Packing Densities and Simulated Tempering for Hard Core Gibbs Point Processes

(proposed running head) Packing Densities and Simulated Tempering

August 31, 2000

S. Mase[1], J. Møller[2], D. Stoyan[3], R.P. Waagepetersen[2], G. Döge[3]

[1] *Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Oh-Okayama, 2-12-1, Meguro-Ku, Tokyo, 152-8552, Japan.*

[2] *Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7E, DK-9220 Aalborg Ø, Denmark.*

[3] *Institute of Stochastics, Bergakademie Freiberg, Bernhard-von-Cotta-Str. 2, D-09596 Freiberg, Germany.*

Abstract: Monotonicity and convergence properties of the intensity of hard core Gibbs point processes are investigated and compared to the closest packing density. For such processes simulated tempering is shown to be an efficient alternative to commonly used Markov chain Monte Carlo algorithms. Various spatial characteristics of the pure hard core process are studied based on samples obtained with the simulated tempering algorithm.

Keywords: closest packing density; hard core Gibbs point processes; intensity; Markov chain Monte Carlo; Metropolis-Hastings; phase transition; simulated tempering; spatial statistics; statistical physics; stochastic geometry.

AMS 1991 MATHEMATICS SUBJECT CLASSIFICATION: 62H11, 60K35, 62M30.

1 Introduction

This paper is concerned with hard core Gibbs point processes from the point of view of primarily spatial statistics and stochastic geometry and secondarily statistical physics. Hard core Gibbs point processes are of interest in spatial statistics and stochastic geometry as they provide models for marked point processes of discs (or balls) exhibiting a much higher degree of regularity than other types of hard core point processes such as Mátern's hard core models and 'simple sequential inhibition' processes (for definitions of these models, see Diggle (1983); Stoyan *et al.* (1995); Stoyan and Schlather, (2000)). In statistical physics, phase transition behavior of hard core Gibbs point processes has been a topic of intense study since Metropolis *et al.* (1953) investigated the classical hard-disc model in two dimensions; see, for example, Strandburg (1988), Fernández *et al.* (1995), Weber *et al.* (1995), and the references therein.

The intensity is a fundamental characteristic of a hard core Gibbs point process. It is known to be a strictly increasing function of the so-called activity parameter $z > 0$ ($-\log z$ is also called the chemical potential) but otherwise only various approximations exist for the intensity, see Stillinger *et al.* (1965), Salsburg *et al.* (1967), Hoover and Ree (1969) and Hansen and McDonald (1986). In Section 2 we consider stationary Gibbs point processes where the pair-potential is hard core and regular. As z tends to infinity, the intensity of the process is shown to attain the closest packing density ρ defined as follows. For $a > 0$, let $\Lambda(a) = [-a, a[^d$ be the d-dimensional hypercube and let $\Lambda_n = \Lambda(n + 1/2)$, $n \in \mathbb{N}_0$. The Lebesgue measure of a Borel set $\Delta \subseteq \mathbb{R}^d$ is denoted $|\Delta|$. Let $v_n = |\Lambda_n|$ and let N_n be the maximal number of mutually disjoint open unit balls which are included in Λ_n. The d-dimensional closest packing density is now defined by

$$\rho^{\max} \equiv \sup_n N_n/v_n = \lim_{n\to\infty} N_n/v_n$$

(see Lemma A.3, Appendix A).

Consider $\eta^{\max} = \rho^{\max}\omega_d$, where ω_d denotes the volume of the d-dimensional unit ball. In the terminology of stochastic geometry, $\eta^{\max}$ is the maximal area fraction ($d = 2$) or the maximal volume fraction ($d = 3$). For $d = 2$, it is known that $\eta^{\max} = \pi/(2\sqrt{3}) \approx 0.907$ is attained by close-packed hard discs whose centers form an equilateral triangular lattice in $\mathbb{R}^2$; see, e.g., Tóth (1972). For $d = 3$, according to Kepler's conjecture, $\eta^{\max} = \pi/\sqrt{18} \approx 0.740$, which corresponds to a lattice of equilateral tetrahedrons in $\mathbb{R}^3$. A proof of this conjecture is given in a series of papers by Hales (1997a, 1997b, 1998a, 1998b, 1998c, 1998d) and Ferguson and Hales (1998).

Various theoretical and simulation results suggest the existence of phase transitions for a pure hard core Gibbs point process on $\mathbb{R}^d$ (i.e. the simplest case of a hard core Gibbs point process, also called a Poisson hard core process), see Alder and Wainwright (1957, 1962) and especially Weber *et al.* (1995). For example, it is known that two physically important characteristics such as the intensity and the pressure may have finite radii of convergence when they are considered as analytic functions of z, see Ruelle (1963, Theorem 4.5.3). It is furthermore believed that there exist discontinuities for the intensity or its derivatives considered as a function of z. Another question, in the two dimensional case, is to what extent realizations of a pure hard core Gibbs point process look like the configuration of vertices of an equilateral triangular lattice.

As hard core Gibbs point processes are analytical intractable, simulations play an important role, though they have been seriously limited by computer running times. In statistical physics mostly the ordinary Metropolis *et al.* (1953) algorithm and molecular dynamics (Strandburg (1988); Allen and Tildesley (1987)) have been used with a fixed number of points/balls (within some bounded region). In Section 3 we consider both the fixed and random number of points cases and we improve the mixing properties of certain Metropolis-Hastings algorithms (Geyer and Møller (1994); Geyer (1999); Møller (1999)) by using simulated tempering (Marinari and Parisi (1992); Geyer and Thompson (1995)). Finally, in Section 4 we discuss some empirical findings for pure hard core Gibbs point processes. Though the number of points in our experiments is relatively small compared to what physicists think is appropriate (usually several thousands of points), it is certainly within the range of what is common for applications in spatial statistics and stochastic geometry — we leave more extensive simulation studies to physicists like Fernández *et al.* (1995), who report on computer work performed on many workstations over a year's time.

2 Hard Core Gibbs Processes

In this section we consider the intensity of stationary Gibbs point processes. Basic definitions are given in Section 2.1. Section 2.2 contains monotonicity and limit results for the intensity. Our main result (Proposition 2) is concerned with the particular case of hard core Gibbs processes.

2.1 Definitions

Let $\mathcal{C}$ be the space of point configurations of $\mathbb{R}^d$, i.e. the set of locally finite subsets of $\mathbb{R}^d$, and let $\mathcal{F}$ be the standard σ-algebra on $\mathcal{C}$, see e.g. Daley and Vere-Jones (1988). A probability measure $\boldsymbol{P}$ on $(\mathcal{C}, \mathcal{F})$ is said to be *n-th order* if $\int N_\Delta^n d\boldsymbol{P} < \infty$ for all bounded Borel sets $\Delta \subseteq \mathbb{R}^d$, where $N_\Delta(\omega)$ denotes the cardinality of $\Delta \cap \omega$, $\omega \in \mathcal{C}$. Let $\mathcal{P}$ be the space of all stationary first-order probability measures on $(\mathcal{C}, \mathcal{F})$. The *intensity* of a point process $P \in \mathcal{P}$ is given by

$$\rho(\boldsymbol{P}) = \frac{1}{|\Delta|} \int N_\Delta d\boldsymbol{P} \tag{2.1}$$

where Δ is any bounded Borel subset of $\mathbb{R}^d$ with $|\Delta| > 0$.

A potential function is an even measurable function $\phi : \mathbb{R}^d \mapsto (-\infty, +\infty]$. The corresponding Hamiltonian H_n on Λ_n with the free boundary condition is given by $H_n(\omega) = \frac{1}{2} \sum_{x,y \in \omega \cap \Lambda_n, x \neq y} \phi(x - y)$, $\omega \in \mathcal{C}$. The potential ϕ is *stable* if there exists a constant b such that $H_n \geq -bN_{\Lambda_n}$ for all $n \geq 0$. It is *superstable* if $\phi = \phi^s + \phi^r$, where ϕ^s is a stable potential and ϕ^r is a purely repulsive potential, that is, ϕ^r is both nonnegative and bounded away from 0 near the origin. The potential is *lower regular* if there exists a decreasing function $\psi : [0, \infty) \mapsto [0, \infty)$ such that $\phi(x) \geq -\psi(\sup_i |x_i|)$ for all $x \in \mathbb{R}^d$ and $\int_0^\infty \psi(s) s^{d-1} ds < \infty$. It is *regular* if, if in addition, there exists a number $r(\phi)$ such that $\phi(x) \leq \psi(\sup_i |x_i|)$ whenever $\sup_i |x_i| \geq r(\phi)$. Finally, it is *hard core* if there exists a number $D > 0$ such that $\phi(x) = \infty$ if $\|x\| < D$. The supremum of such D is called the *hard core distance.* A hard core potential which is lower regular is also superstable.

Define the set of *tempered* configurations by

$$\mathcal{C}^* = \{\omega \in \mathcal{C} : \sum_{\substack{s \in \mathbb{Z}^d \\ \sup_i |s_i| \leq n}} N_{\Lambda_0 + s}^2(\omega) \leq t v_n \text{ for some } t > 0 \text{ and all } n \in \mathbb{N}_0\}.$$

A probability measure is called *tempered* if it is supported by $\mathcal{C}^*$. For $n \geq 1$, let $\mathcal{C}_n$ be the space of point configurations contained in Λ_n. Given a potential ϕ, define for $z, \beta > 0$ and $\zeta \in \mathcal{C}^*$, the local Gibbs measure $\boldsymbol{P}_{n,z,\beta,\zeta}$ on $\mathcal{C}_n$ by

$$\int_{\mathcal{C}_n} g(\omega) d\boldsymbol{P}_{n,z,\beta,\zeta}(\omega) = \frac{1}{Z_{n,z,\beta,\zeta}} \sum_{m=0}^{\infty} \int_{\Lambda_n^m} \frac{z^m}{m!} g((x)_m) e^{-\beta H_{n,\zeta}((x)_m)} d(x)_m$$

where g is nonnegative and measurable, $(x)_n$ and $d(x)_n$ are short hand for $\{x_1, \ldots, x_n\}$ and $dx_1 \cdots dx_n$, respectively,

$$H_{n,\zeta}(\omega) = H_n(\omega) + \sum_{x \in \omega, y \in \zeta \setminus \Lambda_n} \phi(x - y), \ \omega \in \mathcal{C}_n,$$

is the Hamiltonian with boundary condition ζ, and

$$Z_{n,z,\beta,\zeta} = \sum_{m=0}^{\infty} \frac{z^m}{m!} \int_{\Lambda_n^m} e^{-\beta H_{n,\zeta}((x)_m)} d(x)_m$$

is the local grand partition function. A tempered probability measure $\boldsymbol{P}$ on $(\mathcal{C}, \mathcal{F})$ is *Gibbsian* if

$$\int g d\boldsymbol{P} = \int d\boldsymbol{P}(\zeta) \int_{\mathcal{C}_n} g(\omega \cup (\zeta \backslash \Lambda_n)) d\boldsymbol{P}_{n,z,\beta,\zeta}(\omega) \tag{2.2}$$

for all $n \geq 0$ and for any nonnegative measurable g. Gibbs measures with hard core potentials are tempered. Let $\mathcal{G}(z,\beta)$ denote the set of stationary, tempered, first-order Gibbs measures which satisfy (2.2). It is well-known that $\mathcal{G}(z,\beta)$ is nonempty if the potential function is both superstable and regular. If $\mathcal{G}(z,\beta)$ is not a singleton, then a phase transition is said to occur.

2.2 Results

Proposition 1 below states that the intensity $\rho(\boldsymbol{P})$ is strictly increasing in z if $\boldsymbol{P} \in \mathcal{G}(z,\beta)$. This is a well-known result in statistical physics (see e.g. Georgii, (1999)), but since the result and its proof are less known among statisticians, a proof is included in Appendix A. A part of the proof is furthermore necessary in the proof of Lemma A.1 in Appendix A.

Proposition 1. *Suppose that ϕ is superstable and regular and let $\rho(z,\beta)$ denote the intensity for $\boldsymbol{P} \in \mathcal{G}(z,\beta)$, $z,\beta > 0$ (select any if $\#\mathcal{G}(z,\beta) > 1$). For each $\beta > 0$, $\rho(z,\beta)$ is strictly increasing in $z \in]0,\infty[$.*

The following is our main theoretical result, which states that the intensity of a hard core Gibbs point process attains the closest packing density as z tends to infinity. The proof is given in Appendix A. The result itself seems well-known among physicists.

Proposition 2. *Assume that ϕ is regular and hard core with hard core distance 1, and let $\rho(z,\beta)$ be specified as in Proposition 1. Also assume the condition*

$$\int_{|x|>1} \phi_+(x)dx < \infty, \quad \phi_+ = \max\{\phi, 0\}. \tag{2.3}$$

For each $\beta > 0$,

$$\lim_{z\to\infty} \rho(z,\beta) = \rho^{\max}.$$

3 Markov chain Monte Carlo for planar pure hard core Gibbs processes

We consider now the particular case of a planar stationary pure hard core Gibbs process, i.e. $d = 2$ with $\phi(x) = \infty$ for $||x|| \le D$ and $\phi(x) = 0$ otherwise. As this process does not depend on the value of β, we have that $\rho(z, \beta) = \rho(z)$, etc. Propositions 1 and 2 establish that the intensity $\rho(z)$ is monotonically increasing towards the closest packing density. But many other questions concerning the qualitative behavior of $\rho(z)$ are still open, in particular the question concerning existence of discontinuities of $\rho(z)$ or its derivatives. Equivalently we may consider the area fraction

$$\eta(z) = \rho(z)\pi D^2/4 \tag{3.1}$$

of the system of of non-overlapping discs of diameter D and centered at the points in the process; this is a natural characteristic as $\rho(z)$ but not $\eta(z)$ depends on the value of $D > 0$. Discontinuities of $\eta(z)$, if they exist, could be associated with critical points for phase transitions. Experimental results indicate the existence of a critical point between the 'freezing point' $\eta = 0.69$ and the 'melting point' $\eta = 0.716$ (Weber *et al.* (1995); Mitus *et al.*, (1997); Truskett *et al.* (1998)). Changes from short to long range order of the spatial distributional behavior of the Gibbs process are also expected, but it is not clear whether this change happens continuously or abruptly as a function of z. In order to investigate such properties one has to resort to computer simulations of a grand canonical local pure hard core Gibbs process on a bounded region G.

In our simulations $G = [0, a]^2$ is a square and we use the periodic boundary condition. Our target density with respect to the unit rate Poisson process on G is thus proportional to

$$\pi_z(\xi) = z^{\#\xi}\mathbf{1}[\text{for all } x, y \in \xi : ||x - y|| > D \text{ if } x \ne y],\ \xi \in \mathcal{C}(G), \tag{3.2}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function and $\mathcal{C}(G)$ is the set of finite point configurations contained in G. Here and henceforth,

$$||x|| = \sqrt{\big(\min(x_1, a - x_1)\big)^2 + \big(\min(x_2, a - x_2)\big)^2},\ x = (x_1, x_2) \in G$$

denotes geodesic distance when G is wrapped on a torus. Note that if we fix the number of points in (3.2) to be $\#\xi = n$, then π_z can be considered as an unnormalized conditional density with respect to the Bernoulli process with n i.i.d. points in G, and $\pi_z \propto \mathbf{1}[\cdots]$ does not depend on the value of z.

The simulation problem becomes difficult when z increases. In order to obtain a simulation algorithm with good mixing properties, we use simulated tempering (Marinari and Parisi (1992); Geyer and Thompson (1995)); for short simulated tempering is called ST below. Our basic algorithm is the Metropolis-Hastings (MH) algorithm studied in Geyer and Møller (1994), Geyer (1999) and Møller (1999); Section 3.1 provides a short description of the MH algorithm. Note that it covers the canonical ensemble (i.e. the fixed number case) as well as the grand canonical ensemble (i.e. when the number of points fluctuates). The combination of the MH and ST algorithms is studied in Section 3.2. Section 3.3 considers some practical issues concerning the implementation of the ST algorithm. In Section 3.4 we demonstrate the improvement of ST over MH. Furthermore, the ST algorithm is shown in Appendix B to be geometrically ergodic. This ensures that Monte Carlo estimates follow central limit theorems, where the the square root of the asymptotic variance in the limiting normal distribution divided by the sample length yields the standard deviation of a Monte Carlo estimate; see e.g. Møller (1999) and the references therein.

3.1 Basic algorithm

For later purposes it is convenient to describe the MH algorithm for the case of simulation from any generalized unnormalized density g, i.e. when g is a nonnegative integrable function with respect to the distribution of the unit rate Poisson process on G.

The MH algorithm generates a Markov chain as follows. Assume that ξ with $g(\xi) > 0$ is the current state of the Markov chain. It is then proposed to either **(a)** add, **(b)** delete, or **(c)** move a point with probabilities $p_1(\xi)$, $p_2(\xi)$, and $1 - p_1(\xi) - p_2(\xi)$, respectively. The proposal ξ' for the next state in the chain is generated as follows:

(a) $\xi' = \xi \cup \{x\}$ where the new point $x \in G$ is sampled from a density $b(\xi, \cdot)$ on G;

(b) $\xi' = \xi \setminus \{x\}$ where $x \in \xi$ is chosen with probability $d(\xi, x)$ (if $\xi = \emptyset$ we set $\xi' = \xi$);

(c) $\xi' = (\xi \setminus \{x\}) \cup \{y\}$ where $x \in \xi$ is chosen with probability $d(\xi, x)$ and y is sampled from a density $m(\xi \setminus \{x\}, x, \cdot)$ (if $\xi = \emptyset$ we set $\xi' = \emptyset$).

The proposed state ξ' is accepted as the next state of the Markov chain with probability $\min\{1, r(\xi, \xi')\}$, where the Hastings ratio $r(\xi, \xi')$ depends on the type of transition and is given by

(a) $\dfrac{g(\xi')p_2(\xi')d(\xi', x)}{g(\xi)p_1(\xi)b(\xi, x)}$;

(b) $$\frac{g(\xi')p_1(\xi')b(\xi',x)}{g(\xi)p_2(\xi)d(\xi,x)} \qquad (\text{if } \xi=\emptyset \text{ then } r(\xi,\xi')=r(\emptyset,\emptyset)=1);$$

(c) $$\frac{g(\xi')\left(1-p_1(\xi')-p_2(\xi')\right)d(\xi',y)m(\xi\setminus\{y\},y,x)}{g(\xi)\left(1-p_1(\xi)-p_2(\xi)\right)d(\xi,x)m(\xi\setminus\{x\},x,y)}.$$

If ξ' is rejected, the Markov chain remains in ξ.

In our simulations, $p_1(\xi)=p_2(\xi)=p$ are constant with $0\le p\le 1/2$; the densities $d(\xi,\cdot)$ and $b(\xi,\cdot)$ are uniform on ξ and G, respectively; and the density $m(\xi\setminus\{x\},x,\cdot)$ is uniform on a square of side length $2\times\epsilon$ centered in x. Note that the Metropolis algorithm (Metropolis *et al.* (1953)) is the special case with $p=0$ and a fixed number of points.

Theoretical properties of the MH algorithm are studied in Geyer and Møller (1994), Geyer (1999) and Møller (1999). By construction the Markov chain is reversible with invariant density specified by g with respect to the unit rate Poisson process on G if $p>0$ or with respect to a Bernoulli process on G if $p=0$. In particular, if $g=\pi_z$ is the target density (3.2), the Markov chain is uniformly ergodic when $p>0$, and also when $p=0$ provided that D is sufficiently small (this is needed to ensure irreducibility), cf. Geyer (1999) and Møller (1999).

Despite the property of uniform ergodicity when $g=\pi_z$, the MH chain converges very slowly and produces highly autocorrelated samples for large values of z. As an example we applied the MH algorithm with $p=0.1$ and $\epsilon=0.3$ for simulation of the process with $g=\pi_z$, $G=[0,10]^2$, $D=1$ and $\log z=12.62$. Figure 1 shows the corresponding time series and estimated autocorrelations for the random number of points. The time series of length 25000 were obtained by subsampling each 420000th state of a chain given by 10.5×10^9 basic updates, i.e. either insert, delete or move. The Monte Carlo estimate of the intensity $\rho(z)$ based on the time series in Figure 1 is 90.03, and the Geyer (1992) initial positive sequence and initial monotone sequence estimates of the asymptotic variance are 1166.4 and 543.2, respectively. The 'automatic windowing' estimate described in Sokal (1996) gave a value of 1153.5 for the asymptotic variance. The standard error for the Monte Carlo estimate of the intensity based on the subsample of length 25000, and using the initial positive sequence estimate of the asymptotic variance, is 0.22. Better results are obtained when the MH algorithm is combined with simulated tempering as described in the following section.

Fig. 1 about here

3.2 Simulated tempering

The equilibrium distribution of our implementation of the ST algorithm is a mixture of repulsive point process models with unnormalized densities $g_1, \ldots, g_n$, $n \geq 2$, where the MH algorithm for g_i mixes well when i is small, while it produces highly autocorrelated samples when i increases towards n. Simulated tempering generates a Markov chain $(X_l, I_l)_{l\geq 0}$ whose equilibrium distribution is given by the (unnormalized) density

$$\tilde{g}(\xi, i) = g_i(\xi)\delta_i, \quad \xi \in \mathcal{C}(G),\ i = 1, \ldots n,$$

with positive δ_i. Suppose that $(X, I) \sim \tilde{g}$. The component X is a point process with marginal distribution given by the mixture $\sum_{i=1}^n g_i\delta_i$ and I is an auxiliary variable with $P(I = i) \propto \delta_i c_i$ where c_i denotes the normalizing constant (grand partition function) of g_i. The conditional density of $X|I = i$ is proportional to g_i.

Given a current state (ξ, i) of the ST chain (X_l, I_l), the two components are updated in turn as follows. First the MH update for the density $g_i(\cdot)$ is made as described in Section 3.1; suppose that ξ' is the updated state. Second it is proposed to update $i \to i'$ using a proposal kernel Q given by $Q(i, i+1) = Q(i, i-1) = 1/2$ for $1 < i < n$ and $Q(1, 2) = Q(n, n-1) = 1$. We return (ξ', i') with probability $\min\{1, r(i, i'|\xi')\}$ and retain (ξ', i) otherwise, where $r(i, i'|\xi') = \tilde{g}(\xi', i')Q(i', i)/(\tilde{g}(\xi', i)Q(i, i'))$.

By construction the Markov chain (X_l, I_l) is reversible with invariant density $\tilde{g}$; in particular, $(X_l)_{l\geq 1: I_l = n}$ has equilibrium density g_n. Geometric or uniform ergodicity of the ST chain can be established under mild conditions, see Appendix B.

In our application, for $i = 1, \ldots, n$,

$$g_i(\xi) = z_i^{\#\xi} \exp\left(-\frac{\gamma_i}{2} \sum_{\substack{x,y\in\xi \\ x\neq y}} \left[\mathbf{1}(||x-y|| \leq D) + c\frac{|b(x, D/2) \cap b(y, D/2)|}{|b(0, D/2)|} \right] \right)$$

with $0 = \gamma_1 < \gamma_2 < \cdots < \gamma_{n-1} < \gamma_n = \infty$ and $c > 0$, where $0 \times \infty = 0$. The terms $\gamma_i\mathbf{1}(||x-y|| \leq D)$ and $\gamma_i c|b(x, D/2) \cap b(y, D/2)|/|b(0, D/2)|$ both introduce a penalty whenever two discs overlap; the latter term enables us to distinguish between point patterns with the same number of overlapping pairs of discs, but where the degree of overlap differs. In particular, $g_n = \pi_{z_n}$ is the target density with $z = z_n$, while g_1 specifies the Poisson process with rate z_1. The penalizing parameter γ_i is by analogy with physics often referred to as an inverse temperature, so that the Poisson process is the 'hot' distribution and the target process is the 'cold'. For the simulations reported in this paper, the value $c = 10$ was chosen as a result of some pilot simulations.

3.3 Choice of parameter values for the ST algorithm

It is important that the simulated tempering chain does not spend an excessive proportion of time in just one temperature and if c_i was known, one could take $\delta_i = 1/c_i$ to obtain a uniform distribution over the temperatures. In practice one takes $\delta_i = 1/\hat{c}_i$ where $\hat{c}_i$ is an estimate of c_i in order to obtain an approximate uniform mixture. Estimates of c_i can up to a constant of proportionality be obtained in different ways as described in Geyer and Thompson (1995). One possibility is to use stochastic approximation and another is reverse logistic regression (Geyer (1991)) where the normalizing constants are estimated from preliminary samples obtained with Metropolis-coupled Markov chains. Our experience is that stochastic approximation is not feasible for large n, while reverse logistic regression is computationally demanding but secure.

Let (p_i, ϵ_i) denote the parameter values for each of the MH algorithms combined in the ST algorithm, $i = 1, \ldots, n$. We choose the parameter ϵ_i to be decreasing as a function of i so that reasonable acceptance rates for proposed moves are obtained for each temperature. The values of p_i are also taken to be decreasing since insert or delete proposals have low acceptance probabilities for the low temperatures.

The chain $(X_l)_{l \geq 1: I_l = n}$ yields a well-mixed sample from the target model g_n, provided that the pairs of parameter values (z_i, γ_i) and (z_{i+1}, γ_{i+1}) are chosen sufficiently close so that reasonable acceptance rates between 20% and 40% for transitions $(\xi, i) \to (\xi, i \pm 1)$ are obtained. The intensity of the Poisson process with density g_1 is chosen as $z_1 = 1/D^2$. This value corresponds to the area fraction $\eta(z) = \pi/4 = 0.785$ of a hard-disc point process with the same intensity. The remaining parameters are chosen as

$$\log z_i \quad = \quad \log z_1 + t_i \left(\log z_n - \log z_1\right)$$

and

$$\gamma_i \quad = \quad \begin{cases} t_i \gamma^* & \text{for } 1 \leq i < n \\ \infty & \text{for } i = n \end{cases}$$

with n normalized 'temperatures' $0 = t_1 < t_2 < \ldots < t_n = 1$ and a value of γ^* such that there are almost no overlapping discs in the $(n-1)$th chain $(X_l)_{l \geq 1: I_l = n-1}$. Finally, the adjustment of n and $(t_i)_{i=1,\ldots,n}$ to obtain reasonable acceptance rates for transitions $(\xi, i) \to (\xi, i \pm 1)$ are done similarly to Geyer and Thompson (1995, Section 2.3).

3.4 Comparison with MH

For comparison with Figure 1, Figure 2 shows the time series and estimated autocorrelations computed from the coldest states $(X_l)_{l\geq 1:I_l=n}$ of a ST sample with $n = 21$ temperatures and of length 10.5×10^9; the parameters for g_n are again $G = [0, 10]^2$, $D = 1$ and $\log z_n = 12.62$. The time series were obtained by subsampling the cold chain in order to obtain time series with lengths 25000 as in Figure 1; another plot (not shown here) confirms that each (I_l) is approximately uniformly distributed as desired. The Monte Carlo estimate of the intensity based on the time series in Figure 2 is 89.8, the initial positive sequence estimate of the asymptotic variance is 241.0 and the standard error of the Monte Carlo intensity estimate is hence 0.10. In this case the initial positive sequence, the initial monotone sequence, and the automatic windowing estimates gave similar results. The computing times required to obtain the time series in Figures 1 and 2 are approximately the same, but the asymptotic variance is 4.8 times bigger for MH.

We have also generated a ST chain where the basic updates are replaced by regeneration steps at the hottest temperature, that is, if the ST chain reaches a state $(X_l, 1)$, the point pattern X_l is replaced by a completely new point pattern generated from the Poisson process density g_1. Then the segments of the Markov chain sample path between regeneration times are independent, and this is useful for obtaining estimators for Monte Carlo errors as explained in Geyer and Thompson (1995). The estimated standard error based on the independent segments is 0.10, which is similar to the estimated standard error 0.09 computed using the initial positive sequence estimate.

The expected time use for a regeneration update is approximately the expected number of points for density g_1 times the time use of one of the basic updates. The computing time for obtaining the ST chain with regeneration is in the present example around five times bigger than when only basic updates are used, so the introduction of regeneration is here not advantageous in terms of computational efficiency since the standard error is not reduced.

Fig. 2 about here

4 Results obtained with simulated tempering

In the following we report on simulated results for the area fraction and a number of spatial characteristics for the planar pure hard core Gibbs process considered in the previous section. For each considered value of z we used the ST algorithm for the grand canonical ensemble (i.e. all $p_i > 0$) in the simulation study of the area fraction, while for the other characteristics we used the less computer intensive method of ST for the canonical ensemble (i.e. all $p_i = 0$).

4.1 *Area fraction*

By (3.1) and the periodic boundary condition, we estimate the area fraction by

$$\hat{\eta}(z) = \frac{\pi D^2 \overline{N}(z)}{4|G|}$$

where $\overline{N}(z)$ is the observed mean number of discs in the cold chain.

For $G = [0, 10]^2$, $D = 1$ and $\log z_n$ between 3 and 14 we obtained for $\hat{\eta}(z)$ the increasing curve shown in Figure 3. For each considered value of z, we generated a ST chain of length between 5×10^8 and 10^{10}. The calculation of $\overline{N}(z)$ is based on a sample for the cold chain of length between 2×10^7 and 10^8; here we used an appropriate burn-in (about 10% of the sample).

Fig. 3 about here

Figure 3 shows that for a wide range of z values the curve of $\hat{\eta}(z)$ nearly coincides with the curve obtained by a Padé approximation for area fraction (see Appendix C). This indicates that for values of $\log z < 9$ (corresponding to $\eta < 0.65$) both the Padé formula and our simulations yield good approximations of $\eta(z)$. But notice the change in the $\hat{\eta}(z)$ curve at values close to the freezing point $\eta = 0.69$ and the melting point point $\eta = 0.716$; in particular, this indicates a jump in the curve at the melting point.

Fig. 4 about here

4.2 *Other spatial characteristics*

For the other spatial characteristics considered in the sequel we used canonical simulations with $G = [0, 20]^2$, $D = 1$ and determined the number N of points for every value of area fraction $\eta = 0.65$, 0.67, 0.69, 0.696, 0.701, 0.707, 0.71, 0.715, 0.721, 0.735, so that $\eta = \pi D^2 N/(4|G|)$ in accordance with (3.1). Hence N is ranging from 331 to 374. For the estimation of each considered statistic (pair correlation function, hexagonality number, and so on) we used subsamples of 100 point patterns of the cold chain obtained from long runs of the ST chain. A spacing of at least $10Nn$ between the subsampled point patterns was used due to high autocorrelation in the coldest chain.

4.3 *Pair correlation function*

The pair correlation function $\mathrm{pcf}(r)$ is a well-known characteristic for point processes, see e.g. Stoyan and Stoyan (1994), Stoyan *et al.* (1995) and Truskett *et al.* (1998). In $\mathbb{R}^2$, assuming stationarity and isotropy of the hard core Gibbs point process, $\rho(z)^2\mathrm{pcf}(r)(\pi\delta^2)^2$ can be interpreted as the probability of observing a point in each of two infinitesimally small discs of radius δ and with arbitrary but fixed centers located in distance $r > 0$ from each other. Under the target model (3.2), when the number of points is fixed and G is identified with a torus, $(N/|G|)^2\mathrm{pcf}(r)(\pi\delta^2)^2$ has a similar interpretation (using the periodic boundary condition when calculating inter-point distances).

The pair correlation functions were estimated by non-parametric kernel methods as described in Stoyan and Stoyan (1994) apart from a few modifications: the intensity ρ was replaced with $N/|G|$ (since the number of points is fixed). Furthermore, because of the high number of points per sample and since we averaged over 100 samples, we used a very small band width in the kernel (of value 0.03, compare with the remarks in Stoyan and Stoyan (1994), page 285). Reducing the band width reduces the bias in the estimator and by the averaging we still obtain a smooth curve. Furthermore, because of averaging, the variance in the estimator is substantially reduced.

Estimated pair correlation functions corresponding to $\eta = 0.65$ and $\eta = 0.735$ are shown in Figure 4. As expected, with increasing η the pair correlation function reflects more order. The peaks of the estimated pair correlation functions can be compared with the modes at $r = 1, \sqrt{3}, 2, \ldots$ for the pair correlation function of the limiting regular triangular lattice pattern of hard discs with diameter $D = 1$. Clearly the curve for $\eta = 0.735$ is in better agreement with the limiting case than the curve for $\eta = 0.65$. In particular, the second mode for the curve with $\eta = 0.735$ shows the beginning of splitting towards two modes at $r = 1$ and $r = \sqrt{3}$.

Fig. 5 about here

Figure 4 is in agreement with results in Truskett *et al.* (1998) obtained by the molecular dynamics method.

4.4 *Alignment function*

The alignment function $z_B(r)$ is a kind of third-order characteristic, which is well adapted to show if there are linear chains of points as for lattice-like point patterns (Stoyan and Stoyan

(1994)). For $r > 0$, consider any $\boldsymbol{r} \in \mathbb{R}^2$ with $\|\boldsymbol{r}\| = r$ and let B_r be a square centered at $\boldsymbol{r}/2$ and of side length αr, where one side is parallel to $\boldsymbol{r}$ and $0 < \alpha < 1$ is a user-specified parameter. In $\mathbb{R}^2$, assuming stationarity and isotropy of the hard core Gibbs point process, $\rho(z)|B_r|z_B(r)$ can be interpreted as the mean number of points in B_r under the condition that there is a point in each of the locations $\boldsymbol{o} = (0, 0)$ and $\boldsymbol{r}$. The scaling of $z_B(r)$ is so that for a stationary Poisson point process we have that $z_B \equiv 1$, while if e.g. $z_B(r) > 1$, then B_r contains on the average more points than an arbitrarily placed rectangle of the same area. Large or small values of $z_B(r)$ for suitable r may thus indicate a tendency of alignment in the point pattern. For the hard core Gibbs process and small α one may expect $z_B(2)$ to be an increasing function of η with limit $0.2165/\alpha^2$ obtained at the maximal area fraction $\eta = 0.907$.

The statistical estimation of $z_B(r)$ follows Stoyan and Stoyan (1994, page 294) except that we again replace $\rho(z)$ with $N/|G|$ and use the torus convention. After some experimentation we decided to use $\alpha = 0.1$.

Simulations show as expected that $z_B(2)$ increases with increasing η; but it is $z_B(2) = 1.83$ for $\eta = 0.735$ and this is yet far from the expected maximum value 21.65 for $\eta = 0.907$. The alignment of the point patterns is more apparent for slightly increased r, e.g. $r = 2.2$. Figure 5 shows estimates of $z_B(2.2)$ and $z_B(3)$ as functions of η. Also $z_B(2.2)$ is an increasing function of η which is steepest for values of η between the freezing and melting points. The value 14.96 of $z_B(2.2)$ for $\eta = 0.735$ is not very far from the upper bound 17.89 obtained by assuming that $\rho(z)|B_r|z_B(r) \leq 1$ (which holds as $z \to \infty$). However the curve of $z_B(3)$ decreases nearly linearly and slowly towards 0; perhaps surprisingly, this curve does not show any change at the freezing and melting points.

Fig. 6 about here

4.5 *Hexagonality number*

The idea behind any hexagonality characteristic is to look for deviations from the triangular lattice arrangement of hard discs in the plane.

A first possibility is to use Ripley's K function (Ripley (1976), Stoyan and Stoyan (1994)). In $\mathbb{R}^2$, assuming stationarity, $\rho(z)K(r)$ is the mean number of points in a disc of radius r centered at the typical point (which is not counted). It vanishes for $r < 1$ and takes the value 6 for values of r a bit larger than 1 in the case of an equilateral triangular lattice with side length 1. Thus, for the hard core Gibbs point process when r is a bit larger than 1, one should expect an abrupt change of the values of $K(r)$ for η in the phase transition region. This, however, was not observed in our simulations, where we observed a continuous and nearly linear dependence of $K(r)$ on η.

Quite different is the behavior of the 'hexagonality number' $H(r)$, the *probability* that a disc of radius r centered at the typical point contains exactly 6 other points. Figure 6 shows the estimated $H(1.3)$ as an increasing function of η. The curve is steepest when η is between the freezing and melting points.

Weber *et al.* (1995) consider another characteristic $\psi(r)$ defined as the norm of the mean of the following sum taken over all points of the hard core Gibbs point process contained in a disc of radius r centered at a typical point:

$$\sum_j e^{6i\phi_j}$$

where i denotes the imaginary unit and ϕ_j is the angle between the x-axis and the line through the typical point and the jth point contained in the disc. Figure 7, which shows the estimated $\psi(1.3)$ as a function of η, again indicates the changes in point distribution between $\eta = 0.65$ and 0.72.

$\boxed{\text{Fig. 7 about here}}$

Acknowledgement: JM and RW were supported by the European Union's network 'Statistical and Computational Methods for the Analysis of Spatial Data. ERB-FMRX-CT96-0095', by MaPhySto – Center for Mathematical Physics and Stochastics, funded by a grant from the Danish National Research Foundation – and by the Danish Informatics Network in the Agricultural Sciences, funded by a grant from the Danish Research Councils. DS was supported by a grant of Deutsche Forschungsgemeinschaft. The authors are grateful to Hans-Otto Georgii, Klaus Mecke and Salvatore Torquato for helpful comments. We owe the nice condition (2.3) to Hans-Otto Georgii. We are grateful to the referees for many helpful comments, in particular for pointing out the usefulness of Gibbs variational results.

Appendix A: Monotonicity and limit for the intensity of a hard core Gibbs point process

Let $\boldsymbol{Q}$ denote the stationary Poisson point process in $\mathbb{R}^d$ with unit intensity. For $z, \beta > 0$, $\boldsymbol{P} \in \mathcal{P}$ and a potential ϕ define

1. *energy density*:

$$\Phi(\boldsymbol{P}) = \lim_{n\to\infty} \frac{1}{v_n} \int H_n d\boldsymbol{P}.$$

   This limit always exists. It is finite or equal to $+\infty$, see Georgii (1994).

2. *negative entropy density*:

$$I(\boldsymbol{P}) = \lim_{n\to\infty} \frac{1}{v_n} I(\boldsymbol{P}_n; \boldsymbol{Q}_n), \tag{A.1}$$

   where $\boldsymbol{P}_n$ (resp. $\boldsymbol{Q}_n$) is the restriction of $\boldsymbol{P}$ (resp. $\boldsymbol{Q}$) to Λ_n, and $I(\boldsymbol{P}_n; \boldsymbol{Q}_n)$ is the *relative entropy* given by

$$I(\boldsymbol{P}_n; \boldsymbol{Q}_n) = \int \log\left(\frac{d\boldsymbol{P}_n}{d\boldsymbol{Q}_n}\right) d\boldsymbol{P}_n$$

   if $\boldsymbol{P}_n \ll \boldsymbol{Q}_n$ and defined to be $+\infty$ otherwise. The negative entropy density always exists and belongs to $(-\infty, +\infty]$, see Georgii (1994).

3. *pressure*:

$$p(z,\beta) = -\min_{\boldsymbol{P}\in\mathcal{P}} \left[I(\boldsymbol{P}) + \beta\Phi(\boldsymbol{P}) - \rho(\boldsymbol{P})\log z\right], \quad z, \beta > 0, \tag{A.2}$$

   where $I(\boldsymbol{P}) + \beta\Phi(\boldsymbol{P}) - \rho(\boldsymbol{P})\log z$ is the free energy density. The pressure exists if ϕ is regular and superstable, and the function $f(a) \equiv p(e^a, \beta)$ is convex on $(-\infty, +\infty)$ (Georgii (1995)). In the literature, the pressure is sometimes alternatively defined as $\beta^{-1}p(z,\beta)$.

4. *excess free energy density*:

$$I_{z,\beta}(\boldsymbol{P}) = I(\boldsymbol{P}) + \beta\Phi(\boldsymbol{P}) - \rho(\boldsymbol{P})\log z + p(z,\beta). \tag{A.3}$$

   Note that $0 \leq I_{z,\beta} \leq \infty$ from (A.2).

Throughout this appendix it is assumed that ϕ is regular and superstable. A crucial result in the following is the so-called *Gibbs variational principle* (Georgii (1995), Theorem 3.4),

$$I_{z,\beta}(\boldsymbol{P}) = 0 \text{ if and only if } \boldsymbol{P} \in \mathcal{G}(z,\beta) \tag{A.4}$$

that is, the excess free energy density is minimized by the Gibbs measures in $\mathcal{G}(z,\beta)$.

Proof of Proposition 1: Let $\boldsymbol{P}_{z,\beta}$ denote an arbitrary element in $\mathcal{G}(z,\beta)$. The definition (A.2) and the 'if' part of the variational principle (A.4) implies

$$\frac{f(b)-f(a)}{b-a} \leq \rho(\boldsymbol{P}_{e^b,\beta}) \leq \frac{f(c)-f(b)}{c-b} \quad \text{if } a < b < c$$

whereby

$$f'_-(b) \le \rho(\boldsymbol{P}_{e^b,\beta}) \le f'_+(b). \tag{A.5}$$

Monotonicity now follows from standard results on the left-hand derivative f'_- and the right-hand derivative f'_+ of the convex function f: (a) f'_- and f'_+ always exists and are non-decreasing, (b) $f'_-(b) = f'_+(b) = f'(b)$ except for at most a countable number of b's, and (c) $f_-(a) \le f_+(a) \le f_-(b)$ if $a < b$, see Bourbaki (1958). Thus, for each $\beta > 0$, $\rho(\boldsymbol{P}_{z,\beta})$ is increasing in z. To verify strict monotonicity assume that $\rho(\boldsymbol{P}_{z,\beta}) = \rho(\boldsymbol{P}_{z_0,\beta})$ for all $z \in]z_0, z_1[$, where $z_0 < z_1$. This implies $I_{z,\beta}(\boldsymbol{P}_{z_0,\beta}) = 0$ for all $z \in]z_0, z_1[$, so $\boldsymbol{P}_{z_0,\beta} \in \mathcal{G}(z,\beta)$ by the 'only if' part of the characterization (A.4). On the other hand, by definition $\boldsymbol{P}_{z_0,\beta} \in \mathcal{G}(z_0,\beta)$ and $\mathcal{G}(z,\beta) \cap \mathcal{G}(z_0,\beta) = \emptyset$ as $z \ne z_0$. Thereby a contradiction is obtained. □

Notice that we have in fact shown that $\rho(z,\beta) = z\partial p(z,\beta)/\partial z$. We now proceed by verifying Proposition 2, which follows by combining Lemmas A.1 and A.2 below.

Lemma A.1. *Suppose that ϕ is regular and hard core with hard core distance 1 and that for each sufficiently small $\epsilon > 0$ there exists a stationary first-order probability measure $\boldsymbol{P}_\epsilon$ such that $\rho(\boldsymbol{P}_\epsilon) \ge \rho^{\max} - \epsilon$ and both $I(\boldsymbol{P}_\epsilon)$ and $\Phi(\boldsymbol{P}_\epsilon)$ are finite. Let $\rho(z,\beta)$ be as in Proposition 1. Then*

$$\lim_{z\to\infty} \rho(z,\beta) = \rho^{\max}.$$

Proof. From the definition of the pressure (A.2) we have

$$-f(a) \le I(\boldsymbol{P}_\epsilon) + \beta\Phi(\boldsymbol{P}_\epsilon) - \rho(\boldsymbol{P}_\epsilon)a$$

which for $a > 0$ implies

$$\rho(\boldsymbol{P}_\epsilon) \le \frac{1}{a}\left[I(\boldsymbol{P}_\epsilon) + \beta\Phi(\boldsymbol{P}_\epsilon)\right] + \frac{f(a)}{a}.$$

Since $f(a)$ is convex, the limit $\lim_{a\to\infty} f(a)/a$ exists and belongs to $]-\infty,+\infty]$, and

$$\lim_{a\to\infty} f(a)/a = \lim_{a\to\infty} f'_-(a) = \lim_{a\to\infty} f'_+(a),$$

see Bourbaki (1958). Combining this with (A.5) yields $\lim_{a\to\infty} \rho(e^a) = \lim_{a\to\infty} f(a)/a \ge \rho^{\max} - \epsilon$. Trivially, $\rho(e^a) \le \rho^{\max}$ so the result follows. □

Fix a sufficiently small $\epsilon > 0$. Let $\boldsymbol{B}$ denote the d-dimensional open ball with unit diameter and center at the origin. By Lemma A.3 below, there exist numbers m, k, and a set $\xi = \{z_1, \dots, z_k\} \subset \Lambda(m)$ such that

$$k/v_m > \rho^{\max} - \epsilon, \quad z_i + \epsilon\boldsymbol{B} \subseteq \Lambda(m), \quad |z_i - z_j| > 1 + 2\epsilon \text{ if } i \ne j, \quad i,j = 1,\dots,k.$$

For $i = 1, \dots, k$ and $s \in \mathbb{Z}^d$, let Y_{is} be independent random variables uniform on $\epsilon\boldsymbol{B}$, respectively, and independent of the uniform random variable H on Λ_m. Define the stationary point process

$$\tilde{X} \equiv \{z_i + (2m+1)s + Y_{is} + H : 1 \le i \le k, s \in \mathbb{Z}^d\}$$

with distribution $\tilde{\boldsymbol{P}}$. Since $N_{\Lambda_m}(\tilde{X}) = k$, $\tilde{\boldsymbol{P}}$ is first-order and its intensity $\rho(\tilde{\boldsymbol{P}})$ is equal to $k/v_m > \rho^{\max} - \epsilon$.

Lemma A.2. *The negative entropy density* $I(\tilde{\boldsymbol{P}})$ *is finite. The energy density* $\Phi(\tilde{\boldsymbol{P}})$ *is finite if the condition* (2.3) *is satisfied.*

Proof. For a sufficiently small $\epsilon > 0$ and $n \in \mathbb{N}_0$, construct m, k, ξ, and $\tilde{X}$ as above. Let $\{F_1, \dots, F_K\}$, $K = k(2n+3)^d$, be an enumeration of $\{\epsilon\boldsymbol{B} + z_i + (2m+1)s : 1 \le i \le k, s \in \mathbb{Z}^d \cap [-n-1, n+1]^d\}$ and let $\tilde{\boldsymbol{P}}_n$ and $\boldsymbol{Q}_n$ denote the restrictions of $\tilde{\boldsymbol{P}}$ and $\boldsymbol{Q}$ to $B_n = \Lambda_{m+(2m+1)n}$. Let further $\tilde{c} = |\Lambda_m|^{-1}|\epsilon\boldsymbol{B}|^{-K}$ denote the normalizing constant of the joint density of H and Y_{is}, $1 \le i \le k$, $s \in \mathbb{Z}^d \cap [-n-1, n+1]^d$, and let $g : \mathcal{C} \cap B_n \to \mathbb{R}_+$ be a nonnegative measurable function. After some manipulations we get

$$\int g d\tilde{\boldsymbol{P}} = \int_{\Lambda_m} dh \int_{F_1 \times \cdots \times F_K} \tilde{c} g(\{y_1 + h, \dots, y_K + h\} \cap B_n) d(y)_K =$$
$$e^{-|B_n|} \sum_{r=0}^{K} \frac{1}{r!} \int_{B_n^r} \gamma((x)_r) g((x)_r) d(x)_r$$

where

$$\gamma((x)_r) = e^{|B_n|} \tilde{c} \int_{\Lambda_m} \sum_{(\sigma,\tau) \in \Xi_r} \prod_{i=1}^{r} \mathbf{1}(x_i - h \in F_{\sigma_i}) \prod_{l=1}^{K-r} |(F_{\tau_l} + h) \backslash B_n|^{K-r} dh \tag{A.6}$$

and Ξ_r denotes the set of all $(\sigma_1, \dots, \sigma_r, \tau_1, \dots, \tau_{K-r}) \in \{1, \dots, K\}^K$ with $\tau_1 < \tau_2 < \cdots < \tau_{K-r}$ and $\{\sigma_1, \dots, \sigma_r\} = \{1, \dots, K\} \setminus \{\tau_1, \dots, \tau_{K-r}\}$. In particular, $\tilde{\boldsymbol{P}}_n \ll \boldsymbol{Q}_n$.

Since $\sum_{(\tau,\sigma) \in \Xi_r} \prod_{i=1}^{r} \mathbf{1}(x_i - h \in F_{\sigma_i}) \le 1$, we obtain that

$$\gamma((x)_r) \le e^{|B_n|} |\epsilon\boldsymbol{B}|^{K-r} |\Lambda_m| \tilde{c} \le e^{|B_n|} |\epsilon\boldsymbol{B}|^{-K}.$$

Hence

$$I(\tilde{\boldsymbol{P}}_n; \boldsymbol{Q}_n) = e^{-|B_n|} \sum_{r=0}^{K} \frac{1}{r!} \int_{B_n^r} \gamma((x)_r) \log \gamma((x)_r) d(x)_r \le$$
$$e^{-|B_n|} \sum_{r=0}^{K} \frac{1}{r!} \int_{B_n^r} \gamma((x)_r) \log \left(e^{|B_n|} |\epsilon\boldsymbol{B}|^{-K}\right) d(x)_r = |B_n| + K \log |\epsilon\boldsymbol{B}|^{-1}.$$

The limit $I(\tilde{\boldsymbol{P}}) = \lim_{n\to\infty} I(\tilde{\boldsymbol{P}}_n; \boldsymbol{Q}_n)/|B_n|$ is thus finite.

For simplicity, let $N = m + (2m+1)n$ and $W_{is} = z_i + (2m+1)s + Y_{is}$. $\Phi_N(\tilde{\boldsymbol{P}})$ is seen to be dominated by

$$\frac{1}{2v_N}\boldsymbol{E} \sum_{\substack{s,t\in[-n-1,n+1]^d\cap\mathbb{Z}^d\\ 1\le i,j\le k:\,(j,t)\neq(i,s)}} \phi_+(W_{is} - W_{jt})\mathbf{1}(W_{is} + H \in \Lambda_N, W_{jt} + H \in \Lambda_N).$$

On conditional on $W_{is} = w$,

$$\boldsymbol{E} \sum_{\substack{t\in[-n-1,n+1]^d\cap\mathbb{Z}^d\\ 1\le j\le k:\,(j,t)\neq(i,s)}} \phi_+(w - W_{jt})\mathbf{1}(W_{jt} + H \in \Lambda_N) \le \frac{1}{|\epsilon \boldsymbol{B}|}\int_{|x|>1} \phi_+(x)dx,$$

since the expectation of the tj'th summand is dominated by the integral of ϕ_+ over a subset $A_{tj} \subset \{|x| > 1\}$ where the A_{tj}'s are mutually disjoint. Hence we have

$$\Phi_N(\tilde{\boldsymbol{P}}) \le \frac{K}{2v_N|\epsilon \boldsymbol{B}|}\int_{|x|>1} \phi_+(x)dx.$$

Therefore $\Phi(\tilde{\boldsymbol{P}}) = \lim_{N\to\infty} \Phi_N(\tilde{\boldsymbol{P}})$ is finite. □

Lemma A.3. *For any dimension $d \ge 1$,*

$$\sup_n N_n/v_n = \lim_n N_n/v_n.$$

Proof. For each $\epsilon > 0$, there exists an m with $N_m/v_m \ge \rho^{\max} - \epsilon$ and for $n \ge m$ there exist $a \in \mathbb{N}$ and $0 \le b \le 2m$ such that $2n+1 = a(2m+1)+b$. The hypercube Λ_n contains at least a^d mutually disjoint hypercubes which are congruent to Λ_m. Therefore,

$$\frac{N_n}{v_n} \ge \frac{a^d N_m}{v_n} \ge \left(\frac{2n+1-b}{2n+1}\right)^d (\rho^{\max} - \epsilon),$$

so that $N_n/v_n \ge \rho^{\max} - 2\epsilon$ for n large enough. Hence, $\rho^{\max} = \lim_n N_n/v_n$. □

Appendix B : Geometric ergodicity of the simulated tempering algorithm

Recall that for the i^{th} MH algorithm in the ST algorithm described in Section 3.2, we have that $p_i \in [0, 1/2]$ and $b(\xi,\cdot)$, $d(\xi,\cdot)$, $m(\xi\setminus\{x\}, x, \cdot)$ are uniform on G, ξ, $S_{x,i}$, respectively, where $G = [0,a]^2$ is wrapped on a torus and $S_{x,i}$ is a square on the torus of side length $2\times\epsilon_i$ centered in x (see Section 3.1). Under these conditions and setting $p_* = \min_i p_i$ and $p^* = \max_i p_i$ we verify below the following proposition.

Proposition B.1. *If $p_* > 0$, then the ST chain defined on the state space* $\operatorname{supp}(\tilde{g}) = \{(\xi, i) : g_i(\xi) > 0\}$ *is ergodic and satisfies a geometric drift condition which ensures geometric* *fast convergence towards the unique invariant distribution specified by $\tilde{g}$, that is the chain is geometric ergodic. If $p^* = 0$ and we have fixed the number of points to be $m < \infty$, then setting $\tilde{g}_{|m}(\xi, i) = \tilde{g}(\xi, i)\mathbf{1}[\#\xi = m]$, the ST chain restricted to* $\mathrm{supp}(\tilde{g}_{|m})$ *is uniformly ergodic with unique invariant distribution specified by $\tilde{g}_{|m}$.*

Proof. The proof is much inspired by proofs in Geyer and Møller (1994), Møller (1999) and particularly Geyer (1999, Propositions 2 and 3). For background material on the theory of Markov chains we refer to Meyn and Tweedie (1993).

Consider first the case $p_* > 0$. Notice that each of the densities g_i is locally stable in the sense that there exists a constant $K_i > 0$ so that $g_i(\xi \cup \{x\}) \leq K_i g_i(\xi)$ for all $\xi \in \mathcal{C}(G)$ and $x \in G$. Choose any number K so that $K \geq \max_i K_i$ and (for convenience later on) $K > 1$ and $K|G| \geq 1$. Then for the Hastings ratio $r(\xi, \xi'|i)$ in the i^{th} MH chain we have an upper bound for births,

$$r(\xi, \xi \cup \{x\}|i) = \frac{g_i(\xi \cup \{x\})|G|}{g_i(\xi)(\#\xi + 1)} \leq \frac{K|G|}{\#\xi + 1} \quad \text{if } (\xi, i) \in \mathrm{supp}(\tilde{g}),$$

and similarly a lower bound for deaths,

$$r(\xi \cup \{x\}, \xi|i) = \frac{g_i(\xi)(\#\xi + 1)}{g_i(\xi \cup \{x\})|G|} \geq \frac{\#\xi + 1}{K|G|} \quad \text{if } (\xi \cup \{x\}, i) \in \mathrm{supp}(\tilde{g}).$$

Since $z_{i-1} < z_i$ and $\gamma_{i-1} < \gamma_i$ we further get that

$$r(i, i-1|\xi) \geq \left(\frac{z_{i-1}}{z_i}\right)^{\#\xi} \frac{\hat{c}_i}{2\hat{c}_{i-1}} > 0,$$

so for each $k \geq 1$ we can find a $\delta_k > 0$ with $r(i, i-1|\xi) > \delta_k$ for all $(\xi, i) \in \mathrm{supp}(\tilde{g})$ with $\#\xi \leq k$ and $i > 1$.

Let ϕ be the measure on $\mathrm{supp}(\tilde{g})$ defined by $\phi(A) = \mathbf{1}[(\emptyset, 1) \in A]$ and let P^k denote the k-step transition kernel for the ST chain. Suppose $(\xi, i) \in \mathrm{supp}(\tilde{g})$ and $\phi(A) > 0$. Further, for ease of presentation, assume that $\hat{c}_1 < 2\hat{c}_2$ (this is indeed the case in our applications where $\hat{c}_i < \hat{c}_{i+1}$, $i = 1, \ldots, n-1$). Then $r(1, 2|\emptyset) = \hat{c}_1/(2\hat{c}_2) \in]0, 1[$. Choose any integer $k \geq \#\xi$ and set $\rho_k = \min(\delta_k, 1 - r(1, 2|\emptyset)$. Then

$$\begin{aligned} P^{\max(k,n)}((\xi, i), A) &\geq P^{\max(\#\xi, i-1)}((\xi, i), \{(\emptyset, 1)\}) P^{\max(k,n) - \max(\#\xi, i-1)}((\emptyset, 1), \{(\emptyset, 1)\}) \\ &\geq (p_* \rho_k / (2|G|K))^{\max(\#\xi, i-1)} (p_1(1 - r(1, 2|\emptyset))^{\max(k,n) - \max(\#\xi, i-1)} \\ &\geq (p_* \rho_k / (2|G|K))^{\max(k,n)} \phi(A) > 0. \end{aligned} \tag{B.1}$$

From this we get that the ST chain is ϕ-irreducible and hence $\tilde{g}$-irreducible with unique invariant distribution specified by $\tilde{g}$. Since aperiodicity follows from

$$P((\emptyset, 1), \{(\emptyset, 1)\}) = p_1(1 - r(1, 2|\emptyset)) > 0$$

we can also conclude that the ST chain is ergodic. From (B.1) we furthermore get that

$$C_k = \{(\xi, i) \in \mathrm{supp}(\tilde{g}) : \#\xi \leq k\}$$

is a so-called small set (see Meyn and Tweedie (1993)) for every $k \geq 1$.

We turn next to the geometric drift condition (Meyn and Tweedie (1993), Theorem 15.0.1):

$$E[V(X_1, I_1)|X_0 = \xi, I_0 = i] \leq \beta V(\xi, i) + b\mathbf{1}[(\xi, i) \in C] \tag{B.2}$$

for all states $(\xi, i) \in \mathrm{supp}(\tilde{g})$, where $\beta < 1$ and $b < \infty$ are constants, $V \geq 1$ is a measurable real function, and C is a small set.

If $\#\xi \geq K|G|$, $(\xi, i) \in \mathrm{supp}(\tilde{g})$, and we condition on that $(X_0, I_0) = (\xi, i)$, then for the MH algorithm used for generating X_1 we have that: a birth is proposed and accepted with probability at most $p_i \frac{K|G|}{\#\xi+1}$ (since $K|G| \geq 1$); a birth is proposed but rejected with probability at most p_i; a death is proposed (and hence accepted since $\#\xi \geq K|G|$) with probability at most p_i; and a move is proposed and accepted with probability at most $1 - 2p_i$. Consequently

$$E[K^{\#X_1 - \#\xi}|X_0 = \xi, I_0 = i] \leq p_i \left(\frac{K^2|G|}{\#\xi + 1} + \frac{1}{K} + 1 \right) + (1 - 2p_i).$$

If we choose $N \geq K|G|$ such that $K^2|G|/(N + 1) = \epsilon < 1 - 1/K$ (recall that $K > 1$), we see that (B.2) holds with $\beta = 1 + p^*(\epsilon + 1/K - 1)$, $V(\xi, i) = K^{\#\xi}$, $C = C_{N-1}$ and $b = K^{N+1}$.

Finally consider the case $p^* = 0$. Since on $\mathrm{supp}(\tilde{g}_{|m})$, $\inf \tilde{g}_{|m} > 0$ and $\sup \tilde{g}_{|m} < \infty$, we obtain the following lower bounds of the Hastings ratios: For any $(\xi, i) \in \mathrm{supp}(\tilde{g}_{|m})$, $i' \in \{\max(i-1, 1), \min(i+1, n)\}$, and $\xi' = (\xi \setminus x) \cup \{y\}$ with $x \in \xi$ and $y \in \cap_{i=1}^{n} S_{x,i}$, we have that

$$r(\xi, \xi'|i) \geq \delta \quad \text{and} \quad r(i, i'|\xi') \geq \delta/2 \quad \text{with} \quad \delta = \frac{\inf \tilde{g}_{|m}}{\sup \tilde{g}_{|m}} > 0\,.$$

Furthermore, $g_i > 0$ for all $i < n$. It is then not difficult to see that the state space $\mathrm{supp}(\tilde{g}_{|m})$ is a small set; this is equivalent to uniform ergodicity (Meyn and Tweedie (1993), Theorem 16.0.2).

Appendix C: The Padé approximation

Combining the following Padé approximation from Hoover and Ree (1969),

$$\frac{\beta F}{N} = \log \rho - 1 + b\rho \frac{1 - 0.28b\rho + 0.006b^2\rho^2}{1 - 0.67b\rho + 0.09b^2\rho^2}$$

(here β, F, N denote physical parameters) with the following relation from Hansen and McDonald (1986),

$$\log z = \rho \frac{\partial \beta F/N}{\partial \rho} + \frac{\beta F}{N}$$

yields

$$\log z = \log \frac{4\eta}{\pi D^2} + \frac{4\eta - 6.04\eta^2 + 3.1936\eta^3 - 0.59616\eta^4 + 0.03456\eta^5}{(1 - 1.34\eta + 0.36\eta^2)^2}.$$

References

[1] Alder, B. J. and Wainwright, T. E. (1957). Phase transition of a hard sphere system. *J. Chem. Phys.* **27**, 1208–1209.

[2] Alder, B. J. and Wainwright, T. E. (1962). Phase transition in elastic disks. *Phys. Rev.* **127**, 359–361.

[3] Allen, M. P. and Tildesley, D. J. (1987). *Computer Simulation of Liquids.* Oxford University Press, Oxford.

[4] Bourbaki, N. (1958). *Élements de Mathématique, Fonctions d'une variable réelle, (Théorie Élémentaire).* Hermann, Paris.

[5] Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* Springer, New York.

[6] Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press, London.

[7] Ferguson, S. P. and Hales, T. C. (1998). A formulation of the Kepler conjecture. Manuscript.

[8] Fernández, J. F., Alonso, J. J. and Stankiewicz, E. (1995). One-stage continuous melting transition in two dimensions. *Physical Review Letters* **75**, 3477–3480.

[9] Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.*. **7**, 473-511.

[10] Geyer, C. J. (1999). Likelihood inference for spatial point processes. In *Stochastic Geometry: Likelihood and Computations* (eds. O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout), pages 79–140. Chapman and Hall/CRC, London.

[11] Geyer, C. J. and Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.* **21**, 359–373.

[12] Geyer, C. J. and Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to pedigree analysis. *J. Am. Statist. Ass.* **90**, 909–920.

[13] Georgii, H.-O. (1994). Large deviations and the equivalence of ensembles for Gibbsian particle systems with superstable interaction, *Probab. Theory Relat. Fields*, **99**, 171-195.

[14] Georgii, H.-O. (1995). The Equivalence of Ensembles for Classical Systems of Particles, *J. Statist. Phys.* **80**, 1341–1378.

[15] Georgii, H.-O. and Häggström, O. (1996). Phase transition in continuum Potts models. *Commun. Math. Phys.* **181**, 507–528.

[16] Georgii, H.-O. (1999). Phase transition and percolation in Gibbsian particle models, manuscript.

[17] Hales, T. C. (1997a). Sphere packings I. *Discrete Comput. Geom.* **17**, 1–51.

[18] Hales, T. C. (1997b). Sphere packings II. *Discrete Comput. Geom.* **18**, 135–149.

[19] Hales, T. C. (1998a). Sphere packings III. Manuscript.

[20] Hales, T. C. (1998b). Sphere packings IV. Manuscript.

[21] Hales, T. C. (1998c). The Kepler conjecture. Manuscript.

[22] Hales, T. C. (1998d). An overview of the Kepler conjecture. Manuscript.

[23] Hansen, J.-P. and McDonald, I. R. (1986). *Theory of Simple Liquids.* Academic Press, London.

[24] Hoover, W. G. and Ree, F. H. (1969). Melting Transition and Communal Entropy for Hard Spheres. *J. Chem. Phys.* **49**, 3609–3617.

[25] Marinari, E. and Parisi, G. (1992). Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters* **19**, 451–458.

[26] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chemical Physics* **21**, 1087–1092.

[27] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer-Verlag, London.

[28] Mitus, A. C., Weber, H. and Marx, D. (1997). Local structure analysis of the hard-disk fluid near melting. *Phys. Rev.* E **55**, 6855–6859.

[29] Møller, J. (1999). Markov chain Monte Carlo and spatial point processes. In *Stochastic Geometry: Likelihood and Computations* (eds. O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout), pages 141–172. Chapman and Hall/CRC, London.

[30] Nguyen, X. X. and Zessin, H. (1979a). Integral and Differential Characterizations of the Gibbs Process. *Math. Nachr.* **88**, 105–115.

[31] Nguyen, X. X. and Zessin, H. (1979b). Ergodic Theorems for Spatial Processes. *Z. Wahrsheinlichkeitstheorie und verwandte Gebiete* **48**, 133–158.

[32] Preston, C. (1976). *Random Fields.* Lecture Notes in Math., Vol. 534, Springer, Berlin.

[33] Ripley, B. D. (1976). The second-order analysis of stationary point processes. *J. Appl. Prob.* **13**, 255–266.

[34] Ruelle, D. (1970). Superstable Interactions in Classical Statistical Mechanics. *Commun. Math. Phys.* **18**, 127–159.

[35] Ruelle, D. (1963). *Statistical Mechanics: Rigorous Results*. Benjamin, Reading, Massachusetts. (Presently published from Addison-Wesley, California.)

[36] Salsburg, Z. W., Rudd, W. G. and Stillinger, F. H. (1967). Rigid disks at high density, II. *J. Chem. Phys.* **47**, 4534–4539.

[37] Sokal, A. D. (1996). Monte Carlo methods in statistical mechanics: Foundations and New Algorithms. Lecture notes for the Cargèse Summer School on 'Functional Integration: Basics and Applications', September 1996.

[38] Stillinger, F. H., Salsburg, Z. W. and Kornegay, R. L. (1965). Rigid disks at high density. *J. Chem. Phys.* **43**, 932–943.

[39] Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and its Applications*, 2nd edition. Wiley & Sons, New York.

[40] Stoyan, D. and Schlather, M. (2000). On the random sequential absorption model with discs of different sizes. Manuscript.

[41] Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields.* Wiley & Sons, Chichester.

[42] Strandburg, K. J. (1988). Two-dimensional melting. *Rev. Mod. Phys.* **60**, 161–207.

[43] Tóth, L. F. (1972). *Lagerungen in der Ebene, auf der Kugel und im Raum.* Springer-Verlag, Heidelberg.

[44] Truskett, T. M., Torquato, S., Sastry, S., Debenetti, P. G. and Stillinger, F. H. (1998). A structural precursor to freezing in the hard-disk and hard-sphere systems. *Phys. Rev.* E **58**, 3083–3088.

[45] Weber, H., Marx, D. and Binder, K. (1995). Melting transition in two dimensions: A finite-size analysis of bond-orientational order in hard disks. *Phys. Rev.* B **51**, 14636–14651.

Figure captions

Fig. 1: Time series for the number of points and estimated autocorrelations obtained with the Metropolis-Hastings algorithm.

Fig. 2: Time series for the number of points and estimated autocorrelations obtained from the coldest chain in the simulated tempering algorithm.

Fig. 3: Estimated values of $\eta(z)$ using simulated tempering (solid line) and Padé approximation (broken line).

Fig. 4: Estimated pair correlation functions for $\eta = 0.65$ (dashed line) and $\eta = 0.735$ (solid line).

Fig. 5: The alignment functions $z_B(2.2)$ (upper curve) and $z_B(3)$ (lower curve) versus area fraction.

Fig. 6: Estimated hexagonality number $H(1.3)$ versus area fraction.

Fig. 7: Estimated hexagonality statistic $\psi(1.3)$ versus area fraction.

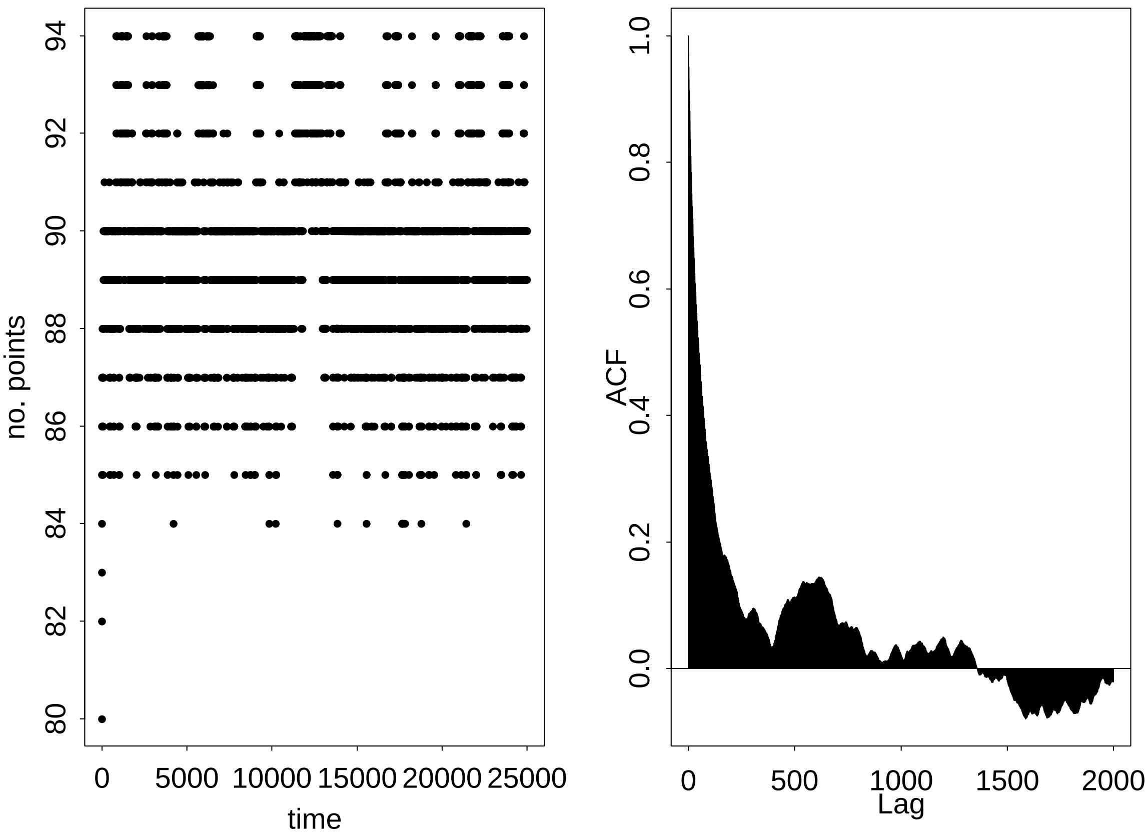

Figure 1: Time series for the number of points and estimated autocorrelations obtained with the Metropolis-Hastings algorithm.

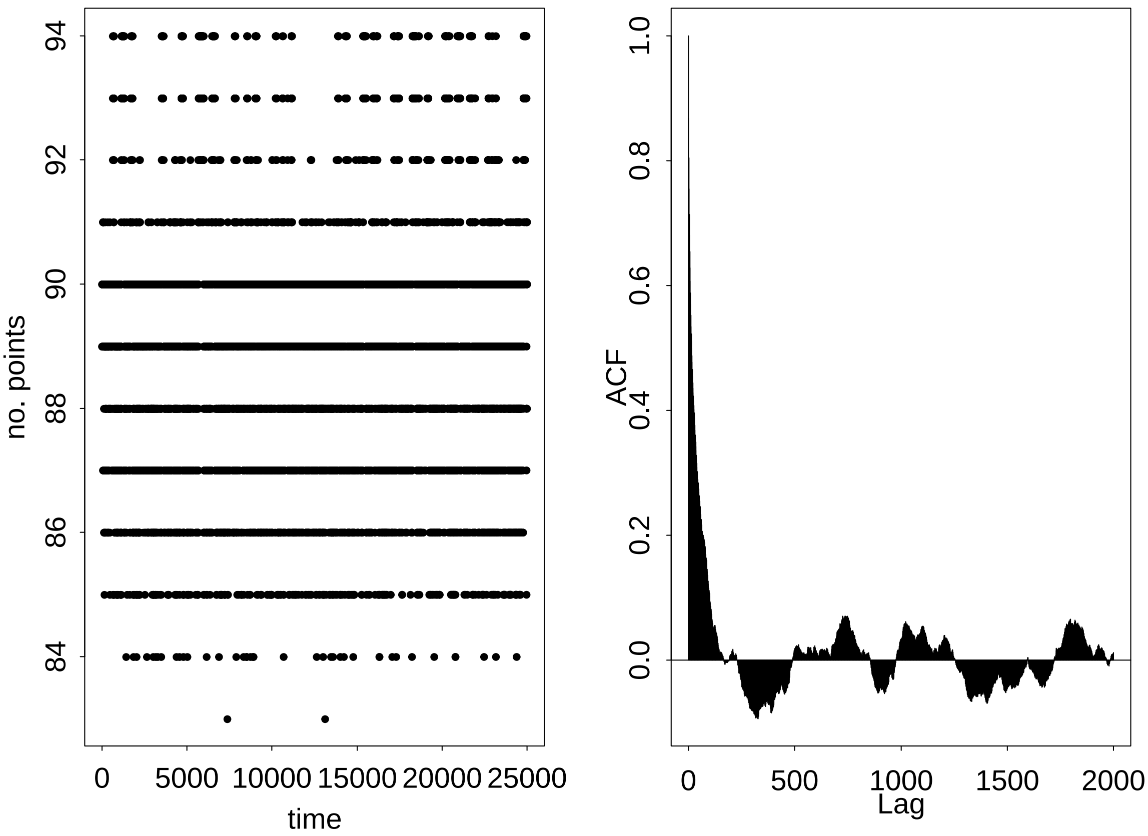

Figure 2: Time series for the number of points and estimated autocorrelations obtained from the coldest chain in the simulated tempering algorithm.

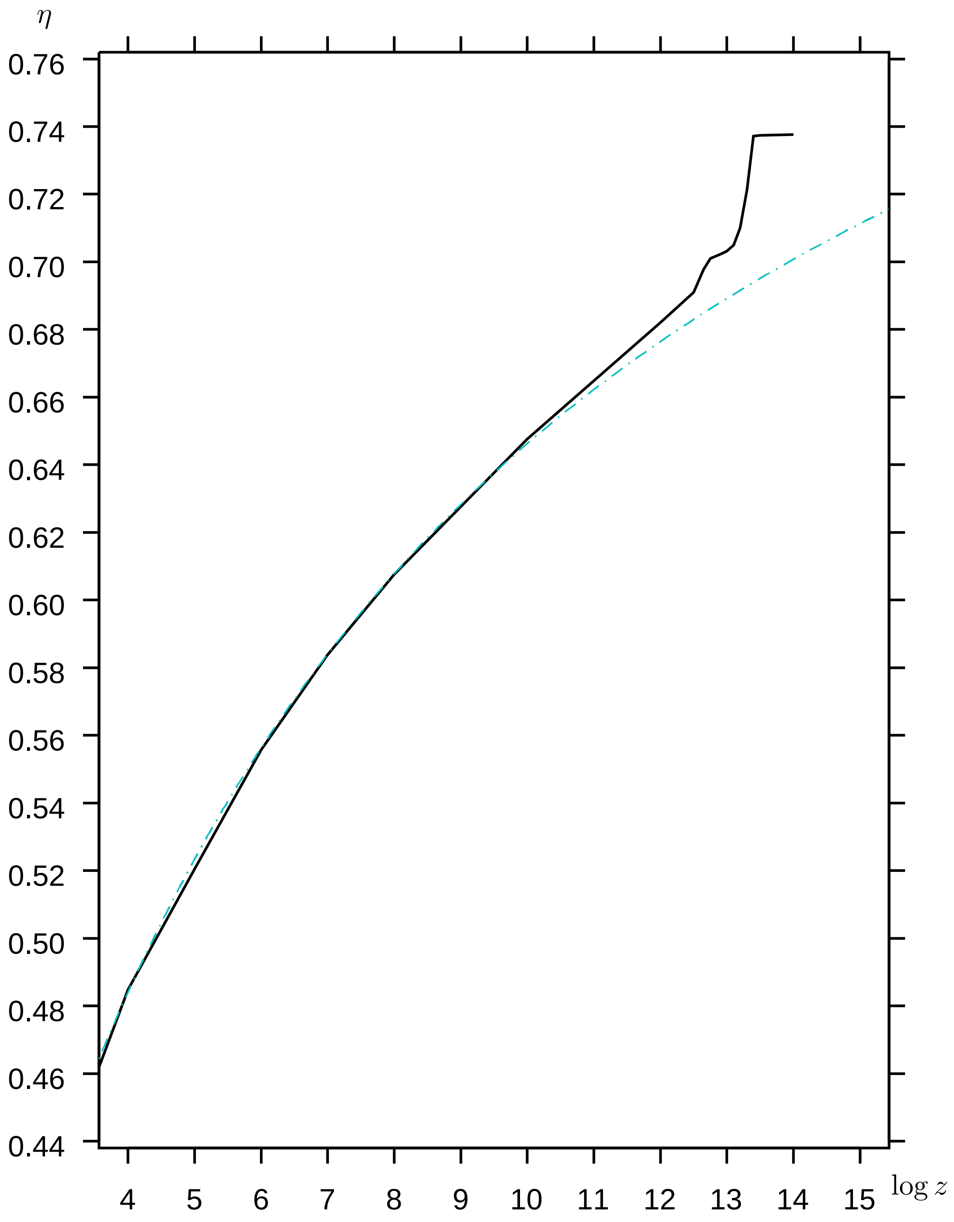

Figure 3: Estimated values of $\eta(z)$ using simulated tempering (solid line) and Padé approximation (broken line).

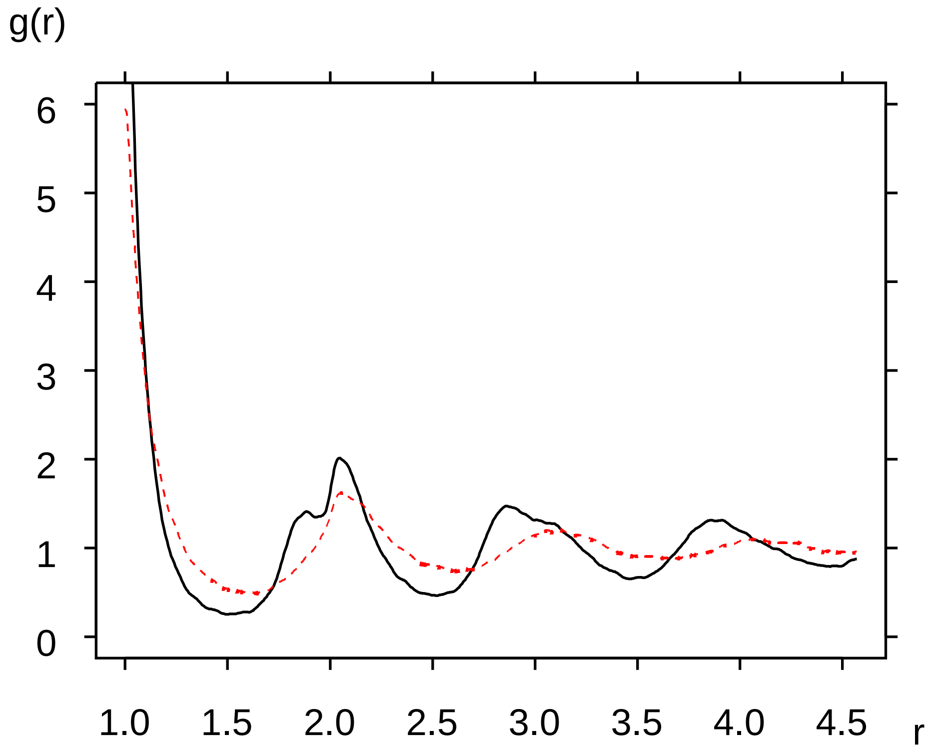

Figure 4: Estimated pair correlation functions for $\eta = 0.65$ (dashed line) and $\eta = 0.735$ (solid line).

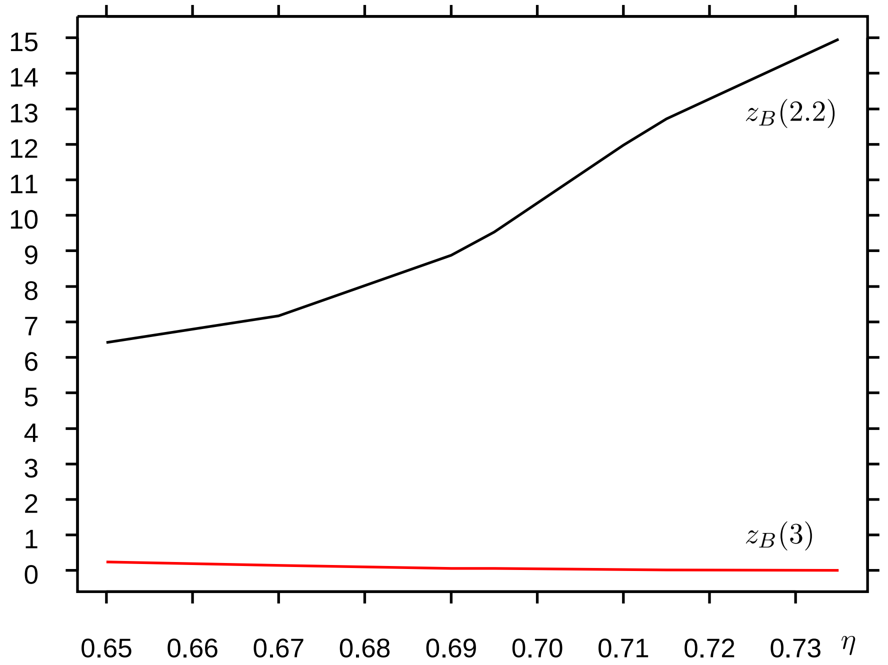

Figure 5: The alignment functions $z_B(2.2)$ (upper curve) and $z_B(3)$ (lower curve) versus area fraction.

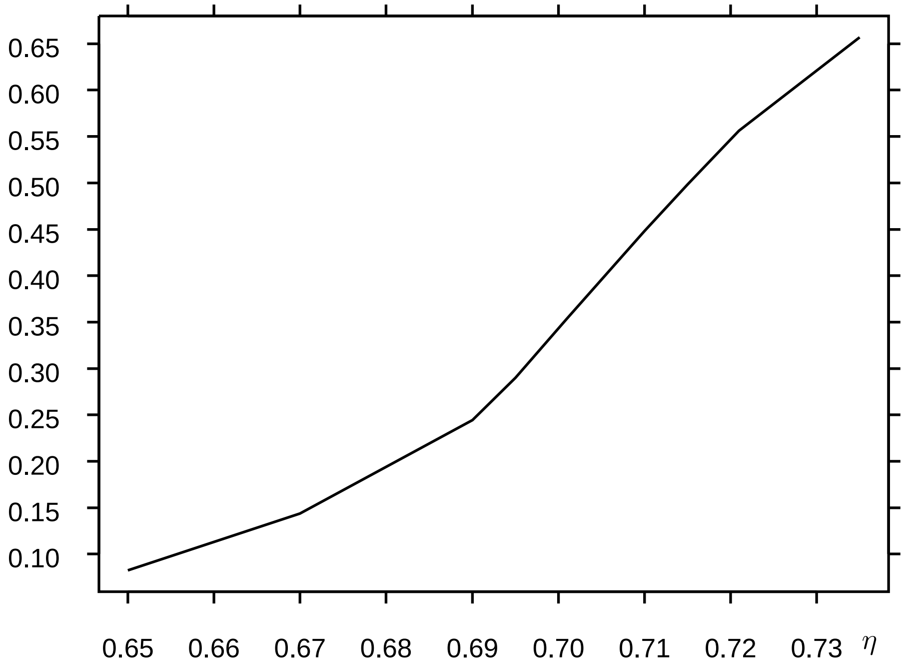

Figure 6: Estimated hexagonality number $H(1.3)$ versus area fraction.

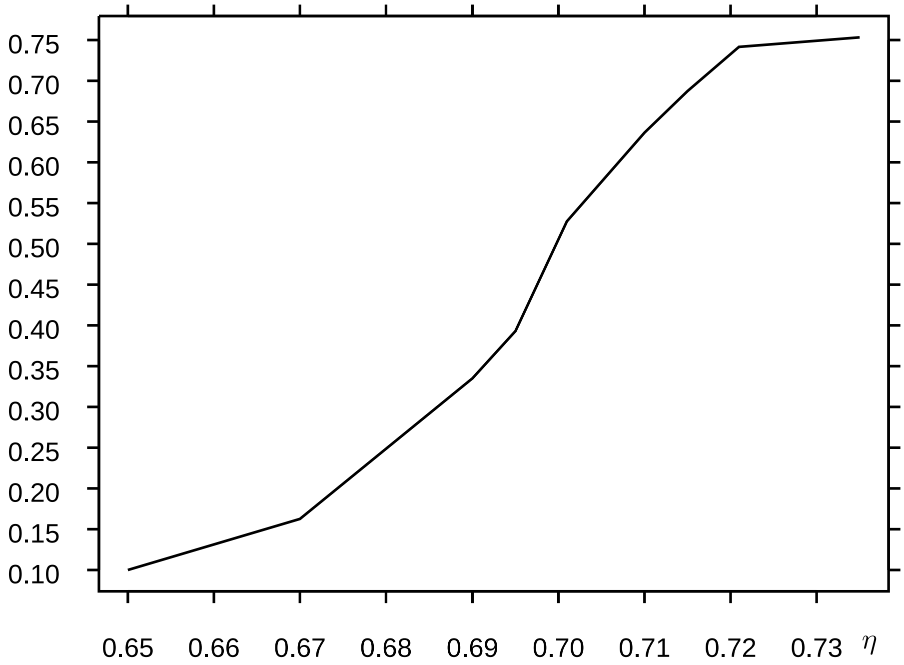

Figure 7: Estimated hexagonality statistic $\psi(1.3)$ versus area fraction.