Combinatorics, Probability and Computing (2009) 18, 123–144. © 2008 Cambridge University Press
doi:10.1017/S0963548308009589 Printed in the United Kingdom

Deterministic Random Walks on the Two-Dimensional Grid

BENJAMIN DOERR and TOBIAS FRIEDRICH

Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany
(e-mail: Tobias.Friedrich@mpi-inf.mpg.de)

Received 23 February 2007; revised 14 June 2008; first published online 24 October 2008

Jim Propp's rotor–router model is a deterministic analogue of a random walk on a graph. Instead of distributing chips randomly, each vertex serves its neighbours in a fixed order. We analyse the difference between the Propp machine and random walk on the infinite two-dimensional grid. It is known that, apart from a technicality, independent of the starting configuration, at each time the number of chips on each vertex in the Propp model deviates from the expected number of chips in the random walk model by at most a constant. We show that this constant is approximately 7.8 if all vertices serve their neighbours in clockwise or anticlockwise order, and 7.3 otherwise. This result in particular shows that the order in which the neighbours are served makes a difference. Our analysis also reveals a number of further unexpected properties of the two-dimensional Propp machine.

1. Introduction

The rotor–router model is a simple deterministic process suggested by Jim Propp. It can be viewed as an attempt to derandomize random walks on graphs. So far, the 'Propp machine' has mainly been analysed on infinite grids $\mathbb{Z}^d$. There, each vertex $x \in \mathbb{Z}^d$ is equipped with a 'rotor' together with a cyclic permutation (called a 'rotor sequence') of the $2d$ cardinal directions of $\mathbb{Z}^d$. While a chip (particle, coin, ...) performing a random walk leaves a vertex in a random direction, in the Propp model it always goes in the direction the rotor is pointing. After a chip is sent, the rotor is rotated according to the fixed rotor sequence. This will ensure that the chips are distributed highly evenly among the neighbours.

The Propp machine has recently attracted considerable attention. It has been shown that it closely resembles a random walk in several respects. The first result is due to Levine and Peres [8, 9], who compared a random walk and the Propp machine in an *aggregating model* called Internal Diffusion-Limited Aggregation (IDLA) [4]. There, each chip starts at the origin of $\mathbb{Z}^d$ and walks until it reaches an unoccupied site, which it then occupies. In the random walk model it is well known that the shape of the occupied locations converges to a Euclidean ball in $\mathbb{R}^d$ [7]. Recently, Levine and Peres [8, 9] proved an analogous result for the Propp machine.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

Surprisingly, the convergence seems to be much faster. Kleber [5] showed experimentally that, for circular rotor sequences, after three million chips the radius of the inscribed and circumscribed circles differs by approximately 1.61. Hence, the occupied locations almost form a perfect circle. Some more results on this aggregating model in two dimensions can be found in Section 8.

Cooper and Spencer [1] compared the Propp machine and the random walk in terms of the *single-vertex discrepancy*. Apart from a technicality, which we defer to Section 2, they place arbitrary numbers of chips on the vertices. Then they run the Propp machine on this initial configuration for a certain number of rounds. A round consists of each chip (in arbitrary order) doing one move as directed by the Propp machine. For the resulting position, for each vertex they compare the number of chips that end up there with the expected number of chips that a random walk in the same number of rounds would have placed there starting from the initial configuration. Cooper and Spencer showed that for all grids $\mathbb{Z}^d$, these differences can be bounded by a constant c_d independent of the initial set-up (in particular, the total number of vertices) and the run-time.

For the case $d = 1$, that is, the graph being the infinite path, Cooper, Doerr, Spencer and Tardos [2] showed, among other results, that this constant c_1 is approximately 2.29. They further proved that to maximize the discrepancy on a particular vertex it suffices that each location has an odd number of chips at at most one time.

In this paper, we rigorously analyse the Propp machine on the two-dimensional grid $\mathbb{Z}^2$. A particular difference from the one-dimensional case is that now there are two non-isomorphic orders in which the four neighbours can be served. The first are clockwise and anticlockwise orders of the four cardinal directions. These are called circular rotor sequences. All other orders turn the rotor by 180° at one time and are called non-circular rotor sequences. We prove $c_2 \approx 7.83$ for circular rotor sequences and $c_2 \approx 7.29$ otherwise. To the best of our knowledge, this is the first paper showing that the rotor sequence can make a difference.

We also characterize the respective worst-case configurations. In particular, we prove that the maximal single-vertex discrepancy can only be reached if there are vertices which send a number of chips not divisible by four at least three different times.

These results raise the question of whether all graphs have a constant single-vertex discrepancy. This is not true. Recently, Cooper, Spencer and the authors [3] showed that, for the graph being an infinite k-ary tree ($k \geqslant 3$), the discrepancy is unbounded.

The remainder of this paper is organized as follows. The basic notation is given in Section 2. In Section 3 we show that, roughly speaking, by suitably choosing the initial configuration, we may prescribe the number of chips on each vertex at each time modulo 4. This will yield sharp lower bounds, since in Section 4 we see that the discrepancy on a vertex can be expressed by exactly this information. In Sections 5 and 6, we derive sufficient information about initial configurations leading to maximal discrepancies on a vertex, so that we can then estimate the maximum possible discrepancy numerically. This estimate is shown to be relatively tight in Section 7. Since the investigation up to this point in particular showed that different rotor sequences lead to different results, we briefly examine the aggregating model in this respect in Section 8. We summarize our results in the last section.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

2. Preliminaries

To bound the single-vertex discrepancy between the Propp machine and a random walk on the two-dimensional grid we introduce several requisite definitions and notational conventions in this section.

First, it will be useful to use a different representation of the two-dimensional grid $\mathbb{Z}^2$. Let $\text{DIR} := \{(+1,+1),(+1,-1),(-1,-1),(-1,+1)\}$. Define a graph $G = (V, E)$ via $V = \{(x_1, x_2) \mid x_1 \equiv x_2 \pmod 2\}$ and $E = \{(\mathbf{x}, \mathbf{y}) \in V^2 \mid \mathbf{x} - \mathbf{y} \in \text{DIR}\}$. Clearly, G is isomorphic to the standard two-dimensional grid $G' = (\mathbb{Z}^2, E')$ with $E' = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{Z}^2 \mid \|\mathbf{x} - \mathbf{y}\|_1 = 1\}$. Therefore, our results on G immediately translate to G'. The advantage of our representation is that now each direction $D \in \text{DIR}$ can be uniquely expressed as $D = \varepsilon_x(1,0) + \varepsilon_y(0,1)$ with $\varepsilon_x, \varepsilon_y \in \{-1, 1\}$. This allows a convenient computation of the probability distribution of the random walk on the grid (see equation (2.1) below). For convenience we will also use the symbols $\{ \nearrow, \searrow, \swarrow, \nwarrow \}$ to describe the directions in the obvious manner.

In order to avoid discussing all equations in the expected sense and thereby to simplify the presentation, one can treat the expectation of the random walk as a *linear machine* [1]. Here, in each time step a pile of k chips is split evenly, with $k/4$ chips going to each neighbour. By the 'harmonic property' of random walks, the (possibly non-integral) number of chips at vertex $\mathbf{x}$ at time t is exactly the expected number of chips in the random walk model.

For $\mathbf{x}, \mathbf{y} \in V$ and $t \in \mathbb{N}_0$, let $\mathbf{x} \sim t$ denote that $x_1 \equiv x_2 \equiv t \pmod 2$ and $\mathbf{x} \sim \mathbf{y}$ denote that $x_1 \equiv x_2 \equiv y_1 \equiv y_2 \pmod 2$. A vertex $\mathbf{x}$ is called *even* or *odd* if $\mathbf{x} \sim 0$ or $\mathbf{x} \sim 1$, respectively.

A *configuration* describes the current 'state' of the linear or Propp machine. A configuration of the linear machine is a function $V \to \mathbb{R}_+$, assigning to each vertex $\mathbf{x} \in V$ its current (possibly fractional) number of chips. A configuration of the Propp machine assigns to each vertex $\mathbf{x} \in V$ its current (integral) number of chips and the current direction of the rotor. A configuration is called *even* (*odd*) if all chips lie on even (odd) vertices.

As pointed out in the Introduction, there is one limitation without which neither the results of [1, 2] nor our results hold. Note that since G is a bipartite graph, chips that start on even vertices never mix with those starting on odd vertices. It looks like we are playing two games at once. However, this is not true, because chips at different parity vertices may affect each other through the rotors. We therefore require the initial configuration to have chips only on *one* parity. Without loss of generality, we consider only even initial configurations.

A random walk on G can be described nicely by its probability density. By $H(\mathbf{x}, t)$ we denote the probability that a chip from vertex $\mathbf{x}$ arrives at the origin after t random steps ('at time t') in a simple random walk. Then,

$$H(\mathbf{x}, t) = 4^{-t} \binom{t}{(t + x_1)/2} \binom{t}{(t + x_2)/2} \tag{2.1}$$

for $\mathbf{x} \sim t$ and $\|\mathbf{x}\|_\infty \leqslant t$, and $H(\mathbf{x}, t) = 0$ otherwise.

We now describe the *Propp machine* in detail. First, we define a rotor sequence by a cyclic permutation $\text{NEXT} : \text{DIR} \to \text{DIR}$. That is, after a chip has been sent in direction $\mathbf{A}$, the rotor moves such that afterwards it points in direction $\text{NEXT}(\mathbf{A})$. Instead of using NEXT directly, it will often be more handy to describe a *rotor sequence* as a 4-tuple $\mathcal{R} = (\nearrow, \text{NEXT}(\nearrow), \text{NEXT}^2(\nearrow), \text{NEXT}^3(\nearrow))$. We distinguish between circular and non-circular rotor sequences.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

Circular rotor sequences are either clockwise ($\nearrow, \searrow, \swarrow, \nwarrow$) or anticlockwise ($\nearrow, \nwarrow, \swarrow, \searrow$). All other rotor sequences are called *non-circular*. Our main focus is on the classical Propp machine in which all vertices have the same rotor sequence. In [1], Cooper and Spencer allow different rotor sequences for each vertex $\mathbf{x}$. Our results also hold in this general setting. However, to simplify the presentation we will typically assume that there is only one rotor sequence for all vertices $\mathbf{x}$.

In the following notation, we implicitly fix the rotor sequence as well as the initial configuration (that is, chips on vertices and rotor directions at time $t = 0$). In one step of the Propp machine, each chip makes exactly one move, that is, it moves in the direction the arrow associated with his current position is pointing and updates the arrow direction according to the rotor sequence. Note that the particular order in which the chips move within one step is irrelevant (as long as we do not label the chips). By this observation, all subsequent configurations are determined by the initial configuration. For all $\mathbf{x} \in V$ and $t \in \mathbb{N}_0$, let $f(\mathbf{x}, t)$ denote the number of chips on vertex $\mathbf{x}$ and let $\text{ARR}(\mathbf{x}, t)$ denote the direction of the rotor associated with $\mathbf{x}$ after t steps of the Propp machine.

To describe the linear machine we use the same fixed initial configuration as for the Propp machine. In one step, each vertex $\mathbf{x}$ sends a quarter of its (possibly fractional) number of chips to each neighbour. Let $E(\mathbf{x}, t)$ denote the number of chips at vertex $\mathbf{x}$ after t steps of the linear machine. This is equal to the expected number of chips at vertex $\mathbf{x}$ after a random walk of all chips for t steps. Note that $E(\mathbf{x}, t) = \frac{1}{4} \sum_{\mathbf{A} \in \text{DIR}} E(\mathbf{x} + \mathbf{A}, t - 1)$ by the harmonic property of random walks.

3. Mod-4-forcing theorem

For a deterministic process like the Propp machine, it is obvious that the initial configuration (that is, the location of each chip and the direction of each rotor) determines all subsequent configurations. The following theorem shows a partial converse, namely that (roughly speaking) we may prescribe the number of chips modulo 4 on all vertices at all times and still find an initial configuration leading to such a game. An analogous result for the one-dimensional Propp machine has been shown in [2].

Theorem 3.1 (mod-4-forcing theorem). *For any initial directions of the rotors and any* $\pi : V \times \mathbb{N}_0 \to \{0, 1, 2, 3\}$ *with* $\pi(\mathbf{x}, t) = 0$ *for all* $\mathbf{x} \not\sim t$, *there is an initial even configuration* $f(\mathbf{x}, 0)$, $\mathbf{x} \in V$, *that results in a game with* $f(\mathbf{x}, t) \equiv \pi(\mathbf{x}, t) \pmod 4$ *for all* $\mathbf{x}$ *and* t.

Proof. Let $\text{ARR}(\mathbf{x}, 0)$ describe the initial rotor directions given in the assumption. The sought-after configuration can be found iteratively. We start with $f(\mathbf{x}, 0) := \pi(\mathbf{x}, 0)$ chips at location $\mathbf{x}$.

Now assume that our initial (even) configuration is such that for some $T \in \mathbb{N}$ we have $f(\mathbf{x}, t) \equiv \pi(\mathbf{x}, t) \pmod 4$ for all $t < T$ and $\mathbf{x} \in V$. We modify this initial configuration by defining $f'(\mathbf{x}, 0) := f(\mathbf{x}, 0) + \varepsilon_{\mathbf{x}} 4^T$ for even $\mathbf{x}$, while we have $f'(\mathbf{x}, 0) = 0$ for odd $\mathbf{x}$. Here, $\varepsilon_{\mathbf{x}} \in \{0, 1, 2, 3\}$ are to be determined such that $f'(\mathbf{x}, t) \equiv \pi(\mathbf{x}, t) \pmod 4$ for all $t \leqslant T$ and $\mathbf{x} \in V$.

Observe that a pile of 4^T chips splits evenly T times. Hence, for all choices of the $\varepsilon_{\mathbf{x}}$ we still have $f'(\mathbf{x}, t) \equiv \pi(\mathbf{x}, t) \pmod 4$ for all $t < T$. At time T, the extra piles of 4^T chips have spread

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

as follows:

$$f'(\mathbf{x},T) = f(\mathbf{x},T) + \sum_{\substack{\mathbf{y}\sim 0\\ \|\mathbf{y}-\mathbf{x}\|_\infty \leqslant T}} \varepsilon_{\mathbf{y}} \binom{T}{\frac{T+x_1-y_1}{2}} \binom{T}{\frac{T+x_2-y_2}{2}}.$$

Initially, let $\varepsilon_{\mathbf{y}} := 0$ for all $\mathbf{y} \in V$. By induction on $\|\mathbf{y}\|_1$, we change the $\varepsilon_{\mathbf{y}}$ to their final value. We keep $\varepsilon_{\mathbf{y}} = 0$ for all $\mathbf{y}$ with $\|\mathbf{y}\|_1 < 2T$.

Assume that for some $\theta \in \mathbb{N}_0$, the current $\varepsilon_{\mathbf{y}}$ fulfil $f'(\mathbf{x},T) \equiv \pi(\mathbf{x},T) \pmod 4$ for all $\mathbf{x}$ with $\|\mathbf{x}\|_1 < \theta$. We now determine $\varepsilon_{\mathbf{y}}$ for all $\mathbf{y}$ with $\|\mathbf{y}\|_1 = 2T+\theta$ in such a way that $f'(\mathbf{x},T) \equiv \pi(\mathbf{x},T) \pmod 4$ for all $\mathbf{x} \in V$ such that $\|\mathbf{x}\|_1 \leqslant \theta$.

Fortunately, to achieve $f'(\mathbf{x},T) \equiv \pi(\mathbf{x},T) \pmod 4$ for some $\mathbf{x} \in V$ such that $\|\mathbf{x}\|_1 = \theta$, it suffices to change a single $\varepsilon_{\mathbf{y}}$, $\mathbf{y} \in V$, $\|\mathbf{y}\|_1 = 2T+\theta$. Without loss of generality, let $\mathbf{x} \in V$, $\|\mathbf{x}\|_1 = \theta$, and $\mathbf{x} \sim T$ such that $x_1, x_2 \geqslant 0$. Let $\mathbf{y} = \mathbf{y}(\mathbf{x}) = (x_1+T, x_2+T)$. Now, choosing $\varepsilon_{\mathbf{y}} \in \{0,1,2,3\}$ such that $\varepsilon_{\mathbf{y}} \equiv \pi(\mathbf{x},T) - f(\mathbf{x},T) \pmod 4$ yields $f'(\mathbf{x},T) = f(\mathbf{x},T) + \varepsilon_{\mathbf{y}} \equiv \pi(\mathbf{x},T) \pmod 4$ and $f'(\mathbf{x},T) = f(\mathbf{x},T)$ for all other $\mathbf{x} \in V$ such that $\|\mathbf{x}\|_1 \leqslant \theta$.

Hence, for each $\mathbf{x} \in V$ such that $\|\mathbf{x}\|_1 = \theta$, we find a $\mathbf{y}(\mathbf{x})$ and a value for $\varepsilon_{\mathbf{y}(\mathbf{x})}$ such that the resulting $f'(\mathbf{x},T)$ are as desired. All other $\varepsilon_{\mathbf{y}}$ with $\|\mathbf{y}\|_1 = 2T+\theta$ remain fixed to zero.

This defines a sequence $(f_\theta)_{\theta\in\mathbb{N}}$ of initial configurations $V \times \{0\} \to \mathbb{N}_0$ such that the resulting games $f_\theta : V \times \mathbb{N}_0 \to \mathbb{N}_0$ satisfy $f_\theta(\mathbf{x},t) \equiv \pi(\mathbf{x},t) \pmod 4$ for all $(\mathbf{x},t)$ such that $t < T$ or $\|\mathbf{x}\|_1 \leqslant \theta$. Note that, by construction, $(f_\theta(\mathbf{x},0))_{\theta\in\mathbb{N}}$ is constant for $\theta \geqslant \|\mathbf{x}\|_1$. Hence the initial configuration $f : V \times \{0\} \to \mathbb{N}_0$, $f(\mathbf{x},0) := \lim_{\theta\to\infty} f_\theta(\mathbf{x},0)$ is well defined and the resulting game $f : V \times \mathbb{N}_0 \to \mathbb{N}_0$ satisfies $f(\mathbf{x},t) \equiv \pi(\mathbf{x},t) \pmod 4$ for all $\mathbf{x} \in V$ and $t \leqslant T$.

Up to this point, we have proved that for all $T \in \mathbb{N}$ there is an even initial configuration $f_T : V \times \{0\} \to \mathbb{N}_0$ such that the resulting game $f_T : V \times \mathbb{N}_0 \to \mathbb{N}_0$ satisfies $f_T(\mathbf{x},t) \equiv \pi(\mathbf{x},t) \pmod 4$ for all $t \leqslant T$ and $\mathbf{x} \in V$. Again, $(f_T(\mathbf{x},0))_{T\in\mathbb{N}_0}$ is constant for T sufficiently large compared to $\|\mathbf{x}\|_1$. Hence, as above, $f : V \times \{0\} \to \mathbb{N}_0$, $f(\mathbf{x},0) := \lim_{T\to\infty} f_T(\mathbf{x},0)$ is well defined and the resulting game $f : V \times \mathbb{N}_0 \to \mathbb{N}_0$ satisfies $f(\mathbf{x},t) \equiv \pi(\mathbf{x},t) \pmod 4$ for all $\mathbf{x} \in V$ and $T \in \mathbb{N}_0$. □

4. The basic method

In this section, we lay the foundations for our analysis of the maximal possible single-vertex discrepancy. In particular, we will see that we can determine the contribution of a vertex to the discrepancy at another one independent from all other vertices.

In the following, we re-use several arguments from [1, 2]. For the moment, in addition to the notation given in Section 2, we also use the following mixed notation. By $E(\mathbf{x},t_1,t_2)$ we denote the (possibly fractional) number of chips at location x after first performing t_1 steps with the Propp machine and then $t_2 - t_1$ steps with the linear machine.

We are interested in bounding the discrepancies $|f(\mathbf{x},t) - E(\mathbf{x},t)|$ for all vertices $\mathbf{x}$ and all times t. Since we aim at bounds independent of the initial configuration, it suffices to consider the vertex $\mathbf{x} = \mathbf{0}$. From

$$E(\mathbf{0},0,t) = E(\mathbf{0},t),$$
$$E(\mathbf{0},t,t) = f(\mathbf{0},t),$$

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

we obtain

$$f(\mathbf{0},t) - E(\mathbf{0},t) = \sum_{s=0}^{t-1} \big(E(\mathbf{0},s+1,t) - E(\mathbf{0},s,t)\big).$$

Now $E(\mathbf{0},s+1,t) - E(\mathbf{0},s,t) = \sum_{\mathbf{x}\in V} \sum_{k=1}^{f(\mathbf{x},s)} (H(\mathbf{x} + \text{NEXT}^{k-1}(\text{ARR}(\mathbf{x},s)), t-s-1) - H(\mathbf{x}, t-s))$ motivates the definition of the *influence* of a Propp move (compared to a random walk move) from vertex $\mathbf{x}$ in direction $\mathbf{A}$ on the discrepancy of $\mathbf{0}$ (t time steps later) by

$$\text{INF}(\mathbf{x},\mathbf{A},t) := H(\mathbf{x}+\mathbf{A}, t-1) - H(\mathbf{x},t).$$

To finally reduce all ARRs involved to the initial arrow settings $\text{ARR}(\cdot,0)$, we define $s_i(\mathbf{x}) := \min\{u \geqslant 0 \mid i < \sum_{t=0}^{u} f(\mathbf{x},t)\}$ for all $i \in \mathbb{N}_0$. Hence, at time $s_i(\mathbf{x})$ the location $\mathbf{x}$ is occupied by its ith chip (where, to be consistent with [2], we start counting with the 0th chip).

Let T be a time at which we analyse the discrepancy at $\mathbf{0}$. Then the above yields

$$f(\mathbf{0},T) - E(\mathbf{0},T) = \sum_{\mathbf{x}\in V} \sum_{\substack{i\geqslant 0,\\ s_i(\mathbf{x})<T}} \text{INF}(\mathbf{x}, \text{NEXT}^i(\text{ARR}(\mathbf{x},0)), T - s_i(\mathbf{x})). \tag{4.1}$$

Since the inner sum of equation (4.1) will occur frequently in the remainder, let us define the *contribution* of a vertex $\mathbf{x}$ to be

$$\text{CON}(\mathbf{x}) := \sum_{\substack{i\geqslant 0,\\ s_i(\mathbf{x})<T}} \text{INF}(\mathbf{x}, \text{NEXT}^i(\text{ARR}(\mathbf{x},0)), T - s_i(\mathbf{x})),$$

where we both suppress the initial configuration leading to the $s_i(\cdot)$ as well as the run-time T. Occasionally, we will write $\text{CON}_{\mathcal{C}}$ to specify the underlying initial configuration.

The first main result of this section, summarized in the following theorem, is that it suffices to examine each vertex $\mathbf{x}$ separately.

Theorem 4.1. *The discrepancy between the Propp machine and linear machine after T time steps is the sum of the contributions* $\text{CON}(\mathbf{x})$ *of all vertices* $\mathbf{x}$*, i.e.,*

$$f(\mathbf{0},T) - E(\mathbf{0},T) = \sum_{\mathbf{x}\in V} \text{CON}(\mathbf{x}).$$

Our aim in this paper is to prove a sharp upper bound for the single-vertex discrepancies $|f(\mathbf{y},T) - E(\mathbf{y},T)|$ for all $\mathbf{y}$ and T. As discussed already, by symmetry we may always assume $\mathbf{x} = \mathbf{0}$. To get rid of the dependency of T, let us define $\text{MAXCON}(\mathbf{x})$ to be the supremum contribution of $\mathbf{x}$ over all initial configurations and all T. We will shortly see that the supremum actually is a maximum (Corollary 6.3), that is, there is an initial configuration and a time T such that $\text{CON}(\mathbf{x}) = \text{MAXCON}(\mathbf{x})$. Since the contribution only depends on $T - s_i(\mathbf{x})$ and the (mod-4)-forcing theorem tells us how to manipulate the $s_i(\mathbf{x})$, we may choose T as large as we like (and still have a configuration leading to $\text{CON}(\mathbf{x}) = \text{MAXCON}(\mathbf{x})$). Provided that $\sum_{\mathbf{x}\in V} \text{MAXCON}(\mathbf{x})$ is finite (which we prove in the remainder), we obtain that $\sum_{\mathbf{x}\in V} \text{MAXCON}(\mathbf{x})$ is a *tight* upper bound for $\sup(f(\mathbf{0},T) - E(\mathbf{0},T))$, where the supremum is taken over all initial configurations and all T.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

To bound $|f(\mathbf{0},T)-E(\mathbf{0},T)|$, we need an analogous discussion for negative contributions. Let $\text{MINCON}(\mathbf{x})$ be the infimum contribution of $\mathbf{x}$ over all initial configurations and all T. Fortunately, using symmetries, we can show that $\sum_{\mathbf{x}\in V}\text{MAXCON}(\mathbf{x})=-\sum_{\mathbf{x}\in V}\text{MINCON}(\mathbf{x})$, hence it suffices to deal with positive contributions. Let us briefly sketch the symmetry argument and then summarize the above discussion.

Observe that sending one chip in each direction at the same time does not change $\text{CON}(\mathbf{x})$. That is, for all $\mathbf{x}$ and t we have

$$\sum_{\mathbf{A}\in\text{DIR}}\text{INF}(\mathbf{x},\mathbf{A},t)=0. \tag{4.2}$$

This follows from the definition of INF and the elementary fact $H(\mathbf{x},t)=\frac{1}{4}\sum_{\mathbf{A}\in\text{DIR}}H(\mathbf{x}+\mathbf{A},t-1)$. Based on equation (4.2) we will ignore piles of four chips (and multiples) at a common time t in the remainder of this section. The remaining one to three chips are called *relevant chips*.

To describe the symmetries of CON, we further distinguish the non-circular rotor sequences. We call $(\nearrow,\nwarrow,\searrow,\swarrow)$ and $(\nearrow,\swarrow,\searrow,\nwarrow)$ *x-alternating*, and we call $(\nearrow,\searrow,\nwarrow,\swarrow)$ and $(\nearrow,\swarrow,\nwarrow,\searrow)$ *y-alternating*. Now a short look at the definition of MAXCON reveals symmetries, such as:

- $\text{MAXCON}((x_1,x_2))=\text{MAXCON}((-x_1,-x_2))$ for circular rotor sequences,
- $\text{MAXCON}((x_1,x_2))=\text{MAXCON}((x_1,-x_2))$ for x-alternating rotor sequences, and
- $\text{MAXCON}((x_1,x_2))=\text{MAXCON}((-x_1,x_2))$ for y-alternating rotor sequences.

The following lemma exhibits symmetries for MAXCON and MINCON. It shows that the discrepancies caused by having too few or too many chips have the same absolute value.

Lemma 4.2. *For all $\mathbf{x}\in V$, the following symmetries hold.*

- *Circular rotor sequences:* $\text{MAXCON}((x_1,x_2))=-\text{MINCON}((-x_1,x_2))$.
- *x-alternating rotor sequences:* $\text{MAXCON}((x_1,x_2))=-\text{MINCON}((-x_1,x_2))$.
- *y-alternating rotor sequences:* $\text{MAXCON}((x_1,x_2))=-\text{MINCON}((x_1,-x_2))$.

Proof. The proofs are not difficult, so we only give the one for the first statement. We show that for each configuration $\mathcal{C}_1$ there is another configuration $\mathcal{C}_3$ and a simple permutation π of V with $\text{CON}_{\mathcal{C}_1}(\mathbf{x})=-\text{CON}_{\mathcal{C}_3}(\pi(\mathbf{x}))$ for all implicit run-times T, and assuming the clockwise rotor sequence $\mathcal{R}:=(\nearrow,\searrow,\swarrow,\nwarrow)$ for both $\mathcal{C}_1$ and $\mathcal{C}_3$. By Theorem 3.1, there is a configuration $\mathcal{C}_2$ which sends, using the rotor sequence $(\nearrow,\nwarrow,\swarrow,\searrow)$, a relevant chip from $(-x_1,x_2)$ in direction $(-A_1,A_2)$ at time t if and only if $\mathcal{C}_1$ sends a relevant chip from (x_1,x_2) in direction (A_1,A_2) at time t. Note that $\text{CON}_{\mathcal{C}_2}((-x_1,x_2))=\text{CON}_{\mathcal{C}_1}(\mathbf{x})$. A configuration $\mathcal{C}_3$ which sends, for each single chip $\mathcal{C}_2$ sends, *three* chips from the same vertex in the same direction at the same time obeys rotor sequence $\mathcal{R}$ and gives by equation (4.2) a contribution $\text{CON}_{\mathcal{C}_3}((-x_1,x_2))=-\text{CON}_{\mathcal{C}_2}((-x_1,x_2))=-\text{CON}_{\mathcal{C}_1}(\mathbf{x})$. Hence, $\text{MINCON}((-x_1,x_2))=-\text{MAXCON}((x_1,x_2))$ for the clockwise rotor sequence $\mathcal{R}$. □

Now Lemma 4.2 immediately yields $\sum_{\mathbf{x}\in V}\text{MINCON}(\mathbf{x})=-\sum_{\mathbf{x}\in V}\text{MAXCON}(\mathbf{x})$. Therefore, it suffices to consider maximal contributions.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

Theorem 4.3.

$$\sup_{\mathcal{C},T} |f(\mathbf{0},T) - E(\mathbf{0},T)| = \sum_{\mathbf{x}\in V} \text{MAXCON}(\mathbf{x})$$

is a tight *upper bound for the single-vertex discrepancies.*

5. The modes of INF

In Theorem 4.3 we expressed the discrepancy as the sum of contributions $\text{CON}(\mathbf{x})$, which in turn are sums of the influences $\text{INF}(\mathbf{x},\mathbf{A},t)$. To bound the discrepancy, we are now interested in the extremal values of such sums. In this section we derive some monotonicity properties of these sums. For this, we define

$$\text{INF}(\mathbf{x},\mathcal{A},t) := \sum_{\mathbf{A}\in\mathcal{A}} \text{INF}(\mathbf{x},\mathbf{A},t)$$

for a finite sequence $\mathcal{A} := (\mathbf{A}^{(1)},\mathbf{A}^{(2)},\ldots)$ of rotor directions ordered according to a fixed rotor sequence. In the remainder of the article all finite sequences of rotor directions for which we use the calligraphic $\mathcal{A}$ are ordered according to their respective rotor sequence.

Let $X \subseteq \mathbb{R}$. We call a mapping $f : X \to \mathbb{R}$ *unimodal* if there is a $t_1 \in X$ such that $f|_{x\leqslant t_1}$ as well as $f|_{x\geqslant t_1}$ are monotone. We call a mapping $f : X \to \mathbb{R}$ *bimodal* if there are $t_1, t_2 \in X$ such that $f|_{x\leqslant t_1}$, $f|_{t_1\leqslant x\leqslant t_2}$, and $f|_{t_2\leqslant x}$ are monotone. We call a mapping $f : X \to \mathbb{R}$ *strictly bimodal* if it is bimodal but not unimodal. In the following, we show that all $\text{INF}(\mathbf{x},\mathcal{A},t)$ are bimodal in t.

From equation (4.2) we see that

$$\begin{aligned} \text{INF}(\mathbf{x},(\mathbf{A}^{(1)},\mathbf{A}^{(2)},\mathbf{A}^{(3)}),t) &= -\text{INF}(\mathbf{x},\text{DIR}\setminus\{\mathbf{A}^{(1)},\mathbf{A}^{(2)},\mathbf{A}^{(3)}\},t) \quad \text{and} \\ \text{INF}(\mathbf{x},(\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(k)}),t) &= \text{INF}(\mathbf{x},(\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(k-4)}),t) \quad \text{for } k \geqslant 4. \end{aligned} \tag{5.1}$$

This shows that it suffices to examine $\text{INF}(\mathbf{x},\mathcal{A},t)$ for $\mathcal{A}$ of length one and two, which is done in Lemmas 5.2 and 5.3, respectively. For both proofs, we need Descartes' Rule of Signs, which can be found in [12].

Theorem 5.1 (Descartes' Rule of Signs). *The number of positive roots counting multiplicities of a non-zero polynomial with real coefficients is either equal to its number of coefficient sign variations* (*i.e., the number of sign changes between consecutive non-zero coefficients*) *or else is less than this number by an even integer.*

With this, we are now well equipped to analyse the monotonicity properties of $\text{INF}(\mathbf{x},\mathcal{A},\cdot)$ for $|\mathcal{A}| \in \{1,2\}$.

Lemma 5.2. *For all* $\mathbf{x} \in V$ *and* $\mathbf{A} \in \text{DIR}$, $\text{INF}(\mathbf{x},\mathbf{A},t)$ *is bimodal in* t. *It is strictly bimodal if and only if*

(i) $\|\mathbf{x}\|_\infty > 6$ *and*
(ii) $-A_1x_1 > A_2x_2 > (-A_1x_1+1)/2$ *or* $-A_2x_2 > A_1x_1 > (-A_2x_2+1)/2$.

Proof. A chip at vertex $\mathbf{x}$ requires at least $\|\mathbf{x}\|_\infty$ time steps to arrive at the origin. Hence, $\text{INF}(\mathbf{x},\mathbf{A},t) = 0$ for $t < \|\mathbf{x}\|_\infty$. We show that $\text{INF}(\mathbf{x},\mathbf{A},\cdot)$ has at most two extrema larger than

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

$\|\mathbf{x}\|_\infty$. The discrete derivative of $\text{INF}(\mathbf{x}, \mathbf{A}, t)$ in t is

$$\text{INF}(\mathbf{x}, \mathbf{A}, t+2) - \text{INF}(\mathbf{x}, \mathbf{A}, t) = \frac{p(\mathbf{x}, \mathbf{A}, t) \cdot \big((t-1)!\big)^2}{4^{t+2}\big(\frac{t+x_1+2}{2}\big)!\big(\frac{t-x_1+2}{2}\big)!\big(\frac{t+x_2+2}{2}\big)!\big(\frac{t-x_2+2}{2}\big)!}$$

with

$$\begin{aligned}
p(\mathbf{x}, \mathbf{A}, t) :=\ & (4A_1x_1 + 4A_2x_2)t^4 \\
& + (-A_1x_1^3 - A_2x_2^3 - A_1x_1x_2^2 - A_2x_2x_1^2 - 6A_1x_1A_2x_2 + 19A_1x_1 + 19A_2x_2)t^3 \\
& + (A_1x_1^3A_2x_2 + A_1x_1A_2x_2^3 - 4A_1x_1^3 - 4A_2x_2^3 - 4A_1x_1x_2^2 \\
& \qquad - 4A_2x_2x_1^2 - 23A_1x_1A_2x_2 + 30A_1x_1 + 30A_2x_2)t^2 \\
& + (A_1x_1^3x_2^2 + A_2x_2^3x_1^2 + 4A_1x_1^3A_2x_2 + 4A_1x_1A_2x_2^3 - 4A_1x_1^3 - 4A_2x_2^3 \\
& \qquad - 4A_1x_1x_2^2 - 4A_2x_2x_1^2 - 32A_1x_1A_2x_2 + 16A_1x_1 + 16A_2x_2)t \\
& - A_1x_1^3A_2x_2^3 + 4A_1x_1^3A_2x_2 + 4A_1x_1A_2x_2^3 - 16A_1x_1A_2x_2.
\end{aligned}$$

We observe that the number of extrema of $\text{INF}(\mathbf{x}, \mathbf{A}, \cdot)$ is exactly the number of roots of $p(\mathbf{x}, \mathbf{A}, \cdot)$. Since this is a polynomial of degree 4 in t, we can use Descartes' Rule of Signs and some elementary case distinctions to show that $p(\mathbf{x}, \mathbf{A}, \cdot)$ has at most two roots larger than $\|\mathbf{x}\|_\infty$. A closer calculation reveals that $p(\mathbf{x}, \mathbf{A}, \cdot)$ has precisely two roots larger than $\|\mathbf{x}\|_\infty$ if $\|\mathbf{x}\|_\infty > 6$ and one of $-A_1x_1 > A_2x_2 > (-A_1x_1 + 1)/2$ and $-A_2x_2 > A_1x_1 > (-A_2x_2 + 1)/2$ hold. □

Lemma 5.3. *For all* $\mathbf{x} \in V$ *and* $\mathbf{A}^{(1)}, \mathbf{A}^{(2)} \in \text{DIR}$ *such that* $\mathbf{A}^{(1)} \neq \mathbf{A}^{(2)}$, $\text{INF}(\mathbf{x}, (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}), t)$ *is unimodal in* t.

Proof. The discrete derivative of $\text{INF}(\mathbf{x}, (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}), t)$ is

$$\begin{aligned}
&\text{INF}(\mathbf{x}, (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}), t+2) - \text{INF}(\mathbf{x}, (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}), t) \\
&\quad = \frac{\big(p(\mathbf{x}, \mathbf{A}^{(1)}, t) + p(\mathbf{x}, \mathbf{A}^{(2)}, t)\big) \cdot \big((t-1)!\big)^2}{4^{t+2}\big(\frac{t+x_1+2}{2}\big)!\big(\frac{t-x_1+2}{2}\big)!\big(\frac{t+x_2+2}{2}\big)!\big(\frac{t-x_2+2}{2}\big)!}
\end{aligned}$$

with $p(\mathbf{x}, \mathbf{A}, t)$ as defined in the proof of Lemma 5.2. Thus, the extrema of INF are the roots of the quartic function $p(\mathbf{x}, \mathbf{A}^{(1)}, t) + p(\mathbf{x}, \mathbf{A}^{(2)}, t)$. Descartes' Rule of Signs now shows that $p(\mathbf{x}, \mathbf{A}^{(1)}, t) + p(\mathbf{x}, \mathbf{A}^{(2)}, t)$ has at most one root larger than $\|\mathbf{x}\|_\infty$ for all $\mathbf{x}$ and $\mathbf{A}^{(1)} \neq \mathbf{A}^{(2)}$. □

6. Maximal contribution of a vertex

We now fix a position $\mathbf{x}$ and a rotor sequence $\mathcal{R}$ to examine $\text{MAXCON}(\mathbf{x})$. Lemmas 5.2 and 5.3 show that $\sum_{\mathbf{A}\in\mathcal{A}} \text{INF}(\mathbf{x}, \mathbf{A}, t)$ is bimodal in t for all finite sequences $\mathcal{A} := (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots)$ of rotor directions ordered according to $\mathcal{R}$. Hence, for all $\mathcal{A}$ there are at most two times at which the monotonicity of $\sum_{\mathbf{A}\in\mathcal{A}} \text{INF}(\mathbf{x}, \mathbf{A}, t)$ changes. A time t at which the monotonicity of $\sum_{\mathbf{A}\in\mathcal{A}} \text{INF}(\mathbf{x}, \mathbf{A}, t)$ changes for some $\mathcal{A}$ is called *extremal*. In case of ambiguities, we define the first such time to be extremal. That is, for unimodal $\sum_{\mathbf{A}\in\mathcal{A}} \text{INF}(\mathbf{x}, \mathbf{A}, t)$, we choose the first time t_1 such that $\sum_{\mathbf{A}\in\mathcal{A}} \text{INF}(\mathbf{x}, \mathbf{A}, t)$ is monotone for $t \leqslant t_1$ and $t \geqslant t_1$. Analogously, for strictly bimodal $\sum_{\mathbf{A}\in\mathcal{A}} \text{INF}(\mathbf{x}, \mathbf{A}, t)$, we choose the first times t_1 and t_2 such that $\sum_{\mathbf{A}\in\mathcal{A}} \text{INF}(\mathbf{x}, \mathbf{A}, t)$ is monotone for $t \leqslant t_1$, $t_1 \leqslant t \leqslant t_2$, and $t \geqslant t_2$. The set of all *extremal times* is denoted by $\text{EX}(\mathbf{x})$.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

$\mathrm{EX}(\mathbf{x})$ can be computed easily. By equation (5.1) it suffices to consider $\mathcal{A}$ of length one and two. The corresponding extremal times are the (rounded) roots of the polynomials $p(\mathbf{x}, \mathbf{A}, t)$ and $p(\mathbf{x}, \mathbf{A}^{(1)}, t) + p(\mathbf{x}, \mathbf{A}^{(2)}, t)$ given in Lemma 5.2. The following lemma shows that the number of extremal times is very limited.

Lemma 6.1. $|\mathrm{EX}(\mathbf{x})| \leqslant 7$

Proof. According to Lemma 5.2, there is at most one rotor direction $\mathbf{A}$ for which $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ is *strictly* bimodal in t. Hence, the number of extremal times of $\mathrm{INF}(\mathbf{x}, \mathcal{A}, t)$ with $|\mathcal{A}| = 1$ is at most five. For a rotor sequence $\mathcal{R} = (\mathbf{R}^{(1)}, \mathbf{R}^{(2)}, \mathbf{R}^{(3)}, \mathbf{R}^{(4)})$, equation (4.2) and Lemma 5.3 show that

$$\mathrm{INF}(\mathbf{x}, (\mathbf{R}^{(1)}, \mathbf{R}^{(2)}), t) = -\mathrm{INF}(\mathbf{x}, (\mathbf{R}^{(3)}, \mathbf{R}^{(4)}), t) \quad \text{and}$$
$$\mathrm{INF}(\mathbf{x}, (\mathbf{R}^{(2)}, \mathbf{R}^{(3)}), t) = -\mathrm{INF}(\mathbf{x}, (\mathbf{R}^{(4)}, \mathbf{R}^{(1)}), t)$$

are unimodal in t. Therefore, the total number of extremal times of $\mathrm{INF}(\mathbf{x}, (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}), t)$ with $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)})$ obeying $\mathcal{R}$ is at most two. □

Between two successive times $t_1, t_2 \in \mathrm{EX}(\mathbf{x}) \cup \{0, T\}$, $\sum_{\mathbf{A}\in\mathcal{A}} \mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ is monotone in t for all $\mathcal{A}$. Such periods of time $[t_1, t_2]$ we call a *phase*. Note that $\sum_{\mathbf{A}\in\mathcal{A}} \mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ could also be constant in a certain phase. This implies that it is monotonically increasing as well as monotonically decreasing. To avoid this ambiguity, we use the terms increasing and decreasing (in contrast to monotonically increasing and decreasing) based on the minima and maxima at extremal times $\mathrm{EX}(\mathbf{x})$, which are unambiguously defined and alternating. We now define precisely when a function $\sum_{\mathbf{A}\in\mathcal{A}} \mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ is increasing or decreasing. Consider the set E of the extremal times of $\sum_{\mathbf{A}\in\mathcal{A}} \mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ as defined above. By Lemmas 5.2 and 5.3 we know that $|E| \in \{1, 2\}$. We call $\sum_{\mathbf{A}\in\mathcal{A}} \mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ *increasing* at t if it has a minimum at the maximal $t' \in E$ with $t' < t$ or a maximum at the minimal $t' \in E$ with $t' > t$. Analogously, we call $\sum_{\mathbf{A}\in\mathcal{A}} \mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ *decreasing* at t if it has a maximum at the maximal $t' \in E$ with $t' < t$ or a minimum at the minimal $t' \in E$ with $t' > t$.

By abuse of language, let us say that $\mathbf{x}$ *sends relevant chips* at time t if $f(\mathbf{x}, T - t) \not\equiv 0 \pmod 4$.

Lemma 6.2. *Let $\mathcal{C}_1$ be an arbitrary configuration with run-time $T \geqslant \max \mathrm{EX}(\mathbf{x})$ and let $\mathrm{CON}_{\mathcal{C}_1}(\mathbf{x})$ be the corresponding contribution of $\mathbf{x}$. Then there is a configuration $\mathcal{C}_2$ with the same run-time and $\mathrm{CON}_{\mathcal{C}_1}(\mathbf{x}) \leqslant \mathrm{CON}_{\mathcal{C}_2}(\mathbf{x})$ that sends relevant chips only at extremal times, i.e., for which the associated f satisfies $f(\mathbf{x}, T - t) \not\equiv 0 \pmod 4$ only if $t \in \mathrm{EX}(\mathbf{x})$.*

Proof. Let $\mathcal{C}_2$ be a configuration with $\mathrm{CON}_{\mathcal{C}_2}(\mathbf{x}) \geqslant \mathrm{CON}_{\mathcal{C}_1}(\mathbf{x})$ and a minimal number of non-extremal times at which relevant chips are sent from $\mathbf{x}$. We assume this number to be greater than zero and show a contradiction.

The sum of the INFs of all chips sent at a certain non-extremal time t is either increasing or decreasing in the phase t lies in.

Let us first assume that it is increasing. Let t' be the minimal t' such that $t' \in \mathrm{EX}(\mathbf{x})$ or there are relevant chips sent at time t' (assume for the moment that such a t' exists). Then, sending the

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

considered pile of relevant chips at time t' instead of time t decreases the number of non-extremal times while not decreasing its contribution. Such a modified configuration exists by Theorem 3.1 and contradicts our assumption on $\mathcal{C}_2$. Therefore, there is no such time t'. This implies that t lies in the last phase and that the relevant chips sent at time t are the last to be sent at all. By $\lim_{t\to\infty} \mathrm{INF}(\mathbf{x}, \mathbf{A}, t) = 0$ for all $\mathbf{A}$, the contribution of the chips sent at time t is negative (since it is an increasing function). Hence, not sending these chips at all does not decrease $\mathrm{CON}_{\mathcal{C}_2}(\mathbf{x})$, but the number of non-extremal times.

The same line of argument holds if the sum of the INFs is decreasing instead of increasing. In this case we use that $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t) = 0$ for all $t < \|\mathbf{x}\|_\infty$. □

Lemma 6.2 immediately gives the following corollary.

Corollary 6.3. *There is an initial configuration and a time T such that* $\mathrm{CON}(\mathbf{x}) = \mathrm{MAXCON}(\mathbf{x})$. *The configuration can be chosen such that $f(\mathbf{x}, T - t) \not\equiv 0 \pmod 4$ only if $t \in \mathrm{EX}(\mathbf{x})$. T can be chosen arbitrarily as long as $T \geqslant \max \mathrm{EX}(\mathbf{x})$.*

Lemma 6.1 and Corollary 6.3 give a simple but costly approach to calculate $\mathrm{MAXCON}(\mathbf{x})$. There are four different initial rotor directions for $\mathbf{x}$ and at each (of the at most seven) extremal times we can either send 0, 1, 2, or 3 relevant chips. As all subsequent rotor directions are chosen according to $\mathcal{R}$, there is only a constant $4 \cdot 4^7 = 65536$ number of configurations to consider. The maximum of the respective $\mathrm{CON}(\mathbf{x})$ will be $\mathrm{MAXCON}(\mathbf{x})$ by Corollary 6.3.

Fortunately, we can also find the worst-case configuration directly. A *block* of a phase $[t_1, t_2]$ is a 4-tuple $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)}) \in \mathrm{DIR}^4$ of rotor directions in the order of $\mathcal{R}$ such that $\sum_{i=1}^{k} \mathrm{INF}(\mathbf{x}, \mathbf{A}^{(i)}, t)$ is increasing in t in this phase for all $k \in \{1, 2, 3\}$. By equation (4.2), this is equivalent to $\sum_{i=k}^{4} \mathrm{INF}(\mathbf{x}, \mathbf{A}^{(i)}, t)$ being decreasing in t within the phase for all $k \in \{2, 3, 4\}$.

Lemma 6.4. *Each phase has a unique block. This is determined by the directions of monotonicity of $t \mapsto \mathrm{INF}(\mathbf{x}, \mathcal{A}, t)$ with $|\mathcal{A}| \in \{1, 2\}$.*

Proof. Consider a fixed phase. We want to show that for all valid combinations of directions of monotonicity of $\mathrm{INF}(\mathbf{x}, \mathcal{A}, t)$ with $|\mathcal{A}| \in \{1, 2\}$ within this phase, there is exactly one permutation $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)})$ of DIR obeying $\mathcal{R}$ such that $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)})$ forms a block.

To describe the *type* of monotonicity of $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ within the phase, we use a function τ with $\tau(\mathbf{A}) := \rightarrow$ if $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ is increasing and $\tau(\mathbf{A}) := \leftarrow$ if it is decreasing. This notation should indicate the direction in which the respective $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ is increasing. As an abbreviation we also use $\tau(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)}) := (\tau(\mathbf{A}^{(1)}), \tau(\mathbf{A}^{(2)}), \tau(\mathbf{A}^{(3)}), \tau(\mathbf{A}^{(4)}))$.

By equation (4.2), we know that there is at least one $\mathbf{A}$ of type $\rightarrow$. If there is exactly one direction $\mathbf{A}$ of type $\rightarrow$, then the unique permutation $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)})$ of DIR obeying $\mathcal{R}$ such that $\tau(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)}) = (\rightarrow, \leftarrow, \leftarrow, \leftarrow)$ is the uniquely defined block. If there are three rotor directions $\mathbf{A}$ of type $\rightarrow$, the block is uniquely defined by $\tau(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)}) = (\rightarrow, \rightarrow, \rightarrow, \leftarrow)$.

It remains to examine the case of exactly two rotor directions of type $\rightarrow$. If these two directions are consecutive in $\mathcal{R}$, then $\tau(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)}) = (\rightarrow, \rightarrow, \leftarrow, \leftarrow)$ again defines the unique block. Otherwise, rotor directions of type $\rightarrow$ and $\leftarrow$ are alternating in the rotor sequence and $(\rightarrow, \leftarrow, \rightarrow, \leftarrow)$ is the only type possible for a block. This allows two blocks $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)})$

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

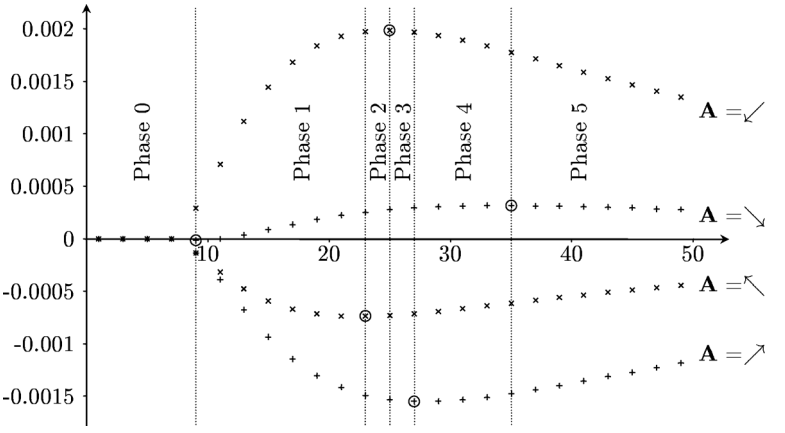


Figure 1. $\text{INF}((5,9),\mathbf{A},t)$ for $\mathbf{A} \in \{\nearrow, \searrow, \swarrow, \nwarrow\}$. The circles indicate the extrema.

and $(\mathbf{A}^{(3)}, \mathbf{A}^{(4)}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)})$. The choice between these two is uniquely fixed by the direction of monotonicity of $\text{INF}(\mathbf{x}, (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}), t)$. Therefore, in all cases there is exactly one unique block. □

We now use Lemma 6.4 to define a particular configuration, which we call a *block configuration*. By Theorem 3.1, to specify a configuration it suffices to fix the number of relevant chips at all times and locations. In a block configuration $\mathcal{B}$, a vertex $\mathbf{x}$ sends relevant chips only at extremal times $t \in \text{EX}(\mathbf{x})$. Let $(\hat{\mathbf{A}}^{(1)}, \hat{\mathbf{A}}^{(2)}, \hat{\mathbf{A}}^{(3)}, \hat{\mathbf{A}}^{(4)})$ and $(\bar{\mathbf{A}}^{(1)}, \bar{\mathbf{A}}^{(2)}, \bar{\mathbf{A}}^{(3)}, \bar{\mathbf{A}}^{(4)})$ denote the blocks in the phases ending and starting at t. Then $\mathbf{x}$ sends k chips at time t in directions $(\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(k)})$, where k is such that $0 \leqslant k \leqslant 3$ and $(\ldots, \hat{\mathbf{A}}^{(4)}, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(k)}, \bar{\mathbf{A}}^{(1)}, \ldots)$ obeys $\mathcal{R}$. This uniquely defines when and in which directions relevant chips are sent. Note that we use the blocks only as a technical tool. There are not necessarily chips sent corresponding to $\hat{\mathbf{A}}^{(1)}, \hat{\mathbf{A}}^{(2)}, \hat{\mathbf{A}}^{(3)}, \hat{\mathbf{A}}^{(4)}$ and $\bar{\mathbf{A}}^{(1)}, \bar{\mathbf{A}}^{(2)}, \bar{\mathbf{A}}^{(3)}, \bar{\mathbf{A}}^{(4)}$. By Theorem 3.1, there are configurations $\mathcal{B}$ as just defined, and for all $\mathbf{x}$, all of them have the same contribution $\text{CON}_{\mathcal{B}}(\mathbf{x})$.

Example. We now derive the block configuration of the position $\mathbf{x} = (5,9)$ with the clockwise rotor sequence $\mathcal{R} = (\nearrow, \searrow, \swarrow, \nwarrow)$. By calculating the roots of the polynomials $p(\mathbf{x}, \mathbf{A}, t)$ and $p(\mathbf{x}, \mathbf{A}^{(1)}, t) + p(\mathbf{x}, \mathbf{A}^{(2)}, t)$ given in Lemma 5.2, it is easy to verify that:

- $\text{INF}(\mathbf{x}, \nearrow, t)$ is unimodal with minimum at $t = 27$,
- $\text{INF}(\mathbf{x}, \searrow, t)$ is bimodal with minimum at $t = 9$ and maximum at $t = 35$,
- $\text{INF}(\mathbf{x}, \swarrow, t)$ is unimodal with maximum at $t = 25$,
- $\text{INF}(\mathbf{x}, \nwarrow, t)$ is unimodal with minimum at $t = 23$,
- $\text{INF}(\mathbf{x}, (\nearrow, \searrow), t)$ and $\text{INF}(\mathbf{x}, (\nwarrow, \nearrow), t)$ are unimodal with minimum at $t = 27$,
- $\text{INF}(\mathbf{x}, (\searrow, \swarrow), t)$ and $\text{INF}(\mathbf{x}, (\swarrow, \nwarrow), t)$ are unimodal with maximum at $t = 27$.

Hence, the extremal points are $\text{EX}(\mathbf{x}) = \{9, 23, 25, 27, 35\}$. In Figure 1 we depict the plots of $\text{INF}(\mathbf{x}, \mathbf{A}, t)$. The modes of $\text{INF}(\mathbf{x}, \mathcal{A}, t)$ listed above uniquely determine the blocks of each phase. Table 1 lists rotor directions and type of the block of each phase. This yields the following

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

Table 1. Rotor directions and type of the block of each phase of $\text{INF}((5,9))$.

	Boundaries of the phase		Block of the phase	
Phase	Lower	Upper	Rotor directions	Type
0	0	9	↙ ↖ ↗ ↘	→←←←
1	9	23	↘ ↙ ↖ ↗	→→←←
2	23	25	↘ ↙ ↖ ↗	→→←←
3	25	27	↘ ↙ ↖ ↗	→←→←
4	27	35	↖ ↗ ↘ ↙	→→→←
5	35	T	↖ ↗ ↘ ↙	→→←←

Table 2. Values of $\text{CON}((5,9))$ for other rotor sequences.

Rotor sequence	Times and directions of relevant chips in a block configuration	$\text{CON}((5,9))$
(↗, ↘, ↙, ↖)	9 :↙ ↖ ↗, 27 :↘ ↙	0.002277 …
(↗, ↖, ↙, ↘)	23 :↙ ↘ ↗, 27 :↖ ↙, 35 :↘	0.002309 …
(↗, ↖, ↘, ↙)	9 :↙ ↗ ↖, 23 :↘ ↙ ↗, 27 :↖ ↘ ↙	0.002302 …
(↗, ↙, ↘, ↖)	25 :↙, 35 :↘	0.002230 …
(↗, ↘, ↖, ↙)	17 :↙ ↗, 27 :↘ ↖ ↙	0.002083 …
(↗, ↙, ↖, ↘)	25 :↙	0.001985 …

(maximal as we will see shortly) contribution at $\mathbf{x} = (5,9)$:

$$\begin{aligned}\text{CON}(\mathbf{x}) &= \text{INF}(\mathbf{x}, \swarrow, 9) + \text{INF}(\mathbf{x}, \nwarrow, 9) + \text{INF}(\mathbf{x}, \nearrow, 9) + \text{INF}(\mathbf{x}, \searrow, 27) + \text{INF}(\mathbf{x}, \swarrow, 27)\\ &= \tfrac{20,506,216,364,597}{9,007,199,254,740,992} \approx 0.002277.\end{aligned}$$

Note that just sending a single chip in the worst direction ↙ at its worst time $t = 25$ gives a smaller contribution of $\text{INF}(\mathbf{x}, \swarrow, 25) \approx 0.001985$. Also, sending two chips in directions ↘ and ↙ at time $27 = \text{argmax}_t \text{INF}(\mathbf{x}, (\searrow, \swarrow), t)$ gives $\text{INF}(\mathbf{x}, (\searrow, \swarrow), 27) \approx 0.002261$. Hence we do profit from sending a chip in the 'wrong' direction ↗ at time 9.

The values of $\text{CON}((5,9))$ for other rotor sequences are shown in Table 2.

Lemma 6.5. *A block configuration yields a contribution of* $\text{MAXCON}(\mathbf{x})$.

Proof. Consider a configuration $\mathcal{C}$ with contribution $\text{CON}_{\mathcal{C}}(\mathbf{x}) = \text{MAXCON}(\mathbf{x})$. By previous considerations, we can further assume the following.

(1) $\mathcal{C}$ only sends relevant chips at times $t \in \text{EX}(\mathbf{x})$ (*cf.* Corollary 6.3).
(2) $\mathcal{C}$ sends at least seven chips at each time $t \in \text{EX}(\mathbf{x})$ (*cf.* equation (4.2)).
(3) Let $t_1, t_2 \in \text{EX}(\mathbf{x})$ such that $[t_1, t_2]$ is a phase and let $k \in \{1,2,3\}$. Let $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(k)}$ be the directions the last k chips are sent from vertex $\mathbf{x}$ at time t_1. If $\sum_{i=1}^{k} \text{INF}(\mathbf{x}, \mathbf{A}^{(i)}, t)$ is increasing (*cf.* the definition on p. 132) in $[t_1, t_2]$, then it is not constant. This is a feasible assumption on $\mathcal{C}$, since otherwise we could send these k chips at time t_2 without changing $\text{CON}_{\mathcal{C}}(\mathbf{x})$.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

(4) Analogously, let $t_1, t_2 \in \text{EX}(\mathbf{x})$ such that $[t_1, t_2]$ is a phase and let $j \in \{1, 2, 3\}$. Let $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(j)}$ be the directions in which the first j chips are sent from vertex $\mathbf{x}$ at time t_2. If $\sum_{i=1}^{k} \text{INF}(\mathbf{x}, \mathbf{A}^{(i)}, t)$ is decreasing in $[t_1, t_2]$, then it is not constant.

Let $\mathcal{B}$ be a block configuration. Aiming at a contradiction, we assume $\text{CON}_{\mathcal{C}}(\mathbf{x}) > \text{CON}_{\mathcal{B}}(\mathbf{x})$. Since by assumption (1) and the definition of $\mathcal{B}$ both configurations send relevant chips only at times in $\text{EX}(\mathbf{x})$, there is a time $t \in \text{EX}(\mathbf{x})$ at which the chips of $\mathcal{C}$ contribute more than the chips of $\mathcal{B}$.

We now closely examine the chips sent from $\mathbf{x}$ at time t by both configurations. We know that $\mathcal{B}$ sends a uniquely determined number $\ell \in \{0, \dots, 3\}$ of relevant chips at time t in some directions $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(\ell)}$. By assumption (2), $\mathcal{C}$ also sends a sequence of chips in directions $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(\ell)}$. Let j and k denote the number of chips sent by $\mathcal{C}$ at time t before and after these ℓ chips, respectively. By ignoring possible piles of four chips, we may assume $j, k \leqslant 3$.

Assume that $k \geqslant 1$. Then the sum of the INFs of the last k chips $\mathcal{C}$ sends at time t is increasing by the definition of a block. Assume first that t is not the last extremal time, that is, there is some $t_2 \in \text{EX}(\mathbf{x})$ such that $[t, t_2]$ form a phase. Then, by assumption (3) above, the sum of the INFs of the last k chips is strictly increasing in $[t, t_2]$. Hence, a configuration which sends these chips instead at t_2 has a larger contribution, in contradiction to the maximality of $\mathcal{C}$. Now let t be the last extremal time. From $\lim_{t\to\infty} \text{INF}(\mathbf{x}, \mathbf{A}, t) = 0$ for all $\mathbf{A}$ and the fact that the sum of the INFs of the last k chips is increasing, we see that it is not positive. Hence the last k chips do not contribute positively to $\text{CON}_{\mathcal{C}}(\mathbf{x})$.

Analogously, assume that $j \geqslant 1$. Assume first that t is not the first extremal time, that is, $[t_1, t]$ form a phase for some $t_1 \in \text{EX}(\mathbf{x})$. By assumption (4), the first j chips $\mathcal{C}$ sends at time t have a strictly monotonically decreasing sum of INFs. Hence sending them at time t_1 instead of t gives a larger contribution, again contradicting the maximality of $\mathcal{C}$. If t is the first extremal time of $\mathbf{x}$, then $\text{INF}(\mathbf{x}, \mathbf{A}, t) = 0$ for all $\mathbf{A}$ and $t < \|\mathbf{x}\|_\infty$ shows, as above, that the contribution of the first j chips is not positive.

We conclude that the first j and last k chips sent from $\mathbf{x}$ and time t in $\mathcal{C}$, if they are present, do not contribute positively to the contribution of $\mathbf{x}$. This contradicts our assumption $\text{CON}_{\mathcal{C}}(\mathbf{x}) > \text{CON}_{\mathcal{B}}(\mathbf{x})$. □

With the help of a computer, we can now calculate $\text{MAXCON}(\mathbf{x})$ for all $\mathbf{x}$. Using about two months on a Xeon 3 GHz CPU, we computed the maximal contribution of all vertices in $[-800, 800]^2$. If we have the same rotor sequence for all vertices then

$$\sum_{\|\mathbf{x}\|_\infty \leqslant 800} \text{MAXCON}(\mathbf{x}) = \begin{cases} 7.832\dots & \text{for a circular rotor sequence,} \\ 7.286\dots & \text{for a non-circular rotor sequence.} \end{cases} \tag{6.1}$$

On the other hand, if we allow a different rotor sequence for each vertex, and further assume that each vertex has a rotor sequence leading to the maximal contribution, then we obtain

$$\sum_{\|\mathbf{x}\|_\infty \leqslant 800} \text{MAXCON}(\mathbf{x}) = 7.873\dots$$

Since $\text{MAXCON}(\mathbf{x})$ is non-negative for all $\mathbf{x} \in V$, all of these values are lower bounds for $\sum_{\mathbf{x}\in V} \text{MAXCON}(\mathbf{x})$, and hence for the single-vertex discrepancy by Theorem 4.3.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

Remark 1. Lemma 6.1 shows that the number of extremal times of a vertex is at most seven. However, a block configuration does not send relevant chips at at all extremal times. Let $\widehat{\mathrm{EX}}(\mathbf{x})$ denote the set of extremal times at which relevant chips are sent by the block configuration. There are vertices $\mathbf{x}$ such that $|\widehat{\mathrm{EX}}(\mathbf{x})| \geqslant 3$. We now sketch a proof that $|\widehat{\mathrm{EX}}(\mathbf{x})| \leqslant 4$ for all $\mathbf{x}$.

Note that $|\widehat{\mathrm{EX}}(\mathbf{x})|$ only depends on the relative order of the extremal points and the initial direction of monotonicity (*i.e.*, increasing or decreasing) of $\mathrm{INF}(\mathbf{x}, \mathcal{A}, t)$ for $|\mathcal{A}| \leqslant 2$. We use the following two properties of INF (derived from equation (4.2)).

- In each phase there is at least one $\mathbf{A} \in \mathrm{DIR}$ such that $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ is increasing (or decreasing).
- If $\mathrm{INF}(\mathbf{x}, \mathbf{A}^{(1)}, t)$ and $\mathrm{INF}(\mathbf{x}, \mathbf{A}^{(2)}, t)$ are both increasing or decreasing in a phase, then so is $\mathrm{INF}(\mathbf{x}, (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}), t)$.

For a vertex $\mathbf{x}$ with only unimodal $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$, there are $6! = 720$ permutations of the extrema of $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ and $\mathrm{INF}(\mathbf{x}, (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}), t)$ and $2^6 = 64$ initial directions of monotonicity (using equation (5.1)). A simple check by a computer shows that for only 384 of these 46080 cases both properties from above are satisfied. For all of them, $|\widehat{\mathrm{EX}}(\mathbf{x})| \leqslant 3$ holds. For vertices $\mathbf{x}$ with $\mathrm{INF}(\mathbf{x}, \mathbf{A}, t)$ strictly bimodal for an $\mathbf{A} \in \mathrm{DIR}$, there are $7!/2! = 2520$ permutations of the extrema and $2^6 = 64$ initial directions of monotonicity. Here, all 408 cases which satisfy both properties only achieve $|\widehat{\mathrm{EX}}(\mathbf{x})| \leqslant 4$. This proves $|\widehat{\mathrm{EX}}(\mathbf{x})| \leqslant 4$ for all $\mathbf{x}$. A computer can easily verify that $|\widehat{\mathrm{EX}}(\mathbf{x})| \leqslant 3$ for all $\|\mathbf{x}\|_\infty \leqslant 800$. Therefore, we actually expect $|\widehat{\mathrm{EX}}(\mathbf{x})| \leqslant 3$ to hold for all $\mathbf{x}$. To bridge this gap, stronger properties of INF seem necessary.

7. Tail estimates

In the previous section, we have calculated the values of $\sum_{\|\mathbf{x}\|_\infty \leqslant 800} \mathrm{MAXCON}(\mathbf{x})$ depending on the rotor sequence. To show that these are good approximations for the maximal single-vertex discrepancy, we need to find an upper bound on

$$E := \sum_{\|\mathbf{x}\|_\infty > 800} \mathrm{MAXCON}(\mathbf{x}).$$

In this section, we will prove $E < 0.16$.

We now fix an arbitrary initial configuration and a time T. A simple calculation based on the definitions of INF and CON gives for all $\mathbf{x}$, $\mathbf{A}$ and t

$$\begin{aligned} \mathrm{INF}(\mathbf{x}, \mathbf{A}, t) &= \big((A_1 x_1 \cdot A_2 x_2) t^{-2} - (A_1 x_1 + A_2 x_2) t^{-1}\big) H(\mathbf{x}, t), \\ \mathrm{CON}(\mathbf{x}) &= \sum_{\substack{i \geqslant 0, \\ s_i(\mathbf{x}) < T}} \left(\frac{A_1^{(i)} x_1 \cdot A_2^{(i)} x_2}{(T - s_i(\mathbf{x}))^2} - \frac{A_1^{(i)} x_1 + A_2^{(i)} x_2}{T - s_i(\mathbf{x})} \right) H(\mathbf{x}, T - s_i(\mathbf{x})), \end{aligned} \tag{7.1}$$

with $s_i(\mathbf{x})$ as defined in Section 4 and $\mathbf{A}^{(i)} := \mathrm{NEXT}^i(\mathrm{ARR}(\mathbf{x}, 0))$. Note that, independent of the chosen rotor sequence, each of the sequences $(A_1^{(i)}(\mathbf{x}))_{i \geqslant 0}$, $(A_2^{(i)}(\mathbf{x}))_{i \geqslant 0}$ and $(A_1^{(i)}(\mathbf{x}) A_2^{(i)}(\mathbf{x}))_{i \geqslant 0}$ is alternating, or alternating in groups of two. To bound the alternating sums in equation (7.1), we use the following fact, which is an elementary extension of Lemma 4 in [2].

Lemma 7.1. *Let $f : X \to \mathbb{R}$ be non-negative and unimodal with $X \subseteq \mathbb{R}$. Let $A^{(0)}, \dots, A^{(n)} \in \{-1, +1\}$ and $t_0, \dots, t_n \in X$ such that $t_0 \leqslant \cdots \leqslant t_n$. If $A^{(i)}$ is alternating, or alternating in groups*

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

of two, then

$$\left|\sum_{i=0}^{n} A^{(i)} f(t_i)\right| \leqslant 2 \max_{x \in X} f(x).$$

It remains to show that $H(\mathbf{x},t)/t$ and $H(\mathbf{x},t)/t^2$ are indeed unimodal. Note that $\textsc{inf}(\mathbf{x},\mathbf{A},t)$ itself is not always unimodal as shown in Lemma 5.2.

Lemma 7.2. *For all* $\mathbf{x} \in V$, $H(\mathbf{x},t)/t$ *and* $H(\mathbf{x},t)/t^2$ *are unimodal in* t *with global maxima at* $t_{\max}(\mathbf{x})$ *and* $t'_{\max}(\mathbf{x})$, *respectively. For the maxima, we have* $(x_1^2+x_2^2)/4-2 \leqslant t_{\max}(\mathbf{x}) \leqslant (x_1^2+x_2^2)/4+1$ *and* $(x_1^2+x_2^2)/6-1 \leqslant t'_{\max}(\mathbf{x}) \leqslant (x_1^2+x_2^2)/6+2$.

Proof. By symmetry, let us assume $x_1 \leqslant x_2$. By definition, $H(\mathbf{x},t)/t = 0$ for $t < x_2$. We show that $H(\mathbf{x},t)/t$ has only one maximum in $t \in [x_2,\infty)$. We compute

$$\frac{H(\mathbf{x},t-2)}{t-2} - \frac{H(\mathbf{x},t)}{t} = \frac{4^{-t} p(t) (t-3)!^2\,(t-2)}{(\frac{t+x_1}{2})!\,(\frac{t-x_1}{2})!\,(\frac{t+x_2}{2})!\,(\frac{t-x_2}{2})!},$$

with $p(t) := 4t^3 - (x_1^2+x_2^2+5)t^2 + 2t + x_1^2x_2^2$. By Descartes' Rule of Signs (see Theorem 5.1), $p(t)$ has at most one real root larger than x_2. Since

$$p\left(\frac{x_1^2+x_2^2}{4}\right) = \tfrac{1}{16}\left(6x_1^2x_2^2 + 8x_1^2 + 8x_2^2 - 5x_1^4 - 5x_2^4\right) < 0,$$
$$p\left(\frac{x_1^2+x_2^2+5}{4}\right) = x_1^2x_2^2 + \frac{x_1^2+x_2^2+5}{2} > 0,$$

we see that $H(\mathbf{x},t)/t$ has a unique extremum, which is a maximum, in $[(x_1^2+x_2^2)/4-2, (x_1^2+x_2^2)/4+1]$. This proves the lemma for $H(\mathbf{x},t)/t$. The analogous proof for $H(\mathbf{x},t)/t^2$ is omitted. □

By equation (7.1), and Lemmas 7.1 and 7.2, we obtain

$$E \leqslant 4E_1 + 2E_2,$$

with

$$E_1 := \sum_{\|\mathbf{x}\|_\infty > 800} \left|\frac{x_1 H(\mathbf{x}, t_{\max}(\mathbf{x}))}{t_{\max}(\mathbf{x})}\right|, \quad E_2 := \sum_{\|\mathbf{x}\|_\infty > 800} \left|\frac{x_1 x_2 H(\mathbf{x}, t'_{\max}(\mathbf{x}))}{(t'_{\max}(\mathbf{x}))^2}\right|.$$

Using Lemma 7.2 and $H(\mathbf{x},t) \leqslant \left(2^{-t}\binom{t}{t/2}\right)^2 \leqslant 1/t$, we now derive upper bounds for $H(\mathbf{x},t)/t$ and $H(\mathbf{x},t)/t^2$ for $\|\mathbf{x}\|_\infty \geqslant 88$:

$$\left|\frac{H(\mathbf{x}, t_{\max}(\mathbf{x}))}{t_{\max}(\mathbf{x})}\right| \leqslant \frac{1}{t_{\max}(\mathbf{x})^2} \leqslant \frac{16}{(x_1^2+x_2^2-8)^2} \leqslant \frac{17}{(x_1^2+x_2^2)^2},$$
$$\left|\frac{H(\mathbf{x}, t'_{\max}(\mathbf{x}))}{t'_{\max}(\mathbf{x})^2}\right| \leqslant \frac{1}{t'_{\max}(\mathbf{x})^3} \leqslant \frac{216}{(x_1^2+x_2^2-6)^3} \leqslant \frac{217}{(x_1^2+x_2^2)^3}.$$

For the calculations in the remainder of this section we need the following estimates. All of them can be derived by bounding the infinite sums with integrals.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

- $\displaystyle\sum_{x>y} \frac{1}{x^k} \leqslant \frac{1}{(k-1)y^{k-1}}$ for all $y > 0$ and all constants $k > 1$.
- $\displaystyle\sum_{x_2=0}^{\infty} \frac{1}{(x_1^2+x_2^2)^2} \leqslant \frac{7}{3\,x_1^3}$ for all $x_1 \geqslant 1$.
- $\displaystyle\sum_{y\geqslant\beta} \frac{1}{(\alpha^2+y^2)y} \geqslant \frac{\ln(\alpha^2+\beta^2)-2\ln(\beta)}{2\alpha^2}$.
- $\displaystyle\sum_{\substack{y>\alpha,\\ y\equiv c(\mathrm{mod}\,2)}} \frac{y}{(y^2+\gamma^2)^2} \leqslant \frac{1}{4(\alpha^2+\gamma^2)}$.
- $\displaystyle\sum_{\substack{y>\alpha,\\ y\equiv c(\mathrm{mod}\,2)}} \frac{1}{(y^2+\gamma^2)^2} \leqslant \frac{\left(\pi-2\arctan(\frac{\alpha}{\gamma})\right)(\alpha^2+\gamma^2)-2\alpha\gamma}{8(\alpha^2+\gamma^2)\gamma^3}$.
- $\displaystyle\sum_{y>\beta} \frac{\pi-2\arctan(\frac{\alpha}{y})}{y^2} \leqslant \frac{\ln(\alpha^2+\beta^2)-2\ln(\beta)}{\alpha} + \frac{\pi-2\arctan(\frac{\alpha}{\beta})}{\beta}$.

With this, we can now bound E_2 easily:

$$\begin{aligned} E_2 &\leqslant \sum_{\|\mathbf{x}\|_\infty>800} \left|\frac{217x_1x_2}{(x_1^2+x_2^2)^3}\right| \leqslant \sum_{\|\mathbf{x}\|_\infty>800} \left|\frac{217}{2(x_1^2+x_2^2)^2}\right| \\ &< \sum_{x_1=1}^{800}\sum_{x_2>800} \frac{434}{(x_1^2+x_2^2)^2} + \sum_{x_1>800}\sum_{x_2\geqslant 0} \frac{434}{(x_1^2+x_2^2)^2} \\ &< \sum_{x_1=1}^{800}\sum_{x_2>800} \frac{434}{x_2^4} + \sum_{x_1>800} \frac{3038}{3\,x_1^3} \\ &\leqslant \frac{434}{3\cdot 800^2} + \frac{1519}{3\cdot 800^2} < 0.0011. \end{aligned} \tag{7.2}$$

Achieving a good bound for E_1 is significantly harder. We divide E_1 into three subsums:

$$\begin{aligned} E_1 < {} & \overbrace{4\sum_{x_1=1}^{800}\sum_{\substack{x_2=801,\\ x_2\equiv x_1(\mathrm{mod}\,2)}}^{\infty} \frac{x_1 H(\mathbf{x}, t_{\max}(\mathbf{x}))}{t_{\max}(\mathbf{x})}}^{\text{see equation (7.4)}} + \overbrace{4\sum_{x_1=801}^{\infty}\sum_{\substack{x_2=1,\\ x_2\equiv x_1(\mathrm{mod}\,2)}}^{800} \frac{x_1 H(\mathbf{x}, t_{\max}(\mathbf{x}))}{t_{\max}(\mathbf{x})}}^{\text{see equation (7.5)}} \\ & + \underbrace{4\sum_{x_1=801}^{\infty}\sum_{\substack{x_2=801,\\ x_2\equiv x_1(\mathrm{mod}\,2)}}^{\infty} \frac{x_1 H(\mathbf{x}, t_{\max}(\mathbf{x}))}{t_{\max}(\mathbf{x})}}_{\text{see equation (7.6)}} < 0.038. \end{aligned} \tag{7.3}$$

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

Now we bound these sums separately as follows:

$$4\sum_{x_1=1}^{800}\sum_{\substack{x_2=801,\\ x_2\equiv x_1 \pmod 2}}^{\infty}\frac{x_1 H(\mathbf{x},t_{\max}(\mathbf{x}))}{t_{\max}(\mathbf{x})} < 68\sum_{x_1=1}^{800} x_1 \sum_{\substack{x_2>800,\\ x_2\equiv x_1 \pmod 2}}\frac{1}{(x_1^2+x_2^2)^2}$$
$$\leqslant \frac{17}{2}\sum_{x_1=1}^{800}\frac{(800^2+x_1^2)\big(\pi-2\arctan(800/x_1)\big)-1600\,x_1}{(x_1^2+800^2)\,x_1^2} < 0.0046, \tag{7.4}$$

$$4\sum_{x_1=801}^{\infty}\sum_{\substack{x_2=0,\\ x_2\equiv x_1 \pmod 2}}^{800}\frac{x_1 H(\mathbf{x},t_{\max}(\mathbf{x}))}{t_{\max}(\mathbf{x})} < 68\sum_{x_2=0}^{800}\sum_{\substack{x_1>800,\\ x_1\equiv x_2 \pmod 2}}\frac{x_1}{(x_1^2+x_2^2)^2}$$
$$\leqslant 17\sum_{x_2=0}^{800}\frac{1}{x_2^2+800^2} < 0.0167, \tag{7.5}$$

$$4\sum_{x_1=801}^{\infty}\sum_{\substack{x_2=801,\\ x_2\equiv x_1 \pmod 2}}^{\infty}\frac{x_1 H(\mathbf{x},t_{\max}(\mathbf{x}))}{t_{\max}(\mathbf{x})} \leqslant 68\sum_{x_1>800} x_1 \sum_{\substack{x_2>800,\\ x_2\equiv x_1 \pmod 2}}\frac{1}{(x_1^2+x_2^2)^2}$$
$$\leqslant \frac{17}{2}\sum_{x_1>800}\frac{\big(\pi-2\arctan(\frac{800}{x_1})\big)(800^2+x_1^2)-1600x_1}{(800^2+x_1^2)x_1^2}$$
$$\leqslant \frac{17}{2}\left(\frac{\ln(2\cdot 800^2)-2\ln(800)}{800}+\frac{\pi/2}{800}-\frac{\ln(2\cdot 801^2)-2\ln(801)}{801}\right)$$
$$< 0.0167. \tag{7.6}$$

Putting this together, we obtain

$$E < 4\cdot 0.038 + 2\cdot 0.0011 < 0.16. \tag{7.7}$$

This upper bound on E is not tight. However, it suffices to prove that the bounds for the single-vertex discrepancy calculated in Section 6 do depend on the rotor sequence. Theorem 4.3 and equations (6.1) and (7.7) yield the following theorem.

Theorem 7.3. *The maximal single-vertex discrepancy between the Propp machine and linear machine is a constant c_2, which depends on the following allowed rotor sequences.*

- *If all vertices have the same circular rotor sequence, then* $7.832 \leqslant c_2 \leqslant 7.985$.
- *If all vertices have the same non-circular rotor sequence, then* $7.286 \leqslant c_2 \leqslant 7.439$.
- *If all vertices may have different rotor sequences, and we assume that each vertex has a rotor sequence leading to a maximal contribution, then* $7.873 \leqslant c_2 \leqslant 8.026$.

8. Aggregating model

Besides the small single-vertex discrepancies examined in the previous sections, the Propp machine and random walk bear striking similarities also in other respects. The historically first research started by Jim Propp studied an aggregating model called Internal Diffusion-Limited Aggregation (IDLA) [4]. In physics this is a well-established model to describe condensation around a source.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

Table 3. $\Delta(n)$ if all rotors are initially set to the left.

Rotor sequence	$(\leftarrow,\uparrow,\rightarrow,\downarrow)$	$(\leftarrow,\uparrow,\downarrow,\rightarrow)$	$(\leftarrow,\rightarrow,\uparrow,\downarrow)$
average $\Delta(n)$ for $2\cdot 10^6 < n \leqslant 3\cdot 10^6$	1.600	0.996	1.810
maximal $\Delta(n)$ for $n \leqslant 3\cdot 10^6$	1.741	1.218	1.967

Table 4. $\Delta(n)$ with random initial directions of the rotors.

Rotor sequence	$(\leftarrow,\uparrow,\rightarrow,\downarrow)$	$(\leftarrow,\uparrow,\downarrow,\rightarrow)$	$(\leftarrow,\rightarrow,\uparrow,\downarrow)$
average $\Delta(n)$ for $2\cdot 10^6 < n \leqslant 3\cdot 10^6$	1.920 ± 0.004	1.782 ± 0.003	1.781 ± 0.003
maximal $\Delta(n)$ for $n \leqslant 3\cdot 10^6$	2.541 ± 0.051	2.351 ± 0.053	2.364 ± 0.067

The process starts with an empty grid. In each round, a particle is inserted at the origin and does a (quasi-)random walk until it occupies the first empty site it reaches. For the random walk, it is well known that the shape of the occupied locations converges to a Euclidean ball [7] in the following sense. Let n be the number of particles and let $\Delta(n)$ denote the difference of the radius of the largest inscribed and the smallest circumscribed circle of an aggregation with n chips. It has been shown by Lawler [6] that the fluctuations around the limiting shape are bounded by $\tilde{O}(n^{1/6})$ with high probability. Moore and Machta [11] observed experimentally that these error terms were even smaller, namely poly-logarithmic.

The analogous model in which the particles do a rotor–router walk instead of a random walk is much less understood. Levine and Peres [8, 9] proved that the shape of occupied locations converges to a Euclidean ball, however, in a weaker sense than before. They showed that the Lebesgue measure of the symmetric difference between the Propp aggregation and an appropriately scaled Euclidean ball centred at the origin is $O(n^{1/3})$. Recently, they improved this and showed that after n particles have been added, the Propp aggregation contains a disc of radius $\sqrt{n/\pi} - O(\log n)$ and is contained in a disc of radius $\sqrt{n/\pi} + O(n^{1/4}\log n)$ [10]. Surprisingly, experimental results indicate much stronger bounds. Kleber [5] computed that for anticlockwise permutations of the rotor directions, $\Delta(3\cdot 10^6) \approx 1.611$ if all rotors initially point to the left. An apparent conjecture is that there is a constant δ such that $\Delta(n) \leqslant \delta$ for all n.

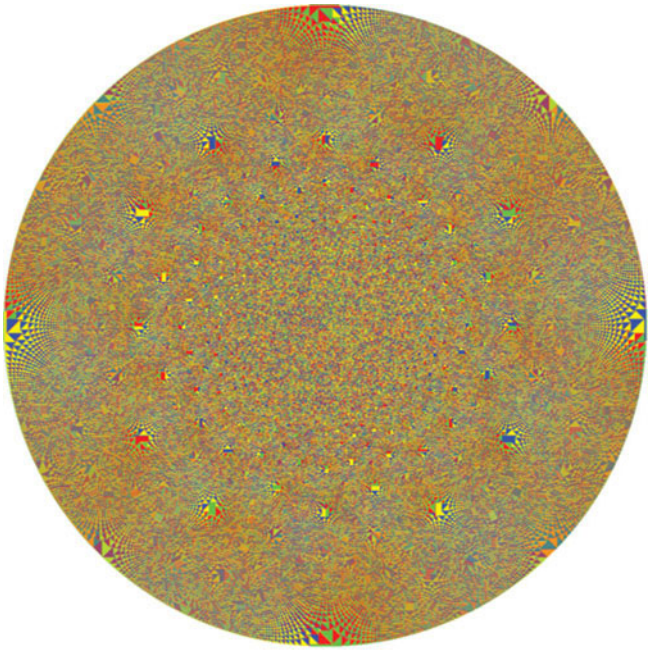
We reran these experiments with different rotor sequences. The aggregations for one million particles are shown in Figure 2. Both aggregations differ not only in the colour patterns but also in the precise value of $\Delta(n)$. If all rotors are initially set to the left, we obtained for $\Delta(n)$ the values shown in Table 3. It is noteworthy that the respective $\Delta(n)$ of both non-circular rotor sequences $(\leftarrow,\uparrow,\downarrow,\rightarrow)$ and $(\leftarrow,\rightarrow,\uparrow,\downarrow)$ differ considerably.

Additionally, we also examined $\Delta(n)$ for random initial rotor directions. This leads to slightly larger Δ-values. Table 4 shows averages and standard deviations of 100 aggregations with random initial directions of the rotors.

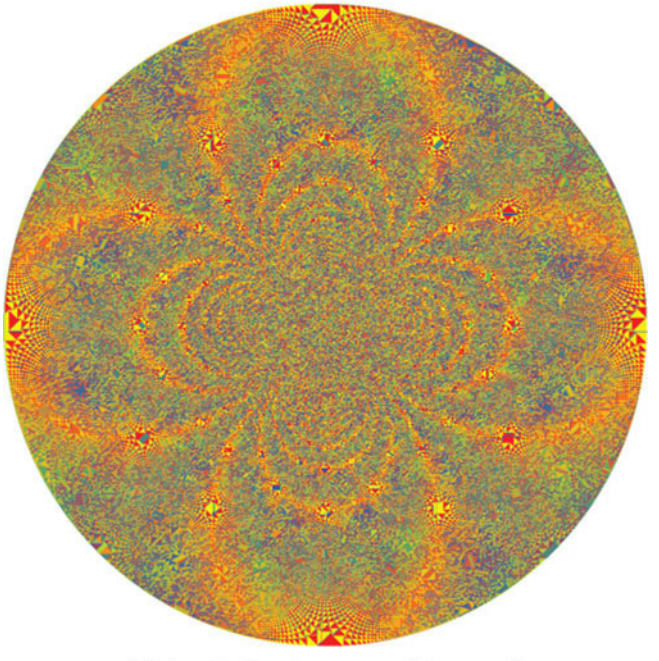
As one might have expected, for random initial rotor directions the two non-circular rotor sequences (columns 1 and 3) are statistically not distinguishable.

The results above again show that different rotor sequences do make a difference. The main open problem, however, remains to show the conjectured constant upper bound for $\Delta(n)$.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press



(a) Anticlockwise rotor sequence $(\uparrow, \leftarrow, \downarrow, \rightarrow)$.



(b) Non-circular rotor sequence $(\uparrow, \leftarrow, \rightarrow, \downarrow)$.

Figure 2. Propp aggregations with one million particles. All rotors initially point to the left. The final rotor directions are denoted by different gray scale shades. (A color version of this figure is available on the journal website – journals.cambridge.org/cpc)

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

9. Conclusion

One way of comparing the Propp machine with a random walk is in terms of the maximal discrepancy that can occur on a single vertex. It has been shown by Cooper and Spencer [1] that for the underlying graph being an infinite grid $\mathbb{Z}^d$, this single-vertex discrepancy can be bounded by a constant c_d independent of the particular initial configuration. For $d = 1$, this constant has been estimated as $c_1 \approx 2.29$ in [2]. Also, the initial configurations leading to a high discrepancy have been described. For $d \geqslant 2$, no such results were known.

In this paper, we analysed the case $d = 2$. We chose the case $d = 2$ out of two considerations. On the one hand, from dimension two on, there is more than one rotor sequence available, which raises the question of whether different rotors sequences make a difference. On the other hand, we restrict ourselves to $d = 2$, because for larger d a nice expression for the probability $H(\mathbf{x}, t)$ that a chip from vertex $\mathbf{x}$ arrives at the origin after t random steps is missing. This probably makes it very hard to find sufficiently sharp estimates for the single-vertex discrepancies.

We were able to give relatively tight estimates for the constants c_2 taking into account different rotor sequences, and obtain several interesting facts about the worst-case initial configurations. The maximal single-vertex discrepancy c_2 satisfies the following. If all vertices have the same circular rotor sequence, then $7.832 \leqslant c_2 \leqslant 7.985$. If all vertices have the same non-circular rotor sequence, then $7.286 \leqslant c_2 \leqslant 7.439$. If all vertices may have different rotor sequences, and we assume that each vertex has a rotor sequence leading to a maximal contribution, then $7.873 \leqslant c_2 \leqslant 8.026$. In particular, we see that non-circular rotor sequences seem to produce smaller discrepancies than circular ones. The gaps between upper and lower bounds stem from the fact that we used a computer to calculate the precise maximal contribution $\mathrm{CON}(\mathbf{x} - \mathbf{y})$ of vertex $\mathbf{x}$ on the discrepancy at $\mathbf{y}$. Hence the lower bounds are the maximal discrepancies obtained from initial configurations such that all vertices $\mathbf{x}$ with $\|\mathbf{x} - \mathbf{y}\|_\infty > 800$ at all times contain only numbers of chips that are divisible by 4.

We also learned that the initial configurations leading to such discrepancies are more complicated than in the one-dimensional case. Recall from [2] that in the one-dimensional case in a worst-case setting, each position needs to have an odd number of chips only once. If we aim at a surplus of chips in the Propp model, these odd chips were always sent towards the position under consideration, otherwise away from it.

In the two-dimensional case, things are more complicated. Here it can be necessary for a position to hold a number of chips not divisible by 4 up to three times. Also, the number of 'relevant' chips (those which cannot be put into piles of four) can be as high as nine. In consequence, it can make sense to send relevant chips in the wrong direction (*e.g.*, away from the position where we aim at a surplus of chips). An example showing this was analysed in Section 6. The reason for such behaviour seems to be that the influences $\mathrm{INF}(\mathbf{x}, A, t)$ of relevant chips sent from $\mathbf{x}$ in direction A at time t are no longer unimodal functions in t (as in the one-dimensional case).

We also briefly considered the IDLA aggregation model. We saw that the surprisingly strong convergence to a Euclidean ball observed in earlier research also holds for non-circular rotor sequences and non-regular initial rotor settings. However, the suspected constant again seems to depend on the rotor sequences, and again, the circular ones seem to behave slightly worse than the non-circular ones.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press

Acknowledgements

We would like to thank Joel Spencer and Jim Propp for several very inspiring discussions.

References

[1] Cooper, J. and Spencer, J. (2006) Simulating a random walk with constant error. *Combin. Probab. Comput.* **15** 815–822.
[2] Cooper, J., Doerr, B., Spencer, J. and Tardos, G. (2007) Deterministic random walks on the integers. *Europ. J. Combin.* **28** 2072–2090.
[3] Cooper, J., Doerr, B., Friedrich, T. and Spencer, J. (2008) Deterministic random walks on regular trees. In *Proc. 19th ACM–SIAM Symposium on Discrete Algorithms* (SODA), pp. 766–772.
[4] Diaconis, P. and Fulton, W. (1990) A growth model, a game, an algebra, Lagrange inversion, and characteristic classes. *Rend. Sem. Mat. Univ. Pol. Torino* **49** 95–119.
[5] Kleber, M. (2005) Goldbug variations. *The Mathematical Intelligencer* **27** 55–63.
[6] Lawler, G. F. (1995) Subdiffusive fluctuations for internal diffusion limited aggregation. *Ann. Probab.* **23** 71–86.
[7] Lawler, G. F., Bramson, M. and Griffeath, D. (1992) Internal diffusion limited aggregation. *Ann. Probab.* **20** 2117–2140.
[8] Levine, L. and Peres, Y. (2008) Spherical asymptotics for the rotor–router model in $\mathbb{Z}^d$. *Indiana Univ. Math. J.*, no. 1, 431–450.
[9] Levine, L. and Peres, Y. (2005) The rotor–router shape is spherical. *The Mathematical Intelligencer* **27** 9–11.
[10] Levine, L. and Peres, Y. (2007) Strong spherical asymptotics for rotor–router aggregation and the divisible sandpile. Submitted *Potential Analysis*; preliminary version available from `arXiv:math/07040688`.
[11] Moore, C. and Machta, J. (2000) Internal diffusion-limited aggregation: Parallel algorithms and complexity. *J. Statist. Phys.* **99** 661–690.
[12] Yap, C. K. (2000) *Fundamental Problems of Algorithmic Algebra*, Oxford University Press.

https://doi.org/10.1017/S0963548308009589 Published online by Cambridge University Press