www.maphysto.dk

ISSN 1398-599X MPS-RR 1999-12

Packing Densities and Simulated Tempering for Hard Core Gibbs Point Processes

March 19, 1999

S. Mase[1], J. Møller[2], D. Stoyan[3], R.P. Waagepetersen[4], 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.*
[4] *Department of Agricultural Systems, P.O. Box 50, Research Centre Foulum, DK-8830 Tjele, Denmark.*

Abstract: Certain monotonicity and convergence properties of the intensity of local and global 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. We study empirically the behaviour of the area fraction and various spatial characteristics of the pure hard core process using 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 standpoint 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, 1999). In statistical physics, phase transition behaviour 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.

So far only a few probabilistic properties of hard core Gibbs point processes have been established, most of them being related to questions of importance in statistical physics. A fundamental characteristic such as the intensity (that is the mean number of points per unit volume) of a hard core Gibbs point process is known until now only by various approximations in the physical literature, see Stillinger *et al.* (1965), Salsburg *et al.* (1967), Hoover and Ree (1969) and Hansen and McDonald (1986). In this paper we establish some monotonicity and convergence properties of the intensity of local and global hard core Gibbs point processes. Moreover, we discuss how such processes can be simulated in an efficient way and we present some experimental results.

In Section 2 we consider finite pairwise interaction point processes defined on a bounded region and where the pair-potential is stable, hard core, and invariant under translations. We show that the intensity of the process is an increasing function of the so-called fugacity parameter $z > 0$ ($-\log z$ is also called the chemical activity). As both z and the region tend to infinity, the intensity of the process is shown to attain the closest packing density ρ defined as follows. For $G \subset \mathbf{R}^d$, define the closest packing number M_G as the maximal number of non-overlapping open unit balls included in G. Further, let $\mathbf{B}$ be the closed unit ball with center at the origin of the d-dimensional space. Then

$$\rho \;=\; \limsup_{r\to\infty} \frac{M_{r\mathbf{B}}}{|r\mathbf{B}|}. \tag{1}$$

Clearly, $\rho \le 1/|\mathbf{B}|$, where $|G|$ denotes the volume of a Borel set G. In the terminology of stochastic geometry, if $d = 2$ then $A_A^{\max} = \rho|\mathbf{B}|$ is the maximal area fraction or maximal packing density, while if $d = 3$ then $V_V^{\max} = \rho|\mathbf{B}|$ is the maximal volume fraction. It is known that $A_A^{\max} = \pi \lim_r M_{r\mathbf{B}}/|r\mathbf{B}| = \pi/(2\sqrt{3}) \approx 0.907$ is attained by close-packed hard discs whose centers form an equilateral triangular lattice in $\mathbf{R}^2$; see, e.g., Tóth (1972). According to Kepler's conjecture, $V_V^{\max} = \pi/\sqrt{18} \approx 0.740$ which corresponds to a lattice of equilateral tetrahedrons in $\mathbf{R}^3$. 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). We can replace "lim sup" in (1) by "lim" for $d \le 3$, but to the best of our knowledge it is open problem if this also holds when $d \ge 4$.

In Section 3 we consider hard core Gibbs point processes defined on $\mathbf{R}^d$; in order to ensure that the process exists, the pair-potential is now also assumed to be of finite range. We verify that the so-called local intensity is an increasing function of z and that in a certain sense it converges to ρ as $z \to \infty$. Further, we establish a similar result for global intensities when restricting attention to stationary hard core Gibbs processes. Furthermore,

in the non-phase transition regime, we show that local intensities converge to the global intensity and that the latter is a non-decreasing function of z.

Various theoretical and simulation results seem to suggest the existence of phase transitions for a pure hard core Gibbs point process defined on $\mathbf{R}^d$ (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 have finite radii of convergence as analytic functions of z, see Ruelle (1983). Physicists believe that so-called critical values of z are scarce for pure hard core processes. We believe but are unable to prove that the global intensity is non-decreasing in the phase transition regime too; possibly there may be some irregularity at the critical values of z. Another question, in the two dimensional case, is to what extend realisations of a pure hard core Gibbs point process look like the configuration of vertices of an equilateral triangular lattice.

Due to their analytical intractability, simulation algorithms play an important role in the study of hard core Gibbs point processes, though simulations 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 constant number of points/balls (within some bounded region). In Section 4.1 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 4.2 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 Local Hard Core Gibbs Processes

In this section we study some properties of the intensity of hard core Gibbs point processes defined on a bounded region. We start by introducing some assumptions, terminology, and notation used throughout this paper.

We consider a potential Φ which is stable and has a hard core distance $R > 0$ — without loss of generality we can assume that $R = 2$, which is the minimal distance between centers of non-overlapping unit balls as considered in Section 1. Specifically, $\Phi : \mathbf{R}^d \to (-B, \infty]$ is an even measurable function with $\Phi(x) = \infty$ if $||x|| \leq R$ and $\Phi(x)$ is finite otherwise, where $B \geq 0$ is a so-called stability constant. For Borel sets $F \subseteq \mathbf{R}^d$, let $\mathcal{C}(F)$ denote the set of configurations (i.e. locally finite subsets) of F. We let $\mathcal{C}(F)$ be equipped with the usual σ-algebra (see, e.g., Daley and Vere-Jones (1988)). The restriction of a configuration ξ to a set G is denoted by $\xi_G = \xi \cap G$. The total energy of a finite configuration ξ is defined

by

$$H(\xi) = \frac{1}{2} \sum_{x,y \in \xi:\, x \neq y} \Phi(x-y)$$

setting $H(\emptyset) = 0$. The mutual interaction energy between two disjoint configurations ξ and η is given by

$$H(\xi,\, \eta) = \sum_{x \in \xi, y \in \eta} \Phi(x-y)$$

setting $H(\xi,\, \eta) = 0$ if $\xi = \emptyset$ or $\eta = \emptyset$.

Henceforth we let $G \subset \mathbf{R}^d$ denote a bounded Borel set, $z > 0$ a so-called fugacity parameter, and $\xi \in \mathcal{C}(\mathbf{R}^d)$ a feasible boundary condition, that is $||x-y|| > R$ for all distinct $x, y \in \xi$. The grand partition function corresponding to z, G, and ξ is defined by

$$Z_{z,G}(\xi) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{G^n} d(x)_n \, e^{-H((x)_n) - H((x)_n,\, \xi_{G^c})},$$

where the conventions $(x)_n = \{x_1, x_2, \ldots, x_n\}$ and $d(x)_n = dx_1 dx_2 \cdots dx_n$ are used. Since Φ is stable, the grand partition function is finite. The local Gibbs measure $\mathbf{P}_{z,G}(\cdot \,|\, \xi)$ with boundary condition ξ is the probability measure on $\mathcal{C}(G)$ defined by

$$\int f(X) \mathbf{P}_{z,G}(dX \,|\, \xi) = \frac{1}{Z_{z,G}(\xi)} \sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{G^n} d(x)_n \, f((x)_n) e^{-H((x)_n) - H((x)_n,\, \xi_{G^c})}$$

for non-negative measurable functions f. Clearly, $Z_{z,G}(\xi)$ and $\mathbf{P}_{z,G}(\cdot \,|\, \xi)$ do not depend on ξ_G. Finally, if $|G| > 0$, we define the conditional local intensity $\lambda_G(z|\, \xi)$ on G as the mean number of points per unit volume under $\mathbf{P}_{z,G}(\cdot \,|\, \xi)$,

$$\begin{aligned} \lambda_G(z|\, \xi) &= \int \frac{\#X}{|G|} \mathbf{P}_{z,G}(dX|\, \xi) \\ &= \frac{1}{|G|} \times \frac{\sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{G^n} d(x)_n \, n e^{-H((x)_n) - H((x)_n,\, \xi_{G^c})}}{\sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{G^n} d(x)_n \, e^{-H((x)_n) - H((x)_n,\, \xi_{G^c})}}, \end{aligned} \tag{2}$$

where $\#X$ denotes the cardinality of $X \in \mathcal{C}(G)$.

Remark 1. As distances between points of a sample configuration with respect to $\mathbf{P}_{z,G}(\cdot|\xi)$ cannot be less than R, the denominator and numerator of the right hand side of (2) are polynomials in z of the same order. Let

$$N_G(\xi) = \max \left\{ n : \int_{G^n} d(x)_n \, e^{-H((x)_n) - H((x)_n,\, \xi_{G^c})} > 0 \right\}$$

be the order of these polynomials. Note that $N_G(\xi)$ depends only on (R, G, ξ_{G^c}) and neither on z or the particular type of hard core process considered.

Lemma 1. *For any bounded Borel set $G \subset \mathbf{R}^d$ with $|G| > 0$, the conditional local intensity $\lambda_G(z|\,\xi)$ is increasing in z, and its supremum $\lim_{z\to\infty} \lambda_G(z|\,\xi)$ is equal to $N_G(\xi)/|G|$.*

Proof. The first property follows as the derivative of the right hand side of (2) with respect to z is equal to $1/|G|$ times the variance of $\#X$ under $\mathbf{P}_{z,G}(\cdot|\xi)$. The second property follows from Remark 1 and by taking the limits of the numerator and the denominator of the right hand side of (2) after dividing each of them by $z^{N_G(\xi)}$. □

In the rest of this section we consider a local Gibbs measure $\mathbf{P}_{z,G}(\cdot) = \mathbf{P}_{z,G}(\cdot\,|\,\emptyset)$ with the free boundary condition $\xi = \emptyset$. We set $\lambda_G(z) = \lambda_G(z, \emptyset)$.

Proposition 1. *The conditional local intensities of local hard core Gibbs processes with free boundary condition attain the closest packing density, that is, we have*

$$\limsup_{r\to\infty} \lim_{z\to\infty} \lambda_{r\mathbf{B}}(z) = \lim_{z\to\infty} \limsup_{r\to\infty} \lambda_{r\mathbf{B}}(z) = \rho\,. \tag{3}$$

Proof. Lemma 1 ensures that, for any $\delta > 0$ and $r > 0$,

$$\left|\lambda_{r\mathbf{B}}(z) - \frac{N_{r\mathbf{B}}}{|r\mathbf{B}|}\right| \;\leq\; \delta$$

for sufficiently large z. We verify first the following useful inequalities,

$$M_{(r-1)\mathbf{B}} \leq N_{r\mathbf{B}} \leq M_{r\mathbf{B}}, \qquad r > 1. \tag{4}$$

Since we can include $M_{(r-1)\mathbf{B}}$ disjoint unit balls in $(r-1)\mathbf{B}$, we can include the same number of disjoint balls with radius $r/(r-1) > 1$ in $r\mathbf{B}$. So we can include not only $M_{(r-1)\mathbf{B}}$ disjoint unit balls but also their translates by vectors h in $r\mathbf{B}$ whenever $||h|| < 1/(r-1)$. Hence the first inequality holds. The second inequality is trivial. Now, (4) gives that

$$\frac{M_{(r-1)\mathbf{B}}}{|r\mathbf{B}|} - \delta \leq \lambda_{r\mathbf{B}}(z) \leq \frac{M_{r\mathbf{B}}}{|r\mathbf{B}|} + \delta$$

for all $r \geq 1$ and for z large enough, whereby (3) follows. □

Remark 2. If "lim sup" in (1) can be replaced by "lim" as in the two-dimensional case, we can also replace "lim sup" by "lim" in Proposition 1. In fact $\lim_{z,r\to\infty} \lambda_{r\mathbf{B}}(z) = \rho$. Also note that the restriction to the case of balls, $G = r\mathbf{B}$, is by no means essential and we can take any homothetic sets $G = r\mathbf{C}$ instead if $\mathbf{C} \subset \mathbf{R^d}$ is a convex set with non-empty interior. If $\mathbf{C}$ is rectangular and we consider the corresponding torus T (i.e. use the periodic boundary condition, which is often imposed in computer simulations), we obtain the same results.

Remark 3. Arguments similar to those in the proof of Lemma 1 lead to the relation

$$\lim_{z\to\infty} \int \left(\frac{\#X}{|G|}\right)^2 \mathbf{P}_{z,G}(dX) = \left(\frac{N_G}{|G|}\right)^2 .$$

This implies

$$\lim_{z\to\infty} \mathrm{Var}_{\mathbf{P}_{z,G}} \left\{ \frac{\#X}{|G|} \right\} = 0 \,.$$

Consequently, the sample intensity of a hard core Gibbs process fluctuates very little if the activity is large enough. Further, if "lim sup" in the relation (3) can be replaced by "lim", it is straightforwardly shown that there exists an r_ϵ for every $\epsilon > 0$ such that

$$\mathbf{P}_{z,r\mathbf{B}} \left\{ \rho - \epsilon \le \frac{\#X}{|r\mathbf{B}|} \le \rho \right\} \ge 1 - \epsilon$$

for every $r \ge r_\epsilon$ if z is sufficiently large.

3 Global Hard Core Processes

In this section we consider intensities of global hard core Gibbs processes.

We let the situation be as in the previous section and consider again a potential Φ which is stable and has a hard core distance $R > 0$; as before we can without loss of generality take $R = 2$. In addition we assume that Φ is of finite range $r_0 \ge R$, i.e. $\Phi(x) = 0$ whenever $||x|| > r_0$. Then there exists a global Gibbs measure $\mathbf{P}_z$, that is a probability measure on $\mathcal{C}(\mathbf{R}^d)$, which assigns probability 1 to the feasible configurations in $\mathcal{C}(\mathbf{R}^d)$, and which satisfies the Dobrushin-Lanford-Ruelle relation

$$\int f(X)\mathbf{P}_z(dX) = \int \mathbf{P}_z(dY) \int_{\mathcal{C}(G)} f(X \cup Y_{G^c})\mathbf{P}_{z,G}(dX \,|\, Y_{G^c}) \tag{5}$$

for every bounded Borel set $G \subset \mathbf{R}^d$ and non-negative measurable function f. In the special case $r_0 = R$, i.e. when $\Phi(x) = \infty$ if $||x|| \le R$ and $\Phi(x) = 0$ otherwise, $\mathbf{P}_z$ is a so-called pure hard core Gibbs process.

The existence of $\mathbf{P}_z$ is ensured as Φ is superstable, lower regular, and $\int |1 - e^{-\Phi(x)}|dx < \infty$, see Ruelle (1970). In fact, at least one stationary Gibbs measure exists for every $z > 0$, where stationarity means that the distribution is invariant with respect to translations of configurations. But there may be more than one Gibbs measure for some z — if this is the case, we say that a phase transition occurs. This phenomena is studied fairly well for the discrete (lattice) case, but although physicists have drawn much attention to the phase transition phenomenon for the continuous state space case (see Section 4) mathematical knowledge about it is rather limited, see Georgii and Häggström (1996).

Now, for bounded Borel sets $G \subset \mathbf{R}^d$ with volume $|G| > 0$, we define the local intensity $\lambda_G^\infty(z)$ on G as the mean number of points in G per unit volume under $\mathbf{P}_z$,

$$\lambda_G^\infty(z) = \int \frac{\#(X_G)}{|G|} \, \mathbf{P}_z(dX) = \int \lambda_G(z \,|\, X) \, \mathbf{P}_z(dX) \,. \tag{6}$$

Note that, if $\mathbf{P}_z$ is stationary, $\lambda_G^\infty(z) = \lambda^\infty(z)$ is independent of G and coincides with the usual definition of the intensity for a stationary point process.

Proposition 2. *The local unconditional intensities of global hard core Gibbs processes attain the closest packing intensity, that is,*

$$\lim_{z\to\infty} \limsup_{r\to\infty} \lambda^{\infty}_{r\mathbf{B}}(z) = \limsup_{r\to\infty} \liminf_{z\to\infty} \lambda^{\infty}_{r\mathbf{B}}(z) = \limsup_{r\to\infty} \limsup_{z\to\infty} \lambda^{\infty}_{r\mathbf{B}}(z) = \rho\,. \tag{7}$$

Moreover, if "$\limsup$*" in (1) can be replaced by "*$\lim$*",*

$$\lim_{z,r\to\infty} \lambda^{\infty}_{r\mathbf{B}}(z) = \rho\,. \tag{8}$$

Finally, if we choose a stationary $\mathbf{P}_z$ *for each* z*,*

$$\lim_{z\to\infty} \lambda^{\infty}(z) = \rho. \tag{9}$$

Proof. By Lemma 1, for any $\delta > 0$, $r > r_0$, and $Y \in \mathcal{C}(\mathbf{R}^d)$, we have that

$$\left| \lambda_{r\mathbf{B}}(z\,|\,Y) - \frac{N_{r\mathbf{B}}(Y)}{|r\mathbf{B}|} \right| \le \delta \tag{10}$$

for sufficiently large z. By the finite-range assumption, $H((x)_n, Y_{(r\mathbf{B})^c}) = 0$ if $(x)_n \subset (r-r_0)\mathbf{B}$. Hence, if $r > r_0$,

$$\int_{(r\mathbf{B})^n} d(x)_n\, e^{-H((x)_n)-H((x)_n,\,Y_{(r\mathbf{B})^c})} \ge \int_{((r-r_0)\mathbf{B})^n} d(x)_n\, e^{-H((x)_n)}\,. \tag{11}$$

Since the maximal number of points with mutual distances ≥ 2 included in G is less than $|G \oplus \mathbf{B}|/|\mathbf{B}|$ ($\oplus$ stands for Minkowski addition of sets), we see that, for $r > r_0 + 1$, the left hand side in (11) is bounded by

$$\int_{(r\mathbf{B})^n} d(x)_n\, e^{(r+1)^d (r+r_0+1)^d B - H((x)_n)}$$

from above, where B is the stability constant. Therefore it can be assumed that (10) holds uniformly in $Y \in \mathcal{C}((r\mathbf{B})^c)$.

Since always

$$M_{(r-2)\mathbf{B}} \le N_{(r-1)\mathbf{B}} \le N_{r\mathbf{B}}(Y) \le N_{(r+1)\mathbf{B}} \le M_{(r+2)\mathbf{B}}\,,$$

it follows from (10) that

$$\frac{M_{(r-2)\mathbf{B}}}{|r\mathbf{B}|} - \delta \le \lambda_{r\mathbf{B}}(z|Y) \le \frac{M_{(r+2)\mathbf{B}}}{|r\mathbf{B}|} + \delta$$

for each $r > \max(2, r_0 + 1)$ if z is large enough. Taking the expectation with respect to $\mathbf{P}_z(dY)$ yields

$$\frac{M_{(r-2)\mathbf{B}}}{|r\mathbf{B}|} - \delta \le \lambda^{\infty}_{r\mathbf{B}}(z) \le \frac{M_{(r+2)\mathbf{B}}}{|r\mathbf{B}|} + \delta\,.$$

Thereby (7) and (8) follow. If $\mathbf{P}_z$ is stationary, $\lambda^{\infty}_G(z)$ is independent of G and (9) follows immediately from the first assertion. □

Remark 4. Using the integral characterization of Gibbs processes due to Nguyen & Zessin (1979a), if $\mathbf{P}_z$ is stationary,

$$\int \#X_{r\mathbf{B}}\, \mathbf{P}_z(dX) = z|r\mathbf{B}| \int e^{-H(\{0\},X)}\, \mathbf{P}_z(dX)\,.$$

Therefore, for any stationary pure hard core process $\mathbf{P}_z$,

$$\lim_{z\to\infty} z\mathbf{P}_z\{X \cap (2\mathbf{B}) = \emptyset\} = \rho\,.$$

Also, as in Remark 3, it can be shown that

$$\lim_{z\to\infty} \limsup_{r\to\infty} \mathrm{Var}_{\mathbf{P}_z} \left\{ \frac{\#X_{r\mathbf{B}}}{|r\mathbf{B}|} \right\} = 0\,.$$

It is plausible that the global intensity like the conditional intensity is increasing in z, but it seems difficult to verify this because of the possible existence of phase transitions. Our final result (Proposition 3 below) shows that the global intensity is indeed non-decreasing possibly except for those values of z for which phase transition occur.

We need first some technical results, where $\{G_l\}$ is a given sequence of increasing bounded convex subsets of $\mathbf{R}^d$ so that the supremum of the radii of balls lying in G_l converges to ∞.

Lemma 2 (Nguyen & Zessin, 1979b, Lemma 3.1). *Let*

$$\Delta_j = \{(x_1, \ldots, x_d); j_i - 1/2 \le x_i < j_i + 1/2, 1 \le i \le d\}$$

be the unit hyper-cube with center $j = (j_1, \ldots, j_d) \in \mathbf{Z}^d$. Further, for $h = 1, 2, \ldots$, let $\bar{G}_l^h$ (resp. $\underline{G}_l^h$) be the union of those $h\Delta_j$ which intersect with (resp. are included in) G_l. Furthermore, let G_l^h be the set of those $x \in \mathbf{R}^d$ which is within distance h from the boundary of G_l. Then $|\bar{G}_l^h|/|G_l| \to 1$, $|\bar{G}_l^h \backslash \underline{G}_l^h|/|G_l| \to 0$, and $|G_l^h|/|G_l| \to 1$ as $l \to \infty$.

The next lemma is the key to the proof of Proposition 3. In the following we only consider the free boundary condition and let $Z_{z,G} = Z_{z,G}(\emptyset)$. Further, $\mathbf{E}_{z,G}$ denotes expectation with respect to $\mathbf{P}_{z,G} = \mathbf{P}_{z,G}(\cdot \mid \emptyset)$.

Lemma 3. *Consider a $z > 0$ so that $\mathbf{P}_z$ is unique and hence stationary. Let $\Delta \subset G$ be two bounded Borel sets with $|\Delta| > 0$ and set*

$$\lambda_{\Delta,G}(z) = \int \frac{\#(X \cap \Delta)}{|\Delta|} \mathbf{P}_{z,G}(dX).$$

Then there is a subsequence $\{G'_l\}$ of $\{G_l\}$ such that

$$\lambda_{\Delta,G'_l}(z) \to \lambda^\infty(z).$$

Proof. This result follows from Ruelle (1970, Theorem 5.5). Let $\rho^m_{\Delta,G}((x)_m)$ be the m-point density function of $X \cap \Delta$ for $\mathbf{P}_{z,G}$, that is, for $m = 0, 1, \ldots,$

$$\rho^m_{\Delta,G}((x)_m) = \frac{1}{Z_{z,G}} \sum_{n=0}^{\infty} \frac{z^{n+m}}{n!} \int_{(G\setminus\Delta)^n} d(y)_n e^{-H((y)_n \cup (x)_m)}$$

if $(x)_m \subset \Delta$ and $\rho^m_{\Delta,G}((x)_m) = 0$ otherwise. It is seen that

$$\int f(X \cap \Delta)\mathbf{P}_{z,G}(dX) = \sum_{m=0}^{\infty} \frac{1}{m!} \int_{\Delta^m} d(x)_m f((x)_m)\rho^m_{\Delta,G}((x)_m).$$

Then Ruelle showed that there is a subsequence $\{G'_l\}$ such that the following limit exists uniformly in $(x)_m$,

$$\lim_{l'\to\infty} \rho^m_{\Delta,G'_l}((x)_m) = \sigma^m_{\Delta}((x)_m)\,.$$

Here $\{\sigma^m_{\Delta}\}$ is the system of m-point density functions of the marginal distribution of $X \cap \Delta$ under $\mathbf{P}_z$. Since $\#(X \cap \Delta)$ is always bounded by $|\Delta \oplus \mathbf{B}|/|\mathbf{B}|$,

$$\begin{aligned} \lambda_{\Delta,G'_l}(z) &= \sum_{m=0}^{\infty} \frac{1}{m!} \int_{\Delta^m} d(x)_m \frac{m}{|\Delta|} \rho^m_{\Delta,G'_l}((x)_m) \\ &\to \sum_{m=0}^{\infty} \frac{1}{m!} \int_{\Delta^m} d(x)_m \frac{m}{|\Delta|} \sigma^m_{\Delta}((x)_m) = \lambda^{\infty}(z). \end{aligned}$$

This completes the proof. □

Remark 5. The construction of the system of m-point density functions $\{\sigma^m_{\Delta}\}$ as a limit of $\{\rho^m_{\Delta,G}\}$ is the basic idea of the distinguished existence proof of global Gibbs processes due to Ruelle (1970). The system $\{\sigma^m_{\Delta,G}\}$ exists even if there occurs a phase transition at z, but it may depend on the choice of subsequences.

Proposition 3. *Let I be the set of values z where the global Gibbs point process $\mathbf{P}_z$ is unique. Then the local intensity converges to the global intensity, that is,*

$$\lim_{l\to\infty} \lambda_{G_l}(z) = \lambda^{\infty}(z) \quad \textit{for} \quad z \in I\,.$$

In particular, the global intensity $\lambda^{\infty}(z)$ is non-decreasing on I.

Proof. The second assertion follows from the first assertion combined with Lemma 1. By Lemma 3, for any given $\epsilon > 0$ there is an $r > 0$ such that

$$|\lambda_{\Delta_0,G_l}(z) - \lambda^{\infty}(z)| \le \epsilon$$

whenever $G_l \supset \Delta_0 \oplus r\mathbf{B}$. Note that also $|\lambda_{\Delta_j,G_l}(z) - \lambda^\infty(z)| \leq \epsilon$ if $G_l \supset \Delta_j \oplus r\mathbf{B}$. For sufficiently large l, G_l can be decomposed as the mutually disjoint union

$$G_l = \Big(\bigcup_{j\in D_l} \Delta_j\Big) \bigcup S_l$$

where D_l is the set of those j for which $\Delta_j \oplus r\mathbf{B} \subset G_l$. Consequently,

$$\begin{aligned} |\lambda_{G_l}(z) - \lambda^\infty(z)| &\leq \frac{1}{|G_l|}\sum_{j\in D_l} |\lambda_{\Delta_j,G_l}(z) - \lambda^\infty(z)| + \frac{|S_l|}{|G_l|}|\lambda_{S_l,G_l}(z) - \lambda^\infty(z)| \\ &\leq \frac{\#D_l}{|G_l|}\epsilon + \frac{|S_l|}{|G_l|}|\lambda_{S_l,G_l}(z) - \lambda^\infty(z)|. \end{aligned}$$

Using Lemma 2 it can be shown that $|S_l|/|G_l| \to 0$, $\#D_l/|G_l| \to 1$, and $\lambda_{S_l,G_l}(z)$ is uniformly bounded in l. This verifies the first assertion. □

4 Markov chain Monte Carlo for hard core Gibbs processes

For specificity consider a planar stationary global pure hard core Gibbs process, i.e. $d = 2$, $\Phi(x) = \infty$ for $||x|| \leq R$ and $\Phi(x) = 0$ otherwise. Proposition 3 establishes the monotonicity of the intensity $\lambda(z) = \lambda^\infty(z)$, but many other questions concerning the qualitative behaviour of $\lambda(z)$ are still open, in particular the question concerning existence of discontinuities of $\lambda(z)$ or its derivatives. Equivalently we may consider the area fraction

$$A_A(z) = \lambda(z)\pi R^2/4 \tag{12}$$

of non-overlapping discs of diameter R and centered at the points in the process; note that $\lambda(z)$ but not $A_A(z)$ depends on R.

Discontinuities of $A_A(z)$, 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" $A_A = 0.69$ and the "melting point" $A_A = 0.716$ (Weber *et al.*, 1995; Mitus *et al.*, 1997; Truskett *et al.*, 1998). It is expected that also changes from short to long range order of the spatial distributional behaviour of the global pure hard core Gibbs process appear at the phase transition point(s), 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 local pure hard core Gibbs process defined on a bounded region G. In our simulations $G = [0, a]^2$ is a square and we use the periodic boundary condition, i.e. our target density with respect to the unit rate Poisson process on G is proportional to

$$\pi_z(\xi) = z^{\#\xi}\mathbf{1}[\forall x, y \in \xi : ||x - y|| > R \text{ if } x \neq y] \tag{13}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function. Here and henceforth,

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

denotes geodesic distance when G is wrapped on a torus. Note that if we fix the number of points in (13) to be $\#\xi = n$, then π_z can be considered as a conditional density with respect to the Bernoulli process with n i.i.d. points in G. Clearly this conditional density 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 combine various Metropolis-Hastings algorithms with simulated tempering in Section 4.1. The experimental results concerning the area fraction, the pair correlation function and some other characteristics are discussed in Section 4.2.

4.1 Algorithms

Our basic algorithm is the Metropolis-Hastings algorithm (MH) studied in Geyer and Møller (1994), Geyer (1999) and Møller (1999). Section 4.1.1 provides a short description of the this 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 algorithm and simulated tempering (Marinari and Parisi, 1992; Geyer and Thompson, 1995) is studied in Section 4.1.2.

4.1.1 Basic algorithm

For latter purposes it is convenient to describe the MH algorithm when we want to simulate from any unnormalized density g, i.e. g is a non-negative integrable function with respect to the unit rate Poisson process on G.

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

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

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

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

The proposed state η' is finally accepted with probabilities $\min\{1, r(\eta, \eta')\}$, where the Hastings ratio $r(\eta, \eta')$ depends on the type of transition and is given by

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

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

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

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

In our simulations, $p_1(\eta)=p_2(\eta)=p$ are constant with $0\le p\le 1/2$; the densities $d(\eta,\cdot)$ and $b(\eta,\cdot)$ are uniform on η and G, respectively; and the density $m(\eta\setminus\{x\},x,\cdot)$ 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 $p=0$ where the number of points is fixed.

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 the Bernoulli process on G if $p=0$. In particular, if $g=\pi_z$ is the target density (13), the Markov chain is uniformly ergodic when $p>0$, and also when $p=0$ provided that R is sufficiently small (this is needed to ensure irreducibility).

For large values of z, however, the chain, despite the property of uniform ergodicity, converges very slowly and produces highly autocorrelated samples. As an example we applied the MH algorithm with $p=0.1$ and $\epsilon=0.3$ for simulation of the process with $g=\pi_z$, $G=[0,10]^2$, $R=1$ and $\log z=12.62$. Figure 1 shows time series and estimated autocorrelations for two statistics, viz. the number of points and the empirical mean of the first coordinates of the points in the point pattern. The time series of length 10000 were obtained by subsampling each 5000th state of the chain given by 5×10^7 basic updates, i.e. either insert, delete or move. The time series are highly autocorrelated and the algorithm has not even converged. In particular the chains get stuck after about the first 7000 subsamples. The estimated acceptance probabilities for insert, delete and move are 2×10^{-5}, 3×10^{-6} and 0.12, respectively. The two first acceptance probabilities had to be equal for a chain in equilibrium.

4.1.2 Simulated tempering

Much better results are obtained when the MH algorithm is combined with simulated tempering as described in the following.

The equilibrium distribution of our implementation of simulated tempering is a mixture of repulsive point process models with unnormalized densities $g_1,\ldots,g_n$, $n\ge2$, where the MH algorithm for g_i mixes well when i is small, while it produces highly autocorrelated samples when i increases towards n. Specifically, for $i=1,\ldots,n$,

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

with $0=\gamma_1<\gamma_2<\cdots<\gamma_{n-1}<\gamma_n=\infty$ and $c>0$, and where we set $0\times\infty=0$. The terms $\gamma_i\mathbf{1}(||x-y||\le R)$ and $\gamma_i c|b(x,R/2)\cap b(y,R/2)|/|b(0,R/2)|$ both introduce a

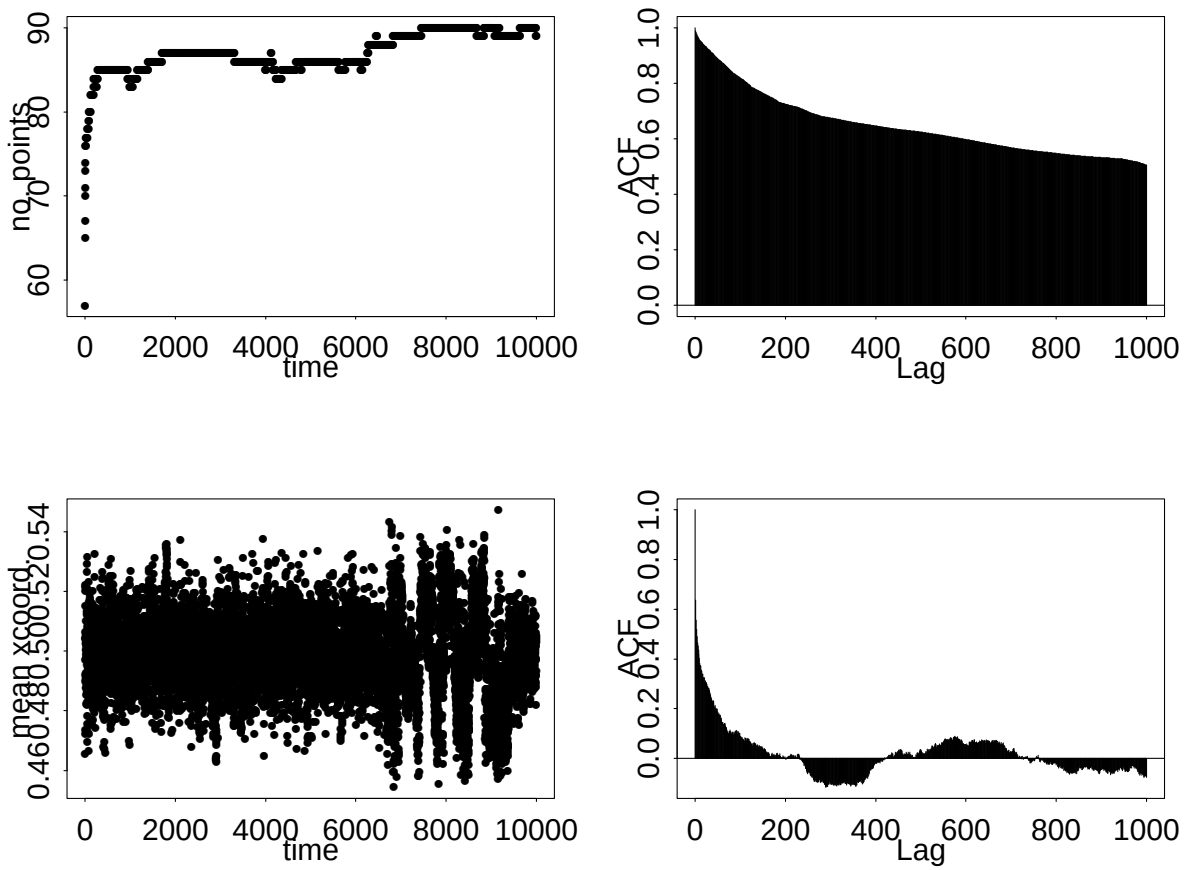

Figure 1: Time series and estimated autocorrelations obtained with the Metropolis-Hastings algorithm. Upper row: number of points. Lower row: statistic given by the empirical mean of the first coordinates in the point pattern.

penalty whenever two hard 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" distribution. For the simulations reported in this paper, the value $c = 10$ was chosen as a result of some pilot simulations. We discuss below how to choose the other parameters in order to obtain an algorithm which inherits the good mixing properties of the MH algorithms for small i.

Simulated tempering generates a Metropolis-Hastings chain $(X_l, I_l)_{l\geq 0}$ whose equilibrium distribution is given by the (unnormalized) density

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

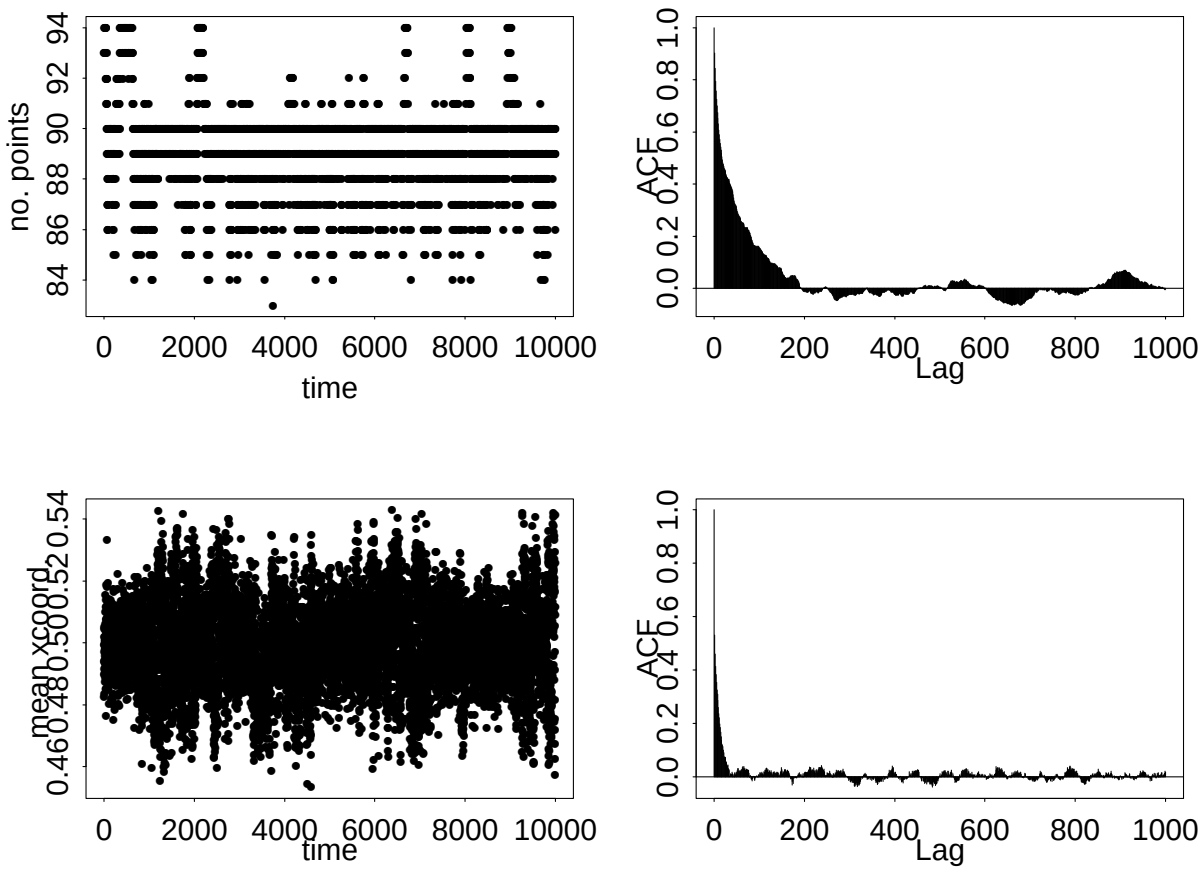

Figure 2: Time series and estimated autocorrelations for the coldest chain in simulated tempering. Upper row: number of points. Lower row: statistic given by the empirical mean of the first coordinates in the point pattern.

where $\delta_i > 0$ is specified as follows. Suppose that $(X, I) \sim \tilde{g}$ where I is a so-called auxiliary variable. The marginal distribution of X is then the mixture $\sum_{i=1}^{n} g_i \delta_i$; g_i is the (unnormalized) conditional density of $X|I = i$; and $P(I = i) \propto \delta_i c_i$ where c_i denotes the normalizing constant (grand partition function) of g_i. Estimates $\hat{c}_i$ of c_i, $i = 1, \ldots, n$, 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, 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. By choosing $\delta_i = 1/\hat{c}_i$ an approximate uniform mixture is obtained.

Now, for the simulated tempering algorithm a proposal kernel Q on $\{1, \ldots, n\}$ is de-

fined by $Q(i, i+1) = Q(i, i-1) = 1/2$ for $1 < i < n$ and $Q(1,2) = Q(n, n-1) = 1$. Given a current state (η, i), the two components are updated in turn using first the MH update $\eta \to \eta'$ for the density $g_i(\cdot)$, and secondly the kernel Q is used to propose an update $(\eta', i) \to (\eta', i')$: we return (η', i') with probability $\min\{1, r(i, i'|\eta')\}$ and retain (η', i) otherwise, where $r(i, i'|\eta') = \tilde{g}(\eta', i')Q(i', i)/(\tilde{g}(\eta', i)Q(i, i'))$. 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. As shown in Appendix A, geometric or uniform ergodicity of the simulated tempering chain can be established under mild conditions; thereby the strong law of large numbers and central limit theorems apply for the Monte Carlo calculations, see e.g. Møller (1999, Appendix A) and the references therein.

Regeneration may be useful for estimation of Monte Carlo errors as explained in Geyer and Thompson (1995). A regeneration step can be done at the hot temperature, i.e. if the simulated tempering chain reaches a state $(X_l, 1)$, the point pattern X_l is replaced by a completely new generated point pattern with the Poisson process density g_1.

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 $(\eta, i) \leftrightarrow (\eta, i \pm 1)$ are obtained, the chain $(X_l)_{l \geq 1: I_l = n}$ yields a well-mixed sample from the target model g_n. Let (p_i, ϵ_i) denote the parameter values for each of the MH algorithms combined in the simulated tempering algorithm, $i = 1, \ldots, n$. We choose the parameter ϵ_i to be decreasing as a function of i so that reasonable accept 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 acceptances probabilities for the low temperatures. The intensity of the Poisson process with density g_1 is chosen as $z_1 = 1/R^2$. This value corresponds to the area fraction $A_A(z) = \pi/4 = 0.785$ of a planar stationary global pure hard core Gibbs process with the same intensity. The remaining parameters are chosen as

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

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 $(\eta, i) \leftrightarrow (\eta, i \pm 1)$ are done similarly to Geyer and Thompson (1995, Section 2.3).

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 simulated tempering sample with $n = 17$ temperatures and of length $17 \times 5 \times 10^7$; the parameters for g_n are again $G = [0, 10]^2$, $R = 1$ and $\log z_n = 12.62$. The time series were obtained by subsampling the cold chain in order to obtain time series with lengths 10000 as in Figure 1; another plot (not shown here) confirms that each (I_l) is approximately uniformly distributed as desired. The estimated acceptance probabilities for the cold chain are 3×10^{-6}, 3×10^{-6} and 0.10,

for insert, delete and move, respectively. The cost of obtaining the simulated tempering sample is approximately 17 times the cost for the usual MH algorithm, but convergence is now achieved and autocorrelations are very much reduced.

In Section 4.2.2 below we further consider results obtained with the simulated tempering algorithm.

4.2 Results

In the following we report on some simulated results for the area fraction (Section 4.2.1) and some other characteristics (Sections 4.2.2–4.2.4) under the target model (13). These characteristics seem to describe various aspects of certain "discontinuities" around the phase transition between the freezing and melting points. Moreover some of the characteristics measure the degree of order in the point patterns compared to an equilateral triangular lattice.

For each considered value of z we used the simulated tempering 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 simulated tempering for the canonical ensemble (i.e. all $p_i = 0$).

4.2.1 Area fraction

By (12) and the periodic boundary condition, we estimate $A_A(z)$ by

$$\hat{A}_A(z) = \frac{\pi R^2 \overline{N}(z)}{4|G|}$$

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

For $G = [0, 10]^2$, $R = 1$ and $\log z_n$ between 3 and 14 we obtained the increasing curve of estimated area fraction $\hat{A}_A(z)$ shown in Figure 3. In the simulated tempering chain, for each considered value of z, we generated a sample of length between 5×10^8 and 10^{10}. The calculation of $\overline{N}(z)$ is based on the 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).

Figure 3 shows that for a wide range of z values the curve of $\hat{A}_A(z)$ nearly coincides with the curve obtained by a Padé approximation (see Appendix B). This indicates that for values of $\log z < 9$ (corresponding to $A_A < 0.65$) both the Padé formula and our simulations yield good approximations of $A_A(z)$. Notice the change in the $\hat{A}_A(z)$ curve at values close to the freezing point $A_A = 0.69$ and the melting point point $A_A = 0.716$ mentioned after equation (12); in particular this indicates a jump in the curve at the melting point. The behaviour of the curve of $\hat{A}_A(z)$ for $\log z > 13.3$ raises doubt about if sufficiently long simulated tempering chains have been used for the values of $\log z > 13.3$. Particularly we believe that the curve should increase further and not show a flat behaviour as for the largest values of $\log z$ in Figure 3.

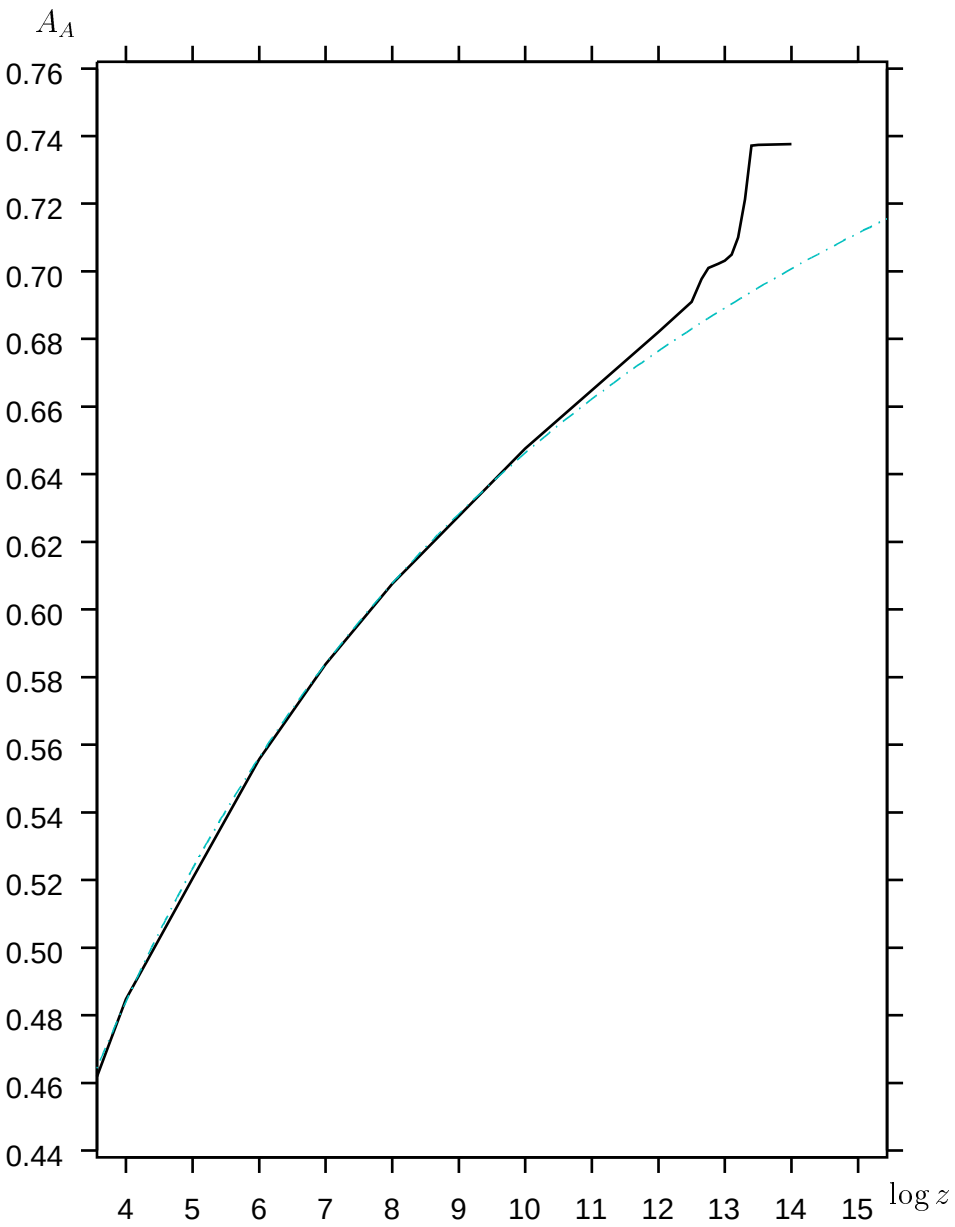

Figure 3: Estimated values of $A_A(z)$ using simulated tempering (solid line) and Padé approximation (broken line).

4.2.2 Pair correlation function

For the results in this and subsequent sections we used $G = [-20, 20]^2$, $R = 1$ and determined the number N of points for every value of area fraction $A_A = 0.65, 0.67, 0.69, 0.696, 0.701, 0.707, 0.71, 0.715, 0.721, 0.735$, so that $A_A = \pi R^2 N/(4|G|)$ in accordance with (12). 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 for various reasons subsam-

ples of 100 point patterns of the cold chain obtained from mucher longer runs of the ST chain: The estimation was first done for each point pattern, then we averaged over the 100 estimates; but the cost of the estimation for one point pattern is about 500 times the cost of one step in the simulated tempering chain. Moreover, because of the small changes in the simulated tempering chain, estimates based on subsequent point patterns look almost the same. Therefore, for the 100 point patterns, we used a spacing of at least $10Nn$.

The pair correlation function pcf 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 $\mathbf{R}^2$, assuming stationarity and isotropy of the hard core Gibbs point process, $\lambda(z)^2\mathrm{pcf}(r)(\pi\delta^2)^2$ 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 (13), 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 function can be estimated by non-parametric kernel methods as described in Stoyan and Stoyan (1994) apart from the following modifications. We replaced the intensity λ with $N/|G|$. Furthermore, because of the 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, see 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 the averaging, the variance in the estimator is substantially reduced.

Estimated pair correlation functions with $A_A = 0.65$ and $A_A = 0.735$ are shown in Figure 4. As expected, with increasing A_A the pair correlation function reflects more order. The peaks of the estimated pair correlation functions can be compared with the modes at $r = 1, 1.732, 2, \ldots$ for the pair correlation function of the limiting regular hexagonal pattern of hard discs with diameter $R = 1$. Clearly the curve for $A_A = 0.735$ is in better agreement to the limiting case than the curve for $A_A = 0.65$. In particular the second mode for the curve with $A_A = 0.65$ splits into two modes as A_A increases.

As mentioned at the end of Section 1, in statistical physics mostly the ordinary Metropolis algorithm (i.e. the MH algorithm for the canonical ensemble) and molecular dynamics have been used for simulations. Molecular dynamics (MD) is based on the equations of motion of N molecules described by a local Gibbs process (see e.g. Strandburg, 1988; Allen and Tildesley, 1987). The theoretical convergence properties of MD are not well understood, but numerical evidence, e.g. obtained by comparison with results produced by the ordinary Metropolis algorithm, supports that MD produces reliable results for pure hard core Gibbs processes (Torquato, 1998, personal communication). This is also supported by our results: Figure 4 is in agreement with the results in Truskett *et al.* (1998) obtained by MD.

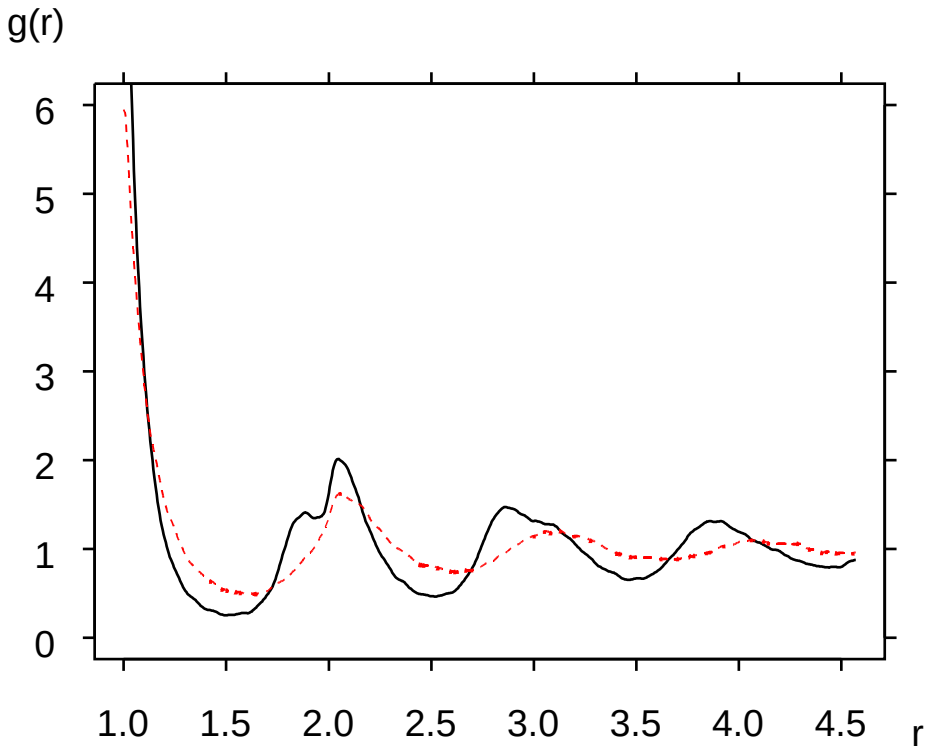

Figure 4: Estimated pair correlation functions for $A_A = 0.65$ (dashed line) and $A_A = 0.735$ (solid line).

4.2.3 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 $\boldsymbol{r} \in \mathbf{R}^2$ with $\|\boldsymbol{r}\| = r$ and let B_r be a square centered at $\mathbf{r}/2$ and of side length αr, where one side is parallel to $\mathbf{r}$ and $0 < \alpha < 1$ is a user-specified parameter. In $\mathbf{R}^2$, assuming stationarity and isotropy of the hard core Gibbs point process, $\lambda(z)|B_r|z_B(r)$ 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 $\mathbf{o} = (0,0)$ and $\boldsymbol{r}$. For $r = 2$, this mean number is exactly 1 in the limiting case of a regular hexagonal pattern of discs with diameter 1. 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 and small values of $z_B(r)$ for suitable r may thus indicate a tendency of alignment in the point pattern. In particular, if α is sufficiently small, one may expect $z_B(2)$ to be an increasing function of A_A with limit $0.2165/\alpha^2$ obtained at the maximal area fraction $A_A = 0.907$.

The statistical estimation of $z_B(r)$ follows the same lines as in Stoyan and Stoyan (1994, page 294) except that we again replace $\lambda(z)$ with $N/|G|$ (since the number of points is fixed) 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 A_A; but $z_B(2) = 1.83$

for $A_A = 0.735$ and this is far from the maximum value 21.65 obtained at $A_A = 0.907$. 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 A_A. Also $z_B(2.2)$ is an increasing function of A_A. Note that the curve of $z_B(2.2)$ is steepest for values of A_A between the freezing and melting points. The value 14.96 of $z_B(2.2)$ for $A_A = 0.735$ is not very far from the upper bound 17.89 obtained by assuming that $\lambda(z)|B_r|z_B(r) \leq 1$ (which holds as $z \to \infty$). However the curve of $z_B(3)$ decreases nearly linearly and slowly towards 0; perhaps surprisingly, this curve is not showing any change at the freezing and melting points.

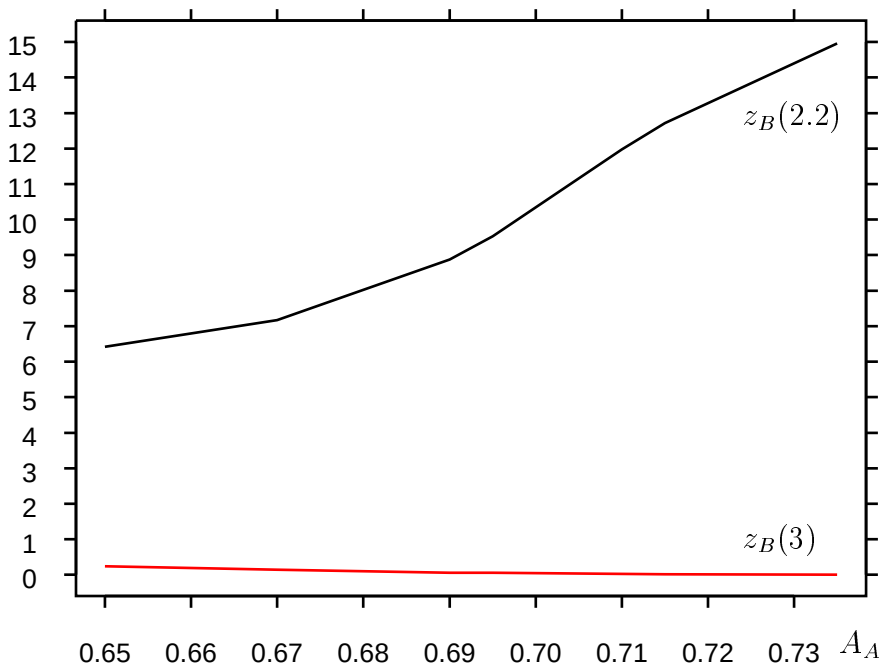

Figure 5: The alignment functions $z_B(2.2)$ (upper curve) and $z_B(3)$ (lower curve) versus area fraction.

4.2.4 Hexagonality number

The idea behind any hexogonality characteristic is to look for deviations from the hexagonal arrangement of neighbouring points to a point in an equilateral triangular lattice.

A first possibility is to use Ripley's K function (see e.g. Stoyan and Stoyan, 1994). In $\mathbf{R}^2$, assuming stationarity, $\lambda(z)K(r)$ is the mean number of points in a disk of radius r centred 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 2. Thus, for the hard core Gibbs point process when r is a bit larger

than 1 and the number of points is fixed, one should expect an abrupt change of the values of $K(r)$ for A_A in the phase transition region. This, however, was not observed in our simulations, where we observed a continuous and nearly linear dependence of A_A (estimation of $(N/|G|)K(r)$ and the other characteristics considered below follow the same lines as in Stoyan and Stoyan, 1994).

Quite different is the behaviour of the "hexagonality number" $H(r)$, the *probability* that a disk of radius r centred at the typical point contains exactly 6 other points. Figure 6 shows the estimated $H(1.3)$ as an increasing function of A_A. The curve is steepest when A_A 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 disk of radius r centred at the 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 disk. Clearly, this characteristic is well adapted to quantify the degree of hexagonality in a point pattern. Figure 7, which shows the estimated $\psi(1.3)$ as a function of A_A, is similar to Figure 6.

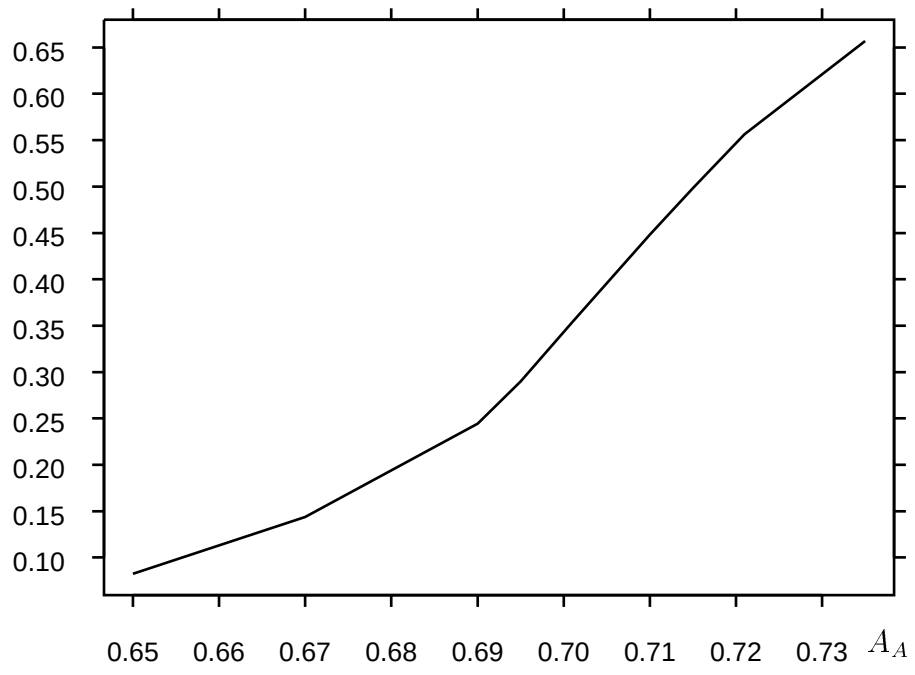

Figure 6: Estimated hexagonality number $H(1.3)$ versus area fraction.

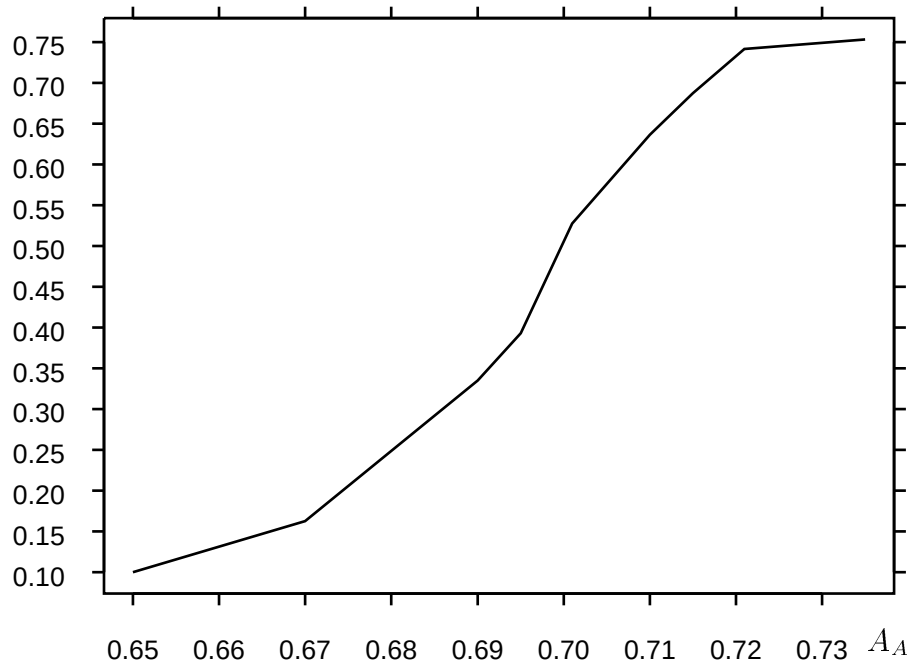

Figure 7: Estimated hexagonality statistic $\psi(1.3)$ versus area fraction.

Acknowledgement: JM and RW was supported by the European Union's network "Statistical and Computational Methods for the Analysis of Spatial Data. ERB-FMRX-CT96-0095", by MaPhySto – Centre 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. We are grateful to Klaus Mecke and Salvatore Torquato for helpful comments.

Appendix A: Geometric ergodicity of the simulated tempering algorithm

Recall that for the i^{th} MH algorithm in the simulated tempering algorithm described in Section 4.1.2, we have that $p_i \in [0, 1/2]$ and $b(\eta, \cdot),\ d(\eta, \cdot),\ m(\eta \setminus \{x\}, x, \cdot)$ are uniform on $G,\ \eta,\ S_{x,i}$, 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 4.1.1). Under these conditions and setting $p_* = \min_i p_i$ and $p^* = \max_i p_i$ we verify below that

- if $p_* > 0$, then the simulated tempering chain defined on the state space $\text{supp}(\tilde{g}) = \{(\eta, i) : g_i(\eta) > 0\}$ 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 $p^* = 0$, so that we have fixed the number of points to be $m < \infty$ then the simulated tempering chain restricted to $\mathrm{supp}(\tilde{g}_{|m})$ is uniformly ergodic with unique invariant distribution specified by $\tilde{g}_{|m}(\eta, i) = \tilde{g}(\eta, i)\mathbf{1}[\#\eta = m]$.

The proofs below are 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 $g_i(\eta \cup \{x\}) \leq K_i g_i(\eta)$ for all $\eta \in \mathcal{C}(G)$ 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 $r(\eta, \eta'|i)$ in the i^{th} MH chain we have an upper bound for births,

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

and similarly a lower bound for deaths,

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

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

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

so for each $k \geq 1$ we can find a $\delta_k > 0$ with $r(i, i-1|\eta) > \delta_k$ for all $(\eta, i) \in \mathrm{supp}(\tilde{g})$ with $\#\eta \leq k$ 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 simulated tempering chain. Suppose $(\eta, i) \in \mathrm{supp}(\tilde{g})$ 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 $k \geq \#\eta$ and set $\rho_k = \min(\delta_k, 1 - r(1, 2|\emptyset)$. Then

$$\begin{aligned} P^{\max(k,n)}((\eta, i), A) &\geq P^{\max(\#\eta, i-1)}((\eta, i), \{(\emptyset, 1)\}) P^{\max(k,n) - \max(\#\eta, i-1)}((\emptyset, 1), \{(\emptyset, 1)\}) \\ &\geq (p_* \rho_k / (2|G|K))^{\max(\#\eta, i-1)} (p_1(1 - r(1, 2|\emptyset))^{\max(k,n) - \max(\#\eta, i-1)} \\ &\geq (p_* \rho_k / (2|G|K))^{\max(k,n)} \phi(A) > 0. \end{aligned} \tag{14}$$

From this we get that the simulated tempering 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 simulated tempering chain is ergodic. From (14) we furthermore get that

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

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):

$$E[V(X_1, I_1)|X_0 = \eta, I_0 = i] \leq \beta V(\eta, i) + b\mathbf{1}[(\eta, i) \in C] \tag{15}$$

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

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

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

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 (15) holds with $\beta = 1 + p^*(\epsilon + 1/K - 1)$, $V(\eta, i) = K^{\#\eta}$, $C = C_{N-1}$ and $b = K^{N+1}$.

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

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

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

Appendix B: The Padé approximation

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

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

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

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

yields

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

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

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

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

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

[8] 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.

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

[10] 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.

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

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

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

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

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

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

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

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

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

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

[21] 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.

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

[23] 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.

[24] 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.

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

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

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

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

[29] Ruelle, D. (1983). *Statistical Mechanics: Rigorous Results*. Addison Wesley, California.

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

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

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

[33] Stoyan, D. and Schlather, M. (1998). On the random sequential absorption model with discs of different sizes. Manuscript.

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

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

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

[37] 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.

[38] 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.