9516

…EAN-VLASOV LIMIT FOR INTERACTING …NDOM PROCESSES IN RANDOM MEDIA

by

Paolo Dai Pra and Frank den Hollander

Report 9516
April 1995
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NIJMEGEN
Toernooiveld
6525 ED Nijmegen
The Netherlands

McKean-Vlasov limit for interacting random processes in random media

Paolo Dai Pra
Dipartimento di Matematica Pura e Applicata
Universitá di Padova
Via Belzoni 7, 35131 Padova, Italy

Frank den Hollander
Mathematical Institute
University of Nijmegen
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Abstract

We apply large deviation theory to particle systems with a random mean-field interaction in the McKean-Vlasov limit. In particular, we describe large deviations and normal fluctuations around the McKean-Vlasov equation. The randomness in the interaction gives rise to new phenomena, which are illustrated for the Kuramoto model (random oscillators) and the Curie-Weiss model (random magnets).

Keywords and phrases: Interacting particle systems; random media; McKean-Vlasov equation; large deviations; central limit theorem.

AMS 1991 *subject classifications*. Primary 60F10. Secondary 82C44.

0 Introduction

In this paper, we consider interacting diffusions and interacting spin-flip systems with a mean-field Hamiltonian that depends on a random medium. In the thermodynamic limit, the dynamics of a typical particle is described by a collection of *coupled* McKean-Vlasov equations indexed by a medium parameter. For finite but large systems there are fluctuations around the McKean-Vlasov limit, which are controlled by the random dynamics and by the random medium.

Our approach to the problem is to do a large deviation analysis for the *double layer empirical measure*

$$L_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x^i_{[0,T]},\omega^i)}. \tag{0.1}$$

Here, N is the size of the system,

$$\begin{array}{lcl} x^i_{[0,T]} & = & \text{the path of the } i\text{-th particle in the time interval } [0,T], \\ \omega^i & = & \text{the } i\text{-th component of the medium.} \end{array} \tag{0.2}$$

Our main results are the following (see Sections 1-3):

1. We derive a large deviation principle for L_N as $N \to \infty$, with an explicit representation for the corresponding rate function I.

2. The McKean-Vlasov limit is the associated law of large numbers, i.e., the McKean-Vlasov equation follows from 1. by identifying the unique zero of I.

3. By a standard contraction argument we derive a large deviation principle for the *double layer empirical flow*

$$\ell_N = \Big(\frac{1}{N}\sum_{i=1}^{N} \delta_{(x_t^i,\omega^i)}\Big)_{t\in[0,T]} \tag{0.3}$$

as $N \to \infty$ and compute the corresponding rate function i.

4. The second order fluctuations around the McKean-Vlasov limit are identified in the form of a central limit theorem, deduced from 1. via a variational computation.

The goal of our paper is two-fold:

I. For homogeneous systems, results as in 1.-4. have been obtained by Dawson (1982), Dawson and Gärtner (1987), Ben Arous and Brunaud (1990). (See also Comets and Eisele (1988) for models with a so-called "local" mean-field interaction.) We show how to generalize the analysis in these papers to systems with a random medium interaction. The random medium leads to some new ingredients in the analysis. It is also responsible for some new effects (see Section 4).

II. We want to give an expository presentation of the large deviation approach to this problem area.

The outline of the paper is as follows. In Section 1 we consider interacting diffusions and state our theorems for this class of models (Theorems 1-4). Section 2 and Appendices A-B are devoted to the proof of the results. In Section 3 we consider spin-flip systems and show how the results have to be modified (Theorems 5-8). Finally, Section 4 describes two applications:

(i) The Kuramoto model of random oscillators (i.e., diffusions on the unit circle).

(ii) The Curie-Weiss model of random magnets (i.e., spin flips between + and −).

Example (i) was studied by Bonilla, Neu and Spigler (1992), and the McKean-Vlasov limit was obtained through heuristic arguments. This model shows the phenomenon of "phase locking" above a critical value for the strength of the interaction (depending on the law of the random medium). Example (ii) was studied by Salinas and Wrezinski (1985), and the equilibrium properties were described in detail. This model shows the phenomenon of "spontaneous magnetization" below a critical value for the temperature (depending on the law of the random medium). In both examples the type of randomness critically determines the phase diagram.

1 Diffusions

1.1 The model

Let $H_N : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be the N-particle random Hamiltonian given by

$$H_N(\underline{x},\underline{\omega}) = \frac{1}{2N} \sum_{i,j=1}^{N} f(x^j - x^i;\omega^i,\omega^j) + \sum_{i=1}^{N} g(x^i;\omega^i), \tag{1.1}$$

where $\underline{x} = (x^i)_{i=1}^N$ is the *state* variable and $\underline{\omega} = (\omega^i)_{i=1}^N$ is the *medium* variable. The ω^i are assumed to be i.i.d. random variables with common law μ. For a fixed realization of $\underline{\omega}$, think of $\underline{x} \to H_N(\underline{x};\underline{\omega})$ as a Hamiltonian in the components x^i with an inhomogeneous mean field interaction parametrized by the components ω^i. The functions f and g play the role of a pair potential resp. external field, and are assumed to satisfy:

- f, f', f'', g, g', g'' exist, are bounded and are jointly continuous in all variables ($'$ denotes derivative w.r.t. the x-variable). [1]

For given $\underline{\omega}$, let $\underline{x}_t = (x_t^i)_{i=1}^N$ be the system of N interacting diffusions evolving according to the Itô stochastic differential equations

$$dx_t^i = -\frac{\partial H_N}{\partial x^i}(\underline{x}_t,\underline{\omega})dt + d\xi_t^i \quad (i = 1,\ldots,N;\ t \in [0,T]), \tag{1.2}$$

where $(\xi_t^i)_{i=1}^N$ are i.i.d. standard Brownian motions on $\mathbb{R}$. For every $\underline{\omega}$, (1.2) has a reversible equilibrium measure proportional to $\exp[-H_N(\underline{x},\underline{\omega})]$. The initial condition $\underline{x}_0$ is assumed to have product distribution $\lambda^{\otimes N}$, with λ having a finite second moment. The time $T > 0$ is fixed but arbitrary. Because f', g' are globally Lipschitz, (1.2) has a unique (strong) solution with continuous trajectories (see Karatzas and Shreeve (1988), Theorem 2.9).

We shall write $P_N^{\underline{\omega}}$ to denote the law of $\underline{x}_{[0,T]} = (\underline{x}_t)_{t\in[0,T]}$ given $\underline{\omega}$, and $W^{\otimes N}$ to denote the law of the solution of (1.2) when $H_N \equiv 0$ (i.e., W is the law of a standard Brownian motion starting with initial distribution λ).

The system in (1.2) will be our object of study. We shall identify its large deviation and central limit behavior in the limit as $N \to \infty$. Our main results are formulated in Theorems 1-4 in Sections 1.2-5 below.

1.2 Empirical measure and large deviations

Define the *double layer empirical measure*

$$L_N(\underline{x}_{[0,T]},\underline{\omega}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x^i_{[0,T]},\omega^i)}. \tag{1.3}$$

[1]The assumptions on f, g are stronger than what is actually needed for proving the results in this paper. However, they allow us to illustrate the use of large deviations without excessive technicalities. A few more restrictions will be imposed later, for the same reason.

For the medium variables $\mathbb{R}$ could be replaced by any Polish space without change in the proofs. For the state variables $\mathbb{R}$ could be replaced by $\mathbb{R}^d$ ($d \geq 1$) with only minor modifications in the proof of Theorem 3 in Section 2.3.

This is a random variable taking values in $\mathcal{M}_1(C[0,T]\times\mathbb{R})$, the set of probability measures on $C[0,T]\times\mathbb{R}$ (where $C[0,T]$ is the path space, i.e., the continuous functions on $[0,T]$). In (1.3), the symbol δ_y denotes the point measure at y, so $L_N(A)=\frac{1}{N}\sum_{i=1}^N 1\{(x^i_{[0,T]},\omega^i)\in A\}$ $(A\subset C[0,T]\times\mathbb{R})$.

Lemma 1 below gives a representation for $P_N^{\underline{\omega}}$ in terms of L_N.

Lemma 1 *For given* $\underline{\omega}$

$$\frac{dP_N^{\underline{\omega}}}{dW^{\otimes N}}(\underline{x}_{[0,T]})=\exp[NF(L_N(\underline{x}_{[0,T]},\underline{\omega}))] \tag{1.4}$$

where for $Q\in\mathcal{M}_1(C[0,T]\times\mathbb{R})$

$$\begin{aligned} F(Q) \;=\; \int Q(dx_{[0,T]},d\omega)\;\Big\{ & -\tfrac{1}{2}\int_0^T dt\Big[\Big(\int Q(dy_{[0,T]},d\pi)\hat{f}'(y_t-x_t;\omega,\pi)+g'(x_t;\omega)\Big)^2 \\ & \qquad -\int Q(dy_{[0,T]},d\pi)\hat{f}''(y_t-x_t;\omega,\pi)+g''(x_t;\omega)\Big] \\ & -\tfrac{1}{2}\int Q(dy_{[0,T]},d\pi)[f(y_T-x_T;\omega,\pi)-f(y_0-x_0;\omega,\pi)] \\ & -[g(x_T;\omega)-g(x_0;\omega)]\Big\} \end{aligned} \tag{1.5}$$

with $\hat{f}$ *given by*

$$\hat{f}(x;\omega,\pi)=\frac{1}{2}[f(x;\omega,\pi)+f(-x;\pi,\omega)]. \tag{1.6}$$

The proof of Lemma 1 will be given in Section 2.1. Note that $Q\to F(Q)$ is nonlinear and contains repeated integrals over the measure Q. A simpler representation for $F(Q)$ will be given in Lemma 2 below.

The representation in (1.4) is the key to the following large deviation principle (LDP), from which we shall deduce various features of the asymptotic behavior of L_N as $N\to\infty$. Define

$$P_N(\cdot)=\int \mu^{\otimes N}(d\underline{\omega})P_N^{\underline{\omega}}(L_N\in\cdot), \tag{1.7}$$

which is the law of L_N under the joint distribution of process and medium. Note that $P_N\in\mathcal{M}_1(\mathcal{M}_1(C[0,T]\times\mathbb{R}))$.

Theorem 1 $(P_N)_{N\geq 1}$ *satisfies the LDP with rate function*

$$I(Q)=H(Q|W\otimes\mu)-F(Q) \tag{1.8}$$

where H *denotes the relative entropy*

$$H(Q|W\otimes\mu)=\int dQ\log\frac{dQ}{d(W\otimes\mu)}. \tag{1.9}$$

The proof of Theorem 1 will be given in Section 2.1. Roughly, the statement in Theorem 1 means that

$$\frac{1}{N}\log P_N(A)\approx-\inf_{Q\in A}I(Q) \tag{1.10}$$

for large N and for A sufficiently regular. For a precise formulation of the LDP we refer to Deuschel and Stroock (1989), pp. 35-36.

One sees from (1.5) that $F \equiv 0$ when $H_N \equiv 0$ (i.e., $f, g \equiv 0$). Thus $H(Q|W \otimes \mu)$ is the rate function for the system without interaction.

1.3 McKean-Vlasov equation

Before we analyze $I(Q)$, we first give an alternative representation for $F(Q)$ in (1.5) that will turn out to be more convenient. For given $\omega \in \mathbb{R}$ and $q \in \mathcal{M}_1(\mathbb{R} \times \mathbb{R})$ define

$$\beta^{\omega,q}(x) = -\int q(dy, d\pi)\hat{f}'(y - x; \omega, \pi) - g'(x; \omega) \quad (t \in [0,T], x \in \mathbb{R}). \tag{1.11}$$

Let $P^{\omega,Q}$ be the law of the unique (strong) solution of the 1-dimensional Itô equation

$$dx_t = \beta^{\omega,\Pi_t Q}(x_t)dt + d\xi_t, \tag{1.12}$$

where ξ_t is a standard Brownian motion on $\mathbb{R}$ and x_0 has law λ. Here $\Pi_t Q$ is the projection of Q at time t, i.e.,

$$(\Pi_t Q)(E \times F) = Q\Big(\{(x_{[0,T]}, \omega) : x_t \in E, \omega \in F\}\Big) \quad (E, F \subset \mathbb{R}). \tag{1.13}$$

For fixed ω the drift in (1.12) has a mean-field form, i.e., the interaction in (1.2) of a single-component diffusion with the other components and with the medium appears in (1.12) as an average w.r.t. the given measure $\Pi_t Q$.

Lemma 2 *For all Q*

$$F(Q) = \int Q(dx_{[0,T]}, d\omega) \log \frac{dP^{\omega,Q}}{dW}(x_{[0,T]}). \tag{1.14}$$

The proof of Lemma 2 will be given in Section 2.2. By combining (1.8), (1.9) and (1.14) we get the following simpler representation for the rate function:

Corollary 1 *For all Q*

$$I(Q) = H(Q|P^Q), \tag{1.15}$$

where $P^Q \in \mathcal{M}_1(C[0,T] \times \mathbb{R})$ is defined by

$$P^Q(dx_{[0,T]}, d\omega) = \mu(d\omega)P^{\omega,Q}(dx_{[0,T]}). \tag{1.16}$$

Since $I(Q) \geq 0$ for all Q, one sees from (1.10) that as $N \to \infty$ the measure P_N tends to concentrate around the zeroes of I, i.e., the solutions of

$$Q = P^Q. \tag{1.17}$$

The next theorem states that (1.17) has a unique solution. Define $\nu^Q \in \mathcal{M}_1(\mathbb{R})$ to be the projection of Q on the medium coordinate, i.e.,

$$\nu^Q(F) = Q\Big(\{(x_{[0,T]}, \omega) : \omega \in F\}\Big) \quad (F \in \mathbb{R}). \tag{1.18}$$

Let $Q^\omega \in \mathcal{M}_1(C[0,T])$ be the regular conditional probability measure obtained from Q after conditioning on ω, i.e.,

$$Q(dx_{[0,T]}, d\omega) = \nu^Q(d\omega) Q^\omega(dx_{[0,T]}). \tag{1.19}$$

The results that follow will be proved under the following assumption on the initial measure λ for the single-component diffusions: [2]

(A1) λ has a density ϕ w.r.t. Lebesgue measure satisfying $\phi \in L^1(dx) \cap L^p(dx)$ for some $p > 1$.

Theorem 2 *Assume (A1). Then (1.17) has a unique solution Q_* which has the following properties:*

1. $\nu^{Q_*} = \mu$.
2. *Q_*^ω is the law of a Markov diffusion process for μ-a.s. all ω.*
3. *Let $q_t^\omega = \Pi_t Q_*^\omega$. Then q_t^ω is the weak solution of the McKean-Vlasov equation* [3]

$$\begin{cases} \frac{\partial}{\partial t} q_t^\omega = \mathcal{L}^\omega q_t^\omega \quad (t \in (0,T], \omega \in \mathbb{R}) \\ q_0^\omega = \lambda \end{cases} \tag{1.20}$$

where $\mathcal{L}^\omega$ is the nonlinear operator

$$\mathcal{L}^\omega q_t^\omega = -\frac{\partial}{\partial x}[\beta^{\omega, q_t} q_t^\omega] + \frac{1}{2}\frac{\partial^2}{\partial x^2} q_t^\omega \quad (\omega \in \mathbb{R}) \tag{1.21}$$

and q_t is defined by $q_t(dx, d\omega) = \mu(d\omega) q_t^\omega(dx)$.

4. *The diffusion process in 2. has generator L_t^ω given by*

$$L_t^\omega = \beta^{\omega, q_t}\frac{\partial}{\partial x} + \frac{1}{2}\frac{\partial^2}{\partial x^2} \quad (\omega \in \mathbb{R}). \tag{1.22}$$

The proof of Theorem 2 will be given in Section 2.2. Note that the equations in (1.20) for different values of ω are *coupled*, because

$$\beta^{\omega, q_t}(x) = -\int \mu(d\pi) \int q_t^\pi(dy) \hat{f}'(y - x; \omega, \pi) - g'(x; \omega) \tag{1.23}$$

depends on the whole family $\{q_t^\pi\}_{\pi \in \mathbb{R}}$ (see (1.11)).

As a corollary to Theorems 1 and 2 we obtain the following law of large numbers:

Corollary 2 *Assume (A1). Then*

$$P_N \Rightarrow \delta_{Q_*} \text{ weakly as } N \to \infty. \tag{1.24}$$

[2] Assumption (A1) could in principle be weakened by using the technique of Lyapunov functions, as in Sznitman (1984). However, we stick to (A1) because it allows us to give a rather elementary proof of uniqueness of the solution of (1.17).

[3] Eqs.(1.20-1.21) mean that $\frac{d}{dt}\int q_t^\omega(dx)\phi(x) = \int q_t^\omega(dx)\beta^{\omega,q_t}(x)\phi'(x) + \frac{1}{2}\int q_t^\omega(dx)\phi''(x)$ for every $\phi \in \mathcal{D}$, the space of infinitely differentiable functions with compact support. By standard arguments this implies that q_t^ω for $t > 0$ has a density that is a classical solution of (1.20).

1.4 Empirical flow and large deviations

With each $Q \in \mathcal{M}_1(C[0,T] \times \mathbb{R})$ is associated the flow of marginals $q_{[0,T]} = (\Pi_t Q)_{t\in[0,T]}$. Define the *double layer empirical flow*

$$\ell_N = \Big(\frac{1}{N}\sum_{i=1}^{N} \delta_{(x_t^i,\omega^i)}\Big)_{t\in[0,T]}. \tag{1.25}$$

This is a random variable taking values in $\mathcal{M}_1(\mathbb{R} \times \mathbb{R})^{[0,T]}$. (The topology on this power set is the one induced by the weak topology on $\mathcal{M}_1(C[0,T] \times \mathbb{R})$ via the map $Q \to q_{[0,T]}$.) Note that both $q_{[0,T]}$ and ℓ_N take values in the subset of $\mathcal{M}_1(\mathbb{R} \times \mathbb{R})^{[0,T]}$ consisting of those flows whose projection on the medium coordinate is independent of t. We shall denote this subset by $\mathcal{M}$. The empirical flow ℓ_N contains less information than the empirical measure L_N (recall (1.3)). Therefore its large deviation behavior can be obtained from Theorem 1 via the *contraction principle* (Varadhan (1984), Theorem 2.4).

To formulate the LDP for $(\ell_N)_{N\geq 1}$ we introduce the following notation. For $q_{[0,T]} \in \mathcal{M}$, let $q^{\omega}_{[0,T]}$ be the conditional flow given ω, i.e.,

$$q_t(dx, d\omega) = \nu^q(d\omega) q_t^{\omega}(dx) \quad (t \in [0,T]), \tag{1.26}$$

where ν^q is the projection of q_t on the medium coordinate (which is independent of t). Let $\mathcal{D}$ be the space of infinitely differentiable functions with compact support, and let $\mathcal{D}^*$ be its dual (the elements of which are distributions). For $\psi^* \in \mathcal{D}^*$ and $p \in \mathcal{M}_1(\mathbb{R})$ define the norm

$$\|\psi^*\|_p^2 = \frac{1}{2} \sup_{\phi\in\mathcal{D}:\langle p,\phi'^2\rangle>0} \frac{\langle \psi^*, \phi\rangle^2}{\langle p, \phi'^2\rangle}, \tag{1.27}$$

where $\langle\cdot\rangle$ denotes the usual inner product. Let $\Delta \subset \mathcal{M}$ be the set of all flows satisfying

$$\begin{aligned} &\nu^q \ll \mu \\ &t \to q_t^{\omega} \text{ is weakly differentiable for } \nu^q\text{-a.s. all } \omega. \end{aligned} \tag{1.28}$$

Finally, let

$$\wp_N(\cdot) = \int \mu^{\otimes N}(d\underline{\omega}) P_N^{\underline{\omega}}(\ell_N \in \cdot), \tag{1.29}$$

which is the law of ℓ_N under the joint distribution of process and medium. Note that $\wp_N \in \mathcal{M}_1(\mathcal{M})$.

Theorem 3 *$(\wp_N)_{N\geq 1}$ satisfies the LDP with rate function*

$$i(q_{[0,T]}) = \begin{cases} \int_0^T dt\Big\{\int \nu^q(\omega)\|\frac{\partial}{\partial t}q_t^{\omega} - \mathcal{L}^{\omega} q_t^{\omega}\|^2_{q_t^{\omega}}\Big\} + H(\nu^q|\mu) & \text{if } q_{[0,T]} \in \Delta \\ \infty & \text{otherwise.} \end{cases} \tag{1.30}$$

The proof of Theorem 3 will be given in Section 2.3. Note that $i(q_{[0,T]}) = 0$ iff $\nu^q = \mu$ and q_t^{ω} is the solution of the McKean-Vlasov equation for μ-a.s. all ω (recall (1.20), (1.21) and (1.23)).

1.5 Central limit theorem

It is possible to deduce from Theorem 1 a central limit theorem (CLT) for the empirical measure L_N in (1.3). The general technique is formulated in Bolthausen (1986). Essentially, what we must do is show that the rate function $Q \to I(Q)$ in (1.8) and (1.15) has a strictly positive and finite curvature at its unique zero Q_*. However, in order to apply Bolthausen's theorem we need a technical assumption, namely: [4]

(A2) There are functions $\alpha_i, \beta_i : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ and numbers $c_i \in \mathbb{R}^+$ such that

$$f(y-x;\omega,\pi) = \sum_{i=0}^{\infty} c_i \alpha_i(x,\omega)\beta_i(y,\pi) \tag{1.31}$$

with
(1) $\sum_i c_i < \infty$
(2) α_i, β_i twice continuously differentiable w.r.t. the variable x resp. y
(3) $\alpha_i, \alpha_i', \alpha_i'', \beta_i, \beta_i', \beta_i''$ bounded uniformly in i.

Our central limit theorem reads:

Theorem 4 *Assume (A2). Let C_b be the set of bounded continuous functions from $C[0,T] \times \mathbb{R}$ to $\mathbb{R}$. As $N \to \infty$ the field*

$$\left(N^{\frac{1}{2}}\left[\int \phi dL_N - \int \phi dQ_*\right]\right)_{\phi \in C_b} \tag{1.32}$$

converges under P_N to a Gaussian field with covariance

$$C(\phi,\psi) = \int Q_*(dx_{[0,T]}, d\omega)\phi[Q_*](x_{[0,T]},\omega)\psi[Q_*](x_{[0,T]},\omega), \tag{1.33}$$

where

$$\begin{aligned} \phi[Q_*](x_{[0,T]},\omega) &= \phi(x_{[0,T]},\omega) - \phi^* \\ &- \int_0^T \left(\int Q_*(dy_{[0,T]}, d\pi)[\phi(y_{[0,T]},\pi) - \phi^*]\hat{f}'(y_t - x_t;\omega,\pi)\right) dw_t^{\omega} \end{aligned} \tag{1.34}$$

with $\phi^* = \int \phi dQ_*$ (similarly for ψ), $w_t^\omega = x_t - \int_0^t \beta^{\omega,\Pi,Q_*} ds$ (which is a Brownian motion under Q_*^ω) and $\hat{f}$ given by (1.6).

The statement in Theorem 4 means the following: for $\phi_1, \phi_2, \ldots, \phi_n \in C_b$ the vector

$$\left(N^{\frac{1}{2}}\left[\int \phi_i dL_N - \int \phi_i dQ_*\right]\right)_{i=1}^{n} \tag{1.35}$$

converges in law to an n-dimensional Gaussian random variable with mean zero and covariance matrix $(C(\phi_i,\phi_j))_{i,j=1}^n$.

The proof of Theorem 4 will be given in Section 2.4. From the proof it will be seen that the covariance matrix is strictly positive definite.

[4] By applying the techniques in Sznitman (1984), the CLT could in principle be proved without assumption (A2). However: (i) Bolthausen's method nicely connects large deviations and CLT; (ii) The proof is easily modified to cover other models, e.g. spin-flip systems (see Section 3); (iii) Assumption (A2) is satisfied in many interesting examples (e.g. the Kuramoto model (see Section 4); see also Ben Arous and Brunaud (1990) for more examples).

2 Proof of Lemmas 1-2 and Theorems 1-3

2.1 Proof of Lemma 1 and Theorem 1

Proof of Lemma 1.
The proof is based on two basic tools in stochastic calculus, namely Girsanov's formula and Itô's rule (see e.g. Karatzas and Shreve (1987), Theorems 3.3.3 and 3.5.1). Girsanov's formula yields (recall (1.2))

$$\frac{dP_N^{\underline{\omega}}}{dW^{\otimes N}}(\underline{x}_{[0,T]}) = \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\int_0^T \left(\frac{\partial H_N}{\partial x^i}(\underline{x}_t,\underline{\omega})\right)^2 dt - \sum_{i=1}^{N}\int_0^T \left(\frac{\partial H_N}{\partial x^i}(\underline{x}_t,\underline{\omega})\right) d\underline{x}_t^i\right]. \quad (2.1)$$

Under the measure $W^{\otimes N}$, the process $\underline{x}_{[0,T]}$ is N-dimensional Brownian motion (see Section 1.1). Thus, by Itô's rule,

$$\sum_{i=1}^{N}\int_0^T \left(\frac{\partial H_N}{\partial x^i}(\underline{x}_t,\underline{\omega})\right) d\underline{x}_t^i = H_N(\underline{x}_T,\underline{\omega}) - H_N(\underline{x}_0,\underline{\omega}) - \frac{1}{2}\sum_{i=1}^{N}\int_0^T \left(\frac{\partial^2 H_N}{\partial (x^i)^2}(\underline{x}_t,\underline{\omega})\right) dt. \quad (2.2)$$

Hence

$$\begin{aligned}\frac{dP_N^{\underline{\omega}}}{dW^{\otimes N}}(\underline{x}_{[0,T]}) &= \exp\Big[-\tfrac{1}{2}\textstyle\sum_{i=1}^{N}\int_0^T \Big\{\Big(\frac{\partial H_N}{\partial x^i}(\underline{x}_t,\underline{\omega})\Big)^2 - \frac{\partial^2 H_N}{\partial (x^i)^2}(\underline{x}_t,\underline{\omega})\Big\} dt \\ &\qquad - \Big(H_N(\underline{x}_T,\underline{\omega}) - H_N(\underline{x}_0,\underline{\omega})\Big)\Big].\end{aligned} \quad (2.3)$$

The rest of the proof simply consists of inserting the definition of H_N (see (1.1)) and rewriting the resulting expression in terms of the empirical measure L_N (see (1.5)). This leads to the expression given in (1.4)-(1.6). ∎

Proof of Theorem 1.
Let R_N be the law of L_N under the measure $W^{\otimes N} \otimes \mu^{\otimes N}$. Under R_N, the pairs $(x^i_{[0,T]}, \omega^i)$ are i.i.d. random variables. It therefore follows from Sanov's Theorem (Deuschel and Stroock (1989) Theorem 3.2.17) that $(R_N)_{N\geq 1}$ satisfies the LDP with rate function $H(Q|W \otimes \mu)$ given in (1.9). Now, using Lemma 1, we have (recall (1.4) and (1.7))

$$\begin{aligned} P_N(\cdot) &= \textstyle\int \mu^{\otimes N}(d\underline{\omega}) P_N^{\underline{\omega}}(L_N(d\underline{x}_{[0,T]},\underline{\omega}) \in \cdot) \\ &= \textstyle\int \mu^{\otimes N}(d\underline{\omega}) \int W^{\otimes N}(d\underline{x}_{[0,T]}) \frac{dP_N^{\underline{\omega}}}{dW^{\otimes N}}(\underline{x}_{[0,T]}) 1\{L_N(d\underline{x}_{[0,T]},\underline{\omega}) \in \cdot\} \\ &= \textstyle\int d\Big(W^{\otimes N} \otimes \mu^{\otimes N}\Big) \exp[NF(L_N)] 1\{L_N \in \cdot\} \\ &= \textstyle\int R_N(dQ) \exp[NF(Q)] 1\{Q \in \cdot\}. \end{aligned} \quad (2.4)$$

Identity (2.4) means that

$$\frac{dP_N}{dR_N}(Q) = \exp[NF(Q)]. \quad (2.5)$$

Our assumptions on f, g in Section 1.1 imply that F is a bounded continuous function w.r.t. the weak topology in $\mathcal{M}_1(C[0,T] \times \mathbb{R})$ (see (1.5)). Therefore, (2.5) allows us to apply Varadhan's Lemma (Varadhan (1984), Theorem 2.2) and conclude that the LDP for $(R_N)_{N\geq 1}$ with rate function $H(Q|W \otimes \mu)$ implies the LDP for $(P_N)_{N\geq 1}$ with rate function $H(Q|W \otimes \mu) - F(Q)$, as was claimed in (1.8) and (1.9). ∎

2.2 Proof of Lemma 2 and Theorem 2

Proof of Lemma 2.
We begin by applying Girsanov's formula to the 1-dimensional Itô-equation in (1.12), namely

$$\log \frac{dP^{\omega,Q}}{dW}(x_{[0,T]}) = -\frac{1}{2}\int_0^T (\beta^{\omega,\Pi_t Q}(x_t))^2 dt + \int_0^T \beta^{\omega,\Pi_t Q}(x_t)dx_t. \tag{2.6}$$

We want to show that the r.h.s. of (2.6), when integrated over $Q(d\underline{x}_{[0,T]}, d\omega)$, yields $F(Q)$ given in (1.5). Recalling (1.11), we see that the first term in the r.h.s. of (2.6) gives rise to the first term in the r.h.s. of (1.5). To check the remaining terms, let us look a bit closer at the stochastic integral in (2.6).

By (1.11) we have

$$\begin{aligned} &\int Q(dx_{[0,T]}, d\omega) \int_0^T \beta^{\omega,\Pi_t Q}(x_t)dx_t \\ &= -\int Q(dx_{[0,T]}, d\omega) \int_0^T \Big[\int Q(dy_{[0,T]}, d\pi)\hat{f}'(y_t - x_t;\omega,\pi) + g'(x_t;\omega)\Big]dx_t. \end{aligned} \tag{2.7}$$

(Note that if $Q \ll W \otimes \mu$ then $x_{[0,T]}$ is a Q-semimartingale, so the stochastic integral in (2.7) makes sense.) Consider the first term in the r.h.s. of (2.7). Since $\hat{f}'$ is an odd function of its first argument, this term equals

$$-\frac{1}{2}\int Q(dx_{[0,T]}, d\omega)\int Q(dy_{[0,T]}, d\pi)\int_0^T \hat{f}'(y_t - x_t;\omega,\pi)[dx_t - dy_t]. \tag{2.8}$$

We can apply Itô's rule to the 2-dimensional semimartingale $(x,y)_{[0,T]}$ and write

$$d\hat{f}(y_t - x_t;\omega,\pi) = \hat{f}''(y_t - x_t;\omega,\pi)dt - \hat{f}'(y_t - x_t;\omega,\pi)dx_t + \hat{f}'(y_t - x_t;\omega,\pi)dy_t. \tag{2.9}$$

By substituting (2.9) into (2.8) we get the expression

$$\begin{aligned} &-\tfrac{1}{2}\int Q(dx_{[0,T]}, d\omega)\int Q(dy_{[0,T]}, d\pi) \\ &\times\Big[\int_0^T \hat{f}''(y_t - x_t;\omega,\pi)dt - \hat{f}(y_T - x_T;\omega,\pi) + \hat{f}(y_0 - x_0;\omega,\pi)\Big]. \end{aligned} \tag{2.10}$$

Next consider the second term in the r.h.s. of (2.7). Itô's rule yields that this term equals

$$-\int Q(dx_{[0,T]}, d\omega)\Big[-\frac{1}{2}\int_0^T g''(x_t;\omega)dt + g(x_T;\omega) - g(x_0;\omega)\Big]. \tag{2.11}$$

From (2.10) and (2.11) the claim in Lemma 2 easily follows after observing that (1.6) gives,

$$\begin{aligned} &\int Q(dx_{[0,T]}, d\omega)\int Q(dy_{[0,T]}, d\pi)\hat{f}(y_t - x_t;\omega,\pi) \\ &= \int Q(dx_{[0,T]}, d\omega)\int Q(dy_{[0,T]}, d\pi)f(y_t - x_t;\omega,\pi) \end{aligned} \tag{2.12}$$

for every t and, in particular, for $t = 0$ and $t = T$. ∎

Proof of Theorem 2.
Observe that $\nu^Q = \nu^{P^Q} = \mu$ (recall (1.16-1.18)) and that $P^{\omega,Q}$ is the law of the solution of (1.12), i.e., the Markov diffusion with generator given in (1.21). It is therefore easy to see that properties 1.-4. in Theorem 2 are satisfied by any solution of (1.17) (note that (1.20) is the Fokker-Planck equation associated with the diffusion Q_*). Now, the *existence* of a solution of (1.17) comes from the general fact that the rate function of an LDP must have *at least one zero* (Deuschel and Stroock (1989), Exercise 2.1.14(i)). The *uniqueness* of the solution will be proved in Appendix A. ∎

2.3 Proof of Theorem 3

Let Π denote the map $\Pi : Q \to q_{[0,T]}$ (remember that $q_t = \Pi_t Q$). The topology on $\mathcal{M}$ has been chosen in such a way that Π is continuous. Since $\ell_N = \Pi L_N$, it follows from the contraction principle (Varadhan (1984), Theorem 2.4) that $(\wp_N)_{N\geq 1}$ satisfies the LDP with rate function

$$j(q_{[0,T]}) = \inf_{\Pi Q = q_{[0,T]}} I(Q). \tag{2.13}$$

We want to show that $j(q_{[0,T]}) = i(q_{[0,T]})$ for every $q_{[0,T]} \in \mathcal{M}$, where i is the rate function given in (1.30). In order to do so, we shall first show that equality holds when $j(q_{[0,T]}) < \infty$ (Steps 1-3 below). After that we shall show that if $i(q_{[0,T]}) < \infty$ then $j(q_{[0,T]}) < \infty$ (Step 4 below), which will complete the proof. The basic ideas are taken from Föllmer (1986) (see also Brunaud (1993)).

Step 1. By a standard argument involving lower semicontinuity and compactness of the level sets of the rate function I, we have that if $j(q_{[0,T]}) < \infty$ then there exists a Q such that $\Pi Q = q_{[0,T]}$ and $I(Q) = j(q_{[0,T]})$. From (1.8) we have

$$I(Q) = \int \nu^q(d\omega) H(Q^\omega|W) + H(\nu^q|\mu) - F(Q). \tag{2.14}$$

Moreover, since $F(Q)$ depends on Q only through $q_{[0,T]}$ (see (1.5) and (1.14)) we have that Q^ω minimizes $H(Q^\omega|W)$ under the constraint $\Pi Q^\omega = q^\omega_{[0,T]}$ for ν^q-a.s. all ω. As shown in Föllmer (1986), Theorem 1.31, the latter fact implies that Q^ω is the law of a Markov diffusion

$$dx_t = b^\omega_t(x_t)dt + dw_t \tag{2.15}$$

for a suitable drift $b^\omega_t(x)$, and that

$$H(Q^\omega|W) = \int Q^\omega(dx_{[0,T]}) \int_0^T dt\, [b^\omega_t(x_t)]^2. \tag{2.16}$$

Substituting (2.16) into (2.14), and using Lemma 2 in combination with (2.6) and (2.15), we obtain

$$\begin{aligned} I(Q) &= \tfrac{1}{2} \int \nu^q(d\omega) \int Q^\omega(dx_{[0,T]}) \int_0^T dt\, [b^\omega_t(x_t) - \beta^{\omega,\Pi_t Q}(x_t)]^2 + H(\nu^q|\mu) \\ &= \tfrac{1}{2} \int_0^T dt \Big\{ \int \nu^q(d\omega) \Big[\int q^\omega_t(dx) \Big(b^\omega_t(x) - \beta^{\omega,\Pi_t Q}(x)\Big)^2 \Big] \Big\} + H(\nu^q|\mu). \end{aligned} \tag{2.17}$$

This equation reduces to the required expression in (1.30) if we can show that for every $t \in (0,T]$ and for ν^q-a.s. all ω

$$\frac{1}{2} \int q^\omega_t(dx) \Big(b^\omega_t(x) - \beta^{\omega,\Pi_t Q}(x)\Big)^2 = \|\frac{\partial}{\partial t} q^\omega_t - \mathcal{L}^\omega q^\omega_t\|^2_{q^\omega_t}. \tag{2.18}$$

Step 2. To prove (2.18) we proceed as follows. According to (2.15), q^ω_t is the weak solution of the Fokker-Planck equation

$$\frac{\partial q^\omega_t}{\partial t} = -\frac{\partial}{\partial x}[b^\omega_t q^\omega_t] + \frac{1}{2}\frac{\partial^2}{\partial x^2} q^\omega_t. \tag{2.19}$$

Together with (1.21) this implies

$$\frac{\partial}{\partial t}q_t^\omega - \mathcal{L}^\omega q_t^\omega = -\frac{\partial}{\partial x}[(b_t^\omega - \beta^{\omega,\Pi_t Q})q_t^\omega]. \tag{2.20}$$

Hence, recalling the definition of $\|\cdot\|$ in (1.27), we get

$$\begin{aligned} \|\tfrac{\partial}{\partial t}q_t^\omega - \mathcal{L}^\omega q_t^\omega\|_{q_t^\omega}^2 &= \tfrac{1}{2}\sup_{\phi\in\mathcal{D}:\langle q_t^\omega,\phi'^2\rangle>0} \frac{\langle (b_t^\omega-\beta^{\omega,\Pi_t Q})q_t^\omega,\phi'\rangle^2}{\langle q_t^\omega,\phi'^2\rangle} \\ &\le \tfrac{1}{2}\langle q_t^\omega,(b_t^\omega-\beta^{\omega,\Pi_t Q})^2\rangle, \end{aligned} \tag{2.21}$$

where we have used the Cauchy-Schwarz inequality (recall that $\langle\cdot,\cdot\rangle$ denotes the usual inner product). Thus, to get (2.18) we must show that in (2.21) equality is attained.

Step 3. It suffices to show that the set $\{\phi' : \phi \in \mathcal{D}\}$ is dense in $L^2(q_t^\omega)$ for all t and ν^q-a.s. all ω. We first note that q_t^ω is absolutely continuous w.r.t. Lebesgue measure for all t and ν^q-a.s. all ω (this follows from the fact that $Q \ll W \otimes \mu$, $\nu^q \ll \mu$ and the marginals of W are absolutely continuous w.r.t. Lebesgue measure). So, it is enough to prove that if ρ is an absolutely continuous probability measure on $\mathbb{R}$, i.e., $\rho(dx) = p(x)dx$, then $\{\phi' : \phi \in \mathcal{D}\}$ is dense in $L^2(\rho)$.

The proof is by contradiction. Suppose $\{\phi' : \phi \in \mathcal{D}\}$ is not dense in $L^2(\rho)$. Then there exists $h \in L^2(\rho)$ such that

$$\int \phi'(x)h(x)p(x)dx = 0 \quad \text{for every } \phi \in \mathcal{D}. \tag{2.22}$$

Since $hp \in L^1(dx)$, it follows from Brezis (1983), Lemma 8.1, that there exists $C \in \mathbb{R}$ such that $hp \equiv C$ a.s. w.r.t. Lebesgue measure. If $C = 0$ then clearly $h \equiv 0$ ρ-a.s. On the other hand, if $C \neq 0$ then $hp \notin L^1(dx)$.

Step 4. To complete the proof of Theorem 3 we need to show that if $i(q_{[0,T]}) < \infty$ then $j(q_{[0,T]}) < \infty$. We use Föllmer (1986), Theorem 1.31, where it is observed that there exists a countable number of bounded continuous functions $(\phi_i)_{i\ge1}$ from $\mathbb{R}\times\mathbb{R}$ to $\mathbb{R}$ and a countable (dense) subset $(t_i)_{i\ge1}$ of $[0,T]$ such that $\Pi Q = q_{[0,T]}$ if and only if

$$\int \Pi_{t_i} Q(dx_{[0,T]}, d\omega)\phi_i(x,\omega) = 0 \quad (i = 0,1,2,\ldots). \tag{2.23}$$

Now, by compactness and lower semicontinuity of H, for every $n \ge 0$ there exists a Q_n such that $H(Q_n|W\otimes\mu) < \infty$ and Q_n minimizes $H(Q|W\otimes\mu)$ under the constraint that (2.23) holds for $i = 1,2,\ldots,n$. Since we have proved that $i(q_{[0,T]}) = j(q_{[0,T]})$ when $j(q_{[0,T]}) < \infty$, it follows from (2.13) that

$$I(Q_n) = \inf\Big\{i(p_{[0,T]}) : \int p_{t_i}^\omega(dx,d\omega)\phi_i(x,\omega) = 0 \ \text{ for } \ i = 1,\ldots,n\Big\}. \tag{2.24}$$

In particular, $I(Q_n) \le i(q_{[0,T]})$. By compactness of the level sets of I, the sequence $(Q_n)_{n\ge1}$ has a limit point Q which, by lower semicontinuity of I, satisfies $I(Q) \le i(q_{[0,T]})$. Moreover, (2.23) holds for Q. Hence, via (2.13) we get $j(q_{[0,T]}) \le I(Q) \le i(q_{[0,T]})$. ∎

2.4 Proof of Theorem 4

The proof essentially amounts to applying the method developed by Bolthausen (1986) to the random variables

$$X_i = \delta_{(x^i_{[0,T]},\omega^i)} \quad (i = 1, \ldots, N). \tag{2.25}$$

Strictly speaking, this method only applies to random variables taking values in certain "nice" Banach spaces, namely Banach spaces of type 2 (such as L^p-spaces with $2 \le p < \infty$). Unfortunately, $\mathcal{M}_1(C[0,T] \times \mathbb{R})$ is not in this class. However, this problem can be circumvented via a trick due to Ben Arous and Brunaud (1990), which consists of mapping $\mathcal{M}_1(C[0,T] \times \mathbb{R})$ into a Banach space of type 2. In this section we *formally* compute the covariance operator according to Bolthausen's recipe (Steps 1-3 below) and check its strict positivity (I-II below), which is the key to having a central limit theorem. The change of variable trick, which provides *rigorous justification* for what is done here and which requires the use of Assumption (A2), will be given in Appendix B.

Step 1. We start by letting ν_* be the law of the $\mathcal{M}_1(C[0,T] \times \mathbb{R})$-valued random variable $\delta_{(x_{[0,T]},\omega)} - Q_*$ induced by Q_*. For $R \in \mathcal{M}_1(C[0,T] \times \mathbb{R})$ and $\phi \in C_b$ we write $\phi(R) = \int \phi dR$ and $\phi^* = \phi(Q_*)$. The *free* covariance operator $(\Gamma(\phi, \psi))_{\phi,\psi \in C_b}$ is defined by

$$\begin{aligned} \Gamma(\phi,\psi) &= \int \phi(R)\psi(R)\nu_*(dR) \\ &= E^{Q_*}\{[\phi(x_{[0,T]},\omega) - \phi^*][\psi(x_{[0,T]},\omega) - \psi^*]\} \\ &= \text{Cov}_{\,Q_*}(\phi,\psi). \end{aligned} \tag{2.26}$$

The meaning of this operator is that the field

$$\left(N^{1/2}[\int \phi dL_N - \phi^*]\right)_{\phi \in C_b} \tag{2.27}$$

converges, under $Q_*^{\otimes N}$ as $N \to \infty$, to a Gaussian field with covariance $\Gamma(\phi, \psi)$. This follows from the standard central limit theorem for i.i.d. $\mathbb{R}$-valued random variables. We remark that

$$\Gamma(\phi,\psi) = D^2 H(Q_*|W \otimes \mu)[\hat{\phi}, \hat{\psi}], \tag{2.28}$$

as is easily proved from (1.9) via direct computation. Here the second derivative $D^2 H$ is defined in the usual directional sense (Fréchet derivative).

Step 2. For a given $\phi \in C_b$, let $\hat{\phi} \in \mathcal{M}_0(C[0,T] \times \mathbb{R})$ be the signed measure on $C[0,T] \times \mathbb{R}$ with zero total mass defined by

$$\hat{\phi} = \int R\phi(R)\nu_*(dR), \tag{2.29}$$

i.e., for $A \subset C[0,T] \times \mathbb{R}$ measurable,

$$\begin{aligned} \hat{\phi}(A) &= \int R(A)\phi(A)\nu_*(dR) \\ &= \int Q_*(dx_{[0,T]}, d\omega)[\delta_{(x_{[0,T]},\omega)}(A) - Q_*(A)][\phi(x_{[0,T]},\omega) - \phi^*] \\ &= \text{Cov}_{\,Q_*}(1_A, \phi) \end{aligned} \tag{2.30}$$

where 1_A is the characteristic function of A. Then Bolthausen's theorem states that (modulo the change of variable trick and some regularity assumptions on the function $Q \to F(Q)$ in

(1.5), all to be discussed in Appendix B) the field in (2.27) converges, under P_N as $N \to \infty$, to a Gaussian field with covariance

$$\Delta(\phi,\psi) = \Gamma(\phi,\psi) - D^2F(Q_*)[\hat{\phi},\hat{\psi}] \tag{2.31}$$

(recall Lemma 1), *provided* $\Delta(\phi,\phi) > 0$ for all ϕ such that $\hat{\phi} \not\equiv 0$.

Step 3. By combining (2.31) and (2.28) with (1.8), we get

$$\Delta(\phi,\psi) = D^2I(Q_*)[\hat{\phi},\hat{\psi}]. \tag{2.32}$$

Thus the requirement $\Delta(\phi,\phi) > 0$ can be interpreted as saying that the rate function $Q \to I(Q)$ must have finite curvature at its unique minimum Q_*.

The rest of the proof consists of showing the following two facts. Let $C(\phi,\psi)$ be the covariance defined in (1.33). Then

$$\begin{aligned} &I.\ \ C(\phi,\psi) = \Delta(\phi,\psi) \\ &II.\ \ C(\phi,\phi) > 0 \ \text{ for all } \ \phi \ \text{ such that } \ \hat{\phi} \not\equiv 0. \end{aligned} \tag{2.33}$$

Proof of I.
For simplicity we assume $\phi = \psi$. The proof for the general case follows the same argument. We first note that, by (2.30), $\hat{\phi} \ll Q_*$ and

$$\frac{d\hat{\phi}}{dQ_*} = \phi - \phi^*. \tag{2.34}$$

Using the expression (recall (1.14) and (2.6))

$$F(Q) = E^Q\Big\{ -\frac{1}{2}\int_0^T (\beta^{\omega,\Pi_t Q}(x_t))^2 dt + \int_0^T \beta^{\omega,\Pi_t Q}(x_t)dx_t \Big\} \tag{2.35}$$

we get, by a lengthy but straightforward computation via (1.11),

$$\begin{aligned} D^2F(Q)[\hat{\phi},\hat{\phi}] \ = \ & -\ E^Q \textstyle\int_0^T [\gamma^{\omega,\Pi_t\hat{\phi}}(x_t)]^2 dt \\ & -\ 2\textstyle\int \hat{\phi}(dx_{[0,T]},d\omega) \int_0^T \beta^{\omega,\Pi_t Q}(x_t)\gamma^{\omega,\Pi_t\hat{\phi}}(x_t)dt \\ & +\ 2\textstyle\int \hat{\phi}(dx_{[0,T]},d\omega) \int_0^T \gamma^{\omega,\Pi_t\hat{\phi}}(x_t)dx_t \end{aligned} \tag{2.36}$$

with

$$\gamma^{\omega,\Pi_t\hat{\phi}}(x) = \int \hat{\phi}(dy_{[0,T]},d\pi)\hat{f}'(y_t - x;\omega,\pi). \tag{2.37}$$

(The computation becomes elementary once we realize that, due to (2.34), the Itô-integrals make sense under $\hat{\phi}$.)

Now let $w_t^\omega = x_t - \int_0^t \beta^{\omega,\Pi_s Q_*} ds$ (which is a Brownian motion under Q_*^ω). Then by (2.26), (2.31), (2.34) and (2.36) we have

$$\begin{aligned}
\Delta(\phi,\phi) &= \Gamma(\phi,\phi) - D^2F(Q_*)[\hat\phi,\hat\phi] \\
&= E^{Q_*}\{[\phi(x_{[0,T]},\omega)-\phi^*]^2\} + E^{Q_*}\{\textstyle\int_0^T [\gamma^{\omega,\Pi_t\hat\phi}(x_t)]^2 dt\} \\
&\quad +2E^{Q_*}\{[\phi(x_{[0,T]},\omega)-\phi^*]\textstyle\int_0^T \gamma^{\omega,\Pi_t\hat\phi}(x_t)dw_t^\omega\} \\
&= E^{Q_*}\{[\phi(x_{[0,T]},\omega)-\phi^*]^2\} + E^{Q_*}\{[\textstyle\int_0^T \gamma^{\omega,\Pi_t\hat\phi}(x_t)dw_t^\omega]^2\} \\
&\quad +2E^{Q_*}\{[\phi(x_{[0,T]},\omega)-\phi^*]\textstyle\int_0^T \gamma^{\omega,\Pi_t\hat\phi}(x_t)dw_t^\omega\} \qquad (2.38)\\
&= E^{Q_*}\left\{\left[\phi(x_{[0,T]},\omega)-\phi^* + \textstyle\int_0^T \gamma^{\omega,\Pi_t\hat\phi}(x_t)dw_t^\omega\right]^2\right\} \\
&= C(\phi,\phi),
\end{aligned}$$

where in the second equality we have used the standard isometry property of integration w.r.t. Brownian motion. [5] ∎

Proof of II.
Suppose $\phi \in C_b$ is such that $C(\phi,\phi)=0$. It is not restrictive to assume $\phi^*=0$. We want to show that $\hat\phi \equiv 0$, i.e., $\phi = 0$ Q_*-a.s. Define the following σ-field on $C[0,T]\times \mathbb{R}$

$$\mathcal{F}_t = \sigma\{x_s : 0\le s\le t\}\otimes\mathcal{B} \qquad (2.39)$$

with $\mathcal{B}$ denoting the Borel σ-field on $\mathbb{R}$. Let

$$\phi_t(x_{[0,t]},\omega) = E^{Q_*}\{\phi|\mathcal{F}_t\}. \qquad (2.40)$$

According to (1.33-1.34), $C(\phi,\phi)=0$ implies

$$\phi(x_{[0,T]},\omega) = \int_0^T \Big[\int Q_*(dy_{[0,T]},d\pi)\phi(y_{[0,T]},\pi)\hat f'(y_t-x_t;\omega,\pi)\Big]dw_t^\omega \quad Q_*-a.s. \qquad (2.41)$$

Taking conditional expectation and using the fact that the integral in the r.h.s. of (2.41) is an $\mathcal{F}_t$-martingale, we get

$$\phi_t(x_{[0,t]},\omega) = \int_0^t \Big[\int Q_*(dy_{[0,T]},d\pi)\phi_t(y_{[0,t]},\pi)\hat f'(y_s-x_s;\omega,\pi)\Big]dw_s^\omega \quad Q_*-a.s. \qquad (2.42)$$

Thus, using again the isometry property of integration w.r.t. Brownian motion, we obtain

$$\begin{aligned}
\|\phi_t\|^2_{L^2(Q_*)} &= \left\| \textstyle\int_0^t \left[\int Q_*(dy_{[0,T]},d\pi)\phi_t(y_{[0,t]},\pi)\hat f'(y_s-x_s;\omega,\pi)\right]dw_s^\omega\right\|^2_{L^2(Q_*)} \\
&= E^{Q_*}\left\{\textstyle\int_0^t \left[\int Q_*(dy_{[0,T]},d\pi)\phi_t(y_{[0,t]},\pi)\hat f'(y_s-x_s;\omega,\pi)\right]^2 dt\right\} \qquad (2.43)\\
&\le t\|\hat f'\|^2_\infty \|\phi_t\|^2_{L^2(Q_*)}
\end{aligned}$$

which implies $\phi_t = 0$ Q_*-a.s. for $t\in[0,1/\|\hat f'\|^2_\infty)$. It is easy to see that this argument can be repeated, and so we get $\phi_t=0$ Q_*-a.s. for $t\in[0,T]$. Since $\phi_T=\phi$ the conclusion follows. ∎

[5] Let $(w_t)_{t\in[0,T]}$ be a Brownian motion. Let $(\xi_t)_{t\in[0,T]}$ be a stochastic process, adapted to the filtration generated by $(w_t)_{t\in[0,T]}$, such that $E(\int_0^T \xi_t^2 dt) < \infty$. Then the following equality holds: $E(\int_0^T \xi_t^2 dt) = E([\int_0^T \xi_t dw_t]^2)$.

3 Spin-flip systems

All the results stated in Section 1, together with their proofs in Section 2, can be modified in an essentially straightforward manner to cover the case of spin-flip systems. In this section we formulate these modifications and indicate which parts of their proofs are not trivially obtained from the corresponding parts for diffusions. We follow the same order as in Section 1.

3.1 The model

Let $H_N : \{-1,+1\}^N \times \mathbb{R}^N \to \mathbb{R}$ be the N-particle Hamiltonian given by

$$H_N(\underline{x},\underline{\omega}) = \frac{1}{2N}\sum_{i,j=1}^{N} f(\omega^i,\omega^j)x^i x^j + \sum_{i=1}^{N} g(\omega^i)x^i, \tag{3.1}$$

where $\underline{x} = (x^i)_{i=1}^N$ is the state variable and $\underline{\omega} = (\omega^i)_{i=1}^N$ is the medium variable. As for diffusions, the ω^i are i.i.d. random variables with common law μ. Moreover, the functions f, g are assumed to be bounded and continuous.

For given $\underline{\omega}$, let $\underline{x}_t = (x_t^i)_{i=1}^N$ be the N-spin system defined to be the Markov chain with infinitesimal generator $\mathcal{G}$, acting on functions $\phi : \{-1,+1\}^N \to \mathbb{R}$ as follows:

$$(\mathcal{G}\phi)(\underline{x}) = \sum_{i=1}^{N} c_N^{\omega}(i,\underline{x})[\phi(\underline{x}^i) - \phi(\underline{x})]. \tag{3.2}$$

Here, $\underline{x}^i$ is the state obtained from $\underline{x}$ by flipping the i-th spin x^i, and

$$\begin{aligned} c_N^{\omega}(i,\underline{x}) &= \exp\left[\tfrac{1}{2}\left\{H_N(\underline{x},\underline{\omega}) - H_N(\underline{x}^i,\underline{\omega})\right\}\right] \\ &= \exp\left[\tfrac{1}{N}\textstyle\sum_{j=1,j\neq i}^{N} \hat{f}(\omega^i,\omega^j)x^i x^j + g(\omega^i)x^i\right] \end{aligned} \tag{3.3}$$

with $\hat{f}(\omega,\pi) = f(\omega,\pi) + f(\pi,\omega)$. For every $\underline{\omega}$, (3.2) has a reversible equilibrium measure proportional to $\exp[-H_N(\underline{x},\underline{\omega})]$. The initial condition $\underline{x}_0$ is assumed to have product distribution $\lambda^{\otimes N}$, where λ is any probability measure on $\{-1,+1\}$. The path space for a single spin is $D[0,T]$, the space of right-continuous piecewise-constant functions from $[0,T]$ to $\{-1,+1\}$. This space has a topology and a Borel σ-field, provided by the Skorohod metric; see e.g. Ethier and Kurtz (1986), p. 117.

We denote by $W^{\otimes N}$ the law of the N-spin system whose generator has the form (3.2) with $c_N^{\omega} \equiv 1$. All other notations introduced in Section 1 ($P_N^{\omega}, L_N, P_N, \ldots$ etc.) are left unchanged.

3.2 Empirical measure and large deviations

The analogues of Lemma 1 and Theorem 1 read as follows.

Lemma 3 *For given* $\underline{\omega}$

$$\frac{dP_N^{\omega}}{dW^{\otimes N}}(\underline{x}_{[0,T]}) = \exp[NF(L_N(\underline{x}_{[0,T]},\underline{\omega})) + O(1)] \tag{3.4}$$

where for $Q \in \mathcal{M}_1(D[0,T] \times \mathbb{R})$

$$\begin{aligned} F(Q) &= \int Q(dx_{[0,T]}, d\omega)\Big\{ \int_0^T dt \Big(1 - \exp\Big[\int Q(dy_{[0,T]}, d\pi)\hat{f}(\omega,\pi)x_t y_t + g(\omega)x_t\Big]\Big) \\ &\quad + \tfrac{1}{2} \int Q(dy_{[0,T]}, d\pi)\Big[\hat{f}(\omega,\pi)(x_T y_T - x_0 y_0) + g(\omega)(x_T - x_0)\Big]\Big\}. \end{aligned} \tag{3.5}$$

The proof of Lemma 3 relies on Girsanov's formula for spin-flip systems, which is easily derived from Girsanov's formula for point processes (see Comets (1987) or Lipster and Shiryaev (1988), Theorem 19.3).

Theorem 5 $(P_N)_{N\geq 1}$ *satisfies the LDP with rate function*

$$I(Q) = H(Q|W \otimes \mu) - F(Q). \tag{3.6}$$

This follows from Lemma 3 as for diffusions. The technical difference is that the martingale term in the Girsanov formula is not driven by a Brownian motion but by a compensated Poisson process.

3.3 McKean-Vlasov equation

Given $Q \in \mathcal{M}_1(D[0,T] \times \mathbb{R})$ and $\omega \in \mathbb{R}$, let $P^{\omega,Q}$ be the law of the single-spin system whose initial distribution is λ and whose rate of flipping from x to $-x$ at time t is given by $c^{\omega,\Pi_t Q}(x)$, where for $q \in \mathcal{M}_1(\{-1,1\} \times \mathbb{R})$

$$c^{\omega,q}(x) = \exp\Big[x\Big(\int q(dy, d\pi) f(\omega,\pi) y + g(\omega)\Big)\Big]. \tag{3.7}$$

In analogy with Lemma 2 and Corollary 1, the next facts are easily proved.

Lemma 4 *For all* Q

$$F(Q) = \int Q(dx_{[0,T]}, d\omega) \log \frac{dP^{\omega,Q}}{dW}(x_{[0,T]}). \tag{3.8}$$

Corollary 3 *For all* Q

$$I(Q) = H(Q|P^Q), \tag{3.9}$$

where $P^Q \in \mathcal{M}_1(D[0,T] \times \mathbb{R})$ *is defined by*

$$P^Q(dx_{[0,T]}, d\omega) = \mu(d\omega) P^{\omega,Q}(dx_{[0,T]}). \tag{3.10}$$

The next theorem is the analogue of Theorem 2. Define ν^Q as in (1.18).

Theorem 6 *Equation (3.9) has a unique solution* Q_* *which has the following properties:*

1. $\nu^{Q_*} = \mu$.
2. Q_*^ω *is the law of a Markov chain on* $\{-1,+1\}$ *for* μ*-a.s. all* ω.

3. *Let* $q_t^\omega = \Pi_t Q_*^\omega$. *Then* q_t^ω *solves the differential equation*

$$\begin{cases} \frac{\partial}{\partial t} q_t^\omega = \mathcal{L}^\omega q_t^\omega & (t \in (0,T], \omega \in \mathbb{R}) \\ q_0^\omega = \lambda \end{cases} \tag{3.11}$$

where $\mathcal{L}^\omega$ *is the nonlinear operator*

$$(\mathcal{L}^\omega q_t^\omega)(x) = q_t^\omega(-x)c^{\omega,q_t}(-x) - q_t^\omega(x)c^{\omega,q_t}(x) \quad (\omega \in \mathbb{R}) \tag{3.12}$$

and q_t *is defined by* $q_t(x, d\omega) = \mu(d\omega)q_t^\omega(x)$.

4. *Under* Q_*^ω *the rate of flipping from* x *to* $-x$ *at time* t *for the Markov chain in 2. is* c^{ω,q_t}.

The only essential difference with the proof of Theorem 2 is the part concerning the uniqueness of the solution of (3.11), which is much easier here. Indeed, via the relation $q_t^\omega(-1)+q_t^\omega(+1)= 1$ for all ω and t, (3.11) can be rewritten as an equation for $q_t^\omega(+1)$, thought of as an element of $L^\infty(\mu)$. The *coupled* family of equations in (3.11), indexed by $\omega \in \mathbb{R}$, is an ordinary differential equation in the Banach space $L^\infty(\mu)$ driven by a locally Lipschitz vector field. Uniqueness follows by classical arguments (Brezis (1983), Theorem VII.3).

3.4 Empirical flow and large deviations

The definitions of ℓ_N and $\wp_N$ are the same as in Section 1 (see (1.25) and (1.26)). For p a probability measure on $\{-1,+1\} \times \mathbb{R}$ and $\omega \in \mathbb{R}$, define $\Psi_p^\omega : \mathbb{R}^{\{-1,+1\}} \to \mathbb{R}^+$ by

$$\Psi_p^\omega(\lambda) = \sup_{\delta \in \mathbb{R}^{\{-1,+1\}}} \left\{ \sum_{x=\pm 1} \left[\lambda(x)\delta(x) - p^\omega(x)c^{\omega,p}(x)\left(e^{\hat{\delta}(x)} - \hat{\delta}(x) - 1\right)\right]\right\} \tag{3.13}$$

where $\hat{\delta}(x) = \delta(-x) - \delta(x)$. Defining Δ as in (1.28), we obtain the following analogue of Theorem 3.

Theorem 7 $(\wp_N)_{N\geq 1}$ *satisfies the LDP with rate function*

$$i(q_{[0,T]}) = \begin{cases} \int_0^T dt\left\{\int \nu^q(d\omega)\Psi_{q_t}^\omega\left(\frac{\partial q_t^\omega}{\partial t} - \mathcal{L}^\omega q_t^\omega\right)\right\} + H(\nu^q|\mu) & \text{if } q_{[0,T]} \in \Delta \\ \infty & \text{otherwise.} \end{cases} \tag{3.14}$$

(For the model *without* random field a different representation for i is given in Comets (1987).)

The proof of Theorem 7 is not a trivial modification of the proof of Theorem 3. We therefore give a sketch here (Steps 1-3 below).

Step 1. Fix a flow $q_{[0,T]} \in \Delta$. Suppose that there exists a $Q \in \mathcal{M}_1(D[0,T] \times \mathbb{R})$ such that $I(Q) < \infty$ and Q minimizes I under the constraint $\Pi_t Q = q_t$ for $t \in [0,T]$. Then, as for diffusions, it can be shown that Q^ω is Markovian for μ almost all ω (e.g. by using the notion of h-process; see Föllmer (1988), Theorem 1.31). Let us denote by $k_t^\omega(x_t)$ the flip rate of this process at time t. Then from Girsanov's formula for spin processes we get

$$I(Q) = \int_0^T dt\left\{\int \nu^q(d\omega)\left[\sum_{x=\pm 1} q_t^\omega(x)\left(c^{\omega,q_t}(x) - k_t^\omega(x) + k_t^\omega(x)\log\frac{k_t^\omega(x)}{c^{\omega,q_t}(x)}\right)\right]\right\}. \tag{3.15}$$

Step 2. Write the identity

$$\sum_{x=\pm1} q_t^\omega(x)\Big(c^{\omega,q_t}(x) - k_t^\omega(x) + k_t^\omega(x)\log\frac{k_t^\omega(x)}{c^{\omega,q_t}(x)}\Big)$$

$$= \sup_{\delta\in\mathbb{R}^{\{-1,+1\}}} \sum_{x=\pm1} q_t^\omega(x)\Big[\delta(x)\Big(k_t^\omega(x) - c^{\omega,q_t}(x)\Big) - c^{\omega,q_t}(x)\Big(e^{\delta(x)} - \delta(x) - 1\Big)\Big], \tag{3.16}$$

which is easily checked by noting that the supremum is attained at $\delta = \delta_*$ given by $\delta_*(x) = \log(k_t^\omega(x)/c^{\omega,q_t}(x))$. We claim that the r.h.s. of (3.16) equals

$$\sup_{\delta\in\mathbb{R}^{\{-1,+1\}}} \sum_{x=\pm1} q_t^\omega(x)\Big[\hat\delta(x)\Big(k_t^\omega(x) - c^{\omega,q_t}(x)\Big) - c^{\omega,q_t}(x)\Big(e^{\hat\delta(x)} - \hat\delta(x) - 1\Big)\Big] \tag{3.17}$$

(which is the same as the r.h.s. of (3.16) but with δ replaced by $\hat\delta$). This will be shown below. From (3.17), together with the identities

$$\begin{aligned}\sum_{x=\pm1} q_t^\omega(x)\hat\delta(x)[k_t^\omega(x) - c^{\omega,q_t}(x)] &= \sum_{x=\pm1} \delta(x)[q_t^\omega(x)(k_t^\omega(x) - c^{\omega,q_t}(x)]\\ &= \sum_{x=\pm1} \delta(x)[\tfrac{\partial}{\partial t} q_t^\omega(x) - \mathcal{L}^\omega q_t^\omega(x)],\end{aligned} \tag{3.18}$$

we get $I(Q) = i(q_{[0,T]})$. The second equality in (3.18) uses (3.11) and (3.12) with k_t^ω replacing c^{ω,q_t}. The proof can now be completed as for Theorem 3.

Step 3. We still have to show that (3.16) equals (3.17), which amounts to verifying that $\delta_* = \hat\gamma$ for some $\gamma \in \mathbb{R}^{\{-1,+1\}}$. This is equivalent to saying that $\sum_{x=\pm1}\delta_*(x) = 0$ or

$$k_t^\omega(x) = c^{\omega,q_t}(x)e^{\lambda_t x} \qquad \text{for some } \lambda_t \in \mathbb{R}. \tag{3.19}$$

There are various ways of checking (3.19). The most direct and elementary way consists of looking for the minimum of (3.15) (w.r.t. the rates $k_t^\omega(x)$) under the constraint

$$\frac{\partial q_t^\omega(x)}{\partial t} = q_t^\omega(-x)k_t^\omega(-x) - q_t^\omega(x)k_t^\omega(x) \quad (t\in(0,T]). \tag{3.20}$$

The classical method of Lagrange multipliers shows that the k_t^ω realizing the minimum must have the form (3.19) (we already know that the minimum exists). The details are straightforward.

Theorem 7 shows that the large deviations for the empirical flow are controlled by the positive convex functions Ψ_p^ω. These are *not* norms squared, unlike for diffusions. To appreciate the analogy between Theorem 3 and Theorem 7, note that we could have used in Theorem 3 the following expression equivalent to (1.27) (Dawson and Gärtner (1987)):

$$\|\psi^*\|_p^2 = \sup_{\phi\in\mathcal{D}}\Big\{\langle\psi^*,\phi\rangle - \frac{1}{2}\langle p,\phi'^2\rangle\Big\}. \tag{3.21}$$

3.5 Central limit theorem

The CLT for spin systems will be proved under the following assumption which, for technical reasons that will be explained in Appendix B, is much stronger than the corresponding Assumption (A2) for diffusions:

(A3) There exist a finite set $X \subset \mathbb{R}$ and functions $\alpha_i, \beta_i : \mathbb{R} \to X$ $(i = 1, \dots, p)$ such that

$$f(\omega, \pi) = \sum_{i=1}^{p} \alpha_i(\omega)\beta_i(\pi). \tag{3.22}$$

We note that Assumption (A3) is satisfied in two relevant cases: (i) when f is constant, i.e., the medium does not affect the interaction (e.g. the Curie-Weiss model in Section 4); (ii) when the support of the medium law μ is finite.

For $x_{[0,T]} \in D[0,T]$, we let $J_t(x_{[0,T]})$ be the number of jumps of the path $x_{[0,T]}$ up to and including time t.

Theorem 8 *Let C_b be the set of bounded continuous functions from $D[0,T] \times \mathbb{R}$ to $\mathbb{R}$. As $N \to \infty$ the field*

$$\left(N^{1/2}\Big[\int \phi dL_N - \int \phi dQ_*\Big]\right)_{\phi \in C_b} \tag{3.23}$$

converges under P_N to a Gaussian field with covariance

$$C(\phi, \psi) = \int Q_*(dx_{[0,T]}, d\omega)\phi[Q_*](x_{[0,T]}, \omega)\psi[Q_*](x_{[0,T]}, \omega), \tag{3.24}$$

where

$$\begin{aligned} \phi[Q_*](x_{[0,T]}, \omega) &= \phi(x_{[0,T]}, \omega) - \phi^* \\ &+ \textstyle\int_0^T \left(\int Q_*(dy_{[0,T]}, d\pi)[\phi(y_{[0,T]}, \pi) - \phi^*]y_t \hat{f}(\omega, \pi)\right) dw_t^\omega \end{aligned} \tag{3.25}$$

with $\phi^* = \int \phi dQ_*$ (similarly for ψ) and $w_t^\omega = J_t(x_{[0,T]}) - \int_0^t c^{\omega, \Pi, Q_*}(x_s)ds$ (which is a martingale under Q_*^ω).

The part of the proof of the CLT for diffusions, contained in Section 2.4, extends readily to spin systems. The part concerning the change of variable trick will be sketched at the end of Appendix B.

4 Two applications

In this section we describe two examples of systems where the random medium controls the phase diagram. The phases of the system correspond to the *stationary* solutions of the McKean-Vlasov equation that are *stable* under small perturbations. [6] We shall assume that the law μ of the random medium components is *symmetric*. More in particular, we shall consider the following two subcases:

Case I. $\mu(d\omega) = \phi(\omega)d\omega$ with $\phi(\omega) = \phi(-\omega)$ and $\omega \to \phi(\omega)$ non-increasing on $\mathbb{R}^+$.
Case II. $\mu = \frac{1}{2}(\delta_\eta + \delta_{-\eta})$ with $\eta > 0$.

[6] Thermodynamically this includes both the stable and the metastable phases.

4.1 Curie-Weiss model

The Curie-Weiss model in random magnetic field is the spin-flip system driven by the Hamiltonian (3.1) with

$$\begin{array}{lcl} f(\omega,\pi) & \equiv & -\beta \\ g(\omega) & = & -\beta\omega \quad (\omega,\pi\in\mathbb{R}) \end{array} \tag{4.1}$$

where $\beta \in (0,\infty)$ is the *inverse temperature*. With this choice, (3.1-3.3) describe a system of mean-field ferromagnetically coupled spins, each with its own random magnetic field and subject to Glauber dynamics. The two terms in the Hamiltonian have opposite effects: f tends to align the spins, g tends to point each spin in the direction of its local field.

The order parameter of the system is the *magnetization*

$$\begin{array}{lcl} m_t(\omega) & = & \sum_{x\pm 1} x q_t^\omega(x) \\ m_t & = & \int_{\mathbf{R}} m_t(\omega)\mu(d\omega), \end{array} \tag{4.2}$$

where $q_t^\omega(x)$ is the probability that a typical spin is in state x at time t in the medium ω (in the McKean-Vlasov limit). Written in terms of (4.2), the McKean-Vlasov equation (3.11-3.12) reads

$$\begin{array}{lcl} \frac{\partial}{\partial t}m_t(\omega) & = & (1-m_t(\omega))\exp[\beta(m_t+\omega)] - (1+m_t(\omega))\exp[-\beta(m_t+\omega)] \\ & = & 2\sinh[\beta(m_t+\omega)] - 2m_t(\omega)\cosh[\beta(m_t+\omega)]. \end{array} \tag{4.3}$$

The stationary solutions of (4.3) have been investigated by Salinas and Wreszinski (1985).

1. Stationary solution(s). Any stationary solution of (4.3) is of the form

$$m(\omega) = \tanh[\beta(m+\omega)], \tag{4.4}$$

where m must satisfy the consistency relation (see (4.2))

$$\begin{array}{lcl} m & = & \Gamma_\beta(m) \\ \Gamma_\beta(m) & = & \int_{\mathbf{R}} \tanh[\beta(m+\omega)]\mu(d\omega). \end{array} \tag{4.5}$$

It follows from (4.5) that

$$\begin{array}{l} \Gamma_\beta(\pm\infty) = \pm 1 \\ \Gamma'_\beta(m) = \beta\int_{\mathbf{R}} \frac{1}{\cosh^2[\beta(m+\omega)]}\mu(d\omega) > 0. \end{array} \tag{4.6}$$

Since μ is symmetric, we have $\Gamma_\beta(0) = 0$ for all β, so that (4.5) always has the paramagnetic solution $m = 0$. To investigate under what conditions ferromagnetic solutions $m > 0$ may occur, we distinguish between the two subcases I and II.

Case I. Γ_β now has the following property:

Fact 1 *For every β:* sign $\Gamma''_\beta(m) = -$sign m.

Proof. Compute, using the symmetry of ϕ,

$$\begin{aligned}\Gamma''_\beta(m) &= \beta \int_{\mathbf{R}} \frac{\partial}{\partial\omega}\left(\frac{1}{\cosh^2[\beta(m+\omega)]}\right)\phi(\omega)d\omega \\ &= -\beta \int_{\mathbf{R}} \frac{1}{\cosh^2[\beta(m+\omega)]} d\phi(\omega) \\ &= -\beta \int_0^\infty \left(\frac{1}{\cosh^2[\beta(m+\omega)]} - \frac{1}{\cosh^2[\beta(m-\omega)]}\right) d\phi(\omega).\end{aligned} \tag{4.7}$$

In the last integral, the difference between brackets has the oppsite sign as m for all $\omega \geq 0$, because $x \to 1/\cosh^2(x)$ is symmetric and unimodal. By the unimodality of ϕ, we have $d\phi(\omega) \leq 0$ and the claim follows. ∎

Thus, by Fact 1, if

$$\Gamma'_\beta(0) = \beta \int_{\mathbf{R}} \frac{1}{\cosh^2[\beta\omega]} \mu(d\omega) > 1, \tag{4.8}$$

then (4.5) has *exactly one* ferromagnetic solution $m = m^*(\beta) > 0$.

Next we investigate (4.8).

Fact 2 *(a) There exists* $1 < \beta_c = \beta_c(\phi) \leq \infty$ *such that (4.8) holds iff* $\beta > \beta_c$.
(b) $\beta_c(\phi) < \infty$ *iff* $\phi(0) > \frac{1}{2}$.

Proof. (a) To prove the existence of a unique critical value β_c, it suffices to show that $\beta \to \Gamma'_\beta(0)$ is non-decreasing. This is done as follows. Compute

$$\frac{\partial}{\partial\beta}\Gamma'_\beta(0) = \int_{\mathbf{R}} h_\beta(\omega)\phi(\omega)d\omega \tag{4.9}$$

where

$$h_\beta(\omega) = \frac{1}{\cosh^2(\beta\omega)}\Big[1 - 2\beta\omega\tanh(\beta\omega)\Big]. \tag{4.10}$$

Since h_β and ϕ are symmetric, we have

$$\frac{\partial}{\partial\beta}\Gamma'_\beta(0) = 2\int_0^\infty h_\beta(\omega)\phi(\omega)d\omega. \tag{4.11}$$

Next, let ω^* be the unique positive solution of the equation $2\beta\omega\tanh(\beta\omega) = 1$. Then $h_\beta(\omega)$ changes from positive to negative as ω increases through ω^*. Hence, by the unimodality of ϕ,

$$\begin{aligned}\int_0^\infty h_\beta(\omega)\phi(\omega)d\omega &\geq \phi(\omega^*)\Big[\int_0^{\omega^*} h_\beta(\omega)d\omega + \int_{\omega^*}^\infty h_\beta(\omega)d\omega\Big] \\ &= \phi(\omega^*)\int_0^\infty h_\beta(\omega)d\omega.\end{aligned} \tag{4.12}$$

But $h_\beta(\omega) = (\partial/\partial\omega)[\omega/\cosh^2(\beta\omega)]$, which makes the last integral equal to zero. This proves the existence of β_c. The inequality $\beta_c > 1$ follows from $\Gamma'_\beta(0) < \beta$.
(b) Simply note that

$$\lim_{\beta\to\infty}\Gamma'_\beta(0) = \lim_{\beta\to\infty}\int_{\mathbf{R}} \frac{1}{\cosh^2(x)}\phi\Big(\frac{x}{\beta}\Big)dx = 2\phi(0). \tag{4.13}$$

Facts 1 - 2 show that in the unimodal case the situation is qualitatively similar to the standard Curie-Weiss model in zero magnetic field (for which $\Gamma_\beta(0) = \beta$ and hence $\beta_c = 1$). The only difference is that possibly $\beta_c = \infty$, which occurs when the peak of ϕ is sufficiently low. This corresponds to large randomness, which destroys the spin ordering at arbitrarily low temperature.

Case II. In the bimodal case the situation is more complex. If

$$\Gamma'_\beta(0) = \frac{\beta}{\cosh^2(\beta\eta)} > 1, \tag{4.14}$$

then obviously there is *at least one* ferromagnetic solution. However, Fact 1 is no longer true in general and therefore there may be a ferromagnetic solution even when (4.14) fails. In fact, then there must be *at least two* ferromagnetic solutions (corresponding to the curve $m \to \Gamma_\beta(m)$ crossing the diagonal first from below and then from above).

The regime defined by (4.14) lies under the curve

$$\beta \to \eta(\beta) = \frac{1}{\beta}\text{arccosh}(\sqrt{\beta}) \quad (\beta \in [1,\infty)). \tag{4.15}$$

This curve is unimodal, with endpoints $\eta(1) = \eta(\infty) = 0$, maximum at $\beta_1 = 1.72\ldots$, and maximal value $\eta_1 = \eta(\beta_1) = 1/2\sqrt{\beta_1(\beta_1 - 1)} = 0.45\ldots$.

An idea of when two ferromagnetic solutions occur may be obtained from the small-m expansion

$$\Gamma_\beta(m) = \frac{1}{c^2}\beta m + \frac{2c^2 - 3}{3c^4}\beta^3 m^3 + O(m^5) \text{ with } c = \cosh(\beta\eta). \tag{4.16}$$

On the curve defined by (4.15) (i.e., $c^2 = \beta$), this expansion reduces to $\Gamma_\beta(m) = m + \beta(\frac{2}{3}\beta - 1)m^3 + O(m^5)$, from which we see that $\beta_2 = \frac{3}{2}$ is a critical value. Indeed, if $\beta > \beta_2$, then as η increases through $\eta(\beta)$ (i.e., β/c^2 decreases through 1) at least two ferromagnetic solutions $m_2 > m_1 > 0$ occur, because $m \to \Gamma_\beta(m)$ is convex for small m.

The full phase diagram is drawn in Figure 1, which is obtained numerically. There are three phases, corresponding to 0, 1 resp. 2 ferromagnetic solutions. The lower separation line is the curve in (4.15). The upper separation line corresponds to the choice of parameters where there exists $m > 0$ such that $\Gamma_\beta(m) = m, \Gamma'_\beta(m) = 1$. (The latter curve moves up to 1 because $\Gamma_\beta(m)$ tends to the step function at $m = \eta$ as $\beta \to \infty$.) Note that the two curves coincide for $\beta \in [1, \beta_2]$ and separate at the "tricritical point" (β_2, η_2) with $\eta_2 = \eta(\beta_2)$. The picture shows that a phase transition occurs at some $\beta_c = \beta_c(\eta) < \infty$ iff $\eta \in (0,1)$. The phase transition is second order when $\eta \in (0, \eta_2]$ and first order when $\eta \in (\eta_2, 1)$ (i.e., the ferromagnetic solution appears discontinuously). Interestingly, if $\eta \in (\eta_2, \eta_1)$, then as β increases we get phases 0, 2, 1 and again 2. [7]

2. Linear stability. A stationary solution corresponds to a phase of the system iff it is stable under small perturbations. To check stability we linearize the McKean-Vlasov equation (4.3) about its stationary solutions, as follows.

[7] Inside phase 2 there is a phase coexistence line (not drawn), above which the paramagnetic solution is stable and the ferromagnetic solution is metastable, and below which the reverse is true. See Salinas and Wrezinski (1985) and recall footnote 6.

Rewrite (4.3) as

$$\frac{\partial}{\partial t} m_t(\omega) = \Theta_\omega(m_t(\omega)). \tag{4.17}$$

Let $m(\cdot)$ be given by (4.4) and (4.5). Then the Fréchet derivative of Θ_ω at $m(\cdot)$ is given by

$$\begin{aligned} & D\Theta_\omega(m(\cdot))[n(\cdot)] \\ &= 2\beta n\Big(\cosh[\beta(m+\omega)] - m(\cdot)\sinh[\beta(m+\omega)]\Big) - 2n(\cdot)\cosh[\beta(m+\omega)] \\ &= 2\beta n \tfrac{1}{\cosh[\beta(m+\omega)]} - 2n(\cdot)\cosh[\beta(m+\omega)], \end{aligned} \tag{4.18}$$

where $n = \int n(\omega)\mu(d\omega)$ and in the last equation we use (4.4). Linear stability means that the spectrum of the operator $D\Theta_\omega(m(\cdot))$ is contained in $\{z \in \mathbb{C} : \text{Re } z < 0\}$. We shall see that only the discrete part of the spectrum is relevant for the stability issue.

Fact 3 *(a) The discrete part of the spectrum consists of a single $\lambda \in \mathbb{R}$ given by the relation*

$$\beta \int_{\mathbb{R}} \frac{1}{\cosh[\beta(m+\omega)](\cosh[\beta(m+\omega)] + \frac{1}{2}\lambda)} \mu(d\omega) = 1, \tag{4.19}$$

which satisfies $\lambda < 0$ iff $\Gamma'_\beta(m) < 1$ (recall (4.6)).
(b) If $\Gamma'_\beta(m) < 1$, then the continuous part of the spectrum is contained in $\{\lambda \in \mathbb{C} : \text{Re } \lambda < 0\}$.

Proof. (a) Elementary. The relation in (4.19) corresponds to the case $n \neq 0$. The case $n = 0$ requires that $n(\omega)(2\cosh[\beta(m+\omega)] + \lambda) = 0$ μ-a.s. This can only occur when $m = 0$ and μ is of the type in Case II. But then $\lambda = -2\cosh(\beta\eta) < 0$. The imaginary part of the integrand in (4.19) has the opposite sign as Im λ. This implies that $\lambda \in \mathbb{R}$. The value of λ is unique because the integrand in (4.19) is strictly decreasing in λ.
(b) Elementary. Check that if Re $\lambda \geq 0$, then $D\Theta_\omega(m(\cdot)) - \lambda I$ (I =identity) is invertible. ∎

From Fact 3 we conclude:

<u>Case I</u>. The paramagnetic solution is linearly stable ($\lambda < 0$) when it is unique and not critical, neutrally stable ($\lambda = 0$) when it is critical, and unstable ($\lambda > 0$) when it is not unique. The ferromagnetic solution is linearly stable iff $\Gamma'_\beta(m^*(\beta)) < 1$, which clearly is true whenever it exists, because of (4.5) and (4.6).

<u>Case II</u>. The paramagnetic solution is linearly stable in phases 0 and 2, unstable in phase 1, and neutrally stable on the boundaries. In phase 2 a stable paramagnetic and a stable *subcritical* ferromagnetic solution coexist (together with an unstable ferromagnetic solution).

4.2 Kuramoto model

The Kuramoto model with random frequencies is the system of diffusions on the unit circle driven by the Hamiltonian in (1.1) with

$$\begin{aligned} f(x;\omega,\pi) &= -K\cos x \\ g(x;\omega) &= -x\omega \qquad (x \in [0,2\pi); \omega,\pi \in \mathbb{R}) \end{aligned} \tag{4.20}$$

where $K \in (0,\infty)$ is the *coupling strength*. [8] With this choice, (1.1-1.2) describe a system of mean-field nonlinearly coupled oscillators, each with its own frequency and external white noise. The two terms in the Hamiltonian have opposite effects: f tends to point the oscillators in the same direction, g tends to make each oscillator rotate at its local frequency.

Let $q_t^\omega(x)$ denote the probability density that a typical oscillator has angle x at time t in the medium ω (in the McKean-Vlasov limit), normalized as

$$\int_0^{2\pi} q_t^\omega(x)dx = 1 \quad \text{for all } t,\omega. \tag{4.21}$$

Then the appropriate order parameter of the system is the complex number

$$r_t e^{i\psi_t} = \int_{\mathbb{R}} \mu(d\omega) \int_0^{2\pi} dx\ e^{ix} q_t^\omega(x). \tag{4.22}$$

Here $r_t \geq 0$ measures the *phase coherence* and $\psi_t \in [0,2\pi)$ measures the *average phase* of the oscillators. In terms of these quantities the McKean-Vlasov equation (1.20-1.21) reads

$$\frac{\partial}{\partial t} q_t^\omega = \frac{1}{2}\frac{\partial^2}{\partial x^2} q_t^\omega - \frac{\partial}{\partial x}[\beta^{\omega,q_t} q_t^\omega] \tag{4.23}$$

with β^{ω,q_t} the drift given by (1.23)

$$\beta^{\omega,q_t}(x) = K r_t \sin(\psi_t - x) + \omega. \tag{4.24}$$

The stationary solutions of (4.23) and (4.24) and their stability properties have been investigated by Strogatz and Mirollo (1990,1991) and Bonilla, Neu and Spigler (1992). We summarize the results here.

1. Stationary solution(s). Abbreviate

$$B^{\omega,q}(x) = 2\int^x \beta^{\omega,q}(y)dy = 2Kr\cos(\psi - x) + 2\omega x. \tag{4.25}$$

Any stationary solution of (4.23) is of the form

$$q^\omega(x) = \frac{1}{Z^\omega} e^{B^{\omega,q}(x)} \int_0^{2\pi} dy\ e^{-B^{\omega,q}(x+y)}, \tag{4.26}$$

where Z^ω is the normalizing constant (see (4.21)) and (r,ψ) must satisfy the consistency relation (see (4.22))

$$r\cos\psi = \int_{\mathbb{R}} \mu(d\omega) \frac{\int_0^{2\pi} dx \int_0^{2\pi} dy\ \cos x\ \exp[B^{\omega,q}(x) - B^{\omega,q}(x+y)]}{\int_0^{2\pi} dx \int_0^{2\pi} dy\ \exp[B^{\omega,q}(x) - B^{\omega,q}(x+y)]}. \tag{4.27}$$

Solutions with $r = 0$ are called *incoherent*, those with $r > 0$ *(partially) synchronized*. It follows from (4.25-4.26) that the only solution with $r = 0$ is the uniform solution

$$q^\omega(x) \equiv \frac{1}{2\pi} \tag{4.28}$$

[8] The state variable x, which was originally $\mathbb{R}$-valued, is wrapped around the unit circle. See footnote 1.

and that this solution exists for all choices of K and μ.

Case I. Define

$$K_c = \left[\int_{\mathbb{R}} \frac{\phi(\omega)}{1+4\omega^2} d\omega\right]^{-1}. \tag{4.29}$$

Then the incoherent solution is the only solution when $K < K_c$, while a synchronized solution bifurcates off as K increases through K_c. Here the critical value K_c comes from the fact that for small r the r.h.s. of (4.27) behaves like $\sim Kr/K_c$ (pick $\psi = 0$).

Case II. The phase diagram is drawn in Figure 2. There are three phases, numbered 0, 1 resp. 2, counting the number of synchronized solutions. The lower curve is

$$K \to \eta(K) = \frac{1}{2}\sqrt{K-1} \quad (K \in [1, K_1]) \tag{4.30}$$

and terminates at the point $(K_1, \eta_1) = (2, \frac{1}{2})$. (Here $K_1 = 2$ turns out to be a boundary value above which non-stationary periodic solutions occur, as will be seen below.) The upper curve is obtained numerically. The two curves coincide for $K \in [1, K_2]$ with $K_2 = \frac{3}{2}$ and separate afterwards. The qualitative features of the phase diagram can be seen from the expansion for small r (and $\psi = 0$) that is obtained by inserting (4.27) into (4.26):

$$\begin{aligned} r &= Kr\left[\tfrac{1}{K_c} - \tfrac{1}{2}CK^2r^2 + O(r^4)\right], \\ K_c &= 1 + 4\eta^2, \\ C &= \tfrac{1-8\eta^2}{(1+\eta^2)(1+4\eta^2)^2}. \end{aligned} \tag{4.31}$$

We see that C changes sign as η increases through the value $\eta_2 = \eta(K_2) = 1/2\sqrt{2}$.

2. Linear stability. We consider $r = 0$ and $r > 0$ separately.

2.a. $r = 0$. The stability of the incoherent solution was studied by Strogatz and Mirollo (1990,1991). They showed that if (4.23) is abbreviated as

$$\frac{\partial}{\partial t} q_t^\omega = \Theta_\omega(q_t^\omega), \tag{4.32}$$

then $D\Theta_\omega(q_t^\omega) : L^1(\mu) \to L^1(\mu)$ has continuous spectrum

$$\left\{\lambda \in \mathbb{C} : \lambda = -\frac{1}{2} - i\omega \ (\omega \in \operatorname{supp}(\mu))\right\} \tag{4.33}$$

and discrete spectrum given by the relation

$$K \int_{\mathbb{R}} \frac{\mu(d\omega)}{1 + 2\lambda + 2i\omega} = 1. \tag{4.34}$$

Thus, the continuous spectrum does not contribute to the stability issue, which therefore all depends on (4.34).

Equations (4.33-4.34) in fact require no assumptions on μ. For μ symmetric, as was assumed, (4.34) reduces to

$$K \int_{\mathbb{R}} \frac{2\lambda + 1}{(2\lambda + 1)^2 + 4\omega^2}\, \mu(d\omega) = 1. \tag{4.35}$$

Again we distinguish between Case I and Case II, as in Section 4.1.

Case I. It can be shown that the unimodality of ϕ implies that (4.35) has at most one solution $\lambda \in \mathbb{R}$, satisfying

$$\begin{array}{ll} K \le K_c^* : & \text{no } \lambda \text{ exists} \\ K_c^* < K < K_c : & -\frac{1}{2} < \lambda < 0 \\ K = K_c : & \lambda = 0 \\ K > K_c : & \lambda > 0 \end{array} \tag{4.36}$$

with K_c given by (4.29) and

$$K_c^* = \frac{2}{\pi \phi(0)} \tag{4.37}$$

(obtained by letting $\lambda \downarrow -\frac{1}{2}$ in (4.35)). Hence the incoherent solution is linearly stable if $K < K_c$, neutrally stable if $K = K_c$, and unstable in $K > K_c$.

Case II. Now (4.35) reduces to $K(1 + 2\lambda)/[(1 + 2\lambda)^2 + 4\eta^2] = 1$, which has two solutions

$$\lambda^{\pm} = -\frac{1}{2} + \frac{K}{4} \pm \frac{1}{4}\sqrt{K^2 - 16\eta^2}. \tag{4.38}$$

Thus we find that

$$\begin{array}{ll} K \le 1 : & \text{Re } \lambda^+ < 0 \text{ for all } \eta \\ 1 < K < 2 : & \text{Re } \lambda^+ < 0 \text{ iff } K < K_c = 1 + 4\eta^2 \\ K > 2 : & \text{Re } \lambda^+ > 0 \text{ for all } \eta. \end{array} \tag{4.39}$$

Thus the incoherent solution is linearly stable when $K < K_1 \wedge K_c$, neutrally stable when $K = K_1 \wedge K_c$, and unstable when $K > K_1 \wedge K_c$.

2.b. $r > 0$. The stability of synchronized states is less well understood.

Case I. A (unique) synchronized state bifurcates off as K increases through K_c and this state is linearly stable.

Case II. The phase diagram is more complex. Bonilla, Neu and Spigler (1992) *heuristically* argue the following:

(1) $\eta \in (0, \eta_2)$: The same bifurcation occurs as in Case I, namely, as K increases through the value $K_c = 1 + 4\eta^2$ one stable synchronized state appears.
(2) $\eta \in (\eta_2, \eta_1)$: There exists $1 < K_c^* < K_c$ such that for $K \in (K_c^*, K_c)$ there is a stable *subcritical* synchronized state that coexists with the stable incoherent state (there is also an unstable synchronized state). As K increases through the value K_c the incoherent state becomes unstable and the synchronized state survives alone.

(3) $\eta \in (\eta_1, \infty)$: As K increases through the value $K_1 = 2$ the incoherent state becomes unstable and a stable *time periodic* state bifurcates off. This is a state where r_t, ψ_t are periodic in time.

Appendix A

We prove here that equation (1.17) has a unique solution. We assume (A1): the initial measure λ has a density ϕ w.r.t. Lebesgue measure satisfying $\phi \in L^1(dx) \cap L^p(dx)$ for some $p > 1$.

Step 1: A priori estimate.
We first prove that if Q_* is a solution of (1.17) then there are constants $A > 0$ and $0 \le \alpha < 1/2$ such that

$$q_t^\omega(x) \le \frac{A}{t^\alpha} \quad \text{for every } x, \omega \in \mathbb{R} \text{ and } t > 0, \tag{A.1}$$

where $q_t^\omega = \Pi_t Q_*^\omega$. To see this, observe that $Q_* = P^{Q_*}$ gives

$$\frac{dQ_*^\omega}{dW} = \frac{dP^{\omega, Q_*}}{dW}. \tag{A.2}$$

The process having law P^{ω,Q_*} is a diffusion whose drift $\beta_t^{\omega, \Pi_t Q_*}$ is the bounded derivative of a bounded function (recall (1.11-1.12)). By the usual argument involving Girsanov's formula and Itô's rule, one sees that there is a constant $B > 0$ such that the Radon-Nikodym derivative in (A.2) is bounded by B uniformly in ω. It follows that

$$q_t^\omega(x) \le B\psi_t(x) \quad (t > 0), \tag{A.3}$$

where $\psi_t = \Pi_t W$, i.e.,

$$\psi_t(x) = \frac{1}{\sqrt{2\pi t}} \int e^{-\frac{1}{2t}(x-y)^2} \phi(y) dy. \tag{A.4}$$

By Hölder's inequality we have

$$\psi_t(x) \le \frac{1}{\sqrt{2\pi t}} \Big[\int e^{-\frac{q}{2t}(x-y)^2} dy \Big]^{\frac{1}{q}} \|\phi\|_p = \frac{C}{t^{\frac{1}{2} - \frac{1}{2q}}} \|\phi\|_p \tag{A.5}$$

with $C > 0$ some constant and $1/p + 1/q = 1$. Now (A.1) follows from (A.3) and (A.5).

Step 2: Uniqueness.
Let Q and $\tilde{Q}$ be two solutions of (17), with $q_t^\omega, \tilde{q}_t^\omega$ denoting the corresponding marginals. As mentioned in footnote 3, these are both classical solutions of the McKean-Vlasov equation (1.20). Define, for $t > 0$,

$$F_t^\omega(x) = q_t^\omega(x) - \tilde{q}_t^\omega(x). \tag{A.6}$$

The following relation is easily checked (see (1.11) and (1.20-1.21)):

$$\frac{\partial F_t^\omega}{\partial t} - \frac{1}{2}\frac{\partial^2 F_t^\omega}{\partial x^2} = \frac{\partial L_t^\omega}{\partial x}, \tag{A.7}$$

where

$$\begin{aligned} L_t^\omega(x) &= F_t^\omega(x) \textstyle\int dy\, \mu(d\pi) q_t^\pi(y) (\hat{f}'(y - x; \omega, \pi) + g'(x; \omega)) \\ &\quad + \tilde{q}_t^\omega(x) \textstyle\int dy\, \mu(d\pi) F_t^\pi(x) (\hat{f}'(y - x; \omega, \pi) + g'(x; \omega)). \end{aligned} \tag{A.8}$$

Now let $G(x,t)$ be the fundamental solution of the heat equation, i.e.,

$$G(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}. \tag{A.9}$$

Then (A.7) yields

$$\begin{aligned} F_t^\omega(x) &= \int_0^t ds \int dy\, G(x-y,t-s) \frac{\partial}{\partial y} L_s^\omega(y) \\ &= -\int_0^t ds \int dy\, L_s^\omega(y) \frac{\partial G}{\partial y}(x-y,t-s), \end{aligned} \tag{A.10}$$

where the last integration by parts is justified since, for $\omega \in \mathbb{R}$ and $t > 0$, $L_t^\omega(x)$ is a bounded function of x. Now define

$$H_t^\omega = \int |F_t^\omega(x)| dx. \tag{A.11}$$

By substituting (A.1) into (A.8), one obtains the following estimate:

$$\int |L_t^\omega(x)| dx \leq \frac{A}{t^\alpha} H_t^\omega + \frac{B}{t^\alpha} \int \mu(d\pi) H_t^\mu \quad \text{for all } \omega \text{ and } t \tag{A.12}$$

with A and B suitable constants independent of t. Moreover, by direct computation one sees that there is a constant K such that

$$\int \left| \frac{\partial G}{\partial y}(x-y,t-s) \right| dx \leq \frac{K}{\sqrt{t-s}} \qquad \text{for all } \ y. \tag{A.13}$$

Putting together (A.10-A.13) and defining $H_t = \int \mu(d\omega) H_t^\omega$, we get

$$H_t \leq C \int_0^t \frac{1}{s^\alpha \sqrt{t-s}} H_s ds \quad (t \in [0,T]), \tag{A.14}$$

with C some constant independent of t. Below we shall show that (A.14) implies $H_t \equiv 0$. We complete the proof by showing how the latter implies $Q = \tilde{Q}$. Indeed, if $H_t \equiv 0$ then (A.6) and (A.11) give $q_t^\omega(x) = \tilde{q}_t^\omega(x)$ for all t and for almost every ω, x. Q and $\tilde{Q}$ being solutions of (1.17), this in turn implies that for almost every $\omega \in \mathbb{R}$ the diffusions with laws Q^ω and $\tilde{Q}^\omega$ have the same bounded and continuous drift and the same initial distribution. By standard uniqueness results for stochastic differential equations, it follows that $Q^\omega = \tilde{Q}^\omega$ ω-a.s., and so $Q = \tilde{Q}$.

Step 3: $H_t \equiv 0$.
Let us define

$$\|H\|_t = \sup_{s \in [0,t]} H_s. \tag{A.15}$$

By (A.14)

$$H_s \leq C \|H\|_t \int_0^t \frac{ds}{s^\alpha \sqrt{t-s}} \qquad \text{for all } \ s \in [0,t]. \tag{A.16}$$

Now, because $\alpha < \frac{1}{2}$ we have

$$\lim_{t \to 0} \int_0^t \frac{ds}{s^\alpha \sqrt{t-s}} = 0. \tag{A.17}$$

This, together with (A.16), implies that there exists $t' > 0$ such that $H_t = 0$ for $t \in [0, t']$. Using (A.14) again we obtain

$$H_s \leq \frac{C}{t'^{\alpha}} \int_{t'}^{t} \frac{ds}{\sqrt{t-s}}. \tag{A.18}$$

It is trivial to see that

$$\lim_{t \to t'} \int_{t'}^{t} \frac{ds}{\sqrt{t-s}} = 0, \tag{A.19}$$

and so there must exist $t'' > 0$ such that $H_t = 0$ also for $t \in [t', t' + t'']$. This argument can be repeated to show that $H_t = 0$ for $t \in [t' + t'', t' + 2t'']$ and so on. Hence $H_t \equiv 0$.

We remark that $\alpha < 1/2$ in (A.1) is a consequence of our assumption (A1) on the initial condition λ. By removing that assumption we would get $\alpha = 1/2$ and the proof would not work.

Appendix B

The proof of Theorem 4 will be completed here, i.e., we carry out the change of variable trick which provides the rigorous justification for the formal computation in Section 2.4. We first give an outline of the proof, which is based on Claims 1-4 below. The proof of these claims comes later. At the end of this Appendix we show what modifications are needed for spin-flip systems.

Let $\mathcal{M}(C[0,T] \times \mathbb{R})$ be the vector space of signed measures on $C[0,T] \times \mathbb{R}$, provided with the weak topology.

Claim 1. *There exists a Banach space $(B, \|\cdot\|)$, a continuous linear map $T : \mathcal{M}(C[0,T] \times \mathbb{R}) \to B$, and a continuous map $\Psi : B \to \mathbb{R}$ that is bounded on $T(\mathcal{M}_1(C[0,T] \times \mathbb{R}))$ and infinitely Fréchet differentiable, such that*

$$\frac{dP_N^{\underline{\omega}}}{dW^{\otimes N}}(\underline{x}_{[0,T]}) = \exp[N\Psi(T(L_N))]. \tag{B.1}$$

Moreover, $\text{Range}(T^*) \subset C_b$, *where $T^* : B^* \to (\mathcal{M}(C[0,T] \times \mathbb{R}))^*$ is the adjoint map of T.*

Next, let

$$Y_i = T(\delta_{(x^i_{[0,T]}, \omega^i)}) \quad (i = 1, \ldots, N) \tag{B.2}$$

and denote by p_N and w_N the laws of $\underline{Y} = (Y_1, \ldots, Y_N)$ induced by P_N resp. $W^{\otimes N} \otimes \mu^{\otimes N}$. Then it follows from (B.1) that

$$\frac{dp_N}{dw_N}(\underline{Y}) = \exp[N\Psi(M_N)] \tag{B.3}$$

with $M_N = N^{-1}\sum_{i=1}^{N} Y_i$.

As we shall see later, the Banach space $(B, \|\cdot\|)$ in Claim 1 satisfies the requirements of Bolthausen's theorem (see (B.11) below), which can therefore be applied to the random variables Y_i with the help of (B.3). Moreover, by the Contraction Principle, the p_N-law of

M_N satisfies the LDP with rate function $J(Y) = \inf_{T(Q)=Y} I(Q)$, which has a unique zero at $Y_* = T(Q_*)$.

To compute the covariance of the corresponding CLT, we begin by defining a probability measure p on B by putting

$$\frac{dp}{dw}(Y) = \frac{1}{Z} \exp[D\Psi(Y_*)[Y]] \tag{B.4}$$

where w is the law of $T(\delta_{(x[0,T],\omega)})$ induced by $W \otimes \mu$, D is the Frechet derivative, and Z is the normalizing constant.

Claim 2. *The measure p is the law of $T(\delta_{(x[0,T],\omega)})$ induced by Q_*.*

Next, let $p_* = p - Y_*$. For $h, k \in B^*$ define

$$\begin{array}{lcll} \gamma(h,k) & = & \int p_*(dY)h(Y)k(Y) & \\ \tilde{h} & = & \int p_*(dY)Yh(Y) & \in B. \end{array} \tag{B.5}$$

Claim 3. *Let Γ be as in (2.28). The following identities hold:*

$$\begin{array}{lcl} \gamma(h,k) & = & \Gamma(T^*h, T^*k) \ = \ D^2H(Q_*)[\widehat{T^*h}, \widehat{T^*h}] \\ \tilde{h} & = & T(\widehat{T^*h}) \\ D^2\Psi(Y_*)[\tilde{h}, \tilde{k}] & = & D^2F(Q_*)[\widehat{T^*h}, \widehat{T^*k}]. \end{array} \tag{B.6}$$

Thus, by what was shown in Section 2.4 (proof of II), we have $\gamma(h,h) - D^2\Psi(Y_*)[\tilde{h},\tilde{h}] > 0$ unless $\tilde{h} \equiv 0$. It follows from Bolthausen's theorem that, under the p_N-law as $N \to \infty$, the field

$$\left(N^{\frac{1}{2}}h(M_N - Y_*)\right)_{h\in B^*} = \left(N^{\frac{1}{2}} \int (T^*h) d(L_N - Q_*)\right)_{h \in B^*} \tag{B.7}$$

converges weakly to a Gaussian field with covariance (recall (2.38))

$$\gamma(h,k) - D^2\Psi(Y_*)[\tilde{h},\tilde{k}] = D^2I(Q_*)[\widehat{T^*h}, \widehat{T^*h}] = C(T^*h, T^*k). \tag{B.8}$$

To complete the proof of Theorem 4 it therefore suffices to show the following fact.

Claim 4. *For given $\phi_1, \ldots, \phi_n \in C_b$, $n \in \mathbb{N}$, the Banach space $(B, \|\cdot\|)$ and the map T can be constructed in such a way that $\{\phi_1, \ldots, \phi_n\} \subset$* Range$(T^*)$.

We next proceed with the proof of Claims 1-4.

Proof of Claims 1 and 4.

By redefining the functions α_i, β_i in Assumption (A2), it is clear that instead of (1.31) we may also write

$$-\hat{f}(y - x; \omega, \pi) - g(x;\omega) = \sum_{i=0}^{\infty} c_i \alpha_i(x,\omega)\beta_i(y,\pi) \tag{B.9}$$

(where $(\alpha_i, \beta_i, c_i)_{i\geq 0}$ have the properties described in Assumption (A2)). Substituting (B.9) into (1.5) we get

$$\begin{aligned}
F(Q) = & -\tfrac{1}{2}\int_0^T dt\Big[\textstyle\sum_{i,j} c_i c_j \Big(\int Q(dx_{[0,T]}, d\omega)\alpha_i'(x_t,\omega)\alpha_j'(x_t,\omega)\Big) \\
& \Big(\int Q(dy_{[0,T]}, d\pi)\beta_i(y_t,\pi)\Big)\Big(\int Q(dy_{[0,T]}, d\pi)\beta_j(y_t,\pi)\Big) \\
& + \textstyle\sum_i c_i \Big(\int Q(dx_{[0,T]}, d\omega)\alpha_i''(x_t,\omega)\Big)\Big(\int Q(dy_{[0,T]}, d\pi)\beta_i(y_t,\pi)\Big)\Big] \\
& + \tfrac{1}{2}\textstyle\sum_i c_i \Big(\int Q(dx_{[0,T]}, d\omega)\alpha_i(x_T,\omega)\Big)\Big(\int Q(dy_{[0,T]}, d\pi)\beta_i(y_T,\pi)\Big) \\
& - \tfrac{1}{2}\textstyle\sum_i c_i \Big(\int Q(dx_{[0,T]}, d\omega)\alpha_i(x_0,\omega)\Big)\Big(\int Q(dy_{[0,T]}, d\pi)\beta_i(y_0,\pi)\Big).
\end{aligned} \tag{B.10}$$

Next, denote by c the finite measure on $\mathbb{N}$ given by $c(\{i\}) = c_i$. We introduce the following Banach spaces:

$$\begin{aligned}
B_1 &= L^3(\mathbb{N}^2 \times [0,T], c^{\otimes 2} \otimes dt) \\
B_2 &= L^2(\mathbb{N} \times [0,T], c \otimes dt) \\
B_3 &= L^2(\mathbb{N} \cup \{-1,-2,\ldots,-n\}, c + \delta_{-1} + \cdots + \delta_{-n}) \\
B &= (B_1)^3 \times (B_2)^2 \times (B_3)^4.
\end{aligned} \tag{B.11}$$

The norm $\|\cdot\|$ on B will be chosen to be the supremum of the norms on the factors. An element $Y \in B$ will be written

$$Y = (Y_1^1, Y_1^2, Y_1^3, Y_2^1, Y_2^2, Y_3^1, Y_3^2, Y_3^3, Y_3^4). \tag{B.12}$$

The map $T : \mathcal{M}(C[0,T] \times \mathbb{R}) \to B$ is now defined as follows: For $i, j \in \mathbb{N}$ and $t \in [0,T]$

$$\begin{aligned}
T(Q)_1^1(i,j,t) &= \int Q(dx_{[0,T]}, d\omega)\alpha_i'(x_t,\omega)\alpha_j'(x_t,\omega) \\
T(Q)_1^2(i,j,t) &= \int Q(dx_{[0,T]}, d\omega)\beta_i(x_t,\omega) \\
T(Q)_1^3(i,j,t) &= \int Q(dx_{[0,T]}, d\omega)\beta_j(x_t,\omega) \\
T(Q)_2^1(i,t) &= \int Q(dx_{[0,T]}, d\omega)\alpha_i''(x_t,\omega) \\
T(Q)_2^2(i,t) &= \int Q(dx_{[0,T]}, d\omega)\beta_i(x_t,\omega),
\end{aligned} \tag{B.13}$$

for $i \in \mathbb{N}$

$$\begin{aligned}
T(Q)_3^1(i) &= \int Q(dx_{[0,T]}, d\omega)\alpha_i(x_T,\omega) \\
T(Q)_3^2(i) &= \int Q(dx_{[0,T]}, d\omega)\beta_i(x_T,\omega) \\
T(Q)_3^3(i) &= \int Q(dx_{[0,T]}, d\omega)\alpha_i(x_0,\omega) \\
T(Q)_3^4(i) &= \int Q(dx_{[0,T]}, d\omega)\beta_i(x_0,\omega),
\end{aligned} \tag{B.14}$$

for $i = 1, 2, \ldots, n$ and $k = 1, 2, 3, 4$

$$T(Q)_3^k(-i) = \int Q(dx_{[0,T]}, d\omega)\phi_i(x_{[0,T]}, \omega). \tag{B.15}$$

A straightforward computation (which we omit) allows us to get an explicit (but rather long) formula for the operator $T^* : B^* = (B_1^*)^3 \times (B_2^*)^2 \times (B_3^*)^4 \to (\mathcal{M}(C[0,T] \times \mathbb{R}))^*$, from which it easily follows that $\mathrm{Range}(T^*) \subset \mathcal{C}_b$. Moreover, we see from (B.15) that

$$\phi_i = T^*(0,0,0,0,0,0,0,0,1_{\{-i\}}), \tag{B.16}$$

which proves Claim 4.

For $Y \in B$ define

$$\begin{aligned}\Psi(Y) &= -\tfrac{1}{2}\int_{\mathbb{N}^2\times[0,T]}(dc^{\otimes 2}\otimes dt)Y_1^1Y_1^2Y_1^3 \\ &\quad -\tfrac{1}{2}\int_{\mathbb{N}\times[0,T]}(dc\otimes dt)Y_2^1Y_2^2 \\ &\quad +\tfrac{1}{2}\int_{\mathbb{N}}dc(Y_3^1Y_3^2 - Y_3^3Y_3^4).\end{aligned} \tag{B.17}$$

Clearly, Ψ is continuous and infinitely Fréchet differentiable. Moreover, Ψ is bounded on $T(\mathcal{M}_1(C[0,T]\times\mathbb{R}))$ because the components of $T(Q)$ are bounded uniformly in $Q \in \mathcal{M}_1(C[0,T]\times\mathbb{R})$. Finally, (B.10), (B.13), (B.14) and (B.15) imply that $F(Q) = \Psi(T(Q))$ (note that F extends to all $\mathcal{M}(C[0,T]\times\mathbb{R})$). This proves Claim 1. [9] ∎

Proof of Claim 2. The main step in the proof is the relation

$$DF(Q_*)[\delta_{(x_{[0,T]},\omega)}] = \log\frac{dQ_*}{d(W\otimes\mu)}(x_{[0,T]},\omega) \qquad \text{for } W\otimes\mu \text{ -a.s. all } (x_{[0,T]},\omega). \tag{B.18}$$

This relation is easily obtained from (1.5) by direct computation using Girsanov formula. We omit the details.

By (B.18) and the fact that T is linear and continuous, we have

$$D\Psi(Y_*)[T(\delta_{(x_{[0,T]},\omega)})] = DF(Q_*)[\delta_{(x_{[0,T]},\omega)}]. \tag{B.19}$$

Thus, for any $\rho : B \to \mathbb{R}$ measurable and bounded, (B.4) gives

$$\begin{aligned}\int p(dY)\rho(Y) &= \tfrac{1}{Z}\int w(dY)\rho(Y)\exp[D\Psi(Y_*)[Y]] \\ &= \tfrac{1}{Z}\int (W\otimes\mu)(dx_{[0,T]},d\omega)\rho(T(\delta_{(x_{[0,T]},\omega)}))\tfrac{dQ_*}{d(W\otimes\mu)}(x_{[0,T]},\omega) \\ &= \tfrac{1}{Z}\int Q_*(dx_{[0,T]},d\omega)\rho(T(\delta_{(x_{[0,T]},\omega)})).\end{aligned} \tag{B.20}$$

Letting $\rho \equiv 1$, we get $Z = 1$. ∎

Proof of Claim 3.
Using Claim 2 and the definition of adjoint operator, we have

$$\begin{aligned}\gamma(h,k) &= \int p(dY)h(Y-Y_*)k(Y-Y_*) \\ &= \int Q_*(dx_{[0,T]},d\omega)h(T(\delta_{(x_{[0,T]},\omega)} - Q_*))k(T(\delta_{(x_{[0,T]},\omega)} - Q_*)) \\ &= \int Q_*(dx_{[0,T]},d\omega)\left[T^*h(x_{[0,T]},\omega) - E^{Q_*}(T^*h)\right]\left[T^*k(x_{[0,T]},\omega) - E^{Q_*}(T^*k)\right] \\ &= \Gamma(T^*h,T^*k).\end{aligned} \tag{B.21}$$

[9] As we mentioned earlier, Bolthausen's theorem can be used with no further assumption in Banach spaces of type 2 (see Ben Arous and Brunaud (1990) for the precise definition). Now, L^p-spaces with $2 \le p < \infty$ are of type 2, and finite products of Banach spaces of type 2 are again of type 2. Thus our $(B, \|\cdot\|)$ defined in (B.11) is a Banach space of type 2.

Similarly,

$$\begin{aligned}\tilde{h} &= \int p(dY)(Y-Y_*)h(Y-Y^*)\\ &= \int Q_*(dx_{[0,T]},d\omega)T(\delta_{(x_{[0,T]},\omega)}-Q_*)h(T(\delta_{(x_{[0,T]},\omega)}-Q_*))\\ &= T\Big(\int Q_*(dx_{[0,T]},d\omega)(\delta_{(x_{[0,T]},\omega)}-Q_*)T^*h(\delta_{(x_{[0,T]},\omega)}-Q_*)\Big)\\ &= T(\widehat{T^*h}),\end{aligned} \tag{B.22}$$

where we again use the notation $(T^*h)(Q)$ for $\int(T^*h)dQ$. The third identity in (B.6) follows from the second and the fact that T is linear and continuous. ∎

We finally sketch the corresponding change of variable trick for spin-flip systems. We only show the key part of the contruction, which consists of defining a linear continuous map T from $\mathcal{M}(D[0,T]\times\mathbb{R})$ to a Banach space $(B,\|\cdot\|)$ of type 2 and a smooth function $\Psi: B\to\mathbb{R}$ such that $F=\Psi\circ T$. The rest of the proof is a simple modification of what we have done above for diffusions.

In order to avoid unnecessary complications, we shall explain the construction for the function F' defined by

$$F'(Q)=\int Q(dx_{[0,T]},d\omega)\int_0^T dt\exp[\int Q(dy_{[0,T]},d\pi)f(\omega,\pi)x_ty_t]. \tag{B.23}$$

The extension of our construction from F' to F (defined in (3.5)) is straightforward.

In the above argument for diffusions, we were able to map $\mathcal{M}(C[0,T]\times\mathbb{R})$ to a Banach space $(B,\|\cdot\|)$ that is a *finite* product of L^p-spaces with $p\geq 2$ and therefore is a Banach space of type 2. In doing so, we used the fact that the function $F(Q)$ in (1.5) is "polynomial" in Q (i.e., $F(\lambda Q)$, $\lambda\in\mathbb{R}$, is a polynomial in λ). Such a property holds neither for F in (3.5) nor for F' in (B.23). Here is where Assumption (A3) plays a crucial role. Since the function $\{-1,+1\}\times\mathbb{R}\to\mathbb{R}$ given by $(x,\omega)\mapsto x\alpha_i(\omega)$ assumes only finitely values, we can find a $q\in\mathbb{N}$ and smooth functions ϕ^i_j,ψ^i_j, $j=1,\ldots,q$, such that for all $z\in\mathbb{R}$

$$e^{\alpha_i(\omega)xz} = \sum_{j=1}^{q}\psi^i_j(x\alpha_i(\omega))\phi^i_j(z)\quad(i=1,\ldots,p). \tag{B.24}$$

Substituting (3.22) into (B.23) and using (B.24), we find

$$\begin{aligned}F'(Q) &= \int_0^T dt\int Q(dx_{[0,T]},d\omega)\textstyle\prod_{i=1}^p\prod_{j=1}^q\psi^i_j(x_t\alpha_i(\omega))\phi^i_j(\int Q(dy_{[0,T]},d\pi)y_t\beta_i(\pi))\\ &= \textstyle\sum_{j_1,\ldots,j_p=1}^q\int_0^T dt\Big\{\Big[\prod_{i=1}^p\int Q(dx_{[0,T]},d\omega)\psi^i_{j_i}(x_t\alpha_i(\omega))\Big]\\ &\qquad\Big[\textstyle\prod_{i=1}^p\phi^i_{j_i}(\int Q(dy_{[0,T]},d\pi)y_t\beta_i(\pi))\Big]\Big\}.\end{aligned} \tag{B.25}$$

Note that the arguments of the functions ϕ^i_j in (B.25) are bounded uniformly in Q. Thus it is not restrictive to assume these functions and all their derivatives to be bounded. We now define

$$B=\left(L^{p+1}\right)^{q^p}\otimes\left(L^{p+1}\right)^p. \tag{B.26}$$

The norm $\|\cdot\|$ on B is taken to be the supremum of the norms on the factors. An element $f \in B$ is written in the form

$$f = \left((f_{\underline{j}}^{(1)}), (f_i^{(2)})\right)_{\underline{j}\in\{1,\ldots,q\}^{\{1,\ldots,p\}},\, i\in\{1,\ldots,p\}}. \tag{B.27}$$

The maps $T : \mathcal{M}(D[0,T] \times \mathbb{R}) \to B$ and $\Psi : B \to \mathbb{R}$ are defined by

$$\begin{aligned} T(Q)_{\underline{j}}^{(1)}(t) &= \int Q(dx_{[0,T]}, d\omega) \prod_{i=1}^{p} \psi_{j_i}^{i}(x_t \alpha_i(\omega)) \qquad (\underline{j} \in \{1,\ldots,q\}^{\{1,\ldots,p\}}) \\ T(Q)_i^{(2)}(t) &= \int Q(dy_{[0,T]}, d\pi) y_t \beta_i(\pi) \qquad (i \in \{1,\ldots,p\}) \\ \Psi(f) &= \sum_{\underline{j}} \int_0^T dt \left[f_{\underline{j}}^{(1)}(t) \prod_{i=1}^{p} \phi_{j_i}^{i}(f_i^{(2)}(t)) \right]. \end{aligned} \tag{B.28}$$

It is easily seen that T is linear and continuous. Moreover, the smoothness of Ψ follows from the fact that the functions ϕ_j^i and their derivatives are Lipschitz continuous. Finally, it is clear that $F' = \Psi \circ T$ and that B, being a finite product of L^p-spaces, is of type 2.

References

[1] G. Ben Arous and M. Brunaud, Méthode de Laplace: étude variationnelle des fluctuations de diffusions de type "champs moyens". Stochastics and Stochastic Reports 31 (1990) pp. 79-144.

[2] E. Bolthausen, Laplace approximations for sums of independent random vectors. Probab. Th. Rel. Fields 72 (1986) pp. 305-318.

[3] L.L. Bonilla, J.C. Neu and R. Spigler, Nonlinear stability of incoherence and collective synchronization in a population of coupled oscillators. J. Stat. Phys. 67 (1992) pp. 313-330.

[4] H. Brezis, *Analyse Fonctionnelle: Theorie et Applications.* Masson, Paris, 1983.

[5] M. Brunaud, Finite Kullback information diffusion laws with fixed marginals and associated large deviations functionals. Stoch. Proc. Appl. 44 (1993) pp. 329-345.

[6] F. Comets, Nucleation for a long range magnetic model. Ann. Inst. H. Poincaré 23 (1987), pp. 135-178.

[7] F. Comets and Th. Eisele, Asymptotic dynamics, non-critical and critical fluctuations for a geometric long-range interacting model. Comm. Math. Phys. 118 (1988) pp. 531-567.

[8] D.A. Dawson, Critical dynamics and fluctuations for a mean field model of cooperative behavior. J. Stat. Phys. 31 (1982) pp. 29-83.

[9] D.A. Dawson and J. Gärtner, Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. Stochastics 20 (1987) pp. 247-308.

[10] J.-D. Deuschel and D.W. Stroock, *Large Deviations.* Academic Press, Boston, 1989.

[11] S.N. Ethier and T.G. Kurtz, *Markov Processes, Characterization and Convergence*, John Wiley & Sons, New York, 1986.

[12] H. Föllmer, Random fields and diffusion processes. In: Ecole d'Eté de Probabilités de Saint Flour XV-XVII, Lecture Notes in Math. 1362, Springer, Berlin, 1988, pp. 101-204.

[13] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*. Springer, New York, 1988.

[14] R.S. Lipster and A.N. Shiryaev, *Statistics of Random Processes*, Vol. 2. Springer, New York, 1988.

[15] S.R. Salinas and W.F. Wrezinski, On the mean-field Ising model in a random magnetic field. J. Stat. Phys. 41 (1985) 299-313.

[16] S.H. Strogatz and R.E. Mirollo, Amplitude death in an array of limit-cycle oscillators. J. Stat. Phys. 60 (1990) 245-262.

[17] S.H. Strogatz and R.E. Mirollo, Stability of incoherence in a population of coupled oscillators. J. Stat. Phys. 63 (1991) 613-635.

[18] A.-S. Sznitman, Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. J. Funct. Anal. 56 (1984) pp. 311-336.

[19] S.R.S. Varadhan, *Large Deviations and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, 1984.

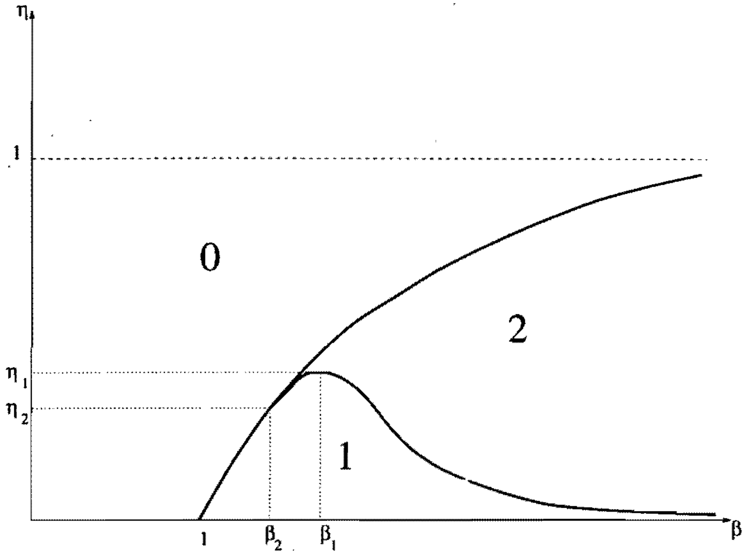

Figure 1: Phase diagram for the Curie-Weiss model

$$\beta_1 = 1.72\ldots \quad \eta_1 = 1/2\sqrt{\beta_1(\beta_1 - 1)}$$

$$\beta_2 = \tfrac{3}{2} \qquad \eta_2 = \tfrac{1}{\beta_2}\ \mathrm{arccosh}\,(\sqrt{\beta_2})$$

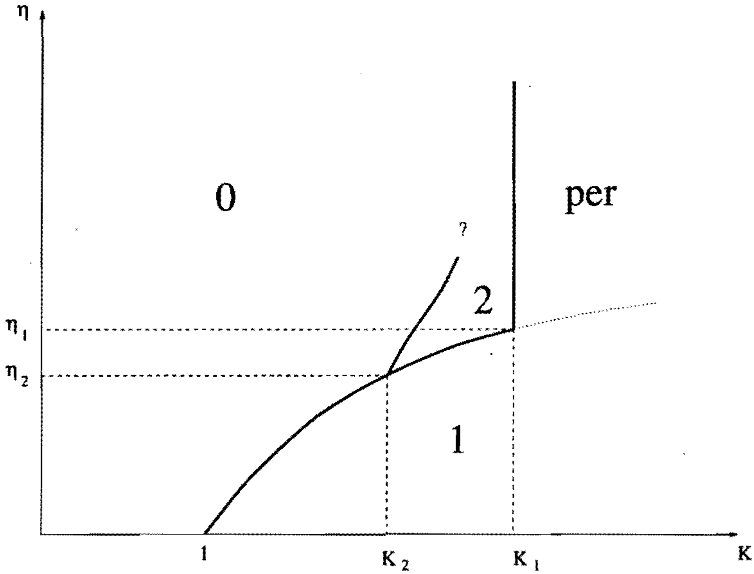

Figure 2: Phase diagram for the Kuramoto model

$$K_1 = 2 \quad \eta_1 = \frac{1}{2}$$
$$K_2 = \frac{3}{2} \quad \eta_2 = \frac{1}{2\sqrt{2}}$$