}{\mu}$ and $\mu > \mu_2$, where μ_2 is sufficiently large, we have

$$\sum_{base(\gamma)=I} G(\gamma)e^{c(\gamma)+d(\gamma)} \leq e^{-(\mu+\mu_4-2\mu_2-5)|I|} \frac{R(\mu_2,\delta)^{|I|}}{1-R(\mu_2,\delta)}.$$

Since

$$|Q(\zeta)| = e^{-\mu}\left|\frac{\sinh\mu}{\cosh\mu - \cosh\mu\zeta}\right| \geq \frac{e^{-\mu}\tanh\mu_2}{1+e^{-\delta}} := e^{-\mu-\mu_3} \tag{2.14}$$

if $Re\zeta < 1 - \frac{\delta}{\mu}$ and $\mu > \mu_2$, we have

$$\sum_{base(\xi)=I} |\hat{\Psi}(\xi;\zeta)|e^{c(\xi)+d(\xi)} \leq e^{-(\mu_4-2\mu_2-\mu_3-5)|I|}\frac{R(\mu_2,\delta)^{|I|}}{1-R(\mu_2,\delta)}. \tag{2.15}$$

Let $\mu_4 > 2\mu_2 + \mu_3 + 5$. For $\mu > \mu_4$ we will estimate the RHS of (2.5). Fix ξ_0 and write $base(\xi_0) = [\hat{l}, \hat{r}]$. Then we have

$$\begin{aligned}\sum_{\xi\, i\, \xi_0} |\hat{\Psi}(\xi;\zeta)|e^{c(\xi)+d(\xi)} &\leq \sum_{x\in[\hat{l}-1,\hat{r}+1]}\sum_{I\ni x}\frac{R(\mu_2,\delta)^{|I|}}{1-R(\mu_2,\delta)} \\ &= \frac{(\hat{r}-\hat{l}+3)}{1-R(\mu_2,\delta)}\sum_{k=1}^{\infty} kR(\mu_2,\delta)^k \\ &\leq 3|base(\xi_0)|\frac{R(\mu_2,\delta)}{\big(1-R(\mu_2,\delta)\big)^3} \leq c(\xi_0).\end{aligned}$$

if μ_2 is large. Thus, (2.5) is proved. From (2.5) to (2.4), we argue in the following way. We call a family of intervals $I_1 = [\hat{l}_1, \hat{r}_1], \dots, I_n = [\hat{l}_n, \hat{r}_n]$ *linked intervals* if for each $1 \leq u \leq n$, $\hat{r}_u < \hat{l}_{u+1} \leq \hat{r}_u + 2$ holds. The base of a polymer chain forms linked intervals. For a fixed polymer chain $\mathcal{C}_0$, let $[base(\mathcal{C}_0)] = [\hat{l}_0, \hat{r}_0]$ be the smallest interval including $base(\mathcal{C}_0)$. Then noting that the distance of $base(\mathcal{C}_0)$ and $base(\mathcal{C})$ is less than 2 if $\mathcal{C}_0$ and $\mathcal{C}$ are incompatible, we have

$$\begin{aligned}&\sum_{\mathcal{C}\, i\, \mathcal{C}_0} |\mathbf{F}_{\hat{\Psi}}(\mathcal{C};\zeta)|e^{c^*(\mathcal{C})+d^*(\mathcal{C})} \\ &\leq \sum_{x\in[\hat{l}_0-2,\hat{r}_0+2]}\sum_{n=1}^{\infty}\sum_{\substack{I_1,\dots,I_n\subset[0,L];\cup I_u\ni x\\ linked\, intervals,}}\sum_{\substack{\xi_1,\dots,\xi_n\in\mathcal{K}_L;\\ base(\xi_u)=I_u,1\leq u\leq n}} \\ &\prod_{u=1}^{n}\left[\hat{\Psi}(\xi_u;\zeta)e^{c(\xi_u)+d(\xi_u)}\right]\mathcal{J}_l(\xi_1)\mathcal{J}_r(\xi_n)\prod_{u=1}^{n-1}\mathcal{J}(\xi_u,\xi_{u+1})\end{aligned}$$

By definition and (2.14), there exists $\mu_3^* > 0 = \mu_3^*(\delta)$ such that $|\mathcal{J}_r|$, $|\mathcal{J}_l|$, $|\mathcal{J}|$ are all bounded by $e^{\mu_3^*}$ from above if $Re(\zeta) < 1 - \frac{\delta}{\mu}$. Therefore from the estimate (2.15), we have

$$\sum_{\substack{\xi_1,\dots,\xi_n\in\mathcal{K}_L;\\ base(\xi_u)=I_u,1\le u\le n}} \prod_{u=1}^{n}\left[\hat{\Psi}(\xi_u;\zeta)e^{c(\xi_u)+d(\xi_u)}\right]\mathcal{J}_l(\xi_1)\mathcal{J}_r(\xi_n)\prod_{u=1}^{n-1}\mathcal{J}(\xi_u,\xi_{u+1})$$

$$\le\prod_{u=1}^{n} e^{-(\mu_4-2\mu_2-\mu_3-2\mu_3^*-5)|I_u|}\frac{R(\mu_2,\delta)^{|I_u|}}{1-R(\mu_2,\delta)}.$$

Assuming that $\mu_4 > 2\mu_2+\mu_3+2\mu_3^*+5$, we have

$$\sum_{\mathcal{C}\,i\,\mathcal{C}_0}|\mathbf{F}_{\hat{\Psi}}(\mathcal{C};\zeta)|e^{c^*(\mathcal{C})+d^*(\mathcal{C})}$$

$$\le(\hat{r}_0-\hat{l}_0+4)\sum_{n=1}^{\infty}\sum_{u=1}^{n}\sum_{\substack{I_1,\dots,I_n\subset[0,L];I_u\ni x,\\ linked\,intervals}}\prod_{u=1}^{n}\frac{R(\mu_2,\delta)^{|I_u|}}{1-R(\mu_2,\delta)}$$

$$\le(\hat{r}_0-\hat{l}_0+4)\frac{R(\mu_2,\delta)}{(1-R(\mu_2,\delta))^3}\sum_{n=1}^{\infty}n(\frac{2R(\mu_2,\delta)}{(1-R(\mu_2,\delta))^2})^{n-1}$$

$$\le\frac{(\hat{r}_0-\hat{l}_0+4)}{2}$$

if μ_2 is large. Since $\sum_{\xi\in\mathcal{C}_0}|base(\xi)|\ge\max\left\{\frac{2}{3}[base(\mathcal{C}_0)],1\right\}$, the RHS of the above inequality is not larger than $c^*(\mathcal{C}_0)$.

This allows us to apply general theory of cluster expansion so that there exists a function

$$\mathbf{F}^T_{\hat{\Psi}} : \mathcal{P}_f(\mathcal{CP})\times\mathbf{C}\to\mathbf{C}$$

such that $\mathbf{F}^T_{\hat{\Psi}}$ is analytic for $Re\zeta<1-\frac{\delta}{\mu}$ and it satisfies

$$\sum_{\substack{\mathcal{C}_1,\dots,\mathcal{C}_r\in\mathcal{CP}_L;\\ compatible}}\prod\mathbf{F}_{\hat{\Psi}}(\mathcal{C}_i;\zeta)=\exp\left\{\sum_{\Delta\in\mathcal{P}_f(\mathcal{CP}_L)}\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta)\right\}\tag{2.16}$$

and

$$\sum_{\Delta\,i\,\mathcal{C}_0}|\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta)|e^{d^*(\Delta)}\le c^*(\mathcal{C}_0),\tag{2.17}$$

where $\mathcal{P}_f(\mathcal{CP}_L)$ is the collection of all finite subsets of $\mathcal{CP}_L$ and $d^*(\Delta)=\sum_{\mathcal{C}\in\Delta}d^*(\mathcal{C})$. If Δ is decomposed into two disjoint subsets Δ_1 and Δ_2 such that $\{\mathcal{C}_1,\mathcal{C}_2\}$ are compatible for every pair $\mathcal{C}_1\in\Delta_1$, $\mathcal{C}_2\in\Delta_2$, then $\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta)=0$. We call $\Delta\in\mathcal{P}_f(\mathcal{CP}_L)$ a *cluster* if there are no such decomposition $\Delta = \Delta_1 \cup \Delta_2$. Also, we note that $\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta)$ is invariant under horizontal translation of Δ. For $\Delta \in \mathcal{P}_f(\mathcal{CP})$, put $base(\Delta) = \cup_{\mathcal{C}\in\Delta} base(\mathcal{C})$. Then (2.16) and (2.17) implies that the limit

$$\begin{aligned}\hat{\varphi}(\zeta) &= \lim_{L\to\infty}\frac{1}{L}\ln \sum_{\mathcal{C}_1,\dots,\mathcal{C}_r\in\mathcal{CP}_L}\prod_{u=1}^{r}\mathbf{F}_{\hat{\Psi}}(\mathcal{C}_u;\zeta) \\ &= \sum_{\substack{\Delta\in\mathcal{P}_f(\mathcal{CP});[base(\Delta)]=[0,k] \\ for\ some\ k\geq 0}} \mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta)\end{aligned}$$

exists and analytic for $\zeta < 1 - \frac{\delta}{\mu}$ if $\mu > \mu_4$.

free energy for a joint distribution

Let $q \geq 1$, and let $0 < t_1 < \cdots < t_{q+1} = 1$. For $\underline{\zeta} = (\zeta_0, \zeta_1, \dots, \zeta_{q+1}) \in \mathbf{C}^{q+1}$, let

$$(2.18) \qquad \varphi^{(q)}(\underline{\zeta}; t_1, \dots, t_{q+1}) = \lim_{L\to\infty}\frac{1}{L}\ln\sum_{\Gamma\in\mathcal{S}_L} e^{\mu\underline{\zeta}\cdot\hat{X}_L^{(q)}(t_1,\dots,t_{q+1})}W(\Gamma)$$

if the limit exists. Here, the random vector $\hat{X}_L^{(q)}(t_1, \dots, t_{q+1})$ is defined by

(2.19)

$$\hat{X}_L^{(q)}(t_1,\dots,t_{q+1}) = \left(\frac{a(\pi(\Gamma))}{L}, X_L(\frac{\lfloor Lt_1\rfloor}{L}), \dots, X_L(\frac{\lfloor Lt_q\rfloor}{L}), X_L(1)\right).$$

With a slight change of the proof of Lemma 2.1, we can prove existence and analyticity of the limit $\varphi^{(q)}(\underline{\zeta}; t_1, \dots, t_{q+1})$. To be more precise, we decompose $a(\pi(\Gamma))$ into terms corresponding to polymers appearing in the decomposition of Γ. Let $\xi = (\gamma, \{C_\nu\}, \{\Lambda_\alpha\})$ be a polymer with base $[a, b]$. The area $area(\xi)$ is then defined by

$$\begin{aligned}area(\xi) =&\#\{x \in [\hat{l}, \hat{r}] \times [-M, M] \cap \mathbf{Z}^{2^*};\ x \text{ is below}\,\gamma\} \\ &- \#\{x \in [\hat{l}, \hat{r}] \times [-M, M] \cap \mathbf{Z}^{2^*};\ x \text{ is above}\,\gamma\}.\end{aligned}$$

This is independent of large M. For a $\Gamma \in \mathcal{S}_L$, denote $\mathcal{D}(\Gamma)$ all polymers, which obtained through any triple $(\pi(\Gamma), \{C_\nu\}, \{\Lambda_\alpha\})$ with its cutting lines, where $\{\Lambda_\alpha\}$ is taken over all families of connected sets such that $\Lambda_\alpha i\Gamma$ for each α. We have

$$(2.20) \qquad a(\pi(\Gamma)) = \sum_{\xi\in\mathcal{D}(\Gamma)}\{area(\xi) + k(\gamma)(L - \hat{r}(\xi))\},$$

where $base(\xi) = [\hat{l}(\xi), \hat{r}(\xi)]$ for $\xi \in \mathcal{D}(\Gamma)$. Therefore,

$$\begin{aligned}\underline{\zeta} \cdot \hat{X}_L^{(q)}(t_1, \dots, t_{q+1}) = \zeta_0 & \sum_{\xi \in \mathcal{D}(\Gamma)} \left\{ \frac{area(\xi)}{L} + k(\gamma)\left(1 - \frac{\hat{r}(\xi)}{L}\right) \right\} \\ & + \sum_{i=1}^{q+1} \zeta_i \sum_{\xi \in \mathcal{D}(\Gamma)} 1_{[\hat{r}(\xi) < Lt_i]} k(\gamma) \\ & + \sum_{i=1}^{q+1} \zeta_i \sum_{\xi \in \mathcal{D}(\Gamma)} 1_{[\hat{l}(\xi) \le Lt_i \le \hat{r}(\xi)]} k(\gamma; t_i L),\end{aligned}$$

where $k(\gamma; t_i L)$ is the maximal hight of the intersection of polygonal line γ and the vertical line $\{x^1 = t_i L\}$.

Proposition 2.2. Let $\mu > \mu_4$. If $\underline{\zeta}$ satisfies

$$(2.21) \quad \begin{cases} \max\{|Re(\zeta_0 + \zeta_{q+1})|, |Re\zeta_{q+1}|\} \le 1 - \frac{2\delta}{\mu}, & \\ |Re\zeta_i| \le \frac{\delta}{4(q+1)\mu}, & i = 1, 2, \dots, q, \end{cases}$$

then the limit $\varphi^{(q)}(\underline{\zeta}; t_1, \dots, t_{q+1})$ exists and is analytic in $\underline{\zeta}$.

Proof. Let ξ be a polymer with base $[\hat{l}(\xi), \hat{r}(\xi)] \subset [0, L]$. We decompose ξ into fragments $\{f_p\}_{p=1}^P$. The hight of a fragment $f = \{x_1, \dots, x_u\}$ is defined by

$$h(f) = x_u^2 - x_1^2$$

and the position of f is given by

$$pos(f) = x_1^1 = x_u^1.$$

Then we have as in [DH2],

$$area(\xi) = \sum_{p=1}^P h(f_p)(\hat{r}(\xi) - pos(f_p)).$$

Since $k(\gamma) = \sum_{p=1}^P h(f_p)$, we have

$$\frac{area(\xi)}{L} + k(\gamma)\left(1 - \frac{\hat{r}(\xi)}{L}\right) = \sum_{p=1}^P h(f_p)\left(1 - \frac{pos(f_p)}{L}\right).$$

Thus, we have

$$
\begin{aligned}
&\left|Re\left[\zeta_0\left(\frac{area(\xi)}{L}+k(\gamma)(1-\frac{\hat{r}(\xi)}{L})\right)+\sum_{i=1}^{q+1}\zeta_i 1_{[\hat{r}(\xi)<Lt_i]}k(\gamma)\right.\right.\\
&\quad\left.\left.+\sum_{i=1}^{q+1}\zeta_i 1_{[\hat{l}(\xi)\le Lt_i\le\hat{r}(\xi)]}k(\gamma;Lt_i)\right]\right|\\
\le&\left|Re(\zeta_0+\zeta_{q+1})\sum_{p=1}^{P}h(f_p)(1-\frac{pos(f_p)}{L})+Re\zeta_{q+1}\sum_{p=1}^{P}h(f_p)\frac{pos(f_p)}{L}\right|\\
&\quad+\sum_{i=1}^{q}|Re\zeta_i|N_v(\gamma)\\
\le&|Re(\zeta_0+\zeta_{q+1})|\sum_{p=1}^{P}|h(f_p)|(1-\frac{pos(f_p)}{L})+|Re\zeta_{q+1}|\sum_{p=1}^{P}|h(f_p)|\frac{pos(f_p)}{L}\\
&\quad+\sum_{i=1}^{q}|Re\zeta_i|N_v(\gamma)\\
\le&\left[\max\{|Re(\zeta_0+\zeta_{q+1})|,|Re\zeta_{q+1}|\}+\sum_{i=1}^{q}|Re\zeta_i|\right]N_v(\gamma).
\end{aligned}
$$

Set

$$
\begin{aligned}
X^{(L)}(\underline{\zeta};\xi)&=X^{(L)}_{t_1,\dots,t_{q+1}}(\underline{\zeta};\xi)\\
&=\zeta_0\left(\frac{area(\xi)}{L}+k(\gamma)(1-\frac{\hat{r}(\xi)}{L})\right)+\sum_{i=1}^{q+1}\zeta_i 1_{[\hat{r}(\xi)<Lt_i]}k(\gamma)\\
&\quad+\sum_{i=1}^{q+1}\zeta_i 1_{[\hat{l}(\xi)\le Lt_i\le\hat{r}(\xi)]}k(\gamma;Lt_i).
\end{aligned}
$$

As before, let

$$
(2.22)\qquad \hat{\Psi}(\xi;\ \underline{\zeta},t_1,\dots,t_{q+1})=\Psi(\xi)e^{\mu X^{(L)}(\underline{\zeta};\xi)}\prod_{\ell=\hat{l}(\xi)}^{\hat{r}(\xi)}Q^{-1}(\zeta_L(\ell)),
$$

where $\zeta_L(\ell)=\zeta_0(1-\frac{\ell}{L})+\sum_{i=1}^{q+1}\zeta_i 1_{[\ell\le Lt_i]}$. For simplicity we write $\hat{\Psi}(\xi;\underline{\zeta})$ for $\hat{\Psi}(\xi;\underline{\zeta};t_1,\dots,t_{q+1})$ for the moment. Then for a polymer chain $\mathcal{C}=$

$\{\xi_1, \dots, \xi_m\}$, we define $\mathbf{F}_{\hat{\Psi}}(\mathcal{C}; \underline{\zeta}) = \mathbf{F}_{\hat{\Psi}}(\mathcal{C}; \underline{\zeta}, t_1, \dots, t_{q+1})$ analogously to $\mathbf{F}_{\hat{\Psi}}(\mathcal{C}; \zeta)$. Namely,

$$\mathbf{F}_{\hat{\Psi}}(\mathcal{C}; \underline{\zeta}) = \prod_{u=1}^{m} \hat{\Psi}(\xi_u; \underline{\zeta}) \mathcal{J}_l^{(q)}(\xi_1) \mathcal{J}_r^{(q)}(\xi_m) \prod_{u=1}^{m-1} \mathcal{J}^{(q)}(\xi_u, \xi_{u+1}),$$

where $\mathcal{J}_l^{(q)}$, $\mathcal{J}_r^{(q)}$ and $\mathcal{J}^{(q)}$ are defined as $\mathcal{J}_l$, $\mathcal{J}_r$ and $\mathcal{J}$ by replacing $\hat{\Psi}(\xi; \zeta)$ with $\hat{\Psi}(\xi; \underline{\zeta})$. If $\underline{\zeta}$ satisfies (2.21), then $Q(\zeta_L(\ell))$ is analytic in $\underline{\zeta}$ and satisfies the estimate

$$|Q(\zeta_L(\ell))^{-1}| \leq e^{\mu + \mu_3} \quad \ell = 0, 1, \dots, L$$

if $\mu > \mu_2$. Therefore as in the proof of Lemma 2.1, for $\mu > \mu_4$ we have convergent cluster expansion:

(2.23)

$$\frac{1}{L} \ln \sum_{\substack{\mathcal{C}_1, \dots, \mathcal{C}_n \in \mathcal{CP}_L; \\ compatible}} \prod_{j=1}^{n} \mathbf{F}_{\hat{\Psi}}(\mathcal{C}_j; \underline{\zeta}) = \frac{1}{L} \sum_{\Delta \in \mathcal{P}_f(\mathcal{CP}_L)} \mathbf{F}_{\hat{\Psi}}^T(\Delta; \underline{\zeta}, t_1, \dots, t_{q+1})$$

such that $\mathbf{F}_{\hat{\Psi}}^T(\Delta; \underline{\zeta}, t_1, \dots, t_{q+1}) = 0$ unless Δ is a cluster, and (2.17) holds uniformly in $\underline{\zeta}$ satisfying (2.21). So, if (2.23) converges uniformly in $\underline{\zeta}$ satisfying (2.21), then the limit is analytic in this region.

For an interval $I \subset [0, L]$, set

$$\hat{\Xi}(I; \underline{\zeta}, t_1, \dots, t_{q+1}) := \sum_{\substack{\mathcal{C}_1, \dots, \mathcal{C}_m; \ compatible \\ base(\mathcal{C}_i) \subset I, \ 1 \leq i \leq m}} \prod_{i=1}^{m} \mathbf{F}_{\hat{\Psi}}(\mathcal{C}_i; \underline{\zeta}, t_1, \dots, t_{q+1}).$$

Then by cluster expansion we have

$$\ln \hat{\Xi}(I; \underline{\zeta}, t_1, \dots, t_{q+1}) = \sum_{\Delta \in \mathcal{P}_f(\mathcal{CP}); base(\Delta) \subset I} \mathbf{F}_{\hat{\Psi}}^T(\Delta; \underline{\zeta}, t_1, \dots, t_{q+1})$$

if $\underline{\zeta}$ satisfies (2.21), where $base(\Delta) = \cup_{\mathcal{C} \in \Delta} base(\mathcal{C})$. Writing

$$(2.24) \qquad \Phi(J; \underline{\zeta}) := \sum_{\Delta \in \mathcal{P}_f(\mathcal{CP}); [base(\Delta)] = J} \mathbf{F}_{\hat{\Psi}}^T(\Delta; \underline{\zeta}, t_1, \dots, t_{q+1})$$

for an interval $J \subset I$, we obtain

$$\ln \hat{\Xi}(I; \underline{\zeta}, t_1, \dots, t_{q+1}) = \sum_{J \subset I} \Phi(J; \underline{\zeta}).$$

From Möbius' inversion formula, we also have

$$(2.25) \qquad \Phi(J;\underline{\zeta}) = \sum_{\tilde{I}\subset J}(-1)^{|J|-|\tilde{I}|}\ln\hat{\Xi}(\tilde{I};\underline{\zeta},t_1,\dots,t_{q+1}).$$

Let us also define

$$\Phi_0(J;\zeta) := \sum_{\Delta\in\mathcal{P}_f(\mathcal{CP});[base(\Delta)]=J}\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta),$$

where $\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta)$ is given in (2.16) through cluster expansion. Then by (2.17) and the definition of $d^*(\Delta)$, $\Phi(J;\underline{\zeta})$ and $\Phi_0(J;\zeta)$ satisfy the following estimate.

$$(2.26) \qquad \max\{|\Phi(J;\underline{\zeta})|,|\Phi_0(J;\zeta)|\} \le 3e^{-(\mu-2\mu_4+\mu_2+1)\lceil\frac{|J|}{3}\rceil}$$

if $\mu > 2\mu_4-\mu_2-1$, $|Re\zeta| \le 1-\frac{\delta}{\mu}$ and $\underline{\zeta}$ satisfies (2.21).

Lemma 2.3. Let $\mu > 2\mu_4-\mu_2-1$. If $\underline{\zeta}$ satisfies (2.21), then

$$(2.27) \qquad \lim_{L\to\infty}\frac{1}{L}\sum_{J=[\hat{l},\hat{r}]\subset[0,L]}|\Phi(J;\underline{\zeta})-\Phi_0(J;\zeta_L(\hat{r}))| = 0,$$

where $\zeta_L(\hat{r}) = \zeta_L(\hat{r};\underline{\zeta}) := \zeta_0(1-\frac{\hat{r}}{L})+\sum_{i=1}^{q+1}\zeta_i 1_{[0,Lt_i]}(\hat{r})$.

Lemma 2.3 implies that

$$\lim_{L\to\infty}\frac{1}{L}\sum_{J=[\hat{l},\hat{r}]\subset[0,L]}\Phi(J;\underline{\zeta}) = \lim_{L\to\infty}\frac{1}{L}\sum_{J=[\hat{l},\hat{r}]\subset[0,L]}\Phi_0(J;\zeta_L(\hat{r})).$$

Note that for ζ satisfying $|Re\zeta| < 1-\frac{\delta}{\mu}$,

$$\hat{\varphi}(\zeta) = \sum_{\substack{J=[-k,0]\\ for\ some\ k\ge 0}}\Phi_0(J;\zeta),$$

which implies that

(2.28)

$$\lim_{L\to\infty}\frac{1}{L}\sum_{J=[\hat{l},\hat{r}]\subset[0,L]}\Phi_0(J;\zeta_L(\hat{r})) = \int_0^1\hat{\varphi}(\zeta_0(1-x)+\sum_{i=1}^{q+1}\zeta_i 1_{[0,t_i]}(x))dx$$

uniformly in $\underline{\zeta}$ satisfying (2.21). As a result of Proposition 2.2 and Lemma 2.3, we obtain

Corollary 2.4 For $\mu > \mu_4$,

(2.29)
$$\varphi^{(q)}(\underline{\zeta}; t_1, \dots, t_{q+1}) = \int_0^1 (\hat{\varphi} + \ln Q)\big(\zeta_0(1-x) + \sum_{i=1}^{q+1} \zeta_i 1_{[0,t_i]}(x)\big)dx$$

if $\underline{\zeta}$ satisfies (2.21). This function is analytic in $\underline{\zeta}$ in this region.

Proof of Lemma 2.3. We first introduce an intermediate weight $\tilde{\Psi}(\xi;\underline{\zeta})$ by

$$\begin{aligned}&\tilde{\Psi}(\xi;\underline{\zeta})\\ &:= \Psi(\xi)\exp\left[\mu\{\zeta_0(\frac{area(\xi)}{L} + (1-\frac{\hat{r}(\xi)}{L})k(\gamma)) + \sum_{i=1}^{q+1}\zeta_i 1_{[\hat{r}(\xi)<Lt_i]}k(\gamma)\}\right]\\ &\quad\times \prod_{\ell=a(\xi)}^{b(\xi)} Q^{-1}(\zeta_L(\ell)).\end{aligned}$$

It is easy to verify that $\tilde{\Psi}(\xi;\underline{\zeta})$ also satisfies (2.5) if $\underline{\zeta}$ satisfies (2.21), and therefore we have corresponding $\tilde{\Phi}$ by

$$\ln \sum_{\substack{C_1,\dots,C_m;\, compatible\\ base(C_p)\subset I,\, 1\le p\le m}} \prod_{p=1}^{m} \mathbf{F}_{\tilde{\Psi}}(C_p;\underline{\zeta}) = \sum_{J\subset I; interval} \tilde{\Phi}(J;\underline{\zeta})$$

for every interval $I \subset [0, L]$. $\tilde{\Phi}$ also satisfies the estimate (2.26). By the Möbius inversion formula $\Phi(I;\underline{\zeta}) = \tilde{\Phi}(I;\underline{\zeta})$ if I contains none of $\{Lt_i\}_{j=1}^{q+1}$. This means by (2.26) that

(2.30)
$$\lim_{L\to\infty}\frac{1}{L}\sum_{J\subset[0,L]} |\Phi(J;\underline{\zeta}) - \tilde{\Phi}(J;\underline{\zeta})| = 0.$$

For $s \in [0, 1]$, let us define

$$\tilde{\Psi}_s(\xi;\underline{\zeta}) := s\tilde{\Psi}(\xi;\underline{\zeta}) + (1-s)\hat{\Psi}(\xi;\zeta_L(\hat{r})),$$

and let $\tilde{\Phi}_s$ be the corresponding function defined through cluster expansion. Then we have

(2.31)
$$\begin{aligned}&|\tilde{\Phi}(J;\underline{\zeta}) - \Phi_0(J;\zeta_L(\hat{r}))|\\ &\le \sum_{\xi\in\mathcal{K}(J)} \sup_{s,\underline{\zeta}} \left|\frac{\partial\tilde{\Phi}_s(J;\underline{\zeta})}{\partial\tilde{\Psi}_s(\xi;\underline{\zeta})}\right| |\tilde{\Psi}(\xi;\underline{\zeta}) - \hat{\Psi}(\xi;\zeta_L(b))|.\end{aligned}$$

Like (2.25) we have

$$\tilde{\Phi}_s(J;\underline{\zeta}) = \sum_{I' \subset J; interval} (-1)^{|J|-|I'|} \ln \tilde{\Xi}(I'),$$

where

$$\tilde{\Xi}(I') := \sum_{\substack{\mathcal{C}_1,\cdots,\mathcal{C}_m \in \mathcal{CP} \\ base(\mathcal{C}_p) \subset I',\, 1 \le p \le m}} \prod_{p=1}^{m} \mathbf{F}_{\tilde{\Psi}_s}(\mathcal{C}_p;\underline{\zeta}).$$

For a polymer chain $\mathcal{C}$ with $base(\mathcal{C}) \subset I'$, we have

$$\left|\frac{\partial \ln \tilde{\Xi}(I')}{\partial \mathbf{F}_{\tilde{\Psi}_s}(\mathcal{C})}\right| \le \exp\Bigg\{ \sum_{\substack{\Delta\ i\ \mathcal{C}; \\ base(\Delta) \subset I'}} \left|\mathbf{F}^{T}_{\tilde{\Psi}_s}(\Delta;\underline{\zeta})\right| \Bigg\}$$

$$\le \exp\big\{c^*(\mathcal{C}) e^{-(\mu - 2\mu_4 + \mu_2 + 1)}\big\}.$$

Therefore for a polymer ξ with $base(\xi) \subset I'$, we have

$$\Big|\frac{\partial \ln \tilde{\Xi}(I')}{\partial \tilde{\Psi}_s(\xi)}\Big|$$

$$\le \sum_{\substack{\mathcal{C} \in \mathcal{CP}, \mathcal{C} \ni \xi, \\ base(\mathcal{C}) \subset I'}} \Big|\frac{\partial \mathbf{F}_{\tilde{\Psi}_s}(\mathcal{C};\underline{\zeta})}{\partial \tilde{\Psi}_s(\xi;\underline{\zeta})}\Big| \exp\big\{c^*(\mathcal{C}) e^{-(\mu - 2\mu_4 + \mu_2 + 1)}\big\}$$

$$\le \sum_{n,m \ge 0} \sum_{\substack{\{I_1,\cdots,I_n\}; \\ I_1,\cdots,I_n, base(\xi)\ form \\ linked\, intervals}} \sum_{\substack{\{I_{n+1},\cdots,I_{n+m}\}; \\ base(\xi), I_{n+1},\cdots,I_{n+m}\ form \\ linked\, intervals}}$$

$$\times \exp\big\{c(\xi) e^{-(\mu - 2\mu_4 + \mu_2 + 1)}\big\} e^{(n+m+2)\mu_3^*}$$

$$\times \prod_{p=1}^{n+m} \Big(\sum_{base(\xi_p) = I_p} |\tilde{\Psi}_s(\xi_p;\underline{\zeta})| e^{d(\xi_p) + c(\xi_p)} \Big)$$

$$\le \sum_{n,m \ge 0} \left\{ \frac{2R(\mu_2,\delta) e^{2\mu_3^*}}{(1 - R(\mu_2,\delta))^2} \right\}^{n+m} \exp\big\{c(\xi) e^{-(\mu - 2\mu_4 + \mu_2 + 1)}\big\}$$

$$\le 4 \exp\big\{c(\xi) e^{-(\mu - 2\mu_4 + \mu_2 + 1)}\big\},$$

if μ_2 is sufficiently large. This implies the uniform bound

$$\left|\frac{\partial \tilde{\Phi}_s(J;\underline{\zeta})}{\partial \tilde{\Psi}_s(\xi;\underline{\zeta})}\right| \le 4|J|^2 \exp\big\{3|base(\xi)| e^{-(\mu - 2\mu_4 + \mu_2 + 1)}\big\} \tag{2.32}$$

for $s \in [0,1]$, $\xi \in \mathcal{K}(J)$ and $\underline{\zeta}$ satisfying (2.21). Let $J = [\hat{l}, \hat{r}]$ be an interval in $[0, L]$ with $|J| \le (\ln L)^2$ and $Lt_i \notin J$ for any $i = 1, \ldots, q+1$, and let $\xi \in \mathcal{K}(J)$ be such that $N_v(\xi) \leq (\ln L)^2$. Let $K > 0$ be an arbitrary positive number and we fix it. We assume that $\underline{\zeta}$ satisfies (2.21) with $|Im\zeta_0| \leq K$. By analyticity, for $\hat{l} \leq \ell \leq \hat{r}$ we have

$$\log Q(\zeta_L(\hat{r})) - \log Q(\zeta_L(\ell)) \leq Const. \frac{(\ln L)^2}{L}.$$

uniformly in $\underline{\zeta}$ in this region. From this and the fact that

$$\begin{aligned} \mu\frac{area(\xi)}{L} + \mu(\frac{\hat{r}}{L} - \frac{\hat{r}(\xi)}{L})k(\gamma) &\leq \frac{\mu}{L}\sum_f |h(f)|(\hat{r} - pos(\xi)) \\ &\leq \mu\frac{(\ln L)^2}{L}N_v(\xi) \leq \mu(\ln L)^4/L, \end{aligned}$$

using the inequality $|e^z - 1| \leq |z|e^{|z|}$ we have

$$\begin{aligned} (2.33) \quad & \frac{|\tilde{\Psi}(\xi;\underline{\zeta}) - \hat{\Psi}(\xi;\zeta_L(\hat{r}))|}{|\hat{\Psi}(\xi;\underline{\zeta})|} \\ &= \left| \frac{Q(\zeta_L(\hat{r}))^{|base(\xi)|}}{\prod_{\ell=\hat{l}(\xi)}^{\hat{r}(\xi)} Q(\zeta_L(\ell))} \exp\left[\mu\zeta_0\frac{area(\xi)}{L} + \mu\zeta_0(\frac{\hat{r}}{L} - \frac{\hat{r}(\xi)}{L})k(\gamma)\right] - 1 \right| \\ &\leq Const. \frac{(\ln L)^4}{L}. \end{aligned}$$

The constant does not depend on L or $\underline{\zeta}$ satisfying $|Im\zeta_0| \leq K$ and (2.23). Hence we have

$$\begin{aligned} & |\tilde{\Phi}(J;\underline{\zeta}) - \Phi_0(J;\zeta_L(\hat{r}))| \\ & \leq Const. \sum_{\substack{\xi;\, base(\xi) \subset J \\ N_v(\xi) \leq (\ln L)^2}} |J|^2 e^{3|base(\xi)|e^{-(\mu - 2\mu_4 + \mu_2 + 1)}} |\hat{\Psi}(\xi;\underline{\zeta})| \frac{(\ln L)^4}{L} \\ & \quad + \sum_{\substack{\xi; base(\xi) \subset J \\ N_v(\xi) \geq (\ln L)^2}} |J|^2 e^{3|base(\xi)|e^{-(\mu - 2\mu_4 + \mu_2 + 1)}} \left(|\tilde{\Psi}(\xi;\underline{\zeta})| + |\hat{\Psi}(\xi;\zeta_L(\hat{r}))|\right) \\ & := I + II. \end{aligned}$$

Since $|J| \leq (\ln L)^2$ and $\underline{\zeta}$ satisfies (2.21), we can bound I and II in the following way.

$$\begin{aligned}
I &\leq Const.|J|^3 \Big\{ \sum_{\substack{\xi;\, base(\xi)=[0,k] \\ for\ some\ k\geq 0}} |\hat{\Psi}(\xi;\underline{\zeta})| e^{c(\xi)+d(\xi)} \Big\} \frac{(\ln L)^4}{L} \\
&= O\Big(\frac{(\ln L)^{10}}{L}\Big), \\
II &\leq |J|^2 e^{-\frac{\delta}{6}(\ln L)^2} \sum_{\xi;\, base(\xi)\subset J} \big[|\tilde{\Psi}(\xi;\underline{\zeta})| + |\hat{\Psi}(\xi;\zeta_L(\hat{r}))|\big] e^{c(\xi)+d(\xi)} \\
&\leq 6(\ln L)^6 e^{-\frac{\delta}{6}(\ln L)^2}.
\end{aligned}$$

Using this and (2.26), we have

$$\begin{aligned}
&\frac{1}{L} \sum_{J=[\hat{l},\hat{r}]\subset[0,L]} \big|\tilde{\Phi}(J;\underline{\zeta}) - \Phi_0(J;\zeta_L(\hat{r}))\big| \\
\leq& \frac{6}{L} \sum_{\substack{J\subset[0,L]; \\ |J|>(\ln L)^2}} e^{-(\mu-2\mu_4+\mu_2+1)\lceil\frac{|J|}{3}\rceil} + \frac{6}{L} \sum_{\substack{J\subset[0,L]; \\ |J|\leq(\ln L)^2, \\ Lt_i\in J \text{ for some } i}} e^{-(\mu-2\mu_4+\mu_2+1)\lceil\frac{|J|}{3}\rceil} \\
&+ \frac{1}{L} \sum_{\substack{J=[\hat{l},\hat{r}]\subset[0,L]; \\ |J|\leq(\ln L)^2, \\ Lt_i\notin J \text{ for any } i=1,\dots,q+1}} \big|\tilde{\Phi}(J;\underline{\zeta}) - \Phi_0(J;\zeta_L(\hat{r}))\big| \\
=& O\Big(\frac{(\ln L)^{10}}{L}\Big)
\end{aligned}$$

uniformly in $\underline{\zeta}$ satisfying (2.21) with $Im\zeta_0 \leq K$. Since we can take $K > 0$ in an arbitrary way, we proved (2.27).

the limiting quadratic form

Let $\underline{\zeta}$ satisfy (2.21). We introduce a $(q+1)\times(q+1)$ matrix $V_L(\underline{\zeta})$ by

$$V_L(\underline{\zeta}) = \frac{1}{\mu^2 L} \text{Hess } \ln \sum_{\Gamma\in\mathcal{S}_L} e^{\mu\underline{\zeta}\cdot\hat{X}_L^{(q)}(t_1,\dots,t_{q+1})} W(\Gamma).$$

This is analytic in $\underline{\zeta}$ satisfying (2.21).

Lemma 2.5. Assume that $\mu > 2\mu_4 - \mu_2 - 1$ and that $\underline{\zeta} \in \mathbf{R}^{q+2}$ and $\underline{\zeta}$ satisfies (2.21). Then uniformly in $\underline{\zeta}$ and $\underline{\eta} = (\eta_0, \dots, \eta_{q+1}) \in \mathbf{R}^{q+2}$

such that $|\underline{\eta}| = 1$,

$$\underline{\eta} \cdot V_L(\underline{\zeta})\underline{\eta} \longrightarrow \underline{\eta} \cdot V(\underline{\zeta})\underline{\eta}$$

as $L \to \infty$, where

$$V(\underline{\zeta}) = \frac{1}{\mu^2}\text{Hess} \int_0^1 (\ln Q + \hat{\varphi})(\zeta(x))dx, \tag{2.34}$$

and

$$\zeta(x) = \zeta_0(1 - x) + \sum_{i=1}^{q+1} \zeta_i 1_{[0,t_i]}(x). \tag{2.35}$$

Further, there exists $\mu_5 > 2\mu_4 - \mu_2 - 1$ such that $V(\underline{\zeta})$ is uniformly positive definite for $\mu > \mu_5$.

Proof. Let $\mu_5 > \mu_4 + 1$ be fixed and let $\mu > \mu_5$. It is easy to see that $\ln Q(\zeta(x))$ is analytic in $\underline{\zeta}$ for every $x \in [0, 1]$, and

$$\underline{\eta} \cdot V(\underline{\zeta})\underline{\eta} = \frac{1}{\mu^2} \int_0^1 (\eta_0(1 - x) + \sum_{i=1}^{q+1} \eta_i 1_{[0,t_i]}(x))^2 (\ln Q + \hat{\varphi})''(\zeta(x))dx.$$

The uniform convergence of

$$\frac{1}{L} \ln \sum_{\Gamma \in \mathcal{S}_L} e^{\mu \underline{\zeta} \cdot X_L^{(q)}(t_1, \dots, t_{q+1})} W(\Gamma)$$

to

$$\int_0^1 (\ln Q + \hat{\varphi})(\zeta(x))dx$$

assures the convergence $V_L(\underline{\zeta}) \to V(\underline{\zeta})$ by Cauchy's formula. What remains to prove is the non-degeneracy of $V(\underline{\zeta})$. First, note that for any $\zeta \in \mathbf{R}$ with $|\zeta| < 1$,

$$\frac{1}{\mu^2}(\ln Q)''(\zeta) = \frac{\cosh \mu \cosh \mu\zeta - 1}{(\cosh \mu - \cosh \mu\zeta)^2} \geq e^{-\mu} \frac{\cosh \mu_5 - 1}{\cosh \mu_5} \tag{2.36}$$

holds if $\mu > \mu_5$.

We prove the lemma in two different cases depending on whether $|\zeta_0 + \zeta_{q+1}|$ and $|\zeta_{q+1}|$ are both small or not.

Case 1) $|\zeta_0 + \zeta_{q+1}| < 1/5$, $|\zeta_{q+1}| < 1/5$.

In this case, we have

$$\begin{aligned}|\zeta(x)| &\leq (1-x)|\zeta_0+\zeta_{q+1}| + x|\zeta_{q+1}| + \sum_{i=1}^{q}|\zeta_i| \\ &\leq \frac{1}{5}+\frac{\delta}{4\mu}\end{aligned}$$

for every $x \in [0,1]$. By Cauchy's formula, we have

$$\hat{\varphi}''(\zeta(x)) = \frac{1}{\pi i}\int_{|z-\zeta(x)|=\frac{1}{5}} \frac{\hat{\varphi}(z)}{(z-\zeta(x))^3}dz$$

If $|z-\zeta(x)| = \frac{1}{5}$, then $|Rez| < \frac{3}{5} < 1-\frac{\delta}{\mu}$. Therefore by (2.26) and (2.28) we have

$$|\hat{\varphi}(z)| \leq 9\sum_{n=1}^{\infty} e^{-(\mu-2\mu_4+\mu_2+1)n}.$$

Therefore as $\mu \to \infty$

$$\left|\frac{1}{\mu^2}\hat{\varphi}''(\zeta(x))\right| \leq \frac{18\cdot 5^2}{\mu^2}e^{-\mu}(1+o(1)) \tag{2.37}$$

uniformly in $x \in [0,1]$. Taking μ_5 sufficiently large, we have

$$\frac{1}{\mu^2}(\ln Q+\hat{\varphi})''(\zeta(x)) \geq \frac{e^{-\mu}}{2} > 0$$

for $\mu > \mu_5$.

Case 2) $|\zeta_{q+1}| > \frac{1}{5}$ or $|\zeta_0+\zeta_{q+1}| > \frac{1}{5}$.

We assume that $|\zeta_0+\zeta_{q+1}| > \frac{1}{5}$. The argument for the case where $|\zeta_{q+1}| > \frac{1}{5}$ is the same. For $x \in [0, \frac{1}{16}]$ we have

$$\begin{aligned}|\zeta(x)| &\geq (1-x)|\zeta_0+\zeta_{q+1}| - x|\zeta_{q+1}| - \sum_{i=1}^{q}|\zeta_i| \\ &\geq \frac{1}{8}-\frac{\delta}{4\mu} > \frac{1}{10}\end{aligned}$$

for $\mu > \mu_5$, if μ_5 is sufficiently large. This means that

$$\frac{1}{\mu^2}(\ln Q)''(\zeta(x)) \geq \frac{e^{-\frac{9}{10}\mu}}{4}\frac{\cosh\mu_5-1}{\cosh\mu_5}$$

for $x \in [0, \frac{1}{16}]$ and $\mu > \mu_5$. Therefore by (2.36)

$$\begin{aligned}(2.38)\qquad &\int_0^{\frac{1}{16}} \frac{1}{\mu^2}(\ln Q)''(\zeta(x))\big(\eta_0(1-x)+\sum_{i=1}^{q+1}\eta_i 1_{[0,t_i]}(x)\big)^2 dx\\ &\geq \frac{e^{-\frac{9}{10}\mu}}{16\cdot 4}\frac{\cosh\mu_5 - 1}{\cosh\mu_5}\int_0^1 \big(\eta_0(1-x)+\sum_{i=1}^{q+1}\eta_i 1_{[0,t_i]}(x)\big)^2 dx.\end{aligned}$$

Since $\underline{\zeta} \in \mathbf{R}^{q+2}$ satisfies (2.21), $|\zeta(x)| < 1 - \frac{7\delta}{4\mu}$ for every $x \in [0,1]$. By Cauchy's formula,

$$\hat{\varphi}''(\zeta(x)) = \frac{1}{\pi i}\int_{|z-\zeta(x)|=\frac{\delta}{2\mu}} \frac{\hat{\varphi}(z)}{(z-\zeta(x))^3}d\zeta.$$

Since the circle $\{|z-\zeta(x)| = \frac{\delta}{2\mu}\}$ lies entirely in the region $\{Rez < 1-\frac{\delta}{\mu}\}$, by (2.26) and (2.28) we have

$$(2.39)\qquad \Big|\int_0^1 \frac{1}{\mu^2}\hat{\varphi}''(\zeta(x))dx\Big| \leq \frac{12}{\delta^2}e^{-\mu}(1+o(1)).$$

Thus, by (2.38) and (2.39) $V(\underline{\zeta})$ is uniformly positive definite.

Let $\hat{P}_L^{(q)}$ be the distribution of $\hat{X}_L^{(q)}(t_1,\dots,t_{q+1})$ under P_L, and $\hat{P}_{L,\underline{\zeta}}^{(q)}$ be given by

$$\hat{P}_{L,\underline{\zeta}}^{(q)}(\underline{\eta}) = E_L\left[e^{\mu\underline{\zeta}\cdot\hat{X}_L^{(q)}(t_1,\dots,t_{q+1})}\right]^{-1} e^{\mu\underline{\zeta}\cdot\underline{\eta}}\hat{P}_L^{(q)}(\underline{\eta})$$

for $\mu > \mu_5$, $\underline{\zeta} \in \mathbf{R}^{q+2}$ satisfying (2.21).

Lemma 2.6. Let $\delta > 0$ be small and $\mu > \mu_5$. Assume that $\underline{\zeta}_L, \underline{\zeta} \in \mathbf{R}^{q+2}$ satisfy (2.21) and $\underline{\zeta}_L \to \underline{\zeta}$ as $L \to \infty$. Then, under $\hat{P}_{L,\underline{\zeta}_L}^{(q)}$ the centralized random vector

$$\hat{Y}_L^{(q)}(t_1,\dots,t_{q+1}) = \frac{1}{\sqrt{L}}\big(\hat{X}_L^{(q)}(t_1,\dots,t_{q+1}) - \hat{E}_{L,\underline{\zeta}_L}^{(q)}\hat{X}_L^{(q)}(t_1,\dots,t_{q+1})\big)$$

converges weakly to a centered Gaussian random vector $\hat{Y}^{(q)}(t_1,\dots,t_{q+1})$ of which covariance matrix is given by $V(\underline{\zeta})$.

Proof. Let

$$g_L(\underline{\eta}) = \hat{E}_{L,\underline{\zeta}_L}^{(q)}\left[e^{i\underline{\eta}\cdot\hat{Y}_L^{(q)}(t_1,\dots,t_{q+1})}\right].$$

Then

$$\ln g_L(\underline{\eta}) = L\varphi_L(\underline{\zeta}_L + \frac{i}{\sqrt{L}\mu}\underline{\eta}) - L\varphi_L(\underline{\zeta}) - \frac{i\underline{\eta}}{\sqrt{L}}\cdot\hat{E}_{L,\underline{\zeta}_L}^{(q)}\left[\hat{X}_L^{(q)}(t_1,\dots,t_{q+1})\right],$$

where $\varphi_L(\underline{\zeta})$ is given by

$$\varphi_L(\underline{\zeta}) = \frac{1}{L}\ln\sum_{\Gamma\in\mathcal{S}_L} e^{\mu\underline{\zeta}\cdot\hat{X}_L^{(q)}(t_1,\dots,t_{q+1})}W(\Gamma).$$

Since $\underline{\zeta}_L$ satisfies (2.21), so does $\underline{\zeta}_L + \frac{i}{\mu\sqrt{L}}\underline{\eta}$, and we have

$$\begin{aligned}&\varphi_L(\underline{\zeta}_L + \frac{i}{\mu\sqrt{L}}\underline{\eta}) - \varphi_L(\underline{\zeta}_L)\\ =&\frac{i}{\mu L\sqrt{L}}E^{(q)}_{L,\underline{\zeta}_L}\left[\hat{X}_L^{(q)}(t_1,\dots,t_{q+1})\right] - \frac{1}{2\mu^2L^2}\sum_{j,k=1}^{q+1}\eta_j\eta_k\frac{\partial^2\varphi_L}{\partial\zeta_j\partial\zeta_k}\bigg|_{\underline{\zeta}=\underline{\zeta}_L} + R_L.\end{aligned}$$

Since

$$\frac{1}{\mu^2L}\sum_{j,k=1}^{q+1}\eta_j\eta_k\frac{\partial^2\varphi_L}{\partial\zeta_j\partial\zeta_k}\bigg|_{\underline{\zeta}=\underline{\zeta}_L} = \sum_{j,k=1}^{q+1}\eta_j\eta_k V_L(\underline{\zeta}_L)_{j,k},$$

this term converges to $-\frac{1}{2}\underline{\eta}\cdot V(\underline{\zeta})\underline{\eta}$. So it remains to show that $LR_L \to 0$ as $L\to\infty$. Formally, R_L has the following integral representation.

$$R_L = \left(\frac{i}{\mu\sqrt{L}}\right)^3\sum_{1\le j\le k\le m\le n} R_{j,k,m}, \tag{2.40}$$

where for $j < k < m$,

$$\begin{aligned}
R_{j,j,j} =& \frac{\eta_j^3}{2\pi i}\int_{C_j} \frac{\varphi_L\big(\underline{\zeta}_L + (\xi_j - \zeta_{L,j})\mathbf{e}_j + \sum_{\nu=j+1}^n(\frac{i}{\mu\sqrt{L}})\eta_\nu \mathbf{e}_\nu\big)}{(\xi_j - \zeta_L, j)^3(\xi_j - \zeta_{L,j} - (\frac{i}{\mu\sqrt{L}})\eta_j)} d\xi_j, \\
R_{j,j,k} =& \frac{\eta_j^2\eta_k}{(2\pi i)^2}\int_{C_j} \frac{d\xi_j}{(\xi_j - \zeta_{L,j})^3}\int_{C_k} d\xi_k \\
&\times \frac{\varphi_L\big(\underline{\zeta}_L + \sum_{\alpha=j,k}(\xi_\alpha - \zeta_{L,j})\mathbf{e}_\alpha + \sum_{\beta=k+1}^n(\frac{i}{\mu\sqrt{L}})\eta_\beta \mathbf{e}_\beta\big)}{(\xi_k - \zeta_{L,k})(\xi_k - \zeta_{L,k} - (\frac{i}{\mu\sqrt{L}})\eta_k)}, \\
R_{j,k,k} =& \frac{\eta_j\eta_k^2}{(2\pi i)^2}\int_{C_j} \frac{d\xi_j}{(\xi_j - \zeta_{L,j})^2}\int_{C_k} d\xi_k \\
&\times \frac{\varphi_L\big(\underline{\zeta}_L + \sum_{\alpha=j,k}(\xi_\alpha - \zeta_{L,\alpha})\mathbf{e}_\alpha + \sum_{\beta=k+1}(\frac{i}{\mu\sqrt{L}})\eta_\beta \mathbf{e}_\beta\big)}{(\xi_j - \zeta_{L,j})(\xi_k - \zeta_{L,k} - (\frac{i}{\mu\sqrt{L}})\eta_k)}, \\
R_{j,k,m} =& \frac{\eta_j\eta_k\eta_m}{(2\pi i)^3}\int_{C_j} \frac{d\xi_j}{(\xi_j - \xi_{L,j})^2}\int_{C_k} \frac{d\xi_k}{(\xi_k - \zeta_{L,k})^2}\int_{C_m} d\xi_m \\
&\times \frac{\varphi_L\big(\underline{\zeta}_L + \sum_{\alpha=j,k,m}(\xi_\alpha - \zeta_{L,\alpha})\mathbf{e}_\alpha + \sum_{\beta=m+1}^n(\frac{i}{\mu\sqrt{L}})\eta_\beta \mathbf{e}_\beta\big)}{(\xi_m - \zeta_{L,m})(\xi_m - \zeta_{L,m} - (\frac{i}{\mu\sqrt{L}})\eta_m)}.
\end{aligned}$$

Here, C_p is a curve composed of the lower half of the circle $\{|\xi_p - \zeta_{L,p}| = \rho\}$, upper half of the circle $\{|\xi_p - \zeta_{L,p} - (\frac{i}{\mu\sqrt{L}})\eta_p| = \rho\}$, and vertical line segments connecting them, and ρ is a small positive number. Let us estimate $|R_{j,j,j}|$. Other terms can be estimated similarly. Set

$$\underline{w}_L(j) := \underline{\zeta}_L + (\xi_j - \zeta_{L,j})\mathbf{e}_j + \sum_{\nu=j+1}^{n} \frac{i}{\mu\sqrt{L}}\eta_\nu \mathbf{e}_\nu.$$

Then it is easy to see that

$$\max\{|Re(w_L(j)_0 + w_L(j)_{q+1})|, |Re(w_L(j)_{q+1})|\} \leq 1 - \frac{2\delta}{\mu} + \rho(\delta_{0,j} + \delta_{q+1,j})$$

and $|Re(w_L(j)_\alpha)| \leq \dfrac{\delta}{4(q+1)\mu} + \rho\delta_{\alpha,j}$, where $\delta_{j,k} = 1$ if $j = k$ and $= 0$ if $j \neq k$. Note that

$$\varphi_L(w_L(j)) = \hat{\varphi}(w_L(j)) + \frac{1}{L}\sum_{\ell=0}^{L} \ln Q(\tilde{\underline{w}}_L(j;\ell)),$$

where

$$\underline{\tilde{w}}_L(j;\ell) := w_L(j)_0(1-\frac{\ell}{L}) + \sum_{p=1}^{q+1} w_L(j)_p 1_{[\ell \leq Lt_p]},$$

and that

$$\begin{aligned}|Re\underline{\tilde{w}}_L(j;\ell)| \leq& \max\{|Re(w_L(j)_0 + w_L(j)_{q+1})|, |Re(w_l(j)_{q+1})|\}\\ &+ \sum_{p=1}^{q} |Re(w_L(j)_p)| < 1 - \frac{7\delta}{4\mu} + \rho.\end{aligned}$$

If $\rho < \delta/4\mu$, then we have analyticity of the integrand in the expression of $R_{j,j,j}$ as in the proof of Proposition 2.2. This is true when $\underline{\zeta}_L$ satisfies (2.21) and $\xi_j \in C_j$. Thus, we can assume that $|\varphi_L(\underline{w}_L(j))|$. From this we easily obtain that

$$|R_{j,j,j}| \leq 2M\frac{|\eta_j|^3}{\rho^3}$$

for some $M > 0$, which is independent of L. This means that $LR_L = O(L^{-1/2})$ uniformly in $\underline{\eta}$.

Let $g_{\underline{\zeta}}$ be the density function of the Gaussian vector $\hat{Y}^{(q)}(t_1, \dots, t_{q+1})$ given in Lemma 2.6.

Proposition 2.7. Let $\mathcal{X}_L^{(q)} = (L^{-1}\mathbf{Z}) \times \mathbf{Z}^{q+1}$. For each $\underline{x}_L \in \mathcal{X}_L^{(q)}$ and $\underline{\zeta}_L \in \mathbf{R}^{q+2}$ satisfying (2.21), let

$$\underline{y}_L := \frac{1}{\sqrt{L}}(\underline{x}_L - \hat{E}^{(q)}_{L,\underline{\zeta}_L}\hat{X}^{(q)}_L(t_1, \dots, t_{q+1})).$$

Then we have

$$2L^{(q+4)/2}\hat{P}^{(q)}_L(\underline{x}_L) - g_{\underline{\zeta}_L}(\underline{y}_L) \to 0$$

uniformly in $\underline{x}_L \in \mathcal{X}_L$ and $\underline{\zeta}_L \in \mathbf{R}^{q+2}$ satisfying (2.21).

The proof is a complete repetition of the proof of Theorem 6.3 in [DH2], so we omit it. Let $h > 0$ and $a \geq \frac{h}{2}$ be such that

(2.41a) $$\frac{1}{\mu}\int_0^1 (1-x)\varphi'((1-x)\zeta_0^* + \zeta_1^*)dx = a$$

(2.41b) $$\frac{1}{\mu}\int_0^1 \varphi'((1-x)\zeta_0^* + \zeta_1^*)dx = h$$

hold for some $(\zeta_0^*, \zeta_1^*) \in \mathbf{R}^2$ with

$$\max\{|\zeta_0^* + \zeta_1^*|, |\zeta_1^*|\} \leq 1 - \frac{2\delta}{\mu}, \tag{2.42}$$

where $\varphi = \ln Q + \hat{\varphi}$. Let also $a_L > 0$ and $h_L > 0$ satisfy

$$\frac{1}{\mu}\frac{\partial \varphi_L}{\partial \zeta_0}(\zeta_{L,0}, 0, \dots, 0, \zeta_{L,1}) = \frac{a_L}{L^2} \tag{2.43a}$$

$$\frac{1}{\mu}\frac{\partial \varphi_L}{\partial \zeta_1}(\zeta_{L,0}, 0, \dots, 0, \zeta_{L,1}) = \frac{h_L}{L} \tag{2.43b}$$

for some $(\zeta_{L,0}, \zeta_{L,1})$ satisfying (2.42), and

$$(\frac{a_L}{L^2}, \frac{h_L}{L}) \to (a, h).$$

For simplicity, we write $\varphi_L(\zeta_0, \zeta_1)$ for $\varphi_L(\zeta_0, 0, \dots, 0, \zeta_1)$. By the argument in the proof of Lemma 2.5, for a sufficiently small $\rho > 0$, $\varphi_L(\zeta_{L,0}, \zeta_{L,1})$ and

$$\mathcal{L}(\zeta_0, \zeta_1) := \int_0^1 \varphi(\zeta_0(1-x) + \zeta_1)dx$$

are analytic in $(\zeta_0, \zeta_1) \in \mathcal{D}_\rho$, where

$$\mathcal{D}_\rho := \{(\zeta_0, \zeta_1) \in \mathbf{C}^2; \max\{|\zeta_0 - \zeta_0^*|, |\zeta_1 - \zeta_1^*|\} \leq \rho\}.$$

Also, $\varphi_L(\zeta_0, \zeta_1)$ converges to $\mathcal{L}(\zeta_0, \zeta_1)$ uniformly in $\mathcal{D}_\rho$. Therefore we also have the convergence;

$$\big(\nabla_{(\zeta_0,\zeta_1)}\varphi_L\big)(\zeta_0^*, \zeta_1^*) \to \big(\nabla_{(\zeta_0,\zeta_1)}\mathcal{L}\big)(\zeta_0^*, \zeta_1^*). \tag{2.44}$$

This convergence is uniform in (ζ_0^*, ζ_1^*) satisfying (2.42). By Lemma 2.5 for $q = 0$, there exist $L_0 \geq 1$ and $\varepsilon = \varepsilon(\rho, \mu, \delta, \zeta_0^*, \zeta_1^*) > 0$ such that

$$\sum_{j,k=0}^{1} [Hess_{(\zeta_0,\zeta_1)}\varphi_L(\zeta_0, \zeta_1)]_{j,k}\eta_j\eta_k \geq \varepsilon(|\eta_0|^2 + |\eta_1|^2),$$

$$\sum_{j,k=0}^{1} [Hess_{(\zeta_0,\zeta_1)}\mathcal{L}(\zeta_0, \zeta_1)]_{j,k}\eta_j\eta_k \geq \varepsilon(|\eta_0|^2 + |\eta_1|^2),$$

for $(\zeta_0, \zeta_1) \in \mathcal{D}_\rho \cap \mathbf{R}^2$, $L \geq L_0$ and $\eta_0, \eta_1 \in \mathbf{R}$. This implies that $\nabla_{(\zeta_0,\zeta_1)}\varphi_L$ and $\nabla_{(\zeta_0,\zeta_1)}\mathcal{L}$ are one-to-one bicontinuous maps on $\mathcal{D}_\rho \cap \mathbf{R}^2$ for every $L \geq L_0$. In particular, we have

(2.45a)

$$\| (\nabla_{(\zeta_0,\zeta_1)}\varphi_L)(\zeta_0, \zeta_1) - (\nabla_{(\zeta_0,\zeta_1)}\varphi_L)(\zeta_0^*, \zeta_1^*) \| \geq \frac{\varepsilon}{2} \| (\zeta_0, \zeta_1) - (\zeta_0^*, \zeta_1^*) \|$$

and

(2.45b)

$$\| (\nabla_{(\zeta_0,\zeta_1)}\mathcal{L})(\zeta_0, \zeta_1) - (\nabla_{(\zeta_0,\zeta_1)}\mathcal{L})(\zeta_0^*, \zeta_1^*) \| \geq \frac{\varepsilon}{2} \| (\zeta_0, \zeta_1) - (\zeta_0^*, \zeta_1^*) \|$$

for every $(\zeta_0, \zeta_1) \in \mathcal{D}_\rho \cap \mathbf{R}^2$. By (2.44) and the definition of (a, h) and (a_L, h_L), we have

$$\| \frac{1}{\mu}\nabla_{(\zeta_{L,0},\zeta_{L,1})}\varphi_L(\zeta_0^*, \zeta_1^*) - (\frac{a_L}{L^2}, \frac{h_L}{L}) \| \to 0.$$

This means that we can find $(\zeta_{L,0}, \zeta_{L,1}) \in \mathcal{D}_\rho$ which solves (2.43a, 2.43b) and by (2.45a, 2.45b) it converges to (ζ_0^*, ζ_1^*).

In order to discuss convergence of $X_L(t)$ from Proposition 2.7, except tightness we need one more estimate which assures that the separating contour itself neither fluctuates a lot nor is fat. To do this, let us define

$$vol(\xi) := |\gamma| + \sum_{\alpha+1}^{u} |C_\alpha| \tag{2.46}$$

for a polymer $\xi = (\gamma, \{C_\alpha\}_{\alpha=1}^u, \{\Lambda_\beta\}_{\beta=1}^v)$.

Lemma 2.8. Let $\mu > \mu_5$, $h > 0$, $a \geq \frac{h}{2}$ and a, h, a_L, h_L be given as above such that $(\frac{a_L}{L^2}, \frac{h_L}{L}) \to (a, h)$ as $L \to \infty$. Then for every $k \in \mathbf{N}$, there exists a constant $L_0 \geq 1$ such that for $L \geq L_0$,

(2.47)

$$P_L\left(\max\{vol(\xi); \xi \in \Delta(\Gamma)\} \geq \frac{6}{\delta}\ln L + k \;\middle|\; a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L\right)$$
$$\leq 1 - \exp\bigl(-4e^{-\frac{\delta}{6}k}\bigr).$$

Proof. Let (ζ_0^*, ζ_1^*) solves (2.41a, 2.41b) satisfying (2.42) and $(\zeta_{L,0}, \zeta_{L,1})$ be a solution of (2.43a, 2.43b) satisfying (2.42) such that $(\zeta_{L,0}, \zeta_{L,1})$

converges to (ζ_0^*, ζ_1^*) as $L \to \infty$. Put

$$\begin{aligned}\hat{X}_L^{(0)}(\Gamma) &:= (\frac{a(\pi(\Gamma))}{L}, k(\Gamma)) \\ &= \sum_{\xi \in \Delta(\Gamma)} \Big(\frac{area(\xi)}{L} + k(\gamma)(1 - \frac{\hat{r}(\xi)}{L}), k(\gamma) \Big).\end{aligned}$$

Then for $N := \frac{6}{\delta} \ln L + k$,

$$\begin{aligned}(2.48) \quad & P_L\big(\max\{vol(\xi); \xi \in \Delta(\Gamma)\} \le N \bigm| a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L\big) \\ =& \Big[\sum_{\Gamma \in \mathcal{S}_L; a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L} e^{\mu(\zeta_{L,0}, \zeta_{L,1}) \cdot \hat{X}_L^{(0)}(\Gamma)} W(\Gamma) \Big]^{-1} \\ & \times \sum_{\substack{\Gamma \in \mathcal{S}_L, a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L \\ vol(\xi) \le N \ for \ every \ \xi \in \Delta(\Gamma)}} e^{\mu(\zeta_{L,0}, \zeta_{L,1}) \cdot \hat{X}_L^{(0)}(\Gamma)} W(\Gamma).\end{aligned}$$

By Proposition 2.7 we have

$$\begin{aligned}(2.49) \quad & \sum_{\Gamma \in \mathcal{S}_L; a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L} e^{\mu(\zeta_{L,0}, \zeta_{L,1}) \cdot \hat{X}_L^{(0)}(\Gamma)} W(\Gamma) \\ =& e^{L \varphi_L(\zeta_{L,0}, \zeta_{L,1})} \hat{P}^{(0)}_{L, (\zeta_{L,0}, \zeta_{L,1})} (\frac{a_L}{L}, h_L) \\ =& e^{L \varphi_L(\zeta_{L,0}, \zeta_{L,1})} \frac{g_{(\zeta_{L,0}, \zeta_{L,1})}(0,0)}{2L^2} \{1 + o(1)\}\end{aligned}$$

as $L \to \infty$.

Let (ζ_0, ζ_1) satisfy (2.42) and

$$\varphi_L^{(N)}(\zeta_0, \zeta_1) := \frac{1}{L} \ln \sum_{\substack{\Gamma \in \mathcal{S}_L, a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L \\ vol(\xi) \le N \ for \ every \ \xi \in \Delta(\Gamma)}} e^{\mu(\zeta_0, \zeta_1) \cdot \hat{X}_L^{(0)}(\Gamma)} W(\Gamma).$$

It is straightforward to check that the estimate (2.4) is still valid when we replace $d(\xi)$ with

$$d_1(\xi) := d(\xi) - \frac{\delta}{6} |\gamma| + \frac{\delta}{6} vol(\xi).$$

The only change is that we introduce

$$G_1(\gamma) := \sum_{\xi; \ \gamma \ is \ the \ backbone \ of \ \xi} \big| \Psi(\xi) e^{\mu(\zeta_0, \zeta_1) \cdot \hat{X}_L^{(0)}} e^{\frac{\delta}{6} \sum_\alpha |C_\alpha|} \big|,$$

in place of $G(\gamma)$, and in estimating $G_1(\gamma)$, we have to put

$$g_2(\mu_2,\mu_0) = 4 \sum_{C \ni 0;\ connected} e^{-(\mu_2 - g_1(\mu_2,\mu_0) - \ln 2 - \delta/6)|C|}.$$

Therefore we have convergent cluster expansion;

$$\varphi_L^{(N)}(\zeta_0,\zeta_1) = \frac{1}{L} \sum_{\Delta \in \mathcal{P}_f(\mathcal{CP}_L(N))} \mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta_0,\zeta_1),$$

where $\mathcal{CP}_L(N) := \{\mathcal{C} \in \mathcal{CP}_L;\ vol(\mathcal{C}) \le N\}$ and

$$\sum_{\Delta\, i\, \mathcal{C}_0,\ \Delta \in \mathcal{P}_f(\mathcal{CP})} |\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta_0,\zeta_1)| e^{d_1^*(\Delta)} \le c^*(\mathcal{C}_0). \tag{2.50}$$

Therefore we have

(2.51)

$$|\varphi_L(\zeta_0,\zeta_1) - \varphi_L^{(N)}(\zeta_0,\zeta_1)| \le \frac{1}{L} \sum_{\Delta \in \mathcal{P}_f(\mathcal{CP}_L) \setminus \mathcal{P}_f(\mathcal{CP}_L(N))} |\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta_0,\zeta_1)|.$$

If $\Delta \in \mathcal{P}_f(\mathcal{CP}_L(N))$, then Δ contains at least one $\xi \in \mathcal{K}_L$ such that $vol(\xi) \ge N$. Therefore by (2.50) the RHS of (2.51) is bounded by

$$\frac{e^{-\frac{\delta}{6}N}}{L} \sum_{\Delta \in \mathcal{P}_f(\mathcal{CP}_L)} |\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta_0,\zeta_1)| e^{-d_1(\Delta)} \le 3e^{-\frac{\delta}{6}N}.$$

This estimate is uniform for (ζ_0,ζ_1) satisfying (2.42). By analyticity of φ_L and $\varphi_L^{(N)}$, we have for $\alpha,\beta \in \{0,1\}$,

$$|\frac{1}{\mu}\frac{\partial}{\partial\zeta_\alpha}[\varphi_L - \varphi_L^{(N)}](\zeta_{L,0},\zeta_{L,1})| \le \frac{3}{\rho}e^{-\frac{\delta}{6}N} \tag{2.52a}$$

and

$$\left|\frac{1}{\mu^2}\frac{\partial^2}{\partial\zeta_\alpha\partial\zeta_\beta}\left[\varphi_L - \varphi_L^{(N)}\right](\zeta_{L,0},\zeta_{L,1})\right| \le \frac{3}{\rho^2}e^{-\frac{\delta}{6}N} \tag{2.52b}$$

where $0 < \rho < \frac{\delta}{4\mu}$. Since $N \to \infty$ as $L \to \infty$,

$$Hess_{(\zeta_0,\zeta_1)}\varphi_L^{(N)}(\zeta_{L,0},\zeta_{L,1}) \to Hess_{(\zeta_0,\zeta_1)}\mathcal{L}(\zeta_0^*,\zeta_1^*)$$

as $L \to \infty$. Let $\hat{P}_{L,(\zeta_{L,0},\zeta_{L,1})}(\Gamma)$ be the probability weight which is proportional to $e^{\mu(\zeta_{L,0},\zeta_{L,1})\cdot\hat{X}^{(0)}(\Gamma)}W(\Gamma)$ restricted to the ensemble

$$\{\Gamma \in \mathcal{S}_L;\ vol(\xi) \le N \text{ for every } \xi \in \Delta(\Gamma)\}.$$

Then by (2.52a, 2.52b) as in the proof of Proposition 2.7, we see that

$$\frac{1}{\sqrt{L}}\Big((\frac{a(\pi(\Gamma))}{L}, k(\Gamma)) - \frac{1}{\mu}E^{(N)}_{L,(\zeta_{L,0},\zeta_{L,1})}(\frac{a(\pi(\Gamma))}{L}, k(\Gamma))\Big)$$

converges to a centered Gaussian vector with covariance matrix

$$\frac{1}{\mu^2}Hess_{(\zeta_0,\zeta_1)}\mathcal{L}(\zeta_0^*, \zeta_1^*)$$

as far as $N \to \infty$ as $L \to \infty$. Further, since $N - \frac{3}{\delta}\ln L \to \infty$,

$$\frac{1}{\mu}|\nabla_{(\zeta_0,\zeta_1)}\varphi_L^{(N)}(\zeta_{L,0}, \zeta_{L,1}) - (\frac{a_L}{L^2}, \frac{h_L}{L})| = o(\frac{1}{\sqrt{L}})$$

as $L \to \infty$ and by this we have

$$\hat{P}^{(N)}_{L,(\zeta_{L,0},\zeta_{L,1})}\Big((\frac{a(\pi(\Gamma))}{L}, k(\Gamma)) = (\frac{a_L}{L}, h_L)\Big) = \frac{g_{(\zeta_{L,0},\zeta_{L,1})}(0,0)}{2L^2}\{1+o(1)\}$$

as in the proof of Proposition 2.7. Combining this with (2.48) and (2.49), we see that there exists an $L_0 \geq 1$ such that for $L \geq L_0$ and $N = \frac{6}{\delta}\ln L + k$,

$$\begin{aligned}
&P_L(vol(\xi) \leq N \text{ for every } \xi \in \Delta(\Gamma) \mid a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L) \\
&\geq \exp\{-L|\varphi_L(\zeta_{L,0}, \zeta_{L,1}) - \varphi_L^{(N)}(\zeta_{L,0}, \zeta_{L,1})|\}\exp\{-e^{-\frac{\delta}{6}k}\} \\
&\geq \exp\{-4e^{-\frac{\delta}{6}k}\}.
\end{aligned}$$

Theorem 2.9. Let $\mu > \mu_5$, $h > 0, a \geq \frac{h}{2}$ and a_L, h_L be given as above. Further, we assume that $aL^2 - a_L = o(\sqrt{L^3})$ and $hL - h_L = o(\sqrt{L})$ as $L \to \infty$. Then the process $Y_L(t)$ under $P_L(\cdot \mid a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L)$ converges in finite dimensional distribution to the process

$$Y(t) = \frac{1}{\mu}\int_0^t \sqrt{\varphi''(\zeta_1 + (1-x)\zeta_0)}dB(x)$$

conditioned that

$$\int_0^1 Y(t)dt = 0, \quad Y(1) = 0.$$

Proof. Let $q \geq 1$, and $0 < t_1 < \cdots < t_{q+1} = 1$ be given arbitrarily. We take $(\zeta_0, \zeta_1) \in \mathbf{R}^2$ which satisfies (2.42) and solves (2.41a, 2.41b). Also, we take $(\zeta_{L,0}, \zeta_{L,1})$ as a solution of (2.43a, 2.43b) which satisfies

(2.42). Then by the above argument $(\zeta_{L,0}, \zeta_{L,1}) \to (\zeta_0, \zeta_1)$ as $L \to \infty$. Let $\underline{\zeta}^\circ_L, \underline{\zeta}^\circ \in \mathbf{R}^{q+2}$ be

$$\begin{aligned} \underline{\zeta}^\circ_L &= (\zeta_{L,0}, 0, \dots, 0, \zeta_{L,1}) \\ \underline{\zeta}^\circ &= (\zeta_0, 0, \dots, 0, \zeta_1). \end{aligned}$$

From the assumption of the theorem and (2.45a) and the uniform boundedness of $Hess_{\underline{\zeta}}\varphi_L$, we have

$$\begin{aligned} &\hat{E}^{(q)}_{L,\underline{\zeta}^\circ_L} \hat{X}^{(q)}_L(t_1, \dots, t_{q+1}) \\ =&(\frac{a_L}{L}, \hat{E}^{(q)}_{L,\underline{\zeta}^\circ_L} X_L(\frac{\lfloor Lt_1 \rfloor}{L}), \cdots, \hat{E}^{(q)}_{L,\underline{\zeta}^\circ_L} X_L(\frac{\lfloor Lt_q \rfloor}{L}), h_L) \\ =&\frac{L}{\mu}(\nabla_{\underline{\zeta}}\varphi_L)(\underline{\zeta}^\circ_L) \\ =&\frac{L}{\mu}(\nabla_{\underline{\zeta}}\varphi^{(q)})(\underline{\zeta}^\circ; t_1, \dots, t_{q+1}) + o(\sqrt{L}). \end{aligned}$$

By proposition 2.7 we have for $-\infty < \hat{l}_j < \hat{r}_j < \infty, 1 \leq j \leq q$,

$$\begin{aligned} &\lim_{L\to\infty} \hat{P}^{(q)}_L \big(y_j \in [\hat{l}_j, \hat{r}_j] \quad 1 \leq j \leq q \mid x_0 = \frac{a_L}{L}, x_{q+1} = h_L\big) \\ =& \lim_{L\to\infty} \hat{P}^{(q)}_{L,\underline{\zeta}^\circ_L} \big(y_j \in [\hat{l}_j, \hat{r}_j] \quad 1 \leq j \leq q \mid x_0 = \frac{a_L}{L}, x_{q+1} = h_L\big) \\ =&\frac{\int_{[\hat{l}_1,\hat{r}_1]\times\cdots\times[\hat{l}_q,\hat{r}_q]} g_{\underline{\zeta}^\circ}(0, y_1, \dots, y_q, 0) dy_1 \cdots dy_q}{\int_{\mathbf{R}^q} g_{\underline{\zeta}^\circ}(0, y_1, \dots, y_q, 0) dy_1 \cdots dy_q}. \end{aligned}$$

Let

$$\hat{Y}^{(q)}(t_1, \dots, t_{q+1}) = (Y_0, Y(t_1), Y(t_2), \dots, Y(t_{q+1}))$$

be a Gaussian random vector with distribution density $g_{\underline{\zeta}}(y_0, \dots, y_{q+1})$. Then its covariance matrix is given by

$$E\left[Y(t_j)Y(t_k)\right] = \frac{1}{\mu^2}\int_0^{t_j \wedge t_k} \varphi''(\zeta_0(1-x) + \zeta_1)dx$$

for $j, k = 1, \dots, q+1$, where $a \wedge b = \min\{a, b\}$, and

$$\begin{aligned} E\left[Y_0 Y(t_j)\right] =& \frac{1}{\mu^2}\int_0^{t_j} \varphi''(\zeta_0(1-x) + \zeta_1)dx \\ E\left[Y_0^2\right] =& \frac{1}{\mu^2}\int_0^1 \varphi''(\zeta_0(1-x) + \zeta_1)dx \end{aligned}$$

for $j = 1, 2, \dots, q+1$. This means that $\{Y_0, \{Y(t)\}_{t \in [0,1]}\}$ is a Gaussian system with covariance given above for every $0 < t_1 < \dots < t_{q+1} = 1$, $q \geq 1$. Finally, by Lemma 2.8 we can replace $\hat{E}^{(q)}_{L, \underline{\zeta}^{\circ}_{L}} X_L(t_j)$ with $\hat{E}^{(q)}_{L, \underline{\zeta}^{\circ}_{L}} X_L(\frac{\lfloor L t_j \rfloor}{L})$ for every $1 \leq j \leq q$ in the above argument.

§3. Tihgtness

As usual, we will estimate the fourth moment of $Y_L(t) - Y_L(s)$ for every $s, t \in [0, 1]$. First, we show the following one polymer estimate. For an integer $x \in [0, L]$ and $\Gamma \in \mathcal{S}_L$, let $\xi(x) = \xi(x, \Gamma)$ be the unique element of $\mathcal{D}(\Gamma)$ whose base contains x.

Lemma 3.1 Let $\mu > \mu_5, h > 0, a \geq \frac{h}{2}$ and a_L, h_L be given as in Lemma 2.8. Then there exist constants $C > 0$ and $L_1 \geq 1$ such that for $L \geq L_1$,

$$E_L \left[e^{\frac{1}{2} d(\xi(x))} \middle| \, a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L \right] \leq C.$$

Proof. Let (ζ_0^*, ζ_1^*) satisfy (2.41a, 2.41b) and (2.42), and $(\zeta_{L,0}, \zeta_{L,1})$ satisfy (2.42) and (2.43a, 2.43b) such that $(\zeta_{L,0}, \zeta_{L,1}) \to (\zeta_0^*, \zeta_1^*)$ as $L \to \infty$. For $\Gamma \in \mathcal{S}_L$ such that $\mathcal{D}(\Gamma) \ni \xi$, let $\Gamma'(\xi)$ denote the set of elements of $\mathcal{S}_L$ such that $\mathcal{D}(\Gamma'(\xi)) = \mathcal{D}(\Gamma) \setminus \{\xi\}$. Also we put for a polymer ξ,

$$\hat{X}_L^{(0)}(\xi) = (\frac{area(\gamma)}{L} + k(\gamma)(1 - \frac{\hat{r}(\xi)}{L}), k(\gamma)),$$

and $\Psi(\xi; \zeta_0, \zeta_1) := \Psi(\xi) \exp\{\mu \hat{X}_L^{(0)}(\xi) \cdot (\zeta_0, \zeta_1)\}$, where γ stands for the backbone of ξ. Then

$$\begin{aligned} &P_{L,(\zeta_{L,0},\zeta_{L,1})}\big[\{\mathcal{D}(\Gamma) \ni \xi\} \cap \{\hat{X}_L^{(0)}(\Gamma) = (\frac{a_L}{L}, h_L)\}\big] \\ =& e^{-L\varphi_L(\zeta_{L,0},\zeta_{L,1})} \Psi(\xi; \zeta_{L,0}, \zeta_{L,1}) \\ &\times \sum_{\substack{\Gamma \in \mathcal{S}_L;\ \mathcal{D}(\Gamma) \ni \xi, \\ a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L}} e^{\mu(\zeta_{L,0},\zeta_{L,1}) \cdot \hat{X}_L^{(0)}(\Gamma'(\xi))} W(\Gamma'(\xi)). \end{aligned}$$

By the cluster expansion we have

$$\begin{aligned} (3.1)\quad & P_{L,(\zeta_{L,0},\zeta_{L,1})}\big[\{\mathcal{D}(\Gamma) \ni \xi\} \cap \{\hat{X}_L^{(0)}(\Gamma) = (\frac{a_L}{L}, h_L)\}\big] \\ &= \sum_{\substack{\mathcal{C} \in \mathcal{CP}_L; \\ \mathcal{C} \ni \xi}} \mathbf{F}_{\hat{\Psi}}(\mathcal{C}; \zeta_{L,0}, \zeta_{L,1}) \exp\{- \sum_{\Delta \in \mathcal{P}_f(\mathcal{CP}_L); \Delta i\, \mathcal{C}} \mathbf{F}^T_{\hat{\Psi}}(\Delta; \zeta_{L,0}, \zeta_{L,1})\} \\ &\quad \times P_{L,(\zeta_{L,0},\zeta_{L,1})}\big[a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L \mid \mathcal{D}(\Gamma) \ni \xi\big], \end{aligned}$$

where

$$\hat{\Psi}(\xi;\zeta_{L,0},\zeta_{L,1}) := \Psi(\xi;\zeta_{L,0},\zeta_{L,1})\prod_{\ell=0}^{L} Q^{-1}(\zeta_{L,0}(1-\frac{\ell}{L})+\zeta_{L,1}).$$

Since the final term in the RHS of (3.1) is not larger than 1, by the same argument to derive (2.32) we have for $C>0$,

$$\begin{aligned}\sum_{\substack{\xi;base(\xi)\ni x,\\ |\gamma|\geq C\ln L}} & e^{\frac{1}{2}d(\xi)}P_{L,(\zeta_{L,0},\zeta_{L,1})}\left[\{\mathcal{D}(\Gamma)\ni\xi\}\cap\{\hat{X}_L^{(0)}(\Gamma)=(\tfrac{a_L}{L},h_L)\}\right]\\ \leq 4 & \sum_{\xi;base(\xi)\ni x,|\gamma|\geq C\ln L} e^{c(\xi)+\frac{1}{2}d(\xi)}\hat{\Psi}(\xi;\zeta_{L,0},\zeta_{L,1}).\end{aligned}$$

As in the proof of Lemma 2.1,

$$\begin{aligned}(3.2)\qquad & \sum_{\xi;\ base(\xi)\ni x,\ |\gamma|\geq C\ln L} e^{c(\xi)+\frac{1}{2}d(\xi)}|\hat{\Psi}(\xi;\zeta_{L,0},\zeta_{L,1})|\\ \leq & e^{-\frac{\delta}{12}C\ln L}\sum_{\xi;\ base(\xi)\ni x} e^{c(\xi)+d(\xi)}|\hat{\Psi}(\xi;\zeta_{L,0},\zeta_{L,1})|\\ \leq & 3e^{-\frac{\delta}{12}C\ln L}.\end{aligned}$$

By (2.49), we have for a constant $C_1>0$ and a sufficiently large L,

$$\begin{aligned}&E_{L,(\zeta_{L,0},\zeta_{L,1})}\left[e^{\frac{1}{2}d(\xi(x))}1_{\{|\gamma(x)|\geq C\ln L\}}\ \middle|\ \hat{X}_L^{(0)}(\Gamma)=(\tfrac{a_L}{L},h_L)\right]\\ \leq & C_1L^2e^{-\frac{\delta}{12}C\ln L},\end{aligned}$$

which goes to zero as $L\to\infty$. Here, $\gamma(x)$ stands for the backbone of $\xi(x)$.

Assume that $|\gamma|\leq C\ln L$ for the backbone γ of ξ. Then since

$$\begin{aligned}\varphi_L(\zeta_{L,0},\zeta_{L,1}\mid\xi) := & \frac{1}{L}\ln\sum_{\Gamma\in\mathcal{S}_L;\mathcal{D}(\Gamma)\ni\xi} e^{\mu(\zeta_{L,0},\zeta_{L,1})\cdot\hat{X}_L^{(0)}(\Gamma)}W(\Gamma)\\ = & \varphi_L(\zeta_{L,0},\zeta_{L,1})-\frac{1}{L}\sum_{\Delta\in\mathcal{K}_L;\ \Delta i\xi}\mathbf{F}^T_{\hat{\Psi}}(\Delta;\zeta_{L,0},\zeta_{L,1}),\end{aligned}$$

$[Hess_{(\zeta_0,\zeta_1)}\varphi_L(\cdot\mid\xi)](\zeta_{L,0},\zeta_{L,1})$ converges to $[Hess_{(\zeta_0,\zeta_1)}\mathcal{L}](\zeta_0^*,\zeta_1^*)$ uniformly in ξ with $|\gamma|\leq C\ln L$. Therefore there exist constants $C_2>0$

and $L_0 \geq 1$ such that for $L \geq L_0$,

$$(3.3) \qquad P_{L,(\zeta_{L,0},\zeta_{L,1})}\big(a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L \mid \mathcal{D}(\Gamma) \ni \xi\big) \leq \frac{C_2}{L^2}$$

uniformly in ξ such that $|\gamma| \leq C \ln L$. Combining (2.49) with (3.3), we can find L_1 such that for $L \geq L_1$,

$$(3.4) \qquad E_{L,(\zeta_{L,0},\zeta_{L,1})}\left[e^{\frac{1}{2}d(\xi(x))}1_{\{|\gamma|\leq C\ln L\}}\middle|\ \hat{X}_L^{(0)}(\Gamma) = \big(\tfrac{a_L}{L}, h_L\big)\right]$$
$$\leq C_1C_2 \sum_{base(\xi)\ni x, |\gamma|\leq C\ln L} |\hat{\Psi}(\xi)|e^{\frac{1}{2}d(\xi)+c(\xi)} \leq 3C_1C_2.$$

This together with (3.3) proves Lemma 3.1.

Now let us turn to the estimate of the fourth moment of $Y_L(t) - Y_L(s)$. It is sufficient to consider the case where Ls, Lt are integers and $s < t$.

Lemma 3.2 There exist constants $C_3 > 0$ and $L_2 \geq 1$ such that for $L \geq L_2$, if $|t-s| \leq L^{-\frac{4}{5}}$, then

$$(3.5) \quad E_L\big(|Y_L(t) - Y_L(s)|^4 | a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L\big) \leq C_3|t-s|^{\frac{3}{2}}.$$

Proof. Since

$$Y_L(t) - Y_L(s) = \frac{1}{\sqrt{L}}\big[X_L(t) - X_L(s) - \frac{L}{\mu}\int_s^t \varphi'(\zeta_0^*(1-x) + \zeta_1^*)dx\big],$$

we estimate

$$E_L(|X_L(t) - X_L(s)|^4 \mid a(\pi(\gamma)) = a_L, k(\gamma) = h_L)$$

and

$$E_L(|L\int_s^t \varphi'(\zeta_0^*(1-x) + \zeta_1^*)dx|^4 \mid a(\pi(\gamma)) = a_L, k(\gamma) = h_L)$$

separately, where (ζ_0^*, ζ_1^*) solves (2.41a), (2.41b) and satisfies (2.42). By analyticity the latter is bounded by $C(L|t-s|)^4$ for some positive constant C. Also, by Lemma 3.1, the former is bounded by

$$C'(L|t-s|)^4$$

for some positive constant C'. It remains to check that

$$L^2|t-s|^4 \leq |t-s|^{\frac{3}{2}},$$

which is true when $|t-s| \leq L^{-\frac{4}{5}}$.

To handle the case where $|t-s| \geq L^{-\frac{4}{5}}$, we introduce a moment generating function $\varphi_L^{(s,t)}$ by

$$\varphi_L^{(s,t)}(\zeta_0, \zeta_1, \zeta_2) := \frac{1}{L} \ln \sum_{\Gamma \in \mathcal{S}_L} e^{\mu \hat{X}_L^{(s,t)}(\Gamma) \cdot \underline{\zeta}} W(\Gamma),$$

where

$$\hat{X}_L^{(s,t)}(\Gamma) := \Big(\frac{a(\pi(\Gamma))}{L}, k(\Gamma), \frac{X_L(t) - X_L(s)}{\sqrt{t-s}}\Big)$$

and $\underline{\zeta} = (\zeta_0, \zeta_1, \zeta_2) \in \mathbf{R}^3$ such that (ζ_0, ζ_1) satisfies (2.42) and

$$|\zeta_2| \leq \frac{\delta}{2\mu} \sqrt{t-s}. \tag{3.6}$$

To complete the proof of the tightness of $\{Y_L(t),\ 0 \leq t \leq 1\}$, it is sufficient to show that there exists a constant ε_0 independent of L such that (3.5) holds for all $s, t \in [0,1]$ with $|t-s| \leq \varepsilon_0$.

Let $a, h, a_L, h_L, (\zeta_0^*, \zeta_1^*), (\zeta_{L,0}, \zeta_{L,1})$ be taken as before; i.e.,

1. (ζ_0^*, ζ_1^*) and $(\zeta_{L,0}, \zeta_{L,1})$ satisfy (2.42),
2. (ζ_0^*, ζ_1^*) solves (2.41a), (2.41b), and
3. $(\zeta_{L,0}, \zeta_{L,1})$ solves (2.43a), (2.43b).

Put

$$v_L := \frac{1}{\mu} \frac{\partial \varphi_L^{(s,t)}}{\partial \zeta_2} (\zeta_{L,0}, \zeta_{L,1}, 0). \tag{3.7}$$

Then as in the proof of Lemma 2.3, we can show that

$$v_L - \frac{1}{\mu\sqrt{t-s}} \int_s^t \varphi'(\zeta_0^*(1-x) + \zeta_1^*) dx = O\big(L^{-\frac{3}{5}} (\ln L)^{10}\big). \tag{3.8}$$

Therefore

$$\frac{Y_L(t) - Y_L(s)}{\sqrt{t-s}} = \frac{X_L(t) - X_L(s)}{\sqrt{L(t-s)}} - \sqrt{L} v_L + o(1).$$

So, we will show that for some $\varepsilon_0 > 0$ and for all $s, t \in [0,1]$ such that $|t-s| < \varepsilon_0$,

$$\sum_{k=0}^{\infty} (k+1)^4 P_L \left(\frac{X_L(t) - X_L(s)}{\sqrt{L(t-s)}} - v_L \sqrt{L} \geq k \;\middle|\; \begin{matrix} a(\pi(\gamma)) = a_L, \\ k(\Gamma) = h_L \end{matrix} \right)$$

converges and is bounded from above by a constant independent of L, s, t. For $k \in \mathbf{N}$, let $\underline{\zeta}_L^{(k)} = (\zeta_{L,0}^{(k)}, \zeta_{L,1}^{(k)}, \zeta_{L,2}^{(k)})$ be the solution of

$$\frac{1}{\mu}[\nabla_{(\zeta_0,\zeta_1,\zeta_2)}\varphi_L^{(s,t)}](\underline{\zeta}_L^{(k)}) = (\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + \frac{k}{\sqrt{L}})$$

and $\zeta_L^{(0)} = (\zeta_{L,0}, \zeta_{L,1}, 0)$. For $\underline{\eta} = (\eta_0, \eta_1, \eta_2)$, let $\varphi_L^{*(s,t)}(\underline{\eta})$ be the Legendre transform of $\frac{1}{\mu}\varphi_L^{(s,t)}$. Then by duality,

$$[\nabla_{\underline{\eta}}\varphi_L^{*(s,t)}](\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + \frac{k}{\sqrt{L}}) = (\zeta_{L,0}^{(k)}, \zeta_{L,1}^{(k)}, \zeta_{L,2}^{(k)})$$

and

$$\zeta_{L,2}^{(k)} = \int_0^{\frac{k}{\sqrt{L}}} \frac{\partial^2 \varphi_L^{*(s,t)}}{\partial \eta_2^2}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + u)du \geq 0.$$

Therefore

(3.9)

$$\begin{aligned}
&P_L\left(\frac{X_L(t) - X_L(s)}{\sqrt{L(t-s)}} \geq v_L\sqrt{L} + k \,\middle|\, \begin{matrix} a(\pi(\Gamma)) = a_L, \\ k(\Gamma) = h_L \end{matrix}\right) \\
&= \sum_{j \geq Lv_L\sqrt{t-s}+k\sqrt{L(t-s)}} \frac{e^{L\varphi_L^{(s,t)}(\underline{\zeta}_L^{(k)}) - \mu(\frac{a_L}{L}, h_L, j)\cdot\zeta_L^{(k)}}}{e^{L\varphi_L^{(s,t)}(\underline{\zeta}_L^{(0)}) - \mu(\frac{a_L}{L}, h_L, Lv_L)\cdot\zeta_L^{(0)}}} \\
&\qquad \times \frac{P_{L,\underline{\zeta}_L^{(k)}}(X_L(t) - X_L(s) = j, a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L)}{P_{L,\underline{\zeta}_L^{(0)}}(a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L)} \\
&\leq \exp\left\{-L\mu[\varphi_L^{*(s,t)}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + \frac{k}{\sqrt{L}}) - \varphi_L^{*(s,t)}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L)]\right\} \\
&\quad \times \frac{P_{L,\underline{\zeta}_L^{(k)}}\left(\frac{X_L(t)-X_L(s)}{\sqrt{L(t-s)}} \geq v_L\sqrt{L} + k, a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L\right)}{P_{L,\underline{\zeta}_L^{(0)}}(a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L)}.
\end{aligned}$$

From Proposition 2.7, the RHS of (3.9) is bounded by

$$\begin{aligned}
&\exp\{-L\mu[\varphi_L^{*(s,t)}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + \frac{k}{\sqrt{L}}) - \varphi_L^{*(s,t)}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L)]\} \\
&\times Const.L^2.
\end{aligned}$$

as $L \to \infty$.

Lemma 3.3. There exist positive constants α_1, α_2, L_0 such that every eigenvalue of

$$\frac{1}{\mu^2} Hess_{(\zeta_0,\zeta_1,\zeta_2)} \left[\varphi_L^{(s,t)}(\zeta_0, \zeta_1, \frac{\zeta_2}{\sqrt{t-s}})\right]$$

is in the interval $[\alpha_1, \alpha_2]$ if $L \geq L_0$ and

$$\text{(3.10)} \qquad \begin{cases} \qquad\qquad |\zeta_2| & \leq \frac{\delta}{3\mu}\sqrt{|t-s|} \\ \max\{|\zeta_0+\zeta_1|, |\zeta_1|\} & \leq 1 - \frac{3\delta}{2\mu} \end{cases}$$

For the moment we take it for granted that Lemma 3.3 is true. Then, since $(\zeta_{L,0}, \zeta_{L,1})$ satisfies (2.42), by Lemma 3.3 and the continuity, we can find $\varepsilon > 0$ such that if $\frac{k}{\sqrt{L}} < \varepsilon\sqrt{t-s}$, then $|\zeta_{L,0}^{(k)} - \zeta_{L,0}|$, $|\zeta_{L,1}^{(k)} - \zeta_{L,1}|$, $|\zeta_{L,2}^{(k)}|$ are all bounded by $\frac{\delta}{4\mu}$ and every eigenvalue of

$$\left[Hess_{\underline{\eta}}\varphi_L^{*(s,t)}\right] (\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + \frac{k}{\sqrt{L}})$$

is in the interval $[\alpha_2^{-1}, \alpha_1^{-1}]$. Thus, we have

$$\text{(3.11)} \qquad \begin{aligned} &\varphi_L^{*(s,t)}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + \frac{k}{\sqrt{L}}) - \varphi_L^{*(s,t)}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L) \\ =&\int_0^{\frac{k}{L}} (\frac{k}{\sqrt{L}} - u) \frac{\partial^2 \varphi_L^{*(s,t)}}{\partial \eta_2^2} (\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + u) du \geq \alpha_2^{-1} \frac{k^2}{2L} \end{aligned}$$

if $k \leq \varepsilon\sqrt{L(t-s)}$. By convexity, this means that the LHS of (3.9) is not less than

(3.12)
$$\frac{k}{\sqrt{L}} \frac{\varphi_L^{*(s,t)}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L + \varepsilon\sqrt{t-s}) - \varphi_L^{*(s,t)}(\frac{a_L}{L^2}, \frac{h_L}{L}, v_L)}{\varepsilon\sqrt{t-s}} \geq \frac{\alpha_2^{-1}}{2} \varepsilon L^{-\frac{9}{10}} k$$

(3.12) proves that

$$\begin{aligned} &\sum_{k \geq \varepsilon\sqrt{L(t-s)}} (k+1)^4 P_L \left(\frac{X_L(t) - X_L(s)}{\sqrt{L(t-s)}} \geq v_L\sqrt{L} + k \middle| \begin{array}{l} a(\pi(\Gamma)) = a_L, \\ k(\Gamma) = h_L \end{array} \right) \\ =&O(L^4 \exp\{-\mu \frac{\alpha_2^{-1}\varepsilon}{2} L^{\frac{1}{5}}\}) \end{aligned}$$

for large L. Also, for $k \le \varepsilon\sqrt{L(t-s)}$, $Hess_{(\zeta_0,\zeta_1,\zeta_2)}[\varphi_L^{(s,t)}(\zeta_0,\zeta_1,\frac{\zeta_2}{\sqrt{t-s}})]$ is uniformly positive definite and by Lemma 3.3,

$$
\begin{aligned}
&P_{L,\underline{\zeta}_L^{(k)}}\left(\frac{X_L(t)-X_L(s)}{\sqrt{L(t-s)}} \ge v_L\sqrt{L}+k, a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L\right) \\
\le &P_{L,\underline{\zeta}_L^{(k)}}\big(a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L\big) \\
\le &\frac{Const.}{L^2}.
\end{aligned}
$$

This and (3.9) together with (3.8) prove

$$
\begin{aligned}
&\sum_{k\le\varepsilon\sqrt{L(t-s)}} (k+1)^4 \\
&\times P_L\left(\frac{X_L(t)-X_L(s)}{\sqrt{L(t-s)}} \ge v_L\sqrt{L}+k \,\middle|\, a(\pi(\Gamma)) = a_L, k(\Gamma) = h_L\right) \\
\le &Const.\sum_{k=0}^{\infty}(k+1)^4 e^{-\frac{k^2}{2\alpha_2}} < \infty
\end{aligned}
$$

Proof of Lemma 3.3. Put

$$
\begin{aligned}
&\Psi_L^{(s,t)}(\xi;\zeta_0,\zeta_1,\frac{\zeta_2}{\sqrt{t-s}}) \\
=&\Psi(\xi)\exp\{\zeta_0(\frac{\hat{l}(\xi)}{L}+(1-\frac{\hat{r}(\xi)}{L}k(\gamma))+\zeta_1 k(\gamma)+\frac{\zeta_2}{\sqrt{t-s}}k(\gamma;Ls,Lt)\}
\end{aligned}
$$

where

(3.13)

$$
k(\gamma;Ls,Lt) = \begin{cases} k(\gamma) & \text{if } base(\xi) \subset [Ls,Lt], \\ k(\gamma)-k(\gamma;Ls) & \text{if } \hat{l}(\xi) < Ls \le \hat{r}(\xi) \le Lt \\ k(\gamma;Lt) & \text{if } Ls \le \hat{l}(\xi) \le Lt < \hat{r}(\xi) \\ k(\gamma;Lt)-k(\gamma;Ls) & \text{if } \hat{l}(\xi) < Ls < Lt < \hat{r}(\xi). \end{cases}
$$

Then as in the proof of Proposition 2.2, we have a convergent cluster expansion

$$\begin{aligned}
&\varphi_L(\zeta_0,\zeta_1,\frac{\zeta_2}{\sqrt{t-s}}) \\
&= \frac{1}{L}\sum_{J=[a,b]\subset[0,L]} \Phi^{(\Delta)}(J;\zeta_0,\zeta_1,\frac{\zeta_2}{\sqrt{t-s}}) \\
&= \frac{1}{L}\sum_{J=[\hat{l},\hat{r}]\subset[0,L]\setminus[s,t]} \Phi(J;\zeta_L(\hat{r})) \\
&\quad + \frac{1}{L}\sum_{\substack{J=[\hat{l},\hat{r}]\subset[0,L]\\ Ls\le\hat{r}\le Lt}} \Phi(J;\zeta_L(\hat{r})+\frac{\zeta_2}{\sqrt{t-s}}) + O(\frac{(\ln L)^{10}}{L}) \\
&= \int_0^1 \varphi(\zeta_0(1-x)+\zeta_1+\frac{\zeta_2}{\sqrt{t-s}}1_{[s,t]}(x))dx + O(\frac{(\ln L)^{10}}{L}).
\end{aligned}$$

Note that

$$\frac{\partial}{\partial\zeta_2}\Phi^{(s,t)}(J;\zeta_0,\zeta_1,\frac{\zeta_2}{\sqrt{t-s}}) = 0$$

if $J\cap[s,t]=\emptyset$. By analitycity, this means that for $\underline{\eta}\in\mathbf{R}^3$

$$\begin{aligned}
(3.14)\qquad &\underline{\eta}\cdot\big[Hess_{(\zeta_0,\zeta_1,\zeta_2)}\varphi_L^{(s,t)}(\zeta_0,\zeta_1,\frac{\zeta_2}{\sqrt{t-s}})\big]\underline{\eta} \\
&= \int_0^1 \{(1-x)\eta_0+\eta_1+\frac{\eta_2}{\sqrt{t-s}}1_{[s,t]}(x)\}^2 \\
&\quad\times\varphi''(\zeta_0(1-x)+\zeta_1+\frac{\zeta_2}{\sqrt{t-s}}1_{[s,t]}(x))dx \\
&\quad+|\eta|^2 O\big(L^{-\frac{1}{5}}(\ln L)^{10}\big)
\end{aligned}$$

as long as $t-s>L^{-\frac{4}{5}}$. If $(\zeta_0,\zeta_1,\zeta_2)$ satisfies (3.10), then as in the proof of Lemma 2.5, we have some $\alpha_1^0>0$ depending only on μ and δ such that

$$\alpha_1^0\le\varphi''(\zeta_0(1-x)+\zeta_1+\frac{\zeta_2}{\sqrt{t-s}}1_{[s,t]}(x))$$

for every $x\in[0,1]$. Also, by analyticity, there exists $\alpha_2^0>0$ depending only on μ and δ such that

$$\varphi''(\zeta_0(1-x)+\zeta_1+\frac{\zeta_2}{\sqrt{t-s}}1_{[s,t]}(x))\le\alpha_2^0$$

for every $x \in [0,1]$. Therefore we have

(3.15)

$$\begin{aligned}
&\alpha_1^0\int_0^1 \Big\{\eta_0(1-x)+\eta_1+\frac{\eta_2}{\sqrt{t-s}}1_{[s,t]}(x)\Big\}^2 dx + O\big(L^{-\frac{1}{5}}(\ln L)^{10}\big)\cdot|\underline{\eta}|^2\\
\le& \text{ the RHS of (3.14)}\\
\le& \alpha_2^0\int_0^1 \Big\{\eta_0(1-x)+\eta_1+\frac{\eta_2}{\sqrt{t-s}}1_{[s,t]}(x)\Big\}^2 dx + O\big(L^{-\frac{1}{5}}(\ln L)^{10}\big)\cdot|\underline{\eta}|^2.
\end{aligned}$$

Further, since

$$\begin{aligned}
&\int_0^1 \Big\{\eta_0(1-x)+\eta_1+\frac{\eta_2}{\sqrt{t-s}}1_{[s,t]}(x)\Big\}^2 dx\\
=&\int_0^1 \big\{\eta_0(1-x)+\eta_1\big\}^2 dx + \eta_2^2 + \frac{2\eta_2}{\sqrt{t-s}}\int_s^t \big\{\eta_0(1-x)+\eta_1\big\}dx
\end{aligned}$$

Since we know that the first term in the RHS of the above equality is bounded from below by $\alpha_1^0(\eta_0^2+\eta_1^2)$, the RHS is bounded from below by

$$\begin{aligned}
&\alpha_1^0(\eta_0^2+\eta_1^2) - 2\sqrt{t-s}(|\eta_0\eta_2|+|\eta_1\eta_2|)+\eta_2^2\\
\ge&(\alpha_1^0-\sqrt{t-s})(\eta_0^2+\eta_1^2)+(1-2\sqrt{t-s})\eta_2^2.
\end{aligned}$$

Set $2\alpha_1 := \min\{\frac{1}{2}\alpha_0^1, \frac{1}{3}\}$. It is obvious that the RHS of the above inequality is larger than $\alpha_1|\underline{\eta}|^2$ if $\sqrt{t-s} < 2\alpha_1$. The existence of α_2 is obvious from (3.14).

References

[DH1] R.L. Dobrushin and O. Hryniv, Fluctuations of shapes of boundaries large areas under paths of random walks, Probab. Theo. Rel. Fields, **105**, 423–458(1996)

[DH2] R.L. Dobrushin and O. Hryniv, Fluctuations of the phase boundary in the 2D Ising Ferromagnet, Commun. Math. Phys. **189**, 395–445(1997)

[Ioffe] D. Ioffe, Ornstein-Zernike behaviour and analyticity of shapes for self-avoiding walks on $\mathbf{Z}^d$, Markov Process. Related Fields **4**, 323–350 (1998)

[KH] R. Kotecký and O. Hryniv Surface tension and the Ornstein-Zernike behaviour for the 2D Blume-Capel model, J. Stat. Phys. **106** (3-4), 431–476(2002)

[KP] R. Kotecký and D. Preis, Cluster expansions for abstract polymer models, Commun. Math. Phys. **103**, 491–498(1986)

Yasunari Higuchi
Department of mathematics, Faculty of Science
Kobe University
Nada-ku
Kobe, 657-8501
Japan

Joshin Murai
Graduate school of Humanities and Social Sciences
Okayama University
Tsushima
Okayama 700-8530
Japan

Jun Wang
Department of Mathematics, College of Science
Northern Jiaotong University
Beijing 100044
P.R.China

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 283–306

Infinite Systems of Non-Colliding Brownian Particles

Makoto Katori, Taro Nagao and Hideki Tanemura

Abstract.

Non-colliding Brownian particles in one dimension is studied. N Brownian particles start from the origin at time 0 and then they do not collide with each other until finite time T. We derive the determinantal expressions for the multitime correlation functions using the self-dual quaternion matrices. We consider the scaling limit of the infinite particles $N \to \infty$ and the infinite time interval $T \to \infty$. Depending on the scaling, two limit theorems are proved for the multitime correlation functions, which may define temporally inhomogeneous infinite particle systems.

§1. Introduction

We consider the process $X(t)$, which represents the system of N Brownian motions in one dimension all started from the origin and conditioned never to collide with each other up to time T. If we take the limit $T \to \infty$, the system becomes a temporally homogeneous diffusion process $Y(t)$, which is the Doob h-transform [3] of the absorbing Brownian motion in a Weyl chamber

$$\mathbf{R}^N_< = \Big\{ \mathbf{x} = (x_1, x_2, \dots, x_N) \in \mathbf{R}^N ; x_1 < x_2 < \cdots < x_N \Big\},$$

with harmonic function $h_N(\mathbf{x}) = \prod_{1 \le i < j \le N} (x_j - x_i)$ [8]. By virtue of the Karlin-McGregor formula [12, 13], its transition density $f_N(t, \mathbf{x}, \mathbf{y})$ from the state $\mathbf{x}$ to $\mathbf{y}$ in $\mathbf{R}^N_<$ in time period $t > 0$ is given by

$$f_N(t, \mathbf{x}, \mathbf{y}) = \det_{1 \le i,j \le N} \Big(p_t(x_i, y_j) \Big),$$

where $p_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t}$. On the other hand, if the non-colliding time interval T remains finite, the process $X(t), 0 \le t \le T$, is temporally inhomogeneous [15].

Received January 6, 2003.
Revised April 5, 2003.

We notice an integral formula found in Harish-Chandra [9], Itzykson and Zuber [10], and Mehta [16],

$$\frac{\det_{1\le i,j\le N}(p_t(x_i,y_j))}{h_N(\mathbf{x})h_N(\mathbf{y})} = c\int dU\ \exp\left[-\frac{1}{2t}\mathrm{tr}(X-U^{\dagger}YU)^2\right]$$

with $c^{-1}=(2\pi)^{N/2}t^{N^2/2}\prod_{i=1}^{N+1}\Gamma(i)$, where X and Y are the $N\times N$ diagonal matrices, $X_{ij}=x_i\delta_{ij}, Y_{ij}=y_i\delta_{ij}$, and the integral is taken over the group of unitary matrix U of size N. This equality implies that the non-colliding Brownian motions such as $X(t)$ and $Y(t)$ can be described by using the eigenvalue-statistics of Hermitian random matrices in Gaussian ensembles [18]. In earlier papers[14, 15], it was shown that $Y(t)$ is identified with Dyson's Brownian motion model with $\beta=2$ [4] and the particle distribution is expressed by the probability density of eigenvalues of random matrices in the Gaussian unitary ensemble (GUE) with variance t, while $\sqrt{\frac{T}{t(2T-t)}}X(t)$ coincides with the distribution of eigenvalues of random matrices in the Pandey-Mehta ensemble [19, 25] with $\alpha=\sqrt{\frac{T-t}{T}}$, and this temporally inhomogeneous process exhibits a transition from the GUE statistics to the Gaussian orthogonal ensemble (GOE) statistics as the time t goes on from 0 to T.

It is known that the eigenvalue distributions of Hermitian random matrices have determinantal expressions. For instance, in the GUE, the probability density of N eigenvalues is expressed by

$$\rho_N(x_1,x_2,\dots,x_N)=\frac{1}{N!}\det_{1\le i,j\le N}(K_N(x_i,x_j)),$$

with $K_N(x,y)=\sum_{\ell=0}^{N-1}\varphi_\ell(x)\varphi_\ell(y)$, where

$$\varphi_\ell(x)=\frac{1}{\sqrt{h_\ell}}e^{-x^2/2}H_\ell(x) \tag{1.1}$$

with the ℓ-th Hermite polynomial $H_\ell(x)$ and $h_\ell=\sqrt{\pi}\,2^\ell\,\ell!$. By the orthogonality of $\varphi_k(x)$, we can prove the equality

$$\begin{aligned}&\int \det_{1\le i,j\le N'} K_N(x_i,x_j)\,dx_{N'}\\ &\quad=(N-N'+1)\det_{1\le i,j\le N'-1}K_N(x_i,x_j)\end{aligned} \tag{1.2}$$

for any $1\le N'\le N$. Such integral property enables us not only to obtain determinantal expressions for correlation functions, but also to argue the $N\to\infty$ limit of the system by studying the large N asymptotic of the function $K_N(x,y)$. With proper scaling limit, determinantal point processes with sine-kernel and Airy-kernel are derived. See [27] and references therein.

In the present paper, we derive the determinantal expressions of the multitime correlation functions for the process $X(t)$. Our aim is to prove limit theorems of the multitime correlation functions in the scaling limits of infinite particles $N \to \infty$ and infinite time interval $T \to \infty$. Depending on the scaling, we derive two kinds of limit theorems, one of which provides a spatially homogeneous but temporally inhomogeneous infinite particle system (Theorem 1), and other of which does the system with inhomogeneity both in space and time (Theorem 2). We remark that it is easier to prove the limit theorems for Dyson's Brownian motion model $Y(t)$. Corresponding to Theorem 1, we will obtain the multitime correlation functions of the homogeneous system, which coincides with the system studied by Spohn [28], Osada [24], and Nagao and Forrester [21]. Similarly, corresponding to Theorem 2, an infinite system with spatial inhomogeneity will be derived, which is related with the Airy process recently studied by Prähofer and Spohn [26] and Johansson [11].

One of the key points of our arguments is that, in order to give the determinantal expressions for the correlation functions for the present processes, we shall prepare matrices with the elements, which are neither real nor complex numbers, but quaternions

$$q = q_0 + q_1 e_1 + q_2 e_2 + q_3 e_3 \in \mathbf{Q}$$

with $q_i \in \mathbf{C}, 0 \le i \le 3$, in which the four basic units $\{1, e_1, e_2, e_3\}$ have the following 2×2 matrix representations, $C : \mathbf{Q} \mapsto \mathrm{Mat}_2(\mathbf{C})$;

$$C(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad C(e_1) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$
$$C(e_2) = \begin{pmatrix} 0 & -\sqrt{-1} \\ -\sqrt{-1} & 0 \end{pmatrix}, \quad C(e_3) = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}.$$

The dual of a quaternion q is defined by $q^\dagger = q_0 - \sum_{i=1}^3 q_i e_i$, and for a quaternion matrix $Q = (q_{ij}), q_{ij} \in \mathbf{Q}$, its dual matrix $Q^\dagger = ((Q^\dagger)_{ij})$ is defined to have the elements $(Q^\dagger)_{ij} = q_{ji}^\dagger$. Following Dyson's definition of the quaternion determinant for self-dual matrices [5, 17, 18], we can give the quaternion determinantal expressions having the similar properties to (1.2) for arbitrary multitime correlation functions for $X(t)$ (Theorem 3). As briefly reported in [23], the present results can be regarded as simple applications of the results given in Nagao and Forrester [22] and Nagao [20] for multimatrix models, and in Forrester, Nagao and Honner [6] for the asymptotic of quaternion determinantal systems, here we give, however, a self-contained explanation for all the formulae and calculus developed in the random matrix theory, which are used to prove our limit theorems.

The theorems proved here mean the convergence of processes in the sense of finite dimensional distributions. As argued in Prähofer and Spohn [26] and in Johansson [11], tightness in time should be confirmed.

§2. Statement of Results

For a given $T > 0$, we define

$$g_N^T(s,\mathbf{x};t,\mathbf{y}) = \frac{f_N(t-s,\mathbf{x},\mathbf{y})\mathcal{N}_N(T-t,\mathbf{y})}{\mathcal{N}_N(T-s,\mathbf{x})} \tag{2.1}$$

for $0 \leq s \leq t \leq T, \mathbf{x},\mathbf{y} \in \mathbf{R}^N_<$, where $\mathcal{N}_N(t,\mathbf{x}) = \int_{\mathbf{R}^N_<} f_N(t,\mathbf{x},\mathbf{y})d\mathbf{y}$, which is the probability that a Brownian motion started at $\mathbf{x} \in \mathbf{R}^N_<$ does not hit the boundary of $\mathbf{R}^N_<$ up to time $t > 0$. The function $g_N^T(s,\mathbf{x};t,\mathbf{y})$ can be regarded as the transition probability density from the state $\mathbf{x} \in \mathbf{R}^N_<$ at time s to the state $\mathbf{y} \in \mathbf{R}^N_<$ at time t, and associated with the temporally inhomogeneous diffusion process, which is the N Brownian motions conditioned not to collide with each other in a time interval $[0,T]$. In [14, 15] it was shown that as $|\mathbf{x}| \to 0$, $g_N^T(0,\mathbf{x},t,\mathbf{y})$ converges to

$$g_N^T(0,\mathbf{0},t,\mathbf{y}) \equiv C(N,T,t)h_N(\mathbf{y})\mathcal{N}_N(T-t,\mathbf{y})\prod_{i=1}^N p_t(0,y_i), \tag{2.2}$$

where $C(N,T,t) = \pi^{N/2}\left(\prod_{j=1}^N \Gamma(j/2)\right)^{-1} T^{N(N-1)/4}t^{-N(N-1)/2}$. Then the diffusion process $X(t)$ starting from $\mathbf{0}$ can be constructed.

We denote by $\mathfrak{X}$ the space of countable subset ξ of $\mathbf{R}$ satisfying $\sharp(\xi \cap K) < \infty$ for any compact subset K. We introduce the map γ from $\bigcup_{n=1}^\infty \mathbf{R}^n$ to $\mathfrak{X}$ defined by $\gamma(x_1,x_2,\dots,x_n) = \{x_i\}_{i=1}^n$. Then $\Xi^N(t) = \gamma X(t)$ is the diffusion process on the set $\mathfrak{X}$ with transition density function $\widetilde{g}_N^T(s,\xi;t,\eta)$, $0 \leq s \leq t \leq T$:

$$\widetilde{g}_N^T(s,\xi;t,\eta) = \begin{cases} g_N^T(s,\mathbf{x};t,\mathbf{y}), & \text{if } s > 0,\ \sharp\xi = \sharp\eta = N, \\ g_N^T(0,\mathbf{0};t,\mathbf{y}), & \text{if } s = 0,\ \xi = \{0\},\ \sharp\eta = N, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the elements of $\mathbf{R}^N_<$ with $\xi = \gamma\mathbf{x}$, $\eta = \gamma\mathbf{y}$. For $\mathbf{x}_N^{(m)} \in \mathbf{R}^N_<$, $1 \leq m \leq M+1$, and $N' = 1,2,\dots,N$, we put $\mathbf{x}_{N'}^{(m)} =$

$\left(x_1^{(m)}, x_2^{(m)}, \dots, x_{N'}^{(m)}\right)$ and $\xi_m^{N'} = \gamma \mathbf{x}_{N'}^{(m)}$. For a given time interval $[0,T]$, we consider the M intermediate times $0 < t_1 < t_2 < \cdots < t_M < T$. Then the multitime transition density function of the process $\Xi^N(t)$ is given by

$$\rho_N^T(t_1, \xi_1^N; \dots; t_{M+1}, \xi_{M+1}^N) = \prod_{m=0}^{M} \widetilde{g}_N^T(t_m, \xi_m^N; t_{m+1}, \xi_{m+1}^N), \tag{2.3}$$

where, for convenience, we set $t_0 = 0$, $t_{M+1} = T$ and $\xi_0^N = \{0\}$. From (2.1) and (2.2) we have

$$\begin{aligned} &\rho_N^T(t_1, \xi_1^N; t_2, \xi_2^N; \dots; t_{M+1}, \xi_{M+1}^N) \\ &= C(N,T,t_1) h_N\left(\mathbf{x}_N^{(1)}\right) \operatorname{sgn}\left(h_N\left(\mathbf{x}_N^{(M+1)}\right)\right) \\ &\times \prod_{i=1}^{N} p_{t_1}\left(0, x_i^{(1)}\right) \prod_{m=1}^{M} \det_{1 \le i,j \le N}\left(p_{t_{m+1}-t_m}\left(x_i^{(m)}, x_j^{(m+1)}\right)\right). \end{aligned} \tag{2.4}$$

For a sequence $\{N_m\}_{m=1}^{M+1}$ of positive integers less than or equal to N, we define the $(N_1, N_2, \dots, N_{M+1})$-multitime correlation function by

$$\begin{aligned} &\rho_N^T\left(t_1, \xi_1^{N_1}; t_2, \xi_2^{N_2}; \dots; t_{M+1}, \xi_{M+1}^{N_{M+1}}\right) \\ &= \int_{\prod_{m=1}^{M+1} \mathbf{R}^{N-N_m}} \prod_{m=1}^{M+1} \frac{1}{(N-N_m)!} \prod_{i=N_m+1}^{N} dx_i^{(m)} \\ &\times \rho_N^T(t_1, \xi_1^N; t_2, \xi_2^N; \dots; t_{M+1}, \xi_{M+1}^N). \end{aligned} \tag{2.5}$$

We will study limit theorems of the correlation functions $\rho_N^{T_N}$ as $N \to \infty$. First, we consider the case $T_N = 2N$. Let

$$\begin{aligned} &\widetilde{\mathbb{S}}(s,x;t,y) \\ &= \begin{cases} \dfrac{1}{\pi} \displaystyle\int_0^1 d\lambda \, \cos(\lambda(x-y)) e^{-\lambda^2(t-s)/2} & \text{if } s > t \\ \dfrac{\sin(x-y)}{\pi(x-y)} & \text{if } s = t \\ -\dfrac{1}{\pi} \displaystyle\int_1^\infty d\lambda \, \cos(\lambda(x-y)) e^{-\lambda^2(t-s)/2} & \text{if } s < t \end{cases} \end{aligned}$$

$$\mathbb{D}(s,x;t,y) = -\frac{1}{\pi}\int_0^1 d\lambda\ \lambda\sin(\lambda(x-y))e^{-(s+t)\lambda^2/2}$$
$$\widetilde{\mathbb{I}}(s,x;t,y) = -\frac{1}{\pi}\int_1^\infty d\lambda\ \frac{1}{\lambda}\sin(\lambda(x-y))e^{(s+t)\lambda^2/2}.$$

And let $\mathbf{q}^{m,n}(x,y)$ be the quaternion, whose 2×2 matrix expression is given by

$$C(\mathbf{q}^{m,n}(x,y)) = \left(\begin{array}{cc} \widetilde{\mathbb{S}}(s_m,x;s_n,y) & \widetilde{\mathbb{I}}(s_m,x;s_n,y) \\ \mathbb{D}(s_m,x;s_n,y) & \widetilde{\mathbb{S}}(s_n,y;s_m,x) \end{array}\right).$$

Let $M\geq 1$ and $\{N_m\}_{m=1}^{M+1}$ be a sequence of positive integers. We denote by $\mathbb{Q}\left(\mathbf{x}_{N_1}^{(1)},\mathbf{x}_{N_2}^{(2)},\dots,\mathbf{x}_{N_{M+1}}^{(M+1)}\right)$ the self-dual $\sum_{m=1}^{M+1}N_m\times\sum_{m=1}^{M+1}N_m$ quaternion matrix whose elements are $\mathbf{q}^{m,n}\left(x_i^{(m)},x_j^{(n)}\right)$, $1\leq i\leq N_m$, $1\leq j\leq N_n$, $1\leq m,n\leq M+1$, that is,

$$\mathbb{Q}\left(\mathbf{x}_{N_1}^{(1)},\mathbf{x}_{N_2}^{(2)},\dots,\mathbf{x}_{N_{M+1}}^{(M+1)}\right)$$
$$= \left[\begin{array}{ccc} \mathbb{Q}^{1,1}\left(\mathbf{x}_{N_1}^{(1)},\mathbf{x}_{N_1}^{(1)}\right) & \cdots & \mathbb{Q}^{1,M+1}\left(\mathbf{x}_{N_1}^{(1)},\mathbf{x}_{N_{M+1}}^{(M+1)}\right) \\ \mathbb{Q}^{2,1}\left(\mathbf{x}_{N_2}^{(2)},\mathbf{x}_{N_1}^{(1)}\right) & \cdots & \mathbb{Q}^{2,M+1}\left(\mathbf{x}_{N_2}^{(2)},\mathbf{x}_{N_{M+1}}^{(M+1)}\right) \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \mathbb{Q}^{M+1,1}\left(\mathbf{x}_{N_{M+1}}^{(M+1)},\mathbf{x}_{N_1}^{(1)}\right) & \cdots & \mathbb{Q}^{M+1,M+1}\left(\mathbf{x}_{N_{M+1}}^{(M+1)},\mathbf{x}_{N_{M+1}}^{(M+1)}\right) \end{array}\right]$$

with blocks of $N_m\times N_n$ quaternion matrices

$$\mathbb{Q}^{m,n}\left(\mathbf{x}_{N_m}^{(m)},\mathbf{x}_{N_n}^{(n)}\right) = \left(\mathbf{q}^{m,n}\left(x_i^{(m)},x_j^{(n)}\right)\right)_{1\leq i\leq N_m,1\leq j\leq N_n},$$

for $1\leq m,n\leq M+1$.

For an $N\times N$ self-dual quaternion matrix Q, the quaternion determinant $\mathrm{Tdet}Q$ is defined by Dyson [5] as

$$\mathrm{Tdet}Q = \sum_{\pi\in S_N}(-1)^{N-\ell(\pi)}\prod_1^{\ell(\pi)} q_{ab}q_{bc}\cdots q_{da},$$

where $\ell(\pi)$ denotes the number of exclusive cycles of the form $(a\to b\to c\to\cdots\to d\to a)$ included in a permutation $\pi\in S_N$.

Theorem 1. *Let* $T_N = 2N$. *For any* $M \geq 1$, *any sequence* $\{N_m\}_{m=1}^{M+1}$ *of positive integers, and any strictly increasing sequence* $\{s_m\}_{m=1}^{M+1}$ *of nonpositive numbers with* $s_{M+1} = 0$,

$$\begin{aligned} & \rho\left(s_1, \xi_1^{N_1}; s_2, \xi_2^{N_2}; \dots; s_M, \xi_M^{N_M}; 0, \xi_{M+1}^{N_{M+1}}\right) \\ \equiv & \lim_{N\to\infty} \rho_N^{T_N}\left(T_N + s_1, \xi_1^{N_1}; T_N + s_2, \xi_2^{N_2}; \dots; T_N, \xi_{M+1}^{N_{M+1}}\right) \\ = & \operatorname{Tdet}\mathbb{Q}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_{M+1}}^{(M+1)}\right). \end{aligned}$$

Remark 1. The above system is spatially homogeneous, since all elements of the quaternion determinant are functions of difference of positions, $x_i^{(m)} - x_j^{(n)}$. This expresses the bulk property of our infinite particle system. When $M = 1$, the present system is equivalent with the $N \to \infty$ limit of the two-matrix model reported by Pandey and Mehta [19, 25]. In the system defined by Theorem 1, if we take the further limit such that $s_m \to -\infty$ with the time difference $s_n - s_m$ fixed, $1 \leq m, n \leq M$, then $\mathbb{D}(s_m, x; s_n, y) \to \infty$, $\widetilde{\mathbb{I}}(s_m, x; s_n, y) \to 0$, while the product $\mathbb{D}(s_m, x; s_n, y)\widetilde{\mathbb{I}}(s_m, x; s_n, y) \to 0$. Therefore, we may replace $\mathbb{D}$ and $\widetilde{\mathbb{I}}$ by zeros in this limit, and the quaternion determinant Tdet $\mathbb{Q}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_M}^{(M)}\right)$ will be reduced to an ordinary determinant det $\mathbb{A}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_M}^{(M)}\right)$ with the elements $\mathbf{a}^{m,n}\left(x_i^{(m)}, x_j^{(n)}\right) = \widetilde{\mathbb{S}}\left(s_m, x_i^{(m)}; s_n, x_j^{(n)}\right)$. Hence, we obtain a temporally and spatially homogeneous system, whose correlation functions are given by

$$\rho'\left(s_1, \xi_1^{N_1}; s_2, \xi_2^{N_2}; \dots; s_M, \xi_M^{N_M}\right) = \det\mathbb{A}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_M}^{(M)}\right).$$

Such a homogeneous system was studied by Spohn [28], Osada [24] and Nagao and Forrester [21] as an infinite particle limit of Dyson's Brownian motion model[4].

Next, we consider the case that $T_N = 2N^{1/3}$. In order to state the result, we have to introduce the following functions. Let $\mathrm{Ai}(z)$ be the Airy function:

$$\mathrm{Ai}(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{\sqrt{-1}(zt+t^3/3)}dt. \tag{2.6}$$

For $s, t < 0$ and $x, y \in \mathbf{R}$, we put

$$\begin{aligned}\mathcal{D}(s,x;t,y) &= \frac{1}{4}\left[\int_0^\infty d\lambda\ e^{s\lambda/2}\mathrm{Ai}(x+\lambda)\frac{d}{d\lambda}\left\{e^{t\lambda/2}\mathrm{Ai}(y+\lambda)\right\}\right.\\&\quad\left.-\int_0^\infty d\lambda\ e^{t\lambda/2}\mathrm{Ai}(y+\lambda)\frac{d}{d\lambda}\left\{e^{s\lambda/2}\mathrm{Ai}(x+\lambda)\right\}\right],\\ \widetilde{\mathcal{I}}(s,x;t,y) &= \int_0^\infty d\lambda\ e^{t\lambda/2}\mathrm{Ai}(y-\lambda)\int_\lambda^\infty d\lambda'\ e^{s\lambda'/2}\mathrm{Ai}(x-\lambda')\\&\quad-\int_0^\infty d\lambda\ e^{s\lambda/2}\mathrm{Ai}(x-\lambda)\int_\lambda^\infty d\lambda'\ e^{t\lambda'/2}\mathrm{Ai}(y-\lambda'),\end{aligned}$$

and

$$\widetilde{\mathcal{S}}(s,x;t,y) = \mathcal{S}(s,x;t,y) - \mathcal{P}(s,x;t,y)1(s<t),$$

with

$$\begin{aligned}\mathcal{S}(s,x;t,y) &= \int_0^\infty d\lambda\ e^{(t-s)\lambda/2}\mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda)\\&\quad+\frac{1}{2}\mathrm{Ai}(y)\int_0^\infty d\lambda\ e^{s\lambda/2}\mathrm{Ai}(x-\lambda),\\ \mathcal{P}(s,x;t,y) &= \int_{-\infty}^\infty d\lambda\ e^{(t-s)\lambda/2}\mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda),\end{aligned}$$

where $1(s<t) = 1$ if $s < t$, and $= 0$ otherwise. And let $\mathfrak{q}^{m,n}(x,y)$ be the quaternion, whose 2×2 matrix expression is given by

$$C(\mathfrak{q}^{m,n}(x,y)) = \begin{pmatrix} \widetilde{\mathcal{S}}(s_m,x;s_n,y) & \widetilde{\mathcal{I}}(s_m,x;s_n,y) \\ \mathcal{D}(s_m,x;s_n,y) & \widetilde{\mathcal{S}}(s_n,y;s_m,x) \end{pmatrix}.$$

Let $M \geq 1$ and $\{N_m\}_{m=1}^{M+1}$ be a sequence of positive integers. We denote by $\mathcal{Q}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_{M+1}}^{(M+1)}\right)$ the self-dual $\sum_{m=1}^{M+1} N_m \times \sum_{m=1}^{M+1} N_m$ quaternion matrix whose elements are $\mathfrak{q}^{m,n}\left(x_i^{(m)}, x_j^{(n)}\right)$, $1 \leq i \leq N_m$, $1 \leq j \leq N_n$, $1 \leq m,n \leq M+1$.

Theorem 2. *Let* $T_N = 2N^{1/3}$ *and* $a_N(s) = 2N^{2/3} - s^2/4$ *for* $s \in \mathbf{R}$. *For any* $M \geq 1$, *any sequence* $\{N_m\}_{m=1}^{M+1}$ *of positive integers, and any strictly increasing sequence* $\{s_m\}_{m=1}^{M+1}$ *of nonpositive numbers with* $s_{M+1} = 0$,

$$\begin{aligned}&\widehat{\rho}\left(s_1, \xi_1^{N_1}; \dots; s_{M+1}, \xi_{M+1}^{N_{M+1}}\right)\\ \equiv\ &\lim_{N\to\infty}\rho_N^{T_N}\left(T_N + s_1, \theta_{a_N(s_1)}\xi_1^{N_1}; \dots; T_N, \theta_{a_N(s_{M+1})}\xi_{M+1}^{N_{M+1}}\right)\\ =\ &\mathrm{Tdet}\,\mathcal{Q}\left(\mathbf{x}_{N_1}^{(1)}, \dots, \mathbf{x}_{N_{M+1}}^{(M+1)}\right),\end{aligned}$$

where $\theta_u\{x_i\} = \{x_i + u\}$.

Remark 2. This theorem may define an infinite particle system, in which any type of space-time correlation function is given by the above quaternion determinant. This quaternion determinantal system is the same as that derived in Forrester, Nagao and Honner [6], and it is inhomogeneous both in space and time. The spatial inhomogeneity is attributed to the fact that this system expresses the edge property of the infinite non-colliding Brownian particles. Thus, if we take the bulk limit, $x_i^{(m)} \to -\infty$ with the position differences $x_i^{(m)} - x_j^{(n)}$ fixed, then the system should recover spatial homogeneity. It is confirmed by observing that the quaternion determinantal system given in Theorem 1 can be derived as the bulk limit of the system of Theorem 2, if we use the asymptotic expansion of the Airy function (2.6) [1],

$$\mathrm{Ai}(-x) \sim \frac{1}{\pi^{1/2}x^{1/4}} \cos\left(\frac{2}{3}x^{3/2} - \frac{\pi}{4}\right) \quad \text{as } x \to \infty.$$

On the other hand, keeping the spatial inhomogeneity, one can consider the limit $s_m \to -\infty$ with the time difference $s_n - s_m$ fixed, $1 \leq m, n \leq M$. In this limit, $\mathcal{D}(s_m, x; s_n, y) \to 0$, $\widetilde{\mathcal{I}}(s_m, x; s_n, y) \to 0$, and

$$\mathcal{S}(s_m, x; s_n, y) \to \int_0^\infty d\lambda\; e^{(s_n - s_m)\lambda/2}\mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda).$$

Hence the off-diagonal elements vanish in the 2×2 matrix expressions of quaternion $\mathfrak{q}^{m,n}(x,y)$ and

$$C(\mathfrak{q}^{m,n}(x,y)) \to \begin{pmatrix} \mathfrak{a}^{m,n}(x,y) & 0 \\ 0 & \mathfrak{a}^{n,m}(y,x) \end{pmatrix}$$

for $1 \leq m, n \leq M$, where

$$\begin{aligned} \mathfrak{a}^{m,n}(x,y) &= \mathfrak{a}(s_m, x; s_n, y) \\ &= \begin{cases} \displaystyle\int_0^\infty d\lambda\; e^{(s_n - s_m)\lambda/2}\mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda) & \text{if } m \geq n \\ \displaystyle -\int_{-\infty}^0 d\lambda\; e^{(s_n - s_m)\lambda/2}\mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda) & \text{if } m < n. \end{cases} \end{aligned}$$

Then the quaternion determinant $\mathrm{Tdet}\mathcal{Q}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_M}^{(M)}\right)$ is reduced to an ordinary determinant $\det \mathcal{A}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_M}^{(M)}\right)$ with the elements $\mathfrak{a}^{m,n}\left(x_i^{(m)},x_j^{(n)}\right)$. In this way, we will obtain the infinite particle system, which is temporally homogeneous but spatially inhomogeneous with the multitime correlation functions

$$\widetilde{\rho}'\left(s_1,\xi_1^{N_1};s_2,\xi_2^{N_2};\dots;s_M,\xi_M^{N_M}\right)=\det\mathcal{A}\left(\mathbf{x}_{N_1}^{(1)},\mathbf{x}_{N_2}^{(2)},\dots,\mathbf{x}_{N_M}^{(M)}\right).$$

In particular, if we set $N_1=N_2=\cdots=N_M=1$, then

$$\widetilde{\rho}'\left(s_1,\{x^{(1)}\};\dots;s_M,\{x^{(M)}\}\right)=\det_{1\le m,n\le M}\left(\mathfrak{a}^{m,n}\left(x^{(m)},x^{(n)}\right)\right).$$

This is the same as the system called the Airy process by Prähofer and Spohn in [26]. (See also [11].)

§3. Quaternion determinantal expressions of the correlations

In this section we give quaternion determinantal expressions for the correlation functions defined in (2.5) along the procedure in [20]. From now on we consider the case N is even, for simplicity of notations. See [20], for necessary modifications for odd case. For $1\le m,n\le M+1$, define

$$(3.1)\qquad F^{m,n}(x,y)=\int_{-\infty}^{\infty}dw\int_{-\infty}^{w}dz\left|\begin{array}{cc} p_{T-t_m}(x,z) & p_{T-t_m}(x,w)\\ p_{T-t_n}(y,z) & p_{T-t_n}(y,w)\end{array}\right|,$$

where $p_0(y,x)dy=\delta_x(dy)$. We introduce an antisymmetric inner products

$$\langle f,g\rangle_m=\int_{\mathbf{R}}dx\int_{\mathbf{R}}dy\ F^{m,m}(x,y)f(x)g(y),$$

and

$$\langle f,g\rangle=\int_{\mathbf{R}}dx\int_{\mathbf{R}}dy\ F^{1,1}(x,y)p_{t_1}(0,x)p_{t_1}(0,y)f(x)g(y).$$

For $k=0,1,\dots$ we consider the polynomials in x of degree k defined by

$$(3.2)\qquad R_k(x)=z_1^{-k}\sum_{j=1}^{k}\alpha_{kj}H_j\left(\frac{x}{c_1}\right)z_1^j,$$

where $c_1=\sqrt{\frac{t_1(2T-t_1)}{T}}$, $z_1=\sqrt{\frac{2T-t_1}{t_1}}$,

$$(3.3)\qquad \alpha_{kj}=\begin{cases}2^{-k}c_1^k\delta_{kj}, & \text{if } k \text{ is even},\\ 2^{-k}c_1^k\Big\{\delta_{kj}-2(k-1)\delta_{k-2\,j}\Big\}, & \text{if } k \text{ is odd},\end{cases}$$

and $H_j(x)$ are the Hermite polynomials. They are monic and satisfy the skew orthogonal relations:

$$\begin{aligned}&\langle R_{2j}, R_{2\ell+1}\rangle = -\langle R_{2\ell+1}, R_{2j}\rangle = r_j \delta_{j\ell},\\ &\langle R_{2j}, R_{2\ell}\rangle = \langle R_{2j+1}, R_{2\ell+1}\rangle = 0, \quad j, \ell = 0, 1, 2, \dots,\end{aligned}$$

where

$$r_j = \frac{\Gamma(j+\frac{1}{2})\Gamma(j+1)}{\pi} \left(\frac{t_1^2}{T}\right)^{2j+1/2}.$$

For $m = 1, 2, \dots, M+1$, and $k = 0, 1, \dots$, put

$$R_k^{(m)}(x) = \int_{\mathbf{R}} dy\ R_k(y) p_{t_1}(0, y) p_{t_m - t_1}(y, x). \tag{3.4}$$

Then we can prove the skew orthogonal relations

$$\begin{aligned}&\langle R_{2j}^{(m)}, R_{2\ell+1}^{(m)}\rangle_m = -\langle R_{2\ell+1}^{(m)}, R_{2j}^{(m)}\rangle_m = r_j \delta_{j\ell},\\ &\langle R_{2j}^{(m)}, R_{2\ell}^{(m)}\rangle_m = \langle R_{2j+1}^{(m)}, R_{2\ell+1}^{(m)}\rangle_m = 0, \quad j, \ell = 0, 1, 2, \dots,\end{aligned}$$

for any $m = 1, 2, \dots, M+1$. For $m = 1, 2, \dots, M+1$, define

$$\Phi_k^{(m)}(x) = \int_{\mathbf{R}} dy\ R_k^{(m)}(y) F^{m,m}(y, x), \quad k = 0, 1, 2, \dots. \tag{3.5}$$

Now we introduce the functions on $\mathbf{R}^2$, $D^{m,n}, I^{m,n}$ and $S^{m,n}, 1 \le m, n \le M+1$, given by

$$D^{m,n}(x, y) = \sum_{k=0}^{(N/2)-1} \frac{1}{r_k} \Big[R_{2k}^{(m)}(x) R_{2k+1}^{(n)}(y) - R_{2k+1}^{(m)}(x) R_{2k}^{(n)}(y)\Big], \tag{3.6}$$

$$I^{m,n}(x, y) = -\sum_{k=0}^{(N/2)-1} \frac{1}{r_k} \Big[\Phi_{2k}^{(m)}(x) \Phi_{2k+1}^{(n)}(y) - \Phi_{2k+1}^{(m)}(x) \Phi_{2k}^{(n)}(y)\Big], \tag{3.7}$$

$$S^{m,n}(x, y) = \sum_{k=0}^{(N/2)-1} \frac{1}{r_k} \Big[\Phi_{2k}^{(m)}(x) R_{2k+1}^{(n)}(y) - \Phi_{2k+1}^{(m)}(x) R_{2k}^{(n)}(y)\Big]. \tag{3.8}$$

Further we define

$$\tilde{S}^{m,n}(x, y) = S^{m,n}(x, y) - p_{t_n - t_m}(x, y) 1(m < n), \tag{3.9}$$

$$\tilde{I}^{m,n}(x, y) = I^{m,n}(x, y) + F^{m,n}(x, y). \tag{3.10}$$

Define the quaternions $q^{m,n}(x,y), 1 \leq m,n \leq M+1, x,y \in \mathbf{R}$ so that these 2×2 matrix expressions $C(q^{m,n}(x,y))$ are given by

$$C(q^{m,n}(x,y)) = \begin{pmatrix} \widetilde{S}^{m,n}(x,y) & \widetilde{I}^{m,n}(x,y) \\ D^{m,n}(x,y) & \widetilde{S}^{n,m}(y,x) \end{pmatrix}.$$

Let $M \geq 1$ and $\{N_m\}_{m=1}^{M+1}$ be a sequence of positive integers less than or equal to N. For $\mathbf{x}_N^{(m)} \in \mathbf{R}_<^N$, $1 \leq m \leq M+1$, we denote by $Q\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_{M+1}}^{(M+1)}\right)$ the self-dual $\sum_{m=1}^{M+1} N_m \times \sum_{m=1}^{M+1} N_m$ quaternion matrix whose elements are $q^{m,n}\left(x_i^{(m)}, x_j^{(n)}\right)$, $1 \leq i \leq N_m$, $1 \leq j \leq N_n$, $1 \leq m,n \leq M+1$. Then we show the following relation.

Theorem 3. *The multitime correlation function* (2.5) *is written as*

$$\rho_N^T\left(t_1, \xi_1^{N_1}; \dots; t_{M+1}, \xi_{M+1}^{N_{M+1}}\right) = \text{Tdet}Q\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_{M+1}}^{(M+1)}\right).$$

In order to prove the theorem, first we introduce the Pfaffian. For an integer N and an antisymmetric $2N \times 2N$ matrix $A = (a_{ij})$, the Pfaffian is defined as

$$\begin{aligned} \text{Pf}(A) &= \text{Pf}_{1 \leq i < j \leq 2N}(a_{ij}) \\ &= \frac{1}{N!} \sum_{\sigma} \text{sgn}(\sigma) a_{\sigma(1)\sigma(2)} a_{\sigma(3)\sigma(4)} \cdots a_{\sigma(2N-1)\sigma(2N)}, \end{aligned}$$

where the summation is extended over all permutations σ of $(1,2,\dots,2N)$ with restriction $\sigma(2k-1) < \sigma(2k), k = 1,2,\dots,N$. If Q is an $N \times N$ self-dual quaternion matrix, then

$$\text{Tdet}Q = \text{Pf}\Big(JC(Q)\Big), \tag{3.11}$$

where J is an $2N \times 2N$ antisymmetric matrix with only non-zero elements

$$J_{2k+1,2k+2} = -J_{2k+2,2k+1} = 1, \quad k = 0,1,2,\dots,N-1.$$

See, for instance, Mehta [17].

For a function $\psi^{m,n}$ defined on $\mathbf{R}^2$ we denote the $N \times N$-matrices whose (i,j)-entry is $\psi^{m,n}\left(x_i^{(m)}, x_j^{(n)}\right)$ by $\psi^{m,n}\left(\mathbf{x}_N^{(m)}, \mathbf{x}_N^{(n)}\right)$, or simply by $\psi^{m,n}$ for short. And we denote by $R^{(m)}\left(\mathbf{x}_N^{(m)}\right)$ the $N \times N$ matrix with $R^{(m)}\left(\mathbf{x}_N^{(m)}\right)_{i,j} = R_{j-1}^{(m)}(x_i)$, and by $\Phi^{(m)}\left(\mathbf{x}_N^{(m)}\right)$ that with

$\Phi^{(m)}\left(\mathbf{x}_N^{(m)}\right)_{i,j} = \Phi_{j-1}^{(m)}(x_i)$. Let L be the $N \times N$ diagonal matrix with $L_{i,i} = \sqrt{r_{[(i-1)/2]}}$, $i = 1, 2, \ldots, N$, and $\widetilde{R}^{(m)}\left(\mathbf{x}_N^{(m)}\right) = L^{-1}R^{(m)}\left(\mathbf{x}_N^{(m)}\right)$. Then we have

$$\widetilde{R}^{(m)}\left(\mathbf{x}_N^{(1)}\right) J \widetilde{R}^{(n)}\left(\mathbf{x}_N^{(1)}\right)^T = D^{m,n}\left(\mathbf{x}_N^{(m)}, \mathbf{x}_N^{(n)}\right). \tag{3.12}$$

As the first step of the proof of Theorem 3. We show that the multitime probability density defined in (2.3) is written as

$$\rho_N^T\left(t_1, \xi_1^N; \ldots; t_{M+1}, \xi_{M+1}^N\right) = \mathrm{Tdet}Q\left(\mathbf{x}_N^{(1)}, \ldots, \mathbf{x}_N^{(M+1)}\right). \tag{3.13}$$

For simplicity of notation, here we give the proof of (3.13) for $M = 2$. It is straightforward to prove (3.13) for general M. Since

$$\mathrm{sgn}\left(h_N\left(\mathbf{x}_N^{(3)}\right)\right) = \mathrm{Pf}_{1 \le i < j \le N}\left(\mathrm{sgn}\left(x_j^{(3)} - x_i^{(3)}\right)\right),$$

and $\mathrm{sgn}(y - x) = F^{3,3}(x, y)$, we have

$$\mathrm{sgn}\left(h_N\left(\mathbf{x}_N^{(3)}\right)\right) = \mathrm{Pf}\left[F^{3,3}\right]. \tag{3.14}$$

Noting that $R_k(x)$ is the monic polynomial of degree k, we have

$$h_N\left(\mathbf{x}^{(1)}\right) = \det_{1 \le i,j \le N}\left(\left(x_j^{(1)}\right)^{i-1}\right) = \det_{1 \le i,j \le N}\left(R_{i-1}\left(x_j^{(1)}\right)\right),$$

and so

$$\prod_{i=1}^{N} p_{t_1}\left(0, x_i^{(1)}\right) h_N\left(\mathbf{x}^{(1)}\right) = \det\left[R^{(1)}\left(\mathbf{x}_N^{(1)}\right)\right] \tag{3.15}$$

Since $\det L = \prod_{k=0}^{N/2-1} r_k = C(N, T, t_1)^{-1}$, from (3.12) and (3.15)

$$\begin{aligned} C(N, T, t_1)\prod_{i=1}^{N} p_{t_1}\left(0, x_i^{(1)}\right) h_N\left(\mathbf{x}^{(1)}\right) &= \det\left[\widetilde{R}^{(1)}\left(\mathbf{x}_N^{(1)}\right)\right] \\ &= \mathrm{Pf}\left[\widetilde{R}^{(1)}\left(\mathbf{x}_N^{(1)}\right) J \widetilde{R}^{(1)}\left(\mathbf{x}_N^{(1)}\right)^T\right] \\ &= \mathrm{Pf}\left[D^{1,1}\left(\mathbf{x}_N^{(1)}, \mathbf{x}_N^{(1)}\right)\right]. \end{aligned} \tag{3.16}$$

Then from (2.4), (3.14) and (3.16) we have

$$\begin{aligned}
&\rho_N^T(t_1,\xi_1^N;t_2,\xi_2^N;t_3,\xi_3^N)\\
&=\mathrm{Pf}[D^{1,1}]\mathrm{Pf}[F^{3,3}]\prod_{m=1}^{2}\det_{1\le i,j\le N}\left[p_{t_{m+1}-t_m}\right]\\
&=(-1)^{3N/2}\mathrm{Pf}\begin{bmatrix} D^{1,1} & O\\ O & -F^{3,3}\end{bmatrix}\prod_{m=1}^{2}\mathrm{Pf}\begin{bmatrix} O & -(p_{t_{m+1}-t_m})^T\\ p_{t_{m+1}-t_m} & O\end{bmatrix}.
\end{aligned}$$

By basic properties of the Pfaffians, we have

$$\begin{aligned}
&\mathrm{Pf}\begin{bmatrix} D^{1,1} & O\\ O & -F^{3,3}\end{bmatrix}\prod_{m=1}^{2}\mathrm{Pf}\begin{bmatrix} O & -(p_{t_{m+1}-t_m})^T\\ p_{t_{m+1}-t_m} & O\end{bmatrix}\\
&=\mathrm{Pf}\begin{bmatrix}
D^{1,1} & O & O & O & O & O\\
O & -F^{3,3} & O & O & O & O\\
O & O & O & -(p_{t_2-t_1})^T & O & O\\
O & O & p_{t_2-t_1} & O & O & O\\
O & O & O & O & O & -(p_{t_3-t_2})^T\\
O & O & O & O & p_{t_3-t_2} & O
\end{bmatrix}\\
&=\mathrm{Pf}\begin{bmatrix}
D^{1,1} & O & O & O & O & O\\
O & O & p_{t_2-t_1} & O & O & O\\
O & -(p_{t_2-t_1})^T & O & O & O & O\\
O & O & O & O & p_{t_3-t_2} & O\\
O & O & O & -(p_{t_3-t_2})^T & O & O\\
O & O & O & O & O & -F^{3,3}
\end{bmatrix}\\
&=\mathrm{Pf}\begin{bmatrix}
D^{1,1} & O & O & O & O & O\\
O & -F^{1,1} & p_{t_2-t_1} & -F^{1,2} & p_{t_3-t_1} & -F^{1,3}\\
O & -(p_{t_2-t_1})^T & O & O & O & O\\
O & -F^{2,1} & O & -F^{2,2} & p_{t_3-t_2} & -F^{2,3}\\
O & -(p_{t_3-t_1})^T & O & -(p_{t_3-t_2})^T & O & O\\
O & -F^{3,1} & O & -F^{3,2} & O & -F^{3,3}
\end{bmatrix}.
\end{aligned}$$

Since $\mathbf{x}_N^{(1)}\in\mathbf{R}_<^N$, $h_N\left(\mathbf{x}_N^{(1)}\right)\neq 0$, and so $\det\left[R^{(1)}\left(\mathbf{x}_N^{(1)}\right)\right]\neq 0$ by (3.15). Hence we can define matrices

$$U^{(m)}=R^{(m)}\left(\mathbf{x}_N^{(m)}\right)R^{(1)}\left(\mathbf{x}_N^{(1)}\right)^{-1},\quad V^{(m)}=\Phi^{(m)}\left(\mathbf{x}_N^{(m)}\right)R^{(1)}\left(\mathbf{x}_N^{(1)}\right)^{-1},$$

which satisfies

$$\begin{aligned}
U^{(m)}D^{1,1}(U^{(n)})^T&=D^{m,n}, & V^{(m)}D^{1,1}(V^{(n)})^T&=-I^{m,n},\\
V^{(m)}D^{1,1}(U^{(n)})^T&=S^{m,n}, & U^{(m)}D^{1,1}(V^{(n)})^T&=-(S^{n,m})^T.
\end{aligned}$$

By repeating elementary operations, we see that the last Pfaffian equals to

$$
\begin{aligned}
&\mathrm{Pf}\begin{bmatrix}
D^{1,1} & (S^{1,1})^T & D^{1,2} & (S^{2,1})^T & D^{1,3} & (S^{3,1})^T \\
-S^{1,1} & -\tilde{I}^{1,1} & -\tilde{S}^{1,2} & -\tilde{I}^{1,2} & -\tilde{S}^{1,3} & -\tilde{I}^{1,3} \\
D^{2,1} & (\tilde{S}^{1,2})^T & D^{2,2} & (S^{2,2})^T & D^{2,3} & S^{3,2} \\
-S^{2,1} & -\tilde{I}^{2,1} & -\tilde{S}^{2,2} & -\tilde{I}^{2,2} & -\tilde{S}^{2,3} & -\tilde{I}^{2,3} \\
D^{3,1} & (\tilde{S}^{1,3})^T & D^{3,2} & (\tilde{S}^{2,3})^T & D^{3,3} & (S^{3,3})^T \\
-S^{3,1} & -\tilde{I}^{3,1} & -S^{3,2} & -\tilde{I}^{3,2} & -S^{3,3} & -\tilde{I}^{3,3}
\end{bmatrix} \\
&= (-1)^{3N/2}\mathrm{Pf}\begin{bmatrix}
A^{1,1} & A^{1,2} & A^{1,3} \\
A^{2,1} & A^{2,2} & A^{2,3} \\
A^{3,1} & A^{3,2} & A^{3,3}
\end{bmatrix},
\end{aligned}
$$

where each $A^{m,n} = (A_{ij}^{m,n})$ is a $2N \times 2N$ matrix which consists of 2×2 blocks

$$
A_{ij}^{m,n} = \begin{pmatrix} D_{ij}^{m,n} & \tilde{S}_{ji}^{n,m} \\ -\tilde{S}_{ij}^{m,n} & -\tilde{I}_{ij}^{m,n} \end{pmatrix}.
$$

We can see that the above matrix $A = (A_{ij}^{m,n})$ satisfies the relation $A = JC(Q)$. Therefore, (3.13) is derived from (3.11).

For square integrable functions ϕ and ψ defined on $\mathbf{R}^2$, put $\phi * \psi(x,y) = \int_{\mathbf{R}} \phi(x,z)\psi(z,y)dz$. Then we have

$$
\begin{aligned}
&S^{m,p} * S^{p,m} = I^{m,p} * D^{p,n} = D^{m,p} * F^{p,n} = S^{m,p}, \\
&D^{m,p} * S^{p,n} = D^{m,n}, \quad S^{m,p} * I^{p,n} = S^{m,p} * F^{p,n} = I^{m,n}, \\
&S^{m,p} * p_{t_n - t_p} = S^{m,n}, \quad D^{m,p} * p_{t_n - t_p} = D^{m,n}, \quad \text{if } p < n.
\end{aligned}
$$

Hence by simple calculation we see that

$$
\begin{aligned}
&\int_{\mathbf{R}} q^{m,m}(z,z)dz = N, \\
&\int_{\mathbf{R}} q^{m,p}(x,z)q^{p,n}(z,y)dz = q^{m,n}(x,y) \\
&\qquad\qquad +q^{m,n}(x,y)\kappa(n,p) - \kappa(p,m)q^{m,n}(x,y),
\end{aligned}
$$

where $\kappa(n,p)$ is a quaternion with

$$
C(\kappa(n,p)) = \begin{pmatrix} 1 - 1(p < n) & 0 \\ 0 & -1(n < p) \end{pmatrix}.
$$

Then by slight modification of Theorem 6 in [22] we have the following integral formula for any $1 \leq N_m \leq N, m = 1, 2, \dots, M+1$,

$$\int_{\mathbf{R}} \mathrm{Tdet}Q\left(\mathbf{x}_{N_1}^{(1)}, \dots, \mathbf{x}_{N_m}^{(m)}, \dots, \mathbf{x}_{N_{M+1}}^{(M+1)}\right) dx_{N_m}^{(m)}$$
$$= (N - N_m + 1)\mathrm{Tdet}Q\left(\mathbf{x}_{N_1}^{(1)}, \dots, \mathbf{x}_{N_m-1}^{(m)}, \dots, \mathbf{x}_{N_{M+1}}^{(M+1)}\right),$$

which is the generalization of the formula (1.2) given in Introduction of the present paper. Successive application of the above relation yields Theorem 3.

§4. Expansion using Hermite polynomials

In this section we show expansions of functions $p_{t_n - t_m}$, $R_k^{(m)}$ and $\Phi_k^{(m)}$ by using Hermite polynomials H_k. Put

$$c_n = \sqrt{\frac{t_n(2T - t_n)}{T}}, \quad \gamma_n = -\frac{T - t_n}{T}, \quad z_n = \sqrt{\frac{2T - t_n}{t_n}},$$

and $\tau^{(n)} = -\log z_n$. By simple calculation we have

$$p_{t_n - t_m}(x, y) = \frac{e^{-(t_m/2T)(x/c_m)^2} e^{(t_n/2T)(y/c_n)^2}}{\sqrt{2\pi(t_n - t_m)}} \times \exp\left(-\frac{\left\{(y/c_n) - e^{-(\tau^{(n)} - \tau^{(m)})}(x/c_m)\right\}^2}{1 - e^{-2(\tau^{(n)} - \tau^{(m)})}}\right)$$

for $1 \leq m < n < M + 1$. Using Mehler's formula [2]

$$\exp\left(-\frac{(y - xz)^2}{1 - z^2}\right) = e^{-y^2}\sqrt{\pi(1 - z^2)} \sum_{k=0}^{\infty} \frac{z^k}{h_k} H_k(x) H_k(y),$$

we will have the following expansions using the Hermite polynomials. For $1 \leq m < n \leq M + 1$,

$$(4.1)\quad p_{t_n - t_m}(x, y) = \frac{\sqrt{T} e^{-\frac{1}{2}(1 + \gamma_m)(x/c_m)^2} e^{-\frac{1}{2}(1 - \gamma_n)(y/c_n)^2}}{\sqrt{t_n(2T - t_m)}} \times \sum_{k=0}^{\infty} \frac{e^{-k(\tau^{(n)} - \tau^{(m)})}}{h_k} H_k\left(\frac{x}{c_m}\right) H_k\left(\frac{y}{c_n}\right),$$

and for $1 < m \le M+1$,

$$p_{t_1}(0,x)p_{t_m-t_1}(x,y) = \frac{\sqrt{T}e^{-(x/c_1)^2}e^{-\frac{1}{2}(1-\gamma_m)(y/c_m)^2}}{\sqrt{2\pi t_1 t_m(2T-t_1)}} \times \sum_{k=0}^{\infty} \frac{e^{-k(\tau^{(m)}-\tau^{(1)})}}{h_k} H_k\left(\frac{x}{c_1}\right) H_k\left(\frac{y}{c_m}\right).$$

Then from (3.2), (3.4) and the orthogonal relation of the Hermitian polynomials, we obtain

$$(4.2) \qquad R_k^{(m)}(x) = \frac{e^{-\frac{1}{2}(1-\gamma_m)(x/c_m)^2}e^{k\tau^{(1)}}}{\sqrt{2\pi t_m}} \sum_{j=0}^{k} \alpha_{kj} e^{-j\tau^{(m)}} H_j\left(\frac{x}{c_m}\right).$$

From the definition (3.1) and the expansion (4.1) we can obtain

$$(4.3) \qquad F^{m,n}(x,y) = \frac{e^{-\frac{1}{2}(1+\gamma_m)(x/c_m)^2}e^{-\frac{1}{2}(1+\gamma_n)(y/c_n)^2}}{\sqrt{(2T-t_m)(2T-t_n)}} \times \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} \frac{e^{k\tau^{(m)}}e^{\ell\tau^{(n)}}}{h_k h_\ell} H_k\left(\frac{x}{c_m}\right) H_\ell\left(\frac{y}{c_n}\right) \times \left\langle H_k\left(\frac{\cdot}{\sqrt{T}}\right), H_\ell\left(\frac{\cdot}{\sqrt{T}}\right)\right\rangle_*,$$

where $\langle \cdot, \cdot \rangle_*$ is the antisymmetric inner product defined by

$$\langle f, g \rangle_* = \int_{-\infty}^{\infty} dw \int_{-\infty}^{w} dz \; e^{-(z^2+w^2)/2T}\Big[f(z)g(w) - f(w)g(z)\Big].$$

Put $R_k^*(x) = \sum_{j=0}^{k} \alpha_{kj} H_j\left(\frac{x}{c_{M+1}}\right)$. Then $\{R_k^*(x)\}$ satisfy the following skew orthogonal relations

$$(4.4) \qquad \begin{aligned} &\langle R_{2j}^*, R_{2\ell+1}^* \rangle_* = -\langle R_{2\ell+1}^*, R_{2j}^* \rangle_* = r_j^* \delta_{j\ell}, \\ &\langle R_{2j}^*, R_{2\ell}^* \rangle_* = 0, \quad \langle R_{2j+1}^*, R_{2\ell+1}^* \rangle_* = 0, \quad \text{for } j, \ell = 0, 1, 2, \dots, \end{aligned}$$

where $r_\ell^* = 4h_{2\ell}T(c_1/2)^{4\ell+1}$. We put

$$\beta_{kj} = \begin{cases} 2^k c_1^{-k} \delta_{j\,k}, & \text{if } \; k \text{ is even}, \\ 2^k \left(\frac{k-1}{2}\right)! \left\{ c_1^j \left(\frac{j-1}{2}\right)! \right\}^{-1}, & \text{if } \; k, j \text{ are odd and } k \ge j, \\ 0, & \text{otherwise}, \end{cases}$$

for nonnegative integers k and j. Then $\sum_{j=s}^{k} \beta_{kj}\alpha_{js} = \delta_{ks}$, if $0 \le s \le k$, and

$$H_k\left(\frac{x}{\sqrt{T}}\right) = \sum_{j=0}^{k} \beta_{kj} R_j^*(x). \tag{4.5}$$

From the definition (3.5) and the equations (4.2) and (4.3) we have

$$\begin{aligned}\Phi_k^{(m)}(x) &= \frac{c_m}{\sqrt{2\pi t_m}\,(2T - t_m)} e^{-\frac{1}{2}(1+\gamma_m)(x/c_m)^2} e^{k\tau^{(1)}} \\ &\times \sum_{\ell=0}^{\infty}\sum_{j=0}^{k} \frac{e^{\ell\tau^{(m)}}}{h_\ell} H_\ell\left(\frac{x}{c_m}\right) \alpha_{kj} \left\langle H_j\left(\frac{\cdot}{\sqrt{T}}\right), H_\ell\left(\frac{\cdot}{\sqrt{T}}\right)\right\rangle_* \\ &= \frac{e^{-\frac{1}{2}(1+\gamma_m)(x/c_m)^2}}{\sqrt{2\pi T(2T - t_m)}} e^{k\tau^{(1)}} \sum_{j=0}^{\infty} \langle R_k^*, R_j^*\rangle_* \sum_{\ell=j}^{\infty} \frac{e^{\ell\tau^{(m)}}}{h_\ell} H_\ell\left(\frac{x}{c_m}\right) \beta_{\ell j}.\end{aligned}$$

Using the skew orthogonal relations (4.4), we show that for $k = 0, 1, 2, \ldots$

$$\begin{aligned}\Phi_{2k}^{(m)}(x) &= \frac{e^{-\frac{1}{2}(1+\gamma_m)(x/c_m)^2} r_k^*}{\sqrt{2\pi T(2T - t_m)}} e^{2k\tau^{(1)}} \\ &\times \sum_{\ell=2k+1}^{\infty} \frac{e^{\ell\tau^{(m)}}}{h_\ell} \beta_{\ell 2k+1} H_\ell\left(\frac{x}{c_m}\right),\end{aligned} \tag{4.6}$$

$$\begin{aligned}\Phi_{2k+1}^{(m)}(x) &= -\frac{e^{-\frac{1}{2}(1+\gamma_m)(x/c_m)^2} r_k^*}{\sqrt{2\pi T(2T - t_m)}} e^{(2k+1)\tau^{(1)}} \\ &\times \sum_{\ell=2k}^{\infty} \frac{e^{\ell\tau^{(m)}}}{h_\ell} \beta_{\ell 2k} H_\ell\left(\frac{x}{c_m}\right).\end{aligned} \tag{4.7}$$

Using above expansions we show the following lemma.

Lemma 4. *For* $1 \le m, n \le M + 1$,

$$F^{m,n}(x,y) = \sum_{k=0}^{\infty} \frac{1}{r_k} \left[\Phi_{2k}^{(m)}(x)\Phi_{2k+1}^{(n)}(y) - \Phi_{2k+1}^{(m)}(x)\Phi_{2k}^{(n)}(y)\right], \tag{4.8}$$

$$\widetilde{I}^{m,n}(x,y) = \sum_{k=N/2}^{\infty} \frac{1}{r_k} \left[\Phi_{2k}^{(m)}(x)\Phi_{2k+1}^{(n)}(y) - \Phi_{2k+1}^{(m)}(x)\Phi_{2k}^{(n)}(y)\right]. \tag{4.9}$$

Proof. By (4.6), (4.7) and the relation

$$r_k = \frac{1}{2\pi T}\left(\frac{t_1}{2T - t_1}\right)^{2k+1/2} r_k^*, \tag{4.10}$$

we have

$$
\begin{aligned}
&\sum_{k=0}^{\infty}\frac{1}{r_k}\left[-\Phi^{(m)}_{2k+1}(x)\Phi^{(n)}_{2k}(y)+\Phi^{(m)}_{2k}(x)\Phi^{(n)}_{2k+1}(y)\right]\\
=\ &\frac{e^{-\frac{1}{2}(1+\gamma_m)(x/c_m)^2}e^{-\frac{1}{2}(1+\gamma_n)(y/c_n)^2}}{\sqrt{(2T-t_m)(2T-t_n)}}\sum_{k=0}^{\infty}r^*_k\\
&\times\left\{\sum_{j=2k}^{\infty}\frac{e^{j\tau^{(m)}}}{h_j}\beta_{j\,2k}H_j\left(\frac{x}{c_m}\right)\times\sum_{\ell=2k+1}^{\infty}\frac{e^{\ell\tau^{(n)}}}{h_\ell}\beta_{\ell\,2k+1}H_\ell\left(\frac{y}{c_n}\right)\right.\\
&\left.-\sum_{j=2k+1}^{\infty}\frac{e^{j\tau^{(m)}}}{h_j}\beta_{j\,2k+1}H_j\left(\frac{x}{c_m}\right)\times\sum_{\ell=2k}^{\infty}\frac{e^{\ell\tau^{(n)}}}{h_\ell}\beta_{\ell\,2k}H_\ell\left(\frac{y}{c_n}\right)\right\}.
\end{aligned}
$$

By (4.4) and (4.5) the right hand side of the above equation equals to

$$
\begin{aligned}
&\frac{e^{-\frac{1}{2}(1+\gamma_m)(x/c_m)^2}e^{-\frac{1}{2}(1+\gamma_n)(y/c_n)^2}}{\sqrt{(2T-t_m)(2T-t_n)}}\sum_{\mu=0}^{\infty}\sum_{\nu=0}^{\infty}\langle R^*_\mu,R^*_\nu\rangle_*\\
&\times\sum_{j=\mu}^{\infty}\sum_{\ell=\nu}^{\infty}\frac{e^{j\tau^{(m)}}e^{\ell\tau^{(n)}}}{h_jh_\ell}\beta_{j\mu}H_j\left(\frac{x}{c_m}\right)\beta_{\ell\nu}H_\ell\left(\frac{y}{c_n}\right)\\
=\ &F^{m,n}(x,y),
\end{aligned}
$$

where we have used (4.3). From the definitions (3.7) and (3.10), (4.9) is derived from (4.8).

§5. Proof of Theorems

The following formulae are known for (1.1) [2, 29]. For $u\in\mathbf{R}$,

$$\lim_{\ell\to\infty}(-1)^\ell\ell^{1/4}\varphi_{2\ell}\left(\frac{u}{2\sqrt{\ell}}\right)=\frac{1}{\sqrt{\pi}}\cos u, \tag{5.1}$$

$$\lim_{\ell\to\infty}(-1)^\ell\ell^{1/4}\varphi_{2\ell+1}\left(\frac{u}{2\sqrt{\ell}}\right)=\frac{1}{\sqrt{\pi}}\sin u, \tag{5.2}$$

$$\lim_{\ell\to\infty}2^{-\frac{1}{4}}\ell^{\frac{1}{12}}\varphi_\ell\left(\sqrt{2\ell}-\frac{u}{\sqrt{2}\,\ell^{1/6}}\right)=\mathrm{Ai}(-u) \tag{5.3}$$

Here we give the proof of Theorem 2 by using (5.3). The proof of Theorem 1 will be easier and given by the similar argument using (5.1) and (5.2).

Let $b^m(x) = \sqrt{2T - t_m}\exp\left\{1/2\gamma_m(x/c_m)^2 - N\tau^{(m)}\right\}$ and $\zeta^m(x)$ be the quaternion with

$$C(\zeta^m(x)) = \begin{pmatrix} b^m(x) & 0 \\ 0 & 1/b^m(x) \end{pmatrix}.$$

For $\mathbf{x}_N^{(m)} \in \mathbf{R}_<^N, 1 \le m \le M+1$, we consider the transformation of the quaternions $q^{m,n}\left(x_i^{(m)}, x_j^{(n)}\right) \mapsto \widehat{q}^{m,n}\left(x_i^{(m)}, x_j^{(n)}\right)$ defined by

$$\widehat{q}^{m,n}(x,y) = \zeta^m(x) q^{m,n}(x,y) \zeta^n(y)^{-1}.$$

We denote by $\widehat{Q}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_{M+1}}^{(M+1)}\right)$ the self-dual $\sum_{m=1}^{M+1} N_m \times \sum_{m=1}^{M+1} N_m$ quaternion matrix whose elements are $\widehat{q}^{m,n}\left(x_i^{(m)}, x_j^{(n)}\right)$, $1 \le i \le N_m$, $1 \le j \le N_n$, $1 \le m,n \le M+1$. By the definition of quaternion determinants, the following invariance is established:

$$\mathrm{Tdet}\, Q\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(2)}, \dots, \mathbf{x}_{N_{N_{M+1}}}^{(M+1)}\right) = \mathrm{Tdet}\, \widehat{Q}\left(\mathbf{x}_{N_1}^{(1)}, \mathbf{x}_{N_2}^{(m)}, \dots, \mathbf{x}_{N_{N_{M+1}}}^{(M+1)}\right).$$

Hence to prove Theorem 2 it is enough to show the following lemma.

Lemma 5. *Let* $T_N = 2N^{1/3}$ *and* $t_m = T_N + s_m$, $1 \le m,n \le M+1$. *Then for any* $x, y \in \mathbf{R}$,

$$(5.4)\ \lim_{N\to\infty} \frac{1}{b^m(x)b^n(y)} D^{m,n}(a_N(s_m)+x, a_N(s_n)+y) = \mathcal{D}(s_m, x; s_n, y),$$

$$(5.5)\ \lim_{N\to\infty} b^m(x)b^n(y)\tilde{I}^{m,n}(a_N(s_m)+x, a_N(s_n)+y) = \widetilde{\mathcal{I}}(s_m, x; s_n, y),$$

$$(5.6)\ \lim_{N\to\infty} \frac{b^m(x)}{b^n(y)} \tilde{S}^{m,n}(a_N(s_m)+x, a_N(s_n)+y) = \widetilde{\mathcal{S}}(s_m, x; s_n, y).$$

We start to prove this lemma by showing

$$(5.7)\quad \lim_{N\to\infty} \frac{b^m(x)}{b^n(y)} p_{t_n - t_m}(a_N(s_m)+x, a_N(s_n)+y) = \mathcal{P}(s_m, x; s_n, y).$$

By (4.1) and the fact

$$\frac{a_N(s_m)+x}{c_m} = \sqrt{2N} + \frac{x}{\sqrt{2}N^{1/6}} + \mathcal{O}(T_N^{-1}),$$

$$\tau^{(n)} = \frac{s_n}{T_N} + \mathcal{O}(T_N^{-2}),$$

for large N, we have

$$\begin{aligned}
&\lim_{N\to\infty} \frac{b^m(x)}{b^n(y)} p_{t_n-t_m}(a_N(s_m)+x, a_N(s_n)+y) \\
&= \lim_{N\to\infty} \sqrt{\frac{1}{T_N}} \sum_{p=-\infty}^{N} e^{\frac{p}{2N^{1/3}}(s_n-s_m)} \varphi_{N-p}\left(\sqrt{2N}+\frac{x}{\sqrt{2}N^{1/6}}\right) \\
&\qquad\qquad \times \varphi_{N-p}\left(\sqrt{2N}+\frac{y}{\sqrt{2}N^{1/6}}\right) \\
&= \lim_{N\to\infty} \frac{1}{N^{1/3}} \sum_{p=-\infty}^{N} e^{\frac{p}{2N^{1/3}}(s_n-s_m)} \mathrm{Ai}\left(x+\frac{p}{N^{1/3}}\right) \mathrm{Ai}\left(y+\frac{p}{N^{1/3}}\right),
\end{aligned}$$

where we have used (5.3). Then we have (5.7).

From (3.8), (4.2), (4.6) and (4.7), we have

$$S^{m,n}(x,y) = S_1^{m,n}(x,y) + S_2^{m,n}(x,y),$$

with

$$\begin{aligned}
S_1^{m,n}(x,y) &= \frac{b^n(y)}{c_n b^m(x)} \sum_{\ell=0}^{N-1} e^{(N-\ell)(\tau^{(n)}-\tau^{(m)})} \varphi_\ell\left(\frac{x}{c_m}\right) \varphi_\ell\left(\frac{y}{c_n}\right), \\
S_2^{(mn)}(x,y) &= \frac{b^n(y)}{c_n b^m(x)} \varphi_{N-1}(y/c_n) \\
&\quad \times \sum_{k=N/2}^{\infty} \frac{B(N/2+k)}{B(N/2-1)} e^{-(N-2k-1)\tau^{(m)}} \varphi_{2k+1}\left(\frac{x}{c_m}\right),
\end{aligned}$$

where $B(k) = \frac{2^k k!}{\sqrt{(2k+1)!}}$. Since

$$\frac{B(k)}{B(\ell)} = \left(\frac{k}{\ell}\right)^{1/4} \left(1+\mathcal{O}\left(\frac{|k-\ell|}{k+\ell}\right)\right),$$

by the same argument to show (5.7) we have

$$\lim_{N\to\infty} \frac{b^m(x)}{b^n(y)} S^{m,n}(a_N(s_m)+x, a_N(s_n)+y) = \mathcal{S}(s_m,x;s_n,y). \tag{5.8}$$

(5.6) is derived from (5.7) and (5.8).

From (4.6) and (4.7), by calculations with (4.10), we have

$$\begin{aligned}&b^m(x)b^n(y)\Phi^{(m)}_{N+2p}(x)\Phi^{(n)}_{N+2p+1}(y)\\&= -2^{3/2}r_{N/2+p}\frac{T}{\sqrt{N+2p+1}}e^{(2p)\tau^{(n)}}\varphi_{N+2p}\left(\frac{y}{c_n}\right)\\&\quad\times\sum_{k=p}^{\infty}\frac{B(N/2+k)}{B(N/2+p)}e^{(2k+1)\tau^{(m)}}\varphi_{N+2k+1}\left(\frac{x}{c_m}\right).\end{aligned}$$

From (4.9) we obtain (5.5) by the same procedure as above.

From (4.2), by calculations with (3.3) and

$$e^{-y^2/2}H_{\ell+1}(y) = -2\frac{d}{dy}\left(e^{-y^2/2}H_\ell(y)\right) + 2\ell e^{-y^2/2}H_{\ell-1}(y),$$

we have

$$\begin{aligned}&\frac{R^{(m)}_{2k}(x)R^{(n)}_{2k+1}(y)}{r_k b^m(x)b^n(y)}\\&= \frac{-1}{2\sqrt{t_m t_n(2T-t_m)(2T-t_n)}}e^{(N-2k)\tau^{(m)}}\varphi_{2k}\left(\frac{x}{c_m}\right)e^{(N-2k+1)\tau^{(n)}}\\&\quad\times\left\{\varphi'_{2k}\left(\frac{y}{c_n}\right)+\sqrt{2k}\left(1-e^{-2\tau^{(n)}}\right)\varphi_{2k-1}\left(\frac{y}{c_n}\right)\right\}.\end{aligned}$$

Using the fact that

$$\frac{(N-p)^{1/12}}{2^{3/4}N^{1/6}}\varphi'_{N-p}\left(\frac{a_N(s_n)+y}{c_n}\right) = \frac{d}{d\lambda}\mathrm{Ai}(y+\lambda)\Big|_{\lambda=p/N^{1/3}} + o(1),$$

we obtain (5.4). This completes the proof of Lemma 5.

References

[1] Abramowitz, M. and Stegun, I. A. : *Handbook of Mathematical Functions*, (1965), Dover.

[2] Bateman, H. : *Higher Transcendental Functions*, (A. Erdélyi Ed.), Vol. 2, (1953), McGraw Hill.

[3] Doob, J. L. : *Classical Potential Theory and its Probabilistic Counterpart*, (1984), Springer.

[4] Dyson, F. J.: A Brownian-motion model for the eigenvalues of a random matrix, *J. Math. Phys.* **3** (1962), 1191-1198.

[5] Dyson, F. J.: Correlation between the eigenvalues of a random matrix, *Commun. Math. Phys.* **19** (1970), 235-250.

[6] Forrester, P. J., Nagao, T. and Honner G.: Correlations for the orthogonal-unitary and symplectic-unitary transitions at the hard and soft edges, *Nucl. Phys.* **B553** (1999), 601-643.

[7] Fulton, W. and Harris, J. : *Representation Theory*, (1991), Springer.

[8] Grabiner, D. J. : Brownian motion in a Weyl chamber, non-colliding particles, and random matrices, *Ann. Inst. Henri Poincaré* **35** (1999), 177-204.

[9] Harish-Chandra, Differential operators on a semisimple Lie algebra, *Am. J. Math.* **79** (1957), 87-120.

[10] Itzykson, C. and Zuber, J.-B.: The planar approximation. II, *J. Math. Phys.* **21** (1980), 411-421.

[11] Johansson, K.: Discrete polynuclear growth and determinantal processes, **math.PR/0206208**.

[12] Karlin, S. and McGregor, L. : Coincidence properties of birth and death processes, *Pacific J.* **9** (1959), 1109-1140.

[13] Karlin, S. and McGregor, L. : Coincidence probabilities, *Pacific J.* **9** (1959), 1141-1164.

[14] Katori, M. and Tanemura, H. : Scaling limit of vicious walkers and two-matrix model, *Phys. Rev.* E **66** (2002), 011105.

[15] Katori, M. and Tanemura, H. : Functional central limit theorems for vicious walkers, **math.PR/0203286**.

[16] Mehta, M. L. : A method of integration over matrix variables, *Commun. Math. Phys.* **79** (1981), 327-340.

[17] Mehta, M. L. : *Matrix Theory*, Editions de Physique, (1989), Orsay.

[18] Mehta, M. L. : *Random Matrices*, second edition, (1991), Academic Press.

[19] Mehta, M.L. and Pandey, A. : On some Gaussian ensemble of Hermitian matrices, *J. Phys. A: Math. Gen.* **16** (1983), 2655-2684.

[20] Nagao, T. : Correlation functions for multi-matrix models and quaternion determinants, *Nucl. Phys.* **B602** (2001), 622-637.

[21] Nagao, T. and Forrester, P. J. : Multilevel dynamical correlation function for Dyson's Brownian motion model of random matrices, *Phys. Lett.* **A247** (1998), 42-46.

[22] Nagao, T. and Forrester, P. J. : Quaternion determinant expressions for multilevel dynamical correlation functions of parametric random matrices, *Nucl. Phys.* **B563[PM]** (1999), 547-572.

[23] Nagao, T., Katori, M. and Tanemura, H. : Dynamical correlations among vicious random walkers, *Phys. Lett.* **A307** (2003), 29-35.

[24] Osada, H. : Dirichlet form approach to infinite-dimensional Wiener processes with singular interactions, *Commun. Math. Phys.* **176** (1996), 117-131.

[25] Pandey, A. and Mehta, M.L. : Gaussian ensembles of random Hermitian intermediate between orthogonal and unitary ones, *Commun. Math. Phys.* **87** (1983), 449-468.

[26] Prähofer, M. and Spohn H.: Scale invariance of the PNG droplet and the Airy process, *J. Stat. Phys.* **108** (2002), 1071-1106.

[27] Soshnikov, A.: Determinantal random point fields, *Russian Math. Surveys* **55** (2000), 923-975.
[28] Spohn H.: Interacting Brownian particles: a study of Dyson's model, in *Hydrodynamic Behavior and Interacting Particle Systems* (G. Papanicolaou Ed.), IMA Volumes in Mathematics and its Applications **9** (1987), Springer.
[29] Szegö, G.: *Orthogonal Polynomials*, 4th edition (1975), American Mathematical Society.

Makoto Katori
Department of Physics
Chuo University
Kasuga, Bunkyo-ku
Tokyo 112-8551
Japan
E-mail address: `katori@phys.chuo-u.ac.jp`

Taro Nagao
Department of Physics
Graduate School of Science
Osaka University
Toyonaka
Osaka 560-0043
Japan
E-mail address: `nagao@sphinx.phys.sci.osaka-u.ac.jp`

Hideki Tanemura
Department Mathematics and Informatics
Chiba University
1-33 Yayoi-cho, Inage-ku
Chiba 263-8522
Japan
E-mail address: `tanemura@math.s.chiba-u.ac.jp`

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 307–324

Diffusive Behaviour of the Equilibrium Fluctuations in the Asymmetric Exclusion Processes

Claudio Landim, Stefano Olla and Srinivasa R. S. Varadhan

Abstract.

We consider the asymmetric simple exclusion process in dimension $d \geq 3$. We review some results concerning the equilibrium bulk fluctuations and the asymptotic behaviour of a second class particle.

§1. Introduction

The asymmetric simple exclusion process is the simplest model of a driven lattice gas. This model is given by the dynamics of infinitely many particles moving on $\mathbb{Z}^d$ as asymmetric random walks with an exclusion rule: when a particle attempts to jump on a site occupied by another particle the jump is suppressed. Of course we consider initial configurations where there is at most one particle per site. We denote the configurations by $\eta \in \{0,1\}^{\mathbb{Z}^d}$ so that $\eta(x) = 1$ if site x is occupied for the configuration η and $\eta(x) = 0$ if site x is empty. The number of particles is conserved and the Bernoulli product measures $\{\nu_\alpha,\ \alpha \in [0,1]\}$ are the ergodic invariant measures.

Rezakhanlou, in [11], proved that the empirical field of particles, after a hyperbolic rescaling of space and time by a parameter ϵ

$$\pi_t^\epsilon = \epsilon^d \sum_x \eta_{t\epsilon^{-1}}(x) \delta_{\epsilon x} \tag{1.1}$$

converges weakly, as $\epsilon \to 0$, to the (entropic) solution of the Burgers equation

$$\begin{cases} \partial_t \rho \; + \; \gamma \cdot \nabla[(1-\rho)\rho] \; = \; 0 \; , \\ \rho(0,\cdot) = \rho_0(\cdot) \; . \end{cases}$$

Received August 16, 2003.
Revised May 5, 2003.

Here ρ_0 is the initial asymptotic profile, i.e.

$$\pi_0^\epsilon \underset{\epsilon \to 0}{\rightharpoonup} \rho_0 \quad \text{in probability,} \tag{1.2}$$

while γ is the average velocity of a single particle. In fact in [11] the initial distribution of the particles is assumed to be an inhomogeneous product measure with a slowly varing profile $\nu_{\rho_0(\epsilon \cdot)}$, such that $\nu_{\rho_0(\epsilon \cdot)}(\eta(x)) = \rho_0(\epsilon x)$.

Suppose now that we start the system with a stationary measure ν_α, for a certain density $\alpha \in]0,1[$. The invariant measure ν_α has macroscopic Gaussian uncorrelated fluctuations, i.e. the fluctuation field

$$\xi_0^\epsilon = \epsilon^{d/2} \sum_{x \in \mathbb{Z}^d} \delta_{\epsilon x}(\eta_0(x) - \alpha) \tag{1.3}$$

converges in law to a white noise field ξ with covariance

$$\mathbb{E}(\xi(q)\xi(q')) = \alpha(1-\alpha)\delta(q-q') \tag{1.4}$$

It is not hard to prove in this context that the macroscopic evolution of these fluctuations, at time $t\epsilon^{-1}$, will converge to the solution of the linearized equation

$$\partial_t \xi + (1-2\alpha)\gamma \cdot \nabla \xi = 0 \tag{1.5}$$

Of course this equation (and the following ones) is to be intended in the weak sense, since ξ is only a distribution-valued field on $\mathbb{R}^d$. This means that an initial fluctuation will evolve macroscopically by a deterministic translation with velocity $(1-2\alpha)\gamma$. In simple exclusion processes there is a simple way to keep track of density fluctuations. Let us condition the stationary measure ν_α to have the site 0 empty and put in this site a *second class particle*, i.e. a particle that has the same jump rates as the other particles but when a normal (*first class*) particle attempts to jump in the site occupied by a second class particle, the particles exchange sites. Then this second class particles evolves like a density fluctuation (cf. section 6) and equation (1.5) corresponds to a law of large numbers for the position X_t of the second class particle

$$\frac{X_t}{t} \underset{t \to \infty}{\longrightarrow} (1-2\alpha)\gamma \ . \tag{1.6}$$

The natural question is about the fluctuation around this law of large numbers. The answer is that, in dimension $d \geq 3$, the recentered position of the second class particle $X_t - (1-2\alpha)\gamma t$ behaves diffusively and

$\epsilon(X_{t\epsilon^{-2}} - (1-2\alpha)\gamma t\epsilon^{-2})$ converges in law to a Brownian motion with a diffusion matrix $D(\alpha)$ (cf. Theorem 6.2).

In dimension $d = 1$ and 2 the asymptotic behavior of the second class particle is superdiffusive (cf. [9, 14]), and one of the most interesting and challenging problem is to determine the corresponding limit process in these cases.

We review in this article results concerning the dimension $d \geq 3$. The diffusive behaviour of the density fluctuations is then stated in the following way. Consider the recentered density fluctuation field at diffusive scaling

$$Y_t^\epsilon = \epsilon^{d/2} \sum_{x \in \mathbb{Z}^d} \delta_{\epsilon x - vt\epsilon^{-1}} (\eta_{t\epsilon^{-2}}(x) - \alpha) \tag{1.7}$$

where $v = (1-2\alpha)\gamma$. We show in section 4 that, as a process with values on the distributions on $\mathbb{R}^d$, it converges (in law) to the solution of the linear stochastic partial differential equation

$$\partial_t Y = \nabla \cdot D(\alpha) \nabla Y + \sqrt{2\alpha(1-\alpha)}\sqrt{D(\alpha)}\nabla \cdot W \tag{1.8}$$

where $W(x,t) = (W_1, \ldots, W_d)$ are independent standard white noises on $\mathbb{R}^{d+1}$.

This result was proved in [1] for the asymmetric exclusion process in $d \geq 3$ and in finite macroscopic volume. The purpose of this article is to give a simpler proof in infinite volume based on the fluctuation-dissipation theorem as stated in [8]. The fluctuation-dissipation theorem, which was first proved in this context in [7], is in fact the the core of the proof for the macroscopic density fluctuation, as we explain in section 4. This theorem states that the *space-time* fluctuations of the current associated to the density are equivalent to to the fluctuations of a gradient of the density times the matrix $D(\alpha)$, in the sense that the variance of the difference is asymptotically small (cf. Theorem 3.2 in section 3).

This fluctuation-dissipation theorem was also applied in order to study the diffusive incompressible limit (cf. [3]) and the first order corrections to the hydrodynamic limit (cf. [5]). It has also been used to study corresponding results for a lattice system of particles with exclusion rule with conservation of number of particles, velocity and energy (cf. [2] and references therein).

The fluctuation-dissipation theorem for the exclusion processes can be proved by using the duality properties of these models (as explained in [8]), that permits to control the size of these fluctuation by estimates on the Green functions of simple random walks (see also [12] where this method is applied in the study of the diffusive behaviour of a tagged particle). At the moment there are not any result of this type for systems that do not have such nice duality.

§2. Density fluctuations

Fix a probability distribution $p(\cdot)$ supported on a finite subset of $\mathbb{Z}^d_* = \mathbb{Z}^d \backslash \{0\}$ and denote by L the generator of the simple exclusion process on $\mathbb{Z}^d$ associated to $p(\cdot)$. L acts on local functions f on $\mathcal{X}_d = \{0,1\}^{\mathbb{Z}^d}$ as

$$(2.1) \qquad (Lf)(\eta) \;=\; \sum_{x,y\in\mathbb{Z}^d} p(y-x)\eta(x)\{1-\eta(y)\}[f(\sigma^{x,y}\eta)-f(\eta)]\;,$$

where $\sigma^{x,y}\eta$ stands for the configuration obtained from η by exchanging the occupation variables $\eta(x)$, $\eta(y)$:

$$(\sigma^{x,y}\eta)(z) \;=\; \begin{cases} \eta(z) & \text{if } z \neq x,\, y \;, \\ \eta(x) & \text{if } z = y \;, \\ \eta(y) & \text{if } z = x \;. \end{cases}$$

Denote by $s(\cdot)$ and $a(\cdot)$ the symmetric and the anti-symmetric parts of the probability $p(\cdot)$:

$$s(x) = (1/2)[p(x)+p(-x)]\;, \quad a(x) = (1/2)[p(x)-p(-x)]$$

and denote by L^s, L^a the symmetric part and the anti-symmetric part of the generator L. L^s, L^a are obtained by replacing p by s, a in the definition of L.

For α in $[0,1]$, denote by ν_α the Bernoulli product measure on $\mathcal{X}$ with $\nu_\alpha[\eta(x)=1]=\alpha$. Measures in this one-parameter family are stationary and ergodic for the simple exclusion dynamics and in the symmetric case, i.e. $p(x)=p(-x)$, these measures are reversible. Expectation with respect to ν_α is represented by $<\cdot>_\alpha$ and the scalar product in $L^2(\nu_\alpha)$ by $<\cdot,\cdot>_\alpha$.

Fix $\varepsilon>0$. For a configuration η, denote by $\pi^\varepsilon = \pi^\varepsilon(\eta)$ the empirical measure associated to η. This is the measure on $\mathbb{R}^d$ obtained assigning mass ε^d to each particle of η :

$$\pi^\varepsilon \;=\; \varepsilon^d \sum_{x\in\mathbb{Z}^d} \eta(x)\delta_{x\varepsilon}\;,$$

where δ_u stands for the Dirac measure concentrated on u. For a measure π on $\mathbb{R}^d$ and a continuous function $G\colon \mathbb{R}^d \to \mathbb{R}$, denote by $<\pi, G>$ the integral of G with respect to π. In particular, $<\pi^\varepsilon, G>$ is equal to $\varepsilon^d \sum_{x\in\mathbb{Z}^d} G(x\varepsilon)\eta(x)$.

Consider a sequence of probability measures μ^ε on the configuration space $\mathcal{X}_d$ and assume that, under μ^ε, π^ε converges in probability to an absolutely continuous measure $\rho_0(u)du$. This means that for every continuous function $G\colon \mathbb{R}^d \to \mathbb{R}$ and every $\delta > 0$,

$$\lim_{\varepsilon\to 0} \mu^\varepsilon \Big\{ \Big| <\pi^\varepsilon, G> - \int G(u)\rho_0(u)\, du \Big| > \delta \Big\} = 0 . \tag{2.2}$$

Consider the hyperbolic equation

$$\left\{ \begin{array}{l} \partial_t \rho + \gamma \cdot \nabla [\rho(1-\rho)] = 0 , \\ \rho(0,\cdot) = \rho_0(\cdot) . \end{array} \right. \tag{2.3}$$

In this formula γ stands for the drift : $\gamma = \sum_y y p(y)$. For $t > 0$, denote by π_t^ε the empirical measure associated to the state of the process at time t : $\pi_t^\varepsilon = \pi^\varepsilon(\eta_t)$. Rezakhanlou [11] proved that, starting from a inhomogeneous product measure $\nu_{\rho(\varepsilon\cdot)}$ satisfying (2.2), then, for every $t \ge 0$, $\pi_{t\varepsilon^{-1}}^\varepsilon$ converges in probability to the measure $\rho(t,u)du$, where the density ρ is the entropy solution of the hyperbolic equation (2.3). More precisely, for a measure μ on $\mathcal{X}_d$, denote by $\mathbb{P}_\mu$ the measure on the path space $D(\mathbb{R}_+, \mathcal{X}_d)$ induced by the Markov process η_t and the measure μ. Then, for every $t \ge 0$, every continuous function G and every $\delta > 0$,

$$\lim_{\varepsilon\to 0} \mathbb{P}_{\nu_{\rho(\varepsilon\cdot)}} \Big\{ \Big| <\pi_{t\varepsilon^{-1}}^\varepsilon, G> - \int G(u)\rho(t,u)\, du \Big| > \delta \Big\} = 0 ,$$

where the density ρ is the entropy solution of the hyperbolic equation (2.3). Notice the Euler rescaling of time in the previous formula.

Let us now consider, for a fixed density $\alpha \in (0,1)$, the system in equilibrium with distribution $\mathbb{P}_{\nu_\alpha}$ on the path space. Let $Y_\cdot^\varepsilon$ be the density fluctuation field that acts on smooth functions H as

$$Y_t^\varepsilon(H) = \varepsilon^{d/2} \sum_{x\in\mathbb{Z}^d} H(x\varepsilon - vt\varepsilon^{-1})(\eta_{t\varepsilon^{-2}}(x) - \alpha) , \tag{2.4}$$

where $v = \gamma(1-2\alpha)$. Notice the diffusive rescaling of time on the right hand side of this identity.

For any $k \ge 0$ and $f, g \in \mathcal{C}^\infty(\mathbb{R}^d)$ consider the scalar product

$$(g,f)_k = \int_{\mathbb{R}^d} g(q) \left(|q|^2 - \Delta\right)^k f(q)\, dq \tag{2.5}$$

and denote by $\mathcal{H}_k$ the corresponding closure. For any positive k we denote by $\mathcal{H}_{-k}$ its dual space with respect to the $L^2(\mathbb{R}^d) \equiv \mathcal{H}_0$ scalar product.

We will show in section 5 that the probability distribution Q^ϵ of $Y^\varepsilon_\cdot$ under $\mathbb{P}_{\nu\alpha}$, is supported and is tight in $D([0,T],\mathcal{H}_{-k})$ for any $k > d+1$. Observe that in [1] the tightness is proved for any $k > d/2+1$. This difference is due to the fact that here we are dealing with distribution on the infinite volume $\mathbb{R}^d$, while in [1] we were working in the finite d-dimensional torus.

We state now the main theorem. Assume that $d \geq 3$. Let $D_{i,j}(\alpha)$ a strictly positive symmetric matrix. Denote by $\mathcal{A}$, $\mathcal{B}$ the differential operators defined by $\mathcal{A} = \sum_{1\le i,j\le d} D_{i,j}(\alpha)\partial^2_{u_i,u_j}$, $\mathcal{B} = \sqrt{2\alpha(1-\alpha)}\,\sigma\,\nabla$ (where the matrix σ is the positive square root of D). Denote by $\{T_t,\, t \geq 0\}$ the semigroup in $L^2(\mathbb{R}^d)$ associated to the operator $\mathcal{A}$. Fix a positive integer $k_0 > d+1$. Let Q be the probability measure concentrated on $C([0,T],\mathcal{H}_{-k_0})$ corresponding to the stationary generalized Ornstein–Uhlenbeck process with mean 0 and covariance

$$E_Q\Big[Y_t(H)Y_s(G)\Big] \;=\; \chi(\alpha)\int_{\mathbb{R}^d} du\,(T_{|t-s|}H)(u)\,G(u)$$

for every $0 \leq s \leq t$ and H, G in $\mathcal{H}_{k_0}$. Here $\chi(\alpha)$ stands for the static compressibility given by $\chi(\alpha) = \mathbf{Var}_{\nu_\alpha}[\eta(0)] = \alpha(1-\alpha)$.

Theorem 2.1. *There exists a strictly positive symmetric matrix $D_{i,j}(\alpha)$ such that the sequence Q^ε converges weakly to the probability measure Q of the corresponding Ornstein–Uhlenbeck process.*

Formally, Y_t is the solution of the stochastic differential equation

$$dY_t \;=\; \mathcal{A}Y_t dt \;+\; dB^\nabla_t\;,$$

where B^∇_t is a mean zero Gaussian field with covariances given by

$$E_Q\Big[B^\nabla_t(H)B^\nabla_s(G)\Big] \;=\; 2\chi(\alpha)(s\wedge t)\int_{\mathbb{R}^d}(\nabla H)\cdot D\,(\nabla G)\,.$$

In section 6 we show that the viscosity matrix $D(\alpha)$ is identified as

(2.6)
$$D_{i,j}(\alpha) = \frac{1}{\alpha(1-\alpha)}\lim_{t\to\infty}\frac{1}{t}\sum_{x\in\mathbb{Z}^d} x_i x_j \mathbb{E}_{\nu_\alpha}\left((\eta_t(x-vt)-\alpha)(\eta_0(0)-\alpha)\right),$$

which is also the asymptotic variance of the second class particle. Other expressions for this matrix can be given (cf. [6]), while in [8] we prove that $D_{i,j}(\alpha)$ are smooth functions of the density $\alpha \in [0,1]$.

§3. The Fluctuation-Dissipation Theorem

We recall in this section some results proved in [8] (cf. also [7]). For local functions u, v, define the scalar product $\ll \cdot\,, \cdot \gg$ by

$$\ll u,v \gg = \sum_{x\in\mathbb{Z}^d} \{ < \tau_x u, v > - < u >< v > \} , \tag{3.1}$$

where $\{\tau_x,\ x \in \mathbb{Z}^d\}$ is the group of translations and $< \cdot >$ stands for the expectation with respect to the measure ν_α. That this is in fact an inner product can be seen by the relation

$$\ll u,v \gg = \lim_{V\uparrow\mathbb{Z}^d} \frac{1}{|V|} < \sum_{x\in V} \tau_x(u- < u >), \sum_{x\in V} \tau_x(v- < v >) > \ .$$

Since $\ll u - \tau_x u\,, v \gg = 0$ for all x in $\mathbb{Z}^d$, this scalar product is only positive semidefinite. Denote by $L^2_{\ll\cdot,\cdot\gg}(\nu_\alpha)$ the Hilbert space generated by the local functions and the inner product $\ll \cdot, \cdot \gg$.

Denote by L^s the symmetric part of the generator. For two local functions u, v, let

$$\ll u,v \gg_1 = \ll u, (-L^s)v \gg$$

and let H_1 be the Hilbert space generated by local functions and the inner product $\ll \cdot, \cdot \gg_1$. To introduce the dual Hilbert spaces of H_1, for a local function u, consider the semi-norm $\| \cdot \|_{-1}$ given by

$$\|u\|_{-1} = \sup_v \Big\{ 2 \ll u,v \gg - \ll v,v \gg_1 \Big\} ,$$

where the supremum is carried over all local functions v. Denote by H_{-1} the Hilbert space generated by the local functions and the semi-norm $\| \cdot \|_{-1}$.

Notice that the function $\eta(0) - \alpha$ does not belong to H_{-1}. Indeed, due to the translations, $\|\eta(0) - \alpha\|_1 = 0$ so that $\|\eta(0) - \alpha\|_{-1} = \infty$. One can show, however, that these linear functions are essentially the only zero-mean local functions that do not belong to H_{-1}.

To make the previous statement precise, we need to introduce some notation. For a local function f, denote by $\tilde f : [0,1] \to \mathbb{R}$ the polynomial defined by

$$\tilde f(\alpha) = E_{\nu_\alpha}[f] \ .$$

Denote by $\mathcal{C}_0 = \mathcal{C}_0(\alpha)$ the collection of local functions such that

$$\tilde f(\alpha) = E_{\nu_\alpha}[f] = 0\ , \quad \tilde f'(\alpha) = \frac{d}{d\beta} E_{\nu_\beta}[f] \Big|_{\beta=\alpha} = 0\ .$$

It is proved in [8] that all functions in $\mathcal{C}_0$ have finite H_{-1} norm:

$$\|f\|_{-1} < \infty \quad \text{if } f \in \mathcal{C}_0 \ . \tag{3.2}$$

Denote by $\mathcal{C}_1$ the space of local functions f in $\mathcal{C}_0$ which are orthogonal to all linear functions:

$$< f, \eta(x) - \alpha > = 0 \quad \text{for all } x \text{ in } \mathbb{Z}^d.$$

The next result states that a local function f in $\mathcal{C}_1$ has a finite space-time variance in the diffusive scaling.

Theorem 3.1. *Fix $T > 0$, a vector v_0 in $\mathbb{R}^d$, a smooth function $G : \mathbb{R}^d \to \mathbb{R}$ with compact support and a local function f in $\mathcal{C}_1$. There exists a finite constant C_0 such that*

$$\limsup_{\varepsilon\to 0} \mathbb{E}_{\nu_\alpha}\Big[\sup_{0\le t\le T}\Big(\varepsilon^{d/2+1}\int_0^{t\varepsilon^{-2}} \sum_{x\in\mathbb{Z}^d} G(\varepsilon[x - rv_0]) f(\tau_x\eta_r)\, dr\Big)^2\Big]$$
$$\le C_0 T \|G\|_{L^2}^2 \, \|f\|_{-1}^2 \ .$$

The theorem is not correct if we replace $\mathcal{C}_1$ by $\mathcal{C}_0$ because $\eta(e_1) - \eta(0)$ has H_{-1} norm equal to 0 and a finite, strictly positive space-time variance.

It is proved in [8] that any local function in $\mathcal{C}_0$ may be approximated in H_{-1} by a local function in the range of the generator. More precisely, for every local function f in $\mathcal{C}_0$ and every $\varepsilon > 0$, there exists a local function u_ε, which may be taken in $\mathcal{C}_1$, such that

$$\|Lu_\varepsilon - f\|_{-1}^2 \le \varepsilon \ . \tag{3.3}$$

The fluctuation-dissipation theorem stated below follows from this result, the estimate stated in Theorem 3.1 and some elementary computations.

Theorem 3.2. *Fix $T > 0$, a vector v_0 in $\mathbb{R}^d$, a local function w in $\mathcal{C}_1$ and a smooth function $G : \mathbb{R}^d \to \mathbb{R}$ with compact support. There exist a sequence of local functions u_m and $D_z(\alpha)$ such that*

$$\limsup_{m\to\infty}\limsup_{\varepsilon\to 0}$$
$$\mathbb{E}_{\nu_\alpha}\Big[\sup_{0\le t\le T}\Big(\varepsilon^{d/2+1}\int_0^{t\varepsilon^{-2}} \sum_{x\in\mathbb{Z}^d} G(\varepsilon[x - rv_0])\tau_x W_m(\eta_s)\, ds\Big)^2\Big] = 0 \ ,$$

where

$$W_m(\eta) = w - Lu_m + \sum_{z\in\mathbb{Z}^d} a(z) D_z(\alpha)\{\eta(z) - \eta(0)\} \ .$$

The idea of the proof is quite simple. By (3.3), we may approximate w by some local function Lu_m in the range of the generator. However, Lu_m might have linear terms and therefore might not be in $\mathcal{C}_1$. Subtracting expressions of type $\eta(z)-\eta(0)$, we convert Lu_m in a $\mathcal{C}_1$-function and apply Theorem 3.1.

§4. Time evolution of density fluctuations

We show in this section how to deduce from the fluctuation-dissipation theorem the equilibrium fluctuations for the density field defined in (2.4).

Fix a smooth function $G : \mathbb{R}^d \to \mathbb{R}$ with compact support and define the martingale $M_t^{1,\varepsilon}$ by the time evolution equation

$$Y_t^\varepsilon(G) - Y_0^\varepsilon(G) = \int_0^t \left(\partial_s + \varepsilon^{-2} L\right) Y_s^\varepsilon(G)\, ds + M_t^{1,\varepsilon} \, .$$

An elementary computation shows that

(4.1)
$$\begin{aligned}(\partial_t + \varepsilon^{-2} L) Y_t^\varepsilon(G) = -\varepsilon^{d/2-1} \sum_x (v \cdot \nabla G)[\varepsilon(x - vt\varepsilon^{-2})](\eta(x) - \alpha) \\ - \varepsilon^{d/2-2} \sum_x G[\varepsilon(x - vt\varepsilon^{-2})] \sum_y j_{x,y} \, ,\end{aligned}$$

where $j_{x,y}$, the instantaneous current between sites x and y, is given by

$$\begin{aligned} j_{x,y} \;&=\; -s(x-y)[\eta(y)-\eta(x)] \\ &\quad - a(x-y)\Big\{\eta(y)(1-\eta(x)) + \eta(x)(1-\eta(y))\Big\} \, .\end{aligned}$$

Let $G_t^\varepsilon(x) = G[\varepsilon(x - vt\varepsilon^{-2})]$. Since the current is anti-symmetric ($j_{x,y} = -j_{y,x}$) and since its expectation with respect to ν_α is equal to $2a(y-x)\alpha(1-\alpha)$, an elementary computation gives that

$$\begin{aligned} -\sum_x G_t^\varepsilon(x) \sum_y j_{x,y} = \sum_{x,y} s(x-y)[G_t^\varepsilon(y) - G_t^\varepsilon(x)][\eta(x) - \alpha] \\ - \sum_{x,y} a(x-y)[G_t^\varepsilon(x) - G_t^\varepsilon(y)][\eta(y) - \alpha][\eta(x) - \alpha] \\ + (1-2\alpha) \sum_{x,y} a(x-y)[G_t^\varepsilon(x) - G_t^\varepsilon(y)][\eta(x) - \alpha] \end{aligned}$$

because

$$\begin{aligned} &\eta(y)(1-\eta(x)) \;+\; \eta(x)(1-\eta(y)) \;-\; 2\alpha(1-\alpha) \\ &\quad=\; -2(\eta(y)-\alpha)(\eta(x)-\alpha) \;+\; (1-2\alpha)\Big\{(\eta(y)-\alpha) + (\eta(x)-\alpha)\Big\} \, .\end{aligned}$$

Finally, since $a(\cdot)$ is anti-symmetric, a Taylor expansion and a change of variables in the summation permit to conclude that

$$\begin{aligned}(4.2)\quad & (\partial_t + \varepsilon^{-2}L)Y_t^\varepsilon(G) \\ & = \varepsilon^{d/2}\sum_x \sum_{i,j} \sigma_{i,j}(\partial_i\partial_j G)[\varepsilon(x - vt\varepsilon^{-2})](\eta(x) - \alpha) \\ & - \varepsilon^{d/2-1}\sum_{x,z} a(z)(z \cdot \nabla G)[\varepsilon(x - vt\varepsilon^{-2})]\Phi_{x,x+z} \; + \; R_\varepsilon(\eta) \, .\end{aligned}$$

In this formula, $\sigma_{i,j}$ is the symmetric matrix defined by

$$\sigma_{i,j}^s \;=\; \sum_z s(z) z_i z_j \; ,$$

$\Phi_{x,y}$ is the zero-mean local function defined by

$$\Phi_{x,y} \;=\; (\eta(x) - \alpha)(\eta(y) - \alpha)$$

and $R_\varepsilon(\eta)$ is a remainder which vanishes in $L^2(\nu_\alpha)$ as $\varepsilon \downarrow 0$. In fact $< R_\varepsilon(\eta)^2 >= O(\varepsilon^2)$.

Since $\Phi_{x,y}$ belongs to $\mathcal{C}_1$, by Theorem 3.2, there exist a sequence of local functions $\{v_m, \, m \geq 1\}$ in $\mathcal{C}_1$ and constants $D_{z,z'}$ such that

$$\begin{aligned}\lim_{m\to\infty}\lim_{\varepsilon\to 0}\mathbb{E}_{\nu_\alpha}\Big[\sup_{0\leq t\leq T}\Big|\varepsilon^{d/2-1}\int_0^t \sum_{x,z} a(z)(z\cdot\nabla G)[\varepsilon(x - vt\varepsilon^{-2})] \\ \Big\{\Phi_{0,z} - Lv_m - \sum_{z'} a(z')D_{z,z'}[\eta(z') - \eta(0)]\Big\}(\tau_x\eta_{\varepsilon^{-2}s})\, ds\Big|^2\Big] \;=\; 0 \, .\end{aligned}$$

This result shows that we may replace $\Phi_{x,x+z}$ in the second term on the right hand side of (2.4) by $Lv_m - \sum_{z'} a(z')D_{z,z'}[\eta(z') - \eta(0)]$. The difference $\eta(z') - \eta(0)$ enables a second summation by parts which cancels a factor ε^{-1}, while the term Lv_m produces an extra martingale.

Let $F(x) = \sum_z a(z)(z \cdot \nabla G)(x)$. For each $m \geq 1$, consider the martingale $M_t^{2,m,\varepsilon}$ defined by

$$\begin{aligned}M_t^{2,m,\varepsilon} = & \varepsilon^{d/2+1}\sum_x \int_0^t F(\varepsilon(x - vs\varepsilon^{-2}))\varepsilon^{-2}Lv_m(\tau_x\eta_{\varepsilon^{-2}s})\, ds \\ & + \varepsilon^{d/2+1}\sum_x \int_0^t \partial_s F(\varepsilon(x - vs\varepsilon^{-2}))v_m(\tau_x\eta_{\varepsilon^{-2}s})\, ds \\ & - \varepsilon^{d/2+1}\sum_x \Big\{F(\varepsilon(x - vt\varepsilon^{-2}))v_m(\tau_x\eta_{\varepsilon^{-2}t}) - F(\varepsilon x)v_m(\tau_x\eta_0)\Big\}\end{aligned}$$

Since v_m are local functions, it is easy to see that the L^2 norm of the third term of the right hand side vanishes as $\varepsilon \to 0$. The second term is equal to

$$\varepsilon^{d/2} \sum_x \int_0^T (v \cdot \nabla F)(\varepsilon(x - vt\varepsilon^{-2})) v_m(\tau_x \eta_{\varepsilon^{-2}t})\, dt\ .$$

Since v_m belongs to $\mathcal{C}_1$ for each $m \geq 1$, it has finite H_{-1} norm in view of (3.2). In particular, by Theorem 3.1, the expectation of the square goes to 0 as $\varepsilon \to 0$.

In conclusion, we have shown so far that

$$\int_0^t \left(\partial_s + \varepsilon^{-2} L\right) Y_s^\varepsilon(G)\, ds = \int_0^t Y_s^\varepsilon(\mathcal{A}G)\, ds + M_t^{2,m,\varepsilon} + R_{m,\varepsilon}(t)\ ,$$

where

$$\lim_{m \to \infty} \lim_{\varepsilon \to 0} E_{\nu_\alpha}\Big(\sup_{0 \leq t \leq T} \Big| R_{m,\varepsilon}(t) \Big|^2 \Big) = 0$$

and the second order differential operator $\mathcal{A}$ is given by

$$\mathcal{A} = \sum_{i,j}^{d} D_{i,j} \partial_i \partial_j\ ,$$

with the matrix $D_{i,j}$ given by

$$D_{i,j} = \sigma_{i,j}\ -\ \sum_{z,z'} a(z) a(z') z_i z'_j D_{z,z'}\ .$$

We turn now to the calculation of the quadratic variation of the martingale $M_t^{1,\varepsilon} + M_t^{2,n,\varepsilon}$. This is given by

(4.3)
$$\int_0^t \sum_y \sum_z p(z) \eta(y) [1 - \eta(y+z)] \varepsilon^{-2} \left[H_s^\varepsilon(\eta_{s\varepsilon^{-2}}^{y,y+z}) - H_s^\varepsilon(\eta_{s\varepsilon^{-2}}) \right]^2 ds\ ,$$

where

$$H_s^\varepsilon(\eta) = Y_s^\varepsilon(G) - \varepsilon^{d/2+1} \sum_x \int_0^T F(\varepsilon(x - vt\varepsilon^{-2})) v_m(\tau_x \eta_{\varepsilon^{-2}t})\ .$$

Elementary computations show that (4.3) is equal to

(4.4)

$$\int_0^t \varepsilon^d \sum_y \sum_z p(z)\eta(y)[1-\eta(y+z)] \times \Bigg(\varepsilon^{-1}[G_s^\varepsilon(y+z) - G_s^\varepsilon(y)] \\ - \sum_x F_s^\varepsilon(x)\left(v_m(\tau_x\sigma^{y,y+z}\eta_{s\varepsilon^{-2}}) - v_m(\tau_x\eta_{s\varepsilon^{-2}})\right)\Bigg)^2 ds\ ,$$

where $F_s^\varepsilon(x) = F(\varepsilon(x - vt\varepsilon^{-2}))$.

Since v_m is a local function, the sum inside the square in the above expression extends over a finite number of x depending only on the support of v_m. We can therefore substitute $F_s^\varepsilon(x)$ by $F_s^\varepsilon(y)$, with an error that is small in view of Theorem 3.1. In the same way, we replace the discrete derivative of G by the actual derivative, obtaining that (4.3) is equal to

$$\int_0^t \varepsilon^d \sum_y \sum_z p(z)\eta(y)[1-\eta(y+z)] \times \Bigg[(z\cdot\nabla G)(\varepsilon(y - vt\varepsilon^{-2})) \\ - F_s^\varepsilon(y)\sum_x \left(v_m(\tau_x\sigma^{y,y+z}\eta_{s\varepsilon^{-2}}) - v_m(\tau_x\eta_{s\varepsilon^{-2}})\right)\Bigg]^2 ds$$

plus a remainder $R_\varepsilon(t)$ which vanishes in L^2 as $\varepsilon \downarrow 0$. Recall the definition of F and take the limit as $\varepsilon \to 0$. By the law of large numbers, we obtain that the previous expression converges to

$$t\int dy \sum_z p(z)(z\cdot\nabla G)(y)^2 \\ \times \left\langle \eta(0)(1-\eta(z))\Big(1 - a(z)\Big[\Gamma_{v_m}(\eta^{0,z}) - \Gamma_{v_m}(\eta)\Big]\Big)^2\right\rangle\ ,$$

where $\Gamma_{v_m}(\eta)$ denotes the formal sum $\sum_x v_m(\tau_x\eta)$.

Since we performed this calculations in equilibrium and since for the invariant product measure the static fluctuations converges to the Gaussian field with covariance operator $\alpha(1-\alpha)(-\Delta)^{-1}$, if

$$b_{i,j} = \lim_{m\to\infty} \sum_z p(z) \\ \times \left\langle \eta(0)(1-\eta(z))\Big(1 - a(z)\Big[\Gamma_{v_m}(\eta^{0,z}) - \Gamma_{v_m}(\eta)\Big]\Big)^2\right\rangle z_i z_j\ ,$$

the fluctuation-dissipation relation for the limit Ornstein-Uhlenbeck process gives

$$b_{i,j} = \alpha(1-\alpha)D_{i,j} \ .$$

§5. Tightness

Recall that we have defined, for any $k \in \mathbb{R}$, $\mathcal{H}_k$ as the closure of $C^\infty(\mathbb{R}^d)$ with respect to the scalar product

$$(g, f)_k = \int_{\mathbb{R}^d} g(q) \left(|q|^2 - \Delta\right)^k f(q) \, dq$$

It is convenient to represent the scalar product $(\cdot, \cdot)_k$ in the orthonormal basis of the Hermite polynomials, which are the eigenfunctions of $|q|^2 - \Delta$. Let $\vec{n}$ be a multi-index of $(\mathbb{Z}^+)^d$ and $|\vec{n}| = \sum_{i=1}^d n(i)$. We denote by $\lambda_{n(i)} = 2n(i) + 1$ for $n(i) \in \mathbb{Z}^+$ and $\lambda_{\vec{n}} = \sum_{i=1}^d \lambda_{n(i)}$. Define $h_{\vec{n}}(q) = \prod_{i=1}^d h_{n(i)}(q_i)$ where h_m is the m^{th} normalized Hermite polynomial of order m in $\mathbb{R}$. We have then for every $k \geq 0$ and $f \in L^2$

$$\|f\|_k^2 = \int_{\mathbb{R}^d} f(q)(|q|^2 - \Delta)^k f(q) \, dq = \sum_{\vec{n} \in (\mathbb{Z}^+)^d} \lambda_{\vec{n}}^k \left(\int_{\mathbb{R}^d} f(q) h_{\vec{n}}(q) \, dq \right)^2$$

This is valid also for negative k. So the $\mathcal{H}_{-k}$-norm of a distribution ξ on $\mathbb{R}^d$ can be written as

$$\|\xi\|_{-k}^2 = \sum_{\vec{n} \in (\mathbb{Z}^+)^d} \lambda_{\vec{n}}^{-k} \, \xi(h_{\vec{n}})^2 \tag{5.1}$$

Observe that, for $k' > k$, the injection J of $\mathcal{H}_{-k}$ in $\mathcal{H}_{-k'}$ is compact. In fact it can be approximated by the finite range operators $J_m \xi = \sum_{|\vec{n}| \leq m} \xi(h_{\vec{n}}) h_{\vec{n}}$, and it is easy to see that the operator norm of the difference is bounded by

$$\|J - J_m\| \leq (2m + d)^{-(k'-k)}$$

By the compactness of the injections $\mathcal{H}_{-k} \hookrightarrow \mathcal{H}_{-k'}$ for $k < k'$, and standard compactness arguments, the tightness of the distribution of Y_t^ϵ is a consequence of the following proposition.

Proposition 5.1. *For any $k > d + 1$ and every $T > 0$, we have that*

1.

$$\sup_{\varepsilon\in(0,1)} \mathbb{E}_\mu \left(\sup_{t\in[0,T]} \|Y_t^\varepsilon\|_{-k}^2 \right) < +\infty$$

2. For any $R > 0$,

$$\lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \mathbb{P}_\mu \left(\sup_{\substack{t,s\in[0,T]\\ |t-s|\le\delta}} \|Y_t^\varepsilon - Y_s^\varepsilon\|_{-k} > R \right) = 0$$

It is easy to see, by using (5.1) and that $k > d+1$, that Proposition 5.1 is a consequence of the following

Proposition 5.2. *For any smooth function G on $\mathbb{R}^d$ with compact support*

$$\sup_\varepsilon \mathbb{E}_\mu \left(\sup_{t\in[0,T]} Y_t^\varepsilon(G)^2 \right) \le TC_G \tag{5.2}$$

$$\lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \mathbb{P}_\mu \left(\sup_{\substack{t,s\in[0,T]\\ |t-s|\le\delta}} |Y_t^\varepsilon(G) - Y_s^\varepsilon(G)| > R \right) = 0 \tag{5.3}$$

Let just sketch the proof of (5.3), the proof of (5.2) will follow a similar argument (cf. [1] for details).

By the same calculation made in the previous section we have

$$\begin{aligned} Y_t^\varepsilon(G) - Y_s^\varepsilon(G) &= \int_s^t \left(\partial_\tau + \varepsilon^{-2} L\right) Y_\tau^\varepsilon(G)\, d\tau + M_{s,t}^{1,\varepsilon} \\ &= \int_s^t Y_\tau^\varepsilon(\mathcal{A}G)\, d\tau + M_{s,t}^{1,\varepsilon} + M_{s,t}^{2,m,\varepsilon} + R_{m,\varepsilon}(s,t)\ , \end{aligned} \tag{5.4}$$

where $M_{s,t}^{1,\varepsilon}$ and $M_{s,t}^{2,m,\varepsilon}$ are the differences of the corresponding martingales defined in the previous section, and

$$\lim_{m\to\infty} \lim_{\varepsilon\to 0} E_{\nu_\alpha} \left(\sup_{0\le s\le t\le T} \left| R_{m,\varepsilon}(s,t) \right|^2 \right) = 0$$

The first term is easy to deal since

$$\mathbb{E}_\mu \left(\sup_{\substack{t,s\in[0,T]\\ |t-s|\le\delta}} |\int_s^t Y_\tau^\varepsilon(\mathcal{A}G)\, d\tau|^2 \right) \le \delta T \left\langle Y^\varepsilon(\mathcal{A}G)^2 \right\rangle \le \delta T C \|\mathcal{A}G\|_2^2. \tag{5.5}$$

About the difference martingale $\widetilde{M}^m_{s,t} = M^{1,\varepsilon}_{s,t} + M^{2,m,\varepsilon}_{s,t}$, for any finite m has a bounded quadratic variation given by (4.3) or (4.4), and it is not difficult to show exponential bounds for it. So it follows that for any m

$$\lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \mathbb{P}_\mu \left(\sup_{\substack{t,s\in[0,T] \\ |t-s|\leq\delta}} |\widetilde{M}^m_{s,t}| > R \right) = 0.$$

§6. The second class particle

The second class particle is a special particle that has the same jump rates of the other particles (with the exclusion rule), furthermore when a normal (first class) particle attempts to jump in the site occupied by the second class particle, the particles exchange site. Therefore the evolution of the first class particles is unaffected by the presence of the second class particles. We will show here that the asymptotic evolution of the second class particle is closely related to the equilibrium fluctuations of the density.

Let ν^0_α the Bernoulli measure on $\mathbb{Z}^d$ with parameter α conditioned to have the site 0 empty. Let X_t be the position at time t of the second class particle, one starting at time 0 from 0, when the other particles are distributed initially with ν^0_α.

Proposition 6.1.

$$E^{\nu^0_\alpha}(1_{[X_t=x]}) = \frac{1}{\alpha(1-\alpha)}\left[E^{\nu_\alpha}(\eta_t(x)\eta_0(0)) - \alpha^2\right] \tag{6.1}$$

Proof. Let $\sigma_0\eta(x) = \eta(x)$ if $x \neq 0$ and $\sigma_0\eta(0) = 0$.

$$\begin{aligned} E^{\nu_\alpha}(\eta_t(x)\eta_0(0)) &= \int \nu_\alpha(d\eta)\eta(0)E^\eta(\eta_t(x)) \\ &= \int \nu_\alpha(d\eta)\eta(0)\left[E^\eta(\eta_t(x)) - E^{\sigma_0\eta}(\eta_t(x))\right] \\ &\quad + \int \nu_\alpha(d\eta)\eta(0)E^{\sigma_0\eta}(\eta_t(x)) \\ &= \int \nu_\alpha(d\eta)\eta(0)E^\eta(1_{[X_t=x]}) + \int \nu_\alpha(d\xi)\frac{\alpha}{1-\alpha}(1-\xi(0))E^\xi(\eta_t(x)) \end{aligned}$$

(where in the last term we have performed the change of variable $\xi = \sigma_0\eta$)

$$= \alpha E^{\nu^0_\alpha}(1_{[X_t=x]}) + \frac{\alpha}{1-\alpha}\alpha - \frac{\alpha}{1-\alpha}E^{\nu_\alpha}(\eta_t(x)\eta_0(0))$$

Reordering the term one obtains (6.1). □

Theorem 6.2. *$\varepsilon(X_{t\varepsilon^{-2}} - vt\varepsilon^{-2})$ converges in law to a zero mean Gaussian r.w. with covariance matrix $tD^s_{i,j}$, where D^s is the symmetric part of the viscosity matrix D introduced in the previous section.*

Proof. Let $H(y)$ a bounded continuous function on $\mathbb{R}^d$ with compact support. Let $G(y)$ a probability density on $\mathbb{R}^d$ and let ν^y_α the Bernoulli measure conditioned to have the site y empty. Then

$$\varepsilon^d \sum_y G(\varepsilon y) E^{\nu^y_\alpha} \left(H(\varepsilon(X^y_{t\varepsilon^{-2}} - vt\varepsilon^{-2})) \right)$$
$$= \varepsilon^d \sum_y G(\varepsilon y) \sum_x H(\varepsilon(x - vt\varepsilon^{-2})) E^{\nu^y_\alpha}(1_{[X^y_{t\varepsilon^{-2}}=x]}).$$

In this formula, X^y_t stands for the position at time t of a second class particle initially at y. By (6.1) this is equal to

$$\frac{1}{\alpha(1-\alpha)} \varepsilon^d \sum_y \sum_x H(\varepsilon(x - vt\varepsilon^{-2})) G(\varepsilon y) \left[E^{\nu_\alpha}(\eta_{t\varepsilon^{-2}}(x)\eta_0(y)) - \alpha^2 \right]$$
$$= \frac{1}{\alpha(1-\alpha)} E^{\nu_\alpha} \Bigg[\varepsilon^d \sum_x H(\varepsilon(x - vt\varepsilon^{-2}))$$
$$\times (\eta_{t\varepsilon^{-2}}(x) - \alpha) \sum_y G(\varepsilon y)(\eta_0(y) - \alpha) \Bigg]$$
$$= \frac{1}{\alpha(1-\alpha)} E^{\nu_\alpha} \left[Y^\varepsilon_t(H) Y^\varepsilon_0(G) \right]$$

and by the convergence result for Y^ε_t proved in the previous section, this converge as $\varepsilon \to 0$ to

$$\int_{\mathbb{R}^d} G(u)\, e^{tA} H(u)\, du.$$

□

An immediate consequence of the above result is the following formula for the symmetric part of the viscosity matrix:

$$D^s_{i,j} = \lim_{t\to\infty} \frac{1}{t} \frac{1}{\alpha(1-\alpha)} \sum_x x_i x_j \left[E^{\nu_\alpha}(\eta_t(x + tv)\eta_0(0)) - \alpha^2 \right] .$$

References

[1] C. C. Chang, C. Landim, S. Olla: Equilibrium fluctuations of asymmetric exclusion processes in dimension $d \geq 3$. Probability Theory and Related Field, **119**, 381-409, (2001).

[2] O. Benois, R. Esposito, R. Marra: Equilibrium fluctuations for lattice gases. To appear in Ann. Inst. H. Poincaré, Probabilités et Statistiques (2003).

[3] R. Esposito, R. Marra, H. T. Yau: Diffusive limit of asymmetric simple exclusion. Rev. Math. Phys. **6**, 1233-1267 (1994).

[4] C. Kipnis, C. Landim: *Scaling Limit of Interacting Particle Systems*, Springer Verlag, Berlin (1999).

[5] C. Landim, S. Olla, H. T. Yau; First order correction for the hydrodynamic limit of asymmetric simple exclusion processes in dimension $d \geq 3$. Communications on Pure and Applied Mathematics **50**, 149–203, (1997).

[6] C. Landim, S. Olla, H. T. Yau; Some properties of the diffusion coefficient for asymmetric simple exclusion processes. The Annals of Probability **24**, 1779–1807, (1996).

[7] C. Landim, H. T. Yau; Fluctuation–dissipation equation of asymmetric simple exclusion processes, Probab. Th. Rel. Fields **108**, 321–356, (1997)

[8] C. Landim, S. Olla, S. R. S. Varadhan; On viscosity and fluctuation-dissipation in exclusion processes. To appear in J. Stat. Phys. (2003).

[9] C. Landim, J. Quastel, M. Salmhofer, H.T. Yau; Superdiffusivity of asymmetric exclusion process in dimension 1 and 2, to appear in Commun. Math. Phys., arXiv:math.PR/0201317, 2002.

[10] T. M. Liggett; *Interacting Particle Systems*, Springer Verlag, New York, 1985.

[11] F. Rezakhanlou: Hydrodynamic limit for attractive particle systems on Z^d. Commun. Math. Phys. **140**, 417-448, (1990).

[12] S. Sethuraman, S. R. S. Varadhan, H. T. Yau: Diffusive limit of a tagged particle in asymmetric exclusion process, Comm.Pure Appl.Math. **53**, 972-1006, (2000).

[13] H. Spohn: *Large Scale Dynamics of Interacting Particles*, Springer 1991.

[14] H. T. Yau: $(\log t)^{2/3}$ law of the two dimensional asymmetric simple exclusion process, preprint, arXiv:math-ph/0201057 v1, 2002.

Claudio Landim
IMPA
Estrada Dona Castorina 110
CEP 22460 Rio de Janeiro, Brasil
and
CNRS UMR 6085, Université de Rouen
76128 Mont Saint Aignan, France.
E-mail address: `landim@impa.br`

Stefano Olla
Ceremade
UMR CNRS 7534
Université de Paris IX - Dauphine
Place du Maréchal De Lattre De Tassigny
75775 Paris Cedex 16 - France.
E-mail address: `olla@ceremade.dauphine.fr`

Srinivasa R. S. Varadhan
Courant Institute of Mathematical Sciences
New York University
251 Mercer street
New York, NY, 10012
U.S.A.
E-mail address: `varadhan@cims.nyu.edu`

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 325–343

Non-Collision and Collision Properties of Dyson's Model in Infinite Dimension and Other Stochastic Dynamics Whose Equilibrium States are Determinantal Random Point Fields

Hirofumi Osada

Dedicated to Professor Tokuzo Shiga on his 60th birthday

Abstract.

Dyson's model on interacting Brownian particles is a stochastic dynamics consisting of an infinite amount of particles moving in $\mathbb{R}$ with a logarithmic pair interaction potential. For this model we will prove that each pair of particles never collide.

The equilibrium state of this dynamics is a determinantal random point field with the sine kernel. We prove for stochastic dynamics given by Dirichlet forms with determinantal random point fields as equilibrium states the particles never collide if the kernel of determining random point fields are locally Lipschitz continuous, and give examples of collision when Hölder continuous.

In addition we construct infinite volume dynamics (a kind of infinite dimensional diffusions) whose equilibrium states are determinantal random point fields. The last result is partial in the sense that we simply construct a diffusion associated with the *maximal closable part* of *canonical* pre Dirichlet forms for given determinantal random point fields as equilibrium states. To prove the closability of canonical pre Dirichlet forms for given determinantal random point fields is still an open problem. We prove these dynamics are the strong resolvent limit of finite volume dynamics.

§1. Introduction

Dyson's model on interacting Brownian particles in infinite dimension is an infinitely dimensional diffusion process $\{(X_t^i)_{i\in\mathbb{N}}\}$ formally

---

Received February 9, 2003.
Revised March 31, 2003.
Partially supported by Grant-in-Aid for Scientific Research (B) 11440029.

given by the following stochastic differential equation (SDE):

$$(1.1)\qquad dX_t^i = dB_t^i + \sum_{j=1,\ j\neq i}^{\infty} \frac{1}{X_t^i - X_t^j}\, dt \quad (i = 1, 2, 3, \dots),$$

where $\{B_t^i\}$ is an infinite amount of independent one dimensional Brownian motions. The corresponding unlabeled dynamics is

$$(1.2)\qquad \mathbb{X}_t = \sum_{i=1}^{\infty} \delta_{X_t^i}.$$

Here $\delta_\cdot$ denote the point mass at $\cdot$. By definition $\mathbb{X}_t$ is a Θ-valued diffusion, where Θ is the set consisting of configurations on $\mathbb{R}$; that is,

$$(1.3)\qquad \Theta = \{\theta = \sum_i \delta_{x_i}\, ;\ x_i \in \mathbb{R},\ \theta(\{|x| \le r\}) < \infty \text{ for all } r \in \mathbb{R}\}.$$

We regard Θ as a complete, separable metric space with the vague topology.

In [11] Spohn constructed an unlabeled dynamics (1.2) in the sense of a Markovian semigroup on $L^2(\Theta, \mu)$. Here μ is a probability measure on $(\Theta, \mathfrak{B}(\Theta))$ whose correlation functions are generated by the sine kernel

$$(1.4)\qquad \mathsf{K}_{\sin}(x) = \frac{\bar{\rho}}{\pi x} \sin(\pi x).$$

(See Section 2). Here $0 < \bar{\rho} \le 1$ is a constant related to the *density* of the particle. Spohn indeed proved the closability of a non-negative bilinear form $(\mathcal{E}, \mathcal{D}_\infty)$ on $L^2(\Theta, \mu)$

$$(1.5)\qquad \begin{aligned} \mathcal{E}(\mathfrak{f}, \mathfrak{g}) &= \int_\Theta \mathbb{D}[\mathfrak{f}, \mathfrak{g}](\theta) d\mu, \\ \mathcal{D}_\infty &= \{\mathfrak{f} \in \mathcal{D}_\infty^{loc} \cap L^2(\Theta, \mu); \mathcal{E}(\mathfrak{f}, \mathfrak{f}) < \infty\}. \end{aligned}$$

Here $\mathbb{D}$ is the square field given by (2.8) and $\mathcal{D}_\infty^{loc}$ is the set of the local smooth functions on Θ (see Section 3 for the definition). The Markovian semi-group is given by the Dirichlet form that is the closure $(\mathcal{E}, \mathcal{D})$ of this closable form on $L^2(\Theta, \mu)$.

The measure μ is an equilibrium state of (1.2), whose formal Hamiltonian $\mathcal{H} = \mathcal{H}(\theta)$ is given by $(\theta = \sum_i \delta_{x_i})$

$$(1.6)\qquad \mathcal{H}(\theta) = \sum_{i \neq j} -2 \log |x_i - x_j|,$$

which is a reason we regard Spohn's Markovian semi-group is a correspondent to the dynamics formally given by the SDE (1.1) and (1.2).

We remark the existence of an L^2-Markovian semigroup does not imply the existence of the associated *diffusion* in general. Here a diffusion means (a family of distributions of) a strong Markov process with continuous sample paths starting from each $\theta \in \Theta$.

In [5] it was proved that there exists a diffusion $(\{\mathsf{P}_\theta\}_{\theta\in\Theta}, \{\mathbb{X}_t\})$ with state space Θ associated with the Markovian semigroup above. This construction admits us to investigate the *trajectory-wise* properties of the dynamics. In the present paper we concentrate on the collision property of the diffusion. The problem we are interested in is the following:

Does a pair of particles (X_t^i, X_t^j) that collides each other for some time $0 < t < \infty$ exist ?

We say for a diffusion on Θ *the non-collision occurs* if the above property does *not* hold, and *the collision occurs* if otherwise.

If the number of particles is finite, then the non-collision should occur at least intuitive level. This is because drifts $\frac{1}{x_i - x_j}$ have a strong repulsive effect. When the number of the particles is infinite, the non-collision property is non-trivial because the interaction potential is long range and un-integrable. We will prove the non-collision property holds for Dyson's model in infinite dimension.

Since the sine kernel measure is the prototype of determinantal random point fields, it is natural to ask such a non-collision property is universal for stochastic dynamics given by Dirichlet forms (1.5) with the replacement of the measure μ with general determinantal random point fields. We will prove, if the kernel of the determinantal random point field (see (2.3)) is locally Lipschitz continuous, then the non-collision always occurs. In addition, we give an example of determinantal random point fields with Hölder continuous kernel that the collision occurs.

The second problem we are interested in this paper is the following:

Does there exist Θ-valued diffusions associated with the Dirichlet forms $(\mathcal{E}, \mathcal{D})$ on $L^2(\Theta, \mu)$ when μ is determinantal random point fields ?

We give a partial answer for this in Theorem 2.5.

The organization of the paper is as follows: In Section 2 we state main theorems. In Section 3 we prepare some notion on configuration spaces. In Section 4 we prove Theorem 2.2 and Theorem 2.3. In Section 5 we prove Proposition 2.9 and Theorem 2.4. In Section 6 we prove Theorem 2.5. Our method proving Theorem 2.1 can be applied to Gibbs measures. So we prove the non-collision property for Gibbs measures in Section 7.

§2. Set up and the main result

Let $\mathsf{E} \subset \mathbb{R}^d$ be a closed set which is the closure of a connected open set in $\mathbb{R}^d$ with smooth boundary. Although we will mainly treat the case $\mathsf{E} = \mathbb{R}$, we give a general framework here by following the line of [10]. Let Θ denote the set of configurations on E, which is defined similarly as (1.3) by replacing $\mathbb{R}$ with E.

A probability measure on $(\Theta, \mathcal{B}(\Theta))$ is called a random point field on E. Let μ be a random point field on E. A non-negative, permutation invariant function $\rho_n : \mathsf{E}^n \to \mathbb{R}$ is called an n-correlation function of μ if for any measurable sets $\{A_1, \dots, A_m\}$ and natural numbers $\{k_1, \dots, k_m\}$ such that $k_1 + \cdots + k_m = n$ the following holds:

$$\int_{A_1^{k_1} \times \cdots \times A_m^{k_m}} \rho_n(x_1, \dots, x_n) dx_1 \cdots dx_n = \int_\Theta \prod_{i=1}^m \frac{\theta(A_i)!}{(\theta(A_i) - k_i)!} d\mu.$$

It is known ([10], [3], [4]) that, if a family of non-negative, permutation invariant functions $\{\rho_n\}$ satisfies

$$\sum_{k=1}^\infty \left\{ \frac{1}{(k+j)!} \int_{A^{k+j}} \rho_{k+j} \, dx_1 \cdots dx_{k+j} \right\}^{-1/k} = \infty, \tag{2.1}$$

then there exists a unique probability measure (random point field) μ on E whose correlation functions equal $\{\rho_n\}$.

Let $K : L^2(\mathsf{E}, dx) \to L^2(\mathsf{E}, dx)$ be a non-negative definite operator which is locally trace class; namely

$$\begin{aligned} &0 \le (Kf, f)_{L^2(\mathsf{E},dx)}, \\ &\mathrm{Tr}(1_B K 1_B) < \infty \quad \text{for all bounded Borel set } B. \end{aligned} \tag{2.2}$$

We assume K has a continuous kernel denoted by $\mathsf{K} = \mathsf{K}(x, y)$. Without this assumption one can develop a theory of determinantal random point fields (see [10], [9]); we assume this for the sake of simplicity.

Definition 2.1. A probability measure μ on Θ is said to be a determinantal (or fermion) random point field with kernel K if its correlation functions ρ_n are given by

$$\rho_n(x_1, \dots, x_n) = \det(\mathsf{K}(x_i, x_j)_{1 \le i,j \le n}) \tag{2.3}$$

We quote:

Lemma 2.2 (Theorem 3 in [10]). *Assume* $\mathsf{K}(x, y) = \overline{\mathsf{K}(y, x)}$ *and* $0 \le K \le 1$. *Then* K *determines a unique determinantal random point field* μ.

We give examples of determinantal random point fields. The first example is the stationary measure of Dyson's model in infinite dimension. The first three examples are related to the semicircle law of empirical distribution of eigen values of random matrices. We refer to [10] for detail.

Example 2.3 (sine kernel). Let $\mathsf{K}_{\sin}$ and $\bar{\rho}$ be as in (1.4). Then

$$\mathsf{K}_{\sin}(t) = \frac{1}{2\pi}\int_{|k|\le\pi\bar{\rho}} e^{\sqrt{-1}kt}\,dk. \tag{2.4}$$

So the $\mathsf{K}_{\sin}$ is a function of positive type and satisfies the assumptions in Lemma 2.2. Let $\hat{\mu}^N$ denote the probability measure on $\mathbb{R}^N$ defined by

$$\hat{\mu}^N = \frac{1}{Z^N} e^{-\sum_{i,j=1}^{N} -2\log|x_i-x_j|} e^{-\lambda_N^2\sum_{i=1}^{N}x_i^2}\,dx_1\cdots dx_N, \tag{2.5}$$

where $\lambda_N = 2(\pi\bar{\rho})^3/3N^2$ and Z^N is the normalization. Set $\mu^N = \hat{\mu}^N \circ (\xi^N)^{-1}$, where $\xi^N : \mathbb{R}^N \to \Theta$ such that $\xi^N(x_1,\dots,x_N) = \sum_{i=1}^N \delta_{x_i}$. Let ρ_n^N denote the n-correlation function of μ^N. Let ρ_n denote the n-correlation function of μ. Then it is known ([11, Proposition 1], [10]) that for all $n = 1,2,\dots$

$$\lim_{N\to\infty} \rho_n^N(x_1,\dots,x_n) = \rho_n(x_1,\dots,x_n) \quad \text{for all } (x_1,\dots,x_n). \tag{2.6}$$

In this sense the measure μ is associated with the Hamiltonian $\mathcal{H}$ in (1.6) coming from the log potential $-2\log|x|$.

Example 2.4 (Airy kernel). $\mathsf{E} = \mathbb{R}$ and

$$\mathsf{K}(x,y) = \frac{\mathcal{A}_i(x)\cdot\mathcal{A}_i'(y) - \mathcal{A}_i(y)\cdot\mathcal{A}_i'(x)}{x-y}$$

Here $\mathcal{A}_i$ is the Airy function.

Example 2.5 (Bessel kernel). Let $\mathsf{E} = [0,\infty)$ and

$$\mathsf{K}(x,y) = \frac{J_\alpha(\sqrt{x})\cdot\sqrt{y}\cdot J_\alpha'(\sqrt{y}) - J_\alpha(\sqrt{y})\cdot\sqrt{x}\cdot J_\alpha'(\sqrt{x})}{2(x-y)}.$$

Here J_α is the Bessel function of order α.

Example 2.6. Let $\mathsf{E} = \mathbb{R}$ and $\mathsf{K}(x,y) = \mathsf{m}(x)\mathsf{k}(x-y)\mathsf{m}(y)$, where $\mathsf{k}:\mathbb{R}\to\mathbb{R}$ is a non-negative, continuous *even* function that is convex in $[0,\infty)$ such that $\mathsf{k}(0) \le 1$, and $\mathsf{m} : \mathbb{R} \to \mathbb{R}$ is nonnegative continuous and $\int_{\mathbb{R}} \mathsf{m}(t)dt < \infty$ and $\mathsf{m}(x) \le 1$ for all x and $0 < \mathsf{m}(x)$ for some

x. Then K satisfies the assumptions in Lemma 2.2. Indeed, it is well-known that k is a function of positive type (187 p. in [1] for example), so the Fourier transformation of a finite positive measure. By assumption $0 \le \mathsf{K}(x,y) \le 1$, which implies $0 \le K \le 1$. Since $\int \mathsf{K}(x,x)dx < \infty$, K is of trace class.

Let A denote the subset of Θ defined by

$$\mathsf{A} = \{\theta \in \Theta;\ \theta(\{x\}) \ge 2 \quad \text{for some } x \in \mathsf{E}\}. \tag{2.7}$$

Note that A denotes the set consisting of the configurations with collisions. We are interested in how large the set A is. Of course $\mu(\mathsf{A}) = 0$ because the 2-correlation function is locally integrable. We study A more closely from the point of stochastic dynamics; namely, we measure A by using a capacity.

To introduce the capacity we next consider a bilinear form related to the given probability measure μ. Let $\mathcal{D}_\infty^{loc}$ be the set of all local, smooth functions on Θ defined in Section 3. For $\mathfrak{f}, \mathfrak{g} \in \mathcal{D}_\infty^{loc}$ we set $\mathbb{D}[\mathfrak{f}, \mathfrak{g}] : \Theta \to \mathbb{R}$ by

$$\mathbb{D}[\mathfrak{f}, \mathfrak{g}](\theta) = \frac{1}{2} \sum_i \frac{\partial f(\mathbf{x})}{\partial x_i} \frac{\partial g(\mathbf{x})}{\partial x_i}. \tag{2.8}$$

Here $\theta = \sum_i \delta_{x_i}$, $\mathbf{x} = (x_1, \dots)$ and $f(\mathbf{x}) = f(x_1, \dots)$ is the permutation invariant function such that $\mathfrak{f}(\theta) = f(x_1, x_2, \dots)$ for all $\theta \in \Theta$. We set g similarly. Note that the left hand side of (2.8) is again permutation invariant. Hence it can be regard as a function of $\theta = \sum_i \delta_{x_i}$. Such f and g are unique; so the function $\mathbb{D}[\mathfrak{f}, \mathfrak{g}] : \Theta \to \mathbb{R}$ is well defined.

For a probability measure μ in Θ we set as before

$$\mathcal{E}(\mathfrak{f}, \mathfrak{g}) = \int_\Theta \mathbb{D}[\mathfrak{f}, \mathfrak{g}](\theta) d\mu,$$
$$\mathcal{D}_\infty = \{\mathfrak{f} \in \mathcal{D}_\infty^{loc} \cap L^2(\Theta, \mu); \mathcal{E}(\mathfrak{f}, \mathfrak{f}) < \infty\}.$$

When $(\mathcal{E}, \mathcal{D}_\infty)$ is closable on $L^2(\Theta, \mu)$, we denote its closure by $(\mathcal{E}, \mathcal{D})$.

We are now ready to introduce a notion of capacity for a *pre-*Dirichlet space $(\mathcal{E}, \mathcal{D}_\infty, L^2(\Theta, \mu))$. Let $\mathcal{O}$ denote the set consisting of all open sets in Θ. For $O \in \mathcal{O}$ we set $\mathcal{L}_O = \{\mathfrak{f} \in \mathcal{D}_\infty\,;\ \mathfrak{f} \ge 1\ \mu\text{-a.e. on } O\}$ and

$$\mathrm{Cap}(O) = \begin{cases} \inf_{\mathfrak{f} \in \mathcal{L}_O} \{\mathcal{E}(\mathfrak{f}, \mathfrak{f}) + (\mathfrak{f}, \mathfrak{f})_{L^2(\Theta,\mu)}\} & \mathcal{L}_O \ne \emptyset \\ \infty & \mathcal{L}_O = \emptyset \end{cases}.$$

For an arbitrary subset $A \subset \Theta$ we set $\mathrm{Cap}(A) = \inf_{A \subset O \in \mathcal{O}} \mathrm{Cap}(O)$. This quantity Cap is called 1-capacity for the pre-Dirichlet space $(\mathcal{E}, \mathcal{D}_\infty, L^2(\Theta, \mu))$.

We state the main theorem:

Theorem 2.1. *Let μ be a determinantal random point field with kernel K. Assume K is locally Lipschitz continuous. Then*

$$\mathrm{Cap}(\mathsf{A}) = 0, \tag{2.9}$$

where A is given by (2.7).

In [5] it was proved

Lemma 2.7 (Corollary 1 in [5]). *Let μ be a probability measure on Θ. Assume μ has locally bounded correlation functions. Assume $(\mathcal{E}, \mathcal{D}_\infty)$ is closable on $L^2(\Theta, \mu)$. Then there exists a diffusion $(\{\mathsf{P}_\theta\}_{\theta \in \Theta}, \{\mathbb{X}_t\})$ associated with the Dirichlet space $(\mathcal{E}, \mathcal{D}, L^2(\Theta, \mu))$.*

Combining this with Theorem 2.1 we have

Theorem 2.2. *Assume μ satisfies the assumption in Theorem 2.1. Assume $(\mathcal{E}, \mathcal{D}_\infty)$ is closable on $L^2(\Theta, \mu)$. Then a diffusion $(\{\mathsf{P}_\theta\}_{\theta \in \Theta}, \{\mathbb{X}_t\})$ associated with the Dirichlet space $(\mathcal{E}, \mathcal{D}, L^2(\Theta, \mu))$ exists and satisfies*

$$\mathsf{P}_\theta(\sigma_{\mathsf{A}} = \infty) = 1 \quad \textit{for q.e. } \theta, \tag{2.10}$$

where $\sigma_{\mathsf{A}} = \inf\{t > 0\,;\, \mathbb{X}_t \in \mathsf{A}\}$.

We refer to [2] for q.e. (quasi everywhere) and related notions on Dirichlet form theory. We remark the capacity of pre-Dirichlet forms are bigger than or equal to the one of its closure by definition. So (2.10) is an immediate consequence of Theorem 2.1 and the general theory of Dirichlet forms once $(\mathcal{E}, \mathcal{D}_\infty)$ is closable on $L^2(\Theta, \mu)$ and the resulting (quasi) regular Dirichlet space $(\mathcal{E}, \mathcal{D}, L^2(\Theta, \mu))$ exists.

To apply Theorem 2.2 to Dyson's model we recall a result of Spohn.

Lemma 2.8 (Proposition 4 in [11]). *Let μ be the determinantal random point field with the sine kernel in Example 2.3. Then $(\mathcal{E}, \mathcal{D}_\infty)$ is closable on $L^2(\Theta, \mu)$.*

We say a diffusion $(\{\mathsf{P}_\theta\}_{\theta \in \Theta}, \{\mathbb{X}_t\})$ is Dyson's model in infinite dimension if it is associated with the Dirichlet space $(\mathcal{E}, \mathcal{D}, L^2(\Theta, \mu))$ in Theorem 2.8. Collecting these we conclude:

Theorem 2.3. *No collision* (2.10) *occurs in Dyson's model in infinite dimension.*

The assumption of the local Lipschitz continuity of the kernel K is crucial; we next give a collision example when K is merely Hölder continuous. We prepare:

Proposition 2.9. *Assume K is of trace class. Then $(\mathcal{E}, \mathcal{D}_\infty)$ is closable on $L^2(\Theta, \mu)$.*

Theorem 2.4. *Let $\mathsf{K}(x, y) = \mathsf{m}(x)\mathsf{k}(x-y)\mathsf{m}(y)$ be as in Example 2.6. Let α be a constant such that*

$$0 < \alpha < 1. \tag{2.11}$$

Assume m and k are continuous and there exist positive constants c_1 and c_2 such that

$$c_1 t^\alpha \le \mathsf{k}(0) - \mathsf{k}(t) \le c_2 t^\alpha \quad \textit{for } 0 \le t \le 1. \tag{2.12}$$

Then $(\mathcal{E}, \mathcal{D}_\infty, L^2(\Theta, \mu))$ is closable and the associated diffusion satisfies

$$\mathsf{P}_\theta(\sigma_{\mathsf{A}} < \infty) = 1 \quad \textit{for q.e. } \theta. \tag{2.13}$$

Unfortunately the closability of the pre-Dirichlet form $(\mathcal{E}, \mathcal{D}_\infty)$ on $L^2(\Theta, \mu)$ has not yet proved for determinantal random point fields of locally trace class except the sine kernel. So we propose a problem:

Problem 2.10. (1) Are pre-Dirichlet forms $(\mathcal{E}, \mathcal{D}_\infty)$ on $L^2(\Theta, \mu)$ closable when μ are determinantal random fields with continuous kernels?
(2) Can one construct stochastic dynamics (diffusion processes) associated with pre-Dirichlet forms $(\mathcal{E}, \mathcal{D}_\infty)$ on $L^2(\Theta, \mu)$.

We remark one can deduce the second problem from the first one (see [5, Theorem 1]). We conjecture that $(\mathcal{E}, \mathcal{D}_\infty, L^2(\Theta, \mu))$ are always closable. As we see above, in case of trace class kernel, this problem is solved by Proposition 2.9. But it is important to prove this for determinantal random point field of *locally* trace class. This class contains Airy kernel and Bessel kernel and other nutritious examples. We also remark for interacting Brownian motions with Gibbsian equilibriums this problem was settled successfully ([5]).

In the next theorem we give a partial answer for (2) of Problem 2.10. We will show one can construct a stochastic dynamics in infinite volume, which is canonical in the sense that (1) it is the strong resolvent limit of a sequence of finite volume dynamics and that (2) it coincides with $(\mathcal{E}, \mathcal{D})$ whenever $(\mathcal{E}, \mathcal{D}_\infty)$ is closable on $L^2(\Theta, \mu)$.

For two symmetric, nonnegative forms $(\mathcal{E}_1, \mathcal{D}_1)$ and $(\mathcal{E}_2, \mathcal{D}_2)$, we write $(\mathcal{E}_1, \mathcal{D}_1) \le (\mathcal{E}_2, \mathcal{D}_2)$ if $\mathcal{D}_1 \supset \mathcal{D}_2$ and $\mathcal{E}_1(\mathfrak{f}, \mathfrak{f}) \le \mathcal{E}_2(\mathfrak{f}, \mathfrak{f})$ for all $\mathfrak{f} \in \mathcal{D}_2$.

Let $(\mathcal{E}^{\mathrm{reg}}, \mathcal{D}^{\mathrm{reg}})$ denote the regular part of $(\mathcal{E}, \mathcal{D}_\infty)$ on $L^2(\Theta, \mu)$, that is, $(\mathcal{E}^{\mathrm{reg}}, \mathcal{D}^{\mathrm{reg}})$ is closable on $L^2(\Theta, \mu)$ and in addition satisfies the following:

$$(\mathcal{E}^{\mathrm{reg}}, \mathcal{D}^{\mathrm{reg}}) \leq (\mathcal{E}, \mathcal{D}_\infty),$$

and for all closable forms such that $(\mathcal{E}', \mathcal{D}') \leq (\mathcal{E}, \mathcal{D}_\infty)$

$$(\mathcal{E}', \mathcal{D}') \leq (\mathcal{E}^{\mathrm{reg}}, \mathcal{D}^{\mathrm{reg}}).$$

It is well known that such a $(\mathcal{E}^{\mathrm{reg}}, \mathcal{D}^{\mathrm{reg}})$ exists uniquely and called the maximal regular part of $(\mathcal{E}, \mathcal{D})$. Let us denote the closure by the same symbol $(\mathcal{E}^{\mathrm{reg}}, \mathcal{D}^{\mathrm{reg}})$.

Let $\pi_r : \Theta \to \Theta$ be such that $\pi_r(\theta) = \theta(\cdot \cap \{x \in \mathsf{E}; |x| < r\})$. We set

$$\mathcal{D}_{\infty,r} = \{\mathfrak{f} \in \mathcal{D}_\infty ; \mathfrak{f} \text{ is } \sigma[\pi_r]\text{-measurable}\}.$$

We will prove $(\mathcal{E}, \mathcal{D}_{\infty,r})$ are closable on $L^2(\Theta, \mu)$. These are the finite volume dynamics we are considering.

Let $\mathbb{G}_\alpha$ (resp. $\mathbb{G}_{r,\alpha}$) $(\alpha > 0)$ denote the α-resolvent of the semi-group associated with the closure of $(\mathcal{E}^{\mathrm{reg}}, \mathcal{D}^{\mathrm{reg}})$ (resp. $(\mathcal{E}, \mathcal{D}_{\infty,r})$) on $L^2(\Theta, \mu)$.

Theorem 2.5. (1) *$(\mathcal{E}^{reg}, \mathcal{D}^{reg})$ on $L^2(\Theta, \mu)$ is a quasi-regular Dirichlet form. So the associated diffusion exists.*
(2) *$\mathbb{G}_{r,\alpha}$ converge to $\mathbb{G}_\alpha$ strongly in $L^2(\Theta, \mu)$ for all $\alpha > 0$.*

Remark 2.11. We think the diffusion constructed in Theorem 2.5 is a reasonable one because of the following reason. (1) By definition the closure of $(\mathcal{E}^{\mathrm{reg}}, \mathcal{D}^{\mathrm{reg}})$ equals $(\mathcal{E}, \mathcal{D})$ when $(\mathcal{E}, \mathcal{D}_\infty)$ is closable. (2) One naturally associated Markov processes on Θ_r, where Θ_r is the set of configurations on $\mathsf{E} \cap \{|x| < r\}$. So (2) of Theorem 2.5 implies the diffusion is the strong resolvent limit of finite volume dynamics.

Remark 2.12. If one replace μ by the Poisson random measure λ whose intensity measure is the Lebesgue measure and consider the Dirichlet space $(\mathcal{E}^\lambda, \mathcal{D})$ on $L(\Theta, \lambda)$, then the associated Θ-valued diffusion is the Θ-valued Brownian motion $\mathbb{B}$, that is, it is given by

$$\mathbb{B}_t = \sum_{i=1}^{\infty} \delta_{B_t^i},$$

where $\{B_t^i\}$ $(i \in \mathbb{N}\ )$ are infinite amount of independent Brownian motions. In this sense we say in Abstract that the Dirichlet form given by (1.5) for Radon measures in Θ *canonical.* We also remark such a type of local Dirichlet forms are often called distorted Brownian motions.

§3. Preliminary

Let $I_r = (-r,r)^d \cap \mathsf{E}$ and $\Theta_r^n = \{\theta \in \Theta; \theta(I_r) = n\}$. We note $\Theta = \sum_{n=0}^{\infty} \Theta_r^n$. Let I_r^n be the n times product of I_r. We define $\pi_r : \Theta \to \Theta$ by $\pi_r(\theta) = \theta(\cdot \cap I_r)$. A function $\mathbf{x} : \Theta_r^n \to I_r^n$ is called a I_r^n-coordinate of θ if

$$(3.1) \qquad \pi_r(\theta) = \sum_{k=1}^{n} \delta_{x_k(\theta)}, \qquad \mathbf{x}(\theta) = (x_1(\theta), \dots, x_n(\theta)).$$

Suppose $\mathfrak{f} : \Theta \to \mathbb{R}$ is $\sigma[\pi_r]$-measurable. Then for each $n = 1, 2, \dots$ there exists a unique permutation invariant function $f_r^n : I_r^n \to \mathbb{R}$ such that

$$(3.2) \qquad \mathfrak{f}(\theta) = f_r^n(\mathbf{x}(\theta)) \quad \text{for all } \theta \in \Theta_r^n.$$

We next introduce mollifier. Let $j : \mathbb{R} \to \mathbb{R}$ be a non-negative, smooth function such that $j(x) = j(|x|)$, $\int_{\mathbb{R}^d} j dx = 1$ and $j(x) = 0$ for $|x| \geq \frac{1}{2}$. Let $j_\epsilon = \epsilon j(\cdot/\epsilon)$ and $j_\epsilon^n(x_1, \dots, x_n) = \prod_{i=1}^n j_\epsilon(x_i)$. For a $\sigma[\pi_r]$-measurable function $\mathfrak{f}$ we set $\mathfrak{J}_{r,\epsilon}\mathfrak{f} : \Theta \to \mathbb{R}$ by

$$(3.3) \qquad \mathfrak{J}_{r,\epsilon}\mathfrak{f}(\theta) = \begin{cases} j_\epsilon^n * \hat{f}_r^n(\mathbf{x}(\theta)) & \text{for } \theta \in \Theta_r^n, n \geq 1 \\ \mathfrak{f}(\theta) & \text{for } \theta \in \Theta_r^0, \end{cases}$$

where f_r^n is given by (3.2) for $\mathfrak{f}$, and $\hat{f}_r^n : \mathbb{R}^{dn} \to \mathbb{R}$ is the function defined by $\hat{f}_r^n(x) = f_r^n(x)$ for $x \in I_r^n$ and $\hat{f}_r^n(x) = 0$ for $x \notin I_r^n$. Moreover $\mathbf{x}(\theta)$ is an I_r^n-coordinate of $\theta \in \Theta_r^n$, and $*$ denotes the convolution in $\mathbb{R}^n$. It is clear that $\mathfrak{J}_{r,\epsilon}\mathfrak{f}$ is $\sigma[\pi_r]$-measurable.

We say a function $\mathfrak{f} : \Theta \to \mathbb{R}$ is local if $\mathfrak{f}$ is $\sigma[\pi_r]$-measurable for some $r < \infty$. For $\mathfrak{f} : \Theta \to \mathbb{R}$ and $n \in \mathbb{N} \cup \{\infty\}$ there exists a unique permutation function f^n such that $\mathfrak{f}(\theta) = f^n(x_1, \dots)$ for all $\theta \in \Theta^n$. Here $\Theta^n = \{\theta \in \Theta\,; \theta(\mathsf{E}) = n\}$, and $\theta = \sum_i \delta_{x_i}$. A function $\mathfrak{f}$ is called smooth if f^n is smooth for all $n \in \mathbb{N} \cup \{\infty\}$. Note that a $\sigma[\pi_r]$-measurable function $\mathfrak{f}$ is smooth if and only if f_r^n is smooth for all $n \in \mathbb{N}$.

§4. Proof of Theorem 2.2

We give a sequence of reductions of (2.9). Let $\mathbf{A}$ denote the set consisting of the sequences $\mathbf{a} = (a_r)_{r \in \mathbb{N}}$ satisfying the following:

$$(4.1) \qquad a_r \in \mathbb{Q} \quad \text{for all } r \in \mathbb{N},$$

$$(4.2) \qquad a_r = 2r + r_0 \quad \text{for all sufficiently large } r \in \mathbb{N},$$

$$(4.3) \qquad 2 \leq a_1, \ 1 \leq a_{r+1} - a_r \leq 2 \quad \text{for all } r \in \mathbb{N}.$$

Note that the cardinality of $\mathbf{A}$ is countable by (4.1) and (4.2).

Let $\mathbb{I} = \{2, 3, \dots , \}^3$. For $(r, n, m) \in \mathbb{I}$ and $\mathbf{a} = (a_r) \in \mathbf{A}$ we set

$$\Theta^{\mathbf{a}}(r, n) = \{\theta \in \Theta \,;\, \theta(I_{a_r}) = n\}$$
$$\Theta^{\mathbf{a}}(r, n, m) = \{\theta \in \Theta \,;\, \theta(I_{a_r}) = n,\ \theta(\bar{I}_{a_r+\frac{1}{m}} \backslash I_{a_r}) = 0\}.$$

Here $\bar{I}_{a_r+\frac{1}{m}}$ is the closure of $I_{a_r+\frac{1}{m}}$, where $I_r = (-r, r)^d \cap \mathsf{E}$ as before. We remark $\Theta^{\mathbf{a}}(r, n, m)$ is an open set in Θ. We set

$$\text{(4.4)} \qquad \mathsf{A}^{\mathbf{a}}_{\epsilon}(r, n, m) = \{\theta = \sum_i \delta_{x_i} \,;\, \theta \in \Theta^{\mathbf{a}}(r, n, m) \text{ and } \theta \text{ satisfy}$$
$$|x_i - x_j| < \epsilon \text{ and } x_i, x_j \in I_{a_r-1} \text{ for some } i \neq j\}.$$

It is clear that $\mathsf{A}^{\mathbf{a}}_{\epsilon}(r, n, m)$ is an open set in Θ.

Lemma 4.1. *Assume that for all* $\mathbf{a} \in \mathbf{A}$ *and* $(r, n, m) \in \mathbb{I}$

$$\text{(4.5)} \qquad \inf_{0<\epsilon<1/2m} \mathrm{Cap}(\mathsf{A}^{\mathbf{a}}_{\epsilon}(r, n, m)) = 0.$$

Then (2.9) *holds.*

Proof. Let

$$\mathsf{A}^{\mathbf{a}}(r, n, m) = \{\theta = \sum_i \delta_{x_i} \,;\, \theta \in \Theta^{\mathbf{a}}(r, n, m) \text{ and } \theta \text{ satisfy}$$
$$x_i = x_j \text{ and } x_i, x_j \in I_{a_r-1} \text{ for some } i \neq j\}.$$

Then $\mathsf{A} = \bigcup_{\mathbf{a}\in\mathbf{A}} \bigcup_{(r,n,m)\in\mathbb{I}} \mathsf{A}^{\mathbf{a}}(r, n, m)$. Since $\mathbf{A}$ and $\mathbb{I}$ are countable sets and the capacity is sub additive, (2.9) follows from

$$\text{(4.6)} \qquad \mathrm{Cap}(\mathsf{A}^{\mathbf{a}}(r, n, m)) = 0 \quad \text{for all } \mathbf{a} \in \mathbf{A},\ (r, n, m) \in \mathbb{I}.$$

Note that $\mathsf{A}^{\mathbf{a}}(r, n, m) \subset \mathsf{A}^{\mathbf{a}}_{\epsilon}(r, n, m)$. So (4.5) implies (4.6) by the monotonicity of the capacity, which deduces (2.9). Q.E.D.

Now fix $\mathbf{a} \in \mathbf{A}$ and $(r, n, m) \in \mathbb{I}$ and suppress them from the notion. Set

$$\text{(4.7)} \quad \mathsf{A}^{-}_{\epsilon} = \mathsf{A}^{\mathbf{a}}_{\epsilon/2}(r, n, m), \quad \mathsf{A}_{\epsilon} = \mathsf{A}^{\mathbf{a}}_{\epsilon}(r, n, m), \quad \mathsf{A}^{+}_{\epsilon} = \mathsf{A}^{\mathbf{a}}_{1+\epsilon}(r, n, m).$$

and let $h_\epsilon : \mathbb{R} \to \mathbb{R}$ $(0 < \epsilon < 1/m < 1)$ such that

$$\text{(4.8)} \qquad h_\epsilon(t) = \begin{cases} 2 & (|t| \leq \epsilon) \\ 2 \log |t| / \log \epsilon & (\epsilon \leq |t| \leq 1) \\ 0 & (1 \leq |t|). \end{cases}$$

We define $\mathfrak{h}_\epsilon : \Theta \to \mathbb{R}$ by $\mathfrak{h}_\epsilon(\theta) = 0$ for $\theta \notin \Theta^{\mathsf{a}}(r,n,m)$ and

$$\mathfrak{h}_\epsilon(\theta) = \sum_{x_i,\, x_j \in I_{a_r - 1},\ j \neq i} h_\epsilon(x_i - x_j) \quad \text{for } \theta \in \Theta^{\mathsf{a}}(r,n,m).$$

Here we set $\mathfrak{h}_\epsilon(\theta) = 0$ if the summand is empty. Let $\mathfrak{g}_\epsilon = \mathfrak{J}_{a_r + \frac{1}{m}, \epsilon/4} \mathfrak{h}_\epsilon$. Here $\mathfrak{J}_{a_r + \frac{1}{m}, \epsilon/4}$ is the mollifier introduced in (3.3).

Lemma 4.2. *For $0 < \epsilon < 1/2m$, $\mathfrak{g}_\epsilon$ satisfy the following:*

$$\mathfrak{g}_\epsilon \in \mathcal{D}_\infty \tag{4.9}$$

$$\mathfrak{g}_\epsilon(\theta) \geq 1 \qquad \textit{for all } \theta \in \mathsf{A}_\epsilon \tag{4.10}$$

$$0 \leq \mathfrak{g}_\epsilon(\theta) \leq n(n+1) \qquad \textit{for all } \theta \in \Theta \tag{4.11}$$

$$\mathfrak{g}_\epsilon(\theta) = 0 \qquad \textit{for all } \theta \notin \mathsf{A}_\epsilon^+ \tag{4.12}$$

$$\mathbb{D}[\mathfrak{g}_\epsilon, \mathfrak{g}_\epsilon](\theta) = 0 \qquad \textit{for all } \theta \notin \mathsf{A}_\epsilon^+ \backslash \mathsf{A}_\epsilon^- \tag{4.13}$$

$$\mathbb{D}[\mathfrak{g}_\epsilon, \mathfrak{g}_\epsilon](\theta) \leq \frac{c_3}{(\log \epsilon \ \min |x_i - x_j|)^2} \qquad \textit{for all } \theta \in \mathsf{A}_\epsilon^+ \backslash \mathsf{A}_\epsilon^- . \tag{4.14}$$

Here $\theta = \sum \delta_{x_k}$ and the minimum in (4.14) *is taken over x_i, x_j such that*

$$x_i,\ x_j \in I_{a_r - 1},\ \epsilon/2 \leq |x_i - x_j| \leq 1 + \epsilon,$$

and $c_3 \geq 0$ is a constant independent of ϵ (c_3 depends on (r,n,m)).

Proof. (4.9) follows from [5, Lemma 2.4 (1)]. Other statements are clear from a direct calculation. Q.E.D.

Permutation invariant functions $\sigma_r^n : I_r^n \to \mathbb{R}^+$ are called density functions of μ if, for all bounded $\sigma[\pi_r]$-measurable functions $\mathfrak{f}$,

$$\int_{\Theta_r^n} \mathfrak{f}\, d\mu = \frac{1}{n!} \int_{I_r^n} f_r^n \sigma_r^n dx. \tag{4.15}$$

Here $f_r^n : I_r^n \to \mathbb{R}$ is the permutation invariant function such that $f_r^n(\mathbf{x}(\theta)) = \mathfrak{f}(\theta)$ for $\theta \in \Theta_r^n$, where $\mathbf{x}$ is an I_r^n-coordinate. We recall relations between a correlation function and a density function ([10]):

$$\rho_n = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{I_r^k} \sigma_r^{n+k}(x_1, \dots, x_{n+k}) dx_{n+1} \cdots dx_{n+k} \tag{4.16}$$

$$\sigma_r^n = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{I_r^k} \rho_{n+k}(x_1, \dots, x_{n+k}) dx_{n+1} \cdots dx_{n+k} \tag{4.17}$$

The first summand in the right hand side of (4.16) is taken to be σ_r^n. It is clear that

$$0 \le \sigma_r^n(x_1, \dots, x_n) \le \rho_n(x_1, \dots, x_n) \tag{4.18}$$

Lemma 4.3. *There exists a constant* c_4 *depending on* r, n *such that*

$$\sigma_r^n(x_1, \dots, x_n) \le c_4 \min_{i \ne j} |x_i - x_j| \quad \textit{for all } (x_1, \dots, x_n) \in I_r^n \tag{4.19}$$

Proof. By (2.3) and the kernel K is locally Lipschitz continuous, we see ρ_n is bounded and Lipschitz continuous on I_r^n. In addition, by using (2.3) we see $\rho_n = 0$ if $x_i = x_j$ for some $i \ne j$. Hence by using (2.3) again there exists a constant c_5 depending on n, r such that

$$\rho_n(x_1, \dots, x_n) \le c_5 \min_{i \ne j} |x_i - x_j| \quad \text{for all } (x_1, \dots, x_n) \in I_r^n. \tag{4.20}$$

(4.19) follows from this and (4.18) immediately. Q.E.D.

Lemma 4.4. (4.5) *holds true.*

Proof. By the definition of the capacity, $\mathfrak{g}_\epsilon \in \mathcal{D}_\infty$, (4.9) and (4.10) we obtain

$$\mathrm{Cap}(\mathsf{A}_\epsilon) \le \mathcal{E}(\mathfrak{g}_\epsilon, \mathfrak{g}_\epsilon) + (\mathfrak{g}_\epsilon, \mathfrak{g}_\epsilon)_{L^2(\Theta, \mu)} \tag{4.21}$$

So we will estimate the right hand side. We now see by (4.13)

$$\begin{aligned} \mathcal{E}(\mathfrak{g}_\epsilon, \mathfrak{g}_\epsilon) &= \int_{\mathsf{A}_\epsilon^+ \backslash \mathsf{A}_\epsilon^-} \mathbb{D}[\mathfrak{g}_\epsilon, \mathfrak{g}_\epsilon] d\mu \\ &= \frac{1}{n!} \int_{B_\epsilon} \{ \frac{1}{2} \sum_{i=1}^n \frac{\partial g_\epsilon^n}{\partial x_i} \frac{\partial g_\epsilon^n}{\partial x_i} \} \sigma_{a_r + \frac{1}{m}}^n dx_1 \cdots dx_n \\ &=: \mathbf{I}_\epsilon. \end{aligned} \tag{4.22}$$

Here g_ϵ^n is defined by (3.2) for $\mathfrak{g}_\epsilon$, and $B_\epsilon = \varpi_{a_r + \frac{1}{m}}^{-1}(\pi_{a_r + \frac{1}{m}}(\mathsf{A}_\epsilon^+ \backslash \mathsf{A}_\epsilon^-))$, where $\varpi : I_{a_r + \frac{1}{m}}^n \to \Theta$ is the map such that $\varpi((x_1, \dots, x_n)) = \sum \delta_{x_i}$.

By using (4.14) and Lemma 4.3 for $a_r + \frac{1}{m}$ it is not difficult to see there exists a constant c_6 independent of ϵ satisfying the following:

$$\mathbf{I}_\epsilon \le \frac{c_6}{|\log \epsilon|}.$$

This implies $\lim_{\epsilon \to 0} \mathcal{E}(\mathfrak{g}_\epsilon, \mathfrak{g}_\epsilon) = 0$. By (4.11) and (4.12) we have

$$(\mathfrak{g}_\epsilon, \mathfrak{g}_\epsilon)_{L^2(\Theta, \mu)} = \int_{\mathsf{A}_\epsilon^+} \mathfrak{g}_\epsilon^2 d\mu \le n^2 (n+1)^2 \, \mu(\mathsf{A}_\epsilon^+) \to 0 \quad \text{as } \epsilon \downarrow 0.$$

Combining these with (4.21) we complete the proof of Lemma 4.4. Q.E.D.

Proof of Theorem 2.1. Theorem 2.1 follows from Lemma 4.1 and Lemma 4.4 immediately. Q.E.D.

§5. Proof of Proposition 2.9

Lemma 5.1. *Let μ be a probability measure on $(\Theta, \mathcal{B}(\Theta))$ such that $\mu(\{\theta(\mathsf{E}) < \infty\}) = 1$ and that density functions $\{\sigma_{\mathsf{E}}^{n}\}$ on E of μ are continuous. Then $(\mathcal{E}, \mathcal{D}_{\infty})$ is closable on $L^2(\Theta, \mu)$.*

Proof. Let $\Theta^n = \{\theta \in \Theta\,;\, \theta(\mathsf{E}) = n\}$ and set

$$\mathcal{E}^n(\mathfrak{f}, \mathfrak{g}) = \sum_{k=1}^{n} \int_{\Theta^k} \mathbb{D}[\mathfrak{f}, \mathfrak{g}] d\mu.$$

By assumption $\sum_{n=0}^{\infty} \mu(\Theta^n) = 1$, from which we deduce $(\mathcal{E}, \mathcal{D}_{\infty})$ is the increasing limit of $\{(\mathcal{E}^n, \mathcal{D}_{\infty})\}$. Since density functions are continuous, each $(\mathcal{E}^n, \mathcal{D}_{\infty})$ is closable on $L^2(\Theta, \mu)$. So its increasing limit $(\mathcal{E}, \mathcal{D}_{\infty})$ is also closable on $L^2(\Theta, \mu)$. Q.E.D.

Lemma 5.2. *Let μ be a determinantal random point field on E with continuous kernel K. Assume K is of trace class. Then their density functions σ^n on E are continuous.*

Proof. For the sake of simplicity we only prove the case $K < 1$, where K is the operator generated by the integral kernel K. The general case is proved similarly by using a device in [10, 935 p.].

Let λ_i denote the i-th eigenvalue of K and φ_i its normalized eigenfunction. Then since K is of trace class we have

$$\mathsf{K}(x, y) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(x) \overline{\varphi_i(y)}. \tag{5.1}$$

It is known that (see [10, 934 p.])

$$\sigma^n(x_1, \dots, x_n) = \det(\mathrm{Id} - K) \cdot \det(L(x_i, x_j))_{1 \le i,j \le n}, \tag{5.2}$$

where $\det(\mathrm{Id} - K) = \prod_{i=1}^{\infty}(1 - \lambda_i)$ and

$$L(x, y) = \sum_{i=1}^{\infty} \frac{\lambda_i}{1 - \lambda_i} \varphi_i(x) \overline{\varphi_i(y)}. \tag{5.3}$$

Since $\mathsf{K}(x,y)$ is continuous, eigenfunctions $\varphi_i(x)$ are also continuous. It is well known that the right hand side of (5.1) converges uniformly. By $0 \le K < 1$ we have $0 \le \lambda_i \le \lambda_1 < 1$. Collecting these implies the right hand side of (5.3) converges uniformly. Hence $L(x,y)$ is continuous in (x,y). This combined with (5.2) completes the proof. Q.E.D.

Proof of Proposition 2.9. Since K is of trace class, the associated determinantal random point field μ satisfies $\mu(\{\theta(\mathsf{E}) < \infty\}) = 1$. By Lemma 5.2 we have density functions σ_{E}^n are continuous. So Proposition 2.9 follows from Lemma 5.1. Q.E.D.

We now turn to the proof of Theorem 2.4. So as in the statement in Theorem 2.4 let $\mathsf{E} = \mathbb{R}$ and $\mathsf{K}(x,y) = \mathsf{m}(x)\mathsf{k}(x-y)\mathsf{m}(y)$, where $\mathsf{k}:\mathbb{R}\to\mathbb{R}$ is a non-negative, continuous *even* function that is convex in $[0,\infty)$ such that $\mathsf{k}(0) \le 1$, and $\mathsf{m}:\mathbb{R}\to\mathbb{R}$ is nonnegative continuous and $\int_{\mathbb{R}} \mathsf{m}(t)dt < \infty$ and $\mathsf{m}(x) \le 1$ for all x and $0 < \mathsf{m}(x)$ for some x. We assume k satisfies (2.12).

Lemma 5.3. *There exists an interval I in E such that*

$$\sigma_I^2(x, x+t) \ge c_7 t^\alpha \quad \text{for all } |t| \le 1 \text{ and } x, x+t \in I, \tag{5.4}$$

where c_7 is a positive constant and σ_I^2 is the 2-density function of μ on I.

Proof. By assumption we see $\inf_{x\in I} \mathsf{m}(x) > 0$ for some open bounded, nonempty interval I in E. By (4.17) we have

$$\sigma_I^2(x, x+t) \ge \rho_2(x, x+t) - \int_I \rho_3(x, x+t, z)dz \tag{5.5}$$

By (2.3) and (2.12) there exist positive constants c_8 and c_9 such that

$$\begin{aligned} c_8 t^\alpha &\le \rho_2(x, x+t) && \text{for all } |t| \le 1 \text{ and } x, x+t \in I \\ \rho_3(x, x+t, z) &\le c_9 t^\alpha && \text{for all } |t| \le 1 \text{ and } x, x+t, z \in I. \end{aligned} \tag{5.6}$$

Hence by taking I so small we deduce (5.4) from (5.5) and (5.6). Q.E.D.

Proof of Theorem 2.4. The closability follows from Proposition 2.9. So it only remains to prove (2.13).

Let $(\mathcal{E}^2, \mathcal{D}^2)$ and $(\mathcal{E}, \mathcal{D})$ denote closures of $(\mathcal{E}^2, \mathcal{D}_\infty)$ and $(\mathcal{E}, \mathcal{D}_\infty)$ on $L^2(\Theta, \mu)$, respectively. Then

$$(\mathcal{E}^2, \mathcal{D}^2) \le (\mathcal{E}, \mathcal{D}) \tag{5.7}$$

Let I be as in Lemma 5.3. Let $\{I_r\}_{r=1,\dots}$ be an increasing sequence of open intervals in E such that $I_1 = I$ and $\cup_r I_r = \mathsf{E}$. Let

$$\mathcal{E}_r^2(\mathfrak{f}, \mathfrak{g}) = \int_{\Theta^2} \sum_{x_i \in I_r} \frac{1}{2} \frac{\partial f(\mathbf{x})}{\partial x_i} \cdot \frac{\partial g(\mathbf{x})}{\partial x_i} d\mu \tag{5.8}$$

Here we set $\mathbf{x} = (x_1, \dots)$, f and $\mathfrak{f}$ similarly as in (2.8). Then since density functions on I_r are continuous, we see $(\mathcal{E}_r^2, \mathcal{D}_\infty)$ are closable on $L^2(\Theta, \mu)$. So we denote its closure by $(\mathcal{E}_r^2, \mathcal{D}_r^2)$. It is clear that $\{(\mathcal{E}_r^2, \mathcal{D}_r^2)\}$ is increasing in the sense that $\mathcal{D}_r^2 \supset \mathcal{D}_{r+1}^2$ and $\mathcal{E}_r^2(\mathfrak{f}, \mathfrak{f}) \le \mathcal{E}_{r+1}^2(\mathfrak{f}, \mathfrak{f})$ for all $\mathfrak{f} \in \mathcal{D}_{r+1}$. So we denote its limit by $(\check{\mathcal{E}}^2, \check{\mathcal{D}}^2)$. It is known ([5, Remark (3) after Theorem 3]) that

$$(\check{\mathcal{E}}^2, \check{\mathcal{D}}^2) \le (\mathcal{E}^2, \mathcal{D}^2). \tag{5.9}$$

By (5.7), (5.9) and the definition of $\{(\mathcal{E}_r^2, \mathcal{D}_r^2)\}$ we conclude $(\mathcal{E}_1^2, \mathcal{D}_1^2) \le (\mathcal{E}, \mathcal{D})$, which implies

$$\mathrm{Cap}_1^2 \le \mathrm{Cap}, \tag{5.10}$$

where Cap_1^2 and Cap denote capacities of $(\mathcal{E}_1^2, \mathcal{D}_1^2)$ and $(\mathcal{E}, \mathcal{D})$, respectively. Let $\mathsf{B} = \Theta^2 \cap \{\theta(\{x\}) = 2 \text{ for some } x \in I\}$. Then by (2.11) and (5.4) together with a standard argument (see [2, Example 2.2.4] for example) we obtain

$$0 < \mathrm{Cap}_1^2(\mathsf{B}). \tag{5.11}$$

Since $\mathsf{B} \subset \mathsf{A}$, we deduce $0 < \mathrm{Cap}(\mathsf{A})$ from (5.10) and (5.11), which implies (2.13). Q.E.D.

§6. A construction of infinite volume dynamics

In this section we prove Theorem 2.5. We first prove the closability of pre-Dirichlet forms in finite volume.

Lemma 6.1. *Let $I_r = (-r, r) \cap \mathsf{E}$ and σ_r^n denote the n-density function on I_r. Then σ_r^n is continuous.*

Proof. Let $M = \sup_{x,y \in I_r} |\mathsf{K}(x, y)|$. Then $M < \infty$ because K is continuous. Let $\mathbf{x}_i = (\mathsf{K}(x_i, x_1), \mathsf{K}(x_i, x_2), \dots, \mathsf{K}(x_i, x_n))$ and $\|\mathbf{x}_i\|$ denote its Euclidean norm. Then by (2.3) we see

$$|\rho_n| \le \prod_{i=1}^{n} \|\mathbf{x}_i\| \le \{\sqrt{n} M\}^n. \tag{6.1}$$

By using Stirling's formula and (6.1) we have for some positive constant c_{10} independent of k and M such that

$$\tag{6.2} \begin{aligned} &|\frac{(-1)^k}{k!}\int_{I_r^k}\rho_{n+k}(x_1,\dots,x_{n+k})dx_{n+1}\cdots dx_{n+k}| \\ &\qquad\le c_{10}^k k^{-k+1/2}(n+k)^{(n+k)/2}M^{n+k}. \end{aligned}$$

This implies for each n the series in the right hand side of (4.17) converges uniformly in $(x_1,\dots,x_n)$. So σ_r^n is the limit of continuous functions in the uniform norm, which completes the proof. Q.E.D.

Lemma 6.2. *$(\mathcal{E},\mathcal{D}_{\infty,r})$ are closable on $L^2(\Theta,\mu)$.*

Proof. Let $I_r = \{x\in\mathsf{E};|x|<r\}$ and $\Theta_r^n = \{\theta(I_r)=n\}$. Let $\mathcal{E}_r^n(\mathfrak{f},\mathfrak{g}) = \int_{\Theta_r^n}\mathbb{D}[\mathfrak{f},\mathfrak{g}]d\mu$. Then it is enough to show that $(\mathcal{E}_r^n,\mathcal{D}_{\infty,r})$ are closable on $L^2(\Theta,\mu)$ for all n.

Since $\mathfrak{f}$ is $\sigma[\pi_r]$-measurable, we have ($\mathbf{x}=(x_1,\dots,x_n)$)

$$\mathcal{E}_r^n(\mathfrak{f},\mathfrak{g}) = \frac{1}{n!}\int_{I_r^n}\sum_{i=1}^n\frac{1}{2}\frac{\partial f_r^n(\mathbf{x})}{\partial x_i}\cdot\frac{\partial g_r^n(\mathbf{x})}{\partial x_i}\sigma_r^n(\mathbf{x})d\mathbf{x},$$

where f_r^n and g_r^n are defined similarly as after (4.15). Then since σ_r^n is continuous, we see $(\mathcal{E}_r^n,\mathcal{D}_{\infty,r})$ is closable. Q.E.D.

Proof of Theorem 2.5. By Lemma 6.2 we see the assumption (A.1*) in [5] is satisfied. (A.2) in [5] is also satisfied by the construction of determinantal random point fields. So one can apply results in [5] (Theorem 1, Corollary 1, Lemma 2.1 (3) in [5]) to the present situation. Although in Theorem 1 in [5] we treat $(\mathcal{E},\mathcal{D})$, it is not difficult to see that the same conclusion also holds for $(\mathcal{E}^{\mathrm{reg}},\mathcal{D}^{\mathrm{reg}})$, which completes the proof. Q.E.D.

§7. Gibbsian case

In this section we consider the case μ is a canonical Gibbs measure with interaction potential Φ, whose n-density functions for bounded sets are bounded, and 1-correlation function is locally integrable. If Φ is super stable and regular in the sense of Ruelle, then probability measures satisfying these exist. In addition, it is known in [5] that, if Φ is upper semi-continuous (or more generally Φ is a measurable function dominated from above by a upper semi-continuous potential satisfying certain integrable conditions (see [7])), then the form $(\mathcal{E},\mathcal{D})$ on $L^2(\Theta,\mu)$ is closable. We remark these assumptions are quite mild. In [5] and [7] only *grand* canonical Gibbs measures with *pair* interaction potential are treated; it is easy to generalize the results in [5] and [7] to the present situation.

Proposition 7.1. *Let* μ *be as above. Assume* $d \geq 2$. *Then* $\mathrm{Cap}(\mathsf{A}) = 0$ *and no collision* (2.10) *occurs.*

Proof. The proof is quite similar to the one of Theorem 2.1. Let $\mathbf{I}_\epsilon$ be as in (4.22). It only remains to show $\lim_{\epsilon \to 0} \mathbf{I}_\epsilon = 0$.

We divide the case into two parts: (1) $d = 2$ and (2) $3 \leq d$. Assume (1). We can prove $\lim \mathbf{I}_\epsilon = 0$ similarly as before. In the case of (2) the proof is more simple. Indeed, we change definitions of A_ϵ^+ in (4.7) and h_ϵ in (4.8) as follows: $\mathsf{A}_\epsilon^+ = \mathsf{A}_{4\epsilon}^{\mathbf{a}}(r, n, m)$

$$h_\epsilon(t) = \begin{cases} 2 & (|t| \leq \epsilon) \\ -(2/\epsilon)|t| + 4 & (\epsilon \leq |t| \leq 2\epsilon) \\ 0 & (2\epsilon \leq |t|). \end{cases} \tag{7.1}$$

Then we can easily see $\lim \mathbf{I}_\epsilon = 0$. Q.E.D.

Remark 7.2. (1) This result was announced and used in [6, Lemma 1.4]. Since this result was so different from other parts of the paper [6], we did not give a detail of the proof there.

(2) In [8] a related result was obtained. In their frame work the choice of the domain of Dirichlet forms may be not same as ours. Indeed, their domains are smaller than or equal to ours (we do not know they are same or not). So one may deduce Proposition 7.1 from their result.

References

[1] Donoghue, W., *Distributions and Fourier Transforms*, Academic Press (1969).

[2] Fukushima, M., Oshima, Y., Takeda, M., *Dirichlet forms and symmetric Markov processes*, Walter de Gruyter (1994).

[3] Lenard A., *States of classical statistical mechanical systems of infinitely many particles. I*, Arch. Rational Mech. Anal. **59** (1975) 219-235.

[4] Lenard A., *States of classical statistical mechanical systems of infinitely many particles. I*, Arch. Rational Mech. Anal. **59** (1975) 240-256.

[5] Osada, H., *Dirichlet form approach to infinitely dimensional Wiener processes with singular interactions*, Commun. Math. Physic. (1996), 117-131.

[6] Osada, H. *An invariance principle for Markov processes and Brownian particles with singular interaction*, Ann. Inst. Henri Poincarë, **34**, n° 2 (1998), 217-248.

[7] Osada, H. *Interacting Brownian motions with measurable potentials.* Proc. Japan Acad. Ser. A Math. Sci. **74** (1998), no. 1, 10–12. v

[8] Röckner, M., Schmuland, B. *A support property for infinite-dimensional interacting diffusion processes*, C. R. Acad. Sci. Paris Sér. I Math. **326** (1998), no. 3, 359–364.

[9] Shirai T., Takahashi Y., *Random point fields associated with certain Fredholm determinant I* (preprint).

[10] Soshnikov, A., *Determinantal random point fields*, Russian Math. Surveys **55**:5 923-975.

[11] Spohn, H., *Interacting Brownian particles: a study of Dyson's model,* In: Hydrodynamic Behavior and Interacting Particle Systems, ed. by G.C. Papanicolaou, IMA Volumes in Mathematics **9** , Springer-Verlag (1987) 151-179.

Graduate School of Mathematics
Nagoya University
Chikusa-ku
Nagoya 464-8602
Japan

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 345–354

Random Point Fields Associated with Fermion, Boson and Other Statistics

Tomoyuki Shirai and Yoichiro Takahashi

Abstract.

We show that the grand canonical ensembles of ideal gas under Fermi, Boson and other statistics give simple examples of the random point fields studied in the previous papers [13, 14, 15]. Also we present two classes of nonsymmetric integral operators for which such random point fields do exist.

§1. Introduction

In the present paper we are concerned with the nonnegativity problem of certain generalization of determinants and permanents denoted by $\det_\alpha$ in our previous paper [14]. The problem is almost equivalent to the existence problem of those random point fields or point processes whose Laplace transforms are given as the Fredholm determinants to the power $-1/\alpha$ of certain integral operators. They are also closely related to the Fermi, Boson and other statistics in quantum statistical mechanics of the ideal gas. Indeed, the grand canonical ensembles under these statistics (if any) are special examples of our random point fields.

Definition 1.1. Let α be a real number and $A = (a_{ij})$ be a square matrix of size n. Given a permutation σ, denote the number of cycles by $\nu(\sigma)$. The following quantity is called the α-permanent of A in [18]:

$$\det_\alpha(A) = \sum_{\sigma \in S_n} \alpha^{n-\nu(\sigma)} \prod_{i=1}^{n} a_{i\sigma(i)}, \tag{1.1}$$

where S_n is the symmetric group.

Received December 20, 2002.
Revised March 4, 2003.
The first and the second authors are partially supported by JSPS under the Grant-in-Aid for Scientific Research No.13740057 and No.13340030, respectively.

For instance,

$$\begin{aligned}
\det_\alpha \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &= a_{11}a_{22} + \alpha a_{12}a_{21}, \\
\det_\alpha \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} &= a_{11}a_{22}a_{33} \\
&\quad +\alpha(a_{12}a_{21}a_{33} + a_{13}a_{31}a_{22} + a_{23}a_{32}a_{11}) \\
&\quad +\alpha^2(a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21}).
\end{aligned}$$

It is immediate to see

$$\det_{-1}(A) = \det(A), \quad \det_1(A) = \mathrm{per}(A) \tag{1.2}$$

and

$$\det_0(A) = a_{11}a_{22}\ldots a_{nn}. \tag{1.3}$$

In [14] we proved the following (though not stated directly there):

Theorem 1.2. *There holds the inequality* $\det_\alpha(A) \geq 0$ *if* α *and* A *satisfy one of the following three conditions* (A^-), (A) *and* (B).
(A^-) $\alpha \in \{-1/m \mid m = 1, 2, \ldots\}$ *and* A *is nonnegative definite.*
(A) $\alpha \in \{2/m \mid m = 1, 2, \ldots\}$ *and* A *is nonnegative definite.*
(B) $\alpha \in (0, \infty)$ *and* A *is a nonnegative matrix.*

The sufficiency of (B) is obvious from Definition 1.1. We give an alternative proof of the case (A) below in Section 3. In [14] we also proposed the following.

Conjecture 1.3. *If* $0 \leq \alpha \leq 2$, $\det_\alpha(A) \geq 0$ *for nonnegative definite matrix* A *of any size.*

The random point fields mentioned above are defined as follows. For simplicity, let R be a locally compact separable metric space and fix a nonnegative Radon measure λ on R. We define the locally finite configuration space Q over R as the set of nonnegative integer-valued Radon measures ξ on R and say that a function f is a test function if it is nonnegative and its support is compact. For a test function f and a (locally finite) configuration $\xi = \sum \delta_{x_i} \in Q$ we denote

$$e_f(\xi) = \exp\left(-\int_R f(x)\xi(dx)\right).$$

Now let α be a real number and K be a locally trace class integral operator on $L^2(R, \lambda)$, i.e., the restriction K_Λ of K to any compact set Λ

is of trace class. The Fredholm determinant $\mathrm{Det}(I+T)$ for a trace class operator T is defined as $\prod_{i=1}^{\infty}(1+\lambda_i)$, where $\lambda_i, i \geq 1$ are the eigenvalues (counting the multiplicity) of T.

Definition 1.4. A probability measure μ on Q will be called a random point field associated with (α, K) if it satisfies for any test function f

$$\int_Q \mu(d\xi)e_f(\xi) = \mathrm{Det}(I + \alpha\varphi K)^{-1/\alpha},$$

where $\varphi = 1 - e^{-f}$. In particular, μ is called a fermion point process and a boson point process in [8, 9] according as $\alpha = -1$ and $\alpha = 1$. Some people use the terminology "determinantal processes" for fermion processes (cf. [7, 16]).

In [13, 14] we essentially proved the following:

Theorem 1.5. *Assume that the kernel $K(x,y)$ is continuous and, in addition, that the operator norm K is so small that $\|\alpha K\| < 1$ when $\alpha < 0$. Set $J = (I + \alpha K)^{-1}K$.*
(i) The random point field μ associated with (α, K) exists and is unique if $\det_\alpha(J(x_i, x_j))_{i,j=1}^n$ is nonnegative for any n and any $x_1, \dots, x_n$.
(ii) If the random point field μ associated with (α, K) exists, then both $\det_\alpha(J(x_i, x_j))_{i,j=1}^n$ and $\det_\alpha(K(x_i, x_j))_{i,j=1}^n$ are nonnegative for any n and any $x_1, \dots, x_n$.

Combining Theorems 1.2 and 1.5 we showed

Theorem 1.6. ([14]) *The random point field μ associated with (α, K) exists and is unique if (α, K) satisfies one of the following conditions:*
(A^-) *$\alpha \in \{-1/m \mid m = 1, 2, \dots\}$, $\|\alpha K\| < 1$ and K is nonnegative definite.*
(A) *$\alpha \in \{2/m \mid m = 1, 2, \dots\}$ and K is nonnegative definite.*
(B) *$\alpha \in (0, \infty)$ and the kernel $J(x,y)$ defined by $J = (I + \alpha K)^{-1}K$ is nonnegative.*
Here $\|\cdot\|$ stands for the operator norm.

The case where $\|\alpha K\| = 1$ with $\alpha < 0$ can also be treated in [14, 15] although the operator J becomes unbounded.

In [13, 14, 15] we did not study the case where K is a nonsymmetric operator. But we can show the following:

Theorem 1.7. *Let $R = \mathbb{R}^1$, λ be the Lebesgue measure and $T_t, t \geq 0$ be the transition semigroup of a one dimensional diffusion process or* *let $R = \mathbb{N}$, λ be a counting measure and $T_t, t \geq 0$ be the transition semigroup of a birth and death process. Then the random point field μ associated with $(-1, T_t)$ exists and is unique.*

We would like to emphasize that T_t can be nonsymmetric in the above Theorem 1.7 (cf. [3, 12]).

We had better to mention here on a rather classical theorem of Karlin and McGregor [4, 5] from which Theorem 1.7 above follows immediately by using Theorem 1.5. Recall that a matrix A is called totally positive if all of its square minors are nonnegative and that a kernel function $K(x, y)$ is called totally positive if $\det(K(x_i, y_j))_{i,j=1}^n$ is nonnegative for any n and any $x_1, \ldots, x_n, y_1, \ldots, y_n$.

Theorem 1.8 (Karlin-McGregor). *Let $p(t, x, y)$ be the transition probability density of a one dimensional diffusion process or a birth and death process. Then for each $t > 0$ the kernel function $p(t, x, y)$ is totally positive.*

The proof of Karlin and McGregor is simple and is based only on the two facts: the strong Markov property and the one dimensionality. So it may also be applied to discrete time nearest neighbor random walks on $\mathbb{Z}^1$ under suitable settings (cf., [6]). Notice that $p(t, x, y)$ is not necessarily symmetric in x and y.

§2. Fermi statistics, Boson statistics and other statistics

Consider quantum statistical mechanics of ideal gas. If the energy levels are E_i, then the grand canonical partition function Z is given in text books, for instance [2] as

$$Z = \prod(1 + ze^{-\beta E_i}) \text{ and } Z = \prod(1 - ze^{-\beta E_i})^{-1}$$

under fermi statistics and boson statistics, respectively. If we introduce the trace class diagonal operator

$$J = \operatorname{diag}(ze^{-\beta E_1}, ze^{-\beta E_2}, \ldots)$$

on $\ell^2(\{1, 2, \ldots\})$ and a parameter α, then they can be written as

$$Z = \operatorname{Det}(I - \alpha J)^{-1/\alpha}$$

with $\alpha = -1$ for fermi statistics and $\alpha = 1$ for boson statistics.

For general values of α we might consider the α-statistics. If the underlying space R consists of a single point, then such statistics exist for

$\alpha \in \{1/m \mid m = 1, 2, \dots\}$ or $\alpha \in (0, \infty)$ and the corresponding distributions are called negative binomial or generalized binomial, respectively. There are some attempts to generalize these distributions to spaces consisting of two or more points [1]. Anyway we need some restriction on α in order to consider α-statistics.

The grand canonical ensemble, say μ, under α-statistics is, if any, described in terms of its Laplace transform as

$$\int_Q \mu(d\xi) e_f(\xi) = \frac{\operatorname{Det}(I - \alpha e^{-f} J)^{-1/\alpha}}{\operatorname{Det}(I - \alpha J)^{-1/\alpha}}. \tag{2.1}$$

For fermion and boson cases it is immediate to see, by setting f to be a linear combination of indicator functions of intervals and then by expanding the (infinite) product into the (infinite) sum, that one can obtain the micro canonical ensembles.

Thus, if we introduce an operator $K = (I - \alpha J)^{-1} J$, then we obtain

$$\int_Q \mu(d\xi) e_f(\xi) = \operatorname{Det}(I + \alpha \varphi K)^{-1/\alpha}.$$

Consequently, the grand canonical ensemble μ is the random point field over the set $\{1, 2, \dots\}$ associated with (α, K) in our terminology. By Theorem 1.6 we may consider the α-statisitcs as a real object at least for $\alpha \in \{-1/m \mid m = 1, 2, \dots\} \cup [0, \infty)$ since J is now nonnegative definite and has nonnegative entries provided that $\|\alpha J\| < 1$ when α is positive.

Moreover, the operator J may not be of trace class nor diagonal. Indeed, J can be a locally trace class operator and one can consider the infinite volume limit of grand canonical ensembles as is shown by the following theorem which is a restatement of results in [14]:

Theorem 2.1. *Let α be a real number and J be a locally trace class operator. The random point field μ satisfying (2.1) exists if (α, J) satisfies one of the following conditions.*
(A$^-$) $\alpha \in \{-1/m \mid m = 1, 2, \dots\}$ *and J is nonnegative definite.*
(A) $\alpha \in \{2/m \mid m = 1, 2, \dots\}$ *and J is nonnegative definite with* $\|\alpha J\| < 1$.
(B) $\alpha \in (0, \infty)$ *and the kernel $J(x, y)$ is nonnegative with* $\|\alpha J\| < 1$.

The points of the proof of Theorem 2.1 are to introduce restrictions J_Λ of J to compact subsets and to show that the operators $(I - \alpha J_\Lambda)^{-1} J_\Lambda$ converge to $K := (I - \alpha J)^{-1} J$ as $\Lambda \to R$. Then it turns out that the grand canonical ensemble over the compact subsets Λ converges to the limiting grand canonical ensemble which is nothing but our random point field associated with (α, K).

Roughly to say, the random point fields associated with powers of Fredholm determinants are Gibbs random fields as the argument above suggests. If R is discrete, say d-dimensional square lattice $\mathbb{Z}^d$, then we obtained the following rigorous result in the fermion case($\alpha = -1$). Since the fermion point fields have no multiple points, we may safely identify the configuration space Q with the power set of R or $\{0,1\}^R$.

Theorem 2.2. ([15]) *Let $R = \mathbb{Z}^d$ and λ be the counting measure. Assume the the operator $K : \ell^2(R) \to \ell^2(R)$ is positive definite with $\|K\| < 1$. Set $J = (I-K)^{-1}K$ and write its restriction to a subset $\Lambda_1 \times \Lambda_2$ by J_{Λ_1,Λ_2}. Then the fermion point field μ associated with K exists and is the unique Gibbs measure for the potential*

$$U(x_0|\xi) = J(x_0,x_0) - J_{\{x_0\},\xi}(J_{\xi,\xi})^{-1}J_{\xi,\{x_0\}}, (x_0 \in R, \xi \in Q).$$

Here the potential $U(x_0|\xi)$ is defined by

$$U(x_0|\xi) = -\log \frac{\mu(\xi\{x_0\} = 1|\mathcal{B}_{\{x_0\}^c})(\xi)}{\mu(\xi\{x_0\} = 0|\mathcal{B}_{\{x_0\}^c})(\xi)}, \tag{2.2}$$

where $\mathcal{B}_{\{x_0\}^c}$ is the σ-algebra generated by $\xi(x), x \neq x_0$.

To conclude this section we want to point out an analogy. The following formula for Schur functions is well known:

$$\prod_{i,j=1}^{n} (1-x_iy_j)^{-1} = \sum_p S_p(x)S_p(y),$$

where the summation is taken over all partition $p = (p_1, \dots, p_n), p_1 \geq \cdots \geq p_k \geq 1$, of the number $|p| := p_1 + \cdots + p_n$, $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ and $S_p(x)$ is the Schur function. Note that the left hand side is the reciprocal of a certain determinant. Similarly, $\prod_{i,j=1}^n(1-2x_iy_j)^{-1/2}$ is expanded in terms of zonal functions $Z_p(x)$ and $Z_p(y)$ with suitable coefficients. Moreover, the case of general α can be expanded in terms of the Jack polynomials which is a new face in the study of symmetric functions. (cf. [10]). The zonal function has been introduced and studied by mathematical statisticians mainly to apply the noncentered Wishart distributions (cf., for instance [11, 17]).

§3. Logarithmic derivatives of Fredholm determinant: Alternative proof of the nonnegativity under Condition (A)

The quantity $\det_\alpha$ can be characterized as follows:

Lemma 3.1. *Let α be a real number and A be a square matrix of size n. For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ write $X = \operatorname{diag}(x_1, \dots, x_n)$. Then,*

$$\det_\alpha(A) = \frac{\partial^n}{\partial x_1 \dots \partial x_n}\Big|_{x=0} \det(I - \alpha AX)^{-1/\alpha}.$$

Proof. Denote by E_i the diagonal matrix with 1 on the ith entry and 0 elsewhere. For a while we write $G = (I - \alpha AX)^{-1}$ for simplicity of notations. Then,

$$\frac{\partial}{\partial x_i} \det(I - \alpha AX)^{-1/\alpha} = \operatorname{Tr}(GAE_i) \det(I - \alpha AX)^{-1/\alpha}.$$

Moreover, since $\frac{\partial}{\partial x_j} G = \alpha GAE_j G$, we obtain

$$\begin{aligned} & \frac{\partial}{\partial x_j} \operatorname{Tr}(GAE_{i_1} \dots GAE_{i_k}) \\ = \ & \alpha\{\operatorname{Tr}(GAE_j GAE_{i_1} \dots GAE_{i_k}) + \operatorname{Tr}(GAE_{i_1} GAE_j GAE_{i_2} \dots GAE_{i_k}) \\ & + \dots + \operatorname{Tr}(GAE_{i_1} \dots GAE_{i_k} GAE_j)\} \end{aligned}$$

for any $j, i_1, \dots, i_k \in \{1, 2, \dots, n\}$ and $k \geq 1$. From these two algorithms we obtain the above formula. Q.E.D.

The above proof also shows the following:

Lemma 3.2. *For all $k, i_1, \dots, i_k \in \{1, 2, \dots, n\}$,*

$$\det_\alpha(A_{i_1,\dots,i_k}) = \frac{\partial^k}{\partial x_{i_1} \dots \partial x_{i_k}}\Big|_{x=0} \det(I - \alpha AX)^{-1/\alpha},$$

where $A_{i_1,\dots,i_k}$ stands for the square matrix of size k whose (j,k)-element is the (i_j, i_k)-element of A.

Example 3.3. *Let $\alpha = 2$, $X = \operatorname{diag}(x_1, \dots, x_n)$ and C be a positive definite matrix of size n. Assume $\max |x_i|$ is sufficiently small. Then,*

$$\begin{aligned} & \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{\det C}} \int_{\mathbb{R}^n} \exp(-\sum_{i=1}^n x_i u_i^2) \exp(-\frac{1}{2}\langle C^{-1}u, u\rangle) du \\ = \ & (\det(I + 2CX))^{-1/2}. \end{aligned}$$

Hence,

$$\det_2(C) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{\det C}} \int_{\mathbb{R}^n} u_1^2 \cdots u_n^2 \exp(-\frac{1}{2}\langle C^{-1}u, u\rangle) du.$$

In other words, if $Z = (Z_1, \dots, Z_n)$ is a Gaussian random variable with mean 0 and covariance matrix C, then

$$\det_2(C) = E[Z_1^2 \cdots Z_n^2].$$

Proposition 3.4. *Let α be a real number and T be a trace class integral operator on $L^2(R, \lambda)$ with kernel $T(x, y)$. Then,*

$$\begin{aligned} &\operatorname{Det}(I - \alpha z T)^{-1/\alpha} \\ = \ & 1 + \sum_{n=1}^{\infty} \frac{z^n}{n!} \int_R \cdots \int_R \det_\alpha(T(x_i, x_j))_{i,j=1}^n \lambda(dx_1) \dots \lambda(dx_n) \end{aligned}$$

if $z \in \mathbb{C}$ and $|z|$ is sufficiently small(so that $|\alpha z| \|T\| < 1$).

Proof. If T is a finite dimensional operator, the assertion follows immediately from the Taylor expansion of $\det(I - \alpha z T)^{-1/\alpha}$ in z based on Lemmas 3.1 and 3.2 where λ is the counting measure. The generalization to the trace class operators is obtained by a routine approximation procedure. Q.E.D.

As an application of Proposition 3.4 one can give a proof to the following well-known formula.

Proposition 3.5. *Let $Z(x), x \in R$ be a Gaussian random field with mean 0 and covariance $K(x, y)$ in the sense that*

$$E[\{\int_R Z(x)\varphi(x)\lambda(dx)\}^2] = \int_R \int_R K(x, y)\varphi(x)\varphi(y)\lambda(dx)\lambda(dy).$$

Then,

$$E[\exp(-\int_R Z(x)^2 \varphi(x)\lambda(dx))] = \operatorname{Det}(I + 2\varphi K)^{-1/2}.$$

Proof. Expand the exponential in the right hand side. Then the expectation of each term is expressed as

$$\begin{aligned} &\int_R \cdots \int_R \lambda(dx_1) \dots \lambda(dx_n) E[Z(x_1)^2 \dots Z(x_n)^2] \\ = \ &\int_R \cdots \int_R \lambda(dx_1) \dots \lambda(dx_n) \det_2(K(x_i, x_j))_{i,j=1}^n. \end{aligned}$$

Consequently, we obtain the desired formula from the previous Proposition 3.4. Q.E.D.

Theorem 3.6. *Let A be a nonnegative definite symmetric matrix and $\alpha \in \{2/m \mid m = 1, 2, \dots\}$. Then,*

$$\det_\alpha(A) \geq 0.$$

Moreover, if $Z = (Z_1, \dots, Z_n)$ be a Gaussian random variable with mean 0 and covariance matrix $(1/m)A$ and $Z^{(1)}, \dots, Z^{(m)}$ be m independent copies of Z, then

$$\det_{2/m}(A) = E[\prod_{i=1}^{n}(\sum_{j=1}^{m}(Z_i^{(j)})^2)]. \tag{3.1}$$

Proof. Consider

$$E[\exp(-\sum_{i=1}^{n} x_i \sum_{j=1}^{m}(Z_i^{(j)})^2)] = \det(I + (2/m)XA)^{-m/2}$$

and differentiate it in $x_1, \dots, x_n$ successively. Then we obtain the formula (3.1) and so the desired nonnegativity. Q.E.D.

References

[1] R. C. Griffiths and R. K. Milne, *A class of infinitely divisible multivariate negative binomial distributions*, J. Multivariate Anal. **22** (1987), 13–23.

[2] K. Huang, Statistical Mechanics, John Wiley, 1963

[3] K. Johansson, *Discrete polynuclear growth and determinantal processes*, available via http://xxx.lanl.gov/abs/math.PR/0206208.

[4] S. Karlin and J. McGregor, *Coincidence properties of birth and death processes*, Pacific J. Math. **9** (1959), 1109–1140

[5] S. Karlin and J. McGregor, *Coincidence probabilities*, Pacific J. Math. **9** (1959), 1141–1164

[6] M. Katori and H. Tanemura, *Functional central limit theorems for vicious walkers*, available via http://xxx.lanl.gov/abs/math.PR/0203286.

[7] R. Lyons, *Determinantal probability measures*, preprint.

[8] O. Macchi. *The coincidence approach to stochastic point processes*, Adv. Appl. Prob. **7** (1975), 83–122

[9] O. Macchi, *The fermion process – a model of stochastic point process with repulsive points*, Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A, pp. 391–398.

[10] I. G. McDonald, Symmetric functions and Hall polynomials. 2nd ed., Clarendon Press, 1995

[11] R. J. Muirhead, Aspects of Multivariants statistical Theory, John Wiley, 1982

[12] M. Prähofer and H. Spohn, *Scale invariance of the PNG droplet and the Airy kernel*, J. of Stat. Phys. **108** (2002), 1071–1106.

[13] T. Shirai and Y. Takahashi, *Fermion process and Fredholm determinant*, Proceedings of the Second ISAAC Congress 1999 Vol.1, (Eds.) H.G.W. Begehr, R.P. Gilbert and J. Kajiwara, 15–23, (2000), Kluwer Academic Publ.

[14] T. Shirai and Y. Takahashi, *Random point fields associated with certain Fredholm determinant I: Fermion, Poisson and Boson processes*, submitted.

[15] T. Shirai and Y. Takahashi, *Random point fields associated with certain Fredholm determinant II : fermion shifts and their ergodic and Gibbs properties*, to appear in Annals of Prob.

[16] A. Soshnikov, *Determinantal random point fields*, Russian Math. Surveys **55** (2000), 923–975.

[17] A. Takemura, Zonal polynomials, IMS Lec. Notes vol.4, ed. S.S.Gupta, 1984.

[18] D. Vere-Jones, *A generalization of permanents and determinants*, Linear Algebra Appl. **111** (1988), 119–124.

Tomoyuki Shirai
Department of Computational Science
Kanazawa University
Kakuma-machi, Kanazawa
Ishikawa 920-1192
Japan
E-mail address: `shirai@kenroku.kanazawa-u.ac.jp`

Yoichiro Takahashi
Research Institute for Mathematical Sciences
Kyoto University
Sakyo-ku
Kyoto 606-8502
Japan
E-mail address: `takahasi@kurims.kyoto-u.ac.jp`

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 355–376

Lévy Processes Conditioned to Stay Positive and Diffusions in Random Environments

Hiroshi Tanaka

Abstract.

Some general properties of Lévy processes conditioned to stay positive are studied. As an application, which is our main concern, a result of localization is obtained for diffusion processes in Lévy environments.

§1. Introduction

We discuss a problem of localization of diffusion processes in Lévy random environments. For this we must first prepare some general properties of Lévy processes conditioned to stay positive, which were studied intensively by Bertoin [1, 2] and Chaumont [4, 5]. Some of our results in § **2** may also be found in [1] and [5] but our method is more or less analytical and different from theirs.

Let **W** denote the space of real valued right continuous functions on $[0, \infty)$ with left limits and vanishing at 0. For an element w of **W** we write $w = (w(t),\ t \geq 0)$ in § **2** and $w = (w(x),\ x \geq 0)$ in § **3**.

Given a one-dimensional Lévy process $W = \{w(t),\ t \geq 0,\ P\}$, we define a function h by $h(x) = \mu([0, x))$, $x > 0$, where μ is the measure in $[0, \infty)$ determined by (2.9). According to Silverstein [20] the function h is sub-invariant for the absorbing process W^- in $(0, \infty)$; it is invariant for W^- if $\sup w(t) = \infty$ a.s. Therefore $H(t, x, dy) = h(x)^{-1} P^-(t, x, dy) h(y)$ is a sub-Markov transition function in $(0, \infty)$ where $P^-(t, x, dy)$ denotes the transition function of W^-. Defining the transition $H(t, 0, dy)$ from 0 in a suitable way, we will have a Markov process with state space $[0, \infty)$, called the h-transform of W^- and denoted by W^h. When $\sup w(t) = \infty$ a.s., the process W^h is what we call the Lévy process W conditioned to stay positive. This definition is the same as that of Bertoin and Chaumont. When $\sup w(t) < \infty$ a.s., Hirano [10] showed that there are two

---

Received December 17, 2002.
Revised March 27, 2003.

different ways of defining Lévy processes conditioned to stay positive (under some additional condition) and so, simply to avoid confusion in this case we do not call W^h the process conditioned to stay positive in this paper, though it seems so (if, in addition, W has no positive jumps, W^h is not the same as the one considered in [2]). By an analytic method we prove that, if W enters immediately into $(0, \infty)$ a.s. and if W enters $(-\infty, 0)$ within a finite time with positive probability, then W^h has a Feller semigroup H_t strongly continuous on $C.[0, \infty)$, the space of continuous functions on $[0, \infty)$ vanishing at infinity. This fact was also noticed by Chaumont [5] as a consequence of more probabilistic arguments. For our application in § **3** we must also prepare some convergence theorems on the reversed pre-minimum and the post-minimum processes $\hat{V}_\lambda$ and V_λ defined in (2.43) and (2.44). Similar results were already obtained by Bertoin [1] and Chaumont [5] but there is a delicate difference and we need extra arguments.

In § **3** we are concerned with diffusion processes in Lévy random environments. Suppose that we are given a Lévy process $W = \{w(x), x \geq 0, P\}$. Let $\Omega = C[0, \infty)$ and for $\omega \in \Omega$ set $X(t) = X(t, \omega) = \omega(t)$ (the value of ω at time t). For each $w \in \mathbf{W}$ we denote by P^w the probability measure on Ω such that $\{X(t), t \geq 0, P^w\}$ is a reflecting diffusion process on $[0, \infty)$ with generator $\frac{1}{2}e^{w(x)}\frac{d}{dx}\left(e^{-w(x)}\frac{d}{dx}\right)$ and starting at 0. The reflecting barrier at $x = 0$ is not essential; it was considered just to simplify the situation; we may (but do not) consider the case where the Lévy environment $w(x)$ is given in the whole of $\mathbf{R}$. We set $\mathbb{P}(dwd\omega) = P(dw)P^w(d\omega)$, which is a probability measure on $\mathbf{W} \times \Omega$. We then regard $\{X(t), t \geq 0, \mathbb{P}\}$ as a process defined on the probability space $(\mathbf{W} \times \Omega, \mathbb{P})$ and call it the (reflecting) diffusion process in the Lévy environment W.

When W is a Brownian environment, Brox [3] and Schumacher [19] proved that $\{X(t), t \geq 0, \mathbb{P}\}$ has the same limiting behavior (or the same localization property) as Sinai's random walk in a random environment ([21]). A result of refinement, which corresponds to that of Golosov [8] for Sinai's random walk, was then obtained by Tanaka [22, 25] and some extension by Kawazu-Tamura-Tanaka [13]. In this paper for a certain class of Lévy environments we obtain such a results of localization, which is similar to those of [8],[22, 25],[13]. To be precise let $w \in \mathbf{W}, \lambda > 0$ and set

(1.1) $N(x) = N(x, w) = \inf\{w(y) : 0 \leq y \leq x\}$, $w^{\#}(x) = w(x) - N(x)$,

(1.2) $a_\lambda = a_\lambda(w) = \inf\{x > 0 : w^{\#}(x) > \lambda\}$,

(1.3) $b_\lambda = b_\lambda(w) =$ the unique x such that $w(x)$ is equal to $N(a_\lambda)$.

In general there may be many x with $w(x) = N(a_\lambda)$ but, in the case we actually discuss, such an x is unique a.s. (see Lemma 7). It will be proved that, under a certain condition on W, the distribution of $X(e^\lambda) - b_\lambda$ under $\mathbb{P}$ tends to some nondegenerate distribution $\bar{\nu}$ as $\lambda \to \infty$. It can happen that the limit distribution of $X(e^\lambda) - b_\lambda$ under $\mathbb{P}$ exists even when the limit distribution of a suitably scaled $X(t)$ (without centering, under $\mathbb{P}$) does not; the latter exists if, in addition, b_λ has a limit distribution under a suitable scaling (without centering). We are also interested in the form of the limit distribution $\bar{\nu}$. Under a certain condition on W our result is that $\bar{\nu}$ can be expressed in terms of two independent Lévy processes conditioned to stay positive starting at 0, the one is related to W and the other to $-W$.

§2. Lévy processes conditioned to stay positive

We use the notation in § **1** and so $W = \{w(t),\, t \geq 0,\, P\}$ is a Lévy process starting at 0. Throughout the paper we exclude the trivial case where $w(t) = 0\ (t \geq 0)$ a.s. The infimum process $N(t)$ and the reflecting process $w^{\#}(t)$ are defined by (1.1) with x replaced by t. The hitting (entrance) times $\sigma(x)$ and $\tau(x)$ are defined by

$$\sigma(x) = \inf\{t > 0 : x + w(t) \leq 0\}, \quad x \geq 0,$$

$$\tau(x) = \inf\{t > 0 : x + w(t) < 0\}, \quad x \geq 0,$$

$$\sigma = \sigma(0), \quad \tau = \tau(0).$$

We often consider $\hat{W} = \{\hat{w}(t), t \geq 0\}$, the dual of W, where $\hat{w}(t) = -w(t)$ and define $\hat{N}(t), \hat{\sigma}, \hat{\tau}$, etc., similarly in terms of $\hat{W}$. The absorbing Lévy process W^- in $(0, \infty)$ is defined as the Markov process $\{x + w(t), 0 \leq t < \sigma(x), x > 0\}$ and its transition function, semigroup and Green operator are denoted by $P^-(t, x, dy)$, T_t^- and G_λ^-, respectively. We set $G^- = G_0^-$. Another absorbing Lévy process $W^=$ on $[0, \infty)$, which does not much differ from W^-, is the Markov process $\{x + w(t), 0 \leq t < \tau(x), x \geq 0\}$; its Green operator of order 0 is denoted by $G^=$. We also define the reflecting Lévy process $W^{\#}$ on $[0, \infty)$ associated with W as the Markov process $\{w^{\#}(t; x), t \geq 0, x \geq 0\}$, where

$$\begin{aligned} w^{\#}(t; x) &= \sup_{0 \leq s \leq t} \{w(t; x) - w(s; x)\} \vee w(t; x) \\ &= \begin{cases} x + w(t) & \text{if } x + N(t) > 0, \\ w(t) - N(t) & \text{if } x + N(t) \leq 0, \end{cases} \end{aligned}$$

wherein $w(t;x) = x + w(t)$ and $a \vee b = \max\{a, b\}$, and we denote by $P^{\#}(t, x, dy)$ the transition function of $W^{\#}$. The reflecting dual Lévy process $\hat{W}^{\#}$ on $[0, \infty)$ and its transition function $\hat{P}^{\#}(t, x, dy)$ are defined in a similar manner from $\hat{W}$. Throughout the paper T denotes an exponential random time with mean $1/\lambda$ and independent of W.

2.1 Preliminaries. In this subsection we present in an elementary way some preliminary and known facts concerning the measure μ such as the (sub-)invariance for $\hat{W}^{\#}$.

Lemma 1. (Silverstein [20, p.556]) (i) *For any fixed* $t > 0, x > 0$ *and* $y \geq 0$ *we have* $\hat{P}^{\#}(t, y, [0, x)) = P^{-}(t, x, (y, \infty))$.
(ii) *For any fixed* $t \geq 0$, $-N(t) \overset{d}{=} \hat{w}^{\#}(t)$ *and* $-\hat{N}(t) \overset{d}{=} w^{\#}(t)$, *where* $\overset{d}{=}$ *is the equality in distribution. In particular,* $-N(T) \overset{d}{=} \hat{w}^{\#}(T)$ *and* $-\hat{N}(T) \overset{d}{=} w^{\#}(T)$.

Let ν_λ be the distribution of $-N(T)$, or equivalently of $\hat{w}^{\#}(T)$, and $\hat{\nu}_\lambda$ be the distribution of $-\hat{N}(T)$, or equivalently of $w^{\#}(T)$. The fluctuation identity

$$\log E\{e^{\xi N(T)+\eta(w(T)-N(T))}\}$$
$$= \int_0^\infty \frac{e^{-\lambda t}}{t} \Big\{E(e^{\xi w(t)} - 1; w(t) < 0) + E(e^{\eta w(t)} - 1; w(t) > 0)\Big\} dt$$

due to Pecherskii-Rogozin [15] (see also Sato [18], Bertoin [2], Doney [6]) implies that $N(T)$ and $w^{\#}(T)$ are independent and for $\xi \geq 0$

$$(2.1) \quad \mathcal{L}(\xi, \nu_\lambda) = \exp \int_0^\infty t^{-1} e^{-\lambda t} E\{e^{\xi w(t)} - 1; w(t) < 0\} dt,$$

$$(2.2) \quad \mathcal{L}(\xi, \hat{\nu}_\lambda) = \exp \int_0^\infty t^{-1} e^{-\lambda t} E\{e^{-\xi w(t)} - 1; w(t) > 0\} dt,$$

where $\mathcal{L}(\xi, \nu)$ denotes the Laplace transform $\int_{[0,\infty)} e^{-\xi x} \nu(dx)$ of a measure ν. The equations (2.1) and (2.2) imply that there exist finite measures μ_λ and $\hat{\mu}_\lambda$ on $[0, \infty)$ such that

$$(2.3) \qquad \mathcal{L}(\xi, \mu_\lambda) = \exp \int_0^\infty t^{-1} e^{-\lambda t} E\{e^{\xi w(t)} - e^{-t}; w(t) < 0\} dt,$$

$$(2.4) \qquad \mathcal{L}(\xi, \hat{\mu}_\lambda) = \exp \int_0^\infty t^{-1} e^{-\lambda t} E\{e^{-\xi w(t)} - e^{-t}; w(t) > 0\} dt,$$

$$(2.5) \qquad \nu_\lambda = c_\lambda \, \mu_\lambda, \qquad \hat{\nu}_\lambda = \hat{c}_\lambda \, \hat{\mu}_\lambda,$$

$$(2.6) \qquad c_\lambda = \exp\left\{- \int_0^\infty t^{-1} e^{-\lambda t} (1 - e^{-t}) P(w(t) < 0) dt\right\},$$

and $\hat{c}_\lambda$ is defined with the replacement of "$w(t) < 0$" by "$w(t) > 0$" in the equation (2.6).

For our elementary and straightforward method the following simple lemma, probably known, is useful.

Lemma 2. *If W is not the zero process, then for any $\xi > 0$*

$$\text{(i)} \qquad E\{e^{-\xi|w(t)|}\} \le \text{const.}t^{-1/4} \qquad (t \ge 0),$$

$$\text{(ii)} \qquad E\{1 - e^{-\xi|w(t)|}\} \le \text{const.}t^{1/2} \qquad (0 \le t \le 1),$$

where const. may depend on ξ.

Proof. Assume that W is not deterministic. Then we can write $w(t) = w_0(t) + w_1(t)$ where $w_0(t)$ and $w_1(t)$ are independent Lévy processes and $E\{w_0(t)\} = 0$, $E\{|w_0(t)|^2\} = \sigma^2 t$, $(\sigma > 0)$, $E\{|w_0(t)|^3\} < \infty$. In fact, the decomposition can be obtained by noting the fact that any Lévy process having Lévy measure with bounded support admits finite absolute moments of all positive orders (e.g. see Sato [18, p.161]). We now make use of the Berry-Esseen theorem (e.g. see Feller [7, p.542]):

$$\sup_{x \in \mathbf{R}} \left| P\{(\sigma\sqrt{t})^{-1} w_0(t) \le x\} - \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-y^2/2)dy \right| = O(t^{-1/2}),$$

as $t \to \infty$. Setting $Y(t) = (\sigma\sqrt{t})^{-1}|w_0(t) + x|$ we have

$$\begin{aligned} &E\{e^{-\xi|w_0(t)+x|}\} \\ &= E\{e^{-\xi\sigma\sqrt{t}Y(t)}; Y(t) < t^{-1/4}\} + E\{e^{-\xi\sigma\sqrt{t}Y(t)}; Y(t) \ge t^{-1/4}\} \\ &\le P\{Y(t) < t^{-1/4}\} + \exp(-\xi\sigma t^{1/4}) \\ &\le \text{const.}t^{-1/2} + \int_{\{|y+(\sigma\sqrt{t})^{-1}x| < t^{-1/4}\}} (2\pi)^{-1/2} e^{-y^2/2} dy + e^{-\xi\sigma t^{1/4}} \\ &\le \text{const.}t^{-1/4} \quad \text{(for large } t\text{)}, \end{aligned}$$

where const. may depend on ξ. Therefore

$$E\{e^{-\xi|w(t)|}\} = \int_{-\infty}^{\infty} E\{e^{-\xi|w_0(t)+x|}\} P\{w_1(t) \in dx\} \le \text{const.}t^{-1/4}.$$

The proof of (ii) is omitted.

2.1.1 *A formula on Green operators of absorbing Lévy processes.* For $\lambda > 0$, $x > 0$ and $f \in C_0[0, \infty)$, the space of continuous functions with compact supports, we have

$$(2.7)\quad \begin{aligned} G_\lambda^- f(x) &= \int_0^\infty e^{-\lambda t} E\{f(x+w(t));\ \sigma(x) > t\} dt \\ &= \lambda^{-1} E\{f(x+N(T)+w^{\#}(T));\ -N(T) < x\} \\ &= \lambda^{-1} \int_{[0,x)} \nu_\lambda(du) \int_{[0,\infty)} \hat{\nu}_\lambda(dv) f(x-u+v), \end{aligned}$$

and similarly

$$(2.8)\quad G_\lambda^= f(x) = \lambda^{-1} \int_{[0,x]} \nu_\lambda(du) \int_{[0,\infty)} \hat{\nu}_\lambda(dv) f(x-u+v), \quad x \geq 0.$$

By Lemma 2 the integrals on the right hand sides of (2.3) and (2.4) are convergent for $\lambda = 0$ and so the measures μ_λ and $\hat{\mu}_\lambda$ converge vaguely as $\lambda \downarrow 0$ to the measures μ and $\hat{\mu}$ in $[0, \infty)$, respectively, which are defined by

$$(2.9)\qquad \mathcal{L}(\xi, \mu) = \exp \int_0^\infty t^{-1} E\{e^{\xi w(t)} - e^{-t};\ w(t) < 0\} dt,$$

$$(2.10)\qquad \mathcal{L}(\xi, \hat{\mu}) = \exp \int_0^\infty t^{-1} E\{e^{-\xi w(t)} - e^{-t};\ w(t) > 0\} dt,$$

where $\xi > 0$. Moreover, using the definition of c_λ and $\hat{c}_\lambda$ we see that $\lambda^{-1} c_\lambda \hat{c}_\lambda \to c^0$ as $\lambda \downarrow 0$ where

$$(2.11)\qquad c^0 = \exp \int_0^\infty t^{-1}(1 - e^{-t}) P\{w(t) = 0\} dt,$$

which is finite by Lemma 2. It is also known that (e.g. see Sato [18, p.372])

$$(2.12)\qquad c^0 = 1 \quad \text{if } W \text{ is not a compound Poisson process.}$$

Thus $\lambda^{-1}\nu_\lambda \otimes \hat{\nu}^\lambda = \lambda^{-1} c_\lambda \hat{c}_\lambda \mu_\lambda \otimes \hat{\mu}_\lambda \to c^0 \mu \otimes \hat{\mu}$ vaguely as $\lambda \downarrow 0$ and hence letting $\lambda \downarrow 0$ in (2.7) and (2.8) we obtain the following theorem.

Theorem 1. *If W is not the zero process, then for $f \in C_0[0, \infty)$*

$$(2.13)\quad G^- f(x) = c^0 \int_{[0,x)} \mu(du) \int_{[0,\infty)} \hat{\mu}(dv) f(x-u+v), \quad x > 0,$$

$$(2.14)\quad G^= f(x) = c^0 \int_{[0,x]} \mu(du) \int_{[0,\infty)} \hat{\mu}(dv) f(x-u+v), \quad x \geq 0.$$

This theorem was obtained by Ray [16] for symmetric stable processes and by Silverstein [20] for general Lévy processes; the present derivation of (2.13) and (2.14) was taken from Tanaka [23, 24] with a slight improvement.

2.1.2 *The measure μ is a sub-invariant measure of the Markov process $\hat{W}^{\#}$.*

Theorem 2. (Silverstein [20]) (i) *If the Markov process $\hat{W}^{\#}$ is recurrent, then μ is an invariant measure of $\hat{W}^{\#}$.*
(ii) *If $\hat{W}^{\#}$ is transient, then μ is a sub-invariant measure of $\hat{W}^{\#}$; more precisely, for any $A \in \mathcal{B}[0,\infty)$*

$$\mu(A) = \bar{c}E\left\{\int_0^\infty \mathbf{1}_A(\hat{w}^{\#}(t))dt\right\}, \tag{2.15}$$

$$\int_{[0,\infty)} \mu(dx)\hat{P}^{\#}(t,x,A) = \mu(A) - \bar{c}E\left\{\int_0^t \mathbf{1}_A(\hat{w}^{\#}(s))ds\right\}, \tag{2.16}$$

$$\bar{c} = \exp\left\{-\int_0^\infty t^{-1}(1-e^{-t})P(w(t) \geq 0)dt\right\}. \tag{2.17}$$

Proof. (i) We assume that $\hat{W}^{\#}$ is recurrent and that W is not an increasing process. Take $a > 0$, let $T^{\#}$ be the time of first return of $\hat{w}^{\#}(t)$ to 0 after visiting (a,∞) and define the measures $\mu_\lambda^{\#}, \lambda \geq 0$, in $[0,\infty)$ by

$$\int f d\mu_\lambda^{\#} = E\left\{\int_0^{T^{\#}} e^{-\lambda t} f(\hat{w}^{\#}(t))dt\right\}. \tag{2.18}$$

Then

$$\mu_\lambda^{\#} = \lambda^{-1}\left\{1 - E(e^{-\lambda T^{\#}})\right\}\nu_\lambda = \lambda^{-1}\left\{1 - E(e^{-\lambda T^{\#}})\right\}c_\lambda \mu_\lambda.$$

Since $\mu_\lambda^{\#} \to \mu_0^{\#}$ and $\mu_\lambda \to \mu$ as $\lambda \downarrow 0$, the above identity implies that the measure $\mu_0^{\#}$ is a constant multiple of μ. On the other hand it is easy to see that $\mu_0^{\#}$ is an invariant measure of the recurrent process $\hat{W}^{\#}$ and so is μ. (ii) If $\hat{W}^{\#}$ is transient, then $\bar{c} > 0$ and the assertion follows from

$$\int f d\mu = \bar{c}E\left\{\int_0^\infty f(\hat{w}^{\#}(t))dt\right\}.$$

We now introduce a function $h(x), x \geq 0$, by

$$h(x) = \begin{cases} \mu([0,x)) & \text{for } x \geq 0, \\ \mu(\{0\}) & \text{for } x = 0. \end{cases} \tag{2.19}$$

Then using Lemma 1 we can rephrase Theorem 2 as follows.

Theorem 3. (Silverstein [20]) (i) *If* $\hat{W}^{\#}$ *is recurrent, then*

$$\int_{(0,\infty)} P^{-}(t,x,dy)h(y) = h(x), \quad x > 0, \tag{2.20}$$

$$\lambda G_{\lambda}^{-}h(x) = h(x), \quad x > 0. \tag{2.21}$$

(ii) *If* $\hat{W}^{\#}$ *is transient, then for any* $x > 0$

$$\int_{(0,\infty)} P^{-}(t,x,dy)h(y) = h(x) - \bar{c}\int_0^t P\{\hat{w}^{\#}(s) < x\}ds, \tag{2.22}$$

$$\lambda G_{\lambda}^{-}h(x) = h(x) - \lambda^{-1}\bar{c}\nu_{\lambda}([0,x)). \tag{2.23}$$

Remark. The following conditions are equivalent to each other (e.g. see Sato [18]).

(i) $\hat{W}^{\#}$ is recurrent. (ii) $\sup_{t \geq 0} w(t) = \infty$, a.s.
(iii) $\int_0^\infty t^{-1}(1-e^{-t})P\{w(t) > 0\}dt = \infty$. (iv) $\bar{c} = 0$.
(v) $\int_1^\infty t^{-1}P\{w(t) > 0\}dt = \infty$. (vi) $\hat{\mu}(\mathbf{R}) = \infty$.

2.2 The Feller property of the semigroup of W^h. We define the superharmonic transform $H(t,x,dy)$ of $P^{-}(t,x,dy)$ by

$$H(t,x,dy) = h(x)^{-1}P^{-}(t,x,dy)h(y). \tag{2.24}$$

We set $H_t f(x) = \int_{(0,\infty)} H(t,x,dy)f(y)$. Then $H_t f(x)$ is well-defined for $f \in C.[0,\infty)$ and for $x > 0$. We will prove that $H_t f$ can be extended to a function in $[0,\infty)$ so that H_t gives rise to a strongly continuous sub-Markov semigroup on $C.[0,\infty)$ provided that $\hat{\tau} = 0$ a.s. and $\tau < \infty$ with positive probability.

We prepare three lemmas.

Lemma 3. (i) If W is not a compound Poisson process, then $\mu(\{x\}) = \nu_{\lambda}(\{x\}) = 0$ for any $x > 0$ and $\lambda > 0$.
(ii) The condition $\tau = 0$ a.s. is equivalent to each of $\mu(\{0\}) = 0$ and $\nu_{\lambda}(\{0\}) = 0$.

The proof is easy; for instance, the equivalence of $\tau = 0$ (a.s.) and $\mu(\{0\}) = 0$ follows from the formula (2.14). The rest are omitted.

In what follows we often use the notation $\nu(f) = \int_{[0,\infty)} f d\nu$.

Lemma 4. *Suppose that W is not the zero process.*
(i) *For any $\lambda > 0$ and $x > 0$*

$$(2.25)\quad \lambda^{-1}\bar{c} + \int_{[0,\infty)} h(v)\hat{\nu}_\lambda(dv) \le \frac{h(x)}{\nu_\lambda([0,x))} \le \lambda^{-1}\bar{c} + \int_{[0,\infty)} h(x+v)\hat{\nu}_\lambda(dv).$$

(ii) *For any $\lambda > 0$*

$$(2.26)\quad \lim_{x\downarrow 0} \frac{\nu_\lambda([0,x))}{h(x)} = \alpha_\lambda,$$

$$(2.27)\quad \alpha_\lambda = \left\{\lambda^{-1}\bar{c} + \hat{\nu}_\lambda(h)\right\}^{-1} \in (0,\infty).$$

Proof. The function $\lambda G_\lambda^- h$ can be expresses in two ways:

$$(2.28)\quad \lambda G_\lambda^- h(x) = \int_{[0,x)} \nu_\lambda(du) \int_{[0,\infty)} \hat{\nu}_\lambda(dv) h(x-u+v).$$

$$(2.29)\quad \lambda G_\lambda^- h(x) = h(x) - \lambda^{-1}\bar{c}\,\nu_\lambda([0,x)).$$

Firstly we remark that the finiteness of $\hat{\nu}_\lambda(h)$ follows from (2.28); moreover, if $\hat{\nu}_\lambda(h) = 0$ then $\hat{\nu}_\lambda$ is the δ-distribution at 0 so W is decreasing and hence $\bar{c} > 0$. Thus $0 < \alpha_\lambda < \infty$ always. Secondly from (2.28) and (2.29) we have

$$h(x) = \lambda^{-1}\bar{c}\nu_\lambda([0,x)) + \int_{[0,x)} \nu_\lambda(du) \int_{[0,\infty)} \hat{\nu}_\lambda(dv) h(x-u+v),$$

and hence

$$\lambda^{-1}\bar{c} + \hat{\nu}_\lambda(h) \le \frac{h(x)}{\nu_\lambda([0,x))} \le \lambda^{-1}\bar{c} + \hat{\nu}_\lambda\left(h^x\right),$$

where $h^x(\cdot) = h(x + \cdot)$, which proves (2.25) and (2.26). The proof of the lemma is finished.

Let $\lambda > 0, f \in C_0[0,\infty)$ and set

$$(2.30)\quad U_\lambda f(x) = \int_0^\infty e^{-\lambda t} H_t f(x) dt, \quad x > 0.$$

Then by (2.7) we have

$$(2.31)\quad U_\lambda f(x) = \lambda^{-1} h(x)^{-1} \int_{[0,x)} \nu_\lambda(du) \int_{[0,\infty)} \hat{\nu}_\lambda(dv)\, \tilde{f}(x-u+v),$$

where $\tilde{f} = fh$. If W is not a compound Poisson process, then $h(x)$ is continuous by Lemma 3 and hence $U_\lambda f(x)$ is also continuous in $x > 0$. Moreover, $U_\lambda f(x)$ tends to $\lambda^{-1}\alpha_\lambda \hat{\nu}_\lambda(\tilde{f})$ as $x \downarrow 0$ by (2.26) and (2.31). On the other hand it is clear that $U_\lambda f(x)$ tends to 0 as $x \to \infty$. Therefore $U_\lambda f(x), x > 0$, can be extended continuously to a function in $C.[0,\infty)$, which we denote by the same notation $U_\lambda f$. Since $\|U_\lambda f\|_\infty \leq \lambda^{-1}\|f\|_\infty$, $U_\lambda f$ is well-defined also for $f \in C.[0,\infty)$. Thus we have a linear operator $U_\lambda : C.[0,\infty) \to C.[0,\infty)$, which clearly satisfies

$$(2.32)\qquad U_\lambda f \geq 0 \quad \text{if } f \geq 0,$$

$$(2.33)\qquad \|U_\lambda f\|_\infty \leq \lambda^{-1}\|f\|_\infty,$$

$$(2.34)\qquad U_\lambda - U_{\lambda'} + (\lambda - \lambda')U_\lambda U_{\lambda'} = 0, \quad \lambda > 0,\ \lambda' > 0.$$

Now we introduce the following conditions.

(A) $\tau = \hat{\tau} = 0$, a.s.

(A$'$) $\hat{\tau} = 0$ a.s. and $0 < \tau < \infty$ with positive probability.

Lemma 5. *If either one of the conditions* (A) *and* (A$'$) *is satisfied, then*

$$(2.35)\qquad \lim_{\lambda\to\infty} \|\lambda U_\lambda f - f\|_\infty = 0 \quad \textit{for} \quad f \in C.[0,\infty).$$

Proof. Making use of (2.7) and (2.26) we have

$$(2.36)\quad \lambda U_\lambda f(0) = \lim_{x\downarrow 0} \lambda U_\lambda f(x) = \lambda \lim_{x\downarrow 0} h(x)^{-1} G_\lambda^- \tilde{f}(x) = \alpha_\lambda \hat{\nu}_\lambda(\tilde{f}).$$

If we set $\rho_\lambda(dx) = \alpha_\lambda h(x)\hat{\nu}_\lambda(dx)$, then ρ_λ is a measure in $[0,\infty)$ with total mass ≤ 1 and (2.36) yields

$$(2.37)\qquad \lambda U_\lambda f(0) = \rho_\lambda(f), \quad f \in C.[0,\infty).$$

We are going to prove that ρ_λ converges vaguely to δ_0 as $\lambda \to \infty$. To prove this, we assume that $f = U_\theta g$ with $g \in C_0[0,\infty)$ and $\theta > 0$. Then the equation (2.34) implies $\|\lambda U_\lambda f - f\|_\infty = \|\theta U_\lambda f - U_\lambda g\|_\infty \to 0$, as $\lambda \to \infty$. In particular,

$$(2.38)\qquad U_\theta g(0) = f(0) = \lim_{\lambda\to\infty} \lambda U_\lambda f(0) = \lim_{\lambda\to\infty} \rho_\lambda(f).$$

Let ρ be any vague limit of ρ_λ as $\lambda \to \infty$ via a sequence $\{\lambda_n\}$. Then (2.38) implies $U_\theta g(0) = \rho(f)$, which can be rewritten, again by making use of (2.7) and (2.26), as follows:

$$\begin{aligned}(2.39)\quad \theta^{-1}\alpha_\theta\, \hat{\nu}_\theta(\tilde{g}) &= \rho(\{0\})\theta^{-1}\alpha_\theta\, \hat{\nu}_\theta(\tilde{g})\\ &+ \theta^{-1}\int_{(0,\infty)} h(x)^{-1}\rho(dx)\int_{[0,x)}\nu_\theta(du)\int_{[0,\infty)}\tilde{g}(x-u+v)\hat{\nu}_\theta(dv).\end{aligned}$$

We now prove that, under the assumption of the lemma, the equation (2.39) holds for $g(x) = h(x)^{-1}e^{-\xi x}$, $\xi > 0$. Since $\hat{\nu}_\theta(\{0\}) = 0$ by Lemma 3, the integration interval $[0,\infty)$ of $\hat{\nu}_\theta$ in (2.39) can be replaced by the open interval $(0,\infty)$. With such a replacement we take $g_n(x) = \min\{h(x)^{-1}e^{-\xi x}, n\}$ for $g(x)$ in (2.39) and then let $n \uparrow \infty$. The result is

$$\begin{aligned}\theta^{-1}\alpha_\theta\int_{(0,\infty)} e^{-\xi x}\hat{\nu}_\theta(dx) &= \rho(\{0\})\theta^{-1}\alpha_\theta\int_{(0,\infty)} e^{-\xi x}\hat{\nu}_\theta(dx)\\ +\theta^{-1}\int_{(0,\infty)} h(x)^{-1}\rho(dx)&\int_{[0,x)}\nu_\theta(du)\int_{(0,\infty)} e^{-\xi(x-u+v)}\hat{\nu}_\theta(dv),\end{aligned}$$

or equivalently,

$$\alpha_\theta = \rho(\{0\})\alpha_\theta + \int_{(0,\infty)} h(x)^{-1}\rho(dx)\int_{[0,x)} e^{-\xi(x-u)}\nu_\theta(du).$$

Letting $\xi \uparrow \infty$ we obtain $\alpha_\theta = \rho(\{0\})\alpha_\theta$ so $\rho = \delta_0$. This proves that ρ_λ converges vaguely to δ_0 as $\lambda \to \infty$. Thus (2.37) implies

$$(2.40)\qquad \lim_{\lambda\to\infty}\lambda U_\lambda f(0) = f(0), \quad f \in C.[0,\infty).$$

On the other hand it is clear that, for any $x > 0$,

$$(2.41)\qquad \lim_{\lambda\to\infty}\lambda U_\lambda f(x) = f(x), \quad f \in C.[0,\infty).$$

From (2.40) and (2.41) we can easily derive (2.35). This completes the proof of Lemma 5.

As an immediate consequence of (2.32) ~ (2.35) we obtain the following theorem.

Theorem 4. *If* $\hat{\tau} = 0$ a.s. *and* $\tau < \infty$ *with positive probability (namely, either one of the conditions* (A) *and* (A′) *is satisfied), then there exists a unique strongly continuous sub-Markov semigroup* H_t *on* $C.[0,\infty)$ *such that, for any* $t > 0, x > 0, f \in C.[0,\infty)$,

$$(2.42)\qquad H_t f(x) = h(x)^{-1}\int_{(0,\infty)} P^-(t,x,dy)f(y)h(y).$$

Denote by $C_\triangle$ the subspace of $C.[0,\infty)$ consisting of those functions f with $f(0) = E\{f\,(w(\hat{\tau}) - w(\hat{\tau}-))\,;\ \hat{\tau} < \infty\}$. We omit the proof of the following theorem since it is not used in our later arguments. Pictorial observation of the sample path of the reversed pre-minimum process of the next subsection suggests the result. .

Theorem 5. *If* $\tau = 0, \hat{\tau} > 0$ *a.s. and* $\hat{\tau} < \infty$ *with positive probability, then there exists a unique strongly continuous sub-Markov semigroup* H_t *on the subspace* $C_\triangle$ *such that* (2.42) *holds for* $f \in C_\triangle$. H_t *can not be strongly continuous at* $t = 0$ *on the whole space* $C.[0,\infty)$.

2.3 The reversed pre-minimum and the post-minimum processes. We assume that our Lévy process $W = \{w(t),\ t \geq 0,\ P\}$ satisfies the following conditions:
Condition (A). $\tau = \hat{\tau} = 0$ a.s.
Condition (B). $\sup\{w(t) : t > 0\} = -\inf\{w(t) : t > 0\} = \infty$ a.s.
So the process W^h is the process W conditioned to stay positive. We denote by W^+ such a process starting at 0. Similarly $\hat{W}^+$ denotes the process $\hat{W}$ conditioned to stay positive starting at 0. We consider the reversed pre-minimum and the post-minimum processes $\hat{V}_\lambda$ and V_λ defined by

$$\hat{V}_\lambda(t) = w((b_\lambda - t)-) - w(b_\lambda),\ 0 \leq t < b_\lambda, \tag{2.43}$$

$$V_\lambda(t) = w(b_\lambda + t) - w(b_\lambda),\ 0 \leq t < c_\lambda, \tag{2.44}$$

where a_λ and b_λ are defined by (1.2) and (1.3) and $c_\lambda = a_\lambda - b_\lambda$. It is known that $\hat{V}_\lambda$ and V_λ are independent for each fixed λ. We are interested in the convergence in law of $\hat{V}_\lambda$ to $\hat{W}^+$ and of V_λ to W^+ (as $\lambda \to \infty$). The proof of the former convergence is considerably easier but we can prove the latter convergence only under an additional condition (C) which is somewhat stronger. We have to omit the details of the latter part since our proof is too lengthy to be included here.

2.3.1 *The reversed pre-minimum process.* To prove the law convergence of $\hat{V}_\lambda$ first we express the sample functions of W and $\hat{V}_\lambda$, à la Itô [11, (6.6) of p.233], in terms of the Poisson point process (P.p.p. for short) of "W-excursions off the zeros of $W^\#$" which was first used by Greenwood-Pitman [9] (see also Bertoin [1]). So let $L(t)$ be the local time of the reflecting process $W^\#$ at 0, let $L^{-1}(s)$ be the right continuous inverse function of $L(t)$ and set

$$\triangle_s = w(L^{-1}(s)) - w(L^{-1}(s-)), \quad \zeta_s = L^{-1}(s) - L^{-1}(s-).$$

Some equations to follow hold under the phrase "a.s." but we shall often omit to write it. We have $L^{-1}(s) = \sum_{r \leq s} \zeta_r$ (the continuous part vanishes under the condition (A)). It can also be proved that the continuous part $N_c(t)$ of $N(t)$ is equal to $-cL(t)$ where c is the nonnegative constant determined by $E\{e^{-cL(T)}\} = E\{e^{-N_c(T)}\}$, T being an exponential random time with mean 1 and independent of W. Thus the decomposition of $N(t)$ to continuous and jump parts yields $N(t-) = -cs+\sum_{r<s} \triangle_r$, $t > 0$, where s is determined by $L^{-1}(s-) \le t \le L^{-1}(s)$. Now let

$$p_s(t) = w(L^{-1}(s-)+t) - w(L^{-1}(s-)) \quad \textit{for} \quad 0 \le t \le \zeta_s.$$

$p_s = \{p_s(t),\ 0 \le t \le \zeta_s\}$ is the W-excursion on $[L^{-1}(s-), L^{-1}(s)]$ that starts at 0 (this is a consequence of the condition (A)) and moves during $[0, \zeta_s]$ with the increments of w on $[L^{-1}(s-), L^{-1}(s)]$, ending at time ζ_s with final value $\triangle_s$. Then $\{p_s,\ s > 0\}$ is a P.p.p. and we have

$$w(t) = p_s(t - L^{-1}(s-)) - cs + \sum\nolimits_{r<s} \triangle_r, \quad t > 0, \tag{2.45}$$

with s such that $L^{-1}(s-) \le t \le L^{-1}(s)$ where we use the convention that $p_s(\cdot) = 0$ whenever $L^{-1}(s-) = L^{-1}(s)$. Thus the process W is constructed from the P.p.p. $\{p_s,\ s > 0\}$. Moreover $b_\lambda = L^{-1}(s_\lambda -)$ where s_λ is the minimum of $s > 0$ such that the excursion p_s can cross the level λ.

Consider the reversed excursion $\hat{p}_s = \{\hat{p}_s(t),\ t \in \{0-\} \cup [0, \zeta_s)\}$ defined by

$$\hat{p}_s(0-) = \triangle_s \quad \textit{and} \quad \hat{p}_s(t) = p_s((\zeta_s - t)-) \quad \textit{for} \quad 0 \le t < \zeta_s.$$

Then $\{\hat{p}_s,\ s > 0\}$ is also a P.p.p., which we now modify as follows. Let $\lambda > 0$ be fixed, let s_λ be the same as before and define $\tilde{p}_s$ by

$$\tilde{p}_s = \begin{cases} \hat{p}_{s_\lambda - s} & \text{for } 0 < s < s_\lambda, \\ \hat{p}_s & \text{for } s \ge s_\lambda. \end{cases}$$

Then $\{\tilde{p}_s,\ s > 0\} \overset{d}{=} \{\hat{p}_s,\ s > 0\}$ and we can prove that, for $0 \le t < b_\lambda$,

$$\hat{V}_\lambda(t) = \tilde{p}_s(t - \tilde{L}^{-1}(s-)) + cs - \sum\nolimits_{r \le s} \tilde{\triangle}_r, \tag{2.46}$$

where s is determined by $\tilde{L}^{-1}(s-) \le t \le \tilde{L}^{-1}(s)$. If we replace $\{\tilde{p}_s\}$ by $\{\hat{p}_s\}$, the right hand side of (2.46) has the form

$$\hat{p}_s(t - L^{-1}(s-)) + cs - \sum_{r \le s} \triangle_r$$

with s such that $L^{-1}(s-) \leq t \leq L^{-1}(s)$, $0 \leq t < b_\lambda$ (it is to be noted that the inverse local time associated with $\{\hat{p}_s\}$ is still $L^{-1}(s)$). From these observations we see that a cadlag process $\{\hat{V}(t),\, t \geq 0\}$ is defined by

$$\hat{V}(t) = \hat{p}_s(t - L^{-1}(s-)) + cs - \sum\nolimits_{r \leq s} \triangle_r, \tag{2.47}$$

with s such that $L^{-1}(s-) \leq t \leq L^{-1}(s)$, and for *each fixed* $\lambda > 0$

$$\{\hat{V}_\lambda(t),\, 0 \leq t < b_\lambda\} \overset{d}{=} \{\hat{V}(t),\, 0 \leq t < b_\lambda\}. \tag{2.48}$$

We are going to prove that $\{\hat{V}(t),\, t \geq 0\}$ is the $\hat{W}^+$-process. Let $\tilde{b}_\lambda$ be the unique t such that $w(t) = N(\lambda)$, define the reversed pre-minimum process $\{\tilde{V}_\lambda(t),\, 0 \leq t < \tilde{b}_\lambda\}$ in a way similar to (2.43) and set $\bar{b}_\lambda = L^{-1}(\bar{s}_\lambda -)$ where $\bar{s}_\lambda$ is determined by $L^{-1}(\bar{s}_\lambda -) \leq \lambda \leq L^{-1}(\bar{s}_\lambda)$. Then as in the case of (2.48) we have

$$\{\tilde{V}_\lambda(t),\, 0 \leq t < \tilde{b}_\lambda\} \overset{d}{=} \{\hat{V}(t),\, 0 \leq t < \bar{b}_\lambda\}.$$

On the other hand $\{\tilde{V}_\lambda(t)\}$ converges in law to $\hat{W}^+$ as $\lambda \to \infty$ by Bertoin [1, Cor.3.2,Th.3.4]. See also Chaumont [5, Th.2]. ($\{\tilde{V}_\lambda(t)\}$ is identical in law to the post-minimum process for the dual process $\hat{W}$ while this will not be true for $\{\hat{V}_\lambda(t)\}$. We may also use Millar [14]; in this case it is better to define $\{\tilde{V}_\lambda(t)\}$ and the related quantities by replacing the constant time λ in $w(t) = N(\lambda)$ and $L^{-1}(\bar{s}_\lambda -) \leq \lambda \leq L^{-1}(\bar{s}_\lambda)$ with an exponential random time of mean λ and independent of W.) Therefore $\{\hat{V}(t),\, t \geq 0\}$ is $\hat{W}^+$ and we have the following:

Theorem 6. *Under the conditions* (A) *and* (B) $\{\hat{V}(t),\, t \geq 0\}$ *defined by* (2.47) *is the* $\hat{W}^+$*-process and* (2.48) *holds for each fixed* λ. *In particular,* $\hat{V}_\lambda$ *converges in law to* $\hat{W}^+$ *as* $\lambda \to \infty$.

2.3.2 *The post-minimum process.* Since we have no formula for V_λ like (2.48), we need extra arguments for the proof of the convergence in law of V_λ. And, assuming only the conditions (A) and (B) we did not succeed (the part (ii) of Theorem 7 in [24] lacked a complete proof); we had unexpected difficulty in proving the tightness concerning $\{V_\lambda\}$ and so we must assume the additional condition (C) below which is somewhat stronger. For $0 < x < \lambda$ let $h_\lambda(x)$ denote *the probability that* $x + w(t)$ *enters* (λ, ∞) *before it enters* $(-\infty, 0)$ and let us state the following lemma.

Lemma 6. *If W satisfies the condition* (A) *and* $\sup w(t) = \infty$ *a.s., then*

$$\lim_{\lambda\to\infty} h_\lambda(x)^{-1} h_\lambda(y) = h(x)^{-1}h(y), \quad x > 0,\, y > 0. \tag{2.49}$$

Outline of proof. We first prepare the result for a random walk analogous to (2.49), in which the right hand side is replaced by the ratio of certain renewal functions (e.g. see [13, Theorem 2.3, p.524]). This ratio is again replaced by the ratio of certain mean occupation measures of the reflecting dual random walk which are defined in a manner similar to (2.18). To go to the case of a Lévy process we make use of the uniform approximation of W by suitable step processes of semi-Markov type.

Remark. When $\sup w(t) < \infty$ a.s. contrary to the assumption of the lemma Hirano [10] proved, under some additional condition, that the limit in (2.49) exists but the equality does not hold so that there are two different processes conditioned to stay positive attached to the same W.

From Lemma 6 it follows that

$$\frac{h_\lambda(1)^{-1}h_\lambda(x)}{h(1)^{-1}h(x)} \to 1 \quad \textit{as } \lambda \to \infty, \tag{2.50}$$

uniformly on any compact subset of $(0, \infty)$.

Condition (C). *The convergence in* (2.50) *is uniform on* $(0, a]$ *for any* $a > 0$.

Theorem 7. *Under* (A),(B) *and* (C), *the post-minimum process* V_λ *converges in law to* W^+ *as* $\lambda \to \infty$.

Key of proof. (i) By Lemma 6 the transition function of V_λ tends to that of W^+ as $\lambda \to \infty$, and (ii) the family of laws of $V_\lambda, \lambda > \lambda_0$, is tight; we used the condition (C) to check this.

§3. Diffusion processes in Lévy environments

In this section we write $w(x)$ instead of $w(t)$. Suppose, as stated in § **1**, that we are given the reflecting diffusion process $\{X(t),\, t \geq 0,\, \mathbb{P}\}$ in a Lévy environment $W = \{w, P\}$.

3.1 Let $w \in \mathbf{W}$ and $x > 0$. Then w is said to be locally right-oscillating (resp. locally left-oscillating) at x if $\sup\{w(y) : x < y < x + \epsilon\} > w(x)$ and $\inf\{w(y) : x < y < x + \epsilon\} < w(x)$ for any $\epsilon > 0$ (resp. if $\sup\{w(y) : x - \epsilon < y < x\} > w(x-)$ and $\inf\{w(y) : x - \epsilon < y < x\} < w(x-)$ for any $\epsilon > 0$). w is said to be locally oscillating at x if it is locally right- and left-oscillating at x. w is said to have a local maximum (resp. local minimum) at x if $\sup\{w(y),\ x-\epsilon<y<x+\epsilon\}=w(x)\vee w(x-)$ (resp. if $\inf\{w(y):\ x-\epsilon<y<x+\epsilon\}=w(x)\wedge w(x-)$) for some ϵ. An extreme point is a point of either local maximum or local minimum.

The following lemma can be proved in the same way as Lemma 3.1 of [13].

Lemma 7. *If the conditions* (A) and (B) *are satisfied, then there exists* $\mathbf{W}_0\subset\mathbf{W}$ *with* $P(\mathbf{W}_0)=1$ *such that any* $w\in\mathbf{W}_0$ *has the following properties:*
(i) $\tau=\hat{\tau}=0$.
(ii) $\sup\{w(x):\ x>0\}=-\inf\{w(x):\ x>0\}=\infty$.
(iii) *w can not have the same value at distinct extreme points.*
(iv) *w is locally oscillating at any point of discontinuity. In particular, w is continuous at any point of local minimum.*

We assume that W satisfies the conditions (A) and (B). We denote by μ_λ^w the distribution of $X(e^\lambda)-b_\lambda$ under P^w and by ν_λ^w the probability measure on $[-b_\lambda,a_\lambda-b_\lambda]$ defined by

$$\nu_\lambda^w(dx)=Z_{\lambda,w}^{-1}\exp\{-(w(x+b_\lambda)-w(b_\lambda))\}dx, \tag{3.1}$$

where $Z_{\lambda,w}=\int_{-b_\lambda}^{c_\lambda}\exp\,dx$ (normalization), a_λ and b_λ are defined by (1.2) and (1.3), and $c_\lambda=a_\lambda-b_\lambda$. In what follows $\|\cdot\|$ stands for the total variation. In computing $\|\mu_\lambda^w-\nu_\lambda^w\|$ we regard ν_λ^w as a probability measure in $(-\infty,\infty)$. Such a convention is often used. Note that $\|\mu_\lambda^w-\nu_\lambda^w\|$ is a random variable on the probability space $(\mathbf{W},P)$. For $\alpha>0,\lambda>0$ and $w\in\mathbf{W}$ we define $w_\lambda^\alpha\in\mathbf{W}$ by $w_\lambda^\alpha(x)=\lambda^{-1}w(\lambda^\alpha x)$, $x\geq 0$. Then $W_\lambda^\alpha=\{w_\lambda^\alpha(x),\ x\geq 0,\ P\}$ is a Lévy process.

Here are the conditions often used in the arguments to follow.

Condition (D_Λ). *Let $\Lambda=\{\lambda_n\}$ be a given positive sequence tending to ∞ and let it be fixed. There exists $\alpha>0$ such that W_λ^α converges in law, as $\lambda\to\infty$ along Λ, to some Lévy process $\tilde{W}=\{w(x),\ x\geq 0,\ \tilde{P}\}$ satisfying the conditions* (A) *and* (B).

Condition (D). *For any positive sequence $\{\lambda_n'\}$ tending to ∞ there exists a subsequence $\Lambda=\{\lambda_n\}$ of $\{\lambda_n'\}$ for which the condition (D_Λ) is satisfied.*

Most of strictly semi-stable Lévy processes satisfy the condition (D). A simple example of W satisfying (D) but is not strictly semi-stable is a Lévy process W with characteristic function

$$E\{e^{i\xi w(1)}\}=\exp\int_0^\infty(\cos i\xi x-1)x^{-\alpha-1}a(x)\,dx\,,$$

where $0<\alpha<2$ and $a(x)$ is a Borel function such that $a(e^t)$ is aperiodic in t and bounded from above and below by positive constants.

Theorem 8. *Suppose that W satisfies the conditions* (A),(B) *and* (D_Λ). *Then* $\|\mu_\lambda^w - \nu_\lambda^w\| \to 0$ *in probability with respect P as $\lambda \to \infty$ along Λ. If, in addition,* (D) *is satisfied, then the phrase "along Λ" is removed.*

3.2 This subsection is for preliminaries to the proof Theorem 8. Let $\Lambda = \{\lambda_n\}$, W_λ^α and $\tilde{W} = \{w(x),\ x \geq 0,\ \tilde{P}\}$ be the ones in the condition (D_Λ). Then by Lemma 7 there exists $\mathbf{W}_0 \subset \mathbf{W}$ with $P(\mathbf{W}_0) = \tilde{P}(\mathbf{W}_0) = 1$ such that any $w \in \mathbf{W}_0$ has the properties (i) $\sim$ (iv) of Lemma 7. Take an arbitrary $w \in \mathbf{W}_0$ and then let $\{w_n,\ n \geq 1\}$ be any sequence in $\mathbf{W}$ converging to w in the Skorohod topology. In the argument of this subsection $\{\lambda_n\}$, w and $\{w_n\}$ are all fixed.

We set $a = a_1$ and $b = b_1$ suppressing the suffix $\lambda = 1$. Then for any small $\epsilon > 0$ there exists a' with the following properties:

(i′) $a < a' < a + \epsilon$. (ii′) w is continuous at a'.

(iii′) $w(a) - \epsilon < w(x) < w(a')$ for any $x \in [a, a')$.

We set $d' = w(a') - w(b)$ and $e' = \sup\{w(y) - w(x) : 0 \leq x < y \leq b\}$. Then $d' > 1$ and $e' < 1$ (as for the latter we have to take $\mathbf{W}_0$ so that $e' < 1$ holds for any $w \in \mathbf{W}_0$ but this is certainly possible). We now employ the coupling method of Brox [3]. We use the notation $\omega(t)$ instead of $X(t)$ for the time being. Consider the product probability measure $P_n^\otimes = P^{\lambda_n w_n} \otimes \hat{P}_n$ on $\Omega \times \hat{\Omega}$ where $\hat{\Omega} = C([0,\infty) \to [0,a'])$ and $\hat{P}_n$ is the probability measure on $\hat{\Omega}$ with respect to which the coordinate process $\{\hat{\omega}(t),\ t \geq 0\}$ is a stationary reflecting diffusion process on $[0,a']$ with (local) generator

$$\frac{1}{2} e^{\lambda_n w_n} \frac{d}{dx}\left(e^{-\lambda_n w_n} \frac{d}{dx}\right).$$

Let $\hat{\nu}^{w_n}$ be the distribution (on $[0,a']$) of $\hat{\omega}(t)$ under $\hat{P}_n$; it is independent of t and has the density const.$\exp\{-\lambda_n w_n(x)\}$, $0 \leq x \leq a'$. We set

$$\tau' = \inf\{t > 0 : \omega(t) = a'\}, \quad \hat{\tau}' = \inf\{t > 0 : \hat{\omega}(t) = a'\},$$
$$\tilde{\sigma} = \inf\{t > 0,\ \omega(t) = \hat{\omega}(t)\}, \quad \tilde{\tau} = \inf\{t > \tilde{\sigma} : \hat{\omega}(t) = a'\}.$$
$$\tilde{\omega}(t) = \begin{cases} \omega(t) & \text{if } 0 \leq t < \tilde{\sigma}, \\ \hat{\omega}(t) & \text{if } t \geq \tilde{\sigma}. \end{cases}$$

Note that $\tilde{\sigma} \leq \tau'$ and $\hat{\tau}' \overset{d}{\leq} \tau'$. The following lemma can be proved as in Brox [3] (see also [12, p.179]).

Lemma 8. (i) *The process $\{\omega(t),\ 0 \leq t < \tau',\ P^{\lambda_n w_n}\}$ is equivalent in law to $\{\tilde{\omega}(t),\ 0 \leq t < \tilde{\tau},\ P_n^\otimes\}$.*

(ii) *Let* $e' < r < d'$. *Then as* $n \to \infty$ *we have*, $P_n^{\otimes}\{\tilde{\sigma} < e^{\lambda_n r}\} \to 1$ *and*

$$(3.2) \qquad P_n^{\otimes}\{\tilde{\tau} > e^{\lambda_n r}\} = P^{\lambda_n w_n}\{\tau' > e^{\lambda_n r}\} \geq \hat{P}_n\{\hat{\tau}' > e^{\lambda_n r}\} \to 1.$$

Denote by $E^{\lambda_n w_n}, \hat{E}_n$ and $E_n^{\otimes}$ the expectations with respect to $P^{\lambda_n w_n}, \hat{E}_n$ and $P_n^{\otimes}$, respectively. Then for any Borel function f in $[0,\infty)$ with $|f| \leq 1$ and for any positive sequence $\{r_n\}$ tending to 1 we have

$$\begin{aligned} E^{\lambda_n w_n}\{f(\omega(e^{\lambda_n r_n}))\} &= E_n^{\otimes}\{f(\tilde{\omega}(e^{\lambda_n r_n})); \tilde{\sigma} < e^{\lambda_n r_n} < \tilde{\tau}\} + \mathrm{o}(1) \\ = E_n^{\otimes}\{f(\hat{\omega}(e^{\lambda_n r_n}))\} + \mathrm{o}(1) &= \int_{[0,a']} f\, d\hat{\nu}^{w_n} + \mathrm{o}(1), \end{aligned}$$

where o(1), which may vary from place to place, denotes a term whose absolute value is dominated by some ϵ_n independent of f and tending to 0 as $n \to \infty$. Therefore, if μ^{w_n} denotes the distribution of $\omega(e^{\lambda_n r_n})$ under $P^{\lambda_n w_n}$, then $\|\mu^{w_n} - \hat{\nu}^{w_n}\| \to 0$ as $n \to \infty$. This can be rephrased as (3.3) below. Let ν^{w_n} be the probability measure on $[0,a]$ with density const.exp$\{-\lambda_n w_n(x)\}$, $0 \leq x \leq a$. Since a' can be taken arbitrarily close to a and since $w_n \to w$ (the Skorohod convergence), we have $\|\hat{\nu}^{w_n} - \nu^{w_n}\| \to 0$ as $n \to \infty$ and hence

$$(3.3) \qquad \|\mu^{w_n} - \nu^{w_n}\| \to 0 \quad \text{as } n \to \infty.$$

3.3 We proceed to the proof of Theorem 8. From now on we use $X(t)$ for $\omega(t)$. As in [3] we have $\{\lambda^{-\alpha}X(\lambda^{2\alpha}t),\, t \geq 0,\, P^w\} \overset{d}{=} \{X(t),\, P^{\lambda w_\lambda^\alpha}\}$ and so

$$\{X(e^\lambda),\, P^w\} \overset{d}{=} \{\lambda^\alpha X(\lambda^{-2\alpha}e^\lambda),\, P^{\lambda w_\lambda^\alpha}\} = \{\lambda^\alpha X(e^{\lambda r(\lambda)}),\, P^{\lambda w_\lambda^\alpha}\},$$

where $r(\lambda) = 1 - 2\alpha\lambda^{-1}\log\lambda$ which tends to 1 as $\lambda \to \infty$. Now let us denote by $\tilde{\mu}_\lambda^w$ the distribution of $X(e^{\lambda r(\lambda)})$ under $P^{\lambda w_\lambda^\alpha}$ and by $\tilde{\nu}_\lambda^w$ the probability measure on $[0, a(w_\lambda^\alpha)]$ with density const.exp$\{-\lambda w_\lambda^\alpha(x)\}, 0 \leq x \leq a(w_\lambda^\alpha)$. Noting that W_λ^α converges in law to $\widetilde{W}$ as $\lambda \to \infty$ along $\{\lambda_n\}$, we first make use of Skorohod's realization theorem of almost sure convergence and then apply (3.3). As a result we have

$$(3.4) \qquad \|\tilde{\mu}_\lambda^w - \tilde{\nu}_\lambda^w\| \to 0 \quad \text{in probability as } \lambda \to \infty \text{ along } \{\lambda_n\}.$$

Since $a_\lambda(w) = \lambda^\alpha a(w_\lambda^\alpha), b_\lambda(w) = \lambda^\alpha b(w_\lambda^\alpha)$ and $\{X(e^\lambda) - b_\lambda,\, P^w\}$ is identical in law to $\{\lambda^\alpha(X(e^{\lambda r(\lambda)})) - b(w_\lambda^\alpha),\, P^{\lambda w_\lambda^\alpha}\}$, we have, for any

Borel function f in $\mathbf{R}$ with $|f| \leq 1$ and as $\lambda \to \infty$ along $\{\lambda_n\}$,

$$\int_{[-b_\lambda,\infty)} f\, d\mu_\lambda^w = \int_{[0,\infty)} f(\lambda^\alpha(x - b(w_\lambda^\alpha)))\tilde{\mu}_\lambda^w(dx) \tag{3.5}$$

$$= \int_{[0,a(w_\lambda^\alpha)]} f(\lambda^\alpha(x - b(w_\lambda^\alpha)))\tilde{\nu}_\lambda^w(dx) + \mathrm{o}(1) \tag{3.6}$$

$$= \text{const.} \int_{[0,a(w_\lambda^\alpha)]} f(\lambda^\alpha(x - b(w_\lambda^\alpha)))e^{-w(\lambda^\alpha x)}dx + \mathrm{o}(1) \tag{3.7}$$

$$= \text{const.}\lambda^{-\alpha} \int_{[0,a_\lambda]} f(x - b_\lambda)e^{-w(x)}dx + \mathrm{o}(1) \tag{3.8}$$

$$= \int_{[-b_\lambda,a_\lambda - b_\lambda]} f\, d\nu_\lambda^w + \mathrm{o}(1), \tag{3.9}$$

where we used (3.4) for (3.6), the definition of $\tilde{\nu}_\lambda^w$ for (3.7), change of variable for (3.8) and the definition (3.1) of ν_λ^w for (3.9). The proof of Theorem 8 is finished.

3.4 Let $\overline{\mathbf{W}}$ be the space of those nonnegative functions w in $\mathbf{R}$ which are right continuous and have left limits with $w(0) = w(0-) = 0$. Let $\bar{P}$ be the probability measure on $\overline{\mathbf{W}}$ such that $\bar{W}^- = \{w(-x-),\ x \geq 0\}$ is $\hat{W}^+$, $\bar{W}^+ = \{w(x),\ x \geq 0\}$ is W^+ and $\bar{W}^-$ and $\bar{W}^+$ are independent. The following lemma can be proved by making use of (2.13).

Lemma 9. *Under* (A) *and* (B), $\bar{E}\{\int_{-\infty}^{\infty} e^{-w(x)}\, dx\} < \infty$.

By this lemma we can define a probability measure $\bar{\nu}^w$ in $\mathbf{R}$, with superfix w outside some $\bar{P}$-negligible subset of $\overline{\mathbf{W}}$, and then $\bar{\nu}$ by

$$\bar{\nu}^w(dx) = Z_w^{-1}e^{-w(x)}dx \quad (Z_w = \int_{-\infty}^{\infty} e^{-w(x)}dx), \quad \bar{\nu} = \int \bar{\nu}^w \bar{P}(dw).$$

Of course ν_λ^w and $\bar{\nu}^w$ are random variables taking values of probability measures in $\mathbf{R}$; the former is governed by P and the latter by $\bar{P}$. From Theorem 6 and 7 it will be expected that ν_λ^w converges in law to $\bar{\nu}^w$ as $\lambda \to \infty$ but, to verify this, we still have to assume the following condition.

Condition (E). *There is a constant C such that, for any $x > 0$ and $y > 0$, the inequality $h_\lambda(x)^{-1}h_\lambda(y) \leq C\, h(x)^{-1}h(y)$ holds.*

In the following theorem μ_λ denotes the distribution of $X(e^\lambda) - b_\lambda$ under $\mathbb{P}$, namely, $\mu_\lambda = \int \mu_\lambda^w P(dw)$.

Theorem 9. *Under the conditions* (A)$\sim$ (E) ν_λ^w *and hence* μ_λ^w, *by* Theorem 8, *converge in law to* $\bar{\nu}^w$ *as* $\lambda \to \infty$. *In particular,* μ_λ *converges to* $\bar{\nu}$ *as* $\lambda \to \infty$.

For the proof we have to show the law convergence of $Z_{\lambda,w}$ (governed by P) to Z_w (governed by $\bar{P}$) as $\lambda \to \infty$. Set

$$Z^-_{\lambda,w} = \int_{-b_\lambda}^0 \exp\{-(w(x+b_\lambda) - w(b_\lambda))\}dx, \quad Z^+_{\lambda,w} = \int_0^{c_\lambda},$$

$$Z^-_w = \int_{-\infty}^0 e^{-w(x)}dx, \quad Z^+_w = \int_0^\infty e^{-w(x)}dx .$$

Then the law convergence of $Z^-_{\lambda,w}$ to Z^-_w follows immediately from Theorem 6 (in particular, from (2.48)). As for $Z^+_{\lambda,w}$, Theorem 7 alone is not enough; in fact, we have to show the uniform smallness (w.r.t. λ) of the tail $\int_{r\wedge c_\lambda}^{c_\lambda}$ for large r and this is done by using the inequality

$$(3.10) \qquad P\{V_\lambda(t_k) \in \Gamma_k,\ 1 \leq \forall k \leq n\} \leq C\bar{P}\{w(t_k),\ 1 \leq \forall k \leq n\},$$

where $0 \leq t_1 < t_2 < \cdots < t_n$ and Γ_k, $1 \leq k \leq n$, are Borel sets in $(0,\infty)$. The condition (E) is used for (3.10).

Remark. Our arguments remain valid when the conditions (D_Λ) and (D) are replaced by the following.

Condition (D'_Λ). *Let $\Lambda = \{\lambda_n\}$ be a given positive sequence tending to ∞ and let it be fixed. There exists a positive sequence $\{\alpha_n\}$ with $\alpha_n = o(\lambda_n/\log\lambda_n)$ and such that $W^{\alpha_n}_{\lambda_n}$ converges in law, as $n \to \infty$, to some Lévy process $\tilde{W} = \{w(x),\ x \geq 0,\ \tilde{P}\}$ satisfying the conditions* (A) *and* (B).

Condition (D'). *For any positive sequence $\{\lambda'_n\}$ tending to ∞ there exists a subsequence $\Lambda = \{\lambda_n\}$ of $\{\lambda'_n\}$ for which the condition (D'_Λ) is satisfied.*
Thus Theorem 9 still holds when the condition (D) is replaced by (D′). On the other hand the conditions (C) and (E) seem too strong and it is desirable to remove or relax these conditions.

Examples. (i) Let W be a strictly stable Lévy process with exponent $\alpha \in (0,2)$ such that $0 < \rho = P\{w(1) > 0\} < 1$. Then W satisfies all the conditions (A) $\sim$ (E) and $h(x) = \text{const.}x^{\alpha(1-\rho)}$. The verification of (C) and (E) is done by using, in detail, the explicit formula on $h_\lambda(x)$ obtained by Rogozin [17].
(ii) Spectrally negative Lévy processes satisfy the conditions (C) and (E) since $h_\lambda(x)^{-1}h_\lambda(y) = h(x)^{-1}h(y)$, $x,y \in (0,\lambda)$, for any λ. So any Lévy process W such that

$$E\{e^{i\xi w(1)}\} = \exp\int_{-\infty}^0 (e^{i\xi x} - 1 - i\xi x) \mid x \mid^{-\alpha-1} a(x)\, dx,$$

with $1 < \alpha < 2$ and $0 < c_1 \leq a(x) \leq c_2$, satisfies all the conditions (A)$\sim$(E).

References

[1] J. Bertoin, *Splitting at the infimum and excursions in half-lines for random walks and Lévy processes*, Stoch. Proc. Appl. 47(1993), 17–35.

[2] J. Bertoin, *Lévy Processes*, Cambridge Univ. Press, Cambridge, 1996.

[3] T. Brox, *A one-dimensional diffusion process in Wiener medium*, Ann. Probab. **14**(1986), 1206–1218.

[4] L. Chaumont, *Sur certains processus de Lévy conditionnés rester à positifs*, Stoch. Stoch. Rep. 47(1994), 1–20.

[5] L. Chaumont, *Conditionings and path decompositions for Lévy processes*, Stoch. Proc. Appl. 64(1996), 39–54.

[6] R. Doney, *Fluctuation theory for Lévy processes*, Lévy Processes: Theory and Application (ed. O. Barndorff-Nielsen et al.), 57–66, Birkhäuser, Boston, 2001.

[7] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd ed., Wiley, New York, 1971.

[8] A.O. Golosov, *Localization of random walks in one-dimensional random environments*, Comm. Math. Phys. **92**(1984), 491–506.

[9] P. Greenwood and J. Pitman, *Fluctuation identities for Lévy processes and splitting at the maximum*, Adv. Appl. Probab. **12**(1980), 893–902.

[10] K. Hirano, *Lévy processes with negative drift conditioned to stay positive*, Tokyo J. Math. **24**(2001), 291–308.

[11] K. Itô, *Poisson point processes attached to Markov processes*, Proc. 6th Berkeley Symp. Math. Statist. Probab. **III**, 1970, 225–239.

[12] K. Kawazu, Y. Tamura and H. Tanaka, *One-dimensional diffusions and random walks in random environments*, Probab. Th. Math. Statist., 5th Japan-USSR Symp. Proc., 1986, Lecture Notes in Math. **1299** (ed. S. Watanabe and Yu.V. Prokhorov), 170–184, Springer-Verlag, 1988.

[13] K. Kawazu, Y. Tamura and H. Tanaka, *Localization of diffusion processes in one-dimensional random environment*, J. Math. Soc. Japan **44**(1992), 515–550.

[14] P.W. Millar, *A path decomposition for Markov processes*, Ann. Probab. **6**(1978), 345–348.

[15] E.A. Pecherskii and B.A. Rogozin, *On joint distribution of random variables associated with fluctuations of a process with independent increments*, Theory Probab. Appl. **14**(1969), 410–423.

[16] D. Ray, *Stable processes with an absorbing barrier*, Trans. Amer. Math. Soc. **89**(1958), 16–24.

[17] B.A. Rogozin, *The distribution of the first hit for stable and asymptotically stable walks on an interval*, Theor. Probab. Appl. **17**(1972), 332–338.

[18] K. Sato, *Lévy processes and Infinitely Divisible Distributions*, Cambridge Univ. Press, Cambridge, 1999.

[19] S. Schumacher, *Diffusions with random coefficients*, Particle Systems, Random Media and Large Deviations (ed. R. Durrett), Contenmp. Math. **41**, 351–35, Amer. Math. Soc. 1985.

[20] M.L. Silverstein, *Classification of coharmonic and coinvariant functions for a Lévy process*, Ann. Probab. **8**(1980), 539–575.

[21] Ya.G. Sinai, *The limiting behavior of a one-dimensional random walk in a random medium*, Theor. Probab. Appl. **27**(1982), 256–268.

[22] H. Tanaka, *Limit theorem for one-dimensional diffusion process in Brownian environment*, Stochastic Analysis, Lecture Notes in Math. **1322** (ed. M. Métivier and S. Watanabe), 156–172, Springer-Verlag, 1988.

[23] H. Tanaka, *Green operators of absorbing Lévy processes on the half line*, Stochastic Processes. A Festschrift in Honour of Gopinath Kallianpur (ed. S. Cambanis et al.), 313–319, Springer, New York, 1993.

[24] H. Tanaka, *Superharmonic transform of absorbing Lévy processes* (in Japanese), Additive Processes and Related Topics. Cooperative Research Report 51, 13–25, The Institut. Statist. Math., Tokyo, 1993.

[25] H. Tanaka, *Localization of a diffusion process in a one-dimensional Brownian environment*, Comm. Pure Appl. Math. **47**(1994), 755–766.

1-4-17-104 Miyamaedaira, Miyamae-ku
Kawasaki 216-0006
Japan

Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 377–395

Zero-Range-Exclusion Particle Systems

Kôhei Uchiyama

§1. Introduction

Let $\mathbf{T}_N$ denote the one-dimensional discrete torus $\mathbf{Z}/N\mathbf{Z}$ represented by $\{1, ..., N\}$. The zero-range-exclusion process that we are to introduce and study in this article is a Markov process on the state space $\mathcal{X}^N := \mathbf{Z}_+^{\mathbf{T}_N}$ $(\mathbf{Z}_+ = \{0, 1, 2, ...\})$. Denote by $\eta = (\eta_x, x \in \mathbf{T}_N)$ a generic element of $\mathcal{X}^N$, and define

$$\xi_x = \mathbf{1}(\eta_x \geq 1)$$

(namely, ξ_x equals 0 or 1 according as η_x is zero or positive). The process is regarded as a 'lattice gas' of particles having energy. The site x is occupied by a particle if $\xi_x = 1$ and vacant otherwise. Each particle has energy, represented by η_x, which takes discrete values $1, 2, \ldots$ If y is a nearest neighbor site of x and is vacant, a particle at site x jumps to y at rate $c_{\text{ex}}(\eta_x)$, where c_{ex} is a positive function of $k = 1, 2, \ldots$ Between two neighboring particles the energies are transferred unit by unit according to the same stochastic rule as that of the zero-range processes. In this article we shall give some results related to the hydrodynamic scaling limit for this model.

To give a formal definition of the infinitesimal generator of the process we introduce some notations. Let $b = (x, y)$ be an oriented bond of $\mathbf{T}_N$, namely x and y are nearest neighbor sites of $\mathbf{T}_N$, and (x, y) stands for an ordered pair of them. Define the *exclusion* operator π_b and *zero-range* operator ∇_b attached to b which act on $f \in C(\mathcal{X}^N)$ by

$$\pi_b f(\eta) = f(S^b_{\text{ex}}\eta) - f(\eta) \quad \text{and} \quad \nabla_b f(\eta) = f(S^b_{\text{zr}}\eta) - f(\eta)$$

where the transformation $S^b_{\text{ex}} : \mathcal{X}^N \mapsto \mathcal{X}^N$ is defined by

$$(S^b_{\text{ex}}\eta)_z = \begin{cases} \eta_y, & \text{if } z = x, \\ \eta_x, & \text{if } z = y, \\ \eta_z, & \text{otherwise,} \end{cases}$$

Received December 26, 2002.
Revised March 24, 2003.

if $\xi_x = 1$ and $\xi_y = 0$; and $S^b_{\text{zr}}\eta$ by

$$(S^b_{\text{zr}}\eta)_z = \begin{cases} \eta_x - 1, & \text{if } z = x, \\ \eta_y + 1, & \text{if } z = y, \\ \eta_z, & \text{otherwise,} \end{cases}$$

if $\eta_x \geq 2$ and $\xi_y = 1$; and in the remaining case of η, both $S^b_{\text{ex}}\eta$ and $S^b_{\text{zr}}\eta$ are set to be η, namely

$$S^b_{\text{ex}}\eta = \eta \quad \text{if} \quad \xi_x(1 - \xi_y) = 0,$$
$$S^b_{\text{zr}}\eta = \eta \quad \text{if} \quad \mathbf{1}(\eta_x \geq 2)\xi_y = 0.$$

Let c_{ex} and c_{zr} be two non-negative functions on $\mathbf{Z}_+$ and define for $b = (x, y)$

$$L_b = c_{\text{ex}}(\eta_x)\pi_b + c_{\text{zr}}(\eta_x)\nabla_b.$$

Let $\mathbf{T}^*_N$ denote the set of all oriented bonds in $\mathbf{T}_N$:

$$\mathbf{T}^*_N = \{b = (x, y) : x, y \in \mathbf{T}_N, |x - y| = 1\}.$$

Then the infinitesimal generator L_N of our Markovian particle process on $\mathbf{T}_N$ is given by

$$L_N = \sum_{b \in \mathbf{T}^*_N} L_b.$$

It is assumed that for some positive constant a_0, $c_{\text{ex}}(k) \geq a_0$ for $k \geq 1$ and $c_{\text{zr}}(k) \geq a_0$ for $k \geq 2$. This especially implies that the lattice gas on $\mathbf{T}_N$ with both the number of particles and the total energy being given is ergodic. We call the Markov process generated by L_N the *zero-range-exclusion* process. For the sake of convenience we set

$$c_{\text{ex}}(0) = 0 \quad \text{and} \quad c_{\text{zr}}(0) = c_{\text{zr}}(1) = 0.$$

We need some technical conditions on the functions c_{ex} and c_{zr}: there exist positive constants a_1, a_2, a_3, a_4 and an integer k_0 such that

$$(1) \qquad |c_{\text{zr}}(k) - c_{\text{zr}}(k + 1)| \leq a_1 \quad \text{for all } k \geq 1;$$

$$(2) \qquad c_{\text{zr}}(k) - c_{\text{zr}}(l) \geq a_2 \quad \text{whenever } k \geq l + k_0;$$

$$(3) \qquad a_3 k \leq c_{\text{ex}}(k) \leq a_4 k \quad \text{for all } k \geq 1.$$

These conditions are imposed mainly for guaranteeing an estimate of the spectral gaps for the local processes ([4]). The conditions (1) and

(2) are the same as in the paper [2] where is carried out an estimation of the spectral gap for the zero-range processes.

We shall also write $\pi_{x,y}, S_{\text{ex}}^{x,y}, L_{x,y}$, etc. for $\pi_b, S_{\text{ex}}^b, L_b$, etc.

Grand Canonical Measures and Dirichlet Form.

For a pair of constants $0 < p < 1$ and $\rho > p$ let $\nu_{p,\rho} = \nu_{p,\rho}^{\mathbf{T}_N}$ denote the product probability measure on $\mathcal{X}^N$ whose marginal laws are given by

$$\nu_{p,\rho}(\{\eta : \eta_x = l\}) := \begin{cases} 1-p & \text{if } l = 0, \\ \dfrac{p}{Z_{\lambda(p,\rho)}} & \text{if } l = 1, \\ \dfrac{p}{Z_{\lambda(p,\rho)}} \cdot \dfrac{(\lambda(p,\rho))^{l-1}}{c_{\text{zr}}(2)c_{\text{zr}}(3)\cdots c_{\text{zr}}(l)} & \text{if } l \geq 2, \end{cases}$$

for all x. Here $Z_\lambda := 1 + \sum_{l=2}^{\infty} \dfrac{\lambda^{l-1}}{c_{\text{zr}}(2)c_{\text{zr}}(3)\cdots c_{\text{zr}}(l)}$ and $\lambda(p,\rho)$ is a positive constant depending on p and ρ and determined uniquely by the relation $E^{\nu_{p,\rho}}[\eta_x] = \rho$, where $E^{\nu_{p,\rho}}$ denotes the expectation under the law $\nu_{p,\rho}$. Clearly $E^{\nu_{p,\rho}}[\xi_x] = p$. The lattice gas is reversible relative to the measures $\nu_{p,\rho}$ (namely L_N is symmetric relative to each of them).

It is convenient to introduce the transformations S^b, $b = (x,y)$ which acts on $\eta \in \mathcal{X}^N$ according to

$$S^b\eta = \begin{cases} S_{\text{ex}}^b\eta & \text{if } \xi_y = 0, \\ S_{\text{zr}}^b\eta & \text{if } \xi_y = 1, \end{cases}$$

and the operators

$$\Gamma_b = \xi_x\pi_b + \mathbf{1}(\eta_x \geq 2)\nabla_b \qquad (b = (x,y)).$$

The latter may also be defined by $\Gamma_b f(\eta) = f(S^b\eta) - f(\eta)$ $(f \in C(\mathcal{X}^N))$. Let $\tau_x\eta$ be the configuration $\eta \in \mathcal{X}$ viewed from x, namely $(\tau_x\eta)_y = \eta_{x+y}$. We let it also act on a function f of η according to $\tau_x f(\eta) = f(\tau_x\eta)$. Setting

$$c_{01}(\eta) = c_{\text{ex}}(\eta_0)(1-\xi_1) + c_{\text{zr}}(\eta_0)\xi_1;$$
$$c_{10}(\eta) = c_{\text{ex}}(\eta_1)(1-\xi_0) + c_{\text{zr}}(\eta_1)\xi_0;$$

and $c_{x,x+1} = \tau_x c_{01}$, $c_{x+1,x} = \tau_x c_{10}$, we can write

$$L_b = c_b\Gamma_b.$$

The Dirichlet form is then given by

$$\mathcal{D}^{p,\rho}\{f\} = \sum_{b\in\mathbf{T}_N^*} E^{\nu_{p,\rho}}[(\Gamma_b f)^2 c_b].$$

(Functions f of configuration η will be always real in this article.)

Diffusion Coefficient Matrix.

Following Varadhan [7] we define the diffusion coefficient matrix. First we introduce some notations. Let $\mathcal{X}$ denote $\mathbf{Z}_+^{\mathbf{Z}}$, the set of all configurations on $\mathbf{Z}$ and $\mathcal{F}_c$ the set of all local functions on $\mathcal{X}$ (namely, $f \in \mathcal{F}_c$ if f depends only on a finite number of coordinates of $\eta \in \mathcal{X}$). For $f \in \mathcal{F}_c$ we use the symbol $\tilde{f}$ to represent the formal sum $\sum_x \tau_x f$. It has meaning if Γ_{01} is acted:

$$\Gamma_{01}\tilde{f} = \sum_x \Gamma_{01}\tau_x f = \sum_x \tau_x \Gamma_{x,x+1} f,$$

where the infinite sums are actually finite sums. Let $\chi(p,\rho)$ denote the covariance matrix of ξ_0 and η_0 under $\nu_{p,\rho}$:

$$\chi(p,\rho) = \begin{pmatrix} (1-p)p & (1-p)\rho \\ (1-p)\rho & E^{\nu_{p,\rho}}|\eta_0-\rho|^2 \end{pmatrix}$$

For each $0 < p < 1, \rho > p$, let $\hat{c}(p,\rho) = (\hat{c}^{i,j}(p,\rho))_{1\le i,j\le 2}$ denote a 2×2 symmetric matrix whose quadratic form is defined by the following variational formula:

$$\begin{aligned} \underline{\alpha}\cdot\hat{c}(p,\rho)\underline{\alpha} &= \hat{c}^{11}(p,\rho)\alpha^2 + 2\hat{c}^{12}(p,\rho)\alpha\beta + \hat{c}^{22}(p,\rho)\beta^2 \\ &= \inf_{f\in\mathcal{F}_c} E^{\nu_{p,\rho}}\left[\left(\Gamma_{01}\{\alpha\xi_0 + \beta\eta_0 + \tilde{f}\,\}\right)^2 c_{01}\right] \end{aligned}$$

where $\underline{\alpha} = (\alpha,\beta)^T$, a two-dimensional real column vector (T indicates the transpose), and $\cdot$ indicates the inner product in $\mathbf{R}\times\mathbf{R}$. Then the diffusion coefficient matrix is defined by

$$D(p,\rho) = \hat{c}(p,\rho)\chi^{-1}(p,\rho),$$

where $\chi^{-1}(p,\rho)$ is the inverse matrix of $\chi(p,\rho)$. The two eigen-values of D are positive (cf. Section 5) and D is diagonalizable.

Let $\nabla^-\xi$ and $\nabla^-\eta$ be the particle and energy gradients:

$$\nabla^-\xi = \xi_0 - \xi_1 \quad \text{and} \quad \nabla^-\eta = \eta_0 - \eta_1$$

and w_{01}^P and w_{01}^E the particle and energy currents, respectively, from the site 0 to the site 1 :

$$w_{01}^P = -L_{\{0,1\}}\{\xi_0\} \quad \text{and} \quad w_{01}^E = -L_{\{0,1\}}\{\eta_0\}.$$

Here $L_{\{0,1\}} = L_{01} + L_{10}$. The explicit form of the currents are

$$\begin{aligned} w_{01}^P &= c_{\text{ex}}(\eta_0)(1-\xi_1) - c_{\text{ex}}(\eta_1)(1-\xi_0) \\ w_{01}^E &= c_{\text{ex}}(\eta_0)(1-\xi_1)\eta_0 + c_{\text{zr}}(\eta_0)\xi_1 - c_{\text{ex}}(\eta_1)(1-\xi_0)\eta_1 - c_{\text{zr}}(\eta_1)\xi_0. \end{aligned}$$

We can show that

$$\begin{pmatrix} w_{01}^P \\ w_{01}^E \end{pmatrix} - D(p,\rho)\begin{pmatrix} \nabla^- \xi \\ \nabla^- \eta \end{pmatrix} \in \overline{\left\{ \begin{pmatrix} Lf_1 \\ Lf_2 \end{pmatrix} : f_1, f_2 \in \mathcal{F}_c^K \text{ for some } K \in \mathbf{N} \right\}}^{p,\rho},$$

where $\overline{\{\cdots\}}^{p,\rho}$ is the closure relative to the central limit theorem variance $V^{p,\rho}$ (see Section 3). This would lead one to expect that the hydrodynamic equation for the limit densities $p = p(t,\theta)$ and $\rho = \rho(t,\theta)$ should be

$$\frac{\partial}{\partial t}\begin{pmatrix} p \\ \rho \end{pmatrix} = \frac{\partial}{\partial \theta} D(p,\rho) \frac{\partial}{\partial \theta}\begin{pmatrix} p \\ \rho \end{pmatrix}.$$

Unfortunately in deriving this equation there arises serious difficulty due to the unboundedness of the spin values. While the marginal of our grandcanonical measure is roughly Poisson, the energy current w_{01}^E involves the term $c_{\text{ex}}(\eta_0)\eta_0$ that is bounded below by $\delta\eta_0^2$ ($\delta > 0$) and cannot be controlled by the grandcanonical measure as in the case of Ginzburg-Landau model, the logarithm of the Poisson density function being of the order $O(\eta_0 \log \eta_0)$. Nagahata [3] studies a similar model and derives a system of diffusion equations of the same form as above: his model is the same as the present one except that the energy values are bounded by a constant.

In the rest of this article we shall state some results on the equilibrium fluctuations and the central limit theorem variances without proof, and give certain asymptotic estimates for the density-density correlation coefficients and for the least upper bound of the spectrum of an operator of the form $V_N + L$ as consequences of these results. In the last part of the paper some upper and lower bounds of the diffusion matrix will be given.

§2. Density-Density Correlation Function

Consider an infinite particle system on the whole lattice $\mathbf{Z}$ whose formal generator is $L = \sum c_b \Gamma_b$. It is well defined on $\mathcal{F}_c$:

$$Lf(\eta) = \sum_{b \in \mathbf{Z}^*} c_b(\eta)\Gamma_b f(\eta), \qquad f \in \mathcal{F}_c.$$

Let $\mathcal{F}_c^\circ$ be the set of all $f \in \mathcal{F}_c$ such that both f and Lf are in $L^2(\nu_{p,\rho}, \mathcal{X})$. Then the operator L with the domain $\mathcal{F}_c^\circ$ is a symmetric and non-negative transformation in $L^2(\nu_{p,\rho}, \mathcal{X})$. Clearly $\mathcal{F}_c^\circ$ is dense in $L^2(\nu_{p,\rho}, \mathcal{X})$. Hence L has the Friedrichs extension, which we denote by $\mathcal{L}$: namely $\mathcal{L}$ is the smallest self-adjoint extension of L. The following theorem is a consequence from the standard theory on the semigroup of operators. Let Λ_K be the finite interval $\{-K, \dots, K\}$ and $L_{\Lambda(K)}$ the generator of the lattice gas on Λ_K, namely

$$L_{\Lambda(K)} = \sum_{b \in \Lambda^*(K)} L_b;$$

also put $\mathcal{X}_{\Lambda(K)} = \mathbf{Z}_+^{\Lambda(K)}$. Here $\Lambda(K)$ is used in stead of Λ_K in sub- or superscripts and $\Lambda^*(K) = (\Lambda(K))^*$ (the set of all oriented bonds in Λ).

Theorem 1. *The operator $\mathcal{L}$ generates a strongly continuous Markov semigroup on $L^2(\nu_{p,\rho}, \mathcal{X})$. Denote by $S(t)$, $t \geq 0$ this semigroup, and by $S_K(t)$ the semigroup on $L^2(\mathcal{X}_{\Lambda(K)})$ generated by $L_{\Lambda(K)}$. Then*

$$\lim_{K \to \infty} S_K(t) f(\eta|_{\Lambda(K)}) = S(t) f(\eta), \qquad f \in \mathcal{F}_c^\circ,$$

strongly in $L^2(\nu_{p,\rho}, \mathcal{X})$. The convergence is locally uniform in t.

Fix $0 < p < 1$ and $\rho > p$. Let $\eta(t)$ be a Markov process on $\mathcal{X}$ whose infinitesimal generator and initial distribution are $\mathcal{L}$ and $\nu_{p,\rho}$, respectively. Denote the probability law of the process $\eta(t)$ by $P_{\mathrm{eq}} = P_{\mathrm{eq}(p,\rho)}$ and the expectation relative to it by $E_{\mathrm{eq}(p,\rho)}$. Define the fluctuation processes $Y_{t,N}^P$ and $Y_{t,N}^E$ by

$$Y_{t,N}^P(J) = \frac{1}{\sqrt{N}} \sum_{x \in \mathbf{Z}} J(x/N)(\xi_x(N^2 t) - p), \qquad J \in C_0^\infty(\mathbf{R}),$$

$$Y_{t,N}^E(J) = \frac{1}{\sqrt{N}} \sum_{x \in \mathbf{Z}} J(x/N)(\eta_x(N^2 t) - \rho), \qquad J \in C_0^\infty(\mathbf{R})$$

respectively. ($C_0^\infty(\mathbf{R})$ is the set of smooth functions with compact supports.) Under the equilibrium measure $P_{\mathrm{eq}(p,\rho)}$ the process $Y_{t,N} =$

$(Y^P_{t,N}, Y^E_{t,N})$ converges in the sense of finite dimensional distributions, namely for each set of $J_1, \dots, J_k \in C_0^\infty(\mathbf{R})$ and $t_1, \dots, t_k \in [0, \infty)$, the joint distribution of $Y_{t_1,N}(J_1), \dots, Y_{t_k,N}(J_k)$ converges ([6]). The limit process $Y_t = (Y_t^P, Y_t^E)$ is an infinite dimensional Ornstein-Uhlenbeck process. The distribution of Y_t is described as follows.

Let K_D denote the fundamental solution for the heat equation

$$\frac{\partial}{\partial t}\underline{u} = D^T \frac{\partial^2}{\partial \theta^2}\underline{u}$$

and U_t a matrix of corresponding convolution operators:

$$U_t \underline{J}(\theta) = \int_{-\infty}^{\infty} K_D(t, \theta - \theta') \underline{J}(\theta') d\theta',$$

where $\underline{J} = (J^1, J^2)^T \in C_0^\infty(\mathbf{R}) \times C_0^\infty(\mathbf{R})$. Let $\underline{J}_1$ and $\underline{J}_2$ be vector functions of the same kind. Then the distribution of the limit process Y_t is given by

$$E\Big[e^{i(Y_0, \underline{J}_1)} e^{i(Y_t, \underline{J}_2)}\Big] = \exp\Big[-\frac{1}{2}\int_0^t Q\{U_r \underline{J}_2\} dr - \frac{1}{2}\sigma^2\{U_t \underline{J}_2 + \underline{J}_1\}\Big];$$

in particular

$$(4) \quad E[(Y_0, \underline{J}_1)(Y_t, \underline{J}_2)] = \sigma^2(U_t \underline{J}_2, \underline{J}_1) = (\chi(p, \rho) U_t \underline{J}_2, \underline{J}_1)_{L^2(\mathbf{R})}.$$

Here E denotes the expectation by the probability law of the limit process and

$$Q\{\underline{J}\} = 2(\underline{J}', \hat{c}\underline{J}')_{L^2(\mathbf{R})}, \quad \sigma^2\{\underline{J}\} = (\underline{J}, \chi \underline{J})_{L^2(\mathbf{R})}.$$

(Also $(Y_t, \underline{J}) = Y_t^P(J_1) + Y_t^E(J_2)$, $(\underline{J}_1, \underline{J}_2)_{L^2(\mathbf{R})} = \int_{\mathbf{R}} (J_1^1 J_2^1 + J_1^2 J_2^2) d\theta$; $\hat{c} = \hat{c}(p, \rho)$ is the matrix appearing in the definition of $D = D(p, \rho)$; $\underline{J}'$ is the (component-wise) derivative of $\underline{J}$; $\sigma^2(\cdot, \cdot)$ is the bilinear form associated with the quadratic form $\sigma^2\{\cdot\}$.) The kernel K_D may be explicitly written down in the form

$$\begin{aligned} K_D(t, \theta) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\{-t\lambda^2 D^T\} e^{-i\lambda\theta} d\lambda \\ &= \sqrt{4\pi t D^T}^{-1} \exp\{-\theta^2 (4tD^T)^{-1}\}. \end{aligned}$$

Here D^T is the transpose of D; for a 2×2 real matrix A whose eigenvalues are positive,

$$\sqrt{A} := \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\{-\theta^2 A^{-1}\} d\theta,$$

which is a real matrix having positive eigenvalues such that $A = (\sqrt{A})^2$.

Define the symmetric matrix $\Sigma(x,t)$ with parameters $(x,t) \in \mathbf{Z} \times [0,\infty)$ by

$$\underline{\alpha} \cdot \Sigma(x,t)\underline{\alpha} = E_{eq(p,\rho)}[u_{\underline{\alpha}}(0,0)u_{\underline{\alpha}}(x,t)]$$

$$\text{where} \quad u_{\underline{\alpha}}(x,t) = \alpha(\xi_x(t) - p) + \beta(\eta_x(t) - \rho).$$

Since $P_{\mathrm{eq}(p,\rho)}$ is invariant under the translation, $\Sigma(x,t)$ is the covariance matrix of $(\xi_x(s), \eta_x(s))$ and its space-time translation $(\xi_{x+y}(s+t), \eta_{x+y}(s+t))$. Hence if we define

$$R(x,t) := \Sigma(x,t)\chi^{-1}(p,\rho),$$

then $R(x-y, t-s)$ is the space-time correlation coefficient of $(\xi_x(t), \eta_x(t))$. The next theorem states that $R(x,t)$ behaves like $R(x,t) \approx K_D(t,x)$ as $x, t \to \infty$, as being expected ([5]).

Theorem 2. *For* $\underline{J} = (J^1, J^2)^T \in C_0^\infty(\mathbf{R}) \times C_0^\infty(\mathbf{R})$

$$\lim_{N\to\infty} \sum_{x\in\mathbf{Z}} \mathbf{R}(x, N^2 t)\underline{J}(x/N) = \int_{-\infty}^{\infty} K_D(t,\theta)\underline{J}(\theta)d\theta.$$

Theorem 2 is deduced from (4). Indeed by (4),

$$\lim_{N\to\infty} \frac{1}{N} \sum_x \sum_y \underline{J}_1(y/N) \cdot R(x-y, N^2 t)\underline{J}_2(x/N) = \int_{-\infty}^{\infty} \underline{J}_1(\theta) \cdot U_t \underline{J}_2(\theta)d\theta \tag{5}$$

because the formula under the limit on the left side equals $E[(Y_{0,N}, \underline{J}_1)(Y_{t,N}, \underline{J}_2)]$. If the delta function could be taken for $\underline{J}_1$, the relation of Theorem 2 would come out. For justification we take Fourier transform in (5). To this end let $\hat{R}$ be the Fourier series with coefficients R:

$$\begin{aligned} \hat{R}(\lambda,t) &= \hat{\Sigma}(\lambda,t)\chi^{-1}, \qquad \lambda \in \mathbf{R} \\ \hat{\Sigma}(\lambda,t) &= \sum_{x\in\mathbf{Z}} e^{i\lambda x}\Sigma(x,t). \end{aligned}$$

Lemma 3.

$$0 \le \hat{\Sigma}(\lambda,t) \le \hat{\Sigma}(\lambda,0) = \chi.$$

Proof. If $a_x = e^{i\lambda x}\Sigma(x,t)$, then

$$\sum_{x=-k}^{k-1}\sum_{y=-k}^{k-1} a_{y-x} = \sum_{u=-2k}^{2k}(2k-|u|)a_u.$$

The right-hand side divided by $2k$ converges, as $k \to \infty$, to $\hat{\Sigma}(\lambda,t)$. Since $S(t)$ is a symmetric operator, the first diagonal component of a_{y-x} may be expressed in the form

$$a^{11}_{y-x} = E^{\nu_{p,\rho}}\left[e^{i\lambda y}S(t/2)\{\xi_y - p\}e^{-i\lambda x}S(t/2)\{\xi_x - p\}\right],$$

and similarly for the other components; hence

$$\underline{\alpha}\cdot\hat{\Sigma}(\lambda,t)\underline{\alpha} = \lim_{k\to\infty}\frac{1}{2k}E^{\nu_{p,\rho}}\left|S(t/2)\left\{\sum_{x=-k}^{k-1} e^{i\lambda x}[\alpha(\xi_x - p) + \beta(\eta_x - \rho)]\right\}\right|^2 .$$

The inequalities of the lemma now follow from the fact that $S(t)$ is contraction in $L^2(\nu_{p,\rho})$. Q.E.D.

Proof of Theorem 2. Rewriting the relation (5) by means of $\hat{R}$, we have

$$\begin{aligned}&\lim_{N\to\infty}\int_{-N\pi}^{N\pi}\hat{\underline{J}}_1^N(\lambda)\cdot\hat{R}(\lambda/N, N^2t)\hat{\underline{J}}_2^N(-\lambda)d\lambda\\ &= \int_{-\infty}^{\infty}\hat{\underline{J}}_1(\lambda)\cdot e^{-t\lambda^2D^T}\hat{\underline{J}}_2(-\lambda)d\lambda.\end{aligned}\tag{6}$$

Here

$$\hat{\underline{J}}^N(\lambda) = \frac{1}{N}\sum\underline{J}(x/N)e^{i\lambda x/N},\quad \hat{\underline{J}}(\lambda) = \int_{-\infty}^{\infty}\underline{J}(\theta)e^{i\lambda\theta}d\theta.$$

By the Poisson summation formula, $\hat{\underline{J}}^N(\lambda) = \sum_{x\in\mathbf{Z}}\hat{\underline{J}}(\lambda + 2\pi N x)$. The class of J_1^i $(i = 1,2)$ in (6) may be extended to the set of rapidly decreasing functions. Let $\delta > 0$, $g_\delta(\theta) = (4\pi\delta)^{-1/2}e^{-\theta^2/(4\delta)}$ and $\underline{J}_1(\theta) = g_\delta(\theta)\underline{\alpha}$. Then, $\hat{g}_\delta(\lambda) = e^{-\delta\lambda^2}$ and

$$e^{-\delta\lambda^2} \le \hat{g}_\delta^N(\lambda) \le e^{-\delta\lambda^2} + \frac{2e^{-\delta(\pi N)^2}}{1-e^{-\delta(\pi N)^2}}\ (|\lambda| \le N\pi);$$

and writing $\underline{J}$ for $\underline{J}_2$ in (6), we infer that

$$\begin{aligned}&\lim_{N\to\infty}\int_{-N\pi}^{N\pi}e^{-\delta\lambda^2}\underline{\alpha}\cdot\hat{R}(\lambda/N, N^2t)\hat{\underline{J}}^N(-\lambda)d\lambda\\ &= \int_{-\infty}^{\infty}e^{-\delta\lambda^2}\underline{\alpha}\cdot e^{-t\lambda^2D^T}\hat{\underline{J}}(-\lambda)d\lambda.\end{aligned}$$

On taking the limit as $\delta \downarrow 0$ this relation is also valid for $\delta = 0$. The proof is complete. Q.E.D.

§3. Central Limit Theorem Variance

The canonical measure for the configurations on Λ_n with the number of particles m and the total energy E is the conditional law

$$P_{n,m,E}[\,\cdot\,] = \frac{\nu_{p,\rho}(\,\cdot\, \cap \{|\xi|_{\Lambda(n)} = m,\ |\eta|_{\Lambda(n)} = E\} \,|\, \mathcal{F}_{\mathbf{Z}\setminus\Lambda(n)}\,)}{\nu_{p,\rho}(|\xi|_{\Lambda(n)} = m,\ |\eta|_{\Lambda(n)} = E\,)}.$$

Here for $\Lambda \subset \mathbf{Z}$, $|\xi|_\Lambda = \sum_{x\in\Lambda} \xi_x$ and $|\eta|_\Lambda = \sum_{x\in\Lambda} \eta_x$; $\mathcal{F}_\Lambda$ stands for the σ-field in $\mathcal{X}$ generated by $\eta_y, y \in \Lambda$. From the reversibility relation it follows that for any functions f and g of η and any bond $b \in \Lambda_n^*$,

$$E_{n,m,E}[c_b(\eta) f(S^b\eta) g(\eta)] = E_{n,m,E}[c_{b'}(\eta) f(\eta) g(S^{b'}\eta)],$$

where b' is the bond obtained from b by reversing its direction. The Dirichlet form for $L_{\Lambda(n)}$ accordingly is given by

$$\begin{aligned} \mathcal{D}_{n,m,E}\{f\} &:= -E_{n,m,E}[f L_{\Lambda(n)} f] \\ &= \sum_{b\in\Lambda^*(n)} \mathcal{D}^b_{n,m,E}\{f\} \end{aligned}$$

where $\mathcal{D}^b_{n,m,E}\{f\} = \frac{1}{2} E_{n,m,E}[(\Gamma_b f)^2 c_b],$; the corresponding bilinear form is given by

$$\mathcal{D}^{01}_{n,m,E}(f,g) = -\frac{1}{2} E_{n,m,E}[f\cdot(L_{01}+L_{10})g] = \frac{1}{2} E_{n,m,E}[(\Gamma_{01}f)(\Gamma_{01}g)c_{01}].$$

We introduce a function space on which the central limit theorem variance is well defined. The numbers p and ρ are fixed so that $0 < p < 1$ and $\rho \geq p$ unless otherwise specified. They will be dropped from the notations if used as sub- or superscripts.

Definition 4. *Let $\mathcal{G}$ denote the linear space of all functions $h \in \mathcal{F}_c$ of the form*

$$L_I H := \sum_{b\in I^*} L_b H = h, \tag{7}$$

where I is an interval of $\mathbf{Z}$ and H is a local function such that for some positive integer K,

$$\sum_{b\in I^*} (\Gamma_b H(\eta))^2 \leq K \sum_{x\in I} (\eta_x)^K, \quad \eta \in \mathcal{X}. \tag{8}$$

(This bound, which may be replaced by a weaker one, is adopted only for convenience sake. We may take I as the minimal of intervals Λ such that $h \in \mathcal{F}_\Lambda$.)

If $h \in \mathcal{F}_c$ satisfies

$$E^\nu[h \mid \mathcal{F}_{\mathbf{Z}\setminus I} \vee \sigma\{|\xi|_I, |\eta|_I\}] = 0 \text{ a.s.},$$

then it admits a representation (7) but the condition (8) may fail to hold. The functions w_{01}^P, w_{01}^E are in $\mathcal{G}$: the requirements are satisfied with $I = \{0,1\}$ and $H = -\xi_0$ and $H = -\eta_0$, respectively. For each positive integer K put

$$\mathcal{F}_c^K = \{f \in \mathcal{F}_c : |f(\eta)| \le K \sum_{|x|\le K} (\eta_x)^K\}$$

Then the linear space $L\mathcal{F}_c^K$ is obviously included in $\mathcal{G}$.

Let $L_{n,m,E}$ denote the restriction of $L_{\Lambda(n)}$ to the space of functions on $\mathcal{X}_{n,m,E} := \{\eta \in \mathcal{X}_{\Lambda(n)} : |\xi|_{\Lambda(n)} = m, |\eta|_{\Lambda(n)} = E\}$, and for h, $g \in \mathcal{G}$, define

$$V_{n,m,E}(h,g) = \frac{1}{2n} E_{n,m,E} \left[\sum_{|x|<n'} \tau_x h \cdot (-L_{n,m,E})^{-1} \sum_{|x|<n'} \tau_x g \right],$$

where n' is the maximal integer among those for which both sums in the brackets are $\mathcal{F}_{\Lambda(n)}$-measurable.

Theorem 5. *For every $h, g \in \mathcal{G}$ and for every $p > 0, \rho \ge p$, there exists a following limit*

$$\lim_{m/2n\to p, E/2n\to\rho} V_{n,m,E}(h,g),$$

where the limit is taken in such a way that n, m and E are sent to infinity so that $m/2n \to p$ and $E/2n \to \rho$. The functional defined by this limit makes a bilinear form on $\mathcal{G}$. If it is denoted by

$$V(h,g) = V^{p,\rho}(h,g),$$

then the subspace

$$\mathcal{G}_\circ := \{\alpha w_{01}^P + \beta w_{01}^E - Lf : \alpha, \beta \in \mathbf{R}, f \in \mathcal{F}_c^K \text{ for some } K\}$$

is dense in $\mathcal{G}$ with respect to the quadratic form $V^{p,\rho}\{h\} := V^{p,\rho}(h,h)$.

Theorem 5 says that every $h \in \mathcal{G}$ can be approximated by an element of $\mathcal{G}_\circ$ in the metric $\sqrt{V^{p,\rho}}$ as accurately as one needs. To apply this to the gradients $\nabla^- \xi := \xi_0 - \xi_1$ and $\nabla^- \eta := \eta_0 - \eta_1$, we need the following lemma (cf. [6]).

Lemma 6. *Suppose that (1) and (2) are satisfied. Then both* $\nabla^- \xi$ *and* $\nabla^- \eta$ *are in* $\mathcal{G}$. *Let* H^P *and* H^E *stand for the corresponding* H*'s (with* $I(h) = \{0, 1\}$ *). Then*

$$\Gamma_{01} H^P = \xi_0 / c_{\text{ex}}(\eta_0) \quad \textit{and} \quad \Gamma_{01} H^E = \eta_0 / c_{\text{ex}}(\eta_0) \quad \textit{if} \quad \xi_0(1 - \xi_1) = 1$$

and $\Gamma_{01} H^P = 0$ *if* $\xi_0(1 - \xi_1) = 0$*; moreover there exists a constant* $\delta > 0$ *such that* $\delta \le \Gamma_{01} H^E \le 1/\delta$ *whenever* $\mathbf{1}(\eta_0 \ge 2)\xi_1 = 1$.

The proof of Theorem 5 may be carried out along the same lines as in [7] or [8].

§4. The Least Upper Bound of Spectrum

In this section we are concerned with the Markov process whose infinitesimal generator is $\mathcal{L}$, a self-adjoint operator on $L^2(\nu_{p,\rho})$ (see Theorem 1). Let $\mathcal{P}(\mathcal{X})$ be the set of all probability measures on $\mathcal{X}$. Define a functional $\mathcal{I}(\mu)$ of $\mu \in \mathcal{P}(\mathcal{X})$ by

$$\mathcal{I}(\mu) = E^\nu[\varphi(-\mathcal{L})\varphi], \text{ where } \varphi = \sqrt{d\mu/d\nu}$$

if μ is absolutely continuous relative to $\nu = \nu_{p,\rho}$ and φ is in the domain of $\sqrt{-\mathcal{L}}$; and $\mathcal{I}(\mu) = \infty$ otherwise. For a local function G on $\mathcal{X}$ let $\Omega_\circ\{G + \mathcal{L}\}$ denote the least upper bound of the spectrum of the operator $G + \mathcal{L}$. It has the variational representation

$$\Omega_\circ\{G + \mathcal{L}\} = \sup_{\mu \in \mathcal{P}(\mathcal{X})} \Big(E^\mu[G] - \mathcal{I}(\mu) \Big).$$

Given a positive integer n and $h \in \mathcal{G}$, let n' be the maximal integer such that $\tau_y h \in \mathcal{F}_{\Lambda(n)}$ if $|y| < n'$, and define a function $G_n = G_n^h$ by

$$G_n = \frac{1}{2n} \sum_{y : |y| < n'} \tau_y h.$$

Theorem 7. *Let* $h \in \mathcal{G}$. *Let the interval* $I = I(h)$ *and the function* H *be chosen so that*

$$\sum_{b \in I^*} (\Gamma_b H)^2 c_b \le A \sum_{x \in I} \eta_x^K \tag{9}$$

where $\eta_x^K = (\eta_x)^K$, *and* A *and* K *are positive constants with* $K \geq 1$. *Let* $G_n = G_n^h$ *be defined as above. Also define a function* $\zeta_n^l(\eta)$ *for* $l \geq 1$ *by*

$$\zeta_n^l(\eta) = \frac{1}{2n} \sum_{x:|x| \leq n} \eta_x^K \mathbf{1}(\eta_x > l).$$

Then, if $\lambda \in (-1,1)$, $J \in C_0^2(\mathbf{R})$, *and* C *is a positive constant such that* $A|I|^2(1-2^{-K})^{-1} \leq C$, *it holds that for all* $n, l \in \mathbf{N}$,

$$\begin{aligned} &\overline{\lim_{N\to\infty}}\ \Omega_\circ \left\{ \sum_{x \in \mathbf{Z}} \left[N^\lambda J(x/N) \tau_x G_n - \frac{C}{N} J^2(x/N) \tau_x \zeta_n^l \right] + N^{1+2\lambda} \mathcal{L} \right\} \\ &\leq ||J||_{L^2}^2 \sup_{m,E:E/m \leq 2l} V_{n,m,E}\{h\}. \end{aligned}$$

where $||J||_{L^2}^2 = \int_{\mathbf{R}} J^2 d\theta$ *and the supremum is taken over all couples of positive integers* m *and* E *such that* $m \leq E \leq 2lm$.

Proof. The proof is divided into three steps.

Step 1. This step is quite similar to a corresponding argument in [7], so we provide only an outline. The supremum of the spectrum $\Omega_\circ$ that is to be estimated may be given by the variational formula

$$\Omega^N = \sup_{\mu \in \mathcal{P}(\mathcal{X})} E^\mu \left[\sum_{x \in \mathbf{Z}} \left[N^\lambda j_x \tau_x G_n - \frac{C}{N} j_x^2 \tau_x \zeta_n^l \right] - N^{1+2\lambda} \mathcal{I}(\mu) \right].$$

where we put $j_x = J(x/N)$.

Let $\varphi = \sqrt{d\mu/d\nu}$ and $\mathcal{D}^\Lambda = \sum_{b \in \Lambda^*} \mathcal{D}^b$, then $\mathcal{I}(\mu) = \sum_{b \in \mathbf{Z}^*} \mathcal{D}^b\{\varphi\} = \frac{1}{2n} \sum_{x \in \mathbf{Z}} \mathcal{D}^{\Lambda(n)}\{\tau_x \varphi\}$. We substitute this into the variational expression given above. To compute the expectation appearing in it we first take the conditional expectation conditioned on $\omega = \eta|_{\Lambda_n^c}$. If $\mu(\,\cdot\,|\omega)$ stands for this conditional law, then $E^\mu[G_n]$ is expressed as an integral of $F(\omega) = E^{\mu(\,\cdot\,|\omega)}[G_n]$ by μ. We have a similar expression for the form $\mathcal{D}^{\Lambda(n)}\{\varphi\}$, which may be naturally restricted to the space $L^2(\nu^{\Lambda(n)}, \mathcal{X}_{\Lambda(n)})$ (ν^Λ is the product measure on $\mathcal{X}_\Lambda$ with the same common one-site marginal as that of $\nu = \nu_{p,\rho}$). Rewriting μ for $\mu(\,\cdot\,|\omega) \in \mathcal{P}(\mathcal{X}_{\Lambda(n)})$ and taking the supremum in μ, we see that Ω^N is not greater than

$$\frac{N^{1+2\lambda}}{2n} \sum_{x \in \mathbf{Z}} \sup_{\mu \in \mathcal{P}(\mathcal{X}_{\Lambda(n)})} \left\{ \frac{2n}{N^{1+2\lambda}} E^\mu \left[N^\lambda j_x G_n - \frac{C}{N} j_x^2 \zeta_n^l \right] - \mathcal{D}^{\Lambda(n)}\{\varphi\} \right\}.$$

Decomposing $\mathcal{X}_{\Lambda(n)}$ into the ergodic classes $\mathcal{X}_{n,m,E}$ we may express $\mathcal{D}^{\Lambda(n)}\{\varphi\}$ in the form $\mathcal{D}^{\Lambda(n)}\{\varphi\} = \sum_m \sum_E p_{m,E} \mathcal{D}_{n,m,E}\{\varphi_{m,E}\}$, where

$p_{m,E} = \mu(\mathcal{X}_{n,m,E})$ and $\varphi_{m,E}$ is the square root of a probability density on $\mathcal{X}_{n,m,E}$. As a consequence we see that if

$$\Omega^N_{n,m,E,x} = \sup_{\mu\in\mathcal{P}(\mathcal{X}_{n,m,E})} \left\{ \frac{2nj_x}{N^{1+\lambda}} E^\mu[G_n] - \frac{2nCj_x^2}{N^{2+2\lambda}} E^\mu[\zeta^l_n] - \mathcal{D}_{n,m,E}\{\varphi\} \right\},$$

then

$$\Omega^N \le \frac{N^{1+2\lambda}}{2n} \sum_{x=1}^{N} \sup_{m,E} \Omega^N_{n,m,E,x}. \tag{10}$$

Step 2. Let $\langle\,\cdot\,\rangle_{n,m,E}$ stand for the expectation by $P_{n,m,E}$. For H introduced in Definition 4 and for any $\mathcal{F}_{\Lambda(n)}$-measurable function u, we have the following identity

$$\langle u\tau_x h\rangle_{n,m,E} = -\frac{1}{2}\sum_{b\in I^*(h)} \Big\langle \Gamma_{b+x}u\cdot\tau_x(c_b\Gamma_b H)\Big\rangle_{n,m,E} \tag{11}$$

or in terms of the Dirichlet form

$$\langle u\tau_x h\rangle_{n,m,E} = -\sum_{b\in I^*(h)} \mathcal{D}^{b+x}_{n,m,E}(u,\tau_x H). \tag{12}$$

(Here $b+x$ is the oriented bond obtained by translating b by x.) From this it follows that

$$E^\mu[G_n] = -\frac{1}{2n}\sum_{|x|<n'}\sum_{b\in I^*(h)} \mathcal{D}^{b+x}_{n,m,E}(\tau_x H,\varphi^2).$$

A simple computation verifies that the terms $|\mathcal{D}^b_{n,m,E}(F,\varphi^2)|$, where $F\in C(\mathcal{X}_{n,m,E})$, are bounded by

$$\sqrt{\tfrac{1}{2}\Big\langle\Big[(\Gamma_b F)^2 c_b + (\Gamma_{b'}F)^2 c_{b'}\Big]\varphi^2\Big\rangle_{n,m,E}}\sqrt{\mathcal{D}^b_{n,m,E}\{\varphi\}}.$$

where b' is the bond b but reversely oriented. By employing Schwarz inequality and the assumption (9) on H it therefore follows that $|E^\mu[G_n]|$ is at most

$$\begin{aligned}&\frac{1}{2n}\sqrt{\sum_{|x|<n'}\sum_{b\in I^*(h)}\Big\langle(\Gamma_{b+x}\tau_x H)^2 c_{b+x}\varphi^2\Big\rangle_{n,m,E}}\sqrt{|I^*|\mathcal{D}_{n,m,E}\{\varphi\}}\\ &\le \frac{|I|}{n}\sqrt{A\sum_{|x|\le n}\Big\langle\eta^K_x\varphi^2\Big\rangle_{n,m,E}}\sqrt{\mathcal{D}_{n,m,E}\{\varphi\}}.\end{aligned}$$

By the inequality $2ab - a^2 \le b^2$ this shows that

$$(13) \qquad \frac{2nj_x}{N^{1+\lambda}} E^{\mu}[G_n] - \mathcal{D}_{n,m,E}\{\varphi\} \le \frac{A|I|^2 j_x^2}{N^{2+2\lambda}} \sum_{|x|\le n} \left\langle \eta_x^K \varphi^2 \right\rangle_{n,m,E}.$$

Since $(m^{-1}\sum \eta_x)^K \le m^{-1}\sum \eta_x^K$, the condition $E = \sum \eta_x > 2lm$ implies the inequality $2^{-K}\sum \eta_x^K \ge l^K m$, which in turn implies that

$$2n\zeta_n^l = \sum \eta_x^K \mathbf{1}(\eta_x > l) \ge \sum \eta_x^K - l^K m \ge (1-2^{-K}) \sum \eta_x^K.$$

This combined with (13) shows that if the constant C is chosen so that $A|I|^2 \le (1-2^{-K})C$, then

$$\Omega_{n,m,E,x}^N \le 0 \ \text{ whenever } E/m > 2l,$$

and accordingly that the supremum over the pairs of m and E in (10) may be restricted to those satisfying $E/m \le 2l$. Consequently

$$(14) \qquad \Omega^N \le \frac{N^{1+2\lambda}}{2n} \sum_{x\in\mathbf{Z}} \sup_{m,E:E/m\le 2l} \Omega_{n,m,E,x}^N.$$

Step 3. Now we apply the following estimate for the spectrum of the Schrödinger type operator $L_{n,m,E} + F$ with $F \in C(\mathcal{X}_{n,m,E})$ satisfying $\langle F\rangle_{n,m,E} = 0$:

$$(15) \qquad \Omega_{\circ}\{F + L_{n,m,E}\} \le \langle F(-L_{n,m,E})^{-1}F\rangle_{n,m,E} + \frac{4}{\kappa_n^2}\|F\|_\infty^3,$$

where $\kappa_n = \kappa_{n,m,E}$ is the second eigenvalue of $-L_{n,m,E}$ (cf. [7],[1] etc.). Taking $F = (2nj_x/N^{1+\lambda})G_{n,m,E}$ in (15), where $G_{n,m,E} = G_n|\mathcal{X}_{n,m,E}$,

$$\begin{aligned} \Omega_{n,m,E,x}^N &\le \Omega_{\circ}\{(2nj_x/N^{1+\lambda})G_{n,m,E} + L_{n,m,E}\} \\ &\le (2n)V_{n,m,E}\left\{\frac{j_x}{N^{1+\lambda}}h\right\} + \frac{4}{\kappa_n^2}\cdot\left[\frac{2nj_x\|G_{n,m,E}\|_\infty}{N^{1+\lambda}}\right]^3 \\ &= \frac{2nj_x^2}{N^{2+2\lambda}}V_{n,m,E}\{h\} + O\left(\frac{1}{N^{3+3\lambda}}\right). \end{aligned}$$

From (14) we thus obtain $\overline{\lim}_{N\to\infty} \Omega^N \le \|J\|_{L^2}^2 \sup_{m,E:E/m\le 2l} V_{n,m,E}\{h\}$, the required bound. Q.E.D.

The next theorem is essentially a corollary of Theorem 7.

Theorem 8. *Let $h \in \mathcal{G}$ and put*

$$F^N(\eta) = \sqrt{N} \sum_{x \in \mathbf{Z}} J(x/N) \tau_x h(\eta).$$

Then there exists a constant C such that for all positive constants β and l,

$$\overline{\lim_{N\to\infty}} E_{\mathrm{eq}} \left| \int_0^T F^N(\eta(N^2t))dt \right| \leq \beta T ||J||_{L^2}^2 \sup_{p_o, \rho_o : \rho_o/p_o \leq l} V^{p_o, \rho_o}\{h\} + (\log 2)/\beta + (C\beta)/l.$$

Proof. We may replace F^N by

$$F_n^N := \sqrt{N} \sum_{x \in \mathbf{Z}} J(x/N) \frac{1}{2n} \sum_{y : |y-x|<n'} \tau_y h.$$

In fact if

$$a_{N,n}^x = \frac{N}{2n^2} \sum_{y:\, |y-x|<n'} [J(x/N) - J(y/N)],$$

then $|a_{N,n}^x| \leq \int_{-n/N}^{n/N} |J''(s + N^{-1}x)| ds$ and the difference

$$F^N - F_n^N = \frac{n}{\sqrt{N}} \sum_{x \in \mathbf{Z}}^{N} a_{N,n}^x \tau_x h$$

is obviously negligible under the equilibrium measure.

Introducing the random variable $X^N = \int_0^T F_n^N(\eta(N^2t))dt$, we may write $E_{\mathrm{eq}}|X^N|$ for what to estimate. Let $K \geq 1$ be a constant for which the condition (9) is satisfied. Let ζ_n^l be a function defined in Theorem 7 and put

$$Y^N = \int_0^T \frac{C}{N} \sum_{x \in \mathbf{Z}} J^2(x/N) \tau_x \zeta_n^l(\eta(N^2t))dt.$$

Then by Jensen's inequality and the Feynman-Kac formula

$$\begin{aligned} & E_{\mathrm{eq}}[|X^N| - \beta Y^N] \\ & \leq \frac{1}{\beta} \log \max_{+,-} E_{\mathrm{eq}}[e^{\pm \beta X^N - \beta^2 Y^N}] + \frac{\log 2}{\beta} \\ & \leq \frac{T}{\beta} \max_{+,-} \Omega_o \Big\{ \pm \beta F^N - \frac{C}{N} \sum_{x \in \mathbf{Z}} |\beta J(x/N)|^2 \tau_x \zeta_n^l + N^2 L \Big\} + \frac{\log 2}{\beta}. \end{aligned}$$

According to Theorems 7 and 5, if C is chosen suitably large, then

$$\varlimsup_{N\to\infty} E_{\text{eq}}[\,|X^N| - \beta Y^N\,] \le \beta T ||J||_{L^2}^2 \sup_{p_o,\rho_o:\rho_o/p_o\le l} V^{p_o,\rho_o}\{h\} + \frac{\log 2}{\beta}.$$

This gives the required inequality since $E_{\text{eq}}[\beta Y^N] \le C_1\beta/l$. Q.E.D.

§5. Upper and Lower Bounds For $D(p,\rho)$

Let $\underline{\kappa} = \underline{\kappa}(p,\rho)$ and $\bar{\kappa} = \bar{\kappa}(p,\rho)$ stand for the eigen-values of $D(p,\rho)$ such that $\underline{\kappa} \le \bar{\kappa}$. We here prove that for some positive constants m and M,

$$\frac{m}{p+(1+\lambda)^{-1}} \le \underline{\kappa} \le \bar{\kappa} \le M(1+\lambda) \qquad (\rho \ge p > 0),$$

where $\lambda = \lambda(p,\rho)$ is the parameter appearing in the definition of $\nu_{p,\rho}$.

Proof of the upper bound. We shall apply the fact that if $\hat{c}_\circ$ is a symmetric 2×2 matrix and $\hat{c}_\circ \ge \hat{c}$, then $\mathrm{Tr}(\hat{c}_\circ\chi^{-1}) \ge \mathrm{Tr}(\hat{c}\chi^{-1})$. Let $\langle\cdot\rangle$ indicate the expectation under $\nu_{p,\rho}$. Then

$$\begin{aligned}\underline{\alpha}\cdot\hat{c}(p,\rho)\underline{\alpha} &\le \left\langle\Big(\Gamma_{01}\{\alpha\xi_0+\beta\eta_0\}\Big)^2 c_{01}\right\rangle \\ &= \Big\langle\{\alpha\xi_0+\beta\eta_0\}^2(1-\xi_1)c_{\text{ex}}(\eta_0)\Big\rangle + \beta^2\langle\xi_0\xi_1 c_{\text{zr}}(\eta_0)\rangle\end{aligned}$$

In view of the conditions (2) and (3), $c_{\text{ex}}(\eta_0) \le C[c_{\text{zr}}(\eta_0)+\mathbf{1}(\eta_0=1)]$. By combining this with the relations $\langle c_{\text{zr}}(\eta_0)\rangle = p\lambda$, $\langle\eta_0 c_{\text{zr}}(\eta_0)\rangle = (\rho+p)\lambda$ and $\langle\eta_0^2 c_{\text{zr}}(\eta_0)\rangle = (\langle\eta_0^2\rangle + 2\rho + p)\lambda$, the last line above is dominated by $\beta^2p^2\lambda$ plus a constant multiple of

$$(1-p)[\alpha^2p\lambda + 2\alpha\beta(\rho+p)\lambda + \beta^2(\langle\eta_0^2\rangle+2\rho+p)\lambda + (\alpha+\beta)^2\langle\mathbf{1}(\eta_0=1)\rangle].$$

Recalling what is remarked at the beginning of this proof, noticing $\det\chi = (p\langle\eta_0^2\rangle - \rho^2)(1-p)$ so that

$$\chi^{-1}(p,\rho) = \frac{1}{(p\langle\eta_0^2\rangle-\rho^2)(1-p)}\begin{pmatrix}\langle\eta_0^2\rangle-\rho^2 & -(1-p)\rho \\ -(1-p)\rho & (1-p)p\end{pmatrix}$$

and carrying out simple computations, we see that

$$\mathrm{Tr}(\hat{c}\chi^{-1}) \le C_1[\lambda + p^2(\lambda^2)(p\langle\eta_0^2\rangle-\rho^2)^{-1}+\lambda].$$

Since $\bar{\kappa}+\underline{\kappa} = \mathrm{Tr}(\hat{c}\chi^{-1})$, these yield the required upper bound, if we can find a positive constant δ so that

$$(16) \qquad p\langle\eta_0^2\rangle - \rho^2 \ge \delta p^2\lambda.$$

(This is certainly true for $\lambda \le 1$.) To this end set $\ell = \ell(\lambda) = \max\{k : c_{\rm zr}(k) \le \lambda\}$ and $p_k = \nu_{p,\rho}\{\eta : \eta_0 = k\}/p$. Noticing that $p_{k+1}/p_k = \lambda/c_{\rm zr}(k+1)$, we infer from $|c_{\rm zr}(k) - c_{\rm zr}(\ell)| \le a_1|k-\ell|$ that for all sufficiently large λ,

$$p_k \ge p_\ell \exp\{-a_1(k-\ell)^2/\lambda\} \quad \text{if} \quad |k-\ell| \le 2\sqrt{\lambda},$$

or, what we are about to apply, $\min\{\sum_{k<\ell-\sqrt{\lambda}} p_k, \sum_{k>\ell+\sqrt{\lambda}} p_k\} \ge \delta$ with some constant $\delta > 0$ independent of λ. Hence

$$\begin{aligned} \langle \eta_0^2 \rangle/p - (\rho/p)^2 &= E^{\nu_{p,\rho}}[|\eta_0 - \rho/p|^2 \,|\, \eta_0 > 0] \\ &\ge \lambda P^{\nu_{p,\rho}}[|\eta_0 - \rho/p| \ge \sqrt{\lambda} \,|\, \eta_0 > 0] \ge \delta\lambda. \end{aligned}$$

Thus we have shown (16).

Proof of the lower bound. Let $A = A(p,\rho)$ be a 2×2 symmetric matrix whose quadratic form is

$$\underline{\alpha} \cdot A\underline{\alpha} = V\{\alpha\nabla^-\xi + \beta\nabla^-\eta\}.$$

Then $D(p,\rho) = \chi(p,\rho)A^{-1}(p,\rho)$ and it holds that $V\{\alpha\nabla^-\xi + \beta\nabla^-\eta\} \le \langle(\Gamma_{01}\{\alpha H^P + \beta H^E\})^2 c_{01}\rangle$ (cf. [6]), where H^P and H^E are functions introduced in Lemma 6. We shall apply the inequality

$$\underline{\kappa} \ge \frac{\det(\chi A^{-1})}{\operatorname{Tr}(\chi A^{-1})} = \frac{1}{\operatorname{Tr}(\chi^{-1}A)}. \tag{17}$$

By employing Lemma 6 as well as the conditions (1) through (3) we see that for some constant C,

$$\begin{aligned} \underline{\alpha} \cdot A\underline{\alpha} &\le \langle(\Gamma_{01}\{\alpha H^P + \beta H^E\})^2 c_{01}\rangle \\ &\le C\left\langle \frac{\xi_0(1-\xi_1)}{c_{\rm zr}(\eta_0+1)}(\alpha\xi_0 + \beta\eta_0)^2 \right\rangle + C\beta^2\langle \xi_1 c_{\rm zr}(\eta_0)\rangle. \end{aligned}$$

One observes that the right-hand side equals C times

$$\begin{aligned} &\alpha^2(1-p)\frac{p}{\lambda}\left(1 - \frac{1}{Z_\lambda}\right) + 2\alpha\beta(1-p)\frac{\rho-p}{\lambda} \\ &+\beta^2\left(\frac{1-p}{\lambda}\langle(\eta_0-\xi_0)^2\rangle + p^2\lambda\right). \end{aligned}$$

Noticing that $Z_\lambda = 1 + \lambda/c_{\rm zr}(2) + O(\lambda^2)$ as $\lambda \downarrow 0$ and $\nu_{p,\rho}\{\eta_0 = 2\} = p\lambda/c_{\rm zr}(2)Z_\lambda$, and applying the inequality used in the preceding proof, we infer that

$$\det(\chi)\operatorname{Tr}(\chi^{-1}A) \le C'p^2(1-p)\lambda \quad \text{for} \quad 0 < \lambda < 1. \tag{18}$$

For large values of λ we make an elementary computation (as we did for the upper bound) to see that $\det(\chi)\mathrm{Tr}(\chi^{-1}A)$ is at most C times

$$\frac{1-p}{\lambda}(2-p)(p\langle\eta_0^2\rangle-\rho^2)+\frac{(1-p)^2p^2}{\lambda}-\frac{(1-p)p}{\lambda Z_\lambda}(\langle\eta_0^2\rangle-\rho^2)+(1-p)p^3\lambda.$$

Hence, in view of (16),

$$\mathrm{Tr}(\chi^{-1}A)\le C'\left[\frac{1}{\lambda}+p\right]\ (\lambda\ge 1).$$

This together with (17) and (18) concludes the asserted lower bound of $\underline{\kappa}$.

References

[1] C. Kipnis and C. Landim, Scaling limits of particle systems, Springer, 1999.

[2] C. Landim, S. Sethuraman and S. Varadhan, Spectral Gap for Zero-Range Dynamics, Ann. Probab. 24 (1996), pp. 1871-1902.

[3] Y. Nagahata, Fluctuation dissipation equation for lattice gas with energy, to appear in Jour. Stat. Phys., Vol. 110 Nos.1/2 (2003) 219-246.

[4] Y. Nagahata and K. Uchiyama: Spectral gap for zerorange-exclusion dynamics, preprint

[5] H. Spohn, Large scale dynamics of interacting particles, Text and Monographs in Physics, Springer, 1991.

[6] K. Uchiyama, Equilibrium fluctuations for zero-range-exclusion processes, preprint

[7] S.R.S. Varadhan, Nonlinear diffusion limit for a system with nearest neighbor interactions - II, Asymptotic problems in probability theory: stochastic models and diffusions on fractals (eds. K.D. Elworthy and N. Ikeda), Longman (1993), pp. 75-128.

[8] S.R.S. Varadhan and H.T. Yau., Diffusive limit of lattice gases with mixing condition, Asian J.Math vol. 1 (1997), 623-678.

Department of Mathematics
Tokyo Institute of Technology
Oh-okayama, Meguro
Tokyo 152-8551
Japan
E-mail address: `uchiyama@math.titech.ac.jp`